

T.C.
BOLU ABANT İZZET BAYSAL ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANA BİLİM DALI



HİBRİT METASEZGİSEL YÖNTEMLER KULLANILARAK
BANKA MÜŞTERİLERİNİN SADAKAT DURUMLARININ
DİNAMİK OLARAK KATEGORİZE EDİLMESİ

BİLGİSAYAR MÜHENDİSLİĞİ YÜKSEK LİSANS TEZİ

OKTAY EFE

TEZ DANIŞMANI
Dr. Öğr. Üyesi Mehmet MİLLİ

BOLU, ARALIK 2024

KABUL VE ONAY SAYFASI

OKTAY EFE tarafından hazırlanan **HİBRİT METASEZGİSEL YÖNTEMLER KULLANILARAK BANKA MÜŞTERİLERİNİN SADAKAT DURUMLARININ DİNAMİK OLARAK KATEGORİZE EDİLMESİ** adlı tez çalışması jürimiz tarafından Bilgisayar Mühendisliği Ana Bilim Dalı Yüksek Lisans Tezi olarak oy birliği kabul edilmiştir. 30/12/2024

Jüri Üyeleri

İmza

Danışman
Dr. Öğr. Üyesi Mehmet MİLLİ
Bolu Abant İzzet Baysal Üniversitesi

.....

Üye
Doç. Dr. Ferhat DEMİRAY
Bolu Abant İzzet Baysal Üniversitesi

.....

Üye
Dr. Öğr. Üyesi Emre ÜNSAL
Sivas Cumhuriyet Üniversitesi

.....

Lisansüstü Eğitim Enstitüsü Onayı

Prof. Dr. İbrahim KÜRTÜL
Lisansüstü Eğitim Enstitüsü Müdürü

ETİK BEYAN

Bolu Abant İzzet Baysal Üniversitesi, Lisansüstü Eğitim Enstitüsü Tez Yazım Kılavuzu ve Şablonuna uygun olarak hazırladığım bu tez çalışmada;

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

Teze ilişkin Turnitin adlı programında enstitü müdürlüğünce belirlenen filtrelemeler uygulanarak alınmış olan benzerlik raporuna göre, tezin benzerlik oranı %30'u geçmemektedir.

.....

OKTAY EFE

ÖZET

**HİBRİT METASEZGİSEL YÖNTEMLER KULLANILARAK BANKA
MÜŞTERİLERİNİN SADAKAT DURUMLARININ DİNAMİK OLARAK
KATEGORİZE EDİLMESİ
YÜKSEK LİSANS TEZİ
OKTAY EFE
BOLU ABANT İZZET BAYSAL ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANA BİLİM DALI
BİLGİSAYAR MÜHENDİSLİĞİ YÜKSEK LİSANS PROGRAMI
(TEZ DANIŞMANI: DR. ÖĞR. ÜYESİ MEHMET MİLLİ)
BOLU, ARALIK - 2024
XI + 51 sayfa**

Günümüzde finansal hizmetler sektörü, rakipleri ile rekabet edebilmek ve devamlı artan müşteri beklentilerini koruyabilmek için sürekli teknolojiye ayak uydurmak ve devamlı kendilerini yenilemek zorundadırlar. Özellikle bankalar, müşteri sadakatini koruma ve artırma konusunda önemli bir mücadele içindedirler. Teknolojik ilerlemeler ve dijital bankacılığın yükselişi ile birlikte, müşterilerin bankalar arasında hareketliliği artmış, bu da bankaları müşteri sadakatini sağlama konusunda daha da dikkatli olmaya yöneltmiştir. Bankalar, müşterilerinin ihtiyaçlarına daha duyarlı, hızlı ve kişiselleştirilmiş hizmetler sunarak rekabet avantajı elde etmek istemektedirler. Bu bağlamda, müşteri sadakatini belirleme sürecinde farklı bakış açıları ve daha etkin yöntemlere ihtiyaç duyulmaktadır. Bankaların müşteri sadakatinde belirlenmede kullanılacak veri setleri içerisinde sayısız öznelite olması geliştirilen sistemlerin performans ve doğruluk oranlarının düşük olmasına sebebiyet vermektedir. Bankaların müşteri sadakatini artırmak için doğru nitelikleri seçmesi, müşteri odaklı stratejiler oluşturması ve rekabet avantajı elde etmesi için hayati öneme sahiptir. Bu bağlamda, müşteri sadakatini artırmak için belirli nitelikleri seçmek ve bu niteliklere dayalı stratejiler oluşturmak bankaların başarısı için kritik bir rol oynar. Bu çalışma kapsamında bankaların müşteri sadakat tespitinde kullanabilecekleri yapay zekâ ve makine öğrenmesi yaklaşımları için kullanılan veri setlerinde bulunan sayısız öznelite içerisinden uygun olanların meta sezgisel yöntemler ile belirlenmesi sağlanmaktadır. Bu çalışmada öznelite seçimi için meta sezgisel algoritmalar Genetik Algoritma ve Yapay Arı Kolonisi Algoritması hibrit olarak kullanılmıştır. Hibrit meta sezgisel yöntemler, farklı meta sezgisel algoritmaların bir araya getirilmesiyle oluşturulan ve optimize edilmiş bir yaklaşımı ifade eder. Bu çalışma kapsamına geliştirilen yöntem, bankaların büyük veri setlerinden müşteri sadakatini daha dinamik ve etkili bir şekilde kategorize etmelerine olanak tanıyabilir. Bu çalışma kapsamında elde edilen nitelikler, müşterilerin ihtiyaçlarına ve tercihlerine uygun stratejiler oluşturmak için temel sağlar ve bankaların uzun vadeli başarısını etkiler.

Anahtar Kelimeler: Genetik Algoritma, Yapay Arı Kolonisi Algoritma, Meta sezgisel Algoritmalar, Yapay Zekâ, Müşteri Sadakati

ABSTRACT

DYNAMICALLY CATEGORIZING THE LOYALTY STATUS OF BANK CUSTOMERS USING HYBRID METAHEURISTIC METHODS

MASTER THESIS

OKTAY EFE

BOLU ABANT İZZET BAYSAL UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF COMPUTER ENGINEERING

(SUPERVISOR: DR. ÖĞR. ÜYESİ MEHMET MİLLİ)

BOLU, DECEMBER 2024

XI + 51 pages

Today, the financial services sector has to constantly keep up with technology and constantly renew itself in order to compete with its competitors and maintain ever-increasing customer expectations. Banks in particular are in a significant struggle to maintain and increase customer loyalty. With technological advances and the rise of digital banking, customer mobility between banks has increased, leading banks to be even more careful about ensuring customer loyalty. Banks want to gain a competitive advantage by offering faster and more personalized services that are more responsive to the needs of their customers. In this context, different perspectives and more effective methods are needed in the process of determining customer loyalty. The fact that there are numerous attributes in the data sets to be used in determining customer loyalty in banks causes the performance and Toplam Sürerates of the developed systems to be low. It is vital for banks to choose the right attributes to increase customer loyalty, create customer-focused strategies and gain competitive advantage. In this context, choosing certain attributes to increase customer loyalty and creating strategies based on these attributes plays a critical role in the success of banks. Within the scope of this study, it is ensured that the appropriate attributes among the numerous attributes found in the data sets used for artificial intelligence and machine learning approaches that banks can use in customer loyalty determination are determined by meta-heuristic methods. In this study, Genetic Algorithm and Artificial Bee Colony Algorithm were used as hybrids for attribute selection from meta-heuristic algorithms. Hybrid meta-heuristic methods refer to an approach created and optimized by combining different meta-heuristic algorithms. The method developed within the scope of this study may allow banks to categorize customer loyalty from large data sets more dynamically and effectively. The attributes obtained within the scope of this study provide the basis for creating strategies suitable for customers' needs and preferences and affect the long-term success of banks.

KEYWORDS: Genetic Algorithm, Artificial Bee Colony Algorithm, Meta-heuristic Algorithms, Artificial Intelligence, Customer Loyalty

İÇİNDEKİLER

Sayfa

KABUL VE ONAY SAYFASI	ii
ÖZET.....	iv
ABSTRACT	v
İÇİNDEKİLER	vi
ŞEKİL LİSTESİ.....	viii
TABLO LİSTESİ	ix
KISALTMA VE SEMBOLLER LİSTESİ	x
TEŞEKKÜR	xi
1. GİRİŞ.....	1
1.1 Tez Çalışmasının Amacı.....	1
1.2 Tez Çalışmasının Organizasyonu	2
1.3 Yapay Zekâ: Güncel Yaklaşımlar ve Uygulamalar	3
1.4 Müşteri Sadakati ve Bankacılıkta Önemi	4
1.5 Öznitelik Seçimi ve Makine Öğrenmesi.....	4
1.6 Meta-Sezgisel Algoritmalar.....	6
1.6.1 Genetik Algoritma (GA).....	6
1.6.2 Yapay Arı Kolonisi (ABC) Algoritması.....	7
1.6.3 Parçacık Sürüsü Optimizasyonu (PSO).....	7
1.6.4 Tabu Arama	8
1.6.5 Simüle Tavlama (Simulated Annealing)	8
1.7 Sınıflandırma Algoritmaları.....	8
1.7.1 Naive Bayes.....	8
1.7.2 Genelleştirilmiş Doğrusal Model (Generalized Linear Model) GLM)	9
1.7.3 Lojistik Regresyon (Logistic Regression)	9
1.7.4 Fast Large Margin	9
1.7.5 Derin Öğrenme (Deep Learning).....	10
1.7.6 Karar Ağacı (Decision Tree)	10
1.7.7 Rastgele Orman (Random Forest)	10
1.7.8 Gradyan Artırmalı Ağaçlar (Gradient Boosted Trees)	10
2. MATERYAL ve YÖNTEM.....	13
2.1 Öznitelik Seçimi	13
2.2 GA ve ABC'nin Hibrit Kullanımı	13
2.3 Büyük Veri Setlerinde Hibrit Algoritmalarla Öznitelik Seçimi	16
2.4 Kullanım Ortamı.....	17
2.5 Avantajlar ve Uygulamalar.....	18
2.6 Bu Çalışma Kapsamında Optimize Edilmiş Alt Veri Setleri.....	19
2.7 Analizde Kullanılan Algoritmalar	20
2.8 Analizde Kullanılan Algoritmaların HiperParametre Bilgileri.....	22
2.8.1 Naive Bayes.....	22
2.8.2 Genelleştirilmiş Doğrusal Model (GLM)	22
2.8.3 Lojistik Regresyon.....	22

2.8.4 Fast Large Margin	22
2.8.5 Derin Öğrenme	23
2.8.6 Karar Ağacı.....	23
2.8.7 Rastgele Orman	23
2.8.8 Gradyan Artırmalı Ağaçlar	23
3. BULGULAR	24
3.1 5 Öznitelikli Veri Bulguları	24
3.2 6 Öznitelikli Veri Bulguları	28
3.3 8 Öznitelikli Veri Bulguları	31
3.4 9 Öznitelikli Veri Bulguları	34
3.5 10 Öznitelikli Veri Bulguları	37
3.6 12 Öznitelikli Veri Bulguları	41
4. TARTIŞMA	46
5. SONUÇ ve ÖNERİLER.....	47
5.1 Sonuç	47
5.2 Öneri	48
5.2.1 Donanım Yatırımı.....	48
5.2.2 Yazılım ve Ortam Optimizasyonu	48
5.2.3 Hibrit Algoritma Kullanımı	49
5.2.4 Öznitelik Sayısının Dengesi	49
KAYNAKLAR	50

ŞEKİL LİSTESİ

Sayfa

Şekil 2.1. Sözde Kod Başlangıç ve Problemin Tanımı.....	14
Şekil 2.2. Sözde Kod – GA Algoritma Tanımı.....	14
Şekil 2.3. Sözde Kod – ABC Algoritma Tanımı	15
Şekil 2.4. Sözde Kod – GA ve ABC Algoritmanın Hibrit Tanımı ve Bitiş.....	16
Şekil 2.5. Seçildi (1) / Seçilmedi (0).....	20
Şekil 3.1. Öznitelik 5 Doğruluk Grafiği	25
Şekil 3.2. Öznitelik 5 Toplam Süre Grafiği	26
Şekil 3.3. Öznitelik 5 Puanlama Süresi Grafiği	26
Şekil 3.4. Öznitelik 5 Korelasyon Sonuçları.....	27
Şekil 3.5. Öznitelik 6 Doğruluk Grafiği	29
Şekil 3.6. Öznitelik 6 Toplam Süre Grafiği	29
Şekil 3.7. Öznitelik 6 Puanlama Süresi Grafiği	30
Şekil 3.8. Öznitelik 6 Korelasyon Sonuçları.....	31
Şekil 3.9. Öznitelik 8 Doğruluk Grafiği	32
Şekil 3.10. Öznitelik 8 Toplam Süre Grafiği	32
Şekil 3.11. Öznitelik 8 Puanlama Süresi Grafiği	33
Şekil 3.12. Öznitelik 8 Korelasyon Sonuçları.....	34
Şekil 3.13. Öznitelik 9 Doğruluk Grafiği	35
Şekil 3.14. Öznitelik 9 Toplam Süre Grafiği	35
Şekil 3.15. Öznitelik 9 Puanlama Süresi Grafiği	36
Şekil 3.16. Öznitelik 9 Korelasyon Sonuçları.....	37
Şekil 3.17. Öznitelik 10 Doğruluk Grafiği.....	39
Şekil 3.18. Öznitelik 10 Toplam Süre Grafiği	39
Şekil 3.19. Öznitelik 10 Puanlama Süresi Grafiği	40
Şekil 3.20. Öznitelik 10 Korelasyon Sonuçları.....	41
Şekil 3.21. Öznitelik 12 Doğruluk Grafiği.....	43
Şekil 3.22. Öznitelik 12 Toplam Süre Grafiği	43
Şekil 3.23. Öznitelik 12 Puanlama Süresi Grafiği	44
Şekil 3.18. Öznitelik 12 Korelasyon Sonuçları.....	45

TABLO LİSTESİ

Sayfa

Tablo 2.1. Sınıflandırma Algoritmaları	21
Tablo 3.1. Öznitelik 5 Alt Veri Seti Performans-Doğruluk Tablosu	25
Tablo 3.2. Öznitelik 6 Alt Veri Seti Performans-Doğruluk Tablosu	28
Tablo 3.3. Öznitelik 8 Alt Veri Seti Performans-Doğruluk Tablosu	31
Tablo 3.4. Öznitelik 9 Alt Veri Seti Performans-Doğruluk Tablosu	34
Tablo 3.5. Öznitelik 10 Alt Veri Seti Performans-Doğruluk Tablosu	38
Tablo 3.6. Öznitelik 12 Alt Veri Seti Performans-Doğruluk Tablosu	42



KISALTMA VE SEMBOLLER LİSTESİ

ABC	: Yapay Arı Kolonisi Algoritması
CNN	: Konvolüsyonel Sinir Ağları
GA	: Genetik Algoritma
GLM	: Genelleştirilmiş Doğrusal Model
PSO	: Parçacık Sürüşü Optimizasyonu
SVM	: Destek Vektör Makineleri
YZ	: Yapay Zekâ



TEŞEKKÜR

Lisansüstü eğitimim süresince, aldığım derslerle birlikte bana kattıkları bilgi birikiminden dolayı öncelikle danışman hocam Sayın Dr. Öğr. Üyesi Mehmet MİLLİ'ye akabinde değerli hocalarım Doç. Dr. Ferhat DEMİRAY, Doç. Dr. Mehmet Dinçer ERBAŞ, Doç. Dr. Önder EYECİOĞLU'na en içten dileklerimle teşekkürlerimi sunarım. Bu süreçte verdiği büyük destekler doğrultusunda her türlü konuda çok iyi yönlendirmesi ve bilgi birikimini aktarması, bu süreçte her zaman hoşgörü ve sabır ile çok büyük destekte bulunan Mehmet Milli Hocam, bu süreçte hem Danışman Hocam hem de Abim olmuştur.

Hayatını çocuklarına adayan okuma yazma bilmediği halde hep 'Okuyacaksınız, hayatınızı kurtaracaksınız.' diyen sevgili annem Sakine EFE'ye, ailesine en iyi şekilde bakmak ve çocuklarını okutmak için yıllarını inşaat sahalarında demirci olarak geçiren her zaman desteğini hissettiğimi sevgili babam İsmet EFE'ye, zorlu süreçlerde yanımda olan ve Lisansüstü eğitimine başlamamdaki en büyük destekçim olan sevgili eşim Semanur ÖZTÜRK EFE'ye sonsuz teşekkür ederim. Ayrıca lisansüstü eğitime başlarken hayatımda olmayan sonrasında hayatıma giren ve eğitimimi tamamladığımda yanımda olacak ayrıca bunu çok uzun süre sonra okuyacak olan daha yaşını doldurmamış sevgili kızım Asya Hüma EFE sevgilerimle.

1. GİRİŞ

Bankalar müşterilerini memnun etmek için sürekli çalışmalar yapmaktalar. Bankalar için müşteri memnuniyetinin önemli olması kadar müşteri sadakatini de bir o kadar önemli. Müşteri sadakatini belirleyen faktörlerin en başında, müşterinin bankada gerçekleştirdiği işlemler ve bu işlemlere dayalı olarak gösterdiği hareketler yer almaktadır. Bu hareketler müşterinin bankaya olan sadakatini yansıtmaktadır. Müşteri sadakatini, finansal hizmetler sektöründe başarılı bir rekabet avantajı sağlamanın anahtarı haline gelmiştir. Bankalar müşterilerini elde tutmak ve onları uzun vadeli sadakatini kazanmak için çeşitli stratejiler uygulamaktadır. Bankaların sadakat programları, müşterilerine sağladıkları ek hizmetler ve avantajlarla birlikte finansal deneyimlerini daha cazip hale getirirken, müşteri memnuniyetini artırmak için etkili bir araç olarak öne çıkıyor. Bu programlar, müşterilerin belirli bankalarla uzun vadeli ilişkiler kurmalarını teşvik ederken, aynı zamanda bankalar için sadık müşteri tabanı oluşturma'nın önemli bir yolunu temsil etmektedir. Asıl amaç bankaların müşteri sadakatini arttırmak ve müşteri kaybını en aza indirmektir.

Makine öğrenmesi ve veri madenciliği alanındaki araştırmalar, veri setlerinin karmaşıklığını azaltarak ve daha etkili modelleme sağlayarak öznitelik çıkarma tekniklerine büyük önem vermektedir. Genetik algoritma ve yapay arı kolonisi algoritması gibi doğa esinli optimizasyon yöntemleri, bu bağlamda dikkat çeken ve başarılı sonuçlar elde eden teknikler arasındadır. Bu makalede, genetik algoritma ve yapay arı kolonisi algoritması birlikte kullanılarak öznitelik çıkarma yönteminin detayları incelenecek ve bu yaklaşımın avantajları üzerinde durulacaktır.

1.1 Tez Çalışmasının Amacı

Bu çalışmanın temel amacı, banka müşterilerinin memnuniyetini artırmak ve müşteri sadakatini ölçmek amacıyla canlı veri akışı üzerinde yüksek performans ve doğruluk oranı sağlayan analizler gerçekleştirmektir. Bu bağlamda, müşteri verileri üzerinde etkili ve hızlı analiz yapılarak, müşteri davranışları ve sadakatine yönelik stratejiler geliştirilmesi hedeflenmektedir.

Bankacılık işlemlerinden elde edilen canlı veri akışı üzerinde analiz gerçekleştirilmesi ve müşteri sadakatinin yüksek doğrulukla ölçülmesi sağlanmıştır.

Verilerin işleme aşamasında Genetik Algoritma (GA) ve Yapay Arı Kolonisi (ABC) algoritmaları ile öznitelik seçimi yapılmıştır. Bu sayede verinin boyutunun optimize edilmesi ve gereksiz özniteliklerin elenmesi sağlanmıştır.

Öznitelik seçimi sonrası Yapay Zekâ (YZ) algoritmaları kullanılarak veriler analiz edilmiştir. Özellikle canlı ortamda analiz yapılacağı göz önünde bulundurularak, doğruluktan minimum ödün verilmiş ve performans odaklı bir yaklaşımla gerçek zamanlı analiz gerçekleştirilmiştir. Bu çalışma, özellikle büyük veri setlerinin işlenmesi ve analizinde yüksek doğruluk oranı ve hızlı performans gereksinimlerine çözüm sunmayı hedeflemektedir.

1.2 Tez Çalışmasının Organizasyonu

Bu tez çalışmasının bundan sonraki bölümlerin aşağıdaki şekilde organize edilmiştir.

İkinci Bölüm: İlgili literatür araştırmalarını kapsamaktadır. Bu bölümde, müşteri sadakati, öznitelik seçimi, Genetik Algoritma (GA) ve Yapay Arı Kolonisi (ABC) algoritmalarının literatürdeki temel uygulamaları ve bu alanda yapılan çalışmalar detaylı bir şekilde incelenmiştir.

Üçüncü Bölüm: Çalışmada kullanılan yöntem ve araçlar açıklanmıştır. Öznitelik seçimi, veri hazırlama süreci ve Yapay Zekâ algoritmalarının kullanımı detaylı bir şekilde ele alınmıştır.

Dördüncü Bölüm: Tez çalışmasında yapılan uygulamalar ve analiz sonuçları sunulmuştur. Bu bölümde, öznitelik seçim sonuçları ve kullanılan algoritmaların performansları karşılaştırmalı olarak değerlendirilmiştir.

Beşinci Bölüm: Tez çalışmasından elde edilen bulguların değerlendirilmesi ve sonuçları içermektedir. Ayrıca, bu bölümde çalışmanın katkıları, sınırlılıkları ve gelecekte yapılabilecek araştırmalar için öneriler sunulmuştur.

Bu organizasyon sayesinde, çalışmanın amaçları, yöntemleri, sonuçları ve önerileri adım adım bir mantık silsilesi içerisinde okuyucuya sunulmuştur. Böylece, okuyucuya çalışmanın her aşamasında net bir bilgi akışı sağlanmış ve araştırmanın bilimsel temellere dayandığı açık bir şekilde gösterilmiştir. Aynı zamanda, bölümlerin birbiriyle ilişkili yapısı, çalışmanın bütünselliğini korumakta ve elde edilen bulguların analizinin daha iyi anlaşılmasını sağlamaktadır. Bu yapı, tezde ele

alınan problemlere yönelik çözümlerin etkili bir şekilde ifade edilmesine de katkıda bulunmaktadır.

1.3 Yapay Zekâ: Güncel Yaklaşımlar ve Uygulamalar

Yapay zekâ (YZ), bilgisayarların insan benzeri karar verme, öğrenme ve problem çözme yeteneklerini taklit etmesini sağlayan bir teknoloji alanıdır. Bu alanda yapılan çalışmalar, özellikle makine öğrenimi ve derin öğrenme gibi alt disiplinlerle hız kazanmıştır. YZ'nin temel amacı, algoritmalar ve veri odaklı modeller aracılığıyla, insan müdahalesi olmaksızın karmaşık problemlere çözüm sunmaktır. 1950'li yıllarda Alan Turing'in "Makine Zekâsı" konseptiyle başlayan bu süreç, günümüzde sağlık, finans, otomotiv ve birçok sektörde çığır açan gelişmelere yol açmıştır (Turing, 1950). Makine öğrenmesi ile büyük veri analitiğini birleştiren yaklaşımlar, YZ'nin özellikle özelleşmiş görevlerde üstün performans göstermesini sağlamaktadır (Russell ve Norvig, 2016).

YZ teknolojileri, veri analitiği, görsel tanıma, doğal dil işleme ve robotik sistemlerde geniş bir uygulama alanına sahiptir. Örneğin, sağlık sektöründe yapay zekâ, tıbbi görüntüleme ile kanser teşhisinde insan uzmanlardan daha yüksek doğruluk oranlarına ulaşmaktadır. Finans sektöründe ise YZ, dolandırıcılık tespiti ve müşteri davranış analizi gibi alanlarda etkili çözümler sunmaktadır. Derin öğrenme algoritmaları, özellikle büyük veri setlerinde, karmaşık yapıları ve örüntüleri öğrenerek başarılı sonuçlar elde etmektedir. Örneğin, LeCun ve arkadaşlarının geliştirdiği konvolüsyonel sinir ağları (CNN) algoritması, görüntü sınıflandırmada çığır açan bir yaklaşım olarak kabul edilmektedir (LeCun vd., 2015).

YZ'nin gelişimiyle birlikte, etik ve sosyal sorumluluk gibi konular da giderek daha fazla tartışılmaktadır. Özellikle önyargılı algoritmalar ve veri gizliliği ihlalleri gibi sorunlar, bu teknolojilerin toplum üzerindeki etkilerini sorgulatmaktadır. Etik yapay zekâ insan merkezli tasarım prensiplerini ön planda tutarak, daha güvenilir ve adil sistemlerin geliştirilmesini amaçlamaktadır. Bu bağlamda, literatürdeki çalışmalar, YZ'nin olumlu etkilerini maksimize ederken olumsuz etkilerini en aza indirmek için öneriler sunmaktadır. Örneğin, Floridi ve Cows (2019), etik yapay zekanın beş temel prensibini (yararlılık, zarar vermeme, özerklik, adalet ve açıklık) önermektedir.

1.4 Müşteri Sadakati ve Bankacılıkta Önemi

Müşteri sadakati, bankacılık sektöründe rekabet avantajı elde etmek için kritik bir faktör olarak kabul edilmektedir. İlk çalışmalar, müşteri sadakatının finansal sonuçlar üzerindeki etkisini ve müşteri davranışlarının sadakat üzerindeki rolünü vurgulamıştır (Reichheld ve Teal, 1996). Bu araştırmalarda müşteri sadakatını artıran etkenlerin tanımlanması ve uzun vadeli müşteri ilişkilerinin önemi ele alınmıştır.

Bankaların müşteri memnuniyetine yönelik stratejileri ve sadakat programları üzerine yapılan çalışmalar, sadık müşteri tabanının bankalar için ekonomik değer yarattığını göstermektedir (Colgate ve Danaher, 2000). Bu bağlamda, bankaların müşteri kaybını en aza indirmeye yönelik stratejik adımları vurgulanmaktadır.

Sadakat, yalnızca finansal kazanç sağlamanın ötesinde, bankaların pazardaki güvenilirliğini ve marka değerini artırmada da kilit bir rol oynamaktadır. Günümüz dijital çağında, müşterilerin farklı bankalar arasında kolayca geçiş yapabildiği bir ortamda, sadakat programlarının etkili bir şekilde uygulanması daha da önem kazanmaktadır. Örneğin, dijital hizmetlerin kişiselleştirilmesi ve müşteri ihtiyaçlarının gerçek zamanlı olarak karşılanması, sadakati artıran önemli faktörlerdendir. Ayrıca, müşteri sadakatını güçlendirmek, yalnızca mevcut müşterilerle sınırlı kalmamakta, aynı zamanda ağızdan ağıza pazarlama etkisiyle yeni müşteri kazanımını da desteklemektedir. Bu bağlamda, bankaların teknolojiyi ve veri analitiğini etkin bir şekilde kullanarak, müşteri beklentilerini proaktif bir şekilde karşılaması, uzun vadeli başarı için kaçınılmaz bir gereklilik olarak öne çıkmaktadır.

1.5 Öznitelik Seçimi ve Makine Öğrenmesi

Öznitelik seçimi, makine öğrenmesi ve veri madenciliği süreçlerinde önemli bir adım olup, veri setindeki gereksiz veya etkisi düşük özniteliklerin elenmesi ve etkili özniteliklerin belirlenmesi için kullanılan bir tekniktir. Bu süreç, hem modelin performansını artırmak hem de hesaplama maliyetlerini düşürmek amacıyla gerçekleştirilir. Özellikle büyük veri setleri ile çalışırken, gereksiz özniteliklerin

varlığı modelin doğruluğunu olumsuz etkileyebileceği gibi, öğrenme süresini de uzatabilir. Öznitelik seçimi, yalnızca veri boyutunu azaltmakla kalmaz, aynı zamanda modelin genelleme yeteneğini artırarak aşırı öğrenme (overfitting) riskini de minimize eder. Bu bağlamda, öznitelik seçiminin makine öğrenmesi performansı üzerindeki olumlu etkilerini vurgulamış ve bu tekniklerin veri madenciliğinde kritik bir rol oynadığını ifade etmiştir (Guyon ve Elisseeff, 2003).

Öznitelik seçiminin makine öğrenmesi algoritmalarının doğruluğunu artırdığı ve modelin eğitim süresini azalttığı üzerine yapılan araştırmalar, veri madenciliğinde bu tekniğin önemini ortaya koymuştur. Özellikle büyük veri setlerinde öznitelik seçimi, algoritmaların performansını artırmak için önemli bir adım olarak kabul edilmektedir (John vd., 1994).

Makine öğrenmesi modellerinde öznitelik seçimi, hem denetimli hem de denetimsiz algoritmalar için uygulanabilmektedir. Örneğin, denetimli öğrenme algoritmalarında, etiketli veriler kullanılarak özniteliklerin hedef değişken üzerindeki etkisi analiz edilirken, denetimsiz öğrenmede verilerin yapısal özelliklerinden faydalanılarak öznitelikler değerlendirilir. Öznitelik seçimi yöntemleri genel olarak üç kategoriye ayrılmaktadır: filtre yöntemleri, sarma (wrapper) yöntemleri ve gömülü (embedded) yöntemler. Filtre yöntemleri, özniteliklerin istatistiksel özelliklerini analiz ederek seçim yaparken, sarma yöntemleri modeli doğrudan sürece dahil eder ve özniteliklerin model performansı üzerindeki etkisine göre seçim yapar. Gömülü yöntemler ise öznitelik seçim sürecini modelin öğrenme süreciyle entegre eder. Bu yöntemlerin her biri, veri setinin yapısına ve problem türüne bağlı olarak farklı avantajlar sunar ve makine öğrenmesi süreçlerini optimize etmek için kritik bir araç olarak kullanılır.

Bu çalışmada, öznitelik seçimi sürecinde genetik algoritma (GA) ve yapay arı kolonisi (ABC) algoritması hibrit bir şekilde kullanılmıştır. Genetik algoritma, doğal seleksiyon ve genetik mutasyon prensiplerini temel alarak popülasyon tabanlı bir optimizasyon tekniği sunarken, yapay arı kolonisi algoritması, arıların besin kaynaklarını bulma davranışlarını taklit ederek optimizasyon sağlar. Bu iki doğa esinli algoritmanın hibrit kullanımı, veri setindeki özniteliklerin etkili bir şekilde seçilmesine olanak tanımıştır. Genetik algoritma, çözüm uzayını keşfetme ve yüksek etkili öznitelik kombinasyonlarını belirleme sürecinde güçlü bir araç olarak kullanılırken, yapay arı kolonisi algoritması, daha lokal düzeyde inceleme yaparak yüksek doğruluk oranları elde edilmesini sağlamıştır. Bu hibrit yaklaşım, veri

setinin performansını artırmakla kalmamış, aynı zamanda daha hızlı ve daha etkili modellerin geliştirilmesine olanak tanımıştır.

1.6 Meta-Sezgisel Algoritmalar

Metasezgisel algoritmalar, karmaşık optimizasyon problemlerini çözmek için geliştirilen, genel amaçlı ve problem bağımsız yöntemlerdir. Bu algoritmalar, doğadan ilham alarak sezgisel yaklaşımları bir araya getirir ve geniş bir çözüm uzayını etkili bir şekilde aramak için tasarlanmıştır. Metasezgisel yöntemler, özellikle çözüm uzayı çok geniş veya geleneksel algoritmaların yetersiz kaldığı durumlarda kullanılır. Bu yöntemler, genellikle popülasyon tabanlıdır ve çoklu çözüm adayları üzerinde çalışarak küresel optimuma ulaşmaya çalışır. Ayrıca, yerel minimumlara takılmayı önlemek için keşif (exploration) ve sömürü (exploitation) arasında bir denge kurar.

Metasezgisel algoritmalar, çeşitli alanlarda kullanılmak üzere farklı ilkelere dayanır. Bunlar arasında genetik algoritma, yapay arı kolonisi algoritması, parçacık sürüsü optimizasyonu ve tabu arama gibi popüler yöntemler bulunmaktadır. Her bir algoritma, belirli bir optimizasyon sürecini modellemek için doğadan veya insan yapımı sistemlerden ilham alır ve kendi güçlü yönleriyle karmaşık problemlere çözümler sunar.

1.6.1 Genetik Algoritma (GA)

Genetik algoritma, biyolojik evrimden ilham alarak doğal seleksiyon ve genetik mutasyon süreçlerini taklit eden bir optimizasyon yöntemidir. İlk olarak John Holland tarafından tanıtılan bu algoritma, popülasyon tabanlı bir yaklaşıma dayanır. Çözüm uzayında çeşitli bireyler (çözüm adayları) oluşturulur ve her biri bir uygunluk (fitness) değeri ile değerlendirilir. Genetik algoritmanın temel mekanizmaları seçim, çaprazlama (crossover) ve mutasyon işlemleridir. Bu işlemler, daha iyi çözümler üretmek için popülasyonu sürekli olarak geliştirir. GA, özellikle karmaşık ve çok boyutlu optimizasyon problemlerinde kullanılır ve küresel optimuma ulaşma potansiyeli yüksektir.

Genetik algoritmanın adaptif sistemler üzerindeki etkisini vurgulamış ve doğal seleksiyon süreçlerinin bilgisayarlı optimizasyona uyarlanabilirliğini göstermiştir (Holland, 1975). Ayrıca genetik algoritmaların mühendislik problemleri ve çoklu optimizasyon gerektiren alanlarda nasıl etkili bir çözüm

sunduğunu ortaya koyarak, bu yöntemin geniş bir uygulama alanına sahip olduğunu belirtmiştir (Goldberg, 1989).

1.6.2 Yapay Arı Kolonisi (ABC) Algoritması

Yapay arı kolonisi algoritması, arıların besin kaynaklarını bulma ve optimize etme davranışlarından ilham alır. 2005 yılında Derviş Karaboğa tarafından geliştirilen bu algoritma, sürekli optimizasyon problemleri için etkili bir araç olarak kabul edilir. ABC algoritması, işçi arılar, izci arılar ve gözlemci arılar olmak üzere üç ana bileşene dayanır. İşçi arılar mevcut besin kaynaklarını iyileştirmeye çalışırken, izci arılar yeni kaynaklar keşfetmek için rastgele araştırma yapar. Gözlemci arılar ise mevcut kaynakların kalitesine göre karar verir. ABC algoritması, düşük hesaplama maliyetleri ve geniş çözüm uzayını etkili bir şekilde araştırma kabiliyetiyle öne çıkar.

Yapay arı kolonisi algoritmasının özellikle sürekli optimizasyon problemleri için rekabetçi bir yöntem olduğunu ve basitliği sayesinde geniş bir uygulama alanına hitap ettiğini ifade etmiştir (Karaboğa, 2005). Bu algoritmanın biyolojik sistemlerden ilham alan doğası, veri analizi ve mühendislik problemlerinde yenilikçi çözümler sunmasını sağlamıştır.

1.6.3 Parçacık Sürüsü Optimizasyonu (PSO)

Parçacık sürüsü optimizasyonu, kuş sürüleri ve balık sürülerinin koordineli hareketlerinden ilham alınarak geliştirilmiş bir algoritmadır. Her bir parçacık, çözüm uzayında bir konuma sahiptir ve hız ile yönlendirilir. Algoritma, her bir parçacığın kendi en iyi konumunu ve sürünün genel en iyi konumunu dikkate alarak hareket etmesini sağlar. PSO, özellikle sürekli ve yüksek boyutlu optimizasyon problemlerinde etkili sonuçlar sunar. Ayrıca, PSO'nun uygulanması kolaydır ve genellikle hızlı bir yakınsama sağlar.

PSO'nun adaptif bir optimizasyon algoritması olarak performansını değerlendirirken, doğal sürü davranışlarının çözümlerin hızlı bir şekilde iyileştirilmesine olan katkısını vurgulamıştır (Kennedy ve Eberhart, 1995). Ayrıca, bu algoritmanın düşük işlem maliyetleri ile geniş çözüm uzaylarını etkili bir şekilde keşfetme yeteneği olduğu belirtilmiştir.

1.6.4 Tabu Arama

Tabu arama, yerel arama tabanlı bir metasezgisel algoritmadır ve belirli bir problemin çözüm uzayında daha iyi çözümler bulmak için kullanılır. Algoritma, belirli bir süre boyunca tekrarlanan çözümleri engellemek için bir "tabu listesi" tutar. Bu yaklaşım, algoritmanın yerel minimumlarda takılmasını önler ve daha geniş bir çözüm uzayını araştırmasına olanak tanır. Tabu arama, özellikle zaman kısıtlaması olan veya kombinatoriyal optimizasyon problemlerinde tercih edilir.

Tabu aramanın optimizasyon problemlerinde yerel minimumların ötesine geçmek için etkili bir strateji sunduğunu ifade etmiş ve bu yöntemin kombinatoriyal problemler üzerinde güçlü bir etkisi olduğunu göstermiştir (Glover, 1986). Tabu arama, özellikle lojistik, planlama ve zamanlama gibi operasyonel alanlarda önemli uygulamalara sahiptir.

1.6.5 Simüle Tavlama (Simulated Annealing)

Simüle tavlama, termodinamikteki tavlama sürecinden ilham alır ve optimizasyon problemleri için kullanılır. Algoritma, yüksek enerjili (daha kötü) çözümleri kabul ederek yerel minimumlardan kaçmaya çalışır ve zamanla daha iyi çözümler bulmaya odaklanır. Tavlama sıcaklığı adı verilen bir parametre, algoritmanın kabul oranını kontrol eder ve zamanla soğur. Simüle tavlama, özellikle yerel arama yöntemleri ile birleştiğinde etkili sonuçlar sunar.

Simüle tavlamayı fiziksel tavlama sürecinden bilgisayarlı optimizasyona uyarlayarak, bu yöntemin karmaşık problemlerde geniş bir çözüm alanı sunduğunu göstermiştir (Kirkpatrick vd., 1983). Ayrıca, algoritmanın kabul ettiği kötü çözümler sayesinde yerel minimumlardan kurtulma yeteneği üzerinde durulmuştur.

1.7 Sınıflandırma Algoritmaları

1.7.1 Naive Bayes

Naive Bayes, Bayes teoremine dayalı bir olasılık modeli olup özellikle sınıflandırma problemlerinde kullanılır. Naive Bayes'in en önemli varsayımı, modelin öznitelikler arasında bağımsızlık varsayımıdır, yani bir öznitelik değerinin diğerinden bağımsız olduğunu kabul eder. Bu basit varsayıma rağmen, Naive Bayes özellikle metin sınıflandırma, spam filtreleme ve duygu analizi gibi uygulamalarda oldukça başarılıdır. Model, çok hızlıdır ve büyük veri setleri üzerinde uygulanabilir.

Dezavantajı, bağımsızlık varsayımının gerçek dünyada her zaman geçerli olmamasıdır, bu da bazı durumlarda modelin performansını düşürebilir. Bu basit ve hızlı yapısı nedeniyle Naive Bayes, özellikle düşük hesaplama maliyetinin önemli olduğu durumlarda tercih edilmektedir, ancak bağımsızlık varsayımının ihlal edildiği karmaşık veri setlerinde daha sofistike algoritmalara ihtiyaç duyulabilir.

1.7.2 Genelleştirilmiş Doğrusal Model (Generalized Linear Model) GLM)

GLM, klasik doğrusal regresyonun bir genellemesidir. Temel farkı, doğrusal regresyonun aksine, GLM'nin bağımlı değişkenin belirli bir dağılımına uyan tahminler yapmasıdır. Link fonksiyonu kullanarak, farklı olasılık dağılımlarına uyum sağlayabilir. Örneğin, logit fonksiyonu kullanarak ikili sınıflandırma yapılabilir. GLM, geniş veri tipleri üzerinde uygulanabilen esnek bir modeldir ve özellikle sigortacılık, biyostatistik ve sosyal bilimlerde yaygın olarak kullanılır. Yinede, doğrusal bağımlılıklara dayanması, çok karmaşık veri ilişkilerinde performansını sınırlandırabilir.

1.7.3 Lojistik Regresyon (Logistic Regression)

Lojistik regresyon, doğrusal bir regresyon modeli gibi çalışır, ancak bağımlı değişkenin iki veya daha fazla kategoride olduğu durumlar için kullanılır. En yaygın uygulaması ikili sınıflandırmadır (örneğin, hasta mı sağlıklı mı). Lojistik fonksiyon, çıktıların bir olasılık dağılımı gibi yorumlanmasını sağlar. Modelin basitliği, onu özellikle küçük veri setlerinde ve kolay yorumlanabilir sonuçlar istenilen alanlarda ideal kılar. Ancak, lojistik regresyonun doğrusal ilişkiler varsayması, karmaşık veri setlerinde düşük performans göstermesine neden olabilir.

1.7.4 Fast Large Margin

Fast Large Margin, destek vektör makinelerine (SVM) benzer şekilde çalışan bir modeldir ve büyük marjlara dayalı olarak sınıflandırma yapar. Bu algoritma, veriyi daha iyi ayırabilmek için sınır çizgilerini olabildiğince büyük tutmaya çalışır. Özellikle sınıflandırma problemlerinde yüksek doğruluk sağlayabilir. Algoritmanın avantajı, hızlı bir şekilde çalışması ve büyük veri setleri üzerinde etkili sonuçlar verebilmesidir. Dezavantajı ise, eğitim sürecinde doğru parametre ayarları yapılmazsa performansın düşebileceğidir.

1.7.5 Derin Öğrenme (Deep Learning)

Derin öğrenme, çok katmanlı yapay sinir ağlarına dayanan bir makine öğrenme yaklaşımıdır. Bu model, özellikle büyük veri setlerinde ve çok karmaşık örüntülerin olduğu durumlarda güçlü sonuçlar sağlar. Derin öğrenme algoritmaları, katmanlı bir öğrenme süreci ile verilerdeki özellikleri otomatik olarak keşfeder ve öğrenir. Görüntü işleme, doğal dil işleme ve zaman serisi tahmini gibi alanlarda sıklıkla kullanılır. Ancak, derin öğrenme modelleri büyük miktarda veri ve hesaplama gücü gerektirir ve modelin aşırı öğrenmeye yatkın olması potansiyel bir dezavantajdır. Bu nedenle, hiperparametre optimizasyonu kritik bir rol oynar.

1.7.6 Karar Ağacı (Decision Tree)

Karar ağaçları, verileri dallara ayırarak sınıflandırma veya regresyon yapmaya çalışan sezgisel bir algoritmadır. Model, her adımda veriyi bir özneliğe göre ayırarak alt dallar oluşturur ve sonuca ulaşır. Karar ağaçlarının en büyük avantajı, görselleştirilmelerinin kolay olmasıdır. Aynı zamanda, eğitim ve tahmin süreçleri oldukça hızlıdır. Ancak, karar ağaçları aşırı öğrenmeye yatkındır ve bu da modelin genelleme yeteneğini olumsuz etkileyebilir. Aşırı büyüyen ağaçlar, hatalı kararlar alabilir, bu yüzden budama teknikleri kullanılarak ağaç boyutları sınırlandırılmalıdır.

1.7.7 Rastgele Orman (Random Forest)

Rastgele Orman, karar ağaçlarının topluluk yöntemine dayalı bir genişlemesidir. Model, birçok farklı karar ağacı oluşturur ve her bir ağaç bağımsız olarak eğitilir. Sonrasında, her ağacın çıktısı oy birliği ile son tahmini oluşturur. Bu yöntem, tek bir karar ağacına kıyasla aşırı öğrenmeyi azaltır ve genelleme yeteneğini artırır. Rastgele Orman, hem sınıflandırma hem de regresyon problemlerinde yüksek doğruluk sağlar. Ancak, büyük miktarda ağaç oluşturulması hesaplama açısından maliyetli olabilir.

1.7.8 Gradyan Artırmalı Ağaçlar (Gradient Boosted Trees)

Aşamalı güçlendirme, her adımda önceki adımların hatalarını azaltmayı amaçlayan bir algoritmadır. Karar ağaçları, ardışık olarak inşa edilir ve her yeni ağaç, önceki ağaçların hatalarını düzeltmeye çalışır. Bu süreç, daha yüksek doğruluk ve daha az hata oranı sağlar. Aşamalı güçlendirilmiş ağaçlar, özellikle

sınıflandırma ve regresyon problemlerinde oldukça etkilidir. Ancak, modelin aşırı öğrenmeye karşı dikkatli bir şekilde optimize edilmesi gerekir.

Bu çalışmada, farklı veri setleri üzerinde öznitelik seçim işlemleri tamamlandıktan sonra, çeşitli makine öğrenme algoritmaları ile modellerin doğruluk ve performanslarını değerlendirdik. Her bir algoritma, sınıflandırma ve regresyon problemlerine farklı yaklaşımlar sunarak, veri setinin yapısına göre en iyi performansı elde etmeyi hedeflemiştir. Kullanılan modeller arasında, hem hızlı ve basit algoritmalar (Naive Bayes, Lojistik Regresyon) hem de daha karmaşık ve derin öğrenme yöntemlerine dayanan modeller (Derin Öğrenme, Aşamalı Güçlendirilmiş Ağaçlar) yer almaktadır. Bu algoritmalar, veri setlerindeki örüntüleri belirlemek, önemli öznitelikleri seçmek ve doğru tahminler yapmak amacıyla kullanılmaktadır.

Modeller, veri setlerindeki özniteliklerin doğru şekilde analiz edilmesini sağlarken, aynı zamanda farklı metriklerle performanslarını değerlendirmek için kullanılmıştır. Örneğin, Naive Bayes gibi hızlı ve basit modeller, düşük boyutlu ve bağımsız özniteliklere sahip veri setlerinde etkili olurken, Derin Öğrenme gibi karmaşık modeller, büyük veri setlerinde ve çok sayıda öznitelige sahip karmaşık veri yapılarında daha iyi sonuçlar elde etmektedir. Karar Ağaçları ve Rastgele Ormanlar gibi ağaç tabanlı modeller ise veri setlerini dallara ayırarak ayrıntılı analizler yapar ve aşırı öğrenme riskini azaltmak için topluluk yöntemlerini kullanır.

Modellerin performansı; doğruluk (accuracy), hassasiyet (precision), hatırlama (recall) ve F1 skoru gibi sınıflandırma metrikleriyle değerlendirilmiştir. Ayrıca, her bir modelin genelleme yeteneği, aşırı öğrenme riskini en aza indirmek için test edilmiş ve çapraz doğrulama yöntemleriyle güvenilirlikleri kontrol edilmiştir. Bu değerlendirmeler sonucunda, her bir modelin güçlü ve zayıf yönleri ortaya konmuş ve belirli veri yapıları için hangi modellerin daha uygun olduğu belirlenmiştir.

Çalışmamızda kullanılan bu geniş model çeşitliliği, veri setlerinin farklı boyut ve karmaşıklıklarına uyum sağlamakta, dolayısıyla her modelin hangi durumlarda daha iyi performans gösterdiğini anlamamızı kolaylaştırmaktadır. Sonuç olarak, her bir algoritmanın belirli veri setlerinde gösterdiği performans, model seçiminde kritik bir rol oynar ve farklı problemler için en uygun çözüm yolunun belirlenmesine katkı sağlar.

Model çeşitliliği, makine öğrenmesi uygulamalarında belirli veri setleri için özelleştirilmiş ve optimize edilmiş çözümler geliştirmek açısından büyük önem taşır. Her modelin belirli veri türlerinde performansını değerlendirmek, yalnızca doğruluk oranını artırmakla kalmaz, aynı zamanda sistemin genel verimliliğini ve kaynak yönetimini optimize eder (Li vd., 2019).

Ayrıca, bu çeşitlilik, farklı veri yapıları ve öznitelik dağılımlarının model performansı üzerindeki etkisini daha iyi değerlendirme olanağı sunar. Bu durum, algoritmaların güçlü ve zayıf yönlerini belirleyerek, gelecekteki çalışmalar için daha optimize stratejiler geliştirilmesine olanak tanır. Bu sayede, doğru model seçimi ile yalnızca yüksek doğruluk elde edilmekle kalmayıp, aynı zamanda işlem süresi ve kaynak kullanımı açısından da verimli çözümler geliştirilmesi mümkün olmaktadır.

Son olarak, model çeşitliliği, makine öğrenimi algoritmalarının genelleme yeteneği üzerinde de önemli bir rol oynar. Özellikle, farklı modellerin topluluk yöntemleriyle birleştirilmesi, genelleme hatalarını azaltarak daha doğru ve kararlı tahminler yapılmasına olanak sağlar. Bu yöntemler, sınıflandırma ve regresyon problemlerinde sıklıkla tercih edilmekte ve bireysel modellerin eksikliklerini gidermek için güçlü bir alternatif sunmaktadır (Dietterich, 2000). Topluluk yöntemleri, bireysel modellerin önyargı ve varyans kaynaklı hatalarını telafi ederek, daha dengeli ve güvenilir sonuçlar elde edilmesini mümkün kılar.

2. MATERİYAL ve YÖNTEM

2.1 Öznitelik Seçimi

Bu çalışmada kullanılan veri seti, bir bankanın müşteri kaybı tahminine yönelik Kaggle platformundan elde edilmiştir (Konstantin, 2020). Veri seti, 23 özellik ve 10.127 gözlemden oluşmakta olup, müşterilerin demografik bilgileri, bankacılık hizmetleri kullanımı ve finansal durumu gibi çeşitli faktörleri içermektedir. Bu veriler, banka müşterilerinin kayıp durumlarını analiz etmek ve tahmin modelleri geliştirmek amacıyla kullanılmıştır.

Veri setinde öncelikle öznitelik seçimi gerçekleştirilerek veri hazırlıkları yapıldı. Bu bağlamda GA ve ABC'nin hibrit kullanılması ile veri setinin sonuçlanan özniteliklerinin seçimleri yapıldı. Farklı öznitelik seçimleri ile birlikte elde edilen farklı veri setleri üzerinde gerekli performans ve doğruluk çalışmaları sonrası farklı sonuçlar elde edilerek çalışmanın amacına ulaşılmasına yarayacaktır.

2.2 GA ve ABC'nin Hibrit Kullanımı

GA ve ABC'nin hibrit kullanımı üzerine yapılan ilk çalışmalar, her iki algoritmanın güçlü yönlerini birleştirilerek daha hızlı ve etkili optimizasyon çözümleri sunabileceğini ortaya koymuştur. Bu bağlamda, Wang ve diğerleri (2011), GA ve ABC'nin hibrit kullanımının, çok boyutlu veri setlerinde öznitelik seçim performansını artırdığını ve böylece yüksek doğruluk oranları elde edildiğini göstermiştir. Hibrit meta-sezgisel yaklaşımlar konusunda yapılan bir diğer çalışma, genetik algoritmanın keşif yeteneğini ve ABC'nin sömürü yeteneğini bir araya getirilerek optimize edilen çözümlerin doğruluğunu artırmayı hedeflemiştir (Kumar ve Singh, 2018). Bu hibrit yöntem, özellikle veri madenciliği uygulamalarında öznitelik seçiminde yüksek doğruluk ve hızlı işlem süresi sağlamıştır.

Bu çalışmada, öznitelik seçimi için Genetik Algoritma (GA) ve Arı Kolonisi Algoritması (ABC) hibrit bir şekilde uygulanmıştır. Aşağıda, bu iki algoritmanın birleşimiyle oluşturulan çözüm sürecinin adımlarını gösteren sözde kod sunulmaktadır. Algoritmaların her biri, başlangıç çözümü için GA kullanırken, çözümün iyileştirilmesi için ABC algoritmasını kullanmaktadır.

BAŞLA

```
CSV dosyasından "BankLoyaltyEtiketli.csv" veri setini YÜKLE
Veri setini öznitelikler (X) ve etiketler (y) olarak AYIR

Öznitelik Seçim Problemini TANIMLA:
- GİRİŞ: X (öznitelikler), y (etiketler)
- ÇIKIŞ: Bir öznitelik seçimi çözümünün değerlendirilmesi
FONKSİYON değerlendirme(çözüm):
    Çözümün (ikili vektör) seçtiği öznitelikleri SEÇ
    EĞER hiç öznitelik seçilmediyse:
        0 (ceza) DÖNDÜR
    Veriyi eğitim ve test setlerine AYIR
    Eğitim seti üzerinde bir Destek Vektör Makinesi (SVC) modelini EĞİT
    Test setinde tahminler yap
    Tahminlerin doğruluğunu (accuracy) DÖNDÜR
```

Şekil 2.1. Sözde Kod Başlangıç ve Problemin Tanımı

İlk adımda, programın çalışabilmesi için gerekli veri seti yüklenir ve işleme başlamak için temel yapı oluşturulur. Veri seti, "BankLoyaltyEtiketli.csv" dosyasından alınarak X (öznitelikler) ve y (etiketler) olarak iki farklı bileşene ayrılır. Ardından, öznitelik seçimi problemi tanımlanır ve her çözümün doğruluğunu ölçen bir değerlendirme fonksiyonu yazılır. Bu fonksiyon, seçilen öznitelikleri kullanarak bir eğitim ve test seti oluşturur, SVM (Destek Vektör Makinesi) modeli ile eğitimi gerçekleştirir ve test verisi üzerinde tahminler yaparak doğruluğu hesaplar. Bu adım, algoritmanın başarısını ölçebilmek için temel bir yapı kurar ve ilerleyen adımlarda bu yapı kullanılarak çözüm optimize edilir.

```
Genetik Algoritmayı (GA) TANIMLA:
- GİRİŞ: Öznitelik seçim problemi, popülasyon boyutu, nesil sayısı, çaprazlama oranı, mutasyon oranı
- ÇIKIŞ: En iyi öznitelik seçimi çözümü
Popülasyonu rastgele başlat (ikili öznitelik vektörleri)
Her bireyin fitness değerini (değerliliğini) DEĞERLENDİR
Her nesilde:
    Ebeveynleri turnuva seçimiyle SEÇ
    Ebeveynler arasında çaprazlama yaparak yeni yavrular oluştur
    Yavrulara mutasyon uygula (belirtilen oranla)
    Yeni popülasyonun fitness değerlerini DEĞERLENDİR
En iyi çözümü ve fitness değerini DÖNDÜR
```

Şekil 2.2. Sözde Kod – GA Algoritma Tanımı

Genetik Algoritma (GA), öznitelik seçimi problemine yönelik en iyi çözümü bulmak için evrimsel bir yaklaşım kullanır. İlk olarak, popülasyon rastgele başlatılır ve her bireyin (ikili vektör) uygunluk değeri, özniteliklerin doğruluğunu

belirleyerek değerlendirilir. Ardından, her nesilde ebeveynler turnuva seçimiyle belirlenir ve çaprazlama işlemiyle yeni yavrular oluşturulur. Yavrular üzerinde mutasyon işlemi gerçekleştirilir ve yeni popülasyonun uygunluk değerleri tekrar hesaplanır. Bu süreç, her nesilde çözümün daha iyi bir hale gelmesini hedefler. Genetik algoritma, doğal seleksiyon ilkelerini taklit ederek, daha iyi çözüm adaylarını bulmak için evrimsel süreçleri kullanır ve problem çözme kapasitesini artırır. Bu adım, GA'nın nasıl çalıştığını ve her nesilde nasıl daha iyi çözümler üretmeye çalıştığını detaylandırmaktadır.

```
Arı Kolonisi Algoritmasını (ABC) TANIMLA:  
- GİRİŞ: Öznitelik seçim problemi, popülasyon boyutu, maksimum iterasyon sayısı, deneme limiti  
- ÇIKIŞ: En iyi öznitelik seçimi çözümü  
Popülasyonu rastgele başlat (ikili öznitelik vektörleri)  
Her bireyin fitness değerini DEĞERLENDİR  
Her iterasyonda:  
    Her işçi arı için:  
        Bir komşu arı seç ve yeni bir aday çözüm oluştur  
        Aday çözüm daha iyi fitness değerine sahipse:  
            Çözümü güncelle  
            Deneme sayısını sıfırla  
    Her gözcü arı için:  
        EĞER deneme sayısı limit değerini aşarsa:  
            Çözümü rastgele yeniden başlat  
            Fitness değerini yeniden değerlendir  
En iyi çözümü ve fitness değerini DÖNDÜR
```

Şekil 2.3. Sözde Kod – ABC Algoritma Tanımı

ABC (Arı Kolonisi Algoritması), biyolojik bir arı kolonisi davranışından esinlenerek geliştirilmiş bir optimizasyon yöntemidir. Bu adımda, her işçi arı rastgele bir komşu çözüm seçer ve yeni bir aday çözüm oluşturur. Eğer bu yeni çözüm, mevcut çözümden daha iyi bir uygunluk değeri sunuyorsa, çözüm güncellenir ve deneme sayısı sıfırlanır. Gözcü arılar ise, deneme sayısının limit değeri aşılması durumunda, mevcut çözümü yeniden başlatır ve uygunluk değerini yeniden değerlendirir. Bu süreç, çözüm alanındaki farklı bölgelerde arama yaparak, lokal optimumlardan kaçınmayı ve daha iyi bir çözüm bulmayı amaçlar. Arı Kolonisi Algoritması, global optimizasyon yetenekleri ile bilinir ve özellikle karmaşık ve çok boyutlu problemlerde yüksek performans gösterir. Bu adımda, çözümün zamanla nasıl evrildiği ve ABC algoritmasının optimizasyon süreci açıkça gösterilmektedir.

```

Hibrit GA-ABC Algoritmasını TANIMLA:
- GİRİŞ: Öznitelik seçim problemi, GA parametreleri, ABC parametreleri
- ÇIKIŞ: En iyi öznitelik seçimi çözümü
İlk çözümü GA ile elde et
Bu çözümü optimize etmek için ABC'yi kullan
En iyi çözümü ve fitness değerini DÖNDÜR

Ana prosedürü TANIMLA:
- Veri setini yükle ve X ile y'ye ayır
- Problemi X ve y ile başlat
- GA parametrelerini TANIMLA (popülasyon boyutu, nesil sayısı, çaprazlama oranı, mutasyon oranı)
- ABC parametrelerini TANIMLA (popülasyon boyutu, maksimum iterasyon sayısı, deneme limiti)
- Hibrit GA-ABC algoritmasını çalıştır
- En iyi çözümü ve doğruluğu yazdır

Ana prosedürü ÇALIŞTIR

BİTİR

```

Şekil 2.4. Söзде Kod – GA ve ABC Algoritmanın Hibrit Tanımı ve Bitiş

Son adımda, Genetik Algoritma (GA) ve Arı Kolonisi Algoritması (ABC) hibrit bir şekilde birleştirilir. İlk olarak, GA algoritması ile başlangıç çözümü elde edilir, ardından bu çözüm ABC algoritması ile optimize edilir. Bu hibrit yaklaşım, her iki algoritmanın güçlü yönlerini birleştirerek daha etkili bir öznitelik seçimi çözümü üretir. Ana prosedürde, veri seti yüklenir ve X ile y'ye ayrılır. GA ve ABC parametreleri belirlenir, ardından hibrit GA-ABC algoritması çalıştırılır. Son olarak, elde edilen en iyi çözüm ve doğruluk değeri yazdırılarak algoritmaların başarısı ölçülür. Bu süreç, öznitelik seçiminde daha doğru ve verimli bir sonuç elde etmek için birleştirilen iki optimizasyon yönteminin nasıl işlediğini gösterir.

Her iki algoritmanın avantaj ve dezavantajlarını göz önünde bulundurarak iç içe bir algoritma yapısı kurarak veri setini her iki algoritmanın da işlevlerini kullanarak öznitelik seçimi gerçekleştirildi. Öznitelik seçimi data analizinde seçilen niteliklerin kullanılması sonrası müşteri sadakati için performansı ve doğruluk oranı yüksek ve en iyi şekilde sonuçlanmasını sağlamaktır.

2.3 Büyük Veri Setlerinde Hibrit Algoritmalarla Öznitelik Seçimi

Verinin en iyi şekilde kullanılmasını sağlamak amacıyla GA ve ABC hibrit şekilde kullanılarak öznitelik seçimi gerçekleştirildi. Bu bağlamda elde edilen data için en iyi sonucu verecek şekilde dizayn edildi.

Hibrit algoritmaların büyük veri setlerinde öznitelik seçimi üzerindeki etkisi, veri madenciliği alanında önemli bir araştırma konusu olmuştur. GA-ABC hibrit algoritmasının büyük veri setlerinde kullanıldığında, gereksiz özniteliklerin

elenmesi ve yüksek etkili özniteliklerin seçilmesiyle hem doğruluğun hem de işlem hızının arttığını göstermiştir (Liu vd., 2020).

Bu konuda yapılan bir diğer çalışma, GA-ABC hibrit yaklaşımının, finansal veri setleri üzerinde öznitelik seçimi için başarılı sonuçlar verdiğini ve özellikle müşteri sadakati gibi karmaşık problemlerde veri setini optimize ederek daha doğru tahminler yapılmasını sağladığını ifade etmiştir (Li ve Zhang, 2019).

Bu çalışma kapsamında geliştirilen hibrit öznitelik çıkarma yaklaşımı, genetik algoritma ve yapay arı kolonisi algoritmasının birleştirilmesini içerir. Bu yaklaşımda, genetik algoritma ve yapay arı kolonisi algoritması bir araya getirilerek, her bir algoritmanın avantajlarından yararlanılır ve birlikte daha güçlü bir optimizasyon sağlanır. Öznitelik çıkarma sürecinde, genetik algoritma ve yapay arı kolonisi algoritması birlikte kullanılarak, veri setinden en önemli özniteliklerin seçilmesi hedeflenir. Genetik algoritma ve Yapay arı kolonisi algoritmasını hibrit şekilde kullanarak öznitelik seçimi gerçekleştirdik. İşlem sonrası etkisi yüksek olan özniteliklerin seçimi ile müşteri sadakatini ölçümü için etkili özniteliklerden oluşan veri seti hazırlandı.

2.4 Kullanım Ortamı

Apple Macbook Pro bilgisayar kullanıldı. 2022 yılı üretimi tamamlanan M2 işlemcili, 16 GB RAM, 256 GB SSD model bilgisayar ortamında çalışmalar yapıldı. GA ve ABC hibrit yazılmış algoritma ise Visual Studio Code da çalıştırıldı. Python dili kullanılarak hibrit algoritma yazılmış ve kullanılmıştır.

Visual Studio Code (VSCode), Microsoft tarafından geliştirilen, açık kaynaklı ve platformlar arası bir kod editörüdür. Geliştiriciler için oldukça popüler olan bu araç, hafif bir yapıya sahip olmasına rağmen güçlü özellikler sunar. VSCode, birçok programlama dilini destekler ve dil eklentileri sayesinde özelleştirilebilir bir yapıya sahiptir. Ayrıca, IntelliSense adı verilen akıllı tamamlama özelliği ile kod yazımını hızlandırır ve hataları en aza indirir. Geliştiriciler, Visual Studio Code'u Python, Java, C++, JavaScript, HTML, CSS gibi dillerde kullanırken, hata ayıklama (debugging) araçlarıyla da kodlarının performansını analiz edebilirler.

Python, açık kaynaklı bir dil olup, esnek yapısı ve geniş kütüphane desteği ile veri analizi ve makine öğrenimi projelerinde yaygın olarak tercih edilmektedir.

Çalışmada kullanılan GA(Genetik Algoritma) ve ABC(Arı Kolonisi Algoritması) gibi evrimsel algoritmaların yanı sıra, modelin doğruluğu ve performansını artırmak için NumPy, Pandas, Matplotlib ve SciPy gibi kütüphaneler etkin bir şekilde kullanılmıştır. Veri ön işleme, modelleme ve değerlendirme adımlarında ise, scikit-learn kütüphanesi önemli bir rol oynamıştır. Özellikle, train_test_split fonksiyonu ile verilerin eğitim ve test setlerine ayrılması sağlanmış, SVM (Destek Vektör Makineleri) algoritması ile sınıflandırma yapılmış ve doğruluk metrikleri ile modelin doğruluğu değerlendirilmiştir. Pandas kütüphanesi, veri manipülasyonu ve ön işleme işlemlerinde; NumPy ise sayısal hesaplamalar ve vektörleşmiş işlemler için kullanılmıştır. Ayrıca, Matplotlib kütüphanesi ile verilerin görselleştirilmesi ve sonuçların görsel olarak analiz edilmesi mümkün olmuştur.

2.5 Avantajlar ve Uygulamalar

Genetik algoritma ve yapay arı kolonisi algoritmasının hibrit bir şekilde kullanımı, büyük veri setlerinden öznitelik çıkarma süreçlerinde dikkat çekici bir avantaj sağlamaktadır. Genetik algoritma, doğal seleksiyon ve genetik mutasyon süreçlerinden esinlenerek çözüm uzayını etkili bir şekilde keşfederken, yapay arı kolonisi algoritması, arıların besin kaynaklarını bulma davranışını taklit ederek optimize edilmiş çözümler üretir. Bu iki algoritmanın hibrit bir şekilde kullanılması, her iki yöntemin güçlü yönlerini birleştirerek veri setlerinin daha verimli analiz edilmesini ve en uygun özniteliklerin seçilmesini mümkün kılar.

Özellikle karmaşık ve büyük boyutlu veri setlerinde, geleneksel öznitelik çıkarma yöntemleri zaman ve hesaplama maliyeti açısından yetersiz kalabilir. Bu tür durumlarda, genetik algoritma ve yapay arı kolonisi algoritması hibrit kullanımı hem zaman verimliliği hem de doğruluk oranı açısından üstün bir performans sergiler. Genetik algoritma, geniş bir çözüm uzayını sistematik bir şekilde tarayarak en uygun öznitelikleri belirlerken, yapay arı kolonisi algoritması bu süreci daha da optimize ederek yerel optimumlara takılma riskini azaltır. Bu kombinasyon, veri setindeki gereksiz veya düşük etkili özniteliklerin elenmesini ve yalnızca hedef modele en yüksek katkıyı sağlayan özniteliklerin seçilmesini sağlar.

Bu hibrit yaklaşım, yalnızca öznitelik seçimi süreçlerini hızlandırmakla kalmaz, aynı zamanda makine öğrenmesi modellerinin genel performansını da artırır. Seçilen özniteliklerin etkisi yüksek olduğunda, makine öğrenmesi modelleri

daha yüksek doğruluk oranlarına ulaşır ve genel genelleme kabiliyeti artar. Ayrıca, bu yöntem, gerçek zamanlı sistemlerde de etkili bir şekilde kullanılabilir, çünkü hızlı ve optimize edilmiş öznitelik çıkarma süreci, veri setlerinin analizini ve modelleme süreçlerini büyük ölçüde hızlandırır.

Sonuç olarak, genetik algoritma ve yapay arı kolonisi algoritmasının hibrit kullanımı, makine öğrenmesi ve veri madenciliği uygulamalarında, özellikle büyük veri setlerinde, etkili bir çözüm sunmaktadır. Bu iki algoritmanın birleşimi, öznitelik çıkarma sürecini daha verimli hale getirirken, aynı zamanda makine öğrenmesi modellerinin başarısını da önemli ölçüde artırır. Literatürde, bu tür hibrit yaklaşımların farklı problem türlerinde üstün performans sergilediği ve model doğruluğunu artırdığına dair birçok çalışma bulunmaktadır (Holland, 1975; Karaboga, 2005).

2.6 Bu Çalışma Kapsamında Optimize Edilmiş Alt Veri Setleri

Hibrit algoritma kullanılması sonrası seçilen özniteliklerden yeni veri setleri oluşturulmaktadır. Şekil 3.1. de ki değerler ise meta sezgisel algoritmaların birlikte kullanılması sonrası seçilen öznitelikleri ifade etmektedir.

GA algoritmasının hiperparametreleri: Popülasyon boyutu (20), Nesil sayısı (50), Çaprazlama oranı (0.8), Mutasyon oranı (0.1). ABC algoritmasının hiperparametreleri: Popülasyon boyutu (10), Maksimum iterasyon sayısı (50), Sabırlı arıların sınırlı değeri (5).

Seçilen özniteliklere göre 6 farklı veri seti ve her bir veri setinde sırası ile 5,6,8,9,10,12 seçilmiş sütun mevcut. Bu alt veri setleri ile ayrı ayrı model yapıları ile incelemeler yapıldı. Bu incelemeler sonrası veri setlerinin performans ve doğruluk değerleri ölçüldü. Seçilen alt veri setleri üzerinde yapılan analizlerde, öznitelik sayısındaki artışın, özellikle model doğruluğu üzerinde olumlu bir etkisi olduğu gözlemlenmiştir. Bununla birlikte, daha fazla özneliğin dahil edilmesi, eğitim ve değerlendirme sürelerini uzatmıştır. Bu durum, performans doğruluğu ile işlem süresi arasında bir denge sağlanmasının gerekliliğini ortaya koymaktadır. Ayrıca, her bir alt veri seti için test edilen modeller, belirli öznitelik kombinasyonlarında farklı başarı oranları göstermiştir, bu da öznitelik seçiminin modellerin genelleme yetenekleri üzerindeki etkisini vurgulamaktadır. Elde edilen sonuçlar, doğru öznitelik seçiminin hem model performansını artırmak hem de

işlem sürelerini optimize etmek açısından kritik öneme sahip olduğunu göstermektedir.

Example	Example 1	Example 2	Example 3	Example 4	Example 5	Example 6
CLIENTNUM	0	0	0	0	0	0
Attrition_Flag	0	0	0	1	0	1
Dependent_count	0	1	1	1	1	1
Education_Level	1	0	0	0	0	0
Income_Category	0	0	1	0	0	1
Card_Category	0	0	1	1	1	1
Months_on_book	0	0	0	0	0	0
Total_Relationship_Count	1	0	0	1	1	1
Months_Inactive_12_mon	1	0	0	0	1	1
Contacts_Count_12_mon	0	1	0	1	0	0
Credit_Limit	0	1	1	1	1	0
Total_Revolving_Bal	0	0	1	1	1	1
Avg_Open_To_Buy	0	0	0	0	1	1
Total_Amt_Chng_Q4_Q1	0	0	0	1	0	1
Total_Trans_Amt	1	1	0	1	0	1
Total_Trans_Ct	0	1	1	0	1	1
Total_Ct_Chng_Q4_Q1	0	0	0	0	0	0
Avg_Utilization_Ratio	0	0	1	0	1	1
Naive_Bayes_Classifier_Attrition_Flag	0	0	0	0	0	0
Naive_Bayes_Classifier_Attrition_Flag	0	0	0	0	1	0
Cluster	1	1	1	1	1	1

Şekil 2.5. Seçildi (1) / Seçilmedi (0)

2.7 Analizde Kullanılan Algoritmalar

Bu çalışmada, öz nitelik seçim sürecinden sonra elde edilen alt veri setlerinin doğruluk ve performans ölçümlerini değerlendirmek için sekiz farklı makine öğrenme algoritması kullanılmıştır. Her bir algoritma, alt veri setlerinin karakteristik özelliklerine uygun olarak farklı avantajlar ve yaklaşımlar sunmaktadır. Kullanılan modeller arasında, hem hızlı sonuçlar üretebilen istatistiksel yöntemler (Naive Bayes, Lojistik Regresyon, Genelleştirilmiş Doğrusal Model) hem de daha karmaşık, çok katmanlı öğrenme gerektiren derin öğrenme yöntemleri (Derin Öğrenme) yer almaktadır. Ayrıca, karar ağaçları ve rastgele ormanlar gibi karar verme süreçlerini dallara ayırarak analiz eden algoritmalar da doğruluk ve genelleme performanslarını artırmak amacıyla kullanılmıştır.

Bu modeller, alt veri setlerinin farklı boyut ve karmaşıklık düzeylerinde test edilmiştir. Her bir modelin performansı; doğruluk, duyarlılık, hatırlama ve F1 skoru gibi temel sınıflandırma metrikleri ile ölçülmüştür. Aynı zamanda, modellerin aşırı öğrenme ve genelleme yeteneği de analiz edilmiştir. Modellerin doğruluk sonuçları, kullanılan algoritmanın alt veri setinin özelliklerine ne kadar iyi uyum sağladığını

gösterirken, her bir algoritmanın hangi durumlarda daha iyi performans sergilediğini karşılaştırmak için ayrıntılı bir analiz yapılmıştır.

Çalışmada elde edilen sonuçlar, basit ve kompleks modellerin farklı alt veri setlerinde nasıl performans gösterdiğini kapsamlı bir şekilde ortaya koymuştur. Basit modeller (örneğin Naive Bayes ve Lojistik Regresyon), düşük boyutlu ve lineer olarak ayrılabilir alt veri setlerinde hızlı ve yeterli sonuçlar verirken, daha gelişmiş yöntemler (örneğin Derin Öğrenme ve Aşamalı Güçlendirilmiş Ağaçlar) büyük ve karmaşık alt veri setlerinde daha yüksek doğruluk oranları elde etmiştir. Bu durum, model seçiminde alt veri setinin özelliklerine göre hareket edilmesinin önemini vurgulamaktadır. Ayrıca, ileri düzey modellerin genelleme yetenekleri, doğru öznitelik seçimi ile birleştirildiğinde, performansın önemli ölçüde artırılabilceği görülmüştür.

Tablo 2.1. Sınıflandırma Algoritmaları

Sınıflandırma Algoritması	
Algoritma- 1	Naive Bayes
Algoritma- 2	Genelleştirilmiş Doğrusal Model(GLM)
Algoritma- 3	Lojistik Regresyon
Algoritma- 4	Fast Large Margin
Algoritma- 5	Derin Öğrenme
Algoritma- 6	Karar Ağacı
Algoritma- 7	Rastgele Orman
Algoritma- 8	Gradyan Artırmalı Ağaçlar

Yapılan bu karşılaştırma, her bir modelin farklı veri yapıları ve öznitelik dağılımları karşısındaki güçlü ve zayıf yönlerini detaylı bir şekilde değerlendirme imkânı sunmuştur. Elde edilen bulgular, model seçimi sürecinde uygulamaya ve alt veri setine özgü faktörlerin dikkate alınmasının, sonuçların doğruluğu ve tutarlılığı açısından kritik öneme sahip olduğunu göstermiştir. Kullanılan modellerin detaylı performans değerlendirmesi ve karşılaştırmaları Tablo 2.1.'de sunulmaktadır. Bu analiz, gelecekte benzer çalışmalarda model seçim süreçlerini yönlendirecek önemli ipuçları sağlamaktadır.

2.8 Analizde Kullanılan Algoritmaların HiperParametre Bilgileri

2.8.1 Naive Bayes

Naive Bayes algoritması, olasılık temelli bir sınıflandırma yöntemidir ve özellikle metin sınıflandırma gibi kategorik veri problemlerinde sıklıkla kullanılır. Bu algoritmanın performansı üzerinde etkili olan ana hiperparametrelerden biri Laplace Smoothing'dir. Bu parametre, sıfır olasılığı önlemek ve sınıflandırmayı daha kararlı hale getirmek için kullanılır. Ayrıca, sürekli değişkenler için Gaussian gibi kernel türü ve minimum standart sapma parametresi algoritmanın doğruluk oranını etkileyen diğer faktörlerdir.

2.8.2 Genelleştirilmiş Doğrusal Model (GLM)

Genelleştirilmiş Doğrusal Model (GLM), hem sınıflandırma hem de regresyon problemleri için esnek bir modeldir. GLM'in en önemli hiperparametreleri arasında link fonksiyonu ve dağılım ailesi (family) yer alır. Link fonksiyonu, bağımlı değişken ile bağımsız değişkenler arasındaki ilişkiyi tanımlar; örneğin, logit fonksiyonu lojistik regresyonda yaygın olarak kullanılır. Ayrıca, düzenlileştirme parametresi (Lambda), modelin aşırı öğrenmesini önlemek için kullanılır.

2.8.3 Lojistik Regresyon

Lojistik Regresyon, iki sınıflı veya çok sınıflı problemler için yaygın bir şekilde kullanılan bir sınıflandırma algoritmasıdır. Bu algoritmanın ana hiperparametrelerinden biri, düzenleme türü (L1 veya L2 düzenlileştirme) ve düzenleme gücü (C)'dir. Bu parametreler, modelin genel performansını kontrol ederek aşırı öğrenme riskini azaltır. Ayrıca, optimizasyon için kullanılan çözücü tür (örneğin, liblinear) parametresi, büyük veri setleri ile çalışırken önem kazanır.

2.8.4 Fast Large Margin

Fast Large Margin algoritması, sınıflar arasındaki ayrımı maksimize eden bir sınıflandırma yöntemidir. Bu algoritmanın önemli hiperparametreleri arasında çekirdek türü (lineer veya polynomial gibi) ve marj genişliği bulunur. Ayrıca, düzenlileştirme parametresi, algoritmanın sınıflar arasındaki sınırı belirlerken daha iyi sonuçlar elde etmesini sağlar.

2.8.5 Derin Öğrenme

Derin Öğrenme algoritması, özellikle karmaşık veri setleri için güçlü bir sınıflandırma algoritmasıdır. Bu algoritmanın performansını etkileyen hiperparametreler arasında katman sayısı, her bir katmandaki nöron sayısı, ve öğrenme oranı (learning rate) bulunur. Ayrıca, kullanılan aktivasyon fonksiyonları (ReLU, sigmoid veya tanh gibi) ve toplam işlem süresi boyunca belirlenen epochs sayısı, modelin doğruluk oranını ve genel başarısını etkiler.

2.8.6 Karar Ağacı

Karar Ağacı algoritması, sınıflandırmada kullanılan en temel algoritmalarından biridir ve ağaç yapısı üzerinden karar verir. Maksimum derinlik (maximum depth), yaprak başına minimum örnek sayısı ve bölme kriteri (split criterion) gibi hiperparametreler, modelin performansını doğrudan etkiler. Bölme kriteri olarak bilgi kazancı veya Gini indeksi kullanılır ve bu parametreler, algoritmanın sınıfları nasıl ayıracağını belirler.

2.8.7 Rastgele Orman

Rastgele Orman algoritması, birden fazla karar ağacının bir araya getirilmesiyle oluşan güçlü bir sınıflandırma yöntemidir. Ormandaki ağaç sayısını belirleyen ağaç sayısı(number of trees) ve her ağaç için seçilecek maksimum özellik sayısını tanımlayan maksimum özellik(maximum features) gibi hiperparametreler bulunmaktadır. Ayrıca, bootstrap örnekleme ve maksimum derinlik gibi parametreler, modelin performansını artırmak için optimize edilebilir.

2.8.8 Gradyan Artırmalı Ağaçlar

Gradyan Artırmalı Ağaçlar algoritması, ağaç tabanlı modellerin ardışık olarak oluşturulduğu güçlü bir sınıflandırma yöntemidir. Bu algoritmanın hiperparametreleri arasında öğrenme oranı (learning rate), maksimum derinlik ve ağaç sayısı yer alır. Ayrıca, eğitim veri setinden alınacak alt kümenin oranını belirleyen alt örnek oranı (subsampling ratio), modelin daha dengeli ve hızlı bir şekilde eğitilmesine olanak tanır.

Bu açıklamalar, her algoritmanın temel hiper-parametrelerini ve bunların analiz performansına olan etkilerini vurgulamaktadır.

3. BULGULAR

Bu çalışma kapsamında gerekli alt veri setleri GA ve ABC'nin hibrit kullanılması öznitelik seçimi sonrası elde edilen alt veri setlerin sınıflandırma algoritmalarından geçerek performans ve doğruluk oranları ölçüldü. Bu sonuçlar doğrultusunda işleme alınan verilerin öznitelik seçimine göre analiz sonuçları farklılık gösteriyor. Genetik algoritma ve yapay arı kolonisi algoritmasının hibrit bir şekilde kullanılarak gerçekleştirilen öznitelik seçimi süreçleri sonucunda elde edilen veriler, model performansı, doğruluk oranları ve işlem süresi açısından kapsamlı bir şekilde analiz edilmiştir. Elde edilen bulgular, önerilen yöntemlerin etkili bir şekilde alt veri seti boyutunu azaltırken, makine öğrenmesi modellerinin genel başarı oranlarını artırdığını göstermektedir.

Öznitelik seçimi sonrası oluşan her alt veri seti, sınıflandırma algoritmaları kullanılarak performans ve doğruluk değerleri ölçüldü. Seçilen öznitelikler sonrası oluşan alt veri setleri haricinde 23 özniteliği bulunan datanın tüm öznitelikleri ile algoritmalarından performans ve doğruluk ölçümleri gerçekleştirildi. Bunun sonunca 23 özniteliğe sahip veri seti üzerinde yapılan analizlerde, modellerin doğruluk performansı ölçülmüş ancak aşırı öğrenme değerinin yüksek olduğu gözlemlenmiştir. Aşırı Öğrenme, modelin eğitim verilerine aşırı uyum sağlaması sonucunda, test verileri üzerinde düşük performans sergilemesine neden olmaktadır. Bu durum, modelin genelleme yeteneğini kaybettiği ve sadece mevcut veri setine uygun çalıştığı anlamına gelir. Çalışmada, bu öznitelik sayısı ile elde edilen modellerin aşırı öğrenmeye yatkın olduğu tespit edilmiş ve bu nedenle 23 öznitelik içeren veri seti, tez dokümanına dahil edilmemiştir. Bunun yerine, daha az sayıda öznitelikle yapılan analizlere odaklanılmış ve daha dengeli bir performans elde edilmiştir. Bu yaklaşım, modelin gerçek dünyadaki veri üzerinde daha tutarlı ve etkili sonuçlar üretmesi için kritik bir adım olarak değerlendirilmiştir.

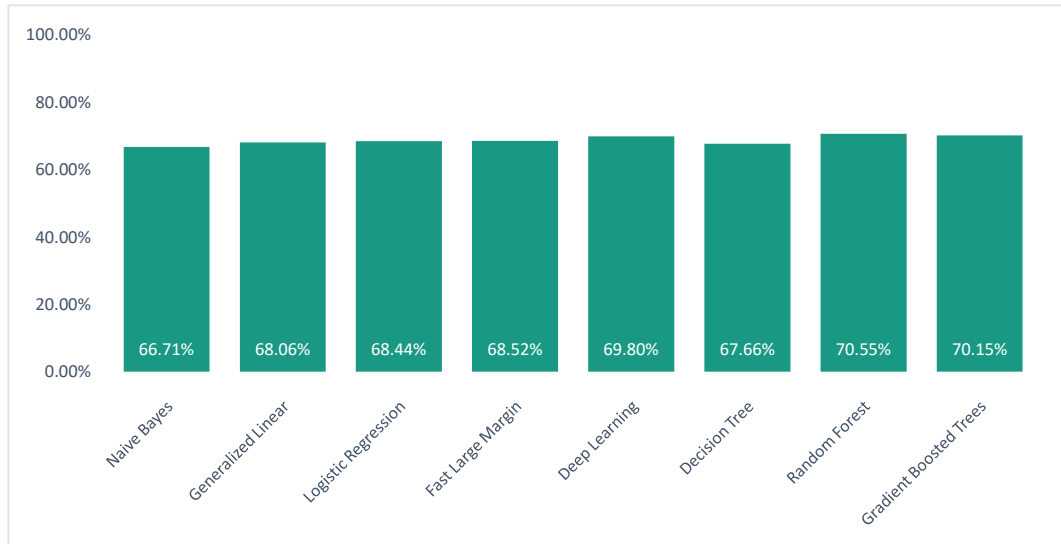
3.1 5 Öznitelikli Veri Bulguları

Tablo 3.1. Öznitelik 6 alt veri seti üzerinde kullanılan çeşitli makine öğrenimi algoritmalarının Tablo 3.1. de gösterildiği gibi doğruluk, toplam süre, eğitim süresi ve puanlama sürelerine göre performanslarını göstermektedir. Ayrıca Şekil 3.1 Doğruluk grafiği, Şekil 3.2 ve 3.3 Performans grafiği gösterilmektedir.

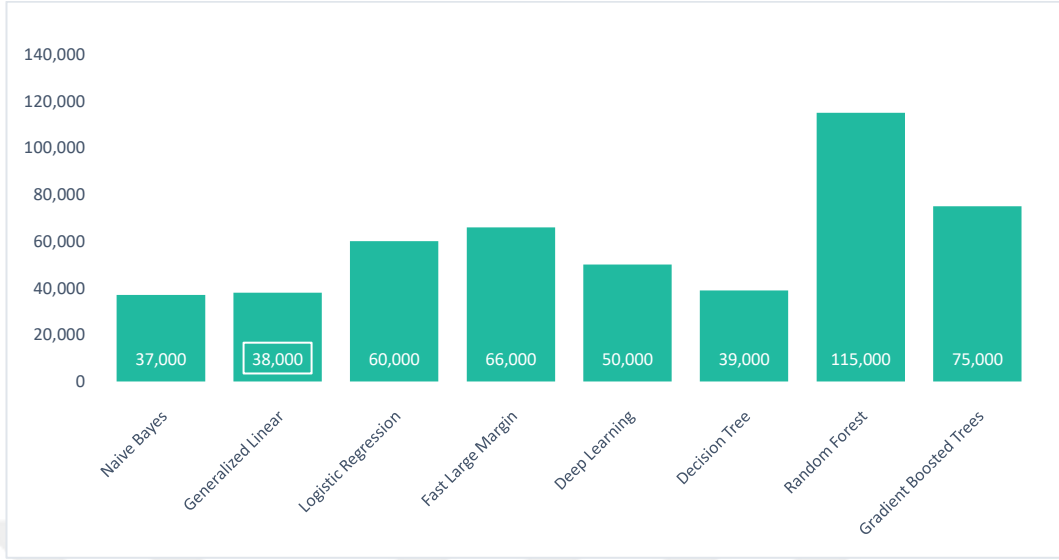
Tabloya bakıldığında, algoritmaların doğruluk oranları ve zaman performansları arasında önemli farklılıklar bulunmaktadır.

Tablo 3.1. Öznitelik 5 Alt Veri Seti Performans-Doğruluk Tablosu

MODEL	Doğruluk (%)	Toplam Süre(s)	Toplam Süre(ms)	Eğitim Süresi(ms)	Puanlama Süresi(ms)
Naive Bayes	66,71%	37	37.000	4	91
Generalized Linear	68,06%	38	38.000	126	86
Lojistik Regresyon	68,44%	60	60.000	13	151
Fast Large Margin	68,52%	66	66.000	60	155
Derin Öğrenme	69,80%	50	50.000	265	112
Karar Ağacı	67,66%	39	39.000	6	95
Rastgele Orman	70,55%	115	115.000	16	220
Gradyan Artırmalı Ağaçlar	70,15%	75	75.000	20	147
Ortalama	68,74%	60	60.000	64	132

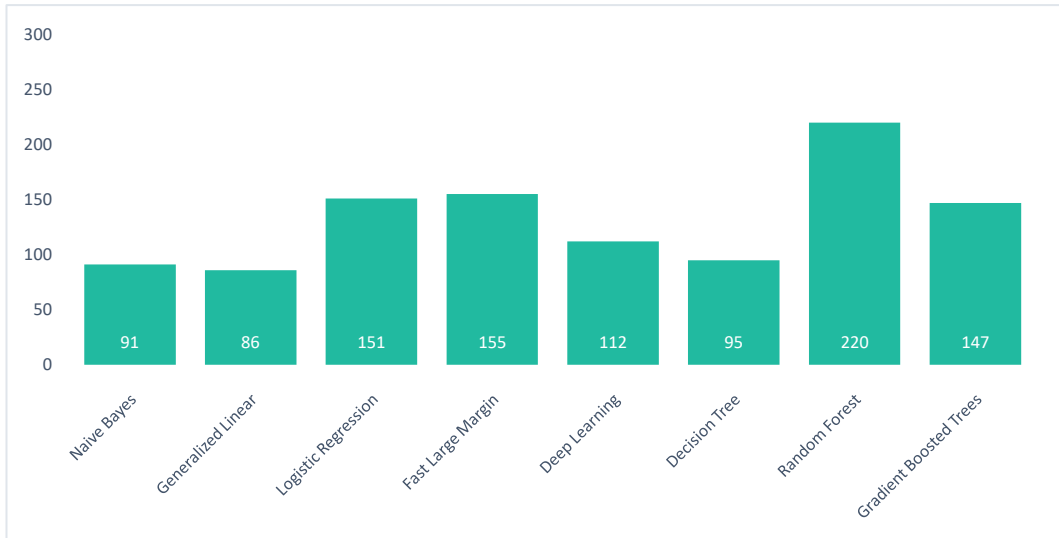


Şekil 3.1. Öznitelik 5 Doğruluk Grafiği



Şekil 3.2. Öznitelik 5 Toplam Süre Grafiği

Bu çalışmada, farklı makine öğrenmesi modellerinin doğruluk oranları ve işlem süreleri incelenmiştir. Tablo, her bir modelin başarı oranını, toplam işlem süresini, eğitim süresini ve tahmin yapma (skoring) süresini göstermektedir. En yüksek doğruluk oranını %70,55 ile Rastgele Orman modeli elde etmiştir, bunu %70,15 doğrulukla Gradyan Artırmalı Ağaçlar ve %69,80 doğrulukla Derin Öğrenme modeli takip etmektedir. Öte yandan, Naive Bayes modeli %66,71 doğrulukla en düşük performansı sergilemiştir.



Şekil 3.3. Öznitelik 5 Puanlama Süresi Grafiği

Modellerin işlem süreleri de dikkate alındığında, en hızlı modeller arasında Naive Bayes ve Karar Ağacı yer alırken, en uzun işlem süresini 115 saniye ile Rastgele Orman modeli almıştır. Modellerin eğitim süreleri ise genellikle 10 ile 265 ms arasında değişiklik göstermektedir; burada en hızlı eğitim süresi Naive Bayes ve Karar Ağacı modelleriyle sağlanırken, Derin Öğrenme modeli en uzun eğitim süresini göstermektedir. Sonuçlar, doğruluk oranı ile işlem süresi arasında bir denge kurmanın önemini vurgulamaktadır. Ayrıca, ortalama doğruluk oranı %68,74 olarak hesaplanmış olup, modellerin genel performansı üzerinde bir genel değerlendirme sunmaktadır.

Attribu...	Cluster...	Cluster...	Contac...	Credit_...	Depen...	Total_T...	Total_T...
Cluster ...	1	-0.652	-0.135	-0.615	-0.068	-0.040	0.156
Cluster ...	-0.652	1	-0.029	0.781	0.066	0.211	0.162
Contact...	-0.135	-0.029	1	0.021	-0.041	-0.113	-0.152
Credit_L...	-0.615	0.781	0.021	1	0.068	0.172	0.076
Depend...	-0.068	0.066	-0.041	0.068	1	0.025	0.050
Total_T...	-0.040	0.211	-0.113	0.172	0.025	1	0.807
Total_T...	0.156	0.162	-0.152	0.076	0.050	0.807	1

Şekil 3.4. Öznitelik 5 Korelasyon Sonuçları

Şekil 3.4. de gösterilen Korelasyon değerleri, hangi özniteliklerin birbirleriyle güçlü bir ilişkisi olduğunu analiz etmede faydalıdır. Güçlü pozitif veya negatif korelasyonlar, öznitelikler arasındaki bağlantıları anlamaya ve model performansını artırmak için kullanılabilecek öznitelik seçim stratejileri geliştirmeye yardımcı olur. Özellikle yüksek korelasyona sahip öznitelikler (örneğin, Toplam işlemler 1 ve Toplam İşlemler 2) modelin öğrenme sürecinde önemli bir etkiye sahip olabilir. Bununla birlikte, bağımsız (korelasyonu düşük) öznitelikler, modele çeşitlilik sağlayabilir.

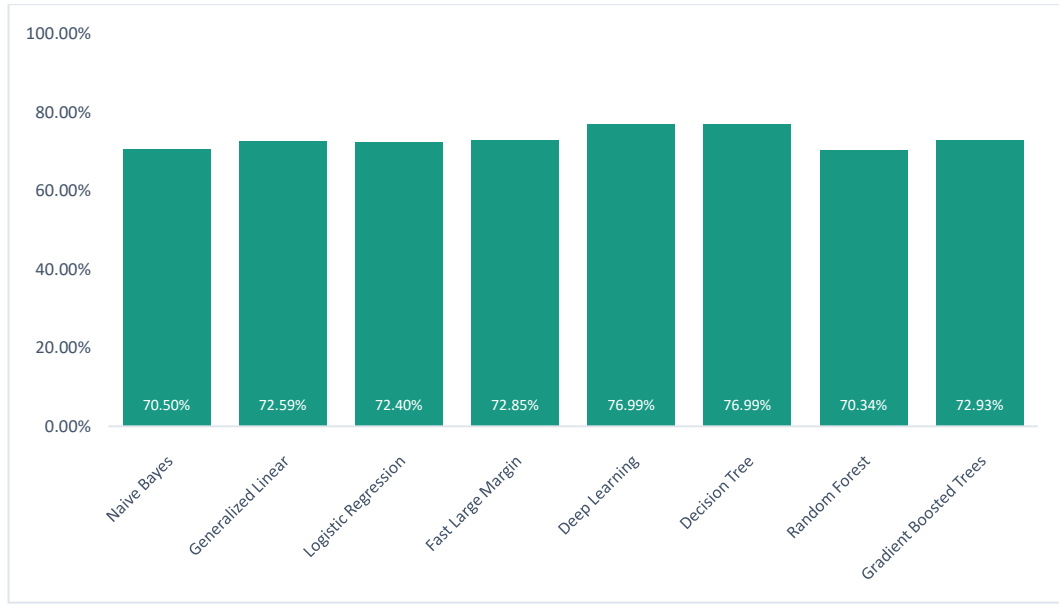
3.2 6 Öznitelikli Veri Bulguları

Tablo 3.2. Öznitelik 6 alt veri seti üzerinde kullanılan çeşitli makine öğrenimi algoritmalarının Tablo 3.2. de gösterildiği gibi doğruluk, standart sapma, toplam süre, eğitim süresi ve puanlama sürelerine göre performanslarını göstermektedir. Ayrıca Şekil 3.5. Doğruluk grafiğini Şekil 3.6. ve Şekil 3.7. Performans grafiği gösterilmektedir.

Tablo 3.2. Öznitelik 6 Alt Veri Seti Performans-Doğruluk Tablosu

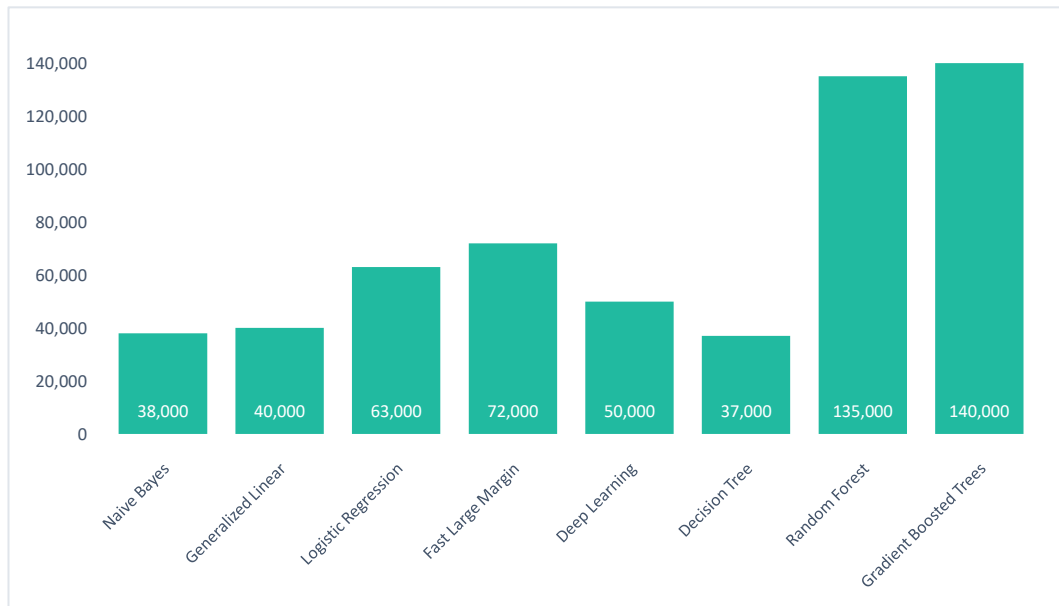
MODEL	Toplam Süre(%)	Toplam Süre(s)	Toplam Süre(ms)	Eğitim Süresi(ms)	Puanlama Süresi(ms)
Naive Bayes	70,50%	38	38.000	11	106
Generalized Linear	72,59%	40	40.000	132	87
Lojistik Regresyon	72,40%	63	63.000	19	153
Fast Large Margin	72,85%	72	72.000	61	150
Derin Öğrenme	76,99%	50	50.000	276	134
Karar Ağacı	76,99%	37	37.000	12	89
Rastgele Orman	70,34%	135	135.000	16	263
Gradyan Artırmalı Ağaçlar	72,93%	140	140.000	21	194
Ortalama	73,20%	72	71.875	69	147

Tabloya bakıldığında, algoritmaların doğruluk oranları ve zaman performansları arasında önemli farklılıklar bulunmaktadır.



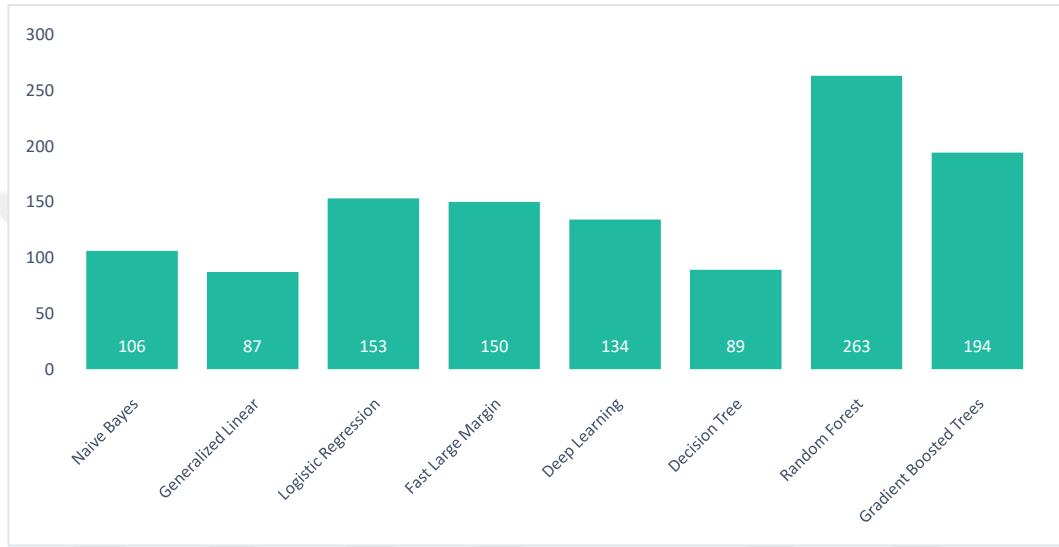
Şekil 3.5. Öznitelik 6 Doğruluk Grafiği

Bu çalışmada, farklı makine öğrenmesi modellerinin doğruluk oranları ve işlem süreleri bir kez daha analiz edilmiştir. Tabloya göre, en yüksek doğruluk oranını %76,99 ile Derin Öğrenme ve Karar Ağacı modelleri elde etmiştir. Bu modelleri sırasıyla %72,93 doğrulukla Gradyan Artırmalı Ağaçlar, %72,85 doğrulukla Fast Large Margin ve %72,59 doğrulukla Genelleştirilmiş Doğrusal Model izlemektedir.



Şekil 3.6. Öznitelik 6 Toplam Süre Grafiği

En düşük doğruluk oranı ise %70,34 ile Rastgele Orman modeline aittir. Modellerin işlem süreleri incelendiğinde, en hızlı model 37 saniye ile Karar Ağacı olarak kaydedilmiştir, ardından Naive Bayes modeli 38 saniye ile hızlı bir performans sergilemiştir. Buna karşılık, en uzun işlem süresine 140 saniye ile Gradyan Artırmalı Ağaçlar sahip olmuştur. Eğitim süreleri açısından bakıldığında, en hızlı eğitim süresi 11 ms ile Naive Bayes modelinde görülürken, en uzun eğitim süresi 276 ms ile Derin Öğrenme modeline aittir.



Şekil 3.7. Öznitelik 6 Puanlama Süresi Grafiği

Puanlama süreleri de modele göre değişiklik göstermekte olup, Rastgele Orman en uzun puanlama süresini (263 ms.) alırken, Karar Ağacı en hızlı puanlama süresini (89 ms.) sağlamaktadır. Ortalama doğruluk oranı %73,20 olarak hesaplanmış olup, modellerin performansı ve işlem süreleri arasındaki dengeyi dikkate alarak, doğruluk ve verimliliğin optimize edilmesi gerektiği sonucuna varılmaktadır.

Şekil 3.8. Öznitelik 6 alt veri setindeki değişkenler arasındaki korelasyon değerlerini göstermektedir. Korelasyon, iki değişken arasındaki doğrusal ilişkinin gücünü ve yönünü ifade eder. Pozitif korelasyon değerleri iki değişkenin aynı yönde hareket ettiğini, negatif değerler ise ters yönde hareket ettiğini gösterir. Korelasyon katsayısının büyüklüğü, ilişkinin gücünü ifade eder; 0'a yakın değerler zayıf ilişkiyi, ± 1 'e yakın değerler ise güçlü ilişkiyi gösterir.

Attribu...	Cluster...	Cluster...	Educati...	Educati...	Educati...	Educati...	Educati...	Educati...	Month...	Total_R...	Total_T...
Cluster ...	1	-0.652	0.004	0.013	0.016	-0.016	-0.011	-0.005	-0.086	0.160	-0.040
Cluster ...	-0.652	1	0.002	-0.007	-0.008	0.009	0.015	-0.003	-0.040	-0.054	0.211
Educati...	0.004	0.002	1	-0.223	-0.166	-0.077	-0.138	-0.140	0.004	-0.014	-0.015
Educati...	0.013	-0.007	-0.223	1	-0.333	-0.155	-0.277	-0.281	0.006	0.005	0.003
Educati...	0.016	-0.008	-0.166	-0.333	1	-0.115	-0.207	-0.209	-0.006	-0.002	-0.000
Educati...	-0.016	0.009	-0.077	-0.155	-0.115	1	-0.096	-0.097	-0.006	0.012	0.008
Educati...	-0.011	0.015	-0.138	-0.277	-0.207	-0.096	1	-0.174	0.010	0.008	0.010
Educati...	-0.005	-0.003	-0.140	-0.281	-0.209	-0.097	-0.174	1	-0.012	-0.004	0.002
Months_...	-0.086	-0.040	0.004	0.006	-0.006	-0.006	0.010	-0.012	1	-0.004	-0.037
Total_R...	0.160	-0.054	-0.014	0.005	-0.002	0.012	0.008	-0.004	-0.004	1	-0.347
Total_T...	-0.040	0.211	-0.015	0.003	-0.000	0.008	0.010	0.002	-0.037	-0.347	1

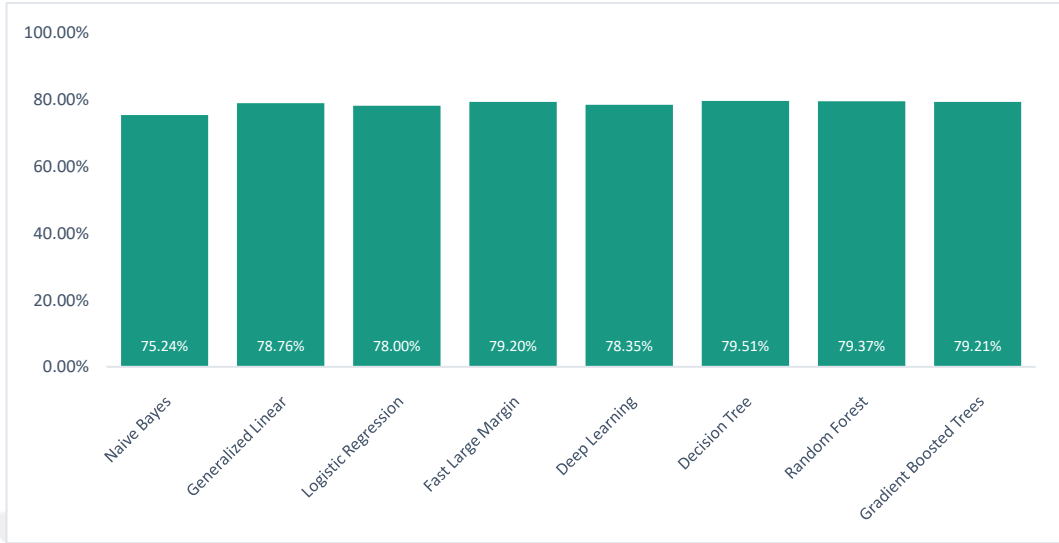
Şekil 3.8. Öznitelik 6 Korelasyon Sonuçları

3.3 8 Öznitelikli Veri Bulguları

Tablo 3.3. Öznitelik 8 alt veri seti ile yapılan modelleme sonuçlarını özetlemektedir. Ayrıca Tablo 4.3.1. farklı makine öğrenme modellerinin doğruluk oranları, standart sapmaları, toplam çalışma süreleri, eğitim süreleri ve test (puanlama) sürelerini içermektedir. Ayrıca Şekil 3.9. Doğruluk grafiğini ve Şekil 3.10 ve Şekil 3.11. Performans grafiğini göstermektedir.

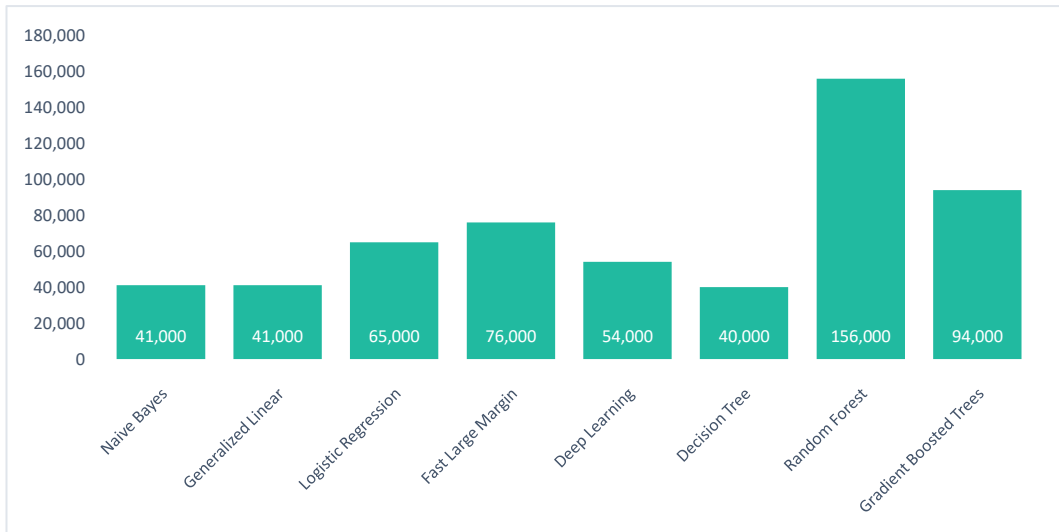
Tablo 3.3. Öznitelik 8 Alt Veri Seti Performans-Doğruluk Tablosu

MODEL	Toplam Süre(%)	Toplam Süre(s)	Toplam Süre(ms)	Eğitim Süresi(ms)	Puanlama Süresi(ms)
Naive Bayes	75,24%	41	41.000	5	94
Generalized Linear	78,76%	41	41.000	142	109
Lojistik Regresyon	78,00%	65	65.000	15	160
Fast Large Margin	79,20%	76	76.000	70	178
Derin Öğrenme	78,35%	54	54.000	270	124
Karar Ağacı	79,51%	40	40.000	6	100
Rastgele Orman	79,37%	156	156.000	20	187
Gradyan Artırmalı Ağaçlar	79,21%	94	94.000	27	198
Ortalama	78,46%	71	70.875	69	144



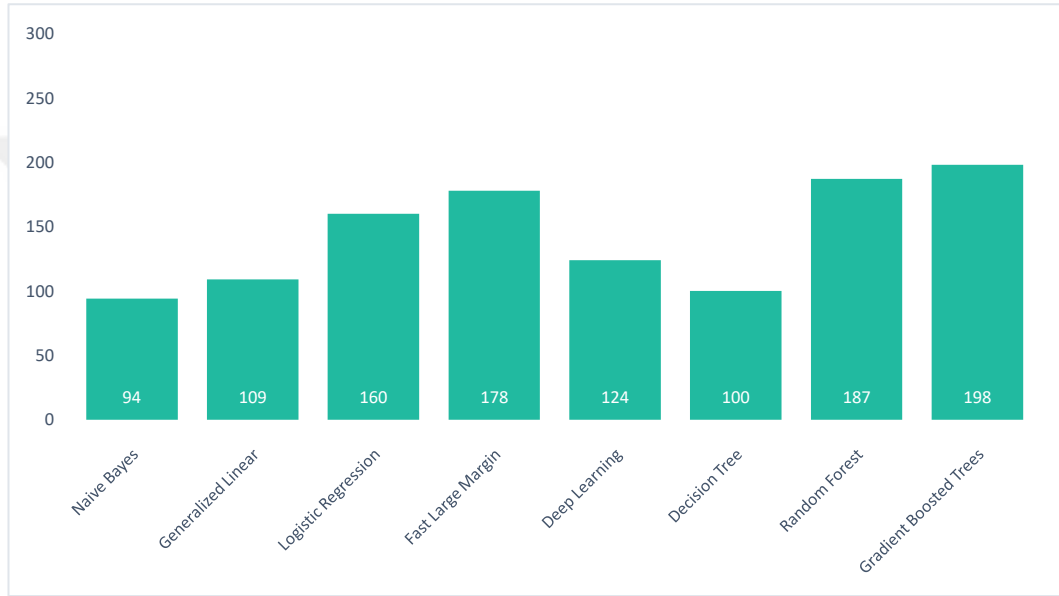
Şekil 3.9. Öznitelik 8 Doğruluk Grafiği

Bu çalışma, farklı makine öğrenmesi modellerinin doğruluk oranları ve işlem sürelerini bir kez daha incelemiştir. Tabloya göre, en yüksek doğruluk oranını %79,51 ile Karar Ağacı modeli elde etmiştir, bunu sırasıyla %79,37 doğrulukla Rastgele Orman, %79,21 doğrulukla Gradyan Artırmalı Ağaçlar ve %79,20 doğrulukla Fast Large Margin modelleri takip etmektedir. En düşük doğruluk oranına sahip model ise %75,24 doğrulukla Naive Bayes'tir.



Şekil 3.10. Öznitelik 8 Toplam Süre Grafiği

Modellerin işlem sürelerine bakıldığında, en hızlı işlem süresi 40 saniye ile Karar Ağacı ve Genelleştirilmiş Doğrusal Modellerine aittir. Buna karşılık, en uzun işlem süresi 156 saniye ile Rastgele Orman modelinde görülmektedir. Eğitim süreleri açısından en hızlı model, 5 ms ile Naive Bayes iken, en uzun eğitim süresi 270 ms ile Derin Öğrenme modeline aittir. Puanlama süreleri, modeller arasında farklılık göstermekte olup, en hızlı puanlama süresi 94 ms ile Naive Bayes modeline, en uzun puanlama süresi ise 198 ms ile Gradyan Artırmalı Ağaçlar modeline aittir.



Şekil 3.11. Öznitelik 8 Puanlama Süresi Grafiği

Ortalama doğruluk oranı %78,46 olarak hesaplanmış olup, bu, kullanılan modellerin genel olarak yüksek doğruluk sağladığını göstermektedir. Ancak işlem süreleri, özellikle eğitim ve puanlama süreleri dikkate alındığında, doğruluk oranının artırılmasının işlem süresi maliyetinde artışa neden olabileceğini göstermektedir. Bu durum, doğruluk ve işlem verimliliği arasında bir denge kurmanın önemini vurgulamaktadır.

Şekil 3.12. Öznitelik 8 alt veri setindeki öznitelikler arasındaki korelasyonları göstermektedir. Korelasyon katsayıları, iki değişken arasındaki doğrusal ilişkiyi ifade eder ve -1 ile +1 arasında değişir. +1, pozitif ve güçlü bir ilişkiyi, -1 ise negatif ve güçlü bir ilişkiyi ifade eder. 0 ise ilişki olmadığını gösterir.

Attribu...	Cluster...	Cluster...	Educati...	Educati...	Educati...	Educati...	Educati...	Educati...	Month...	Total_R...	Total_T...
Cluster ...	1	-0.652	0.004	0.013	0.016	-0.016	-0.011	-0.005	-0.086	0.160	-0.040
Cluster ...	-0.652	1	0.002	-0.007	-0.008	0.009	0.015	-0.003	-0.040	-0.054	0.211
Educati...	0.004	0.002	1	-0.223	-0.166	-0.077	-0.138	-0.140	0.004	-0.014	-0.015
Educati...	0.013	-0.007	-0.223	1	-0.333	-0.155	-0.277	-0.281	0.006	0.005	0.003
Educati...	0.016	-0.008	-0.166	-0.333	1	-0.115	-0.207	-0.209	-0.006	-0.002	-0.000
Educati...	-0.016	0.009	-0.077	-0.155	-0.115	1	-0.096	-0.097	-0.006	0.012	0.008
Educati...	-0.011	0.015	-0.138	-0.277	-0.207	-0.096	1	-0.174	0.010	0.008	0.010
Educati...	-0.005	-0.003	-0.140	-0.281	-0.209	-0.097	-0.174	1	-0.012	-0.004	0.002
Months_...	-0.086	-0.040	0.004	0.006	-0.006	-0.006	0.010	-0.012	1	-0.004	-0.037
Total_R...	0.160	-0.054	-0.014	0.005	-0.002	0.012	0.008	-0.004	-0.004	1	-0.347
Total_T...	-0.040	0.211	-0.015	0.003	-0.000	0.008	0.010	0.002	-0.037	-0.347	1

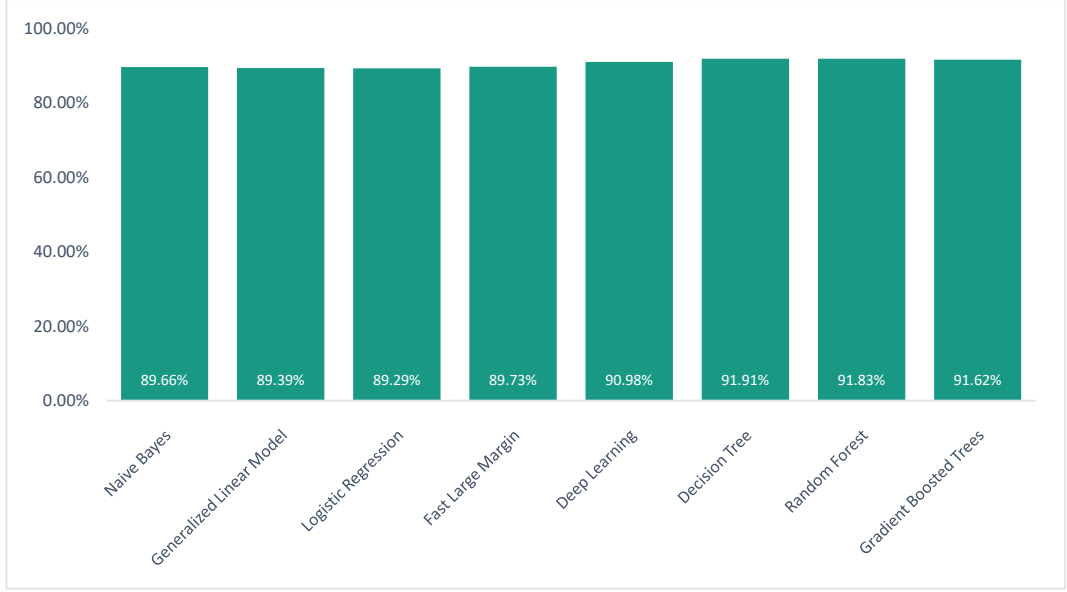
Şekil 3.12. Öznitelik 8 Korelasyon Sonuçları

3.4 9 Öznitelikli Veri Bulguları

Tablo 3.4. Öznitelik 8 alt veri seti ile yapılan modelleme sonuçlarını özetlemektedir. Ayrıca Tablo 3.4. farklı makine öğrenme modellerinin doğruluk oranları, standart sapmaları, toplam çalışma süreleri, eğitim süreleri ve test (puanlama) sürelerini içermektedir. Ayrıca Şekil 3.13. Doğruluk grafiğini ve Şekil 3.14. ve Şekil 3.15. Performans grafiğini göstermektedir.

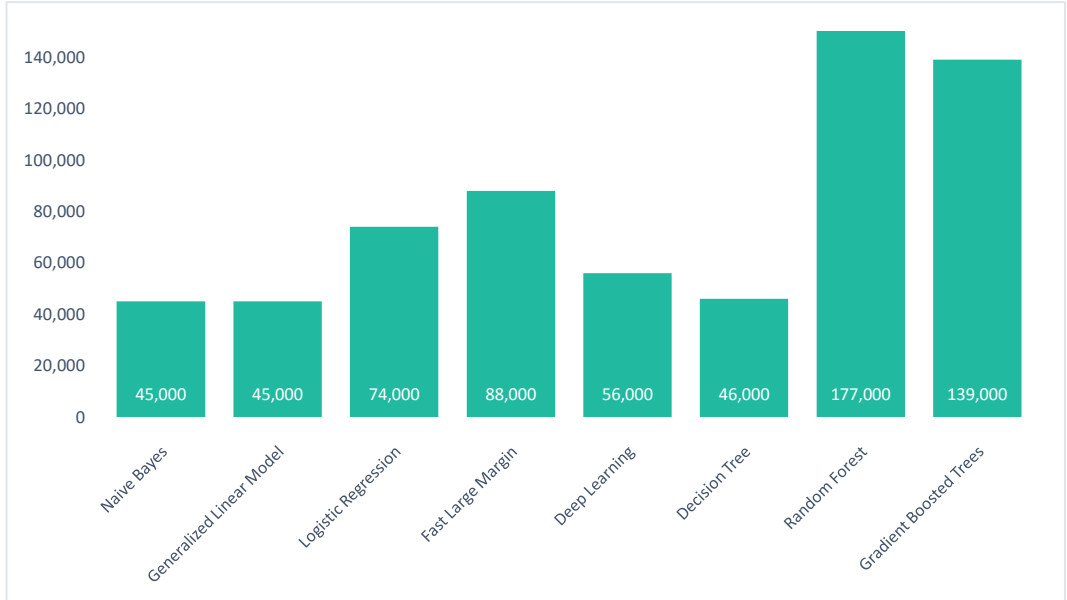
Tablo 3.4. Öznitelik 9 Alt Veri Seti Performans-Doğruluk Tablosu

MODEL	Toplam Süre(%)	Toplam Süre(s)	Toplam Süre(ms)	Eğitim Süresi(ms)	Puanlama Süresi(ms)
Naive Bayes	89,66%	45	45.000	34	105
Genelleştirilmiş Doğrusal Model	89,39%	45	45.000	185	125
Lojistik Regresyon	89,29%	74	74.000	25	180
Fast Large Margin	89,73%	88	88.000	80	200
Derin Öğrenme	90,98%	56	56.000	264	177
Karar Ağacı	91,91%	46	46.000	15	164
Rastgele Orman	91,83%	177	177.000	25	387
Gradyan Artırmalı Ağaçlar	91,62%	139	139.000	61	279
Ortalama	90,55%	84	83.750	86	202



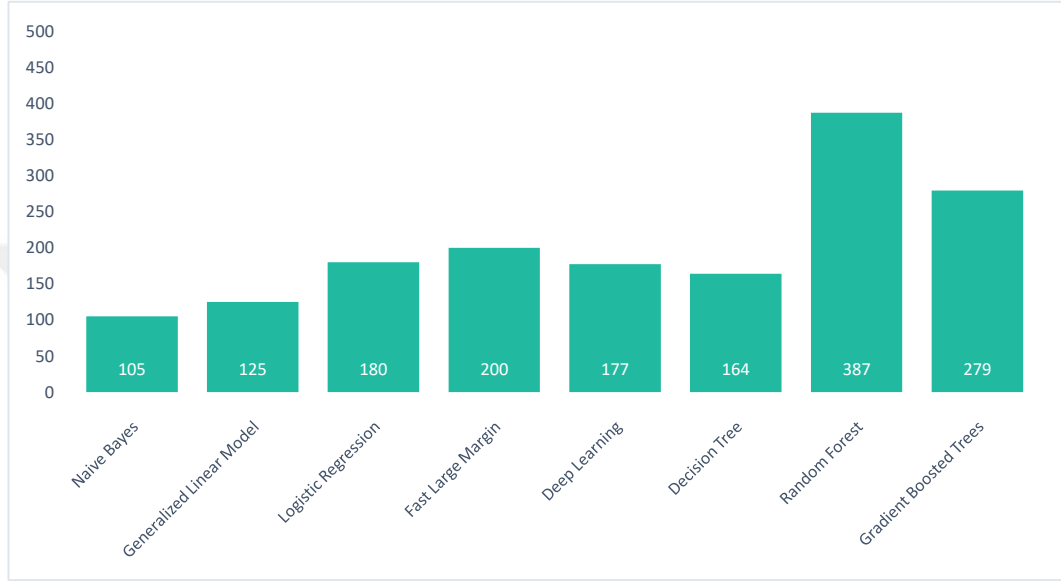
Şekil 3.13. Öznitelik 9 Doğruluk Grafiği

Bu çalışma, farklı makine öğrenmesi modellerinin doğruluk oranları ve işlem sürelerini detaylı bir şekilde incelemiştir. Tabloya göre, en yüksek doğruluk oranını %91,91 ile Karar Ağacı modeli elde etmiştir, bunu sırasıyla %91,83 doğrulukla Rastgele Orman, %91,62 doğrulukla Gradyan Artırmalı Ağaçlar ve %90,98 doğrulukla Derin Öğrenme modelleri takip etmektedir.



Şekil 3.14. Öznitelik 9 Toplam Süre Grafiği

Diğer modeller ise %89,66 ile Naive Bayes, %89,39 ile Genelleştirilmiş Doğrusal Model ve %89,29 ile Lojistik Regresyon olmak üzere yüksek doğruluk oranları sergilemiştir. Modellerin işlem süreleri incelendiğinde, en hızlı işlem süresi 45 saniye ile Naive Bayes ve Genelleştirilmiş Doğrusal Model modellerinde görülmektedir. En uzun işlem süresi ise 177 saniye ile Rastgele Orman modeline aittir.



Şekil 3.15. Öznitelik 9 Puanlama Süresi Grafiği

Eğitim süreleri açısından bakıldığında, en hızlı eğitim süresi 15 ms ile Karar Ağacı modeline, en uzun eğitim süresi ise 264 ms ile Derin Öğrenme modeline aittir. Puanlama süreleri de modele göre değişiklik gösterirken, en hızlı puanlama süresi 105 ms ile Naive Bayes modelinde, en uzun puanlama süresi ise 387 ms ile Rastgele Orman modelinde gözlemlenmiştir. Ortalama doğruluk oranı %90,55 olarak hesaplanmış olup, modellerin çoğu yüksek doğruluk değerlerine ulaşırken, işlem sürelerinin de artan doğrulukla paralel şekilde yükseldiği görülmektedir. Bu durum, doğruluk ve işlem süresi arasındaki dengeyi dikkate alarak, daha verimli bir model seçimi yapmanın önemini ortaya koymaktadır.

Öznitelik 9 da doğruluk oranında bir sıçrama gözükmemektedir. Tam olarak bu kısımda amacımıza yaklaşmış olduğumuzu görmekteyiz. Yüksek doğruluk oranına ulaşmış bulunmaktayız. Sonraki öznitelik çalışmalarında değerler karşılaştırılarak doğruluk ve performans karşılaştırmaları yapılacaktır.

Attribu...	Avg_O...	Avg_Ut...	Cluster...	Cluster...	Credit...	Depen...	Month...	Total_R...	Total_R...	Total_T...
Avg_Op...	1	-0.539	-0.631	0.778	0.996	0.068	-0.017	-0.073	-0.047	0.071
Avg_Util...	-0.539	1	0.443	-0.375	-0.483	-0.037	-0.008	0.068	0.624	0.003
Cluster ...	-0.631	0.443	1	-0.652	-0.615	-0.068	-0.086	0.160	0.183	0.156
Cluster ...	0.778	-0.375	-0.652	1	0.781	0.066	-0.040	-0.054	0.025	0.162
Credit_L...	0.996	-0.483	-0.615	0.781	1	0.068	-0.020	-0.071	0.042	0.076
Depend...	0.068	-0.037	-0.068	0.066	0.068	1	-0.011	-0.039	-0.003	0.050
Months_...	-0.017	-0.008	-0.086	-0.040	-0.020	-0.011	1	-0.004	-0.042	-0.043
Total_R...	-0.073	0.068	0.160	-0.054	-0.071	-0.039	-0.004	1	0.014	-0.242
Total_R...	-0.047	0.624	0.183	0.025	0.042	-0.003	-0.042	0.014	1	0.056
Total_T...	0.071	0.003	0.156	0.162	0.076	0.050	-0.043	-0.242	0.056	1

Şekil 3.16. Öznitelik 9 Korelasyon Sonuçları

Şekil 3.16. Öznitelik 9 alt veri setindeki öznitelikler arasındaki korelasyonları göstermektedir. Korelasyon katsayısı, öznitelikler arasındaki doğrusal ilişkiyi ifade eder ve -1 ile 1 arasında değer alır. Pozitif değerler, iki öznitelik arasında pozitif bir ilişki olduğunu; negatif değerler ise negatif bir ilişki olduğunu gösterir. 0'a yakın değerler ise ilişkisizliği işaret eder.

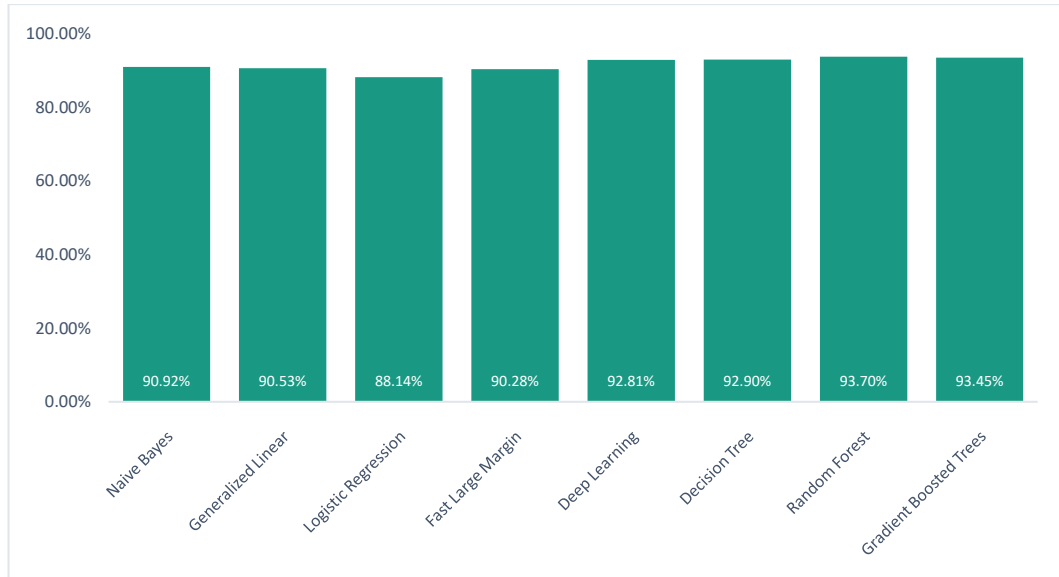
3.5 10 Öznitelikli Veri Bulguları

Tablo 3.5. Öznitelik 10 alt veri seti ile yapılan modelleme sonuçlarını özetlemektedir. Ayrıca Tablo 3.5. farklı makine öğrenme modellerinin doğruluk oranları, standart sapmaları, toplam çalışma süreleri, eğitim süreleri ve test (puanlama) sürelerini içermektedir. Ayrıca Şekil 3.17. Doğruluk grafiğini ve Şekil 3.18 ve Şekil 3.19. Performans grafiğini göstermektedir.

Tablo 3.5. Öznitelik 10 Alt Veri Seti Performans-Doğruluk Tablosu

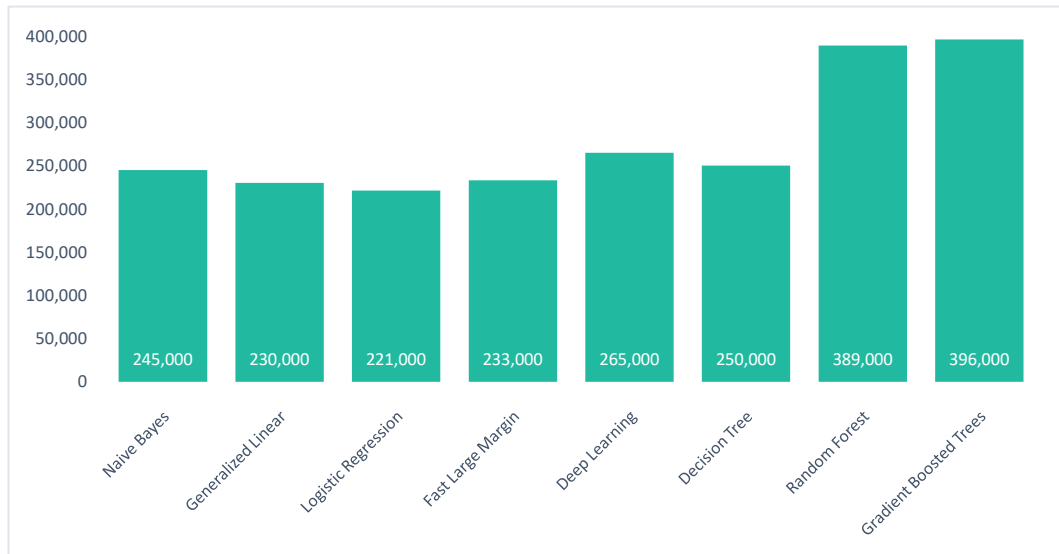
MODEL	Toplam Süre(%)	Toplam Süre(s)	Toplam Süre(ms)	Eğitim Süresi(ms)	Puanlama Süresi(ms)
Naive Bayes	90,92%	245	245.000	50	150
Generalized Linear	90,53%	230	230.000	250	200
Lojistik Regresyon	88,14%	221	221.000	70	400
Fast Large Margin	90,28%	233	233.000	157	390
Derin Öğrenme	92,81%	265	265.000	396	280
Karar Ağacı	92,90%	250	250.000	6	130
Rastgele Orman	93,70%	389	389.000	84	320
Gradyan Artırmalı Ağaçlar	93,45%	396	396.000	289	481
Ortalama	91,59%	279	278.625	163	294

Bu çalışmada, çeşitli makine öğrenmesi modellerinin doğruluk oranları ve işlem süreleri karşılaştırılmıştır. Tabloya göre, en yüksek doğruluk oranını %93,70 ile Rastgele Orman modeli elde etmiştir, bunu %93,45 doğrulukla Gradyan Artırmalı Ağaçlar, %92,90 doğrulukla Karar Ağacı ve %92,81 doğrulukla Derin Öğrenme modelleri takip etmektedir. Diğer modeller ise sırasıyla %90,92 doğrulukla Naive Bayes, %90,53 doğrulukla Generalized Linear, %90,28 doğrulukla Fast Large Margin ve %88,14 doğrulukla Lojistik Regresyon olarak kaydedilmiştir.



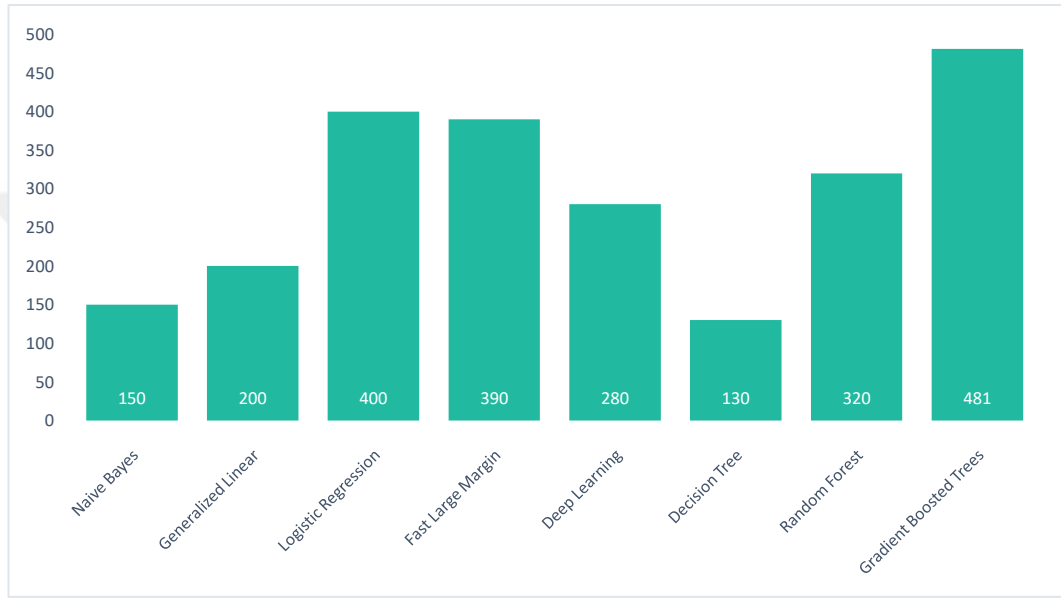
Şekil 3.17. Öznelik 10 Doğruluk Grafiği

Modellerin işlem süreleri incelendiğinde, en hızlı işlem süresi 221 saniye ile Lojistik Regresyon modeline aittir, en uzun işlem süresi ise 396 saniye ile Gradyan Artırmalı Ağaçlar modelinde gözlemlenmiştir. Eğitim süreleri bakımından en hızlı model 6 ms ile Karar Ağacı iken, en uzun eğitim süresi 396 ms ile Derin Öğrenme modeline aittir. Puanlama süreleri de modele bağlı olarak farklılık göstermekte olup, en hızlı puanlama süresi 130 ms ile Karar Ağacı modeline, en uzun puanlama süresi ise 481 ms ile Gradyan Artırmalı Ağaçlar modeline aittir.



Şekil 3.18. Öznelik 10 Toplam Süre Grafiği

Ortalama doğruluk oranı %91,59 olarak hesaplanmış olup, yüksek doğruluk oranları genel olarak modellerin performansını olumlu bir şekilde yansıtmaktadır. Ancak işlem sürelerinin, özellikle eğitim ve puanlama sürelerinin doğruluk oranı ile paralel olarak arttığı ve bu durumun modelin verimliliğini etkileyebileceği görülmektedir. Bu bulgular, doğruluk ve işlem süresi arasında bir denge kurarak, her iki faktörü de optimize etmeye yönelik bir stratejinin gerekliliğini ortaya koymaktadır.



Şekil 3.19. Öznitelik 10 Puanlama Süresi Grafiği

Öznitelik 10 da görüldüğü üzere doğruluk oranı Öznitelik 9 ile çok yakın ve performans değerlerini karşılaştırdığımızda yüksek bir sonuç vermiştir. Bu karşılaştırma durumuna baktığımızda Öznitelik 9 alt veri setimiz doğruluk ve performans konusunda en ideal alt veri seti olduğunu söyleyebiliriz.

Attribu...	Cluster...	Cluster...	Contac...	Credit_...	Depen...	Total_...	Total_R...	Total_R...	Total_T...
Cluster ...	1	-0.652	-0.135	-0.615	-0.068	0.083	0.160	0.183	-0.040
Cluster ...	-0.652	1	-0.029	0.781	0.066	0.023	-0.054	0.025	0.211
Contact...	-0.135	-0.029	1	0.021	-0.041	-0.024	0.055	-0.054	-0.113
Credit_L...	-0.615	0.781	0.021	1	0.068	0.013	-0.071	0.042	0.172
Depend...	-0.068	0.066	-0.041	0.068	1	-0.035	-0.039	-0.003	0.025
Total_A...	0.083	0.023	-0.024	0.013	-0.035	1	0.050	0.058	0.040
Total_R...	0.160	-0.054	0.055	-0.071	-0.039	0.050	1	0.014	-0.347
Total_R...	0.183	0.025	-0.054	0.042	-0.003	0.058	0.014	1	0.064
Total_T...	-0.040	0.211	-0.113	0.172	0.025	0.040	-0.347	0.064	1

Şekil 3.20. Öznitelik 10 Korelasyon Sonuçları

Şekil 3.20. Öznitelik 10 alt veri setindeki öznitelikler arasındaki ilişkileri Pearson korelasyon katsayıları ile göstermektedir. Korelasyon katsayısı, iki değişken arasındaki doğrusal ilişkiyi ifade eder ve -1 ile 1 arasında değerler alır. Pozitif değerler, iki öznitelik arasında pozitif bir ilişki olduğunu; negatif değerler ise negatif bir ilişki olduğunu gösterir. 0'a yakın değerler ise ilişkisizliği işaret eder.

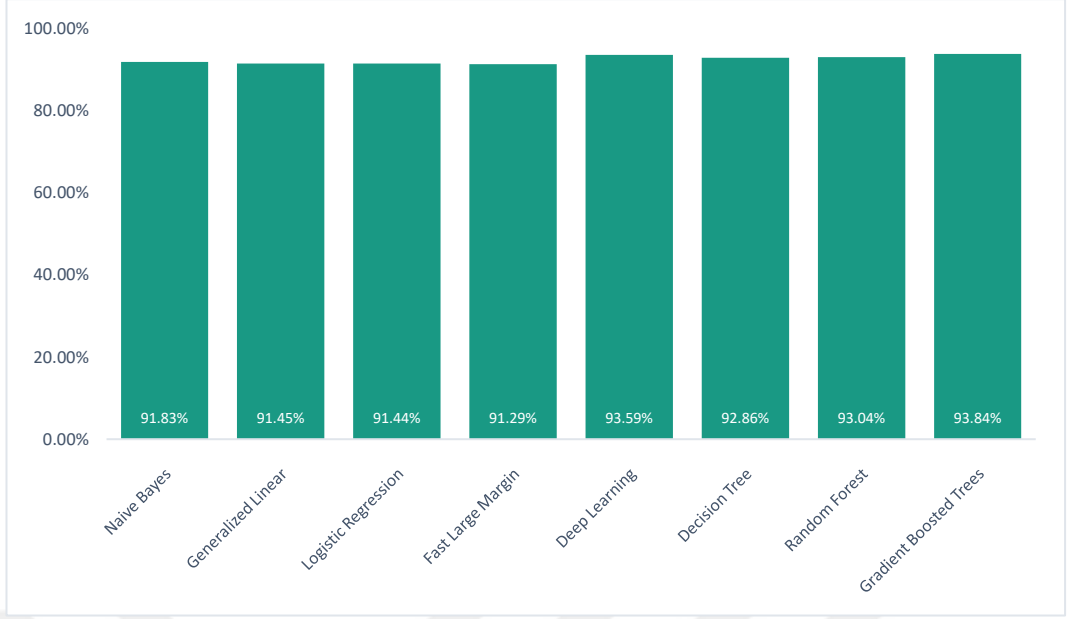
3.6 12 Öznitelikli Veri Bulguları

Tablo 3.6. Öznitelik 10 alt veri seti ile yapılan modelleme sonuçlarını özetlemektedir. Ayrıca Tablo 3.6. farklı makine öğrenme modellerinin doğruluk oranları, standart sapmaları, toplam çalışma süreleri, eğitim süreleri ve test (puanlama) sürelerini içermektedir. Ayrıca Şekil 3.21. Doğruluk grafiğini ve Şekil 3.22. ve Şekil 3.23. Performans grafiğini göstermektedir.

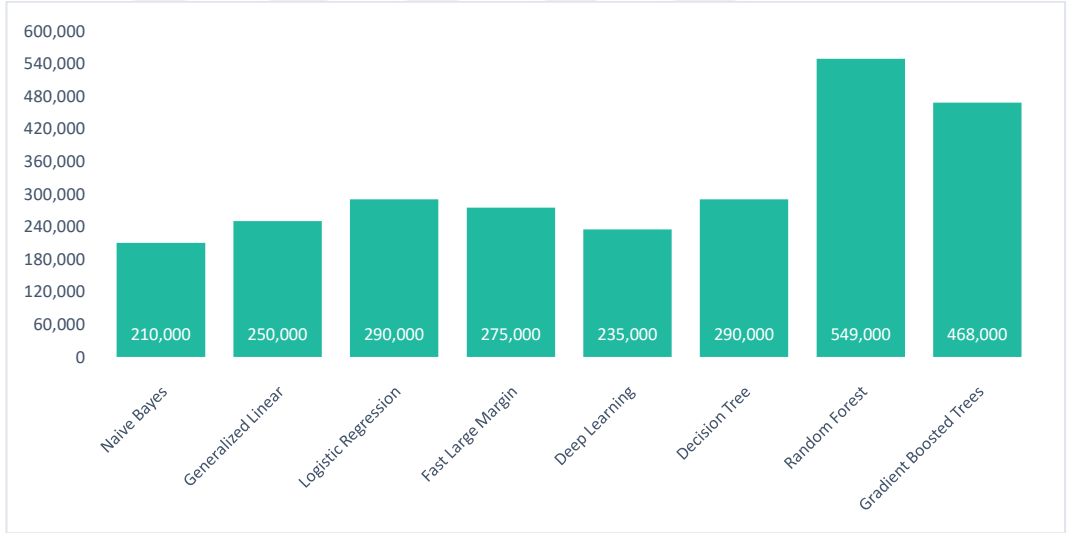
Tablo 3.6. Öznitelik 12 Alt Veri Seti Performans-Doğruluk Tablosu

MODEL	Toplam Süre(%)	Toplam Süre(s)	Toplam Süre(ms)	Eğitim Süresi(ms)	Puanlama Süresi(ms)
Naive Bayes	91,83%	210	210.000	55	270
Generalized Linear	91,45%	250	250.000	260	250
Lojistik Regresyon	91,44%	290	290.000	100	500
Fast Large Margin	91,29%	275	275.000	200	480
Derin Öğrenme	93,59%	235	235.000	402	320
Karar Ağacı	92,86%	290	290.000	6	260
Rastgele Orman	93,04%	549	549.000	116	400
Gradyan Artırmalı Ağaçlar	93,84%	468	468.000	357	651
Ortalama	92,42%	321	320.875	187	391

Bu çalışmada, farklı makine öğrenmesi modellerinin doğruluk oranları ve işlem süreleri detaylı bir şekilde analiz edilmiştir. Tabloya göre, en yüksek doğruluk oranını %93,84 ile Gradyan Artırmalı Ağaçlar modeli elde etmiştir, bunu sırasıyla %93,59 doğrulukla Derin Öğrenme, %93,04 doğrulukla Rastgele Orman, %92,86 doğrulukla Karar Ağacı ve %91,83 doğrulukla Naive Bayes modelleri takip etmektedir. Diğer modeller ise %91,45 doğrulukla Genelleştirilmiş Doğrusal Modelve %91,29 doğrulukla Fast Large Margin olarak sıralanmıştır.



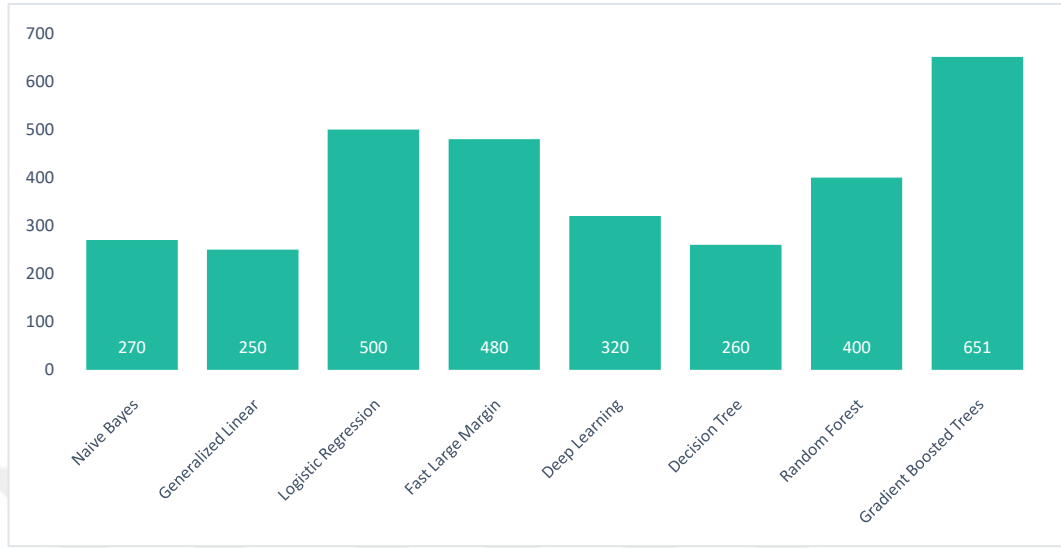
Şekil 3.21. Öznitelik 12 Doğruluk Grafiği



Şekil 3.22. Öznitelik 12 Toplam Süre Grafiği

Modellerin işlem süreleri incelendiğinde, en hızlı işlem süresi 210 saniye ile Naive Bayes modeline aittir, en uzun işlem süresi ise 549 saniye ile Rastgele Orman modeline aittir. Eğitim süreleri açısından, en hızlı eğitim süresi 6 ms ile Karar Ağacı modeline, en uzun eğitim süresi ise 402 ms ile Derin Öğrenme modeline aittir. Puanlama süreleri de modele göre değişiklik göstermekte olup, en hızlı puanlama süresi 260 ms ile Karar Ağacı modeline, en uzun puanlama süresi ise 651 ms ile Gradyan Artırmalı Ağaçlar modeline aittir. Ortalama doğruluk oranı

%92,42 olarak hesaplanmış olup, genel olarak yüksek doğruluk değerleri elde edilmiştir.



Şekil 3.23. Öznitelik 12 Puanlama Süresi Grafiği

Ancak işlem sürelerinin, özellikle eğitim ve puanlama sürelerinin doğruluk oranıyla orantılı olarak arttığı gözlemlenmektedir. Bu durum, doğruluk oranlarının artırılmasının işlem sürelerini önemli ölçüde uzatabileceğini ve model seçiminde doğruluk ile verimlilik arasında bir denge kurmanın önemini vurgulamaktadır.

Şekil 3.18. Öznitelik 12 alt veri setindeki değişkenler arasındaki korelasyonları göstermektedir. Korelasyon katsayıları -1 ile +1 arasında değer alır ve iki değişken arasındaki doğrusal ilişkiyi ifade eder. Pozitif korelasyon (+1'e yakın değerler), değişkenlerin aynı yönde hareket ettiğini, negatif korelasyon (-1'e yakın değerler) ise ters yönde hareket ettiğini ifade eder. 0'a yakın değerler ise iki değişken arasında anlamlı bir ilişki olmadığını gösterir.

Attr...	Avg.Op...	Avg...	Cluster...	Cluster...	Depen...	Income...	Income...	Income...	Income...	Income...	Month...	Total...	Total_R...	Total_R...	Total_T...	Total_T...
Total_T...	0.071	0.003	0.156	0.162	0.050	0.003	-0.028	-0.039	0.041	0.021	-0.043	0.005	-0.242	0.056	0.807	1
Total_T...	0.166	-0.083	-0.040	0.211	0.025	0.000	0.006	0.010	-0.009	-0.013	-0.037	0.040	-0.347	0.064	1	0.807
Total_R...	-0.047	0.624	0.183	0.025	-0.003	0.005	-0.004	0.023	-0.016	-0.021	-0.042	0.058	0.014	1	0.064	0.056
Total_R...	-0.073	0.068	0.160	-0.054	-0.039	-0.009	0.005	0.001	0.005	0.000	-0.004	0.050	1	0.014	-0.347	-0.242
Total_A...	0.008	0.035	0.083	0.023	-0.035	-0.001	0.022	0.013	-0.025	0.008	-0.032	1	0.050	0.058	0.040	0.005
Months...	-0.017	-0.008	-0.086	-0.040	-0.011	-0.021	-0.005	-0.006	0.019	0.011	1	-0.032	-0.004	-0.042	-0.037	-0.043
Income...	0.036	-0.069	-0.019	0.016	-0.019	-0.163	-0.141	-0.148	-0.259	1	0.011	0.008	0.000	-0.021	-0.013	0.021
Income...	-0.394	0.274	0.234	-0.309	-0.047	-0.341	-0.295	-0.311	1	-0.259	0.019	-0.025	0.005	-0.016	-0.009	0.041
Income...	0.332	-0.174	-0.215	0.268	0.048	-0.196	-0.169	1	-0.311	-0.148	-0.006	0.013	0.001	0.023	0.010	-0.039
Income...	0.094	-0.100	-0.056	0.096	0.029	-0.186	1	-0.169	-0.295	-0.141	-0.005	0.022	0.005	-0.004	0.006	-0.028
Income...	-0.162	0.070	0.111	-0.126	-0.013	1	-0.186	-0.196	-0.341	-0.163	-0.021	-0.001	-0.009	0.005	0.000	0.003
Depend...	0.068	-0.037	-0.068	0.066	1	-0.013	0.029	0.048	-0.047	-0.019	-0.011	-0.035	-0.039	-0.003	0.025	0.050
Cluster ...	0.778	-0.375	-0.652	1	0.066	-0.126	0.096	0.268	-0.309	0.016	-0.040	0.023	-0.054	0.025	0.211	0.162
Cluster ...	-0.631	0.443	1	-0.652	-0.068	0.111	-0.056	-0.215	0.234	-0.019	-0.086	0.083	0.160	0.183	-0.040	0.156
Avg_Util...	-0.539	1	0.443	-0.375	-0.037	0.070	-0.100	-0.174	0.274	-0.069	-0.008	0.035	0.068	0.624	-0.083	0.003
Avg_Op...	1	-0.539	-0.631	0.778	0.068	-0.162	0.094	0.332	-0.394	0.036	-0.017	0.008	-0.073	-0.047	0.166	0.071

Şekil 3.24. Öznitelik 12 Korelasyon Sonuçları

Bu çalışmada elde edilen bulgular, genetik algoritma ve yapay arı kolonisi algoritmasının hibrit kullanımıyla yapılan öznitelik seçiminin, makine öğrenmesi modellerinin doğruluğunu ve performansını önemli ölçüde artırdığını göstermiştir. Özellikle büyük alt veri setlerinde, seçilen özniteliklerin modele olan etkisi daha belirgin hale gelmiş ve öznitelik optimizasyonunun, modelin eğitim süresi ile doğruluk arasında kritik bir denge sağladığı görülmüştür. Bulgular, doğru algoritma seçiminin yanı sıra, verilerin etkili bir şekilde işlenmesi ve özniteliklerin uygun bir şekilde optimize edilmesinin önemini bir kez daha vurgulamaktadır.

Araştırma kapsamında yapılan analizler, basit modellerin hızlı sonuçlar verirken daha karmaşık veri setlerinde doğruluk açısından yetersiz kalabileceğini, buna karşın derin öğrenme gibi ileri düzey algoritmaların büyük veri setlerinde yüksek performans sunduğunu ortaya koymuştur. Bununla birlikte, elde edilen sonuçlar, performansı artırmak için öznitelik seçim sürecinin dikkatli bir şekilde yapılandırılması gerektiğini göstermektedir. Verinin boyutu ve karmaşıklığı arttıkça, daha gelişmiş algoritmaların öne çıktığı gözlemlenmiştir.

4. TARTIŞMA

Bu çalışma kapsamında hibrit meta-sezgisel yöntemler kullanılarak öznitelik seçimi gerçekleştirilmiş ve farklı algoritmaların performansı detaylı bir şekilde analiz edilmiştir. Çalışmanın temel bulguları, seçilen özniteliklerin doğruluk oranı, işlem süresi ve model performansı üzerindeki etkilerini açık bir şekilde ortaya koymuştur. Özellikle Genetik Algoritma (GA) ve Yapay Arı Kolonisi (ABC) algoritmalarının hibrit bir yapıda kullanılması, büyük veri setlerindeki gereksiz öznitelikleri eleyerek etkili özniteliklerin belirlenmesini sağlamıştır. Bu durum, veri setindeki gürültünün azaltılması ve makine öğrenmesi algoritmalarının daha hızlı ve doğru sonuçlar üretmesi açısından kritik bir rol oynamaktadır. Elde edilen sonuçlar, hibrit algoritma yapısının hem doğruluk oranlarını artırmada hem de işlem sürelerini optimize etmede başarılı bir yaklaşım olduğunu göstermektedir. Derin öğrenme tabanlı modeller yüksek doğruluk oranları sağlarken, daha basit algoritmaların (Naive Bayes, Lojistik Regresyon gibi) düşük işlem süreleriyle hızlı çözümler sunduğu gözlemlenmiştir. Ancak, veri setinin büyüklüğü ve karmaşıklığı arttıkça daha karmaşık algoritmaların tercih edilmesi gerektiği de görülmüştür. Çalışmada kullanılan hibrit yöntem, büyük veri setlerinde yalnızca algoritmaların performansını artırmakla kalmamış, aynı zamanda modelin genel genelleme yeteneğini de geliştirmiştir.

Ancak, çalışmada karşılaşılan zorluklardan biri, kullanılan bilgisayar ortamının donanımsal sınırlandırmaları olmuştur. Daha güçlü bir donanım altyapısının, bu çalışmada elde edilen sonuçları daha ileriye taşıyabileceği düşünülmektedir. Ayrıca, algoritmaların parametre optimizasyon süreçlerinde kullanılan hiperparametre ayarlarının da sonuçlara önemli ölçüde etki ettiği görülmüştür. Bu nedenle, gelecekteki çalışmalarda hiperparametre optimizasyonuna yönelik daha kapsamlı yöntemlerin uygulanması faydalı olacaktır.

Sonuç olarak, bu çalışmada önerilen hibrit meta-sezgisel yöntemler, müşteri sadakati gibi karmaşık bir problemi analiz etmek için etkili bir araç sunmaktadır. Elde edilen bulgular hem akademik çevrelerde hem de bankacılık sektörü gibi pratik alanlarda önemli katkılar sağlayabilir.

5. SONUÇ ve ÖNERİLER

5.1 Sonuç

Bu çalışmada, algoritmaların çalıştırıldığı ortamların ve kullanılan bilgisayarların donanım özelliklerinin sonuçlara doğrudan etkisinin büyük olduğu görülmüştür. Algoritmaların performansı, yalnızca algoritmanın kendi yapısı ile değil, aynı zamanda işlemci hızı, RAM kapasitesi, depolama birimlerinin erişim süreleri gibi donanımsal faktörler ve arka planda çalışan uygulamalar gibi yazılımsal etkilerle de ilişkilidir. Çalışmanın güvenilir sonuçlar verebilmesi için, mümkün olduğunca izole bir ortam hazırlanması ve tüm sistem kaynaklarının algoritmalara tahsis edilmesi önerilmektedir. Arka planda çalışan uygulamaların kapatılması ve bilgisayarın işlem gücünün tamamen algoritmalar için kullanılması hem işlem sürelerini azaltmakta hem de doğruluk oranlarını artırmaktadır.

Bu çalışmada hibrit algoritmalar kullanılarak öznitelik (özellik) seçimi gerçekleştirilmiş ve sonuçların doğruluğu üzerinde bu yaklaşımın önemli bir etkisi olduğu gözlemlenmiştir. Öznitelik seçiminde yapılan doğru ve etkili seçimler, algoritmaların sonuçlarını iyileştirmenin yanı sıra, analiz sürecinin daha verimli ve hızlı bir şekilde tamamlanmasını sağlamıştır. Öznitelik seçimi sürecinde hem genetik algoritma hem de yapay arı kolonisi algoritmalarının hibrit kullanımının, özelliklerin optimize edilmesine ve en etkili özniteliklerin belirlenmesine olan katkısı, bu tür yaklaşımların önemini bir kez daha vurgulamaktadır.

Çalışmada, öznitelik sayısı arttıkça, algoritmaların performans sürelerinde belirgin bir artış olduğu gözlemlenmiştir. Ancak, bu artış doğruluk oranlarındaki iyileşme ile de doğru orantılıdır. Öznitelik sayısının artırılması, modelin veri setinden daha fazla bilgi öğrenmesini sağlarken, aynı zamanda işlem süresinin ve hesaplama maliyetlerinin artmasına neden olmaktadır. Bu noktada, özellikle büyük veri setleri ile çalışılırken, öznitelik seçimi sürecinde en etkili ve gerekli özelliklerin belirlenmesi, modelin performansını optimize etmek açısından kritik öneme sahiptir. Tüm data olan 23 özniteliğe sahip veri seti üzerinde yapılan analizlerde, doğruluk performansı yüksek olmasına rağmen, modellerin aşırı öğrenme sorunu yaşadığı tespit edilmiştir. Aşırı öğrenme, özellikle gerçek zamanlı ve dinamik sistemlerde büyük bir sorun teşkil edebileceği için, bu veri seti tez kapsamındaki nihai analizlere dahil edilmemiştir. Bunun yerine, daha az öznitelik ile optimize

edilmiş veri setleri üzerinde yapılan çalışmalar, dengeli doğruluk oranları ve düşük aşırı öğrenme ile daha başarılı sonuçlar sunmuştur. Bu yaklaşım hem modelin genelleme yeteneğini artırmış hem de pratik uygulamalar için daha uygun bir çözüm sunmuştur. Öznitelik sayısının artırılmasının yalnızca doğruluk oranını artırmakla kalmayıp, aynı zamanda modelin genelleme yeteneğini de geliştirdiği görülmüştür. Bununla birlikte, her bir eklenen öz niteliğin modeli karmaşıktırdığı ve aşırı öğrenme riskini artırabileceği unutulmamalıdır. Bu nedenle, öz nitelik seçimi sırasında, yalnızca modele anlamlı katkı sağlayan öz niteliklerin seçilmesi gereklidir. Doğru öz nitelik seçimi, hem modelin işlem süresini optimize eder hem de aşırı yüklenme olmadan daha iyi sonuçlar elde edilmesine olanak tanır. Gelecekte yapılacak çalışmalarda, farklı veri setleri üzerinde öz nitelik seçimi stratejilerinin etkisinin daha ayrıntılı olarak araştırılması önerilmektedir.

5.2 Öneri

5.2.1 Donanım Yatırımı

Daha hızlı işlemciler, yüksek kapasiteli RAM ve SSD gibi modern donanımlar, büyük veri setleri üzerinde çalışan algoritmaların performansını önemli ölçüde artırabilir. Özellikle yapay zekâ ve makine öğrenimi algoritmalarında, veri işleme hızının artırılması ve bellek yönetiminin etkin bir şekilde gerçekleştirilmesi, doğruluk oranlarını ve modelleme süreçlerini doğrudan etkiler. Bu bağlamda, paralel işlemcili grafik işlem birimleri (GPU) ve tensör işlem birimleri (TPU) gibi donanımlar, makine öğrenimi çalışmalarında verimliliği artırmak için vazgeçilmezdir. Ayrıca, veri depolama ve erişim süreçlerinin optimize edilmesi amacıyla SSD teknolojileri, veri giriş-çıkış hızlarını yükselterek modelin eğitim sürelerini azaltabilir.

5.2.2 Yazılım ve Ortam Optimizasyonu

Algoritmaların etkili bir şekilde çalışabilmesi için yazılım ortamının optimize edilmesi hayati öneme sahiptir. Arka planda çalışan gereksiz uygulamaların kapatılması, donanım kaynaklarının yalnızca algoritmaların ihtiyaçlarına yönlendirilmesini sağlar. Ayrıca, Python, R gibi programlama dilleri ve TensorFlow, PyTorch gibi modern kütüphaneler kullanılarak paralel işlem destekleyen algoritmalar tercih edilmelidir. Çalışma ortamının sanallaştırma veya

bulut altyapıları ile desteklenmesi, büyük veri setleri ve karmaşık modeller üzerinde çalışırken esneklik ve ölçeklenebilirlik sağlar.

5.2.3 Hibrit Algoritma Kullanımı

Öznitelik seçiminde hibrit algoritmaların kullanımı, doğruluk oranlarını artırmanın yanı sıra işlem sürelerini azaltarak verimliliği önemli ölçüde artırır. Genetik Algoritma (GA) ve Yapay Arı Kolonisi (ABC) gibi algoritmaların birleştirilmesi, farklı problem türleri için esnek ve etkili çözümler sunar. Bu tür hibrit algoritmalar, belirli bir algoritmanın zayıf yönlerini telafi ederek daha sağlam ve güvenilir sonuçlar elde edilmesini sağlar. Gelecekte, hibrit algoritmaların daha karmaşık veri setleri üzerinde test edilmesi ve diğer meta-sezgisel yöntemlerle kombinasyonlarının araştırılması, bu alandaki çalışmalara yeni bir boyut kazandırabilir.

5.2.4 Öznitelik Sayısının Dengesi

Bir modelin performansını artırmak için daha fazla öznitelik kullanılabilir, ancak bu durum hesaplama sürelerini ve işlem maliyetlerini de artırabilir. Özellikle büyük veri setlerinde, fazla sayıda öznitelik, modelin aşırı öğrenme riski ile karşı karşıya kalmasına neden olabilir. Bu bağlamda, etkili bir öznitelik seçimi süreci, yalnızca en anlamlı öznitelikleri belirleyerek model performansını optimize edebilir. Öznitelik seçimi tekniklerinin (örneğin, filtre, sarma ve gömülü yöntemler) daha etkin bir şekilde uygulanması hem doğruluk oranlarının artırılmasına hem de işlem sürelerinin düşürülmesine katkıda bulunabilir. Bu dengeyi sağlamak, özellikle gerçek zamanlı çalışan sistemlerde kritik bir başarı faktörüdür.

Bu bulgular, makine öğrenimi algoritmalarının yalnızca veriler ve modellerle değil, aynı zamanda ortam ve donanım optimizasyonu ile desteklenmesi gerektiğini göstermektedir. Çalışmada kullanılan yaklaşımlar ve gözlemler, gelecekteki veri analizi ve makine öğrenimi projelerine ışık tutacak niteliktedir. Bu bulgular, makine öğrenimi algoritmalarının yalnızca veriler ve modellerle değil, aynı zamanda ortam ve donanım optimizasyonu ile desteklenmesi gerektiğini göstermektedir.

KAYNAKLAR

- Alhijawi, B., & Awajan, A. (2023). Genetic algorithms: Theory, genetic operators, solutions, and applications. *Evolutionary Intelligence*, 17(4), 1245–1256. <https://doi.org/10.1007/s12065-023-00822-6>
- Colgate, M., & Danaher, P. J. (2000). Implementing a customer relationship strategy: The asymmetric impact of poor versus excellent execution. *Journal of the Academy of Marketing Science*, 28(3), 375-387. <https://doi.org/10.1177/0092070300283006>
- Dorigo, M., & Stützle, T. (2004). *Ant colony optimization*. MIT Press.
- Dietterich, T. G. (2000). Ensemble methods in machine learning. In *International workshop on multiple classifier systems* (pp. 1-15). Springer. https://doi.org/10.1007/3-540-45014-9_1
- Floridi, L., & Cows, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1). <https://doi.org/10.1162/99608f92.8cd550d1>
- Ghoualmi, N., & Djebbar, A. (2018). Hybrid artificial bees colony and particle swarm on feature selection. *Computational Intelligence and Its Applications*, 93-105.
- Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley.
- Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3(Mar), 1157-1182.
- Holland, J. H. (1975). *Adaptation in natural and artificial systems*. University of Michigan Press.
- Karaboga, D. (2005). *An idea based on honey bee swarm for numerical optimization* (Technical report TR06). Erciyes University, Engineering Faculty, Computer Engineering Department.
- Karaboga, D., & Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm. *Journal of Global Optimization*, 39(3), 459-471. <https://doi.org/10.1007/s10898-007-9149-x>
- Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks* (pp. 1942-1948). IEEE. <https://doi.org/10.1109/ICNN.1995.488968>
- Konstantin, T. (2020). Bank churn data exploration and churn prediction. Kaggle. <https://www.kaggle.com/code/thomaskonstantin/bank-churn-data-exploration-and-churn-prediction/input>
- Kumar, R., & Singh, A. (2018). Hybrid metaheuristic approach based on genetic algorithm and artificial bee colony for feature selection. *Expert Systems with Applications*, 94, 227-239. <https://doi.org/10.1016/j.eswa.2017.10.041>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Li, Y., & Zhang, W. (2019). A hybrid genetic and artificial bee colony algorithm for feature selection in large financial datasets. *IEEE Transactions on Knowledge and Data Engineering*, 31(4), 667-680. <https://doi.org/10.1109/TKDE.2018.2842186>
- Liu, X., Li, Z., & Zhu, Y. (2020). Big data feature selection using hybrid genetic and artificial bee colony algorithms. *Journal of Computational Science*, 39, 101-109. <https://doi.org/10.1016/j.jocs.2019.101109>

- Londe, M. A., Pessoa, L. S., Andrade, C. E., & Resende, M. G. C. (2024). Early years of biased random-key genetic algorithms: A systematic review. *Journal of Global Optimization*. <https://doi.org/10.1007/s10898-024-01446-5>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
- Russell, S., & Norvig, P. (2016). *Artificial intelligence: A modern approach* (3rd ed.). Pearson Education Limited.
- Swerhun, M., Foley, J., Massop, B., & Mago, V. (2020). A summary of the prevalence of genetic algorithms in bioinformatics from 2015 onwards. *arXiv preprint arXiv:2008.09017*. <https://doi.org/10.48550/arXiv.2008.09017>. <https://doi.org/10.48550/arXiv.2008.09017>
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433–460. <https://doi.org/10.1093/mind/LIX.236.433>
- Vie, A., Kleinnijenhuis, A. M., & Farmer, D. J. (2020). Qualities, challenges, and future of genetic algorithms: A literature review. *arXiv preprint arXiv:2011.05277*. <https://arxiv.org/abs/2011.05277>
- Wang, H., Zhang, L., & Gao, Y. (2011). A hybrid approach for feature selection based on genetic algorithm and artificial bee colony. *Applied Soft Computing*, 11(8), 5856-5862. <https://doi.org/10.1016/j.asoc.2011.06.010>
- Watkins, S. (2011). *The neural basis of attention and perception in the human brain* [Unpublished master's thesis]. University of London.
- Yang, D., Yu, Z., Yuan, H., & Cui, Y. (2022). An improved genetic algorithm and its application in neural network adversarial attack. *PLOS ONE*, 17(5), e0267970. <https://doi.org/10.1371/journal.pone.0267970>