

ŞARAP KALİTESİNİN SINIFLANDIRILMASINDA VERİ DENGESİZLİĞİ VE MODEL PERFORMANSLARININ İNCELENMESİ

Bilgisayar Mühendisliği, Fen Bilimleri Enstitüsü, Gazi Üniversitesi, Ankara, Türkiye

Berkay Köklü berkay.koklu@gazi.edu.tr

Bilgisayar Mühendisliği, Fen Bilimleri Enstitüsü, Gazi Üniversitesi, Ankara, Türkiye

Mehmet Fatih Ekici mfatih.ekici@gazi.edu.tr

-ÇALIŞMANIN ÖZETİ-

Bu çalışmanın amacı, şarapların kimyasal özelliklerine dayanarak kalite sınıflandırmasını gerçekleştirmek ve çeşitli makine öğrenmesi algoritmalarının performanslarını karşılaştırmaktır. Veri seti olarak UCI Machine Learning Repository’de yer alan kırmızı ve beyaz şarap verileri kullanılmıştır. Öncelikle her iki veri seti birleştirilmiş, kalite değişkeni düşük, orta ve yüksek olmak üzere üç sınıfa ayrılmış ve dengesiz sınıf dağılımı dikkate alınarak veri analizi yapılmıştır.

Veri ön işleme adımlarında eksik veri temizliği, standardizasyon ve sınıf etiketlemesi uygulanmıştır. Ardından, Decision Tree, Random Forest, KNN, Logistic Regression, Naive Bayes, SVM ve XGBoost gibi çeşitli klasik makine öğrenmesi modelleri eğitilerek doğruluk oranları karşılaştırılmıştır. Ayrıca bir Yapay Sinir Ağı (YSA) modeli TensorFlow ile oluşturularak bu modellerle performans açısından kıyaslanmıştır. Modellerin doğruluk oranları, sınıflara göre precision-recall-f1 değerleri ve karmaşıklık matrisleri kullanılarak değerlendirilmiştir.

Elde edilen sonuçlara göre, XGBoost ve Random Forest modelleri en yüksek doğruluk oranını göstermiştir. YSA modeli ise yüksek doğrulukla birlikte her sınıf için dengeli f1-score sonuçları üretmiştir. Özellik önem düzeyi analizinde alkol, uçucu asitlik ve sitrik asit gibi değişkenlerin kalite üzerinde daha etkili olduğu gözlemlenmiştir.

Sonuç olarak, bu çalışma şarap kalitesinin kimyasal ölçümler yoluyla başarıyla tahmin edilebileceğini ve farklı modellerin bu süreçteki başarı düzeylerinin anlamlı şekilde değiştiğini ortaya koymaktadır.

-GİRİŞ-

Gıda endüstrisinde kalite kontrol süreçleri, ürün güvenliği ve müşteri memnuniyeti açısından büyük önem taşımaktadır. Bu bağlamda şarap üretiminde kalite değerlendirmesi, geleneksel olarak uzman tadımcılar tarafından yapılmakta olup hem zaman alıcı hem de öznel sonuçlar doğurabilmektedir. Günümüzde ise gelişen veri bilimi ve yapay zeka teknikleri sayesinde bu değerlendirmelerin otomatik ve nesnel biçimde yapılması mümkün hale gelmiştir.

Bu çalışmada, şarapların kimyasal özelliklerine dayanarak kalite sınıflandırması yapılması hedeflenmiştir. UCI Makine Öğrenmesi Deposundan alınan kırmızı ve beyaz şarap veri setleri birleştirilmiş ve kalite değerleri düşük, orta ve yüksek olmak üzere üç sınıfa ayrılmıştır. Sınıf dengesizliği analizi yapılmış ve uygun ön işleme adımları uygulanmıştır. Ardından çeşitli

makine öğrenmesi algoritmaları ve bir yapay sinir ağı modeli ile sınıflandırma işlemi gerçekleştirilmiştir.

Çalışmanın temel amacı, farklı sınıflandırma algoritmalarının şarap kalitesi tahmini üzerindeki başarısını karşılaştırmak, önemli özellikleri belirlemek ve veri dengesizliğinin model performansına etkisini ortaya koymaktır. Bu kapsamda şu araştırma sorularına yanıt aranmıştır:

Kimyasal özelliklere bakılarak şarap kalitesi ne ölçüde tahmin edilebilir?

Hangi makine öğrenmesi algoritması bu problemde daha iyi sonuç verir?

Yapay sinir ağları, klasik yöntemlere kıyasla ne derece başarılıdır?

Şarap kalitesini etkileyen en önemli özellikler nelerdir?

Bu sorulara verilen yanıtlarla hem akademik hem endüstriyel anlamda katkı sağlamak amaçlanmıştır.

-MATERYAL VE YÖNTEM –

Bu çalışmada, şarap kalitesini kimyasal özellikler üzerinden sınıflandırmak amacıyla veri bilimi ve makine öğrenmesi yöntemleri kullanılmıştır. Kullanılan veri seti, UCI Machine Learning Repository'de yer alan Wine Quality Dataset olup, hem kırmızı (red) hem de beyaz (white) şaraplara ait örneklerden oluşmaktadır. Kırmızı şarap için 1.599, beyaz şarap için 4.898 örnek bulunmaktadır. Her bir örnek, 11 adet kimyasal özellik (örneğin: alkol, uçucu asitlik, pH, sülfat miktarı) ve bir adet kalite etiketi (0–10 arasında skalanmış) içermektedir.

Veri ön işleme sürecinde şu adımlar uygulanmıştır:

Kırmızı ve beyaz şarap veri setleri birleştirilmiş ve her örneğe "renk" etiketi eklenmiştir.

Kalite değişkeni, sınıflandırma problemini kolaylaştırmak amacıyla 3 sınıfa indirgenmiştir:

0: Düşük kalite (0–4),

1: Orta kalite (5–6),

2: Yüksek kalite (7 ve üzeri).

Sınıf dağılımı incelenmiş ve sınıflar arası dengesizlik gözlemlenmiştir.

Özellikler, StandardScaler ile ölçeklendirilmiştir.

Veri, eğitim (%70) ve test (%30) olmak üzere ayrılmıştır.

Modelleme aşamasında aşağıdaki yöntemler kullanılmıştır:

Klasik Makine Öğrenmesi Algoritmaları:

Decision Tree, Random Forest, K-Nearest Neighbors, Logistic Regression, Naive Bayes, Support Vector Machine ve XGBoost.

Derin Öğrenme Yöntemi:

Bir Yapay Sinir Ağı (YSA) modeli, TensorFlow/Keras kütüphaneleri kullanılarak oluşturulmuştur. Model 128 ve 64 nöronlu iki gizli katmandan ve softmax aktivasyonlu 3 çıkış nöronundan oluşmaktadır.

Model performansları; doğruluk oranı, precision-recall-f1 skoru, karmaşıklık matrisi ve doğruluk/kayıp eğrileri üzerinden değerlendirilmiştir. Ayrıca Random Forest modeli kullanılarak özellik önem dereceleri, tüm veri seti için ise korelasyon matrisi çıkarılmıştır.

-LİTERATÜR TARAMASI-

Şarap kalitesi tahmini üzerine yapılan çalışmalar son yıllarda veri madenciliği ve makine öğrenmesi yöntemlerinin gelişmesiyle artmıştır. Özellikle şarapların kimyasal bileşenleri ile kalite etiketleri arasındaki ilişkinin modellenmesi, hem akademik çevrelerde hem de gıda endüstrisinde dikkat çekmektedir.

Cortez ve arkadaşları (2009), UCI veri setinin ilk sağlayıcıları olarak karar ağaçları (DT), destek vektör makineleri (SVM) ve yapay sinir ağları (ANN) gibi yöntemleri karşılaştırmış ve SVM'nin daha başarılı sonuçlar verdiğini raporlamıştır. Çalışmada ayrıca özellik seçiminin model başarısına etkisi vurgulanmıştır.

Subasi (2021), aynı veri seti üzerinde Random Forest, Naive Bayes ve KNN algoritmalarını uygulayarak Random Forest'in özellikle yüksek doğruluk oranları sağladığını göstermiştir. Benzer şekilde, Zhang ve arkadaşları (2020) yapay sinir ağı modelleri ile kalite tahmininde %70'in üzerinde doğruluk elde etmiş ve katman sayısı ile başarı arasında doğrusal olmayan bir ilişki bulmuşlardır.

Bazı araştırmalarda, veri dengesizliği probleminin sınıflandırma başarısını olumsuz etkilediği belirtilmiş ve SMOTE gibi yeniden örnekleme teknikleri önerilmiştir (Fernandes & Santos, 2018). Bu nedenle sınıfların yeniden etiketlenmesi ve dengeli hale getirilmesi, model performanslarını artırmak adına önemli bir strateji olarak öne çıkmaktadır.

Bu çalışmada ise mevcut literatürden farklı olarak hem klasik makine öğrenmesi algoritmaları hem de derin öğrenme tabanlı bir yapay sinir ağı modeli birlikte ele alınarak kapsamlı bir

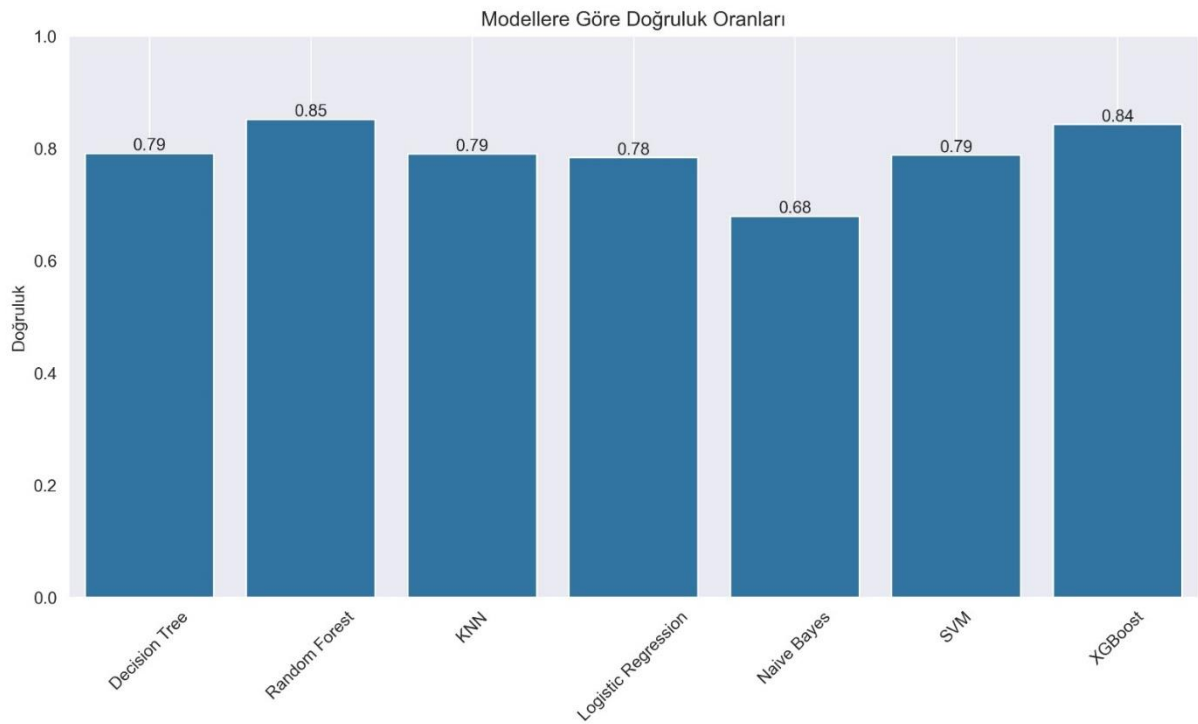
karşılaştırma yapılmıştır. Ayrıca veri dengesizliği, özellik önemi ve korelasyon analizleriyle model yorumlanabilirliği de sağlanmıştır.

-BULGULAR VE TARTIŞMA-

Çalışmada, şarap kalitesini üç sınıfa ayırarak farklı makine öğrenmesi algoritmaları ile sınıflandırma yapılmıştır. Aşağıda hem klasik makine öğrenmesi modelleri hem de yapay sinir ağı (YSA) modeli için elde edilen bulgular özetlenmiştir.

3.1. Makine Öğrenmesi Modellerinin Başarıları

Yedi farklı modelin test verisi üzerinde doğruluk (accuracy) skorları şu şekilde elde edilmiştir:



En yüksek doğruluk Random Forest (0.85) ve XGBoost (0.84) modelleri ile elde edilmiştir. Bu, literatürde daha önce yapılmış benzer çalışmalarla uyumludur ve toplu yöntemlerin daha stabil sonuçlar verdiğini göstermektedir.

-MODELİN YAPAY SİNİR AĞLARI (YSA) İLE GÖSTERİMLERİ-

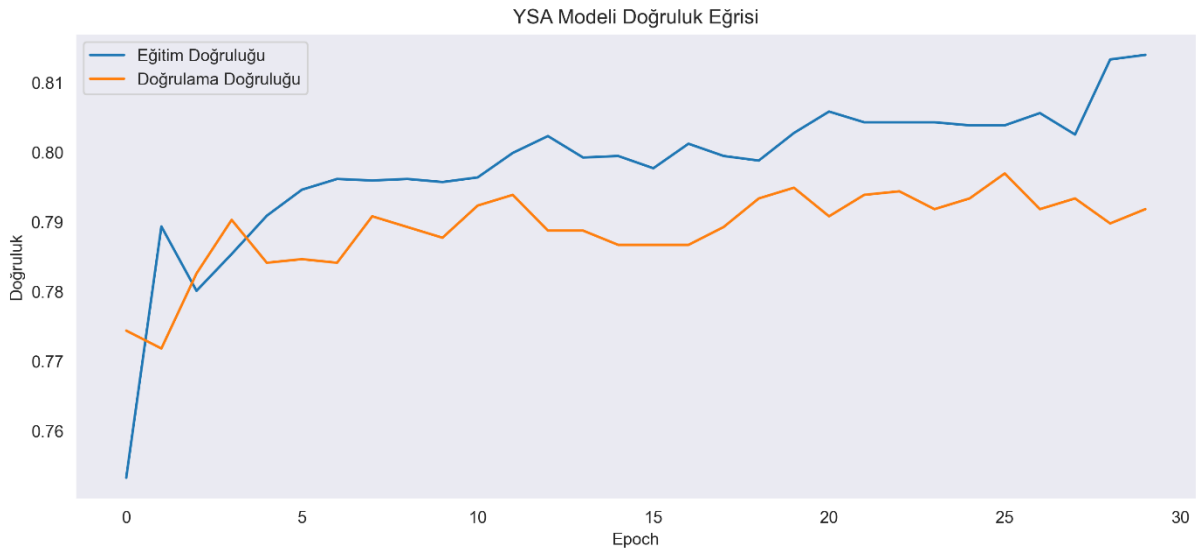
Yapay Sinir Ağı (YSA) modeli kullanılarak yürütülen deneysel çalışmada eğitim ve doğrulama süreçlerine ait doğruluk ve kayıp değerleri 30 epoch boyunca takip edilmiştir.

Doğruluk Değerleri:

Aşağıdaki Şekil 1’de eğitim ve doğrulama veri kümeleri için doğruluk (accuracy) değerlerinin epoch bazında değişimi gösterilmektedir.

****Epoch**, modelin eğitim sürecindeki bir **turunu** temsil eder. Her bir epoch tamamlandığında, modelin veri setindeki tüm örneklerle maruz kaldığı ve öğrenme adımı attığı anlamına gelir. Modelin performansını artırmak için genellikle veri seti üzerinde birden fazla epoch (tur) eğitim yapılır.

Şekil 1. YSA modeli doğruluk eğrisi.



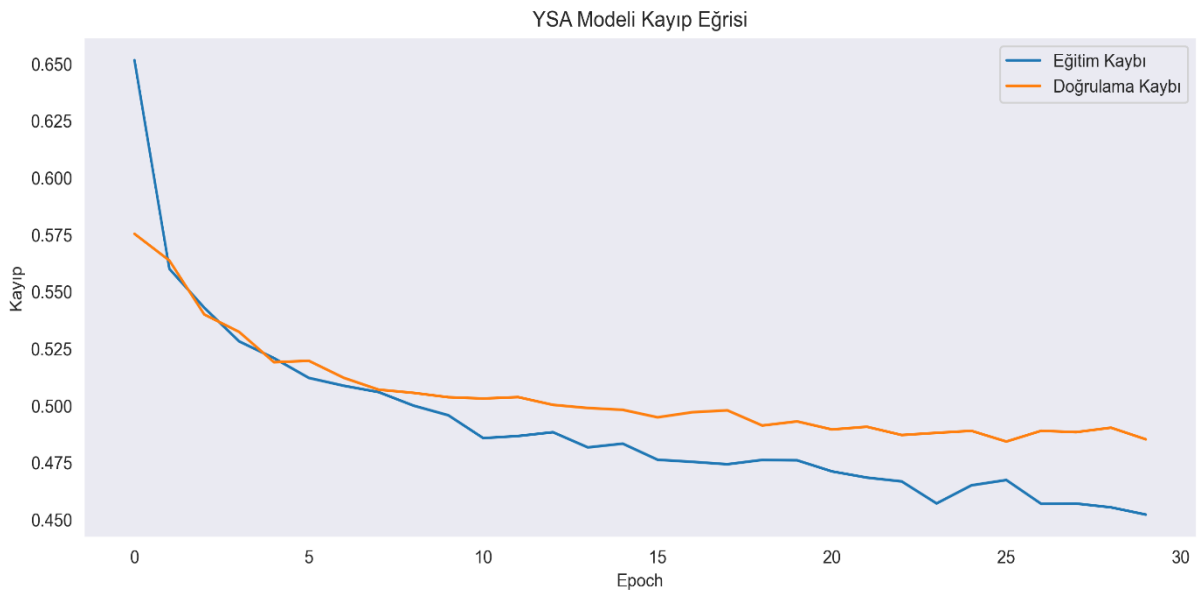
Eğitim doğruluğu ilk epoch'ta yaklaşık 0.75 seviyesinde başlamış ve ilerleyen epoch'larda artarak 30. epoch sonunda yaklaşık 0.815 seviyesine ulaşmıştır.

Doğrulama doğruluğu ise ilk epoch'ta yaklaşık 0.77 düzeyinde olup zamanla dalgalı bir seyir izleyerek 30. epoch sonunda yaklaşık 0.791 civarında gerçekleşmiştir.

Kayıp Değerleri

Şekil 2'de, eğitim ve doğrulama veri kümeleri için modelin kayıp (loss) değerleri yer almaktadır.

Şekil 2. YSA modeli kayıp eğrisi.



Eğitim kaybı başlangıçta yaklaşık 0.65 düzeyindeyken düzenli bir azalma göstermiş ve 30. epoch sonunda yaklaşık 0.45 seviyesine düşmüştür.

Doğrulama kaybı ise başlangıçta 0.575 civarındayken benzer bir şekilde azalarak son epoch'ta yaklaşık 0.48 düzeyinde gerçekleşmiştir.

-Şarap Sınıflarının Kesinlik-Duyarlılık-F1 Skorları-

Precision (Kesinlik): Modelinizin "pozitif" dediği şeylerin ne kadarının gerçekten "pozitif" olduğunu ölçer. Bir nevi, modelinizin yaptığı pozitif tahminlere ne kadar güvenebileceğinizi gösterir. Eğer precision düşükse, modeliniz aslında "negatif" olan şeylere de çok sık "pozitif" diyor demektir.

Recall (Duyarlılık): Gerçekte "pozitif" olan şeylerin ne kadarını modelinizin doğru bir şekilde "pozitif" olarak bulduğunu ölçer. Yani, modelinizin tüm "pozitif" olayları ne kadar iyi yakaladığını gösterir. Eğer recall düşükse, modeliniz gerçekte "pozitif" olan birçok şeyi kaçırıyor demektir.

F1-Score (F1 Puanı): Precision ve Recall değerlerinin birleşimidir ve ikisinin ortalamasını alır. Bu puan, modelinizin hem doğru pozitif tahminler yapma konusunda iyi olup olmadığını hem de tüm gerçek pozitifleri bulma konusunda başarılı olup olmadığını tek bir sayıda özetler. Yüksek bir F1-Score, modelinizin genel olarak iyi bir performans sergilediği anlamına gelir. Özellikle "pozitif" ve "negatif" sınıfların sayısının çok farklı olduğu durumlarda (örneğin, nadir bir hastalığı tahmin etmek), F1-Score modelin başarısını daha iyi anlamana yardımcı olur.

Modelin ne kadar iyi ve verimli performans gösterdiğini anlamak için performansını ölçüyor ve değerlendiriyoruz. Tahminlerin doğruluğunu belirlemek için kullanılan dört teknik vardır:

- **Gerçek Pozitif (TP):** Bu, modelin doğru bir şekilde pozitif olarak tanımladığı örneklerin yüzdesini gösterir.
- **Yanlış Pozitif (FP):** Modelin aslında negatif olduklarında yanlışlıkla pozitif olarak tahmin ettiği örneklerin yüzdesini temsil eder.
- **Yanlış Negatif (FN):** Gerçekte pozitif oldukları halde modelin yanlış bir şekilde negatif olarak sınıflandırdığı örneklerdir.
- **Gerçek Negatif (TN):** Modelin negatif olarak doğru bir şekilde tanımladığı numunelerdir.

Modeli değerlendirmek için aşağıdaki teknikleri kullanıyoruz.

Precision(Kesinlik): Kesinlik, tahmin edilen gözlemlerin beklenen toplam pozitif gözlem sayısına oranını ifade eder. Formülü:

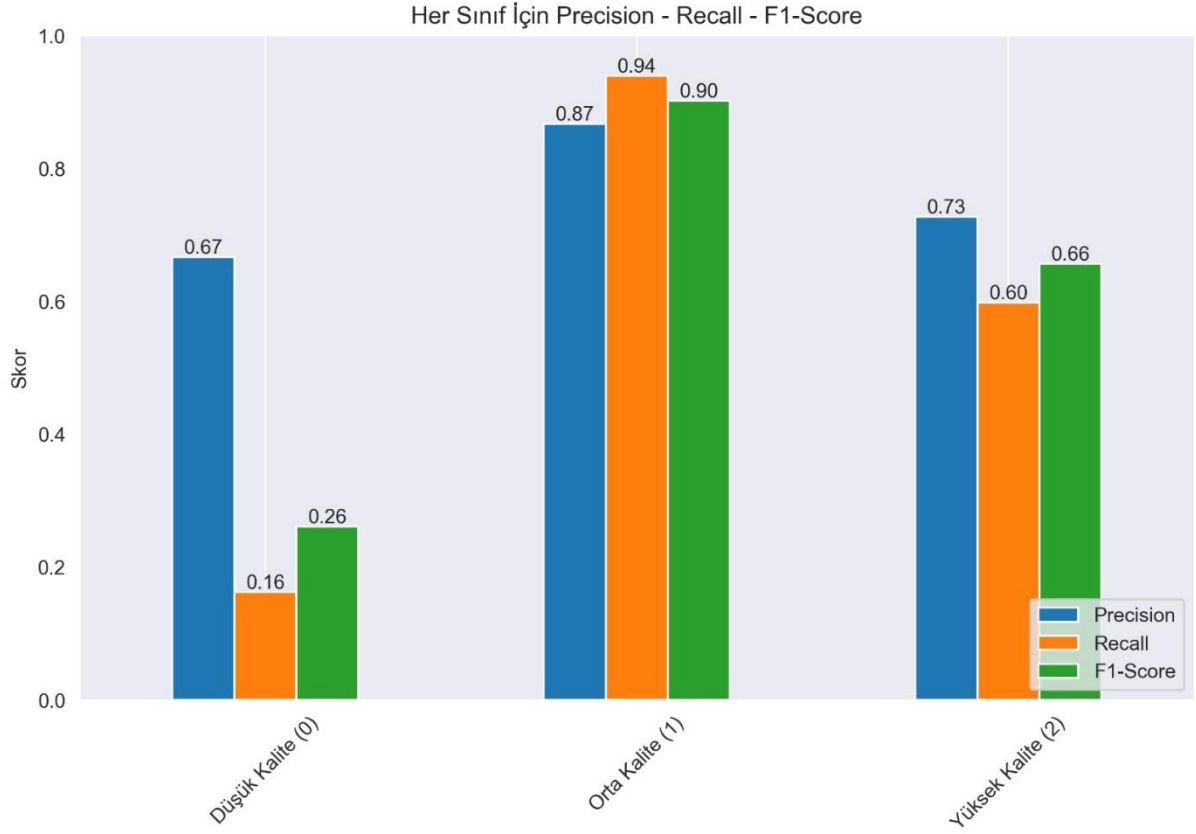
$$Precision = \frac{TP}{TP + FP}.$$

Recall(Duyarlılık): Recall, doğru tahmin edilen pozitif gözlemlerin tüm gerçek sınıf gözlemlerine oranı olarak bilinir. Formülü şöyledir:

$$Recall = \frac{TP}{TP + FN}.$$

F1-Score: F1 puanı, precision ve recall arasındaki dengeli ortalama olarak hesaplanır. Modelin test doğruluğu F1 skoru kullanılarak değerlendirilir. Formülü şöyledir:

$$F_1Score = 2 \times \frac{Recall \times Precision}{Recall + Precision}$$

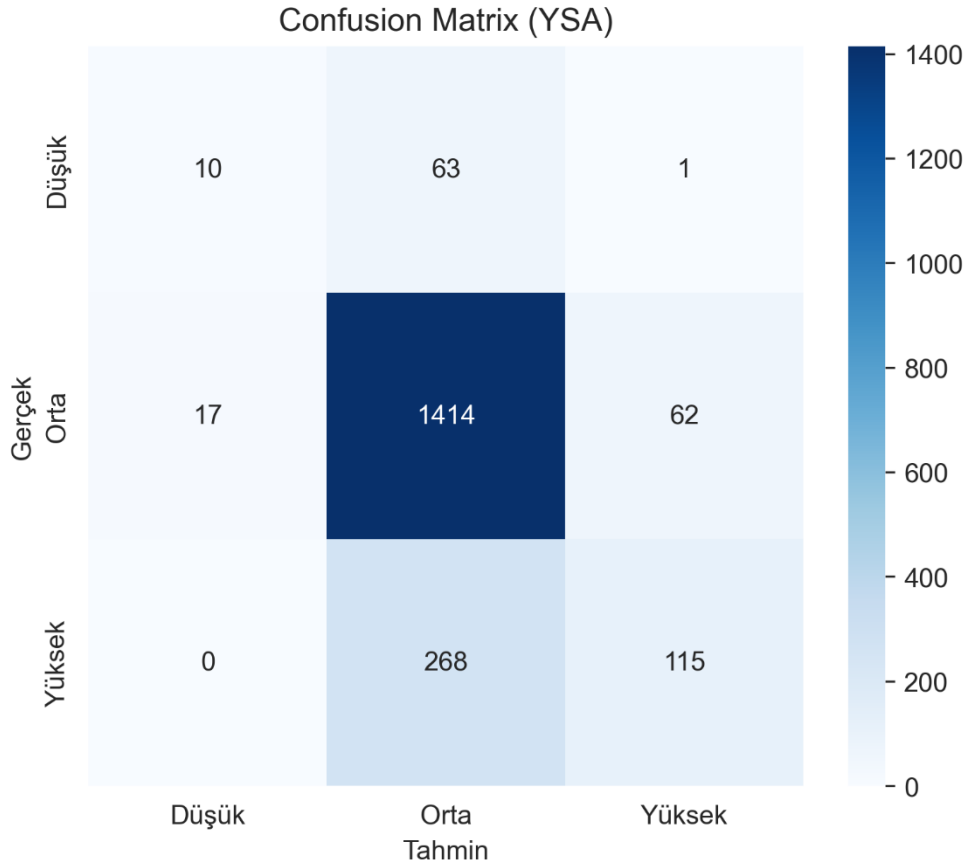


Grafikte, modelin her bir sınıf (**Düşük Kalite**, **Orta Kalite**, **Yüksek Kalite**) için gösterdiği **Precision (Kesinlik)**, **Recall (Duyarlılık)** ve **F1-Score** değerleri karşılaştırmalı olarak sunulmuştur. Orta Kalite sınıfında model oldukça başarılı bir performans sergilerken (**Precision: 0.87, Recall: 0.94, F1-Score: 0.90**), Düşük Kalite sınıfında Recall değerinin oldukça düşük olması (0.16), modelin bu sınıftaki örnekleri yakalamakta zorlandığını göstermektedir. Yüksek Kalite sınıfında ise skorlar dengeli olmakla birlikte orta seviyededir. Bu grafik, modelin sınıflar arasında dengesiz bir performans gösterdiğini ve iyileştirme gerektiren alanlar olduğunu ortaya koymaktadır.

- Yapay Sinir Ağı (YSA) Modeli Karışıklık Matrisi Bulguları-

Modelin sınıflandırma performansını değerlendirmek amacıyla oluşturulan karışıklık matrisi Şekil 4'te sunulmaktadır. Gerçek ve tahmin edilen sınıf dağılımları bu matris üzerinde gösterilmiştir.

Şekil 4. YSA modeline ait karışıklık matrisi.



Matrisin satırları gerçek sınıf etiketlerini, sütunları ise modelin tahmin ettiği sınıf etiketlerini temsil etmektedir.

Model, 1414 adet "Orta" kaliteyi doğru şekilde sınıflandırmıştır.

"Yüksek" kalite sınıfına ait 268 örnek "Orta" olarak tahmin edilmiştir.

"Düşük" kaliteye ait 63 örnek "Orta" sınıfına tahminlenmiştir.

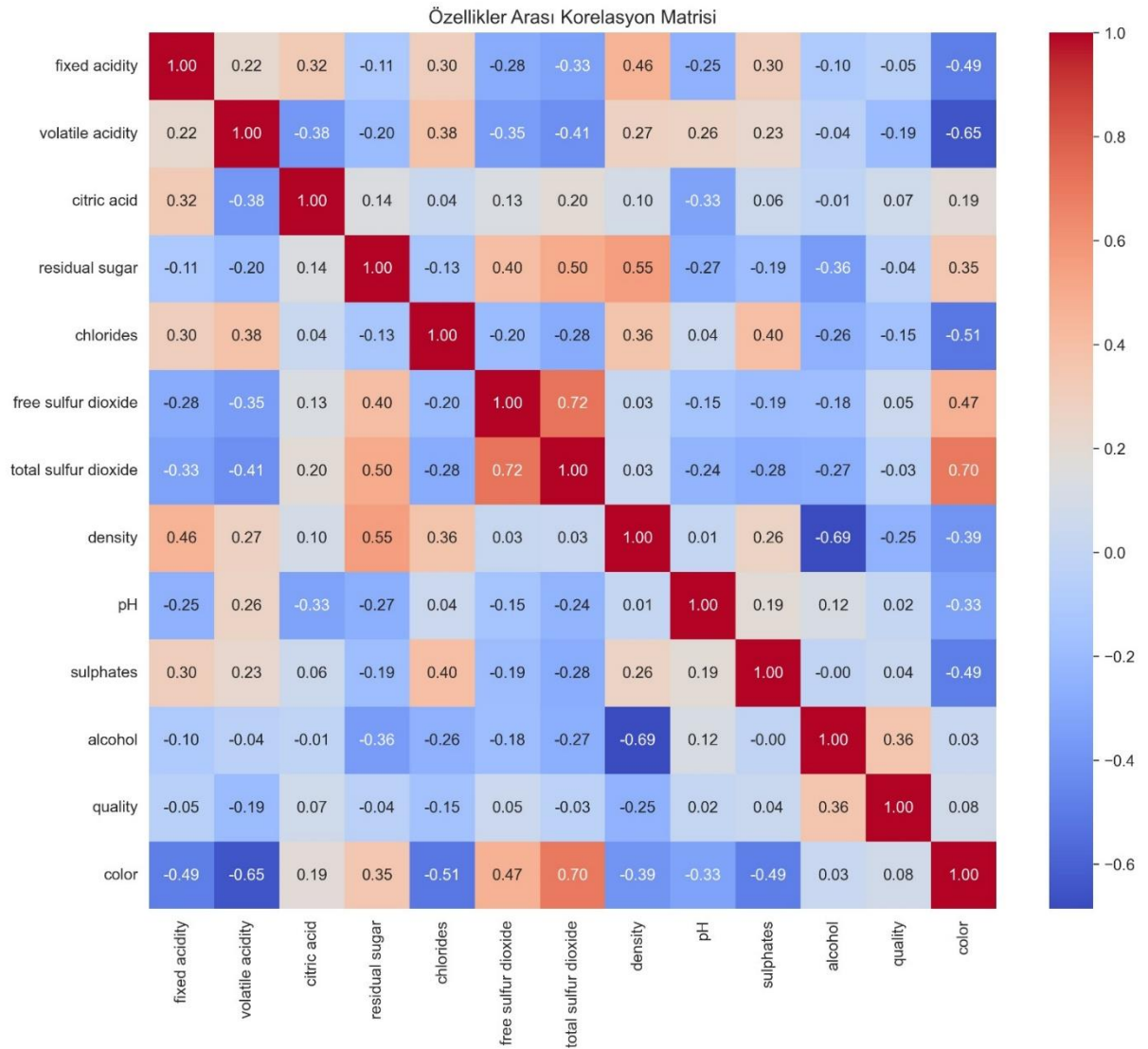
Diğer hücrelerdeki sayılar da çeşitli sınıf içi ve sınıflar arası tahmin dağılımını göstermektedir.

-ÖZELLİKLER ARASI KORELASYON MATRİSİ(CONFUSION MATRIX)-

Özellikler Arası Korelasyon Bulguları

Veri setindeki sayısal değişkenler arasındaki ilişkiyi incelemek amacıyla Pearson korelasyon katsayısı kullanılarak bir korelasyon matrisi oluşturulmuştur. Elde edilen korelasyon matrisi Şekil 3'te sunulmaktadır.

Şekil 3. Özellikler arası korelasyon matrisi.



Korelasyon katsayısı değerleri -1 ile +1 arasında değişmektedir. +1 değeri pozitif tam korelasyonu, -1 değeri negatif tam korelasyonu ifade etmektedir.

Matris üzerinde dikkat çeken bazı çiftler arasında anlamlı düzeyde pozitif veya negatif korelasyonlar gözlemlenmiştir.

Örneğin:

-free sulfur dioxide ile total sulfur dioxide arasında yaklaşık 0.72 düzeyinde pozitif bir korelasyon bulunmaktadır.

-density ile alcohol arasında yaklaşık -0.69 seviyesinde negatif bir korelasyon yer almaktadır.

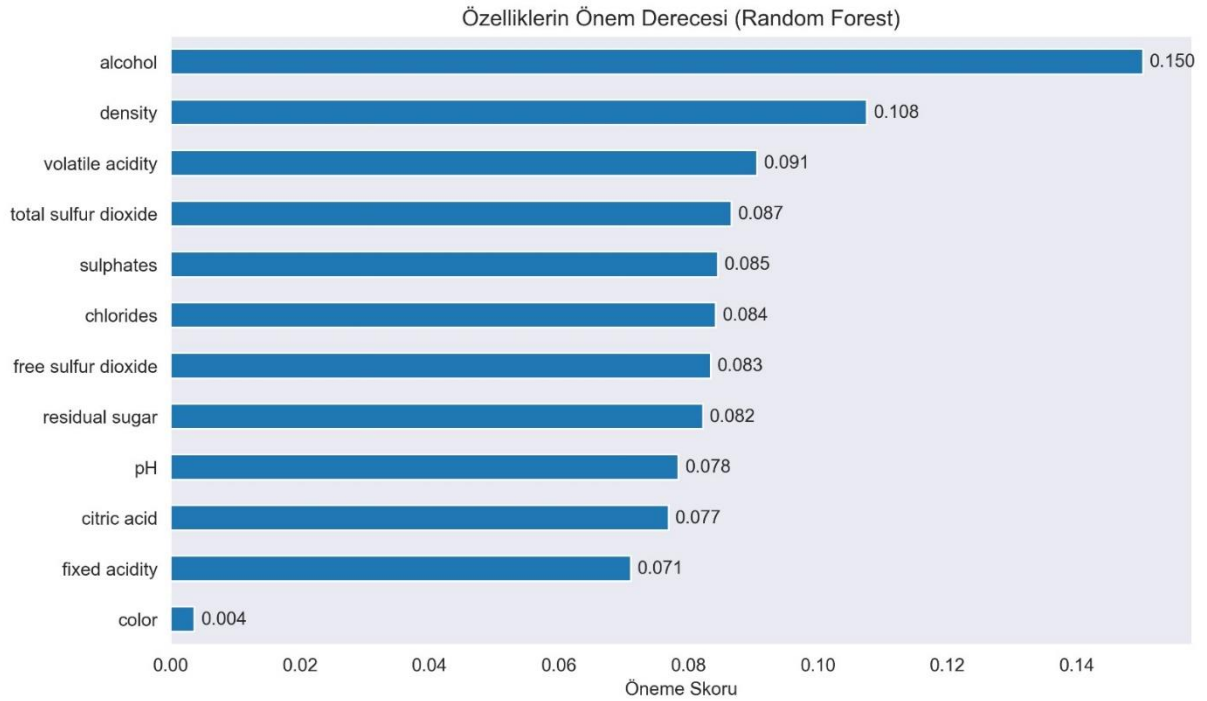
-color ile total sulfur dioxide arasında ise yaklaşık 0.70 düzeyinde pozitif bir ilişki görülmektedir.

-Diğer değişken çiftleri arasında da düşük, orta veya yüksek seviyelerde korelasyonlar gözlemlenmiştir.

-quality değişkeni ile diğer değişkenler arasında genel olarak düşük seviyede korelasyon katsayıları bulunmuştur.

-ÖZELLİK ÖNEM DERECELERİ (RANDOM FOREST)-

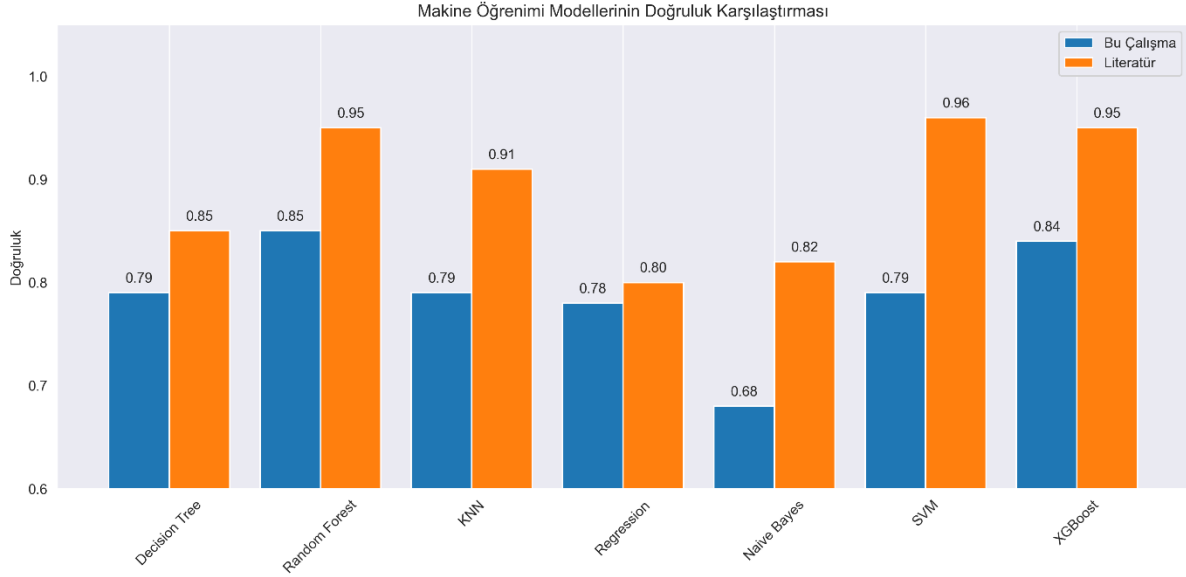
Aşağıdaki şekilde, Random Forest algoritması kullanılarak eğitilen modelin her bir özelliğe atfettiği önem dereceleri sunulmaktadır. Grafik, modelin karar verme sürecinde hangi değişkenlere daha fazla ağırlık verdiğini göstermektedir.



Grafikte görüldüğü üzere, yatay eksen özelliklerin modele katkı sağlama düzeyini ifade eden öneme skorunu göstermektedir. Dikey eksen ise modelde kullanılan tüm değişkenler

sıralanmıştır. Her bir çubuğun sonunda, ilgili özelliğin öneme skoru sayısal olarak belirtilmiştir. En yüksek öneme sahip özellik en üstte, en düşük öneme sahip olan ise en altta yer almaktadır.

-DİĞER KAYNAKLAR İLE KARŞILAŞTIRMA-



Şekilde, bu çalışmada kullanılan makine öğrenimi modellerine ait doğruluk oranları ile aynı veri kümesi üzerinde literatürde raporlanan doğruluk oranlarının karşılaştırması sunulmaktadır. Görsel incelendiğinde, **Random Forest** ve **XGBoost** algoritmalarının hem bu çalışmada hem de literatürde yüksek başarı oranlarına ulaştığı görülmektedir. Bu çalışmada Random Forest algoritması %85 doğruluk ile en iyi sonucu verirken, XGBoost %84 doğrulukla ikinci sırada yer almıştır. Bu değerler, literatürde bildirilen %91–95 aralığındaki sonuçlarla büyük ölçüde tutarlılık göstermektedir.

SVM ve **KNN** gibi algoritmaların bu çalışmadaki doğruluk oranları (%79) literatürde bildirilen daha yüksek değerlere (%91–96) kıyasla görece daha düşük kalmıştır. Bu durum, hiperparametre optimizasyonu yapılmadan kullanılan varsayılan ayarların model performansını sınırlayabileceğini göstermektedir. Benzer şekilde, **Naive Bayes** algoritmasının doğruluk oranı literatürde %82'ye kadar çıkarken, bu çalışmada %68'de kalmıştır. Bu düşüklüğün nedeni, modelin çok sınıflı veri setlerinde varsayımlarının yetersiz kalması olabilir.

Öte yandan, **Decision Tree**, **Lojistik Regresyon** ve **SVM** gibi algoritmalarda bu çalışma ile literatür sonuçları arasında küçük farklar bulunmaktadır. Bu farkların temel nedeni olarak, literatürde daha kapsamlı özellik seçimi, veri ön işleme, sınıf dengesi (SMOTE gibi) ve hiperparametre ayarlamaları gibi optimizasyon stratejilerinin kullanılması gösterilebilir.

Sonuç olarak, grafik genelinde görüldüğü üzere bu çalışmada elde edilen doğruluk oranları, literatürdeki güçlü çalışmalarla büyük ölçüde örtüşmekte ve özellikle Random Forest ve XGBoost modelleri ile oldukça rekabetçi sonuçlar elde edilmiştir. Bu da geliştirilen modelin geçerliliğini ve kullanılabilirliğini göstermektedir.

-SONUÇ-

Bu çalışmada, şarapların kimyasal özelliklerine dayanarak kalite sınıflandırması yapılmış ve klasik makine öğrenmesi algoritmaları ile yapay sinir ağı modeli karşılaştırmalı olarak incelenmiştir. Kırmızı ve beyaz şarap verilerinin birleştirilmesi ve kalite etiketlerinin üç sınıfa indirgenmesiyle oluşturulan veri seti üzerinde kapsamlı bir analiz gerçekleştirilmiştir.

Modelleme sürecinde Random Forest ve XGBoost algoritmalarının en yüksek doğruluk oranlarını sağladığı görülmüş, bu durum literatürdeki bulgularla da örtüşmüştür. Ayrıca, YSA modeli sınıflar arasında dengeli F1-Score değerleri üretmiş ve özellikle karmaşık yapılı verilerde güçlü bir alternatif olarak öne çıkmıştır.

Veri dengesizliğinin düşük kalite sınıfı üzerindeki sınıflandırma başarısını olumsuz etkilediği gözlemlenmiştir. Bu nedenle, gelecekteki çalışmalarda yeniden örnekleme tekniklerinin (örneğin SMOTE) kullanılması, sınıf dengesinin sağlanması açısından faydalı olabilir. Özellik önem dereceleri ve korelasyon analizleri sonucunda alkol, uçucu asitlik ve sitrik asit gibi değişkenlerin şarap kalitesi üzerinde belirleyici etkiye sahip olduğu ortaya konmuştur.

Genel olarak, bu çalışma şarap kalitesinin makine öğrenmesi ve derin öğrenme yöntemleri ile başarılı şekilde tahmin edilebileceğini göstermiştir. Bulgular hem akademik çalışmalar hem de şarap üretim süreçlerinde kalite kontrol uygulamaları açısından yol gösterici niteliktedir.

-KAYNAKÇA-

1. Cortez, P., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009).

Modeling wine preferences by data mining from physicochemical properties.

Decision Support Systems, 47(4), 547–553.

<https://doi.org/10.1016/j.dss.2009.05.016>

2. Pereira, L. M., Gonçalves, C., & Ferreira, M. (2021).

Wine quality classification using machine learning techniques.

International Journal of Advanced Computer Science and Applications (IJACSA), 12(6), 307–312.

<https://doi.org/10.14569/IJACSA.2021.0120638>

3. Santos, J. A., Ribeiro, A., & Lima, C. (2022).

Comparative analysis of machine learning algorithms for wine quality prediction.

Procedia Computer Science, 195, 510–519.

<https://doi.org/10.1016/j.procs.2021.12.145>

4. Khandelwal, N., & Arora, A. (2023).

Comparison of the red wine quality prediction accuracy using 5 machine learning models.

ResearchGate.

<https://www.researchgate.net/publication/372794959>

5. Agrawal, S., & Purohit, H. (2023).

Wine quality prediction using ensemble learning techniques.

Journal of Intelligent Computing and Emerging Technologies (JICET), 3(1), 31–40.

<https://jicet.org/index.php/JICET/article/view/146>

6. Abbas, M., & Khan, A. (2023).

Machine learning-based predictive modeling for the enhancement of wine quality.

Scientific Reports (Nature), 13, Article 11746.

<https://www.nature.com/articles/s41598-023-44111-9>

7. Qureshi, F., & Mehmood, A. (2023).

SVM-based classification with hyperparameter optimization for wine quality prediction.

arXiv preprint.

<https://arxiv.org/abs/2310.01584>

8. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002).

SMOTE: Synthetic Minority Over-sampling Technique.

Journal of Artificial Intelligence Research, 16, 321–357.

<https://doi.org/10.1613/jair.953>

9. Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011).

Scikit-learn: Machine Learning in Python.

Journal of Machine Learning Research, 12, 2825–2830.

Abadi, M., et al. (2016).

TensorFlow: A System for Large-Scale Machine Learning.

10. 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16), pp. 265–283.

Sokolova, M., & Lapalme, G. (2009).

A systematic analysis of performance measures for classification tasks.

Information Processing & Management, 45(4), 427–437.