



T.C.
ÜSKÜDAR ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YAPAY ZEKA ANABİLİM DALI
YAPAY ZEKA MÜHENDİSLİĞİ YÜKSEK LİSANS PROGRAMI
YÜKSEK LİSANS TEZİ

MAKİNE ÖĞRENMESİ YAKLAŞIMLARININ FORMULA 1
YARIŞ SONUÇLARI TAHMİNİ İÇİN KARŞILAŞTIRILMASI

MERVE BAYRAK

Tez Danışmanı
Prof.Dr. Serhat ÖZEKES

İSTANBUL-2024

T.C.
ÜSKÜDAR ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YAPAY ZEKA ANABİLİM DALI
YAPAY ZEKA MÜHENDİSLİĞİ YÜKSEK LİSANS PROGRAMI
YÜKSEK LİSANS TEZİ

MAKİNE ÖĞRENMESİ YAKLAŞIMLARININ FORMULA 1
YARIŞ SONUÇLARI TAHMİNİ İÇİN KARŞILAŞTIRILMASI

MERVE BAYRAK

Tez Danışmanı
Prof.Dr. Serhat ÖZEKES

İSTANBUL-2024

ÖZET

Makine Öğrenmesi Yaklaşımlarının Formula 1 Yarış Sonuçları Tahmini İçin Karşılaştırılması

Bu çalışma, motor sporlarının önemli bir dalı olan Formula 1 yarışlarının sonuçlarının tahmin edilmesi üzerine hazırlanmıştır. Çalışmada, yapay zekanın bu tahmin edilmesi zor alandaki performansı incelenmiştir. İlk olarak, 2014-2023 yılları arasındaki Formula 1 verileri üzerinde çalışılmış ve veri setine SMOTE uygulanarak dengesiz veri yapısı dengelenmiştir. Verilerde öznitelik seçimi gerçekleştirilerek ve öznitelik seçimi yapılmadan çeşitli makine ve derin öğrenme teknikleri kullanılarak modeller eğitilmiştir. Karşılaştırma sonuçları, özellikle Karar Ağacı (DT) ve Rastgele Orman (RFC) algoritmaları için, öznitelik seçimi yapılmadan daha yüksek doğruluk oranlarına ulaşıldığını göstermiştir. Buna karşılık, Destek Vektör Makineleri (SVM) ve LSTM algoritmaları öznitelik seçimi sonrasında da nispeten yüksek doğruluk oranlarını korumuştur. Ancak, öznitelik seçimi sonrası tahmin başarısında önemli düşüşler gözlemlenmiştir. Bu bulgular, Formula 1 gibi birçok etkenin önemli olduğu sporlarda öznitelik seçiminin tahmin doğruluğunu olumsuz etkileyebileceğini ortaya koymaktadır.

Anahtar Kelimeler: Formula 1, Yapay Zeka, Öznitelik Seçimi, Yarış Tahminleme

ABSTRACT

Comparison of Machine Learning Approaches for Predicting Formula 1 Race Results

This study focuses on predicting the results of Formula 1 races, which is a significant branch of motor sports. The performance of artificial intelligence in this challenging prediction task has been examined. Initially, Formula 1 data from the years 2014 to 2023 were used, and SMOTE was applied to balance the imbalanced data structure. Various machine learning and deep learning techniques were employed to train models both with and without feature selection. The comparison results indicated that, particularly for Decision Tree (DT) and Random Forest (RFC) algorithms, higher accuracy rates were achieved without feature selection. In contrast, Support Vector Machines (SVM) and LSTM algorithms maintained relatively high accuracy rates even after feature selection. However, a significant decrease in prediction success was observed after feature selection. These findings suggest that in sports like Formula 1, where many factors play an important role, feature selection can negatively impact prediction accuracy.

Keywords: Formula 1, Artificial Intelligence, Feature Selection, Race Prediction

TEŞEKKÜR

Bu tez programı sürecinde hayatımın en güzel ve en kötü günlerini geçirdim. Hocalarıma bu dönemde yanımda oldukları için teşekkür ederim. Ama en çok aileme. Babam benim eğitimim için çok çabalamış ve her imkanını bana sunmuş bir insandır. Bu tez sürecinin başlangıcını gördü ama maalesef bitişinde yanımda olamadı. O yüzden bu tezi babama adıyorum. En sonunda benim eğitimim için her şeyi yaptığını kabul ediyorum...



BEYAN FORMU

Bu alıřmadaki btn bilgi ve belgeleri akademik kurallar erevesinde elde ettiđimi, grsel, iřitsel ve yazılı tm bilgi ve sonuları bilimsel ahlak kurallarına uygun olarak sunduđumu, kullandıđım verilerde herhangi bir tahrifat yapmadıđımı, yararlandıđım kaynaklara bilimsel normlara uygun olarak atıfta bulunduđumu, tezimin kaynak gsterilen durumlar dıřında zgn olduđunu, tarafımdan retildiđini ve řkdar niversitesi Sađlık Bilimleri Enstits Tez Yazım Kılavuzuna gre yazıldıđını beyan ederim



Tarih
Merve Bayrak
İmzası

İÇİNDEKİLER

ÖZET	i
ABSTRACT.....	ii
TEŞEKKÜR	iii
BEYAN FORMU	iv
İÇİNDEKİLER	v
TABLolar DİZİNİ.....	ix
ŞEKİLLER DİZİNİ.....	x
SİMGELER VE KISALTMALAR DİZİNİ.....	xii
1. GİRİŞ.....	1
2. GENEL BİLGİLER	3
2.1. Yapay Zeka'nın Tanımı ve Tarihsel Gelişimi.....	3
2.1.1. Yapay Zeka Türleri.....	4
2.1.1.1. Dar Yapay Zeka (ANI).....	4
2.1.1.2. Genel Yapay Zeka (AGI)	5
2.1.1.3. Süper Yapay Zeka (ASI)	5
2.1.2. Yapay Zeka Teknolojileri.....	6
2.1.2.1. Makine Öğrenimi (Machine Learning)	6
2.1.2.2. Derin Öğrenme (Deep Learning).....	6
2.1.2.3. Doğal Dil İşleme (Natural Language Processing).....	7
2.1.2.4. Bilgisayarla Görme (Computer Vision)	7
2.1.2.5. Pekiştirmeli Öğrenme (Reinforcement Learning)	7
2.1.3. Yapay Zekanın Çalışma Prensipleri	8
2.1.3.1. Veri ve Veri Temsili	8
2.1.3.2. Öğrenme Algoritmaları	8
2.1.3.3. Algoritmik Optimizasyon.....	9

2.1.3.4.	Geri Bildirim ve Uyarlama	9
2.1.3.5.	Karar Verme ve Öngörü	9
2.1.4.	Yapay Zekada Kullanılan Algoritmalar	10
2.1.4.1.	Karar Ağaçları (Decision Trees)	10
2.1.4.2.	Destek Vektör Makineleri (Support Vector Machines - SVM).....	11
2.1.4.3.	Yapay Sinir Ağları (Artificial Neural Networks - ANN).....	11
2.1.4.4.	K-En Yakın Komşu (K-Nearest Neighbors - KNN)	11
2.1.4.5.	Naive Bayes Algoritması.....	12
2.1.4.6.	Random Forest Algoritması	12
2.1.4.7.	Derin Öğrenme (Deep Learning).....	12
2.1.4.8.	Uzun Kısa Süreli Bellek (LSTM).....	13
2.1.4.9.	Meta-Sezgisel Algoritmalar (Metaheuristic Algorithms).....	13
2.1.5.	Yapay Zekanın Avantajları ve Dezavantajları	15
2.1.5.1.	Yapay Zekanın Avantajları.....	15
2.1.5.2.	Yapay Zekanın Dezavantajları	16
2.2.	Formula 1	17
2.2.1.	Formula 1'in Tarihçesi	17
2.2.2.	Takımlar ve Sürücüler	18
2.2.2.1.	Önde Gelen Takımlar	18
2.2.2.2.	Sürücüler: Yetenek ve Strateji.....	19
2.2.2.3.	Sürücüler ve Takım İçi Rekabet	19
2.2.3.	Formula 1'in Yapısı ve Kuralları	20
2.2.3.1.	Puanlama Sistemi ve Stratejik Yenilikler.....	20
2.2.3.2.	Teknik Kurallar ve Araç Tasarımı.....	21
2.2.3.3.	Sürdürülebilirlik ve Çevresel Etkiler.....	21
2.2.4.	Formula 1'in Sosyoekonomik Etkisi.....	22
2.2.5.	Teknolojik Gelişmeler ve Mühendislik İnovasyonları	23

2.2.5.1.	Formula 1’de Veri Analizi ve Yapay Zeka	23
2.2.5.2.	Formula 1’de Makine Öğrenimi Uygulamaları	26
2.2.5.3.	Yapay Zekanın Performans Tahmini ve Strateji Oluşturmada Kullanımı ...	27
2.2.5.4.	Yapay Zekanın Sürücü Performansını Değerlendirmedeki Rolü	27
2.2.5.5.	Yapay Zeka ile Yarış Simülasyonları	28
2.2.5.6.	Formula 1’de Yapay Zekanın Geleceği	29
3.	GEREÇ VE YÖNTEM	30
3.1.	Verilerin Toplanması	32
3.2.	Verilerin Temizlenmesi ve İşlenmesi	33
3.3.	Öznitelik Seçimi	36
3.4.	Verilerin Modellenmesi	38
4.	BULGULAR	39
4.1.	Ham Veriye Uygulamam Makine Öğrenmesi Algoritmaları	39
4.1.1.	Long Short-Term Memory	39
4.1.2.	Decision Tree	44
4.1.3.	Naive Bayes	46
4.1.4.	Random Forest Classifier	50
4.1.5.	Support Vector Machines	52
4.2.	MASEFE Yöntemi ile Seçilen Özniteliklerin Makine Öğrenimi Algoritmalarına Uygulanması	55
4.2.1.	OriginalSHIO Optimisasyonu	55
4.2.1.1.	Long Short- Term Memory	56
4.2.1.2.	Desicion Tree	59
4.2.1.3.	Navie Bayes	63
4.2.1.4.	Random Forest	66
4.2.1.5.	Support Vektor Machine	71
4.3.	Ham Veri Seti Sonuçlarının Değerlendirilmesi	76

4.4.	Öznitelik Seçimi (OriginalSHIO) Sonuçlarının Değerlendirilmesi	77
4.5.	Öznitelik Seçimi ve Ham Veri Seti Tahmin Sonuçları Karşılaştırması	79
4.6.	SMOTE Uygulamasının Etkisi.....	79
5.	TARTIŞMA.....	80
6.	SONUÇ VE ÖNERİLER	83
KAYNAKLAR		84
EKLER		90
Ek 1. Özgeçmiş		90



TABLÖLAR DİZİNİ

	<u>Sayfa</u>
Tablo 1: Ergast API ile indirilen öznitelikler.....	32
Tablo 2: Ham veri setinin doğruluk sonuçları	76
Tablo 3: Ham veri setinin tahmin sonuçları.....	77
Tablo 4: OriginalSHIO optimizasyonu ile yeni veri setinin doğruluk sonuçları.....	77
Tablo 5: OriginalSHIO optimizasyonu ile yeni veri setinin tahmin sonuçları	79



ŞEKİLLER DİZİNİ

	<u>Sayfa</u>
Şekil 1: LSTM verisinin dağılımı	40
Şekil 2: SMOTE uygulanan eğitim verisinin dağılımı	40
Şekil 3: LSTM eğitim verisi performans değerleri	41
Şekil 4: LSTM test verisi performans değerleri.....	42
Şekil 5: LSTM doğrulama verisi performans değerleri	43
Şekil 6 LSTM örnek tahmin sütunu.....	43
Şekil 7: DT test verisi performans değerleri	44
Şekil 8: DT doğrulama verisi performans değerleri	45
Şekil 9: DT eğitim verisi performans değerleri	45
Şekil 10: DT örnek tahmin sütunu	46
Şekil 11: NB eğitim verisi performans değerleri	47
Şekil 12: NB test verisi performans değerleri.....	48
Şekil 13: NB doğrulama verisi performans değerleri	49
Şekil 14: NB örnek tahmin sütunu.....	49
Şekil 15: RF eğitim verisi performans değerleri.....	50
Şekil 16: RF test verisi performans değerleri	51
Şekil 17: RF doğrulama verisi performans değerleri.....	51
Şekil 18: RF örnek tahmin sütunu	52
Şekil 19: SVM eğitim verisi performans değerleri	53
Şekil 20: SVM test verisi performans değerleri.....	54
Şekil 21: SVM doğrulama verisi performans değerleri	54
Şekil 22: SVM örnek tahmin sütunu.....	55
Şekil 23: SHIO optimizasyonu ile LSTM modelinde kullanılmak için seçilen öznitelikler	56
Şekil 24 LSTM modelinde kullanılacak yeni veri boyutları.....	56
Şekil 25: LSTM ve yeni veri seti ile eğitim verisi performans değerleri	57
Şekil 26: LSTM ve yeni veri seti ile test verisi performans değerleri	58
Şekil 27: LSTM ve yeni veri seti ile doğrulama verisi performans değerleri.....	58
Şekil 28: LSTM ve yeni veri seti ile örnek tahmin sütunu	59
Şekil 29: SHIO optimizasyonu ile DT modelinde kullanılmak için seçilen öznitelikler	60
Şekil 30: DT modelinde kullanılacak yeni veri boyutları	60

Şekil 31: DT ve yeni veri seti ile test verisi performans değerleri	61
Şekil 32: DT ve yeni veri seti ile doğrulama verisi performans değerleri	61
Şekil 33: DT ve yeni veri seti ile eğitim verisi performans değerleri	62
Şekil 34: DT ve yeni veri seti ile örnek tahmin sütunu	62
Şekil 35: SHIO optimizasyonu ile NB modelinde kullanılmak için seçilen öznitelikler	63
Şekil 36: NB modelinde kullanılacak yeni veri boyutları	63
Şekil 37: NB ve yeni veri seti ile eğitim verisi performans değerleri	64
Şekil 38: NB ve yeni veri seti ile test verisi performans değerleri	65
Şekil 39: NB ve yeni veri seti ile doğrulama verisi performans değerleri	65
Şekil 40: NB ve yeni veri seti ile örnek tahmin sütunu	66
Şekil 41: SHIO optimizasyonu ile RF modelinde kullanılmak için seçilen öznitelikler	67
Şekil 42: RF modelinde kullanılacak yeni veri boyutları	67
Şekil 43: RF ve yeni veri seti ile eğitim verisi performans değerleri	68
Şekil 44: RF ve yeni veri seti ile test verisi performans değerleri	69
Şekil 45: RF ve yeni veri seti ile doğrulama verisi performans değerleri	70
Şekil 46: RF ve yeni veri seti ile örnek tahmin sütunu	71
Şekil 47: SHIO optimizasyonu ile SVM modelinde kullanılmak için seçilen öznitelikler	72
Şekil 48: SVM modelinde kullanılacak yeni veri boyutları	72
Şekil 49: SVM ve yeni veri seti ile eğitim verisi performans değerleri	73
Şekil 50: SVM ve yeni veri seti ile test verisi performans değerleri	74
Şekil 51: SVM ve yeni veri seti ile doğrulama verisi performans değerleri	75
Şekil 52: SVM ve yeni veri seti ile örnek tahmin sütunu	75

SİMGELER VE KISALTMALAR DİZİNİ

AR-GE: Araştırma ve Geliştirme

YZ: Yapay Zeka

SMOTE: Synthetic Minority Over-sampling Technique

DT: Decision Tree

SVM: Support Vector Machine

NB: Naive Bayes

LSTM: Long Short-Term Memory

AI: Artificial Intelligence

ANI: Artificial Narrow Intelligence

AGI: Artificial General Intelligence

ASI: Artificial Super Intelligence

ML: Machine Learning

DL: Deep Learning

F1: Formula 1

FIA: Fédération Internationale de l'Automobile

ERS: Energy Recovery Systems

MÖ: Makine Öğrenimi

API: Application Programming Interface

URL: Uniform Resource Locator

CSV: Comma-separated Value

MAFESE: Metaheuristic Algorithms for FEature SElection

OBB: out-of-bag

OriginalSHIO: Success History Intelligent Optimization



1. GİRİŞ

Formula 1, dünyadaki en yüksek seviyede araştırma-geliştirme (AR-GE) gerektiren ve en çok para harcanan spor dallarından biri olarak öne çıkmaktadır. Bu spor, hem mühendisler hem de diğer çalışanlar açısından ciddi emek ve adanmışlık gerektirmektedir. Aynı zamanda, sürücülerin hayatlarını tehlikeye atarak yarıştığı bu alanda, teknoloji ve mühendislik çalışmalarının başarısı doğrudan yarış sonuçlarını etkileyen en önemli faktörlerden biri olarak kabul edilmektedir. Formula 1, sadece spor anlamında değil, aynı zamanda teknolojik gelişim ve inovasyonun sınırlarını zorlayan bir alan olarak görülmektedir. Bu bağlamda, bu tezin seçilmesindeki temel motivasyon, Formula 1'in sağladığı bu cazip AR-GE ve mühendislik imkanlarıdır.

Bu tez çalışmasında, yapay zekanın gelişmiş bir dalı olan derin öğrenme ve öznitelik seçimi gibi konular ele alınmıştır. Bu bağlamda, kullanılan veri seti 2014-2023 yılları arasındaki Formula 1 yarışlarına ait olup, birçok temizlik ve düzenleme işlemi uygulanarak hazır hale getirilmiştir.

Python programlama dili kullanılarak yapılan bu işlemler sonucunda, 117 öznitelik ve 3837 satır içeren bir veri seti oluşturulmuştur. Çalışmada 2014-2022 yılları eğitim veri seti, 2023 yılı ise test veri seti olarak ayrılmıştır. Eğitim veri setinde düzensizliği azaltmak ve modelin dengesiz veri ile eğitilmesini önlemek amacıyla SMOTE (Synthetic Minority Over-sampling Technique) yöntemi uygulanmıştır. Bu adımın ardından, özniteliklerin sadeleştirilmesi amacıyla öznitelik seçimi için kullanılan Mafese kütüphanesi tercih edilmiş ve farklı optimizasyon yöntemleriyle öznitelikler daraltılmıştır.

Veri setinin sadeleştirilmesinin ardından, farklı makine öğrenmesi algoritmaları olan Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Naive Bayes (NB) ve Long Short-Term Memory (LSTM) modelleri uygulanmıştır. Elde edilen sonuçlar, yarış sonuçlarını tahmin etmeye yönelik olarak analiz edilmiştir. Ayrıca, öznitelik seçimi yapılmamış ham veri setine de aynı modeller uygulanarak, öznitelik seçiminin model performansı üzerindeki etkisi karşılaştırılmıştır. Bu analiz, Formula 1 veri seti gibi karmaşık veri yapılarında, öznitelik seçiminin gerekliliği ve önemini

sorgulamak için yapılmıştır. Makine öğrenmesi algoritmalarının daha verimli çalışmasını sağlamak için öz nitelik seçimi oldukça önemli bir adımdır, ancak her veri seti için bu uygulamanın gerekli olup olmadığı da tartışmaya açıktır.

Buna ek olarak, LSTM derin öğrenme algoritmasının sınıflandırma performansı da ölçülmüştür. LSTM, zaman serileri ve ardışık veriler üzerinde güçlü bir performansa sahip olmasına karşın, sınıflandırma problemlerindeki etkinliği incelenmiş ve diğer algoritmalarla karşılaştırılmıştır. Formula 1 yarışlarının sonuçlarını tahmin etmeye yönelik olarak geliştirilen bu makine öğrenmesi algoritmalarının amacı, veri odaklı analizlerle gelecekteki yarış sonuçlarını öngörmek ve performans analizi yapmaktır. Bu çalışmada araç özellikleri değerlendirme dışı bırakılmış olup, tahminler sürücü performansları, hava durumu, pist koşulları ve geçmiş yarış verileri gibi çeşitli faktörler üzerinden yapılmıştır. Araç özellikleri, daha geniş kapsamlı ve farklı veri kaynaklarına ihtiyaç duyduğundan, bu projeye dahil edilmemiştir.

Sonuç olarak, bu tür tahminler, Formula 1 takımları için stratejik kararları destekleyebilir, yarış stratejilerinin optimize edilmesine yardımcı olabilir ve rakiplerin performansını öngörerek avantaj sağlayabilmektedir. Ayrıca, bahis, medya ve izleyici kitlesi için de ilgi çekici tahmin modelleri oluşturularak, daha geniş bir etki yaratılabilmektedir. Bu tezde yapılan çalışmalar, hem akademik hem de pratik anlamda yapay zekanın ve makine öğrenmesinin spor dünyasında nasıl kullanılabileceğine dair önemli bulgular sunmaktadır.

2. GENEL BİLGİLER

2.1. Yapay Zeka'nın Tanımı ve Tarihsel Gelişimi

Yapay zeka (YZ), insan zekasını taklit eden sistemlerin geliştirilmesi ve makinelerin, özellikle bilgisayarların, insan benzeri görevleri yerine getirme yeteneğini ifade eden bir kavramdır. Bu alanda yapılan çalışmalar, makine öğrenimi, derin öğrenme ve doğal dil işleme gibi alt disiplinler aracılığıyla makinelerin, insanlara özgü yetenekleri kazanmalarını hedeflemektedir (Russell ve Norvig, 2020). Yapay zeka, basit karar ağaçlarından karmaşık sinir ağlarına kadar farklı algoritmalar kullanılarak geliştirilmekte ve geniş bir uygulama alanı bulmaktadır. Bu gelişmelerin temel amacı, insan müdahalesi olmadan makinelerin öğrenebilmesi, sorunları çözebilmesi ve öngöründe bulunabilmesidir.

Yapay zekanın kökleri, modern bilgisayar bilimi ile yakından ilişkilidir. Yapay zekanın ilk teorik temelleri, 20. yüzyılın ortalarında Alan Turing tarafından atılmıştır. Turing, makinelerin düşünebileceği fikrini öne sürmüştü ve bu fikir 1950 yılında yayımlanan "Computing Machinery and Intelligence" makalesi ile şekillenmiştir (Turing, 1950). Bu dönemde, makinelerin insan zekasına benzer şekilde problem çözme yeteneklerine sahip olup olamayacağı üzerine tartışmalar başlamıştır. Turing Testi olarak bilinen ve bir makinenin, bir insanın zekâsına denk olup olmadığını ölçmeyi hedefleyen bu test, yapay zeka çalışmalarında önemli bir dönüm noktası olarak kabul edilmiştir.

Yapay zeka terimi ise ilk kez 1956 yılında Dartmouth Konferansı'nda John McCarthy tarafından kullanılmıştır (McCarthy ve ark., 2006). Bu konferans, yapay zekanın bağımsız bir bilim dalı olarak ortaya çıkışını simgelemektedir. McCarthy'nin önderliğindeki araştırmacılar, makinelerin öğrenme, düşünme ve problem çözme kapasitelerini geliştirmek için çalışmalara başlamışlardır. Bu süreçte, sembolik yapay zeka (symbolic AI) yaklaşımı ağırlık kazanmış ve uzman sistemler gibi kural tabanlı sistemler geliştirilmiştir. Ancak, bu sistemlerin karmaşık ve dinamik ortamlarda yetersiz kaldığı görülmüş, bu nedenle makine öğrenimi gibi yeni yaklaşımlar ortaya çıkmıştır (Goodfellow ve ark., 2016).

1980'li yıllarda yapay sinir ağlarının geliştirilmesiyle, YZ'nin ikinci büyük dönemi başlamıştır. Derin öğrenme adı verilen bu yöntem, çok katmanlı sinir ağları aracılığıyla veriden öğrenme kapasitesini artırmıştır. Özellikle Geoffrey Hinton ve Yann LeCun gibi bilim insanlarının öncülük ettiği bu dönem, büyük veri (big data) ve artan hesaplama gücü sayesinde hızla ilerlemiştir (LeCun ve ark., 2015). Bu gelişmeler, görüntü tanıma, doğal dil işleme ve otonom sistemler gibi birçok alanda önemli atılımlar yapılmasına olanak sağlamıştır.

Son dönemde yapay zeka, günlük yaşamın bir parçası haline gelmiştir. Siri, Alexa ve Google Assistant gibi sanal asistanlar, otonom araçlar ve tıbbi teşhis sistemleri gibi teknolojiler, yapay zekanın günlük kullanımlarına örnek olarak gösterilebilir. Bunun yanı sıra, YZ'nin etik ve sosyal boyutları da daha fazla tartışılmaya başlanmıştır. Özellikle karar süreçlerinde insan müdahalesinin azalması ve veriye dayalı önyargılar gibi konular, yapay zeka uygulamalarının etik zeminini sorgulatan tartışmalara yol açmıştır (Bostrom, 2014).

YZ alanı, insan benzeri düşünme ve problem çözme yeteneklerine sahip makineler geliştirmeye odaklanarak, bilgisayar bilimi ve mühendisliği alanlarında büyük bir ilerleme kaydetmiştir. Tarihsel gelişimi boyunca, sembolik yapay zekadan derin öğrenmeye kadar farklı yaklaşımlar benimsenmiş ve her bir dönemde önemli teorik ve pratik katkılar sağlanmıştır.

2.1.1. Yapay Zeka Türleri

Yapay zeka (YZ) türleri, genellikle sistemlerin yeteneklerine ve işleyiş biçimlerine göre sınıflandırılmaktadır. Bu sınıflandırma, YZ'nin farklı seviyelerde öğrenme, karar verme ve problem çözme kapasitelerine dayanmaktadır. Yapay zeka, genellikle üç ana kategoride incelenmektedir: dar yapay zeka (ANI), genel yapay zeka (AGI) ve süper yapay zeka (ASI) (Bostrom, 2014).

2.1.1.1. Dar Yapay Zeka (ANI)

Dar yapay zeka, belirli bir görev veya problem için tasarlanmış sistemleri ifade etmektedir. Bu sistemler, sınırlı bir alanda faaliyet gösterir ve genel zekaya sahip değildir. Yani, bir alan üzerinde oldukça yetenekli olmasına rağmen, başka bir alana adapte olma kapasitesine sahip değildir. Örneğin, satranç oynayan bir yapay zeka, sadece bu oyunda

uzmanlaşmıştır ve satranç dışındaki görevlerde işlev gösterememektedir. Siri, Alexa, Google arama motoru gibi uygulamalar dar yapay zekanın günlük yaşamda yaygın olarak kullanılan örnekleridir (Russell ve Norvig, 2020). Dar yapay zekanın bu sınırlı yeteneği, bugün en çok karşılaşılan yapay zeka türüdür ve pratik uygulamalar açısından büyük bir önem taşımaktadır.

2.1.1.2. Genel Yapay Zeka (AGI)

Genel yapay zeka, insan düzeyinde bilişsel işlevlere sahip olan sistemlerdir. Bu tür bir yapay zeka, insan zekasını taklit ederek birçok farklı görevde başarılı olabilme yeteneğine sahiptir. Genel yapay zekanın en önemli özelliği, birden fazla problemi çözebilmesi, öğrenebilmesi ve bu bilgiyi farklı alanlara aktarabilmesidir (Goertzel, 2014). AGI, bir insanın gerçekleştirebileceği her türlü bilişsel görevi yerine getirebilir ve uyarlanabilir yeteneklere sahiptir. Ancak, bu tür yapay zeka henüz teorik bir kavram olarak kalmakta ve bugünkü teknolojik düzeyde tam anlamıyla gerçekleştirilememiştir. AGI'nin geliştirilmesi, yapay zekanın en büyük hedeflerinden biri olarak görülmektedir.

2.1.1.3. Süper Yapay Zeka (ASI)

Süper yapay zeka, insan zekasını aşan bir yapay zeka türünü ifade etmektedir. Bu tür bir yapay zeka, her açıdan insanın bilişsel yeteneklerini geride bırakabilir ve çok daha karmaşık problemleri çözebilmektedir. ASI, bilim kurgu senaryolarında sıkça yer bulmasına rağmen, günümüzde henüz ulaşılmamış bir seviyedir. Bu seviyeye ulaşılmasının, insanlık üzerinde derin etkileri olacağı ve bu nedenle etik sorunları beraberinde getireceği öne sürülmektedir (Bostrom, 2014). ASI, teorik olarak sınırsız bir öğrenme kapasitesine sahip olabilir ve insanlığın karşılaştığı en büyük sorunları çözme potansiyeline sahip olacaktır. Ancak, aynı zamanda kontrol edilemez hale gelme riski de taşıdığından dolayı büyük bir dikkatle ele alınması gereken bir konudur.

YZ türleri, sistemlerin yetenek seviyelerine göre farklılaşmaktadır. Dar yapay zeka, günümüzde en yaygın olan tür iken, genel yapay zeka ve süper yapay zeka henüz teorik düzeyde kalmaktadır. Bu sınıflandırma, yapay zekanın mevcut kapasitesini ve potansiyel gelişim alanlarını anlamak açısından kritik öneme sahiptir.

2.1.2. Yapay Zeka Teknolojileri

Yapay zeka teknolojileri, algoritmaların ve yazılım sistemlerinin geliştirilmesiyle makinelerin insan benzeri yetenekler kazanmasını sağlayan çeşitli araçları ve teknikleri içermektedir. Bu teknolojiler, veri analizi, doğal dil işleme, makine öğrenimi ve derin öğrenme gibi birçok alanda kullanılmaktadır. YZ teknolojilerinin temel amacı, makinelerin öğrenebilmesi, karar verebilmesi ve öngöründe bulunabilmesidir (Russell ve Norvig, 2020). Aşağıda, yapay zeka teknolojilerinin en önemli bileşenleri açıklanmaktadır.

2.1.2.1.Makine Öğrenimi (Machine Learning)

Makine öğrenimi (ML), yapay zeka sistemlerinin veri kullanarak kendi kendine öğrenmesine olanak tanıyan bir teknolojidir. Geleneksel programlamanın aksine, makine öğrenimi algoritmaları, belirli kurallarla programlanmaz; bunun yerine, veriden öğrenerek kendilerini geliştirirler. Bu süreçte, denetimli öğrenme (supervised learning), denetimsiz öğrenme (unsupervised learning) ve pekiştirmeli öğrenme (reinforcement learning) gibi farklı öğrenme türleri kullanılmaktadır (Goodfellow ve ark., 2016). Makine öğrenimi, veri analitiği, görüntü tanıma, ses tanıma ve öngörücü modelleme gibi birçok alanda uygulanmaktadır. Örneğin, Netflix'in öneri sistemi, kullanıcı davranışlarına dayalı olarak makine öğrenimi kullanarak öneriler sunmaktadır.

2.1.2.2.Derin Öğrenme (Deep Learning)

Derin öğrenme (DL), yapay sinir ağlarına dayalı bir makine öğrenimi alt dalıdır. Bu teknoloji, çok katmanlı yapay sinir ağları kullanarak büyük miktarda veriyi analiz edebilir ve bu veriden karmaşık ilişkiler öğrenebilmektedir. Derin öğrenme, özellikle görüntü ve ses tanıma, doğal dil işleme ve otonom sistemlerde yaygın olarak kullanılmaktadır. 2010'lerden itibaren, büyük veri setleri ve güçlü hesaplama kaynaklarının artmasıyla derin öğrenme alanında önemli atılımlar yapılmıştır (LeCun ve ark., 2015). Bu teknolojinin başarılı uygulamaları arasında Google'ın AlphaGo sisteminin Go oyununu kazanması ve otonom araçların çevreyi algılayıp karar vermesi yer almaktadır.

2.1.2.3.Doğal Dil İşleme (Natural Language Processing)

Doğal dil işleme (NLP), bilgisayarların insan dilini anlamasına ve bu dilde işlem yapmasına olanak tanıyan bir yapay zeka teknolojisidir. Bu teknoloji, dilin işlenmesi, analiz edilmesi ve üretimi gibi alanları kapsamaktadır. Doğal dil işleme, metin sınıflandırma, konuşma tanıma, dil çevirisi ve chatbot'lar gibi birçok uygulama alanına sahiptir. Örneğin, Google Translate, farklı diller arasında çeviri yaparken doğal dil işleme algoritmalarını kullanmaktadır. ChatGPT ve Siri gibi sanal asistanlar da doğal dil işleme teknolojisini kullanarak kullanıcıların sorularını anlamakta ve yanıtlamaktadır (Jurafsky ve Martin, 2020).

2.1.2.4.Bilgisayarla Görme (Computer Vision)

Bilgisayarla görme, bilgisayarların görsel verileri anlamasına ve yorumlamasına olanak tanıyan bir yapay zeka alanıdır. Görüntülerin ve videoların analiz edilmesi, nesnelerin tanınması ve görüntüdeki kalıpların belirlenmesi bu teknolojinin temel hedeflerindendir. Bilgisayarla görme teknolojisi, otonom araçlar, güvenlik kameraları, tıbbi görüntüleme ve yüz tanıma sistemleri gibi birçok alanda kullanılmaktadır (Szeliski, 2021). Özellikle derin öğrenme teknikleri sayesinde, bilgisayarla görme alanında önemli ilerlemeler kaydedilmiştir. Örneğin, Tesla'nın otonom sürüş teknolojileri, çevresini analiz etmek için bilgisayarla görme teknolojilerini kullanmaktadır.

2.1.2.5.Pekiştirmeli Öğrenme (Reinforcement Learning)

Pekiştirmeli öğrenme, bir yapay zeka sisteminin çevresinden aldığı geri bildirimler doğrultusunda öğrenmesini sağlayan bir teknolojidir. Bu teknoloji, ödül ve ceza sistemine dayanır; sistem, yaptığı her eylem sonucunda bir ödül veya ceza alır ve bu bilgilere dayanarak gelecekteki kararlarını optimize etmektedir (Sutton ve Barto, 2018). Pekiştirmeli öğrenme, özellikle otonom robotlar, oyunlar ve kontrol sistemlerinde yaygın olarak kullanılmaktadır. Örneğin, DeepMind tarafından geliştirilen AlphaGo, Go oyununda rakiplerini yenmek için pekiştirmeli öğrenme algoritmalarını kullanmıştır.

YZ teknolojileri, veri işleme, öğrenme ve karar verme süreçlerini geliştirmek amacıyla birçok farklı teknik ve algoritmayı bir araya getirmektedir. Makine öğrenimi, derin öğrenme, doğal dil işleme, bilgisayarla görme ve pekiştirmeli öğrenme, günümüzde

yapay zekanın en yaygın kullanılan teknolojileridir ve bu alanlar, gelecekteki yapay zeka gelişmelerinin temelini oluşturacaktır.

2.1.3. Yapay Zekanın Çalışma Prensipleri

Yapay zeka, insan zekasını taklit eden makinelerin tasarlanması ve bu makinelerin öğrenme, problem çözme, karar alma ve öngöründe bulunma becerileriyle çalışmasını sağlayan belirli prensiplere dayanmaktadır. Yapay zekanın çalışma prensipleri, algoritmaların veri işleme yetenekleri ve bu verilerden anlam çıkarma kabiliyetlerine odaklanmaktadır. Bu prensipler, makine öğrenimi, veri temsili, optimizasyon ve geri bildirim döngülerini kapsayan bir yapıda ele alınmaktadır (Russell ve Norvig, 2020).

2.1.3.1. Veri ve Veri Temsili

Yapay zekanın çalışma prensiplerinin temelinde veri ve veri temsili yer almaktadır. Veriler, yapay zeka sistemlerinin öğrenmesi için gereken en kritik unsurdur. Bu veriler, yapılandırılmış (tablo verileri) ya da yapılandırılmamış (görüntü, ses, metin) olabilmektedir. YZ algoritmaları, bu verilerden anlamlı bilgiler çıkarabilmek için öncelikle verileri uygun bir formatta temsil etmelidir. Veri temsili, algoritmaların verilerdeki örüntüleri tanımlayabilmesi ve çıkarımlar yapabilmesi için gerekli olan adımlardan biridir. Örneğin, makine öğrenimi modelleri genellikle verileri sayısal özelliklere (features) indirger ve bu özellikler üzerinden tahminlerde bulunur (Goodfellow ve ark., 2016). Yapay zeka sistemleri için doğru veri temsili, algoritmaların doğruluğunu ve etkinliğini doğrudan etkileyen bir faktördür.

2.1.3.2. Öğrenme Algoritmaları

Yapay zekanın en önemli prensiplerinden biri, öğrenme algoritmaları aracılığıyla verilerden bilgi çıkarma yeteneğidir. Makine öğrenimi, yapay zeka sistemlerinin verilerden öğrenebilmesini sağlayan ana teknolojidir. Öğrenme algoritmaları, deneyimler (veri) aracılığıyla sistemin performansını iyileştirmeyi amaçlar. Denetimli öğrenme (supervised learning), denetimsiz öğrenme (unsupervised learning) ve pekiştirmeli öğrenme (reinforcement learning) bu prensip altında ele alınan temel öğrenme türleridir (Sutton ve Barto, 2018). Denetimli öğrenmede, sisteme verilen verilerle birlikte hedefler

de sađlanır ve algoritma, bu hedefleri tahmin etmeyi öğrenmektedir. Denetimsiz öğrenme ise, verilerdeki gizli örüntüleri bulmayı hedefler ve hedef etiketleri içermez. Pekiştirmeli öğrenme ise ödöl ve ceza mekanizması aracılığıyla sistemin çevresiyle etkileşim kurarak öğrenmesini sađlamaktadır.

2.1.3.3.Algoritmik Optimizasyon

Yapay zekanın etkin bir şekilde çalışabilmesi için algoritmaların optimize edilmesi gerekmektedir. Optimizasyon, algoritmaların hata oranını en aza indirmek ve doğruluđunu artırmak amacıyla yapılan bir süreçtir. Genellikle makine öğrenimi modellerinde, modelin tahmin doğruluđunu artırmak için parametrelerin düzenlenmesiyle gerçekleştirilir. Yapay zeka sistemleri, optimizasyon teknikleri sayesinde, büyük veri setlerinden öğrenerek karmaşık problemleri çözebilmektedir. Gradient descent gibi optimizasyon yöntemleri, parametrelerin doğru şekilde ayarlanması için kullanılan yaygın tekniklerdir (Ruder, 2017). Optimizasyon süreci, yapay zeka modellerinin hem daha hızlı hem de daha doğru sonuçlar üretmesini sađlamaktadır.

2.1.3.4.Geri Bildirim ve Uyarılma

Yapay zeka sistemlerinin bir diđer önemli prensibi, geri bildirim mekanizmasıdır. Yapay zeka, performansını artırmak için sürekli olarak geri bildirim alır ve bu geri bildirim doğrultusunda kendisini uyarlayabilmektedir. Pekiştirmeli öğrenmede olduđu gibi, sistemler çevreden aldıkları geri bildirimler aracılığıyla eylemlerini düzenlemektedir. Geri bildirim mekanizması, sistemin doğru sonuçlar üretmesini sađlamakla kalmaz, aynı zamanda sistemin dinamik ve deđişken ortamlara uyum sađlamasına da yardımcı olmaktadır (Sutton ve Barto, 2018). Bu adaptasyon yeteneđi, yapay zeka sistemlerinin çevrelerine ve deđişen koşullara göre performanslarını iyileştirmesini sađlamaktadır.

2.1.3.5.Karar Verme ve Öngörü

Yapay zeka sistemlerinin nihai hedeflerinden biri, öğrenilen verilerden karar verme ve öngörüde bulunma yeteneđidir. Yapay zeka algoritmaları, verilerdeki örüntüleri analiz ederek gelecekteki olayları tahmin edebilir ve bu tahminlere dayalı kararlar alabilmektedir. Örneđin, bir YZ modeli, geçmiş finansal verileri inceleyerek gelecekteki piyasa hareketlerini tahmin edebilir ya da bir otonom araç, çevresindeki nesneleri

algılayarak güvenli sürüş kararları alabilmektedir (LeCun ve ark., 2015). Karar verme süreçleri, algoritmaların doğru sonuçlara ulaşması ve bu sonuçlara dayalı olarak eylemde bulunması prensibine dayanmaktadır.

Yapay zekanın çalışma prensipleri, verilerin işlenmesi, öğrenme süreçlerinin optimize edilmesi, geri bildirim mekanizmaları ve doğru karar verme yeteneklerini kapsayan çok yönlü bir yapı üzerine inşa edilmiştir. Bu prensipler, yapay zeka sistemlerinin dinamik, öğrenebilir ve öngörülebilir olmasını sağlayarak, insan benzeri bilişsel yetenekler kazanmalarına olanak sağlamaktadır.

2.1.4. Yapay Zekada Kullanılan Algoritmalar

Yapay zeka alanında kullanılan algoritmalar, makinelerin veri öğrenmesi, öngöründe bulunması ve karar vermesini sağlamak için geliştirilmiştir. Bu algoritmalar, yapay zeka sistemlerinin temeli olarak kabul edilir ve farklı problemlerin çözümünde çeşitli teknikler ve yöntemler sunmaktadır. Yapay zekada kullanılan algoritmalar, makine öğrenimi, derin öğrenme, arama algoritmaları ve doğal dil işleme gibi birçok alt disiplinde kendine yer bulmaktadır (Russell ve Norvig, 2020). Aşağıda, yapay zeka alanında en yaygın olarak kullanılan algoritmaların başlıca türleri açıklanmaktadır.

2.1.4.1. Karar Ağaçları (Decision Trees)

Karar ağaçları, yapay zeka alanında kullanılan popüler bir denetimli öğrenme algoritmasıdır. Bu algoritma, veri setlerini sınıflandırma ve regresyon analizleri için kullanılmaktadır. Karar ağacı, verileri dallara ayırarak ve her bir dalda bir karar vererek çalışmaktadır. Bu süreç, karar ağacının kökünden başlayarak her bir düğümde bir koşul sorgulaması yapılması ve bu koşula bağlı olarak yeni dalların oluşturulması şeklinde ilerlemektedir. Sonuçta, karar ağaçları, sınıflandırma veya regresyon problemlerinde veri kümelerini öngörülerde bulunmak amacıyla kullanılmaktadır (Quinlan, 1986). Karar ağaçları, kolay anlaşılabilir yapısı nedeniyle yaygın olarak tercih edilmektedir ve özellikle küçük veri setlerinde oldukça başarılı sonuçlar verebilmektedir.

2.1.4.2. Destek Vektör Makineleri (Support Vector Machines - SVM)

Destek vektör makineleri (SVM), özellikle sınıflandırma problemleri için kullanılan güçlü bir denetimli öğrenme algoritmasıdır. Bu algoritma, iki sınıf arasındaki en iyi ayırım çizgisini veya hiper düzlemi bulmaya çalışmaktadır. SVM, verilerin en geniş marjla iki sınıfa ayrılmasını hedefler ve bu marjı maksimize ederek sınıflandırmayı gerçekleştirir (Cortes ve Vapnik, 1995). SVM'nin en önemli avantajı, çok boyutlu veri kümelerinde dahi yüksek performans gösterebilmesidir. Ayrıca, çekirdek (kernel) fonksiyonları sayesinde doğrusal olmayan veri setlerinde de kullanılabilir. Bu nedenle, yüz tanıma, biyometrik güvenlik ve genetik veri analizi gibi alanlarda sıkça tercih edilmektedir.

2.1.4.3. Yapay Sinir Ağları (Artificial Neural Networks - ANN)

Yapay sinir ağları, biyolojik sinir sistemini taklit ederek veri öğrenme süreçlerini modelleyen bir algoritma türüdür. Bu algoritma, giriş verilerini çok katmanlı bir yapı aracılığıyla işler ve her katman, önceki katmandan aldığı veriyi daha karmaşık bir hale getirmektedir. Sinir ağları, ileri beslemeli ağlar (feedforward neural networks) ve geri yayılım algoritması (backpropagation) gibi teknikler aracılığıyla öğrenme sürecini gerçekleştirmektedir. Özellikle derin öğrenme (deep learning) alanında, derin sinir ağları (deep neural networks) kullanılarak büyük veri setleri üzerinde oldukça başarılı sonuçlar elde edilmiştir (LeCun ve ark., 2015). Görüntü tanıma, ses tanıma ve dil işleme gibi alanlarda yapay sinir ağları büyük başarılar göstermektedir.

2.1.4.4. K-En Yakın Komşu (K-Nearest Neighbors - KNN)

K-En Yakın Komşu algoritması (KNN), hem sınıflandırma hem de regresyon problemlerinde kullanılan basit ama etkili bir denetimli öğrenme algoritmasıdır. KNN, bir veri noktasını sınıflandırmak için o noktaya en yakın “k” komşuyu inceler ve bu komşuların sınıflarına bakarak yeni veri noktasının sınıfını belirlemektedir (Cover ve Hart, 1967). Bu algoritma, herhangi bir varsayıma dayanmadan çalışır ve hesaplamalar için veri noktalarının mesafesini ölçmektedir. KNN, özellikle küçük veri setlerinde etkili sonuçlar verirken, büyük veri setlerinde yavaş çalışabilmektedir. KNN'nin avantajı, kolay uygulanabilir olması ve parametre sayısının az olmasıdır.

2.1.4.5.Naive Bayes Algoritması

Naive Bayes, olasılıksal bir sınıflandırma algoritmasıdır ve özellikle metin sınıflandırma, spam filtreleme ve duygu analizi gibi uygulamalarda yaygın olarak kullanılmaktadır. Bu algoritma, her bir özelliğin bağımsız olduğunu varsayarak Bayes teoremini kullanır ve her bir sınıfın olasılığını hesaplamaktadır (McCallum ve Nigam, 1998). Naive Bayes'in en büyük avantajı, veriler arasında bağımsızlık varsayımına dayanmasına rağmen oldukça hızlı ve etkili olmasıdır. Olasılıksal yapısı sayesinde, yüksek boyutlu veri setlerinde dahi başarılı sonuçlar verebilmektedir.

2.1.4.6.Random Forest Algoritması

Sınıflandırma ve regresyon problemlerinde kullanılan bir topluluk (ensemble) öğrenme yöntemidir. Bu algoritma, birden fazla karar ağacının bir araya getirilmesiyle çalışır ve her ağaç farklı veri alt kümeleriyle eğitilmektedir. Nihai tahmin, her bir ağacın verdiği tahminlerin ortalaması veya oylaması ile belirlenmektedir (Breiman, 2001). Bu yapı, modelin aşırı öğrenme (overfitting) yapmasını önler ve genelleme kapasitesini artırmaktadır.

Random Forest, yüksek boyutlu ve karmaşık veri setlerinde güçlü performans göstermektedir. Rastgele örnekleme ve özellik seçimi ile modelin çeşitliliği artırılmaktadır. Ayrıca, "out-of-bag" (OOB) hata tahmini ile doğrulama seti olmadan hata oranı tahmin edilebilmektedir (Segal, 2004). Ancak, büyük veri setlerinde ve çok sayıda ağaç kullanıldığında hesaplama maliyetleri artabilmektedir (Geurts ve ark., 2006).

Random Forest, makine öğrenmesinde özellikle sınıflandırma problemlerinde tercih edilen güçlü ve dayanıklı bir algoritmadır. Modelin aşırı öğrenmeye karşı dayanıklı olması ve doğruluk oranlarının yüksek olması, çeşitli uygulama alanlarında geniş bir kullanım sağlamaktadır.

2.1.4.7.Derin Öğrenme (Deep Learning)

Derin öğrenme, yapay sinir ağlarının bir alt dalı olarak kabul edilen ve büyük veri setleri üzerinde yüksek doğrulukla çalışabilen bir algoritma türüdür. Derin öğrenme algoritmaları, çok katmanlı sinir ağları (deep neural networks) kullanarak verilerin karmaşık yapısını öğrenmektedir. Her katman, önceki katmandan aldığı bilgiyi daha

karmaşık bir biçimde işler ve bu katmanların derinliği arttıkça modelin öğrenme kapasitesi de artmaktadır. Özellikle görüntü tanıma, ses tanıma, doğal dil işleme gibi alanlarda derin öğrenme büyük başarılar göstermiştir (LeCun ve ark., 2015). Derin öğrenme algoritmalarının başarısı, yüksek boyutlu verilerdeki örüntüleri öğrenme kapasitesine ve güçlü GPU'lar kullanarak büyük veri setlerini işleyebilme yeteneğine dayanmaktadır. En yaygın kullanılan derin öğrenme algoritmaları şunlardır:

Evrişimsel Sinir Ağları (Convolutional Neural Networks - CNNs): Özellikle görüntü tanıma ve işleme alanında kullanılan CNN'ler, evrişim katmanları aracılığıyla görüntülerdeki yerel özellikleri öğrenir ve yüksek doğrulukta tahminler yapmaktadır (Krizhevsky ve ark., 2012).

Tekrarlayan Sinir Ağları (Recurrent Neural Networks - RNNs): Zaman serisi verilerini veya ardışık verileri işlemek için kullanılan RNN'ler, özellikle doğal dil işleme ve konuşma tanıma gibi alanlarda etkilidir. RNN'lerin bir alt türü olan LSTM (Long Short-Term Memory) ağları, uzun süreli bağımlılıkları öğrenme kapasitesi sayesinde başarılı sonuçlar elde etmektedir (Hochreiter ve Schmidhuber, 1997).

2.1.4.8.Uzun Kısa Süreli Bellek (LSTM)

LSTM, zamana bağlı verilerde uzun vadeli bağımlılıkları öğrenmek için tasarlanmış bir tür RNN modelidir. Geleneksel RNN'ler kısa süreli bağımlılıkları öğrenebilirken, LSTM'ler hücre durumu ve kapı mekanizmaları sayesinde uzun süreli bağımlılıkları da öğrenebilmektedir (Hochreiter ve Schmidhuber, 1997). Formula 1 yarış verilerinde, geçmiş performans verilerine dayanarak sürücü veya takım performansının zaman içindeki değişimlerini analiz etmek için kullanılabilmektedir. LSTM'ler, özellikle zaman serisi verilerinde gelecekteki olayları tahmin etmek için güçlü bir araçtır (Graves, 2012).

2.1.4.9.Meta-Sezgisel Algoritmalar (Metaheuristic Algorithms)

Meta-sezgisel algoritmalar, karmaşık ve büyük çözüm uzaylarına sahip problemleri çözmek için kullanılan genel amaçlı optimizasyon yöntemleridir. Bu algoritmalar, tam çözüm bulmak yerine iyi bir çözüm elde etmek amacıyla sezgisel yaklaşımlar kullanmaktadır. Meta-sezgisel algoritmaların temel amacı, küresel en iyi çözüme ulaşmak için yerel en iyi çözümlerden kaçınarak büyük veri setlerinde veya karmaşık

problemlerde optimizasyon sağlamaktır (Blum ve Roli, 2003). Yaygın kullanılan meta-sezgisel algoritmalar şunlardır:

Genetik Algoritmalar (Genetic Algorithms): Biyolojik evrim sürecinden esinlenen optimizasyon algoritmalarıdır. Bu algoritmalar, doğal seleksiyon ve genetik çaprazlama prensiplerine dayanarak en iyi çözümü bulmayı hedeflemektedir. Genetik algoritmalar, çözüm uzayında arama yaparken bir popülasyon kullanır ve bu popülasyonu her döngüde daha iyi bireyler üretmek için geliştirmektedir (Holland, 1992). Genetik algoritmalar, çok karmaşık ve doğrusal olmayan problemleri çözmek için uygundur ve mühendislik, ekonomi, robotik gibi birçok alanda kullanılmaktadır.

Parçacık Sürü Optimizasyonu (Particle Swarm Optimization - PSO): PSO, bir grup parçacığın çözüm uzayında hareket ederek küresel optimuma ulaşmaya çalıştığı bir algoritmadır. Bu algoritma, her bir parçacığın kendi ve diğer parçacıkların en iyi çözümünü dikkate alarak sürekli güncellenmektedir (Kennedy ve Eberhart, 1995).

Simulated Annealing: Bu algoritma, metalin soğutulma sürecini taklit eder ve yüksek sıcaklıkta rastgele aramalar yaparak en iyi çözümü bulmayı amaçlamaktadır. Soğutma işlemi ilerledikçe arama alanı daralmakta ve küresel optimuma ulaşma olasılığı artmaktadır (Kirkpatrick ve ark., 1983).

Meta-sezgisel algoritmalar, mühendislik, lojistik, finans ve biyoinformatik gibi birçok alanda geniş bir uygulama yelpazesine sahiptir. Karmaşık ve çok boyutlu problemlerin çözümünde etkili oldukları için özellikle optimizasyon süreçlerinde tercih edilmektedirler.

YZ alanında kullanılan algoritmalar, veri setlerinin türüne ve problemin yapısına göre farklılık göstermektedir. Karar ağaçları, SVM, yapay sinir ağları, KNN, Naive Bayes ve genetik algoritmalar, günümüzde en yaygın kullanılan algoritmalar arasında yer almakta ve her biri belirli avantajlara sahip olmaktadır. Yapay zeka sistemlerinin başarılı bir şekilde geliştirilmesi, bu algoritmaların doğru bir şekilde uygulanmasına bağlıdır.

2.1.5. Yapay Zekanın Avantajları ve Dezavantajları

YZ'nın çeşitli endüstrilerde ve alanlarda büyük yenilikler getiren bir teknoloji olarak kabul edilmekte olup, birçok avantajı beraberinde getirmiştir. Bununla birlikte, bu teknolojinin getirdiği bazı riskler ve dezavantajlar da mevcuttur. Aşağıda, yapay zekanın sunduğu başlıca avantajlar ve karşılaşılan dezavantajlar detaylandırılmaktadır.

2.1.5.1.Yapay Zekanın Avantajları

Verimlilik ve Hız: YZ insan yeteneklerinin çok ötesinde bir hız ve doğrulukla veri işleme kapasitesine sahiptir. Büyük veri kümelerinin analiz edilmesi, karmaşık hesaplamaların gerçekleştirilmesi ve büyük çaplı görevlerin tamamlanması gibi süreçlerde yapay zeka teknolojileri oldukça verimlidir (Russell ve Norvig, 2020). Örneğin, derin öğrenme algoritmaları, birkaç dakika veya saat içinde büyük görüntü veri kümelerini işleyebilir ve sonuçlar çıkarabilmektedir. Bu tür süreçlerin insan tarafından yapılması uzun zaman alacak ve hata riski barındıracaktır.

Karar Verme Süreçlerinde Nesnellik: YZ, belirli algoritmalar ve veri setlerine dayanarak karar vermektedir. İnsan hatasından etkilenmemesi ve duygusal unsurlardan bağımsız çalışması nedeniyle YZ, daha nesnel ve tutarlı sonuçlar sunmaktadır. Bu, özellikle finans, tıp ve hukuk gibi alanlarda kritik kararlarda insan hatasının azaltılmasında büyük önem taşımaktadır (Goodfellow ve ark., 2016).

Kesintisiz Çalışma Yeteneği: YZ sistemleri, insanlar gibi dinlenmeye veya molaya ihtiyaç duymadan sürekli olarak çalışabilmektedir. Bu özellik, özellikle üretim ve hizmet sektörlerinde yapay zeka uygulamalarını cazip hale getirmektedir. Üretim hatlarında kullanılan robotlar veya müşteri hizmetleri alanında kullanılan yapay zeka tabanlı sohbet botları, 7/24 hizmet sunarak operasyonların devamlılığını sağlamaktadır.

Özelleştirilmiş Deneyimler: YZ, kullanıcılara kişiselleştirilmiş deneyimler sunarak hizmet kalitesini artırabilmektedir. Örneğin, öneri sistemleri, kullanıcı davranışlarını analiz ederek bireysel tercihlere göre ürün ve hizmetler önerebilmektedir. Amazon, Netflix gibi büyük platformlar, yapay zeka algoritmalarını kullanarak müşteri davranışlarını analiz etmekte ve kişiye özel tavsiyeler sunmaktadır (Ricci ve ark., 2011).

2.1.5.2.Yapay Zekanın Dezavantajları

İşsizlik ve Ekonomik Eşitsizlik: Yapay zekanın en büyük dezavantajlarından biri, insan iş gücünün yerini almasıyla ortaya çıkan işsizlik riskidir. YZ tabanlı otomasyon sistemleri, özellikle tekrarlayan ve rutin görevlerde çalışan işçilerin yerini alarak iş kayıplarına yol açmaktadır. Üretim, lojistik ve müşteri hizmetleri gibi sektörlerde birçok çalışan, yapay zeka sistemleri tarafından yapılan görevleri yerine getiremeyecek duruma gelmiştir (Brynjolfsson ve McAfee, 2014). Bu durum, toplumda ekonomik eşitsizliklerin artmasına ve sosyal problemlerin ortaya çıkmasına neden olabilmektedir.

Gizlilik ve Güvenlik Riskleri: YZ sistemlerinin büyük miktarda veri üzerinde çalışması, gizlilik ve güvenlik risklerini de beraberinde getirmektedir. Yapay zeka sistemleri, kullanıcıların kişisel verilerini analiz ederken, bu verilerin kötüye kullanılması veya sızdırılması durumunda büyük güvenlik ihlalleri yaşanabilmektedir (Bostrom ve Yudkowsky, 2014). Ayrıca, YZ'nin siber saldırılara karşı savunmasız olması, kritik altyapılarda güvenlik açıklarına yol açabilmektedir.

Ayrımcılık ve Önyargılar: YZ sistemleri, eğitildikleri veri setlerinden öğrendiği için, veri setlerindeki önyargılar ve ayrımcılıklar da sonuçlara yansıyabilmektedir. Örneğin, bir YZ modeli cinsiyet, ırk veya sosyoekonomik durumlara dayalı olarak önyargılı kararlar verebilmektedir. Bu durum, özellikle işe alım süreçleri, kredi başvuruları ve yargı süreçlerinde ciddi adaletsizliklere yol açabilmektedir (O'Neil, 2016). Yapay zekanın tarafsız olması beklenirken, yanlış veya önyargılı veri setleriyle eğitildiğinde etik sorunlar ortaya çıkabilmektedir.

Maliyet ve Uygulama Zorlukları: YZ sistemlerinin geliştirilmesi, uygulanması ve bakımı genellikle yüksek maliyetlidir. Gelişmiş yapay zeka modellerinin eğitilmesi, yüksek miktarda veri, güçlü donanım ve deneyimli uzmanlar gerektirmektedir. Ayrıca, yapay zekanın uygulanacağı alanların ihtiyaçlarına uygun hale getirilmesi ve adapte edilmesi uzun zaman ve kaynak gerektirmektedir. Bu nedenle, küçük ve orta ölçekli işletmeler için yapay zeka çözümlerinin benimsenmesi zor olabilmektedir (Russell ve Norvig, 2020).

Yapay Zekanın Kontrol Edilemezliği Riski: Yapay zekanın gelişimi, bazı durumlarda kontrol edilemez hale gelebilmektedir. Yapay zekanın belirli etik ve sosyal

kurallar çerçevesinde hareket etmesi gerekmektedir, ancak özerklik kazanan yapay zeka sistemlerinin insan kontrolünden çıkma riski vardır. Bu, özellikle askeri ve savunma alanlarında kullanılan otonom yapay zeka sistemleri için ciddi bir risk oluşturmaktadır (Bostrom, 2014).

YZ, büyük fırsatlar ve avantajlar sunarken, beraberinde getirdiği riskler ve dezavantajlar da göz ardı edilmemelidir. Yapay zekanın doğru bir şekilde geliştirilmesi ve etik prensipler doğrultusunda kullanılması hem toplumsal hem de ekonomik açıdan denge sağlamak için önemlidir.

2.2. Formula 1

Formula 1 (F1), dünyanın en prestijli motor sporları şampiyonalarından biri olarak kabul edilmektedir. 1950 yılında başlayan bu spor, hız, teknoloji ve mühendislik yenilikleri ile tanınmaktadır. F1, her yıl farklı ülkelerde gerçekleştirilen Grand Prix yarışları ile milyonlarca kişiye ulaşmakta ve küresel bir izleyici kitlesine hitap etmektedir. Bu spor, sadece rekabetin yoğun olduğu bir yarış serisi değil, aynı zamanda ileri mühendislik uygulamalarının test edildiği ve geliştirildiği bir platformdur. F1'in yüksek hızlara sahip araçları ve ünlü pilotları, onu küresel spor arenasının en popüler etkinliklerinden biri haline getirmiştir (Feature, n.d.).

2.2.1. Formula 1'in Tarihçesi

Formula 1'in kökenleri, İkinci Dünya Savaşı sonrası döneme dayanmaktadır. İlk dünya şampiyonası 1950 yılında İngiltere'nin Silverstone pistinde düzenlenmiş ve bu tarihten itibaren her yıl düzenli olarak gerçekleştirilmektedir. Bu yıllarda Ferrari, Maserati ve Mercedes-Benz gibi markalar öne çıkmış, Juan Manuel Fangio gibi sürücüler ise ilk dönemin efsane pilotları arasında yer almıştır (Donaldson, 2012). Fangio, 1950'lerde beş dünya şampiyonluğu kazanarak bu sporun erken dönemlerine damgasını vurmuştur.

Zamanla F1, hem teknoloji hem de güvenlik açısından büyük ilerlemeler kaydetmiştir. 1970'li yıllarda risklerin yüksek olduğu bir dönem yaşanmış olsa da, 1980'lerden itibaren güvenlik standartları geliştirilmiş ve sürücüler için daha güvenli

koşullar oluşturulmuştur. 21. yüzyılda ise F1, sürdürülebilirlik hedefleri doğrultusunda büyük bir dönüşüm yaşamaktadır. 2014'ten itibaren hibrit motor sistemleri kullanılmış, 2026 yılında tamamen sürdürülebilir yakıtlarla yarışılması hedeflenmiştir (Formula 1, 2023).

2.2.2. Takımlar ve Sürücüler

Formula 1'de takımlar ve sürücüler, bu sporun kalbini oluşturmaktadır. Her sezon, takımlar ve sürücüler arasındaki mücadele, şampiyonluk yarışının en belirleyici faktörlerinden biri olarak öne çıkmaktadır. F1'de yarışan her takım, iki araçla katılır ve bu araçları sürmekle görevli iki ana sürücüye sahiptir. Takımların performansları, yalnızca sürücülerin yeteneklerine değil, aynı zamanda araçların teknik donanımına, mühendislik yeniliklerine ve stratejik planlamaya bağlıdır (Feature, n.d.).

2.2.2.1.Önde Gelen Takımlar

Tarih boyunca Ferrari, Mercedes, McLaren ve Red Bull Racing gibi takımlar, Formula 1'in en başarılı ve tanınan ekipleri arasında yer almıştır. Ferrari, F1'in en köklü takımı olarak kabul edilmekte olup, 1950'den bu yana yarışmaktadır. Ferrari'nin efsaneleşmiş pilotları arasında Michael Schumacher, 2000'li yıllarda elde ettiği yedi dünya şampiyonluğu ile dikkat çekmiştir (Red Bull, n.d.).

Mercedes, 2014 yılında hibrit motor teknolojisinin devreye girmesiyle beraber Formula 1'de baskın bir güç haline gelmiştir. Lewis Hamilton, Mercedes takımıyla kazandığı yedi dünya şampiyonluğu ile sporun en başarılı pilotlarından biri olarak tarihe geçmiştir. 2021 sezonunda Red Bull Racing ile Max Verstappen, yeni bir şampiyonluk dönemi başlatarak Mercedes'in üstünlüğüne son vermiştir (Red Bull, n.d.).

Red Bull Racing, özellikle son yıllarda gösterdiği performansla öne çıkmıştır. Verstappen'in genç yaşta kazandığı başarılar, Red Bull'un aerodinamik mühendisliği ve yarış stratejilerinin ne denli başarılı olduğunu göstermektedir (Red Bull, n.d.). McLaren ise geçmişte Ayrton Senna ve Alain Prost gibi sürücülerle büyük başarılarla imza atmış olup, günümüzde eski gücünü geri kazanma hedefiyle yarışmaya devam etmektedir.

2.2.2.2.Sürücüler: Yetenek ve Strateji

F1 sürücüler, yalnızca hız değil, aynı zamanda dayanıklılık, strateji ve takım çalışması gerektiren bir spor dalında yarışmaktadır. Bir Formula 1 sürücüsünün refleksleri ve karar verme yetenekleri, yarışın gidişatını doğrudan etkileyebilir. Örneğin, Michael Schumacher ve Ayrton Senna gibi pilotlar, yalnızca hızlarıyla değil, aynı zamanda yarış sırasında gösterdikleri stratejik zekâ ve adaptasyon yetenekleriyle de efsaneleşmiştir (Donaldson, 2012).

Günümüzün önde gelen sürücülerinden biri olan Lewis Hamilton, yalnızca yarış becerileriyle değil, aynı zamanda sürüş dışındaki sosyal sorumluluk projeleriyle de dikkat çekmektedir. Hamilton, sporun çeşitliliği artırma yönündeki çabaları ile spor tarihinin en ikonik figürlerinden biri haline gelmiştir.

Max Verstappen ise 2021 ve sonrasında kazandığı dünya şampiyonlukları ile yeni nesil F1 sürücüler arasında lider konumundadır. Sürüş stilineki agresiflik ve risk alabilme yetisi, onun yarışlardaki başarısının temel unsurlarındandır (Red Bull, n.d.).

2.2.2.3.Sürücüler ve Takım İçi Rekabet

Takımlar genellikle iki sürücü ile yarışır ve bu iki sürücü arasında takım içi rekabet kaçınılmazdır. Geçmişte Ayrton Senna ve Alain Prost'un McLaren'da yaşadığı rekabet, F1 tarihinde unutulmaz anılardan biri olarak kalmıştır (Donaldson, 2012). Benzer şekilde, Lewis Hamilton ve Nico Rosberg arasında Mercedes'te yaşanan mücadele de aynı dönemde büyük yankı uyandırmıştır. Bu tür rekabetler, takım dinamiklerini etkileyebilir ve yarış sonuçları üzerinde büyük bir etki yaratabilir.

Takım içi rekabet, bazen şampiyonluk mücadelelerinde belirleyici bir faktör olabilmektedir. Takımlar, sürücüler arasındaki dengeyi iyi yönetmek zorundadır. Bir yandan takım içi birlik sağlanırken, diğer yandan iki sürücü de en üst seviyede performans göstermek zorundadır (Bell ve ark., 2016).

Formula 1'de takımlar ve sürücüler, sporun temel yapı taşlarını oluşturmaktadır. Her sezon, teknolojik gelişmelerin yanı sıra sürücülerin bireysel yetenekleri ve takım stratejileri, yarışların sonucunu belirlemektedir. Ferrari, Mercedes ve Red Bull gibi takımların başarıları, tarihe geçmiş sürücülerin performanslarıyla birleşerek Formula 1'in büyüleyici dünyasını oluşturmuştur. Aynı zamanda, takım içi rekabetler ve stratejik

planlamalar, F1'in dinamik ve heyecan verici yapısının bir parçası olmaya devam etmektedir (Dodgins & Webber, 2023).

2.2.3. Formula 1'in Yapısı ve Kuralları

Formula 1, dünyanın en prestijli motor sporları serilerinden biri olarak kabul edilmektedir ve her yıl belirli bir şampiyona formatı dahilinde düzenlenmektedir. Bu format, sezon boyunca ortalama 20 ila 24 yarışın gerçekleştirilmesi ile tamamlanmaktadır. Bu yarışlar, dünyanın dört bir yanındaki ikonik pistlerde düzenlenir ve her bir yarış genellikle 305 kilometrelik bir mesafeyi kapsar. Yarış hafta sonları, serbest antrenmanlar, sıralama turları ve ana yarış olmak üzere birkaç aşamadan oluşmaktadır. Sezon sonunda, en yüksek puanları toplayan sürücü ve takım, dünya şampiyonluğu unvanını elde etmektedir (Feature, n.d.)

Yarışların düzenlendiği pistler, farklı uzunluklar ve zorluk derecelerine sahip olabilir. Örneğin, Monako Grand Prix'si dar ve kıvrımlı sokak pistleri ile ünlüken, İtalya'nın Monza Pisti uzun düzlükleri ve yüksek hızlara olanak tanıyan yapısıyla dikkat çeker. Pistler arasındaki bu çeşitlilik, her yarışın kendine özgü bir strateji gerektirmesine neden olur, bu da Formula 1'in cazibesini artıran unsurlardan biridir.

2.2.3.1.Puanlama Sistemi ve Stratejik Yenilikler

Formula 1'de sürücüler ve takımlar, yarışlarda gösterdikleri performans doğrultusunda puan kazanmaktadır. Yarış sonunda, ilk on sırada bitiren sürücüler belirli puanlar alır ve bu puanlar sezon sonunda genel klasman şampiyonunu belirler. 2022 yılında yapılan düzenlemelerle, yarış esnasında en hızlı turu atan sürücüye ek puan verilmesi kararlaştırılmıştır. Bu yenilik, takımlar ve sürücüler için stratejik bir unsura dönüşmüş, özellikle şampiyonluk mücadelesi sırasında bu puanlar kritik önem kazanmıştır (Racecar Engineering, n.d.).

F1, sadece hız üzerine kurulu bir yarış serisi olmaktan öte, strateji ve takım çalışmasının ön planda olduğu bir spor dalıdır. Yarış esnasında kullanılan lastik stratejileri, yakıt yönetimi ve pit stop süreleri, yarışın sonucunu doğrudan etkileyen unsurlardır. Takımlar, bu stratejileri optimize edebilmek için yapay zeka ve veri analitiği

gibi ileri teknolojilerden faydalanmaktadır. Örneğin, her bir takımın pit duvarında bulunan strateji ekipleri, yarışın gidişatını sürekli olarak analiz eder ve sürücülere anlık geri bildirimler sunmaktadır (Marr, 2023).

2.2.3.2.Teknik Kurallar ve Araç Tasarımı

Formula 1 araçları, Uluslararası Otomobil Federasyonu (FIA) tarafından belirlenen katı teknik kurallara göre tasarlanır ve üretilir. Bu kurallar, aracın güvenliği, performansı ve çevresel etkileri ile ilgilidir. Örneğin, her sezon başında yapılan teknik düzenlemelerle araçların aerodinamik yapısında ve motor performansında önemli değişiklikler yapılır. 2022 sezonunda yapılan büyük aerodinamik düzenlemeler, araçların birbirine daha yakın yarışabilmesini ve geçişlerin artmasını sağlamak amacıyla gerçekleştirilmiştir. Bu düzenlemeler, daha rekabetçi ve seyir zevki yüksek yarışların ortaya çıkmasına katkıda bulunmuştur (FIA, 2023).

Özellikle hibrit motor teknolojisinin 2014 yılında tanıtılması, F1'de devrim niteliğinde bir değişiklik yaratmıştır. Hibrit motorlar, hem içten yanmalı motoru hem de elektrikli motorları birleştirerek enerji verimliliğini artırmakta ve çevresel etkileri azaltmaktadır. Bu sistem, enerji geri kazanım sistemleri (ERS) ile entegre edilmiştir. ERS, frenleme sırasında açığa çıkan enerjiyi depolayarak aracın performansını artırmak için kullanır. Formula 1 araçlarının bu teknolojilerle donatılması, enerji verimliliği ve sürdürülebilirlik konularında sporun ileriye dönük hedeflerine katkı sağlamaktadır.

Ayrıca, güvenlik önlemleri F1'de büyük bir öneme sahiptir. 1970'li yıllarda birçok ölümlü kazanın yaşanmasının ardından, FIA tarafından kapsamlı güvenlik reformları getirilmiştir. Bugün, araçlarda kullanılan "Halo" isimli koruma sistemi, sürücünün başını dış etkilerden koruyarak güvenlik standartlarını en üst seviyeye çıkarmaktadır. Bununla birlikte, yarış pisti çevresindeki güvenlik bariyerleri ve araç tasarımındaki çarpışma bölgeleri de güvenlik önlemleri kapsamında düzenlenmiştir (FIA, 2023).

2.2.3.3.Sürdürülebilirlik ve Çevresel Etkiler

Formula 1'in 2020'li yıllardan itibaren sürdürülebilirlik hedeflerine yönelik attığı adımlar, sporun geleceği açısından kritik öneme sahiptir. FIA, F1'in çevresel ayak izini azaltmak amacıyla 2026 yılı itibarıyla tamamen sürdürülebilir yakıtların kullanımını zorunlu hale getirecek yeni kurallar getirmiştir. Bu yeni düzenlemeler, karbon

emisyonlarını azaltmayı ve çevreye olan zararı minimuma indirmeyi hedeflemektedir. Aynı zamanda, hibrit motorların verimliliği artırılacak ve sporun enerji tüketimi daha sürdürülebilir bir hale getirilecektir (Formula 1, 2023).

Formula 1'in çevresel etkilerini azaltmak için yapılan bu girişimler, sporun sadece yarış performansına değil, aynı zamanda dünya genelindeki çevresel sorumluluklarına da ne kadar önem verdiğini göstermektedir. Bu bağlamda, takımlar ve üreticiler, mühendislik çözümlerinde daha çevre dostu teknolojiler geliştirmeye teşvik edilmektedir. Örneğin, McLaren ve Mercedes gibi takımlar, karbon ayak izlerini azaltma yönünde çeşitli çevresel inisiyatifler başlatmıştır (Formula 1, n.d.).

Formula 1'in yapısı ve kuralları, teknolojik gelişmelerin hızla uygulanabildiği, mühendislik inovasyonlarının test edilebildiği ve stratejik kararların yarışların kaderini belirlediği bir sistem üzerine kuruludur. Aerodinamik düzenlemeler, hibrit motor teknolojileri ve sürdürülebilirlik hedefleri, sporun evriminde önemli kilometre taşları olarak görülmektedir. Takımların mühendislik inovasyonları ve stratejik yaklaşımı ile sürücülerin bireysel yetenekleri, F1'i rekabetçi ve dinamik bir spor dalı haline getirmektedir. Hem teknik hem de çevresel standartların sürekli geliştiği Formula 1, küresel motor sporları sahnesinde her geçen gün daha fazla önem kazanmaktadır.

2.2.4. Formula 1'in Sosyoekonomik Etkisi

Formula 1, küresel çapta büyük bir ekonomik etkiye sahiptir. Yarışların düzenlendiği şehirler ve ülkeler, Formula 1 sayesinde turizm gelirlerinde büyük artışlar yaşamaktadır. Örneğin, 2023 yılında Suudi Arabistan Grand Prix'si, ülkeye 500 milyon dolar civarında ekonomik katkı sağlamıştır (Saudi Motorsport, n.d.). Yarışlar ayrıca, yerel ekonomilere ve altyapı gelişimine de katkıda bulunmaktadır.

Sponsorluk anlaşmaları ve televizyon hakları, F1'in ticari gelirlerinde büyük bir paya sahiptir. Örneğin, Red Bull Racing, 2023 sezonunda 650 milyon dolarlık bir bütçeyle yarışmış ve bu da takımın en yüksek harcamalara sahip olduğunu göstermektedir (F1 Business, 2023). Ayrıca, markalar, F1 takımları ile sponsorluk anlaşmaları yaparak küresel ölçekte tanınırlık kazanmaktadır (Red Bull, n.d.).

2.2.5. Teknolojik Gelişmeler ve Mühendislik İnovasyonları

Formula 1, mühendislik inovasyonlarının test edildiği ve geliştirildiği bir alan olarak kabul edilmektedir. Aerodinamik yapılar, motor güçleri ve elektronik kontrol sistemleri, araçların performansını artıran en önemli faktörler arasında yer almaktadır. 2022 yılında yapılan teknik düzenlemelerle birlikte, araçların aerodinamik performansı iyileştirilmiş ve sürücüler arasındaki mücadeleler daha yakın hale getirilmiştir (Formula 1, n.d.).

Hibrit motor teknolojisinin 2014 yılında devreye girmesi, çevre dostu yakıt kullanımını artırmış ve enerji verimliliğini yükseltmiştir. 2026 yılında yürürlüğe girmesi planlanan yeni kurallar çerçevesinde, tamamen sürdürülebilir yakıtların kullanılması hedeflenmektedir (Formula 1, 2023). Ayrıca, yapay zeka ve simülasyon teknolojileri, takımların yarış stratejilerini geliştirmesinde önemli bir rol oynamaktadır.

2.2.5.1. Formula 1’de Veri Analizi ve Yapay Zeka

Formula 1, yüksek hızda rekabetin yanı sıra veri analitiği ve yapay zeka (YZ) teknolojilerinin yoğun bir şekilde kullanıldığı bir spor dalıdır. Günümüzün gelişmiş teknolojileri, takımlara stratejik kararlar alma, performanslarını optimize etme ve sürüş dinamiklerini geliştirme konularında yardımcı olmaktadır. Veri analizi ve yapay zeka uygulamaları, F1 araçlarından toplanan büyük veri setlerinin işlenmesi ve bu verilerin yarış stratejileri, araç geliştirme ve sürücü performansını iyileştirmek amacıyla kullanılmasında kilit rol oynamaktadır (Autosport, n.d.).

2024 yılında yapay zeka (YZ) ve veri analitiğinin, yarış stratejilerinin ve araç performansının optimize edilmesinde devrim niteliğinde bir etki yaratmaya devam ettiği görülmektedir. Bu doğrultuda, Formula 1 takımlarının yarış dinamiklerini anlamak ve en uygun yarış stratejilerini geliştirmek için ileri düzey YZ algoritmalarını entegre ettiği belirtilmektedir. Bu entegrasyon, yalnızca yarış esnasında değil, sezon boyunca veri analitiği ile elde edilen iç görülerin kullanılmasını da kapsamaktadır. Böylece takımlar, sürücünün frenleme noktalarını, hızlanma stratejilerini ve rakip takımların hamlelerini analiz ederek avantaj sağlama konusunda büyük bir adım atmaktadır (GrandPrix247, 2024).

Veri Analizinin F1'deki Rolü: Formula 1'de her bir araç, yarış esnasında ve antrenmanlar sırasında sensörler aracılığıyla büyük miktarda veri toplamaktadır. Araç üzerindeki yüzlerce sensör, lastik sıcaklıklarından yakıt tüketimine, fren sistemlerinden aerodinamik yüklere kadar birçok veriyi gerçek zamanlı olarak kaydetmektedir. Bu veriler, takım mühendisleri tarafından analiz edilerek yarış stratejilerinin oluşturulmasına katkıda bulunmaktadır. Her bir yarış hafta sonunda yaklaşık 1.5 terabayt veri toplandığı ve bu verilerin anlık analizlerle yarış içi kararların verilmesine olanak sağladığı bilinmektedir (Marr, 2023).

Veri analizi, sadece yarış stratejilerini belirlemekle kalmaz, aynı zamanda araçların performansını optimize etmek ve sürücüler için en uygun ayarları belirlemek amacıyla da kullanılmaktadır. Örneğin, lastik aşınması veya yakıt verimliliği gibi faktörler, veri analizleri yardımıyla takip edilmekte ve sürücülere yarış esnasında bu verilere dayalı olarak geri bildirimler sunulmaktadır. Bu sayede hem yarış içi performans artırılmakta hem de araçların güvenliği en üst seviyede tutulmaktadır (Marr, 2023).

Yapay Zeka ve Makine Öğrenimi Uygulamaları: Yapay zeka, Formula 1'de giderek daha fazla kullanılmaya başlanmış bir teknolojidir. Yapay zeka, büyük veri setlerinin analiz edilmesi, gelecekteki yarış senaryolarının simüle edilmesi ve potansiyel sonuçların tahmin edilmesi için önemli bir araç haline gelmiştir. Takımlar, makine öğrenimi algoritmaları sayesinde geçmiş yarış verilerini analiz ederek gelecekteki performanslarını iyileştirebilecek stratejiler geliştirmektedir. Örneğin, bir yarışın başlangıç anından itibaren, hava koşullarının, pist sıcaklığının ve lastik stratejilerinin etkileri simüle edilmekte ve bu verilerle en iyi yarış senaryoları oluşturulmaktadır (Marr, 2023).

Makine öğrenimi, aynı zamanda sürücü performansının değerlendirilmesinde de kullanılmaktadır. Sürücülerin frenleme noktaları, viraj hızları ve diğer sürüş parametreleri yapay zeka tarafından analiz edilerek, sürücülere en verimli sürüş biçimi sunulabilmektedir. Bu analizler, sürücülerin hangi hataları yaptığını, hangi alanlarda gelişim gösterebileceğini belirlemekte de yardımcı olmaktadır. Örneğin, 2022 sezonunda Red Bull Racing, makine öğrenimi algoritmalarını kullanarak Max Verstappen'in lastik aşınmasını minimize edecek sürüş stratejileri geliştirmiştir (Motorsport.com, n.d.).

YZ'nin Formula 1'deki entegrasyonu, aynı zamanda aracın kurulumunun her yönünü etkileyen kararları otomatikleştirme sürecini de içermektedir. Sürüş sırasında anlık değişkenlere hızlıca uyum sağlama ve performansı optimize etme kabiliyeti, YZ uygulamalarının bu sporun ayrılmaz bir parçası haline gelmesine neden olmuştur. Araç ayarlarından yarış içi stratejilere kadar geniş bir alanda kullanılan YZ, veri analizleri sayesinde araçların en uygun ayarlara sahip olmasını sağlamaktadır (Seabiscuit.ai, 2024).

Yapay Zeka ve Yarış Stratejileri: Yapay zeka, yarış stratejilerinin belirlenmesinde de önemli bir rol oynamaktadır. Bir yarış hafta sonunda hava durumu, pist koşulları, rakip takımların stratejileri ve lastik seçimleri gibi faktörler sürekli değişkenlik göstermektedir. Bu noktada, yapay zeka algoritmaları, bu değişkenleri analiz ederek en uygun stratejiyi belirlemekte yardımcı olmaktadır. Örneğin, bir pit stop zamanlamasının optimize edilmesi ya da farklı lastik stratejilerinin seçilmesi, yapay zeka destekli analizler sayesinde daha isabetli hale gelmiştir (Autosport, n.d.).

Ayrıca, simülasyon yazılımları aracılığıyla yapay zeka, farklı yarış senaryolarını gerçek zamanlı olarak simüle edebilmektedir. Bu simülasyonlar, takımların olası yarış senaryolarına önceden hazırlıklı olmasını sağlamaktadır. Mercedes ve Ferrari gibi üst düzey F1 takımları, bu simülasyonlarla gelecekteki yarışlar için stratejik hazırlıklarını yapmaktadır. Özellikle değişken hava koşullarının olduğu pistlerde, yapay zeka destekli stratejiler, takımlara büyük avantajlar sağlayabilmektedir (Marr, 2023).

Ayrıca, YZ algoritmalarının matematiksel hesaplamalar ile birleşerek yarış analizlerinde daha hassas sonuçlar elde edilmesini sağladığı belirtilmektedir. Bu sistemler, lastik aşınma modelleri, yakıt tüketimi, hava koşullarının etkisi gibi değişkenleri değerlendirerek optimum yarış stratejilerini belirlemekte yardımcı olmaktadır. Bu sayede, takımların veri odaklı ve anlık analizlerle karar alması mümkün olmaktadır (GrandPrix247, 2024).

YZ uygulamaları, yalnızca yarış stratejileri ve araç performansı ile sınırlı kalmamakta, aynı zamanda Formula 1'in dünya çapındaki hayran kitlesiyle etkileşim kurma sürecinde de kullanılmaktadır. YZ ve veri analitiği, sosyal medya platformları ve resmi Formula 1 kanalları aracılığıyla taraftarlarla etkileşimi artırmak ve kişiselleştirilmiş içerik sunmak için kullanılmaktadır. Bu uygulamalar, izleyicilere özel analizler ve yarış içi güncellemeler sağlama konusunda önemli bir rol oynamaktadır (Euronews, 2024).

Veri Analizi ve Yapay Zeka ile Gelecekteki İnovasyonlar: Formula 1’de veri analizi ve yapay zeka kullanımının önümüzdeki yıllarda daha da yaygınlaşması beklenmektedir. Araçların sensör sistemleri ve toplanan veri miktarı her geçen yıl artarken, bu verilerin analizi de daha sofistike hale gelmektedir. Özellikle yapay zeka ve makine öğrenimi algoritmalarının daha gelişmiş versiyonları, yarış performansını tahmin etmekte ve stratejik kararlar almakta daha etkin sonuçlar verecektir.

Gelecekte yapay zeka ve veri analitiğinin F1’deki rolünün daha da artacağı öngörülmektedir. Bu teknolojilerin yalnızca yarış performansını artırmakla kalmayıp, aynı zamanda sürdürülebilirlik hedeflerine ulaşmak için de kullanılması beklenmektedir. Örneğin, çevre dostu motor teknolojilerinin geliştirilmesinde, yapay zeka destekli mühendislik çözümleri önemli bir rol oynayacaktır.

Formula 1, hız ve teknolojinin bir araya geldiği bir spor dalıdır ve bu bağlamda veri analizi ile yapay zeka kullanımı, performans optimizasyonu ve yarış stratejileri için büyük önem taşımaktadır. Günümüzde her F1 takımı, topladığı büyük veri setlerini analiz ederek yarışlarda avantaj sağlamaya çalışmakta ve yapay zeka algoritmalarını kullanarak gelecekteki performansını iyileştirecek stratejiler geliştirmektedir. Teknolojinin hızla geliştiği bu alanda, yapay zeka ve veri analizinin F1’in geleceğinde daha da belirleyici olması beklenmektedir.

2.2.5.2. Formula 1’de Makine Öğrenimi Uygulamaları

Formula 1, günümüzün en teknolojik ve veri odaklı sporlarından biri olarak, makine öğrenimi (MÖ) uygulamalarını kapsamlı bir şekilde kullanmaktadır. MÖ, büyük veri setlerini analiz etme ve bu verilerden öğrenme yeteneği sayesinde takımların performanslarını önemli ölçüde artırmalarına olanak tanımaktadır. Her yarışta toplanan veriler, araçların sensörleri aracılığıyla elde edilir; bu sensörler motor sıcaklığı, lastik aşınması, aerodinamik veriler, yakıt tüketimi ve sürücünün frenleme noktaları gibi birçok parametreyi gerçek zamanlı olarak kaydetmektedirler (Chaudhary, 2023).

MÖ algoritmaları, bu verileri analiz ederek takımların en uygun yarış stratejilerini geliştirmesine yardımcı olmaktadır. Örneğin, Ferrari, lastik aşınması tahminleri yapmak için MÖ uygulamalarını kullanmakta ve bu sayede pit stop zamanlamalarını optimize etmektedir. Bu tür uygulamalar, takımların yarış sırasında anlık kararlar alabilmesini ve rakiplerine karşı stratejik avantaj elde etmesini sağlamaktadır (Marr, 2023).

MÖ teknikleri arasında regresyon analizi, karar ağaçları, yapay sinir ağları ve destek vektör makineleri gibi yöntemler yer almaktadır. Bu yöntemler, geçmiş verilerden öğrenerek gelecekteki performans tahminlerini geliştirmekte ve sürücülerin, araçların ve pist koşullarının etkileşimlerini analiz etmektedir. Böylece takımlar, verimliliklerini artıracak ayarlamalar yapma konusunda daha bilinçli kararlar alabilmektedir (Autosport, n.d.).

2.2.5.3. Yapay Zekanın Performans Tahmini ve Strateji Oluşturmada Kullanımı

Yapay zeka (YZ), Formula 1’de performans tahmini ve strateji oluşturma süreçlerinde merkezi bir rol oynamaktadır. YZ algoritmaları, geçmiş yarış verilerinden öğrenerek sürücülerin ve araçların performansını tahmin etme yeteneğine sahiptir. Bu tahminler, takımların hangi stratejiyi izlemesi gerektiği konusunda karar vermelerine yardımcı olmaktadır (Marr, 2023).

Özellikle hava koşulları, pist yüzeyi ve rakiplerin stratejileri gibi değişkenlerin etkilerini analiz etmek, YZ’nin en büyük avantajlarından biridir. Örneğin, takımlar, geçmiş verilere dayalı olarak hava durumu tahminlerini inceleyebilir ve bu bilgiler ışığında lastik stratejilerini belirleyebilmektedir. YZ destekli analizler, pit stop zamanlamaları ve yakıt yönetimi gibi stratejik kararların alınmasında kritik öneme sahiptir. Red Bull Racing, bu tür YZ uygulamaları ile pit stop zamanlamalarını optimize ederek, yarış sonuçlarını olumlu yönde etkilemiştir (Motorsport.com, n.d.).

YZ’nin sağladığı öngörüler, yarış sırasında gerçek zamanlı strateji güncellemelerine olanak tanımaktadır. Takımlar, YZ algoritmalarını kullanarak sürekli olarak verileri analiz etmekte ve bu veriler doğrultusunda yarış stratejilerini anlık olarak değiştirebilmektedir. Bu, sürücülerin en iyi performansı göstermelerini sağlamak için kritik bir unsurdur.

2.2.5.4. Yapay Zekanın Sürücü Performansını Değerlendirmedeki Rolü

Yapay zeka, sürücü performansını değerlendirmede önemli bir etkiye sahiptir. YZ algoritmaları, sürücülerin frenleme noktaları, viraj alma hızları ve genel sürüş dinamikleri gibi parametreleri analiz ederek hangi alanlarda gelişim gösterilmesi gerektiğini

belirlemektedir. Bu tür analizler, sürücülerin teknik becerilerini artırmalarına ve yarış esnasındaki performanslarını iyileştirmelerine yardımcı olmaktadır (Autosport, n.d.).

YZ, sürücülerin psikolojik durumlarını izleme yeteneği ile de dikkat çekmektedir. Yarış sırasında sürücülerin stres düzeyleri ve konsantrasyonları, araç içi veri toplama sistemleri ve YZ algoritmaları ile analiz edilmekte, bu veriler ışığında sürücülerin performansları hakkında değerlendirmeler yapılmaktadır. Örneğin, Mercedes takımı, sürücülerinin ruh halini ve konsantrasyon düzeyini izleyerek, hangi yarışlarda daha fazla destek gerektiğini belirlemektedir (Chaudhary, 2023).

Bu tür değerlendirmeler, sürücülerin yarış sırasında en iyi performansı sergilemelerini sağlamak için kritik öneme sahiptir. Sürücülerin gerçek zamanlı olarak alınan verilerle geri bildirim alması, performanslarını artırmak için gerekli ayarlamaları yapmalarına olanak tanımaktadır.

2.2.5.5.Yapay Zeka ile Yarış Simülasyonları

Yapay zeka, yarış simülasyonlarında da önemli bir rol oynamaktadır. YZ algoritmaları, farklı yarış senaryolarını simüle ederek takımların hangi stratejilerin daha etkili olabileceğini analiz etmelerine yardımcı olmaktadır. Bu simülasyonlar, çeşitli hava koşulları, pist durumları ve rakip takımların stratejileri dikkate alınarak yapılmaktadır (Marr, 2023).

Simülasyonlar, takımların gelecekteki yarışlara hazırlık yapmalarını sağlamakta ve potansiyel sorunları önceden tespit etmelerine yardımcı olmaktadır. Örneğin, Mercedes takımı, YZ destekli simülasyonlar kullanarak farklı pit stop senaryolarını test etmekte ve bu testler sayesinde daha etkili stratejiler geliştirmektedir. Bu simülasyonlar, sürücülerin ve mühendislerin yarış koşullarında nasıl hareket etmeleri gerektiğine dair iç görüler elde etmelerine olanak tanımaktadır (ITPro, n.d.).

Ayrıca, bu simülasyonlar, farklı sürüş stillerinin etkilerini analiz etmekte ve takımların sürücülerine en uygun ayarları bulmalarına yardımcı olmaktadır. YZ algoritmaları, simülasyon verilerini analiz ederek, sürücülerin hangi ayarlarla daha iyi performans göstereceklerini tahmin etmektedir.

2.2.5.6. Formula 1’de Yapay Zekanın Geleceđi

Formula 1’de yapay zekanın geleceđi, hızla gelişen teknolojilerle birlikte büyük bir potansiyele sahiptir. YZ ve makine öğrenimi uygulamalarının daha da yaygınlaşması beklenmektedir. Özellikle veri analitiđi ve YZ tekniklerinin daha sofistike hale gelmesi, takımların performanslarını artırmalarına ve sürdürülebilirlik hedeflerine ulaşmalarına olanak tanıyacaktır (Marr, 2023).

Yapay zeka, takımların daha çevre dostu motor teknolojileri ve yakıt verimliliđi sağlamalarına yardımcı olacaktır. Formula 1, bu alandaki inovasyonları benimseyerek, daha sürdürülebilir bir spor dalı olma yolunda önemli adımlar atmaktadır (Autosport, n.d.). Örneđin, takımların, yapay zekayı kullanarak gelişen çevresel standartlara uyum sağlaması ve rekabet avantajlarını artırması, gelecekteki yarışların şekillenmesinde kritik bir rol oynayacaktır.

YZ’nin uygulamaları, hayranların deneyimlerini de zenginleştirecek şekilde genişleyecektir. Gerçek zamanlı veri analizi ve görselleştirme teknikleri sayesinde, izleyicilere daha fazla bilgi sunulacak ve yarış deneyimleri daha etkileşimli hale gelecektir. Böylece, Formula 1, sadece bir yarış deđil, aynı zamanda teknoloji ve inovasyonun bir sergisi haline gelecektir.

3. GEREÇ VE YÖNTEM

Bu tezde, mevcut veriler kullanılarak Formula 1 yarışlarının sonuçlarının tahmin edilmesi amaçlanmaktadır. Tahminler, farklı yöntemlerle elde edilen sonuçların karşılaştırılması yoluyla değerlendirilecektir. Bu karşılaştırmalar için veri setine mafese yöntemiyle öznitelik seçimi uygulanmış ve farklı optimizeler kullanılarak çeşitli öznitelik grupları elde edilmiştir. Elde edilen bu öznitelikler, LSTM, SVM, Karar Ağaçları (DT), Random Forest (RF) ve Naive Bayes (NB) algoritmalarına dâhil edilerek tahminleme işlemleri gerçekleştirilmiştir. Aynı zamanda, tüm veri seti kullanılarak da benzer işlemler yapılmış ve sonuçlar arasında en iyi performans gösteren model belirlenmeye çalışılmıştır.

Veri setinde tutarsızlık sorununun giderilmesi amacıyla SMOTE yöntemi kullanılmıştır. Bu projede Formula 1'in hibrit motorlara geçtiği yıla ait veriler esas alınmıştır. Eğitim veri seti 2014-2022 yıllarını, test veri seti ise 2023 yılını kapsamaktadır. Bu yöntemle veri setindeki tutarsızlığın azaltılması hedeflenmiştir.

Tutarsızlıkların giderilmesinin ardından, veri setini proje için hazırlamak amacıyla kullanılan kütüphaneler aşağıda açıklanmaktadır:

Scikit-learn: Scikit-learn (genellikle sklearn olarak kısaltılır), Python programlama dilinde geliştirilmiş ve açık kaynak kodlu bir makine öğrenimi kütüphanesi olarak tanıtılmaktadır. Makine öğrenimi, veri madenciliği ve veri analizi alanlarında geniş bir kullanım yelpazesi sunan scikit-learn, hem gözetimli (supervised) hem de gözetimsiz (unsupervised) öğrenme algoritmalarının uygulanmasına olanak sağlamaktadır. Kütüphanenin, veri ön işleme, model değerlendirme ve doğrulama gibi makine öğrenimi iş akışının önemli adımlarını destekleyen bir dizi özellik sunduğu belirtilmektedirler (Pedregosa ve ark., 2011).

Keras: Keras, açık kaynaklı bir derin öğrenme kütüphanesidir ve özellikle hızlı prototipleme ve kullanımı kolay bir arayüz sunmasıyla öne çıkmaktadır. TensorFlow ve Theano gibi arka uç kütüphanelerle entegre bir şekilde çalışarak, sinir ağları ve diğer makine öğrenimi modellerinin hızlı bir şekilde oluşturulmasını sağlar. Bu çalışmada, Keras özellikle LSTM (Uzun Kısa Süreli Bellek) gibi derin öğrenme modellerinin oluşturulması ve eğitilmesi amacıyla kullanılmıştır (Chollet, 2015).

TensorFlow: TensorFlow, Google tarafından geliştirilen ve makine öğrenimi ile derin öğrenme modellerinin eğitilmesi için kullanılan geniş kapsamlı bir kütüphanedir. TensorFlow, özellikle büyük veri setleri üzerinde yüksek performanslı hesaplamalar ve model eğitimi için optimize edilmiştir. Bu çalışmada, derin öğrenme modellerinin uygulanması ve optimize edilmesi sürecinde TensorFlow, Keras ile birlikte kullanılmaktadır (Abadi ve ark., 2016).

Pandas: Pandas kütüphanesi, yapılandırılmış veri manipülasyonu ve analizi için kullanılan güçlü bir Python kütüphanesidir. Pandas, özellikle büyük veri setleri üzerinde kolay veri temizleme, dönüştürme ve analiz işlemleri yapabilme kabiliyeti ile öne çıkmaktadır. Bu tezde, veri setlerinin işlenmesi, özelliklerin seçimi ve hazırlanması gibi veri manipülasyonu işlemleri için kullanılmıştır. Pandas, veri çerçeveleri ve seriler gibi veri yapıları ile çalışarak veri analizini kolaylaştırmaktadır (McKinney, 2010).

NumPy: NumPy kütüphanesi, çok boyutlu diziler ve matrisler üzerinde yüksek performanslı hesaplamalar gerçekleştirmek için tasarlanmış bir bilimsel hesaplama kütüphanesidir. NumPy, özellikle sayısal analiz ve vektörel hesaplamalar gerektiren işlemler için tercih edilmiştir. Veri kümelerinin verimli bir şekilde işlenmesi ve matematiksel işlemler gerçekleştirilmesi amacıyla bu çalışmada önemli bir araç olarak kullanılmıştır (Harris ve ark., 2020).

Matplotlib: Python'da iki boyutlu grafikler oluşturmak için kullanılan kapsamlı bir görselleştirme kütüphanesidir. Grafiklerin detaylı olarak özelleştirilmesine imkan veren Matplotlib, sonuçların görselleştirilmesi ve analizlerin sunulmasında etkili bir araç olarak kullanılmıştır. Çizgi grafikleri, histogramlar, dağılım grafikleri gibi görselleştirmeler Matplotlib aracılığıyla oluşturulmuş ve analiz sonuçları görselleştirilmiştir (Hunter, 2007).

Seaborn: İstatistiksel verilerin görselleştirilmesine odaklanan, Matplotlib üzerine inşa edilmiş bir yüksek seviyeli Python kütüphanesidir. Seaborn, özellikle veri dağılımlarının ve ilişkilerinin kolayca anlaşılmasını sağlayan zengin grafik türleri ile veri analizinde sıklıkla kullanılmaktadır. Tezde, istatistiksel analiz sonuçlarının daha anlaşılır hale getirilmesi için Seaborn grafiklerinden yararlanılmıştır (Waskom, 2021).

Smote: Veri setlerinde sınıf dengesizliği, makine öğrenmesi modellerinin doğruluğunu olumsuz yönde etkileyen önemli bir sorundur. Dengesiz veri setlerinde,

çoğunluk sınıfına ait örnekler daha fazla olduğunda model, azınlık sınıfını yeterince öğrenemez ve bu durum tahmin performansını düşürür. Bu çalışmada, dengesiz veri setindeki sınıf dağılımını dengelemek amacıyla **Synthetic Minority Over-sampling Technique (SMOTE)** yöntemi kullanılmıştır. SMOTE, azınlık sınıfına ait yeni örnekler oluşturarak veri setindeki dengesizliği azaltır. Böylece, sınıflar arasında daha dengeli bir dağılım sağlanır ve modelin her iki sınıfı da öğrenme yeteneği artar (Chawla ve ark., 2002).

Bu kütüphaneler, özellikle derin öğrenme algoritmalarının uygulanması, performanslarının analiz edilmesi ve sonuçlarının görselleştirilmesi açısından çalışmanın önemli bileşenlerini oluşturmaktadır.

3.1. Verilerin Toplanması

Bu tezde, Formula 1 resmi web sitesinden toplanan bilgileri bir araya getiren Ergast API'sinden alınan kapsamlı verilerden yararlanmaktadır. Ergast API'ı, 1950'den günümüze kadar yarış sonuçları, sürücü ve takım istatistikleri, tur süreleri, pit stoplar ve puan durumlarını içeren geniş bir bilgi yelpazesi sunmaktadır (Ergast, n.d.). Bu veri setini zenginleştirmek amacıyla, Races DataFrame'ine ek olarak yıllık toplam yarış mesafeleri eklenmiştir. Bu işlem, BeautifulSoup kullanılarak Wikipedia'dan kazıma yöntemiyle gerçekleştirilmiş ve ek veriler F1-Fansite.com'dan elde edilmiştir (F1 Fansite, n.d.). Yarış ismi değişkeni manipüle edilerek Selenium undetected chromedriver ve BeautifulSoup kullanılmış ve daha doğru veriler toplanmıştır. Bu birleşik veri noktaları, Formula 1 yarışlarını onlarca yıl boyunca kapsamlı bir şekilde analiz edebilmek için zenginleştirilmiş bir veri seti sunmaktadır (Wikipedia, n.d.). Bu elde edilen veri setinin içinde bulunan özniteliklerin isimleri Tablo 1'de gösterilmiştir.

Tablo 1: Ergast API ile indirilen öznitelikler

Races	Circuits	Drivers	Results	Qualifying
Season	Circuit ID	Driver ID	Season	Season
Round	Circuit Name	Name	Round	Round
Race Name	Latitude	Nationality	Circuit ID	Circuit ID
Circuit ID	Longitude	Driver Code	Driver ID	Driver ID

Latitude	Locality	Date of Birth	Finish Position	Qualifying Position
Longitude	Country		Date of Birth	Constructor
Country	Wikipedia URL		Nationality	Q1
Date			Constructor	Q2
Wikipedia URL			Grid	Q3
			Time	
			Status	
			Points	

3.2. Verilerin Temizlenmesi ve İşlenmesi

Bu projede, hava durumu verilerindeki eksiklikleri tamamlamak amacıyla, scraping yöntemleri kullanılarak f1-fansite.com'dan veri toplanmıştır. Selenium ile bulut güvenlik sistemleri aşılmış ve yıllara göre farklı URL yapıları taranmıştır. BeautifulSoup ile hava durumu bilgileri çıkarılmış ve işlenmemiş veriler bir veri çerçevesine dönüştürülmüştür. Eksik değer içeren satırlar çıkarıldıktan sonra, sıcaklık aralıkları düzenlenmiş ve elde edilen veriler CSV formatında saklanmıştır. Bu yeni hava durumu verileri, daha önceki verilerle birleştirilmiştir.

Tezde kullanılan veri kaynaklarından biri Ergast API olmuştur. Ergast API, 1950'den günümüze kadar olan Formula 1 verilerini içermektedir. Bu API'den elde edilen veriler json formatında çıkarılmış, liste sözlüklerine kaydedilmiş ve daha sonra bir veri çerçevesine yüklenmiştir. İşleme aşamasında, sürücülerin genel yarış süreleri milyonlarca milisaniye olarak hesaplanmıştır. İlk ölçeklendirme amacıyla, bu süreler saniyeye çevrilmek için 1000'e bölünmüştür. Bu işlem, verilerin doğru bir şekilde modellenmesine zemin hazırlamıştır.

Veri işleme aşamasında, sıralama turlarının süreleri `qual_time_formatter` fonksiyonu ile saniye ve milisaniye cinsine dönüştürülmüş, her sürücü için her yarışa ait ortalama ve en iyi sıralama süreleri hesaplanmıştır. Sıralama turları, üç aşamadan oluşmaktadır: Q1'de tüm sürücüler katılır ve en iyi zamanlarını elde etmeye çalışırlar, ardından 16. ile 20. sıradaki sürücüler elenir. Q2'de ilk 15 sürücü mücadele eder ve 11. ile 15. sıradakiler sıralanır. Q3'te ise kalan sürücüler son kez en iyi tur zamanlarını belirler. Ancak, her sürücü tüm turlarda yer almadığından, verilerde Q2 ve Q3 için eksik değerler oluşabilir. Eğer veriler orijinal formatıyla kullanılsaydı, veri setindeki sıralama turlarının yaklaşık yarısında eksik değerler bulunacaktı. Bu eksiklikleri gidermek ve tüm sıralama verilerini kullanabilmek için `Q_mean`, `Q_best` ve `Q_worst` değişkenleri oluşturulmuştur. `Q_mean`, sürücünün katıldığı her sıralama turundaki en iyi zamanların ortalamasını, `Q_best` en iyi zamanını, `Q_worst` ise en kötü zamanını temsil eder. Bu süreler, her sürücünün o hafta sonu ve pistteki genel performansını yansıtacak şekilde hesaplanmış ve Q1, Q2, Q3 sütunları veri setinden çıkarılarak, verilerin bütünlüğü sağlanmıştır.

Ayrıca, bu hesaplamalar sonrasında her bir sürücünün `q_best`, `q_worst` ve `q_mean` değerleri belirli aralıklara bölünerek kategorik değişkenlere dönüştürülmüştür. Bu işlem için `pd.cut` fonksiyonu kullanılmış ve 52 ile 145 arasında belirlenen aralıklara göre veriler 1'den 9'a kadar etiketlenmiştir. Bu dönüşüm, her sürücünün sıralama performansını daha anlamlı gruplar halinde incelemeye olanak sağlamıştır. Son olarak, veri setindeki `q_best`, `q_worst` ve `q_mean` sütunları çıkarılarak, işlenmiş veri setine sadece kategorik değişkenler eklenmiş ve modelleme için hazır hale getirilmiştir. Bu işlem, sürücülerin sıralama turlarındaki tutarlılıklarını belirlemede ve tahmin doğruluğunu artırmada önemli bir rol oynamaktadır.

Wikipedia'dan pistlerle ilgili detaylı bilgiler toplanmıştır. Bu bilgiler arasında pist türü (sokak veya yarış pisti), döngü yönü (saat yönünde veya tersine), tur uzunluğu ve birleştirme anahtarları için gereken isim ve konum bilgileri bulunmaktadır. Verilerin işlenmesi amacıyla BeautifulSoup kullanılarak özel fonksiyonlar geliştirilmiştir.

Yıllar içinde sponsor değişiklikleri ve birleşmeler nedeniyle birçok constructor isminin değiştiği gözlemlenmiştir. Bu değişiklikler genellikle takım yapısında minimal değişikliklerle gerçekleşmiştir. Örneğin, Aston Martin adı, Lawrence Stroll tarafından Force India'nın satın alınması ve Racing Point olarak yeniden markalaşması sürecinde iki

kez değiştirilmiştir. 2020 sezonunun sonuna kadar Racing Point adı kullanılmış ve sonrasında Aston Martin adıyla sponsor olarak birleşilmiştir. 2023 yılı test veri seti için eski veriler güncellenmiştir.

Veri setindeki eksik değerlerin giderilmesi amacıyla, `main_df` veri çerçevesinde bulunan `split_times` ve `status` sütunlarından yola çıkarak yeni bir `filled_splits` sütunu oluşturulmuştur. Bu işlemde, öncelikle `split_times` sütunundaki eksik değerler, bir önceki geçerli değerle doldurulmak amacıyla `ffill` fonksiyonu kullanılarak işlenmiştir. Böylece, eksik değerler, önceki geçerli değerlerle doldurulmuştur. Ardından, `split_compute` adlı bir fonksiyon tanımlanmıştır. Bu fonksiyon, `split` ve `status` parametrelerini alarak, `status` içerisinde 'Laps' ifadesi bulunduğu ve `status` tamsayı olduğu durumda `split` değerini, `status` ile çarparak geri döndürmektedir. Aksi takdirde yalnızca `split` değeri geri döndürülmektedir. Tüm bu işlemler, veri çerçevesindeki her satır için bir döngü yardımıyla uygulanmış ve elde edilen yeni `split` değerleri `new_splits` listesine eklenmiştir. Son aşamada, `new_splits` listesi `filled_splits` adıyla veri çerçevesine yeni bir sütun olarak eklenmiştir. Bu süreç, veri setindeki eksik değerlerin doldurularak, daha tutarlı bir analiz ve modelleme yapılmasına olanak sağlamayı hedeflemiştir.

Veri setinde yer alan podium özelliği, tahmin ölçeğini artırmak amacıyla güncellenmiştir. Bu değişiklik, yarışta ilk üçe giren sürücüler için 1, diğerleri için 0 değeri döndürecek şekilde yeniden düzenlenmiş ve bu değerler podium sütununa eklenmiştir. Bu işlem, eğitim veri setindeki bütün özellikler üzerinden model uygulandığı zaman, yarış sonuçlarını daha iyi analiz etmek ve tahmin doğruluğunu artırmak için gerçekleştirilmiştir.

Bu çalışmada, veri setinde yer alan kategorik değişkenler, makine öğrenimi modellerine uygun hale getirilmek amacıyla `pd.get_dummies` fonksiyonu ile "one-hot encoding" işlemine tabi tutulmuştur. Bu işlem, `country`, `type`, `direction`, `locality`, `race_name`, `circuitId` ve `constructor` gibi kategorik sütunları ikili formatta kodlayarak, her benzersiz kategori için yeni sütunlar oluşturmuştur. Bu dönüşüm, her kategori için ilgili veri noktalarının 1 (var) ya da 0 (yok) olarak temsil edilmesini sağlamıştır.

Oluşturulan bu yeni sütunlar, belirli bir eşik değerine göre filtrelenmiştir. Her sütunun toplam değeri hesaplanarak, 70'ten daha az değere sahip olan sütunlar veri setinden çıkarılmıştır. Bu işlem, yeterince temsili olmayan kategorilerin modelleme sürecinde yer almamasını sağlayarak modelin performansını iyileştirmeyi

amaçlamaktadır. Filtreleme işlemi, country, type, direction, locality, race_name, circuitId ve constructor sütunları için uygulanmıştır.

Ayrıca, season sütunu orijinal veri setinden alınıp herhangi bir dönüşüm uygulanmadan işlenmiş veri setine eklenmiştir. Bu adımlar sonucunda, kategorik değişkenler modellemeye uygun hale getirilmiş ve verilerin temsiliyet gücü artırılarak analiz için gerekli hazırlıklar tamamlanmıştır.

3.3. Öznitelik Seçimi

MAFESE, özellikle yüksek boyutlu veri setleri üzerinde öznitelik seçimi yaparken karşılaşılan zorlukları aşmak için tasarlanmış bir araçtır. Yüksek boyutlu veriler, çok sayıda özelliğin bir arada bulunduğu ve bu özelliklerin çoğunun hedef değişkenle anlamlı bir ilişki göstermediği durumlarda, veri setlerinin aşırı uyuma (overfitting) meyilli olmasına yol açabilir. Bu noktada, MAFESE'nin sunduğu meta-sezgisel algoritmalar (metaheuristic algorithms) devreye girerek, özellik seçim sürecini optimize etmeyi hedeflemektedir. Meta-sezgisel algoritmalar, özniteliklerin tüm olası kombinasyonlarını değerlendirmek yerine, problem uzayını etkili bir şekilde tarayarak en iyi özellik alt kümelerini seçmektedirler. Bu, genetik algoritmalar (Genetic Algorithms), parçacık sürü optimizasyonu (Particle Swarm Optimization), karınca kolonisi optimizasyonu (Ant Colony Optimization) gibi yöntemleri içerir ve bu algoritmalar, MAFESE'nin zengin algoritma havuzunda bulunmaktadır.

Bu çalışmada, MAFESE'nin sunduğu sarma tabanlı yöntemler (wrapper-based methods) özellikle tercih edilmiştir, çünkü bu yöntemler, öznitelik alt kümelerini değerlendirirken aynı zamanda bir tahmin modeli ile performanslarını ölçmektedirler. Sarma tabanlı yaklaşımlar, algoritmanın hem öznitelik seçim sürecini hem de model performansını aynı anda optimize etmesine olanak tanımaktadır. MAFESE'nin yerleşik değerlendirme metrikleri, doğruluk (accuracy), F1 skoru (F1 score), kesinlik (precision) ve hassasiyet (recall) gibi ölçütlerle, algoritmaların performanslarını karşılaştırmayı sağlamaktadır. Bu metrikler, öznitelik seçiminde hangi algoritmanın en iyi sonucu verdiğini anlamada kritik rol oynamaktadır.

Bunun yanı sıra, MAFESE'nin filtre tabanlı öznitelik seçim (filter-based feature selection) yöntemleri, veri kümesindeki özniteliklerin hedef değişken ile bağımsız olarak değerlendirildiği bir süreç sunmaktadır. Filtre tabanlı yaklaşımlar, özellikle büyük veri

setleri için hızlı ve etkili bir yöntemdir çünkü her bir özneliğin istatistiksel ilişkisi göz önünde bulundurularak seçme işlemi yapılmaktadır. Ancak, sarma tabanlı yaklaşımlardan farklı olarak, model performansı göz önüne alınmadığından, özneliliklerin etkisi sadece veri kümesi üzerinde yapılan istatistiksel analizlere dayandırılmaktadır.

Gömülü yöntemler (embedded methods) ise, öznelilik seçimini modelin öğrenme sürecine entegre eder ve belirli algoritmalar tarafından doğal olarak gerçekleştirilir. Örneğin, karar ağaçları (decision trees) veya Lasso regresyonu (Lasso regression) gibi modellerde, öznelilik seçimi modelin kendisi tarafından yapılmaktadır. Gömülü yöntemler, modelin özneliliklerin önem derecesini öğrenmesine ve aynı anda model performansını optimize etmesine olanak tanımaktadır. MAFESE'nin bu tür gömülü yöntemleri desteklemesi, özellikle karmaşık modelleme süreçlerinde veri bilimcilere esneklik kazandırmaktadır (Thieu, n.d.).

Açık kaynaklı yapısı sayesinde MAFESE, topluluk tarafından sürekli geliştirilmeye açıktır ve bu da sürekli olarak yeni algoritmaların ve tekniklerin entegrasyonuna olanak tanımaktadır. Ayrıca, scikit-learn gibi popüler makine öğrenimi kütüphaneleri ile entegrasyon sağlaması, kullanıcıların tanıdık bir arayüz üzerinden çalışmasını kolaylaştırmaktadır. Bu sayede MAFESE, özellikle nesne yönelimli Python programlaması ile basit ve etkili bir kullanım sunarak hem akademik araştırmalar hem de endüstriyel uygulamalarda tercih edilen bir araç haline gelmektedir (Thieu ve ark., 2024).

Bu bağlamda, MAFESE kütüphanesi bu çalışma kapsamında, yüksek boyutlu ve karmaşık veri setleri üzerinde öznelilik seçimi yapılarak model performansını iyileştirmek amacıyla kullanılmıştır. Farklı meta-sezgisel algoritmaların karşılaştırılmasıyla elde edilen sonuçlar, hangi algoritmaların belirli veri setleri üzerinde daha etkili olduğunu göstermiştir. Ayrıca, öznelilik seçimi yapıldığında modelin doğruluğunun arttığı ve aşırı uyum riskinin azaldığı gözlemlenmiştir.

Öznelilik seçimi işlemi için, MAFESE kütüphanesi kullanılarak meta-sezgisel algoritmalar aracılığıyla gerçekleştirilmiştir. Bu süreçte, meta-sezgisel algoritmaların farklı kategorilerinden optimizasyon yöntemleri seçilmiş ve elde edilen sonuçlar karşılaştırılarak, öznelilik seçiminin etkileri üzerine bir değerlendirme yapılması amaçlanmıştır. Formül 1 verisi üzerinde, öznelilik seçiminin sonuçları ile tüm özneliliklerin kullanıldığı durum karşılaştırılarak hangi yaklaşımın daha iyi performans sağladığı incelenecektir.

Aşağıdaki optimizasyon yöntemleri bu tezde kullanılmıştır:

Success History Intelligent Optimizer (SHIO): Optimizasyon süreçlerinin iyileştirilmesi amacıyla başarı geçmişine dayanan bir yöntem olarak tanıtılmaktadır. SHIO, önceki optimizasyon denemelerinden elde edilen başarılı sonuçları analiz ederek ve bu bilgileri kullanarak arama stratejilerini dinamik olarak ayarlamaktadır. Bu yaklaşım, özellikle karmaşık problemlerde daha etkili ve hızlı çözümler elde edilmesini sağlamaktadır (Fakhouri ve ark., 2022).

3.4. Verilerin Modellenmesi

Bu çalışmada, 117 öznitelik ve 3837 satırdan oluşan bir veri seti kullanılmıştır. Test verisi olarak sadece 2023 sezonu seçilmektedir. Bu işlem, veri setindeki 'season' sütununun 2023 yılına eşit olan veriler filtrelenerek yapılmaktadır. Eğitim ve doğrulama veri kümelerini oluşturmak için 2023 ve 2024 sezonları haricindeki veriler 'train_val_data' olarak seçilmektedir. Sonrasında, bu veri kümesi eğitim ve doğrulama kümelerine ayrılmaktadır. Bu ayırma işlemi, veri setinin %80'inin eğitim, %20'sinin doğrulama verisi olarak kullanılmasını sağlayan 'train_test_split' fonksiyonu ile gerçekleştirilmektedir.

Eğitim verisindeki dengesizliği azaltmak için SMOTE yöntemi uygulanmıştır. Ardından, öznitelik seçimi amacıyla Mafese kütüphanesi ve çeşitli optimizasyon yöntemleri kullanılmıştır. Öznitelik seçimi sonrası, DT, RF, SVM, NB ve LSTM algoritmalarıyla modelleme yapılmıştır. Ayrıca, öznitelik seçimi yapılmamış ham veri setiyle de aynı algoritmalar kullanılarak model performansları karşılaştırılmıştır. LSTM algoritmasının sınıflandırma etkinliği de incelenmiştir. Tahminlerde araç özellikleri kullanılmamıştır, sürücü performansı, hava durumu ve pist koşulları gibi faktörler kullanılmıştır.

4. BULGULAR

4.1. Ham Veriye Uygulamam Makine Öğrenmesi Algoritmaları

4.1.1. Long Short-Term Memory

2014-2022 yıllarına ait veri seti, LSTM algoritmasında kullanılmadan önce %80'i eğitim verisi olacak şekilde bölünmüştür. Şekil 1'de verilerin bu şekilde bölünmüş hali gösterilmektedir. Bu işlem sonrasında, eğitim verisine SMOTE yöntemi uygulanmıştır. Formula 1 verilerinin her yarışta yalnızca bir kazananı olduğu için veri seti dengesizdir. Bu dengesizlik, SMOTE uygulanarak giderilmeye çalışılmıştır. Şekil 2'de ise SMOTE uygulanmadan önce ve uygulandıktan sonra 0 ve 1 sınıflarının düzenlenmiş hali gösterilmektedir. Model uygulamasına kadar, ham verilerin hazırlanması ve işlenmesi adımları tüm algoritmalar için aynıdır.

Şekil 1: LSTM verisinin dağılımı

Eğitim Verisi:
(2784, 117)

Doğrulama Verisi:
(697, 117)

Test Verisi (2023 Sezonu):
(356, 117)

Şekil 2: SMOTE uygulanan eğitim verisinin dağılımı

```
0      2366
1       418
Name: count, dtype: int64
0      2644
1      2644
Name: count, dtype: int64
```

Düzenlenen veriler, LSTM algoritması ile modellenmeye başlanmıştır. Modelin eğitimi tamamlandıktan sonra, Şekil 3, 4 ve 5'te sırasıyla eğitim, test ve doğrulama verilerinin değerleri gösterilmektedir.

Modelin eğitim verisi Şekil 3'te görüldüğü üzere doğruluğu %85.95 olarak belirlenmiştir. Sınıflandırma raporunda, sınıf 0 için %91 doğruluk ve %80 hatırlama oranı

ile %85 F1 skoru elde edilmiştir. Sınıf 1 için ise %82 doğruluk ve %92 hatırlama oranı ile %87 F1 skoru hesaplanmıştır.

Hata matrisi incelendiğinde, modelin sınıf 0'da 2122 doğru, 522 yanlış; sınıf 1'de ise 2423 doğru, 221 yanlış sınıflandırma yaptığı görülmektedir. Bu sonuçlar, modelin genel olarak başarılı bir sınıflandırma yaptığını göstermektedir.

Şekil 3: LSTM eğitim verisi performans değerleri

Train Accuracy: 0.8594931921331316

Sınıflandırma Raporu (Train Veri):

	precision	recall	f1-score	support
0	0.91	0.80	0.85	2644
1	0.82	0.92	0.87	2644
accuracy			0.86	5288
macro avg	0.86	0.86	0.86	5288
weighted avg	0.86	0.86	0.86	5288

Train Hata Matrisi :

```
[[2122 522]
 [ 221 2423]]
```

Modelin test verisi Şekil 4'te görüldüğü üzere doğruluğu %75 olarak hesaplanmıştır. Sınıflandırma raporuna göre, sınıf 0 için %100 doğruluk ve %74 hatırlama oranı ile %85 F1 skoru elde edilmiştir. Sınıf 1 için ise %16 doğruluk ve %94 hatırlama oranı ile %28 F1 skoru hesaplanmıştır.

Hata matrisi incelendiğinde, modelin sınıf 0'da 250 doğru, 88 yanlış; sınıf 1'de ise 17 doğru, 1 yanlış sınıflandırma yaptığı görülmektedir. Bu sonuçlar, modelin sınıf 0'da

oldukça başarılı olduğunu, ancak sınıf 1'de düşük doğruluk oranına sahip olduğunu göstermektedir.

Şekil 4: LSTM test verisi performans değerleri

*Test Accuracy: 0.75

Sınıflandırma Raporu (Test Veri):

	precision	recall	f1-score	support
0	1.00	0.74	0.85	338
1	0.16	0.94	0.28	18
accuracy			0.75	356
macro avg	0.58	0.84	0.56	356
weighted avg	0.95	0.75	0.82	356

Test Hata Matrisi:

```
[[250  88]
 [  1  17]]
```

Modelin doğrulama verisi üzerindeki doğruluğu %84.65 olarak hesaplanmıştır. Şekil 5'te gösterilen sınıflandırma raporuna göre, sınıf 0 için %97 doğruluk ve %84 hatırlama oranı ile %90 F1 skoru elde edilmiştir. Sınıf 1 için ise %47 doğruluk ve %86 hatırlama oranı ile %61 F1 skoru hesaplanmıştır.

Şekil 5'te yer alan hata matrisi incelendiğinde, modelin sınıf 0'da 506 doğru, 93 yanlış; sınıf 1'de ise 84 doğru, 14 yanlış sınıflandırma yaptığı görülmektedir. Bu sonuçlar, modelin sınıf 0'da oldukça iyi performans sergilediğini, ancak sınıf 1'de doğruluk oranının daha düşük olduğunu göstermektedir.

Şekil 5: LSTM doğrulama verisi performans değerleri

Doğrulama(Validation) Accuracy: 0.8464849354375896

Sınıflandırma Raporu (Validation Veri):

	precision	recall	f1-score	support
0	0.97	0.84	0.90	599
1	0.47	0.86	0.61	98
accuracy			0.85	697
macro avg	0.72	0.85	0.76	697
weighted avg	0.90	0.85	0.86	697

Validation Hata Matrisi :

```
[[506  93]
 [ 14  84]]
```

Şekil 6 da, örnek bir tahminleme sütunu gösterilmektedir. 2023 yılı test verisinden elde edilen sonuçlara göre, toplamda düzenlenen 18 yarışın 12'sinin doğru tahmin edildiği görülmektedir.

Şekil 6 LSTM örnek tahmin sütunu

	name	actual	predicted
0	Max Verstappen	1	1
1	Sergio Pérez	0	0
2	Lewis Hamilton	0	0
3	George Russell	0	0
4	Logan Sargeant	0	0
5	Alexander Albon	0	0
6	Fernando Alonso	0	0
7	Carlos Sainz	0	0
8	Esteban Ocon	0	0
9	Lando Norris	0	0
10	Oscar Piastri	0	0
11	Nico Hülkenberg	0	0
12	Guanyu Zhou	0	0
13	Valtteri Bottas	0	0
14	Nyck de Vries	0	0
15	Yuki Tsunoda	0	0
16	Lance Stroll	0	0
17	Pierre Gasly	0	0
18	Charles Leclerc	0	0
19	Kevin Magnussen	0	0
Toplam doğru tahmin sayısı: 12			

4.1.2. Decision Tree

LSTM'de uygulanan adımlar, Decision Tree modeli için de geçerlidir. Şekil 7, 8 ve 9'da bu modelden elde edilen test, doğrulama ve eğitim verilerinin sonuçları gösterilmektedir.

DT modelinin test doğruluğu %99.72 olarak hesaplanmıştır. Şekil 7'de yer alan sınıflandırma raporuna göre, sınıf 0 için %100 doğruluk ve %100 hatırlama oranı ile %100 F1 skoru elde edilmiştir. Sınıf 1 için ise %95 doğruluk ve %100 hatırlama oranı ile %97 F1 skoru hesaplanmıştır. Hata matrisi incelendiğinde, modelin 338 sınıf 0 örneğinin 337'sini doğru, 1'ini yanlış; 18 sınıf 1 örneğinin ise tamamını doğru sınıflandırdığı görülmektedir.

Şekil 7: DT test verisi performans değerleri

```
DT Test accuracy degeri : 0.9971910112359551
[[337  1]
 [  0 18]]
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	338
1	0.95	1.00	0.97	18
accuracy			1.00	356
macro avg	0.97	1.00	0.99	356
weighted avg	1.00	1.00	1.00	356

Doğrulama verisi üzerindeki doğruluk değeri %100 olarak hesaplanmıştır ve bu sonuçlar Şekil 8'de gösterilmektedir. Model, sınıf 0 için %100 doğruluk ve %100 hatırlama oranına sahipken, sınıf 1 için de aynı oranlarla %100 F1 skoru elde edilmiştir. Hata matrisi incelendiğinde, 664 sınıf 0 ve 33 sınıf 1 örneğinin tamamının doğru sınıflandırıldığı görülmektedir.

Şekil 8: DT doğrulama verisi performans değerleri

DT Validation Accuracy değeri: 1.0				
[[664 0]				
[0 33]]				
	precision	recall	f1-score	support
0	1.00	1.00	1.00	664
1	1.00	1.00	1.00	33
accuracy			1.00	697
macro avg	1.00	1.00	1.00	697
weighted avg	1.00	1.00	1.00	697

Eğitim verisi için de doğruluk %100 olarak hesaplanmıştır ve sonuçlar Şekil 9'da sunulmuştur. Sınıflandırma raporuna göre, her iki sınıf için de %100 doğruluk ve %100 hatırlama oranları ile %100 F1 skorları elde edilmiştir. Hata matrisi, 2644 adet sınıf 0 ve 2644 adet sınıf 1 örneğinin tamamının doğru sınıflandırıldığını göstermektedir.

Şekil 9: DT eğitim verisi performans değerleri

DT Train Accuracy değeri: 1.0				
[[2644 0]				
[0 2644]]				
	precision	recall	f1-score	support
0	1.00	1.00	1.00	2644
1	1.00	1.00	1.00	2644
accuracy			1.00	5288
macro avg	1.00	1.00	1.00	5288
weighted avg	1.00	1.00	1.00	5288

Şekil 10 da örnek bir tahminleme sütunu sunulmaktadır. 2023 yılı test verisiyle yapılan tahminlerde, toplam 18 yarışın tamamının doğru tahmin edildiği görülmektedir.

Şekil 10: DT örnek tahmin sütunu

	name	actual	predicted
0	Max Verstappen	1	1
1	Esteban Ocon	0	0
2	Pierre Gasly	0	0
3	Yuki Tsunoda	0	0
4	Nico Hülkenberg	0	0
5	Guanyu Zhou	0	0
6	Kevin Magnussen	0	0
7	Nyck de Vries	0	0
8	George Russell	0	0
9	Logan Sargeant	0	0
10	Oscar Piastri	0	0
11	Alexander Albon	0	0
12	Valtteri Bottas	0	0
13	Lance Stroll	0	0
14	Fernando Alonso	0	0
15	Lewis Hamilton	0	0
16	Charles Leclerc	0	0
17	Carlos Sainz	0	0
18	Sergio Pérez	0	0
19	Lando Norris	0	0
Toplam doğru tahmin sayısı(18 Tane): 18			

4.1.3. Naive Bayes

NB modeli için elde edilen değerler Şekil 11, 12 ve 13'te gösterilmektedir. Eğitim verisi üzerindeki doğruluk %88.14 olarak hesaplanmıştır. Şekil 11'de sunulan sınıflandırma raporuna göre, sınıf 0 için %98 doğruluk ve %78 hatırlama oranı ile %87 F1 skoru elde edilmiştir. Sınıf 1 için ise %82 doğruluk ve %98 hatırlama oranı ile %89 F1 skoru hesaplanmıştır. Hata matrisi incelendiğinde, modelin 2644 sınıf 0 örneğinin 2068'ini doğru, 576'sını yanlış; 2644 sınıf 1 örneğinin ise 2593'ünü doğru, 51'ini yanlış sınıflandırdığı görülmektedir.

Şekil 11: NB eğitim verisi performans değerleri

Train Accuracy: 0.88142965204236

Sınıflandırma Raporu (Train Veri):

	precision	recall	f1-score	support
0	0.98	0.78	0.87	2644
1	0.82	0.98	0.89	2644
accuracy			0.88	5288
macro avg	0.90	0.88	0.88	5288
weighted avg	0.90	0.88	0.88	5288

Train Hata Matrisi :

```
[[2068 576]
 [ 51 2593]]
```

Test verisi üzerindeki doğruluk %69.94 olarak hesaplanmıştır ve bu sonuçlar Şekil 12’de gösterilmektedir. Sınıflandırma raporuna göre, sınıf 0 için %100 doğruluk ve %68 hatırlama oranı ile %81 F1 skoru elde edilmiştir. Sınıf 1 için ise %14 doğruluk ve %100 hatırlama oranı ile %25 F1 skoru hesaplanmıştır. Hata matrisi incelendiğinde, 338 sınıf 0 örneğinin 231’inin doğru, 107’sinin yanlış; 18 sınıf 1 örneğinin ise tamamının doğru sınıflandırıldığı görülmektedir.

Şekil 12: NB test verisi performans değerleri

*Test Accuracy: 0.699438202247191

Sınıflandırma Raporu (Test Veri):

	precision	recall	f1-score	support
0	1.00	0.68	0.81	338
1	0.14	1.00	0.25	18
accuracy			0.70	356
macro avg	0.57	0.84	0.53	356
weighted avg	0.96	0.70	0.78	356

Test Hata Matrisi:

```
[[231 107]
 [ 0 18]]
```

Doğrulama verisi üzerindeki doğruluk %81.64 olarak hesaplanmış olup, sonuçlar Şekil 13'te sunulmaktadır. Sınıf 0 için %98 doğruluk ve %80 hatırlama oranı ile %88 F1 skoru elde edilmiştir. Sınıf 1 için ise %43 doğruluk ve %92 hatırlama oranı ile %58 F1 skoru hesaplanmıştır. Hata matrisi incelendiğinde, modelin 599 sınıf 0 örneğinin 479'unu doğru, 120'sini yanlış; 98 sınıf 1 örneğinin ise 90'ını doğru, 8'ini yanlış sınıflandırdığı görülmektedir.

Şekil 13: NB doğrulama verisi performans değerleri

Doğrulama(Validation) Accuracy: 0.8163558106169297

Sınıflandırma Raporu (Validation Veri):

	precision	recall	f1-score	support
0	0.98	0.80	0.88	599
1	0.43	0.92	0.58	98
accuracy			0.82	697
macro avg	0.71	0.86	0.73	697
weighted avg	0.91	0.82	0.84	697

Validation Hata Matrisi :

[[479 120]

[8 90]]

Şekil 14'te, örnek bir tahminleme sütunu sunulmaktadır. 2023 yılı test verisiyle yapılan tahminlerde, toplam 18 yarışın tamamının doğru tahmin edildiği görülmektedir.

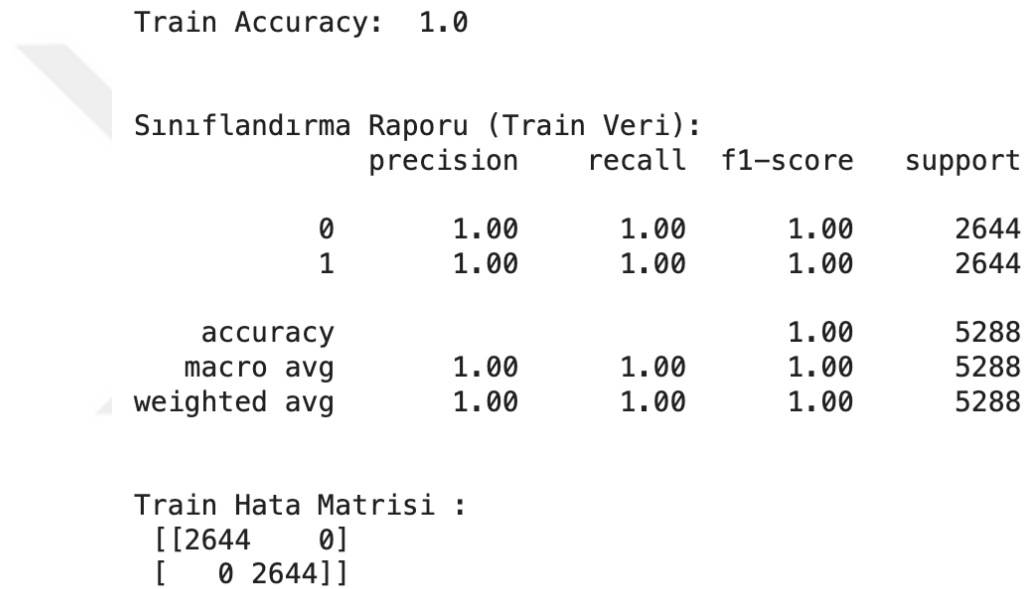
Şekil 14: NB örnek tahmin sütunu

	name	actual	predicted
0	Lewis Hamilton	0	1
1	Max Verstappen	1	0
2	Esteban Ocon	0	0
3	Pierre Gasly	0	0
4	Yuki Tsunoda	0	0
5	Nico Hülkenberg	0	0
6	Guanyu Zhou	0	0
7	Kevin Magnussen	0	0
8	Nyck de Vries	0	0
9	George Russell	0	0
10	Logan Sargeant	0	0
11	Oscar Piastri	0	0
12	Alexander Albon	0	0
13	Valtteri Bottas	0	0
14	Lance Stroll	0	0
15	Fernando Alonso	0	0
16	Charles Leclerc	0	0
17	Carlos Sainz	0	0
18	Sergio Pérez	0	0
19	Lando Norris	0	0
Toplam doğru tahmin sayısı: 16			

4.1.4. Random Forest Classifier

RF modeli için elde edilen değerler Şekil 15, 16 ve 17’te gösterilmektedir. Eğitim verisi üzerindeki doğruluk %100 olarak hesaplanmıştır. Şekil 15’te sunulan sınıflandırma raporuna göre, her iki sınıf için de %100 doğruluk, %100 hatırlama ve %100 F1 skoru elde edilmiştir. Hata matrisi incelendiğinde, modelin 2644 sınıf 0 ve 2644 sınıf 1 örneğinin tamamını doğru sınıflandırdığı görülmektedir.

Şekil 15: RF eğitim verisi performans değerleri



```
Train Accuracy: 1.0
```

Sınıflandırma Raporu (Train Veri):

	precision	recall	f1-score	support
0	1.00	1.00	1.00	2644
1	1.00	1.00	1.00	2644
accuracy			1.00	5288
macro avg	1.00	1.00	1.00	5288
weighted avg	1.00	1.00	1.00	5288

Train Hata Matrisi :

```
[[2644  0]
 [  0 2644]]
```

Test verisi üzerindeki doğruluk %99.72 olarak hesaplanmıştır ve bu sonuçlar Şekil 16’da gösterilmektedir. Sınıflandırma raporuna göre, sınıf 0 için %100 doğruluk ve %100 hatırlama oranı ile %100 F1 skoru elde edilmiştir. Sınıf 1 için ise %95 doğruluk ve %100 hatırlama oranı ile %97 F1 skoru hesaplanmıştır. Hata matrisi incelendiğinde, 338 sınıf 0 örneğinin 337’sinin doğru, 1’inin yanlış; 18 sınıf 1 örneğinin ise tamamının doğru sınıflandırıldığı görülmektedir.

Şekil 16: RF test verisi performans değerleri

*Test Accuracy: 0.9971910112359551

```
Sınıflandırma Raporu (Test Veri):
      precision    recall  f1-score   support

     0         1.00      1.00      1.00       338
     1         0.95      1.00      0.97        18

 accuracy          1.00          356
 macro avg         0.97          356
weighted avg         1.00          356
```

```
Test Hata Matrisi:
[[337   1]
 [   0  18]]
```

Doğrulama verisi üzerindeki doğruluk %100 olarak hesaplanmış olup, sonuçlar Şekil 17'de sunulmuştur. Sınıflandırma raporuna göre, her iki sınıf için de %100 doğruluk, %100 hatırlama ve %100 F1 skorları elde edilmiştir. Hata matrisi, 664 sınıf 0 ve 33 sınıf 1 örneğinin tamamının doğru sınıflandırıldığını göstermektedir.

Şekil 17: RF doğrulama verisi performans değerleri

Doğrulama(Validation) Accuracy: 1.0

```
Sınıflandırma Raporu (Validation Veri):
      precision    recall  f1-score   support

     0         1.00      1.00      1.00       664
     1         1.00      1.00      1.00        33

 accuracy          1.00          697
 macro avg         1.00          697
weighted avg         1.00          697
```

```
Validation Hata Matrisi :
[[664   0]
 [   0  33]]
```

Şekil 18’de, örnek bir tahminleme sütunu sunulmaktadır. 2023 yılı test verisiyle yapılan tahminlerde, toplam 18 yarışın 17 tanesini doğru tahmin edildiği görülmektedir.

Şekil 18: RF örnek tahmin sütunu

	name	actual	predicted
0	Lewis Hamilton	0	1
1	Max Verstappen	1	0
2	Esteban Ocon	0	0
3	Pierre Gasly	0	0
4	Yuki Tsunoda	0	0
5	Nico Hülkenberg	0	0
6	Guanyu Zhou	0	0
7	Kevin Magnussen	0	0
8	Nyck de Vries	0	0
9	George Russell	0	0
10	Logan Sargeant	0	0
11	Oscar Piastri	0	0
12	Alexander Albon	0	0
13	Valtteri Bottas	0	0
14	Lance Stroll	0	0
15	Fernando Alonso	0	0
16	Charles Leclerc	0	0
17	Carlos Sainz	0	0
18	Sergio Pérez	0	0
19	Lando Norris	0	0
Toplam doğru tahmin sayısı: 17			

4.1.5. Support Vector Machines

SVM modeli için elde edilen değerler Şekil 19, 20 ve 21’te gösterilmektedir. Eğitim verisi üzerindeki doğruluk %99.98 olarak hesaplanmıştır. Şekil 19’da sunulan sınıflandırma raporuna göre, sınıf 0 ve sınıf 1 için %100 doğruluk, %100 hatırlama ve %100 F1 skoru elde edilmiştir. Hata matrisi incelendiğinde, 2644 sınıf 0 örneğinin 2643’ünün doğru, 1’inin yanlış; 2644 sınıf 1 örneğinin tamamının doğru sınıflandırıldığı görülmektedir.

Şekil 19: SVM eğitim verisi performans değerleri

Train Accuracy: 0.9998108925869894

Sınıflandırma Raporu (Train Veri):

	precision	recall	f1-score	support
0	1.00	1.00	1.00	2644
1	1.00	1.00	1.00	2644
accuracy			1.00	5288
macro avg	1.00	1.00	1.00	5288
weighted avg	1.00	1.00	1.00	5288

Train Hata Matrisi :

```
[[2643  1]
 [  0 2644]]
```

Test verisi üzerindeki doğruluk %87.64 olarak hesaplanmıştır ve bu sonuçlar Şekil 20'de gösterilmektedir. Sınıflandırma raporuna göre, sınıf 0 için %100 doğruluk ve %87 hatırlama oranı ile %93 F1 skoru elde edilmiştir. Sınıf 1 için ise %29 doğruluk ve %100 hatırlama oranı ile %45 F1 skoru hesaplanmıştır. Hata matrisi incelendiğinde, 338 sınıf 0 örneğinin 294'ünün doğru, 44'ünün yanlış; 18 sınıf 1 örneğinin ise tamamının doğru sınıflandırıldığı görülmektedir.

Şekil 20: SVM test verisi performans değerleri

*Test Accuracy: 0.8764044943820225

Sınıflandırma Raporu (Test Veri):				
	precision	recall	f1-score	support
0	1.00	0.87	0.93	338
1	0.29	1.00	0.45	18
accuracy			0.88	356
macro avg	0.65	0.93	0.69	356
weighted avg	0.96	0.88	0.91	356

Test Hata Matrisi:

```
[[294  44]
 [  0  18]]
```

Doğrulama verisi üzerindeki doğruluk %100 olarak hesaplanmış olup, sonuçlar Şekil 21'de sunulmuştur. Sınıflandırma raporuna göre, her iki sınıf için de %100 doğruluk, %100 hatırlama ve %100 F1 skorları elde edilmiştir. Hata matrisi, 664 sınıf 0 ve 33 sınıf 1 örneğinin tamamının doğru sınıflandırıldığını göstermektedir.

Şekil 21: SVM doğrulama verisi performans değerleri

Doğrulama(Validation) Accuracy: 1.0

Sınıflandırma Raporu (Validation Veri):				
	precision	recall	f1-score	support
0	1.00	1.00	1.00	664
1	1.00	1.00	1.00	33
accuracy			1.00	697
macro avg	1.00	1.00	1.00	697
weighted avg	1.00	1.00	1.00	697

Validation Hata Matrisi :

```
[[664  0]
 [  0  33]]
```

Şekil 22’de, örnek bir tahminleme sütunu sunulmaktadır. 2023 yılı test verisiyle yapılan tahminlerde, toplam 18 yarışın 9 tanesini doğru tahmin edildiği görülmektedir.

Şekil 22: SVM örnek tahmin sütunu

	name	actual	predicted
0	Lewis Hamilton	0	1
1	Fernando Alonso	0	0
2	Max Verstappen	1	0
3	Esteban Ocon	0	0
4	Logan Sargeant	0	0
5	Yuki Tsunoda	0	0
6	Nico Hülkenberg	0	0
7	Guanyu Zhou	0	0
8	Kevin Magnussen	0	0
9	Nyck de Vries	0	0
10	George Russell	0	0
11	Oscar Piastri	0	0
12	Pierre Gasly	0	0
13	Alexander Albon	0	0
14	Valtteri Bottas	0	0
15	Lance Stroll	0	0
16	Charles Leclerc	0	0
17	Carlos Sainz	0	0
18	Sergio Pérez	0	0
19	Lando Norris	0	0
Toplam doğru tahmin sayısı: 9			

4.2. MASEFE Yöntemi ile Seçilen Özniteliklerin Makine Öğrenimi Algoritmalarına Uygulanması

Bu öznitelik seçimi kütüphanesi ile, verinin özniteliklerini sadeleştirmek amacıyla 2020 yılından sonra geliştirilen bazı optimizasyon yöntemlerini kullanmıştır. Ham veriye uygulanan işlemlerin aynısı burada da uygulanmaktadır. Model eğitime geçilmeden önce, bu kütüphane kullanılarak öznitelik sadeleştirilmesi yapılmıştır. Bu yöntemlerin çıktıları, belirlenen makine öğrenimi algoritmalarına uygulanmıştır.

4.2.1. OriginalSHIO Optimizasyonu

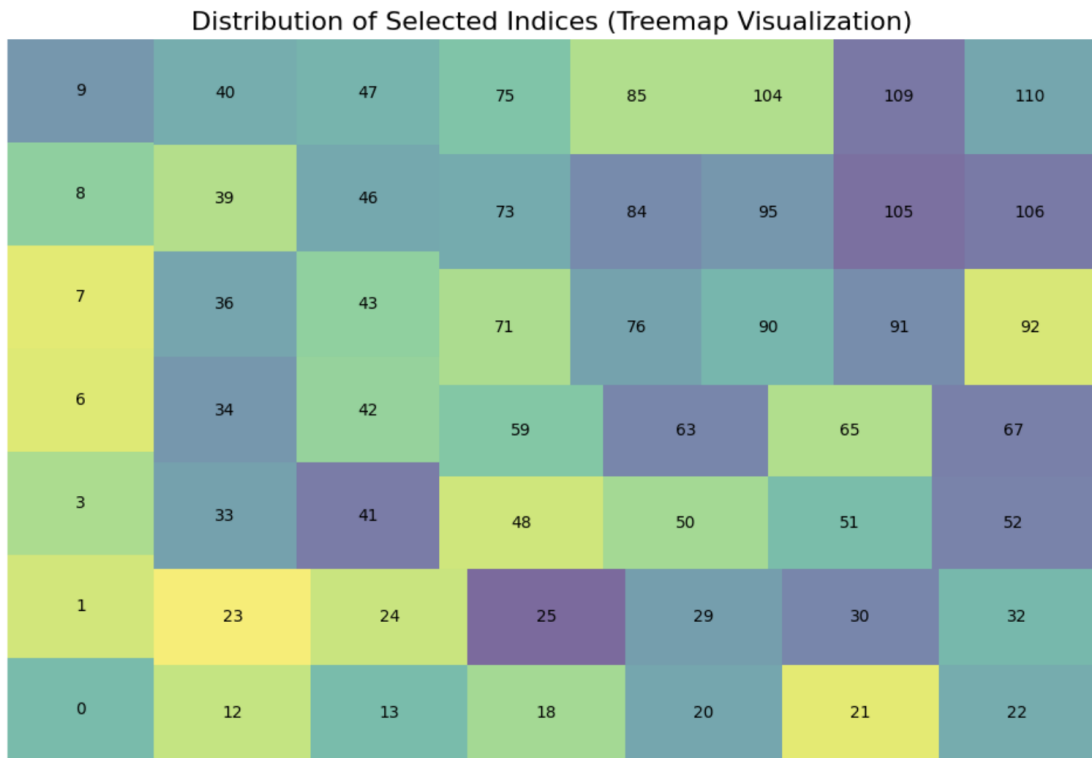
Bu çalışmada, Formula 1 yarış sonuçlarının makine öğrenimi algoritmaları ile tahmin edilmesi amacıyla hem ham veri seti hem de öznitelik seçimi uygulanmıştır. Öznitelik seçiminde optimizasyon olarak OriginalSHIO kullanılmıştır. Ayrıca, veri setine

SMOTE uygulanarak veri dengesizliği giderilmeye çalışılmıştır. SMOTE kullanımı, özellikle dengesiz sınıfların tahmin edilmesinde algoritmaların performansını artırmayı hedeflemektedir. Bu tartışma, elde edilen sonuçları detaylı bir şekilde değerlendirerek farklı algoritmaların performanslarını karşılaştırmakta ve öz nitelik seçiminin etkilerini ele almaktadır.

4.2.1.1. Long Short- Term Memory

Bu optimizasyon yöntemi ile elde edilen yeni veri boyutları, Şekil 23 ve 24'te gösterilmektedir. Bu veri setine uygulanan LSTM modelinin çıktıları ise Şekil 25, 26 ve 27'de sunulmuştur.

Şekil 23: SHIO optimizasyonu ile LSTM modelinde kullanılmak için seçilen öz nitelikler



Şekil 24 LSTM modelinde kullanılacak yeni veri boyutları

Eğitim seti boyutu: (4732, 52)
Doğrulama seti boyutu: (697, 52)
Test seti boyutu: (356, 52)

Eğitim verisi üzerindeki doğruluk %69.25 olarak hesaplanmıştır. Şekil 25'te sunulan sınıflandırma raporuna göre, sınıf 0 için %64 doğruluk ve %89 hatırlama oranı ile %74 F1 skoru elde edilmiştir. Sınıf 1 için ise %81 doğruluk ve %50 hatırlama oranı ile %62 F1 skoru hesaplanmıştır. Hata matrisi incelendiğinde, 2366 sınıf 0 örneğinin 2097'sinin doğru, 269'unun yanlış; 2366 sınıf 1 örneğinin 1180'inin doğru, 1186'sının yanlış sınıflandırıldığı görülmektedir.

Şekil 25: LSTM ve yeni veri seti ile eğitim verisi performans değerleri

Train Accuracy: 0.6925190194420964



Sınıflandırma Raporu (Train Veri):				
	precision	recall	f1-score	support
0	0.64	0.89	0.74	2366
1	0.81	0.50	0.62	2366
accuracy			0.69	4732
macro avg	0.73	0.69	0.68	4732
weighted avg	0.73	0.69	0.68	4732

Train Hata Matrisi :
[[2097 269]
[1186 1180]]

Test verisi üzerindeki doğruluk %86.52 olarak hesaplanmıştır ve bu sonuçlar Şekil 26'da gösterilmektedir. Sınıflandırma raporuna göre, sınıf 0 için %95 doğruluk ve %91 hatırlama oranı ile %93 F1 skoru elde edilmiştir. Sınıf 1 için ise %6 doğruluk ve %11 hatırlama oranı ile %8 F1 skoru hesaplanmıştır. Hata matrisi incelendiğinde, 338 sınıf 0 örneğinin 306'sının doğru, 32'sinin yanlış; 18 sınıf 1 örneğinin 2'sinin doğru, 16'sının yanlış sınıflandırıldığı görülmektedir.

Şekil 26: LSTM ve yeni veri seti ile test verisi performans değerleri

*Test Accuracy: 0.8651685393258427

Sınıflandırma Raporu (Test Veri):

	precision	recall	f1-score	support
0	0.95	0.91	0.93	338
1	0.06	0.11	0.08	18
accuracy			0.87	356
macro avg	0.50	0.51	0.50	356
weighted avg	0.91	0.87	0.88	356

Test Hata Matrisi:

```
[[306  32]
 [ 16   2]]
```

Doğrulama verisi üzerindeki doğruluk %76.61 olarak hesaplanmıştır ve sonuçlar Şekil 27'de sunulmuştur. Sınıf 0 için %85 doğruluk ve %88 hatırlama oranı ile %87 F1 skoru elde edilmiştir. Sınıf 1 için ise %8 doğruluk ve %6 hatırlama oranı ile %7 F1 skoru hesaplanmıştır. Hata matrisi incelendiğinde, 599 sınıf 0 örneğinin 528'inin doğru, 71'inin yanlış; 98 sınıf 1 örneğinin 6'sının doğru, 92'sinin yanlış sınıflandırıldığı görülmektedir.

Şekil 27: LSTM ve yeni veri seti ile doğrulama verisi performans değerleri

Doğrulama(Validation) Accuracy: 0.7661406025824964

Sınıflandırma Raporu (Validation Veri):

	precision	recall	f1-score	support
0	0.85	0.88	0.87	599
1	0.08	0.06	0.07	98
accuracy			0.77	697
macro avg	0.46	0.47	0.47	697
weighted avg	0.74	0.77	0.75	697

Validation Hata Matrisi :

```
[[528  71]
 [ 92   6]]
```

Ayrıca Şekil 28'de, 2023 yılı test verisiyle yapılan tahminlerde toplam 18 yarışın 14'ünün doğru tahmin edildiği görülmektedir.

Şekil 28: LSTM ve yeni veri seti ile örnek tahmin sütunu

1/1 ————— 0s 162ms/step
Round 8: Precision = 1.0

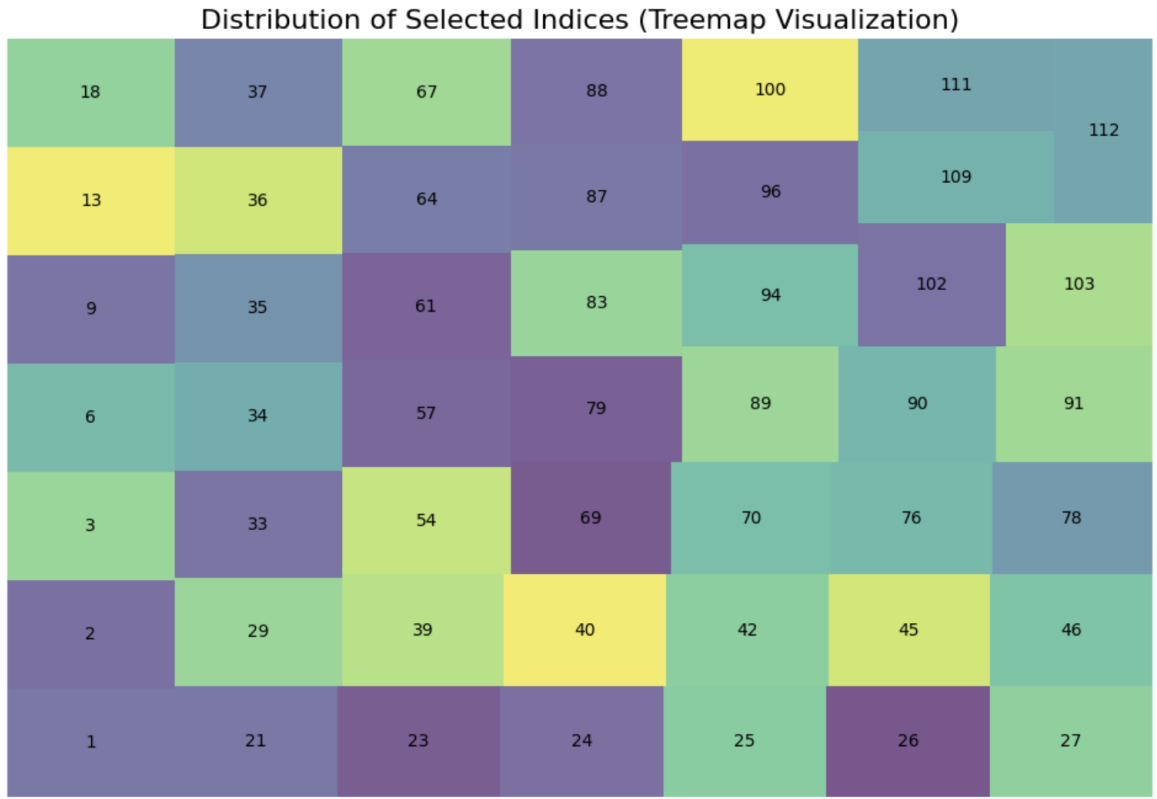
	name	actual	predicted
0	Max Verstappen	1	1
1	Sergio Pérez	0	0
2	Fernando Alonso	0	0
3	Alexander Albon	0	0
4	Esteban Ocon	0	0
5	Valtteri Bottas	0	0
6	Lance Stroll	0	0
7	Oscar Piastri	0	0
8	Pierre Gasly	0	0
9	Lando Norris	0	0
10	Logan Sargeant	0	0
11	Nyck de Vries	0	0
12	Kevin Magnussen	0	0
13	Guanyu Zhou	0	0
14	Nico Hülkenberg	0	0
15	Yuki Tsunoda	0	0
16	Charles Leclerc	0	0
17	Carlos Sainz	0	0
18	George Russell	0	0
19	Lewis Hamilton	0	0

Toplam doğru tahmin sayısı: 14

4.2.1.2. Desicion Tree

Bu optimizasyon yöntemi ile elde edilen yeni veri boyutları, Şekil 29 ve 30'da gösterilmektedir. Bu veri setine uygulanan DT modelinin çıktıları ise Şekil 31, 32 ve 33'te sunulmuştur.

Şekil 29: SHIO optimizasyonu ile DT modelinde kullanılmak için seçilen öznelilikler



Şekil 30: DT modelinde kullanılacak yeni veri boyutları

Eğitim seti boyutu: (4732, 48)
Doğrulama seti boyutu: (697, 48)
Test seti boyutu: (356, 48)

Test verisi üzerindeki doğruluk %87.36 olarak hesaplanmıştır ve bu sonuçlar Şekil 31'de gösterilmektedir. Sınıflandırma raporuna göre, sınıf 0 için %100 doğruluk ve %87 hatırlama oranı ile %93 F1 skoru elde edilmiştir. Sınıf 1 için ise %29 doğruluk ve %100 hatırlama oranı ile %44 F1 skoru hesaplanmıştır. Hata matrisi incelendiğinde, 338 sınıf 0 örneğinin 293'ünün doğru, 45'inin yanlış; 18 sınıf 1 örneğinin tamamının doğru sınıflandırıldığı görülmektedir.

Şekil 31: DT ve yeni veri seti ile test verisi performans değerleri

DT Test accuracy değeri : 0.8735955056179775

[[293 45]
[0 18]]

	precision	recall	f1-score	support
0	1.00	0.87	0.93	338
1	0.29	1.00	0.44	18
accuracy			0.87	356
macro avg	0.64	0.93	0.69	356
weighted avg	0.96	0.87	0.90	356

Doğrulama verisi üzerindeki doğruluk %93.83 olarak hesaplanmıştır ve bu sonuçlar Şekil 32'de sunulmuştur. Sınıf 0 için %95 doğruluk ve %97 hatırlama oranı ile %96 F1 skoru elde edilmiştir. Sınıf 1 için ise %82 doğruluk ve %71 hatırlama oranı ile %77 F1 skoru hesaplanmıştır. Hata matrisi incelendiğinde, 599 sınıf 0 örneğinin 584'ünün doğru, 15'inin yanlış; 98 sınıf 1 örneğinin 70'inin doğru, 28'inin yanlış sınıflandırıldığı görülmektedir.

Şekil 32: DT ve yeni veri seti ile doğrulama verisi performans değerleri

DT Validation Accuracy değeri: 0.9383070301291249

[[584 15]
[28 70]]

	precision	recall	f1-score	support
0	0.95	0.97	0.96	599
1	0.82	0.71	0.77	98
accuracy			0.94	697
macro avg	0.89	0.84	0.86	697
weighted avg	0.94	0.94	0.94	697

Eğitim verisi üzerindeki doğruluk %99.94 olarak hesaplanmış olup, sonuçlar Şekil 33'te sunulmuştur. Sınıflandırma raporuna göre, her iki sınıf için de %100 doğruluk ve

%100 hatırlama oranları ile %100 F1 skorları elde edilmiştir. Hata matrisi, 2366 sınıf 0 örneğinin tamamının ve 2366 sınıf 1 örneğinin 2363'ünün doğru sınıflandırıldığını göstermektedir.

Şekil 33: DT ve yeni veri seti ile eğitim verisi performans değerleri

DT Train Accuracy değeri: 0.9993660185967879

```
[[2366    0]
 [   3 2363]]
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	2366
1	1.00	1.00	1.00	2366
accuracy			1.00	4732
macro avg	1.00	1.00	1.00	4732
weighted avg	1.00	1.00	1.00	4732

Ayrıca Şekil 34'te, 2023 yılı test verisiyle yapılan tahminlerde, toplam 18 yarışın sadece 1'inin doğru tahmin edildiği görülmektedir.

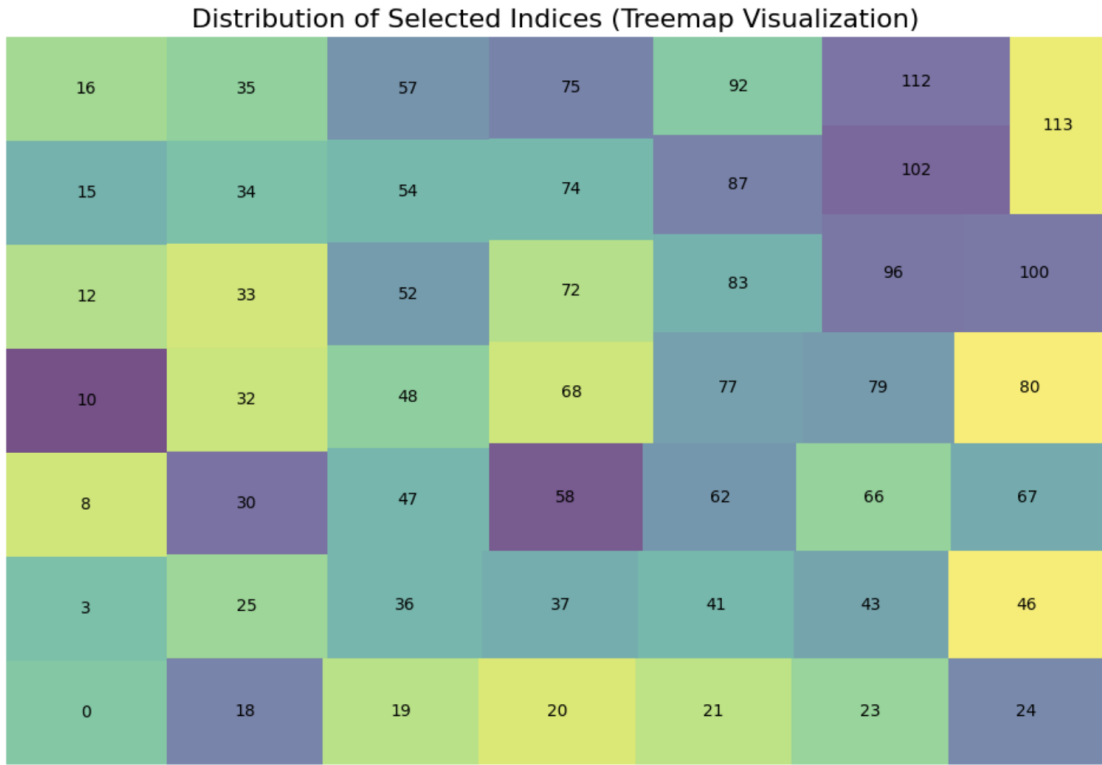
Şekil 34: DT ve yeni veri seti ile örnek tahmin sütunu

	name	actual	predicted
0	Esteban Ocon	0	1
1	Alexander Albon	0	0
2	Sergio Pérez	0	0
3	Carlos Sainz	0	0
4	Charles Leclerc	0	0
5	Lewis Hamilton	0	0
6	Fernando Alonso	0	0
7	Max Verstappen	1	0
8	Lance Stroll	0	0
9	Valtteri Bottas	0	0
10	Oscar Piastri	0	0
11	Pierre Gasly	0	0
12	Logan Sargeant	0	0
13	George Russell	0	0
14	Nyck de Vries	0	0
15	Kevin Magnussen	0	0
16	Guanyu Zhou	0	0
17	Nico Hülkenberg	0	0
18	Yuki Tsunoda	0	0
19	Lando Norris	0	0
Toplam doğru tahmin sayısı: 1			

4.2.1.3. Navie Bayes

Bu optimizasyon yöntemi ile elde edilen yeni veri boyutları, Şekil 35 ve 36'da gösterilmektedir. Bu veri setine uygulanan NB modelinin çıktıları ise Şekil 37, 38 ve 39'da sunulmuştur.

Şekil 35: SHIO optimizasyonu ile NB modelinde kullanılmak için seçilen öznelilikler



Şekil 36: NB modelinde kullanılacak yeni veri boyutları

Eğitim seti boyutu: (5288, 48)
Doğrulama seti boyutu: (697, 48)
Test seti boyutu: (356, 48)

Eğitim verisi üzerindeki doğruluk %70.31 olarak hesaplanmıştır. Şekil 37'de sunulan sınıflandırma raporuna göre, sınıf 0 için %98 doğruluk ve %41 hatırlama oranı ile %58 F1 skoru elde edilmiştir. Sınıf 1 için ise %63 doğruluk ve %99 hatırlama oranı ile %77 F1 skoru hesaplanmıştır. Hata matrisi incelendiğinde, 2644 sınıf 0 örneğinin

1094'ünün doğru, 1550'sinin yanlış; 2644 sınıf 1 örneğinin 2624'ünün doğru, 20'sinin yanlış sınıflandırıldığı görülmektedir.

Şekil 37: NB ve yeni veri seti ile eğitim verisi performans değerleri

Train Accuracy: 0.7031013615733737

Sınıflandırma Raporu (Train Veri):

	precision	recall	f1-score	support
0	0.98	0.41	0.58	2644
1	0.63	0.99	0.77	2644
accuracy			0.70	5288
macro avg	0.81	0.70	0.68	5288
weighted avg	0.81	0.70	0.68	5288

Train Hata Matrisi :
[[1094 1550]
[20 2624]]

Test verisi üzerindeki doğruluk %34.27 olarak hesaplanmıştır ve bu sonuçlar Şekil 38'de gösterilmektedir. Sınıflandırma raporuna göre, sınıf 0 için %100 doğruluk ve %31 hatırlama oranı ile %47 F1 skoru elde edilmiştir. Sınıf 1 için ise %7 doğruluk ve %100 hatırlama oranı ile %13 F1 skoru hesaplanmıştır. Hata matrisi incelendiğinde, 338 sınıf 0 örneğinin 104'ünün doğru, 234'ünün yanlış; 18 sınıf 1 örneğinin tamamının doğru sınıflandırıldığı görülmektedir.

Şekil 38: NB ve yeni veri seti ile test verisi performans değerleri

*Test Accuracy: 0.34269662921348315

Sınıflandırma Raporu (Test Veri):				
	precision	recall	f1-score	support
0	1.00	0.31	0.47	338
1	0.07	1.00	0.13	18
accuracy			0.34	356
macro avg	0.54	0.65	0.30	356
weighted avg	0.95	0.34	0.45	356

Test Hata Matrisi:
[[104 234]
[0 18]]

Doğrulama verisi üzerindeki doğruluk %48.78 olarak hesaplanmış olup, sonuçlar Şekil 39'da sunulmuştur. Sınıf 0 için %99 doğruluk ve %41 hatırlama oranı ile %58 F1 skoru elde edilmiştir. Sınıf 1 için ise %21 doğruluk ve %97 hatırlama oranı ile %35 F1 skoru hesaplanmıştır. Hata matrisi incelendiğinde, 599 sınıf 0 örneğinin 245'inin doğru, 354'ünün yanlış; 98 sınıf 1 örneğinin 95'inin doğru, 3'ünün yanlış sınıflandırıldığı görülmektedir.

Şekil 39: NB ve yeni veri seti ile doğrulama verisi performans değerleri

Doğrulama(Validation) Accuracy: 0.4878048780487805

Sınıflandırma Raporu (Validation Veri):				
	precision	recall	f1-score	support
0	0.99	0.41	0.58	599
1	0.21	0.97	0.35	98
accuracy			0.49	697
macro avg	0.60	0.69	0.46	697
weighted avg	0.88	0.49	0.55	697

Validation Hata Matrisi :
[[245 354]
[3 95]]

Ayrıca Şekil 40'ta, 2023 yılı test verisiyle yapılan tahminlerde, toplam 18 yarışın sadece 1'inin doğru tahmin edildiği görülmektedir.

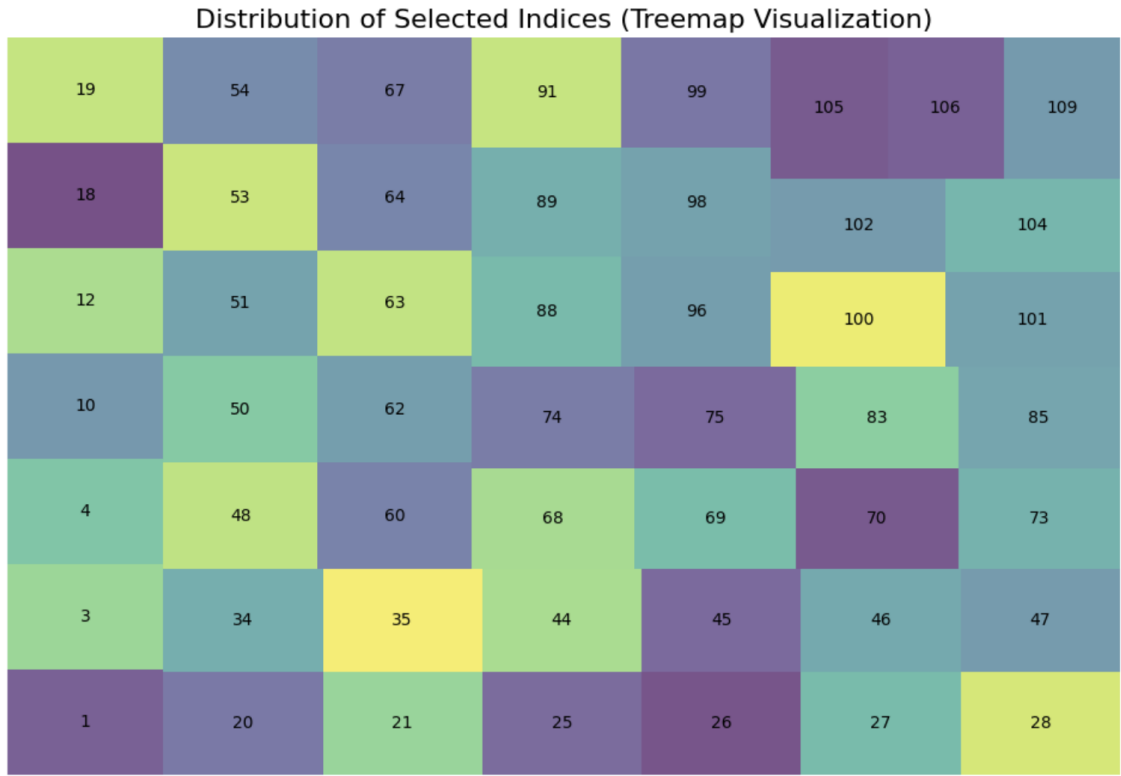
Şekil 40: NB ve yeni veri seti ile örnek tahmin sütunu

	name	actual	predicted
0	Esteban Ocon	0	1
1	Alexander Albon	0	0
2	Yuki Tsunoda	0	0
3	Nico Hülkenberg	0	0
4	Guanyu Zhou	0	0
5	Kevin Magnussen	0	0
6	Nyck de Vries	0	0
7	George Russell	0	0
8	Logan Sargeant	0	0
9	Pierre Gasly	0	0
10	Oscar Piastri	0	0
11	Valtteri Bottas	0	0
12	Lance Stroll	0	0
13	Max Verstappen	1	0
14	Fernando Alonso	0	0
15	Lewis Hamilton	0	0
16	Charles Leclerc	0	0
17	Carlos Sainz	0	0
18	Sergio Pérez	0	0
19	Lando Norris	0	0
Toplam doğru tahmin sayısı:		1	

4.2.1.4. Random Forest

Bu optimizasyon yöntemi ile elde edilen yeni veri boyutları, Şekil 41 ve 42'de gösterilmektedir. Bu veri setine uygulanan RF modelinin çıktıları ise Şekil 43, 44 ve 45'te sunulmuştur.

Şekil 41: SHIO optimizasyonu ile RF modelinde kullanılmak için seçilen öznelilikler



Şekil 42: RF modelinde kullanılacak yeni veri boyutları

Eğitim seti boyutu: (4732, 50)
Doğrulama seti boyutu: (697, 50)
Test seti boyutu: (356, 50)

Eğitim verisi üzerindeki doğruluk %98.96 olarak hesaplanmıştır. Şekil 43'te sunulan sınıflandırma raporuna göre, sınıf 0 ve sınıf 1 için %99 doğruluk ve %99 hatırlama oranı ile %99 F1 skoru elde edilmiştir. Hata matrisi incelendiğinde, 2366 sınıf 0 örneğinin 2344'ünün doğru, 22'sinin yanlış; 2366 sınıf 1 örneğinin 2339'unun doğru, 27'sinin yanlış sınıflandırıldığı görülmektedir.

Şekil 43: RF ve yeni veri seti ile eğitim verisi performans değerleri

Train Accuracy: 0.9896449704142012

Sınıflandırma Raporu (Train Veri):

	precision	recall	f1-score	support
0	0.99	0.99	0.99	2366
1	0.99	0.99	0.99	2366
accuracy			0.99	4732
macro avg	0.99	0.99	0.99	4732
weighted avg	0.99	0.99	0.99	4732

Train Hata Matrisi :

```
[[2344  22]
 [ 27 2339]]
```

Test verisi üzerindeki doğruluk %90.17 olarak hesaplanmıştır ve bu sonuçlar Şekil 44'te gösterilmektedir. Sınıflandırma raporuna göre, sınıf 0 için %100 doğruluk ve %90 hatırlama oranı ile %95 F1 skoru elde edilmiştir. Sınıf 1 için ise %34 doğruluk ve %100 hatırlama oranı ile %51 F1 skoru hesaplanmıştır. Hata matrisi incelendiğinde, 338 sınıf 0 örneğinin 303'ünün doğru, 35'inin yanlış; 18 sınıf 1 örneğinin tamamının doğru sınıflandırıldığı görülmektedir.

Şekil 44: RF ve yeni veri seti ile test verisi performans değerleri

*Test Accuracy: 0.901685393258427

Sınıflandırma Raporu (Test Veri):					
	precision	recall	f1-score	support	
0	1.00	0.90	0.95	338	
1	0.34	1.00	0.51	18	
accuracy			0.90	356	
macro avg	0.67	0.95	0.73	356	
weighted avg	0.97	0.90	0.92	356	

Test Hata Matrisi:

```
[[303  35]
 [  0  18]]
```

Doğrulama verisi üzerindeki doğruluk %93.69 olarak hesaplanmıştır ve sonuçlar Şekil 45'te sunulmuştur. Sınıf 0 için %95 doğruluk ve %97 hatırlama oranı ile %96 F1 skoru elde edilmiştir. Sınıf 1 için ise %82 doğruluk ve %70 hatırlama oranı ile %76 F1 skoru hesaplanmıştır. Hata matrisi incelendiğinde, 599 sınıf 0 örneğinin 584'ünün doğru, 15'inin yanlış; 98 sınıf 1 örneğinin 69'unun doğru, 29'unun yanlış sınıflandırıldığı görülmektedir.

Şekil 45: RF ve yeni veri seti ile doğrulama verisi performans değerleri

Doğrulama(Validation) Accuracy: 0.9368723098995696

Sınıflandırma Raporu (Validation Veri):

	precision	recall	f1-score	support
0	0.95	0.97	0.96	599
1	0.82	0.70	0.76	98
accuracy			0.94	697
macro avg	0.89	0.84	0.86	697
weighted avg	0.93	0.94	0.93	697

Validation Hata Matrisi :

```
[[584  15]
 [ 29  69]]
```

Ayrıca Şekil 46 da 2023 yılı test verisiyle yapılan tahminlerde, toplam 18 yarışın sadece 2 tanesi doğru tahmin edildiği görülmektedir.

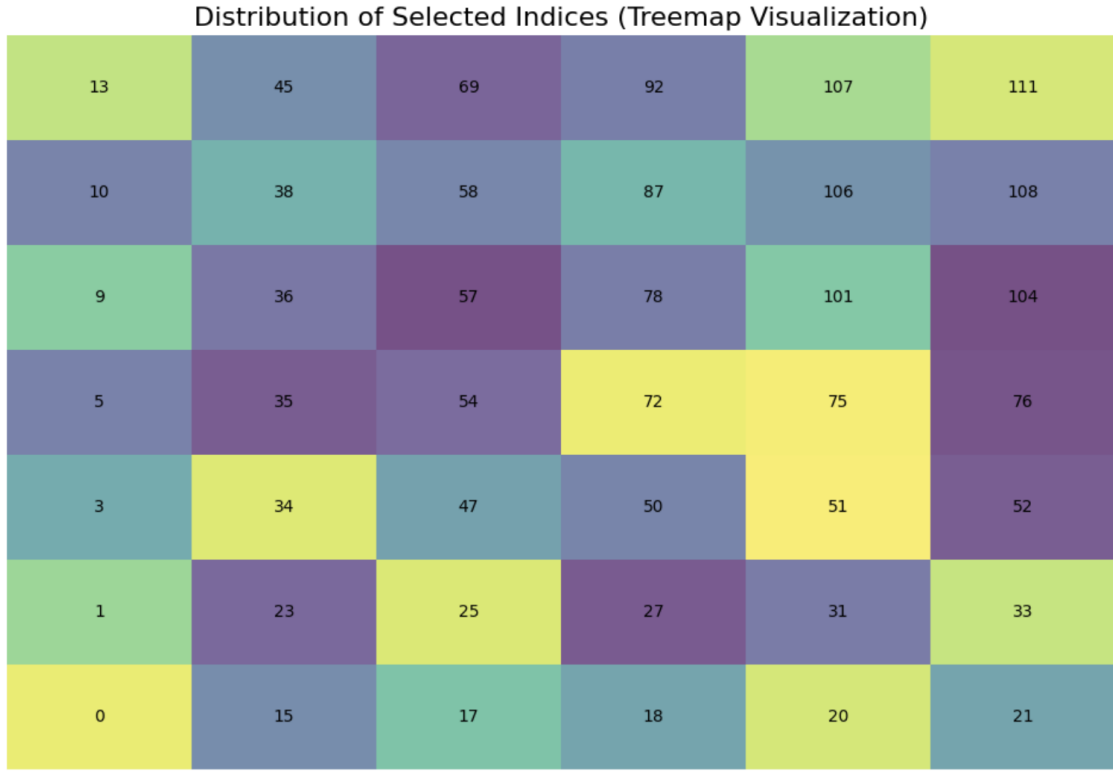
Şekil 46: RF ve yeni veri seti ile örnek tahmin sütunu

	name	actual	predicted
0	Esteban Ocon	0	1
1	Alexander Albon	0	0
2	Sergio Pérez	0	0
3	Carlos Sainz	0	0
4	Charles Leclerc	0	0
5	Lewis Hamilton	0	0
6	Max Verstappen	1	0
7	Valtteri Bottas	0	0
8	Pierre Gasly	0	0
9	Oscar Piastri	0	0
10	George Russell	0	0
11	Logan Sargeant	0	0
12	Lando Norris	0	0
13	Nico Hülkenberg	0	0
14	Kevin Magnussen	0	0
15	Guanyu Zhou	0	0
16	Lance Stroll	0	0
17	Fernando Alonso	0	0
18	Nyck de Vries	0	0
19	Yuki Tsunoda	0	0
Toplam doğru tahmin sayısı:		2	

4.2.1.5. Support Vektor Machine

Bu optimizasyon yöntemi ile elde edilen yeni veri boyutları, Şekil 47 ve 48'de gösterilmektedir. Bu veri setine uygulanan SVM modelinin çıktıları ise Şekil 49, 50 ve 51'de sunulmuştur.

Şekil 47: SHIO optimizasyonu ile SVM modelinde kullanılmak için seçilen öz nitelikler



Şekil 48: SVM modelinde kullanılacak yeni veri boyutları

Eğitim seti boyutu: (4732, 42)
Doğrulama seti boyutu: (697, 42)
Test seti boyutu: (356, 42)

Eğitim verisi üzerindeki doğruluk %95.73 olarak hesaplanmıştır. Şekil 49'da sunulan sınıflandırma raporuna göre, sınıf 0 için %97 doğruluk ve %94 hatırlama oranı ile %96 F1 skoru elde edilmiştir. Sınıf 1 için ise %94 doğruluk ve %98 hatırlama oranı ile %96 F1 skoru hesaplanmıştır. Hata matrisi incelendiğinde, 2366 sınıf 0 örneğinin 2221'inin doğru, 145'inin yanlış; 2366 sınıf 1 örneğinin 2309'unun doğru, 57'sinin yanlış sınıflandırıldığı görülmektedir.

Şekil 49: SVM ve yeni veri seti ile eğitim verisi performans değerleri

Train Accuracy: 0.9573119188503804

Sınıflandırma Raporu (Train Veri):

	precision	recall	f1-score	support
0	0.97	0.94	0.96	2366
1	0.94	0.98	0.96	2366
accuracy			0.96	4732
macro avg	0.96	0.96	0.96	4732
weighted avg	0.96	0.96	0.96	4732

Train Hata Matrisi :

```
[[2221 145]
 [ 57 2309]]
```

Test verisi üzerindeki doğruluk %86.24 olarak hesaplanmıştır ve bu sonuçlar Şekil 50'de gösterilmektedir. Sınıflandırma raporuna göre, sınıf 0 için %100 doğruluk ve %86 hatırlama oranı ile %92 F1 skoru elde edilmiştir. Sınıf 1 için ise %27 doğruluk ve %100 hatırlama oranı ile %42 F1 skoru hesaplanmıştır. Hata matrisi incelendiğinde, 338 sınıf 0 örneğinin 289'unun doğru, 49'unun yanlış; 18 sınıf 1 örneğinin tamamının doğru sınıflandırıldığı görülmektedir.

Şekil 50: SVM ve yeni veri seti ile test verisi performans değerleri

*Test Accuracy: 0.8623595505617978

Sınıflandırma Raporu (Test Veri):					
	precision	recall	f1-score	support	
0	1.00	0.86	0.92	338	
1	0.27	1.00	0.42	18	
accuracy			0.86	356	
macro avg	0.63	0.93	0.67	356	
weighted avg	0.96	0.86	0.90	356	

Test Hata Matrisi:
[[289 49]
[0 18]]

Doğrulama verisi üzerindeki doğruluk %91.97 olarak hesaplanmış olup, sonuçlar Şekil 51'de sunulmuştur. Sınıf 0 için %97 doğruluk ve %94 hatırlama oranı ile %95 F1 skoru elde edilmiştir. Sınıf 1 için ise %68 doğruluk ve %80 hatırlama oranı ile %74 F1 skoru hesaplanmıştır. Hata matrisi incelendiğinde, 599 sınıf 0 örneğinin 563'ünün doğru, 36'sının yanlış; 98 sınıf 1 örneğinin 78'inin doğru, 20'sinin yanlış sınıflandırıldığı görülmektedir.

Şekil 51: SVM ve yeni veri seti ile doğrulama verisi performans değerleri

Doğrulama(Validation) Accuracy: 0.9196556671449068

Sınıflandırma Raporu (Validation Veri):

	precision	recall	f1-score	support
0	0.97	0.94	0.95	599
1	0.68	0.80	0.74	98
accuracy			0.92	697
macro avg	0.82	0.87	0.84	697
weighted avg	0.93	0.92	0.92	697

Validation Hata Matrisi :

```
[[563 36]
 [ 20 78]]
```

Ayrıca Şekil 52’de, 2023 yılı test verisiyle yapılan tahminlerde, toplam 18 yarışın sadece 2 tanesi doğru tahmin edildiği görülmektedir.

Şekil 52: SVM ve yeni veri seti ile örnek tahmin sütunu

	name	actual	predicted
0	Esteban Ocon	0	1
1	Alexander Albon	0	0
2	Sergio Pérez	0	0
3	Carlos Sainz	0	0
4	Charles Leclerc	0	0
5	Lewis Hamilton	0	0
6	Fernando Alonso	0	0
7	Max Verstappen	1	0
8	Lance Stroll	0	0
9	Valtteri Bottas	0	0
10	Oscar Piastri	0	0
11	Pierre Gasly	0	0
12	Logan Sargeant	0	0
13	George Russell	0	0
14	Nyck de Vries	0	0
15	Kevin Magnussen	0	0
16	Guanyu Zhou	0	0
17	Nico Hülkenberg	0	0
18	Yuki Tsunoda	0	0
19	Lando Norris	0	0
Toplam doğru tahmin sayısı:		1	

4.3. Ham Veri Seti Sonuçlarının Değerlendirilmesi

Ham veri seti kullanılarak yapılan analizlerde Tablo 2’de, DT ve RFC algoritmalarının oldukça yüksek doğruluk oranlarına ulaştığı gözlemlenmiştir. Hem eğitim hem de test verileri üzerinde %99.7 ve %100 doğruluk oranları elde edilmiştir. Bu, ham veri seti ile bu iki algoritmanın aşırı öğrenme eğilimi gösterdiğini düşündürmektedir. Eğitim ve test doğruluğunun bu denli yüksek olması, modelin öğrenme sürecinde tüm verileri ezberlediğini ve yeni veri ile karşılaştığında genelleme yapma kapasitesinin düşük olabileceğini göstermektedir.

Tablo 2: Ham veri setinin doğruluk sonuçları

	Test Accuracy	Eğitim Accuracy	Doğrulama Accuracy
LSTM	0.75	0.8594931921331316	0.8464849354375896
DT	0.9971910112359551	1.0	1.0
NB	0.699438202247191	0.88142965204236	0.8163558106169297
RFC	0.9971910112359551	1.0	1.0
SVM	0.8764044943820225	0.9998108925869894	1.0

SVM ise %87.6 test doğruluğu ile nispeten iyi bir performans göstermiştir. Ancak, LSTM ve NB algoritmalarının test doğruluk oranları sırasıyla %75 ve %69.9 olarak kaydedilmiştir; bu, bu algoritmaların ham veri seti üzerinde diğer algoritmalarla kıyasla daha düşük bir performans sergilediğini göstermektedir.

Ham veri seti ile yapılan tahmin sonuçlarına bakıldığında Tablo 3’te, DT algoritmasının 18 yarışın tamamını doğru tahmin ederek (18/18) en başarılı model olduğu belirlenmiştir. RFC algoritması ise 18/17 doğru tahmin ile ikinci sırada yer alırken, NB algoritması 18/16 doğru tahmin ile dikkat çekmiştir. SVM algoritmasının 18/9 tahmin sonucu, özellikle ham veri seti üzerindeki performansının bazı sınırlamalara sahip olduğunu göstermektedir. Bu sonuçlar, ham veri seti üzerinde DT ve RFC'nin etkili olduğunu ortaya koymaktadır, ancak aşırı öğrenme riskinin de göz önünde bulundurulması gerekmektedir.

Tablo 3: Ham veri setinin tahmin sonuçları

	Tahmin Sonuçları
LSTM	18/12
DT	18/18
NB	18/16
RFC	18/17
SVM	18/9

4.4. Öznitelik Seçimi (OriginalSHIO) Sonuçlarının Değerlendirilmesi

OriginalSHIO optimizasyonu ile veri setinin öznitelik sayısının azaltılması, modellerin performansında önemli değişikliklere yol açmıştır. Tablo 4'te doğruluk değerleri karşılaştırılmıştır.

Tablo 4: OriginalSHIO optimizasyonu ile yeni veri setinin doğruluk sonuçları

	Test Accuracy	Eğitim Accuracy	Doğrulama Accuracy
LSTM - OriginalSHIO	0.8651685393258427	0.6925190194420964	0.7661406025824964
DT - OriginalSHIO	0.8735955056179775	0.9993660185967879	0.9383070301291249
NB - OriginalSHIO	0.34269662921348315	0.7031013615733737	0.4878048780487805
RFC - OriginalSHIO	0.901685393258427	0.9896449704142012	0.901685393258427
SVM - OriginalSHIO	0.8623595505617978	0.9573119188503804	0.9196556671449068

Öznitelik seçimi uygulanmış veri seti ile yapılan analizlerde, DT ve RFC algoritmalarının test doğruluğu sırasıyla %87.3 ve %90.1'e düşmüştür. Bu azalma, öznitelik sayısının azalmasıyla birlikte modelin bazı önemli özellikleri kaybetmiş

olabileceğini düşündürmektedir. Ancak, bu doğruluk oranları hala yüksek olup, öznelik seçimi sonrasında modellerin genelleme kapasitesinde kısmi bir artış olduğunu göstermektedir.

Özellikle LSTM algoritmasının doğruluğunda belirgin bir iyileşme gözlemlenmiştir; test doğruluğu %75'ten %86.5'e yükselmiştir. Bu durum, öznelik seçimi ile daha az ancak daha nitelikli verinin kullanılmasının LSTM gibi derin öğrenme algoritmalarının performansını artırabileceğini göstermektedir. Diğer yandan, NB algoritmasının test doğruluğunda (%34.2) belirgin bir düşüş yaşanmıştır. Bu, öznelik seçiminin Naive Bayes algoritmasının performansı üzerinde olumsuz bir etki yarattığını göstermektedir; bu durum, NB'nin özneliklerin olasılıksal bağımsızlığı varsayımına dayalı olmasından kaynaklanabilmektedir.

Öznelik seçimi ile elde edilen tahmin sonuçları incelendiğinde Tablo 5'de, DT ve RFC algoritmalarının tahmin doğruluğunda belirgin bir düşüş yaşandığı gözlemlenmiştir. Bu sonuçlar, öznelik seçiminin bazı algoritmaların performansını olumsuz etkilediğini ve modelin genelleme kapasitesini artırmaya çalışırken doğruluk oranının düşebileceğini göstermektedir. SVM algoritması, öznelik seçimi sonrası %86.2 test doğruluğu ile ham veri seti sonuçlarına oldukça yakın bir performans sergilese de, tahmin doğruluğu 18/9'dan 18/1'e düşmüştür. Bu durum, SVM'nin öznelik seçimi sonrası genel doğruluğunu korumasına rağmen, doğru tahmin yapma kabiliyetinde ciddi bir azalma yaşadığını göstermektedir. Bu düşüş, modelin belirli özneliklerin çıkarılması sonrası tahmin yeteneğinin olumsuz etkilendiğine ve öznelik seçiminin SVM için bu veri setinde istenilen iyileşmeyi sağlamadığına işaret etmektedir.

Tablo 5: OriginalSHIO optimizasyonu ile yeni veri setinin tahmin sonuçları

	Tahmin Sonuçları
LSTM - OriginalSHIO	18/14
DT - OriginalSHIO	18/1
NB - OriginalSHIO	18/1
RFC - OriginalSHIO	18/2
SVM - OriginalSHIO	18/1

4.5. Öznitelik Seçimi ve Ham Veri Seti Tahmin Sonuçları Karşılaştırması

Bu tahmin sonuçları, öznitelik seçimi sonrasında özellikle DT, NB ve RFC algoritmalarının doğru tahmin sayısında önemli bir azalma olduğunu göstermektedir. Bu azalma, öznitelik seçiminin bazı algoritmaların performansını olumsuz etkilediğini ve modelin belirli öznitelikler olmadan yarış sonuçlarını tahmin etmede zorlandığını göstermektedir. Öte yandan, LSTM algoritmasının tahmin doğruluğu bir miktar artış göstererek, öznitelik seçiminin bu modelin genelleme kapasitesini olumlu etkilediğine işaret etmektedir.

4.6. SMOTE Uygulamasının Etkisi

SMOTE tekniğinin uygulanması, veri setindeki dengesizliği gidermeyi ve azınlık sınıflarının daha iyi temsil edilmesini amaçlamıştır. Ancak, elde edilen sonuçlar SMOTE uygulamasının algoritmaların performansı üzerindeki etkisinin karmaşık olduğunu göstermektedir. Örneğin, ham veri seti ile RFC algoritması %99.7 test doğruluğu elde ederken, SMOTE ve öznitelik seçimi uygulandıktan sonra bu oran %90.1'e düşmüştür. Bu, SMOTE'un öznitelik seçimi ile birlikte uygulandığında aşırı öğrenmeyi azaltabileceğini, ancak bu sırada modelin doğruluğunun da etkilenebileceğini göstermektedir.

5. TARTIŞMA

Bu çalışmanın ham veri seti üzerine elde edilen sonuçlarla, daha önce yapılan Formula 1 yarışları ile ilgili makine öğrenimi analizlerine dair yayımlanmış bazı çalışmalar ile karşılaştırılmıştır.

Keertish Kumar ve Preethi (2023), tarafından yapılan çalışmada, çeşitli algoritmaların hem sürücülerin hem de yapımcıların performansını belirlemek için kullanıldığı ve özellikle DT ve RFC algoritmalarının yüksek doğruluk oranlarına ulaştığı belirtilmiştir. Mevcut çalışmada da benzer şekilde DT ve RFC algoritmaları için %99.7 gibi yüksek test doğrulukları elde edilmiştir. Bu sonuçlar, önceki çalışmaların bulgularıyla uyum göstermekte ve DT ile RFC' nin Formula 1 performans tahminlerinde güvenilir algoritmalar olduğunu desteklemektedir.

Trowbridge ve Ohlan (2021), tarafından gerçekleştirilen bir başka çalışmada ise, RFC ve SVM algoritmalarının yarış sonuçlarını tahmin etmede etkili olduğu belirtilmiştir. Mevcut çalışmada da benzer şekilde RFC ve SVM algoritmalarının yüksek doğruluk oranlarına ulaştığı gözlenmiştir. RFC için %99.7, SVM için ise %87.6 test doğruluğu elde edilmiştir. Bu sonuçlar, Trowbridge ve Ohlan'ın bulgularıyla büyük ölçüde uyumludur ve bu algoritmaların yarış tahminlerinde kullanılmasının doğruluğunu desteklemektedir. Ancak, SVM doğruluğu önceki çalışmalara kıyasla biraz daha yüksek bulunmuştur; bu durum, mevcut çalışmadaki veri kümesinin yapısı ve özelliklerinin etkisi olarak açıklanabilir.

Carvalho (2022), tarafından yapılan bir çalışmada, tarihsel Formula 1 verilerinin analizinde özellikle NB algoritmasının kullanımı vurgulanmıştır. Ancak, mevcut çalışmada NB algoritmasının %69.9 gibi nispeten düşük bir test doğruluğu sergilediği gözlenmiştir. Bu farklılık, NB algoritmasının özellikle bu veri kümesinde diğer algoritmalara kıyasla daha düşük performans gösterdiğini işaret edebilir. Bu sonuç, NB'nin diğer algoritmalara göre sınırlı bir tahmin gücüne sahip olabileceğini ve özellikle veri setinin karmaşıklığından etkilendiğini göstermektedir.

Tejada (2023) tarafından yapılan bir çalışmada, pit stop zamanlamasını belirlemek için makine öğrenimi algoritmaları kullanılmıştır. Bu çalışmada, 2019-2022 sezonlarına ait yarış verileri analiz edilmiş ve özellikle lastik aşınması, araç pozisyonları ve genel yarış dinamikleri dikkate alınarak bir model geliştirilmiştir. SVM, Random Forest ve Yapay Sinir Ağları algoritmaları kullanılmış ve performansları F1 skoru üzerinden değerlendirilmiştir. Sonuçlar, pit stop tahminlerinin karmaşıklığını ve bu alanda kullanılan modellerin potansiyelini vurgulamaktadır. Bununla birlikte, bu modellerin tek başına kullanılmak yerine, karar destek araçları olarak daha faydalı olduğu belirtilmiştir.

Sicoie (2022) tarafından gerçekleştirilen çalışmada, Formula 1 yarış galibi ve şampiyona sıralamalarının tahmini için çeşitli makine öğrenimi algoritmaları (SVR, RFR, GBR) kullanılmıştır. Veri işleme sürecinde, verilerin zenginleştirilmesi ve çapraz doğrulama tekniklerinin önemi vurgulanmıştır. Bu tez, öznitelik seçimi ve veri setinin pist uzunluğu, hava durumu gibi çevresel faktörlerle zenginleştirilmesini kapsamaktadır. Ayrıca, modellerin performansını değerlendirmek için Leave-One-Out Cross Validation (LOOCV) gibi yöntemler kullanılmıştır. Mevcut çalışmada ise, benzer şekilde, öznitelik seçimi için mafese algoritması ve veri dengesizliğini gidermek için SMOTE yöntemi kullanılmıştır. Ancak, mevcut çalışmanın farkı, LSTM ve NB gibi farklı algoritmaların performansının incelenmesidir.

Rondelli (2022) tarafından yapılan çalışmada ise, Formula 1 lastik stratejisi tahminleri için derin öğrenme algoritmaları (özellikle LSTM ve GRU) kullanılmıştır. Bu çalışmada, "FastF1" API'sinden elde edilen telemetrik veriler kullanılmış ve hangi lastik değişimlerinin ne zaman yapılması gerektiği tahmin edilmiştir. Derin öğrenme algoritmalarının bu tür bir yüksek performans ortamında uygulanabilirliği ele alınmıştır. Mevcut çalışmada ise, LSTM algoritması yarış sonuçlarını tahmin etmede kullanılmış, ancak sadece lastik stratejisi değil, genel yarış tahminleri üzerinde odaklanılmıştır.

Sonuç olarak, Sicoie (2022) ve Rondelli (2022) çalışmalarının veri zenginleştirme, öznitelik seçimi ve algoritma performans değerlendirmesi açısından mevcut çalışmayla benzer yönleri bulunmaktadır. Ancak mevcut çalışmada kullanılan SMOTE, mafese algoritması ve LSTM modelleri, bu çalışmalarla olan farkları ve katkıları ortaya koymaktadır.

Ek olarak, LSTM algoritması bu çalışmada %75 test doğruluğu ile nispeten başarılı bir sonuç elde etmiştir. Daha önce yapılan çalışmalarla doğrudan bir karşılaştırma olmamakla birlikte, zaman serisi verilerinde kullanılan LSTM'nin yarış sonuçlarını tahmin etmede makul bir performans sergilediği gözlenmiştir. Bu bulgu, özellikle gelecekteki çalışmalarda derin öğrenme algoritmalarının kullanımına yönelik önemli bir katkı sağlamaktadır.

Mevcut çalışmanın bir sınırlaması olarak, bazı algoritmaların yüksek doğruluk oranlarına sahip olmasına rağmen, veri setinin belirli özelliklerinden dolayı genelleştirilebilirliklerinin sorgulanabileceği vurgulanmalıdır. Özellikle, DT ve RFC algoritmalarının eğitim doğruluklarının %100 olması, bu algoritmaların aşırı öğrenme (overfitting) eğilimi gösterebileceğini ve yeni veriler üzerinde performanslarının düşebileceğini düşündürmektedir. Dolayısıyla, bu algoritmaların gerçek dünya uygulamalarında dikkatli bir şekilde optimize edilmesi gerektiği belirtilmelidir.

6. SONUÇ VE ÖNERİLER

Elde edilen bu sonuçlar, ham veri setinin kullanımının özellikle DT ve RFC algoritmaları için yüksek doğruluk oranları sağladığını ortaya koymaktadır. Ancak, aşırı öğrenme riski göz önünde bulundurulmalıdır. Öznitelik seçimi, özellikle LSTM ve SVM gibi modellerin genelleme kapasitesini artırmada olumlu etkiler gösterirken, NB algoritmasının performansını olumsuz yönde etkilemiştir.

Özellikle DT ve RFC algoritmaları için aşırı öğrenmeyi azaltmak amacıyla hiperparametre optimizasyonu yapılması önerilmektedir. Öznitelik seçimi sürecinde, farklı öznitelik seçimi algoritmaları ve optimizasyon yöntemleri kullanılarak sonuçların iyileştirilmesi değerlendirilebilir. Ayrıca, SMOTE gibi veri dengeleme tekniklerinin etkisinin daha iyi anlaşılması için daha detaylı analizler ve farklı kombinasyonlarla testler yapılmalıdır.

Ayrıca öznitelik seçiminin Formula 1 yarış sonuçlarının tahmininde önemli bir rol oynadığını göstermektedir. Formula 1'in karmaşık doğası gereği, seçilecek özniteliklerin sürücünün performansını, araç özelliklerini, hava koşullarını ve pist tipini yansıtacak şekilde dikkatlice belirlenmesi önerilmektedir.

Bu çalışmada, öznitelik seçimi sonrası bazı algoritmalarda tahmin doğruluğunda ciddi düşüşler gözlemlenmiştir. Özellikle Formula 1 gibi sürekli değişen ve dinamik bir sporun verilerini analiz etmek için, genelleme kapasitesi yüksek olan ve çoklu öznitelik etkileşimlerini dikkate alabilen daha karmaşık modeller üzerinde durulabilir.

Sonuç olarak, öznitelik seçimi ve SMOTE gibi veri işleme tekniklerinin makine öğrenimi modellerinin performansı üzerindeki etkileri, model türüne ve veri setinin yapısına bağlı olarak değişmektedir. Öznitelik seçimi ve SMOTE gibi veri işleme tekniklerinin, Formula 1'in karmaşık yapısını yansıtacak şekilde dikkatle uygulanması, gelecekte daha başarılı ve genellenebilir tahmin modelleri geliştirilmesine olanak sağlayacaktır.

KAYNAKLAR

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. In *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)* (pp. 265–283). USENIX Association. <https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf>
- Autosport. (n.d.). *How is artificial intelligence changing Formula 1?* <https://www.autosport.com/f1/news/how-is-artificial-intelligence-changing-formula-1/10659532/>
- Bell, A., Smith, J., Sabel, C. E., & Jones, K. (2016). Formula for success multilevel modelling of Formula One driver and constructor performance, 1950–2014. *Journal of Quantitative Analysis in Sports*, 12(2), 99-112
- Biau, G. (2012). Analysis of a random forests model. *Journal of Machine Learning Research*, 13, 1063-1095. <http://jmlr.org/papers/v13/biau12a.html>
- Blum, C., & Roli, A. (2003). Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Computing Surveys (CSUR)*, 35(3), 268-308. <https://doi.org/10.1145/937503.937505>
- Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press. -> kitap
- Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. In K. Frankish & W. M. Ramsey (Eds.), *The Cambridge Handbook of Artificial Intelligence* (pp. 316-334). Cambridge University Press.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W.W. Norton & Company.
- Carvalho, A. (2022). *Using Historical Data to Identify the Best Driver in Formula 1 History*. National College of Ireland.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321-357. <https://doi.org/10.1613/jair.953>
- Chollet, F. (2021). *Deep learning with Python* (2nd ed., pp. 68-95). Manning Publications.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297. <https://doi.org/10.1007/BF00994018>

- Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21-27. <https://doi.org/10.1109/TIT.1967.1053964>
- Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press.
- Cutler, D. R., Edwards Jr, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2012). Random forests for classification in ecology. *Ecology*, 88(11), 2783-2792. <https://doi.org/10.1890/07-0539.1>
- Dodgins, T., & Webber, M. (2023). *Formula One: The Rivals: F1's Greatest Duels* (ss. 177-187). Ivy Press.
- Donaldson, G. (2012). *Fangio: The Life Behind the Legend*. Random House.
- Ergast. (n.d.). *Motor Racing Data*. <https://ergast.com/mrd/>
- F1 Fansite. (n.d.). *F1 News, Drivers, Results & Standings*. <https://www.f1-fansite.com/>
- Fakhouri, H. N., Hamad, F., & Alawamrah, A. (2022). Success history intelligent optimizer. *The Journal of Supercomputing*, 78(5), 6461-6502. <https://doi.org/10.1007/s11227-021-04093-9>
- Feature. (n.d.). *The history of Formula One racing*. <https://feature.com/blogs/feature-sneaker-boutique/the-history-of-formula-one-racing>
- FIA. (2023). *2026 power unit regulations approved*. <https://www.fia.com/news/2026-power-unit-regulations-approved>
- Formula 1. (2023). *F1 makes significant progress in sustainability as first impact report*. <https://www.formula1.com/en/latest/article/f1-makes-significant-progress-in-sustainability-as-first-impact-report.4jk1NFk81TMXP6EyLdg10k>
- Formula 1. (n.d.). *7 key rule changes for the 2022 season*. <https://www.formula1.com/en/latest/article/7-key-rule-changes-for-the-2022-season.2E7JH9MywymU8xxw6r5yDS>
- García Tejada, L. (2023). *Applying Machine Learning to Forecast Formula 1 Race Outcomes* (Master's thesis, ICT Innovation - Data Science).
- Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1), 3–42. <https://doi.org/10.1007/s10994-006-6226-1>
- Goertzel, B. (2014). Artificial general intelligence: Concept, state of the art, and future prospects. *Journal of Artificial General Intelligence*, 5(1), 1-48. <https://doi.org/10.2478/jagi-2014-0001>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. ->kitab

- GrandPrix247. (2024). *Revolutionizing Formula 1: The Impact of AI & Data Analytics on Formula One in 2024*. <https://www.grandprix247.com/2024/03/26/revolutionizing-formula-1-the-impact-of-ai-data-analytics-on-formula-one-in-2024/>
- GrandPrix247. (2024). *The Role of Mathematical Computing in Formula 1: How AI Artificial Intelligence Helps with Race Analytics*. <https://www.grandprix247.com/2024/09/18/the-role-of-mathematical-computing-in-formula-1-how-ai-artificial-intelligence-helps-with-race-analytics/>
- Graves, A. (2012). Supervised sequence labelling with recurrent neural networks. *Studies in Computational Intelligence*, 385, 5-13. https://doi.org/10.1007/978-3-642-24797-2_book
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... & Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Holland, J. H. (1992). *Adaptation in natural and artificial systems*. MIT Press.-> kitap
- http://www.pacheco.j.com/courses/csc380_fall21/lectures/mlintro.pdf
- <http://yangli-feasibility.com/home/classes/lfd2022fall/media/aaaiws98.pdf>
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90-95. <https://doi.org/10.1109/MCSE.2007.55>
- ITPro. (n.d.). *Mercedes F1 accelerates AI adoption in off-track IT transformation project*. <https://www.itpro.com/cloud/cloud-computing/mercedes-f1-accelerates-ai-adoption-in-off-track-it-transformation-project>
- Jurafsky, D., & Martin, J. H. (2020). *Speech and language processing* (3rd ed.). Pearson.
- Keertish Kumar, M. & Preethi, N. (2023). *Formula One Race Analysis Using Machine Learning*. Lecture Notes in Networks and Systems, 540.
- Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks* (Vol. 4, pp. 1942-1948). IEEE. <https://doi.org/10.1109/ICNN.1995.488968>
- Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598), 671-680. <https://doi.org/10.1126/science.220.4598.671>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097-1105. <https://doi.org/10.1145/3065386>

- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. <https://doi.org/10.1038/nature14539>
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest, *R News*.
- Loh, W.-Y. (2011). Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1), 14-23. <https://doi.org/10.1002/widm.8>
- Marr, B. (2023, July 10). *How artificial intelligence, data, and analytics are transforming Formula One in 2023*. Forbes. <https://www.forbes.com/sites/bernardmarr/2023/07/10/how-artificial-intelligence-data-and-analytics-are-transforming-formula-one-in-2023/>
- McCallum, A., & Nigam, K. (1998). A comparison of event models for naive bayes text classification. In *AAAI-98 Workshop on Learning for Text Categorization* (pp. 41-48).
- McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine*, 27(4), 12–14. <https://doi.org/10.1609/aimag.v27i4.1904>
- McKinney, W. (2010). Data structures for statistical computing in Python. In *Proceedings of the 9th Python in Science Conference* (pp. 51-56). <https://doi.org/10.25080/Majora-92bf1922-00a>
<http://conference.scipy.org.s3.amazonaws.com/proceedings/scipy2010/pdfs/mckinney.pdf>
- Mitchell, T. M. (1997). *Machine learning*. McGraw Hill. kitap
- Motorsport.com. (n.d.). *Red Bull and Oracle partnership to use artificial intelligence in F1*. <https://www.motorsport.com/f1/news/red-bull-artificial-intelligence-oracle-partnership/6271315/>
- O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81-106. <https://doi.org/10.1023/A:1022643204877>
- Racecar Engineering. (n.d.). *Tech explained: 2022 F1 technical regulations*. <https://www.racecar-engineering.com/articles/tech-explained-2022-f1-technical-regulations/>
- Red Bull. (n.d.). *The history of Formula 1*. <https://www.redbull.com/gb-en/history-of-formula-1>

- Reuters. (2024). *F1 Teams Harnessing AI for Speed and Strategy*. <https://www.reuters.com/sports/formula1/f1-teams-harnessing-ai-speed-strategy-2024-06-06/>
- Ricci, F., Rokach, L., & Shapira, B. (2011). *Introduction to recommender systems handbook*. Springer.
- Rish, I. (2001). An empirical study of the naive Bayes classifier. In *IJCAI 2001 Workshop on Empirical Methods in AI* (pp. 41-46).
- Rondelli, M. (2022). *The Future of Formula 1 Racing: Neural Networks to Predict Tyre Strategy* (Bachelor's thesis, Università di Bologna).
- Ruder, S. (2017). An overview of gradient descent optimization algorithms. *arXiv preprint*. <https://arxiv.org/abs/1609.04747>
- Russell, S., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Pearson.-> kitab bende var
- Saudi Motorsport. (n.d.). *Interview: Jeddah gears up to host Formula 1 STC Saudi Arabian Grand Prix*. <https://saudimotorsport.com/en/smc-news/interview-jeddah-gears-up-to-host-formula-1-stc-saudi-arabian-grand-prix>
- Seabiscuit.ai. (2024). *AI in Formula 1 Racing*. <https://www.seabiscuit.ai/ai-in-industry/a/ai-in-formula-1-racing/r/recyyvLGIKXMG9t24>
- Segal, M. R. (2004). Machine learning benchmarks and random forest regression. *Center for Bioinformatics & Molecular Biostatistics*. <https://escholarship.org/uc/item/35x3v9t4>
- Sicoie, H. (2022). *Machine Learning Framework for Formula 1 Race Winner and Championship Standings Predictor* (Bachelor's thesis, Tilburg University).
- Singh, N., & Hachimi, H. (2018). A New Hybrid Whale Optimizer Algorithm with Mean Strategy of Grey Wolf Optimizer for Global Optimization. *Mathematics*, 23(1), 14. <https://doi.org/10.3390/mca23010014>
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press. kitab
- Szeliski, R. (2021). *Computer vision: Algorithms and applications* (2nd ed.). Springer.
- Thieu, N. (n.d.). *Mealpy: A Python library for optimization*. GitHub. <https://github.com/thieu1995/mealpy>
- Thieu, N. T., Nguyen, T. M., & Tran, H. T. (2024). MAFESE: Metaheuristic Algorithms for Feature Selection. *GitHub Repository*. <https://github.com/thieu1995/mafese>
- Trowbridge, K. & Ohlan, R. (2021). *Formula 1 Race Prediction*. MS Data Science, Fordham University.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433-460. <https://doi.org/10.1093/mind/LIX.236.433>

Waskom, M. L. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, 6(60), 3021. <https://doi.org/10.21105/joss.03021>

Wikipedia. (n.d.). *Formula One*. https://en.wikipedia.org/wiki/Formula_One

