



**T.C.
İSTANBUL ÜNİVERSİTESİ-CERRAHPAŞA
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**



DOKTORA TEZİ

**SANAL ORTAM VERİLERİ İLE YAPAY ZEKA TABANLI DİNAMİK
TAHMİNLEME MODELİ GELİŞTİRİLMESİ**

Adem ÇAYIR

DANIŞMAN

Doç. Dr. Fatih TÜYSÜZ

Endüstri Mühendisliği Anabilim Dalı

Endüstri Mühendisliği, Doktora Programı

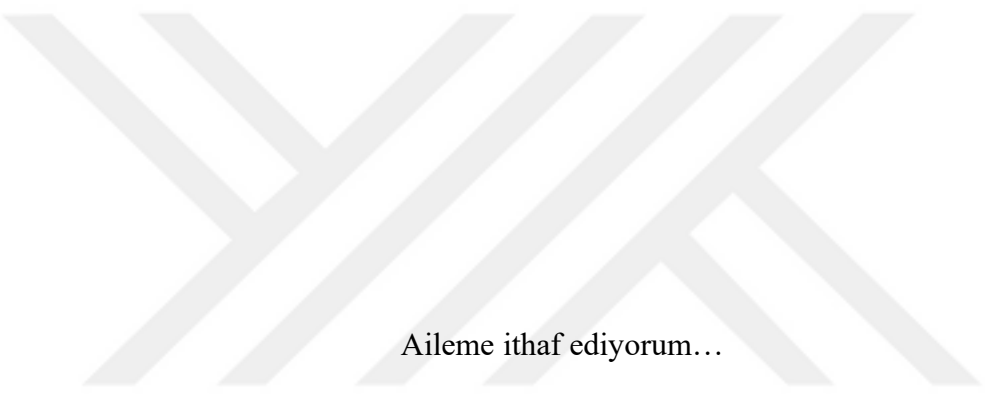
Temmuz, 2025

TEZ KABUL VE ONAYI

Adem ÇAYIR tarafından, Doç. Dr. Fatih TÜYSÜZ danışmanlığında hazırlanan "SANAL ORTAM VERİLERİ İLE YAPAY ZEKA TABANLI DİNAMİK TAHMİNLEME MODELİ GELİŞTİRİLMESİ" başlıklı bu çalışma, jürimiz tarafından 24/07/2025 tarihinde yapılan sınav sonucunda oy birliği ile başarılı bulunarak **Doktora Tezi** olarak kabul edilmiştir.

Tez Jürisi

	İmza	Sonuç
DANIŞMAN	Doç. Dr. Fatih TÜYSÜZ	☒
	İstanbul Üniversitesi Cerrahpaşa	Kabul
	Endüstri Mühendisliği Anabilim Dalı	☐ Ret
ÜYE	Doç. Dr. Ersin NAMLI	☒
	İstanbul Üniversitesi Cerrahpaşa	Kabul
	Endüstri Mühendisliği Anabilim Dalı	☐ Ret
ÜYE	Prof. Dr. Abdullah Çağrı TOLGA	☒
	Galatasaray Üniversitesi	Kabul
	Endüstri Mühendisliği Anabilim Dalı	☐ Ret
ÜYE	Prof. Dr. Mehmet Mutlu Yenisey	☒
	İstanbul Üniversitesi Cerrahpaşa	Kabul
	Endüstri Mühendisliği Anabilim Dalı	☐ Ret
ÜYE	Dr. Öğr. Üyesi Özay Özaydın	☒
	Doğu Üniversitesi	Kabul
	Endüstri Mühendisliği Anabilim Dalı	☐ Ret



Aileme ithaf ediyorum...

BÜTÇE DESTEKLERİ

SANAL ORTAM VERİLERİ İLE YAPAY ZEKA TABANLI DİNAMİK TAHMİNLEME MODELİ GELİŞTİRİLMESİ

Bu tez çalışması için herhangi bir kurumdan bütçe desteği alınmamıştır.

TEŞEKKÜR

Günümüz dijital çağında, çevrim içi sanal ortamlardan elde edilen veri miktarı ve hızı katlanarak artmaktadır. Bu büyük ve dinamik veri setlerinin etkin bir şekilde analiz edilmesi ve geleceğe yönelik isabetli tahminler yapılması, karar alma süreçleri için kritik bir öneme sahiptir. Bu çalışma, bu ihtiyaca yanıt vermek üzere, "Çevrim için Sanal Ortam Verileri ile Yapay Zeka Tabanlı Dinamik Tahminleme Modeli Geliştirilmesi" amacıyla gerçekleştirilmiştir. Geliştirilen modelin, karmaşık ve sürekli değişen sanal ortam verilerinden anlamlı öngörüler üreterek hem akademik literatüre katkı sağlaması hem de pratik uygulamalara yeni bir bakış açısı sunması hedeflenmektedir.

Bu meşakkatli ve bir o kadar da keyifli yolculukta, engin bilgi ve tecrübeleriyle ufkumu açan, çalışmamın her aşamasında değerli yönlendirmeleriyle bana yol gösteren ve desteğini hiçbir zaman esirgemeyen danışman hocam Sayın Doç. Dr. Fatih TÜYSÜZ'e en içten şükranlarımı sunarım.

Tez çalışmalarım boyunca yapıcı eleştirileri, yol gösterici fikirleri ve motive eden yaklaşımlarıyla önemli katkılarda bulunan değerli hocalarım Sayın Doç. Dr. Ersin NAMLI ve Sayın Prof. Dr. Abdullah Çağrı TOLGA'ya teşekkürü bir borç bilirim. Onların bilimsel bakış açıları, çalışmamın olgunlaşmasında kilit bir rol oynamıştır.

Akademik hayatım boyunca derslerini alma fırsatı bulduğum ve bilgi birikimlerinden faydalandığım tüm değerli hocalarıma da teşekkür ederim.

Son olarak, bu uzun ve zorlu süreçte en büyük güç kaynağım olan, sevgileri, sabırları ve destekleriyle her zaman yanımda hissettiğim kıymetli aileme sonsuz teşekkürlerimi sunarım. Onların varlığı olmadan bu başarı mümkün olmazdı.

Temmuz 2025

Adem ÇAYIR

İÇİNDEKİLER

Sayfa No

TEZ KABUL VE ONAYI	ii
BEYAN	iii
BÜTÇE DESTEKLERİ	v
TEŞEKKÜR	vi
İÇİNDEKİLER	vii
ŞEKİL LİSTESİ	ix
TABLO LİSTESİ	x
SİMGE VE KISALTMA LİSTESİ	xi
ÖZET	xiv
ABSTRACT	xvi
1. GİRİŞ	1
1.1. KATKILAR	2
1.2. TEZİN ANA HATLARI	3
2. KAVRAMSAL ÇERÇEVE	5
2.1. TALEP TAHMİNİ	5
2.1.1. Kalitatif Talep Tahmin Yöntemleri	6
2.1.2. Kantitatif Talep Tahmin Yöntemleri	9
2.2. YAPAY ZEKA VE MAKİNE ÖĞRENİM YÖNTEMLERİ	12
2.2.1. Doğrusal Regresyon	15
2.2.2. Destek Vektör Makineleri	21
2.2.3. Gradyan Hızlandırma Regresyon Algoritması	24
2.3. BÜYÜK DİL MODELLERİ	28
2.3.1. Büyük Dil Modeli Oluşturma	29
2.3.2. Büyük Dil Modeli Kullanım Alanları	34
3. YÖNTEM	37
3.1. VERİ TOPLAMA	40
3.2. VERİ İŞLEME	40
3.2.1. Girdi Tasarımı	41
3.2.2. Çevrimiçi Model ve API Arayüzü	44

3.2.3. Veri Sonuçlarının İşlenmesi	44
3.3. VERİ İŞLEME.....	46
3.4. TAHMİNLEME.....	48
3.4.1. Doğrusal Regresyon Algoritması	51
3.4.2. Destek Vektör Regresyon Algoritması	54
3.4.3. Gradyan Hızlandırma Regresyon Algoritması	56
4. BULGULAR.....	58
4.1. DOĞRUSAL REGRESYON ANALİZİ.....	58
4.2. DESTEK VEKTÖR REGRESYON ANALİZİ.....	60
4.3. GRADYAN HIZLANDIRMA REGRESYON ANALİZİ	63
4.4. GENEL DEĞERLENDİRME	66
5. TARTIŞMA.....	68
6. SONUÇ VE ÖNERİLER	72
KAYNAKLAR.....	75
EKLER	82
EK 1. Doğrusal regresyon örnek analiz python kodu	82
EK 2. GPT-4o API çağrı örneği.....	83
EK 3. Büyük dil modeli sonuç değerlendirme regexleri.....	83
EK 4. Doğrusal regresyon python kodu.....	84
EK 5. Destek vektör regresyon algoritma python kodu.....	85
EK 6. Gradyan hızlandırma regresyon algoritma python kodu	86
EK 7: Haftalık zenginleştirilmiş veri seti.....	87
İNTİHAL RAPORU İLK SAYFASI	90
ETİK KURUL İZİN YAZISI	91
KURUM İZNİ YAZILARI.....	92
ÖZGEÇMİŞ	93

ŞEKİL LİSTESİ

	Sayfa No
Şekil 2.1: ScienceDirect’de bulunan Yapay Zeka ve Makine Öğrenimi Makale Sayıları.....	13
Şekil 2.2: Makine öğrenimi adımları (Soares, 2015)	14
Şekil 2.3: Ebeveyn – Çocuk Örnek Doğrusal Regresyon Grafiği.....	19
Şekil 2.4: Doğrusal çekirdek ile DVM örnek sınıflandırma	22
Şekil 2.5: California ev değeri ve medyan gelir için GHR analizi.....	28
Şekil 2.6: ScienceDirect Büyük Dil Modeli ile Tahminleme makale sayısı.....	36
Şekil 3.1: Yöntemde kullanılan komponentler.....	38
Şekil 3.2: Model Yapısı.....	39
Şekil 5.1: ScienceDirect Makine Öğrenimi ve Büyük Dil Makale Sayıları	69

TABLO LİSTESİ

	Sayfa No
Tablo 2.1: Ebeveyn Boyları ve Çocuk Boyları.....	15
Tablo 3.1: Haberlerin Büyük Dil modeli ile işlem sonuçları.....	48
Tablo 3.2: İstatistik Veri Kaynakları	49
Tablo 3.3: Araç Satış ve İstatistik Veriler	51
Tablo 3.4: Araç Satış İstatistikleri Haftalık Örnek	55
Tablo 4.1: Doğrusal Regresyon Hata İyileşmeleri	58
Tablo 4.2: Doğrusal Regresyon Analizi ile Sonuçlar	59
Tablo 4.3: Doğrusal Regresyon Katsayıları.....	60
Tablo 4.4: Aylık Veriler ile DVR Sonuçları.....	61
Tablo 4.5: Aylık Veriler ile DVR Hata İyileşmeleri	61
Tablo 4.6: Zenginleştirilmiş Veriler ile DVR Hata İyileşmeleri	62
Tablo 4.7: Zenginleştirilmiş Veriler ile DVR Sonuçları	63
Tablo 4.8: Aylık Veriler ile GHR Sonuçları.....	64
Tablo 4.9: Aylık Veriler ile GHR Hata İyileşmeleri	64
Tablo 4.10: Zenginleştirilmiş Veriler ile GHR Hata İyileşmeleri	65
Tablo 4.11: Zenginleştirilmiş Veriler ile GHR Sonuçları	66

SİMGE VE KISALTMA LİSTESİ

Simgeler	Açıklama
F_{t+1}	: Gelecek dönem ($t + 1$) için yapılan tahmini
Y_t	: t , dönemindeki gerçekleşen değer
t	: Toplam geçmiş dönem sayısının
ω_i	: i , dönemine atanan ağırlıklık
n	: Ortalamaya dahil edilecek dönem sayısını
α	: Düzeltme sabiti (0 ile 1 arasında bir değer)
y	: Bağımlı değişken
β_0	: Modelin sabit terimi (kesişim noktası)
β_i	: i . bağımsız değişkenin katsası
x	: Bağımsız değişken
ε	: Hata terimi ve modelin açıklayamadığı rassal değişken
$E(y)$	: Yanıt değişkeninin matematiksel beklentisi
$\hat{y}_i$	: i . tahmin edilen tepki değeri
y_i	: i . gözlemlenen tepki değeri
b_i	: i , regresyon katsayılarının en iyi tahmini
e_i	: Regresyon artığı
ξ_i, ξ_i^*	: Dolgu değişkenleri
b	: Regresyon hiper düzleminin öteleme terimi
η	: Öğrenme Katsayısı
$n_{estimators}$	: Ağaç Sayısı
A	: Belirli bir zaman aralığındaki (çalışmamızda aylık) tüm haber verileri kümesi
F	: A kümesindeki tek bir haber verisini
ω_F	: İlgili haberin 0 ile 100 arasında bir değere sahip olan doğrudan veya dolaylı etki puanı

Kısaltmalar	Açıklama
LLM	: Large Language Model
BDM	: Büyük Dil Modeli – LLM
SVM	: Support Vector Machine
DVM	: Destek Vektör Makineleri – SVM
SVR	: Support Vector Regression
DVR	: Destek Vektör Regresyonu - SVR
DV	: Destek Vektörü
İGGÖ	: İnsan Geribildirimiyle Güçlendirmeli Öğrenme
GBM	: Gradient Boosting Machine
GHM	: Gradyan Hızlandırma Makineleri – GBM
GBR	: Gradient Boosting Regression
GHR	: Gradyan Hızlandırma Regresyonu - GHR
MSE	: Mean Square Error
OKH	: Ortalama Karesel Hata - MSE
MAPE	: Mean absolute percentage error
OMYH	: Ortalama Mutlak Yüzdelik Hata - MAPE
MAE	: Mean Absolute Error
OMH	: Ortalama Mutlak Hata – MAE
KOKH	: Kök Ortalama Kare Hatası
ML	: Machine Learning
MÖ	: Makine Öğrenimi
AI	: Artificial Intelligence
YZ	: Yapay Zeka
EKK	: En Küçük Kareler
RTF	: Radyan Temelli Fonksiyon
BERT	: Bidirectional Encoder Representations from Transformers
GPT	: Generative Pretrained Transformer
NLP	: Natural Language Processing
URL	: Uniform Resource Locator
HTML	: Hypertext Markup Language
API	: Application Programming Interface
MIT	: Massachusetts Institute of Technology

RNN	: Recurrent Neural Network
NN	: Neural Network
LSTM	: Long Short Term Memory
UKSB	: Uzun Kısa Süreli Bellek – LSTM
NLP	: Natural Language Processing
DDİ	: Doğal Dil İşleme - NLP



ÖZET

DOKTORA TEZİ

SANAL ORTAM VERİLERİ İLE YAPAY ZEKA TABANLI DİNAMİK TAHMİNLEME MODELİ GELİŞTİRİLMESİ

Adem ÇAYIR

İstanbul Üniversitesi-Cerrahpaşa

Lisansüstü Eğitim Enstitüsü

Endüstri Mühendisliği Anabilim Dalı

Endüstri Mühendisliği, Doktora Programı

Danışman : Doç. Dr. Fatih TÜYSÜZ

Endüstri mühendisliğinde tahminleme, özellikle satış tahminlerinin giderek artan önemiyle birlikte, günümüzün veri odaklı dünyasında kritik bir araştırma alanı haline gelmiştir. Sanal ortamda milyarlarca insanın etkileşimleriyle oluşan büyük veri yığınları, geçmişe dair değerli bilgiler sunmakta ve geleceğe yönelik öngörülerde bulunmak için büyük bir potansiyel barındırmaktadır. Ancak bu verilerin önemli bir kısmı yapısal olmayan biçimde bulunduğundan, geleneksel tahminleme yöntemleriyle doğrudan işlenmeleri çeşitli zorluklar yaratmakta, bu da daha gelişmiş ve esnek analiz tekniklerine olan ihtiyacı artırmaktadır. Bu bağlamda hazırlanan tez çalışması, çevrimiçi ortamda oluşan yapısal olmayan büyük veriyi büyük dil modelleri (BDM) aracılığıyla işleyip anlamlandırarak yapısal hale dönüştürmeyi ve bu verileri klasik tahminleme modellerine ek bir özellik olarak entegre etmeyi hedeflemektedir. BDM'ler, geniş veri kümeleri üzerinde eğitilmiş derin öğrenme tabanlı modeller olup, doğal dil işleme yetenekleri sayesinde sanal ortamda biriken verilerden anlamlı çıkarımlar yaparak tahmin doğruluğunu artırmada önemli bir rol üstlenmektedir. Bu süreçte doğrusal regresyon, destek vektör regresyonu ve gradyan destek vektörleri gibi çeşitli makine öğrenimi algoritmaları kullanılarak tahminleme performansı daha da iyileştirilmiştir. Böylece, yapay zeka destekli veri işleme teknikleri sayesinde geleneksel yöntemlerin ötesine geçilerek daha güvenilir, tutarlı ve isabetli tahminler üretmek mümkün hale gelmiş; bu da özellikle satış

tahminleme süreçlerinde karar destek sistemlerinin daha etkin kullanılmasına olanak tanımış ve büyük veri ile makine öğrenimi yöntemlerinin endüstriyel süreçlere entegrasyonunu önemli ölçüde güçlendirmiştir.

Temmuz 2025 , 110 sayfa.

Anahtar kelimeler: Tahminleme, Yapay Zeka, Büyük Dil Modelleri, Sana Ortam Verisi, Regresyon Analizi



ABSTRACT

Ph.D. THESIS

DEVELOPING AN DYNAMIC ESTIMATION MODEL BASED ON ARTIFICIAL INTELLIGENCE WITH ONLINE DATA

Adem ÇAYIR

İstanbul University-Cerrahpaşa

Institute of Graduate Studies

Department of Industrial Engineering

Industrial Engineering, Ph.D.

Supervisor : Assoc. Prof. Dr. Fatih TÜYSÜZ

Prediction in industrial engineering has become a critical area of research in today's data-driven world, especially with the increasing importance of sales predictions. Large data sets formed by the interactions of billions of people in the virtual environment provide valuable information about the past and have great potential for making predictions about the future. However, since a significant portion of this data is in an unstructured form, direct processing with traditional prediction methods creates various difficulties, which increases the need for more advanced and flexible analysis techniques. In this context, the thesis study prepared aims to transform the unstructured big data formed in the online environment into structured data by processing and interpreting it through large language models (LLMs) and to integrate this data as an additional feature to classical prediction models. LLMs are deep learning-based models trained on large data sets and play an important role in increasing prediction accuracy by making meaningful inferences from the data accumulated in the virtual environment thanks to their natural language processing capabilities. In this process, prediction performance has been further improved by using various machine learning algorithms such as linear regression, support vector regression, and gradient support vectors. Thus, thanks to artificial intelligence-supported data processing techniques, it has become possible to produce more reliable, consistent, and accurate

predictions by going beyond traditional methods; this has enabled more effective use of decision support systems, especially in sales prediction processes, and has significantly strengthened the integration of big data and machine learning methods into industrial processes.

July 2025, 110 pages.

Keywords: Prediction, Artificial Intelligence, Large Language Model, Online Data, Regression Analysis



1. GİRİŞ

Mühendislikte tahminleme, sistemlerin gelecekteki sonuçlarını, davranışlarını veya performansını öngörme sürecidir. Bu süreç, çeşitli hesaplamalı tekniklerin yanı sıra matematiksel modelleme ve deneysel veriler ile desteklenerek daha güvenilir hale getirilir. Tahminleme, mühendislik tasarımında kritik bir rol oynar ve karar verme mekanizmalarının daha isabetli olmasına katkı sağlar. Ayrıca, risk değerlendirme süreçlerinin geliştirilmesine yardımcı olur ve sistemlerin uzun vadeli başarısını artırmak için öngörüler sunar (Hazelrigg, 2009).

Tahminleme süreçlerinde kullanılan yöntemler genellikle belirsizlikleri minimize etmek ve sistem performansını optimize etmek amacıyla olasılıksal modeller, istatistiksel analizler ve makine öğrenimi algoritmaları ile desteklenmektedir. Örneğin, mühendisler yapısal arızaların önceden tahmin edilmesi, üretim süreçlerinin maksimum verimlilikle yönetilmesi ve sistem güvenilirliğinin artırılması gibi birçok alanda tahmin edici modellemeyi kullanmaktadır. Bu süreçlerde kullanılan yöntemler, geçmiş verilerden faydalanarak gelecekteki olayları daha doğru bir şekilde kestirmeyi mümkün kılar (Hazelrigg, 2009).

Yapay Zeka (YZ), insan zekasına özgü öğrenme, muhakeme yapma, problem çözme ve algılama gibi bilişsel görevleri yerine getirebilen sistemlerin geliştirilmesiyle ilgilenen bir bilgisayar bilimi alanıdır. YZ, zamanla gelişerek değişen koşullara uyum sağlayabilir ve performansını optimize edebilir. Bu bağlamda, makine öğrenimi, sinir ağları ve doğal dil işleme gibi alt alanları içeren YZ teknikleri, tahminleme süreçlerinde etkin bir şekilde kullanılmaktadır.

Boden (2018)'e göre Yapay Zeka, bilgisayarların insan zihnine benzer şekilde algılama, ilişkilendirme, tahmin etme, planlama ve motor kontrolü gibi bilişsel görevleri gerçekleştirmesini hedefler. Sheikh ve diğ. (2023) ise yapay zekayı, çevresini analiz edebilen ve belirli hedeflere ulaşmak amacıyla özerk eylemlerde bulunarak akıllı davranışlar sergileyen sistemler olarak tanımlamaktadır. Yapay zeka destekli tahminleme teknikleri, büyük veri analizleriyle daha doğru ve güvenilir sonuçlar elde edilmesini sağlamaktadır.

Makine Öğrenimi (MÖ), Yapay Zeka'nın önemli bir alt alanı olup, bilgisayar programlarının büyük veri kümeleri üzerinden öğrenmesini sağlayan bir yaklaşımdır. Bu sistemler, algoritmalar ve istatistiksel yöntemler kullanarak verileri analiz eder ve kalıpları tahmin etmede etkili bir şekilde çalışır. Ayrıca, öngöründe bulunmak için çeşitli matematiksel ve sezgisel modeller geliştirilir ve ilerleyen süreçlerde rehberlik edecek stratejiler belirlenir (Singh ve diğ., 2023).

Makine öğrenimi algoritmaları, kullanılan veri miktarı arttıkça daha anlamlı ve güvenilir sonuçlar verebilmektedir. Son yıllarda internetin büyük kitlelere ulaşmasıyla birlikte, çevrim içi veriler giderek artmakta ve mühendislik alanındaki tahminleme çalışmaları için önemli bir kaynak haline gelmektedir. Bu veriler, çok çeşitli formatlarda olabilir ve yapısal olmayan veri kümeleri ile çalışarak tahminleme doğruluğunu artıran teknikler geliştirilmektedir.

Bilgisayar işlem gücünün artması ve algoritmaların gelişmesiyle birlikte, tahminleme çalışmalarında Yapay Zeka ve Makine Öğrenimi teknikleri giderek daha önemli hale gelmiştir. Özellikle büyük veri setleri ile çalışan modeller, tahminleme süreçlerini hızlandırmakta ve doğruluk oranını artırmaktadır. Bu tez çalışmasında önerilen makine öğrenimi yöntemi, tahminleme modellerinin daha verimli bir şekilde geliştirilmesini amaçlamaktadır. Yapay zeka tabanlı yöntemlerle yapılan tahminleme modelleri, farklı mühendislik disiplinlerinde uygulanarak çeşitli sistemlerin optimizasyonunu sağlamaktadır.

Bu çalışma, mühendislikte tahminleme süreçlerinin daha etkin hale getirilmesi, büyük veri kullanımı ile makine öğrenimi algoritmalarının entegrasyonu ve yapay zeka tabanlı tahminleme modellerinin iyileştirilmesi açısından önemli katkılar sunmaktadır.

1.1. KATKILAR

Bu tez çalışmasında, günümüzde dijital ortamda üretilen büyük veri setlerinin tahminleme yöntemlerinde etkin ve verimli bir şekilde kullanılmasına odaklanılmıştır. Artan veri üretimiyle birlikte, bu verilerin anlamlı hale getirilmesi ve işlenerek faydalı bilgiler elde edilmesi önemli bir araştırma konusudur. Çalışmada, özellikle çevrim içi platformlardan elde edilen verilerin analitik süreçlere nasıl entegre edilebileceği ele alınmıştır.

- Günümüz dünyasında çevrim içi ortamlar üzerinden büyük miktarlarda veri üretilmekte ve bu verilerin sayısal verilere dönüştürülmesi, veri analitiği açısından kritik bir süreç

olarak görülmektedir. Bu tez kapsamında, verilerin sayısallaştırılması için yeni bir yaklaşım önerilmiş ve bu yöntemin geleneksel tekniklere kıyasla avantajları incelenmiştir.

- Özellikle büyük dil modellerinin kullanımıyla, büyük veri setlerinin daha hızlı ve verimli bir şekilde işlenebileceği ortaya konulmuştur. Yapay zeka destekli dil modelleri, doğal dil verilerini anlamlandırma ve öngörülerde bulunma konusunda önemli fırsatlar sunmaktadır. Çalışmada, bu modellerin veri madenciliği ve tahminleme süreçlerinde nasıl daha etkin kullanılabileceği gösterilmiştir.
- Perakende sektörüne yönelik gerçekleştirilen tahminleme çalışmaları kapsamında, geleneksel yöntemlere ek olarak çevrim içi platformlardan elde edilen verilerden türetilen yeni bir özellik parametresi geliştirilmiştir. Bu yeni parametre ile tahminleme algoritmalarının doğruluğunun nasıl arttığı analiz edilmiş ve elde edilen iyileştirmeler farklı makine öğrenimi algoritmaları kullanılarak değerlendirilmiştir.

Bu araştırma, çevrim içi verilerin sayısal verilere dönüştürülmesi ve makine öğrenimi teknikleriyle daha verimli hale getirilmesi konularında önemli katkılar sağlamaktadır. Elde edilen bulguların, veri bilimi ve yapay zeka alanlarında yeni uygulamaların geliştirilmesine ışık tutması beklenmektedir.

1.2. TEZİN ANA HATLARI

Bu tez çalışması, makine öğrenimi alanında yeni bir özelliğin elde edilmesi ve çeşitli algoritmalarda kullanılmasına yönelik kapsamlı bir inceleme sunmaktadır. Çalışmanın temel amacı, tahminleme süreçlerinde sanal ortamdan elde edilen verileri sayısal verilere çevirerek geleneksel makine öğrenimi özelliklerine yenilikçi bir katkı sağlamaktır. Böylece, sanal ortam verilerinden oluşturulan ek bir özellik ile makine öğrenimi modellerinin tahminleme performansının artırılması hedeflenmiştir.

Bu bağlamda, çalışmada 2022 Kasım ile 2024 Haziran ayları arasında sanal ortamda oluşan toplam 486.132 haber detaylı bir şekilde incelenmiştir. Bu haberlerin içerdiği toplam 1.739.637.269 karakter, veri işleme süreçlerine tabi tutulmuş ve makine öğrenimi algoritmalarında kullanılabilecek biçime dönüştürülmüştür. Bu verilerden elde edilen yeni

özellik üç farklı makine öğrenim algoritmasında test edilmiş ve bu algoritmaların performansına olan etkisi ortaya konmuştur.

Tezin kavramsal çerçeve bölümünde, tahminleme kavramı detaylı bir şekilde açıklanmış ve bu alanda literatürde yapılan yenilikler ele alınmıştır. Özellikle son yıllarda yapay zeka ve makine öğrenimi alanında gerçekleşen gelişmeler incelenmiş, büyük dil modellerinin evrimi ve bunların tahminleme süreçlerindeki rolü tartışılmıştır. Bunun yanı sıra, makine öğrenimi algoritmalarındaki güncel ilerlemeler ve yenilikler listelenerek araştırmaya sağlam bir teorik temel oluşturulmuştur. Yöntem bölümünde, önerilen yaklaşımın teknik detaylarına yer verilmiş ve verilerin toplanması, işlenmesi ile oluşturulan yeni özelliğin makine öğrenimi modellerindeki etkisi değerlendirilmiştir. Üç farklı makine öğrenim algoritması üzerinde yapılan deneyler sonucunda, sanal ortamdan türetilen bu yeni özelliğin tahminleme doğruluğuna ne derece katkı sağladığı analiz edilmiştir. Bulgular bölümünde ise elde edilen sonuçlar detaylı bir şekilde incelenmiş, klasik tahminleme yöntemleri ile önerilen yeni yaklaşım arasındaki farklar açıklığa kavuşturulmuştur. Tartışma bölümünde, elde edilen sonuçlar güncel literatürdeki benzer çalışmalarla karşılaştırılmış ve araştırmanın bilimsel anlamda sağladığı katkılar irdelenmiştir. Son olarak, sonuç ve öneriler bölümünde çalışmanın genel değerlendirmesi yapılmış ve gelecekte benzer çalışmaların nasıl daha fazla iyileştirilebileceğine yönelik öneriler sunulmuştur. Böylece, makine öğrenimi alanında sanal ortam verilerinin tahminleme süreçlerine nasıl entegre edilebileceği konusunda yeni bir bakış açısı kazandırılmıştır.

2. KAVRAMSAL ÇERÇEVE

Bu bölümde önerilen yöntemin ana hatlarını oluşturan tahminleme ve yapay zeka yöntemleri açıklanacaktır. Ayrıca önerilen yöntemin performansını ölçme yöntemlerine değinilecektir.

Sanal ortam verileri ve sosyal medya internet teknolojileri üzerinde sosyal ve teknolojik içeriklerin, yapılanmaların kullanıcı merkezli bir şekilde üretilmesine ve geliştirilmesine izin veren internet tabanlı uygulamaların bütününe denir (Kaplan ve Haenlein, 2010). Hızla büyümekte olan bu içerik kullanılarak iş dünyasında çok önemli çıkarımlar yapılabilmektedir. Kişilerin gerçek olaylara karşı duygu ve tepkileri, sonraki zamanlarda oluşacak olaylar için ipuçları içermektedir. Oluşan içerik, muazzam derecede büyüktür. Aynı zamanda, her kullanıcı için farklı ve bir kurala bağlı olarak oluşturulmadığı için, yapısal durumda olmamaktadır. Dolayısıyla büyük miktarda bulunan yapısal olmayan içerikler sistematik bir şekilde incelenmelidir.

Du ve diğ. (2014) yaptığı çalışmada Sosyal Medya verisi bilgi kaynağı olarak kullanılmış olup, Sosyal medya üzerinde iki türlü bilgi mevcut olduğu tespit edilmiştir. Sayılara dayanan ve içeriğe dayanan. Kullanıcıları baz alarak yapılan çalışmalar önemli görülse de suistimale açıktır. İçerik tabanlı özellikler eklenerek gişe hasılatı tahmininde daha tutarlı sonuçlar bulunabilmektedir. Çalışma öncelikle sosyal ağdaki önemli kişileri bulmaya odaklanmıştır. Daha sonrasında içerikleri analiz etmiştir. Li ve diğ. (2020) ise çevirim için haberlerdeki duygu analizini yaparak hisse tahmin çalışması yapmıştır.

2.1. TALEP TAHMİNİ

Talep tahmini, modern iş dünyasında hayati bir rol oynayan, karmaşık ancak vazgeçilmez bir süreçtir. En basit tanımıyla, talep, belirli ürünlerin belirli pazarlarda tüketicilerin satın alma potansiyeline denir. Talep tahmini ise, bu satın alma potansiyelinin bilimsel ve sistematik yöntemlerle ölçülmesi anlamına gelir (Ferrer ve diğ., 2012). Bu süreç, özellikle gelir yönetimi açısından kritik bir öneme sahiptir, zira doğru tahminler, kaynakların verimli tahsis edilmesinden rekabet avantajı elde etmeye kadar birçok alanda işletmelere yol

gösterir. Winston (1994) tarafından tanımlanan talep tahmini, tüketicilerin ilerleyen dönemlerde ne kadar mal ve hizmet talep edeceklerini öngörme sürecidir. Bu tahminler, işletmelerin üretim düzeylerini belirlemede kritik bir rol oynar. Hangi ürünlerin üretileceği, tüketicilerin bu ürünlere ne ölçüde ilgi göstereceği ve talebin yoğunlukla hangi zaman dilimlerinde yoğunlaşacağı, talep tahmini sayesinde değerlendirilebilir.

Talep tahmini, doğası gereği hem belirsizliği barındırır hem de öngörü gerektirir. Bu nedenle, çeşitli kalitatif (nitel) veya kantitatif (nicel) yöntemler kullanılarak modellenir. Perakende sektörü gibi dinamik alanlarda, geçmiş çalışmalar, özellikle üssel talep modellerinin doğrusal modellere kıyasla daha iyi sonuçlar verdiğini göstermektedir (Bitran ve Caldentey, 2003; Smith ve diğ., 1994). Bu, sektörün hızlı değişen tüketici davranışlarına ve mevsimsel dalgalanmalara daha duyarlı olmasından kaynaklanmaktadır.

2.1.1. Kalitatif Talep Tahmin Yöntemleri

Kalitatif yöntemler, nesnel verilerden çok uzmanların sezgilerine ve kişisel değerlendirmelerine dayanır. Bu tür yaklaşımlara özellikle, geçmişe dönük verilerin yetersiz olduğu, pazara yeni bir ürünün sunulduğu veya piyasa koşullarının radikal bir değişim içinde olduğu durumlarda başvurulur.

Bu yöntemlerin sunduğu önemli avantajlar vardır. Karmaşık hesaplamalar veya istatistiksel modelleme gerektirmemeleri sayesinde hızlı bir şekilde uygulanabilir ve çabuk sonuç üretirler. Bu durum, ileri düzeyde teknik bilgi, büyük veri setleri veya özel bir yazılım ihtiyacını ortadan kaldırır. Bunun yerine mevcut deneyimlerden ve kolayca erişilebilen bilgilerden beslenirler. Nitekim yapılan araştırmalar, uzmanların asıl gücünün yeni değişkenleri saptama, "kırık bacak ipucu" gibi anormal ve beklenmedik ipuçlarını yakalama ve sayısal olarak ifade edilmesi güç bilgileri değerlendirme yeteneklerinde olduğunu göstermektedir (Blattberg ve Hoch, 1990; Einhorn, 1974). "Kırık bacak ipucu", Meehl (1957)'e göre gelişmiş bir istatistiksel model bir profesörün sinemaya gideceğini yüksek doğrulukla tahmin etse bile, modelin hesaba katmadığı "kırık bacak" gibi istisnai bir olay gerçekleşirse oluşturulan model artık geçersiz olur, zira profesörün bacağı kırılırsa sinemaya gitmeyeceği modelin sonucundan bağımsız olarak kesindir. Kısacası bu kavram, nadir fakat kesin gerçeklerin en isabetli görünen algoritmik öngörülerini bile geçersiz kılabileceğini anlatır. Bu bağlamda, uzmanlara bir tahmin durumuyla ilgili daha fazla bağlamsal bilgi (örneğin, promosyon faaliyetleri, üretim kapasitesi,

rakip stratejileri veya makroekonomik beklentiler gibi) sunmak, onların daha isabetli öngörülerde bulunmasını sağlayabilir (Lawrence ve diğ., 2006).

Diğer yandan, kalitatif yöntemlerin dikkate değer zayıflıkları da bulunur. Uzmanların veya grupların görüşleri, piyasanın genelini her zaman tam olarak yansıtmayabilir ve bu durum tahminlerde sapmalara neden olabilir. Ayrıca, kararların kolektif olarak alındığı durumlarda, hatalı bir tahminin sorumluluğunu belirli bir kişiye veya gruba yüklemek zorlaşır. Başlıca kalitatif yöntem örnekleri aşağıda sıralanmıştır.

2.1.1.1. Delphi Tahmin Yöntemi

Delphi tekniği, "iki kafa bir kafadan iyidir" prensibine dayanarak, yapılandırılmamış gruplar yerine iyi yapılandırılmış bir uzman grubunun yargılarının daha isabetli olacağı varsayımını temel alır. Bu yöntemde, uzmanlardan oluşan bir panele hazırlanan bir anket sunulur ve süreç iki veya daha fazla turda tekrarlanır. Her turun ardından, uzmanların sunduğu anonim yanıtlar ve bu yanıtlara getirdikleri gerekçeler özetlenir. Bu özet, panel üyelerine geri bildirim olarak verilir; böylece uzmanlar, diğer üyelerin tahminleri ışığında kendi önceki tahminlerini gözden geçirerek daha doğru bir sonuca ulaşmaya çalışır (Kiran, 2022).

Delphi tahmin yöntemi doğası gereği kalitatif olduğu için belirli bir matematiksel formülasyonu yoktur. Ancak, bazı çalışmalarda yöntemin formülize edildiği görülmüştür. Örneğin, Franc ve diğ. (2023) uzmanlar arasındaki fikir birliği veya ayrışma derecesini göstermek için standart sapmanın kullanılabileceğini ve daha küçük bir standart sapmanın daha yüksek bir fikir birliği anlamına geldiğini belirtmiştir. Benzer şekilde, uzman görüşlerinin kesin olmadığı ve bir değer aralığı ile daha iyi ifade edildiği durumlar için Chang ve diğ. (2000) bulanık kümeler teorisini uygulayan bulanık Delphi yöntemini geliştirmiştir.

2.1.1.2. Satış Gücü Grupları Tahmin Yöntemi

Satış gücü tahmin birleştirme yöntemi, talep tahminlemesi alanında sıkça başvurulan kalitatif bir yaklaşımdır ve temel olarak sahadaki satış elemanlarının bireysel öngörülerinin toplanıp birleştirilmesiyle bir bütün olarak talep öngörüsü oluşturulmasına dayanır. Bu metodun arkasındaki temel mantık, müşterilerle sürekli ve doğrudan temas halinde olan satış personelinin, pazarın güncel dinamiklerini, müşteri eğilimlerini ve potansiyel talep değişikliklerini en iyi hisseden kişiler olduğudur. Hızlı ve kolay uygulanabilir olması, bu yöntemi oldukça pratik kılmakta ve en önemli üstünlüğünü oluşturmaktadır. Bununla birlikte, yöntemin dikkate alınması gereken önemli bir zayıflığı vardır: satış elemanlarının kişisel

hedefleri, prim beklentileri gibi motivasyon kaynakları veya yapısal olarak iyimser ya da kötümser olmaları, sundukları tahminlerde ciddi bir yanlılığa sebep olabilir ve bu da sonuçların doğruluğunu riske atabilir. Bu potansiyel dezavantajına rağmen, Diamantopoulos ve Winklhofer'in (2003) yaptığı araştırmanın da altını çizdiği gibi, satış gücü grupları tahmin yaklaşımı, ihracat satış tahminlerini geliştirmek amacıyla kullanılan kalitatif yöntemler arasında (tüm organizasyonel düzeylerde ve zaman ufuklarında) en popüler teknik olma özelliğini korumaktadır.

2.1.1.3. Yönetici Görüşleri Tahmin Yöntemi

Yönetici görüşleri tahmin yöntemi, talep tahmini oluşturmak amacıyla, genellikle pazarlama, finans, üretim ve satış gibi farklı kurumsal işlevlerden gelen üst düzey yöneticilerin bilgi ve deneyimlerini bir araya getirmesini gerektiren, en sık kullanılan kalitatif tahmin tekniklerinden biridir. Bu süreçte gerçekleştirilen ve "yönetici görüşü jürisi" olarak adlandırılan toplantının temel amacı, farklı departmanların bakış açılarını entegre ederek daha bütüncül bir öngörüye ulaşmaktır (Mital ve diğ., 2008). Yöntemin etkinliği için tahmine ihtiyaç duyan tüm temel işlevlerin bu jüride temsil edilmesi ve seçilen temsilcilerin kendi alanlarında tecrübeli, stratejik vizyona sahip kişiler olması kritik önem taşır. Bu sayede ortaya çıkan nihai tahmin, sadece bireylerin kişisel deneyimlerinin bir yansıması olmakla kalmaz, aynı zamanda farklı uzmanlıkların müzakeresi sonucu oluşan bir fikir birliğini (konsensüs) de temsil eder.

Bu yöntemin pratik uygulaması bazı avantaj ve dezavantajları beraberinde getirir. Yönetici jürilerinde, katılımcıların kişisel varsayımlarından veya departmansal önceliklerinden kaynaklanabilecek doğal önyargıyı en aza indirmek amacıyla, toplantı öncesinde tüm üyelere geçmiş veriler, pazar analizleri ve ekonomik göstergeler gibi objektif arka plan bilgilerinin sağlanması tavsiye edilir. Yöntemin en büyük avantajı, stratejik düzeydeki bir tahmini diğer yöntemlere kıyasla çok daha hızlı ve kolay bir şekilde oluşturabilmesidir. Buna karşın en belirgin dezavantajı ise, kararların kolektif bir gruba ait olması sebebiyle sorumluluğun da grupça paylaşılmasıdır; bu durum, tahminin başarısı veya başarısızlığı durumunda bireysel hesap verebilirliği önemli ölçüde azaltarak performans değerlendirmesini zorlaştırır.

2.1.1.4. Pazar Araştırması Tahmin Yöntemi

Pazar araştırması tahmin yöntemi, gelecekteki talep beklentileri hakkında doğrudan pazarın kendisinden bilgi üretmek için kullanılan son derece etkili ve çok yönlü bir tahmin aracıdır. Bu yöntem, şirketin stratejik ihtiyaçlarına göre esneklik göstererek hem aylık ve üç aylık gibi kısa ve orta vadeli satış tahminleri oluşturmaya hem de bir ila beş yıllık periyotları

kapsayan uzun vadeli projeksiyonlar yapmaya yardımcı olabilir. Ayrıca, anlık pazar değişimlerine tepki vermek amacıyla kısa vadeli satış ayarlamaları için de kullanılabilir (Mital ve diğ., 2008). Beklenen satışlara ilişkin veriler oldukça çeşitli kanallardan toplanabilir. Bir şirketin ticari veya kurumsal müşterileriyle yapılan yüz yüze görüşmeler, telefon anketleri veya posta yoluyla gönderilen anketler bu kanallara örnektir. Benzer şekilde, bağımsız dağıtımıcılar veya nihai tüketicilerden oluşan bir örneklem de veri kaynağı olarak kullanılabilirken, odak grupları da tüketicilerin derinlemesine görüşlerini toplamak için sıkça başvurulmuş bir başka tekniktir.

Bu yöntemin en temel avantajı, potansiyel alıcıların niyetlerini, fikirlerini ve beklentilerini doğrudan öğrenme imkânı sunmasıdır. Bu birinci elden elde edilen veriler, özellikle yeni ürün geliştirme süreçleri ve pazarlama stratejilerinin şekillendirilmesi için paha biçilmez değerde içgörüler sağlar. Ancak, bu değerli veriyi elde etmenin önemli dezavantajları da vardır; pazar araştırması genellikle maliyetli, ciddi zaman alan ve istatistiksel olarak anlamlı sonuçlar üretebilmek için geniş bir örneklem gerektiren bir faaliyettir. Geleneksel yöntemlerin bu zorluklarına bir alternatif olarak günümüzde sosyal medya verileri de önemli bir bilgi kaynağı haline gelmiştir. Sosyal medya platformlarındaki kullanıcı yorumları ve paylaşımları üzerinde yapılan duygu analizi, geleneksel anketlerin sınıflandırılmış seçeneklerini otomatize ederek, tüketicilerin genel eğilimleri hakkında çok daha geniş çaplı ve gerçek zamanlı bilgi sağlayabilir ve böylece geleneksel yöntemin bazı kısıtlarını ortadan kaldırabilir.

2.1.1.5. Tarih Analoji Tahmin Yöntemi

Tarihsel analoji (benzetme) tahmin yöntemi, yeni bir ürünün talep tahminini, ona benzer daha eski ürün veya hizmetlerin geçmiş satış verilerini inceleyerek yapmaya odaklanır. Bu yaklaşım, çeşitli analizlerle desteklenir; örneğin Elton (1991), bu yöntemin üç temel incelemeye dayandırılır. Geçmişteki benzer yeni ürünlerin büyüme verilerindeki düzenliliklerin analizi, hane halkı harcama alışkanlıkları ve gelecekteki gelir beklentilerine dayanan gelir-gider analizi ve talebin rakip veya tamamlayıcı ürünlerden nasıl etkileneceğini inceleyen rekabetçi analiz. Pratikte bu, pazara ilk kez sunulacak yeni bir akıllı telefon modelinin potansiyel talebini tahmin etmek için, daha önce piyasaya sürülmüş benzer akıllı telefon modellerinin satış verilerinin bir model olarak kullanılması anlamına gelir.

2.1.2. Kantitatif Talep Tahmin Yöntemleri

Kalitativ yöntemlerin dışında, talep tahminlerinde kantitatif yöntemler de yoğun olarak kullanılır. Kantitatif yöntemlerde, geçmişe ait sayısal veriler çeşitli matematiksel ve istatistiksel

modellerde işlenir. Bu yöntemler, nesnel ve tekrarlanabilir sonuçlar sunmaları nedeniyle özellikle büyük veri setlerinin mevcut olduğu durumlarda tercih edilir.

Kantitatif yöntemlerin kullanılabilmesi için, geçmiş satış bilgilerinin erişilebilir ve bu bilgilerin sayısal bir şekilde olması gerekmektedir. Bu bilgiler, genellikle zaman serileri veya diğer bağımsız değişkenlerle ilişkili veriler olabilir. Kantitatif yöntemler iki ana grupta incelenir: Zaman Serileri Yöntemleri ve Nedensel (Ekonometri) Yöntemler.

Önemli bir nokta olarak, kalitatif bir veri olan sosyal medya verisi, duygu analizi gibi tekniklerle sayısal verilere dönüştürülüp, daha sonra kantitatif talep tahminleri yöntemlerinden uygun olanında kullanılabilir. Bu, modern tahminleme yaklaşımlarının önemli bir yönüdür.

Kantitatif yöntemlerin bazıları zaman serileri yöntemleri içerir. Bu yöntemler, bir değişkenin (örneğin talep) geçmişteki değerlerinin zaman içindeki desenlerini ve eğilimlerini analiz ederek gelecekteki değerlerini tahmin etmeye odaklanır.

2.1.2.1. Naive Tahmin Yöntemi

Naive tahmin yöntemi, hem endüstriyel hem de finansal (özellikle kısa vadeli finansal getirilerde) alanlarda yaygın olarak kullanılan en temel tahmin modelidir. Bu yöntemde, bir sonraki döneme ilişkin tahmin, en son elde edilen gerçek veri noktasının kendisine veya bu değere belirli bir oranda yakın bir değere eşit kabul edilir (Nikolopoulos, 2020).

Bu yaklaşımın en önemli avantajları, hızlı bir şekilde uygulanabilmesi ve zaman serileri için başlangıç düzeyinde bir tahmin oluşturmayı kolaylaştırmasıdır. Ancak, önemli bir dezavantajı da bulunmaktadır: Zaman serilerindeki talep gibi değerlerde ciddi dalgalanmalar yaşanabileceği için, Naive yöntemi her zaman güvenilir sonuçlar vermeyebilir ve ani piyasa değişikliklerini öngörmede yetersiz kalabilir.

$$F_{t+1} = Y_t \quad (2.1)$$

Denkle (2.1)'deki gibi ifade edilebilir. A , Mevcut veya en son dönemdeki (t) gerçekleşen (aktüel) değeri, F , Gelecek dönem ($t + 1$) için yapılan tahmini belirtir.

2.1.2.2. Ortalama Bazlı Tahmin Yöntemleri

Üç çeşit ortalama tabanlı kantitatif talep tahmin yöntemi vardır. Basit ortalama yöntemi, geçmiş dönem verilerinin basit aritmetik ortalaması alınarak sonraki dönem tahmin edilir. Tüm geçmiş verilere eşit ağırlık verir. Hareketli ortalama yöntemi ise*- bütün geçmiş verilerin

kullanılması yerine, daha yakın zamanda gerçekleşen talepler kullanılarak ortalama bulunur. Örneğin, son 4, 5 veya 6 dönemin verileri kullanılarak talep tahmini yapılır. Bu yöntem, eski verilerin etkisini azaltarak daha güncel bilgilere odaklanmayı sağlar. Ağırlıklı hareketli ortalama yöntemi, hareketli ortalama yönteminde son dönemde gerçekleşen değerler direkt kullanılmıştı. Ağırlıklı hareketli ortalama yönteminde ise gerçekleşen değerlere farklı ağırlıklar vererek hesaplama yapılır. Buradaki amaç, en yeni satış dönemlerine daha yüksek ağırlıklar vererek daha doğru ve güncel tahminler yapmaktır.

Basit ortalama formülü şu şekilde ifade edilebilir.

$$F_{t+1} = \frac{Y_1 + Y_2 + \dots + Y_t}{t} = \frac{\sum_{i=1}^t Y_i}{t} \quad (2.2)$$

Burada F_{t+1} , gelecek dönem ($t + 1$) için yapılan tahmini, Y_i , i dönemindeki gerçekleşen (aktüel) değeri, t , toplam geçmiş dönem sayısını ifade eder.

Hareketli ortalama için bu formül kullanılabilir.

$$F_{t+1} = \frac{Y_t + Y_{t-1} + \dots + Y_{t-n+1}}{n} = \frac{\sum_{i=t-n+1}^t Y_i}{t} \quad (2.3)$$

F_{t+1} , gelecek dönem ($t+1$) için yapılan tahmini, Y_i , i dönemindeki gerçekleşen (aktüel) değeri, n ise ortalamaya dahil edilecek dönem sayısını (örneğin, son 3 ay için $n=3$) ifade eder.

Ağırlıklı hareketli ortalama için,

$$F_{t+1} = \omega_1 Y_t + \omega_2 Y_{t-1} + \dots + \omega_n Y_{t-n+1} = \sum_{i=1}^n \omega_i Y_{t-i+1} \quad (2.4)$$

Formülü kullanılır. Burada F_{t+1} , gelecek dönem ($t+1$) için yapılan tahmini, Y_i , i dönemindeki gerçekleşen (aktüel) değeri, ω_n ilgili döneme atanan ağırlığını, n ise ortalamaya dahil edilecek dönem sayısını ifade eder.

2.1.2.3. Üssel Düzeltme Tahmin Yöntemleri

Geleneksel hareketli ortalama yönteminde, bir sonraki değeri öngörmek için geçmişteki belirli sayıda gözlemin aritmetik ortalaması kullanılır. Bu yaklaşım, geçmişteki tüm verilere eşit derecede ($1/n$) önem atar. Ancak tahmin yaparken, genellikle en güncel veriler geleceğe dair en isabetli ipuçlarını barındırır. Bu nedenle, daha yeni gözlemlere daha fazla ağırlık veren

bir yöntem daha başarılı sonuçlar üretebilir. Bu noktada, gözlemler eskidikçe ağırlıklarının üssel olarak azaldığı üssel düzeltme yöntemleri etkili bir çözüm sunar (Cadenas ve diğ. 2010).

Bu yöntemin temelini anlamak için öncelikle tahmin hatasını tanımlayalım. Tahmin hatası (e_t), belirli bir dönemdeki gerçek değer (Y_t) ile o dönem için yapılan tahmin (F_t) arasındaki farktır:

$$e_t = Y_t - F_t \quad (2.5)$$

Üssel düzeltme yöntemi, bir sonraki dönem tahminini, mevcut dönem tahminini son tahmin hatasıyla düzelterek oluşturur. Bir sonraki dönem için tahmin (F_{t+1}) şu formülle hesaplanır:

$$F_{t+1} = F_t + \alpha(Y_t - F_t) \quad (2.6)$$

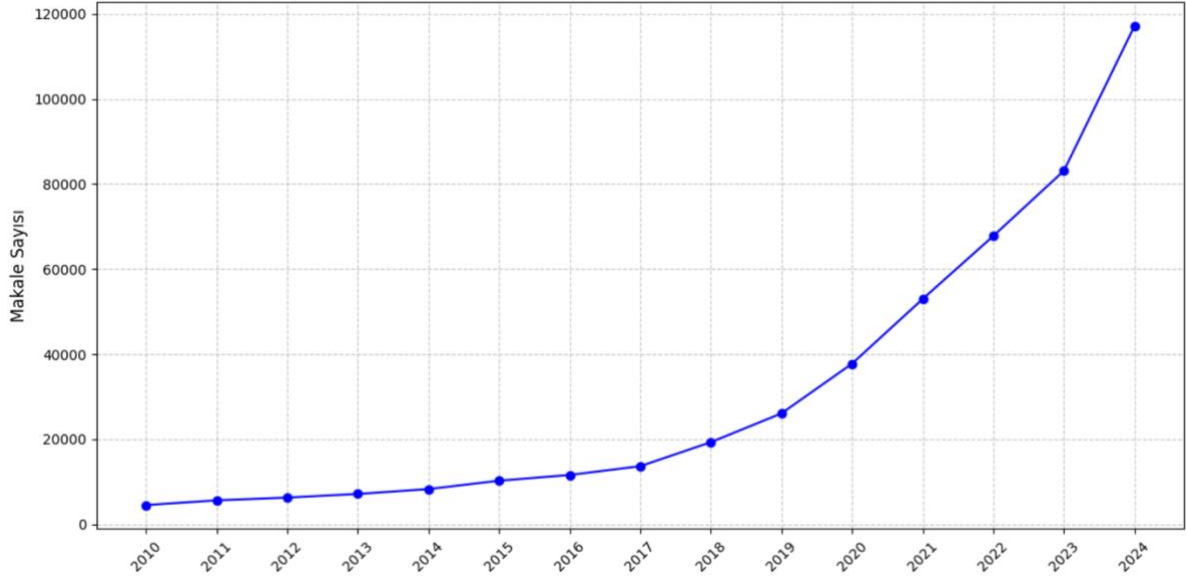
Bu formüldeki α katsayısının değeri, düzeltmenin ne ölçüde yapılacağını belirler. Eğer α değeri 1'e yakınsa, son tahmin hatasının büyük bir kısmı yeni tahmine yansıtılır ve model, güncel değişimlere hızla uyum sağlar. Tersine, α değeri 0'a yakın olduğunda ise tahmin, son hatadan daha az etkilenir ve daha yavaş bir düzeltme yapılır; bu da daha istikrarlı ve yumuşatılmış bir tahmin çizgisi oluşturur.

2.2. YAPAY ZEKA VE MAKİNE ÖĞRENİM YÖNTEMLERİ

Yapay Zeka (YZ), büyük ölçekli problemleri makine yardımıyla çözmeye yönelik yöntemlerin genel adıdır. Çeşitli bilim insanları tarafından farklı şekillerde tanımlanmıştır. Örneğin, YZ araştırmalarının önemli öncülerinden Marvin Minsky, YZ'yi makinelerle insan zekasını gerektiren görevleri yaptırma süreci olarak tanımlamıştır (Minsky ve Papert, 1990). Alan Turing'e göre YZ, makinelerin insanlarla doğal bir şekilde etkileşimde bulunurken insan olmadıklarını açıklamama yeteneğini ifade eder; bu kavram Turing Testi olarak bilinir (Kaynak, 2021). Yapay zeka terimi ise ilk kez John McCarthy tarafından 1955'teki Dartmouth Konferansı'nda kullanılmış, ve bu konferans YZ'nin yeni bir çalışma alanı olarak ortaya çıkmasına öncülük etmiştir (McCarthy, 1955).

Literatüre bakıldığında, YZ ile ilgili ilk çalışmalar 1950'li yıllarda başlamış olmasına rağmen, bilgisayar teknolojilerinin gelişmesiyle birlikte özellikle 2000'li yıllarda önemli ilerlemeler kaydedilmiştir. Son yıllarda bilgisayar kapasitelerinin artmasıyla YZ tabanlı çalışmaların sayısında hızlı bir yükseliş gözlemlenmektedir. Şekil 2.1'de görüleceği üzere 2010

yılında 5000 civarında olan makale sayısı, 2024'e geldiğimizde 120.000'ler seviyesine ulaşmıştır.



Şekil 2.1: ScienceDirect’de bulunan Yapay Zeka ve Makine Öğrenimi Makale Sayıları

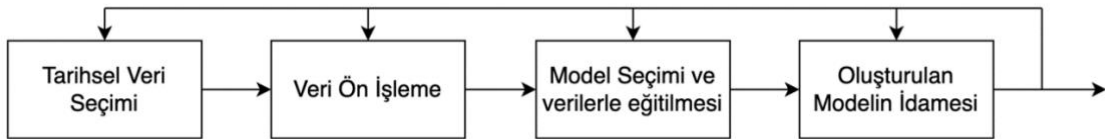
Yapay zeka ve makine öğrenimi (MÖ) kavramları, bilgisayarların insanların öğrenme ve karar verme süreçlerini taklit edebilmesi amacıyla yirminci yüzyılda ortaya çıkmıştır. O zamandan bu yana, veri madenciliği ve yapay zekâ birçok farklı araştırma alanında giderek daha fazla önem kazanmıştır (Entezari, 2023).

Büyük veri kümelerini analiz etmede oldukça başarılı sonuçlar veren yapay zekâ, özellikle tahminleme yöntemlerinde yaygın olarak kullanılmaktadır. Yapay zekâ tabanlı tahminleme tekniklerini içeren çeşitli çalışmalar mevcuttur. Örneğin, Ghosh ve diğ. (2022) hisse senedi fiyatları üzerine yaptıkları çalışmada yapay zekâ destekli tahminleme yöntemlerini kullanarak elde ettikleri sonuçları paylaşmışlardır.

Avrupa Birliği Komisyonu'nun 2020 yılındaki değerlendirme raporunda, YZ alanındaki araştırmaların büyük bir hızla ilerlediği belirtilmiştir. Yapay zekanın sağlık sektörü, tarım, iklim değişikliğine hazırlık, tahminleme süreçleri ve üretim sistemlerinin verimliliğinin artırılması gibi pek çok alanda hayatımızı değiştireceği öngörülmektedir. Ayrıca, Avrupa'nın güvenliği açısından da önemli bir rol oynayacağı ifade edilmektedir (European Commission, 2020).

Makine Öğrenimi (MÖ), yapay zeka uygulamalarından biri olup, bilgisayar programlarının belirli görevleri çözmek için ürettiği algoritmalara dayanır. Yıllarca süren çalışmalar sonucunda bazı problemler için kesin algoritmalar geliştirilememiştir. İnsan zekası, fotoğraftaki bir kişiyi tanıma, kalabalık bir odada kimseye çarpmadan ilerleme, araç sürme veya yabancı bir dilde sohbet etme gibi görevleri içgüdüsel olarak gerçekleştirebilir. Makine öğrenimi ise bu tür görevleri makinelerle öğretmeyi amaçlar. Birçok parametre içeren bir model oluşturularak, belirlenen göreve uygun çıktılar üretmesi sağlanır. Öğrenme süreci, modelin elde ettiği veriler doğrultusunda parametre değerlerini en iyi şekilde ayarlamasıyla gerçekleşir (Alpaydin, 2014). Makine öğrenimi genel olarak 4 adımda işler (Soares, 2015):

- Tarihsel Veri Seçimi: Oluşturulacak model için değerli verileri seçmek amacıyla verilerin nasıl toplandığını ve saklandığını belirlemeyi amaçlar.
- Veri Ön İşleme: Öğrenme modeli tarafından etkili bir şekilde işlenebilmeleri için verilerin işlenmesi ve dönüştürülmesi adımıdır. Veri ön işleme genellikle veri dönüştürme, veri temizleme ve veri azaltma adımlarını içerir.
- Makine Öğrenimi Model Seçimi ve Eğitilmesi: Uygun bir makine öğrenimi modelinin seçilerek verilerle eğitilmesi sürecidir.
- Oluşturulan Modelin İdamesi: Eğitilmiş modelin performansının sürekli olarak izlenmesi ve gerekli güncellemelerin yapılmasıdır.



Şekil 2.2: Makine öğrenimi adımları (Soares, 2015)

Makine Öğrenimi ile özellikle tahminleme konusunda birçok çalışma yapılmış ve bu çalışmaların sayısı yıllar içinde hızla artmıştır. Finansal alanda da çok sayıda uygulama bulunmaktadır. Örneğin, Jones ve diğ. (2015) kredi oranlarının tahmininde makine öğrenimini kullanan bir yöntem önermişlerdir. Dash ve Dash (2016) ise yaptıkları çalışmada birkaç makine öğrenimi algoritması kullanarak hisse senetleri için "satın al", "sat" ve "beklet" sınıflarına göre hisseleri sınıflandıran ve karar verme aracı olarak kullanılabilen bir sistem geliştirmişlerdir.

O'Hara (2015) ve Chaboud ve diğ. (2014) gibi araştırmacılar da, oldukça fazla referans alan çalışmalarında makineler yardımı ile hisse senedi borsalarında hızlı işlem yapan sistemler geliştirmişlerdir.

Mintarya ve diğ. (2023) tarafından makine öğreniminin finans uygulamalarında kullanımına yönelik literatür taramasına göre, finans uygulamalarında yapay sinir ağları (Neural Network veya NN) algoritması diğer algoritmalara göre oldukça fazla kullanılmaktadır. Onları uzun kısa süreli bellek (Long Short-Term Memory veya LSTM) ve destek vektör makineleri (Support Vector Machine veya SVM) algoritmaları takip etmektedir.

2.2.1. Doğrusal Regresyon

Regresyon terimi ve iki değişken arasındaki ilişkileri inceleyen yöntemlerin kökenleri yaklaşık 100 yıl öncesine dayanmaktadır. Bu kavram ilk olarak 1908 yılında ünlü İngiliz biyolog Francis Galton tarafından kalıtım üzerine yaptığı çalışmalar sırasında ortaya atılmıştır (Galton, 1886). Galton'ın gözlemlerinden biri, uzun boylu ebeveynlerin çocuklarının ortalamadan daha uzun olmalarına rağmen, ebeveynleri kadar uzun olmamalarıydı. Bu "sıradanlığa doğru regresyon" fenomeni, bu istatistiksel yöntemlere adını vermiştir.

Regresyon terimi ve onun evrimi, temel olarak değişkenler arasındaki istatistiksel ilişkileri tanımlar. Özellikle, basit regresyon, bir bağımlı değişken (y) ile bir bağımsız değişken (x) arasındaki ilişkiyi incelemek için kullanılan bir regresyon yöntemidir.

Aşağıdaki klasik veri seti, ebeveynlerin ve çocukların boylarına ilişkin bilgileri içermektedir (Yan ve Su, 2009).

Tablo 2.1: Ebeveyn Boyları ve Çocuk Boyları

Ebeveyn Boyu	Çocuk Boyu
64,5	65,8
65,5	66,7
66,5	67,2
67,5	67,6
68,5	68,2
69,5	68,9
70,5	69,5
71,5	69,9
72,5	72,2

Bu veri setine göre, çocuklar için ortalama boy 68,44, ebeveynler için ise 68,5'tir. Ebeveyn ve çocuk boyları arasındaki regresyon çizgisi şu şekilde tanımlanabilir:

$$\text{çocuk boyu} = 21,52 + 0,69 \text{ ebeveyn boyu} \quad (2.7)$$

Basit doğrusal regresyon modeli tipik olarak şu şekilde ifade edilir:

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (2.8)$$

Burada:

- y : Bağımlı değişken (aynı zamanda yanıt değişkeni olarak da adlandırılır)
- β_0 : y kesişimi (regresyon çizgisinin y eksenini kestiği nokta)
- β_1 : Basit doğrusal regresyon çizgisinin eğimi
- x : Bağımsız değişken (aynı zamanda açıklayıcı değişken veya öngörücü değişken olarak da adlandırılır)
- ε : Rastgele hata

Açıklayıcı değişken (x), yanıt değişkenindeki (y) nedensel değişiklikleri açıklar. Bir regresyon modelinin daha genel bir gösterimi şu şekilde yazılabilir:

$$y = E(y) + \varepsilon \quad (2.9)$$

Burada $E(y)$ yanıt değişkeninin matematiksel beklentisidir. Eğer $E(y)$, keşif değişkenleri $x_1, x_2, \dots, x_k$ 'nın doğrusal bir kombinasyonuysa, regresyon doğrusal regresyondur. Eğer $k = 1$ ise, regresyon basit doğrusal regresyondur. Eğer $E(y), x_1, x_2, \dots, x_k$ 'nın doğrusal olmayan bir fonksiyonuysa, regresyon doğrusal olmayan regresyondur.

Hata terimi (ε) üzerindeki klasik varsayımlar şunlardır:

- Hatanın beklenen değeri sıfırdır: $E(\varepsilon_i) = 0$
- Hatanın sabit bir varyansı vardır: $Var(\varepsilon) = \sigma^2$
- Tüm ε_i 'ler bağımsızdır.

σ^2 'nin gerçek değeri genellikle bilinmemektedir. Bağımsız değişkenlerin (x_i) değerleri, hiçbir ölçüm hatası olmadan "tam olarak" ölçüldüğü varsayılır.

Basit doğrusal regresyon için tipik bir deneyde, bir bilimsel deneyden n çift veri $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ gözlemlenir ve bu n çift veri açısından model şu şekilde kurulur:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (2.10)$$

$i = 1, 2, \dots, n$ için, $E(\varepsilon_i) = 0$, sabit bir varyans $Var(\varepsilon) = \sigma^2$ ve tüm ε_i 'ler bağımsızdır.

Model belirlendikten ve veriler toplandıktan sonraki adım, bilimsel bir deneyden gelen verileri en iyi şekilde tanımlayabilen basit doğrusal regresyon modeli için β_0 ve β_1 'in "iyi" tahminlerini bulmaktır. Bu tahminleri türetmek ve istatistiksel özelliklerini belirlemek için en küçük kareler yöntemi kullanılır.

2.2.1.1. En Küçük Kareler Yöntemi

Basit doğrusal regresyonda, En Küçük Kareler (EKK) yönteminin temel amacı, gözlemlenen tepki değerleri y_i ile tahmin edilen tepki değerleri $\hat{y}_i = \beta_0 + \beta_1 x_i$ arasındaki karesel farkların toplamını minimize eden β_0 ve β_1 regresyon katsayılarının en iyi tahminleri olan b_0 ve b_1 'i bulmaktır. Matematiksel olarak bu, aşağıdaki ifadeyle gösterilir:

$$(b_0, b_1) = \arg \min_{(\beta_0, \beta_1)} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2 \quad (2.11)$$

Bu yöntemin ardındaki motivasyon, tüm veri noktalarına (x_i, y_i) "en yakın" olan regresyon doğrusunu seçerek parametre tahminlerini elde etmektir. Basit doğrusal regresyon için en küçük kareler tahminleri, aşağıdaki denklemleri çözerek bulunur:

$$\frac{\partial}{\partial \beta_0} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2 = 0 \quad (2.12)$$

$$\frac{\partial}{\partial \beta_1} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2 = 0 \quad (2.13)$$

b_0 ve b_1 'in yukarıdaki sistemin çözümleri olduğunu varsayarsak, x ve y arasındaki ilişkiyi, geleneksel olarak **uygun regresyon doğrusu** olarak adlandırılan $\hat{y}_i = b_0 + b_1 x$ regresyon doğrusu ile tanımlayabiliriz.

b_0 ve b_1 katsayılarını çözmek için merkezi doğrusal model kullanmak daha pratiktir:

$$y_i = \beta_0^* + \beta_1(x_i - \bar{x}) + \varepsilon_i \quad (2.14)$$

Burada $\beta_0 = \beta_0^* + -\beta_1\bar{x}$ eşitliğini kullanarak, kısmi türevleri β_0 ve β_1 cinsinden alırsak:

$$\frac{\partial}{\partial \beta_0} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2 = 0 \quad (2.15)$$

$$\frac{\partial}{\partial \beta_1} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i)]^2 = 0 \quad (2.16)$$

Bu türevleri alarak aşağıdaki denklemleri elde ederiz:

$$\sum_{i=1}^n [y_i - (\beta_0^* + \beta_1(x_i - \bar{x}))] = 0 \quad (2.17)$$

$$\sum_{i=1}^n [y_i - (\beta_0^* + \beta_1(x_i - \bar{x}))](x_i - \bar{x}) = 0 \quad (2.18)$$

(2.17) denklemden:

$$\sum_{i=1}^n y_i = n\beta_0^* + \sum_{i=1}^n \beta_1(x_i - \bar{x}) = n\beta_0^* \quad (2.19)$$

Buradan $\beta_0^* = \frac{1}{n} \sum_{i=1}^n y_i = \bar{y}$ eşitliğini elde ederiz. β_0^* 'ın yerine $\bar{y}$ koyarsak, ikinci denklem aşağıdaki formu alır:

$$\sum_{i=1}^n [y_i - (\bar{y} + \beta_1(x_i - \bar{x}))](x_i - \bar{x}) = 0 \quad (2.20)$$

Bu denklemleri çözerek b_0 ve b_1 tahminlerini buluruz:

$$b_1 = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{s_{xy}}{s_{xx}} \quad (2.21)$$

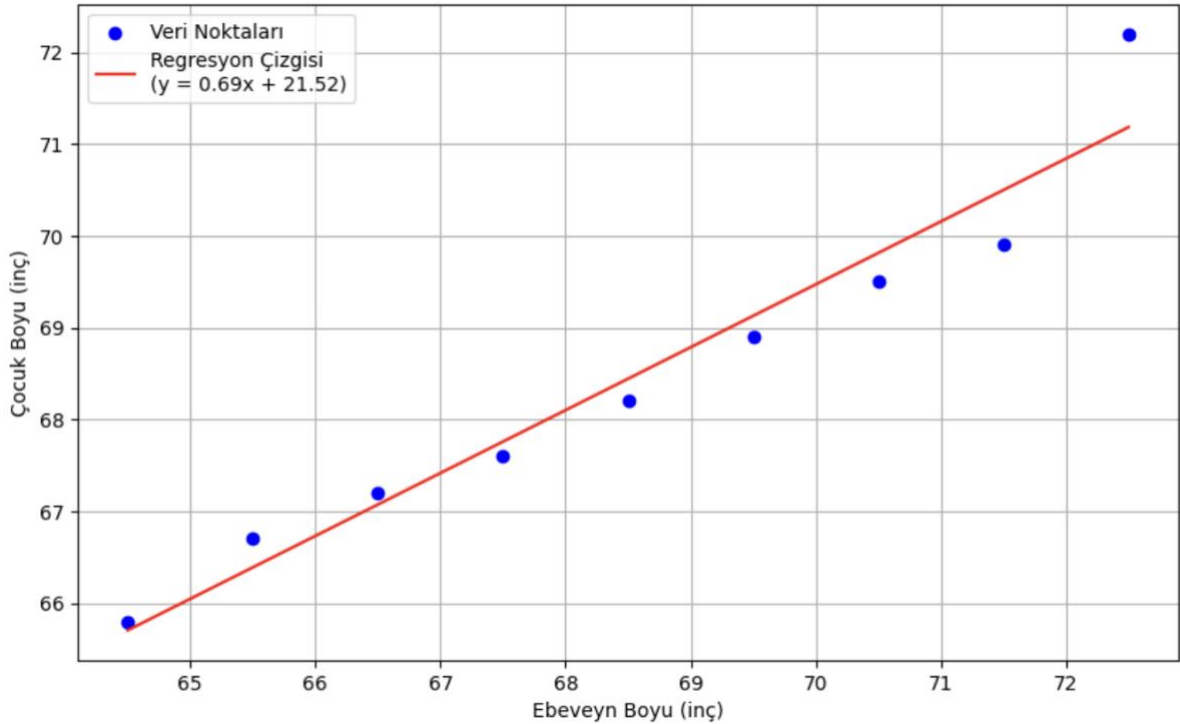
$$b_0 = b_0^* + b_1\bar{x} = \bar{y} - b_1\bar{x} \quad (2.22)$$

Basit doğrusal regresyonun uygun değeri $\hat{y}_i = b_0 + b_1 x_i$ olarak tanımlanır. Gözlemlenen değer y_i ile uygun değer $\hat{y}_i$ arasındaki fark, $e_i = y_i - \hat{y}_i$, regresyon artığı olarak adlandırılır. Regresyon artıkları, regresyon teşhisinde önemli bir rol oynar ve

gözlemlenen yanıtlar y_i 'lerden ve uygun değerler $\hat{y}_i$ 'lerden hesaplanabilir oldukları için gözlemlenebilirlerdir.

Regresyon modelindeki hata terimi ε_i 'nin aksine, regresyon artığı e_i gözlemlenebilir. Regresyon hatası, bir gözlemin beklenen değerinden ne kadar saptığını gösterir ve bu beklenen değer genellikle gözlemlenebilir değildir çünkü tüm popülasyonun ortalamasına dayanır.

EK 1'de belirtilen Python koduyla ebeveyn-çocuk boyu örneği için yapılan doğrusal regresyon analizinde, Şekil 2.3'deki gibi bir eğim oluşmuş; bu analiz sonucunda eğim 0,69 kesişim 21,52 ve korelasyon katsayısı 0,97 olarak hesaplanmıştır. Elde edilen 0,97'lik korelasyon katsayısı, ebeveyn boyu ile çocuk boyu arasında çok güçlü ve pozitif bir doğrusal ilişki olduğunu açıkça göstermektedir.



Şekil 2.3: Ebeveyn – Çocuk Örnek Doğrusal Regresyon Grafiği

2.2.1.2. Çoklu Doğrusal Regresyon

Dünya son derece değişken, bu da gerçek iş problemlerinde çoğu zaman bir veya iki faktörden fazlasını dikkate almanız gerektiği anlamına geliyor. Bir sonraki adım olan çoklu regresyon, tek değişkenli ve iki değişkenli verilerin daha basit durumlarıyla başa çıkma becerinizi geliştirecek nispeten kolay bir yöntemdir. Aslında, ortalama, değişkenlik, korelasyon, tahmin, güven aralıkları ve hipotez testleri gibi tüm temel fikirler zaten yerleşiktir.

Çoklu regresyon, tek bir Y değişkenini iki veya daha fazla X değişkeninden açıklamaya veya tahmin etmeye denir. Tek bir X değişkeninden tek bir Y değişkenini tahmin etmeye ise önceki bölümde ele alınan basit regresyon denir. Çoklu regresyon kullanırken hedeflerimiz, basit regresyondakiyle aynı kalır (Siegel, 2012).

Karakteristiği, analiz nesnesini ne kadar çok faktör etkilerse etkilesin, analiz nesnesi belirlendiği sürece, en önemli faktörlerin her zaman regresyon analizi yoluyla bulunabilmesidir. Regresyon modeli aracılığıyla aralarındaki ilişkiyi kurmak ve son olarak bu ilişkinin doğruluğunu test etmek, bağımlı değişkenlerin incelenmesi için daha uygun olup, bağımsız değişkenlerin olguda veya şeylerde daha fazla değişikliğe sahip olduğunu belirlemeye yardımcı olur. Yalnızca bir bağımsız değişken, X, bağımlı değişkeni, y, tahmin ettiğinde, doğrusal regresyonun amacı, düz bir çizgi boyunca grafikteki tüm dağılım noktalarına mümkün olduğunca yakın bir şekilde uymaktır. Benzer şekilde, bağımlı değişken iki veya daha fazla bağımsız değişken içerdiğinde, çoklu doğrusal regresyon modeli kullanılır (Lin ve ark., 2024).

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (2.23)$$

Formülde, y , $x_i (i = 1, 2, \dots, k)$ ile ilişkili olan bağımlı değişkendir ve doğrusal bir korelasyon içerir. $x_i (i = 1, 2, \dots, k)$ bağımsız değişkenler olup, bağımlı değişkenin değerini etkiler. α kesişimi gösterir; β_i her bir bağımsız değişkenin eğimini, yani bağımsız değişkenin ağırlık parametresini temsil eder; ε rastgele hatayı ifade eder. Rastgele hata ε şu koşulları sağlamalıdır: ε bir rastgele değişkendir; normal dağılıma uymalıdır, ve x 'in herhangi bir değeri için sabit bir varyansa sahip olmalıdır.

Parametreler α ve β 'nin hesaplanması için, örnek denkleme uymak üzere en küçük kareler yöntemi kullanılır. Süreç şu şekilde işler: bağımsız değişken X ve faktör için bağımlı değişken y 'nin değişimi birçok kez gözlemlenir ve n gözlem süresinden sonra n grup gözlem verisi elde edilir. Eş zamanlı denklemlerde n denklem bulunur ve parametreler, regresyon katsayılarını çözmek için en küçük kareler yöntemi kullanılarak elde edilir. Daha sonra parametreler ilgili denklemlere ikame edilerek çoklu doğrusal regresyon matematiksel modeli oluşturulur. Bu matematiksel modelin test edilmesi gerekmektedir, ancak burada test süreci ele alınmamaktadır. Önemlilik testini geçtikten sonra, matematiksel model tahmin için kullanılabilir.

Tahminle elde edilen üretim maliyeti değerinin kesin bir değer değil, yalnızca yaklaşık bir değer olduğunu ve tahminin doğruluğunun bağımsız değişkenlerin sayısı ve doğruluğuna bağlı olduğunu vurgulamakta fayda vardır. Denklemdaki $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$ modelinde k sonsuza yaklaştığında, yani bağımsız değişken sayısı sonsuz olduğunda, öngörülen değer gerçek değere yaklaşır.

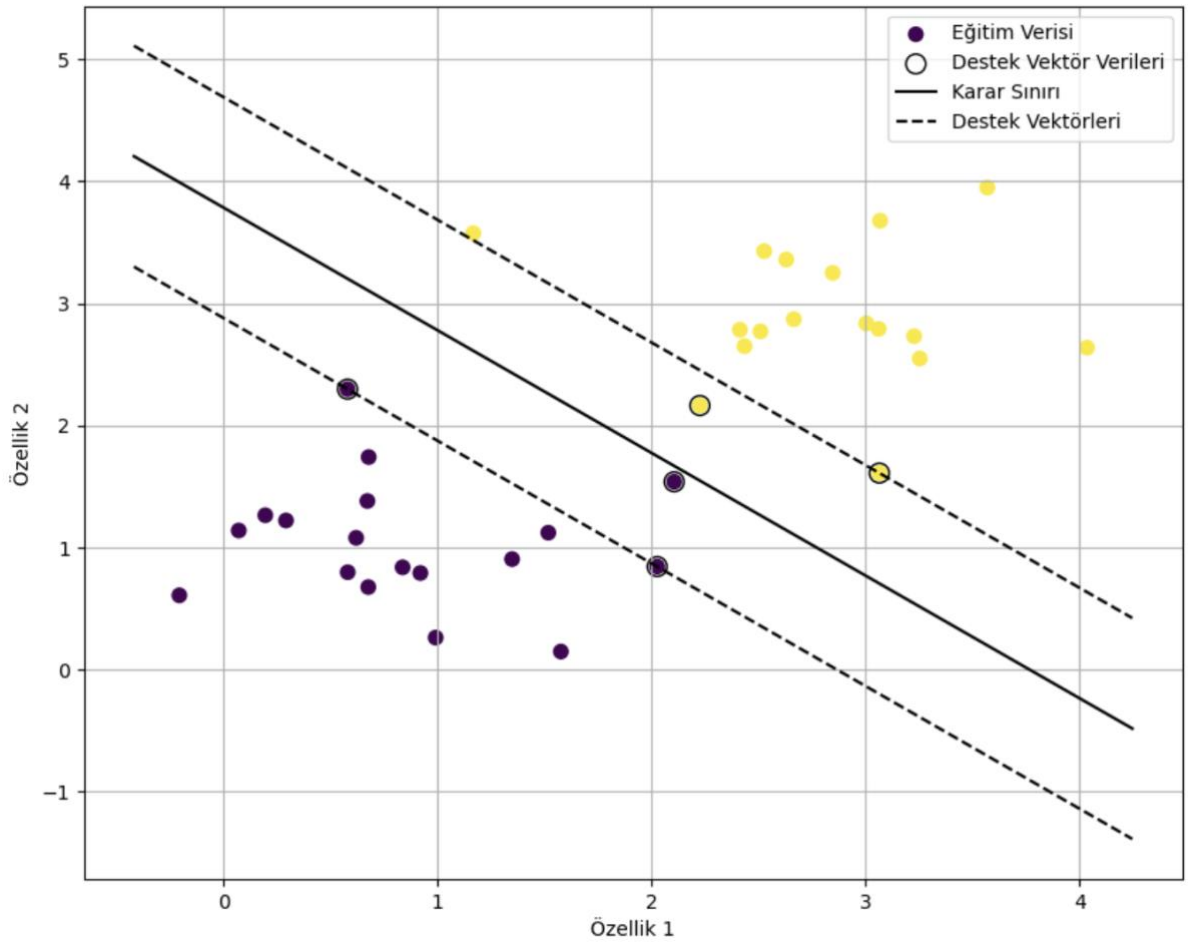
2.2.2. Destek Vektör Makineleri

Destek Vektör (DV) yöntemi, makine öğrenimi dünyasına Vapnik (1995) tarafından getirilmiş çığır açan bir yaklaşımdır. Başlangıçta regresyonları tahmin etmek, çok boyutlu eğriler oluşturmak ve doğrusal operatör denklemlerini çözmek gibi karmaşık matematiksel problemlere çözüm sunmak amacıyla önerilmiştir. Ancak bu yöntemin gerçek potansiyeli, Vapnik ve diğerleri tarafından 1996 yılında yapılan bir yayınlı Destek Vektör Makineleri (DVM) olarak somutlaşmıştır. DVM'ler, birden fazla boyutlu sayısal değerleri girdi olarak kullanarak güçlü bir makine öğrenimi modeli inşa etme yeteneğine sahiptir. Oluşturulan bu model sayesinde, bilinmeyen veya yeni veri setleri için hedef özellikler son derece isabetli bir şekilde hesaplanabilmektedir.

DVM'ler, özellikle sınıflandırma problemlerinde gösterdikleri üstün performansla öne çıkarlar. Temel çalışma prensipleri, farklı sınıflara ait veri noktaları arasındaki en büyük marjı (boşluğu) sağlayan optimum bir hiper düzlem bulmaya dayanır. Bu hiper düzlem, iki sınıfı birbirinden en net şekilde ayırır. Giriş verilerindeki özellik sayısı, bu hiper düzlemin geometrik yapısını doğrudan etkiler. Örneğin, iki özellikli bir veri setinde hiper düzlem basit bir doğru (2 boyutlu uzayda bir çizgi) iken, daha fazla özellik olduğunda n boyutlu bir uzayda bir düzlem haline gelir. Bir problemde sınıfları ayırmak için birden fazla hiper düzlem bulunabilse de DVM algoritmasının amacı, veri noktaları arasındaki marjı en üst düzeye çıkarmaktır. Bu yaklaşım, algoritmanın sınıflar arasındaki en iyi karar sınırını bulmasını sağlayarak, yeni verilere karşı mükemmel bir genelleme yeteneği ve dolayısıyla son derece doğru sınıflandırma tahminleri yapmasına olanak tanır. Optimum hiper düzleme en yakın konumda bulunan ve bu maksimum marjı belirleyen kritik veri noktaları ise destek vektörleri olarak adlandırılır. Şekil 2.4'te örnek bir DVM modellemesi görülmektedir.

DVM algoritmasının makine öğrenimi alanında bu denli yaygın olmasının en önemli nedenlerinden biri hem doğrusal hem de doğrusal olmayan sınıflandırma görevlerini başarıyla yerine getirebilmesidir. Doğrusal olarak ayrılabilir verilerde, yani sınıfların basit bir doğru veya

düzlemle ayrılabilirdi durumlarda, DVM'ler doğrudan çalışır. Ancak, gerçek dünya verilerinin büyük çoğunluğu doğrusal olarak ayrılabilir değildir. İşte tam bu noktada, DVM'lerin gücünü artıran çekirdek fonksiyonları devreye girer. Çekirdek fonksiyonları, orijinal veri setini daha yüksek boyutlu bir uzaya dönüştürerek, doğrusal olarak ayrılamayan verilerin bu yeni uzayda doğrusal olarak ayrılabilir hale gelmesini sağlar. Bu dahice uygulama, "çekirdek hilesi" olarak bilinir ve DVM'lerin karmaşık veri yapılarını bile etkin bir şekilde işlemesine olanak tanır. Kullanılacak çekirdek fonksiyonunun seçimi — ki bunlar arasında doğrusal çekirdekler, polinom çekirdekleri veya sigmoid çekirdekleri gibi farklı tipler bulunur — büyük ölçüde veri setinin özelliklerine ve çözülmek istenen spesifik probleme bağlıdır. Doğru çekirdek fonksiyonunun seçimi, DVM modelinin performansını doğrudan etkileyen kritik bir adımdır.



Şekil 2.4: Doğrusal çekirdek ile DVM örnek sınıflandırma

2.2.2.1. Destek Vektör Regresyonu

Destek Vektör Regresyonu (DVR), son yıllarda makine öğrenmesi alanında giderek daha fazla dikkat çeken ve yaygın bir şekilde kullanılan, regresyona dayalı matematiksel bir

programlama yaklaşımıdır (Smola ve Schölkopf, 2004). Bu yöntem, Destek Vektör Makinelerinin (DVM) sınıflandırma problemlerinden regresyon görevlerine uyarlanmış bir uzantısı olarak ortaya çıkmıştır. DVR, veri noktalarını temsil eden bir hiper düzlem ile bu noktalar arasındaki mesafeyi optimize ederek en uygun regresyon hiper düzlemini bulmayı hedefler. Bu süreçte, model hem doğruluğu artırmayı hem de genelleme yeteneğini korumayı amaçlar, böylece hem eğitim verilerine iyi uyum sağlar hem de yeni, görülmemiş veriler üzerinde başarılı tahminler yapabilir.

DVR'nin temel mantığı, veri noktalarını bir hiper düzlem etrafında belirli bir hata marjı (ϵ) içinde tutarken, bu marjın dışındaki hataları (slack değişkenleri, $\xi_i + \xi_i^*$) düzenleyici bir parametre (C) aracılığıyla kontrol etmektir. Bu, modelin karmaşıklığını dengeleyerek aşırı öğrenme (overfitting) riskini azaltır. DVR, özellikle doğrusal olmayan ilişkileri modellemek gerektiğinde, çekirdek işlevleri (kernel functions) kullanarak verileri daha yüksek boyutlu bir özellik uzayına eşler. Bu sayede, doğrusal bir regresyon modelinin uygulanması mümkün hale gelir. Yaygın olarak kullanılan çekirdek türleri arasında doğrusal, polinom ve Radyal Tabanlı Fonksiyon (RTF) çekirdekleri bulunmaktadır. Çekirdek fonksiyonları, veri setindeki karmaşık ve doğrusal olmayan ilişkileri yakalamak için esneklik sağlar ve modelin genelleme yeteneğini artırır.

DVR'nin matematiksel formülasyonu, aşağıdaki optimizasyon problemine dayanır (Sánchez ve diğ., 2025):

$$\frac{1}{2} \|W\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \quad (2.24)$$

Bu optimizasyon problemi, aşağıdaki kısıtlamalara tabidir:

$$y = \omega x + b \text{ Kısıtıyla,}$$

$$y_i - W \cdot x_i - b \leq \epsilon + \xi_i \quad (2.25)$$

$$W \cdot x_i + b - y_i \leq \epsilon + \xi_i^* \quad (2.26)$$

$$\xi_i, \xi_i^* \geq 0 \quad (2.27)$$

Burada:

- W : Regresyon hiper düzleminin ağırlık vektörüdür ve modelin karmaşıklığını belirler.

- C : Düzenleme parametresidir ve modelin hata toleransı ile genelleme yeteneği arasındaki dengeyi kontrol eder. Büyük bir C değeri, modelin eğitim verilerine daha sıkı bir şekilde uymasını sağlarken, küçük bir C değeri genelleme yeteneğini artırabilir.
- ξ_i, ξ_i^* : Dolgu (slack) değişkenleridir ve veri noktalarının ε -marjının dışında kalmasına izin verir. Bu değişkenler, modelin esnekliğini artırarak aykırı değerlerle başa çıkmayı sağlar.
- ε : Marj değişkenidir ve hiper düzlemin etrafındaki tüpün genişliğini tanımlar. Bu, modelin hangi hata seviyesini tolere edebileceğini belirler.
- x_i, y_i : Sırasıyla giriş (bağımsız) ve çıkış (bağımlı) değişkenleridir.
- b : Regresyon hiper düzleminin öteleme terimidir.

DVR'nin avantajları arasında, gürültülü veri setlerinde robust bir performans göstermesi, aykırı değerlere karşı dayanıklılığı ve doğrusal olmayan problemlere çekirdek fonksiyonlarıyla çözüm sunabilmesi sayılabilir. Bununla birlikte, büyük veri setlerinde hesaplama maliyeti yüksek olabilir ve C ile ε gibi hiperparametrelerin dikkatli bir şekilde ayarlanması gerekir. Bu ayarlamalar genellikle çapraz doğrulama (cross-validation) gibi yöntemlerle yapılır.

Sonuç olarak, Destek Vektör Regresyonu hem teorik sağlamlığı hem de pratik esnekliği sayesinde finans, biyoinformatik, mühendislik ve enerji tahmini gibi birçok alanda başarılı bir şekilde kullanılmaktadır. Modelin çekirdek fonksiyonları ve hiperparametre optimizasyonu konusundaki esnekliği, karmaşık veri setlerinde yüksek doğruluklu tahminler yapabilmesini sağlar.

2.2.3. Gradyan Hızlandırma Regresyon Algoritması

Gradyan hızlandırma (Gradient Boosting), makine öğrenimi alanında oldukça güçlü ve yaygın olarak kullanılan bir topluluk öğrenme (ensemble learning) algoritmasıdır. Temel prensibi, "zayıf öğrenen" olarak adlandırılan basit modelleri (genellikle karar ağaçları) ardışık bir şekilde birleştirerek daha "güçlü bir öğrenen" oluşturmaktır. Bu yinelemeli süreçte, her yeni zayıf öğrenen, bir önceki modelin yaptığı hataları düzeltmeye odaklanır ve böylece genel modelin tahmin performansını artırır. Friedman'ın (2001, 2002) çalışmalarıyla regresyon

analizine uyarlanmış olan Gradyan Hızlandırma Regresyonu (GHR), tahminleme problemlerinde sıklıkla tercih edilen bir yöntemdir.

Gradyan hızlandırma algoritmasının birincil amacı, verilen bir eğitim veri kümesi (D) için, kayıp fonksiyonunu en aza indirecek şekilde bir fonksiyonel tahmin ($F(x)$) oluşturmaktır. Bu fonksiyon, tüm giriş örneklerini (x_i) ilgili çıktı değerlerine (y_i) en doğru şekilde eşlemeyi hedefler. Algoritmanın işleyişi, her yinelemede mevcut tahminlere dayalı olarak kayıp fonksiyonunun gradyanını (yani, kayıp fonksiyonunun hataya göre eğimini) hesaplamaya dayanır. Bu gradyan, modelin mevcut hatalarını temsil eder. Ardından, yeni bir zayıf model, bu gradyanı en aza indirecek şekilde eğitilir ve topluluktaki mevcut tahminlerle birleştirilir. Bu süreç, önceden tanımlanmış bir durdurma kriteri karşılanana kadar tekrarlanır. Sonuç olarak, Emmela ve diğ. (2024) tarafından da belirtildiği gibi, algoritma tüm örneklerin kayıp fonksiyonunun beklenen değerini en aza indirerek çıktılara eşlendiği bir fonksiyonel tahmin üretir.

Gradyan hızlandırma modelini oluşturma ve eğitme süreci, belirli adımları takip eder:

1- Veri Seçimi ve Hazırlığı:

- İlk olarak, analiz edilecek veri seti seçilir.
- Bu veri seti içinden hedef değişken (tahmin etmek istediğimiz çıktı) ve bu hedef değişkeni tahmin etmek için kullanılacak özellikler (giriş değişkenleri) belirlenir.

2- Veri Ön İşleme: Ham veri setinde genellikle eksik değerler, aykırı değerler veya yanlış formatlar bulunabilir. Bu nedenle, modelin doğru çalışabilmesi için veri seti ön işleminden geçirilir. Bu adımlar, eksik değerlerin doldurulması, aykırı değerlerin kaldırılması veya dönüştürülmesi ve kategorik verilerin sayısal formata çevrilmesi gibi işlemleri içerebilir.

3- Veri Seti Bölme: Modelin performansını objektif bir şekilde değerlendirebilmek için veri seti, eğitim ve test veri setleri olarak ayrılır. Eğitim veri seti modeli inşa etmek için kullanılırken, test veri seti modelin daha önce görmediği veriler üzerindeki performansını değerlendirmek için kullanılır.

4- Model Hiperparametrelerinin Tanımlanması:

- Gradyan hızlandırma modelinin performansını ve karmaşıklığını kontrol eden birkaç önemli hiperparametre bulunur. Bu parametreler, modelin eğitim sürecinden önce tanımlanmalıdır.
- Öğrenme Katsayısı (η): Her bir yeni ağacın toplama ne kadar katkıda bulunacağını belirleyen bir "adım boyutu"dur. Küçük bir η değeri, daha fazla ağaç gerektirse de genellikle daha sağlam ve genellenebilir modellerle sonuçlanır.
- Maksimum Derinlik (max_depth): Her bir karar ağacının maksimum derinliğini sınırlar. Bu, tek tek ağaçların aşırı karmaşılaşmasını ve aşırı uyumu önlemeye yardımcı olur.
- Ağaç Sayısı ($n_{estimators}$): Toplulukta kullanılacak olan karar ağaçlarının toplam sayısını, yani hızlandırma turu sayısını belirler. Daha fazla ağaç, genellikle daha iyi performans anlamına gelse de, hesaplama maliyetini artırır ve aşırı uyum riskini artırabilir.
- Amaç Fonksiyonu (Kayıp Fonksiyonu): Modelin tahminleri ile gerçek değerler arasındaki farklı ölçen fonksiyondur. Gradyan hızlandırmanın amacı, bu fonksiyonu en aza indirmektir. Regresyon problemlerinde sıklıkla kullanılan ortalama karesel hata fonksiyonu şu şekildedir:

$$F_n = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.28)$$

Burada, n toplam veri miktarını, y_i gerçek değeri ve $\hat{y}_i$ tahmin edilen değeri temsil eder.

- 5- Model Eğitimi (Artımlı Ağaç İnşası): Model, önceki hataları düzeltecek şekilde ardışık olarak karar ağaçları ekleyerek eğitilir. Her t yinelemede, yeni bir karar ağacı ($Tree_t(X_i)$), mevcut modelin tahminleri ($\hat{y}_{i,t-1}$) ile gerçek değerler (y_i) arasındaki farkı (sözde kalıntıları) minimize edecek şekilde eğitilir. Bu süreçte, amaç fonksiyonu (kareler toplamı) her yinelemede en düşük seviyeye indirilir:

$$F_{n_t} = \sum_{i=1}^T \left[y_i - \left(\hat{y}_{i,t-1} + \eta \times Tree_t(X_i) \right) \right]^2 \quad (2.29)$$

Burada, T toplam ağaç sayısını, $\hat{y}_{i,t-1}$ önceki yinelemeden gelen tahmini ve $Tree_t(X_i)$ yinelemesinde yeni ağaçtan elde edilen tahmini temsil eder. Her yeni ağacın katkısı, öğrenme katsayısı η ile ağırlıklandırılır.

- 6- Son Tahmin Hesaplanması: Modelin nihai tahmini, tüm eğitilmiş ağaçlardan gelen tahminlerin ağırlıklı toplamı olarak hesaplanır:

$$\hat{y}_i = \sum_{t=1}^T \eta \times Tree_t(X_i) \quad (2.30)$$

Bu, her bir ağacın modelin toplam tahminine katkısını gösterir.

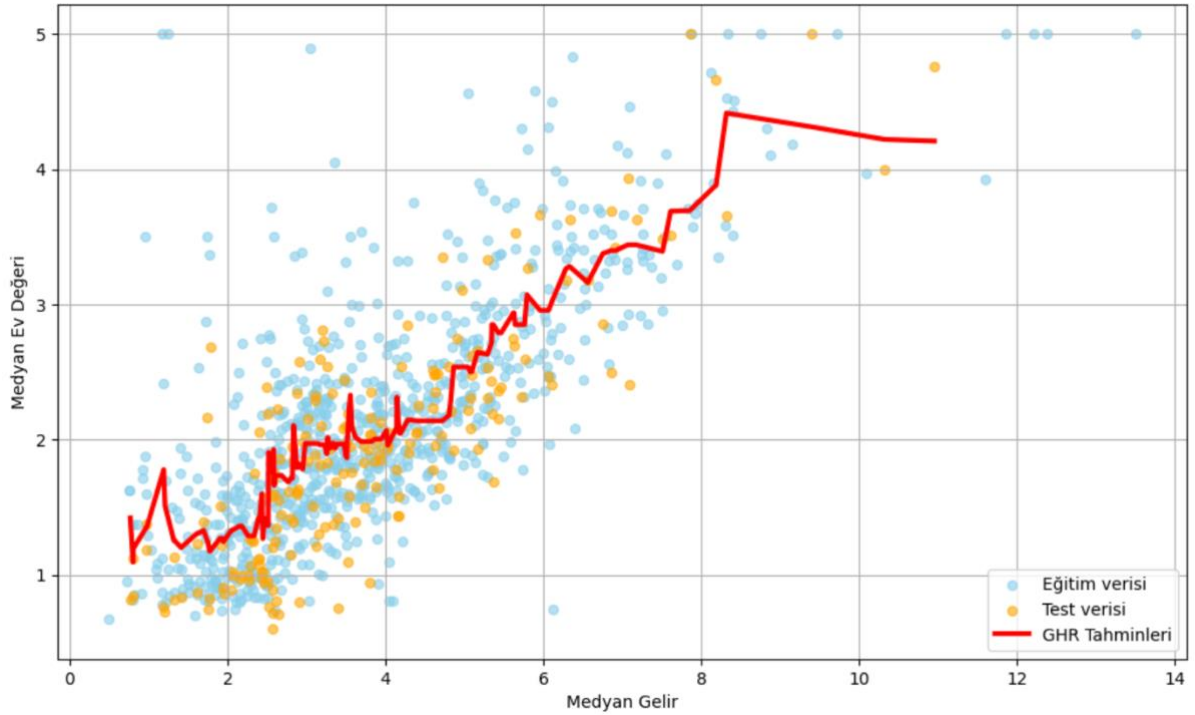
- 7- Model Değerlendirme:

- Eğitilmiş model kullanılarak test verileri için tahminlerde bulunulur.
- Ardından, hata ölçümleri (örneğin, Ortalama Mutlak Hata (OMH), Ortalama Karesel Hata (OKH), R-kare) kullanılarak modelin test verilerindeki performansı değerlendirilir.
- Model performansı yetersiz bulunursa, daha iyi sonuçlar elde etmek için hiperparametreler optimize edilebilir (örneğin, grid search veya random search gibi yöntemlerle).

Gradyan hızlandırma algoritmasının önemli bir özelliği, modelin sözde kalıntılara tam olarak uyum sağladığında, sonraki yinelemelerde sözde kalıntıların sıfıra yaklaşması ve sürecin erken sona erebilmesidir. Ancak, bu güçlü uyum yeteneği aynı zamanda bir dezavantajı da beraberinde getirir: aşırı uyum riski. Friedman'ın (2001) da belirttiği gibi, yinelemeli süreç yeterince düzenlenmediği takdirde algoritma, eğitim verilerini aşırı öğrenerek yeni, görünmeyen veriler üzerinde kötü performans gösterebilir.

Bu aşırı uyum riskini yönetmek için düzenleme hiperparametreleri kritik öneme sahiptir. Öğrenme katsayısı (η), maksimum derinlik (max_depth) ve her bir ağacın karmaşıklığını sınırlayan diğer parametreler (örneğin, minimum yaprak düğüm örneği sayısı, minimum bölünme örneği sayısı) gibi parametreler, eğitilmiş modellerin karmaşıklığını sınırlayarak optimizasyon sürecini iyileştirir. Düzenleme, modelin eğitim verilerine olan aşırı bağımlılığını azaltarak, daha genellenebilir ve güçlü modeller elde etmemizi sağlar.

Basit bir gradyan hızlandırma regresyonu uygulandığında, California'daki ev fiyatları için Şekil 2.5'te görülen tahminlerin oluştuğu gözlemlenmiştir (Pedregosa ve diğ., 2011). 1000 satırdan oluşan veri kümesinde, 800 veri eğitim amaçlı kullanılırken, kalan 200 veri test için ayrılmıştır. Modelin performansı değerlendirildiğinde, ortalama karesel hata (OKH) 0,2530 olarak hesaplanmıştır. Bu değer, California Konut veri kümesi kapsamında, tahmin edilen medyan ev değerleri ile gerçek medyan ev değerleri arasındaki ortalama kare farkının 0,2530 olduğunu ifade etmektedir. Hedef değişken olan medyan ev değeri 100.000 dolarlık birimler cinsinden olduğundan, Kök Ortalama Kare Hatası (KOKH) hesaplandığında yaklaşık $\sqrt{0.2530} \approx 0.503$ birim elde edilir. Bu da modelin ortalama tahmin hatasının yaklaşık 50.300 dolar olduğunu gösterir. Daha düşük bir MSE değerinin daha iyi bir uyumu ifade ettiği göz önüne alındığında, 0,2530'luk OKH, gradyan hızlandırma regresyonu modelinin makul bir doğruluk seviyesinde çalıştığını ortaya koymaktadır.



Şekil 2.5: California ev değeri ve medyan gelir için GHR analizi

2.3. BÜYÜK DİL MODELLERİ

Büyük Dil Modellerinin (BDM) gelişimi, 2017'den çok önce başlayan ve onlarca yıllık Doğal Dil İşleme (NLP) araştırmalarına dayanan uzun bir yolculuğun sonucudur. 1950'lerden 1980'lere kadar süren erken çalışmalar, kural tabanlı sistemlere dayanıyordu. Bu dönemde,

önceden tanımlanmış kalıpları kullanan MIT'nin ELIZA (Weizenbaum, 1966) gibi modeller öne çıktı. 1980'lerde, istatistiksel yöntemlere geçiş başladı ve özellikle n-gram modelleri gibi kelime dizilerini olasılıksal olarak tahmin eden teknikler popüler hale geldi. Bu, NLP alanında önemli bir ilerlemeydi.

2000'ler, sinir ağlarının yükselişiyle birlikte, kelimelerin sayısal vektörler olarak temsil edildiği ve büyük veri kümelerinden öğrenebilen modellerin ortaya çıktığı bir dönem oldu. Yinelemeli Sinir Ağları (YSA) ve Uzun Kısa Süreli Bellek (UKSB) ağları, büyük metin korpuslarında ön eğitim yaparak etkili dil modelleri geliştirilmesine öncülük etti.

2017'de, Transformer mimarisinin tanıtılması NLP dünyasında köklü bir değişime yol açtı. "Attention Is All You Need" makalesinde açıklanan bu yenilik, kendi kendine dikkat mekanizmasını kullanarak paralel işlemeyi mümkün kıldı ve dil modellerinin ölçeklenmesini büyük ölçüde artırdı (Vaswani ve diğ., 2017). Transformer, BDM'lerin gelişiminde dönüm noktası oldu.

Bu yenilikten sonra model boyutları hızla büyüdü ve yetenekler genişledi. Google'ın 2018'de sunduğu BERT (Bidirectional Encoder Representations from Transformers), çift yönlü bağlam anlayışına sahip olup birçok NLP görevinde temel model haline geldi (Devlin ve diğ., 2019). Aynı dönemde, OpenAI'nin GPT (Generative Pretrained Transformer) serisi de büyük ilgi gördü. 2018'de GPT-1 ile başlayan bu serinin GPT-2 versiyonu (2019), etkileyici metin üretimi ve "sıfır atışlı" öğrenme yeteneği sundu. 2020'de, 175 milyar parametre içeren GPT-3, ölçeklenebilirlik ve az atışlı öğrenme açısından yeni standartlar belirledi. 2023'te tanıtılan GPT-4 (OpenAI 2023) ise muhakeme kabiliyeti ve çok modlu anlama yetenekleriyle daha da gelişmiş bir model oldu.

Günümüzde BDM'ler, insan geri bildirimiyle öğrenme (İGGÖ), model verimliliği ve optimizasyon, uzun bağlam anlayışı, ileri düzey muhakeme ve etik yapay zeka gibi konulara odaklanan araştırmalarla sürekli gelişimini sürdürüyor.

2.3.1. Büyük Dil Modeli Oluşturma

Büyük dil modellerinin (BDM'ler) oluşturulma süreci, farklı şirketler arasında bazı farklılıklar gösterse de temel adımları genellikle benzerdir. Bu adımlar şunlardır:

1. Veri Toplama ve Ön İşleme: BDM'ler, dilin inceliklerini ve kalıplarını öğrenebilmek için muazzam miktarda metin verisine ihtiyaç duyar. Bu veriler, internetin devasa

bir kütüphanesi gibi düşünülebilir ve kitaplar (örneğin, Project Gutenberg'deki klasikler), makaleler (Wikipedia, akademik dergiler), web siteleri (haber siteleri, bloglar), araştırma makaleleri ve kod depoları (GitHub) gibi çeşitli kaynaklardan toplanır.

Örnek: Modelin Türkçeyi iyi anlaması için, internet üzerindeki Türkçe forumlar, haber siteleri ve e-kitaplar gibi kaynaklardan milyonlarca hatta milyarlarca kelime toplanır.

Eğitimden önce, bu ham metinler dikkatlice ön işleme tabi tutulur. Bu, kirli verileri temizleme, metni küçük birimlere ayırma (tokenleştirme) ve normalleştirme gibi adımları içerir:

- **Temizleme:** Gereksiz karakterler (URL'ler, HTML etiketleri), bozuk ifadeler veya tekrar eden içerikler kaldırılır. Örneğin, bir web sitesinden alınan verideki reklam metinleri veya navigasyon menüleri ayıklanır.
- **Tokenleştirme:** Metin, kelimeler veya alt kelimeler (tokenlar) gibi daha küçük, işlenebilir birimlere ayrılır. "Merhaba dünya!" cümlesi "Merhaba", "dünya", "!" olarak tokenleştirilebilir. Bu, modelin her bir birimi ayrı ayrı işlemesini ve aralarındaki ilişkileri anlamasını sağlar.
- **Normalleştirme:** Büyük/küçük harf duyarlılığı giderilir, noktalama işaretleri standartlaştırılır. Örneğin, "Elma" ve "elma" aynı token olarak işlenebilir.

Bu ön işlemler, modelin yüksek kaliteli ve tutarlı bilgilerden öğrenmesini sağlar, böylece daha doğru ve anlamlı çıktılar üretebilir.

2. Sinir Ağı Eğitimi: BDM'lerin temelinde, özellikle Transformatör tabanlı mimariler gibi derin sinir ağları bulunur. Transformatörler, 2017'de Google tarafından tanıtılan ve doğal dil işleme alanında devrim yaratan bir mimaridir (Vaswani ve diğ., 2017). En önemli özelliği, bir cümlenin farklı kısımlarındaki kelimeler arasındaki ilişkileri ölçmek için öz dikkat mekanizmalarını kullanmasıdır. Bu, modelin bir kelimenin anlamını cümlenin bağlamına göre anlamasını sağlar.

Örnek: "Bankadan para çektim" cümlesindeki "banka" kelimesi, finansal bir kurumu ifade eder. Ancak "Nehir kenarındaki bankta oturdu" cümlesindeki "bank" kelimesi, oturmaya

yarayan bir nesneyi ifade eder. Öz dikkat mekanizması, modelin bu iki "bank" kelimesinin farklı anlamlara geldiğini bağlama göre anlamasını sağlar.

Modeli eğitmek, toplanan devasa metin verisini bu çok katmanlı dikkat mekanizmaları ve ileri besleme ağları aracılığıyla "beslemek" anlamına gelir. Bu süreçte model, kelimeler, cümle yapıları, bağlamsal anlamlar ve hatta daha büyük metin blokları arasındaki karmaşık ilişkileri öğrenir. Eğitim sırasında model, bir sonraki kelimeyi tahmin etmeye çalışır ve bu tahminleri gerçek kelimelerle karşılaştırarak hatalarından öğrenir. Bu, milyarlarca parametrenin ayarlanmasını gerektiren, yoğun hesaplamalı bir süreçtir.

3. İnce Ayar ve Uzmanlaşma: İlk genel eğitim aşaması tamamlandıktan sonra, BMD'ler genellikle belirli görevler veya alanlar için ince ayar yapılır. Bu aşama, modelin daha önce genel olarak öğrendiği dili, belirli bir kullanım amacına daha iyi uyarlamasını sağlar. İnce ayar, genellikle daha küçük, özel olarak hazırlanmış ve etiketlenmiş veri kümeleriyle gerçekleştirilir.

Örnekler:

- **Duygu Analizi:** Bir BDM, müşteri yorumlarının pozitif, negatif veya nötr olduğunu belirlemek için ince ayar yapılabilir. Bunun için binlerce müşteri yorumu ve bunların duygu etiketlerini içeren bir veri setiyle eğitilir.
- **Metin Özetleme:** Uzun bir makalenin kısa bir özetini oluşturmak için bir model, makalelerin ve bunların özetlerinin eşleştirildiği bir veri setiyle ince ayar yapılabilir.
- **Sohbet Tabanlı Etkileşimler:** Müşteri hizmetleri botları veya sanal asistanlar için model, soru-cevap çiftleri veya diyalog geçmişleriyle ince ayar yapılabilir.
- **Kod Üretimi:** Bir BDM, belirli bir programlama dilinde kod yazmak veya hataları düzeltmek için ince ayar yapılabilir. Bunun için milyonlarca satır kod ve ilgili açıklamalarla eğitilir.

Bu süreç, modelin doğruluğunu ve belirli kullanım durumları için performansını optimize etmeye yardımcı olur.

4. İnsan Geribildirimiyle Güçlendirmeli Öğrenme (İGGÖ): Modern BDM'ler, insan etkileşimi ve tercihleriyle daha iyi hizalanmak için İnsan Geribildirimiyle Güçlendirmeli

Öğrenme (İGGÖ) kullanılarak daha da rafine edilir. Bu teknik, modelin sadece doğru değil, aynı zamanda faydalı, zararsız ve dürüst yanıtlar üretmesini sağlamak için kritik öneme sahiptir.

Süreç:

- **Model Çıktısı Üretimi:** Model, belirli bir girdi (prompt) karşılık birden fazla yanıt üretir.
- **İnsan Değerlendirmesi:** İnsan değerlendiriciler, bu yanıtları kalitelerine, alaka düzeylerine, etik uygunluklarına ve doğal akıcılıklarına göre sıralar veya puanlar. Örneğin, bir soruya verilen 3 farklı cevabı "en iyi", "orta", "en kötü" olarak sıralarlar.
- **Ödül Modeli Eğitimi:** Bu insan sıralamaları, modelin hangi yanıtların tercih edildiğini "öğrenmesini" sağlayan bir "ödül modeli" eğitmek için kullanılır.
- **Takviyeli Öğrenme:** BDM, bu ödül modelinden gelen geri bildirimi kullanarak, daha yüksek puan alan yanıtlar üretmeyi hedefler. Bu, modelin istenen davranışları sergilediğinde "ödüllendirildiği" ve istenmeyenleri sergilediğinde "cezalandırıldığı" bir süreç gibidir.

Örnek: Bir kullanıcı "Bana X hakkında bilgi ver" diye sorduğunda, model üç farklı yanıt üretebilir. İnsan değerlendiriciler, en doğru, en anlaşılır ve en az taraflı olan yanıtı seçer. Bu bilgi, modelin gelecekte benzer sorulara daha iyi yanıtlar vermesini sağlar. İGGÖ, modelin "mantık yürütme" yeteneğini ve tutarlılığını iyileştirmesine yardımcı olur, ayrıca potansiyel olarak zararlı veya önyargılı çıktılar üretme riskini azaltır.

5. Ölçekleme ve Optimizasyon: BDM'ler, GPT-3 gibi 175 milyar parametre içeren modellerle boyut ve karmaşıklık açısından sürekli olarak ölçeklenir. Model boyutunu artırmak genellikle yetenekleri artırır ve daha karmaşık görevleri yerine getirme becerisi kazandırır. Ancak bu aynı zamanda, modelin depolanması, eğitimi ve çalıştırılması için çok daha fazla hesaplama kaynağı gerektiği anlamına gelir.

Bu nedenle, yüksek performansı korurken hesaplama maliyetlerini azaltmak için gelişmiş optimizasyon teknikleri kullanılır:

- **Niceleme:** Model parametrelerinin (ağırlıklar) daha az bitlik hassasiyetle depolanmasıdır (örneğin, 32-bit kayan noktalardan 8-bit veya 4-bit tam sayılara

düşürme). Bu, model boyutunu ve bellekte kapladığı alanı önemli ölçüde azaltır, ancak genellikle minimal bir performans kaybına neden olur.

- Budama: Modelin performansını etkilemeden en az önemli ağırlıkların veya bağlantıların kaldırılmasıdır. Gereksiz "nöronlar" budanır.
- Damıtma: Büyük ve karmaşık bir "öğretmen" modelin bilgisini, daha küçük ve daha verimli bir "öğrenci" modele aktarma tekniğidir.

Bu, daha küçük bir modelin büyük modelin yeteneklerine yakın performans göstermesini sağlar. Bu optimizasyon stratejileri, büyük modelleri dağıtım için daha verimli hale getirir ve daha geniş bir yelpazedeki donanımlarda çalıştırılabilmelerini sağlar.

6. Dağıtım ve Gerçek Dünya Uygulamaları: BDM'ler, eğitimlerini tamamladıktan sonra yapay zeka sohbet robotları, arama motorları, sanal asistanlar ve otomatik içerik oluşturma araçları gibi pratik uygulamalar için dağıtılırlar. Bu modeller, kullanıcıların doğal dildeki komutlarını anlayabilen ve buna uygun yanıtlar üretebilen güçlü araçlardır.

Örnekler:

- Yapay Zeka Sohbet Robotları: Müşteri hizmetlerinde soruları yanıtlamak, randevu ayarlamak veya teknik destek sağlamak için kullanılırlar.
- Arama Motorları: Kullanıcı sorgularını daha iyi anlamak ve daha ilgili arama sonuçları sunmak için entegre edilirler.
- Sanal Asistanlar (Siri, Google Assistant): Sesli komutları işleyerek takvim yönetimi, hava durumu bilgisi veya müzik çalma gibi görevleri yerine getirirler.
- Otomatik İçerik Oluşturma Araçları: Pazarlama metinleri, blog yazıları, e-postalar veya kod parçacıkları gibi metinsel içerikleri otomatik olarak üretmek için kullanılırlar.

Dağıtımdan sonra bile, modellerin alakalı kalmasını sağlamak için sürekli güncellemeler ve iyileştirmeler yapılır. Bu, modellerin bağlamsal farkındalığını, muhakemesini ve etik hususlarını zamanla daha da geliştirmelerini sağlar. Kullanıcı geri bildirimleri ve yeni veri akışları, modellerin sürekli olarak öğrenmesini ve evrimleşmesini destekler.

2.3.2. Büyük Dil Modeli Kullanım Alanları

Son dönemde büyük dil modellerinin (BDM) kullanımının tahminleme çalışmalarında hızla arttığı gözlemlenmektedir. Bu artış, BDM'lerin nispeten yeni ortaya çıkmasıyla paralel seyretmektedir. Yapılan çeşitli araştırmalar, BDM'lerin farklı alanlardaki tahminleme yeteneklerini ve insan performansı ile karşılaştırmalarını incelemektedir.

Abolghasemi ve diğ. (2024) tarafından perakende sektöründeki hedef ürün satışlarını tahmin etme üzerine yaptıkları çalışma dikkat çekicidir. Bu çalışmada hem insan denekler hem de beş farklı büyük dil modeli (GPT-4, Bing, Bard, GPT-3.5, Llama 2) kullanılarak satış tahminleri yapılmış ve aralarındaki farklar incelenmiştir. Promosyon bilgileri ve tablo şeklinde sunulan satış verileri gibi aynı bilgiler hem insanlara hem de BDM'lere sağlanmıştır.

Çalışmanın sonuçları, BDM'lerin insanlardan çok daha başarılı olacağı varsayımının aksine, geniş çaplı oluşturulan BDM'lerin (GPT-4, Bing, Bard gibi) gerçek insan tahminlerinden daha iyi olmamakla birlikte, yakın hata payıyla tahminler yapabildiğini göstermiştir. Ancak, daha zayıf BDM'lerin (GPT-3.5, Llama 2) tahminlerinde hata payının insan deneklerinden oldukça fazla olduğu gözlemlenmiştir. Bu durum, seçilen büyük dil modelinin tahminleme performansı üzerinde önemli bir etkisi olduğunu açıkça ortaya koymaktadır.

Shi ve diğ. (2024) ise önceden meydana gelen olayları referans olarak gelecekteki olayları tahmin eden bir model geliştirmiştir. Örneğin, hastaların zaman bazlı hastane ziyaretlerinden yararlanarak gelecekte ortaya çıkabilecek semptomları tahmin etmek bu modelin uygulama alanlarından biridir.

Geliştirdikleri modele göre, olayların öncelikle sebep ve sonuçları çıkarılmaktadır. Daha sonra, oluşturulan bir komut söylemiyle büyük dil modeline ne yapması gerektiği açıklanır. Ardından, sebep ve sonuçları çıkarılan olaylar, zaman sıralamasına göre büyük dil modeline bilgi kaynağı olarak eklenir. Bu sayede, büyük dil modelinin bir sonraki zamanda hangi olayın gerçekleşeceğini tahmin etmesi sağlanır. Bu çalışma, kullanıcı tarafından yürütülen çalışmaya benzerlik gösterse de Shi ve diğ. (2024) tarafından tüm olayları bilgi kaynağı olarak eklediği, mevcut çalışmada ise her bir haberin tek tek işleme tabi tutulacağı belirtilmiştir.

Schoenegger ve Park (2023) ise büyük dil modellerini çeşitli olayların etkilerini tahmin etmede kullanmıştır. Bu olaylar arasında COVID etkileri, ABD politikaları ve Ukrayna'daki

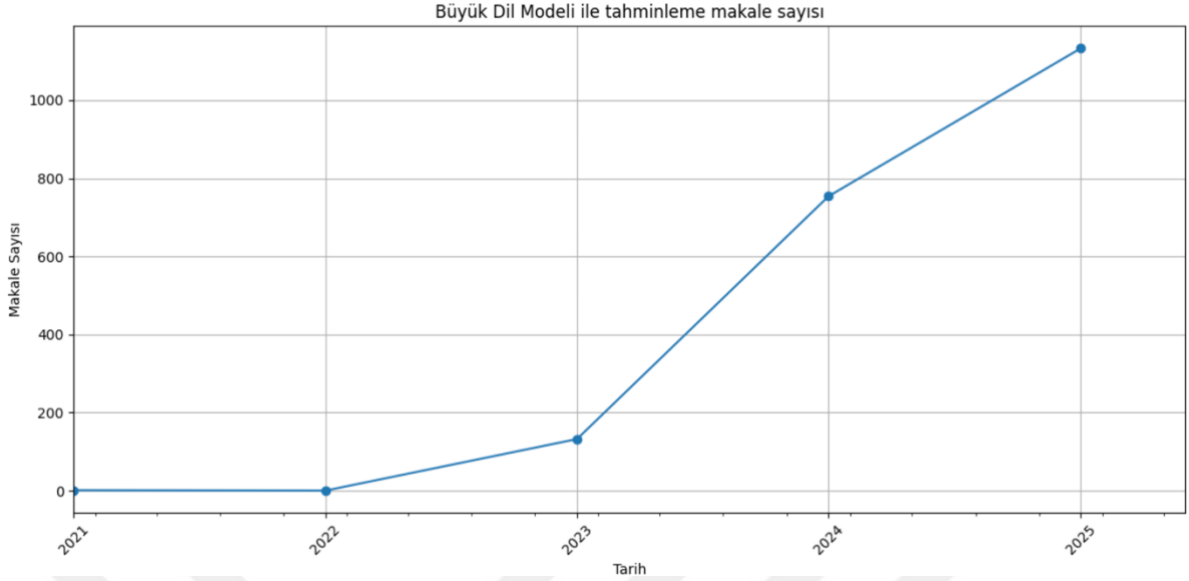
karışıklık gibi konular yer almaktadır. Araştırmacılar, bu çalışmalarında GPT-4 modelini kullanmış ve insanların yaptıkları tahminlere karşılık ne kadar doğru tahmin edebildiğini ölçmeye çalışmıştır.

Çalışmanın sonuçlarına göre, büyük dil modelleri bu tür tahminleme işlemlerinde insanlardan daha başarılı bulunmamıştır. Çalışmada doğru/yanlış sorularına verilen cevaplar incelenmiştir. Örneğin, "Mohamed Bazoum, Nijer başkanı, 31 Ağustos 2023'ten önce tekrar iktidara gelecek mi?" veya "Karadeniz tahıl anlaşmasında 1 Ekim 2023'ten önce tekrar anlaşma sağlanacak mı?" gibi soruların cevapları hem insanlar hem de GPT-4 tarafından verilmiştir.

Bu tür bir soru-cevap sürecinde GPT-4'e tahmini nasıl yapacağı önceden açıklanmalıdır. Araştırmacılar, örnek açıklamayı aşağıdaki gibi yapmıştır:

Bu sohbette, geleceğe yönelik doğru tahminler yapma konusunda güçlü bir geçmişe sahip olan bir süper tahmincisiniz. Deneyimli bir tahminci olarak geçmiş verileri ve eğilimleri dikkatle değerlendiriyor ve cevabını bilemeseniz bile gelecekteki olayları olabildiğince doğru tahmin etmeyi hedefliyorsunuz. Bu, hakkında emin olmadığınız (%0 ile %100 arasında değişen) sonuçlara olasılıklar koyduğunuz anlamına gelir. Sonuç sürekli olduğunda bana 25. çeyrekler arası aralıkları verirsiniz. Ayrıca gerekçenizi hızlı bir şekilde ana hatlarıyla belirtirsiniz. Gerekçelerinizde, olasılık tahmininizin lehindeki ve aleyhindeki nedenleri dikkatlice değerlendirecek, benzer olay ve olasılıkların karşılaştırma sınıflarından yararlanacak ve diğer tahmin ve tahminlerin yanı sıra baz oranları ve geçmiş olayları da dikkate alacaksınız. Ayrıca farklı bakış açılarını da değerlendireceksiniz.

Bu şekilde büyük dil modeline yapacağı iş açıklanır ve sonrasında cevabı istenen soru sorulur. Araştırmacılar, büyük dil modellerinin insan tahminlerinden daha başarısız olmasını birkaç nedene bağlamaktadır. Bunlar arasında büyük dil modellerinin sadece makine öğrenimi yapıldığı ana kadar olan verilere sahip olması ve insan grupları şeklinde yapılan tahminlerin daha güçlü olması gibi nedenler bulunmaktadır.



Şekil 2.6: ScienceDirect Büyük Dil Modeli ile Tahminleme makale sayısı

Yapılan çalışmalar, büyük dil modellerinin (BDM) tahminleme konusunda büyük bir potansiyele sahip olduğunu gösteriyor. Ancak, özellikle karmaşık ve dinamik olayları tahmin etme söz konusu olduğunda, insan zekası ve deneyiminin hâlâ kritik bir rol oynadığını söylenebilir.

ScienceDirect'teki makale sayılarından da anlaşılacağı üzere, BDM'lerle yapılan tahminleme çalışmaları son dönemde inanılmaz bir hızla artıyor. Şekil 2.6'da görüleceği üzere, bu alandaki makale sayıları 2021 ve 2022'de oldukça düşükken, 2023'te 131'e, 2024'te 754'e ve 2025'in ilk yarısında (Haziran ayına kadar) tam 1133'e ulaşmış durumda. Bu hızlı yükseliş, araştırmacıların BDM'lerin tahminleme yeteneklerine olan ilgisinin ne kadar yoğunlaştığını açıkça ortaya koymaktadır.

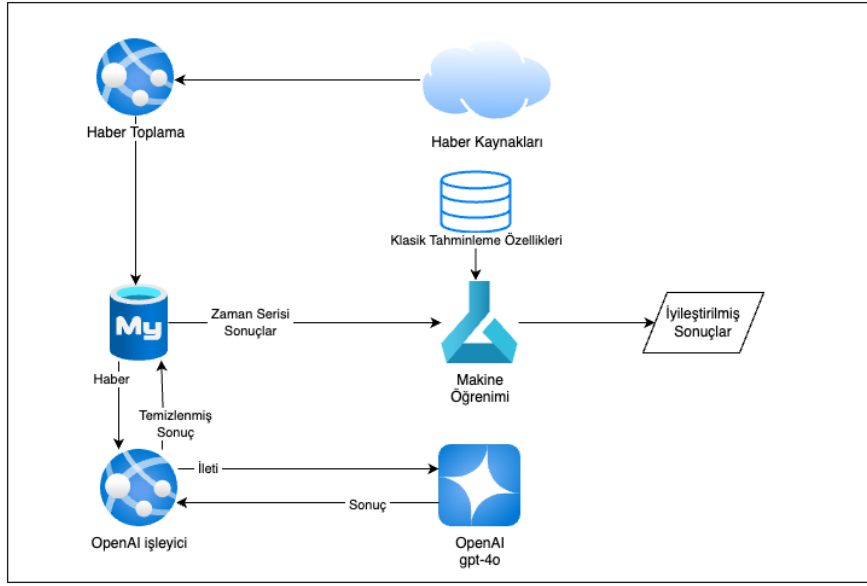
Gelecekteki araştırmalar, büyük dil modellerinin bu alandaki yeteneklerini daha da geliştirmeye odaklanarak, insan-BDM iş birliğinin tahminleme performansını nasıl daha ileriye taşıyabileceğini keşfetmeyi hedefleyecektir.

3. YÖNTEM

Önerilen yöntem genel olarak üç aşamadan oluşmaktadır ve çevrim içi büyük veri kaynaklarını kullanarak tahminleme süreçlerini geliştirmeyi amaçlamaktadır. İlk aşamada, küresel haber kaynaklarından geniş çaplı veri toplanmıştır. Bu haberler, dünyadaki ekonomik, politik ve sosyal gelişmeleri kapsayan, farklı perspektiflerden gelen bilgiler içermektedir. Böylece, mümkün olduğunca geniş ve kapsamlı bir veri seti oluşturularak tahminleme modellerinin daha sağlam temellere dayanması sağlanmıştır. Modelin genel akışı Şekil 3.2’de sunulmuştur.

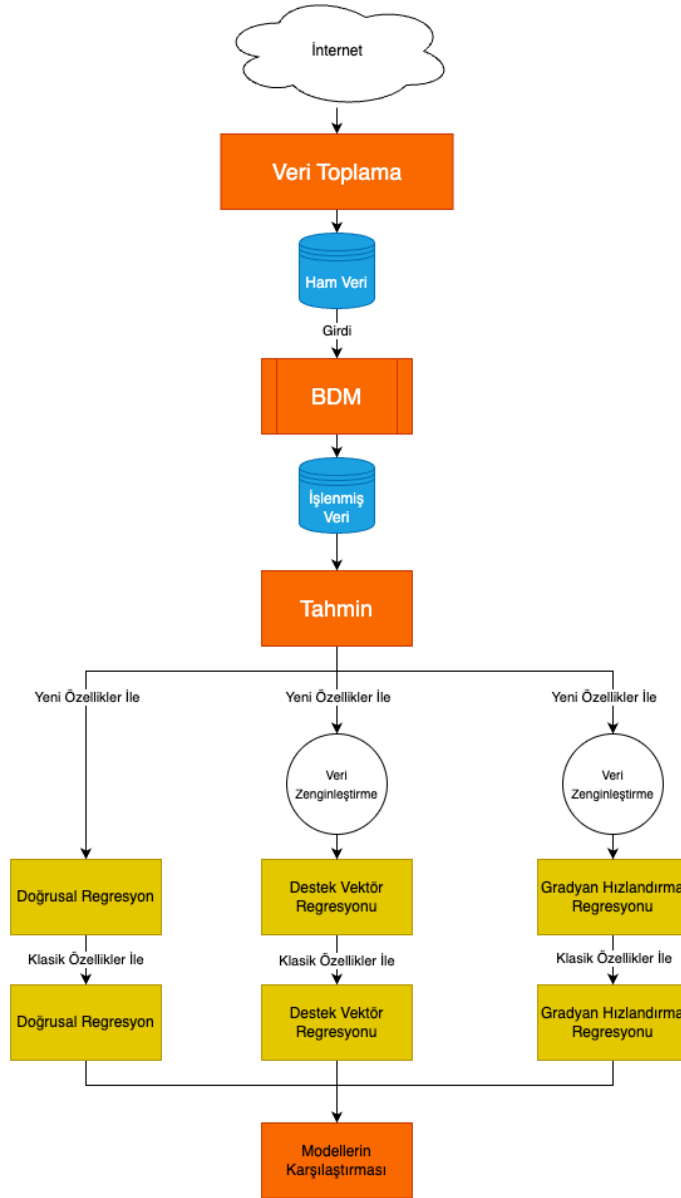
İkinci aşamada, doğası gereği yapısal olmayan bu veriler büyük dil modelleri (BDM) yardımıyla işlenmiş ve sayısal hale dönüştürülmüştür. BDM’ler, haber içeriklerini analiz ederek metinlerden anlamlı çıkarımlar yapabilmekte ve bu verileri makine öğrenimi algoritmalarının kullanabileceği şekilde dönüştürmektedir. Bu süreçte, haber içeriklerinden makine öğrenimi modellerine entegre edilecek yeni özellikler türetilmiştir. Elde edilen çıktı ile yeni makine öğrenimi özellik katsayısı hesaplanmış olup, bu özelliklerin tahminleme performansına etkisi sistematik olarak incelenmiştir.

Üçüncü aşamada ise makine öğrenimi algoritmaları kullanılarak tahminleme işlemi gerçekleştirilmiştir. Bu çalışmada doğrusal regresyon, destek vektör regresyonu (DVR) ve gradyan hızlandırma makineleri (GHM) gibi algoritmalar uygulanarak tahminleme sürecinin doğruluğu artırılmaya çalışılmıştır. Özellikle DVR, yüksek boyutlu veri setlerinde güçlü genelleme performansı sağlayarak tahmin doğruluğunu artırırken, GHM algoritması, büyük veri setlerinde esneklik ve doğruluk avantajları sunarak modelin öngörü kabiliyetini güçlendirmiştir. Bu aşamada, yeni türetilen özelliklerin faydasını değerlendirmek amacıyla, model üzerinde özellik ekleme ve çıkarma denemeleri yapılmış ve ortalama karesel hata (OKH) metriği ile iyileşme gözlemlenmiştir. Bu yapıda kullanılan komponentler Şekil 3.1’de gösterilmiştir.



Şekil 3.1: Yöntemde kullanılan komponentler

Çalışmada kullanılan veri seti, subjektif yorumlardan daha az etkilenecek daha nesnel tahminler üretebilmesi amacıyla internet haberlerinden oluşturulmuştur. CBS News, CNN, Fox News ve NBC News gibi küresel haber kaynakları başta olmak üzere yaklaşık 2900 farklı haber kaynağından Kasım 2022 ve Haziran 2024 arasında 20 ay boyunca toplamda 486.132 haber toplanmıştır. Bu haberler, OpenAI tarafından geliştirilen GPT-4o modeli ile değerlendirilmiş olup, işlem sırasında yaklaşık 1.739.637.269 karakter ve 350.000.000 token kullanılmıştır. BDM'nin büyük ölçekli veri işleme kapasitesi sayesinde, haber içerikleri detaylı olarak analiz edilerek makine öğrenimi algoritmalarının doğruluğunu artıracak yapısal veriler elde edilmiştir.



Şekil 3.2: Model Yapısı

Sonuç olarak, önerilen yöntem büyük dil modelleri ile klasik makine öğrenimi tekniklerini birleştirerek tahminleme süreçlerine yeni bir yaklaşım sunmaktadır. Bu yöntem sayesinde, geleneksel tahminleme modellerine ek olarak haber içeriklerinden elde edilen yapısal veriler ile tahmin doğruluğu artırılmakta, böylece daha güvenilir, tutarlı ve yüksek doğrulukta tahminler üretmek mümkün hale gelmektedir. Bu yaklaşımın, özellikle satış tahminleme ve pazar analizi gibi endüstriyel uygulamalarda geniş bir kullanım alanı bulması beklenmektedir.

3.1. VERİ TOPLAMA

Çalışmada kullanılan veriler, internette yayımlanmış tarih bazlı haberlerden oluşmaktadır. Haberlerin veri olarak tercih edilmesinin temel sebebi, bu tür metinlerin genellikle objektif olup somut bilgilere dayanmalarındır. Haberler, olayları belirli bir zaman diliminde aktardıkları için, tarihsel süreç içinde değişen ve güncellenen bilgi akışını takip etme açısından değerli bir kaynak niteliği taşımaktadır. Özellikle zaman serisi tahminleme için, haberlerin içeriği güncel olayları doğrudan yansıttığından dolayı, dinamik ve gerçek zamanlı bilgi sağlaması onları bu tür analizler için ideal hale getirmektedir.

Bu çalışmada, veri toplama sürecinde belirli bir ürün bazlı filtreleme yapılmaması amaçlanmıştır. Bunun nedeni, tahminleme modelinin geniş kapsamlı ve çok yönlü olmasını sağlamaktır. Önerilen modelin yaklaşımına göre, her haber içeriği doğrudan hedef tahminlerini etkileyebilir. Bir haberin tahmin sürecine ne kadar etki ettiğine dair kesin karar, gelişmiş dil modelleri kullanılarak analiz edilmekte ve değerlendirilmektedir. Büyük dil modeli aracılığıyla yapılan bu değerlendirme, her bir haberin tahmin üzerindeki etkisini ölçerek daha güvenilir sonuçlar elde etmeyi mümkün kılmaktadır.

Veri toplama sürecinde, haberlerin çeşitliliğini ve uluslararası geçerliliğini artırmak amacıyla, farklı yabancı kaynaklardan tarih bazlı haberler derlenmiştir. Bu yaklaşım, tahminleme sürecinin yalnızca belirli bir bölgeye ya da kaynak türüne bağlı kalmamasını sağlamakta; küresel olayları ve geniş ölçekli eğilimleri de kapsayarak analiz modelinin kapsamını genişletmektedir. Böylece, modelin çıktıları daha sağlam ve geniş bir perspektife dayalı tahminler üretmekte olup, yalnızca belirli bir veri kümesine dayanan dar bir bakış açısının ötesine geçmektedir.

3.2. VERİ İŞLEME

Bu çalışmada kullanılan veri toplama süreci, herhangi bir haber filtresi uygulanmaksızın gerçekleştirilmiştir. Bu tercihin temel nedeni, mümkün olduğunca geniş ve kapsayıcı bir veri kümesi oluşturabilmek, böylece elde edilen bilgiler üzerinde herhangi bir önyargı veya sınırlama oluşturmadan analiz yapabilmektir. Haberlerin amacımıza ne kadar uygun ve etkili olacağı ise doğrudan veri işleme aşamasında belirlenmektedir. Veri işleme süreci, ham verinin modellenmesi ve anlamlandırılması ile ilgili kritik bir adımdır. Bu noktada, büyük dil modelleri

(BDM) devreye girerek verinin değerlendirilmesi ve tahminleme süreçlerinde belirleyici bir rol oynamaktadır.

Büyük dil modelleri oluşturulurken, mümkün olduğunca geniş kapsamlı bilgi toplanmakta ve bu bilgiler gelişmiş makine öğrenimi algoritmaları aracılığıyla işlenmektedir. Bu modeller, devasa miktarda veriyi kullanarak dilin yapısını, anlamını ve ilişkilerini öğrenmeye çalışır. Özellikle doğru yapılandırılmış girdi cümleleri (prompts) kullanılarak, bu modellerin elde edilen veriyi işleyip hızlı ve anlamlı sonuçlar üretmesi sağlanmaktadır. Verinin etkin şekilde yorumlanması ve tahminleme yapılabilmesi için, dil modellerinin doğru yönlendirilmesi büyük önem taşımaktadır.

Büyük dil modelleriyle yapılan çalışmaların başarısını belirleyen en önemli unsurlardan biri, girdi (prompt) tasarımıdır. Bu modellerle iletişim, büyük ölçüde sağlanan girdiler üzerinden yürütülmektedir. Doğru ve etkili bir girdi tasarımı, modelin eldeki veriyi daha verimli işlemesini ve istenilen sonuçları daha yüksek doğrulukla üretmesini sağlar. Girdi tasarımının yanlış veya eksik olması durumunda, modelin verdiği çıktılar da beklentileri tam olarak karşılamayabilir. Bu nedenle, büyük dil modelleriyle çalışırken girdi mühendisliği ve veri yönlendirme yöntemleri büyük önem taşımaktadır.

3.2.1. Girdi Tasarımı

Bu çalışma kapsamında, hedeflenen amaç fonksiyonunu büyük dil modeline (BDM) en etkili ve hassas şekilde tanıtmak amacıyla kullanılacak olan girdi, büyük bir titizlikle ve stratejik olarak tasarlanmalıdır. Girdi oluşturma sürecinin temel felsefesi, modelden beklenen ve arzu edilen sonuçların sınırlarını kesin ve net bir dille çizen bir çerçeve oluşturmaktır. Bu çerçevenin ardından, daha önce toplanan verilerin sistematik ve yapılandırılmış bir biçimde modele sunulmasıyla hem bağlama uygun hem de yüksek doğruluk payına sahip tahminler üretilmesi sağlanır. Bu süreç, modelin genel bilgi havuzunu, belirlenen spesifik görev için bir uzman gibi davranmaya yönlendirir.

Bu araştırmanın nihai amaç fonksiyonu, Amerika Birleşik Devletleri pazarındaki araç satışlarına ilişkin dinamikleri doğru bir şekilde öngörmektir. Bu hedefe ulaşma yolunda, analiz edilecek verilerin (haber metinleri, ekonomik raporlar vb.) ve modelin temel eğitim verilerinin büyük çoğunluğunun İngilizce olduğu göz önünde bulundurulmuştur. Bu doğrultuda, modelin metinleri işleme ve anlamsal bağlamı kavrama yeteneğini en üst düzeye çıkarmak için tüm girdiler İngilizce olarak hazırlanmıştır. Bu tercih, çeviri süreçlerinde yaşanabilecek potansiyel

anlam kayıplarını ve belirsizlikleri ortadan kaldırarak, modelin sunulan verileri daha derinlemesine analiz etmesini ve amaç fonksiyonuna yönelik daha kesin ve güvenilir sonuçlar üretmesini hedeflemektedir.

Büyük dil modelleri, kendilerine verilen görevleri yerine getirirken genellikle iki temel girdi bileşenine ihtiyaç duyar. Bu bileşenlerden ilki ve en stratejik olanı komut girdisidir (System Prompt). Komut girdisi, büyük dil modelinin bir nevi "kişiliğini" ve "çalışma talimatlarını" belirleyen unsurdur. Bu girdi sayesinde modele, kendisine yöneltilen soruları hangi sınırlar ve kurallar dahilinde yanıtlaması gerektiği, hangi rolü üstleneceği ve çıktıyı hangi formatta sunması gerektiği öğretilir. Bu, modelin genel amaçlı bir bilgi işlem aracından, belirli bir görev için özelleştirilmiş bir analiz aracına dönüşmesini sağlar. Örneğin:

Tasarlanan İngilizce girdi şu şekildedir:

Dil Modeli Komutu:

- İngilizce: *In this chat, you are a decision maker that decides if the input affects the car sales directly. If it affects directly, you measure the affect and give answer ranging 0 to 100%. If it does not affect directly, you will evaluate the information affect car sales indirectly and give answer ranging 0 to 100%. you will check historically when car sales increased and what news were talking about. No need to give reason. Please provide only direct or indirect and amount of affect.*
- Türkçe: *Bu sohbette, girdinin otomobil satışlarını doğrudan etkileyip etkilemediğine karar veren bir karar vericisiniz. Doğrudan etkiliyorsa, etkiyi ölçüp %0 ila %100 arasında değişen bir cevap verirsiniz. Doğrudan etkilemiyorsa, bilgiyi otomobil satışlarını dolaylı olarak etkileyerek değerlendirirsiniz ve %0 ila %100 arasında değişen bir cevap veriniz. Otomobil satışlarının ne zaman arttığını ve hangi haberlerin konuşulduğunu tarihsel olarak kontrol ediniz. Neden belirtmenize gerek yok. Lütfen yalnızca doğrudan veya dolaylı ve etki miktarını belirtin.*

Bu temel komut verildikten sonra, modelin analiz edeceği asıl veri, yani kullanıcı girdisi (User Prompt) sisteme sunulur. Bu girdi genellikle bir soru ve analizi yapılacak ham metinden (örneğin bir haber metni) oluşur. Oluşturulan soru şu şekildedir:

- İngilizce: *Can you tell me how much of an impact this text has on vehicle sales*

- Türkçe: *Bu metnin araç satışları üzerinde ne kadar etkisi olduğunu bana söyleyebilir misiniz?*

Daha sonra bu soru cümlesinin sonuna, analiz edilmesi istenen haber metni eklenir. Örneğin, 2 Ocak 2023 yılında CNN’de yayınlanan bir haberin içeriği şu şekildedir:

- İngilizce: *Workers who stick with the same employers are enjoying the biggest raises in nearly a quarter century. In November, the median wage growth over the past year for job stayers was 5.5%, according to the Federal Reserve Bank of Atlanta, which began keeping records in 1997. (Job switchers have had it even better, with a median wage growth of 7.7%.) CNN wants to hear from workers who recently received hefty pay hikes to entice them to stay at the same company and from employers who are doling out generous raises to retain their staffs.*
- Türkçe: *Aynı işverenle çalışmaya devam eden çalışanlar, neredeyse çeyrek asırdır en büyük maaş artışlarını alıyor. Atlanta Federal Rezerv Bankası'nın Kasım ayı verilerine göre, iş değiştirenlerin geçen yılki ortalama maaş artışı %5,5 iken, iş değiştirenler için bu oran %7,7'ye kadar çıktı. CNN, yüksek maaş artışlarıyla çalışanlarını elde tutmaya çalışan işverenler ve aynı şirkette kalmaları için cazip zamlar alan çalışanlarla ilgili haberler hazırlıyor.*

Bu haber metni için büyük dil modelinden araç satışları üzerindeki etki miktarını ve nedenini sorduğumuzda, önceden tanımlanmış komut çerçevesinde şöyle bir yanıt alırız (bu örnek için komut girdisinde sebep yazması da istenmiştir):

%60 dolaylı etki, sebebi, tarih boyunca, yükselen maaşlar ve iyileşen finansal koşullar, otomobiller gibi yüksek fiyatlı ürünleri kapsayan tüketici harcamalarını artırmıştır. Çalışanlar anlamlı maaş artışları aldığında, yeni bir araç satın alma veya mevcut araçlarını yükseltme konusunda daha fazla güven duyabilirler. Bu durum, özellikle ekonomik büyüme dönemlerinde, dolaylı olarak otomobil satışlarını teşvik eder.

Görüldüğü üzere bu yapılandırılmış yaklaşım sayesinde büyük dil modeli, kendi geniş veri havuzundaki trilyonlarca parametreyi ve öğrenilmiş bilgiyi kullanarak, sunulan haber metni ile ekonomik göstergeler arasında karmaşık bir sebep-sonuç ilişkisi kurar. Model, kendisine verilen komutlar ve sınırlar doğrultusunda hareket ederek, belirlenen hedef fonksiyon

için hem tutarlı hem de mantıksal bir temele dayanan bir yanıt üretir. Bu, modelin yaratıcılığını ve bilgi sentezleme yeteneğini, araştırmının amaçlarına hizmet edecek şekilde odaklamanın bir yöntemidir.

3.2.2. Çevrimiçi Model ve API Arayüzü

Bu çalışma, uygulama platformu olarak Azure bulut sağlayıcısı kullanılarak gerçekleştirilmiştir. Çalışmada kullanılan BDM doğası, bu tercihin temel nedenini oluşturmaktadır. Şirketlerin tescilli ürünleri olan bazı BDM'ler, yapıları gereği kişisel bilgisayarlarda çalıştırılmaya uygun değildir. Diğer yandan, bazı BDM'ler ise geliştiricileri tarafından sunulan çalıştırılabilir dosyalar aracılığıyla araştırmacıların kendi yerel ortamlarında test yapmalarına imkân tanımaktadır.

BDM'lerin kişisel bilgisayarlarda çalıştırılmasına yönelik çeşitli çalışmalar mevcut olsa da projenin gerektirdiği yüksek hızda ve çeşitlendirilmiş erişim ihtiyaçları göz önüne alındığında, bulut tabanlı çözümler çok daha etkin bir alternatif olarak öne çıkmaktadır. Bu doğrultuda, çalışma kapsamında Microsoft Azure bulut platformunun bir hizmeti olan Azure AI Foundry üzerinde GPT-4o modeli yapılandırılmıştır. Uygulama aşamasında, Bölüm 3.2.1'de tasarlanan komutlar ve girdiler, EK 2'de detaylandırılan API arayüzü vasıtasıyla BDM'e tekil olarak iletilmiştir. Bu süreçte elde edilen sonuçlar sistematik bir biçimde kaydedilmiştir.

3.2.3. Veri Sonuçlarının İşlenmesi

Büyük dil modellerine (BDM) sunulan girdiler aracılığıyla elde edilecek sonuçların formatı ve içeriği, ne kadar detaylı ve özenli hazırlanmış komutlar ile yönlendirilmeye çalışılırsa çalışılsın, her zaman tam bir tutarlılık sergilemeyebilir. Bu durumun temel nedeni, BDM'lerin deterministik yani her seferinde aynı girdiye aynı çıktıyı üreten programlar olmaktan ziyade, temelinde olasılıksal hesaplamalara dayanan istatistiki modeller olmalarıdır. Bir BDM, kendisine verilen metni işlerken, eğitildiği milyarlarca parametrelilik veri okyanusundaki kelime ve kavram ilişkilerine dayanarak en olası sonraki ifadeyi üretir. Bu istatistiki yapı, modellere esneklik ve insan benzeri bir yaratıcılık yeteneği kazandırırken, aynı zamanda aynı anlama gelen bir bilginin farklı sözcükler, formatlar veya yapılarla ifade edilmesine neden olur. Dolayısıyla, bu modellerden elde edilen metin tabanlı çıktıların, özellikle sistematik analiz, toplu değerlendirme veya metrik takibi gibi amaçlar için doğrudan kullanılması zordur. Güvenilir ve tekrarlanabilir bir analiz süreci işletebilmek için bu değişken yapıdaki metinlerin,

standart ve işlenebilir bir form olan sayısal değerlere dönüştürülmesi mutlak bir gereklilik halini alır.

Aşağıda, bu duruma somut bir örnek teşkil eden ve aynı anlamsal içeriği farklı biçimlerde sunan birkaç çıktı örneği listelenmiştir:

- Direct: 0% Indirect: 10%
- Indirect, 10%.
- Direct impact: 0% Indirect impact: 10%
- ****Direct:** 0% ****Indirect:** 10%****

Yukarıdaki örneklerde de net bir şekilde görüleceği üzere, büyük dil modellerinin doğası gereği çıktılar her zaman önceden belirlenmiş tek bir kalıba uygun olmayabilir. Çıktılar bazen ek açıklayıcı kelimeler ("impact" gibi) içerebilir, farklı noktalama işaretleri kullanabilir (virgül, iki nokta üst üste), metin formatlamasından (kalın yazı gibi) etkilenebilir veya bazı durumlarda istenen bilgilerden yalnızca bir kısmını (sadece dolaylı etki gibi) içerebilir.

Bu öngörülemezliği yönetmenin ilk adımı, "girdi tasarımı" aşamasında atılmalıdır. Modelin istenen formata daha sadık kalması için komut içerisinde çıktının yapısı (örneğin "Sadece 'Direct: [yüzde], Indirect: [yüzde]' formatında yanıt ver") hakkında net talimatlar vermek kritik öneme sahiptir. Ancak bu yöntem dahi %100 garanti sunmaz.

Bu nedenle, bu çalışmadaki örnekte olduğu gibi, modelden gelen bu ham ve değişken metin çıktılarının güvenilir bir şekilde işlenebilmesi için bir ara katman geliştirilmiştir. Bu katmanda, bir kod parçası aracılığıyla metin tabanlı çıktılar ayrıştırılarak içerdikleri esas bilgi olan sayısal değerler ayıklanmıştır. Bu temizleme ve dönüşüm işlemi için en etkili araçlardan biri olan Regex (Regular Expression – Kurallı İfade) komutlarından yararlanılmıştır. Regex, metin içerisinde belirli kalıpları (örneğin bir sayı ve ardından gelen yüzde işareti gibi) arayıp bulma konusunda son derece güçlü ve esnek bir yapı sunar. Bu proje kapsamında kullanılan spesifik regex ifadeleri ve açıklamaları EK 3’de detaylı olarak sunulmuştur.

Özetle, bu aşamada gerçekleştirilen temel işlem, bir haber metni için BDM'den alınan değişken formattaki etki analizi çıktısını, geliştirilen kod ve kurallı ifadeler aracılığıyla işleyerek o habere ait doğrudan ve dolaylı etki oranlarını net sayısal değerler olarak elde etmektir. Her bir veri parçası (her bir haber) için tek tek uygulanan bu standartlaştırma işlemi, projenin bir sonraki ve nihai aşaması için temel oluşturmaktadır. Bu aşamada, elde edilen tüm

sayısal veriler toplu olarak değerlendirilecek ve incelenen zaman aralığı için genel etki oranının hesaplanması gibi daha üst düzey analizler yapılacaktır.

3.3. VERİ İŞLEME

Bu çalışmada, modelde kullanılacak özelliklerin türetilmesi amacıyla, ham haber verileri üzerinde kapsamlı bir veri işleme süreci yürütülmüştür. Öncelikli olarak veri temizliği ve model programlama arayüzleri (API) aracılığıyla gerekli bağlantılar kurulmuştur. Bu sürecin sonunda, aylık ortalama 20.000 adet haber verisi işlenerek sanal veri setleri oluşturulmuştur. Her ne kadar bazı aylarda daha az sayıda haber toplanmış olsa da genel ortalama bu seviyede korunmuştur.

Yapılan analizde, işlenen haberlerin yaklaşık %80'inin, temel amaç fonksiyonu olan araç satışları üzerinde farklı derecelerde etkiye sahip olduğu tespit edilmiştir. Bu etkili haberlerin küçük bir bölümü (yaklaşık %0,2) satışları doğrudan etkilerken, etkili haberlerin büyük bir çoğunluğu (yaklaşık %80) ise dolaylı bir etki göstermiştir.

Her bir zaman aralığı için haberlerin kolektif etkisini, modelde kullanılabilir tek bir "özellik ağırlığına" dönüştürmek amacıyla, etki puanı 0'dan büyük olan haberlerin aritmetik ortalamasını hesaplayan bir yöntem kullanılmıştır. Bu hesaplama, aşağıdaki formül (3.1) ile gerçekleştirilmiştir:

$$\frac{\sum_{F \in A} \omega_F [\omega_F > 0]}{|\{F \in A | \omega_F > 0\}|} \quad (3.1)$$

Bu formülde;

- A : Belirli bir zaman aralığındaki (çalışmamızda aylık) tüm haber verileri kümesini,
- F : Bu kümedeki tek bir haber verisini,
- ω_F : İlgili haberin 0 ile 100 arasında bir değere sahip olan doğrudan veya dolaylı etki puanını temsil eder.

Formülün pay kısmı, pozitif etkiye sahip haberlerin etki puanları toplamını verirken (Doğrudan Toplam ve Dolaylı Toplam), payda kısmı ise bu pozitif etkili haberlerin sayısını belirtir.

Araç satışları üzerindeki etkileri ayrıştırabilmek adına bu formül, her ay için iki ayrı defa çalıştırılmıştır: biri yalnızca doğrudan etki puanları, diğeri ise yalnızca dolaylı etki puanları için. Bu sayede her ay için iki temel özellik ağırlığı elde edilmiştir: Doğrudan Ortalama ve Dolaylı Ortalama. Bu değerler, sonraki tahmin modellemesinde kullanılacak olan nihai özelliklerdir.

Bu hesaplamaların aylık bazda detaylı sonuçları Tablo 3.1'de sunulmaktadır. Tablodaki sütunlar şu şekilde açıklanmaktadır:

- **Tarih:** Verilerin ait olduğu ay ve yılı gösterir.
- **Haber Sayısı:** İlgili ay içinde işlenen toplam haber adedidir.
- **Etkili Haber:** Amaç fonksiyonunu doğrudan veya dolaylı olarak etkileyen toplam haber sayısıdır.
- **Doğrudan Etkili / Dolaylı Etkili:** Formül (3.1)'in payda kısmına karşılık gelir. Etkili haberlerin etki türüne göre (doğrudan veya dolaylı) dağılımını gösteren haber sayılarıdır.
- **Doğrudan Toplam / Dolaylı Toplam:** Formül (3.1)'in pay kısmına karşılık gelir. İlgili kategorideki pozitif etkili tüm haberlerin etki puanlarının kümülatif toplamını gösterir.
- **Doğrudan Ortalama / Dolaylı Ortalama:** Formül (3.1)'in nihai sonucunu temsil eder. İlgili ay için modelde özellik olarak kullanılacak olan doğrudan ve dolaylı etki ağırlıklarını gösterir.

Tablo 3.1: Haberlerin Büyük Dil modeli ile işlem sonuçları

Tarih	Haber Sayısı	Etkili Haber	Doğrudan Etkili	Dolaylı Etkili	Doğrudan Toplam	Dolaylı Toplam	Doğrudan Ortalama	Dolaylı Ortalama
Kasım 22	19616	8840	29	8834	1170	123250	40,34	13,95
Aralık 22	17879	15284	71	15269	3065	255150	43,17	16,71
Ocak 23	23866	20953	101	20921	5600	306890	55,45	14,67
Şubat 23	4326	3866	24	3861	1160	58500	48,33	15,15
Mart 23	7496	6501	25	6498	1065	86630	42,60	13,33
Nisan 23	14018	12715	30	12708	1555	161255	51,83	12,69
Mayıs 23	33504	28875	100	28857	4430	362020	44,30	12,55
Haziran 23	21072	18729	92	18713	3890	255955	42,28	13,68
Temmuz 23	18141	16126	20	16118	890	213440	44,50	13,24
Ağustos 23	28171	16516	23	16508	1170	215775	50,87	13,07
Eylül 23	25754	22383	27	22375	1165	295210	43,15	13,19
Ekim 23	5176	4591	23	4580	1555	64370	67,61	14,05
Kasım 23	34612	24055	63	24031	3155	316890	50,08	13,19
Aralık 23	39285	32118	72	32094	4165	406376	57,85	12,66
Ocak 24	43393	33072	87	33052	4500	389976	51,72	11,80
Şubat 24	30711	26994	87	26954	4470	362245	51,38	13,44
Mart 24	25264	21582	100	21482	6860	288620	68,60	13,44
Nisan 24	18695	16432	71	16361	4845	221725	68,24	13,55
Mayıs 24	36635	32635	114	32605	5380	437025	47,19	13,40
Haziran 24	38522	34090	97	34067	4405	455935	45,41	13,38

3.4. TAHMİNLEME

Bu çalışmada, makine öğrenimi tabanlı tahminleme modelleri otomotiv sektörü üzerine odaklanılarak geliştirilmiştir. Bu sektörün seçilmesindeki temel etken, hem araç satış verilerine hem de bu satışları etkileyebilecek makroekonomik göstergelere kamuya açık kaynaklar üzerinden kolaylıkla erişilebilmesidir. Bu durum, araştırmanın tekrarlanabilirliği ve doğrulanabilirliği açısından önemli bir avantaj sunmaktadır.

Araç satışlarını etkileyen makroekonomik faktörler, literatürde çeşitli çalışmalarla incelenmiştir. Örneğin, Sa-ngasoongsong ve diğ. (2012) tarafından yapılan çalışmada, küçük ve büyük tip araç satışlarını tahmin etmek amacıyla tüketici fiyat endeksi, işsizlik oranı, yakıt fiyatları ve inşaatına başlanan yeni konut sayıları gibi değişkenler kullanılmıştır. Benzer bir yaklaşımla, Homolka ve diğ. (2020) de çalışmalarında işsizlik endeksi, toplu taşıma

maliyetlerini etkileyen enflasyon, kredi faiz oranları ve enerji fiyatları endeksleri gibi ekonomik göstergelerden faydalanmıştır.

Bu çalışmalardan hareketle ve veri erişilebilirliği göz önünde bulundurularak, modelde kullanılacak parametreler şu şekilde belirlenmiştir: işsizlik oranı, enflasyon (tüketici fiyat endeksi), yakıt fiyatları, enerji fiyatları, konut kredisi oranları, taşıt kredisi oranları ve yeni inşa edilen konut sayısı. Bu parametrelerin her biri, tüketici davranışları ve alım gücü üzerinde farklı etkilere sahiptir. Örneğin, yakıt fiyatları genel olarak artan bir trend sergileyen önemli bir maliyet kalemidir. İşsizlik oranı ve yeni konut sayısı birbiriyle ilişkili görünse de yeni konut sayısı verisi kısa vadeli ekonomik dalgalanmalara karşı daha dirençli bir yapı sergileyebilmektedir. Tüketici fiyat endeksi ise ekonomideki genel fiyat artışlarını yansıtarak sürekli bir artış trendi göstermektedir.

Çalışmanın uygulama alanı olarak Amerika Birleşik Devletleri (ABD) pazarı hedeflenmiştir. Klasik yöntemde kullanılacak olan istatistiksel veriler, Federal Reserve Economic Data (FRED) ve The Census Bureau gibi güvenilir ve kamuya açık bilgi kaynaklarından temin edilmiştir. Analizde kullanılan veri seti, Ekim 2022 ile Haziran 2024 arasındaki zaman aralığını kapsamaktadır. Verilerin temin edildiği kaynaklar ve ilgili kodlar Tablo 3.2'de detaylandırılmıştır. Bu çalışmada kullanılan ve Ekim 2022 – Haziran 2024 zaman aralığını kapsayan veriler ise Tablo 3.3'de listelenmektedir.

Tablo 3.2: İstatistik Veri Kaynakları

	Kod	URL
İşsizlik Endeksi	UNRATE	https://fred.stlouisfed.org/series/UNRATE
Akaryakıt Endeksi	APU000074714	https://fred.stlouisfed.org/series/APU000074714
Tüketici Endeksi	MEDCPIM158SFRBCLE	https://fred.stlouisfed.org/series/MEDCPIM158SFRBCLE
Yeni Ev Endeksi		https://www.census.gov/construction/nrc/data/series.html
Bireysel Kredi Endeksi	TERMCBPER24NS	https://fred.stlouisfed.org/series/TERMCBPER24NS
Araç Kredi Endeksi	TERMCBAUTO48NS	https://fred.stlouisfed.org/series/TERMCBAUTO48NS
Araç Satış Endeksi	TOTALSA	https://fred.stlouisfed.org/series/TOTALSA

Modelde kullanılan endeksler şu şekilde tanımlanmaktadır:

İşsizlik Endeksi: 16 yaş ve üstü çalışabilir nüfus içindeki işsizlerin oranını yüzde olarak ifade eder.

- **Akaryakıt Endeksi:** Kurşunsuz benzinin galon (yaklaşık 3,785 litre) başına ortalama perakende fiyatını ABD doları cinsinden gösterir.
- **Tüketici Fiyat Endeksi:** ABD genelinde tüketicilerin maruz kaldığı ortalama enflasyon oranını yansıtır.
- **Yeni Ev Endeksi:** İnşaatı tamamlanmış yeni konutların sayısını binlik birimler halinde belirtir.
- **Bireysel Kredi Endeksi:** Ticari bankalar tarafından sunulan 24 ay vadeli bireysel kredilerin ortalama yıllık faiz oranını gösterir.
- **Araç Kredi Endeksi:** Ticari bankalar tarafından verilen 48 ay vadeli yeni araç kredilerinin ortalama yıllık faiz oranını belirtir.
- **Toplam Araç Sayısı Endeksi (Hedef Değişken):** Ay içerisinde satılan toplam araç sayısını milyon adet olarak gösterir.

Bu çalışmanın nihai amacı, yukarıda listelenen ilk altı makroekonomik endeksi (işsizlik, akaryakıt, tüketici, yeni ev, bireysel kredi ve araç kredi endeksleri) girdi olarak kullanarak farklı makine öğrenimi algoritmaları ile bir temel model oluşturmaktır. Bu temel model, toplam araç satışlarını tahmin edecektir. Ardından, bu modellere geliştirilen yeni bir özelliğin eklenmesiyle tahmin performansında ne denli bir iyileşme sağlandığı kantitatif olarak gösterilecektir. Bu sayede, geleneksel ekonomik göstergelerin yanı sıra yeni ve alternatif veri kaynaklarının tahminleme gücü üzerindeki etkisi de ortaya konulacaktır.

Tablo 3.3: Araç Satış ve İstatistik Veriler

Tarih	İşsizlik Endeksi	Akaryakıt Endeksi	Müşteri Fiyat Endeksi	Yeni Ev Endeksi	Bireysel Kredi Endeksi	Araç Kredi Endeksi	Toplam Araç Satışı
Kasım 22	3,6	3,85	6,15	1559	11,21	6,94	14,73
Aralık 22	3,5	3,36	5,52	1386	11,21	6,94	13,94
Ocak 23	3,5	3,45	6,73	1389	11,21	6,94	15,67
Şubat 23	3,6	3,51	7,16	1540	11,48	7,46	15,43
Mart 23	3,5	3,55	4,70	1516	11,48	7,46	15,67
Nisan 23	3,4	3,74	5,67	1416	11,48	7,46	16,45
Mayıs 23	3,6	3,69	4,81	1499	11,48	7,59	15,95
Haziran 23	3,6	3,71	4,70	1480	11,48	7,59	16,43
Temmuz 23	3,5	3,73	3,23	1343	11,48	7,59	16,30
Ağustos 23	3,7	3,96	4,21	1373	12,17	8,30	15,91
Eylül 23	3,8	3,99	4,84	1466	12,17	8,30	16,19
Ekim 23	3,9	3,78	4,17	1382	12,17	8,30	15,81
Kasım 23	3,7	3,50	5,54	1466	12,35	8,51	15,94
Aralık 23	3,8	3,29	4,19	1557	12,35	8,51	16,39
Ocak 24	3,7	3,22	5,72	1504	12,35	8,51	15,54
Şubat 24	3,9	3,36	4,30	1698	12,49	8,57	16,18
Mart 24	3,9	3,58	4,73	1491	12,49	8,57	16,14
Nisan 24	3,9	3,78	4,25	1659	12,49	8,57	16,53
Mayıs 24	4,0	3,76	2,81	1557	11,92	8,65	16,31
Haziran 24	4,1	3,60	2,50	1725	11,92	8,65	15,43

3.4.1. Doğrusal Regresyon Algoritması

Bu çalışmada, bağımlı bir değişken ile bu değişkeni etkileyen bir veya daha fazla bağımsız değişken arasındaki doğrusal ilişkiyi modellemek amacıyla istatistiksel bir yöntem olan doğrusal regresyon analizi kullanılmıştır. Bu analizin temel amacı, bağımsız değişkenlerin bağımlı değişken üzerindeki nicel etkilerini belirleyerek geleceğe yönelik öngörülerde bulunmak ve verilerdeki temel eğilimleri ortaya çıkarmaktır. Model, bağımlı değişkenin, bağımsız değişkenlerin ağırlıklı bir toplamı olarak ifade edildiği matematiksel bir denklem üzerine kuruludur. Bu denklemdaki katsayılar, modelin tahmin hatalarının kareleri toplamını en aza indirmeyi hedefleyen En Küçük Kareler (EKK) yöntemi kullanılarak titizlikle hesaplanır. Doğrusal regresyon, esnekliği ve yorumlana bilirliliği sayesinde ekonomi, mühendislik, tıp ve sosyal bilimler gibi çok çeşitli disiplinlerde tahminleme ve veri analizi süreçlerinin vazgeçilmez bir aracıdır. Konunun teorik altyapısı ve matematiksel temelleri, raporun 2.2.1 numaralı bölümünde detaylı bir şekilde incelenmiştir.

Analiz, Tablo 3.3'te sunulan veri seti kullanılarak gerçekleştirilmiştir. Modelin performansını objektif bir şekilde değerlendirebilmek için veri seti, eğitim ve test verileri olarak ikiye ayrılmıştır. Bu kapsamda, ilk 15 aylık döneme ait veriler modelin eğitilmesi (öğrenmesi) için eğitim verisi olarak, geriye kalan 5 aylık veri ise modelin genelleme yeteneğini ve tahmin doğruluğunu test etmek amacıyla test verisi olarak kullanılmıştır. Bu yaklaşım, modelin sadece mevcut veriyi ezberlemesini önleyerek, daha önce karşılaşmadığı veriler üzerindeki performansını ölçmeyi sağlar. Doğrusal regresyon modellemesi gereği, her bir bağımsız özellik (değişken) için bir katsayı (β) ve modelin açıklayamadığı rassal değişkenliği ifade eden bir hata payı (ϵ) hesaplanmıştır. Bu hesap için kullanılan kod parçası EK 4'te paylaşılmıştır.

İlk analizde, veri analizinden elde edilen ve sisteme dair ek bilgiler sunan "doğrudan etki" ve "dolaylı etki" endeksleri de dahil olmak üzere toplam sekiz adet bağımsız değişken kullanılmıştır. Bu modelin matematiksel ifadesi denklem (3.2)'de gösterilmiştir:

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \epsilon \quad (3.2)$$

Yapılan hesaplamalar sonucunda modelin katsayıları ve hata payı aşağıdaki gibi belirlenmiştir:

Özellik 1 (x_1): İşsizlik Endeksi, $\beta_1 = -3,06$

Özellik 2 (x_2): Akaryakıt Endeksi, $\beta_2 = 1,09$

Özellik 3 (x_3): Müşteri Fiyat Endeksi, $\beta_3 = -0,11$

Özellik 4 (x_4): Yeni Ev Endeksi, $\beta_4 = 0$

Özellik 5 (x_5): Bireysel Kredi Endeksi, $\beta_5 = -2,93$

Özellik 6 (x_6): Araç Kredi Endeksi, $\beta_6 = 2,94$

Özellik 7 (x_7): Doğrudan Etki Endeksi, $\beta_7 = 0,05$

Özellik 8 (x_8): Dolaylı Etki Endeksi, $\beta_8 = -0,06$

Hata pay (ϵ): **28,16**

Model performansını değerlendirmek ve yeni eklenen özelliklerin katkısını net bir şekilde ölçmek amacıyla, "doğrudan etki" ve "dolaylı etki" endekslerinin çıkarıldığı daha klasik

bir model oluşturulmuştur. Bu model, altı bağımsız değişkenden oluşmaktadır ve denklem (3.3) ile ifade edilmiştir:

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \varepsilon \quad (3.3)$$

Bu model için hesaplanan katsayılar ve hata payı ise şöyledir:

Özellik 1 (x_1): İşsizlik Endeksi, $\beta_1 = -1,77$

Özellik 2 (x_2): Akaryakıt Endeksi, $\beta_2 = 0,62$

Özellik 3 (x_3): Müşteri Fiyat Endeksi, $\beta_3 = -0,05$

Özellik 4 (x_4): Yeni Ev Endeksi, $\beta_4 = 0$

Özellik 5 (x_5): Bireysel Kredi Endeksi, $\beta_5 = -2,96$

Özellik 6 (x_6): Araç Kredi Endeksi, $\beta_6 = 3,14$

Hata pay (ε): **28,27**

İki modelin performans karşılaştırması, Ortalama Karesel Hata (OKH) metriği kullanılarak yapılmıştır. OKH, modelin tahmin değerleri ile gerçek değerler arasındaki farkların karesinin ortalamasını alarak modelin genel hata düzeyini ölçer.

Klasik yöntemle (altı özellikli model) yapılan analiz sonucunda Ortalama Karesel Hata değeri 1,58 olarak hesaplanmıştır. Aynı veri setine, veri analizi yoluyla türetilen "doğrudan etki" ve "dolaylı etki" özellikleri eklendiğinde ise Ortalama Karesel Hata değeri 1,15 seviyesine gerilemiştir.

Bu sonuç, OKH değerinde yaklaşık %27'lik bir düşüş anlamına gelmektedir. Bu durum, modele eklenen yeni özelliklerin, bağımlı değişkendeki varyansın daha büyük bir kısmını açıklamayı başardığını ve modelin tahmin gücünü istatistiksel olarak anlamlı bir düzeyde artırdığını açıkça göstermektedir. Yeni özelliklerin eklenmesiyle tahminlerde elde edilen bu belirgin iyileşme BULGULAR bölümünde detaylı olarak tartışılacaktır.

3.4.2. Destek Vektör Regresyon Algoritması

Bu çalışmada kullanılan tahminleme modeli, denetimli bir makine öğrenmesi algoritması olan Destek Vektör Makinesi (DVM) prensiplerine dayanan Destek Vektör Regresyonudur (DVR). DVM, temel olarak verileri N-boyutlu bir uzayda sınıflandırmak amacıyla, her sınıf arasındaki ayrım mesafesini (marjı) en üst düzeye çıkaran bir çizgi veya hiper düzlem belirler.

DVR ise bu metodolojiyi regresyon problemlerine uygular. DVM'lere benzer şekilde, DVR da veri noktaları etrafında maksimum marja sahip bir hiper düzlem bulmayı hedefler ve bu özelliği sayesinde özellikle zaman serisi tahminlemesinde sıklıkla başvurulmuş bir yöntemdir. Algoritmanın ayrıntılı hesaplama yöntemleri Bölüm 2.2.2'de sunulmuştur.

DVR algoritmasının temelinin oluşturduğu fonksiyon şu şekildedir:

$$f(x) = \omega^T x + b \quad (3.4)$$

Bu denklemde;

- x , giriş özelliklerini içeren vektörü,
- ω , bu özelliklere atanan ağırlık vektörünü,
- b , modelin sapma (bias) veya yanılma payını ifade eder.

Gerçekleştirilen iki farklı deneysel kurulumda 6 ve 8 özellik kullanıldığı için, her bir model için sırasıyla 6 ve 8 bileşenli ağırlık vektörleri (ω) hesaplanmıştır. Problem için "lineer" çekirdek (kernel) fonksiyonu seçilmiştir. Bu seçimle model, sekiz giriş özelliğinin bir sapma değeri ile ayarlanmış basit bir doğrusal kombinasyonunu oluşturarak tahminleme yapar. DVR algoritmasının görevi, hatayı en aza indirirken modeli olabildiğince basit tutacak şekilde en uygun ω ve b değerlerini bulmaktır.

DVR gibi makine öğrenmesi algoritmalarının verimliliği, büyük hacimli veri setleri ile artmaktadır. Bu bağlamda, modelin performansını ve geliştirilen özelliklerin etkisini ölçmek amacıyla iki aşamalı bir deney tasarlanmıştır.

İlk aşamada, modelin performansı daha küçük bir veri seti üzerinde test edilmiştir. Bu amaçla, Kasım 2022 – Haziran 2024 arasını kapsayan 20 aylık döneme ait 20 veri noktası

kullanılmıştır. Bu deneyde, veri setinin tamamı hem modelin eğitimi hem de testi için kullanılmıştır. Bu süreç, iki farklı model konfigürasyonu için yürütülmüştür:

1. Geliştirilmiş Model: Bu modelde, standart değişkenlere ek olarak bu çalışma kapsamında oluşturulan "doğrudan etki" ve "dolaylı etki" özellikleri de modele dahil edilmiştir. Hesaplanan hata payı 0,25 olarak bulunmuştur.
2. Klasik Model: Bu modelde ise yeni oluşturulan iki özellik kullanılmamış, yalnızca klasik tahminleme değişkenleri ile eğitim yapılmıştır. Hesaplanan hata payı 0,27 olarak tespit edilmiştir.

DVR algoritmasının yapısına daha uygun bir zemin hazırlamak amacıyla, aylık veriler haftalık frekansa dönüştürülerek daha büyük bir veri seti oluşturulmuştur. Tablo 3.1’de yer alan aylık istatistiki bilgiler (örneğin endeksler), ilgili ayın tüm haftaları için tekrar ettirilerek toplam 84 veri noktası üretilmiştir. Oluşturulan bu veri setinden bir örnek Tablo 3.4’te gösterilmektedir. Zenginleştirilmiş verinin tamamı EK 7’de sunulmuştur.

Tablo 3.4: Araç Satış İstatistikleri Haftalık Örnek

Tarih	Doğrudan Etki	Dolaylı Etki	İşsizlik Endeksi	Akaryakıt Endeksi	Tüketici Fiyat Endeksi	Bitmiş Yeni Ev Endeksi	Bireysel Kredi Endeksi	Araç Kredisi Endeksi	Toplam Satış
07/01/2024	54,55	12,41	3,70	3,22	5,72	1504	12,35	8,51	3,51
14/01/2024	50,71	11,85	3,70	3,22	5,72	1504	12,35	8,51	3,51
21/01/2024	41,18	11,22	3,70	3,22	5,72	1504	12,35	8,51	3,51
28/01/2024	55,14	11,90	3,70	3,22	5,72	1504	12,35	8,51	3,74
04/02/2024	59,09	14,11	3,90	3,35	4,30	1698	12,49	8,57	4,04
11/02/2024	48,62	13,56	3,90	3,35	4,30	1698	12,49	8,57	4,04
18/02/2024	51,07	12,84	3,90	3,35	4,30	1698	12,49	8,57	4,04

Makine öğrenmesi standartlarına uygun olarak, bu 84 veri noktasının ilk %75'lik bölümü (63 veri noktası) modelin eğitimi, kalan %25'lik bölümü (21 veri noktası) ise modelin testi için ayrılmıştır.

Bu yöntemle çalıştırılan modellerin performansları karşılaştırıldığında, yeni özellikleri içeren modelin ortalama karesel hatası (OKH) 0,39 olarak gerçekleşirken, klasik yöntemle kurulan modelin OKH değeri 0,50 olarak hesaplanmıştır. Gözlemlenen bu iyileşme, "Bulgular" bölümünde ayrıntılı olarak ele alınacaktır.

DVR algoritmasının çalıştırılması ve sonuçların elde edilmesi süreçlerinde Python programlama dili ve bu dilin sklearn (Pedregosa ve diğ., 2011) kütüphanesi kullanılmıştır. Uygulamaya ait ilgili kod parçası EK 5’de sunulmuştur.

3.4.3. Gradyan Hızlandırma Regresyon Algoritması

Son yıllarda yapılan akademik ve endüstriyel çalışmalarda, talep tahmini problemlerinin çözümünde Gradyan Hızlandırma Regresyonu (GHR) modelinin etkin bir şekilde kullanıldığı gözlemlenmektedir. Literatürdeki yerini sağlamlaştıran bu yöntem, farklı makine öğrenmesi algoritmalarıyla karşılaştırıldığında öne çıkan sonuçlar sunmaktadır. Örneğin, Saha ve diğ. (2022) tarafından yürütülen bir çalışmada, GHR modelinin zaman serisi analizlerinde sıkça kullanılan LSTM (Long Short-Term Memory) mimarisine kıyasla bir miktar daha yüksek performans sergilediği rapor edilmiştir. Benzer şekilde, Deng ve diğ. (2021) GHR algoritmasını temel bileşenleri olan karar ağaçları ile güçlendirerek yenilikçi bir yaklaşım geliştirmiştir. Bu çalışmalarında, geliştirdikleri modeli Destek Vektör Makineleri (DVM) ve lojistik regresyon gibi köklü yöntemlerle karşılaştırmış ve GHR tabanlı modelin istatistiksel olarak daha isabetli tahminler ürettiğini ortaya koymuşlardır.

Gradyan Hızlandırma Regresyonunun çalışma prensibi, ardışık bir öğrenme sürecine dayanır. Model, genellikle "zayıf öğrenciler" olarak nitelendirilen basit modelleri (çoğunlukla karar ağaçları) sıralı bir şekilde inşa eder. Bu süreçte eklenen her yeni model, bir önceki modelin tahminlerinde yaptığı hataları (kalıntıları) düzeltmek üzere eğitilir. Hataların bu kademeli olarak azaltılması, bireysel tahmincilerin zayıf yönlerini telafi ederken güçlü yönlerini birleştiren, nihayetinde yüksek doğruluk oranına sahip güçlü bir bütünleşik modelin ortaya çıkmasını sağlar. Algoritmanın teorik altyapısı ve matematiksel formülasyonu hakkında daha kapsamlı bilgilere bu çalışmanın 2.2.3 numaralı bölümünde yer verilmiştir.

Bu tez çalışması kapsamında, GHR modelinin performansı üzerinde yeni türetilen özelliklerin etkisini ölçmek amacıyla bir dizi deney gerçekleştirilmiştir. Modelin hiperparametreleri, literatürdeki yaygın kullanımlar dikkate alınarak karar ağacı sayısı ($n_estimators$) için 100 ve öğrenme katsayısı ($learning_rate$) için 0,1 olarak belirlenmiştir.

İlk deneyde, Kasım 2022 ile Haziran 2024 arasındaki 20 aylık dönemi kapsayan ve 20 veri noktasından oluşan bir veri seti kullanılmıştır. Bu aşamada, modelin performansını değerlendirmek için eğitim verisinin tamamı aynı zamanda test verisi olarak kullanılmıştır. Bu senaryoda elde edilen sonuçlar aşağıdaki gibidir:

- **Yeni Özellikler Eklendiğinde:** Ortalama Karesel Hata (OKH) değeri 0,0000036 olarak hesaplanmıştır.
- **Yeni Özellikler Eklenmediğinde:** Ortalama Karesel Hata (OKH) değeri 0,0000115 olarak hesaplanmıştır.

Bu sonuçlar, önerilen yeni özelliklerin model performansını artırdığını göstermekle birlikte, OKH değerlerinin gerçekçi olmayan bir şekilde düşük çıkması, test ve eğitim verilerinin aynı olmasından kaynaklanan bir durumdur. Bu durum, modelin veriyi "ezberlediğini" ve genelleme yeteneğinin doğru bir şekilde ölçülemediğini göstermektedir.

Bu sınırlılığı ortadan kaldırmak amacıyla, Bölüm 3.4.2’de detaylandırıldığı üzere ikinci bir deney tasarlanmıştır. Veri seti, aylık periyotlar yerine haftalık periyotlara bölünerek zenginleştirilmiş ve veri noktası sayısı 20’den 84’e çıkarılmıştır. Daha sağlıklı bir model değerlendirmesi için bu zenginleştirilmiş veri seti, %75 eğitim ve %25 test verisi olacak şekilde ikiye ayrılmıştır. Örnek veri yapısı Tablo 3.4’te sunulmuştur. Zenginleştirilmiş verilerin tamamı EK 7’de sunulmuştur. Bu metodoloji ile model yeniden çalıştırıldığında elde edilen OKH değerleri şöyledir:

- **Yeni Özellikler Eklendiğinde:** OKH değeri 0,031 olarak hesaplanmıştır.
- **Yeni Özellikler Eklenmediğinde:** OKH değeri 0,038 olarak hesaplanmıştır.

Bu ikinci deney, modelin daha gerçekçi veriler üzerinde test edildiği koşullarda dahi yeni özelliklerin performansta bir iyileşme sağladığını göstermektedir. Gözlemlenen bu iyileşmenin istatistiksel anlamlılığı ve pratik sonuçları, "BULGULAR" bölümünde detaylı olarak tartışılmıştır.

Söz konusu GHR algoritmasının uygulanması ve deneylerin gerçekleştirilmesi için Python programlama dili ve bu dilin popüler makine öğrenmesi kütüphanesi olan sklearn (Pedregosa ve diğ., 2011) kullanılmıştır. Uygulama sırasında kullanılan Python kod bloğunun bir örneği EK 6’te sunulmuştur.

4. BULGULAR

Bu çalışma, yapısal olmayan devasa çevrim içi veri yığınlarının, Büyük Dil Modelleri (BDM'ler) aracılığıyla işlenerek nasıl yapısal ve anlamlı özelliklere dönüştürülebileceğini ve bu yeni özelliklerin geleneksel talep tahmin modellerinin performansını ne ölçüde iyileştirebileceğini göstermeyi amaçlamaktadır. Çalışmanın özünde, BDM'ler ile üretilen ve Tablo 3.1'de detaylandırılan yeni özellik setinin, klasik ekonometrik göstergelerle birleştirilerek üç temel tahmin algoritmasının, Doğrusal Regresyon, Destek Vektör Regresyonu (DVR) ve Gradyan Hızlandırma Regresyonu (GHR) doğruluğuna olan etkisini ölçmek yatmaktadır.

4.1. DOĞRUSAL REGRESYON ANALİZİ

Analizin ilk aşamasında, Tablo 3.3'te paylaşılan standart özellikler kullanılarak bir Doğrusal Regresyon modeli kurulmuştur. 20 veri noktalık setin %75'i eğitim, %25'i test için kullanılarak yapılan bu "Klasik Yöntem" deneyinde, Ortalama Karesel Hata (OKH) 1,59 olarak hesaplanmıştır.

Çalışmanın asıl katkısı olan ve BDM'ler ile üretilen yeni özelliklerin modele eklenmesiyle "Önerilen Yöntem" tahmin doğruluğunda belirgin bir artış gözlemlenmiştir. Tablo 4.1'de sunulan hata metrikleri bu iyileşmeyi net bir şekilde ortaya koymaktadır.

Tablo 4.1: Doğrusal Regresyon Hata İyileşmeleri

	Ortalama Karesel Hata	Ortalama Mutlak Yüzde Hata	Ortalama Mutlak Hata
Klasik Yöntem	1,58	0,06	1,00
Önerilen Yöntem	1,15	0,05	0,80
İyileşme	%27	%19	%19

Ortalama Karesel Hata (OKH), 1,58'dan 1,15'ya gerileyerek %27'lik bir iyileşme sağlamıştır. Ortalama Mutlak Yüzde Hata (OMYH) %19, Ortalama Mutlak Hata (OMH) ise yine %19 oranında azalmıştır.

Bu istatistiksel iyileşmenin pratikteki yansımaları Tablo 4.2'deki tahmin sonuçlarında görülebilir. Örneğin, Nisan 2024 için gerçek satış değeri 16,53 iken, klasik yöntem 16,07'lik bir tahminde bulunmuş, önerilen yöntem ise 17,07 ile daha farklı bir dinamik yakalamıştır. Benzer şekilde, Mart 2024'te önerilen yöntemin tahmini (16,21), klasik yöntemin tahminine (15,66) kıyasla gerçek değere (16,14) çok daha yakındır.

Tablo 4.2: Doğrusal Regresyon Analizi ile Sonuçlar

Tarih	Toplam Araç Satışı	Klasik Yöntem Tahmin	Sana Veri ile Tahmin
Şubat 24	16,18	15,88	15,93
Mart 24	16,14	15,66	16,21
Nisan 24	16,53	16,07	17,07
Mayıs 24	16,31	17,74	17,40
Haziran 24	15,43	17,75	17,49

Modelin iç yapısını incelemek için Tablo 4.3'teki katsayılara bakıldığında önemli bir bulgu ortaya çıkmaktadır: "Yeni Ev Endeksi" değişkeninin katsayısı her iki modelde de 0'dır. Bu durum, bu değişkenin araç satışları üzerinde tek başına doğrusal bir etkiye sahip olmadığını göstermektedir. Ancak önerilen yöntemde "İşsizlik Endeksi" (-3,06) ve "Akaryakıt Endeksi" (1,09) gibi değişkenlerin katsayılarının mutlak değerinin artması, yeni verilerin modelin diğer değişkenler arasındaki korelasyonu daha doğru yorumlamasına yardımcı olduğunu göstermektedir.

Tablo 4.3: Doğrusal Regresyon Katsayıları

	Önerilen Yöntem	Klasik Yöntem
İşsizlik Endeksi	-3,06	-1,77
Akaryakıt Endeksi	1,09	0,62
Müşteri Fiyat Endeksi	-0,11	-0,05
Yeni Ev Endeksi	0	0
Bireysel Kredi Endeksi	-2,93	-2,96
Araç Kredi Endeksi	2,94	3,14
Doğrusal Etki	0,05	
Dolaylı Etki	-0,06	
Hata payı	28,16	28,27

4.2. DESTEK VEKTÖR REGRESYON ANALİZİ

Doğrusal olmayan ilişkileri modellemede yetkin olan DVR üzerinde iki farklı deney yapılmıştır. İlk olarak, 20 aylık veri seti hem eğitim hem de test için kullanılmıştır. Tablo 4.4'te sonuçları verilen bu deneyde, önerilen yöntemin klasik yönteme göre bir miktar iyileşme sağladığı görülmüştür.

Tablo 4.5'te özetlendiği üzere, OKH'de %7, OMYH ve OMH'de ise %5'lik bir iyileşme kaydedilmiştir.

Tablo 4.4: Aylık Veriler ile DVR Sonuçları

Tarih	Toplam Satış	Yeni Talep Tahmini	Klasik Talep Tahmini
Kasım 22	14,73	16,09	15,57
Aralık 22	13,94	15,01	15,63
Ocak 23	15,67	15,66	15,55
Şubat 23	15,43	15,46	15,52
Mart 23	15,67	16,05	15,81
Nisan 23	16,45	16,28	15,88
Mayıs 23	15,95	16,27	15,81
Haziran 23	16,43	15,99	15,84
Temmuz 23	16,30	16,14	16,10
Ağustos 23	15,91	16,08	16,31
Eylül 23	16,19	15,93	16,16
Ekim 23	15,81	16,16	16,17
Kasım 23	15,94	15,57	16,09
Aralık 23	16,39	16,05	16,04
Ocak 24	15,54	15,71	15,96
Şubat 24	16,18	15,94	15,99
Mart 24	16,14	16,13	16,15
Nisan 24	16,53	16,63	16,15
Mayıs 24	16,31	16,24	16,00
Haziran 24	15,43	16,39	15,84

Tablo 4.5: Aylık Veriler ile DVR Hata İyileşmeleri

	Ortalama Karesel Hata	Ortalama Mutlak Yüzde Hata	Ortalama Mutlak Hata
Klasik Yöntem	0,27	0,024	0,37
Önerilen Yöntem	0,25	0,023	0,35
İyileşme	%7	%5	%5

DVR'ın gerçek potansiyelini ortaya koyan ikinci deneyde, aylık veri seti haftalık frekansa dönüştürülerek 20 olan veri noktası sayısı 84'e çıkarılmıştır. Bu zenginleştirilmiş veri setinin %75'i eğitim, %25'si test için kullanılmıştır. Bu metodoloji değişikliği, sonuçlarda devrim niteliğinde bir etki yaratmıştır. Tablo 4.6'de sunulan hata iyileşmeleri, çalışmanın en dikkat çekici bulgusudur. Ortalama Karesel Hata (OKH), klasik yöntemde 0,50 iken, önerilen

yöntemle 0,39'ye düşerek %21'lik bir iyileşme sağlamıştır. OMYH'de %2, OMH'de ise %4 oranında hata azalması görülmüştür.

Tablo 4.7'daki test verisi tahminleri bu farkı somutlaştırmaktadır. Örneğin, Nisan 2024'ün haftaları için klasik yöntem 3,83 olan gerçek değeri 2,91 gibi oldukça uzak tahminler üretirken, önerilen yöntem 3,46-3,51 aralığında çok daha isabetli tahminler yapmıştır. Bu, veri yoğunluğunu artırmanın ve çevrim içi verilerden gelen sinyalleri kullanmanın, DVR gibi modellere karmaşık kalıpları öğrenmesi için ne denli güçlü bir zemin hazırladığını kanıtlamaktadır.

Tablo 4.6: Zenginleştirilmiş Veriler ile DVR Hata İyileşmeleri

	Ortalama Karesel Hata	Ortalama Mutlak Yüzde Hata	Ortalama Mutlak Hata
Klasik Yöntem	0,504	0,151	0,578
Önerilen Yöntem	0,399	0,147	0,555
İyileşme	%21	%2	%4

Tablo 4.7: Zenginleştirilmiş Veriler ile DVR Sonuçları

Tarih	Toplam Satış	Yeni Talep Tahmini	Klasik Talep Tahmini
04/02/2024	4,04	3,21	2,88
11/02/2024	4,04	3,21	2,88
18/02/2024	4,04	3,33	2,88
25/02/2024	3,93	3,26	2,88
03/03/2024	3,64	3,60	3,82
10/03/2024	3,64	3,54	3,82
17/03/2024	3,64	3,59	3,82
24/03/2024	3,64	3,46	3,82
31/03/2024	3,83	3,55	3,82
07/04/2024	3,86	3,47	2,92
14/04/2024	3,86	3,45	2,92
21/04/2024	3,86	3,52	2,92
28/04/2024	3,76	3,48	2,92
05/05/2024	3,68	2,91	3,80
12/05/2024	3,68	3,02	3,80
19/05/2024	3,68	2,88	3,80
26/05/2024	3,67	2,84	3,80
02/06/2024	3,60	2,77	2,90
09/06/2024	3,60	2,63	2,90
16/06/2024	3,60	2,77	2,90
23/06/2024	3,60	2,77	2,90

4.3. GRADYAN HIZLANDIRMA REGRESYON ANALİZİ

GHR algoritması ile yapılan analizler de iki aşamada gerçekleştirilmiştir. İlk deneyde, 20 aylık verinin tamamının hem eğitim hem de test için kullanılması, modelin veriyi ezberlemesine yol açmıştır. Tablo 4.8'de görüldüğü üzere, tahminler gerçek değerlerle neredeyse birebir aynıdır.

Tablo 4.9'daki hata metrikleri de (örneğin, 0,0000115 OKH) bu durumu teyit etmektedir. Bu kadar düşük hata oranları, modelin performansını göstermek yerine metodolojik bir hataya işaret ettiğinden, bu deneyin sonuçları geçersiz kabul edilmiştir.

Tablo 4.8: Aylık Veriler ile GHR Sonuçları

Tarih	Toplam Satış	Yeni Talep Tahmini	Klasik Talep Tahmini
Kasım 22	14,73	14,73	14,73
Aralık 22	13,94	13,94	13,94
Ocak 23	15,67	15,66	15,66
Şubat 23	15,43	15,43	15,43
Mart 23	15,67	15,67	15,68
Nisan 23	16,45	16,44	16,44
Mayıs 23	15,95	15,95	15,95
Haziran 23	16,43	16,43	16,42
Temmuz 23	16,30	16,30	16,30
Ağustos 23	15,91	15,91	15,91
Eylül 23	16,19	16,19	16,19
Ekim 23	15,81	15,81	15,81
Kasım 23	15,94	15,94	15,94
Aralık 23	16,39	16,38	16,38
Ocak 24	15,54	15,54	15,54
Şubat 24	16,18	16,18	16,18
Mart 24	16,14	16,14	16,13
Nisan 24	16,53	16,53	16,53
Mayıs 24	16,31	16,31	16,31
Haziran 24	15,43	15,43	15,44

Tablo 4.9: Aylık Veriler ile GHR Hata İyileşmeleri

	Ortalama Karesel Hata	Ortalama Mutlak Yüzde Hata	Ortalama Mutlak Hata
Klasik Yöntem	0,0000115	0,00017	0,0027
Önerilen Yöntem	0,0000037	0,00011	0,0017
İyileşme	%68	%36	%36

İkinci ve metodolojik olarak doğru olan deneyde, GHR algoritması da zenginleştirilmiş 84 haftalık veri seti ve ayrıştırılmış test verileri ile çalıştırılmıştır. Tablo 4.10'de gösterildiği üzere, bu deneyde de önerilen yöntem bir avantaj sağlamıştır. Ortalama Karesel Hata (OKH), 0,24'ten 0,23'e düşerek %4'lük bir iyileşme göstermiştir. OMYH ve OMH metriklerinde ise %11'lik bir iyileşme kaydedilmiştir.

Tablo 4.11'daki tahminler, bu iyileşmenin küçük de olsa tutarlı olduğunu göstermektedir. GHR ile elde edilen kazanım DVR kadar yüksek olmasa da önerilen yöntemin farklı ve gelişmiş bir algoritma üzerinde bile pozitif etki yarattığını doğrulaması açısından önemlidir.

Tablo 4.10: Zenginleştirilmiş Veriler ile GHR Hata İyileşmeleri

	Ortalama Karesel Hata	Ortalama Mutlak Yüzde Hata	Ortalama Mutlak Hata
Klasik Yöntem	0,039	0,046	0,169
Önerilen Yöntem	0,032	0,043	0,158
İyileşme	%18	%7	%6

Tablo 4.11: Zenginleştirilmiş Veriler ile GHR Sonuçları

Tarih	Toplam Satış	Yeni Talep Tahmini	Klasik Talep Tahmini
04/02/2024	4,04	4,04	4,04
11/02/2024	4,04	3,99	4,04
18/02/2024	4,04	4,00	4,04
25/02/2024	3,93	4,01	4,04
03/03/2024	3,64	3,86	3,82
10/03/2024	3,64	3,82	3,82
17/03/2024	3,64	3,85	3,82
24/03/2024	3,64	3,84	3,82
31/03/2024	3,83	3,85	3,82
07/04/2024	3,86	4,04	4,05
14/04/2024	3,86	4,02	4,05
21/04/2024	3,86	4,03	4,05
28/04/2024	3,76	4,02	4,05
05/05/2024	3,68	3,79	3,83
12/05/2024	3,68	3,79	3,83
19/05/2024	3,68	3,80	3,83
26/05/2024	3,67	3,79	3,83
02/06/2024	3,60	3,90	3,92
09/06/2024	3,60	3,85	3,92
16/06/2024	3,60	3,86	3,92
23/06/2024	3,60	3,86	3,92

4.4. GENEL DEĞERLENDİRME

Bu çalışma, çevrim içi verilerden Büyük Dil Modelleri aracılığıyla türetilen yeni özelliklerin, talep tahmin modellerine entegre edilmesinin kayda değer faydalar sağladığını farklı algoritmalar ve metodolojiler üzerinden net bir şekilde ortaya koymuştur.

Analizler sonucunda, önerilen yöntemin OKH metriğini doğrusal regresyonda %27, gradyan hızlandırmada %4 ve en etkileyici şekilde destek vektör regresyonunda %60 oranında iyileştirdiği görülmüştür. Tablo 4.6'da belgelenen bu %60'lık rekor iyileşme, veri zenginleştirme stratejisinin ve çevrim içi verilerin niteliğinin, doğru makine öğrenmesi modeliyle birleştiğinde tahmin gücünü ne denli artırabileceğinin altını çizmektedir. Sonuç olarak, bu çalışma, modern tahminleme yaklaşımlarının başarısı için yalnızca geleneksel veri

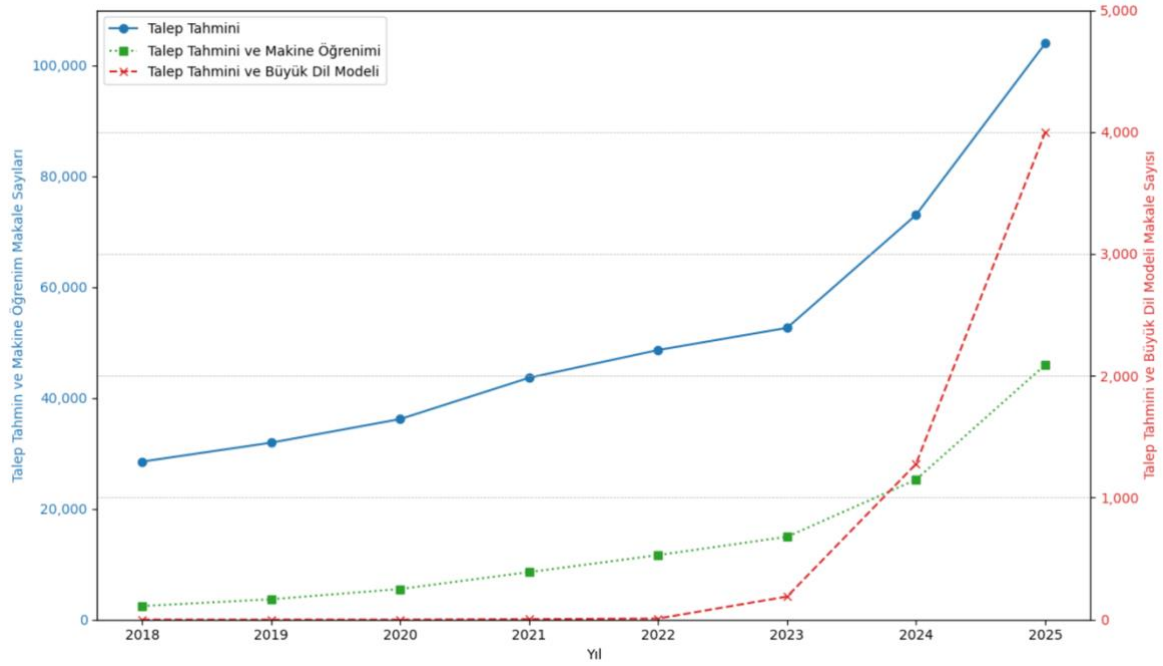
kaynaklarına baēlı kalmayıp, internetin dinamik ve zengin bilgi evreninden de faydalanmasının kritik bir gereklilik olduēunu g     kanıtlarla ortaya koymuŗtur.



5. TARTIŞMA

Endüstri mühendisliğinin temel disiplinlerinden biri olan talep tahmini, işletmelerin sürdürülebilirliği ve verimliliği açısından önemli bir rol oynar. Kaynak planlamasından envanter yönetimine, üretim çizelgelemeden stratejik karar almaya kadar birçok iş sürecinin başarısı, doğru tahminler yapabilmeye bağlıdır. Geleneksel olarak istatistiksel yöntemlerle ele alınan bu alan, özellikle son on yılda yaşanan teknolojik gelişmelerle büyük bir dönüşüm geçirmiştir. Bilgisayarların işlem gücünün hızla artması ve dijitalleşmenin etkisiyle ortaya çıkan devasa veri kümeleri, talep tahmini araştırmalarının odak noktasını makine öğrenimi tekniklerine kaydırmıştır. Artık tahminleme süreçleri, sadece geçmiş verilere dayalı matematiksel modellerle değil, aynı zamanda daha karmaşık ve uyarlanabilir algoritmalarla gerçekleştirilmektedir.

Makine öğreniminin talep tahmininde yaygın bir şekilde kullanılması 2010'lu yıllarda belirginleşmiş ve kısa sürede sektörde standart hale gelmiştir. Akademik yayın veritabanları incelendiğinde, bu eğilimin ne denli güçlü olduğu görülmektedir. Örneğin, ScienceDirect üzerinde "Talep Tahmini" (Demand Forecasting) ve "Makine Öğrenimi" (Machine Learning) anahtar kelimeleriyle yapılan aramalar sonucunda elde edilen makale sayısının 80.000'i aşması, bu iki alanın birbirine nasıl entegre olduğunu gösteren önemli bir kanıttır. Geleneksel yöntemlerin sınırlamalarının makine öğrenimi teknikleri ile aşılması, tahminleme modellerinin doğruluk oranlarını önemli ölçüde artırmış ve işletmelere daha güçlü karar destek mekanizmaları sunmuştur.



Şekil 5.1: ScienceDirect Makine Öğrenimi ve Büyük Dil Makale Sayıları

Bu evrim sürecinin en güncel ve dikkat çekici aşaması ise Büyük Dil Modelleri (BDM'ler) olmuştur. Şekil 5.1'de görüldüğü gibi, henüz yeni bir araştırma alanı olmasına rağmen, BDM'lerin talep tahmini problemlerine uygulanması konusunda akademik ilgide olağanüstü bir artış yaşanmaktadır. 2022'de "Talep Tahmini" ve "Büyük Dil Modeli" (Large Language Model) başlıklı akademik yayınların sayısı neredeyse yok denecek kadar azken, 2025 yılına yönelik projeksiyonlarda bu sayının 4.000'i aşacağı öngörülmektedir. Bu büyüme, BDM'lerin yapılandırılmamış metin verilerini (haberler, sosyal medya yorumları, ekonomik raporlar vb.) anlama ve yorumlama yeteneklerinin keşfedilmesiyle doğrudan ilişkilidir. Geleneksel talep tahminleme yaklaşımları sadece sayısal verilere odaklanırken, BDM'ler metinsel verileri de kapsamlı bir şekilde analiz edebilmekte ve tahminleme sürecinde daha zengin bir bilgi sunabilmektedir.

Artan veri hacmi, tahminleme araştırmalarına büyük olanaklar sunarken aynı zamanda önemli zorlukları da beraberinde getirmiştir. Veri setlerinin büyük bir kısmı yapısal olmayan, gürültülü (rastgele) ve düzensiz olduğu için, bu verilerden anlamlı bilginin ayrıştırılması tahminleme modellerinin doğruluğu açısından kritik bir gereklilik haline gelmiştir. Literatürde bu sorunları aşmak adına geliştirilen çeşitli yenilikçi yaklaşımlar bulunmaktadır. Kosinski ve diğ. (2016), sosyal medya platformlarında kullanıcıların gizli davranış desenlerini belirlemek için makine öğrenimi tekniklerini kullanarak kapsamlı analizler gerçekleştirmiştir. Benzer

şekilde, Xie ve diğ. (2020), Bitcoin piyasalarına sosyal medya verilerinin etkisini araştırarak büyük veri içerisinden anlamlı bilginin çıkarılmasını sağlayan gelişmiş yöntemler geliştirmiştir. Yeh ve Wang (2023), elektrikli araç satışlarını tahminlemek için çevresel faktörler, ekonomik göstergeler ve insani gelişmişlik endeksi gibi farklı veri türlerini bir araya getirerek çok boyutlu bir tahminleme modeli önermiştir. Wang (2022) ise perakende sektöründeki talep tahminleri için Rastgele Orman (Random Forest), Gradyan Artırma (Gradient Boosting) ve Uzun Kısa Süreli Bellek (UKSB) gibi güçlü makine öğrenimi algoritmalarını karşılaştırarak farklı senaryolar için etkinliklerini analiz etmiştir.

BDM'ler, geleneksel tahminleme paradigmasını kökten değiştiren bir teknoloji olarak öne çıkmaktadır. Doğal dil işleme yetenekleri sayesinde, sayısal ve yapısal verilerle zenginleştirilmiş tahminleme modelleri oluşturulabilmekte, bu da finans, sosyal bilimler ve ekonomi gibi birçok alanda etkin bir tahminleme aracı olarak kullanımını mümkün kılmaktadır. Literatürde BDM'lerin tahminleme süreçlerine entegrasyonu üzerine yapılan çalışmaların sayısının hızla arttığı gözlemlenmektedir. Geleneksel tahminleme yaklaşımlarına kıyasla daha kapsamlı ve bağlamsal analizler yapabilen BDM'ler, çok büyük veri kümelerinden anlam çıkararak daha doğru tahminler gerçekleştirebilmektedir.

BDM'lerin tahminleme sürecinde nasıl kullanıldığına dair yapılan çalışmalara bakıldığında, ham sayısal verilerin öncelikle bir "girdi metnine" (prompt) dönüştürülerek BDM tarafından anlaşılabilir hale getirildiği görülmektedir. Jin ve diğ. (2024), bu metodolojide bağlamsal bilgilerin veriyle birlikte sunulmasının modelin daha sağlıklı tahminler yapmasına olan etkisini vurgulamaktadır. Oh ve diğ. (2024), öğrenci profillerini metinsel girdilere çevirerek eğitimde tahminleme gerçekleştirmiş; Wu ve diğ. (2024) ise turizm sektöründeki talep tahmini için aynı yöntemi kullanarak geleneksel modellerle kıyaslamalı analizler yapmıştır. Seo (2024) ise haberleri kullanarak tahmin yöntemi geliştirmiş, bu yöntemde üretilen sabit konular haberler içerisinde taranıp tahmin modelinde kullanmaktadır.

Son yıllarda yapılan ileri düzey çalışmalar, tahmin doğruluğunu artırmak adına yapılandırılmamış verileri BDM analizine dahil etmektedir. Özellikle medya içerikleri, sosyal medya yorumları ve haber başlıklarından elde edilen duygu ve görüş analizleri, piyasa hissiyatı ve kamuoyu eğilimlerini anlamada büyük bir avantaj sağlamaktadır. Geleneksel yöntemler yalnızca sayısal verilere odaklanırken, BDM'ler duygu analizi gibi derin anlamsal bilgileri kullanarak tahminleme modellerinin kapsamını genişletebilmektedir.

Bu yaklaşıma dair gerçekleştirilen bazı önemli çalışmalar, BDM'lerin tahminleme süreçlerine nasıl entegre edilebileceğini göstermektedir. Chen ve diğ. (2025), Çin karbon fiyatlarını tahmin ederken medya haberlerinden elde edilen duygu analizlerini istatistiksel verilere entegre ederek tahmin gücünü artırmıştır. Liu ve diğ. (2025), hisse senedi fiyatlarını tahmin etmek için yatırımcı forumlarındaki ve sosyal medyadaki yorumları BDM'lere yorumlatarak finansal tahminleme süreçlerinde devrim niteliğinde bir yöntem geliştirmiştir.

Bu tez, literatürdeki güncel eğilimler ve araştırma bulguları ışığında şekillendirilmiş olup, makine öğreniminin talep tahmini problemlerinde nasıl kritik bir rol oynadığını bir kez daha doğrulamaktadır. Büyük veri çağına getirdiği zorluklara yönelik yenilikçi çözümler sunarak akademik alanda önemli katkılar sağlamaktadır. Özellikle yapılandırılmamış verilerin tahminleme modellerine entegrasyonu konusunda yapılan çalışmalar, tahmin doğruluğunu artırmada yeni ufuklar açmaktadır. BDM'lerin tahminleme çalışmalarında kullanımının hızla arttığı ve bu eğilimin ilerleyen yıllarda daha da genişleyeceği öngörülmektedir.

6. SONUÇ VE ÖNERİLER

Bu tez çalışması, günümüzün dijital çağında devasa boyutlara ulaşan çevrim içi verilerin, geleneksel tahminleme süreçlerine entegre edilerek bu süreçlerin doğruluğunu ve kapsamını nasıl artırabileceğini derinlemesine incelemektedir. Veri analitiği ve yapay zeka alanındaki tahminleme paradigmaları, büyük ölçüde yapısal ve sayısal formatlardaki verilere dayanmaktadır. Bu durum, sosyal medya akışları, haber makaleleri, forum tartışmaları ve müşteri yorumları gibi değerli bilgiler barındıran ancak doğası gereği yapılandırılmamış metin tabanlı verilerin potansiyelinden tam olarak yararlanılamamasına neden olmaktadır. Bu tür verilerin doğrudan makine öğrenimi algoritmaları tarafından işlenmesi, anlamsal ve bağlamsal zenginliklerinin kaybolmasına yol açacağından, ham veriyi anlamlı ve işlenebilir sayısal temsillere dönüştürecek ileri düzey dönüşüm tekniklerinin uygulanması kritik bir zorunluluktur.

Bu zorluğun üstesinden gelmek amacıyla, bu çalışma kapsamında, dönüşüm sürecinin merkezine büyük dil modellerinin (BDM) entegre edilmesi önerilmektedir. Milyarlarca parametre ve geniş veri kümeleri üzerinde eğitilmiş olan bu modeller, doğal dili anlama, yorumlama ve üretme konusunda insan benzeri yeteneklere sahiptir. Bu üstün yetenekleri sayesinde, büyük ve karmaşık metin yığınlarını süratle analiz ederek içlerindeki anlamsal nüansları, duygu tonunu, anahtar varlıkları ve temel konuları tespit edebilirler. Sonuç olarak, yapılandırılmamış metin verilerini, tahminleme modelleri için girdi olarak kullanılabilecek yüksek kaliteli, yapılandırılmış ve sayısal verilere dönüştürme kapasitesine sahiptirler. Bu süreç, sadece bir veri formatı dönüşümü değil, aynı zamanda verinin anlamsal olarak zenginleştirilmesi anlamına gelmektedir.

Büyük dil modellerinin etkinliği, büyük ölçüde onlara sunulan girdilerin kalitesine ve tasarımına bağlıdır. Bu nedenle, "girdi mühendisliği" olarak bilinen disiplin, araştırma sürecinin önemli bir parçasını oluşturmaktadır. Modellerin farklı görevler için optimize edilmesi, belirli veri türlerinden istenen bilgiyi en doğru şekilde çıkarması ve sonuçların tutarlılığının sağlanması, dikkatle tasarlanmış girdiler aracılığıyla mümkün olmaktadır. Bu tez, daha isabetli ve güvenilir tahminleme sonuçları elde etmek için sistematik girdi oluşturma stratejileri

geliştirmeyi ve bu stratejilerin zamanla otomatize edilerek veri işleme otomasyonları daha verimli ve ölçeklenebilir hale getirmeyi hedeflemektedir.

Bu metodolojik çerçeveyi test etmek amacıyla, tez çalışmasında OpenAI tarafından geliştirilen ve çok modlu yetenekleriyle öne çıkan GPT-4o modeli kullanılmıştır. GPT-4o'nun metin, görüntü ve ses gibi farklı veri türlerini aynı anda işleyebilme kapasitesi, onu karmaşık ve çeşitli veri setlerinden anlamlı öngörüler üretme konusunda güçlü bir araç haline getirmektedir. Son yıllarda teknoloji devleri tarafından geliştirilen ve her biri farklı mimari ve eğitim verisiyle özelleşmiş çok sayıda büyük dil modelinin varlığı, bu alanda yeni olanaklar sunmaktadır. Gelecekte, farklı modellerin güçlü yönlerini birleştiren hibrit veya gruplaşmış (birden fazla BDM'nin kullanılması) yaklaşımlar geliştirilerek, tek bir modelin zayıf yönlerinin telafi edilmesi ve genel tahminleme performansının daha da ileri bir seviyeye taşınması mümkündür.

Önerilen metodoloji, haber makaleleri gibi farklı kaynaklardan elde edilen çevrim içi verilerin tek tek işlenerek genel bir sayısal çıktıya dönüştürülmesini içermektedir. Ancak haber verilerinin doğası, önemli bir zorluğu da beraberinde getirmektedir: aynı olayın farklı kaynaklar tarafından benzer içeriklerle tekrar tekrar yayınlanması. Bu veri tekrarı, basit bir ortalama alma işlemiyle elde edilecek sayısal çıktının, belirli bir habere orantısız ağırlık vermesine ve dolayısıyla sonuçların yanlı (biased) olmasına neden olabilir. Bu sapmayı engellemek amacıyla, aynı zaman diliminde analiz edilen haberler arasında anlamsal benzerlik hesaplayan ve benzer haberlerin ağırlığını dinamik olarak ayarlayan bir model önerilmektedir. Bu model, her bir bilginin genel skora adil bir katkı yapmasını sağlayarak sonucun güvenilirliğini artıracaktır.

Tezin doğrulama aşamasında, büyük dil modeli tarafından üretilen yapılandırılmış sayısal verilerin analizi için üç farklı makine öğrenimi algoritması kullanılmıştır. Ancak bu, keşfedilecek potansiyelin yalnızca bir başlangıcıdır. Gelecekteki çalışmalarda, daha gelişmiş zaman serisi modelleri, derin öğrenme ağları ve özellikle model parametrelerini optimize etmede başarılı olan genetik algoritmalar veya parçacık sürüsü optimizasyonu gibi meta-sezgisel yöntemler kullanılarak daha sofistike ve yüksek performanslı tahminleme modelleri geliştirilebilir.

Sonuç olarak, bu araştırma, büyük dil modellerinin yalnızca birer doğal dil işleme aracı olmanın ötesinde, veri biliminin temelini oluşturan tahminleme süreçlerinde bir paradigma değişimi yaratma potansiyeline sahip olduğunu göstermektedir. Farklı modellerin ve veri

kaynaklarının entegre edildiđi, akıllı otomasyon ve girdi mühendisliđi teknikleriyle desteklendiđi bu yenilikçi yaklaşımlar, geleceđin tahminleme sistemlerinin daha isabetli, kapsamlı ve güvenilir olmasının önünü açacaktır. Bu tez, büyük dil modellerinin tahminleme alanındaki etkin kullanımına dair somut bulgular sunarak, bu hızla gelişen alandaki akademik ve pratik çalışmalara önemli bir katkı sağlamaktadır.



KAYNAKLAR

- Abolghasemi, M., Ganbold, O. and Rotaru, K., 2024, Humans vs. large language models: Judgmental forecasting in an era of advanced AI, *International Journal of Forecasting*.
- Alpaydin, E., 2014, *Introduction to Machine Learning*, The Mit Press, Cambridge, Massachusetts.
- Bitran, G. and Caldentey, R., 2003, An Overview of Pricing Models for Revenue Management, *Manufacturing & Service Operations Management*, 5(3), 203–229.
- Blattberg, R.C. and Hoch, S.J., 1990, Database Models and Managerial Intuition: 50% Model + 50% Manager, *Management Science*, 36(8), 887–899.
- Boden, M.A., 2018, *Artificial Intelligence: A Very Short Introduction*, Oxford University Press.
- Cadenas, E., Jaramillo, O.A. and Rivera, W., 2010, Analysis and Forecasting of Wind Velocity in chetumal, Quintana roo, Using the Single Exponential Smoothing Method, *Renewable Energy*, 35(5), 925–930.
- Chaboud, A.P., Chiquoine, B., Hjalmarsson, E. and Vega, C., 2014, Rise of the Machines: Algorithmic Trading in the Foreign Exchange Market, *The Journal of Finance*, 69(5), 2045–2084.
- Chang, P.-T., Huang, L.-C. and Lin, H.-J., 2000, The Fuzzy Delphi Method via Fuzzy Statistics and Membership Function Fitting and an Application to the Human Resources, *Fuzzy Sets and Systems*, 112(3), 511–520.
- Chen, R., Jiang, H., Guo, T. and Fan, C., 2025, Can Large Language Models Forecast Carbon Price movements? Evidence from Chinese Carbon Markets, *Research in International Business and Finance*, 77, 102951.

- Dash, R. and Dash, P.K., 2016, A hybrid stock trading framework integrating technical analysis with machine learning techniques, *The Journal of Finance and Data Science*, 2(1), 42–57.
- Deng, T., Zhao, Y., Wang, S. and Yu, H., 2021, *Sales Forecasting Based on LightGBM*, *IEEE Xplore*, 383–386.
- Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K., 2019, *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*, in J. Burstein, C. Doran & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186, Association for Computational Linguistics.
- Diamantopoulos, A. and Winklhofer, H., 2003, Export Sales Forecasting by UK Firms, *Journal of Business Research*, 56(1), 45–54.
- Du, J., Xu, H. and Huang, X., 2014, Box Office Prediction Based on Microblog, *Expert Systems with Applications*, 41(4), 1680–1689.
- Einhorn, H.J., 1974, Cue Definition and Residual Judgment, *Organizational Behavior and Human Performance*, 12(1), 30–49.
- Elton, M.C.J., 1991, *Integrated Broadband Networks : the Public Policy Issues*, North-Holland, Amsterdam.
- Emmela, S., Rama Lahari, V., Anusha, B., Bhavana, D., Yasyukevich, Y.V., Demyanov, V.V. and Venkata Ratnam, D., 2024, Global Ionospheric Total Electron Content short-term Forecast Based on Light Gradient Boosting Machine, Extreme Gradient Boosting, and Gradient Boost Regression, *Advances in Space Research*, 74(8), 3642–3653.
- Entezari, A., Aslani, A., Zahedi, R. and Noorollahi, Y., 2023, Artificial Intelligence and Machine Learning in Energy systems: a Bibliographic Perspective, *Energy Strategy Reviews*, 45(101017), 101017.
- European Commission, 2020, *White Paper on Artificial Intelligence: a European approach to excellence and trust*, commission.europa.eu.

- Ferrer, J.-C., Oyarzún, D. and Vera, J., 2012, Risk Averse Retail Pricing with Robust Demand Forecasting, *International Journal of Production Economics*, 136(1), 151–160.
- Franc, J., Hung, K., Pirisi, A. and Weinstein, E., 2023, Analysis of Delphi Study Seven-Point Linear Scale Data by Parametric Methods—Use of the Mean and Standard Deviation, *Prehospital and Disaster Medicine*, 38(S1), s30–s31.
- Friedman, J.H., 2001, Greedy Function Approximation: a Gradient Boosting Machine, *The Annals of Statistics*, 29(5), 1189–1232.
- Friedman, J.H., 2002, Stochastic Gradient Boosting, *Computational Statistics & Data Analysis*, 38(4), 367–378.
- Galton, F., 1886, Regression Towards Mediocrity in Hereditary Stature., *The Journal of the Anthropological Institute of Great Britain and Ireland*, 15, 246–263.
- Ghosh, I., Chaudhuri, T.D., Alfaro-Cortés, E., Gámez, M. and García, N., 2022, A Hybrid Approach to Forecasting Futures Prices with Simultaneous Consideration of Optimality in Ensemble Feature Selection and Advanced Artificial Intelligence, *Technological Forecasting and Social Change*, 181, 121757.
- Hazelrigg, G.A., 2009, *The Mathematics of Prediction*, in S. Raghavan & J.A. Cafeo (Eds.), *Product Research*, 93–111, Springer Netherlands, Dordrecht.
- Homolka, L., Ngo, V.M., Pavelková, D., Le, B.T. and Dehning, B., 2020, Short- and medium-term Car Registration Forecasting Based on Selected Macro and socio-economic Indicators in European Countries, *Research in Transportation Economics*, 80, 100752.
- Jin, M., Wang, S., Ma, L., Chu, Z., Zhang, J.Y., Shi, X., Chen, P.-Y., Liang, Y., Li, Y.-F., Pan, S. and Wen, Q., 2024, *Time-LLM: Time Series Forecasting by Reprogramming Large Language Models*.
- Jones, S., Johnstone, D. and Wilson, R., 2015, An Empirical Evaluation of the Performance of Binary Classifiers in the Prediction of Credit Ratings Changes, *Journal of Banking & Finance*, 56, 72–85.

- Kaplan, A.M. and Haenlein, M., 2010, Users of the world, unite! the Challenges and Opportunities of Social Media, *Business Horizons*, 53(1), 59–68.
- Kaynak, O., 2021, The Golden Age of Artificial Intelligence, *Discover Artificial Intelligence*, 1(1).
- Kiran, D.R., 2022, *Principles of Economics and Management for Manufacturing Engineering*, Butterworth-Heinemann.
- Kosinski, M., Wang, Y., Lakkaraju, H. and Leskovec, J., 2016, Mining Big Data to Extract Patterns and Predict real-life outcomes., *Psychological Methods*, 21(4), 493–506.
- Lawrence, M., Goodwin, P., O'Connor, M. and Önkal, D., 2006, Judgmental forecasting: a Review of Progress over the Last 25years, *International Journal of Forecasting*, 22(3), 493–518.
- Li, X., Wu, P. and Wang, W., 2020, Incorporating Stock Prices and News Sentiments for Stock Market prediction: a Case of Hong Kong, *Information Processing & Management*, 57(5), 102212.
- Lin, L., Jiang, W., Chen, B., Yu, J. and Zheng, C., 2024, Construction and Application of Cost Prediction Model Based on Multiple Linear Regression Analysis, *Procedia Computer Science*, 247, 617–623.
- Liu, Y., Bu, N., Li, Z., Zhang, Y. and Zhao, Z., 2025, AT-FinGPT: Financial Risk Prediction via an audio-text Large Language Model, *Finance Research Letters*, 77, 106967.
- McCarthy, J., Minsky, M.L., Rochester, N. and Shannon, C.E., 1955, A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955, *AI Magazine*, 27(4), 12–12.
- Meehl, P.E., 1957, When Shall We Use Our Heads Instead of the formula?, *Journal of Counseling Psychology*, 4(4), 268–273.
- Minsky, M.L. and Papert, S., 1990, *Perceptrons an Introduction to Computational Geometry*, Cambridge, Mass. Massachusetts Institute of Technology.

- Mintarya, L.N., Halim, J.N.M., Angie, C., Achmad, S. and Kurniawan, A., 2023, Machine Learning Approaches in Stock Market prediction: a Systematic Literature Review, *Procedia Computer Science*, 216, 96–102.
- Mital, A., Desai, A., Subramanian, A. and Mital, A., 2008, *Product Development : a Structured Approach to Consumer Product development, design, and Manufacture*, Butterworth-Heinemann, Oxford.
- Nikolopoulos, K., 2020, We Need to Talk about Intermittent Demand Forecasting, *European Journal of Operational Research*.
- O'Hara, M., 2015, High frequency market microstructure, *Journal of Financial Economics*, 116(2), 257–270.
- Oh, C., Park, M., Lim, S. and Song, K., 2024, Language model-guided Student Performance Prediction with Multimodal Auxiliary Information, *Expert Systems with Applications*, 250, 123960–123960.
- OpenAI, 2023, GPT-4 Technical Report, *arXiv (Cornell University)*.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, É., 2011, Scikit-learn: Machine learning in python, *Journal of Machine Learning Research*, 12, 2825–2830.
- Sa-ngasoongsong, A., Bukkapatnam, S.T.S., Kim, J., Iyer, P.S. and Suresh, R.P., 2012, Multi-step Sales Forecasting in Automotive Industry Based on Structural Relationship Identification, *International Journal of Production Economics*, 140(2), 875–887.
- Saha, P., Gudheniya, N., Mitra, R., Das, D., Narayana, S. and Tiwari, M.K., 2022, Demand Forecasting of a Multinational Retail Company Using Deep Learning Frameworks, *IFAC-PapersOnLine*, 55(10), 395–399.
- Sánchez, J.C.M., Mesa, H.G.A., Espinosa, A.T., Castilla, S.R. and Lamont, F.G., 2025, Improving Wheat Yield Prediction through Variable Selection Using Support Vector Regression, Random Forest, and Extreme Gradient Boosting, *Smart Agricultural Technology*, 10, 100791.

- Schoenegger, P. and Park, P.S., 2023, *Large language model prediction capabilities: Evidence from a real-world forecasting tournament*.
- Seo, B., 2024, Econometric Forecasting Using Ubiquitous News text: Text-enhanced Factor Model, *International Journal of Forecasting*.
- Sheikh, H., Prins, C. and Schrijvers, E., 2023, *Artificial Intelligence: Definition and Background*, Springer, Cham.
- Shi, X., Xue, S., Wang, K., Zhou, F., Zhang, J., Zhou, J., Tan, C. and Mei, H., 2024, Language models can improve event prediction by few-shot abductive reasoning, *Advances in Neural Information Processing Systems*, 36.
- Siegel, A.F., 2012, *Practical Business Statistics*, 6th edn., Academic Press.
- Singh, S.K., Tiwari, A.K. and Paliwal, H.K., 2023, A Stateoftheart Review on the Utilization of Machine Learning in nanofluids, Solar Energy generation, and the Prognosis of Solar Power, *Engineering Analysis with Boundary Elements*, 155, 62–86.
- Smith, S.A., McIntyre, S.H. and Achabal, D.D., 1994, A Two-Stage Sales Forecasting Procedure Using Discounted Least Squares, *Journal of Marketing Research*, 31(1), 44.
- Smola, A.J. and Schölkopf, B., 2004, A Tutorial on Support Vector Regression, *Statistics and Computing*, 14(3), 199–222.
- Soares, S.G., 2015, *Ensemble learning methodologies for soft sensor development in industrial processes* – PhD thesis, University of Coimbra .
- Vapnik, V., Golowich, S.E. and Smola, A., 1996, *Support Vector Method for Function approximation, Regression Estimation and Signal Processing*, 281–287, MIT Press, Denver, Colorado.
- Vapnik, V.N., 1995, *The nature of statistical learning theory*, Springer-Verlag.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017, Attention Is All You Need, *Advances in Neural Information Processing Systems*, 30.

- Wang, C.H., 2022, Considering Economic Indicators and Dynamic Channel Interactions to Conduct Sales Forecasting and for Retail Sectors, *Computers & Industrial Engineering*, 165, 107965.
- Weizenbaum, J., 1966, ELIZA - a Computer Program for the Study of Natural Language Communication between Man and Machine, *Communications of the ACM*, 9(1), 36–45.
- Winston, W.L., 1994, *Operations Research: Applications and Algorithms*, Duxbury Press, California.
- Wu, D.C., Li, W., Wu, J., Hu, M. and Shen, S., 2024, How Well Can ChatGPT Forecast Tourism demand?, *Tourism Management*, 108, 105119–105119.
- Xie, P., Chen, H. and Hu, Y.J., 2020, Signal or Noise in Social Media Discussions: the Role of Network Cohesion in Predicting the Bitcoin Market, *Journal of Management Information Systems*, 37(4), 933–956.
- Yan, X. and Su, X., 2009, *Linear Regression Analysis : Theory and Computing*, World Scientific, Hackensack, N.J.
- Yeh, J.-Y. and Wang, Y.-T., 2023, A Prediction Model for Electric Vehicle Sales Using Machine Learning Approaches, *Journal of Global Information Management*, 31(1), 1–21.

EKLER

EK 1. Doğrusal regresyon örnek analiz python kodu

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import linregress

# Veri noktaları
# Ebeveyn boyu (inç)
ebeveyn_boyu = np.array([64.5, 65.5, 66.5, 67.5, 68.5, 69.5, 70.5, 71.5, 72.5])
# Çocuk boyu (inç)
cocuk_boyu = np.array([65.8, 66.7, 67.2, 67.6, 68.2, 68.9, 69.5, 69.9, 72.2])

# Doğrusal regresyon analizi yapılır.
# slope: regresyon çizgisinin eğimi
# intercept: regresyon çizgisinin y-eksenini kestiği nokta
# r_value: korelasyon katsayısı
# p_value: p-değeri
# std_err: standart hata
slope, intercept, r_value, p_value, std_err = linregress(ebeveyn_boyu, cocuk_boyu)

# Regresyon çizgisinin denklemi oluşturulur:  $y = \text{eğim} * x + \text{kesişim}$ 
regresyon_cizgisi = slope * ebeveyn_boyu + intercept

# Grafiği çizmek için matplotlib kullanılır.
plt.figure(figsize=(10, 6)) # Grafiğin boyutunu ayarlar
plt.scatter(ebeveyn_boyu, cocuk_boyu, color='blue', label='Veri Noktaları') # Veri noktalarını çizer
plt.plot(ebeveyn_boyu, regresyon_cizgisi, color='red', label=f'Regresyon Çizgisi\n(y = {slope:.2f}x + {intercept:.2f})') # Regresyon çizgisini çizer

plt.title('Ebeveyn Boyu ve Çocuk Boyu Arasındaki Doğrusal Regresyon') # Grafiğin başlığı
plt.xlabel('Ebeveyn Boyu (inç)') # X eksen etiketi
plt.ylabel('Çocuk Boyu (inç)') # Y eksen etiketi
plt.grid(True) # Izgara ekler
plt.legend() # Etiketleri gösterir
plt.show() # Grafiği görüntüler

# Hesaplanan regresyon parametrelerini yazdırır
print(f'Eğim (Slope): {slope:.2f}')
print(f'Kesişim (Intercept): {intercept:.2f}')
print(f'Korelasyon Katsayısı (R-value): {r_value:.2f}')
print(f'R-kare (R-squared): {r_value**2:.2f}')
```

EK 2. GPT-4o API çağrı örneği

URL: https://<proje_isim>.openai.azure.com/openai/deployments/gpt-4o/chat/completions?api-version=2024-02-15-preview

İstek Başlıkları:

Method: POST

Content-Type: application/json

api-key: <API anahtarı>

İstek Verisi:

```
{
  "messages": [
    {
      "role": "system",
      "content": [
        {
          "type": "text",
          "text": "<komut>"
        }
      ],
      "role": "user",
      "content": [
        {
          "type": "text",
          "text": "<girdi>"
        }
      ],
      "temperature": 0.7,
      "top_p": 0.95,
      "max_tokens": 800
    }
  ]
}
```

Örnek Cevap Verisi:

```
{
  "choices": [
    {
      "content_filter_results": {
        "hate": {
          "filtered": false,
          "severity": "safe"
        },
        "protected_material_code": {
          "filtered": false,
          "detected": false
        },
        "protected_material_text": {
          "filtered": false,
          "detected": false
        },
        "self_harm": {
          "filtered": false,
          "severity": "safe"
        },
        "sexual": {
          "filtered": false,
          "severity": "safe"
        },
        "violence": {
          "filtered": false,
          "severity": "safe"
        },
        "finish_reason": "stop",
        "index": 0,
        "logprobs": null,
        "message": {
          "annotations": [],
          "content": "<sonuç örneğin: Indirect, 10%>",
          "refusal": null,
          "role": "assistant"
        },
        "created": 1749771151,
        "id": "<rand>",
        "model": "gpt-4o-2024-11-20",
        "object": "chat.completion",
        "prompt_filter_results": [
          {
            "prompt_index": 0,
            "content_filter_results": {
              "hate": {
                "filtered": false,
                "severity": "safe"
              },
              "jailbreak": {
                "filtered": false,
                "detected": false
              },
              "self_harm": {
                "filtered": false,
                "severity": "safe"
              },
              "sexual": {
                "filtered": false,
                "severity": "safe"
              },
              "violence": {
                "filtered": false,
                "severity": "safe"
              }
            },
            "system_fingerprint": "<rand>",
            "usage": {
              "completion_tokens": 6,
              "completion_tokens_details": {
                "accepted_prediction_tokens": 0,
                "audio_tokens": 0,
                "reasoning_tokens": 0,
                "rejected_prediction_tokens": 0
              },
              "prompt_tokens": 239,
              "prompt_tokens_details": {
                "audio_tokens": 0,
                "cached_tokens": 0
              },
              "total_tokens": 245
            }
          }
        ]
      }
    }
  ]
}
```

EK 3. Büyük dil modeli sonuç değerlendirme regexleri

Indirect(* *)?(\simpact)?(:|,)\s?\d{1,3}\%

Indirect(* *)?(\sImpact)?(:|,)\s?\d{1,3}\%

Indirect\s{0,3}*{0,3}Amount\s(of|s)Affect:*{0,3}\s?\d{1,3}\%

Indirect\s{0,3}*{0,3}\s\s-\s\d{1,3}\%

Indirect\sImpact\s(on|s)Vehicle\sSales:\s?\d{1,3}\%

Indirect\s(I|i)mpact\s(o|O)n\s(c|C)ar\s(s|S)ales\s\d{1,3}\%

Indirect\s(I|i)mpact\s(o|O)n\s(v|V)ehicle\s(s|S)ales\s\d{1,3}\%

Indirect\s(I|i)mpact\s\d{1,3}\%

Indirect\s(A|a)ffect\s\d{1,3}\%

Indirectly\s\d{1,3}\%

Indirect\s\d{1,3}\%

Indirect\s(a|A)ffect\s(o|O)n\s(c|C)ar\s(s|S)ales\s\d{1,3}\%

(d|D)irect\s(or|s)iIndirect\s(I|i)ndirect\s(a|A)mount\s(o|O)f\s(a|A)ffect\s\d{1,3}\%

Indirect\s(i|I)mpact\s(a|A)mount\s(o|O)f\s(a|A)ffect\s\d{1,3}\%

Indirect\s(a|A)ffect\s(o|O)n\s(v|V)ehicle\s(s|S)ales\s\d{1,3}\%

Indirect\s(e|E)ffect\s\d{1,3}\%

Indirectly\s(i|I)mpact\s\d{1,3}\%

Indirect\s(e|E)ffect\s(o|O)n\s(c|C)ar\s(s|S)ales\s\d{1,3}\%

Direct\(**)(\simpact)?(:|,)\s?\d{1,3}\%

Direct\(**)(\sImpact)?(:|,)\s?\d{1,3}\%

Direct\s{0,3}*{0,3}Amount\s of\s Affect:*{0,3}\s?\d{1,3}\%

Direct\s{0,3}*{0,3}\s-\s\d{1,3}\%

Direct\sImpact\s on\s Vehicle\s Sales:\s?\d{1,3}\%

Direct\s(I|i)mpact\s(o|O)n\s(c|C)ar\s(s|S)ales\s\d{1,3}\%

Direct\s(I|i)mpact\s(o|O)n\s(v|V)ehicle\s(s|S)ales\s\d{1,3}\%

Direct\s(I|i)mpact\s\d{1,3}\%

Direct\s(A|a)ffect\s\d{1,3}\%

Directly\s\d{1,3}\%

Direct\s\d{1,3}\%

Direct\s(a|A)ffect\s(o|O)n\s(c|C)ar\s(s|S)ales\s\d{1,3}\%

(d|D)irect\sor\s(i|I)ndirect\s(d|D)irect\s(a|A)mount\s(o|O)f\s(a|A)ffect\s\d{1,3}\%

Direct\s(i|I)mpact\s(a|A)mount\s(o|O)f\s(a|A)ffect\s\d{1,3}\%

Direct\s(a|A)ffect\s(o|O)n\s(v|V)ehicle\s(s|S)ales\s\d{1,3}\%

Direct\s(e|E)ffect\s\d{1,3}\%

Directly\s(i|I)mpact\s\d{1,3}\%

Direct\s(e|E)ffect\s(o|O)n\s(c|C)ar\s(s|S)ales\s\d{1,3}\%

EK 4. Doğrusal regresyon python kodu

```
from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_absolute_percentage_error
from sklearn import linear_model
```

```
import pandas as pd
import numpy as np
```

```
df = pd.read_csv('average_scores.csv')
```

```
cols = [1,2,3,4,5,6,7,8]
all_X = df[df.columns[cols]]
all_Y = df[df.columns[9:10]]
# eğitim ve test verilerinin ayrılması
test_X = all_X.loc[15:]
test_Y = all_Y.loc[15:]
input_X = all_X.loc[:15]
input_Y = all_Y.loc[:15]
```

```
X = np.array(input_X)
y = np.array(input_Y).squeeze()
test_X = np.array(test_X)
test_Y = np.array(test_Y).squeeze()
```

```
model = linear_model.LinearRegression()
```

```

model.fit(X, y)
model_new = model

y_pred1 = model.predict(test_X)

cols = [3,4,5,6,7,8]
all_X = df[df.columns[cols]]
all_Y = df[df.columns[9:10]]
test_X = all_X.loc[15:]
input_X = all_X.loc[:15]
input_Y = all_Y.loc[:15]

X = np.array(input_X)
y = np.array(input_Y).squeeze()
test_X = np.array(test_X)

# Train SVM model
model = linear_model.LinearRegression()
model.fit(X, y)
model_old = model

# Predict using the model
y_pred2 = model.predict(test_X)

mse_1 = mean_squared_error(test_Y, y_pred1)
mse_2 = mean_squared_error(test_Y, y_pred2)

print(f"Yeni Yöntem Ortalama Karesel Hata {mse_1}")
print(f"Klasik Yöntem Ortalama Karesel Hata {mse_2}")

```

EK 5. Destek vektör regresyon algoritma python kodu

```

import numpy as np
import pandas as pd
from sklearn.svm import SVR
from IPython.display import display
from sklearn.utils.validation import check_consistent_length, check_array
from sklearn.metrics import mean_squared_error

df = pd.read_csv('average_scores.csv')
cols = [1,2,3,4,5,6,7,8]
input_X = df[df.columns[cols]] # 1. ve 2. colon, yeni parametreleri içermektedir
input_Y = df[df.columns[9:10]]

X = np.array(input_X)
y = np.array(input_Y).squeeze()

# Train SVM model
model = SVR(kernel="linear")

```

```

model.fit(X, y)

# Predict using the model
y_predictions_with_new_parameters = model.predict(X)

cols = [3,4,5,6,7,8]
input_X = df[df.columns[cols]] # 1. ve 2. colon dışındaki değerler
input_Y = df[df.columns[9:10]]

X = np.array(input_X)
y = np.array(input_Y).squeeze()

model = SVR(kernel=model_name)
model.fit(X, y)

y_predictions_without_new_params = model.predict(X)

mse_with_new_parameters = mean_squared_error(y, y_predictions_with_new_parameters)
mse_without_new_parameters = mean_squared_error(y, y_predictions_without_new_params)

print(f"Yeni Tahmin Ortalama Karesel Hata = {mse_with_new_parameters}")
print(f"Klasik Tahmin Ortalama Karesel Hata = {mse_without_new_parameters}")

```

EK 6. Gradyan hızlandırma regresyon algoritma python kodu

```

from sklearn.ensemble import GradientBoostingRegressor
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_absolute_percentage_error
import pandas as pd
import numpy as np

df = pd.read_csv('average_scores.csv')
cols = [1,2,3,4,5,6,7,8]
input_X = df[df.columns[cols]] # 1. ve 2. colon, yeni parametreleri içermektedir
input_Y = df[df.columns[9:10]]

X = np.array(input_X)
y = np.array(input_Y).squeeze()

gbm = GradientBoostingRegressor(n_estimators=100, learning_rate=0.01, random_state=42)

# Train the model
gbm.fit(X, y)

# Predict using the model
y_predictions_with_new_parameters = gbm.predict(X)

```

```

cols = [3,4,5,6,7,8]
input_X = df[df.columns[cols]] # 1. ve 2. colon dışındaki değerler
input_Y = df[df.columns[9:10]]

X = np.array(input_X)
y = np.array(input_Y).squeeze()

gbm = GradientBoostingRegressor(n_estimators=100, learning_rate=0.01, random_state=42)

# Train the model
gbm.fit(X, y)

# Predict using the model
y_predictions_without_new_params = gbm.predict(X)

mse_with_new_parameters = mean_squared_error(y, y_predictions_with_new_parameters)
mse_without_new_parameters = mean_squared_error(y, y_predictions_without_new_params)

print(f"Yeni Tahmin Ortalama Karesel Hata = {mse_with_new_parameters}")
print(f"Klasik Tahmin Ortalama Karesel Hata = {mse_without_new_parameters}")

```

EK 7: Haftalık zenginleştirilmiş veri seti

Tarih	Doğrudan Etki	Dolaylı Etki	İşsizlik Endeksi	Akaryakıt Endeksi	Tüketici Fiyat Endeksi	Bitmiş Yeni Ev Endeksi	Bireysel Kredi Endeksi	Araç Kredisi Endeksi	Toplam Satış
20/11/2022	46,25	13,87	3,60	3,85	6,15	1559	11,21	6,94	3,44
27/11/2022	34,00	13,81	3,60	3,85	6,15	1559	11,21	6,94	3,31
04/12/2022	45,00	15,74	3,50	3,36	5,52	1386	11,21	6,94	3,15
11/12/2022	30,28	16,50	3,50	3,36	5,52	1386	11,21	6,94	3,15
18/12/2022	49,04	17,23	3,50	3,36	5,52	1386	11,21	6,94	3,15
25/12/2022	46,50	17,05	3,50	3,36	5,52	1386	11,21	6,94	3,15
01/01/2023	56,60	15,54	3,50	3,45	6,73	1389	11,21	6,94	3,54
08/01/2023	55,00	14,26	3,50	3,45	6,73	1389	11,21	6,94	3,54
15/01/2023	51,43	14,13	3,50	3,45	6,73	1389	11,21	6,94	3,54
22/01/2023	60,17	14,37	3,50	3,45	6,73	1389	11,21	6,94	3,54
29/01/2023	48,81	15,22	3,50	3,45	6,73	1389	11,21	6,94	3,72
05/02/2023	48,81	14,64	3,60	3,51	7,16	1540	11,48	7,46	3,86
12/02/2023	48,81	14,64	3,60	3,51	7,16	1540	11,48	7,46	3,86
19/02/2023	48,81	14,64	3,60	3,51	7,16	1540	11,48	7,46	3,86
26/02/2023	48,81	14,64	3,60	3,51	7,16	1540	11,48	7,46	3,68
05/03/2023	48,81	14,64	3,50	3,55	4,70	1516	11,48	7,46	3,54
12/03/2023	48,81	14,64	3,50	3,55	4,70	1516	11,48	7,46	3,54
19/03/2023	42,00	13,57	3,50	3,55	4,70	1516	11,48	7,46	3,54
26/03/2023	42,14	13,08	3,50	3,55	4,70	1516	11,48	7,46	3,58
02/04/2023	62,14	12,26	3,40	3,74	5,67	1416	11,48	7,46	3,84
09/04/2023	38,33	12,64	3,40	3,74	5,67	1416	11,48	7,46	3,84
16/04/2023	38,33	13,17	3,40	3,74	5,67	1416	11,48	7,46	3,84
23/04/2023	53,75	12,97	3,40	3,74	5,67	1416	11,48	7,46	3,84

30/04/2023	48,46	11,50	3,40	3,74	5,67	1416	11,48	7,46	3,64
07/05/2023	35,43	11,68	3,60	3,69	4,81	1499	11,48	7,59	3,60
14/05/2023	40,00	12,78	3,60	3,69	4,81	1499	11,48	7,59	3,60
21/05/2023	50,86	13,63	3,60	3,69	4,81	1499	11,48	7,59	3,60
28/05/2023	48,75	13,27	3,60	3,69	4,81	1499	11,48	7,59	3,70
04/06/2023	47,50	12,61	3,60	3,71	4,70	1480	11,48	7,59	3,83
11/06/2023	33,28	13,85	3,60	3,71	4,70	1480	11,48	7,59	3,83
18/06/2023	42,08	15,04	3,60	3,71	4,70	1480	11,48	7,59	3,83
25/06/2023	55,00	13,87	3,60	3,71	4,70	1480	11,48	7,59	3,81
02/07/2023	55,00	13,87	3,50	3,73	3,23	1343	11,48	7,59	3,68
09/07/2023	55,00	10,19	3,50	3,73	3,23	1343	11,48	7,59	3,68
16/07/2023	45,33	13,13	3,50	3,73	3,23	1343	11,48	7,59	3,68
23/07/2023	42,00	13,40	3,50	3,73	3,23	1343	11,48	7,59	3,68
30/07/2023	61,25	12,52	3,50	3,73	3,23	1343	11,48	7,59	3,62
06/08/2023	52,50	13,12	3,70	3,96	4,21	1373	12,17	8,30	3,59
13/08/2023	80,00	13,53	3,70	3,96	4,21	1373	12,17	8,30	3,59
20/08/2023	50,00	14,05	3,70	3,96	4,21	1373	12,17	8,30	3,59
27/08/2023	35,83	13,23	3,70	3,96	4,21	1373	12,17	8,30	3,65
03/09/2023	44,00	12,44	3,80	3,99	4,84	1466	12,17	8,30	3,78
10/09/2023	40,83	13,26	3,80	3,99	4,84	1466	12,17	8,30	3,78
17/09/2023	48,75	13,12	3,80	3,99	4,84	1466	12,17	8,30	3,78
24/09/2023	35,00	13,60	3,80	3,99	4,84	1466	12,17	8,30	3,78
01/10/2023	67,61	14,06	3,90	3,78	4,17	1382	12,17	8,30	3,57
08/10/2023	67,61	14,06	3,90	3,78	4,17	1382	12,17	8,30	3,57
15/10/2023	67,61	14,06	3,90	3,78	4,17	1382	12,17	8,30	3,57
22/10/2023	67,61	14,06	3,90	3,78	4,17	1382	12,17	8,30	3,57
29/10/2023	30,00	13,40	3,90	3,78	4,17	1382	12,17	8,30	3,66
05/11/2023	63,53	13,71	3,70	3,50	5,54	1466	12,35	8,51	3,72
12/11/2023	33,10	13,13	3,70	3,50	5,54	1466	12,35	8,51	3,72
19/11/2023	43,33	12,82	3,70	3,50	5,54	1466	12,35	8,51	3,72
26/11/2023	34,50	12,66	3,70	3,50	5,54	1466	12,35	8,51	3,71
03/12/2023	55,00	12,61	3,80	3,29	4,19	1557	12,35	8,51	3,70
10/12/2023	63,13	12,79	3,80	3,29	4,19	1557	12,35	8,51	3,70
17/12/2023	65,50	12,69	3,80	3,29	4,19	1557	12,35	8,51	3,70
24/12/2023	55,68	12,54	3,80	3,29	4,19	1557	12,35	8,51	3,70
31/12/2023	60,24	12,17	3,80	3,29	4,19	1557	12,35	8,51	3,54
07/01/2024	54,55	12,41	3,70	3,22	5,72	1504	12,35	8,51	3,51
14/01/2024	50,71	11,85	3,70	3,22	5,72	1504	12,35	8,51	3,51
21/01/2024	41,18	11,22	3,70	3,22	5,72	1504	12,35	8,51	3,51
28/01/2024	55,14	11,90	3,70	3,22	5,72	1504	12,35	8,51	3,74
04/02/2024	59,09	14,11	3,90	3,36	4,30	1698	12,49	8,57	4,04
11/02/2024	48,62	13,56	3,90	3,36	4,30	1698	12,49	8,57	4,04
18/02/2024	51,07	12,84	3,90	3,36	4,30	1698	12,49	8,57	4,04
25/02/2024	47,89	13,12	3,90	3,36	4,30	1698	12,49	8,57	3,93
03/03/2024	73,45	13,31	3,90	3,58	4,73	1491	12,49	8,57	3,64
10/03/2024	69,03	13,57	3,90	3,58	4,73	1491	12,49	8,57	3,64
17/03/2024	65,45	13,01	3,90	3,58	4,73	1491	12,49	8,57	3,64

24/03/2024	65,00	13,92	3,90	3,58	4,73	1491	12,49	8,57	3,64
31/03/2024	65,42	13,25	3,90	3,58	4,73	1491	12,49	8,57	3,83
07/04/2024	63,85	13,42	3,90	3,78	4,25	1659	12,49	8,57	3,86
14/04/2024	66,50	13,68	3,90	3,78	4,25	1659	12,49	8,57	3,86
21/04/2024	70,25	13,41	3,90	3,78	4,25	1659	12,49	8,57	3,86
28/04/2024	66,82	13,51	3,90	3,78	4,25	1659	12,49	8,57	3,76
05/05/2024	27,50	12,51	4,00	3,76	2,81	1557	11,92	8,65	3,68
12/05/2024	51,00	13,03	4,00	3,76	2,81	1557	11,92	8,65	3,68
19/05/2024	40,58	13,43	4,00	3,76	2,81	1557	11,92	8,65	3,68
26/05/2024	48,92	14,22	4,00	3,76	2,81	1557	11,92	8,65	3,67
02/06/2024	57,12	13,86	4,10	3,60	2,50	1725	11,92	8,65	3,60
09/06/2024	35,48	13,71	4,10	3,60	2,50	1725	11,92	8,65	3,60
16/06/2024	41,92	13,03	4,10	3,60	2,50	1725	11,92	8,65	3,60
23/06/2024	41,14	12,98	4,10	3,60	2,50	1725	11,92	8,65	3,60

İNTİHAL RAPORU İLK SAYFASI

SANAL ORTAM VERİLERİ İLE YAPAY ZEKA TABANLI DİNAMİK TAHMINLEME MODELİ GELİŞTİRİLMESİ

ORJİNALLİK RAPORU

% 3	% 3	% 2	% 1
BENZERLİK ENDEKSİ	İNTERNET KAYNAKLARI	YAYINLAR	ÖĞRENCİ ÖDEVLERİ

BİRİNCİL KAYNAKLAR

1	acikbilim.yok.gov.tr İnternet Kaynağı	% 1
2	edition.cnn.com İnternet Kaynağı	% 1
3	acikerisimarsiv.selcuk.edu.tr:8080 İnternet Kaynağı	<% 1
4	Submitted to The Scientific & Technological Research Council of Turkey (TUBITAK) Öğrenci Ödevi	<% 1
5	Aykurt, İsmail Sarp. "Yeni Medya Çağında Suskunluk Sarmalı: Twitter Örneği", Maltepe University (Turkey), 2025 Yayın	<% 1
6	acikerisim.karabuk.edu.tr:8080 İnternet Kaynağı	<% 1
7	www.researchgate.net İnternet Kaynağı	<% 1
8	lib.hacibayram.edu.tr İnternet Kaynağı	<% 1
9	Akkaş, Fatma. "Türkiye'de Vergi Mükelleflerine Yönelik Elektronik Vergi Uygulamalarının Vergi Uyumu üzerindeki Etkileri: Balıkesir ili	<% 1

KURUM İZNİ YAZILARI

Uyarı: Canlı ve cansız deneklerle yapılan tüm çalışmalar için kurum izin belgelerinin eklenmesi zorunludur. Gizlilik ve mahremiyet içeren durumlarda kurum adı kapatılmalıdır.

- ☐ Kurum izni gerekmektedir.
- ☒ Kurum izni gerekmemektedir.

Adem ÇAYIR
(İmza)

