YAPAY ZEKA DESTEKLİ TAKİPÇİ PROFİLLEME SİSTEMİ

Kadir Yunus Koç
20 14 02 210

YÜKSEK LİSANS TEZİ
Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Tezli Yüksek Lisans
Danışman: Dr. Öğr. Üyesi Erdal GÜVENOĞLU

İstanbul
T.C. Maltepe Üniversitesi
Lisansüstü Eğitim Enstitüsü
Ağustos, 2022

YAPAY ZEKA DESTEKLİ TAKİPÇİ PROFİLLEME SİSTEMİ

Kadir Yunus Koç
20 14 02 210
ORCID: 0000-0003-0604-2749

YÜKSEK LİSANS TEZİ
Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Tezli Yüksek Lisans
Danışman: Dr. Öğr. Üyesi Erdal GÜVENOĞLU

İstanbul
T.C. Maltepe Üniversitesi
Lisansüstü Eğitim Enstitüsü
Ağustos, 2022



JÜRİ VE ENSTİTÜ ONAYI

Bu belge, Yükseköğretim Kurulu tarafından 19.01.2021 tarihli "Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge" ile bildirilen 6689 Sayılı Kişisel Verilerin Korunması Kanunu kapsamında gizlenmiştir.

ETİK İLKE VE KURALLARA UYUM BEYANI

Bu belge, Yükseköğretim Kurulu tarafından 19.01.2021 tarihli "Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge" ile bildirilen 6689 Sayılı Kişisel Verilerin Korunması Kanunu kapsamında gizlenmiştir.

TEŞEKKÜR

Bu çalışmanın gerçekleştirilmesinde değerli bilgilerini bizlerle paylaşan, kullandığı her kelimenin hayatımıza kattığı önemini asla unutmayacağımız saygıdeğer hocam; Dr. Öğr. Üyesi Erdal GÜVENOĞLU'na ve beraber fikir alışverişinde bulunduğumuz sınıf arkadaşlarımıza sonsuz teşekkürlerimizi sunarız.

Kadir Yunus Koç
08/2022

ÖZ

Yapay Zeka Destekli Takipçi Profilleme Sistemi

Kadir Yunus Koç
Yüksek Lisans Tezi
Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Tezli Yüksek Lisans
Danışman: Dr. Öğr. Üyesi Erdal GÜVENOĞLU
Maltepe Üniversitesi Lisansüstü Eğitim Enstitüsü, 2022

Dünyada sosyal medya kullanımı gün geçtikçe artmaktadır. Sosyal medya alışkanlığının artması ile paylaşımda bulunulan metin ve görüntü verileri de orantılı bir şekilde artmaktadır. Dijital platformlar üzerinde biriken görüntü ve metin verileri yapılandırılmamış veriler olarak görülmektedir. Bu nedenle, yapılandırılmamış metni ve görüntüyü manuel olarak yönetmek zor olduğundan, metin ve görüntü içeriğinin otomatik olarak anlamlandırılması için yeni teknikler üretilmektedir. Bu proje kapsamında, yapay zekâ teknikleri kullanılarak sosyal medya verileri üzerinde (metin ve görüntü verileri) firmaların takipçilerinin analiz edilmesi ve takipçi profilinin çıkarılması hedeflenmektedir. Bu sayede firmalar takipçilerini tanıyabileceği gibi hedef kitlelerine doğru reklam stratejileri yöneltilebilecektir.

Anahtar Sözcükler: Nesne Tespiti, Metin Duygu Analizi, Konu Modelleme, Profilleme, Takipçi Analizi, Doğal Dil İşleme, Derin Öğrenme.

ABSTRACT

ARTIFICIAL INTELLIGENCE BASED FOLLOWER PROFILING SYSTEM

Kadir Yunus Koç
Master Thesis
Department of Computer Engineering
Computer Engineering Master's with Thesis
Thesis Advisor: Assoc. Prof. Dr. Üyesi Erdal GÜVENOĞLU
Maltepe University Graduate Education Institute, 2022

The use of social media in the world is increasing day by day. With the increase in social media habit, the text and image data shared increases proportionally. Image and text data accumulated on digital platforms are seen as unstructured data. Therefore, since it is difficult to manually manage unstructured text and image, new techniques are being produced to make sense of text and image content automatically. Within the scope of this project, it is aimed to analyze the followers of the companies on social media data (text and image data) using artificial intelligence techniques and to create the follower profile. In this way, companies will be able to recognize their followers and direct advertising strategies to their target audiences.

Keywords: Object Detection, Text Sentiment Analysis, Topic Modeling, Profiling, Follower Analysis, Natural Language Processing, Deep Learning.

İÇİNDEKİLER

JÜRİ VE ENSTİTÜ ONAYI i
ETİK İLKE VE KURALLARA UYUM BEYANI ii
TEŞEKKÜR iii
ÖZ iv
ABSTRACT v
İÇİNDEKİLER vi
ŞEKİLLER LİSTESİ ix
KISALTMALAR xi
1. GİRİŞ 1

1.1. Problem 1
1.2. Amaç 1
1.3. Önem 2
2. LİTERATÜR ARAŞTIRMası 4

2.1. Metin Sınıflandırma ve Kümeleme 4
2.2. Nesne Tespiti 5
2.3. Metin Duygu Tahmini 6
3. MATERYAL VE YÖNTEM 8

3.1. Metot ve Hesaplama 8
3.2. Veriler ve Toplanması 9
3.3. Nesne Tespiti Yöntemi 10
3.4. Metin Duygu Tahmini Yöntemi 10
3.5. Metin Konu Çıkarımı Yöntemi 10
3.6. A Tipi Veri Kullanılan Yöntemler 11

3.6.1. Veri Nesne Tespiti Yöntemi 11
3.6.2 Metin Duygu Tahmini Yöntemi 12
3.6.3 Metin Konu Çıkarımı Yöntemi 15

3.7. B Tipi Veri Kullanılan Yöntemler 17

3.7.1. Veri Nesne Tespiti Yöntemi 18
3.7.2 Metin Duygu Tahmini Yöntemi 19
3.7.3 Metin Konu Çıkarımı Yöntemi 22

3.8. C Tipi Veri Kullanılan Yöntemler 24

3.8.1. Veri Nesne Tespiti Yöntemi 25
3.8.2 Metin Duygu Tahmini Yöntemi 26

3.8.3 Metin Konu Çıkarımı Yöntemi 29
3.9. Verilerin Çözümlenmesi ve Yorumlanması 31
4. SONUÇ 33
KAYNAKLAR 35
ÖZGEÇMİŞ 38

TABLOLAR LİSTESİ

Tablo 1. Veri Tipleri..... 9

Tablo 2. Örnek Ham Veri..... 10

Tablo 3. A Tipi Veri Dil Dağılımı..... 11

Tablo 4. B Tipi Veri Dil Dağılımı 17

Tablo 5. C Tipi Veri Dil Dağılımı 24

ŞEKİLLER LİSTESİ

Şekil 1. Sistem Şeması 9

Şekil 2. A tipi pozitif twitter veri nesne tespiti tablosu bar grafiği 12

Şekil 3. A tipi veri Text Blob modeli pozitif twitter veri tablosu bar grafiği 13

Şekil 4. A tipi veri Text Blob modeli pozitif twitter veri tablosu scatter grafiği 13

Şekil 5. A tipi veri Vader modeli pozitif twitter veri tablosu bar grafiği 14

Şekil 6. A tipi veri Vader modeli pozitif twitter veri tablosu scatter grafiği 14

Şekil 7. A tipi veri LDA modeli 1 numaralı konu dağılımı 15

Şekil 8. A tipi veri LDA modeli 2 numaralı konu dağılımı 16

Şekil 9. A tipi veri LDA modeli 3 numaralı konu dağılımı 17

Şekil 10. B tipi pozitif twitter veri nesne tespiti tablosu bar grafiği..... 18

Şekil 11. B tipi veri Text Blob modeli pozitif twitter veri tablosu bar grafiği..... 19

Şekil 12. B tipi veri Text Blob modeli pozitif twitter veri tablosu scatter grafiği..... 20

Şekil 13. B tipi veri Vader modeli pozitif twitter veri tablosu bar grafiği 21

Şekil 14. B tipi veri Vader modeli pozitif twitter veri tablosu scatter grafiği..... 21

Şekil 15. B tipi veri LDA modeli 1 numaralı konu dağılımı..... 22

Şekil 16. B tipi veri LDA modeli 2 numaralı konu dağılımı..... 23

Şekil 17. B tipi veri LDA modeli 3 numaralı konu dağılımı..... 24

Şekil 18. C tipi pozitif twitter veri nesne tespiti tablosu bar grafiği..... 25

Şekil 19. C tipi veri Text Blob modeli pozitif twitter veri tablosu bar grafiği..... 26

Şekil 20. C tipi veri Text Blob modeli pozitif twitter veri tablosu scatter grafiği 27

Şekil 21. C tipi veri Vader modeli pozitif twitter veri tablosu bar grafiği 28

Şekil 22. C tipi veri Vader modeli pozitif twitter veri tablosu scatter grafiği..... 28

Şekil 23. C tipi veri LDA modeli 1 numaralı konu dağılımı..... 29

Şekil 24. C tipi veri LDA modeli 2 numaralı konu dağılımı..... 30

Şekil 25. C tipi veri LDA modeli 3 numaralı konu dağılımı..... 31

KISALTMALAR

NLP	:	Natural Language Processing(Doğal Dil İşleme)
ML	:	Machine Learning(Makine Öğrenmesi)
LDA	:	Latent Dirichlet Allocation(Gizli Dirichlet Tahsisi)
TF-IDF	:	Term Frequency — Inverse Document Frequency(Terim Frekansı — Ters Belge Frekansı)
Word2Vec	:	kelimeleri vektör uzayında ifade etmeye çalışan unsupervised (no labels) ve tahmin temelli(prediction-based) bir modeldir.
FastText	:	Hızlı Metin
BERT K-Means	:	Kümeleme
SVM	:	Support Vector Machine(Destek Vektör Makineleri)
LSTM	:	Long- Short Term Memory(Uzun-Kısa Süreli Bellek)
CNN	:	Convolutional Neural Network(Evrişimsel Sinir Ağı)
GRU-CNN	:	Gated Recurrent Units-Convolutional Neural Network(Kapılı Yinelemeli Üniteler-Evrişimsel Sinir Ağı)
GAN	:	Generative Adversarial Networks(Çekişmeli Üretici Ağlar)
RNN	:	Recurrent Neural Network(Tekrarlayan Sinir Ağı)
YOLO	:	You Only Look Once(Konvolüsyonel sinir ağları kullanarak nesne tespiti yapan bir algoritmadır)
TEXT BLOB	:	(Duygu Analizi)
VADER	:	(Duygu Analizi)
FLAIR	:	(Duygu Analizi)
PyTorch	:	Python tabanlı derin öğrenme kütüphanesi
Selenium Test Aracı	:	Python tabanlı web veri toplama aracı
Regex	:	Python tabanlı veri ön işleme kütüphanesi
langdetect	:	Python tabanlı veri dil dağılım hesaplama kütüphanesi

1. GİRİŞ

Günümüzde veri, firmalar için önemli bir konumda olmaktadır. Nesnelerin İnterneti ve büyük veri ile birlikte, toplanan veri sadece toplanma amacıyla değil, toplanan bu veriden başka katma değerler çıkarmak suretiyle de kullanılabilir duruma gelmektedir. Tez kapsamında Twitter dijital medya platformu üzeride firmaları takip eden takipçilerin metin ve görüntü verileri ile çalışılması hedeflenmektedir. Tez çalışmasında metin verileri üzerinde 3 farklı metin duygu tahmini modeli kullanmaktayız. Bu sayede pozitif anlam içeren metin verilerinin tek modele göre daha doğru tespit edilmesini sağlamaktayız. Pozitif metin verilerinde paylaşılan görüntüler üzerinde nesne tespiti modeli kullanmayı hedeflemekteyiz. Bu sayede pozitif metin verileri ve görüntülerden elde edilen nesneler arasında pozitif ilişki kurmaktayız.

1.1. Problem

Son zamanlarda gelişen bilişim sektörü alanında artarak devam eden verileri kullanarak faydalı bilgiler elde etmek firmaların büyük problemlerinden biri olarak göze çarpmaktadır. Firmalar sahip oldukları veriyi anlamlandırıp bu veriden kar sağlama yoluna gitmektedirler. Veri tabanlarında büyük verilerden anlamlı bilgilere ulaşmak için algoritmalar ve metotlar geliştirilmiştir. Veri Madenciliği algoritmaları örnek olarak verilebilir. Bu tez çalışmasında sosyal medya üzerindeki takipçi kitlesinin görüntü ve metin verilerinin anlamlandırılarak takipçi kitlesine hitap debilecek reklam tercihi sunulması aşamasında ek fayda sağlaması hedeflenmektedir.

1.2. Amaç

Verinin gerçek değeri, denizde yüzen buzdağının görünen parçası olarak tanımlanabilir. Yalnızca ufak bir kısmı görünen bu buzdağının, büyük bölümü deniz seviyesinin altında gizlidir. İnovatif kurum ve kuruluşlar, doğaları gereği bu değeri ortaya çıkarabilmektedirler (Mayer, 2013). Buzdağının altında kalan verinin, yapılan iş ve hizmetler sonucunda oluştuğu ve bunların değerlendirmeye tabi tutulmadığı takdirde yaratılabilecek katma değerden kurum ve kuruluşlar yoksun kalacaktır. Bunun sonucunda da geleceğe yönelik stratejilerini oluştururken bir şeylerin eksik kalacağı, eksik kalan ve görünmeyen bu değerin ise kurum ve kuruluşun varlığı için hayati önem taşıyabileceği

yapılan çalışmalarda ortaya çıkmıştır. Bu durum verinin opsiyon değerinin önemini ortaya çıkarmaktadır. İşletmelerdeki bu sorun müşteri memnuniyetini anlamak açısından büyük bir zorluk teşkil etmektedir.

1.3. Önem

Türkiye'de sosyal medya kullanım alışkanlıklarının gün geçtikçe arttığı bilinmektedir. Son zamanlarda bu alışkanlıkların hangi düzeyde olduğunun tespiti için Next Generation Araştırma şirketi çalışma gerçekleştirdi. 2020 Nisan ayı içerisinde 15 yaş üzeri 2075 kişinin katılımıyla Türkiye genelinde araştırma gerekleştirildi. Araştırma sonuçlarına baktığımızda 20 kişiden 19'u sosyal medya kullanmaktadır. Sosyal medyada, kimimiz düşüncelerini paylaşmak, kimimiz ise yaşantısını fotoğraf veya video yardımı ile paylaşabileceği sosyal medya kanallarına yönelmektedir. Araştırmaya sonuçlarına baktığımızda en sık kullanılan sosyal medya kanalları arasında %58 oranla Instagram büyük bir paya sahiptir. Kullanıcıların %61'i boş zamanını değerlendirmek, %57'si eğlenmek ve %56'sı yeni şeyler öğrenmek için sosyal medya platformlarını kullanmaktadır. Bu bilgiler göz önünde bulundurulduğunda kullanıcıların hobi ve ilgi alanlarını tespit etmek ve bu alanları kullanarak kullanıcıları profillemek için sosyal medya platformlarının kullanılabileceği açıkça görülmektedir (NG Araştırma, 2020). Sosyal medya kullanımının artışı ile birlikte metin ve görüntü verileri de artmaktadır. Kullanıcı hesapları, kullanıcıyı tanımlama ve kullanıcı kitlelerini profillemek için uygun bir kaynak oluşturmaktadır. Firmalar için bu verilerin işlenmesi ve anlamlandırılması, firmaların müşterilerine uygun paylaşımları yapabilmesi adımında önemli bir hal almış durumdadır. Firmalar, kendi kitlelerine uygun reklam stratejileri geliştirmek için takipçilerini analiz edebilmeli ve takipçi profillerini çıkarmalıdır. Bu sayede pazar faaliyetlerinde bir artış gözlemleneceği öngörülmektedir.

Firmalar dijital reklam ajansları üzerinden sosyal medya platformlarında (Instagram, Twitter, Facebook vs.) paylaşım yapmaktadır. Bu tez kapsamında geliştirilmesi planlanan sistem, firmaların sosyal medya hesaplarını yöneten dijital reklam ajanslarında kolaylıkla kullanılabilecek ve birçok işi otomatik hale getirecektir. Deloitte araştırmasına göre 2020 yılında medya yatırımları içerisinde de dijital reklam ajanslarının payının %52 (~300 Milyar Dolar) olduğu gözlemlenmiştir (Deloitte, 2021). Bu maliyeti en verimli bir şekilde kullanmak için müşteri profilinin iyi analiz edilip müşteri profiline uygun reklam tercihi

sunulması oldukça önemlidir. Tez kapsamında, firmaların takipçi kitlelerini profillemek için sosyal medya platformlarından elde edilen veriler makine öğrenmesi ve derin öğrenme teknikleri ile işlenecektir. Bu sayede reklamın hitap edeceği kitlenin analizi yapılmış olup, kitleye uygun reklam stratejisi sağlanmış olmaktadır.

Bu tez kapsamında ortaya çıkacak ürünün, dijital reklam ajansları tarafından kullanılması öngörülmektedir. Bu ürünü kullanan dijital reklam ajansları, sosyal medya üzerinden gerçekleştirecekleri reklam ve paylaşımlarda takipçilerine uygun reklam stratejileri geliştirebilecektir. Dijital reklam ajansları, Twitter üzerinden metin ağırlıklı ve İnstagram üzerinden görüntü ağırlıklı reklam ve paylaşımlar yapmaktadır. Bu nedenle, tez çalışmasında sunulacak prototip çalışmasının Twitter üzerinden gerçekleştirilmesi planlanmaktadır.

Nesne Tespiti aşamasında yolo kullanılması hedeflenmektedir. 80 sınıf kapasitesi ile en geniş sınıflandırma modeli olan yolo5x modeli "%90" üzerinde başarım oranı sağlamaktadır. Metinsel verilerin anlamsal analizinde Text Blob freamework kullanılması hedeflenmektedir. Bu sayede metinlerin pozitif olma durumu tespit edilmesi sağlanmaktadır. Bu tez kapsamında geleneksel profilleme sistemi; hesap üzerinde en çok kullanılan takipçi özellikleri (lokasyon, yaş aralığı, beğeni durumları vs.) yapılmasıda hedeflenmektedir.

Tez raporunun ikinci bölümünde literatürde kullanılan nesne tespiti, metin duygu tahmini ve metin sınıflandırma modelleri hakkında bilgi verilmektedir. Tez raporunun üçüncü bölümünde literatür araştırması sonucu seçmiş olduğumuz modelleri Twitter üzerinden toplamış olduğumuz verilerde uygulamaktayız.

2. LİTERATÜR ARAŞTIRMASI

Tez kapsamında ortaya çıkacak ürünün, dijital reklam ajansları tarafından kullanılması öngörülmektedir. Bu ürünü kullanan dijital reklam ajansları, sosyal medya üzerinden gerçekleştirecekleri reklam ve paylaşımlarda takipçilerine uygun reklam stratejileri geliştirebilecektir. Dijital reklam ajansları, Twitter üzerinden metin ağırlıklı ve İnstagram üzerinden görüntü ağırlıklı reklam ve paylaşımlar yapmaktadır.

Büyük veri analitiğinde, tüm sisteminde başarıya ulaşmasında en etkili aşamalardan biri de veriden doğru içgörülerin elde edilebilmesi için ileri analitik yöntemlerin daha özelinde veriye dayalı yapay zeka modellerinin doğru belirlenmesidir.

2.1. Metin Sınıflandırma ve Kümeleme

Metin sınıflandırma ve Kümeleme aşamasında literatürdeki çalışmalara bakıldığında kullanılan yöntemlerin LDA (Huang, 2021; Wang, 2022, Luo, 2019), TF-IDF (Cahyani, 2021), Word2Vec (Cahyani, 2021), Doc2Vec (Dogru, 2021), FastText (Dharma, 2022), GloVe (Dharma, 2022), BERT K-Means (Tunalı, 2011; Bougteb, 2019), SVM (Bhavani, 2021) ve LSTM (Zhang, 2021) tabanlı yöntemler sıklıkla kullanıldığı görülmüştür. LSTM Model, kelime gömerek metni düşük boyutlu kelime vektörlerine dönüştürür ve metnin yerel özelliklerini çıkarmak için CNN'yi kullanır. LSTM'nin metin dizilerindeki tarihsel bilgileri koruma özellikleriyle birleştiğinde, CNN'nin bağlamla ilgili anlambilimi çıkarmadaki eksikliğini gidermektedir (Zhang, 2021). K-Means model döngüsü içerisinde verilerin kümelere benzerliklerine göre bir teknik geliştirilmiştir. Bu teknik belli ölçüde birden çok kümeye dahil olmasına izin verildiği özgün bir algoritma olan Çoklu Küme Küresel K-Means algoritması geliştirilmiştir. Farklı metinsel veri setleri üzerinde gerçekleştirilen denemeler ile yeni algoritma tekniği işlemci zamanı kullanımında belirgin bir fark oluşturmadan kümeleme kalitesinde büyük bir fayda sağlamakta, temel anlamda kullanılan Küresel K-Means algoritmasının ölçeklenebilirliğini koruyarak büyük doküman koleksiyonları üzerinde uygulanabilir olduğu gösterilmiştir (Tunalı, 2011).

Word2Vec, GloVe ve FastText metin duygu sınıflandırma modeli önermektedir. Bu çalışma, Word2Vec, GloVe ve FastText olmak üzere 3 kelime gömme yönteminin

Konvolüsyonlu Sinir Ağı algoritması kullanılarak metin sınıflandırmasındaki doğruluklarını karşılaştırmayı amaçlamaktadır. Bu çalışmada en iyi performansı FastText modeli kullanılarak elde edilmektedir (Dharma, 2022).

TF-IDF ve Word2Vec metin duygu sınıflandırma modeli önermektedir. Bu çalışma, duygusal metin sınıflandırmasındaki özellikleri temsil etmek için TF-IDF ve Word2Vec modellerinin performansını karşılaştırır. TF-IDF modellemesinin Word2Vec modellemesinden daha iyi performansa sahip olduğunu göstermektedir ve bu çalışma, önceki çalışmalara kıyasla sınıflandırma performansı sonuçlarını iyileştirmektedir (Cahyani, 2021).

LDA ve GRU-CNN modeli bir metin duygu analizi yöntemi önermektedir. LDA ve GRU-CNN'yi birleştirir. Her metin temsil edilir bir kelime alanı yerine gizli bir konu alanı ile ve CNN tarafından uzun vadeli dizileri işleme yeteneği, yüksek düzeyde bir sonuç elde etmek için GRU'nun birleştirilmesiyle geliştirilmiştir (Luo, 2019).

Makine öğrenimi algoritmalarını kullanarak metin sınıflandırmasını gözden geçiren bu çalışmada. Farklı makine öğrenimi algoritmalarının karşılaştırılması, SVM hesaplama için daha fazla doğruluk ve daha kısa zamanda hesaplama yapabildiği görülmüştür. Hiyerarşik sınıflandırma türü için karar ağacı kullanılmış ve KNN, K- değeri doğru verildiğinde daha fazla doğruluk verdiği görülmüştür (Bhavani, 2021).

2.2. Nesne Tespiti

Nesne tespiti aşamasında literatürdeki çalışmalara bakıldığında kullanılan yöntemlerin GAN (Bashir, 2021), CNN (Algur, 2018; Behrendt, 2017; Fang, 2019), R-CNN (Kristo, 2020), RNN (Hanbay, 2017) ve LSTM (iao, 2021) tabanlı yöntemler sıklıkla kullanıldığı görülmektedir.

YOLO VE CNN hibrit modelinde dış mekan görüntülerinde tespit edilen insan hareketlerinin otomatik sınıflandırılması için kullanılmıştır. Bu modelde hibrit bir derin öğrenme yaklaşımı önerilmiştir. YOLO nesne tespit sistemi ile hibrit bir şekilde kullanılarak görüntü içerisindeki insan hareketlerinin otomatik olarak tanınması sağlanmaktadır (Algur, 2018).

YOLO modeli CNN tabanlı bir sınıflandırma yöntemi önermektedir. Bu yazıda Tiny-YOLO-V3'ten türetilen Tinier-YOLO'nun, gelişmiş algılama doğruluğu ve gerçek zamanlı performans elde ederken model boyutunu daha da küçültmesi önerilmiştir. Tinier-YOLO'daki yoğun bağlantılar, özellik yayılımı güçlendirildiği ve ağda maksimum bilgi akışı sağlandığı için algılama doğruluğunun ve gerçek zamanlı performansın iyileştirilmesine yardımcı olduğu gözlenmiştir (Fang, 2019).

Behrendt ve arkadaşları yapmış oldukları çalışmada, zaman bazlı trafik ışıklarının tespiti ve takibi için derin öğrenme tabanlı model geliştirmiştir. Bu model YouOnlyLookOnce (YOLO) yöntemine dayanarak trafik ışıklarını tespit ve takip için farklı olan 2 tane sinir ağı kullanılmaktadır (Behrendt, 2017).

2.3. Metin Duygu Tahmini

Günümüzde gittikçe artan sosyal medya kullanımı ve veri madenciliğinin temel ilkelerinden birisi olan "yenilenebilir ve kullanabilir veriden anlamlı ifadeler çıkarma" düşüncesi son günlerde önemli araştırma konularından birisi olan 'Duygu Analizi'(Sentiment Analysis)'a olan popülerliği arttırmış ve beraberinde kullanım alanlarını genişletmiştir. Bu alanda yapılmış olan çalışmalara bakıldığında Text Blob (Guijar, 2021; Chaudhri, 2021), vader (Elbagir, 2019, Hutto, 2014) ve flair (Bharadi, 2021) gibi apilerin sıklıkla kullanıldığı görülmektedir.

Praveen ve Kumar yaptıkları çalışmada, iş zekası geliştirme alanında Text Blob apisini kullanmaktadır. Yaptıkları çalışma sonucunda;

- Önerilen teknik, iş zekasında oldukça uygulanabilir.
- Önerilen teknik, Kullanıcı ilgisine göre Reklamın yerleştirilmesinde kullanılabilir.
- Önerilen teknik, karar vericilerin ürün ve hizmet kıyaslamasına sahip olmasına yardımcı olabilir.

olduğu gözlenmektedir (Guijar, 2021).

Elbagir ve Yang yaptıkları çalışmada, tweet'leri analiz etmek için bir çoklu sınıflandırma sistemi önermektedir. 2016 ABD seçimleriyle ilgili tweetleri sınıflandırmak için VADER

kullanmaktadır. Sonuçlara bakıldığında, üçlü ve çoklu sınıfları tespit etmede iyi bir doğruluk göstermektedir (Elbagir, 2019).

Bharadi, Poshe, Yadav ve Nandiwadekar yaptıkları çalışmada, drama, romantizm, korku ve aksiyon gibi film türlerini tahmin eden ve aynı zamanda Twitter sosyal ağ web sitesindeki canlı tweet verilerine dayanarak popülerliklerini analiz eden bir model oluşturmak için NLP ve Derin öğrenme yaklaşımını birleştirmektedir. Flair çerçevesi ile birlikte Python Programlamayı kullanmaktadır. Flair çerçevesi, ön işleme ve sınıflandırma amacıyla birçok işlevsellik ve kitaplık sağlayan son teknoloji bir NLP kitaplığı olmaktadır. Diğerlerinden daha iyi sonuç ve gelişmiş doğruluk sağlamaktadır (Bharadi, 2021).

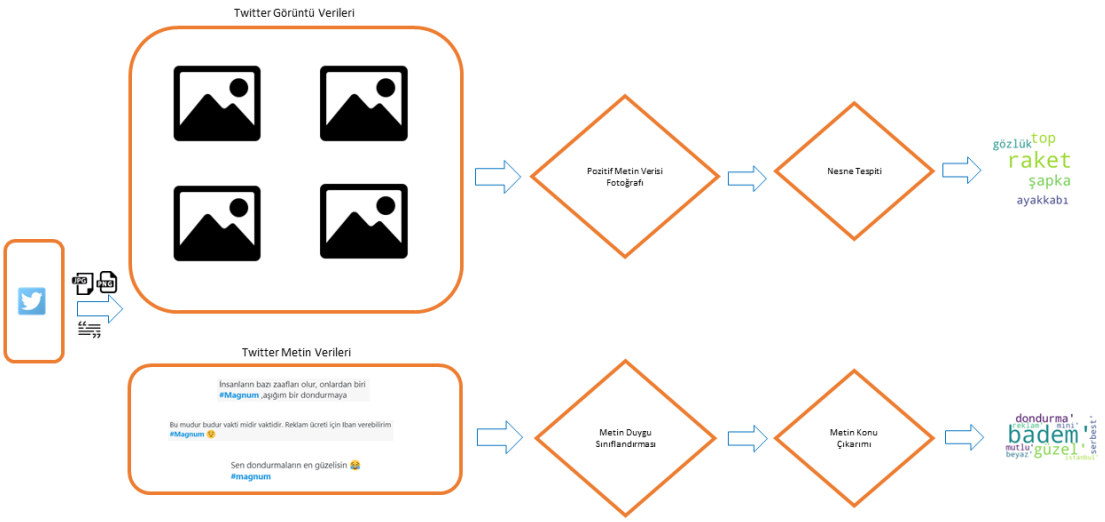
Literatürde bulunan yöntemlerin uygulanacağı ve yüksek başarım sağlanması için gerekli noktalarda hibrit modellerin tasarlanacağı hedeflenmektedir. Dünyada sosyal medya kullanımı gün geçtikçe artmaktadır. Sosyal medya alışkanlığının artması ile paylaşımda bulunulan metin ve görüntü verileri de orantılı bir şekilde artmaktadır. Dijital platformlar üzerinde biriken görüntü ve metin verileri yapılandırılmamış veriler olarak görülmektedir. Bu nedenle, yapılandırılmamış metni ve görüntüyü manuel olarak yönetmek zor olduğundan, metin ve görüntü içeriğinin otomatik olarak anlamlandırılması için yeni teknikler üretilmektedir. Bu tez kapsamında, yapay zekâ teknikleri kullanılarak sosyal medya verileri üzerinde (metin ve görüntü verileri) firmaların takipçilerinin analiz edilmesi ve takipçi profilinin çıkarılması hedeflenmektedir. Bu sayede firmalar takipçilerini tanıyabileceği gibi hedef kitlelerine doğru reklam stratejileri yöneltmesi sağlanmaktadır.

3. MATERYAL VE YÖNTEM

Bu tez çalışmasında Selenium test aracı kullanılarak 3 farklı tipte twitter verileri kullanılmaktadır. Bu veriler üzerinde nesne tespiti, metin duygu tahmini ve metin konu çıkarımı metodu kullanılmaktadır.

3.1. Metot ve Hesaplama

Bu tez kapsamında kullanılması hedeflenen veriler sosyal medya platformu olan twitter verileridir. Twitter verileri Selenium test aracı ve Python Twitter apileri ile toplanmaktadır. Toplanan ham Twitter verileri Python Regex kütüphanesi kullanılarak sembol ifadelerden temizlenmektedir. Sembol ifadelerden temizlenmiş veri Google translate kullanılarak İngilizceye çevrilmektedir. İngilizceye çevrilen veri üzerinde 3 farklı metin duygu analizi modeli kullanılmaktadır. Veri üzerinde kullanılan metin duygu analizi modellerinin sonucunda veri incelenerek eşik değeri 0,1 değerinin üzerinde olan verilere pozitif, eşik değeri -0,1 ve 0,1 arasında olan verilere nötr ve eşik değeri -0,1 değerinin altında olan veriler negatif metin verileri olarak sınıflandırılmaktadır. Metin duygu tahmini modellerinin 0,1 eşik değeri üzerinde bulduğu metin verileri ve görüntü verileri seçilerek nesne tespiti ve konu çıkarımı aşamasında kullanılmaktadır. Nesne tespiti aşamasında seçilen pozitif Twitter verilerinde bulunan görüntü verileri üzerinde nesne tespiti işlemi gerçekleştirilmektedir. Bu sayede pozitif metin verilerinde paylaşılan görüntü verilerinden elde edilen nesneler hakkında kullanıcı adına pozitif bir etkisi vardır yorumu yapılabilmektedir. Konu çıkarımı aşamasında pozitif metin verileri seçilerek LDA modeline verilmektedir. Bu sayede metin verilerinde sıklıkla kullanılan kelimelerin gruplanması ve analiz edilebilmesi sağlanmaktadır. Sistem şeması Şekil1'de gösterilmektedir.



Şekil 1. Sistem Şeması

3.2. Veriler ve Toplanması

Bu tez kapsamında kullanılması hedeflenen veriler sosyal medya platformu olan twitter verileridir. Bu veriler 3 kısımda Tablo1'de gösterilmektedir.

Tablo 1. Veri Tipleri

Veri Tipi	Veri Açıklaması	Kullanıcı Sayısı	Tweet Sayısı
A	Firmayı takip eden kullanıcıların twitter verileri	11287	125504
B	Twitter üzerinde aratılan etiket(hashtag) ile toplanan veriler	610	1002
C	Firmanın beğendiği twitter verileri	857	1208

Tablo1'de gösterilen veri tipleri Python Selenium sanal botu ve twitter api kütüphaneleri kullanılarak toplanmaktadır. Tablo2'de örnek ham veri gösterilmektedir.

Tablo 2. Örnek Ham Veri

Resim linki	Tweet
######.jpeg	Bu dondurmayı çok seviyorum .)
######.jpeg	< En sevdiğim spor >
######.jpeg	Tavsiye ediyorum deneyin <3

3.3. Nesne Tespiti Yöntemi

Tez çalışmasında kullanılan nesne tespiti yöntemi PyTorch kütüphanesi üzerinde çalışan Yolov5x modelidir. Bu model Jocher tarafından yayınlanan CNN tabanlı bir nesne tespit modelidir. Yolov5x modelinin sunmuş olduğu api desteği ile resimlerdeki nesneleri hızlı bir şekilde tespit etmemize olanak tanımaktadır. Yolov5x modeli ile A, B ve C tipi verilerinde bulunan resimler üzerinde nesne tespiti işlemi gerçekleştirilmiştir.

3.4. Metin Duygu Tahmini Yöntemi

Twitter verileri üzerinde metin duygu tahmini modelleri kullanılmıştır. Bu modeller Text Blob, Vader ve Flair modelleri olmaktadır. Bu modeller metin ile görüntülerden elde edilen nesnelerin arasında bir ilişki yakalamak için A, B ve C tipi verilerinde kullanılmaktadır.

3.5. Metin Konu Çıkarımı Yöntemi

Tez kapsamında twitter verileri üzerinde kullandığım metin konu çıkarımı modeli kullanılmıştır. Bu model LDA modeli olmaktadır. Bu model A, B ve C tipi verilerinde metinsel verilerden konu çıkarımı aşamasında kullanılmıştır. Konu çıkarımı yapılırken sentiment analizi doğrultusunda belirlenen pozitif twitter verileri üzerinde uygulanmıştır.

3.6. A Tipi Veri Kullanılan Yöntemler

A tipi veri firmayı takip eden kullanıcıların twitter verileri, bu verilerde 11287 kullanıcı ve 125504 twitter verisi bulunmaktadır. Bu veriler üzerinde Nesne tespiti yöntemi, Metin duygu tahmini yöntemi, Metin konu çıkarımı yöntemi kullanılmaktadır. A tipi veride langdetect kütüphanesi ile ham veri üzerinde dil dağılımı hesaplanmaktadır. Tablo3'te ham veri üzerindeki dil dağılımı görülmektedir.

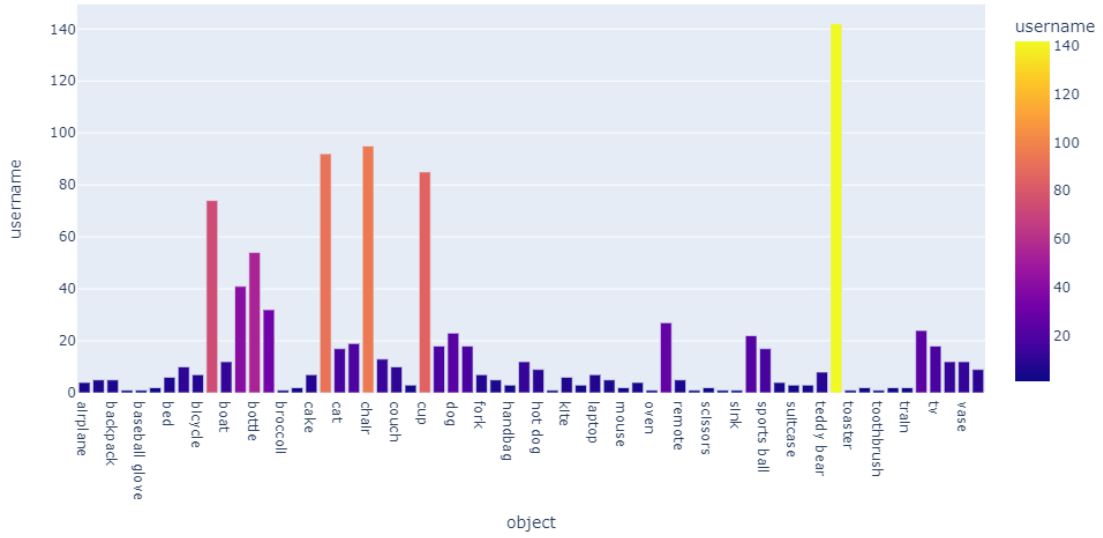
Tablo 3. A Tipi Veri Dil Dağılımı

Dil	Metin Sayısı
Türkçe	85461
İngilizce	6653
Arapça	2172
Başka Diller	31218

Tablo3'te A tipi veri dil dağılımı görülmektedir. A tipi veri üzerinde Türkçe 85461, İngilizce 6653, Arapça 2172 ve başka diller 31218 olacak şekilde dağılım göstermektedir.

3.6.1. Veri Nesne Tespiti Yöntemi

A tipi veride 125504 adet twitter verisi bulunmaktadır. İçerisinde resim olan 37272 twitter verilerinden 81 adet farklı nesne tespiti gerçekleştirilmiştir. Gerçekleştirilen nesne tespiti işleminde eşik değeri olarak 0.65 seçilmiştir. Twitter verisi üzerinde kullanmış olduğumuz metin duygu tahmini yöntemleri ile twitter verileri pozitif ve negatif olarak iki sınıfa ayrılmıştır. Böylelikle resimli twitter verilerinde tespit edilen nesnelerin kullanıcı kitlesi için olumlu yada olumsuz olarak hangi anlam içerdiği belirlenmiştir. Şekil 2'de A tipi twitter verisinde elde edilen nesne tespitleri gösterilmiştir.



Şekil 2. A tipi pozitif twitter veri nesne tespiti tablosu bar grafiği

Şekil 2'de A tipi twitter veri nesne tespiti tablosunda 142 Kravat, 95 Sandalye, 92 Araba 85 Kupa Bardak vb. çıkarımlar yapılmaktadır. Yapılan bu çıkarımlar Metin duygu tahmini yöntemleri ile belirlenen pozitif twitter verileri üzerinden çıkarılmıştır. Bu sayede fotoğraflardan elde edilen nesneler kullanıcılar için pozitif anlama gelmektedir.

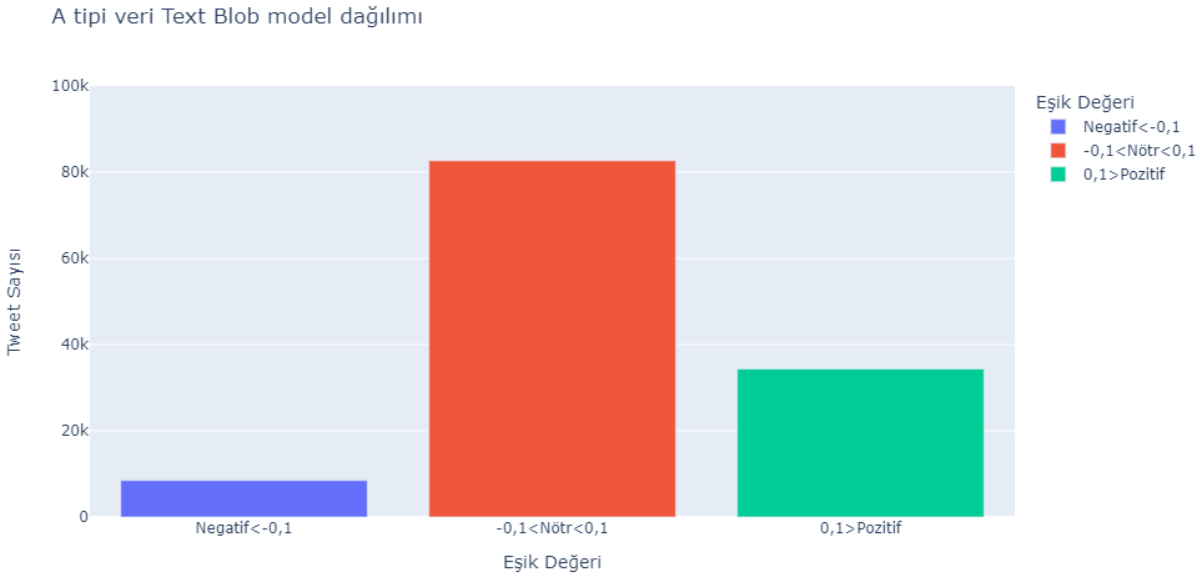
3.6.2 Metin Duygu Tahmini Yöntemi

A tipi veride 125504 adet twitter verisi bulunmaktadır. Twitter verileri üzerinde Text Blob, Vader ve Flair modelleri ile duygu tahmini yapılmıştır. Bu modeller ile ilgili çalışmaların sonuçları aşağıda verilmektedir.

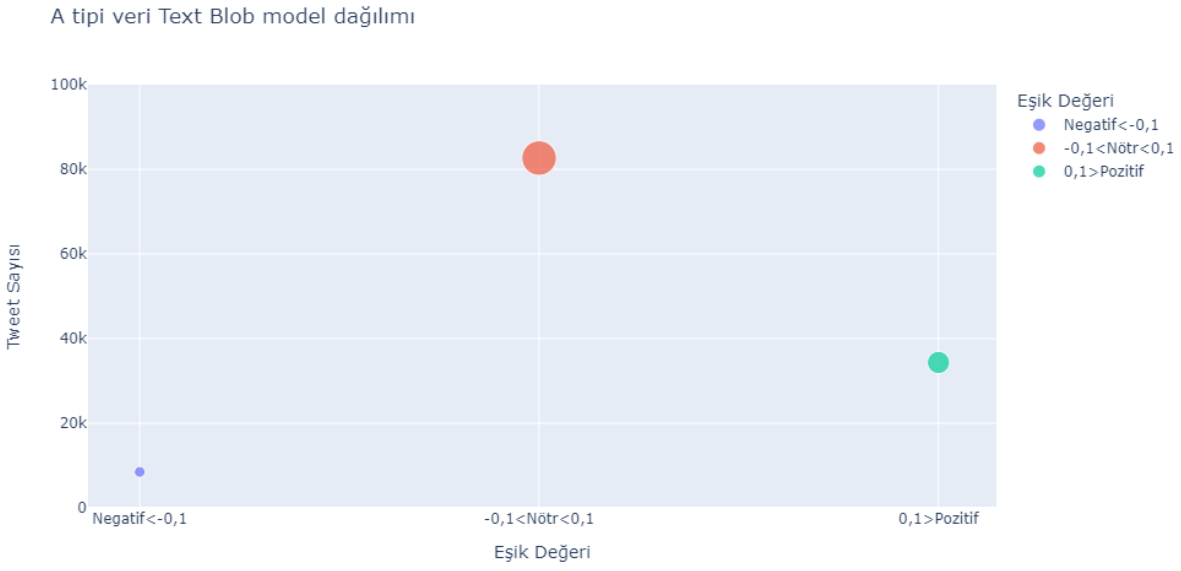
Tez çalışmasında Text Blob modeli için 125504 twitter verisi kullanılmıştır. Model sonucunda twitter verileri 3 sınıfta incelenmiştir.

- Pozitif: Polarity >= 0,1
- Nötr: Polarity >= -0,1 ve Polarity < 0,1
- Negatif: Polarity < -0,1

Sınıflandırma sonucunda 34355 twitter verisi pozitif, 82660 nötr ve 8489 negatif olduğu gözlemlenmiştir. Elde edilen sonuçlar Şekil 3'te Bar grafiği ve Şekil 4'te Scatter grafiği verilmektedir.



Şekil 3. A tipi veri Text Blob modeli pozitif twitter veri tablosu bar grafiği



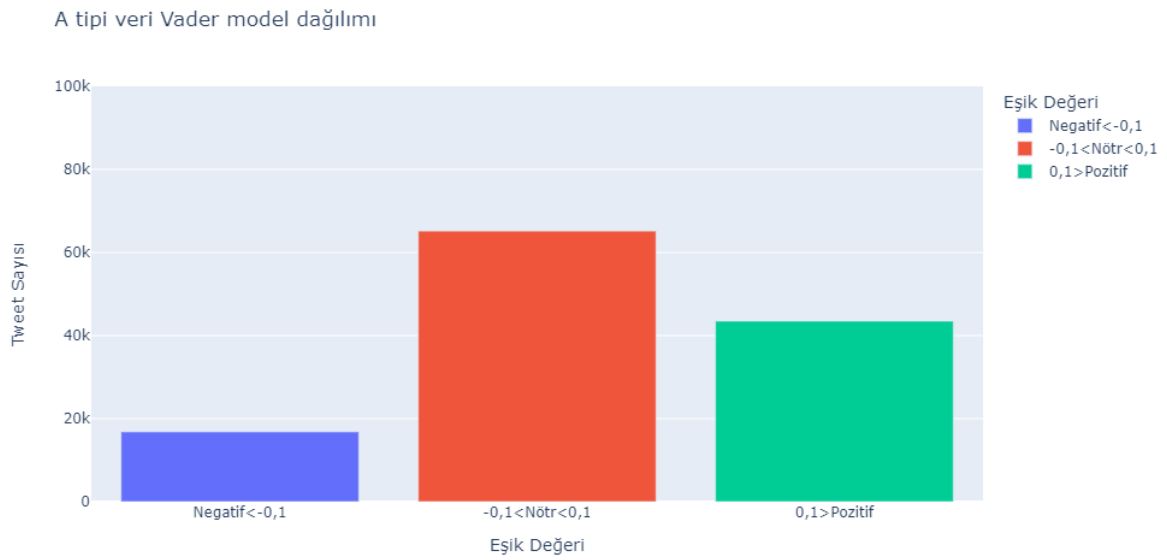
Şekil 4. A tipi veri Text Blob modeli pozitif twitter veri tablosu scatter grafiği

Vader modeli için 125504 twitter verisi kullanılmıştır. Model sonucunda twitter verileri 3 sınıfta incelenmiştir.

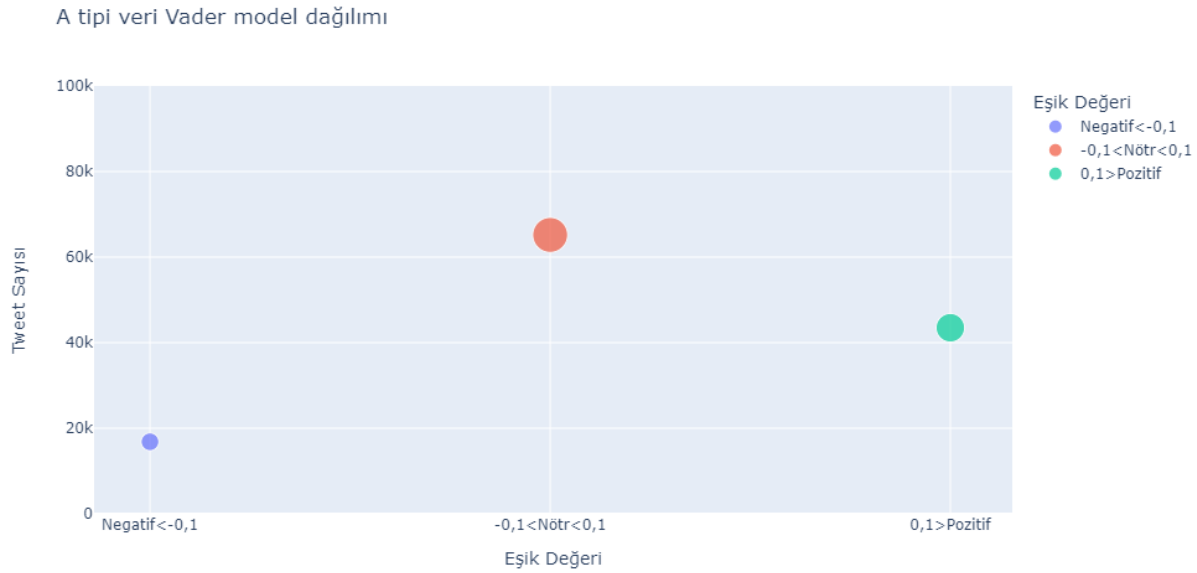
- Pozitif: vader_compound >= 0,1
- Nötr: vader_compound >= -0,1 ve vader_compound < 0,1

- Negatif: vader_compound < -0,1

Sınıflandırma sonucunda 43486 twitter verisi pozitif, 65187 nötr ve 16831 negatif olduğu gözlemlenmiştir. Elde edilen sonuçlar Şekil 5'te Bar grafiği ve Şekil 6'da Scatter grafiği verilmektedir.



Şekil 5. A tipi veri Vader modeli pozitif twitter veri tablosu bar grafiği



Şekil 6. A tipi veri Vader modeli pozitif twitter veri tablosu scatter grafiği

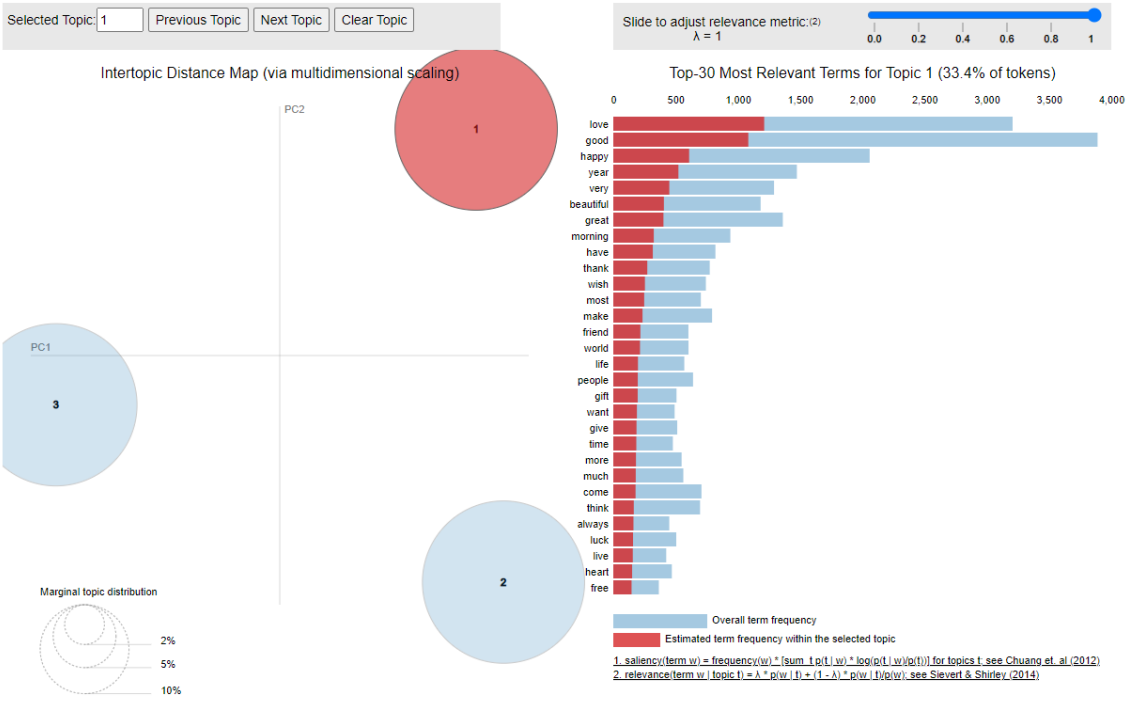
Flair modeli için 125504 twitter verisi kullanılmıştır. Model sonucunda twitter veriler 2 sınıfta incelenmiştir.

- Pozitif: flair_pos >= 0,1
- Negatif: flair_neg < -0,1

Sınıflandırma sonucunda 83336 twitter verisi pozitif, 42168 twitter verisi negatif olduğu gözlemlenmiştir. Dağılıma bakıldığında ağırlıklı olarak pozitif twitter verisi bulunmaktadır.

3.6.3 Metin Konu Çıkarımı Yöntemi

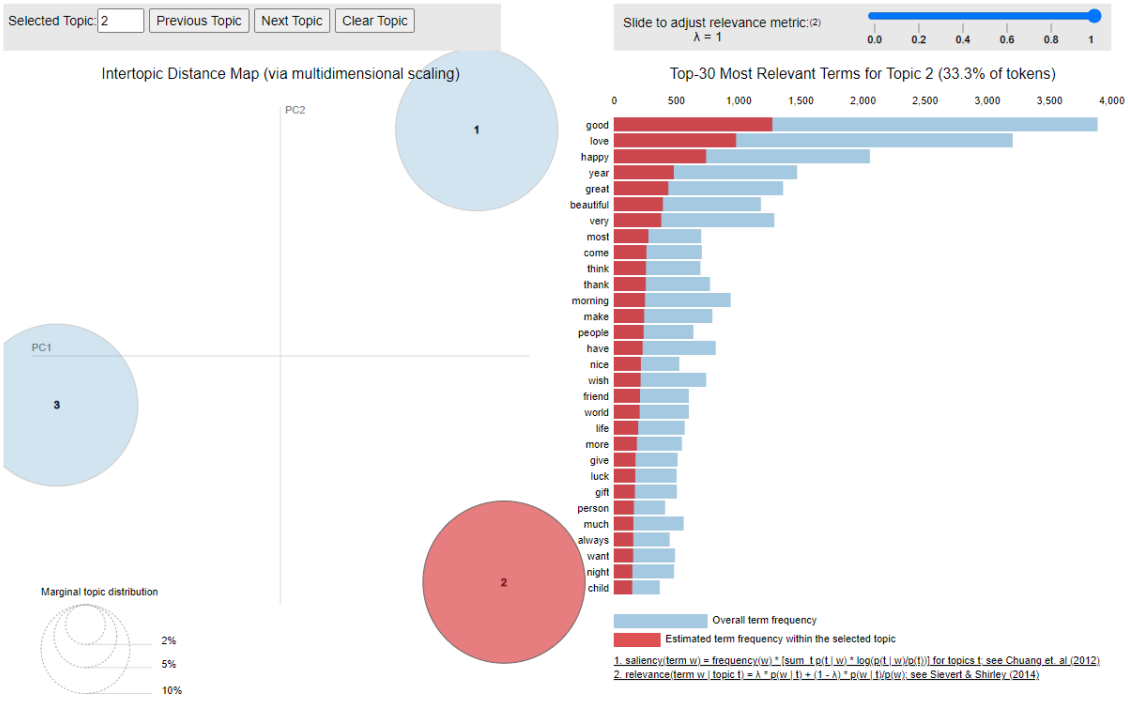
LDA modeli için 125504 twitter verisi kullanılmıştır. Model sonucunda twitter verileri 3 konu başlığı altında incelenmiştir. Elde edilen sonuçlar Şekil 7, Şekil 8 ve Şekil 9'da verilmektedir.



Şekil 7. A tipi veri LDA modeli 1 numaralı konu dağılımı

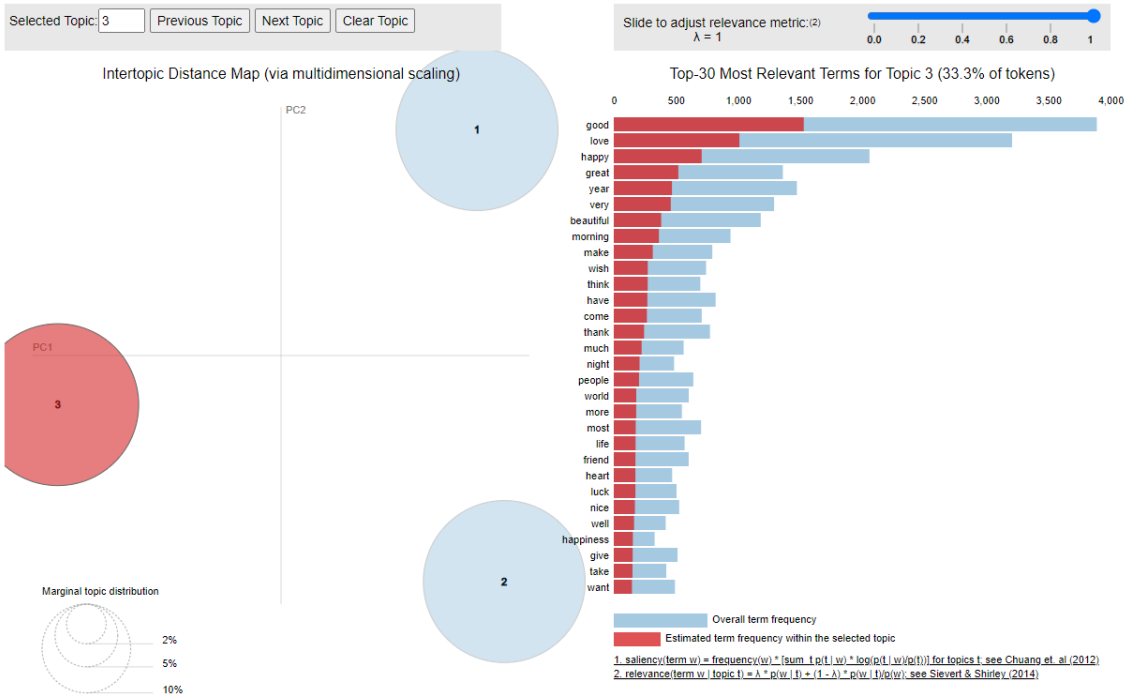
Şekil 7'de A tipi twitter veri LDA 1 numaralı konu dağılımı tablosunda aşk, iyi, mutlu, yıl, güzel gibi çıkarımlar yapılmaktadır. Bu çıkarımlara bakılarak twitter verilerinde

sıklıkla kullanılan kelimelerden ilgi alanları üzerinde yorum yapmak mümkün olmaktadır.



Şekil 8. A tipi veri LDA modeli 2 numaralı konu dağılımı

Şekil 8'de A tipi twitter veri LDA 2 numaralı konu dağılımı tablosunda iyi, aşk, mutlu, yıl, güzel gibi çıkarımlar yapılmaktadır. Bu çıkarımlara bakılarak twitter verilerinde sıklıkla kullanılan kelimelerden ilgi alanları üzerinde yorum yapmak mümkün olmaktadır.



Şekil 9. A tipi veri LDA modeli 3 numaralı konu dağılımı

Şekil 9'da A tipi twitter veri LDA 3 numaralı konu dağılımı tablosunda iyi, aşk, mutlu, yıl, çok, güzel gibi çıkarımlar yapılmaktadır. Bu çıkarımlara bakılarak twitter verilerinde sıklıkla kullanılan kelimelerden ilgi alanları üzerinde yorum yapmak mümkün olmaktadır.

3.7. B Tipi Veri Kullanılan Yöntemler

B tipi veri üzerinde Nesne tespiti yöntemi, Metin duygu tahmini yöntemi, Metin konu çıkarımı yöntemi kullanılmaktadır. B tipi 1002 metin verisi üzerinde langdetect kütüphanesi ile ham veri dil dağılımı hesaplanmaktadır. Tablo4'te ham veri üzerindeki dil dağılımı görülmektedir.

Tablo 4. B Tipi Veri Dil Dağılımı

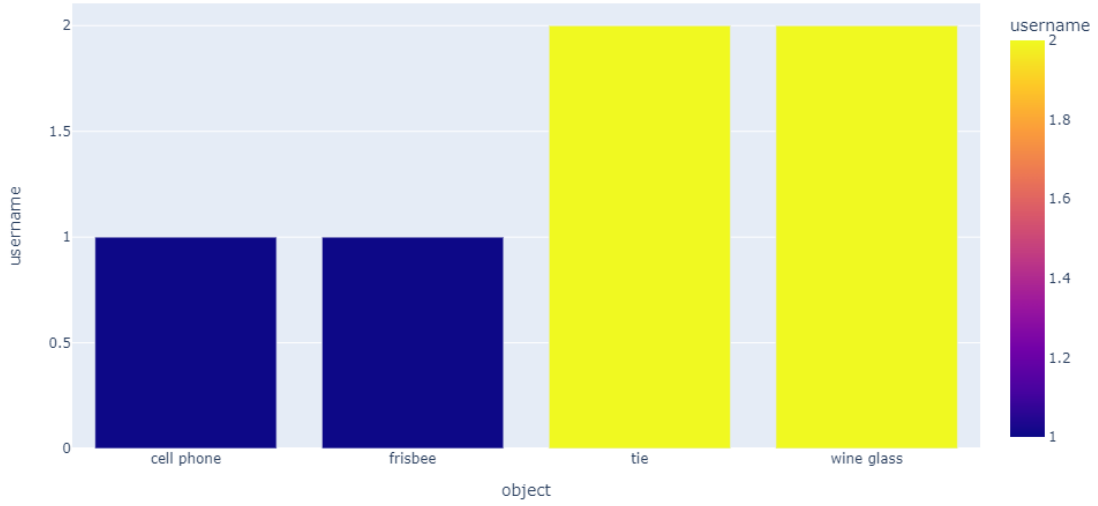
Dil	Metin Sayısı
Türkçe	943
Almanca	15

Estonyaca	9
Başka diller	35

Tablo4'te B tipi veri dil dağılımı görülmektedir. B tipi veri üzerinde Türkçe 943, Almanca 15, Estonyaca 9 ve başka diller 35 olacak şekilde dağılım göstermektedir.

3.7.1. Veri Nesne Tespiti Yöntemi

B tipi veride 1002 adet twitter verisi bulunmaktadır. İçerisinde resim olan 180 twitter verilerinden 29 adet farklı nesne tespiti gerçekleştirilmiştir. Gerçekleştirilen nesne tespiti işleminde eşik değeri olarak 0.65 seçilmiştir. Twitter verisi üzerinde kullanmış olduğumuz metin duygu tahmini yöntemleri ile twitter verileri pozitif ve negatif olarak iki sınıfa ayrılmıştır. Böylelikle resimli twitter verilerinde tespit edilen nesnelerin kullanıcı kitlesi için olumlu yada olumsuz olarak hangi anlam içerdiği belirlenmiştir. Şekil 10'da B tipi twitter verisinde elde edilen nesne tespitleri gösterilmiştir.



Şekil 10. B tipi pozitif twitter veri nesne tespiti tablosu bar grafiği

Şekil 10'da B tipi twitter veri nesne tespiti tablosunda Kravat, Bardak, Frizbi ve Cep Telefonu çıkarımları yapılmaktadır. Yapılan bu çıkarımlar Metin duygu tahmini

yöntemleri ile belirlenen pozitif twitter verileri üzerinden çıkarılmıştır. Bu sayede fotoğraflardan elde edilen nesneler kullanıcılar için pozitif anlama gelmektedir.

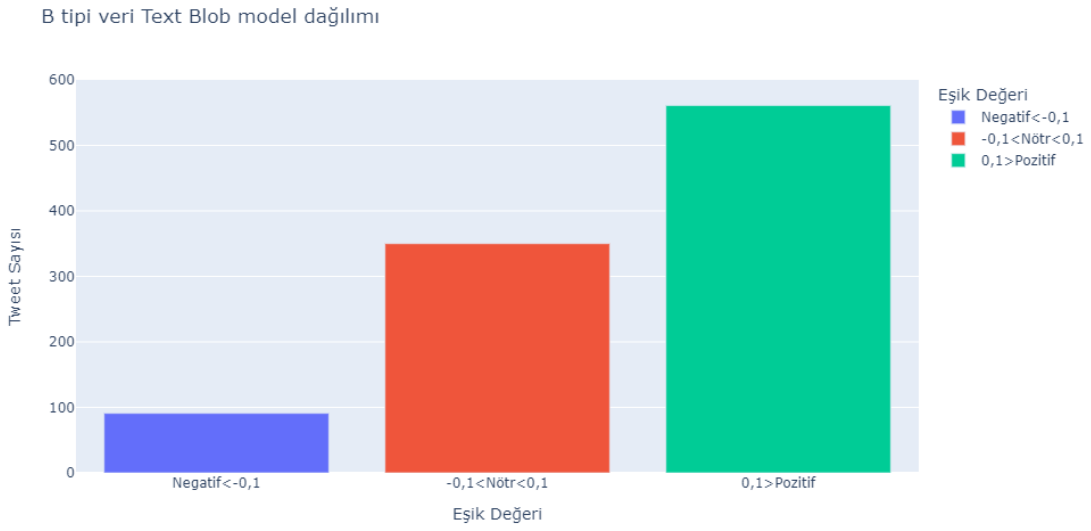
3.7.2 Metin Duygu Tahmini Yöntemi

B tipi veride 1002 adet twitter verisi bulunmaktadır. Twitter verileri üzerinde Text Blob, Vader ve Flair modelleri ile duygu tahmini yapılmıştır.

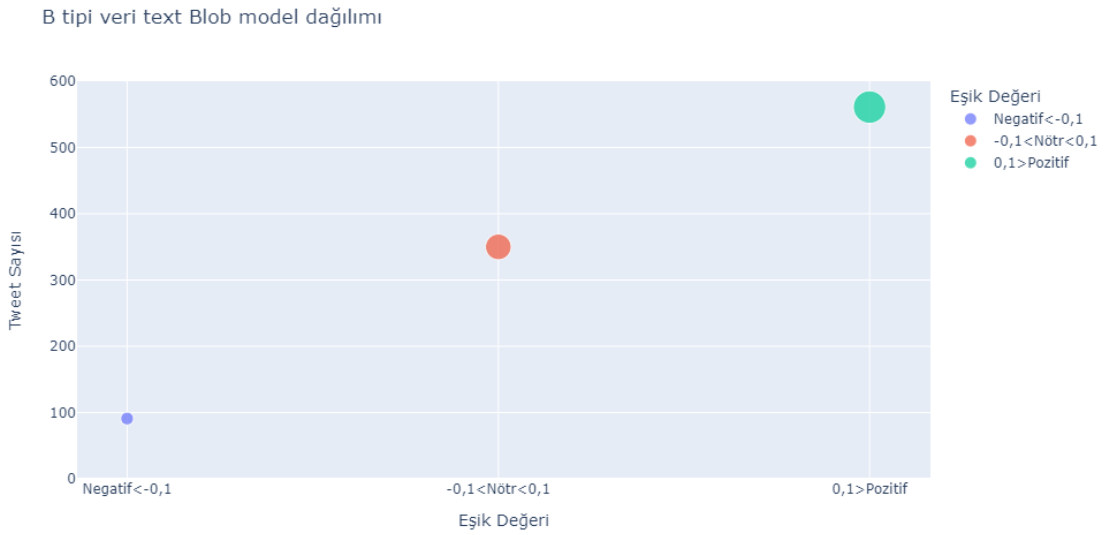
Tez kapsamında Text Blob modeli için 1002 twitter verisi kullanılmıştır. Model sonucunda twitter verileri 3 sınıfta incelenmiştir.

- Pozitif: Polarity >= 0,1
- Nötr: Polarity >= -0,1 ve Polarity < 0,1
- Negatif: Polarity < -0,1

Sınıflandırma sonucunda 561 twitter verisi pozitif, 350 nötr ve 91 negatif olduğu gözlemlenmiştir. Elde edilen sonuçlar Şekil 11'de Bar grafiği ve Şekil 12'de Scatter grafiği verilmektedir.



Şekil 11. B tipi veri Text Blob modeli pozitif twitter veri tablosu bar grafiği

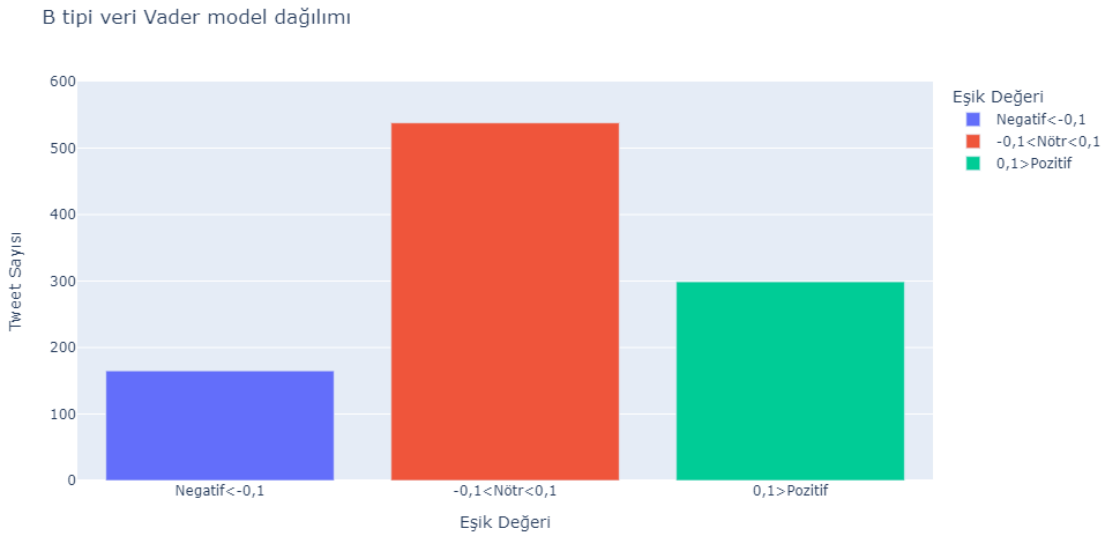


Şekil 12. B tipi veri Text Blob modeli pozitif twitter veri tablosu scatter grafiği

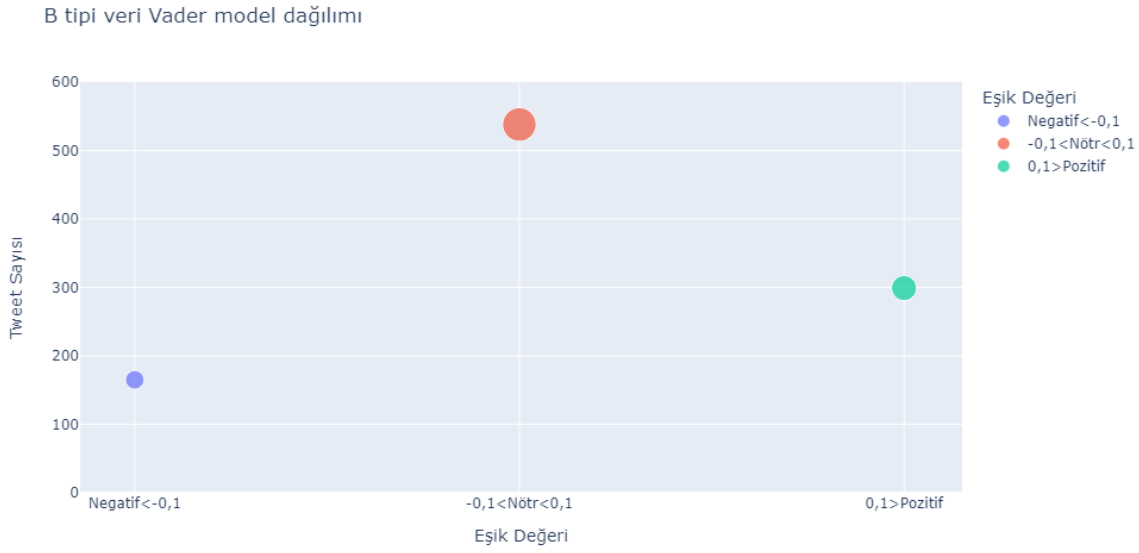
Tez kapsamında Vader modeli için 1002 twitter verisi kullanılmıştır. Model sonucunda twitter verileri 3 sınıfta incelenmiştir.

- Pozitif: vader_compound >= 0,1
- Nötr: vader_compound >= -0,1 ve vader_compound < 0,1
- Negatif: vader_compound < -0,1

Sınıflandırma sonucunda 299 twitter verisi pozitif, 538 nötr ve 165
negatif olduğu gözlemlenmiştir. Elde edilen sonuçlar Şekil 13'te Bar grafiği ve Şekil 14'te Scatter grafiği verilmektedir.



Şekil 13. B tipi veri Vader modeli pozitif twitter veri tablosu bar grafiği



Şekil 14. B tipi veri Vader modeli pozitif twitter veri tablosu scatter grafiği

Tez kapsamında Flair modeli için 1002 twitter verisi kullanılmıştır. Model sonucunda twitter verileri 2 sınıfta incelenmiştir.

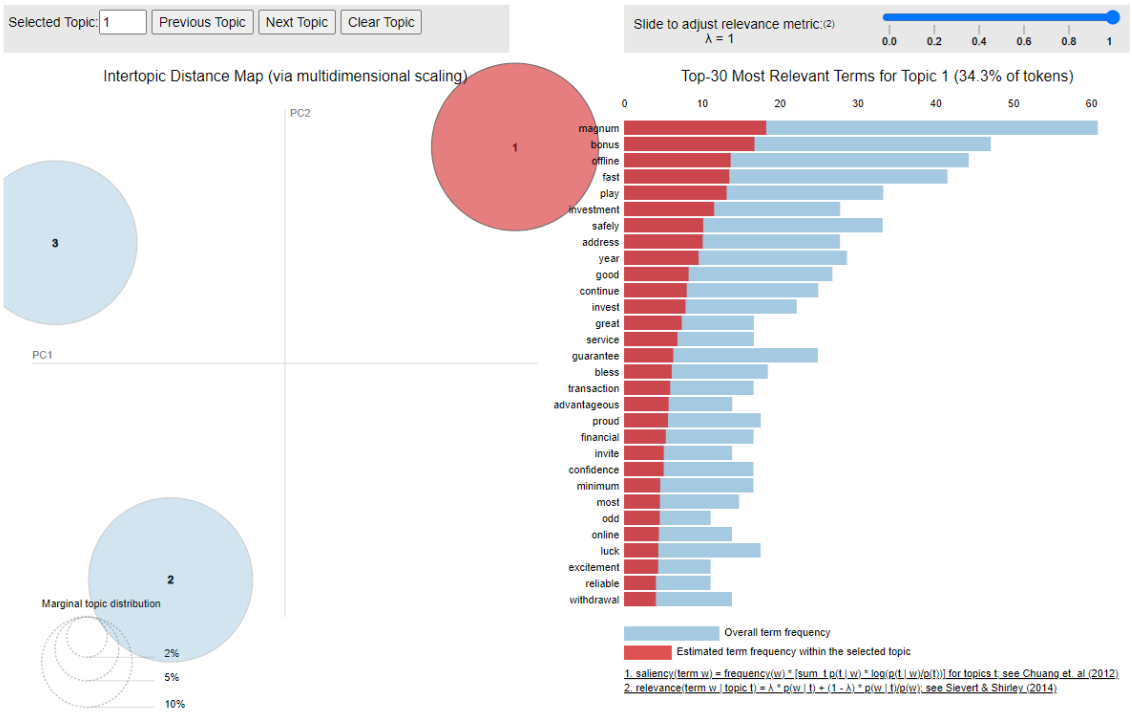
- Pozitif: flair_pos >= 0,1

- Negatif: flair_neg < -0,1

Sınıflandırma sonucunda 388 twitter verisi pozitif, 614
negatif olduğu gözlemlenmiştir. Dağılıma bakıldığında ağırlıklı olarak negatif twitter verisi bulunmaktadır.

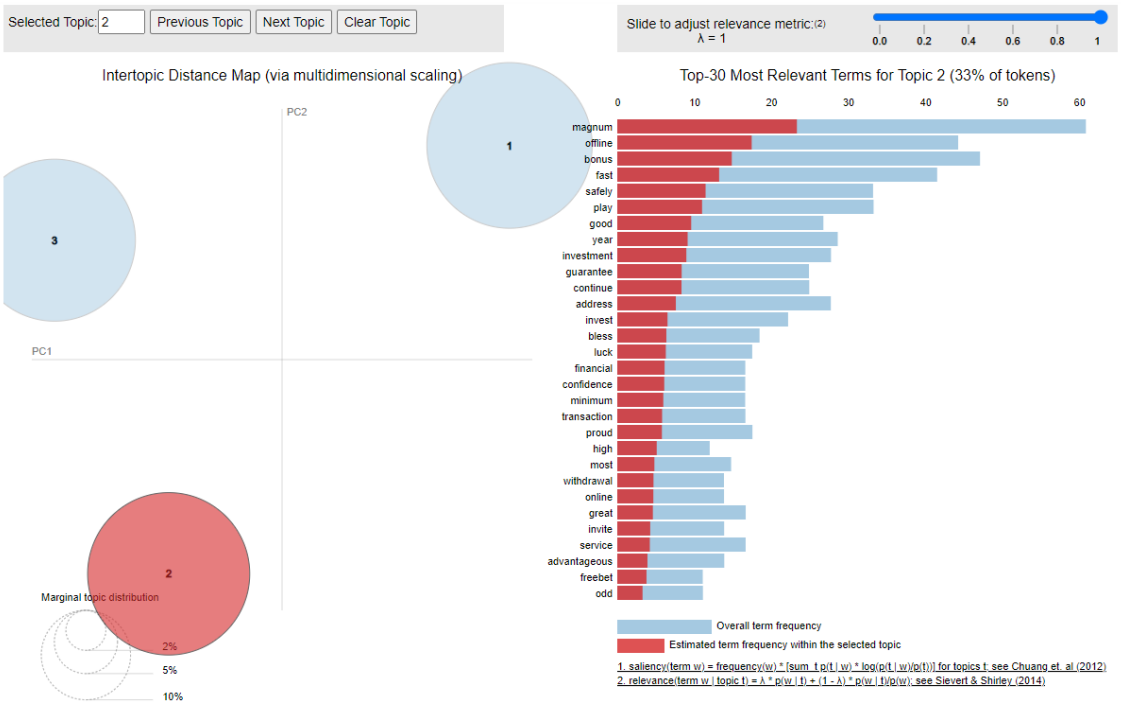
3.7.3 Metin Konu Çıkarımı Yöntemi

Tez kapsamında LDA modeli için 1002 twitter verisi kullanılmıştır. Model sonucunda twitter verileri 3 konu sayısında incelenmiştir. Elde edilen sonuçlar Şekil 15, Şekil 16 ve Şekil 17'de verilmektedir.



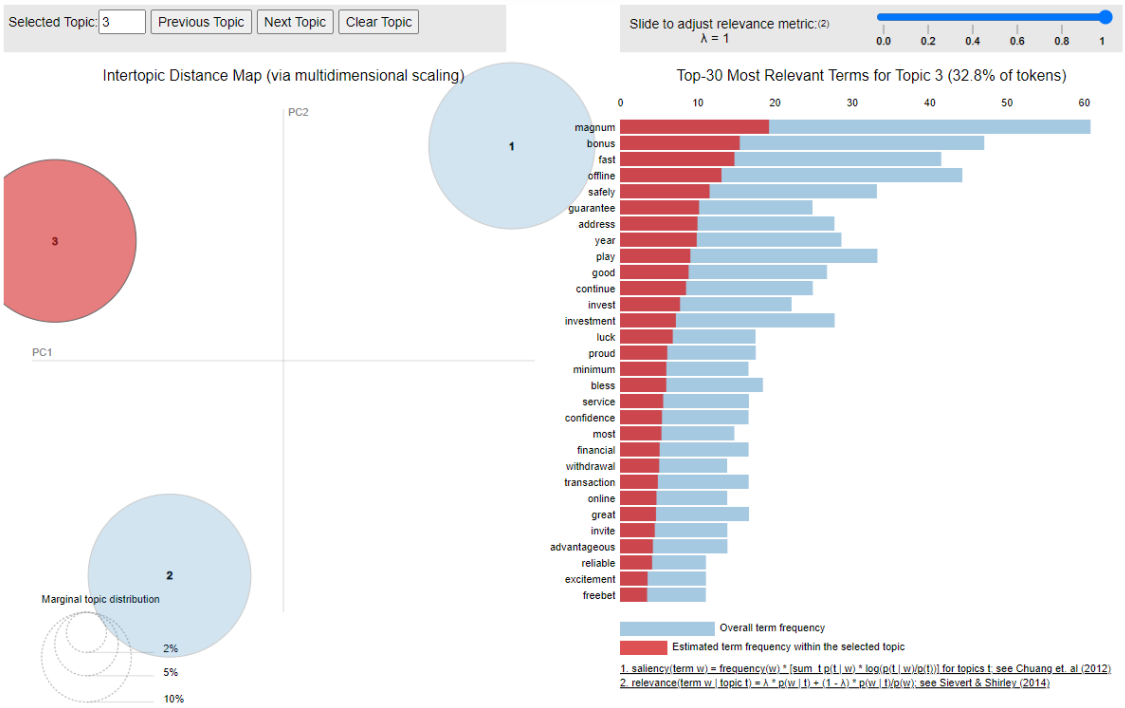
Şekil 15. B tipi veri LDA modeli 1 numaralı konu dağılımı

Şekil 15'te B tipi twitter veri LDA 1 numaralı konu dağılımı tablosunda magnum, bonus, çevrimdışı, hızlı gibi çıkarımlar yapılmaktadır. Bu çıkarımlara bakılarak twitter verilerinde sıklıkla kullanılan kelimelerden ilgi alanları üzerinde yorum yapmak mümkün olmaktadır.



Şekil 16. B tipi veri LDA modeli 2 numaralı konu dağılımı

Şekil 16'da B tipi twitter veri LDA 2 numaralı konu dağılımı tablosunda magnum, çevrimdışı, bonus, hızlı, adres gibi çıkarımlar yapılmaktadır. Bu çıkarımlara bakılarak twitter verilerinde sıklıkla kullanılan kelimelerden ilgi alanları üzerinde yorum yapmak mümkün olmaktadır.



Şekil 17. B tipi veri LDA modeli 3 numaralı konu dağılımı

Şekil 17'de B tipi twitter veri LDA 3 numaralı konu dağılımı tablosunda magnum, bonus, hızlı, çevrimdışı gibi çıkarımlar yapılmaktadır. Bu çıkarımlara bakılarak twitter verilerinde sıklıkla kullanılan kelimelerden ilgi alanları üzerinde yorum yapmak mümkün olmaktadır.

3.8. C Tipi Veri Kullanılan Yöntemler

C tipi veri üzerinde Nesne tespiti yöntemi, Metin duygu tahmini yöntemi, Metin konu çıkarımı yöntemi kullanılmaktadır. C tipi 1208 metin verisi üzerinde langdetect kütüphanesi ile ham veri dil dağılımı hesaplanmaktadır. Tablo5'te ham veri üzerindeki dil dağılımı görülmektedir.

Tablo 5. C Tipi Veri Dil Dağılımı

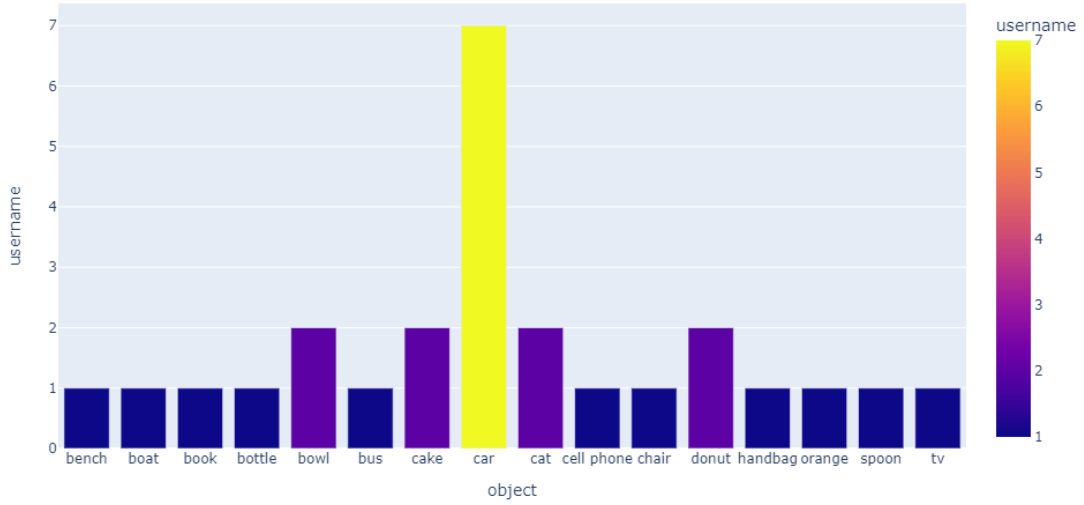
Dil	Metin Sayısı
Türkçe	972
Endonezya dili	68

İngilizce	32
Başka Diller	136

Tablo5'te C tipi veri dil dağılımı görülmektedir. C tipi veri üzerinde Türkçe 972, Endonezya dili 68, İngilizce 32 ve başka diller 136 olacak şekilde dağılım göstermektedir.

3.8.1. Veri Nesne Tespiti Yöntemi

C tipi veride 1208 adet twitter verisi bulunmaktadır. İçerisinde resim olan 384 twitter verilerinden 48 adet farklı nesne tespiti gerçekleştirilmiştir. Gerçekleştirilen nesne tespiti işleminde eşik değeri olarak 0.65 seçilmiştir. Twitter verisi üzerinde kullanmış olduğumuz metin duygu tahmini yöntemleri ile twitter verileri pozitif ve negatif olarak iki sınıfa ayrılmıştır. Böylelikle resimli twitter verilerinde tespit edilen nesnelerin kullanıcı kitlesi için olumlu yada olumsuz olarak hangi anlam içerdiği belirlenmiştir. Şekil 18'de C tipi twitter verisinde elde edilen nesne tespitleri gösterilmiştir.



Şekil 18. C tipi pozitif twitter veri nesne tespiti tablosu bar grafiği

Şekil 18'de C tipi twitter veri nesne tespiti tablosunda Araba, Donut(kek), Kedi, Cep Telefonu vb. çıkarımlar yapılmaktadır. Yapılan bu çıkarımlar Metin duygu tahmini

yöntemleri ile belirlenen pozitif twitter verileri üzerinden çıkarılmıştır. Bu sayede fotoğraflardan elde edilen nesneler kullanıcılar için pozitif anlama gelmektedir.

3.8.2 Metin Duygu Tahmini Yöntemi

C tipi veride 1208 adet twitter verisi bulunmaktadır. Twitter verileri üzerinde Text Blob, Vader ve Flair modelleri ile duygu tahmini yapılmıştır.

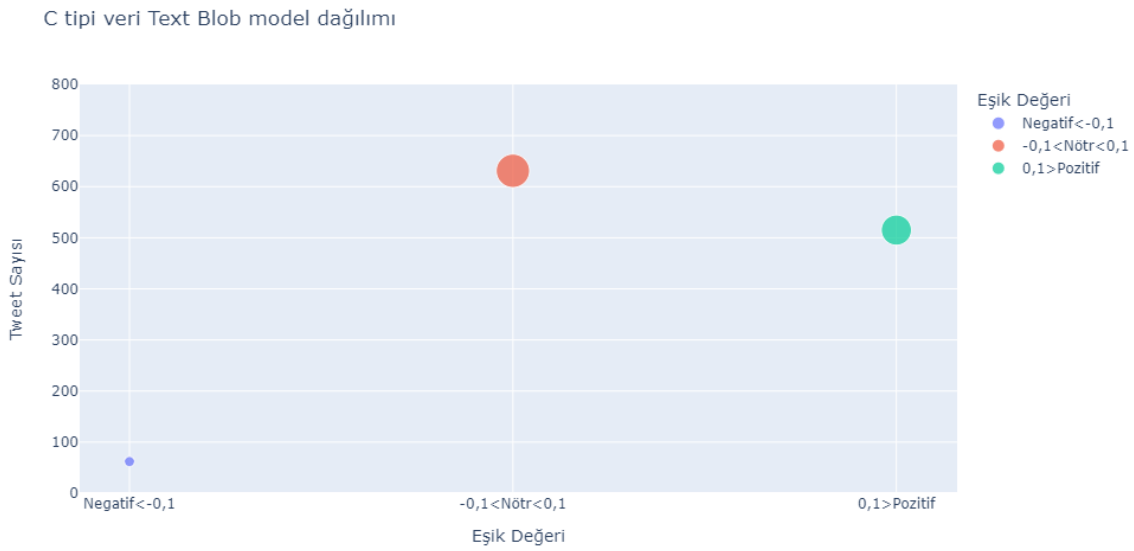
Tez kapsamında Text Blob modeli için 1208 twitter verisi kullanılmıştır. Model sonucunda twitter verileri 3 sınıfta incelenmiştir.

- Pozitif: Polarity >= 0,1
- Nötr: Polarity >= -0,1 ve Polarity < 0,1
- Negatif: Polarity < -0,1

Sınıflandırma sonucunda 515 twitter verisi pozitif, 631 nötr ve 62 negatif olduğu gözlemlenmiştir. Elde edilen sonuçlar Şekil 19'da Bar grafiği ve Şekil 20'de Scatter grafiği verilmektedir.



Şekil 19. C tipi veri Text Blob modeli pozitif twitter veri tablosu bar grafiği

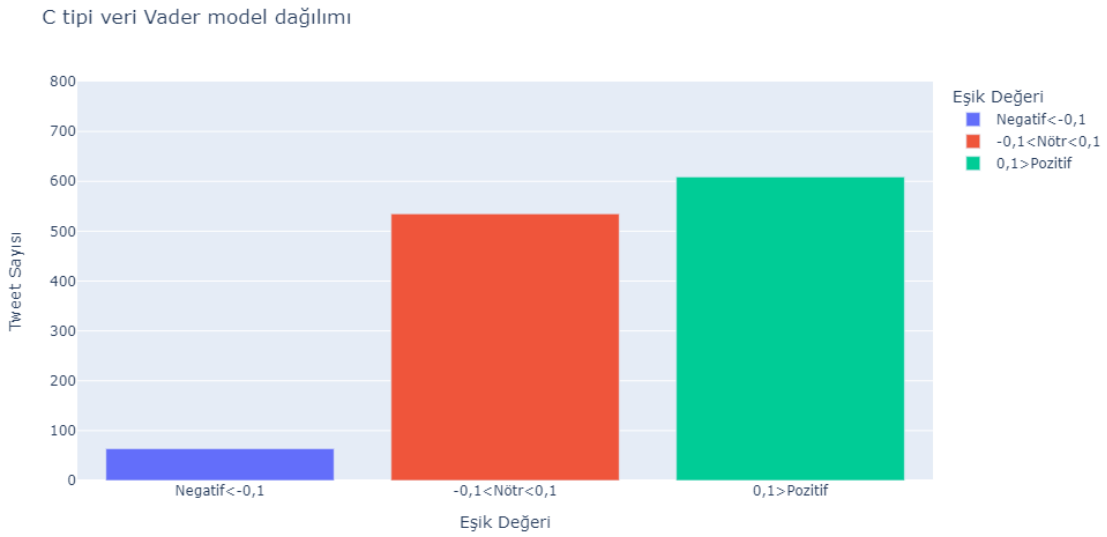


Şekil 20. C tipi veri Text Blob modeli pozitif twitter veri tablosu scatter grafiği

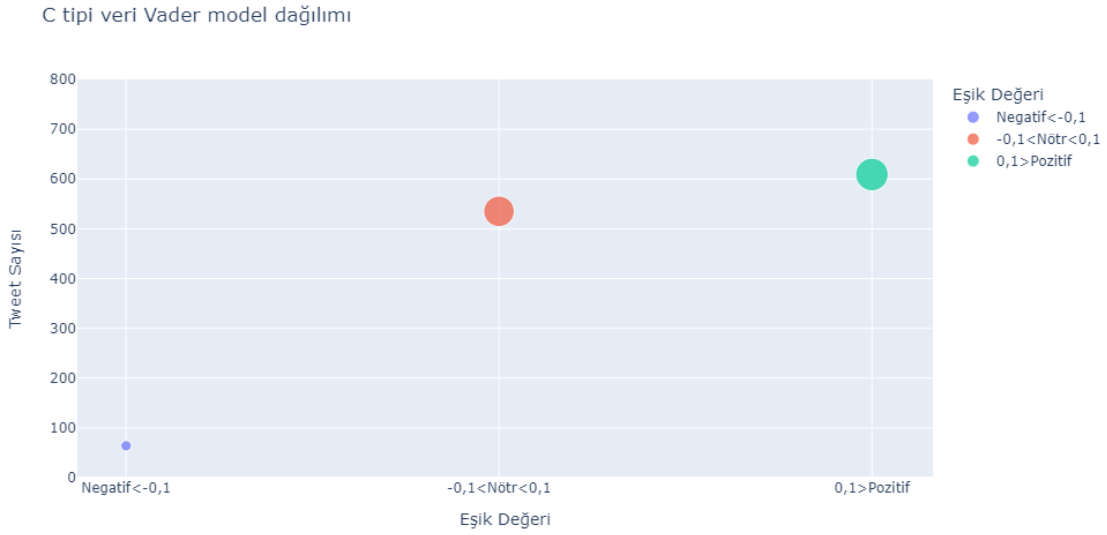
Tez kapsamında Vader modeli için 1208 twitter verisi kullanılmıştır. Model sonucunda twitter verileri 3 sınıfta incelenmiştir.

- Pozitif: vader_compound >= 0,1
- Nötr: vader_compound >= -0,1 ve vader_compound < 0,1
- Negatif: vader_compound < -0,1

Sınıflandırma sonucunda 609 twitter verisi pozitif, 535 nötr ve 64 negatif olduğu gözlemlenmiştir. Elde edilen sonuçlar Şekil 20'de Bar grafiği ve Şekil 21'de Scatter grafiği verilmektedir.



Şekil 21. C tipi veri Vader modeli pozitif twitter veri tablosu bar grafiği



Şekil 22. C tipi veri Vader modeli pozitif twitter veri tablosu scatter grafiği

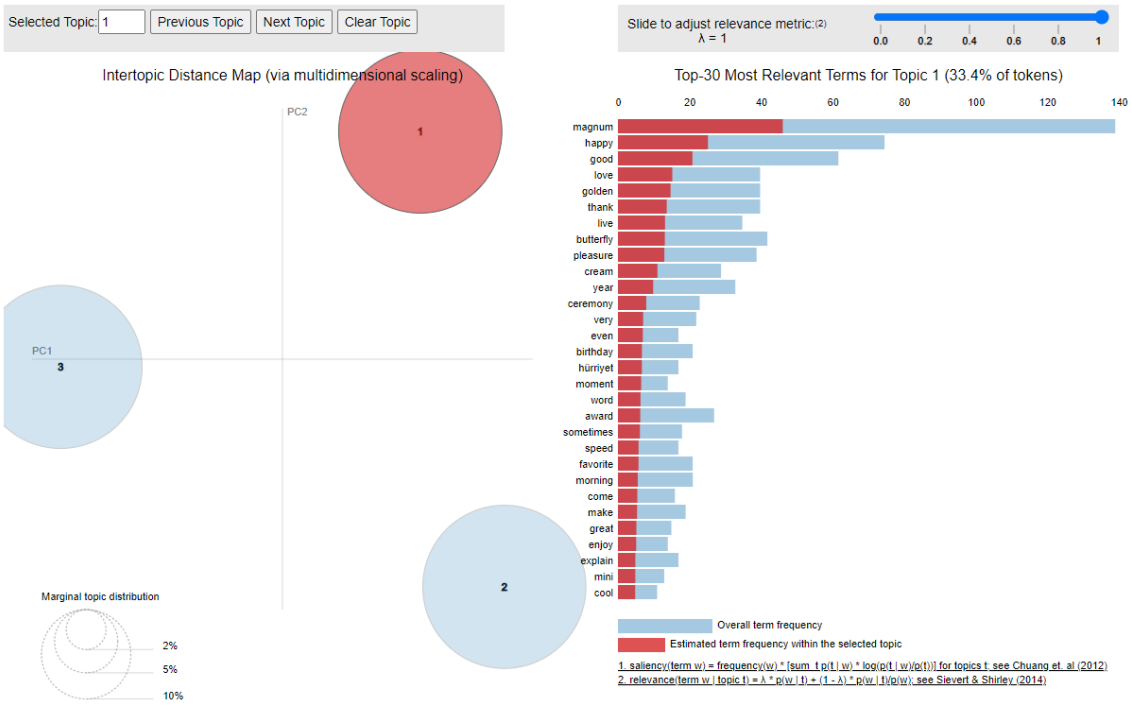
Tez kapsamında Flair modeli için 1208 twitter verisi kullanılmıştır. Model sonucunda twitter verileri 2 sınıfta incelenmiştir.

- Pozitif: flair_pos >= 0,1
- Negatif: flair_neg < -0,1

Sınıflandırma sonucunda 1033 twitter verisi pozitif, 178 negatif olduğu gözlemlenmiştir. Dağılıma bakıldığında ağırlıklı olarak pozitif twitter verisi bulunmaktadır.

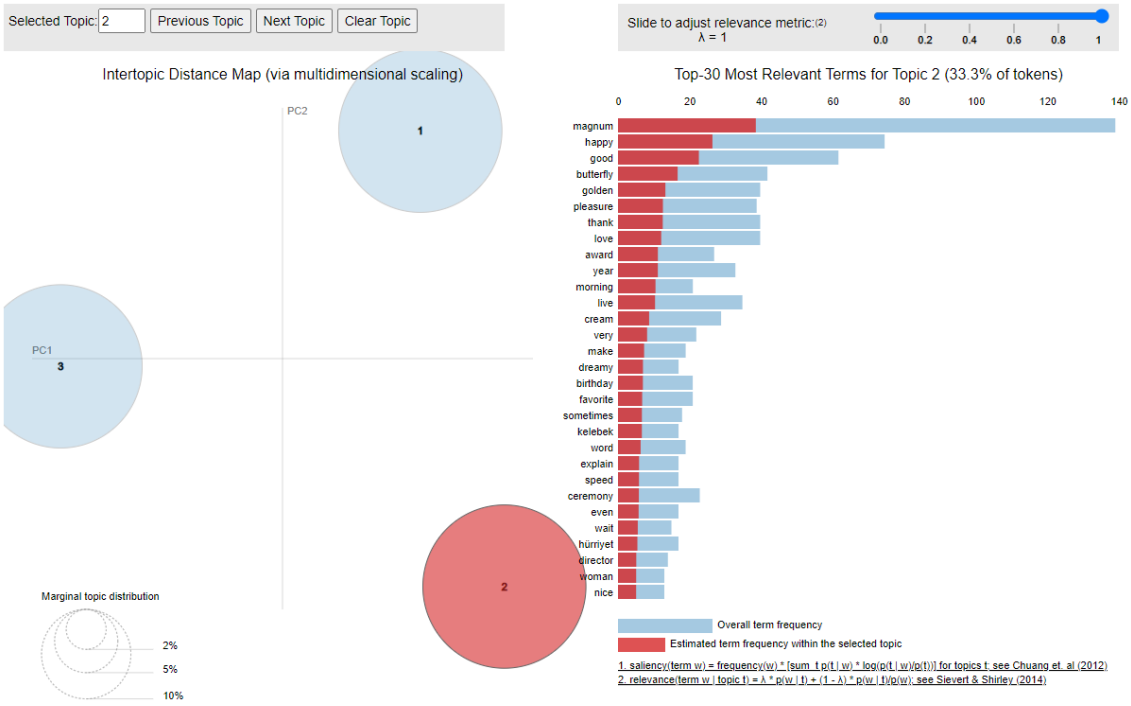
3.8.3 Metin Konu Çıkarımı Yöntemi

Tez kapsamında LDA modeli için 1208 twitter verisi kullanılmıştır. Model sonucunda twitter verileri 3 konu sayısında incelenmiştir. Elde edilen sonuçlar Şekil 23, Şekil 24 ve Şekil 25'te verilmektedir.



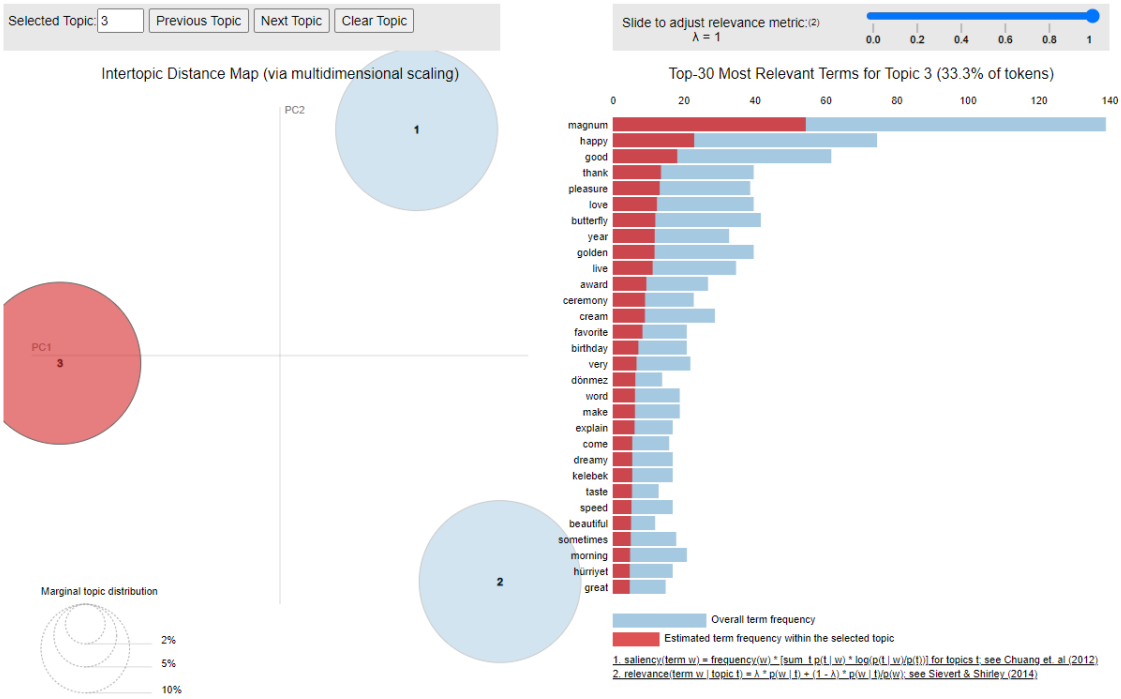
Şekil 23. C tipi veri LDA modeli 1 numaralı konu dağılımı

Şekil 23'te C tipi twitter veri LDA 1 numaralı konu dağılımı tablosunda magnum, mutlu, iyi, aşk, altın(gold), mini gibi çıkarımlar yapılmaktadır. Bu çıkarımlara bakılarak twitter verilerinde sıklıkla kullanılan kelimelerden ilgi alanları üzerinde yorum yapmak mümkün olmaktadır.



Şekil 24. C tipi veri LDA modeli 2 numaralı konu dağılımı

Şekil 24'te C tipi twitter veri LDA 2 numaralı konu dağılımı tablosunda magnum, mutlu, iyi, kelebek, altın(gold), zevk gibi çıkarımlar yapılmaktadır. Bu çıkarımlara bakılarak twitter verilerinde sıklıkla kullanılan kelimelerden ilgi alanları üzerinde yorum yapmak mümkün olmaktadır.



Şekil 25. C tipi veri LDA modeli 3 numaralı konu dağılımı

Şekil 25'te C tipi twitter veri LDA 3 numaralı konu dağılımı tablosunda magnum, mutlu, iyi, teşekkür, kelebek, altın(gold), aşk gibi çıkarımlar yapılmaktadır. Bu çıkarımlara bakılarak twitter verilerinde sıklıkla kullanılan kelimelerden ilgi alanları üzerinde yorum yapmak mümkün olmaktadır.

3.9. Verilerin Çözümlenmesi ve Yorumlanması

Bu tez kapsamında kullanılan veriler sosyal medya platformu olan twitter verileridir. Bu veriler 3 kısımdan oluşmaktadır. Veri tipleri hakkında bilgiye Tablo 1 den ulaşılmaktadır.

Bu 3 kısımdan oluşan veriler Python Selenium sanal botu ve twitter api kütüphaneleri kullanılarak toplanılmıştır. İlk olarak twitter verileri üzerinde bulunan resimler üzerinde nesne tespiti işlemi gerçekleştirilmiştir. Nesne tespiti işleminden sonra twitter verileri üzerinde metin duygu tahmini işlemi yapılmıştır. Metin duygu tahmini işlemi yapılırken 3 tane farklı metin duygu tahmini yöntemi kullanılmıştır. Kullanılan metin duygu tahmini yöntemleri Flair, Vader ve Text Blob yöntemleridir. Bu 3 yöntemin twitter verileri üzerinde etiketlemiş olduğu twitter verilerinden pozitif olma eşiği 3 yöntem içinde 0,1 üzerinde olan twitter verileri seçilmiştir. Bu seçilen twitter verileri üzerinde bulunan

resimlerden elde edilen nesneler grafiksel olarak gösterilmiştir. Bu işlem adımından sonra pozitif twitter verileri üzerinde LDA yöntemi kullanılarak metin konu çıkarımı işlemi gerçekleştirilmiştir. Böylelikle pozitif twitter verileri üzerinde tespit edilen nesneler ile LDA yöntemi ile çıkarılan konu başlıkları arasında pozitif bir ilişki olduğu çıkarımı yapılabilmektedir.

4. SONUÇ

Bu tez kapsamında yapılan çalışmaya bakıldığında yapay zeka teknikleri kullanılarak büyük çaplı sosyal medya kullanıcı kitlelerinin analizinin hızlı ve kapsamlı bir şekilde yapılabileceği görülmektedir. Metin duygu analizi ile twitter resim verisi ve twitter metin verisi arasındaki olumlu yada olumsuz bağlantı tespit edilebileceği görülmektedir. Veriler üzerinde uygulanan 3 farklı metin duygu analizi yönteminin bulmuş olduğu sonuçlar ele alınmaktadır. Bu sayede 3 farklı metin duygu tahmini modelinin pozitif olarak belirlediği metin verileri seçilebilmektedir. 3 farklı metin duygu analizi modelinin pozitif olarak belirlediği metin verileri üzerinde konu modellemesi sonuçlarına bakıldığında çıkan kelimelerin pozitif anlam içerdiği görülmektedir. Metin duygu tahmini çalışmalarına bakıldığında genellikle tek bir model kulanılarak yapıldığı görülmektedir. Bu yöntem tek bir değerlendirme metriğine baktığı için bu çalışmadaki analizlere göre daha düşük doğruluk oranı vermektedir. Bu çalışmada kullandığımız 3 metin duygu tahmini modelinin hesaplamış olduğu sonuçlara bakıldığında 3 modelin aynı metin verileri üzerindeki sonuçlarında pozitif, negatif ve nötr metin verileri sayılarında farklılık olduğu görülmektedir. Çalışma kapsamında 3 metin duygu tahmini modelinin beraber kulanılarak 3 modelinde pozitif olarak kabul ettiği metin verileri üzerinde çalışmak daha doğru analizler yapma noktasında katkı sağlamaktadır. Tez kapsamında kullanılan 3 veri tipine bakıldığında veri miktarı arttıkça modellerin daha kapsamlı sonuçlar verdiği görülmektedir. Bu alanda yapılacak daha kapsamlı çalışmaların olması ve kullanılan tekniklerin geliştirilmesi ile daha iyi sonuçlar alınabilmesi mümkün olmaktadır. Son dönemde artan sosyal medya kullanımı ile düzensiz veri olarak adlandırdığımız işlenmemiş veriler her geçen zaman diliminde artmaktadır. Geleneksel takipçi profilleme sistemlerine alternatif olarak geliştirilen yapay zeka destekli takipçi profilleme sistemi daha kapsamlı ve hızlı analizlerin yapılabilmesine olanak tanımaktadır. Sosyal medya üzerinden yapılabilecek analizler ile bir çok sektör bazında etkili sonuçlar elde edilebileceği görülmektedir. Böylelikle yapay zeka desteği ile yapılacak profilleme işlemi geleneksel yöntemlere göre çok daha hızlı ve daha kısa zamanda bir çok kullanıcıyı profilleme noktasında fayda sağlaması ile ön plana çıkmaktadır. Yapay zeka tekniklerinin zamanla geliştiğini göz önünde bulundurduğumuzda bu alanda yapılacak çalışmaların gelişmesi ile birlikte daha kapsamlı, doğruluk payı daha yüksek ve daha derinlemesine bir analiz imkanı sunacaktır. Kitlelerin profillemesi ve kitlelere uygun reklam

tercihlerinin sunulması pazar faaliyetlerinin artması noktasında olumlu etki oluşturmaktadır. Her geçen gün sosyal medya üzerinde biriken işlenmemiş veriler bu alanda yapılacak çalışmalar veri anlamında kaynak oluşturmaktadır. Bu kaynağın yapay zeka teknikleri ile anlamlı bir hale getirilmesi ve sektör bazında kulanılacak bilgilere dönüştürülmesi pazar faaliyetleri açısından olumlu etki oluşturmaktadır.

KAYNAKLAR

Algur, Ö., Tümen, V., & Yıldırım, Ö. (2018). Dış Ortam Görüntülerindeki İnsan Hareketlerinin Hibrit Derin Öğrenme Yöntemleri Kullanarak Sınıflandırılması. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, *30*(3), 121-129.

Bashir, S. M. A., & Wang, Y. (2021). Small object detection in remote sensing images with residual feature aggregation-based super-resolution and object detector network. *Remote Sensing*, *13*(9), 1854.

Behrendt, K., Novak, L., & Botros, R. (2017, May). A deep learning approach to traffic lights: Detection, tracking, and classification. In *2017 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 1370-1377). IEEE.

Bharadi, D. V. A., Poshe, P. S., Ashok, G., & Yadav, R. S. N. (2021). Movie Genre Prediction and Sentiment Analysis Using Natural Language Processing.

Bhavani, A., & Kumar, B. S. (2021, April). A Review of State Art of Text Classification Algorithms. In *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 1484-1490). IEEE.

Bougteb, Y., Ouhbi, B., & Frikh, B. (2019, October). Deep Learning Based Topics Detection. In *2019 Third International Conference on Intelligent Computing in Data Sciences (ICDS)* (pp. 1-7). IEEE.

Cahyani, D. E., & Patasik, I. (2021). Performance comparison of TF-IDF and Word2Vec models for emotion text classification. *Bulletin of Electrical Engineering and Informatics*, *10*(5), 2780-2788.

Chaudhri, A. A., Saranya, S. S., & Dubey, S. (2021). Implementation paper on analyzing COVID-19 vaccines on twitter dataset using tweepy and Text Blob. *Annals of the Romanian Society for Cell Biology*, 8393-8396.

Deloitte. (Nisan, 2021). Türkiye'de Tahmini Medya ve Reklam Yatırımları. https://www2.deloitte.com/content/dam/Deloitte/tr/Documents/technology-media-telecommunications/medya-ve-reklam-yatirimlari-2020-raporu.pdf.

Dharma, E. M., Gaol, F. L., Leslie, H., Warnars, H. S., & Soewito, B. (2022). The accuracy comparison among word2vec, glove, and fasttext towards convolution neural network (cnn) text classification. *J Theor Appl Inf Technol*, *100*(2), 31.

Dogru, H. B., Tilki, S., Jamil, A., & Hameed, A. A. (2021, April). Deep learning-based classification of news texts using doc2vec model. In *2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA)* (pp. 91-96). IEEE.

Elbagir, S., & Yang, J. (2019, March). Twitter sentiment analysis using natural language toolkit and VADER sentiment. In *Proceedings of the international multiconference of engineers and computer scientists* (Vol. 122, p. 16).

Fang, W., Wang, L., & Ren, P. (2019). Tinier-YOLO: A real-time object detection method for constrained environments. *IEEE Access*, *8*, 1935-1944.

Gujjar, J. P., & Kumar, H. P. (2021). Sentiment analysis: Textblob for decision making. *Int. J. Sci. Res. Eng. Trends*, *7*(2), 1097-1099.

Hanbay, K., & Hüseyin, Ü. Z. E. N. (2017). Nesne tespit ve takip metotları: Kapsamlı bir derleme. *Türk Doğa ve Fen Dergisi*, *6*(2), 40-49.

Huang, Y., Wang, R., Huang, B., Wei, B., Zheng, S. L., & Chen, M. (2021). Sentiment classification of crowdsourcing participants' reviews text based on LDA topic model. *IEEE Access*, *9*, 108131-108143.

Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 216-225).

Jiao, L., Zhang, R., Liu, F., Yang, S., Hou, B., Li, L., & Tang, X. (2021). New generation deep learning for video object detection: A survey. *IEEE Transactions on Neural Networks and Learning Systems*.

Krišto, M., Ivasic-Kos, M., & Pobar, M. (2020). Thermal object detection in difficult weather conditions using YOLO. *IEEE access*, *8*, 125459-125476.

Luo, L. X. (2019). Network text sentiment analysis method combining LDA text representation and GRU-CNN. *Personal and Ubiquitous Computing*, *23*(3), 405-412.

Mayer-Schönberger, V., & Cukier, K. (2013). *Büyük veri: yaşama, çalışma ve düşünme şeklimizi dönüştürecek bir devrim*. Paloma yayınevi.

NG Araştırma. (21 Nisan, 2020). Her 20 kişiden 19'u sosyal medya kullanıyor. Marketing Türkiye, https://www.marketingturkiye.com.tr/haberler/sosyal-medya/her-20-kisiden-19u-sosyal-medya-kullaniyor/.

Tunalı, V. (2011). *Metin madenciliği için iyileştirilmiş bir kümeleme yapısının tasarımı ve uygulaması* (Doctoral dissertation, Marmara Universitesi (Turkey)).

Wang, H., Sun, K., & Wang, Y. (2022). Exploring the Chinese Public's Perception of Omicron Variants on Social Media: LDA-Based Topic Modeling and Sentiment Analysis. *International Journal of Environmental Research and Public Health*, *19*(14), 8377.

Zhang, Y. (2021, April). Research on text classification method based on LSTM neural network model. In *2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)* (pp. 1019-1022). IEEE.

ÖZGEÇMİŞ

Kadir Yunus Koç

Bilgisayar Mühendisliği Anabilim Dalı

Eğitim

Derece	*Yıl*	*Üniversite, Enstitü, Anabilim/Anasanat Dalı*
Y.Ls.	2022	Maltepe Üniversitesi, Lisansüstü Eğitim Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı
Ls.	2019	Girne Amerikan Üniversitesi(Tam burslu), Mühendislik Fakültesi Bilgisayar Mühendisliği Anabilim Dalı
Lise	2014	Sultanbeyli Anadolu Lisesi

İş/İstihdam

Yıl	*Görev*
2019 - &	*IBSS(Arge Mühendisi)*
2019 - Summer	IBSS(Stajyer)

Mesleki Birlik/Dernek Üyelikleri

Yıl	*Kurum*
2019 -	Kurucu Üye: Onu Gülümset Derneği

Alınan Burs ve Ödüller

Yıl	*Burs/Ödül*
2014	Girne Amerikan Üniversitesi Eğitim Bursu

Yayınlar ve Diğer Bilimsel/Sanatsal Faaliyetler

"IBSS Veri Analiz Platformu" (ICENSS2021), May 07-09, 2021 Ankara/ TÜRKİYE .

"Keystroke Dynamics Based Authentication System" (UBMK2021).

"Semantic Similarity Comparison of Production Line Failure Explanations in the Field of Predictive Maintenance" (CAIAME2022).