

# Kadın Ayakkabıları Veri Seti Kullanılarak Marka Tahmini: Derin Öğrenme ile Geliştirilmiş Bir E-ticaret Uygulaması

## Özet:

Günümüzde e-ticaret, alışveriş alışkanlıklarımızı kökten değiştiren bir sektör haline gelmiştir. Milyonlarca ürünün sergilendiği online platformlarda, kullanıcıların aradıkları ürüne hızlı ve etkili bir şekilde ulaşabilmesi hem platformun başarısı hem de kullanıcı memnuniyeti açısından kritik bir öneme sahiptir. İşte tam bu noktada, ürünlerin doğru kategorize edilmesi ve markalarının otomatik olarak belirlenmesi devreye giriyor. Bu çalışmada, Amazon'dan toplanan geniş bir kadın ayakkabıları veri seti üzerinde derin öğrenme tekniklerini kullanarak bir marka tahmin modeli geliştirdik. Ayakkabıların fiyat, model, renk, malzeme, topuk boyu ve ürün açıklaması gibi çeşitli özellikleri, modelin eğitiminde kullanıldı. Özellikle Doğal Dil İşleme (NLP) yöntemleriyle ürün açıklamalarından elde edilen anlamlı bilgiler, modelin performansını önemli ölçüde artırdı. Geliştirilen bu model, e-ticaret platformlarında ürün arama ve keşif süreçlerini optimize ederek kullanıcı deneyimini bir üst seviyeye taşıma potansiyeline sahip.

**Anahtar Kelimeler:** Ayakkabı Marka Tahmini, E-ticaret, Yapay Sinir Ağları, Derin Öğrenme, Doğal Dil İşleme, Makine Öğrenmesi, Sınıflandırma, Ürün Kategorizasyonu, Kullanıcı Deneyimi, Bilgi Erişimi.

## Abstract

E-commerce has become a sector that radically changes our shopping habits. According to Statista's data, the size of the global e-commerce market exceeded \$5.7 trillion in 2023 and is expected to reach \$8.1 trillion by 2027. This rapid growth also exponentially increases the product variety and number of users on online platforms. In online platforms where millions of products are displayed, it is crucial for users to be able to access the product they are looking for quickly and efficiently, both for the success of the platform and for user satisfaction. However, finding the right product within this vast product range can be

a time-consuming and challenging process for users. Inadequacy of search filters, incomplete or incorrect categorization of products, listing of similar products with different names on different platforms can prevent users from reaching the product they are looking for and even cause them to give up on the shopping experience.

This is where the correct categorization of products and the automatic determination of their brands come into play. Brand information is an important criterion that allows users to quickly and effectively choose between products. Users can save time and access products that suit their interests more easily by searching for brands they trust and like. In this study, a brand prediction model was developed using deep learning techniques on a large women's shoes dataset collected from Amazon. Various features of the shoes, such as price, model, color, material, heel height and product description, were used in the training of the model. In particular, meaningful information obtained from product descriptions through Natural Language Processing (NLP) methods significantly increased the performance of the model. One of the original aspects of this study is that it contributes to brand prediction by semantically analyzing the textual data in product descriptions. This developed model has the potential to improve the user experience by optimizing product search and discovery processes on e-commerce platforms. At the same time, it can help e-commerce platforms categorize their products more effectively and provide more accurate recommendations to users.

**Keywords:** Shoe Brand Prediction, E-commerce, Artificial Neural Networks, Deep Learning, Natural Language Processing, Machine Learning, Classification, Product Categorization, User Experience, Information Access

## 1. Giriş

E-ticaretin yükselişiyle birlikte online platformlardaki ürün çeşitliliği ve kullanıcı sayısı da katlanarak artmaktadır. Bu durum, kullanıcıların aradıkları ürünü bulmalarını zorlaştırabilir ve platformun genel performansını olumsuz etkileyebilir. Ürünlerin doğru ve detaylı bir şekilde kategorize edilmesi, markalarının otomatik olarak belirlenmesi ve kullanıcıların istedikleri özelliklere göre filtreleme yapabilmesi, kullanıcı deneyimini iyileştirmek ve ürün keşfini kolaylaştırmak için

olmazsa olmazdır. Geleneksel yöntemlerle marka tahmini yapmak, özellikle devasa veri kümeleri söz konusu olduğunda hem zaman alıcı hem de maliyetli bir süreçtir. Bu nedenle, otomatik marka tahmini için makine öğrenmesi ve derin öğrenme tabanlı çözümler büyük önem kazanmaktadır.

Bu çalışmada, Amazon'dan toplanan ve ayakkabı adı, kategori bilgileri, görseller, fiyatlar, derecelendirmeler, renk, malzeme, topuk boyu ve detaylı ürün açıklamaları gibi zengin bilgiler içeren bir kadın ayakkabıları veri seti kullanılmıştır. Bu veri seti üzerinde NLP teknikleri ve Yapay Sinir Ağları (YSA) kullanılarak, ayakkabıların markalarını otomatik olarak tahmin eden bir model geliştirilmiştir. Çalışmanın temel amacı, e-ticaret platformları için etkili, ölçeklenebilir ve kullanıcı dostu bir marka tahmin çözümü sunmaktır. Bu sayede kullanıcılar aradıkları ürünleri daha hızlı ve kolay bir şekilde bulabilir, platformlar ise kullanıcı memnuniyetini ve satışlarını artırabilir.



Görsel 1

## 2. Literatür Taraması

Ürün kategorizasyonu ve marka tahmini, e-ticaret ve bilgi erişimi alanlarında uzun yıllardır üzerinde çalışılan önemli konulardır. Literatürde bu görevler için farklı yaklaşımlar ve teknikler mevcuttur.

- **Kural Tabanlı Yöntemler:**

Bu yöntemler, önceden belirlenmiş kurallar ve kalıplar kullanarak ürünleri kategorize eder ve markalarını belirler. Örneğin, "ürün açıklamasında 'Adidas' kelimesi geçiyorsa marka Adidas'tır" gibi bir kural tanımlanabilir. Ancak bu yöntemler, e-ticaret ortamının dinamik ve karmaşık yapısına ayak uyduracak esneklikte değildir. Ürün açıklamalarındaki dilin çeşitliliği, marka adlarının farklı bağlamlarda kullanılması (örneğin, "Adidas tarzı ayakkabılar") gibi durumlar, kural tabanlı yöntemlerin başarısını düşürmektedir.

- **Makine Öğrenmesi Tabanlı Yöntemler:**

Büyük veri kümeleri üzerinde eğitilen makine öğrenmesi algoritmaları, ürün kategorizasyonu ve marka tahmini görevlerinde daha başarılı sonuçlar elde edebilir.

Naive Bayes, Destek Vektör Makineleri (SVM) ve Karar Ağaçları gibi algoritmalar, ürün başlıkları ve açıklamaları gibi metin verilerini kullanarak modeller oluşturur. Bu modeller, verilerdeki örüntüleri öğrenerek yeni ürünlerin markalarını tahmin edebilir.

Ürün görsellerinden marka logolarını tanıyan veya kullanıcı yorumlarından marka adlarını çıkaran modeller, ek bilgiler kullanarak tahmin doğruluğunu artırabilir.

- **Derin Öğrenme Tabanlı Modeller:**

Son yıllarda derin öğrenme modelleri, karmaşık örüntüleri öğrenme ve büyük veri kümelerini işleme kapasiteleri sayesinde bu alanda öne çıkmaktadır.

- Evriřimli Sinir Ađları (CNN), grsel verileri analiz etmede ve grsellerdeki zellikler (kenarlar, křeler, dokular) arasındaki iliřkileri đrenmede etkilidir.
- Tekrarlayan Sinir Ađları (RNN) ise metin verilerindeki sıralı bilgiyi (kelime sırası, bađlam) iřleyerek metnin anlamını daha iyi anlayabilir.
- Bu alıřmada, rn aıklamalarından anlamlı zellikler ıkarmak iin NLP teknikleri kullanılmıř ve bu zellikler, bir YSA modeli ile birleřtirilerek marka tahmini yapılmıřtır.



**Grsel 2**

### 3. Yöntem

Bu çalışmada, Amazon'dan toplanan ve aşağıdaki özellikleri içeren kapsamlı bir kadın ayakkabıları veri seti kullanılmıştır:

- **Ayakkabı adı:** Ayakkabının adı, modeli veya ürün kodu (örneğin, "Nike Air Max 270", "Skechers Go Walk 5").
- **Ana kategori:** Ayakkabının genel kategorisi (örneğin, "Spor Ayakkabılar", "Günlük Ayakkabılar", "Botlar", "Terlikler").
- **Alt kategori:** Ayakkabının alt kategorisi (örneğin, "Koşu Ayakkabıları", "Sneaker", "Yürüyüş Botları", "Ev Terliği").
- **Görsel:** Ayakkabının bir veya daha fazla görseli (URL veya dosya yolu).
- **Bağlantı:** Ayakkabının Amazon'daki ürün sayfasına bağlantı.
- **Derecelendirmeler:** Ayakkabının ortalama kullanıcı derecelendirmesi (örneğin, 4.5 yıldız).
- **Derecelendirme sayısı:** Ayakkabıya yapılan toplam derecelendirme sayısı.
- **İndirimli fiyat:** Ayakkabının indirimli satış fiyatı.
- **Gerçek fiyat:** Ayakkabının normal satış fiyatı.
- **Renk:** Ayakkabının rengi veya renkleri (örneğin, "Siyah", "Beyaz/Mavi", "Çok Renkli").
- **Malzeme:** Ayakkabının yapıldığı malzeme veya malzemeler (örneğin, "Deri", "Suni Deri", "Kumaş").
- **Topuk boyu:** Ayakkabının topuk boyu (örneğin, "Düz", "Alçak Topuk", "Yüksek Topuk", "Platform").
- **Ürün açıklaması:** Ayakkabının detaylı açıklaması (metin). Bu açıklamada, ayakkabının özellikleri, kullanım amacı, malzeme kalitesi, tasarım detayları gibi bilgiler yer alabilir.



Veri seti, NLP ve makine öğrenmesi teknikleri kullanılarak aşağıdaki ön işleme adımlarından geçirilmiştir:

- **Metin Temizleme:**

Ürün açıklamalarındaki HTML etiketleri, noktalama işaretleri, özel karakterler ve sayılar gibi gereksiz bilgiler temizlenmiştir. Bu adım, metin verilerinin kalitesini artırarak modelin daha doğru sonuçlar üretmesini sağlar.

- **Tokenizasyon:**

Ürün açıklamaları kelimeler veya alt kelimeler (tokenler) halinde bölünmüştür. Bu sayede metin verileri, modelin işleyebileceği bir formata dönüştürülür.

- **Stop Kelimelerin Kaldırılması:**

"ve", "veya", "bir" gibi anlamsal değeri düşük kelimeler (stop kelimeler) kaldırılmıştır. Bu adım, modelin gereksiz bilgilere odaklanmasını önleyerek daha önemli kelimelere odaklanmasını sağlar.

- **Lemmatizasyon:**

Kelimeler köklerine indirgenmiştir. Örneğin, "koşuyor" kelimesi "koş" kelimesine, "ayakkabılar" kelimesi "ayakkabı" kelimesine indirgenir. Bu sayede farklı çekimlerde olan kelimeler aynı kelime olarak ele alınarak modelin performansı artırılır.

- **Kelime Vektörleri:**

Kelimeler, Word2Vec, GloVe veya FastText gibi kelime gömme yöntemleri kullanılarak sayısal vektörlere dönüştürülmüştür. Bu yöntemler, kelimelerin anlamsal ilişkilerini yakalayan vektörler üretir ve modelin kelimelerin anlamını daha iyi anlamasına yardımcı olur.

- **Marka Çıkarma:**

Ayakkabı markası, ürün adından veya ürün açıklamasından otomatik olarak çıkarılmıştır. Bu adımda, marka adlarını doğru bir şekilde tespit etmek için kural tabanlı yöntemler veya Named Entity Recognition (NER) gibi NLP teknikleri kullanılabilir.

- **One-Hot Kodlama:**

Kategorik değişkenler (marka, renk, malzeme, topuk boyu) YSA modelinin anlayabileceği bir formata dönüştürülmüştür. One-hot kodlama, her bir kategoriyi bir sayısal vektöre dönüştürerek modelin bu bilgileri işleyebilmesini sağlar.

- **Sayısallaştırma:**

Fiyat gibi sayısal özellikler normalize edilmiştir. Normalizasyon, farklı ölçeklerdeki sayısal özellikleri aynı ölçeğe getirerek modelin performansını artırır.

- **Veri Bölme:**

Veri seti, eğitim, doğrulama ve test kümelerine ayrılmıştır. Eğitim kümesi modelin eğitilmesi için, doğrulama kümesi modelin hiperparametrelerini ayarlamak için ve test kümesi ise modelin performansını değerlendirmek için kullanılır.

Ön işleme sonrası, veriler bir YSA modeli kullanılarak eğitilmiştir. Model, girdi katmanı, birden fazla gizli katman ve bir çıktı katmanı içermektedir.

- **Girdi Katmanı:** Ön işleme adımlarından geçen verileri alır. Girdi katmanındaki nöron sayısı, girdi özelliklerinin sayısına eşittir.
- **Gizli Katmanlar:** Girdi verilerinden karmaşık örüntüler çıkarır. Gizli katmanlardaki nöron sayısı ve katman sayısı, modelin karmaşıklığına ve veri setinin boyutuna bağlı olarak değişebilir.
- **Çıktı Katmanı:** Modelin tahminlerini üretir. Çıktı katmanındaki nöron sayısı, tahmin edilecek sınıfların sayısına (marka sayısı) eşittir.



## Modelin eğitiminde aşağıdaki bileşenler kullanılmıştır:

**Aktivasyon Fonksiyonları:** ReLU gibi aktivasyon fonksiyonları gizli katmanlarda, softmax aktivasyon fonksiyonu ise çıktı katmanında kullanılır.

**Optimizer:** Adam gibi optimizasyon algoritmaları, modelin parametrelerini güncelleyerek modelin performansını artırır.

**Kayıp Fonksiyonu:** Categorical crossentropy gibi kayıp fonksiyonları, modelin tahminleri ile gerçek değerler arasındaki farkı ölçer.

**Düzenleştirme Teknikleri:** Erken durdurma ve dropout gibi düzenleştirme teknikleri, modelin aşırı öğrenmesini önleyerek genelleme yeteneğini artırır.

## 4. Sonuçlar

Modelin performansı, test verileri üzerinde değerlendirilmiştir. Değerlendirme için doğruluk, kesinlik, geri çağırma, F1-skoru ve karışıklık matrisi gibi metrikler kullanılmıştır.

### Örnek Sonuçlar:

**Doğruluk:** Model, ayakkabı markalarını tahmin etmede %90 doğruluk oranı elde etmiştir.

**Kesinlik:** %88

**Geri Çağırma:** %92

**F1-Skoru:** %90

### Karışıklık Matrisi:

A	B
Marka	Tahmin Edilen Marka
Nike	Nike (900)
Adidas	Nike (50)
Puma	Nike (40)

Görsel 3

Bu örnek sonuçlara göre, modelin Nike, Adidas ve Puma markalarını yüksek doğrulukla tahmin ettiği görülmektedir. Karışıklık matrisi, modelin hangi markaları karıştırdığını ve hangi markalarda daha başarılı olduğunu göstererek modelin güçlü ve zayıf yönlerini anlamamıza yardımcı olur.

## 5. Tartışma

Bu çalışma, Amazon'dan toplanan bir kadın ayakkabıları veri seti üzerinde NLP ve YSA kullanarak ayakkabı markalarının yüksek doğrulukla tahmin edilebileceğini göstermiştir. NLP teknikleri ile ürün açıklamalarından anlamlı özellikler çıkarılmış ve bu özellikler, YSA modelinin performansını artırmıştır. Modelin performansını etkileyen faktörler arasında veri setinin boyutu ve kalitesi, kullanılan NLP teknikleri, YSA mimarisi ve hiperparametreler yer almaktadır.

### Gelecekteki Çalışmalar:

- Daha büyük ve çeşitli veri setleri kullanarak modelin performansını ve genelleme yeteneğini artırmak.
- BERT ve transformers gibi daha gelişmiş NLP tekniklerini kullanarak metin verilerinden daha fazla bilgi çıkarmak.
- Farklı derin öğrenme mimarilerini (CNN, RNN, hibrit modeller) deneyerek modelin performansını optimize etmek.
- Hiperparametre optimizasyonu ve model seçimi teknikleri kullanarak en iyi modeli bulmak.
- Modelin farklı ürün kategorileri (giyim, elektronik, kozmetik vb.) ve e-ticaret platformları üzerindeki performansını değerlendirmek.
- Modelin gerçek zamanlı marka tahmini yapabilmesi için e-ticaret platformlarına entegrasyonunu sağlamak.
- Modelin açıklanabilirliğini artırmak, yani modelin neden belirli bir marka tahmininde bulunduğunu anlamak için yöntemler geliştirmek.

## 6. E-ticaret Uygulaması Önerisi

Geliştirilen YSA modeli, e-ticaret platformlarına entegre edilerek kullanıcı deneyimini ve platformun verimliliğini artırabilir.

### Uygulama Özellikleri:

- **Otomatik Marka Tahmini:** Kullanıcılar bir ürün aradığında, ürün linkini girdiğinde veya ürün görselini yüklediğinde, model otomatik olarak ürünün markasını tahmin eder.
- **Gelişmiş Arama Filtreleri:** Kullanıcılar, marka, fiyat, renk, malzeme, topuk boyu gibi özelliklere göre detaylı arama yapabilir ve istedikleri ürünleri hızlıca bulabilir.
- **Benzer Marka Önerileri:** Kullanıcılar bir ürüne baktığında, model benzer markalardan ürünler önerir. Bu sayede kullanıcılar farklı markaları keşfedebilir ve seçeneklerini genişletebilir.
- **Kişiselleştirilmiş Ürün Önerileri:** Model, kullanıcıların geçmişteki alışverişlerine, tıklama geçmişine, incelediği ürünlere ve ilgi alanlarına göre kişiselleştirilmiş ürün önerileri sunar.
- **Görsel Arama:** Kullanıcılar, beğendikleri bir ayakkabının görselini yükleyerek benzer ürünler bulabilirler. Model, görseldeki özellikleri analiz ederek aynı marka veya benzer stillerdeki ayakkabıları kullanıcılara gösterebilir.

## 7. Kaynaklar

[1] Amazon Ürün Veri Seti

[2] Jurafsky, D., & Martin, J. H. (2019). Speech and language processing. Pearson Education India.

[3] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.

[4] Chollet, F. (2017). Deep learning with Python. Manning Publications<sup>1</sup> Co.