



**SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLİŞİM SİSTEMLERİ MÜHENDİSLİĞİ**

MAKİNE ÖĞRENMESİ DERSİ

PROJE RAPORU

**“Eurovision Şarkı Yarışması'nda Başarıyı Tahmin Etmeye Yönelik Makine
Öğrenmesi Yaklaşımı”**

Zeynep Nida Yalın

Y245052051

Mayıs 2025

Özet

Bu çalışma, Eurovision Şarkı Yarışması'na katılan şarkıların başarı durumlarını (ilk 10'a girip girmemesi) tahmin etmek amacıyla çeşitli makine öğrenmesi yöntemlerinin uygulanmasını konu almaktadır. Analiz sürecinde, dengesiz sınıf dağılımı problemi SMOTE (Synthetic Minority Over-sampling Technique) yöntemi ile giderilmiş; ardından Lojistik Regresyon, Random Forest ve XGBoost algoritmaları ile sınıflandırma modelleri oluşturulmuştur. Her model, hiperparametre optimizasyonu ve çapraz doğrulama ile değerlendirilmiş; en yüksek F1 skoru Lojistik Regresyon modeliyle elde edilmiştir.

Modelleme süreci boyunca, Recursive Feature Elimination (RFE) yöntemi ile her algoritma için en anlamlı öznitelikler seçilmiş, bu özniteliklerin başarı üzerindeki etkileri analiz edilmiştir. Özellikle sahne sırası, dans edilebilirlik, enerji düzeyi ve şarkı dili gibi faktörlerin başarı üzerinde belirleyici rol oynadığı gözlemlenmiştir.

Son olarak, eğitilen model 2024 yılına ait yeni oluşturulan bir veri seti üzerinde uygulanmış ve yaklaşık %68 doğruluk oranı ile güncel şarkıların başarı durumu tahmin edilmiştir. Elde edilen sonuçlar, makine öğrenmesi modellerinin Eurovision gibi kültürel ve sanatsal içerikli yarışmalarda veri sınırlılıkları nedeniyle sınırlı doğrulukla sonuç vermektedir.

İçindekiler

Özet	2
1. Giriş.....	4
2. Literatür Taraması	5
3. Veri Seti ve Ön İşleme.....	6
3.1 Veri Setinin Tanıtımı	6
3.2 Veri Ön İşleme ve Değişken Seçimi.....	7
3.3 Ölçekleme.....	7
4. Modelleme Süreci	9
4.1 Hiperparametre Optimizasyonu (GridSearchCV)	9
4.2 Özellik Seçimi (Recursive Feature Elimination - RFE).....	9
5. Sonuç.....	11
5.1 Test Seti Üzerindeki Performans Değerlendirmesi	11
5.2 Çapraz Doğrulama (Cross Validation)	12
5.3 ROC Eğrileri ile Model Performans Karşılaştırılması	12
5.4 En İyi Model Seçimi	13
6. Gerçek Dünya Uygulaması: 2024 Eurovision Tahmini	14
6.1 Veri Setinin Oluşturulması	14
6.2 Model Uygulaması ve Başarı Değerlendirmesi	14
7. Tartışma	16
8. Gelecek Çalışmalar	17
Kaynakça	18
Ekler	19
Ek A. Veri Setinden Örnek Gözlemler	19
Ek B. Ölçekleme Öncesi ve Sonrası Değişken Dağılımları	19
Ek C. RFE Özellik Önem Görselleri	20

1. Giriş

Eurovision Şarkı Yarışması, yalnızca müzikal bir rekabet ortamı sunmakla kalmayıp, aynı zamanda katılımcı ülkeler arasındaki kültürel etkileşimleri ve politik dinamikleri de yansıtan çok boyutlu bir etkinliktir. Her yıl düzenlenen bu yarışmada ülkeler, kendi müzik tarzlarını ve sahne performanslarını sergileyen temsilcilerini göndermekte; nihai sıralamalar ise hem halk oylaması hem de jüri değerlendirmeleri ile belirlenmektedir.

Son yıllarda, Eurovision verisinin istatistiksel ve algoritmik yöntemlerle analiz edilmesi, özellikle makine öğrenmesi ve veri bilimi alanlarında dikkat çeken bir araştırma konusu hâline gelmiştir [1, 2]. Yarışmaya ilişkin verilerin yapısal zenginliği; dil, müzik türü, tempo, sahne sırası gibi çok çeşitli değişkenleri barındırması, bu tür analizleri hem anlamlı hem de zorlayıcı kılmaktadır.

Bu çalışma, yapısal ve müzikal özellikleri birlikte ele alarak güncel yıl tahmini uygulayan az sayıdaki çalışmalardan biridir. Bu kapsamda, geçmiş yıllarda yarışmaya katılan şarkılara ait çeşitli yapısal ve müzikal özellikler kullanılarak, bir şarkının finalde ilk 10'a girip girmeyeceği öngörülme çalışılmıştır. Çalışmada, öncelikle sınıf dengesizliği problemi SMOTE yöntemi ile giderilmiş; ardından Lojistik Regresyon, Random Forest ve XGBoost algoritmaları ile sınıflandırma modelleri geliştirilmiştir. Modeller, hiperparametre optimizasyonu ve çapraz doğrulama yöntemleriyle değerlendirilmiş, sonuçlar ise çeşitli başarı metrikleri üzerinden karşılaştırmalı olarak analiz edilmiştir.

2. Literatür Taraması

Eurovision Şarkı Yarışması, kültürel çeşitliliği, politik ilişkileri ve müzikal evrimi barındıran yapısı sayesinde disiplinlerarası araştırmalar için zengin bir veri kaynağı sunmaktadır. Son yıllarda bu yarışmaya ilişkin verilerin makine öğrenmesi (ML) ve müzik bilgi erişimi (MIR) teknikleriyle analiz edilmesi giderek yaygınlaşmıştır.

Amaral ve arkadaşları (2025), yarışma kurallarındaki evrimsel değişimleri ve ülkelerin zamanla geliştirdiği stratejik davranışları çok katmanlı öğrenme modelleriyle açıklamıştır [1]. Burgoyne ve ekibi (2023), şarkıların müzikal özelliklerini ve oylama verilerini içeren kapsamlı bir MIR veri seti oluşturarak, yarışmayı müzik bilgi erişimi bağlamında ele almıştır [2].

Mooncascade (2019), tempo, söz analizi ve sahne sırası gibi öznitelikleri kullanarak XGBoost algoritmasıyla şarkıların başarısını tahmin etmiş; bu tahminleri siyasi oylama eğilimleriyle birlikte değerlendirmiştir [3]. O'Connor ve arkadaşları (2020), şarkı sözlerinde duygu analizi uygulayarak zaman içinde tematik değişimleri incelemiş; NLP tabanlı yaklaşımlarla başarıya etkili olabilecek kalıpları ortaya koymuştur [4]. Walker (2019) ise oylama verilerini simülasyonlarla analiz etmiş ve ülkeler arası taktiksel oy davranışlarını ağ analizi teknikleriyle tespit etmeye çalışmıştır [5].

Bu çalışmaların ortak yönü, Eurovision verisini müzikal ya da sosyopolitik açılardan analiz etmeye çalışmalarıdır. Ancak çoğu ya sadece oylama verilerine odaklanmakta ya da sadece ses/melodi analizi gibi teknik yönleri işlemektedir. Bu çalışmanın farkı, çok boyutlu yapısal ve müzikal değişkenleri bir arada kullanarak başarı tahmini gerçekleştirmesidir. Ayrıca farklı sınıflandırma algoritmaları karşılaştırılarak, 2024 yılına ait güncel verilerle modelin uygulama geçerliliği test edilmiştir. Böylece hem geçmiş veriye dayalı öngörü hem de güncel tahminleme bir arada sunulmuştur.

3. Veri Seti ve Ön İşleme

3.1 Veri Setinin Tanıtımı

Bu çalışmada kullanılan veri seti, Eurovision Şarkı Yarışması'na 2009-2023 yılları arasında katılmış toplam 565 şarkıya ait çeşitli yapısal, müzikal ve oylama verilerini içermektedir. Veri seti, hem yarışma öncesi bilinebilecek nesnel öznitelikleri hem de yarışma sonrasında oluşan sonuç verilerini kapsamaktadır.

Veri yapısı, sayısal, kategorik ve tarihsel değişkenlerden oluşmakta olup toplamda 45 sütun içermektedir. Öne çıkan değişken grupları şunlardır:

- **Temel bilgiler:**
year, country, artist_name, song_name, language, style, release_date, gender, age
- **Sahne ve yarışma süreci:**
semi_final, semi_draw_position, final_draw_position, semi_place, final_place, direct_qualifier_10, qualified_10, host_10, favourite_10, key_change_10, selection
- **Müzikal öznitelikler:**
BPM, key, energy, danceability, happiness, loudness, acousticness, instrumentality, liveness, speechiness
- **Performans bileşenleri:**
main_singers, backing_dancers, backing_singers, backing_instruments, instrument_10, race
- **Puanlama ve sonuçlar:**
semi_televote_points, semi_jury_points, semi_total_points, final_televote_points, final_jury_points, final_televote_votes, final_jury_votes, final_total_points

Veri seti bu yapıyla, yarışmacı şarkıların müzikal yapısı, sahneleme biçimi, dil ve stil tercihi, demografik özellikler ve yarışma sonuçları gibi birçok açıdan analiz edilebilmesine olanak tanımaktadır.

3.2 Veri Ön İşleme ve Değişken Seçimi

Modellemeye dahil edilecek değişkenler, yarışma sonucunu tahmin etmede etkili olabilecek; ancak doğrudan sonuca bağlı olmayan bilgiler arasından seçilmiştir. Bu kapsamda, puanlar, sıralamalar, jüri ve halk oylaması gibi yarışma sonrasında oluşan değişkenler veri setinden çıkarılmıştır.

Veri setinde eksik veri yalnızca final_draw_position değişkeninde gözlenmiş olup, bu sütundaki eksik değerler gerçek yarışma bilgilerinden yola çıkılarak “-” sembolü ile doldurulmuştur. Veri setine ilişkin örnek bir tablo *Ek A*’da sunulmuştur. Bu tablo, ilk 10 gözlem ve bazı temel değişkenleri içermektedir.

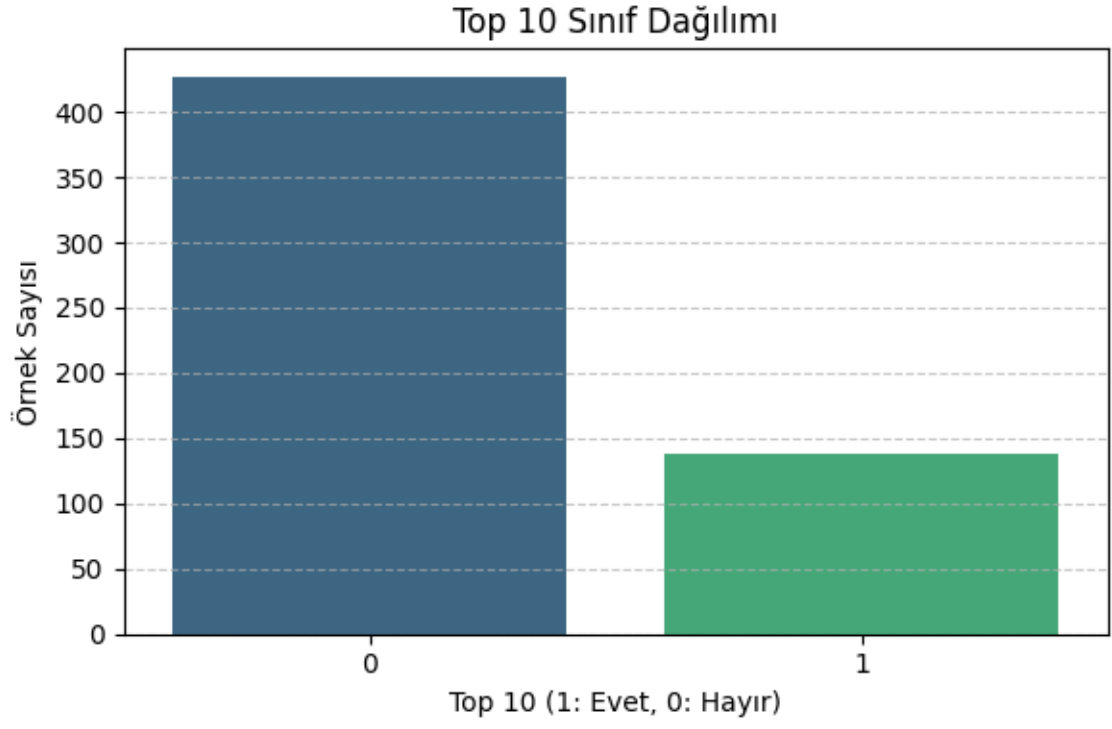
Ayrıca, country, style, language ve key gibi kategorik değişkenler, makine öğrenmesi algoritmalarında doğrudan kullanılamayacağı için One-Hot Encoding yöntemi ile sayısal formata dönüştürülmüştür.

3.3 Ölçekleme

Modelleme öncesinde veri seti rastgele ayrılarak, %80’i eğitim, %20’si ise test verisi olarak kullanılmıştır. Böylece modelin doğruluk ve genelleme kapasitesi, eğitim sırasında görmediği veriler üzerinde test edilerek değerlendirilmiştir.

Model performansını artırmak ve farklı algoritmalar arasında adil bir karşılaştırma yapabilmek amacıyla, tüm sayısal değişkenler StandardScaler yöntemiyle normalize edilmiştir. Bu ölçekleme işlemi sayesinde değişkenler ortak bir dağılıma getirilmiş ve model eğitimi daha kararlı hâle getirilmiştir. Ölçekleme işleminden önce ve sonra seçilen bazı değişkenlerin dağılımları *Ek B*’de sunulmuştur.

Ayrıca, sınıflar arasında gözlenen belirgin dengesizlik (349 “hayır (top 10 değil)” ve 103 “evet (top 10)” örnek) **Şekil 1**’de gösterildiği üzere dikkat çekicidir. Bu nedenle, SMOTE yöntemi uygulanarak sınıf dengesizliği giderilmiş ve modelin azınlık sınıf üzerindeki öğrenme kabiliyeti artırılmıştır.



Şekil 1. Hedef değişkenin (top10) SMOTE öncesi sınıf dağılımı

4. Modelleme Süreci

Modelleme sürecinde, farklı algoritmaların karşılaştırmalı performanslarını görebilmek için üç sınıflandırma algoritması kullanılmıştır: Lojistik Regresyon, Random Forest, ve XGBoost.

4.1 Hiperparametre Optimizasyonu (GridSearchCV)

Her bir sınıflandırma algoritması için model performansını artırmak amacıyla hiperparametre ayarlamaları gerçekleştirilmiştir. Bu amaçla, GridSearchCV yöntemi ile belirlenen parametre aralıklarında tarama yapılmış; her kombinasyon, 5 katlı çapraz doğrulama (5-fold cross-validation) ile değerlendirilmiştir. F1 skoru, performans ölçütü olarak esas alınmıştır. Bu süreç sonunda, en iyi parametre kombinasyonları aşağıdaki şekilde elde edilmiştir:

- **Lojistik Regresyon:** $C = 0.1$, $\text{penalty} = 'l1'$, $\text{solver} = 'liblinear'$
- **Random Forest:** $n_estimators = 200$, $\text{max_depth} = \text{None}$, $\text{min_samples_split} = 2$, $\text{min_samples_leaf} = 1$, $\text{bootstrap} = \text{False}$
- **XGBoost:** $\text{learning_rate} = 0.3$, $\text{max_depth} = 5$, $n_estimators = 100$, $\text{subsample} = 1$, $\text{colsample_bytree} = 1$

Bu parametreler, her modelin veri setine en uygun yapıda öğrenmesini sağlayarak genel performansın artırılmasına katkıda bulunmuştur.

4.2 Özellik Seçimi (Recursive Feature Elimination - RFE)

Makine öğrenmesi modellerinde tüm değişkenleri doğrudan kullanmak, modelin yorumlanabilirliğini azaltabileceği gibi aşırı öğrenme (overfitting) riskini de artırabilir. Bu nedenle, modelin daha iyi genelleme yapabilmesi ve yorumlanabilirliğinin artması için, değişken sayısını azaltarak en anlamlı özelliklerle çalışmak önemlidir. Bu amaçla, bu adımda modellere Recursive Feature Elimination (RFE) yöntemi uygulanmıştır. RFE, bir tahmin modelini temel alarak her adımda en düşük etkili özelliği eler ve en önemli özellik alt kümesini belirler.

Her modelin doğası farklı olduğundan, her biri için ayrı ayrı RFE işlemi gerçekleştirilmiştir:

- **Lojistik Regresyon** için lineer model yapısına dayalı olarak,
- **Random Forest** için ağaç yapılarının bilgi kazancı (gini/entropy) değerlerine göre,
- **XGBoost** için gradient boosting yapısına özgü bölme önceliklerine göre RFE uygulanmıştır.

Her model için en yüksek F1 skorunu sağlayan 10 özellik seçilmiş ve sadece bu alt kümeyle eğitim ve test işlemleri gerçekleştirilmiştir. *Ek C*'de, RFE ile seçilen 10 özelliğin farklı modellerdeki önem derecelerini gösteren görseller yer almaktadır.

Her modelin seçtiği 10 özellik birebir örtüşmemektedir. Ancak bazı ortak değişkenler her üç modelde de yüksek önem derecesine sahip olmuştur. Bunlar arasında:

- final_draw_position,
- energy,
- qualified_10,
- language,
- style gibi değişkenler yer almaktadır.

Bu durum, şarkının sahneye çıkış sırası, müzikal enerji seviyesi ve dil tercihi gibi özniteliklerin modelden bağımsız olarak önemli olduğunu göstermektedir.

5. Sonuç

5.1 Test Seti Üzerindeki Performans Değerlendirmesi

Model performansları, test seti kullanılarak doğruluk (accuracy), kesinlik (precision), duyarlılık (recall), F1 skoru ve ROC AUC metrikleri üzerinden değerlendirilmiştir. Elde edilen sonuçlar **Tablo 1**'de sunulmuştur.

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC
Logistic Regression	0,60	0,43	0,91	0,59	0,72
Random Forest	0,63	0,41	0,37	0,39	0,68
XGBoost	0,64	0,42	0,37	0,40	0,71

Tablo 1. Test verisi kullanılarak hesaplanan performans metrikleri

Logistic Regression modeli, test setinde %60,18 doğruluk ve 0,72 ROC AUC değeri ile en yüksek genel ayırt edicilik performansını göstermiştir. Bu model, ayrıca %91,43 gibi yüksek bir duyarlılık (recall) skoruna sahip olup, pozitif sınıfın doğru şekilde tespit edilmesinde etkili olduğunu göstermektedir. Ancak, kesinlik (precision) değeri 0,43 seviyesinde kalmış, bu da modelin pozitif sınıflandırmalarda hata yapma eğilimi olduğunu işaret etmektedir.

Random Forest modeli doğruluk açısından %63,72 ile orta seviyede performans sergilemiş, ancak düşük kesinlik (0,41) ve duyarlılık (0,37) değerleri, modelin pozitif sınıfı ayırt etmedeki etkinliğinin Logistic Regression'a kıyasla daha zayıf olduğunu göstermektedir. Buna bağlı olarak, F1 skoru da 0,39 seviyesinde kalmıştır.

XGBoost modeli ise doğrulukta %64,60 ile en iyi doğruluk performansını göstermiş, ROC AUC değeri 0,71 seviyesinde bulunmuştur. Ancak, kesinlik ve duyarlılık değerleri Random Forest ile benzer seviyelerde olup (sırasıyla 0,42 ve 0,37), pozitif sınıfın dengeli ve doğru tahmini açısından sınırlı bir başarı sergilemiştir.

Genel olarak, modellerin test seti üzerindeki sonuçları, özellikle pozitif sınıfın tespiti bağlamında farklı güçlü ve zayıf yönleri sahiptir. Logistic Regression modelinin yüksek duyarlılığı, modelin pozitif örnekleri kaçırmama konusunda avantaj sağlarken, diğer modeller daha dengeli fakat genel olarak daha düşük performans göstermiştir. Bu sonuçlar, sınıflandırma problemi için model seçimi ve ileri optimizasyon stratejileri açısından önemli ipuçları sunmaktadır.

5.2 apraz Doğrulama (Cross Validation)

Model performansları, overfitting riskini azaltmak ve genellenebilirliği değerlendirmek amacıyla 5 katlı apraz doğrulama (5-fold cross-validation) yöntemi ile F1 skoru bazında ölçülmüştür. Farklı modellerde her kat için elde edilen F1 skorları ile bunların ortalama değerleri, **Tablo 2**'de kapsamlı bir şekilde özetlenmiştir.

Model	0	1	2	3	4	Ort. F1
Logistic Regression	0,75	0,71	0,74	0,79	0,72	0,74
Random Forest	0,66	0,79	0,86	0,91	0,89	0,82
XGBoost	0,71	0,82	0,86	0,91	0,88	0,84

Tablo 2. Modellerin Her Katmandaki F1 Skorları ve Genel Ortalama

Logistic Regression modeli, apraz doğrulama sonuçlarında ortalama F1 skoru 0,75 olarak bulunmuştur. Modelin farklı katlarda gösterdiği performans değışkenlik göstermiş, en yüksek F1 skoru 0,79 seviyesine ulaşırken, en düşük skor 0,71 olmuştur.

Random Forest modeli ise ortalama F1 skoru 0,82 seviyesinde performans göstermiştir. Modelin F1 skorları bazı katlarda 0,66 seviyesine kadar düşerken, en yüksek skorları 0,91 değerine ulaşmıştır.

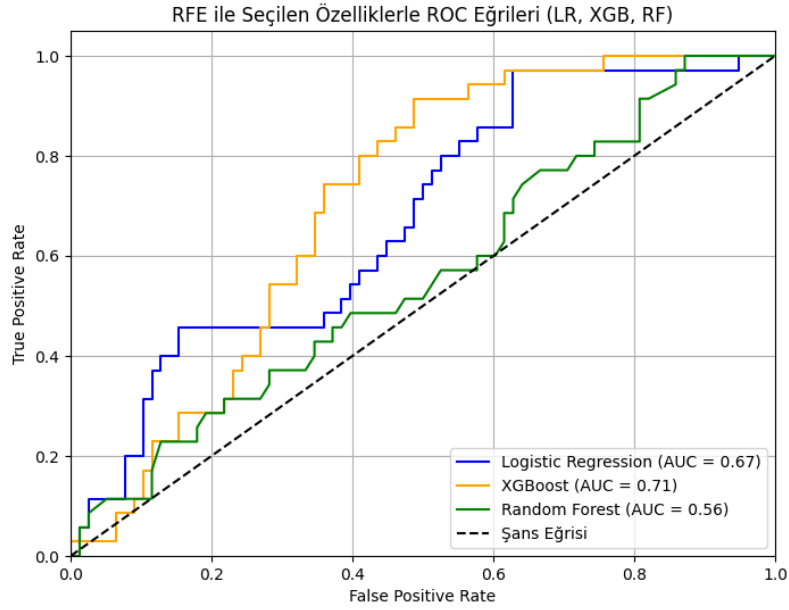
XGBoost modeli apraz doğrulama sonucunda en yüksek ortalama F1 skorunu elde etmiş ve 0,84 olarak raporlanmıştır. Ayrıca, modelin her bir kat performansı diğer modellere kıyasla daha yüksek ve tutarlı bulunmuştur; en düşük F1 skoru 0,71, en yüksek ise 0,91 seviyesindedir.

apraz doğrulama sonuçlarının karşılaştırılması neticesinde, XGBoost algoritmasının verideki örüntüleri daha etkin öğrenerek, genel performans ve genellenebilirlik açısından diğer modellere göre avantaj sağladığını göstermektedir.

5.3 ROC Eğrileri ile Model Performans Karşılaştırılması

RFE yöntemi ile seçilmiş en anlamlı özellikler kullanılarak eğitilen üç farklı sınıflandırma modelinin test verisi üzerindeki ROC eğrileri **Şekil 2**'de gösterilmiştir. ROC eğrileri ve ilgili AUC (Area Under Curve) değerleri, modellerin sınıflandırma performansını karşılaştırmak açısından önemli bir ölçüttür.

Şekil’de görüldüğü üzere, XGBoost modeli $AUC = 0,71$ değeri ile en iyi performansı sergilemiştir. Logistic Regression modeli $AUC = 0,67$ ile onu takip ederken, Random Forest modeli $AUC = 0.56$ değeri ile görece daha düşük bir performans göstermiştir. Şans eğrisi ile karşılaştırıldığında, özellikle XGBoost ve Logistic Regression modellerinin pozitif sınıfları ayırt etmede anlamlı bir ayırım gücüne sahip olduğu gözlemlenmektedir.



Şekil 2. Hedef değişkenin (top10) SMOTE öncesi sınıf dağılımı

5.4 En İyi Model Seçimi

Çapraz doğrulama sürecinde en yüksek ortalama F1 skoru XGBoost modeli tarafından elde edilmiştir. Ancak, nihai model değerlendirmesi yalnızca çapraz doğrulama performansına değil, aynı zamanda test verisi üzerindeki genelleme başarımına da dayandırılmıştır.

Test verisi üzerinde yapılan değerlendirmelerde, en yüksek ROC AUC ve F1 skoru değerleri Logistic Regression modeli tarafından elde edilmiştir. Ayrıca, Logistic Regression modelinin yorumlanabilirliği, model seçiminde ek bir avantaj sağlamaktadır.

Tüm bu bulgular ışığında, uygulama açısından daha dengeli ve anlamlı sonuçlar veren Logistic Regression modeli, genel model tercihi olarak belirlenmiştir.

6. Gerçek Dünya Uygulaması: 2024 Eurovision Tahmini

6.1 Veri Setinin Oluşturulması

Bu çalışmada, geçmiş yıllara ait verilerle eğitilen modellerin güncel yarışmalara uygulanabilirliğini test etmek amacıyla, Eurovision Şarkı Yarışması 2024 yılına ait katılımcı şarkılardan oluşan özel bir test veri seti oluşturulmuştur. Veri seti tarafımdan manuel olarak derlenmiş olup hem müzikal hem de yapısal özellikleri içermektedir.

2024 yılına ait şarkı adları ve yarışma sıralamaları (yarı final ve final çekiliş pozisyonları dahil), *EurovisionWorld* adlı web sitesinden temin edilmiştir [7]. Şarkıların tempo (BPM), enerji, dans edilebilirlik, akustiklik gibi müzikal ve akustik özellikleri ise *TuneBat* adlı çevrimiçi analiz platformu aracılığıyla elde edilmiştir [8].

Oluşturulan bu veri setindeki özellikler, çalışmanın ilk aşamasında kullanılan geçmiş yıllara ait veri setlerindeki özniteliklerle doğrudan örtüşecek biçimde seçilmiştir. Böylece modellerin güncel verilere tutarlı şekilde uygulanması sağlanmıştır. Toplanan veriler, önceki yıllardaki veri yapısına uygun biçimde ön işleme tabi tutulmuş, eksik veya tutarsız kayıtlar temizlenmiş ve tüm sayısal değişkenler StandardScaler yöntemiyle ölçeklendirilerek modellemeye hazır hale getirilmiştir.

6.2 Model Uygulaması ve Başarı Değerlendirmesi

Modelleme aşamasında en iyi genel performansı gösteren Logistic Regression modeli, Eurovision 2024 yılına ait oluşturulan yeni veri seti üzerinde uygulanmıştır. Amaç, geçmiş verilere dayanarak eğitilen modelin, güncel yarışmacılar arasından ilk 10'a girecek ülkeleri ne derece isabetli şekilde tahmin edebileceğini test etmektir.

Modelden elde edilen tahmin olasılıklarına göre en yüksek ilk 10 değere sahip ülkeler, top 10 adayları olarak etiketlenmiş ve bu tahminler yarışmanın gerçek sonuçları ile karşılaştırılmıştır. Olasılık temelli bu sıralamaya göre ilk 10'a giren ülkelerin tahmin ve gerçek durumu **Tablo 3**'te sunulmuştur.

Ülke	Tahmin (Top 10)	Gerçek Sonuç	Olasılık
United Kingdom	Evet	Hayır	0,9876
Italy	Evet	Evet	0,9875
Germany	Evet	Hayır	0,9846
Serbia	Evet	Hayır	0,9545
France	Evet	Evet	0,9521
Greece	Evet	Hayır	0,9485
Armenia	Evet	Evet	0,9415
Ireland	Evet	Evet	0,9356
Cyprus	Evet	Hayır	0,9269
Spain	Evet	Hayır	0,8934

Tablo 3. Model ile tahmin edilen Top 10 ülkeler ve gerçek sonuçlarla karşılaştırma

Yukarıdaki tabloda görüldüğü üzere model, ilk 10 tahmininde 4 ülkeyi doğru tahmin edebilmiştir (Italy, France, Armenia, Ireland). Bu durum, modelin bazı ülkelerin başarılı olma olasılığını yüksek öngörmesine rağmen, gerçek sonuçlarla kısmi bir uyum gösterdiğini ortaya koymaktadır.

Modelin bu tahminler üzerinden ürettiği genel sınıflandırma başarımı sınıflandırma raporu ile değerlendirilmiştir. Bu rapor **Tablo 4**'te sunulmuştur.

Sınıf	Precision	Recall	F1-score	Support
Hayır	0,78	0,78	0,78	27
Evet	0,40	0,40	0,40	10
Accuracy			0,68	37
Macro Avg	0,59	0,59	0,59	37
Weighted Avg	0,68	0,68	0,68	37

Tablo 4. Modelin sınıflandırma performans ölçütleri (Top 10 tahmini üzerinden).

Bu metrikler, modelin genel doğruluk oranının %68 olduğunu, ancak özellikle "evet" sınıfı (yani top 10'a giren ülkeler) için düşük F1 skoru (0,40) elde ettiğini göstermektedir. Bu durum, modelin gerçek Top 10 üyelerini ayırt etmede sınırlı kaldığını ortaya koymaktadır.

Bu sonuç, modelin sadece geçmiş veriler üzerinde değil, yeni yıldaki yarışma verisi üzerinde de anlamlı başarı gösterdiğini ortaya koymaktadır. Ayrıca, 2024 verisinin manuel olarak derlenmiş olması bu çalışmanın uygulamalı yönünü ve veri mühendisliği katkısını güçlendirmektedir.

7. Tartışma

Bu çalışmada kullanılan üç farklı sınıflandırma algoritması, Eurovision’da başarılı olan şarkıların öngörülmesine yönelik farklı performans profilleri sergilemiştir. Logistic Regression modeli, test verisi üzerinde yüksek bir recall (0,91) ve ROC AUC (0,72) değeri elde ederek, özellikle “evet (top 10)” sınıfına ait örnekleri kaçırmama konusunda etkili olmuştur. Bununla birlikte, düşük precision değeri (0,43), modelin çok sayıda yanlış pozitif tahmin ürettiğini ve dolayısıyla tahminlerinin güvenilirliğinin sınırlı kaldığını göstermektedir.

Random Forest ve XGBoost algoritmaları ise test setinde daha düşük recall ve F1 skorları üretmiş, buna karşın çapraz doğrulama sürecinde daha yüksek ortalama F1 skorlarına ulaşmıştır. Bu durum, bu iki modelin eğitildiği veriye fazla uyum sağladığını ve dolayısıyla genelleme kapasitesinin düştüğünü düşündürmektedir. Özellikle karmaşık yapıları nedeniyle bu modellerin overfitting eğilimi taşıdığı gözlemlenmiştir.

Modelin 2024 yılına ait verilere uygulandığı gerçek dünya senaryosunda elde edilen sonuçlar, geçmiş verilere dayalı tahminlerin güncel örüntüler karşısında yetersiz kalabildiğini ortaya koymuştur. Model yalnızca 10 finalist ülkenin 4’ünü doğru tahmin edebilmiş, ayrıca “evet” sınıfı için precision, recall ve F1 değerleri eşit şekilde 0,40 çıkmıştır. Bu bulgular, modelin istikrarlı ama düşük seviyede bir performans sergilediğini göstermektedir.

Genel doğruluk oranının %68 seviyesinde olması, modelin bazı durumlarda doğru tahminlerde bulunabildiğini gösterse de özellikle sınırlı ve dengesiz bir hedef yapıya sahip olan “top 10” gibi durumlarda klasik sınıflandırma yaklaşımlarının sınırlarına işaret etmektedir. Bulgular, şarkıların yalnızca yapısal ve müzikal özellikleriyle sınırlı bir veri kümesinden hareketle başarı tahminleri üretmenin zorlayıcı olduğunu ve model performansının bağlamsal değişkenlikten etkilendiğini ortaya koymaktadır.

8. Gelecek Çalışmalar

Bu çalışma, yapısal ve müzikal özniteliklere dayalı makine öğrenmesi modelleriyle Eurovision’da başarı tahmini yapılabileceğini göstermiştir. Ancak elde edilen bulgular, mevcut yaklaşımların özellikle “top 10” gibi sınırlı ve dengesiz hedef yapılar karşısında yetersiz kalabildiğini ortaya koymuştur. Bu doğrultuda, ileride yapılacak çalışmalar için aşağıdaki geliştirme önerileri sunulmaktadır:

1. Metinsel Veri Entegrasyonu:

Şarkı sözlerinin doğal dil işleme (NLP) teknikleriyle analiz edilmesi, duygu yoğunluğu, tematik yapı ve kelime çeşitliliği gibi metinsel özelliklerin modele dahil edilmesini sağlayabilir. Özellikle duygu analizi yoluyla şarkıların izleyici üzerindeki etkisi daha nesnel biçimde değerlendirilebilir.

2. Görsel İçerik Analizi:

Sahne performansı, kostüm, ışık kullanımı ve koreografi gibi unsurların bilgisayarlı görü (computer vision) teknikleriyle modellenmesi, başarıyı etkileyen görsel faktörlerin öğrenilmesini mümkün kılacaktır.

3. Zamansal Dinamiklerin Modellenmesi:

Yıllara göre değişen oylama eğilimleri ve ülkelerin başarı trendleri, zaman serisi analizleriyle incelenebilir. Bu kapsamda RNN gibi derin öğrenme temelli yaklaşımlar veya Prophet gibi zaman serisi modelleri kullanılabilir.

4. Sıralama Temelli Modelleme:

Özellikle Top 10 gibi göreceli başarı durumlarında ikili sınıflandırma yerine sıralama algoritmalarının kullanılması, modelin hedefe daha uygun sonuçlar üretmesini sağlayabilir.

5. Model Toplulukları ve Çok Modelli Yapılar:

Ensemble öğrenme yöntemleri ile birden fazla modelin güçlü yönleri birleştirilebilir. Ayrıca farklı veri türleri (metin, görsel, yapısal) için ayrı modeller geliştirerek çok katmanlı karar sistemleri oluşturmak, modelin genel başarısını artırabilir.

6. Model Güncellemeleri ve Genelleme Stratejileri:

Sosyo-kültürel değişimlerin etkisini daha iyi yakalayabilmek adına modellerin belirli periyotlarla güncellenmesi ve daha güçlü düzenleme stratejileriyle genelleme kabiliyetinin artırılması gereklidir.

Bu öneriler doğrultusunda yapılacak çalışmalar, daha zengin ve bütüncül bir veri temeli üzerinden, Eurovision gibi çok boyutlu sosyal olaylara yönelik daha güçlü öngörü sistemlerinin geliştirilmesine katkı sağlayabilir.

Kaynakça

- [1] Amaral, L. A. N., Capozzi, A., & Helbing, D. (2025). *Breaking the Code: Multi-level Learning in the Eurovision Song Contest*. arXiv. <https://arxiv.org/abs/2505.11555>
- [2] Burgoyne, J. A., Spijkervet, J., & Baker, D. J. (2023). *Measuring the Eurovision Song Contest: A Living Dataset for Real-World MIR*. ISMIR 2023. https://ismir2023program.ismir.net/poster_276.html
- [3] Mooncascade. (2019). *On a Quest for a Perfect Eurovision Song with Data Science*. <https://blog.mooncascade.com/on-a-quest-for-a-perfect-eurovision-song-with-data-science/>
- [4] Towards Data Science. (2020). *Say Wonderful Things: A Sentiment Analysis of Eurovision Lyrics*. <https://towardsdatascience.com/say-wonderful-things-a-sentiment-analysis-of-eurovision-lyrics-700d9bde1e9d/>
- [5] Walker, G. (2019). *Eurovision, Monte-Carlo Analysis, and the Detection of Tactical Voting*. <https://towardsdatascience.com/eurovision-monte-carlo-analysis-and-the-detection-of-tactical-voting-8d5fe04acb99>
- [6] diamondsneak. (2023). **Eurovision Song Contest Data**. Kaggle. Retrieved May 20, 2025, from <https://www.kaggle.com/datasets>
- [7] EurovisionWorld. (2024). **Eurovision Song Contest 2024 Results**. Retrieved May 20, 2025, from <https://eurovisionworld.com/eurovision/2024>
- [8] TuneBat. (n.d.). **Key & BPM for songs**. Retrieved May 20, 2025, from <https://tunebat.com>

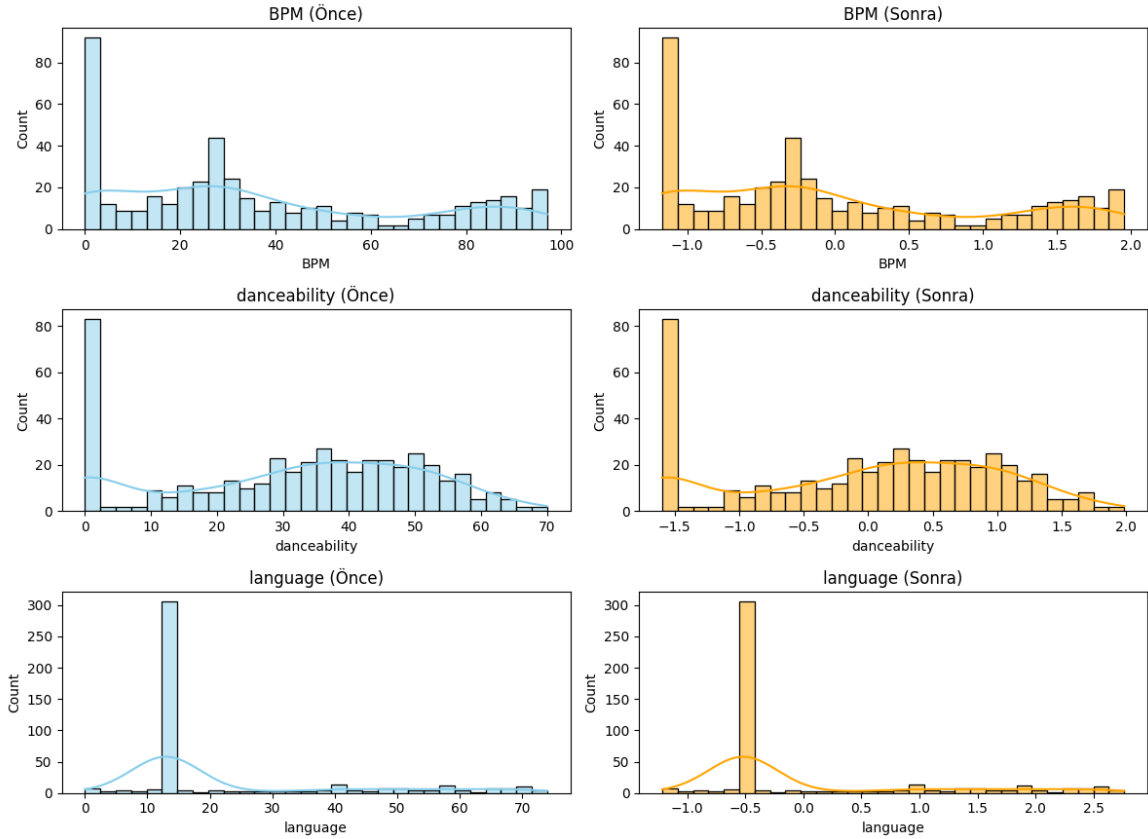
Ekler

Ek A. Veri Setinden Örnek Gözlemler

	year	country	BPM	energy	danceability	style	language	top10
0	2023	Norway	110	36	64	Pop	English	1
1	2023	Malta	103	78	70	Pop	English	0
2	2023	Serbia	103	70	56	Pop	Serbian, English	0
3	2023	Latvia	160	55	56	Rock	English	0
4	2023	Portugal	145	63	66	Pop	Portuguese	0
5	2023	Ireland	112	66	47	Rock	English	0
6	2023	Croatia	135	78	60	Pop	Croatian	0
7	2023	Switzerland	130	87	68	Ballad	English	0
8	2023	Israel	130	87	68	Pop	English	1
9	2023	Moldova	125	86	62	Traditiona	Romanian	0

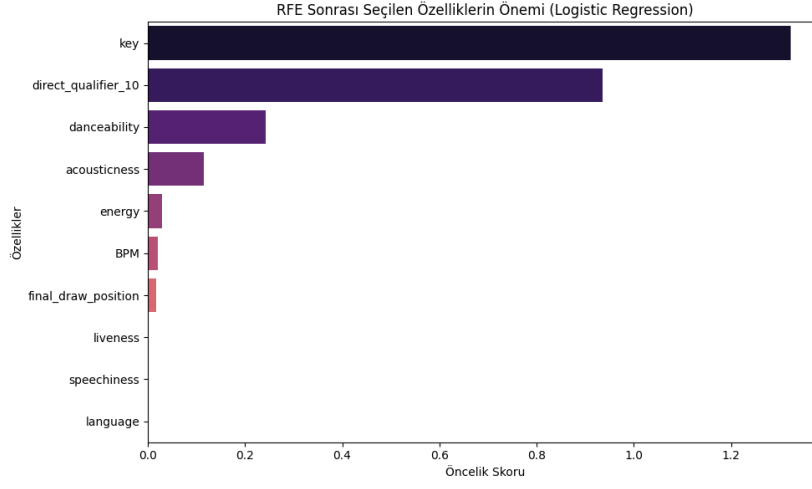
Ek B. Ölçekleme Öncesi ve Sonrası Değişken Dağılımları

Ölçekleme Öncesi ve Sonrası Dağılım Karşılaştırması

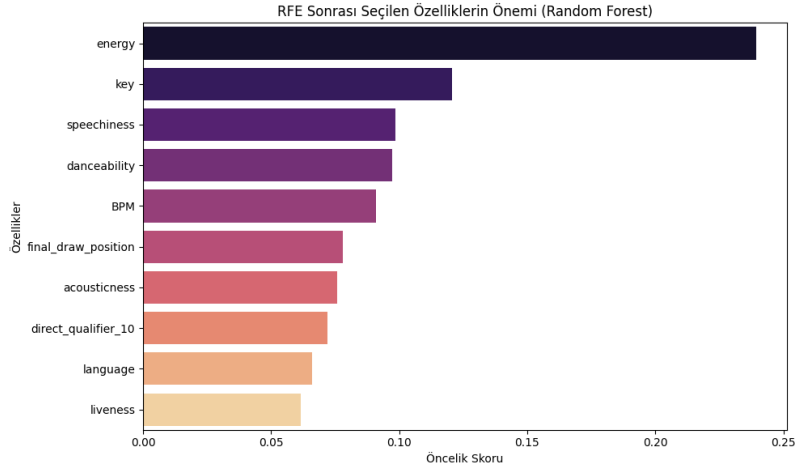


Ek C. RFE Özellik Önem Görselleri

- Ek C.1 Lojistik Regresyon – RFE Sonuçları



- Ek C.2 Random Forest – RFE Sonuçları



- Ek C.3 XGBoost – RFE Sonuçları

