T.C.
İSTANBUL ÜNİVERSİTESİ-CERRAHPAŞA
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

GEMİ TAŞIMACILIĞINDA VERİYE DAYALI YAKIT TÜKETİMİ OPTİMİZASYONU

Tarık TAŞPINAR

DANIŞMAN
Doç. Dr. Zeynep ORMAN

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

Aralık, 2022

TEZ KABUL VE ONAYI

Tarık TAŞPINAR tarafından, **Doç. Dr. Zeynep ORMAN** danışmanlığında hazırlanan *"GEMİ TAŞIMACILIĞINDA VERİYE DAYALI YAKIT TÜKETİMİ OPTİMİZASYONU"* başlıklı bu çalışma, jürimiz tarafından Tarih girmek için tıklayın veya dokunun. tarihinde yapılan sınav sonucunda **oy birliği** ile başarılı bulunarak **Yüksek Lisans Tezi** olarak kabul edilmiştir.

Tez Jürisi

		İmza	Sonuç
DANIŞMAN	Doç. Dr. Zeynep ORMAN İstanbul Üniversitesi- Cerrahpaşa Bilgisayar Mühendisliği Anabilim Dalı		☒ Kabul ☐ Ret
ÜYE	Doç. Dr. Selçuk SEVGEN İstanbul Üniversitesi- Cerrahpaşa Bilgisayar Mühendisliği Anabilim Dalı		☒ Kabul ☐ Ret
ÜYE	Dr. Öğr. Üyesi Şengül BAYRAK Sabahattin Zaim Üniversitesi Yazılım Mühendisliği Anabilim Dalı		☒ Kabul ☐ Ret

BEYAN

Bu tez çalışmasının kendi çalışmam olduğunu, tezin planlanmasından yazımına kadar bütün aşamalarda etik dışı davranışımın olmadığını, bu tezdeki bütün bilgileri akademik ve bilimsel etik kuralları içinde elde ettiğimi, bu tez çalışmasıyla elde edilmeyen bütün bilgi ve yorumlara kaynak gösterdiğimi ve bu kaynakları da kaynaklar listesine aldığımı, yine bu tezin çalışılması ve yazımı sırasında patent ve telif haklarını ihlal edici bir davranışımın olmadığını ve her türlü hukuki sorumluluğu aldığımı kabul ederim.

Öğrenci Tarık Taşpınar

(İmza)

Emeği geçen tüm öğretmenlerime ithaf ediyorum…

BÜTÇE DESTEKLERİ

GEMİ TAŞIMACILIĞINDA VERİYE DAYALI YAKIT TÜKETİMİ OPTİMİZASYONU

Bu tez çalışması için herhangi bir kurumdan bütçe desteği alınmamıştır.

TEŞEKKÜR

Bu tezde emeği geçen sayın hocam Doç. Dr. Zeynep ORMAN'a destekleri için sonsuz teşekkür ederim.

Aralık 2022 **Öğrenci Tarık TAŞPINAR**

İÇİNDEKİLER

Sayfa No

1. GİRİŞ 1
1.1. ARAŞTIRMANIN KONUSU 2
2. KAVRAMSAL ÇERÇEVE 3
2.1. LİTERATÜR ÖZETİ 3
3. YÖNTEM 6
3.1. MATERYALLER 6
3.1.1. Veri Toplama 6
3.1.2. Veri Ön işleme 10
3.1.3. Öznitelik Seçimi 20
3.2. MODELLEME ÇALIŞMALARI 22
3.2.1. Model Performans Değerlendirme Metrikleri 22
3.2.2. Regresyon Analizi İçin Doğrusal Metodlar 22
3.2.3. Doğrusal Olmayan Metodlar 23
3.3. GENETİK ALGORİTMALAR İLE GEMİ HIZ OPTİMİZASYONU 28
3.3.1. Önerilen Genetik Algoritma Modeli 29
3.3.2. Genetik Algoritma Modeli Uygulama Adımları 32
3.3.3. Genetik Algoritma Parametreleri 35
3.3.4. Genetik Algoritmanın Parametre Ayarı 36
4. BULGULARIN DEĞERLENDİRİLMESİ 38
4.1. DOĞRUSAL REGRESYON TEST SONUÇLARI 38
4.2. DOĞRUSAL OLMAYAN METODLARIN SONUÇLARI 41
4.3. GENETİK ALGORİTMALAR İLE HIZ OPTİMİZASYONU SONUÇLARI 46
4.3.1. Rastgele İlklendirme Sonuçları 46
4.3.2. Manuel İlklendirme Sonuçları 53
5. SONUÇ VE ÖNERİLER 56
Ek-1: Method: Scikit-learn LocalOutlierFactor Algoritması 60

ŞEKİL LİSTESİ

Sayfa No

Şekil 1. Yakıt tüketimi ham verisi ve sefer zamanı 8

Şekil 2. Regresyon analizi için seçilen öznitelikler-1 9

Şekil 3. Regresyon analizi için seçilen öznitelikler-2 9

Şekil 4. Regresyon analizi için seçilen öznitelikler-3 10

Şekil 5. Yakıt tüketimi ham verilerinin 30 saniye ile örneklemesi 12

Şekil 6. Standart normal dağılım 13

Şekil 7. IQR ile genel bir kutu grafiği gösterimi 14

Şekil 8. Regresyon özniteliklerinin histogramları ve IQR'leri-1 15

Şekil 9. Regresyon özniteliklerinin histogramları ve IQR'leri-2 16

Şekil 10. Regresyon özniteliklerinin histogramları ve IQR'leri-3 17

Şekil 11. Ulaşılabilirlik mesafeleri [17] 18

Şekil 12. LOF methodu ile aykırı değer sonuçları (farkı k-komşuluk and kirlilik oranlarında) 20

Şekil 13. X1 and X2 serpilme diyagramı üzerinde korelasyon katsayıları 21

Şekil 14. Öznitelikler arasındaki Pearson korelasyon katsayıları 21

Şekil 15. Komşu değerlere karşı test hatası 27

Şekil 16. Hız optimizasyon probleminin akış şeması 31

Şekil 17. Geminin gerçek hız profili 33

Şekil 18. Bireyleri ilklendirmek için rastgele veya manuel hız değişimi 34

Şekil 19. Seçilen özelliklerin yakıt tüketimine karşı dağılım grafikleri, en uygun doğrusal regresyon çizgileri dahil 39

Şekil 20 Doğrusal regresyonun tahmini ve artıkları (a) Tahmine karşı gözlemlenen, (b) Artıkların normal dağılımı, (c) Normalleştirilmiş artıklara karşı tahminler, (d) Örnek niceliklerin eğilim çizgisi 40

Şekil 21. LR ve RF modellerinin öğrenme eğrileri 43

Şekil 22. SVR, KNN, ADABoost ve GBR modellerinin öğrenme eğrileri..... 44

Şekil 23. XGBoost, VR ve SR modellerinin öğrenme eğrileri..... 45

Şekil 24. Test edilen ilk 4 modelin tahmini ve ölçülen yakıt tüketimi (kg/s)..... 46

Şekil 25. Rastgele ilklendirme yönteminde parametre en iyilenmesi..... 48

Şekil 26. Rastgele yaklaşım adım-1'de en iyi hız profilleri, a) çaprazlama ve b) mutasyonlar 49

Şekil 27. Rastgele yaklaşım adım-2'de en uygun hız profilleri ve mutasyon standart sapmaları 50

Şekil 28. Rastgele yaklaşma adım-4'e en uygun hız profilleri, a) değişen hız limiti ve b) kalan mesafe için seyir hızı 51

Şekil 29. ETA gecikmesi 1 saat (sol) ve 2 saat (sağ) için yakıt tüketimi sonuçları..... 52

Şekil 30. Manuel ilklendirme yönteminde parametre en iyilenmesi 54

TABLO LİSTESİ

Sayfa No

Tablo 1 Petrol Tankeri Özellikleri [12] 6
Tablo 2 Regresyon analizi için seçilen parametreler 7
Tablo 3 Eksik veri yüzdeleri 11
Tablo 4 Yakıt tüketimi IQR yöntemi aykırı değerleri (σ = 2) 14
Tablo 5. Hiperparametreler ve optimizasyonun koşma süreleri 24
Tablo 6 Hız optimizasyon modelinin matematiksel terimleri 29
Tablo 7 Genetik algoritma uygulama adımları 32
Tablo 8 Manuel ve rastgele ilklendirmede parametrelerinin karşılaştırılması 36
Tablo 9 Eğitim veri kümesinin performans sonuçları 42
Tablo 10 Test veri kümesinin performans sonuçları 42
Tablo 11 Rastgele ilklendirilen GA'da kontrol parametreleri ve sabitler 47
Tablo 12 Rastgele ilklendirmede 1 ve 2 saatlik gecikmeler için yakıt tüketimi tasarrufları 52
Tablo 13 Manuel olarak başlatılan GA'da kontrol parametreleri ve sabit 53
Tablo 14 Manuel ilklendirmede hız limitlerinin yakıt tasarrufları 55

KISALTMA LİSTESİ

Kısaltmalar	Açıklama
ANN	Yapay Sinir Ağları (Artificial Neural Network)
ADA	Uyarlanabilir Yükseltme (Adaptive Boosting)
EEDI	Enerji Verimliliği Tasarım Indeksi (Energy Efficiency Design Index)
ETA	Tahmini Varış Zamanı (Expected Time of Arrival)
IMO	Uluslararası Denizcilik Örgütü (International Maritime Organization)
IQR	Çeyrekler Açıklığı (Inter Quartile Range)
GA	Genetik Algoritma (Genetic Algorithm)
GB	Gradyan Artırma (Gradient Boosting)
KNN	K-En Yakın Komşu (K-Nearest Neighbors)
LR	Doğrusal Regresyon (Linear Regression)
LOF	Yerel Aykırı Değer Faktörü (Local Outlier Factor)
ML	Makine Öğrenmesi (Machine Learning)
RF	Rastgele Orman (Random Forest)
RMSE	Hata kareler ortalamasının karekökü (Root Mean Square Error)
SEEMP	Gemi Enerji Verimliliği Yönetim Planı (Ship Energy Efficiency Management Plan)
SR	Yığma Regressörü (Stacking Regressor)
SVR	Destek Vektörleri Regresyonu (Support Vector Regressor)
VR	Oylama Regresyonu (Voting Regressor)
XGB	Aşırı Gradyan Artırma (Extreme Gradient Boosting)

ÖZET

YÜKSEK LİSANS TEZİ

GEMİ TAŞIMACILIĞINDA VERİYE DAYALI YAKIT TÜKETİMİ OPTİMİZASYONU

TARIK TAŞPINAR

İstanbul Üniversitesi-Cerrahpaşa

Lisansüstü Eğitim Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

Danışman : Doç. Dr. Zeynep Orman

Gemicilik sektöründe sera gazı emisyonunu azaltmak için yakıt tüketimini azaltmaya yönelik bazı düzenlemeler yapılmaktadır. Bunların başında seyir hızlarının düşürülmesi gelmektedir. Böylelikle, yakıt tüketiminin azaltılması yoluyla sera gazı salınımının düşürülmesi amaçlanmaktadır. Bu çalışmada, gemilerin yakıt tüketimini en aza indiren seyir hızları araştırılmıştır. Bu amaçla, öncelikle gemi hızına etki eden faktörler incelenmiştir. Bu faktörler gemi taşımacılığında karşılaşılan hava ve deniz koşulları, gemi makinelerinin anlık teknik parametreleri, gemi taşımacılığı piyasasındaki anlık değişebilen ekonomik veriler ve işletme faktörleridir. Bu tez kapsamında kullanılmak üzere gemi işletmesine ait hız, anlık yakıt tüketimi, derinlik, trim, rüzgar hızı vb. olmak üzere otuzdan fazla sayıda parametre kullanılmıştır. Kullanılan veri kümesi bir günlük kayıtları kapsamakta olup kayıt sıklığı bir saniyedir. Ham veri kümesi üzerinde çeşitli veri bilimi teknikleri uygulanarak eksik veya hatalı verilerin düzeltilmesi ve yerel anormallik faktör algoritması kullanılarak anormalliklerin temizlenmesi sağlanmıştır. Daha sonra, korelasyon analizi ile regresyon değişkenleri arasındaki ilişkiler belirlenmiştir. En az yakıt tüketimiyle sonuçlanan optimum gemi hızının araştırılması için ön işlemden geçirilmiş veriler kullanılarak yakıt tüketimi tahmin modeli ortaya çıkarılmıştır. Regresyon tahmin modeli için bir dizi makine öğrenmesi teknikleri kullanılmıştır. Bunlar; basit lineer regresyon (LR), K-En Yakın Komşuluk algoritması (KNN), Destek Vektörleri (SVR) ve çeşitli Rastgele Orman (RF), ADABoost, Gradyan Arttırmalar (GRB ve XGB) ve topluluk yöntemleridir. Ayrıca, tüm modeller için hiper parametre optimizasyonu yapılmıştır.

Gemi seferlerinde en az yakıt tüketimine sahip optimum hız profili belirlemek için genetik algoritmaların (GA) seçilmesindeki en büyük neden hız optimizasyonunun çözümünde rastgele

yaklaşımların kullanılabilmesidir. Literatürdeki çalışmalarda optimum hıza etki eden parametreler tam olarak belirlenememiştir. Hava olayları, finansal parametreler, teknik arızalar vb. gemi hızına etki edebilecek parametreler rastgele yaşanabilmektedir. Bu nedenle, gemi hızında bir dağılıma bağlı kalmaksızın rastgele değişimler yapılarak optimizasyon modeli elde edilmiştir. Ayrıca, gemi hızı GA algoritmasında bireylerin ilk değeri olarak ele alınmıştır.] Aralık 2022 , [79] sayfa.

Anahtar kelimeler: [Yerel aykırılık faktörü, en yakın komşuluk, rastgele ağaçlar, gradyan arttırma, destek vektör regresyonu, topluluk öğrenmesi, gemi hız optimizasyonu, genetik algoritma, DEAP, HyperOpt]

ABSTRACT

[M.Sc. THESIS]

[DATA DRIVEN FUEL CONSUMPTION OPTIMIZATION OF VESSELS]

[Tarık Taşpınar]

İstanbul University-Cerrahpaşa

Institute of Graduate Studies

Department of Computer Engineering

Computer Engineering Program

[Supervisor : Assoc. Prof. Dr. Zeynep ORMAN]

[Due to increasing amount of greenhouse gas emissions of ships, an energy efficiency studies are emerged with less fuel consumption. Speed optimization is one of most beneficial ways to decrease fuel consumption of ships. But, there exist many other parameters to cause more fuel consumption for ships. In this thesis, a speed optimization study is implemented using genetic algorithm. Because, genetic algorithms are well suited for natural environmental variables of ships which have stochastic behaviours. Subsequently, a fuel consumption estimator regressor method are also researched. For this purposes, linear and non-linear methods such as KNN, support vectors, random forest, gradient boosting, ada boost, xgboost and ensemble models accuracy results are compared for better fuel consumption predictions in speed optimization using genetic algorithms. Furthermore, hyperparameters are researched for all models. Besides, the speed profiles are modifies by manual and random approach. Because the genetic algorithm with different speed profiles showed higher performance for fuel consumption effieciency. Moreover, Mutations and crossovers supported genetic varieties of speed profiles. The speed limits are caused less fuel consumption efficiencies. Because genetic algorithm losses the new speed profile trials under speed limit conditions. On the other hand, port and voyage delaying

times are so beneficial for allowing to decrease speed and fuel consumption by genetic algorithms.]

December 2022, [79] pages.

Keywords: [Local outlier factor, support vector machines, random forest, gradient boosting, ensemble learning, KNN, genetic algorithm, ship speed optimization, DEAP.]

1. GİRİŞ

Deniz taşımacılığındaki sera gazı emisyonları (GHG), küresel ısınmaya en çok katkıda bulunan etkenlerden biridir. Ayrıca, deniz taşımacılığında sera gazı emisyonlarının 2008 yılındaki değerlerinin, 2050 yılına kadar %130 artması öngörülmektedir [1]. Uluslararası Denizcilik Örgütü (International Maritime Organization - IMO), yayınladığı stratejide enerji verimliliği tasarım endeksinin (Energy Efficiency Design Index - EEDI) uygulamaya konması ile deniz taşımacılığının karbon yoğunluğunu azaltmayı hedeflemiştir. EEDI, farklı gemi tür ve boyutları için ton-mil başına minimum yakıt tüketimini ifade etmektedir. Gemi Enerji Verimliliği Yönetim Planı (Shipboard Energy Efficiency Management Plans and Energy Audits - SEEMP) ise, gemi operatörlerinin yeni sefer planlaması, hava durumuna bağlı rota belirleme, limanlara gecikme, hız optimizasyonları, şaft gücü, trim, balast, pervane ve atık ısı geri kazanım sistemleri gibi operasyonel sistemler kullanılarak yakıt tüketiminin azaltılmasını amaçlamaktadır [2]. Bu sistemler, deniz taşımacılığında teknik, operasyonel ve çevresel verilerin oluşturduğu zorunlu bir veri toplama sistemini gerektirmektedir [3]. Bu nedenlerle 2018 yılında 5000 gros ton üzeri gemiler için zorunlu olarak akaryakıt tüketim verileri toplama sistemi kullanılmaya başlanmıştır. Ayrıca, veri toplama sisteminin ileriki yıllar için planlanan diğer bir amacı ise veri analizinin yapılabilmesi ve gelecekte ihtiyaç duyulması halinde karar verme sistemleri tarafından kullanılabilmesidir [4].

Deniz taşımacılığı sektöründe, hız kısıtlamaları ile ilgili yeni düzenlemeler tartışma konusu olmaktadır. IMO tarafından, yakıt tüketimini ve seyir maliyetlerini azaltmak amacıyla gemilerin tasarlanırken belirlenen ve optimum seyir hızları için hız limitleri önerilmektedir. Bunun gerekçesi ise şimdiye kadar genel bir kural olarak bilinen; seyir hızı %10 azalırsa, ilgili motor gücü ve sefer için gerekli toplam enerji sırasıyla %27 ve %19 azalmakta olduğudur [5].

Ancak, hız azaltma tartışmaları halen devam etmektedir. Yönetmelikte maksimum hız sınırlarının mı yoksa maksimum ortalama hızların mı uygulanacağı hala belirsizliğini korumaktadır. Çünkü hız azaltılmasına endüstri tarafından birçok eleştiri yöneltilmiştir. Örneğin, Stena şirketi, yoğun sezonlarda her yolculuk için hızı %20 oranında düşürmenin mümkün olmadığını iddia etmektedir. Ayrıca, Terntank şirketi, bu durumun sadece yeni modern gemilerin inşasını azaltacağını öngörmektedir. İsveç Orient Line şirketi ise aynı

miktarda malın teslimi için daha fazla geminin gerekeceğini iddia ettikleri için hız azaltmanın maliyetleri arttıracağını düşünmektedir [6].

1.1. ARAŞTIRMANIN KONUSU

Gemi operasyonel verilerini kullanan literatürdeki son çalışmalar, hız azalsa bile yakıt tüketiminin artabileceğini göstermiştir. Ayrıca, çalışmalar aynı zamanda her yolculuk için optimum bir hız profilinin olduğunu ve optimum hızdan sapmaların daha fazla yakıt tüketimine neden olabileceğini göstermektedir [6]. Çünkü, hava şartları, rüzgar yönü, dalga büyüklüğü, limanlardaki trafik, bunker fiyatları ve daha birçok parametre anlık olarak değişerek nakliyede gecikmelere neden olmaktadir. Hava şartları kötü olsa bile, rüzgar yönünün avantajı veya limanlardaki trafikte beklemenin maliyeti nedeniyle yolculuğun belirli bir döneminde hız yükseltmek avantajlı olabilirken, yine hava şartları nedeniyle tersi de geçerli olabilmektedir. Bu yüzden bir yolculuk boyunca sabit hız veya hız limiti uygulamak yerine, enerji verimliliğini arttırma amacıyla kurulan yapay zekalı karar destek sistemleri sayesinde değişken hızların bulunduğu bir hız profili ortaya çıkabilmektedir. Bu nedenle, bu çalışmada, gemi operatörü tarafından müsade edilen gecikmeler için minimum yakıt tüketimine yol açan optimum gemi hızı profilleri araştırılmıştır. Bununla birlikte, optimum gemi hızı profillerini ortaya çıkarmak için optimizasyon problemlerinin çözülmesi gerekir. Bu nedenle, bu çalışmada gemilerin bir yolculuk boyunca kaydedilmiş hız profillerinde rastgele değişiklikler yaparak en iyi hız profilini bulmayı hedefleyen bir genetik algoritma önerilmektedir.

Öte yandan, yine bu çalışmada, her hızdaki yakıt tüketimini tahmin etmek için bir takım makine öğrenmesi modelleri kullanılarak bir yakıt tüketimi tahmin modeli de araştırılmıştır. Sonuç olarak, ham gemi verileri için ilk veri ön işleme aşamasından sonra, bir yakıt tüketim modeli geliştirilmiştir. Sonrasında en iyi sonucu veren tahmin fonksiyonu genetik bir algoritmada amaç fonksiyonu olarak kullanılmıştır.

Bu nedenlerle, bu tezde keşifsel ve tahminsel veri analizi ve hız optimizasyonu uygulanmıştır.

2. KAVRAMSAL ÇERÇEVE

2.1. LİTERATÜR ÖZETİ

Gemi ulaşımında enerji verimliliği çalışmaları literatürde yaygın olarak bulunmaktadır. University College London (UCL) Enerji Enstitüsü (2013) [7] Otomatik Tanımlama Sistemi-OTS (Automatic Identification System - AIS) uydu verilerini kullanarak farklı filoların operasyonel verimliliğinin bir değerlendirmesini yayınlamıştır. AIS veri sonuçları, gemilerin ortalama hızında yapılan %10-15 oranındaki bir azaltmanın, toplu filolar için yakıt verimliliğini yaklaşık %10 ve konteyner filoları için %30'a kadar artırabileceğini göstermiştir.

Ancak, zamanla tam tersi bir görüş de ortaya çıkmıştır. Karl ve diğerleri (2020) [6], İsveç ticaretindeki gemiler için hız azaltmanın sonuçlarını araştırmıştır. Araştırma sonuçları, yakıt tüketiminin düşük hızlarda da artabileceğini göstermiştir. Ayrıca, en önemli bulgulardan biri, her yük seviyesi için optimum hıza göre değişebilen minimum yakıt tüketimidir. Hızı bu optimum noktadan yukarı veya aşağı yönde değiştirmek, daha fazla yakıt tüketimine neden olabilmektedir.

Aslında, daha önceleri baskın görüş, gemi hızının kübik bir fonksiyonu ile, tarihsel verilerin kullanılmadığı zamanlarda, gemilerin yakıt tüketiminin açıklanabildiği varsayılıyordu [8]. Böylece, gemilerin sabit olan tasarım hızları seyir için de optimum hız haline gelmişti. Optimum nokta etrafında hızın esneklik potansiyelini göstermek için Adland ve diğerleri [8] petrol tankerlerinin öğlen raporlarını (noon reports) dikkate almıştır. Çalışma sonuçları ise, hız azaltma değerlendirmelerinin yakıt tüketimi verimliliğini olduğundan daha yüksek tahmin ettiğini göstermiştir. Çünkü, hız esnekliği için kübik yöntemlerin kullanılması, regresyon modellerinden daha fazla yakıt tasarrufuyla sonuçlanmıştır. Ayrıca, ilgili çalışmada doğrusal olmayan bir regresyon modeli önerilmiş olsa bile tahminler için global optimum bulma yeteneğine sahip olmadığı iddia edilmiştir. Lineer regresyon modelinin yakıt tüketimi tahmin doğruluğu ise Aframax ve Suezmax tankerleri için %80 civarındaydı. Ancak, beş gizli katmana sahip sinir ağları daha yüksek doğrulukta sonuçlar vermiştir.

Daha yüksek doğruluk ve daha düşük hata oranlarına sahip yakıt tüketimi tahmini, optimizasyon çalışmaları için bir ön koşuldur. Kim ve diğerleri [9] geminin ana motor hızı, yere göre hızı, rüzgar hızı, dümen açısı, trim, ıslak yüzey alanı ve deplasman parametreleri ile ANN veya çoklu doğrusal regresyon (MLR) modellerini yakıt tüketimini tahmin etmek için

kullanmıştır. Bu çalışmanın sonucunda ANN ve MLR modelleri sırasıyla %99 ve %87 doğruluk sonuçlarıyla çıktı vermiştir.

Gemilerde yakıt tüketimini etkileyen çok sayıda parametre bulunmaktadır. Ancak, bazı değişkenlerin modelde sadece gürültü etkisi vardır ve sonuca anlamlı bir etkisi de bulunmaz. Ayrıca, bazı parametreler aşırı öğrenmeye (overfitting) veya kötü algoritma performansına neden olabilir. Bu nedenle, parametre seçimi, makine öğrenmesi modellerinde ele alınması gereken önemli bir konudur. Bu konuda Corradu ve arkadaşları [10] hesaplamalı olarak düşük performans gösteren rastgele orman yöntemini (random forest method) kullanmıştır. Karar ağaçlarını seçmek için ise kaba kuvvet yöntemini (brute force method) kullanmıştır. Sonunda, kaba kuvvet yöntemleri daha iyi doğruluk sonuçları vermiştir. Daha kesin olarak, pervane adımı ve gemi hızı parametrelerinin yakıt tüketimi tahmininde daha fazla etkisi görülmüştür. Ayrıca, gemi taslağı ve şaft gücü, yakıt tüketimi tahmini için en önemli parametreler arasında bulunurken, rüzgar hızı ve yönü parametreleri ise ciddi bir etkiye sahip olmamıştır.

Parametrik modellerin öznitelik seçimini abartılı tahmin edebilmesi nedeniyle aşırı öğrenmeyi önlemek için parametrik olmayan modeller kullanılabilir. Gkerekos ve arkadaşları [11], en iyi model performansını bulmak için bir grup parametrik ve parametrik olmayan makine öğrenmesi modellerini birlikte test etmişlerdir. Ridge and Lasso regresyon, Destek Vektörleri (SVR) ve ANN parametrik modeller olarak kullanılmıştır. Ayrıca, Karar Ağacı Regresörü (DTR), Ekstra Ağaç Regresörü (ETR), K-En Yakın Komşular (KNN) ve Rastgele Orman Regresörü (RFR) parametrik olmayan modeller olarak ele alınmıştır. Çalışmanın varsayımı ise, başlangıçta parametre varsayımları olmadan çalıştırıldığı için, parametrik olmayan modellerin daha yüksek performans göstereceğidir. Sonuçta en iyi doğruluk sonuçları, randomize ETR'lerden elde edilmiştir.

Yang, Chen [12], bir petrol tankerinin hız optimizasyonu için bir genetik algoritma kullanmıştır. Öncelikle tüm rota birkaç etaba ayrılmıştır. Bu etaplar hız, rüzgar, dalgalar ve okyanus akıntıları dikkate alınarak değiştirilmiştir. Genetik optimizasyon kısıtlamaları için maksimum ve minimum seyir hızları ve beklenen varış süreleri uygulanmıştır. Ayrıca, yakıt tüketimi, minimum tutulacak bir amaç fonksiyonu olarak ele alınmıştır. Sonunda, her segment için doğru hızın seçilmesiyle birlikte yapılan çalışma %2,2 daha az yakıt tüketimi ile sonuçlanmıştır.

Deniz taşımacılığında beklenen varış zamanı pek çok beklenmedik olayla değişebilir. Lineer taşımacılıkta limanlarda yaşanan bekleme ve gecikmeler ekstra maliyete neden olabilmektedir.

Aydın ve Lee [13] değişken tahmini varış tarihi (Estimated time of arrival - ETA) gecikmeleri ve bunker fiyatlarını kullanarak optimum seyir hızını araştırmıştır. Dinamik programlama, belirsiz liman sürelerini göz önünde bulundurarak seyir hızını belirlemek için kullanılmıştır. Çünkü, gemi kaptanı seyahat sırasında bir sonraki liman sıkışıklığını ve yakıt ikmal fiyatlarını anlık kontrol ederek gemi hızını arttırmayı ya da düşürmeyi tercih edebilirler. Ayrıca, alternatif olarak, dinamik programlamada rüzgar hızı, dalga yüksekliği vb. rastgele niceliklerin beklenen değerleri kullanılarak deterministik bir yöntem de uygulanmıştır. Sonuçta, ortalama seyir hızları en deterministik modelde dinamik programlama modeline göre daha yüksek hesaplanmıştır. Ayrıca gecikme cezası arttıkça deterministik modelin performansı düşmüştür.

Tahmini varış zamanı ve yakıt tüketimi birbiriyle ters orantılıdır. ETA'nın kısaltılması daha fazla yakıt tüketimine neden olabilmektedir veya bunun tersi de geçerlidir. Helong, Xiao [14], iki farklı amaç fonksiyonu ile genetik algoritma kullanarak yolculuk optimizasyon modeli geliştirmiştir. Bunlar, yakıt tüketimi minimizasyonu ve varış dakikliğidir. Zorlu deniz koşullarında rotaları sezgisel olarak planlayarak ve aynı ETA'ları koruyarak yakıt tüketiminde %3,4'e varan tasarruf sağlanmıştır. Ayrıca, bu çalışma, genetik algoritma popülasyonlarında genetik çeşitliliği artırmanın, amaç fonksiyonunun başarısını yani yakıt tüketimi tasarrufunu artırdığını göstermiştir.

3. YÖNTEM

3.1. MATERYALLER

3.1.1. Veri Toplama

Bu çalışmada kullanılan veri kümesi 69 ay boyunca çeşitli yük ve ortam koşullarında kaydedilmiştir [15]. Orijinal veri kümesinin örnekleme hızı 1 saniyedir. Hesaplama süresini kısaltmak ve optimizasyon algoritmalarının daha iyi yakınsamasını sağlamak için örnekleme hızı 30 saniyeye düşürülmüştür.

3.1.1.1. Veri Kaynağı

Bu tezde kullanılan veriler proje paydaşları tarafından sağlanmış olup İsveç'teki bir petrol tankerine aittir. Bu petrol tankerinin fiziksel özellikleri Tablo 1'de listelenmiştir. Tanker, sıcak yük taşıma konusunda uzmanlaşmıştır ve bu nedenle daha büyük bir balast suyu tankına sahiptir. Ayrıca geminin ön ve arka tarafı arasındaki suya batma seviyesi farkı trim olarak tanımlanır ve gemi yüksüzken trim daha çok değişir, ancak balast koşullarında da trim çok küçük değerlerde değişir [15].

Tablo 1 Petrol Tankeri Özellikleri [12]

Özellik	Boyut	Değer
Tam Boy	m	99
Dizayn Draftı	m	5.7
Ağırlık	ton	4972
Yapım Yılı		2012
Kargo Kapasitesi	m^2	4300
Balast	m^2	2000
Ana Motor	kW	4000
Şaft Jeneratörü	ekW	760
Servis Hızı	knots	14.0

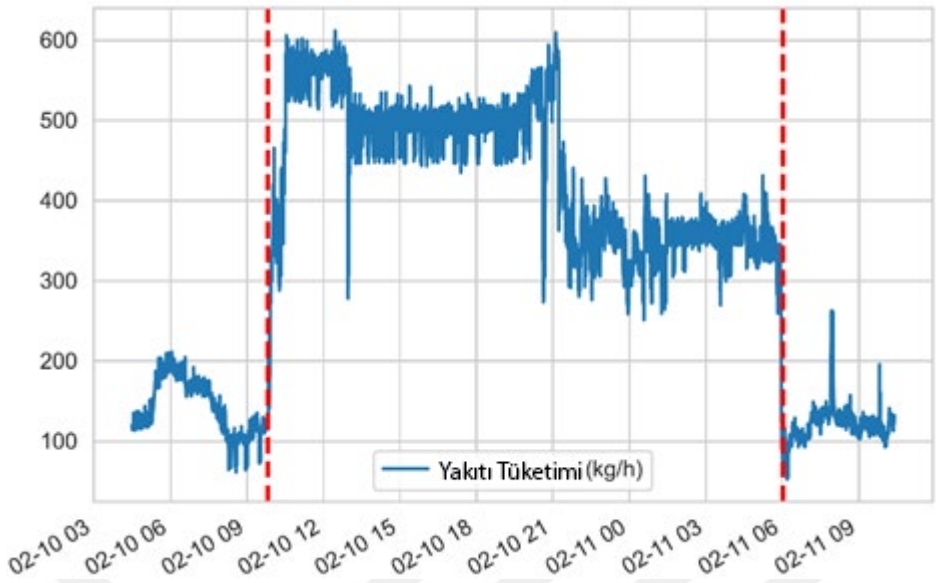
Bu tankere ait birçok çevresel, operasyonel ve sefer veri günlüğü arasından bu çalışma için sadece 14 tanesi seçilmiştir. Çünkü seçim sürecinde yalnızca yakıt tüketimi tahmini ve hız optimizasyonu için zorunlu değişkenler dikkate alınmıştır. Yakıt tüketimi verileri, bu tezdeki ana tahmini hedeftir. Ancak ana motorların çıkış gücü veya pervane hatvesi ile de temsil

edilebilir. Yakıt tüketimi ile ana motorlar veya pervane hatvesi arasındaki korelasyon yüzde yüz olmasa da sonuç olarak yakıt sadece ana motorlar tarafından tüketilmektedir. Bu nedenle, diğer parametrelerin tahmin gücünü kaybetmemek için ana motorlar veya pervane hatvesi veri değişkenleri bu çalışmanın dışında tutulmuştur. Seçilen tüm parametreler Tablo 2'de listelenmiştir.

Tablo 2 Regresyon analizi için seçilen parametreler

Id	Parametre	Boyut
1	Yere Göre Hız	knots
2	Yakıt Tüketimi	kg/h
3	Balast	ton
4	Dizel Yakıt	ton
5	Omurga Derinliği	m
6	Navlun	ton
7	Su	ton
8	Meyil Açısı	deg
9	Seyir Yönü	deg
10	Trim	m
11	Pupa Rüzgarı	m/s
12	Rüzgar Hızı	m/s
13	Rüzgar Yönü	deg

Tanker gemilerinin sefer süreleri limanlar arası mesafeye göre değişmektedir. Şekil 1'de gözlemlenen yakıt tüketimi 1 günlük aralıkta gösterilmiştir ve tüm yolculuğun 24 saat aralığında kaydedilmiştir. Daha doğrusu, sabah 9 ile ertesi gün sabah 6 arasındadır. Bu tezde, yakıt tüketimi analizi, gemi bakım veya manevra zaman aralıklarından çok, sefer süresince operasyonel ve çevresel parametrelere odaklanmıştır. Bu tezde motor parametreleri kullanılmadığı için düşük yakıt tüketim değerleri filtre edilmiştir. Çalışmada kullanılan zaman aralığı Şekil 1'de görülen kırmızı dikey çizgilerle gösterilmiştir.

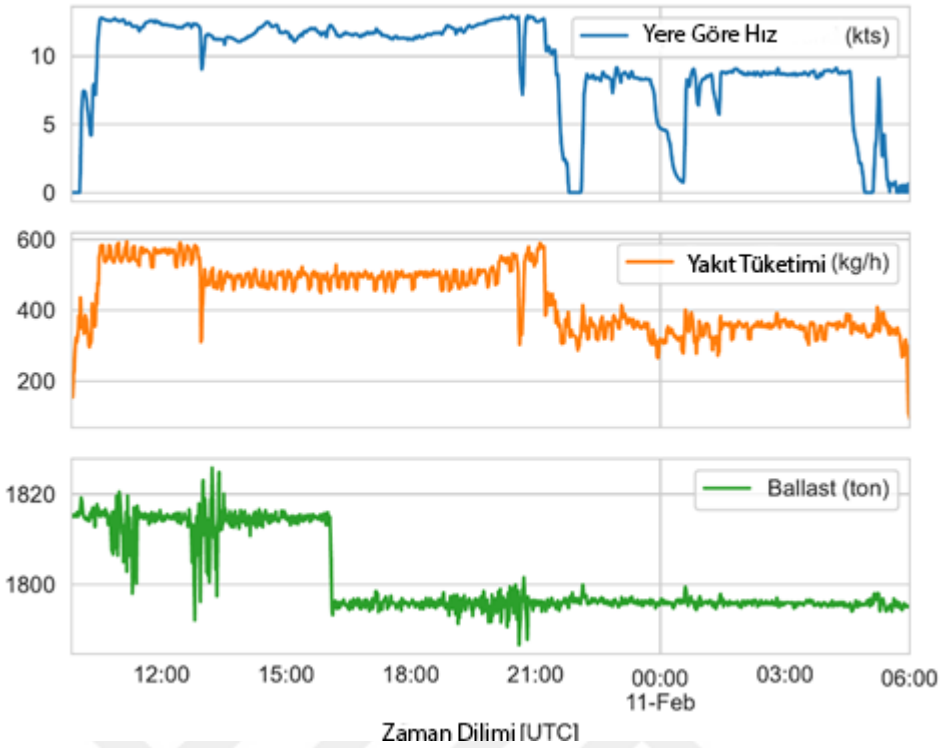


Şekil 1. Yakıt tüketimi ham verisi ve sefer zamanı

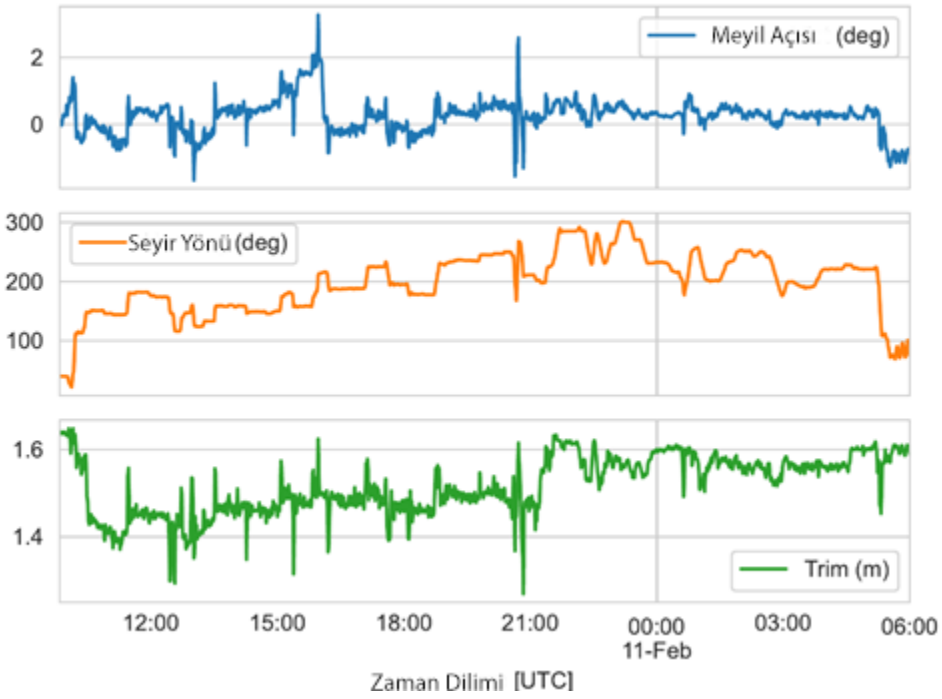
Tablo 2'de regresyon analizi için seçilen değişkenler zaman domeninde verilmiştir. Orijinal örnekleme hızı (1 saniye), tüm veri kümesine uygulanan ortalama 30 saniyelik alt örnekleme işlemi ile değiştirilmiştir.

Şekil 2, 3 ve 4'te veri kümesinde bulunan tüm değişkenlerin 1 günlük grafikleri gösterilmektedir. Rüzgar hızı, yere göre hız ile yakıt tüketimi arasındaki korelasyonlar gözle görülebilmektedir. Ancak diğer değişkenler için (örneğin balast yükü) bunu söylemek zordur.

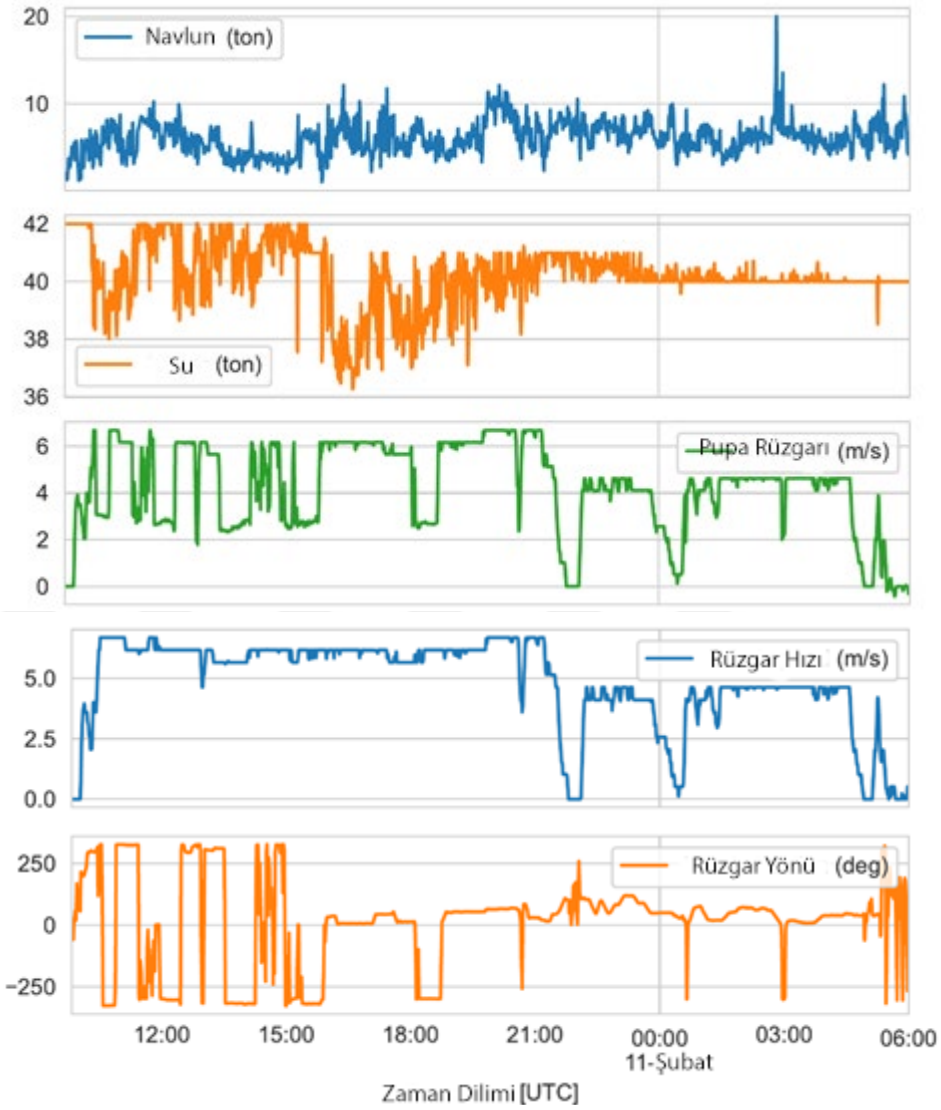
Regresyon algoritmalarının her örneklem için hesaplama yapması ve eğer eksik veri varsa hata vermesi nedeniyle, öncelikle tüm değişkenlerde eksik veya yanlış verilerin araştırılması yapılmalıdır. Bu amaçla bazı veri ön işleme teknikleri kullanılmıştır.



Şekil 2. Regresyon analizi için seçilen öznitelikler-1



Şekil 3. Regresyon analizi için seçilen öznitelikler-2



Şekil 4. Regresyon analizi için seçilen öznitelikler-3

3.1.2. Veri Ön işleme

Veri analizinden önce ham veriler, birçok veri ön işleme ve istatistik araçları kullanılarak hazırlanmıştır. Çünkü ham veriler içerisinde eksik, gürültülü ve regresyon modellerinde veya anormallik problemlerinde daha az doğruluğa neden olabilecek aykırı değerler bulunmaktadır [16].

3.1.2.1. Veri Temizleme

Bir verikümesinde veri eksikliği durumunda esas olarak istatistiksel özellikler kaybolabilir ve hiper parametre optimizasyonunda bir sapma etkisine neden olabilir [17]. Böylece optimum parametrelerle en yüksek doğruluğun elde edilmesini engellenmiş olur. Ayrıca, Python diline ait veri bilimi metodları eksik verilerle çalışmamaktadır.

Eksik veri işleme üç ana başlık altında toplanabilir [17]:

1. **Öznitelik Azaltma**: Bu yöntem, sütunların silinmesini önerir. Bir sütunu yalnızca verilerin %70'i eksikse silmek mümkündür. Bu yöntemi "Downstream (kts)" sütunu için uygulanmıştır (bkz. Tablo 3). Çünkü bu değişkenin %100'ü kayıptır.

2. **Kısmi Silme**: Bu yöntem, eksik verileri içeren satırların silinmesini önerir. Zaman serisi verileri dönemsel etkilere sahip ve periyodik ise parçacık silme, örneklenen verilere ait bir periyotta bozulmaya neden olabilir. Ancak bu tezde kullanılan veri kümesi periyodik değildir. Yalnızca 'Ekonomi' değişkeni %35,23 oranında eksik değere sahiptir (bkz. Tablo 3). Bu değişken, regresyon modelleri için gerekli değildir. Bu nedenle, tüm istatistiksel özellikleri korumak için bu değişkenin tamamı da silinmiştir.

Tablo 3 Eksik veri yüzdeleri

Sütun Adı	Toplam Eksik Değer	Eksiklik Yüzdesi
Ekonomi (kg/nm)	37946	35.23
Downstream (kts)	107716	100

3. **Yerine Koyma**: Yerine koyma, bir veri doldurma yöntemidir ancak daha karmaşık yapılara sahiptir. Temel olarak, yerine koyma, tekli ve çoklu değer yerine koyma olarak iki kategoriye ayrılabilir:

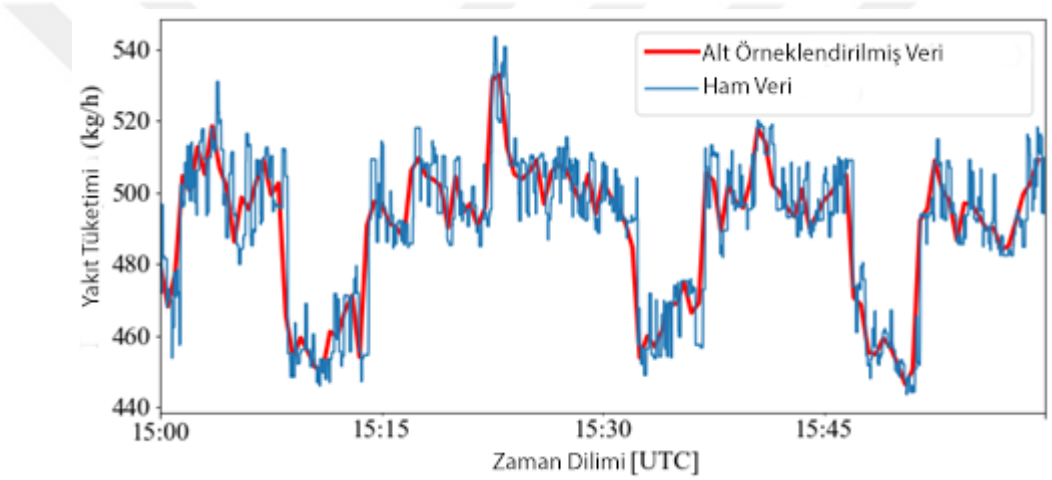
   a. **Tekli Değer Yerine Koyma**: Veride yalnızca bir değer için atama yapılır. Eksik değer, bir sütunun ortalama değeri veya bir regresyon modeli ile değiştirilebilir.

      1. **Ortalama değer**: Modelde sapmaya neden olabilir.
      2. **Regresyon modeli**: XGBoost en çok kullanılan regresörlerden biridir. Mevcut değerler bağımsız değişkenlerdir ve eksik değerler, bunlara uyması için hedef değişkenlerdir. Bu yöntem, sapma etkisi için modele daha az zarar verir.

   b. **Çoklu Değer Yerine Koyma**: Bir veri kümesindeki birden çok değer için yerine koymadır. Tekli değer yerine koymadaki regresyon modeli yinelemeli olarak da kullanılabilir. Veri kümesinde, Python diline ait scikit-learn kütüphanesinde bulunan bir yinelemeli yeri koyma (iterativeimputer) yöntemi aracılığıyla çoklu değer yerine koyma yöntemi kullanılır.

3.1.2.2. Yeniden Örnekleme

Orijinal veri kayıt örnekleri 1 saniye periyoduna sahiptir. Ancak bu zaman aralığı, regresyon modellerinin ve hiperparametre optimizasyonunun daha uzun süre çalışmasına neden olmaktadır. Ayrıca, alt örnekleme aykırı değerlerin azaltılmasına yardımcı olabilir. Böylece, regresyon uyumu problemlerinden büyük ölçüde kaçınılır. Bu nedenlerle, tüm ham veriler, ortalama alma yöntemiyle 30 saniyelik yeniden örneklenmiştir. Şekil 5'te, yeniden örneklenmiş yakıt tüketimi verilerinin orijinal sinyalin trendini başarıyla takip ettiği gösterilebilir. Ayrıca, bu yeniden örnekleme yöntemi, ham verilerin gürültüsünü ve aykırı değerleri önemli ölçüde ortadan kaldırır.



Şekil 5. Yakıt tüketimi ham verilerinin 30 saniye ile örneklemesi

3.1.2.3. Aykırı Değer Tespiti

Aykırı değerler veya anomaliler, regresyon eğri uydurmada birçok soruna neden olabilir. Aykırı değerler, varyans hatasıyla doğru orantılıdır. Birçok aykırı değere sahip regresyon uydurma işlemi, RMSE değerini zorlukla azaltabilir ve yürütme zamanını uzatabilir. Ayrıca, fiziksel bellek ve işlemci kaynağı tüketimi önemli ölçüde artar. Ayrıca, regresyon uydurma sırasında hata metriğine fazladan sapma (hata) eklenir.

Öte yandan, çok değişkenli zaman serilerinde bir aykırı değerin doğrudan kaldırılması diğer değişkenlerden de veri kaybına neden olmaktadır. Eğer aykırı değerlerin sayısı önemli bir miktarda ise, regresyon analizi artık doğru tahmin yapamaz. Literatürde aykırı değeri silmek yerine değişkenin ortalama değerini koymak daha yaygındır; ancak bu yöntem popülasyonun dağılımını bozabilir ve çeşitliliğini azaltabilir [18].

Literatürde birçok anomali tespit yöntemi bulunmaktadır. Kısaca aykırı değer tespit algoritmaları girdi ve aykırı değer türlerine göre değişmektedir. Bu tezde sadece çok değişkenli zaman serisi giriş veri tipi kullanılmıştır. Nokta şeklinde aykırı değerler genellikle değişkenin kendisinin yanlış davranışlarına bağlıdır. Bağlamsal aykırı değerler ise, diğer değişkenlere daha fazla bağımlıdır. Çok değişkenli zaman serisi verileri birden fazla değişkenden oluştuğu için, uzun süreli anomaliler genellikle birkaç veya daha fazla değişken tarafından oluşturulur. Ve bazen, tüm zaman serisi verileri anormallikler içerebilir [18].

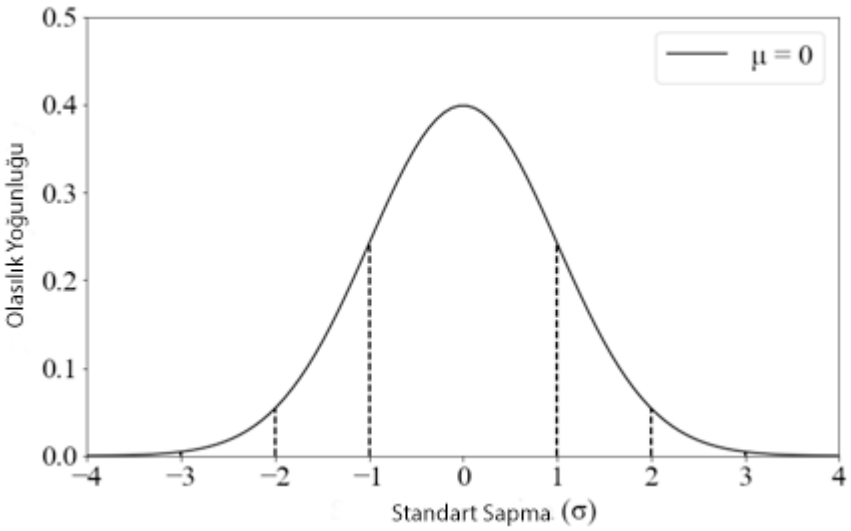
Bu tezde, aşağıda listelenen tek değişkenli veriler için iki aykırı değer saptama tekniği kullanılmıştır:

- Z-skoru
- Çeyrek Açıklığı (IQR)

Z-skoru, kullanılan veri kümesinde kaç tane standart sapmanın geçerli olacağını belirler. Ancak veri kümesinin bir Gauss dağılımına uyduğunu da varsayar. Denklem 1'de, Z-skoru, gözlenen (x) ve ortalama değerler (μ) arasındaki farkın standart sapmaya (σ) bölünmesiyle hesaplanır.

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

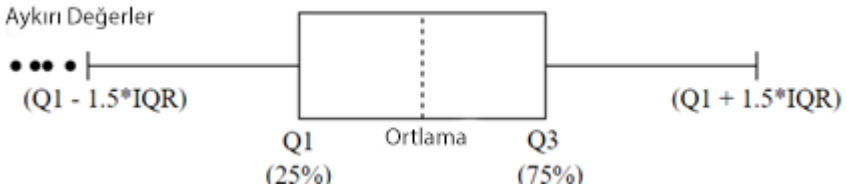
Şekil 6'da, dağıtım altındaki alan yüzde cinsinden veri miktarını temsil eder. Standart sapma azaldıkça aykırı değerlerin sayısı artar.



Şekil 6. Standart normal dağılım

IQR yöntemleri gauss dağılımını kullanmaz. Bu nedenle, parametrik olmayan verilere de uygulanabilirler. IQR değerlerinin genel bir kutu grafiği görünümü Şekil 7'de görülebilir. Bu

şekilde, Q1, medyan ve Q3, verilerin %25, %50 ve %75'lik dilimlerini temsil eder. IQR'deki en küçük değer, Q1'den 1,5 kat daha azdır. En büyük değer, Q3'ten 1,5 kat daha büyüktür. IQR grafikleri çoğunlukla veri analizi projelerinde aykırı değerleri belirginleştirmek için kullanılır.



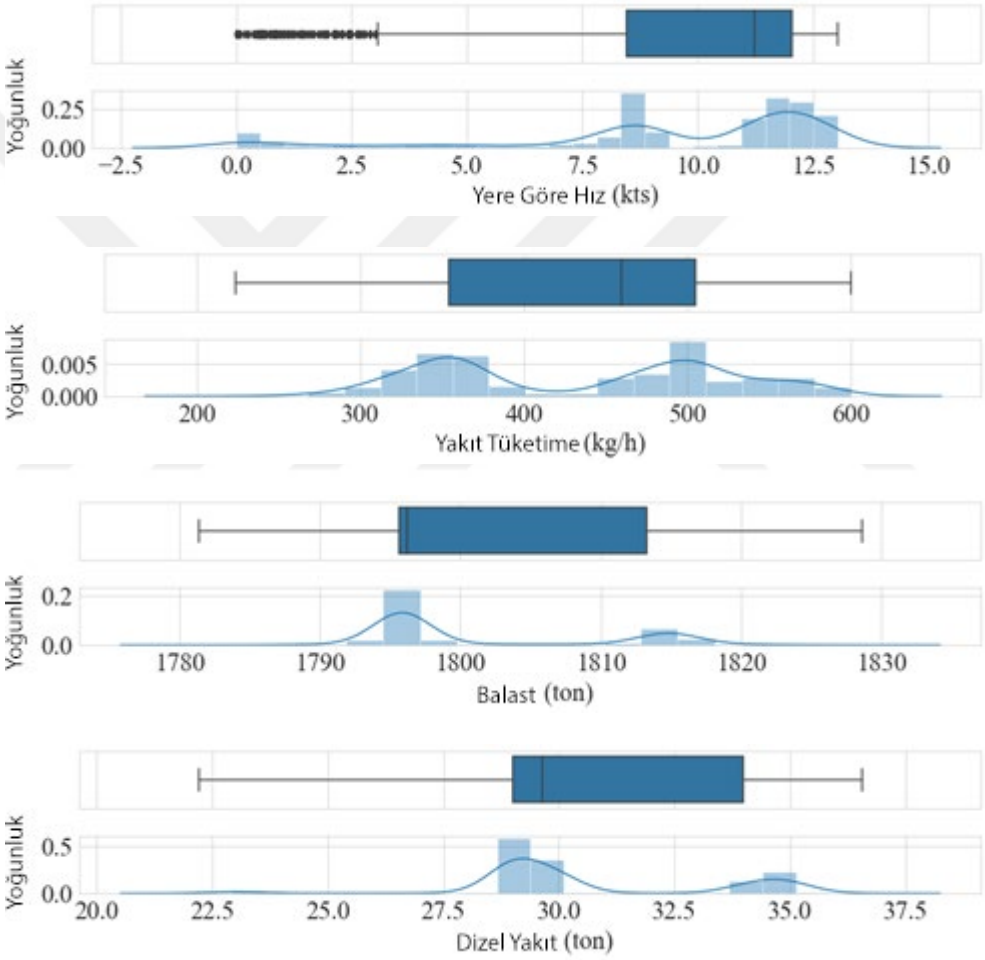
Şekil 7. IQR ile genel bir kutu grafiği gösterimi

Yakıt tüketimi verilerinin aykırı değerleri Şekil 8 kutu grafiğinde görülebilir. 100 kg/h yakıt tüketimi değeri civarında az sayıda aykırı değer bulunmaktadır. Aykırı değerlerin sayısını bir kutu grafiği gösteriminden öğrenmek mümkün değildir. Bu nedenle yakıt tüketim sütununa standart sapma (σ) veya 2 eşik değeri ile Z-skoru yöntemi uygulanabilir. Sonuç olarak, Z-skoru yöntemi uygulaması ile 20 adet aykırı değer bulunmuştur (bkz. Tablo 4).

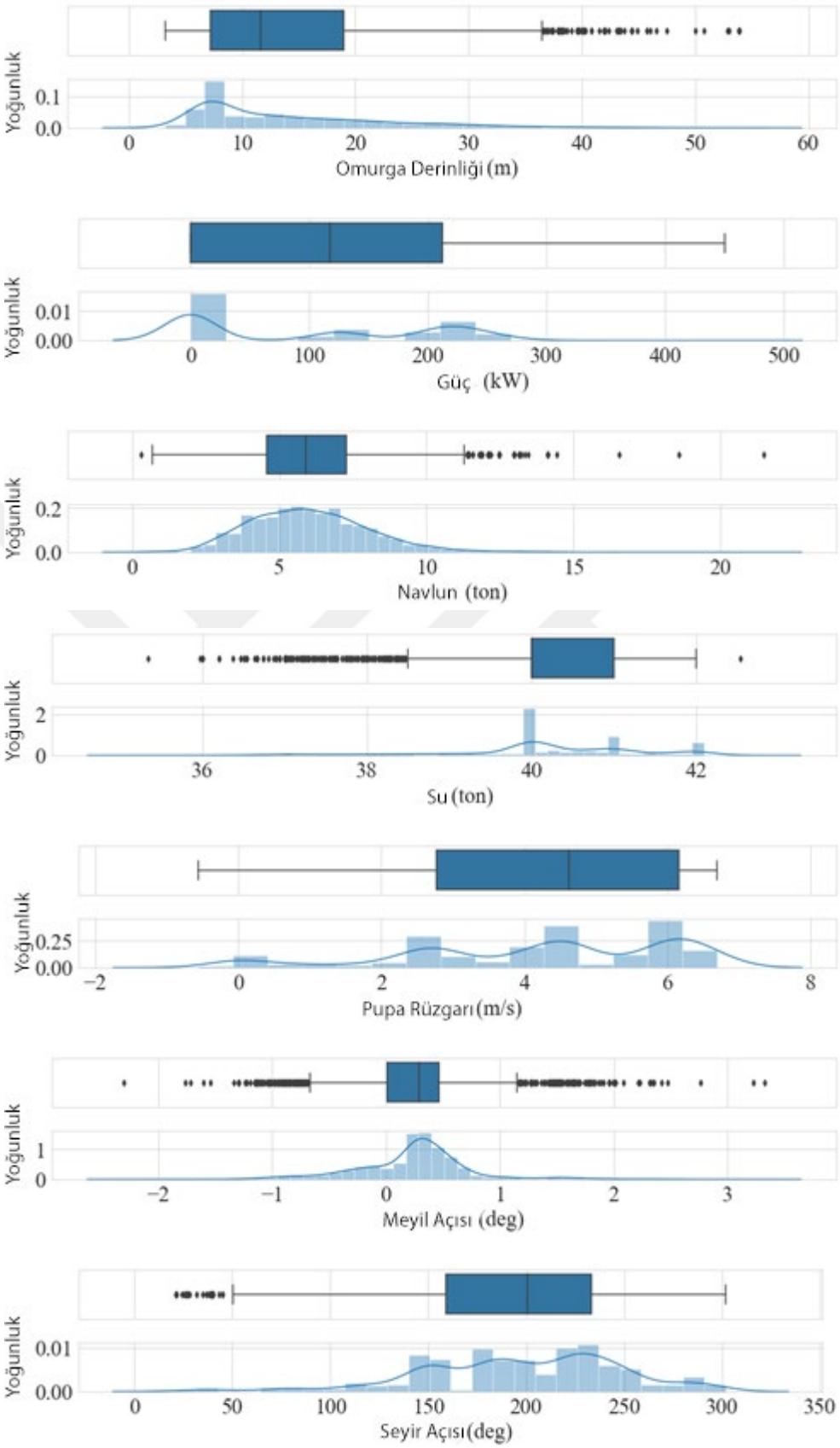
Tablo 4 Yakıt tüketimi IQR yöntemi aykırı değerleri (σ = 2)

	Zaman Dilimi [UTC]	**Yakıt Tüketimi (kg/h)**
1	2021-02-10 09:49:30	124.41
2	2021-02-10 09:50:00	127.62
3	2021-02-10 09:50:30	176.86
4	2021-02-10 09:51:00	164.00
5	2021-02-10 09:51:30	148.87
6	2021-02-10 09:52:00	151.11
7	2021-02-10 09:52:30	159.08
8	2021-02-10 09:53:00	224.50
9	2021-02-10 09:53:30	223.69
10	2021-02-10 09:54:00	184.45
11	2021-02-11 05:57:00	247.81
12	2021-02-11 05:57:30	179.67
13	2021-02-11 05:58:00	127.06
14	2021-02-11 05:58:30	95.58
15	2021-02-11 05:59:00	99.36
16	2021-02-11 05:59:30	109.13
17	2021-02-11 06:00:00	98.78
18	2021-02-11 06:00:30	88.42
20	2021-02-10 09:49:30	124.41

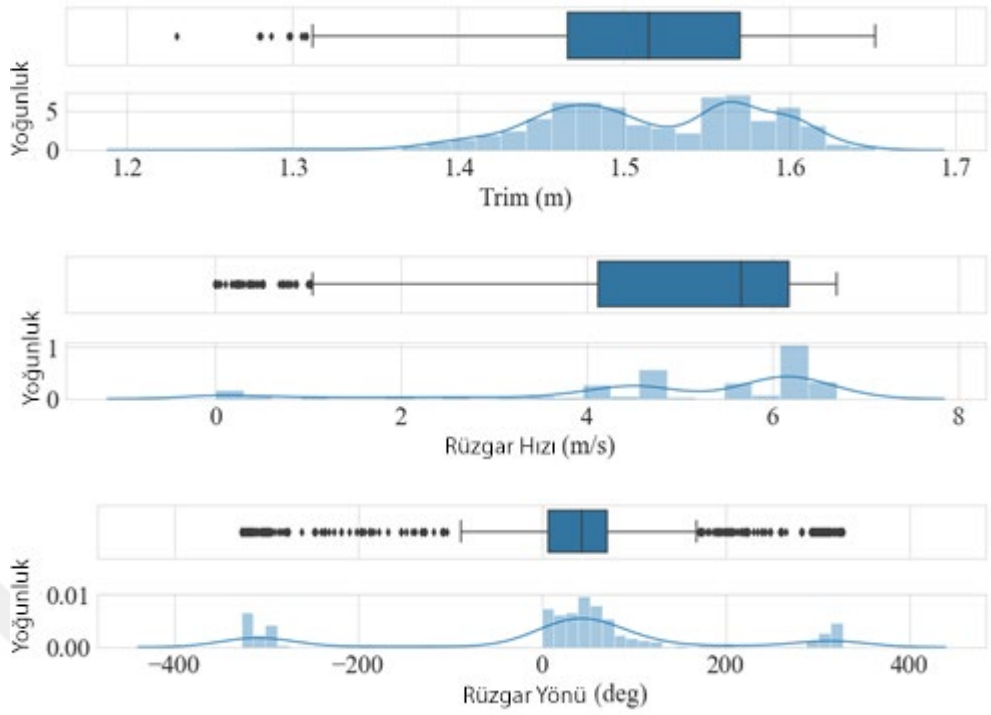
Geri kalan tüm değişkenlerin kutu grafiği Şekil 9 ve 10'da görülebilir. Ayrıca, aynı standart sapma değeri 2'ye göre hemen hemen hepsinin çok sayıda aykırı değer içerdiği kolayca fark edilebilir. Ayrıca, değişkenlerin çoğu, çok sayıda aykırı değer nedeniyle önemli miktarda çarpıklığa sahiptir. Ayrıca Z-skoru yönteminin çok değişkenli bir zaman serisine sahip olması nedeniyle diğer değişkenlere ayrı ayrı uygulanması büyük veri kayıplarına neden olacaktır. Ve azalan veriler, gerçek dünya verilerinin istatistiksel özelliklerinin kaybolmasına neden olabilir. Veri kaybını önlemek için, tüm değişkenler kullanılarak çok değişkenli bir anormallik (aykırı değer) tespit algoritması uygulanmıştır.



Şekil 8. Regresyon özniteliklerinin histogramları ve IQR'leri-1



Şekil 9. Regresyon özniteliklerinin histogramları ve IQR'leri-2

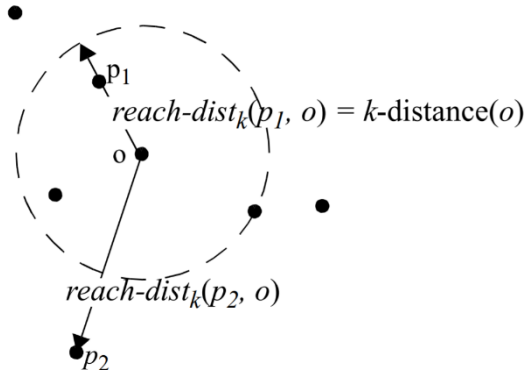


Şekil 10. Regresyon özniteliklerinin histogramları ve IQR'leri-3

3.1.2.4. Yerel Aykırı Değer Faktörü Algoritması

Yerel Aykırı Değer Faktörü (Local Outlier Factor - LOF) algoritması, bir aykırı olma derecesi tanımlayarak aykırı değerleri bulur. Kısaca, LOF algoritması, noktanın komşularıyla karşılaştırmak için bir yoğunluk fonksiyonu kullanır. LOF algoritmasının temel varsayımı, aykırı değerlerin yoğunluğunun, aykırı olmayanlara göre komşuların yoğunluğundan çok daha farklı olduğu için, yoğunluk k-en yakın komşulara ve bunların k-mesafesine bağlıdır.

Şekil 11, LOF algoritmasını açıklamak için gösterilmiştir. K-mesafe o noktasından dairesel bir çizgiye olan mesafe olarak tanımlanırsa, bu dairesel bölge içindeki tüm nesnelerin ulaşılabilirlik mesafeleri k-mesafedir. Örneğin, p_1 noktasından o noktasına olan normal mesafe k-mesafesinden daha düşük olmasına rağmen, ulaşılabilirlik mesafesi $reach - dist_k(p_1, o)$,k-mesafe(o)'ye eşittir (bkz. Şekil 11). Aksi takdirde, p_2 noktasının ulaşılabilirlik mesafesi doğrudan o noktasına olan normal mesafesine eşittir. Ayrıca bu daire içindeki tüm noktaların sayısı k-komşu sayısı $N_k(p)$ olarak tanımlanır.



Şekil 11. Ulaşılabilirlik mesafeleri [19]

Ayrıca, noktaları karşılaştırmak için LOF'ta başka bir terim olan yerel erişilebilirlik yoğunluğu (lrd) kullanılır. Komşularının ortalama erişilebilirlik mesafelerinin toplamının tersi olarak ifade edilen yerel erişebilirlik yoğunluğu Denklem 2'de formüle edilmiştir.

$$lrd_k(p) = \left[sum(reach - dist_k(p,o) / N_k(p) \right]^{-1} \quad \| \forall p, o \in D \tag{2}$$

Son olarak, yerel aykırı değer faktörü, yerel erişilebilirlik yoğunlukları ve komşu nokta sayısı kullanılarak tanımlanabilir. Denklem 3'te, p noktasının ulaşılabilirlik yoğunluklarının komşu sayısına oranının ortalaması olarak tanımlanmaktadır.

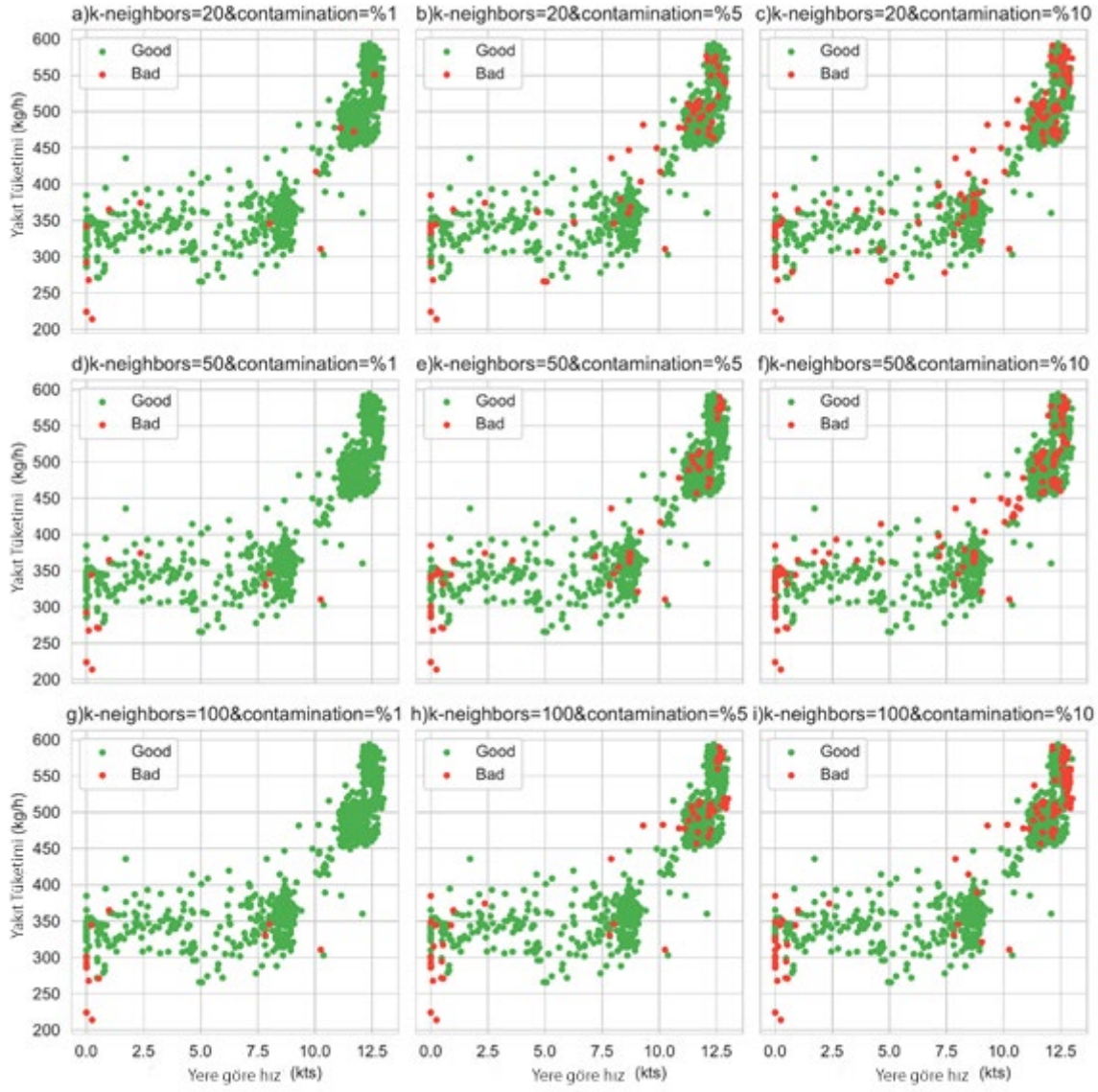
$$LOF_k(p) = \frac{\sum \frac{lrd_k(o)}{lrd_k(p)}}{N_k(p)} \tag{3}$$

Sonuç olarak, eğer komşu noktaların yerel erişilebilirlik yoğunlukları p noktasının ulaşılabilir yoğunluğundan yüksekse, o zaman *p* noktası için yerel aykırı değer faktörünün daha yüksek olacağı açıktır. Bu, daha düşük yerel aykırı değer faktörlerinin, aykırı değer olarak atanmasına daha yakın olduğu anlamına gelir. Başka bir deyişle, yalıtılmış noktaların yerel yoğunlukları doğası gereği komşu noktalardan daha düşük olacak ve LOF algoritmasına göre daha düşük hesaplanacaktır.

Bu çalışmada, LOF algoritmasını uygulamak için scikit-learn kütüphanesi kullanılmıştır. Tablo 2'de listelenen tüm gemi değişkenleri, önerilen yöntem uygulandığında kullanılır. Varsayılan mesafe ölçüsü Minkowski olup Minkowski metriği için yine varsayılan Öklid uzaklığı yöntemi seçilmiştir. En yakın komşuları aramak için otomatik bir yöntem seçim yapılandırması etkinleştirilir. Regresyon uydurma yöntemine iletilen değerlere dayalı olarak BallTree, KDTree veya Kaba Kuvvet arama algoritmalarından birini otomatik olarak seçecektir (daha ayrıntılı

parametre ayarları için Ek A'ya bakınız). Ayrıca, k-komşular ve tahmin edilen uç değer orantı parametreleri tek tek değiştirilerek belirlenir. Bu amaçla, 20, 50 ve 100 komşu sayısı (k-komşu) ile tahmini kontaminasyon oranlarının yüzde 1, 5 ve 10'u sırasıyla ve diğer tüm değişkenler sabit tutulmak koşuluyla yinelemeli olarak birlikte uygulanır. Son olarak, LOF yönteminin sonuçları Şekil 12'de dağılım grafikleri olarak gösterilmektedir. Kırmızı renkli noktalar aykırı değerleri göstermektedir. 14 giriş değişkeni LOF yöntemi için girdi olarak kullanılmasına rağmen, dağılım grafiklerinde yalnızca yakıt tüketimi ve hız değişkenleri gösterilmektedir. Hız değişkeni bu tezin hız optimizasyonu bölümünde bir kontrol parametresi olarak kullanıldığından, ortaya çıkan aykırı değerlerin 14 parametre tarafından belirlendiğini tekrar vurgulamakta fayda vardır. Bu nedenle, komşulardan aşırı derecede izole edilmiş gibi görünen ancak aykırı değerler olduğu için sonuçlanmayan birçok yeşil nokta olması normaldir.

Gemi değişkenleri genellikle bir ay veya daha uzun zaman aralıklarında farklı özellikler gösterebilir. Ancak kullanılan veri kümesinin zaman aralığı 1 gün ile sınırlıdır. Bu nedenle ham veri kümesinden daha fazla veri dolayısıyla da istatistiksel özellik kaybını önlemek için kirlilik oranı %5 seçilerek en fazla %5'lik veri kaybı yapılmıştır. Ayrıca artan kirlilik oranının normal yeşil alanlar arasında birçok aykırı değerin yanlış çıkmasına neden olduğu Şekil 12'den izlenebilir. Öte yandan, sıfır hız bölgelerinde birçok aykırı değer ile artan k-komşu sonuçları daha arzu edilir. Bu nedenle, k-komşuluk ve kirlilik parametreleri sırasıyla 50 ve 5 olarak seçilmiştir, bu seçim Şekil 12'deki alt grafikte (h) görülebilir.

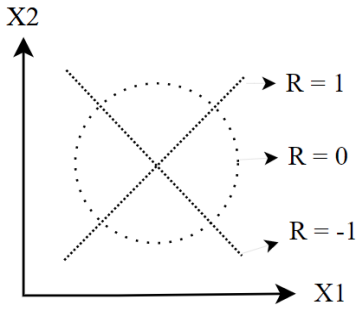


Şekil 12. LOF methodu ile aykırı değer sonuçları (farkı k-komşuluk and kirlilik oranlarında)

3.1.3. Öznitelik Seçimi

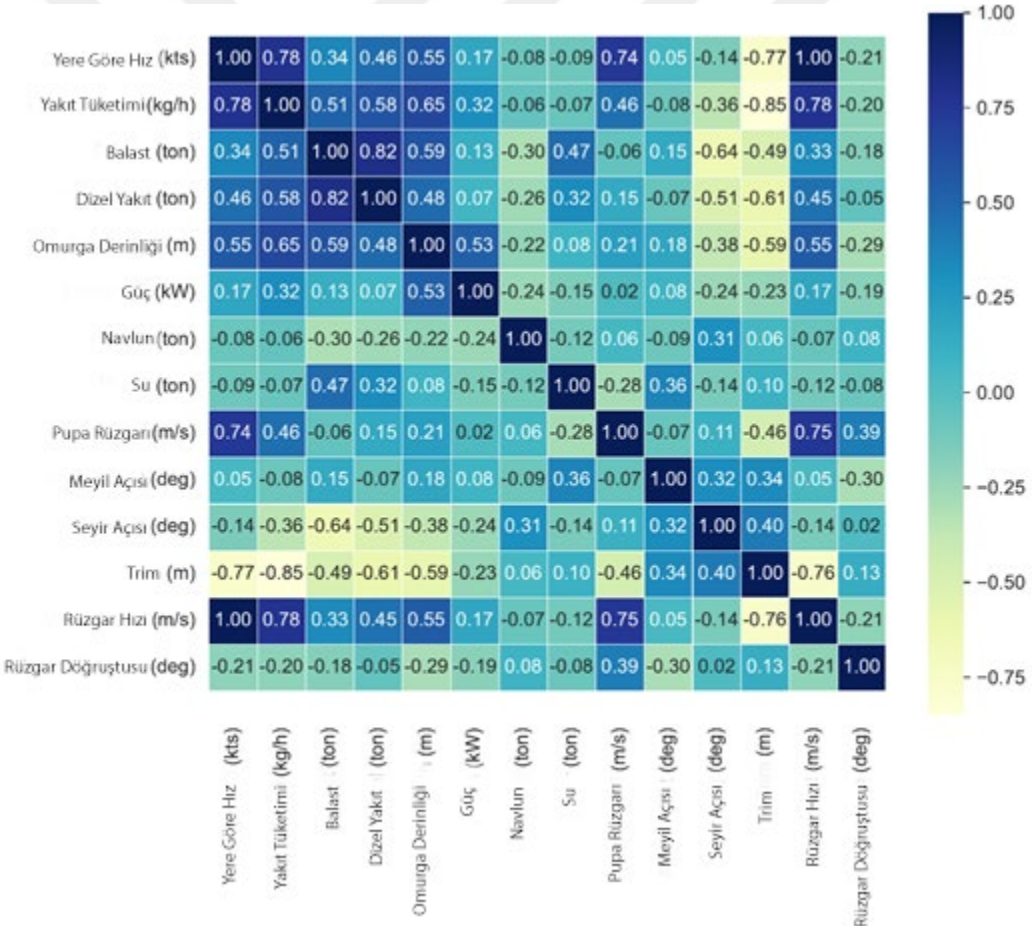
Yakıt tüketimini tahmin etmek için tüm özniteliklerin kullanılması fazla gürültüye ve uzun hesaplama süresine neden olabilir. Ayrıca çok faz kaynak gerektirebilir. Ancak, gereksiz öznitelikleri belirlemek için bir korelasyon analizi gereklidir [20]. Korelasyon analizi ile, yüksek korelasyonlu öznitelikler, tahminciler açısından tek bir değişken olarak kabul edilir [21]. Korelasyon katsayıları (R), yalnızca iki değişken veri kümesi arasındaki doğrusal ilişkileri açıklar. R-değeri 1 ile -1 arasında normalleştirilir. Şekil 13'ün yardımıyla, R'nin 1'e eşit olması durumunda, X1 ve X2'nin tamamen birlikte arttığı veya azaldığı mükemmel bir pozitif doğrusal ilişki vardır. R, -1'e eşitse, X1 ve X2 arasında tamamen zıt yönlerde artan veya azalan

mükemmel bir negatif doğrusal ilişki vardır. Ve R'nin 0'a eşit olması, iki değişken veri kümesi arasında doğrusal bir ilişki olmadığı anlamına gelir [22].



Şekil 13. X1 and X2 serpilme diyagramı üzerinde korelasyon katsayıları

Dolayısıyla Şekil 14'te gösterilen korelasyon matrisi de kare şekilli ve simetrik bir matristir.



Şekil 14. Öznitelikler arasındaki Pearson korelasyon katsayıları

Regresyon parametrelerinin bağımsız değişkenleri arasındaki mevcut doğrusal bağımlılık, regresyon katsayılarının yanlış önemde belirlenmesine ve dolayısıyla regresyonun yanlış

tahminde bulunmasına neden olabilir [7]. Şekil 14'te ilk bakışta geminin yere göre hızı ile rüzgar hızının korelasyon katsayısının 1 olduğu dikkat çekmektedir. Dolayısıyla bu değişkenler arasında bir doğrusal bağlantı mevcuttur. Ancak, bu tezde yere göre hız, hız optimizasyonu için bir kontrol parametresi olarak kullanılacaktır. Bu nedenle onun yerine rüzgar hızı parametresi regresyon analizinden çıkarılmıştır.

3.2. MODELLEME ÇALIŞMALARI

3.2.1. Model Performans Değerlendirme Metrikleri

Tahmin modellerinin performans değerlendirmesi, gerçek ve tahmin edilen değerlerin karşılaştırılması yoluyla yapılır. Bu çalışmada aşağıda belirtilen üç hata metriği kullanılmıştır.

a. Kök Ortalama Kare Hata (RMSE)

RMSE, temel maliyet parametrelerinden biridir. Başka bir deyişle, tahmin edilen ve gözlemlenen değerler arasındaki farkların ortalamasının karaköküdür (bkz. Denklem 4). Regresyon modelleri, her iterative hesaplamada RMSE'yi iterative olarak ölçer ve sezgisel olarak azaltmaya çalışır.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{\hat{y}_i - y_i}{n}} \tag{4}$$

b. Determinasyon Katsayısı (R^2)

R^2 veya doğruluk metriği, bir regresyon modelinin gerçekten gözlenen yakıtı nasıl tahmin edebileceğini açıklar.

c. Çalışma Süresi

Regresyon modellerinin çalışma süresi karşılaştırmaları da bu çalışmada kullanılmıştır. Tüm modeller farklı doğruluk ve RMSE değerlerine sahip olsa da çalışma süreleri düşüktür. Daha büyük veri kümeleri üzerinde çalışırken çalışma süresi önemini arttırmaktadır. Kaynaklar, artan bir çalışma süresinde kolayca aşırı yüklenebilir ve çekirdek durdurma hatalarına neden olabilir.

3.2.2. Regresyon Analizi İçin Doğrusal Metodlar

3.2.2.1. Çoklu Doğrusal Regresyon Varsayımları

Çoklu doğrusal regresyon parametrik bir analizdir. Tüm değişkenlerin halihazırda 5 belirli özelliğe sahip olduğu varsayılmaktadır [23]. Bunlar aşağıda listelenmiştir.

1. Bağımlı ve bağımsız değişkenler arasında doğrusal ilişkiler vardır. Korelasyonun büyüklüğü, lineer regresyonun açıklanabilirliğini ve hatasını belirler.
2. Bağımsız değişkenler arasında korelasyon yoktur. İdeal olarak, R değerlerinin büyüklüğü 0 olmalıdır, ancak bu sık rastlanan bir durum değildir.
3. Hata terimlerinin arasında bir ilişki yoktur. Başka bir deyişle, otokorelasyon yoktur.
4. Hata terimleri sabit varyansa sahiptir. Başka bir deyişle, değişken varyans yoktur.
5. Hata terimleri normal dağılımlıdır.

Doğrusal regresyon modellerini kullanmadan önce, yukarıdaki varsayımların görsel testlerle kanıtlanması gerekmektedir. Bu nedenle sonuçlar bölümde bazı testler maddeler halinde verilmiştir.

3.2.3. Doğrusal Olmayan Metodlar

Doğrusal regresyon sonuçları yeterli doğruluk ve RMSE değerleri göstermeyebilirler. Bu nedenle, doğrusal regresyon modelleri kullanan yakıt tüketimi tahminleri, mevcut veri kümesine bağlı olarak güvenilir ve uygulanabilir olmayabilir. Bu nedenle, özellikler üzerinde normal dağılım varsayımları yapmak yerine, daha doğru tahmin sonuçları için parametrik olmayan yöntemler kullanılabilir. Bu çalışmada Destek Vektörleri, KNN, Rastgele Orman, AdaBoost ve Gradyan Arttırma regresyon modelleri uygulanmaktadır. Ayrıca, hiperparametrelerini ayarlamak için Grid Arama, Rastgele Arama ve Bayesyen Arama yöntemleri kullanılmıştır. Ayrıca, hız optimizasyonunda en iyi modeli belirlemek için modellerin eğitim ve test veri kümesi performans sonuçları sonuç bölümünde karşılaştırılmıştır.

Veri kümesi, bir Scikit-learn veri ayırıcı yöntemiyle eğitim ve test alt kümelerine ayrılmıştır. Veri kümesi, aşırı öğrenmeyi önlemek için bölmeden önce karıştırılmıştır ve test veri kümesi hiçbir zaman model eğitimi için kullanılmamıştır.

3.2.3.1. Hiperparametre Optimizasyon Teknikleri

Grid Arama (Grid Search - GS), tüm girdi uzayını araştıran ve aynı zamanda n olası değer $O(n^k)$ ile k hiperparametre uzayının kartezyen çarpımının hesaplama karmaşıklığına sahip bir kaba kuvvet yöntemidir. Bu nedenle, GS ile daha fazla olası parametre değeri aramak, hesaplama maliyetini üssel kuvvetle artırır. Ancak bu çalışmada KNN, SVR, ADA ve GBR modeller için GS kullanılmıştır.

GS'den farklı olarak, Rastgele Arama (Random Search - RS), tüm girdi uzayını aramaz. Ancak, kullanıcı tarafından tanımlanan yineleme sayısı kadar yalnızca rastgele seçilen değerleri

kullanır [23]. Yinelemeli sayıları artırmak, model için daha iyi sonuçlar verir; ancak, hesaplama maliyetleri de artar. Bu çalışmada RS, RF, XGB ve GBR modelleri için kullanılmıştır.

Hem GS hem de RS yöntemlerinin ana dezavantajı, önceki değerlendirme sonuçlarını hatırlamadan değerlendirmedir [24]. Ancak Bayes yöntemleri, sonraki giriş değerleri için hiperparametrelerin geçmiş değerlendirme sonuçlarını kullanır. Bu çalışmada hiperparametre optimizasyonu gerçekleştirmek için Bayes Optimizasyon tekniklerinden biri olan Python HyperOpt kütüphanesi kullanılmıştır. Bu çalışmada, RF modeli için RS kullanılmıştır.

Bu çalışmada kullanılan hiperparametre değer arama aralıkları ve atanan ayarlayıcı tarafından bulunan optimal değerler Tablo 5'te listelenmiştir.

Tablo 5. Hiperparametreler ve optimizasyonun koşma süreleri

Model	Hiperparametreler (Scikit-learn kütüphanesi)	Aralık	Optimum Değer	Eğitim Süresi (min.)
LR	Hiperparametre opt. yok			<1
SVR	Hiperparametre opt. yok			<1
	Tuner: GridSearchCV			<120
	C	[0.1,1,10,100,1000]	1000	
	gamma	[1,0.1,0.01,0.001,0.0001]	0.1	
	kernel	['rbf','poly','sigmoid','linear']	rbf	
	degree	[1,2,3]	1	
KNN	Hiperparametre opt. yok			<1
	Tuner: GridSearchCV			<1
	Komşuluk derecesi	[5,7,9,11,13,15,20,30,50]	5	
	Ağırlıklar	['uniform','distance']	distance	
	Ölçüm	['minkowski','euclidean', 'manhattan']	manhattan	
ADA	Hiperparametre opt. yok			<1
	Tuner: GridSearchCV			<1
	Tahmin edici sayısı	[2,3,4,5,7,8,9,10,20,50, 100,300]	50	
	Öğrenme oranı	[0.97,0.98,0.99,1,1.01,1.02,1.06]	1.06	
XGB	Hiperparametre opt. yok			<1

	Tuner: RandomizedSearchCV(iteration:200)			<2
	Min. ağırlık	[1,3,5,7,10,15]	10	
	gamma	[0.1,0.5,1,1.5,2,5,8,15]	1	
	örnekleme	[0.2,0.6,0.8,1.0,1.5]	0.8	
	colsample_bytree	[0.2,0.6,0.8,1.0,1.5]	1	
	Max. derinlik	[3,4,5,7,10,15]	15	
	reg_alpha	[50,70,100,120,150]	50	
	Tahmin ediciler	[100,180,300,1000,1500]	180	
RF	Hiperparametre opt. yok			<1
	Tuner: RandomizedSearchCV(iterasyon sayısı:200)			<2
	Tahmin edici sayısı	[5,20,50,100,300,450,500]	300	
	Maks. özellik	['auto', 'sqrt']	auto	
	Maks. derinlik	np.linspace(10,120,num=24)	10	
	Min. düğüm örneği	[2, 4, 6, 8, 10]	2	
	Min. yaprak örneği	[1, 2, 3, 4, 5]	2	
	Tuner: HyperOpt			<15
	Tahmin edici sayısı	uniform(100,1000)	469	
	Maks. derinlik	uniform(5,120)	16	
	Min. düğüm örneği	uniform(1,5)	2	
	Min. yaprak örneği	uniform(2,10)	4	
	Maks. özellik	['auto','sqrt','log2',None]	2	
GBR	Hiperparametre opt. yok			<1
	Tuner: GridSearchCV			<5
	Tahmin edici sayısı	sample(200,1100,step:50)	1050	
	Maks. özellik	['auto', 'sqrt']	auto	
	Maks. derinlik	sample(4, 16, step:2)	12	
	Min. düğüm örneği	sample(2, 20, step:1)	2	
	Min. yaprak örneği	sample(5, 61, step:5)	5	
	Tuner: RandomizedSearchCV(iterasyon sayısı:200)			<5
	Tahmin edici sayısı	rand(low:100, high:1200)	460	
	Maks. özellik	rand(low:5, high:20)	8	
	Maks. derinlik	rand(low:2, high:30)	9	
	Min. düğüm örneği	rand(low:2, high:100)	55	

Min. yaprak örneği	rand(low:2, high:25)	22
Öğrenme hızı	rand(low:0, high:1)	0.0425
Alt örneklem	rand(low:0, high:1)	0.823

3.2.3.2. Destek Vektör Regresyonu

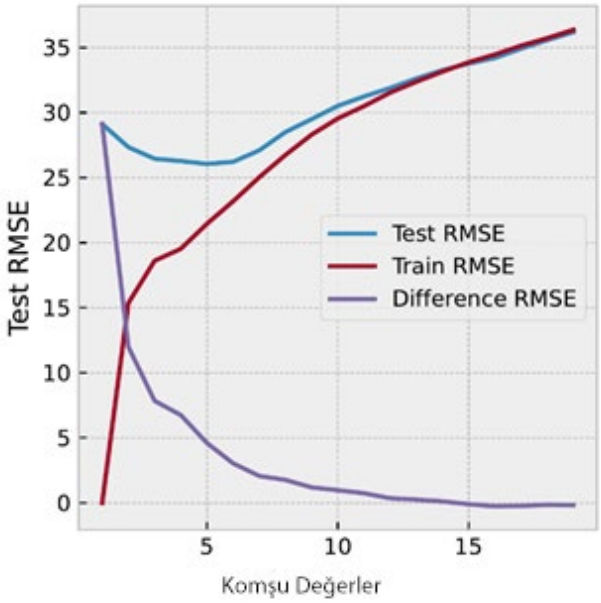
Destek Vektörü Regresyonu (Support Vector Regression - SVR), destek vektör makinelerinin (Support Vector Machines - SVM) bir uygulamasıdır ve aynı zamanda denizcilik literatüründe gemilerin yakıt tüketimi regresyon problemini çözmek için popüler bir araçtır [25]. Kelleher, Mac Namee [26] SVR'nin güçlü yanlarını, "hızlı eğitilebilirler, aşırı öğrenmeye meğilli değildirler ve yüksek boyutlu veriler için iyi çalışırlar" şeklinde tanımlamaktadır.

Bu çalışmada, SVR modelini gerçekleştirmek için Scikit-learn destek vektör makineleri (sklearn.svm) modülünün destek vektör regresyon yöntemi kullanılmıştır. Ayrıca, bu çalışmada aşağıdaki tüm topluluk modellerini uygulamak için Scikit-learn topluluk modülü de kullanılmıştır.

3.2.3.3. K-En Yakın Komşuluk Modeli

K-En Yakın Komşular (K-Nearest Neighbors - KNN), noktaların komşuluk kullanılarak sınıflandırılmasına dayalı parametrik olmayan bir modeldir ve eğitimde geçmiş verileri kullanan diğer modellerden farklıdır [20]. KNN algoritmaları, veri kümesinin dengesiz veya çok büyük olduğu durumlarda düşük performans gösterir. Çünkü KNN, veri kümesindeki büyük bir grup lehine hedefi kolayca belirleyebilir veya büyük veri setlerinde mesafe hesaplama maliyeti çok artar [26].

Bu çalışmada, KNN algoritmasını gerçekleştirmek için Ek-A'da gösterilen Scikit-learn kütüphanesinin komşu sayısı dışındaki varsayılan parametreler kullanılmıştır. Aşırı öğrenmeyi önlemek için komşuluk sayıları, Kare Ortalamalarının Karekökü Hatası (Root Mean Square Error – RMSE) değerleri kontrol edilerek, yinelemeli olarak ayarlanabilir. Artan komşuluk sayısıyla gerçekleştirilen iterasyonların simülasyon sonuçları Şekil 15'te gösterilmiştir. Test ve eğitim hataları arasındaki fark, yirmi komşuya ulaşılarak yeterince azaltılmıştır. Son olarak, önerilen KNN modelinde yirmi adet komşuluk seçilmiştir.



Şekil 15. Komşu değerlere karşı test hatası

3.2.3.4. Rastgele Orman Regresyonu

Rastgele orman (Random Forest - RF), karar ağaçlarına dayalıdır ve bir veri kümesinde rastgele seçilen örnekleri kullanır. Kelleher, Mac Namee [26], "alt uzay örneklemesinin topluluk içindeki karar ağaçlarının çeşitliliğini daha da teşvik ettiğini ve her ağaç için eğitim süresini azaltma avantajına sahip olduğunu" vurgulamaktadır. Bireysel karar ağaçları yüksek varyans ve aşırı öğrenme içerebileceğinden, bununla birlikte, rastgele ormandaki ağaçların rastgele seçilmesiyle hataların ortalaması düşürülebilir [27].

3.2.3.5. Adaptif Yükseltme (ADABoost)

ADABoost, ardışık yinelemelerde eğitim veri kümesinin sonraki zayıf öğrenicilerini kullanarak karar ağacı algoritmasını geliştirir [20]. Yani başka bir deyişle, önceki hatalardan ders çıkarır.

3.2.3.6. Gradyan Yükseltme Regresörü

Gradyent Yükseltme Regresörü (Gradient Boosting Regressor - GBR), önceki zayıf öğrenicilerden de öğrenir. GBR, önceki öğrenicilerin ağırlıklarını ADABoost'ta olduğu gibi değiştirmez, bunun yerine yeni tahmin edicileri önceki artık hatalarıyla eşleştirir [20].

3.2.3.7. Aşırı Gradyan Yükseltme (XGBoost)

XGBoost (XGB), kayıpları en aza indirmek için diferansiyel bir kayıp fonksiyonu ve dereceli azalma optimizasyonu kullanan bir topluluk yöntemidir.

3.2.3.8. Oylama Regresörü

Oylama Regresörü (Voting Regressor - VR) farklı makine öğrenimi regresörlerini birleştiren ve tahmin edilen değerlerin ortalamasını alan başka bir topluluk modelidir. Böylece aşırı öğrenme ortadan kaldırılabilir

3.2.3.9. Yığma Regresörü

Yığma regresörü (Stacking Regressor - SR) içindeki tüm regresyon modellerinin ayrı ayrı eğitildiği bir topluluk yöntemidir. Daha sonra seçilen nihai regresörün uyumu, diğer modellerin eğitim çıktılarına dayalıdır. Bu yöntemle, zincirdeki bir sonraki model için aşırı öğrenme önlenebilir.

3.3. GENETİK ALGORİTMALAR İLE GEMİ HIZ OPTİMİZASYONU

Hız optimizasyonu, ton-mil başına minimum yakıt tüketimiyle sonuçlanan en iyi hız profilini bulmak için değerlendirilir. Çünkü Gemi Enerji Verimliliği Yönetim Planına (Ship Energy Efficiency Management Plan - SEEMP) göre, gemi hızının, optimum hızın altında olması daha yüksek yakıt tüketimine neden olabilir [2, 28]. Ancak optimum hız en düşük hız demek değildir ve ayrıca yolculuk sırasında sabit de değildir [2, 29]. Bununla birlikte, optimum hız profili belirlemek için bir optimizasyon probleminin çözülmesini de gerektirir. Bunker fiyatları, liman varış zamanları, gecikmeler, hava ve deniz çevresel değişkenleri, piyasa değerleri vb. değişkenler gerçek hayatta rastgele değişebilen özellikler içermektedir. Bu nedenle, gemilerin hız ve yakıt tüketimi optimizasyonunu çözmek için var olan deterministik yaklaşımlar, bu rastgele değişkenlerin önceden bilindiklerini varsayarak rastgele olayları göz ardı ederler [13]. Ayrıca literatürde son yıllarda hız optimizasyonu için dışbükey veya kübik fonksiyonlar gibi deterministik yöntemler üzerinde çalışılmaktadır. Ancak, rasgele değişkenler hız optimizasyonu problemi için henüz açıkça dikkate alınmamıştır [13]. Bazı yönlerden, literatürdeki birçok çalışma, hava ve deniz koşulları veya motor parametreleri gibi yakıt tüketimini etkileyen bazı değişkenler için stokastik bir terim kullanır. Ancak, bunlar da yakıt tüketimini etkileyen bu rastgele faktörlerin normal bir dağılım izlediğini varsaymışlardır [13]. Yakıt tüketimini etkileyen daha birçok faktörün olduğunu varsayarsak, optimizasyon problemleri için sadece olasılıksal dağılımları kullanmak etkili çözümler üretmeyecektir.

Bu nedenle, optimizasyon problemlerinde rastgele değişkenlerin kullanılması esastır. Bu çalışmada, minimum yakıt tüketimini elde etmek amacıyla optimum hız profillerini aramak için

genetik optimizasyon algoritmaları kullanılmıştır. Evrimsel bir algoritmanın seçilmesinin ana nedeni, yerel bir optimal noktada yakınsama yapmaktan kaçınmak ve global optimal noktayı bulmak amaçlanmıştır.

Ayrıca, bir önceki bölümde yapılan regresyon analizi sonuçları, yakıt tüketiminin diğer parametrelere göre gemi hızına daha çok bağlı olduğunu göstermiştir. Bu nedenle hız optimizasyonunun diğer faktörlere göre daha fazla yakıt tasarrufu sağlayabileceği kolayca söylenebilir.

Bu çalışmada, ETA gecikmesi, maksimum seyir hızı ve hız ivmesi parametreleri (gemi gövde stresini azaltmak için) hız optimizasyon kısıtlamaları olarak belirlenmiştir. Ayrıca yakıt tasarrufunu desteklemek için limandan ayrılırken hızın aniden aşırı derecede artırılmasına izin verilmiştir [2].

3.3.1. Önerilen Genetik Algoritma Modeli

Bu bölümde, iki liman arasındaki tek bir rota için bir hız optimizasyonu matematiksel modeli geliştirilmiştir. Modelin notasyonları Tablo 6'da verilmiş olup Tablo 6'daki notasyonlara dayalı formülasyonlar türetilmiştir.

Tablo 6 Hız optimizasyon modelinin matematiksel terimleri

Parametre	Açıklama
ETA	The ETA (hedef port için) (h)
d	Seyir mesafesi (mile)
d^{new}	Değiştirilmiş hızın seyir mesafesi (mile)
t	Yolculuk süresi (h)
T^R	Seçilen kalan hız için ek süre (h)
V^R	Kalan hız (mile/h)
F_{total}	Toplan yakıt tüketimi (normal hız için) (kg)
F^R	Kalan hız için yakıt tüketim oranı (kg/h)
V^{new}	Değiştirilmiş hız profili

Denklem 4 değiştirilmiş hız profilinin ulaşabileceği mesafeyi temsil eder:

$$d^{new} = \sum_{i=0}^{n} V_i^{new} \times t \tag{4}$$

Değiştirilen hız profilinin mesafesi ile normal mesafe arasındaki fark, gemi işletmecisi tarafından seçilen kalan hız ile kapatılacaktır. Yeni hız profilinin kalan yolculuğu tamamlaması için gereken ekstra süre Denklem 5'te gösterilmektedir.

$$T^R = (d - d^{new}) / V^R \tag{5}$$

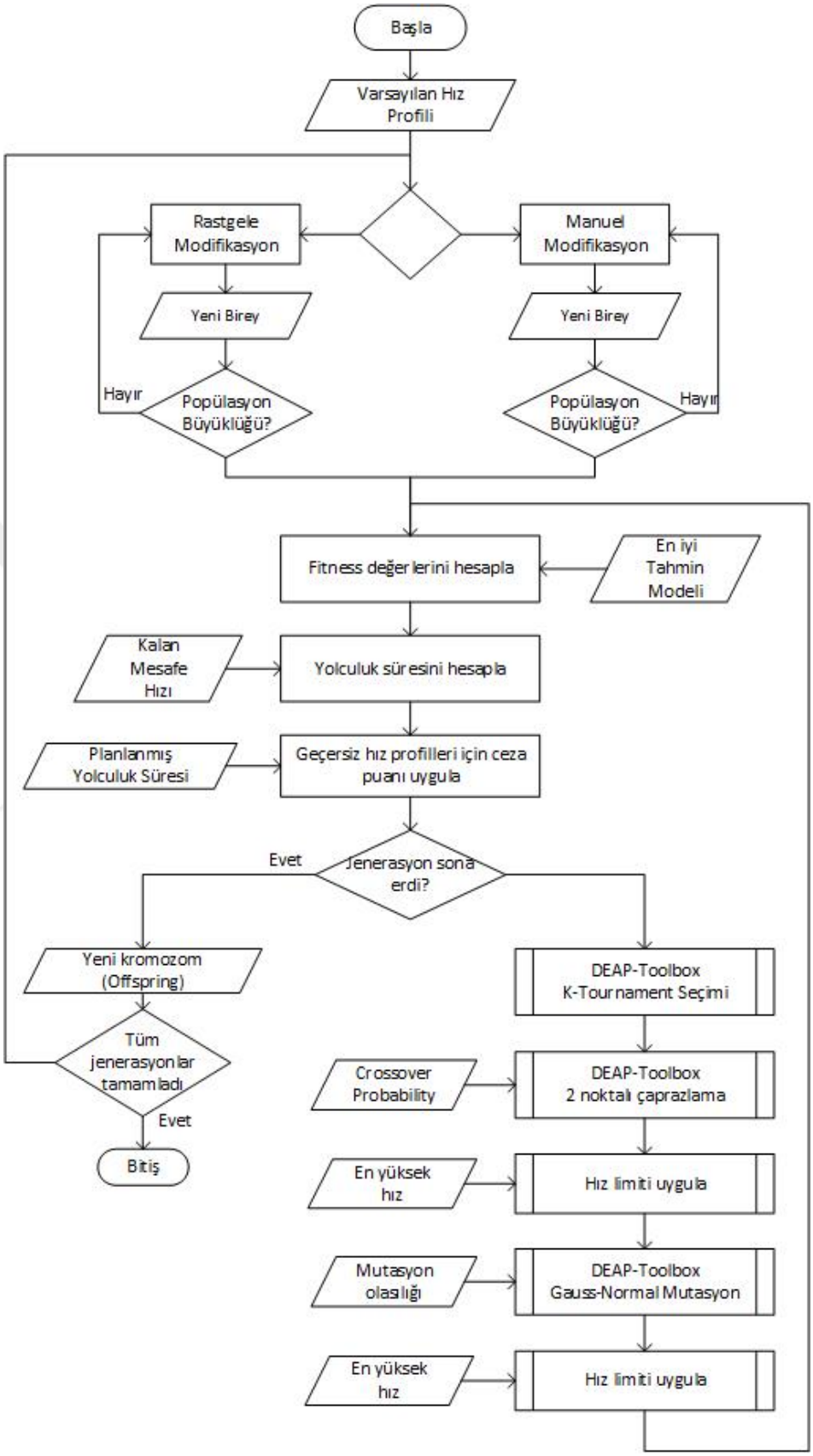
Denklem 6'da gösterilen toplam yakıt tüketimi amaç fonksiyonu, toplam yolculuk mesafesinin sonuna kadar geminin tüm seyir hızlarındaki tahmin edilen yakıt tüketimlerinin toplamıdır.

$$\min \sum F_{total} + T^R \times F^R \tag{6}$$

Denklem 6'daki amaç fonksiyonu, genetik optimizasyon algoritmasında bireylerin uygunluk değerlerini hesaplamak için kullanılır. GA'nın temel amacı, minimum uygunluk değerli bireyleri bulmaktır. Bu nedenle hız profili, gemi varış süresinin ETA'dan uzun olmamasını sağlayarak Denklem 7'deki kısıtlamaları kullanır. Ayrıca, bireysel değişiklik aşamaları için izin verilen maksimum hız limiti, bunun üzerinde aşırı hız olmamasını sağlar.

$$t + T^R < ETA \tag{7}$$

Önerilen hız optimizasyon yönteminin iş akış modeli Şekil 16'da gösterilmiştir.



Şekil 16. Hız optimizasyon probleminin akış şeması

3.3.2. Genetik Algoritma Modeli Uygulama Adımları

Bu bölümde, önceki bölümdeki, optimizasyonun matematiksel olarak uygulanmasına ek olarak algoritma analizi yapılmıştır. Bu çalışmada, genetik optimizasyon algoritması yöntemlerinin kodlanması için DEAP (Python'da Dağıtılmış Evrimsel Algoritmalar) kütüphanesi kullanılmıştır. DEAP, hız değerleri listesi ve en aza indirmek için uygunluk/amaç işlevi gibi bireysel türler oluşturmak için temel olarak oluşturucu ve çeşitli araç kutusu modüllerine sahiptir. DEAP araç kutusu modülü ise, bireylerden oluşan bir popülasyon listesi oluşturmak için genetik algoritmanın seçim, çaprazlama ve mutasyon operatörlerini sağlamaktadır. Ayrıca, araç kutusu modülü ile hız cezaları ve limit kontrol dekoratör fonksiyonları popülasyonlara işlenmiştir.

Uygulanan GA optimizasyon algoritması adımları Tablo 7'de listelenmiştir ve ayrı başlıklar altında ayrıca açıklanmıştır.

Tablo 7 Genetik algoritma uygulama adımları

Adım	Açıklama
1	Rastgele veya manuel yaklaşım kullanarak bireylerin ilklendirilmesi
2	Popülasyon listesini bireylerle doldurma
3	Tüm bireylerin yakıt tüketim değerlerinin (fitness value) değerlendirilmesi
4	Seçme
5	Çaprazlama
6	Mutasyon

3.3.2.1. Bireylerin İlklendirilmesi

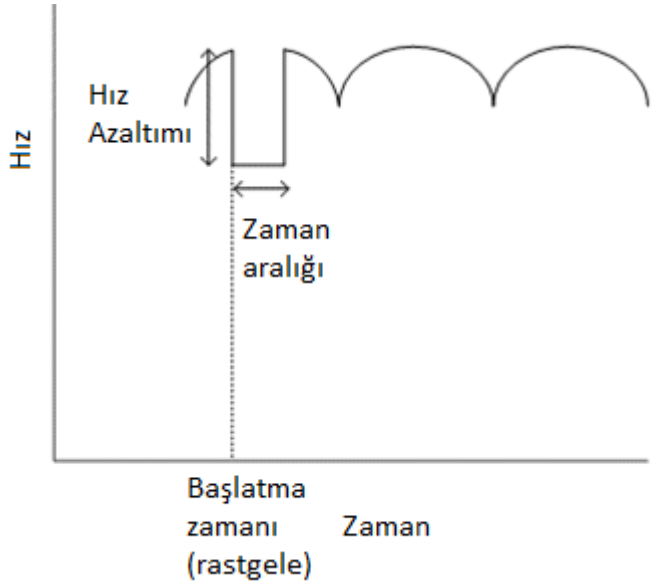
Bu çalışmada gemi hız profilleri, GA'nın bireyleri olarak önerilmiştir. Ayrıca, her birey bir yakıt tüketim puanı ile değerlendirir. Farklı yakıt tüketim puanlarını değerlendirmek için her bireyin bazı farklılıkları olmalıdır. Geminin gerçek hız profili Şekil 17'de sunulmaktadır.



Şekil 17. Geminin gerçek hız profili

Bu çalışmada, önerilen hız optimizasyonu yöntemi, evrimsel algoritmaları kullanarak en uygun seyir hızı profillerini araştırmaktadır. Ancak bu algoritmaların başlatılması, hız optimizasyonunda yakınsama ve dayanıklılık sorunlarına neden olabilir [14]. Bu çalışmada, bu sorunların üstesinden gelmek için bireyler arasında üst düzey bir genetik çeşitlilik oluşturulması amaçlanmıştır.

Bu nedenlerle, yeni bireyler için farklı ilklendirilmiş hız profilleri oluşturmak üzere hız profilinde değişiklikler yapılmıştır. Bu amaçla, iki farklı yaklaşım kullanılmıştır. Birincisi, orijinal hız profilinin rastgele seçilen bir bölümünde rastgele bir değişiklik yapılmasıdır. Böylece orijinal hız profillerinde rastgele seçilen miktarda hız azaltma ve zaman aralığı değişikliği uygulanmıştır. Rastgele hız azaltmaları, yakıt tüketiminin hız değişkenine olan esnekliği olarak kabul edilebilir. Benzer şekilde, sabit azaltmalar sabit esneklik olarak kabul edilebilir [8]. Optimum hız etrafındaki esnekliğin bulunması, bu çalışmada bir araştırma hedefidir. Bu rastgele yaklaşım ise, GA'da seçim, çaprazlama ve mutasyon aşamalarından önce bireyler arasında iyi bir genetik çeşitliliği meydana getirmektedir. İkinci yöntem ise, rastgele yaklaşımla karşılaştırmak amacıyla yapılmış olup yine rastgele bir zamanda ancak, önceden belirlenmiş miktarlarda bir hız değişikliği tanımlamaktır. Şekil 18'de rastgele bir değişikliğin görünümü gösterilmektedir.



Şekil 18. Bireyleri ilklendirmek için rastgele veya manuel hız değişimi

3.3.2.2. Popülasyon ilklendirme

Bir popülasyon, N sayıda bireye karşılık gelir. Bu tezde, her üretim döngüsünde rastgele veya manuel yaklaşımla N sayıda yeni birey ilklendirilir.

3.3.2.3. Uygunluk değerlerinin hesaplanması

Fitness, yakıt tüketimini en aza indirmeyi amaçlayan bir fonksiyon tanımıdır. Rastgele veya manuel hız azaltmaları, çaprazlama ve mutasyonlar, yolculuk için gecikme süresine neden olabilir. GA'da ilk olarak kalan mesafedeki hız dahil edilerek yeni toplam yolculuk süresi hesaplanır. Ve GA tarafından toplam yolculuk süresinin aşılmasına neden olan hız profillerine bir ceza puanı verilir.

3.3.2.4. Seçme

Bir popülasyonda değerlendirilen tüm bireyler, uygunluk değerlerine göre artan düzende sıralanır. Çünkü bu çalışmada uygunluk fonksiyonunun amacı yakıt tüketimini en aza indirmektir. Bir nesildeki ilk adım, yavru seçimidir. Bu çalışmada 3 turnuva büyüklüğünde ve k(1) süreli K-Turnuva seçim yöntemi kullanılmıştır. Çaprazlama adımı için her seferinde en iyi iki birey seçilir.

3.3.2.5. Çaprazlama

En iyi bireyler, çapraz olasılık oranlarına göre eşleştirilir ve çiftleştirilir [24]. En düşük uygunluk değerine sahip mevcut neslin yeni bireylerini tanımlar. Bu çalışmada iki noktalı

çaprazlama kullanılmıştır. Her çaprazlama adımı tamamlandıktan sonra azami hız kontrol edilir ve herhangi bir hız değeri azami hız limitini aşarsa hız limiti ile değiştirilir.

3.3.2.6. Mutasyon

Mutasyon, genetik değişkenliği artırmak için bir başka önemli adımdır. Mümkün olduğu kadar uzun süre en iyi sonuçları aramaya devam etmek için mutasyonlar gereklidir. Aksi takdirde, optimizasyon yanlış bir şekilde yerel bir minimum veya maksimum noktayı global olarak kabul eder. Yanlış varsayım riskini azaltmak için belirli bir büyüklükte mutasyon oranı kullanmakta fayda vardır [14]. Bireylerde rastgele seçilen değerler, sıfır ortalama ve bir standart sapma ile tanımlanan gauss-normal dağılım aralığında mutasyona uğrar. Önceki beş adımın tamamı, tüm nesillerin sonuna kadar tekrarlanır. Toplam nesil sayısı genetik optimizasyonun yaygın kullanımından öğrenilen şekilde yaklaşık 99 olarak tanımlanmıştır.

3.3.3. Genetik Algoritma Parametreleri

1) Mutasyon ve Çaprazlama Olasılıkları

Mutasyon ve çaprazlama parametreleri, genetik algoritmaların ana faktörleridir. Literatürde bazı çalışmalar, yüksek çaprazlama oranlarının üreme için ana faktör olduğunu, ancak yüksek mutasyonların da DNA'yı tahrip edebildiğini ve bu nedenle nadiren kullanıldığını göstermektedir [30]. Bu çalışmada iki noktalı çaprazlama kullanılmıştır. Ayrıca, çaprazlama oranları ampirik olarak yüzde 20 ila 90 arasında araştırılmıştır. Yine de, %80 [31] ve %75 [32] çaprazlama oranları literatürde yaygın olarak kullanılan değerlerdir. Ek olarak, hız değerlerinin miktarını değiştirmek için sıfır ortalamalı Gauss dağılımı kullanılır.

2) Popülasyon Büyüklüğü

Bir popülasyondaki birey sayısı, genetik algoritma performansı üzerinde büyük bir etkiye sahiptir. Problem uzayı, daha yüksek popülasyon boyutlarına sahip genetik algoritmalar ile daha iyi analiz edilebilir. Ancak çözüm, problemin kendisine bağlı olduğu için daha iyi bir çözüm noktası garanti etmez [24]. Ayrıca, çok sayıda bireyle gerçekleşen optimizasyon, daha uzun hesaplama süresi gerektirir.

Optimum birey sayısı, uygulamanın karmaşıklığına bağlıdır. Ek olarak, algoritma karmaşıklığı ve birey sayısının performans üzerinde aynı yönde etkisi vardır. Literatürde tavsiye edilen birey sayısı 30 ile 300 arasındadır. Ancak en iyi performansı gösterecek optimum sayıyı bulmak için

ampirik yöntemler de tavsiye edilmektedir [22]. Bu çalışmada, optimum popülasyon büyüklüğünü bulmak için uyarlamalı boyutlandırma yöntemi kullanılmıştır.

3) En Yüksek Hız

Gecikme süresini belirlemek için öncelikle ETA ve kalan mesafe hızı tanımlanmalıdır. Daha sonra regresyon analizinden seçilen en iyi yakıt tüketimi tahmin modeli tarafından, değiştirilmiş ve kalan mesafedeki hızlar için karşılık gelen yakıt tüketimi eğrisi yaratılır. Sonuçta toplam yakıt tüketimi bir bireyin fitness skoru olup genetik algoritma tarafından minimize edilmesi hedeflenir. Bu nedenle, görece yüksek fitness skoruna sahip bireyler popülasyonlar içerisinde daha düşük oranda seçilme şansına sahip olduğu için gelecek jenerasyonlara aktarılma olasılığı da düşüktür.

3.3.4. Genetik Algoritmanın Parametre Ayarı

Genetik algoritmaların performansı büyük ölçüde parametre ayarlarına bağlıdır [33]. GA'da kullanılan tüm ayar parametreleri Tablo 8'de listelenmiştir. Hız düşüşü, 2 ile 6 knot arasında rastgele bir sayı ile başlatılır ve hız düşürme süresi, rastgele yaklaşım için 30 ile 300 dakika arasında yine rastgele bir zamanda başlatılır. Ancak, manuel yaklaşım için önceden belirlenmiş şekilde manuel olarak başlatılırlar. Bu nedenle, rastgele yaklaşım, manuel yaklaşımdan daha fazla genetik çeşitliliğe sahiptir.

Tablo 8 Manuel ve rastgele ilklendirmede parametrelerinin karşılaştırılması

Parameter	Manuel İlklendirme	Rastgele İlklendirme
Mutasyon olasılığı	var	var
Çaprazlama olasılığı	var	var
Mutasyonun std. sapması	var	var
Hız düşüm süresi	var	yok
Hız düşümü	var	yok
Birey sayısı	var	var
En yüksek hız	var	var
Kalan mesafe	var	var

Parametreler için farklı ayar değerlerinin kullanılması yakıt tüketiminde farklı miktarda bir azalmaya neden olur. Tüm parametrelerin en iyi kombinasyonu için başka bir optimizasyon

çalışmasına ihtiyaç vardır. Bunun yerine bu çalışmada parametrelerin sayısal ve görsel olarak karşılaştırılması tercih edilmiştir.

Ayrıca, bu çalışmada hız profili modifikasyonu için rastgele ve manuel yaklaşımlar kullanılmış, bu nedenle bireylerin rastgele ve manuel ilklendirme yöntemleri arasındaki karşılaştırmayı göstermek için iki farklı parametre ayarlama çalışması yapılmıştır.

3.3.4.1. Rastgele İlklendirme

Rastgele ilklendirme yaklaşımında, hız profili modifikasyon parametreleri (bkz. Şekil 18) daha iyi global optimal noktaları araştırmak amacıyla rastgele seçilir. Hız düşüşü 1 ile 6 knot arasında rastgele seçilir ve hız düşüş süresi 30 ile 300 dakika arasındadır. Bu rastgele süreç, ait olduğu kuşağın tüm bireyleri için her kuşağın başında tekrarlanmıştır. Böylece, genetik algoritma yürütülmeden önce yüksek düzeyde bir genetik çeşitlilik üretilir.

Daha sonra, farklı sayıdaki mutasyon ve çaprazlama olasılıkları, standart sapma, en yüksek hızlar ve birey sayıları kontrol parametreleri olarak kullanılırken diğerleri sabittir. Son olarak, yakıt tüketimi sonuçları görsel olarak karşılaştırılır.

3.3.4.2. Manuel İlklendirme

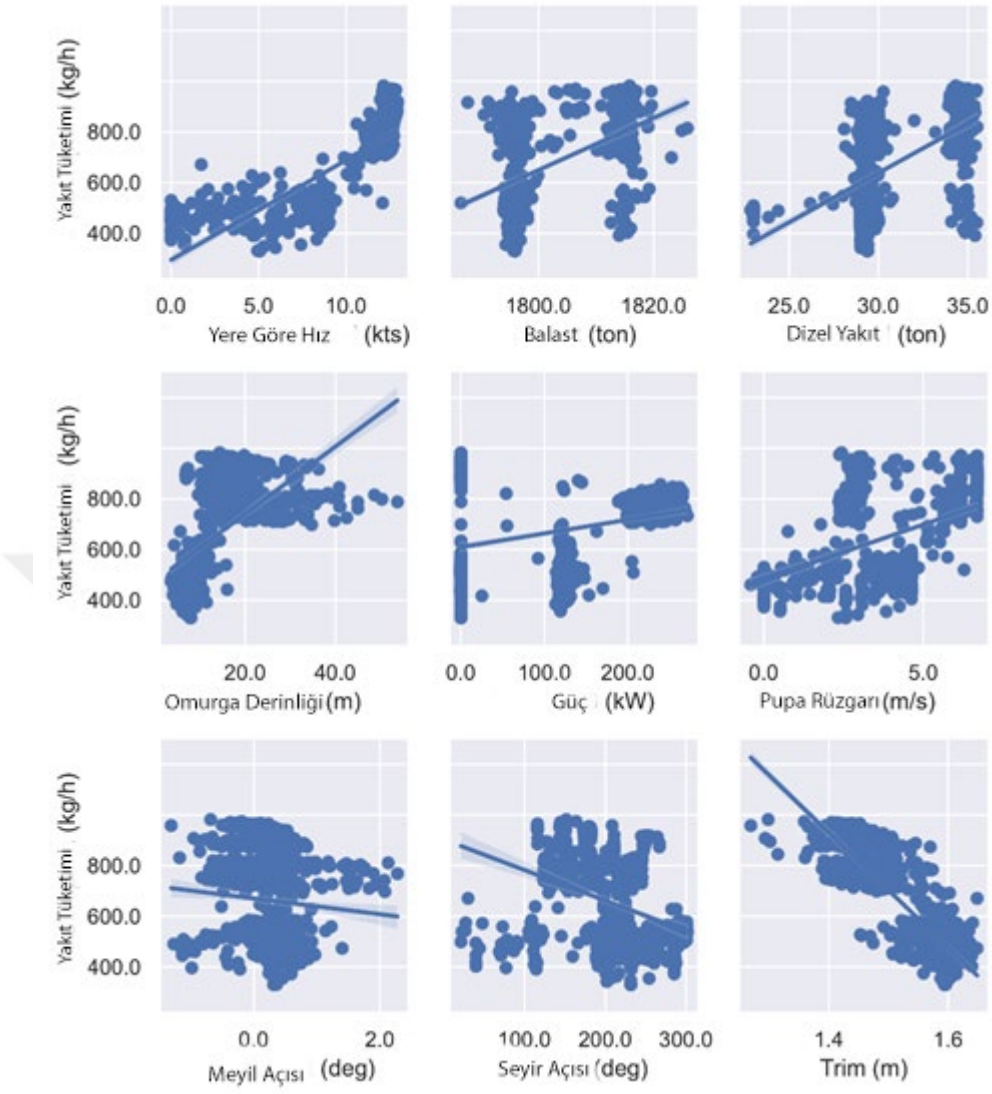
Bu bölümde, bir popülasyondaki bireyleri oluşturmak için hız düşüşü ve zaman aralığı manuel olarak değiştirilir. Rastgele yaklaşımda kullanılan diğer tüm parametreler ise burada da aynı şekilde kullanılır. Hız profillerindeki bir modifikasyonun ilklendirilme zamanı, hem rastgele hem de manuel yaklaşımlarda rastgele seçilir.

4. BULGULARIN DEĞERLENDİRİLMESİ

4.1. DOĞRUSAL REGRESYON TEST SONUÇLARI

a. Doğrusallık Testi

Şekil 19'da tüm regresyon değişkenlerine ait dağılım grafikleri görülmektedir. Y ekseni bağımsız değişken olup yakıt tüketimini temsil eder. Ayrıca, en uygun doğrunun eğimleri diğer bağımsız değişkenlere bağımlı değildir [34]. Dolayısıyla, yere göre hız ve trim değişkenlerinin bağımlı değişkenle daha doğrusal olarak ilişkilenme eğiliminde olduğu Şekil 14'teki korelasyon tablosu kullanılarak iddia edilebilir. Ancak diğer değişkenler için aynı şeyi söylemek zordur. Şekil 14'teki korelasyon matrisine bakıldığında su, yük ve meyil açısı değişkenlerinin korelasyon katsayılarının sırasıyla -0.04, -0.07 ve -0.08 olup değerler sıfıra o kadar yakındırlar ki, bu parametrelerin yakıt tüketimi tahmininde beklenen herhangi bir toplanabilirliği yoktur [34].



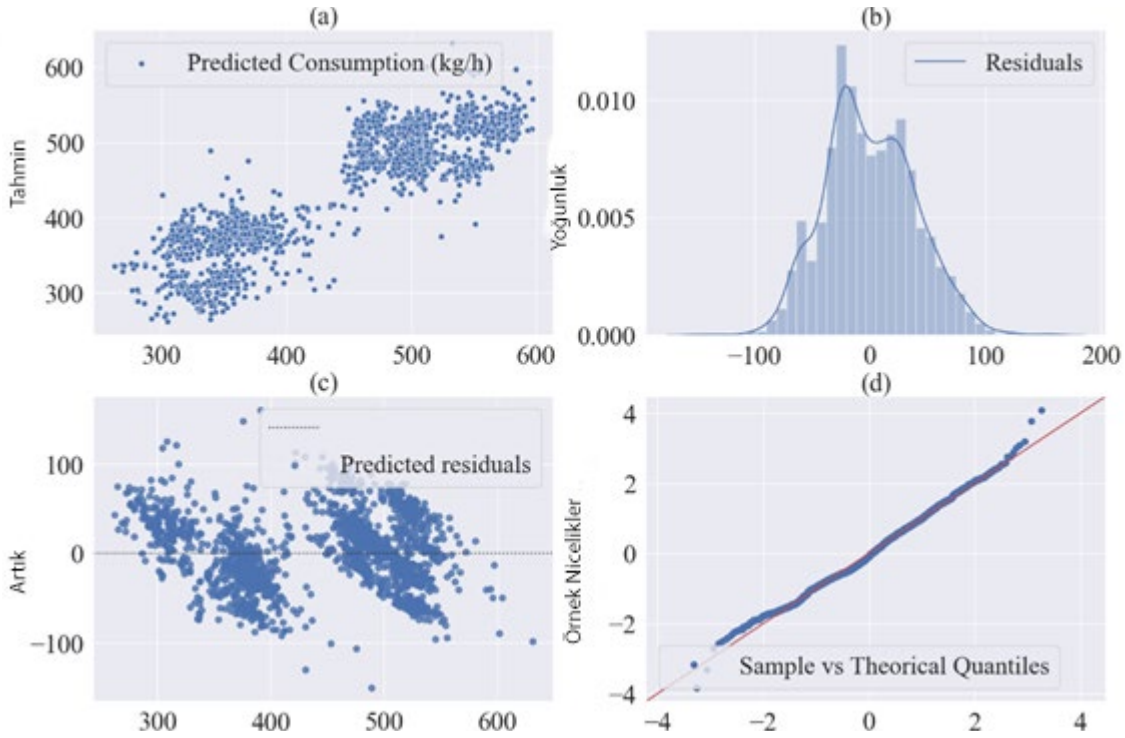
Şekil 19. Seçilen özelliklerin yakıt tüketimine karşı dağılım grafikleri, en uygun doğrusal regresyon çizgileri dahil

b. Çoklu Doğrusal Bağlılık Testi

Şekil 14'teki korelasyon matrisinde koyu renkli hücrelere karşılık gelen değişkenler kullanılarak bağımsız değişkenlerin çoklu doğrusal bağlantısı kontrol edilebilir. Yüksek oranda ilişkili değişkenlerin varlığı, yüksek RMSE hatalarına neden olur. Dolayısıyla tahmin doğruluğu azaldığından güven aralığı da daralır [23]. Ancak bir hatırlatma olarak, rüzgar hızı değişkeni önceki bölümde regresyon analizinden zaten çıkarılmıştı. Bunun dışında, balast ile dizel yakıt arasındaki ve yere göre hız ile trim arasındaki korelasyonlar hala yüksektir. Ancak trim, bu tezde yakıt tüketimi üzerindeki etkiyi araştırmak için bir başka önemli değişkendir. Bu nedenle analizde tutulmuştur.

c. Otokorelasyon, değişen varyans ve normallik veya hata terimlerini kontrol etme (residuals)

Hataların varyansı uydurulan değerlere göre artar veya azalırsa, güven aralıklarının daha küçük bir aralığa girmesine neden olacaktır. Bu sorun, değişen varyans olarak bilinir ve regresyonda yalnızca küçük bir veri alt kümesine çok fazla önem verilmesi ile sonuçlanır [34]. Şekil 20(c)'de, normalleştirilmiş artık değerlerin sıfır ekseni boyunca dağılımının artış veya azalma eğiliminde olmadığı görülmektedir. Yine, dağılım hataları normal dağılırsa, RMSE hata hesaplaması onu daha fazla azaltmakta başarısız olabilir ve güven aralıkları çok dar veya çok geniş olacaktır [34]. Şekil 20b'den artıklar arasında bir korelasyon olmadığı görülmektedir.



Şekil 20 Doğrusal regresyonun tahmini ve artıkları (a) Tahmine karşı gözlemlenen, (b) Artıkların normal dağılımı, (c) Normalleştirilmiş artıklara karşı tahminler, (d) Örnek niceliklerin eğilim çizgisi

Sonuçta, özniteliklere düzeltici dönüşümler uygulansa bile doğrusal regresyon sonuçları yeterli doğrulukta ve yüksek RMSE değerleri ile sonuçlanmamıştır. Bu nedenle, doğrusal regresyon modelleri kullanan yakıt tüketimi tahminleri, mevcut veri kümesi için güvenilir ve uygulanabilir değildir.

4.2. DOĞRUSAL OLMAYAN METODLARIN SONUÇLARI

Değerlendirme sürecinde kullanılan eğitim ve test veri kümesi performans ölçütleri sırasıyla Tablo 9 ve 10'da verilmiştir.

Sonuç olarak, doğrusal regresyon modelinin R2 değerleri yüzde 80 civarındadır, bu nedenle model, literatürdeki çalışmalara kıyasla yakıt tüketimini daha zayıf bir şekilde açıklamaktadır [11, 20, 35]. Ek olarak, Şekil 21'de, LR'nin eğitim doğruluk puanları artan eğitim verisi boyutuyla oldukça düşük bir değere yakınsamaktadır. Bu nedenle, daha fazla eğitim verisi eklemenin doğrusal regresyona fayda sağlamadığı söylenebilir.

Ayrıca SVR modelinin R2 ve RMSE skorları, LR modelinden daha düşüktür (bkz. Tablo 9). Üstelik, SVR literatürdeki çalışmalara kıyasla yakıt tüketimini zayıf bir şekilde açıklamaktadır [11, 36]. Bunun yanında, KNN'nin performans ölçütleri, LR ve SVR modellerinden daha iyi sonuç vermiştir. Ayrıca bu çalışmadaki KNN modeli yakıt tüketimini literatüre göre daha iyi açıklamaktadır . Ayrıca SVR ve KNN modelleri, yeni eğitim verileri ekleyerek doğruluk değerlerini artırmaktadır (bkz. Şekil 22) ve doğruluk puanları, yeni veriler ekleyerek birbirine yakınlaşır, yeni verilerle iyi uyum sağlarlar. Öte yandan, KNN ve SVR'nin hiperparametre optimizasyonu sonucunda, öğrenme eğrilerinin doğruluk puanları hala yüksek olup aralarında bir fark bulunmaktadır. Bu nedenle, KNN ve SVR için aşırı öğrenme yaşanmış olabilir. Özellikle daha az eğitim verisi için, SVR'nin eğitim doğruluk puanı, doğrulama değerinden çok daha yüksektir. Ve yeni eğitim verileri eklemek, genellemeyi artıracaktır.

Eğitim ve test veri kümeleri için VR ve SR topluluk modellerinin doğrulukları birbirine yakındır. Test veri kümesi için VR eğitim doğruluğu (%94,55), SR'den (%93,53) biraz daha yüksektir. Ancak VR ve SR topluluk modellerinde, eğitim ve test puanları arasında hala bir fark olması, aşırı öğrenme sorununu göstermektedir.

Sonuç olarak, RF hiperparametreleri, rastgele arama optimizasyonu ile ayarladıktan sonra eğitim doğruluk değerini %98,07'den %98,39'a yükseltmiştir (bkz. Tablo 9). Ayrıca, test doğruluğu da %94,48'den %94,63'e yükselmiştir ve sonuçta da RF için aşırı öğrenme yaşanmamış olduğu varsayılabilir (bkz. Tablo 10). Bu nedenle, bir sonraki bölümde hız optimizasyonu değerlendirmesi için RF seçilmiştir.

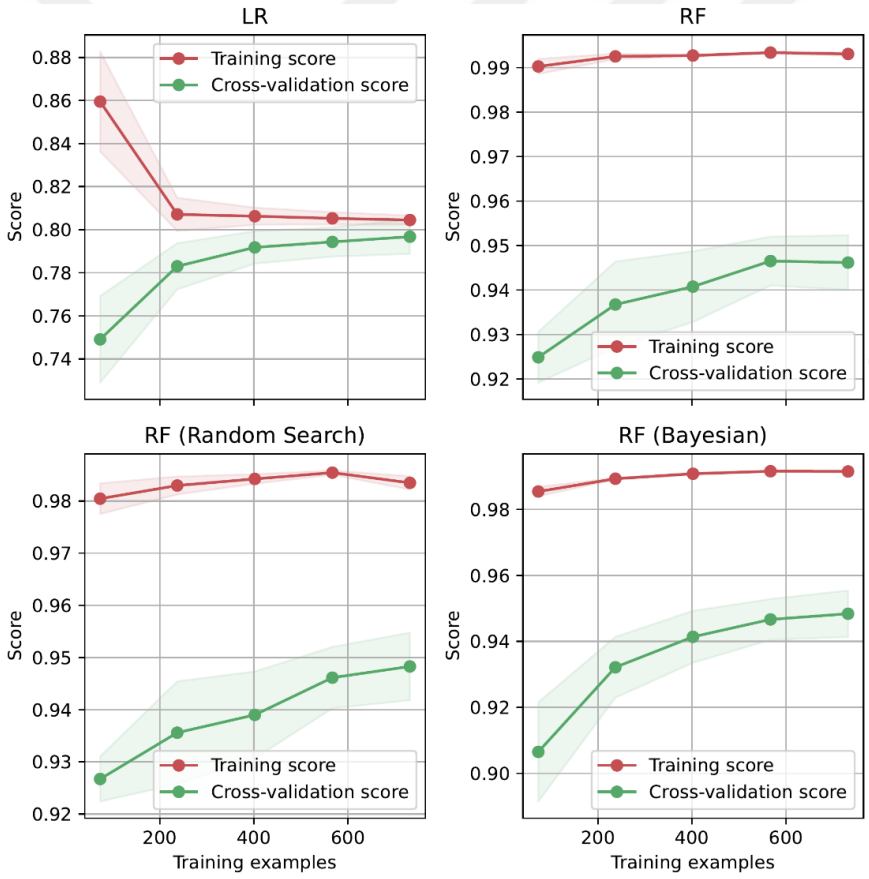
Tablo 9 Eğitim veri kümesinin performans sonuçları

Model	RMSE	R2 Score (%)
SVR	41.37	77.23
LR	38.41	80.38
KNN	28.31	89.34
Tuned ADA (Grid Search)	18.53	95.43
Tuned SVR (Grid Search)	15.11	96.96
ADABoost	15.02	97.00
GBR	13.54	97.56
RF	**12.04**	**98.07**
SR	11.06	98.37
Tuned RF (Random Search)	**11.00**	**98.39**
XGBR	9.85	98.71
VR	7.87	99.18
Tuned RF (Bayesian)	**7.84**	**99.18**
Tuned GBR (Grid Search)	5.09	99.66
Tuned GBR (Random Search)	2.01	99.95
Tuned XGB (Random Search)	1.70	99.96
Tuned KNN (Grid Search)	0.01	99.99

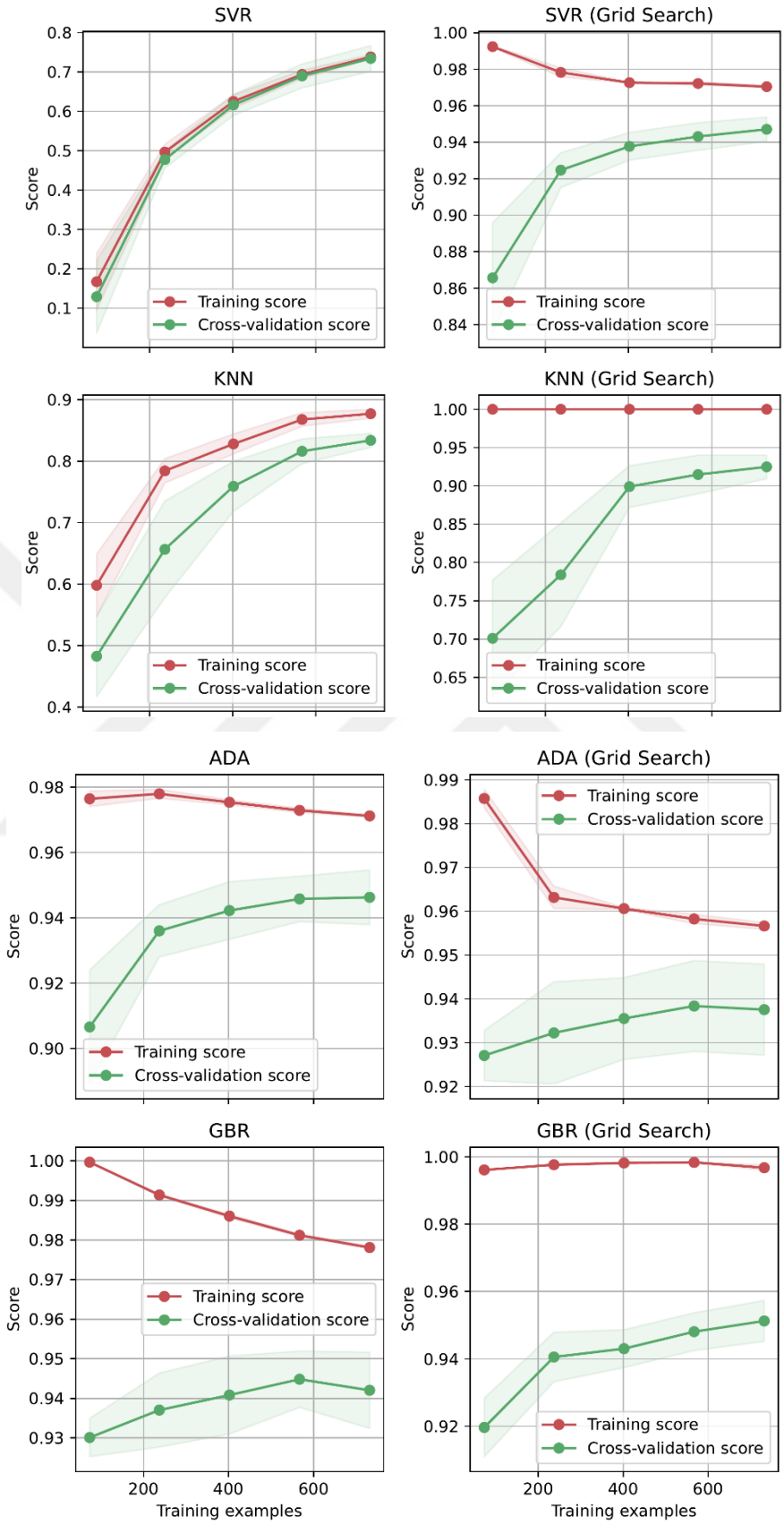
Tablo 10 Test veri kümesinin performans sonuçları

Model	RMSE	R2 Score (%)
SVR	41.81	75.33
LR	39.63	77.83
KNN	29.50	87.72
Tuned KNN (Grid Search)	25.57	90.78
Tuned SVR (Grid Search)	24.86	91.28
Tuned ADA (Grid Search)	23.44	92.25
SR	21.41	93.53
Tuned RF (Bayesian)	**21.07**	**93.73**

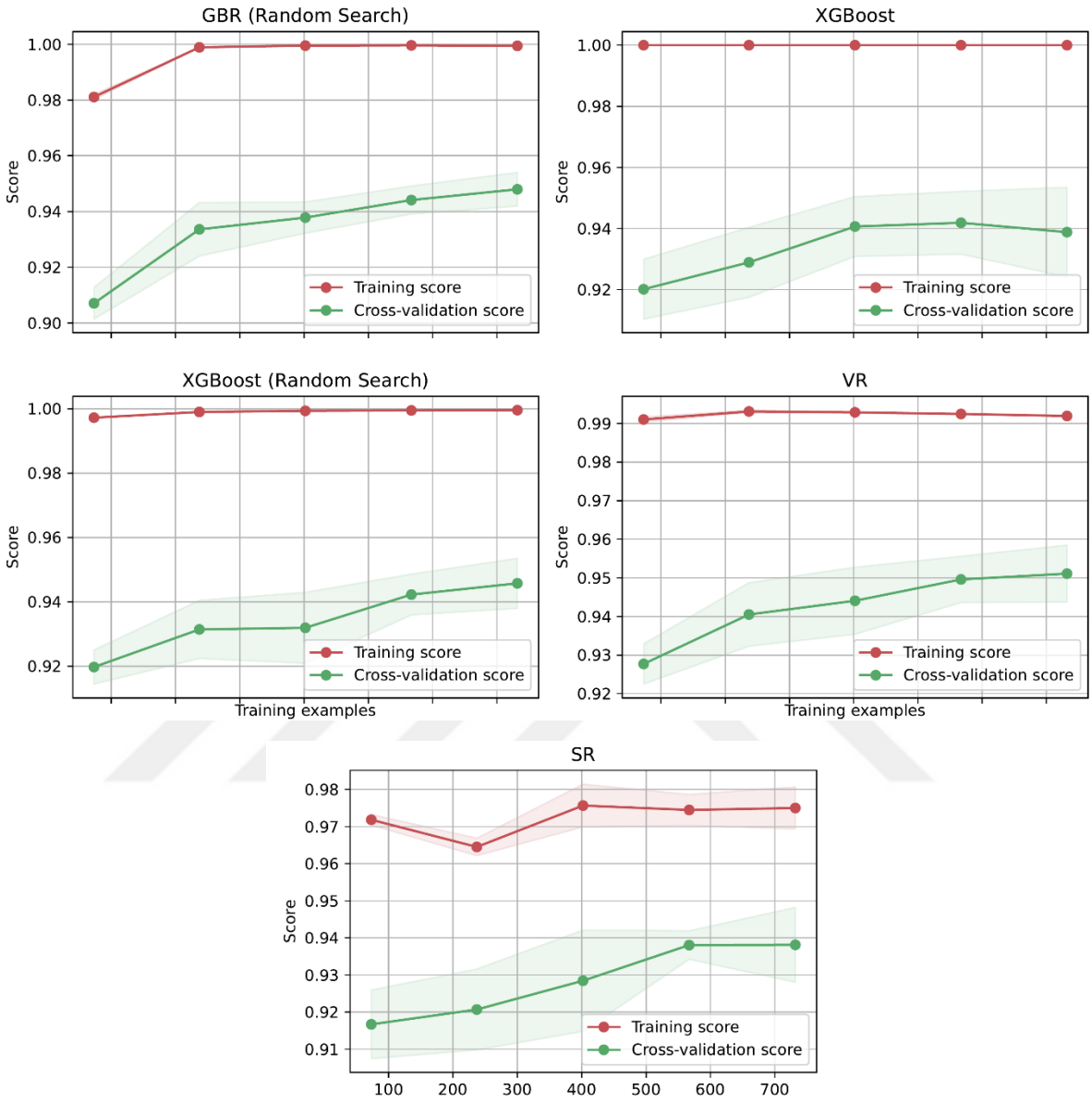
Tuned XGB (Random Search)	20.78	93.90
Tuned GBR (Random Search)	20.75	93.93
ADABoost	20.62	94.00
GBR	20.44	94.10
RF	**19.77**	**94.48**
VR	19.66	94.55
XGBR	19.61	94.57
Tuned GBR (Grid Search)	19.58	94.59
Tuned RF (Random Search)	**19.52**	**94.63**



Şekil 21. LR ve RF modellerinin öğrenme eğrileri

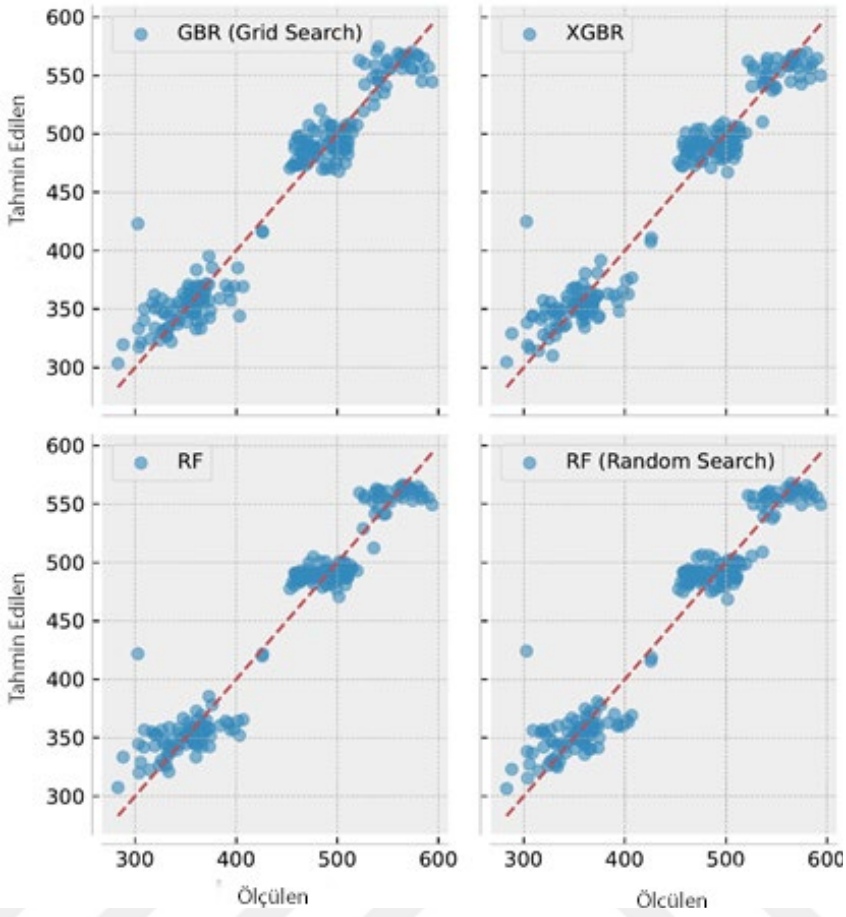


Şekil 22. SVR, KNN, ADABoost ve GBR modellerinin öğrenme eğrileri



Şekil 23. XGBoost, VR ve SR modellerinin öğrenme eğrileri

Test veri kümesindeki en doğru modellerin tahmini ve ölçülen yakıt tüketimi (kg/s) Şekil 24'te gösterilmektedir.



Şekil 24. Test edilen ilk 4 modelin tahmini ve ölçülen yakıt tüketimi (kg/s)

4.3. GENETİK ALGORİTMALAR İLE HIZ OPTİMİZASYONU SONUÇLARI

4.3.1. Rastgele İlklendirme Sonuçları

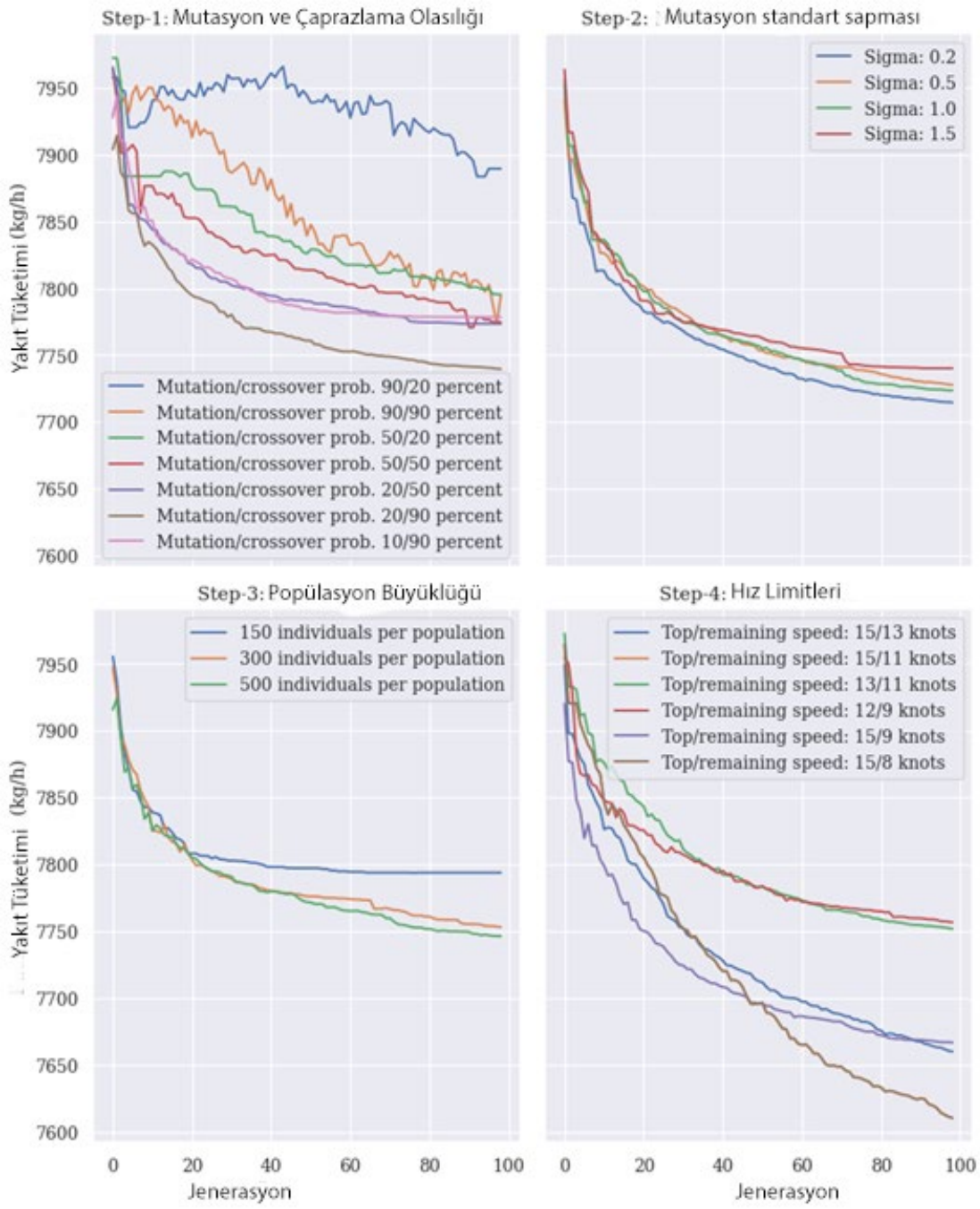
Bireylerin ya da hız profillerinin rastgele ilklendirilmesinde GA parametre ayarı manuel olarak dört adımda uygulanmıştır. Tablo 11'de, her adım için tüm kontrol ve sabit parametreler listelenmiştir. İlk adımda, mutasyon ve çaprazlama olasılık kontrol parametreleri değiştirilir ve GA, olasılıkların %10'u ile %90'ı arasında 7 farklı kombinasyon için tekrar tekrar çalışır. İkinci adımda, mutasyon ve çaprazlama olasılıklarının en iyi parametreleri kullanılır ve mutasyon standart sapma (sigma) kontrol parametreleri dört kez 0,2'den 1,5'e kadar arttırılmıştır. Üçüncü adımda ise popülasyonun birey sayıları 150, 300 ve 500 olmak üzere üç kez değiştirilmiştir. Son olarak hız limiti ve kalan hız parametreleri altı farklı kombinasyonda değiştirilmiştir.

Tablo 11 Rastgele ilklendirilen GA'da kontrol parametreleri ve sabitler

Adım	Mutasyon olasılığı	Çaprazlama olasılığı	Sigma	Popülasyon birey sayısı	Hız limiti (mile/h)	Kalan mesafe hızı (mile/h)
1	değişken	değişken	1.5	300	13	13
2	20%	90%	değişken	300	13	13
3	20%	90%	1	değişken	13	13
4	**20%**	**90%**	**1**	**500**	değişken	değişken

4.3.1.1. Rastgele İlklendirmede Hiperparametre Optimizasyonu Sonuçları

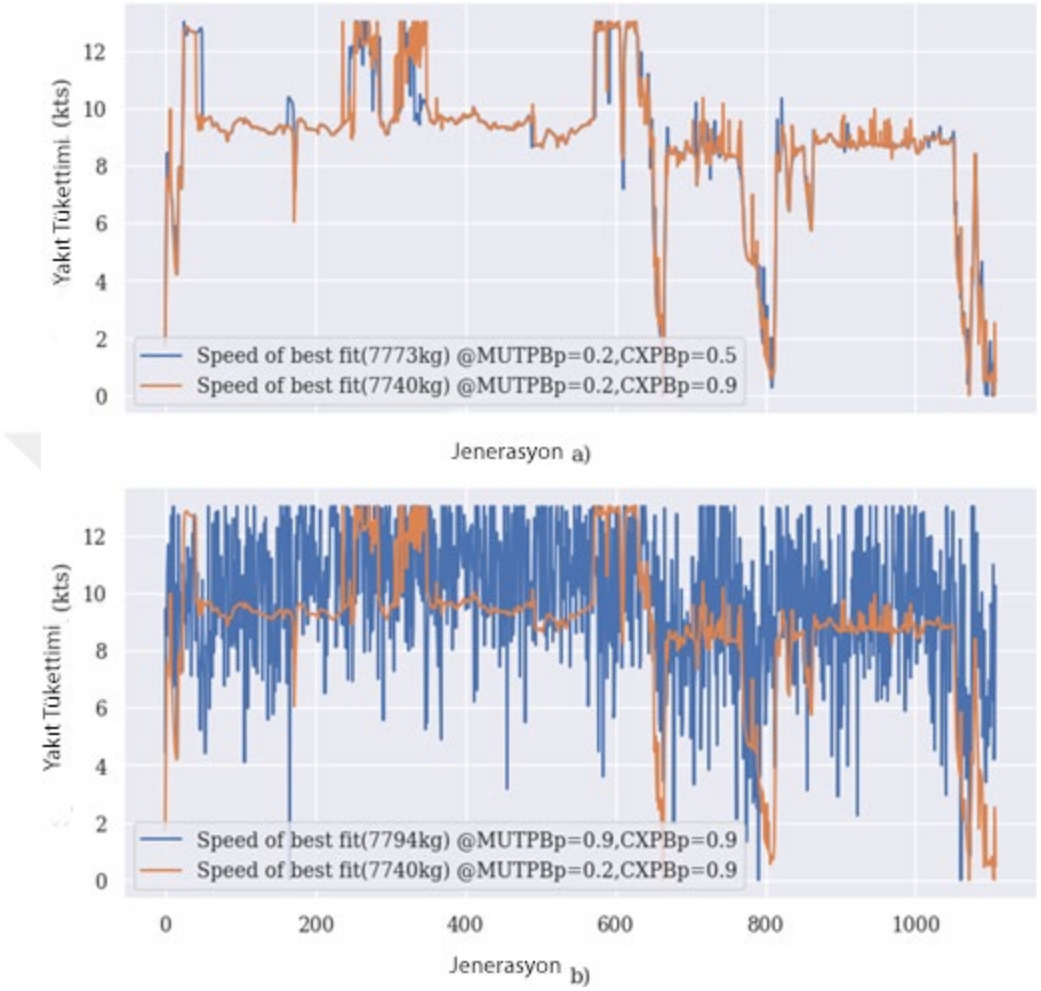
Bu çalışmada GA, kontrol parametreleri değiştirildikten sonra tekrar tekrar çalıştırılmıştır. GA simülasyonlarının minimizasyon veya uygunluk değerlerinin amaç fonksiyonu, yakıt tüketimi değişkenidir. Parametre uyarlama simülasyonları dört adımda tamamlanmıştır. Sonuçlar Şekil 25'te sunulmuştur.



Şekil 25. Rastgele ilklendirme yönteminde parametre en iyilenmesi

Mutasyon ve çaprazlama olasılığı: Şekil 25'e bakarak düşük mutasyon oranlarının, yüksek çaprazlama oranları ile birlikte daha iyi performans gösterdiği söylenebilir. Yine de, çok düşük mutasyon oranları daha fazla yakıt tüketimine neden olmaktadır. Ayrıca, mutasyon oranının artmasının kararsız hız profillerine neden olduğu Şekil 26b'de görülmektedir. Aksine, daha yüksek bir çaprazlama oranı, küresel bir minimum nokta arayışı sırasında hız profilini sabit tutmaktadır. Ayrıca, çaprazlama operatörü hız profilinin her yerine eşit olasılıklı davranır (bkz. Şekil 26a).

Bu nedenle, rastgele yaklaşımda, sonraki adımlarda %20 mutasyon ve %90 çaprazlama oranları seçilmiştir.



Şekil 26. Rastgele yaklaşım adım-1'de en iyi hız profilleri, a) çaprazlama ve b) mutasyonlar

Mutasyon standart sapması (sigma): Sonuçları önemli ölçüde etkilemese de, yüksek gemi hızı yayılımı (sigmanın tersi) daha yüksek performans göstermiştir. Sigma parametresi her seferinde aynı etkiyi yaratmaz. Şekil 27'de gösterilen bazı durumlar ters yönde sonuçlanmıştır. Şekil 27'de farklı sigma değerlerine sahip popülasyonların en uygun bireylerinin hız profilleri ile gerçek yolculuk hız profilleri birlikte gösterilmektedir. Genetik algoritmada sigma değerlerinin arttırılması, hız değerlerinin daha yüksek oranda değiştirilerek daha fazla yakıt tasarrufu sağlanmasına yol açtığı görülmektedir. Ayrıca, Şekil 27'deki gri çizgili pencere çerçevesinde, sigma değerleri arttığında genetik algoritma, hızdaki küçük değişiklikler yaparak yakıt tüketimi tasarrufu için bir çözüm aramaktan vazgeçmektedir. Bu sebeple genetik algoritma için, yerel bir minimum noktaya yığılmayı aşarak, global minimum noktayı aramaya devam etmekte olduğu söylenebilir.



Şekil 27. Rastgele yaklaşım adım-2'de en uygun hız profilleri ve mutasyon standart sapmaları

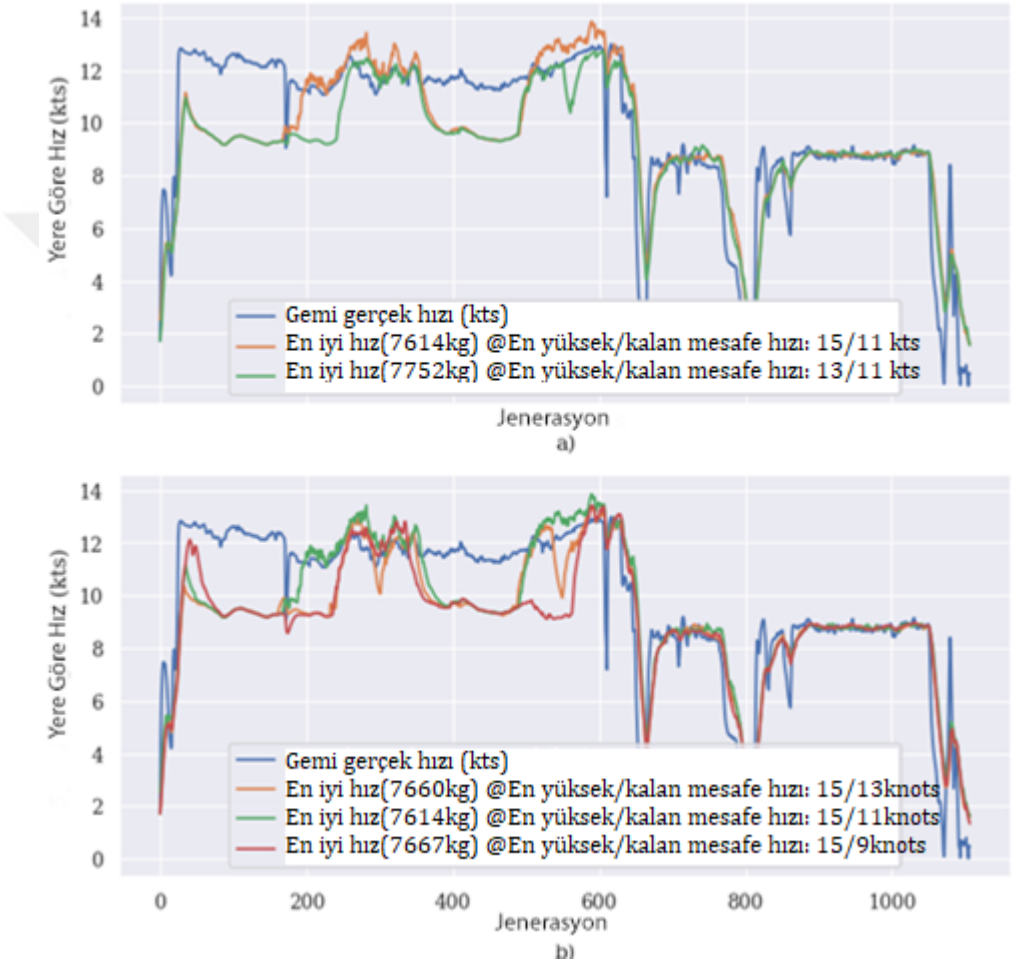
Ayrıca, en iyi bireylere ait hız sinyalleri, pratik kullanım için çok fazla gürültüye sahiptir. Bu nedenle, yüksek frekanslı sinyalleri ortadan kaldırmak için bir alçak geçiren filtre gereklidir. Python Pandas kütüphanesindeki üstel ağırlıklı hareketli ortalama filtresi, Şekil 28'de görüldüğü gibi 0,1 gibi çok düşük bir yumuşatma faktörü (α) ile uygulanmıştır.

Popülasyon büyüklüğü: Şekil 25'e bakıldığında, bir popülasyondaki birey sayısını artırmak bu çalışmada her zaman daha iyi GA performansı göstermektedir. Ayrıca, 500 adet birey sayısı en verimli sonuçları göstermiştir.

Azami/kalan mesafe hızı: Şekil 25'te görüldüğü gibi, yüksek hız sınırı ve düşük kalan mesafe hızları belirlemek daha fazla yakıt tasarrufu ile sonuçlanmıştır.

Hız profilinde aşağı ve yukarı yönlü (maksimum hıza kadar) rastgele değişiklikler yapmak, değiştirilen hız profilinin katedeceği mesafeyi değiştirir. Genetik algoritmanın değerlendirme adımı sonunda hız profilindeki net toplam hız düşüşü nedeniyle kalan mesafe ortaya çıkabilir. Kalan mesafelerin hızı ile azami hız sınırı birbirini etkiler. Adım 4'e ait farklı kombinasyonların

hız profili çıktıları Şekil 28'de gösterilmiştir. Kalan mesafe hızlarını sabitlerken üst hız sınırının düşürülmesi daha az yakıt tasarrufu sağlamaktadır (bkz. Şekil 28a). Ancak bu durumun tersi de geçerli değildir (bkz. Şekil 28b). Örneğin, azami hız limiti 15 knot olarak sabitlenirken, 15 veya 9 knot olan kalan mesafe hızları, 11 knot'tan daha fazla yakıt tüketimine neden olmuştur. Kısaca, en yüksek hız limiti, GA'da kalan mesafe hızı ile doğrusal olmayan bir şekilde ilişkilidir. Ancak, en iyi kombinasyonu bulmak için başka bir optimizasyon rutini gerekir.



Şekil 28. Rastgele yaklaşma adım-4'e en uygun hız profilleri, a) değişen hız limiti ve b) kalan mesafe için seyir hızı

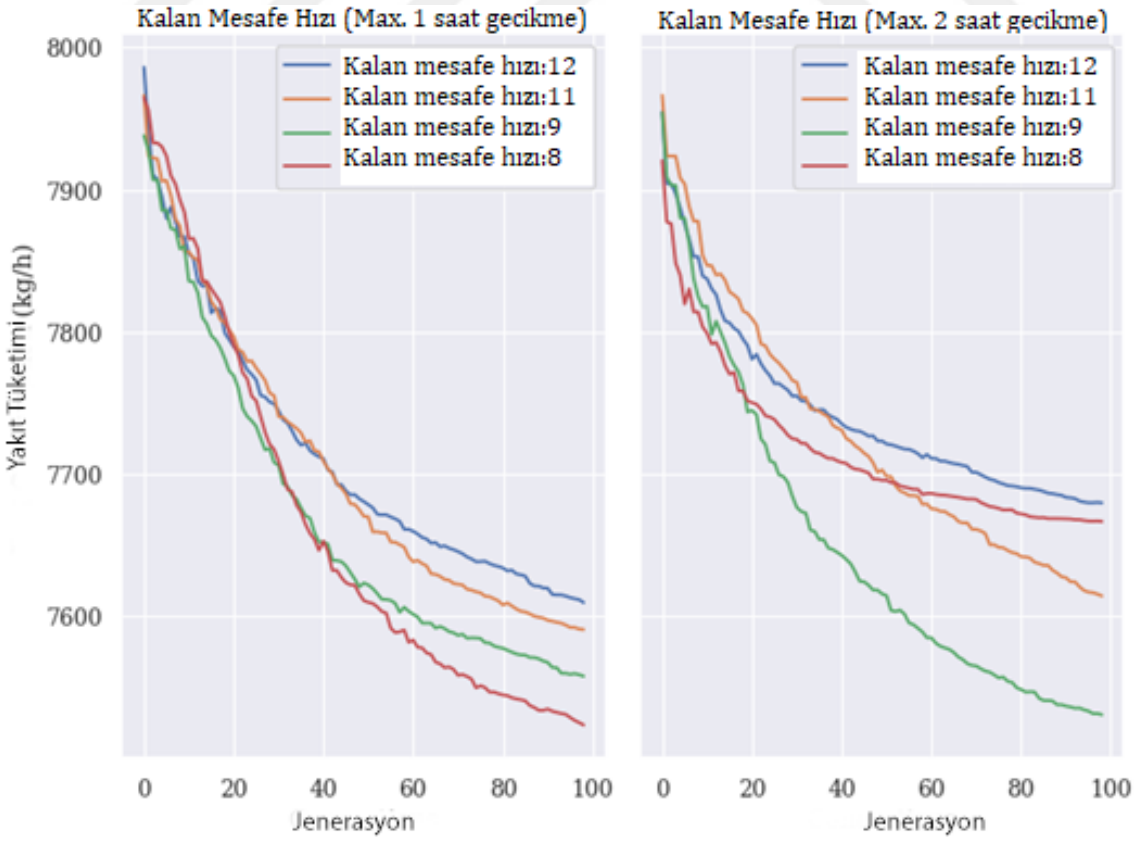
4.3.1.2. ETA Gecikme Durumları

Gerçek hayatta teknik, hava durumu veya yönetimsel sebeplerden ötürü limana varış zamanının gecikmesi çok yaygındır. Buna bağlı olarak, yolculuk hızları da gecikme cezalarından kaçınmak için artabilir. Ayrıca, hava koşulları dışında, denizde yapılan yakıt alım fiyatlarındaki ani değişimler, liman tıkanıklığı nedeniyle bekleme maliyetleri vb. sebepler gemi operatörlerine maddi zarar vermektedir [13]. Ancak limana varış saatlerinin esnek olması bir miktar yakıt tüketimi verimliliği sağlayabilir. Bu amaç doğrultusunda, bu çalışmada genetik algoritma

rutinine bir ve iki saatlik izin verilebilir ETA gecikmeleri tanımlanmıştır. GA ile sadece 15 knot azami hız limiti altında bulunan 1 ve 2 saatlik gecikmeler simüle edilmiştir. Öbür taraftan, yakıt tüketimi üzerindeki etkilerini araştırmak için farklı miktarlarda kalan hız değerleri denenmiştir. Sonuçta, kalan hızın azaltılması ile GA yakıt tasarrufunu ETA'nın yalnızca 1 saatlik gecikme koşulları için her zaman arttırmıştır (bkz. Şekil 29). Ancak 2 saatlik gecikme şartlarında aynı şeyi söylemek mümkün değildir. GA kalan mesafe hızını azaltarak minimum yakıt tüketim noktasını aramaktadır. Ancak, minimum yakıt tüketimi sadece kalan mesafedeki en düşük hız için oluşmamaktadır (bkz. Tablo 12). Bu nedenle, minimum yakıt tüketimi için kalan mesafe hızı, üst hız limiti ve ETA'nın her defasında optimum kombinasyonları başka bir optimizasyon rutiniyle araştırılmalıdır.

Tablo 12 Rastgele ilklendirmede 1 ve 2 saatlik gecikmeler için yakıt tüketimi tasarrufları

Kalan Mesafe Hızı	1 Saat Gecikmede	2 Saat Gecikmede
12	% 4,78	% 4,18
11	% 5,12	% 4,88
9	% 5,52	**% 5,90**
8	**% 6,00**	% 4,22



Şekil 29. ETA gecikmesi 1 saat (sol) ve 2 saat (sağ) için yakıt tüketimi sonuçları

4.3.2. Manuel İlklendirme Sonuçları

Bu bölümde, rastgele yaklaşımda kullanılan bireylerin ilklendirilmesi dışındaki diğer tüm parametreler aynı şekilde uygulanır. Tek fark ise bir popülasyonda birey oluştururken hız azaltımı miktarı ve süresi önceden belirlenmiş miktarlarda değiştirilmiştir. Her adımda değiştirilen parametreler Tablo 12'de sunulmuştur. Ancak, hız profillerinde değişikliğin başlangıç zamanı, hem rastgele hem de manuel yaklaşımlarda her zaman rastgele seçilmiştir.

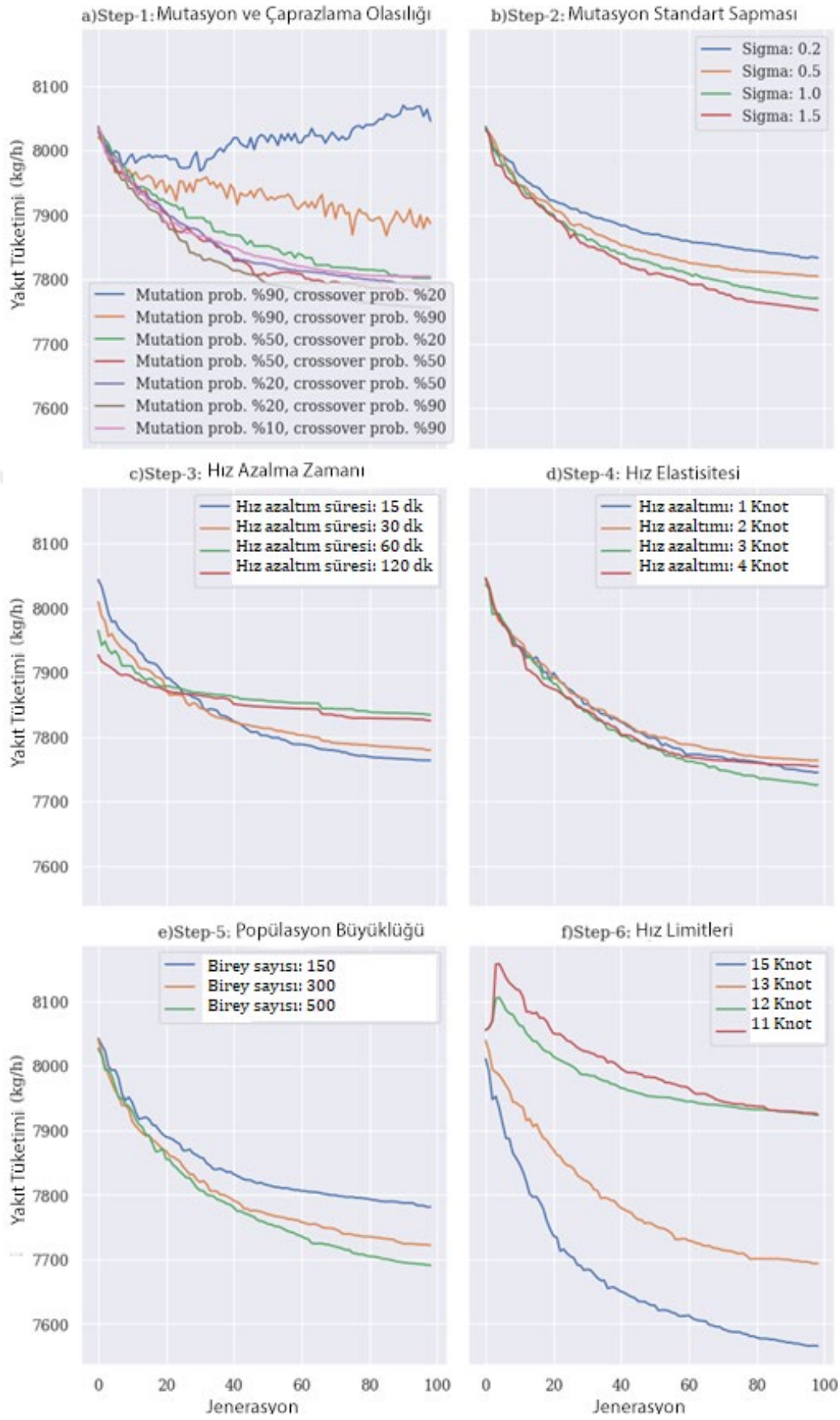
Tablo 13 Manuel olarak başlatılan GA'da kontrol parametreleri ve sabit

Adım	Mutasyon olasılığı	Çağrazlama olasılığı	Sigma	Hız azaltım süresi (mins)	Hız azaltımı (kts)	Popülasyon büyüklüğü	Hız limiti (mile/h)
1	değişken	değişken	1.5	30	2	300	13
2	% 20	% 90	değişken	30	2	300	13
3	% 20	% 90	1.5	değişken	2	300	13
4	% 20	% 90	1.5	15	değişken	300	13
5	% 20	% 90	1.5	15	3	değişken	13
6	**% 20**	**% 90**	**1.5**	**15**	**3**	**500**	değişken

4.3.2.1. Manuel İlklendirmede Hiperparametre Optimizasyonu Sonuçları

GA simülasyonları, Tablo 12'de gösterilen adımlara ait kontrol parametreleri değiştirilerek çok kez tekrarlanmıştır. Ayrıca, simülasyonların yakıt tüketim sonuçları Şekil 30'da verilmiştir. Özetle, mutasyon, çaprazlama ve popülasyon büyüklüğü manuel ve rastgele methodlarda aynı sonuçları vermiştir. Mutasyon ve çaprazlama olasılıkları sırasıyla yüzde 20 ve 90 ve 500 adet popülasyon bireyi ile GA daha fazla yakıt tasarrufu sağlamıştır (bkz. Şekil 30a/e). Mutasyondaki standart sapma (sigma) değerleri artan yönde daha iyi sonuç verdiğinden sonraki en iyileme adımları için sigma değeri 1.5 olarak seçilmiştir (bkz. Şekil 30b).

Hız azaltımı 15 ile 120 dakika ve zaman aralığı ise 1 ile 4 knot arasında kademeli olarak arttırılmıştır. Düşük hız azaltma süreleri ve yüksek hız azaltma miktarlarıyla daha iyi sonuçlar alınmıştır (bkz. Şekil 30c/d). Önemli bir bulgu olarak, manuel ilklendirme yönteminde hız azaltım miktarı ve süresinin birlikte artığı durumlarda kalan mesafe artabilmektedir. Ancak, kalan mesafe için zaman gecikmesi de kalan mesafe hızına doğrudan bağlıdır. Son olarak, maksimum hız limiti GA kısıtlaması olarak uygulanmıştır. Ayrıca, hız limitleri için iki saatlik izin verilebilir gecikme kullanılmış olup bu çalışmada daha yüksek hız limitleri daha fazla yakıt tasarrufu ile sonuçlanmıştır (bkz. Tablo 14).



Şekil 30. Manuel ilklendirme yönteminde parametre en iyilenmesi

Tablo 14 Manuel ilklendirmede hız limitlerinin yakıt tasarrufları

Hız Limiti	Yakıt Tasarrufu
15	**% 5,42**
13	% 3,81
12	% 0,93
11	% 0,93

Manuel ilklendirmede, GA'nın yakıt tüketimi tasarrufunu yapabilmesi için hız azaltımının miktarı ve süresinin izin verilebilir gecikme sürelerine göre GA'nın başında güncellenmesi önemlidir. Aksi taktirde, izin verilebilir gecikme süreleri yeteri kadar uzun olsa bile, GA olası düşük hız profillerini oluşturamadığı için olası en çok yakıt tüketimi tasarrufu verecek hız profillerini oluşturamamaktadır. Ayrıca, GA performansı genel olarak bireylerin ilklendirmesine bağlıdır [14]. Bu çalışmada da hız profillerinin ilklendirilmesinde çeşitliliğin artmadığı durumlarda 1 ve 2 saatlik gecikmenin birbirinden çok farklı sonuçlar üretmediği görülmüştür.

5. SONUÇ VE ÖNERİLER

Bu çalışma, gemiler için hız optimizasyonu yoluyla yakıt tüketimi verimliliği potansiyelini keşfetmeyi amaçlamıştır. Yakıt tüketimi tahmini için doğrusal ve doğrusal olmayan makine öğrenimi regresyon modellerini analiz eden bu tez, gemilerin çeşitli çevresel, operasyonel ve seyir veri günlüklerini kullanarak yakıt tüketiminin nasıl tahmin edilebileceğini göstermiştir. Ham verilere veri ön işleme teknikleri uygulanarak, veri kayıtlarının çoğunda eksik değerler ve yakıt tüketimi ile ilgisiz kayıtlar olduğu gözlemlenmiştir. Bu nedenle, verileri regresyon modellemesine hazırlamak için öznitelik silme ve veri atama teknikleri uygulanmıştır. Öte yandan, orijinal veri kümesinin örnekleme oranı bir saniye olduğu için yüksek hesaplama maliyetleri ve uzun çalışma süreleri oluşabileceği öngörülerek 24 saatlik veri kümesi için örnekleme oranı 30 saniyeye düşürülmüştür. LR, SVR ve KNN yakıt tüketimi modelleri, sırasıyla %90'ın altında en düşük eğitim veri kümesi doğruluğuyla sonuçlanmıştır. Ancak, gradyan artırmaya dayalı tahmin modelleri en yüksek doğrulukları göstermiştir, bununla birlikte yüksek oranda aşırı öğrenme içermektedirler.

Bu çalışma ayrıca, gemilerin hız optimizasyonda kullanılan genetik algoritmalarda rastgele modifikasyonlar ile oluşturulmuş hız profillerinin daha fazla yakıt tüketimi ile sonuçlandığını göstermiştir. Ayrıca, GA'daki yüksek çaprazlama oranları ve popülasyon boyutları, hız limitleri ve düşük mutasyon oranları yakıt tüketimi tasarrufunda daha iyi sonuçlar vermiştir. Ek olarak, bu çalışmanın sonuçlarına bakarak, gemi seyir hız profili dikkate alınmadan hız limitlerinin belirlenmesinin, GA ile daha yüksek yakıt tüketimine neden olabileceği söylenebilir.

GA ile hız optimizasyonu, rastgele seçim kullanan çoklu parametre optimizasyonu ile daha verimli olabilir. Örneğin bu tez kapsamında yer almayan liman gecikmeleri, hava durumu tahminleri, dalga koşulları, bunker, operasyonel emirler vb. yakıt tüketimini etkileyen parametreler GA amaç fonksiyonunda kullanıldığında daha gerçekçi sonuçlar elde edilebilir.

KAYNAKLAR

1. Faber, J., et al., *Fourth IMO GHG Study.* International Maritime Organization, London, UK, 2020.

2. IMO, *Guidelines for the development of a Ship Energy Efficiency Management Plan (SEEMP)*, in *Resolution MEPC.282(70)*. 2016.

3. IMO. *Energy Efficiency Measures*. 2022 2022-05-06]; Available from: https://www.imo.org/en/OurWork/Environment/Pages/Technical-and-Operational-Measures.aspx.

4. IMO. *Data collection system for fuel oil consumption of ships*. 2022 2022-05-06]; Available from: https://www.imo.org/en/OurWork/Environment/Pages/Data-Collection-System.aspx.

5. Faber, J., et al., *Regulated Slow Steaming in Maritime Transport. An assessment of options, costs and benefits.* 2012.

6. förstudie initierad av Lighthouse, E., *Consequences of speed reductions for ships.*

7. Aldous, L.G., *Ship operational efficiency: performance models and uncertainty analysis*. 2016, UCL (University College London).

8. Roar, A., C. Pierre, and W. Francois-Charles, *Optimal ship speed and the cubic law revisited: Empirical evidence from an oil tanker fleet.* Transportation Research Part E: Logistics and Transportation Review, 2020. **140**: p. 101972.

9. Kim, Y.-R., M. Jung, and J.-B. Park, *Development of a Fuel Consumption Prediction Model Based on Machine Learning Using Ship In-Service Data.* Journal of Marine Science and Engineering, 2021. **9**(2): p. 137.

10. Andrea, C., et al., *Vessels fuel consumption forecast and trim optimisation: A data analytics perspective.* Ocean Engineering, 2017. **130**: p. 351-370.

11. Christos, G., L. Iraklis, and T. Gerasimos, *Machine learning models for predicting ship main engine Fuel Oil Consumption: A comparative study.* Ocean Engineering, 2019. **188**: p. 106282.

12. Yang, L., et al., *Ship Speed Optimization Considering Ocean Currents to Enhance Environmental Sustainability in Maritime Shipping.* Sustainability, 2020. **12**(9): p. 3649.

13. Aydin, N., H. Lee, and S.A. Mansouri, *Speed optimization and bunkering in liner shipping in the presence of uncertain service times and time windows at ports.* European Journal of Operational Research, 2017. **259**(1): p. 143-154.

14. Helong, W., L. Xiao, and M. Wengang, *Voyage optimization combining genetic algorithm and dynamic programming for fuel/emissions reduction.* Transportation Research Part D: Transport and Environment, 2021. **90**: p. 102670.

15. Hüffmeier, J., J. Lundman, and F. Elern, *Trim and Ballast Optimisation for a Tanker Based on Machine Learning.* 2020.

16. Xiao, C., *Using machine learning for exploratory data analysis and predictive models on large datasets*, in *Computer science*. 2015, University of Stavanger: Norway.

17. Kang, H., *The prevention and handling of the missing data.* Korean journal of anesthesiology, 2013. **64**(5): p. 402-406.

18. Blázquez-García, A., et al., *A review on outlier/anomaly detection in time series data.* arXiv preprint arXiv:2002.04236, 2020.

19. Breunig, M.M., et al., *LOF: identifying density-based local outliers.* SIGMOD Rec., 2000. **29**(2): p. 93–104.

20. Uyanık, T., Ç. Karatuğ, and Y. Arslanoğlu, *Machine learning approach to ship fuel consumption: A case of container vessel.* Transportation Research Part D: Transport and Environment, 2020. **84**: p. 102389.

21. Abebe, M., et al., *Machine Learning Approaches for Ship Speed Prediction towards Energy Efficient Shipping.* Applied Sciences, 2020. **10**(7): p. 2325.

22. Holmes, A., et al., *The Correlation Coefficient r*, in *STAT 462*. 2022.

23. Karthe, *Going Deeper into Regression Analysis with Assumptions, Plots & Solutions*. 2016.

24. De Andrade, L.C.O., *Genetic Algorithms Application In Line Simplification.* 2014.

25. Vorkapić, A., R. Radonja, and S. Martinčić-Ipšić, *Predicting Seagoing Ship Energy Efficiency from the Operational Data.* Sensors, 2021. **21**(8): p. 2832.

26. Kelleher, J.D., B. Mac Namee, and A. D'arcy, *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. 2020: MIT press.

27. Pedregosa, F., et al., *Scikit-learn: Machine learning in Python.* the Journal of machine Learning research, 2011. **12**: p. 2825-2830.

28. Psaraftis, H., *Speed Optimization vs Speed Reduction: the Choice between Speed Limits and a Bunker Levy.* Sustainability, 2019. **11**: p. 2249.

29. Arslan, O., E. Besikci, and A. Olcer, *Improving energy efficiency of ships through optimisation of ship operations.* No. FY2014-3 IAMU, 2014.

30. Johansson, C. and G. Evertsson, *Optimizing genetic algorithms for time critical problems*. 2003.

31. Chaal, M., *Ship operational performance modelling for voyage optimization through fuel consumption minimization*, in *Maritime Affairs*. 2018, World Maritime University: Malmö, Sweden.

32. Changnan, W., et al. *Optimization Analysis of USV Based on Genetic Algorithm*. in *Proceedings of the 2017 5th International Conference on Mechatronics, Materials, Chemistry and Computer Engineering (ICMMCCE 2017)*. 2017. Atlantis Press.

33. Mosayebi, M. and M. Sodhi, *Tuning genetic algorithm parameters using design of experiments.* 2020: p. 1937-1944.

34. Robert, N. *Regression diagnostics: Testing the assumptions of linear regression*. Statistical forecasting: Notes on regression and time series analysis 2022 1/23/2022]; Available from: https://people.duke.edu/~rnau/testing.htm.

35. Kee, K.K., S.L. Boung Yew, and K.-H. Renco, *Prediction of Ship Fuel Consumption and Speed Curve by Using Statistical Method*. 2018.

36. Hu, Z., et al., *A Novel Hybrid Fuel Consumption Prediction Model for Ocean-Going Container Ships Based on Sensor Data.* Journal of Marine Science and Engineering, 2021. **9**(4): p. 449.

EKLER

Ek-1: Method: Scikit-learn LocalOutlierFactor Algoritması

Girdi parametreleri	Seçenekler	Varsayılan değer
n_neighbors	integer	*overridden*
algorithm	'auto', 'ball_tree', 'kd_tree', 'brute'	auto
metric	'cityblock', 'cosine', 'euclidean', 'l1', 'l2', 'manhattan' 'minkowski'	minkowski
P	integer	2 (Euclidean distance)
contamination	'auto' or float	0.05

İNTİHAL RAPORU İLK SAYFASI

UYARI: Bu bölümde intihal programında çıkan benzerlik yüzdesinin bulunduğu detaylı raporun ilk sayfasının yer alması tüm öğrenciler için zorunludur! Bu uyarıyı okuduktan sonra siliniz!

ETİK KURUL İZİN YAZISI

Uyarı: Canlı denekler üzerinde yapılan tüm araştırmalar için Etik Kurul Belgesi alınması zorunludur.

☐ **Etik Kurul izni gerekmektedir.**

☒ **Etik Kurul izni gerekmemektedir.**

Öğrenci Tarık TAŞPINAR
(İmza)

KURUM İZNİ YAZILARI

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

☐ **Kurum izni gerekmektedir.**

☒ **Kurum izni gerekmemektedir.**

Öğrenci Tarık TAŞPINAR
(İmza)