T.C.
TRAKYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

TÜRKÇE TWITTER VERİLERİ ÜZERİNDE GÜVENİLİRLİK SKORLAMA İLE SAHTE HABER TESPİTİ

HÜSEİN KANTARCİ

YÜKSEK LİSANS TEZİ

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

Tez Danışmanı: Dr. Öğr. Üyesi ÖZLEM AYDIN

EDİRNE, 2023

HÜSEİN KANTARCİ'nin hazırladığı "**TÜRKÇE TWITTER VERİLERİ ÜZERİNDE GÜVENİLİRLİK SKORLAMA İLE SAHTE HABER TESPİTİ**" başlıklı bu tez, tarafımızca okunmuş, kapsam ve niteliği açısından **Bilgisayar Mühendisliği** Anabilim Dalında bir **Yüksek lisans tezi** olarak kabul edilmiştir.

Jüri Üyeleri (Ünvan, Ad, Soyad):	İmza
Dr. Öğr. Üyesi Özlem AYDIN	……
Prof. Dr. Erdinç UZUN	……
Dr. Öğr. Üyesi Emir ÖZTÜRK	……

Tez Savunma Tarihi: 17/08/2023

Bu tezin Yüksek Lisans tezi olarak gerekli şartları sağladığını onaylarım.

İmza

Dr. Öğr. Üyesi Özlem AYDIN
Tez Danışmanı ……

Trakya Üniversitesi Fen Bilimleri Enstitüsü onayı

……

Prof. Dr. Hüseyin Rıza Ferhat KARABULUT
Fen Bilimleri Enstitüsü Müdürü

T.Ü.FEN BİLİMLERİ ENSTİTÜSÜ

BİLGİSAYAR MÜHENDİSLİĞİ YÜKSEK LİSANS PROGRAMI

DOĞRULUK BEYANI

Trakya Üniversitesi Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmasında, tüm verilerin bilimsel ve akademik kurallar çerçevesinde elde edildiğini, kullanılan verilerde tahrifat yapılmadığını, tezin akademik ve etik kurallara uygun olarak yazıldığını, kullanılan tüm literatür bilgilerinin bilimsel normlara uygun bir şekilde kaynak gösterilerek ilgili tezde yer aldığını ve bu tezin tamamı ya da herhangi bir bölümünün daha önceden Trakya Üniversitesi ya da farklı bir üniversitede tez çalışması olarak sunulmadığını beyan ederim.

.... / .... / ....

Hüsein Kantarci

İmza

Yüksek Lisans Tezi
Türkçe Twitter Verileri Üzerinde Güvenilirlik Skorlama ile Sahte Haber Tespiti
T.Ü. Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

ÖZET

Günümüzde, internet üzerindeki veri yığını özellikle sosyal medya aracılığıyla hızla büyümektedir. Bu veri yığını, kullanıcılar tarafından kasıtlı veya kasıtsız bir şekilde paylaşılan ve doğruluğu ile güvenilirliği sorgulanması gereken bilgilerle doludur. Özellikle yanlış bilginin tekrarlandığında insanların inançlarını çarpıtabileceği gerçeği göz önünde bulundurulduğunda, sosyal mecralarda karşılaşılan bu tekrarlı ve tehlikeli haberlerin önüne geçmeye yönelik yapılan çalışmaların büyük önemi ortaya çıkmaktadır.

Bu çalışmada, Twitter veri setindeki tweetlerin güvenilirlik skorlaması yapmak amacıyla denetimsiz makine öğrenmesi kullanılarak sahte haberlerin tespiti yapılmıştır. Türkçe Twitter haber veri setinin bir kısmı, önceden belirlenmiş kullanıcılar tarafından atılmış tweetler toplanarak manuel bir şekilde etiketlenmiştir. Daha sonra K-Means modeli eğitilerek, manuel olarak etiketlenmiş verilerle karşılaştırılmıştır. Elde edilen sonuçlar değerlendirildiğinde, modelin F1 skoru olarak 0.33 başarı elde etmiştir. Bu sonuçlar, modelin sosyal medya platformlarında yanlış bilgi tespitinde potansiyel bir etkinlik gösterdiğini ve kullanıcıları yanıltıcı içeriklerden korumada önemli bir adım olduğunu ortaya koymaktadır.

Yıl : 2023

Sayfa Sayısı : 43

Anahtar Kelimeler : Sahte Haber Tespiti, Denetimsiz Öğrenme, K-Means Modeli

Master’s Thesis
Fake News Detection with Credibility Scoring on Turkish Twitter Data
Trakya University Institute of Natural and Applied Sciences
Department of Computer Engineering

ABSTRACT

In today's world, the data on the internet is rapidly growing, especially through social media. This data is filled with information shared by users, intentionally or unintentionally, that needs to be questioned for its accuracy and reliability. Particularly, considering the fact that misinformation can distort people's beliefs when repeated, the importance of efforts to combat repetitive and dangerous news encountered on social platforms becomes evident.

In this study, unsupervised machine learning is employed to perform credibility scoring on tweets from the Twitter dataset. A portion of the Turkish Twitter news dataset is manually labeled by collecting tweets from pre-determined users. Subsequently, the K-Means model is trained and compared with the manually labeled data. The results show that the model achieved an F1 score of 0.33, indicating its potential effectiveness in detecting false information on social media platforms and representing a significant step towards safeguarding users from misleading content.

Year : 2023

Number of Pages : 43

Keywords : Fake News Detection, Unsupervised Learning, K-Means Model

TEŞEKKÜR

Bu çalışma sürecinde bana gösterdiği sürekli destek ve motivasyon için minnettar olduğum danışman hocam Dr. Özlem Aydın'a teşekkürlerimi sunarım. Her aşamada sizin rehberliğiniz ve yönlendirmeleriniz sayesinde bu çalışmayı tamamlayabildim. Sizlerin bilgi birikimi ve ilgisi, beni daha da ileriye taşıdı ve bu deneyimi unutulmaz kıldı. Emeğiniz, katkınız ve özveriniz için sonsuz teşekkürlerimi sunuyorum.

İÇİNDEKİLER

SİMGELER VE KISALTMALAR DİZİNİ

BERT Bidirectional Encoder Representations from Transformers (Dönüştürücülerden İki Yönlü Kodlayıcı Temsilleri)

C Credible (Güvenilir)

CNN Convolutional Neural Network (Evrişimli Sinir Ağı)

DDİ Doğal Dil İşleme

GPT Generative Pre-trained Transformer (Üretici Ön-Eğitimli Dönüştürücü)

GRU Gated Recurrent Units (Kapılı Yinelemeli Üniteler)

HC High Credibility (Çok güvenilir)

LC Low Credibility (Az güvenilir)

NC Not Credible (Güvenilir Değil)

RFC Random Forest Classifier (Rastgele Orman Sınıflayıcı)

RNN Recurrent Neural Network (Tekrarlayan Sinir Ağı)

SVM Support Vector Machine (Destek Vektör Makinesi)

t-SNE t-Distributed Stochastic Neighbor Embedding (t-Dağıtık Stokastik-Rasgele Komşu Gömme)

UFNDA Unsupervised Fake News Detection Method Based on Auto Encoder (Otokodlayıcıya Dayalı Denetimsiz Sahte Haber Tespit Yöntemi)

ÇİZELGELER DİZİNİ

ŞEKİLLER DİZİNİ

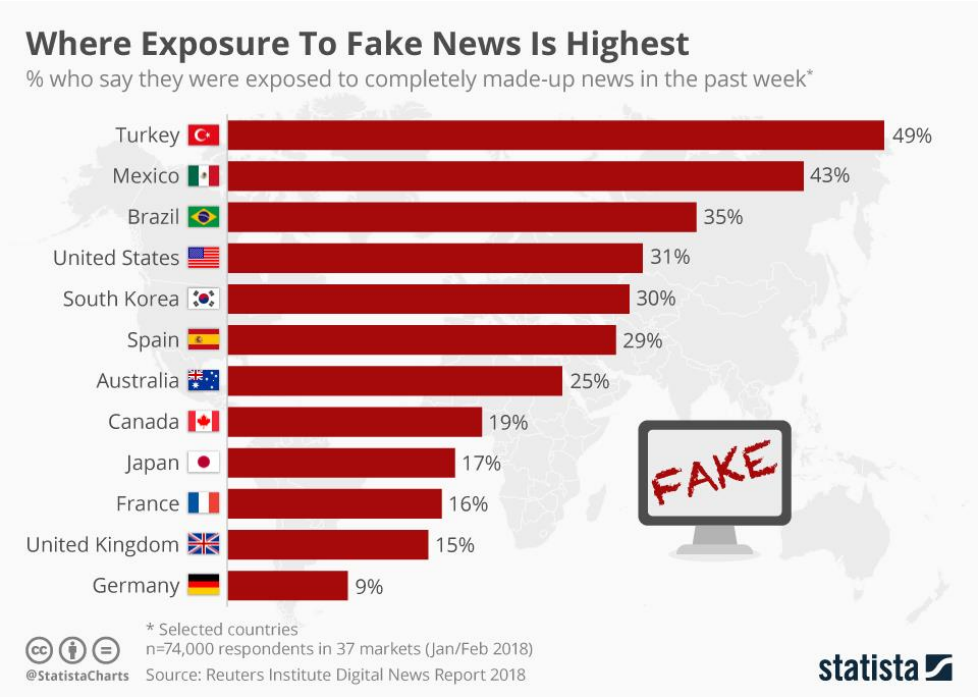
BÖLÜM 1

GİRİŞ

Internet ve sosyal medya platformları, haber ve bilgi paylaşımında önemli bir rol oynamaktadır. Ancak, bu platformların hızla büyümesi ve yaygınlaşmasıyla birlikte sahte haberlerin yayılma hızı ve etkisi de artmıştır. Sahte haberler, "genellikle politik görüşler üzerinde etki yaratmak amacıyla internet ya da diğer medyayı kullanarak yayılan haber gibi görünen düzmece, sahte hikayeler" olarak tanımlanır. (Akyüz, Kazaz, & Gülnar, 2021) Bu tür haberler, toplumun görüşleri, inançları ve kararları üzerinde olumsuz etkilere yol açabilir ve hatta toplumsal istikrarı tehdit edebilir. Bunun dışında, bu haberlerin toplumdaki yanlış bilgi yayılımını artırması nedeniyle insanların gerçekleri ayırt etme yeteneği zayıflayabilir. Sahte haberler, toplumsal kutuplaşmaya neden olabilir ve toplumu ayrıştırıcı bir etkiye sahip olabilir. Ayrıca, siyasi, ekonomik ve toplumsal alanda ciddi sonuçlar doğurabilecek sahte haberlerin manipülatif ve yanıltıcı doğası, demokratik süreçlere ve kamusal güvene zarar verebilir. Sahte haberlerin tespiti için yapılan çalışmalarda, çoğu zaman metinlerin yazım stillerine odaklanarak gerçek ve sahte haberleri sınıflandırmaya yönelik çabalar görülmüştür (Kai, Amy, Suhang, Jiliang, & Huan, 2017). Ancak, bu tür yaklaşımların, incelenen haberlerde tam bir netlik sağlayamadığı gözlemlenmiştir.

Hızlı paylaşım ve viral yayılma mekanizmaları, sahte haberlerin gerçek haberlerle karıştırılmasına ve hızla yayılmasına yol açar. Ayrıca, sahte haberlerin dikkat çekici başlıklar ve şok edici içeriklerle donatılması, insanların haberleri paylaşma ve etkileşime geçme isteğini artırır. Sahte haberlerin yayılmasına karşı etkili bir denetim olmaması, bu tür haberlerin yaygınlaşmasında önemli bir sorundur. Geleneksel haber kaynakları genellikle

gerçeklik kontrolünden geçerken, sosyal medya platformlarındaki haberlerin denetimi daha zayıftır. Bu nedenle, sahte haberlerin sosyal medya platformlarında yayılma olasılığı daha yüksektir. Türkçe içerikli sosyal medya platformları ve haber sitelerinde sahte haberlerin yaygınlığı endişe vericidir. Toplumun farklı kesimlerini etkileyen siyasi, toplumsal ve ekonomik konularda sahte haberler sık sık karşılaşılan bir durumdur.



Şekil 1.1 Reuters Insitute 2018 yılı dijital haber raporu

Şekil 1.1'de Reuters Institute tarafından yayınlanan dijital haber raporu, dünya genelinde haber güveni ve yanlış bilgi tespiti üzerine önemli bulgular sunmaktadır (McCarthy, 2018). Araştırma, haber yazarlığının gerçeği kurgudan ayırma ve sahte haberleri tanıma süreçlerinde kritik bir rol oynadığını ortaya koymaktadır. 2018'de yapılan ankete katılanların kendi beyanlarına göre, sahte haberle karşılaşma oranı Türkiye'de oldukça yüksek seviyededir ve %49 olarak belirtilmiştir. Benzer şekilde, Amerika Birleşik Devletleri'nde bu oran %31 düzeyindedir ve bu durum, 2016 seçimlerinde sahte haberlerin yaygınlaşmasından kaynaklanmaktadır. Şaşırtıcı bir şekilde, Birleşik Krallık'ta sahte haberle

karşılaşma oranı %15 olarak düşük bulunmuştur, bu da ülkenin Brexit oylaması sırasındaki yanlış bilgi tartışmalarına rağmen ilgi çekicidir (McCarthy, 2018).

Sahte haberlerle mücadelede geleneksel yöntemlerin yetersiz kalması, yapay zeka ve makine öğrenmesi gibi yeni teknolojilere yönelik ilgiyi artırmıştır. Denetimsiz makine öğrenmesi, sahte haber tespitinde umut verici bir yol olarak karşımıza çıkar. Bu yaklaşım, büyük veri kütlelerini analiz ederek, sahte ve gerçek haberleri ayırt etmeye çalışır ve güvenilirlik skorlaması yaparak sahte haberlerin yayılmasının önüne geçebilir. Bu çalışma, Twitter veri setindeki tweetlerin güvenilirlik skorlaması ile sahte haberlerin tespitine yönelik önemli bir adım olarak görülebilir. Denetimsiz makine öğrenmesi yöntemi sayesinde sosyal medya platformlarında yanlış bilgi tespitine katkı sağlayabilir ve toplumu yanıltıcı içeriklerden korumada önemli bir rol üstlenebilir.

BÖLÜM 2

KAYNAK ARAŞTIRMASI

Özellikle son zamanlarda, tweetlerin sınıflandırılması, etiketlenmesi, tweetler üzerinde sahte haber tespiti yapılması, Twitter'ı bilgi analizinde kullanmak isteyen birçok uygulama için önemli konular olmuştur. Trendleri belirlemek, tweetleri filtrelemek, sınıflandırmak, sahte haberleri tanımlamak veya bir tweetin güvenilirliğini ölçmek gibi birçok amacın, tweetlerin içeriklerine bağlı olarak gerçekleştirilmekte olduğu görülmüştür.

Bu konuda yapılan bir çalışmada, tweetlerin konularını belirlemek için etiketlerin (hashtaglerin) önemli bir rol oynadığı vurgulanmıştır. Araştırmacılar, metin, kullanıcı ve sosyal güvenilirlik ölçümlerine dayalı bir güvenilirlik modelini geliştirmiş ve bu modeli PHEME veri kümesi kullanılarak sahte haberleri tespit etmek amacıyla değerlendirmişlerdir. Bu çalışmada elde edilen bulgular, konu belirleme algoritmalarının entegrasyonuyla genişletilen güvenilirlik modelinin etkinliğini göstermiştir. PHEME veri kümesi üzerinde yapılan deneyler sonucunda, güncellenen modelin F1 puanında %3.04'lük bir artış sağladığı görülmüştür. Özellikle, PHEME veri kümesindeki hashtag içeren tweetler üzerinde yapılan değerlendirmelerde ise F1 puanında %9.60'lık bir artış gözlemlenmiştir. Bu sonuçlar, etiketlerin konu belirleme sürecindeki önemine işaret ederek, güvenilirlik analizindeki başarıyı artırmak için bu tür unsurların dikkate alınmasının önemini vurgulamaktadır (Aguilera, Dongo, Mendoza, Lupa, & Cardinale, 2022).

Yapılan bir diğer çalışmada ise sosyal medyada yayılan sahte haberlerin tespit edilmesi ve önlenmesi amacıyla Türkçe dilinde Twitter üzerindeki içerikler analiz edilmiştir.

Doğal Dil İşleme (DDİ) yöntemleri kullanılarak sahte haberlerin tespiti için çeşitli denetimli ve denetimsiz öğrenme algoritmaları test edilmiştir. En yüksek başarı skoru, destek vektör makineleri algoritması ile elde edilerek %90 oranında sahte haberlerin tespit edilebildiği belirlenmiştir. Ayrıca, sosyal ağ analizi yöntemleriyle sahte ve gerçek haberleri paylaşan kullanıcıların takipçi-takip edilen ilişkileri incelenmiştir. Sonuçlar, sahte haberlerin sosyal ağlardaki yayılımının, gerçek haberlerden farklı karakteristiklere sahip olduğunu göstermiştir. Bu çalışma, sahte haberlerin önlenmesi ve güvenilir bilgi paylaşımı için veri tabanlı yöntemlerin etkili bir şekilde kullanılabileceğini ve sosyal ağ analizinin sahte haberlerin yayılma mekanizmalarını anlamada yardımcı olduğunu ortaya koymuştur (Taşkın, Küçüksille, & Topal, 2021).

Sosyal medyanın bilgiye erişimi nasıl değiştirdiğini ve sahte haberlerin yayılmasının yaygınlaştığına odaklanan bir diğer çalışmada, var olan literatürdeki sahte haber tespiti çalışmalarının çoğunun kullanıcı veya içerik odaklı olduğu görülmüştür. Ancak bu çalışmaya göre, yapılan son araştırmalar Twitter'da gerçek ve sahte haberlerin farklı şekillerde yayıldığını ortaya koymaktadır. Bu çalışma, sahte haberlerin tespiti için yayılma özelliklerinin kullanılma potansiyelini vurgulayarak literatürdeki açığı doldurmaya katkı sağlamıştır. Çalışmanın sonuçları, gerçek haberlerin sahte haberlere oranla daha büyük boyutta olduğunu, daha fazla takipçiye sahip ama diğer kullanıcılar tarafından daha az takip edilen hesaplar tarafından aktif bir şekilde uzun süre boyunca Twitter'da yayıldığını göstermektedir. Araştırmacılar yayılma özelliklerine dayalı Rastgele Orman Sınıflayıcı (Random Forest Classifier - RFC) modeli ile %87 başarı elde etmiştir. Ayrıca, Geometrik Derin Öğrenme yöntemi kullanarak yapılan grafik sinir ağının, %73.3 başarıya ulaştığı gözlemlenmiştir. Bu çalışma, sahte haberlerin yayılma mekanizmalarını anlama ve yayılma özelliklerini kullanarak sahte haberlerin tespitine odaklanarak literatüre önemli katkılar sağlamaktadır (Meyers, Weiss, & Spanakis, 2020).

Sosyal medya platformlarının gelişimiyle birlikte sahte haberlerin hızla yayılması, insanların günlük yaşantısına ciddi olumsuz etkiler taşımakta ve hatta toplumsal kargaşaya yol açabilmektedir. Bu nedenle, farklı bir çalışmada sahte haberlerle mücadelede etkili bir çözüm olarak "Oto kodlayıcıya (Auto Encoder) Dayalı Denetimsiz Sahte Haber Tespit

Yöntemi (Unsupervised Fake News Detection Method Based on Auto Encoder, UFNDA)" adlı bir çalışma önerilmiştir. Bu çalışmada, özellikle sosyal medyadaki farklı haber türleri dikkate alınarak, haberlerin metin içeriği, görselleri, yayılma süreçleri ve kullanıcı bilgileri bir araya getirilerek sahte haberlerin tespit performansı artırılmıştır. Ayrıca, özellikler arasındaki gizli bilgiyi ve içsel ilişkileri elde etmek amacıyla oto kodlayıcıya Çift Yönlü Kapılı Yinelemeli Üniteler (Bidirectional Gated Recurrent Units, Bi-GRU) katmanı ve Öz-Dikkat (Self-Attention) katmanı eklenerek sahte haber tespiti daha kesin hale getirilmiştir. Elde edilen sonuçlar, gerçek dünya veri kümeleri üzerinde diğer dört yöntemle karşılaştırılarak değerlendirilmiştir ve UFNDA'nın diğer yöntemlere göre daha olumlu sonuçlar sunduğu belirlenmiştir (Li, Guo, Wang, & Zheng, 2021).

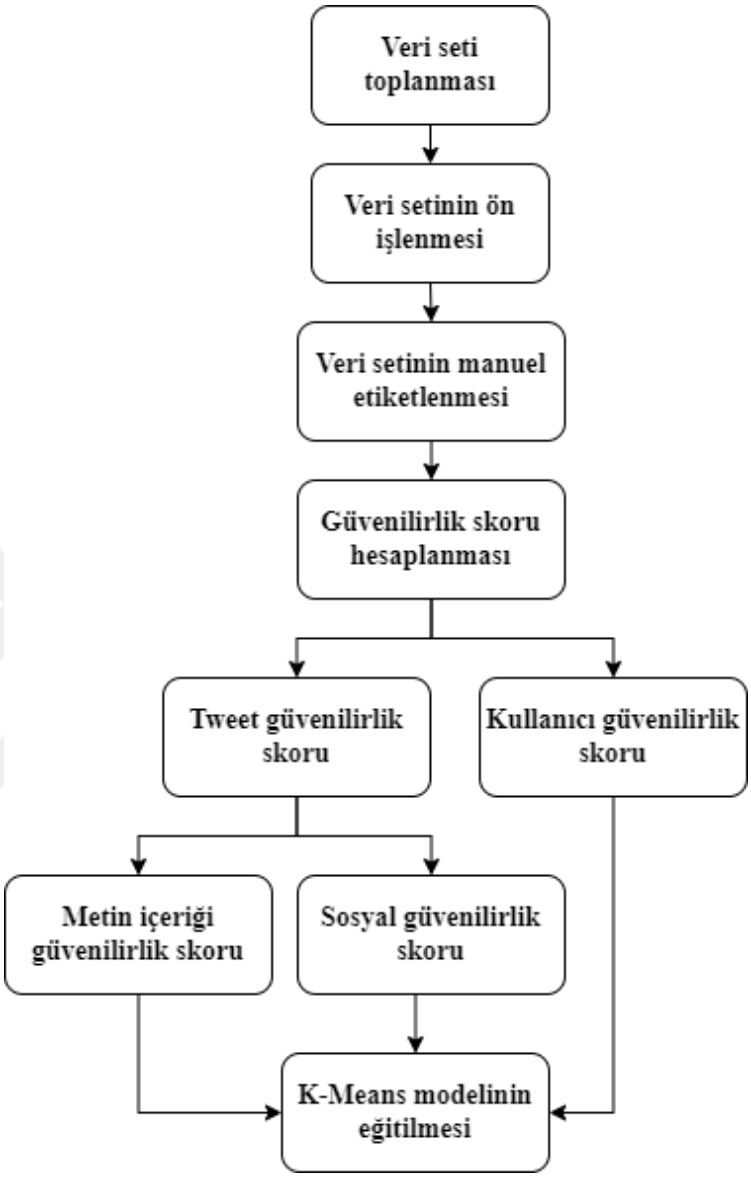
Sahte haberlerin tespiti için mevcut veri kaynaklarının sınırlı olduğu ve bu durumun tespit sürecini zorlaştırdığı belirtilen bir diğer çalışmada, sahte haberlerin tespiti için denetimsiz öğrenme yöntemlerinden Tek Sınıf SVM (Support Vector Machine - Destek Vektör Makinesi) ve derin öğrenme yöntemlerinden Hibrit CNN-RNN (Convolutional Neural Network - Evrişimli Sinir Ağı)-(Recurrent Neural Network - Tekrarlayan Sinir Ağı) algoritmalarının kullanıldığı ifade edilmiştir. Deneyler, bu yöntemlerin NEWS veri kümesi üzerinde uygulandığını göstermiş ve Tek Sınıf SVM için %58, Hibrit CNN-RNN için ise %96.4 doğruluk elde edildiği belirtilmiştir. Hibrit CNN-RNN yönteminin mevcut makine öğrenimi algoritmalarına göre daha yüksek bir uygulama performansı sergilediği ve sahte haberlerin tespiti konusunda daha etkili sonuçlar elde etmeye yönelik potansiyeli olduğu ortaya konmuştur (John, 2022).

BÖLÜM 3

MATERYAL VE YÖNTEM

Bu bölümde, araştırmada kullanılan veri seti ve yöntemler detaylı bir şekilde sunulmuştur.

Şekil 3.1'de gösterilen akış diyagramı, sahte haber tespiti uygulamasının temel adımlarını içerir. İlk adım, Twitter API'si aracılığıyla sahte haberleri içeren veri setini toplamaktır. Toplanan veri seti daha sonra ön işleme aşamasından geçer, burada alan isimleri düzenlenir, boş satırlar veri setinden kaldırılır ve veri seti normalize edilir. Ardından, veri seti manuel olarak etiketlenir. Bu etiketleme, modelin eğitiminde kullanılacak doğruluk metriklerini elde etmek için önemlidir. Sonraki adımda, haber içeriklerinin güvenilirlik düzeyini değerlendirmek için güvenilirlik skorlaması yapılır. Bu skorlamalar, haberlerin sahte veya gerçek olduğu olasılığını tahmin etmeye yardımcı olur. Son olarak, K-Means modeli eğitilir ve kümeleme uygulanarak haber içerikleri etiketlenmiş olur.



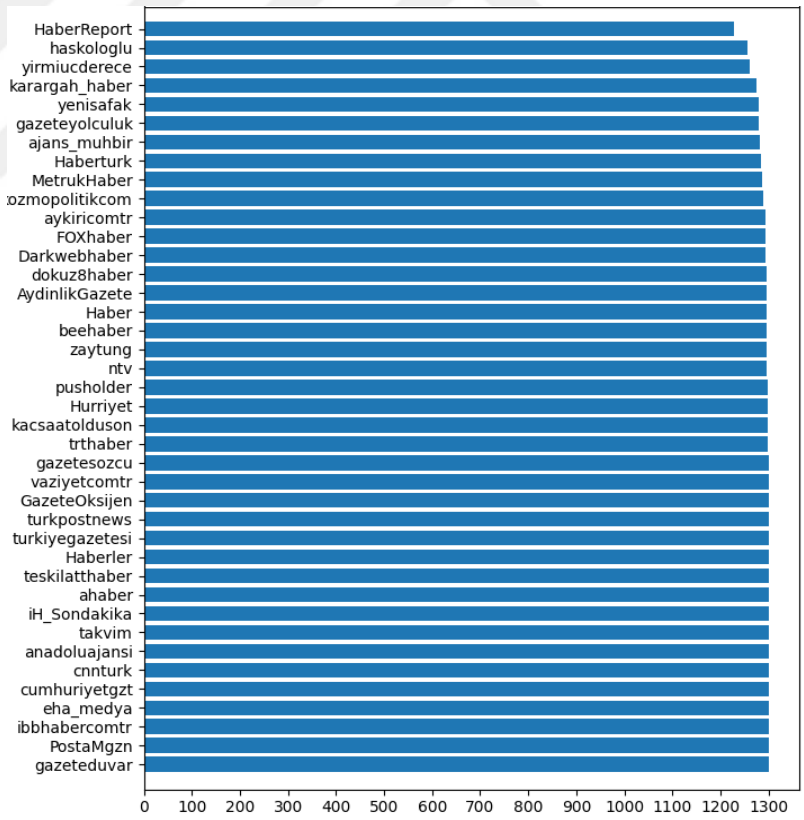
Şekil 3.1 Uygulamanın akış diyagramı

3.1 Veri Seti

Bu çalışma kapsamında, Türkçe haber yayını yapan belirli kullanıcıların Twitter platformunda en son paylaştıkları tweetlerden oluşan bir veri seti toplanmıştır. Twitter'ın resmi API'si kullanılarak gerçekleştirilen bu toplama işlemiyle, önceden belirlenmiş olan kullanıcı listesi üzerinden veriler elde edilmiştir. Veri setinin bir bölümü, sahte haber tespiti modelinin performansını değerlendirmek amacıyla manuel etiketlenmiştir, çünkü veri setinde

önceden etiketlenmiş veriler bulunmamaktadır. Bu manuel etiketleme işlemi, toplam veri setinin %10'luk bir kısmını kapsamaktadır.

Veri toplama süreci, bir sahte haber tespiti uygulaması geliştirmek amacıyla Twitter API'sine erişim gerekliliği doğurmuştur. Bu erişim izni, Twitter API'sine akademik bir uygulama geliştirmek amacıyla yapılan başvurunun kabul edilmesiyle elde edilmiştir. Twitter tarafından sağlanan API erişim tokeni, Python dilindeki Tweepy kütüphanesi aracılığıyla kullanılarak Twitter API fonksiyonlarına erişim sağlanmıştır. Veri toplama süreci, önceden belirlenen 38 kullanıcının son 1300 tweetini içeren bir veri setinin oluşturulması amacıyla gerçekleştirilmiştir. Bu kullanıcı listesi, çalışmanın hedeflerine uygun bir şekilde haber paylaşımı yapan hesaplar arasından seçilmiştir. Şekil 3.2'de, tweetlerin toplandığı kullanıcılar ve bu tweetlerin kullanıcılar arasındaki dağılımları verilmiştir.



Şekil 3.2 Veri seti kullanıcı listesi ve toplanan tweet sayısı

Veri toplama süreci sonucunda toplam 51,674 satırdan oluşan bir tweet veri seti elde edilmiştir. Bu veri setinde hem tweet hakkında hem tweet sahibi kullanıcı hakkında birden fazla alan bulunmaktadır. Bu veri setinin %10'u yani 5,169 tweet manuel olarak etiketlenmiştir. Geri kalan %90'lık kısım ise, K-Means modelinin eğitiminde kullanılmak üzere ayrılmıştır. Manuel etiketlenen %10'luk kısım performans değerlendirmesi için kullanılmıştır.

3.2 Ön İşleme

Veri setinin ön işleme adımlarında, Python Pandas kütüphanesi kullanılmıştır. Aşağıda, toplanan veri setinin ön işleme adımları detaylı bir şekilde açıklanmıştır:

1. Boş satırların ve değerlerin kaldırılması:

   Veri setindeki boş satırlar tespit edilmiş ve bu boş verilere sahip satırlar veri setinden kaldırılmıştır. Bu adım, veri setinin bütünlüğünü ve doğruluğunu sağlamak için yapılmıştır.

2. Alan isimlerinin okunabilir hale dönüştürülmesi:

   Twitter API'si tarafından döndürülen alan isimleri, daha anlaşılır, okunabilir ve kolay kullanılabilir hale getirilmiştir. Çizelge 3.1’de alan isimlerinin önceki ve sonraki halleri gösterilmiştir:

Çizelge 3.1 Alan isimlerinin ön işleme karşılaştırması

Ön işleme öncesi	Ön işleme sonrası
entities.hashtags	hashtags
entities.mentions	mentions
entities.urls	urls
public_metrics.retweet_count	retweet_count
public_metrics.reply_count	reply_count
public_metrics.like_count	like_count
public_metrics.quote_count	quote_count
public_metrics.impression_count	impression_count
attachments.media_keys	media_keys
id	tweet_id
entities.annotations	annotations

3. Hashtag ve mention gibi alanların sayısal hale getirilmesi:

   Veri setindeki hashtag ve mention gibi alanlar, bu alanlarda kaç adet hashtag veya mention olduğuna dayalı olarak sayısal değerlere dönüştürülmüştür. Bu sayısal değerler, analiz ve işleme süreçlerinde kullanılmak üzere elde edilmiştir. Çizelge 3.2'de bu alanların dönüşüm öncesi ve sonrası halleri gösterilmiştir:

Çizelge 3.2 Hashtag ve mention sayısal dönüşüm işlemi karşılaştırması

Önişleme öncesi	**Sayısal karşılığı**
[{'start': 0, 'end': 8, 'tag': 'Türkiye' }, {'start': 54, 'end': 66, 'tag': 'FOXAnaHaber'}]	2
[{'start': 0, 'end': 7, 'tag': '8M2023'}]	1
[{'start': 18, 'end': 26, 'tag': 'ADSvKSP'}]	1

4. Sayısal değere sahip alanların normalize edilmesi:

   Bu adım, veri setindeki sayısal değerleri aynı ölçekte ifade etmek ve analiz sürecinde doğru sonuçlar elde etmek için önemlidir. Normalize etme işlemi, veri setinin her bir sayısal sütunu için ayrı ayrı uygulanmıştır. İlk adımda, veri setindeki en küçük ve en büyük değerler belirlenmiştir. Ardından, her bir sayısal değer, aşağıdaki formül kullanılarak 0 ile 100 arasında bir değere dönüştürülmüştür. Bu dönüşüm için kullanılan denklem aşağıda verilmiştir:

$$Y = \frac{X - X_{min}}{X_{max} - X_{min}} * 100 \qquad 3.1$$

   Çizelge 3.3'de veri setinin reply alanı üzerinde uygulanan normalizasyon işleminin öncesi ve sonrası gösterilmiştir:

Çizelge 3.3 Reply alanı üzerinde normalizasyon işlemi karşılaştırması

Tweet içeriği	Normalize edilmemiş	Normalize edilmiş
Metrobüs kazası sonrası olay yerine giden AK Parti Gençlik Kolları üyesi bir grup: 'Binlerce yaralı ve sayısı belli olmayacak kadar ölü var.' https://t.co/eRx4jLA2tZ	13486	100
Seçimlere bir ay kala Cumhurbaşkanı Erdoğan, CNN TÜRK-Kanal D ortak yayınında. Fulya Kalfa moderatörlüğünde Ahmet Hakan, Hande Fırat, Abdulkadir Selvi ve Zafer Şahin soracak; Cumhurbaşkanı Erdoğan canlı yayında yanıtlayacak. Cumhurbaşkanı Özel Yayını bugün 22.00'de https://t.co/C3fqrR28YQ	10107	74.94
Buket Aydın: "AK Parti seçmeni sokak röportajlarında konuşmaktan çekiniyor. AK Parti'ye sesleniyorum seçmeninizi ezdirmeyin. Seçmeniniz perişan olmuş."	5536	41.04
Türkiye ve dünya gündemi, lider programları ve uluslararası etkinlikler Gündem sayfasında: https://t.co/jVoEMzcfVy https://t.co/pFnYClR4SO	1	0.007
Kanada, yapay zeka sohbet robotu ChatGPT hakkında "kişisel verilerin izinsiz olarak toplanması, kullanılması ve ifşa edilmesi" şikayetleri üzerine soruşturma başlattı https://t.co/2MJ8hUnuKm https://t.co/wVYPRFwb6n	0	0

Bu ön işleme adımları, toplanan veri setinin doğruluğunu artırmak, verilerin daha anlaşılır ve kullanılabilir hale getirilmesini sağlamak amacıyla gerçekleştirilmiştir.

3.3 Manuel Etiketleme

Veri setinin %10'luk kısmı olan 5,169 tweet, manuel olarak etiketlenmiştir. Bunun nedeni, eldeki veri setinde güvenilir veya güvenilir değil şeklinde etiketlenmiş verilerin bulunmamasıdır.

Etiketleme sürecinde, tüm tweetler NC, LC, C ve HC olmak üzere dört kategoride etiketlenmiştir:

- **NC (Not Credible) - Güvenilir Değil**: Bu etiket, doğruluğu kanıtlanamayan ve herhangi bir kaynağı bulunmayan haberler için kullanılmıştır.

- **LC (Low Credibility) - Az Güvenilir**: Bu etiket, doğruluğu tam olarak tespit edilemeyen ancak birkaç kaynakta paylaşılan haberler için kullanılmıştır.
- **C (Credible) - Güvenilir**: Bu etiket, doğruluğu tespit edilmiş ancak çok fazla kaynak tarafından paylaşılmamış veya tweet içeriğinde bir kısmının güvenilir, bir kısmının ise çok güvenilir olmadığı metinler için kullanılmıştır.
- **HC (High Credibility) - Çok Güvenilir**: Bu etiket, birçok kaynak tarafından paylaşılan ve doğruluğu tespit edilmiş tweet metinleri için kullanılmıştır.

Etiketleme sürecinde, tweet içerisindeki haberlerin en doğru şekilde etiketlenmesi için çaba gösterilmiştir. Teyit.org gibi doğrulama platformları, internet haber siteleri ve arama motorları kullanılarak mümkün olduğunca doğru bir etiketleme yapılması hedeflenmiştir.

Çizelge 3.4'te veri setinde yapılan etiketleme için birkaç örnek verilmiştir:

Çizelge 3.4 Manuel etiketlenmiş veriler

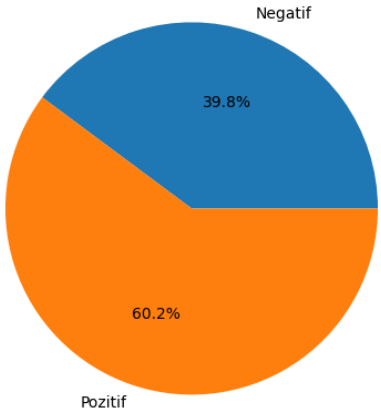
Kullanıcı	Tweet içeriği	Etiket
DarkWeb Haber	Selvi Kılıçdaroğlu: Eğitici oyun setlerini deprem bölgesi çocuklarına getirdik. Çocukların yüzlerindeki gülümseme tarif edilemeyecek kadar güzel. Ama ne yalan söyleyeyim, buralar iyi değil. Çocuklara gülümsedim ama içim çok yaralı. Çok çalışmamız lazım, çok. https://t.co/TU8ooEnYZ0	HC
IBB Haber	Metrobüs seferleri normale döndü.	LC
Yeni Şafak	Türkiye ekonomisinin büyümesine katkıda bulunacak ve finansal piyasaların gelişimini destekleyecek İstanbul Finans Merkezi pazartesi günü hizmete girecek. https://t.co/XQkxV5FIPj	C
zaytung	Eski Kiracı Olduğu İçin Ucuza Oturduğu 35 Yıllık Binanın Deprem Kontrolüne Gireceğini Öğrenen Ezgi Kuray(29), Google'a Girdi: "Kolon nasıl güçlendirilir? Enter..." https://t.co/zWxxS4eVVj https://t.co/lf69Be68AL	NC
Teşkilat Haber	ANKARA METROSUNDA REZALET! ANKARALI PERİŞAN! Koltuğu Devraldığı Günden itibaren Ankara'yı Yönetemeyen CHP'li Belediye Başkan'ı Mahsur Yavaş'tan bir rezalet daha! Ankara Metrosu'nda saatlerce çözülemeyen arıza! Yolda kalan Ankaralılar'dan CHP'li Mansur Yavaş'a istifa çağrısı. https://t.co/llHPm8iZGH	NC

Etiketleme işlemi tamamlandıktan sonra, duygu analizi yapmak için "bert-base-turkish-sentiment-case" (Yıldırım, 2020) adlı bir model kullanılmıştır. Bu model, Türkçe duygu analizi için bir makine öğrenimi modelidir. Duygu analizi, her bir tweetin içerdiği duygusal eğilimi belirlemeyi amaçlar.

Kullanılan duygu analizi modeli, temel olarak BERT (Bidirectional Encoder Representations from Transformers - Dönüştürücülerden İki Yönlü Kodlayıcı Temsilleri) adlı popüler bir dil modelinin Türkçe duygu analizi için özelleştirilmiş halidir. BERT, bir metni anlamak ve onun içerdiği duygusal, anlamsal veya dilbilgisel yapıları çözebilmek için önceden eğitilmiş bir dil modelidir (Devlin, Chang, Lee, & Kristina, 2019). Bu sayede, verilen metinlerin anlamlarını ve içerdikleri duyguları daha doğru bir şekilde anlamak mümkün olur.

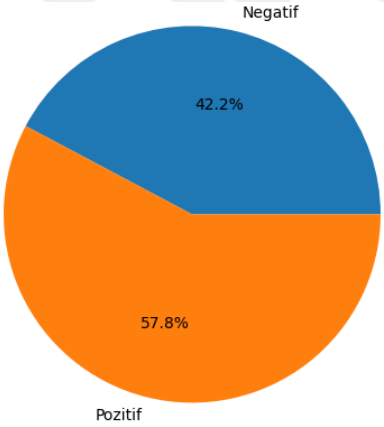
"bert-base-turkish-sentiment-case" modeli, tweetlerin içerdiği duyguları tespit edebilmek için eğitilmiştir. Model, metinde pozitif, negatif veya nötr duygusal eğilimleri saptayabilmektedir. Bu şekilde, her bir etiket altında yer alan tweetlerin içerdiği duygusal içerikleri anlamak ve analiz etmek mümkün olmaktadır.

Duygu analizi işlemi, her etiket kategorisine ilişkin duygusal içeriğin ne olduğunu anlamak amacıyla gerçekleştirilmiştir. Yani, her bir etiket altında yer alan tweetlerin içerdiği duyguların tespit edilmesi ve analiz edilmesi hedeflenmiştir. Bu sayede, belirli etiketlerle ilişkilendirilen duygusal eğilimler ve temaslar daha ayrıntılı bir şekilde anlaşılmış olur.



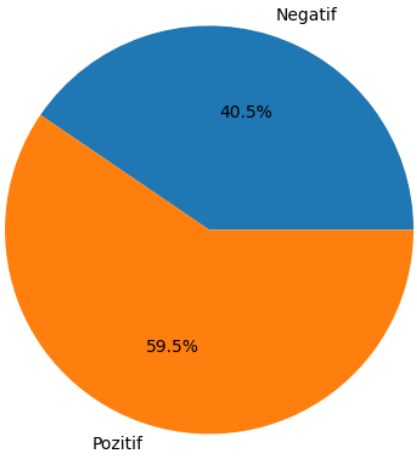
Şekil 3.3 Güvenilir değil etiketine sahip tweetlerin duygu analizi dağılımı

Şekil 3.3'de Güvenilir değil (NC) etiketine sahip tweetlerin %60.2 oranında pozitif duygu analizi içerdiğini gösterilmektedir.



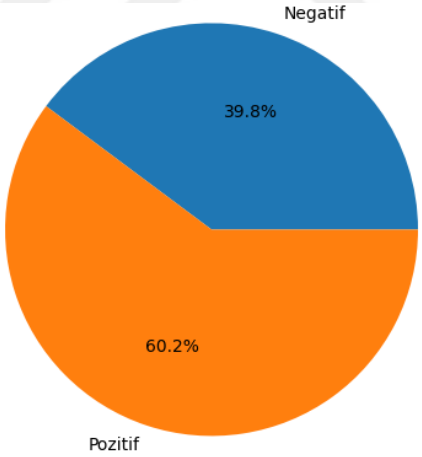
Şekil 3.4 Az güvenilir etiketine sahip tweetlerin duygu analizi dağılımı

Şekil 3.4'de Az Güvenilir (LC) etiketine sahip tweetlerin %57.8 oranında pozitif duygu analizi içerdiğini gösterilmektedir.



Şekil 3.5 Güvenilir etiketine sahip tweetlerin duygu analizi dağılımı

Şekil 3.5'de Güvenilir (C) etiketine sahip tweetlerin %59.5 oranında pozitif duygu analizi içerdiğini gösterilmektedir.

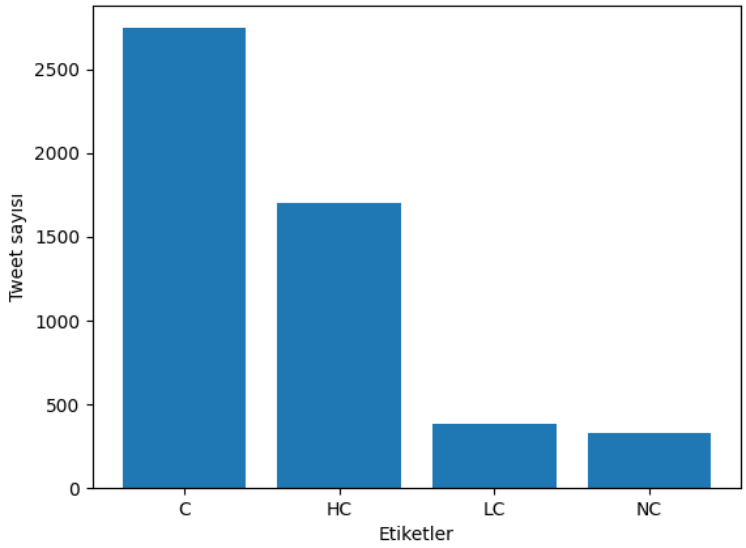


Şekil 3.6 Çok güvenilir etiketine sahip tweetlerin duygu analizi dağılımı

Şekil 3.6'te Çok Güvenilir (HC) etiketine sahip tweetlerin %60.2 oranında pozitif duygu analizi içerdiğini gösterilmektedir.

Yukarıda verilen, duygu analizi dağılımları sonucunda tüm etiket kategorilerinin genel olarak pozitif duygu analizine sahip olduğunu göstermektedir. Yani, etiketlenmiş

Twitter verilerinde yer alan tweetlerin çoğunlukla olumlu yönde duygusal içerik taşıdığı görülmüştür.



Şekil 3.7 Tweetler arasında etiketlerin dağılımı

Şekil 3.7 etiketlenmiş veri setinde yer alan tweetlerin etiketleri arasındaki dağılımı gösteren çubuk grafiği temsil eder. Etiketler ekseni, tweetlerin farklı kategorilere ait olduğunu temsil ederken, y eksenindeki sayılar, her bir etiketin veri kümesinde kaç kez tekrarlandığını göstermektedir. Grafik, etiketler arasındaki dağılımı anlamak ve verilerin etiket kategorileri üzerindeki dağılımını görsel olarak incelemek için kullanılmıştır.

3.4 Güvenilirlik Skoru Hesaplaması

Bu çalışmada, tweetlerin ve tweetleri paylaşan kullanıcıların güvenilirlik skorları ayrı ayrı hesaplanmıştır.

3.4.1 Tweet Güvenilirlik Skoru

Tweet güvenilirlik skorları, iki farklı bileşen üzerinden hesaplanmıştır: tweet metin yazısı güvenilirlik skoru ve tweet sosyal güvenilirlik skoru. Tweet metin yazısı güvenilirlik skoru, tweet içeriğinin güvenilirliğini değerlendirmek amacıyla kullanılmıştır. Bu skor, metin

içerisindeki ifade ve bilgilerin güvenilirlik düzeyini temsil etmektedir. Öte yandan, tweet sosyal güvenilirlik skoru, tweetin etkileşim aldığı sosyal alanlardan (Örneğin; beğeniler, retweetler, yanıtlar) elde edilen bilgileri kullanarak tweetin popülerliğini, kullanıcıların tepkisini ve güvenilirlik düzeyini değerlendirmeyi amaçlamaktadır.

3.4.1.1 Tweet Metin İçeriği Güvenilirlik Skoru

Tweet metin yazısı güvenilirlik skorunu hesaplamak için belirli bir formül kullanılmıştır. Formül, tweet içeriğinin farklı özelliklerinin ağırlıklı toplamını alarak bir skor üretir.

Her tweet için hesaplanan güvenilirlik skoru, şu şekilde ifade edilebilir:

$$(w_1 * x_1) + (w_2 * x_2) + .. + (w_n * x_n) \quad 3.2$$

Burada, w_1 w_2 w_n tweet içeriğinin farklı özelliklerine atanmış olan ağırlıkları temsil eder. Özellikler arasında tweette bulunan URL'ler, içerik uzunluğu, medya varlığı, öznel kelimelerin varlığı ve diğer benzer faktörler yer almaktadır. Bu ağırlıklar, her bir özelliğin önem düzeyini yansıtır. Çalışmada, güvenilirlik skorlamasında hangi alanların önem düzeyinin daha fazla olması gerektiği ile ilgili önceden yapılmış bir çalışma olmadığı için, deneme-yanılma yöntemi tercih edilmiştir. Skorlama yapılacak alana en uygun olduğu düşünülen özelliklere daha fazla ağırlık verilerek önem düzeyleri saptanmıştır. Örneğin, tweetlerde bulunan URL'lerin güvenilirlik skorlamasında daha önemli olduğu gözlemlenmiştir. Çünkü URL'ler, kaynak referansı sağlayarak tweet içeriğinin doğrulanabilirliğini artırabilir ve güvenilirlik açısından önemli bir bilgi sağlar. Benzer şekilde, metnin uzunluğunun da güvenilirlik skorlamasında önemli olduğu belirlenmiştir. Kısa metinlerin eksik veya yanıltıcı olabileceği, daha uzun metinlerin ise daha fazla içerik ve anlam sağlayabileceği düşünülmektedir.

x_1 x_2 x_3 ise, her bir özelliğin değerleridir. Örneğin; URL'lerin sayısı, içerik uzunluğu, medya varlığı gibi özelliklerin değerleri bu formülde kullanılır.

Çizelge 3.5'te Tweet metin güvenilirlik skoru formülünde kullanılan alanların ağırlıkları ve açıklamaları verilmiştir:

Çizelge 3.5 Tweet metin güvenilirlik skoru ve alan ağırlıkları

w	x	Açıklama
0.29	urls	URL'ler tweet içeriğindeki kaynakların referanslarını sağlar ve bu referanslar, tweetin güvenilirliğini artırır.
0.25	context_annotations	Metin etiketleri, tweetin içeriğini daha fazla kategorize etmek ve anlamını belirlemek için kullanılır ve tweetin güvenilirliğine katkıda bulunur.
0.18	text_len	Daha uzun tweetler genellikle daha fazla bilgi ve içerik barındırırken, daha kısa tweetler daha kısıtlı ve eksik olduğu sayılmıştır.
0.18	media_keys	Medya öğeleri, tweetin daha görsel ve etkileşimli olmasını sağlar ve tweetin ilgi çekici ve güvenilir olmasına katkıda bulunur.
0.04	has_subjective_words	Tweet içeriklerindeki öznel ifadeler, güvenilirliği etkileyebilecek ve nesnellikten uzaklaşmaya yol açabilir.
0.03	has_bad_words	Küfürlü veya uygun olmayan içerik, tweetin güvenilirlik düzeyini olumsuz etkileyerek düşük bir güvenilirlik skoruna neden olabilir.
0.03	possibly_sensitive	Eğer bir tweet, hassas veya uygunsuz içeriğe sahipse, genel güvenilirlik düzeyini etkileyebilir ve daha düşük bir güvenilirlik skoru alabilir.

3.4.1.2 Tweet Sosyal Güvenilirlik Skoru

Tweet metin yazısı güvenilirlik skoru hesabına benzer şekilde, tweet sosyal güvenilirlik skoru da hesaplanmıştır. Tweet sosyal güvenilirlik skoru, tweetin sosyal etkileşim düzeyini yansıtmayı hedeflemektedir.

$$(w_1 * x_1) + (w_2 * x_2) + .. + (w_n * x_n) \quad 3.3$$

Çizelge 3.6'da Tweet sosyal güvenilirlik skoru formülünde kullanılan alanların ağırlıkları ve açıklamaları verilmiştir:

Çizelge 3.6 Tweet sosyal güvenilirlik skoru ve alan ağırlıkları

w	x	Açıklama
0.30	retweet_count	Yüksek bir retweet sayısı, tweetin popülerliğini ve etkisini gösterir. Bu nedenle, retweet sayısının sosyal skora yüksek bir ağırlık verilmesi, tweetin toplumdaki etkisini ve güvenilirliğini yansıtmayı amaçlar.
0.25	impression_count	Bir tweetin yüksek izlenme sayısı, tweetin geniş bir kitleye ulaştığını ve ilgi gördüğünü gösterir. Bu nedenle, izlenme sayısının sosyal skor üzerindeki etkisi önemlidir.
0.20	like_count	Beğeni sayısı, bir tweetin toplum tarafından olumlu bir şekilde karşılandığını göstermektedir. Dolayısıyla, beğeni sayısının sosyal skora etkisi, tweetin popülerliğini ve kabul gördüğünü yansıtmayı hedefler.
0.15	quote_count	Bir tweetin alıntılanması, tweetin içeriğinin dikkate alındığını ve paylaşılan bir konu olduğunu gösterebilir. Alıntılama sayısının sosyal skora katkısı, tweetin etkisini ve kullanıcılar tarafından değer verildiğini yansıtmayı hedefler.
0.10	reply_count	Bir tweetin alıntılama, paylaşma veya yanıt alma gibi etkileşimler alması, tweetin önemli ve tartışmaya açık bir konuyu temsil edebileceğini gösterebilir. Dolayısıyla, yanıt sayısının sosyal skora etkisi önemlidir.

3.4.2 Kullanıcı Güvenilirlik Skoru

Kullanıcı güvenilirlik skorları, kullanıcının genel güvenilirliğini yansıtmak üzere hesaplanmıştır. Bu skor, kullanıcının tweetlerindeki içerik kalitesi, etkileşim düzeyi, kullanıcının hesap sürekliliği ve diğer benzer faktörler göz önünde bulundurularak belirlenmiştir. Kullanıcı güvenilirlik skoru, kullanıcının paylaşımlarının güvenilirliğine dair bir ölçü sağlamaktadır. Bu skor, kullanıcının güvenilirlik düzeyini temsil eden bir sayısal değerdir.

Çizelge 3.7'de kullanıcı güvenilirlik skoru formülünde kullanılan alanların ağırlıkları ve açıklamaları verilmiştir:

Çizelge 3.7 Kullanıcı güvenilirlik skoru ve alan ağırlıkları

w	x	Açıklama
0.25	verified	Kullanıcının doğrulanmış hesap olması, kullanıcının güvenilirlik düzeyini yansıtır.
0.20	followers_count	Yüksek takipçi sayısı, kullanıcının popülerlik ve etkileyicilik düzeyini gösterir.
0.10	days_on_twitter	Twitter'da geçirilen süre, kullanıcının deneyim düzeyini gösterir.
0.10	tweet_count	Yüksek tweet sayısı, kullanıcının aktifliğini ve içerik paylaşımını yansıtır.
0.075	url	Profilde paylaşılan web sitesi URL'si, kullanıcının içerik kalitesi ve güvenilirliğini etkileyebilir.
0.075	location	Konum, kullanıcının yerel veya bölgesel güvenilirliğini temsil eder.
0.025	following_count	Takip edilen hesap sayısı, kullanıcının ilgi alanlarını gösterir.
0.015	listed_count	Listelenme sayısı, kullanıcının uzmanlık veya öneme sahip olduğunu ifade eder.
0.015	names_not_equal	Kullanıcı adı ve gerçek adının farklı olması, kullanıcının kimlik doğrulama düzeyini yansıtabilir.
0.01	description_urls	Profil açıklamasındaki URL'ler, kullanıcının daha fazla bilgi sağlama potansiyelini ifade eder.
0.005	mentions	Profilde bahsedilen hesap sayısı, kullanıcının sosyal etkileşim düzeyini yansıtır.

3.5 Denetimsiz Makine Öğrenmesi

Denetimsiz Öğrenme, etiketsiz veri setlerindeki gizli yapıları veya örüntüleri keşfetmek için kullanılan bir makine öğrenme yöntemidir. Bu yöntem, verilerin etiketleri hakkında bilgi olmadığı durumlarda kullanılır. Sadece verilerin kendisi elimizde bulunur ve bu verilerden sonuçlar çıkarılmaya çalışılır. Denetimsiz Öğrenme algoritması, verileri kendi başına analiz ederek öğrenme sürecini gerçekleştirir. Ancak, verilerin etiketsiz olması nedeniyle elde edilen çıktıları kesin doğrular olarak kabul etmek doğru olmayabilir (Candan, 2021).

Denetimsiz öğrenme yöntemleri arasında yaygın olarak kullanılan bir yöntem kümelemedir. Bu çalışmada, güvenilirlik skorları elde edilmiş olan tweet verilerinin üzerinde

K-Means kümeleme yöntemi uygulanmıştır. Kümeleme, benzer özelliklere veya niteliklere sahip veri noktalarını gruplandırarak veri setini farklı kümeler veya gruplar halinde bölme işlemidir. Bu gruplandırma, veriler arasındaki benzerlikleri ve farklılıkları anlamamıza yardımcı olur.

3.6 K-Means

K-Means kümeleme, n tane gözlemi K tane küme şeklinde gruplandırmak için kullanılan bir yöntemdir. Her gözlem, en yakın merkeze sahip kümeye atanır. Bu yöntem, veri noktalarını benzerliklerine göre K kümelere bölmeyi amaçlar ve küme içi homojenliği artırırken farklı gruplamalar elde eder (Sharma, 2019). Her veri noktası sadece bir gruba aittir ve amaç, küme içindeki veri noktalarını benzerleştirmek ve aynı zamanda farklı kümeler arasındaki uzaklığı en üst düzeye çıkarmaktır. Veri noktaları, küme merkezi ile aralarındaki karelerin toplamını en aza indiren bir kümeye atanır. Küme içindeki değişkenlik ne kadar az olursa, aynı kümedeki veri noktaları birbirine o kadar benzer olur. K-Means algoritması, verileri kümeler halinde düzenleyerek veri setinin içindeki yapıları ve benzerlikleri anlamamıza yardımcı olur (Dabbura, 2018).

K-Means algoritmasının uygulanma adımları şu şekildedir:

1. Oluşturulmak istenen küme sayısı (K) belirlenir.
2. Başlangıç merkezlerini belirlemek için veri kümesini karıştırılır ve K adet veri noktası tekrarsız şekilde rastgele seçilir.
3. Merkezler değişmeyene kadar iterasyonlara devam edilir. Her iterasyonda:

    a. Veri noktaları ile merkezler arasındaki karelerin toplamı hesaplanır.

    b. Her veri noktası en yakın merkeze göre bir küme içerisine atanır.

    c. Her küme için yeni merkezler, kümeye ait veri noktalarının ortalaması alınarak güncellenir.

K-Means algoritması, Beklenti-Maksimizasyon yöntemini kullanarak kümeleme problemini çözer. Beklenti (E) adımında, veri noktalarını en yakın merkeze göre kümelere atarken, Maksimizasyon (M) adımında her kümenin merkezini hesaplar. Bu şekilde

merkezlerin tekrar tekrar güncellenmesi, benzer veri noktalarını bir araya getirerek tutarlı ve farklı kümelemeler elde eder (Dabbura, 2018).

3.7 K-Means Uygulanması

Bu çalışmada, sosyal medya içeriğinin güvenilirlik analizini gerçekleştiren ve tweetleri dört farklı kümeye kategorize eden bir uygulama sunulmuştur. Bu uygulama, K-Means algoritmasını kullanarak tweetlerin metin güvenilirlik skorunu, sosyal güvenilirlik skorunu ve kullanıcı güvenilirlik skorunu değerlendirir ve tweetlerin güvenilirlik seviyesine göre kümelenmesini sağlar. K-Means algoritmasını kullanarak bu üç güvenilirlik skoruna dayalı olarak veri seti dört farklı kümeye ayrılmıştır. Bu kümeler, etiketleme aşamasında olduğu gibi; NC, LC, C ve HC olarak isimlendirilmiştir.

K-Means algoritması, veri setindeki tweetlerin güvenilirlik skorlarına dayalı olarak uzayda benzerlikleri ve farklılıkları belirler. Her bir tweet, en yakın olan güvenilirlik kümesine atanır ve böylece veri seti dört ayrı gruba bölünmüş olur. Bu yöntem sayesinde, tweetlerin güvenilirlik seviyelerine göre sınıflandırılması elde edilmiştir.

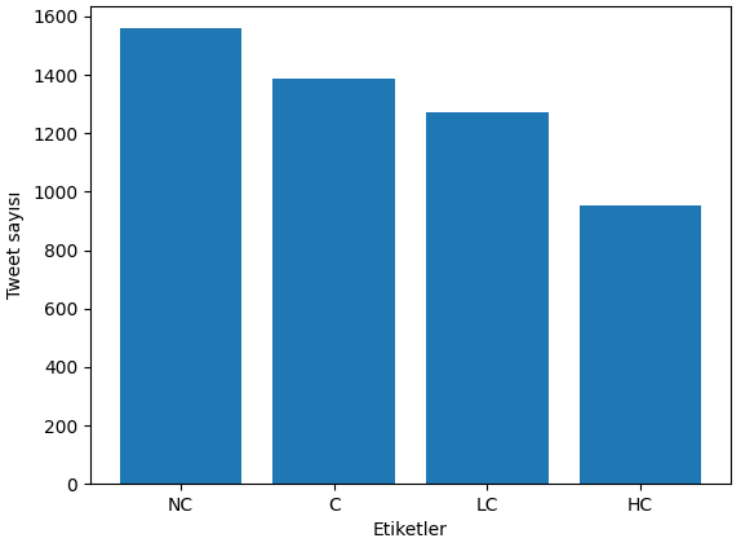
Çizelge 3.8'de, eğitilmiş K-Means modelinin güvenilirlik skorlarına göre veri setinde bulunan bazı tweetlerin manuel etiketleri ve tahmin ettiği güvenilirlik etiketleri verilmiştir:

Çizelge 3.8 K-Means modeli ile tahmin edilen etiketler

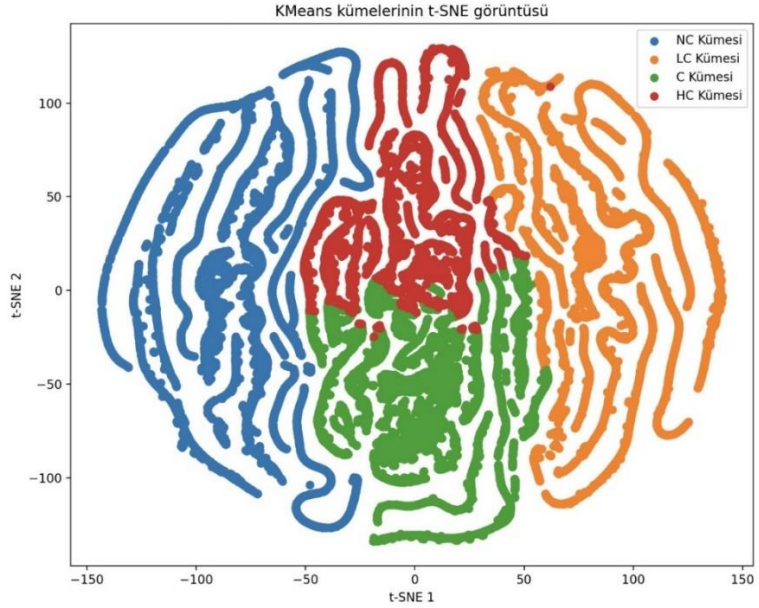
Kullanıcı	Tweet içeriği	Manuel etiket	K-Means etiketi
DarkWeb Haber	Selvi Kılıçdaroğlu: Eğitici oyun setlerini deprem bölgesi çocuklarına getirdik. Çocukların yüzlerindeki gülümseme tarif edilemeyecek kadar güzel. Ama ne yalan söyleyeyim, buralar iyi değil. Çocuklara gülümsedim ama içim çok yaralı. Çok çalışmamız lazım, çok. https://t.co/TU8ooEnYZ0	HC	HC
Haberler.com	Gezi davasında ağırlaştırılmış müebbete mahkum edilen Osman Kavala'nın 14 Mayısla ilgili paylaşımı dikkat çekti.	C	LC
23 Derece	Japonya Başbakanı Fumio Kişida'nın konuşması sırasında büyük bir patlama meydana geldi. Yerel medya, patlama sesinin Kişida'nın bulunduğu noktaya fırlatılan boru benzeri bir nesneden kaynaklandığını ve şüphelinin yakalandığını aktardı. https://t.co/PbMDakPFLM	C	HC
TRT HABER	Elon Musk, Twitter'ın verilerini izinsiz kullandığı iddiasıyla Microsoft'a dava açacak. https://t.co/eeE2uG3f8n	C	C
Son Dakika	#Sondakika İngiltere'de büyük kriz! Başbakan Boris Johnson istifa etti! https://t.co/lr9tinkIQ8 https://t.co/eyJq0JIDIp	HC	NC

Çizelge 3.8'te Haberler.com kullanıcısının paylaştığı tweet için manuel olarak verilen etiket "C" olarak belirlenmişken, K-Means modeli tarafından otomatik olarak verilen etiket "LC" olarak belirlenmiştir. Bu durum, kullanıcı skoru, tweet metin skoru ve tweet sosyal skorunun değerlendirildiği bir analiz sonucunda ortaya çıkmıştır. Özellikle tweetin sosyal skorunun çok düşük olması (0.006), bu otomatik etiketlemenin düşük güvenilirlik kategorisine dahil edilmesine yol açmıştır. Bu tür durumlarda, tweet metin skoru veya tweet sosyal skoru düşük olan tweetler, K-Means modeli tarafından daha düşük etiket kategorilerine atanabilmektedir.

K-Means modeli uygulandıktan sonra elde edilen veri setinde, tahmin edilen etiketlerin dağılımları Şekil 3.8'de gösterilmiştir. Çubuk grafiğe göre, K-Means modelinin tahmin ettiği etiketler arasında NC etiketi C etiketine göre daha yaygın olarak bulunmaktadır. Bu durum, modelin sahte haberlerin oranının daha yüksek olduğunu tahmin ettiğini göstermektedir.



Şekil 3.8 Tweetler arasında tahmin edilen etiketlerin dağılımı



Şekil 3.9 K-Means kümelerinin t-SNE görüntüsü

Şekil 3.9'da K-Means kümelemelerinin t-SNE (t-Distributed Stochastic Neighbor Embedding- t-Dağıtık Stokastik-Rasgele Komşu Gömme) ile gösterimi bulunmaktadır. t-SNE, yüksek boyutlu verileri daha düşük boyutlu bir uzayda görselleştirmek için kullanılan

bir yöntemdir (Maaten & Hinton, 2008). Özellikle verilerin benzerliklerini korumak için tasarlanmıştır ve t-SNE ile yapılan görselleştirmede her veri noktası orijinal veri kümesindeki bir veri noktasını temsil eder (Erdem, 2020). t-SNE, verileri iki boyutlu bir uzayda temsil etmek için genellikle "t-SNE 1" ve "t-SNE 2" olarak adlandırılan iki yeni boyut üretir.

"t-SNE 1" ve "t-SNE 2", t-SNE algoritmasının yaptığı dönüşümler sonucu oluşturulan yeni iki boyutlu uzaydaki koordinatlardır. Bu yeni boyutlar, veri noktalarının görselleştirildiği düzlemdeki konumlarını belirtir.

Bu yeni boyutlardaki değerlerin anlamı, doğrudan orijinal veriyle ilişkilidir ve bu değerlerin belirli bir ölçeği veya özgün bir anlamı yoktur. Örneğin, "t-SNE 1" değeri 50 ve "t-SNE 2" değeri -20 olan bir veri noktası, t-SNE görselleştirmesinde diğer veri noktalarına göre nasıl konumlandığına dair bir bilgi verir, ancak bu değerlerin tek başına anlamı yoktur. t-SNE'nin önemli özelliklerinden biri, yakın olan veri noktalarının görselleştirmede birbirine yakın konumlanmasıdır. Aynı şekilde, uzak olan veri noktaları görselleştirmede birbirinden uzak konumlanır. Bu, veri noktalarının benzerliklerini koruyarak verilerin yapısını görsel olarak anlaşılabilir bir şekilde temsil etmesini sağlar.

BÖLÜM 4

SONUÇLAR VE TARTIŞMA

Bu çalışma, sosyal medya üzerinde Türkçe yanlış veya yanıltıcı bilgilerin yayılmasının engellenmesi ve bu konuda farkındalık yaratılması amacıyla gerçekleştirilmiştir. Özellikle sosyal mecralarda tekrarlayan ve yanıltıcı içeriklerin yayılması, insanların inançlarını çarpıtabilir ve bilgi karmaşasına neden olabilir. Bu çalışma, kullanıcıları yanlış bilgilerden koruyarak, sosyal medyada karşılarına çıkan tweetlerin güvenilirlik seviyelerini belirlemelerine yardımcı olmayı hedeflemektedir.

Bu amaçla, Türkçe Twitter veri seti üzerinde gerçekleştirilen güvenilirlik skorlaması çalışması, kullanıcı güvenilirlik skoru, tweet metin güvenilirlik skoru ve tweet sosyal güvenilirlik skoru hesaplanarak her tweet dört farklı kategoriye (NC, C, LC, HC) bölünmüştür. Daha sonra, veri setinin %10'luk manuel etiketlenmiş bölümü ile K-Means modelinin tahmin ettiği etiketler karşılaştırılmıştır.

Elde edilen sonuçlar, modelin performansını değerlendirmek için kullanılan önemli metriklerle birlikte incelenmiştir. K-Means modeli ile elde edilen başarım değerleri Çizelge 4.1'de verilmiştir.

Çizelge 4.1 Modelin performans sonuçları

Metrik	**Skor**
Doğruluk (Accuracy)	0.283
Kesinlik (Precision)	**0.474**
Duyarlılık (Recall)	0.283
F1	0.333

- **Doğruluk (Accuracy)**:

  Doğruluk, doğru tahmin edilen örneklerin toplam veri sayısına oranını ifade eder. Yani, doğru sınıflandırılmış örneklerin tüm örnekler içindeki yüzdesini belirtir. Doğruluk, aşağıdaki formülle hesaplanır:

$$Doğruluk = \frac{Gerçek\ Pozitifler + Gerçek\ Negatifler}{Toplam\ Örnek\ Sayısı} \quad (4.1)$$

- **Kesinlik (Precision)**:

  Kesinlik, bir sınıfa ait olduğu tahmin edilen örneklerin gerçekten o sınıfa ait olan örneklerin oranını gösterir. Yani, doğru pozitiflerin (TP - Gerçek Pozitifler) yanlış pozitiflere (FP - Yanlış Pozitifler) oranını belirtir. Kesinlik, aşağıdaki formülle hesaplanır:

$$Kesinlik = \frac{Gerçek\ Pozitifler}{Gerçek\ Pozitifler + Yanlış\ Pozitifler} \quad (4.2)$$

- **Duyarlılık (Recall)**:

  Duyarlılık, gerçekten bir sınıfa ait olan örneklerin, model tarafından o sınıfa ait olarak doğru bir şekilde tahmin edilme oranını ifade eder. Yani, doğru pozitiflerin (TP) gerçek pozitiflere (TP + Yanlış Negatifler - FN) oranını belirtir. Duyarlılık, aşağıdaki formülle hesaplanır:

$$Duyarlılık = \frac{Gerçek\ Pozitifler}{Gerçek\ Pozitifler + Yanlış\ Negatifler} \quad (4.3)$$

- **F1 Değeri (F1-score)**:

  F1 Değeri, Kesinlik ve Duyarlılık metriklerinin harmonik ortalaması olarak hesaplanır ve dengeli bir metrik olarak kullanılır. Bu metrik, hem Kesinlik hem de Duyarlılık değerlerini dikkate alarak bir sınıflandırma modelinin performansını ölçmek için kullanılır. F1 Değeri, aşağıdaki formülle hesaplanır:

$$F1 = 2 * \frac{Kesinlik * Duyarlılık}{Kesinlik + Duyarlılık} \quad 4.4$$

Bu çalışmanın sonuçları, güvenilirlik skorlaması ve tweet sınıflandırma alanında önemli bir potansiyele sahip olduğunu ortaya koymaktadır. Elde edilen %47.4'lük yüksek Kesinlik değeri, yanlış pozitiflerin azaltılması ve yanlış haberlerin önlenmesi için önemli bir avantaj sağlar. Aynı şekilde, %28.3'lük Duyarlılık değeri, yanlış negatiflerin azaltılmasını ve gerçekten doğru haberlerin kaçırılmamasını sağlar. Bu sonuçlar, modelin sosyal medya kullanıcılarının yanlış bilgilendirici tweetleri tespit etmelerine ve güvenilir içeriklere daha bilinçli bir şekilde erişmelerine yardımcı olabileceğini göstermektedir.

Gelecekteki çalışmalarda, modelin performansını daha da geliştirmek ve genelleştirilebilirliğini artırmak amacıyla farklı dil ve veri setleri üzerinde test edilmesi önemlidir. Ayrıca, denetimsiz makine öğrenmesi algoritmalarının yanı sıra denetimli makine öğrenmesi algoritmalarının (Örneğin, RFC veya SVM) veya derin öğrenme tekniklerinin kullanılması (Örneğin, BERT veya GPT (Generative Pre-trained Transformer - Üretici Ön Eğitimli Dönüştürücü), modelin performansını artırmak için değerlendirilmelidir. Daha büyük ve çeşitli veri setlerinin kullanılması da, modelin gerçek dünya senaryolarına uygunluğunu ve güvenilirlik analitiğinin etkinliğini artırabilir. Bu şekilde, sosyal medya platformlarında kullanıcıların daha güvenli ve bilinçli bir deneyim yaşamasına katkı sağlayacak daha etkili güvenilirlik analitiği ve yanlış bilgi tespiti yöntemleri geliştirilebilir.

KAYNAKLAR

Aguilera, A., Dongo, I., Mendoza, M. H., Lupa, J. C., & Cardinale, Y. (2022). Credibility Analysis on Twitter Considering Topic Detection. *Applied Sciences, 12*, 9081.

Akyüz, S. S., Kazaz, M., & Gülnar, B. (2021). İletişim Fakültesi Öğrencilerinin Sahte/Yalan Haberlerle İlgili Görüşlerine Yönelik Betimleyici Bir Çalışma. *SELÇUK İLETİŞİM DERGİSİ, 14*, 216-239.

Candan, H. (2021, Dec 26). *Medium*. Retrieved from Adım Adım Makine Öğrenmesi Bölüm 3 : Denetimsiz Öğrenme Nedir?: https://medium.com/machine-learning-t%C3%BCrkiye/ad%C4%B1m-ad%C4%B1m-makine-%C3%B6%C4%9Frenmesi-b%C3%B6l%C3%BCm-3-denetimsiz-%C3%B6%C4%9Frenme-nedir-f890ada49a40

Dabbura, I. (2018, September 17). *K-means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks*. (Towards Data Science) Retrieved from https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a

Devlin, J., Chang, M.-W., Lee, K., & Kristina. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL-HLT*, 4171–4186.

Erdem, K. (2020, Nisan 14). *t-SNE clearly explained*. (Towards Data Science) Retrieved from https://towardsdatascience.com/t-sne-clearly-explained-d84c537f53a

John, A. (2022). Fake News Detection Using Unsupervised and Deep Learning Algorithms. *International Journal of Computer Science and Software Engineering, 7*, 28-37.

Kai, S., Amy, S., Suhang, W., Jiliang, T., & Huan, L. (2017). Fake News Detection on Social Media. *ACM SIGKDD Explorations Newsletter, 19*, 22-36.

Li, D., Guo, H., Wang, Z., & Zheng, Z. (2021). Unsupervised Fake News Detection. *IEEE Access, 3*, 29356 - 29365.

Maaten, L. v., & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research, 9*, 2579-2605.

McCarthy, N. (2018). *Where Exposure To Fake News Is Highest.* Reuters Institute.

Meyers, M., Weiss, G., & Spanakis, G. (2020). Fake News Detection on Twitter Using Propagation Structures. *Multidisciplinary International Symposium on Disinformation in Open Online Media. 12259.* The Netherlands: Springer, Cham.

Sharma, P. (2019, August 19). *The Ultimate Guide to K-Means Clustering: Definition, Methods and Applications*. (Analytics Vidhya) Retrieved from https://www.analyticsvidhya.com/blog/2019/08/comprehensive-guide-k-means-clustering/#What_Is_K-Means_Clustering?

Taşkın, S. G., Küçüksille, E. U., & Topal, K. (2021). Detection of Turkish Fake News in Twitter with Machine Learning Algorithms. *Arabian Journal For Science and Engineering, 47*, 2359–2379.

Yıldırım, S. (2020). *savasy/bert-base-turkish-sentiment-cased.* Retrieved from Hugging Face: https://huggingface.co/savasy/bert-base-turkish-sentiment-cased