A black background with a black square

Description automatically generated with medium confidence

**MÜHENDİSLİK MİMARLIK FAKÜLTESİ**

**BİLGİSAYAR MÜHENDİSLİĞİ**

**BİTİRME PROJESİ VİZE RAPORU**

**ÖĞRETİM GÖREVLİSİ**

Prof. Dr. YAŞAR HOŞCAN

**HAZIRLAYANLAR**

ALİMCAN ÇAKIR / 20212013060

ERHAN TARHANA / 20212013035

ÖMER FARUK ŞENOL / 20212013044

İçindekiler

[1. Proje Ön Tanıtım 3](#_Toc184304844)

[2. Giriş 3](#_Toc184304845)

[a. Proje Amacı 3](#_Toc184304846)

[b. Problem Tanımı 3](#_Toc184304847)

[**c.** Projenin Önemi 3](#_Toc184304848)

[3. Kullanılan Teknolojiler ve Araçlar 4](#_Toc184304849)

[a. Scrum Yönetimi: 4](#_Toc184304850)

[b. Python Kütüphaneleri: 4](#_Toc184304851)

[c. IDE ve Editörler 4](#_Toc184304852)

[4. Yöntem ve Çalışma Süreci 5](#_Toc184304853)

[a. Scrum Süreci 5](#_Toc184304854)

[b. Veri Toplama 5](#_Toc184304855)

[c. Veri Analizi ve Modelleme 5](#_Toc184304856)

# Proje Ön Tanıtım

**Proje Sahipleri:** Alimcan Çakır, Erhan Tarhana, Ömer Faruk Şenol

**Proje Konusu:** Yapay Zeka Destekli Haber Doğruluk Analizi ve Tespiti

# Giriş

## Proje Amacı

Dijital çağda bilgiye ulaşımın kolaylaşması, doğru ve güvenilir haberlerin yanı sıra yanlış ve yanıltıcı içeriklerin de hızla yayılmasına neden olmaktadır. Bu durum, bireylerin doğru kararlar almasını engelleyerek toplumsal kutuplaşma, ekonomik zararlar ve güven kaybı gibi olumsuz sonuçlar doğurabilir. Projemizin amacı, haberlerin doğruluk analizini yapabilen bir yapay zeka destekli sistem geliştirmek ve kullanıcıların güvenilir bilgiye ulaşmasını sağlamaktır. Bu sistem, haber metinlerini analiz ederek sahte veya manipülatif içerikleri tespit etmeyi hedeflemektedir. Böylece hem bireylerin hem de toplumsal yapının bilgi kirliliğinden korunmasına katkı sağlanacaktır.

## Problem Tanımı

Yanlış bilgilerin (sahte haberlerin) yayılması, bireysel düzeyde yanıltıcı kararlar alınmasına neden olurken, toplumsal düzeyde de ciddi sonuçlar doğurabilmektedir. Örneğin:

* Sağlık alanında: Yanlış sağlık bilgileri halk sağlığını tehlikeye atabilir.
* Siyasi alanda: Manipülatif haberler, seçim süreçlerini etkileyebilir ve toplumda kutuplaşmaya yol açabilir.
* Ekonomik alanda: Yanlış yatırım bilgileri veya spekülasyonlar ekonomik kayıplara neden olabilir.

Bu problem, sahte haberlerin otomatik olarak tespit edilmesini sağlayacak araçlara olan ihtiyacı ortaya koymaktadır. Geliştirmekte olduğumuz sistem, doğal dil işleme ve makine öğrenmesi tekniklerini kullanarak haberlerin doğruluk durumunu analiz edecek ve bu soruna pratik bir çözüm sunacaktır.

## Projenin Önemi

Bilgi kirliliği, bireylerin günlük hayatında doğrudan etkili olan küresel bir sorundur. Bu proje, sahte haberlerin tespiti konusunda bir adım atarak bireylerin doğru bilgiye erişmesini kolaylaştırmayı ve bilgi güvenliğine katkıda bulunmayı amaçlamaktadır. Özellikle gazeteciler, araştırmacılar ve bilgi doğruluğu üzerine çalışan kurumlar için bu sistem, önemli bir araç olabilir. Toplumsal düzeyde güvenilir bilgiye erişim artacak ve sahte haberlerin yayılma hızının düşürülmesine yardımcı olunacaktır.

Proje, yalnızca teknik bir başarı sağlamayı değil, aynı zamanda toplumda bilgi güvenliğinin ve şeffaflığın artırılmasına yönelik bir farkındalık oluşturmayı hedeflemektedir.

# Kullanılan Teknolojiler ve Araçlar

## Scrum Yönetimi:

* + **Git:** Projede yazılan kodların versiyon kontrolü ve ekip içi iş birliği için kullanılmıştır. Kod değişiklikleri düzenli olarak takip edilmiş, böylece ekip üyeleri arasında paralel çalışma imkânı sağlanmıştır.
  + **Trello:** Görevlerin yönetimi ve Scrum metodolojisiyle sprint takibi için tercih edilmiştir. Projede yapılacaklar (To-Do), devam eden işler (In Progress) ve tamamlanan işler (Done) şeklinde görevler organize edilmiştir. Trello sayesinde ekip içi koordinasyon ve görev dağılımı etkin bir şekilde sağlanmıştır.

## Python Kütüphaneleri:

* + **Pandas:** Haber verilerini işlemek, analiz etmek ve düzenlemek için kullanılmıştır. Örneğin, haber metinlerinin temizlenmesi, tabloya dökülmesi ve özet istatistiklerin çıkarılması gibi işlemler gerçekleştirilmiştir.
  + **NumPy:** Sayısal veriler üzerinde hızlı ve verimli hesaplamalar yapmak için kullanılmıştır. Model eğitiminde gerekli matris işlemleri ve veri manipülasyonu gibi süreçlerde destek sağlamıştır.
  + **Matplotlib:** Veri görselleştirme için tercih edilmiştir. Veri setindeki haber kategorilerinin dağılımını veya doğruluk analizine ilişkin istatistiksel sonuçları grafiklerle sunmak amacıyla kullanılmıştır.
  + **Scikit-learn:** Makine öğrenmesi modellerinin oluşturulması ve değerlendirilmesi için kullanılmıştır. Modeller bu kütüphane yardımıyla eğitilmiş ve test edilmiştir. Ayrıca model doğruluğu, F1 skoru ve hataları gibi performans metrikleri hesaplanmıştır.
  + **Natural Language Toolkit (NLTK):** Haber metinlerini işlemek ve doğal dil işleme görevlerini gerçekleştirmek için kullanılmıştır. Metinlerdeki durak kelimelerin (stopwords) temizlenmesi, kök bulma (stemming) ve sözcük frekansı analizleri gibi işlemler gerçekleştirilmiştir.

## IDE ve Editörler

* + **Google Colab:** Projenin makine öğrenmesi modellerinin geliştirilmesi, veri analizi ve görselleştirme işlemleri için kullanılmıştır. Colab, bulut tabanlı bir platform olduğundan ekip üyelerinin kodları paylaşmasını ve ortak çalışmasını kolaylaştırmıştır. Ayrıca, GPU desteği sayesinde büyük veri setleriyle çalışırken model eğitimi hızlandırılmıştır.
  + **Visual Studio Code (VSCode):** Projenin kodlama aşamalarında kullanılan bir diğer geliştirme ortamıdır. Özellikle Python kodlarının düzenlenmesi ve lokal makinede test edilmesi için tercih edilmiştir. VSCode'un sunduğu eklentiler, hata ayıklama ve kod tamamlamada büyük kolaylık sağlamıştır.

# Yöntem ve Çalışma Süreci

## Scrum Süreci

* + **Sprint Planlamaları:** Projenin başlangıcında, yapılacak işler detaylı olarak analiz edilerek küçük ve yönetilebilir görevler haline getirildi. Her sprint yaklaşık 1 hafta sürdü. Sprintlerin sonunda, tamamlanan görevler gözden geçirildi ve bir sonraki sprint için yeni görevler planlandı.
  + **Trello Kullanımı:** Görevlerin organizasyonu ve takibi için Trello kullanıldı. Görevler, *To-Do (Yapılacaklar)*, *In Progress (Devam Edenler)* ve *Done (Tamamlananlar)* şeklinde listelere ayrıldı. Ekip üyeleri, görevlerine atandı ve ilerleme durumlarını düzenli olarak güncelledi. Trello, ekip içi iletişimi ve iş birliğini artırarak zamanında teslimatı kolaylaştırdı.

## Veri Toplama

* + **Kaynaklar:** Haber verileri, kamuya açık veri setlerinden (ör. Kaggle, Google Dataset Search) ve çeşitli haber sitelerinden çekildi. Veri setleri, doğruluğu kontrol edilmiş haberler ve sahte haberleri içerecek şekilde seçildi.
  + **Veri Temizleme ve Düzenleme:** 
    - Veri setindeki eksik ve tekrarlanan veriler temizlendi.
    - Haber metinlerinden durak kelimeler (stopwords) kaldırıldı.
    - Doğal dil işleme teknikleriyle metinlerde kök bulma (stemming) ve lemmatization işlemleri uygulandı.
    - Veriler, modele uygun hale getirilmesi için etiketlendi ve eğitim/test ayrımı yapıldı.

## Veri Analizi ve Modelleme

* + **Veri Analizi:** Pandas ve Matplotlib kullanılarak haber verilerinin yapısı incelendi. Örneğin:
    - Haber kategorilerinin dağılımı.
    - Sahte ve doğru haberlerin uzunluklarına ilişkin istatistiksel özetler.
    - Grafikler kullanılarak veri görselleştirmesi yapıldı.
  + **Modelleme:**  
    Scikit-learn kütüphanesi ile farklı makine öğrenmesi modelleri eğitildi ve değerlendirildi: REVİZE
  + **Performans Değerlendirme:** REVİZE  
    Modeller, doğruluk oranı, F1 skoru, precision (kesinlik) ve recall (duyarlılık) gibi metriklerle değerlendirildi. Şu ana kadar en iyi sonuçları SVM modelinden aldık:

# İlerleme ve Hedefler

## Şu Ana Kadarki İlerleme

* + **Veri Toplama ve Temizleme:**
    - Çeşitli kaynaklardan haber metinleri toplandı.
    - Veriler, eksik ve tekrarlayan kayıtlar açısından temizlendi.
    - Doğal dil işleme adımlarıyla (stopwords kaldırma, lemmatization) haber metinleri analize uygun hale getirildi.
  + **Temel Veri Analizi:**
    - Pandas ve Matplotlib kullanılarak haber verilerinin istatistiksel analizi yapıldı.
    - Sahte ve gerçek haberlerin metin uzunluğu, sözcük dağılımı ve diğer özellikleri incelendi ve görselleştirildi.
  + **Makine Öğrenmesi Modelleri:**
    - Naive Bayes, Lojistik Regresyon ve SVM gibi modellerle sınıflandırma denemeleri yapıldı.
    - Modellerin doğruluk ve F1 skoru gibi performans metrikleri hesaplandı. İlk sonuçlara göre SVM modeli %87 doğruluk oranı ile en iyi performansı gösterdi.

## Gelecek Hedefler

* + **Model Performansını Artırma:**
    - Daha gelişmiş modeller (ör. Derin öğrenme tabanlı yaklaşımlar) ve hiperparametre optimizasyonları ile sınıflandırma performansını artırmak.
  + **Daha Büyük ve Çeşitli Veri Setleri:**
    - Sahte ve gerçek haberlerin farklı dillerdeki versiyonlarını içeren daha kapsamlı veri setleri toplayarak modelin genelleştirme yeteneğini geliştirmek.
  + **Kullanıcı Arayüzü Geliştirme:**
    - Model sonuçlarını görselleştiren ve kullanıcıların haber metinlerini kontrol edebileceği bir web tabanlı arayüz tasarlamak.
  + **Sonuçların Değerlendirilmesi:**
    - Modelin çıktılarını analiz ederek gerçek dünyada kullanım senaryoları için optimizasyon yapmak.

# Sonuç

Bu proje, haber doğruluk analizi ve tespiti konusunda önemli bir adım atmamıza olanak tanıdı. Şu ana kadar yapılan çalışmalar, hem teknik hem de yönetimsel anlamda bize değerli deneyimler kazandırdı. Özellikle veri toplama, temizleme ve modelleme süreçlerinde karşılaşılan zorluklar, problem çözme becerilerimizi geliştirdi. Makine öğrenmesi modellerinin uygulanması, çeşitli algoritmaların etkinliğini değerlendirme ve sonuçları optimize etme konusunda önemli bir öğrenme süreci oldu.

Projede öğrendiklerimiz ve kazandığımız deneyimler:

## Veri Temizleme ve Düzenleme:

Veri seti üzerinde yaptığımız temizlik ve düzenleme işlemleri, doğru sonuçlar elde etmenin ne kadar kritik olduğunu gösterdi. Eksik veriler ve yanlış etiketler, model doğruluğunu doğrudan etkileyebileceğinden, bu aşamanın titizlikle yapılması gerektiğini öğrendik.

## Makine Öğrenmesi Modelleri:

Farklı makine öğrenmesi algoritmalarını kullanarak, her birinin güçlü ve zayıf yönlerini gözlemledik. Özellikle, basit modellerin (Naive Bayes) hızlı sonuçlar sunduğunu ancak daha karmaşık modellerin (SVM) daha doğru sonuçlar verdiğini gördük. Bu, makine öğrenmesi sürecinde model seçiminin önemini vurguladı.

## Veri Analizi ve Görselleştirme:

Pandas ve Matplotlib gibi araçlarla veri analizi ve görselleştirme yaparken, veriyi daha iyi anlamanın ve görsel sunumların karar verme sürecine nasıl katkı sağladığını deneyimledik.

## Ekip Çalışması ve Yönetim:

Scrum metodolojisi ve Trello kullanarak, ekip içindeki görevlerin daha verimli bir şekilde yönetilmesini sağladık. Bu yöntemler, zaman yönetimi, iş birliği ve proje takibi konusunda bize önemli katkılar sundu.

Sonuç olarak, projemiz, haber doğruluk tespiti için etkili bir yapay zeka çözümü geliştirmeyi hedeflerken, bu süreçte karşılaşılan zorlukları aşmak için çeşitli teknik beceriler kazanmamıza olanak tanıdı. Gelecek adımlarda daha kapsamlı veri setleri ve gelişmiş modellerle bu çözümü daha güçlü ve güvenilir hale getirmeyi planlıyoruz.

# Ekler

## GitHub