

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

UÇAK PARÇALARI BAKIM VE TAMİR
FİRMASINDA KARLILIK DİNAMİKLERİNİN
VERİ BİLİMİ YÖNTEMLERİYLE İNCELENMESİ

Nur Şahver USLU

DOKTORA TEZİ
İstatistik Anabilim Dalı
İstatistik Programı

Danışman
Prof. Dr. Ali Hakan BÜYÜKLÜ

Mayıs, 2025

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

UÇAK PARÇALARI BAKIM VE TAMİR FİRMASINDA
KARLILIK DİNAMİKLERİNİN VERİ BİLİMİ
YÖNTEMLERİYLE İNCELENMESİ

Nur Şahver USLU tarafından hazırlanan tez çalışması 26.05.2025 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı İstatistik Programı **DOKTORA TEZİ** olarak kabul edilmiştir.

Prof. Dr. Ali Hakan BÜYÜKLÜ
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Prof. Dr. Ali Hakan BÜYÜKLÜ, Danışman
Yıldız Teknik Üniversitesi

Prof. Dr. Karun NEMLİOĞLU, Üye
İstanbul Üniversitesi

Prof. Dr. Ersoy ÖZ, Üye
Yıldız Teknik Üniversitesi

Prof. Dr. Semra ERPOLAT TAŞABAT, Üye
Mimar Sinan Güzel Sanatlar Üniversitesi

Doç. Dr. Elif TUNA, Üye
Yıldız Teknik Üniversitesi

Danışmanım Prof. Dr. Ali Hakan BÜYÜKLÜ sorumluluğunda tarafımda hazırlanan Uçak Parçaları Bakım ve Tamir Firmasında Karlılık Dinamiklerinin Veri Bilimi Yöntemleriyle İncelenmesi başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Nur Şahver USLU

İmza



Aileme



TEŞEKKÜR

Bu uzun ve yoğun çalışma döneminde bilgi ve tecrübeleriyle bana yol gösteren ve çalışmalarımın şekillenmesinde katkı sağlayan danışmanım Prof. Dr. Ali Hakan Büyüklü'ye teşekkürlerimi sunarım. Ayrıca, tezimin değerlendirme sürecinde yapıcı eleştirileri ve katkılarıyla çalışmamın niteliğini artıran jüri üyem Prof. Dr. Karun Nemlioğlu'na ve Çok Değişkenli Yöntemleri bana öğreten ve sevdiren diğer jüri üyem Dr. Öğr. Üyesi Doğan YILDIZ'a teşekkür ederim.

Bu süreçte moral ve motivasyon kaynağım olan aileme derin bir minnet duygusuyla teşekkür etmek istiyorum. Özellikle sevgili annem Sevim Turangil, kayınvalidem Müesser Uslu ve babam Rıfat Turangil, fedakarlıklarınız ve sevginiz, bu zorlu sürecin üstesinden gelmemde en büyük destek oldu.

Sevgili eşim Selman Uslu, havacılık sektöründeki tecrübelerini benimle paylaşarak tezime sağladığınız katkılar ve bu süreç boyunca bana gösterdiğiniz anlayış için sonsuz teşekkürler. Sevgili çocuklarım Ayşe Neslişah Uslu ve Rıfat Halim Uslu, sabrınız ve sevginiz bu yolculuğu benim için çok daha anlamlı kıldı.

Son olarak, akademik hayatım boyunca birlikte çalıştığım tüm hocalarıma teşekkür etmek isterim. Özellikle bu süreçte bana ilham veren ve akademik gelişimime katkı sağlayan değerli hocalarım Doç. Dr. Atıf Ahmet EVREN ve Doç. Dr. Elif TUNA'ya şükranlarımı sunarım.

Hepinize minnettarım.

Nur Şahver USLU

İÇİNDEKİLER

SİMGE LİSTESİ	viii
KISALTMA LİSTESİ	ix
ŞEKİL LİSTESİ	x
TABLO LİSTESİ	xii
ÖZET	xiv
ABSTRACT	xvi
1 GİRİŞ	1
1.1 Literatür Özeti	1
1.2 Tezin Amacı	3
2 VERİ, KULLANILAN YÖNTEMLER VE TEMEL TANIMLAYICI İSTATİSTİKLER	5
2.1 Veri ve Veri Hazırlığı	5
2.1.1 Değişken Tanımları	5
2.1.2 Veri Hazırlığı	6
2.2 Kullanılan Yöntemler	11
2.2.1 İnteraktif Gruplama	11
2.2.2 Değişken Kümeleme	12
2.2.3 Gradient Boosting Algoritması	13
2.2.4 Lojistik Regresyon	14
2.3 Temel Tanımlayıcı İstatistikler	16
2.3.1 Veri Seti Olay Dağılımı ve Korelasyon Matrisi	16
2.3.2 Tanımlayıcı İstatistikler	19
2.3.3 Standartlaştırılmış ve Gruplandırılmış Değişkenler için Kutu Grafikleri	20
2.3.4 Gruplanmış Değişkenler ve Kar Marjı İncelemesi	29
2.3.5 Standartlaştırılmış ve Gruplanmış Değişkenler için İlişki Matrisi	37

2.3.6	Standartlaştırılmış ve Gruplanmış Değişkenler Arasında İlişki ve Yoğunluk Analizi	39
3	UYGULAMA	43
3.1	İş İhtiyaçlarını ve Veriyi Anlama	43
3.2	Veri Hazırlama	44
3.3	Veri Seçimi - İnteraktif Gruplama Uygulaması	44
3.3.1	Parça Tanımı - Part Description	44
3.3.2	Parça Maliyeti - Parts Billed Current	46
3.3.3	İşçilik Maliyeti - Labour Billed Current	47
3.3.4	İlgili Yönetici - Manager	49
3.3.5	Çevrim Süresi - Diff Shipping Entry	50
3.3.6	İlgili Bölüm - Department	51
3.3.7	Fiyat Teklif Süresi - Diff Quote Entry	53
3.3.8	Parça Durumu - Status	54
3.3.9	İnteraktif Gruplama Sonuçlarının Özeti	55
3.4	Değişken Kümeleme Uygulaması	57
3.5	Gradient Boosting Uygulaması	59
3.5.1	Standartlaştırılmış Değişkenler için Gradient Boosting ile Sürekli Hedef Değişken Uygulaması	62
3.5.2	Gruplandırılmış Değişkenler için Gradient Boosting ile İkili Hedef Değişken Uygulaması	64
3.6	Lojistik Regresyon Uygulaması	69
3.7	Lojistik Regresyon - Hava Yolu Firmaları ve Tedarikçi Firmalar İçin Uygulama	80
3.7.1	Hava Yolu Firmaları İçin Lojistik Regresyon Analizi Uygulaması	80
3.7.2	Tedarikçi Firmalar İçin Lojistik Regresyon Analizi Uygulaması	87
4	SONUÇ	94
	KAYNAKÇA	101
	TEZDEN ÜRETİLMİŞ YAYINLAR	104

SİMGE LİSTESİ

$R_{nearest}^2$	En Yakın Kümeye Kare Uzaklığı
x_i	Girdi Verisi
$\alpha \nabla_{\theta}$	Gradyan Güncellemesi
y_i	Hedef Değeri
θ	Karar Ağacı Parametreleri
R_{own}^2	Kendi Kümesine Kare Uzaklığı
$logit(\pi_i)$	Lojistik Model
$\frac{\pi_i}{1-\pi_i}$	Odds Oranı (Olasılık Oranı)
$n_i^{non-event}$	Olay Olmayanların Sayısı
n_i^{event}	Olay Sayısı
R_{ratio}^2	Oransal Kare Uzaklık
α	Öğrenme Hızı
$\hat{y}_i$	x_i için tahmin değeri

KISALTMA LİSTESİ

AIC	Akaike Information Criterion (Akaike Bilgi Kriteri)
ASE	Average Squared Error (Ortalama Hata Karesi)
IQR	Çeyrekler Arası Aralık
D	Doğrulama veri seti
E	Eğitim veri seti
IFS	Input Feature Selection (Girdi Değişkeni Seçimi)
KOBİ	Küçük ve Orta Ölçekli İşletmeler
MCR	Missclassification Rate (Yanlış Sınıflama Oranı)
ML	Machine Learning (Makine Öğrenmesi)
MRO	Maintenance, Repairment, Overhaul (Bakım, Tamir, Onarım)
OEMs	Original Equipment Manufacturers (Orijinal Parça Üreticileri)
RASE	Root Average Squared Error (Ortalama Hata Kareleri Kökü)
ROC	Receiver Operating Characteristic
TAT	Turn Around Time (Çevrim Süresi)
TTQ	Time to Quotation (Fiyat Teklif Süresi)

ŞEKİL LİSTESİ

Şekil 2.1	Veri seti hedef değişken dağılımı	17
Şekil 2.2	Standartlaştırılmış değişkenler için korelasyon matrisi	18
Şekil 2.3	Standartlaştırılmış parça maliyeti - kutu grafiği	21
Şekil 2.4	Gruplanmış parça maliyeti - kutu grafiği	23
Şekil 2.5	Standartlaştırılmış işçilik maliyeti - kutu grafiği	24
Şekil 2.6	Gruplanmış işçilik maliyeti - kutu grafiği	25
Şekil 2.7	Standartlaştırılmış çevrim süresi - kutu grafiği	26
Şekil 2.8	Gruplanmış çevrim süresi - kutu grafiği	27
Şekil 2.9	Standartlaştırılmış fiyat teklif süresi - kutu grafiği	28
Şekil 2.10	Gruplanmış fiyat teklif süresi - kutu grafiği	29
Şekil 2.11	Fiyat teklif süresi ile hedef değişken - kutu grafiği	30
Şekil 2.12	Çevrim süresi ile hedef değişken - kutu grafiği	32
Şekil 2.13	İşçilik maliyeti ile hedef değişken - kutu grafiği	33
Şekil 2.14	Parça durumu ile hedef değişken - kutu grafiği	34
Şekil 2.15	Parça tanımı ile hedef değişken - kutu grafiği	35
Şekil 2.16	Yönetici ile hedef değişken - kutu grafiği	36
Şekil 2.17	Departman ile hedef değişkeni - kutu grafiği	37
Şekil 2.18	Standartlaştırılmış değişkenler arası ilişki matrisi	38
Şekil 2.19	Yoğunluk dağılımlı standartlaştırılmış parça maliyeti ve işçilik maliyeti regresyon grafiği	40
Şekil 2.20	Yoğunluk dağılımlı gruplanmış parça maliyeti ve işçilik maliyeti regresyon grafiği	41
Şekil 3.1	Gruplanmış değişkenler için Gini indeksi	55
Şekil 3.2	Değişkenler için farklı gruplar arasında olay oranlarının karşılaştırılması	56
Şekil 3.3	Parça tanımı değişkeni için grup olay oranlarının karşılaştırılması	57
Şekil 3.4	Değişken Kümeleme diyagramı	58
Şekil 3.5	Değişkenlerin dendrogramı	59
Şekil 3.6	Boost79 - önem değeri saçılım grafiği	64
Şekil 3.7	Boost86 - farklı Gradient Boosting modelleri için ROC Eğrileri	65
Şekil 3.8	Yanlış sınıflandırma oranı ve yineleme sayısı	66

Şekil 3.9	Boost86 - farklı Gradient Boosting modellerinde yanlış sınıflandırma oranları	67
Şekil 3.10	Boost86 - önem değeri saçılım grafiği	68
Şekil 3.11	Lojistik Regresyon - kümülatif kazanç grafiği	71
Şekil 3.12	Lojistik Regresyon - kümülatif yüzdelik dilimi grafiği	73
Şekil 3.13	Lojistik Regresyon - skor dağılım grafiği	75
Şekil 3.14	Lojistik Regresyon - mutlak katsayı grafiği	76
Şekil 3.15	Lojistik Regresyon - tahmin seçim grafiği	78
Şekil 3.16	Lojistik Regresyon - model seçim adım sayısına göre ortalama kare hata değişimi	79
Şekil 3.17	Lojistik Regresyon havayolu - kümülatif kazanç grafiği	82
Şekil 3.18	Lojistik Regresyon havayolu - yanlış sınıflandırma oranları grafiği	83
Şekil 3.19	Lojistik Regresyon havayolu - kümülatif kazanç ve temel kümülatif kazanç karşılaştırması	84
Şekil 3.20	Lojistik Regresyon havayolu - kümülatif yüzdelik dilimi grafiği .	86
Şekil 3.21	Lojistik Regresyon tedarikçi - kümülatif kazanç grafiği	89
Şekil 3.22	Lojistik Regresyon tedarikçi - olasılık oranı grafiği	90
Şekil 3.23	Lojistik Regresyon tedarikçi - yanlış sınıflandırma oranları grafiği	91
Şekil 3.24	Lojistik Regresyon tedarikçi - kümülatif kazanç ve temel kümülatif kazanç karşılaştırması	92
Şekil 3.25	Lojistik Regresyon tedarikçi - kümülatif yüzdelik dilimi grafiği . .	93

TABLO LİSTESİ

Tablo 2.1	Uzun değişken listesi	8
Tablo 2.2	Kısa değişken listesi	10
Tablo 2.3	Nümerik değişkenler için tanımlayıcı istatistikler	19
Tablo 2.4	Kategorik değişkenler için tanımlayıcı istatistikler	20
Tablo 3.1	Parça tanımının gruplara ayrılması	45
Tablo 3.2	Parça maliyetinin gruplara ayrılması	46
Tablo 3.3	İşgücü maliyetinin gruplara ayrılması	48
Tablo 3.4	Yöneticilerin gruplara ayrılması	49
Tablo 3.5	Çevrim süresinin gruplara ayrılması	50
Tablo 3.6	İlgili bölümlerin gruplara ayrılması	52
Tablo 3.7	Fiyat teklif sürelerinin gruplara ayrılması	53
Tablo 3.8	Parça durumunun gruplara ayrılması	54
Tablo 3.9	Boost86 - eğitim veri seti için karmaşıklık matrisi	60
Tablo 3.10	Boost86 - doğrulama veri seti için karmaşıklık matrisi	61
Tablo 3.11	Boost86 - eğitim veri seti üzerindeki model için performans metrikleri	61
Tablo 3.12	Boost86 - doğrulama veri seti üzerindeki model için performans metrikleri	62
Tablo 3.13	Boost 79 - eğitim ve doğrulama veri seti için uyum istatistikleri - sürekli hedef değişken	63
Tablo 3.14	Boost79 - değişken önemi	63
Tablo 3.15	Boost86 - eğitim ve doğrulama veri seti için uyum istatistikleri - ikili hedef değişken	65
Tablo 3.16	Boost86 - değişken önemi	68
Tablo 3.17	Lojistik Regresyon - eğitim ve doğrulama veri setleri için yanlış sınıflama oranı	69
Tablo 3.18	Lojistik Regresyon modeline giren değişkenler	70
Tablo 3.19	Lojistik Regresyon havayolu - eğitim ve doğrulama veri setleri için uyum istatistikleri	80
Tablo 3.20	Lojistik Regresyon havayolu - gruplanmış değişkenler için adım adım etki giriş istatistikleri	81

Tablo 3.21	Lojistik Regresyon tedarikçi - eğitim ve doğrulama veri setleri için uyum istatistikleri	87
Tablo 3.22	Lojistik Regresyon tedarikçi - adım adım etki giriş istatistikleri .	88



ÖZET

Uçak Parçaları Bakım ve Tamir Firmasında Karlılık Dinamiklerinin Veri Bilimi Yöntemleriyle İncelenmesi

Nur Şahver USLU

İstatistik Anabilim Dalı

Doktora Tezi

Danışman: Prof. Dr. Ali Hakan BÜYÜKLÜ

Bu çalışma, bir küçük ve orta ölçekli işletme (KOBİ) MRO (Bakım, Onarım ve Revizyon) şirketinin 2013-2021 yılları arasındaki operasyonel verilerini analiz ederek havacılık sektöründe kâr marjı dinamiklerini anlamayı ve bu dinamiklerin stratejik karar süreçlerine entegrasyonunu sağlamayı amaçlamaktadır. Kâr marjını etkileyen değişkenlerin istatistiksel olarak incelenmesi, gelişmiş analitik yetkinliklere sahip büyük firmalara karşı rekabet avantajı elde etmek isteyen KOBİ MRO şirketleri için büyük önem taşımaktadır. Çalışmada, veri temizleme, dönüştürme ve eksik veri yönetimi süreçlerinin yanı sıra değişkenlerin anlamı, erişilebilirliği ve güvenilirliği ayrıntılı biçimde değerlendirilmiştir. Ayrıca, parça tamiri süreçlerini temsil eden Fiyat Teklif Süresi (TTQ) ve Çevrim Süresi (TAT) gibi yeni değişkenler oluşturularak analiz kapsamı genişletilmiştir. Bu süreçler, benzer ölçekteki firmalar için rehber niteliğinde yöntemsel bir katkı da sunmaktadır.

Analitik modelleme kapsamında Gradient Boosting ve Lojistik Regresyon yöntemleri kullanılmış; kâr marjını etkileyen değişkenler ile değişkenlerin önem düzeyleri ayrıntılı biçimde incelenmiştir. Müşteriler havayolu firmaları ve tedarikçi firmalar olarak iki gruba ayrılmış ve her grup için ayrı Lojistik Regresyon analizleri gerçekleştirilmiştir. İnteraktif Gruplama yöntemi ile değişkenler gruplandırılmış, Gini endeksi değerlerine dayalı olarak değişken seçimi yapılmış ve model doğruluğu artırılmıştır. Bulgular, kâr marjını artırmada parça tanımları, maliyetler, departmanlar ve atölye standardizasyonunun kritik olduğunu göstermektedir. Havayolu firmalarında ise kâr marjının artırılması için operasyonel süreç iyileştirmelerinin maliyet kontrolü ile birlikte ele alınması

gerektiđi sonucuna varılmıřtır. Deđiřken K meleme analizi ile deđiřkenler arası iliřkiler deđerlendirilmiř, bu bulgular model  ıktıları ile birlikte stratejik  nerilerin geliřtirilmesinde kullanılmıřtır.  alıřmanın, havacılık sekt r ndeki KOB  MRO'lara veri temelli karar alma s re lerinde yol g sterici olması hedeflenmektedir.

Anahtar Kelimeler: Havacılık - MRO, gradient boosting, interaktif gr plama, lojistik regresyon, veri bilimi



ABSTRACT

Analysis of Profitability Dynamics in an Aircraft Parts Maintenance and Repair Company with Data Science Methods

Nur Şahver USLU

Department of Statistics
Doctor of Philosophy Thesis

Supervisor: Prof. Dr. Ali Hakan BÜYÜKLÜ

This study aims to analyze the operational data of a small and medium-sized enterprise (SME) MRO (Maintenance, Repair, and Overhaul) company from the period 2013 to 2021, in order to understand profit margin dynamics in the aviation sector and to integrate these dynamics into strategic decision-making processes. The statistical examination of variables affecting profit margins is of great importance for SME MRO companies seeking to gain a competitive advantage against large firms equipped with advanced analytical capabilities. In the study, detailed evaluations were conducted regarding data cleaning, transformation, missing data management, and the meaning, accessibility, and reliability of variables. Additionally, new variables representing part repair processes, namely Time to Quotation (TTQ) and Turnaround Time (TAT), were derived to broaden the analytical scope. These processes also provide methodological guidance for companies of similar scale.

Gradient Boosting and Logistic Regression methods were employed within the analytical framework to examine the factors influencing profit margins and assess the relative importance of variables. Customers were categorized into airline and non-airline groups, and separate Logistic Regression analyses were performed for each. Interactive Grouping was applied to cluster variables, Gini index values were used for variable selection, and model accuracy was improved. The findings indicate that part descriptions, costs, departments, and workshop standardization are critical for increasing profitability. For airline customers, profit margin

improvements require not only cost control but also enhancements in operational processes. For non-airline customers, operational processes did not show a significant impact on profit margins.

Finally, Variable Clustering was conducted to analyze relationships between variables, and these findings were integrated with model outputs to formulate strategic recommendations. The study aims to serve as a guide for SME MRO companies in the aviation sector to support data-driven decision-making and enhance their competitiveness through analytics.

Keywords: Aviation - MRO, gradient boosting, interactive grouping, logistic regression, data science



1.1 Literatür Özeti

Bakım, bir sistemin veya cihazın performansını iyileştirme ve eski haline getirme süreçlerini kapsar ve zamanla bozulmaya uğrayan sistemin beklenen işletim seviyelerine geri dönmesini sağlar [1]. Uçak bakımının temel amacı, sistem arızalarından kaynaklanan operasyonel aksaklıkları en aza indirerek uçuş operasyonları için uçağın güvenliğini ve kullanılabilirliğini sağlamaktır. Bu hedef, bakım kaynaklarının en iyi şekilde kullanılmasıyla ekonomik olarak en etkin biçimde gerçekleştirilir [2]. Havacılıkta bu kapsam, özel olarak bakım, onarım ve revizyon (MRO) endüstrisi olarak adlandırılmaktadır. MRO endüstrisi, uçakların güvenli ve verimli bir şekilde işletilmesini sağlamak için bakım ve onarım hizmetleri sunan Airbus ve Boeing gibi sektör devlerinden daha küçük MRO firmalarına kadar geniş bir yelpazeyi kapsar.

MRO ortamı, küresel ekonomi, jeopolitik olaylar, teknik sorunlar, çevresel faktörler ve pandemiler gibi makro faktörlere oldukça duyarlıdır. Pandemi sonrası zorluklar ve jeopolitik gerilimler tarafından şekillense de, havayolları, MRO'lar ve orijinal ekipman üreticileri (OEM'ler) için önemli engeller devam etmektedir [3]. MRO endüstrisinin 2024 yılı itibarıyla toparlanması ve küresel havacılık MRO pazarının COVID-19 pandemisinden kurtulma yolunda önemli ilerlemeler kaydetmesi beklenmektedir. Oliver Wyman'a göre, ticari uçak MRO pazar büyüklüğü 2023 yılında 101 milyar ABD dolarına ulaşarak 2019 öncesi COVID zirvesinin %98'ine (enflasyona göre düzeltilmiş) ulaşmıştır. 2024 yılında MRO harcamalarının 104 milyar ABD doları ile yeni bir rekor kırması öngörülmektedir. Önümüzdeki yıllarda MRO endüstrisinin yıllık %1,8 oranında büyüyerek 2034 yılı itibarıyla 124 milyar ABD dolarına ulaşacağı tahmin edilmektedir[4].

Uçak MRO pazarı, bakım türlerine göre gövde, motor, komponent ve hat bakımı olarak segmentlere ayrılmaktadır. Pazar, dört ana MRO kategorisinden oluşmaktadır: Motor MRO - %46, Komponent MRO - %20, Gövde MRO

- %20 ve Hat MRO - %14[5]. Uçakların güvenliği ve kullanılabilirliği, bu segmentlerdeki kapsamlı bakım faaliyetlerine bağlıdır [6]. Dünyada binlerce MRO şirketi, üreticiler, yetkili MRO'lar ve bağımsız MRO'lar dâhil olmak üzere faaliyet göstermekte olup Federal Havacılık İdaresi (FAA) ve Avrupa Birliği Havacılık Güvenliği Ajansı (EASA) gibi havacılık otoriteleri tarafından izlenen belirli prosedürlere uymaktadır [7]. MRO endüstrisinde ekonomik olarak uçak güvenliğini sağlamak için reaktif, proaktif ve agresif bakım stratejileri gibi çeşitli yöntemler kullanılmaktadır. Sektördeki eğilimler giderek proaktif yaklaşımın bir parçası olan önleyici ve kestirimci bakımı tercih etmektedir.

Önleyici bakım, sistemi bozulmadan korumak ve kalan kullanım ömrünü uzatmak için belirli aralıklarla yapılan arıza tespit görevlerini içerir. Öte yandan kestirimci bakım, arıza zamanlarını tahmin etmek ve sistem bozulmalarını tespit etmek için kestirimci algoritmalar ve gerçek zamanlı durum izlemeyi kullanmaktadır [8]. Veri analitiği, önleyici ve kestirimci bakımda kritik bir rol oynayarak MRO şirketlerinin rekabetçi olmalarına ve pazarda farklılaşmalarına rehberlik etmektedir [9].

MRO'daki analitik yöntemler genellikle betimleyici, kestirimci ve yönlendirici analitik olarak kategorize edilmektedir[10]. Havacılıkta Yoğunluk Tabanlı Gürültülü Uygulama Kümelemesi (DBSCAN) ve k-ortalama kümeleme (k-means clustering) uçuş verileri için [9], bakım verileri için makine öğrenmesi (ML) [11], havayolu operasyonları için sistem dinamikleri [6] ve kestirimci bakım için Bayes ağları gibi çeşitli teknikler uygulanmaktadır [12].

MRO pazarında OEM'ler, onarım prosedürleri, veri ve malzemeler üzerindeki kontrolleri nedeniyle avantaj sağlamaktadır. Ancak, talep ve kapasite kısıtlamaları aynı zamanda küçük ve orta ölçekli işletme (KOBİ) MRO şirketlerini de sürdürülebilir kılmaktadır[13]. Bu küçük firmalar, büyük MRO oyuncuları ve OEM'ler kadar veri bilimi kaynaklarına ve uzmanlığına sahip değildir[11]. Bu çalışma, literatürde daha önce incelenen bakım yöntemlerindeki verinin genel öneminden ziyade, güçlü veri analizinin bir KOBİ MRO'nun iş modelleri üzerindeki etkisine odaklanmaktadır.

Oliver Wyman, MRO şirketlerinin iş modellerini en çok etkileyen unsurlar arasında maliyet yönetimi, iş gücü yetersizlikleri ve ekonomik faktörleri belirlemektedir[4]. Bu baskılar, KOBİ MRO şirketlerinin bu rekabetçi sektörde farklılaşmak için değer önerilerine ağırlık vermelerini gerektirmektedir.

1.2 Tezin Amacı

Orta ve büyük ölçekli MRO şirketleri ile OEM'lerde veri bilimi faaliyetleri yoğun bir şekilde yürütülmektedir. Bu kuruluşlarda, veri ve ilgili uygulamaların miktarı zamanla hızla artış göstermektedir. Buna karşın, daha küçük ölçekli şirketler, sınırlı kaynaklar ve uzmanlık eksikliği nedeniyle veri bilimi faaliyetlerini yeterli düzeyde gerçekleştirmekte zorluk yaşamaktadır.[11]. Bu çalışma, 2013-2021 yılları arasında toplanan, küçük ve orta ölçekli bir KOBİ MRO şirketine ait operasyonel verileri kullanarak havacılık MRO endüstrisindeki kâr marjı dinamiklerini incelemektedir. KOBİ MRO şirketindeki kâr marjlarının analiz edilmesi, sektörde öncü konumda bulunan büyük şirketlere kıyasla rekabet avantajı sağlamaya olanak tanımaktadır. Bu dinamiklere odaklanan çalışma, şirketin kârlılığını ve sürdürülebilirliğini artırmaya yönelik stratejik değerlendirmeler sunmayı amaçlamaktadır.

Bu çalışma, KOBİ MRO şirketinden alınan ve 19621 gözlem içeren yapılandırılmış bir veri seti kullanılarak gerçekleştirilmiştir. Başlangıçta, ham operasyonel verilerde 83 değişken ve 21708 gözlem bulunmaktadır. Her bir değişken, veri anlamı, veri bulunabilirliği, eksik veri oranı ve veri güvenilirliği açısından detaylı bir şekilde analiz edilmiştir. Veri anlama sürecinin ardından, veri temizleme işlemi çerçevesinde, başlangıçtaki 83 değişkenden 30 tanesi ile devam edilmesine karar verilmiştir[14]. Kalan 30 değişkenden, benzer bilgi içerenlerin elenmesi ve kâr marjını iyi açıklayacağı düşünülen yeni türetilmiş değişkenlerin oluşturulmasının ardından, Gini Endeksi kriteri kullanılarak kâr marjı ile ilişkisi yüksek olduğu tespit edilen nihai 8 değişken ile modelleme çalışmaları yürütülmüştür. Ayrıca, detayları ilerleyen bölümlerde verilecek olan veri temizleme adımları sonrasında, nihai 19621 gözlem kullanılarak Gradient Boosting ve Lojistik Regresyon modelleri kurulmuştur.

Kâr marjlarını analiz etmek amacıyla öncelikle Gradient Boosting algoritmaları ve Lojistik Regresyon modeli uygulanmış, ardından istatistiksel doğrulama süreçleri gerçekleştirilmiştir. Bunun yanı sıra, MRO firmasının müşterileri "havayolu firmaları" ve "tedarikçi firmalar" olarak iki gruba ayrılmış; kâr marjını etkileyen değişkenlerin bu iki grup arasında farklılık gösterip göstermediği detaylı bir şekilde incelenmiştir. Tüm veri hazırlık adımları tamamlandıktan sonra, kâr marjı ile çeşitli operasyonel değişkenler arasındaki ilişkileri belirlemek, süreçleri optimize etmek ve kârlılığı maksimize edecek kararları önceliklendirmek amacıyla 19621 operasyonel veri, eğitim ve doğrulama veri kümelerine ayrılarak analiz edilmiştir. Ayrı bir test seti ayrılmadığından, modelin dış veri üzerindeki başarımı da doğrulama seti üzerinden değerlendirilmiştir.

Modelin doğruluğunu ve etkinliğini artırmak amacıyla anlamlı değişkenleri

seçmek için İnteraktif Gruplama metodolojisi kullanılmıştır. Değişken gruplama sonrasında, Gradient Boosting modeline ve Lojistik Regresyon modeline dahil edilmesi gereken değişkenleri belirlemek için her bir bağımsız değişken için hesaplanan Gini indeksi kullanılmıştır. Ayrıca, Gini indeksi, eşik değeri olarak belirlenen 15 üzerinde olan değişkenlerin kendi içlerinde nasıl kümelendiği analiz etmek, stratejik karar alma süreçleri için girdilerin yorumlanmasını kolaylaştırmak amacıyla Variable Clustering gerçekleştirilmiştir. Farklı Gradient Boosting modellerinin ve Lojistik Regresyon modellerinin doğrulama perspektifinden karşılaştırılması için yablış sınıflama oranı (MCR), ROC Curve ve RASE istatistikleri kullanılmıştır.

Bu çalışmanın temel katkısı, analizlerin KOBİ MRO şirketine ait gerçek veriler üzerinden yapılmış olmasıdır. İnteraktif Gruplama yöntemlerinin değişkenlere uygulanmasıyla, karmaşık veri setlerinin karar alma süreçlerinde kullanılabilirliği sağlanmış, Gradient Boosting algoritmasının ve Lojistik Regresyon modellerinin doğruluğu artırılmıştır. Kâr marjı dinamikleri, KOBİ MRO şirketinin kârlılığını artırmak için stratejik kararlar alırken potansiyel olarak faydalanabileceği bilgiler sunmaktadır.

VERİ, KULLANILAN YÖNTEMLER VE TEMEL TANIMLAYICI İSTATİSTİKLER

Bölüm 2.1, verinin modellere uygun hale getirildiği veri hazırlama sürecinde yapılan çalışmaları detaylandırmaktadır. Bölüm 2.2, Bölüm 3'te sonuçları sunulacak yöntemlerin teorik altyapılarını içermektedir. Bölüm 2.3, değişkenlere ait temel tanımlayıcı istatistikleri içermektedir.

2.1 Veri ve Veri Hazırlığı

2.1.1 Değişken Tanımları

Kar marjı (Billed Margin Calculated), analizlerde kullanılacak olan hedef değişkendir. Genel olarak, operasyon kârı Faiz ve Vergi Öncesi Kâr (EBIT - Earnings Before Interests and Taxes) olarak da bilinmektedir. Başka bir deyişle, operasyon kârı, finansman için katlanılan faiz giderleri ve devlete ödenen vergiler hariç tutularak hesaplanan kârdır. Operasyon kârını hesaplamak için, yalnızca işletmenin normal işleyişi sırasında katlandığı giderler olan operasyonel giderler, brüt kârdan çıkarılır[15].

Parça Tanımı (Part Description), tip sertifikalı (type certified) uçaklara montajı onaylanmış bileşenin, konumu, şekli veya işlevine dayalı olarak adını ve tanımını ifade eder. Parça Tanımı, OEM (Orijinal Ürün Üreticisi) tarafından sayı ve harflerle kodlanmış benzersiz parça numaralarının tanımlanmasına yardımcı olur.

Parça Maliyeti (Parts Billed Current), parça maliyeti hakkında bilgi içeren sayısal bir değişkendir. Bu değişken, onarım sürecinde tüketilen ve kullanılan malzeme maliyetlerini temsil eder. Gerçek zamanlı olarak kaydedilen bu doğrudan maliyetler, her an gerçek maliyetlerin hesaplanabilmesi için onarım sürecinin izlenmesi ve takibi amacıyla kullanılmaktadır.

İşçilik Maliyeti (Labour Billed Current) değişkeni, parçanın onarımı sırasında

kaydedilen doğrudan işçilik saatlerine göre hesaplanan doğrudan işçilik maliyetlerini temsil etmektedir. Bu değişken, onarım sürecinde toplam işçilik maliyetlerinin izlenebilmesi için gerçek zamanlı olarak kaydedilmektedir. İşgücü Maliyeti değişkeni, şirketin iş gücü verimliliğini ve doğrudan işçilik performansını, yani kontrolü altında olan unsurları yansıtırken; diğer yandan Parça Maliyeti değişkeni, malzeme fiyatlarının konuma ve satın alma zamanlamasına bağlı olarak piyasa dalgalanmalarından etkilenen dış bir maliyet unsurudur.

Yönetici (Manager) değişkeni, MRO şirketindeki belirli teknik departmanın liderini veya teknisyenini temsil etmektedir. Şirket, hidrolik, mekanik, pnömatik, elektrik ve aviyonik gibi parça türlerine göre uzmanlaşmış alt departmanlara sahiptir. Bu değişken, departmanın performansını veya belirli bir parçanın şirkete olan katkısını, aynı zamanda teknik personelin uzmanlık düzeyini izlemek ve ölçmek amacıyla kullanılmaktadır. Yönetici değişkeni, operasyonla ilişkili yöneticilerin kimliği hakkında bilgi sağlamakta olup 58 yöneticiyi kapsamaktadır.

Çevrim Süresi (Diff Shipping Entry) değişkeni, çevrim süresi (TAT - turn around time) olarak bilinen, gönderi tarihi ile giriş tarihi arasındaki farkı temsil etmektedir. TAT, bir parça veya bileşenin MRO şirketi bünyesinde işlemde olduğu süreyi ölçerek operasyonel verimliliğin kritik bir göstergesi olarak hizmet etmektedir.

İlgili Bölüm (Department) değişkeni, operasyonu tamamlamaktan sorumlu departman hakkında bilgi sağlamakta olup çeşitli departmanları kapsamaktadır.

Fiyat Teklif Süresi (Diff Quote Entry), bir teklifin müşteriye gönderildiği tarih ile parça veya bileşen için iş emrinin oluşturulduğu tarih arasında geçen süreyi temsil etmektedir. Bu metrik, müşteriye teklif sunma süresini ölçmekte olup hem müşteri hem de şirket için önemli bir performans göstergesi olan fiyat teklif süresi (TTQ - time to quote) olarak adlandırılmaktadır.

Parça Durumu (Status) değişkeni, bir parçanın MRO şirketinde onarım sürecindeyken, giriş tarihinden gönderim tarihine kadar belirli bir zamanda müşterinin siparişinin durumunu göstermektedir.

2.1.2 Veri Hazırlığı

Bu çalışma, bir MRO şirketinin kar marjını etkileyen operasyonel değişkenleri belirlemeye odaklanılmıştır. Bu tür operasyonel verilerin simülasyonundaki zorluklar nedeniyle, gerçek dünya verileri ile çalışılmasına karar verilmiştir. Bu veri seti, gerçek zamanlı operasyonel veri girişlerine dayalı olarak, bir MRO şirketinden türetilen nominal, kategorik ve sayısal değişkenler içeren yapılandırılmış verilerden

oluşmaktadır.[14] İş İhtiyaçlarını ve Veriyi Anlama adımı kapsamında, başlangıçta ham operasyonel verilerde 83 değişken bulunmaktadır. Her bir değişken, veri anlamı, veri bulunabilirliği, eksik veri oranı ve veri güvenilirliği açısından analiz edilmiştir. Veri anlama sürecinin ardından, veri temizleme işleminin devamında kullanılmak üzere başlangıçtaki 83 değişkenden 30 tanesi ile devam edilmesine karar verilmiştir.

30 değişkenden oluşan KOBİ MRO şirketine ait operasyonel ham veri kullanılarak yürütülen araştırmada, 2013'ten 2022'ye kadar olan dönemi kapsayan 21708 gözlem dahil edilmiştir. Sektör standartlarına uygun veri temizleme ve dönüştürme aksiyonları ardından analiz, 19621 gözlem ve 30 değişken ile devam etmiştir. Veri temizleme adımları gözlem bazlı olarak tüm veri setine uygulanmıştır. Uygulanan veri temizleme aşamaları aşağıda listelenmiştir.

- Kar Marjı 0 veya daha küçük olan gözlemler çıkarılmıştır (kalan: 19780 gözlem). Bu kriterleri karşılayan veriler incelendiğinde, bu kritere sahip parça onarım işlemlerinin gerçekleşmediği belirlenmiştir. Modelleme veri setine dahil edilmemişlerdir.
- Müşteri Adı olarak çalışmaya konu olan MRO şirketi adı geçen gözlemler çıkarılmıştır (kalan: 19766 gözlem). Kayıtların şirket içi işlemler için tutulduğu tespit edilmiş olup, bu gözlemlerin kâr marjını açıklamada herhangi bir önemi bulunmamaktadır. Modelleme veri setine dahil edilmemişlerdir.
- Son Gönderi Tarihi değeri 0 olan gözlemler çıkarılmıştır (kalan: 19755 gözlem). Bu sayısal verinin 0'dan küçük olmaması beklenir ve işlem detayları incelendiğinde hatalı kayıt oldukları tespit edilmiştir. Modelleme veri setine dahil edilmemişlerdir.
- İhtiyaç Tarihi eksik olan gözlemler çıkarılmıştır (kalan: 19744 gözlem). Bu kriteri karşılayan gözlemlerde hatalar tespit edilmiş ve modelleme veri setine dahil edilmemişlerdir.
- İşgücü Maliyeti değeri 40000'den fazla olan gözlemler çıkarılmıştır (kalan: 19622 gözlem). Şirketin iş kapsamına göre, işçilik maliyetlerinin bu miktarın üzerinde olması mümkün olmadığından, bu gözlemler modelleme veri setine dahil edilmemişlerdir.
- İşçilik Saatleri değişkeni değeri 11000'den fazla olan gözlemler "eksik" olarak işaretlenmiştir (kalan: 19621 gözlem). İşçilik saatlerinin bu miktarın üzerinde olması mümkün olmadığından, bu gözlemler modelleme veri setine dahil edilmemişlerdir.

İlk veri temizleme aşaması sonrasında veri setinde kalmasına karar verilen 30 değişkenden, benzer bilgiyi içeren değişkenler ve tanımlayıcı benzersiz numaralara sahip veriler de bu aşamada çıkarılmıştır. Bunun sonucuna göre Gradient Boosting ve Lojistik Regresyon modellerine girmesi muhtemel olan verileri (uzun değişken listesi) Tablo 2.1’de yer almaktadır.

Tablo 2.1 Uzun değişken listesi

Değişken	Tanım
Billed Margin Original	Kar marjı
Billed Profit	Kar
Billing Date Last	Fatura tarihi
Customer Name	Müşteri adı
Entry Date Last	Parçanın giriş tarihi (Tarih)
Entry Date Num	Pandemi öncesi - sonrası (Kukla)
Invoice Number	Fatura numarası
Labour Hours	İşçilik saatleri
Last Ship Date	Parçanın gönderi tarihi (Tarih)
Last Ship Date Num	Parçanın gönderi tarihi (Nümerik)
Part Number	Parça numarası
Quote Date Last	Fiyat teklif tarihi (Tarih)
Quote Date Num	Parçanın teklif tarihi (Nümerik)
Sales Invoice	Fatura numarası
Total Billed Current	Toplam fatura tutarı
WO Number	İş emri numarası
WO Type	İş emri türü
Work Type Description	İş türü açıklaması
Labor Billed Current	İşçilik maliyeti
Parts Billed Current	Parça maliyeti
Department	Parçadan sorumlu departman
Part Description	Parça tanımı
Status	Parçanın durumu

Tablo 2.1’de yer alan değişkenlerin modele dahil edilip-edilmemesine ilişkin değerlendirmeler aşağıda yer almaktadır. Bu kararlar sonucunda, Tablo 2.1’deki uzun değişken listesinden modele girecek nihai değişkenler listesi seçilmiştir. Oluşturulan Kısa değişken listesi Tablo 2.2’de yer almaktadır.

Kar marjı, Billed Margin Calculated olarak adlandırılan, manuel olarak hesaplanmış ve standartlaştırılmıştır. Tablo 2.2’deki bu değişken, Tablo 2.1’de yer alan Billed Margin Original değişkeni olmasına rağmen, manuel hesaplanmıştır. Veri kontrolü aşamasında, Billed Margin Original değişkeninin ham veride yer alan otomatik hesaplamasında bazı hatalı sonuçlar elde edildiği gözlemlendiğinden, bu değişken model geliştirmede kullanılmamıştır. Tablo 2.1’deki Billed Profit ve Total

Billed Current değişkenleri, Tablo 2.2’de yer alan standartlaştırılmış Billed Margin Calculated, kar marjı, hedef değişkenini türetmek için kullanılmıştır.

Tablo 2.1’deki Last Ship Date ve Billing Date Last aynı bilgiyi içerdiğinden, Billing Date kısa listeye dahil edilmemiştir. Aşağıdaki paragraflarda açıklanacağı üzere, Last Ship Date, parçanın gönderi tarihi, diğer değişkenlerin türetilmesinde kullanılmış (Diff Quote Entry ve Diff Shipping Entry’de olduğu gibi) ve doğrudan modele dahil edilmemiştir.

Tablo 2.1’de yer alan Customer Name, Invoice Number, Part Number ve Work Order (WO) Number değişkenleri modelde kullanılmamıştır; yalnızca veri birleştirme, veri hazırlama ve veri kontrolü aşamalarında anahtar değişkenler olarak kullanılmışlardır.

Pandemi öncesi ve sonrası yapılan tamir işlemlerini ayırt etmek amacıyla, Tablo 2.1’deki Entry Date Last’dan Entry Date Num adında bir kukla değişken oluşturulmuştur. Entry Date Num değişkeninin, hedef değişken kar marjı ile ikili ilişkisi incelendiğinde Gini indeksi eşik değeri olarak belirlenen 15 değerinin altında olması nedeniyle, modelde kullanılan değişkenlerin listelendiği Tablo 2.2’ye dahil edilmemiştir. Bu değişken kar marjı ile ilişkisi düşük olduğu için modele girmeye hak kazanmamıştır.

Tablo 2.1’deki Labour Hours değişkeni yerine eksik değer sayısı daha az olduğu için Tablo 2.2’deki Labour Billed Current değişkeninin modele dahil edilmesine karar verilmiştir.

Tablo 2.1’deki Quote Date Last değişkeni modelde doğrudan kullanılmamış; ancak, Tablo 2.2’ye dahil edilen standartlaştırılmış Diff Quote Entry değişkenini oluşturmak için kullanılmıştır.

Yukarıda belirtilen veri temizleme sürecinden (kalan 19621 gözlem) sonra Tablo 2.1’deki WO Type değişkeninde yalnızca bir kategori kaldığından, diğer kategoriler veri temizleme aşamasında silindiğinden, WO Type modelde kullanılacak değişkenler arasında yer almamıştır.

Tablo 2.1’deki Labour Billed Current, Parts Billed Current, Department, Part Description ve Status değişkenleri İnteraktif Gruplama ardından Gini indeksinin 15’in üzerinde olması nedeniyle modelde kullanılmış ve Tablo 2.2’de kısa değişken listesine dahil edilmiştir.

Tablo 2.1’deki Work Type Description değişkeni hedef değişkene göre hesaplanan Gini indeksinin 15’in altında olması nedeniyle kısa değişken listesine girememiş ve

Tablo 2.2’de yer almamıştır.

Veri Hazırlama adımı kapsamında, sektör profesyonelleri ile yapılan görüşmeler sonrasında, veri setinde mevcut değişkenleri içeren Tablo 2.1’de nihai modelde yer alan Billed Margin Calculated, Diff Quote Entry ve Diff Shipping Entry gibi türetilmiş değişkenler eklenmiştir. Türetilen Billed Margin Calculated değişkeni, ilgili gözlem için elde edilen kar marjını temsil eder. Diff Quote Entry, parçanın sisteme giriş tarihi olan Entry Date Last ile parçanın fiyat teklif tarihini ifade eden Quote Date Last arasındaki süreyi nümerik formatta hesaplayan bir türetilmiş değişkendir. Diff Shipping Entry, parçanın sisteme kaydedildiği Entry Date Last ile parçanın gönderildiği tarih olan Last Ship Date arasındaki süreyi sayısal olarak hesaplayan başka bir türetilmiş değişkendir.

Tablo 2.1’deki verilere ilişkin yukarıda sıralanan kararlar sonucunda, modelde yer alacak nihai değişkenlerin listesi elde edilmiştir. Modele girecek değişkenlerin seçilmesi için Gini indeksi, Bölüm 2.3’de açıklanan İnteraktif Gruplama teorisi doğrultusunda, değişken seçimi kriteri olarak kullanılmıştır. Tablo 2.2, değişken seçiminden sonra modele dahil edilmesine karar verilen kısa değişken listesini içermektedir.

Tablo 2.2 Kısa değişken listesi

Değişken	Tanım
Billed Margin Calculated (Hedef)	Kar Marjı (Türetilmiş)
Diff Quote Entry	Fiyat teklif süresi (Türetilmiş)
Diff Shipping Entry	Çevrim süresi (Türetilmiş)
Labor Billed Current	İşçilik maliyeti
Parts Billed Current	Parça maliyeti
Department	Departman
Part Description	Parça tanımı
Status	Parçanın durumu

Tablo 2.2, tüm veri temizliği ve değişken seçimi yöntemleri uygulandıktan sonra Gradient Boosting algoritması ve Lojistik Regresyon modellerinde kullanılacak değişkenler listesini içermektedir. Bu değişkenler, ölçüm birimlerini aynı skalaya çekmek için analizlere başlanmadan standardize edilmişlerdir. Bölüm 2.3’de standartlaştırılmış değişkenler ve gruplanmış değişkenlere ilişkin temel tanımlayıcı istatistikler yer almaktadır.

2.2 Kullanılan Yöntemler

2.2.1 İnteraktif Gruplama

İnteraktif Gruplama, sürekli dağılımların çok modlu yapılarının doğrusal olmayan fonksiyon modellerini kullanır. İlk grupları kantiller yoluyla hesaplar ve bu grupların etkileşimli olarak bölünmesine ve birleştirilmesine izin verir. Bu yöntem, hem kategorik hem de aralıklı tüm girdi değişkenleri için gruplar veya sınıflar oluşturmak amacıyla kullanılır. Gruplar, benzersiz seviye sayısını azaltmak ve her girdinin öngörü gücünü artırmak amacıyla oluşturulur. Sağlam özelliklerin seçimi Gini istatistiğine dayanır ve seçilen özellikler iş gereksinimlerine göre gruplandırılır. Bu analiz, aşırı uyuma yol açabilecek model kendine has özellikleri değil, kar marjı sıralamasındaki eğilimleri yansıtmak amacıyla veriyi yapılandırmaya yardımcı olur. Bazı durumlarda, özelliklerin gruplandırılması öngörü gücünü de artırır.

Modelleme sürecinde, İnteraktif Gruplama, Tablo 2.3’de tanımlanan aykırı değerlerin etkisini en aza indirmede önemli bir rol oynar. Bu yöntem, aykırı değerlerin olumsuz etkilerini azaltır ve veriyi gruplandırarak model performansını önemli ölçüde artırır. Gini indeksi, İnteraktif Gruplama’da manuel olarak gruplandırılan değişkenlerin farklı kategorilerdeki olay oranlarının tekdüzeliğini değerlendirmek için yaygın olarak kullanılan önemli bir seçim kriteri olarak kabul edilir ve değişken seçimi sürecinde kullanılmıştır[16]. Gini indeksinin dağılım eşitliğini ölçme, yanlılığı azaltma ve karşılaştırma kolaylığı sağlama konusundaki dikkate değer yeteneği bu seçimi destekler[17].

Gini indeksi bir yüzde olarak ifade edilir. Gini katsayısı 0 ile 1 arasında değişir, dolayısıyla Gini indeksi 0 ile 100 arasında değişir. Gini indeksi bir dağılımın tekdüzeliğini veya tekdüze olmama durumunu ölçer. Gini indeksi ne kadar düşükse, bir değişken o kadar tekdüze dağıtılmıştır. Gradient Boosting ve Lojistik Regresyon uygulanarak kar marjı modeli bağlamında, Gini indeksi, bir özelliğin nitelikleri arasında olay oranlarının ne kadar tekdüze dağıldığını değerlendirmek için kullanılır. Özellik önemini hesaplamak için Gini indeksinin kullanılması, veri seti içindeki verimsiz değişkenlerin çıkarılmasına da katkı sağlamıştır[18].

İlk olarak, veri her bir nitelik ile ilişkili olay oranına göre azalan sırada sıralanır. Bir özellik m niteliğe sahipse, sıralanan nitelikler 1, 2, ..., m grupları olarak temsil edilir. Her grup bir niteliğe karşılık gelir ve grup 1, en yüksek olay oranına sahiptir.

Daha sonra, bu sıralanan grupların her biri için, grup i ’deki olay (n_i^{event}) ve olay dışı ($n_i^{\text{non-event}}$) sayıları hesaplanır. Son olarak ise, Gini indeksini hesaplanır.

Gini indeksi hesaplaması eşitlik 2.1’de gösterildiği gibi yapılabilir:

$$\text{Gini İndeksi} = \left(1 - \frac{2 \sum_{i=2}^m \left(n_i^{\text{event}} \times \left(\sum_{j=1}^{i-1} n_j^{\text{non-event}} \right) \right)}{N^{\text{event}} \times N^{\text{non-event}}} + \frac{\sum_{k=1}^m \left(n_k^{\text{event}} \times n_k^{\text{non-event}} \right)}{N^{\text{event}} \times N^{\text{non-event}}} \right) \times 100 \quad (2.1)$$

Burada, olay N^{event} ve olay dışı $N^{\text{non-event}}$, sırasıyla verideki toplam olay ve olay dışı sayısını temsil etmektedir.

2.2.2 Değişken Kümeleme

Kümeleme analizi, birbirine benzeyen nesneleri (veya değişkenleri) gruplandırmayı içerir. Burada benzerlik, işleme başlamadan önce tanımlanması gereken bir kavramdır. Genellikle bir mesafe ölçütü kullanılır, böylece benzerlik, mekansal yakınlık haline gelir[19].

Kümeleme analizi, 'denetimsiz' öğrenme için önemli bir araçtır - bir yanıt değişkeni yardımı olmadan verilerde grupları bulma problemidir. Kümeleme analizindeki en büyük zorluklardan biri, 'kümelerin' optimum sayısını tahmin etmektir[20].

Değişken kümeleme, değişken sayısını azaltmak amacıyla gerçekleştirilebilir. Kural, küme temsilcisi olarak $1 - R_{ratio}^2$ değerinin minimum olduğu değişkeni seçmeyi gerektirir. $1 - R_{ratio}^2$ değeri eşitlik 2.2’de verilmiştir.

$$1 - R_{ratio}^2 = \frac{1 - R_{own}^2}{1 - R_{nearest}^2} \quad (2.2)$$

Sezgisel olarak, küme temsilcisinin kendi kümesiyle olabildiğince yüksek korelasyona sahip olması ($1 - R_{own}^2$ 1’e yakın) ve en yakın kümeyle olabildiğince düşük korelasyona sahip olması ($1 - R_{nearest}^2$ 0’a yakın) beklenir. Bu nedenle, bir kümenin en uygun temsilcisi, $1 - R_{ratio}^2$ değerinin sıfıra yaklaştığı değişkendir[21].

Değişken kümeleme her ne kadar değişken azaltma yöntemi olarak kullanılsa da bu çalışmada, değişkenlerin benzerliğini daha iyi anlamak amacıyla KOBİ düzeyindeki MRO şirketi için stratejik kararların oluşturulmasına katkı sağlamak amacıyla uygulanmıştır. Bu yöntem, girdi değişkenlerinin nasıl gruplandığını görsel olarak sunmuştur. Bu kümeler ve bireysel değişkenler baz alınarak Bölüm 4’te önerilerde

bulunulmuştur. Bu kümeler, şirket için stratejik kararların önerilmesi sırasında model çıktıları yanında ek bir bilgi sağlamıştır.

2.2.3 Gradient Boosting Algoritması

Gradient Boosting ağaçları yöntemi, özellikle temiz olmayan verilerin madenciliği için uygun olan, rekabetçi, son derece sağlam ve yorumlanabilir işlemler üretir[22].

Gradient Boosting, analiz veri setini birden fazla kez yinelemeli olarak yeniden örnekleyen ve sonuçları yeniden örneklenen veri setlerinin ağırlıklı bir ortalaması şeklinde birleştiren gelişmiş bir artırma tekniğidir. Ağaç artırma, bir dizi karar ağacının inşasını içerir ve bu ağaçlar kolektif olarak tek bir tahmin modeli oluşturur. Dizideki her bir ağaç, bir önceki ağaçlar tarafından üretilen tahminlerin artık değerlerine uyum sağlar; bu artık değerler, bir kayıp fonksiyonunun türevine göre tanımlanır.

Algoritmanın her bir yinelemesinde veri kullanılarak bir karar ağacı oluşturulur ve ortaya çıkan ağacın doğruluğu değerlendirilir. Sonraki örnekler, önceden belirlenmiş hataları ele almak için ayarlanır. Her bir örneğin ağırlığı, önceki modellerin sınıflandırma doğruluğuna göre ayarlandığından, bu yaklaşım bazen stokastik Gradient Boosting olarak da adlandırılır. Artırma, ikili, nominal ve aralıklı hedeflere uygulanabilir.

Stokastik Gradient Boosting kayıp optimizasyonu, eğitim sürecinde üretilen kayıpları azaltmayı hedefler[23]. Aralıklı hedefler için artık değer, kare hata kayıp fonksiyonu kullanılarak tanımlanır. Aralıklı bir hedef için kare hata kaybını kullanarak artık değeri hesaplamak için tahmin değeri, hedef değerden çıkarılır. İkili hedefler için artık değer, lojistik kayıp olarak da bilinen negatif binom log-olasılık kaybı fonksiyonu kullanılarak tanımlanır.

Karar ağaçlarına benzer şekilde, artırma yöntemi de veri dağılımı hakkında varsayımlar yapmaz. Aralıklı girdiler için model yalnızca değerlerin sıralarına dayanır. Aralıklı hedeflerde aşırı değer teorisinin etkisi seçilen kayıp fonksiyonuna bağlıdır. Artırma, tek bir karar ağacına kıyasla aşırı uyuma daha az eğilimlidir. Bir karar ağacı veriyi makul bir şekilde açıklıyorsa, genellikle artırma modelin uyumunu artırır.

Tüm veri seti üzerindeki toplam kayıp, eşitlik 2.3'te genel örneklerin bireysel kayıplarının toplamı olarak tanımlanır[23]:

$$L_{\text{total}} = \sum_i L(f(x_i, \theta), y_i) \quad (2.3)$$

burada (x_i) girdi verisi, (y_i) ise hedef değeri ifade eder. Model, parametreleri θ olan bir karar ağacı kullanarak (x_i) için tahmin değeri $\hat{y}_i$ 'yi üretir. Çıktı, parametrelerin ve girişin herhangi bir fonksiyonu olabilir ve $\hat{y}_i = f(x_i, \theta)$ olarak temsil edilir.

Optimizasyon algoritması, bu toplam kaybı en aza indiren θ parametrelerinin değerlerini tahmin etmeye odaklanır. Genellikle, parametreleri θ 'nın toplam kayıp ile ilgili gradyanının tersi yönde güncellenmesiyle gerçekleşir, bu işlem eşitlik 2.4'te gösterilmiştir[23]:

$$\theta_{\text{new}} = \theta_{\text{old}} - \alpha \nabla_{\theta} L_{\text{total}} \quad (2.4)$$

Burada, α öğrenme hızını kontrol eder ve $\alpha \nabla_{\theta}$ toplam kayıp ile ilgili gradyanı temsil eder[23].

2.2.4 Lojistik Regresyon

Lojistik regresyon, genellikle basit ikili (binary) bağımlı değişkenleri modellemek için en sık kullanılan yöntemdir[24].

Eğer bağımsız değişken yalnızca iki değer alabiliyorsa (“olay” ve “olay olmayan”), doğrusal regresyon için gerekli koşullar sağlanmaz; özellikle, hatalar ikili (binary) ve normal dağılıma sahip değildir. Bu durumu ele almak için İkili Lojistik Regresyon geliştirilmiştir.

Hedef değişkeni Y_i , her bir gözlem $i = 1, \dots, n$ için olayları 1 ve olay olmayanları 0 olarak kodluyor. Ayrıca p adet açıklayıcı değişken $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ olarak tanımlanmıştır. Bir olayın olasılığını ise $\pi_i = \Pr(Y_i = 1)$ olarak tanımlansın. Bu durumda lojistik model şu şekilde yazılır:

$$\text{logit}(\pi_i) = \log \left(\frac{\pi_i}{1 - \pi_i} \right) = \alpha + \mathbf{x}_i \beta = \alpha + x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{ip}\beta_p \quad (2.5)$$

Yanıtların beklenen davranışı, kovaryatların doğrusal bir kombinasyonu olan $C + X$ ile modellenir. Oran $\frac{\pi_i}{1 - \pi_i}$, *odds* (olasılık oranı) olarak adlandırılır ve oranların logaritması ise *logit* veya *log odds* olarak bilinir. Bu model, β ve α parametreleri

için doğrusaldır ve X kovaryatlarının farklı kombinasyonları, yanıtın bağımsız olarak binom dağılım gösterdiği alt popülasyonları tanımlar.

Alt popülasyonlar içinde basit rastgele örnekleme yapıldığında, veriler, bir olasılık fonksiyonu türetebilecek bir çarpımsal binom dağılıma sahiptir ve diğer tüm sonuçlar da bu dağılımdan çıkarılabilir.

Birçok makine öğrenimi (ML) probleminde, Çoklu Lojistik Regresyon modeli için kullanılabilecek geniş bir giriş değişkeni yelpazesi bulunmaktadır. Ancak, bu değişkenlerin hangilerinin en alakalı ya da faydalı olduğunu belirlemek genellikle zordur. Durum, bazı giriş değişkenleri arasında bağımlılıklar olduğunda daha da karmaşık bir hale gelir. Bu tür durumlarda, uygun alt kümeler, iyi bir performansa sahip bir model bulmak için yeterli olabilir [25].

Girdi değişkeni seçimi (Input Feature Selection - IFS), model performansına anlamlı bir katkı sağlamayan giriş değişkenlerini belirlemeyi ve bunları modelden çıkarmayı amaçlayan çeşitli teknikleri kapsar. Genellikle algoritmalar ve Çoklu Lojistik Regresyonun bir arada kullanıldığı teknikler tercih edilmektedir. Bu teknikler, değişkenleri kademeli olarak ekleyen veya çıkaran adım adım algoritmalar olabileceği gibi, budanmış bir regresyon ağacı da kullanılabilir. Bu algoritmalar, diğer yöntemlerle keşfedilemeyen giriş değişkeni alt kümelerini bulabilir.

Özellik dönüşümünden ziyade özellik seçimi, orijinal değişkenlerin birimlerinin ve anlamlarının kullanıcılar için önemli olduğu durumlarda tercih edilir. Ayrıca, modelleme amacı etkili bir alt kümeyi tanımlamak olduğunda da özellik seçimi öne çıkar. Kategorik değişkenlerin bulunduğu ve sayısal dönüşümlerin uygun olmadığı durumlarda ise, boyut indirgeme (*dimension reduction*) tercih edilen özellik seçimi yöntemidir.

FSR (Forward Stepwise Regression) tekniği, özellikle en küçük kareler yöntemi için tasarlanmış ardışık bir özellik seçimi tekniğidir. Bu teknik, yalnızca en küçük kareler kriteri ile mümkün olan optimizasyonları kullanır. Genelleştirilmiş ardışık özellik seçiminin aksine, ileriye doğru adım adım regresyon (forward stepwise regression) daha önce eklenmiş olan özellikleri çıkarabilir veya çıkarılmış olan özellikleri tekrar ekleyebilir. Özellikler, hata kriterini daha fazla azaltmayan bir özelliğin eklenmesine kadar, boş bir aday kümesine ardışık olarak eklenir [26].

Forward stepwise yöntemi algoritmasında ilk adımda model, sadece kesim (intercept) terimini içerir; başlangıçta hiçbir bağımsız değişken modele dahil edilmez. İkinci adımda Modelin başlangıç durumu temel alınarak, ilk adımda

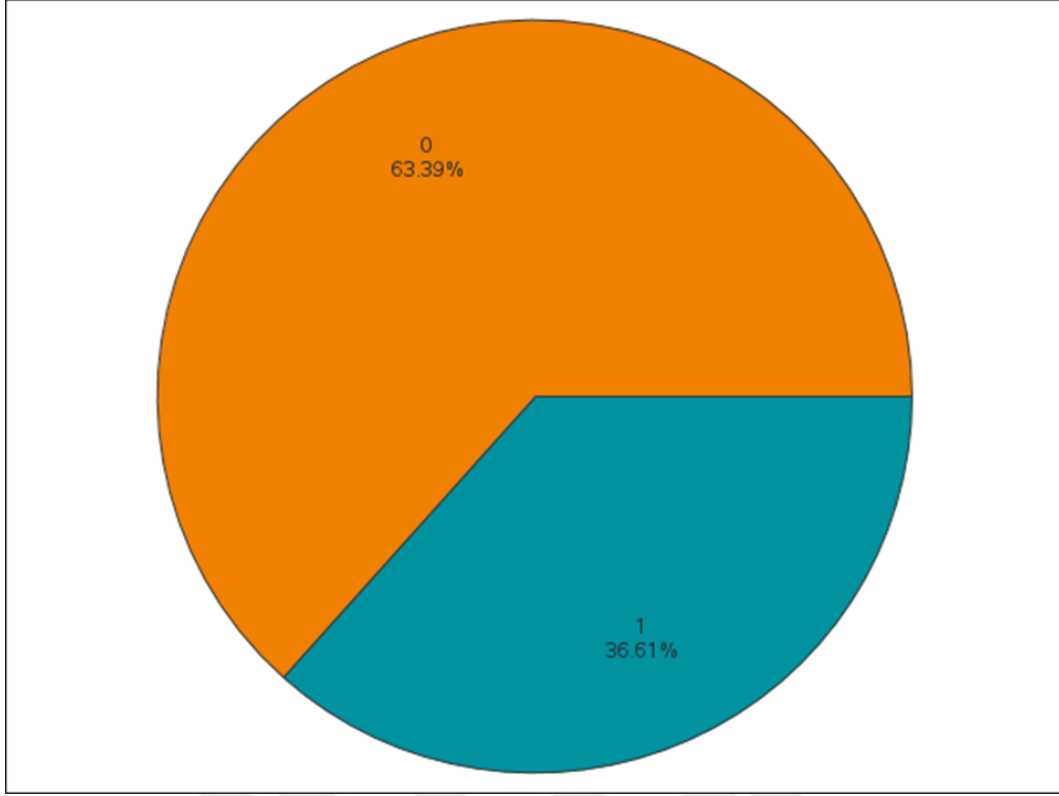
hangi deęişkenin modele eklenmesi gerektięine karar verilir[27]. Her bir bağımsız deęişkenin modele eklenmesi durumunda, model performansında ne kadar iyileşme sağlanacağı hesaplanır. Modele en çok katkı sağlayan bağımsız deęişken seçilir (örneğin, p-deęeri belirli bir eşik deęerinden küçük olmalı). Bu deęişken modele eklenir ve model yeniden deęerlendirilir. Deęişkenlerin eklenmeye devam edilmesi için durdurma kriterine karar verilir. Bu çalışmada durdurma kriteri olarak yanlış sınıflama oranı (MCR) kabul edilmiştir. Deęişkenin modele eklenmesi ile yanlış sınıflama oranında iyileşme gözlemlenmemesi modeli durdurma kriteri olarak kabul edilmiştir. Tüm aday deęişkenler bu şekilde tek tek modele eklenerek nihai model elde edilir.

2.3 Temel Tanımlayıcı İstatistikler

2.3.1 Veri Seti Olay Dağılımı ve Korelasyon Matrisi

Modelleme öncesinde modele giren deęişkenlerin, hedef deęişken olan kar marjı ile ilişkisini daha iyi açıklayabilmek için sürekli kâr marjı deęişkeni, ikili deęişkene dönüştürülmüştür. Bu dönüşümün amacı, hem İnteraktif Gruplama işlemi sırasında bağımsız deęişkenler ile hedef deęişken arasındaki tek yönlü ilişkiyi daha kolay analiz edebilmek hem de Gradient Boosting ve Lojistik Regresyon modellerinde hedef deęişkeni sınıf olarak deęerlendirebilmektir. Sürekli kâr marjı deęişkeni için kesme noktası olarak 0,5 belirlenmiştir. Buna göre, 0,5'in üzerindeki deęerler "1", 0,5'in altındaki deęerler ise "0" olarak ikili deęişkene dönüştürülmüştür.

Bu dönüşüm sonrasında çalışılan veri setinin hedef deęişkeni bazındaki dağılımı Şekil 2.1'de gösterilmektedir.



Şekil 2.1 Veri seti hedef değişken dağılımı

Şekil 2.1, "0" kategorisinin örneklemdeki oranı yaklaşık %63.4'tür (toplam gözlem sayısı 12438). "1" kategorisinin örneklemdeki oranı yaklaşık %36.6'dır (toplam gözlem sayısı 7183).

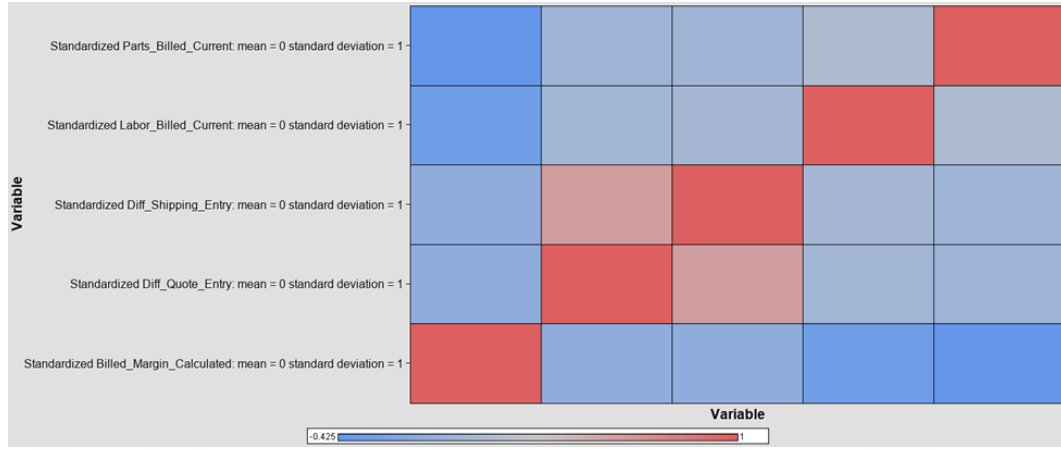
%36,610'lık bir olay oranı (event rate - kar marjının 0,5'den büyük olması), dengesiz bir veri seti olarak kabul edilmez ve lojistik regresyon için uygundur. Çok düşük olay oranlarında (%5 veya altı), model performansı düşebilir ve veri dengesizliğiyle başa çıkmak için özel yöntemler gerekebilir. %35 gibi orta düzeyde bir oran, modelin hem olayları hem de olay dışı durumları iyi bir şekilde öğrenmesine olanak tanır.

%35 olay oranı, Lojistik Regresyon için bir dengesizlik problemi yaratmaz. Ancak, oran çok düşük olsaydı (örneğin %1-%10 aralığında), modeli iyileştirmek için farklı stratejiler gerekebilirdi (örneğin, yeniden örnekleme veya Ağırlıklı Lojistik Regresyon).

Lojistik regresyon, olay oranı makul seviyede olduğunda etkin bir modeldir. Ancak, non-lineer ilişkileri daha iyi yakalamak veya daha karmaşık veri yapılarıyla çalışmak için alternatif modeller (örneğin, karar ağaçları, random forest veya Gradient Boosting) de değerlendirilebilir. Bu çalışmada modellemede Lojistik

Regresyon ve Gradient Boosting kullanılacaktır.

Standartlaştırılmış değişkenler için korelasyon matrisi Şekil 2.2’de sunulmuştur.



Şekil 2.2 Standartlaştırılmış değişkenler için korelasyon matrisi

Şekil 2.2, bağımlı değişken olan Standartlaştırılmış Billed Margin Calculated, Kar Marjı değişkeni ile diğer bağımsız değişkenler arasındaki korelasyonları bir ısı haritası şeklinde göstermektedir. Değişkenler arası ilişkilerin incelenmesi, modelleme sürecinde hangi değişkenlerin daha etkili olduğunu anlamak ve stratejik kararlar almak açısından önem taşımaktadır.

Standartlaştırılmış Parts Billed Current, Parça Maliyeti, bağımlı değişken olan Kar Marjı ile pozitif bir korelasyon göstermektedir. Bu durum, parça maliyetlerinin artmasının kâr marjını olumlu etkileyebileceğini göstermektedir. Ancak korelasyonun gücü sınırlıdır, bu da etkisinin tamamen belirleyici olmadığını ifade eder.

Standartlaştırılmış Labor Billed Current, işçilik maliyetleri ile bağımlı değişken arasında negatif bir ilişki vardır. Bu ilişki, işçilik maliyetlerinin artmasının kâr marjını düşürdüğünü işaret etmektedir. Bu nedenle, işçilik maliyetlerini azaltmak kâr marjını artırma açısından kritik bir strateji olarak öne çıkmaktadır.

Standartlaştırılmış Diff Shipping Entry, çevrim süresi ile bağımlı değişken arasında zayıf bir negatif korelasyon görülmektedir. Daha kısa parça teslim süreleri, kâr marjını artırmada olumlu bir faktör olarak değerlendirilebilir.

Standartlaştırılmış Diff Quote Entry, fiyat teklif süresi ile bağımlı değişken arasında belirgin bir negatif korelasyon bulunmaktadır. Daha kısa teklif süreleri, kârlılığı artırma açısından önemli bir rol oynar.

Bağımsız değişkenler arasındaki ilişkiler incelendiğinde ise Standartlaştırılmış Parts

Billed Current ve Standartlaştırılmış Labor Billed Current değişkenleri arasında pozitif bir ilişki gözlemlenmiştir. Parça maliyetleri ile işçilik maliyetlerinin birlikte hareket ettiği, yani birinin artmasının diğerini de artırabileceği anlamına gelir.

Standartlaştırılmış Diff Quote Entry ve Standartlaştırılmış Diff Shipping Entry değişkenleri arasında da pozitif bir ilişki görülmektedir. Daha uzun teklif sürelerinin, daha uzun parça teslim süreleri ile ilişkili olduğu söylenebilir.

Şekil 2.2, bağımlı değişken olan Standartlaştırılmış Billed Margin Calculated ile bağımsız değişkenler arasındaki korelasyonları ortaya koysa da, korelasyonların gücünün nispeten düşük olması, bağımlı değişkeni açıklamak için bu değişkenlerin doğrudan yeterli olamayabileceğini göstermektedir. Bu durum, bağımlı değişkeni daha iyi açıklayabilecek yeni türetilmiş ve gruplanmış değişkenlerin oluşturulmasını gerekli kılmıştır.

Modelleme sürecinde, İnteraktif Gruplama yöntemi kullanılarak mevcut değişkenler üzerinde detaylı analizler yapılmış ve bağımlı değişkenin farklı kategorilerdeki dağılımını daha iyi yansıtabilecek gruplar oluşturulmuştur. İnteraktif Gruplama, bağımsız değişkenlerin olay ve olaydışı kategorilerdeki dağılımını inceleyerek, değişkenlerin hedef değişken üzerindeki etkilerini en iyi şekilde yansıtan gruplar üretmeyi amaçlamaktadır.

2.3.2 Tanımlayıcı İstatistikler

Tablo 2.2’de listelenen ve modele dahil edilen sayısal değişkenlerin ortalama ve medyan değerleri Tablo 2.3’te sunulmuştur.

Tablo 2.3 Nümerik değişkenler için tanımlayıcı istatistikler

Değişken	Ortalama	Medyan
Billed_Margin_Calculated	55,71	55,25
Diff_Quote_Entry	19,38	8,00
Diff_Shipping_Entry	58,64	31,00
Labor_Billed_Current	383,39	246,00
Parts_Billed_Current	2433,43	204,49

Tablo 2.3’de, Diff Quote Entry, Diff Shipping Entry, Labour Billed Current ve Parts Billed Current değişkenlerinin ortalama ve medyan değerleri, verilerin çoğunluğunun düşük değerlerde yoğunlaştığını, birkaç aşırı yüksek değer ise sağa çarpık bir dağılım yarattığını göstermektedir.

Tablo 2.2’de listelenen ve modele dahil edilen kategorik değişkenlerin tanımlayıcı istatistikleri Tablo 2.4’te sunulmuştur. Amaç, İnteraktif Gruplama uygulayarak

aykırı değer problemini yönetmektir. Bölüm 3.3'te ise gruplama sonrası verinin olay oranları ayrıntılı olarak açıklanmaktadır.

Tablo 2.4 Kategorik değişkenler için tanımlayıcı istatistikler

Değişken	Seviye Sayısı	Mod Yüzdesi	Mod
Department	8	48,25	ELECTROMECHANICAL
Manager	17	11,75	Manager1
Part_Description	1670	19,80	PINION
Status	5	51,65	SHIP

Tablo 2.4, her bir kategorik değişken için seviye sayısını, mod değerini ve mod yüzdesini göstermektedir.

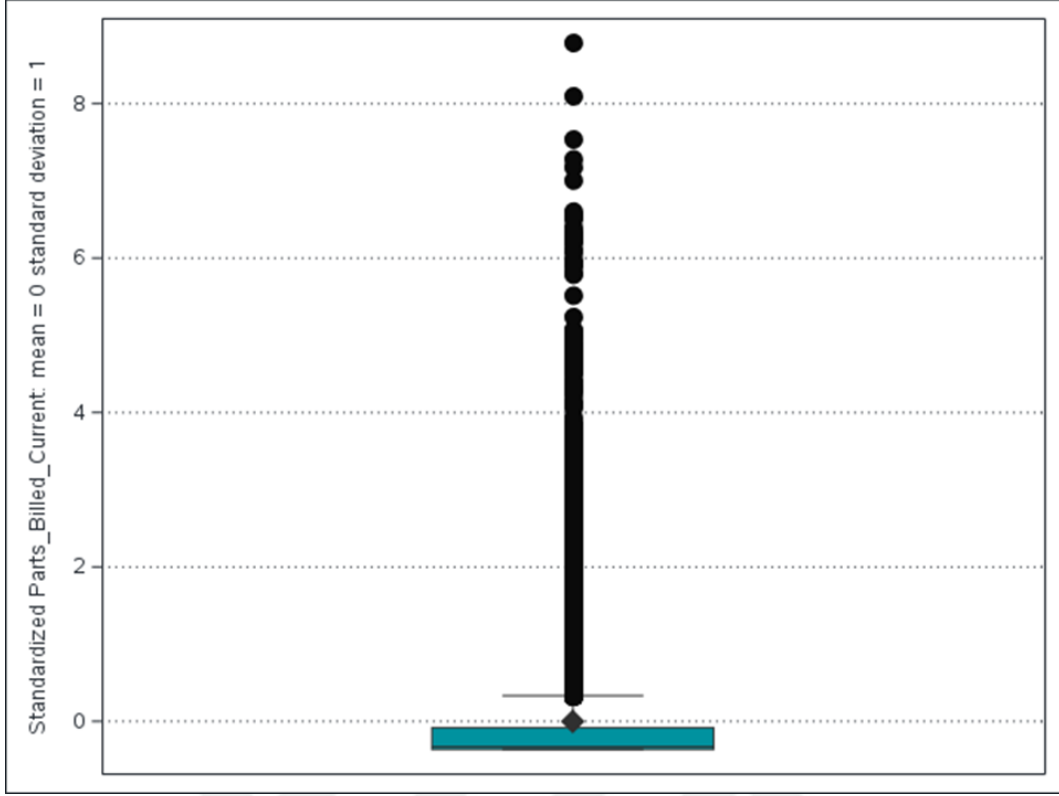
Model çalıştırılmadan önce sayısal değişkenler standartlaştırılmış ve ardından tüm değişkenler gruplandırılmıştır. Değişken seçimi, Gini indeksi uygulanarak gerçekleştirilmiştir [18]. Modelleme adımı kapsamında, Gradient Boosting ve Lojistik Regresyon Modelleri eğitim ve doğrulama veri setleri üzerinde sırasıyla %70 ve %30 oranında çalıştırılmıştır. Bu çalışmada ayrı bir test seti ayrılmadığından, doğrulama veri seti aynı zamanda modelin performansını değerlendirmek amacıyla test işlevi görmüştür.

Tüm veri temizleme, ön işleme ve model uygulamaları için Python ve SAS yazılımları kullanılmıştır.

2.3.3 Standartlaştırılmış ve Gruplandırılmış Değişkenler için Kutu Grafikleri

2.3.3.1 Parça Maliyeti

Standartlaştırılmış Parça Maliyeti değişkeni kutu grafiği Şekil 2.3'te sunulmuştur.



Şekil 2.3 Standartlaştırılmış parça maliyeti - kutu grafiği

Şekil 2.3, Standartlaştırılmış Parts Billed Current, standartlandırılmış parça maliyeti değişkenine ait dağılımı ve uç değerlerini standartlaştırılmış bir formatta göstermektedir. Değişken, ortalama = 0 ve standart sapma = 1 olacak şekilde normalize edilmiştir. Bu tür bir standartlaştırma, farklı ölçeklerdeki değişkenlerin karşılaştırılabilirliğini artırır.

Merkezi eğilimi incelemek için, kutunun içindeki siyah elmas şekli, verinin ortalamasını temsil etmektedir. Kutunun alt sınırına yakın bir medyan (yatay çizgi) değeri bulunmaktadır. Bu durum, dağılımın çarpık (asymmetrical) olduğunu gösterebilir.

Çeyrekler arası aralık (IQR) incelendiğinde, kutunun alt ve üst sınırları, birinci çeyrek (%25) ve üçüncü çeyrek (%75) değerlerini temsil eder. Çeyrekler arası aralık oldukça dar görünmektedir, bu da verinin büyük kısmının düşük değerlerde toplandığını göstermektedir.

Uç değerler (Outliers) incelendiğinde, kutunun üst sınırından çok uzağa dağılan birden fazla uç değer (outliers) bulunmaktadır. Verinin önemli bir kısmı 0 ile 1 arasında yoğunlaşırken, 2'nin üzerinde çok sayıda uç değer bulunmaktadır. 6'nın üzerindeki değerler, ekstrem uç değerler olarak sınıflandırılabilir. Bu durum,

modelleme sırasında dikkat edilmesi gereken bir veri özelliklerine işaret eder.

Maksimum ve minimum değerler incelendiğinde, maksimum değer, 8'in üzerine çıkmaktadır. Bu, veri setinde bazı gözlemlerin standart sapma birimlerine göre oldukça yüksek değerlere sahip olduğunu gösterir. Minimum değer grafikte net bir şekilde görünmemektedir, ancak kutu ve alt çizgi (alt whisker), minimumun 0 civarında toplandığını ima etmektedir.

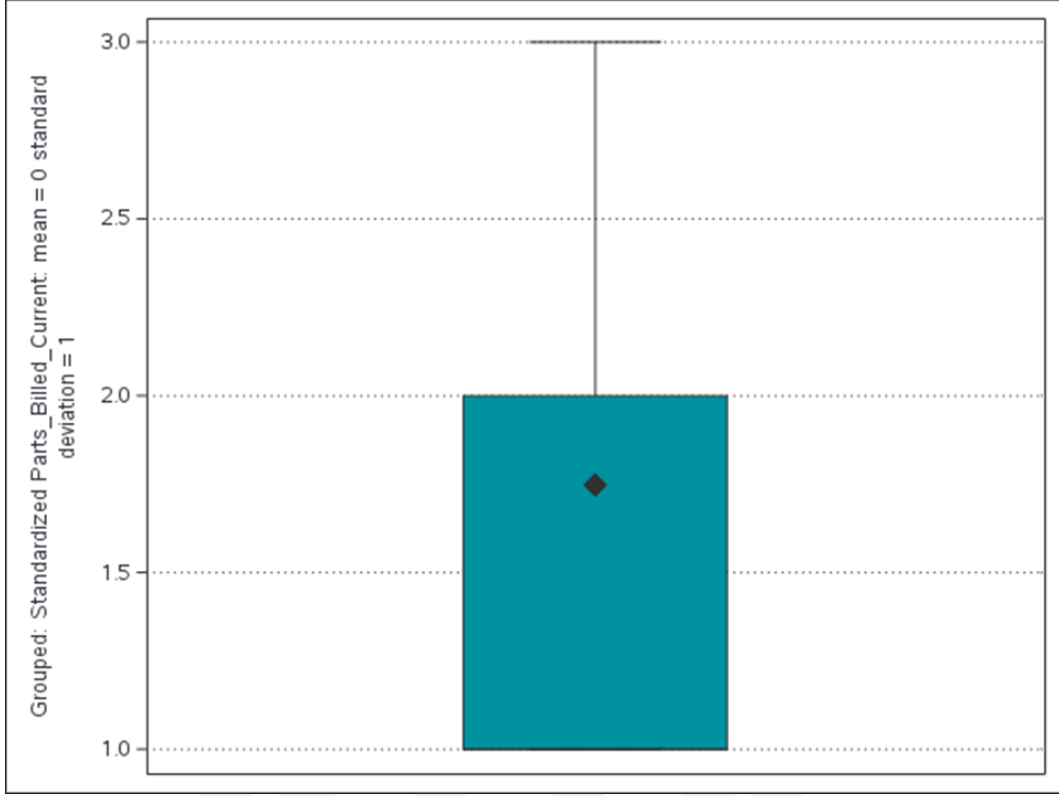
Dağılımın Çarpıklığı incelendiğinde, grafikteki uzun üst whisker ve çok sayıda uç değer, verinin sağa çarpık (positively skewed) olduğunu göstermektedir. Bu, bazı gözlemlerin diğerlerine kıyasla çok yüksek değerlere sahip olduğunu ve bu gözlemlerin standart sapma birimlerine göre oldukça uzak olduğunu işaret eder.

Uç değerlerin modelleme sonuçlarına etkisi analiz edilmelidir. Bu nedenle değişkenlere gruplama yapılmasına karar verilmiştir.

Uç değerlerin kaynağı analiz edilmelidir. Uç değerler bir veri hatası değilse (örneğin, olağandışı durumlar veya gerçek sistem özellikleri), bu gözlemlerin çıkarılmaması önemlidir.

Çeyrekler arası aralığın dar olması, değişkenin büyük kısmının düşük değerlerde toplandığını gösterir. Bu durum, modelleme sırasında değişkenin etkisinin baskın olmamasına neden olabilir.

Gruplanmış Parça Maliyeti Değişkeni Kutu Grafiği Şekil 2.4'te sunulmuştur.



Şekil 2.4 Gruplanmış parça maliyeti - kutu grafiği

Şekil 2.4, Parça Maliyeti değişkenin gruplanmış ve standartlaştırılmış değerlerini göstermektedir. Grafiği şu şekilde yorumlanabilir:

Merkezî Eğilim (Ortanca ve Ortalama) incelendiğinde, verinin genelde simetrik ve aşırı bir eğilimden uzak olduğunu gösterir.

Dağılım (İlk Çeyrek - Üçüncü Çeyrek) incelendiğinde, kutu, veri dağılımının %25 ile %75 arasında kalan kısmını (IQR - interquartile range) ifade eder. Bu, veri setinin çoğunun yoğunlaştığı bölgedir. Bu grafikte, kutunun alt sınırı yaklaşık 1,0 civarındayken, üst sınırı 2,0 civarındadır.

Değişkenlik (Whisker'lar ve Aykırı Değerler) incelendiğinde, gruplanmış ve standartlaştırılmış veride uç değerlerin etkili bir şekilde kontrol altına alındığını ve büyük uç noktalar olmadığını gösterir.

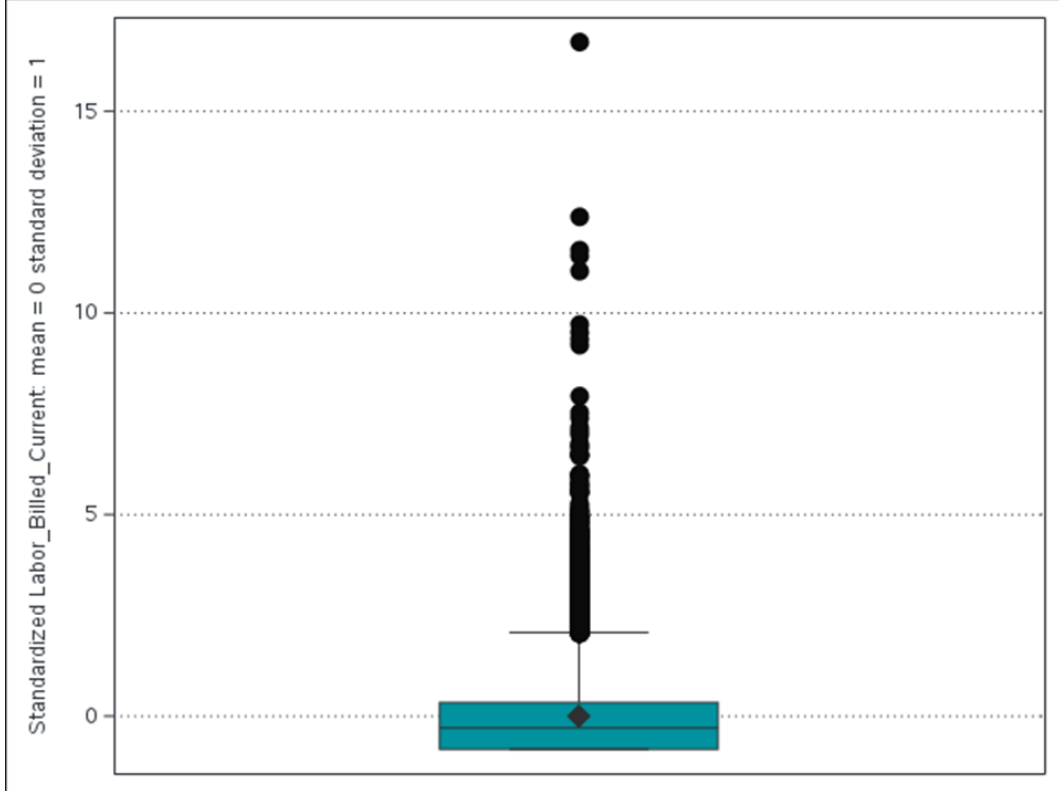
Gruplamanın Etkisi incelendiğinde, gruplama sayesinde veri standartlaştırılmış (ortalama = 0, standart sapma = 1) ve uç değerlerin etkisi minimize edilmiştir. Verideki çeyrekler arasındaki dağılım oldukça yoğun bir hale gelmiştir. Bu, modelleme ve analiz açısından daha stabil ve öngörülebilir sonuçlar sağlayabilir.

Şekil 2.4, verinin iyi bir şekilde gruplanıp standartlaştırıldığını göstermektedir.

Verinin büyük kısmı 1 ile 2 aralığında yoğunlaşmıştır ve aykırı değerlerin olmaması, analizlerde ve modelleme süreçlerinde daha sağlam sonuçlar elde edilmesine olanak tanır. Bu yapı, model performansını artırabilir ve istatistiksel analizlerde daha güvenilir bir temel sağlar.

2.3.3.2 İşçilik Maliyeti

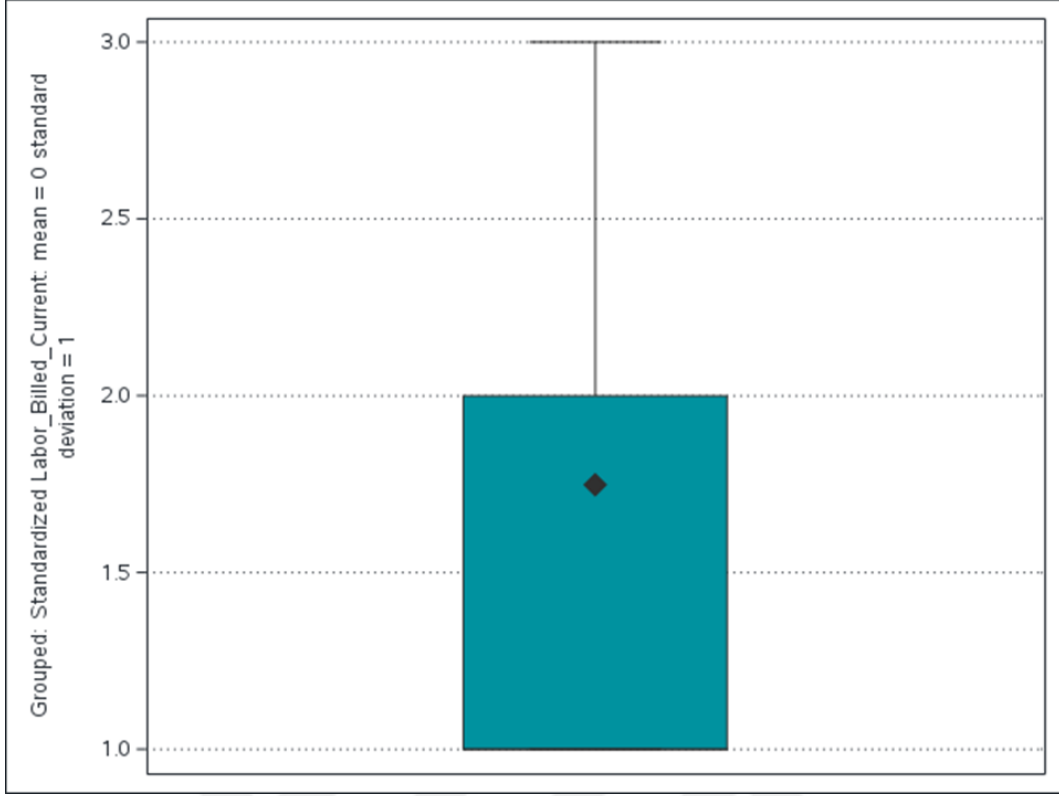
Standartlaştırılmış İşçilik Maliyeti değişkeni kutu grafiği Şekil 2.5'te sunulmuştur.



Şekil 2.5 Standartlaştırılmış işçilik maliyeti - kutu grafiği

Şekil 2.5, standartlaştırılmış işçilik maliyeti değişkeninin kutu grafiğidir ve verinin büyük kısmının düşük değerler arasında yoğunlaştığını, ancak çok sayıda aykırı değer bulunduğunu göstermektedir. Kutunun alt ve üst sınırları, veri dağılımının %25 ile %75'ini temsil ederken, üst whisker sınırı aşan birçok uç değer gözlenmektedir. Özellikle 10'un üzerinde yoğunlaşan uç değerler, veri setinin bu kısmında aşırı değişkenliğe işaret eder. Bu durum, modelleme süreçlerinde uç değerlerin etkilerinin dikkate alınmasını veya gerekli dönüşüm/gruplama işlemlerinin yapılmasını gerektirebilir. Aykırı değerlerin etkili bir şekilde ele alınması, daha sağlam bir analiz ve modelleme sağlar.

Gruplanmış İşçilik Maliyeti değişkeni kutu grafiği Şekil 2.6'de sunulmuştur.



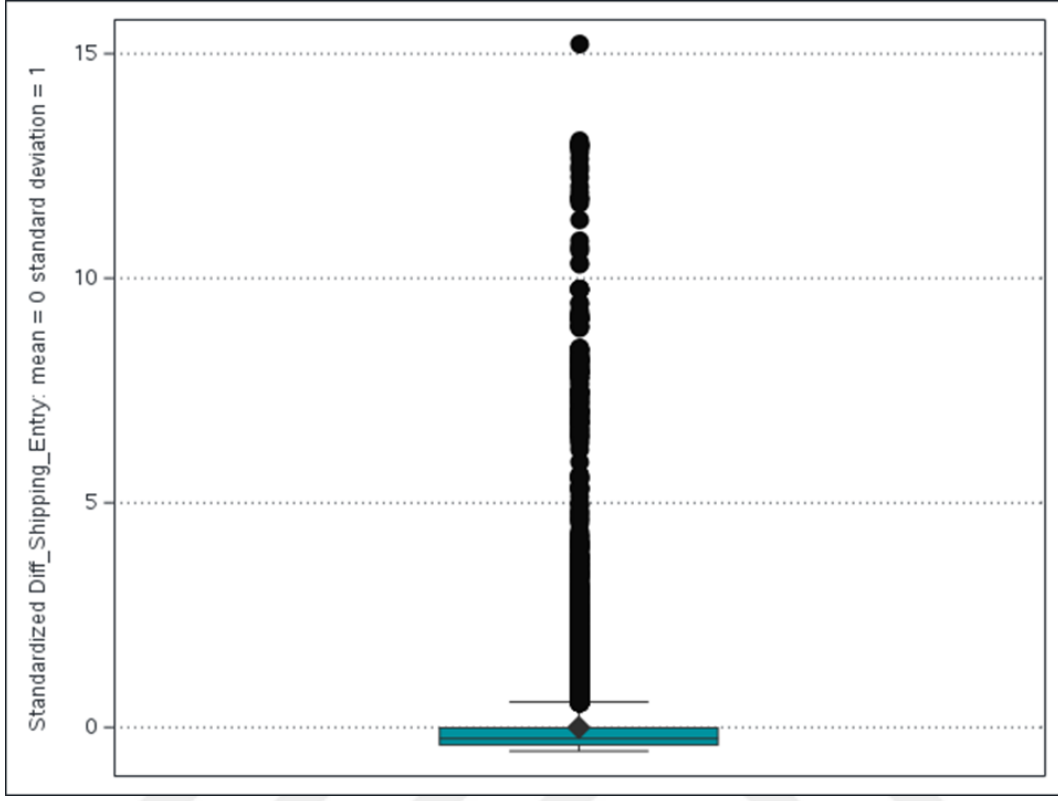
Şekil 2.6 Gruplanmış işçilik maliyeti - kutu grafiği

Şekil 2.6, standartlaştırılmış ve sonrasında gruplanmış bir değişkenin dağılımını göstermektedir. Kutu grafiği, verinin büyük kısmının 1 ile 2 arasında yoğunlaştığını, medyanın (kutunun ortasındaki çizgi) ve ortalamanın (siyah elmas) birbirine yakın olduğunu, dolayısıyla verinin nispeten simetrik bir dağılıma sahip olduğunu ortaya koymaktadır. Üst whisker yaklaşık 3 değerine kadar uzanmakta olup, aykırı değerlerin olmadığı veya etkilerinin büyük ölçüde kontrol altına alındığı görülmektedir. Gruplama sayesinde uç değerlerin etkisi minimize edilmiş ve veri daha dengeli bir yapı sergilemektedir, bu da modelleme sürecinde daha tutarlı sonuçlar elde edilmesini sağlayabilir.

İşçilik Maliyeti değişkeninin gruplanmış ve standartlaştırılmış kutu grafikleri, modelleme açısından farklı avantajlar sunar. Standartlaştırma, ölçek farklılıklarını gidererek regresyon, Destek Vektör Makineleri (Support Vector Machine - SVM) ve En Yakın Komşu Algoritması (K-Nearest Neighbour – KNN) gibi ölçek duyarlı modeller için uygun hale gelirken, uç değerlerin etkisini tam olarak ortadan kaldıramaz. Gruplama ise uç değerleri etkili bir şekilde minimize ederek veriyi daha dengeli ve kompakt bir yapıya dönüştürür, bu da Karar Ağaçları ve Rastgele Orman gibi algoritmaların performansını artırır. Gruplama, veri kaybına yol açabileceği için sürekli bilginin önemli olduğu regresyon gibi modellerde sınırlı kullanılabilir. Seçim, kullanılan modelleme yöntemine ve değişkenin özelliklerine bağlıdır.

2.3.3.3 Çevrim Süresi

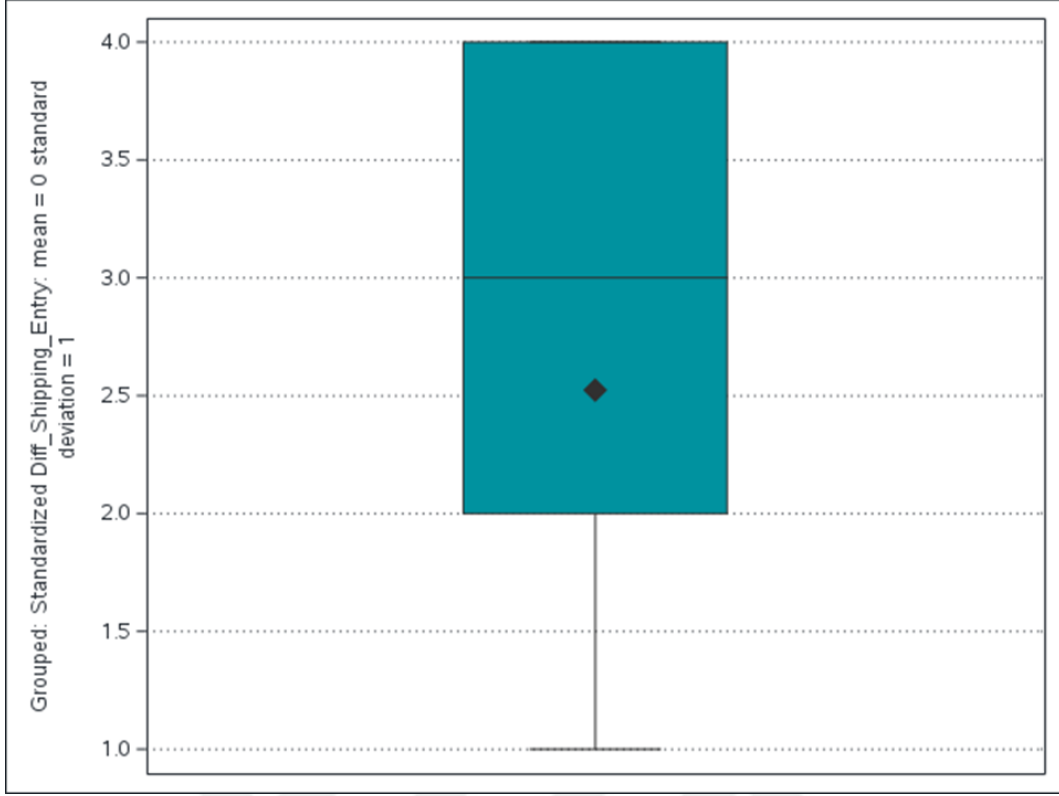
Standartlaştırılmış Çevrim Süresi değişkeni kutu grafiği Şekil 2.7’de sunulmuştur.



Şekil 2.7 Standartlaştırılmış çevrim süresi - kutu grafiği

Şekil 2.7, standartlaştırılmış Diff Shipping Entry, Çevrim Süresi, değişkeninin dağılımını değerlendirirken, veride ciddi bir uç değer sorunu olduğunu ve verinin büyük kısmının düşük değerlere yoğunlaştığını açıkça göstermektedir. Standartlaştırma, veri ölçeklerini normalize ederek modelleme için uygun bir temel oluşturmuş olsa da, üst whisker sınırını aşan ve özellikle 10-15 arasında yoğunlaşan uç değerlerin, analiz ve modelleme süreçlerinde dikkatlice ele alınması gerekmektedir.

Gruplanmış Çevrim Süresi değişkeni kutu grafiği Şekil 2.8’de sunulmuştur.

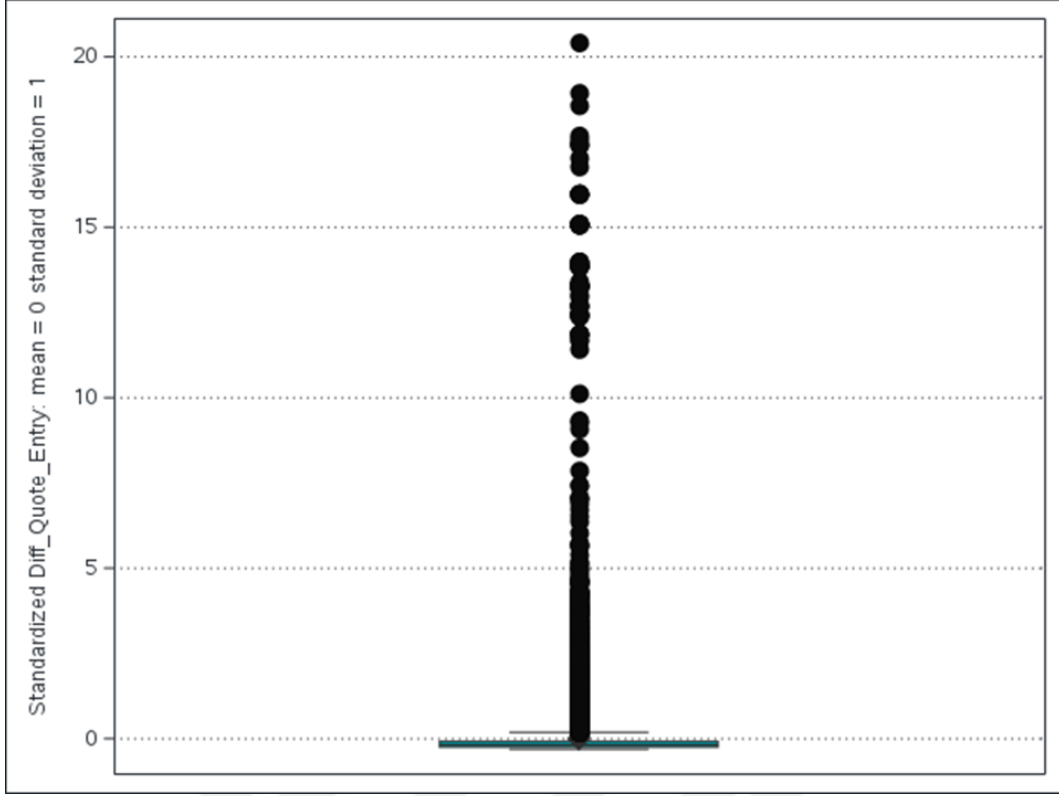


Şekil 2.8 Gruplanmış çevrim süresi - kutu grafiği

Şekil 2.8, gruplama sonucunda verinin daha dengeli bir hale geldiği ve uç değerlerin etkisinin ortadan kaldırıldığı görülmektedir. Verinin büyük bir kısmı 2 ile 4 arasında yoğunlaşmış olup, medyan ve ortalama değerler (siyah elmas) birbirine oldukça yakındır, bu da dağılımın simetrik olduğunu göstermektedir. Alt whisker sınırı yaklaşık 1'e kadar uzanırken, üst sınırda herhangi bir uç değer bulunmamaktadır. Bu yapı, uç değerlerin etkisiz hale getirilmesiyle modelleme sürecinde daha stabil ve öngörülebilir sonuçlar elde edilebileceğini göstermektedir. Bu dönüşüm, özellikle regresyon ve ağaç tabanlı modellerde performans iyileştirmesi sağlayabilir.

2.3.3.4 Fiyat Teklif Süresi

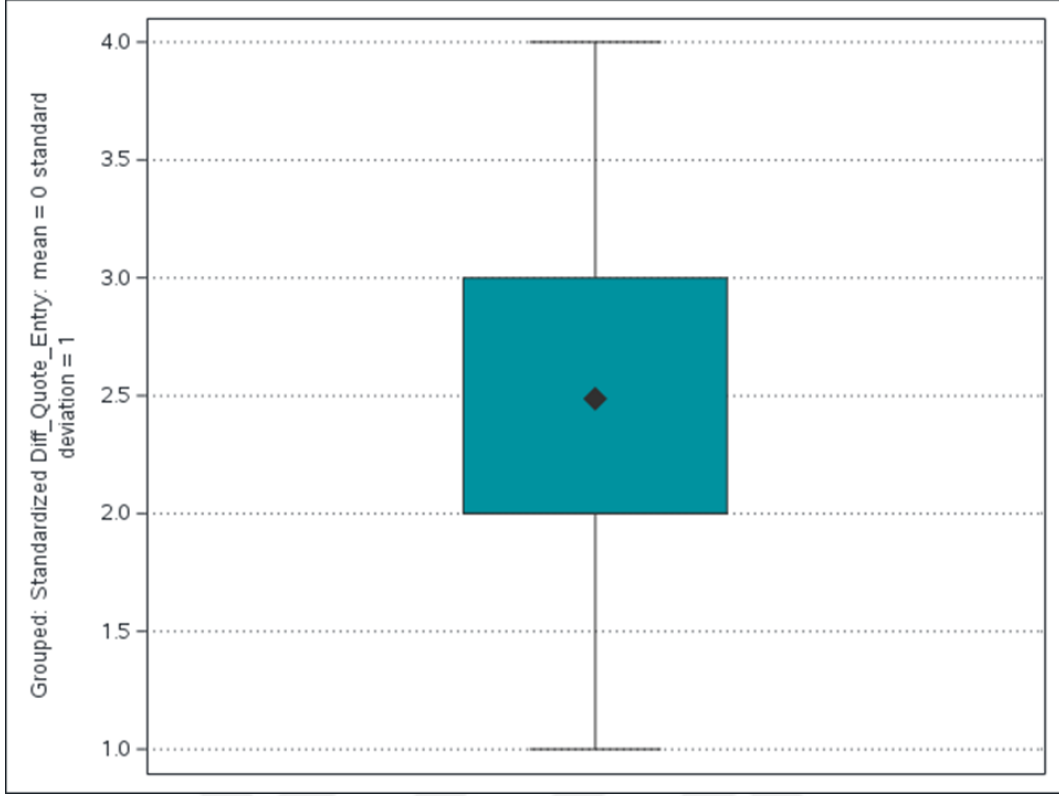
Standartlaştırılmış Fiyat Teklif Süresi değişkeni kutu grafiği Şekil 2.9'da sunulmuştur.



Şekil 2.9 Standartlaştırılmış fiyat teklif süresi - kutu grafiği

Şekil 2.9, standartlaştırılmış Diff Quote Entry, Fiyat Teklifi Süresi, değişkeninin dağılımını göstermekte olup, verinin büyük bir kısmının düşük değerlerde yoğunlaştığını ancak önemli sayıda uç değer bulunduğunu ortaya koymaktadır. Değerlerin çoğu 0'a yakın bir aralıkta toplanmış, ancak 10 ile 20 arasında dikkate değer sayıda uç değer gözlemlenmiştir. Bu durum, değişkenin modelleme süreçlerinde uç değerlere duyarlı olabilecek yöntemlerde (örneğin, regresyon) sapmalara yol açabileceğini göstermektedir. Veri, standartlaştırma ile ölçek açısından normalize edilmiş olsa da uç değerlerin etkisinin azaltılması için dönüşüm (logaritmik veya karekök) veya uç değer tespiti yöntemleriyle ek işlemler gerekebilir.

Gruplanmış Fiyat Teklif Süresi değişkeni kutu grafiği Şekil 2.10'da sunulmuştur.

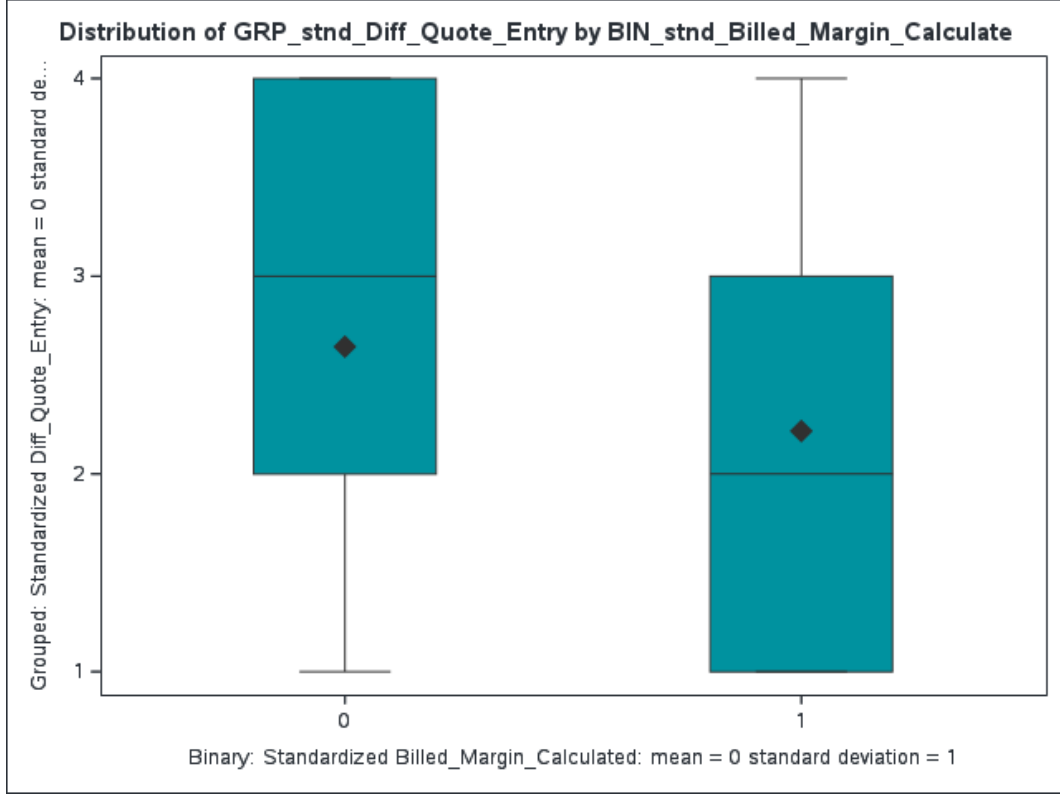


Şekil 2.10 Gruplanmış fiyat teklif süresi - kutu grafiği

Şekil 2.10, gruplama sonucunda verinin daha dengeli bir hale geldiği ve uç değerlerin etkisinin ortadan kaldırıldığı görülmektedir. İlk grafik standartlaştırılmış değişkenlerle, ikinci grafik ise gruplanmış değişkenlerle çizildiği için, veri yapısındaki farklılıklar bu iki grafiğin dağılımlarına yansımıştır. İlk grafik olan Diff Quote Entry, Fiyat Teklif Süresi, standartlaştırılmış olduğu için veri değerleri arasındaki uçurum daha belirgin hale gelmiş ve bu da yüksek sayıda aykırı değer (10-20 arasında) oluşmasına neden olmuştur. Bu durum, değişkenin geniş bir aralıkta dağıldığını göstermektedir. Buna karşın, ikinci grafik olan Diff Quote Entry, gruplanmış değişkenlerle çizildiği için veriler daha dar bir aralıkta yoğunlaşmış (1-4 arasında) ve düzenli bir dağılım sergilemiştir. Ayrıca, gruplanmış yapının etkisiyle herhangi bir aykırı değer görülmemiştir. Bu farklılık, standartlaştırma ile verinin uç noktalarının daha görünür hale geldiğini, gruplamanın ise veriyi daha kontrollü ve özet bir şekilde temsil ettiğini açıkça ortaya koymaktadır.

2.3.4 Gruplanmış Değişkenler ve Kar Marjı İncelemesi

Fiyat Teklif Süresi değişkeni için kutu grafiği Şekil 2.11’de sunulmuştur.



Şekil 2.11 Fiyat teklif süresi ile hedef değişken - kutu grafiği

Şekil 2.11, GRP stnd Diff Quote Entry değişkeninin (Fiyat Teklif Süresi), hedef değişken olan BIN stnd Billed Margin Calculate (0 ve 1 sınıfları) değişkenine göre dağılımını göstermektedir. Bu tür bir kutu grafiği (boxplot), değişkenlerin iki farklı sınıf arasındaki dağılımlarını karşılaştırmak için etkili bir araçtır.

GRP stnd Diff Quote Entry değişkeni, gruplanmış Fiyat Teklifi Süresi, iki sınıfta (0 ve 1) farklı dağılım göstermektedir. Kutu grafiği, verinin medyanını, dörtte birlik dilimlerini (Q1 ve Q3) ve uç değerleri vurgulamaktadır. Grafiğin y eksenini, Fiyat Teklif Süresi değişkeninin standartlaştırılmış değerlerini temsil etmektedir.

Sınıf 0 (BIN stnd Billed Margin Calculate = 0) için dağılım ve medyan incelendiğinde, bu sınıfta gruplanmış fiyat teklif süresi değerlerinin medyanı, sınıf 1'e kıyasla daha yüksek bir seviyededir. Medyan değer, kutunun orta çizgisiyle gösterilmektedir.

Değişim Aralığı (IQR) incelendiğinde, Sınıf 0'daki IQR (interquartile range), yani Q1 ile Q3 arasındaki mesafe, daha geniştir. Bu durum, verilerin daha fazla varyasyona sahip olduğunu göstermektedir.

Aykırı Değerler incelendiğinde, Sınıf 0'da bazı uç değerler (whiskers'ın üst sınırının ötesinde) gözlemlenebilir.

Sınıf 1 (BIN stnd Billed Margin Calculate = 1) incelendiğinde ise, bu sınıfta GRP stnd Diff Quote Entry, gruplanmış Fiyat Teklif Süresi, değerlerinin medyanı, sınıf 0'a göre daha düşük bir seviyededir. Bu, daha düşük fiyat teklif sürelerinin, kar marjı (Billed Margin Calculated) açısından pozitif bir etkisi olabileceğini göstermektedir.

Değişim Aralığı (IQR) incelendiğinde, sınıf 1'deki IQR daha dar bir aralık göstermektedir, bu da verilerin daha odaklanmış olduğunu ve düşük varyasyona sahip olduğunu işaret eder.

Aykırı Değerler incelendiğinde, Sınıf 1'de uç değerler sınırlıdır ve veriler daha homojen bir dağılım sergilemektedir.

Sınıflar arasındaki karşılaştırma yapıldığında, Sınıf 1'de gruplanmış Fiyat Teklif Süresi'nin medyanı ve genel dağılımı, Sınıf 0'dan belirgin şekilde daha düşüktür. Bu, daha kısa fiyat teklif sürelerinin (TTQ) yüksek kar marjıyla ilişkili olabileceğini gösterir.

Varyasyon incelendiğinde, sınıf 0'daki geniş varyasyon, bu sınıftaki işlemlerin daha heterojen olduğunu, yani fiyat teklifi sürelerinde daha fazla belirsizlik veya kontrol eksikliği olduğunu işaret edebilir.

Daha kısa fiyat teklif sürelerinin (TTQ), kar marjını artırmada etkili bir faktör olduğu gözlemlenmektedir. Bu durum, müşterilere daha hızlı geri dönüş sağlamanın operasyonel verimliliği ve kârlılığı artırabileceğini işaret eder.

Operasyonel Verimlilik açısından, Sınıf 1'in daha düşük varyasyon göstermesi, bu işlemlerin daha standart hale getirildiğini veya süreçlerin daha iyi yönetildiğini işaret eder. Operasyonel süreçlerde bu standardizasyonun tüm sınıflara uygulanması önerilebilir.

Uç Değerlerin Yönetimi açısından, Sınıf 0'daki geniş varyasyon ve olası uç değerler, operasyonel süreçlerdeki kontrol eksikliğini gösterebilir. Bu durumun, süreçlerin gözden geçirilmesiyle iyileştirilmesi önerilebilir.

Şekil 2.11, GRP stnd Diff Quote Entry değişkeninin kar marjı üzerindeki etkisini açıkça göstermektedir.

Çevrim Süresi değişkeni için kutu grafiği Şekil 2.12'de sunulmuştur.



Şekil 2.12 Çevrim süresi ile hedef değişken - kutu grafiği

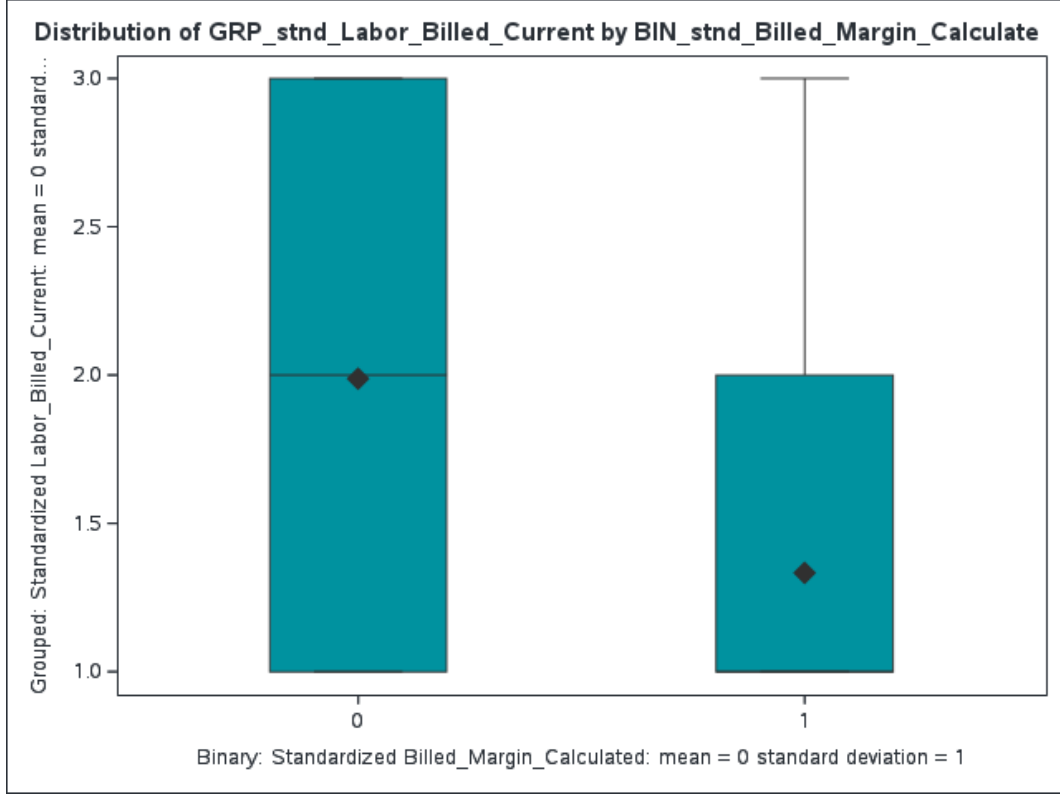
Şekil 2.12, standartlaştırılmış ve gruplanmış Çevrim Süresi değişkeninin kar marjı üzerindeki etkisini açıkça ortaya koymaktadır.

Daha kısa parça teslim sürelerinin (TAT) yüksek kar marjı ile ilişkili olması, operasyonel süreçlerin zaman yönetimi açısından önemini vurgular.

Ortalama ve Medyan Farkı incelendiğinde, sınıf 0'da çevrim süreleri (TAT) genel olarak daha uzunken, Sınıf 1'de bu süreler kısalmaktadır. Bu, daha kısa teslim sürelerinin kar marjını artırmada önemli bir faktör olduğunu gösterebilir.

Varyasyon incelendiğinde, Sınıf 0'daki daha yüksek varyasyon, bu süreçlerin daha az optimize edildiğini veya daha karmaşık bir yapıya sahip olduğunu işaret eder.

İşçilik Maliyeti değişkeni için kutu grafiği Şekil 2.13'te sunulmuştur.



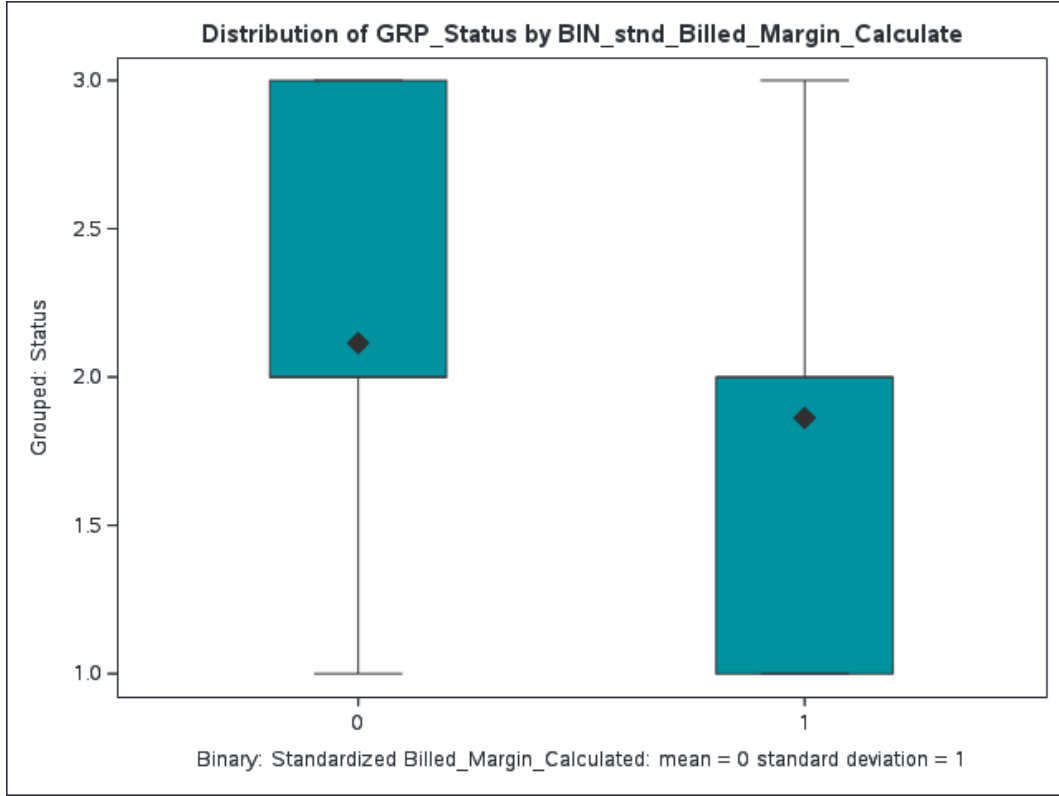
Şekil 2.13 İşçilik maliyeti ile hedef değişken - kutu grafiği

Şekil 2.13, işçilik maliyetlerinin (GRP stnd Labor Billed Current) kar marjı üzerindeki etkisini net bir şekilde göstermektedir. Daha düşük işçilik maliyetleri, daha yüksek kar marjı ile ilişkilidir. Bu durum, işçilik maliyetlerini azaltma ve süreçleri optimize etme konusundaki stratejik önceliği vurgular.

Yüksek işçilik maliyetlerinin düşük kar marjı ile ilişkili olduğu görülmektedir. Bu nedenle, işçilik maliyetlerini azaltmak ve kontrol etmek, kârlılığını artırmak için kritik bir strateji olacaktır.

Sınıf 0'daki geniş varyasyon, standartlaştırılmış süreçlerin eksikliğini ve işçilik maliyetlerinin farklı projelerde büyük ölçüde değişkenlik gösterdiğini ifade eder. Standart operasyonel prosedürlerin uygulanması önerilir.

Parça Durumu değişkeni için kutu grafiği Şekil 2.14'te sunulmuştur.



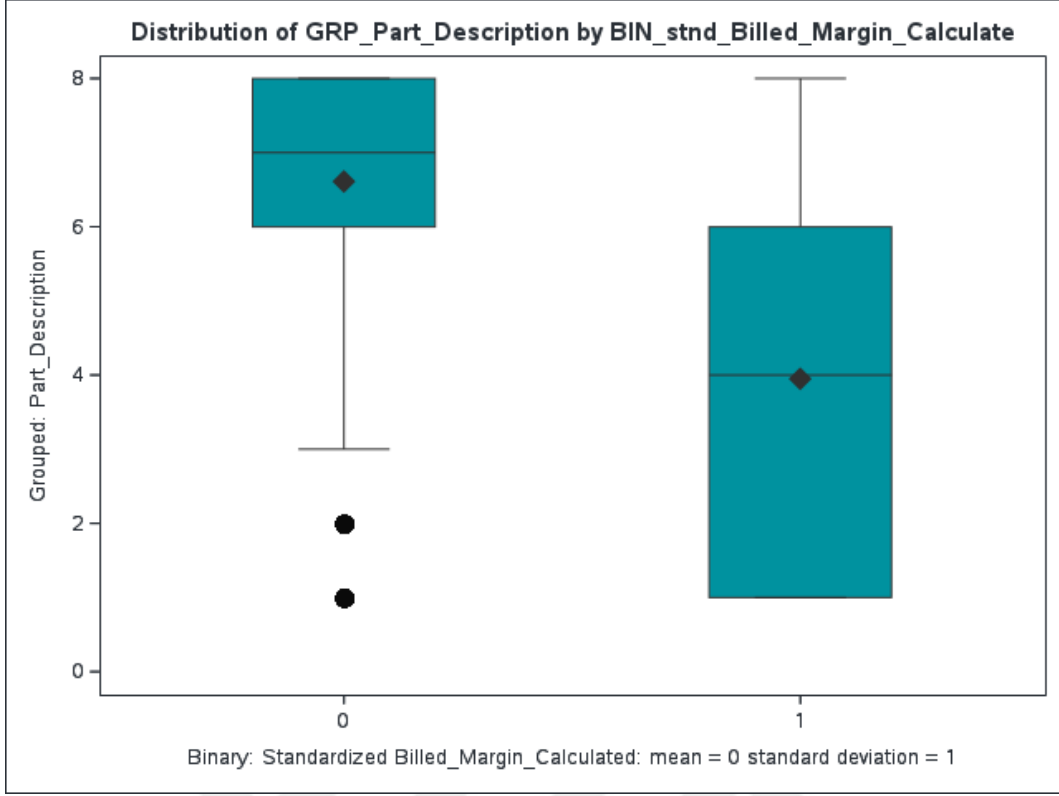
Şekil 2.14 Parça durumu ile hedef değişken - kutu grafiği

Şekil 2.14, Sınıf 0 (düşük kâr marjı) ile ilişkilendirilen örneklerde, Parça Durumu değişkeninin genel olarak daha geniş bir varyasyona sahip olduğu görülmektedir. Bu durum, bu sınıfta yer alan işlemlerin daha karmaşık ve çeşitli bir yapıya sahip olduğunu veya süreçlerin optimize edilmediğini gösterebilir.

Sınıf 1 (yüksek kâr marjı) ile ilişkilendirilen örneklerde, Parça Durumu değişkeni daha düşük bir medyana ve daha dar bir dağılıma sahiptir. Bu, yüksek kâr marjına sahip işlemlerin daha standart ve düzenli süreçlerle gerçekleştirildiğini ifade edebilir.

Parça Durumu değişkenindeki geniş varyasyon, özellikle sınıf 0'daki süreçlerin daha iyi optimize edilmesi gerektiğini vurgulamaktadır. Standart operasyonel prosedürlerin geliştirilmesi ve süreçlerin düzenlenmesi, kâr marjını artırmada önemli bir adım olabilir.

Parça Tanımı değişkeni için kutu grafiği Şekil 2.15'de sunulmuştur.



Şekil 2.15 Parça tanımı ile hedef değişken - kutu grafiği

Şekil 2.15, Sınıf 0 (düşük kâr marjı) için Parça Tanımı değişkeninin medyan değeri daha yüksektir ve verilerin üst kısmında aykırı değerler bulunmaktadır. Bu durum, belirli parça tanımlarıyla ilişkili işlemlerin kârlılığının düşük olduğunu ve süreçlerin karmaşık veya verimsiz olduğunu gösterebilir.

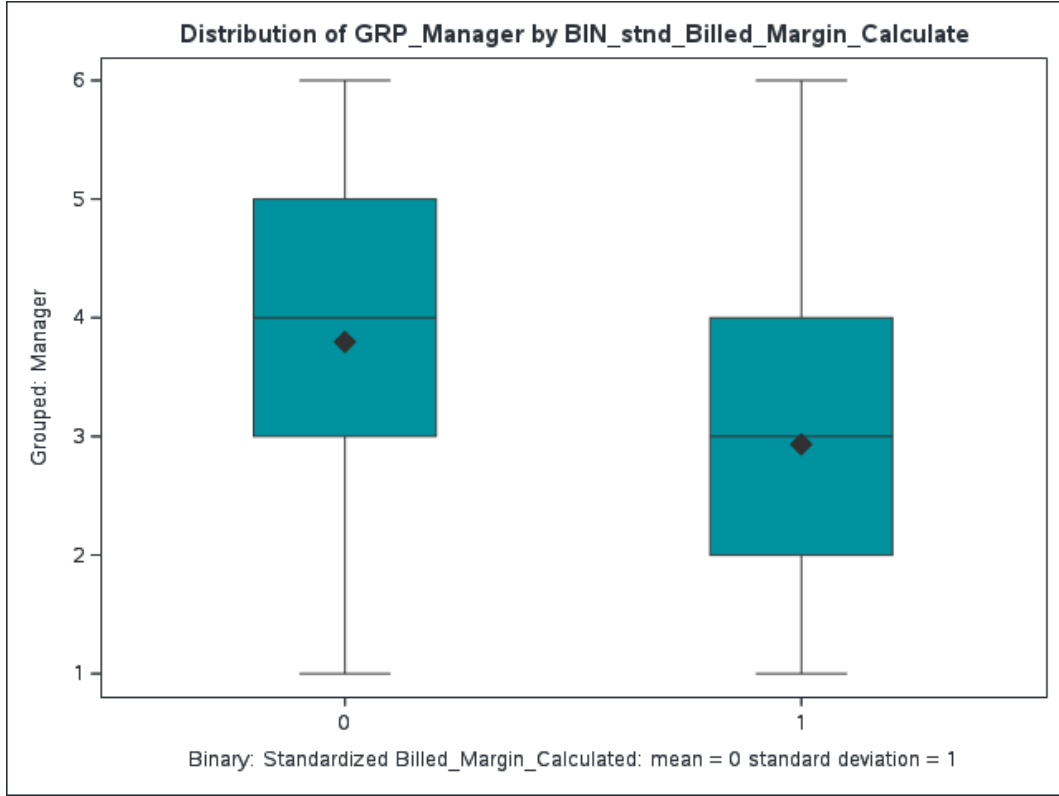
Sınıf 1 (yüksek kâr marjı) için medyan değer daha düşüktür ve veri dağılımı daha geniştir. Bu durum, belirli parça tanımlarının yüksek kâr marjına katkı sağladığını ve bu sınıftaki süreçlerin daha verimli bir şekilde yönetildiğini gösterebilir.

Sınıf 0'daki aykırı değerler, belirli parçalarla ilgili süreçlerin maliyetli veya düşük kârlılık getirdiğini işaret etmektedir. Bu parçaların analiz edilmesi, operasyonel verimlilik açısından önemli içgörüler sağlayabilir.

Sınıf 1'de Parça Tanımı değişkeninin daha düşük medyan değere ve geniş bir dağılıma sahip olması, yüksek kâr marjı sağlayan parçaların daha çeşitli olduğunu göstermektedir. Operasyonel süreçlerde bu parçaların önceliklendirilmesi, kârlılığın artırılması için stratejik bir yaklaşım olabilir.

Bu analiz, işletmelerin hangi parça tanımlarına odaklanmaları gerektiğini ve süreçlerini nasıl optimize edebileceklerini belirlemelerine yardımcı olmaktadır.

Yönetici değişkeni için Kutu Grafiği Şekil 2.16’da sunulmuştur.



Şekil 2.16 Yönetici ile hedef değişken - kutu grafiği

Şekil 2.16, Sınıf 0 (düşük kâr marjı) için Yönetici değişkeninin medyan değeri daha yüksektir, ancak veri dağılımı geniştir. Bu durum, düşük kâr marjı ile ilişkilendirilen süreçlerde yöneticilerin performansındaki farklılıkları veya süreçlerin karmaşıklığını işaret edebilir.

Sınıf 1 (yüksek kâr marjı) için medyan değer daha düşüktür ve veri dağılımı daha dardır. Bu, yüksek kârlılık sağlayan süreçlerin daha iyi optimize edildiğini ve yöneticilerin daha etkin bir rol oynadığını gösterebilir.

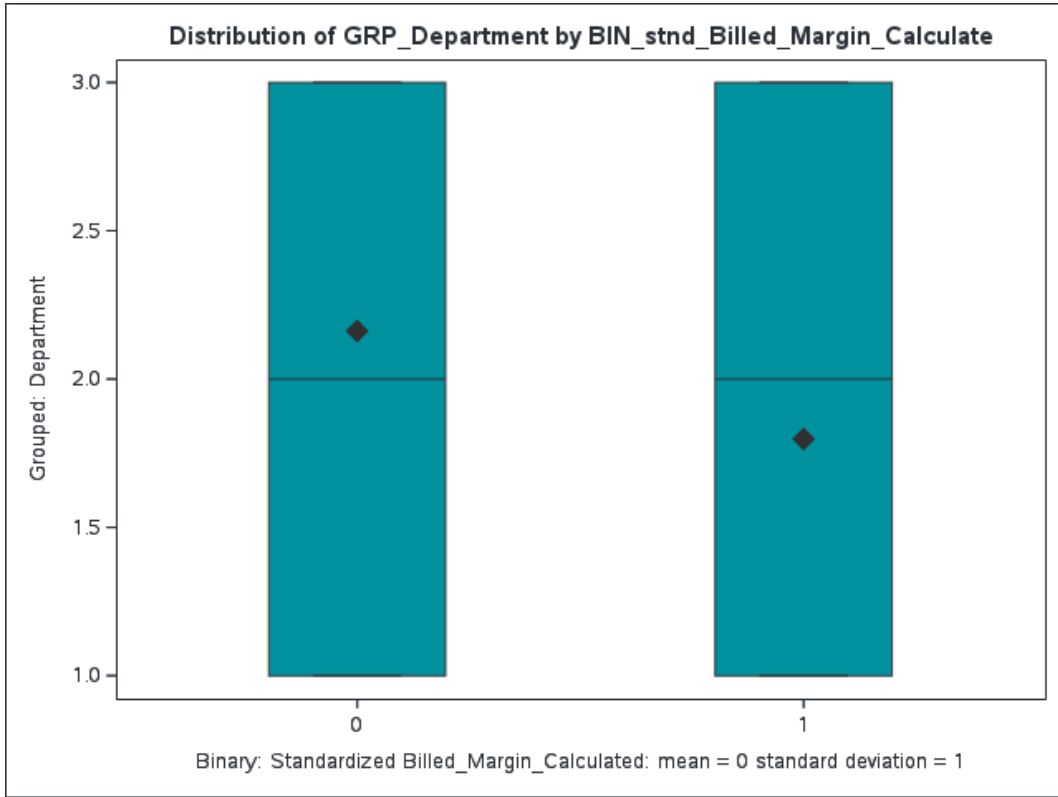
Sınıf 0'daki daha geniş varyasyon, yöneticiler arasında süreç yönetimi ve sonuçları açısından farklılıkların daha belirgin olduğunu göstermektedir. Bu durum, süreçlerin standardizasyonu ve iyileştirilmesi gerektiğine işaret edebilir.

Sınıf 1'de varyasyonun daha düşük olması, belirli yöneticilerin yüksek kârlılık sağladığını ve süreçlerin daha tutarlı bir şekilde yönetildiğini göstermektedir.

Yöneticilerin performansının kârlılığa etkisi, iş süreçlerinin yönetiminde liderlik rollerinin önemini vurgulamaktadır. Sınıf 1'deki yöneticilerin belirlenmesi ve bu yöneticilerle süreç optimizasyonu yapılması, işletmeler için stratejik bir avantaj

sağlayabilir.

İlgili Bölüm değişkeni için Kutu Grafiği Şekil 2.17’de sunulmuştur.



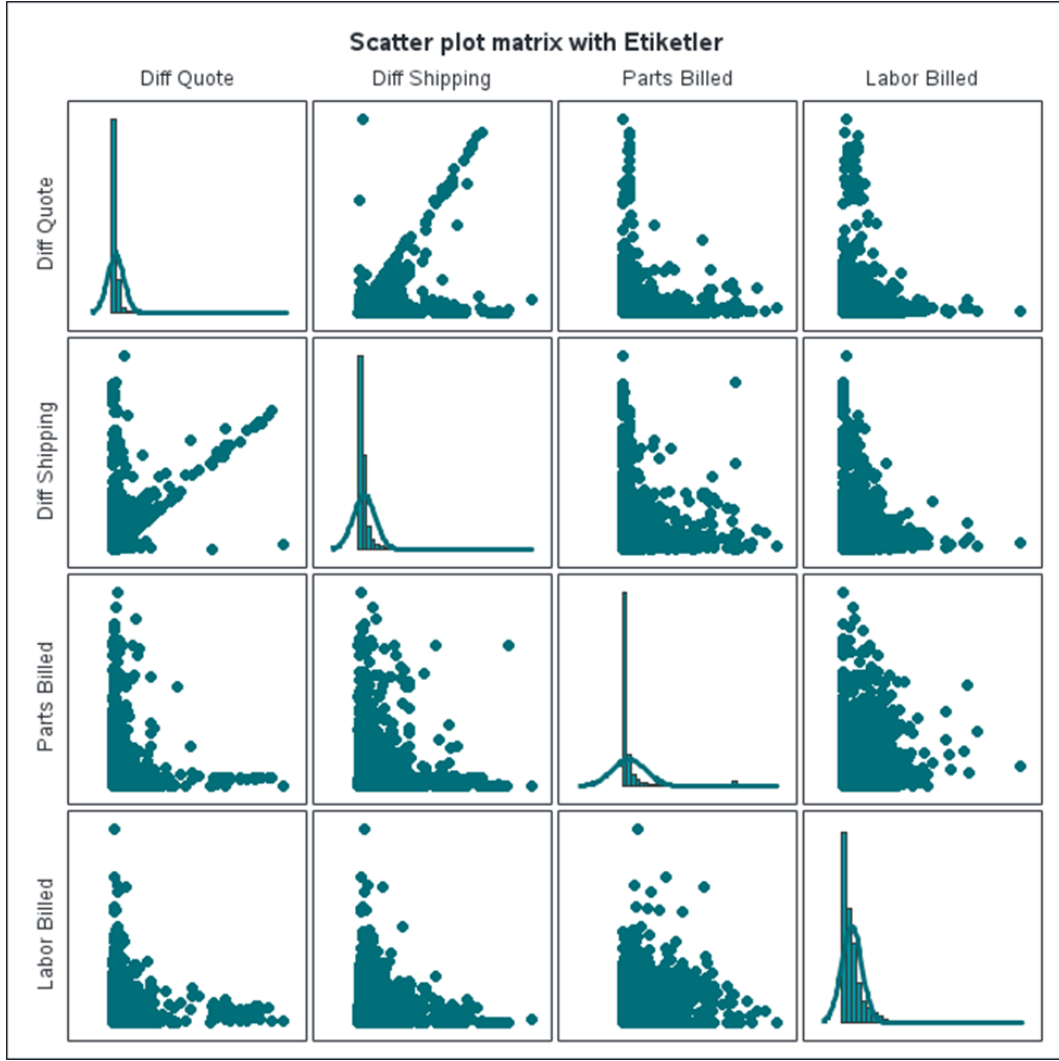
Şekil 2.17 Departman ile hedef değişkeni - kutu grafiği

Şekil 2.17, Sınıf 0 (düşük kâr marjı) ve Sınıf 1 (yüksek kâr marjı) arasında İlgili Bölüm değişkenine ilişkin medyan değerler benzerdir. Bu durum, İlgili Bölüm faktörünün genel kâr marjı üzerinde direkt bir ayrıştırıcı etkiye sahip olmadığını gösterebilir. Ancak, her iki sınıfta da varyasyon gözlenmektedir.

Her iki sınıfta da İlgili Bölüm değişkenine ilişkin dağılım geniştir ve varyasyon oranları birbirine yakındır. Bu, bölümlerin süreç yönetimi ve sonuçlar açısından homojen bir yapıya sahip olmadığını gösterebilir. Özellikle, bölümler arası standartizasyon eksikliğine işaret edebilir.

2.3.5 Standartlaştırılmış ve Gruplanmış Değişkenler için İlişki Matrisi

Tüm sürekli değişkenler için standartlaştırılmış değişkenler arası ilişki matrisi Şekil 2.18’de sunulmuştur.



Şekil 2.18 Standartlaştırılmış değişkenler arası ilişki matrisi

Şekil 2.18, bir dağılım grafiği matrisi olarak adlandırılır ve verideki değişkenler arasındaki ilişkileri görselleştirmek için kullanılır. Grafik, standartlaştırılmış verilerle oluşturulmuş ve değişkenler arasındaki çift yönlü ilişkileri incelerken dağılımlarını da ortaya koyar.

Grafikteki her hücre, bir değişken çiftinin dağılım grafiğini temsil eder. Eğilim çizgileri veya veri yoğunlukları, bu değişkenler arasındaki ilişkiyi değerlendirir. Standartlaştırılmış Parts Billed Current, Parça Maliyeti ile Standartlaştırılmış Labor Billed Current, İşçilik Maliyeti, arasında yoğun veri kümeleri gözlenmektedir, ancak doğrusal bir ilişki zayıf görünmektedir.

Diagonal hücreler, her bir değişkenin kendi dağılımını gösterir. Standartlaştırılmış Parts Billed Current, Parça Maliyeti ve Standartlaştırılmış Labor Billed Current, İşçilik Maliyeti değişkenlerinin dağılımları sağa çarpık bir şekil gösterir, bu da büyük maliyet değerlerinin sınırlı sayıda olduğunu ancak etkilerinin önemli

olabileceğini gösterir.

Dağılım grafiklerinde bazı veri yoğunluklarının yanı sıra uç değerler de gözlemlenebilir. Özellikle Standartlaştırılmış Diff Shipping Entry, Çevrim Süresi, ve diğer değişkenler arasında veri kümelerinin yoğunlaştığı noktalar mevcuttur.

Veriler standartlaştırıldığından, her bir değişkenin ortalaması 0 ve standart sapması 1'dir. Bu, farklı birimlerdeki değişkenlerin karşılaştırılabilir hale gelmesini sağlamıştır.

Fiyat Teklifi Süresi ve Çevrim Süresi arasındaki ilişkiler, süreç iyileştirmelerinde kullanılabilir. Örneğin, kısa fiyat teklif süresi, maliyetleri etkileyebilir ve daha yüksek kâr marjlarına katkıda bulunabilir.

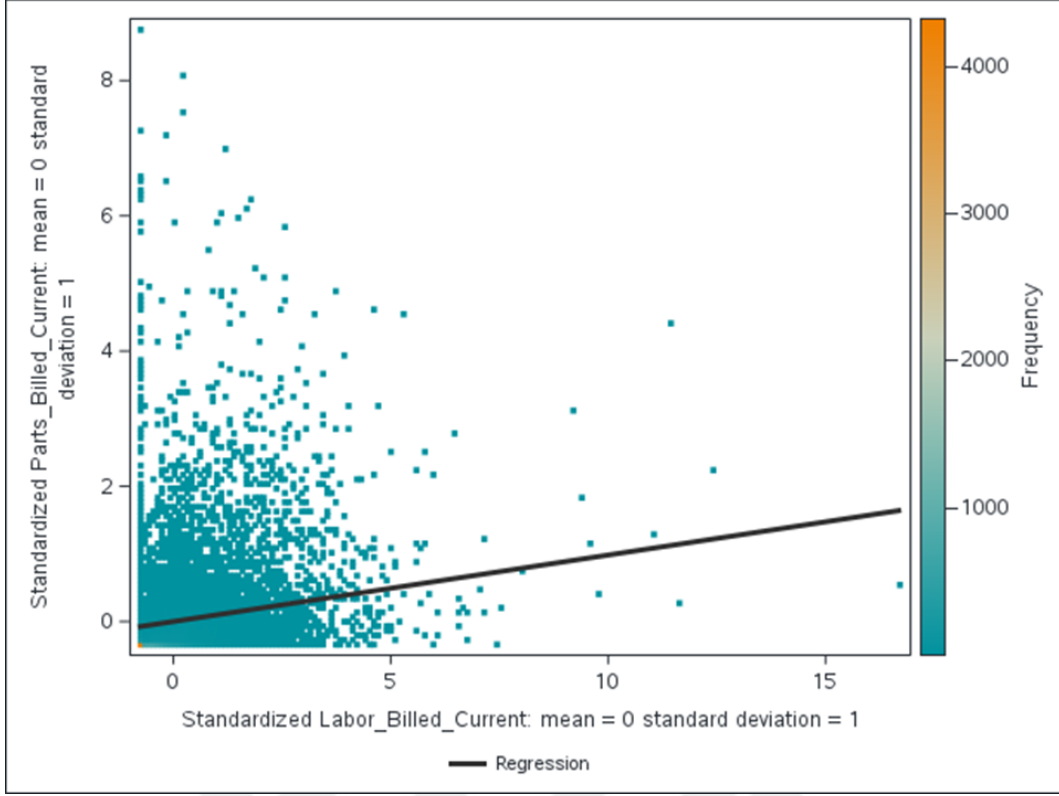
Parça ve işçilik maliyetleri arasındaki ilişki, doğrudan kâr marjını etkileyen faktörler olarak değerlendirilebilir.

Dağılım ve yoğunluk arasındaki korelasyonları anlamak, MRO süreçlerinde stratejik kararlar almak için kullanılabilir.

Şekil 2.18, değişkenlerin birbirleriyle nasıl etkileşimde bulunduğunu anlamak için güçlü bir araçtır. Görülen eğilimler ve korelasyonlar, işletme süreçlerini optimize etmek ve kâr marjlarını artırmak için rehberlik sağlayabilir.

2.3.6 Standartlaştırılmış ve Gruplanmış Değişkenler Arasında İlişki ve Yoğunluk Analizi

Standartlaştırılmış Parça Maliyeti ve İşçilik Maliyeti ilişki grafiği Şekil 2.19'da sunulmuştur.



Şekil 2.19 Yoğunluk dağılımlı standartlaştırılmış parça maliyeti ve işçilik maliyeti regresyon grafiği

Şekil 2.19, Standartlaştırılmış İşçilik Maliyeti (X eksen) ile Standartlaştırılmış Parça Maliyeti (Y eksen) değişkenleri arasındaki ilişkiyi ve yoğunluğu göstermektedir. Her iki değişken standartlaştırılmış olduğundan, ortalamaları 0 ve standart sapmaları 1'dir. Renk skalası, veri yoğunluğunu temsil eder; mavi düşük yoğunluğu, turuncuya yakın tonlar ise yüksek yoğunluğu ifade eder.

Verilerin büyük kısmı düşük X ve Y değerlerinde yoğunlaşmıştır (özellikle İşçilik Maliyeti değerinin 0 ile 5 arasında olduğu bölgede). Yüksek X değerlerinde (10 üzeri) gözlem sayısı oldukça azdır ve uç değerler mevcuttur. Y eksenindeki Parça Maliyeti değişkeni ise 0-2 arasında yoğunlaşmış ancak 8'e kadar uzanan gözlemler de bulunmaktadır.

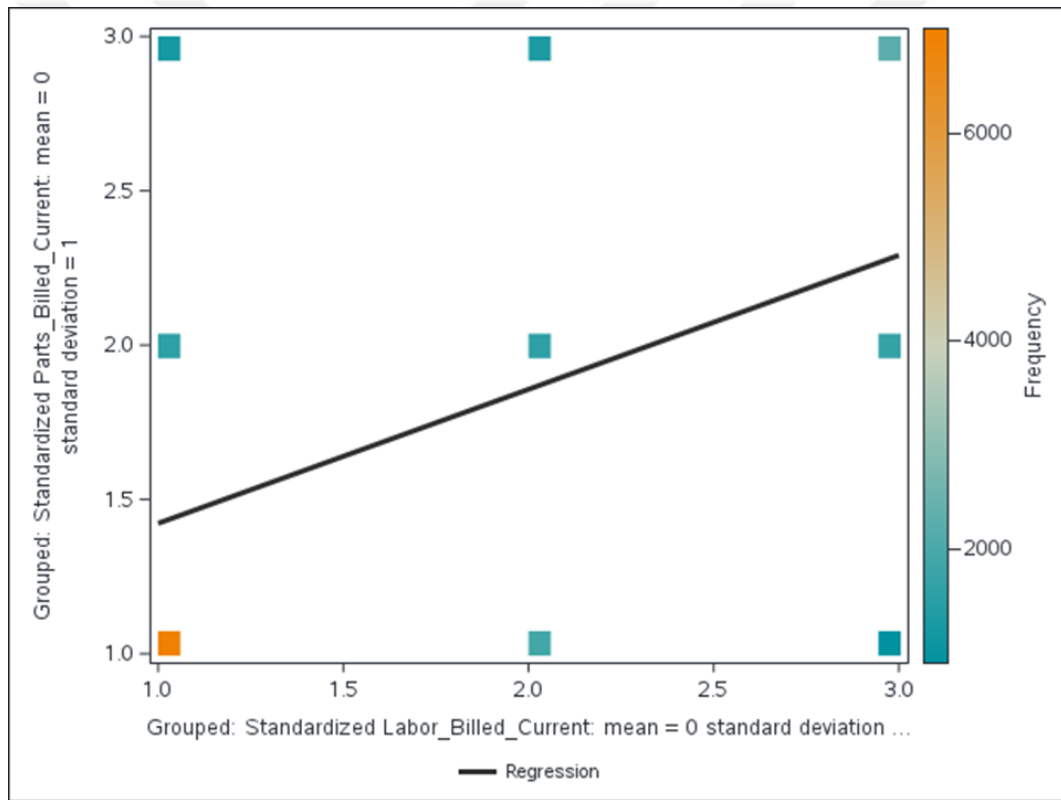
Siyah regresyon çizgisi, iki değişken arasındaki genel doğrusal eğilimi göstermektedir. Çizgi hafif pozitif eğimli olup, iki değişken arasında zayıf bir pozitif ilişki olduğunu göstermektedir. Ancak bu ilişkinin güçlü olmadığı söylenebilir.

Her iki değişkende uç değerler mevcuttur ve bu durum regresyon çizgisini etkileyebilir. Özellikle İşçilik Maliyeti değişkeninde 15 üzerindeki gözlemler, modele olan katkılarından çok sapma yaratma potansiyeline sahiptir.

İşçilik Maliyeti'ndeki yüksek değerler ve parça maliyetindeki geniş dağılım, modele sapma katabilir. Bu nedenle, uç değer yönetimi veya dönüşüm uygulanması gerekebilir.

Yoğunluğun düşük değerlerde olması, modelin düşük X ve Y değerleri üzerinde daha fazla bilgiye sahip olduğunu gösterir. Yüksek değerler için modelin daha fazla hata yapması muhtemeldir. Sonuç olarak, bu grafik, iki değişkenin modelleme sürecinde zayıf bir pozitif ilişki sunduğunu ancak uç değerlerin ve yoğunluk dengesizliğinin dikkatle ele alınması gerektiğini göstermektedir. Özellikle doğrusal modellerde, bu değişkenlerin etkisi dönüşüm yöntemleriyle optimize edilmelidir.

Gruplanmış parça maliyeti ve işçilik maliyeti ilişki grafiği Şekil 2.20'de sunulmuştur.



Şekil 2.20 Yoğunluk dağılımlı gruplanmış parça maliyeti ve işçilik maliyeti regresyon grafiği

Şekil 2.20, gruplanmış ve standartlaştırılmış İşçilik Maliyeti (X eksen) ile gruplanmış ve standartlaştırılmış Parça Maliyeti (Y eksen) değişkenleri arasındaki ilişkiyi ve yoğunluk dağılımını göstermektedir. Grafik, her iki değişkenin belirli gruplar halinde analiz edildiğini ve yoğunluğun renk skalasıyla ifade edildiğini ortaya koymaktadır. Siyah regresyon çizgisi, iki değişken arasındaki doğrusal ilişkiyi göstermektedir.

Veriler üç grup (1, 2, 3) halinde incelenmiştir. X eksenini (İşçilik Maliyeti) grupları, Y ekseninde (Parça Maliyeti) karşılık gelen gruplara göre analiz edilmiştir. Renk skalası, grupların yoğunluğunu ifade etmektedir. En yoğun grup, X eksenini 1 ve Y eksenini 1 olan bölgedir (turuncu renk, yaklaşık 6000 frekans).

Regresyon çizgisi pozitif eğimlidir, yani Labor Billed Current, Parça Maliyeti ve Parts Billed Current, Parça Maliyeti arasında pozitif bir ilişki olduğunu göstermektedir. Labor Billed Current, İşçilik Maliyeti, grubu arttıkça, Parts Billed Current'in, Parça Maliyeti artış eğiliminde olduğu görülmektedir. Ancak eğimin nispeten düşük olması, ilişkinin güçlü olmadığını ve değişkenler arasında sınırlı bir doğrusal bağ olduğunu göstermektedir.

Gruplama, uç değerlerin etkisini azaltarak ve veriyi daha düzenli bir hale getirerek analiz sürecini kolaylaştırmıştır. Regresyon çizgisi, bu iki değişken arasında pozitif bir ilişki bulunduğunu ancak bunun güçlü bir doğrusal ilişki olmadığını göstermektedir.

Bu yapı, özellikle Karar Ağaçları ve Rastgele Orman gibi kategorik grupları iyi öğrenen algoritmalar için daha uygundur. Ancak, doğrusal modellerde bilgi kaybı olabileceği dikkate alınmalıdır.

Şekil 2.20, iki değişkenin pozitif bir ilişki içinde olduğunu, ancak ilişkinin güçlü olmadığını göstermektedir. Gruplamanın sağladığı düzenli yapı sayesinde uç değerler minimize edilmiş ve değişkenler arasındaki genel eğilim net bir şekilde gözlemlenmiştir. Modelleme sürecinde, bu gruplanmış yapı stabilite sağlayarak daha öngörülebilir sonuçlara olanak tanıyabilir.

Şekil 2.19 ve Şekil 2.20, Standartlaştırılmış İşçilik Maliyeti ve Standartlaştırılmış Parça Maliyeti değişkenleri arasındaki ilişkiyi farklı perspektiflerden değerlendirmektedir. Şekil 2.19, sürekli değişkenlerle çalışarak pozitif ancak zayıf bir doğrusal ilişkiyi gösterirken, uç değerlerin bu ilişkiyi zayıflattığı ve modelleme sürecinde sapmalara yol açabileceği gözlemlenmektedir. Şekil 2.20 ise değişkenlerin gruplandırılmış haliyle uç değerlerin etkisini minimize ederek daha kararlı bir ilişki sunar, ancak gruplama sürekli değişkenlerin sağladığı bazı detayların kaybolmasına neden olur. Sürekli yapı, regresyon gibi doğrusal modeller için uygunken, gruplama kararlı sonuçlar sunarak kategorik modeller için avantaj sağlar.

3

UYGULAMA

Bu çalışmada, KOBİ MRO şirketinden 2013-2022 yılları arasında toplanan operasyonel veriler kullanılmıştır. Sektör profesyonelleri ile yapılan görüşmeler sonrasında, bazı ek değişkenlerin türetilmesine karar verilmiştir. Türetilen bu yeni değişkenlerin kar marjını açıklayıcı özellikte olacağı beklenmektedir. Modele girecek değişkenlerin seçimi için İnteraktif Gruplama uygulanmıştır. İnteraktif Gruplama ile gruplanmış değişkenler arasından Gini indeksi 15'ten büyük olanlar modelde değerlendirilmeye hak kazanmışlardır. Model geliştirme için kullanılan nihai veri seti, veri temizleme sürecinin ardından 19621 gözlem içermektedir. Sayısal değişkenler İnteraktif Gruplama öncesinde standartlaştırılmış ve analiz edilmişlerdir. Metodoloji, değişken seçimi için İnteraktif Gruplama, modelleme için Gradient Boosting ve Lojistik Regresyon kullanılarak geliştirilmiştir. Model doğrulama analizi, verilerin %70 eğitim ve %30 doğrulama veri setlerine ayrılmasıyla yapılmıştır. Model uyum istatistikleri, standartlaştırılmış değişkenlerle geliştirilen Gradient Boosting modelleri ile gruplanmış değişkenlerle geliştirilen Gradient Boosting modelleri arasında karşılaştırılmıştır. Ayrıca İnteraktif Gruplanmış değişkenler ile Lojistik Regresyon kurulmuş ve Gradient Boosting ile performans kıyaslaması yapılmıştır. Son bir uygulama olarak ise KOBİ MRO firması için kullanılan operasyon verileri müşteriler adına iki segmente, hava yolu firmaları ve tedarikçi firmalar, ayrılarak Lojistik Regresyon uygulanmıştır. Yapılan bu son modellemedeki amaç, farklı segmentteki müşteriler için kar marjı dinamiklerinin nasıl farklılaştığını analiz etmektir.

3.1 İş İhtiyaçlarını ve Veriyi Anlama

Başlangıçta, ham operasyonel verilerde yer alan 83 değişken, veri anlamı, kullanılabilirlik, eksik veri oranı ve güvenilirlik açısından analiz edilmiştir. Veri ve iş ihtiyaçlarının anlaşılması aşamasının ardından, veri temizleme sürecinde 83 değişkenden 30 değişken seçilmiştir. Veri temizleme prosedürleri, 21708 gözlemden oluşan ham veri setine uygulanmıştır. Veri temizliği sonrasında nihai

veri seti 19621 gözleme düşmüştür.

3.2 Veri Hazırlama

Sektör profesyonelleriyle yapılan görüşmeler sonucunda, ham verilerdeki bazı değişkenlerin tek başına kâr marjını açıklamak için yeterli olmayabileceği değerlendirilmiştir. Bu nedenle, bazı türetilmiş değişkenlerin oluşturulmasına karar verilmiştir. Ayrıca, analizlere başlamadan sayısal değişkenler standartlaştırılmıştır.

3.3 Veri Seçimi - İnteraktif Gruplama Uygulaması

Nihai veri setindeki her bir değişken ile kâr marjı arasındaki ilişkiyi kapsamlı bir şekilde analiz etmek için İnteraktif Gruplama kullanılmıştır. Sürekli nitelikteki kâr marjı değişkeni, İnteraktif Gruplama sürecinde ikili bir değişkene dönüştürülmüştür. Şirketin ortalama kâr marjı olan 0,56 göz önünde bulundurularak, hedef değişkeni olan kâr marjı için ikili hedef değişken oluşturulurken, 0,5'ten yüksek kâr marjı kesim noktası olarak kullanılmıştır. 0,5 üzeri 1, 0,5 altı ise 0 atanarak hedef değişken kâr marjı, İnteraktif Gruplama'da diğer değişkenler ile ilişkisi analiz edilebilir bir ikili değişkene dönüştürülmüştür.

Tüm aralıklı girdi değişkenleri bu şekilde ele alınmıştır. Her bir grubun yaklaşık olarak aynı frekansa sahip olacak şekilde sıralama yapılarak gruplar oluşturulmuştur. Başlangıçta grup sayısı dört olarak belirlenmiştir.

İnteraktif Gruplamada ilk gruplar elde edildikten sonra, manuel düzenlemeler ile nihai gruplar oluşturulmuştur. Manuel düzeltme sonrasında elde edilen gruplar kâr marjı ile ilişkiyi daha iyi açıkladığından, bu gruplar Gradient Boosting ve Lojistik Regresyon modellerinde kullanılmıştır.

Bölüm 3.3.1 ve Bölüm 3.3.8 arasında, Gradient Boosting ve Lojistik Regresyon modellerine dahil edilen her bir değişken için İnteraktif Gruplama sonucunda elde edilen gruplar ve Gini indeksleri yer almaktadır. Gini indeksi, her bir değişkenin kâr marjı değişkeniyle olan ilişkisini ölçmek için kullanılmış; Gini indeksi 15'in altında olan değişkenler Gradient Boosting ve Lojistik Regresyon modellerinde dikkate alınmamıştır.

3.3.1 Parça Tanımı - Part Description

Başlangıçta, kategorik değişken olan Parça Tanımı için dört grup oluşturulmuştur. Parça Tanımı değişkeni, MRO şirketinin havayolu ve havacılık şirketleri için bakım

ve onarım yapabileceği 1670 farklı parça numarasını kapsayan uçak parçalarının adlarına ilişkin bilgi sağlar. Bu parçalar için MRO yetenekleri oldukça geniştir. İkili kâr marjı değişkeni (kâr marjı 0,5'ten büyük olduğunda 1, 0,5'e eşit veya küçük olduğunda 0 olarak etiketlenmiştir) ile kategorik değişken olan Parça Tanımı arasındaki ilişkiyi doğrusal olarak monoton azalan bir biçimde ifade edebilmek için, 1670 parça arasında benzer olay oranlarına sahip parçalar gruplanmıştır. Bu gruplama manuel olarak yapılmıştır. Table 3.1, manuel müdahale ile nihai hali verilen grupları göstermektedir.

Tablo 3.1 Parça tanımının gruplara ayrılması

Grup	Olay Sayısı	Olay Dışı Sayısı	Grup Olay Oranı	Gini Katsayısı
1	1563	90	0,946	65,246
2	442	167	0,726	65,246
3	319	182	0,637	65,246
4	621	558	0,527	65,246
5	495	776	0,389	65,246
6	671	1371	0,329	65,246
7	558	1719	0,245	65,246
8	342	3861	0,081	65,246

Table 3.1'de, Grup 1'e dahil olan parçalar için, bakım veya onarımı tamamlanmış ve kâr marjı değeri 0,50'nin üzerinde olan parçaları temsil eden olay sayısı 1563 iken, kâr marjı değeri 0,50 veya altında olan parça sayısı yalnızca 90'dır. Bu durum, olay oranı 0,946 olan Grup 1'deki parçaların en kârlı parçalar olduğunu göstermektedir.

Grup 2'de, kâr marjı değeri 0,50'nin üzerinde olan 442 parça ve 0,50 veya altında olan 167 parça bulunmaktadır; bu da 0,726'lık bir grup olay oranı ile sonuçlanmaktadır.

Grup 3'te, kâr marjı değeri 0,50'nin üzerinde olan 319 parça ve 0,50 veya altında olan 182 parça bulunmaktadır; bu da 0,637'lik bir grup olay oranı ile sonuçlanmaktadır.

Grup 4'te, kâr marjı değeri 0,50'nin üzerinde olan 621 parça ve 0,50 veya altında olan 558 parça bulunmaktadır; bu da 0,527'lik bir grup olay oranı ile sonuçlanmaktadır.

Grup 5'te, kâr marjı değeri 0,50'nin üzerinde olan 495 parça ve 0,50 veya altında olan 776 parça bulunmaktadır; bu da 0,389'luk bir grup olay oranı ile sonuçlanmaktadır.

Grup 6'da, kâr marjı değeri 0,50'nin üzerinde olan 671 parça ve 0,50 veya altında olan 1371 parça bulunmaktadır; bu da 0,329'luk bir grup olay oranı ile

sonuçlanmaktadır.

Grup 7’de, kâr marjı değeri 0,50’nin üzerinde olan 558 parça ve 0,50 veya altında olan 1719 parça bulunmaktadır; bu da 0,245’lik bir grup olay oranı ile sonuçlanmaktadır.

Grup 8’de, kâr marjı değeri 0,50’nin üzerinde olan 342 parça ve 0,50 veya altında olan 3861 parça bulunmaktadır; bu da 0,081’lik bir grup olay oranı ile sonuçlanmaktadır.

İnteraktif Gruplamanın ilk aşamasında, tüm parçalar için ayrı gruplar oluşturulmuş ve bu grupların Gini indeksi 69,169 olarak hesaplanmıştır. Grup sayısını 1670’den 8’e düşürmek amacıyla manuel gruplama yapılmış ve bu Gini indeksinde hafif bir azalmaya yol açarak değeri Table 3.1’de 65,246 olarak kaydedilmiştir. Yüksek bir Gini indeksi, kâr marjı ile Parça Tanımı değişkeni arasında güçlü bir ilişki olduğunu gösterir. Bu manuel gruplama, Gini indeksini biraz düşürmüş olsa da İnteraktif Gruplama yapılan Parça Tanımı değişkeni artık olay oranı ile monotonik bir ilişki gösterebilmekte, yani kâr marjının 0,50’nin üstünde veya altında olup olmamasına göre sınıflanabilmektedir. Bu yaklaşım, hedef değişken olan kâr marjını daha iyi açıklamaktadır. Dolayısıyla, gruplandırılmış parça türleri ile kâr marjı arasında güçlü bir ilişki olduğu söylenebilir.

3.3.2 Parça Maliyeti - Parts Billed Current

Kar marjı ile parça maliyeti arasındaki ilişki, standartlaştırılmış parça maliyetleri kategorize edilerek analiz edilmiştir. Sayısal değişkenin belirli grupları için, kar marjı değeri 0,50 ve altı ile 0,50 üzeri olarak 0 ve 1 şeklinde oluşturulan ikili hedef değişkene göre aralıklar belirlenmiştir.

Tablo 3.2 Parça maliyetinin gruplara ayrılması

Grup	Olay Sayısı	Olay Dışı Sayısı	Grup Olay Oranı	Gini Katsayısı
1	3847	2986	0,563	49,050
2	1001	2491	0,287	49,050
3	163	3247	0,048	49,050
4	0	0	0	49,050

Tablo 3.2, Parça Maliyeti değişkeninin uyarlanmış gruplarını göstermektedir. Bu gruplar şu şekilde tanımlanmıştır: Parça Maliyeti $< -0,36$, $-0,36 \leq$ Parça Maliyeti $< -0,33$, $-0,33 \leq$ Parça Maliyeti $< -0,08$, $-0,08 \leq$ Parça Maliyeti ve eksik değerler.

Bu değişken için Gini indeksi başlangıçta 50,222 olarak hesaplanmıştır. Gini

indeksinin yüksek olmasına rağmen, hedef değişken olarak tanımlanan ikili kar marjı değişkeni ile monotonik bir ilişki gözlenmemiştir. Bu nedenle, başlangıç gruplarında bazı manuel düzenlemeler yapılmıştır.

Tablo 3.2’de, eksik değerlerin sınıflandırıldığı grup, Grup 4 olarak etiketlenmiştir. Bu değişkende eksik değer bulunmadığından, bu gruptaki gözlemler analiz edilmemiştir. Grup 3’teki düşük olay sayısı nedeniyle, Grup 2 ve Grup 3, Grup 1 ile birleştirilmiştir. Birleştirilen Grup 1’de Parça Maliyeti $< -0,33$ olan gözlemler bulunmakta olup, bunların 3847’si olay sayısına (yani kar marjı 0,50’nin üzerinde) ve 2986’sı olay dışı sayıya (yani kar marjı 0,50 veya altında) sahiptir. Grup 1 için grup olay oranı 0,563 olarak hesaplanmıştır. Bu durum, parça maliyetlerinin düşük olduğu Grup 1’deki gözlemlerin iyi kar marjlarına sahip olduğunu göstermektedir.

Parça Maliyeti değerinin -0,33 ile -0,08 arasında olduğu Grup 2’de, 1001 olay sayısı (yani kar marjı 0,50’nin üzerinde) ve 2491 olay dışı sayısı (yani kar marjı 0,50 veya altında) gözlemlenmiştir. Grup 2’nin olay oranı 0,287 olarak hesaplanmıştır. Bu durum, Grup 2’de parça maliyetlerinin Grup 1’e göre nispeten daha yüksek olması nedeniyle kar marjlarının daha düşük olduğunu göstermektedir.

Parça Maliyeti değerinin -0,08’den büyük olduğu Grup 3’te, 163 olay sayısı (yani kar marjı 0,50’nin üzerinde) ve 3247 olay dışı sayısı (yani kar marjı 0,50 veya altında) gözlemlenmiştir. Grup 3’ün olay oranı 0,048 olarak hesaplanmıştır. Bu durum, Grup 3’te parça maliyetlerinin Grup 1 ve Grup 2’ye göre nispeten daha yüksek olması nedeniyle en düşük kar marjlarına sahip olduğunu göstermektedir.

Bu gruplar için Tablo 3.2’de Gini indeksi 49,050 olarak hesaplanmıştır. Başlangıç gruplarına ait Gini indeksine göre biraz daha düşük olsa da, bu Gini indeksi, olay oranı (yani kar marjının 0,50’nin üzerinde veya altında olması) ile etkileşimli gruplandırılmış Parça Maliyeti değişkeni arasında monotonik bir ilişki olabileceğini göstermektedir. Bu yaklaşım, hedef değişkeni daha iyi açıklamakta olup, malzeme maliyetlerindeki artışın kar marjlarını düşürdüğünü ortaya koymaktadır.

3.3.3 İşçilik Maliyeti - Labour Billed Current

Başlangıçta, İşgücü Maliyeti değişkeni şu şekilde beş gruba ayrılmıştır: İşgücü Maliyeti $< -0,82$, $-0,82 \leq$ İşgücü Maliyeti $< -0,29$, $-0,29 \leq$ İşgücü Maliyeti $< 0,34$, $0,34 \leq$ İşgücü Maliyeti ve eksik değerler.

Tablo 3.3 İşgücü maliyetinin gruplara ayrılması

Grup	Olay Sayısı	Olay Dışı Sayısı	Grup Olay Oranı	Gini Katsayısı
1	3757	3133	0,545	42,664
2	858	2570	0,250	42,664
3	396	3021	0,116	42,664
4	0	0	0	42,664

Tablo 3.3, doğrudan işçilik maliyetleri için uyarlanmış grupları göstermektedir. Bu değişken için Gini indeksi 42,664 olarak hesaplanmıştır. Gini indeksi yüksek olmasına rağmen, hedef değişken olarak tanımlanan ikili kar marjı değişkeni ile monotonik bir ilişki gözlemlenmemiştir. Bu nedenle, ilgili değişkene manuel gruplama yapılmıştır. Aşağıda, başlangıç gruplarına yapılan manuel düzenlemeler açıklanmaktadır.

Tablo 3.3’de, eksik değerlerin sınıflandırıldığı grup, Grup 4 olarak belirlenmiştir. Bu değişkende eksik değer bulunmadığından, bu gruptaki gözlemler analiz edilmemiştir. Grup 2’deki düşük olay sayısı nedeniyle, Grup 2 ve Grup 3, Grup 1 altında birleştirilmiştir. Birleştirilen Grup 1 kapsamında, İşgücü Maliyeti $< -0,29$ olan gözlemler bulunmakta olup, bunların 3757’si olay sayısına (yani kar marjı 0,50’nin üzerinde) ve 3133’ü olay dışı sayıya (yani kar marjı 0,50 veya altında) sahiptir. Grup 1 için grup olay oranı 0,545 olarak hesaplanmıştır. Bu durum, işçilik maliyetlerinin düşük olduğu Grup 1’deki gözlemlerin en iyi kar marjına sahip olduğunu göstermektedir.

İşgücü Maliyeti değeri -0,29 ile 0,34 arasında olan gözlemlerden oluşan Grup 2’de, 858 olay sayısı (yani kar marjı 0,50’nin üzerinde) ve 2570 olay dışı sayısı (yani kar marjı 0,50 veya altında) gözlemlenmiştir. Grup 2’nin olay oranı 0,250 olarak hesaplanmıştır. Bu durum, Grup 2’de işçilik maliyetlerinin Grup 1’e göre nispeten daha yüksek olması nedeniyle kar marjlarının daha düşük olduğunu göstermektedir.

İşçilik Maliyeti değeri 0,34’ten büyük olan gözlemlerden oluşan Grup 3’te, 396 olay sayısı (yani kar marjı 0,50’nin üzerinde) ve 3021 olay dışı sayısı (yani kar marjı 0,50 veya altında) gözlemlenmiştir. Grup 3’ün olay oranı 0,116 olup, işçilik maliyetlerinin Grup 1 ve Grup 2’ye göre nispeten daha yüksek olduğu bu grupta, en düşük kar marjlarının gözlemlendiğini göstermektedir.

Tablo 3.3’deki etkileşimli gruplandırılmış İşgücü Maliyeti değişkeni için Gini indeksi 42,664 olarak kalmıştır; bu, olay oranı (yani kar marjının 0,50’nin üzerinde veya altında olması) ile monotonik bir ilişki olabileceğini göstermektedir. Bu durum, işçilik maliyetlerindeki artışın kar marjını azalttığını daha etkili bir şekilde

açıklamak için bir fırsat sunmaktadır.

3.3.4 İlgili Yönetici - Manager

Kar Marjı ikili değişkeni (kar marjı 0,50'nin üzerindeyse 1, 0,50 ve altındaysa 0 olarak etiketlenmiştir) ile İlgili Yönetici kategorik değişkeni arasındaki ilişkiyi lineer olarak azalan bir monotonik yapıda ifade etmek amacıyla, 58 yönetici arasında benzer olay oranlarına sahip değişkenler gruplandırılmıştır. Bu gruplandırma manuel olarak yapılmıştır. Yönetici değişkeni eksik olan gözlemler ayrı bir grupta toplanmıştır.

Tablo 3.4 Yöneticilerin gruplara ayrılması

Grup	Olay Sayısı	Olay Dışı Sayısı	Grup Olay Oranı	Gini Katsayısı
1	750	195	0,794	33,178
2	1530	1944	0,440	33,178
3	804	1233	0,395	33,178
4	1325	2669	0,332	33,178
5	415	1356	0,234	33,178
6	187	1327	0,124	33,178

Tablo 3.4'de, Grup 1'e dahil edilen yöneticiler için olay sayısı 0,50'nin üzerinde bir kar marjı değerine sahip olup, bu sayı 750 iken, 0,50 ve altındaki kar marjı değerine sahip yönetici sayısı yalnızca 195'tir. Bu durum, 0,794'lük bir grup olay oranına sahip Grup 1'deki yöneticilerin, en yüksek kar marjına sahip operasyonlara yönlendirildiğini göstermektedir.

Grup 2'de, kar marjı değeri 0,50'nin üzerinde olan 1530 ve 0,50'nin altında olan 1944 yönetici bulunmaktadır. Grup 2'nin olay oranı 0,44'tür.

Grup 3'te, Yönetici değişkeninde eksik değer olan gözlemler sınıflandırılmıştır. Kar marjı değeri 0,50'nin üzerinde olan 804 ve 0,50 ve altında olan 1233 yönetici bulunmaktadır. Grup 3'ün olay oranı 0,395'tir.

Grup 4'te, kar marjı değeri 0,50'nin üzerinde olan 1325 ve 0,50 ve altında olan 2669 yönetici bulunmaktadır. Grup 4'ün olay oranı 0,332'dir.

Grup 5'te, kar marjı değeri 0,50'nin üzerinde olan 415 ve 0,50 ve altında olan 1356 yönetici bulunmaktadır. Grup 5'in olay oranı 0,234'tür.

Grup 6'da, kar marjı değeri 0,50'nin üzerinde olan 187 ve 0,50 ve altında olan 1327 yönetici bulunmaktadır. Grup 6'nın olay oranı 0,124'tür.

Bu gruplar dikkate alınarak hesaplanan Gini indeksi, Tablo 3.4'de 33,178 gibi

oldukça yüksek bir değerdir. Yüksek bir Gini indeksi, kar marjı ile İlgili Yönetici değişkeni arasında güçlü bir ilişki olduğunu göstermektedir. Ayrıca, İnteraktif Gruplama'nın başlangıç aşamasında tüm yöneticiler için ayrı gruplar oluşturulmuş ve bu grupların Gini indeksi 34,677 olarak hesaplanmıştır. 59 gruptan 6 gruba geçiş için manuel gruplama yapılmış ve bu manuel gruplama Gini indeksini 33,178'e hafifçe düşürmüştür. Biraz daha düşük Gini indeksi elde edilmiş olsa da, 0,50'nin üzerinde kar marjı olup olmadığına göre olay oranı ile monotonik olarak ilişkilendirilebilen gruplanmış İlgili Yönetici değişkeni, kar marjı değişkeninin daha kapsamlı bir şekilde anlaşılmasına olanak sağlayacaktır. Gini indeksi yorumlanarak İlgili Yönetici değişkeninin kar marjı ile ilişkili olduğu söylenebilir.

3.3.5 Çevrim Süresi - Diff Shipping Entry

Tablo 3.5 Çevrim süresinin gruplara ayrılması

Grup	Olay Sayısı	Olay Dışı Sayısı	Grup Olay Oranı	Gini Katsayısı
1	1842	1396	0,569	30,367
2	1418	2097	0,403	30,367
3	953	2587	0,269	30,367
4	798	2644	0,232	30,367
5	0	0	0	30,367

Tablo 3.5, Çevrim Süresi değişkeni için manuel düzeltme sonrası nihai grupları göstermektedir. Başlangıçta, Çevrim süresi değişkeni şu şekilde beş gruba ayrılmıştır:

$\text{Çevrim Süresi} < -0,39$, $-0,39 \leq \text{Çevrim Süresi} < -0,25$, $-0,25 \leq \text{Çevrim Süresi} < -0,01$, $-0,01 \leq \text{Çevrim Süresi}$ ve eksik değerler.

Bu değişken için Gini indeksi 30,367 olarak hesaplanmıştır. Gini indeksi yüksek olmasına rağmen, hedef değişken olarak tanımlanan ikili kar marjı değişkeni ile monoton bir ilişki bulunmamıştır. Bu nedenle, başlangıç gruplarında bazı manuel düzenlemeler yapılmıştır.

Tablo 3.5'de, eksik değerlerin sınıflandırıldığı grup, Grup 5 olarak belirlenmiştir. Bu değişkende eksik değer bulunmadığından, bu gruptaki gözlemler analiz edilmemiştir. Grup 1'de, Çevrim süresi değeri -0,39'dan küçük olan gözlemler bulunmakta olup, bunların 1842'si olay sayısına (yani kar marjı 0,50'nin üzerinde) ve 1396'sı olay dışı sayıya (yani kar marjı 0,50 veya altında) sahiptir. Grup 1 için grup olay oranı 0,569 olarak hesaplanmıştır. Bu durum, Çevrim süresi değerlerinin

en düşük olduđu Grup 1'deki gözlemlerin en iyi kar marjlarına sahip olduğunu göstermektedir.

Çevrim süresi değeri -0,39 ile -0,25 arasında olan gözlemlerden oluşan Grup 2'de, 1418 olay sayısı (yani kar marjı 0,50'nin üzerinde) ve 2097 olay dışı sayısı (yani kar marjı 0,50 veya altında) gözlemlenmiştir. Grup 2'nin olay oranı 0,403 olarak hesaplanmıştır. Bu durum, Grup 2'de Çevrim süresi değerlerinin Grup 1'e göre nispeten daha yüksek olması nedeniyle kar marjlarının daha düşük olduğunu göstermektedir.

Çevrim süresi değeri -0,25 ile -0,01 arasında olan gözlemlerden oluşan Grup 3'te, 953 olay sayısı (yani kar marjı 0,50'nin üzerinde) ve 2587 olay dışı sayısı (yani kar marjı 0,50 veya altında) gözlemlenmiştir. Grup 3'ün olay oranı 0,269 olarak hesaplanmıştır. Bu durum, Grup 3'te çevrim süresi değerlerinin Grup 1 ve Grup 2'ye göre nispeten daha yüksek olması nedeniyle kar marjlarının daha düşük olduğunu göstermektedir.

Çevrim süresi değeri -0,01'den büyük olan gözlemlerden oluşan Grup 4'te, 798 olay sayısı (yani kar marjı 0,50'nin üzerinde) ve 2644 olay dışı sayısı (yani kar marjı 0,50 veya altında) gözlemlenmiştir. Grup 4'ün olay oranı 0,232 olup, Çevrim süresi değerlerinin Grup 1, Grup 2 ve Grup 3'e göre nispeten daha yüksek olduğu bu grupta, en düşük kar marjlarının gözlemlendiğini göstermektedir.

Manuel gruplama düzenlemesinin ardından, Gini indeksi başlangıçtaki İnteraktif Gruplama ile aynı kalarak Tablo 3.5'de 30,367 olarak hesaplanmıştır. 0,50'nin üzerinde veya altında olup olmadığına göre olay oranı ile monotonik olarak ilişkilendirilebilen etkileşimli gruplandırılmış Çevrim Süresi değişkeni, kar marjını daha etkili bir şekilde açıklamak için bir fırsat sunmaktadır. Çevrim süresi arttıkça, kar marjı azalmaktadır. Daha uzun TAT, kârlılığın daha fazla düşmesine yol açmakta olup, bu veri gruplaması TAT'ın kar marjı üzerindeki olumsuz etkisiyle uyumludur.

3.3.6 İlgili Bölüm - Department

Başlangıçta, İlgili Bölüm kategorik değişkeni için dört grup oluşturulmuştur. Kar Marjı ikili değişkeni (kar marjı 0,50'nin üzerinde ise 1, 0,50 ve altındaysa 0 olarak etiketlenmiştir) ile İlgili Bölüm değişkeni arasındaki ilişkiyi lineer olarak azalan bir monotonik yapıda ifade etmek amacıyla, 11 kategori (eksik kategori dahil) arasındaki benzer olay oranlarına sahip değişkenler gruplandırılmıştır. Bu gruplandırma manuel olarak yapılmıştır. Tablo 3.6 uyarlanmış grupları göstermektedir.

Tablo 3.6 İlgili bölümlerin gruplara ayrılması

Grup	Olay Sayısı	Olay Dışı Sayısı	Grup Olay Oranı	Gini Katsayısı
1	2452	2594	0,486	23,390
2	1171	2089	0,359	23,390
3	1388	4041	0,256	23,390

Tablo 3.6’de, Grup 1’e dahil edilen ilgili bölümler için olay sayısı (kar marjı 0,50’nin üzerinde) 2452 iken, olay dışı sayısı (kar marjı 0,50 veya altında) 2594’tür. Bu durum, 0,486’lık grup olay oranına sahip olan Grup 1’deki bölümlerin, diğer gruplar arasında en kârlı bölümler olduğunu göstermektedir.

Grup 2’de, kar marjı 0,50’nin üzerinde olan 1171 olay sayısı ve 0,50 veya altında olan 2089 olay dışı sayısı gözlemlenmiştir. Grup 2’nin olay oranı 0,359 olup, Grup 2’nin Grup 1’e göre daha düşük kar marjları elde eden bölümleri içerdiğini göstermektedir.

Grup 3’te, kar marjı 0,50’nin üzerinde olan 1388 olay sayısı ve 0,50 veya altında olan 4041 olay dışı sayısı gözlemlenmiştir. Grup 3’ün olay oranı 0,256’dır. Bu durum, Grup 3’ün, Grup 1 ve Grup 2’ye göre en düşük kar marjlarına sahip bölümleri içerdiğini göstermektedir.

İnteraktif gruplamanın başlangıç aşamasında tüm ilgili bölümler için ayrı gruplar oluşturulmuş ve bu grupların Gini indeksi 25,080 olarak hesaplanmıştır. 11 gruptan 3 gruba geçiş yapmak için manuel gruplama yapılmış ve Gini indeksi 23,39’a hafifçe düşmüştür (Tablo 3.6). Gini indeksindeki küçük bir düşüğe rağmen, olay oranı ile monotonik olarak ilişkilendirilebilen (yani kar marjının 0,50’nin üzerinde veya altında olması) etkileşimli gruplanmış İlgili Bölüm değişkeni, kar marjını daha kapsamlı bir şekilde açıklamaktadır. Bu Gini indeksini yorumlayarak İlgili Bölüm değişkeninin kar marjı ile ilişkili olduğu sonucuna varılabilir. Bazı bölüm gruplarının diğerlerine kıyasla daha yüksek kar marjlarına sahip olduğu söylenebilir.

3.3.7 Fiyat Teklif Süresi - Diff Quote Entry

Tablo 3.7 Fiyat teklif sürelerinin gruplara ayrılması

Grup	Olay Sayısı	Olay Dışı Sayısı	Grup Olay Oranı	Gini Katsayısı
1	1533	1301	0,541	23,193
2	1769	2964	0,374	23,193
3	879	2023	0,303	23,193
4	830	2436	0,254	23,193
5	0	0	0,000	23,193

Tablo 3.7 Fiyat Teklif Süresi değişkeni için uyarlanmış grupları göstermektedir. Başlangıçta, Fiyat Teklif Süresi değişkeni şu şekilde beş gruba ayrılmıştır:

Fiyat Teklif Süresi $< -0,25$, $-0,25 \leq$ Fiyat Teklif Süresi $< -0,18$, $-0,18 \leq$ Fiyat Teklif Süresi $< -0,07$, $-0,07 \leq$ Fiyat Teklif Süresi ve eksik değerler.

Bu değişken için Gini indeksi 23,193 olarak hesaplanmıştır. Gini indeksi yüksek olup, hedef değişken olarak tanımlanan ikili kar marjı değişkeni ile monotonik bir azalma gözlenmiştir. Başlangıç gruplarında, eksik grup için bazı manuel düzenlemeler yapılmıştır.

Tablo 3.7’de, eksik değerlerin sınıflandırıldığı grup, Grup 5 olarak belirlenmiştir. Bu değişkende eksik değer bulunmadığından, bu gruptaki gözlemler analiz edilmemiştir. Grup 1 kapsamında, Fiyat Teklif Süresi değeri -0,25’ten küçük olan gözlemler bulunmakta olup, bunların 1533’ü olay sayısına (yani kar marjı 0,50’nin üzerinde) ve 1301’i olay dışı sayıya (yani kar marjı 0,50 veya altında) sahiptir. Grup 1’in olay oranı 0,541 olarak hesaplanmıştır. Bu durum, Fiyat Teklif Süresi düşük olan Grup 1’deki gözlemlerin iyi kar marjlarına sahip olduğunu göstermektedir.

Fiyat teklif süresi değeri -0,25 ile -0,18 arasında olan gözlemlerden oluşan Grup 2’de, 1769 olay sayısı (yani kar marjı 0,50’nin üzerinde) ve 2964 olay dışı sayısı (yani kar marjı 0,50 veya altında) gözlemlenmiştir. Grup 2’nin olay oranı 0,374 olup, fiyat teklif süresi değerinin Grup 1’e göre nispeten daha yüksek olduğu bu grupta kar marjlarının daha düşük olduğu sonucuna varılabilir.

Fiyat teklif süresi değeri -0,18 ile -0,07 arasında olan gözlemlerden oluşan Grup 3’te, 879 olay sayısı (yani kar marjı 0,50’nin üzerinde) ve 2023 olay dışı sayısı (yani kar marjı 0,50 veya altında) gözlemlenmiştir. Grup 3’ün olay oranı 0,303’tür. Bu durum, Grup 3’te fiyat teklif süresi değerinin Grup 1 ve Grup 2’ye göre nispeten daha yüksek olması nedeniyle kar marjlarının daha düşük olduğunu göstermektedir.

Fiyat Teklif Süresi değeri -0,07’den büyük olan gözlemlerden oluşan Grup 4’te, 830

olay sayısı (yani kar marjı 0,50'nin üzerinde) ve 2436 olay dışı sayısı (yani kar marjı 0,50 veya altında) gözlemlenmiştir. Grup 4'ün olay oranı 0,254 olup, Fiyat Teklif Süresi değerinin Grup 1, Grup 2 ve Grup 3'e göre en yüksek olduğu bu grupta kar marjlarının en düşük olduğu sonucuna varılabilir.

Tablo 3.7'de Gini indeksi, 23,193 olarak değişmeden kalmıştır. Fiyat Teklif Süresi değişkeninin İnteraktif Gruplaması, başlangıçta olay oranı ile monotonik bir ilişki göstermekte olup (yani kar marjı 0,50'nin üzerinde veya altında), bu durum hedef değişken olan kar marjını daha net açıklama imkanı sunmaktadır. Gruplamalarda bir değişiklik yapılmamıştır. Fiyat teklif süresi arttıkça kar marjı azalmakta olup, teklif süresinin kârlılık üzerindeki olumsuz etkisini ortaya koymaktadır.

3.3.8 Parça Durumu - Status

Tablo 3.8, Parça Durumu değişkeni için uyarlanmış grupları göstermektedir.

Tablo 3.8 Parça durumunun gruplara ayrılması

Grup	Olay Sayısı	Olay Dışı Sayısı	Grup Olay Oranı	Gini Katsayısı
1	1345	1467	0,478	19,876
2	2999	4762	0,386	19,876
3	667	2495	0,211	19,876
4	0	0	0,000	19,876

Tablo 3.8'de, Grup 1, sevk edilmiş durumda olan parçaları içermektedir. Kar marjı 0,50'nin üzerinde olan olay sayısı 1345, kar marjı 0,50 veya altında olan olay dışı sayısı ise 1467'dir. 0,478'lik bir grup olay oranına sahip olan ve sevk edilen durumdaki parçaları içeren Grup 1'in, tüm durum grupları arasında en kârlı grup olduğu görülmektedir.

Grup 2'de, iş tamamlanmış durumdaki parçalar yer almakta olup, kar marjı 0,50'nin üzerinde olan 2999 olay sayısı ve 0,50 veya altında olan 4762 olay dışı sayısı gözlemlenmiştir. Grup 2'nin olay oranı 0,386'dır.

Grup 3'te, iş tamamlanmış ve sevk edilmiş parçaların geri kalanları bulunmaktadır. Bu grupta kar marjı 0,50'nin üzerinde olan 667 olay sayısı ve 0,50 veya altında olan 2495 olay dışı sayısı yer almaktadır. Grup 3'ün olay oranı 0,211'dir.

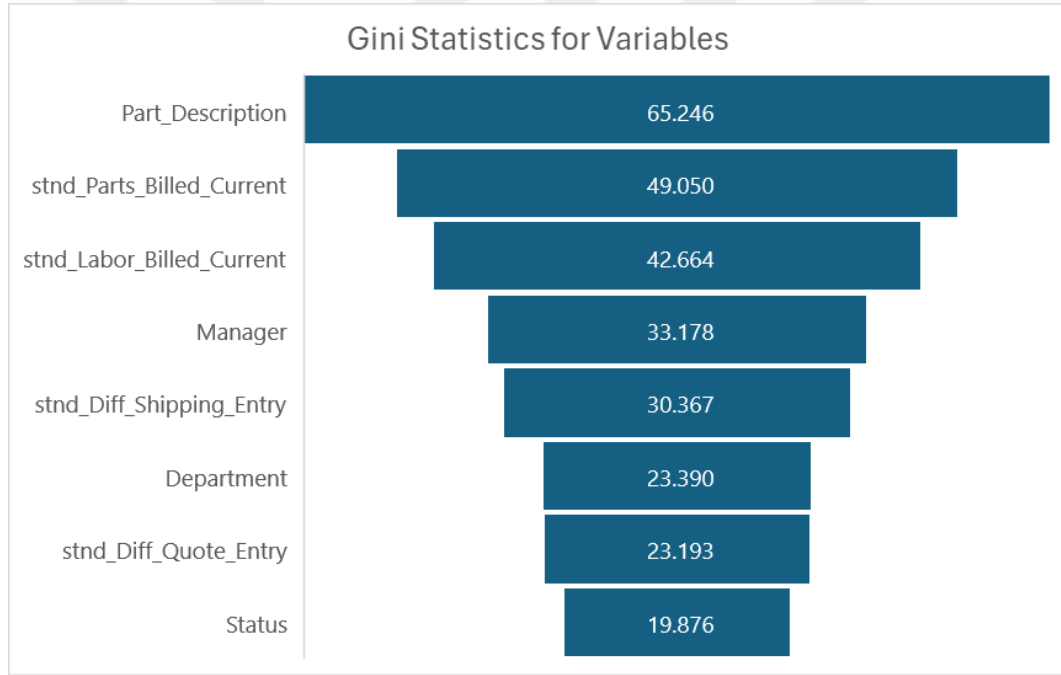
Bu gruplar dikkate alınarak hesaplanan Gini indeksi 19,876'dır. Kabul edilebilir bir Gini indeksi, Kar Marjı ve Parça Durumu değişkenleri arasında bir ilişki olduğunu göstermektedir. Ayrıca, İnteraktif Gruplamanın başlangıç aşamasında tüm durumlar için ayrı gruplar oluşturulmuş ve bu grupların Gini indeksi 20,865 olarak hesaplanmıştır. 11 gruptan 4 gruba geçiş yapmak için manuel gruplama

yapılmış ve bu manuel gruplama Tablo 3.8’de Gini indeksini 19,876’ya hafifçe düşürmüştür.

Gini indeksi biraz daha düşük olsa da, olay oranı ile monoton olarak artan bir ilişki gösteren etkileşimli Parça Durumu değişkeni, kar marjını daha iyi açıklama fırsatı sunmaktadır. Bu Gini indeksini yorumlamak, Parça Durumu değişkeni ile kar marjı arasında belirgin bir ilişki olduğunu göstermektedir. Sonuç olarak, müşteri siparişlerinin tamamlama seviyesi arttıkça kar marjının da artma eğiliminde olduğu söylenebilir.

3.3.9 İnteraktif Gruplama Sonuçlarının Özeti

Her bir değişken için açıklanan İnteraktif Gruplama sonuçları, tüm değişkenler için Şekil 3.1’de toplu olarak sunulmuştur.

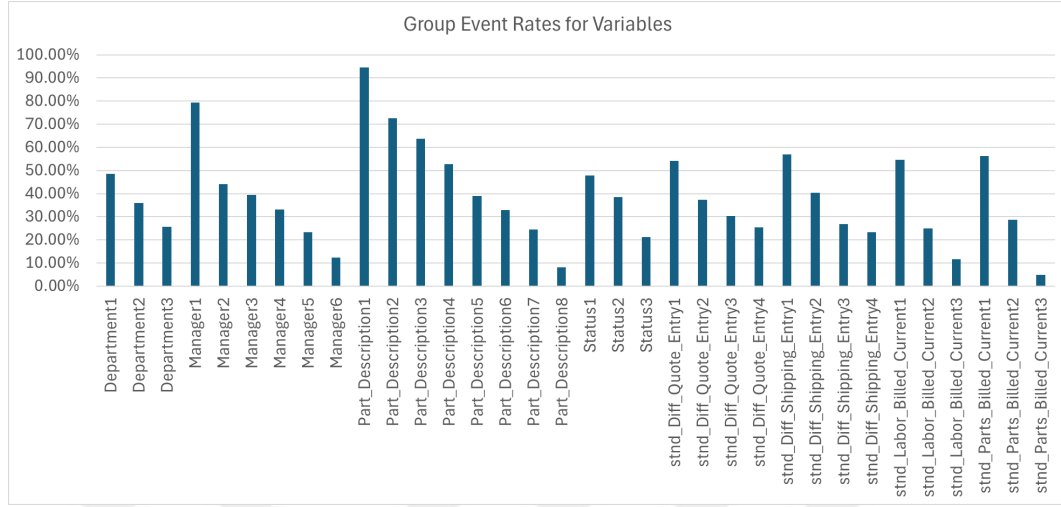


Şekil 3.1 Gruplanmış değişkenler için Gini indeksi

Şekil 3.1, modele dahil edilen gruplanmış değişkenlerin hedef değişken ile ilişkisini açıklayan Gini indekslerini göstermektedir. Şekil 3.1’den görüldüğü üzere, kar marjını en iyi açıklayan değişken Parça Tanımı değişkeni iken, en az açıklayan değişken Parça Durumu değişkeni olmuştur. İnteraktif Gruplama öncesinde, standartlaştırılmış girdi değişkenleri ile hedef değişken arasında ciddi bir ilişki gözlemlenmemiştir.

Değişken gruplarında yapılan manuel düzenlemeler sonrasında, değişken

grup numarası arttıkça kar marjının azaldığını gösteren monoton bir eğilim gözlemlenmiştir. Girdi ve hedef değişken arasındaki hedeflenen monoton eğilim Şekil 3.2’de gösterilmektedir.

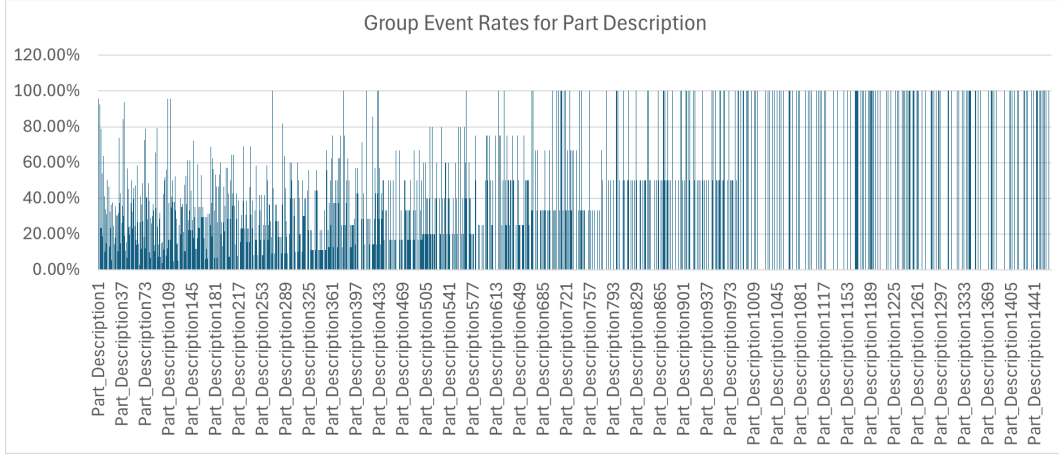


Şekil 3.2 Değişkenler için farklı gruplar arasında olay oranlarının karşılaştırılması

Şekil 3.2’de, x eksenindeki değişkenlerin sonunda yer alan sayılar grupları göstermektedir. Grup numaraları 1’den n’e doğru ilerledikçe, 0,5’in üzerindeki kar marjı oranının azaldığı söylenebilir.

Değişken gruplarında yapılan manuel düzenlemeler sonrasında, değişken grup sayısı arttıkça kar marjının azaldığını gösteren monoton bir eğilim gözlemlenmiştir. Girdi ve hedef değişken arasındaki hedeflenen monoton eğilim Şekil 3.2’de gösterilmektedir.

Parça Tanımı değişkeninin manuel grupta ile hedef değişkeni açıklamada ne kadar başarılı olduğunu göstermek amacıyla, manuel düzenleme öncesinde Parça Tanımı değişkenindeki 1670 parçanın olay oranı ile ilişkisi Şekil 3.3’te gösterilmiştir.



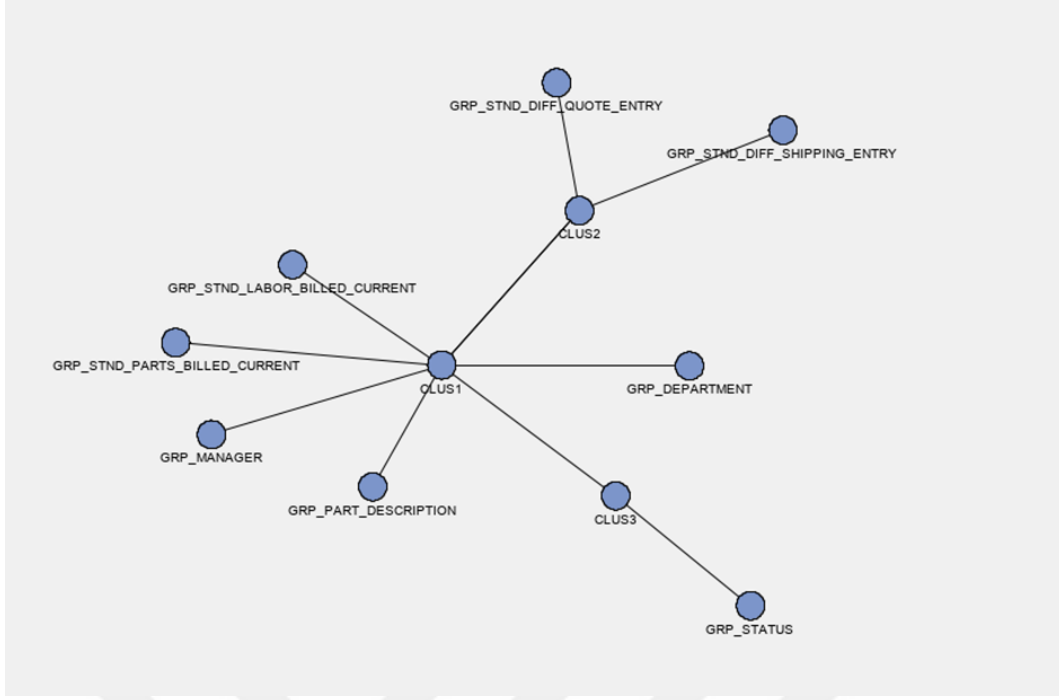
Şekil 3.3 Parça tanımı değişkeni için grup olay oranlarının karşılaştırılması

Şekil 3.3'te, manuel gruplama öncesinde 1670 parça ile kar marjı arasında bir ilişki kurmanın zorluğu görülmektedir.

Modele girmeye hak kazanan ve İnteraktif Gruplama sonrası analizlerine yer verilen değişkenler dışında değinilmesi gereken bir değişken daha vardır. Veri hazırlama aşamasında, pandemi öncesi ve sonrası yapılan tamirleri ayırt edebilmek amacıyla Entry Date Num, Pandemi Öncesi ve Sonrası, adında bir kukla değişken oluşturulmuştur. Bu değişkenin kar marjı ile ilişkisi sınırlı olduğundan, değişken seçiminde eşik değeri olarak belirlenen Gini indeksi 15 değerini aşamamıştır, modelde kullanılmamıştır. Pandemi öncesi ve sonrasında yapılan tamirlerin kar marjı analizi açısından farklılaşmadığı söylenebilir.

3.4 Değişken Kümeleme Uygulaması

Modeller, değişkenlere İnteraktif Gruplama dönüşümünün uygulanmasının katkısını göstermek amacıyla başlangıçta standartlaştırılmış girdi değişkenleri ve sürekli hedef değişken ile çalıştırılmıştır. Daha sonra, yeni gruplandırılmış değişkenler ve ikili hedef değişken ile çalıştırılmıştır. Modellemeleri çalıştırmadan önce, modele girmeye aday olarak kabul edilen değişkenlerin kendi aralarındaki mesafeleri ölçmek için Variable Clustering (Değişken Kümeleme) uygulanmıştır. Sonuçlar Şekil 3.4'te sunulmaktadır.

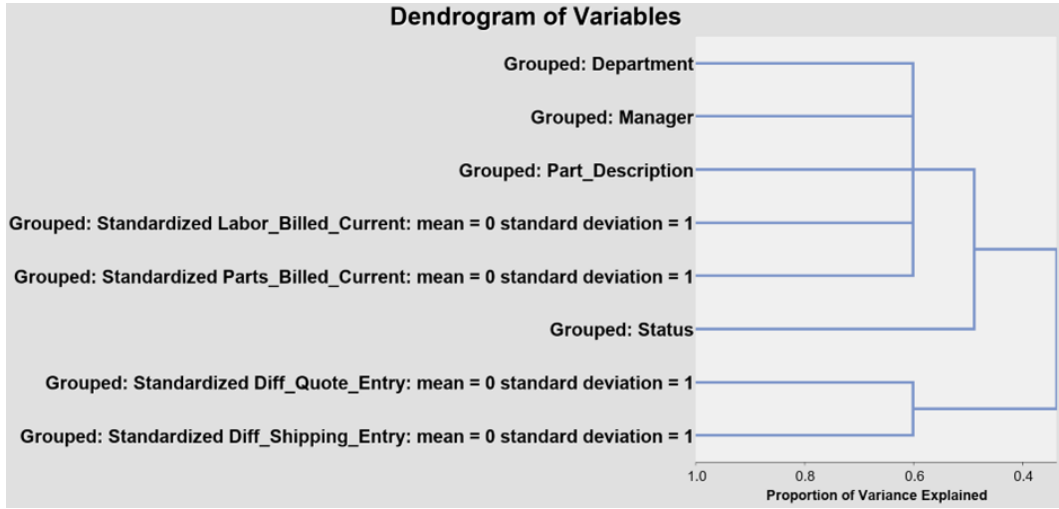


Şekil 3.4 Değişken Kümeleme diyagramı

Şekil 3.4, değişkenler ve değişkenlerin küme merkezlerini temsil etmektedir. Gruplandırılan değişkenlerin isimleri belirgin benzerlik ilişkilerini göstermektedir. Benzer niteliksel özelliklere sahip değişkenler bir araya kümelendiğinden, Şekil 3.4 görsel olarak kolaylıkla yorumlanabilir ve analiz edilebilir[21]. Değişken kümeleme, bir grubun içindeki değişkenlerin birbirine benzer, farklı gruptaki değişkenlere ise olabildiğince farklı olduğu değişken grupları sağlar. Bu çalışmada, değişken kümeleme yöntemi bir değişken azaltma tekniği olarak kullanılmamıştır. Bunun yerine, değişkenleri kümelerle organize ederek veriyi daha anlaşılır hale getirmek için kullanılmıştır. Bu kümeler içindeki değişkenler, Bölüm 4’te paylaşılan stratejik kararlar hakkında içgörüler sağlamak amacıyla kullanılmıştır.

Diyagram incelendiğinde, Diff Quote Entry, Fiyat Teklif Süresi ve Diff Shipping Entry, Çevrim Süresi, değişkenlerinin ile birlikte Clus2’de yer aldığı görülmektedir. Bu küme, operasyonel süreci temsil eden TAT ve TTQ kavramlarını açıklayan bir kümedir[28]. Teklif verme süresi ve TAT, Clus2’de benzer bilgileri tanımlamaktadır. TAT ve TTQ’nun diğer değişkenlerden farklı olduğu sonucuna varılabilir. Parts Billed Current, Manager, Part Description, Labour Billed Current ve Department değişkenleri sırasıyla Parça Maliyeti, İlgili Yönetici, Parça Tanımı, İşçilik Maliyeti ve İlgili Bölüm olarak Clus1’de gruplanmıştır. Clus1’in parça ile ilgili bilgileri içeren bir küme olduğu söylenebilir. Status, Parça Durumu değişkeni ise Clus3’te tek başına yer almaktadır. Parçanın durumu hakkındaki bilgi, ayrı bir küme içinde sunulmuştur ve diğer değişkenlerden daha bağımsız hareket ettiği

söylenebilir. Değişkenlerin dendrogramı Şekil 3.5'te gösterilmektedir.



Şekil 3.5 Değişkenlerin dendrogramı

Şekil 3.5, gruplanmış noktaları ve grup üyelerinin benzerliğini görselleştirerek Parça Teklif Süresi, TAT ve Çevrim Süresi, TTQ, değişkenlerinin diğer değişkenlerden ayrıldığını göstermektedir.

3.5 Gradient Boosting Uygulaması

Bu çalışmada, veri temizleme sürecinden model doğrulama aşamasına kadar izlenen adımlar, KOBİ düzeyindeki MRO şirketlerinin kâr marjını artırmaya yönelik kararlarını belirlemede analitik yöntemlerin nasıl etkili bir şekilde kullanılabileceğini ortaya koymaktadır. Gruplandırılmış değişkenler ve standartlaştırılmış değişkenler için ise ayrı olmak üzere iki farklı Gradient Boosting algoritması uygulanmış ve sonuçları karşılaştırılmıştır. Gradient Boosting modeli, sınıflandırma problemlerinde yüksek performans göstermesi nedeniyle tercih edilmiştir[22, 29].

Bu araştırma sırasında, hiperparametre ayarları manuel olarak, deneme-yanılma yöntemiyle yapılmıştır[30]. Öncelikle, anahtar hiperparametreler için bir değer aralığı belirlenmiştir. Bu hiperparametrelerin her kombinasyonu, modelin eğitim verisi üzerinde eğitilmesi ve doğrulama verisindeki performansının değerlendirilmesiyle ardışık olarak test edilmiştir. Performans sonuçlarına bağlı olarak hiperparametrelerde ayarlamalar yapılmıştır. Bu yinelemeli süreç, eğitim ve doğrulama performans metrikleri arasında en iyi dengeyi sağlayan optimal hiperparametre kümesi belirlenene kadar devam etmiştir. Boost79 ve Boost86 modelleri için ayarlanan hiperparametreler sırasıyla Bölüm 3.5.1 ve Bölüm 3.5.2'de ele alınmıştır. Hiperparametre seçiminde alternatif olarak

optimizasyon yöntemlerinin kullanılması da tercih edilebilirdi.

Ayrıca, veri seti 2013-2021 yıllarını kapsayan belirli bir KOBİ düzeyindeki MRO şirketiyle sınırlıdır. Bu zamansal ve bağlamsal sınırlılık, elde edilen sonuçların farklı bağlamlara veya zaman dilimlerine genellenebilirliğini sınırlayabilir.

Modellenme aşaması kapsamında, iki farklı Gradient Boosting algoritması çalıştırılmış ve sonuçları bu bölümde karşılaştırılmıştır. Standartlaştırılmış değişkenlerle oluşturulan Boost79 Modeli, doğrulama veri setinde Tablo 3.13’de görülen 0,637 RASE değeri ile oldukça yüksek bir hata oranına sahiptir. Gruplandırılmış değişkenlerle geliştirilen Boost86 Modeli’nde, Tablo 3.15’te yer aldığı gibi 0,164 yanlış sınıflandırma oranı ve 0,821 Gini değeri, bu veri dönüşümü ile (İnteraktif Gruplama) daha düşük hata oranına sahip bir sınıflandırma modeli kurulabileceğini göstermektedir. Değişken önem değerlerinin modellerdeki eğitim ve doğrulama veri setleri ile uyumluluğunu analiz etmek için bir saçılım grafiği çizilmiştir. Boost79 Modeli için çizilen saçılım grafiği Şekil 3.6, eğitim ve doğrulama veri setlerinde Boost86 Modeli için çizilen Şekil 3.10 saçılım grafiğine göre daha az uyum göstermektedir. Sürekli ve kategorik verilerin anlamlı gruplara dönüştürülmesinin Boost86 Modeli’nde, Bölüm 3.5.2, Bölüm 3.5.1’e göre kar marjını daha iyi açıkladığı belirlenmiştir. İnteraktif Gruplama kullanılarak daha yüksek doğruluğa sahip bir model elde edildiği söylenebilir.

Daha iyi tahmin performansı gösteren Boost86 modelinin karmaşıklık matrisleri (Confusion Matrix), sırasıyla eğitim ve doğrulama veri setleri için Tablo 3.9 ve Tablo 3.10’de sunulmaktadır.

Tablo 3.9 Boost86 - eğitim veri seti için karmaşıklık matrisi

	Gerçek Pozitif (1)	Gerçek Negatif (0)	Toplam
Tahmin Edilen Pozitif (1)	3825	834	4659
Tahmin Edilen Negatif (0)	1202	7872	9074
Toplam	5027	8706	13733

Tablo 3.9, modelin eğitim veri setindeki performansına ait karmaşıklık matrisini göstermekte olup, gerçek pozitif ve negatif vakaların tahmin edilen pozitif vakalar ve negatif vakalar ile karşılaştırılmasını sunmaktadır. Matristeki özet şu şekildedir:

Gerçek Pozitif (1): 3825 örnek doğru bir şekilde pozitif olarak tahmin edilmiştir.
Gerçek Negatif (0): 834 örnek yanlış bir şekilde pozitif olarak tahmin edilmiştir.
Gerçek Pozitif (1): 1202 örnek yanlış bir şekilde negatif olarak tahmin edilmiştir.
Gerçek Negatif (0): 7872 örnek doğru bir şekilde negatif olarak tahmin edilmiştir.

Tablo 3.10 Boost86 - doğrulama veri seti için karmaşıklık matrisi

	Gerçek Pozitif (1)	Gerçek Negatif (0)	Toplam
Tahmin Edilen Pozitif (1)	1589	397	1986
Tahmin Edilen Negatif (0)	567	3335	3902
Toplam	2156	3732	5888

Tablo 3.10, modelin doğrulama veri setindeki performansına ait karışıklık matrisini göstermekte olup, gerçek pozitif ve negatif vakaların tahmin edilen pozitif ve negatif vakalar ile karşılaştırılmasını sunmaktadır. Matristeki özet şu şekildedir:

Gerçek Pozitif (1): 1589 örnek doğru bir şekilde pozitif olarak tahmin edilmiştir.
Gerçek Negatif (0): 397 örnek yanlış bir şekilde pozitif olarak tahmin edilmiştir.
Gerçek Pozitif (1): 567 örnek yanlış bir şekilde negatif olarak tahmin edilmiştir.
Gerçek Negatif (0): 3335 örnek doğru bir şekilde negatif olarak tahmin edilmiştir.

Tablo 3.11 ve Tablo 3.12, modelin eğitim veri setindeki temel performans metriklerini sunmaktadır. Bu metrikler, bir sınıflandırma modelinin etkinliğini değerlendirmek için kritik olan Kesinlik (Precision), Duyarlılık (Recall) ve F1 Skoru'nu içermektedir.

Kesinlik (Precision): Pozitif tahminlerin doğruluğunu ölçer. Yüksek bir kesinlik, düşük yanlış pozitif oranına işaret eder.

Duyarlılık (Recall): Modelin tüm ilgili örnekleri tanıma yeteneğini ölçer. Yüksek bir duyarlılık, düşük yanlış negatif oranına işaret eder.

F1 Skoru: Kesinlik ve duyarlılığın harmonik ortalamasıdır ve ikisi arasında bir denge sağlamak istendiğinde kullanılır. Bu metrik, kesinlik ve duyarlılık arasında bir denge sağlamak için faydalıdır.

Tablo 3.11 Boost86 - eğitim veri seti üzerindeki model için performans metrikleri

Metrik	Değer
Kesinlik (Precision)	0,821
Duyarlılık (Recall)	0,761
F1 Skoru	0,790

Tablo 3.11'de, 0,821'lik bir kesinlik değeri, pozitif olarak sınıflandırılan örneklerin yaklaşık %82'sinin gerçekten pozitif olduğunu göstermektedir. Bu yüksek kesinlik, modelin çok az yanlış pozitif hata yaptığını göstermektedir. 0,761'lik bir duyarlılık değeri, modelin tüm gerçek pozitif örneklerin %76'sını başarıyla tanıdığını ifade etmektedir. Bu iyi bir skor olmasına rağmen, pozitif örnekleri tespit etme

konusunda hâlâ iyileştirme için bir alan bulunduğunu göstermektedir. 0,790'lık bir F1 skoru, kesinlik ve duyarlılık arasında bir denge sağlayarak yaklaşık 0,79'luk bir değer sunmaktadır. Bu skor, modelin kesinlik ve duyarlılık arasında iyi bir denge sağladığını ve verilen görev için güvenilir bir sınıflandırıcı olduğunu göstermektedir.

Tablo 3.12 Boost86 - doğrulama veri seti üzerindeki model için performans metrikleri

Metrik	Değer
Kesinlik (Precision)	0,800
Duyarlılık (Recall)	0,737
F1 Skoru	0,767

Tablo 3.12'de, %80'lik bir kesinlik değeri, pozitif olarak sınıflandırılan örneklerin yaklaşık %80'inin gerçekten pozitif olduğunu göstermektedir. Bu yüksek kesinlik, modelin çok az yanlış pozitif hata yaptığını göstermektedir. 0,737'lik bir duyarlılık değeri, modelin tüm gerçek pozitif örneklerin %73,7'sini başarıyla tanıdığını ifade etmektedir. Bu iyi bir skor olmakla birlikte, pozitif örneklerin tespitinde hâlâ iyileştirme için bir alan bulunduğunu göstermektedir. Kesinlik ve duyarlılık arasında bir denge sağlayan 0,767'lik bir F1 skoru, yaklaşık olarak 0,767 olup modelin bu iki ölçüt arasında iyi bir denge sağladığını ve verilen görev için güvenilir bir sınıflandırıcı olduğunu göstermektedir.

Sunulan performans metrikleri, modelin yüksek bir kesinlik ve kesinlik ile duyarlılık arasında makul bir denge ile iyi performans gösterdiğini ortaya koymaktadır. Eğitim ve doğrulama veri setlerindeki bu metriklerin tutarlılığı, modelin dayanıklılığını ve genellenebilirliğini daha da doğrulamaktadır. Ancak, pozitif örneklerin doğru tanımlanmasını sağlamak amacıyla özellikle duyarlılığı artırmak için daha fazla iyileştirme potansiyeli bulunmaktadır.

Boost79 ve Boost86 modellerinin detaylı sonuçları sırasıyla Bölüm 3.5.2, Bölüm 3.5.1'de incelenecektir.

3.5.1 Standartlaştırılmış Değişkenler için Gradient Boosting ile Sürekli Hedef Değişken Uygulaması

Gradient Boosting modeli, farklı hiperparametre ayar stratejileri kullanılarak birden fazla kez çalıştırılmıştır. En iyi model (Boost79) dışında geliştirilen modeller için parametre değerlerine değinilmemiştir. Boost79 Modeli'ni oluşturmak için, eğitim oranı olarak %70 kullanılmıştır. Bölme kuralı olarak yalnızca kare hata kaybı fonksiyonu kullanılmıştır. Taban öğrenici, maksimum üç dal ve beş derinlik olacak

şekilde tasarlanmıştır. Yineleme sayısı ve küçültme parametresi sırasıyla 40 ve 0,1 olarak ayarlanmıştır.

Tablo 3.13, gruplandırılmamış ancak standartlaştırılmış değişkenler kullanılarak kar marjı değişkeninin sayısal bir değişken olarak kullanıldığı durumdaki model sonuçlarına ait uyum istatistiklerini sunmaktadır.

Tablo 3.13 Boost 79 - eğitim ve doğrulama veri seti için uyum istatistikleri - sürekli hedef değişken

Model	Eğitim RASE	Doğrulama RASE	Eğitim ASE	Doğrulama ASE
Boost79	0,593	0,637	0,405	0,351
Boost83	0,609	0,640	0,410	0,372
Boost84	0,621	0,645	0,416	0,386
Boost82	0,747	0,776	0,603	0,558

Tablo 3.13, ASE (Ortalama Kare Hata) ve RASE (Karekök Ortalama Kare Hata) değerlerinin çok yüksek olduğunu ve modelin eğitim verilerine iyi uymadığını göstermektedir.

Tablo 3.14, Boost79 Modeli için eğitim ve doğrulama veri setlerinin önem düzeylerini sunmaktadır.

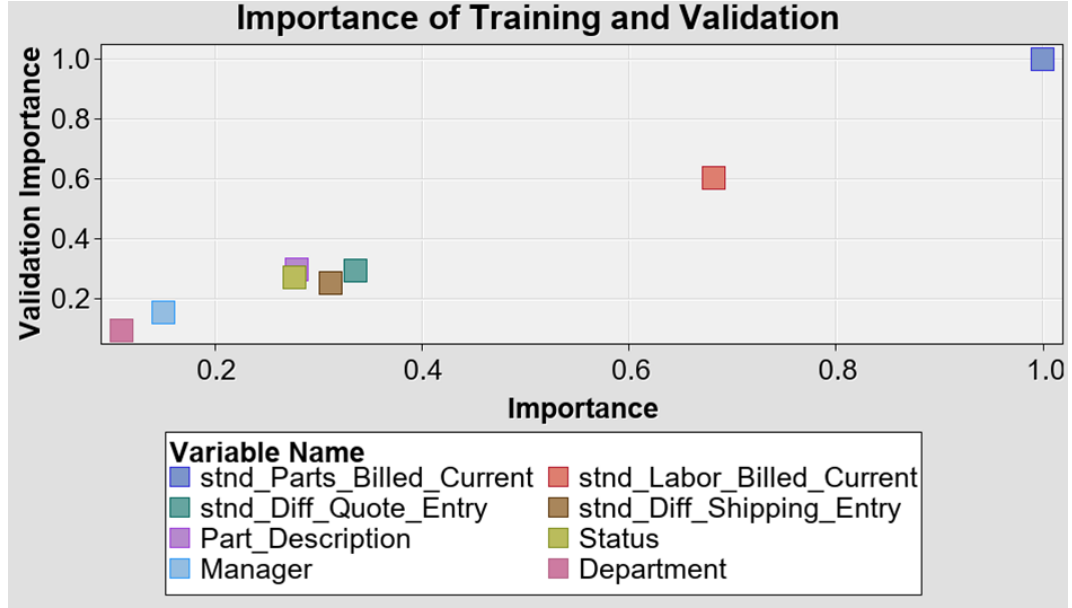
Tablo 3.14 Boost79 - değişken önemi

Değişken Adı	Eğitim Önemi	Doğrulama Önemi
Parts Billed Current	1,00	1,00
Labor Billed Current	0,681	0,603
Diff Quote Entry	0,335	0,294
Diff Shipping Entry	0,311	0,252
Part Description	0,278	0,297
Status	0,277	0,272
Manager	0,149	0,155
Department	0,108	0,095

Tablo 3.14'te, gruplama öncesinde standartlaştırılmış değişkenler ile kurulan modelin değişken önem sırasına bakıldığında sırasıyla, Parts Billed Current, Labour Billed Current, Diff Quote Entry, Diff Shipping Entry, Part Description, Status, Manager ve Department olduğu görülmektedir. Parça Maliyeti, İşçilik Maliyeti, Fiyat Teklif Süresi, Çevrim Süresi, Parça Tanımı, Parça Durumu, İlgili Yönetici ve İlgili Bölüm değişkenleri önem sırasına göre modele giren değişkenleri göstermektedir.

Şekil 3.6, eğitim ve doğrulama veri setlerindeki değişken öneminin saçılım grafiğini sunmaktadır. Değişken önemi, bir düğümün bölünmesinden kaynaklanan kareler

toplamındaki azalmanın kullanılması ile hesaplanır; önem değeri, bu yaklaşımda tüm düğümler üzerinde toplanarak hesaplanmaktadır[22].



Şekil 3.6 Boost79 - önem değeri saçılım grafiği

Şekil 3.6’da, değişkenlerin modeldeki göreceli önemini görselleştirerek, modelin hangi değişkenlere daha çok dayandığını açıkça ortaya koymaktadır. stnd Parts Billed Current, Parça Maliyeti ve stnd Labor Billed Current, İşçilik Maliyeti, değişkenleri hem eğitim hem de doğrulama setlerinde en kritik değişkenlerdir. Bu, bu değişkenlerin modele dahil edilmesinin doğru bir karar olduğunu gösterirken, daha az önemli değişkenlerin model karmaşıklığını artırmaması için dikkatli bir şekilde değerlendirilmesi gerektiğini ortaya koymaktadır. Boost79 modelinin değişken önem değerlerinin, bir sonraki alt bölümde gösterilecek olan Boost86 modeline kıyasla eğitim ve doğrulama veri setlerinde nispeten daha az benzerlik gösterdiği söylenebilir.

3.5.2 Gruplandırılmış Değişkenler için Gradient Boosting ile İkili Hedef Değişken Uygulaması

Bölüm 3.3, İnteraktif Gruplama altında, tüm değişkenler gruplanmış ve Gini indeksi 15’ten büyük olanlar Gradient Boosting modellemesi için aday değişkenler olarak belirlenmiştir. Değişken gruplaması için hedef değişken olan Kar Marjı da sayısal bir değişkenden ikili bir değişkene dönüştürülmüştür. Bu alt bölümde Gradient Boosting modelleri, dönüştürülmüş değişkenler kullanılarak oluşturulmuştur.

Her modelde, farklı hiperparametre ayar stratejileri uygulanmış ve en iyi modele ulaşmak amacıyla model birden fazla kez çalıştırılmıştır. Boost86 Modeli’ni

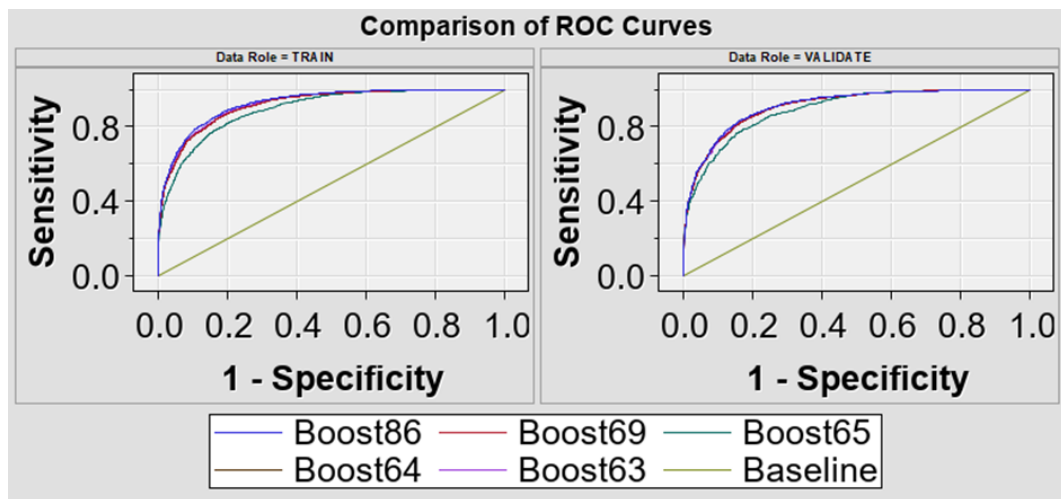
oluşturmak için %70 oranında bir eğitim oranı kullanılmıştır. Bölme kuralı olarak yalnızca kare hata kaybı fonksiyonu kullanılmıştır. Taban öğrenici, maksimum üç dal ve beş derinlik olacak şekilde tasarlanmıştır. Yineleme sayısı ve küçültme parametresi sırasıyla 40 ve 0,1 olarak ayarlanmıştır. Tablo 3.15, ikili hale dönüştürülmüş Kar Marjı hedef değişkeni ile İnteraktif Gruplandırılmış değişkenlere uygulanan Gradient Boosting modellerinden bir sonuç alt kümesini sunmaktadır.

Tablo 3.15 Boost86 - eğitim ve doğrulama veri seti için uyum istatistikleri - ikili hedef değişken

Model	E MCR	D MCR	E ROC	D ROC	E Gini	D Gini
Boost86	0,148	0,164	0,925	0,911	0,849	0,821
Boost63	0,154	0,167	0,919	0,909	0,839	0,818
Boost64	0,156	0,168	0,916	0,908	0,833	0,815
Boost65	0,183	0,186	0,893	0,889	0,785	0,779

Tablo 3.15, E eğitim veri seti ve D doğrulama veri seti olmak üzere, yanlış sınıflandırma oranı (MCR), ROC indeksi ve Gini değerinin farklı modeller için eğitim ve doğrulama veri setlerinde benzer sonuçlar verdiği görülmektedir. Tablo 3.15 incelendiğinde, en iyi modeli seçmek için doğrulama veri seti için yanlış sınıflandırma oranı (MCR) ana kriter olarak değerlendirilmiştir. Yanlış sınıflandırma oranı, yanlış tahmin edilen sonuçların yüzdesini temsil etmektedir. Hedef değişken ve girdi verisinin İnteraktif Gruplama kullanılarak dönüştürülmesi, modelin doğruluğunu olumlu yönde etkilemiştir.

Şekil 3.7, farklı modellerin tahminlerine dayalı ROC eğrilerini göstermektedir.

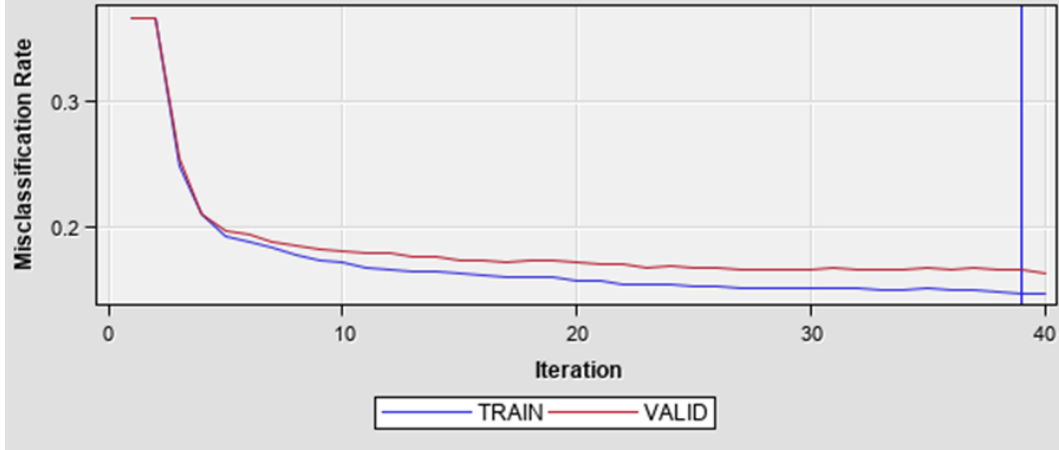


Şekil 3.7 Boost86 - farklı Gradient Boosting modelleri için ROC Eğrileri

Şekil 3.7’de, grafikteki mor çizgi Boost86 Modeli’ni temsil etmektedir. Hem eğitim hem de doğrulama veri setleri için ROC Eğrisi altındaki alanın diğer

modellerden daha büyük olduğu görülmektedir. Boost86 Modeli olarak adlandırılan modelin, diğer modeller arasında her iki veri seti için de en iyi tahminleri yaptığı gözlemlenmiştir.

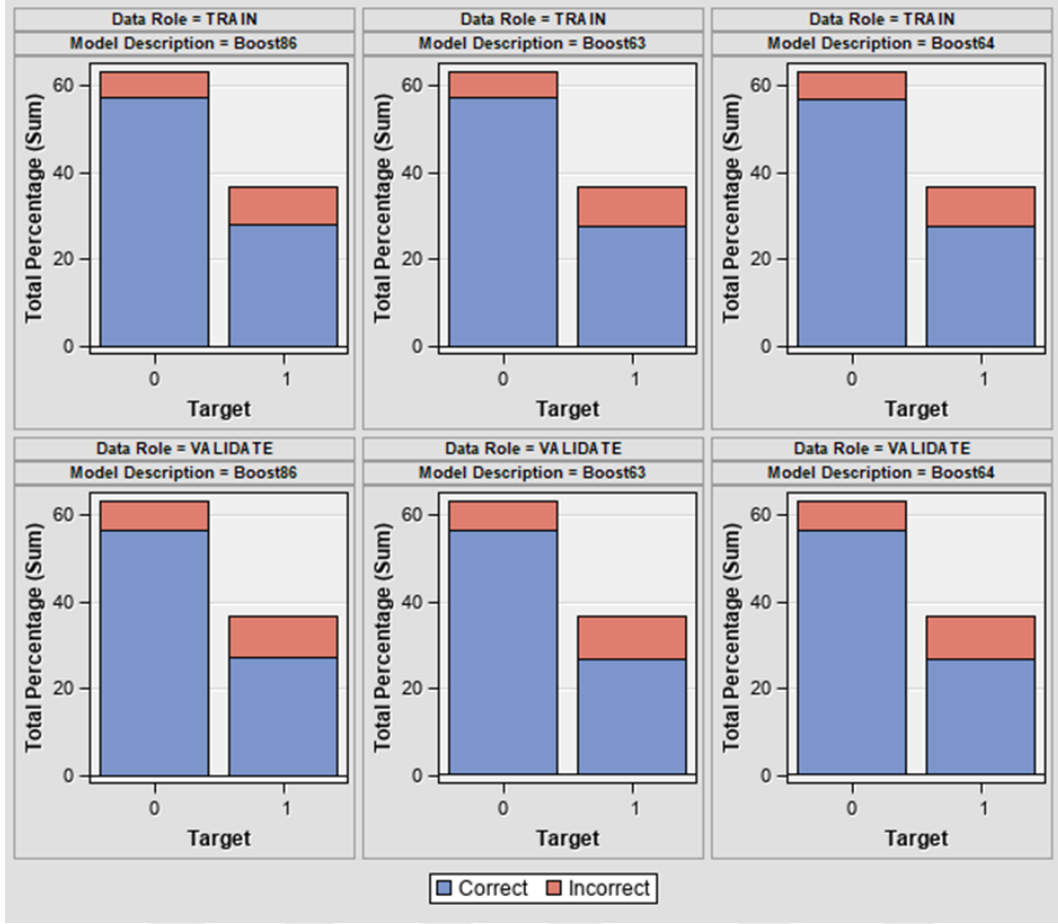
Şekil 3.8, Boost86 Modeli için yanlış sınıflandırma oranını yineleme sayısına göre göstermektedir.



Şekil 3.8 Yanlış sınıflandırma oranı ve yineleme sayısı

Şekil 3.8’da, modelin eğitim (mavi çizgi) ve doğrulama (kırmızı çizgi) veri setlerindeki yanlış sınıflandırma oranlarının iterasyon sayısı ile nasıl değiştiğini göstermektedir. İterasyon sayısı arttıkça her iki veri setindeki yanlış sınıflandırma oranları hızla düşmekte, yaklaşık 10. iterasyondan sonra ise oranlar stabil hale gelmektedir. Eğitim seti, doğrulama setine kıyasla daha düşük bir yanlış sınıflandırma oranı sergilemekte, ancak iki veri seti arasındaki fark minimaldir, bu da modelin fazla öğrenme (overfitting) yapmadığını göstermektedir. Grafik, modelin genel olarak iyi bir doğruluk seviyesine ulaştığını ve iterasyon sayısının artmasıyla performansın stabil hale geldiğini ortaya koymaktadır.

Şekil 3.9, farklı Gradient Boosting modelleri için hedef kategorilerdeki (1: kar oranı %0,5’in üzerinde, 0: kar oranı %0’ın altında) yanlış sınıflandırma oranlarını göstermektedir.



Şekil 3.9 Boost86 - farklı Gradient Boosting modellerinde yanlış sınıflandırma oranları

Şekil 3.9, üç farklı modelin (Boost86, Boost63, Boost64) eğitim (TRAIN) ve doğrulama (VALIDATE) veri setlerinde sınıflandırma başarısını göstermektedir. Her bir grafikte, doğru sınıflandırma (mavi) ve yanlış sınıflandırma (kırmızı) yüzdeleri toplam yüzdelik olarak sunulmuştur. Tüm modellerde, sıfır sınıfındaki (Target = 0) gözlemlerin daha yüksek bir doğru sınıflandırma oranına sahip olduğu, bir sınıfındaki (Target = 1) gözlemlerde ise doğru sınıflandırma oranlarının daha düşük olduğu görülmektedir. Eğitim ve doğrulama veri setleri arasında doğru ve yanlış sınıflandırma oranlarının benzer bir eğilim sergilediği, bu durumun modellerin tutarlılığını ve doğrulama setlerinde de genelleme yeteneklerini ortaya koyduğu söylenebilir. Ancak, Target = 1 sınıfındaki yanlış sınıflandırma oranlarının görece daha yüksek olması, modellerin bu sınıfı tahmin etmede zorlandığını göstermektedir. Analizler ikili hedef değişkene göre yapılmıştır. Bu grafik, Tablo 3.15'teki MCR değerlerinin görsel bir temsidir. Diğer modellere (Boost63 ve Boost64) kıyasla hata oranları birbirine yakın olmakla birlikte, Boost86 Modeli eğitim veri setinde %0,5'in altındaki kar marjı tahminlerinde %6,07 ve doğrulama veri setinde %6,74 oranıyla daha az hata yapmaktadır. Ayrıca, Boost86 Modeli

eğitim veri setinde %0,5'in üzerindeki kar marjı sınıfı tahminlerinde %8,75 ve doğrulama veri setinde %9,63 oranıyla daha az hata yapmaktadır.

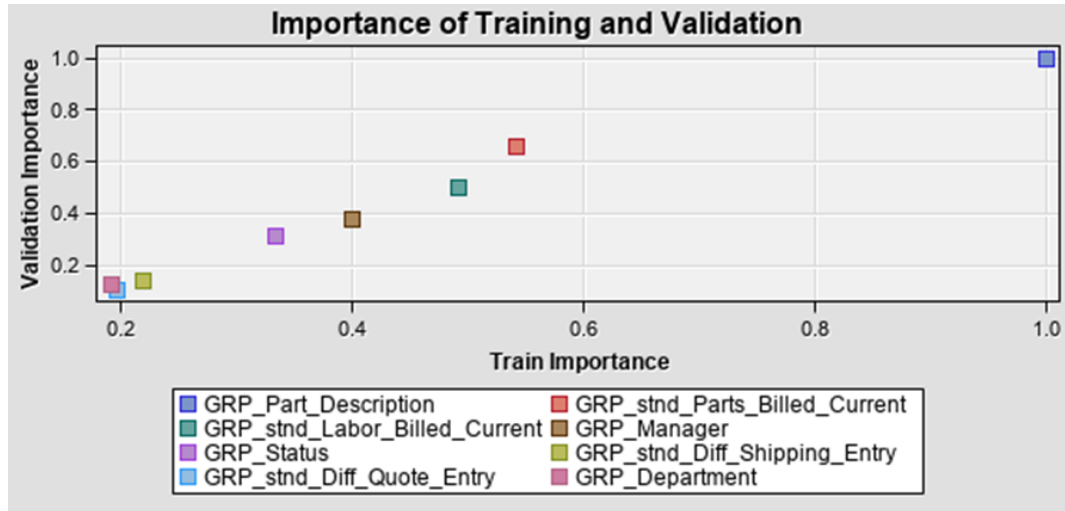
Tablo 3.16, Boost86 Modeli için eğitim ve doğrulama veri setlerinin önem düzeylerini sunmaktadır.

Tablo 3.16 Boost86 - değişken önemi

Değişken Adı	Eğitim Önemi	Doğrulama Önemi
GRP Part Description	1,00	1,00
GRP Parts Billed Current	0,542	0,655
GRP Labor Billed Current	0,493	0,499
GRP Manager	0,402	0,377
GRP Status	0,335	0,309
GRP Diff Shipping Entry	0,219	0,140
GRP Diff Quote Entry	0,197	0,107
GRP Department	0,192	0,122

Tablo 3.16'da, Boost86 Modeli'nin değişken önem değerlerinin eğitim ve doğrulama veri setlerinde birbirine çok yakın olduğu görülmektedir. Bu durum, model doğruluğu için olumlu kriterlerden biridir. Modele dahil olan gruplandırılmış değişkenlerin önem sırası şu şekildedir: Part Description, Parts Billed Current, Labour Billed Current, Manager, Status, Diff Shipping Entry, Diff Quote Entry ve Department. Parça Tanımı, Parça Maliyeti, İşçilik Maliyeti, İlgili Yönetici, Parça Durumu, Çevrimiçi Süre (TAT), Fiyat Teklif Süresi (TTQ) ve İlgili Departman değişkenleri önem derecesine göre modele giren değişkenlerdir.

Şekil 3.10, eğitim ve doğrulama veri setlerinde değişken öneminin saçılım grafiğini sunmaktadır.



Şekil 3.10 Boost86 - önem değeri saçılım grafiği

Şekil 3.10, eğitim (Train Importance) ve doğrulama (Validation Importance) veri setlerinde değişkenlerin önemini karşılaştırmaktadır. X eksenini eğitim veri setindeki önem düzeyini, Y eksenini ise doğrulama veri setindeki önem düzeyini göstermektedir. GRP stnd Parts Billed Current (Parça Maliyeti) değişkeni, her iki veri setinde de en yüksek önem düzeyine sahip olup, modelin kâr marjını tahmin etmede en etkili değişken olduğunu göstermektedir. GRP stnd Labor Billed Current (İşçilik Maliyeti) değişkeni, ikinci sırada yer almakta ve her iki veri setinde de önemli bir katkı sağlamaktadır. GRP stnd Diff Quote Entry (TTQ) ve GRP stnd Diff Shipping Entry (TAT) gibi değişkenler orta düzeyde bir öneme sahiptir. Ancak GRP Manager (İlgili Yönetici), GRP Status (Parça Durumu) ve GRP Department (İlgili Departman) gibi değişkenlerin hem eğitim hem de doğrulama veri setlerinde düşük önem düzeyine sahip olduğu gözlemlenmiştir. Bazı değişkenlerin modelin genelleme kapasitesine olan katkısının sınırlı olduğunu, dolayısıyla bu değişkenlerin gelecekteki model optimizasyon süreçlerinde daha düşük önceliklendirilmesi gerektiğini göstermektedir. Eğitim ve doğrulama setleri arasındaki paralellik, modelin tutarlı bir performans sergilediğini ve aşırı öğrenme (overfitting) yapmadığını ortaya koymaktadır[22].

3.6 Lojistik Regresyon Uygulaması

Bu çalışmada uygulanan Lojistik Regresyon modellerinde değişken seçme kriteri olarak Forward Stepwise metodolojisi kullanılmıştır. Bu yöntem, bir modelin bağımsız değişkenlerini seçmek için kullanılan iteratif bir süreçtir. Bu yöntem, başlangıçta hiçbir bağımsız değişkeni içermeyen bir modelle başlar ve her iterasyonda modele bir değişken eklenir. Eklenen değişken, modelin belirli bir performans ölçütüne (örneğin, AIC, BIC veya log-olabilirlik) en fazla iyileştirme sağlayan değişkendir. Süreç, modele eklenebilecek anlamlı bir değişken kalmayana kadar devam eder. Bu yöntem, modelin yalnızca anlamlı ve önemli değişkenleri içermesini sağlayarak aşırı karmaşıklık önler ve modelin genelleme yeteneğini artırır. Ancak, multikolineerlik gibi durumlarda yanlış değişkenlerin seçilmesine neden olabileceği için dikkatli bir şekilde değerlendirilmelidir. İnteraktif gruplama yapılmış değişkenler için kurulan Lojistik Regresyon modeli sonuçları, aşağıda Tablo 3.17’te değerlendirilmiştir.

Tablo 3.17 Lojistik Regresyon - eğitim ve doğrulama veri setleri için yanlış sınıflama oranı

Uyum İstatistiği	İstatistik Etiketi	Eğitim	Doğrulama
MISC	Yanlış Sınıflandırma Oranı	0,197	0,198

Tablo 3.17, bir modelin eğitim (Train) ve doğrulama (Validation) veri setlerinde çeşitli uyum ve hata istatistiklerini göstermektedir.

Eğitim verisi için hata oranı değeri (MCR) 0,197, doğrulama verisi için hata oranı değeri 0,198 yanlış sınıflandırma oranıdır ve modelin doğruluğunu ölçmek için kullanılan bir diğer metriktir. Bu oranların yakın olması, modelin doğrulama setinde de eğitim setindeki gibi performans gösterdiğini ifade eder.

Bu istatistik, modelin eğitim ve doğrulama veri setlerinde dengeli bir performans sergilediğini ve genelleme yeteneğinin yüksek olduğunu göstermektedir. Eğitim ve doğrulama setlerinde benzer hata oranları ve istatistikler elde edilmesi, modelin aşırı uyum yapmadığını ve doğrulama setinde de tahmin gücünün güçlü olduğunu ortaya koymaktadır. Bu, modelin yeni veri setlerinde de benzer sonuçlar verebileceği anlamına gelir.

Düşük hata oranı ve yanlış sınıflandırma (MCR - misclassification) oranları modelin etkin bir sınıflandırma yaptığına işaret etmektedir. Modelin performansını daha fazla artırmak için daha büyük bir veri seti veya daha fazla özellik eklenmesi önerilebilir, ancak mevcut durumda dengeli ve güçlü bir performans sergilediği söylenebilir.

Forward Stepwise Lojistik regresyonda modele giren değişkenler Tablo 3.18’de yer almaktadır.

Tablo 3.18 Lojistik Regresyon modeline giren değişkenler

Parametre	DF	Tahmin	Pr > ChiSq	Exp(Est)
Intercept	1	5,8357	<0,0001	342,310
G Part Description	1	-0,4799	<0,0001	0,619
G Status	1	-0,5978	<0,0001	0,550
G stnd Labor Billed Current	1	-0,5888	<0,0001	0,555
G stnd Parts Billed Current	1	-1,0112	<0,0001	0,364

Tablo 3.18, Lojistik Regresyon modeline ait Maximum Likelihood Estimates (MLE) sonuçlarını sunmaktadır. Modelde yer alan her bir parametrenin bağımlı değişken üzerindeki etkisi, istatistiksel anlamlılığı ve ilgili Exp(Est) yani olasılık oranı değeri detaylandırılmıştır.

Parametreler (Değişkenler), modelde yer alan değişkenlerin tahmini katsayıları, bağımlı değişken üzerindeki etkilerini göstermektedir. Pozitif katsayılar bağımlı değişkeni artırıcı, negatif katsayılar ise azaltıcı etkiye sahiptir. Tablodaki tüm parametrelerin p-değeri <0.0001 olduğundan, değişkenlerin modelde istatistiksel

olarak anlamlı olduđu gör÷lmektedir. $\text{Exp}(\text{Est})$, tahmin edilen katsayıların üssel dönüşümüdür ve bağımlı değışkenin olasılığındaki değışimi oran olarak ifade eder. 1'in altındaki $\text{Exp}(\text{Est})$ değıerleri bağımlı değışken üzerinde azaltıcı etkiyi, 1'in üzerindeki değıerler ise artırıcı etkiyi gösterir.

Intercept (Kesim Noktası) için katsayı tahmini 5.8357 ve $\text{Exp}(\text{Est})$ değıeri 342.310'dur. Bu, bağımsız değışkenler sıfır olduğunda bağımlı değışkenin temel olasılığının oldukça yüksek olduğunu göstermektedir.

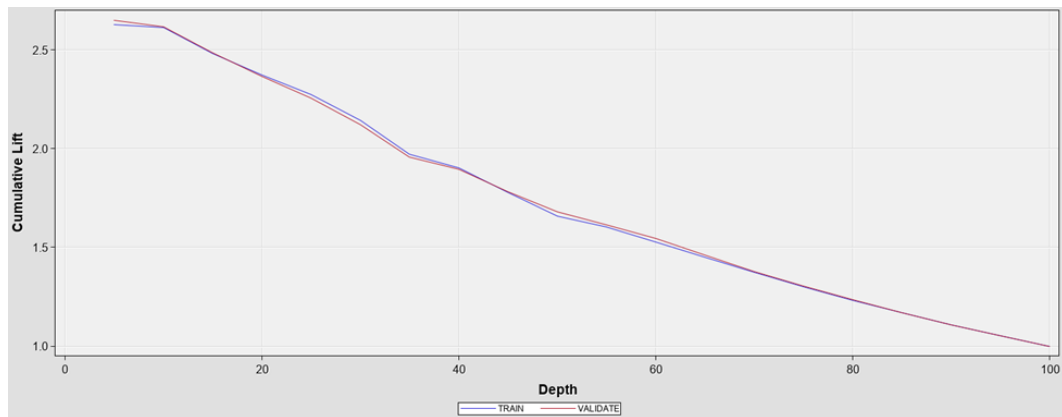
Gruplanmış Part Description (Parça tanımı) için katsayı tahmini -0.4799'dur. Negatif katsayı, bu değışkenin bağımlı değışken üzerindeki etkisinin azaltıcı olduğunu göstermektedir. $\text{Exp}(\text{Est})$ değıeri 0.619'dur, bu da bu değışkenin bir birimlik artışının bağımlı değışkenin olasılığını yaklaşık %38.1 oranında azalttığını ifade eder.

Gruplanmış Status (Parça Durumu) için katsayı tahmini -0.5978 ve $\text{Exp}(\text{Est})$ değıeri 0.550'dir. Bu değışkenin bağımlı değışken üzerindeki etkisi oldukça anlamlıdır ve bir birimlik artış, bağımlı değışkenin olasılığını yaklaşık %45 oranında azaltır.

Gruplanmış stnd Labor Billed Current (İşçilik Maliyeti) için katsayı tahmini -0.5888'dir. Bu değışkenin bağımlı değışken üzerindeki etkisi negatif olup, $\text{Exp}(\text{Est})$ değıeri 0.555'tir. Bu, bir birimlik artışın bağımlı değışken olasılığını yaklaşık %44.5 oranında azalttığını göstermektedir.

Gruplanmış stnd Parts Billed Current (Parça Maliyeti) için katsayı tahmini -1.0112'dir ve bağımlı değışken üzerindeki en güçlü negatif etkiye sahiptir. $\text{Exp}(\text{Est})$ değıeri 0.364'tür, bu da bu değışkenin bir birimlik artışının bağımlı değışkenin olasılığını %63.6 oranında azalttığını ifade eder.

Kümülatif kazanç grafiğı Şekil 3.11'de yer almaktadır.



Şekil 3.11 Lojistik Regresyon - kümülatif kazanç grafiğı

Şekil 3.11 modelin "Eğitim" (TRAIN) ve "Doğrulama" (VALIDATE) veri setlerindeki kümülatif kazanç değerlerini karşılaştırmaktadır. Grafikte, x eksen derinliği (Depth) gösterirken, y eksen kümülatif lift değerini temsil etmektedir. Kümülatif lift, modelin belirli bir yüzdelik dilimdeki performansının rastgele seçimden ne kadar daha iyi olduğunu ifade eden bir ölçüttür. Bu grafik, modelin performansının veri setleri arasında tutarlı olup olmadığını ve genelleme yeteneğini değerlendirmek amacıyla incelenmektedir.

Grafik, hem eğitim (mavi çizgi) hem de doğrulama (kırmızı çizgi) veri setleri için benzer bir kümülatif lift eğilimi göstermektedir. Başlangıçta (düşük derinliklerde), kümülatif lift değeri yüksek başlamakta ve derinlik arttıkça kademeli olarak azalmaktadır. Başlangıçtaki yüksek lift değerleri, modelin en yüksek performansı düşük derinlik seviyelerinde gösterdiğini, yani modelin en iyi tahminleri ilk yüzdelik dilimlerde sağladığını ifade eder. Bu durum, modelin sınıflandırma yeteneğinin en üst düzeyde olduğu ilk dilimlerde hedef sınıfı başarılı bir şekilde yakalayabildiğini göstermektedir.

0-20 Derinlik Aralığı: Modelin en yüksek skorlara sahip ilk %20'lik dilimi temsil eder. Bu dilimde, modelin en iyi tahmin ettiği, en güvenilir pozitif örnekler bulunmaktadır. "Cumulative Lift" değeri burada yüksektir, çünkü model, ilk dilimde hedef sınıfı başarılı bir şekilde ayırt edebilmektedir.

20-100 Derinlik Aralığı: Modelin daha düşük skorlara sahip, yani modelin hedef sınıfı ayırt etme konusunda daha az emin olduğu dilimlere doğru ilerlediğini gösterir. Bu derinlik aralığında, modelin ayırt etme gücü azalmaktadır.

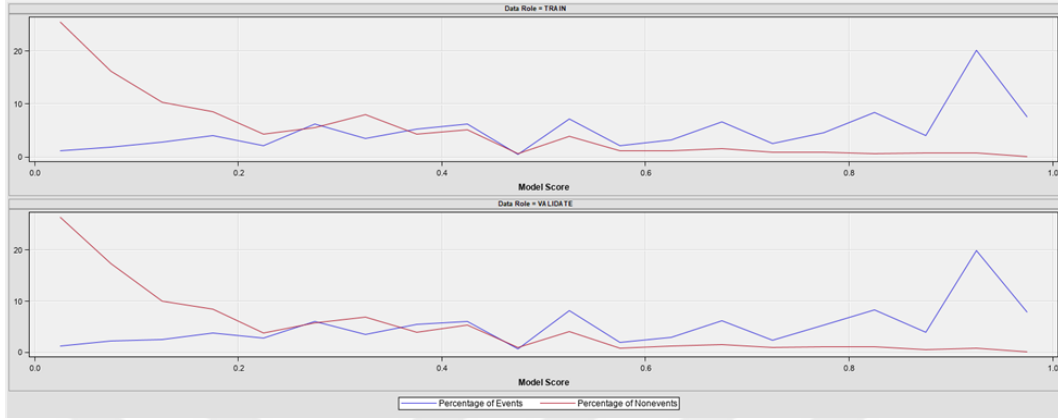
Kısacası, derinlik burada modelin tahmin sonuçlarının güvenilirliğini ve hedef sınıfı ayırt etme performansını değerlendirmek için kullanılan veri dilimlerinin boyutlarını ifade etmektedir. Derinlik arttıkça modelin ayırt edici gücü düşer, çünkü en yüksek performanslı dilimler (en yüksek skorlu tahminler) başlangıç dilimlerinde yer almaktadır.

Şekil 3.11'da eğitim ve doğrulama veri setlerinde kümülatif kazanç eğrilerinin birbirine oldukça yakın olması, modelin genelleme yeteneğinin güçlü olduğunu ve eğitim sırasında öğrendiği örüntüleri doğrulama setinde de sürdürebildiğini göstermektedir. Her iki eğri de benzer bir azalma eğilimi sergilediğinden, modelin hem eğitim hem de doğrulama veri setlerinde tutarlı bir performans gösterdiği söylenebilir. Bu uyum, modelin eğitim verisine aşırı uyum sağlamadığını ve yeni veri setlerine de benzer performansı aktarabileceğini gösterir.

Bu analiz, modelin performansını değerlendirmede önemli bir gösterge

sunmaktadır. Özellikle düşük derinliklerdeki yüksek kazanç değerleri, modelin hedef sınıfı doğru bir şekilde ayırt edebildiğini ve bu nedenle modelin işlevsel ve genellenebilir bir sınıflandırıcı olarak kullanılabileceğini göstermektedir.

Olayların kümülatif yüzdeler dilimi analizi Şekil 3.12'deyi almaktadır.



Şekil 3.12 Lojistik Regresyon - kümülatif yüzdeler dilimi grafiği

Şekil 3.12, modelin "Eğitim" (TRAIN) ve "Doğrulama" (VALIDATE) veri setlerinde elde ettiği "Model Skoru"na göre olay (Events) ve olay olmayan (Nonevents) yüzdesini karşılaştırmaktadır. Grafiklerde, mavi çizgi olayların yüzdesini, kırmızı çizgi ise olay olmayanların yüzdesini göstermektedir. Bu analiz, modelin farklı skor aralıklarında olay ve olay olmayanları ayırt etme yeteneğini değerlendirmeye yardımcı olur.

Eğitim Veri Seti (Data Role = TRAIN)

Olay Yüzdesi (Mavi Çizgi):

Mavi çizgi, düşük skor seviyelerinde oldukça düşük bir düzeyde seyretmekte, ancak skor arttıkça hafif bir yükselme eğilimi göstermektedir. Skoru 0,8'den sonra ciddi bir artış gösterdiği gözlemlenmektedir. Bu durum, modelin yüksek skorlar atadığı örneklerin çoğunlukla olay sınıfına ait olduğunu, yani modelin yüksek skor aralığında olayları doğru bir şekilde tanımlama eğiliminde olduğunu göstermektedir.

Olay Olmayan Yüzdesi (Kırmızı Çizgi):

Kırmızı çizgi ise düşük skor seviyelerinde yüksek bir yüzdeler oranda seyretmekte, model skoru arttıkça hızla düşmektedir. Bu, modelin düşük skorları genellikle olay olmayan sınıfa atadığını ve bu sınıf için doğru tahminlerde bulunduğunu göstermektedir.

Eğitim verisinde, mavi ve kırmızı çizgi arasında belirgin bir ayrım olduğu gözlemlenmektedir. Özellikle yüksek skor aralıklarında, olayların yüzdesinin (mavi çizgi) olay olmayanlardan (kırmızı çizgi) belirgin şekilde ayrıldığı görülmektedir. Bu durum, modelin eğitim verisinde olay ve olay olmayanları ayırt etme yeteneğinin güçlü olduğunu ifade eder.

Doğrulama setinde, olay yüzdesini temsil eden mavi çizgi, eğitim setindeki gibi düşük skor seviyelerinde düşük kalmakta, ancak yüksek skor seviyelerinde belirgin bir şekilde yükselmektedir. Bu eğilim, modelin doğrulama setinde de yüksek skorları olaylara atadığı ve olayları doğru bir şekilde tanımlama eğiliminde olduğunu göstermektedir.

Doğrulama setindeki kırmızı çizgi, düşük skorlarda yüksek bir yüzdede başlayarak skor arttıkça azalmaktadır. Bu desen, eğitim seti ile benzerlik gösterir ve modelin düşük skorlarda olay olmayan sınıfını doğru bir şekilde tanımlayabildiğini ifade eder.

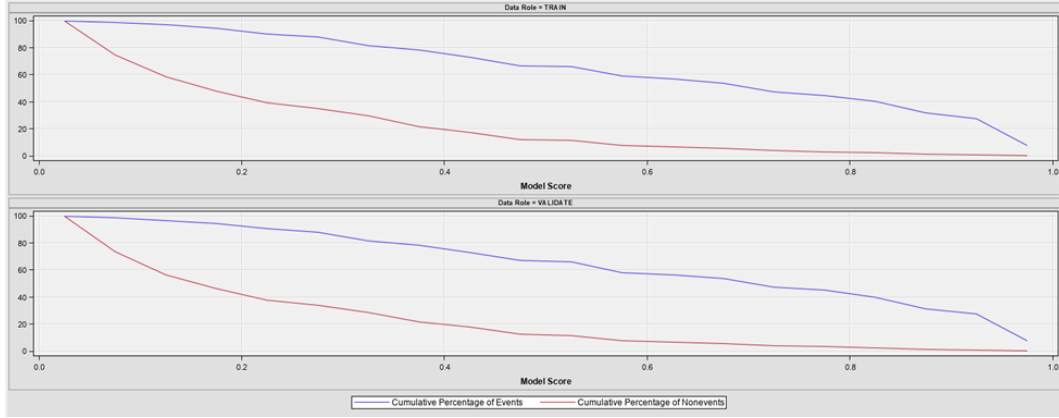
Doğrulama setinde de mavi ve kırmızı çizgi arasındaki ayrım korunmaktadır. Özellikle yüksek skor aralıklarında olay yüzdesinin olay olmayan yüzdesinden belirgin şekilde ayrıldığı görülmektedir. Bu durum, modelin doğrulama setinde de eğitim setindeki gibi olayları ve olay olmayanları ayırt edebilme yeteneğinin güçlü olduğunu gösterir.

Bu grafikler, modelin eğitim ve doğrulama veri setlerinde olay ve olay olmayan sınıflarını ayırt etme yeteneğini etkili bir şekilde ortaya koymaktadır. Eğitim ve doğrulama setleri arasındaki benzerlik, modelin genelleme yeteneğinin yüksek olduğunu ve aşırı uyum yapmadığını gösterir.

Özellikle yüksek skor aralıklarında olay ve olay olmayanlar arasında gözlemlenen belirgin ayrışma, modelin olayları (yüksek skorlar) ve olay olmayanları (düşük skorlar) başarılı bir şekilde sınıflandırdığını ifade etmektedir. Bu sonuç, modelin gerçek dünyadaki veri setlerinde de benzer performans gösterebileceğine dair güven vermektedir.

Ancak, düşük skor aralıklarında olay ve olay olmayan yüzdelерinin yakın seyretmesi, modelin bu aralıkta biraz daha fazla hata yapabileceğini düşündürmektedir. Bu durumun iyileştirilmesi için modelde bazı ayarlamalar yapılabilir. Genel olarak, modelin performansı eğitim ve doğrulama veri setlerinde oldukça tutarlıdır ve bu da modelin güvenilir ve genellenebilir bir sınıflandırıcı olduğunu göstermektedir.

Lojistik Regresyon modeli için oluşturulan skor dağılım grafiği Şekil 3.13'te yer almaktadır.



Şekil 3.13 Lojistik Regresyon - skor dağılım grafiği

Şekil 3.13, modelin performansını eğitim (TRAIN) ve doğrulama (VALIDATE) veri setlerinde kümülatif olay ve olay olmayan yüzdeleri üzerinden analiz etmektedir. Model skoru ile kümülatif yüzdelere arasındaki ilişkiyi gösteren bu grafikler, modelin olayları ve olay olmayanları ayırt etme kapasitesini değerlendirmemizi sağlar.

Eğitim Veri Seti (Üst Grafik)

Kümülatif Olay Yüzdesi (Mavi Çizgi): Mavi çizgi, model skoru düşükten yükseğe doğru ilerledikçe üstten alta doğru yavaş bir azalma göstermektedir. Bu durum, eğitim veri setindeki olayların kümülatif yüzdesinin, düşük skorlarla daha az ilişkilendirildiğini ancak yüksek skorlarla daha fazla ilişkilendirildiğini göstermektedir. Yani, model, olayların çoğunluğunu yüksek skor aralıklarında doğru bir şekilde tanımlamıştır.

Kümülatif Olay Olmayan Yüzdesi (Kırmızı Çizgi): Kırmızı çizgi ise başlangıçta hızlı bir düşüş sergileyip, düşük skorlarda daha yüksek bir eğimle azalmakta, daha yüksek skorlarda ise yatay bir seyir izleyerek azalmaktadır. Bu sonuç, olay olmayanların düşük skorlarla ilişkilendirildiğini ve modelin düşük skorları olay olmayanlara doğru bir şekilde tahsis ettiğini göstermektedir.

Olay ve Olay Olmayanların Ayrımı: Grafik boyunca mavi ve kırmızı çizgi arasındaki ayrım özellikle 0,2 skor seviyesinden itibaren belirginleşmektedir. Bu, modelin belirli bir skor eşliğinden sonra olay ve olay olmayanlar arasında daha güçlü bir ayrım yapabildiğini göstermektedir. Yüksek skor aralıklarında mavi çizginin kırmızı çizgiden uzaklaşması, modelin olayları doğru bir şekilde tespit

etme yeteneğini vurgulamaktadır.

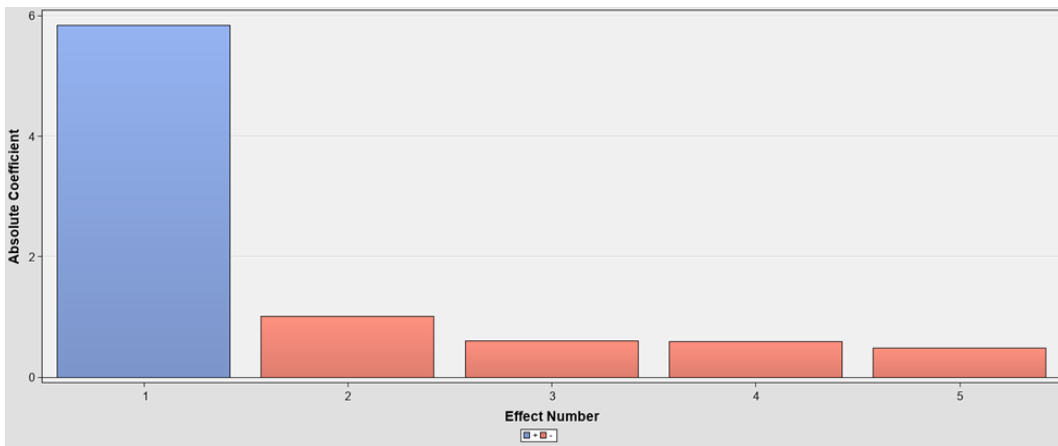
Doğrulama Veri Seti (Alt Grafik) Eğitim Setine Benzer Performans: Alt grafik, doğrulama veri setindeki model performansını göstermektedir ve eğitim seti ile benzer bir desen izlemektedir. Bu benzerlik, modelin doğrulama setine aşırı uyum (overfitting) yapmadığını ve öğrenilen örüntülerin doğrulama verisine de başarılı bir şekilde uygulandığını göstermektedir.

Kümülatif Yüzde Eğrileri Arasındaki Uyum: Mavi ve kırmızı çizgilerde eğitim setine benzer bir ayrışma eğilimi gözlenmektedir. Olay olmayanlar düşük skorlarda yüksek bir yoğunluk gösterirken, olaylar yüksek skorlarda daha fazla toplanmaktadır. Bu durum, modelin doğrulama setinde de olayları ve olay olmayanları başarılı bir şekilde ayırt edebildiğini ve olaylara yüksek skorlar, olay olmayanlara ise düşük skorlar atadığını ifade etmektedir.

Genelleme Yeteneği: Eğitim ve doğrulama verilerinde gözlemlenen benzer desenler, modelin eğitim verisinde öğrendiği ayrıştırma yeteneğini doğrulama setinde de uygulayabildiğini ve genelleme yeteneğinin yüksek olduğunu göstermektedir. Bu, modelin eğitim verisine aşırı bağımlı olmadığını ve gerçek dünyada benzer veri setlerine uygulandığında benzer performans gösterebileceğini ima etmektedir.

Özetle, bu analiz, modelin eğitim ve doğrulama veri setlerinde olay ve olay olmayan ayrımını başarıyla gerçekleştirdiğini ve genelleme yeteneğinin tatmin edici düzeyde olduğunu ortaya koymaktadır. Modelin düşük skor aralıklarında daha fazla ayrıştırma yapabilmesi için ek düzenlemeler önerilebilir.

Mutlak katsayı grafiği Şekil 3.14’de yer almaktadır.



Şekil 3.14 Lojistik Regresyon - mutlak katsayı grafiği

Şekil 3.14, modeldeki bağımsız değişkenlerin etkilerini, "Absolute Coefficient" (Mutlak Katsayı) değeri ile göstermektedir. Grafikte, her bir değişkenin model üzerindeki etkisi, mutlak katsayı büyüklüklerine göre sıralanmıştır. Y eksenı mutlak katsayı değerlerini, X eksenı ise değişken numarasını (Effect Number) temsil etmektedir. Bu analiz, modelin hangi değişkenlere daha fazla ağırlık verdiğini ve hangi değişkenlerin modelde daha güçlü etkiler oluşturduğunu belirlemeye yardımcı olur.

Değişkenlerin Etki Analizi Birinci Değişkenin Etkisi (Effect Number 1):

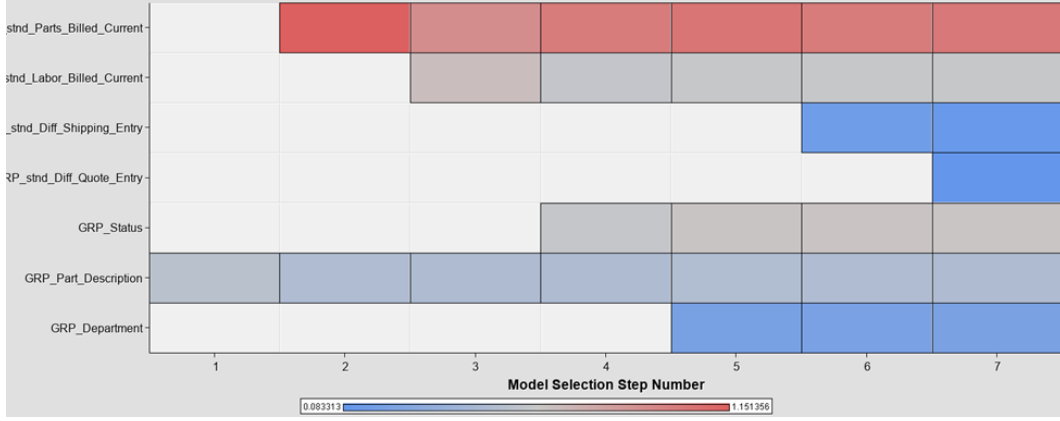
Birinci değişkenin mutlak katsayısı diğer tüm değişkenlere göre oldukça yüksek olup yaklaşık 6 değerindedir. Bu değer, birinci değişkenin modeldeki en güçlü etkene sahip olduğunu ve diğer değişkenlere kıyasla modelin sonuçlarını daha fazla etkilediğini göstermektedir. Bu durum, birinci değişkenin bağımlı değişken üzerindeki açıklayıcılığının oldukça yüksek olduğunu ve modelde en belirleyici unsur olduğunu ifade eder. Bu değişken, modelin tahmin gücüne önemli bir katkı sağlamaktadır.

İkinci, üçüncü, dördüncü ve beşinci değişkenlerin mutlak katsayı değerleri, birinci değişkene kıyasla oldukça düşüktür ve birbirine yakın seviyelerdedir. Bu değişkenlerin mutlak katsayıları yaklaşık olarak 1'in altında olup, model üzerinde daha sınırlı bir etkiye sahip olduklarını göstermektedir. Bu durum, modelin bağımlı değişkeni tahmin etme sürecinde bu değişkenlerden daha az yararlandığını ve bu değişkenlerin model için ikincil önem taşıdığını ifade eder.

Grafik genelinde birinci değişkenin açık ara farkla en etkili değişken olduğu, diğer değişkenlerin ise model üzerinde göreceli olarak daha düşük etkiler sağladığı görülmektedir. Bu sonuç, modeldeki tahmin sürecinde ilk değişkene verilen önemin, diğer değişkenlerden çok daha fazla olduğunu göstermektedir. Dolayısıyla, modelin doğruluğu ve performansı üzerinde birinci değişkenin etkisinin kritik olduğu söylenebilir.

Bu tür bir analiz, model optimizasyonu ve değişken seçimi açısından önemli bilgiler sunmaktadır. Özellikle birinci değişkenin modele olan katkısı göz önünde bulundurulduğunda, bu değişkenin verimliliğini artırmak veya daha iyi temsil edilmesini sağlamak model performansını olumlu yönde etkileyebilir. Diğer değişkenlerin düşük etkisi göz önüne alındığında, modelin sadeliğini korumak amacıyla bu değişkenler üzerinde daha az odaklanılabilir veya modeldeki etkilerini artırmak için bu değişkenlerde ek düzenlemeler yapılabilir.

Tahmin seçim grafiği Şekil 3.15'te yer almaktadır.



Şekil 3.15 Lojistik Regresyon - tahmin seçim grafiği

Şekil 3.15, model seçim sürecinde farklı değişkenlerin katkısını gösteren bir ısı haritasıdır. Y ekseninde, modelde yer alan bağımsız değişkenleri (Etki), X ekseninde ise model seçim adımlarını (Model Seçim Adım Numarası) temsil etmektedir. Bu grafik, lojistik regresyon modeline değişkenlerin hangi adımlarda dahil edildiğini ve katkı büyüklüklerini göstermektedir. Y ekseninde değişkenler, X ekseninde ise modelin seçim adımları yer alır. Renk tonları, maviden kırmızıya gittikçe daha yüksek katkıyı gösterir.

GRP stnd Parts Billed Current, Parça Maliyeti ve GRP stnd Labor Billed Current (İşçilik Maliyeti) değişkenleri, model seçim sürecinin başından itibaren modele katkı sağlamaktadır. Negatif Etkili Değişkenler:

GRP stnd Diff Shipping Entry (Çevrimiçi Süre) ve GRP stnd Diff Quote Entry (Fiyat Teklif Süresi) değişkenleri model seçim sürecinde, özellikle 5. adımdan itibaren dahil edilmişlerdir. Tahmin performansını arttırmadıkları söylenebilir. Forward Stepwise yöntemi adımlarında elendikleri için nihai modele dahil edilmemişlerdir.

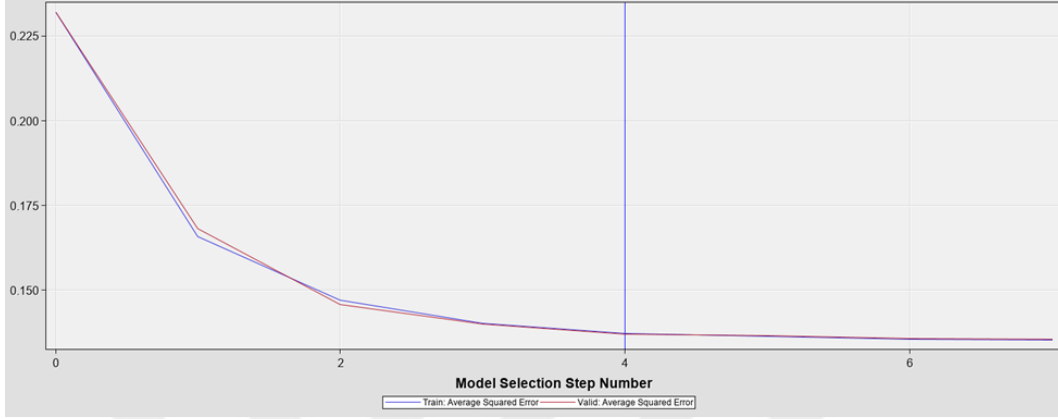
GRP Department (İlgili Departman) değişkeni daha açık tonlarda görüldüğünden model üzerinde nispeten daha zayıf bir etkiye sahiptir. Etkisinin sınırlı olması, bu değişkenin modelin toplam performansını anlamlı bir şekilde değiştirmedikçe göstermektedir. Forward Stepwise yöntemi sırasında elenerek modele dahil olmamıştır.

GRP Part Description (Parça Tanımı) değişkeni ise 1. adımda modele girmiş ve nihai modele kadar modelde kalmayı başarmıştır.

Bu analiz, modelin seçim sürecinde hangi değişkenlere daha fazla ağırlık verildiğini açıklamaktadır.

Bu tür bir analiz, modelde yer alan değişkenlerin seçimi ve etkilerinin anlaşılması açısından değerli bir kaynaktır ve modelin doğruluğuna katkıda bulunan temel unsurların daha iyi kavranmasını sağlar.

Model seçim adım sayısına göre ortalama kare hata değişimi" Şekil 3.16'da yer almaktadır.



Şekil 3.16 Lojistik Regresyon - model seçim adım sayısına göre ortalama kare hata değişimi

Şekil 3.16, model karmaşıklığı arttıkça hata oranındaki değişimi inceleyerek modelin doğruluk performansını ve genelleme yeteneğini değerlendirmeye imkan tanır.

Grafik, model seçim adımı arttıkça hem eğitim hem de doğrulama veri setlerinde hatanın azaldığını göstermektedir. Bu durum, modelin her adımda daha fazla özellik ekleyerek veya ayarları iyileştirerek hata oranını azalttığını ve dolayısıyla tahmin doğruluğunu artırdığını göstermektedir. Özellikle ilk adımlarda hızlı bir azalma gözlemlenmekte olup, bu durum modelin başlangıçta önemli iyileştirmeler kaydettiğini ve ilk birkaç adımın model performansı üzerinde büyük bir etkiye sahip olduğunu göstermektedir.

Eğitim ve doğrulama veri setlerine ait hata oranları birbirine oldukça yakındır ve benzer bir düşüş eğilimi sergilemektedir. Bu paralel hareket, modelin eğitim setindeki öğrenimini doğrulama setinde de başarılı bir şekilde sürdürebildiğini ve aşırı uyum (overfitting) eğiliminin düşük olduğunu göstermektedir. Bu uyumluluk, modelin eğitim setinde öğrendiği örüntüleri doğrulama setinde de genellebildiğine işaret eder; bu da modelin sağlam bir genelleme yeteneğine sahip olduğunu gösterir.

Grafikte yaklaşık olarak 4. adımda dikey bir çizgiyle işaretlenmiş bir dönüm noktası bulunmaktadır. Bu nokta, muhtemelen modelin en uygun performansı sağladığı

adımı temsil etmektedir. Bu noktadan sonra hata değerindeki düşüş minimal seviyeye ulaşmakta ve her iki veri seti için de grafik yatay bir seyir izlemektedir. Bu durum, modelin optimum karmaşıklık seviyesine ulaştığını ve daha fazla özellik veya karmaşıklığın model performansına anlamlı bir katkı sağlamadığını ifade etmektedir.

Bu grafik, modelin performansını ve genelleme yeteneğini ölçmek için değerli bilgiler sunmaktadır. Model seçim sürecinde ortalama kare hatasının her adımda düşmesi, modelin doğruluk oranını artırdığını göstermektedir. Eğitim ve doğrulama veri setlerindeki hata oranlarının birbirine yakın olması, modelin aşırı uyum yapmadığını ve yeni verilerde de benzer performans sergileyebileceğini ifade eder.

Yaklaşık olarak 4. adımda gözlemlenen dönüm noktası, modelin en iyi performans gösterdiği optimum adım olarak değerlendirilmelidir. Bu adım sonrasında daha fazla özellik eklenmesi veya modelin daha karmaşık hale getirilmesi, doğruluk performansına anlamlı bir katkı sağlamamaktadır.

Bu analiz, model seçim sürecinde optimum adım sayısının belirlenmesi ve modelin genelleme yeteneğinin değerlendirilmesi açısından önemli bir içgörü sunmaktadır.

3.7 Lojistik Regresyon - Hava Yolu Firmaları ve Tedarikçi Firmalar İçin Uygulama

İnteraktif gruplama yapılmış değişkenler ile geliştirilen Gradient Boosting ve Lojistik Regresyon modelleri sonrasında, kâr marjının ve kâr marjını etkileyen dinamiklerin müşteri tiplerine göre farklılık gösterip göstermediği analiz edilmek istenmiştir. SME MRO firmasına ait müşteriler, havayolu firmaları ve tedarikçi firmalar olarak iki farklı gruba ayrılmış ve uygulanan Lojistik Regresyon modelleri ile kâr marjını etkileyen değişkenler analiz edilmiştir.

3.7.1 Hava Yolu Firmaları İçin Lojistik Regresyon Analizi Uygulaması

İnteraktif Gruplama yapılmış değişkenler ile hava yolu firmaları için kurulan Lojistik Regresyon modeli sonuçları aşağıda, Tablo 3.19’da, değerlendirilmiştir.

Tablo 3.19 Lojistik Regresyon havayolu - eğitim ve doğrulama veri setleri için uyum istatistikleri

Uyum İstatistiği	İstatistik Etiket	Eğitim	Doğrulama
MISC	Yanlış Sınıflandırma Oranı	0,203	0,192

Tablo 3.19, bir modelin eğitim ve doğrulama veri setlerinde yanlış sınıflandırma

oranı (MCR). Aşağıda her bir istatistiğin anlamı ve yorumlanmasına verilmiştir.

MISC (Misclassification Rate - Yanlış Sınıflandırma Oranı), eğitim veri setinde 0,203, doğrulama veri setinde 0,192 değerlerini almıştır. Yanlış Sınıflandırma Oranı, modelin hatalı sınıflandırdığı örneklerin oranını ifade eder. Bu oran ne kadar düşükse, modelin doğruluğu o kadar yüksektir. Eğitim veri setinde yanlış sınıflandırma oranı %20.34, doğrulama veri setinde ise %19.20 olarak hesaplanmıştır. Bu oranların birbirine yakın olması, modelin genelleme yeteneğinin iyi olduğunu ve doğrulama veri setinde de benzer bir performans gösterdiğini belirtir. Yani model, yalnızca eğitim veri setinde değil, doğrulama veri setinde de makul bir doğrulukla tahmin yapmaktadır. Bu durum, modelin doğrulama verisi üzerinde aşırı öğrenme yapmadığını ve genelleme yeteneğinin iyi olduğunu gösterir. Modelin, her iki veri setinde de benzer performans göstermesi, eğitim sırasında öğrendiği kalıpları yeni verilere başarılı bir şekilde uygulayabildiği anlamına gelir.

Forward Stepwise Lojistik regresyonda modeline giren değişkenler Tablo 3.20’de yer almaktadır.

Tablo 3.20 Lojistik Regresyon havayolu - gruplanmış değişkenler için adım adım etki giriş istatistikleri

Adım	Girilen Etki	Etki Sayısı	P Değeri	AIC	O AIC
0	Kesim (Intercept)	1	NaN	11248,79	0
1	Part_Description	2	0	9764,228	0
2	Parts_Billed_Current	3	1,82E-256	8519,652	0
3	Labor_Billed_Current	4	1,75E-69	8206,353	0
4	Diff_Quote_Entry	5	4,12E-18	8132,738	0
5	GRP_Manager	6	8,40E-08	8106,022	0
6	Diff_Shipping_Entry	7	1,36E-04	8093,481	1

Tablo 3.20 incelendiğinde,

Adım : Her bir model adımını ifade eder.

Girilen Etki (Effect Entered): Modele her adımda eklenen değişken veya etki.

Etki Sayısı (Number of Effects): Modele dahil olan toplam etki sayısı.

P Değeri (P Value): İlgili değişkenin p-değeri.

AIC: Akaike Bilgi Kriteri değeri.

O AIC: Modeldeki AIC’nin optimal olup olmadığını belirtir (1 optimal, 0 değil).

Modelin Aşamaları ve Anlamlı Değişkenler:

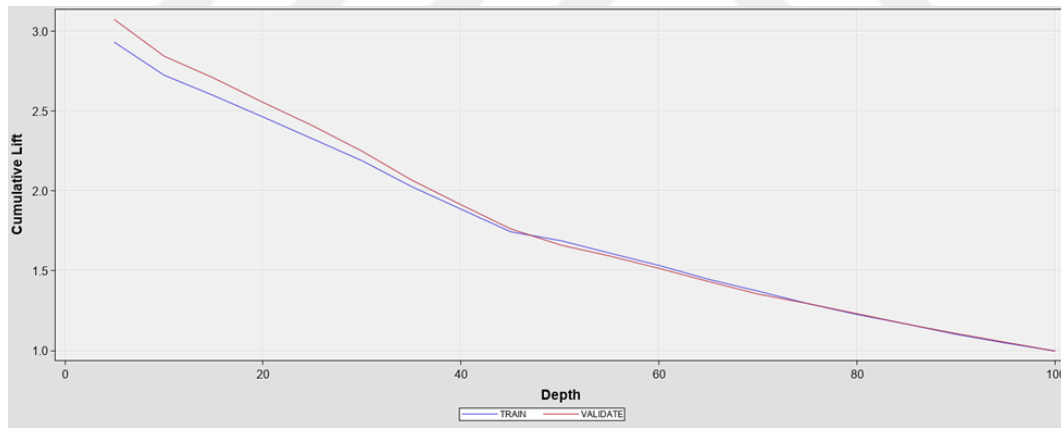
İlk adımlarda eklenen değişkenler (GRP stnd Part Description - Parça Tanımı ve GRP stnd Parts Billed Current - Parça Maliyeti) modele önemli katkılar sağlamaktadır ve düşük p-değerlerine sahiptirler. Bu değişkenlerin eklenmesi, modelin genel uyumunu artırmış ve AIC değerinde anlamlı düşüşler sağlamıştır.

AIC ve Modelin Optimal Aşaması:

AIC değeri, her bir yeni değişkenin modele eklenmesiyle birlikte düzenli olarak düşmüştür ve Adım 6'da en düşük seviyeye ulaşmıştır. Bu aşamadan sonra yeni bir değişken eklemek model performansını anlamlı bir şekilde iyileştirmemiştir; dolayısıyla, mevcut değişkenler ile optimal bir modele ulaşılmıştır.

Forward Stepwise modelleme süreci, modelin hem doğruluk hem de karmaşıklık açısından en uygun seviyeye ulaşması için gerekli değişkenleri seçmeyi hedeflemektedir. GRP stnd Parts Billed Current (Parça Maliyeti), GRP stnd Labor Billed Current (İşçilik Maliyeti) ve GRP Part Description (Parça Tanımı) gibi değişkenlerin modele büyük katkılar sağladığı gözlemlenmiştir. Model seçim süreci, her adımda daha iyi uyum sağlayarak optimal model yapısına ulaşmıştır.

Hava yolu firmaları için kümülatif kazanç grafiği Şekil 3.17'de yer almaktadır.



Şekil 3.17 Lojistik Regresyon havayolu - kümülatif kazanç grafiği

Şekil 3.17, eğitim ve doğrulama setleri arasında oldukça yakın bir uyum olduğunu ortaya koymaktadır. Her iki veri seti için kümülatif ölçüm değerleri birbirine paralel bir şekilde azalmaktadır. Bu durum, modelin eğitim ve doğrulama setlerinde benzer performans gösterdiğini ve aşırı uyum (overfitting) riskinin düşük olduğunu düşündürmektedir. Bu paralellik, modelin genelleme kabiliyetinin yüksek olduğuna dair bir gösterge olarak yorumlanabilir.

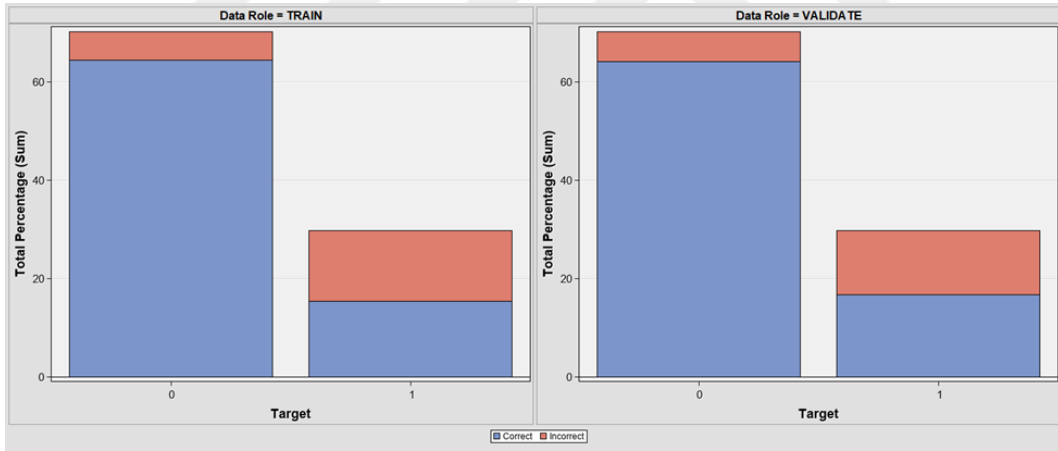
Derinlik arttıkça kümülatif ölçümde belirgin bir azalma görülmektedir. Bu azalma, modelin derinleştikçe daha detaylı bilgilere odaklandığını, ancak kümülatif

performansın azalmakta olduğunu işaret etmektedir. Derinlik seviyesinin artmasıyla birlikte elde edilen bu düşüş, çok derin modellere yönelmenin sınırlı fayda sağladığını ve belirli bir derinlik seviyesinden sonra model performansında azalan getiriler görülebileceğini ortaya koymaktadır.

Derinlik 0 ile 20 arasında sabit bir eğilim gözlenirken, bu aralıktan sonra kümülatif ölçümde sürekli bir azalma meydana gelmektedir. Lift = 2.5 civarında başladığına göre, model en üst yüzde dilimlerde (örn. ilk %10–%20) hedeflenen sınıfı rastgele seçime göre 2.5 kat daha iyi yakalamaktadır. Bu, modelin özellikle yüksek skorlu gruplarda güçlü tahmin yaptığını gösterir. X eksenini ilerledikçe (daha geniş kitlelere uygulandıkça) lift değeri 1'e yaklaşmaktadır. Bu, modelin ayırıcı gücünün üst dilimlerde yoğunlaştığını, daha alt gruplarda azaldığını gösterir. Beklenen bir durumdur; çünkü rastgeleliğe yaklaşılr.

Bu analizde eğitim ve doğrulama verilerinin benzer performans sergilemesi, modelin sağlam bir genelleme yeteneğine sahip olduğunu göstermektedir.

Hava yolu firmaları için yanlış sınıflandırma oranları grafiği Şekil 3.18'de yer almaktadır



Şekil 3.18 Lojistik Regresyon havayolu - yanlış sınıflandırma oranları grafiği

Şekil 3.18, modelin eğitim (TRAIN) ve doğrulama (VALIDATE) veri setlerindeki performansını "hedef" (Target) sınıfları açısından değerlendirmektedir. Grafik, doğru ve yanlış sınıflandırma oranlarını görsel olarak ifade etmektedir. Grafik üzerindeki bölümler, modelin doğruluk ve hata oranlarını kıyaslama olanağı sunar.

Genel Grafik Yapısı:

X Eksenini: "Target" değişkeni yani Kar Marjı değişkeni, hedef sınıfları 0 ve 1 olarak temsil eder.

0: Negatif sınıf (örneğin, olayın gerçekleşmediği durum).

1: Pozitif sınıf (örneğin, olayın gerçekleştiği durum).

Y Eksen: Yüzdelik dağılım, her bir hedef sınıfındaki toplam doğru (Correct) ve yanlış (Incorrect) sınıflandırma oranlarını gösterir.

Renkler:

Mavi: Doğru sınıflandırmalar (Correct).

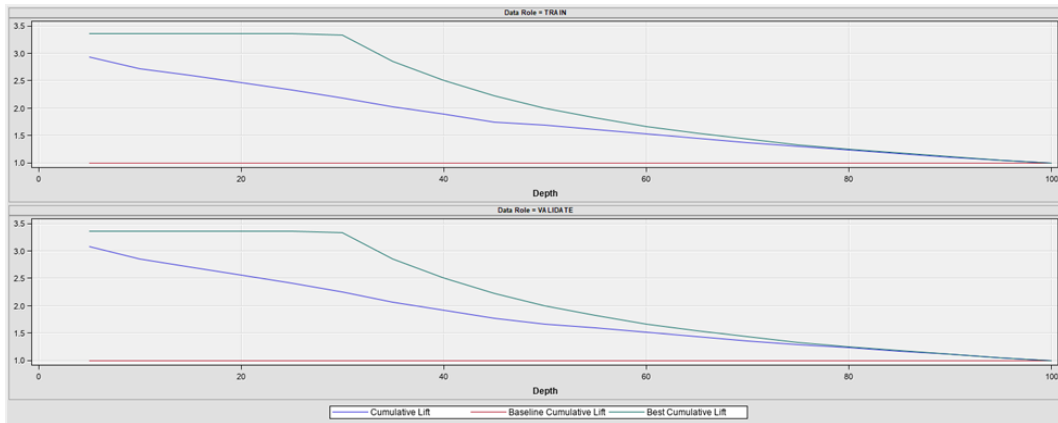
Kırmızı: Yanlış sınıflandırmalar (Incorrect).

Eğitim ve doğrulama veri setlerinde, modelin performansı oldukça benzerdir. Bu, modelin genelleme yapma yeteneğinin iyi olduğunu ve aşırı öğrenme (overfitting) probleminin olmadığını gösterir.

Sınıf 0'da Üstün Performans: Model, hem eğitim hem de doğrulama veri setlerinde sınıf 0'ı tahmin etmede daha başarılıdır.

Sınıf 1'de Düşük Doğruluk: Sınıf 1 için yanlış sınıflandırma oranı, sınıf 0'a kıyasla daha yüksektir. Bu, modelin pozitif sınıfları tahmin etmede biraz daha iyileştirilmesi gerektiğini gösterir.

Hava yolu firmaları için kümülatif kazanç ve temel kümülatif kazanç karşılaştırması grafiği Şekil 3.19'da yer almaktadır.



Şekil 3.19 Lojistik Regresyon havayolu - kümülatif kazanç ve temel kümülatif kazanç karşılaştırması

Şekil 3.19, eğitim (TRAIN) ve doğrulama (VALIDATE) veri setlerinde Kümülatif Kaldıraç (Cumulative Lift) değerlerini derinlik (Depth) ile ilişkilendirerek göstermektedir. Grafikler, modelin belirli bir derinlik seviyesindeki performansını

değerlendirmek amacıyla kullanılır. Cumulative Lift, modelin bir segmentteki olayları (örneğin, belirli bir hedef sınıfı) rastgele tahminlere kıyasla ne kadar iyi tahmin ettiğini ifade eder.

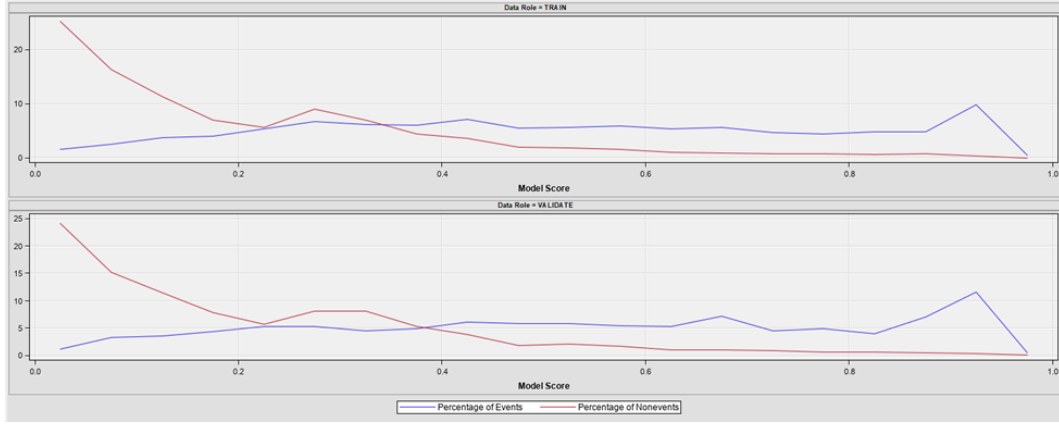
Eğitim Veri Seti (Data Role = TRAIN) Başlangıç (0-20 Derinlik Aralığı) için kümülatif kazanç değeri yaklaşık 3,5 civarında başlamakta, bu da modelin hedef sınıfları rastgele tahminlere kıyasla 3,5 kat daha iyi bir performans sergilediğini göstermektedir. Bu yüksek performans, modelin bu derinlik aralığında hedef sınıfları en iyi şekilde ayırt edebildiğini ve tahmin gücünün en yüksek olduğu noktayı temsil eder. Eğitim ve doğrulama veri setleri arasındaki tutarlılık, modelin genelleme yapma yeteneğinin güçlü olduğunu ortaya koymaktadır.

Orta Aşamalar (20-60 Derinlik Aralığı) için kümülatif kazanç değerleri düzenli bir şekilde azalmaktadır. Bu düşüş, derinlik arttıkça modelin tahmin gücünün kademeli olarak azaldığını göstermektedir. Yine de modelin, rastgele tahminlerden daha üstün performans sergilemeye devam ettiği görülmektedir. Düşüş eğrisi eğitim veri setine oldukça benzer bir yapıdadır. Bu durum, modelin doğrulama veri setinde de tahmin performansının benzer bir şekilde azaldığını göstermektedir.

Son Aşamalar (60-100 Derinlik Aralığı) için derinlik 100'e yaklaştıkça, kümülatif kazanç değeri 1'e yaklaşmaktadır. Bu, modelin rastgele bir tahmin sistemi ile aynı seviyeye indiğini ve ayırt edici gücünün bu aşamada azaldığını ifade eder.

Bu grafik, modelin performansının derinlik arttıkça nasıl değiştiğini ve rastgele tahminlere kıyasla ne kadar etkili olduğunu göstermektedir. Eğitim ve doğrulama veri setlerinde tutarlı sonuçlar, modelin genelleme yapma kabiliyetini desteklemektedir. Modelin başlangıç aşamalarındaki yüksek performansı, belirli hedef sınıfları ayırt etmede etkili olduğunu gösterirken, derinlik arttıkça performanstaki azalma modelin tahmin gücünün sınırlı olduğunu ifade eder. Bu nedenle, modelin performansını artırmak için özellikle yüksek derinliklerde daha etkin stratejiler geliştirilebilir.

Hava yolu firmaları için kümülatif yüzdelik dilimi grafiği Şekil 3.20'de yer almaktadır.



Şekil 3.20 Lojistik Regresyon havayolu - kümülatif yüzdelik dilimi grafiği

Şekil 3.20, modelin tahmin performansını değerlendirmek için eğitim (TRAIN) ve doğrulama (VALIDATE) veri setlerindeki Olayların Yüzdesi (Percentage of Events) ve Olay Olmayanların Yüzdesi (Percentage of Nonevents) dağılımlarını Model Skoru (Model Score) ekseninde göstermektedir.

Mavi çizgi, model skorunun artışıyla birlikte olayların yüzdesindeki değişimi temsil etmektedir. Model skorunun düşük olduğu alanlarda olay yüzdesinin daha yüksek olduğu, skor arttıkça olay yüzdesinin azaldığı gözlemlenmektedir. Bu durum, modelin düşük skorlu kayıtları doğru bir şekilde olay olarak tahmin etme kabiliyetini göstermektedir.

Kırmızı çizgi, olay olmayanların yüzdesini temsil eder. Düşük model skorlarında olay olmayanların yüzdesi düşüktür; ancak skor arttıkça bu yüzdede artış gözlenmektedir. Bu, modelin yüksek skorlu kayıtları olay olmayan olarak tahmin etme kabiliyetini yansıtmaktadır.

Eğitim veri setinde, olaylar ve olay olmayanlar arasındaki ayrımın belirgin olduğu, modelin iyi bir şekilde sınıflandırma yaptığı görülmektedir.

Doğrulama veri setinde, modelin eğitim veri setine benzer bir şekilde düşük skorlu kayıtlar için olay yüzdesini yüksek tuttuğu görülmektedir. Ancak, mavi çizginin eğimi, eğitim veri setindeki kadar keskin değildir. Bu durum, modelin doğrulama veri setindeki performansının eğitim veri setine kıyasla biraz daha düşük olduğunu gösterebilir.

Doğrulama veri setinde, olay olmayanların yüzdesi düşük model skorlarında düşük, yüksek model skorlarında ise yüksektir. Bu, modelin doğrulama veri setinde de olay olmayanları ayırt edebildiğini göstermektedir.

Doğrulama veri setinde de olaylar ve olay olmayanlar arasında ayırım yapılabilmekte, ancak eğitim veri setindeki kadar keskin bir ayırım gözlenmemektedir.

Grafikler, modelin hem eğitim hem de doğrulama veri setlerinde olaylar (pozitif sınıf) ve olay olmayanlar (negatif sınıf) arasında ayırım yapabildiğini göstermektedir.

Model skoruna bağlı olarak olay ve olay olmayan yüzdelerindeki değişim, modelin sınıflandırma performansının tutarlı olduğunu göstermektedir. Doğrulama veri setindeki daha az keskin eğim, modelin doğrulama verisine olan genellenebilirliğinin bir miktar daha düşük olduğunu gösterebilir, ancak performans yine de kabul edilebilir seviyededir.

3.7.2 Tedarikçi Firmalar İçin Lojistik Regresyon Analizi Uygulaması

İnteraktif gruplanmış değişkenler ve tedarikçi firmalar için kurulan Lojistik Regresyon modeli sonuçları aşağıda, Şekil 3.21’de, değerlendirilmiştir.

Tablo 3.21 Lojistik Regresyon tedarikçi - eğitim ve doğrulama veri setleri için uyum istatistikleri

Uyum İstatistiği	İstatistik Etiketi	Eğitim	Doğrulama
MISC	Yanlış Sınıflandırma Oranı	0,196	0,191

Şekil 3.21, modelin eğitim (Train) ve doğrulama (Validation) veri setlerinde performansını değerlendiren MCR istatistiğini göstermektedir.

Eğitim (Train) veri setinde Yanlış Sınıflandırma Oranı (MISC - Misclassification Rate) 0,196’dır. Bu, modelin eğitim verisindeki gözlemlerin 0,196’sını yanlış sınıflandırdığını ifade eder. Doğrulama (Validation) veri setinde Yanlış Sınıflandırma Oranı (MISC - Misclassification Rate) ise 0,191 olarak hesaplanmıştır. Eğitim verisine yakın bir orana sahip olması, modelin genelleme yeteneğini koruduğunu ve aşırı öğrenmeden kaçındığını gösterir.

Eğitim ve doğrulama veri setleri için MISC değerleri oldukça yakındır. Bu, modelin hem eğitim hem de doğrulama verisi üzerinde tutarlı performans gösterdiğini işaret etmektedir.

Forward Stepwise Lojistik Regresyonda modele giren gruplanmış değişkenlerin sıralaması Şekil 3.22’de yer almaktadır.

Tablo 3.22 Lojistik Regresyon tedarikçi - adım adım etki giriş istatistikleri

Adım	Girilen Etki	Etki Sayısı	P Değeri	AIC
0	Kesim (Intercept)	1	NaN	6224,015
1	Part_Description	2	0	4219,603
2	Labor_Billed_Current	3	6,77E-80	3834,053
3	Parts_Billed_Current	4	1,58E-25	3721,365
4	Diff_Shipping_Entry	5	0,402	3722,665
5	Diff_Quote_Entry	6	0,831	3724,620

Şekil 3.22, bir model geliştirme sürecinde adım adım eklenen etkilerin (değişkenlerin) istatistiksel anlamlılık ve model uyumu üzerindeki etkilerini göstermektedir.

Intercept (Kesim Noktası) için AIC (Akaike Bilgi Kriteri) 6224,014'tür. Bu, yalnızca kesim noktasını içeren modelin temel AIC değeridir. Bu adım başlangıç noktası olarak değerlendirilir.

GRP Part Description (Gruplanmış Parça Tanımı) değişkeni için p değeri 0'dır. Bu, değişkenin modelde yüksek istatistiksel anlamlılığa sahip olduğunu ve modele eklenmesinin anlamlı olduğunu gösterir, AIC değeri 4219,603. Kesim noktasına kıyasla bu değişkenin eklenmesiyle model daha iyi hale gelmiştir, AIC değerinde ciddi bir düşüş görülmektedir.

GRP stnd Labor Billed Current (Gruplanmış İşçilik Maliyeti) değişkeni için p değeri $6,77 \times 10^{-80}$ (son derece düşük). Bu durum, bu değişkenin modelde oldukça anlamlı olduğunu gösterir. AIC değeri 3834,053 ile modelin uyumu, bu değişkenin eklenmesiyle daha da artmıştır, AIC azalmıştır.

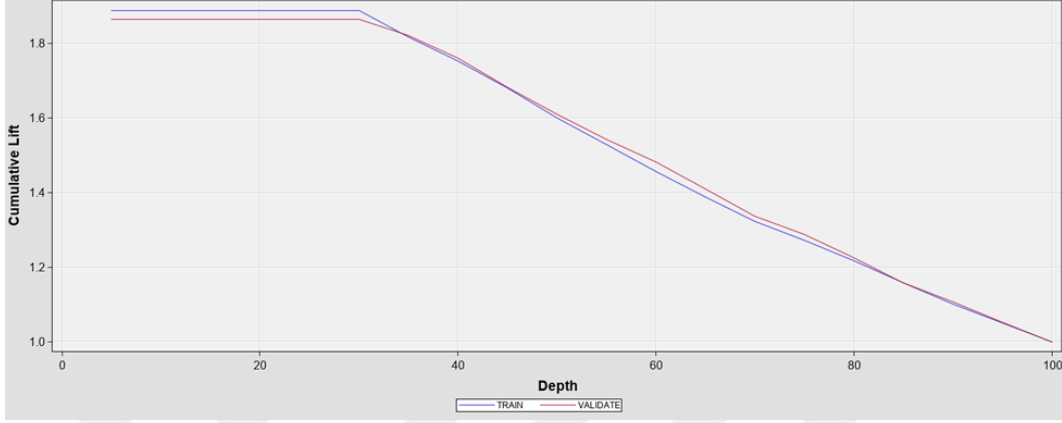
GRP stnd Parts Billed Current (Gruplanmış Parça Maliyeti) değişkeni için p değeri $1,58 \times 10^{-25}$. Bu değişken de oldukça anlamlıdır. AIC değeri 3721,365. Bu değişkenin de modele girişi ile AIC değerinde bir iyileşme görülmektedir.

GRP stnd Diff Shipping Entry (Gruplanmış Çevrim Süresi) değişkeni için p değeri 0,402'dir. Bu değerin yüksek olması, bu değişkenin modelde istatistiksel olarak anlamlı olmadığını gösterir. AIC değeri 3722,665 bir önceki adıma göre artmıştır.

GRP stnd Diff Quote Entry (Gruplanmış Fiyat Teklif Süresi) değişkeni için p değeri 0,8314'tür. Bu değerin oldukça yüksek olması, bu değişkenin modelde anlamlı olmadığını göstermektedir. AIC değeri 3724,62 artması, bu değişkenin model performansını olumsuz etkilediğini işaret eder.

Bu analiz, modelin doğruluğunu artırmak için hangi değişkenlerin önemli olduğunu ve hangi değişkenlerin modelden çıkarılması gerektiğini anlamak açısından kritik bilgiler sunmaktadır.

Tedarikçi firmalar için kümülatif kazanç grafiği Şekil 3.21’de yer almaktadır.



Şekil 3.21 Lojistik Regresyon tedarikçi - kümülatif kazanç grafiği

Şekil 3.21, modelin performansını değerlendirmek için kullanılan önemli bir araçtır.

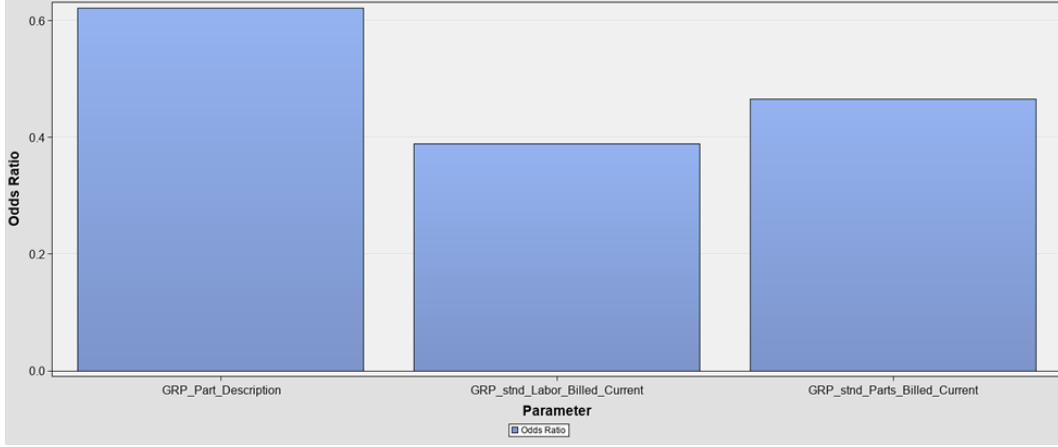
Başlangıç Noktasında (0-20 Derinlik Aralığı) hem eğitim hem de doğrulama veri setlerinde Kümülatif Lift (Cumulative Lift) değerleri yaklaşık olarak 1,9 civarındadır. Bu, modelin hedef sınıfı etkili bir şekilde ayırt edebildiğini ve tahmin performansının yüksek olduğunu göstermektedir. Eğitim ve doğrulama veri setleri arasında kümülatif lift değerleri açısından önemli bir fark bulunmamaktadır, bu da modelin genelleme yeteneğinin iyi olduğunu ve aşırı öğrenme (overfitting) olmadığını işaret eder.

Derinlik Arttıkça (20-50 Aralığı), her iki veri seti için de Kümülatif Kazanç (Cumulative Lift) değerlerinin kademeli olarak azaldığını göstermektedir. Bu azalma, modelin performansının derinlik arttıkça yavaş yavaş düştüğünü, yani hedef sınıfları ayırt etme gücünün azaldığını ifade eder.

50-100 Derinlik Aralığında Kümülatif Kazanç (Cumulative Lift) değeri hem eğitim hem de doğrulama veri setlerinde yaklaşık olarak 1,0’a yaklaşmaktadır. Bu durum, modelin bu noktadan sonra hedef sınıfları ayırt etme yeteneğinin ortalama seviyeye indiğini ve modelin ayrıştırma gücünün düşük olduğunu göstermektedir.

Eğitim ve doğrulama veri setleri arasındaki çizgilerin paralel olması ve birbirine yakın ilerlemesi, modelin tutarlılığını ve doğrulama performansının güvenilir olduğunu ortaya koymaktadır.

Tedarikçi firmalar için Olasılık Oranı (Odds Ratio) Grafiği Şekil 3.22’de yer almaktadır.



Şekil 3.22 Lojistik Regresyon tedarikçi - olasılık oranı grafiği

Şekil 3.22, çeşitli parametrelerin Odds Ratio (Olasılık Oranı) değerleri karşılaştırılmaktadır. Grafikte GRP Part Description (Parça Tanımı), GRP stnd Labor Billed Current (İşçilik Maliyeti), GRP stnd Parts Billed Current (Parça Maliyeti) değişkenleri bulunmaktadır. Bu parametrelerin, belirli bir olayın meydana gelme olasılığı üzerindeki etkisi Olasılık Oranı (Odds Ratio) değerleri aracılığıyla değerlendirilir.

GRP Part Description (Parça Tanımı) değişkeni için en yüksek olasılık oranı değeri yaklaşık 0,6 olarak görülmektedir. Bu, Parça Tanımı değişkeninin, diğer parametrelere kıyasla, incelenen olay üzerinde daha güçlü bir etkiye sahip olduğunu gösterir. Ancak, olasılık oranı değerinin 1’den küçük olması, bu parametrenin olayın meydana gelme olasılığını azalttığını gösterir.

GRP stnd Labor Billed Current (İşçilik Maliyeti) değişkeni için bu parametrenin Odds Ratio değeri yaklaşık 0,4 civarındadır. Bu değer, işçilik maliyetinin olay üzerinde daha düşük bir etkiye sahip olduğunu ve olayın meydana gelme olasılığını azaltma yönünde bir etki gösterdiğini belirtir.

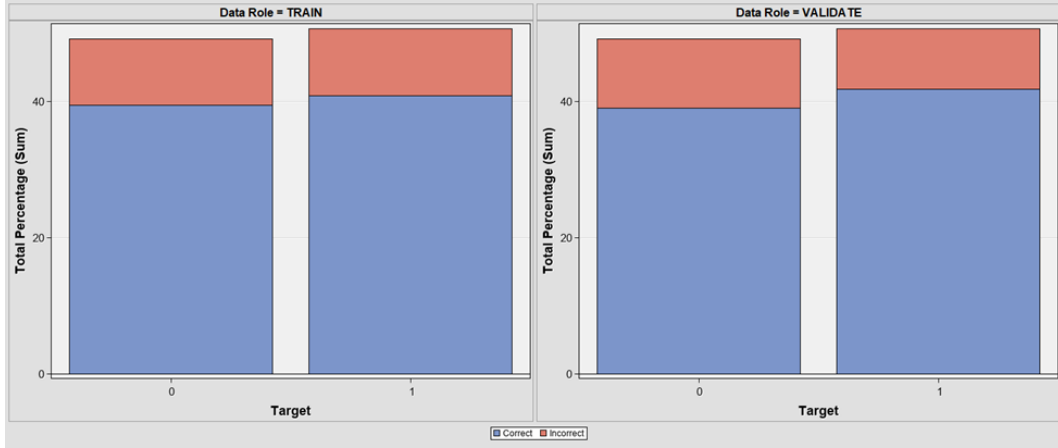
GRP stnd Parts Billed Current (Parça Maliyeti) için bu parametre de yaklaşık 0,45 seviyesinde bir olasılık oranı değerine sahiptir. İşçilik maliyeti ile benzer bir etkiye sahip olduğu, ancak parça tanımına göre olay üzerindeki etkisinin daha düşük olduğu görülmektedir.

Grafikte görülen tüm parametreler için olasılık oranı değeri 1’in altındadır. Bu durum, tüm bu parametrelerin olayın meydana gelme olasılığını azaltıcı bir etkiye sahip olduğunu gösterir. En güçlü etkiye sahip olan parametre Parça Tanımı iken,

en düşük etkiye sahip olan parametre İşçilik Maliyeti olarak gözlemlenmiştir.

Bu tür bir analiz, modelin hangi bağımsız değişkenlerin sonuca olan etkisinin daha ayrıntılı bir şekilde anlaşılmasına.

Tedarikçi firmalar için yanlış sınıflandırma oranları grafiği Şekil 3.23'te yer almaktadır



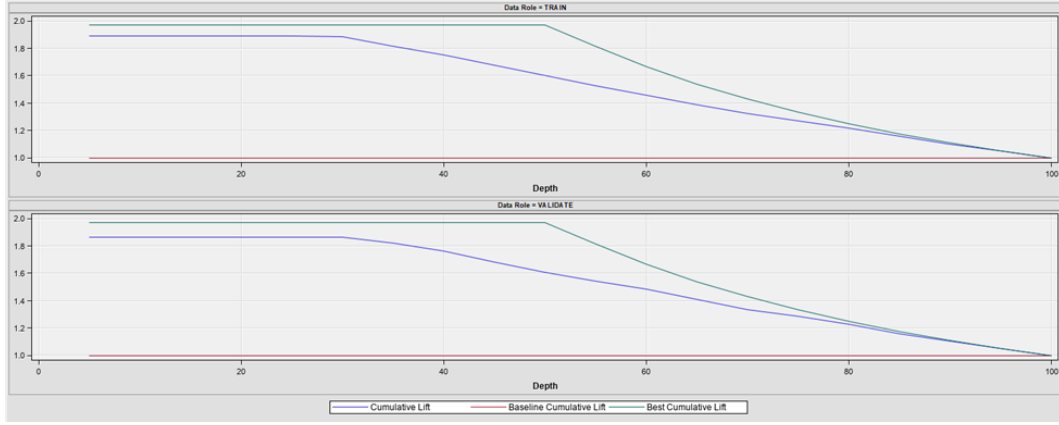
Şekil 3.23 Lojistik Regresyon tedarikçi - yanlış sınıflandırma oranları grafiği

Şekil 3.23, modelin eğitim (TRAIN) ve doğrulama (VALIDATE) veri setlerindeki doğru ve yanlış sınıflandırma oranlarını yüzdeler olarak karşılaştırmaktadır.

Eğitim ve doğrulama veri setlerindeki sınıflandırma oranlarının birbirine yakın olması, modelin tutarlılığını ve genelleme kapasitesini ortaya koymaktadır. Hedef sınıflar arasında bir denge olduğu ve modelin aşırı öğrenme (overfitting) sorununa eğilim göstermediği gözlemlenmektedir.

Modelin hem eğitimde hem de doğrulamada istikrarlı bir performans sergilediğini ve hedef sınıfları ayırıştırma kabiliyetinin güvenilir olduğunu göstermektedir. Bu analiz, modelin genel doğruluğunu ve uygulanabilirliğini teyit eden önemli bir göstergedir.

Tedarikçi firmalar için Kümülatif Kazanç ve Temel Kümülatif Kazanç Karşılaştırması Grafiği Şekil 3.24'te yer almaktadır



Şekil 3.24 Lojistik Regresyon tedarikçi - kümülatif kazanç ve temel kümülatif kazanç karşılaştırması

Şekil 3.24, Kümülatif Kazanç (Cumulative Lift) değerlerinin Depth (Derinlik) ile olan ilişkisini eğitim (TRAIN) ve doğrulama (VALIDATE) veri setleri için göstermektedir. İki grafikte de üç çizgi bulunmaktadır.

Mavi Çizgi (Cumulative Lift): Modelin belirli bir derinlikteki performansını gösterir. Daha yüksek kümülatif kazanç değerleri, modelin hedef sınıfları etkili bir şekilde ayırt ettiğini gösterir.

Kırmızı Çizgi (Baseline Cumulative Lift): Rastgele bir seçim yapıldığında elde edilecek kümülatif kazanç değerini temsil eder ve 1’de sabitlenmiştir. Bu çizgi, model performansının karşılaştırılması için bir referans noktası olarak kullanılır.

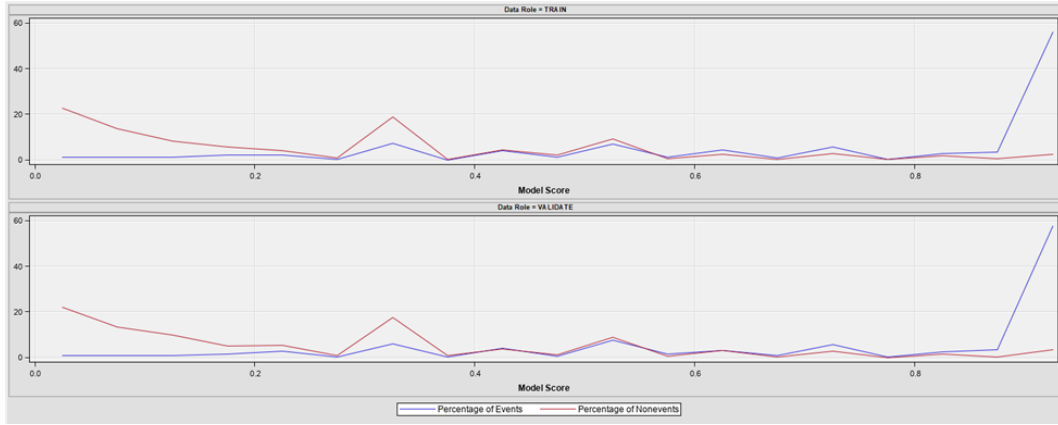
Yeşil Çizgi (Best Cumulative Lift): Modelin teorik olarak ulaşabileceği en iyi kümülatif kazanç değerini temsil eder.

Her iki veri setinde de model, ilk derinliklerde (0-20 aralığında) yüksek bir kümülatif kazanç değeri göstermektedir. Yaklaşık 2,0 seviyesinde olan bu değer, modelin en güçlü ayırt etme yeteneğine sahip olduğu derinlik aralığını işaret eder.

20 - 100 Derinlik Aralığı arası derinlik arttıkça kümülatif kazanç değerlerinde belirgin bir azalma gözlenmektedir. Her iki veri setinde de 100’e yaklaştığında, kümülatif kazanç değeri 1’e yaklaşmaktadır ve bu, modelin ayırt edici gücünün neredeyse rastgele bir seçime eşdeğer hale geldiğini gösterir.

Eğitim (mavi çizgi) ve doğrulama (kırmızı çizgi) veri setleri için kümülatif kazanç eğrilerinin birbirine oldukça yakın olduğu görülmektedir. Bu benzerlik, modelin eğitim setinde öğrendiği bilgiyi doğrulama setinde de başarılı bir şekilde genelleylebildiğini, yani aşırı uyum (overfitting) yapmadığını gösterir.

Tedarikçi firmalar için kümülatif yüzdelik dilimi grafiği Şekil 3.25'te yer almaktadır



Şekil 3.25 Lojistik Regresyon tedarikçi - kümülatif yüzdelik dilimi grafiği

Şekil 3.25, modelin tahmin skorlarına göre olay ve olay dışı durumların yüzdelerini eğitim (TRAIN) ve doğrulama (VALIDATE) veri setleri üzerinde karşılaştırmaktadır.

Olay Yüzdesi (Mavi Çizgi), model skoru 0,9 ile 1,0 arasında olan alanlarda olay yüzdesinin belirgin şekilde arttığı gözlenmektedir. Bu artış, modelin yüksek skor verdiği durumlarda olay gerçekleşme olasılığının daha yüksek olduğunu göstermektedir. Eğitim ve doğrulama veri setlerinde, yüksek skor aralıklarında olay yüzdelerinin benzer bir artış gösterdiği ve birbirine paralel ilerlediği görülmektedir.

Olay Dışı Durum Yüzdesi (Kırmızı Çizgi), düşük skor aralıklarında (özellikle 0.0-0.3) olay dışı durumların yüzdesi oldukça yüksektir ve daha sonra kademeli olarak azalmaktadır. Bu da modelin düşük skor verdiği durumlarda olayın gerçekleşmeme olasılığının yüksek olduğunu işaret eder. Bu eğilim hem eğitim hem de doğrulama setinde benzer bir şekildedir.

Her iki veri setinde de olay ve olay dışı durum yüzdeleri benzer bir trend göstermektedir. Özellikle yüksek skorlarda olay yüzdesinin artması ve düşük skorlarda olay dışı yüzdesinin yüksek olması, modelin farklı veri setlerinde benzer şekilde performans gösterdiğini ve aşırı uyum yapmadığını göstermektedir.

4 SONUÇ

Veri analitiği, havacılık sektöründe, özellikle KOBİ düzeyindeki firmalar için yeni bir alan olmasına rağmen, MRO (Bakım, Onarım ve Revizyon) şirketlerinin verilerini daha planlı, etkili, öngörülebilir ve kârlı bir şekilde kullanmalarını sağlama açısından giderek daha kritik bir hale gelmektedir. MRO şirketlerinin sahip olduğu teknik bilgiler, uçak performans verileri, arıza detayları ve arızaların nedenleri gibi veriler büyük ölçüde uçak üreticileri, OEM'ler (Orijinal Ekipman Üreticileri) ve hava yolu firmalarının elindedir ve bu verilerin standartlaştırılması hâlâ devam etmektedir.

Bu çalışmada gerçek dünya verileri kullanılmıştır. Elde edilen operasyonel verilerin ham hali modelleme için uygun olmadığından, veri hazırlama süreçleri (veri temizleme, yeni değişkenler türetme, veri dönüştürme, veri seçimi vb.) titizlikle uygulanmıştır. Bu zorlu süreçler, Bölüm 2.1'de detaylı olarak tartışılmıştır. Yüksek veri kalitesine sahip değişkenler elde etmek amacıyla değişkenler incelenmiş; veri setinden hata ve tutarsızlıklar ayıklanmıştır. Ayrıca, sektör profesyonelleri ile yapılan görüşmelerin ardından kâr marjını açıklamakta etkili olabileceği düşünülen bazı yeni değişkenler türetilmiştir.

Gerçek dünya verilerine erişimin zorluğu nedeniyle, bu çalışmanın sonuçları, KOBİ düzeyindeki MRO şirketinin parça bakım segmentine özgü nitelikler taşımaktadır. Çalışma, 2013-2021 yıllarını kapsayan veriler kullanılarak gerçekleştirilmiştir. Şirketin organizasyon yapısı, pazardaki konumu ve özel operasyonel zorluklar gibi benzersiz faktörler, elde edilen bulgular üzerinde belirleyici bir etkiye sahip olabilir. Bu nedenle, bulgular şirket için değerli içgörüler sunmakla birlikte, MRO sektörünün geneline uygulanırken dikkatli bir değerlendirme yapılması gerekmektedir.

Değişken dönüşümü için İnteraktif Gruplama yöntemi kullanılmıştır. İnteraktif Gruplama, sürekli değişkenlerdeki aykırı değer sorununu da ele alarak aşırı değerlerin etkisinin en aza indirilmesini ve daha sağlam ve güvenilir modeller elde

edilmesini sağlamıştır.

Variable Clustering (Değişken Kümeleme) ile modele giren değişkenler düzeyinde kümeler oluşturulmuş ve aynı küme içindeki ve ayrı kümelerdeki değişkenler Şekil 3.4'te analiz edilmiştir. Değişken kümeleri, verilerin daha iyi anlaşılmasını sağlayarak, Gradient Boosting ve Lojistik Regresyon modellerine dayanan stratejik çıkarımlarda yorum gücünü artırmıştır.

Bölüm 3.5'te detaylandırıldığı üzere, Gradient Boosting uygulamasında hiperparametre ayarlamaları optimizasyon yöntemleri kullanılmadan manuel olarak yapılmıştır [30]. Eğitim ve doğrulama veri setlerinde benzer ve minimum yanlış sınıflandırma oranları hedeflenerek çeşitli hiperparametre kombinasyonları için birçok model geliştirilmiştir.

Bu çalışmada uygulanan yöntemler içinde Gradient Boosting'e yer verilmesinin nedeni, sınıflandırma görevlerinde kanıtlanmış etkinliğidir [22, 29].

Gradient Boosting modelinin karmaşıklığı ile yorumlanabilirliği arasındaki dengeyi sağlamak için değişken önem değerleri Bölüm 3.5.1 ve Bölüm 3.5.2'de tartışılmıştır.

Çalışmada Boost79 ve Boost86 olmak üzere iki Gradient Boosting algoritması karşılaştırılmıştır ve detaylar Bölüm 3.5'te açıklanmıştır. Gruplandırılmış değişkenleri kullanan Boost86 Modeli, standartlaştırılmış değişkenleri kullanan Boost79 Modeli ile karşılaştırıldığında, daha düşük bir yanlış sınıflandırma oranı ve daha yüksek Gini indeksi ile daha iyi performans göstermiştir.

Alternatif bir yöntem olarak uygulanan Lojistik Regresyon modeli, Gradient Boosting ile karşılaştırma amaçlı kurulmuştur. Bu karşılaştırma, her iki modelin performansını değerlendirerek, hangi modelin hedef değişkeni daha iyi açıkladığını belirlemek için gerçekleştirilmiştir. Lojistik regresyonun, açıklayıcı yapısı ve değişkenlerin etkilerini doğrudan yorumlama imkânı, model sonuçlarının anlaşılabilirliğini artırmıştır. Öte yandan, Gradient Boosting modeli, genellikle doğruluk oranı açısından üstün performans sergileyen bir yöntem olup, karmaşık ilişkileri öğrenme kabiliyeti ile öne çıkmaktadır. Bu çalışmada, her iki modelin de yanlış sınıflandırma oranı (MCR), ROC eğrisi ve Gini katsayısı gibi istatistiklerle değerlendirilmesi sağlanmıştır. Gradient Boosting modeli için yanlış sınıflandırma oranı (MCR) 0,164 iken, Lojistik Regresyon modeli için 0,197'dir. Bu kapsamda, Lojistik Regresyon modeli, daha basit bir yapı sunarak karşılaştırma için bir referans modeli olarak görev yapmıştır. Bu analiz, farklı modelleme yaklaşımlarının avantaj ve sınırlamalarını daha iyi anlamayı mümkün kılmıştır. SME MRO firmasının tüm müşterileri için geliştirilen Gradient Boosting ve Lojistik Regresyon

modellerine giren deęişkenlerin ise farklılaştığı görölmektedir.

Tüm veri seti için kurulan Gradient Boosting ve Lojistik Regresyon modelleri yanı sıra, SME MRO firmasına ait müşteriler, havayolu firmaları ve tedarikçi firmalar olmak üzere iki gruba ayrılmış ve bu iki müşteri grubu için ayrı ayrı Lojistik Regresyon modelleri kurulmuştur. Bu ayırım, farklı müşteri grupları için kâr marjını etkileyen faktörlerdeki farklılıkların ortaya konulmasını sağlamak amacıyla yapılmıştır. Her bir müşteri grubu için oluşturulan modeller, kâr marjını etkileyen deęişkenlerin hangi düzeyde ve nasıl farklılaştığını belirlemek için analiz edilmiştir.

Çalışma, Gradient Boosting ve Lojistik Regresyon modellerinin çıktılarına dayanarak, kârlılığın yalnızca Parça Maliyeti ve İşgücü Maliyeti gibi doğrudan deęişkenlerle deęil, aynı zamanda Fiyat Teklif Süresi (TAT), Çevrim Süresi (TTQ), İlgili Departman, İlgili Yönetici ve Parça Durumu gibi dolaylı deęişkenlerle de ilişkili olduğunu ortaya koymaktadır. Standart bakış açısında kârlılık, doğrudan maliyetlerin (işgücü ve malzemeler) fiyattan çıkarılmasıyla matematiksel olarak hesaplanır ve şirketler bu dar felsefede fiyatları artırmaya ve maliyetleri düşürmeye odaklanır. Ancak, COVID-19 sonrası tedarik zinciri zorlukları ve işgücü yetersizlikleri, MRO şirketlerini, özellikle de KOBİ düzeyindeki MRO'ları, kârlılığı artırmak ve rakiplerinden ayrılmak için alternatif alanlarda benzersiz stratejiler geliştirmeye teşvik etmektedir.

Çalışma, benzersiz ve özel parça kodunu ifade eden Parça Tanımı deęişkenine dikkat çekmektedir. Uçak sistemlerinde binlerce parça bulunmaktadır ve her uçak modeli ve tipi farklı bakım prosedürleri, özel ekipman ve araçlar, malzemeler ve özel iş gücü nitelikleri gerektiren farklı parçalar kullanır. MRO kapasite listesinde çok sayıda parça numarasına sahip olmak, farklı araçlar, ekipmanlar, makineler, malzemeler ve işgücünü yönetmeyi gerektirir. Bu durum ise yönetim için büyük bir çaba ve maliyet yükü oluşturur. En uygun ve en kârlı parça numarasını bulmanın, kârlılık modelindeki en önemli deęişken olduğunu kanıtlar. Gradient Boosting ve Lojistik Regresyon modeline göre kârlılığı en iyi açıklayan deęişken Parça Tanımıdır. İnteraktif Gruplama sonrası hedef deęişken ile ilişkisi incelendiğinde de en yüksek Gini indeksine, 65,246, sahip deęişken Parça Tanımıdır. Dolayısıyla, geliştirilen her iki model önemli stratejik iş sonuçları vermektedir. MRO'lar, parça numarası kapasite listelerini belirlemeye odaklanmalı ve servis vermeyi seçtikleri parça numaralarına öncelik vermelidir.

Modeller, Parça Maliyeti ve İşgücü Maliyeti gibi kâr marjıyla doğrudan ilgili deęişkenlerin kârlılığı önemli ölçüde etkilediğini doğrulamaktadır. Malzeme maliyetleri azaltıldığında ve daha standart bir iş sağlandığında, yani aynı

parçalardan daha fazla sayıda iş sağlandığı sürece, doğrudan işgücü maliyetleri azalacak ve kârlılığa olumlu katkı sağlanacaktır. Parça Maliyeti değişkeni 49,050 Gini indeksi ve İşçilik Maliyeti değişkeni 42,664 Gini indeksi ile bağımlı değişkenle ilişkisi sırasıyla ikinci ve üçüncü en yüksek olan değişkenlerdir. Parça Maliyeti Değişkeni ve İşçilik Maliyeti değişkenleri, gruplanmış değişkenler için geliştirilen Gradient Boosting modeline ikinci ve üçüncü önem sıraları ile girmişlerdir. Aynı değişkenler Lojistik Regresyon modeline sırasıyla dördüncü ve üçüncü sırada dahil olmuşlardır. Yani Lojistik Regresyon modelinde işçilik maliyeti, parça maliyetinden daha fazla öneme sahiptir.

Bu çalışmanın en önemli sonuçları, kârlılığı etkileyen gizli faktörlerin tanımlanması ve doğrulanmasıdır.

Parça Durumu: Onarım sürecindeki parçaların durumu, kârlılığı etkiler. MRO'lar, onarım sürecinin tamamlanmasının ardından parçalar için servise sunma sertifikası düzenler ve faturayı keser. Bu süreç, müşteriye sevkiyat ile birlikte kâr elde edilmesini sağlar. Standartlaştırılmış değişkenlerin kullanıldığı Boost79 Modeli'nde, Parça Durumu değişkeninin doğrulama önemi 0,272'dir (Tablo 3.14 ve Şekil 3.6) ve sıralamada son üç sırada yer almaktadır. Buna karşılık, İnteraktif Gruplama yöntemi kullanılan Boost86 Modeli'nde bu değişken, doğrulama önemi 0,309 ile daha yüksek bir sırada yer alır Tablo 3.16 ve Şekil 3.10. Lojistik Regresyon modeline Parça Durumu değişkeni ikinci önem düzeyinde dahil olmuştur (Şekil 3.18. İşlerin zamanında tamamlanması, daha yüksek bir devir oranına ve etkili zaman yönetimine olanak tanır; bu durum, çalışan verimliliğini artırır ve iş yönetimine olumlu katkı sağlar. Bunun aksine, eksik görevler ve atölye sistemindeki uzun süreli işlerin tamamlanmamış durumu, kârlılığı azaltır ve genel performansı olumsuz etkiler.

Çevrim Süresi (TAT): Boost79 Modelinde Çevrim Süresi değişkeninin doğrulama önemi 0,252 iken, Boost86 Modelinde bu değer 0,140'a düşmektedir. Boost86 modeli, daha hızlı tamamlanan işlerin daha fazla kâr sağladığını, ancak daha uzun süren işlerin daha az kârlı olduğunu göstermektedir. Bu değişken, iş gücü ve malzeme kaynaklarını optimize ederek zaman yönetimini iyileştirmesi nedeniyle, müşteriler ve MRO şirketleri açısından kritik önemini korumaktadır. Çevrim Süresi değişkeni Lojistik Regresyon modeline girmemiştir.

Fiyat Teklif Süresi (TTQ): Bu değişken, atölyeye girişten incelemenin tamamlanmasına kadar geçen süreyi ölçer ve TTQ (Time to Quote) olarak adlandırılır. İnceleme tamamlandıktan sonra MRO, onarım için gerekli malzeme ve iş gücünü belirler. Bu faktörler, kâr marjını da içerecek şekilde bir teklif

fiyatına dönüştürülür ve müşterinin onayına sunulur. Parçalar hızlı bir şekilde incelendiğinde, MRO şirketi teklif fiyatını daha kısa sürede hesaplayabilir ve ilgili parça numarası için öngörülen kâr marjını belirleyebilir. Gruplanmış değişkenler için kurulan Gradient Boosting modeline giren TTQ değişkeni, Lojistik Regresyon modelinde yer almamaktadır.

Bunun yanında, Variable Clustering (Değişken Kümeleme) analizine dayanarak, Çevrim Süresi (TAT) ve Fiyat Teklif Süresi (TTQ) değişkenleri aynı sınıfta gruplanmışlardır Şekil 3.5. Bu sınıf, operasyonel süreleri göz önünde bulunduran sınıf olarak nitelendirilebilir.

İlgili Departman: Bu değişken, Variable Clustering (Değişken Kümeleme) sınıflamasına göre Parça Tanımı ve Yönetici değişkenleri ile aynı grupta yer almakta olup, bu durum Şekil 3.4 ve Şekil 3.5 ile gösterilmektedir. İlgili Departman değişkeni Boost 86 modeline 8. önem derecesi ile son giren değişken olmuştur. Bu bulgular, önem düzeyi çok yüksek olmasa da, bazı atölye birimlerinin daha standart hale geldiğini ve daha spesifik süreçlere sahip olduğunu, bu durumun daha kârlı bir işletme modeli oluşturduğunu ortaya koymaktadır. İlgili Departman değişkeni Lojistik Regresyon modeline ise dahil olmamıştır.

Yönetici: Her departman, kendi spesifik parçaları için bakım hizmeti sağladığından, departman ve lideri kârlılık üzerinde kritik bir rol oynayabilir; bu faktör çalışmada ele alınan modellere dayanmaktadır[31]. Gruplanmış değişkenler ile geliştirilen Gradient Boosting modeli çıktıları, liderin planlama, süreci yönetme, görevleri tamamlama ve parçaları hizmete sunma gibi performansının kârlılıkla ilişkili olduğunu göstermektedir. Doğrulama ROC değeri 0,911 olan Boost86 Modeli, 0,377 doğrulama önemi ile yöneticinin güçlü etkisini gösterirken, Lojistik Regresyon modelinde ise Yönetici değişkeni modele girmemiştir.

Tüm müşteriler için geliştirilen Gradient Boosting ve Lojistik Regresyon modelleri yanında, KOBİ MRO firmasının müşterileri havayolu firmaları ve tedarikçi firmalar olarak iki ayrı gruba ayrılarak uygulanan Lojistik Regresyon modeli sonuçları da incelenmiştir. Havayolu Firmaları için Parça Tanımı, Parça Maliyeti, İşçilik Maliyeti, Fiyat Teklif Süresi (TTQ), Yönetici ve Çevrim Süresi(TAT) değişkenleri modele girmiştir. Tedarikçi firmalar için kurulan Lojistik Regresyon modeline ise sadece Parça Tanımı, İşçilik Maliyeti ve Parça maliyeti değişkenleri girmiştir.

Havayolu firmaları ve tedarikçi firmalar için kurulan Lojistik Regresyon modeli sonuçları incelendiğinde, havayolu firmaları gibi kurumsal firmalardan sağlanan kar marjını, Fiyat Teklifi Süresi (TTQ) ve Çevrim Süresi (TAT) gibi operasyonel süreçleri açıklayan değişkenlerin de etkilediği gözlemlenmektedir. SME MRO

firması, tedarikçi firmalardan daha yüksek kar marjı elde edebilmektedir. Operasyonel süreçlerindeki aksaklıklar nedeniyle, daha kurumsal olan havayolu firmalarına hızlı hizmet veremediği için karlılıkta rekabet dezavantajına sahip olmaktadır. Bu nedenle havayolu firmalarından daha az kar elde etmektedir.

Modelleme süreçlerinden bağımsız olarak, veriyi daha iyi anlamak adına modele girmeye hak kazanan değişkenlere Variable Clustering (Değişken Kümeleme) uygulandığında, 3 temel küme oluşturulmuştur. Clus1, Parça Tanımı, Parça Maliyeti, İşçilik Maliyeti, Departman ve Yönetici değişkenlerini içermektedir, Şekil 3.4). Clus1'in tamir parçası ile ilgili bilgileri içerdiği ifade edilebilir. Yüksek kâr marjına sahip parçaların yetenek listesi ile güncellenmesi ve yüksek kâr marjı sağlayan departmanların yüksek performanslı yöneticilerle birleştirilmesi, kârlılığı artırabilir. Ayrıca, Clus1'de yer alan diğer değişkenlerdeki değişimlerin, Clus1'deki parçaların işçilik maliyetini olumlu yönde etkileyebileceği belirtilmektedir.

İkinci küme olan Clus2, Çevrim Süresi(TAT) ve Fiyat Teklif Süresi (TTQ) değişkenlerini içermektedir. Clus2'nin yalnızca iki değişken içermesi, bu süreçlerin Parça Tanımı, Parça Maliyeti, İşçilik Maliyeti, Departman ve Yönetici değişkenlerinden bağımsız olduğu şeklinde yorumlanabilir. Bu durum, parçanın ne olduğundan bağımsız olarak genel süreçlerde iyileştirmeler yapılmasını önermektedir. Bu süreçlerin optimize edilmesi, kar marjı performansını artırabilir.

Bu çalışma değerli içgörüler sunsa da, bazı sınırlamaların olduğunu kabul etmek önemlidir. Öncelikle, kullanılan veri, yalnızca bir tane KOBİ MRO şirketine ait olup, bileşen bakım segmentine özeldir. Bu durum, elde edilen bulguların havacılık sektöründeki diğer segmentlere veya şirketlere genellenebilirliğini sınırlandırabilir. Ayrıca, veri belirli bir zaman aralığını kapsamaktadır. Bu, analiz edilen dinamiklerin uzun vadede değişebileceği veya farklı zaman dilimlerinde farklı sonuçlar verebileceği anlamına gelebilir. Bu sınırlamalar, bulguların yorumlanmasında dikkat edilmesi gereken önemli unsurlardır.

Gelecekteki araştırmaların, birden fazla KOBİ MRO şirketinden ve farklı coğrafi bölgelerden verileri dahil ederek çalışmalarını sürdürmeleri önerilebilir. Havacılık sektöründeki diğer segmentlerin incelenmesi, farklı operasyonel alanlardaki kâr marjlarını etkileyen faktörlerin belirlenmesine yardımcı olabilir. Bu modeller, motor, uçak gövdesi ve hat bakımı gibi diğer MRO segmentlerine de uygulanabilir. Bu çalışmada geliştirilen modeller, daha hassas ve sağlam bir MRO iş tahmin modeli olarak dönüştürülebilir. Bu yaklaşım, modelin tüm MRO endüstrisini kapsayıcılığının değerlendirilmesini sağlayacaktır.

Bu araştırma öncü bir çalışma olup, farklı şirketlerden benzer operasyonel

verilere erişim imkânı olsaydı, daha kapsamlı bir araştırma yapılabilirdi. Bu durum, bulguların genellenebilirliğini ve çalışmanın sektörel kapsamını daha da güçlendirebilirdi.

Bu sınırlamaların ele alınması ve önerilen araştırma yönlerinin takip edilmesiyle, gelecekteki çalışmalar bu bulguların üzerine inşa edebilir ve KOBİ MRO sektöründeki kâr marjı dinamiklerinin daha kapsamlı bir şekilde anlaşılmasına katkıda bulunabilir. Bu yaklaşım, sektör genelinde daha etkili stratejilerin geliştirilmesine ve MRO şirketlerinin operasyonel performansını optimize etmelerine olanak sağlayacaktır.



- [1] V. Deac, G. Cârstea, C. Bâgu F. Pârvu, “The modern approach to industrial maintenance management,” *Informatica Economică*, c. 14, no. 2, s. 133, 2010.
- [2] J. Sheng D. Prescott, “A coloured Petri net framework for modelling aircraft fleet maintenance,” *Reliability Engineering & System Safety*, c. 189, ss. 67–88, 2019.
- [3] J. M. Berger, “MRO Industry Forecast & Trends,” *Proceedings of the IATA Maintenance Cost Conference, Geneva, Switzerland*, c. 5, 2022.
- [4] *MRO-Survey-2024-Aviation-MRO-Growth-Amid-Turbulence*, <https://www.oliverwyman.com/our-expertise/insights/2024/apr/mro-survey-2024-aviation-mro-grows-amid-rising-costs-supply-chain-woes.html>, Erişim tarihi: 26 May 2024.
- [5] Statista.com, *Global civil air transport MRO market share by segment*, <https://www.statista.com/statistics/387977/global-civil-air-transport-mro-market-segmentation/>, Erişim tarihi: 25 May 2024.
- [6] A. Tokgöz, S. Bulkan, S. Zaim, D. Delen N. G. Torlak, “Modeling airline MRO operations using a systems dynamics approach: A case study of Turkish Airlines,” *Journal of Quality in Maintenance Engineering*, c. 24, no. 3, ss. 280–310, 2018.
- [7] F. De Florio, *Airworthiness: An introduction to aircraft certification and operations*. Butterworth-Heinemann, 2016.
- [8] Ö. F. Ulu, “Data Analytics Methods Used for the Issues of Civil Aviation Maintenance Repair and Overhaul Industry-a Literature Review,” *PQDT-Global*, 2022.
- [9] A. Apostolidis, M. Pelt K. P. Stamoulis, “Aviation data analytics in mro operations: Prospects and pitfalls,” *2020 Annual Reliability and Maintainability Symposium (RAMS)*, IEEE, 2020, ss. 1–7.
- [10] R. Sharda, D. Delen E. Turban, *Business intelligence, analytics, and data science: a managerial perspective*. pearson, 2018.
- [11] M. Pelt, K. Stamoulis A. Apostolidis, “Data analytics case studies in the maintenance, repair and overhaul (MRO) industry,” *MATEC Web of Conferences*, EDP Sciences, c. 304, 2019, s. 04 005.

- [12] D. Dinis, A. Barbosa-Póvoa Â. P. Teixeira, “Valuing data in aircraft maintenance through big data analytics: A probabilistic approach for capacity planning using Bayesian networks,” *Computers & Industrial Engineering*, c. 128, ss. 920–936, 2019.
- [13] M. Intelligence, *Global Aircraft Maintenance, Repair, and Overhaul Market Industry*, <https://www.mordorintelligence.com/industry-reports/global-aircraft-maintenance-repair-and-overhaul-market-industry>, Erişim tarihi: 24 May 2024.
- [14] R. Patriarca, G. Di Gravio, G. Antonioni, N. Paltrinieri ve diğ., “Investigating occupational and operational industrial safety data through Business Intelligence and Machine Learning,” *Journal of Loss Prevention in the Process Industries*, c. 73, s. 104 608, 2021.
- [15] A. K. Jayathilaka, “Operating profit and net profit: measurements of profitability,” *Open Access Library Journal*, c. 7, no. 12, ss. 1–11, 2020.
- [16] V. M. L. Sánchez, F. G. M. Freires, J. C. G. dos Reis J. M. C. M. das Dore, “Industrial Engineering and Operations Management,” 2022.
- [17] C. Strobl, A.-L. Boulesteix T. Augustin, “Unbiased split selection for classification trees based on the Gini index,” *Computational Statistics & Data Analysis*, c. 52, no. 1, ss. 483–501, 2007.
- [18] F. O. Cortés-Ibañez ve diğ., “Prediction of incident cancers in the lifelines population-based cohort,” *Cancers*, c. 13, no. 9, s. 2133, 2021.
- [19] R. Tibshirani, G. Walther T. Hastie, “Estimating the number of clusters in a data set via the gap statistic,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, c. 63, no. 2, ss. 411–423, 2001.
- [20] R. Lleti, M. C. Ortiz, L. A. Sarabia M. S. Sánchez, “Selecting variables for k-means cluster analysis by using a genetic algorithm that optimises the silhouettes,” *Analytica Chimica Acta*, c. 515, no. 1, ss. 87–100, 2004.
- [21] R. Sanche K. Lonergan, “Variable reduction for predictive modeling with clustering,” *Casualty Actuarial Society Forum*, Citeseer, 2006, ss. 89–100.
- [22] J. H. Friedman, “Greedy function approximation: a gradient boosting machine,” *Annals of statistics*, ss. 1189–1232, 2001.
- [23] A. P. Jawalkar ve diğ., “Early prediction of heart disease with data analysis using supervised learning with stochastic gradient boosting,” *Journal of Engineering and Applied Science*, c. 70, no. 1, s. 122, 2023.
- [24] B. Derr, “Ordinal response modeling with the LOGISTIC procedure,” *SAS Global Forum*, Citeseer, 2013, ss. 1–20.
- [25] A. Soroush, A. Bahreininejad J. van den Berg, “A hybrid customer prediction system based on multiple forward stepwise logistic regression mode,” *Intelligent Data Analysis*, c. 16, no. 2, ss. 265–278, 2012.
- [26] N. Draper, *Applied regression analysis*. McGraw-Hill. Inc, 1998.
- [27] C. Park, J.-Y. Koo, P. T. Kim J. W. Lee, “Stepwise feature selection using generalized logistic loss,” *Computational statistics & data analysis*, c. 52, no. 7, ss. 3709–3718, 2008.

- [28] M. Esposito, M. Lazoi, A. Margarito L. Quarta, “Innovating the maintenance repair and overhaul phase through digitalization,” *Aerospace*, c. 6, no. 5, s. 53, 2019.
- [29] L. Guelman, “Gradient boosting trees for auto insurance loss cost modeling and prediction,” *Expert Systems with Applications*, c. 39, no. 3, ss. 3659–3667, 2012.
- [30] D. Paul, A. K. Goswami, R. L. Chetri, R. Roy P. Sen, “Bayesian optimization-based gradient boosting method of fault detection in oil-immersed transformer and reactors,” *IEEE Transactions on Industry Applications*, c. 58, no. 2, ss. 1910–1919, 2021.
- [31] D. P. Pereira, I. L. Gomes, R. Melicio V. M. Mendes, “Planning of aircraft fleet maintenance teams,” *Aerospace*, c. 8, no. 5, s. 140, 2021.



TEZDEN ÜRETİLMİŞ YAYINLAR

Makale

1. N. Ş. Uslu and A. H. Büyüklü, “The dynamics of the profit margin in a component maintenance, repair, and overhaul (MRO) within the aviation industry: an analytical approach using gradient boosting, variable clustering, and the Gini index,” *Sustainability*, vol. 16, no. 15, p. 6470, 2024. doi: 10.3390/su16156470

Konferans Bildirisi

1. N. Ş. Uslu and A. H. Büyüklü, “Challenges in achieving profitability for MRO companies in the aviation industry: an analytical approach,” 11th International Conference on Computational and Experimental Science and Engineering (ICCESEN 2024), Antalya, Türkiye, Oct. 25–28, 2024.

Ödül

1. 100/2000 YÖK Doktora Bursu (Ulaştırma)