



**T.C.  
KÜTAHYA DÜMLUPINAR ÜNİVERSİTESİ  
MÜHENDİSLİK FAKÜLTESİ**

**SAHTE HABER TAHMİNİ MAKİNE ÖĞRENİMİ UYGULAMASI**

**OSMAN CANBAZOĞLU  
201913172036**

**BATUHAN ÖZTÜRK  
201913172024**

**MAKİNE ÖĞRENİMİ PROJE RAPORU  
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ**

**DANIŞMAN  
DR. ÖĞR. ÜYESİ PINAR ÖZEN**

**KÜTAHYA, 2023**

---

*Proje Raporu*

**Sahte Haber Tahmini Makine Öğrenimi Uygulaması**

---

*Proje Teslim Tarihi: 27.12.2023*

---

**ÖZ**

Bu rapor, sahte haber tespiti için makine öğrenimi uygulamasını incelemektedir. Sahte haberlerin toplumsal etkileri ve bu konudaki teknolojik çözüm yollarından bahsedilmektedir. Literatür taraması sayesinde, çeşitli dillerde sahte haber tespiti için yapılmış çalışmalar ve kullanılan teknolojilerden bahsedilerek, gerekli bilgilendirmeler yapılmıştır. Araştırma, farklı makine öğrenimi modellerinin sahte haber veri seti üzerindeki etkisini araştırmayı amaçlamaktadır. Aynı zamanda çalışma, veri toplama ve model eğitimi süreçlerini kapsamaktadır. Bulgular, çeşitli modellerin performans parametreleriyle değerlendirilmiştir.

**Anahtar kelimeler:** Sahte Haber, Makine Öğrenimi, Yapay Zeka, Veri Analizi

## GİRİŞ

Günümüz medyasında habercilik, sürekli olarak gelişen teknoloji karşısında evrimleşmektedir. Sahte haberlerin yayılmasının toplumsal etkileri, bilgi çağının en büyük sorunlarından biri olarak öne çıkmaktadır. Yanıltıcı içeriklerin hızla yayılması toplumda yanlış algılar yaratmaktadır. (Farha, 2023)

Son yıllarda, yapay zeka ve makine öğrenimi gibi teknolojilerin, bu soruna karşı çözüm bulmakta önemli bir rol oynadığı gözlemlenmiştir. Ancak, bu teknolojik araçların yeterliliği ve etkinliği de sürekli olarak sorgulanmakta ve geliştirilmektedir. (Golbeck et al., 2018)

Bu çalışma, sahte haberlerin tespiti ve kontrol altına alınması için yapılan çeşitli araştırmaların önemini vurgulamakta ve bu sorunu çözmek için gerekli çalışmalar barındırmaktadır. Bu çalışmanın önemli parçası, sahte haberlerin toplumsal etkisini anlamak ve bu etkileri azaltmak için geliştirilmiş olmasıdır. Aynı zamanda, sahte haberlerin algılanmasında insan faktörünün rolünü ve teknolojik araçların faydalarını anlatmaktadır. Sahte haberlerle mücadele etmek için makine öğrenimi, derin öğrenme ve bir çok model kullanımı önemli bir rol oynamakla birlikte, hayatımızda kullanılabilir bir teknoloji olarak yer almaya başlamaktadır.

## LİTERATÜR TARAMASI

Makine öğrenmesi ve sahte haber üzerine yapılan literatür taramasında farklı makale, bildiriler göz önünde bulundurularak 20 adet tarama sonucu elde edilmiş olup, bu sonuçlar doğrultusunda ilgili yayınların başlık, konu, kullanılan teknolojiler, yöntem ve sonuç bulguları ortaya çıkarılmıştır.

### 1. Kurdish Fake News Detection Based on Machine Learning Approaches

(Salh & Nabi, 2023)

- a. **Konu:** Yazar bu çalışmada, Kürtçe dilinde sahte haberleri tespit etmeye yönelik bir makine öğrenimi projesini ele almaktadır.
- b. **Kullanılan Teknolojiler:**
  - Makine öğrenimi algoritmaları (Rastgele Orman (Random Forest), Destek Vektör Makineleri (Support Vector Machine), ve Evrişimli Sinir Ağları (Convolutional Neural Networks)).

- Özellik çıkarma teknikleri (Kelime Gömme (Word Embedding), Terim Frekansı-Ters Belge Frekansı (Term Frequency-Inverse Document Frequency), ve Sayım Vektörü (Count Vector)).
- c. **Yöntem:** Kürtçe sahte haber veri setini kullanarak haber metinlerinin özelliği çıkarılmıştır. Makine öğrenimi ve derin öğrenme sınıflandırıcısı kullanılmıştır.
- d. **Sonuç:** Evrişimli Sinir Ağlarının (CNN), diğer modellere göre daha yüksek doğruluk ve F1 skoru ile daha iyi performans göstermiştir. Sahte haberlerin metinsel içeriğinin, özellikle CNN kullanılarak, başarılı bir şekilde tespit edildiği sonucuna varılmıştır.

## 2. Building a Dataset for Detecting Fake News in Amharic Language

(Tazeze & R, 2021)

- a. **Konu:** Yazar bu çalışmasında Amharca dilinde sahte haberleri tespit etmek için bir veri seti oluşturmayı hedeflemiştir.
- b. **Kullanılan Teknolojiler:**
  - Selenium WebDriver
  - Facebook Scraper kütüphanesi
  - TF-IDF
  - Pasif Agresif Sınıflandırıcı
  - Naive Bayes
  - Destek Vektör Makineleri
  - Lojistik Regresyon
  - Stokastik Gradyan
- c. **Yöntem:** Amharca haber kaynaklarından ve Facebook sayfalarından haberler toplanmış, etiketlenmiş ve makine öğrenimi modelleri ile sınıflandırılmıştır.
- d. **Sonuç:** Naive Bayes ve pasif agresif sınıflandırıcı modellerinin TD-IDF ve CountVectorizer ile yüksek doğruluk ve F1 puanı elde ettiği gözlemlenmiştir.

## 3. A Blockchain System For Fake News Detection

(Bobulski, 2024)

- a. **Konu:** Yazar bu çalışmasında blockchain teknolojisi kullanarak sahte haberlerin tespiti uygulaması üzerinde çalışmıştır.

#### **b. Kullanılan Teknolojiler:**

- Blockchain
- Eşler arası ağ (peer-to-peer network)
- Konsensüs algoritmaları
- İş kanıtı (Proof of Work)

**c. Yöntem:** Blockchain tabanlı sahte haber tespiti için bir sistem tasarımı ve bu sistemde kullanılan çeşitli algoritmaların detaylı açıklamaları.

**d. Sonuç:** Blockchain teknolojisinin, sahte haberlerinin tespiti ve doğruluğunun doğrulanmasında etkili olduğu, bu teknolojinin bilgi yayılımını ve içerik doğrulama süreçlerinde manipülasyonlara karşı dirençli olduğu sonucuna varılmıştır.

#### **4. The PolitiFact-Oslo Corpus: A New Dataset for Fake News Analysis and Detection** (Pöldvere et al., 2023)

**a. Konu:** Yazar bu çalışmada PolitiFact-Oslo Corpus ile İngilizce dilinde, ABD odaklı sahte ve gerçek haberler içeren yeni bir veri seti oluşturmayı hedeflemiştir.

#### **b. Kullanılan Teknolojiler:**

- Derin Öğrenme Ağları (DBN)
- Uzun Kısa Süreli Hafıza (LSTM)
- Bidirectional LSTM
- Transformers modelleri (BERT, RoBERTa, DistilBERT, XLNet)

**c. Yöntem:** PoliFact.com'dan veri toplama, manuel ve otomatik tekniklerle veri analizi, doğruluk için uzman etiketlemesi, metin türü ve kaynak bilgileriyle zenginleştirme çalışması yapılmıştır.

**d. Sonuç:** Farklı metin türlerinde (sosyal medya, haber ve bloglar) sahte ve gerçek haberlerin dilbilimsel özelliklerinin analizi yapıp, derin öğrenme tabanlı sahte haber tespiti modellerin performans değerlendirmeleri karşılaştırılmıştır.

#### **5. Building an Optimal Dataset for Arabic Fake News Detection**

(Bsoul et al., 2022)

**a. Konu:** Yazar bu çalışmada Arapça dilindeki sahte haberleri tespit etmek için bir veri seti oluşturmuştur.

#### **b. Kullanılan Teknolojiler:**

- Makine öğrenim modelleri (Lojistik Regresyon, Destek Vektör Makineleri, Rastgele Orman, Naive Bayes, Stokastik Gradyan Azalışı, En Yakın Komşu ve Karar Ağaçları)

**c. Yöntem:** Arapça haber başlıklarının toplanılması, etiketlenmesi ve çeşitli makine öğrenimi algoritmalarıyla sınıflandırılması.

**d. Sonuç:** Farklı makine öğrenimi modellerini uygulanması sonucunda elde edilen F1 puanları ve modellerinin sahte haber tespiti için kullanılabilirliği test edilmiştir.

### **6. Twitter'da Makine Öğrenmesi Yöntemleriyle Sahte Haber Tespiti**

(KAYAKUŞ & YİĞİT AÇIKGÖZ, 2023)

**a. Konu:** Yazar bu çalışmada Twitter üzerindeki sahte haberlerin makine öğrenmesi yöntemleri kullanılarak tespiti üzerinde çalışmalar yapmıştır.

#### **b. Kullanılan Teknolojiler:**

- Karar ağaçları
- Naive Bayes Sınıflandırıcıları

**c. Yöntem:** Twitter'dan seçilen konularda atılan tweetlerin toplanması, metin ön işlemleri ve analizler sonucu sahte ve gerçek haberler olarak etiketlenmesi.

**d. Sonuç:** Naive Bayes sınıflandırıcısının F1 skoru 0.883 ile karar ağaçlarına göre daha yüksek performans gösterdiği sonucuna varılmıştır.

### **7. Twitter Üzerinde Türkçe Sahte Haber Tespiti**

(TAŞKIN et al., 2021)

**a. Konu:** Yazar bu çalışmada Twitter üzerinde Türkçe dilindeki sahte haberlerin tespiti üzerinde çalışma yapmıştır.

#### **b. Kullanılan Teknolojiler:**

- Denetimsiz öğrenme algoritmaları (K-ortalamlar, Non-Negative Matrix Factorization, Doğrusal Diskriminant Analizi)
- Denetimli öğrenme algoritmaları (K En Yakın Komşu, Destek Vektör Makinaları, Rassal Orman)

**c. Yöntem:** Twitter'dan Türkçe tweetlerin toplanması, denetimli ve denetimsiz makine öğrenimi algoritmalarının kullanımı ile analizi.

**d. Sonuç:** Denetimli öğrenme algoritmalarında 0.86 F1-metrik değeriyle başarılı sonuçlar alınmış, denetimsiz öğrenme algoritmalarının F1-metrik değeri 0.72'de kalmıştır.

## 8. Examining the Models Used for Fake News Detection in the Scope of Social Context

(KAYABAŞI KORU & ULUYOL, 2023)

- a. **Konu:** Yazar bu çalışmasında sosyal medya bağlamında sahte haber tespiti için kullanılan farklı modelleri incelemiştir.
- b. **Kullanılan Teknolojiler:**
  - Makine öğrenmesi
  - Derin öğrenme
  - Çeşitli algoritmalar (SVM, Random Forest, Decision Trees, Logistic Regression)
- c. **Yöntem:** Projede makine öğrenmesi ve derin öğrenme yaklaşımlarını kullanarak sahte haber tespiti modellerinin performans kıyaslamaları yapılmıştır.
- d. **Sonuç:** Makine öğrenmesi ve derin öğrenme modellerinin farklı veri kümelerindeki performans değerlendirmeleri; en iyi performans gösteren modellerin belirlenmesi.

## 9. Sahte Haber Tespiti için Derin Bağlamsal Kelime Gömülmeleri ve Sinirsel Ağların Performans Değerlendirmesi

(HARK, 2022)

- a. **Konu:** Yazar bu çalışmasında sahte haber tespiti için derin bağlamsal kelime gömülmeleri ve çeşitli sinirsel ağların performansının değerlendirilmesi üzerinde çalışma yapmıştır.
- b. **Kullanılan Teknolojiler:**
  - GloVe kelime gömülmeleri
  - Çok Katmanlı Algılayıcı (MLP)
  - Uzun Kısa Süreli Bellek (LSTM)
  - Yinelemeli Sinir Ağları (RNN) ve Evrişimli Sinir Ağları (CNN)
- c. **Yöntem:** COVID-19 sahte haber veri seti kullanılarak, GloVe ile ön-eğitilmiş metin gömülmeleri ve farklı sinirsel ağ modellerinin uygulanması yapılmıştır.
- d. **Sonuç:** LSTM modeli en yüksek sınıflandırma doğruluğuna (0.91) sahipken, MLP modeli en düşük performansı göstermiştir.

## 10. Hızlandırılmış Makine Öğrenmesi Algoritmaları İle Türkçe Sahte Haber Tespiti

(Danışmanı & Findik, n.d.)

a. **Konu:** Yazar bu çalışmasında Türkçe dilindeki sahte haberlerin hızlandırılmış makine öğrenmesi algoritmaları kullanılarak tespit edilmesi yönünde çalışma yapmıştır.

b. **Kullanılan Teknolojiler:**

- CatBoost
- AdaBoost
- Gradient Boosting
- XGBoost Algoritmaları

c. **Yöntem:** Makine öğrenmesi algoritmalarının kullanımı ve sahte haberlerin tespiti için modellerin geliştirilmesi.

d. **Sonuç:** Boosting algoritmalarının, diğer algoritmalara göre daha iyi sonuçlar verdiği belirtilmiştir.

## 11. Fake News vs Satire: A Dataset and Analysis

(Golbeck et al., 2018)

a. **Konu:** Yazar bu çalışmasında sahte haberler ile mizahi haberler arasındaki farkları inceleyen bir proje gerçekleştirmiştir.

b. **Kullanılan Teknolojiler:**

- El ile kodlanmış bir veri seti
- Tematik içerik analizi
- Naive Bayes tabanlı sınıflandırma modeli

c. **Yöntem:** Proje geliştirilirken ilk olarak 283 adet sahte haber ve 203 adet mizahi haberler arasındaki farkları incelemek için veri toplanmıştır. Daha sonrasında haberlerin tematik ve dil özellikleri üzerinden analizler yapılmıştır.

d. **Sonuç:** Bu proje sayesinde sahte haberler ve mizahi haberler arasında önemli dil ve tema farklılıkları saptanmıştır. Proje sonucunda Naive Bayes modeli, iki haber türünü %79.1 doğruluk oranıyla ayırt edilebilmiştir.

## 12. A comprehensive survey on machine learning approaches for fake news detection

(Alghamdi et al., 2023)

a. **Konu:** Yazar bu çalışmasında sahte haber tespiti için makine öğrenimi ve derin öğrenme yöntemlerinin kapsamlı bir şekilde incelemesini gerçekleştirmiştir.

b. **Kullanılan Teknolojiler:**

- Doğal Dil İşleme (NLP)



- Makine öğrenimi (ML)
  - Derin Öğrenme (DL)
  - Transformer tabanlı modeller
- c. **Yöntem:** Projede sahte haber ilk olarak sahte haber tanımları ve ilgili terimler incelenmiştir. Ardından içerik, bağlam ve hibrit tabanlı özellik kategorilerine odaklanarak ML ve DL tekniklerini detaylı analizleri yapılmıştır.
- d. **Sonuç:** Projenin sonucu olarak mevcut sahte haber tespiti çalışmalarının zorlukları ve gelecekteki araştırmalara yönelik öneriler geliştirilmiştir. Sahte haber tespiti için etkili mekanizmaların geliştirilmesine yönelik kapsamlı bir genel bakış sunumu yapılmıştır.

### 13. Analysis of Fake News Detection Methods

(Amjad et al., 2023)

- a. **Konu:** Yazar bu çalışmada Urduca sahte haberleri tespit etmek için makine ve derin öğrenme yaklaşımlarının analizini yapmıştır.
- b. **Kullanılan Teknolojiler:**
- Kelime vektör gömme
  - Geleneksel sinir ağları
  - XLNet
  - BERT
- c. **Yöntem:** Projede ilk olarak Urduca sahte haberlerin sınıflandırılması için veri setlerinin toplanması ve işaretlenmesi gerçekleştirilmiştir. Ardından farklı algoritmaların analizi ve sınıflandırıcıların değerlendirmeleri sunulmuştur.
- d. **Sonuç:** Projenin sonucunda 2021’de en iyi performansı gösteren sistem lineer sınıflandırıcı kullanırken, 2020’de en yüksek performansı BERT tabanlı sistem göstermiştir. Kontekstsel temsiller ve büyük sinir ağları, standart özellik tabanlı modellere göre daha iyi performans sergilemektedir.

### 14. Augmented Fake News Detection Model Using Machine Learning

(Farha, 2023)

- a. **Konu:** Yazar bu çalışmada sahte haberlerin tespiti için gelişmiş bir makine öğrenimi modelini geliştirmiştir.
- b. **Kullanılan Teknolojiler:**
- Naive Bayes

- Decision Tree
  - Multilayer Perceptron (MLP)
  - TF-IDF Vektörizer
- c. **Yöntem:** Projede ilk olarak veri seti hazırlanmıştır. Veri seti Kaggle sitesinden indirilmiştir. Ardından veri setinin işlenmesi, öznitelik çıkarımı ve sınıflandırma algoritmaları kullanarak model eğitimi yapılmıştır.
- d. **Sonuç:** Modelin, sahte haberleri etkili bir şekilde tespit edebildiği ve yüksek doğruluk oranına sahip olduğu sonucuna varılmıştır.

## 15. BanFakeNews: A Dataset for Detecting Fake News in Bangla

(Hossain et al., 2020)

- a. **Konu:** Yazar bu çalışmada Bangla dilinde yazılmış yaklaşık 50.000 haber içeren etiketli bir veri seti oluşturularak, düşük kaynaklı dillerde sahte haber tespiti için bir sistem geliştirilmesi hedeflenmiştir.
- b. **Kullanılan Teknolojiler:**
- Çeşitli nöral ağ modelleri
- c. **Yöntem:** Makine öğrenmesi algoritmalarının kullanımı ve sahte haberlerin tespiti için modellerin geliştirilmesi.
- d. **Sonuç:** Makale, haber makalelerinden çıkarılan dilsel özellikler ve bu özellikleri kullanarak sahte haberleri sınıflandıran makine öğrenimi modellerinin geliştirilmesi detaylanmaktadır.

## 16. Fake News Detection on Social Networks: A Survey

(Shen et al., 2023)

- a. **Konu:** Yazar bu çalışmada sosyal ağlarda sahte haberleri tespit etmek için kullanılan yöntemlerin, teknolojilerinin ve yaklaşımların kapsamlı bir analizini sunmuştur.
- b. **Kullanılan Teknolojiler:**
- Yapay Zeka (AI)
  - Makine Öğrenimi (ML)
  - Derin Öğrenme (DL)
- c. **Yöntem:** Sahte haber tespiti için kullanılan çeşitli yöntemler ve modellerin sınıflandırılmasını ve karşılaştırılmasını içerir. Bu, sahte haberlerin özelliklerini ve yayılım biçimlerini analiz etmeyi kapsar.

- d. Sonuç:** Projenin sonucunda sahte haber tespitinde içerik, yayılım ve kaynak tabanlı yaklaşımların her birinin avantajları ve dezavantajları sunulmuştur. Ayrıca, bu alanda ileriye yönelik araştırmalar için öneriler ve gelecek vizyonlar ele alınmışlardır.

## 17. Fake News Detection Using LSTM in TensorFlow and Deep Learning

(Singh & Singh, 2023)

- a. Konu:** Yazar bu çalışmasında LSTM (Uzun Kısa Süreli Bellek), Bidirectional LSTM (Çift Yönlü LSTM) ve Convolutional Neural Network (CNN) tabanlı derin öğrenme algoritmalarını kullanarak sosyal medyada sahte haberleri tespit etmeyi amaçlamaktadır.
- b. Kullanılan Teknolojiler:**
- LSTM
  - Bidirectional LSTM
  - CNN
  - TensorFlow
- c. Yöntem:** Sahte haberleri tespit etmek için, farklı veri setleri üzerinde LSTM, BiLSTM ve CNN modelleri eğitilmiş ve bu modellerin performansı değerlendirilmiştir.
- d. Sonuç:** Yapılan bu çalışmada farklı veri setlerinde uygulanan modellerin performans analizlerini sunmuştur. CNN-BiLSTM modelinin diğer modellere göre daha yüksek doğruluk ve F1 skoru ile üstün olduğunu ortaya koymaktadır.

## 18. Fake News Detection: A Study

(Trivedi et al., 2023)

- a. Konu:** Yazar bu çalışmasında sosyal medyada sahte haberleri tespit etmek için kullanılan çeşitli teknik ve yaklaşımları incelemiştir.
- b. Kullanılan Teknolojiler:**
- Makine Öğrenimi (ML)
  - Derin Öğrenme (DL)
- c. Yöntem:** Projede sırası ile veri toplama işlemi, ön işleme, özellik çıkarma ve farklı öğrenme teknikleri uygulanmıştır.
- d. Sonuç:** Sahte haber tespitinde en iyi sonuçları, farklı öğrenme tekniklerini birleştiren ansambl yöntemleri ile elde edildiği tespit edilmiştir.

## 19. IFND: a benchmark dataset for fake news detection

(Sharma & Garg, 2023)

- a. **Konu:** Yazar bu çalışmada sahte haber tespiti için büyük ölçekli bir Hint veri seti olan IFND'nin geliştirilmesini ve tanıtımını amaçlamaktadır.
- b. **Kullanılan Teknolojiler:**
  - Makine Öğrenimi (ML)
  - Derin Öğrenme (DL)
  - LSTM
  - Bi-LSTM
  - VGG16
  - Resnet-50
- c. **Yöntem:** Veri seti oluşturulurken, Parsehub aracı kullanılarak çeşitli haber sitelerinden veriler toplanmış ve bir metin arttırma algoritması uygulanarak sahte haber miktarı artırılmıştır. Ayrıca, LDA konu modellemesi uygulanarak haberler kategorize edilmiştir.
- d. **Sonuç:** IFND veri setinin, sahte haber tespitinde etkili olduğu çeşitli makine öğrenimi ve derin öğrenme modelleri üzerinde iyi sonuçlar verdiği belirtilmiştir.

## 20. A systemic literature overview of Fake News Challenge (FNC-1) dataset and its use in fake news detection schemes

(Jawad & Obaid, 2023)

- a. **Konu:** Yazar bu çalışmada FNC-1 veri setinin sahte haber tespitinde kullanımının sistematik bir incelemesini sunmaktadır.
- b. **Kullanılan Teknolojiler:**
  - Doğal Dil İşleme (NLP)
  - Yapay Zeka (AI)
  - Makine Öğrenimi (ML)
  - Derin Öğrenme (DL)
- c. **Yöntem:** FNC-1 veri seti kullanılarak farklı sahte haber tespiti tekniklerinin ve yaklaşımlarının analizleri yapılmıştır.
- d. **Sonuç:** FNC-1 veri setiyle yapılan çalışmaların değerlendirilmesi ve sahte haber tespitindeki mevcut yöntemlerin etkinliği sorgulanmıştır. Sahte haber tespitinde daha ileri teknolojilere ve yöntemlere ihtiyaç olduğu vurgulanmıştır.

## KULLANILAN TEKNOLOJİLER VE YÖNTEM

Literatür taramalarından edinilen bulgular sonucunda kullanılan teknolojiler ve yöntem konusunda genel bilgilendirmeye sahip olunmuş olup, literatür taramalarında kullanılan teknoloji ve yöntemlere ek olarak farklı teknolojilere ve yöntemlerin yapay zeka üzerinde etkisinin araştırılmasına yer verilmiştir.

### Araştırmanın Amacı ve Türü

Araştırma basit makine öğrenimi projesinde farklı eğitim modelleri kullanılarak sahte haber veri seti üzerindeki etkisini araştırmaktır. Eğitim ve test verileri üzerinden modelleri teste sokarak sınıflandırma parametreleriyle çeşitli bulgular üzerinde tartışılarak sonuca varılması hedeflenmektedir.

### Verilerin Toplanması ve Analizi

Kaggle sitesindeki “Fake News” adlı veri seti tercih edildi. Veri seti UTM makine öğrenimi kulübü tarafından hazırlanmış olup 26 Ocak 2018 yılında veri seti hazırlanmaya başlanıp 23 Şubat 2018 yılında geliştirilmesi sonlanmıştır. Veri seti toplam 20800 adet veriden oluşmaktadır.

“Fake News” Veri Seti Sütunları				
Id	Title	Author	Text	Label

**Tablo 1.1** “Fake News” Veri Seti Sütunları

<b>“Fake News” Veri Setindeki Sütunların Anlamı</b>	
<b>Sütunlar</b>	<b>Anlamları</b>
Id	Numaralandırma
Title	Haberlerin başlığını içerir
Author	Haberi yazan kişinin isim ve soy isim bilgisini içerir
Text	Haberin metnini içerir
Label	Eğitim verisi için; etiket bilgisini içerir 0 => Güvenilir 1 => Güvenilmez

**Tablo 1.2** “Fake News” Veri Setindeki Sütunların Anlamı

## Kullanılan Teknolojiler

İçinde bulunan çeşitli kütüphaneler, geliştirmeye açık olması ve popülaritesi göz önünde bulundurularak Python programlama dili ile makine öğrenimi yapılmıştır. Visual Studio Code içinde; yapay zeka için büyük artıları bulunan Jupyter Notebook eklentisi kurularak tümleşik geliştirme ortamı hazırlanmıştır.

## Yöntem

Veri seti eğitim ve test verisi olarak ayrılarak, vektörize edildi. Özellik seçme yapıldıktan sonra K-en yakın komşu (KNN), Lojistik Regresyon, Karar Ağacı, Destek Vektör Makinesi (SVM), Sinir Ağı modelleri kullanılarak eğitildi. Eğitilen modelde önce eğitim verisi daha sonra test verisi üzerinden sonuçlar karşılaştırıldı.

## BULGULAR VE TARTIŞMA

Kullanılan yöntem eşliğinde performans parametreleri göz önünde bulunduruldu. Her model için hem eğitim hem de test verileri üzerinden model testi gerçekleştirildi ve karmaşıklık matrisi, doğruluk değeri, duyarlılık, özgüllük, kappa, f-ölçümü, auc ve roc eğrisi sonuçlarına erişildi.

Etiket	Dağılım Sayısı
Label = 0	10987
Label = 1	10413

**Tablo 2.1** Veri Setinin Tür Değişkenine Göre Dağılımı

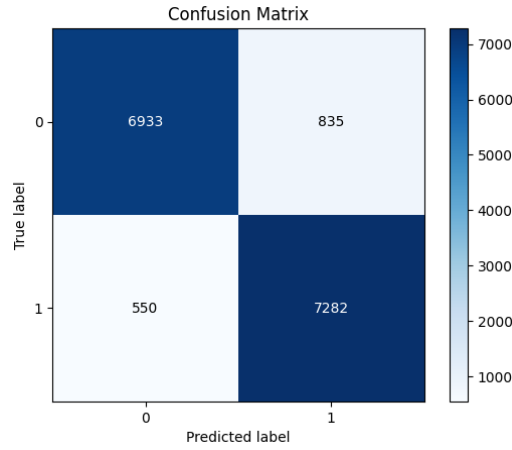
## KNN Modeli – Eğitim Verisi Testi

TN	FP	FN	TP
6933	835	550	7282

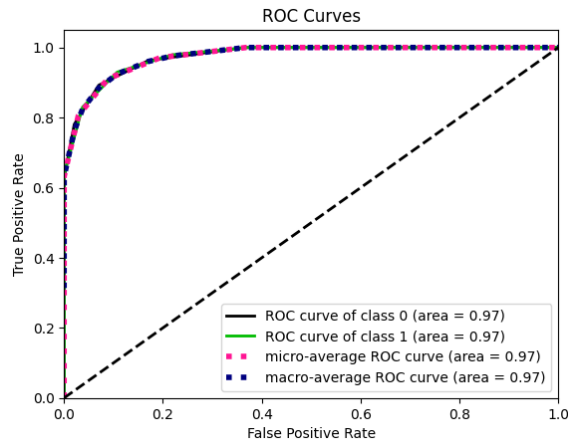
**Tablo 2.1.1** KNN Modeli Eğitim Verisi Testi Karmaşıklık Matrisi Değerleri

Performans Parametresi	Sonuç
Doğruluk Değeri	0.911121
Duyarlılık	0.929775
Özgüllük	0.892507
Kappa	0.822406
F-ölçümü	0.910760
AUC	0.911141

**Tablo 2.1.2** KNN Modeli Eğitim Verisi Testi Performans Parametreleri



**Şekil 2.1.1** KNN Modeli Eğitim Verisi Testi Karmaşıklık Matrisi



**Şekil 2.1.2** KNN Modeli Eğitim Verisi Testi ROC Eğrisi



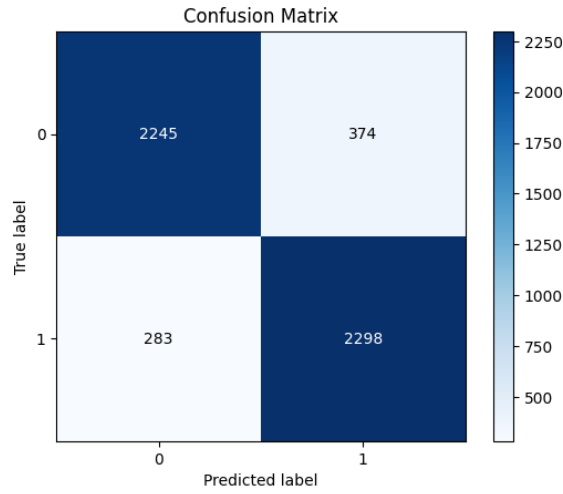
## KNN Modeli – Test Verisi Testi

TN	FP	FN	TP
2245	374	283	2298

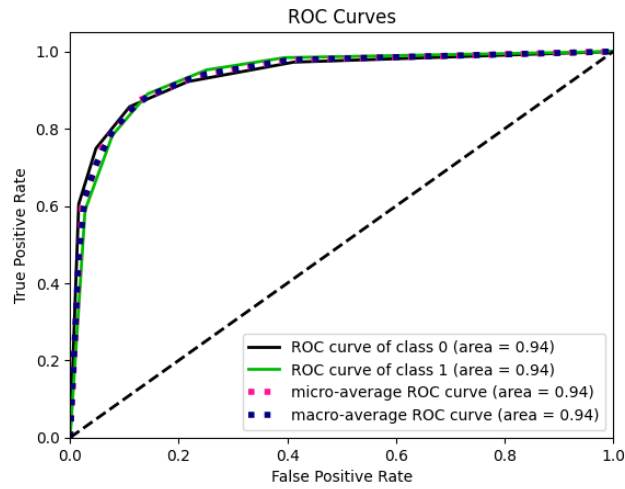
**Tablo 2.2.1** KNN Modeli Test Verisi Testi Karmaşıklık Matrisi Değerleri

Performans Parametresi	Sonuç
Doğruluk Değeri	0.873653
Duyarlılık	0.890352
Özgüllük	0.877197
Kappa	0.747358
F-ölçümü	0.873460
AUC	0.873774

**Tablo 2.2.2** KNN Modeli Test Verisi Test Performans Parametreleri



**Şekil 2.2.1** KNN Modeli Test Verisi Testi Karmaşıklık Matrisi



**Şekil 2.2.2** KNN Modeli Test Verisi Testi ROC Eğrisi

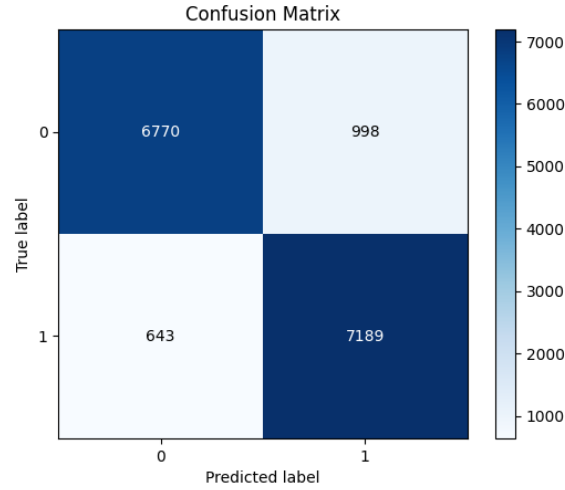
### Lojistik Regreasyon Modeli – Eğitim Verisi Testi

TN	FP	FN	TP
6770	998	643	7189

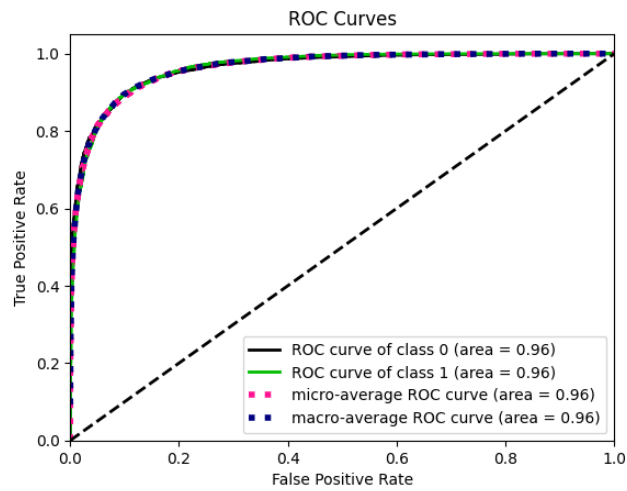
**Tablo 2.3.1** Lojistik Regreasyon Modeli Eğitim Verisi Testi Karmaşıklık Matrisi Değerleri

Performans Parametresi	Sonuç
Doğruluk Değeri	0.894807
Duyarlılık	0.917900
Özgüllük	0.871524
Kappa	0.789525
F-ölçümü	0.894111
AUC	0.894712

**Tablo 2.3.2** Lojistik Regreasyon Modeli Eğitim Verisi Testi Performans Parametreleri



**Şekil 2.3.1** Lojistik Regreasyon Modeli Eğitim Verisi Testi Karmaşıklık Matrisi



**Şekil 2.3.2** Lojistik Regreasyon Modeli Eğitim Verisi Testi ROC Eğrisi

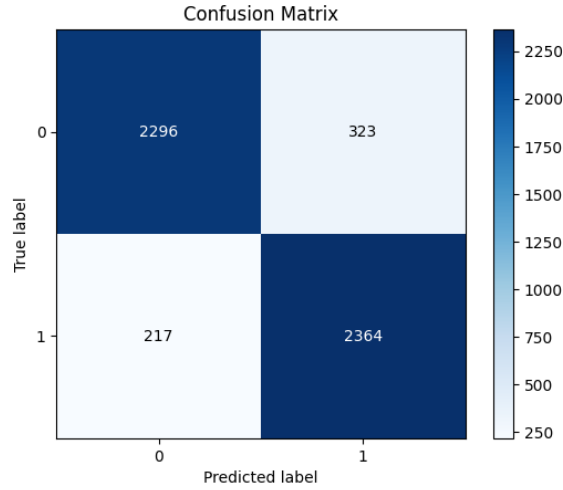
## Lojistik Regreasyon Modeli – Test Verisi Testi

TN	FP	FN	TP
2296	323	217	2364

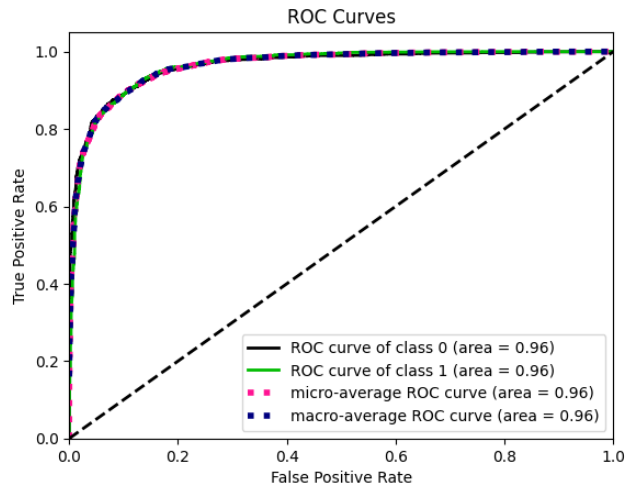
**Tablo 2.4.1** Lojistik Regreasyon Modeli Test Verisi Testi Karmaşıklık Matrisi Değerleri

Performans Parametresi	Sonuç
Doğruluk Değeri	0.896153
Duyarlılık	0.915924
Özgüllük	0.876670
Kappa	0.792358
F-ölçümü	0.895867
AUC	0.896297

**Tablo 2.4.2** Lojistik Regreasyon Modeli Test Verisi Testi Performans Parametreleri



**Şekil 2.4.1** Lojistik Regreasyon Modeli Test Verisi Testi Karmaşıklık Matrisi



**Şekil 2.4.2** Lojistik Regreasyon Modeli Test Verisi Testi ROC Eğrisi

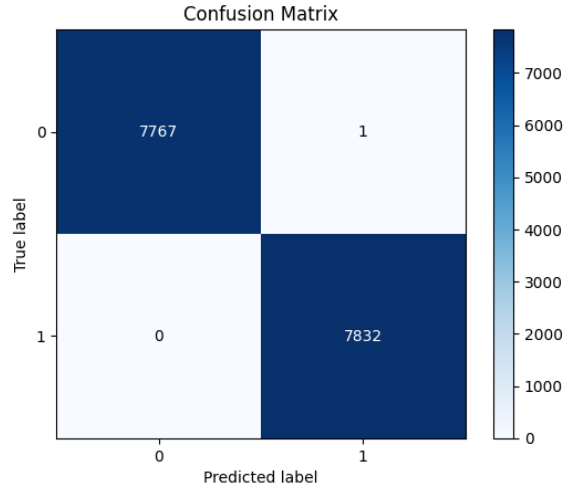
## Karar Ağacı Modeli – Eğitim Verisi Testi

TN	FP	FN	TP
7767	1	0	7832

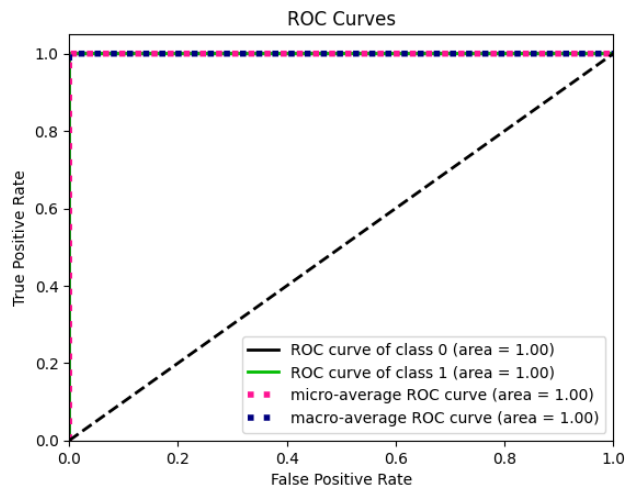
**Tablo 2.5.1** Karar Ağacı Modeli Eğitim Verisi Testi Karmaşıklık Matrisi Değerleri

Performans Parametresi	Sonuç
Doğruluk Değeri	0.999935
Duyarlılık	1.0
Özgüllük	0.999871
Kappa	0.999871
F-ölçümü	0.999935
AUC	0.999935

**Tablo 2.5.2** Karar Ağacı Modeli Eğitim Verisi Testi Performans Parametreleri



**Şekil 2.5.1** Karar Ağacı Modeli Eğitim Verisi Testi Karmaşıklık Matrisi



**Şekil 2.5.2** Karar Ağacı Modeli Eğitim Verisi Testi ROC Eğrisi

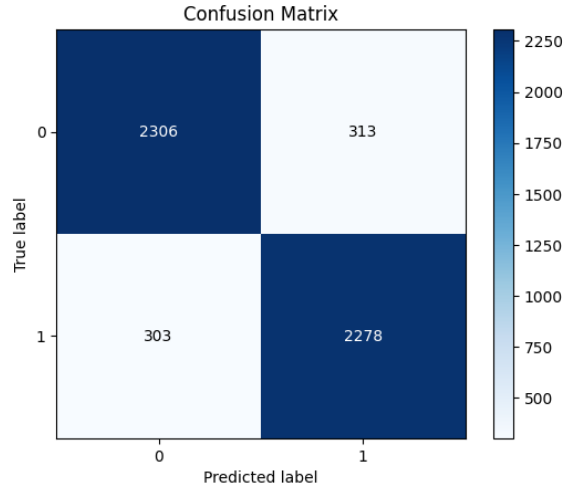
## Karar Ağacı Modeli – Test Verisi Testi

TN	FP	FN	TP
2306	313	303	2278

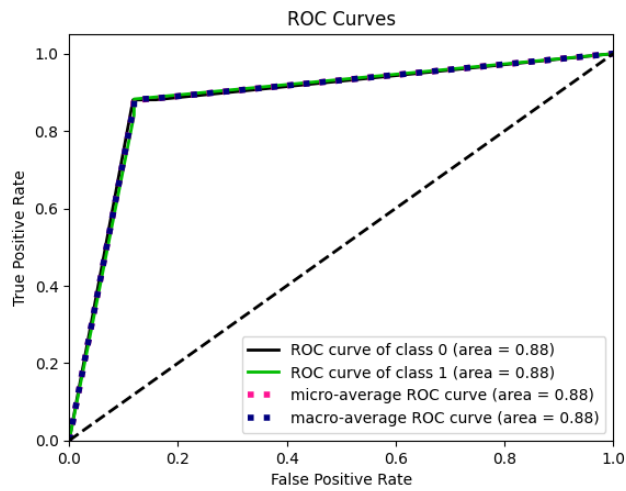
**Tablo 2.6.1** Karar Ağacı Modeli Test Verisi Testi Karmaşıklık Matrisi Değerleri

Performans Parametresi	Sonuç
Doğruluk Değeri	0.881538
Duyarlılık	0.882603
Özgüllük	0.880488
Kappa	0.763070
F-ölçümü	0.881544
AUC	0.881546

**Tablo 2.6.2** Karar Ağacı Modeli Test Verisi Testi Performans Parametreleri



**Şekil 2.6.1** Karar Ağacı Modeli Test Verisi Testi Karmaşıklık Matrisi



**Şekil 2.6.2** Karar Ağacı Modeli Test Verisi Testi ROC Eğrisi

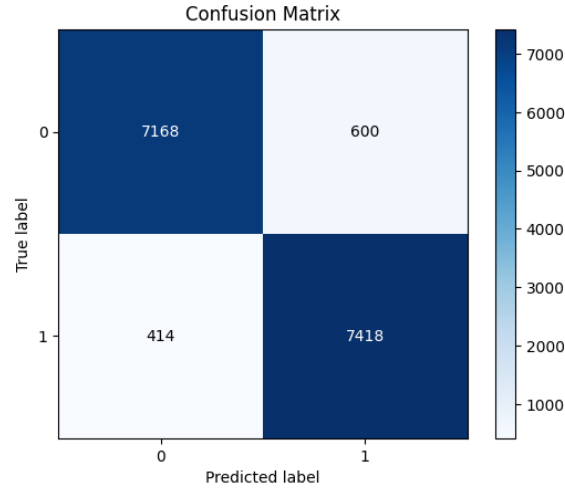
## SVM Modeli – Eğitim Verisi Testi

TN	FP	FN	TP
7168	600	414	7418

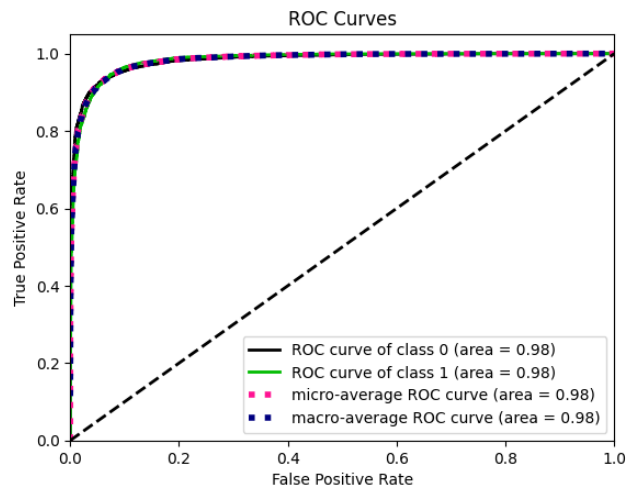
**Tablo 2.7.1** SVM Modeli Eğitim Verisi Testi Karmaşıklık Matrisi Değerleri

Performans Parametresi	Sonuç
Doğruluk Değeri	0.935
Duyarlılık	0.947139
Özgüllük	0.922760
Kappa	0.869985
F-ölçümü	0.934791
AUC	0.934949

**Tablo 2.7.2** SVM Modeli Eğitim Verisi Testi Performans Parametreleri



**Şekil 2.7.1** SVM Modeli Eğitim Verisi Testi Karmaşıklık Matrisi



**Şekil 2.7.2** SVM Modeli Eğitim Verisi Testi ROC Eğrisi

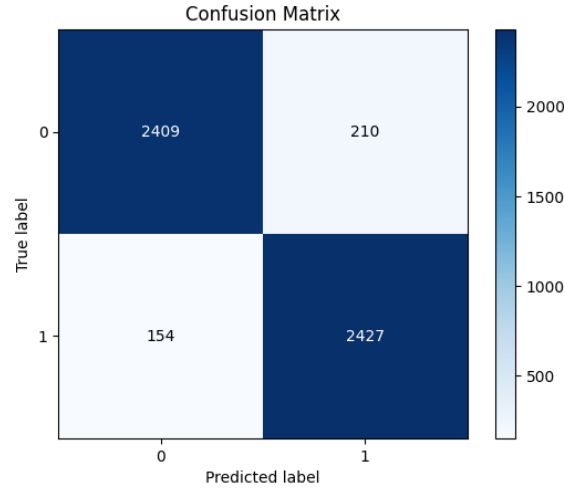
## SVM Modeli – Test Verisi Testi

TN	FP	FN	TP
2409	210	154	2427

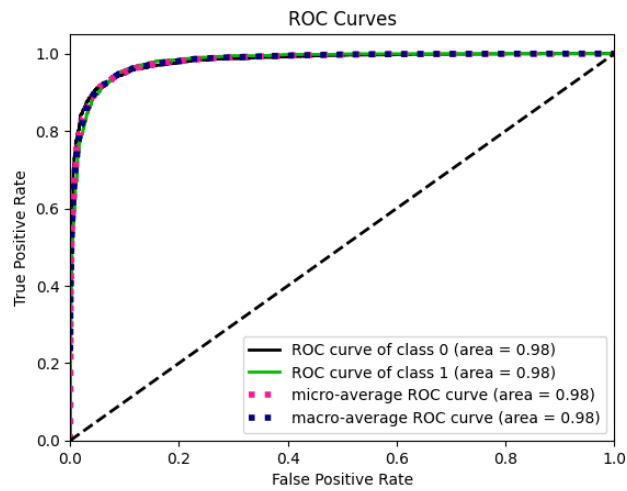
**Tablo 2.8.1** SVM Modeli Test Verisi Testi Karmaşıklık Matrisi Değerleri

Performans Parametresi	Sonuç
Doğruluk Değeri	0.93
Duyarlılık	0.940333
Özgüllük	0.919816
Kappa	0.860014
F-ölçümü	0.929961
AUC	0.930074

**Tablo 2.8.2** SVM Modeli Test Verisi Testi Performans Parametreleri



**Şekil 2.8.1** SVM Modeli Test Verisi Testi Karmaşıklık Matrisi



**Şekil 2.8.2** SVM Modeli Test Verisi Testi ROC Eğrisi

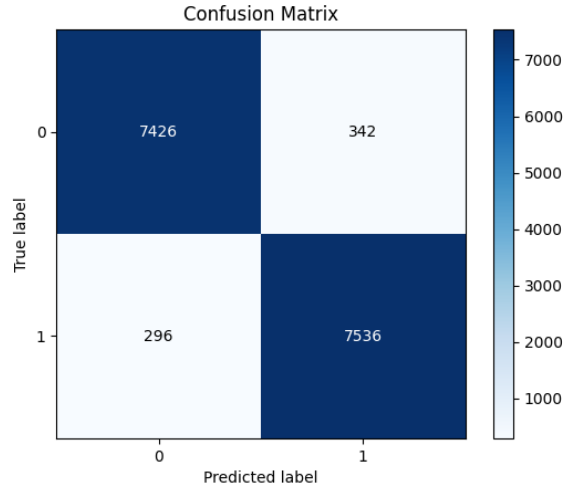
## Sinir Ağı Modeli – Eğitim Verisi Testi

TN	FP	FN	TP
7426	342	296	7536

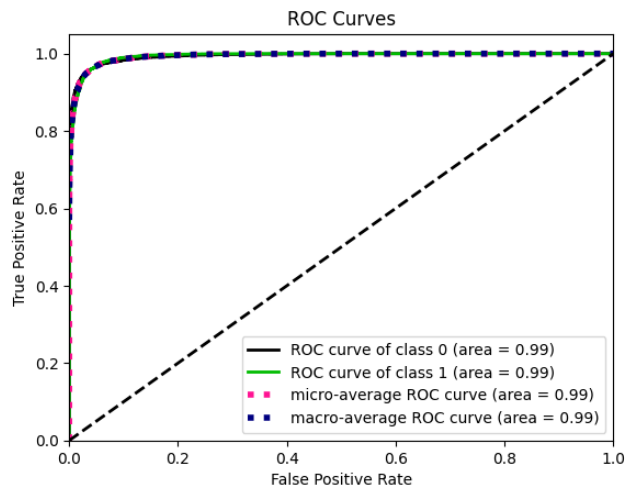
**Tablo 2.9.1** Sinir Ağı Modeli Eğitim Verisi Testi Karmaşıklık Matrisi Değerleri

Performans Parametresi	Sonuç
Doğruluk Değeri	0.959102
Duyarlılık	0.962206
Özgüllük	0.955973
Kappa	0.918201
F-ölçümü	0.959079
AUC	0.959089

**Tablo 2.9.2** Sinir Ağı Modeli Eğitim Verisi Testi Performans Parametreleri



**Şekil 2.9.1** Sinir Ağı Modeli Eğitim Verisi Testi Karmaşıklık Matrisi



**Şekil 2.9.2** Sinir Ağı Modeli Eğitim Verisi Testi ROC Eğrisi



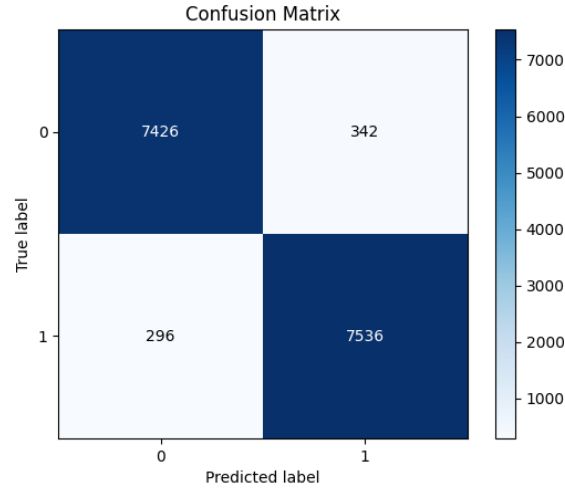
## Sinir Ağı Modeli – Test Verisi Testi

TN	FP	FN	TP
2415	204	125	2456

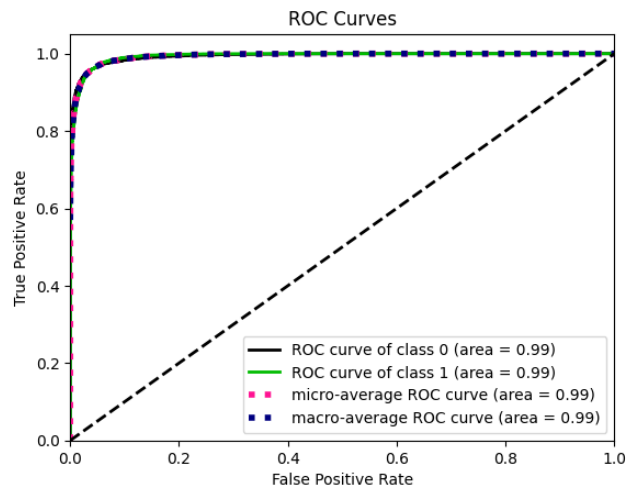
**Tablo 2.10.1** Sinir Ağı Modeli Test Verisi Testi Karmaşıklık Matrisi Değerleri

Performans Parametresi	Sonuç
Doğruluk Değeri	0.936730
Duyarlılık	0.951569
Özgüllük	0.922107
Kappa	0.873482
F-ölçümü	0.936606
AUC	0.936838

**Tablo 2.10.2** Sinir Ağı Modeli Test Verisi Testi Performans Parametreleri



**Şekil 2.10.1** Sinir Ağı Modeli Test Verisi Testi Karmaşıklık Matrisi

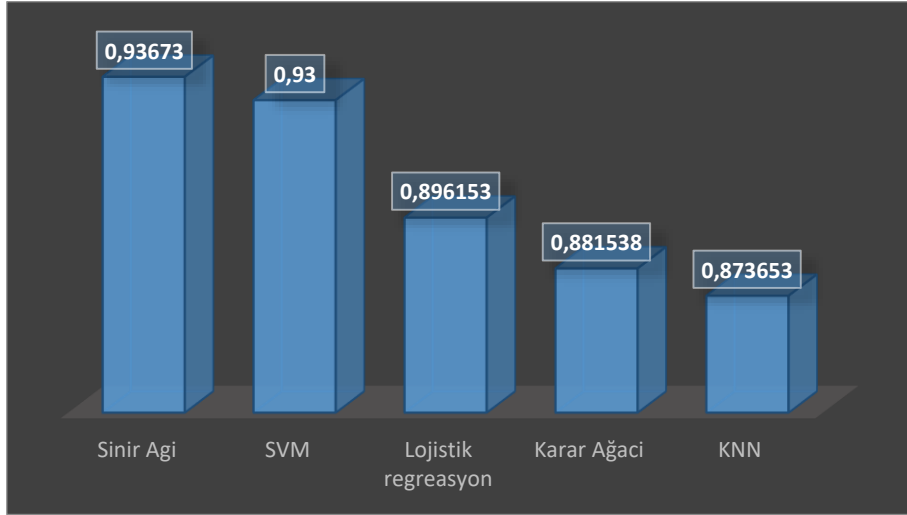


**Şekil 2.10.2** Sinir Ağı Modeli Test Verisi Testi ROC Eğrisi

K-en yakın komşu (KNN), Lojistik Regreasyon, Karar Ağacı, Destek Vektör Makinesi (SVM) ve Sinir Ağı modellerinin eğitim ve test verileri ile test edilmesi sonucunda bulunan bulgular eşliğinde sinir ağı ve svm modelinin en iyi doğruluk sonucuna ulaştığı anlaşılmıştır.

Sıralama	Model	Değer
1	Sinir Ağı	0.936730
2	SVM	0.93
3	Lojistik Regreasyon	0.896153
4	Karar Ağacı	0.881538
5	KNN	0.873653

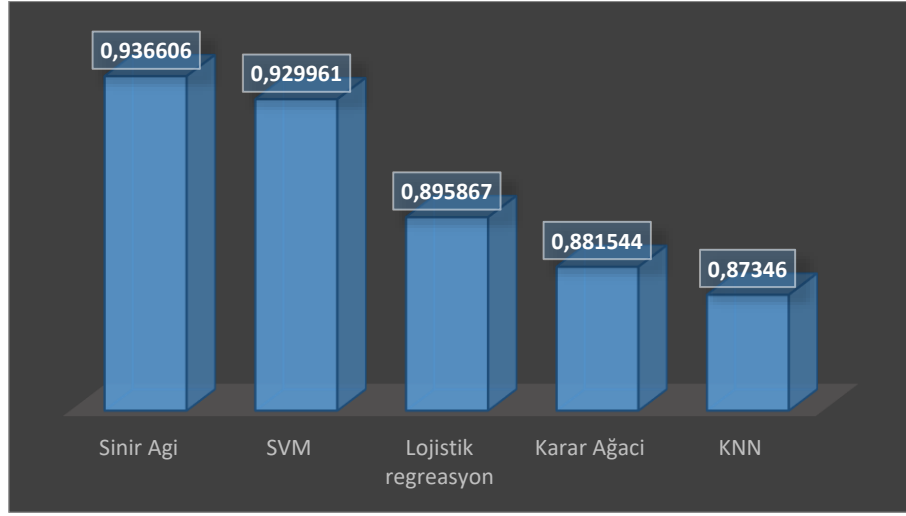
**Tablo 2.11** Modellerin Test Verisi Testi Doğruluk Değerine Göre Sıralaması



**Şekil 2.11** Modellerin Test Verisi Doğruluk Grafiği

Sıralama	Model	Değer
1	Sinir Ağı	0.936606
2	SVM	0.929961
3	Lojistik Regreasyon	0.895867
4	Karar Ağacı	0.881544
5	KNN	0.873460

**Tablo 2.12** Modellerin Test Verisi Testi F-Ölçümü Değerine Göre Sıralaması



Şekil 2.12 Modellerin Test Verisi Testi F-Ölçümü Grafiği

## SONUÇ VE ÖNERİLER

Bu çalışmada, sahte haber tespiti için çeşitli makine öğrenimi modellerinin uygulanması ve karşılaştırılması yapılmıştır. K-en yakın komşu (KNN), Lojistik Regreasyon, Karar Ağacı, Destek Vektör Makinesi (SVM) ve Sinir Ağı gibi farklı modellerin performansları değerlendirilmiştir. Elde edilen bulgular, her bir modelin güçlü ve zayıf yönlerini ortaya koymakla birlikte sahte haber tespitindeki etkinliklerini belirlemektedir. Sinir ağı modeli 0.936730 doğruluk ve 0.936606 F skoruyla en verimli çalışan model olduğu sonucuna varılmıştır. Bu çalışma, sahte haber tespiti konusunda makine öğrenimi yaklaşımlarının potansiyelini göstermekte ve gelecekteki araştırmalara yol gösterici olmaktadır. Sonuç olarak, makine öğrenimi modellerinin, sahte haberlerin tespitinde etkili olduğu anlaşılmıştır.

Bu çalışmanın sonuçları, sahte haber tespitinde makine öğreniminin potansiyelini ortaya koymaktadır. Daha geniş ve çeşitli veri setleri kullanarak modellerin doğruluğunu arttırmak, derin öğrenme teknikleri ve ансамbl modelleri gibi daha gelişmiş yöntemlerin keşfedilmesi, farklı dillerde ve kültürel bağlamlarda sahte haber tespiti için model adaptasyonları yapmak gelecekteki araştırmalar için öneri olarak sunulmaktadır.

## Kaynakça

- Alghamdi, J., Luo, S., & Lin, Y. (2023). A comprehensive survey on machine learning approaches for fake news detection. *Multimedia Tools and Applications*.  
<https://doi.org/10.1007/s11042-023-17470-8>
- Amjad, M., Vitman, O., Sidorov, G., Zhila, A., & Gelbukh, A. (2023). Analysis of Fake News Detection Methods. In *Studies in Fuzziness and Soft Computing* (Vol. 423, pp. 131–144). Springer Science and Business Media Deutschland GmbH. [https://doi.org/10.1007/978-3-031-23476-7\\_13](https://doi.org/10.1007/978-3-031-23476-7_13)
- Bobulski, J. (2024). A Blockchain System for Fake News Detection. *Lecture Notes in Electrical Engineering, 1112 LNEE*, 1–11. [https://doi.org/10.1007/978-981-99-8211-0\\_1](https://doi.org/10.1007/978-981-99-8211-0_1)
- Bsoul, M. A., Qusef, A., & Abu-Soud, S. (2022). Building an Optimal Dataset for Arabic Fake News Detection. *Procedia Computer Science, 201(C)*, 665–672.  
<https://doi.org/10.1016/j.procs.2022.03.088>
- Danışmanı, T., & Findik, O. (n.d.). HIZLANDIRILMIŞ MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE TÜRKÇE SAHTE HABER TESPİTİ Elif YILDIRIM 2022 YÜKSEK LİSANS TEZİ BİLGİSAYAR MÜHENDİSLİĞİ.
- Farha, A. (2023). Augmented Fake News Detection Model Using Machine Learning. *Article in International Journal of Engineering and Management Research, 13(3)*.  
<https://doi.org/10.31033/ijemr.13.3.27>
- Golbeck, J., Mauriello, M., Auxier, B., Bhanushali, K. H., Bonk, C., Bouzaghrane, M. A., Buntain, C., Chanduka, R., Cheakalos, P., Everett, J. B., Falak, W., Gieringer, C., Graney, J., Hoffman, K. M., Huth, L., Ma, Z., Jha, M., Khan, M., Kori, V., ... Visnansky, G. (2018). Fake news vs satire: A dataset and analysis. *WebSci 2018 - Proceedings of the 10th ACM Conference on Web Science*, 17–21. <https://doi.org/10.1145/3201064.3201100>
- HARK, C. (2022). Sahte Haber Tespiti için Derin Bağlamsal Kelime Gömümleri ve Sinirsel Ağların Performans Değerlendirmesi. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi, 34(2)*, 733–742. <https://doi.org/10.35234/fumbd.1126688>
- Hossain, M. Z., Rahman, M. A., Islam, M. S., & Kar, S. (2020). *BanFakeNews: A Dataset for Detecting Fake News in Bangla*. <http://arxiv.org/abs/2004.08789>
- Jawad, Z. A., & Obaid, A. J. (2023). A systemic literature overview of Fake News Challenge (FNC-1) dataset and its use in fake news detection schemes. *Journal of Discrete Mathematical Sciences and Cryptography, 26(4)*, 1197–1206. <https://doi.org/10.47974/JDMSC-1567>
- KAYABAŞI KORU, G., & ULUYOL, Ç. (2023). Sahte Haber Tespiti için Kullanılan Modellerin Sosyal Bağlam Kapsamında İncelenmesi. *Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji, 11(1)*, 39–54. <https://doi.org/10.29109/gujsc.1145516>
- KAYAKUŞ, M., & YİĞİT AÇIKGÖZ, F. (2023). Twitter’da Makine Öğrenmesi Yöntemleriyle Sahte Haber Tespiti. *Abant Sosyal Bilimler Dergisi, 23(2)*, 1017–1027.  
<https://doi.org/10.11616/asbi.1266179>

- Pöldvere, N., Uddin, Z., & Thomas, A. (2023). The PolitiFact-Oslo Corpus: A New Dataset for Fake News Analysis and Detection. *Information*, 14(12), 627. <https://doi.org/10.3390/info14120627>
- Salh, D. A., & Nabi, R. M. (2023). Kurdish Fake News Detection Based on Machine Learning Approaches. *Passer Journal of Basic and Applied Sciences*, 5(2), 262–271. <https://doi.org/10.24271/PSR.2023.380132.1226>
- Sharma, D. K., & Garg, S. (2023). IFND: a benchmark dataset for fake news detection. *Complex and Intelligent Systems*, 9(3), 2843–2863. <https://doi.org/10.1007/s40747-021-00552-1>
- Shen, Y., Liu, Q., Guo, N., Yuan, J., & Yang, Y. (2023). Fake News Detection on Social Networks: A Survey. *Applied Sciences*, 13(21), 11877. <https://doi.org/10.3390/app132111877>
- Singh, Y., & Singh, P. (2023). TensorFlow and Deep Learning. *Journal of Applied Sci-Ence and Education (JASE)*, 03(005), 1–14. <https://doi.org/10.54060/jase.v3i2>
- TAŞKIN, S. G., KÜÇÜKSİLLE, E. U., & TOPAL, K. (2021). Twitter üzerinde Türkçe sahte haber tespiti. *Balıkesir Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 23(1), 151–172. <https://doi.org/10.25092/baunfbed.843909>
- Tazeze, T., & R, R. (2021). Building a Dataset for Detecting Fake News in Amharic Language. *International Journal of Advanced Research in Science, Communication and Technology*, 76–83. <https://doi.org/10.48175/ijarsct-1362>
- Trivedi, S., Jain, M. K., Gopalani, D., Meena, Y. K., & Gupta, Y. (2023). *Fake News Detection: A Study* (pp. 395–408). [https://doi.org/10.1007/978-981-99-1373-2\\_31](https://doi.org/10.1007/978-981-99-1373-2_31)