

ANKARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

ÇOK DİLLİ DUYGU ANALİZİ

Dilek DEMİR

İSTATİSTİK ANABİLİM DALI

ANKARA
2024

Her hakkı saklıdır

ÖZET

Yüksek Lisans Tezi

ÇOK DİLLİ DUYGU ANALİZİ

Dilek DEMİR

Ankara Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Prof. Dr. Olçay ARSLAN

Duygu analizi, makine öğrenmesi ve doğal dil işleme tekniklerini kullanarak metinlerdeki duyguları ortaya çıkarmayı amaçlayan bir çalışma alanıdır. Sosyal medya paylaşımları, ürün ve hizmet yorumları, otel ve film yorumları, haber metinleri ve kitaplar duygu analizi için başlıca veri kaynaklarıdır. Sosyal medya başta olmak üzere sosyal platformlarda kullanıcı ve paylaşım sayıları hızla artmaktadır. Sosyal platformlara dünyanın her yerinden erişim imkanı olması sonucu ortaya çok dilli veri yığınları çıkmaktadır. Ürün, hizmet veya kişiler hakkında ifade edilen görüşleri analiz edip karar süreçlerine dahil etmek önemli avantajlar sağlamaktadır ancak kaynaklar ve geliştirilen yöntemler her dil için yeterli düzeyde bulunmamaktadır. Bununla başa çıkmak için çeşitli çok dilli duygu analizi sistemleri geliştirilmiştir. Bu tez çalışmasında çok dilli duygu analizi yapmak amacıyla sağlık, turizm, pazarlama ve eğitim alanlarında İngilizce, Almanca, Felemenkçe dillerindeki tüketici görüşlerini içeren MultiEmo veri seti kullanılmıştır. MultiEmo çok dilli duygu analizi görevleri için oluşturulmuş bir veri setidir. Tüketici görüşleri Türkçe diline çevrilerek veri setine dahil edilmiştir. Cümle düzeyinde temsil yapan ve önceden eğitilmiş bir derin öğrenme modeli olan çok dilli evrensel cümle kodlayıcı (MUSE) kullanılarak veri seti vektörler ile temsil edilmiştir. Daha sonra kosinüs benzerliği kullanılarak cümleler pozitif veya negatif olarak sınıflandırılmıştır. Bu yaklaşım ile her dil ve her domain için ayrı ayrı model geliştirmek yerine derin öğrenme yaklaşımları kullanılarak başarılı sonuçlar elde edildiği görülmüştür. Yapılan sınıflandırma sonucunda İngilizce dilinde 0.75, Türkçe dilinde 0.72 doğruluk oranı elde edilmiştir.

Ocak 2024, 50 sayfa

Anahtar Kelimeler: Çok dilli duygu analizi, makine öğrenmesi, derin öğrenme, çok dilli evrensel cümle kodlayıcı

ABSTRACT

Master's Thesis

MULTILINGUAL SENTIMENT ANALYSIS

Dilek DEMİR

Ankara University
Graduate School of Natural and Applied Science
Department of Statistics

Advisor: Prof. Dr. Olçay ARSLAN

Sentiment analysis is a field of study that aims to reveal emotions in texts using machine learning and natural language processing techniques. Social media posts, product and service reviews, hotel and movie reviews, news texts, and books are the main data sources for sentiment analysis. The number of users and shares is rapidly increasing on social platforms, especially social media. Multilingual data stacks emerge as a result of access to social platforms from all over the world. Analyzing the opinions expressed about products, services, or people and including them in decision-making processes provides significant advantages, but the resources and developed methods are not sufficient for every language. Various multilingual sentiment analysis systems have been developed to deal with this. In this thesis, the MultiEmo dataset, which contains consumer opinions in English, German and Dutch in the fields of health, tourism, marketing and education, was used to conduct multilingual sentiment analysis. MultiEmo is a dataset created for multilingual sentiment analysis tasks. Consumer opinions were translated into Turkish and included in the data set. The data set was represented with vectors using the multilingual universal sentence encoder (MUSE), which is a pre-trained deep learning model that represents at the sentence level. Sentences were then classified as positive or negative using cosine similarity. With this approach, it has been observed that successful results have been achieved by using deep learning approaches instead of developing separate models for each language and each domain. As a result of the classification, an accuracy rate of 0.75 in English and 0.72 in Turkish was obtained.

January 2024, 50 pages

Key Words: Multilingual sentiment analysis, machine learning, deep learning, Multilingual Universal Sentence Encoder

ÖNSÖZ ve TEŞEKKÜR

Tez çalışmam boyunca bilgi birikimi, önerileri ve yardımlarıyla beni yönlendiren, her zaman yol gösterici ve destekleyici olan, kendisinden değerli bilgiler edindiğim kıymetli hocam Sayın Prof. Dr. Olçay Arslan'a teşekkürlerimi, saygılarımı ve sevgilerimi sunuyorum.

Her zaman yanımda olan, beni destekleyen ve bugünlere getiren canım annem ve canım babama teşekkürlerimi ve en derin sevgilerimi sunuyorum. Her zaman örnek aldığım, fikirleri ve bilgileriyle beni aydınlatan canım abime de teşekkürlerimi ve sevgilerimi sunuyorum. Canım kuzenim Nil Buse'yi de çok seviyorum ve çok öpüyorum. İyi ki varsınız ve iyi ki benim ailemsiniz.

Dilek DEMİR

Ankara, Ocak 2024

İÇİNDEKİLER

TEZ ONAY SAYFASI	
ETİK	i
ÖZET	ii
ABSTRACT	iii
ÖNSÖZ VE TEŞEKKÜR.....	iv
KISALTMALAR DİZİNİ	vi
ŞEKİLLER DİZİNİ	vii
ÇİZELGELER DİZİNİ	viii
1. GİRİŞ	1
2. TEMEL KAVRAMLAR	7
2.1 Veri	7
2.2 Büyük Veri.....	10
2.3 Yapay Zeka	10
2.4 Doğal Dil İşleme	10
2.5 Makine Öğrenmesi	11
2.6 Yapay Sinir Ağları	12
2.7 Derin Öğrenme	14
2.8 Veri Bilimi.....	15
2.9 Metin Madenciliği	16
3. DUYGU ANALİZİ.....	17
3.1 Duygu	17
3.2 Duygu Analizi Nedir?	17
3.2.1 Duygu analizi düzeyleri	19
3.3 Duygu Analizi Kullanım Alanları.....	20
3.4 Duygu Analizinde Karşılaşılan Zorluklar	22
4. ÇOK DİLLİ DUYGU ANALİZİ UYGULAMASI.....	23
4.1 Çok Dilli Duygu Analizi Nedir?	23
4.2 Veri Seti Belirleme	25
4.3 Verilerin Çevrilmesi.....	26
4.4 Ön İşleme	27
4.4.1 Tokenizasyon	27
4.4.2 Durak kelimeleri temizleme	28
4.4.3 Kök bulm	28
4.4.4 N-gram	29
4.4.5 Kelime Çantası (Bag of Words)	30
4.5 Çok Dilli Evrensel Cümle Kodlayıcı (MUSE)	32
4.6 Değerlendirme	38
4.6.1 Doğruluk (Accuracy)	38
4.6.2 Kesinlik (Precision)	38
4.6.3 Duyarlılık (Recall).....	38
4.6.4 F-1 Ölçüsü	39
4.6.5 Dengeli Doğruluk	39
5. SONUÇ VE ÖNERİLER.....	41
KAYNAKLAR	43
ÖZGEÇMİŞ.....	50

KISALTMALAR DİZİNİ

MUSE	Multilingual Universal Sentence Encoder
USE	Universal Sentence Encoder
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
TF-IDF	Term Frequency- Inverse Document Frequency



ŞEKİLLER DİZİNİ

Şekil 1.1 Maslow'un ihtiyaçları hiyerarşisi	1
Şekil 1.2 Sosyal ihtiyaçlar ve karşılanabileceği platformlar	2
Şekil 2.1 İnternet kullanımına dair istatistikler	8
Şekil 2.2 Bilgi keşfi yöntemleri	9
Şekil 2.3 Yapay sinir ağı katmanlar	13
Şekil 2.4 Derin öğrenme ve makine öğrenmesi için veri miktarı ve doğruluk ilişkisi....	14
Şekil 2.5 Ürün yorumunun konulara göre bölünmesi	15
Şekil 2.6 Kotu ve Deshpande (2019) tarafından görselleştirilen yapay zeka ve diğer kavramlar arası ilişkiler.....	16
Şekil 3.1 Duygu analizi seviyeleri	19
Şekil 4.1 İnternette en fazla kullanılan diller	23
Şekil 4.2 Makine çevirisine dayalı duygu analizi örneği	24
Şekil 4.3 En sık kullanılan kelimeler	27
Şekil 4.4 Ön işleme öncesi veri seti	28
Şekil 4.5 Ön işleme sonrası veri seti	29
Şekil 4.6 Farklı n değerleri için n-gram örnekleri.....	29
Şekil 4.7 Kelime çantası yöntemi ile sık kullanılan kelimelerin, kelime bulutu ile görselleştirilmesi	31
Şekil 4.8 Kelime çantası yöntemi ile belirlenen sık kullanılan kelimeler.....	31
Şekil 4.9 Kelime temsili.....	32
Şekil 4.10 Cümle temsili.....	33
Şekil 4.11 Çok dilli cümle temsili.....	33
Şekil 4.12 Vektörlerin 2 ve 3 boyut ile temsilleri	34
Şekil 4.13 Sınıflandırma örneği	35
Şekil 4.14 Kosinüs benzerliği	35
Şekil 4.15 4 dil için hata matrisleri	36
Şekil 4.16 Hata matrisi.....	37
Şekil 4.17 Tip-1 ve Tip-2 hata örnekleri.....	37

ÇİZELGELER DİZİNİ

Çizelge 2.1 Yapılandırılmamış veri örneği	7
Çizelge 2.2 Yapılandırılmış veri örneği	8
Çizelge 2.3 Yapılandırılmamış metnin yapılandırılması örneği	8
Çizelge 4.1 Bigram ve trigram çıktıları.....	30
Çizelge 4.2 Farklı referans cümleler için sınıflandırma sonuçları	40
Çizelge 4.3 Model değerlendirme metrikleri karşılaştırması	40



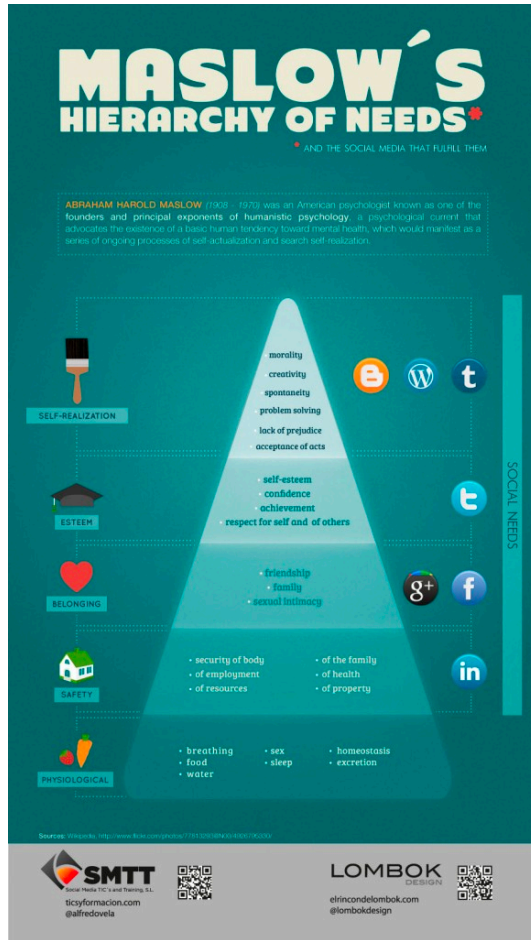
1. GİRİŞ

Şekil 1.1’de Maslow’un ihtiyaçlar hiyerarşisi verilmiştir. Buna göre yaşamımızda doyurulması gereken ihtiyaçlar vardır ve bunlar öncelik sırasına göre sıralanmıştır. Yüksek düzeydeki ihtiyaçları karşılamadan önce alt düzeydeki ihtiyaçların doyuma kavuşturulması gerekmektedir. Kişilerin ilgisi karşılanmayan gereksinimlerindedir, doyuma ulaştıktan sonra da ilgi odağı olmaktan çıkmaktadır. Fiziksel ihtiyaçlar hiyerarşinin ilk basamağında yer almaktadır. Buna göre beslenme, uyku ve diğer temel gereksinimlerin karşılanması gerekmektedir. Bir sonraki adımda güvenlik ihtiyacı yer almaktadır; emniyet içinde olmayı, korkudan ve kargaşadan kurtulma ihtiyacını içermektedir. Ait olma ve sevgi ihtiyacı basamağında arkadaşlık kurmak, yakın ilişkiler içinde olmak ve samimiyet duygusuna duyulan gereksinim ön plana çıkmaktadır. Süreç burada sona ermemekte yeni ihtiyaçlar kendini göstermektedir. Değer ihtiyacı adı verilen yeni basamakta kendini yeterli bulma, beğenilme gereksinimi ön plana çıkmaktadır. Buraya kadar her ihtiyacı doyurmuş bir birey bu kez de kendini gerçekleştirme ihtiyacı içinde olup potansiyelini geliştirmek istemektedir (Burger 2016).



Şekil 1.1 Maslow’un ihtiyaçlar hiyerarşisi (Gürbüz 2020)

Şekil 1.2’de gösterildiği üzere teknolojik gelişmeler temel ve sosyal ihtiyaçları karşılamamanın yeni yollarını da beraberinde getirmiştir. Basamaklarda yer alan ihtiyaçları karşılamaya yönelik farklı sosyal mecralar bulunmaktadır ve giderek çeşitlenmektedir. Sosyal ağlar arasında sosyal medya kullanım süresi, paylaşım ve etkileşim sayıları ile öne çıkmaktadır. Sosyal ağların yaygınlaşması ile alışveriş, iletişim gibi pek işlem dijitalleşmiş; bu işlemlerin istenilen yerde ve zamanda yapılması mümkün hale gelmiştir. Bu ağlara erişimin ve kullanımının kolay olması da eklendiğinde yaşamımızın bir parçası olması kaçınılmaz hale gelmiştir.



Şekil 1.2 Sosyal ihtiyaçlar ve karşılanabileceği platformlar (İnce 2012)

Sosyal ağların artan popülerliği ile finans, sağlık, pazarlama, politika ve diğer pek çok alanda veri yığınları ile karşı karşıya kalınmıştır. Ürün değerlendirmeleri, haber metinleri, sosyal medya paylaşımları, müzik albümü ve film yorumları veri yığınlarını oluşturmaktadır. Kullanıcıların etkileşim içerisinde olduğu bu paylaşımlar diğerlerinin tercihlerine etki etmekte, duygu ve davranışlarına yön verebilmektedir. Diğer yandan sektör yetkilileri ürün ve hizmetlerin geliştirilmesi için müşterilerin görüşlerine ihtiyaç duymaktadır. Bu nedenle bu veri yığınlarından anlamlı bilgiler çıkarmak ticari bir ihtiyaç haline gelmiştir (Sagnika vd. 2020).

Duygu analizi belirli kurum, kişi, hizmet, olay veya konu hakkında ifade edilen görüşleri analiz etmek, ortaya çıkarmak için makine öğrenmesi, doğal dil işleme, metin madenciliği tekniklerini kullanmayı içeren yöntemler bütünüdür. Çeşitli algoritma ve teknikler kullanılarak kitlelerin tutum ve görüşlerini belirlemek amaçlanmaktadır.

Geçmişte tüketicilerin ürün ve hizmetler hakkında görüşlerini almak için anket ve röportaj gibi yöntemler kullanılıyordu. Günümüzde bu görüş alışverişinin yerini bloglar ve web siteleri almıştır. Bu, hem hizmet sağlayıcıların müşterilerle iletişim kurma gereksinimini ortadan kaldırmış hem de müşterilerin görüşlerini istedikleri zaman paylaşma olanağı sağlamıştır (Kou vd. 2016).

Duygu analizi alanında çok sayıda çalışma yapılmıştır. Büyükeke vd. (2020) tarafından yapılan çalışmada turizm alanında kullanıcı yorumlarını içeren bir platform olan TripAdvisor'dan toplanan veriler analiz edilmiştir. Antalya ilindeki turizm işletmelerine yapılan değerlendirmelerin duygu yoğunlukları bölgesel farklılıklar göz önüne alınarak incelenmiştir. Yapılan yorumlar “yapılacak şeyler”, “deneyim” ve “eğlence” gibi çeşitli kategorilere ayrılmıştır. Daha sonra hangi bölgenin hangi yönüyle olumlu veya olumsuz değerlendirildiği belirlenmiştir. Örneğin butik otellerin veya pansiyonların yoğunlukta olduğu bölgelerde en fazla “yapılacak şeyler” konusunda yorum yapıldığı tespit edilmiştir ve bahsedilen çevre için öneriler sunulmuştur. Çalışmanın sonuç ve öneriler

kısımında, sosyal medya verileri; anket, söyleşi gibi yöntemlere kıyasla büyük miktarda ve ekonomik olması yönleriyle üstün görülmüştür. Ayrıca kullanıcılar görüş ve düşüncelerini kendi istekleriyle paylaştıkları için daha objektif oldukları şeklinde değerlendirilmiştir.

Sosyal medyada herhangi bir ürün, kişi veya konu hakkında duygu ve düşünceleri paylaşmak bir kültür haline gelmiştir (Bhargava ve Sharma 2017). Dünyanın her yerinden erişilebilen Twitter, Facebook ve YouTube gibi platformlar kişilerin bilmedikleri dillerde dahi, çeviri araçlarının yardımı ile, içerikleri izleme, yorumlama ve tartışma fırsatı vermiştir. Bu yaygın kullanım, çeşitli alanlarda çok dilli ve çok kültürlü veri yığınlarının oluşmasına neden olmuştur (Becker vd. 2017). Farklı dillerde yapılan yorumlar, farklı kültürlerde mal veya hizmetlerin nasıl algılandığına ve nasıl kullanıldığına dair bilgi sağlamaktadır. Böylece farklı kültürler için ürün veya hizmet geliştirilmesi sürecinde kıymetli bilgiler elde edilmektedir (TransPerfect 2020). Bu zengin içerik nedeniyle, başta Twitter olmak üzere sosyal medya platformları duygu analizi için sıklıkla kullanılmaktadır (Pak ve Paroubek 2010).

Çok dilli duygu analizi farklı dillerdeki metinlerin duygusal içeriğini incelemek için kullanılmaktadır. Duygu analizi özellikle İngilizce gibi yaygın kullanım alanına sahip ve yüksek kaynaklı dillerde önemli gelişmeler gösterse de (Balahur ve Turchi 2012) yetersiz kaynaklara sahip diller için aynı gelişme düzeyinden bahsedilememektedir. Yetersiz kaynaklara sahip diller, dijital dil teknolojileri geliştirmek için yeterli kaynaklara sahip olmayan dilleri ifade etmektedir (Mabokela vd. 2023). İhtiyaç duyulan bu kaynaklar etiketli külliyat, veri seti veya duygu sözlükleri olabilmektedir (Agüero-Torales vd. 2021). Yetersiz kaynaklı bir dilde duygu analizi yapmak için üç yöntem önerilmiştir. İlk yöntemde yetersiz kaynaklı dildeki veri makine çevirisi ile İngilizceye çevrilip İngilizce dilinde duygu sınıflandırılması yapılmaktadır. (Kim ve Hovy 2006). İkinci yöntemde İngilizce külliyat makine çevirisi ile hedef dil veya dillere çevrilmekte ve model çevrilmiş veri ile eğitilmektedir (Mboutayeb 2021). Son yöntemde İngilizce

duygu sözlüğü hedef dile çevrilmekte ve sözlük tabanlı duygu sınıflandırması yapılmaktadır (Kim ve Hovy 2006).

Fuady ve Ibrahim (2019) tarafından, bir afet döneminde yapılan sosyal medya paylaşımlarından, halkın verdiği tepkiyi anlamak amacıyla bir çalışma yürütmüştür. Çok kültürlü bir yapıya sahip Malezya’da insanlar İngilizce ve Malayca’yı kullanarak paylaşım yapmaktadır. Bu nedenle, her iki dilde afet ile ilgili sosyal medya verileri toplanmış, sözlük tabanlı ve derin öğrenme yaklaşımları ile sınıflandırılmıştır. Sonuçlar derin öğrenme modelinin sözlük tabanlı yaklaşıma kıyasla metnin polaritesini tahmin etmede daha etkili performans sergilediğini göstermiştir. Sınıflandırma sonucunda metinlerde %67.9 nötral, %27.1 olumlu ve %5 olumsuz duygunun yer aldığı belirlenmiştir. Büyük bölümde tarafsız duygular yer alması afet döneminde sosyal medyanın, olumlu ve olumsuz duyguları yaymak için değil, bilgi paylaşımı için kullanıldığı şeklinde yorumlanmıştır.

Mboutayeb vd. (2021) tarafından derin öğrenme yaklaşımli çok dilli duygu analizi yapılmıştır. Metin ve duygu analizinin dile olan bağılıkları analizlerde zorluklara sebep olmaktadır. Algoritma ve kütüphanelerin İngilizce dilinde yoğunlaştığı görülmektedir. Her dil için model geliştirmek zaman alıcı ve zahmetli olduğu gibi her dil için yeterli bir veri setine ulaşmak mümkün olmayabilmektedir. Bu nedenlerle İngilizce dilinde bir duygu analizi modeli eğitilmiş; Arapça, Portekizce ve Fransızca dillerindeki veri setleri İngilizceye çevirilerek duygu tahmini yapılmıştır. CNN (convolutional neural network/ evrişimli sinir ağları), RNN (recurrent neural network/ tekrarlayan sinir ağları) ve CNN-RNN modellerinin kullanıldığı çalışmada CNN modelinin, %86 doğruluk ve %85 F1 skoru ile en iyi performansı gösterdiği belirlenmiştir. Daha sonra 5 farklı dilden oluşan bağımsız bir veri seti, eğitilen CNN modeli üzerinde test edilmiştir. Sonuçta İngilizce, Fransızca, İspanyolca ve Portekizce dillerinde %90’ın üzerinde doğruluk değerleri elde edilmiştir.

Bu tez çalışmanın ikinci bölümünde duygu analizini açıklamada gerekli olacak temel kavramlar kısaca tanıtılmıştır. Bu kavramlar veri, büyük veri, yapay zeka, makine öğrenmesi, doğal dil işleme, metin madenciliği ve diğer bazı tanımları içermektedir. Üçüncü bölümde duygu analizinin ne olduğu kullanım alanları, önemi ve analiz sürecinde karşılaşılan zorluklar açıklanmıştır. Son bölümde çok dilli duygu analizi hakkında açıklamalar yapıldıktan sonra çok dilli duygu analizi uygulaması sunulmuştur. Uygulama kapsamında çok dilli duygu analizi görevleri için oluşturulmuş MultiEmo veri seti kullanılmıştır. Sağlık, turizm, pazarlama ve eğitim alanlarında İngilizce, Almanca, Felemenkçe ve Türkçe dillerindeki kullanıcı görüşleri Çok Dilli Evrensel Cümle Kodlayıcı (Multilingual Universal Sentence Encoder/ MUSE) kullanılarak vektörler ile temsil edilmiştir. MUSE cümle düzeyinde temsil yapan önceden eğitilmiş bir derin öğrenme modelidir. Kosinüs benzerliği skorları kullanılarak veri seti pozitif ve negatif olarak sınıflandırılmıştır. Yapılan sınıflandırma sonucunda İngilizce dilinde 0.75 doğruluk, Türkçe dilinde 0.72 doğruluk skorları elde edilmiştir.

Türkçe dilinde çok dilli duygu sınıflandırması çalışmaları sınırlı sayıda olduğundan alana katkı sağlamak hedeflenmiştir. Bunun yanında her dil için ve her domain için ayrı ayrı model geliştirmek yerine derin öğrenme modellerinden yararlanarak başarılı sonuçlar elde edildiği görülmüştür.

2. TEMEL KAVRAMLAR

2.1 Veri

Veri; bilginin taşıyıcısı ve araştırmaların temel unsurudur. Araştırmacılar verilerin işlenmesi yoluyla yeni bilgiler elde etmektedir (Zülfikar 2022). Veriler ham haldedir ve işlenmeden tek başına bir anlam ifade etmemektedir (Durna ve Demirel 2008).

Veriler yapılandırılmış ve yapılandırılmamış olmak üzere ikiye ayrılmaktadır. Yapılandırılmamış veriler standart olmayan yapıdaki verileri ifade etmektedir. Kitap, makale gibi metin türünde olabileceği gibi, fotoğraf video gibi görsel materyaller de olabilmektedir. Verilerin büyük kısmı yapılandırılmamış formdadır. Karmaşık ve büyük hacimli oldukları için analizleri daha zor ve maliyetlidir (Çağlayan Akay 2020). Çizelge 2.1’de yapılandırılmamış veri örneği paylaşılmıştır.

Yapılandırılmış veriler satır ve sütunlarla düzenlenmiş, standart forma sahip verileri ifade etmektedir. Bu verilerin düzenlenmesi ve analiz edilmesi kolaydır. Bu verilerde satırlar gözlemleri, sütunlar ölçülebilen özellikleri temsil etmektedir. İstatistik ve veri madenciliği yapılandırılmış veriler ile ilgilenmektedir (Çağlayan Akay 2020). Çizelge 2.2’de yapılandırılmış veri örneği paylaşılmıştır. Çizelge 2.3’te ise yapılandırılmamış bir verinin yapılandırılması örneği verilmiştir.

Çizelge 2.1 Yapılandırılmamış Veri Örneği (Atan 2020)

Donör	Hikaye
A	“Donör uyuşturucu, alkol ve sigara kullanmıyor “
B	“Eroin ve sigara kullanmış”
C	“Esrar kullanıyor, cigarayı bırakmış”

Çizelge 2.2 Yapılandırılmış Veri Örneği (Atan 2020)

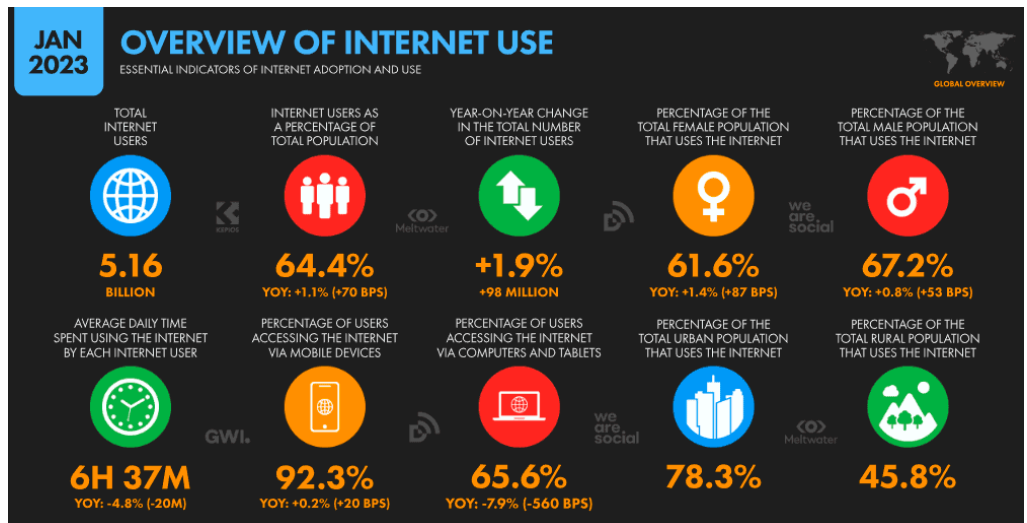
Donör	Sigara	Alkol	Uyuşturucu
A	1	0	0
B	1	1	NA
C	1	1	0

Çizelge 2.3 Yapılandırılmamış Metnin Yapılandırılması Örneği (Uzun 2021)

	Cümleler	Kişi	Gün	Saat	Yer
1	“Burak aradı, çarşamba günü öğleden sonra 3’te buluşmak istediğini söyledi.”	Burak	Çarşamba	15:00	Park
2	“Cuma sabahı Emre’nin ofisinde saat 9’daki toplantıyı unutmayın.”	Emre	Cuma	09:00	Emre’nin Ofisi
3	“Restoranda cumartesi günü akşam 7’de Alper ile buluşacağım.”	Alper	Cumartesi	19:00	Restoran

Metinlerin hem yazılımlar hem insanlar tarafından daha kolay anlaşılmasını sağlamak için satır ve sütun formatında düzenlenmesi yapılandırma olarak tanımlanmaktadır (Uzun 2021).

Şekil 2.1’de internet kullanıma dair istatistikler verilmiştir. Toplam internet kullanıcısı, ortalama internette geçirilen zaman gibi bilgiler elde edilmektedir.

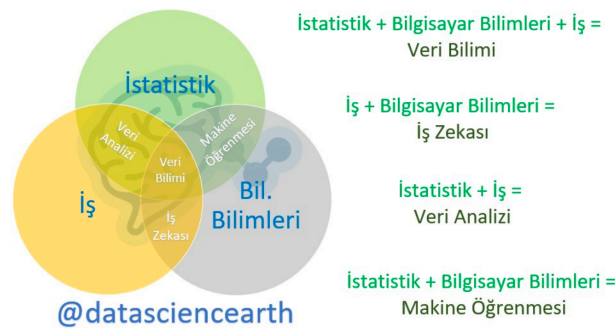


Şekil 2.1 İnternet kullanımına dair istatistikler (We Are Social 2023)

İstatistiklerle ortaya konduğu üzere pek çok kişi aktif olarak sosyal mecralarda vakit geçirmektedir. Sosyal mecraların yoğun kullanımıyla beraber çok sayıda yapılandırılmamış veri ortaya çıkmaktadır. Bu nedenle sosyal medya son yıllarda önemli bir veri kaynağı haline gelmiş, bu verileri analiz etmek ve anlam çıkarmak ihtiyaç haline gelmiştir (Erkırtay ve Ünal 2021).

Geçmişte diğerlerinin görüşlerine başvurmak gerektiğinde anketler, kamuoyu yoklamaları gerçekleştirilmekteydi. Uluslararası ilişkilerde bilgi edinmek amacıyla gazete, radyo gibi geleneksel medya araçlarını takip etmek, casuslar göndermek gibi yöntemler tercih edilirdi. Günümüzde geçmişteki görüş alışverişinin yerini çeşitli bloglar ve web siteleri almıştır. Böylece dünyanın her yerinden bilgi alma ve görüş bildirebilme imkanı oluşmuştur. Paylaşılan bir düşünce hızla yayılıp diğerlerinin tercihlerine etki edebilmektedir (Lei ve Liu 2021).

Verileri kullanarak bilinmeyen olaylar hakkında ve geleceğe yönelik tahminler elde etmek amacıyla algoritmalar geliştirmek için istatistik, matematik ve bilgisayar bilimlerinden yararlanılmaktadır ve aralarındaki ilişki Şekil 2.2’de verilmiştir. Makine öğrenmesi ve veri analizinin temeli istatistik bilimidir (Aksu 2020).



Şekil 2.2 Bilgi keşfi yöntemleri (Aksu 2020)

2.2 Büyük Veri

Büyük veri, büyük miktarda, çok çeşitli ve frekansı yüksek veriler için kullanılan bir terimdir (Çağlayan Akay 2020). Geleneksel veri işleme teknolojileri bu tür verileri işlemek için yetersiz kalmaktadır (Anonim 2022). Veri sayısının çok fazla olması, kapsayıcılık açısından olumlu bir durum olarak görülebileceği gibi zaman ve donanım açısından zorluklara neden olabilmektedir (Zizak vd. 2020).

2.3 Yapay Zeka

Yapay zeka terimi 1956 yılında ortaya çıkmıştır. Yapay zeka, makinelerin zekaya sahip olmasına denmektedir ve canlıların öğrenme süreci taklit edilerek gerçekleştirilmektedir (Altunkaynak 2022). “Yapay zeka, insana özgü olduğu varsayılan akıl yürütme, anlam çıkarma, genelleme ve geçmiş deneyimlerden öğrenme gibi yüksek zihinsel süreçlere ilişkin görevleri yerine getirme yeteneğidir” (Nabiyev 2021). Yapay zeka günlük hayatımızda giderek daha fazla yer edinmektedir. Sosyal medya tahminleri, çeviri araçları, sanal asistanlar örnek olarak gösterilmektedir. Birçok sektörde kullanımı yaygınlaşmaktadır. Örneğin sağlık sektöründe bir hastalığın tanısı, teşhisi, tedavisinde kullanılabildiği gibi robotik cerrahi ile operasyon sürecinde de yararlanılmaktadır (Anonymous 2022).

2.4 Doğal Dil İşleme

Doğal dil işleme doğal insan dilini işlemek, analiz etmek amacıyla bilgisayarların programlanmasını, yazılım sistemlerinin geliştirilmesini içeren bir çalışma alanıdır (Draskovic vd. 2022). Bu çalışma alanında doğal dillerin işlev ve yapısının anlaşılması, bilgisayarlar ve insanlar arası iletişimin kolaylaştırılması ve bilgisayarlar ile dil çevirisi

yapılması amaçlanmaktadır (Nabiyev 2021). Ancak doğal dilin karmaşık yapısı süreçte zorluklara yol açmaktadır. Argo, iğneleme, deyimler ve imla hataları analiz sürecini karmaşık hale getirmektedir (Oğuzlar ve Kızılkaya 2019).

Doğal dil işleme hem metin hem konuşma verilerine tüm dillerde uygulanmaktadır. Spam e-mail filtresi, otomatik metin/konuşma çevirisi, duygu analizi doğal dil işlemenin kullanıldığı araçlardan bazılarıdır (Anonymous 2023f).

Doğal dil işleme alanında en fazla çalışma İngilizce dilinde yapılmıştır. İngilizce dilindeki doğal dil işleme kaynaklarına ulaşım kolay ve çeşitli olduğundan bu dille yapılan araştırmalar artmaktadır (El-Haj vd. 2008).

2.5 Makine Öğrenmesi

Makine öğrenmesi, “insanların öğrenme şekillerini taklit etmek için veri ve algoritmaların kullanımını sağlayan ve doğruluğunu kademeli olarak artıran bir yapay zeka ve bilgisayar bilimi dalıdır” şeklinde tanımlanmaktadır (Anonymous 2023e). Makine öğrenmesi veriden öğrenen algoritmalar ve çeşitli türlerde veriler ile işlem yapabilmektedir. Makine öğrenmesi bilgisayarların verideki karmaşık örüntüleri keşfetme ve veriye dayanarak rasyonel karara varma görevlerini yerine getirmesini sağlamak için çalışılan bir alandır (Anonymous 2023h). Shalev-Shwartz ve Ben David (2014) tarafından yapılan tanımlamaya göre makine öğrenmesi algoritmaları verilen örnek olayların meydana gelme sürecini öğrenir ve genelleme yapar. Böylece olayın meydana gelme sürecini öğrenen algoritmalar olay meydana gelmeden önce veya yaklaşan bir olay için tahmin yapabilmektedir (Keskin 2018).

Öğrenme Çeşitleri:

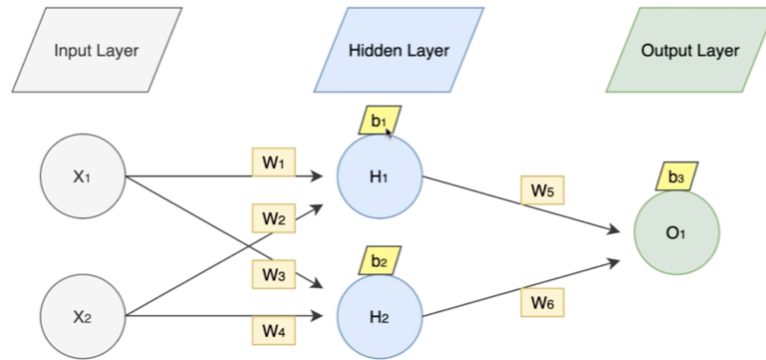
- Denetimli (Gözetimli/Kontrollü) Öğrenme: Mohri vd. (2012) tarafından girdi değerlerine karşılık çıktı değerleri bulunan diğer bir ifade ile etiketli veri setleri kullanılarak gerçekleştirilen öğrenme modeli olarak tanımlanmaktadır. Bu öğrenme modelinde algoritmalar girdi ve çıktı değerleri arasındaki ilişkiyi temsil eden fonksiyonlar üretmeye çalışır. Böylece yeni girdi örnekleri verildiğinde çıktı değeri için tahmin yapılabilmektedir (Keskin 2018).
- Denetimsiz Öğrenme: Bu öğrenme yönteminde etiketsiz veri setleri kullanılmaktadır. Modelin, veri setindeki gizli örüntüleri keşfetmesi beklenmektedir (Cebeci, 2020). Kümeleme, boyut indirgeme gibi yöntemlerle gözlemlerin veya değişkenlerin alt gruplara ayrılması amaçlanmaktadır. Grup üyelikleri daha önceden bilinmediğinden atamanın doğruluğu ölçülemez. Bu nedenle modelin etkinliği ölçülürken grup homojenlikleri incelenmektedir (Altunkaynak 2022). Bu yöntemde benzer örneklerin gruplandırılması, boyut azaltılması, yoğunluk tahmin edilmesi gibi görevler gerçekleştirilmektedir (Sebetci 2023).
- Yarı Denetimli Öğrenme: Bu öğrenme yönteminde etiketli verilerin yanında bir miktar da etiketsiz veri yer alır. Az sayıdaki etiketli veri etkisiz veriler ile nasıl başa çıkılacağı konusunda rehberlik yapmaktadır (Zizak vd. 2020).

2.6 Yapay Sinir Ağları

Yapay sinir ağları insan sinir hücrelerini biçimsel ve işlevsel olarak taklit ederek oluşturulan makine öğrenmesi türüdür (Yurtoğlu 2005). Yapay sinir ağları girdi katmanı, gizli katman ve çıktı katmanlarından oluşmaktadır. Veriler girdi katmanından ağa girmektedir. Bu katmanda analiz veya sınıflandırma sürecinden geçen veriler bir sonraki katmana aktarılmaktadır. Gizli katman, verilerini girdi katmanından veya kendinden önceki gizli katmandan almaktadır. Gizli katman sayısı ve bu katmandaki nöronların

sayısı artırılıp azaltılabilmektedir. Bu katmanda toplanan veriler işlenerek bir sonraki gizli katmana ve nihayetinde çıktı katmanına gönderilmektedir ve çıktı katmanında analiz sonuçlandırılmaktadır. Çıktı katmanındaki nöron sayısı da ihtiyaca göre düzenlenebilmektedir. Örneğin, bir duyguyu olumlu veya olumsuz olarak sınıflandırmak için çıktı katmanında tek düğüm bulunmakta ve cümlelerin polaritesine karar verilmektedir (Anonymous 2023b). Çok sınıflı sınıflandırma işlemlerinde bu sayı artmaktadır.

Yapay sinir ağı eğitimi sürecinde temel amaç ağırlık ve yanlılık değerlerini en uygun şekilde belirleyerek modelin doğru tahmin yapmasını sağlamaktır. Şekil 2.3'te bir yapay sinir ağı yapısı gösterilmiştir.



Şekil 2.3 Yapay sinir ağı katmanları (Cansız 2023)

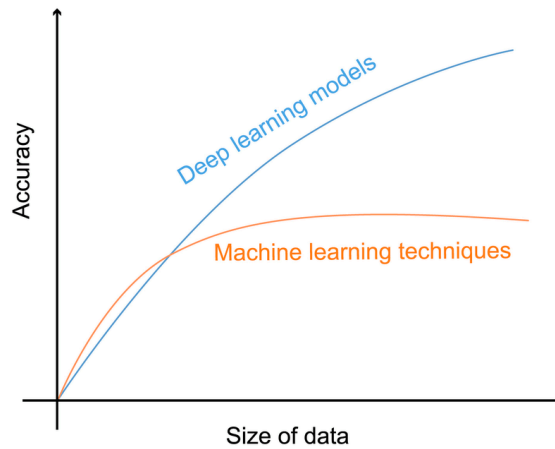
Yapay sinir ağlarının bazı avantaj ve dezavantajları vardır. Karmaşık ve büyük boyutlu verilerle işlem yapabilmesi, gürültülü veya eksik verilere dayanıklı olması, manuel olarak özellik çıkarmaya gerek kalmadan bunu yapabilmesi ve veri setindeki gizli örüntüleri ortaya çıkarabilmesi yapay sinir ağlarının avantajlarıdır (Ivankov 2023). Diğer yandan yapay sinir ağlarının öğrenme süresinin uzun olması, bazı ağlarda parametre değerlerinin rastgele belirlenmesi ve bir kural olmaması gibi dezavantajları vardır (Öztemel 2006). Çok fazla veriye ve güçlü bilgisayar donanımlarına ihtiyaç

duyulması da zorluklara neden olabilmektedir. Ağlar sadece sayısal verilerle çalışabildiğinden metin veya görsel türündeki verilerin sayısala dönüştürme işlemleri de ağın performansı üzerinde etki oluşturmaktadır.

2.7 Derin Öğrenme

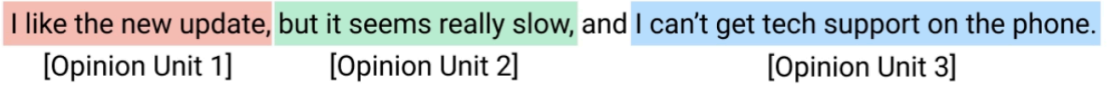
Derin öğrenme, sinir ağlarının katmanlarının desteklediği, insan beyninin çalışma şeklini modelleyen bir makine öğrenmesi yaklaşımıdır. Doğal dil işleme, ses tanıma, tıbbi görüntü analizi gibi çeşitli alanlarda kullanılmaktadır. Çok katmanlı yapay sinir ağları derin öğrenme olarak isimlendirilmektedir (Anonymous 2023g). Derin öğrenme etiketli, etiketsiz veya yarı etiketli veriler ile işlem yapabilmektedir.

Derin öğrenme modellerinin eğitimi zaman alıcıdır ve çok miktarda veri gerektirmektedir. Öğrenme süreci bir çocuğun gelişim sürecine benzetilmektedir. Bir bebek doğru bir şekilde cümle kurmayı öğrenene kadar çok sayıda cümle duymaktadır. Ancak konuşmayı öğrendikten sonra süreçlere ve kurallara hakim olduklarından ileri düzey dil ve konuşma görevlerini yerine getirebilmektedirler (Roldos 2020). Şekil 2.4’de gösterildiği gibi veri miktarı arttıkça derin öğrenme modelinin doğruluğu artmaktadır.



Şekil 2.4 Derin öğrenme ve makine öğrenmesi için veri miktarı ve doğruluk ilişkisi (Es-sabery vd. 2021)

Duygu analizinde derin öğrenme yöntemleri sıklıkla kullanılmaktadır ve başarılı sonuçlar elde edilmektedir (Zhang vd. 2018, akt. Mercha ve Benbrahim 2023). Cümleleri basitçe pozitif veya negatif olarak sınıflandırmaktan ziyade Şekil 2.5’de verildiği şekliyle bir ürüne ait yapılan yorumda, ürünün hangi bileşeni için hangi duygunun var olduğuna ait detaylı bilgiler elde edilmektedir.



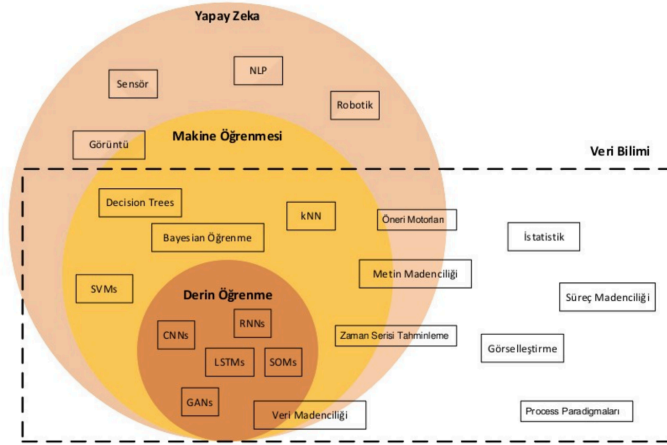
I like the new update, but it seems really slow, and I can't get tech support on the phone.
[Opinion Unit 1] [Opinion Unit 2] [Opinion Unit 3]

Şekil 2.5 Ürün yorumunun konulara göre bölünmesi (Roldos 2020)

Cümle Şekil 2.5’de verildiği gibi bölümlere ayrıldıktan sonra birinci bölüm kullanışlılık, ikinci bölüm işlevsellik, üçüncü bölüm de destek olarak işaretlenmiştir. Sonra bu bölümlerin polaritesi belirlenmiştir. Bir ürün hakkında yapılan binlerce yoruma aynı işlem uygulandığı düşünüldüğünde, örneğin işlevsellik konusu için bildirilen negatif görüşler ürün hakkında düzenleme yapmak kıymetli bir bilgiyi ortaya çıkarmış olacaktır (Roldos 2020).

2.8 Veri Bilimi

Yapılandırılmış veya yapılandırılmamış verilerin işlemlerden geçirilerek kullanıma uygun hale getirilmesi ve bu verilerden anlamlı bilgiler çıkarılması işlemine veri bilimi denir. Veri bilimi süreçlerinde matematik, istatistik, makine öğrenmesinden yararlanır (Altunkaynak 2022). Veri bilimi büyük miktardaki verilerden kullanışlı bilgiler elde etmek amacıyla gerçekleştirilen işlemleri içeren bir çalışma alanıdır. Şekil 2.6’da temel kavramlar bölümünde açıklanan yapay zeka, makine öğrenmesi ve diğer kavramların birbirleriyle ilişkilerine dair bir görsel paylaşılmıştır.



Şekil 2.6 Kotu ve Deshpande (2019) tarafından görselleştirilen yapay zeka ve diğer kavramlar arası ilişkiler (Özdemir 2022)

2.9 Metin Madenciliği

Metin madenciliği, metin türündeki veri yığınlarından kıymetli örüntüleri keşfetmeyi amaçlayan bir çalışma alanıdır. Yapılandırılmamış verileri işleyerek anlamlı bilgiler elde etmek hedeflenmektedir.

Linguamatics (2021) tarafından metin madenciliği için bir tanımlama yapılmıştır. “Metin madenciliği, dokümanlarda ve veri tabanlarında bulunan yapılandırılmamış haldeki metinleri, analizler ve makine öğrenmesi algoritmaları için uygun olacak şekilde yapılandırılmış veriye dönüştüren doğal dil işleme uygulamalarını kullanan yapay zeka teknolojisi” şeklinde ifade edilen bu tanımda temel kavramlar bölümünde değinilen kavramların birbirleri ile ilişkisi ortaya konmuştur (Özdemir 2022).

3. DUYGU ANALİZİ

3.1 Duygu

Duygu terimi pek çok disiplinin araştırmalarına konu olmuştur. Duygunun ne olduğunu açıklamak için çeşitli görüşler sunulmuş olsa da bir fikir birliğine varılamamıştır ve duyguların kökeni, işlevi ve diğer etkilerine yönelik farklı görüşler ortaya atılmıştır. Duygu, Vikipedi (2022) tarafından “biyeyin ruh halinde biyokimyasal (içsel) ve çevresel tesirlerle etkileşimden doğan kompleks psikofizyolojik bir değişim” olarak tanımlanmıştır. Etnograflar kültürlerin kendine özgü duyguları olduğuna ve farklı zamanlarda kültürlerin duyguların ifade edilışinin değiştiğine dikkat çekmiştir. Tıp ve biyoloji literatüründe duygunun evrenselliği, beden diline yansımaları değerlendirilirken, psikoloji literatüründe “kişinin kendisi için önemli olan bir durum ya da etkileşim içinde olduğu zaman meydana gelen his ve duygulanım” şeklinde tanımlanmıştır (Güldü 2023).

3.2 Duygu Analizi Nedir?

Duygu analizi, belirli bir kişi, kurum, hizmet, olay veya konu hakkında ifade edilen tutum ve görüşleri ortaya çıkarmak amacıyla doğal dil işleme ve metin madenciliği yöntemlerini içeren bir dizi yöntemden oluşmaktadır. Sağlık, politika, pazarlama ve finans alanlarında duygu analizi sıklıkla kullanılmaktadır. Sosyal medya paylaşımları, çevrimiçi alışveriş sitelerindeki yorumlar, film ve müzik albümü yorumları duygu analizi için başlıca veri kaynaklarıdır.

Duygu analizi metin madenciliğinde sıklıkla yararlanılan bir çalışma alanıdır. Kişilerin duygularının polaritesi belirlenerek ürün pazarlama veya olaylara ilişkin anlamlı bilgiler çıkarmak amaçlanmaktadır. Diğer yandan video ve görsellerden yararlanarak yürütülen

alışmalar da bulunmaktadır. Gvenlik ve davranış analizi alanlarında kullanıldığı gibi glmseme algılaması ile otomatik fotoğraf ekimi gibi uygulama alanları bulunmaktadır (Safalı ve Avarođlu 2021).

2000’li yıllarda internetin ve sosyal medyanın yaygınlaşmasının bir sonucu olarak duygu ieren metinlere erişim artmıştır (Cebeci 2020) ve bu metinlerden bilgi ıkarmak bir ihtiyaç haline gelmiştir. Sonu olarak internet ve sosyal medyanın gelişimi ile duygu analizi alışmalarının artışı arasında bir ilişki grlmektedir (Lei ve Liu 2021).

Duygu analizi alışmaları ikili sınıflandırma (pozitif veya negatif) veya l sınıflandırma (pozitif, negatif veya ntr) şeklinde yapılabilmektedir. Bunların dıřında metinlerin mutluluk, neře, znt gibi spesifik duyguları ierme durumları arařtırılmaktadır (Can vd. 2018).

Duygu analizinde kullanılan duygu sınıflandırma teknikleri szlēe dayalı ve makine ērenmesine dayalı olmak zere ikiye ayrılmaktadır. Szlēe dayalı yntemlerde metinlerde geen duygu kelimeleri nceden belirlenmiř duygu skorları ile deēerlendirilmektedir. Cmlede yer alan duygu kelimeleri puanlanıp toplandıktan sonra elde edilen toplam sıfırdan bykse cmlenin pozitif, 0 ise ntr kk ise negatif duyguya sahip olduēuna karar verilir (Oēuzlar ve Kızılkaya, 2019). Etiketli veri seti ihtiyacı olmadığından veriyi toplama ve işleme maliyetleri daha azdır. Bu ynteme ait bazı dezavantajlar da bulunmaktadır. Kelimeler baēlamdan baēımsız olarak deēerlendirildiēi iin anlam kayıpları yařanabilir (Tokcaer 2021). Var olan duygu kelimesi bir duygu ifade etmiyor olabilir, alaycı ifadeler zıt polariteye işaret ediyor olabilir (Liu 2020). Her dil iin ayrı bir duygu szlē oluřturmak ihtiyacı bir diēer sınırlılıktır (Kaynar vd. 2017).

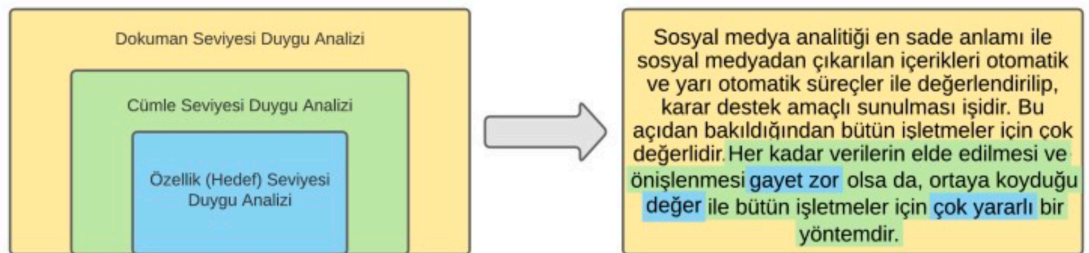
Makine ērenmesine dayalı yntemlerden denetimli makine ērenmesi yaklaşımda etiketli eēitim verisi kullanılarak duygu kategorisini belirleyecek modeli tahmin etmek

amaçlanmaktadır. Naive Bayes, Destek Vektör Makineleri, Maksimum Entropi, Yapay Sinir Ağları denetimli makine öğrenmesi sınıflandırıcılarından bazılarıdır (Tokcaer 2021).

Denetimsiz makine öğrenmesi yaklaşımlarında ise etiketsiz veri setleri kullanılmaktadır. Algoritmadan beklenen duygu ifadelerine kümeleme, özetleme veya sınıflandırma yapmasıdır. Noktasal Karşılıklı Bilgi Katsayısı (PMI), Örtülü Anlam Dizineleme (LSI), Örtülü Dirichlet Dağılımı kullanılan yöntemlerden bazılarıdır (Tokcaer 2021). Denetimsiz öğrenme modelleri denetimli öğrenme modellerine göre daha fazla veriye ihtiyaç duymaktadır (Mercha ve Benbrahim 2023).

3.2.1 Duygu analizi düzeyleri

Duygu analizi metin, cümle ve hedef kategori olmak üzere üç farklı düzeyde gerçekleştirilmektedir (Cebeci 2020). Şekil 3.1’de duygu analizi seviyeleri gösterilmiştir.



Şekil 3.1 Duygu analizi seviyeleri (Cebeci 2020).

Doküman seviyesinde duygu analizinde incelenen metin bir bütün olarak değerlendirilip duygu polaritesi ortaya konmaktadır. Doküman seviyesinde duygu analizinde, metinde tek bir varlık hakkında ve tek bir kişinin görüşünün olduğu

varsayılmaktadır. Birden fazla varlık veya kişi olduğunda anlamda bir kayıp yaşanabileceğinden bu analize dair tereddütler bulunmaktadır (Cebeci 2020). Doküman bir yorum, tweet, mesaj, haber metni olabilir. Bir tüketici yorumu genellikle tek bir ürün hakkında ve bir kullanıcı tarafından yapılan yorumları içerir. Diğer yandan çeşitli blog paylaşımları birden fazla varlık hakkında görüş içerebileceğinden doküman seviyesinde analizlerde tüketici yorumları tercih edilmektedir (Lei ve Liu 2021).

Cümle seviyesinde duygu analizinde tek bir cümlemin duygu polaritesi (pozitif/negatif/nötr) belirlenmeye çalışılır. Bu seviyede her cümlede tek bir varlık hakkında görüş olduğu kabul edilmektedir. Doküman seviyesindeki kadar olmasa da bu düzeye ait tereddütler de bulunmaktadır (Cebeci 2020).

Özellik düzeyinde, varlık hakkında yapılan görüşlerin, varlığın hangi özelliği için negatif hangi özelliği için pozitif görüş bildirildiği analiz edilmeye çalışılmaktadır. Bir ürünün hem olumlu hem olumsuz yönleri değerlendirildiğinde hangi yönlerden olumlu hangi yönlerden olumsuz olduğunun bilinmesi kurumlar için avantaj sağlayabilir. Bir ürün için bildirilen görüşler ve bir ürünün farklı boyutları ile analiz edilmesi farklı seviyelerde kararlar alınmasında etkili olacaktır (Şeker 2016).

3.3 Duygu Analizi Kullanım Alanları

Pazarlama alanında duygu analizi, müşterilerin ürün ve hizmetler hakkındaki görüşlerini öğrenmek amacıyla kullanılmaktadır. Şirketler sosyal medya, blog, haber siteleri gibi sanal ortamlarda ürün ve hizmetleri için yapılan yorumları takip edip analizini yaparak olumsuz yorumlarla ilişkili problemleri çözmek ve olumlu yorumlardan yararlanmak için kullanabilir. Çevrimiçi alışveriş sitelerinde, ürün hakkındaki geri bildirimlerin müşterilerin tercihlerine etki ettiği gözlenmektedir. Bu nedenle şirketler ürünleri ve marka imajı hakkında bilgi edinmek, potansiyel sorunları engellemek, ürüne yönelik eleştirilerden yola çıkarak gerekli düzenleme ve düzeltmenin yapılması amacıyla duygu

analizinden faydalanabilmektedir. Bir diğerkatkısı ise günümüzün hızla değışen koşullarına erişebilmektir Çevrimiçi alışveriş sitelerindeki değlendirmeler, anketler, haberler, bloglar veri kaynaklarıdır (Lei ve Liu 2021).

Diğeryandan, kitlelerin görüşlerine etki etmek isteyen sahtekar kullanıcılar, sosyal medya aracılığı ile aldatıcı içerikler paylaşmakta, sahte ve bot hesaplar kullanılarak manipölasyon yapılabilmektedir. Bu paylaşımları belirlemek ve dezenformasyonu engellemek günümüzde önemli bir ihtiyaç olarak görölmektedir (Lei ve Liu 2021).

Politika alanında halkın veya seçmenlerin siyasi partiler, adaylar, kanunlar hakkındaki görüşlerini belirlemek için duygu analizi kullanılmaktadır. Seçim öncesinde, hem partiler hem seçmenler için değerkil bilgiler elde edilebileceğı gibi seçim sonuçları için de tahmin yapılabilir. Adayların konuşmaları, anketler ve sosyal medya paylaşımları veri kaynaklarıdır (Lei ve Liu 2021). Sosyal medya verileri analizi ile edinilen bilgiler siyasilerin konularını, mesajlarını değıştirme imkanı sunabilir. Seçmen ile etkili bir iletişim kurabilmek için, örneğın bir adayın nasıl algılandığının belirlenmesi, etkili bir seçim kampanyası yürütmek için faydalı olacaktır. Belli konularda halkın görüşlerini inceleyerek atılacak adımlara karar vermek gerekmektedir. Bunları duygu analizi gibi sistemlerle gerçekleştirmek kriz anlarında stratejiyi hızla güncelleyerek sorunların çözölmesine olanak tanımaktadır (Edwards 2021).

Pandemi döneminde sosyal medya platformlarında aşılar, hastalığın belirtileri vb. konularda fikir alışverişı ve tartışmalar olmuştur. Araştırmacılar tarafından bu duyguların anlaşılması için çalışmalar yapılması tıp alanında duygu analizi kullanımına örnektir. Ayrıca intihar düşüncesinin belirlenmesi amacıyla sosyal medya verileri analiz edilmektedir. Diğeryandan hastalık öyküsünü, belirtileri, tedavi planını içeren tıbbi öyküler de veri kaynağı olarak değlendirilmektedir (Denecke ve Reichenpfader 2023).

3.4 Duygu Analizinde Karşılaşılan Zorluklar

Duygu analizinde karşılaşılan zorluklardan biri veri yetersizliğidir. Analiz yapılacak dilde ve alanda yeterli miktarda veya güncel veri olmaması modelin hatalı sonuçlar üretmesine sebep olmaktadır (Pervan Akman 2020).

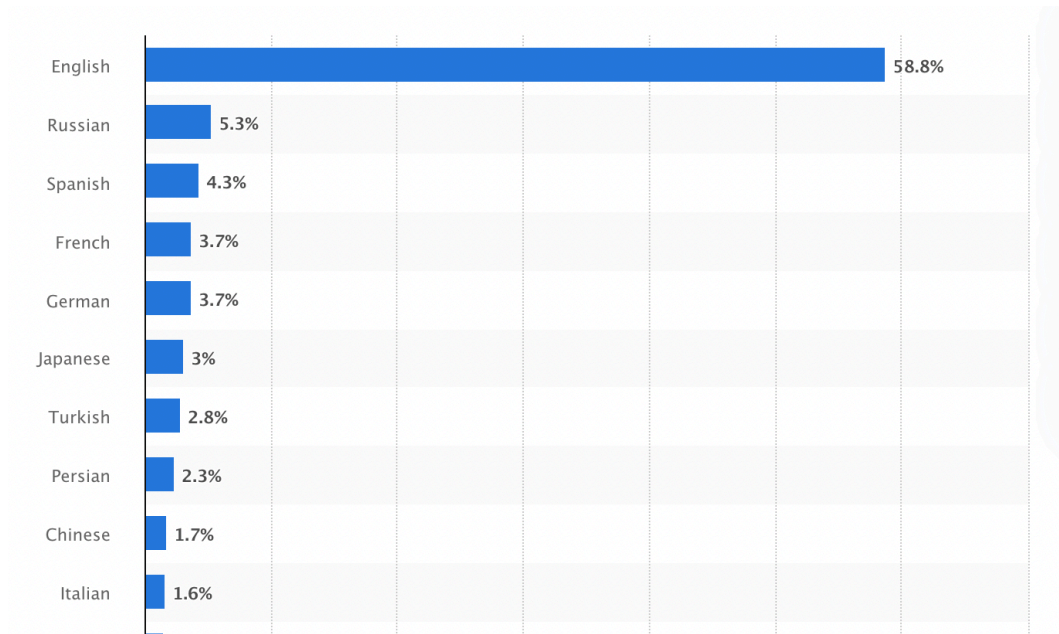
Bir diğeri alan bağımlılık problemidir. Bu, bir kelimenin farklı alanlarda farklı anlamlarda kullanılmasını ifade etmektedir. “Bugün hava çok sıcak nefes alamıyorum cümlesinde “sıcak” kelimesi negatif bir duyguya işaret etmekte iken “Sıcak satış yeteneği ile ön plana çıkıyor” cümlesinde pozitif veya negatif bir duyguyu işaret etmemektedir. Diğer yandan yapıcı olumlu anlamca olumsuz cümleler de analizin başarımında etkili olabilir. “Bu televizyon fazlasıyla elektrik kullanıyor” cümlesinde duygu içeren bir kelime veya ek olmamasına rağmen anlamca olumsuzdur (Cebeci 2020). Alaycı ve iğneleyici sözcükler de zorluklara neden olmaktadır. Bu cümleler genellikle yapıcı olumlu anlamca olumsuz olduklarından gerçekte negatif bir ifade pozitif olarak işaretlenebilmektedir (Yılmaz 2023).

Duygu analizi çalışmaları popüler olsa da geliştirilen algoritma ve kütüphanelerin İngilizce, Almanca, Fransızca gibi yaygın kullanım alanına sahip dillerde yoğunlaştığı görülmektedir. (Tokcaer 2021). Diğer yandan küresel çapta müşterilere sahip şirketler farklı dillerde geri bildirimler alabilmektedirler. Duygu analizi sistemleri tek dilde sınıflandırma yapmak için eğitildiğinden farklı dilde yapılacak analiz bilgi kayıplarına sebep olacaktır (Yılmaz 2023).

4. ÇOK DİLLİ DUYGU ANALİZİ UYGULAMASI

4.1 Çok Dilli Duygu Analizi Nedir?

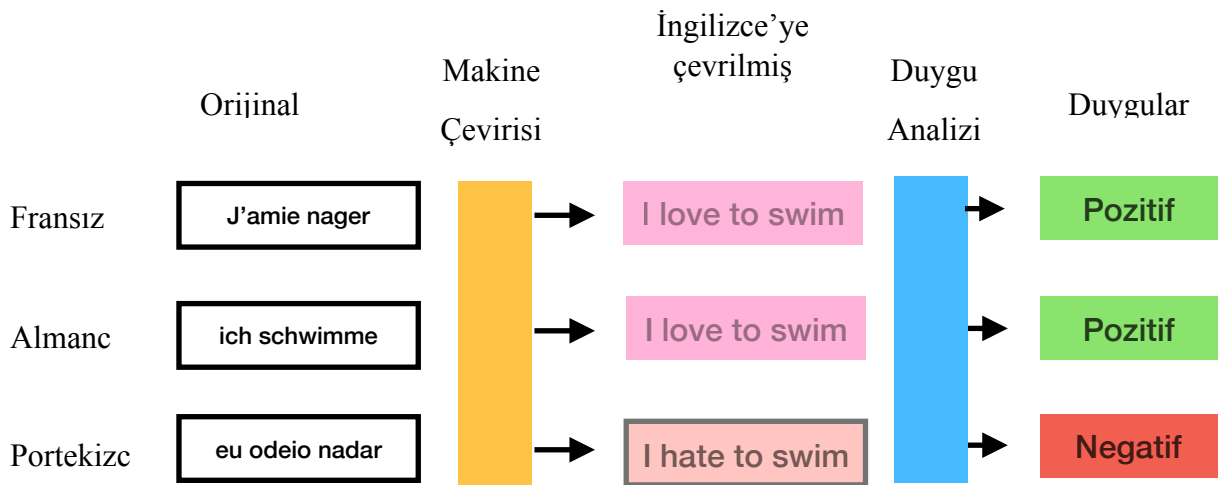
Sosyal medya platformlarının kullanım alanının genişlemesiyle barındırdığı içerikler ve kullanım alanları da çeşitlenmektedir. Örneğin COVID19 pandemisi döneminde sosyal medya, bilgi paylaşımı ve haberlerin yayılmasında etkin bir rol oynamıştı (Goel ve Gupta 2020). Paylaşımlar, alt yazıların ve çeviri araçlarının yardımıyla, o dili bilmeyen kişilerce dahi izlenip yorumlanabilmektedir. Sonuç olarak ortaya çok dilli ve yüksek hacimli bir içerik ortaya çıkmaktadır. Çok dilli duygu analizi de birden fazla dil içeren metinlerin duygusal içeriğini anlamak için kullanılmaktadır. Duygu analizi çalışmalarında büyük hacimli veri setlerine ihtiyaç olduğundan bu mecralar araştırmacıların ilgi odağı haline gelmiştir (Şahinaslan vd. 2022).



Şekil 4.1 İnternette en fazla kullanılan diller (Statista 2023)

Şekil 4.1’de internette en fazla kullanılan dillere dair oranlar paylaşılmıştır. Buna göre internette en çok kullanılan dil İngilizce olsa da, diğer dillerde yazılmış metinler de mevcuttur (Statista 2023). Duygu analizi dile bağımlı olduğundan her dil için duygu analizi modeli geliştirmek gerekmektedir. Durak kelimeler, kelimeleri köklerine indirme işlemleri her dilde farklıdır. Her dil için bu işlemleri yapmak ve duygu analizi modeli geliştirmek zor ve zaman alıcı olmaktadır (Mboutayeb 2021). Çok dilli duygu analizi ile bir dil için geliştirilmiş duygu sözlükleri ve diğer araçlar kullanılarak sınırlı kaynağa sahip dillerde de çalışmaların yürütülmesi amaçlanmaktadır (Taboada vd. 2011).

Diller yüksek kaynaklı veya yetersiz kaynaklı olarak sınıflandırılmaktadır. Yetersiz kaynaklı dil ifadesi; duygu sözlüğü, açıklamalı derlem, bol miktarda veri gibi kaynaklara erişimde sınırlılıkları ifade etmektedir. Yetersiz kaynaklı bir dilde duygu analizi yapmak için üç yöntem önerilmiştir. İlk yöntemde yetersiz kaynaklı dildeki veri, makine çevirisi ile İngilizceye çevrilip İngilizce duygu sınıflandırılması yapılmaktadır. (Kim ve Hovy 2006). İkinci yöntemde İngilizce külliyat makine çevirisi ile hedef dil veya dillere çevrilmekte ve model çevrilmiş veri ile eğitilmektedir (Mboutayeb 2021). Son yöntemde İngilizce duygu sözlüğü hedef dile çevrilmekte ve sözlük tabanlı duygu sınıflandırması yapılmaktadır (Kim ve Hovy 2006). Şekil 4.2’de makine çevirisine dayalı bir duygu analizi örneklendirilmiştir. Bu akışta farklı dillerde oluşturulmuş metinler İngilizce’ye çevrilmiş ardından duygu sınıflandırması yapılmıştır (Mabokela 2023).



Şekil 4.2 Makine çevirisine dayalı duygu analizi örneği Mabokela vd. (2023)’ten çeviri

Çok dilli duygu analizi sürecinde bazı zorluklar ile karşılaşmaktadır. Google Translate gibi araçlar ile yapılan çevirilerin doğruluk oranı analiz başarımı üzerinde etkili olmaktadır. Farklı dillerde yer alan kısaltma, argo, deyim gibi dilin kendine has özellikleri tercüme başarısı üzerinde olumsuz etki oluşturabildiğinden duygu analizi performansını da olumsuz etkilenmektedir (Şahinaslan vd. 2022). Her dilin kendine özgü yapısı olduğundan bir dil için geliştirilen modeli diğer diller için kullanmak problemlere sebep olmaktadır.(Mercha ve Benbrahim 2023).

4.2 Veri Seti Belirleme

Duygu analizine problemin ve çözülmesi beklenen soruların ortaya konması ile başlanır. Bunlar doğrultusunda da bir veri setine ihtiyaç duyulur. Veri setini elde etmenin çeşitli yolları vardır. Dahili veri tabanı veya veri arşivi gibi şirketin iç kaynaklarından elde edilebileceği gibi web kazıma yoluyla dışardan da elde edilebilir. Web kazıma programlama yoluyla internetten içeriğin elde edilmesini ifade etmektedir. Ayrıca bazı web tabanlı sistemlerin API'si kullanılarak anlık veri çekilebilmektedir. Bir diğer seçenek ise hazır veri setlerini kullanmaktır (Zizka vd. 2020).

Bu tez çalışmasında MultiEmo veri seti kullanılmıştır. MultiEmo, 11 dili içeren çok dilli duygu analizi görevleri için oluşturulmuş bir veri setidir. Tıp/sağlık, turizm, pazarlama ve eğitim alanlarında tüketici görüşleri yer almaktadır. Orijinal veri seti Felemenkçe dilindeki 57,466 cümleden oluşmaktadır, bu cümleler daha sonra otomatik olarak İngilizce, Çince, İtalyanca, Japonca, Rusça, Almanca, İspanyolca, Fransızca, Felemenkçe ve Portekizce'ye çevrilmiştir (Clarín, 2015). Bu tez çalışmasında kullanılmak üzere 4 alanı da içeren İngilizce, Almanca, Felemenkçe dillerindeki tüketici görüşleri seçilmiştir. Ek olarak 4 alandaki tüketici görüşleri Türkçe diline çevrilecek veri setine dahil edilmiştir. Böylece çok dilli duygu analizi alanında Türkçe literatüre katkı yapmak hedeflenmiştir.

4.3 Verilerin Çevrilmesi

Veri setinin Türkçe diline çeviri işlemi Python programında yer alan “deep_translator” kütüphanesi kullanılarak gerçekleştirilmiştir. Metinler Google Çeviri API’si kullanılarak çevrilmiştir. Hedef dile çevrilen metinler yeni bir dizine kaydedilerek işlem tamamlanmıştır.

Makine çevirisi, bilgisayarların bir metni bir dilden başka bir dile çevirmesidir (Anonymous 2023a). Makine çevirisinin tarihi 1950’li yıllara kadar uzanmaktadır. II. Dünya Savaşı sonrasında başlayan soğuk savaş döneminde tarafların birbirlerine ait belgelerden bilgi edinme çabaları bu alanda yapılan çalışmaları yoğunlaştırmıştır. Georgetown Üniversitesi’nde IBM ortaklığıyla yürütülen proje makine çevirisi tarihinin ilki olması yönüyle öne çıkmaktadır. Bu deneyde Rusça cümlelerin İngilizce’ye bilgisayar aracılığıyla çevrilmesi amaçlanmıştır. O dönemde 60’tan fazla cümlelerin tam otomatik tercümesi bilim camiasında büyük bir etki yaratmıştır. Çalışmada kullanılan IBM 701 tipi bilgisayar roket yörüngeleri, hava tahmini gibi askeri problemleri çözmek için geliştirilmiştir. Tennis kortu büyüklüğündeki bu bilgisayar dört işlem, kaydırma gibi 33 farklı işlemi gerçekleştirebilmekteydi (Hutchins 2004, akt. Aslan 2019).

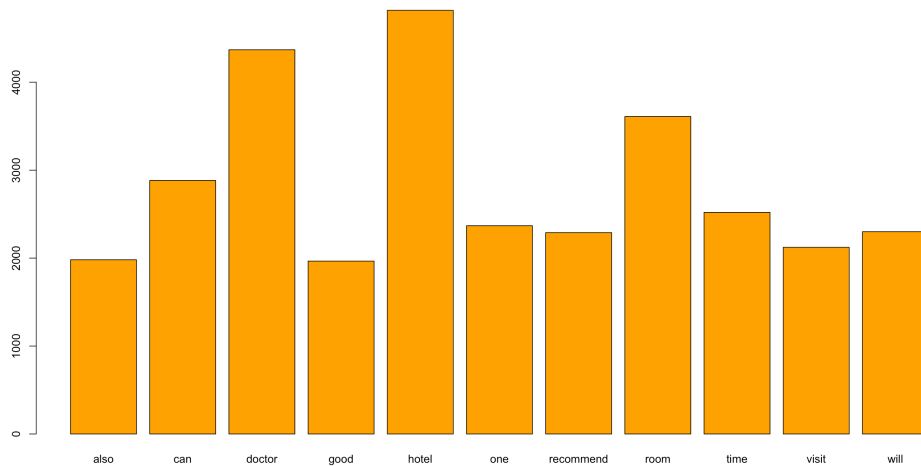
Çevirilerin doğruluk oranı yapılan metin analizlerinin başarımı üzerinde etkili olduğundan, çeviri araçlarının artan doğruluk performansı metin ve duygu analizlerinde etkisini göstermektedir. Geliştirilecek yeni yöntemlerle çevirilerin insan çevirilerindeki doğruluğu yakalayacağı beklenmektedir. Öyleki makine çevirilerinin insan çevirilerinden daha kaliteli sonuçlar ortaya koyması da şaşırtıcı olmayacaktır (Sadıkov ve Sarıgül 2021).

4.4 Ön İşleme

Metinler çoğu zaman analize uygun formda değildir ve bir ön işleme sürecinden geçmesi gerekmektedir. Bu aşamada analizde bir etkisi olmayacak detaylar çıkarılarak verilerin kalitesini artırmak amaçlanmaktadır (Sebetci 2023). Herhangi bir işlem yapılmadan ve veri arındırılmadan analize başlamak hem süreci hem de sonuçları olumsuz etkilemektedir. Ön işleme aşamasında gereksiz işaretlerden ve boşluklardan arındırma, noktalama işaretlerini temizleme, metni küçük harfe çevirme, metni tokenize etme, durak kelimeler çıkarma, kelimeleri köke indirgeme vb. işlemler yapılmaktadır (Çelik vd. 2021).

4.4.1 Tokenizasyon

Tokenizasyon metni küçük birimlere ayırma işlemidir. Belirtkeleme adı da verilen bu işlem kelime, karakter, cümle düzeylerinde uygulanabilmektedir. Kelime tokenizasyonu en yaygın kullanılan biçimdir, metin kelimelere bölünmektedir. Kelimelere ayrılan metnin kelime sıklıkları belirlenerek metnin ne hakkında olduğu gibi bilgiler elde



Şekil 4.3 En sık kullanılan kelimeler

edilebilmektedir (Burchfiel 2022). Şekil 4.3'te veri setinde en sık kullanılan kelimeler belirlenerek görselleştirilmiştir.

4.4.2 Durak kelimeleri temizleme

Durak kelimeleri bir dilde en çok kullanılan kelimeler olarak tanımlanmaktadır. Ancak duygu analizinde bir etkisi olmayacak, polarite içermeyen kelimeler olabileceği gibi tek başına anlam ifade etmeyen (ve, ile, tarafından vb.) kelimeler de olabilir. Bu kelimelerin veri setinden çıkarılması daha anlamlı sonuçların elde edilmesini sağlamaktadır (Yayla vd. 2021). Python, R gibi programlama dillerinde durak kelimeleri temizlemek için paketler bulunmaktadır ancak analizi yapılacak dil için veya spesifik bir alan (domain) için işlem gerçekleştirilemiyorsa duraklama kelimelerinden oluşan bir vektör kullanılarak gerekli filtreleme işlemi yapılabilmektedir.

4.4.3 Kök bulma

Bu işlem ile kelimeler cümledeki anlamlarına göre kök haline getirilmektedir. Böylece aynı köke sahip kelimeler belirlenmektedir. Kök bulma yapılırken yeterli ek çıkarılmaması durumunda kelimenin köküne inmemek veya fazla ek çıkararak başka bir kelime köküne gitmek gibi hatalara düşmemek için dikkatli olmak gerekmektedir (Oğuzlar ve Kızılkaya 2019). Şekil 4.4'de ön işleme işlemlerinden önce veri gösterilmektedir. Şekil 4.5'de ise ön işlemeden sonra elde edilen veri görülmektedir.

```
> head(sentence_train_en, 10)
[1] "The taxi arrives on time, it seems to be, it is a completely random, one-time
z_zero"
[2] "This lady and our 2-month-old daughter were brought in with a suspicion of ne
ity. __label__z_zero"
[3] "A distinctive feature, as well as an interesting element of the whole decorat
laced in the central place, forming a mini library. __label__z_zero"
[4] "Depending on the time of day, our bodies undergo very large changes in the ca
in arterial pressure, in the secretion of stress hormones or thyroid hormones. __la
[5] "I recommend it to everyone, especially those for whom a visit to the doctor i
label__z_plus_m"
[6] "Today, the visit to the E.R. ended with pain killers and the intern recommend
cal consultations. __label__z_zero"
[7] "Everything old and horrible. __label__z_minus_m"
```

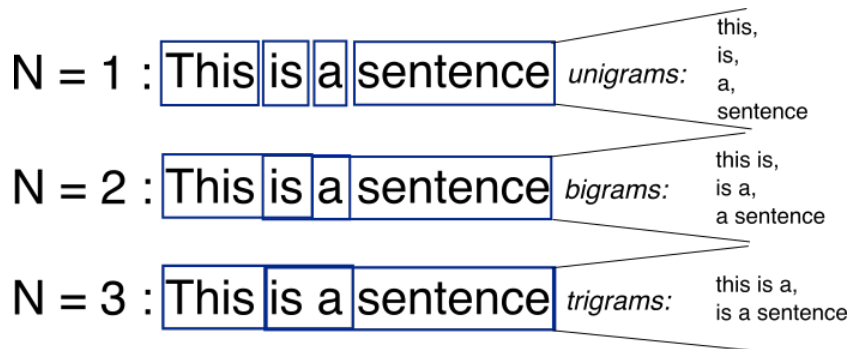
Şekil 4.4 Ön işleme öncesi veri seti


```
[1] "taxi arriv time seem complet random onetim incid"
[1] "ladi monthold daughter brought suspicion nervous system immatur"
[1] "distinct featur well interest element whole decor bookcas place central place form mini librari"
[1] "depend time day bodi undergo larg chang cardiovascular system arteri pressur secret stress hormon thyroid hormon"
[1] "recommend everyon especı visit doctor great stress"
```

Şekil 4.5 Ön işleme sonrası veri seti

4.4.4 N-gram

Bir derlemde seçilen n elemandan oluşan dizilere n -gram denmektedir. Uygulamaya göre heceler, harfler, kelimeler eleman olarak seçilmektedir. n değeri 1 seçildiğinde “unigram”, 2 olduğunda “bigram”, 3 olduğunda “trigram” adını almaktadır. Şekil 4.6’da farklı n değerleri için n -gram örnekleri paylaşılmıştır. Daha fazlası için özel bir adlandırma yoktur. Doğal dil işleme, hesaplamalı dilbilim gibi alanlarda kullanılmaktadır (Akıncı 2020). Bu yöntem ile ardışık bir dizi içindeki tekrar oranını bulmak amaçlanmaktadır (Şeker 2011). Telefonda bir metin mesaj oluştururken karşımıza çıkan kelime tahmin sistemleri (örneğin “iyi” yazdıktan sonra “günler”, “bayramlar”, “geldi” gibi öneriler sunulması) n -gram kullanım alanlarına örnek olarak gösterilebilir (Akıncı 2020).



Şekil 4.6 Farklı n değerleri için n -gram örnekleri (Anonymous 2023c)

Çizelge 4.1’de bigram ve trigram uygulamalarının çıktıları verilmiştir. Daha önce belirtildiği üzere veri seti turizm, sağlık, eğitim ve pazarlama alanlarında müşteri yorumlarını içermektedir. Bu paylaşımlarda tavsiye edilir veya tavsiye edilmez gibi ifadelerin en sık kullanılan ifadelerde yer aldığı görülmüştür. Bu gibi ifadeler diğer kullanıcılara yol göstermekte kullanışlı olacağı gibi yetkili kişiler tarafından tavsiye edilen yönleri geliştirmek ve tavsiye edilmeyen yönleri düzeltmek için bilgi sağlamaktadır. Yüzme havuzu, otel lokasyonu ve klima ifadelerin de sıklıkla kullanılan kelimeler arasında yer alması müşterilerin otel tercihine eden detaylar olarak değerlendirilebilmektedir.

Çizelge 4.1 Bigram ve Trigram Çıktıları

Bigram	Sıklık	Trigram	Sıklık
reception desk	302	lady reception desk	38
swimming pool	252	polish academy sciences	26
dont recommend	224	dont recommend doctor	22
hotel located	165	service nice helpful	20
air conditioning	145	national health fund	20
highly recommend	139	breakfast included price	20

4.4.5 Kelime Çantası (Bag of Words)

Bu yöntemin amacı metni sayısal forma dönüştürmektir. Bunun için metin terimlerine ayrılarak her terimin metinde geçme sıklığı belirlenip elde edilen değer terimin değeri olarak işaretlenmektedir (Aksu ve Karaman 2020). Ancak metindeki kelimelerin frekanslarını belirlemek bir metni temsil etmek için her zaman yeterli değildir. Metinlerde sıklıkla yer alan ancak bir anlam ifade etmeyen “bir”, “gibi”, “de” gibi ifadeler pek çok metnin en yüksek frekans sıklığına sahip olabileceğinden terimler belge

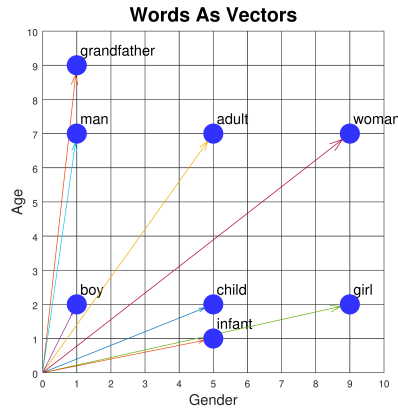
Bag-of-words

[illegible]

31

4.5 Çok Dilli Evrensel Cümle Kodlayıcı (Multilingual Universal Sentence Encoder/ MUSE)

Bilgisayarlar sadece sayılarla işlem yapabildikleri için metinlerin sayılarla ifade edilmeleri gerekmektedir. Bunun bir yolu kodlama modelleridir. Kodlama modelleri kelime, ngram veya cümle gibi metin parçalarını vektörel uzayda tek bir nokta olarak temsil etmektedir. Bunu yaparken de metinlerdeki bilgilerin kaybedilmemesi amaçlanmaktadır. Kelimelerin sayılara atanması işlemine kelime temsili (word embeddings) adı verilmektedir. Bu konumlama, benzer kelimelerin yakın, farklı kelimelerin uzak olacağı şekilde yapılmaktadır. Temsil yapılırken x ve y eksenlerinde farklı özellikler taşınmaktadır. Şekil 4.9'da verilen örnekte dede, kadın ve oğlan gibi kelimeler cinsiyet ve yaş özellikleri kullanılarak temsil edilmiştir. Yeni eksenler ekleyerek farklı özellikleri incelemek de mümkündür (Serrano 2023).

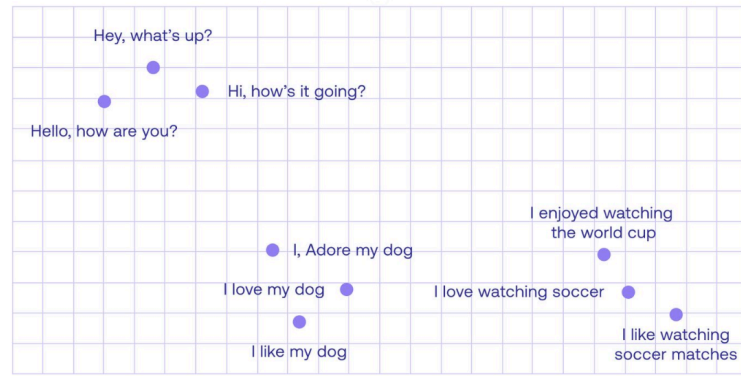


Şekil 4.9 Kelime temsili (Anonymus 2023d)

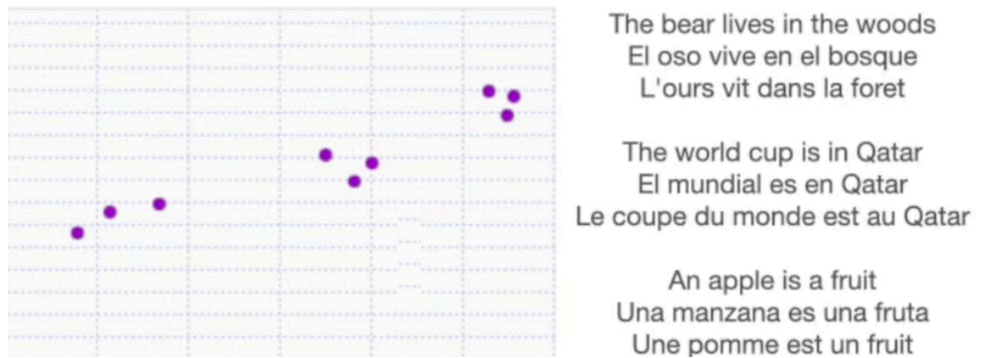
Kelime temsili ile oldukça başarılı sonuçlar elde edilmiştir. Ancak kelimeler bir araya geldiğinde bambaşka bir anlam ortaya çıkabileceği gibi aynı kelimelerin farklı dizilimi ile de farklı anlamlar elde edilebilmektedir. Bu gibi anlam kayıpları analizin başarımı üzerinde etkili olacağından kelime temsili yerine cümle temsili kullanılmaktadır. Kelime temsiline benzer olarak her cümle bir vektör ile ilişkilendirilmektedir. Benzer

cümleler benzer vektörlere, farklı cümleler farklı vektörlere atanmaktadır. Şekil 4.10’da görüleceği üzere benzer anlama sahip cümleler yakın konumlanmıştır. Farklı kelimeler kullanılarak cümle kurulmuş olsa da anlamsal benzerlik yakalanmıştır. Şekil 4.11’de ise birden fazla dil olduğuadaki durum gösterilmektedir. Farklı dillerde ve benzer anlamlardaki cümleler bir arada konumlanmıştır (Serrano 2023).

Output



Şekil 4.10 Cümle temsili (Serrano 2023)

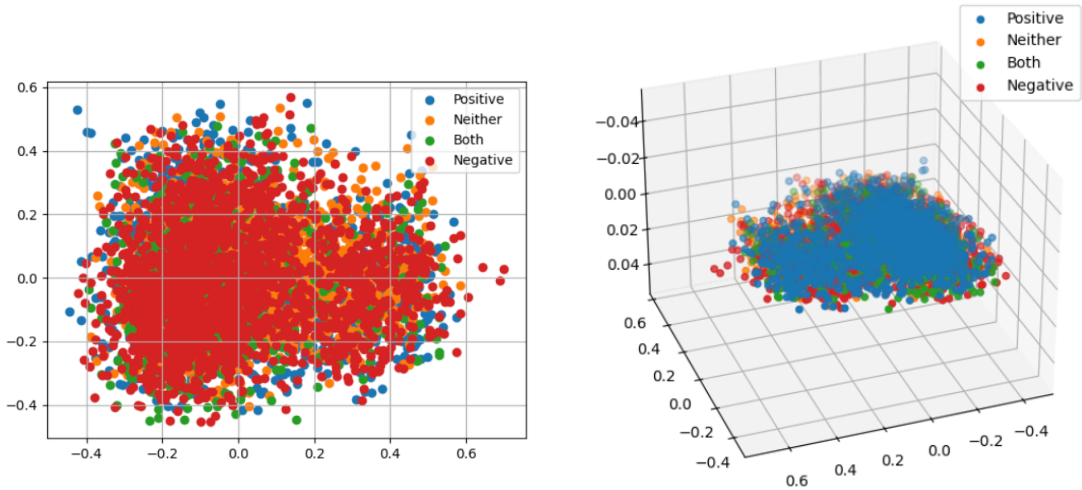


Şekil 4.11 Çok dilli cümle temsili (Serrano 2023)

Bu tez çalışmasında Google AI tarafından 2018 yılında yayınlanmış cümle düzeyinde temsil yapan modelleri içeren bir aile olan Universal Sentence Encoder (USE) modeli

kullanılmıştır. USE önceden eğitilmiş bir derin öğrenme modelidir. Metni duygu analizi vb. doğal dil işleme görevlerinde kullanabilmek için yüksek boyutlu vektörlere kodlamaktadır. Bu modelin TensorFlowHub’da farklı amaçlar için kullanılmak üzere 5 türü bulunmaktadır. Bu modeller Python3 ve TensorFlow modülü ile kullanılabilmektedir (Salvaggio, 2019).

Bu tez çalışmasında USE modelinin çok dilli türü olan MUSE kullanılmıştır ve her cümle 512 ögeli bir dizi ile eşleştirilmiştir. Temsil işleminden sonra boyut azaltmak amacıyla temel bileşenler analizi uygulanmıştır. Temel bileşenler analizi (principal component analysis/ PCA) çok boyutlu verilerde en yüksek varyans ile boyut indirgemek amacıyla kullanılmaktadır. Böylece verinin genel özellikleri tespit edilerek boyutları küçültülmüş olmaktadır. Boyut azaltma bazı özelliklerin kaybolmasına sebep olsa da bunlar kitle hakkında az bilgi içeren özelliklerdir. PCA temel olarak çok boyutlu bir verideki temel özellikleri yakalayarak veriyi daha az özellik ile temsil etmektedir (Öğündür 2020). Veriyi 2 ve 3 boyuta indirmek amacıyla uygulama yapılmış olsa da beklenen ayrışma gerçekleşmemiştir. Uygulama çıktıları Şekil 4.12’de verilmiştir. Şekil 4.13’te veri setindeki bir cümlelerin sınıflandırma örneği verilmiştir.

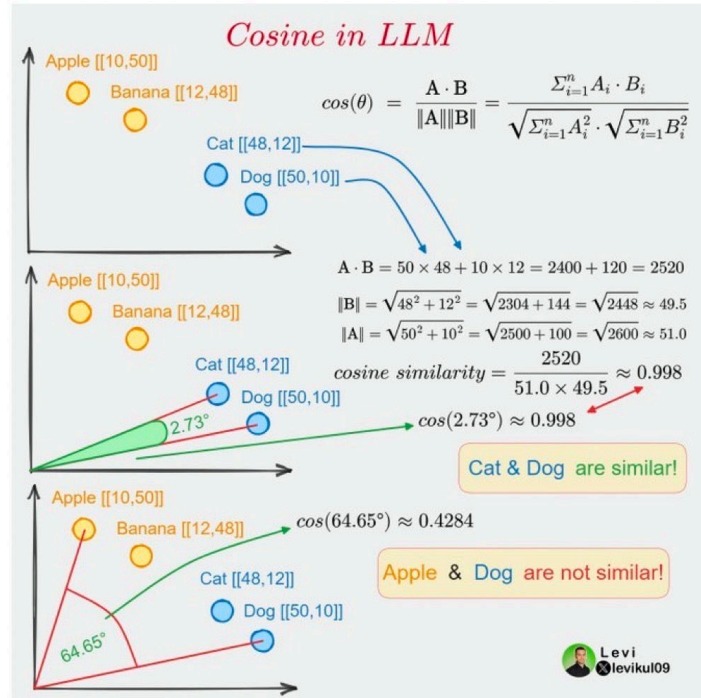


Şekil 4.12 Vektörlerin 2 ve 3 boyut ile temsilleri



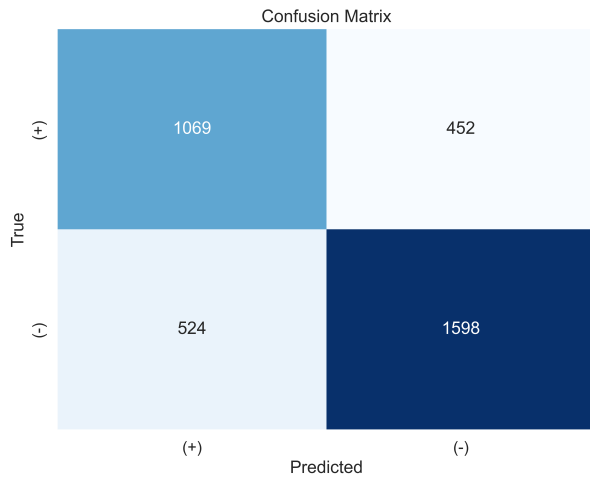
Şekil 4.13 Sınıflandırma örneği

Kosinüs Benzerliği (Cos Similarity Scores), vektörlerin birbirlerine olan benzerliklerini ölçmek amacıyla kullanılmaktadır. İki vektör arasındaki açının kosinüsünün alınması ile hesaplanmaktadır. Aralarındaki açı ne kadar küçükse o kadar benzer oldukları şeklinde yorumlanmaktadır. Açı küçüldükçe değer 1'e yaklaşmaktadır (Cerman 2021). Metinler vektör ile temsil edildikten sonra benzerliğinin niceliği bu eşitlik ile belirlenmektedir. Şekil 4.14'de kosinüs benzerliğinin nasıl hesaplandığı gösterilmektedir.

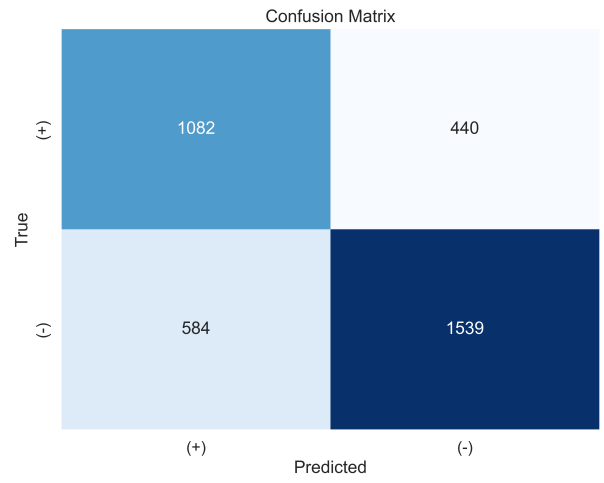


Şekil 4.14. Kosinüs benzerliği (Levi 2023)

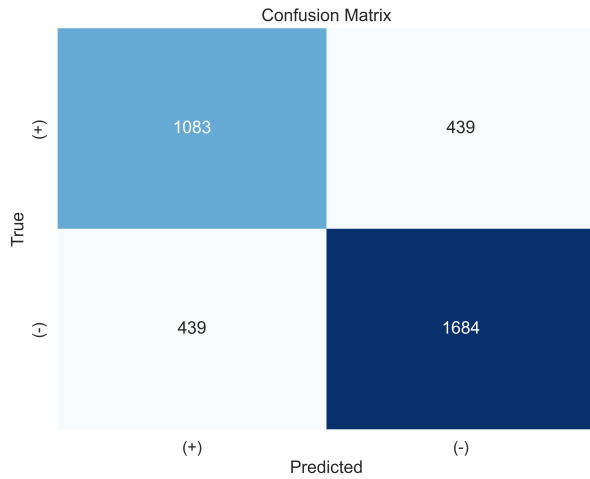
Bu tez çalışmasında veri setindeki cümleler MUSE modeli ile temsil edildikten sonra pozitif ve negatif sınıfı temsil eden iki cümle belirlenerek, veri setindeki cümleler bu cümlelerden hangisine yakınsa, benzerlik skorlarına bakılarak, pozitif veya negatif olarak sınıflandırılmıştır. Daha sonra yapılan tahminler ile gerçek değerler karşılaştırılarak 4 dil için Şekil 4.15’de verilen hata matrisi oluşturulmuş ve çeşitli metrikler ile model değerlendirilmiştir.



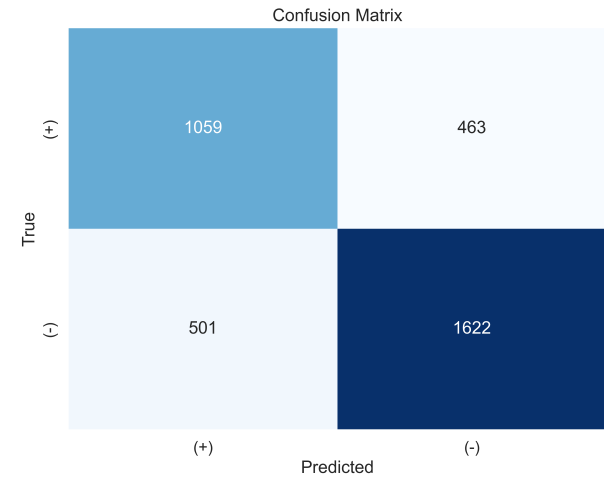
Hata matrisi Türkçe



Hata matrisi Felemenkçe



Hata matrisi İngilizce



Hata matrisi Almanca

Şekil 4.15 4 dil için hata matrisleri

Hata matrisi sınıflandırma problemlerinde modeli değerlendirmek amacıyla kullanılmaktadır. TP modelin gerçekte doğru olanı doğru olarak sınıflandırıldığını, TN ise gerçekte yanlış olanı yanlış olarak sınıflandırdığını ifade etmektedir. FP yani Tip-1 hata, gerçekte yanlış olan ama modelin doğru olarak sınıflandırdığı ifadelerdir. **Şekil 4.17**'de verilen örnekte, hamile olmayan bir kişinin hamile olarak sınıflandırıldığı görülmektedir. FN yani Tip-2 hata ise gerçekte doğru olan ama modelin yanlış olarak sınıflandırdığı ifadelerdir. **Şekil 4.17**'de hamile olan bir kişinin hamile değil olarak sınıflandırıldığı görülmektedir. **Şekil 4.15**'te yapılan sınıflandırma sonucunda elde edilen değerler 4 farklı dil için ayrı ayrı sunulmuştur (Kundu 2022).

		Gerçek	
		Var	Yok
Tahmin	Var	TP	FP
	Yok	FN	TN

Şekil 4.16 Hata matrisi



Şekil 4.17 Tip-1 ve Tip-2 hata örnekleri (Schmarzo 2018).

4.6 Değerlendirme

Sınıflandırma işleminden sonra modelin performansını değerlendirmek için literatürde yaygın olarak da kullanılan metrikler aşağıda verilmiştir (Altunkaynak 2022).

4.6.1 Doğruluk (Accuracy)

Yapılan tahminin ne kadarının doğru olarak sınıflandırıldığını belirlemek için bu metrik kullanılmaktadır. Hesaplamak için “doğru” sınıflandırılan ögelerin sayısı tüm ögelerin sayısına bölünmektedir (Köse 2019, akt. Aksu ve Karaman 2020).

$$Dogruluk = \frac{TP + TN}{TP + TN + FP + FN}$$

4.6.2 Kesinlik (Precision)

“Doğru” olan (P) ve “doğru” (T) şekilde tahmin edilen ögelerin, toplam pozitif ögelerin sayısına bölünmesi ile elde edilir. (Hamile olanlar / hamile olanlar ve hamile değilken hamile çıkanlar)

$$Kesinlik = \frac{TP}{TP + FP}$$

4.6.3 Duyarlılık (Recall)

Doğru olduğu bilinen pozitif ögelerin (TP) tüm doğru olduğu bilinen gözlemlere (TP ve FN) oranıdır. (Hamile olanları doğru tespit etme oranıdır, hamile olanlar hamile, hamile olmayanlar hamile değil çıkmalı)

$$Duyarlilik = \frac{TP}{TP + FN}$$

4.6.4 F-1 Ölçüsü

Duyarlılık ve kesinlik ölçülerinin harmonik ortalamasıdır.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN}$$

4.6.5 Dengeli Doğruluk

İkili ve çoklu sınıflandırma problemlerinde kullanılabilen bir metriktir. Duyarlılık ve özgüllük metriklerinin aritmetik ortalaması alınarak hesaplanır. Özgüllük (Specifity veya TN rate) model tarafından doğru şekilde sınıflandırılan gerçekte olan örneklerin oranıdır.

$$Ozgulluk = \frac{TN}{TN + FP} \quad \quad \quad BalancedAccuracy = \frac{duyarlilik + ozgulluk}{2}$$

Çizelge 4.2'de sınıflandırma yapıldıktan sonra oluşturulan hata matrisi kullanılarak hesaplanmış dengeli doğruluk değerleri verilmiştir. Farklı referans cümleler için farklı sınıflandırma sonuçları elde edilmiştir. “This sentence is rather positive/negative” cümlesi ile en yüksek skorlar elde edildiğinden diğer metrikler bu cümle kullanılarak yapılan sınıflandırma sonuçları ile hesaplanmıştır.

Çizelge 4.2 Farklı referans cümleler için sınıflandırma sonuçları

Referans Cümleler	Positive / Negative	This sentence is positive / negative	This sentence is rather positive, negative
Almanca	0.6703	0.6883	0.7299
İngilizce	0.6818	0.7098	0.7523
Felemenkçe	0.6529	0.6759	0.7179
Türkçe	0.6492	0.69131	0.7279

Bu tez çalışmasında veri setindeki cümleler MUSE modeli ile temsil edildikten sonra pozitif ve negatif sınıfı temsil eden iki cümle belirlenerek, “This sentence is rather positive/negative” , veri setindeki cümleler bu cümlelerden hangisine yakınsa, benzerlik skorlarına bakılarak, pozitif veya negatif olarak sınıflandırılmıştır. Daha sonra yapılan tahminler ile gerçek değerler karşılaştırılarak 4 dil için Şekil 4.15’de verilen hata matrisi oluşturulmuş ve Çizelge 4.3’te verilen model değerlendirme metriklerinin sonuçları elde edilmiştir. 4 farklı dilden ve 4 farklı domaininden oluşan veri seti farklı modeller geliştirilmeden sınıflandırılmış ve başarılı sonuçlar elde edilmiştir. Veri setine dahil olan diller arasında İngilizce dilinde en yüksek skorlar elde edilmiştir.

Çizelge 4.6.2 Model değerlendirme metrikleri karşılaştırması

	Türkçe	Felemenkçe	İngilizce	Almanca
Doğruluk	0.73	0.72	0.76	0.74
Kesinlik	0.67	0.65	0.71	0.68
Duyarlılık	0.70	0.71	0.71	0.70
F-1 Ölçüsü	0.68	0.68	0.71	0.69
Balanced Acc.	0.73	0.71	0.75	0.73

5. SONUÇ VE ÖNERİLER

Duygu analizi belirli kurum, kişi, hizmet, olay veya konu hakkında ifade edilen görüşleri analiz etmek, ortaya çıkarmak için makine öğrenmesi, doğal dil işleme, metin madenciliği tekniklerini kullanmayı içeren yöntemler bütünüdür. Çeşitli algoritma ve teknikler kullanılarak kitlelerin tutum ve görüşlerini belirlemek amaçlanmaktadır. Bu tez çalışmasında çok dilli duygu analizi uygulaması yapılmıştır. Çok dilli duygu analizi de birden fazla dil içeren metinlerin duygusal içeriğini anlamak için kullanılmaktadır. Herhangi bir konu veya ürün için dünyanın her yerinden görüş bildirebilme imkanı ortaya çok dilli veri yığınları çıkarmaktadır. Bunlardan bilgi çıkarmak da ihtiyaç haline gelmiştir.

Kocoń vd. (2021) tarafından MultiEmo veri seti kullanılarak duygu sınıflandırması yapılmıştır. Bu çalışmada XLM-RoBERTa, MultiFiT ve LASER+BiLSTM modelleri kullanılmıştır. 11 dilde yapılan çalışmada MultiFiT modeli ile en iyi sonuçlar elde edilmiştir. Ancak MultiFiT'in sadece 7 dili destekleyen bir model olması dezavantaj olarak görülmüştür. Ayrıca çevirinin kalitesinden ziyade, çevrilen metnin duygusunun değişmesi ilgi çeken bir nokta olarak paylaşılmıştır.

Korczyński ve Kocoń (2022) tarafından MultiEmo veri seti kullanılarak bir duygu analizi çalışması yapılmıştır. BERT gibi transformer modelleri kullanılarak doğal dil işleme alanında başarılı sonuçlar elde edilmiş olsa da hesaplama zamanı ve maliyetleri uygulamada zorluklara neden olmaktadır. Bununla başa çıkabilmek için çeşitli sıkıştırma yöntemleri önerilmektedir. Bu yöntemler kullanılarak modelin boyutlarının düşürülmesi ve hesaplama süresinin kısaltılması hedeflenmektedir. Bunları yaparken de sınıflandırma başarısı da göz ardı edilmemektedir. XtremeDistil metodu ile doküman seviyesinde %87 doğruluk oranı elde edilmiştir. Aynı yöntem ile cümle seviyesinde %78 doğruluk oranı elde edilmiştir.

AL-Smadi vd. (2023) MUSE ve havuzlanmış kapılı tekrarlayan birimler derin öğrenme modellerini kullanarak duygu analizi yapmıştır. Sonuçta duygu sınıflandırması görevi için %90.86 F-1 puanı elde edilmiştir.

Bu tez çalışmasında duygu analizi görevleri için oluşturulmuş MultiEmo veri seti kullanılarak çok dilli duygu analizi yapılmıştır. Sağlık, turizm, pazarlama ve eğitim alanlarında 4 dildeki kullanıcı görüşleri MUSE modeli kullanılarak vektörler ile temsil edilmiştir. Bilgisayarlar metinler ile işlem yapamadığından metinlerin sayısal forma dönüştürmek amacıyla bu işlem yapılmıştır. MUSE cümle düzeyinde temsil yapan önceden eğitilmiş bir derin öğrenme modelidir. Temsil işlemi ardından sınıflandırma yapmak için kosinüs benzerliği skorları kullanılmıştır. Belirlenen pozitif ve negatif sınıfı temsil eden cümlelere uzaklıklar belirlenerek hangisine yakınsa ona göre sınıflandırma yapılmıştır. Yapılan sınıflandırma sonucunda İngilizce dilinde 0.75, Türkçe dilinde 0.72 doğruluk skorları elde edilmiştir.

Türkçe dilinde çok dilli duygu sınıflandırması çalışmaları sınırlı sayıda olduğundan alana katkı sağlamak hedeflenmiştir. Bunun yanında her dil için ve her domain için ayrı ayrı model geliştirmek yerine derin öğrenme yaklaşımları kullanılarak başarılı sonuçlar elde edildiği görülmüştür. Sonraki çalışmalarda araştırmacılar cümle düzeyinde temsil yapan diğer derin öğrenme modellerini kullanarak alana katkı sağlayabilirler.

KAYNAKLAR

- Agüero-Torales, M. M., Salas, J. I. A. ve López-Herrera, A. G. 2021. Deep learning and multilingual sentiment analysis on social media data: An overview. *Applied Soft Computing*, (107).
- Akıncı, M. 2020. Web Sitesi: <https://starlangyazilim.com/ngram-nedir/>. Erişim Tarihi: 23 Nisan 2023.
- Aksu, Ç. 2020. Web Sitesi: <https://www.datascienceearth.com/veri-bilimi-ve-istatistik>. Erişim Tarihi: 7 Kasım 2022.
- Aksu, M. Ç. Ve Karaman, E. 2020. FastText ve kelime çantası kelime temsil yöntemlerinin turistik mekanlar için yapılan Türkçe incelemeler kullanılarak karşılaştırılması. *Avrupa Bilim ve Teknoloji Dergisi*, (20), 311-320.
- AL-Smadi, M., Hammad, M. M., Al-Zboon, S. A., AL-Tawalbeh, S. ve Cambria, E. 2023. Gated recurrent unit with multilingual universal sentence encoder for Arabic aspect-based sentiment analysis. *Know.-Based Syst.* 261, C (Feb 2023). <https://doi.org/10.1016/j.knosys.2021.107540>
- Altunkaynak, B. 2022. Veri Madenciliği Yöntemleri Ve R Uygulamaları Kavramlar Modeller Algoritmalar. Seçkin Yayıncılık, 294, Ankara.
- Anonim. 2022. Web Sitesi: <https://www.masplus.com.tr/big-data-nedir-hangi-sektorlerde-kullanilir/>. Erişim Tarihi: 13 Ocak 2023.
- Anonymous. 2022. Web Sitesi: <https://globalaihub.com/yapay-zeka-nedir/>. Erişim Tarihi 26 Ekim 2023.
- Anonymous. 2023a. Web Sitesi: <https://aws.amazon.com/tr/what-is/machine-translation/>. Erişim Tarihi : 2 Mart 2023.
- Anonymous. 2023b. Web Sitesi: <https://aws.amazon.com/tr/what-is/neural-network/>. Erişim Tarihi: 1 Temmuz 2023.
- Anonymous. 2023c. Web Sitesi: <https://deeplai.org/machine-learning-glossary-and-terms/n-gram>. Erişim Tarihi: 15 Ağustos 2023.
- Anonymous. 2023d. Web Sitesi: <https://www.cs.cmu.edu/~dst/WordEmbeddingDemo/tutorial.html>. Erişim Tarihi: 24 Ekim 2023.
- Anonymous. 2023e. Web Sitesi: <https://www.ibm.com/topics/machine-learning>. Erişim Tarihi: 30 Ekim 2023.

Anonymous. 2023f. Web Sitesi: <https://oracle.com/tr/artificial-intelligence/what-is-natural-language-processing/>. Eriřim Tarihi: 13 Haziran 2023.

Anonymous. 2023g. Web Sitesi: <https://www.oracle.com/tr/artificial-intelligence/machine-learning/what-is-deep-learning/>. Eriřim Tarihi: 8 Ağustos 2023.

Anonymous. 2023h. Web Sitesi: https://en.wikipedia.org/wiki/Machine_learning. Eriřim Tarihi: 30 Ekim 2023.

Aslan, E. 2019. Makine çevirisi. Hiperyayın, 176, İstanbul.

Atan, S. 2020. Metin madencilięi: imkanlar, yöntemler ve kısıtlar. Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, (31), 220-239.

Balahur, A. ve Turchi M. 2012. Multilingual sentiment analysis using machine translation. Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis. pp. 52–60.

Berber, B. Web Sitesi: <https://teknofesor.com/bag-of-words-modeli-nedir-nasil-calisir-2-ornek/>. Eriřim Tarihi: 18 Ağustos 2023.

Bhargava, R. ve Sharma, Y. 2017. MSATS: Multilingual sentiment analysis via text summarization, in Proc. 7th Int. Conf. Cloud Comput., Data Sci. Eng.-Confluence, Jan. 2017, pp. 71–76.

Burchfiel, A. 2022. Web Sitesi: <https://www.tokenex.com/blog/ab-what-isnlp-natural-language-processing-tokenization/>. Eriřim Tarihi: 21 Mayıs 2023.

Burger, J. M. 2016. Kiřilik. Kaknüs Yayınları, 796. İstanbul.

Büyükeke, A., Sökmen, A. ve Gencer, C. 2020. Metin madencilięi ve duygu analizi yöntemleri ile sosyal medya verilerinden rekabetçi avantaj elde etme: turizm sektöründe bir araştırma. Journal of Tourism and Gastronomy Studies, 8(1), 322-335.

Can, E. F., Ezen-Can, A. ve Can, F. 2018. Multilingual sentiment analysis: An RNN-based framework for limited data, arXiv:1806.04511.

Cansız, S. 2023. Web Sitesi: <https://www.udemy.com/course/r-veri-bilimi-istatistik-ve-makine-ogrenmesi-rstudio/learn/lecture/22033842#overview>. Eriřim Tarihi: 9 Kasım 2023.

Cebeci, H. 2020. Sosyal Medya Verileri ile Duygu Analizi. Mühendislikte Yapay Zeka Uygulamaları, (3), 191-211.

- Cerman, A. 2021. Web Sitesi: <https://medium.com/machine-learning-t%C3%BCrkiye/cosine-similarity-python-uygulaması-926ff395c47e#:~:text=Cosine%20Similarity%20veri%20biliminde%20yaygın,benzerliği%20büyük%20değil%20yönelim%20ölçüsür>. Erişim Tarihi: 20 Ekim 2023.
- Clarín. 2015. Web Sitesi: <https://clarin-pl.eu/index.php/en/multiemo-en/>. Erişim Tarihi: 15 Ocak 2023.
- Çağlayan Akay. 2020. Ekonometride büyük veri ve makine öğrenmesi: temel kavramlar. Der Yayınları, 208, İstanbul.
- Çelik, E., Dal, D. Ve Aydın, T. 2021. Duygu analizi için veri madenciliği sınıflandırma algoritmalarının karşılaştırılması. Avrupa Bilim ve Teknoloji Dergisi, (27), 880-889.
- Denecke, K. ve Reichenpfader, D. 2023. Sentiment analysis of clinical narratives: A scoping review. Journal of Biomedical Informatics, 140(2023), 104336.
- Draskovic, D., Zecevic, D. ve Nikolic, B. 2022. Development of a Multilingual Model for Machine Sentiment Analysis in the Serbian Language. Mathematics 2022, 10, 3236. <https://doi.org/10.3390/math10183236>.
- Durna, U. ve Demirel, Y. 2008. Bilgi Yönetiminde Bilgiyi Anlamak . Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi , 0 (30) , 129-156 . Retrieved from <https://dergipark.org.tr/tr/pub/erciyesiibd/issue/5887/77867>
- Edwards, K. 2021. Web Sitesi: <https://www.determ.com/blog/social-media-sentiment-analysis-in-politics/>. Erişim Tarihi: 28 Ekim 2023.
- El-Haj, M., Kruschwitz, U. Ve Fox, C. Creating language resources for under-resourced languages: Methodologies, and experiments with Arabic. Lang. Resour. Eval. 2015, 49, 549–580.
- Erkırtay, O. Ş. ve Ünal, C. 2021. Toplum Çevirmenliğinde Fikir Madenciliği ve Duygu Analizi. Manisa Celal Bayar Üniversitesi Sosyal Bilimler Dergisi, 19 (3), 168-185
- Es-sabery, F., Hair, A., Qadir, J., Sainz de Abajo, B., Garcia-Zapirain, B. Ve De la Torre Díez, I. 2021. Sentence-Level Classification Using Parallel Fuzzy Deep Learning Classifier. IEEE Access, vol. 9, pp. 17943-17985, doi: 10.1109/ACCESS.2021.3053917.
- Fuady, M. J. Ve Ibrahim, R. 2019. Multilingual Sentiment Analysis Social Media Disaster Data, The 6th International Conference on Electrical, Electronics and

- Information Engineering (ICEEIE 2019), Denpasar, Indonesia, 2019, pp. 269-272, doi: 10.1109/ICEEIE47180.2019.8981479.
- Goel, A. ve Gupta, L. 2020. Social media in the times of COVID-19. JCR: Journal of Clinical Rheumatology, 26(6), 220-223.
- Güldü, Ö. 2023. Web Sitesi: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://acikders.ankara.edu.tr/pluginfile.php/7654/mod_resource/content/0/KONU%204.pdf. Erişim Tarihi: 23 Mart 2023.
- Gürbüz, U. 2020. Web Sitesi: <https://www.modernbilgelik.com/post/maslow>. Erişim Tarihi: 30 Ekim 2023.
- Ivankov, A. 2023. Web Sitesi: <https://www.profolus.com/topics/advantages-disadvantages-artificial-neural-networks/>. Erişim Tarihi: 10 Kasım 2023.
- İnce, P. 2012. Web Sitesi: <https://www.digitalfabrika.com/blog/maslowun- ihtiyaclar-hiyerarshisi-uzerinden-sosyal-medya/>. Erişim Tarihi 11 Nisan 2023.
- Kaynar, O., Arslan, H., Görmez, Y. ve Demirkoparan, F. 2017. Feature selection methods in sentiment analysis," 2017 International Artificial Intelligence and Data Processing Symposium (IDAP), Malatya, Turkey, 2017, pp. 1-5, doi: 10.1109/IDAP.2017.8090187.
- Keskin, M. V. 2018. Büyük Veride Makine Öğrenmesi Uygulaması. Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Uygulamalı İstatistik Anabilim Dalı, 71, İstanbul.
- Kim, S. M. Ve Hovy. E. 2006. Identifying and analyzing judgment opinions. In Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (HLT-NAACL '06). Association for Computational Linguistics, USA, 200–207. <https://doi.org/10.3115/1220835.1220861>
- Ko, Y. 2012. A study of term weighting schemes using class information for text classification. In Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval (SIGIR '12). Association for Computing Machinery, New York, NY, USA, 1029–1030. <https://doi.org/10.1145/2348283.2348453>
- Kocoń, J., Miłkowski, P. ve Kanclerz. K. 2021. MultiEmo: Multilingual, Multilevel, Multidomain Sentiment Analysis Corpus of Consumer Reviews. In Computational Science – ICCS 2021: 21st International Conference, Krakow, Poland, June 16–18, 2021, Proceedings, Part II. Springer-Verlag, Berlin, Heidelberg, 297–312. https://doi.org/10.1007/978-3-030-77964-1_24

- Korczyński, W. ve Kocoń, J. 2022. Compression Methods for Transformers in Multidomain Sentiment Analysis," 2022 IEEE International Conference on Data Mining Workshops (ICDMW), Orlando, FL, USA, 2022, pp. 419-426, doi: 10.1109/ICDMW58026.2022.00062.
- Kou, F., Du, J., He, Y., ve Ye, L. 2016. Social network search based on semantic analysis and learning. CAAI transactions on intelligence technology, 1(4), 293-302.
- Kundu, R. 2022. Web sitesi: <https://www.v7labs.com/blog/confusion-matrix-guide>. Erişim Tarihi: 20 Eylül 2023.
- Lei, L. ve Liu, D. 2021. Conducting Sentiment Analysis. Cambridge University Press, 104, Online.
- Liu, B. 2020. Sentiment analysis: Mining opinions, sentiments, and emotions, Cambridge University Press. <https://doi.org/10.1017/9781108639286>
- Levi. 2023. Web sitesi: <https://twitter.com/levikul09/status/1695372473896714423>. Erişim Tarihi: 30 Ekim 2023.
- Mabokela, K. R., Çelik, T. Ve Raborife, M. 2023. Multilingual Sentiment Analysis for Under-Resourced Languages: A Systematic Review of the Landscape. IEEE Access, 11, 15996-16020
- Mboutayeb, S., Majda, A. ve Nikolov, N. (2021). Multilingual Sentiment Analysis: A Deep Learning Approach. 27-32. 10.5220/0010727700003101.
- Mercha, E. M. ve Benbrahim, H. 2023. Machine Learning and Deep Learning for sentiment analysis across languages: A survey. Neurocomputing. (531), 195-216. doi: <https://doi.org/10.1016/j.neucom.2023.02.015>
- Nabiyev, V. 2021. Yapay Zeka: Makine Öğrenmesi, Stratejili Oyunlar, Örüntü Tanıma, Doğal Dil İşleme. Seçkin Yayınları, 816, Ankara.
- Oğuzlar, A. ve Kızılkaya, Y. M. (2019). Metin madenciliğinde duygu analizi. Dora Yayıncılık, 134. Bursa.
- Öğündür, G. 2020. Web Sitesi: <https://medium.com/@gulcanogundur/pca-principal-component-analysis-temel-bile%C5%9Fenler-analizi-bf9098751c62>. Erişim Tarihi: 29 Ekim 2023.

Özdemir, E. 2022. Havayolu Yolcularının Çevrimiçi Değerlendirmelerinin Metin Madenciliği İle Analizi. Doktora Tezi, Anadolu Üniversitesi, Sosyal Bilimler Enstitüsü, Sivil Havacılık Yönetimi Anabilim Dalı, 121, Eskişehir.

Öztemel, E. 2006. Yapay Sinir Ağları. Papatya Yayıncılık, 232, İstanbul.

Pak, A. ve Paroubek, P. 2011. Twitter for Sentiment Analysis: When Language Resources are Not Available. 2011 22nd International Workshop on Database and Expert Systems Applications, Toulouse, France: IEEE, Ağu. 2011, ss. 111-115. doi: [10.1109/DEXA.2011.86](https://doi.org/10.1109/DEXA.2011.86).

Pervan Akman, N. 2020. Web Sitesi: <https://www.sisasoft.com.tr/duygu-analizi-neden-onemli#:~:text=Duygu%20Analizi%20Avantajlar%C4%B1&text=Duygu%20analizi%20modelleri%2C%20firmalar%C4%B1n%20i%C5%9F,yapma%20durum%20bertaraf%20etmek%20m%C3%BCmk%C3%BCnd%C3%BCr>.

Roldos, İ. 2020. Web Sitesi: <https://monkeylearn.com/blog/sentiment-analysis-deep-learning/>. Erişim Tarihi: 29 Ekim 2023.

Sadıkov, T. ve Sarıgül, K. 2021. Makine çeviri yöntemleri ve makine çevirisinin bugünkü durumu. Uluslararası Türkçe Edebiyat Kültür Eğitim (TEKE) Dergisi , 10 (1) , 192-205 . Retrieved from <https://dergipark.org.tr/tr/pub/teke/issue/60878/902717>.

Safalı, Y. ve Avaroğlu, E. 2021. Derin Öğrenme ile Yüz Tanıma ve Duygu Analizi. Avrupa Bilim ve Teknoloji Dergisi, (31), 764-770.

Salvaggio, D. 2019. Web Sitesi: <https://d-salvaggio.medium.com/multilingual-universal-sentence-encoder-muse-f8c9cd44f171>. Erişim Tarihi: 15 Ekim 2023.

Schmarzo, B. 2018. Web Sitesi: <https://www.linkedin.com/pulse/using-confusion-matrices-quantify-cost-being-wrong-bill-schmarzo/> Erişim Tarihi: 14 Haziran 2023.

Serrano, L. Web Sitesi: <https://txt.cohere.com/sentence-word-embeddings/>. Erişim Tarihi: 23 Ekim 2023

Statista. 2023. Web Sitesi: <https://www.statista.com/statistics/262946/most-common-languages-on-the-internet#:~:text=English%20as%20the%20leading%20online%20language>. Erişim Tarihi: 20 Ağustos 2023.

Şahinaslan, Ö, Dalyan, H. ve Şahinaslan, E. 2022. Naive Bayes Sınıflandırıcısı Kullanılarak YouTube Verileri Üzerinden Çok Dilli Duygu Analizi. International Journal of Informatics TechnologiesBilişim Teknolojileri Dergisi, (15), 221-229.

- Şeker, Ş. E. 2011. Web Sitesi: <https://bilgisayarkavramlari.com/2011/04/23/n-gram/>. Erişim Tarihi: 23 Nisan 2023.
- Şeker, S. E. 2016. Web Sitesi: https://ybsansiklopedi.com/wp-content/uploads/2016/09/duygu_analizi.pdf. Erişim Tarihi: 27 Ekim 2023.
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., ve Stede, M. 2011. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2), 267-307.
- Tokcaer, S. 2021. Türkçe metinlerde duygu analizi. *Journal of Yasar University*, 16(63), 1514-1534.
- TransPerfect. 2020. Web Sitesi: <https://www.transperfect.com/blog/value-social-listening>. Erişim Tarihi: 26 Ekim 2023.
- Uzun, E. 2021. Web Sitesi: <https://www.vudols.com/yapilandirilmis-veri-nedir/>. Erişim Tarihi: 23 Eylül 2023.
- We Are Social. 2023. Web Sitesi: <https://wearesocial.com/wp-content/uploads/2023/03/Digital-2023-Global-Overview-Report.pdf>. Erişim Tarihi: 23 Mart 2023.
- Yayla, M., Diyar Demirkol, M. ve Alqaraleh, S. 2021. CNN vs. LSTM for Turkish Text Classification, 2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA), Kocaeli, Turkey, 2021, pp. 1-6, doi: 10.1109/INISTA52262.2021.9548407.
- Yılmaz, B. 2023. Web Sitesi: <https://research.aimultiple.com/sentiment-analysis-challenges/>. Erişim Tarihi: 29 Ekim 2023.
- Yurtoğlu, H. 2005. Yapay Sinir Ağları Metodolojisi ile Öngörü Modellemesi: Bazı Makroekonomik Değişkenler İçin Türkiye Örneği. *Uzmanlık Tezi, Devlet Planlama Teşkilatı, Ekonomik Modeller ve Stratejik Araştırmalar Genel Müdürlüğü*, 104.
- Zülfikar, H. 2022. Web Sitesi: https://ansiklopedi.tubitak.gov.tr/ansiklopedi/veri_data. Erişim Tarihi: 14 Ekim 2022