



**KTO KARATAY ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI
ELEKTRİK VE BİLGİSAYAR MÜHENDİSLİĞİ TEZLİ YÜKSEK LİSANS
PROGRAMI**

**TARIMDA GÜNLÜK AKARYAKIT İHTİYACININ MAKİNE ÖĞRENMESİ
İLE TAHMİNLENMESİ**

Mustafa ÇOBAN

Yüksek Lisans

**KONYA
Temmuz 2021**

TARIMDA GÜNLÜK AKARYAKIT İHTİYACININ MAKİNE ÖĞRENMESİ İLE
TAHMİNLENMESİ

Mustafa ÇOBAN

KTO Karatay Üniversitesi
Lisansüstü Eğitim Enstitüsü
Elektrik-Elektronik Mühendisliği Anabilim Dalı
Elektrik ve Bilgisayar Mühendisliği Tezli Yüksek Lisans Programı

Yüksek Lisans

Tez Danışmanı: Dr. Öğr. Üyesi Semih YUMUŞAK
İkinci Danışman: Dr. Öğr. Üyesi Hüseyin Oktay ALTUN

Konya
Temmuz 2021

KABUL VE ONAY

Mustafa ÇOBAN tarafından hazırlanan “Tarımda Günlük Akaryakıt İhtiyacının Makine Öğrenmesi ile Tahminlenmesi” başlıklı bu çalışma, 14 Temmuz 2021 tarihinde yapılan savunma sınavı sonucunda başarılı bulunarak jürimiz tarafından Yüksek Lisans Tezi olarak kabul edilmiştir.

Jüri Başkanı: **Doç Dr. Kasım TOPRAK** _____
Konya Gıda ve Tarım Üniversitesi

Jüri Üyesi: **Dr. Öğr. Üyesi Ali Osman ÇIBIKDİKEN** _____
KTO Karatay Üniversitesi

Tez Danışmanı: **Dr. Öğr. Üyesi Semih YUMUŞAK** _____
KTO Karatay Üniversitesi

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

Prof. Dr. Hüseyin Bekir YILDIZ
Enstitü Müdürü

BİLDİRİM

Enstitü tarafından onaylanan Yüksek Lisans tezimin tamamını veya herhangi bir kısmını basılı veya dijital biçimde arşivleme ve aşağıda belirtilen koşullar dahilinde erişime açma iznini KTO Karatay Üniversitesine verdiğimi bildiririm. Bu izinle, Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak ve gelecekteki çalışmalar (makale, kitap, lisans, patent vb.) için tezimin taamının veya bir bölümünün kullanım hakları yalnızca bana ait olacaktır.

Tezimin bütünüyle kendi çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Telif hakkı bulunan ve sahiplerinden yazılı izinle kullanılması zorunlu olan kaynakları, yazılı izin alarak kullandığımı ve istenildiğinde izinlerin suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayımlanan “Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge” kapsamında, tezim, aşağıda belirtilen koşullar haricince, YÖK Ulusal Tez Merkezi ve KTO Karatay Üniversitesi Açık Erişim Sisteminde erişime açılır.

☐ Enstitü / Fakülte Yönetim Kurulu kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren 2 yıl ertelenmiştir.¹

☐ Enstitü / Fakülte Yönetim Kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren ... ay ertelenmiştir.²

☐ Tezimle ilgili gizlilik kararı verilmiştir.³⁴

14 Temmuz 2021

Mustafa ÇOBAN

¹ MADDE 6(1) Lisansüstü teze ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir.

² MADDE 6(2) Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internetten paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç imkanı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.

³ MADDE 7(1) Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.

⁴ MADDE 7(2) Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir.

ETİK BEYAN

KTO Karatay Üniversitesi Lisansüstü Eğitim Enstitüsü Tez Hazırlama ve Yazım Kurallarına uygun olarak Dr. Öğr. Üyesi Semih YUMUŞAK danışmanlığında tarafımdan üretilen bu tez çalışmasında; sunduğum tüm veri, enformasyon, bilgi ve belgeleri bilimsel etik kuralları çerçevesinde elde ettiğimi, tüm değerlendirme, analiz, bulgu ve sonuçları bilimsel usullere uygun olarak sunduğumu, tez çalışmasında yararlandığım kaynakların tümüne bilimsel normlara uygun biçimde atıfta bulunarak kaynak gösterdiğimi, tezimin kaynak gösterilen durumlar dışında özgün olduğunu bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

14 Temmuz 2021

Mustafa ÇOBAN

TEŞEKKÜR

Tez çalışmam boyunca yardımlarını esirgemeyen ve beni yönlendiren danışman hocalarım Dr. Öğr. Üyesi Semih YUMUŞAK ve Dr. Öğr. Üyesi Hüseyin Oktay ALTUN'a teşekkürlerimi ve saygımı sunarım. Ayrıca vaktini ayırıp desteğini sağlayan Arş. Gör. Şüheda SEYHAN'a teşekkür ederim.

Temmuz, 2021

Mustafa ÇOBAN

ÖZET

Mustafa ÇOBAN

Tarımda Günlük Akaryakıt İhtiyacının Makine Öğrenmesi İle Tahminlenmesi

Yüksek Lisans

Konya, 2021

Tarımsal faaliyetler için kullanılan akaryakıt miktarı çeşitli etkilere göre değişim göstermektedir. Gelecekte kullanılacak olan akaryakıtın tahmin edilmesi, dağıtım firmalarının stok takibi veya lojistik ihtiyacını iyileştirmesinde katkı sağlayabilmektedir. Bu tez çalışmasında, bir sonraki güne ait akaryakıt kullanımı olup olmayacağı araştırılmış ve akaryakıt tahminleme için makine öğrenmesi algoritmaları kullanılarak karşılaştırmalı olarak analiz edilmiştir. Bunun yanı sıra, akaryakıt kullanımını etkileyen dış etkilere araştırılmıştır. Bu dış etkilerin (örn. hava durumu, akaryakıt ücreti) çiftçilerin aldığı akaryakıt miktarını etkileyip etkilemediğini görmek adına havanın yağmurlu olup olmadığı, akaryakıtın indirim veya zam alıp almadığı, hava sıcaklığındaki değişim gibi değerler veri kümesine eklenerek etkileri gözlemlenmiştir. Akaryakıt ücretlerine oluşan değişimlerin, bir sonraki gün alımlarında yüksek oranda etkisi olduğu, ayrıca yağmur yağış bilgisinin de önemli bir özellik olduğu tespit edilmiştir. Analizlerde K-en yakın komşu (K-Nearest neighbors), karar ağaçları regresyon (decision tree regression), karar ağaçları sınıflandırması (decision tree classifier), lojistik regresyon (logistic regression), destek vektör makinesi (support vector machine), rassal orman sınıflandırması (random forest classifier), sinir ağı (neural network) ve ekstrem artırılmış gradyan (Extreme Gradient Boosting, XGB) makine öğrenmesi algoritmaları uygulanmıştır. Algoritmaların kıyaslanmasında doğruluk, duyarlılık, kesinlik ve f1 skoru değerleri kullanılmıştır. Sonuçların doğruluğu için hata matrisleri kontrol edilmiştir. Buna göre en yüksek doğruluk skorunu XGB algoritması %82,514 skor ile olmuştur. Bu bağlamda, oluşturulan veri kümesinin ve önerilen veri zenginleştirme yöntemlerinin makine öğrenmesi algoritmalarına uyarlanabilir olduğu sonucuna ulaşılmıştır.

Anahtar Kelimeler:

Veri seti hazırlanması, makine öğrenmesi, akaryakıt kullanım tahmini

ABSTRACT

Mustafa ÇOBAN

Prediction of Daily Fuel Needs in Agriculture with Machine Learning

Master's Degree

Konya, 2021

The amount of fuel used for agricultural activities varies according to various effects. Estimating the fuel to be used in the future can contribute to the improvement of the inventory tracking or logistics needs of the distribution companies. In this thesis, it was investigated whether there would be fuel use for the next day and comparatively analyzed using machine learning algorithms for fuel estimation. In addition, external effects affecting the use of fuel were investigated. In order to see whether these external effects (e.g. weather, fuel price) affect the amount of fuel purchased by the farmers, values such as whether it is rainy, whether the fuel is discounted or increased, and the change in air temperature are added to the dataset and their effects are observed. It has been determined that the changes in fuel prices have a high effect on the purchases of the next day, and rain precipitation information is also an important feature. In the analysis, K-Nearest neighbors, decision tree regression, decision tree classifier, logistic regression, support vector machine, random forest classifier, neural network and Extreme Gradient Boosting, XGB machine learning algorithms were applied. Accuracy, sensitivity, precision and f1 score values were used to compare the algorithms. Confusion matrices were checked for the accuracy of the results. Accordingly, the highest accuracy score was the XGB algorithm with a score of 82.514%. In this context, it has been concluded that the created dataset and the proposed data enrichment methods are adaptable to machine learning algorithms.

Keywords:

Dataset preparation, machine learning, fuel usage forecast

İÇİNDEKİLER

KABUL VE ONAY	i
BİLDİRİM	ii
ETİK BEYAN.....	iii
TEŞEKKÜR.....	iv
ÖZET	v
ABSTRACT.....	vi
TABLolar DİZİNİ	ix
ŞEKİLLER DİZİNİ.....	xi
KISALTMALAR DİZİNİ.....	xvii
1. GİRİŞ	1
2. LİTERATÜR	2
3. METODOLOJİ	4
3.1. Analiz Yöntemleri	4
3.1.1. Veri Ön İşleme ve Genişletme.....	4
3.1.2. Önemli Veri Özelliklerinin Bulunması (Feature Importance).....	7
3.1.3. Yüzdelik Ayrım Metodu.....	16
3.1.4. Sıralı Veri Kullanımı	16
3.1.5. Izgara Araması Parametre Optimizasyon Yöntemi (Grid Search Hyperparameter Optimization).....	17
4. DENEYSEL ÇALIŞMA	22
4.1. 2019-2020 Verisi Üzerinden Tam Yıl Analizleri: Mevcut Gün Tahmini	22
4.1.1. Sıralı Veri Tahminleme Analizi.....	22
4.1.2. Yüzdelik Ayrım Analizi	30
4.1.3. Karşılaştırmalı Algoritma Skorları	37
4.1.4. Akaryakıt Miktarı Tahminleme Skorları	40
4.2. 2019-2020 Verisi kullanılarak 2021 ilk çeyrek tahmini: Mevcut Gün Tahmini	41
4.2.1. Sıralı Veri Tahminleme Analizi.....	41
4.2.2. Yüzdelik Ayrım Analizi	48
4.2.3. Karşılaştırmalı Algoritma Skorları	57
4.2.4. Akaryakıt Miktarı Tahminleme Skorları	59
4.3. 2019-2020 Verisi Üzerinden Tam Yıl Analizleri: Sonraki Gün Tahmini.....	60
4.3.1. Sıralı Veri Tahminleme Analizi.....	60
4.3.2. Yüzdelik Ayrım Analizi	68

4.3.3. Karşılaştırmalı Algoritma Skorları	75
4.3.4. Akaryakıt Miktarı Tahminleme Skorları	78
4.4. 2019-2020 Verisi Kullanılarak 2021 İlk Çeyrek Tahmini: Sonraki Gün Tahmini	79
4.4.1. Sıralı Veri Tahminleme Analizi.....	79
4.4.2. Yüzdelik Ayrım Analizi	86
4.4.3. Karşılaştırmalı Algoritma Skorları	94
4.4.4. Akaryakıt Miktarı Tahminleme Skorları	96
5. SONUÇ	98
KAYNAKLAR	100
ÖZGEÇMİŞ	102
EK 1. HATA MATRİSLERİ	103
5.1. 2019-2020 Verisi Üzerinden Tam Yıl Analizleri: Mevcut Gün Tahmini	103
5.1.1. Sıralı Veri Tahminleme Analizi.....	103
5.1.2. Yüzdelik Ayrım Analizi	104
5.2. 2019-2020 Verisi kullanılarak 2021 ilk çeyrek tahmini: Mevcut Gün Tahmini	106
5.2.1. Sıralı Veri Tahminleme Analizi.....	106
5.2.2. Yüzdelik Ayrım Analizi	108
5.3. 2019-2020 Verisi Üzerinden Tam Yıl Analizleri: Sonraki Gün Tahmini.....	110
5.3.1. Sıralı Veri Tahminleme Analizi.....	110
5.3.2. Yüzdelik Ayrım Analizi	112
5.4. 2019-2020 Verisi kullanılarak 2021 ilk çeyrek tahmini: Sonraki Gün Tahmini	114
5.4.1. Sıralı Veri Tahminleme Analizi.....	114
5.4.2. Yüzdelik Ayrım Analizi	116

TABLÖLAR DİZİNİ

Tablo 1. DTC Algoritmasının Önemli Özellik Sıralaması ve Adları (Üye Numaraları Maskelenmiştir).....	9
Tablo 2. DTC hiper parametreleri	18
Tablo 3. DTC* hiper parametreleri	18
Tablo 4. DTR hiper parametreleri	19
Tablo 5. DTR* hiper parametreleri	19
Tablo 6. KNN hiper parametreleri	19
Tablo 7. KNN* hiper parametreleri	19
Tablo 8. NN hiper parametreleri	19
Tablo 9. NN* hiper parametreleri	20
Tablo 10. RFC hiper parametreleri	20
Tablo 11. RFC* hiper parametreleri	20
Tablo 12. SVM hiper parametreleri	20
Tablo 13. SVM* hiper parametreleri	20
Tablo 14. XGB hiper parametreleri	21
Tablo 15. XGB* hiper parametreleri	21
Tablo 16. LR hiper parametreleri	21
Tablo 17. LR* hiper parametreleri	21
Tablo 18. Algoritmaların Litre Tahminleme İçin Aldıkları MSE Skorları	40
Tablo 19. Algoritmaların Litre Tahminleme İçin Aldıkları MSE Skorları	59
Tablo 20. Algoritmaların Litre Tahminleme İçin Aldıkları MSE Skorları	78
Tablo 21. Algoritmaların Litre Tahminleme İçin Aldıkları MSE Skorları	96
Tablo 22. Algoritmaların sıralı veri tahminleme yöntemi ile kullanımı	103
Tablo 23. Algoritmaların sıralı veri tahminleme ve hiper parametre optimizasyonu yöntemi ile kullanımı	104
Tablo 24. Algoritmaların yüzdelerle ayırımı yöntemi ile kullanımı	105
Tablo 25. Algoritmaların yüzdelerle ayırımı yöntemi ve hiper parametre ile kullanımı ...	106
Tablo 26. Algoritmaların sıralı veri tahminleme yöntemi kullanımı	107
Tablo 27. Algoritmaların sıralı veri tahminleme ve hiper parametre optimizasyonu yöntemi ile kullanımı	108
Tablo 28. Algoritmaların yüzdelerle ayırımı yöntemi ile kullanımı	109
Tablo 29. Algoritmaların yüzdelerle ayırımı yöntemi ve hiper parametre ile kullanımı ...	110
Tablo 30. Algoritmaların sıralı veri tahminleme metodu ile kullanımı	111

Tablo 31. Algoritmaların sıralı veri tahminleme ve hiper parametre optimizasyonu yöntemi ile kullanımı	112
Tablo 32. Algoritmaların yüzdellik ayırım yöntemi ile kullanımı	113
Tablo 33. Algoritmaların yüzdellik ayırım yöntemi ve hiper parametre ile kullanımı ...	114
Tablo 34. Algoritmaların sıralı veri tahminleme metodu ile kullanımı	115
Tablo 35. Algoritmaların sıralı veri tahminleme ve hiper parametre optimizasyonu yöntemi ile kullanımı	116
Tablo 36. Algoritmaların yüzdellik ayırım yöntemi ile kullanımı	117
Tablo 37. Algoritmaların yüzdellik ayırım yöntemi ve hiper parametre ile kullanımı ...	118

ŞEKİLLER DİZİNİ

Şekil 1. Tarih genişletme işlemi günlerin karşılığı	5
Şekil 2. Tarih birleştirme ve tatil günleri hesaplamaları Python kod bloğu.....	6
Şekil 3. Üyelerin aldıkları akaryakıtın veriye işlenmesi için Python kod bloğu.....	6
Şekil 4. Önemli özellik tespiti kodu örneği.....	7
Şekil 5. DTR normal veri grafik üzeri önem göstergesi	8
Şekil 6. DTR* önemli veri ayıklaması sonucu özelliklerin önem seviyeleri	8
Şekil 7. DTC normal veri grafik üzeri önem göstergesi	8
Şekil 8. DTC* önemli veri ayıklaması sonucu özelliklerin önem seviyeleri	9
Şekil 9. KNN normal veri özelliklerin önem seviyeleri.....	10
Şekil 10. KNN* önemli veri ayıklaması sonucu özelliklerin önem seviyeleri	11
Şekil 11. LR normal veri özelliklerin önem seviyeleri	11
Şekil 12. LR* önemli veri ayıklaması sonucu özelliklerin önem seviyeleri.....	12
Şekil 13. NN normal veri özelliklerin önem seviyeleri.....	12
Şekil 14. NN* önemli veri ayıklaması sonucu özelliklerin önem seviyeleri	13
Şekil 15. RFC normal veri özelliklerin önem seviyeleri.....	13
Şekil 16. RFC* önemli veri ayıklaması sonucu özelliklerin önem seviyeleri	14
Şekil 17. SVM normal veri özelliklerin önem seviyeleri.....	14
Şekil 18. SVM* önemli veri ayıklaması sonucu özelliklerin önem seviyeleri	15
Şekil 19. XGB normal veri özelliklerin önem seviyeleri.....	15
Şekil 20. XGB* önemli veri ayıklaması sonucu özelliklerin önem seviyeleri	16
Şekil 21. Örnek algoritma yüzdelik ayırım metodu	16
Şekil 22. Örnek algoritma sıralı veri metodu kullanımı.....	17
Şekil 23. Örnek algoritma için ızgara araması uygulanması.....	18
Şekil 24. Algoritmaların sıralı veri tahminleme metodu ile doğruluk skor kıyaslanması	23
Şekil 25. Algoritmaların sıralı veri tahminleme metodu ile kesinlik skor kıyaslanması	24
Şekil 26. Algoritmaların sıralı veri tahminleme metodu ile duyarlılık skor kıyaslanması	24
Şekil 27. Algoritmaların sıralı veri tahminleme metodu ile f1 skor kıyaslanması.....	25
Şekil 28. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu doğruluk skor kıyaslanması.....	25
Şekil 29. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu kesinlik skor kıyaslanması	26

Şekil 30. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu duyarlılık skor kıyaslanması.....	26
Şekil 31. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu f1 skor kıyaslanması.....	27
Şekil 32. Algoritmaların sıralı veri tahminleme yöntemi ile kullanımında doğruluk skoru kıyaslanması.....	28
Şekil 33. Algoritmaların sıralı veri tahminleme metodu ile kullanımında kesinlik skoru kıyaslanması.....	28
Şekil 34. Algoritmaların sıralı veri tahminleme metodu ile kullanımında duyarlılık skoru kıyaslanması.....	29
Şekil 35. Algoritmaların sıralı veri tahminleme metodu ile kullanımında f1 skoru kıyaslanması.....	30
Şekil 36. Algoritmaların yüzdelik ayırım yöntemi kullanımı ile doğruluk skor kıyaslanması.....	31
Şekil 37. Algoritmaların yüzdelik ayırım yöntemi kullanımı ile kesinlik skor kıyaslanması.....	31
Şekil 38. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu doğruluk skor kıyaslanması.....	32
Şekil 39. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu kesinlik skor kıyaslanması	32
Şekil 40. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu duyarlılık skor kıyaslanması.....	33
Şekil 41. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu f1 skor kıyaslanması.....	33
Şekil 42. Algoritmaların yüzdelik ayırım yöntemi kullanımı ile duyarlılık skor kıyaslanması.....	34
Şekil 43. Algoritmaların yüzdelik ayırım yöntemi kullanımı ile f1 skor kıyaslanması ...	34
Şekil 44. Algoritmaların yüzdelik ayırımı yöntemi ile kullanımında doğruluk skoru kıyaslanması.....	35
Şekil 45. Algoritmaların yüzdelik ayırımı yöntemi ile kullanımında kesinlik skoru kıyaslanması.....	36
Şekil 46. Algoritmaların yüzdelik ayırımı yöntemi ile kullanımında duyarlılık skoru kıyaslanması.....	36
Şekil 47. Algoritmaların yüzdelik ayırımı yöntemi ile kullanımında f1 skoru kıyaslanması.....	37
Şekil 48. Doğruluk skorlarının kıyaslanması	38
Şekil 49. Kesinlik skorlarının kıyaslanması.....	38
Şekil 50. Duyarlılık skorlarının kıyaslanması	39
Şekil 51. F1 skorlarının kıyaslanması	40

Şekil 52. Veri Ve XGBoost Algoritmasının Tahmininin Grafiksel Gösterimi.....	41
Şekil 53. Algoritmaların sıralı veri tahminleme metodu ile doğruluk skor kıyaslanması	42
Şekil 54. Algoritmaların sıralı veri tahminleme metodu ile kesinlik skor kıyaslanması	42
Şekil 55. Algoritmaların sıralı veri tahminleme metodu ile duyarlılık skor kıyaslanması	43
Şekil 56. Algoritmaların sıralı veri tahminleme metodu ile f1 skor kıyaslanması.....	43
Şekil 57. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu doğruluk skor kıyaslanması.....	44
Şekil 58. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu kesinlik skor kıyaslanması	44
Şekil 59. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu duyarlılık skor kıyaslanması.....	45
Şekil 60. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu f1 skor kıyaslanması.....	45
Şekil 61. Algoritmaların sıralı veri tahminleme metodu ile kullanımında doğruluk skoru kıyaslanması.....	46
Şekil 62. Algoritmaların sıralı veri tahminleme metodu ile kullanımında kesinlik skoru kıyaslanması.....	47
Şekil 63. Algoritmaların sıralı veri tahminleme metodu ile kullanımında duyarlılık skoru kıyaslanması.....	47
Şekil 64. Algoritmaların sıralı veri tahminleme metodu ile kullanımında f1 skoru kıyaslanması.....	48
Şekil 65. Algoritmaların yüzdelik ayırım yöntemi kullanımı ile doğruluk skor kıyaslanması.....	49
Şekil 66. Algoritmaların yüzdelik ayırım yöntemi kullanımı ile kesinlik skor kıyaslanması.....	49
Şekil 67. Algoritmaların yüzdelik ayırım yöntemi kullanımı ile duyarlılık skor kıyaslanması.....	50
Şekil 68. Algoritmaların yüzdelik ayırım yöntemi kullanımı ile f1 skor kıyaslanması ...	50
Şekil 69. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu doğruluk skor kıyaslanması.....	51
Şekil 70. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu kesinlik skor kıyaslanması	52
Şekil 71. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu duyarlılık skor kıyaslanması.....	52
Şekil 72. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu f1 skor kıyaslanması.....	53

Şekil 73. Algoritmaların yüzdelik ayrımı yöntemi ile kullanımında doğruluk skoru kıyaslanması.....	54
Şekil 74. Algoritmaların yüzdelik ayrımı yöntemi ile kullanımında kesinlik skoru kıyaslanması.....	55
Şekil 75. Algoritmaların yüzdelik ayrımı yöntemi ile kullanımında duyarlılık skoru kıyaslanması.....	56
Şekil 76. Algoritmaların yüzdelik ayrımı yöntemi ile kullanımında f1 skoru kıyaslanması.....	56
Şekil 77. Doğruluk skorlarının kıyaslanması.....	57
Şekil 78. Kesinlik skorlarının kıyaslanması.....	58
Şekil 79. Duyarlılık skorlarının kıyaslanması.....	58
Şekil 80. F1 skorlarının kıyaslanması.....	59
Şekil 81. Veri Ve Svm* Algoritmasının Tahmininin Grafıksel Gösterimi.....	60
Şekil 82. Algoritmaların sıralı veri tahminleme metodu ile doğruluk skor kıyaslanması.....	61
Şekil 83. Algoritmaların sıralı veri tahminleme metodu ile kesinlik skor kıyaslanması.....	62
Şekil 84. Algoritmaların sıralı veri tahminleme metodu ile duyarlılık skor kıyaslanması.....	62
Şekil 85. Algoritmaların sıralı veri tahminleme metodu ile f1 skor kıyaslanması.....	63
Şekil 86. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu doğruluk skor kıyaslanması.....	63
Şekil 87. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu kesinlik skor kıyaslanması.....	64
Şekil 88. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu duyarlılık skor kıyaslanması.....	64
Şekil 89. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu f1 skor kıyaslanması.....	65
Şekil 90. Algoritmaların sıralı veri tahminleme metodu ile kullanımında doğruluk skoru kıyaslanması.....	66
Şekil 91. Algoritmaların sıralı veri tahminleme metodu ile kullanımında kesinlik skoru kıyaslanması.....	66
Şekil 92. Algoritmaların sıralı veri tahminleme metodu ile kullanımında duyarlılık skoru kıyaslanması.....	67
Şekil 93. Algoritmaların sıralı veri tahminleme metodu ile kullanımında f1 skoru kıyaslanması.....	68
Şekil 94. Algoritmaların yüzdelik ayrım yöntemi kullanımı ile doğruluk skor kıyaslanması.....	69
Şekil 95. Algoritmaların yüzdelik ayrım yöntemi kullanımı ile kesinlik skor kıyaslanması.....	69

Şekil 96. Algoritmaların yüzdelik ayırım yöntemi kullanımı ile duyarlılık skor kıyaslanması.....	70
Şekil 97. Algoritmaların yüzdelik ayırım yöntemi kullanımı ile f1 skor kıyaslanması ...	70
Şekil 98. Algoritmaların hiper parametre tespitleriyle yüzdelik ayırım yöntemi kullanımı, doğruluk skor kıyaslanması.....	71
Şekil 99. Algoritmaların hiper parametre tespitleriyle yüzdelik ayırım yöntemi kullanımı, kesinlik skor kıyaslanması	71
Şekil 100. Algoritmaların hiper parametre tespitleriyle yüzdelik ayırım yöntemi kullanımı, duyarlılık skor kıyaslanması	72
Şekil 101. Algoritmaların hiper parametre tespitleriyle yüzdelik ayırım yöntemi kullanımı, f1 skor kıyaslanması	72
Şekil 102. Algoritmaların yüzdelik ayırımı yöntemi ile kullanımında doğruluk skoru kıyaslanması.....	73
Şekil 103. Algoritmaların yüzdelik ayırımı yöntemi ile kullanımında kesinlik skoru kıyaslanması.....	74
Şekil 104. Algoritmaların yüzdelik ayırımı yöntemi ile kullanımında duyarlılık skoru kıyaslanması.....	74
Şekil 105. Algoritmaların yüzdelik ayırımı yöntemi ile kullanımında f1 skoru kıyaslanması.....	75
Şekil 106. Doğruluk skorlarının kıyaslanması	76
Şekil 107. Kesinlik skorlarının kıyaslanması.....	76
Şekil 108. Duyarlılık skorlarının kıyaslanması.....	77
Şekil 109. F1 skorlarının kıyaslanması	78
Şekil 110. Veri Ve XGBoost Algoritmasının Tahmininin Grafikselleştirilmesi.....	79
Şekil 111. Algoritmaların sıralı veri tahminleme metodu ile doğruluk skor kıyaslanması	80
Şekil 112. Algoritmaların sıralı veri tahminleme metodu ile kesinlik skor kıyaslanması	80
Şekil 113. Algoritmaların sıralı veri tahminleme metodu ile duyarlılık skor kıyaslanması	81
Şekil 114. Algoritmaların sıralı veri tahminleme metodu ile f1 skor kıyaslanması.....	81
Şekil 115. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu doğruluk skor kıyaslanması.....	82
Şekil 116. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu kesinlik skor kıyaslanması	82
Şekil 117. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu duyarlılık skor kıyaslanması.....	83
Şekil 118. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu f1 skor kıyaslanması.....	83

Şekil 119. Algoritmaların sıralı veri tahminleme metodu ile kullanımında doğruluk skoru kıyaslanması.....	84
Şekil 120. Algoritmaların sıralı veri tahminleme metodu ile kullanımında kesinlik skoru kıyaslanması.....	85
Şekil 121. Algoritmaların sıralı veri tahminleme metodu ile kullanımında duyarlılık skoru kıyaslanması.....	85
Şekil 122. Algoritmaların sıralı veri tahminleme metodu ile kullanımında f1 skoru kıyaslanması.....	86
Şekil 123. Algoritmaların yüzdelik ayırım yöntemi kullanımı ile doğruluk skor kıyaslanması.....	87
Şekil 124. Algoritmaların yüzdelik ayırım yöntemi kullanımı ile kesinlik skor kıyaslanması.....	87
Şekil 125. Algoritmaların yüzdelik ayırım yöntemi kullanımı ile duyarlılık skor kıyaslanması.....	88
Şekil 126. Algoritmaların yüzdelik ayırım yöntemi kullanımı ile f1 skor kıyaslanması .	88
Şekil 127. Algoritmaların hiper parametre tespitleriyle yüzdelik ayırım yöntemi kullanımı, doğruluk skor kıyaslanması	89
Şekil 128. Algoritmaların hiper parametre tespitleriyle yüzdelik ayırım yöntemi kullanımı, kesinlik skor kıyaslanması	89
Şekil 129. Algoritmaların hiper parametre tespitleriyle yüzdelik ayırım yöntemi kullanımı, duyarlılık skor kıyaslanması	90
Şekil 130. Algoritmaların hiper parametre tespitleriyle yüzdelik ayırım yöntemi kullanımı, f1 skor kıyaslanması	90
Şekil 131. Algoritmaların yüzdelik ayırımı yöntemi ile kullanımında doğruluk skoru kıyaslanması.....	91
Şekil 132. Algoritmaların yüzdelik ayırımı yöntemi ile kullanımında kesinlik skoru kıyaslanması.....	92
Şekil 133. Algoritmaların yüzdelik ayırımı yöntemi ile kullanımında duyarlılık skoru kıyaslanması.....	93
Şekil 134. Algoritmaların yüzdelik ayırımı yöntemi ile kullanımında f1 skoru kıyaslanması.....	93
Şekil 135. Doğruluk skorlarının kıyaslanması	94
Şekil 136. Kesinlik skorlarının kıyaslanması	95
Şekil 137. Duyarlılık skorlarının kıyaslanması	95
Şekil 138. F1 skorlarının kıyaslanması	96
Şekil 139. Veri Ve XGBoost* Algoritmasının Tahmininin Grafiksel Gösterimi.....	97

KISALTMALAR DİZİNİ

Kısaltma	Açıklama
DTC	Decision Tree Classifier (karar ağaçları sınıflandırması) Makine Öğrenmesi Algoritması
DTR	Decision Tree Regression (karar ağaçları regresyonu) Makine Öğrenmesi Algoritması
KNN	K-Nearest Neighbors (k-en yakın komşu) Makine Öğrenmesi Algoritması
LR	Logistic Regression (lojistik regresyon) Makine Öğrenmesi Algoritması
RFC	Random Forest Classifier (rassal orman sınıflandırması) Makine Öğrenmesi Algoritması
SVM	Support Vector (destek vektör makinesi) Machine Makine Öğrenmesi Algoritması
NN	Neural Network (sinir ağı) Makine Öğrenmesi Algoritması
XGB	Extreme Gradient Boosting (ekstrem artırılmış gradyan) Makine Öğrenmesi Algoritması

1. GİRİŞ

Ülkelerin temel yapı taşlarından birisi olan tarım sektöründe akaryakıt kullanımı günümüzde zorunlu ihtiyaçtır. Türkiye devleti coğrafyasında farklı rakım ve ulaşımı zor olan yerlerde tarımsal faaliyetler gerçekleşmektedir. Çiftçinin akaryakıt ihtiyacının karşılanamadığı durumlarda faaliyeti aksayıp durabilmektedir. Bu tez ile çiftçinin akaryakıt ihtiyacının önceden tahmin edilmesi amaçlanmıştır. Tez içerisinde işlenen bu veri daha önce işlenmemiştir. İlk kez makine öğrenmesi algoritmaları ile analizi yapılan bu veri, kurumlara ve çiftçiye uygulama üzerinde yardımcı olacaktır.

Bu tez çalışmasında veri zenginleştirilmesi için araştırmalar yapılmıştır. Çiftçiler ve kurum personelleri ile görüşmeler yapılmıştır. Hazırlanan veri makine öğrenmesi algoritmaları ile analiz edilmiştir. Veri üzerinde farklı deneyler yapılmıştır. Yapılan deneyler kıyaslanmıştır.

Bu tez çalışması beş bölümden oluşmaktadır. Birinci bölümde, girişe yer verilmiştir.

İkinci bölümde makine öğrenmesi algoritmaları, veri madenciliği yöntemleri, zaman serisi algoritmaları ve tarımsal faaliyetlerde petrol kullanımı ile ilgili literatür araştırması yapılmıştır.

Üçüncü bölümde veri kümesinin alımı, veriye eklenen özelliklerin tespiti, makine öğrenmesi algoritmalarının kod kümeleri ile açıklanması, algoritma verimliliğini arttırmak için kullanılan yöntemler açıklanmıştır.

Dördüncü bölümde deneysel çalışmalar anlatılmıştır. Bu bölümde analiz edilecek verinin nasıl kullanıldığı ayrıntılı anlatılmıştır. Veri 2019 eğitim 2020 test ve 2019-2020 eğitim 2021 test olmak üzere iki farklı deney için kullanılmıştır. Deney içerisinde algoritma bazlı önemli özellikler tespit edilmiştir. Mevcut gün tahmini ve sonraki gün tahmini yapılmıştır. Bu ayrımların her ikisinde de yüzdelik ayırım ve sıralı veri tahminleme yöntemleri kullanılmıştır.

Beşinci bölüme sonuç yazılmıştır. Elde edilen skorlar gösterilmiştir.

2. LİTERATÜR

Tarımsal faaliyetlerde kullanılan akaryakıt miktarları hakkında çeşitli bilimsel araştırmalar yapılmaktadır. Kullanılan akaryakıtın miktarını etkileyen faktörler arasında, tarımsal mekanizasyon önemli bir faktör olarak görülebilir. Bu bağlamda, akaryakıt tarım alanında yapılan araştırmalarda kullanılan araçların mekanizasyon oranlarını gösteren bir çalışma gerçekleştirilmiştir (Koçtürk, D., & Avcıoğlu, 2007). Akaryakıtın kullanıldığı mekanik araçlarda, akaryakıtı en aza indirgeyerek çiftçi kalkınmak istemektedir. Ayrıca akaryakıt ve ekonomi büyüme arasında da araştırmalar mevcuttur (Sefer; Usapbeyli, 2015). Çiftçinin aldığı akaryakıt fiyatının tespiti için zaman serisi modelleri kullanılmıştır. İmalat sanayi kapasite kullanım oranı üzerinde hem kısa hem de uzun dönemde yaklaşık % 15 oranında, imalat sanayi istihdam oranı üzerinde ise hem kısa hem de uzun dönemde % 3 oranında bir dalgalanmaya neden olduğu sonucuna ulaşılmıştır (Azazi, H., & Topkaya, 2017). Kullanılacak makine öğrenmesi algoritmaları ile intiharların tahmini, tektonik ortamların incelenmesi, uçucu yağ kimyasal bileşimi ve biyofilm araştırmaları, akaryakıt rezervuarlarının gözenekliliğini ve geçirgenliğini tahmin etmeye yönelik, yeraltı akış davranışında karbonat oluşumunun geçirgenliğini tahmin etmek için, elektrikli araç ilgisinin altında yatan nedenleri araştırmak ve elektrikli araç alıcılarının bir sonraki potansiyel dalgasını belirlemek için, boru hattında bir basınç düşümü hesaplanmasında her segmentteki basınç gradyanının iyi bölümlere ayrılması ve hesaplanması konusunda araştırmalar yapılmıştır (Walsh et al., 2017)(Petrelli & Perugini, 2016) (Hajizadeh, 2019)(Walsh et al., 2017)(Patsilnakos et al., 2019) (Ahmadi & Chen, 2019) (Tran et al., 2020) (Zarazua de Rubens, 2019) (Kanin et al., 2019). Çeşitli çalışmalarda veri madenciliği teknikleri ile, mevduat bankalarının etkinlik ölçümü, tıp ve sağlık hizmetleri kullanımı ve uygulamaları, elektronik ticaret, lise eğitimi, ders geçme notlarının tahmini, üniversite sistemi üzerindeki hareketlilik ile akademik başarı düzeyleri arasındaki ilişki konularında çalışmalar yapılmıştır. (Seyrek & Ata, 2010) (Koyuncugil, A., & Özgülbaş, 2009) (Vahaplar, A., & İnceoğlu, 2001) (Ayık, Y. Z., Özdemir, A., & Yavuz, 2007) (Aydemir, 2019) (Özbay & Ersoy, 2017).

Yararlanılacak zaman serisi analizi yöntemleri, Türkiye’de enflasyon-büyüme ilişkisi, ihracat, ithalat ve ekonomik büyüme, turizm talepleri araştırmaları yapılmıştır (Özer,

M., & Erdoğan, 2006)(Çuhadar, M., Güngör, İ., & Göksu, 2009). Yapılan regresyon analizi ise enflasyonun büyümeyi negatif etkilediini göstermiştir. Örnek döneminde enflasyondaki her 1 puanlık artışın büyüme oranını 0.37 puan düşürdüğü bulunmuştur (Karaca, 2003). Zaman serisi analizlerinden ARIMA ile Isparta'ya ait sıcaklık, buharlaşma ve yağış serilerinin uzun yıllık trendleri dikkate alınarak, 2015 yılına kadar eğilimleri tespit edilmiştir(Yaman et al., 2001). ARIMA yöntemi ile G20 ülkelerinin bankacılık sektörünün gelecekteki büyüklüğü tahmin edilmiştir (Erkan, E. T. İ., Dinçer, H., & Yüksel, 2019). ARMA yöntemi ile elektrik enerjisi yük tahmini yapılmıştır (Demirel, Ö., Kakilli, A., & Tektaş, 2010).

Kullanılacak metotların araştırmaları ile elde olan verinin uyumu grafiksel olarak kontrol edilmiştir. Kullanılacak metotların güncellenmesi ile veriyle arasındaki etkileşimin kuvvetleneceği de tahmin edilmektedir.

3. METODOLOJİ

Akaryakıt kullanımı ile alakalı elde edilen veriler üzerinde, K-en yakın komşu (K-Nearest neighbors), karar ağaçları regresyon (decision tree regression), karar ağaçları sınıflandırması (decision tree classifier), lojistik regresyon (logistic regression), destek vektör makinesi (support vector machine), rassal orman sınıflandırması (random forest classifier), sinir ağı (neural network) ve ekstrem artırılmış gradyan (Extreme Gradient Boosting, XGB) algoritmaları ile ikili sınıflandırma (kullanım var, kullanım yok) olarak çeşitli analizler gerçekleştirilmiştir. Ham veri üzerinde yapılan analizlere ek olarak önemli olan özelliklerin seçildiği yeni özellik vektörleri ile iyileştirilmiş analizler yapılmıştır. Önemli özelliklerin seçildiği algoritmalar yıldız (*) işareti ile belirtilmiştir. Örneğin, KNN algoritması önemli özelliklerin seçimi yapıldıktan sonra KNN* şeklinde gösterilmektedir. Her algoritma için kod çalıştırılmış ve önemli özellikler (feature importance) kaydedilmiştir. Algoritmalar çalıştırılırken, verinin ham hali ve önemli özelliklerin seçilmiş hali algoritmalara gönderilmiştir.

3.1. Analiz Yöntemleri

Algoritmaların veri kümesi üzerine uygulanması işlemi iki farklı analiz yöntemi kullanarak yapılmaktadır. Bu yöntemler yüzdelik ayırım yöntemi ve sıralı veri analizi yöntemi olarak isimlendirilmiştir.

3.1.1. Veri Ön İşleme ve Genişletme

Elde bulunan veriye çeşitli ön işleme yöntemleri uygulanmıştır. Veri kümesi üzerinde çiftçinin akaryakıt alım miktarı, dağıtım şirketi akaryakıt fiyatı ve hava durumu bilgileri tek bir veri kümesi şeklinde birleştirilmiştir. Akaryakıt verisi olarak günlük alınan akaryakıt miktarı baz alarak tahminleme yapılması hedeflenmiştir. Bu hedef doğrultusunda, gün bazlı, sıcaklığı, yağmur olup olmadığı, akaryakıt fiyatı, zam ve indirim olup olmadığı ve miktarı toplanmıştır. Yağmur verisi, toprağın ıslaklığı göz önünde bulundurularak önceki 7 güne kadar olup olmadığı kontrol edilmiştir. Akaryakıt fiyat hareketliliğindeki kontrol önceki güne kadar kontrol edilmiştir.

Veri toplama işleminin yanı sıra sonucu güçlendirmek için eklenebilecek diğer özellikler araştırılmıştır. Örneğin, veride bulunan tarih bilgisi üzerinde özellik genişletme yöntemi uygulanmıştır.

```
dates = df.iloc[:, 0:1]
datesValues = dates.values.tolist()
countDayWeek = 0
daysOfWeek = []
while countDayWeek < len(dates):
    daysOfWeek.append((pd.Timestamp(datesValues[countDayWeek][0]).dayofweek))
    countDayWeek = countDayWeek + 1

countDayYear = 0
daysOfYear = []
while countDayYear < len(dates):
    daysOfYear.append((pd.Timestamp(datesValues[countDayYear][0]).dayofyear))
    countDayYear = countDayYear + 1

countDayMonth = 0
daysOfMonth = []
while countDayMonth < len(dates):
    daysOfMonth.append((pd.Timestamp(datesValues[countDayMonth][0]).day))
    countDayMonth = countDayMonth + 1

firstdayOfMonth = []
for firstday in daysOfMonth:
    if firstday == 1:
        firstdayOfMonth.append(1)
    else:
        firstdayOfMonth.append(0)

weekendOfWeek = []
for weekend in daysOfWeek:
    if weekend == 5 or weekend == 6:
        weekendOfWeek.append(1)
    else:
        weekendOfWeek.append(0)
```

Şekil 1. Tarih genişletme işlemi günlerin karşılığı

Veri içerisindeki ayıklanan günlerin, haftanın kaçıncı günü, ayın kaçıncı günü, yılın kaçıncı günü, hafta sonu olup olmadığı işlemi uygulanmıştır.


```

holidays = ['2019-01-01', '2019-04-23', '2019-05-01', '2019-05-19',
            '2019-06-03', '2019-06-04', '2019-06-05', '2019-06-06',
            '2019-07-15', '2019-08-10', '2019-08-11', '2019-08-12',
            '2019-08-13', '2019-08-14', '2019-08-30', '2019-10-28',
            '2019-10-29',

            '2020-01-01', '2020-04-23', '2020-05-01', '2020-05-19',
            '2020-05-23', '2020-05-24', '2020-05-25', '2020-05-26',
            '2020-07-15', '2020-07-30', '2020-07-31', '2020-08-01',
            '2020-08-02', '2020-08-03', '2020-08-30', '2020-10-28',
            '2020-10-29',
            '2021-01-01'
            ]

holidaysDatetime = []
for i in range(len(holidays)):
    holidaysDatetime.append(datetime.datetime.strptime(holidays[i], '%Y-%m-%d'))

dfHolidays = pd.DataFrame(holidaysDatetime, columns=['Value'])
dfHolidays.columns = ['tarih']

dfDaysOfMonth = pd.DataFrame(daysOfMonth, columns=['gunler_ay'])
dfDaysOfWeek = pd.DataFrame(daysOfWeek, columns=['gunler_hafta'])
dfDaysOfYear = pd.DataFrame(daysOfYear, columns=['gunler_yil'])
dfFirstOfMonth = pd.DataFrame(firstdayOfMonth, columns=['ilkgun_ay'])
dfWeekendOfWeek = pd.DataFrame(weekendOfWeek, columns=['haftasonu'])
allData = pd.concat([df, dfDaysOfMonth, dfDaysOfWeek, dfDaysOfYear, dfFirstOfMonth,
                    dfWeekendOfWeek], axis=1)
allData["holiday"] = allData["tarih"].isin(dfHolidays["tarih"])
hol = []
for holDay in allData["holiday"]:
    if holDay == True:
        hol.append(1)
    else:
        hol.append(0)
allData["holiday"] = hol

```

Şekil 2. Tarih birleştirme ve tatil günleri hesaplamaları Python kod bloğu

Veri içerisine günlerin tatil olup olmadığı kontrolü eklenmiştir. İşlem öncesi gerekli tip değişiklikleri yapılmış ve listeler oluşturulmuştur. *allData* değişkeni ile veri genişletmesindeki işlemler birleştirilmiş ve dataframe elde edilmiştir.

```

pivot_table = dfDb.pivot_table(index=["tarih"], columns=["uye_no"],
                               values="miktar").fillna(0)
dfData = allData.set_index("tarih").join(pivot_table, lsuffix='_caller',
                                         rsuffix='_other', on='tarih').fillna(0)

```

Şekil 3. Üyelerin aldıkları akaryakıtın veriye işlenmesi için Python kod bloğu

Veri hazırlanmasında, algoritmaların tahminleme yeteneğini arttırmak için çiftlerin aldıkları akaryakıt miktarları, veri kümesine özellik olarak eklenmiştir.

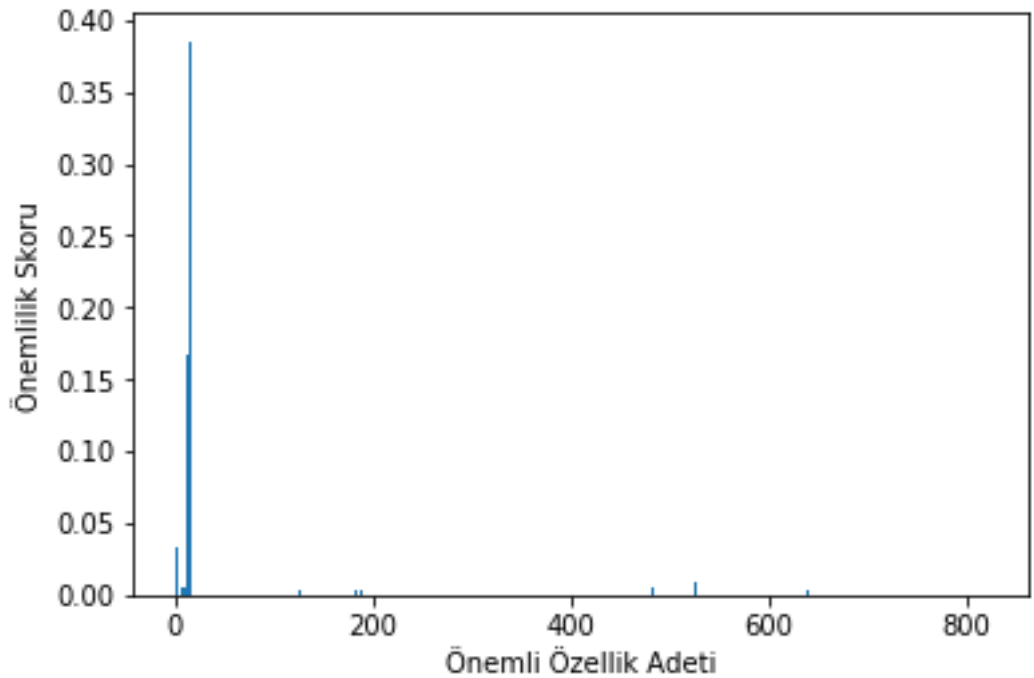
3.1.2. Önemli Veri Özelliklerinin Bulunması (Feature Importance)

Veri kümesi, farklı makine öğrenmesi algoritmaları için özellik seçimi yöntemlerine tabi tutulmuştur. Kullanılan algoritmaların içerisindeki özelliklerden her bir algoritma özelinde önemli özellikleri tespit edilmiştir.

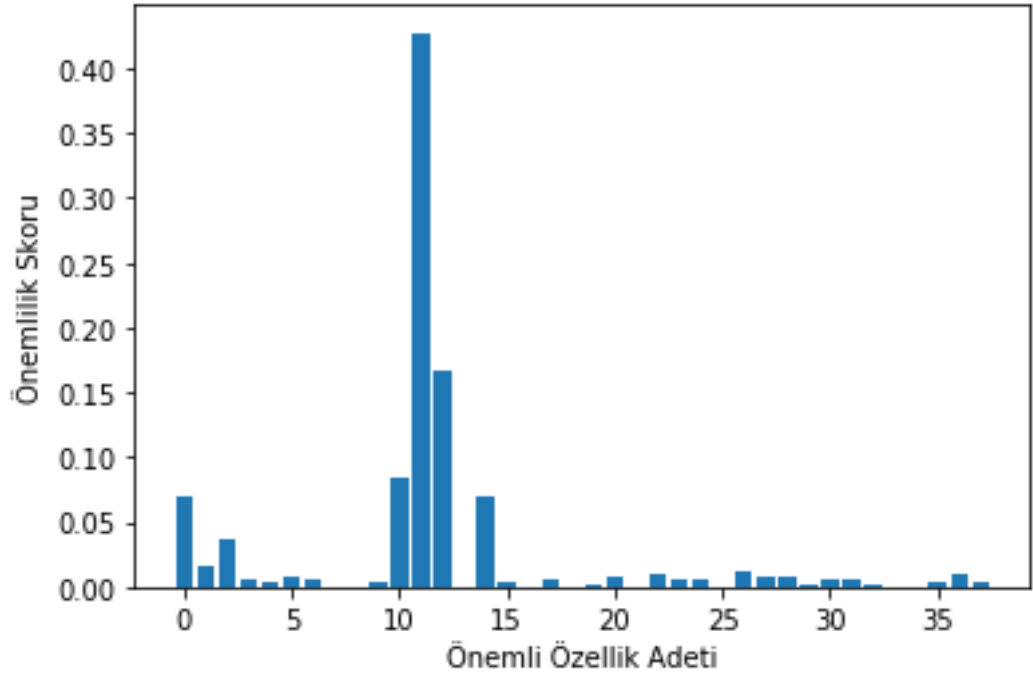
```
def dtrImportanceX(x,y):  
    from sklearn.tree import DecisionTreeRegressor  
    model = DecisionTreeRegressor(random_state=0)  
    importance = model.fit(x,y)  
    importanceFeature = importance.feature_importances_  
    importanceFeature = list(importanceFeature)  
    dtrRegFeatureIndex = []  
    for i in importanceFeature:  
        if i > 0.0:  
            dtrRegFeatureIndex.append(importanceFeature.index(i))  
    xForDtrReg = x[dtrRegFeatureIndex]  
    return xForDtrReg
```

Şekil 4. Önemli özellik tespiti kodu örneği

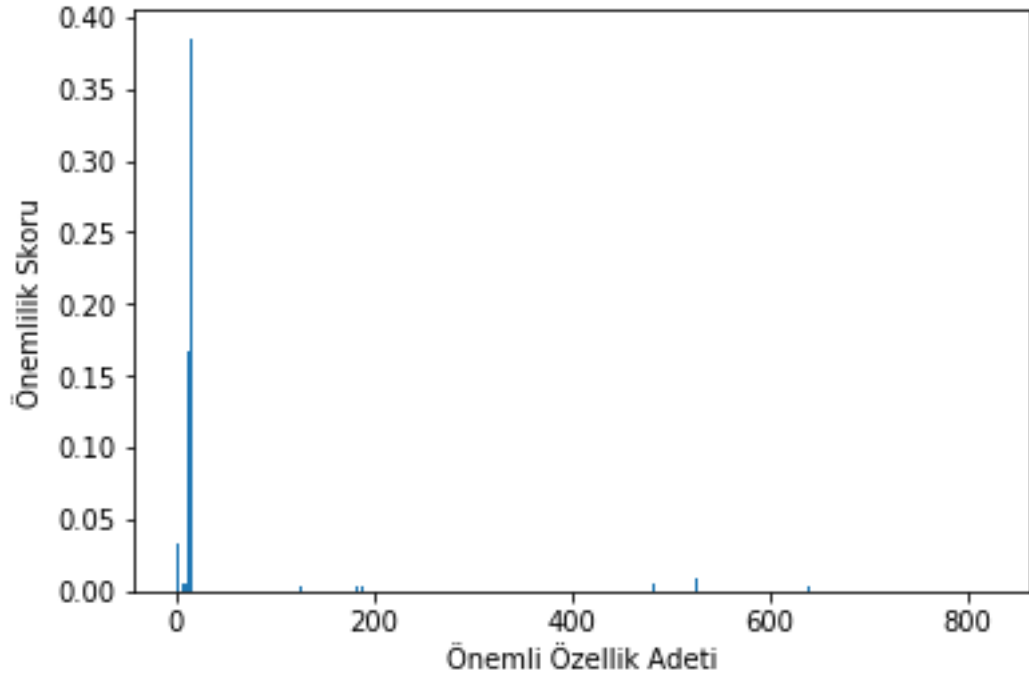
DTR* için yapılmış olan önemli veri tespiti örneği verilmiştir. Önemli özelliklerin tespiti için sklearn kütüphanesinden faydalanılmıştır. Önemli özellik tespiti ile oluşan değişiklikler şöyledir;



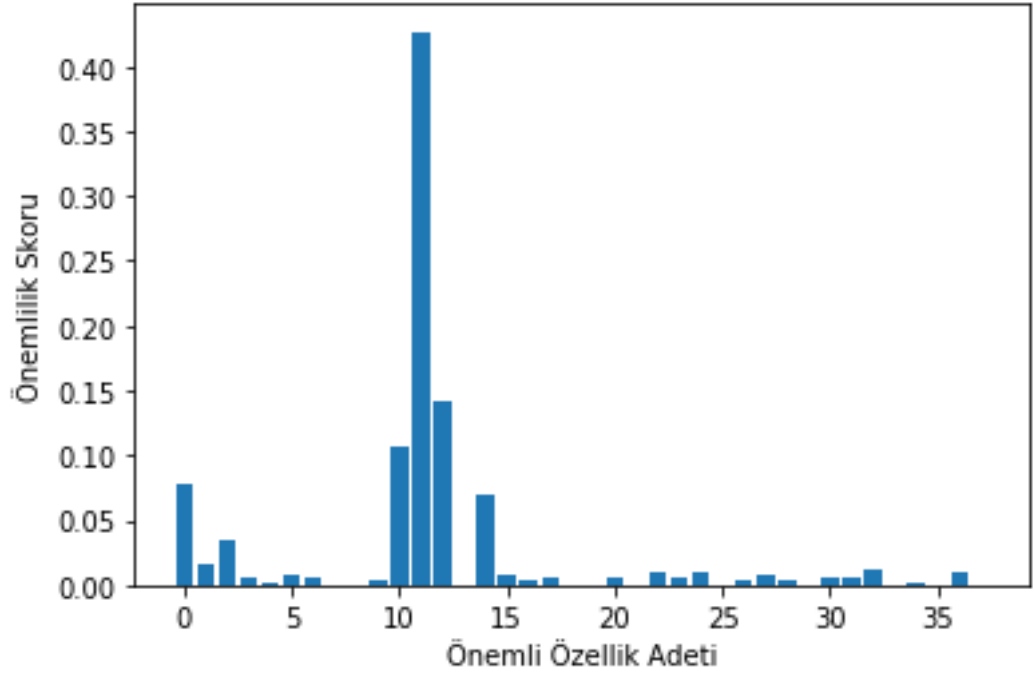
Şekil 5. DTR normal veri grafik üzeri önem göstergesi



Şekil 6. DTR* önemli veri ayıklaması sonucu özelliklerin önem seviyeleri



Şekil 7. DTC normal veri grafik üzeri önem göstergesi



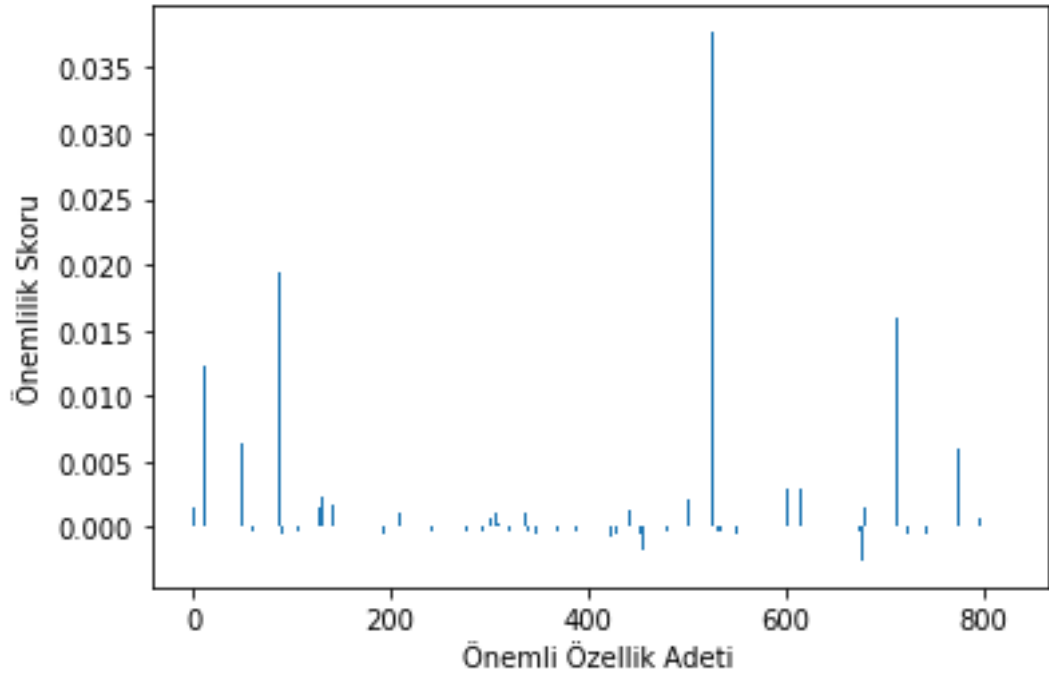
Şekil 8. DTC* önemli veri ayıklaması sonucu özelliklerin önem seviyeleri

DTC algoritmasının önemli özellikleri tespit edilmiştir. Önemli özelliklerinin sıralı olduğu özellik adı bulunmuştur. Önemli özelliklerin bulunduğu adı sütunundaki sayılar üye numarasını temsil etmektedir.

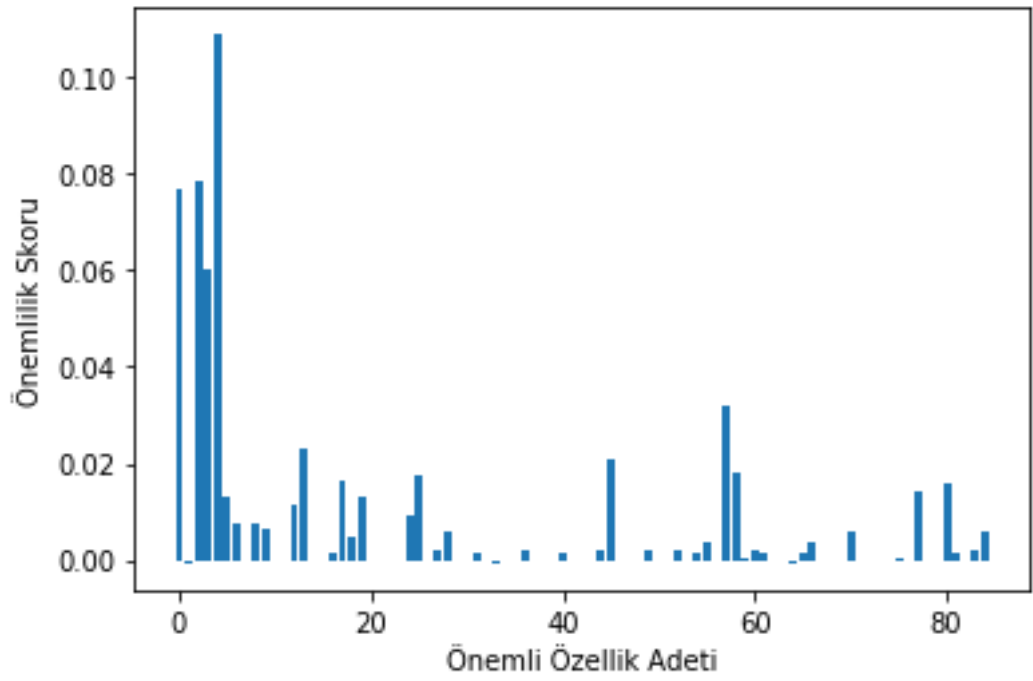
Tablo 1. DTC Algoritmasının Önemli Özellik Sıralaması ve Adları (Üye Numaraları Maskelenmiştir)

Özellik Sırası	Özellik Kısa Adı	Özellik Açıklaması	Özellik Önem Skoru
11	indirim var/yok haberi	İlgili gün indirim olup olmadığı bilgisi	0.426
12	indirim 2 gün öncesi	2 gün önceki indirim bilgisi	0.142
10	akaryakıt fiyat hareketliliği	ilgili gün akaryakıt fiyat hareketlilik bilgisi	0.105
0	sıcaklık	sıcaklık bilgisi	0.077
14	indirim miktarı	akaryakıt indirim miktarı	0.070
2	yağış var/yok bir gün öncesi	ilgili günden önceki bir gün yağış bilgisi	0.035
1	yağış var/yok	ilgili gün yağış bilgisi	0.016
32	X numaralı üye	İlgili üye numarası	0.012

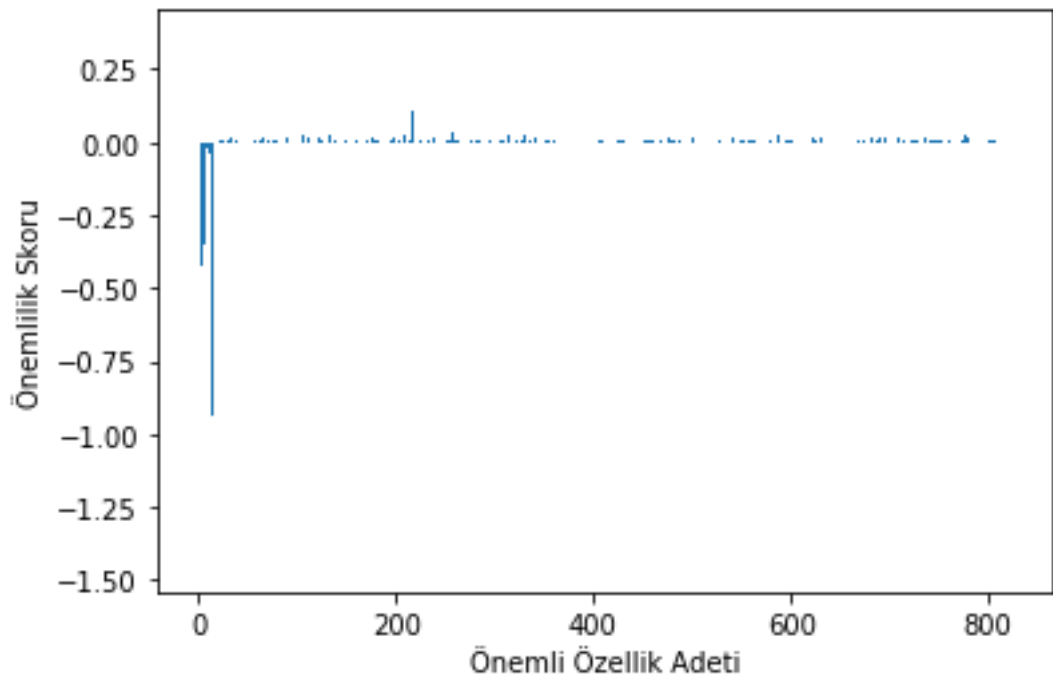
24	X numaralı üye	üye numarası	0.010
36	X numaralı üye	üye numarası	0.009
22	X numaralı üye	üye numarası	0.009
27	X numaralı üye	üye numarası	0.008
5	yağış var/yok dört gün öncesi	ilgili günden önceki dört gün yağış bilgisi	0.008
15	yağış var/yok yedi gün öncesi	ilgili günden önceki yedi gün yağış bilgisi	0.007
3	yağış var/yok iki gün öncesi	ilgili günden önceki iki gün yağış bilgisi	0.006
6	yağış var/yok altı gün öncesi	ilgili günden önceki altı gün yağış bilgisi	0.006
20	X numaralı üye	üye numarası	0.005
31	X numaralı üye	üye numarası	0.005
30	X numaralı üye	üye numarası	0.005
17	X numaralı üye	üye numarası	0.005
23	X numaralı üye	üye numarası	0.004
28	X numaralı üye	üye numarası	0.004
16	X numaralı üye	üye numarası	0.003
26	yağış var/yok üç gün öncesi	ilgili günden önceki üç gün yağış bilgisi	0.003
34	X numaralı üye	üye numarası	0.002

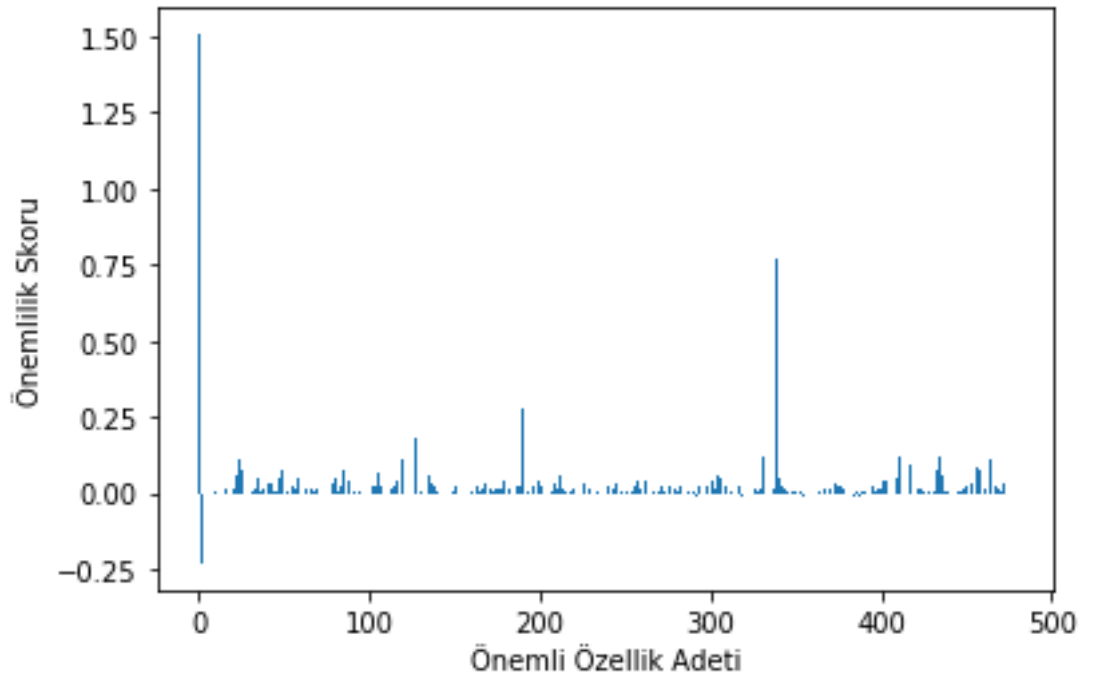


Şekil 9. KNN normal veri özelliklerin önem seviyeleri

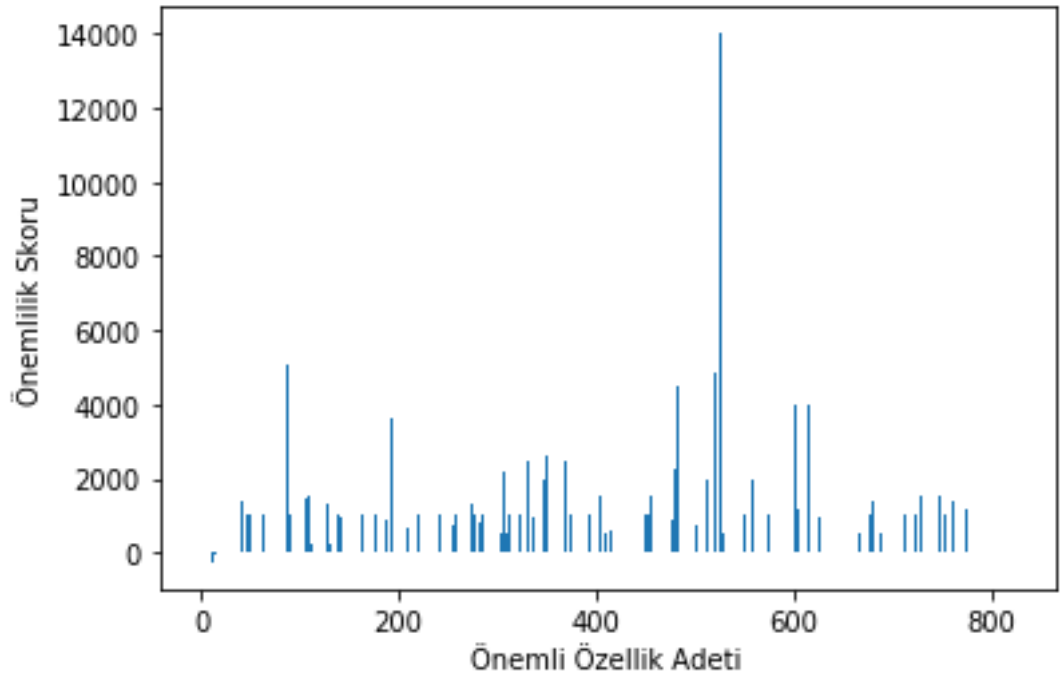


Şekil 10. KNN* önemli veri ayıklaması sonucu özelliklerin önem seviyeleri

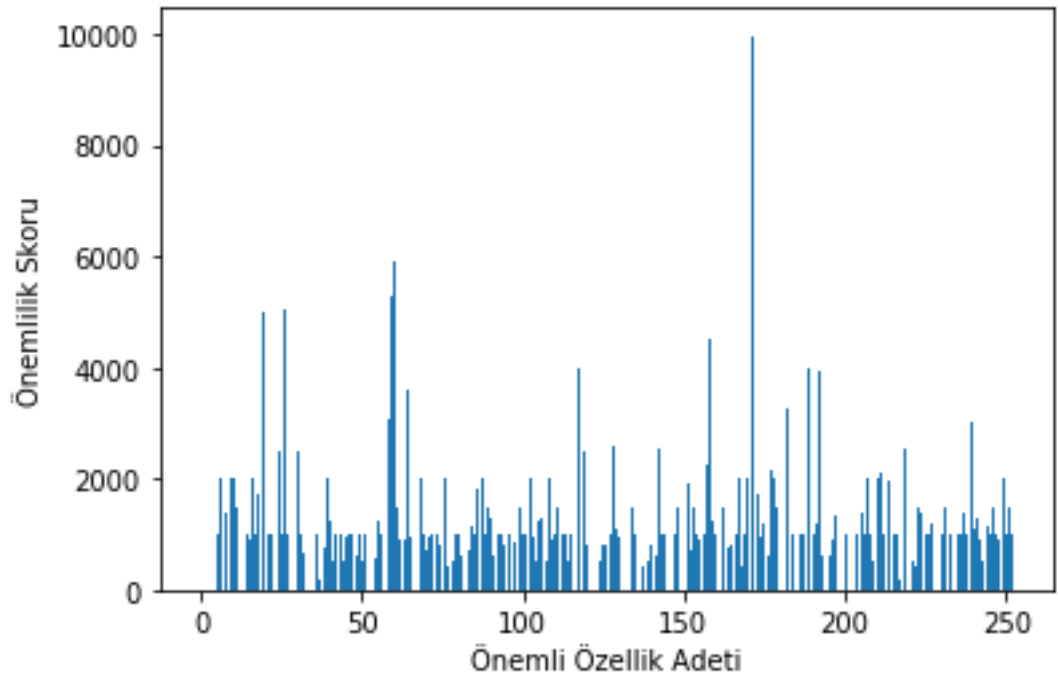




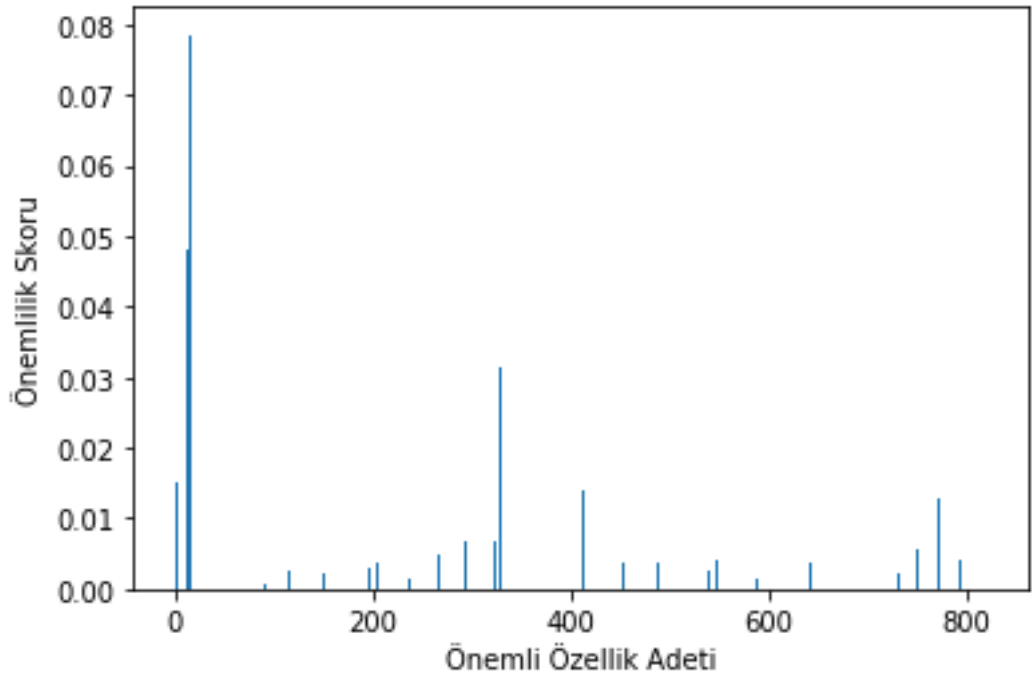
Şekil 12. LR* önemli veri ayıklaması sonucu özelliklerin önem seviyeleri



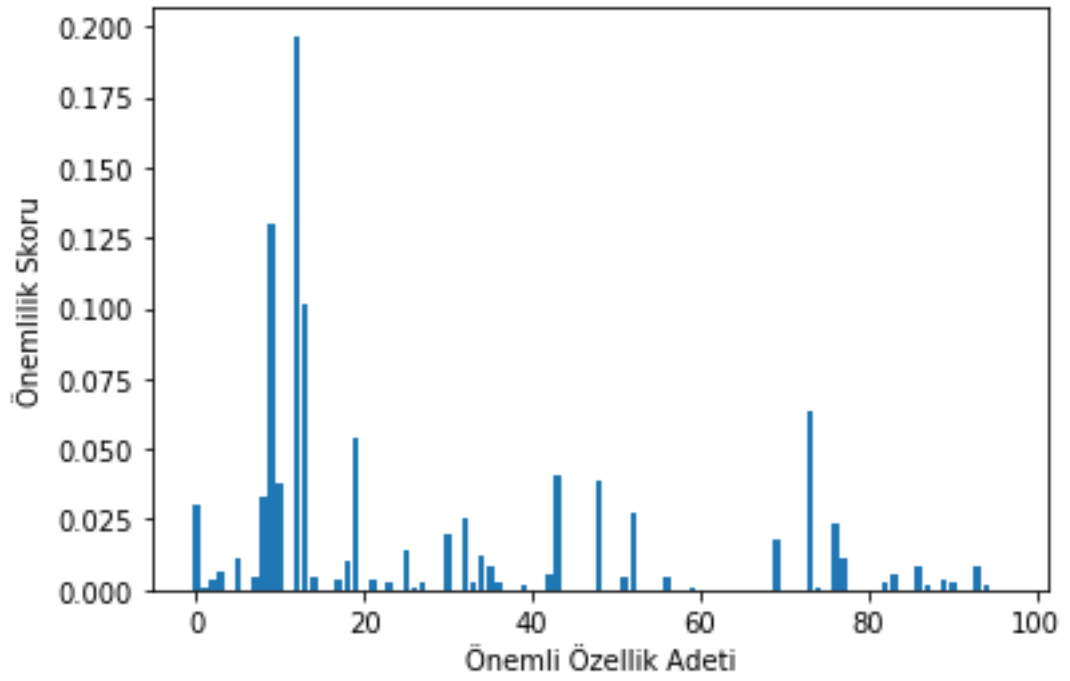
Şekil 13. NN normal veri özelliklerin önem seviyeleri



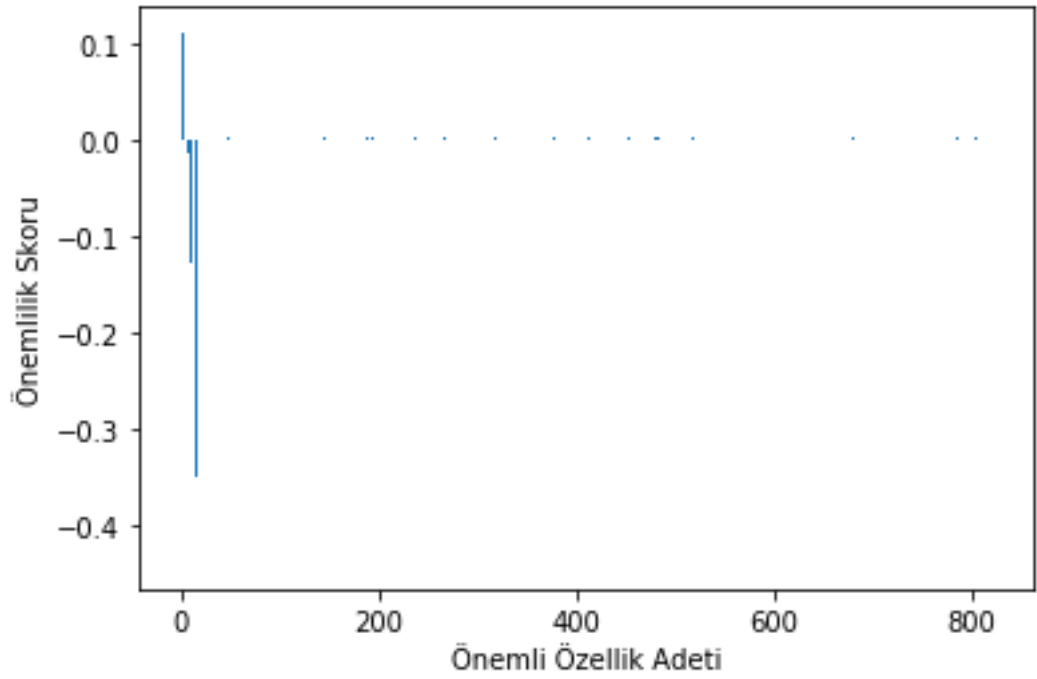
Şekil 14. NN* önemli veri ayıklaması sonucu özelliklerin önem seviyeleri



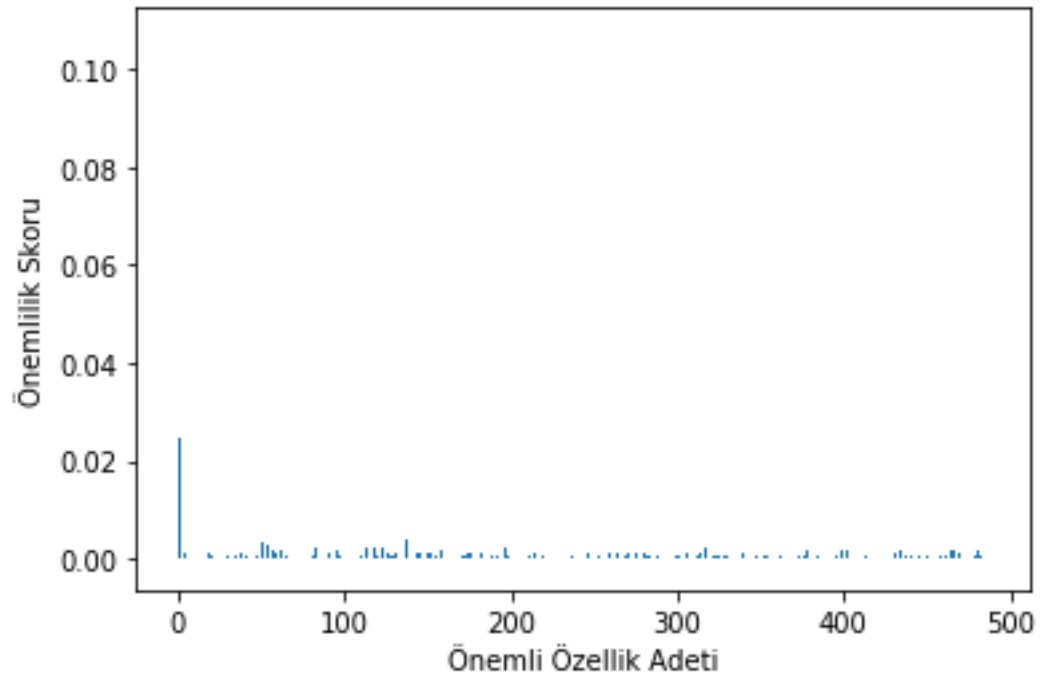
Şekil 15. RFC normal veri özelliklerin önem seviyeleri



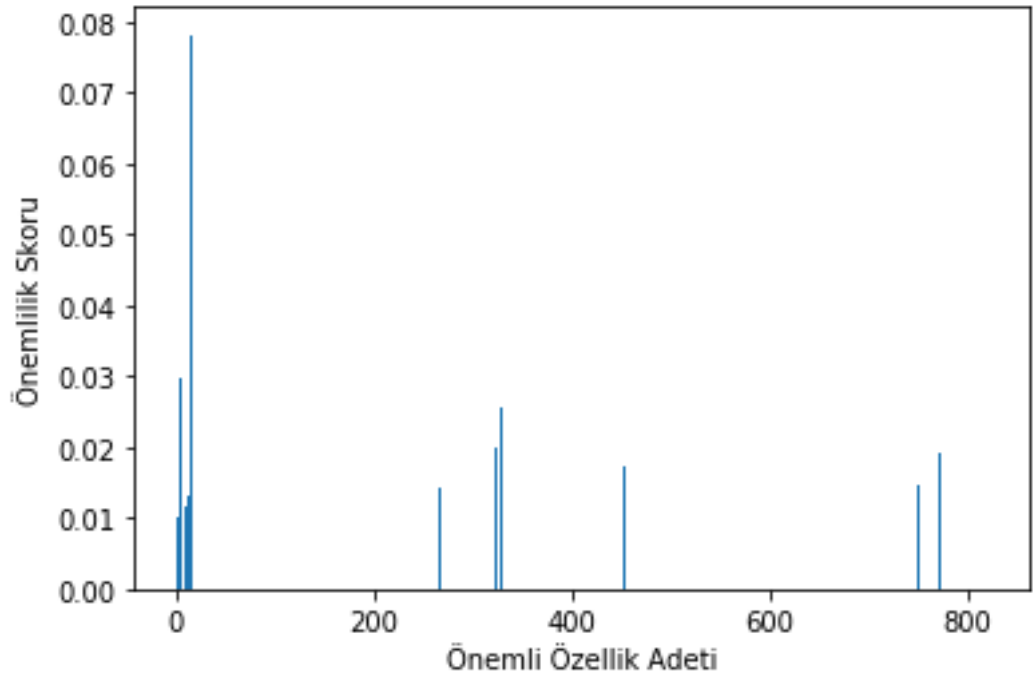
Şekil 16. RFC* önemli veri ayıklaması sonucu özelliklerin önem seviyeleri



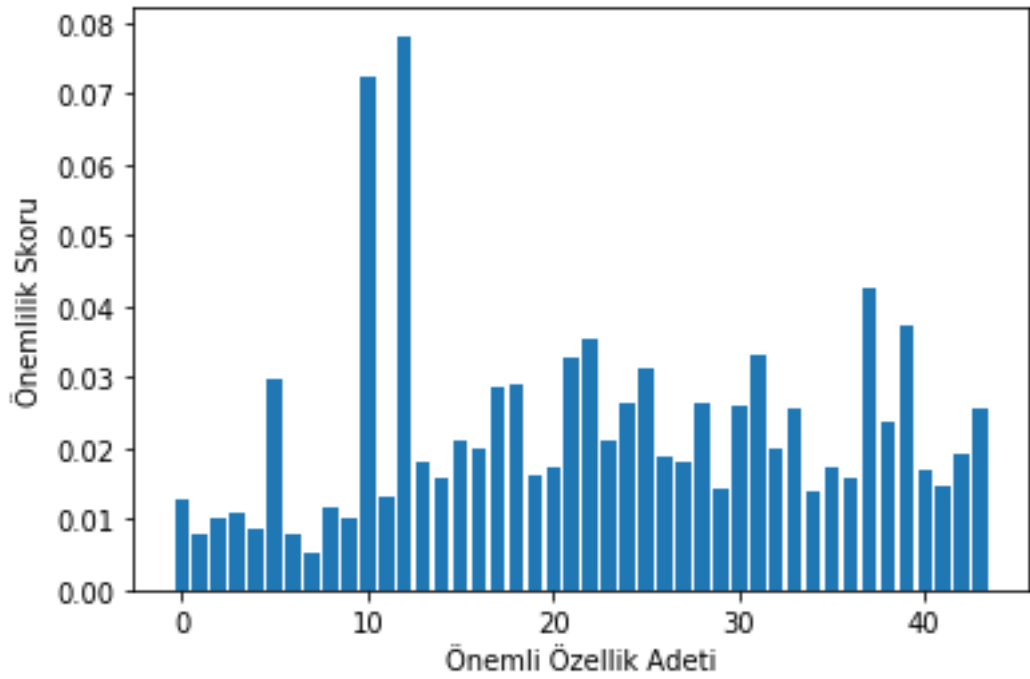
Şekil 17. SVM normal veri özelliklerin önem seviyeleri



Şekil 18. SVM* önemli veri ayıklaması sonucu özelliklerin önem seviyeleri



Şekil 19. XGB normal veri özelliklerin önem seviyeleri



Şekil 20. XGB* önemli veri ayıklaması sonucu özelliklerin önem seviyeleri

3.1.3. Yüzdelik Ayrım Metodu

Sklearn kütüphanesi içerisinde bulunan makine öğrenmesi algoritmalarının veri kümesi üzerinde uygulaması gerçekleştirilmiştir. 2019 eğitim 2021 test verisinde %50 eğitim %50 test kullanılmıştır. 2019 ve 2020 eğitim 2021 test verisinde %89 eğitim %11 test kullanılmıştır.

```
def DtrRegPred(xTrain,xTest,yTrain):
    from sklearn.tree import DecisionTreeRegressor
    model = DecisionTreeRegressor(random_state=0)
    model.fit(xTrain,yTrain)
    yPred = model.predict(xTest)
    return yPred
```

Şekil 21. Örnek algoritma yüzdelik ayırım metodu

3.1.4. Sıralı Veri Kullanımı

Sıralı veri metodu için fonksiyon oluşturulmuştur. Fonksiyon içerisine veri x ve y olarak tanımlanmıştır. X verisi tahmini etkileyen özellikler, y verisi ise test için ayrılan veri setidir. Cut parametresi ile verinin kaçınıcı sıradan ayrılacağını belirlemiştir. Örneğin 2019 eğitim 2020 test verisinde cut değeri 365'tir. Sıralı veri metodu ile verinin ilk 365

günü 366. Günü tahmin edip, tahminlediği günü *yPred* listesine atması ve tahminlediği 366. günü 365 güne dahil edecek 367. günü tahminlemesi sağlanmıştır. Fonksiyon içerisinde döngü sağlandıktan sonra *yPred* listesi fonksiyonun sonucu olmuştur.

```
def DtrClassPred(x,y,cut):
    xTrain = []
    xTest = []
    yTrain = []
    yTest = []
    yPred = []
    for i in range (1,lastNum+1):
        xTrain.append(x.iloc[0:cut-1+i, :].values)
        xTest.append(x.iloc[cut-1+i:cut+i, :].values)
        yTrain.append(y.iloc[0:cut-1+i, :].values)
        yTest.append(y.iloc[cut-1+i:cut+i, :].values)

        from sklearn.tree import DecisionTreeClassifier
        model = DecisionTreeClassifier(random_state=0)
        model.fit(xTrain[i-1], yTrain[i-1])
        yPred.append(model.predict(xTest[i-1]))

    return yPred
```

Şekil 22. Örnek algoritma sıralı veri metodu kullanımı

3.1.5. Izgara Araması Parametre Optimizasyon Yöntemi (Grid Search Hyperparameter Optimization)

Izgara araması yöntemi algoritmalara uygulanmıştır. Algoritmaların parametreleri arasında, veriye en uygun olanlar tespit edilmiştir. Analiz işleminde, algoritmanın sklearn kütüphanesindeki varsayılan hali ve ızgara araması yöntemi ile tespit edilen en uygun parametrelerin kullanıldığı hali kıyaslanmıştır.

Izgara araması için sklearn kütüphanesi kullanılmıştır. Izgara araması kodlaması, veri kümesindeki önemli özelliklerin seçilerek oluşturulan filtrelenmiş veri kümesi ve veri kümesinin önemli özelliklerinin filtrelenmemiş hali için kullanılmıştır. Önemli özelliklerin seçildiği veri kümeleri üzerine tespit edilen parametreler uygulanmıştır. Aynı işlem filtrelenmemiş veri kümesi içinde uygulanmıştır.

```

def XgbGridSearch (X,y):
    import xgboost as xgb
    parameters = {
        'min_child_weight': [1, 5, 10],
        'gamma': [0.5, 1, 1.5, 2, 5],
        'subsample': [0.6, 0.8, 1.0],
        'colsample_bytree': [0.6, 0.8, 1.0],
        'max_depth': [3, 4, 5],
        'objective': ['binary:logistic', 'binary:hinge']
    }

    algorithm=xgb.XGBClassifier()
    model = GridSearchCV(algorithm, parameters)
    model.fit(X, y)
    return model.best_params_

HyperParamsXgb = XgbGridSearch(x,y)
HyperParamsXfbImportance = XgbGridSearch(xForXgb,y)

```

Şekil 23. Örnek algoritma için ızgara araması uygulanması

Izgara araması parametre optimizasyonu yöntemi ile elde edilen parametre değerleri aşağıda listelenmiştir;

Tablo 2. DTC hiper parametreleri

DTC	
PARAMETRE	DEĞER
Criterion	entropy
Max_depth	4
Max_features	sqrt
Max_leaf_nodes	4
Min_samples_leaf	3
Min_samples_split	2
Splitter	Best

Tablo 3. DTC* hiper parametreleri

DTC*	
PARAMETRE	DEĞER
Criterion	Entropy
Max_depth	3
Max_features	Sqrt
Max_leaf_nodes	5
Min_samples_leaf	5
Min_samples_split	4
Splitter	Best

Tablo 4. DTR hiper parametreleri

DTR	
PARAMETRE	DEĞER
Criterion	Mse
Random_state	9
Splitter	Best

Tablo 5. DTR* hiper parametreleri

DTR*	
PARAMETRE	DEĞER
Criterion	Mse
Random_state	1
Splitter	Random

Tablo 6. KNN hiper parametreleri

KNN	
PARAMETRE	DEĞER
algortihm	Auto
Leaf_size	1
N_jobs	-1
N_neighbors	4

Tablo 7. KNN* hiper parametreleri

KNN*	
PARAMETRE	DEĞER
Algorithm	auto
Leaf_size	1
N_jobs	-1
N_neighbors	4

Tablo 8. NN hiper parametreleri

NN	
PARAMETRE	DEĞER
Activation	relu
alpha	0.0001
hidden_layer_sizes	(50,100,50)
Learning_rate	adaptive
solver	Sgd

Tablo 9. NN* hiper parametreleri

NN*	
PARAMETRE	DEĞER
Activation	Relu
alpha	0.05
hidden_layer_sizes	(50,100,50)
Learning_rate	Adaptive
solver	Sgd

Tablo 10. RFC hiper parametreleri

RFC	
PARAMETRE	DEĞER
max_feature	Sqrt
Min_samples_split	6
N_estimators	250

Tablo 11. RFC* hiper parametreleri

RFC*	
PARAMETRE	DEĞER
max_feature	Sqrt
Min_samples_split	6
N_estimators	50

Tablo 12. SVM hiper parametreleri

SVM	
PARAMETRE	DEĞER
C	0.1
Multi_class	Ovr
penalty	l2

Tablo 13. SVM* hiper parametreleri

SVM*	
PARAMETRE	DEĞER
C	1
Multi_class	Crammer_singer
penalty	l1

Tablo 14. XGB hiper parametreleri

XGB	
PARAMETRE	DEĞER
Colsample_bytree	0.8
gamma	5
Max_depth	4
Min_child_weight	5
objective	Binary:logistic
subsample	1.0

Tablo 15. XGB* hiper parametreleri

XGB*	
PARAMETRE	DEĞER
Colsample_bytree	0.6
gamma	5
Max_depth	4
Min_child_weight	1
objective	Binary:logistic
subsample	0.6

Tablo 16. LR hiper parametreleri

LR	
PARAMETRE	DEĞER
C	1.8329807108324339
penalty	12
solver	Lbfgs

Tablo 17. LR* hiper parametreleri

LR*	
PARAMETRE	DEĞER
C	0.004281332398719391
penalty	12
solver	Lbfgs

4. DENEYSEL ÇALIŞMA

Sıralı veri metodu için eğitim ve test veriler ayrılmıştır. Ayrılan verileri eğitim için oluşturulan verinin üzerine tahminlenen veri sırasındaki test verisinde bulunan veriler eğitim verisinin sonuna eklenmiştir. Örneğin 2019 verilerinin eğitim 2020 verilerinin test için kullanılan deneyinde, 2019 yılındaki 365 günlük eğitim verisi 2020 1 ocağı tahmin etmektedir. Tahmin edildikten sonra 2020 yılının 1 Ocak tarihindeki veriler, eğitim verisine eklenerek 365 günden oluşan eğitim veri seri 366 güne çıkarılmıştır. 366 günlük veri, 2020 2 ocağı tahmin etmiştir. Bu işlem test verisindeki bütün veriler için döngü içerisinde tekrarlanmıştır.

Yüzdelik ayırım metodu, verinin eğitim ve test kısmını yüzdelik ilerek bölerek tahminlememizi sağlamıştır. Örneğin 2019 eğitim 2020 test deneyinde 2019 verileri tamamen eğitim için ve 2020 verileri tamamen test için kullanılmıştır. 2019-2020 verisi üzerinden %50 yüzdelik ayırım kullanılmıştır. 2019-2020 verisi kullanılarak 2021 ilk çeyrek tahmini için 90 test, 732 eğitim değer kullanılmıştır. %11 test, %89 eğitim yüzdelik ayırımı yapılmıştır.

Karşılaştırmalı algoritma skorlarında yapılan deneyler kendi aralarında kıyaslanmıştır. Tespit edilen doğruluk, kesinlik, hassasiyet ve f1 skorları deney raporlarında karşılaştırmalı algoritma sonuçları başlığında gösterilmiştir. Grafik içerisinde sıralı veri metodu ve yüzdelik ayırım metodunun varsayılan algoritma girdileri ve uygulamalı parametre optimizasyon yapılmış sonuçların kıyaslanmasını gösterilmiştir.

4.1. 2019-2020 Verisi Üzerinden Tam Yıl Analizleri: Mevcut Gün Tahmini

Çiftçilerin aldıkları akaryakıt miktarları veri genişletme ve ön işleme ile zenginleştirilmiştir. Elde edilen veri yıl bazlı tahminlenmiştir. Bu deneyde 2019 yılı verileri eğitim 2020 yılı verileri test için kullanılmıştır.

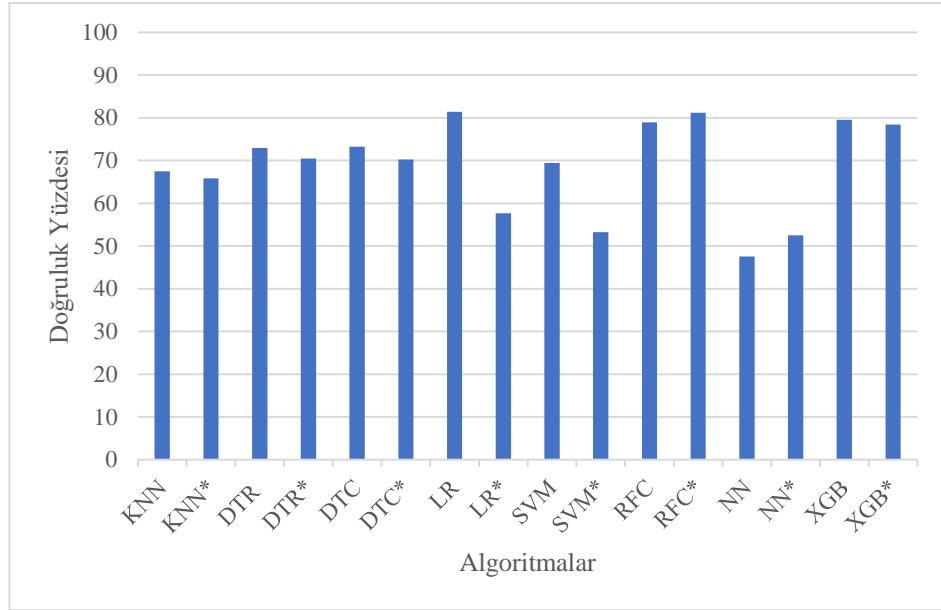
4.1.1. Sıralı Veri Tahminleme Analizi

Veriyi tahminlemek için 2019 verilerinin gün bazlı sıra ile tespitleri fonksiyon içerisinde oluşturulan dizine aktarılmıştır. Eğitim dizisinde bulunan veri miktarı arttıkça döngü ile test edilmek istenen bir sonraki güne etki etmektedir. Elde edilen sonuç

tahmin dizisi test dizisi ile kıyaslanmıştır. Doğruluk skorunun verimliliği için elde edilen hata matrisleri ektedir.

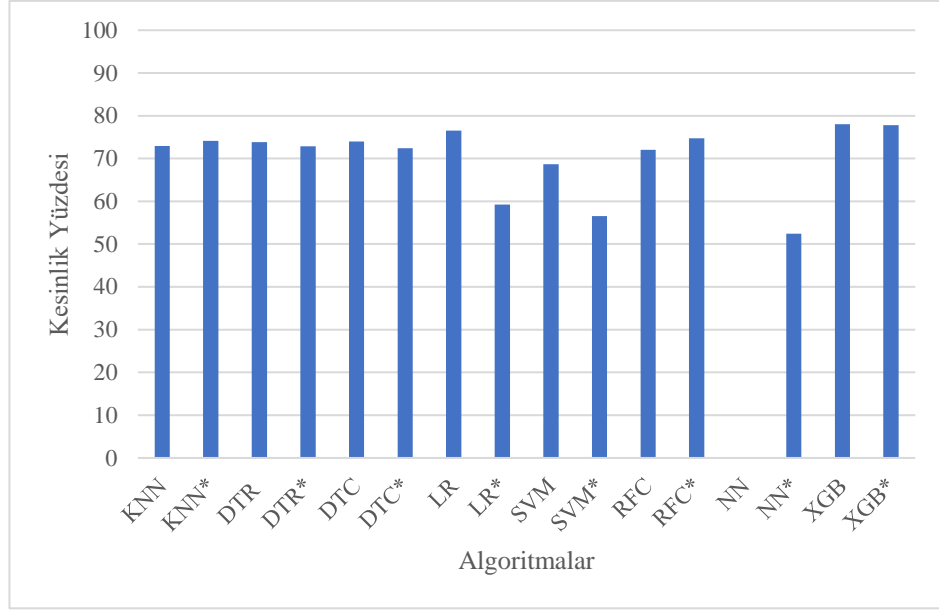
4.1.1.1. Varsayılan algoritma girdi parametreleri ile alınan sonuçlar

Verinin analiz yöntemlerinden birisi olan sıralı veri tahminleme analizinin en yüksek doğruluk skorunu %81,421 ile RFC* algoritması almıştır.



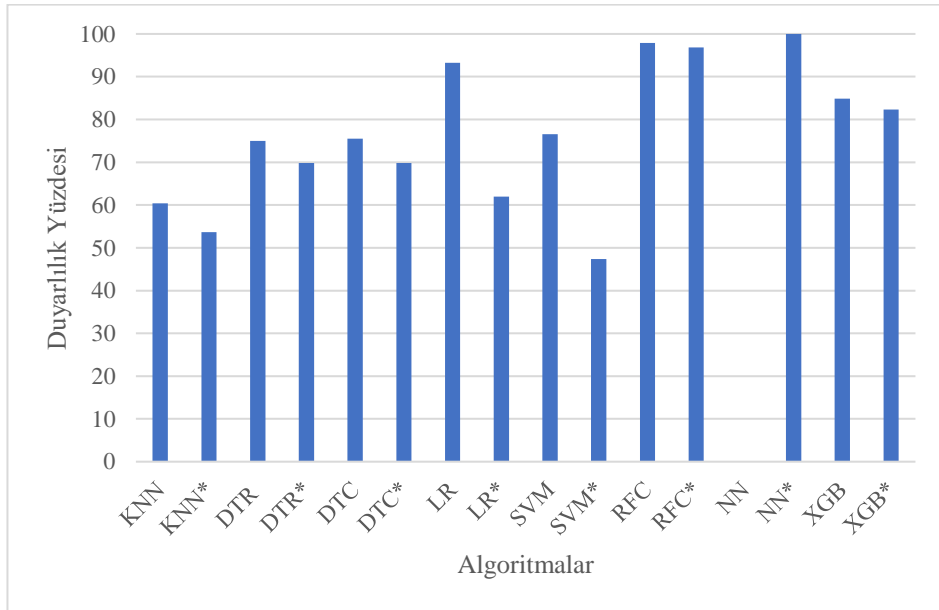
Şekil 24. Algoritmaların sıralı veri tahminleme metodu ile doğruluk skor kıyaslanması

Algoritmaların sıralı veri tahminleme metodu ile kıyaslanmasında kesinlik skorları için en yüksek skoru alanlar, %77,833 ile XGB* olmuştur.



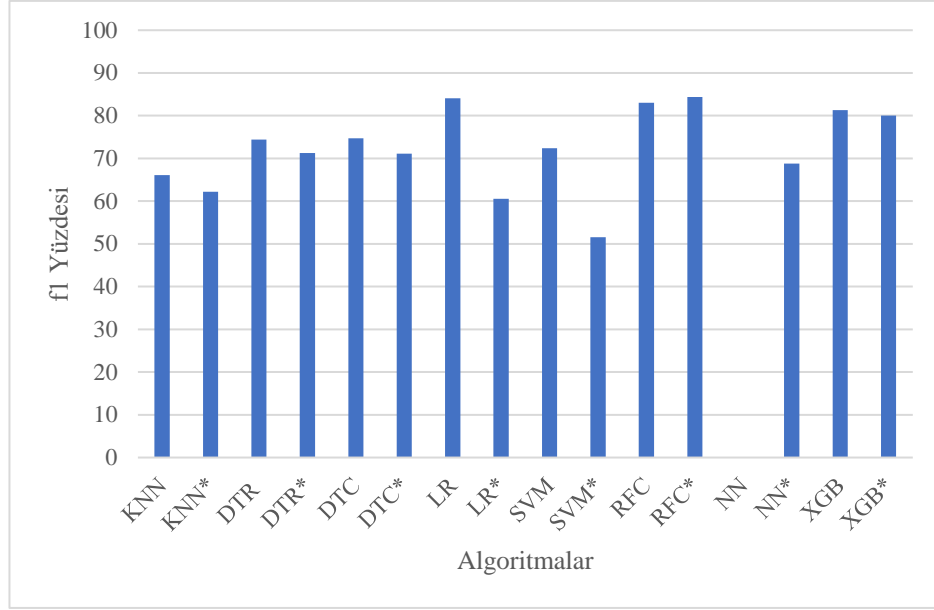
Şekil 25. Algoritmaların sıralı veri tahminleme metodu ile kesinlik skor kıyaslanması

Kullanılan sıralı veri tahminleme metodu sonucunda algoritmalar arasındaki en iyi duyarlılık skoru %100 ile NN* almıştır.



Şekil 26. Algoritmaların sıralı veri tahminleme metodu ile duyarlılık skor kıyaslanması

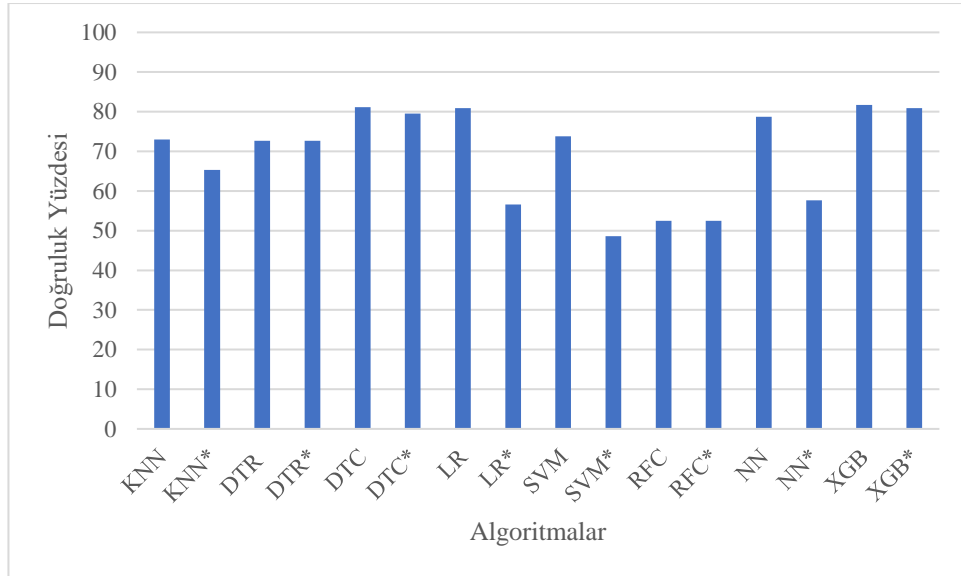
Algoritmalarından RFC*, %84,354 skor ile en iyi f1 skoru yapmıştır.



Şekil 27. Algoritmaların sıralı veri tahminleme metodu ile f1 skor kıyaslanması

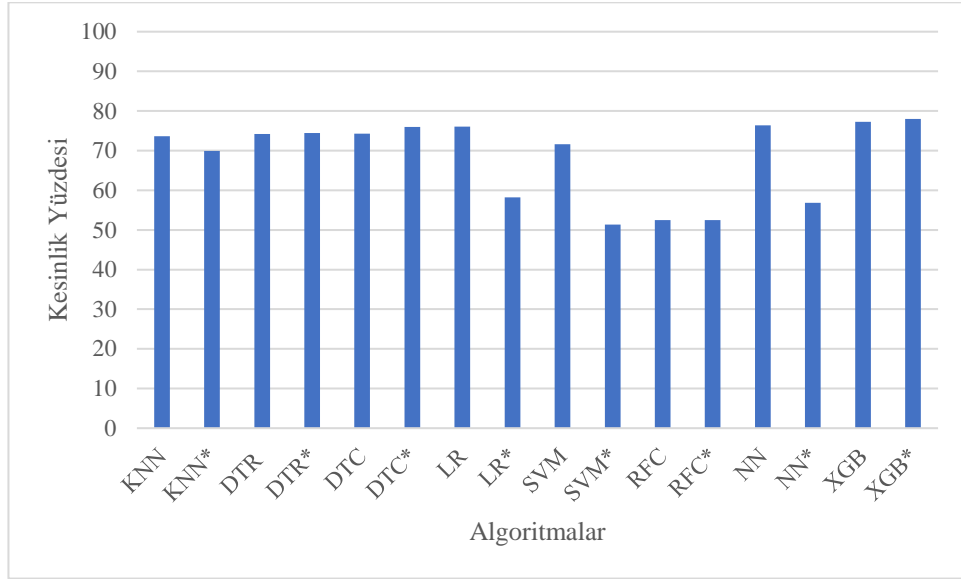
4.1.1.2. Hiper parametre optimizasyonu yapıldıktan sonra alınan sonuçlar

Tespit edilen parametreler sıralı veri tahminleme metodu ile uygulanmıştır. Bu analize göre en yüksek başarıyı %81,694 ile XGB algoritması sağlamıştır.



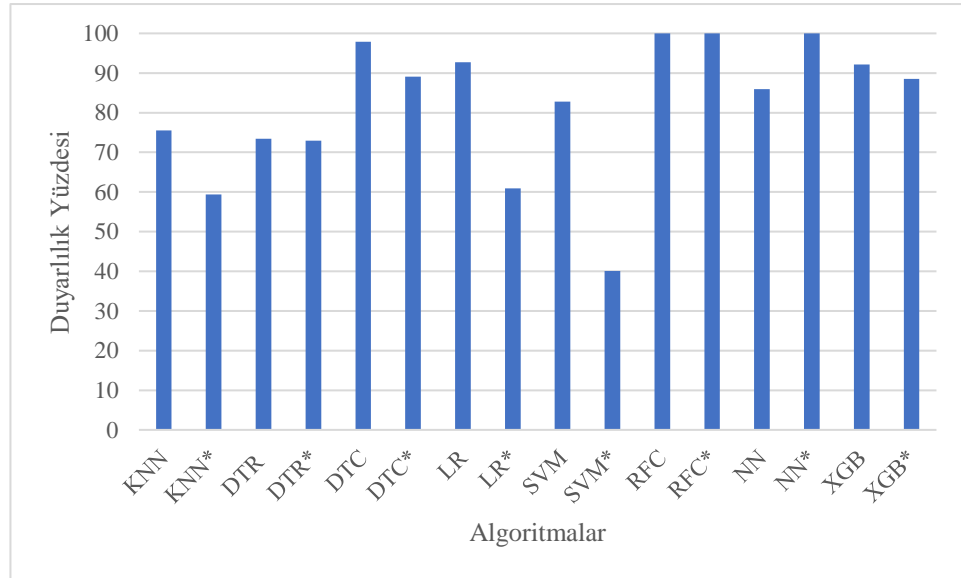
Şekil 28. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu doğruluk skor kıyaslanması

Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu kullanımında en yüksek kesinlik skoru %77,982 ile XGB* sağlamıştır.



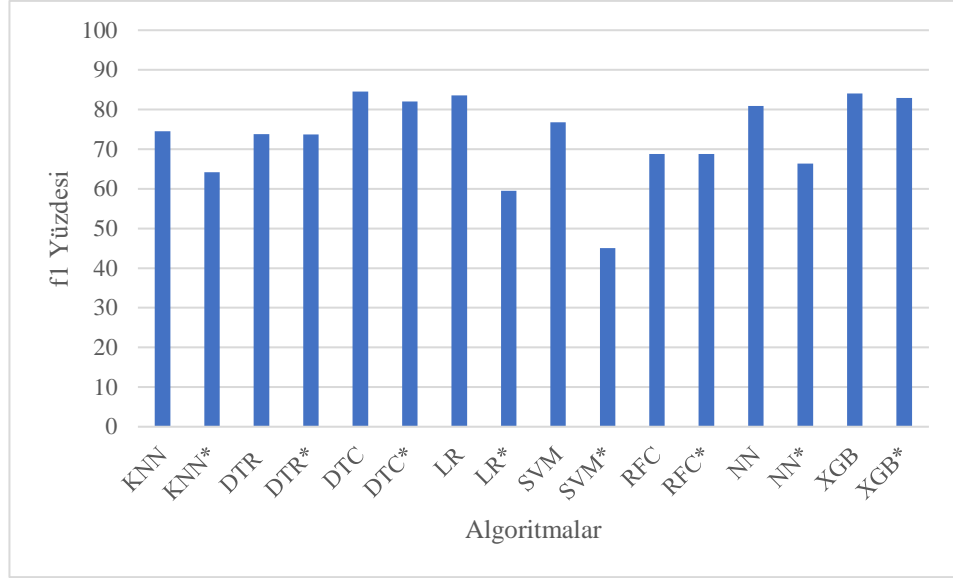
Şekil 29. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu kesinlik skor kıyaslanması

Algoritmaların duyarlılık skorlarının kıyaslanmasında en yüksek skorları %100 ile NN* sağlamıştır.



Şekil 30. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu duyarlılık skor kıyaslanması

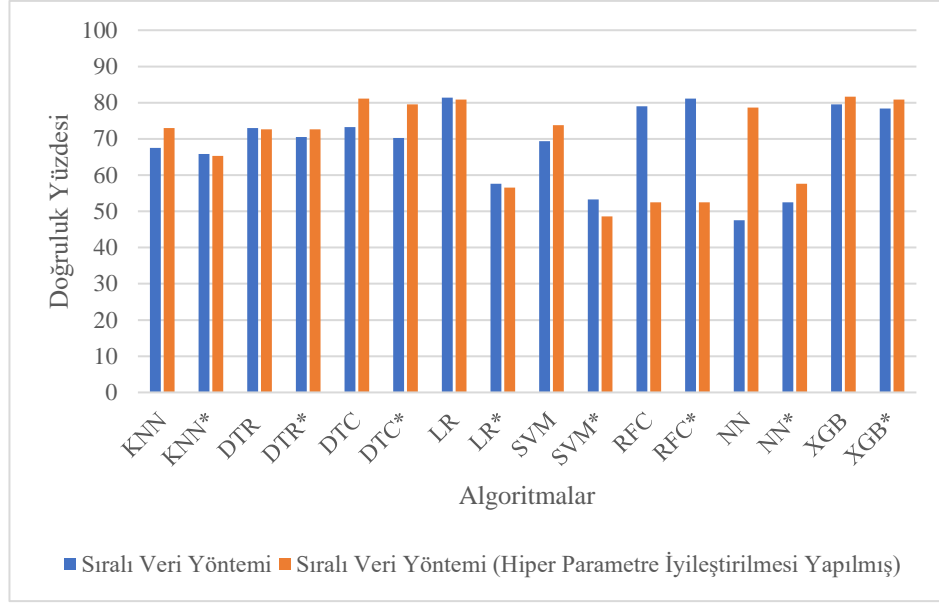
DTC %84,494 skor ile en yüksek f1 skoru almıştır.



Şekil 31. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu f1 skor kıyaslanması

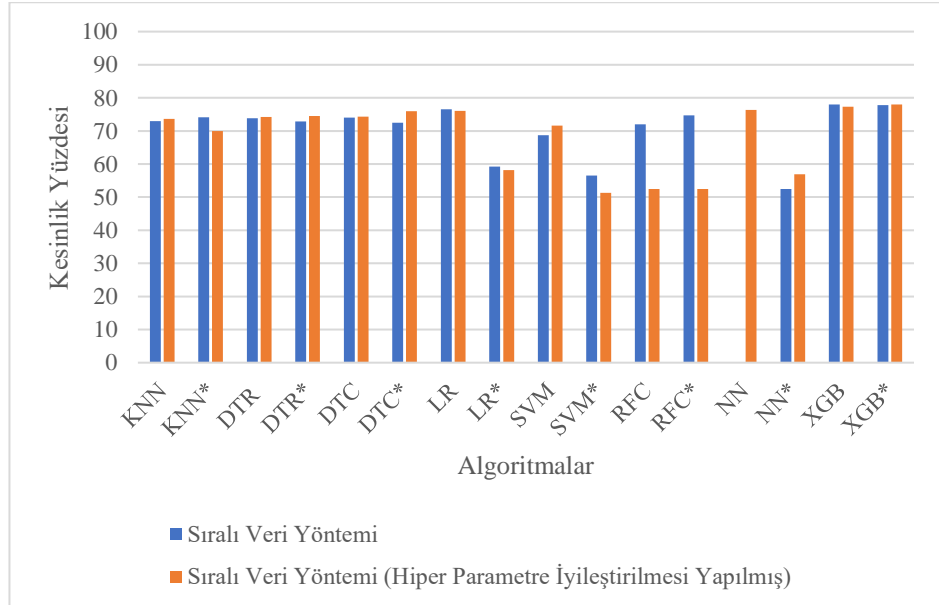
4.1.1.3. Karşılaştırmalı analiz

Algoritmaların sıralı veri tahminleme metodu ile kullanımında, hiper parametre optimizasyonu yapılmış ve hiper parametre optimizasyonu yapılmamış algoritma modelleri kıyaslanmıştır. Modellerin kıyaslanması sonucunda hiper parametre optimizasyonu yapılan algoritmaların doğruluk yüzdesini arttırdığı görülmüştür. Optimize edilmiş parametrelerin kullanımı ile en yüksek skoru %81,694 ile XGB almıştır. Optimize edilmiş parametrelerin kullanılmadığı modelde ise %81,421 ile en yüksek yüzdeliği LR almıştır. Kıyaslanma sonucunda, KNN*, DTR, LR, LR*, SVM*, RFC, ve RFC* algoritmaları optimize edilmiş parametreler ile yüzdelik skorları düştüğü tespit edilmiştir.



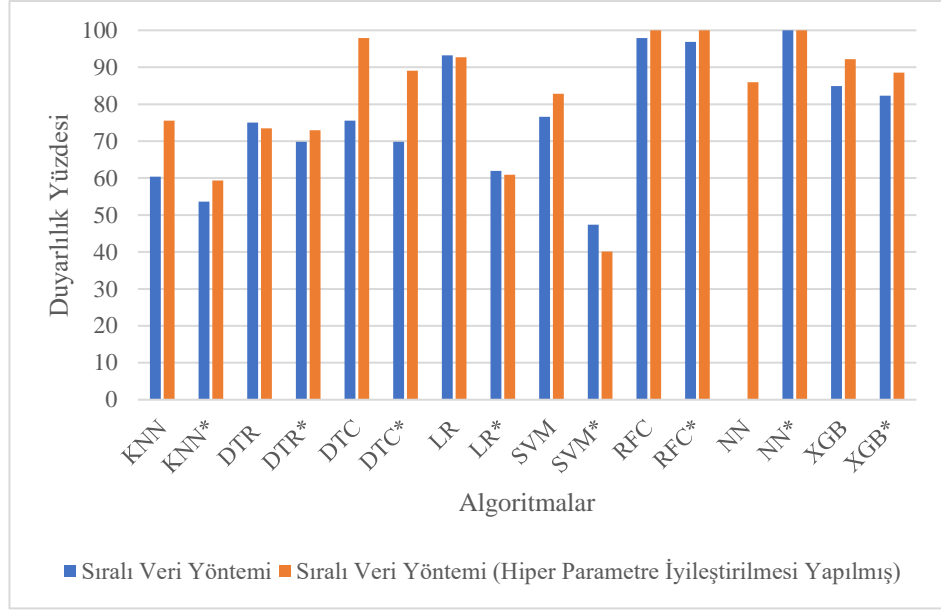
Şekil 32. Algoritmaların sıralı veri tahminleme yöntemi ile kullanımında doğruluk skoru kıyaslanması

Kesinlik skorları kıyaslanmasında, hiper parametre optimizasyonu yapılarak en yüksek skorun %77,982 ile XGB* aldığı görülmüştür. Kesinlik skoru için hiper parametre optimizasyonu yapılmasının daha başarılı sonuç elde ettiği görülmüştür. hiper parametre optimizasyonu kullanılması sonucunda, KNN*, LR, LR*, SVM*, RFC, RFC* ve XGB algoritmalarının kesinlik skorlarının düştüğü görülmüştür.



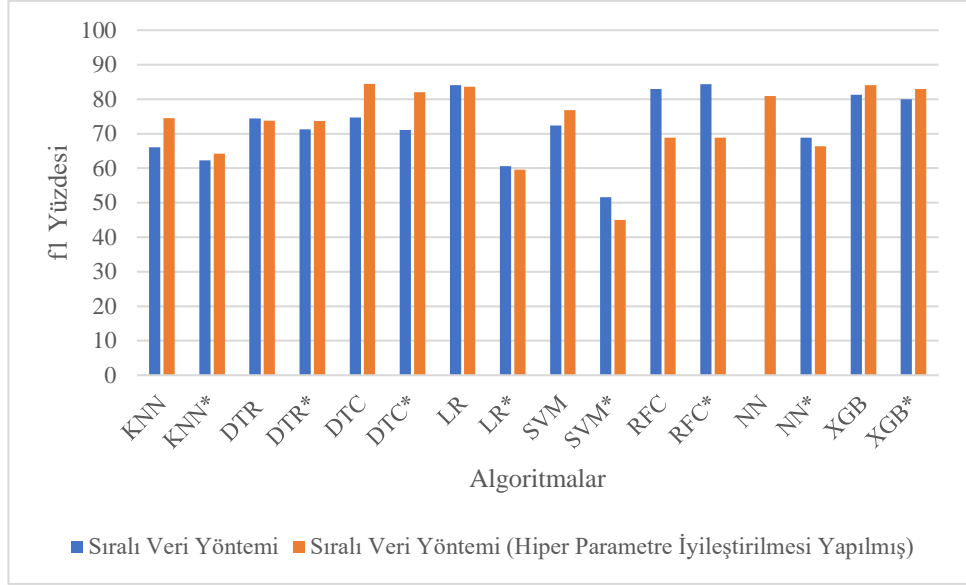
Şekil 33. Algoritmaların sıralı veri tahminleme metodu ile kullanımında kesinlik skoru kıyaslanması

Duyarlılık skoru optimize edilmiş hiper parametrelerin kullanımı ile artmıştır. Hiper parametre optimizasyonu olmadan %100 skor ile NN* algoritması en yüksek başarıyı sağlamıştır. Hiper parametre optimizasyonu ile yine NN* %100 skor sağlamıştır. Bunun yanı sıra, hiper parametre optimizasyonu sonrasında, DTR, LR, LR* ve SVM* algoritmalarının duyarlılık yüzdeleri düşmüştür.



Şekil 34. Algoritmaların sıralı veri tahminleme metodu ile kullanımında duyarlılık skoru kıyaslanması

F1 skorları arasında en yüksek başarı, %84,494 ile hiper parametre optimizasyonu uygulanmış DTC algoritması olmuştur. DTR, LR, LR*, SVM*, RFC, RFC* ve NN* algoritmaları, hiper parametre optimizasyonu uygulandığı durumda f1 skorları düşmüştür.



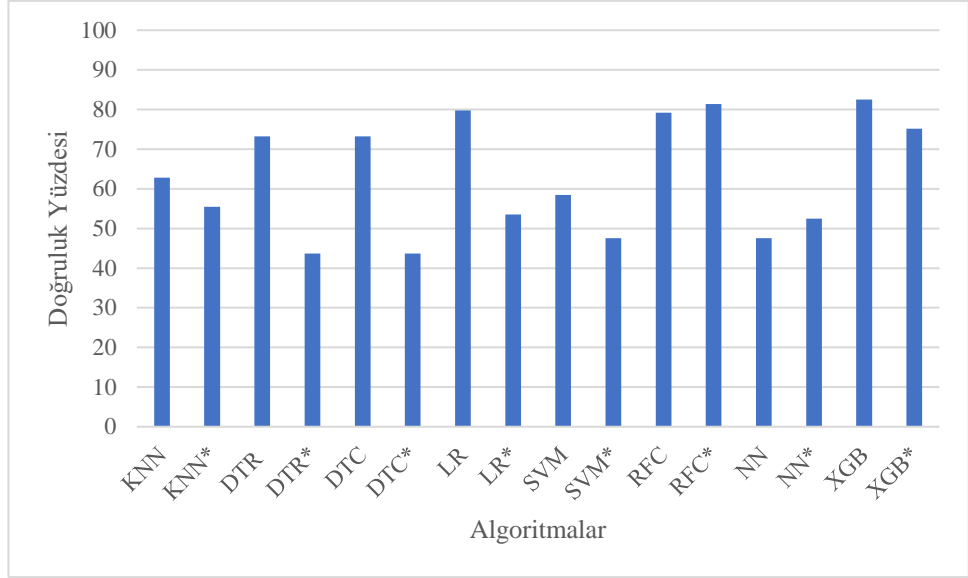
Şekil 35. Algoritmaların sıralı veri tahminleme metodu ile kullanımında f1 skoru kıyaslanması

4.1.2. Yüzdelik Ayrım Analizi

Yıl bazlı yapılan tahminleme için mevcut günün tahmininde %50 yüzdelik ayrım oranı kullanılmıştır. Algoritmalar, varsayılan parametre ve hiper parametre optimizasyonu yapılarak 2 farklı sonuç elde edilmek üzere kullanılmıştır. Doğruluk skorunun başarısının ölçülmesi için çıkartılan hata matrisleri ekteedir.

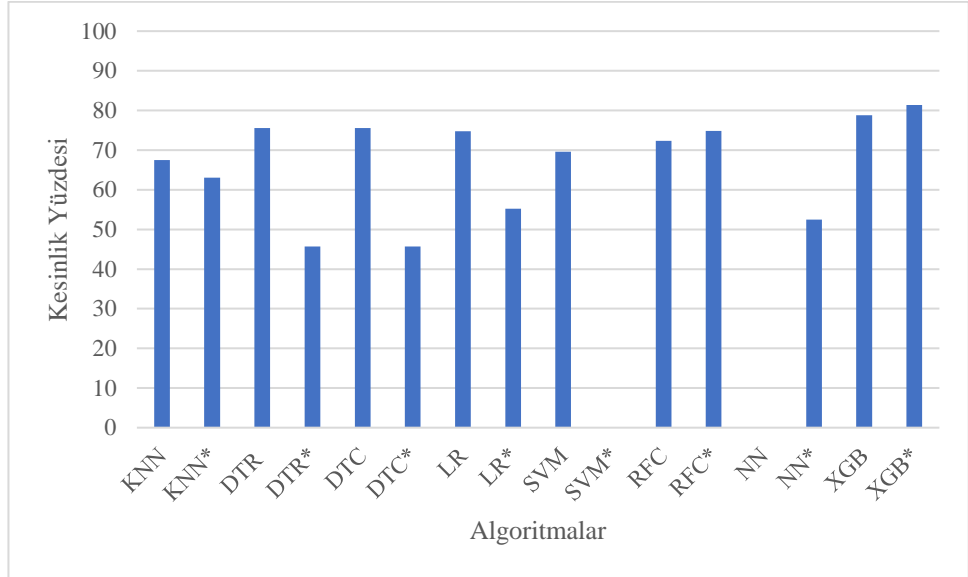
4.1.2.1. Varsayılan algoritma girdi parametreleri ile alınan sonuçlar

Bu analiz sonucunda en iyi doğruluk skorunu XGB ile %82,514 olmuştur.



Şekil 36. Algoritmaların yüzdelerik ayırım yöntemi kullanımı ile doğruluk skor kıyaslanması

Yüzdelerik ayırım yöntemi ile algoritmaların kesinlik skor kıyaslanması yapılmıştır. XGB* algoritması %81,366 skor ile en yüksek başarıyı sağlamıştır.

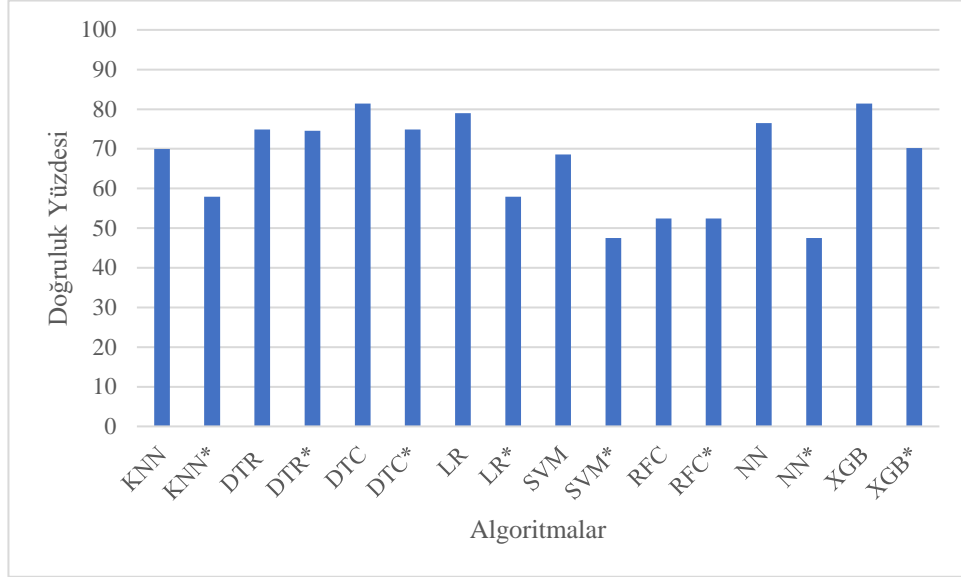


Şekil 37. Algoritmaların yüzdelerik ayırım yöntemi kullanımı ile kesinlik skor kıyaslanması

Sklearn kütüphanesinin ham hali ile yapılan kıyaslamada algoritmaların duyarlılık skorları kıyaslanmıştır. NN* %100 başarı ile en yüksek sonucu vermiştir.

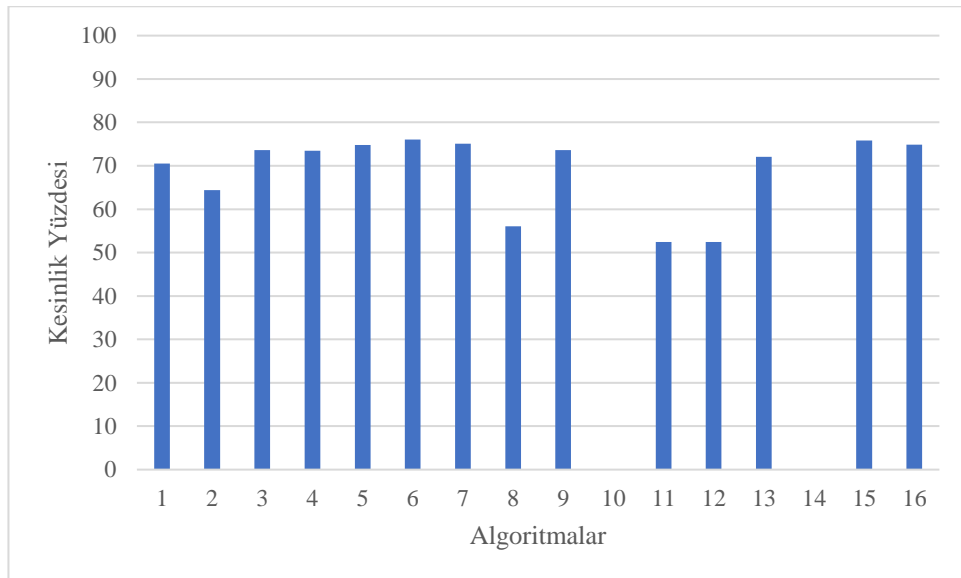
4.1.2.2. Hiper parametre optimizasyonu yapıldıktan sonra alınan sonuçlar

Tespit edilen parametreler sıralı veri tahminleme metodu ile uygulanmıştır. Bu analize göre en yüksek başarıyı %81,421 ile DTC ve XGB algoritması sağlamıştır.



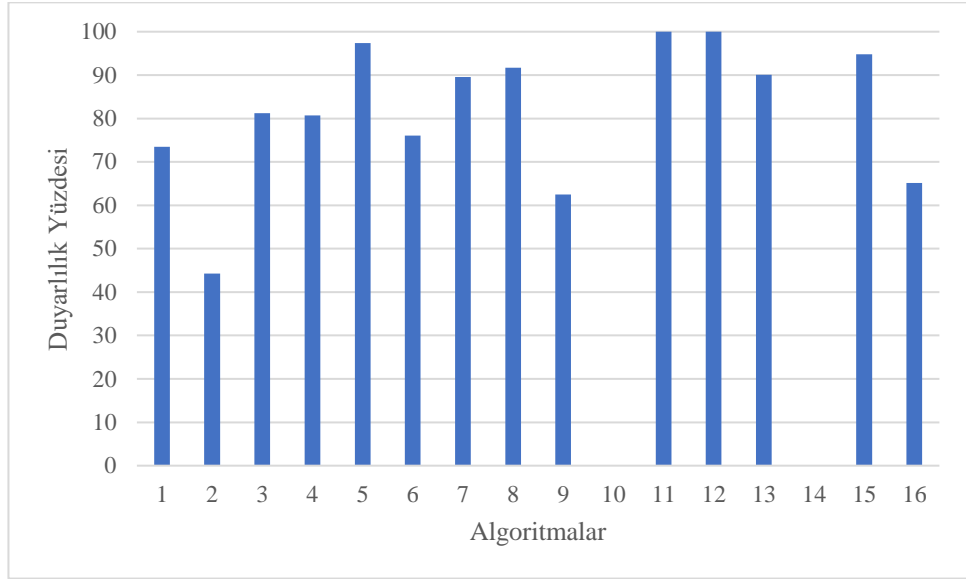
Şekil 38. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu doğruluk skor kıyaslanması

Algoritmaların hiper parametre optimizasyonu ile sıralı veri tahminleme metodu kullanımında en yüksek kesinlik skoru %76,042 ile DTC* sağlamıştır.



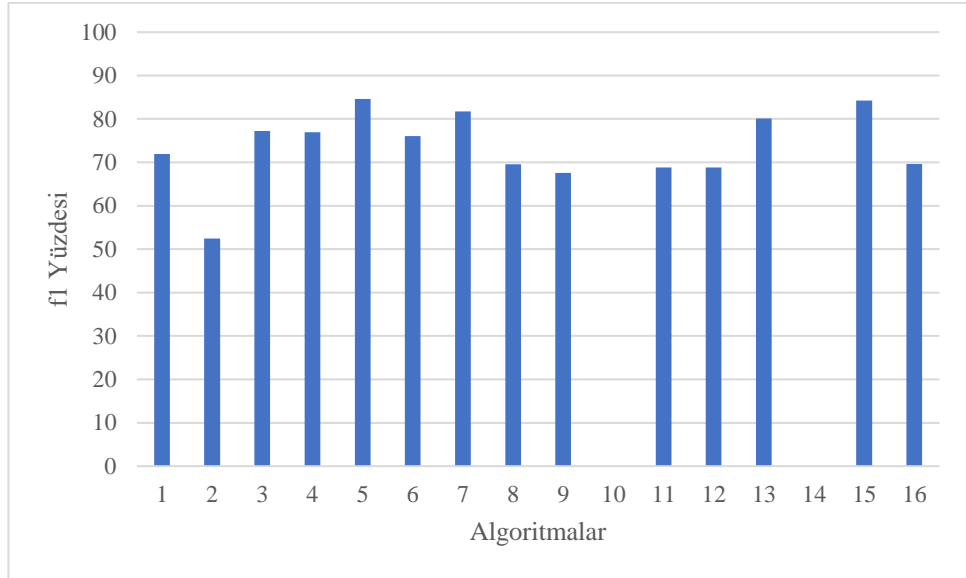
Şekil 39. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu kesinlik skor kıyaslanması

Algoritmaların duyarlılık skorlarının kıyaslanmasında en yüksek skorları %100 ile RCF ve RCF* sağlamıştır.

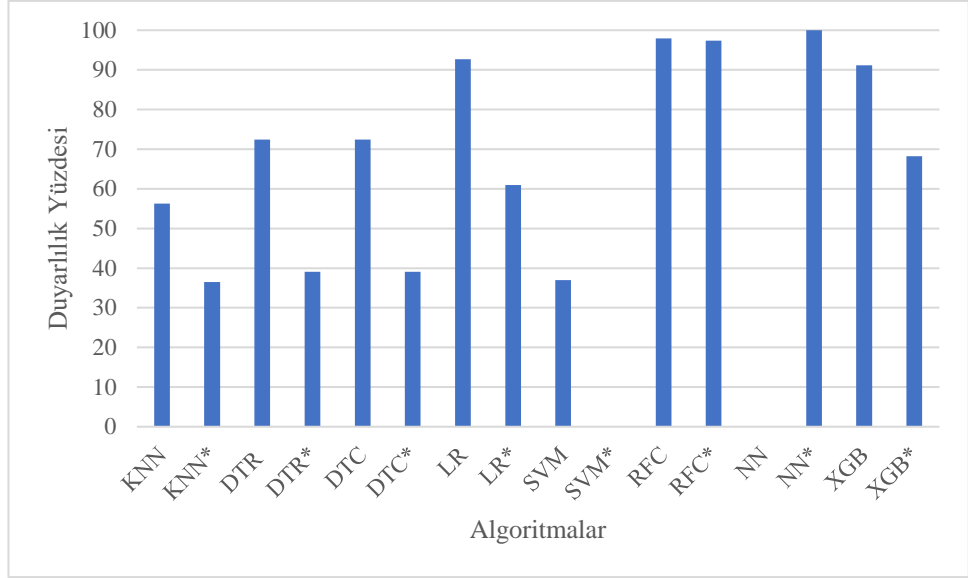


Şekil 40. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu duyarlılık skor kıyaslanması

DTC %84,615 skor ile en yüksek f1 skoru almıştır.

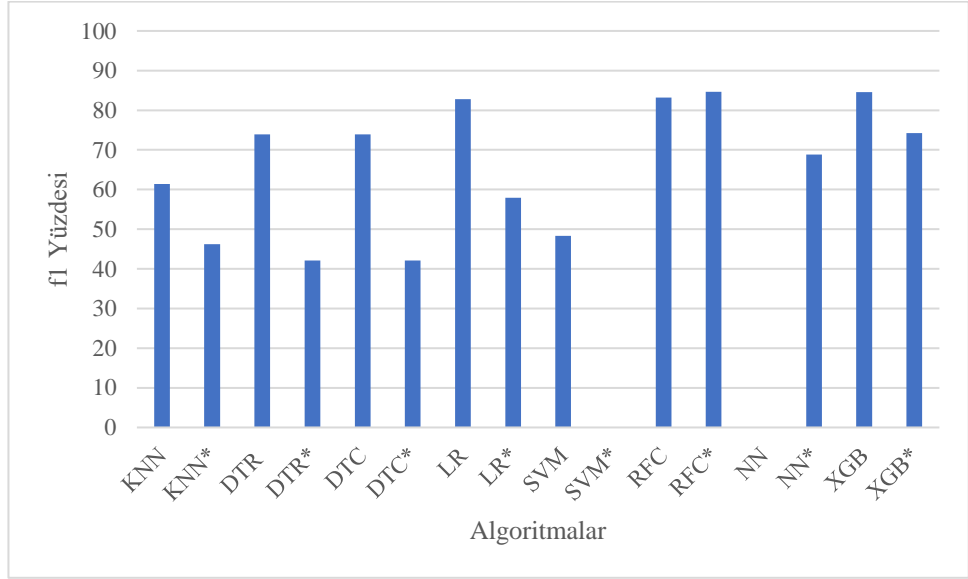


Şekil 41. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu f1 skor kıyaslanması



Şekil 42. Algoritmaların yüzelik ayırım yöntemi kullanımı ile duyarlılık skor kıyaslanması

Hiper parametre optimizasyonu yapılmamış algoritmaların f1 skor kıyaslanmasında RFC* %84,615 alarak en başarılı algoritma olduğunu göstermiştir.

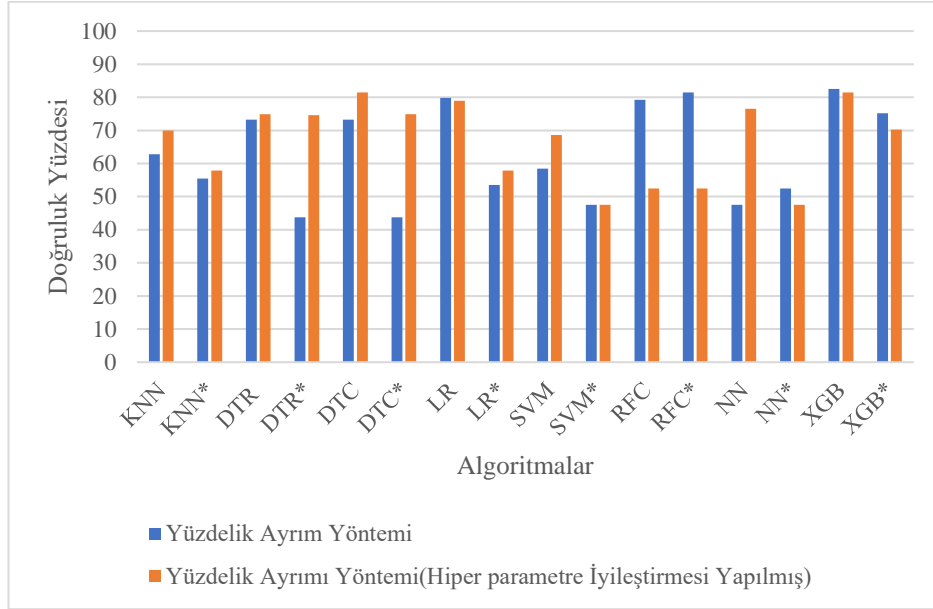


Şekil 43. Algoritmaların yüzelik ayırım yöntemi kullanımı ile f1 skor kıyaslanması

4.1.2.3. Karşılaştırmalı Analiz

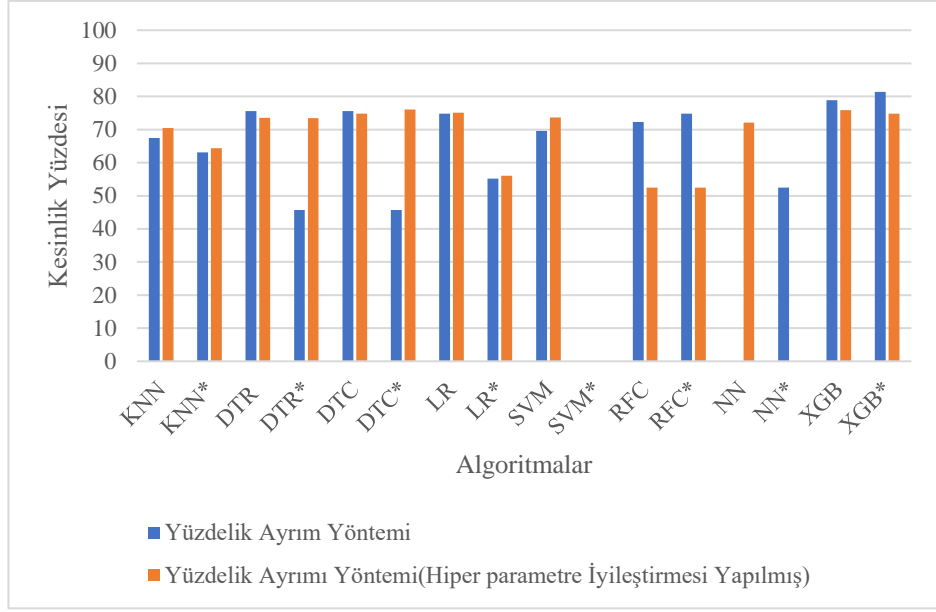
Algoritmaların yüzelik ayırım yöntemi ile kullanımında, hiper parametre optimizasyonu yapılmış ve hiper parametre optimizasyonu yapılmamış modeller kıyaslanmıştır. Modellerin kıyaslanması sonucunda hiper parametre optimizasyonu

yapılmış algoritmalarından KNN, KNN*, DTR, DTR*, DTC, DTC*, LR*, SVM ve NN' nin algoritmalarında doğruluk yüzdesinin yükseldiği görülmüştür. Hiper parametre optimizasyonu ile en yüksek skoru %81,421 ile DTC ve XGB almıştır. Hiper parametre optimizasyonu yapılmamış modelde ise %82,514 ile en yüksek yüzdeliği XGB almıştır.



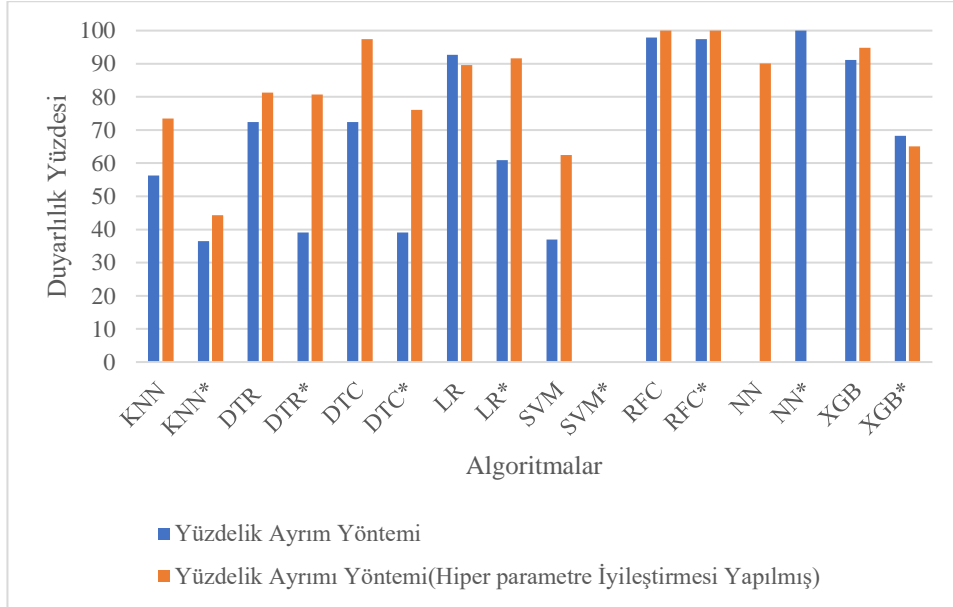
Şekil 44. Algoritmaların yüzdelik ayrımı yöntemi ile kullanımında doğruluk skoru kıyaslanması

Kesinlik skorları kıyaslanmasında, hiper parametre optimizasyonu ile en yüksek skorun %76,042 ile DTC* aldığı görülmüştür. Hiper parametre optimizasyonu yapılmadan alınan en yüksek skor %81,366 ile XGB* olmuştur. Kesinlik skoru için hiper parametre optimizasyonu yapılmasının skoru arttırdığı görülmüştür. Hiper parametre optimizasyonu yapılması sonucunda, KNN, KNN*, DTR*, DTC*, LR, LR*, SVM ve NN algoritmalarının kesinlik skorlarının arttığı görülmüştür.



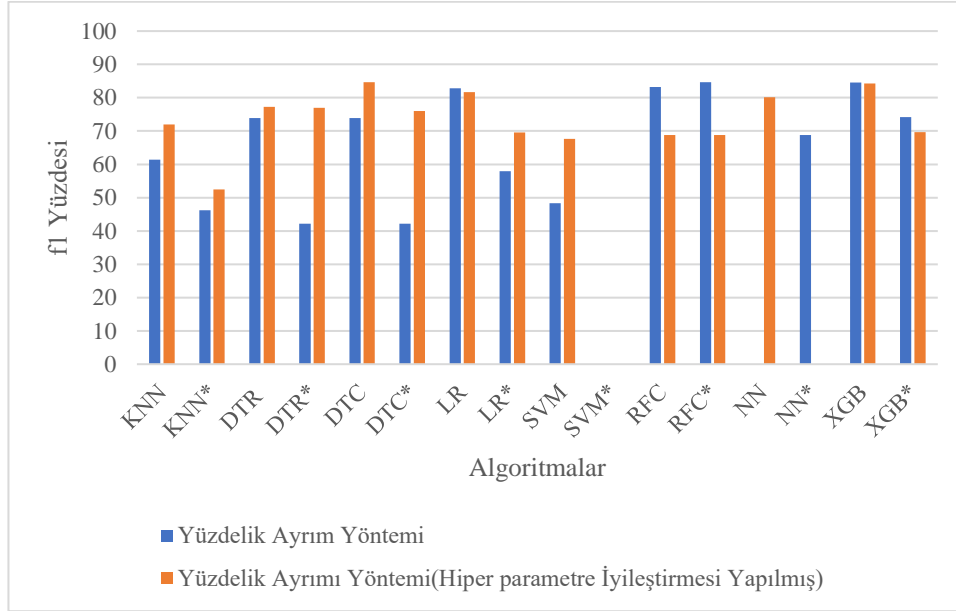
Şekil 45. Algoritmaların yüzdelik ayrımı yöntemi ile kullanımında kesinlik skoru kıyaslanması

Duyarlılık skoru hiper parametre optimizasyonu ile artmıştır. %100 skor ile RFC ve RFC* algoritmaları hiper parametre optimizasyonu ile en yüksek başarıyı sağlamıştır. %100 skor ile NN* algoritmaları hiper parametre optimizasyonu yapılmadan en yüksek skorları almıştır. Bunun yanı sıra, hiper parametre optimizasyonu ile, LR ve NN* algoritmalarının duyarlılık skorları düşmüştür.



Şekil 46. Algoritmaların yüzdelik ayrımı yöntemi ile kullanımında duyarlılık skoru kıyaslanması

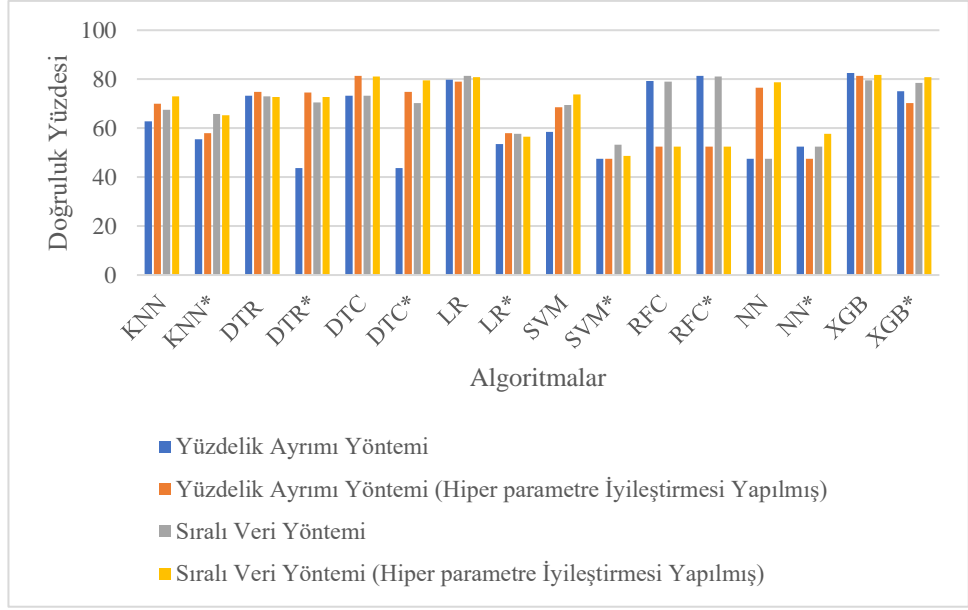
F1 skorları arasından en yüksek başarı, %84,615 ile hiper parametre optimizasyonu yapılmadan RFC ve yüzelik ayırım yöntemi (hiper parametre optimizasyonu yapılmış) DTC algoritması olmuştur. KNN, KNN*, DTR, DTR*, DTC, DTC*, LR*, SVM ve NN algoritmaları hiper parametre optimizasyonu ile yüzelikleri artmıştır.



Şekil 47. Algoritmaların yüzelik ayırımı yöntemi ile kullanımında f1 skoru kıyaslanması

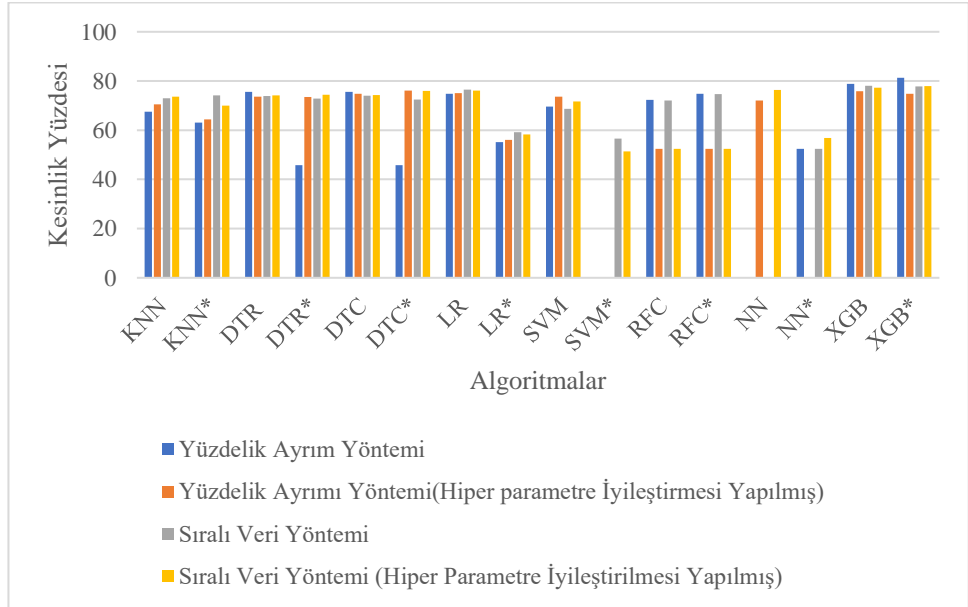
4.1.3. Karşılaştırmalı Algoritma Skorları

Yüzelik ayırım yöntemi ve sıralı veri tahminleme yöntemi kıyaslanmıştır. Yöntemlerin hiper parametre optimizasyonu yapılan ve yapılmayan durumları göz önünde bulundurulmuştur. Buna göre; doğruluk skoru için en başarılı algoritmaların DTC, LR ve XGB olduğu gözlenmiştir. DTC ve XGB yüzelik ayırım yöntemi (hiper parametre iyileştirmesi yapılmış) ile %81,421 skorunu almıştır. XGB yüzelik ayırım yöntemi %82,514 başarı yüzdesine ulaşmıştır. Sıralı veri tahminleme metodunda en yüksek başarıyı LR %81,421 ile elde etmiştir. Sıralı veri tahminleme metodu (hiper parametre optimizasyonu yapılmış) en yüksek başarıyı %81,694 ile XGB sağlamıştır. En başarılı veri ayırma yönteminin yüzelik ayırım yöntemi olduğu tespit edilmiştir.



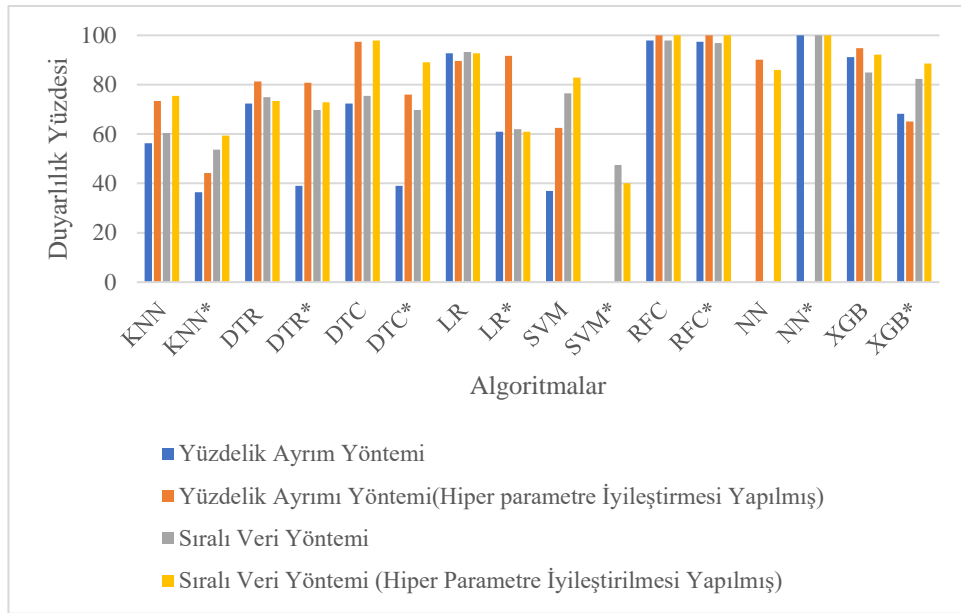
Şekil 48. Doğruluk skorlarının kıyaslanması

Kesinlik skor kıyaslamasında en yüksek skoru, %81,366 XGB* algoritması yüzdelik ayırım yöntemi ile almıştır. Bu algoritmayı takip eden skorlar, %77,99 ile XGB sıralı veri tahminleme metodu, %77,982 ile XGB* sıralı veri tahminleme metodu (hiper parametre optimizasyonu yapılmış) takip etmiştir. Yüzdelik ayırım yöntemi (hiper parametre optimizasyonu yapılmış) en yüksek başarıyı %76,042 ile DTC* sağlamıştır.



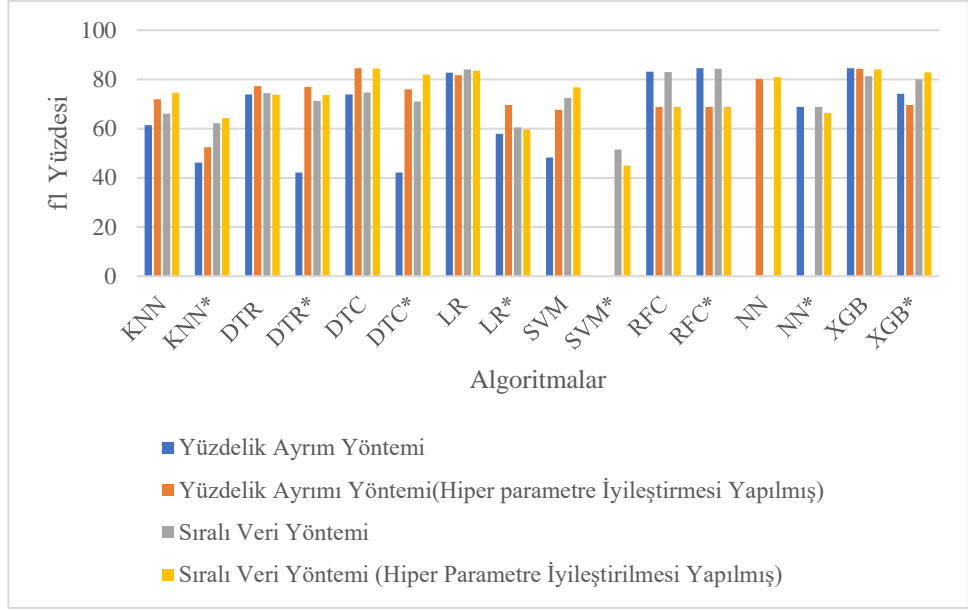
Şekil 49. Kesinlik skorlarının kıyaslanması

Duyarlılık skor kıyaslamasında en yüksek skorları, %100 ile NN* algoritması yüzdelik ayırım yöntemi ile, RFC ve RFC* algoritmaları yüzdelik ayırım yöntemi (hiper parametre optimizasyonu yapılmış) ile, NN* sıralı veri tahminleme metodu ile, RFC, RFC* ve NN* sıralı veri tahminleme metodu (hiper parametre optimizasyonu yapılmış) ile almıştır.



Şekil 50. Duyarlılık skorlarının kıyaslanması

F1 skor kıyaslanmasında en yüksek skoru %84,615 RFC* algoritması yüzdelik ayırım yöntemi ve DTC algoritması yüzdelik ayırım yöntemi (hiper parametre optimizasyonu yapılmış) ile almıştır. Bu algoritmaları takip eden skorlar, %84,494 ile DTC sıralı veri tahminleme metodu (hiper parametre optimizasyonu yapılmış), %84,354 ile RFC* sıralı veri tahminleme metodu takip etmiştir.



Şekil 51. F1 skorlarının kıyaslanması

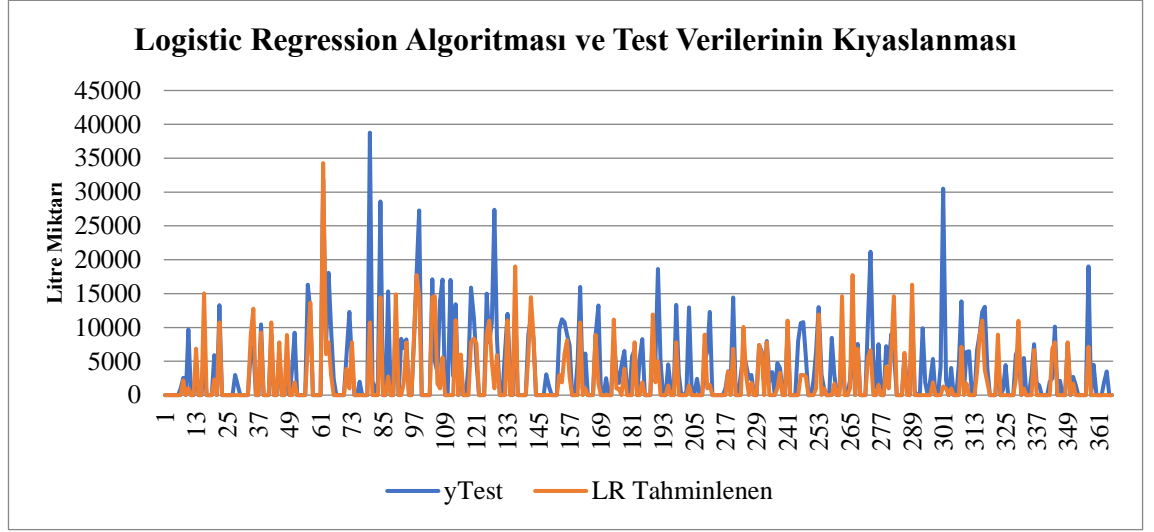
4.1.4. Akaryakıt Miktarı Tahminleme Skorları

2019 eğitim 2020 test edilen veri üzerinden mevcut gün litre tahminlemesi yapılmıştır. Yapılan tahminlemenin, algoritmalar baz alınarak MSE skorları tespit edilmiştir. MSE skorları alımı içi veri standart ölçeklendirilmiş olup sklearn kütüphanesinden faydalanılmıştır.

Tablo 18. Algoritmaların Litre Tahminleme İçin Aldıkları MSE Skorları

MSE Skorları	
Algoritma	Skor
KNN	1,63357
KNN*	1,17077
DTR	1,35425
DTR*	3,16113
DTC	1,63348
DTC*	2,95461
LR	0,83092
LR*	1,93860
SVM	2,11628
SVM*	1,07571
RFC	1,77494
RFC*	1,77494
NN	1,70505
NN*	1,55426
XGB	1,45713
XGB*	1,49869

Algoritmalarından en iyi skoru 0,83092 olduğu ve LR algoritmasının aldığı tespit edilmiştir.



Şekil 52. Veri Ve XGBoost Algoritmasının Tahmininin Grafıksel Gösterimi

4.2. 2019-2020 Verisi kullanılarak 2021 ilk çeyrek tahmini: Mevcut Gün Tahmini

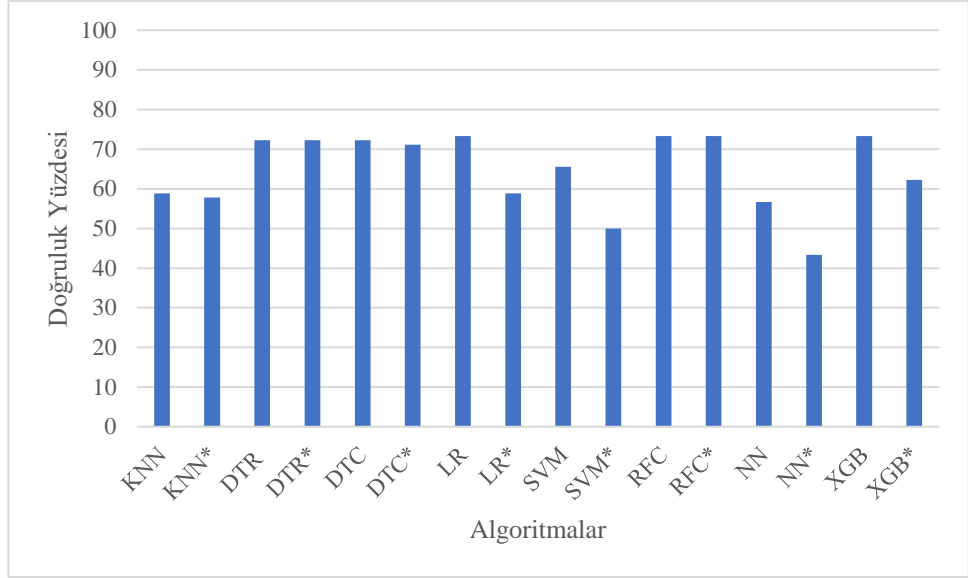
Çiftçilerin 2019, 2020, ve 2021 ilk çeyrek akaryakıt alım verileri toplanmıştır. 2019 ve 2020 verileri eğitim için kullanılıp 2021 verisi test için kullanılmıştır.

4.2.1. Sıralı Veri Tahminleme Analizi

2019 ve 2020 verileri eğitim olarak kullanılıp 2021 ilk çeyrek verileri test olarak kullanılmıştır. Yapılan modelde eğitim verileri 2021 1 Ocak tarihindeki alımı tahminlemiştir. Devam eden döngüde tahminlenen veri tahmin dizisine atanmıştır. Tahminlenen verinin test içerisinde olan değeri eğitimi verimize dahil edilmiştir. 2 Ocak 2021 tarihinin tahminlemesi 1 Ocak 2021 verisinin değeri eğitime dahil olarak yapılmıştır. 2021 ilk çeyrek bulunan veriler döngü halinde eğitim 1 arttırılarak devam etmiştir. Elde edilen sonuçlar raporlanmıştır. Hata matrisleri ektedir.

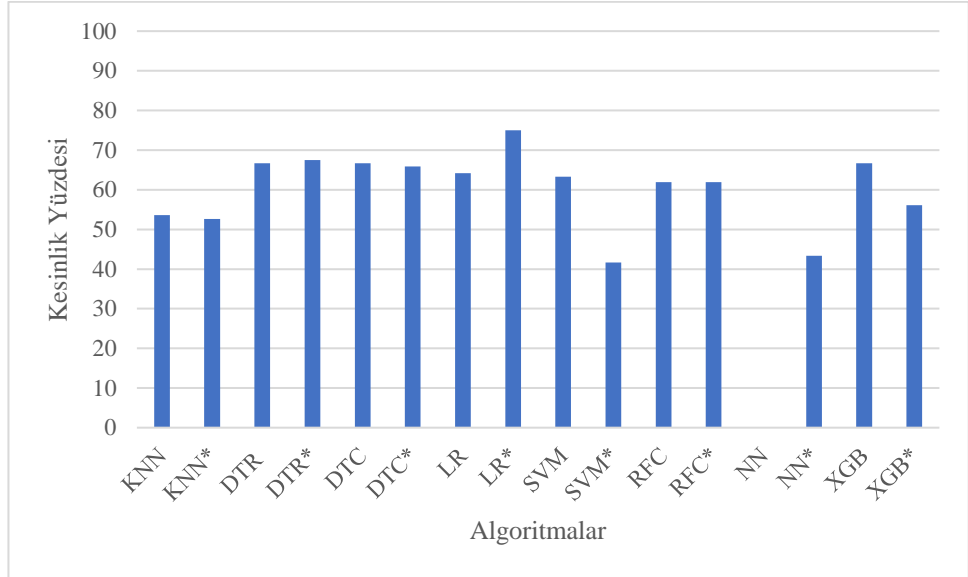
4.2.1.1. Varsayılan algoritma girdi parametreleri ile alınan sonuçlar

Sıralı veri tahminleme metodu kullanımında en yüksek doğruluk skorunu %73,333 ile XGB, RFC ve RFC* algoritmaları almıştır.



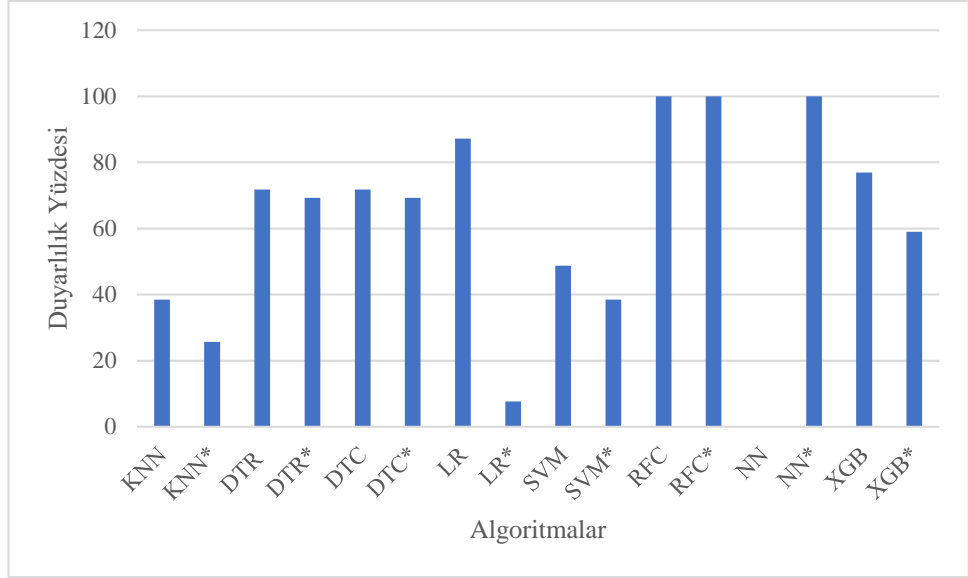
Şekil 53. Algoritmaların sıralı veri tahminleme metodu ile doğruluk skor kıyaslanması

Algoritmaların sıralı veri tahminleme metodu ile kıyaslanmasında kesinlik skorları için en yüksek skoru alanlar, %66,667 ile DTR, DTC ve XGB olmuştur.



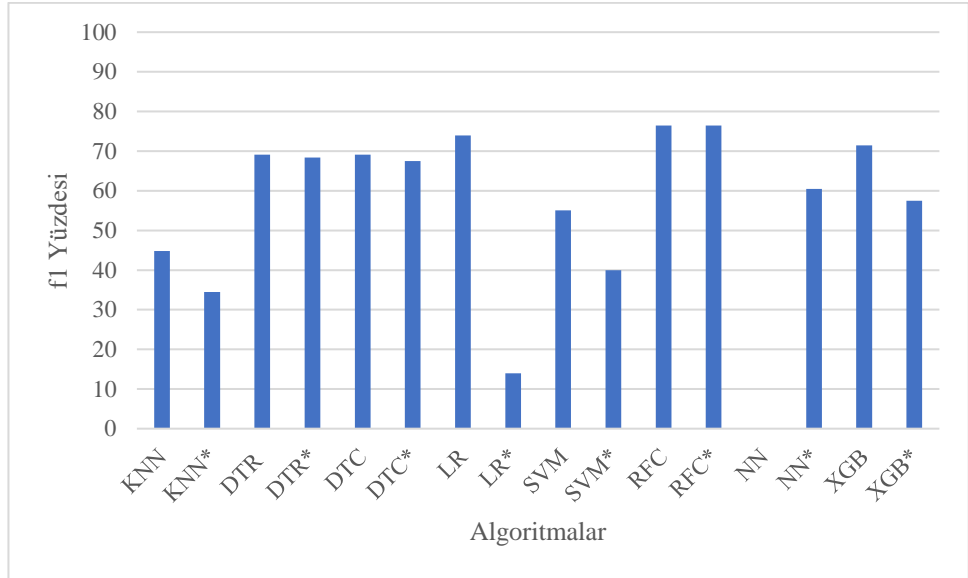
Şekil 54. Algoritmaların sıralı veri tahminleme metodu ile kesinlik skor kıyaslanması

Kullanılan sıralı veri tahminleme metodu sonucunda algoritmalar arasındaki en iyi duyarlılık skoru %100 ile RFC, RFC* ve NN* almıştır.



Şekil 55. Algoritmaların sıralı veri tahminleme metodu ile duyarlılık skor kıyaslanması

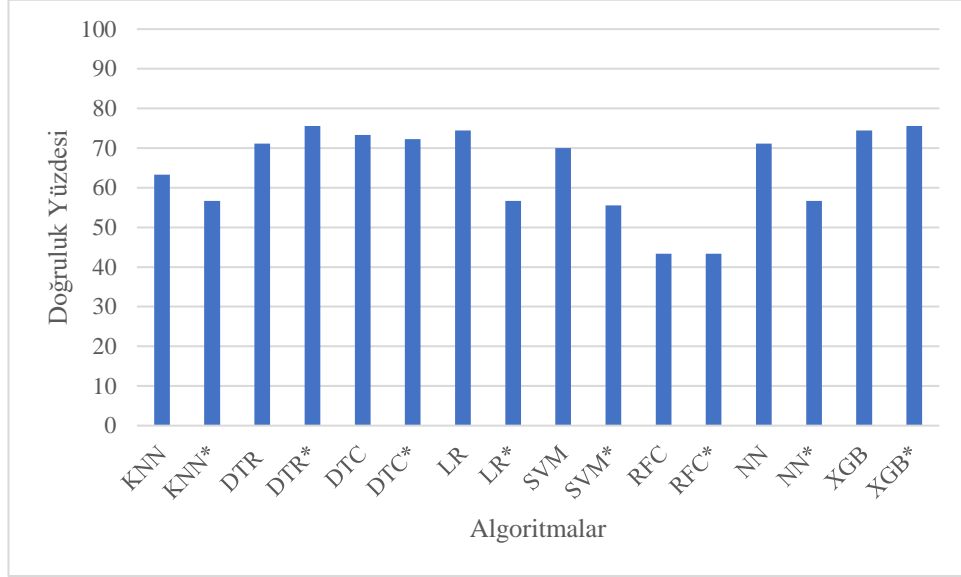
Algoritmalarından RFC ve RFC*, %76,471 skor ile en iyi f1 skoru yapmıştır.



Şekil 56. Algoritmaların sıralı veri tahminleme metodu ile f1 skor kıyaslanması

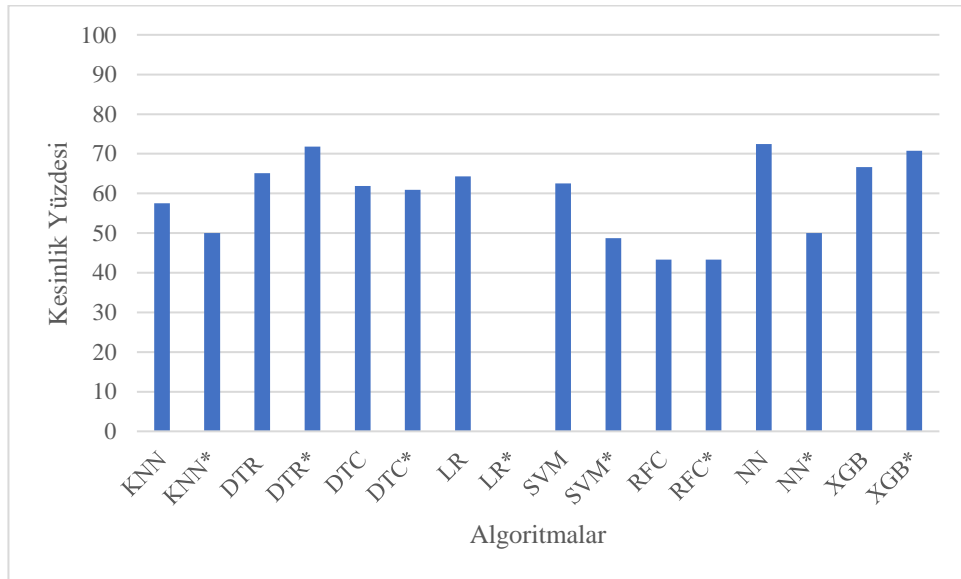
4.2.1.2. Hiper parametre optimizasyonu yapıldıktan sonra alınan sonuçlar

Hiper parametre optimizasyonu yapılan modellere sıralı veri tahminleme metodu uygulanmıştır. Bu analize göre en yüksek başarıyı %75,556 ile DTR* ve XGB* algoritmaları sağlamıştır.



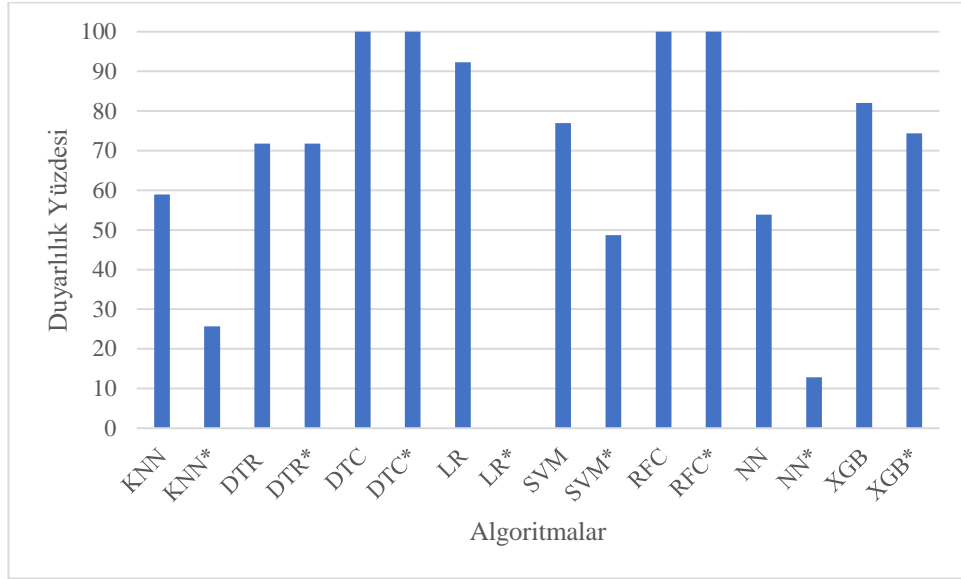
Şekil 57. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu doğruluk skor kıyaslanması

Algoritmaların hiper parametre optimizasyonu ile sıralı veri tahminleme motodu kullanımında en yüksek kesinlik skoru %72,414 ile NN sağlamıştır.



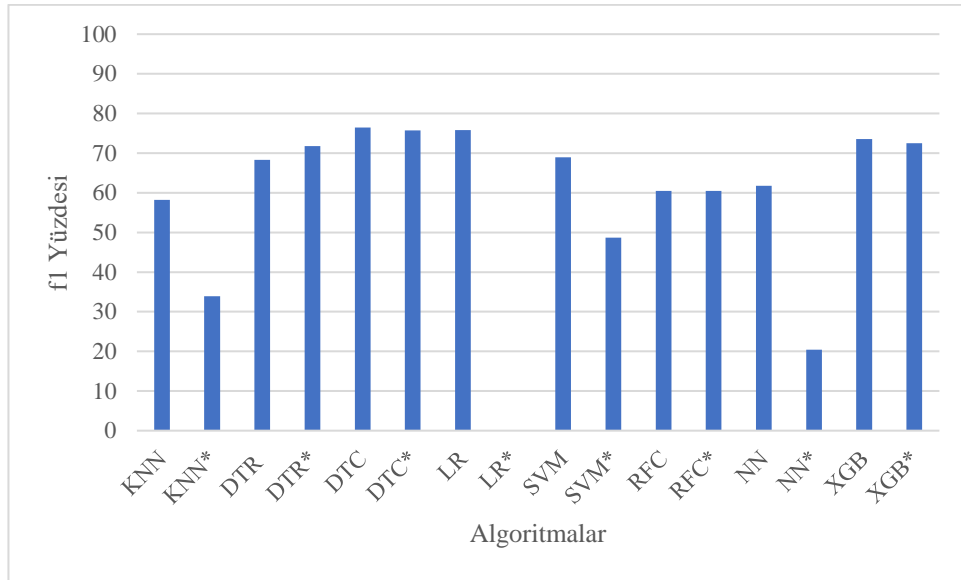
Şekil 58. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu kesinlik skor kıyaslanması

Algoritmaların duyarlılık skorlarının kıyaslanmasında en yüksek skorları %100 ile DTC, DTC*, RCF ve RCF* sağlanmıştır.



Şekil 59. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu duyarlılık skor kıyaslanması

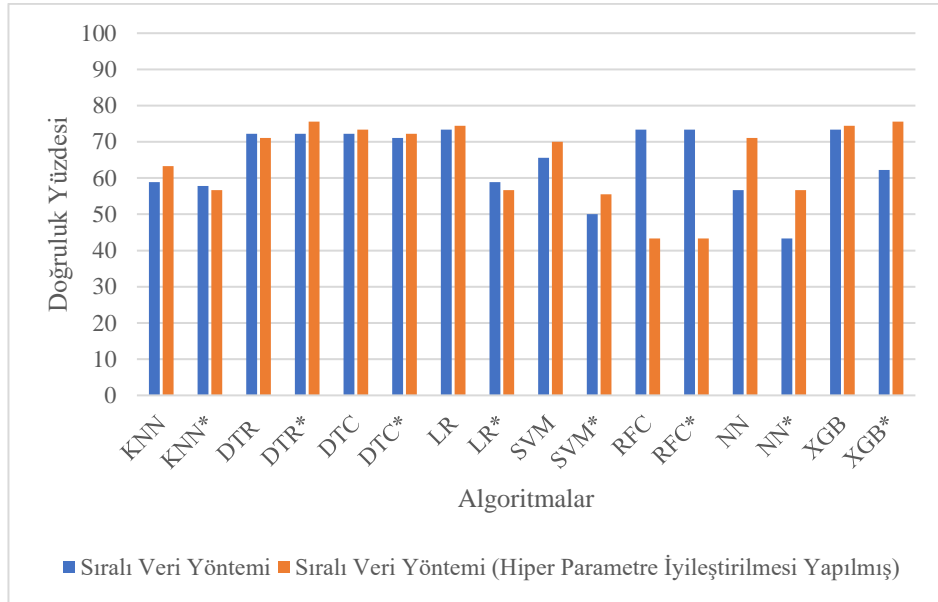
DTC %76,471 skor ile en yüksek f1 skoru almıştır.



Şekil 60. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu f1 skor kıyaslanması

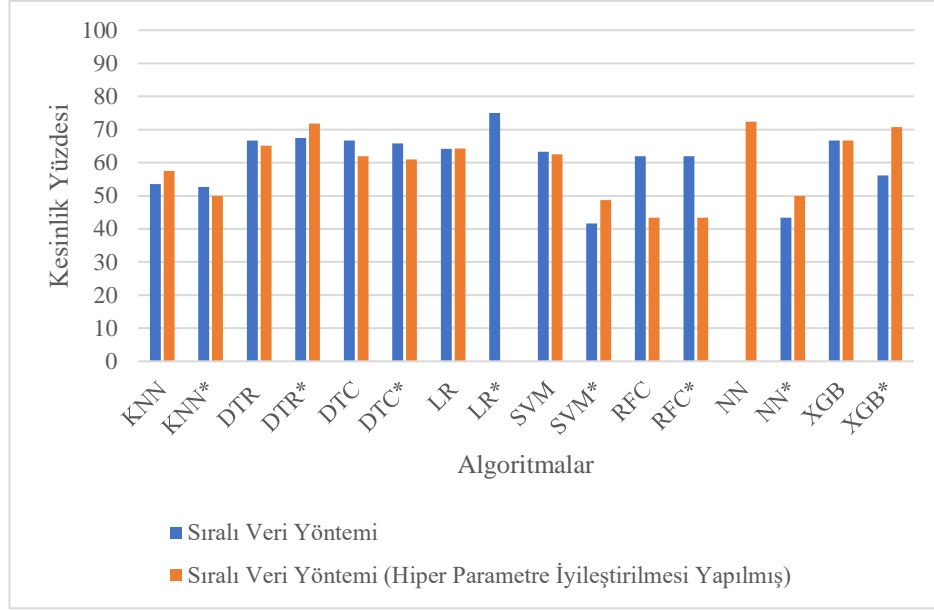
4.2.1.3. Karşılaştırmalı analiz

Algoritmaların sıralı veri tahminleme metodu ile kullanımında, hiper parametre optimizasyonu yapılmış ve yapılmamış modeller kıyaslanmıştır. Modellerin kıyaslanması sonucunda hiper parametre optimizasyonu yapılmış modellerde doğruluk yüzdesinin arttırdığı görülmüştür. Hiper parametre optimizasyonu kullanımı ile en yüksek skoru %75,556 ile DTR* ve XGB* almıştır. Hiper parametre optimizasyonu yapılmayan modelde ise %73,333 ile en yüksek yüzdeliği LR, RFC, RFC* ve XGB almıştır. Kıyaslanma sonucunda, KNN*, DTR, LR*, RFC ve RFC* algoritmaları hiper parametre optimizasyonu ile yüzdelik skorları düştüğü tespit edilmiştir.



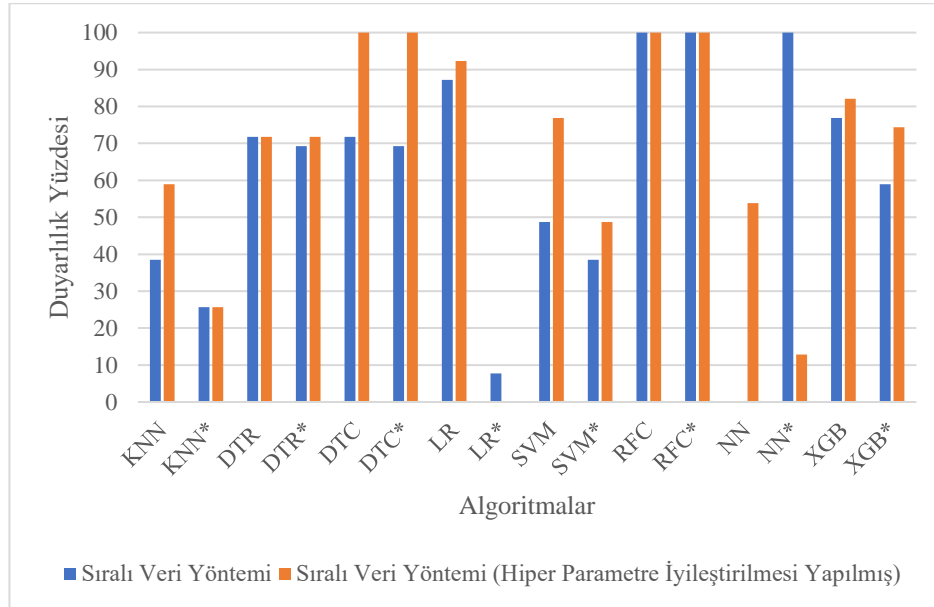
Şekil 61. Algoritmaların sıralı veri tahminleme metodu ile kullanımında doğruluk skoru kıyaslanması

Kesinlik skorları kıyaslanmasında, hiper parametre optimizasyonu yapılarak en yüksek skorun %72,414 ile NN aldığı görülmüştür. Kesinlik skoru için hiper parametre optimizasyonunun daha başarısız sonuç elde ettiği görülmüştür. Hiper parametre optimizasyonu sonucunda, KNN*, DTR, DTC, DTC*, LR*, SVM, RFC ve RFC* algoritmalarının kesinlik skorlarının düştüğü görülmüştür.



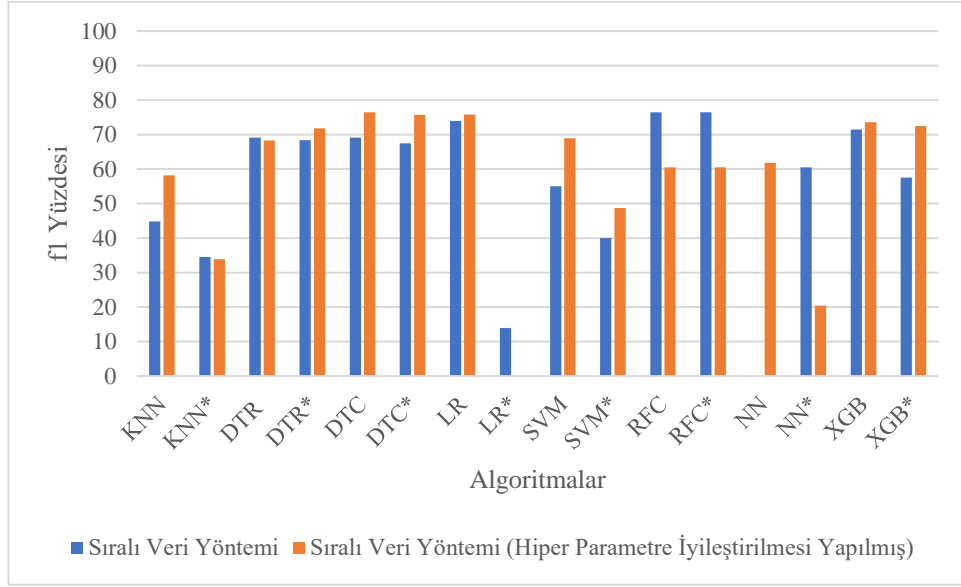
Şekil 62. Algoritmaların sıralı veri tahminleme metodu ile kullanımında kesinlik skoru kıyaslanması

Duyarlılık skoru hiper parametre optimizasyonu ile artmıştır. Hiper parametre optimizasyonu yapılmadan %100 skor ile RFC, RFC* ve NN* algoritmaları en yüksek başarıyı sağlamıştır. Hiper parametre optimizasyonu ile DTC, DTC*, RFC ve RFC* %100 skor sağlamıştır. Bunun yanı sıra, hiper parametre optimizasyonu ile, LR ve NN* algoritmalarının duyarlılık yüzdeleri düşmüştür.



Şekil 63. Algoritmaların sıralı veri tahminleme metodu ile kullanımında duyarlılık skoru kıyaslanması

F1 skorları arasından en yüksek başarı, %76,471 ile hiper parametre optimizasyonu yapılmış DTC ve hiper parametre optimizasyonu yapılmamış RFC ve RFC* algoritmaları olmuştur. KNN*, DTR, LR*, RFC, RFC* ve NN* algoritmaları hiper parametre optimizasyonu yapılması durumunda yüzdelikleri düşmüştür.



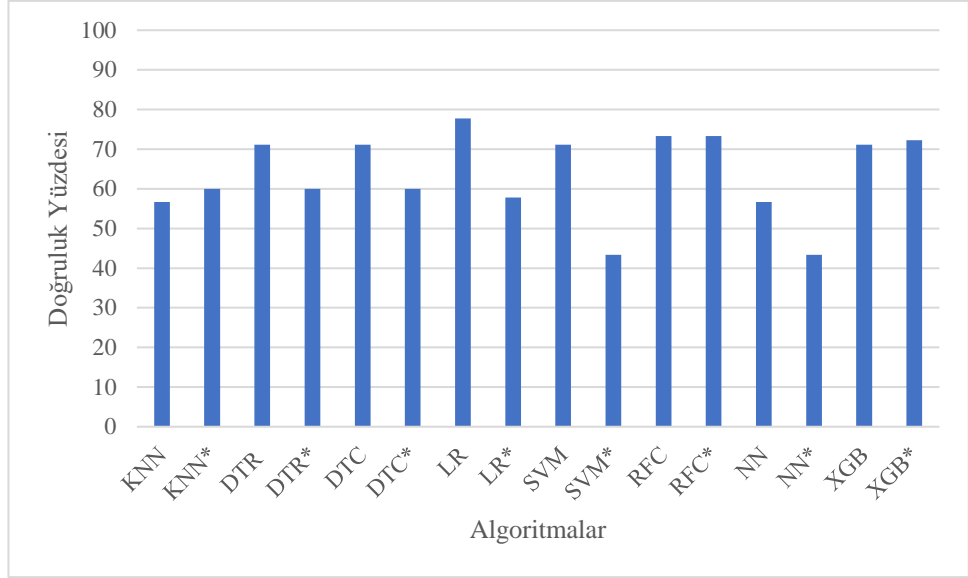
Şekil 64. Algoritmaların sıralı veri tahminleme metodu ile kullanımında f1 skoru kıyaslanması

4.2.2. Yüzdelik Ayrım Analizi

Yüzdelik ayrımı yöntemi için 2019 ve 2020 verileri eğitim 2021 ilk çeyrek verileri test olarak ayrılmıştır. %11 test, %89 eğitim yüzdelik ayrımı yapılmıştır. Hata matrislere eke eklenmiştir.

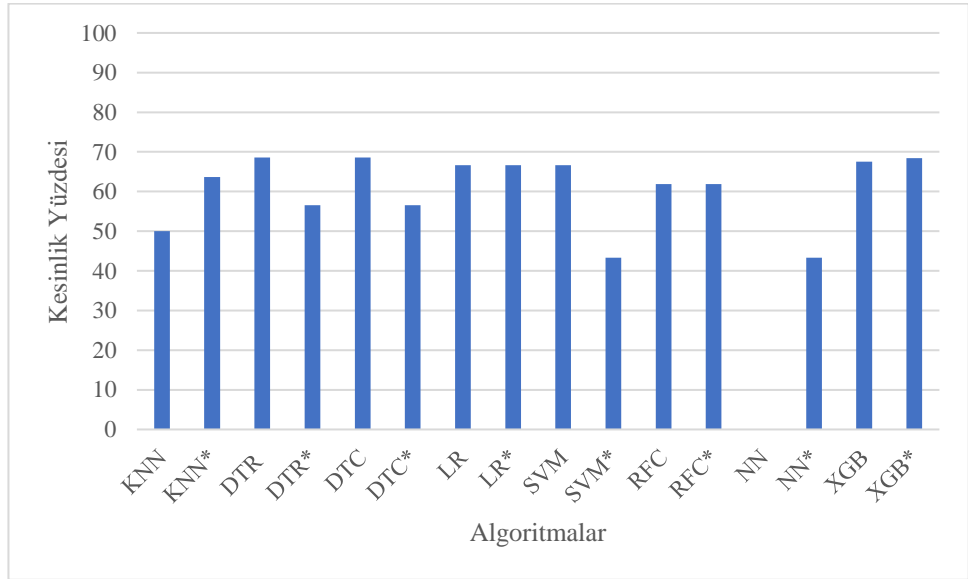
4.2.2.1. Varsayılan algoritma girdi parametreleri ile alınan sonuçlar

Yüzdelik ayrım yöntemi, algoritmaların varsayılan hali ile yapılan analizin doğruluk skorları tespit edilmiştir. Bu analiz sonucunda en iyi skor RFC ve RFC* ile %73,333 olmuştur.



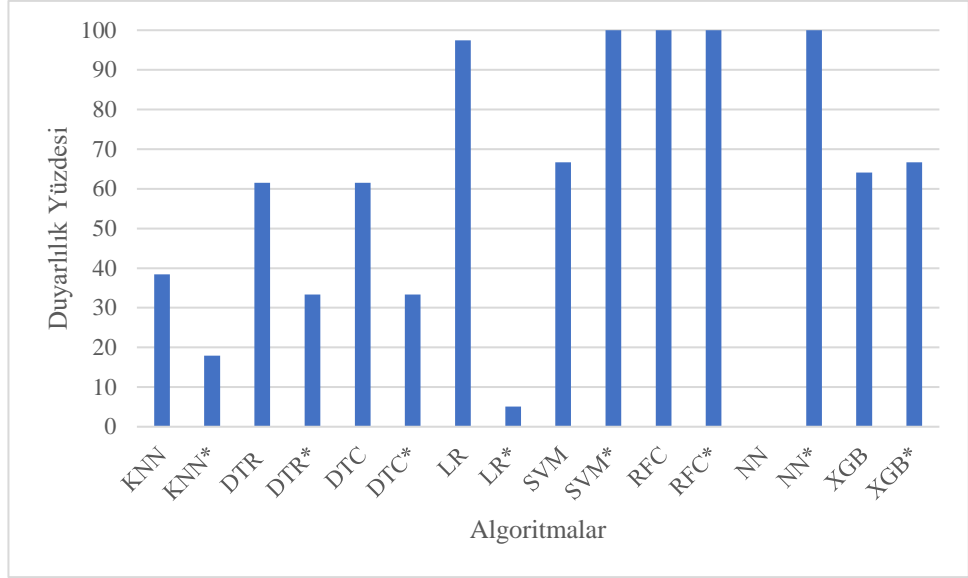
Şekil 65. Algoritmaların yüzelik ayırım yöntemi kullanımı ile doğruluk skor kıyaslanması

Yüzelik ayırım yöntemi ile algoritmaların kesinlik skor kıyaslanması yapılmıştır. XGB* algoritması %68,421 skor ile en yüksek başarıyı sağlamıştır.



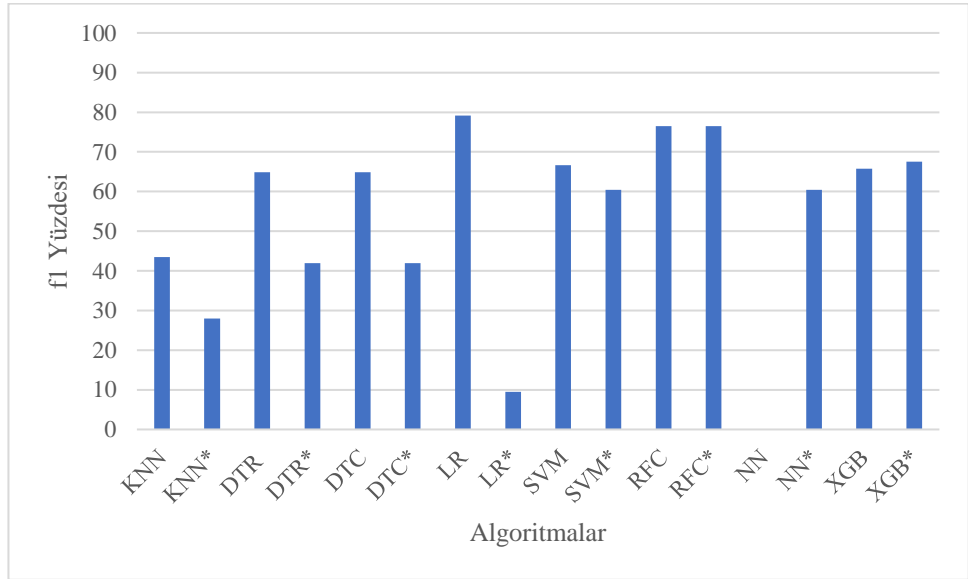
Şekil 66. Algoritmaların yüzelik ayırım yöntemi kullanımı ile kesinlik skor kıyaslanması

Algoritmaların duyarlılık skorları kıyaslanmıştır. SVM* ve NN* %100 başarı ile en yüksek sonucu vermiştir.



Şekil 67. Algoritmaların yüzelik ayırım yöntemi kullanımı ile duyarlılık skor kıyaslanması

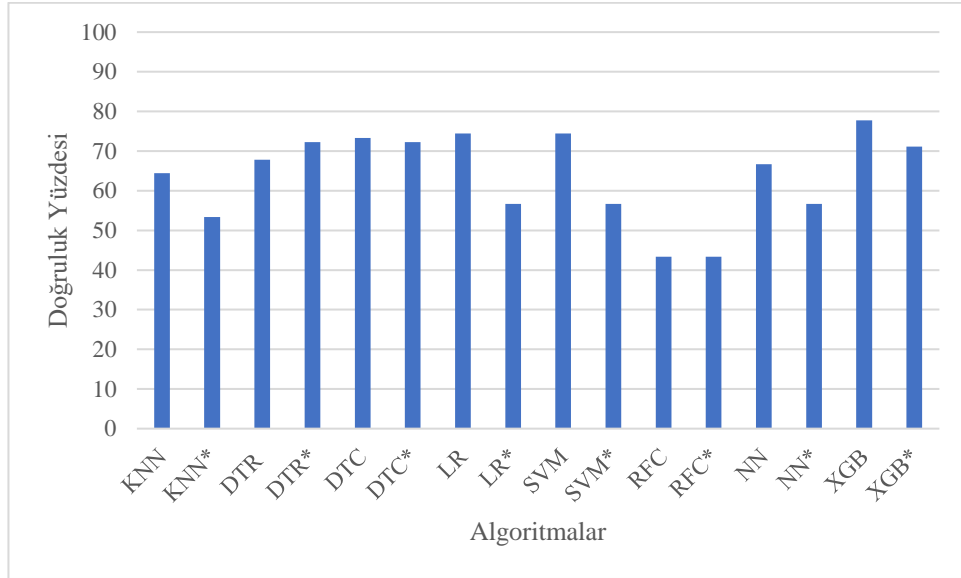
Algoritmaların f1 skor kıyaslanması yapılmıştır. LR %79,167 olarak en yüksek f1 skoruna sahip olmuştur.



Şekil 68. Algoritmaların yüzelik ayırım yöntemi kullanımı ile f1 skor kıyaslanması

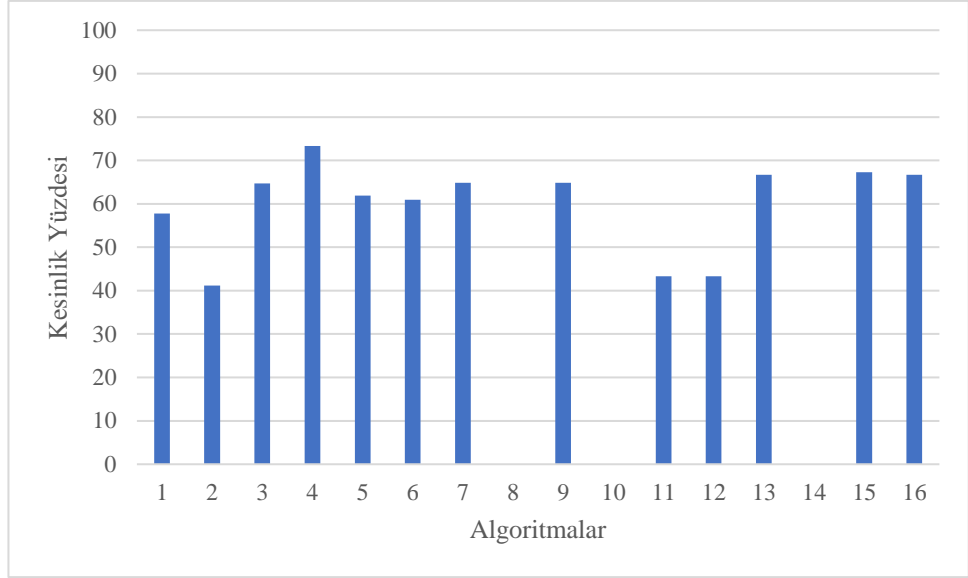
4.2.2.2. Hiper parametre optimizasyonu yapıldıktan sonra alınan sonuçlar

Yapılan hiper parametre optimizasyonu sonucundaki parametrelere sıralı veri tahminleme metodu uygulanmıştır. Bu analize göre en yüksek başarıyı %77,778 ile XGB algoritması sağlamıştır.



Şekil 69. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu doğruluk skor kıyaslanması

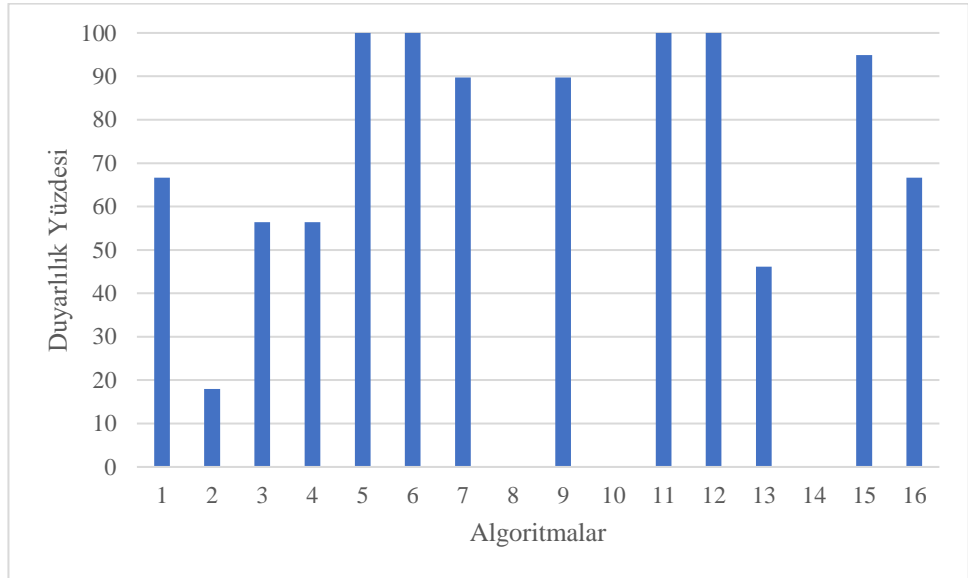
Algoritmaların hiper parametre optimizasyonu ile sıralı veri tahminleme metodu kullanımında en yüksek kesinlik skoru %73,333 ile DTR* sağlamıştır.



Şekil 70. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu kesinlik skor kıyaslanması

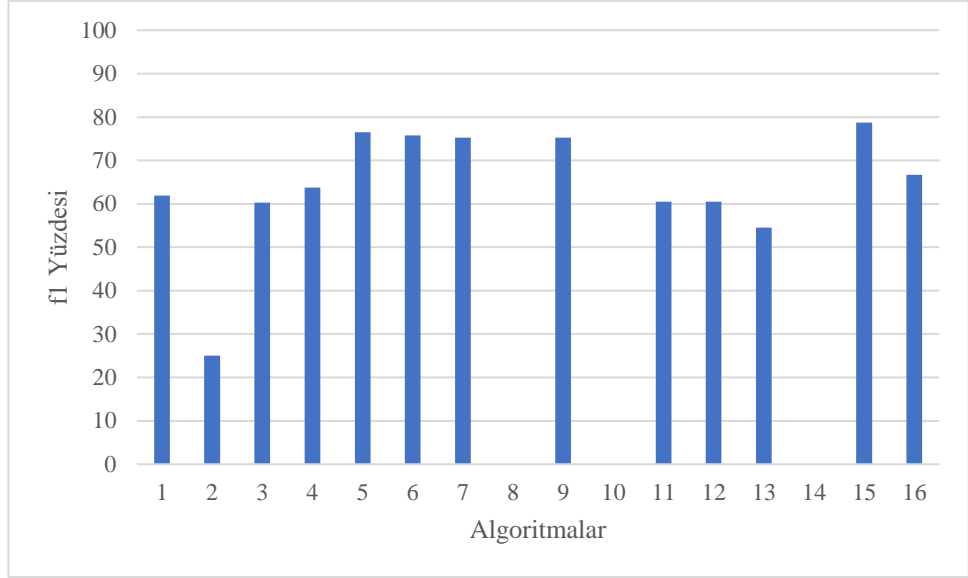
Algoritmaların duyarlılık skorlarının kıyaslanmasında en yüksek skorları %100 ile

DTC, DTC*, RCF ve RCF* sağlanmıştır.



Şekil 71. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu duyarlılık skor kıyaslanması

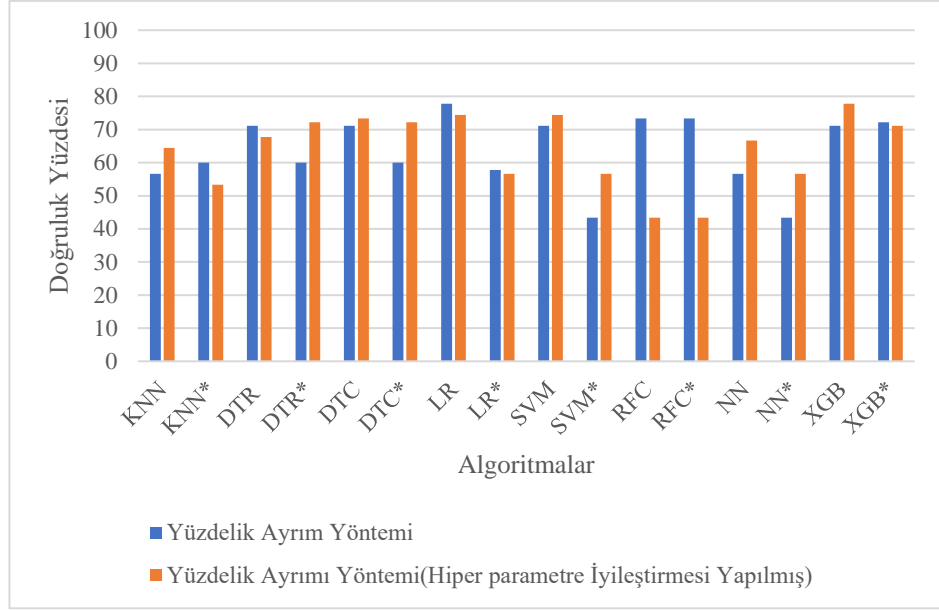
XGB* %78,723 skor ile en yüksek f1 skoru almıştır.



Şekil 72. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu f1 skor kıyaslanması

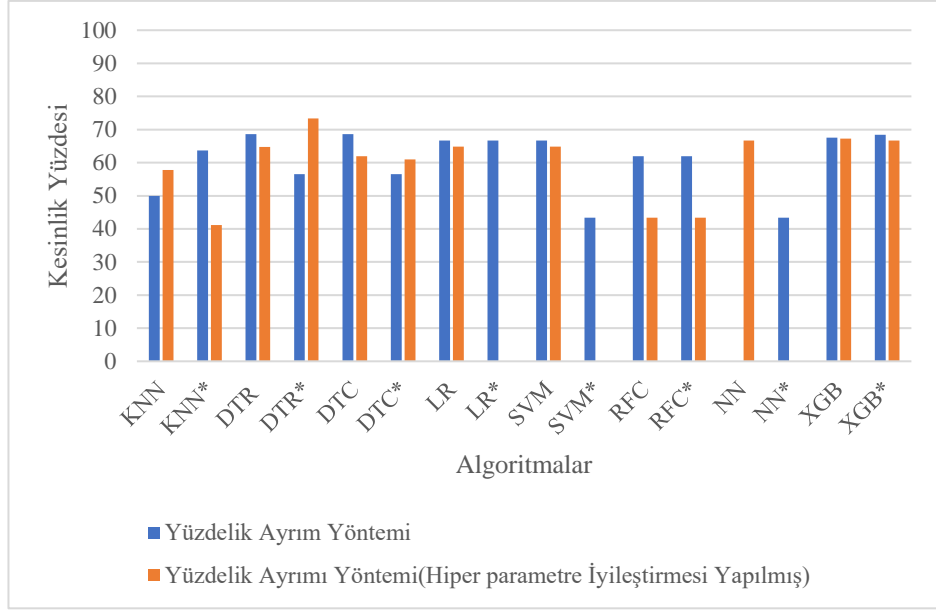
4.2.2.3. Karşılaştırmalı analiz

Algoritmaların yüzdelik ayırım yöntemi ile kullanımında, hiper parametre optimizasyonu yapılmış ve yapılmamış modellerin sonuçları kıyaslanmıştır. Modellerin kıyaslanması sonucunda hiper parametre optimizasyonunun doğruluk yüzdesini KNN, DTR*, DTC, DTC*, SVM, SVM*, NN, NN* ve XGB algoritmalarında yükseldiği görülmüştür. Hiper parametre optimizasyonu kullanımı ile en yüksek skoru %77,778 ile XGB almıştır. Hiper parametre optimizasyonunun kullanılmadığı modelde ise %77,778 ile en yüksek yüzdeliği LR almıştır.



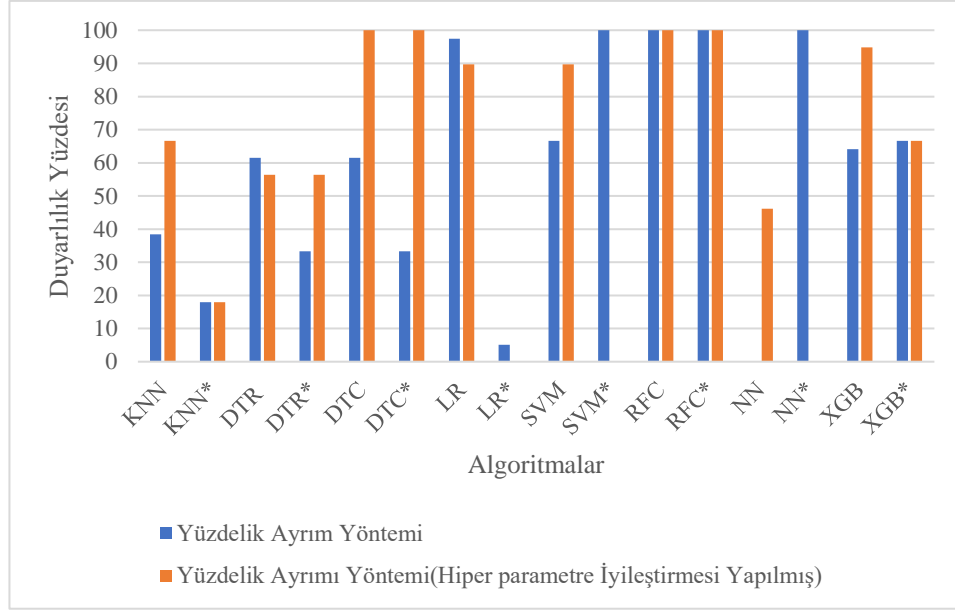
Şekil 73. Algoritmaların yüzdelik ayırımı yöntemi ile kullanımında doğruluk skoru kıyaslanması

Kesinlik skorları kıyaslanmasında, hiper parametre optimizasyonu ile en yüksek skoru %73,333 ile DTR* aldığı görülmüştür. Hiper parametre optimizasyonu ile kesinlik skorunun arttırdığı görülmüştür. Hiper parametre optimizasyonu kullanılması sonucunda, KNN, DTR*, DTC* ve NN algoritmalarının kesinlik skorlarının arttığı görülmüştür.



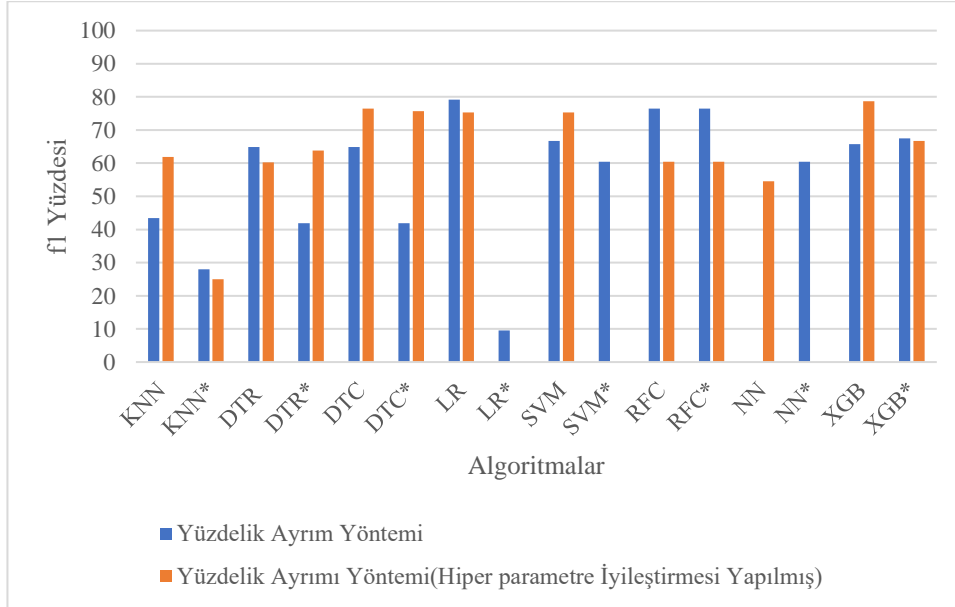
Şekil 74. Algoritmaların yüzdelik ayrımı yöntemi ile kullanımında kesinlik skoru kıyaslanması

Duyarlılık skoru hiper parametre optimizasyonu kullanımı ile artmıştır. %100 skor ile DTC, DTC* RFC ve RFC* algoritmaları hiper parametre optimizasyonu ile en yüksek başarıyı sağlamıştır. %100 skor ile SVM, RFC, RFC* ve NN* algoritmaları hiper parametre optimizasyonu yapılmadan en yüksek skorları almıştır. Bunun yanı sıra, hiper parametre optimizasyonu ile, DTR, LR, LR*, SVM* ve NN* algoritmalarının duyarlılık yüzdelikleri düşmüştür.



Şekil 75. Algoritmaların yüzdelik ayırımı yöntemi ile kullanımında duyarlılık skoru kıyaslanması

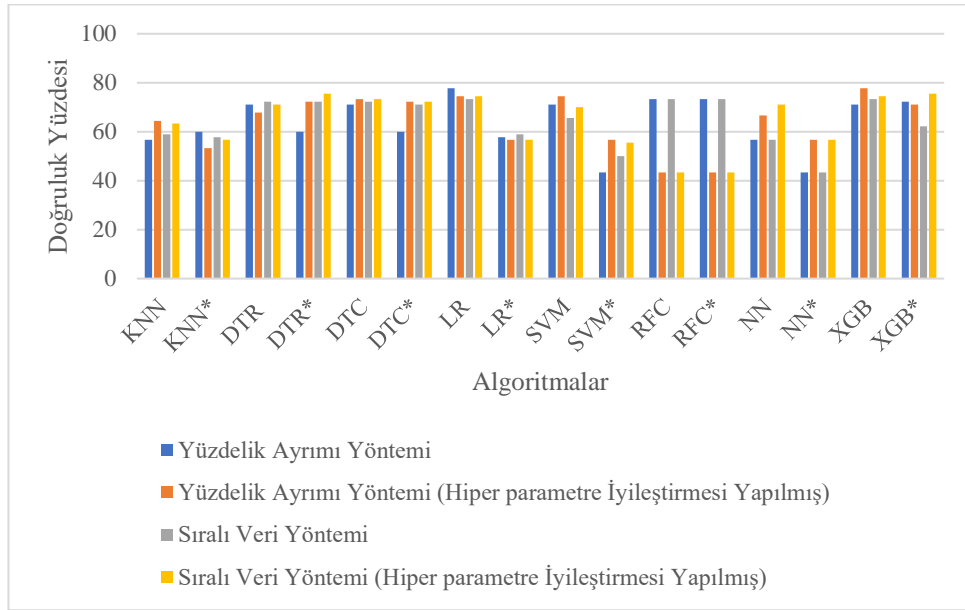
F1 skorları arasından en yüksek başarı, %79,167 ile hiper parametre optimizasyonu yapılmamış LR algoritması olmuştur. Hiper parametre optimizasyonu ile en yüksek skor %78,723 ile XGB algoritmasının olmuştur. KNN, DTR*, DTC, DTC*, SVM, NN ve XGB algoritmaları hiper parametre optimizasyonu kullanılmış durumda yüzdelikleri artmıştır.



Şekil 76. Algoritmaların yüzdelik ayırımı yöntemi ile kullanımında f1 skoru kıyaslanması

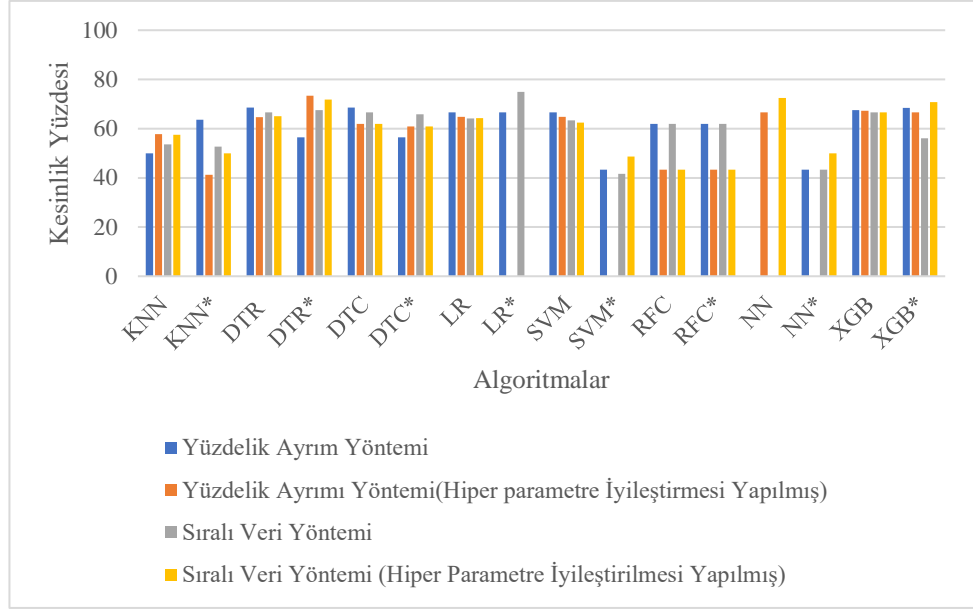
4.2.3. Karşılaştırmalı Algoritma Skorları

Yüzdelik ayırım yöntemi ve sıralı veri tahminleme metodu kıyaslanmıştır. Yöntemlerin hiper parametre optimizasyonu kullanıldığı ve kullanılmadığı durumlar göz önünde bulundurulmuştur. Buna göre; doğruluk skoru için en başarılı algoritmaların LR, XGB ve XGB* olduğu gözlenmiştir. LR yüzdelik ayırım yöntemi ve XGB yüzdelik ayırım yöntemi (hiper parametre optimizasyonu yapılmış) %77,778 başarı yüzdesine ulaşmıştır.



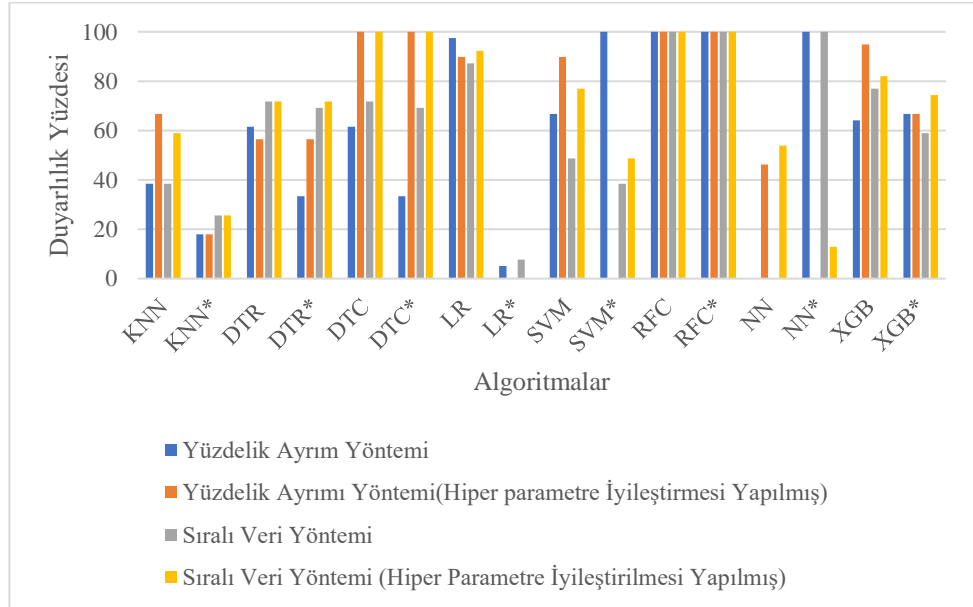
Şekil 77. Doğruluk skorlarının kıyaslanması

Kesinlik skor kıyaslamasında en yüksek skoru, %73,333 DTR* algoritması yüzdelik ayırım yöntemi (hiper parametre optimizasyonu yapılmış) ile almıştır. Bu algoritmayı takip eden skorlar, %72,414 ile NN sıralı veri tahminleme metodu (hiper parametre optimizasyonu yapılmış), %68,421 ile XGB* yüzdelik ayırım yöntemi takip etmiştir.



Şekil 78. Kesinlik skorlarının kıyaslanması

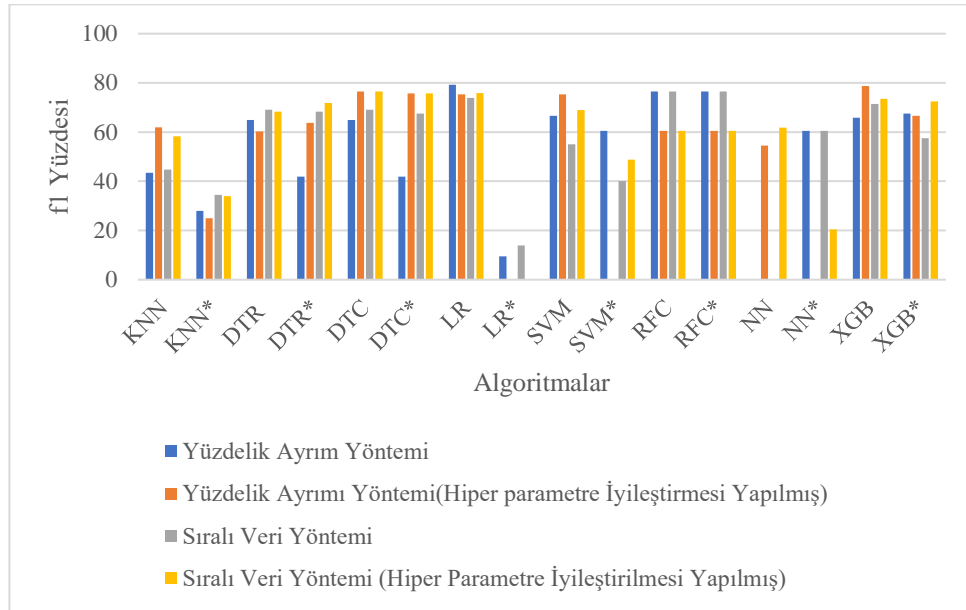
Duyarlılık skor kıyaslamasında en yüksek skorları, %100 ile SVM*, RFC, RFC* ve NN* algoritması yüzdelik ayırım yöntemi ile, DTC, DTC*, RFC ve RFC* algoritmaları yüzdelik ayırım yöntemi (hiper parametre optimizasyonu yapılmış) ve RFC, RFC* ve NN* sıralı veri tahminleme metodu, DTC, DTC*, RFC ve RFC* sıralı veri tahminleme metodu (hiper parametre optimizasyonu yapılmış) ile almıştır.



Şekil 79. Duyarlılık skorlarının kıyaslanması

F1 skor kıyaslanmasında en yüksek skoru yüzdelik ayırma yöntemi ile %79,167 olarak LR olmuştur. Bu skoru takip eden algoritmalar; yüzdelik ayırma yöntemi (hiper

parametre optimizasyonu yapılmış) kullanılarak %78,723 skoru alan XGB ve sıralı veri tahminleme metodu kullanılarak %76,471 skor sağlayan RFC ve RFC* olmuştur.



Şekil 80. F1 skorlarının kıyaslanması

4.2.4. Akaryakıt Miktarı Tahminleme Skorları

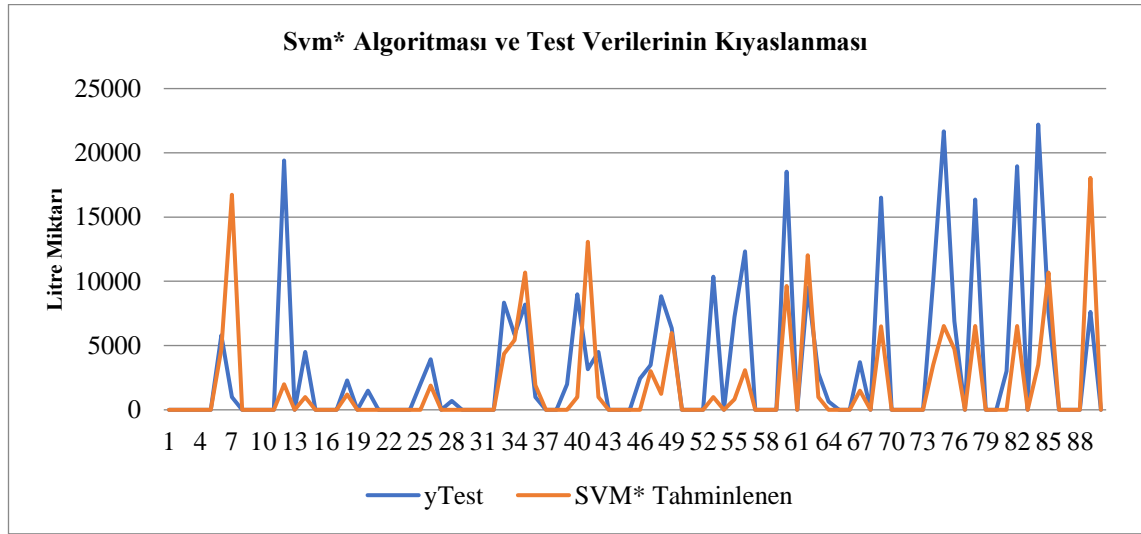
2019 ve 2020 verilerinin eğitim 2021 ilk çeyrek test edilen veri üzerinden mevcut gün litre tahminlemesi yapılmıştır. Yapılan tahminlemenin, algoritmalar baz alınarak MSE skorları tespit edilmiştir. MSE skorları alımı içi veri standart ölçeklendirilmiş olup sklearn kütüphanesinden faydalanılmıştır.

Tablo 19. Algoritmaların Litre Tahminleme İçin Aldıkları MSE Skorları

MSE Skorları	
Algoritma	Skor
KNN	2,59498
KNN*	0,90373
DTR	2,70791
DTR*	1,76341
DTC	2,83301
DTC*	1,19871
LR	1,49653
LR*	1,67912
SVM	2,57028
SVM*	0,82526
RFC	1,46302
RFC*	1,46302
NN	1,46302

NN*	1,46302
XGB	1,49210
XGB*	1,24391

Algoritmalarından en iyi skoru 0,82526 olduğu ve SVM* algoritmasının aldığı tespit edilmiştir.



Şekil 81. Veri Ve Svm* Algoritmasının Tahmininin Grafiksel Gösterimi

4.3. 2019-2020 Verisi Üzerinden Tam Yıl Analizleri: Sonraki Gün Tahmini

Sonraki günün tahmini için elde edilen özellikler arttırılmıştır. Çiftçilerin üye numaralarına göre ayrımı yapılmıştır. Çiftçilerin verileri pivot tablo haline getirilmiş ve veriye eklenmiştir. Çiftçilerin aldıkları akaryakıtın miktarları üye numarası bazlı veriye eklendikten sonra, veri içerisinde toplam 844 özellik bulunmaktadır. Tahminlenmek istenen etiket kaydırılmış ve tahminlemesi yapılmıştır. 2019 eğitim, 2020 test olmak üzere veri hazırlanmıştır. Elde edilen hata matrisleri ektedir.

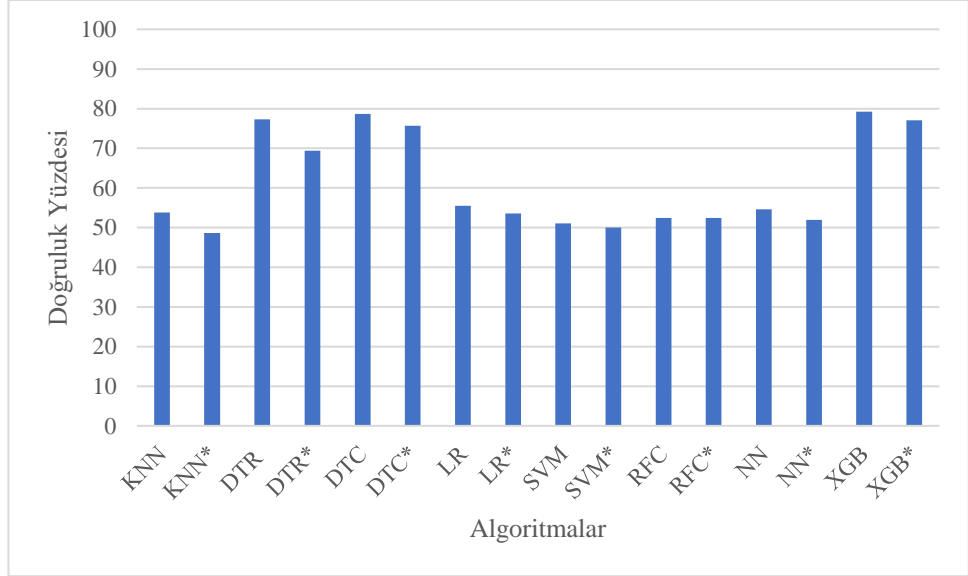
4.3.1. Sıralı Veri Tahminleme Analizi

Hazırlanan fonksiyona 2019 verileri eğitim 2020 verileri test olarak gönderilmiştir. Fonksiyon, 2019 eğitim verisini analiz edip 2020 1 Ocak tarihini tahmin etmiştir. Edilen tahmin tahminlenen dizisine aktarılmıştır. 2020 1 Ocak verileri, 2019 eğitim veri setine aktarılmış ve 2020 2 Ocak tarihi tahmin edilmiştir. Döngü test verisindeki son

tahminleme yapılana kadar devam etmiştir. Fonksiyon içerisindeki tahminlenen dizi test ile kıyaslanmış ve sonuçlar elde edilmiştir. Hata matrisleri ektedir.

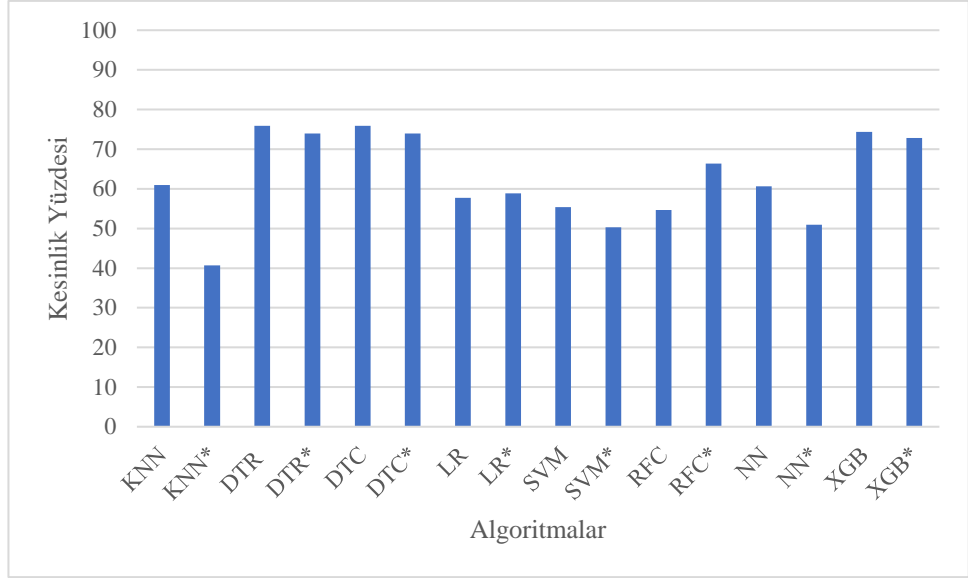
4.3.1.1. Varsayılan algoritma girdi parametreleri ile alınan sonuçlar

Sıralı veri tahminleme metodu ile yapılan analizde en yüksek doğruluk skorunu %79,235 ile XGB almıştır.



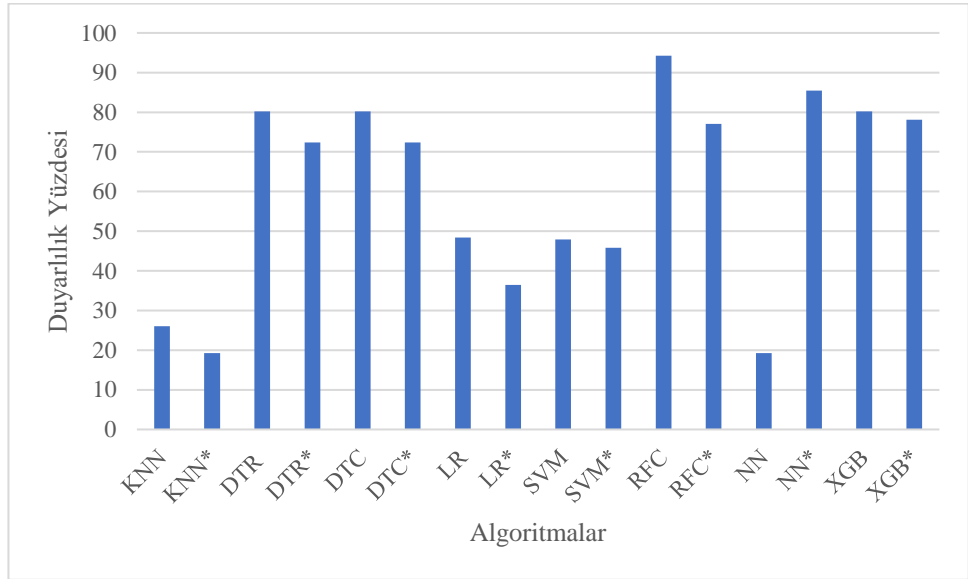
Şekil 82. Algoritmaların sıralı veri tahminleme metodu ile doğruluk skor kıyaslanması

Algoritmaların sıralı veri tahminleme metodu ile kıyaslanmasında kesinlik skorları için en yüksek skoru alanlar, %75,862 ile DRT ve DTC olmuştur.



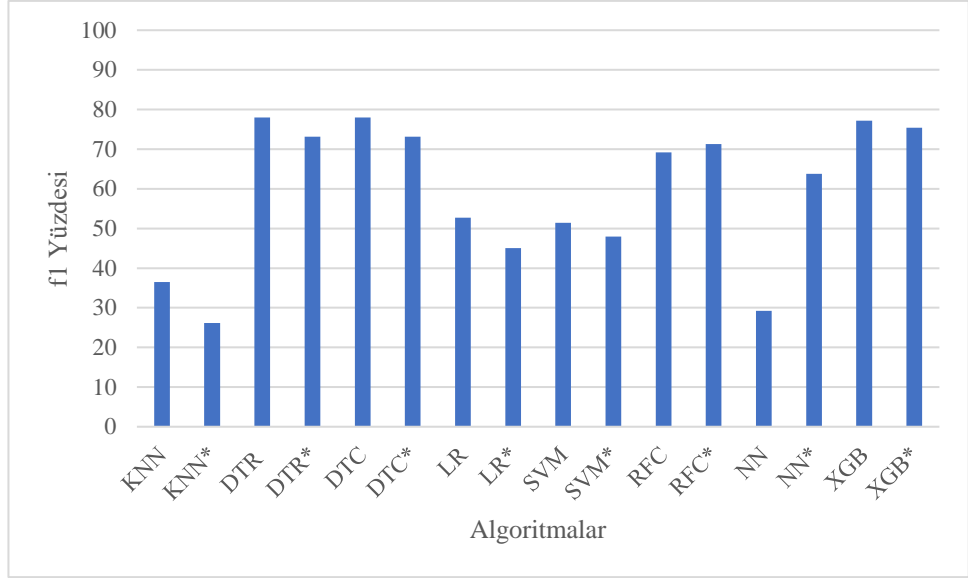
Şekil 83. Algoritmaların sıralı veri tahminleme metodu ile kesinlik skor kıyaslanması

Kullanılan sıralı veri tahminleme metodu sonucunda algoritmalar arasındaki en iyi duyarlılık skoru %85,417 ile NN* almıştır.



Şekil 84. Algoritmaların sıralı veri tahminleme metodu ile duyarlılık skor kıyaslanması

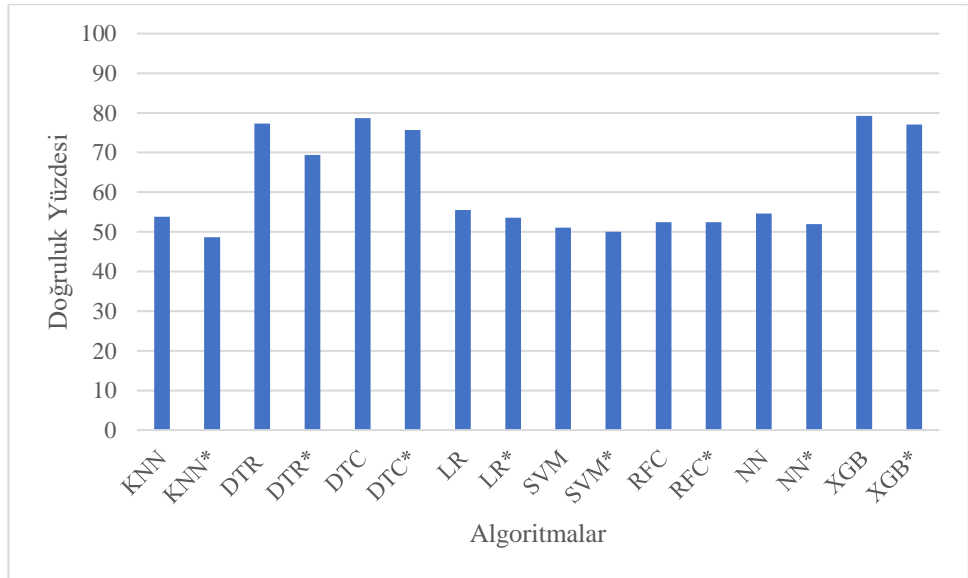
Algoritmalarından DTR, %77,975 skor ile en iyi f1 skoru yapmıştır.



Şekil 85. Algoritmaların sıralı veri tahminleme metodu ile f1 skor kıyaslanması

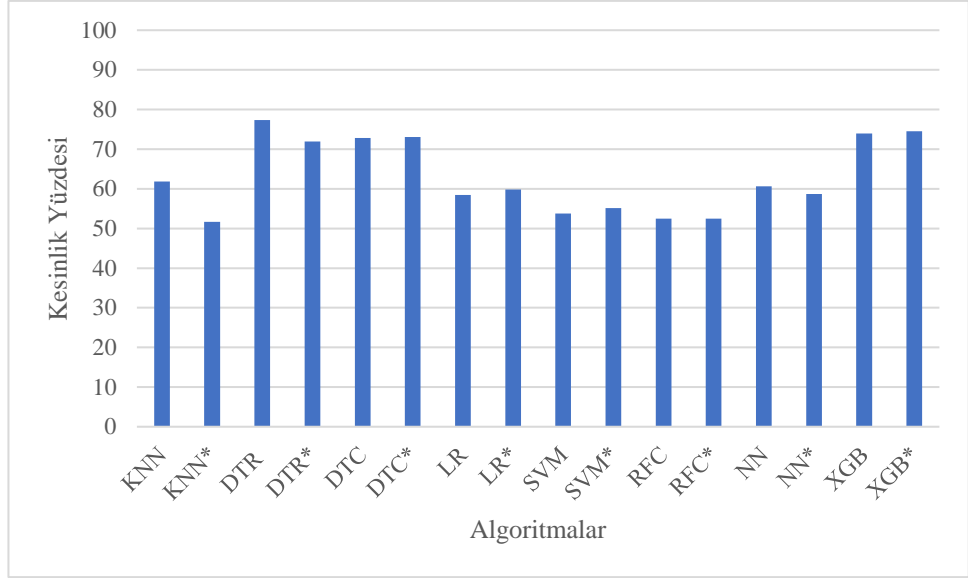
4.3.1.2. Hiper parametre optimizasyonu yapıldıktan sonra alınan sonuçlar

Yapılan hiper parametre optimizasyonu parametreleri sıralı veri tahminleme metodu ile uygulanmıştır. Bu analize göre en yüksek başarıyı %79,235 ile XGB algoritması sağlamıştır.



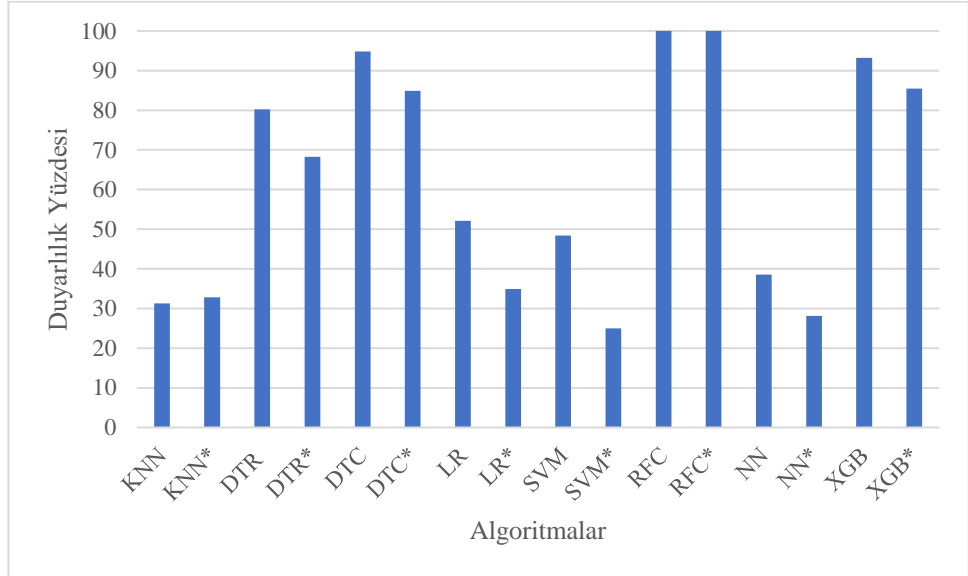
Şekil 86. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu doğruluk skor kıyaslanması

Algoritmaların hiper parametre optimizasyonu ile sıralı veri tahminleme metodundaki en yüksek kesinlik skoru %77,387 ile DTR sağlamıştır.



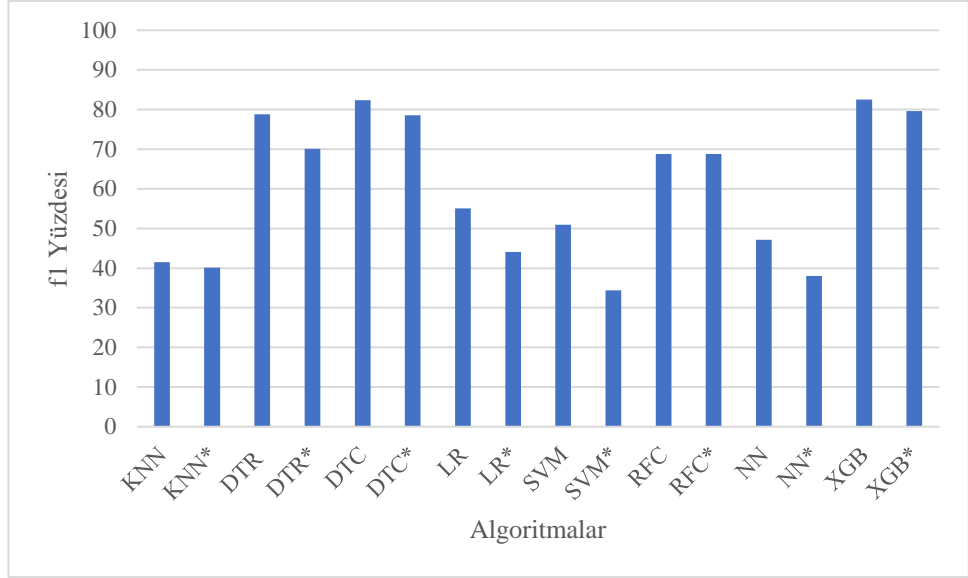
Şekil 87. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu kesinlik skor kıyaslanması

Algoritmaların duyarlılık skorlarının kıyaslanmasında en yüksek skorları %100 ile RCF ve RCF* sağlamıştır.



Şekil 88. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu duyarlılık skor kıyaslanması

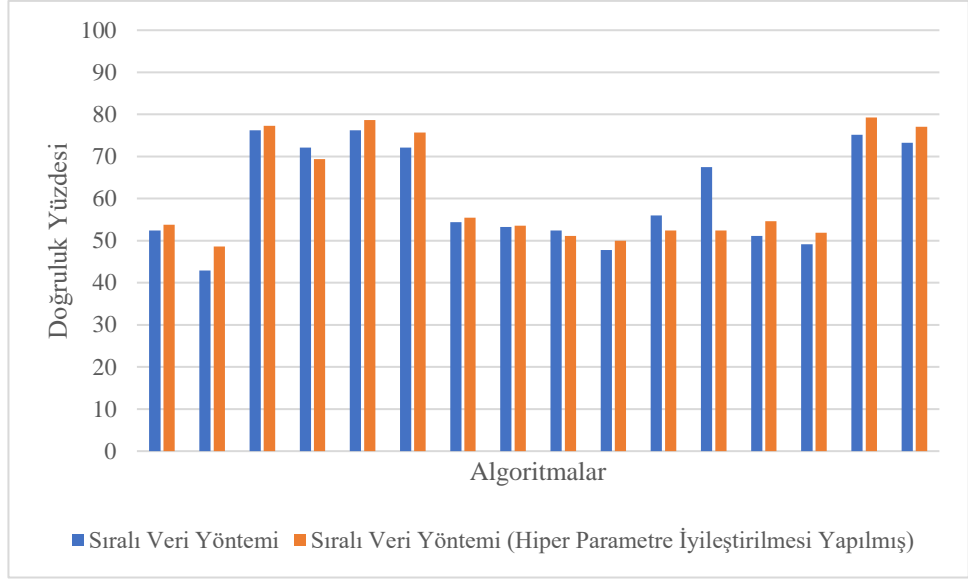
XGB %82,488 skor ile en yüksek f1 skoru almıştır.



Şekil 89. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu f1 skor kıyaslanması

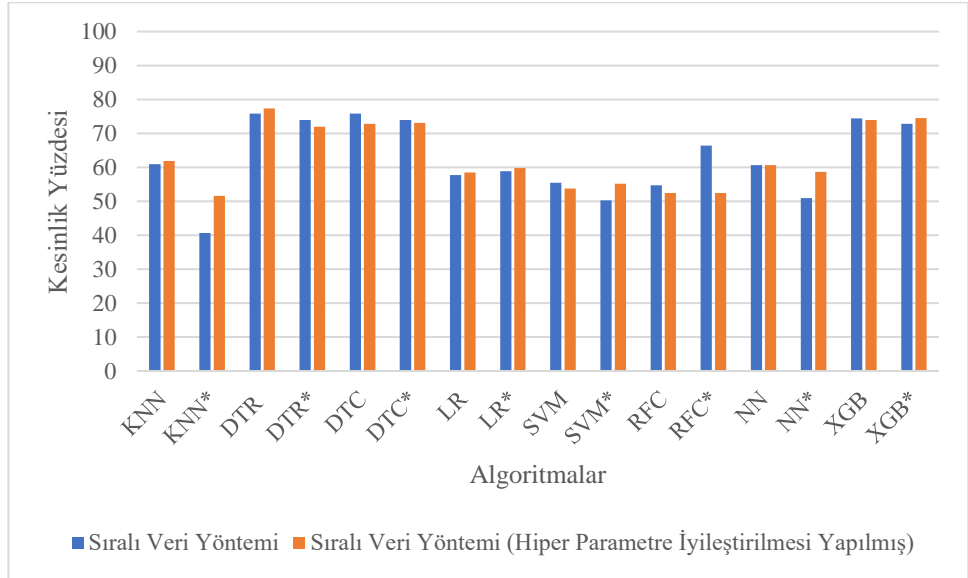
4.3.1.3. Karşılaştırmalı analiz

Algoritmaların sıralı veri tahminleme metodu ile kullanımında, hiper parametre optimizasyonu yapılmış ve yapılmamış modeller kıyaslanmıştır. Modellerin kıyaslanması sonucunda hiper parametre optimizasyonu yapılmış modellerde doğruluk yüzdesinin arttırdığı görülmüştür. Hiper parametre optimizasyonu kullanımı ile en yüksek skoru %79,235 ile XGP almıştır. Hiper parametre optimizasyonu kullanılmadığı modelde ise %76,23 ile en yüksek yüzdeliği DTR almıştır. Kıyaslanma sonucunda, DTR*, SVM, RFC ve RFC* algoritmaları hiper parametre optimizasyonu ile yüzdelik skorları düştüğü tespit edilmiştir.



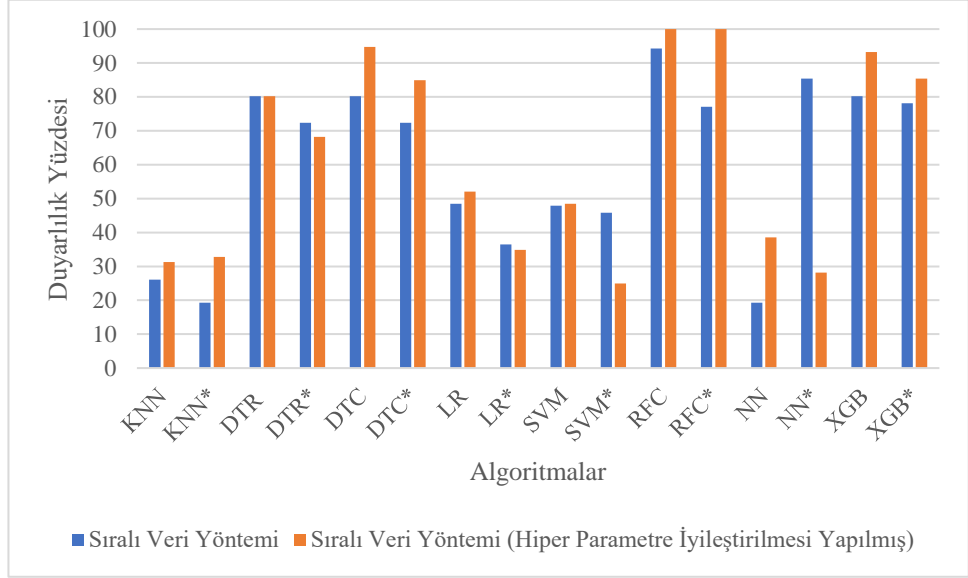
Şekil 90. Algoritmaların sıralı veri tahminleme metodu ile kullanımında doğruluk skoru kıyaslanması

Kesinlik skorları kıyaslanmasında, hiper parametre optimizasyonu yapılarak en yüksek skorun %75,862 ile DTR ve DTC aldığı görülmüştür. Kesinlik skoru için hiper parametre optimizasyonu daha başarılı sonuç elde ettiği görülmüştür. Hiper parametre optimizasyonu kullanılması sonucunda, DTC, DTC*, SVM, RFC ve XGB algoritmalarının kesinlik skorlarının düştüğü görülmüştür.



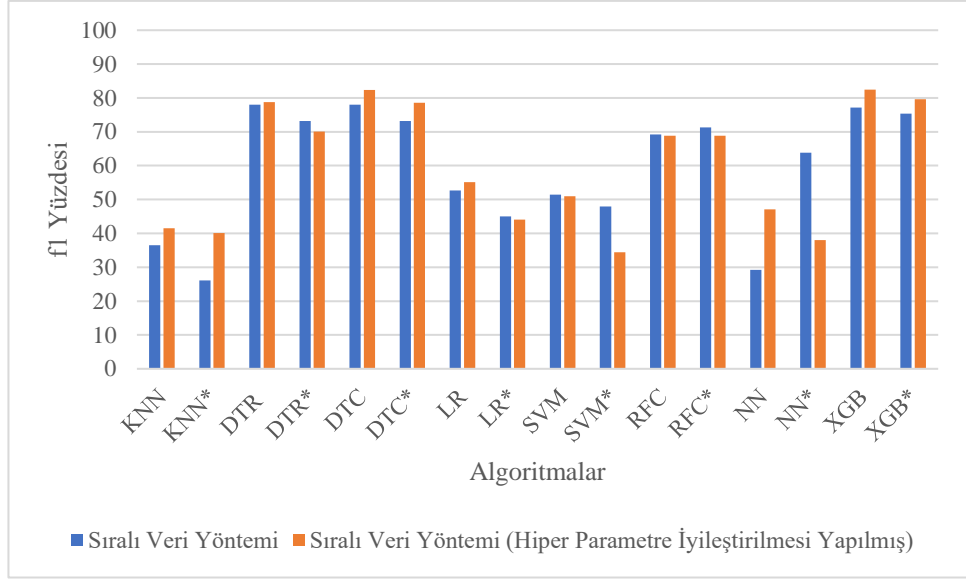
Şekil 91. Algoritmaların sıralı veri tahminleme metodu ile kullanımında kesinlik skoru kıyaslanması

Duyarlılık skoru hiper parametre optimizasyonu kullanımı ile artmıştır. %100 skor ile RFC ve RFC* algoritmaları en yüksek başarıyı sağlamıştır. Bunun yanı sıra, hiper parametre optimizasyonu kullanımı ile, KNN, DTR*, LR*, SVM* ve NN*, algoritmalarının kesinlik yüzdeleri düşmüştür.



Şekil 92. Algoritmaların sıralı veri tahminleme metodu ile kullanımında duyarlılık skoru kıyaslanması

F1 skorları arasından en yüksek başarı, %82,488 ile hiper parametre optimizasyonu kullanılmış XGB algoritması olmuştur. DTR*, LR*, SVM*, RFC, RFC* ve NN* algoritmaları hiper parametre optimizasyonu kullanılmış durumda yüzdeleri düşmüştür.



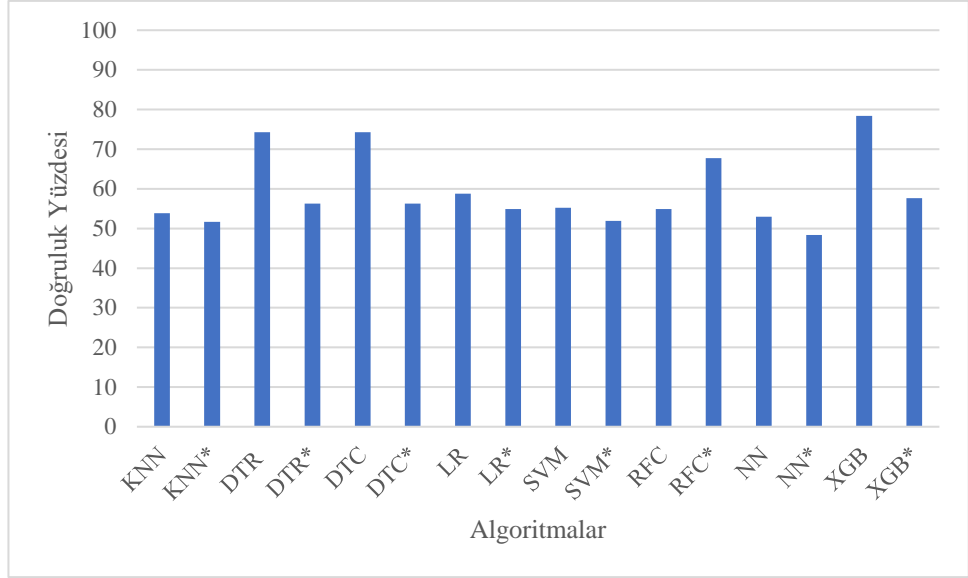
Şekil 93. Algoritmaların sıralı veri tahminleme metodu ile kullanımında f1 skoru kıyaslanması

4.3.2. Yüzdelik Ayrım Analizi

2019 eğitim ve 2020 test olmak üzere %50 ayrılmıştır. Bulunan hata matrisleri ektedir.

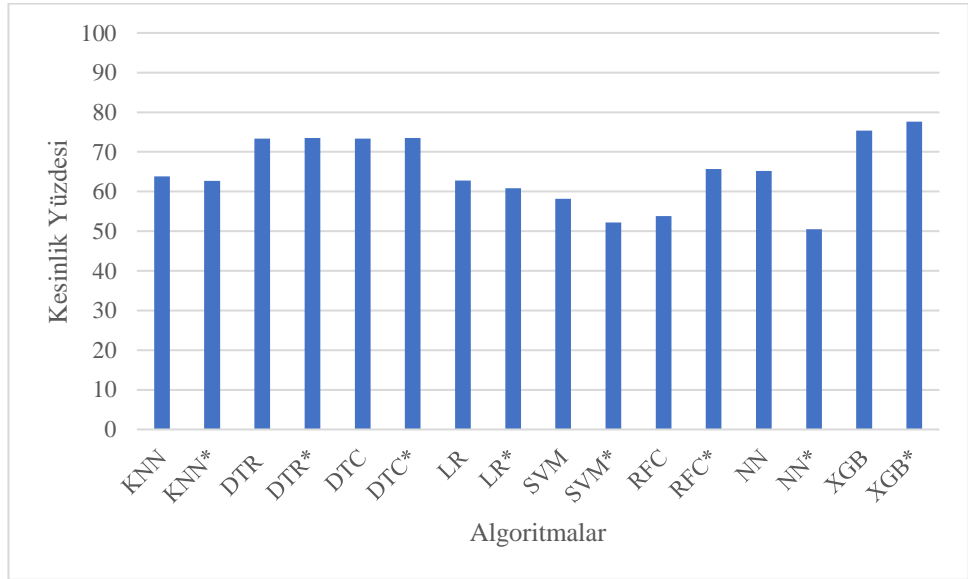
4.3.2.1. Varsayılan algoritma girdi parametreleri ile alınan sonuçlar

Algoritmaların sklearn kütüphanesinde bulunan varsayılan hali ile yapılan analizin doğruluk skorları tespit edilmiştir. Bu analiz sonucunda en iyi skor XGB ile %78,415 olmuştur.



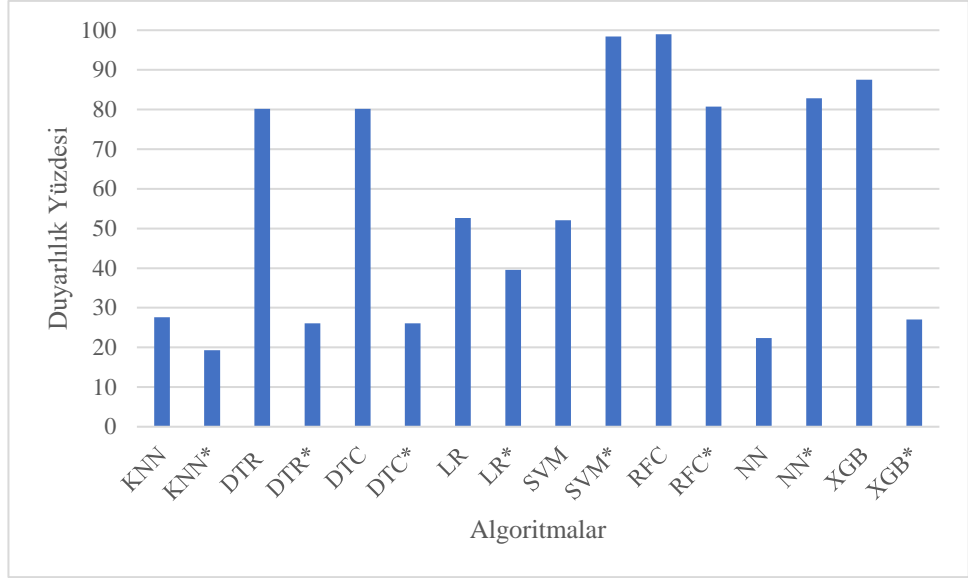
Şekil 94. Algoritmaların yüzdelerik ayırım yöntemi kullanımı ile doğruluk skor kıyaslanması

Yüzdelerik ayırım yöntemi ile algoritmaların kesinlik skor kıyaslanması yapılmıştır. XGB* algoritması %77,778 skor ile en yüksek başarıyı sağlamıştır.



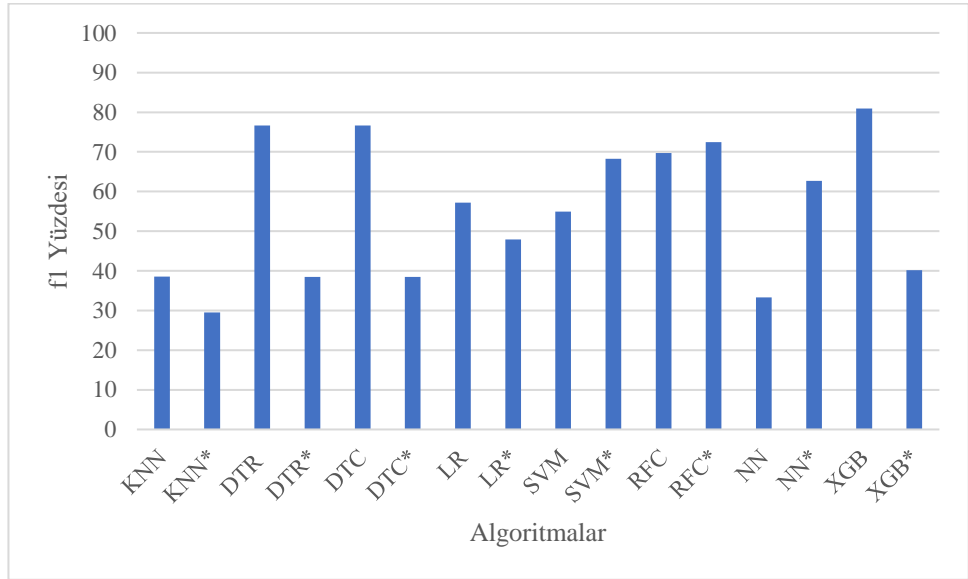
Şekil 95. Algoritmaların yüzdelerik ayırım yöntemi kullanımı ile kesinlik skor kıyaslanması

Duyarlılık skoru kıyaslamasında. SVM* %100 başarı ile en yüksek sonucu vermiştir.



Şekil 96. Algoritmaların yüzdelerik ayırım yöntemi kullanımı ile duyarlılık skor kıyaslanması

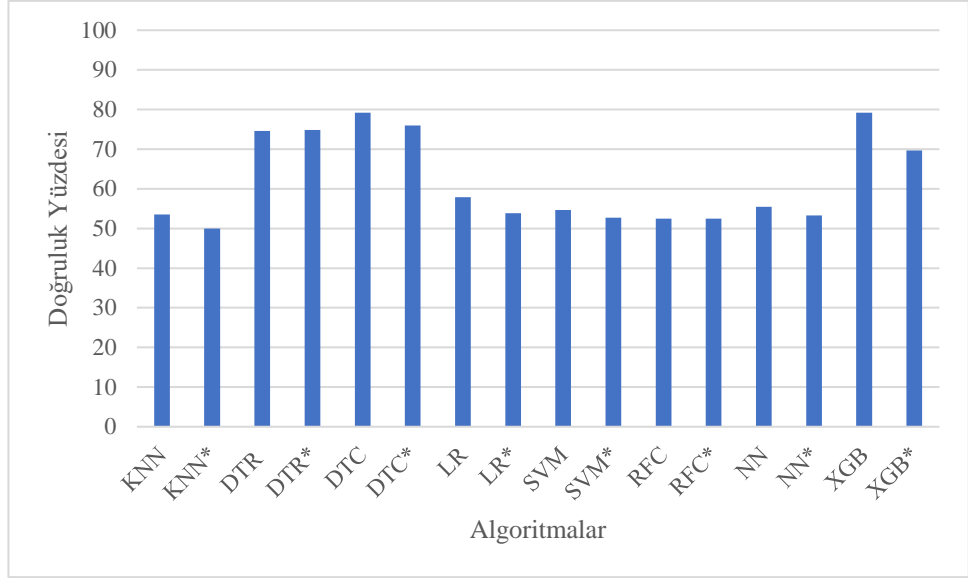
Algoritmaların f1 skor kıyaslanmasında XGB %80,964 alarak en yüksek skoru elde etmiştir.



Şekil 97. Algoritmaların yüzdelerik ayırım yöntemi kullanımı ile f1 skor kıyaslanması

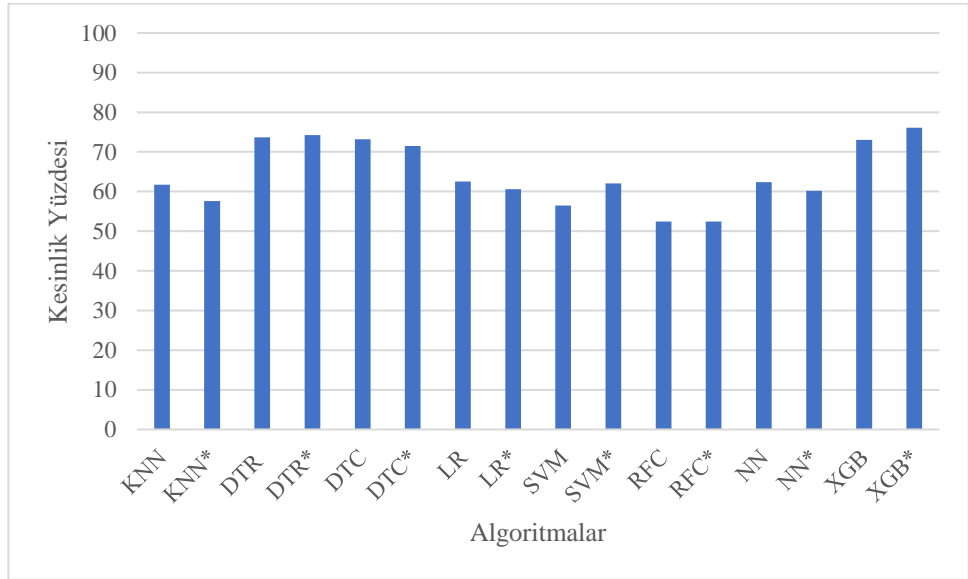
4.3.2.2. Hiper parametre optimizasyonu yapıldıktan sonra alınan sonuçlar

Hiper parametre optimizasyonu ile tespit edilen parametreler kullanılarak algoritmaların skorları kıyaslanmıştır. %79,235 ile en yüksek doğruluk skorunu XGB almıştır.



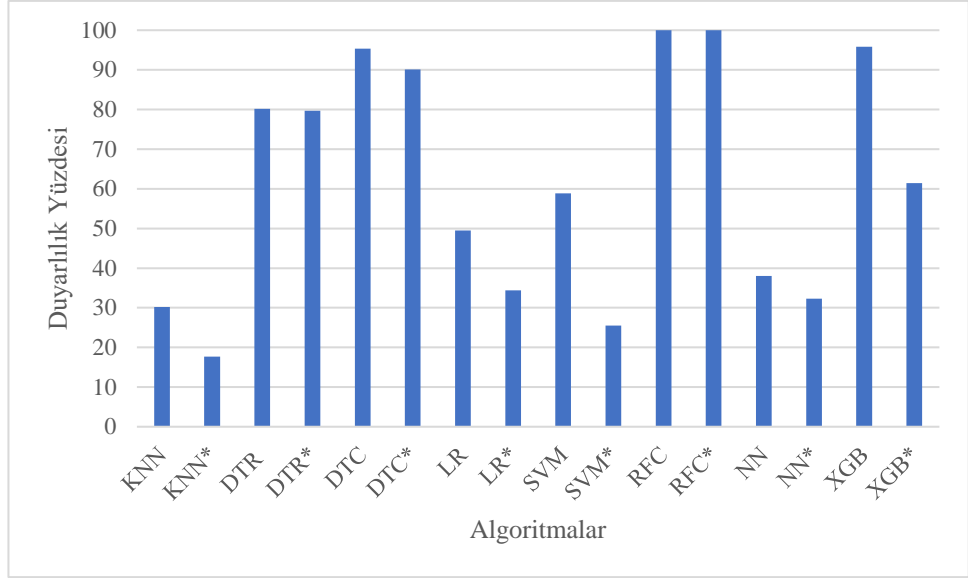
Şekil 98. Algoritmaların hiper parametre tespitleriyle yüzdelerik ayırım yöntemi kullanımı, doğruluk skor kıyaslanması

Hiper parametre optimizasyonu ile kesinlik skor kıyaslanmasında %76,129 skor ile en yüksek skoru XGB* almıştır.



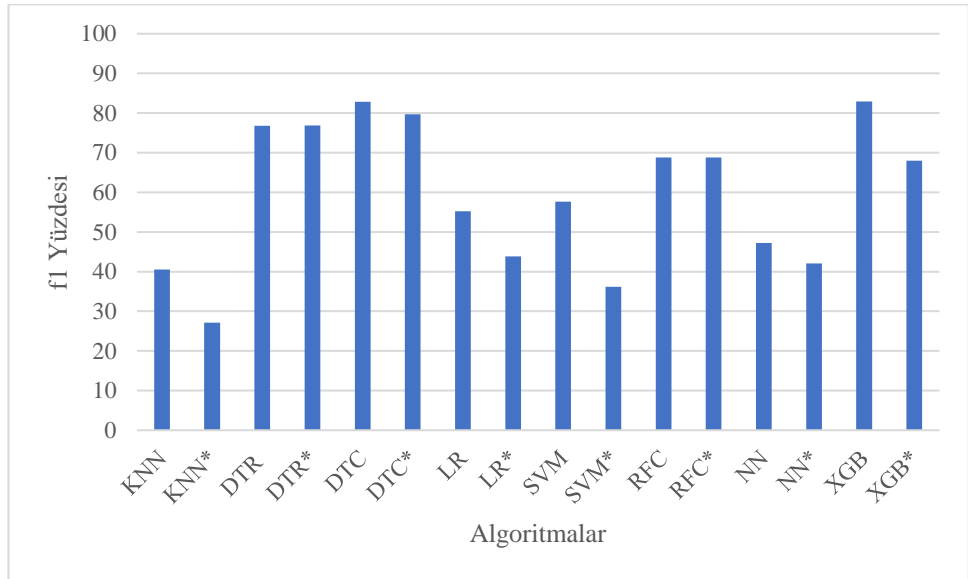
Şekil 99. Algoritmaların hiper parametre tespitleriyle yüzdelerik ayırım yöntemi kullanımı, kesinlik skor kıyaslanması

Algoritmaların hiper parametreler optimizasyonu ile analizlerinde en yüksek duyarlılık skoru %100 ile RFC ve RFC* almıştır.



Şekil 100. Algoritmaların hiper parametre tespitleriyle yüzdelik ayırım yöntemi kullanımı, duyarlılık skor kıyaslanması

Algoritmalar arasında en yüksek f1 skoru %82,883 ile XGB almıştır.

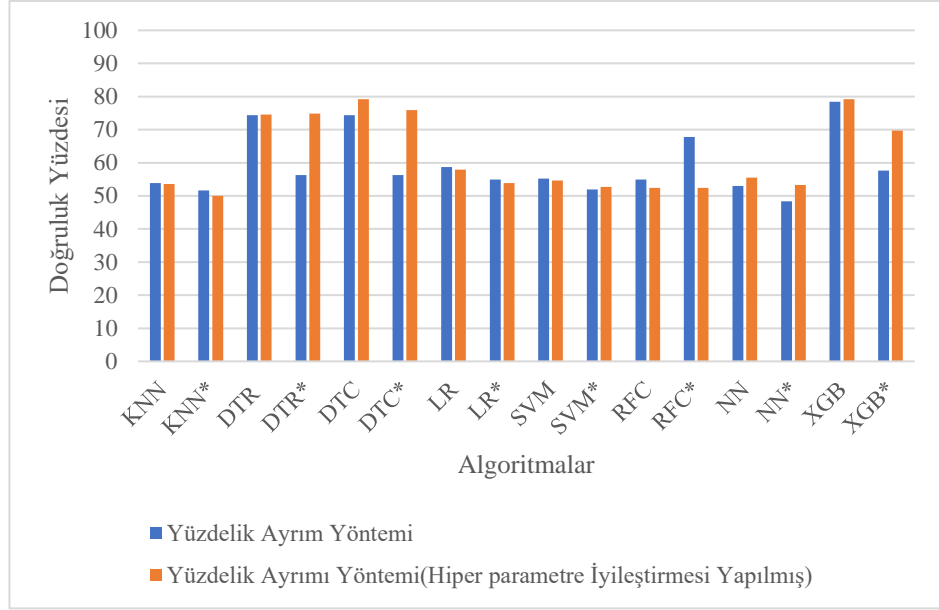


Şekil 101. Algoritmaların hiper parametre tespitleriyle yüzdelik ayırım yöntemi kullanımı, f1 skor kıyaslanması

4.3.2.3. Karşılaştırmalı analiz

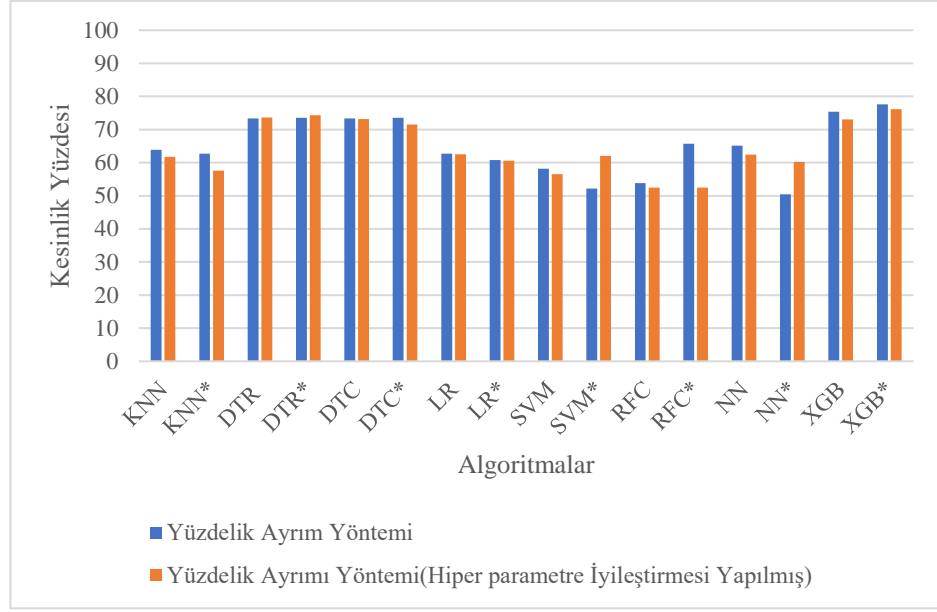
Algoritmaların yüzdelik ayırım yöntemi ile kullanımında, hiper parametre optimizasyonu yapılmış ve yapılmamış modeller kıyaslanmıştır. Modellerin kıyaslanması sonucunda hiper parametre optimizasyonu yapılan modellerin doğruluk

yüzdesini arttırdığı görülmüştür. Hiper parametre optimizasyonu kullanımı ile en yüksek skoru %79,235 ile XGP ve DTC almıştır. Hiper parametre optimizasyonu kullanılmadığı modelde ise %78,415 ile en yüksek yüzdeliği XGB almıştır. Kıyaslanma sonucunda, KNN, KNN*, LR, LR*, SVM, SVM*, RFC ve RFC* algoritmaları hiper parametre optimizasyonu ile yüzdelik skorları düştüğü tespit edilmiştir.



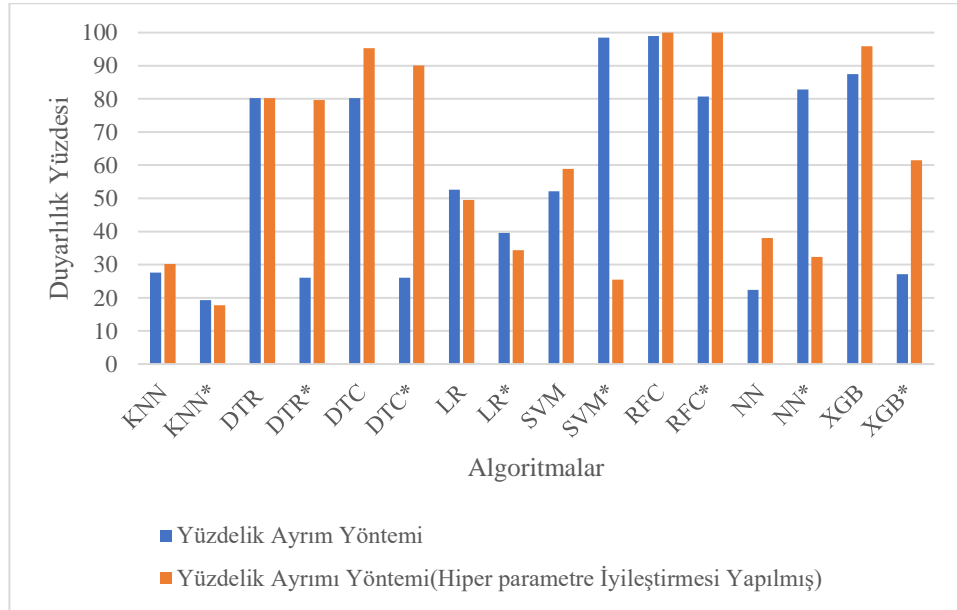
Şekil 102. Algoritmaların yüzdelik ayırımı yöntemi ile kullanımında doğruluk skoru kıyaslanması

Kesinlik skorları kıyaslanmasında, hiper parametre optimizasyonu yapılmadan en yüksek skoru %77,778 ile XGB* aldığı görülmüştür. Kesinlik skoru için hiper parametre optimizasyonu yapılmamasının daha başarılı sonuç elde ettiği görülmüştür. Hiper parametre optimizasyonu kullanılması sonucunda, KNN, KNN*, DTC, DTC*, LR*, LR, NN, XGB ve XGB* algoritmalarının kesinlik skorlarının düştüğü görülmüştür.



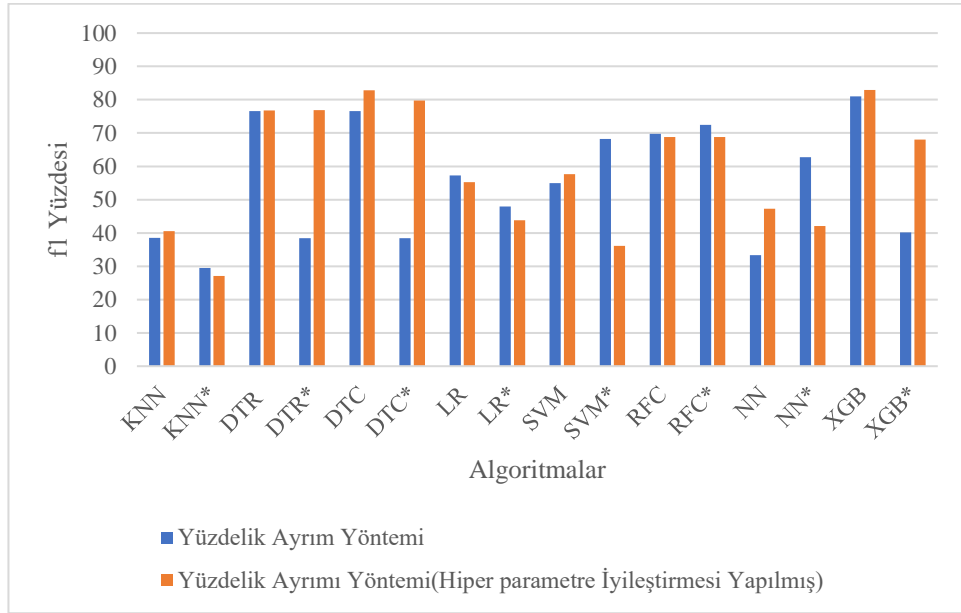
Şekil 103. Algoritmaların yüzdelik ayırımı yöntemi ile kullanımında kesinlik skoru kıyaslanması

Duyarlılık skoru hiper parametre optimizasyonu kullanımı ile artmıştır. %100 skor ile RFC ve RFC* algoritmaları en yüksek başarıyı sağlamıştır. Bunun yanı sıra, hiper parametre optimizasyonu kullanımı ile, KNN, KNN*, LR, LR*, SVM, SVM* ve NN* algoritmalarının kesinlik yüzdelikleri düşmüştür.



Şekil 104. Algoritmaların yüzdelik ayırımı yöntemi ile kullanımında duyarlılık skoru kıyaslanması

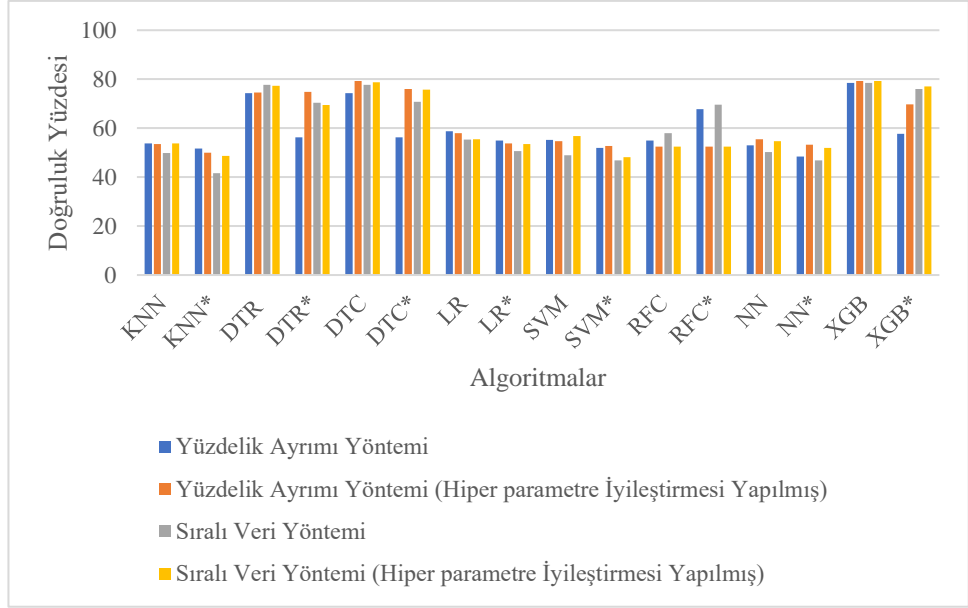
F1 skorları arasından en yüksek başarı, %82,883 ile hiper parametre optimizasyonu kullanılmış XGB algoritması olmuştur. KNN*, DTR*, LR, LR*, SVM, SVM*, RFC, RFC* ve NN* algoritmaları hiper parametre optimizasyonu kullanılmış durumda yüzdeleri düşmüştür.



Şekil 105. Algoritmaların yüzdelik ayırımı yöntemi ile kullanımında f1 skoru kıyaslanması

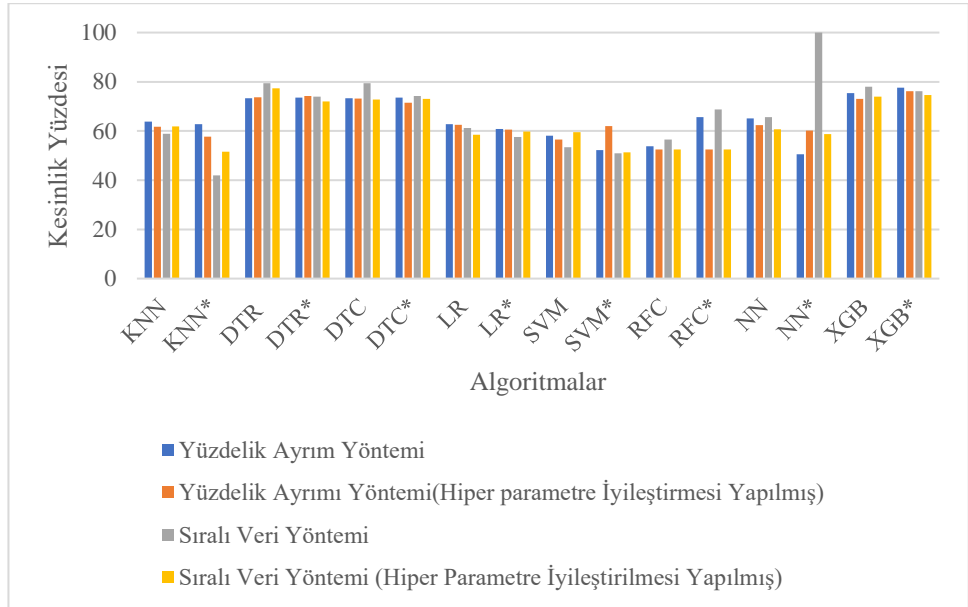
4.3.3. Karşılaştırmalı Algoritma Skorları

Yüzdelik ayırım yöntemi ve sıralı veri tahminleme metodu kıyaslanmıştır. Yöntemlerin hiper parametre optimizasyonu kullanıldığı ve kullanılmadığı durumlar göz önünde bulundurulmuştur. Buna göre; doğruluk skoru için en başarılı algoritmaların XGB ve DTC olduğu gözlenmiştir. XGB yüzdelik ayırım yöntemi ve sıralı veri tahminleme metodunda (hiper parametre optimizasyonu yapılmış) %79,235 başarı yüzdesine ulaşmıştır. DTC algoritması %79,235 başarı yüzdesine, yüzdelik ayırım yöntemi (hiper parametre optimizasyonu yapılmış) ile ulaşmıştır.



Şekil 106. Doğruluk skorlarının kıyaslanması

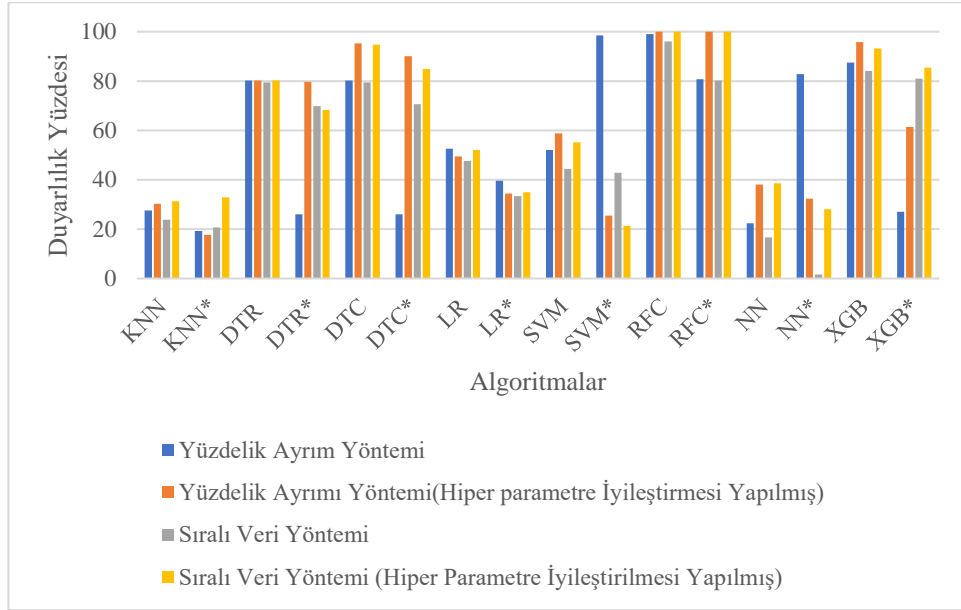
Kesinlik skor kıyaslamasında en yüksek skoru, %77,778 ile XGB* algoritması yüzdelik ayırım yöntemi ile almıştır. Bu algoritmayı takip eden skorlar, %77,387 ile DTR sıralı veri tahminleme metodu (hiper parametre optimizasyonu yapılmış), %76,129 ile XGB* yüzdelik ayırım yöntemi (hiper parametre optimizasyonu yapılmış) takip etmiştir.



Şekil 107. Kesinlik skorlarının kıyaslanması

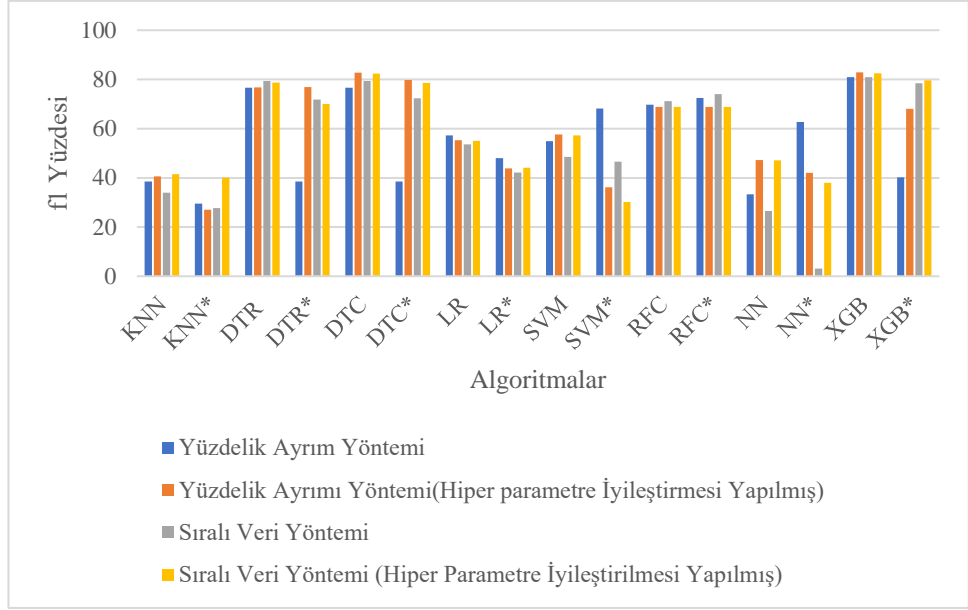
Duyarlılık skor kıyaslamasında en yüksek skorları, %100 ile SVM* algoritması yüzdelik ayırım yöntemi ile, RFC ve RFC* algoritmaları yüzdelik ayırım ve sıralı veri tahminleme metodu (hiper parametre optimizasyonu yapılmış) ile almıştır. Bu

algoritmayı takip eden skorlar, %85,417 ile NN* sıralı veri tahminleme metodu olmuştur.



Şekil 108. Duyarlılık skorlarının kıyaslanması

F1 skor kıyaslanmasında en yüksek skoru yüzdelik ayırma yöntemi (hiper parametre optimizasyonu yapılmış) ile %82,883 olarak XGB olmuştur. Bu skoru takip eden algoritmalar; sıralı veri tahminleme metodu (hiper parametre optimizasyonu yapılmış) kullanılarak %82,488 skoru alan XGB, yüzdelik ayırma yöntemi kullanılarak %80,964 skor sağlayan XGB ve sıralı veri tahminleme metodu ile %77,975 alan DTR olmuştur.



Şekil 109. F1 skorlarının kıyaslanması

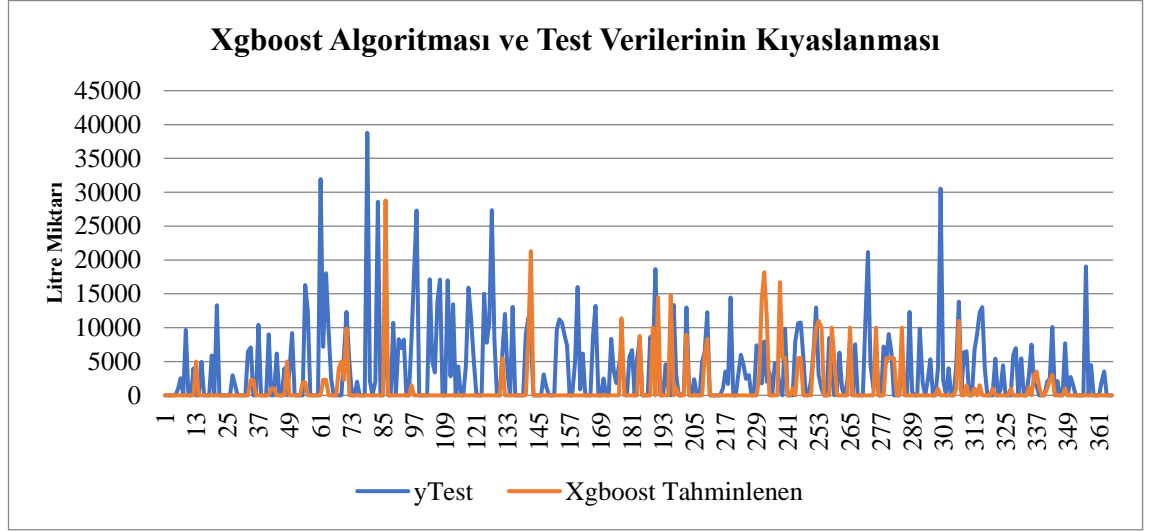
4.3.4. Akaryakıt Miktarı Tahminleme Skorları

2019 eğitim 2020 test edilen veri üzerinden sonraki gün litre tahminlemesi yapılmıştır. Yapılan tahminlemenin, algoritmalar baz alınarak MSE skorları tespit edilmiştir. MSE skorları alımı içi veri standart ölçeklendirilmiş olup sklearn kütüphanesinden faydalanılmıştır.

Tablo 20. Algoritmaların Litre Tahminleme İçin Aldıkları MSE Skorları

MSE Skorları	
Algoritma	Skor
KNN	1,43694
KNN*	1,43123
DTR	1,79385
DTR*	2,97767
DTC	1,82539
DTC*	1,98225
LR	1,63848
LR*	2,02708
SVM	7,76153
SVM*	1,58754
RFC	1,49353
RFC*	1,49353
NN	1,49353
NN*	1,49353
XGB	1,34521
XGB*	1,28921

Algoritmalarından en iyi skoru 1,28921 olduğu ve XGB* algoritmasının aldığı tespit edilmiştir.



Şekil 110. Veri Ve XGBoost Algoritmasının Tahmininin Grafikselleştirilmesi

4.4. 2019-2020 Verisi Kullanılarak 2021 İlk Çeyrek Tahmini: Sonraki Gün Tahmini

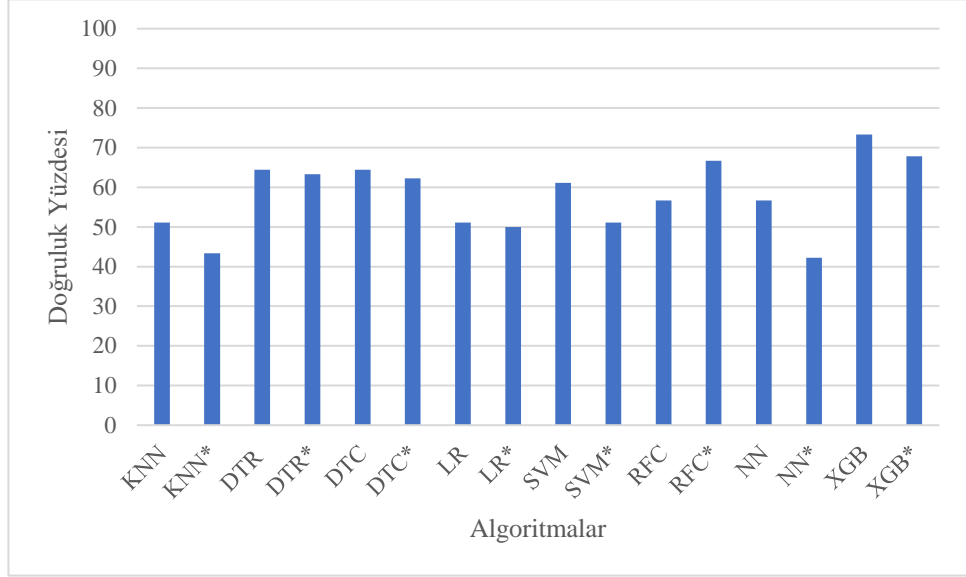
Sonraki gün tahmini için geliştirilen pivot tablo veriye eklenmiştir. Üye numarası bazlı ekleme sonucunda 844 adet özellik bulunmuştur. Bütün özellikler algoritma bazlı önemli özellik bulma yöntemlerinden geçirilmiştir. Algoritmalar için önemli veriler ayrılmıştır. Algoritma bazlı veri haline getirilmiştir. 2019 ve 2020 verileri eğitim, 2021 ilk çeyrek verisi test olarak kullanılmıştır.

4.4.1. Sıralı Veri Tahminleme Analizi

2019 ve 2020 verilerinin tahmin ettiği 2021 1 Ocak verisini tahminleme dizisine atadıktan sonra, 2021 1 Ocak verisini eğitime dahil edilmiştir. Bu döngü 90 adet test verisi bitene kadar devam etmiştir. Elde edilen test dizi, 2021 birinci çeyrek etiket verileri ile kıyaslanmıştır. Hata matrisleri ektedir.

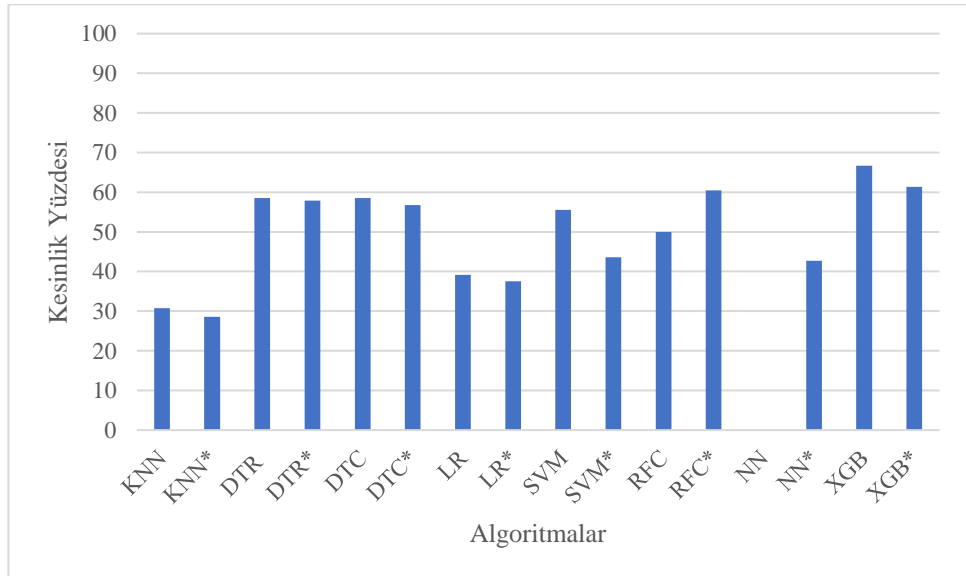
4.4.1.1. Varsayılan algoritma girdi parametreleri ile alınan sonuçlar

Algoritmalarından en yüksek doğruluk skorunu %73,333 ile XGB almıştır.



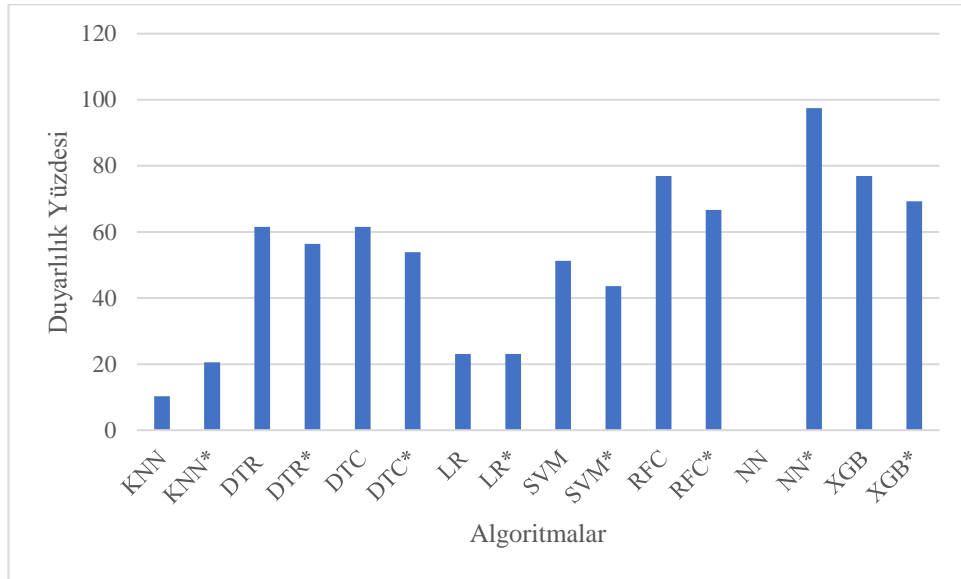
Şekil 111. Algoritmaların sıralı veri tahminleme metodu ile doğruluk skor kıyaslanması

Algoritmaların sıralı veri tahminleme metodu ile kıyaslanmasında kesinlik skorları için en yüksek skoru alanlar, %61,364 ile XGB* olmuştur.



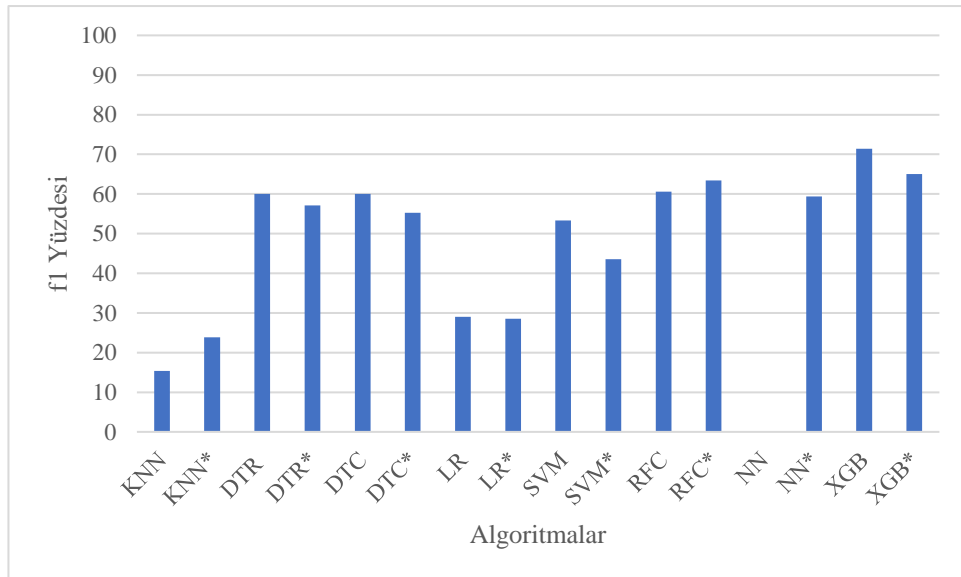
Şekil 112. Algoritmaların sıralı veri tahminleme metodu ile kesinlik skor kıyaslanması

Kullanılan sıralı veri tahminleme metodu sonucunda algoritmalar arasındaki en iyi duyarlılık skoru %97,436 ile NN* almıştır.



Şekil 113. Algoritmaların sıralı veri tahminleme metodu ile duyarlılık skor kıyaslanması

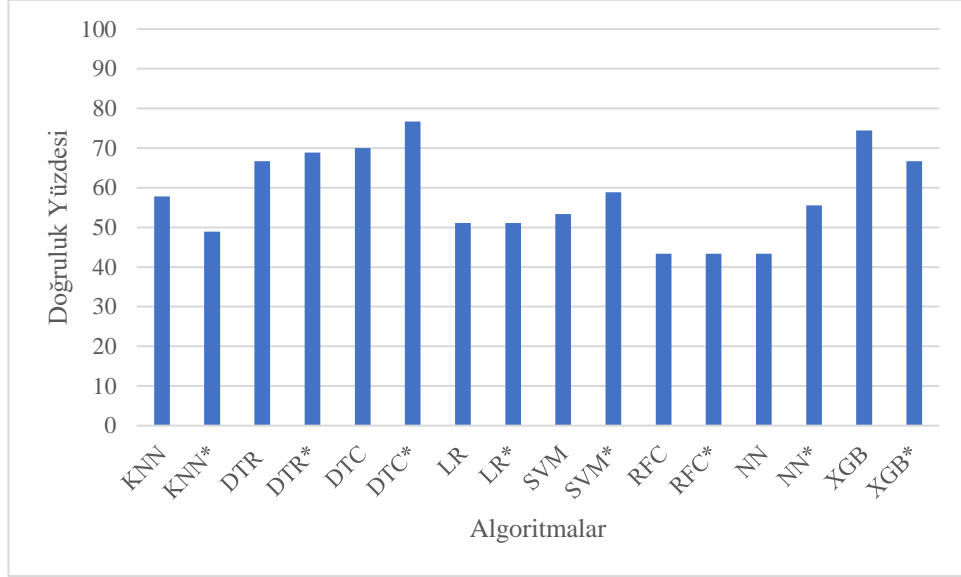
Algoritmalarından XGB, %71,429 skor ile en iyi f1 skoru yapmıştır.



Şekil 114. Algoritmaların sıralı veri tahminleme metodu ile f1 skor kıyaslanması

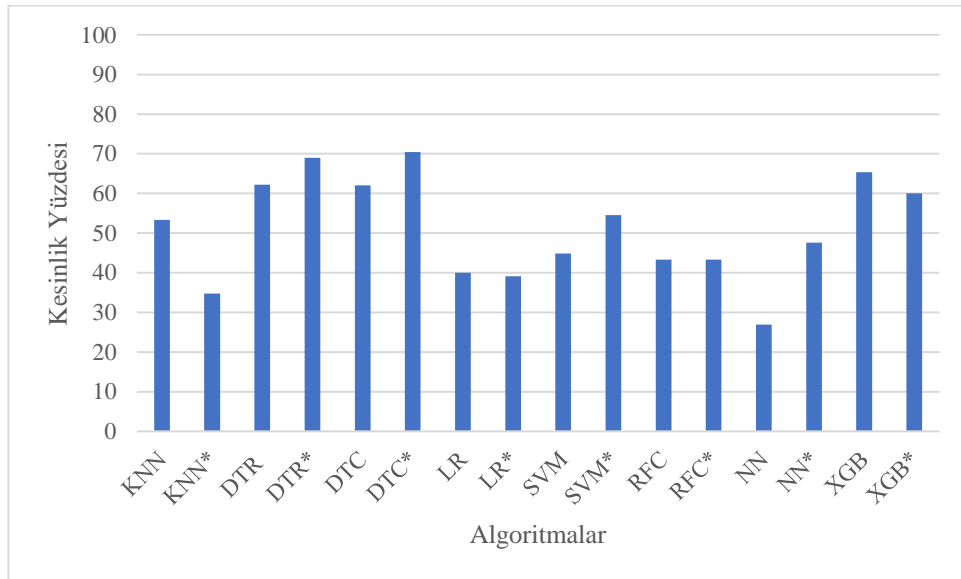
4.4.1.2. Hiper parametre optimizasyonu yapıldıktan sonra alınan sonuçlar

Hiper parametre optimizasyonu yapılarak tespit edilen parametreler sıralı veri tahminleme metoduna ile uygulanmıştır. Bu analize göre en yüksek başarıyı %76,667 ile DTC* algoritması sağlamıştır.



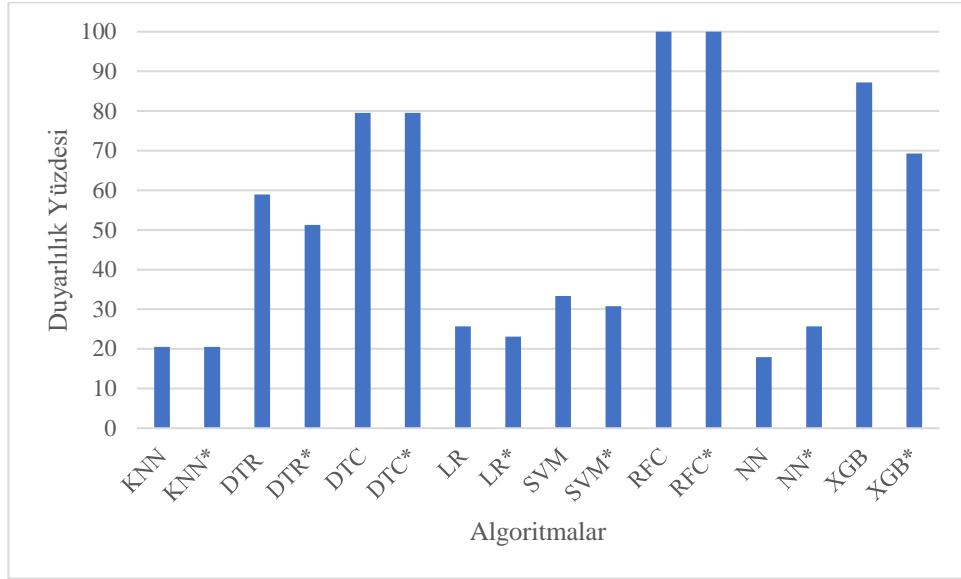
Şekil 115. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu doğruluk skor kıyaslanması

Algoritmaların hiper parametre ile sıralı veri tahminleme metodu kullanımında en yüksek kesinlik skoru %70,455 ile DTC* sağlamıştır.



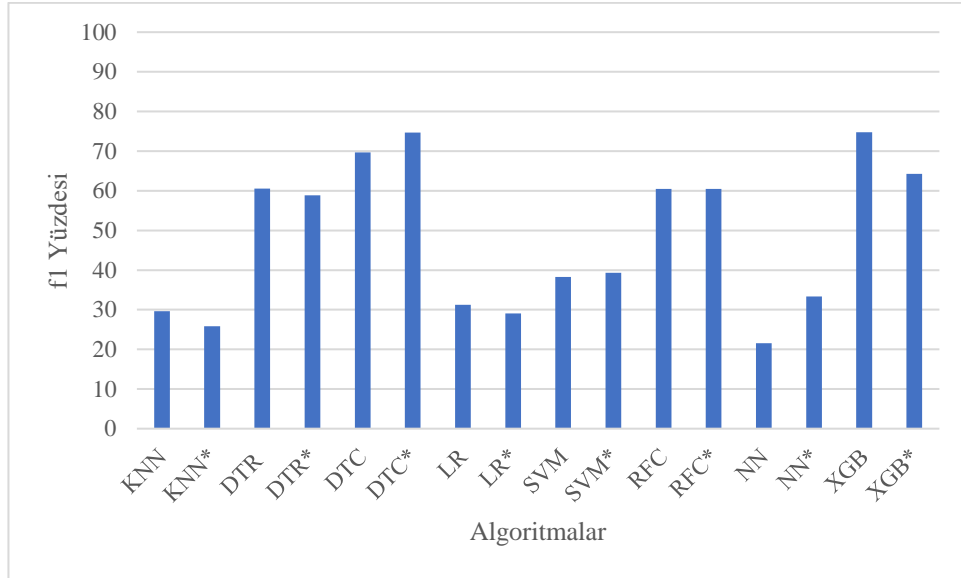
Şekil 116. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu kesinlik skor kıyaslanması

Algoritmaların duyarlılık skorlarının kıyaslanmasında en yüksek skorları %100 ile RCF ve RCF* sağlamıştır.



Şekil 117. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu duyarlılık skor kıyaslanması

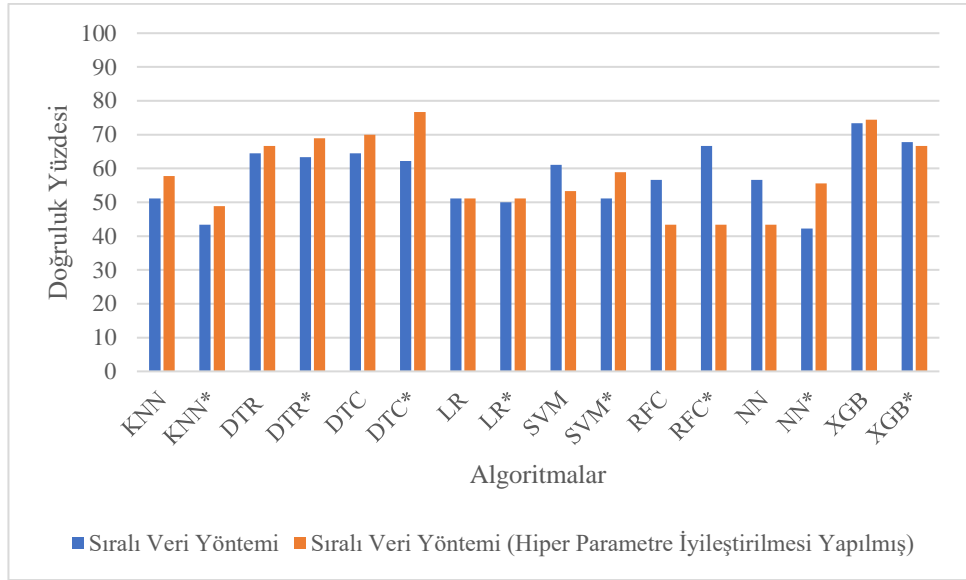
DTC* %74,699 skor ile en yüksek f1 skoru almıştır.



Şekil 118. Algoritmaların hiper parametre optimizasyonu uygulanarak, sıralı veri tahminleme metodu f1 skor kıyaslanması

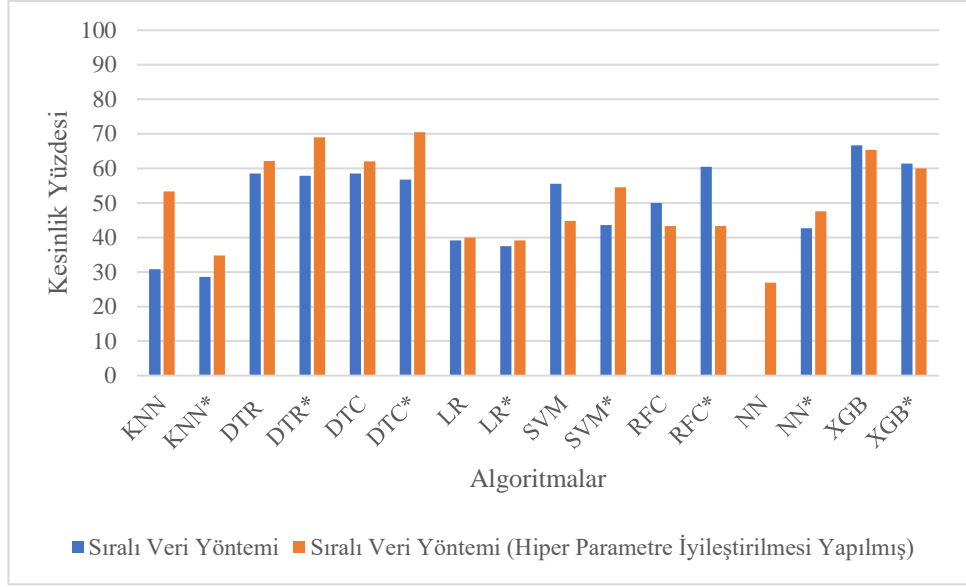
4.4.1.3. Karşılaştırmalı analiz

Algoritmaların sıralı veri tahminleme metodu ile kullanımında, hiper parametre optimizasyonu yapılmış ve yapılmamış modeller kıyaslanmıştır. Modellerin kıyaslanması sonucunda hiper parametre optimizasyonu yapılmış modellerde doğruluk yüzdesinin arttırdığı görülmüştür. Hiper parametre optimizasyonu kullanımı ile en yüksek skoru %76,667 ile DTC almıştır. Hiper parametre optimizasyonu kullanılmadığı modelde ise %73,333 ile en yüksek yüzdeliği XGB almıştır. Kıyaslanma sonucunda, SVM, RFC, RFC*, NN ve XGB* algoritmaları hiper parametre optimizasyonu ile yüzdelik skorları düştüğü tespit edilmiştir.



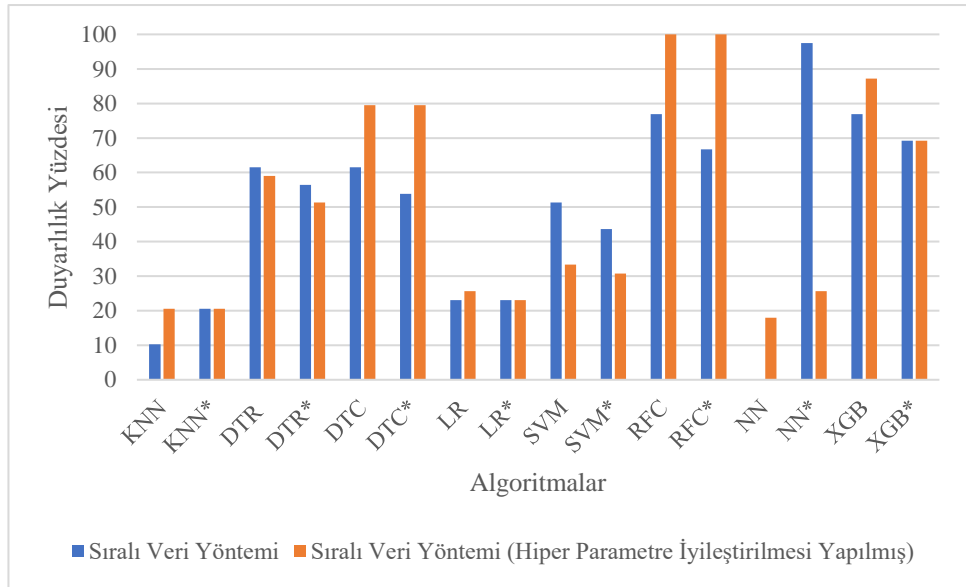
Şekil 119. Algoritmaların sıralı veri tahminleme metodu ile kullanımında doğruluk skoru kıyaslanması

Kesinlik skorları kıyaslanmasında, hiper parametre optimizasyonu yapılarak en yüksek skorun %70,455 ile DTC* aldığı görülmüştür. Kesinlik skoru için hiper parametre optimizasyonu ile daha başarılı sonuç elde ettiği görülmüştür. Hiper parametre optimizasyonu kullanılması sonucunda, SVM, RFC, RFC*, XGB ve XGB* algoritmalarının kesinlik skorlarının düştüğü görülmüştür.



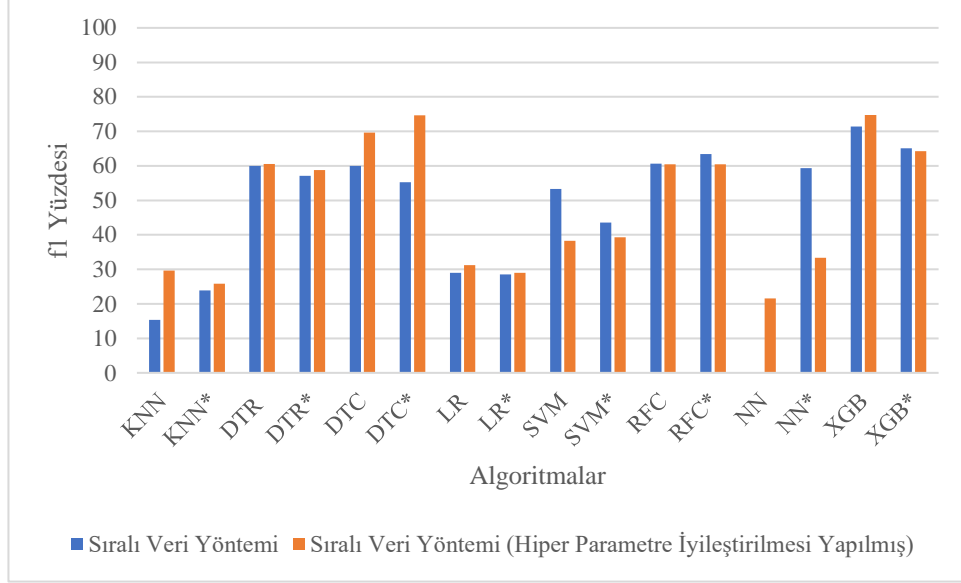
Şekil 120. Algoritmaların sıralı veri tahminleme metodu ile kullanımında kesinlik skoru kıyaslanması

Duyarlılık skoru hiper parametre optimizasyonu kullanımı ile artmıştır. %100 skor ile RFC ve RFC* algoritmaları en yüksek başarıyı sağlamıştır. Bunun yanı sıra, hiper parametre optimizasyonu ile, KNN*, DTR, DTR*, LR*, SVM, SVM* ve NN*, algoritmalarının kesinlik yüzdeleri düşmüştür.



Şekil 121. Algoritmaların sıralı veri tahminleme metodu ile kullanımında duyarlılık skoru kıyaslanması

F1 skorları arasından en yüksek başarı, %74,699 ile hiper parametre optimizasyonu kullanılmış DTC* algoritması olmuştur. SVM, SVM*, RFC, RFC*, NN* ve XGB* algoritmaları hiper parametre optimizasyonu yapılmış durumda yüzdelikleri düşmüştür.



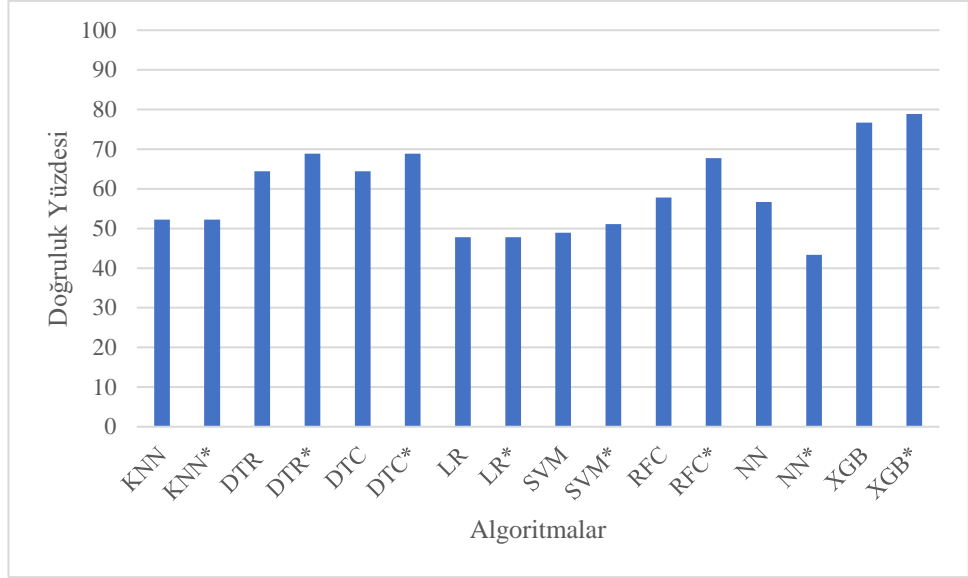
Şekil 122. Algoritmaların sıralı veri tahminleme metodu ile kullanımında f1 skoru kıyaslanması

4.4.2. Yüzdelik Ayrım Analizi

Yüzdelik ayırım yöntemi ile 2019 ve 2020 verileri eğitim, 2021 birinci çeyrek verisi test olarak ayrılmıştır. %89 eğitim ve %11 test verisi elde edilmiştir. Yapılan analizler raporlanmıştır. Hata matrisleri ektedir.

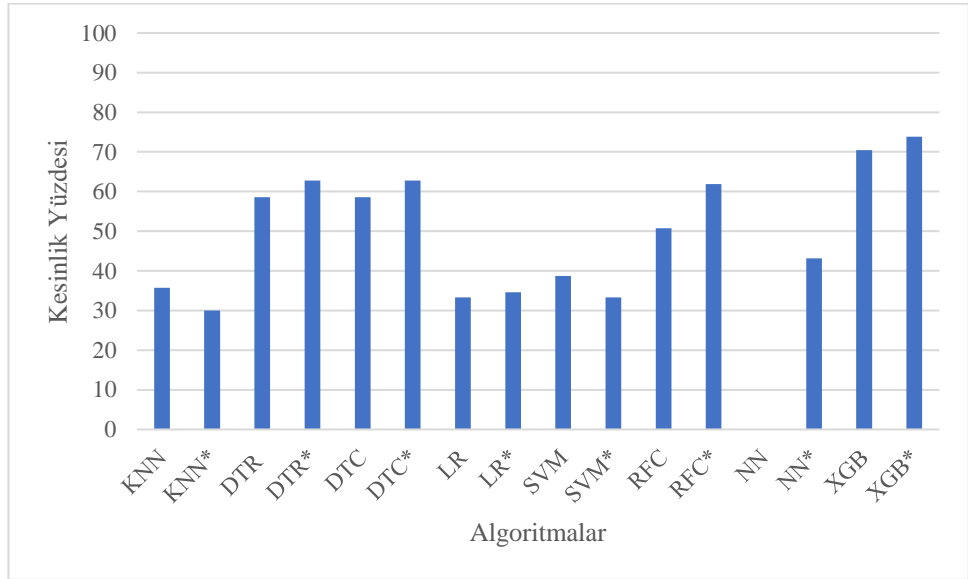
4.4.2.1. Varsayılan algoritma girdi parametreleri ile alınan sonuçlar

Yapılan analizlerdeki doğruluk skoru için en iyi skor XGB* ile %78,889 olmuştur.



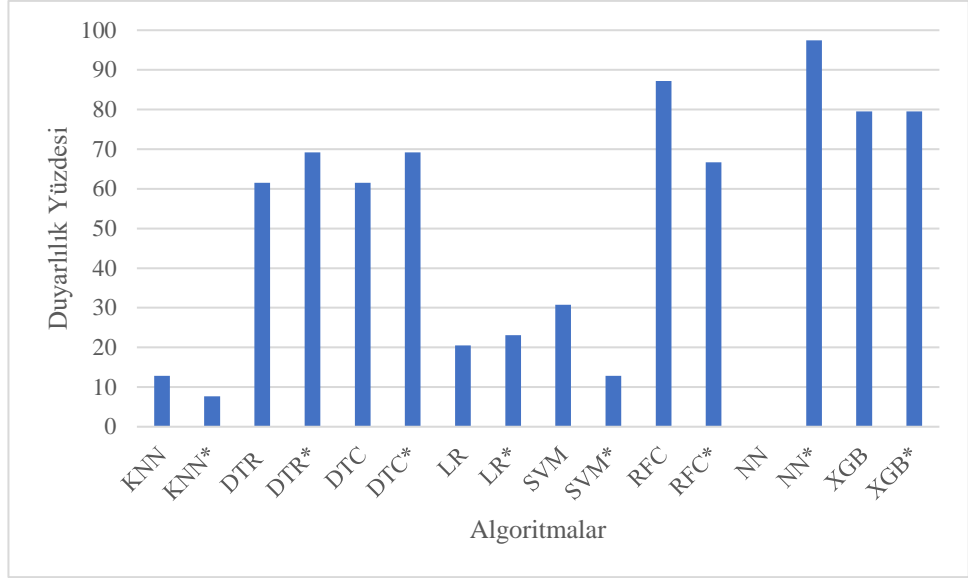
Şekil 123. Algoritmaların yüzelik ayırım yöntemi kullanımı ile doğruluk skor kıyaslanması

Yüzelik ayırım yöntemi ile algoritmaların kesinlik skor kıyaslanması yapılmıştır. XGB* algoritması %73,333 skor ile en yüksek başarıyı sağlamıştır.



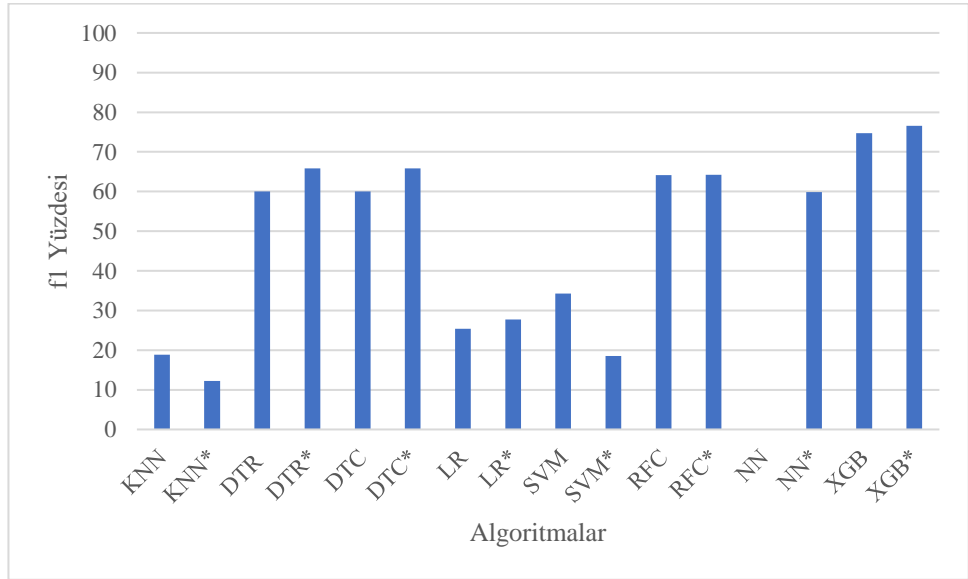
Şekil 124. Algoritmaların yüzelik ayırım yöntemi kullanımı ile kesinlik skor kıyaslanması

Algoritmaların duyarlılık skorları kıyaslanmıştır. NN* %97,436 başarı ile en yüksek sonucu vermiştir.



Şekil 125. Algoritmaların yüzelik ayırım yöntemi kullanımı ile duyarlılık skor kıyaslanması

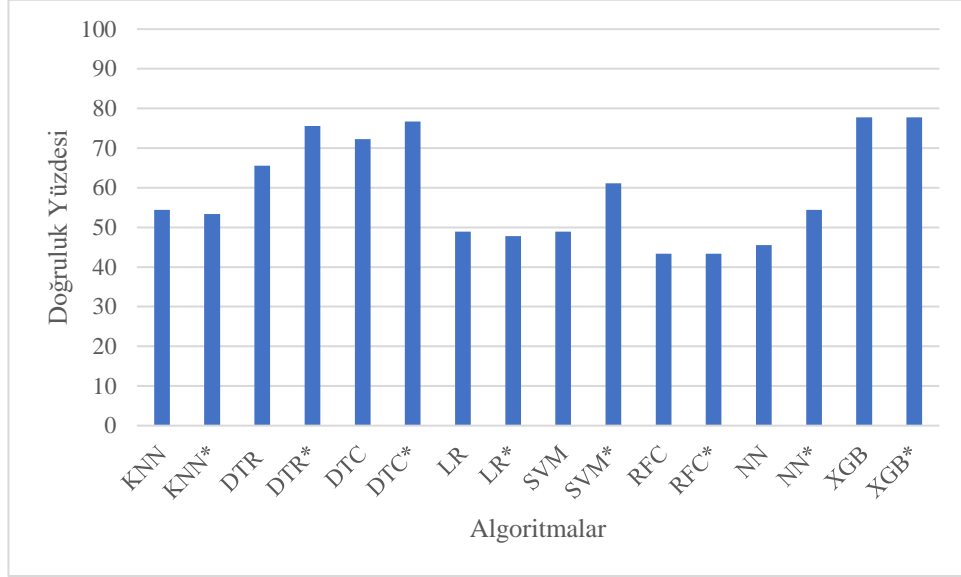
Algoritmaların f1 skor kıyaslanmasında XGB* %76,543 olarak en başarılı algoritma olduğunu göstermiştir.



Şekil 126. Algoritmaların yüzelik ayırım yöntemi kullanımı ile f1 skor kıyaslanması

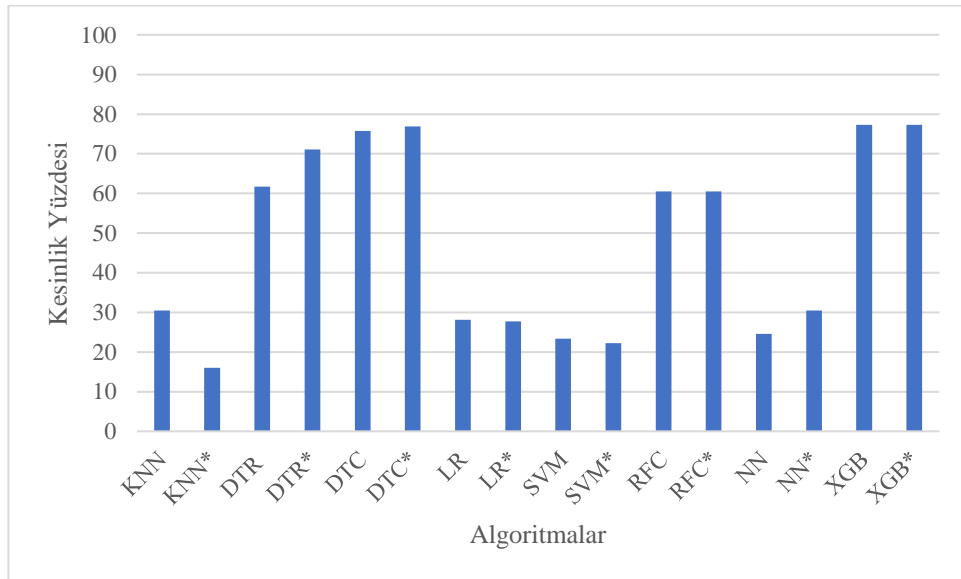
4.4.2.2. Hiper parametre optimizasyonu yapıldıktan sonra alınan sonuçlar

Yüzdelik ayırım yöntemi uygulanırken, hiper parametre optimizasyonu yapılmıştır. Tespit edilen parametreler kullanılarak algoritmaların skorları kıyaslanmıştır. %77,778 ile en yüksek doğruluk skorunu XGB ve XGB* almıştır.



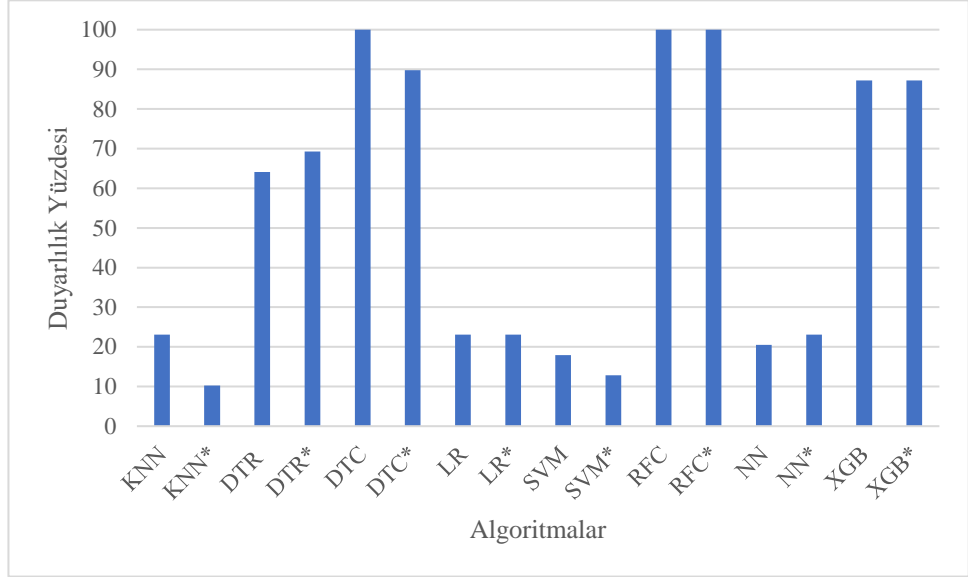
Şekil 127. Algoritmaların hiper parametre tespitleriyle yüzdelik ayırım yöntemi kullanımı, doğruluk skor kıyaslanması

Hiper parametre optimizasyonu ile kesinlik skor kıyaslanmasında %77,273 skor ile en yüksek skoru XGB ve XGB* almıştır.



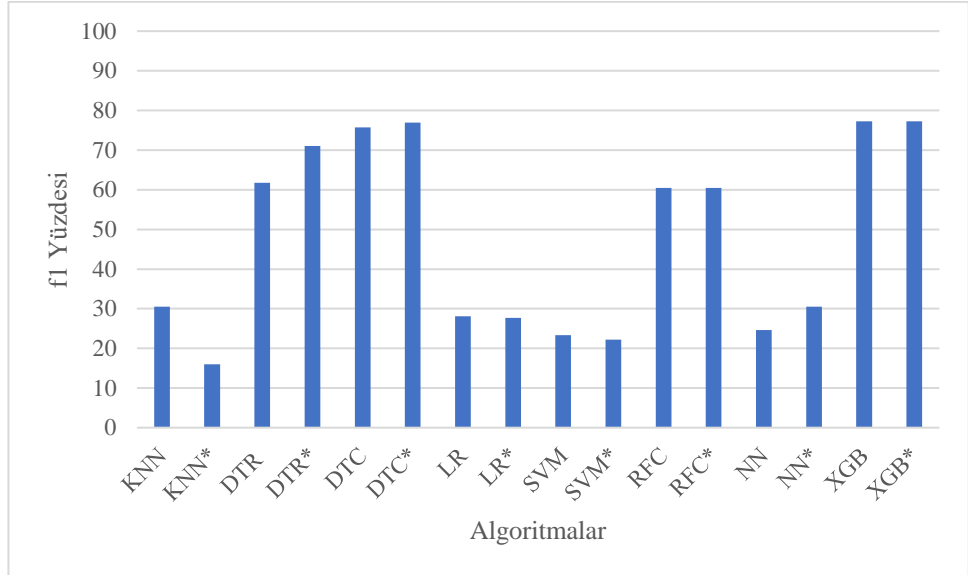
Şekil 128. Algoritmaların hiper parametre tespitleriyle yüzdelik ayırım yöntemi kullanımı, kesinlik skor kıyaslanması

Algoritmaların hiper parametre optimizasyonu ile analizlerinde en yüksek duyarlılık skorunu %100 ile DTC, RFC ve RFC* almıştır.



Şekil 129. Algoritmaların hiper parametre tespitleriyle yüzdelik ayırım yöntemi kullanımı, duyarlılık skor kıyaslanması

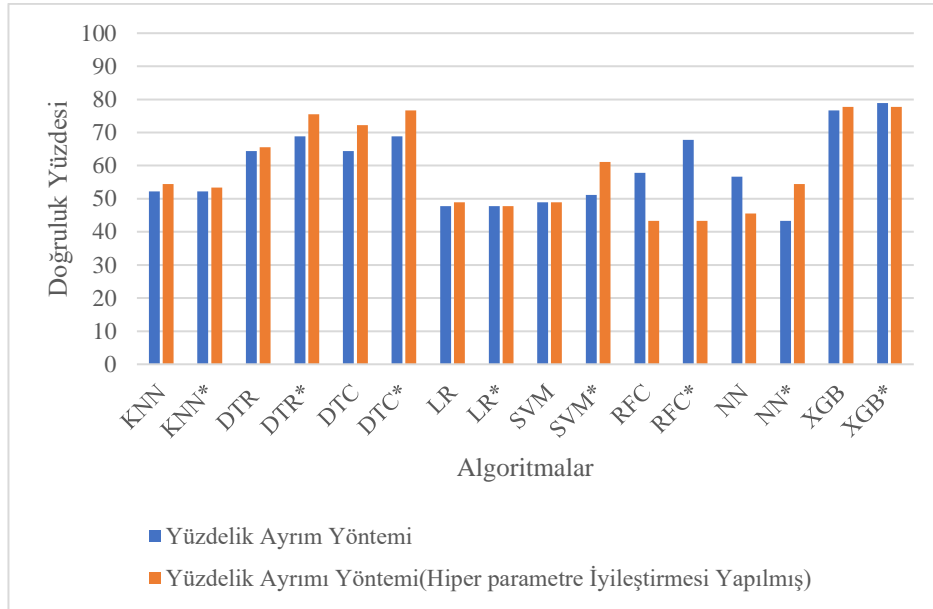
Algoritmalar arasında en yüksek f1 skoru %74,699 ile XGB almıştır.



Şekil 130. Algoritmaların hiper parametre tespitleriyle yüzdelik ayırım yöntemi kullanımı, f1 skor kıyaslanması

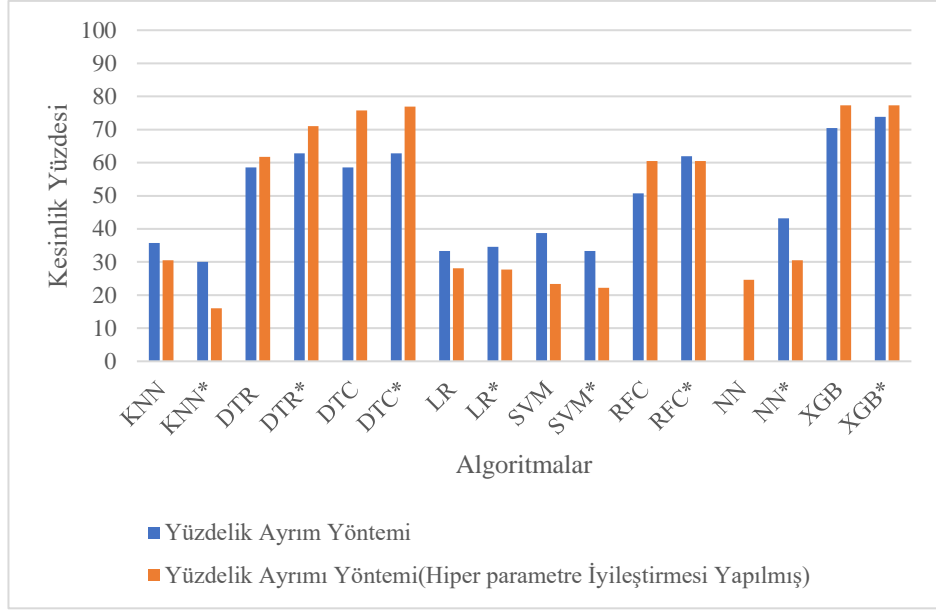
4.4.2.3. Karşılaştırmalı analiz

Algoritmaların sıralı veri tahminleme metodu ile kullanımında, hiper parametre optimizasyonu yapılmış ve yapılmamış modeller kıyaslanmıştır. Modellerin kıyaslanması sonucunda hiper parametre optimizasyonu yapılmış modellerde doğruluk yüzdesinin arttırdığı görülmüştür. Hiper parametrelerin optimizasyonu kullanımı ile en yüksek skoru %77,778 ile XGB ve XGB* almıştır. Hiper parametre optimizasyonu kullanılmadığı modelde ise %78,889 ile en yüksek yüzdeliği XGB* almıştır. Kıyaslanma sonucunda, RFC, RFC*, NN ve XGB* algoritmaları hiper parametre optimizasyonu ile yüzdelik skorları düştüğü tespit edilmiştir.



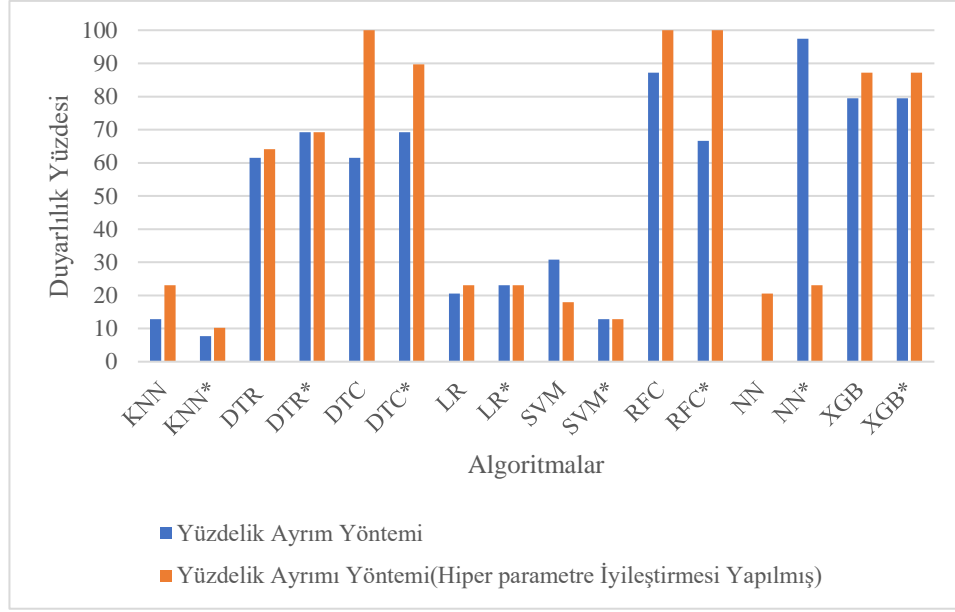
Şekil 131. Algoritmaların yüzdelik ayırımı yöntemi ile kullanımında doğruluk skoru kıyaslanması

Kesinlik skorları kıyaslanmasında, hiper parametre optimizasyonu ile XGB* algoritmasının %77,273 aldığı görülmüştür. Kesinlik skoru için hiper parametre optimizasyonu ile daha başarılı sonuç elde edildiği görülmüştür. Hiper parametre optimizasyonu kullanılması sonucunda, KNN, KNN*, LR*, LR, SVM, SVM* ve NN* algoritmalarının kesinlik skorlarının düştüğü görülmüştür.



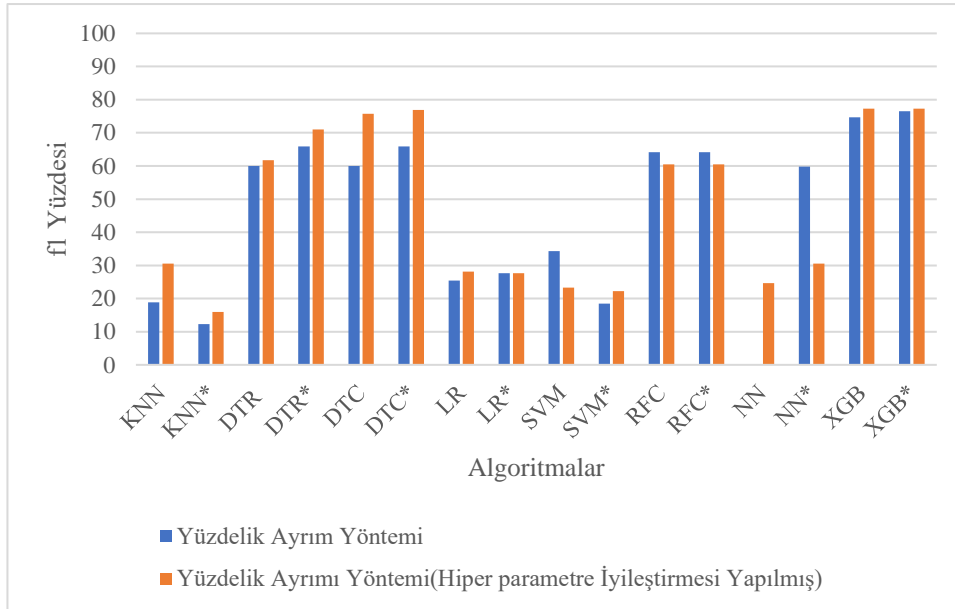
Şekil 132. Algoritmaların yüzdelik ayırımı yöntemi ile kullanımında kesinlik skoru kıyaslanması

Duyarlılık skoru hiper parametre optimizasyonu kullanımı ile artmıştır. %100 skor ile RFC ve NN* algoritmaları hiper parametre optimizasyonu kullanılmadan en yüksek başarıyı sağlamıştır. %100 skor ile DTC, RFC ve RFC* algoritmaları hiper parametre optimizasyonu yapılarak en yüksek skorları almıştır. Bunun yanı sıra, hiper parametre optimizasyonu kullanımı ile, LR* SVM SVM* ve NN* algoritmalarının duyarlılık yüzdelikleri düşmüştür.



Şekil 133. Algoritmaların yüzdelik ayrımı yöntemi ile kullanımında duyarlılık skoru kıyaslanması

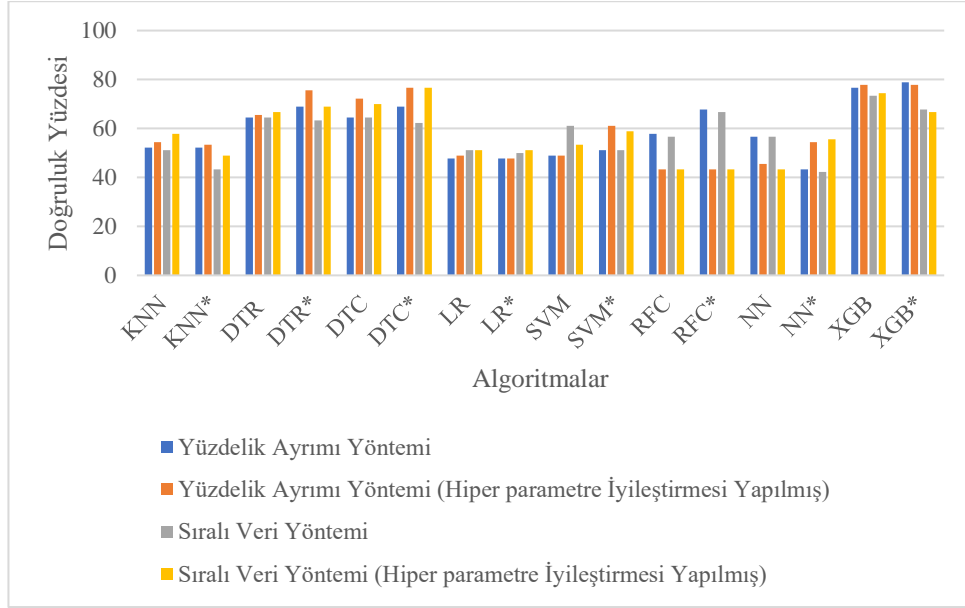
F1 skorları arasından en yüksek başarı, %77,273 ile hiper parametre optimizasyonu kullanılmış XGB ve XGB* algoritmaları olmuştur. LR*, SVM, SVM*, RFC, RFC* ve NN algoritmaları hiper parametre optimizasyonu kullanılmıř durumda yüzdelikleri düşmüřtür.



Şekil 134. Algoritmaların yüzdelik ayrımı yöntemi ile kullanımında f1 skoru kıyaslanması

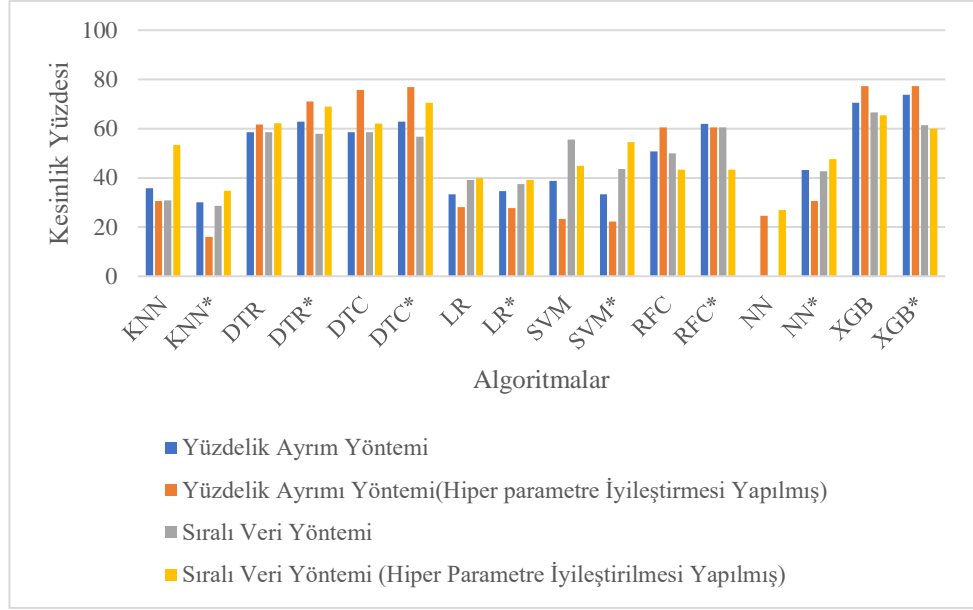
4.4.3. Karşılaştırmalı Algoritma Skorları

Yüzdelik ayırım yöntemi ve sıralı veri analizi metodu kıyaslanmıştır. Yöntemlerin hiper parametre optimizasyonu ile kullanıldığı ve kullanılmadığı durumlar göz önünde bulundurulmuştur. Buna göre; doğruluk skoru için en başarılı algoritmaların XGB ve XGB* olduğu gözlenmiştir. XGB* yüzdelik ayırım yöntemi %78,889 başarı yüzdesine ulaşmıştır.



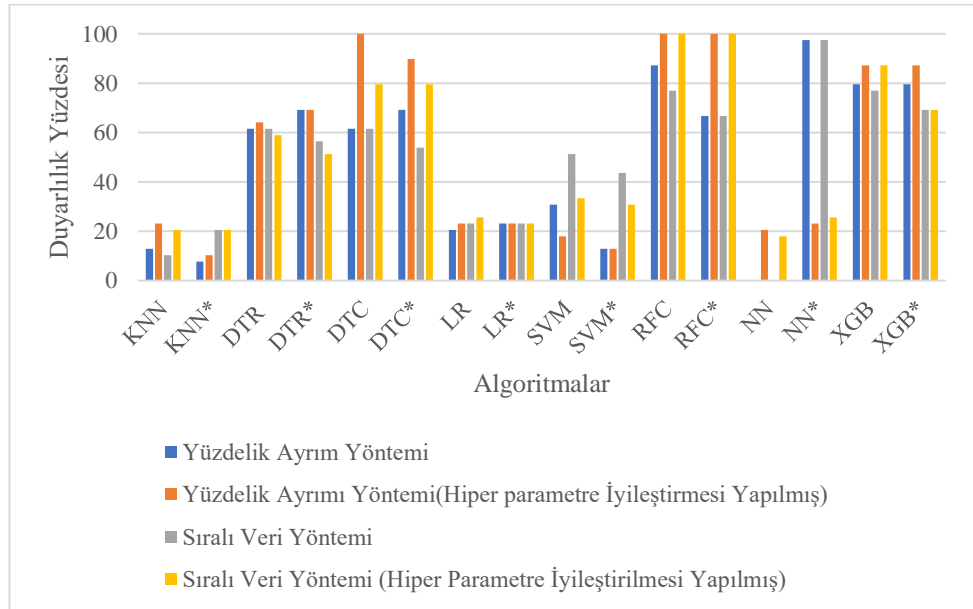
Şekil 135. Doğruluk skorlarının kıyaslanması

Kesinlik skor kıyaslamasında en yüksek skoru, %77,273 XGB ile XGB* algoritması yüzdelik ayırım yöntemi (hiper parametre optimizasyonu yapılmış) ile almıştır. Bu algoritmayı takip eden skorlar, %73,333 ile XGB* sıralı veri tahminleme metodu (hiper parametre optimizasyonu yapılmış), %70,455 ile DTC* tahminleme metodu (hiper parametre optimizasyonu yapılmış) takip etmiştir.



Şekil 136. Kesinlik skorlarının kıyaslanması

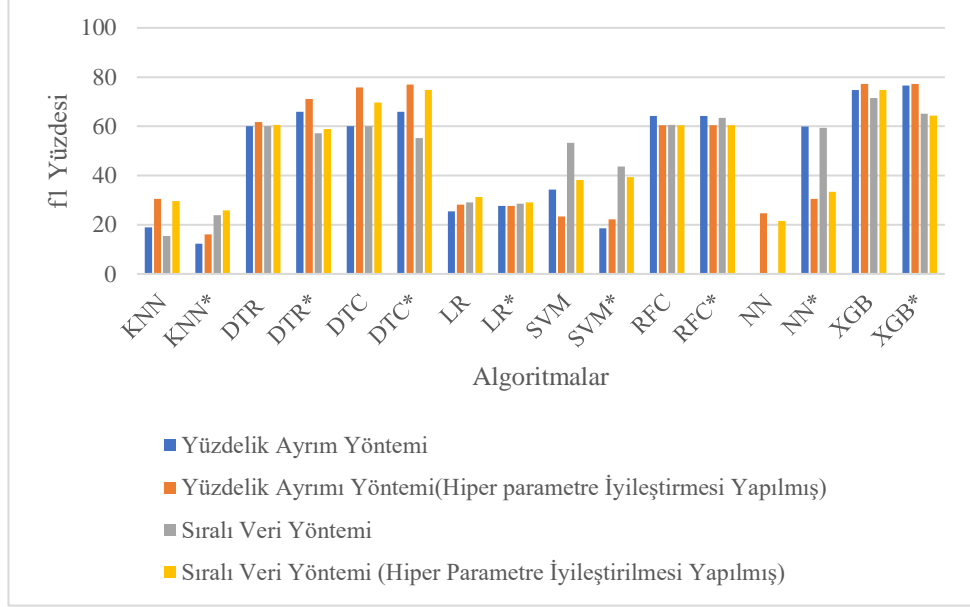
Duyarlılık skor kıyaslamasında en yüksek skorları, %100 ile SVM* ve NN* algoritması yüzdelik ayırım yöntemi ile, RFC, RFC* ve DTC algoritmaları yüzdelik ayırım yöntemi (hiper parametre optimizasyonu yapılmış) ve RFC, RFC* sıralı veri tahminleme metodu (hiper parametre optimizasyonu yapılmış) ile almıştır. Bu algoritmaları takip eden skor, %97,436 ile NN* sıralı veri tahminleme metodu olmuştur.



Şekil 137. Duyarlılık skorlarının kıyaslanması

F1 skor kıyaslanmasında en yüksek skoru yüzdelik ayırma yöntemi (hiper parametre optimizasyonu yapılmış) ile %77,273 olarak XGB olmuştur. Bu skoru takip eden

algoritmalar; sıralı veri tahminleme metodu (hiper parametre optimizasyonu yapılmış) kullanılarak %74,699 skoru alan DTC* ve yüzdelik ayırma yöntemi ile XGB, sıralı veri analizi yöntemi kullanılarak %71,429 skor sağlayan XGB olmuştur.



Şekil 138. F1 skorlarının kıyaslanması

4.4.4. Akaryakıt Miktarı Tahminleme Skorları

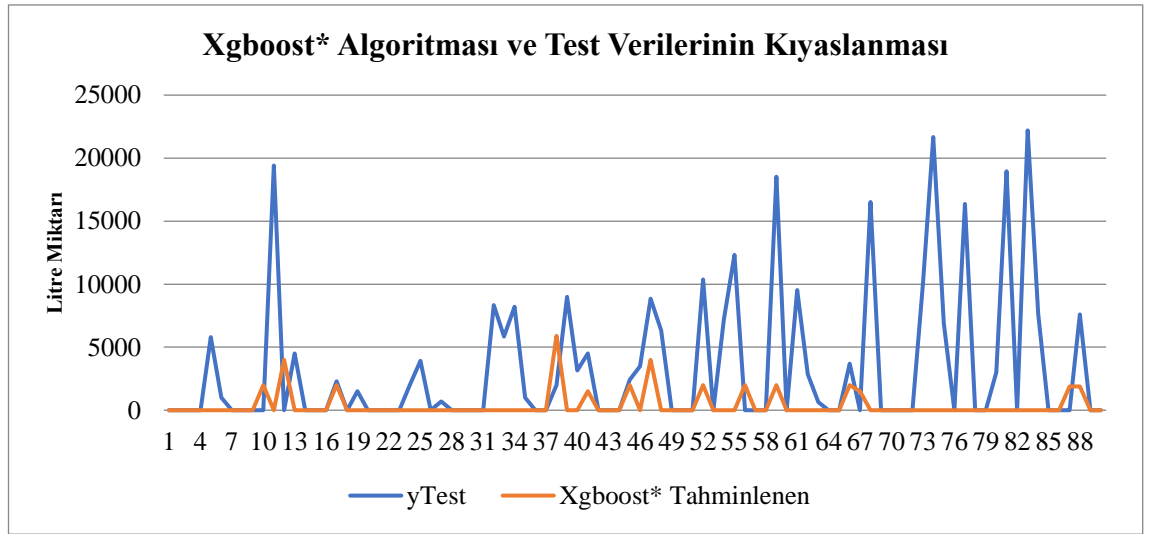
2019 ve 2020 verilerinin eğitim 2021 ilk çeyrek test edilen veri üzerinden sonraki gün litre tahminlemesi yapılmıştır. Yapılan tahminlemenin, algoritmalar baz alınarak MSE skorları tespit edilmiştir. MSE skorları alımı içi veri standart ölçeklendirilmiş olup sklearn kütüphanesinden faydalanılmıştır.

Tablo 21. Algoritmaların Litre Tahminleme İçin Aldıkları MSE Skorları

MSE Skorları	
Algoritma	Skor
KNN	2,03077
KNN*	1,82571
DTR	2,57479
DTR*	2,93652
DTC	1,56863
DTC*	2,71909
LR	1,90929
LR*	2,16695
SVM	2,15847
SVM*	2,05158
RFC	1,33085

RFC*	1,33085
NN	1,33085
NN*	1,33085
XGB	1,35458
XGB*	1,31283

Algoritmalarından en iyi skoru 1,31283 olduğu ve XGB* algoritmasının aldığı tespit edilmiştir.



Şekil 139. Veri Ve XGBoost* Algoritmasının Tahmininin Grafikselsel Gösterimi

5. SONUÇ

Bu çalışmada çiftçi akaryakıt alım verileri ile oluşturulan veri kümesi üzerinde, makine öğrenmesi algoritmaları ile akaryakıt alım tahminlemesi gerçekleştirilmiştir. Yapılan deneylerden 2019-2020 verileri için mevcut gün tahminlemesindeki en yüksek doğruluk başarısını yüzdelik ayırım yöntemi ile %82,514 başarısını sağlayan XGB algoritması elde etmiştir. En yüksek kesinlik skorunu %81,366 ile XGB* algoritması yüzdelik ayırım yöntemi ile almıştır. Duyarlılık skor kıyaslamasında en yüksek skorları, %100 ile NN* algoritması yüzdelik ayırım yöntemi ile, RFC ve RFC* algoritmaları yüzdelik ayırım yöntemi (hiper parametre optimizasyonu yapılmış) ile, NN* sıralı veri tahminleme metodu ile, RFC, RFC* ve NN* sıralı veri tahminleme metodu (hiper parametre optimizasyonu yapılmış) ile almıştır. En yüksek f1 skorunu %84,615 skorunu alan yüzdelik ayırım yöntemi ile RFC* ve hiper parametre optimizasyonu yapılmış yüzdelik ayırım yöntemi DTC algoritmaları olmuştur. Deney bazlı alınan sonuçlar ile veri adetlerinin kıyaslanması yapılabilmektedir. 2019-2020-2021 verilerinin kullanıldığı mevcut gün tahminleme deneyinde en yüksek doğruluk skorunu %77,778 olmuştur. Algoritmaların ve hedeflerin kıyaslanma imkanı sunulmuştur. Yöntemsel farklılıklarda alınan en yüksek sonuçlar mevcuttur. Mevcut sonuçların kıyaslanmaları yapılmıştır. LR yüzdelik ayırım yöntemi ve hiper parametre optimizasyonu yapılmış yüzdelik ayırım yöntemi XGB almıştır. En yüksek kesinlik skorunu hiper parametre optimizasyonu kullanılan yüzdelik ayırım metotlu %73,333 skoru ile DTR* almıştır. Duyarlılık skor kıyaslamasında en yüksek skorları, %100 ile SVM*, RFC, RFC* ve NN* algoritmaları yüzdelik ayırım yöntemi ile, DTC, DTC*, RFC ve RFC* algoritmaları hiper parametre optimizasyonu kullanılan yüzdelik ayırım yöntemi, RFC, RFC* ve NN* sıralı veri tahminleme metodu, DTC, DTC*, RFC ve RFC* hiper parametre optimizasyonu kullanılan sıralı veri tahminleme metodu ile almıştır. En yüksek f1 skor olan %79,167, LR algoritması yüzdelik ayırım yöntemi ile almıştır. Yapılan deneyler ile hiper parametre optimizasyonu yapılmadan ve yapıldıktan sonra algoritmaların verdikleri skorlar kıyaslanabilmektedir. Algoritma bazlı farkındalıklar gözlemlenebilmektedir. 2019 – 2020 verileri kullanılarak yapılan sonraki gün tahminleme de en yüksek doğruluk skorunu %79,235 ile XGB ve DTC algoritmaları almıştır. XGB hiper parametre optimizasyonu yapılarak yüzdelik ayırım ve sıralı veri tahminleme metotlarında skora ulaşmıştır. DTC

algoritması hiper parametre optimizasyonu yapılarak yüzdelerik ayırım metodunda ulaşmıştır. En yüksek kesinlik skoru olan %77,778'i XGB* algoritması yüzdelerik ayırım yöntemi ile almıştır. En yüksek duyarlılık skoru %100'ü SVM* algoritması yüzdelerik ayırım yöntemi ile, RFC ve RFC* algoritmaları hiper parametre optimizasyonu yapılarak yüzdelerik ayırım ve sıralı veri tahminleme metodu ile almıştır. En yüksek f1 skoru hiper parametre optimizasyonu kullanılarak yüzdelerik ayırım metodu ile XGB alınmıştır. En yüksek f1 skoru %82,883'tür. Yapılan 2 ana veri kümesinde kıyaslamalar yapılmıştır. 2 ana veri kümesindeki mevcut ve sonraki günlerin makine öğrenmesi modellerine işlenip sonuçları alınabilmektedir. 2019-2020 ve 2021 verileri kullanılarak yapılan sonraki gün analizinden en yüksek doğruluk skorunu yüzdelerik ayırım yöntemi ile XGB* almıştır ve skoru %78,889'dir. En yüksek kesinlik skoru %77,273'tür. En yüksek kesinlik skorunu alan XGB ve XGB* algoritmalarında hiper parametre optimizasyonu ile yüzdelerik ayırım yöntemi kullanılmıştır. Duyarlılık skor kıyaslamasında en yüksek skorları, %100 ile SVM* ve NN* algoritması yüzdelerik ayırım yöntemi ile, hiper parametre optimizasyonu yapılmış RFC, RFC* ve DTC algoritmaları yüzdelerik ayırım yöntemi ve hiper parametre optimizasyonu yapılmış RFC, RFC* algoritmaları sıralı veri tahminleme metodu ile almıştır. En yüksek duyarlılık skorunu hiper parametre optimizasyonu yapılan XGB algoritması yüzdelerik ayırım yöntemi ile %77,273 skor alarak sağlamıştır.

Bu tez ile optimum modeller tespit edilerek uygulamaya geçirilmesi hedeflenmiştir.

KAYNAKLAR

- Ahmadi, M. A., & Chen, Z. (2019). Comparison of machine learning methods for estimating permeability and porosity of oil reservoirs via petro-physical logs. *Petroleum*, 5(3), 271–284.
- Aydemir, E. (2019). Ders Geçme Notlarının Veri Madenciliği Yöntemleriyle Tahmin Edilmesi. *European Journal of Science and Technology*, 70–76.
- Ayık, Y. Z., Özdemir, A., & Yavuz, U. (2007). Lise Türü ve Lise Mezuniyet Başarısının, Kazanılan Fakülte İle İlişisinin Veri Madenciliği Tekniği İle Analizi. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 10(2), 441–454.
- Azazi, H., & Topkaya, Ö. (2017). Petrol Fiyatlarındaki Değişikliğin Türkiye İmalat Sanayi Ve İstihdamı Üzerindeki Etkileri. *Selçuk Üniversitesi Sosyal Bilimler Meslek Yüksekokulu Dergisi*, 20(1), 14–26.
- Çuhadar, M., Güngör, İ., & Göksu, A. (2009). Turizm Talebinin Yapay Sinir Ağları İle Tahmini Ve Zaman Serisi Yöntemleri İle Karşılaştırmalı Analizi: Antalya İline Yönelik Bir Uygulama. *Suleyman Demirel University Journal of Faculty of Economics & Administrative Sciences*.
- Demirel, Ö., Kakilli, A., & Tektaş, M. (2010). Anfis ve Arma Modelleri İle Elektrik Enerjisi Yük Tahmini. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 25(3).
- Erkan, E. T. İ., Dinçer, H., & Yüksel, S. (2019). G20 Ülkelerinde Bankacılık Sektörünün 5 Yıllık Geleceğinin Arima Yöntemi ile Tahmin Edilmesi. *Uluslararası Hukuk ve Sosyal Bilim Araştırmaları Dergisi*, 26–38.
- Hajizadeh, Y. (2019). Machine learning in oil and gas; a SWOT analysis approach. *Journal of Petroleum Science and Engineering*, 176, 661–663.
- Kanin, E. A., Osipov, A. A., Vainshtein, A. L., & Burnaev, E. V. (2019). A predictive model for steady-state multiphase pipe flow: Machine learning on lab data. *Journal of Petroleum Science and Engineering*, 180, 727–746.
- Karaca, O. (2003). Türkiye’de enflasyon-büyüme ilişkisi: zaman serisi analizi. *Doğuş Üniversitesi Dergisi*, 247–255.
- Koçtürk, D., & Avcıoğlu, A. (2007). Türkiye’de Bölgelere ve İllere Göre Tarımsal Mekanizasyon Düzeyinin Belirlenmesi. *Tarım Makinaları Bilimi Dergisi*, 3(1), 17–24.
- Koyuncugil, A., & Özgülbaş, N. (2009). Veri Madenciliği: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları. *INTERNATIONAL JOURNAL OF INFORMATICS TECHNOLOGIES*, 2(2).
- Özbay, Ö., & Ersoy, H. (2017). Öğrenme Yönetim Sistemi Üzerindeki Öğrenci Hareketliliğinin Veri Madenciliği Yöntemleriyle Analizi. *GEFAD / GUJGEF*, 37(2), 523–558.
- Özer, M., & Erdoğan, L. (2006). Türkiye’de ihracat, ithalat ve ekonomik büyüme arasındaki ilişkilerin zaman serisi analizi. *Ekonomik Yaklaşım*, 93–110.

- Patsilnakos, A., Artini, M., Papa, R., Sabatino, M., Božović, M., Garzoli, S., Vrenna, G., Buzzi, R., Manfredini, S., Selan, L., & Ragno, R. (2019). Machine Learning Analyses on Data including Essential Oil Chemical Composition and In Vitro Experimental Antibiofilm Activities against Staphylococcus Species. *Molecules*, 24(5), 890. <https://doi.org/q>
- Petrelli, M., & Perugini, D. (2016). Solving petrological problems through machine learning: the study case of tectonic discrimination using geochemical and isotopic data. *Contributions to Mineralogy and Petrology*, 171(10), 1–15.
- Sefer; Usapbeyli, U. (2015). Türkiye’de petrol tüketimi ve ekonomik büyüme arasındaki nedensellik ilişkisi. *Ankara Üniversitesi SBF Dergisi*, 70(3), 769–787.
- Seyrek, İ. H., & Ata, H. A. (2010). Veri Zarflama Analizi ve Veri Madenciliği ile Mevduat Bankalarında Etkinlik Ölçümü. *BDDK Bankacılık ve Finansal Piyasalar*, 4(2), 67–84.
- Tran, H., Kasha, A., Sakhaee-Pour, A., & Hussein, I. (2020). Predicting carbonate formation permeability using machine learning. *Journal of Petroleum Science and Engineering*, 195.
- Vahaplar, A., & İnceoğlu, M. M. (2001). Veri madenciliği ve elektronik ticaret. *Türkiye’de İnternet Konferansları VII*.
- Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2017). Predicting Risk of Suicide Attempts Over Time Through Machine Learning. *Clinical Psychological Science*, 5(3), 457–469.
- Yaman, K., Sarucan, A., Atak, M., & Aktürk, N. (2001). Dinamik çizelgeleme için görüntü işleme ve ARIMA modelleri yardımıyla veri hazırlama. *Journal of the Faculty of Engineering and Architecture of Gazi University*, 16(1–2), 19–40.
- Zarazua de Rubens, G. (2019). Who will buy electric vehicles after early adopters? Using machine learning to identify the electric vehicle mainstream market. *Energy*, 172, 243–244.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı : Mustafa ÇOBAN

EĞİTİM DURUMU

Lisans Öğrenimi :2018, KTO Karatay Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği

Yüksek Lisans Öğrenimi : Kto Karatay Üniversitesi Lisansüstü Eğitim Enstitüsü Elektrik-Elektronik Mühendisliği Anabilim Dalı Elektrik ve Bilgisayar Mühendisliği Tezli Yüksek Lisans Programı

Bildiği Yabancı Diller : İngilizce

Bilimsel Faaliyetleri : Öztürk, A., Çoban, M., (2019, April). Evaluation of Student Academics Performance via Machine Learning Algorithms International Conference on Advanced Technologies, Computer Engineering and Science (pp. 355-359).

İŞ DENEYİMİ

Stajlar : Türksat A.Ş stajyer, Utena Koleji stajyer

Projeler : 2018 D4R, 2020 Foressight Tarnet A.Ş.

Çalıştığı Kurumlar : 2021, Veri Analisti, Türk Kızılay

Tarih: 14 Temmuz 2021

EK 1. HATA MATRİSLERİ

5.1. 2019-2020 Verisi Üzerinden Tam Yıl Analizleri: Mevcut Gün Tahmini

5.1.1. Sıralı Veri Tahminleme Analizi

5.1.1.1. Varsayılan algoritma girdi parametreleri ile alınan sonuçlar

Sıralı veri tahminleme yöntemi ile yapılan analizlerin hata matrisleri tespit edilmiştir.

Tablo 22. Algoritmaların sıralı veri tahminleme yöntemi ile kullanımı

KNN		KNN*	
142	32	120	54
142	50	155	37
DTR		DTR*	
125	49	125	49
38	154	53	139
DTC		DTC*	
125	49	125	49
38	154	53	139
LR		LR*	
106	68	125	49
99	93	122	70
SVM		SVM*	
100	74	87	87
100	92	104	88
RFC		RFC*	
24	150	99	75
11	181	44	148
NN		NN*	
150	24	16	158
155	37	28	164
XGB		XGB*	
121	53	118	56
38	154	42	150

5.1.1.2. Hiper parametre optimizasyonu yapıldıktan sonra alınan sonuçlar

Tespit edilen hiper parametreler ile sıralı veri tahminleme yöntemi kullanılarak tespit edilen hata matrisleri gösterilmiştir.

Tablo 23. Algoritmaların sıralı veri tahminleme ve hiper parametre optimizasyonu yöntemi ile kullanımı

KNN		KNN*	
137	37	115	59
132	60	129	63
DTR		DTR*	
129	45	123	51
38	154	61	131
DTC		DTC*	
106	68	114	60
10	182	29	163
LR		LR*	
103	71	129	45
92	100	125	67
SVM		SVM*	
94	80	135	39
99	93	144	48
RFC		RFC*	
0	174	0	174
0	192	0	192
NN		NN*	
126	48	136	38
118	74	138	54
XGB		XGB*	
111	63	118	56
13	179	28	164

5.1.2. Yüzdelik Ayrım Analizi

5.1.2.1. Varsayılan algoritma girdi parametreleri ile alınan sonuçlar

Algoritmaların varsayılan hali ile kullanımı sonucunda hata matrisleri çıkarılmıştır.

Tablo 24. Algoritmaların yüzdelik ayrımı yöntemi ile kullanımı

KNN		KNN*	
144	30	152	22
139	53	155	37
DTR		DTR*	
118	56	156	18
38	154	142	50
DTC		DTC*	
118	56	156	18
38	154	142	50
LR		LR*	
114	60	125	49
91	101	116	76
SVM		SVM*	
102	72	1	173
92	100	3	189
RFC		RFC*	
11	163	93	81
2	190	37	155
NN		NN*	
151	23	18	156
149	43	33	159
XGB		XGB*	
119	55	159	15
24	168	140	52

5.1.2.2. Hiper parametre optimizasyonu yapıldıktan sonra alınan sonuçlar

Algoritmaların varsayılan hali ile kullanılması sonucunda oluşan doğruluk oranlarının hata matrisleri alınmıştır.

Tablo 25. Algoritmaların yüzdelik ayırım yöntemi ve hiper parametre ile kullanımı

KNN		KNN*	
138	36	149	25
134	58	158	34
DTR		DTR*	
119	55	121	53
38	154	39	153
DTC		DTC*	
107	67	105	69
9	183	19	173
LR		LR*	
117	57	131	43
97	95	126	66
SVM		SVM*	
87	87	144	30
79	113	143	49
RFC		RFC*	
0	174	0	174
0	192	0	192
NN		NN*	
130	44	133	41
119	73	130	62
XGB		XGB*	
106	68	137	37
8	184	74	118

5.2. 2019-2020 Verisi kullanılarak 2021 ilk çeyrek tahmini: Mevcut Gün Tahmini

5.2.1. Sıralı Veri Tahminleme Analizi

5.2.1.1. Varsayılan algoritma girdi parametreleri ile alınan sonuçlar

Sıralı veri tahminleme yöntemiyle yapılan analizlerin hata matrisleri tespit edilmiştir.

Tablo 26. Algoritmaların sıralı veri tahminleme yöntemi kullanımı

KNN		KNN*	
42	9	31	20
35	4	31	8
DTR		DTR*	
34	17	35	16
15	24	17	22
DTC		DTC*	
34	17	35	16
15	24	18	21
LR		LR*	
37	14	36	15
30	9	30	9
SVM		SVM*	
35	16	29	22
19	20	22	17
RFC		RFC*	
21	30	34	17
9	30	13	26
NN		NN*	
51	0	0	51
39	0	1	38
XGB		XGB*	
36	15	34	17
9	30	12	27

5.2.1.2. Hiper parametre optimizasyonu yapıldıktan sonra alınan sonuçlar

Hiper parametre optimizasyonu ile sıralı veri tahminleme yöntemi kullanılarak tespit edilen hata matrisleri gösterilmiştir.

Tablo 27. Algoritmaların sıralı veri tahminleme ve hiper parametre optimizasyonu yöntemi ile kullanımı

KNN		KNN*	
44	7	36	15
31	8	31	8
DTR		DTR*	
37	14	42	9
16	23	19	20
DTC		DTC*	
32	19	38	13
8	31	8	31
LR		LR*	
36	15	37	14
29	10	30	9
SVM		SVM*	
35	16	41	10
26	13	27	12
RFC		RFC*	
0	51	0	51
0	39	0	39
NN		NN*	
32	19	40	11
32	7	29	10
XGB		XGB*	
33	18	33	18
5	34	12	27

5.2.2. Yüzdelik Ayrım Analizi

5.2.2.1. Varsayılan algoritma girdi parametreleri ile alınan sonuçlar

Algoritmaların varsayılan hali ile kullanımı sonucunda hata matrisleri çıkarılmıştı

r.

Tablo 28. Algoritmaların yüzdelik ayırım yöntemi ile kullanımı

KNN		KNN*	
42	9	44	7
34	5	36	3
DTR		DTR*	
34	17	35	16
15	24	12	27
DTC		DTC*	
34	17	35	16
15	24	12	27
LR		LR*	
35	16	34	17
31	8	30	9
SVM		SVM*	
32	19	41	10
27	12	34	5
RFC		RFC*	
18	33	35	16
5	34	13	26
NN		NN*	
51	0	1	50
39	0	1	38
XGB		XGB*	
38	13	40	11
8	31	8	31

5.2.2.2. Hiper parametre optimizasyonu yapıldıktan sonra alınan sonuçlar

Hiper parametre optimizasyonu ile kullanılması sonucunda oluşan doğruluk oranlarının hata matrisleri alınmıştır.

Tablo 29. Algoritmaların yüzdelik ayırım yöntemi ve hiper parametre ile kullanımı

KNN		KNN*	
40	11	44	7
30	9	35	4
DTR		DTR*	
34	17	41	10
14	25	12	27
DTC		DTC*	
26	25	34	17
0	39	4	35
LR		LR*	
35	16	34	17
30	9	30	9
SVM		SVM*	
37	14	50	1
32	7	34	5
RFC		RFC*	
0	51	0	51
0	39	0	39
NN		NN*	
33	18	40	11
31	8	30	9
XGB		XGB*	
36	15	36	15
5	34	5	34

5.3. 2019-2020 Verisi Üzerinden Tam Yıl Analizleri: Sonraki Gün Tahmini

5.3.1. Sıralı Veri Tahminleme Analizi

5.3.1.1. Varsayılan algoritma girdi parametreleri ile alınan sonuçlar

Sıralı veri tahminleme analizi yöntemi ile sonraki gün yapılan analizlerin hata matrisleri tespit edilmiştir.

Tablo 30. Algoritmaların sıralı veri tahminleme metodu ile kullanımı

KNN		KNN*	
131	43	138	36
76	116	89	103
DTR		DTR*	
123	51	124	50
48	144	58	134
DTC		DTC*	
123	51	123	51
47	145	58	134
LR		LR*	
119	55	92	82
13	179	73	119
SVM		SVM*	
107	67	104	70
45	147	101	91
RFC		RFC*	
101	73	111	63
4	188	6	186
NN		NN*	
174	0	0	174
192	0	0	192
XGB		XGB*	
128	46	129	45
29	163	34	158

5.3.1.2. Hiper parametre optimizasyonu yapıldıktan sonra alınan sonuçlar

Hiper parametre optimizasyonu ile sıralı veri tahminleme yöntemi kullanılarak tespit edilen hata matrisleri gösterilmiştir.

Tablo 31. Algoritmaların sıralı veri tahminleme ve hiper parametre optimizasyonu yöntemi ile kullanımı

KNN		KNN*	
122	52	125	49
47	145	78	114
DTR		DTR*	
125	49	126	48
51	141	52	140
DTC		DTC*	
109	65	120	54
4	188	21	171
LR		LR*	
118	56	90	84
14	178	75	117
SVM		SVM*	
111	63	101	73
33	159	115	77
RFC		RFC*	
0	174	0	174
0	192	0	192
NN		NN*	
123	51	58	116
27	165	39	153
XGB		XGB*	
122	52	126	48
15	177	22	170

5.3.2. Yüzdelik Ayrım Analizi

5.3.2.1. Varsayılan algoritma girdi parametreleri ile alınan sonuçlar

Algoritmaların varsayılan hali yüzdelik ayrım yöntemi ile kullanılmıştır. Hata matrisleri çıkartılmıştır.

Tablo 32. Algoritmaların yüzdelik ayırım yöntemi ile kullanımı

KNN		KNN*	
122	52	133	41
84	108	122	70
DTR		DTR*	
129	45	85	89
53	139	117	75
DTC		DTC*	
129	45	85	89
53	139	117	75
LR		LR*	
114	60	79	95
14	178	75	117
SVM		SVM*	
143	31	174	0
121	71	192	0
RFC		RFC*	
102	72	111	63
4	188	5	187
NN		NN*	
174	0	0	174
192	0	0	192
XGB		XGB*	
127	47	144	30
17	175	61	131

5.3.2.2. Hiper parametre optimizasyonu yapıldıktan sonra alınan sonuçlar

Hiper parametre optimizasyonu yapılmış algoritmalara yüzdelik ayırım yöntemi uygulanmıştır. Hata matrisleri alınmıştır.

Tablo 33. Algoritmaların yüzdelik ayırım yöntemi ve hiper parametre ile kullanımı

KNN		KNN*	
115	59	127	47
51	141	107	85
DTR		DTR*	
118	56	118	56
36	156	37	155
DTC		DTC*	
111	63	128	46
5	187	46	146
LR		LR*	
117	57	36	138
20	172	16	176
SVM		SVM*	
131	43	174	0
72	120	192	0
RFC		RFC*	
0	174	0	174
0	192	0	192
NN		NN*	
107	67	174	0
19	173	192	0
XGB		XGB*	
116	58	132	42
10	182	67	125

5.4. 2019-2020 Verisi kullanılarak 2021 ilk çeyrek tahmini: Sonraki Gün Tahmini

5.4.1. Sıralı Veri Tahminleme Analizi

5.4.1.1. Varsayılan algoritma girdi parametreleri ile alınan sonuçlar

Sıralı veri tahminleme analizi yöntemi ile sonraki gün yapılan analizlerin hata matrisleri tespit edilmiştir.

Tablo 34. Algoritmaların sıralı veri tahminleme metodu ile kullanımı

KNN		KNN*	
38	13	42	9
24	15	29	10
DTR		DTR*	
37	14	38	13
11	28	12	27
DTC		DTC*	
37	14	37	14
11	28	12	27
LR		LR*	
32	19	50	1
5	34	36	3
SVM		SVM*	
40	11	30	21
20	19	24	15
RFC		RFC*	
27	24	27	24
0	39	0	39
NN		NN*	
51	0	0	51
39	0	0	39
XGB		XGB*	
36	15	33	18
9	30	16	23

5.4.1.2. Hiper parametre optimizasyonu yapıldıktan sonra alınan sonuçlar

Hiper parametre optimizasyonu yapılan algoritmalara sıralı veri tahminleme analizi uygulanmıştır. Hata matrisleri gösterilmiştir.

Tablo 35. Algoritmaların sıralı veri tahminleme ve hiper parametre optimizasyonu yöntemi ile kullanımı

KNN		KNN*	
34	17	41	10
16	23	29	10
DTR		DTR*	
36	15	40	11
11	28	11	28
DTC		DTC*	
27	24	26	25
0	39	0	39
LR		LR*	
31	20	51	0
3	36	39	0
SVM		SVM*	
33	18	31	20
9	30	20	19
RFC		RFC*	
0	51	0	51
0	39	0	39
NN		NN*	
43	8	46	5
18	21	34	5
XGB		XGB*	
35	16	39	12
7	32	10	29

5.4.2. Yüzdelik Ayrım Analizi

5.4.2.1. Varsayılan algoritma girdi parametreleri ile alınan sonuçlar

Algoritmaların varsayılan hali yüzdelik ayrım yöntemi ile kullanılmıştır. Hata matrisleri çıkartılmıştır.

Tablo 36. Algoritmaların yüzdelik ayırım yöntemi ile kullanımı

KNN		KNN*	
36	15	47	4
24	15	32	7
DTR		DTR*	
40	11	41	10
15	24	26	13
DTC		DTC*	
40	11	41	10
15	24	26	13
LR		LR*	
32	19	50	1
1	38	37	2
SVM		SVM*	
38	13	0	51
13	26	0	39
RFC		RFC*	
27	24	27	24
0	39	0	39
NN		NN*	
51	0	0	51
39	0	0	39
XGB		XGB*	
39	12	39	12
14	25	13	26

5.4.2.2. Hiper parametre optimizasyonu yapıldıktan sonra alınan sonuçlar

Hiper parametre optimizasyonu yapılmış algoritmalara yüzdelik ayırım yöntemi uygulanmıştır. Hata matrisleri alınmıştır.

Tablo 37. Algoritmaların yüzdelik ayırım yöntemi ve hiper parametre ile kullanımı

KNN		KNN*	
32	19	41	10
13	26	32	7
DTR		DTR*	
39	12	43	8
17	22	17	22
DTC		DTC*	
27	24	26	25
0	39	0	39
LR		LR*	
32	19	51	0
4	35	39	0
SVM		SVM*	
32	19	51	0
4	35	39	0
RFC		RFC*	
0	51	0	51
0	39	0	39
NN		NN*	
42	9	51	0
21	18	39	0
XGB		XGB*	
33	18	38	13
2	37	13	26