

Coğrafi referanslı sosyal medya verilerinden doğal dil işleme ve derin öğrenme kullanılarak duygu analizi: Kahramanmaraş depremleri örneği

Dilan Gözdem Dolu^{1*}, Alper Şen¹

¹Yıldız Teknik Üniversitesi, Davutpaşa Kampüsü, İnşaat Fakültesi, Harita Mühendisliği Bölümü, Esenler, İstanbul.

Öz: Doğal afet yönetiminde afet öncesi hazırlık, afet anı müdahale ve afet sonrası iyileştirme aşamalarında Coğrafi Bilgi Sistemleri (CBS) kullanılarak uygulanan yönetim biçimleri ile doğal afetlerin yol açtığı zararlar en aza indirgenebilmektedir. Bu çalışmanın amacı, 6 Şubat 2023 tarihli Kahramanmaraş depremleri ile ilgili X (eskiden bilinen adıyla Twitter) sosyal medya verilerinin Python programlama dilinde Selenium ve BeautifulSoup kütüphaneleri kullanılarak kazınması ve doğal dil işleme ve derin öğrenme yöntemleri kullanılarak depremden etkilenen insanların afet sonrası duygu durumlarının incelenmesidir. Bu sayede, deprem anında bölge için genel duygu durumu ve sonrası için yapılacak sosyal ve psikolojik iyileştirme çalışmaları planlamalarının CBS ortamında daha hızlı ve kolay bir şekilde yapılmasına katkı sağlanabilecektir. Bu çalışmada, doğal dil işleme kapı yinlemeli birimler (Gated Recurrent Units, GRU) derin ağ modeli kullanılarak X platformundan kazanılmış ve düzenlenmiş veri kümesi üzerinde %87 test doğruluğu ile bir duygu analizi gerçekleştirilmiştir. Ayrıca CBS ortamında sıcak nokta analizi yapılarak yaşanan deprem ile ilgili X iletilerindeki duygu durumuna ait kümelenme örüntüsü istatistiksel olarak incelenmiştir. Böylece, yaşanabilecek depremler ile ilgili X platformuna ait coğrafi referanslı sosyal medya verilerinin duygu analizinde kullanılabileceği saptanmıştır.

Anahtar Sözcükler: Deprem, Coğrafi referanslı sosyal medya verileri, Web kazıma, Duygu analizi, Doğal dil işleme, Derin öğrenme

Sentiment analysis from georeferenced social media data using natural language processing and deep learning: The case of Kahramanmaraş earthquakes

Abstract: In natural disaster management, the damages caused by natural disasters can be minimized by using Geographic Information Systems (GIS) in pre-disaster preparation, disaster response and post-disaster recovery stages. The aim of this study is to scrape X (formerly known as Twitter) social media data related to the Kahramanmaraş earthquakes on February 6, 2023 using Selenium and BeautifulSoup libraries in Python programming language, and to examine the post-disaster emotional states of people affected by the earthquake using natural language processing and deep learning methods. Thus, it will be possible to contribute to the planning of the general emotional state of the region at the time of the earthquake and the social and psychological rehabilitation activities to be carried out afterwards in a faster and easier way in the GIS environment. In this study, a sentiment analysis was performed with 87% test accuracy on the scraped and organized dataset from the X platform using the Gated Recurrent Units (GRU) deep network model in natural language processing. In addition, by performing hot spot analysis in the GIS environment, the clustering pattern of the emotional state in X messages related to the earthquake occurred was statistically analyzed. Thus, it was determined that georeferenced social media data of the X platform related to possible earthquakes can be used in sentiment analysis.

Keywords: Earthquake, Georeferenced social media data, Web scraping, Sentiment analysis, Natural language processing, Deep learning

* Sorumlu Yazar/Corresponding Author: Tel: +90 212 383 5344

Geliş Tarihi/Received: 14.01.2024

Kabul Tarihi/Accepted: 26.03.2024



0000-0002-9585-4054, gozlem.dolu@std.yildiz.edu.tr (Dolu D. G.)*
0000-0002-7236-6701, alpersen@yildiz.edu.tr (Şen A.)

1. Giriş

Modern teknolojinin hızlı gelişmeleri sayesinde sensör ağları ve uzaktan algılama tekniklerinin kurulması yoluyla doğal afetlerin izlenmesi, önceden uyarılması ve afetlerden sonra yaşanan durumların genel incelenmesi mümkün hale gelmiştir (Feng vd., 2022). Doğal afet yönetiminde de afet öncesi hazırlık, afet anı müdahale ve afet sonrası iyileştirme aşamaları mevcuttur (Neal, 1997). Doğal afetlerin yönetiminde Coğrafi Bilgi Sistemlerinin (CBS) 1960'larda teknoloji ile mantık arasındaki bağlantıyı ortaya çıkarmasıyla başlayan süreç (Clarke, 1997), günümüzde afet yönetimi aşamalarında yaygın bir şekilde diğer sistemlerle entegre edilen CBS'nin kullanımıyla devam etmektedir. Bu bütünleşmiş sistemlerden biri de sosyal medyadır. Sosyal medya, X (eskiden bilinen adıyla Twitter) ve Threads gibi küresel ölçekte yaygın kullanılan platformlarıyla geniş bir kullanıcı kitlesine erişim sağlar. Kullanıcılar paylaşımlarıyla sosyal medyayı haberleri, etkinlikleri veya ürünleri değerlendirdikleri, kendi düşüncelerini ve duygularını aktardıkları bir platform olarak görürler. Sosyal medya verileri, işletmeciler için ürünleri ve piyasadaki marka izlenimleri hakkında bilgi verebileceği gibi sosyal olaylardan sonra vatandaşların duygu durumlarını da analiz etmek için kullanılabilir. Elde edilen verilerin analizleri piyasa fiyat izleme ve planlama çalışmalarında oldukça yaygın kullanılmaktadır. Bu durum, sosyal medya veri işleminin çalışmalarda sık kullanılmasına ve kullanımının artmasına neden olmaktadır (Bhardwaj vd., 2015).

Sosyal medya kaynağının sağladığı çeşitli mekânsal verilerin kullanılması ile ilgi noktalarının belirlenmesi, olayların zamansal ve mekânsal dağılımının değerlendirilmesi, belirli bir zamanda ve mekânda olayların yoğunluğunun belirlenmesi gibi farklı mekânsal problemlerin analizlerine katkı sağlamak mümkündür. Sosyal medya üzerinden anlık veri çekilebilmesi, büyük ve çeşitli veriler sağlanabilmesi nedeniyle potansiyel uygulama alanlarından biri doğal afetlerdir (İlhan & Sağaltıcı, 2020). Sosyal medya içerisindeki bilgiler veri kazıma ile elde edildikten sonra farklı analizlerde kullanılarak müdahale ve iyileştirme çalışmalarına katkı sağlanabilir. Sosyal medya platformları uzun metinler, net fotoğraflar, mekânsal bilgiler ve uzun videolar gibi kaliteli bilgileri içerisinde barındırmaktadır. Sosyal medyadan bilgi verimli bir şekilde elde edilebilir; metin ve görsellerin yorumlanmasıyla yeni içerikler elde edilebilir (Feng vd., 2022). Mekânsal veriler çalışmalarda coğrafi etiketli veriye dayalı uygulamalar için kullanılabilirken, metinsel veriler duygu analizinin yapılması gibi çeşitli uygulama alanlarında kullanılabilir. Sosyal medya verileri, doğal dil işleme (Natural Language Processing, NLP) tekniklerinden biri olan duygu analizinde de kullanılabilir (Adalı, 2012). Duygu analizi yapılırken birçok makine öğrenmesi yöntemi uygulanabilir. Bunlardan biri olan derin evrişimli sinir ağları sayesinde bilgisayarlı görü (computer vision) ve doğal dil işlemede büyük ilerleme sağlanmıştır (Feng vd., 2022).

Son yıllarda deprem gibi doğal afetler için sosyal medya verilerinin araştırılması ve analizi sayesinde elde edilen bilgilerin coğrafi bilgi sistemleri ile kullanılması önem arz eder. Sosyal medya platformları metinler, coğrafi konumlar, fotoğraflar, videolar ve sesler gibi veri çeşitlerini içerisinde barındırır. Bu platformlar, kullanıcıların farklı içerikleri paylaşmasına ve etkileşimde bulunmasına olanak tanır ve böylece kullanıcılar arasında bilgi paylaşımı sağlanır. Sosyal medya veri kazıma çalışmalarından elde edilen metinler üzerinden zamansal ve mekânsal (geometrik ve semantik) veriler elde edilebilir. Veri kazıma sonucu elde edilen iletilerin zaman dilimi, konuları ve kullanıcıların iletilerinin gönderimi sırasında paylaşma açtıkları konum bilgileri elde edilebilir (Kaya, 2021). Sosyal medya ortamında mevcut olabilen konum bilgisi, kullanıcılarının iletiyi paylaştıkları yerin coğrafi konum bilgileri, sözel olarak yer alan kullanıcı profilindeki yerleşim yeri, paylaşımlarında bahsedilen konumlar ve doğrudan paylaşılan enlem-boyam gibi konumlar aracılığıyla belirlenebilir (Gulnerman & Karaman, 2020). Mekânsal veri analizi, verilerin mekânda nasıl etkileşime girdiğini yöntemlerle incelerken, aynı zamanda bu verilerin diğer mekânsal vakalarla olan bağlantılarını araştıran bir yaklaşımdır. Bu tür analizler, değişkenlerin mekânsal dağılımını anlamayı, örüntülerin tanımlanmasını, mekânsal kümelenmeyi ve değişkenler arasındaki bağlantıları çözümlenmeyi amaçlayan çeşitli teknikleri içerir (Özgür & Aydın, 2011).

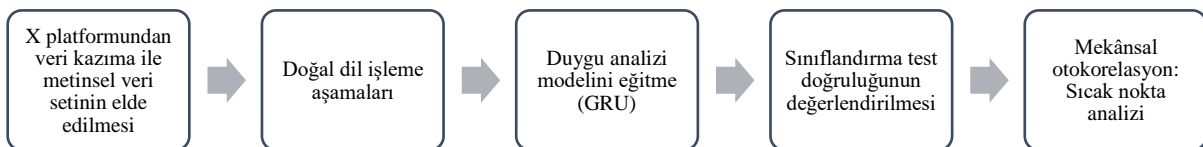
Sosyal medya verileri kullanılarak yapılan mekânsal analiz çalışmaları şu şekilde sıralanabilir: Zhou ve Xu (2017) tarafından, New York City - Washington DC bölgesindeki olayların X verileri üzerinden mekânsal ve zamansal dağılımları incelenmiştir. Bu çalışmada, Geo API'den toplanan Küresel Konumlama Sistemi (Global Positioning System, GPS) bilgisini içeren iletiler kullanılmıştır. Geo API'den toplanan iletilerde konum içeriği, zaman dilimleri, yer adları ve GPS ölçümleri gibi en az bir tür konum verisi bulunur. Elde edilen veriler içerisinde yağmur gibi rastgele olayların yanı sıra planlı etkinlikler ile ilgili, örneğin Papa'nın ziyareti gibi, konuların X kullanıcıları üzerindeki etkileşimleri belirlenmiştir. Rastgele orman algoritması ve mekânsal-zamansal analizler kullanılarak incelenen olayların, X platformundaki iletilerde artışa neden olduğu tespit edilmiştir. X üzerinden olaylara olan anlık tepkileri ve ilgili mekânsal-zamansal örüntüleri belirleyerek sosyal medyadaki etkileşimleri anlamaya yönelik bir metodoloji sunmuştur. Wang vd. (2018) tarafından, doğal afetlerden biri olan sel konusundaki sorunlar hakkında X üzerinden API ile elde edilen veriler kullanılarak NLP ve bilgisayarlı görü teknikleri ile bir yaklaşım belirlenmiş; X tabanlı verilerin sokak adları ölçeğinde hassasiyet sağladığı tespit edilmiştir. Gulnerman ve Karaman (2020) tarafından yakınlık analizi kapsamında kent yol ağı poligon kırık noktalarından Voronoi alanları üretilerek bu alanlarda gönderilen iletilerden yer adları elde edilmiş ve bir coğrafi sözlük üretilmiştir. Gulnerman vd. (2020a) tarafından, coğrafi ileti indiricisi (Geo Tweets Downloader, GTD) adlı sistem ile elde edilen coğrafi etiketli verilerden mekânsal bilgilere ulaşılmıştır. GTD, coğrafi etiketi olmayan iletileri atlayabilen, X API'leri kullanarak gerçek zamanlı olarak genel durum iletilerini sağlayabilen ve bu iletilerin coğrafi etiketleri de dahil olmak üzere çeşitli bilgilerini toplayabilen bir sistemdir. Bu çalışmada, İstanbul'daki kullanıcıların sosyal medya verileri ile davranışlarını anlamaya yönelik bir inceleme yapılmıştır. Veri analizi ile kullanıcı etkinliklerinin temsil düzeylerinin, mekân-zaman önyargılarının ve şehirdeki trendlerin belirlenmesi çalışmaları gerçekleştirilmiştir. Aktif kullanıcıların verilerinin temizlenmesi, farklı temsil düzeylerinin vurgulanması ve mekânsal-zamansal anomali eğilimlerinin tespit edilmesi gibi yöntemler kullanılarak elde edilen bulguların, şehirdeki günlük rutinlerin tarafsız bir şekilde ortaya çıkarılmasına yönelik potansiyel bir çerçeve sunduğu gösterilmiştir. Sonuç olarak sosyal medya verilerinin acil durumlar veya olağan dışı olayların tespitinde ve şehirdeki vatandaş davranışlarını izleme sistemi oluşturulmasında kullanılabilecek önemli verileri sağladığı vurgulanmıştır. Camacho vd. (2021) tarafından kullanıcıların profilinde belirttikleri konum bilgileri ve iletilerin içerdiği yer adları kullanılarak coğrafi kodlama işlemi gerçekleştirilmiştir. Kullanıcıların profil bilgilerinde belirttikleri konumlar ve iletilerin içeriğinde geçen yer adları, bir coğrafi kodlama hizmeti olan Geocodio aracılığıyla koordinatlara dönüştürülerek mekânsal bilgiler elde edilmiştir. Bu çalışmada, sosyal medyadaki doğal gazla ilgili paylaşımların duygusal içeriği incelenerek mekânsal dağılımı ele alınmıştır. Çeşitli duyarlılık analizi yöntemleri ve makine öğrenmesi modelleri kullanılarak paylaşımlar duygusal kategorilere ayrılmış ve coğrafi konumları belirlenmiştir. Amerika Birleşik Devletleri'nin farklı bölgelerinde doğal gazla ilgili paylaşımların duygusal dağılımı gösterilerek, zaman içinde duygusal içeriğin değişimi de ortaya konmuştur. Duygu analizi ve mekânsal analiz yöntemleri entegre edilerek sosyal medyadaki duygusal içeriğin coğrafi dağılımı sunulmuştur. Gulnerman vd. (2020b) tarafından, sosyal medya üzerinden GTD kullanılarak veriler elde edilmiş; coğrafi koordinatlara sahip iletilerin bilgisi mekânsal analiz ve metin analizi ile birleştirilerek olayları anlama sürecine katkı sağlanması amaçlanmıştır. Mekânsal verilerden daha kesin bilgi çıkarımını belirlemek için optimize edilmiş sıcak nokta analizi ile belirli bir zamanda ve mekânda olayların yoğunluğunun belirleme çalışması yapılmıştır. Bu çalışmada, coğrafi referanslı sosyal medya verilerinin acil durum izlemede faydalı olabileceği ele alınmıştır. Feng vd. (2022) tarafından, doğal afetlerle ilgili sosyal medya içeriği analiz edilmiş ve bu verilerden hangi bilgilerin elde edilebileceği ve bunların nasıl kullanılabileceği incelenmiştir. İlk olarak, sosyal medya üzerindeki doğal afet odaklı yayınlar sekiz farklı adımda sınıflandırılmış ve bu verilerin kullanımındaki zorluklar ve fırsatlar tartışılmıştır. Yapılan derleme çalışmasında, sosyal medyanın doğal afetlerle ilgili bilgi çıkarımında ve kullanımında karşılaşılan çeşitli zorluklara ışık tutmuşlardır. Sosyal medyanın doğal afetlerle ilgili veri toplamadaki ve veri analizindeki potansiyelini aynı zamanda da zorluklarını ortaya koyarak mevcut çözümler ile birlikte yeni araştırma fırsatlarını vurgulamışlardır.

X platformu sunduğu özelliklerden dolayı olayların anlık takibini sağlamada eşsiz bir değer sunmaktadır. Kullanıcılar, herhangi bir zamanda ve konumdayken anlık bilgi paylaşımı yaparak, adeta sosyal sensörler gibi işlev görmektedirler (Zhou & Xu, 2017). Afetlerin ardından, toplumun normale dönebilmesi için sadece temel ihtiyaçların değil, bireylerin psikolojik olarak da desteklenmesi gereklidir (Koç & Yalçın, 2023). Aile ve Sosyal Hizmetler Bakanlığının afet süreçlerinde sorumlu olduğu üç ana görev; aynı bağışların depolanması ve dağıtılması, psikososyal desteklerin sağlanması ve risk altındaki gruplara yönelik sosyal çalışmaların yapılmasıdır (2828 sayılı Sosyal Hizmetler Kanunu, 1983; URL-1; URL-2). Psikososyal destek, afet sürecinde oluşabilecek psikolojik sorunları önlerken, insanların topluma uyum sağlamasına, günlük hayata geri dönmesine ve afetle başa çıkma becerilerini geliştirmesine yardımcı olmaktadır (URL-3). AFAD'ın 1 Mart 2023 tarihli basın açıklamasına göre Aile ve Sosyal Hizmetler Bakanlığı, psikososyal destek hizmetleri kapsamında; Kahramanmaraş, Hatay, Osmaniye ve Malatya illerine dört adet mobil sosyal hizmet merkezi kurmuştur (URL-4).

Bu çalışmanın amacı, deprem sonrasında sosyal medya platformuna gönderilen coğrafi referans içeren iletilerin derin öğrenme kullanılarak yapılan duygu analizinin performansının tespit edilmesidir. Bu çalışmada, 6 Şubat 2023 tarihli Kahramanmaraş depremlerinin bölge ve insanlar üzerindeki etkilerinin incelenmesi amacıyla X (eskiden bilinen adıyla Twitter) sosyal medya verileri kullanılarak bir duygu analizi gerçekleştirilmiştir. Python programlama dilinde Selenium ve BeautifulSoup kütüphaneleri kullanılarak verilerin kazınmasının ardından, doğal dil işleme ve kapı yinelenmeli birimler derin ağ modeli kullanılarak depremden etkilenen insanların afet sonrası duygu durumlarının incelenmesi gerçekleştirilmiştir. Bu sayede, deprem sonrasında bölge için genel duygu durumu elde edilebilir. Deprem sonrası için yapılan duygu analizi CBS ortamında sıcak nokta analizleri, zaman-değişim analizleri, yoğunluk analizleri gibi analizlerle kullanılabilir. Bu analizler sayesinde belirli ihtiyaçların ve duygusal tepkilerin yoğunlaştığı bölgeler belirlenebilir. Kaynakların bu ihtiyaçlara ve bölgelere göre verimli kullanılması sayesinde psikososyal destek ekiplerinin güçlerini en etkin şekilde kullanmalarına yardımcı olması sağlanabilir. Ayrıca, deprem bölgesinden gönderilen ileti sayıları CBS ortamında koropleit (renk tonlu) harita ile gösterilmiş ve sıcak nokta analizi yapılarak iletilerdeki kümelenme örüntüsünün varlığı da incelenmiştir. Doğal dil işleme ve derin öğrenme yöntemleri kullanılarak geliştirilen modelin eğitimi sonucunda %87 test doğruluğu elde edilmiştir. Böylece deprem sonrası duygu analizinde sosyal veri kazıma kaynağı olarak X platformunun kullanılabilirliği ve elde edilen sosyal medya verilerinin duygu analizi ile sınıflandırılabilirliği saptanmıştır.

2. Yöntem

Bu çalışmada, 6 Şubat 2023 tarihli Kahramanmaraş depremlerinin bölge ve insanlar üzerindeki etkilerinin incelenmesi amacıyla X (eskiden bilinen adıyla Twitter) sosyal medya verileri kullanılmıştır. Çalışma; X platformundan veri kazıma ile metinsel veri setinin elde edilmesi, doğal dil işleme aşamaları, duygu analizi modelini eğitime, doğruluğunu test etme ve sıcak nokta analizi aşamalarından oluşmaktadır. Bu çalışmada kullanılan akış diyagramı Şekil 1'de gösterilmiştir.



Şekil 1: Çalışmaya ait akış diyagramı

2.1 Veri Kazıma

İnsanların fikirlerini ve bilgi dağarcığını sosyal medya platformları aracılığıyla paylaşmaları büyük veri setlerinin keşfedilmesine yol açmıştır. Bu veri setleri, makine öğrenimi ve veri madenciliği gibi yöntemlerle incelenerek ileriye

öngörmesi aynı zamanda da gizli örüntüleri çıkarması gibi amaçlar için kullanılmaktadır (Albayrak vd., 2017). Bu çalışmada, X sosyal medya platformu üzerinden veri kazıma işlemi gerçekleştirilmiştir. Veriler elde edilirken X platformundan veriyi otomatik ve hızlı bir şekilde kazımak adına Adana, Adıyaman, Batman, Bingöl, Diyarbakır, Elâzığ, Gaziantep, Hatay, Kahramanmaraş, Kayseri, Kilis, Malatya, Mardin, Niğde, Osmaniye, Şanlıurfa ve Tunceli bölgelerinden 685 adet gönderinin kazıma işlemi yapılmıştır.

Büyük verilerin hızlı, otomatik ve sürekli bir şekilde elde edilmesini sağlayan internette veri kazıma işlemi iki adımda yapılabilir. Bu adımlardan ilki internet kaynaklarının elde edilmesi, diğeri ise elde edilen verilerden istenen bilgilerin çıkarılması işlemidir (Zhao, 2017).

İnternet sitelerinin metin tabanlı bir biçimlendirme dili vardır. Bunlardan en yaygın olanı HTML ve XML'dir. Metin tabanlı biçimlendirme dili sayfanın içeriğini tanımlar ve bu sayede içeriğinde olan evrensel biçimleri ve işaretleme dillerinin etiketleri sayesinde istenen bilgiyi almak kolaylaşır. BeautifulSoup kütüphanesi de sayfa kaynağı HTML veya XML olan belgeleri ayrıştırmak için kullanılır (Kılınç vd., 2022). Bu sayede işlenmekte olan ayrıştırmanın kodlamasını otomatik bir şekilde algılayabilir ve bunu istemci tarafından okunabilir bir kodlamaya dönüştürerek anlaşılabilir olmasını sağlar (Zhao, 2017). Bu işlem gerçekleştirilirken sayfa kaynağında bulunan XPath, sınıf ve işaretleme etiketlerinden yararlanır. İşleme sırası ise ilk olarak sayfa içerisinde bütün bilgileri içeren kaynağı indirip içerisinden özellikle istenilen bilgilerin "find_all" çağırısıyla elde edilmesi şeklinde ilerlemektedir (URL-6).

BeautifulSoup kütüphanesi HTML kodundaki bilgileri ayrıştırmaya ve çıkarmaya yönelik bir modüldür. Selenium ise Google Chrome ve Internet Explorer gibi web tarayıcısı ile bağlantı kurup bunun üzerinden kimlik doğrulama, yeniden yönlendirme, çerezler gibi http istekleri ve sosyal medya platformlarına giriş yapmada istenen bilgilerinin otomatik girilmesini sağlayan Python kütüphanesidir. Temelde tarayıcı sürücülerini (web driver) üzerinden komutları takip ve kabul eden, bunları bir tarayıcıya gönderen otomasyon çerçevesidir (Zhao, 2017). Bu çalışmada, tarayıcı sürücülerinden Google Chrome seçilmiştir. Bu bağlantılarda önemli olan bir diğeri faktör ise kullanılan tarayıcının güncel olmasıdır.

Python programlama dili ile yazılan veri kazıma işleminin akış diyagramı Şekil 2'de gösterilmiştir. X platformundan veri kazıma işleminde tarayıcı sunucu üzerinden işlemlerin optimizasyonu için Selenium kütüphanesi, sayfa kaynağını indirmek için BeautifulSoup ve kazıma sonucu elde edilen verileri bir veri çerçevesine aktarmak için Pandas kütüphaneleri eklenir (Şekil 2a). Tarayıcı sürücünün başlatılması işlemi için Google Chrome sunucularına bağlanılarak X uygulamasına ait internet sayfasına bağlanılır (Şekil 2b). Oturum açma işlemi için kullanıcı adı ve şifrenin koda eklenmesi ile internet sayfasında uygun girişlere gerekli bilgilerin girilmesi işlemi gerçekleştirilir (Şekil 2c). 6 Şubat 2023 anahtar kelimesi (örneğin, "6 Şubat 2023" geocode:39.1070000,39.5472000,30km and "6 Şubat 2023", since:2023-06-01 until:2023-09-30) enlem, boylam, arama çapı ve tarih aralığı gibi bilgilerle birlikte kullanılarak kodun arama sonuçlarını açması işlemi gerçekleştirilir (Şekil 2ç). Sayfa kaynağını yüklemek amacıyla BeautifulSoup kütüphanesi kullanılarak XML formatında sayfanın içeriği indirilir. Bu sayede istenilen anahtar kelime ile elde edilen arama sonuçlarına ait sayfa kaynağı sağlanır. Sayfa kaynağı içerisinden XPath'ler sayesinde istenilen metin ve konum bilgilerinin elde edilmesi için ortam hazırlanır (Şekil 2d). X uygulamasından alınacak iletilerin bir liste içine eklenmesi için ve döngülerde kullanılan döngü tekrar başlangıç sayısı ve maksimum döngü tekrar sayısının tanımlanması işlemi gerçekleştirilir (Şekil 2e). İlk sayfadaki iletilerin XPath'leri alınır ve sayfa kaynağında bulunan ilk sayfadaki iletiler listeye eklenerek gönderi listesi oluşturulur (Şekil 2f). İletilerin gönderi listesine ekleme işlemi gerçekleştirilir. İlk gönderiden başlayarak sayfada bulunan tüm iletilerin teker teker listeye eklenmesi için döngü oluşturulur (Şekil 2g). İlk sayfadan alınan iletilerden sonra diğeri sayfalardan da iletilerin alınması için otomatik kaydırma işlemi gerçekleştirilir (Şekil 2ğ). Tekrarlayan gönderilerden arınmış bir liste oluşturmak amacıyla yeni bir gönderi listesi oluşturulur.

Yeni gönderi listesine, eski gönderi listesinde birden fazla aynı olan iletilerden yalnızca birinin alınması sağlanır. Böylece tekrarlı iletilerden arındırılır (Şekil 2h). Gönderi listesinin 200 adet iletiyi geçip geçmediği kontrol edilir. 200 adet ileti listeye alındıktan sonra kod çalışmayı durdurur (Şekil 2ı). Eğer 200 adetten az ileti varsa döngü 10 kez çalıştırılır. Döngü sayısı maksimum döngü sayısına ulaştığında kod çalışmayı durdurur (Şekil 2i). Döngüden gelen her ileti gönderi listesine eklenir (Şekil 2j). Tekrardan arındırılmış gönderi listesi bir veri çerçevesi içerisine yazdırılır (Şekil 2k).

2.2 Doğal Dil İşleme

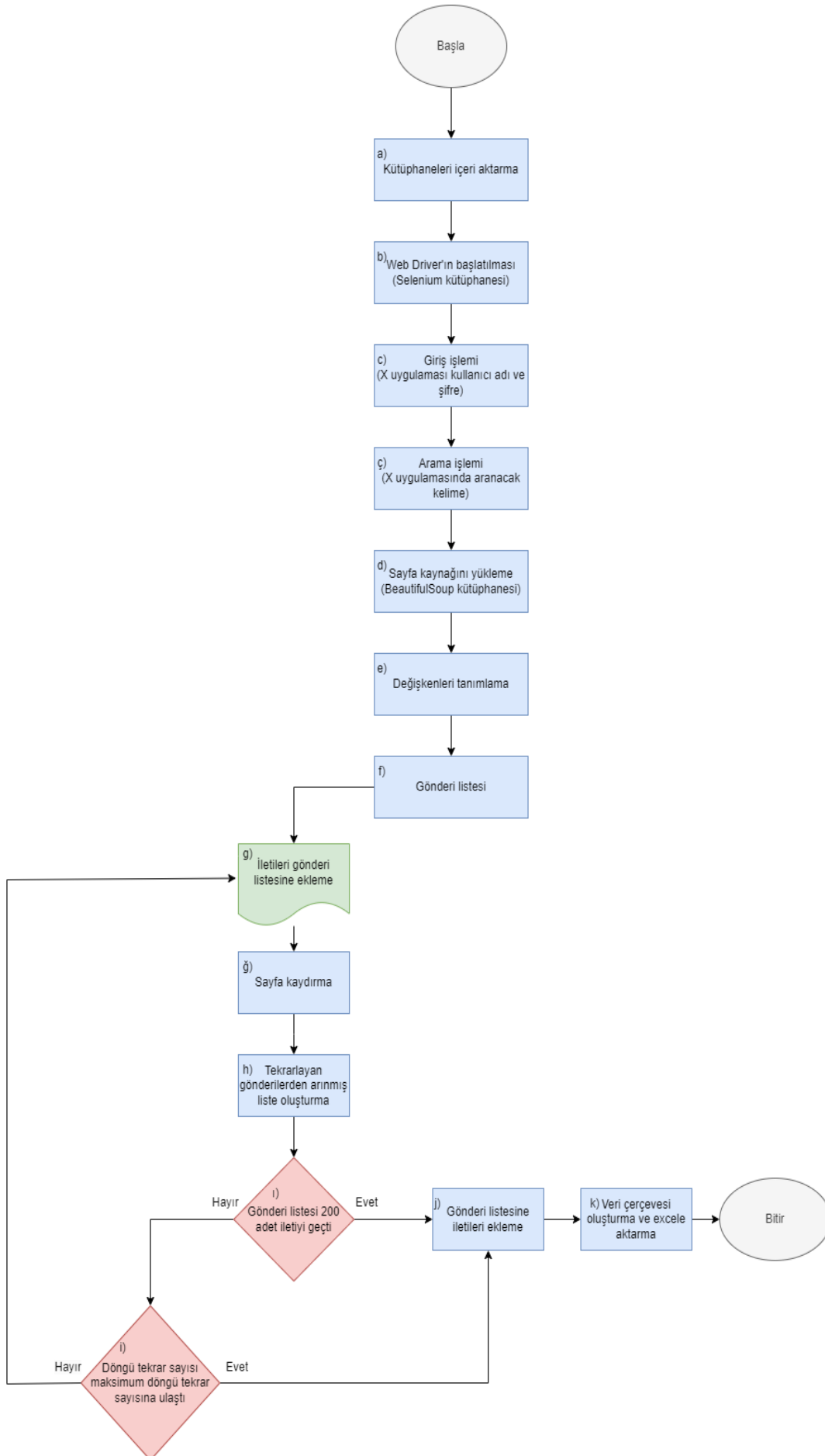
Diller, makine dili ve insanlar tarafından kullanılan doğal dil olarak ikiye ayrılmaktadır. Doğal dillerin makineler tarafından algılanması ve gerektiğinde diller arası bağlantının kurulup çözümlenmesini doğal dil işleme sağlamaktadır (Şeker, 2015). Doğal dil işleme; yapay zekâ alanında bilgisayarların insan dilini anlama, yorumlama ve bu dili kullanarak metinleri veya konuşmaları analiz etme kabiliyetlerini geliştiren bir alt daldır. Bu teknoloji, metinleri sadece kelime düzeyinde anlamakla kalmayıp, cümlelerin ve paragrafların içerdiği anlamı da çözümlenmeyi hedefler (Häberle vd., 2019). Doğal dil işleme günümüzde yaygın olarak diller arası çeviri, mobil telefonlardaki akıllı asistanlar ve çözüm odaklı kullandığımız ChatGPT gibi yapay zekâ asistanlarının temelinde kullanılır (Adalı, 2012). Doğal dil işleme, metinlerin analizlerde kullanılabilir hale getirilmesini sağlayan dil işleme tekniklerini içermektedir (Şeker, 2015).

Hazırlanan veri setine duygu analizi uygulandığında metnin duygusal olarak ne ifade ettiğinin çıkarımı yapılır. Duygu analizi temelde bir metnin pozitif, negatif veya nötr duygularla yazılmış olma durumunu ortaya çıkarır (Albayrak vd., 2017). Pozitif ve negatif duygu çıkarımı yapılabildiği gibi mutlu, üzgün, şaşkın gibi duygu ifadelerinin çıkarımı da yapılabilir (Bhardwaj vd., 2015). Duygu analizinde metin, öncelikle alt kelime birimlerine (tokens) ayrılmaktadır. Metinler içerisindeki kelimelerde edat, bağlaç ve zarfların belirlenmesi ve noktalama işaretlerin temizlenmesi duygu analizinin yapılabilmesi için gerekli işlem adımlarındandır. Böylece metinler üzerinde ayrıştırmalar yapılır ve yüksek frekansa sahip alt kelime birimlerinin yanlılık (bias) oluşturmasının önüne geçilerek metinde sadeleştirme sağlanır. Bu sayede metin içinde her bir kelime bir anlamı ifade edecek şekilde listede yalın şekilde barınmış olur. Bu ön işlemler sonrasında çeşitli makine öğrenmesi sınıflandırma yöntemleri kullanılarak duygu analizi gerçekleştirilmektedir (İlhan & Sağaltıcı, 2020).

Bu çalışmada oluşturulan duygu analizi modeline ait akış diyagramı Şekil 3'te gösterilmiştir. Bu aşamada, iletileri ve olumlu-olumsuz anlamda olduğunu ifade eden etiketleri yükleme işlemi gerçekleştirilir (Şekil 3a). Veri ön işleme amacıyla edat, bağlaç, zarf ve gereksiz kelimeler (stop words) gibi cümlede bulunan fakat tek başına anlamlı olmayan terimlerden verinin arındırılması işlemi gerçekleştirilir (Şekil 3b). K-kat çapraz doğrulama (k-fold cross validation) yöntemi uygulanarak eğitim ve test veri setleri belirlenir (Şekil 3c). Kelime hazinesi belirlenir (en sık geçen 10000 kelime belirlenir). Veri kümesindeki metin verisi kullanılarak bir sözlük oluşturulur. Veri setindeki metinler kelime birimlerine ayrılır. Sözlük içerisinde her bir kelimeye benzersiz bir sayı atanır. Bu sayede metin verisi sayısal bir formata dönüştürülür (Şekil 3ç). Veri setindeki kelime birim sayısı hesaplanır. Her iletinin boyutu farklı olduğundan belirli bir ileti boyutu belirlemek için veri setindeki aritmetik ortalama (ort) kelime birim sayısı (num_{tokens})'na iki kat standart sapma (std) eklenerek bir eşik değeri elde edilir. Eşitlik 1 ile belirli bir ileti boyutunu belirlemek için kullanılacak maksimum kelime birimi sayısı ($maks(tokens)$) hesaplanır.

$$maks(tokens) = ort(num_{tokens}) + 2 \times std(num_{tokens}) \quad (1)$$

Kelime birim sayılarının istatistiksel özellikleri kullanılarak bir maksimum sınırlama getirme amaçlanır. Veri setindeki iletilerin kelime birim sayıları belirlenen eşik değerine göre eşitlenir. Bu işlem, bazı iletilerin sonlarına sıfırlar ekleyerek gerçekleştirilir. Bu sayede tüm iletiler aynı boyuta getirilir. Veri seti, sinir ağı modelinde eğitim ve tahmin etme aşamaları için hazır hale getirilir (Şekil 3d).



Şekil 2: İletilerin elde edilmesi amacıyla veri kazıma işlemi için akış diyagramı

2.3 Kapı Yinelemeli Birimler (Gated Recurrent Units, GRU) Derin Ağ Modeli

Bu çalışmada bir yinelemeli sinir ağı yöntemi olan Kapı Yinelemeli Birim Ağları (Gated Recurrent Units, GRU) kullanılmıştır. Yapay sinir ağları başta olmak üzere tüm sinir ağı yöntemleri nöronlara dayanmaktadır. Yapay sinir ağlarında nöronun girdileri farklı ağırlıklarla doğrusal bir şekilde birleştirilir. Bu kombinasyonun sonucunda ise doğrusal olmayan bir aktivasyon birimine aktarılır (Chen vd., 2020). Ağın yüksek bağlanabilirliği hata etkisinin minimum olmasını sağlar. Yinelenen sinir ağları, veri akışlarını analiz edebilen gizli katmanlara sahip özel bir sinir ağı türüdür. Bu ağlar, çıktılarının önceki hesaplamalara bağlı olduğu problemleri çözmek için oldukça uygun ve etkilidirler. Williams ve Zipser (1989) tarafından tanıtılan bu yöntem, özellikle doğal dil işleme alanında çeşitli problemlerin üstesinden gelmede başarılı olmuştur. Yinelenen sinir ağları, veri akışlarını dinamik bir şekilde analiz edebilme yetenekleriyle öne çıkarlar. Bu özelliği sayesinde, dilin karmaşıklığını ve sürekli değişen yapısını anlama ve işleme konusunda etkili bir araç haline gelmişlerdir. Bu teknoloji, metin analizi, çeviri, duygu analizi ve daha birçok doğal dil işleme problemine yenilikçi çözümler sunmuştur (Küçük & Arıcı, 2018).

Yinelenen sinir ağlarında öğrenme aşamasında karşılaştığı ağırlıkların yok olması sorunu nedeniyle, bu ağların uzun kısa süreli bellek (Long Sort Term Memory, LSTM) (Hochreiter & Schmidhuber, 1997) ve kapı yinelemeli birimler (Gated Recurrent Unit, GRU) (Cho vd., 2014) gibi çözümleri ortaya çıkmıştır (Küçük & Arıcı, 2018). LSTM birimine benzer şekilde, GRU'nun da birim içinde bilginin akışını modüle eden kapı birimleri vardır; ancak ayrı bir bellek hücresi bulunmaz. Ayrıca LSTM'den farklı olarak sadece iki kapısı vardır. Bunlar, unutulacak geçmiş bilginin miktarına karar vermek için kullanılan sıfırlama kapısı ve atılacak veya eklenecek bilgilere karar veren güncelleme kapısıdır (Sachin vd., 2020). Bu çalışmada model, yinelenen sinir ağındaki gradyanların yok olması probleminden dolayı GRU çözümü ile eğitilmiş ve eğitimde GRU katmanları oluşturulmuştur.

$$h_t^j = (1 - z_t^j)h_{t-1}^j + z_t^j h_{tt}^j \quad (2)$$

$$z_t^j = \sigma(W_z x_t + U_z h_{t-1}^j) \quad (3)$$

$$h_{tt}^j = \tanh(W x_t + U(r_t \odot h_{t-1}^j)) \quad (4)$$

$$r_t^j = \sigma(W_r x_t + U_r h_{t-1}^j) \quad (5)$$

Burada; x_t , giriş vektörünü; h_t^j , önceki h_{t-1}^j aktivasyonu ile aday h_{tt}^j aktivasyonu arasında t anındaki doğrusal enterpolasyonu (çıkı vektörünü); z_t^j , güncelleme geçidi vektörünü; U , W_z , W_r , W , ağırlık matrislerini; σ , aktivasyon fonksiyonunu; r_t , sıfırlama geçidi vektörünü; \odot , vektörler arası elementer çarpımı ifade eder. Bu eşitlikler kullanılarak, modelin geçmiş bilgileri hatırlaması, anlık bilgiyi işlemesi ve önemli bilgileri vurgulaması sağlanır (Sachin vd., 2020).

Bu aşamada, model oluşturmak amacıyla modelin ilk katmanına bir gömme katmanı (embedding layer) eklenir. Bu katman, metin verilerini temsili vektörlere dönüştürmek için kullanılır. Üç adet GRU katmanı eklenir. İkili sınıflandırmalı bir sonuç almak için yoğun katman içinde sigmoid aktivasyon fonksiyonu kullanılır (Şekil 3e). Modelin daha iyi öğrenmesi için büyük veri setleri ve karmaşık modellerle çalışırken etkili olan "Adam" optimizasyonu kullanılır. Modelin hazır hale getirilmesi için kayıp fonksiyonu olarak ikili sınıflandırma görevi seçilir. Bu fonksiyon, modelin tahminlerinin referans etiketlerden ne kadar uzak olduğunu ölçer (Şekil 3f). Veri seti, k-kat çapraz doğrulama yöntemi kullanılarak 5 katmana bölünür. Her bir iterasyon için eğitim ve test kümeleri oluşturulur. Model veri üzerinde eğitilir (Şekil 3g). Model, test verileri üzerinde tahminler yapar (Şekil 3ğ). Tahmin olasılıklarında eğer bir tahmin olasılığı 0.5'ten büyükse olumlu (1) sınıfına atama, küçükse

olumsuz (0) sınıfına atama işlemi gerçekleştirilir (Şekil 3h).

2.4 Sınıflandırmada Test Doğruluğunun Değerlendirilmesi

Sınıflandırma performansı çeşitli metriklerle değerlendirilebilir. Sınıflandırma modellerinin test verileri üzerindeki performansını değerlendirmek için hata matrisi kullanılır. Hata matrisi tahmin sonuçlarını gösterir. Hata matrisi içerisindeki Doğru Pozitif (DP) (True Positive, TP), Doğru Negatif (DN) (True Negative, TN), Yanlış Pozitif (YP) (False Positive, FP) ve Yanlış Negatif (YN) (False Negative, FN), modelin tahminlerinin doğruluğunu belirlemede kullanılır. Bu değerlendirmeler, test edilen modelin, referans (ground truth) veriler üzerindeki başarısını gösterir. N, test edilen örneklerin toplam sayısıdır. DP, modelin doğru bir şekilde pozitif olarak sınıflandırdığı durumlardır. DN, modelin doğru bir şekilde negatif olarak sınıflandırdığı durumlardır. YP, modelin yanlış bir şekilde pozitif olarak sınıflandırdığı ve YN, modelin yanlış bir şekilde negatif olarak sınıflandırdığı durumlardır (Markoulidakis vd., 2021). Değerlendirmede doğruluk, kesinlik, duyarlılık ve F1 Skor gibi metrikler kullanılmıştır. Doğruluk ölçütü, doğru tahminlerin toplam test sayısına oranıdır ve başarıyı yansıtır.

$$\text{Doğruluk (Accuracy)} = \frac{DP+DN}{N} \quad (6)$$

Kesinlik (Precision) ölçütü, pozitif olarak sınıflandırılmış örnekleri değerlendirir. Model ne kadar yüksek bir kesinlikle çalışıyorsa, pozitif olarak sınıflandırılan örneklerin gerçekten pozitif olma olasılığı o kadar yüksektir. Kesinlik, doğru olarak sınıflandırılmış pozitif örneklerin, tahmin edilen toplam pozitif örneklere olan oranını ifade eder.

$$\text{Kesinlik (Precision)} = \frac{DP}{DP+YP} \quad (7)$$

Duyarlılık (Recall) ölçütü, pozitif olarak sınıflandırılmış örnekleri değerlendirir. Bu ölçü, doğru olarak sınıflandırılmış pozitif örneklerin, referans verisindeki toplam pozitif örneklere olan oranını ifade eder.

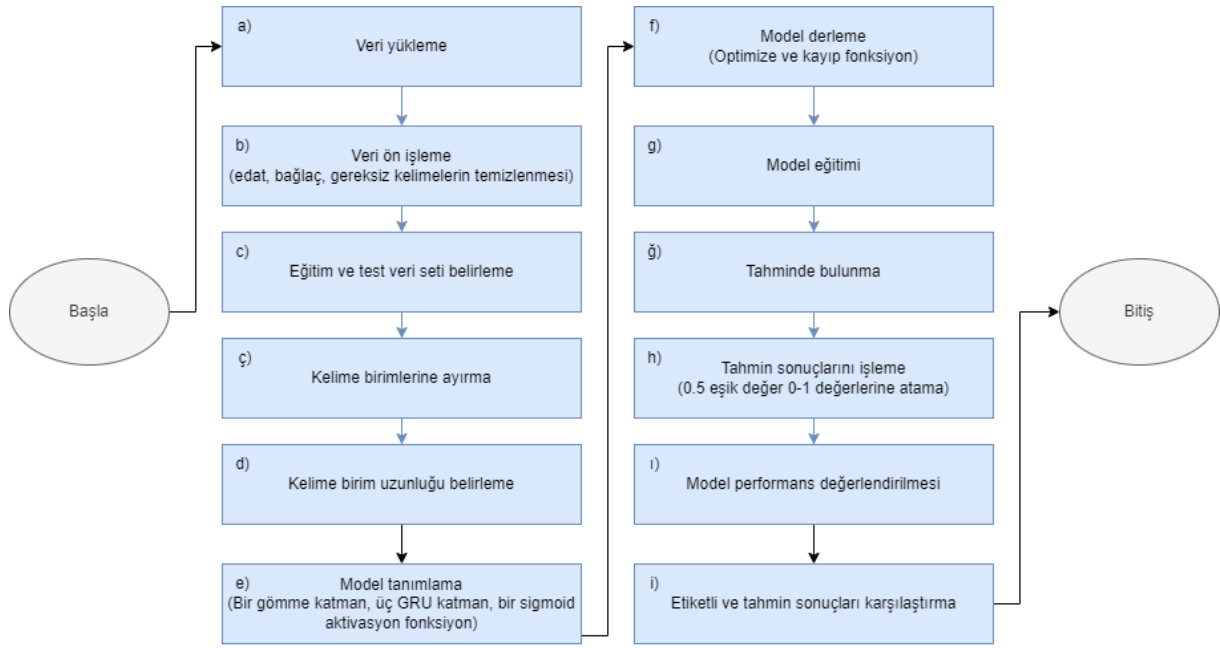
$$\text{Duyarlılık (Recall)} = \frac{DP}{DP+YN} \quad (8)$$

F1 Skor (F1 Score), kesinlik ve duyarlılık ölçütlerinin birlikte hesaplanmasıyla elde edilir. Bu ölçüt, kesinlik ve duyarlılık değerlerinin ağırlıklı harmonik ortalamasının alınmasıyla belirlenir. Yüksek F1 Skor, sınıflandırma performansının kalitesini ifade eder (Gündüz, 2013).

$$\text{F1 Skor} = \frac{2 \times \text{Duyarlılık} \times \text{Kesinlik}}{\text{Duyarlılık} + \text{Kesinlik}} \quad (9)$$

Bu çalışmada, duygu analizi modeli üzerinde eğitim ve test veri setinin belirlenmesi için k-kat çapraz doğrulama yöntemi uygulanmıştır. Çapraz doğrulama yöntemi, bir öğrenme setini eğitmek ve değerlendirmek için kullanılan bir yöntemdir. Öğrenme seti, yaklaşık olarak eşit büyüklükte k adet alt kümeye bölünür. Her bir alt küme sırasıyla doğrulama seti olarak kullanılırken, geriye kalanlar eğitim seti olarak kullanılır. Model k defa eğitilir ve her seferinde bir farklı alt küme doğrulama için ayrılır. Sonuçların ortalaması alınarak çapraz doğrulama performansı elde edilir (Berrar, 2019).

Bu aşamada, model performans değerlendirilmesi işlemi gerçekleştirilir. Performansın değerlendirilmesi için doğruluk, kesinlik, duyarlılık ve F1 Skor sonuçları hesaplama işlemi gerçekleştirilir (Şekil 3i). Etiketli veri ve model tahmin sonuçları il bazında karşılaştırılır (Şekil 3i).



Şekil 3: Duygu analizine ait kod akış diyagramı

2.5 Sıcak Nokta Analizi

Sıcak nokta (hot spot) analizi (Getis-Ord G_i^*), her bir özneteliğin çevresindeki diğer özneteliklerle olan ilişkisini değerlendirerek mekânsal kümelenme desenlerini araştırır. Bir noktanın istatistiksel olarak anlamlı bir sıcak nokta olabilmesi için, bu noktanın yüksek bir değere sahip olması gerektiği gibi, aynı zamanda çevresindeki diğer noktaların da yüksek değerlere sahip olması gerekir. Elde edilen z-istatistiği ve p-değeri, yüksek veya düşük değerlere sahip nesnelerin mekânsal olarak kümelenmesini gösterir (URL-5). Eşitlik 10’da verilen G_i^* yerel istatistiği z-istatistiği değeridir. Örneklemdaki tüm noktalar için G_i^* hesaplanarak, çalışma alanındaki (belli anlamlılık düzeylerinde) sıcak ve soğuk noktalar belirlenir. Pozitif ve belli bir anlamlılık düzeyinde kritik z değerinden büyük G_i^* değerlerine sahip noktalar sıcak nokta; negatif ve belli bir anlamlılık düzeyinde kritik z değerinden küçük G_i^* değerlerine sahip noktalar ise soğuk noktaları oluşturur. Bunun dışında kalan durumlar ise rastlantısal dağılımı gösterir (Çubukçu, 2015).

$$G_i^* = \frac{\sum_{j=1}^n w_{i,j} x_j - \bar{X} \sum_{j=1}^n w_{i,j}}{S \sqrt{\frac{n \sum_{j=1}^n w_{i,j}^2 - (\sum_{j=1}^n w_{i,j})^2}{n-1}}} \quad (10)$$

$$\bar{X} = \frac{\sum_{j=1}^n x_j}{n} \quad (11)$$

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2} \quad (12)$$

Burada; n, örneklemdaki nokta sayısını; x_j , j noktasına ait değişken değerini; $w_{i,j}$, i noktası ve j noktası arasındaki yakınlık ilişkisini veren mekânsal ağırlık değerini; \bar{X} , değişkene ait ortalama değeri belirtir.

3. Bulgular ve Tartışma

Veri kümesi, X platformundan 685 adet iletinin kazıma işlemi ile elde edilmiştir. Veri kazıma aşamasında “6 Şubat 2023” anahtar kelimesini içeren ve Adana, Adıyaman, Batman, Bingöl, Diyarbakır, Elâzığ, Gaziantep, Hatay, Kahramanmaraş,

Kayseri, Kilis, Malatya, Mardin, Niğde, Osmaniye, Şanlıurfa ve Tunceli illerinden gönderilen iletilerin elde edilmesi sağlanmıştır. BeautifulSoup ve Selenium kütüphanesi ile kazıma işlemi yapıldıktan sonra Pandas kütüphanesi ile bir veri çerçevesi oluşturulmuştur.

İletilerden oluşan veri setinin kalitesinin iyileştirilmesi adına, elde edilen verilerin kontrolü gönderilerin özelliklerine göre yapılmıştır. Tablo 1’de bu çalışmada gönderilere ait özellikler verilmiştir. Çalışmada kullanılan iletiler; konusu, beğeni sayısı, gönderi sahibinin birden fazla gönderisinden sadece birinin alınması ve gönderi tarih aralıkları gibi özelliklerine göre filtrelenerek seçilmiştir.

Tablo 1: Gönderilere ait özellikler ve açıklamalar

Gönderiye ait özellikler	Açıklama
Gönderi konusu	6 Şubat 2023 depremleri hakkında bilgi ve görüş içermesi
Gönderi beğeni sayısı	Gönderilerin genel kullanıcılar arasında desteklenme durumu
Gönderi sahibi	Aynı gönderinin aynı kişiden birden fazla atılması
Gönderi tarihi	Haziran 2023 – Eylül 2023

Ön işlemde geçen veri seti ikili olarak olumlu veya olumsuz düşünce belirtmesi durumuna göre etiketlenmiştir. Duygu analizi amacıyla kapı yinelemeli birimler derin öğrenme ağları ile veriler sınıflandırılmıştır. Veriler sınıflandırılırken ilk olarak sinir ağına girdi olarak bir ileti eklenmiştir ve eklenen bu ileti, kelimelerin vektörlerinden oluşan kelime birimleriyle (tokenlerle) temsil edilmektedir. Gömme (embedding) katmanı, ileti içindeki kelimelere karşılık gelen vektörleri içermektedir. Bu vektörler, bir sonraki GRU katmanına çıkış olarak aktarılmaktadır. Bu çalışmada, veri sayısının az olmasından dolayı 3 adet GRU katmanı oluşturulmuştur. Kelime vektörleri, üç GRU katmanında işlenmiştir ve en son çıktı katmanındaki tek bir nörona iletilmiştir. Bu nöron, değerleri tek bir sayıya dönüştürerek sigmoid aktivasyon fonksiyonundan geçirmiştir ve 0 ile 1 arasında bir değer döndürmüştür. Bu değer, modelin çalışılan ileti (X gönderisi) hakkındaki tahminidir. Eğer sonuç 1’e yakınsa, model o iletiyi olumlu olarak değerlendirir; 0’a yakınsa, o gönderiyi olumsuz olarak değerlendirir (Chen vd., 2020).

Duygu analizinde kullanılan sınıflandırmada veriye k-kat çapraz doğrulama uygulanmış ve performansın değerlendirme sonuçları belirlenmiştir. Model, verilen veri setini k-kat çapraz doğrulama yönteminden geçirerek eğitimi tamamlamıştır. Bu çalışmada k = 5 alınmıştır. Burada amaç modelin eğitimini doğru bir şekilde tamamlamış olduğunu test etmektir. Eğitim sonucunda test verisinin doğruluğu Bölüm 2.4’te belirtilen metriklere göre belirlenmiştir. Tablo 2 ve Tablo 3’te derin öğrenme ağında eğitim sonucu modelin performansını değerlendiren test verisinin tahmini hakkında metrik sonuçları genel olarak verilmiştir. Sonuçlara göre, referans veride olumsuz (0) olarak etiketli iken modelin de olumsuz (0) olarak tahmin ettiği ileti sayısı 475 olarak bulunmuştur. Referans veride olumsuz (0) olarak etiketli iken modelin olumlu (1) olarak tahmin ettiği bir sonuç bulunmamaktadır. Referans veride olumlu (1) olarak etiketli iken modelin olumsuz (0) olarak tahmin ettiği ileti sayısı 90 olarak bulunmuştur. Referans veride olumlu (1) olarak etiketli iken modelin de olumlu (1) olarak tahmin ettiği ileti sayısı 120 olarak bulunmuştur. Tablo 2’de hata matrisi olumsuz (0) sınıfını doğru bir şekilde tahmin etme yeteneğinin, olumlu (1) sınıfını tahmin etme yeteneğinden daha yüksek olduğu görülmektedir. Bu durum, kesinlik değerinin yüksek (1.00); duyarlılık değerinin ise düşük (0.57) olmasına yol açmıştır. Bir başka deyişle, referans veride (gerçekte) olumsuz olan iletilerin tümü doğru sınıflandırılmasına rağmen, gerçekte olumlu olan iletilerin %57’si doğru olarak sınıflandırılmıştır. Olumsuz (0) olarak etiketli ileti sınıfının daha net ifadeler ve kısa cümleler içermesi (korkuyoruz, uyuyamıyoruz vb.) modelin bu sınıfı daha iyi öğrenmesine sebep olmuştur. Ancak, olumlu (1) olarak etiketli ileti sınıfının daha karışık ifadeler (örneğin, “üzgünüz ve yanınızdayız” gibi hem olumsuz hem de olumlu duygu durumlarını bir arada bulunduran) ve daha uzun cümleler

içermesi, modelin bu sınıfı daha zor öğrenmesine neden olmuştur. Eğitimini tamamlayan modele test veri seti olarak Adana, Adıyaman, Batman, Bingöl, Diyarbakır, Elâzığ, Gaziantep, Hatay, Kahramanmaraş, Kayseri, Kilis, Malatya, Mardin, Niğde, Osmaniye, Şanlıurfa ve Tunceli konumlarından elde edilen etiketli X iletileri gönderilmiştir. Derin öğrenme ağının Adıyaman, Hatay ve Kahramanmaraş illeri üzerindeki tahmin doğruluk sonuçları Tablo 4’te gösterilmiştir. Adıyaman ili için ortalama test tahmin doğruluğu %82, Hatay için %70 ve Kahramanmaraş ili için ise %90 olduğu tespit edilmiştir. İl bazında ortalama test doğruluklarının birbirine yakın olduğu ve %70-%90 arasında değiştiği görülmüştür. Test doğruluğunun düşük olduğu illerde nötr olarak da değerlendirilebilecek yorumların (olumlu ya da olumsuz ifadelerin daha belirsiz) olması nedeniyle eğitim modelinin daha düşük performans ile çalıştığı gözlemlenmiştir.

Tablo 2: Derin öğrenme ağında test sonucu model performansını değerlendiren hata matrisi

		Tahmin		
		Olumlu (Pozitif)	Olumsuz (Negatif)	Toplam
Referans (Gerçek)	Olumlu (Pozitif)	120 (DP)	90 (YN)	210
	Olumsuz (Negatif)	0 (YP)	475 (DN)	475
	Toplam	120	565	685

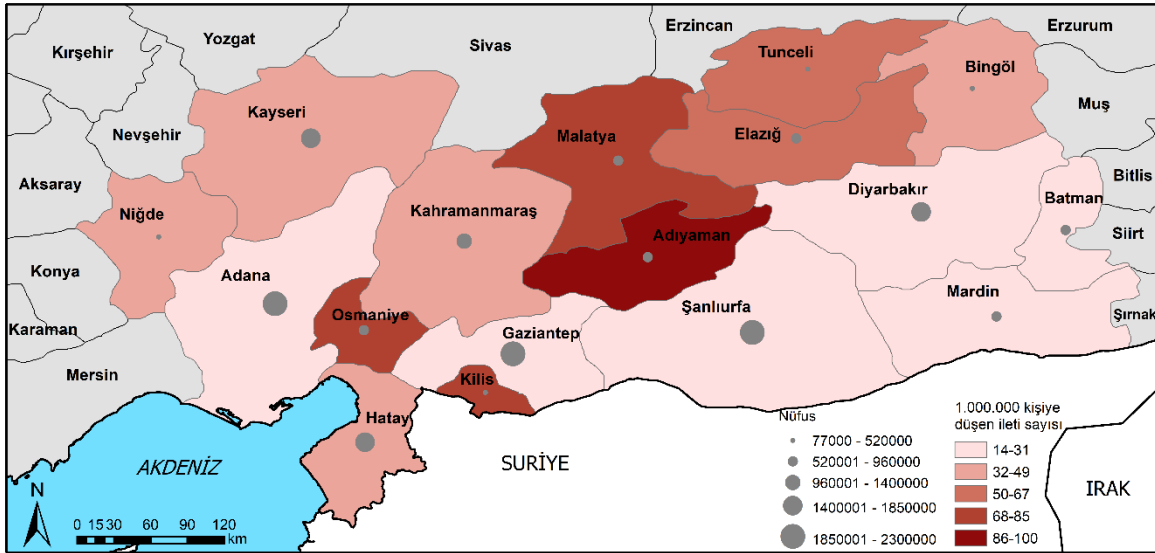
Tablo 3: Derin öğrenme ağında test sonucu model performansını değerlendiren genel metrik sonuçları

Metrik	Genel sonuç
Doğruluk	0.87
F1 Skor	0.73
Duyarlılık	0.57
Kesinlik	1.00

Tablo 4: Derin öğrenme ağının Adıyaman, Hatay ve Kahramanmaraş veri seti üzerinde tahmin performansını değerlendiren metrik sonuçları

Metrik	Adıyaman	Hatay	Kahramanmaraş
Doğruluk	0.82	0.70	0.90
F1 Skor	0.74	0.58	0.87
Duyarlılık	0.82	0.70	0.90
Kesinlik	0.67	0.49	0.83

Duygu analizi modeli sonucunda elde edilen verilerin ileti sayıları, gönderilerin konum bilgileri ve Türkiye İstatistik Kurumu (TÜİK) 2023 illere göre elde edilen nüfus sayıları CBS ortamına (ArcGIS Pro kullanılarak) aktarılmıştır (URL-7). Öncelikle ileti sayılarının nüfusa göre normalize edilmesi işlemi gerçekleştirilmiştir. 6 Şubat 2023 deprem bölgeleri için normalize edilmiş olan iletilerin koroplet (renk tonlu) haritası Şekil 4’te verilmiştir. Ayrıca haritada ikinci bir bilgi olarak 6 Şubat 2023 depreminden etkilenen illerin nüfus bilgisi orantılı işaret haritası olarak sunulmuştur. İletiler Haziran-Eylül 2023 zaman dilimini kapsamaktadır. Şehirlerin nüfus sayılarına oranla 6 Şubat 2023 anahtar kelime iletilerin en az paylaşımının Şanlıurfa, Diyarbakır ve Mardin illerinden gönderildiği ortaya çıkmıştır. Nüfusun daha az olduğu Kilis’te ise nüfusa oranla daha fazla iletinin paylaşıldığı görülmektedir.

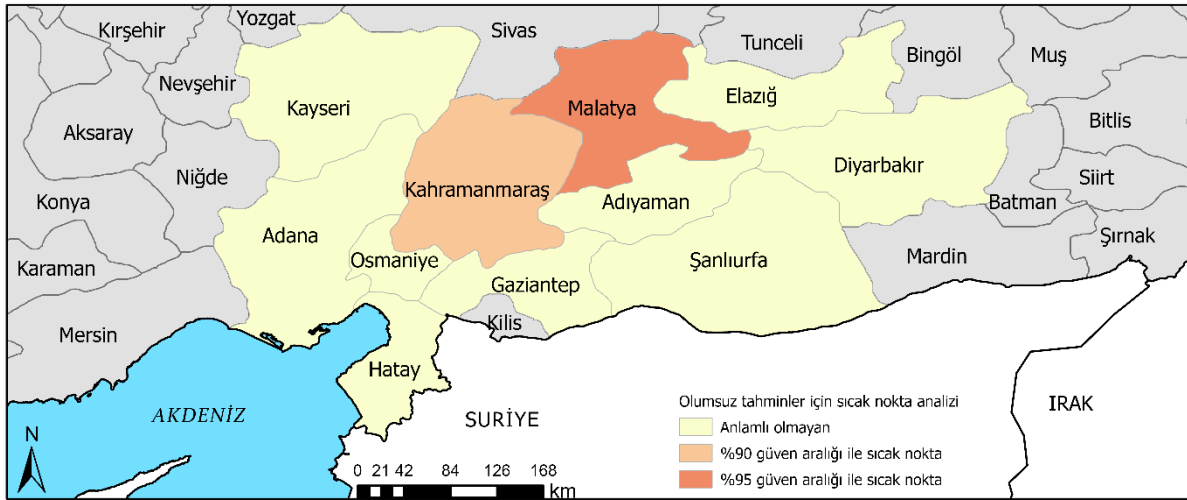


Şekil 4: 6 Şubat 2023 depreminden etkilenen illerin nüfusa oranla X uygulamasındaki depremlerle ilgili ileti sayıları

Referans olumlu-olumsuz ve tahmin olumlu-olumsuz ileti sayıları ait olduğu ilin nüfusuna göre normalize edilmiştir. İllere ait sıcak nokta analizinde kullanılacak olan olumlu-olumsuz ileti sayıları için 17 ilin (Adana, Adıyaman, Batman, Bingöl, Diyarbakır, Elâziğ, Gaziantep, Hatay, Kahramanmaraş, Kayseri, Kilis, Malatya, Mardin, Niğde, Osmaniye, Şanlıurfa ve Tunceli) ileti sayılarının aritmetik ortalaması alınmıştır. Aritmetik ortalama sonucuna göre bir eşik değeri belirlenmiştir. Batman, Bingöl, Kilis, Mardin, Niğde ve Tunceli illerine ait ileti sayılarının ortalama ileti sayısından düşük olmasının verebileceği yanlışlık nedeni ile 6 il sıcak nokta analizinden muaf tutulmuştur. Elde edilen nüfusa göre normalize edilmiş referans ve tahmin olumlu-olumsuz ileti sayıları üzerinden Adana, Adıyaman, Diyarbakır, Elâziğ, Gaziantep, Hatay, Kahramanmaraş, Kayseri, Malatya, Osmaniye ve Şanlıurfa illeri için sabit mesafe bant (fixed distance band) ilişkisi ile sıcak nokta analizi yapılmıştır. Referans ve tahmin olumlu-olumsuz iletilerin sıcak nokta analiz sonucuna göre her iki ileti grubunun da aynı anlamlılık düzeylerinde ($\alpha=0.10$; $\alpha=0.05$; $\alpha=0.01$) aynı kümelenme sonucunu verdiği gözlemlenmiştir. İllere göre duygu analizi modelinin olumsuz tahmin ileti sayılarının sıcak nokta analizi sonucu Şekil 5'te verilmiştir. Şekil 5'te Malatya için 0.05 anlamlılık düzeyinde, Kahramanmaraş için 0.10 anlamlılık düzeyinde diğer illere kıyasla olumsuz iletilerin kümelenmiş (sıcak nokta) olduğu öne çıkmaktadır. Referans ve tahmin ileti grupları üzerindeki sıcak nokta analizlerinin sonucunda aynı anlamlılık düzeyinde sonuç vermesi referans ve duygu analizi modelinin verdiği tahmin sonucu ileti sayılarının arasında anlamlı bir fark olmadığını ortaya çıkarmaktadır. Bu çalışmada, sırasıyla Malatya ve Kahramanmaraş illerinde nüfusa oranla olumsuz ileti sayılarının istatistiki olarak yüksek değer gösterdiği anlaşılmaktadır. Diğer iller nüfusa oranla olumsuz ileti sayısına göre istatistiksel olarak rastlantısal dağılım göstermektedir. Tablo 5'te görülen il bazında ileti sayılarının artması, sıcak nokta analizinde kullanılan z-testinin daha güçlü olmasını sağlayacaktır. Ayrıca, psikososyal destek planlamasında, kitlenin duygu durumu ile birlikte ihtiyaç ve taleplerin mekânsal analizleri de dikkate alınarak planlamaların yapılması gerekmektedir. Bu bağlamda, bu çalışma, psikososyal destek kapsamına alınacak öncelikli illerin belirlenmesine dair bir tespit yapmamakta; doğal dil işleme ve derin öğrenme ile yapılan duygu analizlerinin ve duygu durumları kullanılarak yapılan sıcak nokta analizlerinin CBS tabanlı psikososyal destek planlaması çalışmalarına katkı sağlayabileceğini vurgulamaktadır.

Tablo 5: İl bazında ileti sayıları

İl	İleti Sayısı
Tunceli	4
Kilis	11
Mardin	12
Niğde	12
Bingöl	14
Batman	15
Elâzığ	36
Osmaniye	46
Diyarbakır	49
Hatay	54
Gaziantep	58
Şanlıurfa	58
Adana	59
Kahramanmaraş	59
Kayseri	65
Adıyaman	66
Malatya	67



Şekil 5: 6 Şubat 2023 depreminden etkilen illerden edinilen olumsuz ileti sayıları ile ilgili sıcak nokta analizi (referans ve tahmin haritaları ayrıdır)

Literatürde X platformu üzerinden belli ilgi noktalarından elde edilen veya toplumsal (ünlü birinin bir şehri ziyaret etmesi), siyasal (darbe girişimleri, ülkede yaşanan protestolar vs.) ve doğal olaylar (deprem, sel, yoğun yağış vs.) ile ilgili konular hakkındaki iletilerden yapılan duygu analizi çalışmaları oldukça yaygındır. Bu çalışmanın literatürdeki çalışmalardan farkı, 6 Şubat 2023 depremleri ile ilgili insanların duygu durumlarını ve değişimlerini analiz edilmesidir. Bu çalışmada kullanılan veri kazıma ve duygu analizi metodolojisi özgün olarak kurgulanmış ve diğer afet konularına adapte edilebilir esnek bir algoritma ile ortaya çıkmıştır. Olumlu ve olumsuz olmak üzere veriler iki sınıf olacak şekilde etiketlenmiştir. Ancak, duygu durumlarının (korkma, üzülme, şaşkınlık, mutluluk gibi) çeşitlilik göstermesi nedeniyle verilerin daha fazla sınıfa ayrılarak etiketlenmesi duygu durumlarının daha detaylı tanımlanmasını sağlayabilir. Bu nedenle, modelin eğitiminde kullanılacak

etiketlerin daha fazla olması modelin tahmin doğruluğunu geliştirebilir. Kullanılan veri kaynağının sosyal medya olması ve kitle kaynaklı toplanan verilerin çelişkili veya anlaşılması güç metinleri içerebilmesi duygu analizinde kullanılan makine öğrenimi yöntemlerinin performansını olumsuz yönde etkileyen zorluklardan biridir. Bu çalışmada gözlemlendiği gibi duygu durumu belirsiz ifadeler sınıflandırıcının performansını düşürebilmektedir. Ayrıca sosyal medya verilerinin CBS ortamında analiz edilmesi ile ilgili bir diğer zorluk olarak kullanıcıların konumu kapalı bir biçimde paylaştıkları iletilerin CBS ortamına dahil edilememesi olarak değerlendirilebilir.

4. Sonuç ve Öneriler

Bu çalışmada, X platformu üzerinden Kahramanmaraş depremlerine ait Haziran-Eylül 2023 tarihleri arasındaki 685 adet gönderiye erişilmiştir. Veri kazıma kütüphanelerinden BeautifulSoup ve Selenium kullanılarak otomatik bir şekilde istenen sayıda gönderinin elde edilmesi işlemi gerçekleştirilen bir Python kodu hazırlanmıştır. Hazırlanan derin öğrenme eğitim modeli sayesinde yaşanmış veya yaşanabilecek depremler sonrasında anlık olarak istenilen platform üzerinden veri kazıma ve duygu analizi yapılabilir hale getirilmiştir.

Elde edilen veriler olumlu ve olumsuz anlamlarını taşıyan 1 ve 0 etiketlerine sahip veri seti haline getirilmiştir. Verileri doğal dil ön işleme tekniklerinden geçiren ve duygu analizini yapan hazır bir model geliştirilmiştir. Model, verilen veri setinin eğitimini ortalama %87 doğrulukla tamamlamıştır. Model eğitimini tamamlayıp veri seti üzerinde tahminlerini gerçekleştirdiğinde ise verinin %82.5'ini olumsuz etiketli, %17.5'ini olumlu etiketli olarak sınıflandırmıştır. Veri setinde olumlu etikete sahip olan fakat modelin tahmin sonucunda olumsuz etiket sınıfına eklenen gönderiler incelenmiştir. Bu gönderilerin modelin karıştırabileceği duygu durumu belirsiz ifadeler içerdiği fark edilmiştir. Bu ifadeler örnek olarak bir cümle içerisinde depremin yıkıcı etkilerinden bahsedilirken diğer yandan deprem sonrası yardımlaşma mesajı ile iletilerini bitiren uzun gönderiler verilebilir.

Bu çalışmada, Kahramanmaraş depremleri ile ilgili sosyal medya verilerinin doğal dil işleme ve kapı yinelemeli birimler derin öğrenme modeli kullanılarak duygu analizinin gerçekleştirilebileceği gösterilmiştir. Herhangi bir deprem sonrasında duygu analizi yapabilen bu model sayesinde, afet hakkında duygu analizi güncel olarak gerçekleştirilebilecektir. Ayrıca tahmin verilerinden sıcak nokta analizi ile iletiler arasındaki kümelenme örüntülerinin istatistiki olarak tespit edilebileceği ve bu kümelerin psikososyal planlama aşamasında fikir verebileceği görülmektedir. Gelecek çalışmalarda, sosyal medya üzerinden paylaşılan yeni iletiler de mevcut giriş veri setine eklenerek daha geniş bir sözlük kullanılarak derin öğrenme modelinin eğitimi geliştirilebilir ve sonuçlar iyileştirilebilir. Ayrıca, duygu analizinde daha fazla etiket üretmek (korkma, üzülmeye, şaşkınlık, mutluluk gibi) farklı derin öğrenme eğitim modelleri test edilebilir. Duygu analizinin sonuçları CBS ortamına aktararak duygu durum değişimleri mekân-zaman kapsamında analiz edilebilir.

Yazar Katkısı

Dilan Gözdem Dolu: Literatür taraması, Veri toplama, Analiz ve yorumlama, Yazım. **Alper Şen:** Fikir, Tasarım, Makale değerlendirme, Denetleme, Yazım.

Çıkar Çatışması Beyanı

Yazarlar, bu çalışmada bilinen ilgili herhangi bir finansal veya finansal olmayan çıkar çatışması olmadığını beyan eder.

Kaynaklar

- Adalı, E. (2012). Doğal dil işleme. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 5(2).
- Albayrak, M., Topal, K., & Altıntaş, V. (2017). Sosyal medya üzerinde veri analizi: Twitter. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 22(Kayfor 15 Özel Sayısı), 1991-1998.
- Berrar, D. (2019). Cross-validation. *Encyclopedia of Bioinformatics and Computational Biology*, 1, 542-545.
- Bhardwaj, A., Narayan, Y., & Dutta, M. (2015). Sentiment Analysis for Indian Stock Market Prediction Using Sensex and Nifty. *Procedia Computer Science*, 70, 85-91.
- Camacho, K., Portelli, R., Shortridge, A., & Takahashi, B. (2021). Sentiment Mapping: Point Pattern Analysis of Sentiment Classified Twitter Data. *Cartography and Geographic Information Science*, 48(3), 241-257.
- Chen, L. C., Lee, C. M., & Chen, M. Y. (2020). Exploration of Social Media for Sentiment Analysis Using Deep Learning. *Soft Computing*, 24, 8187-8197.
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. *arXiv preprint arXiv*, 1406-1078.
- Clarke, K.C. (1997). *Getting started with geographic information systems* (25-30). New Jersey: Prentice Hall.
- Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural Language Processing (Almost) from Scratch. *Journal of Machine Learning Research*, 12, 2493-2537.
- Çubukçu, K.M. (2015). *Planlamada ve Coğrafyada Temel İstatistik ve Mekânsal İstatistik*. Ankara: Nobel Akademik Yayıncılık.
- Feng, Y., Huang, X., & Sester, M. (2022). Extraction and Analysis of Natural Disaster Related Vgi from Social Media: Review, Opportunities and Challenges. *International Journal of Geographical Information Science*, 36(7), 1275-1316.
- Gulnerman, A. G., & Karaman, H. (2020). Sosyal Medyanın Gönüllü Coğrafi Veri Olarak Kullanımı ve Sosyal Medya Verilerinden Coğrafya Sözlüğü Üretimi. *Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, 20(2), 276-286.
- Gulnerman, A. G., Karaman, H., Pekaslan, D., & Bilgi, S. (2020a). Citizens' Spatial Footprint on Twitter-Anomaly, Trend and Bias Investigation in Istanbul. *ISPRS International Journal of Geo-Information*, 9(4), 222.
- Gulnerman, A. G., Karaman, H., & Basiri, A. (2020b). New age of crisis management with social media. *Open Source Geospatial Science for Urban Studies: The Value of Open Geospatial Data* (131-160). Berlin:Springer.
- Gündüz, H. (2013). Borsa İstanbul (BIST) 100 Endeksi Yönünün Ekonomi Haberleri ile Tahmin Edilmesi (Yüksek Lisans Tezi). İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, İstanbul.
- Häberle, M., Werner, M., & Zhu, X. X. (2019). Building type classification from social media texts via geo-spatial textmining. *2019 IEEE International Geoscience And Remote Sensing Symposium* (10047-10050). Yokohama, Japonya.
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735-1780.
- İlhan, N., & Sağaltıcı, D. (2020). Twitter'da Duygu Analizi. *Harran Üniversitesi Mühendislik Dergisi*, 5(2), 146-156.
- Kaya, A. (2021). Bir Araştırma Kaynağı Olarak Arşivlenen Sosyal Medya Verilerinin Kullanımı. *Bilgi ve Belge Araştırmaları*, (16), 49-79.
- Kılınç, M., Aydın, C., & Tarhan, Ç. (2022). Türkiye'de Sosyal ve Dijital Girişimcilik: Veri Kazıma Teknikleriyle Kitle Fonlaması Platformlarının İçerik Analizi. *Acta Infologica*, 6(1), 83-97.
- Koç, M., & Yalçın, S. (2023). Afetlerde Krize Müdahale: Kahramanmaraş Depremi'nde Aile ve Sosyal Hizmetler Bakanlığı'nın Çalışmaları. *Uluslararası Sosyal Hizmet Araştırmaları Dergisi*, 3(2), 93-105.
- Küçük, D., & Arıcı, N. (2018). Doğal Dil İşlemede Derin Öğrenme Uygulamaları Üzerine Bir Literatür Çalışması. *Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi*, 2(2), 76-86.
- Markoulidakis, I., Kopsiaftis, G., Rallis, I., & Georgoulas, I. (2021). Multi-class confusion matrix reduction method and its application on net promoter score classification problem. *In The 14th Pervasive Technologies Related to Assistive Environments Conference* (412-419). Korfu, Yunanistan.
- Neal, D.M., (1997). Reconsidering the Phases of Disasters. *International Journal of Mass Emergencies and Disasters*, 15 (2), 239-264.
- Özgür E. M. & Aydın, O. (2011). Türkiye'de Evlilik Göçünün Mekânsal Veri Analizi Teknikleriyle Değerlendirilmesi. *Coğrafi Bilimler Dergisi*, 9(1), 29-40.
- Sachin, S., Tripathi, A., Mahajan, N., Aggarwal, S., & Nagrath, P. (2020). Sentiment analysis using gated recurrent neural networks. *SN Computer Science*, 1, 1-13.

- Şeker, S. E. (2015). Metin Madenciliği (Text Mining). *YBS Ansiklopedi*, 2(3), 30-32.
- Wang, R. Q., Mao, H., Wang, Y., Rae, C., & Shaw, W. (2018). Hyper-Resolution Monitoring of Urban Flooding with Social Media and Crowdsourcing Data. *Computers & Geosciences*, 111, 139-147.
- Williams, R. J., & Zipser, D. (1989). A Learning Algorithm for Continually Running Fully Recurrent Neural Networks. *Neural Computation*, 1(2), 270-280.
- Zhao, B. (2017). Web Scraping. *Encyclopedia of Big Data*. Cham: Springer.
- Zhou, X., & Xu, C. (2017). Tracing the Spatial-Temporal Evolution of Events Based on Social Media Data. *ISPRS International Journal of Geo-Information*, 6(3), 88.
- URL-1: Afet ve Acil Durum Müdahale Hizmetleri Yönetmeliği (2013), Resmî Gazete. <https://www.resmigazete.gov.tr/eskiler/2013/12/20131218-13-1.pdf> (Erişim Tarihi: 02 Ocak 2024).
- URL-2: AFAD Türkiye Afet Müdahale Eylem Planı (TAMP) (2014). https://www.afad.gov.tr/kurumlar/afad.gov.tr/e_Kutuphane/Planlar/TAMP.pdf (Erişim Tarihi: 02 Ocak 2024).
- URL-3: Aile ve Toplum Hizmetleri Genel Müdürlüğü (2023), Psikososyal Destek Rehberi. <https://www.aile.gov.tr/uploads/athgm/uploads/pages/goc-afet-ve-acil-durumlarda-psikososyal-destek/psi-kososyal-destek-nedi-r.pdf> (Erişim Tarihi: 02 Ocak 2024).
- URL-4: AFAD (2023), Basın Bülteni. 36. <https://www.afad.gov.tr/kahramanmarasta-meydana-gelen-depremler-hk-36> (Erişim Tarihi: 02 Ocak 2024).
- URL-5: Esri (2024), How Hot Spot Analysis (Getis-Ord Gi*) Works. <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm> (Erişim Tarihi: 13 Şubat 2024).
- URL-6: Python (2021) Real Python. Beautiful Soup: Build a Web Scraper With Python – Real Python (Erişim Tarihi: 02 Ocak 2024).
- URL-7: Türkiye İstatistik Kurumu. <https://data.tuik.gov.tr/Kategori/GetKategori?p=Nufus-ve-Demografi-109> (Erişim Tarihi: 02 Ocak 2024).