Avrupa Bilim ve Teknoloji Dergisi
Özel Sayı 40, S. 143-150, Eylül 2022

Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
Special Issue 40, pp. 143-150, September 2022

Research Article

Müzik Veri Setinin Analizi ve Sınıflandırma Algoritmaları Kullanılarak Şarkı Türü Tahminleme Çalışması

Berke Bartuğ Sevindik[1], Vahide Bulut[2]

[1]Yazılım Mühendisliği /Fen Bilimleri Enstitüsü, İzmir Katip Çelebi Üniversitesi, Türkiye (ORCID: 0000-0002-5147-5300), bartugsevindik@gmail.com
[2] İzmir Kâtip Çelebi Üniversitesi Mühendislik-Mimarlık Fakültesi Mühendislik Bilimleri Bölümü, (ORCID: 0000-0002-0786-8860), vahide.bulut@ikcu.edu.tr

(1st International Conference on Innovative Academic Studies ICIAS 2022, September 10-13, 2022)

(**DOI:** 10.31590/ejosat.1174115)

ATIF/REFERENCE: Bartuğ, B.B., Bulut, V. (2022). Müzik Veri Setinin Analizi ve Sınıflandırma Algoritmaları Kullanılarak Şarkı Türü Tahminleme Çalışması. *Avrupa Bilim ve Teknoloji Dergisi*, (40), 143-150.

Özet – Bu araştırmanın amacı, Spotify müzik platformunda yer alan 42305 şarkı ve 15 farklı türe sahip veri setini analiz edip şarkının türlerle olan ilişkisini incelemektir. Bu türlerle olan ilişkiler veri setinden tür tahminleme çalışması için ön değerlendirme olarak analiz edilmiştir. Veri setindeki türlere ait özellikler değerlendirilip, kategorik olarak özelliklerine göre veri madenciliği sınıflandırma algoritmalarından; En yakın K-Komşu, rastgele orman, torbalama ve lojistik regresyon kullanılmıştır. Şarkının özelliklerine göre şarkıların türlerini tahmin etme çalışması gerçekleştirilmiştir. %55 ve %77 arasında doğruluk değerleri elde edilmiştir. Sınıflandırma algoritmalarının en iyi performans ölçüm değerine sahip bir model ele alınarak sonuçları değerlendirilmiştir.

Anahtar Kelimeler: Anahtar Kelimeler – Tahminleme, Analiz, En yakın k-komşu, Rasgele orman, torbalama, Lojistik regresyon, Sınıflandırma algoritmaları

Song Genre Estimation Study Using Music Data Set Analysis and Classification Algorithms

Abstract

The aim of this research is to analyze the dataset of 42305 songs and 15 different genres on the Spotify music platform and examine the relationship of the song with the genres. Relationships with these species were analyzed from the dataset as a preliminary assessment for the species prediction study. The features of the species in the data set are evaluated and categorically according to their features, from data mining classification algorithms; Nearest K-Neighbor, random forest, bagging and logistic regression were used. The study was carried out to predict the types of songs according to the characteristics of the song. Accuracy values between 55% and 77% were obtained. A model with the best performance measurement value of the classification algorithms was considered and the results were evaluated.

Keywords: Prediction, Analysis, K-nearest k-neighbors, Random forest, Logistic regression, bagging, Classification algorithms

1. Giriş

Günümüzde insanların en çok tercih ettiği etkinliklerden biri de müzik dinlemektedir. Kullanıcının sevdiği müzik türleri çeşitli özellikler içermektedir. Bu özellikler müziğin türünü belirlemeye yardımcı olmaktadır. Bu özellikler sayesinde çeşitli platformlar kullanıcının sevdiği türleri tespit edip bir sonraki dinleyeceği müziği önerebilmektedir.

Çeşitli müzik platformları içerisinde önemli bir yere sahip olan Spotify 2006 yılında İsveç'te kurulmuş olan bir dijital müzik platformudur. Açık şekilde paylaştıkları verilerle birlikte çok çeşitli çalışmalar yapılabilmektedir.

Bilginin daha özlü bir forma dönüştürülmesiyle elde edilen bir türü daha vardır. Bu bilgi, öz bilgi olarak tanımlanmaktadır Gürsakal (2001).

Günümüzün gelişmiş toplumlarında bilginin yönetiminde, daha çok öz bilgi ile ilgilenilmektedir (Gürsakal, 2001).

Veri tabanlarında öz bilgi keşfi süreci içerisinde, model oluşturma ve değerlendirme aşamalarını içeren veri madenciliği en önemli kısımdır (Akpınar, 2000).

Büyük bir veri setinden faydalı ve anlamlı bilgilerin ortaya çıkarmak son dönemlerde oldukça önem kazanmıştır. Veri madenciliği; makine öğrenmesi, istatistik veri görselleştirilmesi gibi alanlardaki teknikleri kullanan disiplinler arası bir alandır. Temel olarak veri madenciliği, veri kümeleriyle ilgili kalıpların veya komutların, veri analizinin ve yazılım tekniklerinin kullanılmasıdır. Amaç, daha önce tanınmayan veri kalıplarını tanımlamaktır. Veri madenciliği ve öz bilgi, verilerdeki değerli, anlamlı ve önceden bilinmeyen bilgileri elde etme sürecidir (Yıldırım vd, 2007).

Genel olarak veri madenciliği, büyük miktarda veriden, gizli kalmış, değerli ve kullanılabilir bilgilerin ortaya çıkarılmasını amaçlar (Koyuncugil, 2007).

Veri madenciliğinin yanı sıra verilerden anlamlı bilgiler ve tahminler çıkarmayı amaçlayan diğer bir alan da istatistiktir. Veri madenciliği ile istatistik arasında birçok yönden yakın ilişki bulunmaktadır. Veri madenciliği ve istatistiğin ortak özelliği "veriden öğrenme" (Ganesh, 2002) veya "verileri bilgiye dönüştürme"dir (Kuonen, 2004). İstatiksel tekniklerin temel veri ön işleme aşamalarında ve çıktıların değerlendirilmesinde faydaları gözlemlenmektedir.

Makine öğrenmesi, bir veri seti içerisinden önceden bilinmeyen yapıları (kalıpları) ortaya çıkarmaya yarayan otomatik süreçler olarak tanımlanabilir (Kelleher vd, 2015). Makine öğrenmesi, özellikle çok fazla miktarda verinin erişilebilir olduğu günümüzde, bu verilerden anlamlı çıkarımlar yapmak amacıyla geniş bir biçimde kullanılmaktadır. Makine öğrenimi, robotik, el yazısı ve konuşma tanıma, doğal dil işleme, beyin-makine arayüzleri ve daha fazlası dahil olmak üzere geniş bir uygulama yelpazesine sahiptir.

Sınıflandırma algoritmaları, bilgisayar bilimlerinde veri madenciliği konusunda kullanılan bir kavramdır. Bir veri kümesi üzerinde tanımlı olan sınıflar arasında veriyi dağıtmak sınıflandırma kavramının temel amacıdır.

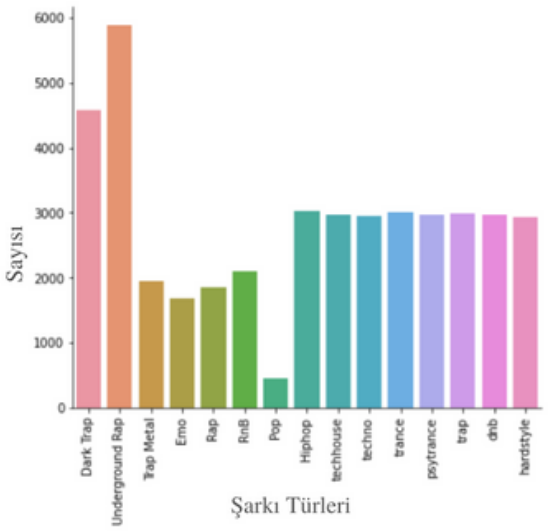
Araştırmanın temel amacı ise seçilmiş olan veri setinin incelenip tür ve özelliklerin analiz edilmesi, özelliklere göre tür tahminleme için kullanılan dört farklı sınıflandırma modellerinin kıyaslanmasıdır.

2. Materyal ve Metot

2.1. Müzik Veri Seti

Spotify büyük bir müzik platformudur. Kaggle adlı veri paylaşımlarının da yer aldığı web sitesinde bulunan Spotify veri seti 42305 şarkı ve 15 farklı türden oluşmaktadır. Veri setinde bulunan türler; Trap, Techno, Techhouse, Trance, Psytrance, Dark Trap, DnB, Hardstyle, Underground Rap, Trap Metal, Emo, Rap, RnB, Pop ve Hiphop müzik türlerinden oluşmaktadır. Veri setinde yer alan şarkı türleri ve bu türe ait olan şarkıların sayısı Şekil 1'de gösterilmiştir.

Şekil 1. Veri Setindeki Tür Dağılımı



Şarkıların da; dans edilebilirlik, enerji, canlılık, vocal miktarı, modu, ses yüksekliği ve benzeri özellikleri yer almaktadır.

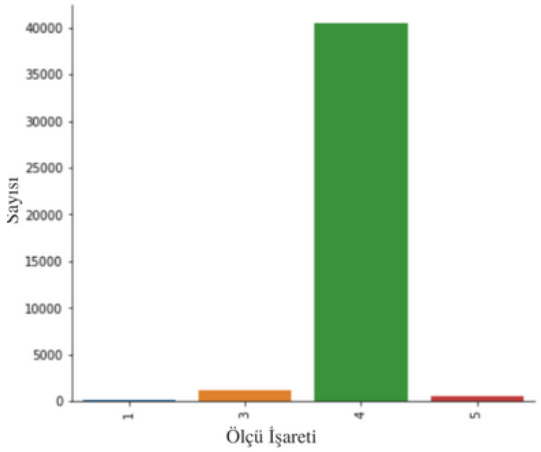
2.1. Veri Setinin Tür Analizi

Veri setinin tür analizi kısmındaki amacımız da farklı özelliklere sahip tüm şarkı türlerini analiz edip bu türlerle nasıl ilişkiler kurulduğunu gözlemleyebilmektir. Veri setindeki türler ve özelliklerinin ilişkileri sınıflandırılma öncesi incelenip analiz edilmektedir. Bu özelliklerden, ölçü işareti, dans edilebilirlik, enerji, ses yüksekliği, kelime miktarı, vokal miktarı ve canlılık çalışma kapsamında analiz edilip ilişkileri açıklanmıştır. Analiz kapsamında kutu grafikleri kullanılmıştır. Kutu grafiği, bir değişkenin dağılımının istatistiksel bir temsilidir. Kutunun uçları alt ve üst çeyrekleri temsil ederken, medyan (ikinci çeyrek) kutunun içinde bir çizgi ile işaretlenmiştir.

2.1.1. Ölçü İşareti ve Tür İlişkisi

Ölçü işareti, batı müzik notasyonunda her bir ölçüde kaç tane vuruş olduğunu belirlemek için kullanılır. Müzik notasyonunun başına konur ve vuruşlardaki nota değerini belirtmek için kullanılan sayı veya semboldür (Eğitim Sistem, n.d.). Veri setinde listelenen tüm şarkılar ve ölçü işaretlerinin dağılımı Şekil 2'de gösterilmiştir.

Şekil 2. Veri setindeki ölçü işareti dağılımı

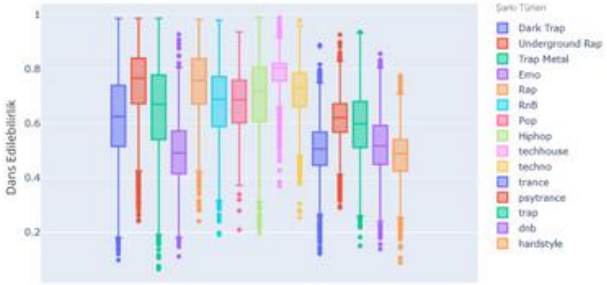


Veri setinde oldukça az olarak 1, 3 ve 5 dağılımı yer alırken, 4 müzik ölçü işaret özelliği çok sayıda yer almaktadır.

2.1.2. Dans Edilebilirlik ve Tür İlişkisi

Dans edilebilirlik, tempo, ritim istikrarı, vuruş gücü gibi müzik öğelerinin bir kombinasyonuna dayalı olarak bir parçanın dans etmek için ne kadar uygun olduğunu tanımlamaktadır. 0.0 değeri en az dans edilebilir ve 1.0 en çok dans edilebilir değeri temsil etmektedir (Santos, J. D. D. ; 2017). 0 ve 1.0 arasında yer alan müzik türlerinin dağılımı Şekil 3'te belirtilmiştir.

Şekil 3. Dans edilebilirlik ve Tür Dağılımı

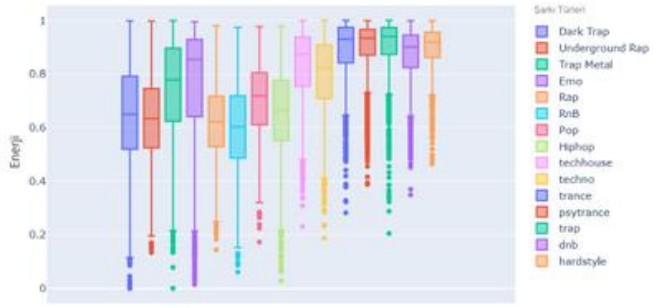


Şekil 3'te görüldüğü gibi techhouse türündeki şarkılar maksimum ortalama dans edilebilirliğe sahiptir. Bunu Underground Rap takip ederken hardstyle şarkıların minimum dans edilebilirliği bulunmaktadır.

2.1.3. Enerji ve Tür İlişkisi

Enerji, yoğunluk ve aktivitenin algısal bir ölçüsünü temsil etmektedir. Tipik olarak, enerjik parçalar hızlı ve gürültülü hissedilebilmektedir. Enerji ve tür ilişkisinin kutu grafiği Şekil 4'te verilmiştir.

Şekil 4. Enerji ve Tür Dağılımı

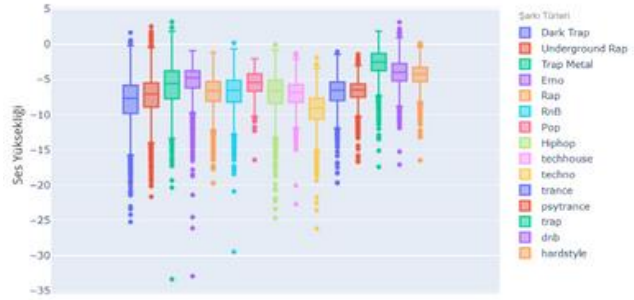


Şekil 4'te görüldüğü gibi trap, hardstyle, trance, dnb ve psytrance türündeki şarkıların en yüksek enerjiye, ve RnB, Rap Underground Rap şarkılarının da en düşük enerjiye sahip olduğu gözlenmektedir.

2.1.4. Ses Yüksekliği ve Tür İlişkisi

Bir şarkının ses seviyesi şarkının türünü belirlemekte kullanılan özelliklerden bir tanesidir. Şarkı türüne göre ses yükseklikleri ile ilgili grafik Şekil 5'te gösterilmiştir.

Şekil 5. Ses Yüksekliği ve Tür Dağılımı

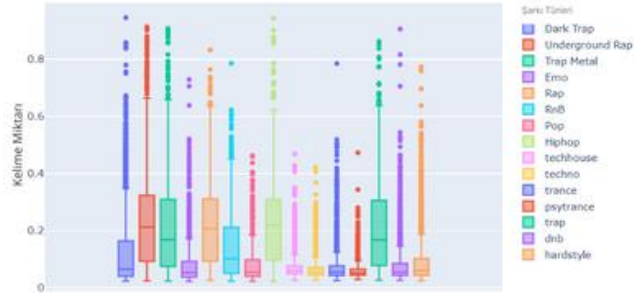


Şekil 5'te görüldüğü gibi tüm türlerdeki şarkıların ses yüksekliği puanları negatiftir. Trap türü en yüksek ses yüksekliği özelliğine sahipken tekno türü ise en düşük ses yüksekliği özelliğine sahiptir.

2.1.5. Kelime Miktarı ve Tür İlişkisi

Bir şarkının konuşma gücü 0,66'nın üzerindeyse, konuşulan kelimelerden oluşmaktadır, 0,33 ile 0,66 arasındaki değer hem müziği hem de kelimeleri içeren bir şarkıdır ve 0,33'ün altındaki değer şarkının herhangi bir konuşmasının olmadığı anlamına gelmektedir (Santos, 2017). 0.0 ve 1.0 arasında değerlendirilen şarkı türü ve kelime miktarı Şekil 6'da gösterilmiştir.

Şekil 6. Kelime Miktarı ve Tür Dağılımı

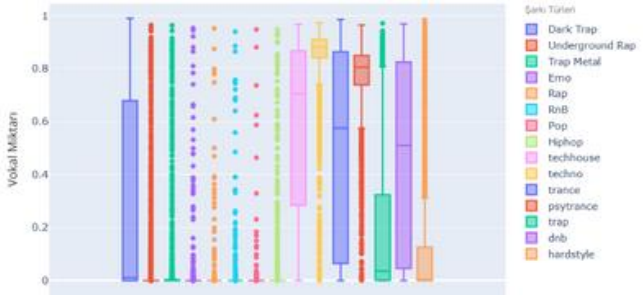


Şekil 6'da görüldüğü gibi rap, underground rap, ve hiphop en iyi üç türdür. Trance, techno ve psytrance ise en düşüklerdir.

2.1.6. Vokal Miktarı ve Tür İlişkisi

Şarkı türlerinin içerdiği enstrümantal seviyesi, veri seti içerisinde vocal miktarı olarak belirtilmektedir. Vokal miktarı ve şarkı türü arasındaki ilişki Şekil 7'de gösterilmiştir.

Şekil 7. *Vokal Miktarı ve Tür Dağılımı*

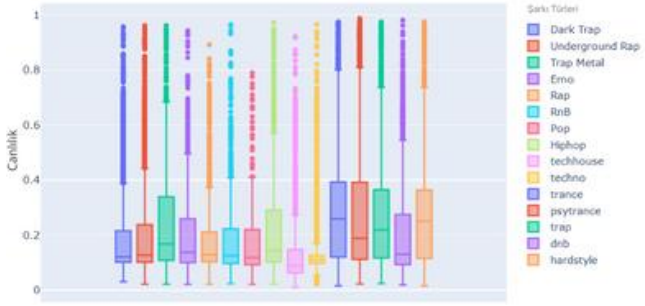


Şekil 7'de görüldüğü gibi techno türündeki şarkılar maksimum *vokal* puanına sahiptir, ve ardından da psytrance türü gelmektedir. Buradaki değer 1 değerine yaklaştıkça daha fazla enstrümantal şarkı içerdiğini belirtmektedir.

2.1.7. Canlılık ve Tür İlişkisi

Şarkı türlerine ait canlılık değerleri 0.0 ile 1.0 arasında tutulmaktadır. Resmi Spotify değerlerine göre 0.8'den yukarıda bir değer, şarkının daha canlı olduğuna dair güçlü bir olasılık katmaktadır. Şekil 8'de canlılık ve şarkı türlerinin dağılımı verilmiştir.

Şekil 8. Canlılık ve Tür Dağılımı



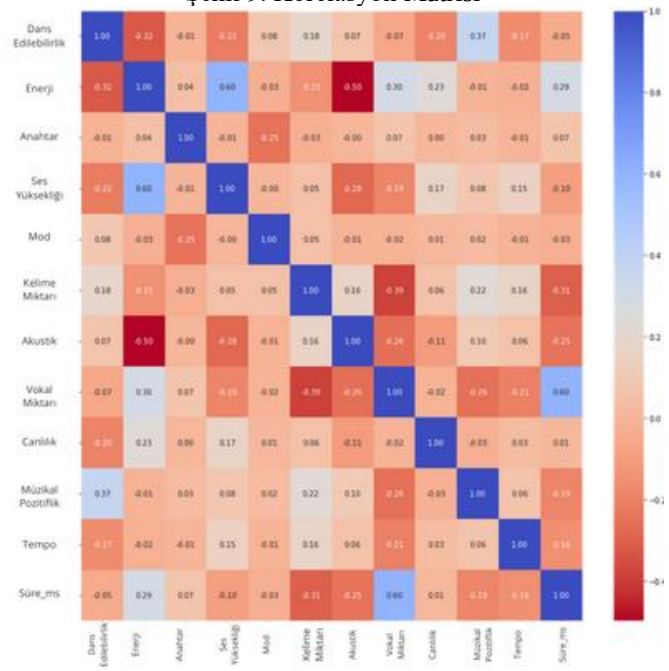
Şekil 8'de görüldüğü gibi trance ve psytrance en iyi canlılık skorlarına sahip türlerdir.

Çalışma kapsamında veri setinde listelenen 42305 şarkıya ait bazı özellikler ve bu şarkıların türlerinin ilişkisi yukarıda açıklanan ilişkiler baz alınarak incelenmiştir.

3. Tür Analizinin Sonuçları

Şarkı türleri analizi çalışması kapsamında, türlere ait özelliklerin korelasyon matrisi Şekil 9'da gösterilmiştir.

Şekil 9. Korelasyon Matrisi



Şekil 9'da görüldüğü gibi dans edilebilirlik ile müzikal pozitiflik özelliği arasında pozitif bir ilişki bulunmaktadır. Enerji ile ses yüksekliği, vokal miktarı, canlılık ve zaman aralığı pozitif bir ilişkiye sahiptir.

4. Sınıflandırma Algoritmaları

Sınıflandırma problemlerinde çıktı kümesinde her bir eleman birer sınıf iken bu problemi çözen algoritmaya da sınıflandırıcı denmektedir. Sınıflandırma problemleri, makine öğrenmesinin popüler ve en temel görevlerinden bir tanesidir. Bilinmeyen bir veri parçasının bilinen bir gruba yerleştirilmesinde kullanılmaktadır (Harrington, 2012).

Örneklerden çıkarım olarak da adlandırılan sınıflandırmada temel amaç, kavram tanımı elde edildikten sonra, daha önce algoritmaya belirtilmemiş örnekleri maksimum doğrulukla etiketleyecek sınıflandırıcıyı geliştirmektir (Çakır, 2005).

Makine öğrenmesi sınıflandırması, literatürde örüntü tanımı olarak da adlandırılmaktadır (Cen vd, 2010; Alpaydın, 2006).

3.1. Modeller

Çalışma kapsamında lojistik regresyon, rastgele orman, en yakın K-Komşu ve torbalama sınıflandırma algoritmaları kullanılmıştır.

Sınıflandırma algoritmaları için Scikit-learn kütüphanesinden Lojistik regresyon, rastgele orman, en yakın k-komşu ve torbalama için gerekli sınıflandırıcılar kullanılmıştır.

3.1.1. Lojistik Regresyon

Lojistik Regresyon çalışmada kullanılan sınıflandırma algoritmalarından bir tanesidir. Temel amacı minimum değişken kullanarak en iyi uyuma sahip olabilecek bağımlı ve bağımsız olarak değişkenler arasındaki ilişkiyi tanımlamak için en uygun modeli bulmaktır.

Lojistik regresyon algoritmasını önemli kılan etkenler şunlardır: log oranlarında uygulanan tahmini değişkenlere ait katsayıların tahmin edilebilme kolaylığı, belirli bir konuya ait sonucu geliştirme olasılığını tahmin edebilme kabiliyeti ve yaygın kullanım alanıdır (Barkus vd., 2006).

Çalışmada, Sklearn kütüphanesinde yer alan linear_model kütüphanesindeki lojistik regresyon, maksimum iterasyon sayısı 1000 olacak şekilde kullanılmıştır.

3.1.2. Rastgele Orman

Rastgele orman algoritması, torbalama algoritmasına dayalı bir karar ağacı tabanlı bir topluluk öğrenme algoritmasıdır (Bonaccorso, 2020; Breiman, 2001).

Rastgele orman algoritması, birden fazla karar ağacı üzerinden her bir karar ağacını farklı bir gözlem örneği üzerinde eğiterek çeşitli modeller üretip, sınıflandırma oluşturulmasını sağlamaktadır. Sınıflandırma işlemi sırasında, her bir karar ağacı, o ağacın eğitiminde kullanılmayan örneklerin sınıflandırılmasına oy verir. İşlemin sonunda ise en çok oyu alan sınıf döndürülür. (Han ve Ark., 2012; Kotu ve Deshpande, 2019).

Çalışmada, Sklearn kütüphanesinde yer alan RandomForestClassifier kütüphanesi, rastgele durumu 1 ve maksimum derinliği 10 olarak kullanılmıştır.

3.1.3. En Yakın K-Komşu

KNN, yöntem denetimli öğrenme metotları arasında yer alıp, sınıflandırma problemlerini çözmek için kullanılan bir modeldir. Bu yöntemde, sınıflandırılan verilerin eğitim setinin normal veri kümeleri ile benzerliği hesaplanarak ve verilen n en yakın verinin ortalaması alınarak elde edilen eşik değerine göre sınıflandırma yapılır. Bu algoritmanın temeli, sınıflandırma yapılmadan önce her sınıfın niteliklerinin önceden açıkça belirtilmesidir (Kaymaz 2007).

Çalışma kapsamında, Sklearn kütüphanesinde yer alan KneighborsClassfier kütüphanesi, komşu sayısı olan n parametresi 5 kabul edilerek kullanılmıştır. Farklı komşu parametreleri kullanılsa da en iyi sonuç komşu sayısı 5 olarak seçildiğinde alınmıştır.

3.1.4. Torbalama Sınıflandırma

Torbalama sınıflandırma yönteminin amacı genel olarak performansı iyileştirmek amacıyla birden fazla modelden alınan kararı iyileştirmektir. Tahminlerin birleştirilmesi aşamasında regresyon ağaçları için ortalama alınırken sınıflandırma ağaçlarında sonuçlar oylama ile belirlenmektedir (Şevket, 2019).

Çalışmada, torbalama olarak da nitelendirilen bagging Sklearn kütüphanesinde yer alan baggingClassifier kütüphanesi kullanılmıştır.

5. Karşılaştırma Yöntemleri

Sonuçları karşılaştırmak için doğruluk değeri, kesinlik değeri, duyarlılık değeri ve F1 değeri kullanılmıştır.

Doğruluk değeri aşağıdaki formülde belirtildiği gibi hesaplanmaktadır.

$$Doğruluk\ Değeri = \frac{TP + TN}{TP + FP + TN + FN}$$

Burada, doğru pozitif (TP) ve doğru negatif (TN) modelin doğru olarak tahminlendiği, yanlış pozitif (FP) ve yanlış negatif (FN) ise modelin yanlış olarak tahminlendiği alanlardır.

Kesinlik değeri ise pozitif olarak tahminlediğimiz değerlerin gerçekte kaç adedinin pozitif olduğunu göstermektedir. Kesinlik değeri aşağıda belirtilen formül ile hesaplanmaktadır.

$$Kesinlik\ Değeri\ = \frac{TP}{TP + FP}$$

Duyarlılık değeri, pozitif olarak tahmin etmemiz gereken işlemlerin kaç tanesini pozitif olarak tahmin edildiğinin bir ölçüsüdür. Duyarlılık değeri aşağıda belirtilen formül ile hesaplanmaktadır.

$$Duyarlılık\ Değeri = \frac{TP}{TP + FN}$$

F1 skor değeri bize, kesinlik ve duyarlılık değerlerinin uyum ortalamasını göstermektedir. F1 skor değeri aşağıdaki belirtilen formül ile hesaplanmaktadır.

$$F1 = 2 * \frac{Kesinlik * Duyarlılık}{Kesinlik + Duyarlılık}$$

6. Araştırma Sonuçları ve Tartışma

Şarkı türleri ve şarkılara ait dans edilebilirlik, enerji, canlılık, vocal miktarı, modu, ses yüksekliği ve benzeri özellikleri ile bağlantılı olduğu gözlenmektedir. Şarkıların özelliklerine göre şarkı türü tahminleme çalışması sonuçlarında, veri seti üzerinde her türün içerdiği şarkı farkları da doğruluk oranında yanılsamalara sebep olmaktadır. Sınıflandırma algoritmalarında ise Topluluk öğrenimi (Ensemble Learning) algoritmalarında iyi sonuç alındığı gözlenmektedir.

Bunu en yakın K-komşu takip etmektedir. Komşu sayısı arttıkça doğruluk oranı da düşmektedir. Genel olarak zayıf doğruluk bulunmaktadır.

Şarkı türü tahminleme çalısmasında kullanılan lojistik regresyon, rastgele orman, en yakın K-komşu ve torbalama sınıflandırma algoritmalarının doğruluk değerleri Tablo 1.'de listelenmiştir.

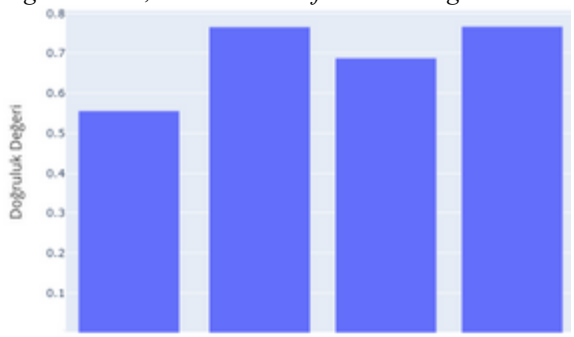
Tablo 1. Model Sonuçları

Model Adı	Doğruluk Değerleri
Lojistik Regresyon	0.5555744680851064
Rastgele Orman	0.6882269503546099
En yakın K-Komşu	0.7657872340425532
Torbalama Sınıflandırma	0.7668652482269503

Tablo 1'de verilen sonuçlara göre en iyi doğruluk değerini torbalama algoritması vermektedir. Tablo 1'deki sonuçlar grafik olarak ta Şekil 10'da verilmiştir.

Şekil 10'da görüleceği gibi en iyi sonuçlar en yakın K-Komşu ve torbalama sınıflandırma yöntemlerinde alınmaktadır. Torbalama sınıflandırma da aldığımız en iyi sonuçtur.

Şekil 10. Sırasıyla; Lojistik regresyon, En yakın K-Komşu, Rastgele orman, torbalama sınıflandırma algoritmaları



Veri setinde yer alan şarkı türleri ve veri setinde türlerin şarkı adetleri şekil 11'de listelenmiştir.

Şekil 11. Tür Sayıları

Underground Rap	5875
Dark Trap	4578
Hiphop	3028
trance	2999
trap	2987
techhouse	2975
dnb	2966
psytrance	2961
techno	2956
hardstyle	2936
RnB	2099
Trap Metal	1956
Rap	1848
Emo	1680
Pop	461

Torbalama yöntemine göre doğruluk ve ağırlıklı F1 değerleri Tablo 2.'de sunulmuştur.

Tablo 2. En iyi algoritma sonuçları

Algoritma	*Doğruluk Değeri*	*Ağırlıklı F1 Skoru*
Torbalama	*0.74*	*0.73*

Doğruluk, duyarlılık ve F1 skoru hesaplanmasında model çıktısına bakılmaktadır.

Sonuçları karşılaştırmak için kullanılan doğruluk değeri, kesinlik değeri, duyarlılık değeri ve F1 değeri Tablo 3'de listelenmiştir.

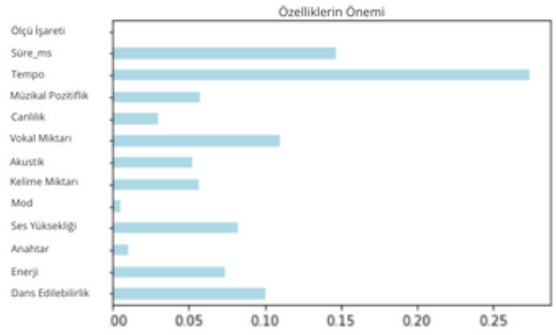
Tablo 3. Tür Sınıflandırma Sonuçları

Tür	Kesinlik	Duyarlılık	F1 Skoru	Destek Değeri
Dark Trap	0.53	0.43	0.47	1211
Emo	0.78	0.90	0.83	1132
Hiphop	0.59	0.62	0.61	1211
Pop	0.79	0.97	0.87	1140
Rap	0.68	0.83	0.74	1270
RnB	0.67	0.70	0.69	1168
Trap Metal	0.67	0.75	0.71	1225
Underground Rap	0.42	0.12	0.19	1149
dnb	0.95	0.96	0.96	1162
hardstyle	0.83	0.87	0.85	1107
psytrance	0.87	0.92	0.90	1187
techhouse	0.90	0.88	0.89	1185
techno	0.87	0.89	0.88	1166
trance	0.83	0.82	0.82	1169
trap	0.88	0.84	0.86	1143
Doğruluk Değeri			0.77	17625
Makro Ortalama	0.75	0.77	0.75	17625
Ağırlıklı Ortalama	0.75	0.77	0.75	17625

Veri setinde yer alan şarkı türlerin torbalama algoritmasına göre kesinlik, duyarlılık, f1 skoru ve destek değerleri Tablo 3'de verilmiştir.

Analiz ettiğimiz özelliklerin sonuçları ise lojistik regresyon algoritmasına göre incelediğimizde şekil 12'de gözükmektedir.

Şekil 12. Özelliklerin Önemi



Veri setinde şarkı türlerine ait özellikleri, çalışmanın tur analizindeki önemi Şekil 12'de belirtilmiştir.

7. Sonuç

Çalışma kapsamında şarkı türü analizi için dans edilebilirlik, enerji, canlılık, vokal miktarı, modu, ses yüksekliği ve benzeri özellikleri incelenmiştir. Bu inceleme sonucunda Şekil 9.'da aktarıldığı gibi tür analizinde özelliklerin birbirleri ile olan ilişkisi görülmektedir. Şarkı Türü tahminlemesi için yapılan çalışmada bu özellikler önemli bir yer almaktadır. Şekil 10'da görüldüğü gibi tür tahminlemesi için kullanılan lojistik regresyon, en yakın K-komşu, rastgele orman, torbalama sınıflandırma algoritma sonuçlarından en iyi sonuçlar K-komşu ve torbalama algoritmalarından alınmıştır.

Tablo 2'de gösterildiği üzere en iyi sonuç torbalama algoritmasında alınmaktadır.

Türlere ait özellikler bir türün tahmininde benzerlik gösterdiği için skorlara da yansımaktadır. Şekil 11'de de görüleceği gibi veri setinde yer alan türlerin sayıları sonuçlarda dengesizliğe sebep olmaktadır. Şarkı türleri benzer özellikleri içerdiği için tahmin sonuçları da yüksek değerler vermemektedir.

8. Teşekkür

Çalışmamda müzik öneri sistemi öncesinde yapmam gereken tür özellik analizi ve sınıflandırma algoritmaları ile tür tahminleme içeriği fikirleri ve destekleri için sayın Doç. Dr. Vahide Bulut'a teşekkürlerimi sunarım.

Kaynakça

[1] *Sklearn.svm.LinearSVC*. scikit. (n.d.). Retrieved September 11, 2022, from https://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html

[2] *About Spotify*. Spotify. (2022, July 27). Retrieved September 11, 2022, from https://newsroom.spotify.com/company-info/.

[3] Mavuduru, A. (2021, February 10). *How to build an amazing music recommendation system*. Medium. Retrieved September 11, 2022, from https://towardsdatascience.com/how-to-build-an-amazing-music-recommendation-system-4cce2719a572

[4] T., Tibshirani, R. & Friedman, J. (2008). The Elements of Statictical Learning: Data Mining, Inference, and Prediction, Second Edition, Springer.

[5] Rajavarman, V.N. ; Rajagopalan, S.P. ; Comparison between Traditional data mining Techniques and Entropy-based Adaptive Genetic Algorithm for Learning Classification Rules; International Journal of Soft Computing Vol 2 Issue 4; 2007; 555-561.

[6] Öztemel, E. (2012). Yapay sinir ağları. (3.baskı). İstanbul: Papatya Yayıncılık.

[7] Han, J., Kamber, M. and Pei, J. (2012). Data mining: Concepts and techniques. (3rd Edition). Waltham: Morgan Kaufmann.

[8] J. Khairnar and M. Kinikar, "Machine learning algorithms for opinion mining and sentiment classification," International Journal of Scientific and Research Publications, vol. 3, no. 6, pp. 1–6, 2013.

[9] N. Mishra and C. K. Jha, "Classification of opinion mining techniques," International Journal of Computer Applications, vol. 56, no. 13, pp. 1–6, 2012.

[10] Watts JD, Lawrence RL. 2008. Merging random forest classification with an object-oriented approach for analysisof agricultural lands, The International Archives of the Photogrammetry, Remote Sensing and Spatial InformationSciences, XXXVII(B7)

[11] Loh WY, Shih YS. 1997. Split selection methods for classification trees. Statistica Sinica 7: 815-840.

[12] Wang, C., Long, Y., Li, W. et al. (2020). Exploratory study on classification of lung cancer subtypes through a combined K-nearest neighbor classifer in breathomics. Sci Rep, 3;10(1):5880. doi: 10.1038/s41598-020-62803-4.

[13] ULUSLARARASI SAĞLIK YÖNETİMİ VE STRATEJİLERİ ARAŞTIRMA DERGİSİ http://dergipark.gov.tr/usaysad (VERANYURT, Ü /DEVECİ, AF /ESEN, MF /VERANYURT,

[14] Mercaldo, F., Nardone, V., Santone, A. (2017). Diabetes Mellitus Affected Patients Classification and Diagnosis through Machine Learning Techniques. Procedia Computer Science, 112: 2519-228.

[15] Mujumdar, A., Vaidehi, V. (2019). Dibetes Prediction Using Machine Learning Algorithms. Procedia Computer Science, 165: 292–299.

[16] Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification,"Information Theory, IEEE Transactions, 13: 21-27.

[17] Breiman, L. (2001). Random forest. Mach. Learn, 45: 5–32. doi: 10.1023/A:1010933404324.

[18] "Analysis of Top Tracks in Spotify." Https://Web.stanford.edu/, 25 Oct. 2018, web.stanford.edu/~kjytay/courses/stats32-aut2018/Session%208/Spotify_final.html.

[19] Ay, Yamac Eren. "Spotify Dataset 1921-2020, 160k+ Tracks." Kaggle, 24 Jan. 2021, www.kaggle.com/yamaerenay/spotify-dataset-19212020-160k-tracks.

[20] Alpaydın, E., 2014, Introduction to Machine Learning, MIT Press, ISBN: 978-0-262-02818-9.

[21] Alpaydın, E., 2006, Projects in Machine Learning, http://web.eecs.utk.edu/~parker/Courses/CS594spring06/handouts/Introduction.pdf,

[22] Harrington, P., 2012, Machine Learning in Action, 1st Edition, Manning Publications Shelter Island, NY, ISBN: 978-1-61729-018-3.

[23] *Ölçü ve ölçü çizgisi nedir, ölçü işareti Nedir*. Eğitim Sistem. (n.d.). Retrieved September 11, 2022, from https://www.egitimsistem.com/olcu-isareti-nedir-86365h.htm

[24] Gürsakal, N. (2001) Sosyal Bilimlerde Araştırma Yöntemleri, Uludağ Üniversitesi Basımevi, Bursa.

[25] Gürsakal, N. (2007) Betimsel İstatistik Minitab, Spss, Statistica, Excel Uygulamalı, Nobel Yayın Dağıtım, Ankara.

[26] Akpınar, H. (2000) "Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği", İstanbul Üniversitesi İşletme Fakültesi Dergisi, Cilt 29, Sayı 1/Nisan, s. 1–22.

[27] Yıldırım, P., Uludağ, M. ve Görür, A. (2007), "Hastane Bilgi Sistemlerinde Veri Madenciliği", Akademik Bilişim Kongresi, Çanakkale Onsekiz Mart Üniversitesi, Çanakkale, 30 Ocak-1 Şubat 2007.

[28] Ganesh, S. (2002) "Data Mining: Should it be included in the 'Statistics' cirriculum?", The Sixt International Conference on Teaching Statistics, Cape Town, South Africa, 7–12 July.

[29] Koyuncugil, A. S. (2007) "Veri Madenciliği ve Sermaye Piyasalarına Uygulaması", Sermaye Piyasası Kurulu Araştırma Raporu, Araştırma Dairesi, 28.02.2007 ASK/1

[30] Santos, J. D. D. (2017, May 31). *Is my Spotify music boring? an analysis involving music, data, and machine learning*.

Medium. Retrieved September 11, 2022, from https://towardsdatascience.com/is-my-spotify-music-boring-an-analysis-involving-music-data-and-machine-learning-47550ae931de

[31] Ay, Şevket. (2019, December 16). *Ensemble learning-bagging VE boosting*. Medium. Retrieved September 11, 2022, from https://medium.com/deep-learning-turkiye/ensemble-learning-bagging-ve-boosting-50643428b22b