



**BİR HAVAYOLU İŞLETMESİNDE KABİN EKİPLERİNİN PERFORMANS
DEĞERLENDİRMESİNE İLİŞKİN BİR VERİ MADENCİLİĞİ
UYGULAMASI**

Büşra YILDIZ KORKMAZ

**YÜKSEK LİSANS TEZİ
ENDÜSTRİ MÜHENDİSLİĞİ ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

ŞUBAT 2019

Büşra YILDIZ KORKMAZ tarafından hazırlanan “BİR HAVAYOLU İŞLETMESİNDE KABİN EKİPLERİNİN PERFORMANS DEĞERLENDİRMESİNE İLİŞKİN BİR VERİ MADENCİLİĞİ UYGULAMASI” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi Endüstri Mühendisliği Ana Bilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Danışman: Doç. Dr. Uğur ÖZCAN

Endüstri Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.



Başkan: Doç. Dr. Hakan ÇERÇİOĞLU

Endüstri Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

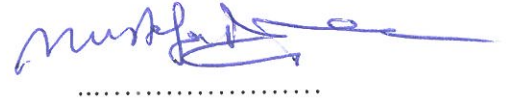
Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.



Üye: Dr. Öğr. Üyesi Mustafa İsa DOĞAN

Yöneylem Araştırması Ana Bilim Dalı, Düzce Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.



Tez Savunma Tarihi: 22.02.2019

Jüri tarafından kabul edilen bu tezin Yüksek Lisans Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

.....

Prof. Dr. Sena YAŞYERLİ

Fen Bilimleri Enstitüsü Müdürü

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
 - Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
 - Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
 - Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
 - Bu tezde sunduğum çalışmanın özgün olduğunu,
- bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.


Büşra YILDIZ KORKMAZ

22/02/2019

BİR HAVAYOLU İŞLETMESİNDE KABİN EKİPLERİNİN PERFORMANS DEĞERLENDİRMESİNE İLİŞKİN BİR VERİ MADENCİLİĞİ UYGULAMASI

(Yüksek Lisans Tezi)

Büşra YILDIZ KORKMAZ

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Şubat 2019

ÖZET

Günümüzde işletmeler gerek piyasaya tutunmak gerekse her geçen gün gelişmekte olan teknolojiyi yakalamak adına yoğun bir rekabet içerisindeyler. Yoğun rekabet ortamı mevcut müşteriyi tutma ve yeni müşteri kazanma amacını da beraberinde getirmektedir. Havayolu işletmelerinde yolculara beklentilerinin ötesinde hizmet sunma noktasında kabin ekibinin etkisi büyüktür. Bir havayolu işletmesinde 3764 kabin memurunun 2015 yılında performans değerlendirmeleri incelenmiştir. Yapılan bu performans değerlendirmelerinin sonucunda karne düzeyleri belirlenmektedir. Bu çalışmanın amacı 2015 yılındaki karne düzeyleri için; kabin memurlarının yetkinlik bazlı değerlendirme puanları ile demografik özellikleri arasında anlamlı bir kural oluşturmaktır. Bu çalışmada, açık kaynak kodlu JAVA dilinde geliştirilmiş WEKA programı ile veri madenciliği yöntemlerinden karar ağacı algoritmaları kullanılmıştır. Oluşturulan karar ağacı algoritmalarından sınıflandırma doğruluğu açısından en başarılı algoritma olarak Random Forest ve ikinci olarak J48 algoritması tespit edilmiştir. Random Forest algoritma çıktısı görsel bir sonuç vermeyip algoritma adımlarını görülmeyecek şekilde vererek karmaşık bir yapı oluşmasından dolayı çalışma J48 algoritmasına göre yorumlanmıştır. Ayrıca, WEKA programında nitelik seçimi özelliği ile InfoGainAttributeEval algoritması ile “Ranker” metodu uygulanması sonucunda çıktıların J48 algoritması çıktıları ile aynı doğrultuda olduğu tespit edilmiştir. Bu bağlamda kabin memurlarının karne düzeylerini en çok etkileyen niteliğin “sürekli öğrenme ve kişisel gelişim” olduğu ve demografik özellikler ile karne düzeyi arasında anlamlı kural olmadığı tespit edilmiştir.

Bilim Kodu : 90608

Anahtar Kelimeler : Veri madenciliği, karar ağaçları, havayolu, performans değerlendirme, kabin memuru

Sayfa Adedi : 87

Danışman : Doç. Dr. Uğur ÖZCAN

A DATA MINING APPLICATION FOR PERFORMANCE EVALUATION OF CABIN CREW MEMBERS IN AN AIRLINE COMPANY

(M. Sc. Thesis)

Büşra YILDIZ KORKMAZ

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

February 2019

ABSTRACT

Nowadays, the companies are in competition fiercely both for keeping market and catching the developing technology. The fierce competition environment aims to keep current customers and gain new customers. In airline companies, the cabin crew has great affect on providing passengers with services beyond their expectations. Performance evaluations of 3764 cabin crew members were examined in 2015 in an airline company. The levels of scorecard are determined as a result of making these evaluations. The objective of this study is making a meaningful rule between evaluation scores based on competence and demographic features for the levels of scorecard in 2015. In this study, WEKA was used, which is developed in open source code JAVA, and decision tree algorithms which is one of data mining methods. It was explored that Random Forest algorithm was the best algorithm and second one was J48 algorithm in terms of true positive rate in generated decision tree algorithms. This study was interpreted according to the J48 algorithm because Random Forest algorithm output was not suitable for this study due to nonvisual output and complex structure in steps of the algorithm. In addition, it was used for attribute selection with Ranker method in “ Info Gain Attribute Eval” algorithm and the results was detected similar to J48 algorithm outputs. In this regard, it was determined that the most important attribute affecting cabin crew scorecard levels was “Continuous learning and personnel development” and was also determined no meaningful rule between demographic attributes and scorecard level.

Science Code : 90608

Key Words : Performance evaluation, cabin crew, data mining, decision tree, airline

Page Number : 87

Supervisor : Assoc. Prof. Dr. Uğur ÖZCAN

TEŞEKKÜR

Bu tezin hazırlanması aşamasında benden yardımlarını esirgemeyen ve çalışmamın tüm aşamalarında yol gösteren Hocam Doç. Dr. Uğur ÖZCAN'a, geliştirme aşamalarında çalışmanın uygulanabilirliği konusunda yüreklendiren değerli mesai arkadaşlarıma ve sonsuz desteklerini esirgemeyen çok kıymetli aileme teşekkürü bir borç biliyorum.



İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	x
ŞEKİLLERİN LİSTESİ	xi
SİMGELER VE KISALTMALAR.....	xii
1. GİRİŞ.....	1
2. PERFORMANS KAVRAMI VE YÖNTEMLERİ	3
2.1. Performans Yönetimi	3
2.2. Performans Ölçümü ve Performans Değerlendirme	4
2.2.1. Finansal performans ölçüm yöntemleri.....	6
2.2.2. Üretim esaslı performans ölçüm yöntemleri	6
2.2.3. Kalite esaslı performans ölçüm yöntemleri.....	6
2.2.4. Pazarlama esaslı performans ölçüm yöntemleri.....	7
2.2.5. Stratejik yönetim esaslı performans ölçüm yöntemleri.....	7
2.2.6. Yönetim muhasebesindeki performans ölçüm yöntemleri.....	7
2.3. Dengeli Performans Değerlendirme Modeli (Balanced Scorecard Model)	7
2.3.1. Kurumsal karnenin boyutları.....	8
2.4. Geleneksel Performans Değerlendirme Modellerinin Yetersizlikleri.....	9
3. BİR HAVAYOLU İŞLETMESİNDE PERFORMANS YÖNETİMİ	11
3.1. Temel Kavramlar ve Açıklamalar	11
3.2. Kabin Ekiplerinin Performans Yönetimi	11
3.2.1. Kabin ekiplerinin performans yönetiminin kazanımları	14
3.2.2. Değerlendirme hataları.....	15

	Sayfa
3.3. Literatür Taraması	15
4. VERİ MADENCİLİĞİ.....	19
4.1. Veri Madenciliği Süreci	21
4.1.1. Problemin tanımlanması	23
4.1.2. Verilerin hazırlanması (veri ön işleme).....	23
4.1.3. Modelin kurulması ve değerlendirilmesi	24
4.1.4. Modelin kullanılması	25
4.1.5. Modelin izlenmesi.....	25
4.2. Veri Madenciliği Yöntemleri	25
4.2.1. Sınıflama ve regresyon.....	26
4.3. Sınıflandırma Algoritmaları	28
4.3.1. Karar ağaçları	28
4.3.2. Diskriminant analizi	33
4.3.3. Naive bayes	34
4.3.4. Sinir ağları	34
4.3.5. Kaba kümeler	35
4.3.6. Genetik algoritma.....	35
4.3.7. Regresyon analizi	36
4.4. Kümeleme	36
4.5. Birliktelik Kuralları ve Ardışık Zamanlı Örüntüler	37
4.6. Veri Madenciliğinde Karşılaşılan Problemler.....	39
4.7. Veri Madenciliği Uygulama Alanları.....	40
5. KABİN MEMURLARININ PERFORMANS DEĞERLENDİRMESİNE İLİŞKİN BİR VERİ MADENCİLİĞİ UYGULAMASI	43
5.1. Problemin Tanımlanması ve Amacın Belirlenmesi	43
5.1.1. İş kavrama	44

	Sayfa
5.1.2. Veriyi kavrama.....	44
5.1.3. Veriyi hazırlama.....	46
5.2.4. WEKA ‘da sınıflandırma algoritmalarının uygulanması ve algoritma sonuçları.....	51
5.2.5. Sınıflandırma algoritmalarının karşılaştırılması	57
5.2.6. J48 algoritmasına göre sonuçların yorumlanması.....	58
5.2.7. Random Forest algoritmasına göre sonuçların yorumlanması.....	61
6. SONUÇ.....	65
KAYNAKLAR	69
EKLER.....	75
EK-1. WEKA explorer ekran görüntüsü.....	76
EK-2. Çalışmada ilk kullanılan veri türleri.....	77
EK-3. Çalışmada ilk kullanılan veri türleri sınıflandırma algoritma sonuçlarının karşılaştırılması ve J48 karar ağacı algoritmasının boyutları.....	78
EK-4. Çalışmada yer alan değişkenlerin WEKA görselleri.....	79
EK-5. J48 algoritması karar ağacı görseli.....	83
EK-6. Random Forest algoritması (50 Ağaç)	84
EK-7. Random Forest algoritması (100 Ağaç)	85
EK-8. Random Forest algoritması (150 Ağaç)	86
ÖZGEÇMİŞ	87

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 3.1. Havayolu işletmelerinin rekabet unsurları.....	12
Çizelge 3.2. Değerlendirme skalası	14
Çizelge 3.3. Kademe düzeyi tanımları	14
Çizelge 4.1. Hata matrisi.....	27
Çizelge 4.2. CART, CHAID, ID3 ve C4.5 karşılaştırması	30
Çizelge 4.3. Veri Madenciliği uygulama alanları ve oranları	41
Çizelge 5.1. Yaş değişkeninin tanımlanması	47
Çizelge 5.2. Eğitim değişkeninin tanımlanması	48
Çizelge 5.3. Karne düzey değişkeninin tanımlanması	49
Çizelge 5.4. Veri dönüşümü sonrası veri tablosu	49
Çizelge 5.5. Çalışmada kullanılan nihai veri türleri	50
Çizelge 5.6. Sınıflandırma algoritmalarının karşılaştırılması	57
Çizelge 5.7. Random Forest algoritmasının ağaç sayılarına göre karşılaştırılması	62
Çizelge 5.8. Nitelik seçme ve sonuçlar	64

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 2.1. Performans yönetiminde performans ölçümü.....	4
Şekil 2.2. Kurumsal karne boyutları	9
Şekil 4.1. Veri tabanlarında bilgi keşfi adımları	20
Şekil 4.2. CRISP-DM süreci.....	22
Şekil 4.3. Hu ve Jha ‘nın veri madenciliği süreci	23
Şekil 4.4. ROC eğrisi	28
Şekil 4.5. Karar ağacı örneği.....	32
Şekil 4.6. Yapay sinir ağları modeli.....	35
Şekil 4.7. Genetik algoritma blok diyagramı	36
Şekil 4.8. Apriori algoritması.....	39
Şekil 5.1. Algoritma görseli (J48).....	52
Şekil 5.2. Algoritma görseli (OneR)	53
Şekil 5.3. Algoritma görseli (ZeroR)	54
Şekil 5.4. Algoritma görseli (Rastgele Ağaç)	54
Şekil 5.5. Algoritma görseli (Çok Katmanlı Algılayıcı).....	55
Şekil 5.6. Algoritma görseli (Rastgele Orman)	56
Şekil 5.7. Algoritma görseli (Naive Bayes)	56
Şekil 5.8. Algoritma görseli (DecisionStump).....	57
Şekil 5.9. KF1 değişkenin 15’ten küçük veya eşit olması durumunda oluşan karar kuralı	59
Şekil 5.10. KF1 değişkeninin 15’ten büyük olması durumunda oluşan karar kuralı.....	60
Şekil 5.11. MO1 değişkeninin 15’ten büyük olması durumunda oluşan karar kuralı	61

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Kısaltmalar	Açıklamalar
AP	Analiz Etme ve Problem Çözme
CRISP-DM	Çapraz Endüstri Veri madenciliği Standart Süreci
CRM	Müşteri İlişkileri Yönetimi
EI1	Etkili İletişim_1
EI2	Etkili İletişim_2
KF1	Kurumsal Farkındalık_1
KF2	Kurumsal Farkındalık_2
KPK	Kurumsal Performans Karnesi
MO1	Müşteri Odaklılık_1
MO2	Müşteri Odaklılık_2
RDBMS	İlişkisel Veri Tabanı Yönetim Sistemi
RF	Random Forest
ROC	Receiver Operating Characteric
SO	Sürekli Öğrenme ve Kişisel Gelişim
TR1	Takım Ruhu ve Dayanışma_1
TR2	Takım Ruhu ve Dayanışma_2
VM	Veri Madenciliği
VTBK	Veri Tabanı Bilgi Keşfi
WEKA	Waikato Environment for Knowledge Analysis
YSA	Yapay Sinir Ağları

1. GİRİŞ

Günümüzde işletmeler gerek piyasaya tutunmak gerekse her geçen gün gelişmekte olan teknolojiyi yakalamak adına yoğun bir rekabet içerisinde dirler. Yoğun rekabet ortamı mevcut müşteri tutma ve yeni müşteri kazanma amacını da beraberinde getirmektedir.

İşletmeler öngördükleri sonuçlara ulaşarak etkinliği veya kaynakları uygun bir şekilde kullanarak verimliliği sağlama amaçları, performans değerlendirmesini önemli kılmaktadır. Tüm çalışanların ortak bir paydada buluşarak etkin kılınan performans yönetimi, amaç ve hedeflere ulaşmada gerekli olan planlama, ölçme, kontrol gibi işlemlerini devamlı olarak değerlendiren yönetim biçimidir. Performans yönetimi, belirlenen amaçlara ulaşmak için mevcut durum ve istenen durum arasındaki kıyaslamayı yaparak sürekli iyileştirme sağlayan bir mekanizmadır. İşletmeler değişen çevre koşullarında rekabet güçlerini arttırmak amacıyla stratejik planlarına ve hedeflerine ne derece ulaşıldıklarının tespiti için sadece finansal boyutta değerlendirme yapan geleneksel performans ölçütleri ile değil çok ölçütlü performans değerlendirmelerini kullanmaya başlamışlardır. Finansal boyutun yanı sıra müşteri boyutu, içsel işlemler boyutu, bilgi ve öğrenme boyutu ile performans değerlendirmelerinin alt yapısı oluşmaktadır.

İşletmelerin ayrıca geleneksel performans ölçütlerini kullanmamasının sebepleri arasında, kısa dönem odaklı olması, geçmişe yönelik olması ve işletmedeki birçok bölüm ile alakalı olmaması gibi bu yetersizlikler de yer almaktadır. Kurumsal Karne (Dengeli Başarı Göstergesi), geçmişe yönelik verileri iz gösterge olarak belirleyerek işletmeyi bir bütün olarak değerlendiren ve stratejileri uygulamaya dönüştüren tüm boyutları ile dengeli bir performans ölçümüdür.

Bu çalışmada, 2. bölümde performans yönetimi ve değerlendirmeleri, kurumsal karne kavramı ve boyutlarından bahsedilmiştir. 3. bölümde kabin ekiplerinin performans yönetimi, yetkinliklerin oluşması ve detayları, tez çalışmasında sıklıkla geçen temel kavramların açıklamaları ve literatür taraması yer almıştır. 4. bölümde çalışmada yer alan performans değerlendirmesinde bir yöntem olarak kullanılan veri madenciliği algoritmalarından bilhassa sınıflandırma algoritmalarından biri olan karar ağaçları aktarılmıştır. 5. bölümde çalışmanın uygulama safhasından ve faydalanılan karar ağacı algoritmaları sonuçlarından bahsedilmiştir. 6. bölümde uygulamanın sonucundan ve katkısından bahsedilmiştir.

Türkiye’de faaliyet gösteren bir havacılık işletmesinde; kabin memurları gerçekleştirdikleri tarifeli dış hat seferlerinde kabin amirleri tarafından performans değerlendirmesine tabi tutulmaktadırlar. Bu performans değerlendirmeleri sonrasında üçer aylık dönemlerde ve yıl sonunda karne verileri oluşturulmaktadır. Bu değerlendirme ölçütleri, işletmenin misyon, vizyon ve stratejik hedefleri doğrultusunda mesleki yetkinlik bazlı kriterlerden oluşmaktadır. Mesleki yetkinlikler işletmedeki yöneticiler tarafından kabin memurlarından beklenen yetkinlikler olarak belirlenmiştir. Başka bir deyişle, yetkinlikler; yolcu ile bire bir iletişime geçen kabin memurlarının görev ve sorumlulukları doğrultusunda kurumsal farkındalık izlenimi yaratmaları ve hızlı aksiyon almaları amacı ile belirlenmiştir.

Literatürde havacılık sektöründe performans değerlendirmesi ile ilgili yapılmış çalışmalar mevcuttur. Genel olarak istatistiksel yöntemler ile yapılmış çalışmalar bulunmaktadır. Bu tez çalışmasında veri madenciliği yöntemlerinden sınıflandırma algoritmaları ile çalışma sürdürülmüştür. Son yıllarda bilgi endüstrilerinde veri madenciliğinin büyük dikkat çekmesinin en büyük sebebi çok büyük miktarlardaki verinin elde edilebilirliği ve böyle verilerin yararlı bilgilere dönüştürülmesine duyulan ihtiyaçtır. Elde edilen bilgi işletme yönetimi, üretim kontrol ve pazar analizi, mühendislik tasarımı ve bilimsel keşiflere kadar değişen uygulamalar için kullanılmaktadır [1].

Bu çalışmada kabin memurlarının demografik ve mesleki yetkinlikleri ile birlikte 18 adet değişken değerlendirilmiştir. 2015 yılında her üç aylık dönemdeki mesleki yetkinlik değerlendirmelerinin sonucunda, yıl sonunda yayımlanan karne düzeylerinin; hangi mesleki yetkinliğe ve/veya hangi demografik özelliğe bağlı olduğu WEKA programı aracılığıyla sınıflandırma algoritmalarından karar ağaçları yöntemleri ile belirlenmiştir. Bu bağlamda kabin memurlarının yıllık karne düzeylerini etkileyen değişkenlerin geliştirilmesi ve elde edilen bilgi sayesinde kabin memurlarının uçuş bazında performanslarında iyileşme sağlanması ile birlikte global rekabet ortamında fark yaratmak amaçlanmıştır.

2. PERFORMANS KAVRAMI VE YÖNTEMLERİ

Performans kavramının birçok kaynakta farklı şekillerde tanımlaması yapılmaktadır. Performans genel bir ifade ile belli bir amaç doğrultusunda yürütülen amaç ve hedeflere yönelik erişilen noktayı, diğer bir deyişle başarılan hedefin kalite ve miktar yönleri ile belirlenmesidir. Bir diğer ifadeyle söylenecek olursa; performans, belirlenmiş olan bir hedefe mutlak ya da görelî ulaşım seviyesinin ölçümüdür [8].

Yenice, performans ölçümünü “bir işletmenin yararlandığı kaynak, üretmiş olduğu ürün ve hizmetleri, ortaya çıkan çıktıların takip edilmesi amacıyla; sistematik ve düzenli bir halde veri toplanması, bu verilerin analiz edilmesi ve raporlanması süreci” olarak tanımlamaktadır [70]. Performans ölçümünün birçok tanımı olmakla birlikte en kısa tanımıyla, verimliliğin ölçülmesidir olarak tanımlanmaktadır [68].

Ayrıca firma faaliyetleri için; hedef edinilmiş sonuçlara ulaşılabilirlik, hizmetlerde verimlilik, etkinlik ve karlılık durumlarının gerçekleşme değerlendirmesi amacıyla performans ölçümü önem arz eden bir araç olarak da tanımlanmaktadır [5].

2.1. Performans Yönetimi

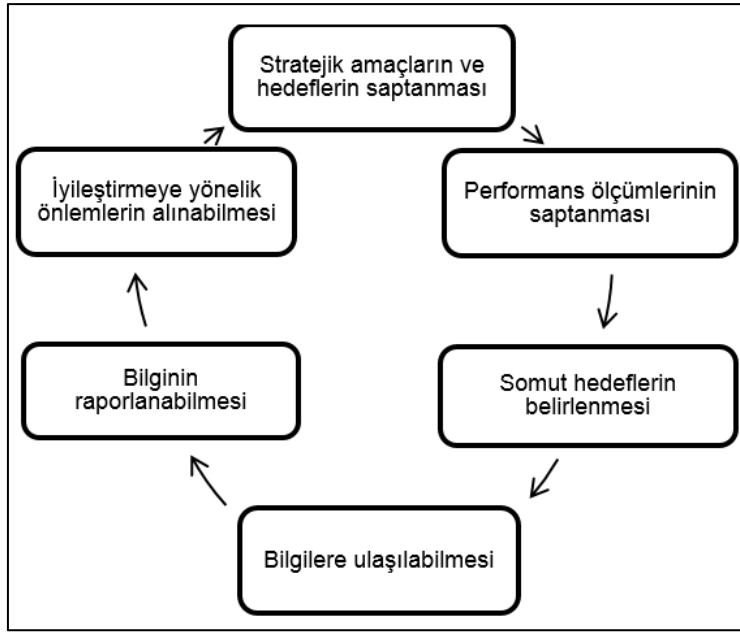
Performans yönetimi; işletmenin önceden belirlemiş olduğu amaç ve hedeflere varırken sürdürdüğü rota ve bu sürecin sonucunda edinilen sonuçların muhakeme edilmesidir [7].

Kuruluşlar açısından bakıldığında ise performans yönetimi, “örgütlerdeki bütün çalışanların, performansın devamlı ilerlemesini hedefleyen ortak amaçlar ile entegre etmeyi ve bu amaçlar için gereken planlama, ölçme, yönlendirme, kontrol işlerini ve yönetimin diğer yolları ile eşzamanlı halde devam etmeyi sağlayan bir yönetim biçimi” olarak tanımlanmaktadır [8]. Ayrıca, kurumda istenilen amaç için mevcutta ve gelecekte olması gereken durumlara ilişkin bilginin toplanması, bu verileri kıyaslama ve kurum performansının devamlı gelişimi için ilerlemeleri sağlama ve izleyen durumları kapsayan yönetim süreci olarak da tanımı yapılmaktadır [6].

Performans yönetiminin temelde iki boyutu olduğu söylenebilir. İlk olarak; kurumdaki belirlenmiş amaç ve hedeflere kaynak atamasının yapılmasının sonrasında sonuçların analiz

edilerek raporlandığı kurumsal performans yönetimidir. İkinci olarak, genelde insan kaynakları ile ilişkili olarak çalışan performans değerlendirmesini amaç edinen bireysel performans yönetimidir [9].

Performans yönetiminin amacı ise; kurum içerisindeki rollerin ve sorumlulukların açık bir şekilde tanımlanması ve bu tanımlamalara paralel iş süreçlerinin oluşturulmasıdır [4].



Şekil 2.1. Performans yönetiminde performans ölçümü [4]

2.2. Performans Ölçümü ve Performans Değerlendirme

Performans ölçümü işletmelerin misyon ve vizyon değerlerine göre saptanan amaç ve hedeflere göre oluşan ürün, hizmet ve/veya çıktıları beraber değerlendirildiği bir süreçtir [4].

Etkili performans yönetiminin temelini performans ölçümü oluşturmaktadır. Performans ölçümünün, yönetim için bir karar destek mekanizması işlevi görmesi amaçlanmaktadır. Bir işletmenin performans seviyesini belirlemek için performans ölçümü yaklaşımı, verimlilik ve etkinlik ilkelerine ne ölçüde ulaştığını, iyileşmeye açık alanların meydana çıkarılmasını gösterir. Performans ölçümünün belirlenmesi için, öncelikle amaçların ve optimum performans göstergelerinin karşılaştırmaya olanak sağlayacak şekilde belirlenmesi, sonrasında veri toplama ve ölçme aşamalarına geçilmesi gerekmektedir. Performans ölçümlerini genel olarak “performansı geliştirmeye yönelik amaçlara ulaşmak için önceden

belirlenen performans göstergelerine göre, kurumun kendi içinde gerçekleştirdiği sonuçları ve çıktıları ölçmek için kullanılan yöntemler” olarak da değerlendirmek mümkündür [8].

Tam bir performans değerlendirme yöntemi kurumlara etkin stratejiler tespit etme ve uygulama, çalışan davranışlarını yönlendirme, yönetsel etkinliği izleme ve ödüllendirmeler için bir zemin oluşturma olanağı sağlamaktadır [13].

Değerlendirmenin yapılabilmesi için öncelikle değerlendirmeyi ve ölçmeyi yapacak görevlilerin belirlenmesi gerekmektedir. Değerlendirmenin ilk amir tarafından yapılmasının en yaygın yöntem olduğu ileri sürülmektedir. Ancak tek amir değerlendirmesinin sakıncalarının önüne geçmek için veya minimize etmek amacıyla birden fazla hiyerarşik amirin bireysel ya da grup olarak değerlendirme yöntemi, kişinin kendi kendini değerlendirmesi, iş arkadaşları, astları ve müşteriler tarafından değerlendirme yapılması gibi usullerden faydalanılmaktadır [4].

Performans ölçümünde “ölçülemeyenin yönetilemeyeceği” düşüncesinden yola çıkılmaktadır. Performans ölçüm ve yönetiminde değişmekte olan koşullara daha hızlı adapte olma, kontrol mekanizması meydana getirilerek yönetimin iyileştirilmesi hedeflenmektedir. Ölçmek ve değerlendirmek arasındaki bağıntı ile devamlı geri bildirim oluşması ve yapının düzenli halde denetlenmesi ile mevcut durumdan daha iyi bir seviyeye ulaşılabilesine katkı sağlamaktadır [8].

Bu tez çalışmasında bir havayolu işletmesinde misyon, vizyon ve stratejik değerler ile yönetim tarafından, bir kabin ekibinde olması gereken yetkinlikler belirlenerek uçuş bazında performans ölçümü yapılmaktadır. Bu bağlamda, kabin memurlarının işletmede hali hazırda uygulanmakta olan çeyrek dönemlerde yayımlanan mesleki yetkinlik bazlı performans değerlendirme puanları ve yıl bazında performans değerlendirmelerinin neticesinde kurumsal karne verileri yayımlanmaktadır.

Kurumsal karne verileri birden fazla kabin ekibinin değerlendirmeleri ile nihai sonuca ulaşmaktadır. Bu sayede yukarıda bahsedilen sakıncalar ile birlikte objektif olmayan ve tekdüze değerlendirmelerin önüne geçilmektedir.

Performans değerlendirme modelleri genel hatları ile aşağıdaki gibi sınıflandırılmaktadır.

- Finansal performans ölçüm yöntemleri

- Üretim esaslı performans ölçüm yöntemleri
- Kalite esaslı performans ölçüm yöntemleri
- Pazarlama esaslı performans ölçüm yöntemleri
- Stratejik yönetim esaslı performans ölçüm yöntemleri
- Yönetim muhasebesindeki performans ölçüm yöntemleri

Bu yöntemlerden “Yönetim muhasebesindeki performans ölçüm yöntemleri” üzerinde durulmakta olup diğer yöntemlerde kısaca bahsedilmektedir.

2.2.1. Finansal performans ölçüm yöntemleri

Kurumdaki finansal, yatırım durumları odak olarak alınmakta olup geçmiş verileri değerlendirme ve belli finansal çizelgelerdeki veriler kullanarak geleceğe yönelik kararlar alınmaktadır. Bu yöntem en sık performans değerlendirmeye katkı sağlayan yöntem olup en büyük dezavantajı sadece maddi varlıklar ile sınırlı olmasıdır.

2.2.2. Üretim esaslı performans ölçüm yöntemleri

Üretim sektöründe var olan işletmeler tarafından üretim sürecinde kullanılan hammadde, iş gücü gibi kaynaklar ile etkinlik açısından performans değerlendirmesi yapılmaktadır. Bu yöntemin ise dezavantajı olarak finansal performans ölçüm yöntemlerinde ki gibi işletmenin bütününe değil belli bir alan ile sınırlı olmasıdır.

2.2.3. Kalite esaslı performans ölçüm yöntemleri

Bu performans değerlendirme yönteminin temelini toplam kalite yönetimi oluşturmaktadır. Toplam kalite yönetimi şirketin tümünü ele alarak kusursuz üretimini amaç edinmiştir. Kalite esaslı performans değerlendirmesinde kurum içi veya kurum dışı otoriteler tarafından öz değerlendirme ile yapılmaktadır.

2.2.4. Pazarlama esaslı performans ölçüm yöntemleri

İşletmelerin kullanmakta olduğu kaynakları nasıl ve ne kadara elde ettikleri; müşterilerine nasıl ve ne kadar maliyet üzerinden sundukları v.b. gibi maliyet ve fayda ilişkisi ile ilgili sonuçlar üzerinden değerlendirme yapmaktadır [14].

2.2.5. Stratejik yönetim esaslı performans ölçüm yöntemleri

İşletmenin hali hazırdaki performansı, işletmedeki yönetim kabiliyetleri ile misyon ve belirlemiş olduğu stratejiler ışığında denetleyerek ölçüm sağlanmaktadır.

2.2.6. Yönetim muhasebesindeki performans ölçüm yöntemleri

Başlıca performans ölçümleri; bütçeleme, standart maliyetleme, faaliyet tabanlı maliyetleme ve dengeli performans değerlendirme modeli (kurumsal karne) dir. Bu tez çalışmasında kabin ekiplerinin performans değerlendirmesinde kullanılmakta olan yöntem kurumsal karnedir.

2.3. Dengeli Performans Değerlendirme Modeli (Balanced Scorecard Model)

İşletme başarısını sadece finansal boyutlar ile gözlemleme yerine finansal olmayan boyutların da dahil edildiği modeller geliştirilmiştir. Bu modellerden biri olan “Dengeli Performans Modeli” ilk defa 1992 yılında Harvard Business Review’de yayımlanmıştır. Ardından ilgili makaleyi yazan Kaplan ve Norton 1993 yılında “Balanced Scorecard’ı Uygulamak” adında çalışmayı yayımlayarak Balanced Scorecard modelini daha geliştirmişlerdir.

Balanced scorecard; kurumsal performans karnesi (KPK), kurumsal performans yöntemi, kurumsal karne, dengeli performans değerlendirme tablosu gibi çok farklı terimler ile dilimizde karşılık bulmuştur. Çalışma da “kurumsal karne” terimi kullanılacaktır.

Kaplan ve Norton'a göre, KPK; sadece bir performans ölçüm sistemi olmamakla birlikte; ayrıca işletmenin uzun dönem boyunca stratejik amaçlara ulaşılması amacıyla çalışanların sahip olduğu enerji, yetenek, özellikli bilgilere yol gösteren yönetim sistemidir [15].

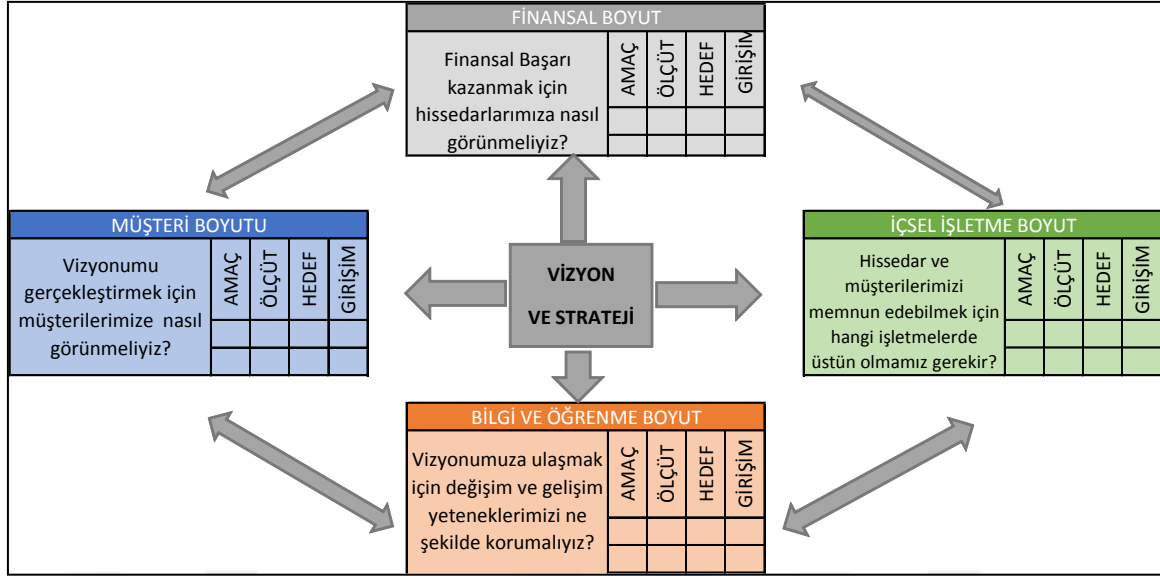
KPK kurumun vizyon, misyon ve strateji bakımından finans değerlerinin dahil etmediği ölçütlerini de kapsayan; performans değerlerine çevrildiği hat sunarak, kurumların performans izlemelerinde tecrübe ettikleri sorunlara çözüm bulmayı amaçlamaktadır [13]. Kaplan ve Norton' un ifade ettiği Kurumsal Karne, örgütlerin uzun dönemli rekabet ortamında müşteri odaklılık, ihtiyaçları daha efektif bir şekilde karşılama, kalite geliştirme, takım çalışmasını teşvik etme, iş geliştirme ve etkin yönetim gibi birbiri ile farklılık gösteren faktörleri aynı çerçevede sunan model olarak tanımlamaktadır [12].

Kurumsal Karne, karne olarak ifade edilen bir araya getirilmiş ve gruplanmış göstergelerden oluşmaktadır. Göstergeler dört ana yönetsel perspektif ile ilişkilendirilmiş olup üst yönetime yapmakta oldukları işler ve yürütülmekte olan faaliyetlerle ilgili olarak kapsamlı bir bakış sunmaktadır. Karneler; finansal, pazarlama, operasyonel ve stratejik boyutlar üzerinde organizasyonel performansın dengeli biçimde değerlendirilmesine imkân tanır. Kurumsal Karne; finansal ve operasyonel ölçütleri bir araya getirerek organizasyonun hem kısa hem de uzun vadeli amaçlarına odaklanmasına imkân tanır. Geleneksel finansal ölçütlerin tek başına organizasyonun performansı hakkında genel ve uygun bir bakış sağlayamayacağının farkına varılması, kurumsal karneyi tetikleyen etken olmuştur [11]

Kurumsal karnenin oluşturulması işlemleri, şirketin stratejik hedeflerinin kesin bir şekilde belirlenmesini ve bu hedeflere ulaşmayı sağlayacak kritik etkenlerin tanımlanmasını sağlar. Kurumsal karne, fikir birliği ve ekip çalışması ile eksiklikleri açık bir halde belirtirken, problemlerin sonuçlanmasını da sağlamaktadır. Kurumsal karne uygulaması işletmelerdeki üst düzey yöneticilerin birlikte çalışmasını zorunlu hale getirmektedir çünkü tüm yöneticilerin belli ilgilendikleri alan neticesinde katkı sağlamalarını gerektirmektedir. Sonuç olarak ortaya çıkan ise işletmenin tüm birimlerindeki faaliyetleri içeren bir model olarak tanımlanmaktadır.

2.3.1. Kurumsal karnenin boyutları

İşletmelerin misyon, vizyon ve stratejik kararlarının sonuçlarının ölçülmesi bazında 4 ayrı boyutta hedeflerin belirlenmesi ve ölçülmesi gerçekleştirilmektedir. Kurumsal karnenin boyutları Şekil 2.2'de gösterilmiştir [10].



Şekil 2.2. Kurumsal karne boyutları [10]

Kurumsal karnenin hedeflediği; işletmelerde üst düzey yöneticiler aşağıda yer alan 4 değişik yönden değerlendirerek ilgili cevapları araştırmaktadır [10].

- Finansal başarı amacıyla ortaklarımıza nasıl görünmemiz gerekmektedir? Finansal Boyut
- Hissedar ve müşterilerimizi memnun etmek amacıyla iyileştirilecek süreç tanımlamaları ve hangi işlemlerde mükemmelliğe ulaşmalıyız? İşletme İçi Yöntem Boyutu
- İşletme vizyonunu sağlamak amacıyla müşterilerimize nasıl görünmemiz gerekmektedir? Müşteri Boyutu (işletmenin müşteriler tarafından değerlendirilmesidir).
- Vizyona ulaşmak amacıyla mevcut değişim ve gelişim yeteneklerimizi nasıl arttırmamız gerekmektedir, değer yaratmanın ve iyileştirmenin devamlılığını ne şekilde sağlamamız gerekmektedir? Bilgi ve Öğrenme Boyutu [16].

2.4. Geleneksel Performans Değerlendirme Modellerinin Yetersizlikleri

Geleneksel yöntemler ile yapılan performans değerlendirmelerinde ana baz olarak “finansal” göstergeler kullanılmaktadır. Yalnızca finansal ölçümlerin kullanıldığı performans ölçümler dinamik ortamlara uyum sağlayamamaktadır. Geleneksel performans yönetim sisteminde birimler, çalışanlar bazında olmakta müşteriler dikkate alınmamaktadır.

Bu tez çalışmasında yapılmakta olan kabin ekiplerinin performans değerlendirmesinden en iyi sonuç alma beklentisinin amacı rakip havayolu işletmelerine göre müşteriye daha fazla

tutunma ve müşteriye kaybetmemedir. Bu sebeptendir ki bu tez çalışmasında performans değerlendirme yönteminde geleneksel performans değerlendirmesi baz alınmamıştır.

Toplam Kalite Yönetimi yaklaşımına önderlik eden Dr. William Edwards Deming’e ait söz olan; “Kâra bakarak bir firmayı yönetmek dikiz aynasına bakarak araba kullanmaya benzer. Dikiz aynası nerede olduğunuzu gösterir ancak nereye gittiğinizi göstermez.” ifadesiyle finansal performans parametreleri kadar finansal olmayan performans parametrelerini de saptayarak firma başarısını değerlendirmek ve hedeflerle belirlenen neticelerin bu parametrelere göre mukayese etmektir [73].



3. BİR HAVAYOLU İŞLETMESİNDE PERFORMANS YÖNETİMİ

3.1. Temel Kavramlar ve Açıklamalar

Kabin Memuru: Sivil Havacılık Genel Müdürlüğü'ne göre; “Yolcu taşımacılığında hava araçları için gerekli emniyet ve güvenlik önlemlerini uygulayıp ve yolcu konforundan sorumlu olan görevlilere “kabin memuru” ya da “kabin görevlisi” olarak” tanımlanmaktadır. Uçuş emniyetinin, güvenilirliğinin ve konforun gerçekleşebilmesi amacıyla hava aracının kabin bölümünde görevli olan kabin memurlarını; Sivil Havacılık Genel Müdürlüğü tarafından yayımlanmış ilgili mevzuatlarda belirtilen gereklilikleri sağlayıp eğitimleri başarı ile tamamlayıp ve akabinde “Kabin Memuru Sertifikası” almaya hak kazanan kişiler olarak tanımlaması yapılmaktadır [43].

Tarifeli Uçuş: Belli program çerçevesinde ülke içi ya da ülkeler arası belli frekanslarla yapılan yasal uçuş olarak tanımlanabilmektedir.

Dar gövde uçak: Teknik bir terim olmakla birlikte tek koridorlu olan uçaklara verilen bir isimdir.

Geniş gövde uçak: En az 1 çift koridora sahip olan ve haliyle dar gövde uçak tipine göre yolcu kapasitesi fazla olan uçak tipidir.

Yetkinlik: İşletmenin misyon ve vizyon değerleri doğrultusunda kabin memurlarının sahip olması gereken kriterlerdir.

3.2. Kabin Ekiplerinin Performans Yönetimi

Bir havayolu işletmesinde kabin memurlarına dış hat tarifeli uçuşlarda performans değerlendirmesi yapılmaktadır. Kabin memurlarının görev ve yetkinliklerine ilişkin belirlenmiş olan standartlara ne kadar uyum sağladıklarının ölçülmesinde yararlanılan performans değerlendirmesinin sonuçları büyük önem arz etmektedir.

İşletmenin vizyon ve misyon hedefleri doğrultusunda belli yetkinlikler çerçevesinde kabin memurları gerçekleştirdikleri dış hat tarifeli uçuşlarda kabin amirleri tarafından değerlendirilmektedirler.

Bu sistemin avantajları; tek bir değerlendirenin değil çok sayıda değerlendiren personel olması sebebiyle insani faktörlerden oluşabilecek hataları minimize etmekte ve işletmenin hedefleri doğrultusunda kaliteyi artırmanın yolunu açmaktadır.

Havayolu işletmeleri için aşağıda yer alan tabloda beş temel kategoride daha iyi bir hizmet sunabilmek ve rekabette üstün olmak maksadıyla yürütülmüş çalışmalar belirtilmiştir [80].

Çizelge 3.1. Havayolu işletmelerinin rekabet unsurları [38]

<u>Güvenilirlik ve İyi Hizmet</u> *Havaalanı Servis Hizmeti *Havaalanındaki hizmetlerinin iyi olması *Güvenilir olması *Kaza sayısının az olması *Büyük bir filo ya sahip olması *Filonun genç ve bakımlı olması *Müşteri menfaati konusunda titiz olması *Personelin müşteri isteklerine hızlı reaksiyonu	<u>Avantaj ve Uygunluk</u> *Ulaşım ağı genişliği *Fiyatların uygun olması *Uçuş saatlerinin uygun olması *Personelin müşteri sorularına tatmin edici cevaplar vermesi *Personelin müşterilere yardımcı olma konusundaki istekliliği
<u>Uzmanlık ve Esneklik</u> *Zamanında hareket etmesi *Sefer sayısının fazla olması *Personelin uzman olması *İletişimin kolay olması *Hizmetin taahhüt edildiği şekilde gerçekleşmesi *Hizmetin taahhüt edilen zamanda yerine getirilmesi *Sunulacak hizmetin zamanının tam olarak bildirilmesi	<u>Rahatlık ve Çeşitlilik</u> *Vakit geçirme olanakları (video, dergi vs) *Yiyecek ve içeceklerin çeşitliliği *Uçağın temizliği *Uçağın rahatlığı *Çalışma saatlerinin müşterilere uygun olması
	<u>Güvenlik ve Kalite</u> *Uçuş güvenliği (iniş-kalkış) *Kabin görevlilerin tutumu *Yiyecek ve içeceklerin kalitesi *Uçuş bilgilendirme

Bu tez çalışmasında da havayolu işletmelerinde sürdürülen, rekabette fark yaratmak amacıyla yukarıda belirtilen çizelgedeki bilgilerin ışığında kabin ekiplerinden beklenen kriterler ortaya konmaktadır. Ayrıca, yöneticilerin norm çalışmaları ile yolcu ile bire bir

etkileşimde olan kabin memurlarının mesleki yetkinlikleri belirlenmiştir. Bu sayede kabin memurlarının performansları üç ayda bir neticelenen karne çıktıları ile gözlemlenmekte ve hızlı aksiyon almaya yardımcı olabilmektedir.

Yöneticilerin şirketin misyon ve vizyon değerlerine göre belirledikleri aşağıda yer alan temel ve fonksiyonel yetkinlikler ve alt kırılımları ile kabin memurları değerlendirilmektedir.

- Kurumsal Farkındalık
 - a) Görüntü ve kişisel bakım açısından örnek bir kabin memurudur.
 - b) Davranış konusunda ve havayolumuzu temsil etme açısından örnek bir kabin memurudur.
- Etkili İletişim
 - a) Yolcularımızı güler yüzle karşıladı, uçuş boyunca güler yüzle hizmet etti.
 - b) Ekip arkadaşlarıyla ve yolcularımızla doğru iletişim kurdu
- Analiz Etme ve Problem Çözme
 - a) Görev esnasında karşılaştığı problemleri çözebildi. Görevi boyunca yolcularımıza alternatifler sundu
- Sürekli Öğrenme ve Kişisel Gelişim
 - a) Öğrenmeye ve eleştiriye açık bir kabin memurudur
- Takım Ruhu ve Dayanışma
 - a) Ekip ile uyumlu çalıştı, görev alanının dışında da ekip arkadaşlarına yardımcı oldu
 - b) Sorumluluk bilinciyle görevlerini tam ve zamanında yerine getirdi
- Müşteri Odaklılık
 - a) Ürün, hizmet ve hat bilgilerine hâkim, yolcularımıza doğru ve zamanında hizmet sundu
 - b) Yolcularımızın talebi olmadan hizmet sunmayı teklif edebildi. Servis öncesinde ve sonrasında yolcularımızın taleplerinin karşılanabilmesi için ulaşılabilir durumdaydı. Farklı kültürlerdeki yolcularımızın beklentilerini karşıladı.

Bu yetkinlikler bazında formlar 5 dereceli ölçüt ile değerlendirmeye alınmaktadır. Bir başka deyişle kabin memurları yukarıda yer alan her bir kriterde 5 üzerinden puan almaktadır.

Çizelge 3.2. Değerlendirme skalası

DERECE	TANIM
Çok İyi	Kişi tüm alt davranış göstergeleri ile yüksek derecede yetkinliğin kanıtlarını sergiler.
İyi	Kişi ilgili yetkinliğe ait alt davranış göstergelerinin çoğunu sergiler.
Yeterli	Kişi ilgili yetkinliğe ait alt davranış göstergelerinin bir kısmını sergiler. Pozitif davranış göstergelerinin tersi davranış göstergeleri sergilemez.
İyileştirme gerekli	Kişi ilgili yetkinliğe ait alt davranış göstergelerinin tersi davranış göstergelerini de bir ölçüde sergiler. Bir takım pozitif davranış göstergeleri de bulunur.
Yetersiz	Kişi ilgili yetkinliğe ait alt davranış göstergelerinin tersi davranış göstergelerini önemli ölçüde sergiler. Pozitif davranış göstergeleri de bulunur veya hiç bulunmaz.

Kabin memurlarının mesleki yetkinliklerde 3 aylık dönem boyunca değerlendirme formları üzerinden aldıkları puanların ortalaması ham başarı puanını vermektedir. Kabin memurlarının kendi sınıflarına ait diğer kabin memurlarının ham puanlarına göre sınıf ortalaması hesaplanır ve dağılımına bakılır. Bu ölçütlere göre ham başarı puanı harfli puana dönüştürülür.

Çizelge 3.3. Kademe düzeyi tanımları

Kademe Düzeyi Tanımları				
Çok İyi	İyi	Yeterli	İyileştirme Gerekli	Yetersiz
“A”	“B1”- “B2”	“C1”	“C2”- “D1”	“D2”- “E”

Performans değerlendirmelerinin neticesinde kabin memurlarının nerede olduklarını ve kendilerini nasıl geliştirmeleri gerektiğini yansıtmaları amacıyla üç aylık kurumsal karne yayımlaması yapılmaktadır. Bu bağlamda Şekil 2.2. de belirtilen kurumsal karnenin 4 boyutu da sağlanmaktadır.

3.2.1. Kabin ekiplerinin performans yönetiminin kazanımları

Hali hazırda kullanılmakta olan bu uygulama ile kabin memurlarının başarı durumlarına istinaden verilere kısa zamanda ulaşılabilmekte ve hızlı aksiyon alınabilmektedir. 2. bölümde performans yönetiminde tek bir amirin değerlendirme sakıncalarının bulunmasından ve bu durumu minimize etmek amacıyla hiyerarşik amirlerin değerlendirme

yapmasının olumlu sonuçlar doğurmasından bahsedilmişti. Bu durumdan hareketle kabin memurları gerçekleştirdikleri her uçuş seferinde farklı ekipler ile görev icra etmelerinden dolayı, performans değerlendirmelerinde fazla sayıda değerlendirici bulunmaktadır. Bu bağlamda değerlendirici hataları en aza indirilmeye çalışılmaktadır.

Ayrıca, şirketin kazanımları ile orantılı olarak stratejilere göre belirlenen organizasyon hedefleri kabin memurlarına ulaşılmasını sağlayarak hizmet kalitesini arttırmaya katkıda bulunmaktadır.

3.2.2. Değerlendirme hataları

İlgili değerlendirmelerde aktif görev üstlenen kabin amirlerinin, kabin memurlarının kişisel ve mesleki gelişimleri için katkı sağladıkları yadsınamaz bir gerçektir. Bu bağlamda, genel olarak değerlendirme hataları aşağıdaki gibidir.

- Müsamaha ve Katılık: Yapılan performans değerlendirmelerinde mevcut durumdakinin üstünde veya altında puanlama eğilimi olarak bilinmektedir.
- Hale Etkisi: Kabin ekiplerinin performans değerlendirmesinde konu olarak birbirinden farklı yetkinlikler yer almaktadır. Değerlendiricinin birbirinden farklı olan yetkinlikleri göz ardı ederek hatalı değerlendirme yapması demektir.
- Yakın Geçmişteki Olaylardan Etkilenme/Kişisel Önyargılar: Performans değerlendirmelerinde mesleki yetkinliklerin değil, değerlendirmeleri kişisel olarak algılayıp geçmişte vuku bulan durumların tesiri altında kalınarak yapılan değerlendirmelerdir.

3.3. Literatür Taraması

Literatürde havacılık sektöründe kabin ekiplerinin performans yönetimi ve ayrıca ilgili sektördeki kurum içinde yapılan değerlendirmeler ve kuruma iletilen geri bildirimler geri ile ilgili çalışmalar aşağıda yer almaktadır. Bunlar;

- Furnham, A. (1991) tarafından yapılan çalışmada; “öngörülen mesleki başarı da verimlilik testinin etkililiğinin belirlenmesi” için başlangıçta 136 kabin memuru 16 PF Kişilik Envanterini (16 temel kişilik Özelliğini ve 5 genel kişilik eğilimini ölçen kişilik envanteri) tamamlamışlardır. 6 aylık bir zamanda uçuştaki davranışları günlük olarak rapor halinde tutulmuştur. Bu rapor daha sonra ortalama altı, ortalama ve ortalama üstü

olarak kategorize edilmek için havayolundan kimlerin ayrılacağını saptamak için kullanılmıştır. Bazı trendler belli olmasına rağmen, Pearson korelasyonu ve tek yönlü ANOVA analizi anlamlı farklılıklar için çeşitli gruplarda (ortalama altı-ortalama) başarısız olmuştur [17].

- Nameghi ve Ariffin (2013) tarafından yapılan çalışmada; Malezya havayollarında yapılmış olan çalışmada havacılık misafirperverliğinin altında yatan boyutları açığa çıkarılmıştır. Bu çalışma havacılık misafirperverliğini kabin ekibi performansına veya ev sahipliği davranışına dayandırarak tarif etmiştir. Bugünün havayolu taşımacılık endüstrisi oldukça yoğun bir şekilde rekabet içerisinde olmasından dolayı havayolu şirketleri sadece ayırt edici servis sunma değil uçuş boyunca misafirperver servis yollarını bulmak zorundadır. 391 yolcunun içerdiği ilk data seti üzerinde yapılan keşfedici faktör analizi göstermiştir ki; misafirperverlik total varyansın %67'sinden gelen 4 boyut içermiş olduğu tespit edilmiştir [18].
- Gencer ve Çetin'in (2011) yapmış olduğu çalışmada; "Havacılık sektöründe faaliyetlerini devam ettiren bir işletmede stratejik planların değerlendirilmesinde; Kurumsal Performans Karne çalışması sürdürülmüştür. Halihazırda var olan durum analiz edilerek, bir sonraki altı aylık zaman dilimi için model kullanılmış ve ilgili zaman aralığı neticesinde değerlendirmeler yapılmıştır. Saptanan neticeler tetkik edildiğinde işletmenin karne notu % 88 olarak elde edilmiştir" [4].
- Okumuş, A. ve Asil, H. (2007), X Hava Yolları'nda dış hat uçuşları esnasında 511 yolcu ile yapılmıştır. Araştırmanın sonucu olarak; hizmet kalitesinin algılamasında, yolcuların memnuniyet seviyelerinde en etkili faktörün "anında hizmet faktörü" olduğu tespit edilmiştir. Bu faktörü; "güvenilir ve doğru hizmet", "güven telkin etme" ve "müşteriyi tanıma-anlama" faktörleri devam ettirmiştir.
- Hatipoğlu, S. ve Işık, E. S. (2015), 2008 yılında kurulmuş olan Anadolu Jet'in iç hatlardaki müşteri profilini meydana çıkarmak ve hizmet kalitesini ölçümlemek amaçlanmıştır. Yolculara, SERVQUAL kalite ölçüm modeli ile oluşturmuş bir anket uygulaması yapılmıştır. Anket sonuçlarına göre hizmet kalemlerinin boşluk analizleri hesaplanmış, Anadolu Jet'in müşteri memnuniyeti açısından beklentinin yüksek, algının düşük olduğu hizmet kalemleri ortaya çıkarılmıştır. SERVQUAL skorlarına göre; yolcuların genel anlamda; çalışanların zamanında, hızlı ve doğru bilgi vermeleri, saygılı olmaları, teknik bilgi ve yeteneklerinin yeterli olması, uçuş arızası, uçuş iptal

edilme/ertelenme olasılığının düşük olması önermelerinden memnun olduklarını göstermektedir.

- Alkoç, H. (2004), THY dış hat yolcularıyla yapmış olduğu çalışmada, SERVQUAL modeli ile hizmet kalitesini değerlendirmiştir. Dış hat yolcularının, kalite boyutları içerisinde sırayla isteklilik, güvenilirlik, teminat ve empati boyutlarına önem verdiği; fiziki unsurların, memnuniyet üzerinde anlamlı bir etkisinin olmadığı saptanmıştır. Ek olarak, Türk ve yabancı yolcu profillerinin daha ayrı bir kalite beklentisine sahip olduğunu ortaya çıkarmıştır.

Literatürde ayrıca havacılık sektöründe yapılmış olan veri madenciliği uygulamaları da incelenmiştir. Bu çalışmalar;

- IBM'in Cathay Pacific Havayollarına yaptığı bir çalışmada tüm sık uçan müşterilerinin verisi veri ambarına taşınıp veri madenciliği algoritmaları uygulanmış. Burada yapılan çalışmada müşteriler farklı özelliklerine göre segmentlere bölünmüştür. Yapılan bu segmentasyon havayolu şirketine özel müşteri segmentlerine odaklanabilme imkânı sağlamıştır [51].
- Pritscher'e göre havayolu işletmelerinde veri madenciliğinin en açık bir şekilde uygulanabileceği alan sık uçanlar programlarıdır. Avrupa'nın ana havayolu ittifaklarından biri olan Qualiflyer'de yapılan bir çalışmada CRM'in başarılı bir şekilde uygulanmasını desteklemek için veri madenciliği yöntemleri kullanılmıştır. Buradaki çalışmada ilk iş yüksek kâr potansiyeline sahip müşterilerin pazar segmentini belirlemek olmuştur [51].
- Harris'e göre İngiltere havayolları (British Airways) tek yöne uçan ve geriye dönüşte farklı havayolu şirketini tercih eden önemli müşterileri hakkında araştırma yapmış. Elde ettiği sonuçlara göre bu müşterilerine her iki yönde de İngiltere havayolları ile uçmalarını özendirmek için özel promosyonlar sunmuş, bu uygulama ile özel müşterilerinin her iki yönde de İngiltere havayolları ile uçmalarında artış sağlamıştır [51].
- Gobena veri madenciliği tekniklerini kullanarak Ethiopian Havayolları ile uçanların kazanacakları puanları tahmin etmeye çalışmıştır. Bu çalışmada yapay sinir ağları, çok katmanlı perceptron, geriye yönelik ağlar ve radyal temelli fonksiyon mimarisi kullanılmıştır. Yapılan bu çalışmanın sonunda ortalama %33 – 37 hata oranına ulaşılmıştır. Sistemin daha fazla eğitilmesiyle daha başarılı sonuçlar elde edilebileceğine inanılmaktadır. Gobena, Ethiopian Havayollarında karar vermeye

destek olabilmek için veri madenciliği tekniklerinin daha fazla kullanılması gerektiğine inanmaktadır [51].

- Türkiye’de havacılık sektöründe yer alan bir işletmede parça söküm raporları ile ilgili veri madenciliği uygulamasında; herhangi bir problem yaşamadan önce uçaklardaki parçaların düzeltici ve önleyici işlemlerinde ikaz seviyelerini saptamak için kural geliştirilmiştir. Öncelikle regresyon analizi sonucunda birbirleriyle bağlantılı olan parametreler bulunmuş akabinde sınıflandırma analizi kullanılmış anlamlı bir kural elde edilmiştir [19].
- Eskişehir’de havayolu yolcu potansiyelinin tespit edilmesi amacıyla bir karar ağacı uygulamasında; havayolu yolcu taşımacılığında olan potansiyel yolcu talebinin ölçülmesi ve kentin sosyoekonomik gelişimine yarar sağlayacak çapraz havayolu uçuşlarının gerçekleşmesi için gerekli verileri karar ağacı modeli ile seyahat etme oranının en yüksek olduğu yolcu kesimi tespit edilmek istenilmiştir [45].

Literatürde incelenen çalışmalarda kabin ekiplerinin performans değerlendirmeleri belli istatistikî yöntemler ile saptanmıştır. Havacılık sektöründe uygulanan veri madenciliği yöntemleri bu çalışmaya ışık tutmuş ve kabin ekiplerinin mevcut kurumsal karnelerini veri madenciliği metodu kullanılarak anlamlı kurallar oluşturmak hedeflenmiştir.

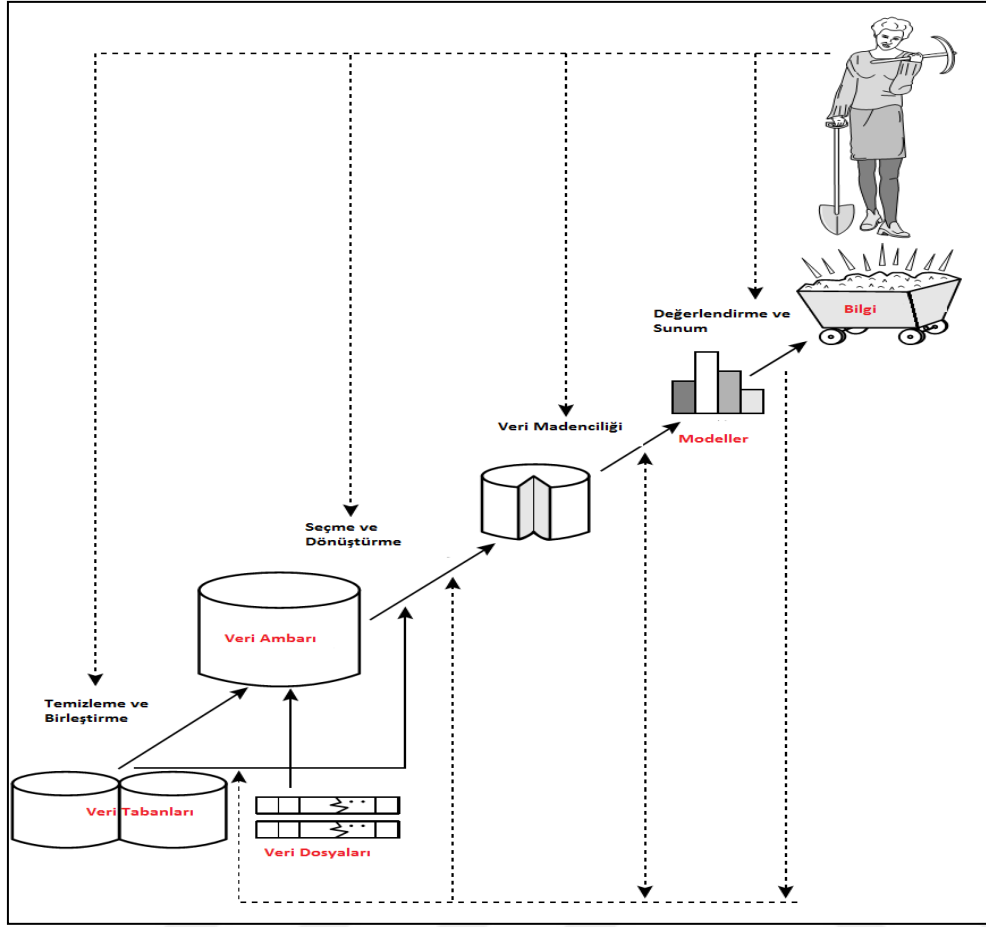
4. VERİ MADENCİLİĞİ

1960'lı yıllarda verileri bir araya getirmek ile başlayan veri tabanı teknolojilerinin gelişmesi, 1970'lerdeki veri tabanlarının meydana getirilmesi ile sürmüştür. 1980'lerde gelişen “İlişkisel Veri tabanı Yönetim Sistemi-RDBMS”, gelişmiş veri modelleri ile veri ambarları almıştır.

Veri tabanlarından bilgi keşfinin konusu ilk kez 1989 yılında A.B.D' nin Detroit Eyaleti'nde gerçekleştirilen 11. Uluslararası Müşterek Yapay Zekâ Konferansı'nda ele alınmıştır [20]. Fayyad ve arkadaşlarına göre veri tabanından bilgi keşfi yararlı bilginin keşfedilmesinin genel sürecini ifade ederken, veri madenciliği bu sürecin belirli bir kısmını temsil etmektedir [21]. Bu kısım örüntünün ortaya çıkarılabilmesi için belirli bir algoritmanın uygulanmasıdır. Fakat ilgili yayınlarda bu iki kavramın birbirinin yerine kullanılması daha yaygın bir anlayıştır. Bunun sebebi ise veri madenciliğinin, veri tabanından bilgi keşfinin en önemli kısmını oluşturmasıdır [22].

Prof. Dr. Usama Fayyad'a göre Veri Tabanlarında Bilgi Keşfi sürecindeki basamaklar aşağıdaki sıradadır [21];

1. Veri Seçimi: Veri kümeleri içerisinde sorguya yakışır verilerin seçilme işlemidir. Belirlenen veriler; örneklem küme olarak bilinmektedir.
2. Veri Temizleme- Ön işleme: Örneklem kümesinin belirlenmesinin akabinde, örneklem kümesi içerisinde bulunan noksan verilerin çıkarılıp ve bu nitelik değerlerinin değiştirilmesidir. Bu mertebede uygun bulunan veri madenciliği sorgusunun çalışma zamanı iyileştirilmektedir.
3. Veri Madenciliği: Yukarıda belirtilen işlemlerin akabinde örneklem kümesinde veri madenciliği algoritmalarının (kümeleme, sınıflandırma, ilişkilendirme gibi) uygulanmasıdır.
4. Yorumlama: Veri madenciliğinde uygulanan algoritmalarından meydana çıkan neticelerin yorumlama aşamasıdır. Bu basamakta geçerlilik, yenilik, yararlılık ve basitlik yönlerinden çıkan neticeler yorumlanmaktadır.



Şekil 4.1. Veri tabanlarında bilgi keşfi adımları [23]

Veri madenciliği, veri ambarlarında bulunan verilerden otomatikleşmiş modellere göre anlamlı kuralları, ilişkileri, davranışları elde etme gibi ayrıca bir tanımlama yapılamamaktadır. Bu aşamada, veri içinde daha önceden bilinmeyen veya muhakeme edilemeyen desenler (pattern) ilk olarak elde edilmektedir. Bu desenler çoğunlukla kurallar arasındaki ilişkilerin, sıralamanın, sınıflandırmanın, veri birlikteliğinin ve tahminlemenin neticesinde belirlenmektedir [24].

Literatürde veri madenciliği ile ilgili yapılmış tanımlamalardan bazıları aşağıda yer almaktadır:

- Veri madenciliği, büyük hacimli veri tabanları içerisindeki bilginin ortaya çıkarılması amacıyla makine öğrenmesi, örüntü tanımanın, istatistiki, görselleştirme gibi alanlarını entegre eden disiplinler arası bir alandır [25].
- Veri madenciliği, ham verinin yeterli olmadığı bilgiyi meydana getiren veri analizi sürecidir [26].

- Veri madenciliği, büyük hacimli veriler içerisindeki örüntüleri değerlendiren matematiksel algoritmaları kullanır. Veri madenciliği hipotezleri ortaya çıkarır, neticeleri entegre etmek amacıyla insani kapasiteden faydalanır. Veri madenciliği sadece bir bilim dalı değil, ayrıca bir sanat olarak da görülebilir [27].
- Veri madenciliği, büyük hacimli veri tabanlarındaki birliktelikleri araştırır [28].
- Veri madenciliği, istatistikî, makine öğrenimi gibi disiplinler ile etkileşim içinde olan yeni nesil disiplin ve çok boyutlu veri tabanları içerisinde önceden bilinmeyen örüntülerin analizi olarak tanımlanmaktadır [29].
- Veri madenciliği, mevcut veriler içerisinde açık olarak ifade edilmeyen, net olmayan, önceden muhakeme edilemeyen bilgilerden potansiyel yararlı bilginin elde edilmesidir [30].

Veri madenciliği, bilgi teknolojilerindeki gelişme ve küresel rekabet sayesinde gün geçtikçe büyüyen ve önemi daha da artan ya da bilinen bir alan olmaya başlamıştır. Bilimsel çevrelerde uzun yıllardır var olan ancak sektörel ilgiyi çok daha geç bulan bu alanda yapılan çalışmalar giderek çeşitlilik kazanmaktadır [31]. Son yıllarda bilgi endüstrilerinde veri madenciliğinin dikkat çekmesinin sebebi çok büyük miktarlardaki verinin elde edilebilirliği ve böyle verilerin yararlı bilgilere dönüştürülmesine duyulan ihtiyaçtır [2]. Elde edilen bilgi işletme yönetimi, üretim kontrol ve pazar analizi, mühendislik tasarımı ve bilimsel keşiflere kadar değişen uygulamalar için kullanılabilir.

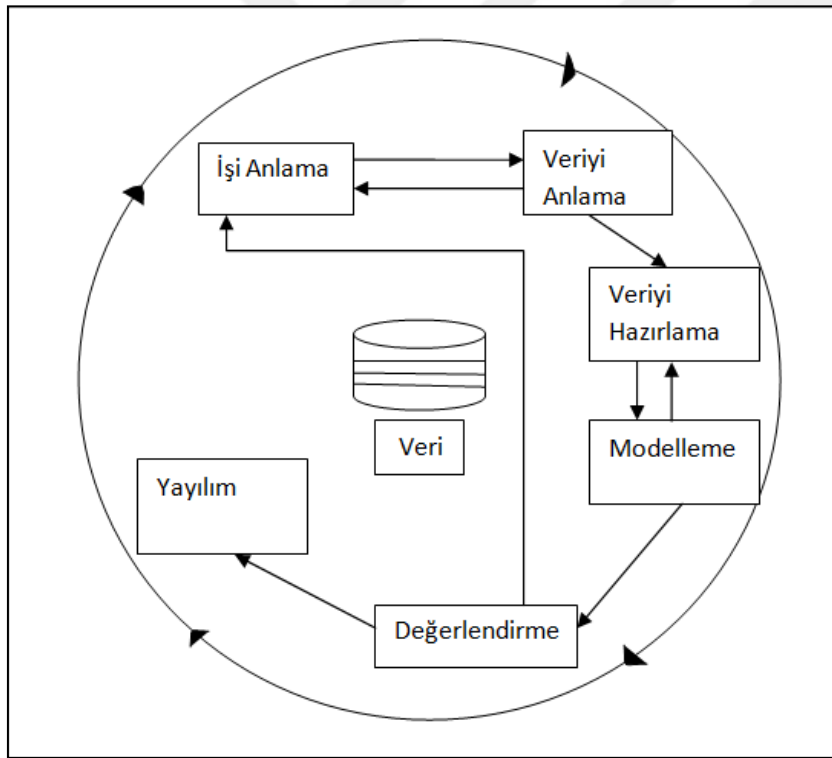
4.1. Veri Madenciliği Süreci

Veri tabanındaki bilgi keşfi süreci, bazı aşamalardan meydana gelen etkileşim içerisinde olan bir süreç olarak bilinmektedir. Bu süreç, uygulamadaki alanın öğrenildiği, uygulamanın amaçlarıyla hedef veri setinin belirlenmesiyle başlamaktadır. Akabinde, gürültülü, eksik ve ilgisiz verilerden elimine edilen veri temizleme ve ön işleme aşaması gelmektedir. Belli vakalarda veri; veri madenciliğine uyan bir forma dönüştürülmektedir. İleriki aşama olan veri madenciliği, algoritmalar doğrultusunda fazla hacimli veriden anlamlı bilgilerin çıkarılması sürecidir. Çıkarılan örüntüler, kullanışlı olan örüntülerin belirlenmesi için değerlendirilir. Veri tabanlarında bilgi keşfinin son adımı; ortaya çıkan bilgiyi görüntülemek ve bilgi gösterimi metodunu kullanılarak kullanıcıya arz edilmesidir [32].

Veri madenciliği süreci çok hızlı bir biçimde karmaşık hale gelebilir. Bu sebeple veri madenciliği için ilgili süreç şirketler birliği tarafından tanımlanmıştır. Çapraz Endüstri Veri Madenciliği Standart Süreci (CRISP-DM), 1996 Chrysler, SPSS ve NCR firmaları tarafından kurulmuştur [33].

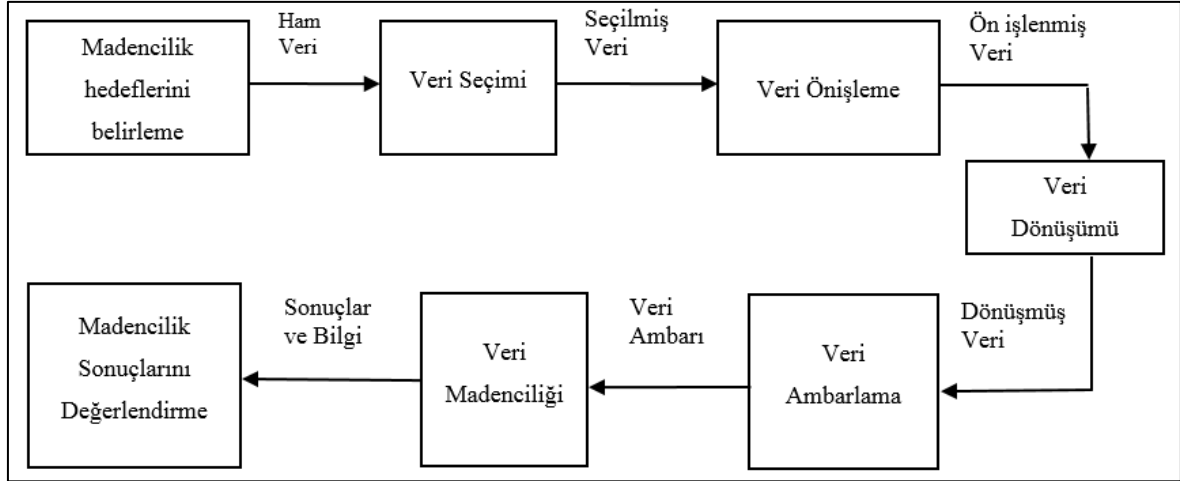
CRISP-DM süreci 6 aşamadan oluşmaktadır:

- Problemin Tanımlanması
- Verinin Hazırlanması
- Modelin Kurulması
- Modelin Değerlendirilmesi
- Modelin Kullanılması
- Modelin İzlenmesi



Şekil 4.2. CRISP-DM süreci [35]

Başka bir örnek veri madenciliği süreci Hui ve Jha tarafından Şekil 4.3'te ifade edilmiştir [36]:



Şekil 4.3. Hu ve Jha ‘nın veri madenciliği süreci [36]

4.1.1. Problemin tanımlanması

Veri madenciliğinde başarı sağlamanın öncelikli koşulu, çalışma amacının net olarak belirtilmesidir. Amaç, probleme odaklanan ve net ifade edilen olmalıdır. Meydana gelecek durumların başarı seviyelerinin nasıl ölçümleneceği belirlenmelidir [39].

Temel olarak aşağıdaki faaliyetlerin gerçekleşmesi beklenmektedir [37]

- İşe yönelik hedeflerin belirlenmesi
- Mevcut durumun değerlendirilmesi
- Veri madenciliği hedeflerinin belirlenmesi
- Proje planının yapılması

4.1.2. Verilerin hazırlanması (veri ön işleme)

Verinin hazırlanması aşaması dört adımdan meydana gelir. Bu adımlar:

- Verinin toplanması,
- Verinin birleştirilmesi,
- Verinin temizlenmesi,
- Verinin dönüştürülmesidir.

Veri hazırlanması sürecinde, veri toplamak en önemli basamaklardan birisini oluşturmaktadır. Bu adımda problemin amaçlara uygun olarak verilerin oluşturulması gerek veri hazırlanması adımını gerekse bütün veri madenciliği çevrimini direkt olarak

etkilemektedir. Veri birleştirme süreci, farklı kaynaklardan alınan verilerin aynı formatta olmasını gerektirmektedir. Böylece farklı kaynaklardan alınan verilerin tümüne aynı analizler uygulanabilmektedir.

Veri temizleme sürecinde; analizi yanlış yönlere sevk edebilecek olan noksan ya da aykırı verilerin veri madenciliği sürecine etkisini ortadan kaldırmak için, veri topluluğundan çıkarılması gerekmektedir. Veri dönüştürme süreci, verilerin farklı niteliklerini analize uygun olarak dönüştürmedir [40].

4.1.3. Modelin kurulması ve değerlendirilmesi

Bu adımda çeşitli modelleme teknikleri seçilir ve uygulanır. Ayrıca bu tekniklerin parametreleri en uygun değerlere ayarlanır. Veri hazırlama ve modelleme arasında sıkı bir ilişki vardır [41].

Veri madenciliği problemleri için birçok teknik kullanılabilmekte, probleme uygun olan yöntem veya yöntemlerin bulunması amacıyla birden fazla yöntem oluşturup uygun olanları seçilmektedir. Model meydana getirilmesinin akabinde uygulanan yöntemin ihtiyaçlarına göre veri hazırlama adımına tekrar dönüp ilgili değişiklikler yapılabilmektedir [42].

Bir model doğruluğu için bazı test yöntemleri kullanılmaktadır. Bunlardan en basit yöntem “basit geçerlilik testi” olarak bilinmektedir. İlgili testte; verilerden %5-%33 arası testte kullanılacak veriler olarak ayrılır ve verilerde test işlemi gerçekleştirilir. Bir sınıflama modeli içinde yanlış sınıflanmış olay sayısı, tüm olay sayısına bölüldüğünde hata oranı; doğru sınıflanmış olay sayısı tüm olay sayısına bölüldüğünde doğruluk oranı hesaplanması yapılmaktadır [42]. (Doğruluk Oranı = 1 – Hata Oranı)

Veri içeriğinin büyük hacimli olmadığı durumlarda, kullanılacak başka metot ise veri kümesinin rastgele iki tam eşit parçaya bölünmesiyle uygulanan çapraz geçerlilik (cross validation) testidir. Çapraz geçerlilik yöntemi uygulamalarında danışmanlı öğrenme analizleri için çoğunlukla yararlanılan bir metottur. Sıklıkla '10 parça çapraz geçerlilik' uygulaması kullanılmaktadır. Veri kümesinden rassal olarak 10 eşit parça ayrılmaktadır. Başlangıçta; birinci kısım test verisi olarak ayrılır, diğer dokuz kısım eğitim seti şeklinde kullanılmaktadır. Akabinde, ikinci ise kısım test veri seti olarak kullanılır, diğer dokuz kısım

ile model bulunmaya devam edilir. Süreç, ilk başta ayrılan 10 kısmın ayrı ayrı test seti halinde kullanılmasına kadar devam etmektedir. Çapraz geçerlilik sonucunda tüm kısımlardan meydana gelen doğru sınıfın ya da ROC (Receiver Operating Characteric) eğrisi altındaki alanın ortalaması alınır. Sonuç olarak kullanılan metodun performansını göstermektedir [34].

4.1.4. Modelin kullanılması

İlgili süreçlerin akabinde değerlendirilen ve geçerli olan modelin kullanılması evresidir.

4.1.5. Modelin izlenmesi

Süreç boyunca edinilen sağlıklı modeller, farklı amaçlarla çeşitli iş operasyonlarına uygulanabilir. Bu modeller koşullar değiştikçe kontrol edilmelidir. Çünkü bugün doğru olan sonuçlar bundan bir yıl sonra yine doğru olacağı anlamına gelmez. Eğer koşullarda önemli değişiklikler meydana gelirse model yeniden inşa edilmelidir. Sonuçları belgelendirmek ve kayıt altına almak gelecek adına akıllıca bir adım olacaktır [37].

4.2. Veri Madenciliği Yöntemleri

Veri madenciliği uygulamalarında yararlanılan modeller; tahmin edici (Predictive) ve tanımlayıcı (Descriptive) modeller halinde ayrılabilir. Tahmin edici modeller; neticeleri belli olan verilerden başlayarak bir model geliştirerek neticeleri bilinmeyen verilerin sonuçlarını öngörülmesi hedeflenmektedir. Tanımlayıcı modeller; karar vermeye öncülük etmede faydalananabilecek hali hazırda verilerdeki örüntülerin bilinmesidir. Veri madenciliği, sınıflandırma, kümeleme, kural çıkarımı, benzerlik tespiti ve tanımlama gibi fonksiyonları uygulamaktadır [46].

Sınıflama ve regresyon modelleri tahmin edici (predictive) modeller olarak bilinir. Kümeleme ve birliktelik kuralları modelleri ise tanımlayıcı (descriptive) modeller olarak nitelendirilirler [47]. Veri madenciliği modelleri yaptıkları işlemlere göre 3 temel başlık altında toplanabilir.

1. Sınıflama (Classification) ve Regresyon (Regression).
2. Kümeleme (Clustering).

3. Birliktelik Kuralları (Association Rules) ve Ardışık Zamanlı Örüntüler (Sequential Patterns).

4.2.1. Sınıflama ve regresyon

Sınıflama kategorik değerlerin tahmin edilmesinde kullanılırken regresyon süreklilik gösteren değerleri tahmin etmek için kullanılmaktadır. Sınıflama ve regresyon modelleri arasında bulunan farklılık, tahmin edilen bağımlı değişkenin kategorik ya da sürekli olup olmadığını gösteren durumdur. Bununla birlikte bazı tekniklerde her iki modelin gitgide birbirlerine yaklaşması; aynı tekniklerden faydalanılmasına sevk etmektedir [48].

Örneğin, sınıflama algoritmalarında bankadaki kredi uygulamalarının ya güvenli ya riskli olmalarını kategorize edebilmek için, regresyon algoritmasında ise gelir, meslek gibi nitelikleri belli olan mevcut müşterilerde bilgisayar ürün alımı yaparken harcama tahmini için kurulmaktadır [49].

Uygulamada kullanılmakta olan WEKA programı ile sınıflandırma algoritmaları ile analiz edilen model çıktıları, aşağıda bahsedilen değerlendirme kriterleri ile karşılaştırılmaktadır. 5. bölümde tez çalışmasının uygulama safhasında aşağıda yer alan değerlendirme kriterleri kullanılarak sınıflandırma algoritmalarının değerlendirilmesi yapılmaktadır.

- Number of Rules / Number of Leaves: Sınıflandırma algoritması tarafından çıkartılan kural sayısı veya oluşturulan karar ağacının boyutunu göstermektedir.
- True Positive Rate (TP Rate): Sınıflandırma algoritmasında herhangi bir sınıfa atanan kayıtlardan gerçekte o sınıfa ait olanların oranını yüzdesel olarak gösterir.
- False Positive Rate (FP): Sınıflandırma algoritmasında herhangi bir sınıfa atandığı halde gerçekte o sınıfa ait olmayan kayıtların oranı olarak gösterilmektedir.
- Precision: Gerçekte herhangi bir sınıfta mevcut olan kayıtların hangi oranda sınıflandırma algoritması tarafından ilgili sınıfa atandığı belirtir.
- Kappa İstatistiği: Tahmin doğruluğunun ölçüsüdür.
- Confusion Matrix: Doğru sınıflandırılması yapılan örneklerin sayısı bu matriste yer alan diyagonal elemanların toplamıdır.

Veri madenciliğinde sınıflama metotları kullanıldığında ve bu metotların performans kıyaslamasında; sınıflama doğruluğu önemli bir göstergedir. Sınıflama doğruluğunu doğru

sınıflanmış birey sayısının bütün birey sayısına oranı şeklinde tanımı yapılmaktadır. Uygulama alanın ve içeriğine göre ne kadar da değişiklik gösterse de %75 lik doğruluk kabul edilebilir seviyedir [34].

Duyarlılık ise gerçekte pozitiflerin arasındaki pozitif sonucun meydana gelme oranını; seçicilik (özgüllük) ise gerçekte negatiflerin negatif sonuç verme oranını belirtmektedir. Uygulanan yöntemlerin bir başka performanslarının karşılaştırıldığı metot ise ROC Eğrisi altında kalan alandır. Bu eğrinin altındaki alanın alabileceği maksimum değeri 1 olarak çıkarımı yapılmıştır. Eğrinin altındaki alan 1'e yaklaştıkça yeni testin ayırt edilme özelliği artmaktadır.

Bir veri kümesi içinde, dört olası sonuç mevcuttur; pozitif örneğin doğru sınıflandırılması Doğru Pozitif (DP), yanlış sınıflandırılması Yanlış Negatif (YN), Negatif örneğin doğru sınıflandırılması Doğru Negatif (DN) ve yanlış sınıflandırılması Yanlış Pozitif (YP) dir. Bu bağlamda hata matrisi çizelge 4.1 de belirtilmektedir [72].

$$\text{Doğru Pozitif Oranı} = \frac{DP}{DP + YN} \quad (4.1)$$

$$\text{Yanlış Pozitif Oranı} = \frac{YP}{YP + DN} \quad (4.2)$$

Çizelge 4.1. Hata matrisi

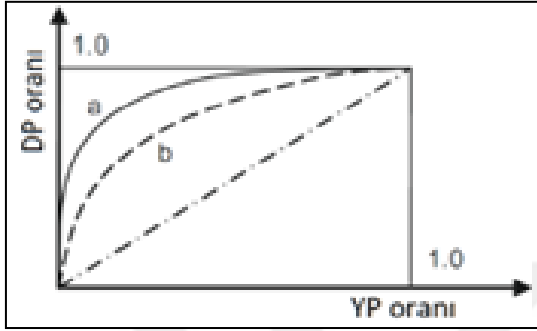
		Gerçek Sınıf	
		Pozitif	Negatif
Tahmin Edilen Sınıf	Pozitif	Doğru Pozitif (DP)	Yanlış Pozitif (YP)
	Negatif	Yanlış Negatif (YN)	Doğru Negatif (DN)

$$\text{Sınıflandırma Doğruluğu} = \frac{DP + DN}{DP + YP + DN + YN} \quad (4.3)$$

$$\text{Duyarlılık (Sensitivity)} = \frac{TP}{TP + FN} \quad (4.4)$$

$$\text{Kesinlik (Specificity)} = \frac{TP}{TP + FP} \quad (4.5)$$

$$\text{F-Ölçütü} = \frac{2 \times \text{Duyarlılık} \times \text{Kesinlik}}{\text{Duyarlılık} + \text{Kesinlik}} \quad (4.6)$$



Şekil 4.4. ROC eğrisi

4.3. Sınıflandırma Algoritmaları

Sınıflandırma (veya sınıflama) algoritmaları aşağıda sıralanmıştır [50]:

- Karar ağaçları
- NaiveBayes
- Diskriminant analizi
- Sinir ağları
- Kaba kümeler
- Genetik algoritmalar
- Regresyon analizi

Bu tez çalışmasında yukarıda belirtilen karar ağacı algoritmalarından faydalanılmaktadır. Diğer algoritmalarından kısaca bahsedilmektedir.

4.3.1. Karar ağaçları

Karar ağacı tekniğinin havayolu taşımacılığı alanında uygulamalarına bakıldığında birçok farklı konuda sınıflandırmaya dayalı karar problemlerinin çözümlenmesine olanak sağladığı görülmektedir [45]. Havayolu yolcularının değerlemesi ve pazar segmentasyonu [69],

havayolu yolcularının gelecekteki değerlerinin tahminlemesi [44] konularında karar ağacı algoritmaları kullanmıştır.

Karar ağaçları, uygulanmasının, yorumlanmasının ve entegrasyonunun kolaylığı nedeniyle en yaygın kullanılan sınıflandırma tekniklerinden biridir [55]. Bilgi teorisine dayanan karar ağaçları, karmaşık ve bilinmeyen verilerin kolay yorumlanmasını sağlayan modeller meydana getirmektedir. Bu bağlamda, bu tez çalışmasında kolay yorumlanabilmesi, tahmin açısından güçlü olması sebebi ile karar ağacı algoritmaları kullanılmaktadır.

Karar ağaçları genellikle yaprakları ve gövdesi ile ağaç yapısında sunulmaktadır. Gövdeler özelliklerin koşullarını gösterirken yapraklar sınıflandırma sonuçlarını ortaya koyar. Karar ağaçları halihazırdaki kullanımıyla en popüler tümevarım metodudur. Karar ağaçları genellikle iki aşamada oluşturulmaktadır. Büyüme olduğu zaman bu algoritma her bir düğümde sınıflar arasındaki en iyi özellikteki ayırt ediciyi (veri alt seti) ortaya çıkarmakta ve daha sonra o özelliğe dayalı olarak bu verileri iki yeni düğüm halinde bölümlere ayırmaktadır. Bu süreç, her bir tabaka için bir sınıf tahsis edilinceye kadar ortaya çıkan veri alt setine tekrar tekrar uygulanmaktadır. İndirgemenin ikinci aşaması en iyi dengeye erişebilmek için ağacın en az yararlı dallarını kesmek suretiyle işletilmektedir. Daha basit bir model genellikle daha sağlam olmaktadır. Bir başka deyişle, yeni veriler hakkında daha doğru sonuçlar ortaya çıkarmaktadır. Nihai ağaç etiketlenmiş birkaç bölgedeki bu özel alanı bölümlere ayırmaktadır [59].

Karar ağaçlarının en bilinen algoritmaları olan; CART, CHAID, ID3, C4.5 algoritmalarının özellikleri Çizelge 4.2 de yer almaktadır [56].

Çizelge 4.2. CART, CHAID, ID3 ve C4.5 karşılaştırması [56]

Algoritma	Yazar	Veri tipi	Ağaç Budama metodu	Her bir düğümdeki dal sayısı	Kayıp değer metodu	Bölünme kriteri
CHAID	Kass (1980)	Kesikli	Budama yok	İki veya daha fazla	Kayıp değer dallanması	Ki-Kare testi için P değeri
C&RT	Breiman ve ark. (1984)	Kesikli ve sürekli	Tüm hata oranı	İki	Sıralı/ yerine geçen bölünme (alternate/surrogate)	Gini değeri, entropi
ID3	Quinlan (1986)	Kesikli	Budama yok	İki veya daha fazla	Elde edilemeyen	Bilgi kazancı
C 4.5	Quinlan (1986)	Kesikli ve sürekli	Tahmini hata oranı	İki veya daha fazla	Olasılıklı ağırlık	Kazanç oranı

ID3 ve C4.5 algoritmaları

Karar ağaçları olarak da adlandırılan ID3 ve C4.5 algoritmaları, sınıflandırma modellerini işlemek için Quinlan tarafından geliştirilmiştir. ID3 tahmin edicileri ve onların bölünüm değerlerindeki sağlanan bilgi kazancına göre seçmektedir. Kazanç, bölünüm öncesi ve sonrasında doğru tahmin yapabilmek için ihtiyaç duyulan enformasyon miktarındaki farkı anlatmaktadır. Kazanç, orijinal segmentin entropisi ile sonuçta elde edilen bölünüm segmentlerinin toplam entropileri arasındaki farktır [3]. C4.5 algoritması ise ID3 algoritmasının geliştirilmiş halidir ki WEKA programında bu algoritma J48 olarak anılmaktadır.

C&RT

Breiman ve Friedman tarafından 1984 yılında geliştirilmiştir. C&RT ağacını oluşturmada tahmin ediciler, farklı tahminlerin kayıtları nasıl etkilediği temeline göre seçilmektedir. C&RT' nin büyük avantajlarından birisi algoritmanın modelin geçerliliğini sınaması ve genel optimal modelin keşfedilmesidir. C&RT çok karmaşık bir ağaç oluşturup geriye doğru budanarak çapraz geçerlilik sınaması ve test kümesi sınaması sonuçlarını temel alması sonucunda optimal genel ağaca ulaşmaktadır [27].

CHAID

Karar ağacı tekniği grupları oluşturmak için ki-kare analizinden yararlanmaktadır. Bu algoritma esasen benzer verileri birleştirerek farklı olanlar ile analize devam etmektedir. Böylece ulaşılmak istenen değerler kesin ve kolay bir şekilde elde edilmektedir. Karar ağacının ilk dallarını oluşturmak için en iyi tahmin edici değişkenler seçilmektedir. Bu algorithmada değişkenler sürekli ise F-testi değerleri; nominal ya da ordinal bir değişken ise ki-kare test değerleri kullanılmaktadır [57].

Sınıflandırma ağaçları, verideki en önemli karakteristikleri tanımlar ve amaç değişkenine ulaşmak için en iyi tahmini sağlayan özelliklerin kombinasyonunu belirler. Ağacı yapılandırmak için, öncelikle popülasyon birbirinden mümkün olduğu kadar farklı 2 alt popülasyona ayrılır. Bunu yaparken, her bir karakteristiğe bakılır ve amaç değişkeninin çıktılarının farklılaştıracak özelliklerin farklılaştırılması ile optimumu sağlayacak ayırım tanımlanır. Bu süreç, yavru popülasyonlara kadar tekrarlanır [58].

Karar ağacı yöntemlerinde parametrik istatistiksel varsayımlar yapılmaz. Öngörüler terminal düğümlerde birkaç mantıksal if-then şartıyla sunulabilirler. Normal bir veri dağılımında veya değişkenler ve tepki değişkeni arasındaki lineer ilişkilerde örtülü varsayımlar yoktur. Bu nedenle karar ağacı yöntemleri ilişkileri açığa çıkarabilmektedir [59].

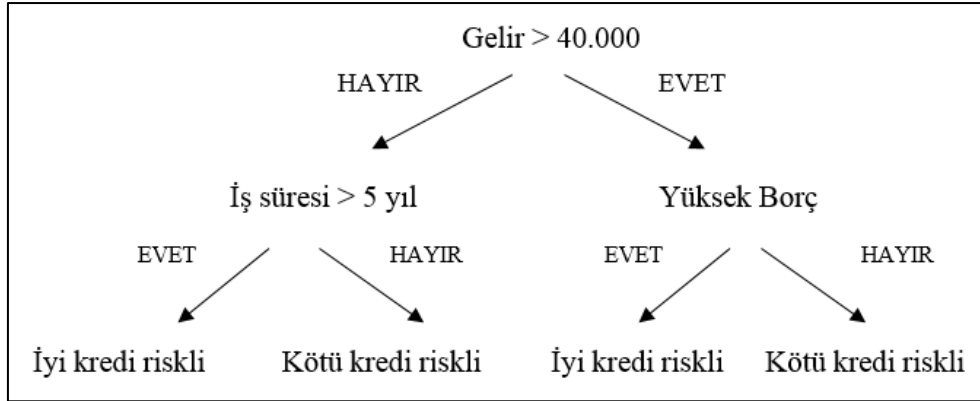
Bağımlı değişkendeki farklılıkları en üste çıkaracak şekilde veriyi sıralı bir biçimde farklı gruplara ayırmak karar ağaçlarının hedefidir. Karar ağacı, ağaç görünümünde olan bir teknik olarak bilinir. Aynı zamanda yukarıdan aşağıya doğru öğrenme sistemi olarak da bilinmektedir.

Karar düğümleri; dallar ve yapraklardan oluşmaktadır. Karar ağaçlarının yapısını oluşturan unsurlar [55]:

- Karar düğümü: Veri setine uygulanan testi tanımlar. Her düğüm bir özelliğe sahip testi belirtmektedir. Test neticesinde ağacın dalları meydana gelir. Veri kaybının yaşanmasını önlemek için tüm verinin dahil olacağı sayıda dal belirlenmelidir.
- Dal: Test sonucunu göstermektedir. Ortaya çıkan her dalın amacı; tanımlanacak sınıfın belirlenmesidir. Şayet, elde edilen dal sonucunda sınıflandırma belirlenemez ise karar düğümü tekrar oluşturulur.

- Yaprak: Dal; şayet bir sınıflandırmayı meydana getiriyor ise yaprak belirlenmiş olur. Yaprak, veriler ile istenen sınıflandırmanın; ortaya çıkardığı sınıflardan birini tanımlamaktadır.

Karar ağacı, kök düğümden aşağı doğru yaprak düğümlerinde sonlanıncaya kadar uzayan, dallar tarafından bağlantıları sağlanmış, karar düğümlerinin bir koleksiyonunu içerir. Karar ağacı diyagramının en üstüne yerleştirilmiş olan kök düğümden başlayarak, dalın her bir olası sonucu ile karar düğümleri test edilir. Her bir dal daha sonra diğer bir karar düğümü ya da sonlandırma yaprak düğümüne yol açmaktadır [2].



Şekil 4.5. Karar ağacı örneği

Bilgi kazancı ölçütü ağaçtaki her bir düğümdaki test alanını seçmek için kullanılmaktadır. Bu tür bir ölçüt alan, seçim ölçütü olarak anılmaktadır. En yüksek bilgi kazancı değerine sahip olan ele alınan düğüm için test alanı olarak seçilmektedir. Bu alan sonuç ayrımlarındaki örnekleri sınıflamak için gerekli olan bilgiyi en aza indirmektedir. Böyle bir teorik bilgi yaklaşımı bir nesneyi sınıflamada ihtiyaç duyulan beklenen test sayısını minimize etmekte ve basit bir ağacın bulunacağını garantilemektedir.

Eğitim verisinde yer alan hangi alanların, hangi sıra ile kullanılarak karar ağacının meydana getirileceği belirlenmelidir. Bu minvalde en sık kullanılan ölçüm Entropi ölçümü olarak bilinmektedir. Ortaya çıkan sonuçların belirli ve kararlı olması entropi ölçüsünün oranına bağlı olmaktadır. Entropi ölçüsünün fazla olması ortaya çıkan sonuçların o kadar kararsız olduğunu işaret etmektedir. Bu sebeple karar ağacının kökünde entropi ölçüsünün en az olduğu alanlar kullanılmaktadır [47].

S, s adet veri örneğini barındıran bir veri seti olduğu; sınıf etiketi alanının m adet farklı $C_i (i=1,2,...,m)$ sınıfı tanımlayan m farklı değere sahip olduğunu ve s_i 'nin C_i sınıfında S'nin örneklerinin sayısı olduğu farz edilmektedir. Verilen örneği sınıflamak için ihtiyaç duyulacak beklenen bilgi aşağıda verilmektedir.

$$\text{Beklenen Bilgi: } I(s_1, s_2, \dots, s_m)^n = - \sum_{i=1}^m p_i \log_2^{p_i} \quad (4.7)$$

Burada p_i keyfi bir örneğin C_i sınıfına ait olması olasılığıdır ve s_i/s ile tahmin edilir.

A alanı k farklı değere sahip olduğu varsayılmaktadır $\{a_1, a_2, \dots, a_k\}$. Belirli bir A alanının entropi ölçüsünü veren formüller aşağıda yer almaktadır [47]:

$E(C \setminus A)$ = Sınıflama özelliğine sahip A alanının Entropi ölçüsünü,

$p(a_k, j)$ = a_k 'nin j değerinde olma olasılığını,

$p(c_i \setminus a_k, j)$ = a_k 'nin j. Değerinde olması durumunda sınıf değerinin c_i olma olasılığını,

M_k = a_k 'nin değerlerinin sayısının; $j=1,2,\dots, M_k$, olmasını

N = farklı sınıfların sayısının; $i=1,2,\dots, N$, olması

k = alanların sayısının ; $k=1,2,\dots, k$, olmasını göstermektedir.

$$\text{Entropi: } E(A) = \sum_{j=1}^k \frac{s_{1j} + \dots + s_{mj}}{s} I(s_{1j} + \dots + s_{mj}) \quad (4.8)$$

4.3.2. Diskriminant analizi

Bu yöntem matematiksel açıdan değerlendirilen en eski sınıflandırma tekniği olarak görülmektedir. Sonuçları yorumlamak kolaydır. Tıp, sosyal bilimler ve saha biyolojisinde yaygın olarak kullanılmaktadır [33].

Amaç olarak, çok değişkenli problemin tek değişkenli biçime dönüştürülmesidir. Başka bir deyişle, tüm değişkenlerin uygun ağırlıklarla bulunduğu tek bir değişkenin (fonksiyon) elde edilmesidir [52].

4.3.3. Naive Bayes

Naive Bayes, modelin öğrenilmesi sırasında, her çıktının öğrenme kümesinde kaç defa oluştuğunu matematiksel olarak ortaya çıkarmaktadır. Ortaya çıkan sonuç, öncelikli olasılık olarak tanımlanır. Naive Bayes ayrıca her bağımsız değişken ya da bağımlı değişken kombinasyonunun ortaya çıkma frekansını bulmaktadır. İlgili frekanslar; öncelikli olasılıklar ile entegre edilerek tahminde kullanılmaktadır [49].

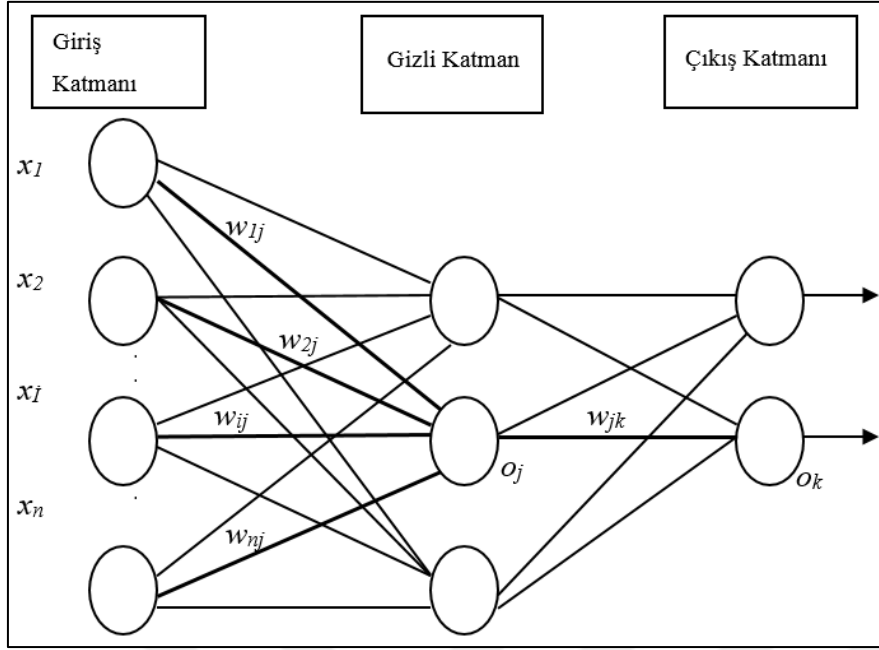
Naive bayes sınıflandırıcısı test verisi üzerinden öğrenme işlemini gerçekleştirir ve en yüksek orandaki örneğini sınıfa dahil eder. C 'nin bir sınıfı gösterdiği düşünölsün. $x = \langle x_1, x_2, x_3, \dots, x_m \rangle$ değeri ise gözlemlenen niteliklerin değeri olduđu, c bilinen bir sınıf etiketini temsil ettiđi farz edilmektedir. Nihai olarak en yüksek olasılık ile sınıf tahminin yapılması akabinde [53];

$$p(C = c_j \mid X = x) = p(C = c_j) \prod_{i=1}^m p(X_i = x_i \mid C = c_j) \quad (4.9)$$

Eşitlikte (4.9) belirtilen denklemi kullanarak test örneklerini hesaplamak ve öğrenme verilerinden tahminde bulunmak daha kolaydır [54].

4.3.4. Sinir ağıları

Genel bir anlamda yapay sinir ağlarının tanımı olarak; insan beyninin çalışma ve düşünebilme yeteneğinden yola çıkılmış ve bu şekilde oluşturulmuş bir bilgi işlem teknolojisidir [60]. İnsan beyninin hesaplama mantığından yola çıkılan (yapay) sinir ağları, karar ağaçları gibi yeni nesil yöntemlerden biridir. Yapay sinir ağları insan beynindeki sinir hücrelerinin işlevini modelleyen bir yapıdır ve birbiri ile bağlantılı katmanlardan oluşur [55].



Şekil 4.6. Yapay sinir ağı modeli

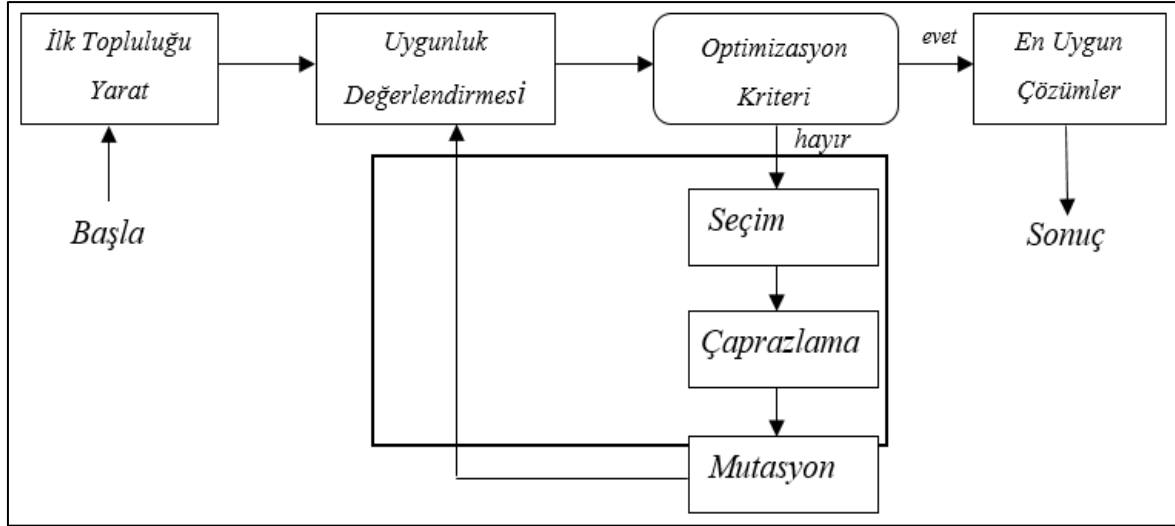
4.3.5. Kaba kümeler

Pawlak tarafından geliştirilen bu algorithmada; yaklaşırtma uzayı ile kümeye ait alt ve üst yaklaşırtmalar bulunmaktadır. Yaklaşırtma uzayı, odak alınan bölgeyi farklı biçimlerde sınıflandırmaktadır. Belli bir alt kümeye ait olduğu net bir şekilde bilinen nesnelerin tanımı; alt yaklaşırtma olarak bilinir. Üst yaklaşırtma; alt kümeye ait olması kesin olarak değil; ihtimaller dahilinde belirlenmiş olan nesneler tanımı olarak bilinmektedir. Alt ve üst sınırların arasındaki tanımlı nesne kaba kümedir [50]. Veri madenciliği kapsamında kaba küme teorisi temel alınarak çok sayıda algoritma geliştirilmiştir. Verinin indirgenmesi, niteliklerin minimal alt kümelerinin bulunması, kuralların en iyi minimal kümesinin hesaplanması, uygulamalarıyla kullanılmakta olan bu algorithmalar son yıllarda en fazla verinin sınıflandırılması amacıyla kullanılmaktadır [61].

4.3.6. Genetik algoritma

Genetik algorithmalar problemlerin çözümü için evrim sürecinin bilgisayar ortamına uyarlanmış biçimidir. Genetik algorithmaların kullanıldığı veri madenciliği uygulamalarında; çözüm uzayının hepsi yerine belirli bir kısmını incelediği için diğer algorithmalara göre daha hızlı çalışmaktadır. İşlemlerde mali kayıpların araştırılması, finansal opsiyonların geliştirilmesi gibi konularda da kullanılmaktadır [62]. Diğer eniyileme yöntemlerindeki

gibi çözüme ait bir yapı geliştirmek yerine, bu yapılardan elde edilen bir küme meydana getirilir. Problem için olabilecek kadar fazla sayıda çözüm gösteren kümeye popülasyon adını verilmektedir. Popülasyon vektör, kromozom veya birey adı verilen sayı dizilerinden oluşmaktadır. Her kuşakta, genetik algoritma tarafından kullanılan; çaprazlama, mutasyon gibi araçlar ile yeni bir popülasyon oluşturur [62].



Şekil 4.7. Genetik algoritma blok diyagramı

4.3.7. Regresyon analizi

Veriler arasındaki korelasyonu bulmak amacıyla sık gidilen yaklaşımlardan birisi lineer regresyondur [63]. Regresyon teriminde, tahmin edilecek değişkene “bağımlı değişken”; bağımlı değişkeni tahmin etmek amacıyla kullanılan değişkene veya değişkenlere ise “bağımsız değişken” olarak isimlendirilir [64].

Regresyon analizindeki amaç bir veya daha fazla bağımsız değişkeni bağımlı değişken ile ilişkili regresyon modeli veya tahmin denklemi kurmaktır. Model, ilgili değişkeni bağımsız değişken temelinde tanımlamak, tahmin ve kontrol amacıyla kullanılabilir [65].

4.4. Kümeleme

Kümeleme, veri tabanından ilginç örüntülerin keşfedildiği bir madencilik tekniğidir. Kümelemenin genel düşüncesi, veri tabanını çok sayıda kümeye ayırmak ve aynı kümeye ait verilerin mümkün olduğu kadar yakın ilişkide olmalarının sağlanmasıdır [66]. Sınıflama

ve kümeleme arasındaki en belirgin fark; kümeleme işlemi, sınıflama işlemine benzer şekilde belirlenmiş birtakım sınıflar doğrultusunda bölme yapmaması olarak gösterilir. Sınıflamada her bir veri; önceden bilinen çıktılar sonucunda denetimli olarak öğrenim gerçekleştirmektedir. Buna rağmen, kümeleme işleminde tanımlanmış sınıflar mevcut değildir [67].

Kümelemenin uygulama alanı çok geniştir. İş zekâsı, görüntü biçimi algılama, web aramaları, biyoloji, güvenlik gibi. Örneğin iş zekâsı uygulamalarında kümeleme, çok sayıda müşteriyi benzer karakteristik özellikler gösteren müşterileri bir arada toplamak suretiyle organize etmek için kullanılabilir [2].

Kümeleme algoritmasının seçimi, veri tipine ve çalışma amacına bağlı olarak uygulanmaktadır. Kümeleme yöntemleri aşağıda yer alan metotlar halinde sınıflandırılır [77].

- Bölümleme yöntemleri (Partitioning methods): Genel anlamda yaygın kullanılmakta olan metotlar; k-means yöntemi, k-medoids yöntemi ve bu yöntemleri varyasyonları olarak bilinmektedir.
- Hiyerarşik yöntemler (Hierarchical methods): Nesnelerdeki yakın olma ilişkisine göre belirlenen kümelerden bir ağaç meydana gelir.
- Model tabanlı yöntemler (Model-based methods)
- Yoğunluk Tabanlı Yöntemler (Density-based methods)
- Izgara Tabanlı Yöntemler (Grid-Based methods)

4.5. Birliktelik Kuralları ve Ardışık Zamanlı Örüntüler

Hareketli veri tabanlarındaki nesneler seti arasındaki yaygın örüntüleri, birliktelikleri, korelasyonları ve tesadüfi yapıları bulmak için kullanılan tekniklerdir. İlk olarak 1993 yılında uygulanmaya başlanmış olan birliktelik kuralları çok geniş bir araştırma sahasına ulaşmıştır [71]. Birliktelik kurallarının kullanıldığı en yaygın ve en bilinen örnek uygulaması olarak market sepeti bilinmektedir. Müşterilerin yaptıkları alışverişlerde satın alınan ürünlerin arasında yer alan birliktelik durumlarını tespit ederek satın alma eğilimlerinin ortaya çıkarılmasıdır. Birliktelik durumlarının açığa çıkarılması, müşterilerin hangi ürünleri hangi ürünler ile satın almış oldukları tespit edilir ve işletme yönetimleri tarafından değerli bilgiye odaklanarak etkin satış stratejileri geliştirebilirler [72]. Örneğin, bir alışveriş

merkezinde, ekmek alan müşterilerin 80%'i süt de almaktadır. İlişki kuralları algoritmasını uygulamanın asıl amacı verilerin analizi ile eş zamanlı ilişkileri ortaya çıkarmak ve karar verirken referans olarak kullanmaktır [66].

Markette mevcut olan bütün ürünlerin kümesini evren olarak düşünüldüğünde, her ürünün varlığını ya da yokluğunu gösteren boolean bir değişkeni olması gerekmektedir. Bu sayede her sepeti boolean değerlerden oluşan vektör şeklinde gösterebilmektedir. İlgili vektörlerden edinilen örnekler hangi ürünlerin birlikte satın alındığını meydana çıkarmaktadır. Bu örnekler ilişkisel kurallar formunda gösterilmektedir [78].

Örneğin bir A ürünü satıldığında B ürünü de onunla birlikte satılıyorsa; koşul aşağıda yer alan birliktelik kuralı ile belirtilir [74]:

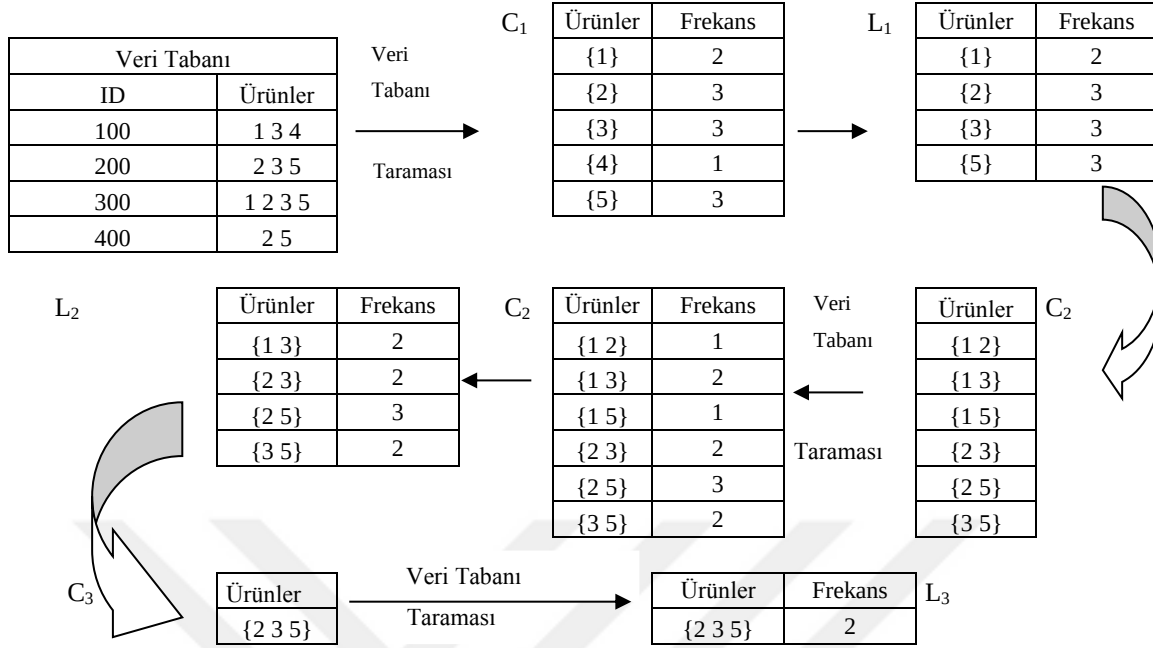
$$A \Rightarrow B \text{ [destek} = \%2, \text{güven} = \%60] \quad (4.10)$$

Yukarıda yer alan bağıntıdaki destek ve güven terimleri, sırayla belirtilen kuralın kullanılabilirliğini ve doğruluğunu işaret etmektedirler. İlgili kural için %2 oranında destek değeri yani analiz edilmiş olan bütün alışverişler içinde %2 oranında A ve B ürünlerinin bir arada satın alındığını gösterir. %60 oranını oluşturan güven değeri ise; satılan A ürününün %60'ında (aynı alışverişte) B ürününün de satıldığını meydana çıkarır.

Veri tabanları için birliktelik kurallarını ortaya çıkarmak için aşağıda yer alan 2 adım yapılmaktadır:

1. Veri tabanı içerisinde öğelerden her birinin önceden belirli olan destek sayısı kadar sık tekrarlanmış olması.
2. En az destek ve güven değerlerinin karşılandığı sık tekrar eden öğelerden birliktelik kurallarının oluşturulması [74].

Birliktelik kuralı için en çok bilinen algoritma aprioridir. Apriori de kullanıcı minimum destek eşikini verir ve algoritma bu eşik değerinden büyük tüm veri kümesini aramaktadır. İkinci adımda, ilk adımda bulunan veri kümelerinden kurallar oluşturulmaktadır. Algoritma her kural için güven değerini hesaplamakta ve kullanıcı tarafından tanımlanan güven eşik değerini aşan kuralları saklamaktadır. Uygulamada görülen en önemli problemlerden biri ise destek ve güven eşiklerinin belirlenmesi olmaktadır [75].



Şekil 4.8. Apriori algoritması

4.6. Veri Madenciliğinde Karşılaşılan Problemler

Boyutu büyük olan verilerin yer aldığı ortamlarda önemli problemler meydana çıkabilmektedir. Bu sebeple küçük veri kümelerinde, benzetim ortamında sağlanmış sistemlerde; eksik, gürültülü, aykırı veya belirsiz veri kümelerinde ortaya çıkan sonuçlar yanlış durumlara sevk edebilir. Bundan dolayı ilgili sistemlere zemin hazırlarken bahse konu olan problemlerin ortadan kaldırılması gerekir. Veri madenciliği uygulamalarında meydana çıkan problemler: [76]

1. Artık veri: Nihai sonucu elde etmek için faydalanan veri kümesindeki gerek olmayan nitelikler olarak bilinmektedir. Devam eden işlemler sırasında çıkabilmektedir.
2. Belirsizlik: Yanlış olan durumların şiddet ve veride yer alan gürültünün derecesi ile bağlantılıdır.
3. Boş veri: Veri kümesinde; kendisinin de dahil edilebileceği hiçbir değere eşit olmayan ve yer almayan birimdir.
4. Eksik veri: Veri kümesinin büyük hacimli olmasından kaynaklanmaktadır. Eksik veriler
 - Eksik verinin yer aldığı kayıtlar veri tabanından çıkarılabilmektedir.
 - İlgili değişkenin tüm kayıtlardaki ortalaması eksik olan verinin yerine kullanılabilmektedir.

- Mevcut verilere göre uygun deęer kullanılabilir.
- 5. Farklı tipteki verileri ele alma: Yalnızca sembolik ya da kategorik veri türleri deęil gerçek hayatta yer alan uygulamalarda tamsayı, kesirli sayı, coęrafi bilgi içeren veri gibi farklı nitelięe sahip verilerde de uygulanmasını gerektirmektedir.
- 6. Gürültü ve kayıp deęerler: Verinin girişinde veya verinin toplanmasında meydana gelen sistem dışında olan hatalar gürültü deęer olarak tanımlanmaktadır. Verinin toplanmasında meydana gelen hatalar; ölçüm hatalarını da içerebilmektedir. Hatalar yüzünden niteliklerin deęeri yanlış olabilir ve bu yanlışlarda veri madencilięinin amacına ve problemin çözümüne ulaşmasına engel olabilmektedir.
- 7. Sınırlı bilgi: Veri tabanları çoęunlukla veri madencilięi uygulamaları için deęil bunun yerine temel öğrenme niteliklerini sunma amacıyla hazırlanmışlardır. Bu sebeple, öğrenme veri madencilięi uygulamalarını kolaylaştıracak bazı nitelikler mevcut olmayabilmektedir.
- 8. Veri tabanı boyutu: Gün geçtikçe artan bilgi birikimi ile orantılı olan veri tabanlarının hacimleri de hızla artış göstermektedir. Veri tabanı algoritmaları birçok küçük örnekleme uygulanacak şekilde geliştirilirken aynı algoritmaların çok daha büyük kullanılabilmesi için çok dikkat gerekmektedir.

4.7. Veri Madencilięi Uygulama Alanları

Günümüzde gittikçe yoğunlaşan veri karşısında istenilen öz bilginin çıkarım konusuna gereksinim duyulan her yerde veri madencilięi yöntemleri kullanılabilir. Aşağıda yer alan Çizelge 3.2’de CRM ve saęlık/İK olarak belirtilmiş oranlardan yola çıkarak bu tez çalışmasında veri madencilięi yöntemlerinin saęlıklı bir şekilde kullanılabileceęi görülmektedir. Bu tez çalışmasında kullanılan yetkinliklerin içerięi düşünöldüğünde özellikle çizelgeye göre CRM konusunun bir hayli uygulama alanında olduęu görülmektedir.

Son yıllarda veri madencilięi uygulama alanları Çizelge 4.3’de yer almaktadır [78].

Çizelge 4.3. Veri Madenciliği uygulama alanları ve oranları

Bankacılık (51)	12%
Bioteknoloji / Genetik (11)	3%
Kredi skorlama (35)	8%
CRM (52)	12%
Doğrudan pazarlama (34)	8%
E-Ticaret (11)	3%
Eğlence/ Müzik (4)	1%
Sahtekarlık tespiti (31)	7%
Şans oyunu (2)	0,01 %
Kamu uygulamaları (12)	3%
Sigortacılık (24)	6%
Yatırım / Hisse senedi (5)	1%
Junkemail / Anti-spam (5)	1%
Sağlık/ İK (15)	4%
İmalat (19)	5%
Tıp/ Farmakoloji (12)	3%
Perakende (25)	6%
Bilim (17)	4%
Güvenlik / Anti-terörizm(5)	1%
Telekomünikasyon (23)	5%
Seyahat (8)	2%
Web (9)	2%
Diğer (11)	3%



5. KABİN MEMURLARININ PERFORMANS DEĞERLENDİRMESİNE İLİŞKİN BİR VERİ MADENCİLİĞİ UYGULAMASI

Wright Kardeşler' in ilk hava taşıtını yapmalarından itibaren havacılık sektöründe devamlı ilerlemeler görülmüştür. Her daim gelişen teknoloji ve çevre koşullarının da ayrıca dahil edildiği bu ilerlemede günümüz havacılık sektöründe bir rekabet oluşmaktadır. Havayolu işletmeleri, ayrıcalığı birçok alanda yakalamaya çalışmaktadır.

Bu tez çalışmasının kapsamında; bir havayolu şirketinde 2015 yılında üçer aylık dönemlerde kabin memurlarının performans değerlendirmelerini; yıllık bazda değerlendirerek mesleki yetkinlik ve demografik verilerinin karar ağacı algoritmaları uygulanarak anlamlı kurallar elde edilmeye çalışılmıştır.

Kabin memurlarının performans değerlendirmelerinde; demografik verileri, mesleki yetkinliklere bağlı değerlendirme sonuçları ve yıllık döneme bağlı karne sonuçları yer almaktadır.

5.1. Problemin Tanımlanması ve Amacın Belirlenmesi

Yapılan araştırmalar için havayolu işletmelerinde rekabeti sağlamak ve müşteri memnuniyetini arttırmak için öncelikle emniyet ve hizmet faktörleri ile çalışılmıştır. Bu faktörlerin devamını, uzmanlık ve esneklik, avantaj ve uygunluk, güvenlik ve kalite, rahatlık ve çeşitlilik gibi belli temel kategoriler üzerinde çalışmalar sürdürülmüş ve rekabette farklılık göstermek hedeflenmiştir.

Globalleşen rekabet ortamında kabin ekiplerinin performanslarını iyileştirme hedefi; gerçekleştirilen değerlendirmeleri etkin ve verimli kılmaya sevk etmiştir.

Gerçekleştirilen tarifeli dış hat seferleri sonrası kabin amirleri tarafından 6 mesleki yetkinlik ve bu yetkinliklerden meydana gelen 10 alt yetkinliğe göre kabin memurları değerlendirilmektedir.

Değerlendirmeler neticesinde harf bazında, istatistiki hesaplamalar doğrultusunda kabin memurlarının puanları belirlenmektedir. Bu puanlar harf bazında her 3 ayda bir kurumsal

karne ve akabinde yıl sonunda yıllık kurumsal karne düzeyi olarak hali hazırda yayımlanmaktadır.

Yıllık yayımlanan karne düzeylerinin; hangi alt yetkinliğe ve hangi demografik özelliğe bağlı olduğu sınıflandırma algoritmalarından karar ağaçları yöntemleri ile belirlenmiştir. Bu sayede kabin memurlarının karne düzeylerini etkileyen yetkinliklerin geliştirilmesi ve akabinde kabin memurlarının performanslarında iyileşme gözlenmesi ile global rekabet ortamında; müşteri ile bire bir iletişimde olan kabin memuru faktörünü yükseltmek amaç olarak saptanmıştır.

Çalışmanın 3. bölümde bahsedilen CRISP-DM süreci aşama olarak aktarılacak ve bulgular paylaşılacaktır.

5.1.1. İşi kavrama

Havacılık sektöründe müşteri memnuniyeti her zaman yadsınamaz bir hedef olmuştur. Müşteri memnuniyetinde kabin ekiplerinin önemi oldukça fazladır. Kabin ekiplerinin; tutum ve davranışları, iyi hizmet sunma, etkili iletişim, emniyet ve güvenlik tedbirlerini uygulama ile birlikte müşteriye dokunma ve marka değerini gösterme noktasında payı büyüktür. Bu çalışmada kabin ekiplerinin performans değerlendirmeleri yetkinlikler bazında, işletmenin misyon ve vizyon göstergeleri doğrultusunda gerçekleşmektedir. Son yıllarda verilerin git gide artış göstermesi ile birlikte öz bilgiye ulaşım konusunu kolaylaştıran veri madenciliği uygulaması kullanılmıştır. Bu yöntem ile geliştirilen algoritmalar neticesinde kabin ekiplerinin performans değerlendirmelerinde anlamlı kurallar bulunması amaçlanmıştır. Özetle, performans değerlendirmelerinin hangi değişkenlere bağlı olduğunu gösteren kusursuz bir model geliştirmek hedeflenmiştir.

5.1.2. Veriyi kavrama

Verilen toplanması safhasında veri tabanında gerçekleşen işlemler özetlenmiştir. Veri madenciliği sürecinin en zaman alıcı adımlarından birisi veri temizleme ve ön işleme süreci olmasından dolayı tez kapsamında kullanılacak verilere karar verildikten sonra elde edilen veri üzerinde temizleme ve hazırlama süreci üzerinde durulmuştur.

Verilerin temin edildiği kaynak güvenli bir kaynak olmasının avantajı ile kapsamlı bir şekilde veriler analiz edilmiştir. Değişkenler demografik ve yetkinlik bazda olmasından dolayı 18 adet değişkenin anlamı ve açıklamaları irdelenmiştir.

1. Kabin memuru numarası: Kabin memurlarına ait gizlilik prensibinin korunması amacıyla; her bir kabin memurunun numarası 1'den başlatılarak sırası ile 1 arttırılarak nümerik hale getirilmiştir. Bu sayede çalışmada kullanılan kabin memuru verilerinin kime ait olduğunun bilinmesinin önüne geçilmiştir.
2. Sınıf: Kabin memurlarının tecrübe ve performanslarına göre sınıf değişiklikleri yapılmaktadır.
 - Z sınıfı kabin memuru: Tecrübeli dar gövde uçaklarda görevli kabin memuru.
 - E / Y sınıfı: Tecrübeli geniş gövde uçaklarda görevli kabin memuru.
 - Q sınıfı: Business sınıfında tecrübeli kabin memuru.
 - N sınıfı: Mahalli dil bilen kabin memuru.
3. Tecrübe: Kabin memurlarının işe başladıkları tarihten itibaren yıl olarak tecrübe süresi.
4. Yaş: Kabin memurlarının yaşı.
5. Cinsiyet: Kabin memurlarının cinsiyeti.
6. Medeni Hal: Kabin memurlarının medeni durumu.
7. Eğitim: Kabin memurlarının öğrenim durumu.
8. KF1 (Kurumsal Farkındalık_1): Kabin memurlarının performans değerlendirmelerinde yetkinlik ölçütlerinden biri olup “Davranış konusunda ve havayolumuzu temsil etme açısından örnek bir kabin memurudur” kriterinin ölçümlendiği değişkendir.
9. KF2 (Kurumsal Farkındalık_2): Kabin memurlarının performans değerlendirmelerinde yetkinlik ölçütlerinden biri olup “Görüntü ve kişisel bakım açısından örnek bir kabin memurudur” kriterinin ölçümlendiği değişkendir.
10. EI1 (Etkili İletişim_1): Kabin memurlarının performans değerlendirmelerinde yetkinlik ölçütlerinden biri olup “Ekip arkadaşlarıyla ve yolcularımızla doğru iletişim kurdu” kriterinin ölçümlendiği değişkendir.
11. EI2 (Etkili İletişim_2): Kabin memurlarının performans değerlendirmelerinde yetkinlik ölçütlerinden biri olup “Yolcularımızı güler yüzle karşıladı, uçuş boyunca güler yüzle hizmet etti” kriterinin ölçümlendiği değişkendir.
12. AP (Analiz Etme ve Problem Çözme): Kabin memurlarının performans değerlendirmelerinde yetkinlik ölçütlerinden biri olup “Görev esnasında karşılaştığı

problemleri çözebildi. Görevi boyunca yolcularımıza alternatif sundu.” kriterinin ölçümlendiği değişkendir.

13. SO (Sürekli Öğrenme ve Kişisel Gelişim): Kabin memurlarının performans değerlendirmelerinde yetkinlik ölçütlerinden biri olup “Öğrenmeye ve eleştiriye açık bir kabin memurudur.” kriterinin ölçümlendiği değişkendir.
14. TR1 (Takım Ruhu ve Dayanışma_1): Kabin memurlarının performans değerlendirmelerinde yetkinlik ölçütlerinden biri olup “Ekiple uyumlu çalıştı, görev alanının dışında da ekip arkadaşlarına yardımcı oldu” kriterinin ölçümlendiği değişkendir.
15. TR2 (Takım Ruhu ve Dayanışma_2): Kabin memurlarının performans değerlendirmelerinde yetkinlik ölçütlerinden biri olup “Sorumluluk bilinciyle görevlerini tam ve zamanında yerine getirdi” kriterinin ölçümlendiği değişkendir.
16. MO1 (Müşteri Odaklılık_1): Kabin memurlarının performans değerlendirmelerinde yetkinlik ölçütlerinden biri olup “Ürün, hizmet ve hat bilgilerine hâkim, yolcularımıza doğru ve zamanında hizmet sundu” kriterinin ölçümlendiği değişkendir.
17. MO2 (Müşteri Odaklılık_2): Kabin memurlarının performans değerlendirmelerinde yetkinlik ölçütlerinden biri olup “Yolcularımızın talebi olmadan hizmet sunmayı teklif edebildi. Servis öncesinde ve sonrasında yolcularımızın taleplerinin karşılanabilmesi için ulaşılabilir durumdaydı.” kriterinin ölçümlendiği değişkendir.
18. Yıllık karne düzeyi: 2015 yılı kabin memurlarının performans değerlendirmelerinin sunulduğu karne düzeyidir. Hedef değişken (bağımlı değişken) olarak tanımlanacaktır.

5.1.3. Veriyi hazırlama

2015 yılı Ocak-Aralık ayları için 18 farklı değişken için 3764 x 18’lik bir veri kümesi elde edilmiştir.

Her üç aylık dönemde; 10 alt yetkinliğe ait kabin memurlarının puanları belirlenmektedir. Çalışmada ilk olarak her üçer aylık dönemde belirli olan 10 alt yetkinliğin puanları yıllık bazda ortalama olarak hesaplanmıştır.

Her mesleki yetkinlik kriterinde her bir dönem için en az 1 en fazla 5 puan üzerinden değerlendirme yapılmaktadır. 2015 yılı boyunca 4 adet dönem bulunmasından dolayı her dönemde her kriter için yapılan puanların ortalamaları alınmıştır.

Her bir mesleki yetkinliğin; yıl boyunca toplam 4 dönemdeki puanları ortalama olarak hesaplanmıştır. Bir başka deyişle, her bir mesleki yetkinliğin 4 dönemdeki ayrı ayrı olan ham puanı, ortalama alınarak yıllık puanı belirlenmiştir. Bu bağlamda veri hazırlama aşamasında ortaya şöyle bir problem çıkmıştır:

4 dönem boyunca mesleki yetkinlik puanlamaları için; puan aralığı 4-20 olurken (her dönem için mesleki yetkinlik puanlamaları en az 1 puan en fazla 5 puan olmak zorundadır); sadece 2 dönem mesleki yetkinlik puanları mevcut olan kişilerin ilgili yetkinliğinde puan aralığı 2-10 arasında olmuştur. Bu biçimde olan “eksik veriler” için veri hazırlaması gerçekleşmiştir. Veri hazırlama sonucunda nihai hale gelen verilerin WEKA Explorer ’da görünümü EK-1 de yer almaktadır.

Eksik ve hatalı veriler elimine edilmesinden sonra veri dönüştürme değişiklikleri aşağıda sıralanmaktadır:

1. Kabin memuru numaraları gizlilik sebebiyle 1’den başlayarak aritmetik bir şekilde sıralanmıştır.
2. Medeni Hal: Evli ve bekar olarak iki kategoride ele alınmıştır. Sırasıyla “E” ve “B” olarak tanımlanmıştır.
3. Yaş: Kabin memurlarının doğum tarihlerinden yola çıkarak hesaplanmıştır. En küçük yaş; 22, en büyük yaş: 54 olmasından dolayı uygulanacak karar ağacı algoritmalarından fazla dallanmaya sebep olmamasından dolayı kategorik hale getirilmiştir.

Çizelge 5.1. Yaş değişkeninin tanımlanması

Yaş Aralığı	Uygulama Tanımı
“20-29”	“22-30 YAS”
“30-39”	“31-39 YAS”
“40-49”	“40-49 YAS”
“50”	“50 YASVEUSTU”

4. Sınıf: Kabin memurlarının sınıflarına göre; “Z, E, Y,N, Q” olarak belirtilmiştir.
5. Eğitim: Öğrenim durumlarına göre 5 farklı tanımlama yapılmıştır.

Çizelge 5.2. Eğitim değişkeninin tanımlanması

Öğrenim Durumu	Uygulama Tanımı
Lise	3
Yüksekokul	4
Üniversite	5
Yüksek Lisans	6
Doktora	7

6. Yetkinliklerin (KF1, KF2, EI1, EI2, AP, SO, TR1, TR2, MO1, MO2) 2015 yılı 4 dönem içerisinde aldıkları ham puan verileri; 2015 yılı için ortalama alınarak birleştirilmiştir.

Veri hazırlama evresi veri madenciliği çalışmalarının en temel aşamalardan biri olması sebebi ile modelin kurulması ve yorumlanması aşamasında ortaya çıkacak sorunlar neticesinde tekrar bu aşamaya dönülmesi gerekmiştir.

Nitekim, bu çalışma ilk olarak; EK-2'de ayrıca belirtilmiş olan KF1, KF2, EI1, EI2, AP, SO, TR1, TR2, MO1, MO2 nominal değişken; karne seviyeleri ham veriden ön işleme alınmadan direkt bağlı karne düzeyleri (harf bazında) olarak tanımlanmış ve ardından bir takım sınıflandırma algoritmaları çalıştırılmış ve test edilmiştir.

İlgili algoritma sonuçlarına göre ayrıca EK-3'te gösterilmiş olan doğru olarak sınıflandırılan kayıt oranının en yüksek olduğu algoritma J48 olarak seçilmiştir. J48 çıktılarına bakıldığında ağaç boyutunun 215 olarak sonuçlanmasından dolayı algoritma çıktısının yorumlanması güçlenmiştir. Bu sebeple veri hazırlama/ ön işleme evresine tekrar dönülmüş bazı değişkenlerin türleri aşağıda belirtildiği gibi değiştirilmiştir:

7. KF1, KF2, EI1, EI2, AP, SO, TR1, TR2, MO1, MO2 mesleki yetkinlik değişken türleri numerik olarak değiştirilmiştir. Bunun sebebi yukarıda belirtildiği üzere ağacın çok fazla dal çıktısı oluşturmamasından dolayı algoritma çıktısı yorumlanamamıştır. Bu bağlamda her karar ağacı düğümünden çıkan dallardaki tüm nominal değerler için ayrı ayrı çıktı oluşturmamıştır.
8. 2015 yılında kabin memurlarının gerçekleştirdikleri dış hat tarifeli uçuşlarda kabin amirleri tarafından değerlendirmeleri neticesinde yayımlanan yıllık karne sonuçları aşağıda yer alan sayısal değerlere dönüştürme işlemi yapılmıştır.

Çizelge 5.3. Karne düzey değişkeninin tanımlanması

Karne Düzeyi	Uygulama Tanımı
A	4
B1	3
B2	3
C1	2
C2	1
D1	1

Veri dönüştürme işlemi tamamlandıktan sonra veri tablosunun gösterimi Çizelge 5.4’te gösterilmektedir.

Çizelge 5.4. Veri dönüşümü sonrası veri tablosu

KM Numaraları	Eğitim	Sınıf	TECRÜBE	YAŞ	Cinsiyet	Medeni Hal	KF1	KF2	E1	E2	AP	SO	TR1	TR2	MO1	MO2	2015
133	3	E	6-10	30-39	E	B	16	16	16	16	16	16	16	17	16	16	3
134	3	E	6-10	30-39	E	E	16	16	16	16	15	15	16	16	15	15	3
135	3	E	6-10	30-39	E	B	14	14	15	14	14	14	15	14	14	14	2
136	5	E	6-10	30-39	E	E	15	15	15	15	15	15	15	15	15	15	2
137	5	E	6-10	30-39	E	E	15	15	15	14	15	15	15	15	15	15	2
138	3	Z	6-10	30-39	E	B	16	16	16	16	16	16	16	16	16	16	3
139	3	Z	6-10	40-49	E	B	15	15	16	16	16	16	16	16	16	16	3
140	3	E	6-10	30-39	E	B	16	17	16	16	16	16	16	17	17	16	3
141	3	E	6-10	30-39	E	B	16	16	16	16	16	16	16	16	16	16	3
142	3	Q	6-10	30-39	E	E	16	16	17	16	16	16	17	17	16	16	3
143	4	E	6-10	30-39	E	E	16	16	16	16	16	16	16	16	16	16	3
144	3	E	6-10	30-39	E	E	16	16	16	16	16	16	16	16	16	16	3
145	3	E	6-10	30-39	E	E	16	16	16	16	16	16	16	16	16	16	3
146	3	E	6-10	30-39	E	E	16	16	16	16	16	16	16	16	16	16	3
147	4	E	6-10	30-39	E	E	16	15	16	16	16	15	16	16	16	16	3
148	3	E	6-10	30-39	E	B	15	15	15	15	15	15	15	15	15	15	2
149	5	Q	1-5	20-29	E	B	17	17	17	16	16	16	16	16	16	16	3
150	3	E	6-10	30-39	E	E	17	17	17	17	17	17	17	17	17	17	3
151	3	E	6-10	30-39	E	E	15	14	15	15	15	15	15	15	15	15	2
152	4	Q	6-10	30-39	E	E	17	17	17	17	17	17	17	17	17	17	3
153	5	Q	6-10	30-39	E	E	17	18	17	17	17	17	17	17	17	17	3
154	5	E	6-10	40-49	E	B	16	16	16	16	16	16	16	16	16	16	3
155	4	E	1-5	20-29	E	B	16	16	15	15	15	15	15	15	15	15	2

Özellikler (attributes)

Bu çalışmada değişken olarak kabin memurlarının demografik özellikleri ve performans değerlendirmelerinde kullanılan mesleki yetkinlik kriterleri yer almaktadır. Çalışmada kullanılan veri türlerinin nihai hali (attribute types) Çizelge 5.5’te gösterilmiştir.

Çizelge 5.5. Çalışmada kullanılan nihai veri türleri

@relation	2015
@attribute	EGITIM {3,4,5,6,7}
@attribute	SINIF {Z,E,Y,N,Q}
@attribute	TECRUBE {1-5,6-10,11-15,16-20}
@attribute	YAS {20-29,30-39,40-49,50}
@attribute	CINS {E,K}
@attribute	MEDENIHAL {E,B}
@attribute	KF1 numeric
@attribute	KF2 numeric
@attribute	EI1 numeric
@attribute	EI2 numeric
@attribute	AP numeric
@attribute	SO numeric
@attribute	TR1 numeric
@attribute	TR2 numeric
@attribute	MO1 numeric
@attribute	MO2 numeric
@attribute	2015 {1,2,3,4}
@Data	
	3,E,11-15,50,K,E,16,16,16,16,16,16,16,16,16,3
	3,E,11-15,40-49,K,E,16,17,16,16,16,16,16,16,16,3

2015 yılı için kullanılan değişkenlere göre dağılımlar EK-4’de yer alan WEKA ekran görüntüleri ile yorumlanmıştır:

Eğitim

Lise mezunu 902 kişi, yüksekokul mezunu 627 kişi, üniversite mezunu 2143 kişi, yüksek lisans mezunu 90 kişi ve doktora mezunu 2 kişi olduğu görülmektedir.

Sınıf

Kabin memurlarının işe giriş tarihlerinden itibaren tecrübe ve performans değerlendirmeleriyle orantılı olarak sınıf değişiklikleri olmaktadır. Analize katılan 5 grup sınıf yer almaktadır.

Tecrübe

Kabin memurlarının işe girişlerinden itibaren geçirdikleri yıl bazında çalışma süredir. Kategorik olarak analiz edilen *tecrübe yılı* en az çalışma süresi 2 yıl, en fazla çalışma süresi 19 yıl olarak tespit edilmiştir.

Yaş

Veri tabanında doğum tarihleri belli olan kabin memurlarının Excel yardımıyla yaşları hesaplanmıştır. Akabinde anlamlı şekilde kategorik hale dönüştürme işlemi yapılmıştır.

Cinsiyet

Veri tabanından “Bay” ve “Bayan” olarak çekilmiştir.%65 oranında bayan kabin memuru, %35 oranında bay kabin memuru bulunmaktadır.

Medeni hal

Veri tabanından “Evli” ve “Bekar” olarak çekilmiştir. Analizde kategorik değişken olarak kullanılmıştır. Evli personel % 34 oranında olurken bekar olan personel %66 oranında hesaplanmıştır.

Denetimli öğrenme algoritmalarında gerek duyulan hedef değişken (bağımlı değişken) çalışmada 2015 yılı yıllık karne düzeyine karşılık gelmiştir. Bahse konu olan karne puanları ise “karne düzeyi tanımlamalarına” istinaden 1-4 arasında kategorik olarak analiz edilmiştir.

5.2.4. WEKA’da sınıflandırma algoritmalarının uygulanması ve algoritma sonuçları

Kabin memurlarının demografik özelliklerinin ve yetkinlik bazlı performans değerlendirmeleri ile hedef değişken olan 2015 yılı karne düzeyi, arasında anlamlı kural oluşturma amaçlanmıştır. WEKA programında yararlanılan sınıflandırma algoritmaları aşağıda yer almaktadır.

J48 Algoritması

C4.5 karar ağacının WEKA tarafından Java dilinde kodlanan versiyonudur. Nümerik özellikler, kayıp değerler, gürültülü veri ile başa çıkabilmekte, kesikli ve sürekli veri üzerinde karar ağacı oluşturabilmektedir [64].

```

=== Summary ===

Correctly Classified Instances      3565              94.7131 %
Kappa statistic                    0.8814
Mean absolute error                0.0368
Root mean squared error            0.1487
Relative absolute error            16.4467 %
Root relative squared error        44.4315 %
Total Number of Instances          3764

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
      0,480    0,002    0,632    0,480    0,545     0,548    0,943    0,619     1
      0,916    0,037    0,922    0,916    0,919     0,880    0,963    0,915     2
      0,967    0,078    0,962    0,967    0,965     0,892    0,965    0,968     3
      0,000    0,000    0,000    0,000    0,000     0,000    0,436    0,001     4
Weighted Avg.   0,947    0,064    0,946    0,947    0,946     0,886    0,964    0,948

=== Confusion Matrix ===

  a    b    c    d  <-- classified as
12  13    0    0 |  a = 1
 7 1118   95    0 |  b = 2
 0   82 2435    0 |  c = 3
 0    0    2    0 |  d = 4

```

Şekil 5.1. Algoritma görseli (J48)

OneR Algoritması

Basit sınıflandırma kuralları bulunmasını sağlayan, tek bir özellik üzerinde tek seviyeli karar ağacı oluşturan bir algoritmadır. Değişkenlerden baskın olanı seçilir ve o değişkene göre tek seviye dallar oluşturulur. Bu algoritmada baskın olan değişken “Sürekli Öğrenme ve Kişisel Gelişim” olarak seçilmiştir.

SO:

< 13.5 -> 1

< 15.5 -> 2

< 18.5 -> 3

>= 18.5 -> 4

(3537/3764 instances correct)

=== Summary ===

Correctly Classified Instances	3535	93.916 %
Kappa statistic	0.864	
Mean absolute error	0.0304	
Root mean squared error	0.1744	
Relative absolute error	13.5775 %	
Root relative squared error	52.1311 %	
Total Number of Instances	3764	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,200	0,000	1,000	0,200	0,333	0,446	0,600	0,205	1
	0,920	0,052	0,895	0,920	0,907	0,862	0,934	0,849	2
	0,956	0,078	0,961	0,956	0,958	0,875	0,939	0,948	3
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	4
Weighted Avg.	0,939	0,069	0,940	0,939	0,938	0,868	0,935	0,911	

=== Confusion Matrix ===

a	b	c	d	<-- classified as
5	20	0	0	a = 1
0	1123	97	0	b = 2
0	112	2405	0	c = 3
0	0	0	2	d = 4

Şekil 5.2. Algoritma görseli (OneR)

ZeroR Algoritması

Hedefe dayalı olan ve tahminleme içermeyen en basit sınıflama yöntemidir.

```

=== Summary ===

Correctly Classified Instances      2517          66.8704 %
Kappa statistic                    0
Mean absolute error                 0.224
Root mean squared error             0.3346
Relative absolute error             100 %
Root relative squared error         100 %
Total Number of Instances          3764

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          0,000    0,000    0,000    0,000    0,000    0,000    0,450    0,006    1
          0,000    0,000    0,000    0,000    0,000    0,000    0,500    0,324    2
          1,000    1,000    0,669    1,000    0,801    0,000    0,499    0,668    3
          0,000    0,000    0,000    0,000    0,000    0,000    0,100    0,001    4
Weighted Avg.    0,669    0,669    0,447    0,669    0,536    0,000    0,498    0,552

=== Confusion Matrix ===

  a   b   c   d   <-- classified as
  0   0   25   0 |   a = 1
  0   0 1220   0 |   b = 2
  0   0 2517   0 |   c = 3
  0   0   2   0 |   d = 4

```

Şekil 5.3. Algoritma görseli (ZeroR)

Rastgele Ağaç (Random Tree) Algoritması

Her düğümde belli bir sayıda rastgele seçilmiş özellikler ile ağaç meydana getiren bir sınıflama algoritmasıdır. Budama işlemi bulunmamaktadır. Bununla birlikte veri setine dayalı sınıf olasılıklarının tahminine izin veren bir alternatifi mevcuttur.

```

=== Summary ===

Correctly Classified Instances      3529          93.7566 %
Kappa statistic                    0.8611
Mean absolute error                 0.0332
Root mean squared error             0.1713
Relative absolute error             14.818 %
Root relative squared error         51.2029 %
Total Number of Instances          3764

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          0,600    0,003    0,536    0,600    0,566    0,564    0,858    0,456    1
          0,909    0,049    0,899    0,909    0,904    0,858    0,936    0,864    2
          0,955    0,079    0,961    0,955    0,958    0,873    0,940    0,950    3
          1,000    0,000    1,000    1,000    1,000    1,000    1,000    1,000    4
Weighted Avg.    0,938    0,068    0,938    0,938    0,938    0,866    0,938    0,919

=== Confusion Matrix ===

  a   b   c   d   <-- classified as
 15   10   0   0 |   a = 1
 13 1109   98   0 |   b = 2
 0   114 2403   0 |   c = 3
 0     0     0   2 |   d = 4

```

Şekil 5.4. Algoritma görseli (Rastgele Ağaç)

Çok Katmanlı Algılayıcı (Multilayer Perceptron) Algoritması

Geriye yayılımı kullanan bir sinir ağıdır ve girdi katmanı, saklı katman ve sonuç katmanı ile üç katmandan oluşmaktadır.

```

=== Summary ===
Correctly Classified Instances      3557          94.5005 %
Kappa statistic                    0.8761
Mean absolute error                 0.0292
Root mean squared error            0.1531
Relative absolute error            13.0387 %
Root relative squared error        45.7603 %
Total Number of Instances          3764

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,000	0,000	0,000	0,000	0,000	0,000	0,985	0,581	1
	0,920	0,042	0,912	0,920	0,916	0,876	0,982	0,957	2
	0,967	0,079	0,961	0,967	0,964	0,891	0,983	0,987	3
	0,000	0,000	0,000	0,000	0,000	0,000	0,037	0,000	4
Weighted Avg.	0,945	0,067	0,938	0,945	0,942	0,879	0,982	0,974	

```

=== Confusion Matrix ===

```

a	b	c	d	<-- classified as
0	25	0	0	a = 1
0	1123	97	0	b = 2
0	83	2434	0	c = 3
0	0	2	0	d = 4

Şekil 5.5. Algoritma görseli (Çok Katmanlı Algılayıcı)

Rastgele Orman (Random Forest) Algoritması

Bu yöntem çok sayıda karar ağacından oluşan orman sınıflayıcısı olarak tanımlanmaktadır. Sınıflama, regresyon ve kümeleme uygulamaları yapılabilmektedir. Sınıf değişkeni kategorik ise sınıflama, sürekli ise regresyon ağaçları kurulmaktadır. Ormanda yer alan her bir karar ağacı, bootstrap tekniği ile orijinal veri setinden örneklem seçmekte; her karar düğümünde ise tüm değişkenler içinden belli sayıda rastgele değişkenin seçilmesi ile oluşturulmaktadır. Bu yöntemde CART algoritması ile ağaçlar oluşturulur ve ağaçlar da budama işlemi yapılmaz. CART algoritması “bilgi kazancını” kullanarak veride hangi değişkenden başlanılarak dallara ayrılacağını tespit etmektedir. Dallara ayrılmak için seçilen değişkenin uygun test ölçütü “Gini Katsayısı” ile tespit edilir.

```

=== Summary ===

Correctly Classified Instances      3576          95.0053 %
Kappa statistic                    0.8881
Mean absolute error                 0.0334
Root mean squared error             0.1358
Relative absolute error             14.9075 %
Root relative squared error         40.5849 %
Total Number of Instances          3764

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          0,600    0,001    0,750    0,600    0,667      0,669    0,956    0,661    1
          0,921    0,036    0,925    0,921    0,923      0,886    0,987    0,970    2
          0,968    0,074    0,964    0,968    0,966      0,896    0,989    0,993    3
          0,500    0,000    1,000    0,500    0,667      0,707    1,000    1,000    4
Weighted Avg.  0,950    0,061    0,950    0,950    0,950      0,891    0,988    0,983

=== Confusion Matrix ===

  a    b    c    d  <-- classified as
15    10     0     0 |    a = 1
 5 1124    91     0 |    b = 2
 0   81 2436     0 |    c = 3
 0     0     1     1 |    d = 4

```

Şekil 5.6. Algoritma görseli (Rastgele Orman)

NaiveBayes Algoritması: Bayes teoremine dayanan standart olasılığa dayalı sınıflandırma algoritmasıdır.

```

=== Summary ===

Correctly Classified Instances      3557          94.5005 %
Kappa statistic                    0.8791
Mean absolute error                 0.0279
Root mean squared error             0.1564
Relative absolute error             12.4716 %
Root relative squared error         46.7394 %
Total Number of Instances          3764

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          1,000    0,016    0,291    1,000    0,450      0,535    0,998    0,806    1
          0,898    0,033    0,930    0,898    0,914      0,874    0,984    0,968    2
          0,967    0,051    0,975    0,967    0,971      0,913    0,992    0,996    3
          1,000    0,000    1,000    1,000    1,000      1,000    1,000    1,000    4
Weighted Avg.  0,945    0,044    0,956    0,945    0,949      0,898    0,990    0,986

=== Confusion Matrix ===

  a    b    c    d  <-- classified as
25     0     0     0 |    a = 1
61 1096    63     0 |    b = 2
 0   83 2434     0 |    c = 3
 0     0     0     2 |    d = 4

```

Şekil 5.7. Algoritma görseli (Naive Bayes)

Decision Stump Algoritması: Tek seviyeli bir karar ağacı algoritmasıdır. Tahminleme bazlı olan bu algoritma ilk giriş özelliğine çeşitleme yapmaktadır.

```

=== Summary ===
Correctly Classified Instances      3506           93.1456 %
Kappa statistic                    0.8464
Mean absolute error                 0.0608
Root mean squared error             0.1781
Relative absolute error             27.1171 %
Root relative squared error         53.2478 %
Total Number of Instances          3764

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,000	0,000	0,000	0,000	0,000	0,000	0,815	0,017	1
	0,910	0,057	0,884	0,910	0,897	0,846	0,922	0,830	2
	0,952	0,090	0,955	0,952	0,954	0,861	0,927	0,943	3
	0,000	0,000	0,000	0,000	0,000	0,000	0,398	0,001	4
Weighted Avg.	0,931	0,079	0,925	0,931	0,928	0,850	0,924	0,900	

```

=== Confusion Matrix ===
  a    b    c    d  <-- classified as
0  25    0    0 |    a = 1
0 1110  110    0 |    b = 2
0  121 2396    0 |    c = 3
0    0    2    0 |    d = 4

```

Şekil 5.8. Algoritma görseli (DecisionStump)

5.2.5. Sınıflandırma algoritmalarının karşılaştırılması

8 farklı sınıflandırma algoritması denenmiş ve doğruluk oranları ve model oluşum zamanını içeren elde edilen sonuçlar Çizelge 5.6' da yer almaktadır.

Çizelge 5.6. Sınıflandırma algoritmalarının karşılaştırılması

	Doğru Sınıflandırılan örnek	Kappa istatistiği	Ort. mutlak hata	Ort. mutlak hata karekök	Görelî mutlak hata (%)	Görelî hata karekök (%)	Model Zamanı	TP	FP	F-Ölçütü	ROC
J48	3565	0,88	0,04	0,15	16,45	44,43	0,36 sn	0,947	0,064	0,946	0,964
Naive Bayes	3557	0,88	0,03	0,16	12,47	46,74	0,06 sn	0,945	0,044	0,949	0,99
OneR	3535	0,86	0,03	0,17	13,58	52,13	0,05 sn	0,939	0,069	0,938	0,935
ZeroR	2517	0,00	0,22	0,33	100,00	100,00	0,06 sn	0,669	0,669	0,536	0,498
Random Tree	3529	0,86	0,03	0,17	14,82	51,20	0,11 sn	0,938	0,068	0,938	0,938
Multilayer Perceptron	3557	0,88	0,03	0,15	13,04	45,76	154 sn	0,945	0,067	0,942	0,982
Random Forest	3576	0,89	0,03	0,14	14,91	40,58	4 sn	0,95	0,061	0,95	0,988
Decision Stump	3506	0,85	0,06	0,18	27,12	53,25	0,08 sn	0,931	0,079	0,928	0,924

İlgili çizelgeye göre; belirtilmiş olan sınıflandırma algoritmaları kullanılarak elde edilmiş performans sonuçlarına yer verilmiştir. 10 katlı çapraz doğrulama ile yapılan sınıflandırma analizinde 100 ağaç türetilerek edinilen Random Forest birinci sınıftaki doğru sınıflandırılan kayıt TP (True Positive) oranı %95,0053, ikinci sınıftaki kayıtların sayısını veren FP (False Positive) 0,061 olarak en iyi sınıflandırma başarısını elde etmiştir. Sınıflandırma doğruluğu açısından 2. sırada ise J48 karar ağacı algoritması bulunmaktadır. ROC eğrisine göre; eğri altında kalan alan 1'e yaklaştığı takdirde testin doğruluğu artmasından dolayı Random Forest ve J48 algoritmasının ROC eğrisi altında kalan alanları karşılaştırılmıştır. RF algoritması için 0,988 olarak, J48 algoritması için 0,964 olarak tespit edilmiştir.

Bu 2 algoritmayı ayrıca model oluşum zamanları bakımından mukayese edildiğinde J48 algoritması, Random Forest algoritmasına göre daha kısa sürede sonuç verdiği tespit edilmiştir. Her 2 algoritmatya göre sonuçlar yorumlanmıştır.

5.2.6. J48 algoritmasına göre sonuçların yorumlanması

WEKA' da analiz edilen J48 algoritmasına göre ilk dallanmanın “Sürekli Öğrenme ve Kişisel Gelişim” olduğu görülmektedir. Yetkinliklere göre değerlendirilen kabin memurlarının; ilgili değişkenlerden hangilerinin ne ölçüde karne düzeylerini etkilediğini hedefleyen çalışmada ilk dallanma Sürekli Öğrenme ve Kişisel Gelişim'in 15 değerinden büyük veya küçük eşit değerler almasından sonra farklılık göstermiştir. EK-5'de J48 algoritmasının görsel hali yer almakta olup elde edilen sonuçların bir bölümü özet halinde aşağıda gösterilmiştir.

```

SO <= 15
|   KF1 <= 15
|   |   KF2 <= 14
|   |   |   SO <= 13: 1 (7.0)
|   |   |   SO > 13
|   |   |   |   EI2 <= 14
|   |   |   |   |   EGITIM = 3: 2 (12.0/1.0)
|   |   |   |   |   EGITIM = 4: 2 (6.0/2.0)
|   |   |   |   |   EGITIM = 5
|   |   |   |   |   |   MEDENIHAL = E: 1 (6.0)
|   |   |   |   |   |   MEDENIHAL = B
|   |   |   |   |   |   |   YAS = 20-29: 2 (5.0)
|   |   |   |   |   |   |   YAS = 30-39
|   |   |   |   |   |   |   |   SINIF = Z: 1 (0.0)
|   |   |   |   |   |   |   |   SINIF = E: 2 (3.0/1.0)
|   |   |   |   |   |   |   |   SINIF = Y: 1 (7.0/2.0)
|   |   |   |   |   |   |   |   SINIF = N: 1 (0.0)
|   |   |   |   |   |   |   |   SINIF = Q: 1 (0.0)
|   |   |   |   |   |   |   |   YAS = 40-49: 2 (0.0)
|   |   |   |   |   |   |   |   YAS = 50: 2 (0.0)
|   |   |   |   |   |   |   |   EGITIM = 6: 2 (2.0)
|   |   |   |   |   |   |   |   EGITIM = 7: 2 (0.0)
|   |   |   |   |   |   |   |   EI2 > 14: 2 (14.0)
|   |   |   |   |   |   |   |   KF2 > 14: 2 (985.0/26.0)

```

Şekil 5.9. KF1 değişkeninin 15'ten en küçük veya eşit olması durumunda oluşan karar kuralı

- Eğer SO değeri 15'ten küçük veya eşit ise KF1 değeri 15'ten küçük veya eşit, KF2 14'ten büyük ise karne seviyesi 2 (yeterli) olarak tespit edilmiştir.
- Eğer SO değeri 15'ten küçük veya eşit ise KF1 değeri 15'ten küçük veya eşit, KF2 14'ten küçük veya eşit ise SO ya bakılmalıdır. SO değeri 13'ten küçük veya eşit ise 1(yetersiz) olduğunu göstermektedir.
- Eğer SO değeri 15'ten küçük veya eşit ise KF1 değeri 15'ten küçük veya eşit olduğu durumda üniversite mezunu bekar 30-39 yaş grubu arasında Y sınıfı kabin memurlarının karne seviyesi 1 (yetersiz) olarak gözlenmiştir. (KF2 ve EI2 değerleri 14 ten küçük veya eşit olması durumunda) temsil kabiliyeti ile yolcuları güler yüzle karşılama arasında güçlü bir bağıntı olduğu ve karne seviyesini etkilediği görülmüştür.

KF1 değerinin 15'ten küçük veya eşit olduğu tüm durumlarda karne seviyesinin en fazla 2 (yeterli) olduğu görülmektedir. Karne düzeyinin iyi ve çok iyi olduğu durum tespit

edilmemiştir. Bu durumda kabin memurlarının davranış konusunda temsil kabiliyetlerini geliştirmeleri öncelikli olarak belirlenen durum olmuştur.

```

| KF1 > 15
| | TR2 <= 15
| | | EI2 <= 15: 2 (41.0/2.0)
| | | EI2 > 15
| | | | KF2 <= 15: 2 (6.0)
| | | | KF2 > 15
| | | | | TECRUBE = 1-5
| | | | | MO2 <= 15
| | | | | AP <= 15: 2 (42.0/8.0)
| | | | | AP > 15: 3 (4.0/1.0)
| | | | | MO2 > 15
| | | | | AP <= 15: 3 (6.0/1.0)
| | | | | AP > 15: 2 (3.0)
| | | | | TECRUBE = 6-10: 3 (5.0/1.0)
| | | | | TECRUBE = 11-15: 2 (0.0)
| | | | | TECRUBE = 16-20: 2 (0.0)
| | TR2 > 15
| | | TR1 <= 15
| | | | EI2 <= 15: 2 (8.0)
| | | | EI2 > 15
| | | | | KF2 <= 15: 2 (2.0)
| | | | | KF2 > 15: 3 (5.0/1.0)
| | | TR1 > 15
| | | | KF2 <= 15
| | | | | TECRUBE = 1-5: 2 (14.0/4.0)
| | | | | TECRUBE = 6-10: 3 (4.0/1.0)
| | | | | TECRUBE = 11-15: 3 (1.0)
| | | | | TECRUBE = 16-20: 2 (0.0)
| | | | KF2 > 15: 3 (72.0/17.0)

```

Şekil 5.10. KF1 değişkeninin 15'ten büyük olması durumunda oluşan karar kuralı

- Eğer SO değeri 15’den küçük veya eşit ise KF1 değeri 15’den büyük, sırasıyla TR2, TR1 ve KF2 değerleri 15 den büyük ise karne seviyesi 3 (iyi) olarak tespit edilmiştir. Temsil ve görüntü bakımından (yolcuya karşı izlenim); ekip çalışması ile sorumluluk bilinciyle görevlerin tam ve zamanında yerine getirilmesi (yolcuya en az beklenen hizmet sunulması) durumları karne seviyesinde “iyi” olarak belirlenmiştir.
- Eğer SO değeri 15’den küçük veya eşit ise KF1 değeri 15’den büyük, TR2 15 den küçük veya eşit olduğu durumda yolcuları temsil kabiliyeti ile güler yüzle karşılama kriterleri 15’ten yüksek olduğu durumda tecrübesi 6-10 yıl arasında değişen kabin memurlarının karne düzeyleri “iyi” olarak tespit edilmiştir. Görevlerini tam ve zamanında yerine

getirme konusunda aksaklık gösteren bir kabin memurunun yolcuya karşı güler yüzlü hizmet vermesi ve davranışları ile temsil kabiliyetinin yüksek olması tecrübesi ile orantılı olarak karne seviyesini olumlu etkilemiştir. Bu durumda olan kabin memurları için zaman yönetimi eğitimi önerilebilmektedir.

```

SO > 15
|   MO1 <= 15
|   |   KF1 <= 15: 2 (32.0/6.0)
|   |   KF1 > 15
|   |   |   MO2 <= 15
|   |   |   |   EI2 <= 15
|   |   |   |   |   TR1 <= 15: 3 (3.0/1.0)
|   |   |   |   |   TR1 > 15: 2 (10.0/1.0)
|   |   |   |   |   EI2 > 15
|   |   |   |   |   |   KF2 <= 15
|   |   |   |   |   |   |   TR1 <= 15: 3 (2.0)
|   |   |   |   |   |   |   TR1 > 15: 2 (7.0/1.0)
|   |   |   |   |   |   |   KF2 > 15: 3 (44.0/14.0)
|   |   |   |   |   |   MO2 > 15: 3 (22.0/3.0)
|   |   |   MO1 > 15: 3 (2384.0/40.0)

```

Şekil 5.11. MO1 değişkeninin 15’ten büyük olması durumunda oluşan karar kuralı

- Eğer SO değeri 15’ten büyük ve MO1 değeri 15’ten büyük ise karne seviyesi 3 (iyi) olduğu tespit edilmiştir. Bir kabin memurunun sürekli öğrenme konusunda gelişim göstermesi ile ürün ve hizmet bilgilerine daha hâkim olması ve bunun sonucunda yolculara doğru ve zamanında hizmet sunması beraberinde gelmektedir. Bu durum karne seviyesine beklenildiği gibi “iyi” düzeyde yansımaktadır.
- Eğer SO değeri 15’ten büyük ve sırasıyla MO1 ve KF1 değişken değerlerinin 15’ten küçük veya eşit olması durumunda karne seviyesi 2 olduğu gözlemlenmiştir. Bu durum, davranış konusunda havayolunu temsil etme ve müşteri odaklılık konularının bütün halinde olduğu ve aynı seviyede puan sonucunda karne seviyesinin 2 olarak gözlemlendiği tespit edilmiştir.

5.2.7. Random Forest algoritmasına göre sonuçların yorumlanması

Random Forest algoritmasında 50, 100, 150 ağaç türetilmesi ile 3 farklı algoritma kullanılarak edinilen karar ağaçları ve bu algoritmaların sonuçları karşılaştırılmıştır. Bu 3 algoritmaya ait WEKA çıktıları ekte yer almaktadır.

Çizelge 5.7. Random Forest algoritmasının ağaç sayılarına göre karşılaştırılması

	Doğru Sınıflandırılan kayıt (%)	Kappa istatistiği	Ort. mutlak hata	Ort. mutlak hata karekök	Görelî mutlak hata (%)	Görelî hata karekök (%)	Model Oluşum Zamanı
RandomForest (50 Ağaç)	95,1116	0,8904	0,0334	0,136	14,9037	40,6567	1,84 saniye
RandomForest (100 Ağaç)	95,0053	0,8881	0,0334	0,1358	14,9075	40,5849	3,19 saniye
RandomForest (150 Ağaç)	94,9522	0,887	0,0334	0,1358	14,9029	40,6041	5,02 saniye

Random Forest yönteminde en iyi performans sonuçlarının sınıflama doğruluğu bakımından 50 ağaç türetilerek edinildiği belirlenmiştir. Ağaç sayısının artmasıyla beraber doğru sınıflandırılan kayıt yüzdesinin azaldığı tespit edilmiştir. Ayrıca modelleme sürelerine bakıldığında sınıflandırma doğruluğu ile aynı orantıda olduğu gözlemlenmiştir.

Ayrıca Random Forest algoritması için aşağıda yer almakta olan üstün ve kısıtlı yönleri doğrultusunda belirtilmiştir:

Üstün Yönleri:

- Aşırı öğrenmeye ve veri setindeki kayıp verilere karşı oldukça güçlüdür. Çok fazla kayıp veri olduğu durumda da başarılı bir şekilde sınıflama yapmaktadır.
- Random Forests yöntemi kategorik, sürekli ya da hem kategorik hem sürekli değişkenlerin yer aldığı veri setinde uygulaması yapılabilmektedir.
- Oluşturulan ağaçlar için budama işlemi yapmaya gerek yoktur.
- Veri setlerinin boyutu sonuçları etkilememektedir. Yani, veri setinin büyük ya da küçük olması sonucunda da doğru sonucu verebilmektedir.
- Orijinal veri seti, öğrenme ve test şeklinde ayrılmadan da uygulanacak olan modeli test edebilmektedir. Model orijinal veri setini kullanarak, iç hata oranını hesaplamaktadır. Hesaplanan iç hata oranı sapmasızdır.
- Çok büyük sayıda değişken içeren veri setleri için de uygun bir yöntemdir.
- Veri setindeki sınıflama değişkeninde (bağımlı değişken) sınıf sayısı çok fazla olsa da bu yöntem sınıflama yapabilmektedir.

- Random Forest yönteminde sınıflamayı gerçekleştiren değişkenlerin önem dereceleri hesaplanabilmektedir.
- Değişkenler arasındaki ilişkiler ve mesafe, proximity matrisi oluşturarak tespit edilebilmektedir. Proximity, kümeleme yaparken, aşırı değerleri (outlier) tespit etmek için kullanılmasının yanında verinin görsel olarak gösterilmesinde de oldukça yararlıdır.
- Sınıflara düşen örnek sayısı (sınıf dağılımları) dengesiz olduğunda, sınıfları dengeli olacak şekilde ele alabilmektedir.
- Uygulama da analistin sadece iki parametre seçmesi gerekir. Bunlar “oluşturulacak ağaç sayısı” ve “her düğümde rastgele seçilecek değişken sayısı”dır. Ancak bu değerler ne olursa olsun sonuç çok fazla değişmemektedir.

Kısıtları:

- Karar ağaçlarında tek bir ortaya çıkan sonuç ile karşılaştırıldığında, bu algorithmada ağaç yapısı sebebiyle görsel bir biçimde görülmemektedir. Karmaşık bir model olarak çıktı oluşturan algoritma işlem adımlarını görülmeyecek (blackbox) şekilde verir.
- Oluşan sonuç için bir güven aralığı bilgisi vermemektedir.
- Random Forest metodunda var olan sistemler, meydana gelen her ağacın bellekte tutulmasından dolayı oldukça fazla bellek kullanmaktadır ve bundan dolayı düşük belleğe sahip bilgisayarlar için uygulama zorlaşmaktadır.

Karar ağacı gibi görsel bir sonuç vermeyip algoritma adımlarını görülmeyecek şekilde vererek karmaşık bir yapı oluşmasından dolayı çalışma J48 algoritmasına göre yorumlanmıştır.

Öyle ki, kabin memurlarının karne verilerini (A, B1, B2, C1, C2, D1) etkileyen özelliklerin belirlenmesi için, Çizelge 5.8 de aktarılmış olan nitelik seçimi ile yorumlanması gereken nitelikler tespit edilmiştir. WEKA uygulamasında nitelik seçimi özelliği ile “BestFirst” ve “Ranker” metodları uygulanması sonucunda çıktıların J48 algoritması çıktıları ile aynı doğrultuda olduğu tespit edilmiştir.

Çizelge 5.8. Nitelik seçme ve sonuçlar

Nitelik Seçimi(Attribute Evaluator)	Metod (Search Method)	Sonuç
CfsSubsetEval	BestFirst	SINIF, KF1, KF2,EI2, AP,SO,TR2,MO1,MO2
InfoGainAttributeEval	Ranker	0.66246 SO 0.65239 MO1 0.635273 TR2 0.632395 KF1 0.628692 MO2 0.617676 AP 0.601472 EI2 0.583548 TR1 0.566257 EI1 0.477559 KF2 0.169042 SINIF 0.005775 EGITIM 0.005161 YAS 0.004515 MADENİHAL 0.003407 TECRUBE 0.000109 CINSİYET

“InfoGainAttributeEval” algoritması sonucunda en yüksek nitelik olarak belirlenen SO, J48 algoritmasında da ilk dallanma değişkenidir. Kabin memurlarının karne düzeyleri için demografik verilerin anlamlı bir değişken olmadığı görülmüştür. Bu bağlamda J48 ve nitelik seçme algoritmaları da aynı doğrultuda olduğu tespit edilmiştir.

6. SONUÇ

Günümüz işletmelerinde mevcut müşterileri elde tutma çabaları, onların memnuniyetleri için taleplerini değerlendirme ve beklentileri karşılama çalışmaları rekabet ortamında var olabilmesi gerekli şartlardan biridir.

Her geçen gün insanların daha fazla tercih ettiği havayolu taşımacılığında yolcuların her bir seyahat deneyiminin bir öncekinden daha iyi olması ve ihtiyaçlarının karşılanması, bir sonraki seyahatinde aynı havayolu işletmesine seçmesine neden olmaktadır.

Bir havayolu işletmesinde aktif olarak uygulanmakta olan dış hat tarifeli uçuşlardan sonra kabin memurları 6 ayrı yetkinlikte ve bu yetkinliklerin alt yetkinliği olan toplam 10 ayrı alt yetkinlik ile değerlendirilmektedirler. Bir yılda üçer aylık toplam 4 dönem için bu yetkinliklerde değerlendirme puanları neticesinde kurumsal karneleri yayımlanmaktadır. Kurumsal karne bağlı düzeyleri, ham puanlara göre ortalama alınarak ve sınıf dağılımlarına göre meydana gelmektedir.

Her bir yetkinliğin yıllık puanını bulabilmek amacıyla üçer aylık dönemsel puanının ortalaması alınarak hesaplanmıştır. Bu bağlamda mesleki yetkinlikler ve demografik veriler ile yıllık karne düzeyleri arasında anlamlı kurallar çıkarma noktasında veri madenciliği yöntemlerinden biri olan karar ağaçları kullanılmış ve bu uygulama ile literatüre katkı sağlayacağı düşünülmüştür. Öyle ki, havacılık sektöründe, performans değerlendirme uygulamalarında veri madenciliği yöntemlerinin uygulanması konularında literatürde oldukça az çalışma bulunmaktadır.

Bu çalışmada, kabin memurlarının 10 adet mesleki yetkinlik bazlı değerlendirme puanlarını kapsayan değişkenlerin yanı sıra eğitim, sınıf, cinsiyet, medeni hal, yaş, tecrübe gibi demografik değişkenlerde yer almaktadır. 2015 yılı Ocak- Aralık ayları içerisinde tarifeli dış hat uçuşlarda değerlendirilen kabin memurlarının her üçer aylık dönemler için belli olan mesleki yetkinlik puanları yıllık bazda ortalama alınarak hesaplanmıştır. 4 dönem boyunca mesleki yetkinlik puanlamaları için; puan aralığı 4-20 olurken (her dönem için mesleki yetkinlik puanlamaları en az 1 puan en fazla 5 puan olmak zorundadır); sadece 2 dönem mesleki yetkinlik puanları mevcut olan kişilerin ilgili yetkinliğinde puan aralığı 2-10 arasında olmuştur. Bu biçimde olan “eksik veriler” için veri hazırlaması gerçekleştirilmiştir.

CRISP-DM süreci ile ilerlenen veri madenciliği çalışmasında veri hazırlama (veri ön işleme) safhası bu çalışma için ayrıca önem arz etmektedir. Nitekim veri ön işleme safhasında başlangıçta 10 adet yetkinlik bazlı değişkenlerin özellik türleri (attribute types) nominal olarak belirlenmiş ve kabin memurlarının karne seviyeleri bağlı karne notu (harf seviyeleri ile) gösterilmiştir. WEKA programında çalıştırılan sınıflandırma algoritmaları neticesinde en yüksek sınıflama doğruluğuna sahip olan J48 algoritmasında çok fazla sayıda ağaç kuralı çıkmıştır. Bu karar ağacının yorumlanması daha basite indirgenmesi ve dallanmanın azalması için veri ön işleme safhasına geri dönmüştür. Mesleki yetkinlik bazlı değerlendirme puanlarına sahip 10 adet değişkenin türleri “nümerik” olarak değiştirilmiş ve kabin memurlarının karne seviyeleri bağlı düzey yerine 4 ayrı seviyeye indirgenmiştir. Bu kısımda tekrar bir budama yöntemine gidilmemiştir. Bunun sebebi budama işlemi sırasında veri kaybının olmasının istenmemesidir.

Veri ön işleme safhasından sonra karar ağacı algoritmalarından; J48, Naive Bayes, OneR, ZeroR, Random Tree, Multilayer Perceptron, Random Forest ve Decision Stump algoritmaları karşılaştırılmıştır. 10 katlı çapraz doğrulama ile yapılan sınıflandırma analizinde 100 ağaç türetilerek edinilen Random Forest birinci sınıftaki doğru sınıflandırılan kayıt TP (True Positive) oranı %95,0053, ikinci sınıftaki kayıtların sayısını veren FP (False Positive) 0,061 olarak en iyi sınıflandırma başarısını elde etmiştir. Sınıflandırma doğruluğu açısından 2. sırada ise J48 karar ağacı algoritması bulunmuştur. ROC eğrisine göre; eğri altında kalan alan 1'e yaklaştığı takdirde testin doğruluğu artmakta olduğu için Random Forest ve J48 algoritmasının ROC eğrisi altında kalan alanları karşılaştırıldığında; RF algoritması için 0,988 olarak, J48 algoritması için 0,964 olarak tespit edilmiştir.

Random Forest uygulamasında karar ağacı sayılarının 50,100 ve 150 ağaç türetilerek algoritma karşılaştırılmıştır. Ağaç sayısının artmasıyla beraber doğru sınıflandırılan kayıt yüzdesinin azaldığı tespit edilmiştir. Ayrıca modelleme sürelerine bakıldığında sınıflandırma doğruluğu ile aynı orantıda olduğu gözlemlenmiştir.

Random Forest algoritmasının J48 algoritmasına göre daha uzun modelleme süresi olması ve karar ağacı gibi görsel bir sonuç vermeyip algoritma adımlarını görülmeyecek şekilde vererek karmaşık bir yapı oluşmasından dolayı J48 algoritma çıktıları gözlemlenmiştir. İlk dallanma Sürekli Öğrenme ve Kişisel Gelişim (SO) değişkeni ile başlamıştır. Ayrıca, WEKA programında nitelik seçme algoritmalarından biri olan “Info Gain Attribute Eval”

algoritması sonucunda en yüksek nitelik olarak “SO” belirlenmiştir. Bu bağlamda “SO” değişkeninin J48 algoritmasında nitelik seçme algoritmasında da ortak değişken olduğu tespit edilmiştir. Ayrıca, kabin memurlarının karne düzeyleri için demografik verilerin anlamlı bir değişken olmadığı görülmüştür.

KF1 (*Görüntü ve kişisel bakım açısından örnek bir kabin memurudur*) değerinin 15 den küçük veya eşit olduğu ve KF2 (*Davranış konusunda ve havayolumuzu temsil etme açısından örnek bir kabin memurudur*) değerinin 14 den küçük veya eşit olduğu tüm durumlarda karne seviyesinin en fazla 2 (yeterli) olduğu görülmektedir. Bir başka deyişle karne düzeyinin iyi ve çok iyi olduğu durum tespit edilmemiştir. Bu durumda kabin memurlarının davranış konusunda temsil kabiliyetlerini geliştirmeleri öncelikli olarak belirlenen durum olmuştur. Öyle ki, rastgele seçilen numarası 1170 olan kabin memurunun KF1 ve KF2 puanının 14 olduğunu yıllık karne puanının da 2 olduğu en az iyi seviyesinde olmadığı tespit edilmiştir. Temsil kabiliyeti ile yolcuları güler yüzle karşılama arasında güçlü bir bağıntı olduğu ve karne seviyesini etkilediği görülmüştür. KF1 değeri 15 den büyük, sırasıyla TR2 (*Sorumluluk bilinciyle görevlerini tam ve zamanında yerine getirdi*), TR1(*Ekiple uyumlu çalıştı, görev alanının dışında da ekip arkadaşlarına yardımcı oldu*) ve KF2 (*Davranış konusunda ve havayolumuzu temsil etme açısından örnek bir kabin memurudur*) değerleri 15 den büyük ise karne seviyesi 3 (iyi) olarak tespit edilmiştir. Temsil ve görüntü bakımından (yolcuya karşı izlenim); ekip çalışması ile sorumluluk bilinciyle görevlerin tam ve zamanında yerine getirilmesi (yolcuya en az beklenen hizmet sunulması) durumları karne seviyesinde “iyi” olarak belirlenmiştir. 2015 yılı yıllık karne seviyesi 3 olarak belirlenmiş random olarak seçilen numarası 82 olan kabin memurunun yukarıda belirtilmiş olan KF1, TR2, TR1 ve KF2 yetkinlik puanları en az 15 olarak tespit edilmiştir. Görevlerini tam ve zamanında yerine getirme konusunda aksaklık gösteren bir kabin memurunun yolcuya karşı güler yüzlü hizmet vermesi, davranışları ile temsil kabiliyetinin yüksek olması; tecrübesi ile orantılı olarak tespit edilmiş ve karne puanı iyi seviyesinde görülmüştür. Bu durumda olan kabin memurları için sadece zaman yönetimi eğitimi önerilebilmektedir. Bir kabin memurunun sürekli öğrenme konusunda gelişim göstermesi ile ürün ve hizmet bilgilerine daha hâkim olması, bunun sonucunda yolculara doğru ve zamanında hizmet sunmasını beraberinde gelmektedir. Bu durum karne seviyesine beklenildiği gibi “iyi” düzeyde yansımaktadır.

MO1 (*Ürün, hizmet ve hat bilgilerine hâkim, yolcularımıza doğru ve zamanında hizmet sundu*) ve KF1 değişken değerlerinin 15 den küçük veya eşit olması durumunda karne seviyesi 2 olduğu gözlemlenmiştir. Bu durum, davranış konusunda havayolunu temsil etme ve müşteri odaklılık konularının bütün halinde olduğu ve aynı seviyede puan sonucunda karne seviyesinin 2 olarak gözlemlendiği tespit edilmiştir. Demografik verilerin karne puanlarına etkisi en erken 5. kırılım olan tecrübe ve eğitimde görülmektedir.

Sonuç olarak, havayolu işletmelerinin yolcularını tutma ve en az beklentileri doğrultusunda hizmet sunma noktasında kabin memurlarının yolcu ile birebir iletişimleri yadsınamaz bir gerçektir. Kabin memurlarının performans değerlendirme sonuçlarında demografik özelliklerin anlamlı kural oluşturmadığı, ilk kırılımın sürekli öğrenme ve kişisel gelişim olarak başladığını ve akabinde kurumsal temsil ve yolcuya doğru ve zamanında hizmet sunma konuları karne seviyesini önemli etkileyen değişkenler olarak belirlenmiştir. Uçuş operasyonu bazında ilgili değişkenlerin kabin memurlarının performans gelişimlerine katkı sağlayacağı ve eğitim, yaş, tecrübe gibi demografik verilerin yıllık karne puanına öncelikli olarak etki etmediği tespit edilmiştir. Bu bağlamda kabin memurlarının 10 alt yetkinlik içerisinde, teknik ve mesleki uzmanlık alanında sürdürülebilir bir gelişim göstermesi (iç kaynaktan beslenerek ihtiyaca göre eğitim planlamalarının yapılması), temsil kabiliyeti ile marka değerini yansıtmayı ve zamanında doğru bir hizmet sunması yıllık karne puanını en çok etkileyen değişkenler olmuştur. Ayrıca, incelenen literatürde performans değerlendirmesi, kurumsal karne yöntemlerine ilişkin veri madenciliği uygulamaları az sayıda olmasından dolayı bu sektörlerde de gelişim sağlanabilmesi ve çalışmanın bu bağlamda katkı sağlayacağı görülmektedir.

KAYNAKLAR

1. Han, J., and Kamber, M. (2001). *Data Mining: Concepts and Techniques 1st ed.* Morgan Kaufmann, USA, 3-16.
2. Larose, D. T. (2006). *Data Mining Methods and Models.* A John Wiley & Sons, Inc., Publication, New Jersey, 205-235.
3. Koyuncugil, A. S. (2006). *Bulanık Veri Madenciliği ve Sermaye Piyasalarına Uygulanması.* Doktora Tezi, Ankara Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 70–80.
4. Çetin, T. (2010). *Kurumsal Performans Karnesi ve Havacılık Sektöründe Bir Uygulama.* Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 9, 13.
5. Elitaş, C., ve Ağca, V. (2006). Firmalarda Çok Boyutlu Performans Değerleme Yaklaşımları: Kavramsal Bir Çerçeve, *Sosyal Bilimler Dergisi*, 8(2),343-370.
6. Usta, A. (2010). Kamu Kurumlarında Örgütsel Performans Yönetim Süreci, *Sayıştay Dergisi*, 78, 12-14.
7. Ateş, H. ve Çetin, D. (2004). *Kamuda Performans Yönetimi ve Performansa Dayalı Bütçe, Çağdaş Kamu Yönetimi II.* Ankara: Nobel Yayınları, 256.
8. Canitez, F. (2015). *Toplu Ulaşım İşletmelerinde Kurumsal Performans Yönetimi: İETT’de Kurumsal Karne Uygulaması.* Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 5-8.
9. Yenice, E. (2007). Performans Ölçümünde Karşılaşılan Sorunlar ve Kurumsal Karne (Balanced Scorecard) Yaklaşımı, *Bütçe Dünyası*, 2, 25.
10. Kaplan, R., ve Norton, D. (1999). *Balanced Scorecard Şirket Stratejisini Eyleme Dönüştürmek,* İstanbul: Sistem Yayıncılık, 71-79.
11. Eilat, H, Golany, B. and Shtub, A. (2008). R&D project evaluation: an integrated DEA and balanced scorecard approach, *Omega*, 36, 895 – 912.
12. Çukurçayır, M. A. ve Eroğlu, T. H. (2004). Yerel yönetimlerde yeniden yapılanmaya farklı bir yaklaşım: verimlilik ve başarı karnesi (balanced scorecard), *Sayıştay Dergileri*, 53, 41-67
13. Güner, M. F. (2006). *Stratejik Performans Değerlemede Dengeli Sonuç Kartı: Bir Sanayi İşletmesinde Uygulama.* Doktora Tezi, Çukurova Üniversitesi Sosyal Bilimler Enstitüsü, Adana, 1-66.
14. Coşkun, A. (2007). *Stratejik Performans Yönetimi ve Performans Karnesi,* İstanbul: Literatür Yayınları, 24-25.

15. Gautreau, A., and Kleiner, B. H. (2001). Recent Trends in Performance Measurement Systems-The Balanced Scorecard Approach, *Management Research News*, 24 (3): 153-156.
16. Bertan, S. (2009). Otel İşletmelerinde Yönetim Aracı Olarak Kurumsal Karne. *Journal of Yasar University*, 4(16), 2525-2538.
17. Furnham, A. (1991). Personality and occupational success: 16PF correlates of cabin crew performance. *Personality and Individual Differences*, 12(1), 87-90.
18. Nameghi, E. N., ve Ariffin, A. A. M. (2013). The measurement scale for airline hospitality: Cabin crew's performance perspective. *Journal of Air Transport Management*, 30, 1-9.
19. Gürbüz, F., Özbakır, L., ve Yapıcı, H. (2009). Türkiye’de bir havayolu işletmesine ait parça söküm raporlarına ilişkin veri madenciliği uygulaması. *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, 24(1).
20. Piatetsky-Shapiro, G. (1990). Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Workshop, *AI Magazine*, 11(5), 68-70.
21. Fayyad, U., Piatetsky-Shapiro, G., ve Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases, *AI Magazine* 17(3), 37-54.
22. Akpınar, H. (2000). Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği, *İ.Ü. İşletme Fakültesi Dergisi*, 29(11), 1-22.
23. Han, J., and Kamber, M. (2006). *Data Mining Concepts and Techniques* (2.Baskı) USA: Morgan Kaufmann Publisher, 2-7(70-71), 398-401.
24. Mitra, S., and Acharya, T. (2003). *Data mining : multimedia, soft computing , and bioinformatics*, John Wiley & Sons Publisher, 1-30.
25. Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., and Zanasi, A. (1998). *Discovering data mining: from concept to implementation*, Prentice Hall, Upper Saddle River, NJ, 517.
26. Jacobs, P. (1999). Data mining: what general managers need to know, *Harvard Management Update*, 4(10), 8.
27. Davis, B. (1999). Data mining transformed, *Information Week*, 751, 86.
28. Dumouchel, W. (1999). Bayesian data mining in large frequency tables, with an application to the FDA spontaneous, *The American Statistician*, 53(3), 177-190.
29. Hand, D. J. (1998). Data mining: statistics and more?, *The American Statistician*, 52, 112-118 .

30. Alataş, B., ve Akın, E. (2004). *Veri madenciliğinde yeni yaklaşımlar*, Ya/Em-2004-Yöneylem Araştırması/Endüstri Mühendisliği XXIV Ulusal Kongresi, Gaziantep-Adana.
31. Giudici, P. (2003), *Applied Data Mining: Statistical Methods for Business and Industry* 1st ed., John Wiley and Sons, England, 1-15, 85-110.
32. Sever, H., ve Oğuz, B. (2003). Veri Tabanlarında Bilgi Keşfine Formel Bir Yaklaşım Kısım I: Eşleştirme Sorguları ve Algoritmalar. *Bilgi Dünyası*, 3(2), 173-204.
33. Aytaç, M. B. (2013). *Doğrudan Pazarlama Aracı Olarak Tele Pazarlama İçin Veri Madenciliği Çözümleri: Banka Müşterileri Üzerine Bir Uygulama*. Yüksek Lisans Tezi, Gazi Üniversitesi Bilişim Enstitüsü, Ankara, 20-28.
34. Coşgun, E. ve Karağaoğlu, E. (2011). Veri Madenciliği Yöntemleriyle Mikrodizilim Gen ifade Analizi, *Hacettepe Tıp Dergisi*, 42, 180-189
35. Chapman, P., Clinton, J., Kerber, R., Khabaz T., Reinartz, T., Shearer, C. and Wirth, R. (2000). *CRISP-DM 1.0 Step-by-step data mining guide* SPSS, Inc.
36. Hui, S.C. ve Jha, G. (2000). Data mining for customer service support, *Information and Management*, 38, 1-13.
37. Olson, D. L., ve Delen, D. (2008). *Advanced Data Mining Techniques*, Springer, Berlin Heidelberg.
38. Yıldırım, S. C. (2015). Havayolu Taşımacılığında Rekabet Unsurlarının Değişkenliği ve Hizmet İnovasyonu İle Müşteri Memnuniyetinin Artırılması: Yeni Bir “Mil+ Süre Puanı” Uygulaması. *Journal of Business Research Turk*. 7(1), 340-359.
39. Dalkılıç, G., ve Türkmen, F. (2002). *Karınca Kolonisi Optimizasyonu*, Yüksek Performanslı Bilişim Sempozyumu, Kocaeli.
40. Kasap, E. (2007). *Sigortacılık sektöründe müşteri ilişkileri yönetimi yaklaşımıyla veri madenciliği teknikleri ve bir uygulama*, Yüksek Lisans Tezi, Marmara Üniversitesi, İstanbul, 35-37.
41. Wirth, R., ve Hipp, J. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining, Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining, Manchester, 29-39.
42. Bilen, H. (2009). *Bankacılık Sektöründe Personel Seçimi ve Performans Değerlendirmesine İlişkin Veri Madenciliği Uygulaması*, Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 19-20.
43. İnternet: Sivil Havacılık Genel Müdürlüğü. URL:<http://web.shgm.gov.tr/tr/havacilik-personeli/2138-kabin-memuru>.<http://www.webcitation.org/query?url=http%3A%2F%2Fweb.shgm.gov.tr%2Ftr%2Fhavacilik-personeli%2F2138-kabin-memuru&date=2019-03-19>, Son Erişim Tarihi: 10.02.2019

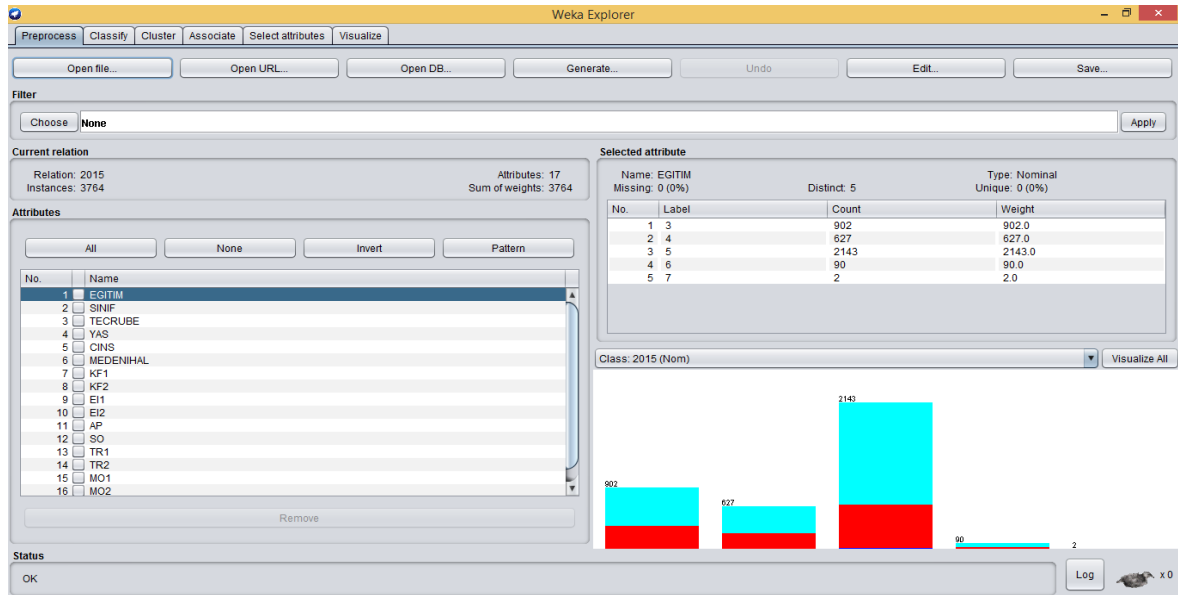
44. Tirenni, G., Kaiser, C., and Herrmann, A. (2007). Applying decision trees for value-based customer relations management: Predicting airline customers'future values. *Journal of Database Marketing and Customer Strategy Management*, 14(2), 130-142.
45. Yılmaz, A. K., Durak, M. Ş., Özdemir, E., Kiracı, K., ve Tanrıverdi, G. (2017). Eskişehir İlinde Havayolu Yolcu Potansiyelinin Belirlenmesi Üzerine Bir Karar Ağacı Uygulaması. *Anadolu University of Sciences & Technology-B: Theoretical Sciences*, 5(1).
46. Emekci, F., Sahin, O. D., Agrawal, D., and El Abbadi, A. (2007). Privacy preserving decision tree learning over multiple parties, *Data and Knowledge Engineering*, 63(2), 348-361.
47. Özekes, S., ve Çamurcu, A. Y. (2002). Veri Madenciliğinde Sınıflama ve Kestirim Uygulaması, *T.C. Marmara Üniversitesi Fen Bilimleri Dergisi*, 18, 159-174.
48. Kalikov, A. (2006). *Veri madenciliği ve bir e-ticaret uygulaması*, Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Ankara, 22-38.
49. Altıntaş, T. (2006). *Veri madenciliği metotlarından olan kümeleme algoritmalarının uygulamalı etkinlik analizi*, Yüksek Lisans Tezi, Sakarya Üniversitesi Fen Bilimleri Enstitüsü, Sakarya, 13-14.
50. Akbulut, S. (2006). *Veri Madenciliği Teknikleri ile Bir Kozmetik Markanın Ayrılan Müşteri Analizi ve Müşteri Segmentasyonu*, Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 20-23.
51. Hasan, A. (2011). Havayolu Şirketlerinde Müşteri İlişkileri Yönetimini (CRM) Desteklemek ve Müşteri Sadakatini Değerlendirmek İçin Veri Madenciliğinin Kullanılması. Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimler Enstitüsü, İstanbul, 31-32.
52. Şimşek, U. T. (2006). *Veri madenciliği ve müşteri ilişkileri yönetiminde (CRM) bir Uygulama*, Doktora Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul, 25-46.
53. Sağbaş, E. A., ve Ballı, S. (2016). Akıllı telefon algılayıcıları ve makine öğrenmesi kullanılarak ulaşım türü tespiti. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 22(5), 376-383.
54. Chandra, B, Gupta, M, and Gupt, M. P. (2007). Robust approach for estimating probabilities in Naive-Bayes classifier. *Pattern Recognition and Machine Intelligence*. Kolkata, India,18-22.
55. Argüden, Y., ve Erşahin, B. (2008). *Veri Madenciliği Veriden Bilgiye, Masraftan Değere*. İstanbul: ARGE Danışmanlık A.Ş., 22-23.
56. Chien, C. F., and Chen, L.F. (2008). Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry, *Expert Systems With Applications*, 34(1), 280-290.

57. Kasap, E. (2007). *Sigortacılık Sektöründe Müşteri İlişkileri Yönetimi Yaklaşımıyla Veri Madenciliği Teknikleri ve Bir Uygulama*, Yüksek Lisans Tezi, Marmara Üniversitesi Bankacılık ve Sigortacılık Enstitüsü, İstanbul, 50- 52.
58. Seow, H.V., and Thomas, L.C. (2007). *To ask or not to ask, that is the question*, *European Journal of Operational Research*, 183, 1513 – 1520.
59. Nisbet, R., Elder, J., and Miner, G. (2009). *Handbook of Statistical Analysis and Data Mining Applications*, Elsevier Inc, Burlington.
60. Öztemel, E. (2003). *Yapay Sinir Ağları*, İstanbul: P apatya Yayınları, 25-27.
61. Aydoğan, E.K. ve Gencer, C. (2007). Veri Madenciliği Problemlerinde Kaba Küme Yaklaşımı Kullanılarak Sınıflandırma Amaçlı Yapılmış Olan Çalışmalar, *Kara Harp Okulu Savunma Bilimleri Dergisi*, 6(2), 17-32.
62. Emel, G. G., ve Taşkın, Ç. (2002). Genetik Algoritmalar ve Uygulama Alanları. *Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, XXI(1), 129-152.
63. Zhang, C., and Zhang, S. (2002). *Association Rule Mining, Models and Algorithms*. Springer.
64. Çil, F. (2010). *Banka yatırım fonu müşteri hareketlerinin belirlenmesine yönelik bir veri madenciliği uygulaması*. Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Ankara, 28-70.
65. Aaker, D. A., Kumar, V., Day, G. S., ve Leone, R. P. (2011). *Marketing Research, International Student Version*, 10th edition, Wiley, Hoboken, NJ, 618,445.
66. Hsu, M. H. (2008). A personalized English learning recommender system for ESL students, *Expert Systems With Applications*, 34, 683-688.
67. Dolgun, M. Ö. (2006). *Büyük alışveriş merkezleri için veri madenciliği uygulamaları*, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 23-27.
68. Apan, A. (2008). Yeni Kamu İşletimi ve Performans Yönetimi, *Türk İdare Dergisi*, (460), 60
69. Chiang, W. Y. (2012). Applying a new model of customer value on international air passengers' market in Taiwan. *International Journal of Tourism Research*. 14(2), 116-123.
70. Yenice, E. (2006). Kamu Kesiminde Performans Ölçümü ve Bütçe İlişkisi. *Sayıştay Dergisi*, 61, 57-68.
71. Chen, Y. L., Tang, K., Shen, R. J., and Hu, Y. H. (2005). *Market basket analysis in a multiple store environment Decision Support Systems*, 40(2), 339-354.
72. Özekes, S. (2001). Veri Madenciliği Modelleri ve Uygulama Alanları, *İstanbul Ticaret Üniversitesi Dergisi*, 2003, 65-82.

73. Ağca, V. (2005). *İç girişimcilik yapısı ve firma performansına etkileri: Denizli tekstil sektöründeki firmalarda bir araştırma*. Doktora Tezi, Afyon Kocatepe Üniversitesi Sosyal Bilimler Enstitüsü, Afyon, 167-168.
74. Zaki, M. J. (1999). Parallel and Distributed Association Mining: A Survey, *IEEE Concurrency, special issue on Parallel Mechanisms for Data Mining*, 7(5), 14-25.
75. Plasse, M., Niang, N., Saporta, G., Villeminot, A., and Leblond, L. (2007). *Combined use of association rules mining and clustering methods to find relevant links between binary rare attributes in a large data set*, *Computational Statistics and Data Analysis*, 52, 596- 613.
76. Savaş, S., Topaloğlu, N., ve Yılmaz, M. (2012). *Veri madenciliği ve Türkiye’deki Uygulama Örnekleri*.
77. Çankırı, S., Kartal, E., Yıldırım, K., ve Gülseçen, S. (2009). *Organizasyonlarda Bilgi Yönetimi Sürecinde Veri Madenciliği Yaklaşımı*. ÜNAK ’09 Bilgi Çağında Varoluş: “Fırsatlar ve Tehditler” Sempozyumu, İstanbul, 148-167.
78. Gülce, A. C. (2010). *Veri madenciliğinde apriori algoritması ve apriori algoritmasının farklı veri kümelerinde uygulanması*. Yüksek Lisans Tezi. Trakya Üniversitesi Fen Bilimleri Enstitüsü, Edirne, 14-27.



EK-1. WEKA explorer ekran görüntüsü



EK-2. Çalışmada ilk kullanılan veri türleri

@relation	2015
@attribute	EGITIM {3,4,5,6,7}
@attribute	SINIF {Z,E,Y,N,Q}
@attribute	TECRUBE {1-5,6-10,11-15,16-20}
@attribute	YAS {20-29,30-39,40-49,50}
@attribute	CINS {Bay,Bayan}
@attribute	MEDENIHAL {Evli,Bekar}
@attribute	KF1 {13,14,15,16,17,18,19}
@attribute	KF2 {13,14,15,16,17,18,19}
@attribute	EI1 {13,14,15,16,17,18,19}
@attribute	EI2 {12,13,14,15,16,17,18,19}
@attribute	AP {12,13,14,15,16,17,18,19}
@attribute	SO {12,13,14,15,16,17,18,19}
@attribute	TR1 {12,13,14,15,16,17,18,19}
@attribute	TR2 {12,13,14,15,16,17,18,19}
@attribute	MO1 {13,14,15,16,17,18,19}
@attribute	MO2 {12,13,14,15,16,17,18,19}
@attribute	2015 {A,B1,B2,C1,C2,D1}
@data	
	3,E,11-15,50,Bayan,Evli,16,16,16,16,16,16,16,16,16,16,B2
	3,E,11-15,40-49,Bayan,Evli,16,17,16,16,16,16,16,16,16,16,B2

EK-3. Çalışmada ilk kullanılan veri türleri sınıflandırma algoritma sonuçlarının karşılaştırılması ve J48 karar ağacı algoritmasının boyutları

		Doğru sınıflandırılan kayıt (%)	Kappa istatistiği	Ort. mutlak hata	Ort. mutlak hata karekök	Görelî mutlak hata (%)	Görelî hata karekök (%)
J48		92,4814	0,8468	0,0368	0,1444	21,8053	49,7294
PART		91,6578	0,8323	0,0345	0,1486	20,4418	51,2069
NaiveBayes		86,6366	0,7597	0,0443	0,2048	26,2801	70,5389
OneR		89,9309	0,7919	0,0336	0,1832	19,8924	63,1113
ZeroR		62,221	0	0,1687	0,2903	100	100

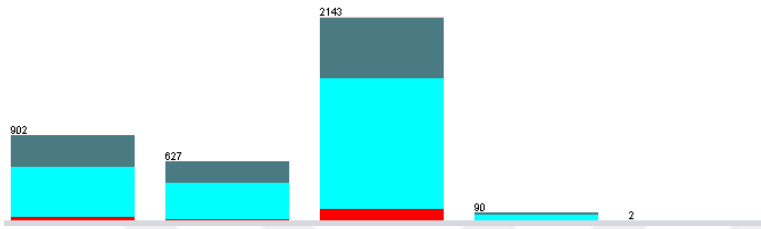
Number of Leaves : 186

Size of the tree : 215

EK-4. Çalışmada yer alan değişkenlerin WEKA görselleri

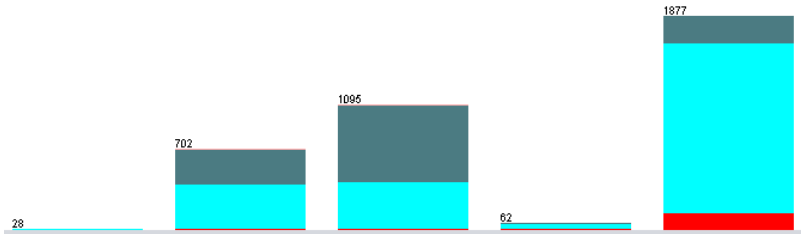
Name: EGITIM		Type: Nominal	
Missing: 0 (0%)		Unique: 0 (0%)	
		Distinct: 5	
No.	Label	Count	Weight
1	3	902	902.0
2	4	627	627.0
3	5	2143	2143.0
4	6	90	90.0
5	7	2	2.0

Class: 2015 (Nom) Visualize All



Name: SINIF		Type: Nominal	
Missing: 0 (0%)		Unique: 0 (0%)	
		Distinct: 5	
No.	Label	Count	Weight
1	Z	28	28.0
2	E	702	702.0
3	Y	1095	1095.0
4	N	62	62.0
5	Q	1877	1877.0

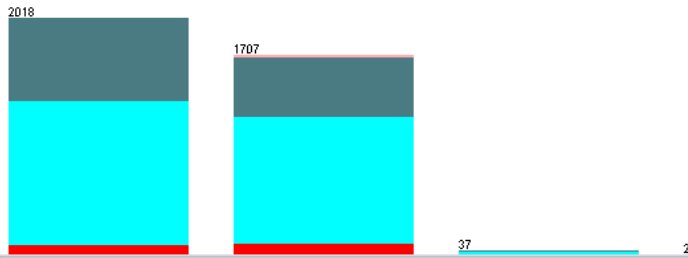
Class: 2015 (Nom) Visualize All



EK-4. (devam) Çalışmada yer alan değişkenlerin WEKA görselleri

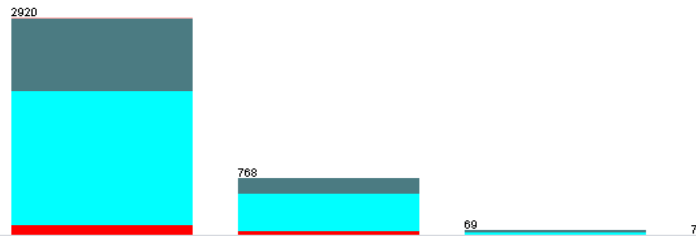
Name: YAS		Type: Nominal	
Missing: 0 (0%)		Unique: 0 (0%)	
Distinct: 4			
No.	Label	Count	Weight
1	20-29	2018	2018.0
2	30-39	1707	1707.0
3	40-49	37	37.0
4	50	2	2.0

Class: 2015 (Nom) Visualize All



Name: TECRUBE		Type: Nominal	
Missing: 0 (0%)		Unique: 0 (0%)	
Distinct: 4			
No.	Label	Count	Weight
1	1-5	2920	2920.0
2	6-10	768	768.0
3	11-15	69	69.0
4	16-20	7	7.0

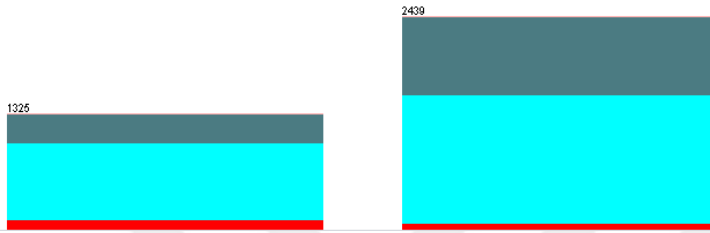
Class: 2015 (Nom) Visualize All



EK-4. (devam) Çalışmada yer alan değişkenlerin WEKA görselleri

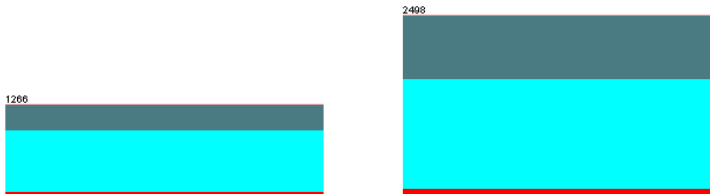
Name: CINS		Type: Nominal	
Missing: 0 (0%)		Unique: 0 (0%)	
Distinct: 2			
No.	Label	Count	Weight
1	Bay	1325	1325.0
2	Bayan	2439	2439.0

Class: 2015 (Nom) Visualize All

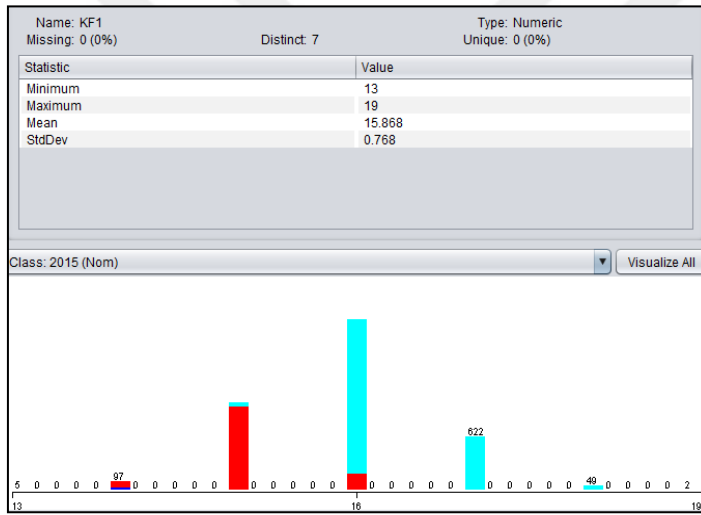
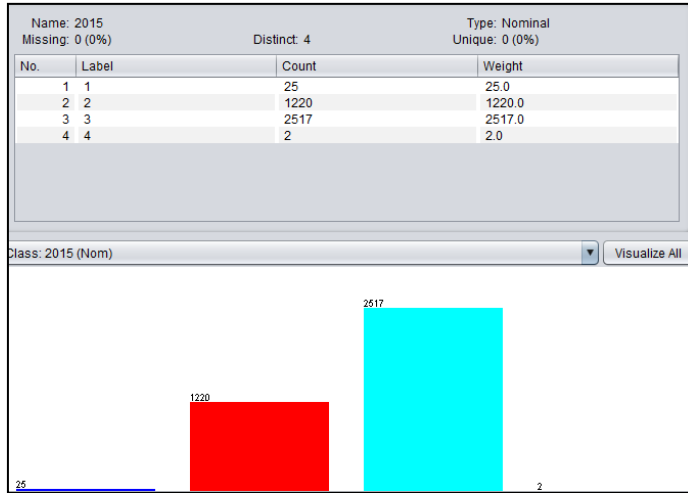


Name: MEDENIHAL		Type: Nominal	
Missing: 0 (0%)		Unique: 0 (0%)	
Distinct: 2			
No.	Label	Count	Weight
1	Evli	1266	1266.0
2	Bekar	2498	2498.0

Class: 2015 (Nom) Visualize All



EK-4. (devam) Çalışmada yer alan değişkenlerin WEKA görselleri



EK-6. Random Forest algoritması (50 Ağaç)

```

=== Summary ===

Correctly Classified Instances      3580           95.1116 %
Kappa statistic                    0.8904
Mean absolute error                 0.0334
Root mean squared error             0.136
Relative absolute error             14.9037 %
Root relative squared error         40.6567 %
Total Number of Instances          3764

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
      0,640    0,002    0,667    0,640    0,653      0,651    0,937    0,649     1
      0,919    0,033    0,930    0,919    0,925      0,889    0,985    0,968     2
      0,970    0,074    0,964    0,970    0,967      0,900    0,988    0,992     3
      0,500    0,000    1,000    0,500    0,667      0,707    1,000    1,000     4
Weighted Avg.   0,951    0,060    0,951    0,951    0,951      0,894    0,987    0,982

=== Confusion Matrix ===

  a    b    c    d  <-- classified as
16     9     0     0 |    a = 1
 8 1121    91     0 |    b = 2
 0   75 2442    0 |    c = 3
 0     0     1     1 |    d = 4

```

EK-7. Random Forest algoritması (100 Ağaç)

```

=== Summary ===

Correctly Classified Instances      3576           95.0053 %
Kappa statistic                     0.8881
Mean absolute error                 0.0334
Root mean squared error             0.1358
Relative absolute error             14.9075 %
Root relative squared error         40.5849 %
Total Number of Instances          3764

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          0,600    0,001    0,750     0,600    0,667      0,669    0,956     0,661     1
          0,921    0,036    0,925     0,921    0,923      0,886    0,987     0,970     2
          0,968    0,074    0,964     0,968    0,966      0,896    0,989     0,993     3
          0,500    0,000    1,000     0,500    0,667      0,707    1,000     1,000     4
Weighted Avg.   0,950    0,061    0,950     0,950    0,950      0,891    0,988     0,983

=== Confusion Matrix ===

  a    b    c    d  <-- classified as
15    10     0     0 |    a = 1
 5 1124    91     0 |    b = 2
 0    81 2436     0 |    c = 3
 0     0     1     1 |    d = 4

```

EK-8. Random Forest algoritması (150 Ağaç)

```

=== Summary ===

Correctly Classified Instances      3574          94.9522 %
Kappa statistic                    0.887
Mean absolute error                 0.0334
Root mean squared error            0.1358
Relative absolute error            14.9029 %
Root relative squared error        40.6041 %
Total Number of Instances         3764

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          0,640   0,002   0,696     0,640   0,667     0,665   0,957    0,659    1
          0,920   0,036   0,924     0,920   0,922     0,885   0,987    0,970    2
          0,967   0,073   0,964     0,967   0,965     0,896   0,989    0,993    3
          0,500   0,000   1,000     0,500   0,667     0,707   1,000    1,000    4
Weighted Avg.   0,950   0,061   0,949     0,950   0,949     0,891   0,988    0,983

=== Confusion Matrix ===

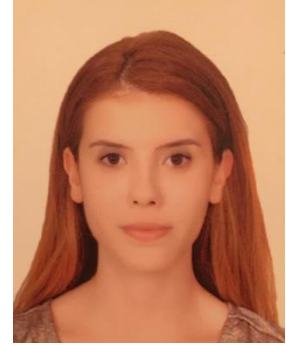
  a    b    c    d  <-- classified as
16     9     0     0 |    a = 1
 7 1123    90     0 |    b = 2
 0   83 2434     0 |    c = 3
 0     0     1     1 |    d = 4

```

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı, adı : YILDIZ KORKMAZ, Büşra
 Uyruğu : T.C.
 Doğum tarihi ve yeri : 05.12.1990, Adana
 Medeni hali : Evli
 E-mail : busraydz@gmail.com



Eğitim

Derece	Eğitim Birimi	Mezuniyet Tarihi
Yüksek Lisans	Gazi Üniversitesi/Endüstri Mühendisliği	Devam Ediyor
Lisans	Süleyman Demirel Üniversitesi/ Endüstri Mühendisliği	2013
Lise	Giresun Fen Lisesi	2007

İş Deneyimi

Yıl	Yer	Görev
2015-Halen	Türk Hava Yolları AO	Uzman
2015	Gırsan Makine ve Hafif Silah Sanayi A.Ş.	Endüstri Mühendisi

Yabancı Dil

İngilizce

Yayınlar

1. Yıldız Korkmaz, B., ve Özcan, U. (2018). Bir Havayolu İşletmesinde Kabin Ekiplerinin Performans Değerlendirmesine İlişkin Bir Veri Madenciliği Uygulaması. *Karadeniz Fen Bilimleri Dergisi*, 8(2), 154-170.

Hobiler

Yüzme, Tenis, Koleksiyon



GAZİ GELECEKTİR..