T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

WEB PORTALLARINDA KULLANICI DAVRANIŞLARININ YERİNDE TESPİTİ VE WEB MADENCİLİĞİNDE KULLANIMI İÇİN YENİLİKÇİ BİR YAKLAŞIM

DOKTORA TEZİ

Özkan CANAY

Bilgisayar ve Bilişim Mühendisliği Anabilim Dalı

OCAK 2024

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

WEB PORTALLARINDA KULLANICI DAVRANIŞLARININ YERİNDE TESPİTİ VE WEB MADENCİLİĞİNDE KULLANIMI İÇİN YENİLİKÇİ BİR YAKLAŞIM

DOKTORA TEZİ

Özkan CANAY

Bilgisayar ve Bilişim Mühendisliği Anabilim Dalı

Tez Danışmanı: Prof. Dr. Ümit KOCABIÇAK

OCAK 2024

Özkan CANAY tarafından hazırlanan “Web portallarında kullanıcı davranışlarının yerinde tespiti ve web madenciliğinde kullanımı için yenilikçi bir yaklaşım” adlı tez çalışması 26.01.2024 tarihinde jüri tarafından oy birliği ile Sakarya Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar ve Bilişim Mühendisliği Anabilim Dalı’nda Doktora tezi olarak kabul edilmiştir.

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Sakarya Üniversitesi Fen Bilimleri Enstitüsü Lisansüstü Eğitim-Öğretim Yönetmeliğine ve Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesine uygun olarak hazırlamış olduğum "Web portallarında kullanıcı davranışlarının yerinde tespiti ve web madenciliğinde kullanımı için yenilikçi bir yaklaşım" başlıklı tezin bana ait, özgün bir çalışma olduğunu; çalışmamın tüm aşamalarında yukarıda belirtilen yönetmelik ve yönergeye uygun davrandığımı, tezin içerdiği yenilik ve sonuçları başka bir yerden almadığımı, tezde kullandığım eserleri usulüne göre kaynak olarak gösterdiğimi, bu tezi başka bir bilim kuruluna akademik amaç ve unvan almak amacıyla vermediğimi ve 20.04.2016 tarihli Resmi Gazete'de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince Sakarya Üniversitesi'nin abonesi olduğu intihal yazılım programı kullanılarak Enstitü tarafından belirlenmiş ölçütlere uygun rapor alındığını, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun ortaya çıkması halinde doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi beyan ederim.

26/01/2024

Özkan CANAY

Ailemin sonsuz desteği ve babam Özcan CANAY'ın anısına, sevgi ve saygıyla...

TEŞEKKÜR

Tez çalışmam boyunca beni motive eden ve yardımlarını esirgemeyen danışmanım, saygıdeğer hocam Prof. Dr. Ümit KOCABIÇAK'a, tezin olgunlaşmasına katkı sunan jüri üyesi kıymetli hocalarıma ve bu uzun süreçte bilgi birikimlerini aktararak bana rehberlik eden değerli dostlarıma teşekkür ederim.

Ayrıca, bu çalışmayı lisansüstü tez projesi kapsamında destekleyen Sakarya Üniversitesi Bilimsel Araştırma Projeleri (BAP) Komisyonu Başkanlığına (Proje No: 2010-50-02-024) teşekkürlerimi sunarım.

Özkan CANAY

İÇİNDEKİLER

Sayfa

KISALTMALAR

ABİS : Akademik Bilgi Sistemi
ACID : Atomicity, Consistency, Isolation, Durability
AJAX : Asynchronous JavaScript and XML
API : Application Programming Interface
APM : Application Performance Monitoring
BDMS : Business Decision Making System
CAWAL : Combined Analytics and Web Application Log
CAWIS : Campus Automation Web Information System
CCPA : California Consumer Privacy Act
CLF : Common Log Format
CLS : Cumulative Layout Shift
CPU : Central Processing Unit
CSV : Comma-Separated Values
DB : Database
DCL : DOM Content Loaded
DDL : Data Definition Language
DM : Data Mining
DML : Data Manipulation Language
DOM : Document Object Model
DQL : Data Query Language
DW/DWH : Data Warehouse
ECLF : Extended Common Log Format
ERD : Entity-Relationship Diagram
ETL : Extract, Transform, Load
FCP : First Contentful Paint
GA : Google Analytics
GDPR : General Data Protection Regulation
HTTP : Hypertext Transfer Protocol
I/O : Input/Output
ID : Identity
IoT : Internet of Things

IP : Internet Protocol
JS : JavaScript
KDD : Knowledge Discovery in Databases
KPI : Key Performance Indicator
LCP : Largest Contentful Paint
MDA : Multidimensional Array
ML : Machine Learning
N/A : Not Available
NAS : Network-Attached Storage
NoSQL : Not Only SQL
OBİS : Öğrenci Bilgi Sistemi
OLAP : Online Analytical Processing
OLTP : Online Transaction Processing
OS : Operating System
OWA : Open Web Analytics
PBİS : Personel Bilgi Sistemi
PIPEDA : Personal Information Protection and Electronic Documents Act
PpS : Pages per Session
PWA : Progressive Web Application
RAM : Random Access Memory
RDBMS : Relational Database Management System
SaaS : Software as a Service
SAÜ : Sakarya Üniversitesi
SPA : Single Page Application
SQL : Structured Query Language
TCP : Transmission Control Protocol
TTI : Time to Interactive
UML : Unified Modeling Language
URL : Uniform Resource Locator
WA : Web Analytics
WCM : Web Content Mining
WM : Web Mining
WSM : Web Structure Mining
WUM : Web Usage Mining

SİMGELER

A	: Kullanıcı eylemleri
$A(\zeta)$	: Belirli bir dönemdeki ham verilerin toplamı
B	: Web kullanıcılarının tüm tarayıcılarının bir kümesi
B_f	: Tarayıcı parmak izi belirsizlikleri durumu
b_i	: Her bir web kullanıcısının tarayıcısı
c	: HTTP durum kodu
C	: IP veya tarayıcı gibi bağlamsal veriler
C_c	: Bağlam değişikliği dönemleri
C_d	: Çerezlerin silinmesi durumu
CL	: Temizlenmiş ve ayıklanmış günlük dosyası
cl_i	: Temizlenmiş ve ayıklanmış günlük dosyası girdisi
d	: GET/POST gibi istek yöntemi
D	: IP adresleri veya çerezler gibi veri noktaları
D_m	: Birden fazla cihazdan erişim durumu
D_o	: Veri tutarsızlıkları
D_{oltp}	: OLTP veritabanından alınan ham veriler
D_s	: Bir oturumun bekleme süresi
$DW_{önc}$	: Veri ambarının önceki durumu
DW_{yeni}	: Veri ambarının yeni durumu
e_i	: Kullanıcının ziyareti sırasında gerçekleştirdiği tüm etkinlikler
Eyl	: Olası eylemler
f	: Bir kullanıcıyı benzersiz bir şekilde tanımlayan görev fonksiyonu
F_b	: Sonuçlar arasındaki sapmayı gösteren geri bildirim değeri
g	: Etkinlikleri bağlamsal oturumlar olarak sınıflandıran fonksiyon
I	: Kullanıcıları birden fazla etkileşim üzerinden tanımayan fonksiyon
I_a	: Kullanıcı hareketsizliği
I_d	: Dinamik IP tahsisi
IP	: Web sitesine erişen kullanıcıların tüm IP adreslerinin bir kümesi
ip_i	: Web sitesine erişen her bir kullanıcının IP adresi
K	: Web sitesi dışındaki bağlantıların bir kümesi

k_i : Web sitesi dışındaki her bir bağlantı

L : Bir web sunucusu günlük (log) dosyası

l_i : ECLF yapısındaki bir günlük (log) girdisi

m : Sistemdeki sunucu sayısı

M : Çeşitli cihazları veya tarayıcıları içerebilen etkileşim yöntemleri

ms : Milisaniye [Birim]

n_i : Sunucuya gelen taleplerin ölçülebilir alt kümesi

O : Zaman karmaşıklığı

$P(A)$: İstatistiksel fonksiyonlarla işlenen birleştirilmiş veriler

p_i : Belirli bir oturumdaki her bir sayfa isteği

p_n : Belirli bir oturumdaki toplam sayfa isteği

p_{ps} : Oturum başına sayfa görüntüleme sayısı

P_s : Belirli bir oturumdaki sayfa istekleri kümesi

R : Web sitesinin tüm kaynaklarının bir kümesi

Rap : İşlenmiş verilere dayalı kapsamlı rapor oluşturma

ref_i : Her bir yönlendiren URL adresi

r_i : Web sitesinin her bir kaynağı

R_p : Bir sayfa isteğinin istek zamanı

R_{wa} : Web analitiği için özel dönüşüm kural seti

s : Aktarılan bayt boyutu

S : Oturum tanımlama fonksiyonu

$S_{dış}$: Dış kaynak verileri

$S_{dön}$: Dönüştürülmüş web analitiği verileri

$S_{iç}$: Temizlenmiş OLTP verileriyle dış veri kaynaklarının entegrasyonu

S_{oltp} : OLTP yapıları için çıkarma fonksiyonu sonucu elde edilen veriler

S_{temiz} : Temiz veri

t : Zaman damgası olarak istek zamanı

T : Etkinlikler arasındaki zaman aralıkları

Tav : Raporu kullanarak eyleme geçirilebilir tavsiyeler türetme

T_c : Temizleme dönüştürme fonksiyonu

T_o : Oturum zaman aşımı

U : Web sitesine erişen tüm kullanıcıların kümesi

u_i : Web sitesine erişen her bir kullanıcı

v : HTTP sürümü

V : Web sitesi kullanıcılarının ziyaretleri

vs_i	: Her bir kullanıcının web sitesini ziyareti
$\sim$	: Yaklaşık değer
δ	: Beklenen ve gerçekleşen sonuçlar arasındaki sapmayı ölçen fonksiyon
$\mathcal{D}_i$	: OLTP oturum tablosundaki işlenecek gün
$\mathcal{D}_{tüm}$	: OLTP oturum tablosunda birikmiş tüm günler
ϕ	: İşlenmiş verilere dayalı kapsamlı rapor oluşturma fonksiyonu
ψ	: Eyleme geçirilebilir tavsiyeler türetme fonksiyonu
ω	: Tavsiyeler ve eylemlerle ilgili optimizasyon fonksiyonu
Ω	: Tavsiyeler ve eylemlerle ilgili optimizasyon değeri
ζ	: Çeşitli metriklerden oluşan ham veriler

TABLO LİSTESİ

Sayfa

ŞEKİL LİSTESİ

Sayfa

WEB PORTALLARINDA KULLANICI DAVRANIŞLARININ YERİNDE TESPİTİ VE WEB MADENCİLİĞİNDE KULLANIMI İÇİN YENİLİKÇİ BİR YAKLAŞIM

ÖZET

Web portalları ve kurumsal web uygulamalarından elde edilen etkileşim verilerinin analizi, kullanıcı davranışlarının anlaşılması açısından büyük önem taşımaktadır. Temel amacı sunucuya yapılan erişimleri kaydetmek olan sunucu günlükleri, kullanıcı etkileşimlerini açıklayabilmek için detaylı veri sağlayamazlar. Bunun yanında, uygulama günlüğü bakış açısına sahip olmayan güncel web analitik araçları, özellikle kurumsal web portallarında kullanıcıların bireysel olarak hizmetleri kullanma biçimlerini ve oturumları boyunca yaptıkları işlemleri takip etme ve anlamlandırma açısından zayıftırlar. Son yıllarda artan veri gizliliği ve egemenliği endişeleri de bulut tabanlı üçüncü parti araçların kullanımında çekincelere yol açmaktadır. Diğer taraftan, verinin değerini bilen ve ondan sonuna kadar yararlanmayı amaçlayan kuruluşlar için üzerinde web kullanım madenciliği teknikleri uygulayabilecekleri verilere sahip olmak son derece önemlidir.

Bu tezde, web portalları ve büyük ölçekli kurumsal web uygulamaları için uygulama günlükleri ile web analitiğini bütüncül bir yapıda birleştiren CAWAL adında bir model ve onun pratik bir uygulaması olan yazılım çerçevesi önerilmiştir. Özellikle çok sunuculu ortamlar için tasarlanan model, alt alanlar arası oturum takibi, gerçek zamanlı analiz, yüksek doğruluğa sahip veri toplama, analitik çıkarımı ve veri depolama yetenekleriyle web analitiğinde alternatif bir çözüm sunmaktadır. Yalın mimarisine karşın zengin veri seti sunan çerçeve, gerçek bir web portalı üzerinde uygulanarak etkinliği sınanmıştır. Uygulama sonucu elde edilen bulgular çerçevenin farklı kullanıcı niteliklerini ve gezinme davranışlarını çok boyutlu ve kapsamlı biçimde analiz edebildiğini göstermiştir. Gerçekleştirilen yük testlerinde CAWAL, yaygın kullanılan açık kaynaklı web analitik araçları OWA ve Matomo'ya kıyasla çok daha düşük yanıt süreleri sunarak performansını kanıtlamıştır.

Veri ambarında depolanan yapılandırılmış ilişkisel verilerin web kullanım madenciliğinde veri kaynağı olarak kullanımı ise bu alana özgün ve yenilikçi bir yaklaşım getirmiştir. Yöntem, WUM'da sunucu günlüklerinin kullanımını ve veri ön işleme aşamasını ortadan kaldırarak süreci önemli ölçüde kısaltmış; aynı zamanda zenginleştirilmiş oturum ve sayfa gösterimi verileri ile madencilik faaliyetlerinin başarımını arttırmıştır. CAWAL'ın gelişmiş veri toplama ve analiz kabiliyetleri multidisipliner araştırmaları destekleyerek gelecekteki çalışmalarda yeni ve farklı uygulama senaryolarının ortaya çıkmasına katkı sağlayacaktır.

AN INNOVATIVE APPROACH FOR ON-PREMISES DETECTION OF USER BEHAVIORS ON WEB PORTALS AND ITS USE IN WEB MINING

SUMMARY

The increasing use of web portals, corporate web applications, and online platforms necessitates new approaches to business strategies and academic research. This situation has made it imperative to analyze the application data and better understand user behavior. Web analytics and web usage mining (WUM) play an essential role in corporate decision-making processes based on the analysis of data obtained to understand user behavior and integrate this information into business strategies.

Web usage mining uses server logs as a traditional data source. While these logs provide valuable data about user interactions, analyzing and processing this data often requires complex and time-consuming processes. Existing tools often provide superficial analysis, not allowing for a deep understanding of user interactions. This shortcoming has led to the need for the development of new and more effective methods in the field of web usage mining and analytics. However, today's growing concerns about data privacy and security have also led organizations to seek greater control and ownership over user data. Cloud-based web analytics tools, in particular, have limitations in terms of security and data sovereignty, while client-side user monitoring tools face challenges such as data accuracy and lack of server-side metrics. For all these reasons, there is a need for an approach to overcome the challenges faced in web analytics and usage mining from a different perspective.

This study proposes an innovative model to detect user behavior on web portals and use this information in web mining processes. It also aims to test its effectiveness by developing a software framework that practically implements the model. This model, which we call CAWAL (Combined Analytics and Web Application Log), represents a holistic approach that integrates application logs and web analytics by overcoming the limitations of traditional web analytics tools. This integration aims to analyze the data obtained from web portals more effectively and to use the information obtained from these analyses in strategic decision-making processes. This innovative approach fills an essential gap in the web analytics and usage mining literature, expanding the application and research possibilities in these fields and opening new research avenues. The proposed model is expected to solve challenges that traditional methods cannot overcome, especially in complex and multi-server web environments.

Another thesis objective is to apply the CAWAL model to a genuine system, evaluate its performance, and show that the data obtained can be effectively used in web mining. The development of the model will enable a more detailed analysis of user behavior for corporate web applications and portals and contribute to strategic decision-making processes. The central hypothesis of this study is that the application of the proposed model and the framework developed based on this model leads to significant improvements in the fields of web analytics and web mining and provides a more effective and comprehensive data collection and analysis process by overcoming the limitations of traditional methods used in these fields.

The academic importance of the study lies in the fact that it fills crucial gaps in the fields of web analytics and web usage mining and brings an innovative approach to methodologies in these disciplines. The model's integration with application logs and its success in session identification, user identification, and data accuracy has primarily eliminated the preprocessing stage in WUM and made the analysis processes more efficient and effective. This approach offers new opportunities in user behavior analysis, enabling web portals and enterprise web applications to understand their users better and effectively integrate these insights into their business strategies. In practical terms, the CAWAL framework has provided new ways of understanding user interactions in web portal management and using this information in strategic decision-making processes. Its capacity to effectively track various user engagements, especially in multi-server environments and web portals with different (sub)domains, provides a significant advantage in enterprise-level applications. In addition, the detailed data collection and analysis capabilities offered by the model will enable organizations to be more effective in their strategic decision-making processes and help them better target their business strategies and web content using this data.

The CAWAL model is designed to cover various platforms and interaction types. The model uses a server-side logging API for data collection and processing. The use of a server-side API enables both detailed analysis of user behavior and the improvement of application performance and security. The model was implemented on a university web portal, where user interactions and portal performance were analyzed extensively. These analyses included various metrics and visualization techniques to understand user behavior, preferences, and interaction patterns. Furthermore, the data collected during the implementation of the model was analyzed along various dimensions such as user types, visit durations, and preferred services. This detailed data analysis plays a critical role in understanding how users navigate the web portal, which content and services they are more interested in, and how to improve the user experience. Thus, the methodology makes valuable contributions to web analytics and web usage mining, both theoretically and practically.

Integration of the CAWAL framework into an enterprise web application and performance tests have demonstrated the model's capacity to work effectively. Moreover, when compared to well-known open-source tools such as Open Web Analytics (OWA) and Matomo, CAWAL outperformed OWA with approximately 24% lower response time and Matomo with 85% lower response time. These results highlight CAWAL not only as a tool that provides accurate and detailed information, but also as a high-performance and scalable solution. The analytical findings of the framework have provided in-depth insights into user behavior and web portal interactions. Analyses of user types, demographics, interaction times, and preferred content have provided valuable insights into optimizing the web portal's design and functionality according to user needs and preferences. Revealing behavioral differences between various user groups, which parts and services of the portal attract more attention, and how users use the portal provides crucial strategic information to improve the user experience of the web portal. Furthermore, CAWAL's superior performance is characterized by user tracking, session management, and data accuracy. Thanks to the model's detailed data analysis and application log integration, software bugs and performance issues were detected more effectively, which contributed to improving the overall quality of web services and user satisfaction.

In the future work section of this dissertation, further research on the applicability and impact of the CAWAL model is recommended. Future work is expected to examine

how the model can be applied to various user interaction scenarios in different web portals and further develop the model's data collection and analysis capabilities. Studies can be conducted on how the model can be applied in different industries and sectors, for example, e-commerce, healthcare, and public services. Such work will expand the model's potential and provide a broader range of applications in web analytics and web usage mining. In addition, detailed studies of the model's current limitations and challenges would improve its applicability and effectiveness. For example, research on the model's scalability, flexibility, and adaptability to various technological infrastructures can play an essential role in its adaptation to a wide range of applications. These studies could provide practical solutions to address technical and operational challenges. In addition, studies on the long-term effects and sustainability of the model are also critical. Further model development will enable it to rapidly adapt to technological innovations and changing user needs, making it a practical and applicable solution in web analytics and usage mining.

1. GİRİŞ

Web teknolojilerinin hızlı evrimi ve dijital veri akışındaki artış, iş dünyasını ve akademik çevreleri önemli ölçüde etkilemektedir. Bu değişim, bilgi teknolojilerinin ve web tabanlı uygulamaların kullanımını daha da kritik hale getirmektedir. Özellikle web portalları, kurumsal uygulamalar ve çevrimiçi platformlara yönelik artan talep, yeni iş stratejilerinin ve akademik metodolojilerin geliştirilmesini zorunlu kılmaktadır. Bu bağlamda, uygulamalardan toplanan verilerin detaylı analizi ve kullanıcı davranışlarının kapsamlı bir şekilde incelenmesi büyük önem taşır. Web analitiği ve web kullanım madenciliği gibi alanlar, kullanıcı etkileşimlerini derinlemesine anlamayı ve elde edilen verileri stratejik karar verme süreçlerine etkin bir şekilde entegre etmeyi amaçlar.

Geleneksel web kullanım madenciliğinde sunucu günlükleri temel bir veri kaynağı olarak kullanılır. Bu günlükler, kullanıcı etkileşimleri hakkında değerli veriler sağlasa da bu verilerin analizi ve işlenmesi karmaşık ve zaman alıcı süreçler gerektirir. Mevcut araçlar, genellikle yüzeysel analizler sunar ve kullanıcı etkileşimlerini derinlemesine anlamaya izin vermez. Bu eksiklikler, web kullanım madenciliği ve analitiği alanında daha etkili yöntemlerin geliştirilmesi gerekliliğini ortaya koymaktadır. Kuruluşların verileri daha etkin bir şekilde kullanma isteğiyle birlikte, günümüzde artan veri gizliliği ve güvenlik endişeleri, bu kuruluşların kullanıcı verileri üzerinde daha fazla kontrol ve mülkiyet talep etmelerine neden olmuştur. Bulut tabanlı web analiz araçları güvenlik ve veri egemenliği konularında sınırlamalar içerirken, istemci tarafı kullanıcı izleme araçları veri doğruluğu ve sunucu tarafı metriklerinin eksikliği gibi zorluklarla karşı karşıyadır.

Tüm bu nedenlerden dolayı web analitiği ve kullanım madenciliği alanlarındaki zorlukları aşmak için farklı bir yaklaşıma ihtiyaç duyulmaktadır. Bu tezde, her iki disiplindeki yaklaşımları genişletmek ve web portallarından elde edilen verilerin analizini iyileştirmek amacıyla CAWAL (Combined Analytics and Web Application Log) adında yenilikçi bir model ve bu model temelinde geliştirilen bir yazılım çerçevesi önerilmektedir. CAWAL modeli, geleneksel web analitiği araçlarının

sınırlarını aşarak, uygulama günlüğü ile web analitiği fonksiyonlarını bütünleştiren bir yaklaşım ortaya koymaktadır. Bu entegrasyon, kullanıcı etkileşimlerinin daha derinlemesine anlaşılmasına ve bu bilgilerin stratejik karar alma süreçlerine etkili bir şekilde entegre edilmesine olanak sağlamaktadır.

Çalışmanın akademik önemi, web analitiği ve web kullanım madenciliği alanlarına yönelik önemli boşlukları doldurması ve bu disiplinlerdeki metodolojilere yenilikçi bir bakış getirmesindedir. Modelin uygulama günlüğü fonksiyonuyla veri doğruluğu, oturum ve kullanıcı tanımlama konularındaki başarısı, WUM'daki ön işleme aşamasını büyük ölçüde ortadan kaldırmış ve analiz süreçlerini daha ekin ve verimli hale getirmiştir. Bu yenilikçi yaklaşım, kullanıcı davranış analizlerinde yeni fırsatlar sunarak, web portallarının ve kurumsal web uygulamalarının kullanıcılarını daha iyi anlamalarına ve bu anlayışları iş stratejilerine etkili bir şekilde entegre etmelerine olanak tanımaktadır. Pratik açıdan ise CAWAL çerçevesi, web portalı yönetiminde kullanıcı etkileşimlerini anlama ve bu bilgileri stratejik karar alma süreçlerinde kullanma konusunda yeni yöntemler sunmuştur. Özellikle çok sunuculu ortamlarda ve farklı (alt)alan adlarına sahip web portallarında çeşitli kullanıcı katılımlarını etkili bir şekilde izleme kapasitesine sahip olması, kurumsal düzeydeki uygulamalarda önemli bir avantaj sağlamaktadır. Ayrıca, modelin sağladığı detaylı veri toplama ve analiz yetenekleri, kurumların stratejik karar alma süreçlerinde daha etkili olmalarını sağlayacak ve bu verilerin kullanımıyla iş stratejilerini ve web içeriğini daha iyi hedeflemelerine yardımcı olacaktır.

1.1. Tezin Amacı

Tezin birincil amacı, web analitiğine yenilik getirecek sunucu tarafında çalışan uygulama tabanlı bir model önererek kullanıcı davranışlarının daha etkin şekilde analiz edilmesini ve elde edilen bilgilerin stratejik karar alma süreçlerine daha fazla katkıda bulunmasını sağlamaktır. Tezin ikincil amacı ise önerilen modelin geleneksel veri kaynağı olan sunucu günlüklerinin yerine geçerek web kullanım madenciliği faaliyetlerinde veri doğruluğunu artırması, veri ön işleme gereksinimini azaltması ve analiz süreçlerini hızlandırması sayesinde madencilik faaliyetlerinin etkinliğini ve başarımını yükseltme potansiyelini göstermektir.

1.2. Tezin Kapsamı

Tezin kapsamı, mevcut yöntemlerin sınırlamalarını ve zayıflıklarını irdelemenin yanında web analitiği ve web kullanım madenciliği süreçlerinde yenilikçi bir yaklaşım sunan CAWAL modelinin geliştirilmesi, uygulanması ve değerlendirilmesini içerir. Çalışma, özellikle web portallarında ve kurumsal web uygulamalarında kullanıcı davranışlarının detaylı analizine odaklanmaktadır. Bu bağlamda, modelin sunucu tarafında çalışan uygulama tabanlı veri toplama yaklaşımı ve bu yaklaşımın web analitiği ile entegrasyonu incelenmektedir. Tez kapsamında ayrıca model kapsamında geliştirilen çerçeve yazılımının modelinin temel özellikleri, performansı, analitik süreçleri ve web analitiği ile WUM alanlarına katkıları irdelenecektir.

Önerilen modelin geleneksel veri kaynakları olan sunucu günlüklerine alternatif olarak nasıl kullanılabileceği, bu sayede web kullanım madenciliği süreçlerinde veri doğruluğunu nasıl artırabileceği, veri ön işleme ihtiyacını nasıl azaltabileceği ve analiz süreçlerini nasıl hızlandırabileceği üzerinde durulacaktır. Ayrıca, modelin gerçek bir kurumsal web portalında uygulanması sonucu elde edilen bulgular, analizler ve sonuçlar da tez kapsamında değerlendirilecektir. Bu uygulama, modelin pratikteki etkinliğini ve web kullanım madenciliği faaliyetlerine katkısını göstermek için kullanılacaktır. Tez, CAWAL modelinin teorik temelleri ile başlayıp modelin uygulama aşamalarını, kullanılan teknolojileri ve elde edilen sonuçları kapsayacak şekilde geniş bir perspektif sunmaktadır. Bu çalışma, modelin sadece teorik bir çerçeve olarak kalmayıp gerçek dünya verileri ile test edildiğini ve değerlendirildiğini göstermektedir. Tez ayrıca, önerilen modelin web analitiği ve web kullanım madenciliği alanlarındaki mevcut yaklaşımları nasıl genişlettiğini ve bu alanlarda araştırma ve uygulama olanaklarını nasıl artırdığını ortaya koymayı amaçlamaktadır.

1.3. Literatür Araştırması

Web analitiği, veri yorumlama, teknolojik uygulamalar ve kullanıcı davranışları çalışmalarını kapsayan karmaşık ve çok boyutlu bir disiplin olarak kabul edilmektedir. Fundingsland ve ark. (2022) çalışmalarında web analitiğinin bilgisayar bilimleri alanındaki teknik arka planının yanı sıra işletme, pazarlama, sosyal bilimler, sağlık, finans, turizm, lojistik gibi birçok farklı alanda uygulanabilirliğini vurgulamışlardır. Ahadi ve ark. (2022) ile Amalina ve ark. (2020) da bu çok yönlülüğü destekleyen çalışmalar yapmışlardır. Harika ve Sudha (2019) ise çalışmalarında web üzerindeki

büyük veri kümelerini analiz etmek, kullanıcı davranışlarını anlamak ve tahminlerde bulunmak için bilgisayar bilimleri yöntemleri ve algoritmalarının kullanımını vurgulamışlardır.

Yazılım geliştirme tarihi kadar eski olan günlük (log) tutma ihtiyacı ile ilgili literatürde birçok çalışma bulunmaktadır. Gholamian ve Ward (2021) yazılımlarda kullanılan genel log yapılarını incelemiş ve birçok log uygulamasını karşılaştırmalı olarak analiz etmiştir. Srivastava ve ark. (2019), sunulan sınırlı erişimli verileri içeren web logların literatürdeki birçok çalışmaya ilham kaynağı olduğunu ortaya koymuştur. Paredes ve Bolanio (2020) bu tür kayıtlardan bilgi çıkarmaya yönelik bazı çalışmaları derlemiştir. Bu çalışmalarda büyük ölçüde başarılı sonuçlar elde edildiği ifade edilse de, oldukça zahmetli olan süreçler çevrimdışı olarak yapılmakta ve her zaman doğru sonuçlar vermemektedir. Abdalla ve ark. (2016) ile Deshpande ve ark. (2018) bu konuda yaşanan zorluklara değinmiştir. Özellikle mevcut algoritmalardan kaynaklanan modelleme sorunları önceki araştırmalarda incelenmiş ve daha verimli ön işleme yöntemlerinin geliştirilmesi için çeşitli önerilerde bulunulmuştur. Kumar ve Ogunmola (2020) araştırmalarında temel web analitik araçlarının karşılaştırmalı bir analizini sunarak analitik uygulamalarının karşılaştığı zorlukları açıklamıştır. Kundu ve Garg (2017) çalışmalarında weblog analiz uygulamalarını derlemiş ve karşılaştırmışlardır. Web analitiği alanında, Knight-Davis (2017) AWStats kullanarak genel erişim kayıtlarının analiz edilmesi üzerine çalışmıştır.

Son dönemdeki web analitiği çalışmaları, farklı metodolojileri ve konuları ele almıştır. Alby (2023) çalışmasında Google Analytics'in web analitiği sektöründeki önemli etkisine odaklanmış, Kumar ve ark (2022) ise e-ticaret web sitelerinin kullanılabilirliğinin değerlendirilmesi için Google Analytics kullanımını incelemiştir. Jansen ve ark. (2022) ise Google Analytics ve Similarweb gibi iki yaygın analiz yaklaşımını karşılaştırmışlardır. Cheng ve Chen (2022) çalışmalarında devlet web sitelerinde kullanıcı davranışlarını anlamak için web analizleri ve derinlemesine görüşmeleri bir arada kullanmıştır. Hasan (2022), kullanıcı davranışını analiz etmek için "Web Analyzer" adlı bir yazılım aracını tanıtmış, De-Groot ve ark. (2022) web analitiği ve sosyal bilim yöntemlerini bir arada kullanarak vatandaş bilimini araştırmak için bir çerçeve geliştirmiştir. Önder ve Berbekova (2021) ise çalışmalarında Avrupa Destinasyon Yönetim Kuruluşlarında web analitiği araçlarının benimsenmesini analiz etmişlerdir.

Alanda en çok tercih edilen açık kaynaklı araçlar olan Matomo (eski adıyla Piwik) ve Open Web Analytics (OWA)'nın veri modellerini, veri toplama ve depolama yöntemlerini ve çalışma prensiplerini açıklayan bir çalışmaya literatürde rastlanmamıştır. Öte yandan, içerik sağlayıcılar, sunucu yöneticileri için ilgi çekici olan standart sistem istatistiklerinin yanı sıra, eğilimleri ve nedenselliği anlamak için esas olarak ziyaretçilerin davranışlarıyla ilgilenmektedir (Canay ve Kocabıçak, 2016). Bu tür gereksinimlerin çoğu basit istatistiksel araçlarla karşılanamaz ve daha sofistike tekniklere ihtiyaç duyar.

Web analitiğinin yanında, web kullanım madenciliği alanı ve özellikle veri ön işlemede sürecindeki zorluklara yönelik son yıllarda önemli çalışmalara imza atılmıştır. Srivastava ve ark. (2019) çalışmalarında web sunucusu günlüğünün veri önişlemesinde uygulanan teknikleri özetlemişlerdir. Svec ve ark. (2020) web kullanım madenciliğinde veri ön işlemenin önemini ve hataların veri analizine etkisini açıklamıştır. Sukumar ve ark. (2016) literatürdeki ön işleme tekniklerini aktararak bu çalışmalar için örnek sözde kodlar sunmuştur. Roy ve Rao (2020), weblog verilerinin kalitesini ve etkinliğini artırmak için detaylı bir ön işleme sürecini incelemişlerdir. Bunun yanı sıra, Ibrahim ve Obaid (2021) tarafından ziyaretçi davranışlarını analiz etme ve ön işleme için iki algoritma sunulmuştur. Jain ve Kashyap (2021) web günlüğü verilerinin ön işlemesine odaklanan bir algoritma önermiş, Srivastava ve ark. (2018) ise web kullanım madenciliğinde MapReduce tabanlı bir kullanıcı tanımlama algoritması geliştirmişlerdir.

Ön işlemedeki kullanıcı ve oturum tanımlama süreçlerine odaklanan çalışmalarda, Bayir ve Toroslu (2022) web kullanıcı oturumlarını oluşturmak için graf tabanlı bir yöntem sunmuş, Munk ve Benko (2018) ise entropiyi kullanarak oturum tanımlama sürecini iyileştirmiştir. Ayrıca, Srivastava ve ark. (2021), web kullanım madenciliğinde robot tespiti yaklaşımları ile birlikte MapReduce tabanlı paralel veri ön işleme algoritmasının performansını değerlendirmişlerdir. Fatima ve ark. (2016), WUM amacıyla sunucu loglarından oturum tanımlama konusunda yapılan çalışmaları ortaya koymuşlar, Deshpande ve ark. (2018) ise çalışmalarında kullanıcı tanımlama konusunda yapılan çalışmaları özetlemişlerdir.

Web kullanım madenciliği ve web analitiği alanındaki bu önemli çalışmaların yanı sıra, sosyal medya veri analitiği ve log verilerinin işlenmesi konuları da araştırmacıların odak noktasında yer almaktadır. Bu bağlamda, iş karar verme

süreçlerinde sosyal medya verilerinin potansiyelinden yararlanma ve çapraz alan log veri analizleri gibi yeni yaklaşımlar, literatürde dikkat çekici çalışmalara konu olmuştur. Yang ve ark. (2022) sosyal medya veri analitiğinin iş karar verme sürecindeki potansiyelini araştırmış, özellikle İş Karar Verme Sistemi (BDMS) çerçevesini tanıtarak, iş geliştirme stratejilerinde sosyal platformlardan elde edilen içgörülerin kullanımının önemini vurgulamışlardır. Tao ve ark. (2020) tarafından yapılan çalışmada, farklı alt sistemlerdeki günlük verilerin çapraz alan log veri birleşimi ile kullanıcı davranışlarının analizi üzerine odaklanılmıştır. Bu çalışma, log verilerinin birleştirilmesini içeren bir yaklaşım sunmaktadır. Ayrıca, Ehikioya ve Lu (2019) tarafından sunulan ve tam yolları temel alan bir yol analizi modeli, etkili e-ticaret işlemleri için önemli bir katkı sağlamaktadır.

Teknik hususların yanı sıra kullanıcı takibi ve izleme ile ilgili etik ve yasal boyutlar yine son zamanlarda önem kazanan konular arasındadır. Utz ve ark. (2023) web analitiği alanındaki etik ve yasal çerçevenin, özellikle kullanıcı rızası, veri şeffaflığı ve çıkış seçenekleri için geliştirilen mekanizmalar açısından kritik olduğunu belirtmiştir. Regueiro ve ark. (2021), GDPR, CCPA ve PIPEDA gibi düzenleyici çerçevelere uyumun, veri yönetimi ve etik uyumluluğu geliştirmede temel olduğunu vurgulamıştır. Demirkan ve Delen (2013) çalışmalarında analitik tasarım seçenekleri ve reklam engelleyicilerin neden olduğu veri tutarsızlıkları ile gerçek zamanlı analitik ihtiyacının sunduğu hesaplama zorluklarına değinerek bir web analitiği sisteminin teknolojik, etik ve yasal düşünceler arasında etkili bir denge kurmasını gerektirdiğini vurgulamışlardır.

Samarasinghe ve ark. (2022) çalışmalarında, bulut tabanlı web analitik araçları kullanan kuruluşlar için en önemli sorunun, verilerin üçüncü taraf sunucularda saklanması nedeniyle artan gizlilik endişeleri olduğunu belirtmişlerdir. Everest-Phillips'in 2019 yılında gerçekleştirdiği çalışma, ülkelerin ve kuruluşların, özellikle uluslararası şirketlerle kullanıcı verileri üzerindeki kontrol ve sahiplik konusunda yaşadıkları tereddütleri ortaya koymuştur. Binns (2017), veri gizliliği yasalarının ve düzenlemelerinin hem kuruluşları hem de bireyleri veri ihlallerine karşı koruma amacı taşıdığını ve uyum sağlamanın önemli bir kurumsal uzmanlık gerektirdiğini vurgulamıştır. Aartsen ve ark. (2020) ise araştırmalarında, web kullanım madenciliği alanında makale yayımlayan yazarların, WUM kullanarak anormal trafik tespiti ve

web güvenliği için çeşitli yöntemlerin yanı sıra veri gizliliğini bu alanın önemli gelecek araştırma konuları olarak gördüklerini ortaya koymaktadır.

Akademik kütüphanelerde web analizi için Matomo'nun potansiyel kullanımını araştıran Quintel ve Wilson (2020) çalışması, Google Analytics gibi geleneksel araçlara bağlı kullanıcı gizliliği endişelerine dikkat çekerek, bu sorunları hafifletmek için açık kaynaklı alternatiflerin kullanımını önermiştir. Bu bakış açısını destekleyen Gamalielsson ve ark. (2021) çalışması ise, açık hükümet bağlamında web analitiği için Matomo gibi açık kaynaklı yazılımların kullanılmasının avantajlarını tartışmıştır. Araştırmacılar, veri gizliliğinin önemini vurgulayarak, açık kaynaklı çözümlerin başka şirketlere olan bağımlılığı önleyebileceğini ve sürdürülebilir analitik uygulamalar elde etme potansiyelini vurgulamışlardır. Veri sahipliğine odaklanan bu çalışmalar, web analitiği ve kullanım madenciliği alanlarında gizlilik ve veri egemenliği konularının giderek daha fazla önem kazandığını göstermektedir.

Bu bölümde ele alınan çalışmaların her biri, alandaki gelişmeler, zorluklar, yeni teknikler ve yaklaşımlar hakkında derinlemesine bilgi sağlayarak web analitiği ve web kullanım madenciliği uygulamalarında daha etkin ve verimli metodolojilerin geliştirilmesine katkıda bulunmuştur. Mevcut çalışmalar genellikle web analitiği alanında belirli bir veya birkaç özgül alanı vurgulamış, ancak bu kritik boyutların hepsini kapsayan çözümler sunmamıştır. Literatürde Čegan ve Filip'in (2017) geliştirdiği Webalyt gibi az sayıda web analitik platformu örneği olmasına rağmen, piyasada kullanılan mevcut araçların teknik altyapılarını ve çalışma yöntemlerini açıklayan yeterli çalışma bulunmamaktadır. Bu eksikliğin muhtemel nedenleri, web analitiği alanının ticari şirketlerin hakimiyeti altında olması ve bu tür büyük projelerin geliştirilmesinin çok fazla zaman ve işgücüne ihtiyaç duyması nedeniyle akademik çevreler tarafından ele alınmamasıdır. Bu tez çalışması, literatürde yeterince detaylandırılmamış olan analitik sistemlerin altında yatan teknolojik yapıları ve işleyiş prensiplerini kapsamlı bir biçimde açıklayarak bu sistemlerin daha iyi anlaşılmasına ve bu alandaki bilgi boşluklarının doldurulmasına katkıda bulunmayı hedeflemektedir.

1.4. Hipotez

Çalışmanın temel hipotezi, önerilen CAWAL modeli ve bu model üzerine inşa edilen yazılım çerçevesinin, web analitiği ve web madenciliği alanlarında önemli ilerlemeler sağlayacağı yönündedir. Geliştirilen modelin, geleneksel yöntemlerin ötesine geçerek

daha etkin ve geniş kapsamlı bir veri toplama ve analiz sürecine imkân tanıyacağı öngörülmektedir. Bu modelin, özellikle web portalları ve çok sunuculu ortamlarda, kullanıcı davranışlarının detaylı analizini mümkün kılması ve bu analizlerin stratejik karar alma süreçlerine entegre edilmesini kolaylaştırması beklenmektedir. Ayrıca, CAWAL modelinin uygulama günlükleriyle entegrasyonu, oturum ve kullanıcı tanımlama süreçlerinde veri kalitesini ve doğruluğunu artıracağı varsayılmaktadır. Bu işlevsellik, WUM'un ön işleme aşamasını büyük ölçüde azaltarak süreç verimliliğini iyileştirecektir. Bu hipotez, çalışmanın ilerleyen bölümlerinde modelin web analitiği ve WUM üzerindeki etkilerini destekleyen kanıtlarla test edilerek değerlendirilecektir.

2. WEB ANALİTİĞİ VE WEB KULLANIM MADENCİLİĞİ

Web analitiği ve web kullanım madenciliği, web portallarındaki kullanıcı etkinliklerini anlamada ve bu etkinliklerden anlamlı kalıplar çıkarmada önemli roller üstlenmektedir. Web portalları ise kullanıcıların hizmetlere erişimden bilgi aramasına ve içerik tüketimine kadar geniş bir yelpazede etkileşimde bulunduğu dinamik platformlardır. Web analitiği, bu etkileşimlerin niceliksel ve niteliksel özelliklerini değerlendirirken, WUM bu verileri kullanarak kullanıcı davranışlarını daha geniş bir perspektiften analiz eder.

Web sunucusu erişim günlükleri, WUM çalışmalarında sıkça kullanılan, önemli bir veri kaynağıdır. Bu günlükler, kullanıcıların site içindeki hareketlerini, hangi sayfaları ziyaret ettiklerini ve hangi içeriklerle etkileşimde bulunduklarını kaydeder. Kolaylıkla erişilebilir ve analiz edilebilir olmaları nedeniyle Apache web sunucusu günlükleri üzerinden yapılmış çalışmalara literatürde sıkça rastlanır. Ancak bu günlükler, kullanıcı davranışlarının daha kapsamlı ve derinlemesine analizleri için her zaman yeterli detay ve doğruluk sağlamazlar. Bu durum, web analitiği ve WUM alanlarında, kullanıcı etkileşimlerini daha detaylı ve doğru bir şekilde anlamak adına alternatif veri toplama ve analiz yöntemlerinin araştırılmasını gerektirir. Web analitiği alanında, sunucu günlüklerinin kullanımı günümüzde çoğunlukla yerini istemci taraflı JavaScript ile gerçekleştirilen kullanıcı takibine bırakmışken, web kullanım madenciliği faaliyetleri için hâlâ yoğun bir şekilde kullanılmaktadır.

2.1. Web Portalı

Web portalı, çeşitli hizmetleri, kaynakları ve bilgileri tek bir noktadan sisteme giriş yapan kullanıcıların erişimine sunan web tabanlı ve geniş çaplı bir platform olarak tanımlanır (Pinho ve ark., 2018). Ana kapı anlamına gelen portal kavramı, haberlerden e-postaya, çevrimiçi alışverişten veritabanlarına erişime ve forumlara kadar çeşitli interaktif modülleri bir arada bulunduran web siteleri için kullanılır. Web portalları, kullanıcıların bilgiye hızlı ve kolay erişimini sağlamak, çeşitli online hizmetleri entegre etmek ve kişiselleştirilmiş bir kullanıcı deneyimi sunmak üzere merkezi bir yapıda tasarlanırlar.

Web portalları, kullanıcılara bilgi, uygulama ve diğer kaynaklara erişim için kişiselleştirilmiş ve bütünleşik bir platform sağlar. İçerik organizasyonu, arama işlevi, kişiselleştirme yetenekleri ve sanal kullanıcı toplulukları için destek gibi özelliklerle kullanıcı deneyimini zenginleştirirler. Ayrıca, bilgi paylaşımı ve erişimi için sundukları özellikler ve yeteneklerle veri şeffaflığını ve açık erişimi destekleyerek, çeşitli kullanıcı ihtiyaçlarına cevap verebilen dinamik ve etkileşimli ortamlar oluştururlar.

Portal kavramı ilk kez, 1996 yılında "MyYahoo!" kişiselleştirilmiş ana kapı hizmetinin devreye alınmasıyla hayatımıza girmiştir. Bu sitede kullanıcılar şifreleriyle sisteme giriş yaparak kendi ilgi alanlarına göre profil oluşturmakta ve sunulan hizmetlere kolayca ulaşmaktaydılar. Günümüzde yerli yabancı birçok örneği bulunan web portalı özelliğine sahip siteler, gelişen web standartları dâhilinde artmaya devam etmektedir. Web portalları, erişim haklarına dayanarak kullanıcıların hizmetlere etkin katılımlarını sağlamanın yanı sıra, bu hizmetlerin kurulum ve yapılandırılmasını da kolaylaştırarak uzak kaynaklara erişim prosedürlerini daha basit hale getirir (Chung ve ark., 2019).

Web portallarının altyapısı incelendiğinde, sisteme girişin tek noktadan yapılması ana esastır. Portala bağlı servisler genelde sitenin alan adına bağlı alt alan adları üzerinden hizmet verilirler ve kullanıcının doğrudan bir servisin web sitesine girmesi durumunda sistem kullanıcıyı giriş yapması için ana kapıya yönlendirir. Portal sistemler, farklı servislerin tümünde bulunması gereken sistem güvenliği, oturum yönetimi, kullanıcı doğrulama, yetkilendirme, işlem kayıtları tutma, hata yakalama ve raporlama, şablon işlevleri gibi ortak temel süreçlerin bütünleşik olarak tek bir noktadan ele alındığı, bu nedenle geliştirilmesi daha zor fakat genişlemesi ve yönetilmesi kolay yapıdadırlar.

Web portallarının özellikleri, kullanıcı memnuniyetini ve kullanım olasılığını etkilemede önemli bir rol oynamaktadır (Al-zagheer ve ark., 2022). Bu özellikler arasında içerik, tasarım, kişiselleştirme yetenekleri ve sanal kullanıcı topluluklarının oluşturulmasına yönelik destek yer almaktadır. Özellikle devlet kurumları için geliştirilen web portallarının kullanılabilirliği ve kalitesi, işlevsel ve işlevsel olmayan özellikler açısından standart kaliteyi karşılamak için gereklidir (Lourenço, 2015).

Web portalları, farklı kullanıcı gruplarına hizmet vermek için geniş bir yelpazede tasarlanabilir. Kurumsal portallar, çalışanların işle ilgili kaynaklara erişimi için kullanılırken, eğitim portalları öğrencilerin çevrimiçi derslere, ders materyallerine,

duyurularına ve sınav sonuçlarına erişimini sağlar. Ayrıca, hükümet portalları vatandaşların kamu hizmetlerine erişimini kolaylaştırır ve e-ticaret portalları tüketicilere çeşitli ürün ve hizmetler sunar (Pinho ve ark., 2018). Web portallarının bu geniş işlevselliği, web analitiği ve web kullanım madenciliği açısından önemli bir zenginliktir. Portallardaki kullanıcı etkileşimleri, ziyaretçi davranışlarını anlamak, kullanıcı deneyimini iyileştirmek ve içerik stratejilerini geliştirmek için değerli veri kaynakları sağlar.

2.2. Web Kullanıcı Erişim Günlükleri

Web'in doğası gereği, kullanıcılar web sitelerindeki bağlantılara tıklayarak bir sayfadan diğerine geçerler. Her sayfa geçişi, sunucuya yapılan yeni bir istektir ve bu istekler web sunucu yazılımı tarafından kayıt altına alınır. İlk web tarayıcısı Mosaic'in 1993 yılında kullanıma sunulmasıyla birlikte web sunucu erişim günlük dosyaları (access log), "hit" olarak da adlandırılan web isteklerini takip etmek için kullanılmaya başlandı (Zheng ve Peltsverger, 2015). Sonrasında, kullanıcıları izlemek (takip etmek) ve erişim verilerini toplamak için çeşitli yöntemler ortaya çıktı. Bu yöntemler genellikle sunucu seviyesi ve uygulama seviyesi olmak üzere iki ana grupta ele alınırlar.

Sunucu seviyesi günlükler (log), web, uygulama veya vekil (proxy) gibi çeşitli hizmetler sağlayan sunucular tarafından sistemin çalışması sırasında arka planda elde edilen ve saklanan erişim kayıtlarıdır. Bir web sunucusunun erişim günlüğü, tarayıcılar tarafından yapılan tüm HTTP isteklerinin yarı yapılandırılmış verilerini içerir. Dünya çapında en çok kullanılan web sunucularından biri olan Apache, varsayılan olarak NCSA'nın Ortak Günlük Biçimi'ni (CLF) kullanır, ancak sunucu yöneticisi daha farklı bir biçim tanımlayabilir. Apache web sunucusunun erişim günlükleri için varsayılan CLF yapılandırması şu şekildedir (Srivastava ve ark., 2015):

LogFormat "%h %l %u %t "%r" %s %b"

Burada yer alan bağımsız değişkenlerin anlamları aşağıdaki gibidir:

%h – İstemcinin IP adresi.

%l – HTTP kimlik doğrulaması tarafından belirlendiği şekilde belgeyi talep eden kişinin kimliği.

%u – İsteğin kimliği HTTP ile doğrulanmışsa istemcinin kimliği.

%t – İsteğin alındığı zaman.

%r – HTTP yöntemini, protokolünü ve kaynak yolunu içeren istek satırı.

%s – Sunucunun istemciye geri gönderdiği durum kodu (200: başarılı, 404: bulunamadı, vb.).

%b – İstek yapılan nesnenin bayt cinsinden boyutu.

Apache web sunucusu ayrıca, W3C Genişletilmiş Günlük Biçimi'ne (ELF) benzeyen ve NCSA Birleşik Günlük Biçimi olarak da bilinen Genişletilmiş Ortak Günlük Biçimi'ni (ECLF) destekler. Bu günlük biçimi, CLF'ye ek olarak yönlendirici ve web tarayıcı/kullanıcı aracısı (user-agent) olmak üzere iki bilgi daha içerir (Srivastava ve ark., 2015).

Tipik bir ECLF yapılandırması olan Apache web sunucusunun erişim günlüğündeki tek satırlık basit bir girdi örneği aşağıdaki gibidir (Das, 2008):

193.140.253.80 - - [15/Aug/2021:17:30:51 +0300] “GET /index.php HTTP/1.1” 200 1246 http://www.server.com/ “Mozilla/5.0 (X11; Ubuntu; Linux x86 64; rv:15.0) Gecko/ 20100101 Firefox/15.0.1”

Belirli bir dizi analiz gerektiren ve Weblogic, WebSphere veya Tomcat gibi ticari uygulama sunucularından elde edilen “uygulama sunucusu günlükleri”, sunucu düzeyinde günlüklerin diğer örnekleridir (Mishra ve Srivastava, 2021). Vekil sunucular ve çeşitli hizmetler sağlayan diğer sunucular da operasyonları sırasında yaygın olarak log tutmaktadır. Öte yandan, uygulama seviyesinde toplanan günlüklerde verinin elde edilmesi ve saklanması ayrı aşamalar olarak ele alınmaktadır.

Veri saklama aşaması sunucular üzerinde (yerinde ya da kurum dışında) gerçekleşirken veri elde etme aşaması istemci tarafı veya sunucu tarafı web teknolojiler aracılığıyla yürütülür. Uygulama düzeyindeki kullanıcı erişim verileri genellikle JavaScript (JS) parçacıkları kullanılarak istemci tarafı sayfa etiketleme yöntemiyle elde edilir ve günlükler çoğunlukla -üçüncü parti olarak da adlandırılan- kurum dışındaki uzak sunucularda saklanır (Zheng ve Peltsverger, 2015). Ancak bu yöntem, verilerine serbestçe web madenciliği teknikleri uygulamak isteyen kuruluşlar için ideal bir seçenek değildir.

2.3. Web Analitiği

Web analitiği, "Web Analitiği Derneği" (Web Analytics Association) tarafından "web kullanımını anlamak ve optimize etmek için internet verilerinin ölçümü, toplanması, analizi ve raporlanması" olarak tanımlanmaktadır. Site içi ve site dışı web kullanımını başarılı bir şekilde tespit etmek ve iyileştirmek için büyük miktarda veri biriktirmeyi ve değerlendirmeyi gerektiren bir süreç olan web analitiği, bu alanda kritik bir rol oynamaktadır (Clifton, 2012). Önder ve Berbekova (2021) web analitiğini aşağıdaki dört adımdan oluşan bir süreç olarak tanımlamıştır:

- Veri toplama (ziyaretçi sayısı, sayfada kalma süresi, vb.)
- Analitik çıkarma (metrikler oluşturarak verileri bilgiye dönüştürme)
- KPI'ların geliştirilmesi (web analitiği metriklerinden elde edilen bilgilerin iş zekâsı için kullanılması)
- Çevrimiçi stratejinin formüle edilmesi (örneğin, çevrimiçi pazarlama kampanyaları oluşturma)

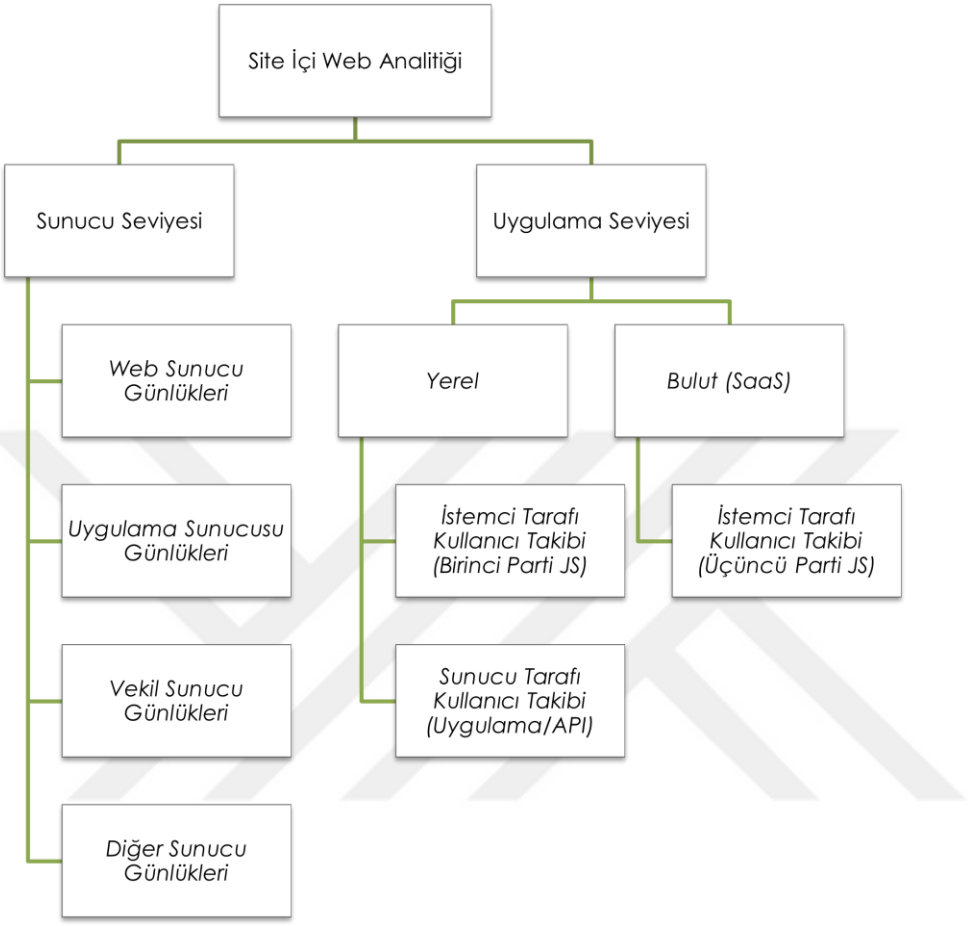
Web analitiği ayrıca dijital varlıkların performansını ve kullanıcı etkileşimlerini değerlendirmek için kritik bir araçtır. Bu tür analizler, web sitesi sahiplerinin ve pazarlamacıların ziyaretçi davranışlarını derinlemesine anlamalarına ve çevrimiçi varlıklarını daha etkili bir şekilde yönetmelerine olanak tanır.

2.3.1. Web analitiği yöntemleri ve araçları

Web analitiği temel olarak site dışı ve site içi web analitiği olmak üzere ikiye ayrılır. Site dışı web analitiği, bir web sitesini etkileyen arama motoru sıralamaları ve sosyal bahsedenler gibi dış faktörleri değerlendirir. Buna karşılık site içi web analizi, kullanıcı davranışlarını ve doğrudan web sitesinden toplanan metrikleri analiz etmeye odaklanır (Rafiq, 2022). Metrikler, ölçülebilir, anlamlı ve nicel veriler içeren göstergelerdir. Bir sistemin performansını ölçen metrikler, performans göstergeleri olarak adlandırılırken iş alanına özgü kritik önemdeki metrikler ise Temel Performans Göstergeleri (KPI) olarak isimlendirilirler.

Web analitiği araçları, kullanıcı yollarını, tek tek sayfalarda harcanan zamanı ve site öğeleriyle etkileşimleri izleyerek ayrıntılı kullanıcı davranışı analizi sağlar ve kuruluşların içeriklerini ve tasarımlarını optimize etmelerine olanak tanır (Pilz ve ark., 2020). Günümüzde daha yaygın kullanılan site içi web analitiği yöntemleri, kullanıcı davranışı, web sitesi performansı ve etkileşim metrikleri hakkında veri toplamak için

ağırlıklı olarak sunucu tarafı günlük kaydı ya da istemci tarafı etiketlemeye dayanır. Site içi web analitiğinde yaygın kullanılan izleme ve veri toplama yöntemleri Şekil 2.1'de ağaç diyagramı biçiminde verilmiştir.

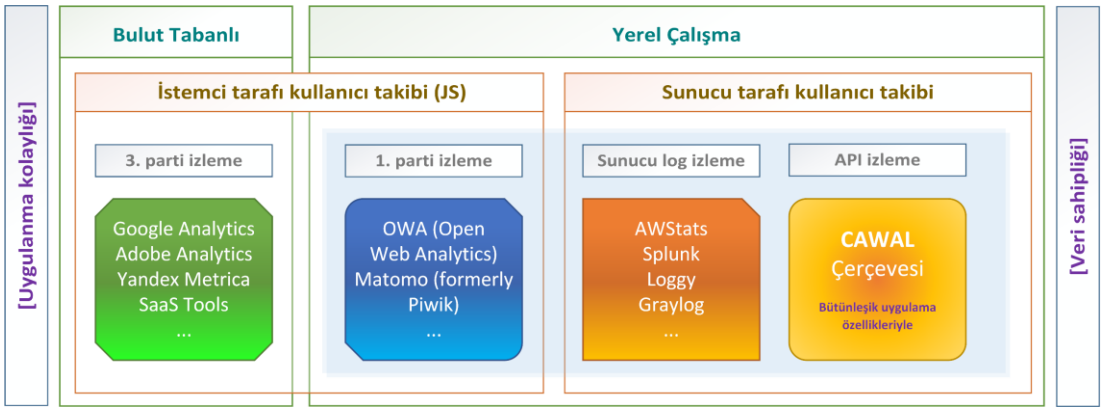


Şekil 2.1. Site içi web analitiği kullanıcı izleme yöntemleri.

Web analitiğinin ilk dönemlerinde trafik ölçümü büyük ölçüde günlük dosyası analizini kullanan isabet (hit) sayıları gibi ilkel metriklere dayanmaktaydı (Kousik ve ark., 2021). Web sunucuları, sayfa görüntülemeleri ve dosya indirmeleri gibi kullanıcı faaliyetlerini kaydediyor, analistler de temel kullanıcı davranışı içgörülerini toplamak için bunları manuel olarak inceliyordu. Bu günlük merkezli yaklaşım, insan ve otomatik trafik arasında doğru bir ayrım yapamama ve kullanıcı oturum yollarını izlemedeki eksiklikler de dahil olmak üzere çeşitli sınırlamalarla karşı karşıya kalmaktaydı (Cahyanto ve ark., 2022). Veriler ne kadar zengin olursa, analiz edilmesi de o kadar karmaşık hale geldiğinden, bu tür karmaşıklıkları çözmek, verileri işlemek ve görselleştirmek sağlam araçlar gerektirmektedir (Walny ve ark., 2020).

Web günlüğü analizörleri, kullanıcı davranışı ve site performansı hakkında önemli metrikler sunmak için tıklama akışlarını işler ve bilinçli karar verme için hassas analitik sağlamak üzere aritmetik ve istatistiksel yöntemler kullanırlar (Sakas ve ark., 2021). Günümüzde Splunk ve Graylog gibi isabet filtreleri ve örüntü eşleştirme işlevlerine sahip, son derece yetenekli günlük analizörleri geliştirilmesine rağmen, geleneksel günlük tabanlı yöntemin popülerliği ve kullanımı giderek azalmaktadır.

1990'ların sonu ve 2000'lerin başında log tabanlı analitiklerden önce tek piksellik resim kullanımı, ardından JavaScript etiketlemeye geçişle birlikte web analitiğinde önemli bir değişim yaşanmıştır. Tekil ziyaretçiler ile oturum yolları gibi daha ayrıntılı verilerin yakalanması mümkün hale getiren bu teknolojik evrim, günümüz pazarına hâkim olan Google Analytics gibi bulut tabanlı Hizmet Olarak Yazılım (SaaS) çözümlerinin öncüsü olmuştur (Alby, 2023). Kişiselleştirmenin önem kazanmasıyla birlikte işletmeler artık kitleleri demografik özellikler, coğrafi konum, davranış ve kullanılan cihazlar gibi çeşitli parametrelere göre segmentlere ayırabilmektedir (Vamsee ve ark., 2023). Ayrıca, makine öğrenimi ve yapay zekâ entegrasyonu, gelecekteki kullanıcı davranışını tahmin etmeye ve daha etkili bir şekilde strateji oluşturmaya yardımcı olan tahmine dayalı analitiği mümkün kılmıştır. Günümüzde site içi web analitiğinde kullanıcı takibi ve veri toplama için yaygın olarak kullanılan yöntemler ve CAWAL'ın bunlar arasındaki yeri Şekil 2.2'de gösterilmiştir.



Şekil 2.2. Web analitiğinde kullanıcı izleme ve veri toplama yöntem ve araçları.

Sayfa yükleme hızı, yüksek kullanılabilirlik ve düşük bellek gereksinimleri, özellikle büyük ölçekli web uygulamalarının geliştirilmesinde dikkate alınan önemli kriterlerdir. Web kullanımında daha yüksek yanıt gecikmesi, kullanıcı katılımının azalmasına, görev tamamlama oranlarının düşmesine ve kullanıcıların web sitesini terk etme olasılığının artmasına neden olmaktadır (Szalek ve Borzemski, 2018;

Nurshuhada ve ark., 2019). Araştırmalar, sayfa yükleme süresinin üç saniyenin üzerinde olması halinde ziyaretçilerin %40'ından fazlasının web sitesini terk ettiğini göstermektedir (Stringam ve Gerdes, 2019; Pourghassemi ve ark., 2019). Bu sorunların daha da kötüleşmesini önlemek için web analitik araçlarının hassas bir şekilde entegre edilmesi şarttır. Bu gerçeğin farkında olan tasarım ve yazılım geliştirme ekipleri, web uygulamalarının performansını artırmaya özel önem vermektedir.

Literatürde çeşitli web analitik araçlarını ve seçim stratejilerini araştıran çalışmalar bulunmaktadır (Almatrafi ve Alharbi, 2023; Boufenneche ve ark., 2022; Kumar ve Ogunmola, 2020). Bir web analitik aracının seçilmesi, veri toplama yöntemleri, konum ve maliyet yapısı dahil olmak üzere birçok faktörü içerir. Ek hususlar, izleme kapasitesi, gerçek zamanlı veri sağlama ve belirlenmiş metrik izlemeyi kapsar. Aracın veri segmentasyon yetenekleri, gösterge paneli özelleştirmesi, birlikte çalışabilirlik ve maliyet etkinliği de kararı etkiler (Al-Otaibi ve ark., 2018). Tüm bu hususlar göz önüne alındığında, seçilecek analitik platform, kullanıcı deneyimi ve performans ölçümlerinden ödün vermeden, kurumsal ihtiyaçlar ve web uygulamasının hedefleriyle uyumlu olmalıdır.

Google Analytics gibi bulut tabanlı ücretsiz hizmetler, başlangıçta maliyet tasarrufu sağlayan avantajlar ve temel bilgiler sunuyor olsalar da yüksek web trafiğine sahip kuruluşların ihtiyaç duyduğu kapsamlı analizlerden yoksundurlar (Haaksma ve ark., 2018; Rafiq, 2022). Öte yandan, bulut tabanlı rakipleri gibi kullanıcı takibi için istemci tarafı JavaScript kullanan Open Web Analytics (OWA) ve Matomo gibi yerel konuşlu platformlar, veri sahipliğine odaklanan alternatifler sunmaktadır (Alby, 2023). Bu nedenle, bir web analitik aracının seçimi yalnızca kuruluşun özel ihtiyaçları ve hedefleriyle değil, aynı zamanda veri sahipliği, rekabetçi kıyaslama ve gelişmiş raporlama gibi hususlarla da uyumlu olmalıdır. Bu uyum, stratejik kararlar vermek için hassas ve kapsamlı analitiklere ihtiyaç duyan kuruluşlar için son derece önemlidir.

2.3.2. Web analitiği veri sorunları

Web analitiğinde doğruluk ve bütünlük, verilere dayanarak anlamlı sonuçlar ve veri odaklı kararlar almak için kritik öneme sahiptir. Web analitik verilerinin doğruluğunu ve bütünlüğünü değerlendirmeye odaklanan yakın tarihli bir çalışmada, web siteleriyle kullanıcı etkileşimlerinin ölçülmesine ilişkin bilgiler sunulmuş ve endüstri standardı olan iki analiz yaklaşımı karşılaştırılmıştır (Jansen ve ark., 2022).

Veri doğruluğunun yanı sıra, verileri gerçek zamanlı olarak analiz etme yeteneği, önceki statik günlük dosyası tabanlı web analitiğine göre önemli bir gelişmeyi temsil etmektedir (Boufenneche ve ark., 2022). Mevcut araçlar, anlık içgörüler sağlayarak değişen kullanıcı davranışlarına veya pazar koşullarına hızlı bir şekilde yanıt verilmesine yardımcı olmaktadır. Bununla birlikte, JavaScript ile ilgili sorunlar veri bütünlüğünü zayıflatabilirken, çok sunuculu ve çok alanlı web sistemleri benzersiz veri uyumlaştırma ve analiz zorlukları ortaya çıkarmaktadır (Biørn-Hansen ve ark., 2017).

Bir web çiftliğindeki etki alanları ve alt etki alanları arasında kullanıcıları izlemedeki zorluklar, merkezi olmayan sunucu mimarileri ve web protokolleri tarafından zorunlu kılınan etki alanı kısıtlamaları nedeniyle daha da ağırlaşmaktadır (Krishnan ve ark., 2020). Yük dengelemeli bir web çiftliğindeki i^{nci} olarak tanımlanan her sunucu, gelen taleplerin ölçülebilir bir alt kümesi olan n_i'yi işler ve analiz için çeşitli metrikleri kaydeder. Sistemde m sunucu bulunduğunda, veri toplamaya yönelik basit yaklaşım, alt kümeler üzerinden toplama işlemi yapmayı gerektirir. Ancak bu yaklaşım, birden fazla sunucuya dağılmış kullanıcı oturumlarından kaynaklanan karmaşıklıkları gözden kaçırabilir.

Veri birleştirme ve normalleştirme adımları, en kötü senaryoda $O(m \times n)$ zaman karmaşıklığı ile tüm m sunucudan gelen günlükleri birleştirmeyi amaçlayan çok önemli aşamalar haline gelir. Birden fazla sunucu kullanıcı etkileşimlerini işleyebileceğinden, yük dengeleme mekanizmaları kullanıcı takibini bozabilir ve veri senkronizasyonu için birleşik bir sunucu tarafı aracı gerektirir.

Bir diğer zorunlu görev de tek bir kullanıcıdan kaynaklanan, ancak birden fazla sunucudaki parçalanmış oturumların tanımlanmasını gerektiren oturum uzlaştırmasıdır (Mortazavi ve ark., 2020). Bu tür bir tanımlama genellikle çerezler gibi istemci tarafı mekanizmalarda depolanan oturum tanımlayıcılarının kullanılmasını içerir. Çerezler, kullanıcı faaliyetlerini izlemede ayrılmaz bir rol oynamakla birlikte, kullanımları etki alanına özgü sınırlamalara tabidir. Alt etki alanları içinde izleme için çerez etki alanı özniteliğinin birincil etki alanına ayarlanması tutarlı bir izleme sağlayabilir. Bununla birlikte, farklı etki alanları arasında izleme, daha fazla karmaşıklık getirir. Üçüncü taraf çerezlerin, bunları sıklıkla engelleyen modern gizlilik korumalarına atfedilebilecek azalan etkinliği, alternatif izleme yöntemlerinin araştırılmasına yol açmıştır (Durántez, 2023).

Web analitiğinde oturum ve kullanıcı tanımlama, kullanıcı davranışlarını doğru anlamak için hayati derecede önemlidir. Oturum tanımlama (Maslennikova ve ark., 2022) matematiksel olarak bir $S(A,T,C)$ fonksiyonu olarak modellenebilir. A kullanıcı eylemlerini, T etkinlikler arasındaki zaman aralıklarını ve C, IP veya tarayıcı gibi bağlamsal verileri gösterir. Birincil amaç, $g:A\times T\times C\rightarrow S$ olacak şekilde bir g fonksiyonu formüle etmektir. Bu fonksiyon, A eylemlerini T zamanı ve C bağlamındaki yakınlıklarına göre S uyumlu oturumlar halinde harmanlamayı amaçlar. Ancak bunu başarmak, bir dizi karmaşık faktörün üstesinden gelmeyi gerektirir. Örneğin, Oturum Zaman Aşımı (T_o) önceden tanımlanmış bir aralıktır ve eğer T, T_o'yu aşarsa, önceki oturumun sona erdiği varsayılır. Benzer şekilde, Kullanıcı Hareketsizliği (I_a) veya önemli Bağlam Değişikliği (C_c) dönemleri bir oturumun sınırlarını gösterebilir.

İkinci önemli görev olan kullanıcı tanımlama (Salminen ve ark., 2020), kullanıcıları birden fazla etkileşim üzerinden tanımayı amaçlar ve $I(U,M,D)$ ile temsil edilebilir. U tüm kullanıcıların kümesidir, M çeşitli cihazları veya tarayıcıları içerebilen etkileşim yöntemlerini temsil eder ve D, IP adresleri veya çerezler gibi mevcut veri noktalarını içerir. $f:M\times D\rightarrow U$ olarak tanımlanan f görev fonksiyonu, M etkileşim yöntemlerine ve D veri noktalarına dayanarak bir U kullanıcısını benzersiz bir şekilde tanımlamaya çalışır. Ancak, çeşitli faktörler bu görevi zorlaştırmaktadır.

Dinamik IP tahsisi (I_d), çerezlerin silinmesi (C_d), tarayıcı parmak izi belirsizlikleri (B_f) ve birden fazla cihazdan erişim (D_m), f 'nin etkili bir şekilde gezinmesi gereken belirsizlikleri ortaya çıkarır. Dolayısıyla hem g hem de f bu görevlerin doğasında bulunan karmaşıklıkları ve zorlukları bünyesinde barındırır. Bu zorlukların başında gelen oturum sınırı sorunu, web analizlerinin, özellikle de kullanıcı etkileşimi metriklerinin doğruluğunu önemli ölçüde etkiler. Geleneksel teknikler, kullanıcı oturumlarını bölümlere ayırmak için sabit bir 30 dakikalık boşta kalma süresi kullanır (Jansen ve ark., 2022), bu da etkinliği yanlış sınıflandırabilen bir yaklaşımdır. Örneğin, boşta kalma süresi boyunca ek sunucu istekleri olmadan bir makalenin okunması, etkin olmayan bir oturum olarak kabul edilebilir (Maslennikova ve ark., 2022). Bu tür yanlışlıklar temel performans göstergelerini bozar ve analitik verilere gürültü katarak karar verme ve kullanıcı deneyimi optimizasyon stratejilerini zayıflatır. Olasılıksal modeller ve makine öğrenimi, kullanıcı davranışı ve bekleme süresi gibi değişkenleri

hesaba katarak oturum tanımlama doğruluğunu artırmak için kullanılmaktadır (Alhlou ve ark., 2016).

Bir diğer önemli sorun ise gece yarısını geçen oturumları birden fazla oturum olarak yanlış sınıflandıran ve oturum sayıları ile ortalama site süresi gibi ölçümleri etkileyen gece yarısı sınırı sorunudur. Bazı yerleşik araçlar uyarlanabilir oturum zaman aşımı gibi kısmi çözümler sunsalar da zaman dilimleri arasındaki karmaşık kullanıcı davranışları nedeniyle sorun devam etmektedir (Jansen ve ark., 2022). Bu nedenle, güvenilir web analitiği için hem oturum sınırı hem de gece yarısı sınırı sorunu hala geliştirilmesi gereken önemli zorluklardır.

Web analitiği, aynı zamanda iş zekâsı gelişiminin ilk aşaması olarak kabul edilmektedir (Król ve Zdonek, 2020). Ancak, bazı nedenlerden dolayı bulut tabanlı web analitiği araçları her kuruluş için mükemmel bir çözüm olmayabilir. Bu nedenlerin başında veri gizliliği ve egemenliği endişeleri gelmektedir (Samarasinghe ve ark., 2022). Gizlilik konuları günümüzün bilgi toplumlarında giderek daha hassas ve önemli hale gelmiştir. Ülkeler ve kuruluşlar, özellikle farklı ülkelerdeki şirketlerle kullanıcı verileri üzerindeki kontrol ve sahiplik rollerini paylaşma konusunda tereddüt etme eğilimindedir (Everest-Phillips, 2019). Veri gizliliği yasaları ve düzenlemeleri, kuruluşları ve bireyleri veri ihlallerinden korumak için hayati önem taşırken, uyumluluğun sağlanması genellikle önemli kurumsal uzmanlık gerektirmektedir (Binns, 2017).

2.4. Web Madenciliği

Veri tabanlarından bilgi keşfinin bir adımı olan veri madenciliği, çeşitli teknikler kullanarak verilerdeki yararlı bilgileri elde etme sürecidir. Veri madenciliği, ilişkisel, uzamsal (coğrafi), multimedya veya zamansal veri tabanları, veri ambarları ve web verileri gibi çeşitli veri türlerine uygulanabilir (Reddy, 2021). Veri madenciliğinin çeşitli uygulama alanları olmakla birlikte web alanına uygulanan ve web verilerini girdi olarak kullanan türüne "web madenciliği" adı verilmektedir. Web madenciliği (WM), üç kategoride ele alınmaktadır (Das ve Turkoglu, 2009):

- Web İçerik Madenciliği (WCM), web sayfalarındaki metin, resim ve video gibi içeriklerden bilgi madenciliği yapmakla ilgilidir (Shah ve Pandit, 2022).
- Web Yapısı Madenciliği (WSM), web sitelerinin yapısal özetini üretmeyi ifade eder ve sayfalardaki bağlantıların yapısıyla ilgilenir (Tyagi ve Gupta, 2018).

- Web Kullanım Madenciliği (WUM) ise tıklama akışı modellerini veya ilişkili verileri keşfetmeye ve analiz etmeye odaklanır (Lim ve ark., 2021).

2.4.1. Web kullanım madenciliği

Web kullanım madenciliği, kullanıcıların tarama geçmişini içeren web erişim günlüklerinde değerli kalıpları arama sürecidir (Cooley ve ark., 1999). Web kullanım madenciliği amaç, hedef, yöntem ve teknikleri Tablo 2.1'de verilmiştir (Román ve ark., 2010; Farid ve ark., 2018).

Tablo 2.1. WUM amaç, yöntem ve teknikleri.

WUM Amaçları	Örnek Hedefler	Yöntemler/Metotlar	Önemli Teknikler
Kullanıcı Profilleme	Kullanıcı Davranışının Anlaşılması	Sınıflandırma	Makine Öğrenimi Algoritmaları (SVM, Karar Ağaçları, vs.)
	Anomali Tespiti	Kümeleme	Veri Ön İşleme Teknikleri, DBSCAN, OPTICS
	Kişiselleştirilmiş İçerik ve Öneri Sistemleri	İşbirlikçi Filtreleme	Matris Çarpanlarına Ayırma, Kullanıcı-Öğe Benzeşim Analizi
	Kullanıcı Segmentasyonu ve Demografik Analiz	K-En Yakın Komşuluk (KNN)	Apriori Algoritması
Etkileşim Modelleri	Etkileşim Modellerinin Keşfi	Birliktelik Kuralı Madenciliği	İstatistiksel Analizler, Kaldırma ve Güven Ölçütleri
	Sayfa Görüntüleme Sürelerinin Analizi	Sıralı Desen Madenciliği (SPM)	Zaman Serisi Analizi
	Kullanıcı Navigasyon Yollarının Keşfi	Temel Bileşenler Analizi (PCA)	Gizli Markov Modelleri
Navigasyon Analizi	Giriş/Çıkış Sayfalarının Analizi	Tahmin	Regresyon Modelleri, Lojistik Regresyon
	Mevsimsellik ve Trend Analizi	Rastgele Orman	Mevsimsel Ayrıştırma
	Kullanıcı Yolculuklarının Optimizasyonu	Zaman Serisi Analizi	Trend Analizi, ARIMA

Aynı veri türünü kullanmalarına rağmen, web kullanım madenciliği ve web analitiği birbirinden belirgin şekilde ayrılır (Canay ve Kocabıçak, 2016). Web analitiği, günlük verilerinin toplanmasını ve ardından istatistiksel analiz yoluyla görsel temsilini kapsar. Buna karşılık, web kullanım madenciliği öncelikle mevcut verilerin temizlenmesi ve işlenmesinin ardından bilgi keşif tekniklerinden yararlanarak verilerden geçerli, yenilikçi, yararlı ve yorumlanabilir kalıpları tespit etmeyi amaçlar. Dolayısıyla web kullanım madenciliği süreci, verilerin elde edilmesi ve biriktirilmesinden ziyade var olan verilerin işlenmesiyle başlar.

Web kullanım madenciliği üzerine yapılan çalışmaların büyük çoğunluğu web sunucusu günlükleri kullanılarak gerçekleştirilmiştir. Bu araştırmalar, günlüklerden bilgi çıkarımını kolaylaştırabilecek çeşitli metodolojilerin ve araçların tartışılmasını kapsamaktadır (Harika ve Sudha, 2019; Jin ve Lin, 2022; Abilio ve ark., 2021). Ham günlükleri temizlemek ve bunları işlem öğeleri kümelerine veya bir veri küpüne dönüştürmek için çeşitli veri madenciliği teknikleri uygulanmalıdır. Web günlüklerinden bilgi çıkarmak için web kullanım madenciliği sürecinin şu üç aşaması gerçekleştirilmelidir:

- Ön işleme, mevcut eksik veya tutarsız verilerin bir sonraki aşamanın ihtiyaçlarına göre işleneceği ilk adımdır (Manchanda ve Gupta, 2018).
- Örüntü keşfi, verideki örüntüleri elde etmek için istatistik, veri madenciliği, makine öğrenmesi ve örüntü tanıma gibi çeşitli yöntem ve algoritmaların uygulandığı ikinci aşamadır (Jokar ve ark., 2016).
- Örüntü analizi ise bulunan örüntülerin analiz edilerek anlaşılabilir yorumlamalar ve görselleştirmelerle sergilendiği son aşamadır (Kumar ve Kumar, 2021).

2.4.2. Ön işleme aşaması

Veri ön işleme (data preprocessing), web kullanım madenciliğindeki ilk ve en karmaşık görevdir. Genellikle madencilik için harcanan tüm çabanın %60'ından fazlasını alan bu önemli aşama, örüntü keşfi için yapısal, güvenilir ve entegre bir veri kaynağı sağlar (Kewen, 2012; Mobasher, 2009). Veri ön işleme süreci çeşitli alt aşamalardan oluşur:

- Veri filtreleme, statik dosyalar ve siteyi tarayan arama motoru robotları gibi kullanıcılara yönelik olmayan girişleri kaldırmak için verilerin temizlenmesi sürecidir (Slanzi ve ark., 2017).
- Kullanıcı tanımlama, IP adreslerine ve bazen kullanıcı aracılarına dayalı olarak bireysel kullanıcıları tanımlama işlemidir (Nandal, 2018).
- Sayfa görüntüleme tanımlama, hangi sayfa dosyası erişimlerinin tek bir tarayıcı görüntüsüne katkıda bulunduğunu belirleme sürecidir (Slanzi ve ark., 2017).

- Oturum tanımlama (oturumlaştırma), siteye yapılan tek bir ziyareti temsil etmek için her kullanıcının etkinlik kaydını oturumlara bölme işlemidir (Fatima ve ark., 2016).
- Yol tamamlama, kullanıcı isteklerini maskeleyen yerel önbellek yönteminin veya "geri" düğmesinin kullanılması nedeniyle günlüğe kaydedilmeyen erişimlerin tanımlanması sürecidir (Joachims ve ark., 2017).

2.4.3. Ön işlemedeki zorluklar

Web madenciliği faaliyeti, öncelikle uygun olmayan veriler gibi teknik sorunlarla karşı karşıyadır. Başarılı ve güvenilir bir madencilik için toplanan verilerin uygun bir biçimde (formatta) olmaları gerekir. Ancak bu veriler genellikle eksik ya da kullanılamaz durumdadır, dolayısıyla doğrulukları sağlanamaz. Madencilik çalışmalarının büyük çoğunluğu veri kalitesini iyileştirmek için harcanmaktadır (García ve ark., 2015).

Web sunucuları tarafından sağlanan ham erişim günlük dosyaları, örüntü keşfi ve analizinin doğruluğunu etkileyen alakasız unsurlar içerir. Bu önemsiz girdiler arasında, sayfanın içeriğiyle ilgisi olmayan resim, CSS veya JavaScript dosyaları gibi kaynaklar, HTTP durum kodu 200'den farklı olan başarısız istekler ve arama motoru indeksleme robotlarının bıraktığı izler yer almaktadır (Kaur ve Aggarwal, 2017). Filtrelenmesi gereken ilgisiz kayıtlar için aşağıdaki dosya istek örnekleri ve HTTP durum kodları incelenebilir:

“GET /site HTTP/1.1” 301

“GET /favicon.ico HTTP/1.1” 200

“GET /app/admin/adm.php HTTP/1.1” 404

“GET /app/images/ccc.png HTTP/l.1” 200

“GET /app/css/login.css HTTP/1.1” 200

“GET /website/ HTTP/1.1” 302

Web sunucusu günlüklerinden kullanıcıları tanımlamak başlı başına zorlu bir iştir. ECLF günlük biçimi için bir kullanıcı tanımlama problemi şöyle formüle edilebilir:

$IP = \{ip_1, ip_2, ..., ip_n\}$ web sitesine erişen kullanıcıların tüm IP adreslerinin bir kümesi, $R = \{r_1, r_2, ..., r_n\}$ web sitesinin tüm kaynaklarının bir kümesi, $B = \{b_1, b_2, ..., b_n\}$ web

kullanıcılarının tüm tarayıcılarının bir kümesi ve $K = \{k_1, k_2, ..., k_n\}$ web sitesi dışındaki bağlantıların bir kümesi olsun. Bu durumda, ECLF yapısındaki bir satırlık günlük girişi $l_i = \langle\ ip_i;\ t;\ d;\ r_i;\ v;\ c;\ s;\ [ref_i];\ [b_i];\ [çerezler]\ \rangle$ şeklinde tanımlanır. Burada $ip_i \in IP$, $r_i \in R$, $ref_i \in R \cup K$, $b_i \in B$, t zaman damgası olarak istek zamanını, d GET/POST gibi istek yöntemini, v HTTP sürümünü, c HTTP durum kodunu ve s aktarılan bayt boyutunu temsil eder. ref_i, b_i ve *çerezler* özelleştirilmiş özniteliklerdir.

Bir web sunucusu günlüğü $L = \{l_1, l_2, ..., l_n\}$'den oluşur. Temizlenmiş ve ayıklanmış günlük $CL = \{cl_1, cl_2, ..., cl_n\}$ ilgili girdileri içerir ve $cl_i = \langle\ ip_i;\ t;\ r_i;\ [ref_i];\ [b_i]\ \rangle$ olarak tanımlanabilir. $U = \{u_1, u_2, ..., u_n\}$ web sitesine erişen tüm kullanıcıların bir kümesi olduğunu varsayalım. Bir u_i kullanıcısının web sitesini ziyareti $vs_i = \langle\ u_i;\ e_i\ \rangle$ olarak tanımlanabilir. Burada e_i, kullanıcının web sitesi ziyareti sırasında gerçekleştirdiği tüm etkinliklerdir ve $e_i = \langle\ (t_1, r_1, [ref_1]);\ (t_2, r_2, [ref_2]);\ ...;\ (t_n, r_n, [ref_n])\ \rangle$; $t_{i+1} \geq t_i$ işlemi ile zaman sırasına göre düzenlenen etkinlik dizisi, kullanıcının oturumunu tanımlar. Sunucu günlüğü L için, web sitesi kullanıcılarının ziyaretleri $V = \{vs_1, vs_2, ..., vs_n\}$ temizlenmiş günlük CL 'den alınır ve ardından kullanıcı etkinlik dosyasına yazılır (Srivastava ve ark., 2018).

Oturum tanımlama, web günlüklerinin ön işlemesinde önemli bir sorundur (Fatima ve ark., 2016). Sunucu taraflı bir programlama dili için benzersiz bir oturum, kapatılana kadar aynı tarayıcıdan yapılan tüm isteklerin izlenmesini sağlar. Web sunucuları, kullanıcıları takip etmek ve oturum sürecini uygun şekilde yürütmek için çerezler kullanır. Ancak, web sunucuların erişim kayıtlarını tutma prensibi oturumlara değil, yapılan isteklerin satır satır kaydedilmesine dayanır.

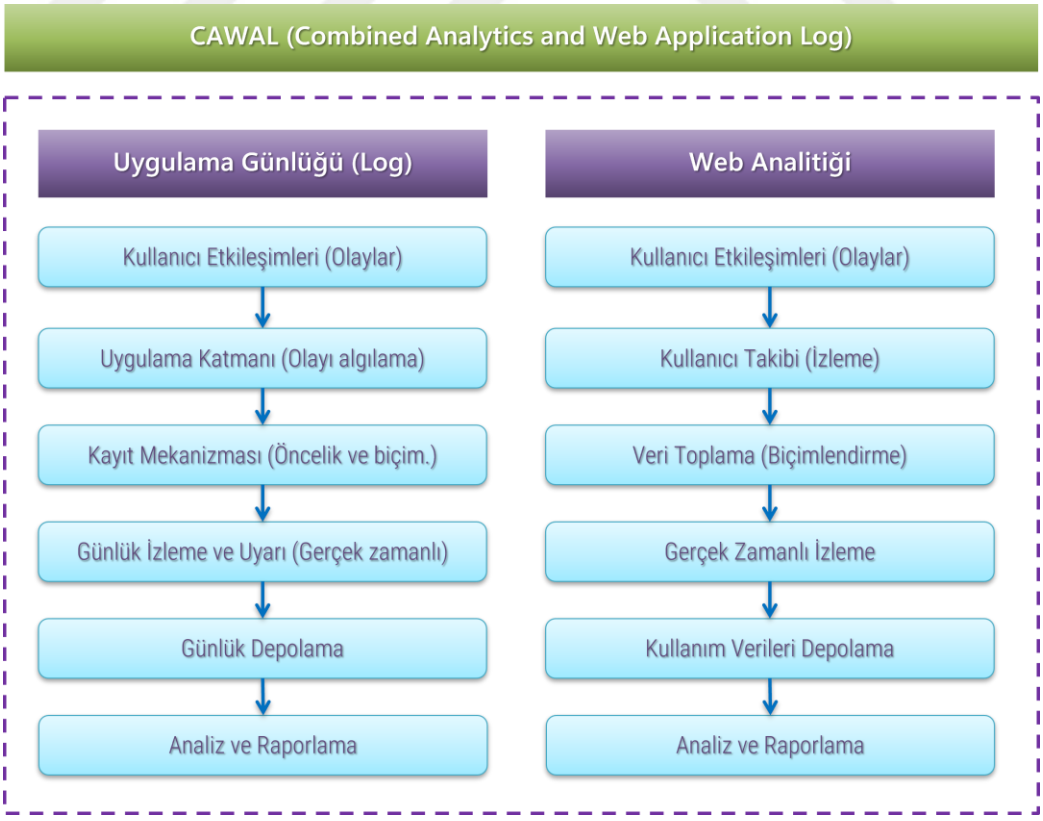
Web sunucusu günlüklerinden kullanıcıları ve oturumları tanımlamak için karmaşık bir dizi işlem gereklidir (Mughal, 2018). Bu sürecin sıkıntılı olmasının temel nedeni, web sunucusu günlüklerinde istemcinin IP adresi ve isteğin zamanı dışında kullanıcıyı veya oturumu ayırt edecek başka bir veri bulunmamasıdır. Web sunucusu günlüklerindeki sınırlı bilgi, örneğin tarayıcının kapatılıp yeniden açıldığını anlamak için yetersizdir. Genellikle, bir günlük kaydı çözümü, aynı IP adresinden ve web tarayıcısı imzasından gelen tüm istekleri tek bir kişiye atfederek ziyaretçi oturumlarını izler. Bu durum, İnternet servis sağlayıcıların oturum boyunca farklı adresler atadığı dinamik IP dağıtımıyla birlikte bir sorun haline gelir (Clifton, 2012).

Son istekten bu yana geçen süre kontrol edilerek kullanıcının ayrı bir oturumda istekte bulunup bulunmadığı tespit edilmeye çalışılır. Kullanıcı bir sayfada belirli bir süreden fazla kalmayacağı için belirli bir süreden sonraki istekler farklı bir oturum olarak değerlendirilir. Bu süre bilimsel çalışmalarda genellikle ardışık sayfalar için 10 dakika, oturum süresi için ise 30 dakika olarak kabul edilmektedir (Varnagar ve ark., 2013). Bir diğer sorun ise gece yarısından önce başlayan ve ertesi gün de devam eden oturumların ayırt edilmesidir (Clifton, 2012). Bütün bu zorluklar, web kullanım madenciliğinde daha etkin ve doğru yöntemlerin geliştirilmesine olan ihtiyacı artırmaktadır.

3. ÖNERİLEN ANALİTİK MODELİ VE ÇERÇEVESİ: CAWAL

Tek bir giriş noktasından erişilen çeşitli hizmetleri bir arada sunan kurumsal web uygulamaları ve portallarında ziyaretçi hareketlerinin takibi ve kullanılan servislerin analizi büyük önem taşımaktadır. Ziyaretçilerin web kullanımlarına dayalı iş geliştirme stratejileri uygulayan kurumlar için bu analizlerin yorumlanması ve gelecekteki eğilimlerin belirlenmesi kritik bir rol oynamaktadır. Ayrıca, elde edilen verilerin doğru şekilde yorumlanması, web geliştiricilere ve yöneticilere kullanıcı katılımı ve etkileşimleri hakkında değerli bilgiler sunar.

Web geliştiricileri genellikle çeşitli amaçlar için uygulamalarında kullanıcı eylemlerini günlüğe kaydederler. Bu çalışmada, bu geleneksel pratik temel alınarak CAWAL adlı yeni bir model ve çerçeve önerisi yapılmıştır. Şekil 3.1'de gösterildiği gibi uygulama günlükleri ve web analitiği süreçleri temelde benzeşen özelliklere sahiptir.



Şekil 3.1. Uygulama günlüğü ve web analitiği süreçleri.

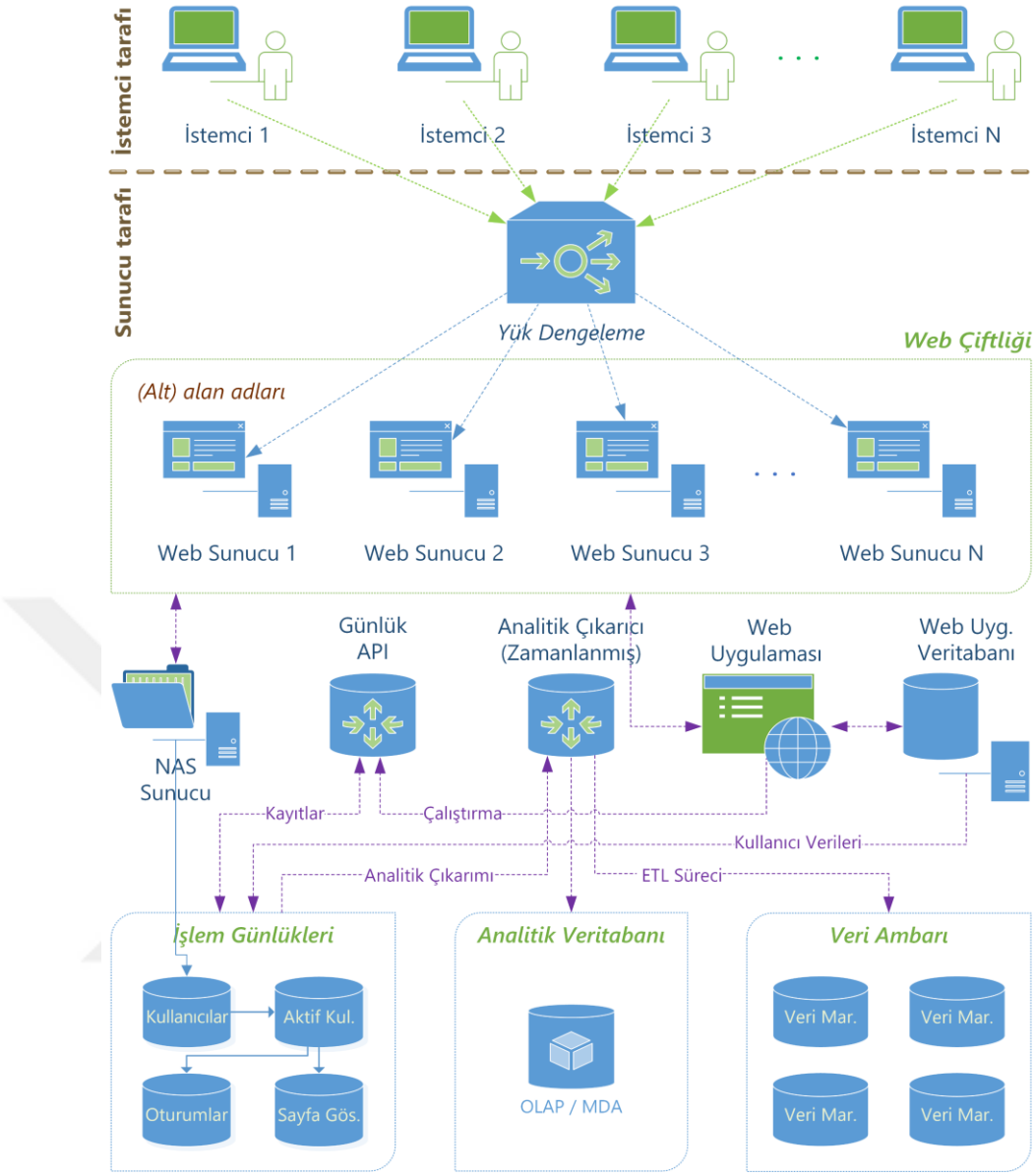
Uygulama günlüğü ile web analitiğini bütünsel bir yaklaşımla birleştiren CAWAL modeli, önerdiğimiz özel veri toplama yaklaşımını (Canay ve Kocabıçak, 2023) kurumsal web portalları ve çok sunuculu mimari ölçeğinde çok daha geniş bir perspektiften ele almaktadır. Modelin pratik bir uygulaması olarak geliştirilen analitik yazılım çerçevesi ise verilerin toplanması ve uygun şekilde depolanması için bir API, verilerin depolanması için bir veritabanı modeli ve analitik bilgileri üretmek için özel bir yöntem içermektedir. Özellikle web portallarında, sayfa hareketlerine dair detaylı verilerin toplanması ve bunların veri madenciliği ile iş zekâsı uygulamalarında kullanılması, bu sayede mümkün ve kolay hale gelmektedir.

CAWAL, piyasadaki bir araç ya da platformdan ziyade, gerçek dünya ortamında uygulanan ve test edilen bir modeldir. Bu nedenle, onu diğer analitik uygulamaları gibi bir araç ya da platform olarak adlandırmak yerine, geniş bir kavramsal bağlamda model ve çerçeve olarak tanımlamak daha doğru bir yaklaşımdır. Belirli bir süreci veya olguyu temsil eden bir soyutlama olarak tanımlanan model bağlamında CAWAL, web analitiği ve günlük yönetimi için veri toplama, depolama, analiz ve raporlama prosedürlerini içeren yapıyı tanımlamaktadır.

CAWAL aynı zamanda araştırmamızın hipotezini ampirik olarak test etmek ve modelin veri doğruluğu, veri sahipliği ve yönetişimi, verimlilik ve çoklu sunucu uyumluluğu gibi kritik alanlarda mevcut web analitiği araçlarına göre avantajlarını göstermek için bu model temel alınarak geliştirilen yazılım çerçevesinin de adıdır. CAWAL çerçevesi, önerilen kavramsal modeli veri toplama API'si, veritabanı yapısı, analitik oluşturma ve kurumsal düzeyde web analitiği işlemleri için özel olarak tasarlanmış yardımcı bileşenlerin bir kombinasyonu aracılığıyla uygulamaktadır.

3.1. CAWAL Modeli ve Çerçeve Yazılımının Tasarımı

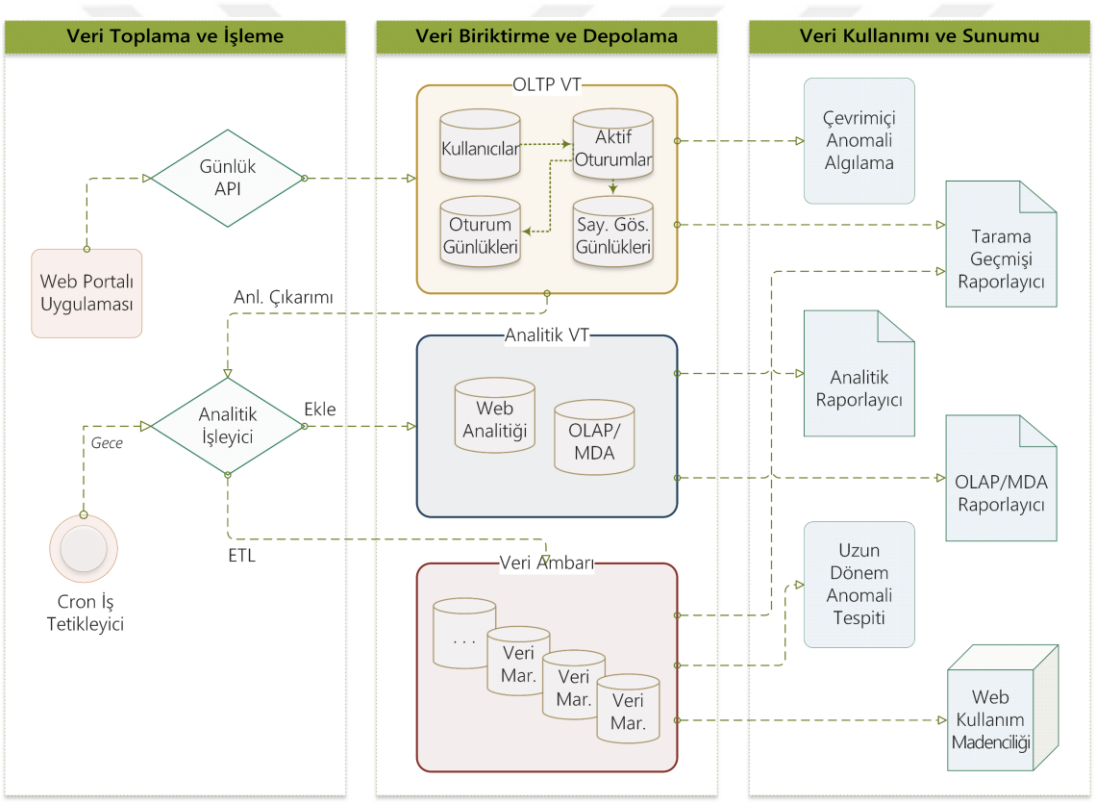
CAWAL, çoklu sunucu destekli web analitiği için kapsamlı bir mimari olarak hizmet vermek üzere tasarlanmış bir modeldir. Sunucu tarafı çalışacak şekilde yapılandırılan model, gelen talepleri çeşitli alt alanlardan oluşan bir web çiftliğine dağıtmak için sunucu alanında bir yük dengeleyiciyi destekler. Şekil 3.2, mimarinin farklı katmanlarını ve işlevlerini tanımlayan ayrıntılı bir gösterimini sunmaktadır.



Şekil 3.2. Önerilen çoklu sunucu destekli CAWAL modeli mimarisi.

CAWAL'ın kavramsal tasarımının ardından, modeli uygulanabilir kılmak için Linux sunucularda çalışacak bir çerçeve yazılım geliştirilmiştir. Mimari, gelişmiş performans ve ölçeklenebilirlik elde etmek için Linux'un sağlam ve güvenli ortamından yararlanmaktadır. Mimariye göre, çok sunuculu bir kurumsal web ortamında, istemcilerden gelen talepler bir yük dengeleyici tarafından web çiftliğinin uygun sunucuları arasında eşit olarak dağıtılır. Bu sunucular, bir NAS (Ağa Bağlı Depolama) sunucusu üzerinde paylaşılan yapılandırma dosyalarına ve oturum dizinlerine erişebilir. Web uygulaması bu isteklere yanıt verdikçe, uygulamaya gömülü olarak çalışan günlük API'si bu etkileşimleri eşzamanlı olarak günlük tablolarına kaydeder.

Bir zamanlanmış görev tarafından her gece tetiklenerek çalıştırılan analitik çıkarım betiği ise bu verileri özel bir analitik veritabanında işler. Çerçeve ayrıca veri analitiğini takiben bir ETL süreci (Al-Rahman ve ark., 2023) içermekte ve sonuçları daha sonraki gereksinimler için bir veri ambarında biriktirmektedir. CAWAL çerçevesi, gerçek zamanlı kullanıcı verilerinin işlenmesi için bir OLTP veritabanını (Li ve ark., 2022) ve kümülatif veri depolama için bir veri ambarını entegre ederek karmaşık web analitiği işlemleri için ideal bir platform sunar. Modele yönelik süreçler, Şekil 3.3'te gösterildiği gibi verilerin toplanması ve işlenmesi, biriktirilmesi ve saklanması, kullanımı ve sunumu şeklinde üç bölümde ele alınabilir. Model, anlık veri toplama ile uzun vadeli analitik depolamayı birleştirerek çağdaş web analitiği zorlukları için uyarlanabilir bir çözüm ortaya koyar.



Şekil 3.3. CAWAL modelinin veri işleme hattını gösteren süreç diyagramı.

CAWAL çerçevesi, nesne yönelimli PHP ile geliştirilmiştir ve MySQL ilişkisel veritabanıyla entegre çalışır. PHP'nin oturum yönetimi de dahil olmak üzere işlevleri sunucu tarafı işlemlerinin omurgasını oluştururken, MySQL verimli veri depolama ve sorgu yürütme için birincil veritabanı sistemi olarak hizmet vermektedir (Alya ve Ikhwan, 2022).

MySQL indekslemedeki özel optimizasyonlar, gerçek zamanlı analitik için temel bir özellik olan yüksek hızlı veri alımı için tasarlanmıştır. Çerçeve, güvenilir veri kalıcılığı için MySQL'in ACID (Atomicity, Consistency, Isolation, and Durability) uyumlu işlem yeteneklerini kullanmakta ve veri işleme, istatistik ve zaman serisi analizi için yerleşik SQL işlevlerinden yararlanmaktadır. Mimari ayrıca, bir API (Ofoeda ve ark., 2019) aracılığıyla dağıtılmış yakalama aracılarını da barındırır ve ölçeklenebilirlik, güvenilirlik ve performansı dengeleyen kapsamlı bir web analitiği çözümü ortaya çıkarır.

3.2. Veri Modelinin Tasarımı

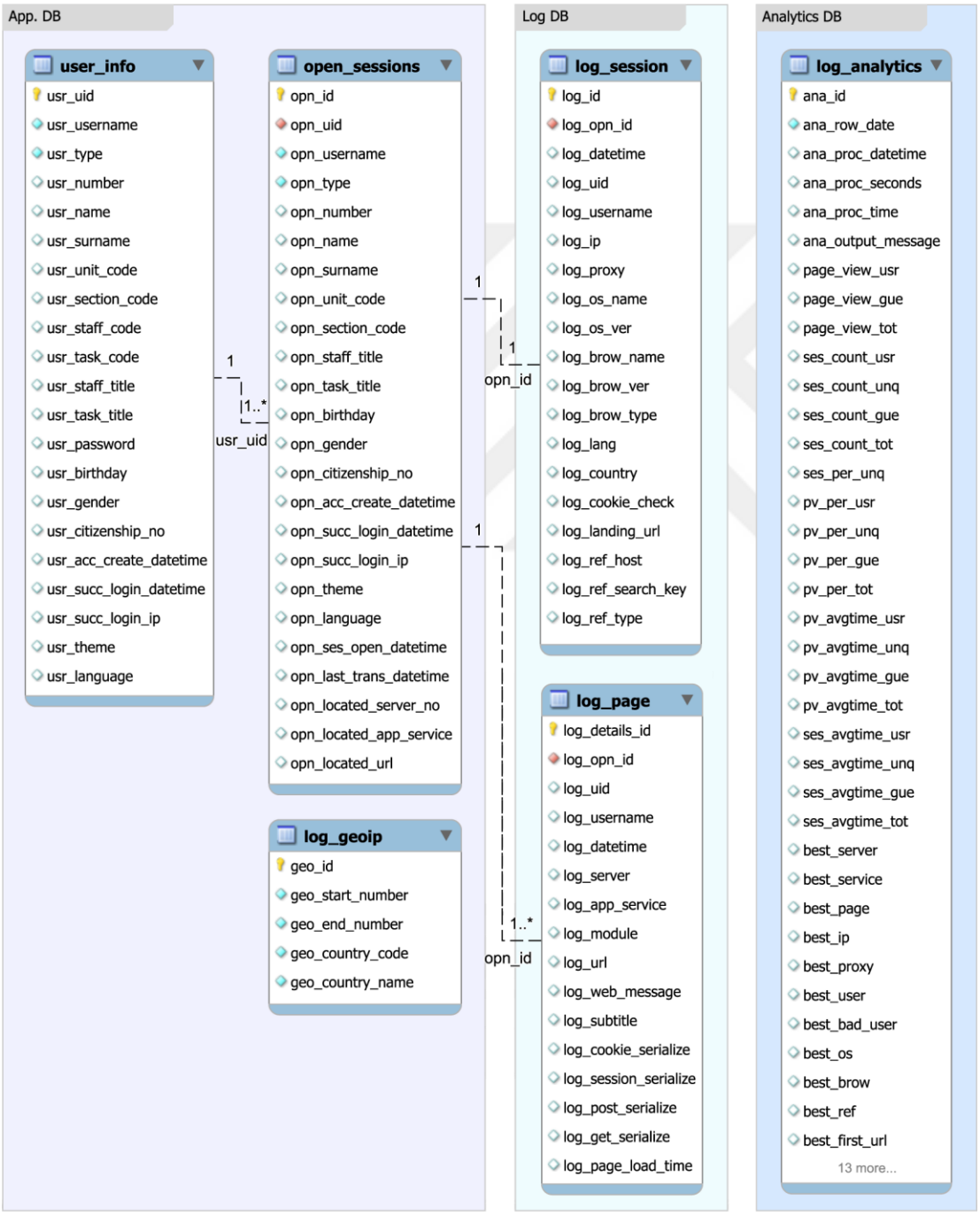
Web uygulamalarında günlük işlemler sırasında üretilen operasyonel veriler, sürekli veri giriş-çıkış işlemlerinin yapıldığı, yazma ağırlıklı OLTP veri tabanlarında tutulmaktadır (Milosevic ve ark., 2021). CAWAL, depolama ek yükünü en aza indirirken verimli analitik sağlamak için optimize edilmiş bir yalın veri modeli uygulamaktadır. Şekil 3.4, CAWAL modelinin veri akışını ve işleme süreçlerini detaylandıran bir şema sunmaktadır.

Şekil 3.4. CAWAL modelinde veri akışı ve işleme süreçleri.

CAWAL modelinin veri akışı ve işleme süreçlerini ayrıntılı bir biçimde ortaya koyan şemanın veri kaynakları bölümünde HTTP istekleri ve ağ protokolleri gibi çeşitli veri kaynakları gösterilmiştir. Veri toplama aşamasında veri füzyonu ile bir araya getirilen kullanıcı ve kullanım verileri işlenmeye hazır hale getirildikten sonra veri saklama (OLTP) sistemlerinde muhafaza edilir. Bu veriler gerçek zamanlı olarak sistem izleme için kullanılabileceği gibi makine öğrenmesi algoritmaları aracılığıyla anomali tespiti ya da çeşitli tahminler için de kullanılabilir. Diğer yandan, OLTP veritabanlarında saklanan veriler yığın işleme adımında analitik çıkarımı için kullanılmalarının ardından ETL süreçleri ile dönüştürülerek veri ambarına kaydedilir. Veri ambarında

biriktirilen veriler daha sonra gerçekleştirilecek WUM analiz ve tahminleri için değerli bir kaynak olarak işlev görür.

Modele uygun olarak tasarlanan çerçeve, verilerin yapılandırılmış bir şekilde elde edilmesi ve depolanması için bir API ve veri saklama için bir veritabanı şeması içermektedir. Şekil 3.5'teki UML diyagramı, analitik veri tabanı da dahil olmak üzere mantıksal veri modelinin kapsamlı bir görselleştirmesini sunmaktadır.



Şekil 3.5. Mantıksal veri modelinin UML diyagramı.

UML diyagramında yer alan tabloların kullanım amaçları aşağıdaki gibidir:

- user_info: Web uygulaması tarafından yönetilen kullanıcı verileri.
- open_sessions: Devam eden (açık) oturumların verileri.
- log_geoip: IP aralıkları olarak tanımlanmış ülke verileri.
- log_session: Oturum bilgilerini içeren istemci verileri.
- log_page: Oturumdaki her istek için sayfa verileri.
- log_analytics: Gün bazlı hesaplanmış analitik verileri.

Oturum yönetimi süreci, uygulama ve web sunucusunun bütünleşik bir şekilde çalışmasıyla gerçekleştirilir. Web sitesine ilk kez bir istek yapıldığında, sunucu tarafı programlama dili tarafından istemci için otomatik olarak 32 karakterlik yeni bir alfanümerik oturum değişkeni oluşturulur. Öte yandan, web uygulaması devam eden oturumları yönetmek için veritabanında yer alan "open_sessions" (açık oturumlar) tablosunu kullanır. Günlük tablolarındaki tüm oturum ve sayfa verileri, bu tablodaki her kayıt için otomatik olarak artan bir sayıdan oluşan "opn id" adlı bir yabancı anahtar aracılığıyla ilişkilendirilir. Kullanıcı tanımlaması, uygulama verilerine de erişebilen API tarafından gerçekleştirilir.

Çerçeveyi destekleyen veri toplama altyapısı ve kullanım verilerini barındıran iki temel OLTP veritabanı tablosu "log_session" ve "log_page" dir (Canay ve Kocabıçak, 2023). Kullanıcı oturum bilgilerinin kaydedildiği "log_session" tablosu, sistemdeki kullanıcı etkileşimlerini anlamak ve izlemek için kritik bir bileşendir. Her bir kullanıcı oturumunu ayrı bir satır olarak temsil edecek şekilde tasarlanan bu tablo, kullanıcı ve teknik ortamla ilgili kapsamlı bilgiler barındırır. Web analitiğinin anahtarı olan "log_session" tablosu oturum kimliği, kullanıcı kimliği, IP adresi ve vekil sunucu (proxy) adresi gibi ayrıntıları içerir. Tablo ayrıca, işletim sistemi ve sürümü, tarayıcı adı, sürümü, türü, dili, ülkesi (GeoIP tablosu aracılığıyla), site içinde erişilen ilk URL, yönlendiren site, -arama motoru aracılığıyla geldiyse- kullanılan arama motoru ve arama anahtar kelimeleri gibi bilgileri de kapsar. Bu veritabanı tablosu, çeşitli parametrelerin detaylı şekilde kaydedilmesiyle birlikte birçok amaca hizmet edecek şekilde tasarlanmıştır.

Oturum tablosunun yanı sıra, kullanıcıların sayfa görüntüleme etkinliklerini detaylandıran "log_page" tablosu da hem web analitiği hem de web kullanım madenciliği süreçlerinde merkezi bir rol oynar. İstek yapılan her sayfada, tablo

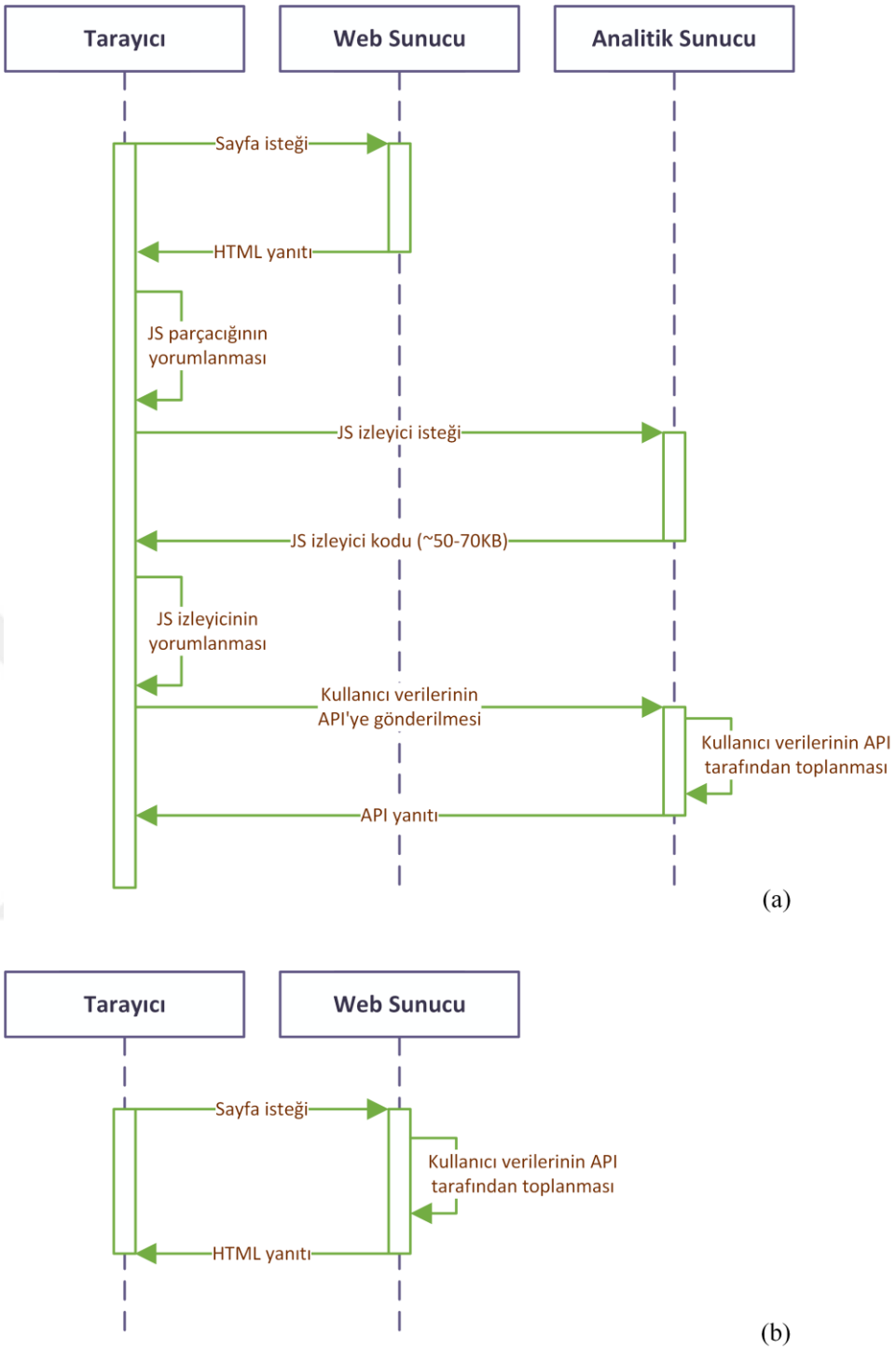
satırındaki en az bir bilgi değiştiği için veri modeli bu yapıda kurgulanmıştır; böylece kullanıcı ve oturum tespiti tam başarıyla sağlanmaktadır. Çerçevenin işlevselliğini artırmak için özel olarak tasarlanan "log_page" tablosu ayrıca uygulamanın çeşitli yönleri hakkında ayrıntılı bilgi sağlayarak hata tespit süreçlerinin kolaylaştırılmasına ve yazılım kalitesinin iyileştirilmesine yardımcı olur. Diğer taraftan, web yöneticisi, sistem kullanımını izlemek ve gerçek zamanlı analiz bilgileri aracılığıyla içgörüler elde etmek için bu veritabanı tablolarından yararlanabilir. Çerçevenin dikkat çeken bir özelliği de web geliştiricileri için kritik bir faktör olan sayfa üretim sürelerinin elde edilmesidir. Sayfa gösterimleri tablosundaki her bir istek kaydında yer alan bu bilgi, yazılım geliştiricilerin veritabanı sorguları ya da kod kaynaklı yavaşlık ile sunucu erişim sorunlarını tespit etmelerini sağlayarak performans optimizasyonunu geliştirir.

3.3. Veri Toplama Süreci

CAWAL'ın veri toplama yeteneğinin temel bileşenleri olan metrikler ve boyutsal veriler dört farklı kaynaktan elde edilmektedir: HTTP isteği, ağ seviyesi, uygulama seviyesi ve harici veriler (Canay ve Kocabıçak, 2023). HTTP istek verileri, HTTP başlıkları aracılığıyla protokolün istek mesajlarından yakalanır. Sunucu tarafından oluşturulmasına ve HTTP istekleriyle ilişkilendirilmesine rağmen, istemci IP adresi gibi ağ düzeyindeki veriler HTTP isteğinin bir parçası değildir, ancak başarılı iletimler için gereklidirler. HTTP istekleriyle birlikte iletilen uygulama düzeyindeki veriler, web programlama dili tarafından yönetilen oturum ve yönlendirme bilgilerini içerir. Tipik olarak web uygulaması veritabanında bulunan harici veriler, diğer kategorilerle ilişkilendirilebilir ve IP adresinden türetilen ülke kimliği ve kullanıcı profili bilgileri gibi unsurları içerir. Kullanıcı takibi sürecinde farklı kaynaklardan toplanan bu bilgiler veri füzyonu ile bir araya getirilir ve daha sonra işlenerek metriklere ve diğer değerli bilgilere dönüştürülür.

3.3.1. Kullanıcı takibi ve veri toplama

CAWAL çerçevesi, verileri almak, yorumlamak, biçimlendirmek ve depolamak için sunucu tarafı ve uygulama tabanlı çalışan bir veri toplama API'si aracılığıyla birden fazla kaynaktan bu tür bilgileri toplar ve işler. Şekil 3.6'da gösterildiği gibi, üçüncü taraf araçlar kullanıcı takibi ve veri toplama için günlük sunucusuna fazladan iki istek gönderirken, CAWAL yalın yapısı nedeniyle sunucuyla çok daha az etkileşime girmektedir. Bu sayede çerçeve, verimliliği artırırken sunucu yükünü azaltmaktadır.

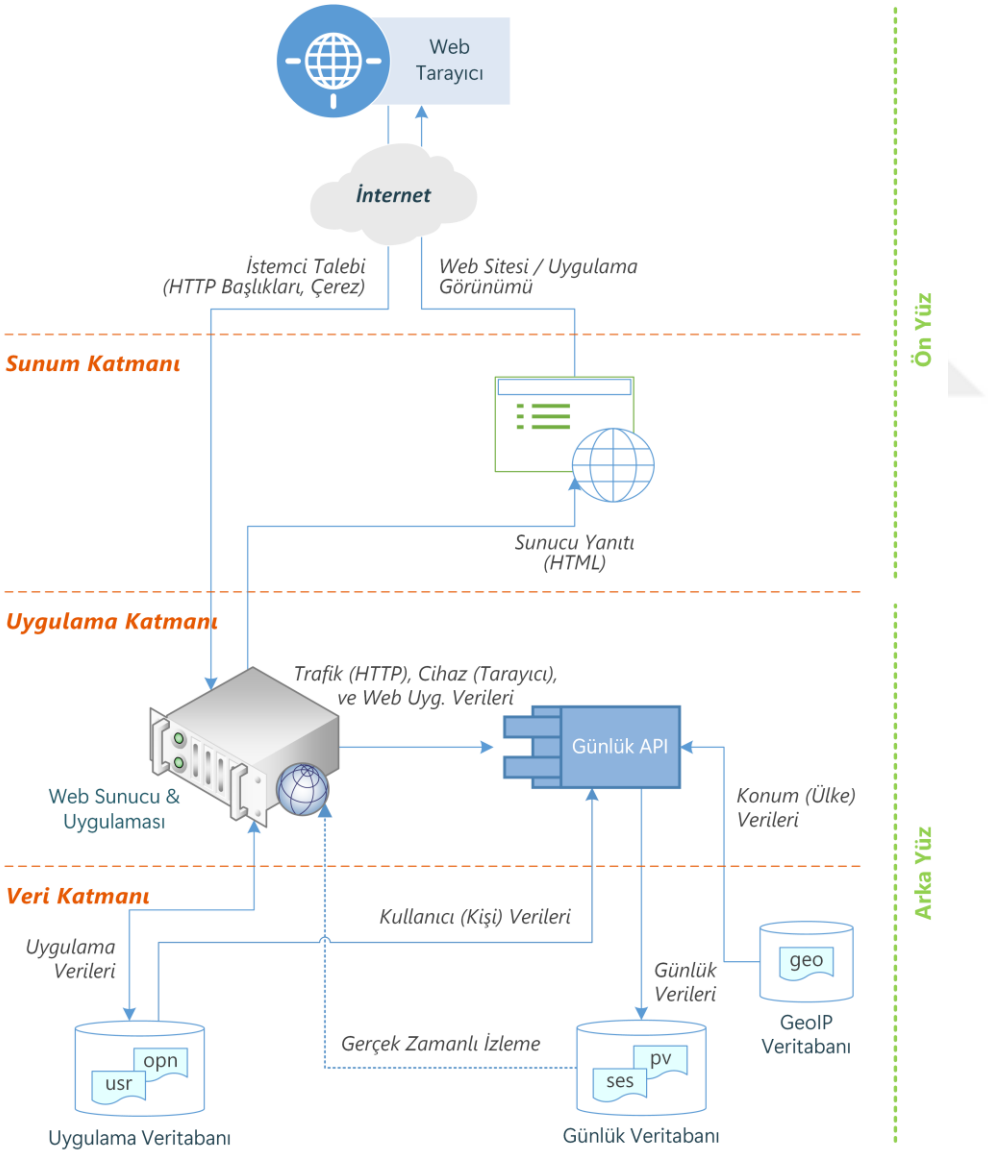


Şekil 3.6. İstemci (a) ve sunucu (b) tarafı kullanıcı takibi sıra diyagramları.

Önerilen veri toplama metodolojisi, geleneksel web sunucusu günlüklerinden, bulut tabanlı sistemlerden ya da istemci tarafı diğer yöntemlerden ayrışarak, sunucu tarafı ve uygulama tabanlı çalışma esasına dayanmaktadır. Bu yaklaşım, veri toplama, yorumlama, biçimlendirme ve depolama süreçlerinde uygulama tabanlı bir API'ye güvenerek, daha derinlemesine ve bütünleşik bir veri analizi imkânı sunar.

3.3.2. Veri toplama yöntemin üç katmanlı organizasyonu

Üç katmanlı model kapsamında geliştirilen API, bilgi güvenliğinin gizlilik, bütünlük ve erişilebilirlik ilkelerine bağlı kalarak yoğun ziyaretçi trafiği altında kullanılabilirlik ve güvenilirliği sağlamak üzere tasarlanmıştır. API, web uygulamalarının yazılım çerçeveleri ile entegrasyonu düşünülerek oluşturulmuş ve bu sayede kullanıcı etkileşim verilerinin başarılı bir şekilde toplanması hedeflenmiştir. Yöntemin üç katmanlı organizasyonu ve log API'sinin web uygulamalarına nasıl entegre edilebileceği Şekil 3.7'de sunulmuştur. Bu yapı, uygulama geliştiricilerine kullanıcı davranışlarını anlama ve analiz etme konusunda güçlü bir araç sunarken aynı zamanda veri toplama süreçlerinde esneklik ve ölçeklenebilirlik sağlar.



Şekil 3.7. Önerilen veri toplama yöntemin üç katmanlı organizasyonu.

Web tabanlı sistemlerin verimliliği ve kullanıcı deneyimi, altında yatan mimari organizasyonun ne kadar iyi tasarlandığına büyük ölçüde bağlıdır. Bu bağlamda, önerilen veri toplama yönteminin üç katmanlı organizasyonu, sistemlerin bu iki kritik bileşeni arasında bir denge kurmayı amaçlamaktadır. Her bir katman, web tabanlı uygulamaların farklı yönlerini ele alır ve birlikte çalışarak genel bir çözüm oluşturur. İlk katman olan sunum katmanı, kullanıcıların doğrudan etkileşimde bulunduğu arayüzleri kapsar. Bu katman, HTML, CSS ve JavaScript gibi ön yüz web teknolojilerini kullanarak, sunucu tarafından üretilen ve kullanıcı tarafından tarayıcıda görüntülenen web uygulamalarının kullanıcı arayüzlerini barındırır. Kullanıcı deneyiminin temelini oluşturan bu katman, uygulamanın estetik ve işlevsel yönlerini birleştirir.

Uygulama katmanı, sunucu üzerinde çalışan web uygulaması ve log API'si dahil olmak üzere, istemci ve sunucu arasındaki etkileşimi yönetir. Bu katman, HTTP, ağ ve web uygulaması gibi çeşitli kaynaklardan gelen istemci ve istek verilerini toplar ve işler. Uygulama katmanının, hem kullanıcı arayüzünü sunan sunum katmanı hem de verilerin saklandığı veri katmanı ile doğrudan bir ilişkisi vardır. Son katman olan veri katmanı, uygulamanın kalıcı veri depolama işlemlerini içerir. Bu katmanda, uygulama verileri, GeoIP bilgileri ve log API veritabanı gibi veriler saklanır. Veri katmanı, uygulama ve veritabanlarının modele göre çapraz erişimine olanak tanır ve sistem verimliliği için kritik bir rol oynar.

3.3.3. Veri kaynakları ve bilgi füzyonu

Geleneksel web analitiği araçlarından farklı olarak, bu yöntemde kullanıcı takibi için istemci tarafı web teknolojileri kullanılmamıştır. Günlük API'si sunucu tarafında çalışır ve yazılım çerçevesinin bir parçası gibi hareket eder. Uygulama düzeyinde etkileşim kayıtları tutmak, web uygulamalarının işlevsel bir özelliği olarak çok uzun zamandır kullanılan yaklaşımdır. CAWAL bu yaklaşımı geliştirerek web analitiği özelliklerini de kapsayacak şekilde daha ileriye taşımıştır. Web analitiğinin ilk aşaması olan kullanıcı takibi için CAWAL tarafından toplanacak veriler temel olarak HTTP istek verileri, ağ düzeyinde veriler, uygulama düzeyinde veriler ve harici veriler olmak üzere dört ana kategoriden birine aittir. Günlük API'si her istekle birlikte Tablo 3.1'deki ana verileri toplamaktadır.

Tablo 3.1. Toplanan temel veriler ve kaynakları.

Toplanan Veriler	Veri Kaynağı
İstemci profili/Kullanıcı aracısı (tarayıcı, işletim sistemi)	HTTP isteği
İstemci IP adresi	Ağ protokolü
Yönlendiren URL adresi (önceki web sayfası)	HTTP isteği
Yönlendirme (kanal tanımlama)	Uygulama
Coğrafi Konum	Harici
Sayfa görüntüleme	HTTP isteği
Kullanıcı profili	Uygulama
Uygulamaya özel veriler	Uygulama
Ziyaret veya oturum	Uygulama

Burada HTTP isteği ve ağ protokolü kaynaklarından elde edilen veriler API'ye sunucu tarafı programlama dili tarafından sağlanır. API ayrıca, yazılım çerçevesi tarafından paylaşılan kullanıcı bilgilerini içeren uygulama verilerine de erişir. İstemci coğrafi konum (ülke) bilgisi, bir SQL sorgusu ile GeoIP tablosundaki IP veri aralığından alınır. Günlük API'si, tarayıcı ve işletim sistemi bilgilerini içeren kullanıcı aracısı verilerini topladıktan sonra çeşitli dize işleme işlevlerini kullanarak tarayıcı adı, tarayıcı sürümü, işletim sistemi adı ve işletim sistemi sürümü gibi ayrıntılı verileri elde eder. Tablodaki diğer veriler ise işlenmeden, doğrudan kaynaklarından elde edildiği şekilde saklanır.

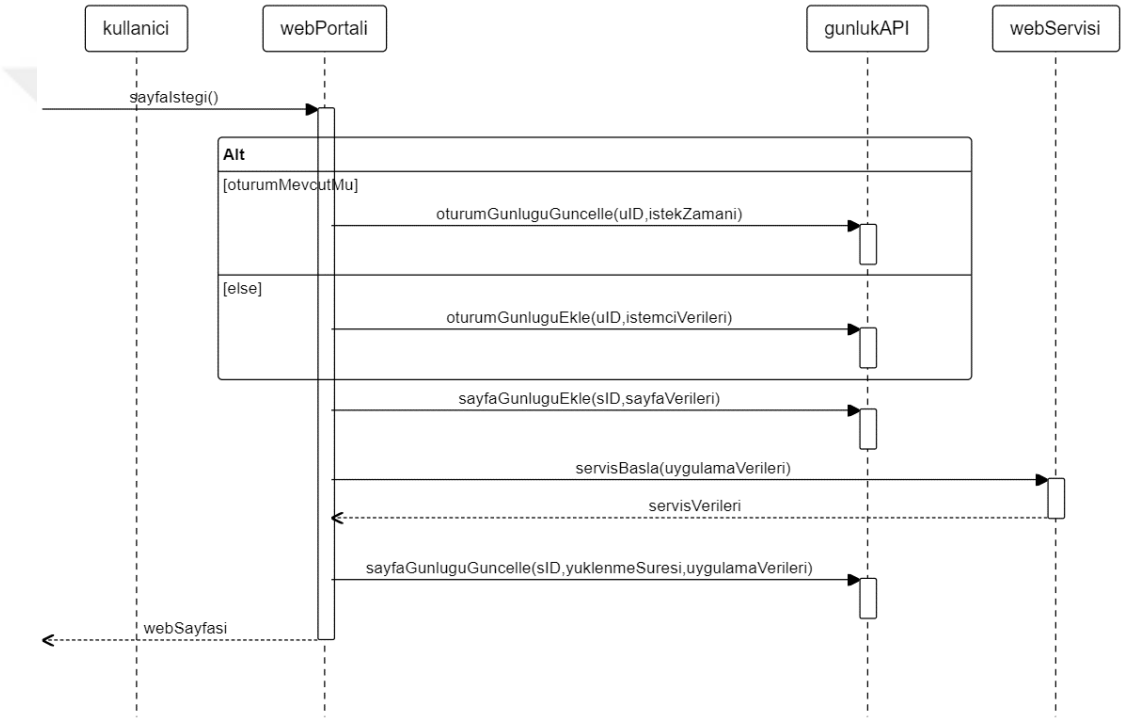
Yöntem, alışılagelmiş ya da basit çözümler yerine daha gelişmiş ve standart bir yapı ortaya koymaktadır. Böylece web kullanım madenciliği sürecinin ön işleme aşamasında veri dönüştürme, veri temizleme, veri filtreleme, kullanıcı ve oturum tanımlama, yol tamamlama ve veri biçimlendirme işlemlerine ihtiyaç olmadan örüntü keşfi aşaması için uygun yapıda veri sağlanmış olur. Ayrıca, web sunucusu günlüklerinin içermediği birçok nicel ve nitel veri, daha sonra işlenmek üzere yapılandırılmış bir biçimde kaydedilir. Bu veriler anlık sistem izleme, analitik çıkarma ve kullanıcı davranış analizi gibi çeşitli amaçlar için de kullanılabilmektedir.

Modelin öne çıkan bir diğer özelliği ise dinamik IP'lere ait oturumları ayırt edebilmesidir. Bu yöntemde oturumlar web sunucusu tarafından oluşturulmakta ve iki farklı cihaz aynı IP adresini kullansa bile ayrı oturumlar olarak tanımlanmaktadır. Bu oturumlar, kullanıcı sistemden güvenli çıkış yaptığında veya belirli bir süre sayfa aktivitesi olmadığında devreye giren uyarı penceresine yanıt vermediğinde sonlandırılmaktadır. Ayrıca gece yarısından önce başlayıp ertesi güne devam eden oturumlar, diğer analitik çözümlerinden farklı olarak gün sonunda sonlandırılmamakta

ve devam eden oturum olarak değerlendirilmektedir. Tüm bu özellikler modelin geleneksel yöntemlere göre çok daha başarılı kullanıcı ve oturum tanımlaması yapmasını ve yüksek veri doğruluğu sunmasını sağlamaktadır.

3.3.4. Günlük API yapısı ve çalışma prensibi

Web portalları ve büyük ölçekli kurumsal web uygulamaları genellikle bir yazılım çerçevesi olarak tasarlanır. CAWAL'ın günlük (log) API'si, ana uygulama çerçevesi kodunun içerisine, baş kısımda çağırılacak şekilde yerleştirilir. Böylece yapılan her sayfa isteği için API de çalışmaktadır. API'nin uygulama katmanındaki çalışma yöntemi Şekil 3.8'de gösterilmiştir.



Şekil 3.8. Günlük API çalışma yönteminin sıra diyagramı.

Yazılım çerçevesi uygun bir şekilde oluşturulduktan ve API doğru bir şekilde entegre edildikten sonra, uygulama geliştiricilerin günlük işlemleriyle hiçbir ilgisi kalmamaktadır. Geliştiriciler, uygulama kodlarını yazıp yayınlarken, API her zaman katmanlı çerçevenin bir parçası olarak çalışır. Ana yazılım çerçeve kodu bitmeden önce aktif sayfa üzerinde gerçekleştirilen işlemlerle ilgili uygulama verileri sayfa log tablosunda güncellenir. Bu sayede sayfa yüklenme süresi, veritabanı sorgularındaki gecikmeler, oluşan hatalar gibi uygulamanın iyileştirilmesine yönelik veriler sayfa gösterimi tablosunda saklanır.

Veri toplama süreci hem nitel hem de nicel bilgilerin toplanması ve analiz edilmesine dayanır. Niteliksel özellikler veya boyutlar, istemcinin tarayıcısı ve işletim sistemi, IP adresi, önceki web sayfaları, kanal tanımlama, coğrafi konum ve kullanıcı profili gibi hususları içerir. Buna karşılık nicel ölçüm ve metrikler, HTTP isteklerinden veya harici kaynaklardan elde edilen sayfa görüntülemelerini ve uygulamaya özgü verileri içerirler. Bu boyutların ve metriklerin çeşitli temizleme, çıkarma veya hesaplama işlemlerinden geçirildikten sonra CAWAL çerçevesine dahil edilmesi, kullanıcı etkileşimlerinin ve davranışlarının kesin bir şekilde anlaşılmasını sağlar. API'nin veri toplama süreci, istek sayısına bağlı olarak *O(N)* zaman karmaşıklığı gösterirken, yüksek trafik durumunda sunucunun işlemci gücü ve depolama ihtiyacı da isteklere orantılı olarak artmaktadır. Çerçeve, uygun kaynaklarla desteklendiğinde ölçeklenebilir bir yapı sunarak işlem kapasitesini genişletebilmektedir.

Günlük kayıt API'si ilk aşamada ağ protokolü, web tarayıcı, web uygulaması, uygulama veritabanı ve GeoIP veritabanı gibi farklı kaynaklardan istemci verilerini elde eder ve veri füzyonu ile birleştirir. Toplanan veriler bir sonraki aşamada temizlendikten ve işlendikten sonra veri modeline göre oturum ve sayfa görüntüleme veritabanı tablolarında saklanır. Açık oturumları barındıran "open_sessions" veritabanı tablosu, oturumları yönetmek ve aktif kullanıcıları gerçek zamanlı olarak takip etmek için kullanılır. Uygulamaya giriş yapan her kullanıcı için bu tabloda bir kayıt oluşturulur. Kullanıcı çıkış yaptığında veya zaman aşımı süresi dolduğunda ilgili kayıt silinir. Böylece sistemde dolaşan ziyaretçileri içeren tablonun sürekli güncel olması sağlanır. Sistemin gerçek zamanlı izlenmesi bu tablo aracılığıyla sağlanır.

3.4. Veri Temizleme ve İşleme Süreci

Günlük API'si veri toplamanın yanı sıra verileri temizleme ve saklama işlevlerini de yerine getirir. İlk olarak, elde edilen tüm veriler temizlendikten ve veri türüne göre biçimlendirildikten sonra veritabanında depolanacak bir forma dönüştürülür. Bu sırada tarayıcı verileri ve istek türü gibi bazı bilgiler kodlanır. Buna ek olarak, istemcinin konumunu belirlemek için ülke bilgisi IP adresi aracılığıyla GeoIP tablosundan elde edilir. Daha sonra oturumla ilişkili tüm yapılandırılmış veriler, API tarafından yürütülen sorgular aracılığıyla referans bütünlüğüne sahip oturum ve sayfa günlüğü tablolarına eklenir.

CAWAL çerçevesi içindeki oturumlar, kullanıcıların web sitesindeki her ziyareti tanımlayan ve izleyen veri kayıtlarından oluşur. Burada, $S = \{s_1, s_2, ..., s_n\}$ şeklinde ifade edilen her bir ziyaretçiye ait oturumlar kümesi bulunur. S'de yer alan her bir s_i oturum verisi, OLTP veritabanının oturum günlüğü tablosunda belirli bir satırda saklanır. Oturumlar, kullanıcı nitelikleri, işletim sistemi, web tarayıcı, IP adresi, yönlendirici, arama motoru gibi sayfalar arasında gezinirken sabit kalan ilk varış verilerinden oluşur. Tablo 3.2, CAWAL çerçevesinin kullanıcı oturumlarıyla ilgili verileri toplamak için kullandığı kapsamlı yaklaşımı yansıtan "log_session" veritabanı tablosundan örnek bir kaydı göstermektedir.

Tablo 3.2. Oturum tablosu alanları ve örnek bir veri satırı.

Alan Adı	Kaynak	Örnek Veri
log_id	Veritabanı	19423375
log_opn_id	Uygulama	21091539
log_datetime	Web sunucu	10.05.2020 22:42
log_uid	Veritabanı	540
log_username	Veritabanı	canay
log_ip	HTTP	193.140.253.81
log_proxy	HTTP	185.169.181.13
log_os_name	Web tarayıcı	Windows
log_os_ver	Web tarayıcı	10
log_brow_name	Web tarayıcı	Chrome
log_brow_ver	Web tarayıcı	81
log_brow_type	Web tarayıcı	1
log_lang	HTTP	tr
log_country	GeoIP VT	Turkiye
log_cookie_check	HTTP	1
log_landing_url	Web sunucu	https://www.gate.sakarya.edu.tr/
log_ref	HTTP	Google
log_ref_host	HTTP	www.google.com.tr
log_ref_search_key	HTTP	sau gate
log_ref_type	HTTP	4

Oturum tablosu, kullanıcı davranışlarındaki kalıpları ve eğilimleri tespit etmeye, kullanıcıların demografik dağılımını anlamaya ve tarayıcı tercihleri ile işletim sistemi

kullanımı gibi teknik bilgileri elde etmeye aracılık eder. İlk URL'yi ve yönlendiren siteleri izleme yeteneği, kullanıcı gezinme modellerinin ve kaynak tercihlerinin daha derinlemesine anlaşılmasına olanak tanıyan başka bir karmaşıklık katmanı ekler.

Çerçevenin bir diğer temel veritabanı tablosu olan “log_page”, sayfalar arasındaki işlemleri takip etmek için kapsamlı bilgileri sağlar. Tablo, $P_s = \{p_1, p_2, ..., p_m\}$ olarak ifade edilen ve bir oturumdaki sayfa isteklerinin kümesini temsil eder. Bir P_s'deki her p_i, sayfa günlüğü tablosunda benzersiz bir oturum kimliği ile ilişkilendirilerek ayrı bir kayıt olarak saklanır.

Tablo 3.3. Sayfa gösterimi tablosu alanları ve örnek bir veri satırı.

Alan Adı	**Örnek Veri**
log_details_id	114803693
log_opn_id	21091539
log_uid	540
log_username	canay
log_datetime	2020-10-05 22:42:38
log_date	2020-10-05
log_server	7
log_app_service	gate
log_module	info
log_url	http://www.gate.sakarya.edu.tr/?page=info
log_web_message	Welcome to WebGate
log_subtitle	Info :: You can review your access and security information here
log_cookie_serialize	a:13:{s:6:"AUTHID"; s:36:"56b01c19-6aa8-497b-9290-ca8ebb9e6bbc"; s:6:"Apache"; s:27:"10.9.54.19.1630566730862646"; s:9:"PHPSESSID"; s:32:"5282b8b77b0b2cbfe2e854b459f31a7b"; s:10:"cookie sid"; s:32:"1765c04ed3a62487d1df2790e078a1ad"; s:11:"cookie theme"; s:1:"0"; s:10:"cookie lang"; s:1:"0"; s:12:"cookie limit"; s:2:"15"; }
log_session_serialize	a:8:{s:7:"ses uid"; s:6:"166553"; s:6:"ses id"; s:5:"81291"; s:8:"ses theme"; seris:1:"0"; s:8:"ses lang"; s:1:"0"; s:9:"ses limit"; s:2:"15"; s:12:"ses security"; N; s:9:"ses child"; b:0; s:8:"ses error"; s:0:""; }
log_post_serialize	a:0:{}
log_get_serialize	a:1:{s:4:"page";s:4:"info";}
log_page_load_time	0,0266

Otomatik olarak artan istek kimliği (ID), oturum kimliği, kullanıcı kimliği ve adı, işlem süresi, sunucu kimliği, uygulama hizmeti, sayfa/modül ve URL gibi birincil veriler "log_page" tablosunda tutulur. Bu veritabanı tablosu aynı zamanda uygulama başlığı, uygulama mesajı, çerez, oturum, gönderme ve alma verileri ile sayfa yükleme süresi gibi uygulamaya özgü verileri de içerir. Sayfa gösterimi kayıtlarını barındıran "log_page" veritabanı tablosundan örnek bir veri satırı Tablo 3.3'te verilmiştir. Veritabanında "log_page" tablosunda yer alan erişim kayıtları, oturumlarla ilişkileri bakımından karışık bir düzende görünse de istek sırasına göre bir zaman serisi olarak saklanmaktadır. Ancak bu durum, yapılacak çeşitli veritabanı sorguları ile oturumlar ve ilgili oturumda ziyaret edilen sayfalar hakkında bilgi edinilmesine engel değildir.

Denklem (3.1)'deki ilişki cebiri (relational algebra) formülasyonu, ayrı oturumlarda gezilen sayfaların verilerini günlük veritabanı tablolarından bir SQL iç birleşimi (join) olarak elde etmek için kullanılır.

$$\rho_s \, log\_session \bowtie_{s.opn\_id = p.opn\_id} \rho_p \, log\_page \tag{3.1}$$

Ayrıca, her bir oturumun uzunluğu (süresi), ziyaretçinin o oturum için sayfalarda geçirdiği sürelerin toplanmasıyla elde edilir. Tablodaki her sayfa için talep edilen süre R mevcut olduğundan, p_{nci} sayfada kalma süresi, p_{nci} sayfanın talep edilen süresi R_p 'den $p+1_{nci}$ sayfanın talep edilen süresi R_{p+1} çıkarılarak hesaplanır (Malarvizhi ve Sathiyabhama, 2016). Buna göre, bir oturumun bekleme süresi D_s, Denklem (3.2)'de verildiği gibi, m adet sayfanın bekleme sürelerinin toplamı ile zaman ölçeği kullanılarak hesaplanır.

$$D_s = \sum_{p=1}^{m} \left(R_{p+1} - R_p \right) \tag{3.2}$$

Oturum başına sayfa görüntüleme sayısı (PpS) ise Denklem (3.3)'e göre hesaplanır.

$$P_{ps} = \frac{\sum P}{m} \tag{3.3}$$

Sayfa günlüğü tablosu, yazılım kalitesini iyileştirme ve hata tespiti amacıyla uygulama geliştiricilere bilgi sağlamak için uygulama başlığı, uygulama mesajı, oturum, get, post, cookie ve sayfa yükleme süresi gibi operasyonel verileri de içerir. Web yöneticisi, sistem kullanımını izlemek ve analiz bilgilerini gerçek zamanlı olarak takip etmek için bu veritabanı tablolarını kullanabilir.

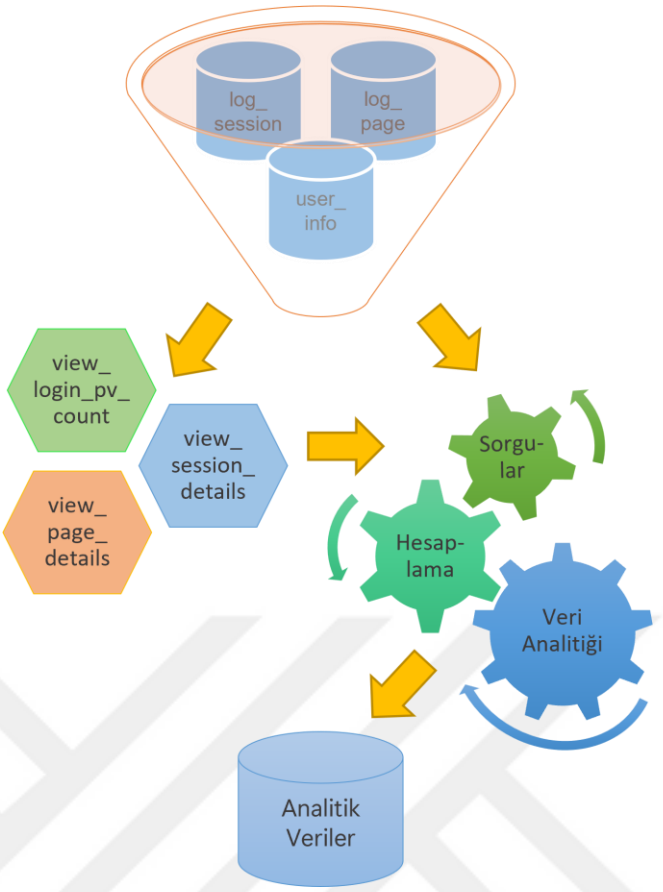
3.5. Veri Saklama Süreci

CAWAL çerçevesinde işlem günlükleri, sonraki veri işleme ve depolama için rutin işlemler sırasında sistematik olarak toplanır. Bu günlükler, tek bir veritabanı kaydında saklanan değerli istatistikler oluşturmak için günlük olarak işlenir. Mimari, sunucu yükünü ve kullanıcı yoğunluğunu saat bazında hesaplama, sayfa ziyaretleri ve bekleme süreleri gibi kullanıcı gezinme ayrıntılarını değerlendirme, optimizasyon için yavaş yüklenen sayfaları belirleme ve yetkisiz erişim girişimlerini tespit etme gibi çeşitli amaçlara hizmet eder. Ayrıca, sayfadaki önemli mesajları ve başlıkları günlüğe kaydeder ve çerezlerden, oturumlardan ve POST ile GET gibi HTTP yöntemlerinden veri yakalar. Bu düzenli veriler yalnızca kapsamlı raporlar ve istatistikler oluşturmak için değil, aynı zamanda sistem performansını ve güvenlik önlemlerini geliştirmek için de temel oluşturur.

Özellikle büyük ölçekli sistemlerde, günlük işlemlerde üretilen kayıt sayısı milyonları aşabilir. Bu ölçekteki bir veritabanı tablosunun OLTP yapısında sürdürülebilirliği, analiz amaçlı kullanılabilirliği ve yedekleme süreçlerinin yönetilebilirliği verimli olmayacaktır. Bu sorunu çözmek için CAWAL modeli kapsamlı bir ETL süreci kullanır. Her gece icra edilen analitik çıkarma sürecinin ilk aşamalarında önceki güne ait günlük kayıtları geçici işlem tablolarına aktarılır ve OLTP tabloları boşaltılarak optimize edilir. Geçici tablolar üzerinde yürütülen analitik çıkarma sürecinin ardından, işlenen veriler uzun süreli depolama ve gelecekte kullanım için veri ambarına aktarılır. Böylece OLTP tablolarındaki veri yığılması ve yavaşlama sorunu ortadan kaldırılmış olur.

3.6. Analitik Çıkarımı Süreci

CAWAL çerçevesinde analitik çıkarımı, gece yarısından sonra gerçekleştirilen otomatik bir süreçtir. Bu süreci yürütecek yığın işleyici betiğin her gece aynı saate düzenli olarak çalışması için Unix tabanlı işletim sisteminde yer alan zaman tabanlı bir iş planlayıcı olan “cron” sistemi yapılandırılır. Bu betik, Şekil 3.9'da gösterildiği gibi günlük web analitiği verilerini kapsamlı veritabanı sorguları kullanarak üretir ve sonuçları analitik tablosuna kaydeder. Bu sürecin devamında, işlenen operasyonel veriler ETL süreciyle birlikte veri ambarındaki ilgili aya ait veri marketine (datamart) aktarılır. Otomatik çalıştırılan komut dosyası tarafından yürütülen analitik çıkarma süreci veritabanı kontrolü ve bakım işlemleriyle başlar.



Şekil 3.9. Web günlüğü verilerinden analitik çıkarma işlemi.

Komut dosyası, Algoritma 3.1'de tanımlı işlemleri adım adım gerçekleştirir. Veritabanı kontrolü ve bakım işlemlerinin ardından metrikler ve çok boyutlu veriler de dahil olmak üzere bir günlük (24 saatlik) web analitiği verileri oluşturulur. Bu algoritma, sunucuların en az meşgul olduğu sabahın erken saatlerinde Linux Cron iş planlama yardımcısı tarafından tetiklenen bir betik ile uygulanmaktadır. Betik sabah erken saatlerde çalıştırıldığı için operasyonel tablolarda önceki ve mevcut güne ait olmak üzere en az iki günlük kayıtlar bulunmaktadır. Analitik çıkarımı yapılırken sorgular, işlem görecek gün başlamış oturumlar üzerinden yürütüldüğü için bir önceki gece başlayıp sonraki güne sarkan oturumları takip etmek ve analitik bilgileri sağlıklı bir şekilde hesaplamak mümkün hale gelmektedir. Ancak, beklenmedik sorunlar nedeniyle betiğin zamanında ya da düzgün çalışmadığı durumlar nedeniyle operasyonel tablolarda birden fazla güne ait günlük kayıtları birikebilir. Böyle bir duruma karşı tedbir amacıyla betik çalıştığında -o gün hariç- bekleyen tüm günleri, $\mathcal{D}_{tüm}$ tek tek işleyecek şekilde yapılandırılmıştır.

Algoritma 3.1. Analitik çıkarımı yapan toplu iş dosyasının yürüttüğü işlemler.

Girdiler:
- Tabloları içeren operasyonel veritabanı (OLTP)
- Operasyonel günlük tabloları: log_ses, log_page
- Geçici tablolar: tmp_ses, tmp_page
- Veri seti: $\mathcal{D}_{tüm}$

Çıktılar:
- Güncellenmiş analitik tablo
- Veri ambarındaki aylık veri marketleri (datamart)
- Süreç meta verilerini barındıran günlük dosyası

1: **Başla:** Zaman bilgisini al ve operasyonel VT'dan tablo adlarını almak için SQL sorgusunu çalıştır.
2: **Veritabanı sağlığını kontrol et:** Operasyonel tablolar üzerinde CHECK, REPAIR (gerekirse), OPTIMIZE ve ANALYZE işlemlerini yürüt.
3: **Süreyi kaydet:** Adım 2'de harcanan zamanı metin tipindeki günlük dosyasına kaydet.
4: **Tarihleri al:** Günlük tablosunda kayıtların bulunduğu geçerli günden önceki tüm $\mathcal{D}_{tüm}$ tarihleri artan sırada sıralanmış olarak elde etmek için SQL sorgusunu çalıştır (birikmiş günlere yönelik).
5: **For döngüsüne başla** ($\mathcal{D}_{tüm}$ içindeki her bir $\mathcal{D}_i$ için):
6: **Geçici tabloları kontrol et:** tmp_ses ve tmp_page mevcutsa TRUNCATE et, değilse CREATE et.
7: **Verileri geçici tablolara aktar:** $\mathcal{D}_i$ kayıtlarını operasyonelden geçici tablolara aktarmak için bir INSERT-SELECT sorgusu çalıştır.
8: **Yıl değişimini kontrol et:** Eğer $\mathcal{D}_i$ 1 Ocak ise, log_ses ve log_page'i TRUNCATE et; aksi takdirde $\mathcal{D}_i$'nin aktarılan kayıtlarını tablolardan DELETE et.
9: **Tabloları optimize et:** log_ses ve log_page tablolarını OPTIMIZE et.
10: **Analitik sorguları gerçekleştir:** $\mathcal{D}_i$ için çeşitli analiz bilgileri elde etmek üzere kapsamlı SELECT sorgularını yürüt.
11: **Ek hesaplamaları gerçekleştir:** Veritabanı sorgularıyla doğrudan elde edilemeyen analitik oranları hesapla.
12: **Çok boyutlu dizileri oluştur:** Farklı türdeki "en-k" listeleri için SELECT sorguları yürüt.
13: **Serileştirme yap:** Çok boyutlu dizi (MDA) verilerini depolamayabilmek için serileştir.
14: **Analitik tablosunu güncelle:** $\mathcal{D}_i$ için mevcut kayıtları analitik tablodan DELETE et (mevcutsa) ve yeni hesaplanan değerleri INSERT et.
15: **Veri marketini oluştur:** $\mathcal{D}_i$'nin ait olduğu aya ait veri marketleri mevcut değilse CREATE et.
16: **Verileri veri marketine aktar:** İşlenen kayıtları veri marketine aktarmak için bir INSERT-SELECT sorgusu çalıştır.
17: **İşlem süresini kaydet:** Analitik tabloda $\mathcal{D}_i$ için "işlem süresi" alanına yazılacak değeri saniye cinsinden hesapla ve güncelle.
18: **İşlemleri kaydet:** $\mathcal{D}_i$'nin işlemlerini ve sürelerini metin günlük dosyasına kaydet.
19: **For döngüsünü bitir**
20: **İşlemi sonlandır:** Analitik tablodaki geçerli tarih için "işlem çıktısı" alanını yürütülen sürece ait kritik bilgilerle güncelle.

Metrik ve boyutların hesaplanması, belirli bir tam gün zaman dilimi içerisinde biriken detaylı veriler üzerinde çalışan, karmaşık bir süreçtir. Bu süreçte ilk olarak belirlenen bir gün penceresi, $\mathcal{D}_i$ için ham veriler, D_{oltp} toplanır. Bu veriler, o gün oturum ve sayfa gösterimi tablolarında toplanan tüm etkileşimleri ve metrikleri kapsar. Bazı veriler ise dönüştürme ve kategorizasyon işlemlerine tabi tutulur.

Sonraki aşamada sayfa görüntüleme ve oturumlar gibi her kritik metrik için o güne ait kümülatif değerleri hesaplamak üzere toplama ve ortalama işlemlerini gerçekleştirir. Son adım, zamansal eğilimlere veya önceden belirlenmiş eşiklere karşı değerlendirilerek bu verilerden çeşitli içgörülerin türetilmesidir. İşlenecek gün olan $\mathcal{D}_i$ verileri, üzerlerinde sorgu çekilmeden önce operasyonel tablolardan alınır ve geçici tablolara aktarılır. Bunun nedeni, operasyonel tabloları mümkün olan en kısa sürede serbest bırakmaktır. Böylece, geçici tablolarda ne kadar uzun sürerse sürsün, bu işlemler operasyonel tablolara ve dolayısıyla sistemde o anda bulunan kullanıcılara etki etmez.

Oturum ve sayfa görünümü tablolarından aynı yapıdaki geçici tablolara aktarılan verileri birleştiren veritabanı sorgusu üç ana bileşende incelenebilir; birleştirme, gruplama ve yansıtma (projeksiyon):

Sorgu, tmp_session (x) ve tmp_page (y) tabloları arasında sol birleştirme (left join) yaparak başlar. Birleştirme koşulu, "log_opn_id" adındaki ortak oturum anahtarının eşit olmasını gerektirir ve veritabanı sorgulama işlemlerini matematiksel bir çerçevede formüle eden ilişkisel cebir ile Denklem (3.4)'teki gibi ifade edilebilir:

$$\sigma x.log\_opn\_id=y.log\_opn\_id(y\times x) \tag{3.4}$$

Bu işlem, "tmp_session" ve "tmp_page" tablolarındaki satırları eşleşen "log_opn_id" değerlerine göre birleştirir. Birleştirme tamamlandığında sonraki adım, elde edilen kayıtları "x.log_opn_id" 'ye göre gruplamaktır. Bu işlem, Denklem (3.5)'teki gösterimle ifade edilir:

$$\gamma x.log\_opn\_id;\min(x.log\_date),\max(x.log\_date),\text{count}(0) \tag{3.5}$$

Gruplama işlemi, aynı "log_opn_id" 'yi paylaşan kayıtları toplayarak oturum süresi, sayfa sayısı ve sayfa başına ortalama süre gibi sonraki hesaplamalar için olanak sağlar.

Daha sonra, sorgu çıktısı için belirli sütunlar yansıtılır. Bunlar arasında *x*'den "log_uid", *y*'den "log_opn_id", minimum ve maksimum "log_date" arasındaki saniye cinsinden fark (SesTime), sayfa gösterimi sayısı (PageCount) ve sayfa başına ortalama geçirilen süre (PageTimeAvg) gibi değerler bulunur. Veri yansıtma matematiksel olarak Denklem (3.6)'teki gibi temsil edilebilir:

$$\pi x.log\_opn\_id, y.log\_uid, \text{timestampdiff}(SECOND,\min(x.log\_date), \max(x.log\_date)), \text{count}(0), \text{timestampdiff}(SECOND,\min(x.log\_date), \max(x.log\_date)) / \text{count}(0) \quad (3.6)$$

Bu projeksiyonlar, "log_opn_id" 'ye göre kullanıcı etkileşimlerini kapsamlı şekilde sunarak, oturum süresi, ziyaret edilen sayfa sayısı ve sayfalarda geçirilen ortalama süre gibi belirli metrikleri sağlar. Analitik çıkarımı, bu örnek sorgunun ötesinde pek çok sofistike SQL sorgusu kullanmayı gerektiren karmaşık bir süreçtir. Ay ve yıl geçişlerinde ise veri marketlerinin düzenlenmesi ve çeşitli optimizasyonlar gibi özel bir dizi kontrol uygulanır. Bu süreçte SQL veritabanı yönetim sistemlerinin bir parçası olan DDL, DML ve DQL komutları (Brahmia ve ark., 2022) kullanılarak oldukça kapsamlı sorgular gerçekleştirir. Analitik çıkarımının en kritik bileşeni olan bu sorgular, önemli miktarda tek veya çok boyutlu veri üretir. Dizi türündeki çok boyutlu veriler ise serileştirilmiş düz metinler haline getirilerek veritabanında saklanırlar. Bu kapsamda, CAWAL aracılığıyla web analitik verilerinin toplanması ve bu verilerden anlamlı bilgilerin elde edilmesine yönelik adımlar Şekil 3.10'da verilmiştir.

Şekil 3.10. CAWAL çerçevesinin veri toplama ve bilgi çıkarımı süreci.

Veritabanı indeks alanlarını ve kapsamlı sorguları optimize etmek, günlük milyonlarca kaydı işleyen sistemlerde veri toplama ve analiz performansı için kritik önem taşır. CAWAL analitik çıkarımı sürecinin optimizasyonu için gerek veritabanlarında gerekse yürütülen algoritmada önemli geliştirmeler yapılmış ve nihayetinde son derece verimli bir komut dosyası elde edilmiştir. Diğer taraftan, betik tarafından gerçekleştirilen bazı kritik işlem bilgilerinin kayıt altına alınması amacıyla bu bilgiler olası sorunlara karşı veritabanından bağımsız bir günlük (log) dosyasına da yazılmaktadır. Önemli olayların metin tabanlı bir dosyada zaman damgası kullanılarak saklanması, bu bilgilerin güvenilir bir şekilde korunmasını ve web üzerinden zahmetsizce izlenebilmesini sağlamaktadır.

Analitik çıkarımı sürecinde hesaplanan günlük istatistiksel veriler her bir gün için bir kayıt olacak şekilde analitik veritabanı tablosunda saklanır. Kapsamlı günlük analitik bilgilerin saklanması için tasarlanan “log_analytics” tablosunun alanları, örnek kayıt yapısı ve açıklamaları Tablo 3.4’te verilmiştir. Analitik tablonun amacı, saatlik servis ve sunucu yoğunlukları ile kullanım bilgilerine dair istatistiklerin yanı sıra, kullanıcı etkinlik ve yoğunluk bilgilerini de kapsayan geniş analitik veri setleri üretmektir. Bu tablo aynı zamanda, sistemde gerçekleşen sayfa ziyaretleri, bu ziyaretlerin süreleri ve kullanıcıların sayfalarda geçirdikleri vakitler üzerine detaylı analitik bilgiler sağlar.

Bütün bu veri noktaları, sistem performansını değerlendirmek ve kullanıcı deneyimini geliştirmek amacıyla çeşitli istatistiklerin ve raporların oluşturulmasına olanak sağlar. Bu sayede, sistemin verimliliğini artırmak ve proaktif iyileştirmeler yapmak için gereken analitik altyapı sağlanmış olur. Analitik tablo, çok boyutlu veriler de dahil olmak üzere yaklaşık 50 alan içerir. Tabloda bulunan bir güne ait kayıt boyutu yaklaşık 68 KB olup, bu durum toplanan verilerin ve analitik çıkarımının kapsamını etkileyici bir şekilde ortaya koymaktadır. Analitik tablosunda “ana_id” haricindeki tüm alanlar, analitik çıkarımı API'sinin hesaplamalı prosedürleri sonucunda oluşturulmuş verilerdir. Bunlar, API'nin özgün algoritmaları ve iş kuralları doğrultusunda üretilir ve çoğunlukla dinamik bir yapının sonucudur.

Öte yandan, otomatik artan (auto increment) benzersiz bir tanımlayıcı olan “ana_id”, “ana_row_date” ile birlikte veritabanında veri sıralamasını ve hızlı erişimi sağlayan bir indeks olarak işlev görür. Bunların dışında, “ana_row_date” ile “ana_output_message” alanları arasında kalan veriler ise analitik çıkarımı için gerekli olan yığın işlemi gerçekleştiren zamanlanmış göreve yönelik bilgileri içerir.

Tablo 3.4. Analitik tablonun alanları, açıklamaları ve örnek bir veri satırı.

Alan Adı	Örnek Veri	Açıklama
ana_id	889	Günlük kaydının benzersiz numarası.
ana_row_date	16.03.2018	İşlenme tarihi (satırın ait olduğu tarih).
ana_proc_datetime	17.03.2018 03:34	İşlem tarihi-saati (kaydedilen satır).
ana_proc_seconds	287	Saniye cinsinden işlem süresi.
ana_proc_time	00:04:47	Zaman biçiminde işlem süresi.
ana_output_message	Tablo onarıldı.	İşlem hakkında uygulama mesajı.
page_view_usr	119149	Oturum açmış kullanıcıların sayfa görüntüleme sayısı.
page_view_gue	42523	Misafir kullanıcıların sayfa görüntüleme sayısı.
page_view_tot	161672	Tüm ziyaretçilerin sayfa görüntüleme sayısı.
ses_count_usr	17449	Oturum açmış tüm kullanıcıların oturum sayısı.
ses_count_unq	10834	Oturum açmış benzersiz kullanıcıların oturum sayısı.
ses_count_gue	4655	Misafir kullanıcıların oturum sayısı.
ses_count_tot	22104	Tüm ziyaretçilerin oturum sayısı.
ses_per_unq	1,6106	Oturum açmış tekil kullanıcı başına oturum sayısı.
pv_per_usr	6,83	Oturum açmış tüm kullanıcılar başına sayfa gör. sayısı.
pv_per_unq	11	Oturum açmış tekil kullanıcı başına sayfa gör. sayısı.
pv_per_gue	9,13	Misafir kullanıcı başına sayfa görüntüleme sayısı.
pv_per_tot	7,31	Tüm ziyaretçiler başına sayfa görüntüleme sayısı.
pv_avgtime_usr	24,92	Oturum açmış tüm kullanıcılar için ort. sayfa gör. süresi.
pv_avgtime_unq	40,14	Oturum açmış tekil kullanıcı başına ort. sayfa gör. süresi.
pv_avgtime_gue	11,71	Konuk kullanıcı başına ortalama sayfa gör. süresi.
pv_avgtime_tot	24,19	Tüm ziyaretçiler başına ortalama sayfa gör. süresi.
ses_avgtime_usr	218,06	Oturum açmış tüm kullanıcılar için ort. oturum süresi.
ses_avgtime_unq	351,21	Oturum açmış tekil kullanıcı başına ort. oturum süresi.
ses_avgtime_gue	22,66	Konuk kullanıcı başına ortalama oturum süresi.
ses_avgtime_tot	176,91	Tüm ziyaretçiler için ortalama oturum süresi.
best_server	3	Web çiftliğinde en çok kullanılan sunucu numarası.
best_service	obis	En çok kullanılan site hizmet adı.
best_page	mail-list	En çok kullanılan site modülü adı.
best_ip	10.8.XX.XX (gizlendi)	Sistemi en yoğun kullanan istemci IP adresi.
best_proxy	78.189.XXX.XX (gizlendi)	Sistemi en yoğun kullanan istemci vekil sunucu adresi.
best_user	sabXXX (gizlendi)	Sistemde oturum başına en fazla sayfa görün. kullanıcı
best_risky_user	tkiXXX (gizlendi)	Bir oturumda en çok sayfa görün. kullanıcı (tehlikeli).
best_os	win-10	En çok kullanılan işletim sistemi.
best_browser	chrome	En çok kullanılan tarayıcı.
best_referer	www.google.com.tr	Sitenin en çok yönlendirme aldığı web sitesi.
best_landing_url	http://www.obis...	Kullanıcıların ilk kez görüntülediği sayfa (açılış sayfası).
best_login_url	http://www.obis...	En çok giriş yapılan sayfa adresi.
best_last_url	http://www../?page=logout	En çok çıkış yapılan sayfa adresi.
best_search_engine	www.google.com.tr	Siteye en çok yönlendiren arama motoru.
best_search_key	sau obis	Siteye erişmek için arama mot. en çok kullanılan ifade.
best_pageview_hour	11	Günün en yüksek sayfa görüntüleme saati.
best_sess_hour	15	Günün en yüksek ziyaret (oturum oluşturma) saati.
best_user_hour	15	Günün en yüksek kullanıcı oturum açma saati.
top_site	Seri çok boyutlu dizi	Sunucular ve alt alan adlarıyla ilgili sayfa g. ve oturum.
top_user	Seri çok boyutlu dizi	Kullanıcı türü ve cinsiyete göre sayfa g., otur. ve süreler.
top_hour	Seri çok boyutlu dizi	Kullanıcı türü ve cinsiyete göre saatlik sayfa g. ve ot. say.
top_client	Seri çok boyutlu dizi	IP, proxy, tarayıcı, ülke, dil vb. gibi istemci bilgileri.
top_ref	Seri çok boyutlu dizi	Yönlendiren site ve türü, açılış sayfası gibi yön. bilgileri.
top_search	Seri çok boyutlu dizi	Kul. siteye erişirken kullandığı arama motoru kelimeleri.

Analitik tablosunun sonunda yer alan altı adet serileştirilmiş verinin sahip olduğu boyutlar (dimensions) ve açıklamaları Tablo 3.5'te verilmiştir.

Tablo 3.5. Analitik tablosunda yer alan çok boyutlu diziler ve açıklamaları.

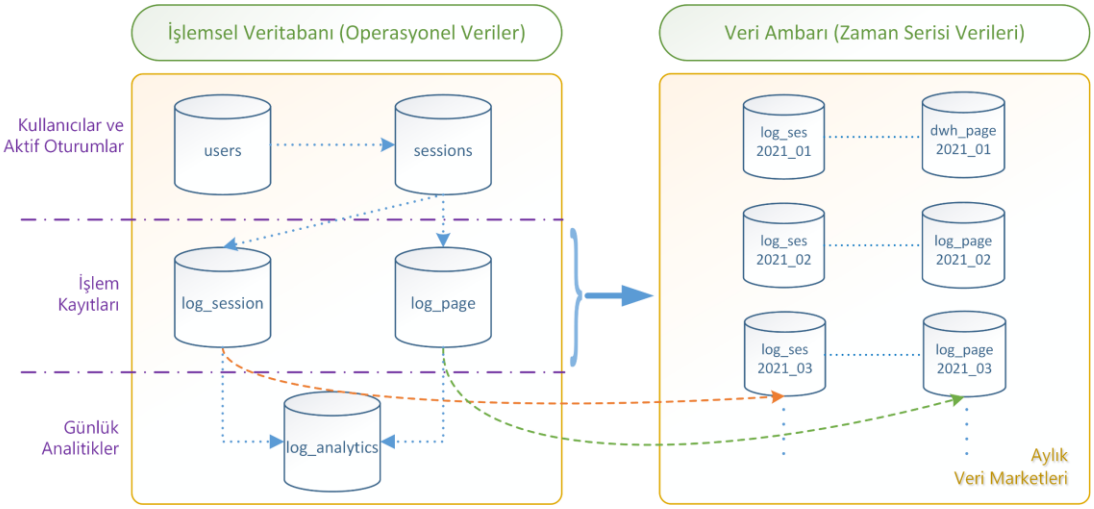
Dizi Adı	Açıklama
[top_site][top_server][server_no]	Sunucu numarasına göre kullanıcı sayfa görüntüleme ve oturum verileri.
[top_site][top_service][sname][utype][sex]	Kullanıcı türü ve cinsiyetine göre sayfa görüntüleme ve oturumlara göre tüm portal hizmetlerinin verileri.
[top_site][top_module][module_name]	En çok sayfa görüntülemesine sahip ilk 20 portal servisi modülünün sayfa görüntüleme ve oturum verileri.
[top_user][top_used_user][user_name]	Tüm oturumlarda en çok sayfa görüntüleme sayısına göre ilk 20 kullanıcının sayfa görüntüleme ve oturum sayısı verileri.
[top_user][top_dangerous_user][user_name]	Bir oturumda en çok sayfa görüntüleme sayısına göre ilk 20 kullanıcının sayfa görüntüleme ve oturum sayısı verileri (tehlikeli).
[top_user][page_view_group][visitor_type]	Sayfa görüntüleme sıklığına göre beş kategoriye ayrılmış ziyaretçi türleri ve kullanım sayısı verileri.
[top_user][utype_sex][usertype][sex]	Kullanıcı türü ve cinsiyete göre kullanıcıların sayfa görüntüleme sayısı, oturum sayısı ve sürelerine ilişkin veriler.
[top_user][usr_ses][int_ext]	Tüm kullanıcıların kurum içi, ülke içi ve ülke dışı sayfa görüntüleme ve oturum sayılarına ilişkin veriler.
[top_hour][detail][hour][vtype][utype][sex]	Ziyaretçi türü, kullanıcı tipi ve cinsiyete göre saatlik kullanıcı sayfa görüntüleme ve oturum sayıları verileri.
[top_hour][hour_arr][hour]	Tüm kullanıcıların saatlik sayfa görüntüleme ve oturum sayılarına ilişkin veriler.
[top_client][top_ip][IP]	Oturum sıklığına göre sıralanmış ilk 20 IP adresinin ziyaret ve sayfa görüntüleme sayılarının verileri.
[top_client][top_proxy][proxy_IP]	Oturum sıklığına göre sıralanmış ilk 20 vekil IP adresinin ziyaret ve sayfa görüntüleme sayılarının verileri.
[top_client][top_os]	Oturum sıklığına göre sıralanmış ilk 20 kullanıcının işletim sistemi verileri.
[top_client][top_brow]	Oturum sıklığına göre sıralanmış ilk 20 kullanıcının tarayıcı verileri (adı ve sürüm bilgisi).
[top_client][top_brow_type]	Oturum sıklığına göre sıralanmış ilk 20 kullanıcının tarayıcı türleri verileri.
[top_client][top_lang]	Oturum sıklığına göre sıralanmış ilk 20 kullanıcının tarayıcı dili verisi.
[top_client][top_country]	Oturum sıklığına göre sıralanmış ilk 20 kullanıcının menşe ülke verisi (GeoIP tablosundan).
[top_ref][top_ref_state]	Oturum sıklığına göre sıralanmış yönlendirici türü verileri.
[top_ref][top_ref]	Oturum sıklığına göre sıralanmış ilk 20 yönlendirici sitenin verileri.
[top_ref][top_first_url]	Erişim sıklığına göre sıralanmış ilk 20 varış sayfası (ilk kullanılan URL) verisi.
[top_ref][top_last_url]	Erişim sıklığına göre sıralanmış ilk 20 ayrılış sayfası (son kullanılan URL) verisi.
[top_ref][top_login_url]	Erişim sıklığına göre sıralanmış ilk 20 giriş sayfası (girişe yönlendiren URL) verisi.
[top_ref][top_logout_type]	Oturum sıklığına göre sıralanmış dört gruptaki oturum kapatma türlerine ait veriler.
[top_search][top_search_engine]	Erişim sıklığına göre sıralanmış yönlendiren arama motoru verileri.
[top_search][top_search_key]	Erişim sıklığına göre sıralanmış arama motoru anahtar kelime verileri.
[top_search][top_src_eng2key][src_eng]	Siteye erişmek için kullanılan arama motoruna göre anahtar kelimelerin kullanım sıklığı verileri.
[top_search][top_key2src_eng][keyword]	Siteye erişmek için kullanılan anahtar kelimelere göre arama motorlarının kullanım sıklığı verileri.
[top_search][top_src_eng_landing][src_eng]	Erişim sıklığına göre sıralanmış arama motorları aracılığıyla doğrudan erişilen portal alt alanlarının verileri.

3.7. ETL Süreci ve Veri Ambarı Yapısı

ETL süreci, veri ambarlarının etkin kurulumu ve işletilmesi için kritik öneme sahiptir ve CAWAL modelinin temel bileşenlerinden biridir. Milyonları bulan günlük işlem kayıtlarının büyüklüğü, OLTP sistemlerinde bu verilerin yönetimini, analizini ve yedeklemesini sürdürülemez kılmaktadır. Bundan dolayı, veriler öncelikle işlenip gerekli analitik çıkarımlar yapıldıktan sonra, ETL süreçleri aracılığıyla uzun vadeli depolama ve ileride gerçekleştirilecek analizler için veri ambarlarına aktarılır.

Sürekli birikmeye devam eden web kullanım verileri, yoğun kullanılan sistemlerde büyük veri sınıfına girecek kadar fazla miktarda olabilir. Çevrimiçi işlem işleme (OLTP) sistemlerinde genellikle doğrudan ve karmaşık olmayan sorguları içeren hızlı işlem oranları desteklenir. Öte yandan, veri ambarları büyük ve durağan veri kümeleri üzerinde karmaşık sorguları yönetmek ve yürütmek için tasarlanmıştır. Veri ambarı ve içerisinde yer alan veri marketleri yapısı, web analitiği ve kullanım madenciliği uygulamaları için özel olarak tasarlanmış CAWAL çerçevesinin önemli bir parçasıdır.

ETL süreci, verileri merkezi bir veri ambarında toplayan analitik çıkarımı ile başlar. Daha sonra, bu ham verileri raporlama ve madenciliğe elverişli bir analitik şemaya dönüştürmek için özel dönüşümler gerçekleştirilir. Dönüşüm sonrasında operasyonel veriler, aylar ve yıllar arasındaki geçişleri yönetmek için özel prosedürler kullanılarak veri ambarı içindeki ilgili aylık veri marketine yönlendirilir. Şekil 3.11, verilerin operasyonel veri tabanından veri ambarına nasıl aktarıldığını görselleştirmektedir.



Şekil 3.11. Operasyonel kayıtların veri ambarına ETL süreciyle aktarımı.

Kullanıcıların aktif oturumları ve işlem kayıtları, günlük analitiklerle birlikte her gece yarısı düzenli olarak veri ambarına ETL süreci aracılığıyla taşınır. Bu süreç, veri ambarındaki zaman serisi verilerini düzenler ve her ay için ayrı veri marketleri oluşturarak verilerin zamana göre segmentasyonunu ve analiz için hazır hale getirilmesini sağlar.

Ham verilerin zamanla birikmesi ve analitik süreçler için hazır hale getirilmesi esasına dayanan veri ambarı yapısı, OLAP (Keskin ve Yazıcı, 2022) gibi çok boyutlu analizlere elverişli bir ortam sağlayarak CAWAL'ın analitik yeteneklerini güçlendirmektedir. Oturumlar ve kullanıcılardan sayfalara, zaman damgalarına ve trafik kaynaklarına kadar çeşitli metriklerin ilişkilendirilmesine izin verir. Böylece ayrıntılı uzun vadeli raporların oluşturulmasına, eğilimlerin ve kalıpların ortaya çıkarılmasına, segmentasyon analizlerinin yapılmasına ve çeşitli veri madenciliği algoritmalarının uygulanmasına olanak tanır.

Operasyonel verilerin dönüşümünde ve zaman serisi veri analizinde büyük rol oynayan ETL sürecinin bu stratejik uygulaması, aşağıda açıklanan çıkarma, dönüştürme ve yükleme şeklindeki üç aşamada gerçekleşir:

Çıkarma sırasında, OLTP veritabanından alınan veriler D_{oltp} olarak gösterilir ve gerçek zamanlı işlemsel bilgileri temsil eder. OLTP yapıları için bir çıkarma fonksiyonu f olarak sembolize edilebilir, böylece $S_{oltp}=f(D_{oltp})$ olur.

Dönüştürme süreci, web trafiğinin doğasındaki değişkenlik göz önüne alındığında, D_o ile temsil edilen veri tutarsızlıklarının giderilmesini gerektirir. Temizleme dönüştürme fonksiyonu T_c, $S_{temiz}=T_c(S_{oltp}-D_o)$ fonksiyonu ile temiz veri sağlar. Ayrıca, temizlenmiş OLTP verileriyle dış veri kaynaklarının ($S_{dış}$) entegrasyonu $S_{iç}=S_{temiz} \cup S_{dış}$ olarak gerçekleştirilir. Web analitiği için özel dönüşüm kural seti R_{wa} veri kümesini $S_{dön}=R_{wa}(S_{iç})$ olarak verir.

Yükleme aşamasında ise veriler DW veri ambarına yüklenir. Yükleme süreci $DW_{yeni}=DW_{önc} \cup S_{dön}$ denklemi ile tanımlanır ve burada $DW_{önc}$ veri ambarının önceki durumunu ifade eder.

Yürütülen bu ETL süreci, web analitiği verilerinin verimli bir şekilde işlenmesini ve depolanmasını garanti ederek hem OLTP sistemlerinin hem de veri ambarının verimliliğini ve faydalarını en üst düzeye çıkartır.

3.8. CAWAL'ın Özellikleri ve İşlevselliği

Çağdaş web analitiğinde sağlam, uyarlanabilir ve duyarlı araçlara duyulan ihtiyaç, özellikle büyük ölçekli ve zorlu web ortamlarında son derece önemli hale gelmiştir. CAWAL çerçevesi, bu talepleri ele alırken modern web sistemlerinin karmaşıklığını yansıtan özellikler ve işlevler sağlayarak verimliliği, güvenilirliği ve kullanıcı deneyimini geliştirmeyi amaçlamaktadır. Çerçevenin etkinliğine katkıda bulunan belirli özellikler ve işlevler aşağıda alt başlıklar halinde detaylıca açıklanmıştır.

3.8.1. Çok sunuculu ortam adaptasyonu

CAWAL çerçevesi, özellikle büyük ölçekli sistemlerle ve web çiftliği olarak adlandırılan çok sunuculu ortamla uyumlu olacak şekilde tasarlanmıştır ve bir sunucu ağı üzerinden dağıtık bilgi işlem yapılmasını sağlar. Bu mimari sadece yük dengelemeyi kolaylaştırmakla kalmaz, aynı zamanda yüksek derecede ölçeklenebilirlik ve güvenilirlik sağlar. İzleme verilerinin gerçek zamanlı olarak işlenmesi gerektiğinde, web çiftliğinin dağıtık yapısı hızlı hesaplamalara olanak tanır. CAWAL bu dağıtık yapıyı kullanarak talep yükünü verimli bir şekilde dengeler, hesaplama görevlerini yönetir ve potansiyel olarak sistem performansını ve güvenirliğini artırır. Web çiftliği yapılandırması, CAWAL çerçevesinin veri toplama süreçlerini desteklemedeki temel bir unsurdur. Bu yapılandırma, çağdaş teknolojik gereksinimler ve optimizasyon ilkeleriyle uyum içinde, büyük ölçekli ve dağıtık sistemlerde veri toplama işlemlerinin etkin ve verimli bir şekilde gerçekleştirilmesine olanak tanır.

3.8.2. Alanlar arası kullanıcı takibi

CAWAL çerçevesi, birden fazla web sunucusunun kullanıldığı ortamlarda oturumları etkin bir şekilde yönetmek üzere tasarlanmıştır. CAWAL, ağa bağlı bir depolama (NAS) sunucusunda bulunan ortak bir oturum dizini kullanarak portalın çeşitli servislerinde gezinen kullanıcılar için veri tutarlılığı sağlar. Çerçeve, farklı etki alanları ve alt etki alanlarındaki oturumları izlemek amacıyla, her bir isteği işleyen sunucuyu tam olarak belirlemek için "sunucu numarası" ve her alt etki alanına karşılık gelen portal servisini tanımlamak için "servis adı" gibi belirli veritabanı alanları içerir. Bu metodolojik yaklaşım, oturum veri yönetimini iyileştirir ve sistemin analitik kapsamını genişleterek sunucular, etki alanları ve portal servisleri arasında kullanıcı etkileşimlerinin kapsamlı, çok boyutlu analizlerini mümkün kılar.

3.8.3. Dinamik IP uyumluluğu

CAWAL, web sunucusu tarafından başlatılan oturumları kullanıcı takibinin temel referansı olarak kullanır. Bu sistemde, aynı IP adresini kullanan iki farklı cihaz, ayrı oturumlar olarak kabul edilir. Dinamik IP adresleri, bir ağa bağlı cihazların her bağlantıda farklı IP adresleri alması durumudur. Bu, geleneksel takip sistemlerinde, özellikle aynı ağ üzerinden İnternet'e erişen birden fazla kullanıcının takibini zorlaştırır. Ancak, CAWAL, bu dinamik IP dağılımının oluşturduğu zorlukların üstesinden gelir ve oturumları IP adreslerine dayalı olarak doğru bir şekilde ayırt edebilir. Web sunucu günlüklerinde, dinamik IP adreslerine bağlı olarak aynı IP'yi kullanan farklı kullanıcıların karıştırılması yaygın bir sorundur. CAWAL bu sorunu aşarak, her kullanıcıyı doğru bir şekilde tanımlar ve ayırt eder. Dinamik IP adreslerinin yaygın kullanımı göz önüne alındığında, CAWAL'ın bu özelliği, kullanıcı takibi ve veri analizi konusunda yüksek derecede veri doğruluğu ve güvenilirlik sağlar.

3.8.4. Oturum zaman aşımı bildirimi

CAWAL, kullanıcı katılımını artırmak ve oturum sürekliliğini korumak için önemli bir özellik olan oturum zaman aşımı bildirim yöntemini içerir. Çerçeve bu yöntemi kullanarak, tanımlanmış bir zaman dilimi içinde kullanıcının hareketsizliğini izler ve zaman aşımı eşiğine yaklaşıldığında kullanıcıya gösterilecek bir bildirim penceresini tetikler. Oturumun sonlanacağını bildiren pencere, kullanıcıya devam etmek istemesi durumunda seçenek sunar. Bu yaklaşım, yanlışlıkla oturumun sonlandırılmasını en aza indirerek potansiyel veri kaybını azaltır ve genel kullanıcı deneyimini iyileştirir. Özellikle İnternet bankacılığı gibi hassas bilgilerin işlendiği uygulamalarda kullanılan bu yöntem, kullanıcıyı oturum yönetimine aktif olarak dahil ederek kullanıcı farkındalığını güçlendirmekte, güvenli ve kontrollü bir ortam sağlanmasına yardımcı olmaktadır. Yöntem ayrıca, diğer araçların genellikle 30 dakikalık bir varsayımla (Jansen ve ark., 2022; Maslennikova ve ark.,2022) ele aldığı bir sorun olan oturum süresi sorununu da çözmekte ve böylece oturumların kesin bir tanımını sunmaktadır.

3.8.5. Tutarlı oturum ve kullanıcı tanımlama

CAWAL, web analitiği süreçlerinde tutarlı oturum ve kullanıcı bilgileri elde etme konusunda özel bir yeteneğe sahiptir. Çerçeve, web uygulamasından ve farklı web platformlarından gelen verileri etkili bir şekilde birleştirerek, kullanıcıların çeşitli oturumlarını tek bir kimlik altında toplar. Kullanıcı ve oturumların tam ve doğru şekilde ayrıştırılması sayesinde aynı kullanıcı tarafından gerçekleştirilen farklı

ziyaretler ve etkileşimler, benzersiz bir kullanıcı profili altında entegre edilir. Böylelikle, kullanıcıların çeşitli cihazlar ve oturumlar boyunca gösterdikleri davranışların daha tutarlı ve bütünsel bir analizi sağlanır. CAWAL, bu yöntemle, web analitiklerinde sıklıkla karşılaşılan veri parçalanması ve tutarsızlık sorunlarını aşarak, daha isabetli ve güvenilir kullanıcı verileri sunar.

3.8.6. Gece yarısı sınır problemi çözümü

Gece yarısını geçen oturumların doğru şekilde tanımlanması, yanlış sınıflandırmadan kaynaklanan yanıltıcı ölçümlerin önlenmesi ve web analizlerinin bütünlüğünün korunması açısından önemlidir. Web analitik araçlarının birçoğu, bu sorunu çözmek amacıyla "oturum zaman aşımı" süresini uzatır veya yeni bir güne geçildiğinde bunu farklı bir oturum olarak tanımlar. Ancak CAWAL, tek tek sayfa istek süreleri yerine oturumun başlangıcına odaklanır. Sabaha karşı otomatik çalışan analitik çıkarımı komut dosyası, gece yarısından önce başlayan ve bir sonraki güne sarkan oturumları algılar ve bunları önceki günün istatistiklerine dahil eder. Bu, kullanıcı etkileşimlerinin daha doğru ve gerçekçi bir biçimde değerlendirilmesini sağlayarak yanlış stratejik kararlara yol açabilecek metrik bozulma risklerini ortadan kaldırır. Ayrıca, işlemsel takip gerektiren belirli kullanıcı etkileşimlerinin sağlıklı şekilde izlenebilmesi bu sayede mümkün hale gelir.

3.8.7. Yüksek veri doğruluğu

Etkili web analitiği için güvenilir ve yüksek kaliteli veri kullanımı esastır. CAWAL çerçevesi, bu alanda yaygın olan tekrarlanan ziyaretçi sayımı gibi sorunları, kullanıcı oturumlarını farklı alanlarda birleştirme ve bütünleşik bir veri ambarı mimarisi kullanarak çözmektedir. Çerçeve, çok boyutlu analizler ve detaylı kullanıcı yolculuğu haritaları aracılığıyla, ziyaretçi davranışlarını frekansa dayalı olarak derinlemesine analiz etme ve segmentasyon yapma imkânı sunar. Analistler bu sayede, haftalık, aylık ve yıllık gibi farklı zaman aralıklarında tekrar ziyaret oranlarını değerlendirerek dijital etkileşim stratejilerinin ve kullanıcı tutunma oranlarının etkinliğini detaylı bir şekilde inceleyebilirler. CAWAL, tarihsel verileri bütünleştiren güçlü bir yapıya sahiptir, bu sayede kuruluşlar geçmişteki kullanıcı ziyaretleri ve etkileşim kalıplarını analiz ederek, eyleme geçirilebilir ve somut içgörüler elde edebilirler. Çerçeve, zaman içinde kullanıcı davranışlarının analizi ve uzun vadeli veri değerlendirmeleri için sunduğu yenilikçi yaklaşımla, diğer çözümlerin karşılaştığı analitik belirsizliklerin üstesinden başarıyla gelir.

3.8.8. Veri sahipliği ve yönetişim

Veri sahipliği ve yönetişimi, günümüzün veri güvenliği ve mahremiyet ortamında hayati öneme sahiptir. Yerinde (on-premises) bir çözüm olarak CAWAL, kuruluşlara analitik veri varlıkları üzerinde tam kontrol sağlar. Veriler, gizlilik ve güvenlik için dahili politikalara bağlı kalarak kurumsal altyapı içinde saklanır. Bu kurulum, yerel veri düzenlemelerine uyumu daha yönetilebilir hale getirir. CAWAL'ın veri yönetimi özellikleri, erişim kontrolleri, denetim kayıtları ve şifreleme ile hem güveni hem de şeffaflığı artırır. Kurumlar veri saklama stratejilerini özelleştirebilir ve analitik süreçler üzerinde tam kontrol sahibi olabilir. Böylece CAWAL, etik standartlar ve değerlerle uyumlu bir şekilde web zekâsında kurumsal özerkliği teşvik ederken, veri sahipliğini koruyarak kapsamlı analitikler yapılmasını sağlar.

3.8.9. Hesap güvenliğini arttırma

CAWAL'ın web portalına entegrasyonu sadece analitik süreçleri değil aynı zamanda açık oturumların takip edildiği ve kullanıcı bilgilerinin saklandığı veritabanı tablolarını da zenginleştirir. Bu entegrasyon sürecinde, kullanıcıların hesap güvenliğiyle ilgili bilgilerini, örneğin en son başarısız giriş denemeleri ve şifre değişiklik tarihleri gibi önemli verileri tutmak için yeni alanlar eklenmiştir. Uygulama veritabanına ilave edilen alanlar, kullanıcı güvenliğini izleme ve iyileştirme kapasitesini artırırken aynı zamanda kullanıcılara ve sistem yöneticilerine kullanıcı hesap güvenliğinin sürekli olarak değerlendirilmesi ve yönetilmesi için daha detaylı bilgiler sağlamaktadır.

3.8.10. Hesap verilebilirliği sağlama

Kullanıcı etkileşimlerini hesap bazlı takip edebilmesi CAWAL'ın önde gelen özelliklerinden biridir. Bu yetenek sayesinde, kullanıcılar yaptıkları işlemleri iddia etme imkanına kavuşurken, web ve sistem yöneticileri kullanıcıların yaptıkları işlemleri doğrulama olanağına sahip olurlar. Örneğin, bir öğrencinin ders kaydını yaptığını iddia ettiği halde yapmamış olduğunun ortaya çıkartılması ya da öğrencinin ders seçimi yaparken belirli bir dersin karşısına gelmediği iddiasını ispatlayabilmesi bu yolla mümkün hale gelmektedir. Ayrıca kanuni bir zorunluluk olarak web hizmeti veren kuruluşlar kullanıcıların yaptıkları işlemleri kayıt altına almak ve talep etmeleri halinde kolluk kuvvetlerine sunmak zorundadırlar. Kullanıcıların erişim ve kullanım bilgilerinin yönetim arayüzü üzerinden zaman aralığı ya da kullanıcı adı gibi filtreler kullanılarak kolaylıkla çıkartılması, sistem üzerinden gerçekleşen eylemlerin tespitini, analizini ve ispatını olanaklı kılmaktadır.

3.8.11. Uygulama ve sistem güvenliğini arttırma

Oturumların sunuculara dağılımlarını takip edebilme, hata yakalama, anormallik tespiti ve istenmeyen kullanıcı etkileşimlerini saptayabilme, CAWAL'ın kendine has, önemli yetenekleridir. Hatalı, aşırı ya da dengesiz kullanım durumlarının kolayca tespit edilerek bu tür etkileşimleri engellemeye yönelik önlemler alınmasına imkân tanıyan bu yetenekler, web portalının güvenlik katmanını daha da güçlendirir ve kullanıcıların çevrimiçi deneyimlerini daha korumalı hale getirir. Web geliştiricileri ve sistem yöneticilerinin anormal etkileşimleri hızla tanımlayıp müdahalede bulunabilmeleri sayesinde uygulama ve sistem güvenliği daha etkin bir şekilde sağlanmış olur.

3.8.12. Yönetim arayüzü

CAWAL çerçevesinin yönetim arayüzü, web analitiği sürecinin çeşitli yönlerini komuta ve kontrol etmek için gelişmiş işlevler sunan bir merkezdir. Kullanıcı dostu bir arayüze sahip olan gösterge paneli, yöneticilere kullanıcı yönetimi, güvenlik ayarları, veri yapılandırması ve raporlama özelliklerinin özelleştirilmesi gibi kritik işlevlere kolay erişim sağlar. Gösterge panelinin modüler tasarımı, sistemin parametrelerinde ve ayarlarında etkili ayarlamalar yapılmasına olanak tanıyan, kolaylaştırılmış gezinme sunar. Kurumsal gereksinimlere esnek bir şekilde yanıt verecek şekilde tasarlanan panel, web analitiğinin etkin bir şekilde kullanılmasına katkıda bulunur.

3.8.13. Gerçek zamanlı izleme

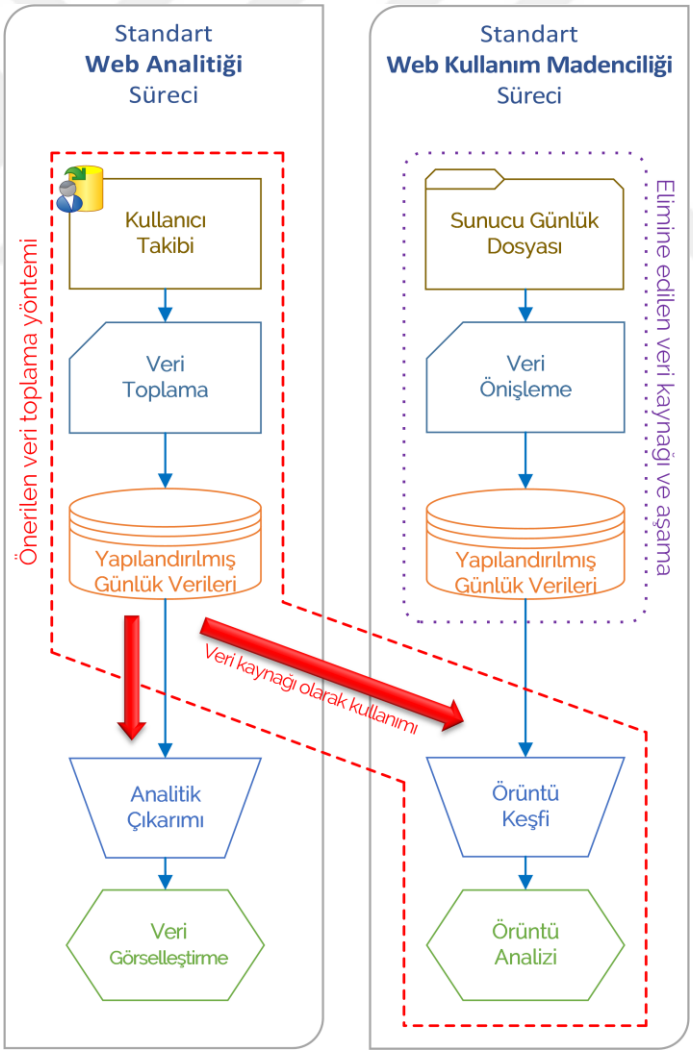
Web ve sistem yöneticileri için sunucular, web hizmetleri, kullanıcılar ve oturumlar dahil olmak üzere bir web sisteminin gerçek zamanlı durumunu izlemek, hızlı sorun tespiti ve çözümü için kritik öneme sahiptir. Bu gerçek zamanlı analiz, sorunlar büyümeden ve kullanıcı deneyimini ciddi şekilde etkilemeden önce proaktif yanıtlar verilmesini sağlar. CAWAL çerçevesi sayesinde hangi sunucu, servis ya da modülde kaç ziyaretçi olduğu, bunların kaçının misafir veya kullanıcı olduğu ya da aralarında aynı IP adresine sahip kaç ziyaretçi bulunduğu gibi sistemin anlık durumu hakkında kapsamlı bilgiler yönetici paneli üzerinden kolaylıkla takip edilebilmektedir. Ayrıca gerçek zamanlı kullanıcı izleme arayüzü üzerinden yetkiler dahilinde kullanıcıları sistemden atma ya da süre bazlı uzaklaştırma gibi yönetimsel işlemler de kolaylıkla gerçekleştirilebilmektedir.

3.8.14. Raporlama ve görselleştirme

CAWAL çerçevesi, karmaşık web analitiği verilerini erişilebilir bir şekilde sunmak için gelişmiş raporlar sağlar. Kullanıcı katılımı, trafik kaynakları ve davranış kalıpları gibi çeşitli ölçümleri çıkarmak için uyarlanan raporlama bileşeni, belirli iş hedefleri ve stratejileriyle uyum sağlamak için özelleştirme sunar. Bunu tamamlayan görselleştirme özelliği, ham verileri çizelgeler, grafikler veya ısı haritaları gibi çeşitli etkileşimli görsel biçimlere dönüştürerek web olaylarına ilişkin gerçek zamanlı içgörü sağlar. Raporlama ve görselleştirmenin bu kombinasyonu sezgisel anlayışı, eğilimleri tanımlamayı ve veriye dayalı taktik ve stratejik kararları destekler.

3.8.15. Web kullanım madenciliği entegrasyonu

Erişim verilerinin doğruluğu ve bütünlüğü, WUM sürecinin başarılı ve güvenilir sonuçlar vermesini sağlamak için kritik önem taşımaktadır.



Şekil 3.12. CAWAL ile toplanan verilerin WA ve WUM sürecinde kullanımı.

Hedeflediği web madenciliği faaliyetlerini gerçekleştirmek için web sunucu günlüklerini yetersiz bulan ve kullanıcı erişim verilerini başkalarıyla paylaşmak istemeyen kuruluşlar için bir çözüm olarak önerilen CAWAL sayesinde kullanıcıların erişim bilgileri kuruluşun kendi veritabanı sunucularında ve yapılandırılmış biçimde saklanır. Şekil 3.12, CAWAL'ın sağlam veri toplama yöntemiyle elde edilen verilerin hem web analitiği hem de web kullanım madenciliği için bir veri kaynağı olarak verimli şekilde kullanılabileceğini göstermektedir.

CAWAL çerçevesi ile elde edilen zamansal veriler, ilişkisel veritabanında homojen bir yapıda saklanmaktadır. Bu verilerin kullanılması, WUM için ön işleme aşamasına olan ihtiyacı da ortadan kaldırmakta ve doğrudan bir sonraki adım olan örüntü keşfi için sağlıklı, temiz ve kullanıma hazır veriler sunmaktadır. CAWAL çerçevesi aracılığıyla toplanan veriler, yüksek veri doğruluğu ve güvenilirliği ile analitik süreçlerde hızlı ve kesin sonuçlar elde edilmesini sağlarken web kullanım madenciliği süreçlerini de optimize eder. Bu özellikler, kullanıcı davranışlarının kapsamlı ve derinlemesine analizine önemli ölçüde katkıda bulunur ve modelin web kullanım madenciliği disiplinindeki stratejik önemini pekiştirir.

3.9. CAWAL'ın Sunduğu Stratejik Kabiliyetler

CAWAL, gelişmiş çok boyutlu fonksiyonları sayesinde web kullanıcıları, web geliştiricileri ve web yöneticileri gibi farklı kullanıcı profilleri için özelleşmiş çözümler sunmaktadır. Her bir profil için çerçevenin sağladığı temel işlevsellikler, kullanıcı etkileşimlerini, sistem performansını ve güvenliğini optimize etmeyi amaçlar. CAWAL'ın farklı kullanıcı profillerine yönelik sergilediği stratejik yetenek ve kabiliyetler Tablo 3.6'da detaylı şekilde listelenmiştir.

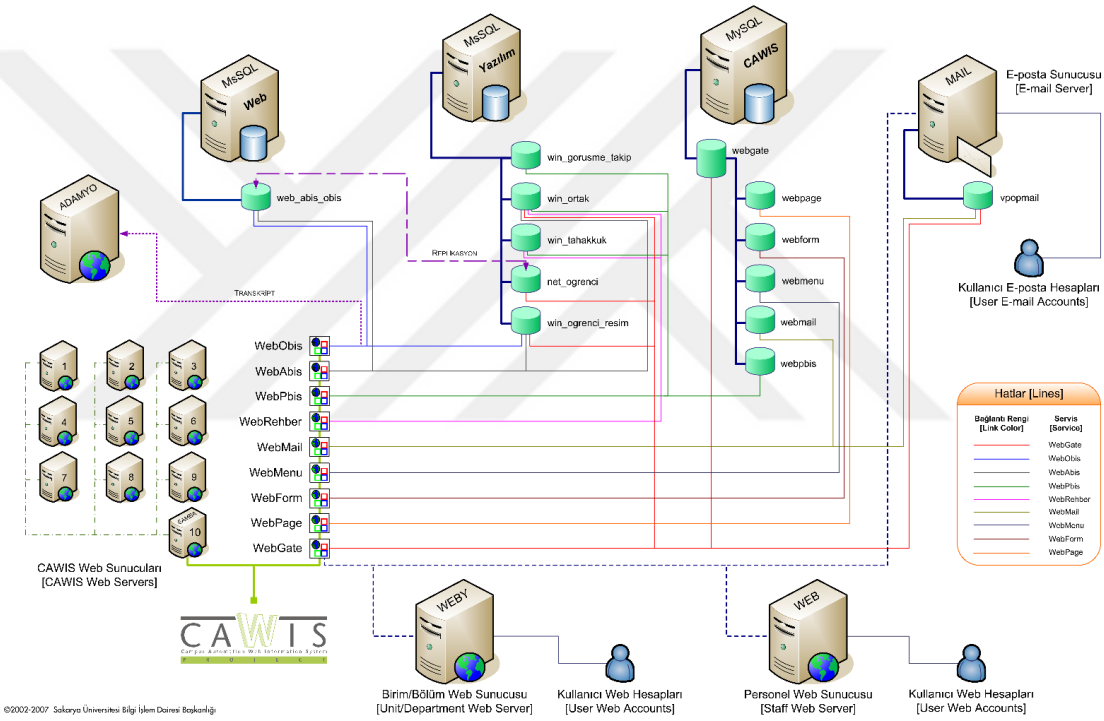
Bu kabiliyetler, çerçevenin kullanıcı gereksinimlerini ne ölçüde karşıladığını ve web portalının verimliliğini nasıl artırdığını ortaya koymaktadır. CAWAL, her bir kullanıcı profilinin özel ihtiyaçlarını göz önünde bulundurarak sistem performansı, güvenlik, kullanıcı deneyimi ve veri analitiği konusunda gelişmiş çözümler sağlar. Böylece çerçeve, kullanıcıların ve yöneticilerin taleplerine hızlı ve etkili bir şekilde yanıt vererek, web portalının genel işleyişini ve kullanıcı memnuniyetini iyileştirmektedir.

Tablo 3.6. CAWAL'ın farklı profillere sağladığı stratejik yetenek ve kabiliyetler.

Profiller	Stratejik Yetenek ve Kabiliyetler
Web kullanıcıları	• Mevcut aktif kullanıcı sayısını görüntüleyerek sistem yüküne ilişkin gerçek zamanlı görünürlük sağlar. • Kullanıcılara kişisel güvenlik ve kullanım verilerine erişim sağlayarak bireysel hesap verebilirliği artırır. • Kullanıcıların başarılı ve başarısız giriş denemelerini gözden geçirmelerine olanak tanıyarak proaktif güvenlik önlemlerini kolaylaştırır. • Kullanıcının ilk oturum açma ve üyelik süresi hakkında bilgi sağlayarak uzun vadeli katılımı ölçmeye yardımcı olur. • Kullanıcıları en son parola değişikliği konusunda bilgilendirerek güvenlik farkındalığı kültürünü teşvik eder. • Güvenlik gözetim özelliklerinin bir parçası olarak, IP adresi ve saati de dahil olmak üzere bir kullanıcının son oturum açma ayrıntılarını sunar. • Kullanıcılara önlem almalarına yardımcı olmak için son başarısız oturum açma girişiminden kritik güvenlik verileri sağlar. • Oturum açma ve kapatma işlemi sırasında tüm kullanıcı işlemlerini kaydederek kullanıcıların sistem faaliyetlerini kanıtlamalarına olanak tanır.
Web geliştiricileri	• Kullanıcıların gezinme davranışlarının incelenerek site organizasyonunun optimize edilmesi için ihtiyaç duyulan verileri sağlar. • IP adresi kaynağı, konum ve kullanıcı aracısı türü değişkenlerine göre içeriği uyarlamak için koşullu mantık kullanılabilmesine imkân tanır. • Kullanıcı arayüzünü tarayıcı özelliklerine ve dil ayarlarına göre uyarlayarak hedeflenen işlevler ve uyarılar sunar. • Sunucu tarafı sayfa oluşturma süresini ölçer ve günlüğe kaydeder, böylece performans değerlendirmesine katkıda bulunur. • Veritabanı sorgularıyla ilgili gecikmelerin belirlenmesini ve giderilmesini kolaylaştıran metrikler sağlar. • Web hizmetleri genelinde kullanıcı oturum uzunlukları ve gezinme modellerinin nicel analizlerini sunar. • Kullanıcıların uygulamada gerçekleştirdiği işlemleri ve bunların sonuçlarını tam ve kesin olarak (kanıt niteliğinde) doğrulama imkânı sunar. • Tanılama süreçlerine yardımcı olmak için kullanıcı uyarılarının ve hata mesajlarını kayıt altına alır. • Sorun giderme ve hata çözümüne yardımcı olmak için sayfalar arasında iletilen GET, POST ve Çerez (Cookie) verilerini yakalar. • Uygulama hizmetleri ya da modülleri ile kullanıcı etkileşimlerini detaylı analiz ederek hizmet sunumu yeteneklerini arttırır. • Tarayıcıda "yenile" veya "geri" işlevleri nedeniyle oluşan sayfa tekrarlarının tespit edilmesini ve incelenebilmesini sağlar.
Web/Sistem yöneticileri	• Ücretsiz ve sınırsız veri saklama imkânı sunar. • Analitik için güvenilir bir temel olarak hizmet veren yüksek veri doğruluğu sağlar. • Veri depolama için kurum sunucusunu kullandığından veri sahipliğini garanti eder. • Detaylı zaman tabanlı veri üretimini destekleyerek saatlik, günlük, aylık ve yıllık veri analitiğine olanak sağlar. • Analitik esnekliği arttıran sınırsız özel boyut ve metrik kullanımına olanak tanır. • Yüksek veri tazeliği sunarak anlık (ad-hoc) analizlere olanak tanır. • Gerçek zamanlı ya da geriye dönük istenen periyot için sistem değerlendirmelerine olanak tanıyan esnek izleme seçenekleri sunar. • Sunucular arasındaki kullanıcı dağılımına anında görünürlük sağlar ve etkin olmayan sunucuları verimli bir şekilde tanımlar. • Dinamik IP adreslerini tespit edebilir ve böylece veri bütünlüğünü korur. • Aynı IP veya kullanıcı hesabından gelen aşırı tekrarlayan talepler yoluyla güvenlik tehditlerini belirleyebilir. • Bot güdümlü ve insan güdümlü faaliyetler arasında ayrım yaparak veri kalitesini iyileştirir. • Kullanıcı tarafından talep edilen faaliyetleri doğrulamak ve uygulama hesap verebilirliğini artırmak için özellikler sunar. • Kullanıcı etkileşimlerinin uygulama başına ve hizmet başına düzeyde izlenmesine olanak tanıyarak hassas analizi kolaylaştırır. • Suç faaliyetlerine ilişkin bilgilerin kolluk kuvvetlerinin kullanımına sunulması için gerekli altyapıyı sağlar. • İstenmeyen kayıtların tespit edilmesini, çıkarılmasını ve kaldırılmasını kolaylaştırır. • Toplanan verilerin çok boyutlu tablolarda veya grafik formatlarda görselleştirilmesine olanak tanır. • Gelişmiş web analitiği ve web kullanımı madenciliği için yapılandırılmış veriler sunar.

4. UYGULAMA VE DEĞERLENDİRME

Tasarlanan CAWAL çerçevesi, tezin çok boyutlu hipotezini doğrulamak için Sakarya Üniversitesi'nin kurumsal web uygulaması olan "Kampüs Otomasyonu Web Bilgi Sistemi"ne (CAWIS) uygulanmış ve uzun dönemli erişim verileri toplanmıştır. Her servis için ayrı bir alt alan adı kullanan, web portalı yapısındaki CAWIS sisteminin mimari şeması Şekil 4.1'de sunulmuştur (Canay ve ark., 2011).



Şekil 4.1. CAWIS web portalı mimari şeması.

CAWAL çerçevesi, sunucu ve web hizmeti yoğunlukları da dahil olmak üzere tüm sistemin izlenmesinde, kullanıcı davranışlarının ve kullanım pratiklerinin anlaşılmasında, web uygulamasının buna göre yapılandırılmasında, uygulama kalitesinin ve güvenliğinin iyileştirilmesinde son derece etkili olmuştur. Uygulamadan elde edilen değerli sonuçlar, çerçevenin performans ve veri yönetişimi gibi çeşitli boyutlarda mevcut araçlara göre avantajlarını doğrulamaktadır.

4.1. Veri Anonimleştirme ve Zaman Düzeltmesi

Çalışmada, Sakarya Üniversitesi'nde geliştirilen ve günümüzde kullanımda olmayan CAWIS web portalından toplanan veriler kullanılmıştır. Veri toplama sürecinde, ilgili tarihte tüm portal kullanıcılarının onayladığı 'Sakarya Üniversitesi İnternet Servisleri Kullanım Politikası Sözleşmesi'nin hükümlerine uygun hareket edilmiştir. Bu sözleşmenin 13. maddesi, üniversiteye her türlü hizmeti denetleyebilme ve kullanıcı etkileşimlerini takip edebilme yetkisi vermektedir.

Araştırmanın sorumlu bir şekilde gerçekleştirilmesi için Sakarya Üniversitesi'nden gerekli izinler alınmıştır. Bu izinler, verilerin toplanması ve kullanılmasının, üniversitenin iç düzenlemelerine ve Türkiye Cumhuriyeti'nin yasal mevzuatına uygun olduğunu temin etmektedir. Ayrıca, bu çalışma Sakarya Üniversitesi Bilimsel Araştırma Projeleri (BAP) Komisyonu Başkanlığı tarafından desteklenmiştir (Proje No: 2010-50-02-024). BAP Komisyonu'nun desteği, araştırmanın yüksek bilimsel ve metodolojik standartlara uygun olarak yürütülmesine olanak sağlamıştır.

Kullanıcı gizliliğini korumak amacıyla, bu çalışmada veri anonimleştirme yöntemleri uygulanmıştır. Anonimleştirme sürecinin bir parçası olarak, toplanan verilerin zaman damgalarında değişiklik yapılmıştır. Orijinal veriler 2008-2012 yılları arasındayken, bu verilere +7 yıl eklenerek 2015-2019 yılları arasındaymış gibi gösterilmiştir. Bu değişiklik, kullanıcıların zamanla ilişkilendirilebilir aktivitelerini daha anonim hale getirmek ve olası veri ihlallerinde riski azaltmak amacıyla yapılmıştır. Bu zaman düzeltmesinin analiz sonuçları üzerinde olumsuz bir etkisi beklenmemektedir; zira çalışmanın odak noktası, mutlak zaman damgalarından ziyade, belirli bir zaman aralığındaki eğilimler ve davranış kalıplarıdır. Araştırmanın tüm aşamaları, katılımcıların gizliliğini koruma ve veri güvenliğini sağlama amacıyla dikkatle yürütülmüştür. Çalışmanın metodolojisi, veri toplama süreçleri ve analiz yöntemleri, çalışmanın şeffaflığını ve güvenilirliğini artırmak için tezde detaylıca açıklanmıştır.

4.2. Web Portalı Entegrasyonu

CAWAL çerçevesi, istisnalar, kullanıcı akışları, durum değişiklikleri, vb. gibi harici izleyiciler tarafından görülemeyen olayları kaydetmek için portal ile bütünleşik bir yapıda çalışır. Veri toplama API'si, form alanları gibi uygulamaya özel verilerin izleme günlüklerine eklenmesine olanak tanıyacak şekilde tasarlanmıştır. Oturum verileri, eksiksiz bir etkileşim takibi için uygulama sunucularının izlenmesiyle

desteklenir. CAWAL'ın veri toplama API'si web portalına, özellikle de kodun başlangıç kısmına entegre edilmiştir. Bu entegrasyon, portalın sistematik tasarımına uygun olarak API'nin her sayfa isteğinde çağrılmasını sağlar. Portal doğru bir şekilde yapılandırıldığında ve API sorunsuz bir şekilde entegre edildiğinde, günlüğe kaydetme karmaşıklığını soyutlayarak uygulama, özellikle de servis geliştiricilerin temel işlevlerine odaklanmalarına olanak tanır.

Portal kodlarının yazılması ve dağıtılması sırasında, CAWAL'ın veri toplama API'si sürekli olarak katmanlı çerçeve içinde çalışır ve varlığını hissettirmeden veri toplar. Portal genel şablonunun son kısmında, sayfa kodu üretimi tamamlanmadan hemen önce, sayfa yüklenme (oluşturulma) süresi, veritabanı sorgusu gecikmeleri, hata ve uyarı mesajları gibi üretilen sayfadaki işlevlerle ilgili ayrıntılar API aracılığıyla sayfa görüntüleme tablosunda güncellenir. Kod düzeyinde görünürlük sayesinde CAWAL, başka türlü elde edilmesi zor olan benzersiz içgörüler sağlar. Bu sistematik yaklaşım, temel verilerin aralıksız toplanmasını kolaylaştırarak uygulamanın sürekli iyileştirilmesine ve optimizasyonuna katkıda bulunur.

4.3. Web Çiftliği Entegrasyonu

CAWAL çerçevesinin on adet web sunucusundan oluşan bir web çiftliği ve çeşitli web hizmetlerini sunmak üzere birçok alt alan adını kapsayan gerçek bir kurumsal web portalı üzerinde uygulanması, kapsamlı web trafiği ve analiz gereksinimlerinin ele alınmasına yönelik benzersiz bir yaklaşımın örneğidir. CAWAL, farklı sunuculardaki işlemleri aynı anda takip etmek için iyi yapılandırılmış bir yük dengeleme mekanizması ile uyumlu bir şekilde çalışarak yoğun kullanıcı etkinliği dönemlerinde dahi sistemin güçlü performansını korumasını sağlamıştır.

Web çiftliğinde sunucuların oturum ve ayar klasörlerinin yönlendirildiği bir NAS sunucusunun kullanılması, CAWAL uygulamasına bir başka tutarlılık katmanı daha eklemiştir. Bu merkezi depolama çözümü, web çiftliği genelinde sürekliliği garanti ederek tek tip ve düzenli oturum veri yönetimi sağlamaktadır. Bu mimarinin yük dengeleme ve NAS sunucuları ile desteklenmesi, etkin bir ölçeklendirmeyi, kesintisiz hizmet sunumunu ve karmaşık web uygulamalarına esnek çözümler sağlamasını mümkün hale getirerek sistemin verimliliğini ve uygulamaların ihtiyaçlarına uyum sağlama kapasitesini arttırmaktadır.

4.4. Çerçevenin Zaman Karmaşıklığı

CAWAL çerçevesi aracılığıyla yapılandırılan sunucu tarafı uygulama tabanlı web analitiği, sunucuya yapılan n kullanıcı isteğinin her birinin bir günlük girişi oluşturduğu günlük toplama ile başlar. Bu, doğrusal bir zaman karmaşıklığı olan $O(n)$ ile sonuçlanır. Bir sonraki aşama veri dönüşümüdür; burada her bir günlük girdisi, IP adreslerinin coğrafi konumlara dönüştürülmesi gibi k adet dönüşümden geçebilir ve bu da $O(kn)$ karmaşıklığına neden olur. Bu işlenmiş bilgilerin veritabanında depolanması, veritabanının yapısı ve indeksleme gibi faktörlere bağlı olarak başka bir karmaşıklık katmanı ekler. Sonraki aşamalarda daha karmaşık işlemlerle karşılaşılır. Normalleştirme sırasında verilerin oturumlara veya kullanıcılara göre gruplandırılması $O(n^2)$ karesel karmaşıklığa yol açsa da optimize edilmiş algoritmalar bunu $O(n\log n)$ veya doğrusal zamana indirir.

Analitik çıkarma süreci ise eyleme geçirilebilir içgörüler elde etmek için veri toplama, işleme ve yorumlamanın karmaşık bir etkileşimidir. Toplama ve özet işlemleri, değişken karmaşıklığa sahip veri özetleri üretir. Basit toplamalar doğrusal olabilirken, karmaşık olanlar hesaplama açısından daha zahmetlidir. Bir web etkileşiminden toplanan ham verileri $\zeta = \{U_i, S_d, P_v, ...\}$ olarak temsil edelim. Bu veriler kullanıcı etkileşimleri U_i, oturum süreleri S_d, sayfa görüntülemeleri P_v vb. gibi çeşitli metriklerden oluşur.

Birleştirilmiş veriler $A(\zeta) = \sum_{t=başla}^{t=bitiş} \zeta_t$ şeklinde hesaplanır. Burada t belirli bir dönemdir ve ham verilerin belirli aralıklar veya kriterler üzerinden birleştirilmesiyle elde edilir. Bu birleştirilmiş veriler daha sonra $P(A)=f(A(\zeta))$ olarak temsil edilen örüntüleri ayırt etmek için istatistiksel fonksiyonlara (f) tabi tutulur. Daha sonra $Rap=\phi(P(A)$ kullanılarak kapsamlı bir rapor (Rap) oluşturulur ve bu rapor $Tav=\psi(Rap)$ şeklinde eyleme geçirilebilir tavsiyelerin türetilmesine rehberlik eder. Bu tavsiyeler, olası eylemler (Eyl) ile birlikte, optimizasyon fonksiyonu $\Omega=\omega(Tav, Eyl)$ hakkında bilgi verir. Sürecin etkinliği, beklenen ve gerçekleşen sonuçlar arasındaki sapmayı ölçen bir geri bildirim fonksiyonu $F_b=\delta(\Omega_{beklenen}, \Omega_{mevcut})$ ile değerlendirilir ve genel amaç, yinelemeli döngüler boyunca F_b'yi en aza indirerek istenen sonuçlarla uyumu sağlamaktır.

4.5. Özellik Karşılaştırması

Web analitiği için kullanılan yöntem ve araçların karşılaştırılması niteliksel ve niceliksel olarak yapılabilir. Seçilen yöntem ve araçları nitel özellikleri açısından çeşitli kriterlere göre inceleyerek ampirik bir değerlendirme yapmak mümkündür. Web sunucu günlükleri, Google Analitik (ücretsiz sürüm), açık kaynak web analiz araçları (OWA ile Matomo) ve CAWAL çerçevesinin özellik karşılaştırması Tablo 4.1'de verilmiştir.

Tablo 4.1. Çeşitli analitik araçların özellik ve yeteneklerine göre karşılaştırılması.

Özellik / Araç	Web Sun. Log (Apache)	Google Analitik (Bulut)	OWA/Matomo (Yerel)	CAWAL (Yerel)
Platform Bağımsızlığı	Hayır	Evet	Hayır	Hayır
Uygulama Bağımsızlığı	Evet	Evet	Evet	Hayır
Yoğun Yerel Kaynak Kullanımı	Hayır	Hayır	Evet	Evet
Kur. Gereksin. ve Kodlama Çabası	Düşük	Düşük	Orta	Yüksek
Kullanıcı İzleme (Taraf)	Sunucu-tarafı	İstemci-tarafı	İstemci-tarafı	Sunucu-tarafı
Kullanıcı İzleme (Taban)	Web S. tabanlı	JS tabanlı	JS tabanlı	Uyg. tabanlı
Web Sunucusuna Tek HTTP İsteği	Evet	Evet	Hayır	Evet
İstemci Tarayıcısına Ek Yük	Hayır	Evet	Evet	Hayır
Web Sunucusuna Ek Yük	Hayır	Hayır	Evet	Evet
(Alt)alanlar Arası Kullanıcı İzleme	Sınırlı	Sınırlı	Sınırlı	Evet
Oturum Zaman Aşımı Bildirimi	Hayır	Hayır	Hayır	Evet
Oturum Sınırı Tespiti	Kısıtlı	İyileştirilmiş	İyileştirilmiş	Uyarlanabilir
Gece Yarısı Sınırı İşleme	Basit	Kısmi	Kısmi	Tam
SPA ve PWA Entegrasyonu	Doğal destek	Çaba ile	Çaba ile	Doğal destek
Mobil ve IoT Cihaz Uyumluluğu	Doğal destek	Çaba ile	Kısmen	Doğal destek
Uygulamaya Özgü (Özel) Veriler	Hayır	Sınırlı	Sınırlı	Sınırsız
Yerel Veri Ambarı Altyapısı	Hayır	Hayır	Hayır	Evet
Veri Birikimi Yöntemi	Yerel (Hizmet)	Bulut (SaaS)	Yerel (API)	Yerel (API)
Ham Verilere Ücretsiz Erişim	Evet	Hayır	Evet	Evet
Veri Egemenliği ve Sahipliği	Evet	Hayır	Evet	Evet
Toplanan Veri Türü	Yarı yapılandır.	Yapılandırılmış	Yapılandırılmış	Yapılandırılmış
Veri Dönüşümünün Karmaşıklığı	Yüksek	Orta	Düşük	Orta
WA için Verinin Uygunluğu	Düşük	Yüksek	Yüksek	Yüksek
WUM için Verinin Uygunluğu	Orta	Hayır	Orta	Yüksek

Burada ücretsiz çözümlerle bir karşılaştırma yapılmış olsa da Google Analitik'in ücretsiz sürümündeki bazı kısıtlamaların ücretli sürüm olan GA360'ta bulunmadığını belirtmek gerekir. Bu yöntemler, gerçek zamanlı analiz yapma, ağır yük altında çalışma, standart dışı ve uygulama bazlı veri elde etme, gizlilik ve veri güvenliği gibi yetenekler açısından da değerlendirilmelidir.

CAWAL, veri doğruluğu, eksiksizliği, güncelliği, gizliliği, ölçeklenebilirliği ve gelecekteki veri ihtiyaçlarına ve içgörü araştırmalarına uyum sağlama yeteneği gibi avantajları açısından geleneksel yöntemlere kıyasla oldukça başarılıdır. Kuruluşun kullandığı diğer araç ve sistemlerle entegre olan CAWAL çerçevesi, zaman içinde artan veri miktarını ve kullanıcı trafiğini kaldırabilecek bir yapıda tasarlanmıştır.

4.6. Performans Karşılaştırması

Doğalarındaki farklılıklar nedeniyle, istemci tarafı ve sunucu tarafı web analiz araçlarını karşılaştırmak ve etkinliklerini birbirlerine karşı ölçmek karmaşık zorluklar içerir. Nitel yaklaşıma ek olarak, yöntemler arasında ortak payda arayışı çerçevesinde incelenecek istemci tarafındaki sayfa yükleme sürelerindeki farklılıklar nicel bir performans göstergesi olabilir. Bu bölümde, önerilen CAWAL çerçevesinin performansı, önde gelen açık kaynaklı analitik araçları olan OWA ve Matomo ile karşılaştırılmaktadır. Ancak, bu sistemlerin Google Analytics ve Adobe Analytics gibi SaaS şeklinde çalışan bulut tabanlı web analiz araçlarıyla doğrudan karşılaştırılmasını kısıtlayan doğal sınırlamaların vurgulanması zorunludur.

Deneysel ortamı oluşturmak için CAWAL çerçevesiyle birlikte OWA ve Matomo platformları, Intel i7 işlemci ve 16 GB RAM'e sahip bir bilgisayarda örnek kurumsal web portalıyla yerel olarak çalışacak hale getirilmiştir. Sunucu ve istemci rolleri de dahil olmak üzere tüm performans ölçümleri, üzerinde Windows 11 işletim sistemi çalışan bu iş istasyonunda gerçekleştirilmiştir. Web portalının düzgün çalışması için, alt alan adları da dahil olmak üzere tüm ana bilgisayar adları, Windows işletim sisteminin "hosts" dosyası aracılığıyla yerel bilgisayara (localhost) yönlendirilmiştir. İstek sırasında çalışacak günlük uygulamasının seçimi, deneysel web sitesinin ana sayfasına eklenen bir GET parametresi ile sağlanmıştır. Gelen parametre değerine göre ilgili analitik aracına ait JavaScript kod parçacığı (snippet) HTML içeriğine dahil edilerek istemciye gönderilmiştir.

4.6.1. İstemci tarafı performans analizi

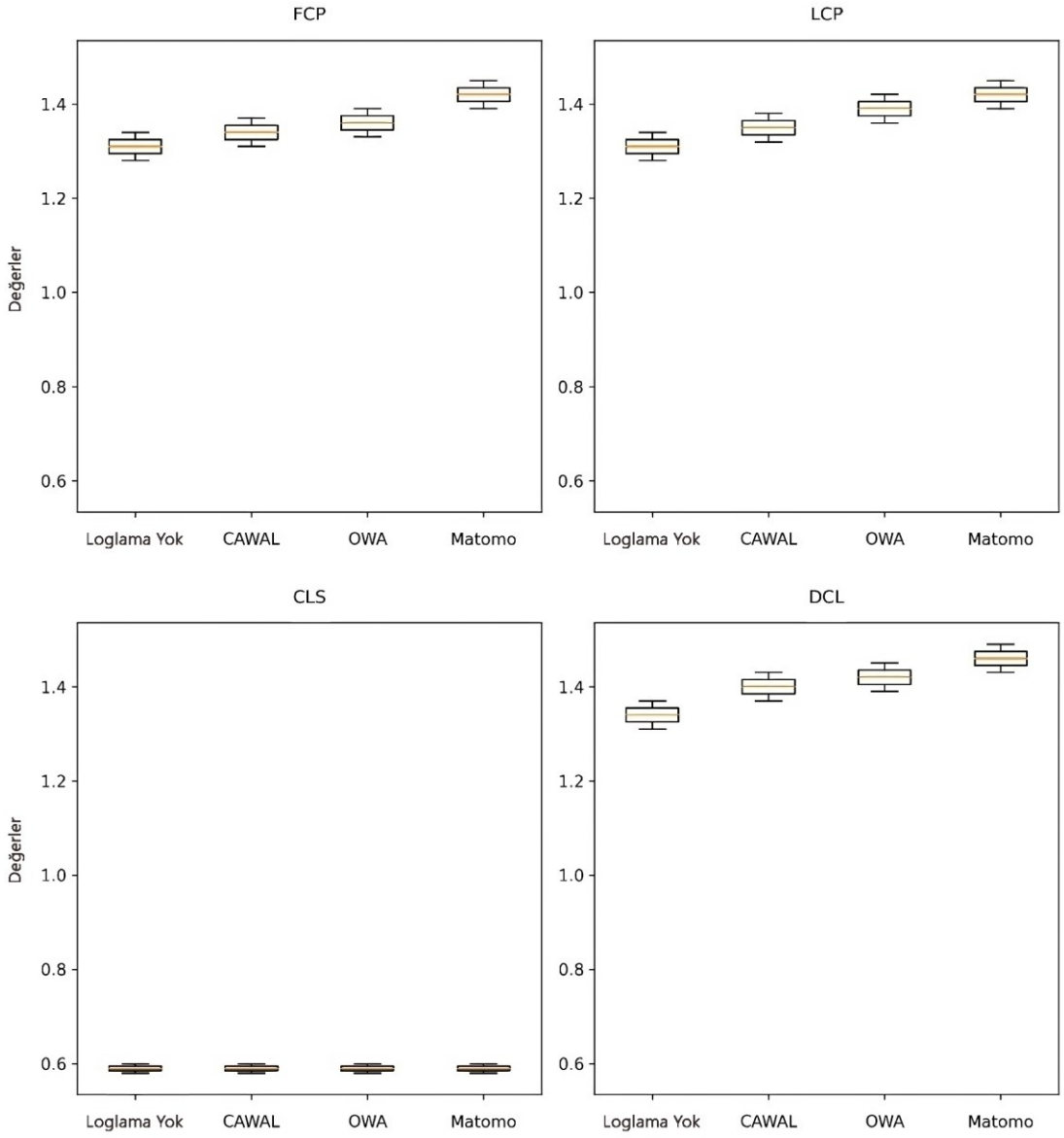
İstemci tarafında kullanıcı takibi yapan web analitik araçlarında, JavaScript tarafından oluşturulan oturum ve kullanıcı kimliği, yönlendiren, ekran çözünürlüğü ve iletilmesi gereken diğer bilgiler, GET veya POST türündeki API isteğine parametre olarak eklenir. Veri toplayıcı API, web uygulamasıyla aynı sunucuda çalışabileceği gibi farklı bir sunucuda da bulunabilir. Ancak bu deneysel çalışmada, seçilen web analitik

araçlarının hem istemci hem de sunucu tarafındaki yük performansı belirleneceği için tüm ortam tek bir bilgisayar üzerinde çalışacak şekilde kurgulanmıştır.

CAWAL günlük çerçevesi herhangi bir istemci tarafı işlemi içermezken, diğer iki istemci tarafı araçta biri JavaScript diğeri API olmak üzere iki adet birinci taraf isteği yapılmaktadır. Kullanıcıyı izlemek için 50 kilobaytın üzerinde bir JavaScript dosyasını indirip yorumladıktan sonra API'ye istek göndermek, istemci tarafında bir miktar gecikmeye neden olur. İzleme komut dosyaları genellikle eşzamansız olarak çalışsa da sunucuya ve istemcinin işlemcisine ilave yük bindirir. Ancak eşzamansız çalışmalarının bir avantajı olarak, kullanıcıların etkileşimi hissetme süreleri (TTI) (Hossfeld ve ark., 2018) üzerindeki olumsuz etkileri nispeten düşüktür.

Bu bağlamda gerçekleştirilen sayfa yüklemeleri, hızlı 3G ve CPU kısıtlaması olmadan Chrome DevTools (Bielak ve ark., 2022) içindeki performans içgörüleri aracı aracılığıyla ölçülmüştür. N=10 ardışık tekrar gerçekleştirilerek ortalama ölçüm süreleri saniye cinsinden elde edilmiştir. Çok sayıda çalışma, sonuçları ardışık beş çalıştırmanın ortancası üzerinden değerlendirmenin değişkenliği önemli ölçüde azalttığını ve test sürecini makul bir zaman dilimine indirgemek için yeterli olduğunu göstermektedir (Chan-Jong-Chu ve ark., 2020; Yu ve Benson, 2021; Heričko ve ark., 2021). Yüksek performanslı bir masaüstü ve düşük performanslı bir dizüstü bilgisayar gibi farklı cihazlarda test yapmak, ölçümleri dalgalandıracaktır. Ancak aynı bilgisayar ortamında araçların birbirlerine kıyasla performanslarının ölçülmesi ve bu işlemin birçok kez tekrarlanarak ortalamasının alınması bu etkiyi ortadan kaldırır ve ölçümleri daha anlamlı hale getirir.

Web uygulamalarının geliştirilmesinde kullanıcı deneyiminin önemi artmaya devam ettikçe, sayfa yükleme hızını ve yanıt verebilirliği doğru bir şekilde değerlendirmek ve optimize etmek için bir dizi performans ölçütü oluşturulmuştur (Armaini ve ark., 2022). FCP, LCP ve DCL toplu olarak bir web sayfasının performansının ve istemci tarafındaki kullanıcı deneyiminin kapsamlı bir resmini sunar (Dewi ve Nurdin, 2023; Saif ve ark., 2021). Şekil 4.2, bu metriklere göre ele alınan analitik araçların performans ölçümlerini sunmaktadır.



Şekil 4.2. İstemci performans ölçütlerine göre WA araçlarının karşılaştırılması.

Ölçüm sonuçları, araçların istemci tarafında küçük farklılıklarla birbirlerine yakın performans gösterdiğini ortaya koymaktadır. İstemci tarafındaki en hızlı sayfa yüklemesi, beklendiği gibi, web portalı herhangi bir analitik aracı olmadan çalıştırıldığında gerçekleşmiştir. Yalın ve etkin yapısıyla CAWAL, sayfanın diğer araçlardan daha hızlı yüklenmesini sağlamış, onu OWA ve Matomo takip etmiştir. Matomo genellikle FCP, LCP ve DCL metriklerinde en yüksek değerleri sergilerken, CAWAL'ın performansı daha hızlı sayfa yükleme kabiliyeti beklentileriyle uyumludur. CLS metriği, çerçeveler arasında minimum varyasyon göstererek tek tip düzen kararlılığına işaret etmektedir.

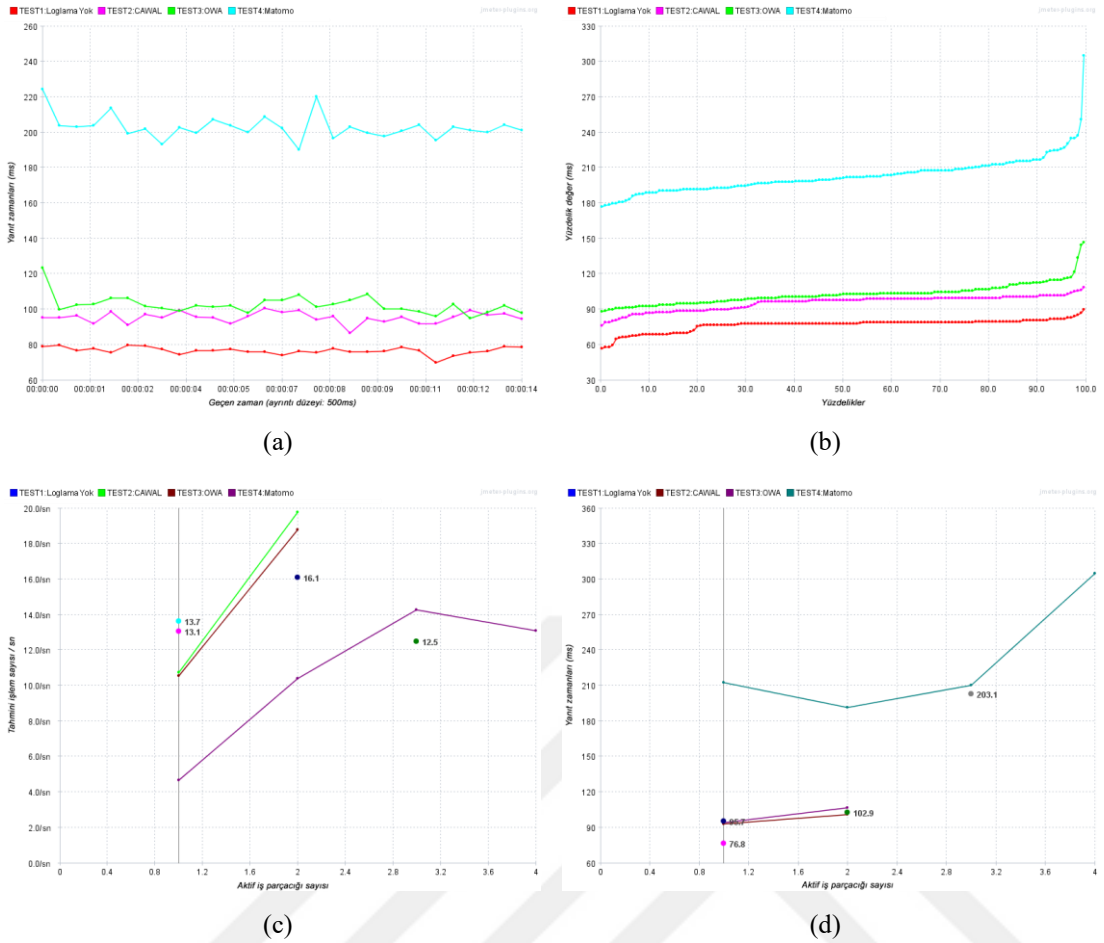
Elde edilen bilgiler, belirli performans gereksinimlerine ve optimizasyon hedeflerine dayalı olarak uygun platformun seçilmesinde değerli bir rehberlik sağlar. Kontrollü koşullar altında gerçekleştirilen bu analiz, CAWAL çerçevesinin tarayıcıda ek bir yük oluşturmadığını göstererek istemci performansı açısından etkin bir çözüm sunduğunu ortaya koymaktadır.

4.6.2. Sunucu tarafı performans analizi

Çalışma kapsamında, sunucu tarafı performans karşılaştırması için oluşturulan test ortamı üzerinde Apache JMeter (Junmei ve Jihong, 2019) aracının 5.5 sürümü kullanılarak Uygulama Performans Ölçümü (APM) gerçekleştirilmiştir. Apache JMeter, performans testi ve yazılım hizmetlerinin işlevsel davranışını ölçmek için tasarlanmış, bu alanda önde gelen bir açık kaynak aracıdır. JMeter, protokol düzeyinde faaliyet göstererek, bir ya da birden fazla tarayıcının işlevlerini andıran tarayıcı benzeri davranışları başarıyla simüle eder (Vasylyshyn ve ark., 2023). Bununla birlikte, JMeter'in işlevlerinin tarayıcı yeteneklerinin tamamını kapsamadığını vurgulamak önemlidir. JMeter, HTML sayfalarına gömülü JavaScript kodunu çalıştıramaz ve HTML sayfalarını geleneksel tarayıcıların karakteristik özelliklerine uyumlu şekilde oluşturamaz. Diğer taraftan, uygun şekilde tasarlanmış bir JMeter testi, tarayıcı kullanan gerçek kullanıcıların etkileşimlerini yakından yansıtabilir.

Çalışmada, sunucu tarafı performans analizini gerçekleştirmek amacıyla istemci tarafı çalışan web analitik araçlarının JavaScript ve API istekleri dikkate alınarak uygun şekilde yapılandırılmış özel bir test komut dosyası oluşturulmuştur. İlk olarak, web portalının analitik barındırmayan sürümü ve ardından karşılaştırılacak araçlar, GET parametresine verilen farklı bir değer kullanılarak ayrı ayrı test edilmiştir. Her test için 15 saniyede 150 iş parçacığı (istek) yürütülmüş ve bu sonuçlar daha sonra Şekil 4.3'te gösterilen grafikleri oluşturmak üzere birleştirilmiştir.

Grafik (a), web portalının herhangi bir günlük kaydı olmaksızın temel yanıt süresinin ortalama 76 ms olduğunu göstermektedir. CAWAL entegrasyonu ile bu süre bir miktar artarak 95 ms'ye çıkmaktadır ki bu OWA'nın 101 ms'lik yanıt süresine kıyasla çok daha düşük bir artıştır. En büyük sapma ise Matomo ile yaşanmakta, burada ortalama yanıt süresi 204 ms'ye yükselmektedir.

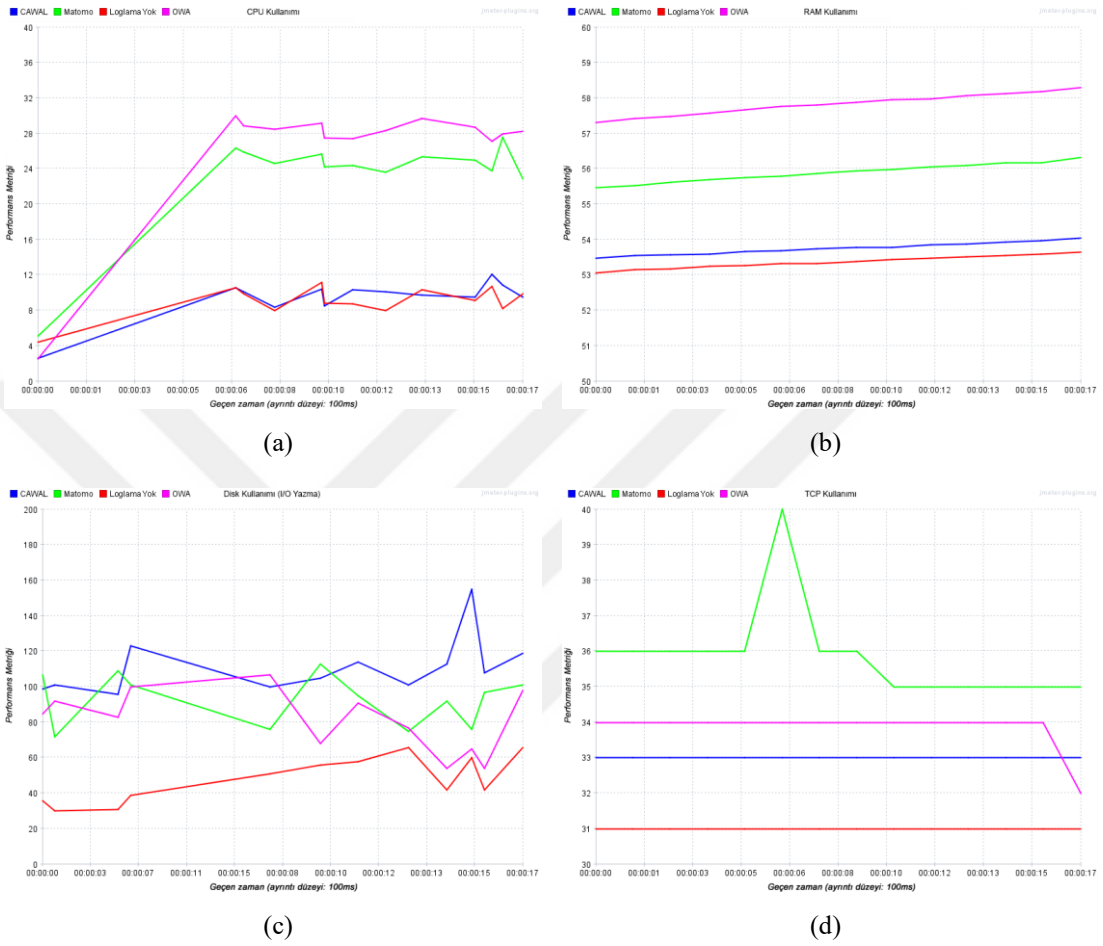


Şekil 4.3. Seçilen analitik araçların yük testi yanıt performansı karşılaştırması.

Diğerleriyle kıyaslandığında CAWAL, OWA'ya göre yaklaşık %24, Matomo'ya göre ise %85,16 daha hızlı yanıt süresi sunmaktadır. Grafik (b), farklı çözümlerin kümülatif yüzdelik dilimlerine odaklanarak benzer sonuçları ortaya koymaktadır. Saniye başına işlem sayısını inceleyen Grafik (c)'de CAWAL'ın performansı OWA'ya yakın ancak ondan daha yüksek, Matomo'nun performansı ise belirgin şekilde tümünden daha düşüktür. Zira bir sistemin saniyede daha fazla talebi karşılayabilmesi, daha yüksek performansı ifade eder.

Son olarak Grafik (d), aktif iş parçacığı sayısına göre yanıt sürelerini karşılaştırmaktadır. Bulgular, CAWAL ile OWA'nın benzer süreler sergilediğini, Matomo'nun ise neredeyse iki kat daha uzun sürede yanıt verdiğini göstermektedir. Bu analiz sonuçları, CAWAL'ın farklı performans metriklerine dayalı olarak OWA ve bilhassa Matomo'dan daha iyi bir performans sergilediğini ortaya koymaktadır.

Yük testinin yanı sıra, sistem performansının bütünsel bir şekilde anlaşılması için donanım kaynak kullanımının incelenmesi de önemlidir. Şekil 4.4, yük testi sırasında JMeter'in PerfMon modülü aracılığıyla ölçülen CPU, bellek, disk ve TCP gibi sunucu kaynaklarının kullanımını göstermektedir.



Şekil 4.4. Seçilen analitik araçların yük testi kaynak kullanımı karşılaştırması.

Grafik (a) incelendiğinde, CAWAL'ın CPU tüketimindeki performansının, etkin işlem akışını yansıtarak, günlük kaydı olmayan karşılaştırma senaryosuna oldukça yakın olduğu görülmektedir. Buna karşılık Matomo, hesaplama karmaşıklıkları veya yoğun veritabanı işlemleri nedeniyle artan kullanım sergilemektedir. Grafik (b)'deki bellek kullanımı bilgileri, Matomo'nun daha fazla hafıza kullandığına ve potansiyel olarak iyileştirilmesi gereken önbellekleme stratejilerine işaret ederken, CAWAL ve OWA mimari farklılıklara rağmen birbirine yakın verimlilik sunmaktadır. Grafik (c)'de gözlemlenen, CAWAL'ın veritabanında daha az tablo kullanımına rağmen diğer çözümlere kıyasla daha yoğun yazma işlemi gerçekleştirmesi, kaydettiği veri miktarının daha fazla olmasıyla ilişkilendirilebilir. Grafik (d)'deki ağ kullanımı,

Matomo'nun karmaşık tasarımını ve istemci taraflı çalışan her iki yöntemin daha yüksek ağ trafiğine neden olduğunu ortaya koymaktadır.

Bu bulgular CAWAL, OWA ve Matomo'nun farklı iç tasarımlarına dayanan beklentileri doğrular niteliktedir. Bu noktada CAWAL, değerlendirilen diğer araçlardan veri kaydı ve işleme mekanizmaları açısından sistematik bir şekilde ayrıldığını göstermektedir. CAWAL'da kullanıcı izleme verileri iki temel veritabanı tablosunda toplanırken, analitik çıkarımı gece yarısından sonra, sunucu yükünün çok daha düşük olduğu zamanlarda gerçekleşir. Optimum kaynak kullanımı sağlayan bu yaklaşım, gerçek zamanlı veri işleme gerektiğinde hızlı hesaplamalara olanak tanıyan bir web çiftliği üzerinden sürdürülen dağıtılmış bir yapı ile daha da güçlendirilmiştir. Kaynakların bu şekilde stratejik yönetimi ve işlem görevlerinin etkin tahsisi, sistem performansını ve taleplerinin bütünsel olarak anlaşılmasını örneklemekte ve analitik için optimize edilmiş ideal bir yaklaşımı yansıtmaktadır.

Yanıt süreleri, işlem sayıları ve aktif iş parçacıkları gibi çeşitli performans göstergeleri üzerinden yapılan detaylı analizler, CAWAL'ın yoğun veri işleme ihtiyacı olan yüksek trafikli koşullarda, kıyaslandığı diğer açık kaynaklı web analitik araçlarına göre daha üstün bir performans sergilediğini ortaya koymaktadır. Bu sonuçlar, CAWAL'ın düşük kaynak tüketimiyle birleştiğinde, karmaşık web ortamlarında yüksek hacimli veri analitiğini verimli bir şekilde gerçekleştirebilme kapasitesini ve etkin performans sergileme yeteneğini açıkça ortaya koymaktadır.

5. UYGULAMANIN ANALİTİK SONUÇLARI

Yürütülen tez çalışmasında hiçbir varsayımsal ya da sentetik veri kullanılmamıştır. Aksine, analizler modelin canlı bir kurumsal web portalına uygulanmasından elde edilen gerçek verilere dayanmaktadır. İlişkisel veritabanında toplanan veriler, oluşturulan karmaşık sorgularla analiz edilebilmektedir. Verilerin depolandığı tablolardaki bir kaydın boyutu, verinin içeriğine bağlı olarak değişmekle birlikte, oturum tablosu için yaklaşık 250 bayt (~0,25KB), sayfa görüntüleme tablosu için ise yaklaşık 1000 bayt (~1KB) civarındadır.

Oluşturulan veri kümesi, site katılımı, demografik profiller, cihaz türleri, ağ yönlendirmesi, sunucu yapılandırmaları ve farklı erişim modları dahil ancak bunlarla sınırlı olmamak üzere çok çeşitli ölçümleri ve özellikleri kapsamaktadır. Bu geniş veri seti, zamana dayalı eğilimlerden karşılaştırmalı değerlendirmelere kadar çeşitli analitik incelemeleri mümkün kılar. Amaç ve hedeflere bağlı olarak bu veriler, herhangi bir WUM tekniğine de kolayca uyarlanarak anomali tespiti ya da sistem yükü tahmini gibi amaçlar için de kullanılabilir.

Çalışmada, kullanıcı davranış kalıpları, coğrafi eğilimler ve çok sunuculu bir mimaride farklı alanlardaki kullanıcıları izlemenin verimliliği gibi birçok faktör incelenmiştir. Sağlam veri toplama yoluyla elde edilen içgörüler, CAWAL'ın web analitiği ve web kullanım madenciliğindeki potansiyel uygulamalarının, sonuçlarının ve gelecekteki araştırma yönlerinin derinlemesine analizine zemin hazırlamaktadır. Toplanan verilerin analizi, metodolojinin güvenilirliğinin yanında, tutarlı sonuçlar veren ve gelecekteki araştırmalar için yeni yönler öneren sağlamlığını da doğrulamaktadır.

Kurumsal bir web portalı olan CAWIS'te uygulanan CAWAL çerçevesi, sistemde kimlerin olduğu, hangi portal servisinde ya da sunucuda kaç kullanıcının işlem yaptığı gibi gerçek zamanlı bilgileri detaylı olarak sunabilmektedir. Çerçevenin analitik çıkarımı işlemlerinden sonra elde edilen geniş çaplı sonuçlar ise kısa, orta ve uzun vadeli analitikler olmak üzere üç başlık altında incelenecektir.

CAWAL tarafından elde edilen analitik verilerin miktarı ve kalitesi, bu bölümde ele alınan analizlerden çok daha fazlasının gerçekleştirilmesine imkân verecek düzeydedir. Yapılan analizlerin derinliği ise toplanan verilerin standart bir web analitik aracıyla elde edilemeyecek düzeyde kapsamlı olduğunu ispatlar niteliktedir.

5.1. Kısa Vadeli Analitikler

Kısa vadeli analitikler, "log_analytics" tablosundaki bir kayda karşılık gelen günlük verilerdir ve istemci, sunucu, servis ve kullanım gibi elde edilmesi mümkün olan tüm bilgileri çok boyutlu biçimde özetler. Modele göre toplanan verilerin anlamlılığını göstermek için 2018 yılında hafta içi bir gün boyunca toplanan veriler incelenmiştir. Analiz edilen 24 saatlik örnek veri, 22.104 oturum ve 161.672 sayfa görüntülemesini içermektedir. Bu kısa vadeli tanımlayıcı (descriptive) analitiklerin sonuçları, sonraki bölümlerde ele alınacak orta ve uzun vadeli periyotlara yönelik tahminsel (predictive) analizler için temel bir referans noktası oluşturmaktadır.

5.1.1. Oturum ve sayfa gösterimi sayısına göre kullanım analizi

Portalda ziyaretçiler ve kayıtlı kullanıcılar tarafından gösterilen sayfa sayıları, sistemin nasıl kullanıldığı hakkında önemli bilgiler sağlamaktadır. Genel olarak "ziyaretçi" (bazen de "kullanıcı") kavramı sistemdeki herkesi ifade ederken, kullanıcı türünün ayırt edildiği durumlarda "kullanıcı" (ya da kayıtlı kullanıcı) oturum açmış, diğer bir ifadeyle sisteme giriş yapmış kullanıcıları, "misafir" ise oturum açmamış kullanıcıları tanımlar. "Misafir" olarak sisteme adım atan ziyaretçiler, sisteme giriş yapmalarının ardından "kullanıcı" olarak adlandırılırlar. Misafir ve kayıtlı kullanıcıların incelenen periyotta ziyaret ettikleri sayfa sayıları ve frekans dağılımları, Tablo 5.1'de verilmiştir.

Tablo 5.1. Ziyaretçi türlerine göre kategorik sayfa gösterimi sayıları.

Ziyaretçi Tipi	Sayfa Gösterimi	Frekans
Misafir	1-3 pages	20.721
	4-10 pages	1.195
	11-30 pages	62
	31-100 pages	3
	101+ pages	1
Kullanıcı	1-3 pages	5.619
	4-10 pages	9.298
	11-30 pages	2.262
	31-100 pages	282
	101+ pages	13

Bu analiz, kayıtlı kullanıcıların ve ziyaretçilerin hemen çıkma oranını belirlemek ve sayfa tarama alışkanlıklarını ortaya çıkarmak açısından önemlidir. Bir oturumda 100'den fazla sayfa gezen kullanıcıların riskli davranışları ayrıca incelemeye değerdir.

5.1.2. Kullanıcı türü ve cinsiyete göre kullanım analizi

Kullanıcı türü ve cinsiyete göre kullanıcı sayısı, oturum, sayfa görüntüleme, oturum başına sayfa görüntüleme ve toplam oturum süreleri saniye, dakika ve saat olarak Tablo 5.2'de verilmiştir. Ayrıca, ortalama oturum ve sayfa süreleri de kolaylıkla ortaya konulabilmekte ve kullanıcı davranışları üzerine değerlendirmeler yapılabilmektedir.

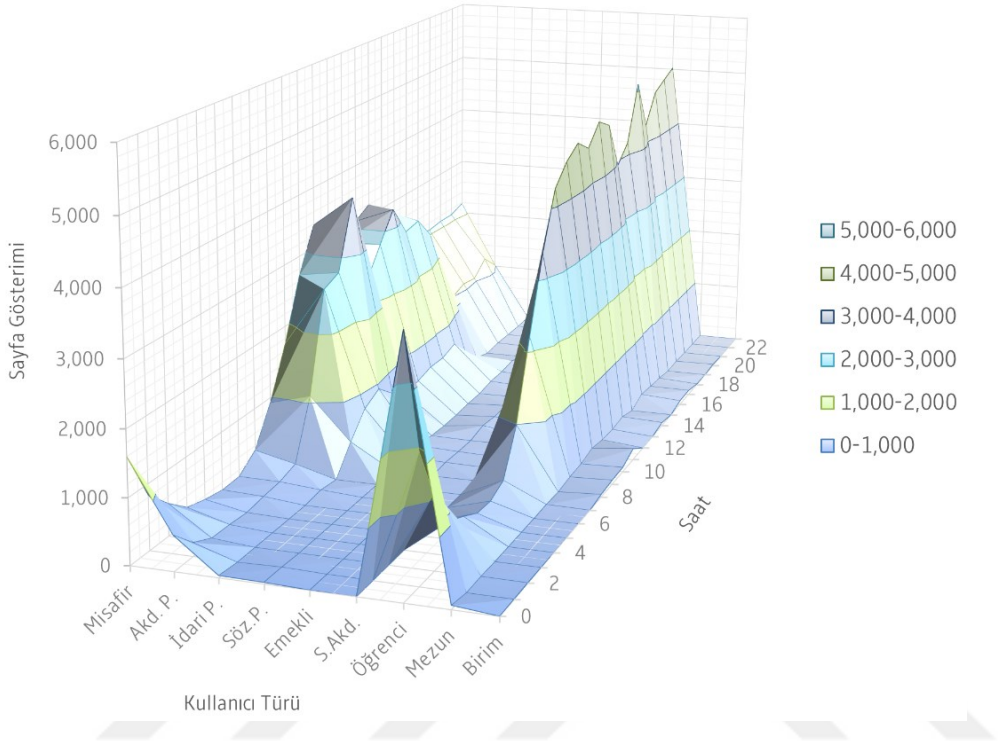
Tablo 5.2. Kullanıcı türleri ve cinsiyetlere göre kullanım bilgileri.

Kullanıcı Tipi	Cinsiyet	Kul. Sayısı	Oturumlar	Sayfa Gör.	Otur. Başına Sayfa Gör.	Toplam Oturum Süresi (sn)	Toplam Otur. Süresi (dk)	Toplam Otur. Süresi (sa)
Misafir	Belirsiz	5.343	4.655	9.006	1,93	-	-	-
Akademik Personel	Erkek	678	2.355	28.893	12,27	778.495	12.975	216,2
	Kadın	261	966	12.413	12,85	363.953	6.066	101,1
İdari Personel	Erkek	217	548	5.606	10,23	123.551	2.059	34,3
	Kadın	74	205	2.193	10,70	79.217	1.32	22,0
Sözleşmeli Personel	Erkek	14	44	479	10,89	24.864	414	6,9
	Kadın	7	16	99	6,19	2.918	49	0,8
Emekli Personel	Erkek	24	47	438	9,32	7.874	131	2,2
	Kadın	4	6	67	11,17	5.103	85	1,4
Öğretim Elemanı (Sözleşmeli)	Erkek	11	24	309	12,88	11.576	193	3,2
	Kadın	1	4	44	11,00	2.579	43	0,7
Öğrenci	Erkek	5.259	7.284	57.596	7,91	1.383.957	23.066	384,4
	Kadın	4.165	5.747	42.479	7,39	950.976	15.85	264,2
Yüksek Lisans	Erkek	50	81	706	8,72	16.18	270	4,5
	Kadın	32	46	445	9,67	18.054	301	5,0
Birim/Bölüm	Belirsiz	41	76	899	11,83	35.658	594	9,9
Toplam / Ort.		16.181	22.104	161.672	7,31	3.804.955	63.416	1.057

Ortalama sayfa izleme ve oturum süreleri, genelde portalın, özelde ise bir portal servisi ya da modülünün kullanıcıları ne kadar süre tutabildiğini ve etkileşim düzeyini göstermesi açısından önemlidir. Portal servisleri farklı kullanıcı türlerine çeşitli hizmetler sunarlar ve gerek kullanıcı ilgisi gerekse işlem süreleri servislerin karakteristiklerini belirler. Örneğin, "obis" ve "mail" gibi uzun etkileşim gerektiren servisler kullanıcıların portalda geçirdikleri süreyi artırırken, "menu" ve "rehber" gibi daha az adım gerektiren ve hızlı işlem yapılan servisler bu süreyi azaltabilirler. Servis bazlı daha kapsamlı ve detaylı analizler yapılması, servis niteliklerini ve kullanım karakteristiklerini daha net oraya koyacaktır.

5.1.3. Kullanıcı türüne göre saatlik kullanım analizi

Sistemin saatlik kullanımını kullanıcı türlerine göre çok boyutlu olarak elde etmek mümkündür. Şekil 5.1'de sunulan veriler, kullanıcı türlerinin günün farklı saatlerinde nasıl değişken bir kullanım örüntüsüne sahip olduğunu ortaya koymaktadır.

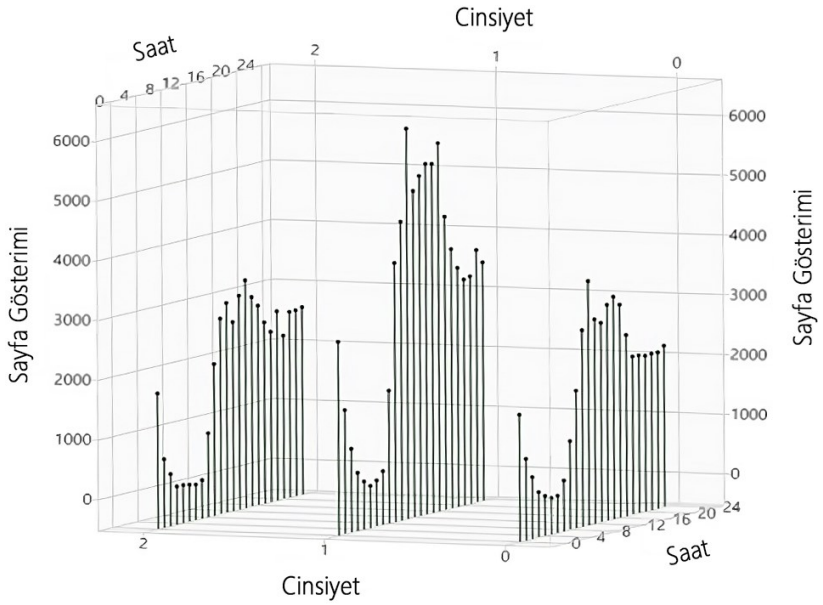


Şekil 5.1. Kullanıcı türlerine göre saatlik sayfa görüntüleme sayıları.

Akademik ve idari personelin pik kullanım saatleri net bir şekilde iş saatlerine denk gelirken, öğrencilerin daha esnek saatlerde sistem etkileşiminde bulunduğu gözlemlenmiştir. Bu kullanım örüntüleri, sunucu kapasitesinin zamanlaması ve boyutlandırılması ile güvenlik önlemlerinin stratejik olarak planlanması için doğrudan bir yönlendirme sağlar. Ayrıca, beklenmeyen kullanım artışları, sistem üzerindeki anormal etkinliklerin erken uyarı sinyalleri olarak değerlendirilebilir, böylece proaktif güvenlik önlemlerinin alınmasına olanak tanır. Bu veriler, sistem yönetimi ve güvenlik stratejilerinin geliştirilmesinde değerli bir rehber olarak işlev görür.

5.1.4. Cinsiyete göre saatlik kullanım analizi

Sistemin kullanımını belirli metrikler ve oransal değerler aracılığıyla anlamak, verimliliğini ve erişilebilirliğini değerlendirmek için çok önemlidir. CAWAL'ın verileri çok boyutlu olarak analiz etme becerisini gösteren Şekil 5.2, günlük sayfa görüntülemelerinin saat ve cinsiyete göre zaman dağılımını gösteren bir görselleştirme sunmaktadır.

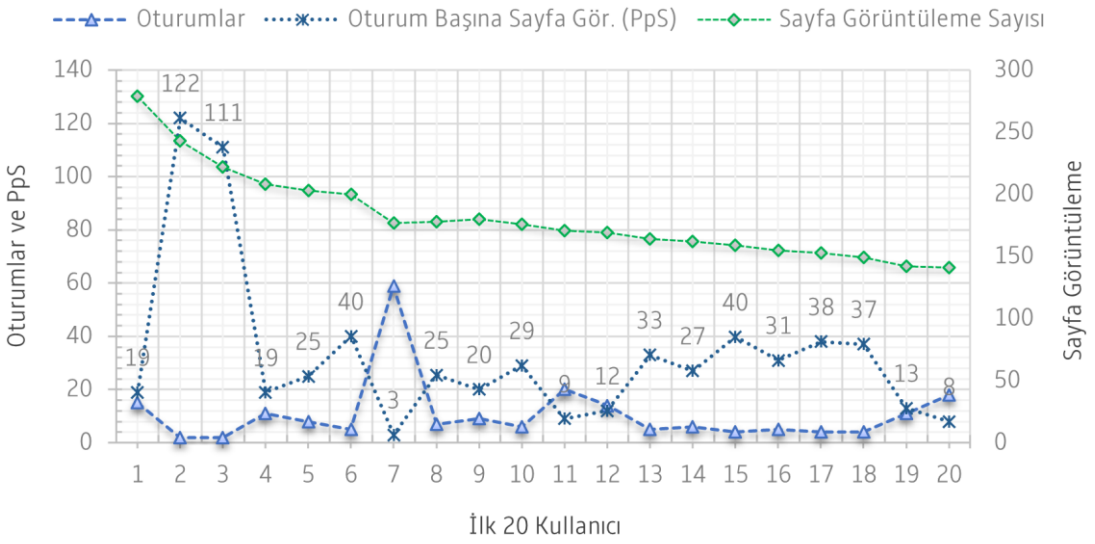


Şekil 5.2. Günlük sayfa görüntülemelerinin saat ve cinsiyete göre dağılımı.

Farklı noktalarla işaretlenmiş her bir çizgi, üç kategori için sayfa görüntüleme sayısını temsil etmektedir: Belirsiz (0), Erkek (1) ve Kadın (2). Bu grafiksel gösterim, farklı cinsiyetteki kullanıcıların gün boyunca web sitesiyle etkileşim kalıplarını etkili bir şekilde tasvir etmektedir. Hem birim, bölüm ya da görev hesaplarını hem de oturum açmamış ziyaretçileri içeren "belirsiz" kategorisi kendine özgü bir eğilim göstermektedir ve 24 saatlik döngü boyunca tarama modellerindeki farklılıklar belirgindir. Örneğin, erkekler öğlen saatlerinde zirve yaparken, kadınlar için örüntü daha dağınıktır. Belirsiz kategorisinin daha yakından incelenmesi, birim hesapların davranışsal özelliklerini sıradan ziyaretçilerden ayırarak daha fazla segmentasyona yol açabilir. Bu tür içgörüler kullanıcı katılımı uygulamalarına derinlik katarak web analitiği ve kişiselleştirilmiş kullanıcı deneyimlerinde yenilikçi araştırmalara kapı aralayacaktır.

5.1.5. Kullanıcı hesabına göre yoğun kullanım analizi

Tüm oturumlar içinde en çok sayfa görüntüleme yapan ilk 20 kullanıcının (kullanıcı adları gizlenmiştir) sayfa görüntüleme ve oturum sayıları Şekil 5.3'te gösterilmiştir. Bu analiz ile kullanıcıların sistemi ne kadar kullandıkları ya da bir anlamda meşgul ettikleri tespit edilebilmektedir. Benzer bir analiz de bir oturumda en çok sayfa görüntülemesine göre ilk 20 kullanıcının sayfa görüntülemeleridir. Tek bir oturumda çok yüksek sayfa görüntülemeleri tehlikeli kullanıcı eylemlerine işaret edebilir ve anomalileri tespit etmek için kullanılabilir.



Şekil 5.3. En çok sayfa görüntüleme yapan 20 kullanıcının kullanım bilgileri.

5.1.6. IP adresine göre yoğun kullanım analizi

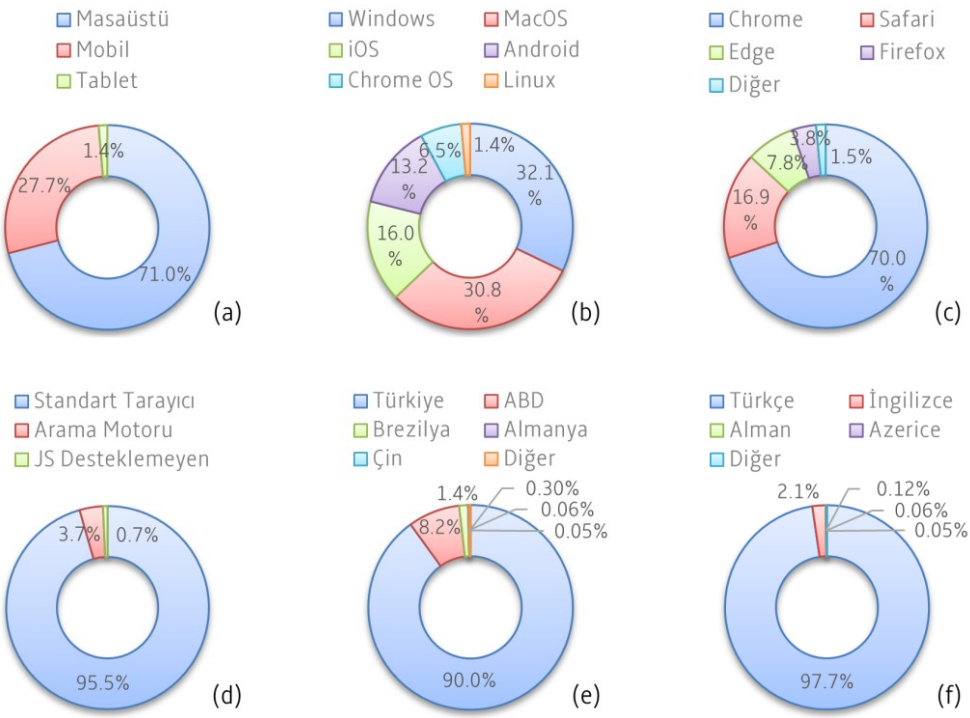
Açılan oturum sayısına göre sıralanmış ilk 20 IP adresinin (gizlenmiştir) ziyaret ve sayfa görüntüleme sayıları Tablo 5.3'te verilmiştir. Vekil sunucu IP adreslerine göre de yapılabilen bu tür analizler, yoğun veya tehlikeli kullanımların tespitinde önemlidir.

Tablo 5.3. En çok ziyaret eden 20 IP adresinin etkileşim sayıları.

IP adresi	Oturumlar	Sayfa Gösterimleri	Oturum Başına Sayfa Gösterimleri
IP1	416	933	2,24
IP2	134	706	5,27
IP3	108	163	1,51
IP4	104	642	6,17
IP5	102	556	5,45
IP6	98	474	4,84
IP7	85	416	4,89
IP8	82	501	6,11
IP9	81	455	5,62
IP10	78	422	5,41
IP11	76	303	3,99
IP12	69	464	6,72
IP13	66	368	5,58
IP14	65	408	6,28
IP15	63	391	6,21
IP16	61	374	6,13
IP17	57	346	6,07
IP18	56	333	5,95
IP19	55	86	1,56
IP20	54	128	2,37

5.1.7. Cihaz ve tarayıcı aracı verilerinin analizi

İstemci odaklı cihaz ve web tarayıcı (user-agent) verilerinin teknik analizi Şekil 5.4'te sunulmuştur. Bu tür teknik analizler site erişilebilirliği, desteklenebilirlik ve kullanılabilirlik gereksinimleri açısından kritik öneme sahiptir. Grafik (a)'da kullanılan cihaz türleri, Grafik (b)'de işletim sistemleri, Grafik (c)'de tarayıcı adları ve Grafik (d)'de tarayıcı türleri ile ilgili dağılımlar sisteme erişen istemciler analiz edilerek elde edilmiştir.



Şekil 5.4. İstemci odaklı cihaz ve web tarayıcı verilerinin teknik analizi.

Aynı şekilde Grafik (e)'de kullanıcıların geldiği ülkeler, Grafik (f)'de ise tarayıcı dilleri oransal olarak sunulmuştur. Tarayıcı dağılımı ya da dili gibi bilgiler, web sitesi kodlandıktan sonra farklı tarayıcılar üzerinde gerçekleştirilen tasarım ve betik (script) testlerinin gerekliliğini ortaya koymaktadır. Web sitelerinin kurumların dünyaya açılan kapıları olduğu düşünüldüğünde bu bilgi daha da anlamlı hale gelmektedir.

5.1.8. Konum tabanlı analiz

Web analitiğinin önemli bir yönü, kullanıcıların coğrafi dağılımını ve etkileşim modellerini anlamayı içerir. Tablo 5.4, gelen IP adreslerinin kaynağına göre kullanıcı etkileşimlerinin bölünmesini vurgulamakta ve verileri kuruluş içi, ülke içi ve ülke dışı olmak üzere üç farklı gruba ayırmaktadır. Bu kategorizasyon, sistemin coğrafi

erişimini ve dahili kullanımını anlamayı kolaylaştırarak kullanıcı katılımı ve yerelleştirme stratejilerine rehberlik eder.

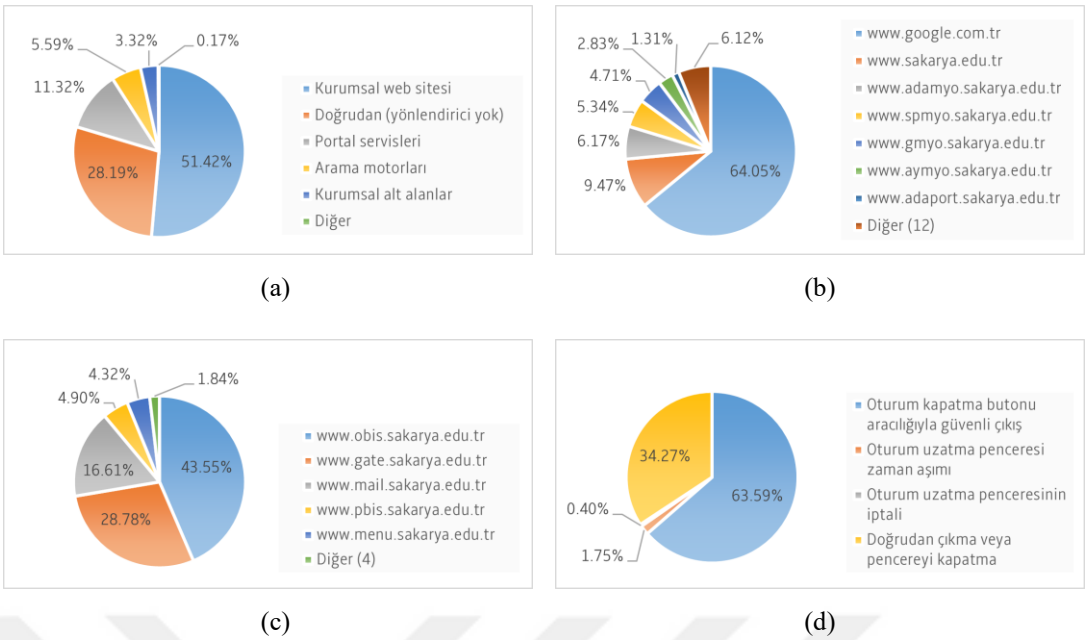
Tablo 5.4. İstemci konumuna göre oturum ve sayfa görüntüleme sayıları.

Gelen IP'nin Kökeni	Kullanıcı Sayısı	Oturum Sayısı	Sayfa Gör. Sayısı	Kullanıcı Başına Sayfa Gör.	Kullanıcı Başına Oturum	Oturum B. Sayfa Görün.
Ülke içi	8.706	11.197	76.069	8,74	1,29	6,79
Kuruluş içi	5.773	8.701	69.443	12,03	1,51	7,98
Ülke dışı	1.702	2.206	16.160	9,49	1,30	7,33
Toplam/Ort.	16.181	22.104	161.672	9,99	1,37	7,31

Veriler, portala kurum içerisinden erişen kullanıcıların oturum başına daha yüksek sayfa görüntülemelerine sahip olduğunu göstermektedir. Ülke içi kullanıcıların baskınlığı, sitenin yerel bağlamdaki önemini vurgulamakta ve daha fazla yerelleştirme ve kullanıcı katılımı için fırsatlar sunmaktadır. Ayrıca, kuruluş içi ve diğer kategoriler arasındaki farklılaşma, kuruluş içi operasyonlar ve kullanıcıların konuma bağlı olarak sistemle nasıl etkileşime girdiği hakkında değerli bilgiler sağlar. Bu bilgiler yöneticilere kaynak tahsisi, kullanıcı deneyimi geliştirme ve coğrafi eğilimlere ve belirli kurumsal ihtiyaçlara göre uyarlanmış stratejik planlama konularında bilinçli kararlar alma konusunda yol gösterir.

5.1.9. Etkileşim analizi

Web sistemiyle kullanıcı etkileşimlerinin incelenmesi, kullanıcı navigasyonunu ve davranışını etkileyen çeşitli unsurların ayrıntılı bir analizini gerektirir. Şekil 5.5'te kullanıcı etkileşimlerinin kapsamlı bir görünümü, yönlendiren türleri, en çok yönlendiren siteler, açılış alt alanları ve oturum sayısına göre çıkış türleri olmak üzere dört farklı açıdan detaylandırılmaktadır. Altı farklı yönlendirici türünün frekans dağılımı Grafik (a)'da gösterilmektedir. Beklendiği gibi, en yüksek yüzde kurumsal ana web sitesine aittir ve bunu yönlendirici olmadan doğrudan erişim izlemektedir. Grafik (b), siteye erişimin ağırlıklı olarak Google arama motoru üzerinden olduğunu göstermektedir. Bunu kurumsal web sitesi ve diğer çeşitli kurumsal alt alanlar takip etmektedir.

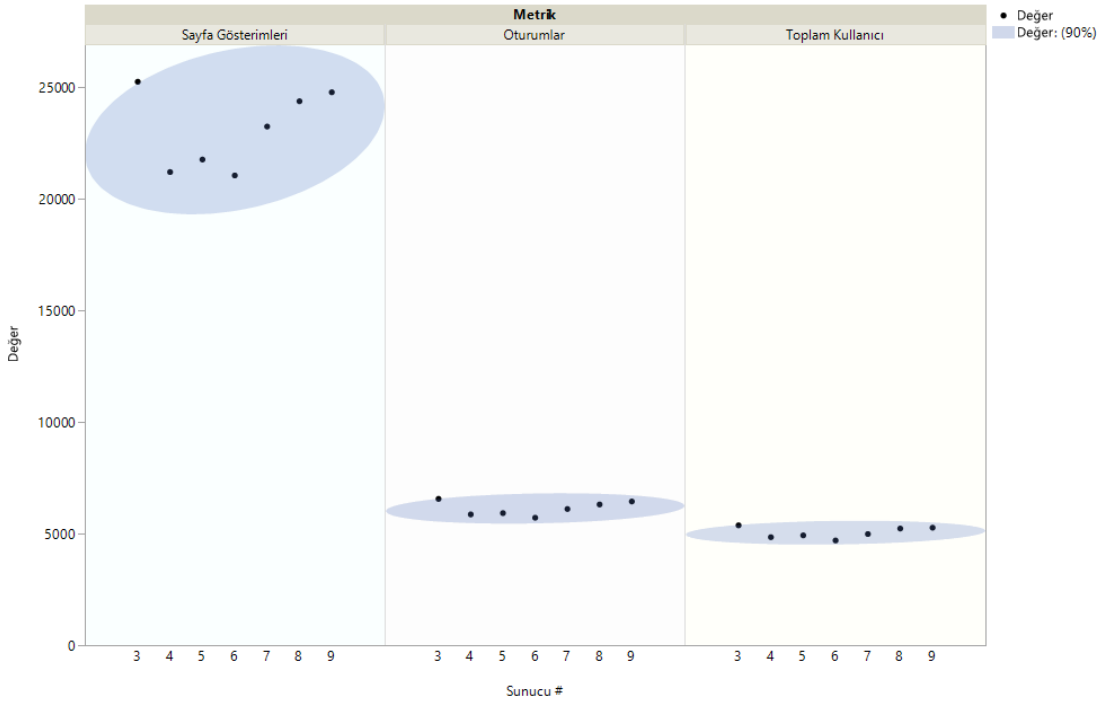


Şekil 5.5. Yönlendiren türleri, yönlendiren siteler, açılış alt alanları ve çıkış türleri.

Grafik (c), uygulamaya ilk gelişte kullanılan alt alan adlarını göstermektedir. CAWIS'in bir üniversite web portalı olduğu düşünüldüğünde, beklendiği gibi Öğrenci Bilgi Sistemi (OBİS) en yüksek orana sahiptir. Grafik (d), kullanıcıların üçte ikisine yakınının portaldan ayrılırken güvenli çıkış yapmayı tercih ettiğini, üçte birlik kısmının ise doğrudan pencereyi kapatarak ya da başka bir adrese yönlenerek siteden ayrıldığını göstermektedir. Bu bilgi, çıkış noktalarındaki kullanıcı davranışını anlamak ve çıkış sürecini optimize etmek için oldukça önemlidir.

5.1.10. Sunucu tabanlı analiz

CAWAL'ın web çiftliği ortamındaki yeteneklerini gösteren Şekil 5.6, sayfa görüntüleme, oturum ve tekil kullanıcı sayılarını sunucu bazlı ortaya koymaktadır. Web çiftliğinde yapılan analiz sırasında ilk iki sunucu geçici olarak hizmet dışı durumdadır. On numaralı sunucu ise test, dağıtım ve NAS görevlerini üstlenmiş, ancak doğrudan kullanıcılara web hizmeti sunmamıştır. İncelenen verilere göre portalın ortalama oturum sayısı yaklaşık 6.130 iken, bu oturumlar 5.045 farklı kullanıcı tarafından gerçekleştirilmiştir. Buna göre kullanıcı başına düşen ortalama oturum sayısı yaklaşık 1,22 olarak gerçekleşmiştir. Portala bağlılığı ve etkileşim sıklığını ifade eden bu önemli metrik, kullanıcıların ilgili periyotta portalı birden fazla kez ziyaret ettiklerini ve bu oranın %22 civarında olduğunu göstermektedir.



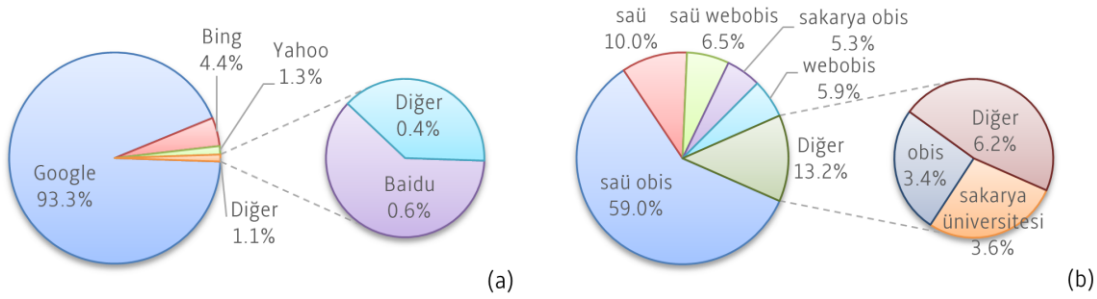
Şekil 5.6. Sayfa gösterimleri ve oturumların sunuculara göre dağılımı.

Grafikte ayrıca, oturumların sunuculara dağılımının yüzde 13,3 ile 15,3 arasında, sayfa görüntülemelerin sunuculara dağılımının yüzde 13,3 ile 15,2 arasında gerçekleştiği görülmektedir. Bu oranlar, istek yükünün sunuculara dengeli bir şekilde dağıldığını kanıtlamaktadır.

Oturum başına sayfa görüntüleme (PpS) oranları incelendiğinde ise sunucular arasındaki dağılımın birbirine oldukça yakın olduğu ve 3,62 ile 3,86 arasında değiştiği görülmektedir. İsteklerin sunuculara doğru bir şekilde dağıtıldığını ve sistemin tüm sunucularla dengeli bir hizmet verdiğini gösteren bu sonuçlar, CAWAL'ın web çiftliği mimarisine uyarlanabilirliğini göstermesi açısından son derece kıymetlidir.

5.1.11. Arama motoru verilerinin analizi

Ziyaretçilerin web sitesine erişirken kullandıkları arama motorlarının (a) ve ziyaretçilerin arama motorları aracılığıyla siteye erişmek için kullandıkları arama anahtar kelimelerinin dağılımı (b), Şekil 5.7'de verilmiştir. Site yöneticileri, arama motorlarından erişilen sponsorlu bağlantıları tespit etme ve kullanıcıların siteyi bulma yöntemini anlama çabaları nedeniyle bu verileri dikkate almalıdır. Hangi ifadeler için sponsor bağlantı ücreti ödemenin daha verimli olacağının değerlendirilmesi ya da site organizasyonunun bu veriler ışığında gözden geçirilmesi etkin site yapılanması için büyük önem taşımaktadır.



Şekil 5.7. Arama motorlarının ve arama anahtar kelimelerinin dağılımı.

5.1.12. Servis bazında çok boyutlu kullanım analizi

Kurumsal web portalındaki önemli bir hizmet olan web tabanlı e-posta servisi "mail", çok boyutlu verileri göstermek amacıyla seçilmiştir. Tablo 5.5'teki metrikler, portal servisinin kullanıcı türü ve cinsiyete göre ayrıntılı bir incelemesini sunmaktadır. Burada, her bir özellik için kullanıcı sayısı, oturum ve sayfa gösterimleri ayrı ayrı tespit edilerek farklı kullanım modelleri gözlemlenebilmektedir. Bu bilgiler, kullanıcı etkileşimi ile portal hizmetlerinin detaylı anlaşılmasını kolaylaştırır ve kullanıcı ihtiyaçlarına özel uyarlanmış deneyimleri geliştirmek için stratejik yaklaşımların oluşturulmasında kritik bir rol oynar.

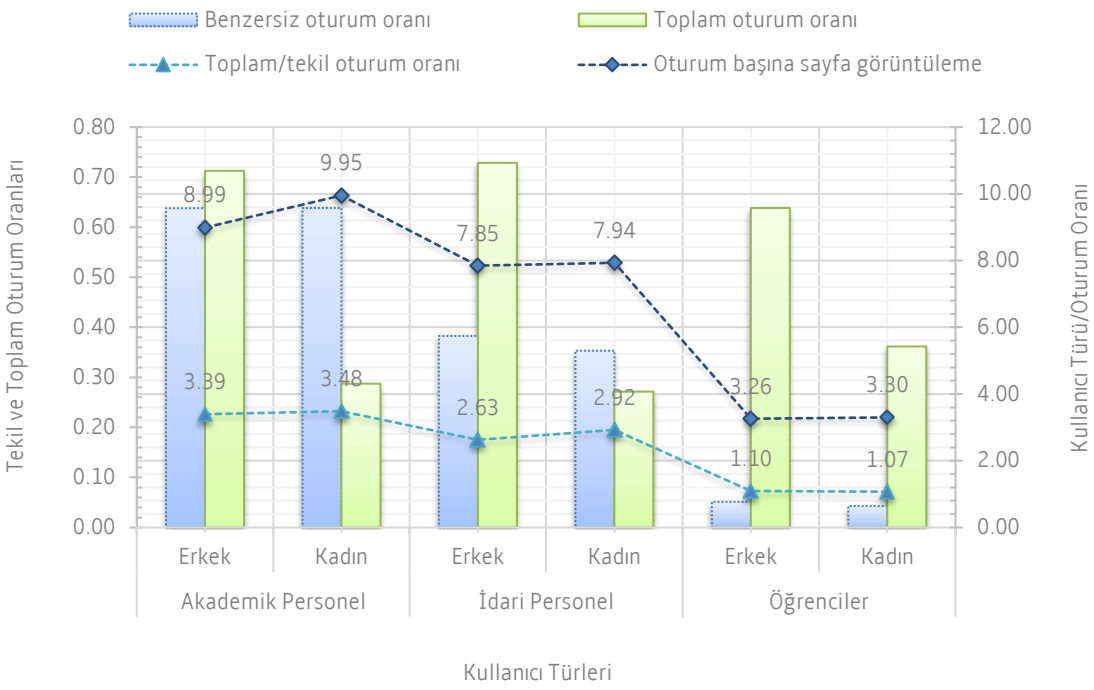
Tablo 5.5. Mail servisinin kullanıcı türü ve cinsiyete göre çok boyutlu verileri.

Cinsiyet	Metrik	Kullanıcı Tipi									
		0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**9**	**Toplam**
Belirsiz	Kullanıcı Sayısı	390	0	0	0	0	0	0	0	0	**390**
	Oturum Sayısı	390	0	0	0	0	0	0	0	67	**457**
	Sayfa Gös. Say.	432	0	0	0	0	0	0	0	575	**1.007**
Erkek	Kullanıcı Sayısı	0	652	194	0	0	8	1.620	39	0	**2.513**
	Oturum Sayısı	0	2.213	510	0	0	0	1.774	62	0	**4.559**
	Sayfa Gös. Say.	0	19.889	4.002	385	337	195	5.782	332	5	**30.927**
Kadın	Kullanıcı Sayısı	0	256	65	0	0	0	937	22	0	**1.280**
	Oturum Sayısı	0	892	190	0	0	0	1.005	0	0	**2.087**
	Sayfa Gös. Say.	0	8.876	1.508	0	57	0	3.320	151	27	**13.939**
Toplam	Kullanıcı Sayısı	390	908	259	0	0	8	2.557	61	0	**4.183**
	Oturum Sayısı	390	3.105	700	0	0	0	2.779	62	67	**7.103**
	Sayfa Gös. Say.	432	28.765	5.510	385	394	195	9.102	483	607	**45.873**

Analiz, farklı kullanıcı gruplarının "mail" servisiyle olan etkileşimlerinin net bir profilini çizmekte ve kullanıcı türü ile cinsiyetin kullanım üzerindeki etkisini ortaya koymaktadır. İncelenen dağılımlar, portal veritabanında kayıtlı 1.022 erkek akademik kullanıcıdan 652'sinin bu portal hizmetini kullandığını göstermektedir. Bu da erkek akademisyen kullanıcıların %64'ünün benzersiz oturum oranı olarak hizmetle etkileşime geçtiği sonucunu doğurmaktadır. Bu bilgiler, kullanıcı davranışının

derinlemesine anlaşılmasına katkıda bulunurken, hizmet tasarımı ve sunumu için geliştirilecek stratejilerin optimizasyonuna yardımcı olma potansiyeli taşımaktadır.

Şekil 5.8'de yer alan grafik ise aynı portal servisiyle en çok etkileşime giren üç kullanıcı türünü ve cinsiyetlere göre katılım özelliklerini karşılaştırmaktadır.

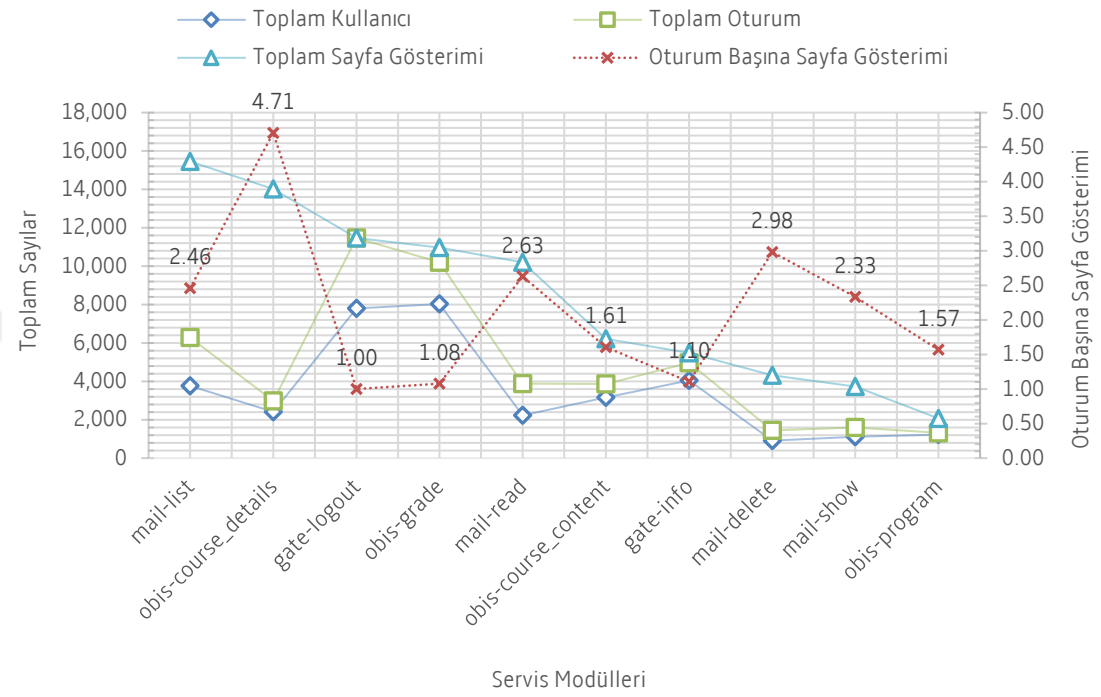


Şekil 5.8. Seçilen kullanıcı türlerine göre "mail" servisi kullanım verileri.

Farklı cinsiyetlerin hizmetlerle nasıl etkileşim kurduğunu anlamak, daha derinlemesine ve gerçekçi bir perspektif sunar. Bu analize göre, akademik personel arasında erkeklerin tekil oturum açma oranları kadınlara göre daha yüksek olmasına karşın, kadın akademik personelin oturum başına gösterdikleri sayfa sayısı daha fazladır. İdari personelde ise, kadınlar her iki kategoride de daha yüksek oranlar sergileyerek portal kullanımında daha aktif olduklarını belirtmektedir. Öğrencilerde cinsiyet temelli farklılıklar daha az belirgindir, ama yine de erkeklerin tekil oturum açma oranları kadınlardan bir miktar fazladır. Bu tür nicel veriler, kullanıcı demografilerinin daha iyi anlaşılmasına yardımcı olur ve portal hizmetlerinin çeşitli kullanıcı grupları ve cinsiyetler arasında daha etkili bir şekilde erişilmesini ve kullanılmasını sağlamak için yapılacak analizlere ve strateji geliştirmelere zemin hazırlar.

5.1.13. Modül bazında çok boyutlu kullanım analizi

Web portalında "modül" terimi, bir servisin içerisindeki ayrı ve özelleşmiş alt işlevleri veya bölümleri tanımlar. CAWIS portalında en çok kullanılan on modülün kullanıcı, oturum ve sayfa gösterimi sayıları ile bu modüllerin oturum başına sayfa gösterimi oranları Şekil 5.9'da verilmiştir.



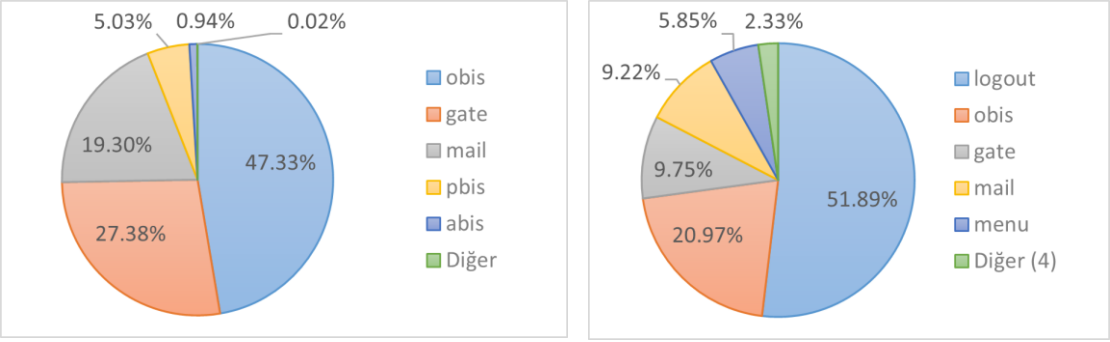
Şekil 5.9. En çok sayfa görüntülemesine sahip ilk on modülün kullanım değerleri.

Portalının çeşitli modüllerindeki kullanıcı sayıları, oturum ve sayfa gösterimleri arasındaki ilişki, kullanıcıların portal içindeki etkileşim düzeylerini ve modüllerin popülerliğini açığa çıkarır. Örneğin, "mail-list" ve "gate-info" modülleri en yüksek kullanıcı ve oturum sayılarına sahipken, "mail-delete" ve "obis-program" modülleri bunlara göre çok daha az etkileşim göstermektedir. Oturum başına yapılan sayfa gösterimi sayıları ise, kullanıcıların hangi modüllerde daha fazla içeriğe eriştiğini ve portalın hangi bölümlerinin yoğun kullanıldığını belirtmektedir. Bu verilerin analizi, kullanıcı davranışını anlamada ve portalın kullanıcı deneyimini iyileştirme çabalarında kritik rol oynar.

5.1.14. Servis bazında kullanıcıların portala erişim ve ayrılış analizi

Kullanıcıların web portalına hangi servis üzerinden ilk erişimi sağladıkları ve ayrılırken en son hangi servis üzerinden portalı terk ettikleri Şekil 5.10'da gösterilmiştir. Grafik (a)'da yer alan veriler, kullanıcıların çoğunlukla "obis", "gate"

ve “mail” servisleri üzerinden portala adım attıklarını ortaya koymaktadır. Grafik (b)’de ise kullanıcıların portaldan en çok “logout” (güvenli çıkış) yaparak ya da “obis”, “gate” ve “mail” servisleri üzerinden ayrıldıkları görülmektedir.



(a) (b)

Şekil 5.10. Sıklığına göre en çok geliş ve ayrılış yapılan portal hizmetleri.

Gerçekte “logout” bir servis olmayıp, “gate” servisinin güvenli çıkış sayfasıdır ve sistemden ayrılmak için en fazla tercih edilen yöntemdir. Bununla birlikte, “logout” oranının yüksekliği, kullanıcıların güvenlik bilincine sahip olduğunu ve oturumlarını düzgün bir şekilde kapattıklarını göstermektedir. Portal yöneticileri ve tasarımcıları için, bu hizmetlerin kullanılabilirliğini ve performansını iyileştirmeye yönelik stratejiler geliştirmek, sistem güvenliği ve kullanıcı memnuniyetini artırma potansiyeli taşır.

5.1.15. Uygulama temelli hata, uyarı ve bilgilendirme analizleri

CAWAL çerçevesinin uygulamaya özel verileri toplama ve analiz etme yeteneklerinin bir yansıması olarak portalda yer alan servis ve modüllerin kullanımı sırasında ortaya çıkan hatalar, uyarılar ve bilgilendirme mesajları üzerinde kapsamlı bir analiz gerçekleştirilmiştir. Burada ele alınan analizler, daha anlamlı çıkarımlar için 24 saatlik veriler yerine sistemin 7 günlük operasyonel verileri üzerinden yürütülmüş olup, bu süre zarfında kullanıcılar tarafından deneyimlenen yaygın sorunları ve karşılaşılan bildirimleri derinlemesine incelemeyi amaçlamaktadır.

Bu süreçte, servis bazlı hata analizi, modül bazlı uyarı ve bilgilendirme analizi ile belirli bir modüle yönelik hedef odaklı uyarı analizi olmak üzere üç ana kategori altında değerlendirmeler yapılmıştır. Bu analizlerin amacı, kullanıcı deneyimini iyileştirmek, sistem hatalarını azaltmak ve kullanıcıların ihtiyaçlarına daha iyi yanıt verebilmek için gerekli düzenlemeleri yapmak üzere değerli içgörüler elde etmektir.

5.1.15.1. Servis bazlı hata analizi

Web portalından CAWAL sayesinde toplanan uygulamaya özel hata ve uyarı mesajı verilerinin incelenmesine dayanan bu analiz, kullanıcıların yaygın olarak karşılaştıkları hataların belirlenmesine odaklanmaktadır. Elde edilen veriler, uygulamanın belirli bölümlerindeki kullanıcı deneyiminin iyileştirilmesi için kritik öneme sahiptir. Tablo 5.6, bu analizlerin sonucunda elde edilen örnek bir veri kümesini sunmaktadır. Sistemdeki hata ve uyarı verileri, bu tür durumların en sık "gate" ve "mail" servislerinde yaşandığını frekans değerleriyle birlikte ortaya koymaktadır.

Tablo 5.6. Portala özel hata ve uyarıların servis bazlı analizi.

Hata Açıklama	Servis	Frekans
Hatalı kullanıcı adı ya da şifre	gate	11.476
Zaman aşımı uyarısına yanıt vermediğiniz için oturumunuz kapanmıştır	gate	3.519
Hatalı kullanıcı adı ya da şifre. CAPS LOCK tuşunu açık unutmadığınıza emin olunuz!	gate	2.372
Hatalı kullanıcı adı ya da şifre. Kullanıcı adı bölümüne öğrenci numaranızı yazmayınız! SSS'i inceleyiniz.	gate	680
Zaman aşımı uyarısına devam onayı vermediğiniz için oturumunuz kapanmıştır	gate	116
E-posta gönderme hatası: sorry, no mailbox here by that name (#5.1.1 - chkuser) RCPT first (#5.5.1)	mail	61
Hatalı şifre girdiniz, lütfen tekrar deneyiniz	gate	34
E-posta gönderme hatası: sorry, can't find a valid MX for rcpt domain (#5.1.1 - chkuser) RCPT first (#5.5.1)	mail	28
E-posta gönderme hatası: DNS temporary failure (#4.5.1 - chkuser) RCPT first (#5.5.1)	mail	17
E-posta gönderme hatası: sorry, you must specify a domain (#5.1.1 - chkuser) RCPT first (#5.5.1)	mail	10
E-posta gönderme hatası: MAIL first (#5.5.1) MAIL first (#5.5.1)	mail	7
E-posta gönderme hatası: sorry, recipient address has invalid format (#5.1.1 - chkuser) RCPT first (#5.5.1)	gate	6
Hatalı şifre girdiniz, lütfen tekrar deneyiniz. CAPS LOCK tuşunu açık unutmadığınıza emin olunuz!	mail	6
Zaman aşımı uyarısını en fazla 3 kez devam ettirebilirsiniz	gate	5
E-posta gönderme hatası: sorry, you must specify a domain (#5.1.1 - chkuser)	mail	3
Çok fazla hatalı şifre girdiğiniz için sistemden (1) gün uzaklaştırıldınız. 'Şifremi Unuttum!' bölümünü kullanın.	gate	2
E-posta gönderme hatası: sorry, no mailbox here by that name (#5.1.1 - chkuser)	mail	2
Bu serviste hata oluştu, lütfen bir süre sonra tekrar deneyiniz	abis	2

Sıklığa göre sıralanan veriler incelendiğinde, hataların büyük bir bölümünün kullanıcı kimlik doğrulama ve e-posta gönderim süreçlerinde meydana geldiği görülmektedir. Bu durum, belirlenen noktalarda yürütülen süreçlerin iyileştirilmesine yönelik çalışmalar yapılması gerektiğine işaret etmektedir. Ayrıca, bu hata ve uyarıların zaman içindeki değişimleri analiz edilerek sistemdeki gelişmelerin etkin bir şekilde takip edilmesi ve böylece sürekli iyileştirme süreci kapsamında kullanıcı memnuniyetinin ve sistem performansının artırılması mümkün olacaktır.

5.1.15.2. Modül bazlı uyarı ve bilgilendirme analizi

Portalın kullanımı sırasında en yaygın karşılaşılan uyarı ve bilgilendirmelerin sıklığına göre servis ve modül bazlı bir özetini sunan analiz Tablo 5.7'de sunulmuştur. Bu tabloda, kullanıcıların deneyimlerini ve sistemdeki etkileşimlerini doğrudan etkileyen mesajlar detaylandırılmıştır.

Tablo 5.7. Portala özel uyarıların servis ve modül bazlı analizi.

Uyarı Açıklama	Servis	Modül	Frekans
Personel kaydınız bulunamadığı için bu servisi kullanamazsınız. Personel Dairesi Başkanlığı'na Başvurunuz.	pbis	pbis	22.284
Posta başarıyla silindi	mail	sil	12.254
Hatalı kullanıcı adı ya da şifre	gate	gate	11.465
Postalar başarıyla silindi	mail	list	6.774
Son 1 yıl içerisinde iletişim bilgilerinizi değiştirmediğiniz için güncelleme yapmanız gerekmektedir.	obis	obis	3.659
Zaman aşımı uyarısına yanıt vermediğiniz için oturumunuz kapanmıştır	gate	logout	3.519
Buraya gireceğiniz veriler Öğrenci İşleri tarafından kontrol edilerek sisteme aktarılacaktır	obis	bilgiguncelle	3.317
İletişim bilgileriniz kontrollü olarak aktarılmak üzere Öğrenci İşlerine gönderilmiştir	obis	bilgiguncelle	2.838
Tarayıcınız desteklemediği için RichText yazım arayüzü aktif değildir	mail	yeni	2.796
Posta başarıyla silindi	mail	list	2.753
Hatalı kullanıcı adı ya da şifre. CAPS LOCK tuşunu açık unutmadığınıza emin olunuz!	gate	gate	2.372
Tarayıcınız desteklemediği için RichText yazım arayüzü aktif değildir	mail	yanitla	1.902
İletişim bilgilerinizi güncellemeniz gerekmektedir. Güncellemeden sonra diğer bölümleri kullanabilirsiniz	obis	obis	1.780
Postalar başarıyla silindi	mail	sil	1.681
Tercihleriniz kaydedilmiştir	gate	pref	1.591
Yayınlanmış ders programı bulunamadı	obis	program	1.457
Girdiğiniz bilgiler doğrulanamadı, lütfen tekrar deneyiniz	gate	forget	1.102
Yeni notlar başarıyla kaydedildi	abis	notgirisi2	794
Hatalı kullanıcı adı ya da şifre. Kullanıcı adı bölümüne öğrenci numaranızı yazmayınız! SSS'i inceleyiniz.	gate	gate	671
Postalar başarıyla silindi	mail	read	638
Arama seçeneklerinize göre kayıt bulunamamıştır	rehber	arama	629
Kullanıcı şifrenizi belirli aralıklarla değiştirmeniz gerektiğini unutmayınız	gate	pwd	516
Tanımladığınız çalışma başarıyla eklendi	abis	katkipaylari	504
Posta hesabınızın kotası dolmak üzere; sakladığınız gereksiz mesajları en kısa sürede silmeniz önerilir	mail	mail	501

Örneğin, “pbis” modülünde en sık karşılaşılan uyarı, personel kaydının bulunmamasıyla ilgili olup, bu durum yetkisiz kullanıcıların sisteme eriştiklerine ve yönlendirme konularında iyileştirmeler yapılması gerektiğine işaret etmektedir. Diğer yandan, “mail” servisinde sıkça rastlanan "Posta başarıyla silindi" bildirimi, kullanıcıların bu özelliği aktif bir şekilde kullandığını ve web tabanlı “mail” servisinin kullanıcı dostu olduğunu göstermektedir. Bu tür analizler, kullanıcı deneyiminin geliştirilmesine ve sistem hatalarının azaltılmasına katkı sağlayacak değerli içgörüler sunmaktadır.

Bu analizden elde edilen bulgular, sistemin kullanıcı arayüzü ve işlevselliğinin nasıl algılandığını ve hangi alanlarda iyileştirmeler yapılması gerektiğini ortaya koymaktadır. Örneğin, “obis” modülündeki güncelleme hatırlatmaları, kullanıcıların kişisel bilgilerini düzenli olarak güncellemeleri gerektiğini vurgulamakta ve bu sürecin kullanıcılar için daha basit ve anlaşılır hale getirilmesi gerektiğine işaret etmektedir.

Diğer yandan, “gate” servisinde sıkça karşılaşılan hatalı giriş uyarıları, kimlik doğrulama süreçlerinin daha etkin hale getirilmesi ihtiyacını göstermektedir. Bu analiz sonuçları, kullanıcıların sistemle etkileşimini iyileştirmek ve daha verimli bir kullanıcı deneyimi sağlamak için faydalı bir temel oluşturmaktadır. Her bir uyarı ve bilgilendirmenin detaylı incelenmesi, web uygulamasının kullanıcı dostu ve etkili bir şekilde işlemesi için gereken stratejik değişikliklerin yapılmasına yardımcı olacaktır.

5.1.15.3. Modül bazlı hedef odaklı uyarı analizi

Portalın giriş kapısı niteliğindeki “gate” servisinin “şifre işlemleri” modülünde karşılaşılan uyarı mesajlarına odaklanılarak detaylı bir analiz gerçekleştirilmiştir. Tablo 5.8, bu modülde en çok karşılaşılan uyarı ve bilgilendirme mesajlarının sıklığına göre bir listesini sunmaktadır. Şifre işlemleri, kullanıcıların sistemdeki güvenliğinin temelini oluşturduğundan dolayı bu modülün analizi, kullanıcıların şifre oluşturma ve güncelleme süreçlerinde karşılaştıkları zorlukları ve problemleri anlamak adına büyük önem taşımaktadır. Analiz ayrıca, kullanıcıların şifre politikalarına ne kadar uyduklarını, sıkça yapılan hataları ve şifre değişikliği sürecindeki kullanıcı davranışlarını detaylı bir şekilde ortaya koymaktadır.

Tablo 5.8. Portal şifre işlemleri modülünün uyarı mesajları analizi.

Uyarı Açıklama	Servis	Modül	Frekans
Kullanıcı şifrenizi belirli aralıklarla değiştirmeniz gerektiğini unutmayınız	gate	pwd	516
Yeni şifrede harf ve rakam dışında karakter bulunmamalıdır	gate	pwd	408
Şifre hatırlatma sistemi yeni şifrenizi kaydetti	gate	pwd	347
Yeni şifrede en az 2 harf bulunmak zorundadır	gate	pwd	334
Yeni şifrenizde adınızı anımsatan heceler bulunmamalıdır	gate	pwd	290
Şifreniz başarıyla değiştirildi	gate	pwd	176
Yeni şifrenizde doğum yılınızı anımsatan sayılar bulunmamalıdır	gate	pwd	170
Yeni şifrede en az 2 rakam bulunmak zorundadır	gate	pwd	158
Yeni şifre, eski şifre ile aynı olmamalıdır	gate	pwd	142
Yeni şifrenizde isminizi anımsatan heceler bulunmamalıdır	gate	pwd	130
Yeni şifrenin uzunluğu 6-15 karakter arası olmalıdır	gate	pwd	94
Yeni şifrenizde doğum tarihinizi anımsatan sayılar bulunmamalıdır	gate	pwd	82
Yeni şifre alanları birbirine eşit olmalıdır	gate	pwd	80
Geçerli şifrenizi yanlış girdiniz, lütfen tekrar deneyiniz	gate	pwd	72
Aktivasyon işlemi başarıyla tamamlandı	gate	pwd	58
Yeni şifrenizde soyadınızı anımsatan heceler bulunmamalıdır	gate	pwd	42
Şu an sistemde olduğunuz için 'Şifremi Unuttum!' bölümüne giremezsiniz	gate	pwd	3

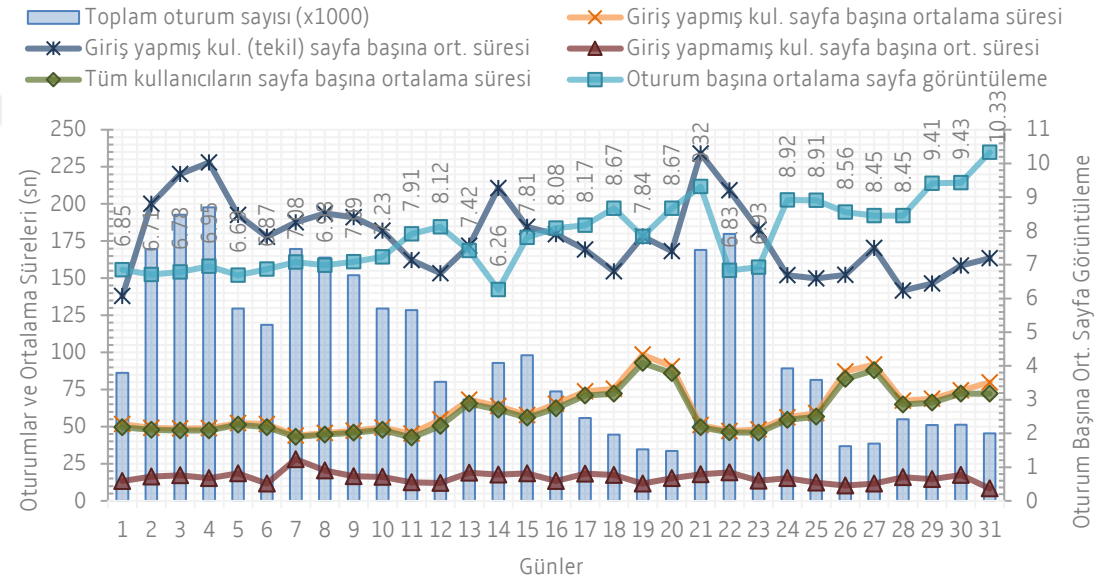
Analiz sonuçları, kullanıcıların şifre oluşturma sürecinde en sık karşılaştıkları zorlukları ve yapılan hataları gözler önüne sermektedir. Örneğin, şifre politikalarına uygun şifre oluşturmakta yaşanan güçlükler, kullanıcıların bu süreçte ne kadar dikkatli olmaları gerektiğini ve sistem tarafından belirlenen şifre kurallarının ne derece anlaşılabilir olduğunu göstermektedir. Bu bulgular, şifre politikalarının daha net ve kullanıcı dostu hale getirilmesi için gerekli düzenlemelerin yapılmasına yardımcı olabilir. Bulgular ayrıca, şifre değiştirme sürecinde kullanıcıların karşılaştığı zorlukları azaltmak ve bu işlemin daha sorunsuz bir şekilde gerçekleştirilmesini sağlamak adına yapılacak iyileştirmeler için de yol gösterici niteliktedir. Bu analiz, kullanıcıların şifre yönetimi süreçlerinde daha sağlıklı ve güvenli bir deneyim yaşamalarını sağlamak amacıyla önemli bir adım teşkil etmektedir.

5.2. Orta Vadeli Analitikler

Orta vadeli analitikler, bir aylık veriler üzerinden yapılan analizleri içerir ve günlük bazda değişen durumları incelenmeyi hedefler. Bu analizler, operasyonel optimizasyonlar ve kısa dönemli eğilimlerin yanı sıra, performans değerlendirmeleri ve taktiksel karar verme süreçleri için de kritik bilgiler sağlar. 2019 yılı Ocak ayına ait veriler üzerinden gerçekleştirilen bu analizler, hem gün bazında bir aylık analitik bilgileri hem de analitik çıkarım işleminin kendisini incelemektedir.

5.2.1. Gün bazında bir aylık kullanım analizi

Kullanıcı türlerine göre günlük sayfa görüntüleme sayıları ve oturum başına ortalama kullanım süreleri Şekil 5.11'de sunulmuştur. Bir aylık analitik verilerine dayanarak oluşturulan grafikte, özellikle hafta sonlarındaki oturum sayısında belirgin düşüşler göze çarpmakta, bu da portalın hafta içi eğitim günlerinde daha yoğun kullanıldığını işaret etmektedir. Kayıtlı kullanıcıların sayfa başına geçirdikleri süre genellikle kayıtsız kullanıcılardan daha fazla olup, bu durum kayıtlı kullanıcıların daha zengin bir içerikle karşılaştıkları ve portalda daha fazla etkileşimde bulundukları anlamına gelmektedir.



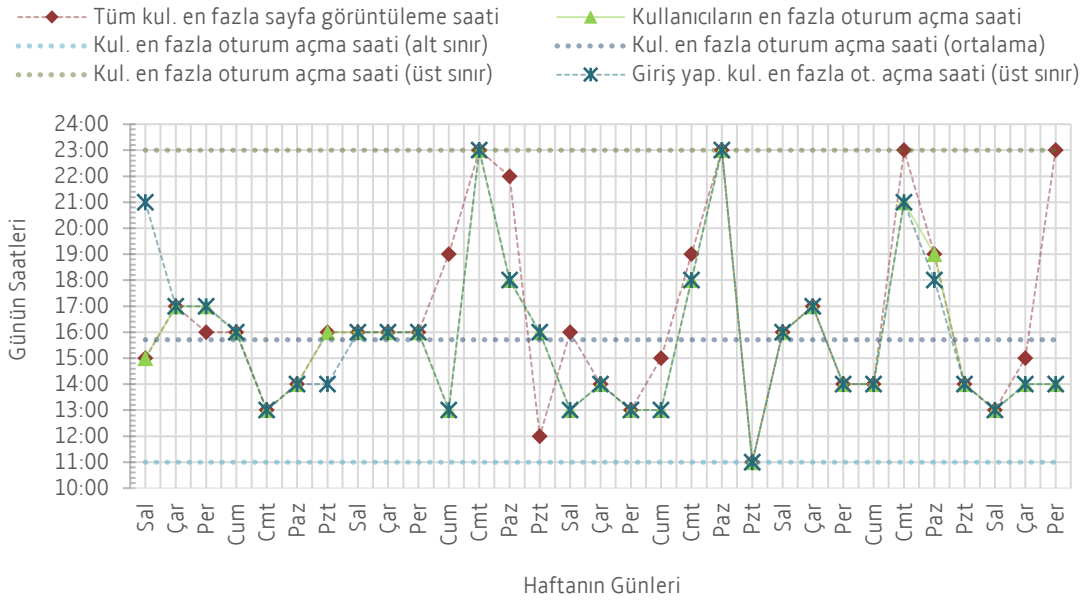
Şekil 5.11. Kullanıcı türüne göre sayfa görüntüleme ve süre bilgileri.

Bu veriler, web portalının kullanıcı etkileşimleri üzerine etkili iyileştirmeler yapmak için yöneticilere ve tasarımcılara önemli bilgiler sağlamaktadır. Analiz sonuçları, kullanıcı deneyiminin iyileştirilmesi ve site içi etkileşimin kalitesinin artırılması için stratejik kararlar alınmasına yardımcı olabilir.

5.2.2. Gün bazında bir aylık sistem yoğunluk zamanları analizi

Bir ay boyunca farklı günlerde kullanıcıların web portalında ne zaman en aktif olduklarını detaylandıran bir etkinlik haritası Şekil 5.12'de verilmiştir. Oturum sayılarının günün saatlerine göre dağılımını gösteren bu grafik, kullanıcı etkinliklerinin zirve yaptığı ve minimum etkinlik gösterilen zamanları net bir biçimde ortaya koyar. En yüksek kullanıcı etkinliğinin akşam saatlerinde yoğunlaştığı gözlemlenirken, sabah saatlerinin ise daha sakin ve düşük etkinlik seviyelerine sahip

olduğu açıkça görülmektedir. Bu analiz, portalın kullanım yoğunluğunun ve kullanıcıların katılım tercihlerinin anlaşılması açısından değerlidir.



Şekil 5.12. Aylık oturum zamanlarının ve sayfa görüntülemelerin günlük dağılımı.

Analiz sonuçları, web portalı kullanımının, kullanıcıların okul ve iş sonrası serbest zamanları ile uyumlu olduğunu göstermektedir. Hafta içi ve hafta sonu etkinlikleri arasındaki farklılıklar, kullanıcıların hafta sonları daha az aktif olduklarını, bu durumun ise portal yöneticileri tarafından içerik ve etkileşim stratejilerini planlarken göz önünde bulundurulması gerektiğini işaret etmektedir. Ayrıca, zirve kullanım saatlerinin belirlenmesi, potansiyel bakım çalışmalarının planlanması ve kullanıcı deneyimini iyileştirecek özelliklerin optimizasyonu için de kritik öneme sahiptir. Bu bilgiler ışığında, web portalının kullanıcı ihtiyaçlarına daha iyi hizmet verecek şekilde nasıl uyarlanabileceğine dair stratejik kararlar alınabilir.

5.2.3. Gün bazında bir aylık ayrıntılı kullanım analizi

Web portalındaki kullanıcı etkileşimlerini derinlemesine incelemek amacıyla bir aylık ayrıntılı kullanım analizi Tablo 5.9'da sunulmuştur. Tablonun sütun başlıkları uzun olduğu için, anlaşılırlığı artırmak ve alana sığdırmak amacıyla kısaltmalar kullanılmıştır (GY.: Giriş Yapmış, GYm.: Giriş Yapmamış, Kul. ve K.: Kullanıcıların, TK.: Tekil Kullanıcıların, Otr.: Oturum, SG.: Sayfa Görüntüleme, Say.: Sayısı, Orn.: Oranı, Ort.: Ortalama, Bşn.: Başına). Sunulan metrikler, portalın trafik yoğunluğunun gün bazlı dağılımını ve kullanıcıların oturum açma alışkanlıklarını tanımlamada hayati öneme sahiptir. Oturum açan ve açmayan kullanıcıların etkinliklerinin analitik

karşılaştırması, portalın farklı kullanıcı segmentlerine sunduğu performansın anlaşılmasına imkân tanır. Bu derinlemesine inceleme, yönetim ve analiz ekiplerine kullanıcı deneyimini geliştirme ve portal kullanılabilirliğini optimize etme konusunda stratejik kararlar alma sürecinde yol gösterici bilgiler sunar.

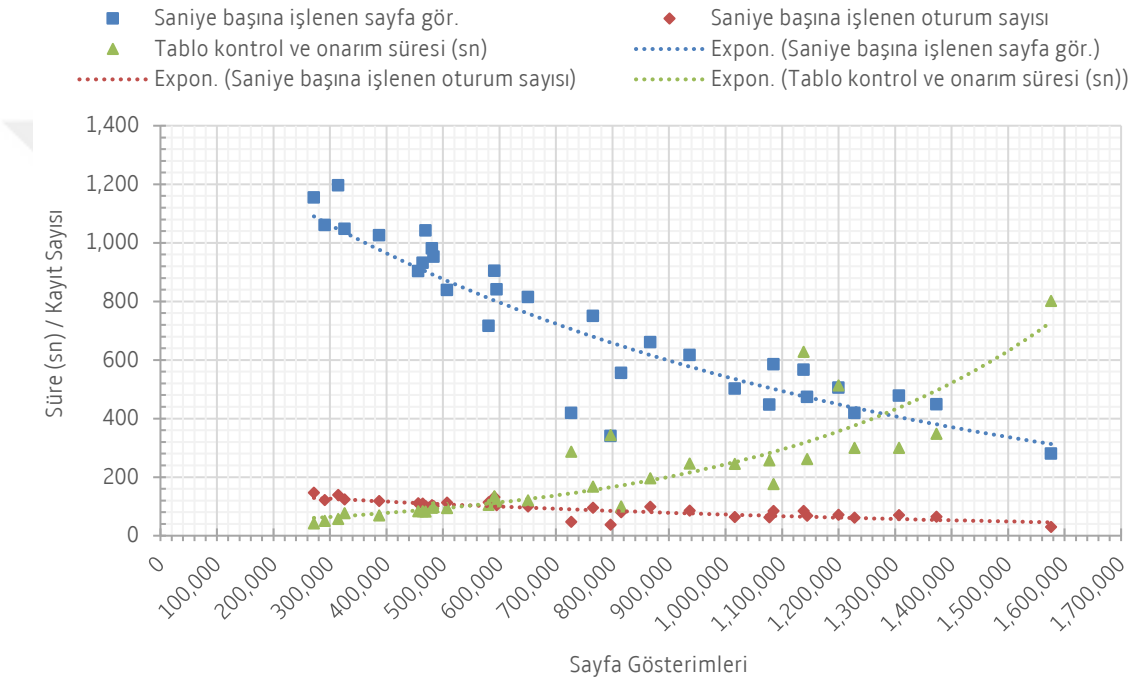
Sayfa görüntüleme ve oturum açma oranları arasındaki korelasyon, kullanıcıların portalda geçirdikleri süre ve ilgi gösterdikleri içerikleri açığa çıkarmaktadır. Oturum açan ve açmayan kullanıcılar arasındaki davranışsal farklılıklar, hedef kitlenin ihtiyaçlarına daha uygun hizmetler sunulmasında rehberlik edebilir. Ortalama sayfa görüntüleme ve oturum sayıları, kullanıcıların navigasyon kalıpları hakkında bilgi sunarken, sadık kullanıcıların etkinlik düzeyleri de bu analiz ile ölçülmektedir. Bu tür bir analitik inceleme, portal performansını yükseltme ve kullanıcı memnuniyetini maksimize etme yolunda atılacak adımlar için kıymetli bir temel oluşturmaktadır.

Tablo 5.9. Bir aylık verilerin analitik çıkarımından seçilen kritik bilgiler.

Tarih	Gün	Top. SG.	Top. Otr.	GY. Kul. SG. Say.	GY. Kul. SG. Orn.	GYm. Kul. SG. Say.	GYm. Kul. SG. Orn.	GY. Kul. Otr. Say.	GY. Kul. Otr. Orn.	GY. TK. Otr. Say.	GY. TK. Ort. Otr. Say.	GYm. Kul. Otr. Say.	GYm. Kul. Otr. Orn.	GY. Kul. Ort. SG. Say.	GY. TK. Ort. SG. Say.	GYm. Kul. Ort. SG. Say.	Otr. Bşn. Ort. SG. Say.
1.01.2019	Sal	590.812	86.281	399.668	0,68	191.144	0,32	76.837	0,89	28.836	2,66	9.444	0,11	5,20	13,86	20,24	6,85
2.01.2019	Çar	1.138.356	169.591	800.391	0,70	337.965	0,30	150.257	0,89	36.897	4,07	19.334	0,11	5,33	21,69	17,48	6,71
3.01.2019	Per	1.306.940	192.770	923.980	0,71	382.960	0,29	170.461	0,88	37.664	4,53	22.309	0,12	5,42	24,53	17,17	6,78
4.01.2019	Cum	1.373.132	197.609	976.727	0,71	396.405	0,29	175.454	0,89	37.653	4,66	22.155	0,11	5,57	25,94	17,89	6,95
5.01.2019	Cmt	866.902	129.551	603.007	0,70	263.895	0,30	116.165	0,90	31.641	3,67	13.386	0,10	5,19	19,06	19,71	6,69
6.01.2019	Paz	815.540	118.679	557.963	0,68	257.577	0,32	106.726	0,90	31.074	3,43	11.953	0,10	5,23	17,96	21,55	6,87
7.01.2019	Pzt	1.200.108	169.582	850.507	0,71	349.601	0,29	149.794	0,88	35.037	4,28	19.788	0,12	5,68	24,27	17,67	7,08
8.01.2019	Sal	1.144.630	163.939	814.842	0,71	329.788	0,29	144.456	0,88	34.110	4,24	19.483	0,12	5,64	23,89	16,93	6,98
9.01.2019	Çar	1.077.565	151.898	767.326	0,71	310.239	0,29	134.745	0,89	33.280	4,05	17.153	0,11	5,69	23,06	18,09	7,09
10.01.2019	Per	936.349	129.509	668.705	0,71	267.644	0,29	114.924	0,89	31.153	3,69	14.585	0,11	5,82	21,47	18,35	7,23
11.01.2019	Cum	1.016.275	128.545	652.598	0,64	363.677	0,36	102.576	0,80	28.639	3,58	25.969	0,20	6,36	22,79	14,00	7,91
12.01.2019	Cmt	650.756	80.147	425.780	0,65	224.976	0,35	68.336	0,85	24.382	2,80	11.811	0,15	6,23	17,46	19,05	8,12
13.01.2019	Paz	507.444	68.357	355.761	0,70	151.683	0,30	56.716	0,83	22.526	2,52	11.641	0,17	6,27	15,79	13,03	7,42
14.01.2019	Pzt	580.858	92.823	388.388	0,67	192.470	0,33	80.590	0,87	24.461	3,29	12.233	0,13	4,82	15,88	15,73	6,26
15.01.2019	Sal	765.524	98.066	568.442	0,74	197.082	0,26	85.585	0,87	26.786	3,20	12.481	0,13	6,64	21,22	15,79	7,81
16.01.2019	Çar	595.172	73.693	440.044	0,74	155.128	0,26	62.143	0,84	22.565	2,75	11.550	0,16	7,08	19,50	13,43	8,08
17.01.2019	Per	456.209	55.839	335.304	0,73	120.905	0,27	46.007	0,82	20.058	2,29	9.832	0,18	7,29	16,72	12,30	8,17
18.01.2019	Cum	386.918	44.613	279.819	0,72	107.099	0,28	36.692	0,82	17.929	2,05	7.921	0,18	7,63	15,61	13,52	8,67
19.01.2019	Cmt	271.529	34.643	201.120	0,74	70.409	0,26	25.773	0,74	14.244	1,81	8.870	0,26	7,80	14,12	7,94	7,84
20.01.2019	Paz	290.577	33.519	213.960	0,74	76.617	0,26	27.131	0,81	14.618	1,86	6.388	0,19	7,89	14,64	11,99	8,67
21.01.2019	Pzt	1.575.898	169.082	1.199.220	0,76	376.678	0,24	135.610	0,80	29.615	4,58	33.472	0,20	8,84	40,49	11,25	9,32
22.01.2019	Sal	1.228.485	179.796	870.697	0,71	357.788	0,29	161.539	0,90	36.339	4,45	18.257	0,10	5,39	23,96	19,60	6,83
23.01.2019	Çar	1.084.954	156.659	748.225	0,69	336.729	0,31	138.804	0,89	36.502	3,80	17.855	0,11	5,39	20,50	18,86	6,93
24.01.2019	Per	796.771	89.307	616.879	0,77	179.892	0,23	73.094	0,82	27.088	2,70	16.213	0,18	8,44	22,77	11,10	8,92
25.01.2019	Cum	727.003	81.585	555.609	0,76	171.394	0,24	67.043	0,82	26.430	2,54	14.542	0,18	8,29	21,02	11,79	8,91
26.01.2019	Cmt	314.699	36.770	244.766	0,78	69.933	0,22	24.605	0,67	14.077	1,75	12.165	0,33	9,95	17,39	5,75	8,56
27.01.2019	Paz	325.970	38.599	246.428	0,76	79.542	0,24	29.859	0,77	16.126	1,85	8.740	0,23	8,25	15,28	9,10	8,45
28.01.2019	Pzt	464.008	54.925	348.465	0,75	115.543	0,25	44.970	0,82	21.447	2,10	9.955	0,18	7,75	16,25	11,61	8,45
29.01.2019	Sal	480.827	51.097	373.966	0,78	106.861	0,22	41.624	0,81	19.534	2,13	9.473	0,19	8,98	19,14	11,28	9,41
30.01.2019	Çar	483.167	51.256	379.451	0,79	103.716	0,21	43.549	0,85	20.463	2,13	7.707	0,15	8,71	18,54	13,46	9,43
31.01.2019	Per	469.335	45.434	333.734	0,71	135.601	0,29	34.634	0,76	16.925	2,05	10.800	0,24	9,64	19,72	12,56	10,33
Toplam		**23.922.713**	**3.174.164**	**17.141.772**		**6.780.941**		**2.726.699**	**0,86**	**818.099**		**447.465**					
Ortalama		**771.700**	**102.392**	**552.960**	**0,72**	**218.740**	**0,28**	**87.958**	**0,84**	**26.390**	**3,08**	**14.434**	**0,16**	**6,85**	**20,15**	**14,78**	**7,86**
Minimum		271.529	33.519	201.120	0,64	69.933	0,21	24.605	0,67	14.077	1,75	6.388	0,10	4,82	13,86	5,75	6,26
Maksimum		1.575.898	197.609	1.199.220	0,79	396.405	0,36	175.454	0,90	37.664	4,66	33.472	0,33	9,95	40,49	21,55	10,33
Mak-Min		1.304.369	164.090	998.100	0,14	326.472	0,14	150.849	0,23	23.587	2,91	27.084	0,23	5,13	26,63	15,80	4,07

5.2.4. Bir aylık veriler üzerinden analitik çıkarım sürecinin analizi

Büyük veri setleri üzerinde yapılan işlemlerin performansını değerlendirmek ve süreç optimizasyonu için gerekli analizleri sağlamak üzere tasarlanan Şekil 5.13, sayfa görüntüleme sayılarına bağlı olarak analitik çıkarım sürecinin farklı aşamalarındaki işlem sürelerini göstermektedir. Elde edilen verilerin işlenme süresi, sayfa görüntüleme ve oturum bilgilerinin hesaplanıp transfer edilme hızlarıyla doğrudan ilişkilendirilmiştir. Aynı zamanda, veritabanı tablolarının kontrol ve onarım süreleri de bu analizde dikkate alınmıştır.

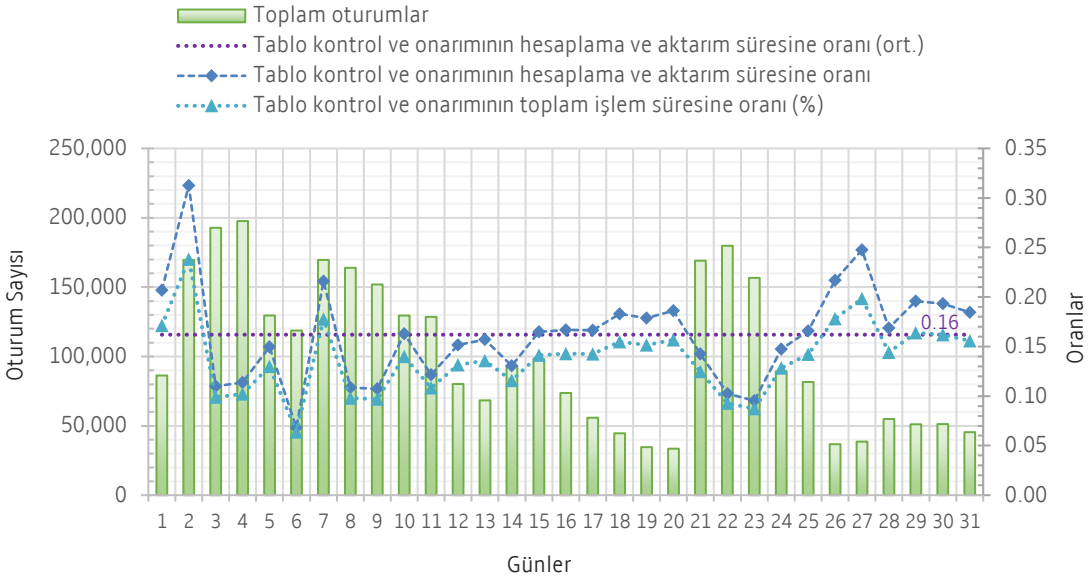


Şekil 5.13. Sayfa görüntüleme sayısına göre analitik çıkarımı işlem süreleri.

Gözlemlenen eğilimler, sayfa görüntüleme ile oturum bilgilerinin hesaplanması ve transferi işlemlerinin, artan sayfa görüntüleme sayılarına rağmen belirli bir verimlilikte gerçekleştiğini göstermektedir. Veritabanı tablosu kontrolü ve onarım süreçlerinin zaman alıcılığı, sayfa görüntüleme kayıt sayısının artışıyla birlikte daha belirgin hale gelmiş ve bu durum sistemin ölçeklenebilirlik sınırlarını ortaya koymuştur.

İşlem sürelerinin üstel eğilimleri, sistem kaynaklarının yönetimi ve veritabanı işlemlerinin optimizasyonu için önemli ipuçları sunmaktadır. Bu analiz, yüksek hacimli verilerin işlenmesinde karşılaşılabilecek zorlukların üstesinden gelmek için teknik iyileştirmelerin yapılmasının gerekliliğini vurgulamaktadır.

Veritabanı kontrol ve onarım işlemlerinin oturum sayılarıyla olan ilişkisi, ayın her günü için detaylı bir şekilde Şekil 5.14'te görselleştirilmiştir. Bu görselde, günlük oturum sayıları ve bu oturumlar sırasında gerçekleştirilen veritabanı işlemlerinin süreleri karşılaştırmalı olarak incelenmiştir. Bu metriklerin toplam işlem süresine oranı ile ortalama hesaplama ve transfer sürelerinin analizi, analitik çıkarımı sürecinin performansına dair önemli bilgiler sunmaktadır.



Şekil 5.14. Oturum sayısına göre kontrol ve transfer işlemlerinin ilişkisi.

Gün bazında incelenen veriler, oturum sayılarındaki dalgalanmaların, veritabanı kontrol ve onarım işlemlerinin süresi üzerinde belirgin bir etkiye sahip olduğunu ortaya koymaktadır. Özellikle, oturum sayılarının yüksek olduğu günlerde, bu işlemlerin toplam işlem süresine oranının da arttığı görülmektedir. Ortalama işlem süresinin ay boyunca görece sabit kaldığı, ancak bazı günlerde beklenmedik piklerin yaşandığı gözlemlenmiştir. Bu durum, sistemin belirli günlerde yoğun kullanımdan kaynaklanabilecek performans değişimlerine karşı hassasiyetini ve muhtemel ölçeklendirme ihtiyacını vurgulamaktadır.

5.3. Uzun Vadeli Analitikler

Uzun vadeli analitikler, veri ambarında yer alan ve üç aydan daha geniş zaman dilimlerini kapsayan verileri aylık ya da yıllık bazda ele alan analizlerdir. Bu tür analitik yaklaşımlar, veri setlerindeki zaman serisi değişimlerini, önceki ayların, yılların veya belirli dönemlerin verileri ile karşılaştırmalı bir biçimde değerlendirerek,

eğilimleri ve desenleri anlamayı hedefler. Uzun vadeli analitikler, kuruluşların hedef belirleme ve planlama için eğilimleri ve büyüme patikalarını değerlendirerek geleceğe dair stratejik tahminler yapmalarını mümkün kılar.

5.3.1. Veri ambarındaki verilerin fiziksel analizi

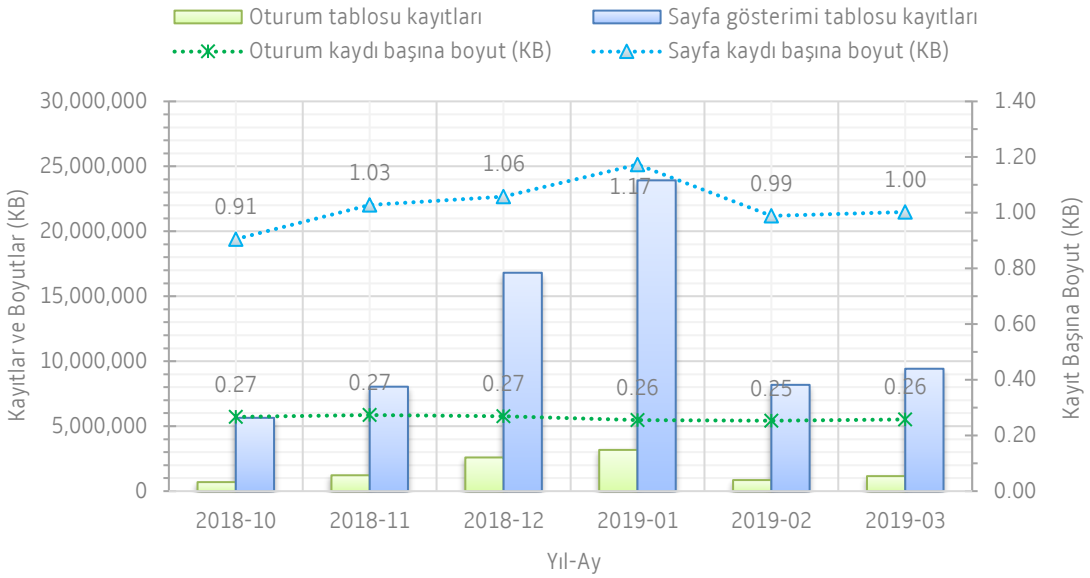
Fiziksel analizde, oturum ve sayfa gösterimlerine odaklanan altı aylık veri marketlerindeki kayıtların veri ve indeks boyutları teknik perspektiften incelenmektedir. Tablo 5.10, veri marketleri içerisindeki veri ve indeks boyutları ile bunların satır (kayıt) sayısı ve birbirleri arasındaki orantısal ilişkileri detaylandırmaktadır. Bu analiz, veri depolama ve sorgulama süreçlerinin etkinliğini değerlendirerek, veri yönetimi stratejilerinin gelişimine önemli katkılarda bulunmakta ve sistem kaynaklarının optimizasyonuna yönelik yol gösterici bilgiler sağlamaktadır.

Tablo 5.10. Veri marketlerindeki kayıtların fiziksel boyutları ve oranlar.

Veri Marketi (Yıl-Ay)	Kayıt Say. (Satır)	Veri Boyutu (B)	İndeks Boyutu (B)	VeriB/ İnd.B	İnd.B/ VeriB	VeriB/ Satır	İnd.B/ Satır
sess-2018-10	694.738	154.626.180	35.011.584	4,42	0,23	223	50,40
sess-2018-11	1.220.966	281.045.560	60.709.888	4,63	0,22	230	49,72
sess-2018-12	2.591.631	587.699.776	127.237.120	4,62	0,22	227	49,10
sess-2019-01	3.174.164	679.858.644	151.209.984	4,50	0,22	214	47,64
sess-2019-02	852.115	177.300.020	43.403.264	4,08	0,24	208	50,94
sess-2019-03	1.155.486	248.394.060	56.126.464	4,43	0,23	215	48,57
page-2018-10	5.643.064	4.883.231.552	347.798.528	14,04	0,07	865	61,63
page-2018-11	8.039.509	7.964.980.436	501.941.248	15,87	0,06	991	62,43
page-2018-12	16.810.614	17.137.541.376	1.071.361.024	16,00	0,06	1.019	63,73
page-2019-01	23.922.713	27.238.149.500	1.499.052.032	18,17	0,06	1.139	62,66
page-2019-02	8.181.662	7.784.880.790	503.005.391	15,48	0,06	952	61,48
page-2019-03	9.420.590	9.092.402.902	578.600.916	15,71	0,06	965	61,42

Burada görülen değerler, veri ve indeks boyutları arasındaki ilişkinin zaman içerisinde nasıl değiştiğini ve bu değişimin kayıt sayısıyla ilişkisini ortaya koymaktadır. Oturum ("sess") ve sayfa gösterimi ("page") veri marketlerinde kayıt sayıları arttıkça hem veri boyutlarının hem de indeks boyutlarının orantılı bir şekilde arttığı görülmektedir. Ancak bu artış, veri ve indeks boyutları arasındaki oranlarda belirgin bir değişikliğe neden olmayarak tutarlı bir eğilim sergilemektedir. Bu durum, kullanılan veri tabanı yönetiminin kararlılığını ortaya koymaktadır. Kayıt başına sayfa gösterimi verilerinde dikkat çeken farklılıklar ise saklanan verilerin içeriğine bağlı olarak değişen durumu yansıtmaktadır.

Oturum veri marketlerinde, veri boyutu ile indeks boyutu arasındaki oranlar nispeten sabitken, sayfa gösterimi veri marketlerinde bu oranlar daha yüksek gerçekleşmiştir. Bu durum, oturumlara göre çok daha fazla alan içeren sayfa gösterimi veri marketlerinin karmaşık indeksleme ihtiyaçlarına işaret etmektedir. Bu gözlemler, veri ambarı ve veri marketi yönetimi için kapasite planlaması ve performans optimizasyonu açısından önemli bilgiler sağlamaktadır. Şekil 5.15'te ise veri marketlerinde aynı döneme ait kayıt sayıları ile kayıt başına düşen boyut miktarları görsel olarak sunulmaktadır.



Şekil 5.15. Seçilen altı aylık veri marketlerinin kayıt sayıları ve boyutları.

Buna göre, veritabanındaki bir satırlık oturum verisi 0.25-0.27KB aralığında olup oturum sayısına göre değişim göstermemektedir. Ancak sayfa gösterimleri kayıt başına 0.91-1.17KB arasında yer kaplamakta ve sayfa gösterimleri arttıkça kayıt başına boyutta da bir miktar artma eğilimi gözlemlenmektedir. Bazı aylarda sayfa görüntüleme sayılarının 24 milyona kadar ulaşması, portalın oldukça yoğun bir biçimde kullanıldığını ortaya koymaktadır. Bu düzeyde yoğun kullanıma sahip bir sistem için kesintisiz hizmet sunumu, kriz senaryolarına hazırlık, sistem bakımı ve yedekleme gibi altyapı faaliyetlerinin dikkatle planlanması büyük önem taşımaktadır.

5.3.2. Aylık oturum verileri analizi

Seçilen üç yılın son üç ayına ait oturum verileri bu analizde özetlenmiştir. Tablo 5.11, her bir ay için minimum, ortalama, maksimum ve toplam oturum sayılarını içermektedir. Oturum verileri, kullanıcı etkileşiminin yoğunluğu ve web portalının

zaman içindeki kullanım eğilimlerini gözlemlemek amacıyla incelenmiştir. Bu veriler, web portalı trafiğinin dönemsel dalgalanmalarını ve yıllık kullanım kalıplarını anlamada oldukça önemlidir ve sistem kaynaklarının yönetimi ile kapasite planlamasında stratejik kararlar almak için temel teşkil eder.

Tablo 5.11. Üç yılın son üç ayına göre toplam ve sınır oturum sayıları.

Yıl/Ay	Minimum Oturum S.	Ortalama Oturum S.	Maksimum Oturum S.	Toplam Oturum S.
2017/10	9.734	17.641	24.759	546.857
2017/11	14.578	27.743	42.397	832.283
2017/12	37.418	52.277	68.793	1.620.599
2018/10	12.811	22.411	32.990	694.738
2018/11	10.348	40.699	87.771	1.220.966
2018/12	47.212	83.601	135.568	2.591.631
2019/10	6.703	20.815	34.167	645.267
2019/11	17.356	44.490	71.025	1.334.697
2019/12	40.610	75.819	122.857	2.350.378

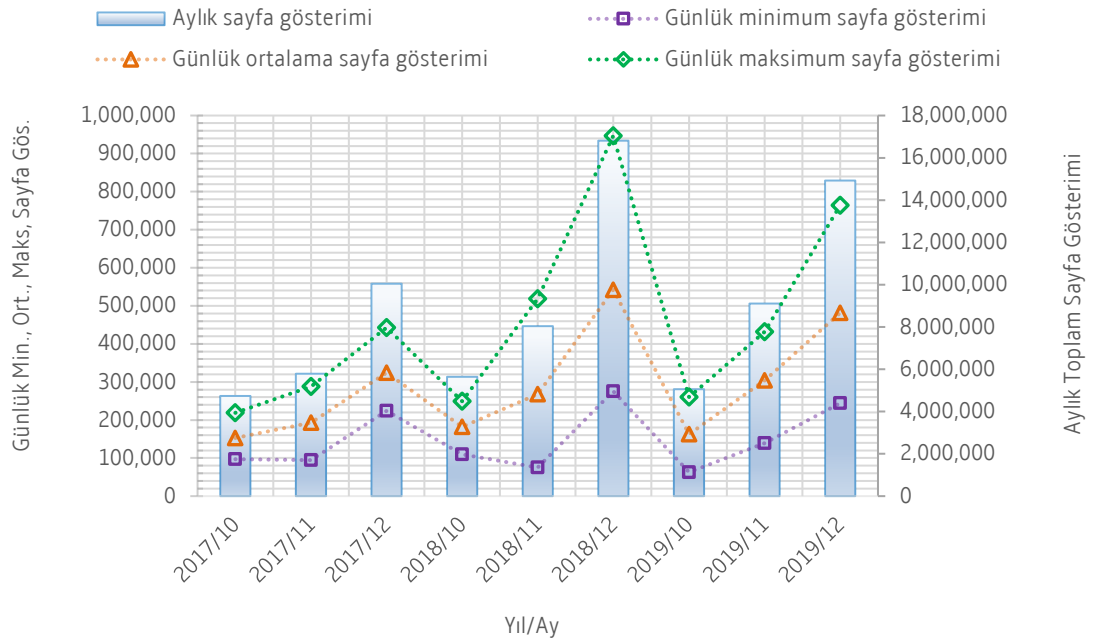
İncelenen dönem boyunca oturum sayılarındaki değişkenlik, belirli aylarda kullanıcı etkileşiminde önemli artışlar olduğunu göstermektedir. Özellikle 2017 ve 2018 yılının Aralık ayında kaydedilen maksimum oturum sayılarındaki büyük sıçramalar, kullanıcı aktivitesindeki dönemsel yükselişleri yansıtmakta ve bu dönemlerde web portalının daha yoğun kullanıldığını ortaya koymaktadır. Bu tür zirve dönemler, portal yöneticileri için altyapıyı güçlendirmek ve kullanıcı memnuniyetini maksimize etmek amacıyla hazırlıklı olunması gereken zamanlar olarak değerlendirilebilir.

Diğer yandan, minimum ve ortalama oturum sayılarında gözlemlenen dalgalanmalar, farklı aylar ve yıllar arasında kullanıcı davranışlarında çeşitlilik olduğunu gösterir. Toplam oturum sayılarının sürekli artış göstermesi, genel kullanıcı tabanının büyümesini ve web portalının giderek daha fazla kullanıcıyı çekmeyi başardığını işaret etmektedir. Bu veriler, portal içeriğinin ve kullanıcı deneyiminin sürekli iyileştirilmesi gerektiğini vurgularken, stratejik planlama çabaları için güçlü bir veri temeli oluşturur. Bu eğilimler ayrıca, hedeflenen kullanıcı etkileşim stratejilerinin etkinliğini ve web portalının uzun vadeli büyüme potansiyelini değerlendirmek için önemli bilgiler sunar.

5.3.3. Aylık sayfa görüntüleme verileri analizi

Veri ambarında birikmiş olan altı aylık verilere dayanarak, 2017-2019 arası yılların son üç ayı için sayfa görüntüleme istatistikleri Şekil 5.16'da detaylandırılmıştır. Bu dönemler boyunca kaydedilen minimum, ortalama, maksimum ve toplam sayfa görüntüleme sayılarını görselleştiren grafik, kullanıcıların gün ve ay bazında web

portalında gezinme eğilimlerini ve içeriğin kullanıcı tarafından nasıl algılandığını anlamak için kritik bir araçtır. Oturum verileriyle uyumlu bir şekilde, sayfa görüntülemeleri de portalın çekiciliği ve içerik erişilebilirliği hakkında önemli bilgiler sunar.

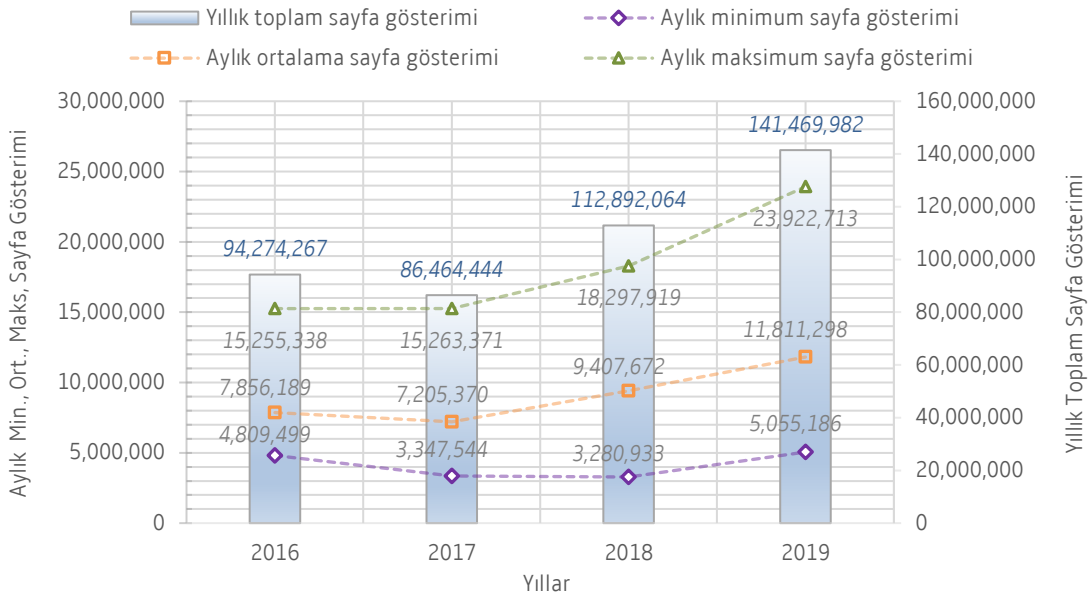


Şekil 5.16. Üç yılın son üç ayına göre sayfa gösterimi sınır değerleri.

İncelenen dönemde, günlük sayfa görüntüleme sayılarının minimum ve maksimum değerleri arasında belirgin değişkenlikler gözlenmiştir. Aralık 2018'deki sayfa görüntüleme sayısında yaşanan göze çarpan artış, kullanıcı aktivitesinin belirli dönemlerde yoğunlaştığına işaret etmektedir. Üniversitenin web portalındaki bu trafik pikleri, genellikle ders kayıtlarının yapıldığı dönem başlangıçları, sınav dönemleri ve dönem sonunda öğrencilerin not sonuçlarını sorguladıkları zamanlar gibi akademik takvimle ilişkili zamanlarda gözlemlenmektedir. Ortalama değerlerin genel eğilimi, web portalının içeriğine olan ilginin artan bir eğilim gösterdiğini işaret ederken, bu da portalın içeriği ve kullanıcı deneyimi açısından sürekli bir gelişim ihtiyacını ortaya koymaktadır. Oturum verileri ile paralellik gösteren bu artışlar, kullanıcı katılımının ve etkileşimlerinin sadece oturum sayılarıyla sınırlı olmadığını, aynı zamanda sayfa görüntülemelerinde de benzer bir artış eğilimi olduğunu göstermektedir. Bu gözlemler, portal yöneticilerinin kullanıcı davranışlarını anlama, içerik stratejilerini şekillendirme, hedeflenen kullanıcı deneyimini iyileştirme ve altyapı planlama konularındaki çabalarına rehberlik edebilir.

5.3.4. Yıllık sayfa görüntüleme verileri analizi

Veri ambarında incelenen dört yıllık süre zarfında, her ayın sayfa görüntüleme verileri Şekil 5.17'de görselleştirilmiştir. Bu görselleştirme, her yılın aylık bazda en düşük, ortalama ve en yüksek sayfa görüntüleme sayılarını ve aynı zamanda her yılın toplam sayfa görüntüleme sayılarını içermektedir. Veriler, portal kullanımının ve erişiminin zaman içerisinde nasıl değiştiğini göstermenin yanında, içerik ve kullanıcı etkileşim stratejilerinin etkinliğini değerlendirme imkânı da sunmaktadır.



Şekil 5.17. Dört yıl için sayfa gösterimlerinin aylık ve yıllık değerleri.

Görselleştirilen veriler, yıllar itibarıyla sayfa görüntüleme sayılarında önemli dalgalanmalar sergilemektedir. Özellikle 2019 yılında görülen toplam sayfa görüntüleme sayısındaki dikkat çekici artış, portalın giderek artan popülerliğini ve kullanıcı etkileşimindeki olumlu eğilimi göstermektedir. Bu artış, portala yeni hizmetlerin eklenmesinden olabileceği gibi, kullanıcı tabanındaki genişlemeden de kaynaklanabilir. Ortalama sayfa görüntüleme değerlerindeki istikrarlı artış, kullanıcıların sitede geçirdikleri sürenin arttığını ve daha fazla içerikle etkileşimde bulunduklarını gösterirken, bu da kullanıcıların sitedeki deneyimlerinin genel olarak iyileştiğini düşündürmektedir.

Bunun yanı sıra, minimum sayfa görüntüleme sayılarındaki yıllara göre değişim, potansiyel olarak düşük trafiğe sahip dönemleri işaret etmektedir. 2016'dan 2019'a doğru genel bir artış eğilimi olmasına rağmen, bu düşük değerler dikkate alınmalı ve içerik stratejilerinin bu dönemlerde nasıl optimize edilebileceği üzerine

düşünülmelidir. Yıl bazında maksimum sayfa görüntüleme sayılarının değişimi, özel etkinliklerin veya belirli olayların trafiği nasıl etkileyebileceği konusunda önemli ipuçları sunabilir. Bu gözlemler, web portalının içeriğini ve kullanıcı etkileşimini yönlendirme stratejilerinin yanı sıra, teknik altyapının ölçeklenmesi ve kaynak tahsisi için de önemli bilgiler sağlar.

5.3.5. Yıllık oturum ve sayfa görüntüleme verilerinin ay bazlı analizi

Aylık ve yıllık bazdaki oturum ve sayfa görüntüleme verileri, 2017-2019 yılları arasındaki üç yıllık zaman dilimini kapsayacak şekilde Tablo 5.12'de sunulmuştur. Her sütun, ilgili yıla ait oturum sayılarını, sayfa görüntülemelerini ve oturum başına düşen ortalama sayfa görüntüleme sayılarını içermektedir. Bu veriler, web portalının performansının yıllık ve aylık değişimini zamana göre değerlendirmede kullanılır ve kullanıcıların içerikle etkileşim kalıplarını anlamada yardımcı olur. Ayrıca, bu bilgiler portalın içerik yönetimi stratejilerinin ve kullanıcı deneyimini iyileştirme çabalarının etkinliğini değerlendirmede kritik bir rol oynar.

Tablo 5.12. Üç yıl boyunca aylık ve yıllık oturum ve sayfa görüntüleme bilgileri.

	Oturum Sayıları			Sayfa Gösterimi Sayıları			Ot. Başına S. Gös.		
Ay/Yıl	**2017**	**2018**	**2019**	**2017**	**2018**	**2019**	**2017**	**2018**	**2019**
1	2.517.516	2.837.916	3.174.164	14.269.343	18.297.919	23.922.713	5,67	6,45	7,54
2	666.811	769.748	852.115	4.345.731	5.896.126	8.181.662	6,52	7,66	9,60
3	632.259	919.937	1.155.486	4.607.032	6.081.282	9.420.590	7,29	6,61	8,15
4	1.323.293	1.925.391	2.755.775	7.925.933	10.462.232	13.656.601	5,99	5,43	4,96
5	2.435.856	2.508.402	4.237.729	15.263.371	15.094.886	22.678.098	6,27	6,02	5,35
6	666.466	1.165.502	1.131.752	4.617.261	7.756.947	6.973.326	6,93	6,66	6,16
7	522.414	514.844	1.531.188	3.347.544	3.280.933	8.833.204	6,41	6,37	5,77
8	847.058	952.648	1.401.080	5.463.985	6.244.478	6.749.277	6,45	6,55	4,82
9	744.245	1.142.624	1.447.650	6.050.950	9.284.074	12.230.150	8,13	8,13	8,45
10	546.857	694.738	645.267	4.737.600	5.643.064	5.055.186	8,66	8,12	7,83
11	832.283	1.220.966	1.334.697	5.789.176	8.039.509	9.108.563	6,96	6,58	6,82
12	1.507.783	2.591.631	2.350.378	9.351.167	16.810.614	14.926.206	6,20	6,49	6,35
Toplam	13.242.841	17.244.347	22.017.281	85.769.093	112.892.064	141.735.576	-	-	-
Min.	522.414	514.844	645.267	3.347.544	3.280.933	5.055.186	5,67	5,43	4,82
Ort.	1.103.570	1.437.029	1.834.773	7.147.424	9.407.672	11.811.298	6,79	6,76	6,82
Mak.	2.517.516	2.837.916	4.237.729	15.263.371	18.297.919	23.922.713	8,66	8,13	9,60

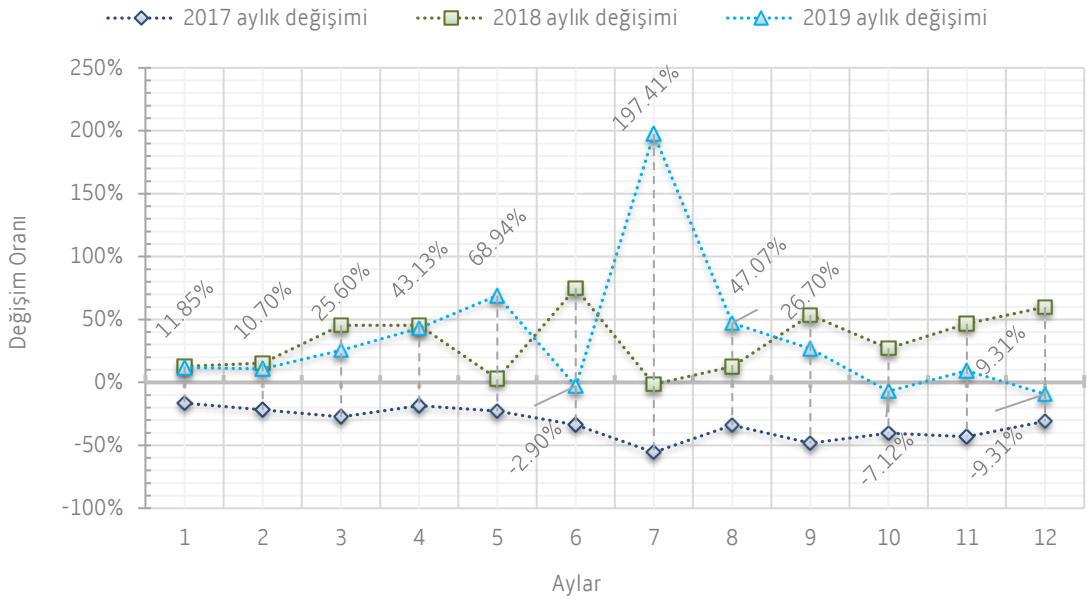
İncelenen dönem boyunca oturum ve sayfa görüntüleme sayılarındaki artış, portalın kullanımının genişlediğini ve içeriğin daha fazla kullanıcı tarafından tüketildiğini göstermektedir. Ortalama sayfa görüntüleme sayılarındaki yıllık artış, kullanıcıların portal içinde daha fazla içerikle etkileşime girdiğini işaret ederken, oturum başına sayfa görüntüleme oranlarındaki dalgalanma, farklı aylar ve yıllar boyunca değişen

kullanıcı katılımı derinliğini yansıtmaktadır. Özellikle oturum başına maksimum sayfa görüntüleme sayısının 2019 yılında elde edilen 9,60 değeri, kullanıcıların portalda geçirdikleri zamanın kalitesinin arttığını ve daha derin bir etkileşime girdiklerini göstermektedir.

Bu tür bir veri seti, portal yöneticilerine, içerik ve sunum stratejilerinin hangi aylarda ve yıllarda gözden geçirilmesi gerektiği konusunda fikir verir. Yıllık oturum sayılarındaki artış eğilimi, portalın genişleyen kullanıcı tabanını ve artan popülerliğini gösterirken, sayfa görüntüleme oranlarındaki değişkenlik, portalın içeriğinin kullanıcı ihtiyaçlarına ve tercihlerine ne derece uyum sağladığının bir göstergesi olarak değerlendirilebilir. Bu bilgiler, gelecekteki içerik geliştirme ve kullanıcı etkileşimi stratejilerinin tasarlanması için önemli bir temel oluşturur.

5.3.6. Önceki yıllara göre oturum sayılarındaki değişimin analizi

Bu analizde, yıllık bazda aylık oturum sayılarının nasıl değiştiği incelenmektedir. Şekil 5.18'de, bir önceki yılın aynı ayına göre oturum sayılarındaki yüzdesel değişim oranları gösterilmektedir.



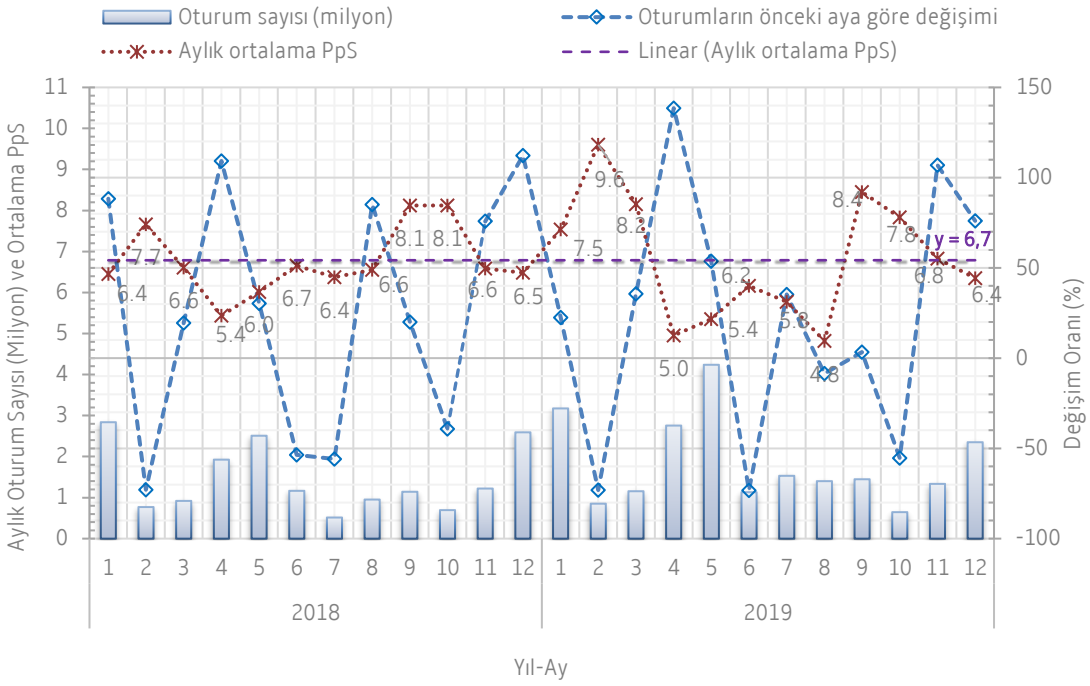
Şekil 5.18. Bir önceki yılın aynı ayına kıyasla oturum sayılarındaki değişim.

Bu karşılaştırmalı gösterim, web portalının kullanımındaki dönemsel dalgalanmaları ve yıllık gelişim eğilimlerini ortaya koymaktadır. Verilerin bu şekilde analizi, çeşitli olayların etkilerini ve kullanıcı etkileşimindeki diğer potansiyel değişiklikleri değerlendirmek için önemli bir yöntemdir.

İncelenen 2017 ve 2019 yılları arasındaki belirli aylarda oturum sayılarında belirgin artışlar gözlenmiştir. Değişim, özellikle 2019 yılı Ağustos ayında, önceki yıla göre iki kata yakın bir artışla zirveye ulaşmıştır. Bu durum muhtemelen akademik takvimdeki bir kayma ya da web portalının işlevselliğindeki önemli bir güncellemenin bir sonucu meydana gelmiş olabilir. Buna karşılık, bazı aylarda yaşanan azalışlar, kullanıcıların etkileşiminde dönemsel düşüşleri veya belirli içeriklere olan ilginin azalmasını göstermektedir. Bu gözlemler, portalın yönetimi için stratejik planlama ve içerik sunumunun yanı sıra teknik altyapıya yatırım yapma kararları üzerinde yöneticilere fikir vermektedir.

5.3.7. Önceki aylara göre oturum ve PpS sayılarındaki değişimin analizi

Yirmi dört aylık periyotta oturum sayıları ve bu oturumlar sırasında gerçekleşen sayfa görüntülemelerinin aylık oranlarındaki değişimi Şekil 5.19'da görselleştirilmiştir. Bu iki metrik arasındaki ilişki, kullanıcı etkileşiminin kalitesi ve web portalı içeriğinin etkinliği hakkında değerli bilgiler sağlar. Aynı zamanda, önceki aya göre oturum sayılarındaki değişim oranları, kullanıcı etkileşimindeki aylık eğilimler ve dönemsel dalgalanmalar hakkında bilgi verir. Doğrusal eğilim çizgisi ise, sayfa görüntüleme oranlarının genel eğilimini belirleyerek, uzun vadeli kullanıcı davranışı analizlerine yardımcı olur.

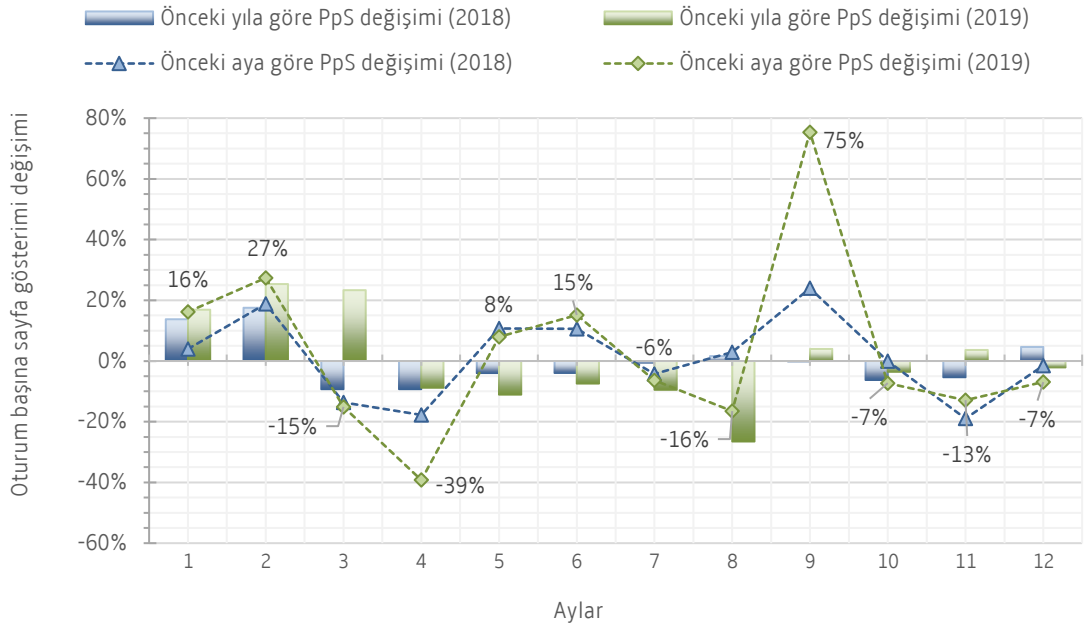


Şekil 5.19. Yirmi dört ay boyunca oturum ve PpS oranlarındaki değişim.

Analiz edilen dönemde, oturum sayıları ve oturum başına sayfa görüntüleme oranlarında dikkate değer aylık dalgalanmalar göze çarpmaktadır. Özellikle sayfa görüntüleme oranlarında, yılın belli dönemlerinde artışlar görülürken, diğer zamanlarda bu oranlarda düşüşler yaşanmıştır. Bu değişkenlik, akademik takvimde yıllara göre yaşanan zaman kaymaları, tatil dönemleri veya portal üzerindeki diğer etkinlikler gibi faktörlerle açıklanabilir. Oturum sayılarındaki düşüş ve artışlar, kullanıcıların siteye olan ilgisinin nasıl değiştiğini gösterirken, sayfa görüntüleme oranlarındaki değişimler, kullanıcıların sitelerde geçirdikleri zamanın ve etkileşim derinliğinin bir göstergesidir. Bu bilgiler, web portalının içerik ve kullanıcı etkileşim stratejilerinin daha da geliştirilmesi için temel oluşturur.

5.3.8. Önceki aya ve yıla göre PpS sayılarındaki değişimin analizi

Oturum başına sayfa görüntüleme sayısı (PpS), kullanıcıların web portalındaki her oturumda ne kadar içerik tükettiğinin bir göstergesi olarak değerlendirilir ve kullanıcı etkileşiminin kalitesi üzerine derinlemesine bir bakış sağlar. Şekil 5.20, oturum başına düşen sayfa görüntüleme sayısının, önceki aya ve önceki yıla kıyasla aylık değişimini göstermektedir. Önceki yılın aynı ayına ve önceki aya göre değişim yüzdeleri, kullanıcı davranışlarındaki ve etkileşim derinliğindeki değişiklikleri yansıtarak, içerik yönetimi ve kullanıcı deneyimi stratejilerinin etkinliğini ölçmede yardımcı olur.

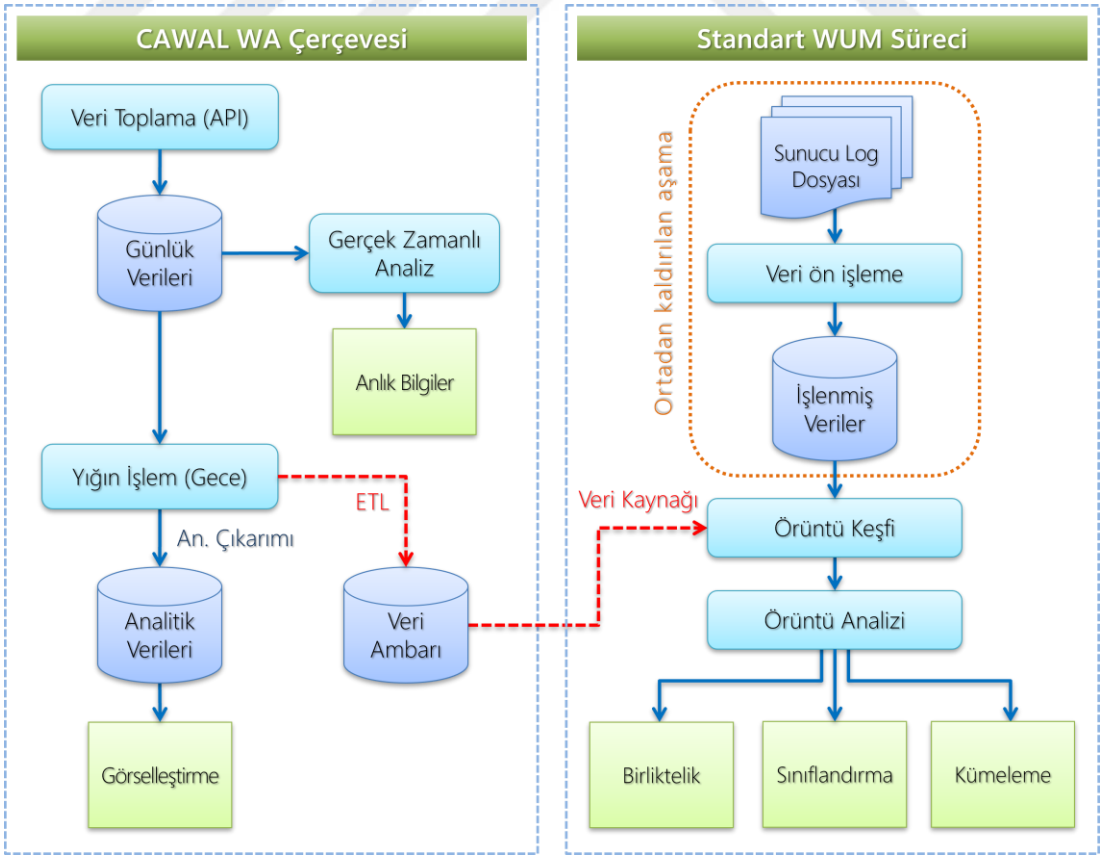


Şekil 5.20. Önceki aya ve yıla kıyasla PpS oranındaki değişim.

2018 ve 2019 yıllarında PpS oranlarındaki dalgalanmalar, kullanıcıların sitelerdeki etkileşimlerinin zaman içinde nasıl değiştiğini açıkça göstermektedir. Özellikle 2019 yılının Eylül ayında gözlemlenen %75'lik artış dikkat çekici olup, bu, belirli bir olayın kullanıcı etkileşimini önemli ölçüde artırdığını işaret etmektedir. Bazı aylarda yaşanan bu tür yükselişler derse yazılma, sınav sonuçlarının ilanı gibi portalın kullanım nedenlerinden kaynaklanabilmektedir. Yine de kullanım analizleri yapılırken bir saldırı ya da anomali durumunun varlığı da göz ardı edilmemeli ve bu tür durumlar daha detaylı analizler ile incelenmelidir. Diğer taraftan, bazı aylarda görülen düşüşler, kullanıcı ilgisindeki dönemsel değişimleri göstermektedir. Bu eğilimler, içerik geliştirme ve site optimizasyonu açısından alınacak kararlar için değerli bilgiler sunar ve web portalının kullanıcı etkileşimini artırmaya yönelik gelecekteki eylemleri şekillendirmeye yardımcı olur.

6. CAWAL İLE WEB KULLANIM MADENCİLİĞİ

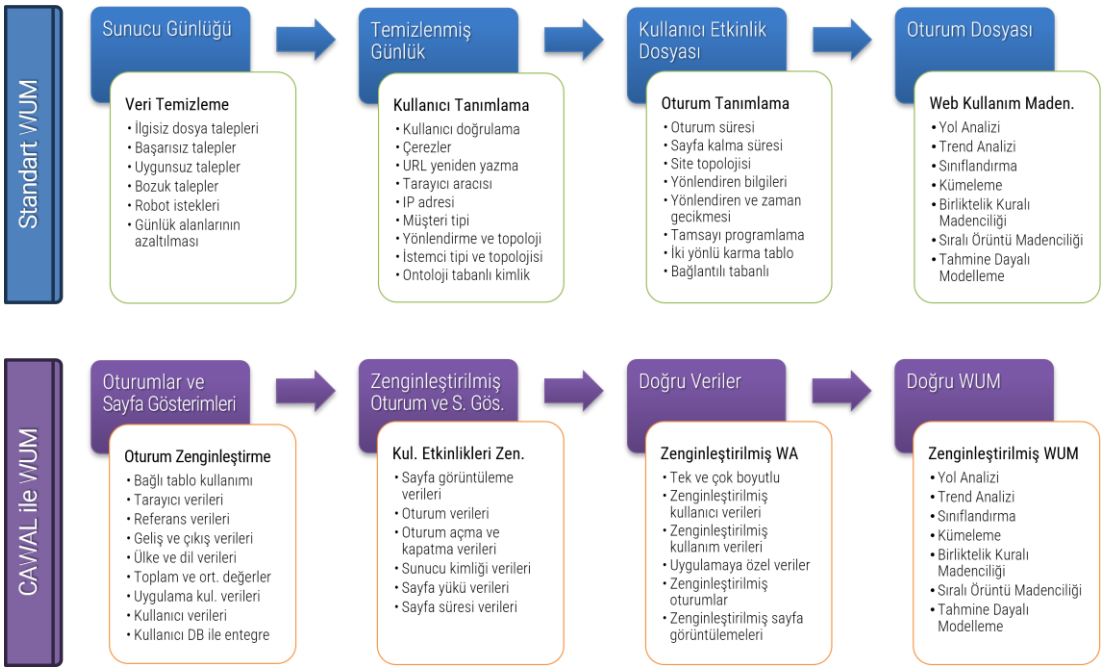
Madencilik faaliyetlerinde kullanılan verilerin kalitesi, analizlerin doğruluğunu ve verimliliğini etkileyen kilit bir faktördür. Web kullanım madenciliğinin geleneksel veri kaynağı olarak kabul edilen sunucu günlüklerinde ham veriler yarı yapılandırılmış biçimde bulunur. Bu yapı, WUM'un ilk ve kritik aşamalarından biri olan veri ön işlemede önemli zorluklar ortaya çıkartır. Veri temizleme, dönüştürme, kullanıcı ve oturum tanımlama gibi süreçler yüksek çaba gerektirir ve çoğu zaman istenen başarı düzeyine ulaşamaz. Ancak, CAWAL modeliyle elde edilen veriler oturum ve uygulama verilerinin de dahil edilmesiyle başlangıçtan itibaren daha zengin ve organize bir yapıda saklanırlar. Bu veriler, WUM süreçlerinde kullanım için benzersiz ve yüksek kaliteli bir veri kaynağı olarak kullanıcı etkileşimlerinin daha detaylı ve doğru bir şekilde izlenmesini sağlar. CAWAL'ın WUM ön işleme aşamasını büyük ölçüde ortadan kaldırarak süreci nasıl hızlandırdığı Şekil 6.1'de gösterilmiştir.



Şekil 6.1. CAWAL ile toplanan verilerin WUM veri kaynağı olarak kullanımı.

CAWAL'ın veri ambarında saklanan, yüksek kalite ve doğruluk standartlarına uygun veriler, web kullanım madenciliğinde geleneksel olarak başvurulan sunucu günlüklerine olan ihtiyacı tamamen ortadan kaldırmakta ve ön işleme adımını gereksiz kılmaktadır. Bu yenilikçi yaklaşım, veri temizleme, normalizasyon ve dönüştürme işlemlerinin CAWAL tarafından toplanan veriler üzerinde daha az zaman ve kaynak harcanmasını sağlayarak, analiz süreçlerini hızlandırır ve veri madenciliği sonuçlarının doğruluğunu iyileştirir. Aynı zamanda, yapılandırılmış veri formatının benimsenmesi, karmaşık kullanıcı davranışı desenlerinin anlaşılmasını ve etkili bir şekilde analiz edilmesini kolaylaştırır.

Veri ambarındaki verilerin WUM uygulamalarına yönelik hazırlanmasında veri seçimi ve zenginleştirme işlemleri önemli bir rol oynamıştır. Veri ambarı içerisinde bir aylık süreyi kapsayan, toplam 8.5GB boyutundaki veri marketleri üzerinde gerçekleştirilen karmaşık SQL sorguları sayesinde zenginleştirilmiş oturum ve sayfa gösterimi verileri elde edilmiştir. Elde edilen bu zenginleştirilmiş veri setleri üzerinde, kullanıcı davranışlarının analizi, kullanıcı ihtiyaçlarının daha iyi anlaşılması ve web portalının kullanım etkinliğinin artırılması hedeflenmiştir. Srivastava ve ark. (2019) tarafından detaylandırılan standart WUM süreci ile CAWAL çerçevesi aracılığıyla gerçekleştirilen zenginleştirilmiş WUM süreci karşılaştırma amacıyla Şekil 6.2'de verilmiştir.



Şekil 6.2. Standart WUM süreci ve CAWAL ile zenginleştirilmiş WUM.

CAWAL çerçevesi aracılığıyla toplanan ve veri ambarında biriktirilen bilgiler üzerinde, oturum ve sayfa gösterimlerinin detaylı analizi için web kullanım madenciliği teknikleri uygulanmıştır. Bu süreçte sık kullanılan desen analizi, kullanıcı segmentasyonu ve davranışsal modelleme yöntemleri etkin bir şekilde kullanılmıştır. İlk olarak, "Zenginleştirilmiş Oturum ve Sayfa Gösterimi Verilerinin Oluşturulması" bölümünde, veri ambarındaki ham verilerin nasıl işlendiği ve analiz için hazırlandığı detaylandırılmaktadır. Daha sonra, "Oturum Verilerine Uygulanan WUM Analizleri ve Tahminleri" başlığı altında hemen çıkma oranları, istemci niteliklerinin analizi, sistemden ayrılma yöntemlerinin servis bazlı analizi, sistemden ayrılma yöntemlerinin ve son terk edilecek servisin tahmini gibi önemli bulgulara ulaşılmıştır. Ayrıca, "Sayfa Gösterimi Verilerine Uygulanan WUM Analizleri ve Tahminleri" başlığı içerisinde portal servis etkileşimlerinin ve geçiş yollarının analizi, birliktelik kurallarının keşfi ve sunucu ile sayfa yüklenme süresi bazlı anomali tespiti gibi çeşitli analizler yapılmıştır.

CAWAL'ın hassas veri toplama yeteneği ve veri doğruluğundaki başarısı, kullanıcı davranışlarının ayrıntılı ve güvenilir bir biçimde incelenmesine imkân tanımış, önemli bulgular ve stratejik çıkarımlar elde edilmesine zemin hazırlamıştır. Yapılan analiz ve tahminler, CAWAL'ın web portalları ve kurumsal ölçekli web uygulamalarının kullanım etkinliğinin ve kullanıcı deneyiminin iyileştirilmesinde önemli bir potansiyele sahip olduğunu göstermektedir.

6.1. Zenginleştirilmiş Oturum ve Sayfa Gösterimi Verilerinin Oluşturulması

Web analitiği ve web kullanım madenciliği süreçleri genellikle büyük ve karmaşık veri setlerinin işlenmesini gerektirir. Bu süreçlerde, verilerden anlamlı bilgiler ve değerli içgörüler elde etmek için detaylı analizler yapılması gerekmektedir. CAWAL çerçevesi tarafından toplanan verilerin işlenmesi için de kapsamlı bir süreç yürütülmüştür. Bu bağlamda, zenginleştirilmiş oturum ve sayfa gösterimi tablo görünümleri (view) için oluşturulan sofistike SQL sorguları, veri setlerinin etkin bir şekilde işlenmesi ve anlamlı içgörülerin elde edilebilmesi için temel bir yapı taşı görevi görmektedir. Heterojen yapıdaki büyük hacimli veri kümelerinin etkin bir şekilde sorgulanmasını, filtrelenmesini ve dönüştürülmesini sağlayan bu kapsamlı sorguların kullanımı, gelişmiş analitik sonuçlar ve çok boyutlu veri setleri elde etmek için hayati derecede önemlidir.

Yapılan sorgular sonucunda karmaşık ilişkileri çözümlemekte ve çok katmanlı veri ilişkilerini ortaya çıkarmakta kullanılacak oturum ve sayfa gösterimleriyle ilgili mümkün olan tüm veriler detaylı şekilde elde edilmiştir. Veritabanında tablo görünümü yapısıyla oluşturulan bu veriler daha sonra ikinci bir işleme tabi tutulmuş ve WUM uygulamaları için özellikle sayısal ve kategorik verileri içerecek şekilde ayrı birer tablo görünümü şeklinde kaydedilmiştir. Son adımda, Tablo 6.1'de verilen zaman aralıklarındaki veriler tablo görünümünden alınarak zenginleştirilmiş oturum ve sayfa gösterimi veri setleri oluşturulmuştur. Bu veri setleri, Python programlama diliyle yazılan kodlar tarafından kolayca işlenebilir hale getirilmek üzere virgülle ayrılmış CSV biçimindeki ayrı dosyalara kaydedilmiştir.

Tablo 6.1. Farklı periyotlarda oluşturulan CSV veri dosyaları ve özellikleri.

Zaman Dilimi	Zaman Aralığı	Sess. Dosya Adı .CSV	Sess. Kayıt Sayısı	Sess. Dosya Boyutu MB	Page Dosya Adı .CSV	Page Kayıt Sayısı	Page Dosya Boyutu MB
4 saat	2018-11-22 13.00 - 17.00	va_sess1	13.820	2,66	va_page1	95.933	8,08
1 gün	2018-11-22 00.00 – 23.59	va_sess2	122.908	23,60	va_page2	787.637	66,30
3 gün	2018-11-21 - 2018-11-23	va_sess3	247.046	47,60	va_page3	1.572.360	132,00
1 hafta	2018-11-21 - 2018-11-27	va_sess4	514.879	99,20	va_page4	3.158.694	266,00
1 ay	2018-11-01 - 2018-11-30	va_sess5	1.220.916	235,00	va_page5	8.039.339	678,00

Zenginleştirilmiş oturum verileri, her bir kayıt için oturum tablosundan alınan verilerin yanı sıra, kullanıcıya ait bilgiler ve ilgili oturuma ait sayfa gösterimi verilerinden SQL sorguları aracılığıyla elde edilen ilk, son, toplam, ortalama gibi değerli özet bilgiler içerir. Web kullanım madenciliği işlemlerinde kullanmak üzere sadece sayısal ve kategorik oturum bilgilerini (sess) içeren CSV dosyalardaki alan adları, açıklamaları ve örnek veriler Tablo 6.2'de sunulmuştur. Bu alanların devamında ise kullanıcıların ilgili oturumda portal servisleriyle olan etkileşimlerini özetleyen alanlar yer almaktadır. "s_" ile başlayan alanlar (s_gate, s_mail, s_obis, vb.), kullanıcının ilgili servisi ziyaret edip etmediğini gösterir. Burada "1" ziyaret edildiğini, "0" ise ziyaret edilmediğini belirtir. "p_" ile başlayan alanlar (p_gate, p_mail, p_obis, vb.), her serviste kullanıcının gezdiği sayfa sayısını kaydeder. Son olarak, "r_" ile başlayan alanlar (r_gate, r_mail, r_obis, vb.), her serviste gezilen sayfa sayısının oturumda gezilen toplam sayfa sayısına oranını gösterir. Bu metrikler, kullanıcı davranışlarını ve etkileşimleri detaylı bir şekilde analiz etmek için kullanılır.

Tablo 6.2. Sayısal oturum bilgilerini içeren CSV dosya yapısı.

Alan Adı	Açıklama	Örnek Veri
Log_ID	Log kimlik numarası.	12359285
Session_ID	Oturum kimlik numarası.	83665107
Log_Date_Time	Log tarih ve saat bilgisi.	22.11.2018 13:00
Log_Date	Log tarih bilgisi.	22.11.2018
User_ID	Kullanıcı kimlik numarası.	184922
Session_Login_Status	Oturumda sisteme giriş durumu.	1
Logins_During_Period	Periyod boyunca yapılan girişlerin sayısı.	16
User_Type	Kullanıcı türü (Akd./İdari/Öğr./Birim vb).	6
Sex	Cinsiyet (Belirsiz/Erkek/Kadın).	2
Age	Yaş.	18
Age_Group	Yaş grubu (Kategorik, 1-4).	1
User_Language_TR	Tarayıcı dili.	1
User_Location	IP konumu.	1
Browser_Type	Tarayıcı türü.	1
Referer_Type	Yönlendirici türü.	6
Landing_Srv_ID	İlk erişilen servis kimlik numarası.	7
Exit_Srv_ID	Çıkış yapılan servis kimlik numarası.	3
Exit_Type	Çıkış türü.	0
Total_Session_Duration	Toplam oturum süresi.	730
Avg_Page_Duration	Ortalama sayfa süresi.	48,67
Total_Page_Load	Toplam sayfa yüklemesi.	2,88
Avg_Page_Load	Ortalama sayfa yüklemesi.	0,19
Page_Count	Oturumda gezilen toplam sayfa sayısı.	15
Visitor_PageView	Ziyaretçi olarak gezilen sayfa sayısı.	2
User_PageView	Kullanıcı olarak gezilen sayfa sayısı.	13
Service_Count	Gezilen servis (hizmet) sayısı.	2
Page_per_Service	Servis başına düşen sayfa sayısı.	7.5
Visited_Service_IDs	Ziyaret edilen servis kimlik numaraları.	“1,3”

Oturum verilerinin ardından her bir kayıt için sayfa gösterim verilerinin devamına kullanıcı ve oturum bilgileri eklenerek zenginleştirilmiş sayfa gösterimi verileri oluşturulmuştur. Web kullanım madenciliği işlemlerinde kullanmak üzere sadece sayısal ve kategorik sayfa görüntüleme bilgilerini (page) içeren CSV dosyalardaki alan adları, açıklamaları ve örnek veriler Tablo 6.3’te verilmiştir.

Tablo 6.3. Sayısal sayfa gösterimi bilgilerini içeren CSV dosya yapısı.

Alan Adı	Açıklama	Örnek Veri
Detail_ID	Detay kimlik numarası.	89010871
Session_ID	Oturum kimlik numarası.	83665107
Detail_Date_Time	Detayın tarih ve saat bilgisi.	22.11.2018 13:01
Detail_Date	Detayın tarih bilgisi.	22.11.2018
User_ID	Kullanıcı kimlik numarası.	184922
Current_Login_Status	Anlık sisteme giriş durumu.	1
Session_Login_Status	Oturumda sisteme giriş durumu.	1
User_Type	Kullanıcı türü.	6
Sex	Cinsiyet.	2
Age	Yaş.	18
Age_Group	Yaş grubu.	1
User_Language_TR	Tarayıcı dili.	1
User_Location	IP konumu.	1
Browser_Type	Tarayıcı türü.	1
Referer_Type	Yönlendirici türü.	6
Server_ID	Sunucu kimlik numarası.	4
Service_ID	Servis kimlik numarası.	3
Page_Duration	Sayfada kalma süresi.	41
Page_Load_Time	Sayfa yüklenme süresi.	0,122

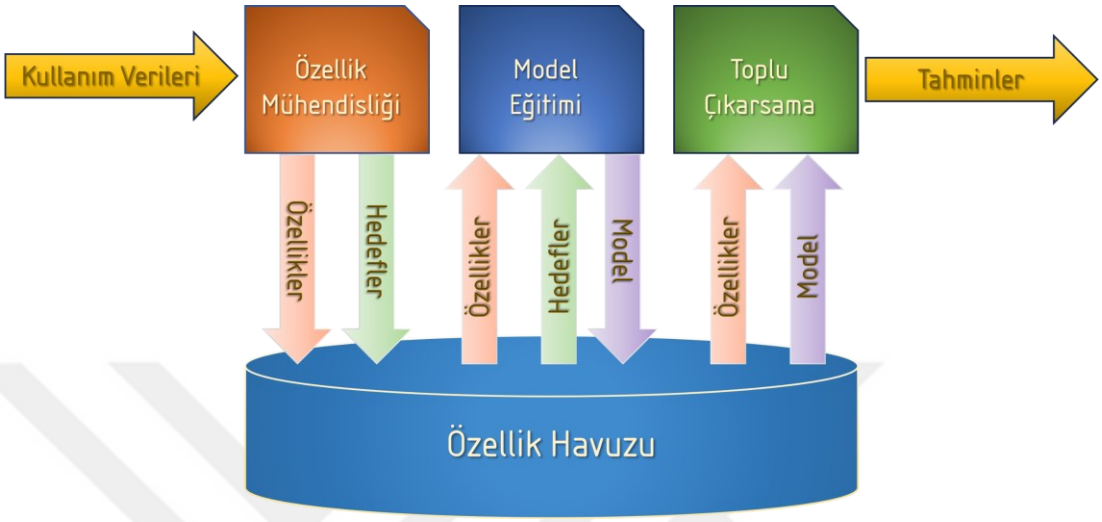
Bu veri setleri, kullanıcıların etkileşimlerinin ve portalın çeşitli servislerine olan ilgilerinin zaman içinde nasıl evrildiğini anlamak için kritik veri kaynaklarıdır. Oluşturulan bu veri setleri ayrıca, belirli günlerdeki veya dönemlerdeki kullanım örüntülerine yönelik detaylı analizler için de temel oluştururlar.

6.2. Oturum Verilerine Uygulanan WUM Analizleri ve Tahminleri

Web kullanım madenciliği, kullanıcıların web siteleri ile etkileşimlerini anlamak ve bu etkileşimlere dayalı olarak kullanıcı deneyimini iyileştirmek için zengin veri kaynaklarından faydalanır. CAWAL çerçevesi, bu zengin veri kaynaklarını sistemli bir şekilde bir araya getirerek, web kullanım verilerinin analitik çıkarımlar için işlenmesinde önemli bir rol oynar. Bu alt bölümde, CAWAL çerçevesi kullanılarak toplanan ve işlenen zenginleştirilmiş oturum verileri üzerinde gerçekleştirilen bazı analiz ve tahminler ele alınmaktadır.

Veri madenciliği süreçlerinin başarısı, ham verinin doğru bir şekilde işlenip özelliklere dönüştürülmesine bağlıdır. Yapılan analiz ve tahminlerde, oturum verilerinin

temizlenmesi, özelliklerin seçimi ve bu özelliklerin modelleme süreçlerine hazırlanması aşamalarını içeren kapsamlı süreçler uygulanmıştır. Şekil 6.3, kullanım verilerinin özellik mühendisliği ile tahminlere dönüşüm sürecinin kavramsal bir özetini sunmaktadır.



Şekil 6.3. Kullanım verilerinin tahminlere dönüşüm süreci.

Bu süreçte kullanım verileri ilk olarak özellik mühendisliği aşamasına taşınır ve burada kullanıcı etkileşimlerinin karmaşıklığını yakalayabilecek önemli öznitelikler belirlenir. Daha sonra bu öznitelikler, veri içerisindeki desenleri ve eğilimleri ortaya çıkarmak amacıyla tahmin modellerini eğitmek için kullanılır. Eğitilen modeller, toplu tahminlerde bulunmak ve karar verme süreçlerine yönelik öngörülerde bulunmak üzere kullanılır. Özellik havuzu olarak adlandırılan bu merkezi depo, özellik mühendisliği ve model eğitiminin doğru ve tutarlı bilgilerle gerçekleştirilmesini sağlayarak WUM sürecinin etkinliğini artırır.

6.2.1. Hemen çıkma oranı ve istemci özelliklerinin analizi

Önemli bir web analitiği metriği olan hemen çıkma oranı (bounce rate), kullanıcıların bir web sitesinde sadece tek bir sayfayı ziyaret ettikleri ve daha fazla etkileşimde bulunmadan siteyi terk ettikleri oturumların yüzdesel oranını ifade eder. Bu oran, kullanıcıların içerikle olan etkileşim seviyelerini ve web sitesi tasarımının kullanıcı deneyimi üzerindeki etkisini anlamak için kritik bir göstergedir. Bir aylık veriler üzerinde yapılan detaylı istatistiksel analiz, oturum ve sayfa gösterimleri açısından tek sayfa ve çok sayfa oturumları arasındaki farklılıkları ortaya koymaktadır.

İncelenen veri setinde yer alan toplam 1.220.916 oturumun 156.707'si tek sayfa oturumu, 1.064.209'u ise çok sayfa oturumudur. Tek sayfa oturumlarının toplam oturumlar içindeki oranı %12,84 iken, çok sayfa oturumları %87,16 oranında gerçekleşmiştir. Sayfa gösterimi sayılarına bakıldığında, tek sayfa oturumlarında oturum sayısı kadar, yani 156,707 sayfa gösterimi bulunurken, çok sayfa oturumlarında toplam 7.882.632 sayfa gösterimi bulunmaktadır. Bu da toplam sayfa gösterimlerinin %1,95'ini tek sayfa ve %98,05'ini çok sayfa oturumlarının oluşturduğunu göstermektedir.

Bu bulgular, kullanıcıların site üzerindeki etkileşim düzeyleri ve içerikle olan bağlantıları hakkında önemli bilgiler sunmakta ve web sitesinin kullanıcı deneyimi açısından değerlendirilmesi için kritik veriler sağlamaktadır. Kullanıcıları sunucuya erişimleri sırasında ayırt etmek için başvurulan dört kategorik özellik ile bu özelliklerin tek sayfa ve çok sayfa gezen kullanıcı grupları arasındaki dağılımı, Tablo 6.4'te detaylı bir şekilde sunulmuştur.

Tablo 6.4. İstemci özelliklerinin tek ve çok sayfa ziyaretlerine göre dağılımları.

Alan Adı ve Özellikleri	Tek Sayfa	Çok Sayfa
Browser_Type:		
1-Standart Tarayıcı	47,55%	99,17%
2-Arama Motoru	14,63%	0,01%
3-Metin Tabanlı Tarayıcı	37,82%	0,82%
Referer_Type:		
1-Ana Sayfadan	1,88%	7.62%
2-Portal Hizmetlerinden	20,32%	15,44%
3-Kurumsal Alt Alanlardan	4,02%	16,95%
4-Arama Motorundan	4,23%	24,44%
5-Diğer Yönlendiriciler	0,31%	1,25%
6-Yönlendirici Yok	69,24%	34,29%
User_Language_TR:		
0-Diğer	1,53%	2,47%
1-Türkçe	98,47%	97,53%
User_Location:		
0-Türkiye İçi	21,37%	19,61%
1-Kurum İçi (SAÜ)	77,56%	79,51%
2-Türkiye Dışı	1,07%	0,87%

Veriler, tek sayfa oturum sayılarının ve bu oturumlar tarafından üretilen sayfa gösterimlerinin, çok sayfa oturum sayılarına ve sayfa gösterimlerine kıyasla çok daha az olduğunu göstermektedir. Bu durum, çok sayfa oturumlarının web sitesinde çok daha fazla etkileşim ve içerik tüketimiyle ilişkili olduğunu ortaya koymaktadır. İncelenen verilere göre standart tarayıcılar, çok sayfa görüntüleyen oturumlar için yüksek bir kullanım oranına sahiptir. Bu da standart tarayıcı kullanıcılarının içeriğe daha fazla ilgi gösterdiğini ve site ile daha etkileşimli olduğunu göstermektedir.

Metin tabanlı tarayıcı ve arama motorlarından gelen ziyaretlerde ise hemen çıkma oranı daha yüksektir. Bu durum daha çok arama motoru indeksleme robotları ve web gezgini (crawler) adı verilen otomatik site inceleme yazılımları nedeniyle ortaya çıkmakta ve bu tür tarayıcıların çerez (cookie) desteği olmaması nedeniyle sistemde tek sayfa hareketi olarak gözükmelerine neden olmaktadır. Ancak yine de "menu" servisi gibi tek bir sayfada kullanıcı ihtiyacını karşılayan portal servislerinin varlığı ve portala ait herhangi bir sayfanın yanlışlıkla açılması durumunda hemen kapatılması gibi durumlar da bu grup içerisinde değerlendirilmelidir.

Yönlendirici olmadan gelen tek sayfa görüntüleyen oturumların yüksek oranı, kullanıcıların doğrudan URL girişi veya yer imlerini kullanarak siteye geldiğini göstermektedir. Arama motorlarından gelinen ve çok sayfa gezilen oturumların yüksek oranı, kullanıcıların yer imleri ya da doğrudan site adını yazarak portala ulaşmak yerine arama motoru üzerinden siteye ulaştıklarını ortaya koymaktadır. Bu durum ayrıca SEO çalışmalarının etkin olduğunu ve kullanıcıların aradıklarını bulduktan sonra sitede gezinmeye devam ettiklerini de göstermektedir. İstemci tarayıcısı dili açısından bakıldığında, beklendiği şekilde, web portalını çoğunlukla Türkçe konuşan kullanıcıların tercih ettiği görülmektedir. Bu, sitenin yerel dildeki içerik optimizasyonunun başarılı olduğunu ve yerel kullanıcıların ihtiyaçlarını karşıladığını göstermektedir. Konum dağılımı, çoğunlukla kurum içi ve ülke içi kullanıcıları işaret etmektedir, bu da sitenin coğrafi olarak yakın kullanıcılar için daha alakalı olduğunu ortaya koymaktadır.

Tek sayfa ve çok sayfa oturum verileri üzerinde yapılan derinlemesine analizler sonucu elde edilen chi-kare testi sonuçları, farklı kategorilerdeki dağılımlar arasındaki istatistiksel olarak anlamlı farklılıkları ortaya koymaktadır. Dört grupta incelenen istemci özelliklerinin (alan adlarıyla) hemen çıkma oranı üzerindeki etkisinin incelendiği test sonuçları Tablo 6.5'te verilmiştir.

Tablo 6.5. İstemci özelliklerinin hemen çıkma oranı üzerindeki etkileri.

İstemci Özellikleri	Chi^2	p-değeri	Serbestlik Derecesi
Browser_Type	68,19	0,00	2
Referer_Type	38,72	0,00	5
User_Language_TR	0,00	1,00	1
User_Location	0,12	0,94	2

Test sonuçlarına göre, tek sayfa ve çok sayfa oturumları arasındaki tarayıcı tipi ve referans türü dağılımında istatistiksel olarak anlamlı bir fark bulunduğu görülmektedir. Bu iki özelliğe ait yüksek chi-kare değerleri ve düşük (<0,05) p-değerleri, bu özelliklerin hedef değişkenle anlamlı bir ilişkileri olduğunu ve tek sayfa ile çok sayfa oturumları arasındaki farkın rastgele bir varyasyondan kaynaklanmadığını göstermektedir.

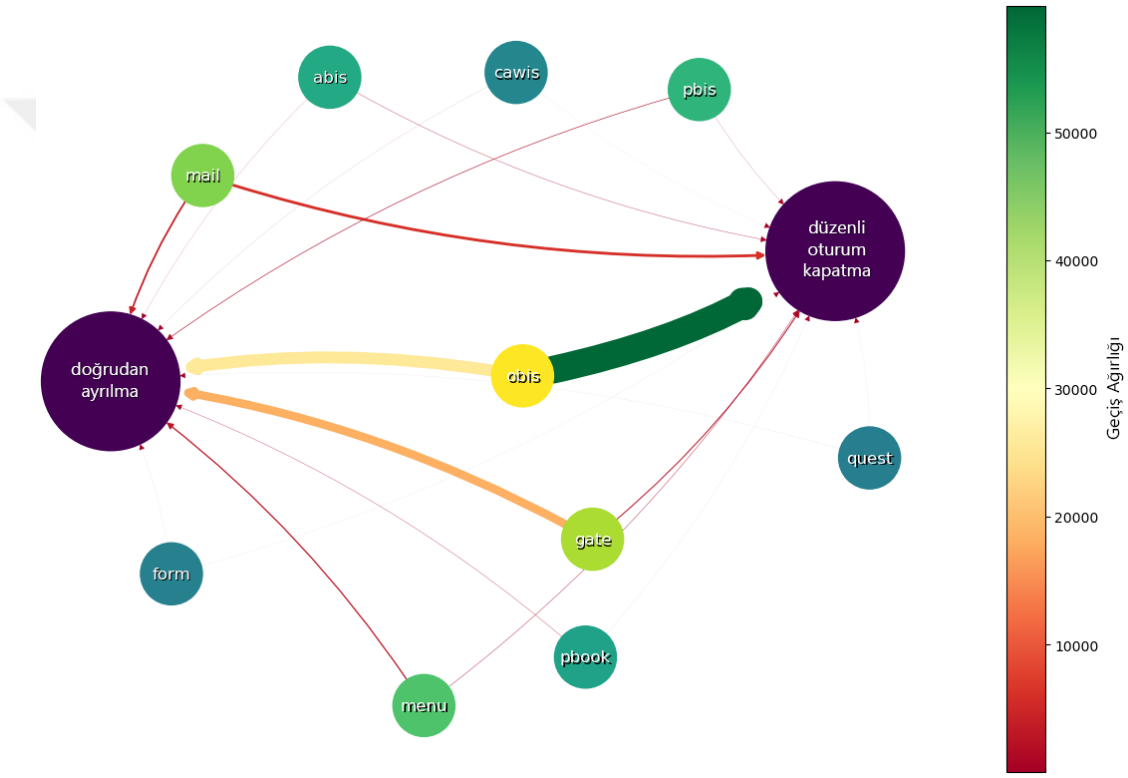
Diğer taraftan, kullanıcı dili (Türkçe) ve kullanıcı konumu özelliklerinin düşük chi-kare değerleri ve yüksek p-değerleri, bu özelliklerin hedef değişkenle belirgin bir ilişkisi olmadığını ve var olan küçük farkın rastgele bir varyasyondan kaynaklanabileceğini göstermektedir. Serbestlik derecesi ise test edilen özelliğin kategorik çeşitlilik sayısının bir eksiğini ifade eder ve bu anlamlılık değerlendirmesinde chi-kare ve p-değeriyle birlikte önemli bir rol oynar.

Hemen çıkma oranı, genellikle bir web sitesinin ilk izlenim kalitesi, içeriğin alakalılığı veya kullanıcı niyetiyle ilişkilendirilir. Yüksek bir hemen çıkma oranı çoğunlukla olumsuz bir göstergedir, ancak her zaman kötü bir kullanıcı deneyimini yansıtmayabilir. Örneğin, bir kullanıcı aradığı bilgiyi ilk sayfada bulursa (örneğin "menu" servisi) ve siteyi terk ederse, bu olumlu bir sonuç olarak değerlendirilebilir. Buna karşılık, eğer yüksek hemen çıkma oranı belirli sayfalarda veya hedef kitle segmentlerinde yoğunlaşıyorsa, bu durum kullanıcı deneyiminde veya içerik stratejisinde iyileştirme yapılması gerektiğine işaret edebilir.

Bu tür analizler, kullanıcıların web sitesi etkileşimlerini ve bu etkileşimlerin altında yatan nedenleri daha iyi anlama açısından önemlidir. Ayrıca, bu analizler sonucunda elde edilen bulgular, site tasarımı ve içerik stratejilerini iyileştirmek için somut adımlar atılmasına zemin hazırlar.

6.2.2. Sistemden ayrılma yöntemlerinin servis bazlı analizi

Bir web uygulamasına giriş yaptıktan sonra kullanıcıların işleri bittiğinde sistemi terk etmeleri için ideal olan "güvenli çıkış" yöntemi ile oturumu kapatmaları ve sistemi o şekilde terk etmeleridir. Ancak web'in doğası gereği, kendi bilinçleri dışında kullanıcıları güvenli çıkış yapmaya zorlayacak bir yöntem bulunmamaktadır. Sistemin güvenli çıkış yapılarak terk edilmemesi güvenlik sorunlarına neden olabileceği gibi analitik bilgilerinin de doğruluğu zayıflatmaktadır. Bir haftalık oturum verilerinin analizi sonucu oluşturulan Şekil 6.4, kullanıcıların web portalındaki farklı servislerden ayrılırken hangi yöntemleri tercih ettiğini gösteren bir ağ görselleştirmesi sunmaktadır.



Şekil 6.4. Portaldan ayrılma yöntemlerinin servis bazlı analizi.

Kullanıcıların portal içindeki davranışsal eğilimlerini ve çıkış stratejilerini tartışmak için temel olarak kullanılabilecek önemli bir veri setini temsil eden bu görsel, kullanıcı deneyimini geliştirmeye yönelik stratejik kararlara kritik katkılar sunabilir. Portaldan çıkış noktaları olarak güvenli çıkış butonu aracılığıyla "düzenli oturum kapatma" ve oturumu kapatmadan "doğrudan sistemden ayrılma" olmak üzere iki ana yöntem öne çıkmaktadır. Az sayıdaki oturum uyarı penceresi aracılığıyla yapılan işlemler düzenli oturum kapatma grubuna dahil edilmiştir. Tercih yoğunluğunu belirten mor renk ile temsil edilen her iki temel yöntemin yaklaşık aynı ağırlıkta kullanıldığı görülmektedir.

Ayrılma tercihleri portal servisleri bazında incelendiğinde “gate”, “obis”, “mail” ve “pbis” gibi servislerin çıkış noktası olarak kullanımının yoğunluğa bağlı olarak daha kalın ve renkli kenarlarla belirginleşmiştir. Bu durum kullanıcıların bu servislerden sistemi terk ederken genellikle güvenli çıkış yapmayı tercih ettiklerini ortaya koymaktadır. Servisler ile ayrılma yöntemleri arasındaki bağlantıların (kenarların) renkleri ve kalınlıkları, geçiş ağırlıklarını göstermektedir. Renk skalasındaki yeşilden kırmızıya değişim, geçiş sıklığının yoğunluğunu ifade etmektedir. Yapılan analizde, kullanıcıların “gate” servisinden sistemi terk ederken çoğunlukla pencereyi doğrudan kapatma ya da başka bir adrese gitme yolunu tercih ettiği, “obis” ve “mail” gibi kişiye özel bilgileri ihtiva eden servislerde ise güvenli çıkış yöntemiyle sistemi terk etmeye özen gösterdikleri görülmektedir. Özellikle “obis”in öğrencilere hizmet veren bir servis olduğu ve bu servisi kullanan öğrencilerin bir bölümünün kurumun ortak kullanım noktalarındaki bilgisayarlardan giriş yaptıkları düşünüldüğünde güvenli çıkış yöntemini tercih etmelerinin beklenen bir durum olduğunu söylemek yanlış olmayacaktır.

Kullanıcı davranışlarını iyi anlamak, web portalı tasarımını kullanıcı etkileşimine göre optimize etmek için değerli bilgiler sunar. Örneğin, güvenli çıkış seçeneğinin ağırlıklı kullanımı, kullanıcıların oturumlarını bilinçli bir şekilde sonlandırdıklarını ve bu işlevin kolay erişilebilir olması gerektiğini göstermektedir. Doğrudan ayrılma seçeneğinin de önemli olması, kullanıcıların oturumu kapatmadan portalı terk etme eğiliminde olduğunu ortaya koymaktadır. Bu da oturum zaman aşımı sürelerinin veya otomatik çıkış işlevlerinin incelenmesi gerektiğine işaret etmektedir.

6.2.3. Sistemden ayrılma yöntemlerinin tahmini

WUM tekniklerinin derinlemesine incelenmesi ve uygulanması, kullanıcı davranışlarının ve sistem etkileşimlerinin anlaşılmasında kritik bir rol oynamaktadır. Çalışmanın bu bölümünde kullanıcıların sistemi hangi yöntemle terk edeceklerini tahmin etmeye odaklanılmıştır. Model, çeşitli kullanıcı ve oturum bilgilerini kullanarak Rastgele Orman Sınıflandırıcı (Random Forest Classifier) ile eğitilmiştir. Modelin eğitiminde çeşitli oturum bilgileri (User_Type, Sex, Age, User_Language_TR, User_Location, Browser_Type, Landing_Srv_ID, Exit_Srv_ID, Session_Login_Status, Page_Count, Service_Count, Total_Session_Duration, Avg_Page_Duration, Total_Page_Load, p_gate, p_mail, p_obis, p_abis, p_pbis, p_menu) kullanılmıştır.

Toplam 1.220.916 satırdan oluşan bir aylık zenginleştirilmiş oturum verilerinin %70'i eğitim, %30'u test için kullanılmış ve oldukça başarılı tahmin sonuçları elde edilmiştir. Tablo 6.6, modelin her bir sınıf için hassasiyet (precision), hatırlama (recall) ve F1-puanı (F1-score) değerlerini, ayrıca her sınıfın destek (support) sayılarını göstermektedir. Ağırlıklandırılmış ortalama, modelin tüm sınıfların oranına göre hesaplanmış genel performansını temsil etmektedir.

Tablo 6.6. Çoklu sınıflandırma modelinin genel performans metrikleri.

Sınıf	Hassasiyet	Hatırlama	F1-Puanı	Destek
0 (doğrudan ayrılma)	0,96	0,91	0,93	163.187
1 (güvenli çıkış butonu)	0,92	0,96	0,94	192.866
2 (oturum bildirim penceresi)	0,77	0,86	0,81	10.222
Ağırlıklı Ortalama	0,93	0,93	0,93	366.275

Modelin performansı bireysel sınıfların tanımlanması ve genel tahminler açısından incelendiğinde, dikkate değer bir doğruluk oranı (accuracy) sağlamaktadır. Bu, modelin hem spesifik sınıfları ayırt etme yeteneğinin hem de genel veri seti üzerindeki tahmin başarısının göstergesidir. Sınıf 0 (doğrudan ayrılma) için elde edilen yüksek hassasiyet ve iyi bir hatırlama değeri, modelin bu sınıfı belirlemede oldukça etkili olduğunu işaret etmektedir. Sınıf 1 (güvenli çıkış butonu) için de benzer bir şekilde yüksek hassasiyet ve daha da yüksek hatırlama değerleri, modelin bu sınıfı tahmin etme konusunda oldukça başarılı olduğunu göstermektedir.

Her iki sınıf için F1-puanları, modelin bu sınıfları tahmin etmede dengeli bir performans gösterdiğine işaret etmektedir. Diğer yandan, sınıf 2 (süre aşımı bildirim penceresi) için hassasiyet ve hatırlama değerleri, diğer sınıflara göre daha düşüktür, ancak bu sınıf için elde edilen F1-puanı, modelin bu sınıfı tahmin etmede yeterli kabul edilebilecek bir performans ortaya koyduğunu göstermektedir. Bu sınıfın görece az olan örnek sayısının, modelin bu sınıfta elde ettiği sonuçların diğer sınıflara göre bir miktar daha düşük olmasına neden olduğunu söylemek mümkündür. Sonuç olarak, modelin genel performansının ağırlıklandırılmış ortalama F1-puanı üzerinden yapılan değerlendirmesi, tüm sınıflar arasında dengeli bir performansla birlikte yüksek doğruluk seviyeleri sunulduğunu ortaya koymaktadır.

6.2.4. Son terk edilecek servis tahmini

Çalışmanın bu bölümünde, kullanıcıların sistemi hangi servis üzerinden terk edeceklerine yönelik tahminlere odaklanılmıştır. Bu amaçla model, bir haftalık oturum bilgileri içeren kapsamlı bir veri setinden çeşitli kullanıcı ve oturum özellikleri kullanılarak Gradyan Artırma Sınıflandırıcısı (Gradient Boosting Classifier) ile eğitilmiştir. Modelin eğitiminde, öğrenme hızı (learning rate) olarak 0,1, maksimum derinlik (maximum depth) olarak 3 ve tahminci sayısı (number of predictors) olarak 100 değerleri tercih edilmiştir.

Tablo 6.7. Modelde kullanılan özellikler ve önem dereceleri.

Özellik	Anlam	Önem
Day_of_Week	Haftanın Günü	0,0182
Hour_of_Day	Günün Saati	0,0330
User_Type	Kullanıcı Tipi	0,0322
Sex	Cinsiyet	0,0190
Age	Yaş	0,0421
Browser_Type	Tarayıcı Türü	0,0326
Referer_Type	Referans Türü	0,0150
Landing_Srv_ID	İlk Gelinen Servis	0,0525
p_gate	Serviste gezilen sayfa sayısı	0,0969
p_mail	Serviste gezilen sayfa sayısı	0,1304
p_obis	Serviste gezilen sayfa sayısı	0,4033
p_abis	Serviste gezilen sayfa sayısı	0,0176
p_pbis	Serviste gezilen sayfa sayısı	0,0277
p_menu	Serviste gezilen sayfa sayısı	0,0595
p_pbook	Serviste gezilen sayfa sayısı	0,0109
p_form	Serviste gezilen sayfa sayısı	0,0029
p_quest	Serviste gezilen sayfa sayısı	0,0027
p_cawis	Serviste gezilen sayfa sayısı	0,0035

Modelde kullanılan özellikler ve bu özelliklerin modelin tahmin yeteneği üzerindeki önem dereceleri Tablo 6.7'de detaylı şekilde sunulmuştur. Modelin test sonuçlarında doğruluk 0.9557, hassasiyet 0.9561, hatırlama 0.9557 ve F1-puanı 0.9555 olarak elde edilmiştir. Bu sonuçlar, modelin kullanıcıların hangi servis üzerinden sistemi terk edeceklerini oldukça başarılı bir şekilde tahmin edebildiğini göstermektedir. Bu sonuçlar, web portal kullanımı üzerine yapılan detaylı analizlerin kullanıcı davranışlarını tahmin etme ve sistemin kullanımını optimize etme potansiyeline sahip olduğunu ortaya koymaktadır.

6.2.5. Belirli bir servise erişim tahmini

Portal etkileşimlerinin daha derinlemesine anlaşılmasını sağlamada WUM yöntemleri önemli bir rol oynar. Bir kullanıcının bir oturumda belirli bir servise girip girmeyeceği tahmin modelleri ile belirlenebilir. Çalışmanın bu bölümünde "mail" (web tabanlı e-posta) servisine erişim tahmini gerçekleştirmek için çeşitli oturum bilgileri (Log_Date_Time, Log_Date, User_Type, Sex, Age, Avg_Page_Duration, User_Language_TR, User_Location, Browser_Type, Referer_Type) dört farklı model ile eğitilmiştir. En iyi başarımı elde etmek amacıyla hiper parametre optimizasyonu yapılmış ve dört modelin en iyi parametreleri tespit edilmiştir. Tablo 6.8, sınıflandırma modellerinin amaçlanan tahmine yönelik performansını ve parametrik yapılandırmalarını karşılaştırmalı bir biçimde sunmaktadır.

Tablo 6.8. Çeşitli sınıflandırma modellerinin performans karşılaştırması.

Metrik / Model	Gradyan Arttırma	Rastgele Orman	Destek Vektör Makinesi	Lojistik Regresyon
En İyi Parametreler	Learn. Rate: 0,1, Max Depth: 3, N Estim.: 100	Max Depth: 10, Min Sam. Leaf: 2, Min Sam. Split: 5, N Estimators: 100	C: 10, Kernel: 'rbf'	C: 10
Ort. ÇD Puanı	0,9182	0,8947	0,9184	0,9143
Doğruluk	0,9252	0,9240	0,9238	0,9178
Hassasiyet	0,9234	0,9218	0,9214	0,9156
Geri Çağırma	0,9252	0,9240	0,9238	0,9178
F1-Puanı	0,9188	0,9176	0,9174	0,9094

Her bir modelin en iyi parametre kombinasyonlarını ve ortalama çapraz doğrulama (cross-validation) puanı, doğruluk, hassasiyet, geri çağırma ve F1 puanı gibi performans metriklerini yan yana sunan bu gösterim, modeller arası performans farklılıklarını karşılaştırmalı olarak ortaya koymaktadır. Model performans metrikleri incelendiğinde, dört modelin de yüksek doğruluk oranları ve dengeli F1-puanları sergilediği görülmektedir. Rastgele Orman (Random Forest) ve Gradyan Arttırma (Gradient Boosting) modellerinin hem doğruluk hem de F1-puanı açısından öne çıkması, bu iki modelin veri seti üzerinde daha iyi bir genelleme kabiliyetine sahip olduğunu göstermektedir. Rastgele Orman modelinin en iyi parametreleri olarak belirlenen değerlerin kombinasyonu, modelin karmaşıklığını ve öğrenme kabiliyetini dengeli bir şekilde yönettiğini göstermektedir.

Lojistik Regresyon modeli, en iyi C parametresi ile yüksek bir Ortalama Çapraz Doğrulama Puanı (Average Cross-Validation Score) sunarken, F1-Puanı açısından biraz daha düşük bir değer sergilemektedir. Bu, modelin belirli sınıflar arasında ayrım yapmada biraz daha zorlandığını gösterebilir. Destek Vektör Makinesi (SVM) modeli ise RBF çekirdeği (kernel) ve C parametresi ile yüksek bir genelleme kabiliyetine sahip olduğunu göstermektedir. Her iki model de sınıflandırma probleminde dengeli bir yaklaşım sergilemektedir.

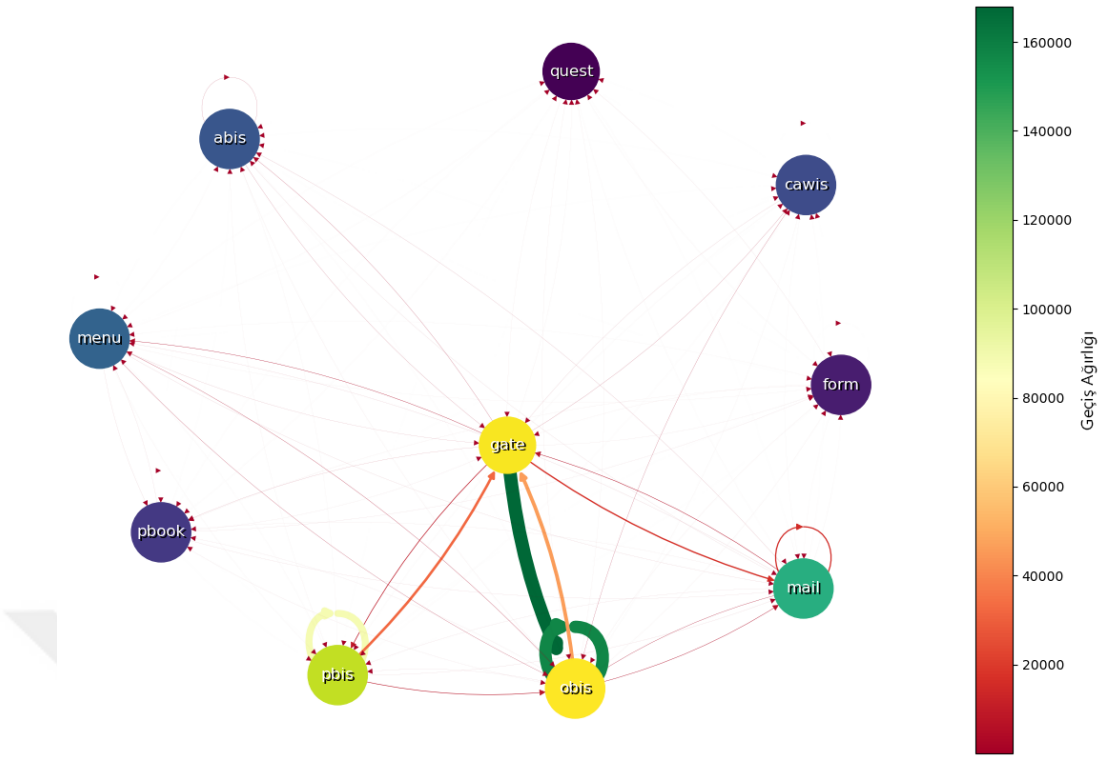
Gradyan Arttırma modelinin performansı, öğrenme oranı, maksimum derinlik ve tahminci sayısı parametrelerinin dikkatli bir şekilde ayarlanmasıyla daha da iyileştirilmiş görünmektedir. Bu model, yüksek doğruluk ve F1-Puanı ile veri setinin karmaşık yapılarını başarıyla yakalamıştır. Bu modellerin parametre ayarlarının model performansı üzerinde belirleyici bir rol oynadığı ve bu probleme özgü en uygun yapılandırmaların bu modeller tarafından sağlandığını söylemek mümkündür. Özetle, dört model de bu sınıflandırma probleminde etkili sonuçlar sergilemiş ve bazı özelliklerine bakılarak bir kullanıcının bir servise erişip erişmeyeceği oldukça yüksek bir başarım ile tahmin edilebilmiştir.

6.3. Sayfa Gösterimi Verilerine Uygulanan WUM Analizleri ve Tahminleri

Web sitelerindeki her bir sayfa gösterimi, kullanıcıların içerikle olan etkileşimlerinin ve tercihlerinin anlaşılması için değerli veri noktalarını içerir. CAWAL çerçevesi kullanılarak elde edilen ve sistemli bir biçimde işlenen sayfa gösterimi verileri, kullanıcı deneyiminin derinlemesine analiz edilmesi ve kişiselleştirilmiş kullanıcı yolculuklarının modellenmesi için zemin hazırlar. Bu alt bölümde, kullanıcıların web portalında nasıl bir gezinme modeli izlediklerinin, hangi içeriklerin dikkat çektiğinin ve kullanıcıların web sitesi üzerindeki davranışsal örüntülerinin incelenmesi üzerine yoğunlaşılacaktır.

6.3.1. Portal servis etkileşimlerinin ve geçiş yollarının analizi

WUM yaklaşımlarının etkin uygulanması, kullanıcıların web portalı içindeki hareket ve etkileşim örüntülerini anlamada önemli bir araç sağlar. Çalışmanın bu bölümünde portal servis etkileşimleri ve geçiş yollarının analizi, kullanıcıların navigasyon örüntülerini ve servisler arası etkileşimleri kapsamlı bir şekilde incelemek için kullanılmıştır. Şekil 6.5, web portalındaki servisler arasındaki kullanıcı geçişlerini temsil eden ağ görselleştirmesini sunmaktadır.



Şekil 6.5. Portal servisleri arasındaki geçiş yolları ve ağırlıkları.

Servisler arası geçişler ve ağırlıkları, kullanıcıların portal içindeki gezinme modellerini anlamak açısından son derece önemlidir. Bulgular, hangi servislerin en sık kullanıldığını ve kullanıcı aktivitesinin akışını analiz etmek, kullanıcı davranışlarının nedenlerine ve olası etkilerine dair hipotezler geliştirmek gibi amaçlarla kullanılabilir. Bu geçiş örüntülerinin portal tasarımı üzerindeki etkilerini incelemek, bu örüntüleri iyileştirme çalışmaları yapmak, kullanıcı deneyimini iyileştirmek için sık kullanılan yolları optimize etmek ve daha az kullanılan servislere erişimi kolaylaştırma stratejileri geliştirmek için de değerli bilgiler sağlar.

Servislerin üzerinde görünen kendi kendine bağlanan yay şekli ve kalınlığı, ilgili servisten sisteme giriş yapmak için "gate" servisine gidilip geri dönülmesini temsil etmektedir. Merkezi düğüm ("gate"), çok sayıda kenarla (geçiş yollarıyla) bağlantılıdır ve bu özelliği yoğun bir kullanıma işaret etmektedir. Bu, kullanıcıların portal içinde en sık etkileşimde bulunduğu ve sisteme giriş noktası olarak kullandığı servis olmasından kaynaklanmaktadır. Yoğun kullanılan "pbis", "obis" ve "mail" gibi servisler parlak düğümlere sahipken, diğer servislere geçiş ağırlıkları belirgin çizgilerle kendini göstermektedir. Diğer yandan, "cawis", "quest" ve "form" gibi koyu renkli düğümlere sahip ve geçişlerin daha az olduğu servisler, daha küçük bir kullanıcı grubu tarafından spesifik amaçlar için kullanılmakta olduklarını belirtmektedir.

6.3.2. Birliktelik kurallarının keşfi

Çalışmanın bu bölümünde kullanıcı davranışlarının derinlemesine analizi için birliktelik (ilişkilendirme) kuralları madenciliği uygulanmıştır. Analiz, bir aylık sayfa gösterimi içeren "va_page5.csv" veri seti kullanılarak gerçekleştirilmiştir. İlk olarak, Service_ID sütunu anlamlı İngilizce isimlerle eşleştirilmiş, böylece her servisin kolayca tanınması sağlanmıştır. Daha sonra, kullanıcı türü, yaş grubu, konum, tarayıcı tipi ve referans tipi gibi özellikler seçilmiş ve bu özelliklere göre kullanıcı oturumlarındaki servis erişim sıraları tespit edilmiştir.

Veri seti, işlem kodlayıcı (transaction encoder) yardımıyla işlem listelerine dönüştürülmüş ve tek seferlik kodlama (one-hot encoding) uygulanarak makine öğrenimi algoritmaları tarafından işlenebilir hale getirilmiş, ardından Apriori algoritması kullanılarak en az %25 destek değerine sahip sık öğe setleri bulunmuştur. Bu öğe setlerinden en az %98 güven eşiğine ve 1'den büyük veya eşit kaldırma değeri (lift) değerine sahip ilişkilendirme kuralları türetilmiş ve kriterlere uyan ilk otuz kural Tablo 6.9'da sunulmuştur.

Burada belirtilen kurallar, "abis", "cawis", "form" gibi önce gelen servisler ile "gate", "mail" gibi sonraki servisler arasındaki kuvvetli ilişkileri açığa çıkarmaktadır. Kullanıcıların bir servisten diğerine geçiş eğilimlerini anlamak için, önce gelen destek, sonraki destek, destek, güven, kaldırma değeri (lift), kaldıraç etkisi (leverage), ikna gücü (convince) ve Zhang ölçütü gibi metrikler detaylı bir şekilde incelenmiştir. Örneğin, "abis" servisini ziyaret eden kullanıcıların büyük bir çoğunluğunun daha sonra "gate" servisine gittiği (%99,44 güven ile) ve bu kuralın rastgele beklenenden %3,43 daha sık görüldüğü (1,0343 kaldırma değeri) tespit edilmiştir. Diğer taraftan, "form" ve "mail" arasındaki ilişki de yüksek bir kaldırma değeri (1,1784) ile dikkat çekici bir bağlantı sergilemektedir.

Bu kuralların yüksek ikna gücü değerleri, önce gelen olmadan sonrakinin oluşma olasılığının son derece düşük olduğunu göstermekte olup, özellikle "form" ve "cawis" için sonsuz ikna gücü değerleri bu durumu vurgulamaktadır. Zhang ölçütü değerleri ise bu ilişkilerin güvenilirliğini ve yönünü değerlendirmede kullanılır. Genel olarak, bu kuralların yüksek destek, güven ve kaldıraç değerleri, kullanıcıların portal içindeki gezinme modellerini anlamak ve portalın kullanıcı deneyimini iyileştirmek için önemli bilgiler sunmaktadır.

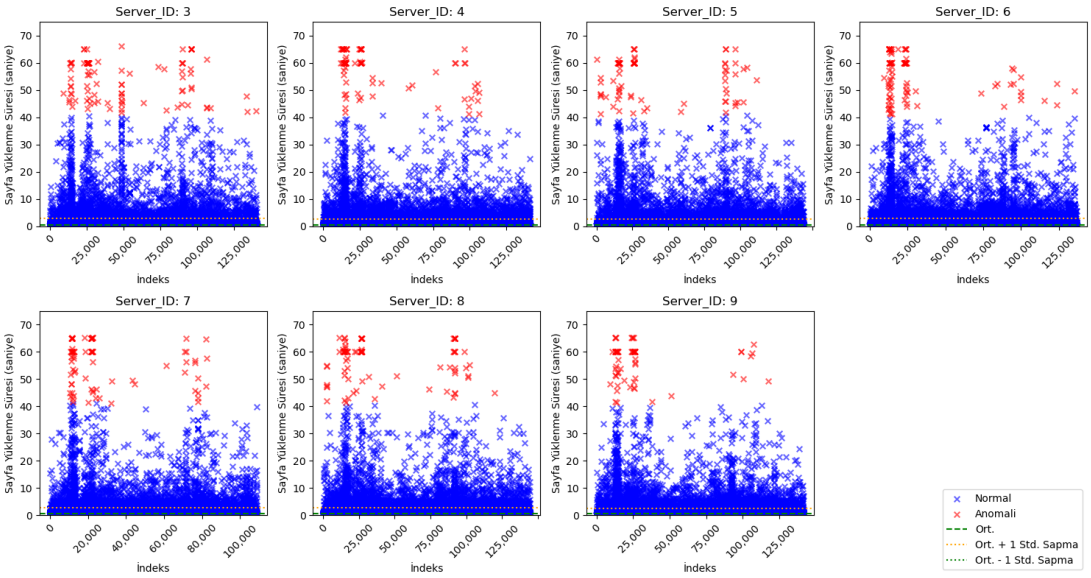
Tablo 6.9. Belirlenmiş kriterlere uyan ilk 30 ilişkilendirme kuralı.

Öncüller	Sonuçlar	Öncül Destek	Sonuç Destek	Destek	Güven	Kaldırma Kuv. (Lift)	Kaldıraç E. (Leverage)	İkna Gücü (Convince)	Zhangs Metriği
{'abis'}	{'gate'}	0,4562	0,9613	0,4536	0,9944	1,0343	0,0151	6,8428	0,0610
{'cawis'}	{'gate'}	0,4046	0,9613	0,4046	1,0000	1,0402	0,0156	Sonsuz	0,0649
{'form'}	{'gate'}	0,3763	0,9613	0,3763	1,0000	1,0402	0,0145	Sonsuz	0,0620
{'form'}	{'mail'}	0,3763	0,8428	0,3737	0,9932	1,1784	0,0566	22,9536	0,2427
{'menu'}	{'gate'}	0,4742	0,9613	0,4742	1,0000	1,0402	0,0183	Sonsuz	0,0735
{'pbook'}	{'gate'}	0,4330	0,9613	0,4330	1,0000	1,0402	0,0167	Sonsuz	0,0682
{'quest'}	{'gate'}	0,3222	0,9613	0,3222	1,0000	1,0402	0,0125	Sonsuz	0,0570
{'menu'}	{'mail'}	0,4742	0,8428	0,4665	0,9837	1,1672	0,0668	9,6426	0,2725
{'pbook'}	{'mail'}	0,4330	0,8428	0,4304	0,9940	1,1795	0,0655	26,4124	0,2684
{'quest'}	{'mail'}	0,3222	0,8428	0,3222	1,0000	1,1865	0,0506	Sonsuz	0,2319
{'cawis', 'abis'}	{'gate'}	0,3015	0,9613	0,3015	1,0000	1,0402	0,0117	Sonsuz	0,0554
{'cawis', 'abis'}	{'mail'}	0,3015	0,8428	0,2990	0,9915	1,1764	0,0448	18,3943	0,2147
{'form', 'abis'}	{'gate'}	0,2758	0,9613	0,2758	1,0000	1,0402	0,0107	Sonsuz	0,0534
{'form', 'abis'}	{'mail'}	0,2758	0,8428	0,2758	1,0000	1,1865	0,0434	Sonsuz	0,2171
{'abis', 'gate'}	{'mail'}	0,4536	0,8428	0,4459	0,9830	1,1663	0,0636	9,2234	0,2610
{'mail', 'abis'}	{'gate'}	0,4459	0,9613	0,4459	1,0000	1,0402	0,0172	Sonsuz	0,0698
{'abis', 'menu'}	{'gate'}	0,3299	0,9613	0,3299	1,0000	1,0402	0,0128	Sonsuz	0,0577
{'obis', 'abis'}	{'gate'}	0,3222	0,9613	0,3222	1,0000	1,0402	0,0125	Sonsuz	0,0570
{'pbis', 'abis'}	{'gate'}	0,3995	0,9613	0,3969	0,9935	1,0335	0,0129	5,9923	0,0540
{'pbook', 'abis'}	{'gate'}	0,3273	0,9613	0,3273	1,0000	1,0402	0,0127	Sonsuz	0,0575
{'abis', 'menu'}	{'mail'}	0,3299	0,8428	0,3299	1,0000	1,1865	0,0519	Sonsuz	0,2346
{'obis', 'abis'}	{'mail'}	0,3222	0,8428	0,3196	0,9920	1,1771	0,0481	19,6521	0,2219
{'pbis', 'abis'}	{'mail'}	0,3995	0,8428	0,3943	0,9871	1,1712	0,0577	12,1843	0,2435
{'pbook', 'abis'}	{'mail'}	0,3273	0,8428	0,3273	1,0000	1,1865	0,0515	Sonsuz	0,2337
{'form', 'cawis'}	{'gate'}	0,2809	0,9613	0,2809	1,0000	1,0402	0,0109	Sonsuz	0,0538
{'form', 'cawis'}	{'mail'}	0,2809	0,8428	0,2809	1,0000	1,1865	0,0442	Sonsuz	0,2186
{'form', 'cawis'}	{'menu'}	0,2809	0,4742	0,2758	0,9817	2,0700	0,1425	28,6546	0,7189
{'mail', 'cawis'}	{'gate'}	0,3866	0,9613	0,3866	1,0000	1,0402	0,0149	Sonsuz	0,0630
{'cawis', 'menu'}	{'gate'}	0,3402	0,9613	0,3402	1,0000	1,0402	0,0132	Sonsuz	0,0586
{'obis', 'cawis'}	{'gate'}	0,3428	0,9613	0,3428	1,0000	1,0402	0,0133	Sonsuz	0,0588

6.3.3. Sunucu ve sayfa yüklenme süresi bazlı anomali tespiti

Web çiftliği (web farm) yapısında sunucu performansı ve sayfa yüklenme sürelerinin sistematik analizi, olası anomalilerin ve uygulamadaki sorunların tespit edilmesinde hayati bir öneme sahiptir. Bu bölümde çoklu sunucu mimarisinde yük dengeleme ile dağıtılan isteklerin 24 saatlik sayfa gösterimi verileri ve yüklenme süreleri analiz edilmiştir.

İzolasyon Ormanı algoritması kullanılarak eğitilen model ile gerçekleştirilen analizin sonuçları Şekil 6.6'da sunulmuştur. Her bir alt grafiğin farklı bir sunucudaki (Server_ID) sayfa yüklenme sürelerinin dağılımını ve anomali durumlarını ortaya koyan görsel, sunucular arasındaki performans farklılıklarını izlemek ve potansiyel problemleri tanımlamak açısından kritik bilgiler sağlamaktadır.



Şekil 6.6. Farklı sunucularda sayfa yükleme süreleri üzerinden anomali tespiti.

Alt grafiklerin X ekseninde yer alan "İndeks" değeri ilgili sunucuda gözlemlenen sayfa gösterimlerinin sayısını zaman serisi biçiminde ifade etmektedir. Grafiklerde mavi renkli x işaretleri normal gözlemleri ifade ederken kırmızıyla işaretlenmiş anomaliler, yüklenme süresinin beklenenden çok daha uzun olduğu durumları göstermektedir. Sistemin ve sunucuların yoğun olduğu hafta içi bir güne ait veriler üzerinde yapılan analizde çok sayıda isteğin sayfa yüklenme (üretilme) süresinin yüksek olduğu görülmektedir. Her sunucuda farklı sayıda anomali gözlemlenmekle birlikte dengeli bir dağılım izlediği söylenebilir.

Anomalilerin dağılımı ve sıklığı, sunucu kaynaklarının yetersizliği, yazılım kalitesindeki sorunlar, ağ problemleri veya yüksek trafik gibi faktörlere ışık tutmaktadır. Bu örnek veri setinde gözlenen gecikmeler üzerinde yapılan incelemede e-posta ve veritabanı sunucularına yönelik isteklerin cevap sürelerinin trafik yoğunluğu ve kaynak darboğazları nedeniyle uzadığını görülmüştür. Bu tür analizler, kaynak planlaması, yük dengeleme ve altyapı geliştirmeleri için kritik içgörüler sağlayarak, sistem performansını optimize etmek ve gelecekteki ihtiyaçlar için proaktif çözümler geliştirmek açısından stratejik bir öneme sahiptir.

7. SONUÇ VE ÖNERİLER

Uygulama günlükleri ile web analitiğini bütünleştiren CAWAL modelinin özgün veri toplama yöntemi ve kullandığı uygulama tabanlı API aracılığıyla elde edilen yüksek doğruluğa sahip veriler, geleneksel yöntemlerle ulaşılamayan ölçüde detaylı bilgiler elde edilmesini sağlamıştır. Her bir sayfa gösteriminde yaklaşık 1KB ve gün bazında toplanan analitik verilerin veritabanındaki 50 farklı alanda toplamda yaklaşık 68KB veri üretmesi, modelin detay seviyesi ve veri toplama kapasitesinin somut bir göstergesidir. Model, kullanıcı etkileşimlerinin detaylı ve kapsamlı bir şekilde incelenmesine zemin hazırlayarak zenginleştirilmiş veri setleri üzerinden daha keskin analizler gerçekleştirilmesini mümkün kılmıştır. Bu sayede, kullanıcı gezinme modellerinden tercih edilen içeriklere, etkileşim sürelerinden uygulama mesajlarına kadar geniş bir yelpazede analizler yapılabilmesine ve kullanıcı etkileşimlerinin derinlemesine incelenmesine olanak tanımıştır.

Ölçeklenebilirlik ve uygulama esnekliği, modelin tasarımında öne çıkan diğer temel özellikleridir. CAWAL, farklı büyüklüklerdeki kurumsal web portalları ve uygulamaları için uygun bir çözüm sunarak, geniş bir kullanım alanına sahip olma potansiyelini göstermiştir. Ölçeklenebilirlik açısından artan veri hacmi ve kullanıcı etkileşimlerini etkin bir şekilde işleyecek şekilde tasarlanan model, web uygulamaları ve sunucularının büyüme ve değişim ihtiyaçlarına dinamik olarak uyum göstermesini sağlar. Modelin API tabanlı mimarisi, geliştiricilere veri toplama süreçlerini kendi uygulama mantıklarına ve iş akışlarına uyacak şekilde özelleştirme esnekliği sunar. Bu esneklik, modelin geniş bir yelpazedeki uygulama senaryolarında kullanılabilmesini sağlar ve uygulamaların belirli gereksinimlere göre düzenlenmesine olanak tanır.

CAWAL, bu çalışmada bir üniversite web portalında uygulanmış olsa da farklı türdeki web uygulamalarına da adapte edilebileceğini göstermiştir. Çerçevenin uygulamadaki başarısı, bütünleştirildiği web uygulamalarından elde edilen detaylı kullanıcı katılımı verileri ile kanıtlanan güçlü ve etkin veri sağlama yeteneği üzerinden değerlendirilmektedir. Özellikle, kullanıcıların gezinme yolları, sayfa görüntüleme süreleri ve tercih edilen içerikler gibi davranışsal metriklerin detaylı kaydı, portal

deneyimini optimize etme ve içerik stratejilerini geliştirme çabalarında kritik bilgiler sunmuştur. Uygulama senaryolarında CAWAL'ın gerçek zamanlı veri işleme kapasitesi, web site yöneticilerine ve içerik üreticilerine anlık kullanıcı davranışları üzerine aksiyon alabilme imkânı vermiş, bu da kullanıcı memnuniyetini ve site etkileşimini arttıran hızlı iyileştirmeler yapılmasına olanak tanımıştır. Çerçeve aracılığıyla elde edilen veriler üzerinde yapılan analizler, farklı kullanıcı türlerinin web sitesi ile etkileşim sürelerinde belirgin farklılıklar olduğunu göstermiştir. CAWAL, özellikle öğrenciler, akademik ve idari personel gibi çeşitli kullanıcı türleri arasındaki davranışsal farklılıkları aydınlatmıştır.

Teknik ve demografik özellikler, etkileşim süreleri ve tercih edilen içerikler hakkında yapılan analizler, web portalının tasarımı ve işlevselliğinin kullanıcı ihtiyaçlarına ve tercihlerine göre nasıl optimize edilebileceği konusunda zengin bilgi kaynakları sağlamıştır. Öğrenciler için önemli olan sınav programları veya ders materyalleri gibi bölümlerin daha erişilebilir hale getirilmesi, akademik ve idari personel için ise araştırma ve idari işlemlerle ilgili bilgilerin daha fazla vurgulanması gibi iyileştirmeler bunlara örnek olarak verilebilir. Bu tür optimizasyonlar, web sitesinin farklı kullanıcı türlerinin ihtiyaçlarını daha etkin bir şekilde karşılamasına ve genel kullanıcı memnuniyetini artırmasına yardımcı olacaktır. CAWAL çerçevesi ile toplanan verilerin analizi, kullanıcıların web sitesindeki belirli hedeflere ulaşma sürecinde karşılaştıkları zorlukları da aydınlatmaktadır. Örneğin, bir hizmetin kullanımı sırasında kullanıcıların çoğunun belirli bir adımda süreci terk etmesi, bu adımda bir kullanılabilirlik sorunu veya kullanıcılar için bir engel olabileceğini gösterir. Toplanan bu tür detaylı ve bağlamsal veriler, portal yöneticilerine ve uygulama geliştiricilere hedef odaklı kararlar almak için gereken somut ve karşılaştırılabilir ölçüm bilgileri sağlamaktadır.

Uzun dönemli kullanımda CAWAL'ın büyük veri hacimlerini işleme kapasitesinin yüksek olduğu ve kullanıcı trafiğinin yoğun olduğu dönemlerde dahi veri akışını sorunsuz bir şekilde yürütebildiği gözlenmiştir. İstemci tarafı performans testleri, çerçevenin sunucu tarafı çalışması nedeniyle diğer birçok analitik aracın aksine kullanıcı deneyimini olumsuz etkileyen gecikmeler yaratmadığını ortaya koymuştur. Yapılan sunucu tarafı yük testlerinde ise çerçeve, Open Web Analytics (OWA) ve Matomo gibi tanınmış açık kaynaklı araçlarla karşılaştırılmıştır. Bu testlerde CAWAL, yaklaşık %24 daha düşük yanıt süresi ile OWA'yı ve %85 daha düşük yanıt süresi ile

Matomo'yu geride bırakarak üstün performans göstermiştir. Ayrıca, çerçevenin sistem kaynakları üzerindeki etkisinin karşılaştırılan diğer açık kaynaklı araçlara göre daha az olduğu gözlenmiştir. Performans testleri, çerçevenin web sunucuları ve veritabanları üzerinde aşırı yük oluşturmadan, etkin bir şekilde çalışabildiğini ortaya koymuştur. Bu durum, web sitelerinin ve uygulamaların mevcut altyapılarını önemli ölçüde değiştirmeye gerek kalmadan çerçeveyi entegre edebilecekleri anlamına gelmektedir. Çerçevenin veri toplama API'si, yüksek trafik altında kullanıcı isteklerini kaydetme ve bu verileri anlamlı bilgilere dönüştürme konusunda üstün performans sergilemiştir. Bu yetenekler, büyük kuruluşlar için CAWAL'ı sadece doğru ve detaylı bilgi sağlayan bir araç değil, aynı zamanda yüksek performanslı ve ölçeklenebilir bir çözüm olarak cazip bir seçenek haline getirmektedir.

Performans bakımından gösterdiği üstün yanıt sürelerinin yanında CAWAL, veri kalitesi, oturum yönetimi ve kullanıcı takibi konularında da öne çıkmaktadır. Model, özellikle karmaşık kullanıcı davranışlarının incelenmesi ve kullanıcı etkileşimlerinin doğru bir şekilde yorumlanması açısından kritik öneme sahip olan yüksek veri doğruluğu sağlayarak analizlerin güvenilirliğini artırmaktadır. CAWAL, başarılı oturum yönetimi sayesinde kullanıcıların portalda gezinmelerini farklı sunucular ve alan adları boyunca sorunsuz izleyerek çok sunuculu ortamlardaki etkileşimleri verimli bir biçimde takip edebilmiştir. Bu kabiliyet, kullanıcı deneyiminin sürekli olarak izlenmesi ve iyileştirilmesi için gerekli zengin veri setlerinin oluşturulmasına olanak tanımıştır. Kullanıcı takibi açısından ise çerçevenin sağladığı detaylı kullanıcı profilleri ve davranış modelleri, özelleştirilmiş içerik sunumu ve hedefli pazarlama stratejilerinin geliştirilmesinde temel bir rol oynar. Bu yetenekler, portal tasarımcılarına, web geliştiricilere ve içerik üreticilerine kullanıcı ihtiyaçlarını ve beklentilerini daha iyi karşılayabilmek için gereken kapsamlı bilgileri sunar.

Veri sahipliği ve gizlilik özellikleri ile yapılandırılmış biçimde sakladığı verileri farklı amaçlar doğrultusunda kullandırma yeteneği, modelin özellikle bulut tabanlı yöntemlere karşı öne çıkan kabiliyetleri arasındadır. Bu yetenekleri sayesinde CAWAL'ın web kullanım madenciliği alanında yaptığı dönüşüm ise ayrı ve özel bir etkiye sahiptir. Sunucu günlüklerine dayalı geleneksel yöntemlerin yerine bu modelin kullanımı, zenginleştirilmiş veri setleri ile yüksek veri doğruluğu, hızlandırılmış madencilik süreci ve kullanıcı davranışlarının çok daha detaylı analizi gibi avantajlar sunmaktadır. Modelin makine öğrenimi ve veri madenciliği teknikleriyle uyumlu

veriler sağlayarak sunucu günlüklerine dayalı WUM ön işleme aşamasını ortadan kaldırması, madencilik sürecinin daha hızlı ve doğru şekilde sonuçlanmasını mümkün kılmıştır. Model, web analitiği ve web kullanım madenciliği alanlarına yenilikçi bir bakış açısı kazandırarak derin ve somut katkılarda bulunmuştur. Dönüştürücü etkileri dikkate alındığında CAWAL, bu disiplinlerde gelecek çalışmalara yön verecek yeni yollar açmıştır.

7.1. Modelin Sınırlamaları ve Zorlukları

CAWAL modeli web analitiği ve web kullanım madenciliği alanlarında önemli avantajlar sağlasa da uygulama bazı sınırlamalar ve zorluklar barındırmaktadır. Modelin web uygulamasıyla entegrasyonu için teknik altyapı ve uzmanlık gereksinimi, özellikle kaynakları sınırlı olan küçük ve orta ölçekli kuruluşlar için önemli bir engel oluşturmaktadır. Bu durum, yöntemin geniş çaplı kabulünü ve uygulanabilirliğini kısıtlayan önemli bir faktör olarak ortaya çıkmaktadır. Buna karşın, modelin kurumsal ortamlar için geliştirildiği ve büyük kuruluşların kullanımına daha uygun olduğu çalışmanın çeşitli bölümlerinde ifade edilmiştir. Avantajlarından dolayı bu modeli tercih edecek büyük kuruluşlar geniş kaynaklara sahip olmalarından dolayı bu tür zorlukları aşma kapasitesine sahiptirler. Dolayısıyla, modelin kurumsal düzeyde adaptasyonu ve entegrasyonu, mevcut kaynakların ve organizasyonel yapıların stratejik kullanımıyla doğrudan ilişkilidir.

Kullanıcı verilerinin toplanması ve işlenmesi sırasında ortaya çıkan veri gizliliği ve güvenlik endişeleri, CAWAL'ın uygulanmasını zorlaştırabilecek bir başka önemli faktördür. Modelin yasal düzenlemelere ve veri koruma standartlarına uyum mecburiyeti, mevzuattaki değişikliklere çabuk adapte olma gerekliliği ve kullanıcı bilgilerinin güvenli işlenmesi ve saklanması gibi ek zorlukları beraberinde getirir. Bu zorlukları aşmak, teknik altyapının güçlendirilmesi, süreçlerin gözden geçirilmesi ve hukuki danışmanlık alınması gibi ek kaynak yatırımlar gerektirerek modelin etkin bir şekilde uygulanması için harcanacak çabayı arttıracaktır.

Modelin mevcut web altyapıları ve sistemlerle entegrasyonu da teknik zorluklar ve uyumluluk sorunları doğurmakta, bu da modelin farklı web portallarına ve kullanıcı senaryolarına uygun şekilde özelleştirilmesini zorunlu kılmaktadır. Platform bağımlılığına sahip modelin kullanılacağı ortama yüksek düzeyde uyum sağlaması için çerçevenin web portalıyla aynı dilde geliştirilmesi ve aynı veritabanı yönetim sistemini

kullanması gibi faktörler dikkate alınmalıdır. Diğer yandan CAWAL modeli, veri hacmi ve çeşitliliği açısından sınırlamalara sahip olmamakla birlikte, uygulandığı web portalının kendine has gereksinimlerine uyum sağlamak amacıyla sıkı bir adaptasyon gerektirir. Özellikle hızla büyüyen ve sürekli değişen web veri setleri karşısında, modelin verimliliğini ve doğruluğunu koruması zorlaşabilir, bu da veri işleme ve analiz süreçlerinde verimlilik sorunlarına yol açabilir. Ayrıca, yaşayan her sistemde olduğu gibi düzenli bakım, teknik altyapının güncellenmesi ve veri güvenliği standartlarının titizlikle uygulanması modelin sürdürülebilirliği ve güvenilirliği açısından büyük önem taşır.

7.2. Gelecek Çalışmalar İçin Öneriler

Modelin uygulama ve analiz süreçleri gelecekteki araştırmalara yön verecek önemli bir potansiyel barındırmaktadır. CAWAL'ın akademik ve kurumsal web portallarının ötesinde farklı sektörlerde nasıl uygulanabileceğinin incelenmesi araştırma yelpazesini genişletecek ve modelin kullanıcı davranışlarına ve etkileşim senaryolarına adaptasyonunu test etme fırsatı sunacaktır. E-ticaret, sağlık hizmetleri ve kamu hizmetleri gibi alanlarda modelin uygulamasını değerlendirmek kullanıcı deneyimini ve etkileşimini anlamak adına yeni perspektifler kazandıracaktır. CAWAL'ın mevcut sınırlamalarına ve karşılaşılan zorluklara yönelik detaylı çalışmalar, modelin uygulanabilirliğini ve etkinliğini arttıracaktır. Teknik ve operasyonel zorlukları ele alan araştırmalar, modelin ölçeklenebilirliğini, esnekliğini ve teknolojik altyapılara adaptasyonunu iyileştirebilir. Bu çabalar, modelin geniş bir kullanım alanına entegre edilmesini kolaylaştırarak teknik engelleri aşma yönünde pratik çözümler sunabilir.

Modelin veri analizi ve işleme kapasitelerinin geliştirilmesi üzerine odaklanan çalışmalar daha karmaşık kullanıcı etkileşim desenlerinin keşfedilmesine ve kullanıcı davranışlarının daha detaylı analizine olanak tanıyacaktır. Yapay zekâ, makine öğrenmesi ve büyük veri analizi tekniklerinin modelle entegrasyonu, analitik ve bilgi çıkarım yeteneklerin genişletilmesine katkı sağlayacaktır. Modelin uzun vadeli etkileri ve sürdürülebilirliğine dair farklı kullanım senaryoları ve web altyapılarına adaptasyonu üzerine yapılacak araştırmalar ise CAWAL'ın etkisini ve uygulanabilirliğini daha da derinleştirecek ve yeni kullanım alanlarına kapı aralayacaktır.

KAYNAKLAR

Aartsen, B., El-Gayar, O. F. ve Noteboom, C. (2020). A systematic review of web usage mining techniques and future research options. *MWAIS 2020 Proceedings* (ss. 1–6) içinde. MWAIS volume 25.

Abdalla, A., Ahmed, T. ve Seliaman, M. (2016). Web usage mining and the challenge of big data: A review of emerging tools and techniques. I. R. M. Association (Ed.), *Big Data: Concepts, Methodologies, Tools, and Applications* chapter 42. (p. 899–928) içinde. IGI Global volume 6. https://doi.org/10.4018/978-1-4666-9840-6.ch042.

Abilio, R., Garcia, C. M., & Fernandes, V. (2021). Data mining applied on web robots detection: A systematic mapping. *Anais do 15º Congresso Brasileiro de Inteligência Computacional*. https://doi.org/10.21528/cbic2021-60.

Ahadi, A., Singh, A., Bower, M. ve Garrett, M. (2022). Text mining in education—a bibliometrics-based systematic review. *Education Sciences*, *12*, 210. https://doi.org/10.3390/educsci12030210.

Alby, T. (2023). Popular, but hardly used: Has google analytics been to the detriment of web analytics? In *Proceedings of the 15th ACM Web Science Conference 2023* (ss. 304-311). ACM. https://doi.org/10.1145/3578503.3583601.

Alhlou, F., Asif, S. ve Fettman, E. (2016). *Google Analytics Breakthrough* volume 1. Wiley. https://doi.org/10.1002/9781119266365.

Almatrafi, A. M. ve Alharbi, Z. H. (2023). The impact of web analytics tools on the performance of small and medium enterprises. *Engineering, Technology and Applied Science Research*, *13*, 11753–11762. https://doi.org/10.48084/etasr.6261.

Al-Otaibi, S., Alnassar, A., Alshahrani, A., Al-Mubarak, A., Albugami, S., Almutiri, N. ve Albugami, A. (2018). Customer satisfaction measurement using sentiment analysis. *International Journal of Advanced Computer Science and Applications*, *9*. https://doi.org/10.14569/ijacsa.2018.090216.

Al-Rahman, S. Q. A., Hasan, E. H. ve Sagheer, A. M. (2023). Design and implementation of the web (extract, transform, load) process in data warehouse application. *IAES International Journal of Artificial Intelligence (IJ-AI)*, *12*, 765. https://doi.org/10.11591/ijai.v12.i2.pp765-775.

Al-Zagheer, H., Hassan, M. ve AlDa'jah, A. (2022). Investigating the Effects of Organizational E-portals on Customer Satisfaction. *Journal of Southwest Jiaotong University*. https://doi.org/10.35741/issn.0258-2724.57.6.54.

Alya, R. ve Ikhwan, A. (2022). Mobile-based design of information system for vocational internship activities for vocational students. *Sinkron*, *7*, 2361–2368. https://doi.org/10.33395/sinkron.v7i4.11764.

Amalina, F., Hashem, I. A. T., Azizul, Z. H., Fong, A. T., Firdaus, A., Imran, M. ve Anuar, N. B. (2020). Blending big data analytics: Review on challenges and a recent study. *IEEE Access*, *8*, 3629–3645. https://doi.org/10.1109/access.2019.2923270.

Armaini, I., Dar, M. H. ve Bangun, B. (2022). Evaluation of labuhanbatu regency government website based on performance variables. *Sinkron*, *7*, 760–766. https://doi.org/10.33395/sinkron.v7i2.11404.

Bayir, M. A. ve Toroslu, I. H. (2022). Maximal paths recipe for constructing web user sessions. *World Wide Web*, (ss. 1–31). https://doi.org/10.1007/s11280-022-01024-3.

Bielak, K., Borek, B. ve Plechawska-Wójcik, M. (2022). Web application performance analysis using angular, react and vue.js frameworks. *Journal of Computer Sciences Institute*, *23*, 77–83. https://doi.org/10.35784/jcsi.2827.

Binns, R. (2017). Data protection impact assessments: a meta-regulatory approach. *International Data Privacy Law*, *7*, 22–35. https://doi.org/10.1093/idpl/ipw027.

Biørn-Hansen, A., Majchrzak, T. A. ve Grønli, T. M. (2017). Progressive web apps: The possible web-native unifier for mobile development. *Proceedings of the 13th International Conference on Web Information Systems and Technologies* (ss. 344–351) içinde. SCITEPRESS-Science and Technology Publications. https://doi.org/10.5220/0006353703440351.

Boufenneche, W., Hebboul, M. ve Benabderrahmane, O. (2022). Web analytics tools for e-commerce: An overview and comparative analysis. *International Conference on Managing Business Through Web Analytics* (ss. 51–72) içinde. Springer International Publishing. https://doi.org/10.1007/978-3-031-06971-0_5.

Brahmia, Z., Brahmia, S., Grandi, F. ve Bouaziz, R. (2022). JUpdate: A JSON update language. *Electronics*, *11*, 508. https://doi.org/10.3390/ electronics11040508.

Cahyanto, K. A., Hilmi, M. A. A. ve Mustamiin, M. (2022). Pengujian rule-based pada dataset log server menggunakan support vector machine berbasis linear discriminat analysis untuk deteksi malicious activity. *Jurnal Teknologi Informasi dan Ilmu Komputer*, *9*, 245. https://doi.org/10.25126/jtiik.2022924107.

Canay, Ö., Meriç, S., Evirgen, H. ve Varan, M. (2011). Realization of campus automation web information system in context of service unity architecture. *International Symposium on Computing in Science & Engineering (ISCSE)* (p. 173–179) içinde. Izmir, Turkey.

Canay, Ö. ve Kocabıçak, Ü. (2023). An innovative data collection method to eliminate the preprocessing phase in web usage mining. *Engineering Science and Technology, an International Journal*, *40*, 1–14. https://doi.org/10.1016/j.jestch.2023.101360.

Canay, Ö. ve Kocabıçak, Ü. (2016). A new data collection model for information extraction from web click logs. *International Artificial Intelligence and Data Processing Symposium (IDAP)* (p. 489–492) içinde. Inonu University.

Čegan, L. ve Filip, P. (2017). Webalyt: Open web analytics platform. *2017 27th International Conference Radioelektronika* (ss. 1–5) içinde. IEEE. https://doi.org/10.1109/radioelek.2017.7937605.

Chan-Jong-Chu, K., Islam, T., Exposito, M. M., Sheombar, S., Valladares, C., Philippot, O., Grua, E. M. ve Malavolta, I. (2020). Investigating the correlation between performance scores and energy consumption of mobile web apps. *Proceedings of the Evaluation and Assessment in Software Engineering* (ss. 190—199) içinde. ACM. https://doi.org/10.1145/3383219.3383239.

Cheng, Y. J. ve Chen, K. H. (2022). Website analytics for government user behavior during COVID-19 pandemic. *Aslib Journal of Information Management*, *75*, 90–111. https://doi.org/10.1108/ajim-11-2021-0329.

Chung, Y., Liu, S., Wang, C. ve Pang, C. (2019). Applying Fuzzy MCDM Methods to the Evaluation on Portal Website Service Quality. *The SIJ Transactions on Computer Science Engineering & its Applications (CSEA)*. https://doi.org/10.9756/SIJCSEA/V7I4/03010020401.

Clifton, B. (2012). *Advanced web metrics with Google Analytics* (3rd ed.). John Wiley & Sons.

Cooley, R., Tan, PN. ve Srivastava, J. (2000). Discovery of Interesting Usage Patterns from Web Data. *Web Usage Analysis and User Profiling* içinde. WebKDD 1999. Lecture Notes in Computer Science, vol 1836. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-44934-5_10

Daş, R. (2008). *Web kullanıcı erişim kütüklerinden bilgi çıkarımı* [Doktora tezi]. Fırat Üniversitesi

Das, R. ve Turkoglu, I. (2009). Creating meaningful data from web logs for improving the impressiveness of a website by using path analysis method. *Expert Systems with Applications*, *36*, 6635–6644. https://doi.org/10.1016/j.eswa.2008.08.067.

De-Groot, R., Golumbic, Y. N., Martínez, F. M., Hoppe, H. U. ve Reynolds, S. (2022). Developing a framework for investigating citizen science through a combination of web analytics and social science methods—the CS track perspective. *Frontiers in Research Metrics and Analytics*, *7*. https://doi.org/10.3389/frma.2022.988544.

Demirkan, H. ve Delen, D. (2013). Leveraging the capabilities of service-oriented decision support systems: Putting analytics and big data in cloud. *Decision Support Systems*, *55*, 412–421. https://doi.org/10.1016/j.dss.2012.05.048.

Deshpande, D., Deshpande, S. ve Thakare, V. (2018). Web user identification: Analysis of heuristic solutions. *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)* (p. 1790–1795) içinde. IEEE. https://doi.org/10.1109/ICCONS.2018.8662893.

Dewi, M. J. ve Nurdin, N. (2023). Analisis performa platform sosial media menggunakan perbandingan software automated testing. *Information Management for Educators and Professionals*, *7*, 164–173. https://doi.org/10.51211/imbi.v7i2.2343.

Durántez, A. B. (2023). Implications of zero party data on user decision-making in digital advertising. *Behanomics*, *1*. https://doi.org/10.55223/bej.7.

Ehikioya, S. A. ve Lu, S. (2019). A path analysis model for effective e-commerce transactions. *African Journal of Computing and ICT*, *12*, 55–71.

Everest-Phillips, M. (2019). Big data-driven public service in the twenty-first century: The politics of big data. *Public Service Excellence in the 21st Century* (ss. 275–318) içinde. Springer Singapore. https://doi.org/10.1007/978-981-13-3215-9_9.

Farid, M., Elgohary, R., Moawad, I. ve Roushdy, M. (2018). User Profiling Approaches, Modeling, and Personalization. *SSRN Electronic Journal* içinde. Elsevier BV. https://doi.org/10.2139/ssrn.3389811

Fatima, B., Ramzan, H. ve Asghar, S. (2016). Session identification techniques used in web usage mining: A systematic mapping of scholarly literature. *Online Information Review*, *40*, 1033–1053. https://doi.org/10.1108/OIR-08-2015-0274.

Fundingsland, E. L. J., Fike, J., Calvano, J., Beach, J., Lai, D. ve He, S. (2022). Methodological guidelines for systematic assessments of health care websites using web analytics: Tutorial. *Journal of Medical Internet Research*, *24*, e28291. https://doi.org/10.2196/28291.

Gamalielsson, J., Lundell, B., Butler, S., Brax, C., Persson, T., Mattsson, A., Gustavsson, T., Feist, J. ve Lönroth, E. (2021). Towards open government through open source software for web analytics: The case of matomo. *JeDEM-eJournal of eDemocracy and Open Government*, *13*, 133–153. https://doi.org/10.29379/jedem.v13i2.650.

García, S., Luengo, J. ve Herrera, F. (2015). *Data preprocessing in data mining* volume 72. Springer. https://doi.org/10.1007/978-3-319-10247-4.

Gholamian, S. ve Ward, P. (2021). A comprehensive survey of logging in software: From logging statements automation to log mining and analysis. *arXiv preprint arXiv:2110.12489*, . https://doi.org/10.48550/arXiv.2110.12489.

Haaksma, T. R., de Jong, M. D. T. ve Karreman, J. (2018). Users' personal conceptions of usability and user experience of electronic and software products. *IEEE Transactions on Professional Communication*, *61*, 116–132. https://doi.org/10.1109/tpc.2018.2795398.

Harika, B. ve Sudha, T. (2019). Extraction of knowledge from web server logs using web usage mining. *Asian Journal of Computer Science and Technology*, *8*, 12–15. https://doi.org/10.51983/ajcst-2019.8.s3.2113.

Hasan, S. B. (2022). Web analyzer software tool in order to identify an know the behavior of users that visit website. *Academic Journal of Nawroz University*, *11*, 75–81. https://doi.org/10.25007/ajnu.v11n4a506.

Heričko, T., Sumak, B. ve Brdnik, S. (2021). Towards representative web performance measurements with google lighthouse. *Proceedings of the 2021 7th Student Computer Science Research Conference (StuCoSReC)* (ss. 39–42) içinde. University of Maribor Press. https://doi.org/10.18690/978-961-286-516-0.

Hossfeld, T., Metzger, F. ve Rossi, D. (2018). Speed index: Relating the industrial standard for user perceived web performance to web QoE. *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)* (pp. 1–6) içinde. IEEE. https://doi.org/10.1109/qomex.2018.8463430.

Ibrahim, K. K. ve Obaid, A. J. (2021). Web mining techniques and technologies: A landscape view. *Journal of Physics: Conference Series*, *1879*, 032125. https://doi.org/10.1088/1742-6596/1879/3/032125.

Jain, V. ve Kashyap, K. (2021). An efficient algorithm for web log data preprocessing. *Machine Vision and Augmented Intelligence—Theory and Applications* (p. 505–514) içinde. Singapore: Springer Singapore. https://doi.org/10.1007/978-981-16-5078-9_41.

Jansen, B. J., Jung, S. ve Salminen, J. (2022). Measuring user interactions with websites: A comparison of two industry standard analytics approaches using data of 86 websites. *PLOS ONE*, *17*, e0268212. https://doi.org/10.1371/journal.pone.0268212.

Jin, J. ve Lin, X. (2022). Web Log Analysis and Security Assessment Method Based on Data Mining. *Computational Intelligence and Neuroscience*, Vol. 2022, Article ID 8485014 (p. 1–9). https://doi.org/10.1155/2022/8485014

Joachims, T., Granka, L., Pan, B., Hembrooke, H. ve Gay, G. (2017). Accurately interpreting clickthrough data as implicit feedback. *ACM SIGIR Forum* (p. 4–11) içinde. ACM volume 51. https://doi.org/10.1145/1076034.1076063.

Jokar, N., Honarvar, A. R., Aghamirzadeh, S. ve Esfandiari, K. (2016). Web mining and web usage mining techniques. *Bulletin de la Société des Sciences de Liège*, *85*, 321–328. https://doi.org/10.25518/0037-9565.5371.

Junmei, W. ve Jihong, W. (2019). Research on performance automation testing technology based on jmeter. *2019 International Conference on Robots & Intelligent System (ICRIS)* (ss. 55–58) içinde. IEEE. https://doi.org/10.1109/icris.2019.00023.

Kaur, N. ve Aggarwal, H. (2017). A novel semantically-time-referrer based approach of web usage mining for improved sessionization in pre-processing of web log. *International journal of advanced computer science and applications*, *8*. https://doi.org/10.14569/IJACSA.2017.080122.

Keskin, S. ve Yazıcı, A. (2022). Modeling and querying fuzzy SOLAP-based framework. *ISPRS International Journal of Geo-Information*, *11*, 191. https://doi.org/10.3390/ijgi11030191.

Kewen, L. (2012). Analysis of preprocessing methods for web usage data. *Proceedings of 2012 International Conference on Measurement, Information and Control* (p. 383–386) içinde. IEEE volume 1. https://doi.org/10.1109/MIC.2012.6273276.

Knight-Davis, S. (2017). Using awstats to analyze logs from ezproxy and from the public opac logs. *Spring Forum: Collection Management and Technical Services Committees* (p. 228) içinde.

Kousik, N. V., Sivaram, M., Yuvaraj, N. ve Mahaveerakannan, R. (2021). Improved density-based learning to cluster for user web log in data mining. *Inventive Computation and Information Technologies* (ss. 813–830) içinde. Springer Singapore. https://doi.org/10.1007/978-981-33-4305-4_59.

Krishnan, A., Das, M., Bendre, M., Yang, H. ve Sundaram, H. (2020). Transfer learning via contextual invariants for one-to-many cross-domain recommendation. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (p. 1081–1090) içinde. Association for Computing Machinery. https://doi.org/10.1145/3397271.3401078.

Król, K. ve Zdonek, D. (2020). Analytics maturity models: An overview. *Information*, *11*, 142. https://doi.org/10.3390/info11030142.

Kumar, B., Roy, S., Sinha, A., Iwendi, C. ve Strážovská, L. (2022). E-commerce website usability analysis using the association rule mining˘ and machine learning algorithm. *Mathematics*, *11*, 25. https://doi.org/10.3390/math11010025.

Kumar, S. ve Kumar, R. (2021). A study on different aspects of web mining and research issues. *IOP Conference Series: Materials Science and Engineering* (ss. 012–018) içinde. IOP Publishing volume 1022. https://doi.org/10.1088/1757-899X/1022/1/012018.

Kumar, V. ve Ogunmola, G. A. (2020). Web analytics for knowledge creation: A systematic review of tools, techniques, and practices. *International Journal of Cyber Behavior, Psychology and Learning*, *10*, 1–14. https://doi.org/10.4018/ijcbpl.2020010101.

Kundu, S. ve Garg, L. (2017). Web log analyzer tools: A comparative study to analyze user behavior. *2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence* (p. 17–24) içinde. IEEE. https://doi.org/10.1109/CONFLUENCE.2017.7943117.

Li, J., Lu, Y., Zhang, Y., Wang, Q., Cheng, Z., Huang, K. ve Shu, J. (2022). SwitchTx: scalable in-network coordination for distributed transaction processing. *Proceedings of the VLDB Endowment*, *15*, 2881–2894. https://doi.org/10.14778/3551793.3551838.

Lim, Z. Y., Ong, L. Y. ve Leow, M. C. (2021). A review on clustering techniques: Creating better user experience for online roadshow. *Future Internet*, *13*, 233. https://doi.org/10.3390/fi13090233.

Lourenço, R. P. (2015). An analysis of open government portals: a perspective of transparency for accountability. *Government Information Quarterly*, 32(3), 323-332. https://doi.org/10.1016/j.giq.2015.05.006

Malarvizhi, S. ve Sathiyabhama, B. (2016). Frequent pagesets from web log by enhanced weighted association rule mining. *Cluster Computing*, *19*, 269–277. https://doi.org/10.1007/s10586-015-0507-z.

Manchanda, M. ve Gupta, N. (2018). Web usage mining: Dynamic methodology to preprocessing web logs. *HELIX*, *8*, 3810–3815. https://doi.org/10.29042/2018-3810-3815.

Maslennikova, A., Rotelli, D. ve Monreale, A. (2022). Visual analytics for session-based time-windows identification in virtual learning environments. *2022 26th International Conference Information Visualisation (IV)* (ss. 251–258) içinde. IEEE. https://doi.org/10.1109/iv56949.2022.00050.

Milosevic, B., Regodic, D. ve Saso, V. (2021). Big data management processes in business intelligence systems. *Economic and Social Development: Book of Proceedings* (ss. 182–192) içinde. Varazdin Development and Entrepreneurship Agency (VADEA).

Mishra, S. ve Srivastava, S. (2021). Web development frameworks and its performance analysis—a review. *Smart Computing*, (p. 337–343). https://doi.org/10.1201/9781003167488-39.

Mobasher, B. (2009). Web mining overview. J. Wang (Ed.), *Encyclopedia of Data Warehousing and Mining, Second Edition* chapter 319. (ss. 2085–2089) içinde. IGI Global volume 3. (2nd ed.). https://doi.org/10.4018/978-1-60566-010-3.ch319.

Mortazavi, S. H., Salehe, M., Balasubramanian, B., Puzhavakath de Lara, E. ve Narayanan, S. (2020). SessionStore: A session-aware datastore for the edge. *2020 IEEE 4th International Conference on Fog and Edge Computing (ICFEC)* (ss. 59–68) içinde. IEEE. https://doi.org/10.1109/icfec50348.2020.00014.

Mughal, M. J. H. (2018). Data mining: Web data mining techniques, tools and algorithms: An overview. *International Journal of Advanced Computer Science and Applications*, *9*. https://doi.org/10.14569/IJACSA.2018.090630.

Munk, M. ve Benko, L. (2018). Using entropy in web usage data preprocessing. *Entropy*, *20*, 67. https://doi.org/10.3390/e20010067.

Nandal, R. (2018). A systematic review on data preprocessing and pattern discovery of web usage mining. *International Journal of Advanced Research in Computer Science*, *9*. https://doi.org/10.26483/ijarcs.v9i2.5763.

Nurshuhada, A., Yusop, R. O. M., Azmi, A., Ismail, S. A., Sarkan, H. M. ve Kama, N. (2019). Enhancing performance aspect in usability guidelines for mobile web application. *2019 6th International Conference on Research and Innovation in Information Systems (ICRIIS)* (ss. 1–6) içinde. IEEE. https://doi.org/10.1109/icriis48246.2019.9073617.

Ofoeda, J., Boateng, R. ve Effah, J. (2019). Application programming interface (API) research. *International Journal of Enterprise Information Systems*, *15*, 76–95. https://doi.org/10.4018/ijeis.2019070105.

Önder, I. ve Berbekova, A. (2021). Web analytics: more than website performance evaluation? *International Journal of Tourism Cities*, *8*, 603–615. https://doi.org/10.1108/ijtc-03-2021-0039.

Paredes, R. ve Bolanio, J. (2020). Analyzing logs from proxy server and captive portal using k-means clustering algorithm. *Middle East Journal of Applied Science & Technology*, *3*, 10–31. https://doi.org/10.4018/IJCBPL.2020010101.

Pilz, M., Burke, R., Redmond, E., Sauer, M. ve Weichelt, B. (2020). JA:2021-23. ideal tracking and analytics: Incorporating modern toolsets to enhance analysis of user acquisition and behavior on AgInjuryNews.org. *Journal of Agromedicine*, *25*, 248–249. https://doi.org/10.1080/1059924x.2020.1765575.

Pinho, C., Franco, M. ve Mendes, L. (2018). Web portals as tools to support information management in higher education institutions: A systematic literature review. *Int. J. Inf. Manag.*, 41, 80-92. https://doi.org/10.1016/j.ijinfomgt.2018.04.002.

Pourghassemi, B., Sani, A. A. ve Chandramowlishwaran, A. (2019). What-if analysis of page load time in web browsers using causal profiling. *Abstracts of the 2019 SIGMETRICS/Performance Joint International Conference on Measurement and Modeling of Computer Systems* (ss. 1–23) içinde. ACM. https://doi.org/10.1145/3309697.3331483.

Quintel, D. ve Wilson, R. (2020). Analytics and privacy. *Information Technology and Libraries*, *39*, 1–11. https://doi.org/10.6017/ital. v39i3.12219.

Rafiq, U. (2022). Towards understanding analytics in software startups. *Proceedings of the 5th International Workshop on Software Intensive Business: Towards Sustainable Software Business* (ss. 31–38) içinde. ACM. https://doi.org/10.1145/3524614.3528632.

Reddy, G. (2021). A review of data warehouses multidimensional model and data mining. *Information Technology in Industry*, *9*, 310–320.

Regueiro, C., Seco, I., de Diego, S., Lage, O. ve Etxebarria, L. (2021). Privacy-enhancing distributed protocol for data aggregation based on blockchain and homomorphic encryption. *Information Processing & Management*, *58*, 102745. https://doi.org/10.1016/j.ipm.2021.102745.

Román, P. E., L'Huillier, G. ve Velásquez, J. D. (2010). Web Usage Mining. *Advanced Techniques in Web Intelligence - I* (ss. 143-165) içinde. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-14461-5_6

Roy, R. ve Rao, G. A. (2020). Survey on pre-processing web log files in web usage mining. *International Journal of Advanced Science and Technology*, *29*, 682–691.

Saif, D., Lung, C. H. ve Matrawy, A. (2021). An early benchmark of quality of experience between HTTP/2 and HTTP/3 using lighthouse. *ICC 2021 - IEEE International Conference on Communications* (ss. 1–6) içinde. IEEE. https://doi.org/10.1109/icc42927.2021.9500258.

Sakas, D. P., Giannakopoulos, N. T., Reklitis, D. P. ve Dasaklis, T. K. (2021). The effects of cryptocurrency trading websites on airlines' advertisement campaigns. *Journal of Theoretical and Applied Electronic Commerce Research*, *16*, 3099–3119. https://doi.org/10.3390/ jtaer16070169.

Salminen, J., Jung, S. G., Chowdhury, S., Sengün, S. ve Jansen, B. J. (2020). Personas and analytics: A comparative user study of efficiency and effectiveness for a user identification task. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* içinde. ACM. https://doi.org/10.1145/3313831.3376770.

Samarasinghe, N., Adhikari, A., Mannan, M. ve Youssef, A. (2022). Et tu, brute? Privacy analysis of government websites and mobile apps. *Proceedings of the ACM Web Conference 2022* (ss. 564–575) içinde. ACM. https://doi.org/10.1145/3485447.3512223.

Shah, P. ve Pandit, H. B. (2022). A review: Web content mining techniques. *Data Engineering for Smart Systems*, (ss. 159–172). https://doi.org/10.1007/978-981-16-2641-8_15.

Slanzi, G., Pizarro, G. ve Velásquez, J. D. (2017). Biometric information fusion for web user navigation and preferences analysis: An overview. *Information Fusion*, *38*, 12–21. https://doi.org/10.1016/j.inffus.2017.02.006.

Srivastava, M., Garg, R. ve Mishra, P. (2018). A mapreduce-based user identification algorithm in web usage mining. *International Journal of Information Technology and Web Engineering (IJITWE)*, *13*, 11–23. https://doi.org/10.4018/IJITWE.2018040102.

Srivastava, M., Garg, R. ve Mishra, P. K. (2015). Analysis of data extraction and data cleaning in web usage mining. *Proceedings of the 2015 International Conference on Advanced Research in Computer Science Engineering & Technology (ICARCSET 2015)* (p. 1–6) içinde. Association for Computing Machinery. https://doi.org/10.1145/2743065.2743078.

Srivastava, M., Srivastava, A. ve Garg, R. (2019). Data preprocessing techniques in web usage mining: A literature review. *Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM)* (ss. 466–476) içinde. Amity University Rajasthan, Jaipur-India. https://doi.org/10.2139/ssrn.3352357.

Srivastava, M., Srivastava, A. K., Garg, R. ve Mishra, P. (2021). Performance evaluation of the mapreduce-based parallel data preprocessing algorithm in web usage mining with robot detection approaches. *IETE Technical Review*, (ss. 1–15). https://doi.org/10.1080/02564602.2021.1918584.

Stringam, B. ve Gerdes, J. (2019). Service gap in hotel website load performance. *International Hospitality Review*, *33*, 16–29. https://doi.org/10.1108/ihr-09-2018-0012.

Sukumar, P., Robert, L. ve Yuvaraj, S. (2016). Review on modern data preprocessing techniques in web usage mining (wum). *2016 International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)* (p. 64–69) içinde. IEEE. https://doi.org/10.1109/CSITSS.2016.7779441.

Svec, P., Benko, L., Kadlecik, M., Kratochvil, J. ve Munk, M. (2020). Web usage mining: Data pre-processing impact on found knowledge in predictive modelling. *Procedia Computer Science*, *171*, 168–178. https://doi.org/10.1016/j.procs.2020.04.018.

Szalek, K. ve Borzemski, L. (2018). Conversion rate gain with web performance optimization: a case study. *Advances in Intelligent Systems and Computing* (ss. 312–323) içinde. Springer International Publishing. https://doi.org/10.1007/978-3-319-99981-4_29.

Tao, Y., Guo, S., Shi, C. ve Chu, D. (2020). User behavior analysis by cross-domain log data fusion. *IEEE Access*, *8*, 400–406. https://doi.org/10.1109/access.2019.2961769.

Tyagi, N. ve Gupta, S. K. (2018). Web structure mining algorithms: A survey. *Big Data Analytics* (ss. 305–317) içinde. Springer. https://doi.org/10.1007/978-981-10-6620-7_30.

Utz, C., Amft, S., Degeling, M., Holz, T., Fahl, S. ve Schaub, F. (2023). Privacy rarely considered: Exploring considerations in the adoption of third-party services by websites. *Proc. Priv. Enhancing Technol.*, *2023*, 5–28. https://doi.org/10.56553/popets-2023-0002.

Vamsee, C. S., Rakesh, D., Prathyusha, I., Dinesh, B. ve Bharathi, C. (2023). Demographic and psychographic customer segmentation for ecommerce applications. *2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)* (ss. 615–622) içinde. https://doi.org/10.1109/ICAAIC56838.2023.10140861.

Varnagar, C. R., Madhak, N. N., Kodinariya, T. M. ve Rathod, J. N. (2013). Web usage mining: A review on process, methods and techniques. *2013 International Conference on Information Communication and Embedded Systems (ICICES)* (ss. 40–46) içinde. IEEE. https://doi.org/10.1109/ICICES.2013.6508399.

Vasylyshyn, S., Susukailo, V., Opirskyy, I., Kurii, Y. ve Tyshyk, I. (2023). A model of decoy system based on dynamic attributes for cybercrime investigation. *Eastern-European Journal of Enterprise Technologies*, *1*, 6–20. https://doi.org/10.15587/1729-4061.2023.273363.

Walny, J., Frisson, C., West, M., Kosminsky, D., Knudsen, S., Carpendale, S. ve Willett, W. (2020). Data changes everything: Challenges and opportunities in data visualization design handoff. *IEEE Transactions on Visualization and Computer Graphics*, *26*, 12–22. https://doi.org/10.1109/tvcg.2019.2934538.

Yang, J., Xiu, P., Sun, L., Ying, L. ve Muthu, B. (2022). Social media data analytics for business decision making system to competitive analysis. *Information Processing and Management*, *59*, 102751. https://doi.org/10.1016/j.ipm.2021.102751.

Yu, A. ve Benson, T. A. (2021). Dissecting performance of production QUIC. *Proceedings of the Web Conference 2021* (ss. 1157—1168) içinde. ACM. https://doi.org/10.1145/3442381.3450103.

Zheng, G. ve Peltsverger, S. (2015). Web analytics overview. *Encyclopedia of Information Science and Technology, Third Edition* (p. 7674–7683) içinde. IGI Global. https://doi.org/10.4018/978-1-4666-5888-2.ch756.

ÖZGEÇMİŞ

Ad-Soyad : Özkan CANAY

ÖĞRENİM DURUMU:

- **Ön Lisans** **:** 1996, Ankara Üniversitesi, KMYO, Bilgisayar Programcılığı
- **Lisans** **:** 2002, Sakarya Üniversitesi, Mühendislik Fak., Endüstri Müh.
- **Yüksek Lisans** **:** 2005, Sakarya Üniv., FBE, Endüstri Mühendisliği ABD
- **Doktora** **:** 2024, Sakarya Üniv., FBE, Bilgisayar ve Bilişim Müh. ABD

MESLEKİ DENEYİM VE ÖDÜLLER:

- 2018-..., Öğretim Görevlisi, Sakarya Meslek Yüksekokulu, SUBÜ
- 2003-2018, Öğretim Görevlisi, Adapazarı Meslek Yüksekokulu (e-Eğitim), SAÜ
- 2011-2012, Müdür, Bilgisayar Araştırma ve Uygulama Merkezi, SAÜ
- 2008-2012, Başkanvekili ve Kurucu Yönetim Kur. Üyesi, Sakarya Bilişim Derneği
- 2005-2011, Kurucu Müdür Yrd., Bilgisayar Araştırma ve Uyg. Merkezi, SAÜ
- 2005-2008, Bilgi İşlem Şube Müdürü, Bilgi İşlem Dairesi Başkanlığı, SAÜ
- 2002-2005, Web Teknolojileri Grup Başkanı, Bilgi İşlem Dairesi Başkanlığı, SAÜ
- 2002-2003, Öğretim Görevlisi, Sakarya Meslek Yüksekokulu (e-Eğitim), SAÜ
- 1996-2002, Web ve Sistem Yöneticisi, Bilgi İşlem Dairesi Başkanlığı, SAÜ
- 2008, SAÜ Rektör v. Prof. Dr. Muzaffer Elmas tarafından üstün başarı ödülü
- 2007, SAÜ Rektörü Prof. Dr. Mehmet Durman tarafından takdir belgesi
- 2007, SAÜ Genel Sekreteri Dr. Zafer Demir tarafından takdir belgesi
- 2006, TSK MEBS Okulu Kom. Muh. Lab. Destek G. Bşk. tarafından takdir belgesi
- 2005, SAÜ Genel Sekreteri Dr. Zafer Demir tarafından teşekkür yazısı
- 1999, SAÜ Genel Sekreteri Dr. Zafer Demir tarafından teşekkür yazısı
- 1998, SAÜ Rektörü Prof. Dr. İsmail Çallı tarafından teşekkür plaketi
- 1998, SAÜ Rektörü Prof. Dr. İsmail Çallı tarafından teşekkür yazısı

TEZDEN TÜRETİLEN ESERLER:

- Canay Ö. ve Kocabıçak Ü. (2016, 17-18, Eylül). Web Tıklama Kayıtlarından Bilgi Çıkarımı İçin Yeni Bir Veri Toplama Modeli. *International Artificial Intelligence and Data Processing Symposium (IDAP)*, Malatya, Türkiye
- Canay, Ö. ve Kocabıçak, Ü. (2023). An Innovative Data Collection Method to Eliminate the Preprocessing Phase in Web Usage Mining. *Engineering Science and Technology, an International Journal*, 40, 101360. https://doi.org/10.1016/j.jestch.2023.101360.
- Canay, Ö. ve Kocabıçak, Ü. (2024). CAWAL: A Novel Unified Analytics Framework for Enterprise Web Applications and Multi-Server Environments. *Information Processing and Management*, 61(3), 103617. https://doi.org/10.1016/j.ipm.2023.103617.

DİĞER ESERLER:

- Arslan, H. ve Canay, Ö. (2023). Cyber Threat Intelligence Systems and Applications. Academic Research and Reviews in Engineering Sciences (ss. 103-116) içinde. Platanus Publishing. https://doi.org/10.5281/zenodo.10060775.
- Durdu, A., Çınar, İ. ve Canay, Ö. (2023). Utilization of Information Systems to Enhance the Efficiency of Internal Audit Activity Management and Audit Testing Processes. Current Debates on Natural and Engineering Sciences 11 (ss. 311-320) içinde. Bidge Publications.
- Durdu, A. ve Canay, Ö. (2023). Digital Transformation and Distance Education Systems of Universities in Türkiye During the Pandemic. Current Debates on Natural and Engineering Sciences 11 (ss. 288-310) içinde. Bidge Publications.
- Arslan, H. ve Canay, Ö. (2019). Comparison of Data Transfer Performance of BitTorrent Transmission Protocols. *Cumhuriyet Science Journal, 40*(3), 762-767.
- Canay, Ö. (2019). Dokunmatik Cihaz Fonksiyonlarına Erişim İçin Çift Sensör Yönetimli Aygıt ve İşlev Yöntemi. Patent numarası: TR201617552-B.
- Canay, Ö. (2018). Uzaktan Kontrollü Bir Deklanşöre İlişkin Çalışma Yöntemi. Patent numarası: TR201409965-B.
- Varan, M., Haidary, S., Öylek, İ. ve Canay, Ö. (2018). Akıllı Şebekelerde Güvenli Haberleşme Tabanlı Güç Akışı Analizi. *Journal of New Results in Engineering and Natural Sciences, 8*(1), 63-74.
- Arslan, H., Yüksek, A. G., Elyakan, M. L. ve Canay, Ö. (2018). Usability and Quality Tests in Software Products to Oriented of User Experience. *The Online Journal of Quality in Higher Education, 5*(3), 79-83.
- Canay, Ö. (2017). Are Our Universities Ready for The Changing Web User Profile in Every Aspect? *The Online Journal of Communication and Media, 3*(1), 1-5.

- Canay, Ö. (2017). Infrastructure of SCI Journal Search Engine. Researches on Science and Art in 21st Century Turkey Vol. 2 (ss. 1955-1961) içinde. Gece Publishing.
- Canay, Ö. ve Güngörsün, T. (2016). *Bilgisayar Donanımı ve Bileşenleri*. Değişim Yayınları. ISBN: 978-605-4925-91-9.
- Canay, Ö. (2016). *Web Tasarımı ve Ön Yüz Web Teknolojileri*. Değişim Yayınları. ISBN: 978-605-4925-90-2.
- Can, E. ve Canay, Ö. (2016). A Planar Robot Design and Construction with Maple. *The Online Journal of Science and Technology, 6*(2), 1-5.
- Cakar, T., Köker, R. ve Canay, O. (2015). A New Neuro-Dominance Rule for Single-Machine Tardiness Problem with Double Due Date. *Neural Computing & Applications, Springer London, 26*(6), 1439-1450. https://doi.org/10.1007/s00521-014-1789-4.
- Taskin, H., Göztepe, K., Taskin, M.F. ve Canay, Ö. (2008). Agent Based University Planning of Sakarya University: CAWIS Project. *Journal of Theoretical and Applied Information Technology, 4*(8), 753-759.
- Arslan, H. ve Canay, Ö. (2007). Sakarya Üniversitesi Bağıl Değerlendirme Sisteminde Yapay Sinir Ağları Kullanılarak Bağıl Aritmetik Ortalama ve Üst Değer Tespiti. *Electronic Letters on Science & Engineering, 3*(2), 18-26.