

TEKNOLOJİ FAKÜLTESİ

BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

YAPAY ÖĞRENMENİN TEMELLERİ DERSİ

DOĞAL DİL İŞLEME TABANLI SAHTE HABER TESPİTİ

Gökhan Topdanış – B200109039

Emircan Bağdu - B200109041

1. **PROBLEM**

Son yıllarda dijital çağın çok hızlı gelişmesiyle birlikte, insanların bilgi haber alma kaynakları da değişmiştir. Radyo, televizyon ve dergi gibi geleneksel medya platformları yerine sosyal medya kullanımı giderek artmaktadır. Sosyal medyadaki haberlerin kalitesi geleneksel medyadaki haberlerin kalitesine göre daha düşüktür. Geleneksel medyada sadece gazeteciler haber yaparken sosyal medyada herkes haber yapabilmektedir. Bu da sosyal medyada sahte haber yayılımını arttırmaktadır.

Sahte haber; insanları yanıltma veya propaganda amacıyla sahte kullanıcılar tarafından yayılan haberlerdir. Sahte haberler yanlış olmasının yanı sıra ilgi çekici olduklarından sosyal medyada hızla yayılmaktadır.

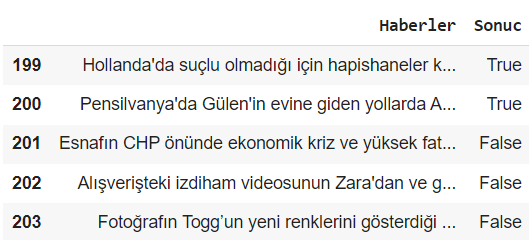
Sosyal medya platformlarında sahte haber tespiti uzman kişiler tarafından yapılmaktadır. Sosyal medya platformlarında sahte haberlerin uzmanlar tarafından tespit edilmesi, yoğun paylaşım trafiği yüzünden mümkün olmamaktadır. Bu yüzden sahte haber kısa sürede birçok işi tarafından bilinçsizce paylaşılıp sahte haberin tespiti anlamında negatif etkiler oluşturmaktadır.

Günümüz medyasının en büyük sorunlarından biri haline gelen dezenformasyon, yayılma hızı nedeniyle gerçek bilgilerin önünü kapatmaktadır. Dezenformasyonla Mücadele Yasası hayatımıza çok sayıda yenilik getirerek yalan haberin yayılmasını engelleme konusunda önemli bir rol oynamaktadır.

Projemizde, insanlar tarafından paylaşılan sahte haber sayısını azaltmak için geliştireceğimiz yazılım ile Dezenformasyonla Mücadele Yasası’na bir katkıda bulunmayı planlanmaktadır. Bu kapsamda projemizde yapay zekâ algoritmaları kullanarak yazılan haberlerin sahte haber kapsamına girip girmediğini sınıflandırarak sahte haberin yayılmasını önlemeyi hedeflenmektedir.

1. **VERİ SETİ**

Haberler Teyit.org ve Github ortamlarından manuel olarak toplanarak veri seti oluşturulmuştur. Veri seti 642 doğru haber, 926 yanlış etiketli olmak üzere 1568 haberden oluşmaktadır.



1. **ALGORİTMA**

Projede haberleri doğru veya yanlış olarak sınıflandırmak için oluşturulan algoritma metin ön işleme, öznitelik çıkarımı ve sınıflandırma algoritması olmak üzere 3 adımdan oluşmaktadır.

* 1. **Metin Ön İşleme**

Metin ön işlemede Python dili üzerinde, NLTK ve ZEMBEREK kütüphaneleri ağırlıklı olmak üzere bu kütüphanelere ek olarak bazı farklı modüller ve Python’da gömülü gelen fonksiyonlar kullanılacaktır. NLTK ile cümlelerdeki noktalama işaretlerini sayısal ifadeleri, durak kelimeleri çıkarmak için ve tüm karakterleri küçük karakterlere çevrilmesi gibi işlemleri yapılacaktır. ZEMBEREK kütüphanesini ise kelimeleri, köklerine ve gövdelerine ayırmak için kullanılacaktır. İki ayrı kütüphane kullanılmasının nedeni ise NLTK kütüphanesi Türkçe gibi sondan eklemeli dillerde kelimenin köklerini ve gövdelerini çoğu kez doğru bir şekilde bulamamasıdır.

* 1. **Öznitelik Çıkarımı**

Önişleme adımlarından sonra eldeki metin verisi makine öğrenmesi algoritmaların işleyeceği hale TF-IDF doküman temsili ile getirilecektir.

* + 1. **Vektörizasyon**

Doğal Dil İşleme çalışmalarında, matematiksel işlemlerin yapılabilmesi için metinsel ifadelerin vektörel ifadelere dönüştürülmesi gerekmektedir.[4]

* + 1. **Terim Frekansı (TF)**

Bir dokümanda geçen terimlerin terim ağırlıklarını hesaplamak için kullanılır. Metindeki bir terimin metinde toplam kullanılma sayısının, metindeki tüm terimlerin toplam sayısına oranıdır. [5]

* + 1. **Ters Doküman Frekansı (IDF)**

Kelimenin kaç farklı dokümanda geçme sayısını bularak bu kelimenin terim olup olmadığını bağlaç vb. (Durak Kelimeler) olduğu anlamaya çalışır. Toplam doküman sayısının, terimi içeren doküman sayısına oranıdır. [5]

* + 1. **TF-IDF**

Doğal dil işleme çalışmalarında, metinsel ifadelerin vektörel ifadelere dönüştürülmeleri gerekmektedir. Bu şekilde matematiksel işlemler yapılabilmektedir. Bu işlem için TF-IDF, doğal dil işlemede sıkça kullanılmaktadır. TF-IDF, tahmin yapmak amacıyla makine öğrenmesi algoritmalarına verilecek olan verilerde, verinin içerisindeki bir kelimenin ne kadar öneme sahip olduğunu hesaplamaktadır. Terim frekansı (TF) ile Ters Doküman Frekansı (IDF) çarpılarak hesaplanmaktadır. Bu işlem dokümanlardaki her kelime için yapılmaktadır.[6]

t: kelime indisi, j: doküman indisi

* 1. **Sınıflandırma**

Sınıflandırma işlemleri yapay zekanın alt dalı olan makine öğrenmesi çatısı altındaki sınıflandırma algoritmaları kullanılarak yapılmıştır. Bu projede sınıflandırma, ikili sınıflandırma (Binary Classification) türüdür. Veri gerekli ön işlem ve öznitelik çıkarımlarından sonra doğru veya yanlış olarak sınıflandırma işlemine tabi tutulmaktadır. Sınıflandırma algoritması için Sklearn kütüphanesi içerisindeki sınıflandırma algoritmaları knn ve naive bayes kullanılmıştır.

* + 1. **K-NN**

Sınıflandırma problemleri için kullanılan denetimli bir makine öğrenmesi algoritmasıdır.  
K-NN sınıflandırmasında, çıktı sınıf üyeliğidir. Bir nesne, komşularının çoğunluk oyuyla sınıflandırılır; nesne, en yakın komşuları arasında en yaygın olan sınıfa verilir.

* + 1. **Naive Bayes**

Bayes‘in teoremine dayanır ve her değeri diğer değerlerden bağımsız olarak sınıflandırır. Olasılık kullanarak ve belirli bir dizi özelliğe dayanarak bir sınıfı / kategoriyi tahmin etmemizi sağlar.

* + 1. **Sınıflandırma Algoritmasının Değerlendirilmesi**

Karmaşıklık matrisi veri setindeki var olan durum ile sınıflandırma modelimizin doğru ve yanlış tahminlerinin sayısını tablo olarak göstermektedir. Accuracy (Doğruluk): Sistemde doğru olarak yapılan tahminlerin tüm tahminlere oranıdır.

tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

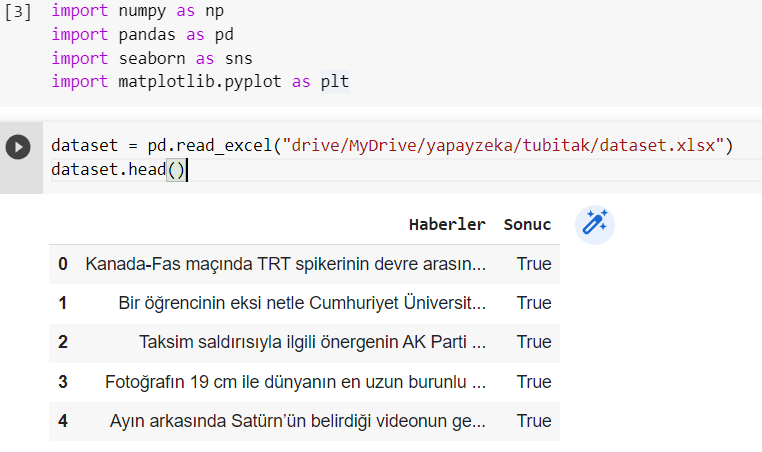
1. **UYGULAMA**

Projenin çözümü için tasarlanan algoritma, Python programlama dili üzerinde kodlanmıştır.

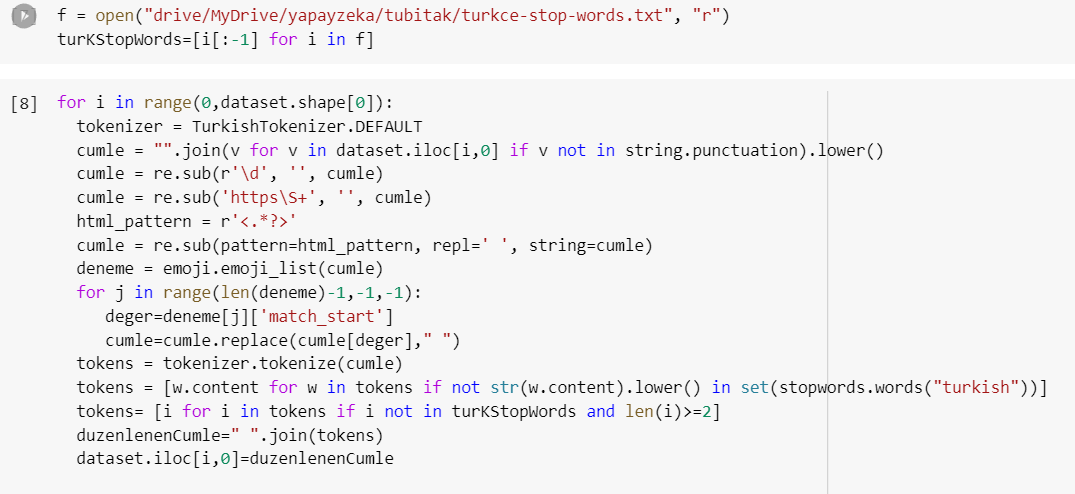
* 1. **Metin Ön İşleme Aşaması**
* **Gerekli Kütüphanelerin Dahil Edilmesi**



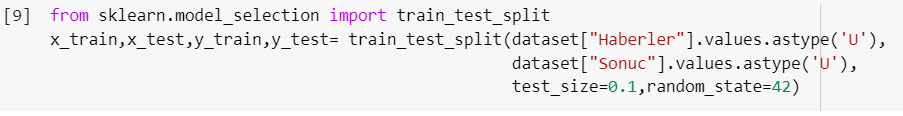
* **Veri Setinin Dahil Edilmesi**

****

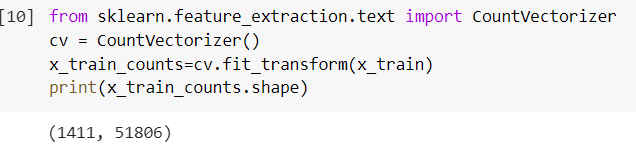
* **Metin Düzenlenme İşlemi**

****

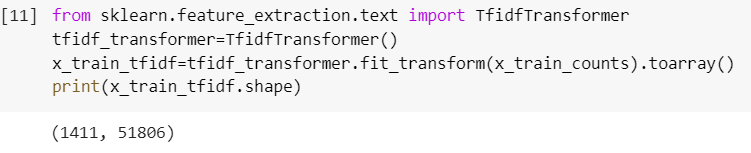
* 1. **Öznitelik Çıkarım Aşaması**
* **Veri Setinin Train-Test Olarak Bölünmesi**

****

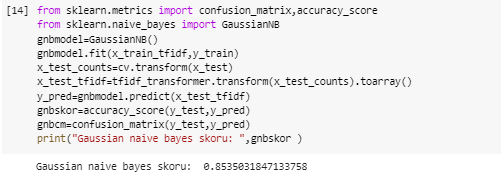
* **Vektörizasyon İşlemi**

****

* **TF-IDF Skoru Hesaplama İşlemi**

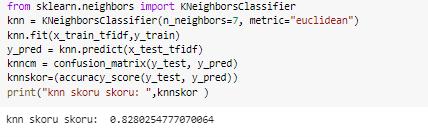
****

* 1. **Sınıflandırma Aşaması**
* **NAİVE BAYES Sınıflandırma Modelinin Test Edilmesi**





* **KNN Sınıflandırma Modelinin Test Edilmesi**





**Kaynakça**

[1] Toğaçar, M., Eşidir, K. A., & Ergen, B. (2022). Yapay Zekâ Tabanlı Doğal Dil İşleme Yaklaşımını Kullanarak İnternet Ortamında Yayınlanmış Sahte Haberlerin Tespiti. Journal of Intelligent Systems: Theory and Applications, 5(1), 1-8.

[2] YILDIRIM, E. (2022). HIZLANDIRILMIŞ MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE TÜRKÇE SAHTE HABER TESPİTİ (Doctoral dissertation).

[3] TAŞKIN, S. G., Küçüksille, E. U., & Topal, K. Twitter üzerinde Türkçe sahte haber tespiti. Balıkesir Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 23(1), 151-172.

[4] <https://yalansavar.org/2020/03/03/podcast-22-uydurma-haberler/>

[5] Onur Dayıbaşı. (Aug 13, 2015). TF-IDF (Term Frequency — Inverse Document Frequency). Medium. <https://medium.com/algorithms-data-structures/tf-idf-term-frequency-inverse-document-frequency-53feb22a17c6>

[6] Şevval Ayşe Yurtekin. (23 OCA 2022). Doğal dil işlemede temel kavramlara hızlı bir bakış. miuul. <https://www.miuul.com/not-defteri/dogal-dil-islemede-temel-kavramlara-hizli-bir-bakis>