

## Oyun Pazarı

Serhat GÜNEY<sup>1</sup>, Pınar KARADAYI ATAS<sup>2,\*</sup>

<sup>1</sup> İstanbul Arel Üniversitesi, Mimarlık ve Mühendislik Fakültesi, Bilgisayar Mühendisliği, Türkiye

<sup>2</sup> İstanbul Arel Üniversitesi, Mimarlık ve Mühendislik Fakültesi, Bilgisayar Mühendisliği, Türkiye

<sup>1</sup> ORCID: <https://orcid.org/0000-0003-3194-5405>

<sup>2</sup> ORCID: <https://orcid.org/0000-0002-9429-8463>

### Öz

Bu veri seti, video oyunlarının global pazarını incelemek için kullanılan bir veri setidir[1]. Veri setinde oyunların adı, platformları, ülkeler, tarihler, satışlar ve diğer bilgiler yer almaktadır. Bu veri seti, makine öğrenimi yöntemleri kullanılarak oyunların satışlarının hangi faktörler tarafından etkilendiğini ve hangi oyunların daha yüksek satışlar elde ettiğini tahmin etmek için kullanılabilir[1]. Özellikle, bu veri seti Decision tree gibi karar ağaçları yöntemi ile oyunların satışlarını tahmin etmek için kullanılabilir. Bu yöntem oyunların satışlarını etkileyen faktörleri ve oyunların satışları arasındaki ilişkiyi görselleştirmek için kullanılabilir.

**Anahtar Kelimeler:** *Oyun Satışları; Karar Ağacı; Makine Öğrenmesi; Naïve Bayes.*

### Abstract

This dataset is used for analyzing the global market of video games[1]. It contains information such as game names, platforms, countries, dates, sales, and other data. This dataset can be used with machine learning techniques to predict the factors that affect game sales and which games have higher sales[1]. Specifically, it can be used with methods such as decision trees to predict game sales and identify the relationship between factors and game sales. The decision tree method can be used to visualize the factors that affect game sales and the relationship between game sales.

**Keywords:** *Game Sales; Decision Tree; Machine learning; Naïve Bayes.*

### 1. Giriş

Bu çalışmada, oyunların adı, platformları, ülkeler, tarihler, satışlar ve diğer bilgiler gibi bilgileri içeren bir veri seti kullanarak video oyunlarının global pazarını analiz etmeyi amaçlıyoruz. Oyunların satışlarını etkileyen faktörleri ve hangi oyunların daha yüksek satışlar elde ettiğini tahmin etmek için, decision tree ve Naive Bayes algoritmalarını kullanacağız.

Decision tree, oyunların satışları ve faktörler arasındaki ilişkiyi anlamak için güçlü bir yöntemdir ve oyunların satışlarını etkileyen faktörleri görselleştirmek için kullanılabilir. Bu algoritma, veriyi girdi özelliklerinin değerlerine göre daha küçük alt kümeler olarak böler ve ağaç benzeri bir model oluşturur. Ağaç dalları son tahminlere yol açan koşulları temsil ederken, yapraklar son tahminleri temsil eder[3].

Diğer taraftan, Naive Bayes bir olasılık tabanlı algoritmadır ve Bayes Teoremi'ne dayanır. Özellikle sınıflandırma problemleri için kullanılır ve girdi özelliklerinin belirli bir sınıfa ait olasılığına dayanarak tahminler yapar. Örneğin, bir oyunun satışlarını tahmin etmek için bir oyunun belirli bir platforma veya ülkeye ait olasılığını kullanır [2].

Bu çalışmada, bu algoritmaları oyunların satışlarını tahmin etmek ve faktörler ile satışlar arasındaki ilişkiyi belirlemek için kullanacağız. Model performansları karşılaştırılacak ve daha ileri analizler için en iyi algoritma seçilecektir. Veri setinde çeşitli kategorik özellikler ve sayısal özellikler bulunmaktadır, bu yüzden bu algoritmalar bu veri tipi için uygun seçenekler olarak kullanılabilir.

## 2. Materyal

### 2.1. Veri Seti

Kullandığımız veri setinde, dünyanın her yerinden, farklı platformlar, türler ve bölgelerden video oyunları için satış verilerini içerir. Çeşitli oyunlar, türleri ve platformları üzerine kapsamlı analizi sayesinde bu veri seti, video oyunlarının küresel başarıya nasıl ulaşabileceğine dair ayrıntılı içgörüler sunarken oyun kültürünün sürekli değişen trendlerine harika bir pencere sunuyor [1].

Veri setindeki sütunların açıklaması şu şekildedir;

- Rank: Her oyunun sıralamasını küresel satışlarına göre (en yüksekten en düşüğe) tanımlar. Bu, hangi oyunların dünya çapında en popüler olduğunu belirlemenize yardımcı olur.
- Game Title: Bir oyunu diğerinden kolayca ayırt etmenizi sağlayan her bir video oyununun adını içerir.
- Platform: Her oyunun piyasaya sürüldüğü platformun türünü listeler, örneğin PlayStation 4 veya Xbox One, böylece platformlar arasında ve her platforma özel oyunlar arasında karşılaştırma yapabiliriz.
- Year: Yıllardan yıla karşılaştırmalar yapmak ve oyun satışlarında zaman içindeki değişiklikleri izlemek için ek bir yol sağlar.
- Publisher: Oyunu yayınlayan yayıncı adıdır.
- Review: İnceleme skorlarını belirtir.
- North America: Kuzey Amerikadaki satış miktarını belirtir.
- Japan: Japonyadaki satış miktarını belirtir.
- Europe: Avrupadaki satış miktarını belirtir.
- Rest of World: Yukardaki farklı diğer ülkelerin satış miktarını belirtir.
- Global: Toplam satış miktarını belirtir.

Veri setinde 1907 tane satır 11 sütun bulunmaktadır bu sütunların bilgileri Tablo 1 de gösterildiği üzere veri setinde year sütununda biraz kayıp değerler (missing value) var. İki tane integer, 4 tane nominal ve 4 tane continues değişken var.

**Tablo 1.**Attribute Tipleri.

Attribute	Count	Data Type
index	1907 non-null	int64
Rank	1907 non-null	int64
Game Title	1907 non-null	object
Platform	1907 non-null	object
Year	1878 non-null	float64
Genre	1907 non-null	object
Publisher	1905 non-null	object
North America	1907 non-null	float64
Europe	1907 non-null	float64
Japan	1907 non-null	float64
Rest of World	1907 non-null	float64

## 2.2 Preprocessing İşlemleri

Preprocessing, verinin analiz veya modelleme için hazır hale getirilmesi işlemidir. Bu işlem, veri setinde eksik değerlerin bulunması, irrelevant özelliklerin kaldırılması, veri setinin düzenlenmesi, veri setinde yer alan değişkenlerin dönüştürülmesi gibi işlemleri içerebilir [2]. Preprocessing, veri setinin temizlenmesi, düzenlenmesi ve hazır hale getirilmesi işlemlerinin yapıldığı adım olarak da tanımlanabilir. Bu işlem, veri setinin daha sonra analiz veya modelleme için kullanılmasını kolaylaştırır ve daha iyi sonuçlar elde etmemizi sağlar[2].

Year sütununda bulunan kayıp değerler bulunan satırlarını silme işlemi yaptık çünkü veri setinde çok fazla kayıp değerler bulunmadığından sıkıntı çıkmadı. Index sütunu bir alakasız bir sütundu çünkü sadece dizin belirtiyordu bunu python üzerinden kolayca öğrenebiliriz ve bu sütun kaldırıldı. Rank özelliği için, kategorik veriye çevrilerek yüksek satış başarmış oyunlar ve düşük satış yapmış olarak 5 farklı kategoriye çevrilir. Decision Tree ve Naïve Bayes learner 'ların çalışması için en sonunda nominal veriler numeric veriye dönüştürdük.

## 3. Metodoloji

Veriyi analiz etmek için grafikler ve learner kullanıldı. Bu tip veri seti için önerilen algoritmalar genelde Knn, Naïve Baye, Decision Tree, Random Forest v.b. algoritmalarıdır [2].

### 3.1 Gaussian Naive Bayes

Gaussian Naive Bayes, Naive Bayes algoritmasının bir varyantıdır. Bu algoritma, Bayes Teoremi'ne dayalı olasılık tabanlı bir algoritmadır. Gaussian Naive Bayes, inputların sürekli değişken olduğu sınıflandırma problemleri için kullanılır. Bernoulli Naive Bayes ve Multinomial Naive Bayes gibi diğer varyantlarına göre, Gaussian Naive Bayes inputların Gaussian dağılımı izlediğini varsayar [4].

Gaussian Naive Bayes algoritması, belirli bir sınıf için inputların olasılığını hesaplayarak çalışır ve daha sonra Bayes Teoremi'ni kullanarak sınıfın inputlarına göre olasılığını hesaplar. En yüksek olasılığa sahip sınıf son tahmin olarak seçilir [4].

**Tablo 2.** *Gaussian Naive Bayes Metrics*

Accuracy	Mean Squared Error	Mean Absolute Error
0.90558	0.66223	0.23936

### 3.2 Decision Tree

Decision Tree, veri setinde bulunan faktörler arasındaki ilişkiyi anlamak ve tahmin yapmak için kullanılan bir makine öğrenimi algoritmasıdır. Bu algoritma, veriyi küçük alt kümeler olarak böler ve sonuçları ağaç benzeri bir yapıda görselleştirir. Ağaç dalları, veriyi bölmede kullanılan koşulları temsil ederken, yapraklar sonuçları temsil eder[2].

Decision tree algoritması, veri setinde bulunan özellikleri kullanarak, veriyi sınıflandırmak için bir karar ağacı oluşturur. Bu ağaç, veriyi sürekli olarak böler ve her bölmede verinin nerede yoğun olduğunu gösterir. Bu, veriyi anlamak ve tahmin yapmak için kullanılabilir.

Decision Tree algoritması, kategorik veya numeric verileri işleyebilir. Ayrıca, veri setinde eksik değerler olsa bile işleyebilir. Decision Tree algoritması, veri setinin boyutunun büyük olması durumunda da etkili sonuçlar verebilir.

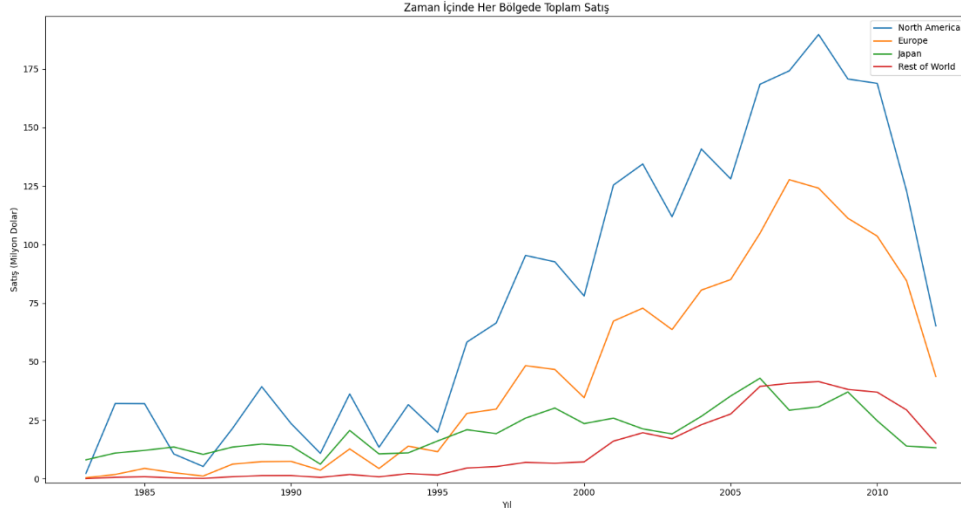
**Tablo 2.** *Decision Tree Metrics.*

Accuracy	Mean Squared Error	Mean Absolute Error
0.99202	0.06515	0.02260

### 3.3 Association

#### 3.3.1 Zaman Göre Her Bölgede Toplam Satış

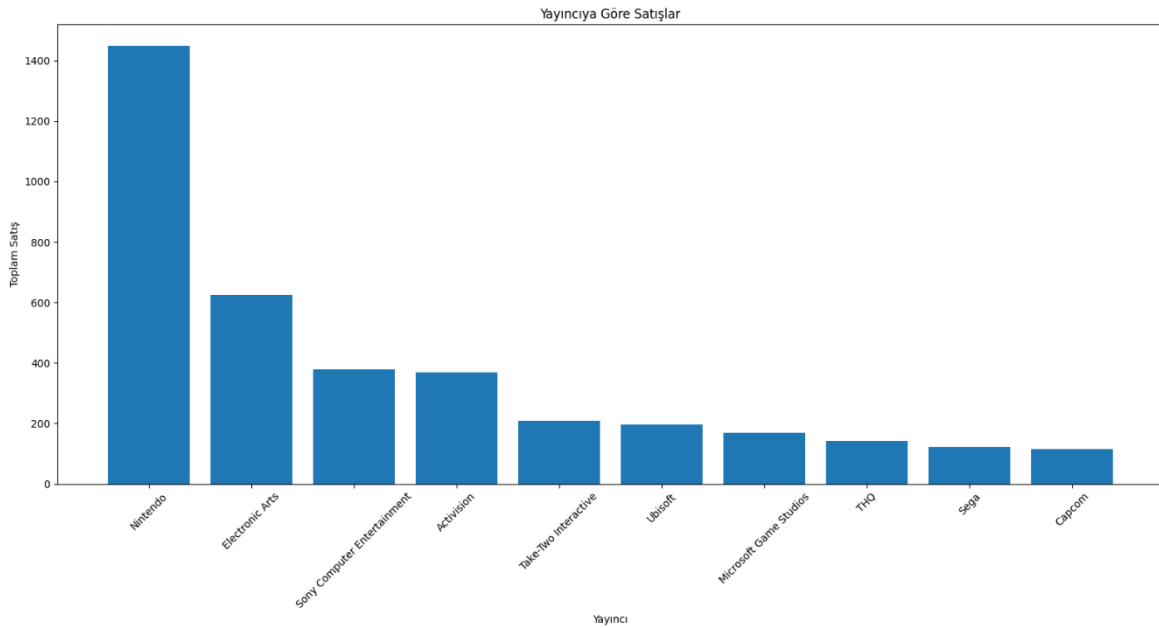
Burada bölgeden bölgeye endüstrinin hangi bölgelerde daha fazla geliştiği görülmekte bu bilgiyle bölgeye göre pazarlama stratejileri geliştirilebilir. Kuzey Amerika en gelişim göstermiştir.



Şekil 1. Zamana Göre Her Bölgede Toplam Satış.

#### 3.3.2 Yayıncıya Göre Satış

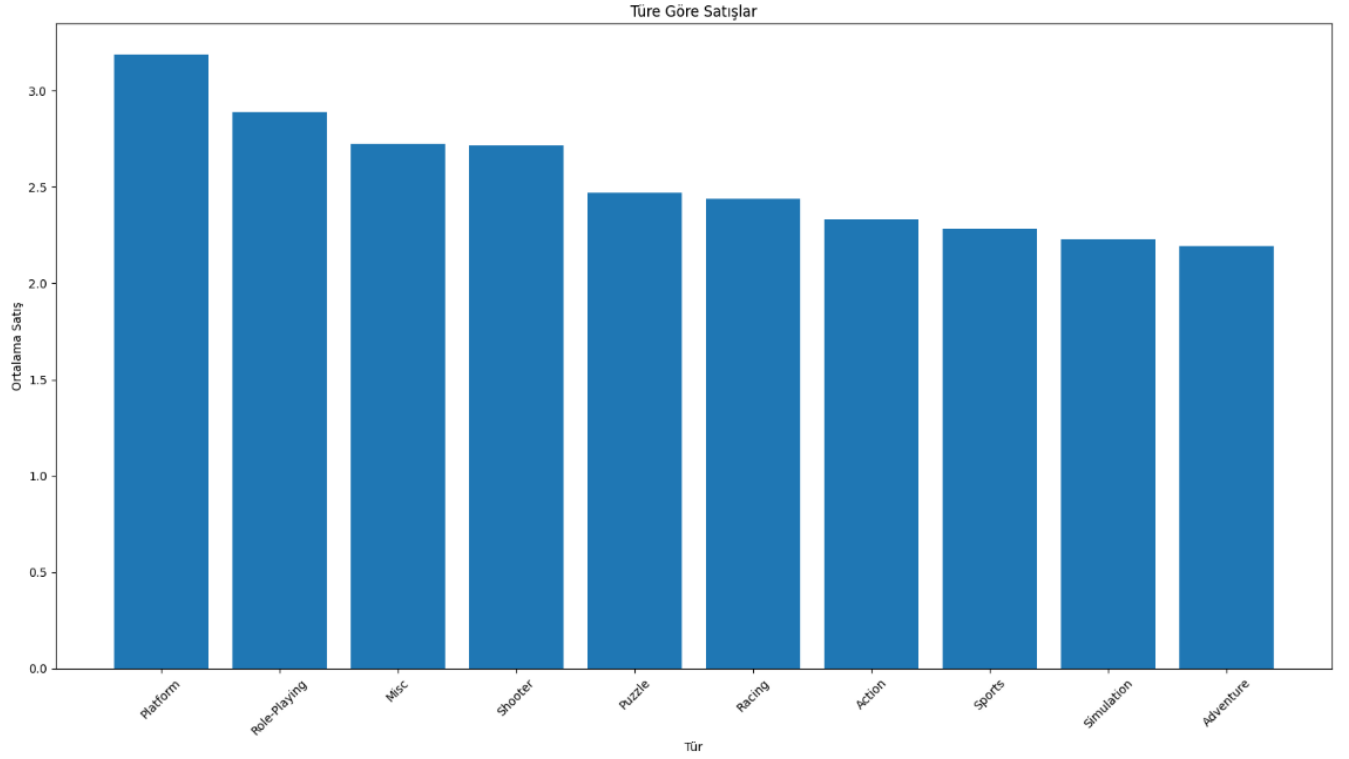
Burada hangi yayıncıların global satışta en çok para kazandığını gösteriyor hangi yayıncıyla çalışmak istediğinizi buradaki veriler göre seçilebilir. Nintendo firması piyasayı domine ediyor .



Şekil 2. Yayıncıya göre satış

### 3.3.3 Oyun Türüne Göre Satış

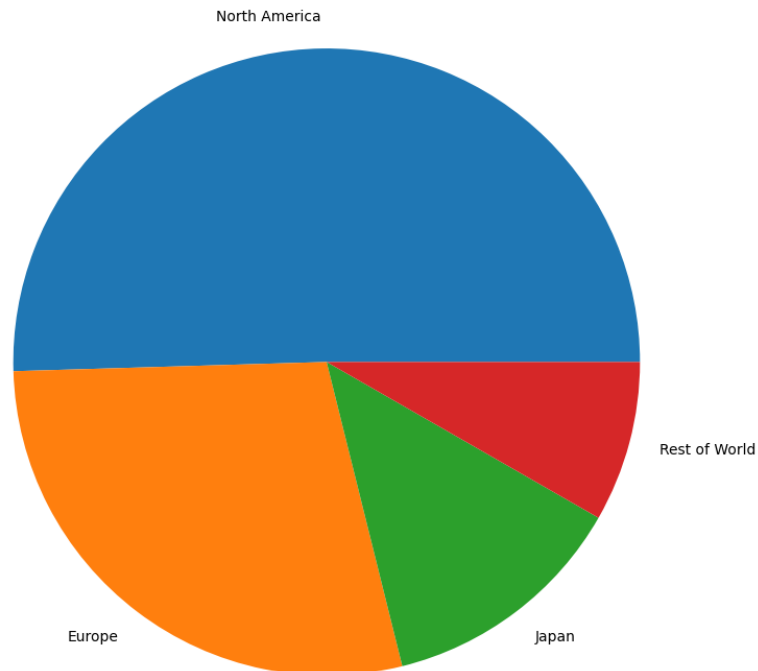
Hangi oyun türünün daha fazla piyasası olduğuna karar vermek için bakılır. Her türün satış oranı benzer olduğundan direkt bir etki etmez.



Şekil 3. Oyun türüne göre satış .

### 3.3.4 Her Bölgenin Satış Piyasası

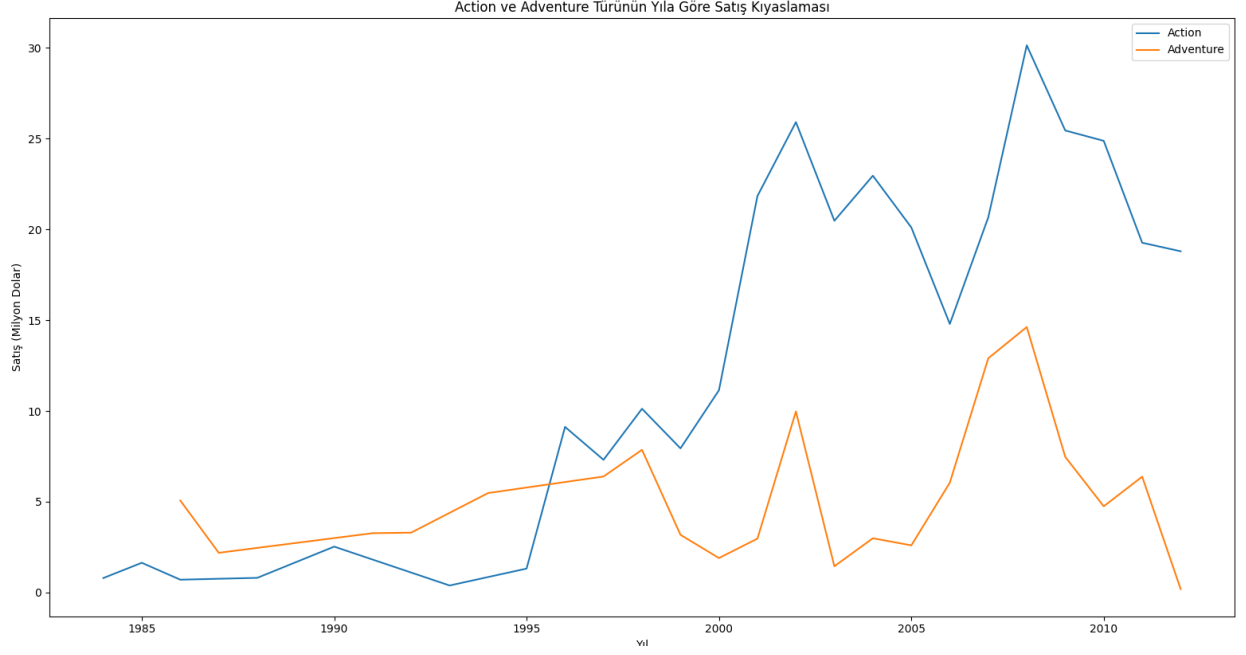
Piyanın satış oranı bakımından Kuzey Amerika pastanın en büyük pay vardır.



Şekil 4. Pazar pasta tablosu .

### 3.3.5 Aksiyon ve Adventure Oyun Türü Satışı Yıla Göre

Yıldan yıla aksiyon türü daha büyük kitlelere erişip gelişirken adventure türü gün geçtikçe daha bir niş kitleye düşmektedir.



Şekil 5. Zamana göre oyun türü satışı.

## 4. Sonuç

Oyun satışlarında yıldan yıla daha fazla Pazar payının büyüdüğünü, en büyük pasta payının Kuzey Amerikada olduğunu ve bazı oyun türlerinin hayran kitlesini zamandan zamana azlamasıyla daha az geliştiğini çıkarımını yaptık.

Learnerlar'la hangi oyun'un daha yüksek rankı olduğunu yüksek bir oranla Decision Tree ve Naïve Bayes algoritmasıyla % 90 üzeri bir doğruluk değeriyle bulabildik.

## Kaynak

- [1] Braimwell A. Discovering Hidden Trends in Global Video Games  
Retrieved from <https://www.kaggle.com/datasets/thedevastator/discovering-hidden-trends-in-global-video-games>
- [2] Alpaydın E. (2014) Introduction to Machine Learning third edition <https://www.cmpe.boun.edu.tr/~ethem/i2ml3e/>
- [3] Kingsford C. & Salzberg S. L. En Popüler 7 (2008) What are decision trees?  
Retrieved from <https://www.nature.com/articles/nbt0908-1011>
- [4] Webb, G. I., Keogh, E., & Miikkulainen, R. (2010). Naïve Bayes. Encyclopedia of machine learning, 15, 713-714.
- [5] Chong, C. Y. (2012, July). Graph approaches for data association. In 2012 15th international conference on information fusion (pp. 1578-1585). IEEE.