

**T.C.
BAHÇEŞEHİR ÜNİVERSİTESİ**

**TWİTTER VERİLERİYLE DUYGU ANALİZİ VE TÜRKÇE
DUYGU KÜTÜPHANESİ**

Yüksek Lisans Tezi

SEMRA YILDIRIM

İSTANBUL, 2018

**T.C.
BAHÇEŞEHİR ÜNİVERSİTESİ**

**FEN BİLİMLERİ ENSTİTÜSÜ
MÜHENDİSLİK YÖNETİMİ**

YÜKSEK LİSANS TEZİ

SEMRA YILDIRIM

Tez Danışmanı: Yrd. Doc. Dr. Serkan AYVAZ

İSTANBUL, 2018

T.C.
BAHÇEŞEHİR ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
MÜHENDİSLİK YÖNETİMİ

Tezin Adı: Twitter Verileriyle Duygu Analizi ve Türkçe Duygu Kütüphanesi
Öğrencinin Adı Soyadı: Semra Yıldırım
Tez Savunma Tarihi: 12 Ocak 2018

Bu tezin Yüksek Lisans tezi olarak gerekli şartları yerine getirmiş olduğu Fen Bilimleri Enstitüsü tarafından onaylanmıştır.

Prof. Dr. Nafiz ARICA
Enstitü Müdürü
İmza

Bu tezin Yüksek Lisans tezi olarak gerekli şartları yerine getirmiş olduğunu onaylarım.

Doç. Dr. Ahmet BEŞKESE
Program Koordinatörü
İmza

Bu Tez tarafımızca okunmuş, nitelik ve içerik açısından bir Yüksek Lisans tezi olarak yeterli görülmüş ve kabul edilmiştir.

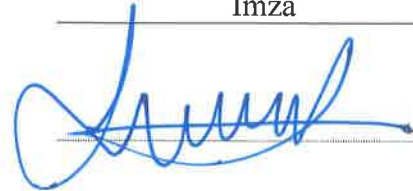
Jüri Üyeleri

Tez Danışmanı
Yrd. Doç. Dr. Serkan AYVAZ

Üye
Yrd. Doç. Dr. Yücel Batu SALMAN

Üye
Yrd. Doç. Dr. Sabri Serkan GÜLLÜOĞLU

İmza



ÖNSÖZ

Bu tez kapsamında Türkçe kelimelerden ve emojilerden oluşan duygu kütüphanesi oluşturulmuştur. Türkçe tweet paylaşan kullanıcıların paylaştıkları bu tweetler üzerinden de oluşturulan yeni kütüphaneden faydalanılarak Türkçe Duygu Analizi çalışmalarına katkı sağlamak hedeflenmiştir.

Öncelikle tez konusu seçiminden bugüne kadar benden desteklerini, yardımlarını ve bilgi birikimini esirgemeyen çalışmalarımda sürekli pozitif olarak beni destekleyen ve anlayış gösteren Tez Danışmanın Yrd. Doc. Dr. Serkan Ayvaz'a teşekkürlerimi sunarım.

Bu tez çalışmasının araştırılmasında, yürütülmesinde ve sonuçlanmasında büyük katkıları olan aileme ve değerli arkadaşlarıma sonsuz teşekkürlerimi sunarım.

İstanbul, 2018

Semra YILDIRIM

ÖZET

TWİTTER VERİLERİYLE DUYGU ANALIZI VE TÜRKÇE DUYGU KÜTÜPHANESİ

Semra Yıldırım

Mühendislik Yönetimi

Tez Danışmanı: Yrd. Doc. Dr. SERKAN AYVAZ

Ocak 2018, 64 Sayfa

Düşünceler davranışlarımıza ilham verir ve insan aktivitelerinin merkezinde yer alırlar. Bizim gerçekte ilgili algılarımız, inançlarımız ve yaptığımız seçimler genelde başkalarının nasıl gördüğüne göre değişkenlik gösterir. Bu nedenle, karar vermemiz gerektiği durumlar da genelde başkalarının düşüncelerini dikkate alırız. Bu durum sadece bireysel olarak geçerli değildir; aynı zamanda toplumsal olarak da başkalarının düşüncelerini çok önemseriz.

Düşünceler, ahlaki değerler, tavırlar ve duygular birbirleriyle bağlantılıdır bu unsurlar duygu analizi ve düşünce madenciliğinin bileşenlerini oluştururlar.

Forumlar, bloglar, mikrobloglar ve sosyal ağlar alanlarındaki teknolojik ilerlemeler ve veri artışları, fikrî verilerin dijital ortamlarda aşırı derecede artışına neden olmuştur. Böylelikle 2000’li yılların başından itibaren duygu analizi çeşitli alanlarda araştırma konusu olarak ele alınmıştır. Bununla birlikte, text mining, data mining ve web mining konuları duygu analizi için kullanılmaya başlanmıştır. Bu konuların gelişimine katkı sağlanabilmesi için birçok yeni iş girişimleri ortaya çıkmıştır. Büyük şirketler kendi bünyelerine bu konuda bilgi sahibi olan kişilerle beraber özel departmanlar oluşturmuşlardır. Böylelikle sentimental analiz iş ve sosyal alanlarda kullanılmaya başlanmıştır.

Duygu Analizi, doğal dil ile oluşturulan belgelerin bilgisayar yazılımları tarafından işlenerek istatistiksel ve algoritmik çözümler ile analiz yapılmasına dayanmaktadır.

Bu tez kapsamında Twitter’den belli etiketlerle oluşturulan tweetler toplanmıştır ve bu tweetler üzerinden Duygu Analizi çalışmaları yapılmıştır. Analiz çalışmaları birbirinden bağımsız 3 farklı konu başlığında gerçekleştirilmiştir. Bu analiz kapsamında yapılan çalışmalardan birincisi hava değişikliklerinin kullanıcılar üzerinde ki etkisini ölçmek adına yapılmıştır bu çalışma kapsamında yaz aylarında ve kış aylarında toplanan veri kümeleri incelenmiştir. Bu kapsamda Türkçe tweet paylaşan kullanıcıların kış aylarında “olumsuz” tweetler paylaştıkları, bahar ve yaz aylarında ise kış aylarında paylaşılan

tweetlere göre daha “olumlu” paylaşımlar yaptıkları gözlemlenmiştir. Ayrıca kullanıcıların mevsim geçişlerinde yoğun duygulu tweetler paylaştıkları gözlemlenmiştir. Gerçekleştirilen analiz çalışmalarından ikincisi ise Türkçe tweet paylaşan kullanıcıların haftanın günlerine göre duygularında ki olumlu-olumsuz-nötr olma durumları ölçümlemeye çalışılmıştır. Bu çalışmalar sonrasında kullanıcıların Çarşamba günleri diğer günlere göre daha az oranda ve “nötr” tweetler paylaştıkları, Cuma Cumartesi ve Pazar günleri ise genelde “olumlu” duygu taşıyan tweetler paylaştıkları gözlemlenmiştir. Son yapılan analiz çalışması ise Türk televizyonlarında belirli bir süre boyunca hemen her gün yayınlanan, günlerce Trend Topic listesinin en üst sırasında yer alan survivor yarışma programı kitlesi tarafından Türkçe olarak paylaşılan tweetler üzerinden yapılmıştır. Burada amaçlanan kullanıcıların takip etmekte oldukları bu programla ilgili “olumsuz” sonuçlar görüntülemektir ve karşılaşılan sonuç genelde “olumsuz” tweetler olmuştur. Bu çalışma kapsamında ispat edilmek istenen durum: kullanıcıların bozuk Türk dili kullanma alışkanlıkları ve de genelde olumsuz duyguları paylaşma oranının olumlu duyguları paylaşma oranından yüksek olduğunun gözlenmesiydi. Analiz çalışmaları esnasında da kullanıcıların bozuk dil kullanımları ve eksik harf yazma alışkanlıklarına çözüm üretebilmek düzenlemeler ve tahminlemeler için algoritmalar yazılmıştır. Tüm bu deneyler ve analizler için tek bir Duygu Kütüphanesi kullanılmıştır. Ayrıca Türkçe Düşünce Kütüphanesi üzerinden duygulara verilen değerleriyle her bir tweet için sonuç üretilmiştir.

Anahtar Kelimeler: Duygu Analizi, Sentimental Analiz, Türkçe Düşünce Kütüphanesi, Türkçe Lexicon, Twitter

ABSTRACT

SENTIMENTAL ANALYSIS BY USING TWEETS AND TURKISH LEXICON

Semra Yıldırım

ENGINEERING MANAGEMENT

Thesis Supervisor: Asst. Prof. SERKAN AYVAZ

Jan 2018, 64 pages

Thoughts inspire acts and they are central to human activities. Our sense of reality, beliefs and decisions are often influenced by other people's views and perceptions. Therefore, we generally consider others' opinions and thoughts when we are making a decision. This is not only the case for individuals, but also true for the society.

Thoughts, moral values, attitudes and feelings are all linked to each other, constitute the core components of sentimental analysis and opinion mining.

Technological advancements in fields of internet forums, blogs, microblogs and social networks have led to exponential growth of digital intellectual data. Thus, since the early 2000s, sentimental analysis has been considered as a research topic in various fields. Consequently, many mining techniques including text mining, data mining and web mining have been used for sentimental analysis. To help to improve this field, many start-ups have been emerged by entrepreneurs and contributed to the advancements in the field. Large corporations have created new departments consisting people who have knowledge about sentimental analysis. Thus, sentimental analysis has begun to be used for both professional and social use.

Sentimental analysis is based on processing the documents created with natural language using computer software and then analyzing the text by using statistical and algorithmic solutions.

Within the scope of this thesis, a sentimental analysis study has been conducted by collecting tweets based on specific hashtags. The topics of those experimental analyses have been selected from three different hashtags. The very first stage of this study was measuring the effects of weather changes on users' emotions. To achieve that, data set has been generated by collecting tweets in both summer and winter seasons and analyzed. It has been observed that, majority of users post more positive tweets in spring and summer seasons comparing to tweets they posted in winter. Furthermore, it is established that,

users, who post in Turkish, post more emotional tweets during mid-seasons. The second stage of the study was measuring users' emotions in different days of the weeks. According to this study, tweets of Turkish users show less neutral emotions than the other days of week, and generally, more positive feelings on Friday to Sunday. The last but not least stage of the study was related to a reality show that aired almost everyday on Turkish televisions, and stays at top of the trend topic lists during the show, and its effects on Turkish Twitter users' emotions. This study relies on "Survivor" followers have usually negative feelings. In briefly, this study demonstrates that Turkish Twitter users who follow 'Survivor', tends to use the language negatively. A Turkish emotional lexicon has been generated and used for all stages of the study.

Keywords: Emotion Analysis, Sentimental Analysis, Sentimental Dictionary for Turkish, Turkish lexicon, Twitter

İÇİNDEKİLER

TABLOLAR	x
ŞEKİLLER	xi
KISALTMALAR	xii
1. GİRİŞ	1
2. LİTERATÜR TARAMASI	5
2.1 SENTİMENTAL ANALİZ	5
2.2 SENTİMENTAL ANALİZDE KULLANILAN ARAŞTIRMA	7
SEVİYELERİ VE GRUPLAMA YÖNTEMLERİ	7
2.2.1 Doküman Seviyesi (Document Level)	7
2.2.2 Cümle Seviyesi (Sentence Level)	8
2.2.3 Varlık ve Görüş Seviyesi (Entity and Aspect Level)	8
2.2.4 Sözlüksel Benzeşim Yöntemi	9
2.2.5 İstatistiksel Gruplama Yöntemi	9
2.2.6 Kelime Yakalama Yöntemi	9
2.3 SENTİMENTAL ANALİZLE İLGİLİ YAPILAN ÇALIŞMALAR	10
2.3.1 Makine Öğrenmesi Gruplama Yöntemleriyle Yapılan Çalışmalar	11
2.3.2 Sözlüksel Benzeşim ve Kelime Yakalama Yöntemleriyle Yapılan Çalışmalar ...	12
2.4 TWITTER	13
2.5 TÜRKÇE TWEETLER ÜZERİNDEN YAPILAN DUYGU ANALİZİ	14
ÇALIŞMALARI	14
2.6 DOĞAL DİL İŞLEME	17
3. VERİ VE YÖNTEM	19
3.1 TEKNİK TASARIM DETAYLARI	19
3.1.1 Düşünce Kütüphanesi	21
3.1.2 Kullanıcı Ara yüzü (UI Layer) Detayları	25
3.1.3 İş Mantığı Katmanı (Business Logic Layer) Detayları	27
3.1.4 Veri Erişim Katmanı (Data Access Layer) Detayları	29
4. ANALİZ VE BULGULAR	31
4.1 HAVA DEĞİŞİKLİĞİ VE MEVSİMLERİN KULLANICILAR	34
ÜZERİNDEKİ ETKİSİNİN ANALİZİ	34
4.1.1 Kısıtlı Veri Kümesi Üzerinden Hava Durumuyla İlgili Analiz Çalışmaları	35
4.1.2 Büyük Veri Kümesi Üzerinden Mevsimsel Değişimlerle ilgili Analiz	42

Çalışmaları.....	42
4.2 HAFTANIN GÜNLERİNİN KULLANICILAR ÜZERİNDEKİ	44
ETKİSİNİN ANALİZİ	44
4.2.1 Deneyisel Kısıtlı Veri Kümesi Üzerinden Gün Bazlı Analiz Çalışmaları	45
4.2.2 Deneyisel Kapsamlı Veri Kümesi Üzerinden Gün Bazlı Analiz Çalışmaları	48
4.3 SURVIVOR İZLEYEN KULLANICILARIN DUYGU ANALİZİ	52
5. DEĞERLENDİRME	57
5.1 KISITLAMALAR VE EKSİKLİKLER	60
6. SONUÇ	62
KAYNAKÇA	65
EKLER.....	69
EK A1 Database Tabloları ve Detaylar	70
ÖZ GEÇMİŞ.....	72

TABLÖLAR

Tablo 3:1: Düşünce Kütüphanesi Örnek Set.....	24
Tablo 4:1: "hava" Kelimesine Göre Duygu Yoğunluğu Verileri.....	37
Tablo 4:2: "bahar ve yazbaslasın" Kelimelerine Göre Duygu Yoğunluğu Verileri.....	40
Tablo 4:3: Haftanın Günlerine Göre Duygu Yoğunluğu Verileri.....	47
Tablo 4:4 Kısıtlı ve Kapsamlı Veri Kümelerinden Bazı Değerler	52
Tablo 4:5: "survivor" Kelimelerine Göre Duygu Yoğunluğu Verileri	54

ŞEKİLLER

Şekil 2.2:1 Sentimental Analizin Gelişimi	7
Şekil 3.1:1 OOP Diyagramı	21
Şekil 3.1:2 Arayüz Ekran Görüntüsü	26
Şekil 3.1:3 Ara yüz Ekran Görüntüsü(Compare)	27
Şekil 3.1:4 Ara yüz Ekran Görüntüsü(About).....	27
Şekil 4.1:1 Kış Aylarında Atılan Tweetlerin Duygu Yoğunluk Dağılımı	36
Şekil 4.1:2 Bahar Aylarında Atılan Tweetlerin Duygu Yoğunluk Dağılımı	39
Şekil 4.1:3 Mevsimsel Olarak Paylaşılan Tweet Oranları	43
Şekil 4.1:4 Mevsimlere Göre Duygu Yoğunluğu	44
Şekil 4.2:1 Gün Bazlı Tweet Oranları Grafiği	45
Şekil 4.2:2 Günlere Göre Duygu Dağılımı Grafiği	46
Şekil 4.2:3 Gün Bazlı Tweet Oranları.....	49
Şekil 4.2:4 Günlere Göre Duygu Dağılımı	50
Şekil 4.2:5 Kapsamlı ve Kısıtlı Veri Kümeleri Tweet Paylaşım Oranları	51
Şekil 4.3:1 “survivor” Etiketine Göre Duygu Yoğunluk Dağılımı	53

KISALTMALAR

ASA	:	Arapça Sentimental Analiz
CSV	:	Comma-Separated Values
DA	:	Duygu Analizi
DAL	:	Data Access Layer
DB	:	Database/Veri tabanı
MSSQL	:	Microsoft SQL Server
NB	:	Naive Bayes
NLP	:	Doğal Dil İşleme
SA	:	Sentimental Analiz
SVM	:	Support Vector Machine
RDMS	:	İlişkisel Veritabanı
TDK	:	Türk Dil Kurumu
UIL	:	User Interface Layer

1. GİRİŞ

Milenyumdan sonra internet üzerinde oluşan devasa verileri yönetebilmek ve analiz edebilmek için insan gözüyle ya da dokümanlar aracılığıyla yapılacak olan işlemler çok yetersiz kalmıştır. Böylelikle bu toplanan verilerden hangilerinin duygu içerikli olduğunun anlaşılması ve bu duygu yüklerine göre gruplanması konuların bilim adamları çalışmalar yapmıştır. Yapılan bu çalışmalar sonrasında Duygu Analizi ortaya çıkmıştır. Duygu analizi, doğal dil ile oluşturulan metinlerin ya da belgelerin bilgisayar yazılımları tarafından işlenerek istatistiksel ve algoritmik çözümler ile analiz yapılmasına dayanmaktadır.

Duygu analizi büyük verinin analiz edilip, kullanılabilir verilerinin içinden anlamlı veriyi çıkartmak için oluşturulan yöntemlerdendir. “Sentimental Analiz (Sentimental Analysis)” ya da “Düşünce Madenciliği (Opinion mining)” gibi birçok isimlendirmeler kullanılmaktadır [1].

Duygu analizi (Sentimental Analiz), programatik olarak cümlelerin, dokümanın ya da metnin ifade etmekte olduğu duyguyu keşfedip analiz etme sanatına verilen addır [2].

İnternetin yaygın kullanılması ve her geçen gün büyüyen devasa bir dünya olması nedeniyle artık el yordamıyla takipler ya da kişiler üzerinden yapılabilecek olan takipler söz konusu olmaktan çıkmıştır. Örneğin daha öncelerde bir kurumla ilgili çıkan tüm haberleri takip etmek ya da bu iş için birkaç kişilik ekipler kurmak mümkün iken günümüzde söz konusu bile değildir. Her geçen gün yeni bloglar, haber siteleri ve paylaşım siteleri ortaya çıkarak bu devasa dünyayı zenginleştirmektedir. Sürekli güncellenen ve artan veri setleri artık eski yapılarında saklanamamaktadır. Bu büyük verileri saklayabilmek ve ihtiyaç dahilinde kullanabilmek ve ayrıca sonuçlar çıkartarak yeni yollar bulmak günümüzün en popüler konularından biridir. Sınırsız bilgi, yorum akışının olduğu bir dünyada, verileri sınıflandırma işini otomasyona bırakmaktan başka bir yol bulunmamaktadır.

Duygu analizi, kişilerin olaylar, hizmetler, kurumlar ya da başka kişiler hakkındaki duygularının ve düşüncelerinin incelenmesini ve buna uyumlu olarak sonuçlar üretilmesini sağlar [3].

İnsanlar tarih boyunca her zaman başkalarının kendileri hakkında yaptıkları yorumları merak etmişlerdir. Bu yorumlar firmalar hakkında, kişiler hakkında ya da dünyada

gerçekleşmekte olan olaylar hakkında olabilmektedir. Günümüzde bu tip verileri en güzel elde edebildiğimiz mecra sosyal medya ortamlarıdır. Örneğin: Twitter, Facebook gibi birçok platform bulunmaktadır.

Sentimental analiz modelleri Twitter üzerinde satış performans tahminleri, ürünleri kademelendirme, kamuoyu yoklamaları, oylama sonucu tahminleri, politik bakış açılarının keşfi, filmlerin gişe tahminleri gibi birçok alanda kullanılmaktadır. Bu çoklu verinin analiz edilebilmesi için çok fazla firma kurulmuştur. Bu firmalar ve çalışanlar birçok araştırmalar yapmışlardır ve günümüzde de bu çalışmalar her geçen gün büyüyerek devam etmektedir.

İnternet yaygınlaşmadan önce veriler dijital ortamda bulunmadığı için ve bu verilere ulaşım çok kısıtlıydı. İnternetin yaygınlaşıp kullanılmaya başlanmasıyla beraber insanlar diğer insanlar, kurumlar, ürünler gibi birçok konu hakkında ki görüşlerini ifade etmeye başlamışlardır. Günümüzde artık bu veriler üzerinden birçok analiz çalışması yapılabilmektedir.

Büyük kuruluşlar kendileriyle, marka değerleriyle ilgili yorumları takip edebilmek sektöründe değerini arttırmak ve reklamcılık paylarını öne çıkartmak için birçok sosyal medya ürünü kullanmaktadır. Bu nedenlerle “Sosyal Medya Uzmanlığı” adı verilen meslek dalı ortaya çıkmıştır.

Geçmişte sadece edebiyatçıların, şairlerin yazdığı şiirler okunurken günümüzde birçok isimsiz Sosyal Medya Fenomenleri bulunmaktadır ve bunları birçok kişi takip etmektedir. Bu kişilerin de etkilediği birçok kişi bulunmaktadır.

Son yıllarda artan sosyal medya kullanımı nedeniyle “sosyal duygu analizleri” çok popüler olmuştur. İnsanların sosyal medya üzerinden paylaştıkları metinlerden kişiler hakkında olumlu, olumsuz gibi analizler yapmak mümkün olmaktadır. Günümüzde kullanılan sosyal medya uygulamaları arasından en fazla duygu yüklü kelime içeren platformlardan biri Twitter’dır. Twitter¹ da kullanıcılar dünya da daha önceden gerçekleşmiş ya da şu anda gerçekleşmekte olan herhangi bir konu hakkında anında paylaşım yapmaktadır. Bu paylaşımlar sayesinde kitleler olumlu ya da olumsuz olarak etkilenmektedir. Bu etkilenmeleri ölçümlendirmek ve sonuçlar üretebilmek için sürekli veri akışı olan platform Twitter kullanılmıştır.

1 TWITTER, <https://twitter.com>, Retrieved September 24, 2017.

Bu tezde amaçlanan Türkçe için kapsamlı Duygu Kütüphanesi üretmek ve bunu hemen her konuda yapılabilecek analiz çalışmalarında kullanarak Türkçe için doğru çalıştığını ispat etmektir. Bu kapsamda daha öncelerden hazırlanan kaynaklardan ve Türk Dil Kurumu'ndan faydalanılmıştır. Ayrıca oluşturulan diğer kütüphanelerden bazıları [4] incelenmiş bunlara benzer mantıkla kütüphane oluşturulmuştur. Daha önceden yapılmışlara ek olarak kütüphaneye basit emoji ve puanlama altyapısı eklenmiştir. Kelime bazlı sonuç üretmek yerine cümlelerin tamamı üzerinden analiz çalışmaları yaparak sonuçlar üretilmiştir. Esnek bir dil olan Türkçe'de negatif anlam içeren kelimelerin cümle içindeki durumuna uygun olarak pozitif sonuç çıkartma durumlarını yönetebilmek konusunda her bir duygu içeren kelimeye değer verilmiş ve böylelikle de puanlama alt yapısı kurgulanmıştır.

Bu tez kapsamında Twitter da yer alan Türkçe veriler üzerinden duygu analizi yapılmaktadır. Türkçe Twitter kullanıcılarının belli etiketler toplanan verileri incelenmiştir. Etiketlerin(keyword) belirlenebilmesi için Twitter'ın Trend Topic listeleri ve çok takipçiye hitap eden kullanıcıların tweetleri incelenmiştir ve tüm bu çalışmaların ardından etiket seçimleri yapılmıştır. Bu incelenen veriler toplanarak veri tabanında saklanmıştır. Verilerde kullanıcı bilgisi, tweet metni, tarih, tweet'in aldığı beğeni sayısı gibi detaylı bilgiler yer almaktadır. Bu bilgiler sayesinde farklı etiketlerle (hava durumu, yazgelsin, ilkbahar, sonbahar, kış, pazartesi, salı, çarşamba, perşembe, cuma, cumartesi, pazar, survivor) ve farklı başlıklarda analiz çalışmaları yapılmıştır. Tüm etiketler için tek bir Düşünce Kütüphanesi kullanılmıştır.

Türkçe'nin esnek bir dil olması ve Twitter'da çok fazla kirli verilerin olması nedeniyle temiz veriler elde edebilmek için yaklaşık bir yıl boyunca farklı farklı etiketlerle tweetler toplanmıştır. Toplanan kayıtlar üzerinden incelemeler yapıp kütüphanede zenginleştirme ve iyileştirmeler gerçekleştirilmiştir. Bu incelemeler esnasında Türkçe tweet paylaşan kullanıcıların genelde İngilizce klavye kullanmalarından dolayı oluşan aksaklıklarla karşılaşmıştır. Bu durumları yönetebilmek adına algoritmalar üretilmiştir. Örneğin kelimelerin tamamı küçük harfe dönüştürülmüş ve ardından kütüphaneden kontrol edilmiştir. Tek tek boşluk kontrolleri ve noktalama işaretleri gibi kontroller yapılmıştır. Bu düzenlemelerin ardından bazı kelimelerin büyük küçük harfle yazımından

oluşan hatalar giderilmiştir. Buna ek olarak tweet metninde ki kelimeler hem Türkçe alfabeye hem de İngilizce alfabeye çevrilmiştir. Bu dönüştürmelerin ardından düşünce kütüphanesinden eşleşmeler kontrol edilmiştir ve sadece bir eşleme üzerinden puanla yapılmıştır. Kısaca hem İngilizce metne hem Türkçe metne puanlama yapılmamıştır. Ayrıca Türkçemizde yer alan ikilemelerden dolayı da oluşan hesaplama hatalarını yönetebilmek için de ek algoritmalar üretilmiştir. Her bir kelime grubuna sadece 1 puan alacak şekilde düzenlemeler gerçekleştirilmiştir. Bu çalışmalar sonrasında yapılan deneylerden anlamlı birçok sonuç elde edilmiştir. Örneğin: kış aylarında kişilerin daha umutsuz ve karamsar oldukları sonucu elde edilmişken, yaz aylarının gelmesiyle kişilerin daha umutlu ve baharı heyecanla karşılayan duygu değerlerine sahip oldukları sonucu görülmektedir. Bu sonuçlarda gerçeğe yakın istatistiki değerler olarak elde edilmiştir.

2. LİTERATÜR TARAMASI

İnternetin yaygın kullanımıyla beraber 2000’li yılların başlarından itibaren Duygu Analizi çalışmaları başlamıştır. Duygu analizi doğal dil ile oluşturulan belgelerin bilgisayar yazılımları tarafından işlenerek istatistiksel ve algoritmik çözümler ile analiz yapılmasına dayandığından Doğal Dil işleme, Yapay Zekâ, Makine Öğrenmesi ve Veri Madenciliği bakış açıları birleştirilerek binlerce çalışma yapılmıştır. Bu oluşturulan yöntemlerden bazıları: istatistiki işlemler, sözlüksel benzeşim ve kelime yakalama gibidir [5]. Kelime yakalama yöntemi metni benzersiz sözcüklerin varlıklarına göre gruplamaktadır. Sözlüksel benzeşim yöntemi var olan bir duygu kütüphanesi yardımıyla, duygu puanlaması yapıp belli değerlere göre gruplama yapmaktadır [6]. SVM (Support Vector Machine) ve NB (Naive Bayes) gibi kullanılan yöntemler ile de duygusal gruplamalar yapılır. Bunlar genel olarak istatistiksel gruplamalara örnek gösterilir [7]. Ayrıca karma gruplama teknikleri de kullanılmıştır [8]. Kelime yakalama ve istatistiksel yöntemler ya da kelime yakalama ile sözlüksel benzeşimler ya da hepsi bir arada kullanılarak da duygu analizi çalışmaları yapılabilmektedir. Sözlüksel benzeşim gruplama yönteminde duygu analizi çalışmaları yapmak kolay ve hızlı analiz edebilme imkânı sunmaktadır [3]. Makine öğrenmesi gruplama yönteminde ise duygu kütüphanesine ihtiyaç duyulmaktadır.

Bu tez kapsamında gruplama yöntemi olarak istatistiksel, sözlüksel benzeşim ve kelime yakalama yöntemleri karma olarak kullanılarak duygu analiz çalışmaları yapılmıştır. Ayrıca araştırma seviyesi olarak da cümle seviyesi ve varlık görüş seviyesi kullanılmıştır. Böylelikle kelime bazlı değil cümle bazlı sonuç elde edilmiştir.

2.1 SENTİMENTAL ANALİZ

İlk olarak sentimental analiz ve fikir madenciliği terimleri 2003 yılında ortaya çıkmıştır [9]. 2000’li yılların öncesinde bakış açısı etkileri, duygu sıfatları, metafor yorumlamaları gibi bazı fikirler ortaya atılmış, fakat esas çalışmalar daha sonra gerçekleştirilmiştir [10]. Sentimental Analiz ile ilgili olarak birçok isimlendirme kullanılmaktadır. Bunlardan bazıları aşağıdaki gibidir:

Sentiment analysis (Duygu Analizi), opinion mining (Düşünce Madenciliği), opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, review mining vb. [10].

En yaygın olarak kullanılan isim Sentimental Analiz ya da Duygu Analizidir. Fakat duygu analizi için sadece kelimelere ihtiyaç yoktur, duyguları ifade etmek için kullanılan emojiler sayesinde de duygu analiz çalışmaları yapılabilmektedir. Kişilerin mutsuzluk durumlarda kullandıkları “☹” ifadesi gibi mutluluk durumlarında da kullandıkları birçok emoji bulunmaktadır. Bu konularda birçok analiz çalışması yapılmıştır. Örneğin: Mohammed O. Shiha ve Serkan Ayvaz tarafından yapılan araştırmada R dili ile Twitter üzerinden toplanan verilerde kullanılan emojilerin Sentimental analize olan etkileri araştırılmıştır ve çıkan sonuçlara göre kullanıcıların genel olarak “olumlu” tweetlerinde emoji kullandıklarını keşfetmişlerdir [11].

Bing Liu’ya göre gerçek dünyada büyük organizasyonlar ve tüm iş dalları hizmetleri ya da ürünleri hakkında tüketicilerin görüşlerini bilmek isterler, aynı zamanda karşılıklı olarak da tüketici de ürünlerin mevcut kullanıcılar tarafından yapılan görüşleri bilmek ister [10]. Bu gibi nedenlerle duygu analizi çalışmaları birçok organizasyon ve iş yerlerinde hızla gelişmiştir ve her geçen gün daha da ivme kazanarak gelişme göstermektedir.

Ayrıca sosyal medyanın çığır açan büyümesiyle kişiler ve organizasyonlar daha iyi kararlar almak için daha sıklıkla sosyal medya kullanmaya başladılar. Bu nedenle sosyal medya üzerinden paylaşılan devasa veri çok büyük önem kazandı ve hem tüketici hem üretici bu verilerin anlamlandırılabilir sonuçlarını görmek istediler böylelikle de duygu analiz çalışmaları daha hız kazanmış oldu. Örneğin: Bugünlerde bir tüketici bir ürünü satın almak için arkadaş çevresi ya da ailesiyle sınırlı kalmıyor; diğer kullanıcıların yorumlarını okumak için halka açık(public) forumlardaki ya da bloglarda ki yorumları inceliyor ve yapılan duygu paylaşımları üzerinden kendi kararını veriyor.

Kısacası günümüzde birçok kurum, kamuoyu yoklaması ve anketler yerine web üzerinden araştırmalar yapılmasını ve düşüncelerin incelenmesini tercih etmektedir. Bu gibi büyük araştırmaları insanların manuel olarak yapması, duyguları analiz etmesi ve

tüm verilerden detaylı sonuç çıkartması çok zordur. Bu nedenle, Sentimental Analiz sistemleri oluşturulmuştur.

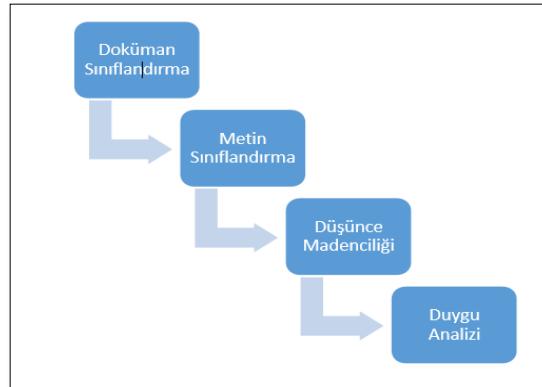
2.2 SENTİMENTAL ANALİZDE KULLANILAN ARAŞTIRMA SEVİYELERİ VE GRUPLAMA YÖNTEMLERİ

Sentimental analiz uygulamaları birçok alanda kullanılmaktadır. Tüketici hizmetleri, sağlık, siyasal seçimler, sosyal organizasyonlar, finans hizmetleri alanları örnek olarak verilebilir.

SA'da farklı araştırma seviyeleri ve gruplama yöntemleri kullanılmıştır. Bu gruplama yöntemleri ve araştırma seviyeleri sayesinde analiz çalışmaları yapılmıştır.

Sentimental analizin kronolojik gelişimi ilk olarak Döküman Sınıflandırma daha sonra Metin sınıflandırma ardından Düşünce Madenciliği ve son olarakta Duygu Analizi şeklindedir(Bkz Şekil 2.2:1). [12] :

Şekil 0:1 Sentimental Analizin Gelişimi



Araştırma seviyesi olarak 3 farklı seviye kullanılmaktadır [10]:

2.2.1 Doküman Seviyesi (Document Level)

Bu yöntemle bir dokümana göre tüm düşüncüyü negatif ya da pozitif sonuç çıkaran yöntemdir [3].

Örneğin: Bir ürün incelemesine göre sistem tüm kelimeleri kontrol eder ve sonuç olarak negatif ya da pozitif bir değer çıkartır.

Bu analiz biçimi birden çok ürünü ya da durumu karşılaştıran dokümanlar için uygun değildir. Çünkü birden fazla durum ya da ürün karşılaştırması yapıldığında birden fazla sonuç çıkması beklenmektedir. Halbuki bu yöntem ile sadece bir sonuç çıkmaktadır. Fakat olması gereken ise birden fazla sonuç çıkmasıdır. Bahsedilen her bir ürünün negatif olması ya da pozitif olması sonuçta yanılısamaya neden olmaktadır. Bu nedenle bu yöntem genellikle tek bir konuyla ilgili doküman ya da ürünler üzerinde yapılan araştırmalarda kullanılmaktadır.

2.2.2 Cümle Seviyesi (Sentence Level)

Bu yöntemle cümle bazında pozitif, negatif ya da nötr sonucunu çıkarma işlemi yapılmaktadır.

Nötr genelde duygu veya fikir belirtmeyen cümle olduğu anlamına gelmektedir.

Birçok objektif cümle düşünce içerebilir fakat sübjektif düşünce içermeyebilir bu durumu küçük bir örnekle açıklamak gerekmektedir [14].

Örneğin: “Geçen ay araba aldık ve arabanın cam sileceği düştü. “Cümlesi aslında olumsuzdur fakat cümle içinde olumsuzluk ifade eden bir kelime bulunmamaktadır. Bu nedenle de objektif bir çıkarım olarak görüntülenmektedir ve bir düşünce içermediği düşünülebilir ancak bu durum tam olarak böyle değildir [15].

2.2.3 Varlık ve Görüş Seviyesi (Entity and Aspect Level)

Doküman seviyesi ve cümle seviyesi analizleri insanların neyi sevip neyi sevmediğini tam olarak keşfedemez. Varlık ve Görüş Seviyesi ise küçük taneli analiz yapmayı sağladığından daha doğru analizler yapmayı sağlamaktadır [16].

Aspect Level daha öncelerde Feature Level (Özellik Seviyesi) olarak bilinmekteydi.

Bu analiz yöntemiyle dil yapılarına bakmak yerine direkt duygu ile ilgilenir. Bu durumda da Doküman ve cümle seviyesi analizlerden farklı olmasını sağlamaktadır [10].

Örneğin: “Servisi mükemmel olmamasına rağmen bu restorandı hala seviyorum.” Cümlesinin pozitif bir tonu olsa da tam anlamıyla pozitif olduğunu söyleyemeyiz. Burada Restoran’la ilgili pozitif düşünce olsa da negatif düşünce içermektedir.

Bu tez kapsamında varlık - görüş seviyesinde ve cümle seviyesinde analiz çalışmaları yapılmıştır. Bu tez kapsamında kelimelere değer verilmiştir. Bu değerlerin kümülatif toplamaları alınarak hesaplamalar yapılmıştır. Yukarıda bahsi geçen örnekte yer alan kelimelerin puanları aşağıdaki gibidir:

Mükemmel=5 puan, Seviyor=5 puandır. Cümlelerin puanı=+10 puan Olumlu olarak sonuçlanmaktadır.

Gruplama yöntemi olarak 3 farklı yöntem kullanılmaktadır [5]:

2.2.4 Sözlüksel Benzeşim Yöntemi

Sözlüksel benzeşim yöntemi var olan bir duygu kütüphanesi yardımıyla, duygu puanlaması yapıp belli değerlere göre gruplama yapmaktadır [6].

Sözlüksel benzeşim gruplama yönteminde duygu analizi çalışmaları yapmak kolay ve hızlı analiz edebilme imkânı sunmaktadır [17], [18].

2.2.5 İstatistiksel Gruplama Yöntemi

SVM (Support Vector Machine) ve NB (Naive Bayes) gibi yöntemler ile de duygusal gruplamalar yapılır [7]. Bu kullanılan yöntemler genel olarak istatistiksel gruplamalara örnek olarak gösterilmektedir.

Makine öğrenmesi gruplama yönteminde duygu kütüphanesine ihtiyaç duyulmaktadır. İstatistiki gruplama yöntemleriyle farklı diller ve farklı veriler için yapılmış birçok çalışma bulunmaktadır [3], [19], [20].

2.2.6 Kelime Yakalama Yöntemi

Kelime yakalama yöntemi metni benzersiz sözcüklerin varlıklarına göre gruplamaktadır [5]. Karma gruplama tekniklerinde de Kelime Yakalama yönteminden faydalanılmaktadır. Ek olarak Kelime yakalama ve istatistiksel yöntemler bir arada kullanılarak analiz çalışmaları yapılabilmektedir.

Ya da kelime yakalama, sözlüksel benzeşimler ve makine öğrenmesi yöntemlerinin hepsi bir arada kullanılarak da duygu analizi çalışmaları yapılabilmektedir [21], [22].

Bu tez kapsamında kelime yakalama, sözlüksel benzeşim ve istatistiksel gruplama yöntemleri kullanılmaktadır.

2.3 SENTİMENTAL ANALİZLE İLGİLİ YAPILAN ÇALIŞMALAR

Duygular sübjektif hisler ve düşüncelerdir. Duygular üzerine psikoloji felsefe gibi bugüne kadar birçok alanda çalışmalar yapılmıştır. Bu çalışmalar çok geniş kapsamlıdır. Örneğin: psikolojik tepkiler (kan basıncı, terleme, psikolojik tepkiler, mimikler gibi). Bilim adamları insanların duygularını kategorize etmişlerdir. Buna rağmen mutabık kalınmış kesin çizgiler yoktur.

Ayrıca günümüzde duygu analizi politik ve siyasi konular içinde kullanılmaya başlanmıştır [23]. Bu günlerde artık oy verecek kişiler siyasi parti adaylarıyla ilgili olarak diğer insanların görüşlerini bilmek istemektedirler. Geçmişte bu gibi düşünce ihtiyaçları aile, arkadaş ve yakın çevreyle sınırlıydı. Fakat şimdilerde web üzerinden her türlü bilgi kolayca elde edilebilmektedir. Bu gibi ve dahası amaçlara binaen toplumsal ya da tüketicilerin düşüncelerini öğrenmek için anketler, kamuoyu yoklamaları yapılmaktadır. Var olan bütün verileri bir araya getirip bu veriler üzerinden sonuçları görüntüleyebilmek için Duygu Analizi çalışmaları yapılmaktadır [24]. Bunun en büyük örneklerinden birkaçı olarak Amerika Seçimleri ve Almanya Seçimleri gösterilmektedir [25]. Örnek 1: Eylül 2009 yılında Almanya seçimleri için Angela Merkel ve ekibi tarafından Twitter üzerinden duygu analizi çalışmaları yapılmıştır [26]. Honeycutt ve Herring'in 2009 da yapmış oldukları araştırmalar sonrasında direkt olarak kişiler referans(@mention) olarak gösterdiklerini keşfetmişlerdir. Böylelikle analiz çalışmalarında sonuçlar çıkartmışlardır [27].

Örnek 2: 2008 yılında Barack Obama ve ekibi tarafından Twitter, Facebook², MySpace³ gibi birçok sosyal medya platformundan veriler toplanmış ve bunlar üzerinden seçim sonuçları tahminleyebilmek için analiz çalışmaları yapılmıştır [18], [28], [29]. Obama bu

2 Fleming, Ryan (December 29, 2011). "Google and Facebook top the most visited websites of 2011". Digital Trends.

Retrieved July 13, 201

3 "MySpace History". FreeMySpace. Retrieved July 24, 2010.

ve ekibi bu başarılı sosyal medya analizi sayesinde 500 milyon dolar bağış elde etmiştir [30], [31].

Son yıllarda fikir içeren sosyal medya paylaşımlarının organizasyonları tekrar şekillendirdiğine toplumsal düşünce ve duyguları sarstığına şahit olduk fakat sosyal medya toplum da ki ayaklanmaları sağlayan platform değildir sadece kişilerin iletişimlerini hızlandıran ve paylaşılan bilgilerin daha kısa sürede daha çok kişiye ulaşmasına yardımcı olan platformlar bütünüdür [32], [33]. Örneğin:17 Aralık 2010'da bazı Arap ülkelerinde sosyal medyanın katkılarıyla politik değişiklikler olduğu görülmüştür [34]. Arap ülkelerinde ki sosyal medya kullanımının artması isyanların büyümesine ve halkın örgütlenmesine katkı sağlamıştır. Bu nedenle sosyal medya devrimi olarak adlandırılmıştır. Bu durum sonucunda da artık sosyal medya üzerinde ki bu fikirleri toplama ve bunları analiz etme zorunluluğu ortaya çıkmıştır.

2.3.1 Makine Öğrenmesi Gruplama Yöntemleriyle Yapılan Çalışmalar

Pang ve Lee SVM algoritmaları ile yıldız (*) ve sembollerle etiketlenmiş verilerde gruplama yapma problemini araştırmışlardır ve oluşturdukları çoklu sınıf SVM regresyon algoritmaları ile yüzde 60 gibi bir başarı elde etmişlerdir [19]. Pang ve Lee daha sonra ise veriden öznel ifadeleri çıkartarak tekrardan hesaplamalar yapmış ve ortalama yüzde 85 olarak başarı oranlarını güncellemişlerdir.

Ayrıca Pang yaptığı bir diğer çalışmada ise NB ve ME (Maximum Entropy) ve SVM algoritmalarıyla analiz çalışmaları yapmıştır. Elle seçtiği 1400 yorumlu veri kümeleri içinden duygu analizi çalışması yapmış ve yüzde 82 başarı elde etmiştir [3].

Turney web üzerinden topladığı 410 yorumu, cümlelerin polaritesine göre sınıflandırmıştır. PMI-IR algoritmasını kullanmıştır. Turney, bu araştırmasında 4 farklı konu üzerine yoğunlaşmıştır. Çalışma sonucunda yüzde 74 gibi bir başarı elde etmiştir [35].

Arapça Sentimental Analiz (ASA) çalışmalarının popülerliği hızla artmaktadır. Bu nedenle de Duwairi çalışmasında diyalektik kelimeler kullanarak Arapça tweetler üzerinden sentimental analiz çalışmaları gerçekleştirmiştir. Makine öğrenmesi gruplama yöntemi ile analiz çalışmaları yapmıştır [36]. Ayrıca Arapça için NLP ile de birçok çalışmalar yapılmıştır [37].

Shankar Setty Facebook’da bulunan Haber Akışı (New Feed) üzerinde Gmail’in “Gelen Kutusunda” (Inbox) kullanıldığı gibi gruplama önerisinde bulunmuştur. Önerdiği bu yapı Haber Akışından gönderilen fikirlerin önem durumlarına göre sınıflanması üzerinedir [17].

Wöllmer, Wenninger, Sun ve ekibi tarafından YouTube’da özellikle film incelemeleri içeren videolar üzerine çalışmalar gerçekleştirmiştir. Yaptığı çalışma videoda bulunan konuşmayı alıp bu konuşma üzerinden duygu analiz çalışmaları yapmayı hedeflemiştir [38].

2.3.2 Sözlüksel Benzeşim ve Kelime Yakalama Yöntemleriyle Yapılan Çalışmalar

Steinberger ve grubu çalışmalarında farklı dillerde duygu kütüphanesi oluşturabilmek için yarı otomatik bir yöntem oluşturmuşlardır. İki farklı dilde oluşturulan sözlükler otomatik algoritmalar sayesinde üçüncü bir dile çeviri kabiliyeti kazandırmıştır. Çalışmalarında İspanyolca, Arapça, İtalyanca, İngilizce ve Rusça dillerini kullanmışlardır. Türkçe için de çalışmalar başlanmış fakat tamamlanamamıştır. Yaklaşık 2000 civarı kelimedenden oluşan sözlükler insanlar kontrolü ile oluşturulmuş ve ortalama yüzde 76 başarı yakalanmıştır [39].

Hu ve grubu el yordamıyla topladıkları olumlu ve olumsuz kelimeleri çalışmalarında listelerler [16]. WordNet gibi zıt ve eş anlamlıları gösteren bir sözlük sayesinde listelerini genişletirler [21]. Daha sonra el yordamıyla bu liste düzenlenir ve hatalı kayıtları istatistiksel yöntemlerle temizlerler. Bu yöntemle yüzde 84 başarı elde etmişlerdir. Kelimelere gelen eklerle zıt anlam tespit etmeyi deneyen Mohammed ise bazı zıt anlamlı kelimeleri bu listeye eklemeyi başarmıştır [40]. Daha sonrasında ise Kamps kullandığı

farklı teknikler ile mesafe tabanlı algoritmaları ve her kelimenin olumlu ya da olumsuzluk durumunu mutlak uzaklıklarına göre tespit etmeyi denemiştir [41].

Esuli ve Sebastiani yaptıkları çalışmalarda makine öğrenmesi yöntemiyle Turney ve Kamps tarafından oluşturulan kelime kümesini kullanmışlardır [22]. Sözlükteki tüm kelimeleri negatif ve pozitif olarak sınıflandırarak, negatif ve pozitif puanlarını ve nesnel puanlarını hesaplamışlardır. Makine öğrenmesi teknikleri ile WordNet sözlüğündeki tüm kelimelere nesnel puanlarını atamışlardır. Böylelikle genel bir duygu sözlüğü olan SentiWordNet'i yayınlamışlardır [42].

2.4 TWITTER

Twitter 2006'lı yıllarda Jack Dorsey tarafından küçük bir proje grubunun iletişimi sağlamak amacıyla SMS sağlayıcı olarak ortaya çıkmış bir mikroblogdur.

2016 analiz sonuçlarına göre yaklaşık olarak 695 milyon Twitter hesabı bulunmaktadır. Bu hesaplardan yüzde 40'ının sadece izleme amacıyla kullanıldığı ve hiç tweet paylaşmadıkları paylaşılan istatistiklerde görüntülenmektedir [43].

Twitter 2017 Eylül ayına kadar 140 karakterle sınırlı olarak insanların duygu ve düşüncelerini paylaştıkları sosyal mesajlaşma ağı olarak bilinmekteydi fakat 26 Eylül ayında yapılan değişiklikle karakter sayısı Japonca, Korece ve Çince hariç geriye kalan tüm diller için iki kat arttırarak 280 karakter olarak güncelledi [44].

Twitter insanlar arasında etkileşim aracı olarak kullanılmaktadır. Twitter üzerinden toplanan verilerle birçok araştırmalar yapılmıştır. Bu çalışmalardan bazı örnekler:

2008'de Twitter Barack Obama sosyal medya analizlerini kullanmıştır [18], [34], 2009 Mayıs/Aralık ayı içerisinde atılmış olan tweetlerden salgın hastalıkları önceden tahminleyen [45], 2009 Mayıs ile 2010 Ekim tarihleri arasında atılmış olan tweetlerden ilaçlar ve onların bilinmeyen yan etkilerinin neler olabileceğini analiz eden [46], [47], sosyal medya dinamikleri ile insan algısının zamanla değişimini zaman serilerini kullanarak yüzde 85 başarı ile tahminleyen [48], turistik bir şehre (Meksika-Cancun) gelen turistlerin atmış olduğu 70 milyon tweet üzerinden algı analizi yapabilen [49] bir çok çalışma bulunmaktadır.

Twitter diğer sosyal mecralara kıyasla metin ağırlıklı platformdur. Ayrıca Tweetlerde kullanıcıların paylaştığı metinler genelde duygu yüklüdür bu nedenle de çalışmalarda Twitter kullanmak daha faydalı olmaktadır. Örneğin Facebook, Instagram, Snapchat gibi sosyal paylaşım siteleri daha çok bir resmin üzerinden duygu analizinin yapılmasına uygundur, bu da bambaşka bir süreç gerektirmektedir. Bu nedenlerden dolayı Twitter üzerinden duygu analizi çok fazla yapılmaktadır.

2.5 TÜRKÇE TWEETLER ÜZERİNDEN YAPILAN DUYGU ANALİZİ ÇALIŞMALARI

Türkçe tweetler üzerinden birçok deney ve araştırmalar yapılmıştır. Fakat bu çalışmalar esnasında birçok engelle karşılaşmıştır. Bu engeller genel olarak iki grup altında incelenebilir. İlk engel dil yapısıdır ve ikinci engel ise kullanıcıların alışkanlıklarıdır. Türkçe'nin diğer dillerden daha farklı yapıya sahip olması, daha esnek olması ve kullanıcıların yaygın olarak İngilizce klavye kullanma alışkanlıkları nedeniyle Türkçe duygu analiz yapma çalışmaları çok zaman almaktadır. Ayrıca toplumumuzun genelde iyi düşünceleri paylaşmak yerine kötü düşünceleri paylaşması da yapılan analiz sonuçlarının genelde negatif sonuçlar doğurmasına neden olmuştur. Ek olarak kullanıcıların Twitter'da yer alan karakter kısıtlaması nedeniyle kelimelerin bazı harflerini eksik yazma alışkanlıkları bulunmaktadır. Bunca engellere rağmen Türkçe kelimeler üzerinden duygu analizi konusunda birçok çalışma yapılmıştır ve halen de çalışmalar yapılmaktadır. Bu konuda firmalar ve üniversiteler uzmanlaşmak için departmanlar kurmuştur. Çünkü internette yer alan binlerce satırlık Türkçe verinin analiz edilebilecek olması ve böylelikle yeni hizmetlerin ve iş dallarının oluşacak olması tüm araştırmacılar için ve sektörler için fayda içermektedir.

Yapılan bu çalışmalarda farklı farklı gruplama teknikleri, analiz seviyeleri ve teknolojiler kullanılmıştır. Çalışmalar farklı teknikler ve farklı metodolojilerle gerçekleştirilmiştir bazıları R programlama diliyle, bazıları C# programlama dilleriyle kodlanmıştır. Çalışmalarda farklı farklı veri tabanları kullanılmıştır. Bazıları İlişkisel veri tabanları (RDMS) kullanmıştır bazıları ise İlişkisel Olmayan veri tabanları (NoSQL) kullanmıştır. Bu durum tamamen analiz edilmesi planlanan dataya ve teknik ekibin bilgisine göre belirlenmektedir.

Yapılan bazı çalışma örnekleri aşağıda yer almaktadır:

Erkut Evirgen tez çalışmasında Türkçe tweetlerde duygu analizi konusunda araştırmalar yapmıştır. Tez çalışması kapsamında R programlama dilini kullanmıştır. Tweetlerde ki düzgün formatlı olmayan kelimeleri düzenleyip işlediğinden bahsetmiştir. Ayrıca da yaygın makine öğrenme metotları olan SVN, Karar Ağaçları, Boosting, Maksimum Entropy, Yapay Sinir Ağları karşılaştırmalarını yapmıştır [50].

Dilek Kayahan Aslı Sergen ve Banu Diri yaptıkları duygu analizi çalışmalarıyla Twitter üzerinden Türkiye de ki TV programlarının reytinglerinin tahminlemesini gerçekleştirmişlerdir. Bu çalışma için kendi belirledikleri hashtagler(etiketler) sınıflandırılmış ve sınıflandırmalar sonrasında da duygu kütüphanesinden kontroller yapılmıştır [51].

B. İbrahim Sevindi Türkçe film yorumlarının duygu kutupları çeşitli yöntemler kullanılarak belirlemeye çalışmıştır. Makine öğrenmesi ve sözlük tabanlı yaklaşımlar kullanılarak karşılaştırmalar yapmıştır. Türkçe film yorumlarının duygu analizinde makine öğrenmesi yöntemleri daha başarılı olması için sonuçlar üretmiştir [20].

Mehmet U. Şimşek Twitter verileri her gün için ayrı olarak duygu belirten kelimelerin kullanılmasına göre analiz etmiştir. Türk Twitter kullanıcılarının duygusal durum değişiminin IMKB indeksi ile ilişkili olup olmadığına araştırmıştır. Twitter mesajlarını analiz etmek için frekans yöntemi ve ortalama mutluluk analizi yöntemini kullanılmış ve bu iki metot karşılaştırmıştır. Tweet veri seti içerisinde mesajları olan kullanıcıların yaptıkları yorumlara göre benzerlikleri incelemiştir. Bu tez çalışması, Türkçe tweetler ve borsa indeksi üzerinde yapılmış ilk çalışma olması açısından önem taşımaktadır [13].

Yasin Görmez ve ekibi, IMDB’de yer alan film yorumlarını içeren veri seti üzerinde çeşitli makine öğrenme teknikleri kullanarak duygu analizi çalışması yapmıştır. Makine öğrenmesinde sıkça kullanılan sınıflandırma algoritmalarından yapay sinir ağları, SVM, NB ve merkez tabanlı sınıflayıcı algoritmalarını kullanılmıştır [52].

Sinem Demirci tarafından yapılan arařtırmalar esnasında “finer-grained emotion analysis/paralı duygu analizi” ynteminin zellikle Trke kaynaklı yapılan analiz alıřmalarının da etkili olduėu gzlemlenmiřtir. Trke iin hali hazırda bulunabilecek bir veri kmesi kullanılması, duygu analizi iřleminde yetersiz kalmaktadır. Bu durumun nne geebilmek amacıyla, kullanılan kelimeler duygulara gre sınıflandırılmış ve bu veri kmeleri yardımıyla tweet gnderen kiřinin duygu analizi gerekleřtirilebilmiřtir [53].

Eyp S. Akgl, Caner Ertona ve Banu Diri taraflarınca yapılan arařtırmalarda Twitter verileri kullanılmıştır. Duygusal Twitter programları aracılıėıyla veriler analiz edilmiřtir ve gruplanmıřtır. Gruplamalar “olumlu”, “olumsuz” veya “ntr” olarak yapılmıřtır. Bylelikle Duygu Ktphanesi kullanılmıştır [4].

Bu arařtırma kapsamında oluřturulan szlkten bu tezdten Dřnce Ktphanesi zenginleřtirme ařamasında yararlanılmıştır. Duygusal tweetlerden farklı olarak kelimelere deėerler verilmiřtir ve bu duygu deėerleri sayesinde tam metnin puanı hesaplanmıřtır bunlara ek olarak emojiiler de puanlanmıřtır. Ayrıca Duygusal Twitter’den farklı arařtırma yntemi kullanılmıştır. Bu tez kapsamında Varlık ve grř seviyesi analiz yntemi kullanılmıştır.

nder oban, Barıř zyer ve Glřah Tmkl zyer taraflarınca yapılan arařtırmalar metin sınıflandırma alanında kullanılan temel yntemler kullanılarak Twitter ortamından elde edilen mesajlar zerinde duygu analizi yapılabilirliėi kanıtlanmış ve literatre temel yntemler kullanılarak yapılan alıřmalarda elde edilen bařarı oranı Trke iin yakalanmıřtır. Sonular duygu analizi alıřmalarının bir metin sınıflandırma problemi olarak ele alınabileceėini gstermekle beraber bařarı oranının ykseltilebilmesi iin harici yntemler uygulanması gerektiėi keřfedilmiřtir. Bu arařtırmadan elde edilen bulgulara gre, deneysel sonular Twitter mesajlarının makine ėrenmesi yntemleriyle sınıflandırılabilceėi tezini doėrulamıřtır [54].

Nazan ztrk ve Serkan Ayvaz tarafından gerekleřtirilen Sentimental analiz alıřmalarında Suriyeli mltecilerle ilgili olarak İngilizce ve Trke tweetler R

programlama dili yardımıyla toplanmış ve bu toplanan tweetlerden yapılan analiz çalışmalarında çıkan sonuçlar 2 farklı şekilde Türkçe paylaşılanlarının daha pozitif oldukları gözlenlenmiştir. İngilizce paylaşılan tweelerin ise genelde nötr ve negatif oldukları sonucuna varmışlardır. Bu araştırma kapsamında da Sentimental Lexicon kullanılmıştır [55].

2.6 DOĞAL DİL İŞLEME

Doğal Dil İşleme (Natural Language Processing), bilgisayar bilimi, yapay zekâ ve bilişimsel dil biliminin birleşiminin bir araya gelmesi ile doğmuş bir alandır. Bu alanın asıl kaynağı bilgisayar kullanarak iki farklı dili birbirine çevirme işlemine (machine translation) olan meraktan kaynaklanmaktadır. 1954 yılında Georgetown deneyi olarak adlandırılan Georgetown Üniversitesi ve IBM'in ortaklaşa olarak geliştirdiği bilgisayarlı çeviri deneyinde 60'dan fazla Rusça cümlelerin başarılı bir şekilde İngilizceye çevrilebildiği kaydedilmiştir [56]. Bu gelişmenin ardından, her ne kadar bilgisayarlı çeviriye duyulan ilgi artmış olsa da süreç boyunca yaşanan olumsuz gelişmeler birçok yatırımcıyı da alandan uzaklaştırmıştır. 1980'li yıllarda ise hesaplama gücünün hızla artması ile Doğal Dil İşleme (NLP) alanındaki gelişmeler hızlanmıştır. Bilgisayarların hızlanması, öğrenme algoritmalarının dil işlemede kullanılmaya başlamasında büyük bir etki göstermiştir. İlk başlarda kullanılan bilgisayar öğrenme algoritmaları basit ve elle kodlanan durum setlerinden (if-else) oluşsa da ilerleyen yıllarda hızla gelişmiştir.⁴ NLP metin tabanlı çalışmalarda çoğu zaman kullanılmaktadır. İstatistiksel olarak metnin üzerinden sonuçlar üretmeyi kapsamaktadır. Genellikle yapay zekâ altındaki dil bilim bilgisine dayalı çalışmaları kapsamaktadır [57].

Genel olarak NLP kullanım alanları aşağıdaki gibidir [58]:

- i. Gruplama ve sınıflandırma (spam detection)
- ii. Metni etiketleme ve metni parçalara ayırma nesne, fiil sıfat vs. (PartOfSpeech Tagging)
- iii. Varlık ismi tanımlama (Named Entity Recognition)

4 Goldberg, Yoav (2016). A Primer on Neural Network Models for Natural Language Processing. Journal of Artificial Intelligence Research 57 (2016) 345–420

Türkçe üzerinden NLP çalışmaları ancak son 15-20 yıldır yapılmaktadır. Türkçe NLP, Türkçenin yapısı nedeniyle sorunludur. Her ne kadar bu konuda birçok çalışma yapılsa da tam bir kaynak oluşturulamamıştır. Fakat son 10 yılda hem Türkiye de hem de yurtdışında yapılan çalışmaların daha da hız kazanması ümit vaat etmektedir. Örneğin: Kemal Oflazer bu konuda birçok araştırma yapmıştır. Bu çalışmasında Türkçe'nin doğal dil işlemede yarattığı zorluklarından bahsetmiştir [59].

Bu tez kapsamın sözlük el yordamıyla oluşturulduğu için ve eklemelerinde analizler sonrasında yapılmasından dolayı gruptama-sınıflandırma işlemleri için NLP kullanılmamıştır.

Türkçe kelimelerin İngilizce klavye ile yazılması durumunda oluşacak olan versiyonlarının tespiti aşamasında NLP benzeri bir yapı geliştirilerek kelimelerin İngilizce klavyeyle yazılmaları tahminlenmiştir.

3. VERİ VE YÖNTEM

Bu kısımda geliştirilen programdan, veri tabanından, algoritmaların çalışma mantıklarından ve düşünce kütüphanesinin yapısından bahsedilmektedir.

Gruplama yöntemi olarak kelime yakalama, sözlüksel benzeşim ve istatistiksel kullanılmıştır. El yordamı ile düşünce kütüphanesi oluşturulmuş ve her bir kelime, kelime grubu ve basit emojiler olumlu ve olumsuz olarak gruplanmıştır. Kelimenin nötr olma durumu söz konusu değil iken cümlelerin nötr olma durumu mümkün olabilecek şekilde hesaplamalar yapılmıştır. Bu tip hesaplamaların tamamı için algoritmalar geliştirilmiştir. Gereksinim dahilinde düşünce kütüphanesine el ile ekleme ve güncelleme işlemleri yapılmıştır. Böylelikle hem cümle seviyesinde (Sentence level) hem de varlık ve görüş seviyesinde (Entity and Aspect level) analiz yapma kabiliyeti kazandırılmıştır. Ayrıca Türkçe karakterlerde yaşanan problemler için yeni algoritmalar üretilmiş ve sözlükte de düzenlemeler yapılmıştır.

3.1 TEKNİK TASARIM DETAYLARI

Bu tez kapsamında kullanılacak veriyi çekebilmek, yönetebilmek ve saklayabilmek için birçok araştırma ve inceleme yapılmıştır. Bu incelemelerin ardından performans olarak yüksek, verimli, esnek, kaynak çokluğu gibi kriterler belirlenmiştir. İlk önce kodlar C#.NET ile geliştirilmiştir. Daha sonra ise yaşanan performans sorunları nedeniyle kodlama dili değiştirilmiş ve Python kullanılmaya başlanmıştır.

C#.NET, Microsoft tarafından üretilmiştir. C ve C++ dillerinin etkileşimiyle türetilmiş nesne yönelimli yazılım dilidir ve .NET framework'ü üzerinde çalışmaktadır. Python, Guido Van Rossum tarafından geliştirilen modüler yazılım dilidir. Açık kaynak kod lisanslıdır. Nesne yönelimli programlama, fonksiyonel ya da yapısal programları geliştirebilmeyi desteklemektedir.⁵

⁵ "The History of Python: A Brief Timeline of Python". Retrieved 20 March 2016

Veri tabanı olarak SQL Management Studio ve SQLite kullanılmaktadır. SQL Management Studio Microsoft tarafından üretilmiş, ilişkisel veri tabanı yönetim sistemidir (RDMS).⁶ SQLite kaynak kodları tamamen açık olan ve hem işlemsel hem ilişkisel olan veri tabanıdır.⁷

Çok katmanlı mimari(OOP) kullanılmaktadır. Object Orianted Programming- Nesne Yönelimli Programlama metodolojisi.

Pyhton dilinin kullanım kolaylığı, modüler ve dinamik yapısı dolayısıyla bu tez kapsamında kullanılmıştır. Esas amaç Twitter da yer alan verilere en hızlı şekilde ve sorunsuz şekilde erişim sağlamaktır. Bu nedenle kodlamalar için Ubuntu 16.04⁸ versiyonu kullanan bir sanal makine kullanılmıştır ve böylelikle Windows işletim sisteminden çok daha performanslı olduğu gözlemlenmiştir.

Bu tez kapsamında 3 Katmanlı yapı esas alınmıştır. Bu katmanlar: User Interface(UI) Layer, Business Layer ve Data Access Layer'dır. Teknik mimari diyagramı Şekil 3.1:1 de görüntülenmektedir. Kodların tamamı Github üzerinden paylaşılmıştır.⁹ Ayrıca bu tez kapsamında Düşünce Kütüphanesi geliştirmeleri yapılmıştır.

6 <https://docs.microsoft.com/en-us/sql/database-engine/configure-windows/database-engine-instances-sql-server> Retrieved 15

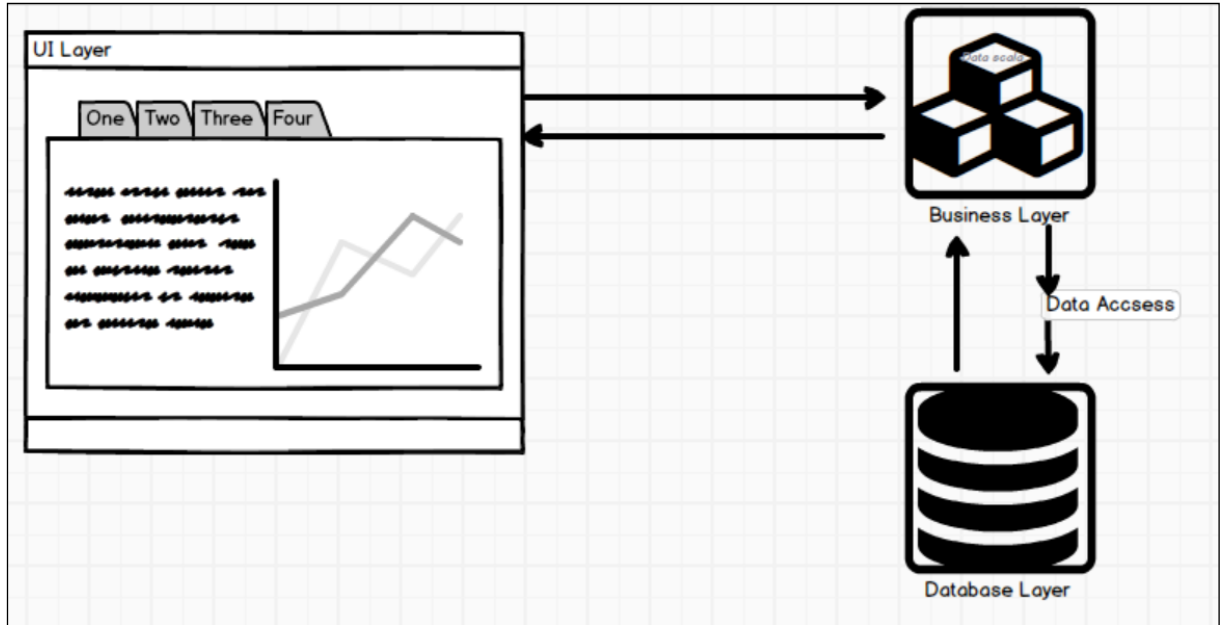
Jan 2018

7 <https://www.sqlite.org/about.html> Retrieved 15 Jan 2018

8 "About Ubuntu. The Ubuntu Story". Ubuntu.com. Canonical Ltd. Retrieved 21 August 2012

9 <https://github.com/SemraYildirim/SENTIMENTAL-ANALYSIS-BY-USING-TWEETS-AND-TURKISH-SENTIMENTAL-LEXICON>

Şekil 3.1:1 OOP Diyagramı



3.1.1 Düşünce Kütüphanesi

Düşüncelerin en önemli göstergeleri düşünce içeren kelimelerdir. Bu kelimeler genellikle pozitif ya da negatif düşünce belirtmek için kullanılırlar. Duygu Analizi bir cümlenin ya da bir metnin duygu içerip içermediğini ve bu duygunun pozitif ya da negatif olma durumunun saptanması süreci olarak da adlandırılır [60].

Duygu barındıran metinler genellikle görüş ya da değerlendirme içerirler. Bu görüş ve değerlendirmeler herhangi bir konu, şahıs, marka ya da siyasi görüş hakkında olabilir. Psikoloji bilim dalı DA’nde sıklıkla kullanılmıştır. Kişilerin duygusal durumuna göre kullandığı kelimeler ve bu kelimelerin kullanma tarzları, araştırılmıştır. Daha sonra uygulanan seanslarda hastaların bu kelimeleri kullanım aralıkları ve şekilleri incelenmiş ve tespitler üzerine çalışmalar yapılmıştır [61].

Ayrıca toplumsal ve sosyal araştırmalar kapsamında yapılan anketler ve incelemelere bakıldığında DA’nin sosyoloji bilim dalını da yakından ilgilendirdiği görülmektedir. Örneğin, toplumsal tepkilerin ya da ayaklanmaların yoğun olduğu dönemlerde genellikle ilgili konularda olumsuz bir dil kullanıldığı görülmektedir. Bu tepkiyi doğuran kişi,

kurum ya da kuruluşlar hakkında birçok duygu barındıran metin sosyal medya üzerinden incelenmesi durumunda da keşfedilebilir ve nihai olarak analiz edilebilir. Bu gibi birçok amaçlarda da duygu barındıran kelimeler duygu yönelimlerine uygun olarak gruplanmış ve taşıdıkları duygu yoğunluklarına uygun olarak puanlandırılmıştır. Örneğin: Güzel, harika, mükemmel pozitif duygu içeren kelimelerdir. Kötü, zayıf, korkunç ise negatif duygu içeren kelimeleridir. Bunun gibi kelimelerin ve ifadelerin liste olarak saklanması sonrasında duygu kütüphaneleri (sentiment lexicon) oluşturulur. Duygu kütüphaneleri DA’nde çok büyük bir değer taşımaktadır.

Geçen yıllar boyunca araştırmacılar sentimental kütüphaneleri derlemek için birçok algoritmalar tasarlamışlardır [62]. Düşünce kelimeleri 2 türde olurlar: baz (base) kelimeler ve karşılaştırmalı (comparative) kelimeler. Örneğin: İyi, kötü baz kelimedir. Daha iyi, daha kötü karşılaştırmalı kelimedir [63].

Bu tez kapsamında düşünce kütüphanesi oluşturulmuştur. Kütüphanenin zenginleştirme süreci ise aşağıdaki gibidir:

- i. Türk Dil Kurumu: Kütüphanesinin zenginleştirme çalışmalarında TDK¹⁰ üzerinden birçok kelime incelenmiş ve duygu analiz çalışmalarına katkı sağlayacağını öngörülenler ayıklanarak Düşünce Kütüphanesine ekleme yapılmıştır.
- ii. Daha önceden oluşturulan veriler: Eyüp Akgül, Caner Ertona, Banu Diri tarafından “Twitter verileri ile duygu analizi” makalesi kapsamında üretilmiş olan kütüphaneden de birçok kelime alınarak Düşünce Kütüphanesine eklemeler yapılmıştır [4].
- iii. Emoji Listesi: Saif Mohammad and Peter Turney, tarafından oluşturulan liste Düşünce Kütüphanesine eklenmiştir [64].
 - a. Emojiler listesinde düzenlemeler yapılmıştır. İlk olarak her bir emoji için Olumlu, Olumsuz veya Nötr olarak gruplama yapılmıştır ve ardından da emojilerde kelimeler gibi değerler verilerek puanlama yapılmıştır. Değerler -5 ile +5 arasındadır.

10 http://www.tdk.gov.tr/index.php?option=com_bts&view=bts, Retrieved June 30, 2016.

Örneğin: “(o: =2(olumlu)”, “): =-4(olumsuz)” olarak gruplanmış ve puanlanmıştır.

- iv. Twitter verileri incelenerek eklenmesinin faydalı olduğu keşfedilen duygu yüklü kelimeler Düşünce Kütüphanesine eklenerek zenginleştirme çalışmaları yapılmıştır.
- v. Türkçe karakter sorunundan kaynaklanan problemleri göz ardı etmemek için yeni metotlar yazılmıştır ve bu algoritmalar sayesinde basit NLP geliştirmesi yapılmıştır.
 - a. Kelimelerde yer alan Türkçe harflerin tamamı İngilizce harflerle değiştirilerek yeni bir kolona KeyText kaydedilmiştir. Her bir kelime için Düşünce Kütüphanesinden kontrol yapılırken öncelikli olarak Keyword tablosunda kontrol yapılır eğer ki Text kolonunda kelime bulunamazsa KeyText kolonundan kontrol yapılır ve kelimelerin eşleşmesi durumunda kelimeye ilgili puan eşleştirilir.

Bu oluşturulan Düşünce Kütüphanesi veri tabanında KeyWord tablo olarak saklanmaktadır. Tabloda yer alan kayıt sayısı 1580’dir. Her bir kelime ‘impression’ kolonunda gruplanır. Gruplamalarda 3 kategori bulunmaktadır. Bunlar: Olumlu, olumsuz ve nötr’dür. Olumlu kelimeler veri tabanında “1”, olumsuz kelimeler “2” olarak, nötr kelimeler ise “0” olarak kaydedilmiştir. Kelimelerde kullanılan kategoriler olumlu ya da olumsuzdur. Basit emojielerde ise olumlu, olumsuz ve nötr kategorisi kullanılmıştır.

Ek olarak her bir kelimeye taşıdığı duygu yoğunluğuna uygun olarak değer verilmiştir.

Bu değerler:

- i. Olumlu kategorisine +1 ile +5 arasındadır.
 - a. Örneğin: mutluluk +5 puan
 - b. Duygu yoğunluğuna göre bazı kelimelere 11 puan verilmiştir. Örneğin: en güzel 11 puandır.
- ii. Olumsuz kategorisinde -1 ile -5 arasındadır.
 - i. Örneğin: sevimsiz -5 puan
 - ii. Küfürler ve argo sözcüklere -11 puan verilerek Olumsuz olarak gruplanmıştır.

- iii. Nötr emoji'lere ise 0 puan verilmiştir.
- iv. Ek olarak kelime grupları da olumlu ve olumsuz olarak gruplanmıştır.
 - iii. Olumsuzlara -6, -10 arası puan verilmiştir. Örneğin: burnu büyük -8 puan
 - iv. Olumlulara 6, 10 arası puan verilmiştir.

Ayrıca bazı özel kelime grupları için farklı puanlama algoritması geliştirilmiştir. Bu kelime gruplarında duygu yoğunluğu daha fazla olduğu için puanlama sistemi de farklı olarak oluşturulmuştur. Olumlu kelime grupları +6 ile +11 arasında değer verilmiştir. Olumsuz kelime gruplarına ise -6 ile -11 arasında değer verilmiştir. Örneğin: “burnu büyük” kelime grubu olumsuz olarak kategorize edilmiştir ve duygu yoğunluk puanı -8 olarak belirlenmiştir. Argo/Kötü söz içeriklerine ise -11 puan verilmiştir.

Bu puanlamalarda ki esas amaç argo kelimeler ve mükemmel olumluluğu ifade eden kelimelerin bir arada kullanılması durumunda çıkacak olan duygu değerlerinin hesaplanması işleminde daha gerçekçi bir sonuç elde edebilmektedir.

Aşağıdaki 3:1 Düşünce Kütüphanesi Data Set tablosunda bazı örnek kelimelerin olumlu (1) olumsuz (2) statüleri, ayrıca bu kelimelerin İngilizce alfabede ki karşılıkları ve kelimelerin duygu değerleri görüntülenmektedir.

Tablo 3:1: Düşünce Kütüphanesi Örnek Set

Key ID	Impression	Text	Key Value	Key Text
359	2	minimum	-3	minimum
35	2	alkolik	-5	alkolik
64	2	az	-3	az
65	2	açgözlü	-5	acgozlu
118	2	cüzi	-2	cuzi
152	2	düztaban	-4	duztaban
216	2	haram	-5	haram
240	2	hödük	-5	hoduk
804	1	enerjik	5	enerjik
805	1	enfes	5	enfes

279	2	kahpe	-5	kahpe
427	2	serseri	-3	serseri
574	2	kokoş	-2	kokos
739	1	bedava	2	bedava
728	1	ağırbaşlı	1	agirbasli
1121	2	burnu büyük	-8	burnu buyuk
1122	2	bölük pörçük	-7	boluk porcuk
1123	2	eli sopalı	-8	eli sopali
1124	2	kuş beyinli	-8	kus beyinli
1125	1	en güzel	11	en guzel

3.1.2 Kullanıcı Ara yüzü (UI Layer) Detayları

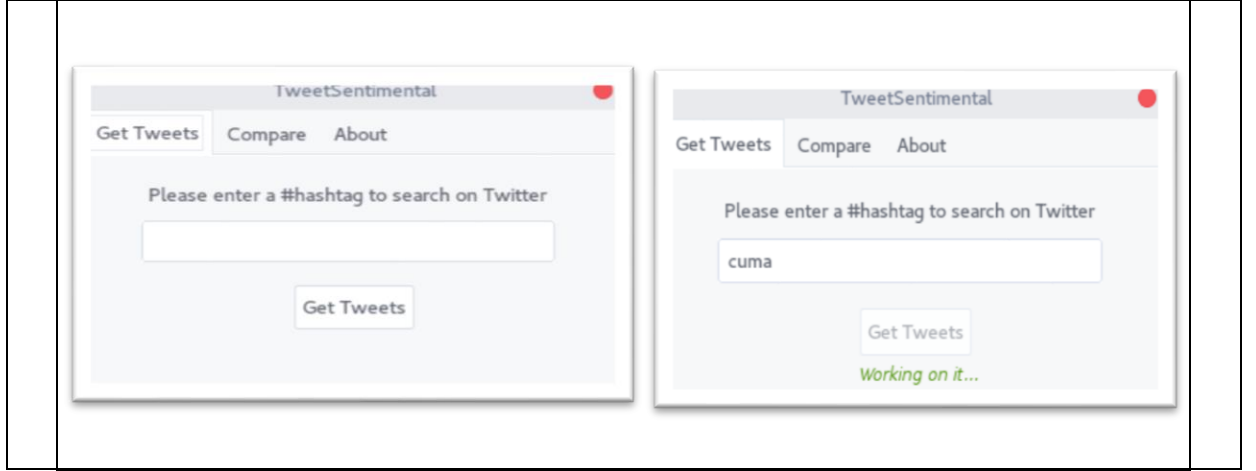
Çok katmanlı yapılarda kullanıcılardan veri almaya ve uygulama ile kullanıcılar arasındaki iletişimin sağlanmasını gerçekleştiren katmandır. Bu katmandan alınan veriler öncelikle Business Logic Layer katmanına iletilir. Bu katmanda işlenir ve ardından da Data Access Layer katmanına iletilir. Bu katmanda da kayıt etme işlemleri gerçekleştirilir. Son olarak tekrar Data Access Layer'dan dönen sonuçlar kullanıcının ekranına yansıtılır.

Bu tez kapsamında ki gereksinimleri karşılayabilmek için basit bir ara yüz tasarlanmıştır. Ara yüz ekran görüntüleri Şekil 3.1:2 ve Şekil 3.1:3 ve Şekil 3.1:4 de görüntülenmektedir.

Ekranda yer alan “Get Tweets” sekmesinde bulunan Etiket(hashtag) textbox (metin kutusu)’na sorgulanmak istenen kelime yazılarak Get Tweets butonuna tıklanır. Böylelikle ekranda ki bu GetTweets butonu inaktif(desible) olur böylelikle tekrardan tıklanamaz ve hemen alt kısmında tweetlerin toplandığı bilgisi için “working on it..” şeklinde bir ifade görüntülenir. Aşağıda yer alan örnek Şekil 3.1:2 Cuma hashtagli tweetler toplanırken görüntülenen metin gösterilmiştir.

Twitter üzerinden ilgili etiket bilgisine sahip veriler döner ve bu veriler veri tabanında ki ilgili tablolarına ayrıştırılarak ve düzenlenerek kaydedilir.

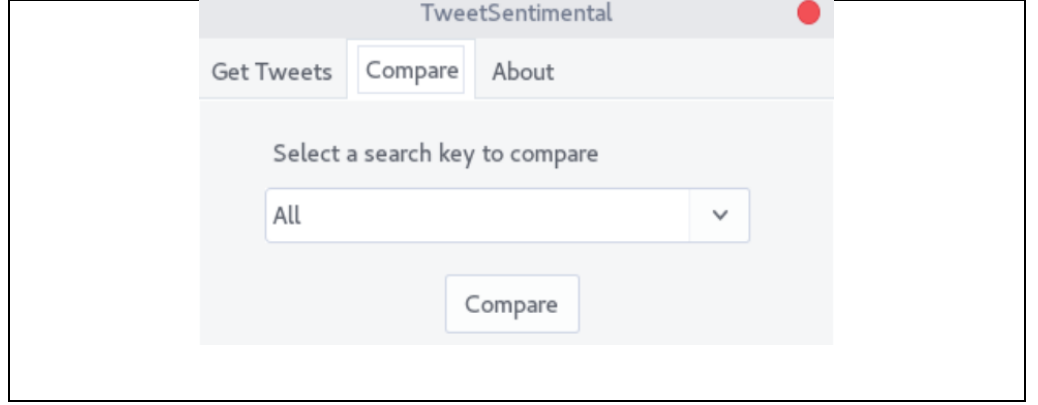
Şekil 3.1:2 Arayüz Ekran Görüntüsü



Takibe başlama etiketleri birden fazla belirlenebilir. Ayraç olarak virgül işareti (,) kullanılarak birden fazla etiketle ilgili tweet verileri çekilebilmektedir.

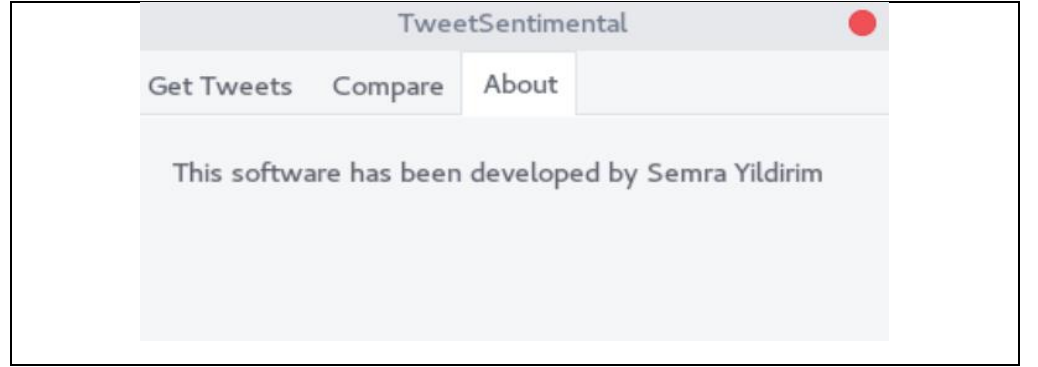
Ayrıca ekranda yer alan “Compare” sekmesinde bulunan açılır listede(combobox) daha önceden toplanan etiketler (searchKey kolonu) liste olarak görüntülenmektedir. Bu listeden seçim yaparak Compare butonuna tıklanması durumunda da daha önceden veri tabanına kaydedilmiş olan tüm tweetler tek tek ayrıştırılır ve bu tweetler içinde yer alan kelimeler tek tek parçalara bölünür. Bu işlemin ardından ayrıştırılan tüm kelimeler Düşünce Kütüphanesinde ki kelimelerle teker teker kıyaslanır. Her bir tweet bazında duygu değeri ve duygu yoğunluğu değeri hesaplanır. Bu değerler Result tablosuna eklenir ya da daha önceden Duygu Değeri Hesaplanan bir tweet için işlem yapılıyorsa Result tablosunda ki değeri güncellenir. Duygu Kütüphanesine yeni ekleme yapılması durumunda daha önceden saptanmış olan duygu yoğunluğu değerlerinde değişiklikler gerçekleşebilir. Örneğin: Eklenen kelime daha önceden duygu yoğunluk değeri hesaplanan bir tweet içinde yakalanırsa bu durumda ilgili tweet için tekrardan hesaplamalar yapılabilir ve veri tabanında ki kaydı güncellenebilir. Birden fazla etiket ile aynı anda Duygu yoğunluk değerleri hesaplanamaz. Tweetlerin toplanmasında birden fazla etiket ile işlem yapılabilirken, tweetlerin analizinde sadece bir etiket üzerinden işlem (duygu değeri hesaplaması) yapılabilir. Bunun sebebi ise analiz hesaplamalarının hızlı sonuç vermesini sağlamaktır. Analiz işlemleri tamamlandıktan sonra UI dan bittiğini göstermek için “Done” şeklinde mesaj gösterilir.

Şekil 3.1:3 Ara yüz Ekran Görüntüsü(Compare)



“About” sekmesinde ise sadece bilgi bulunmaktadır.

Şekil 3.1:4 Ara yüz Ekran Görüntüsü(About)



3.1.3 İş Mantığı Katmanı (Business Logic Layer) Detayları

İş mantığı katmanı, çok katmanlı tasarlanan uygulamalarda mantıksal işlemlerin yapıldığı katmandır. Kullanıcıdan, Kullanıcı Ara yüzü Katmanı(UIL) vasıtasıyla aldığı verileri işleyip Veri Erişim Katmanına (DAL) iletir. Bu iletişim iki yönlü de olabilmektedir. DAL’da yer alan verileri işleyerek UIL yardımıyla kullanıcı karşısına çıkartır.

Bu tez kapsamında, iş mantığı katmanı çeşitli metod ve algoritmalar barındırmaktadır. İlk olarak Twitter’a bağlanma aksiyonunu gerçekleştirmek adına “<https://apps.twitter.com/>” adresinden elde edilen “Access Token, Access Token Secret, Consumer Key,

Consumer Secret” verileri elde edilerek API (Application Programming Interface) erişimine olanak sağlanmıştır. Ardından bu bilgiler kullanılarak Twitter’a giriş yapılmıştır.

Geliştirilen birtakım algoritmalar ve çalışma detaylarından aşağıda bahsedilmiştir:

Twitter üzerinden tweet çekmek adına getTwits metodu tasarlanıp programlanmıştır. Bu metot içerisinde verilen bir kelime (etiket) kullanılarak, Twitter üzerinden tweetleri çekebilmek adına “LinqToTwitter” kütüphanesi kullanılmıştır. UI üzerinde yer alan textbox yardımıyla “SearchKey” parametresi olarak gelen kelimeyi Twitter üzerinde arar. Twitter üzerinden dönen veriyi bir liste olarak hafızada (memory) da saklar.

Hafızada bir liste olarak tutulan tweet listesi, saveTwits metodu içerisinde ayıklanıp, “SearchKey” bazında kategorize edilerek veri tabanında ilgili tablolarına normalizasyon işlemi yapıldıktan sonra kaydedilir. Ayrıca coğrafi konumu (Geocode) ve dili (lang) bazında tweetleri tamamı kategorize edilerek düzenlenir.

Ardından processCompareOperationAndSave metodu yardımıyla tweetler, bu tweetlerin hangi kelime (etiket) ile elde ettiğimiz bilgisi ve etiket bilgisiyle ayrıştırılarak kaydedilir. Türkçe tweetlerde kullanılan klavyenin mobil cihaza ait ve bu yüzden İngilizce olması nedeniyle yaşanan problemleri en aza indirmek adına elde edilen tweet içerisindeki karakterler küçük harflere çevrilir. (ToLower) Sonrasında ise Türkçe ’ye özgü olan (ı, ç, ö vb.) harfleri, İngilizce klavyede ki karşılıklarına (i, c, o vb.) uygun olarak güncellenir. Tweet içinde yer alan her bir kelime, Keyword tablosunda yer alan kelimeler ile tek tek kontrol edilir. Kelimeler eşleştiğinde keyValue değerleri toplanarak hesaplama yapılır. Hesaplama işlemi bir kelimenin, bir tweet içerisinde birden fazla geçmesi durumunda sadece ilk kelimeyi alacak şekilde kurgulanmıştır. Ayrıca tweet içerisinde birden fazla duygu içeren kelimelerin, Keyword tablosundaki kelimeler ile eşleşmesi halinde keyValue değerleri toplanılarak Result tablosundaki resultValue değeri oluşturulur. Örneğin: “Acı bizim için umuttur.” Şeklindeki bir tweet için yapılan hesaplamada “acı” - 5 puan, “umut” 4 puan ise resultValue değeri $-5 + 4 = -1$ olarak hesaplanır. Tweet’in duygusu (impression) ise 2 (olumsuz) olarak belirlenecektir. Bu aksiyonu

gerçekleştirmek için saveResult metodu oluşturulmuştur. ResultValue değeri sıfırdan büyük ise duygu (impression) değeri olumlu, sıfırdan küçük ise duygu değeri olumsuz ve sıfıra eşit ise duygu değeri nötr olarak kaydedilir.

3.1.4 Veri Erişim Katmanı (Data Access Layer) Detayları

Twitter üzerinden belirli etiketlerle çekilen datalar direk olarak veri tabanına kaydedilmektedir. Bu kaydedilme aşamasında her bilgi kendiyile ilgili olan tablolara işlenmektedir. Ayrıca Veri kirliliğini ve düzensizliğini en aza indirgeyebilmek için dağınık katmanlı ve ilişkisel veri tabanı olan MsSQL ve SQLite tercih edilmiştir.

Bu tez çalışması kapsamında geliştirilen yazılımda veri tabanı yönetim sistemi olarak Microsoft SQL Server (MsSQL) ve SQLite kullanılmıştır. Dağınık katmanlı olarak tasarlanan proje için ilişkisel (relational) veri tabanı tasarımı yapılmıştır. Her iki kullanılan veri tabanı mimarisi tamamen aynıdır. Kısıtlı veri olarak veri kümeleriyle ilgili çalışmalar için MsSQL kullanılmıştır fakat büyük veri kümelerinde yapılan işlemler için ise SQLite tercih edilmiştir. İlk olarak Python dili yardımıyla toplanan tweetler performans bakımından hızlı olması nedeniyle öncelikle csv (comma-separated values) formatında toplanmıştır. Daha sonrasında bu verilerde tekil olma(unique) kontrolleri yapılmıştır. Eğer mevcut durumda birden fazla aynı ID'ye sahip tweet var ise bu kayıtlar silinmiştir. Ardından da bu cvs dosyaları MS SQL'e aktarılma(export) istenmiştir fakat alınan hatalar ve satır sayısı sorunları nedeniyle bu işlemde vazgeçilmiştir. Bu işlemlerde esas amaç veri kirliliğinden kurtulmak, veri düzenini ve tekilliğini sağlamaktır. Ardından yapılan araştırmalarda ise bu işlemle çok vakit kaybettiğimiz için Python yardımıyla toplanan tweetler direkt olarak SQLite veri tabanına atılmaya başlanmıştır. Böylelikle kısıtlı verilerin tamamı MS SQL üzerinde analiz çalışmalarında kullanılmıştır. MS SQL veritabanında saklanan veri setlerine “kısıtlı veri kümesi” ve “deneysel kısıtlı veri kümesi” olarak bahsedilmiştir. Diğer veriler ise SQLite üzerinden analiz çalışmaları için kullanılmıştır. Bu verileri kümelerinde veri sayısının boyutu dolayısıyla “büyük veri kümesi” ve “deneysel kapsamlı veri kümesi” olarak adlandırılmıştır.

“SaveTwits” metodu içinde veri tabanıyla bağlantıyı sağlamak amacıyla connection string yazılmıştır.

Yapılan aksiyonlar:

- i. Connection string de ki deęerler girilerek connection aılır.
- ii. Database oluřturulur. Database adı=Twitter 'dır.
 - a. Main tablosu
 - i. Tweet'in hangi arama kelimesiyle(searchKey) ekildięini ve hangi tarihte ekildięi gibi bilgiler bu tabloda kaydedilmektedir.
 - ii. Kolonlar: Twit_ID, Creation_Time, Geo_Code, SearchKey
 - b. Tweets tablosu
 - i. Toplanan tweetlerin metinlerinin ve dil bilgisinin saklandığı tablodur.
 - ii. Kolonlar: Twit_ID, Twit_Text, Lang, RT_Count, Like_Count, Reply_Count
 - c. UserData tablosu
 - i. Hangi kullanıcının hangi tweet'i paylařtığı ve bu kullanıcıların takipi sayısı gibi detaylarının saklandığı tablodur.
 - ii. Kolonlar: Twit_ID, User_Name, Follower_Count, Following_Count
 - d. Results tablosu
 - i. Tweetlerin olumlu, olumsuz ve nötr olarak deęerlerinin görünlendięi ve tweet ierisinde bulunan duygu kütüphanesiyle eřleřen kelimelerin saklandığı tablodur. User interface de yer alan Compare sekmesinde ki Compare butonuna tıklandığında her bir tweet tek tek iřleme alınır ve bu iřlemler sonucunda tweet bazında Result tablosuna duygu yoğunlukları ve deęeri eklenerek kayıt edilir.
 - ii. Kolonlar: Twit_ID, Impression, ResultValue, ContainingKeys
 - e. Keywords tablosu düşünce kütüphanesi olarak oluřturulmuřtur.
 - i. Türke sentimental analiz kütüphanesi olarak kullanılan tablodur.
 - ii. Kolonlar: Key_ID, Impression, Text, KeyValue, KeyText
- iii. İřlemler tamamlandıktan sonra connection kapatılır.

Twitter tarafından yaratılan Tweet'in ID'si aynı řekilde kullanılmıřtır ve ilgili tablolarda Birincil anahtar olarak kullanılmıřtır. Twitter veri tabanındaki tablolar, alanların aıklamaları ve alanların tipleri Bkz. EK 1: Database Detayları tablosundan detaylandırılmıřtır.

4. ANALİZ VE BULGULAR

Bu tez kapsamında yapılan analiz çalışmalarında birçok Twitter verisi incelenmiştir. Bu incelemeler de genelde çok kullanıcısı olan profiller ve Trend Topic listesi kullanılmıştır. Bu incelemeler sonrasında bazı etiketler(hashtagler) belirlenmiştir (#havadurumu, #yaz, #winter, #monday, # pazar gibi). Böylelikle kararlaştırılan etiketler küme küme Twitter dan lokal veri tabanlarına aktarılmıştır. Verileri çekme, ayıklama ve düzenleme işlemleri için C#.NET ve Python yazılım dilleri kullanılmıştır. Verilerin tamamı Sql'e aktarılmış ve veri tabanında saklanmıştır. Gözle tweetler üzerinde gözlemler yapılmış ve bu incelemeler sayesinde Duygu Kütüphanesine eklemeler gerçekleştirilmiştir. Her yeni eklenen kelime için olumlu-olumsuz değerleri ve duygu yoğunlukları düzenlenmiştir. Duygu Kütüphanesine ayrıca basit emojilerde eklenmiştir. Ardından da her bir tweet içinde yer alan kelimeler ayrıştırılarak duygu değeri saptanmıştır. Yaratılan bu duygu değerinin olabildiğince gerçekçi çalıştığını tek bir duygu kütüphanesi kullanılarak simüle edebilmek için birbirinden tamamen farklı 3 farklı başlık altında analiz çalışmaları yapılmıştır. Ek olarak da veri sayısına göre analiz sonuçlarında farklılık olup olmadığına dair incelemeler yapılmıştır. Analiz çalışmalarında 5 ayrı veri kümesi kullanılmış olup bu veri kümeleri hiçbir şekilde birleştirilmemiştir. Hava durumu için 2 ayrı veri kümesi (hava durumu ve mevsimler), haftanın günleri için 2 ayrı veri kümesi (2500lük küme ve 14binlik küme) ve survivor için 1 veri kümesi kullanılmıştır. Tez kapsamında toplanan tweet sayısı ise 42.468'dir. 2016 Eylül-2017 Kasım aylarınca tweet toplama işlemleri yapılmıştır.

A. Hava Durumu Analiz Çalışması

Hava durumunun mevsimsel bazla Twitter kullanıcılarına olan etkilerini ölçmek için analiz çalışmaları yapılmıştır. Bu çalışmalar yaklaşık 1 yıllık zaman dilimi içinde toplanan veriler üzerinde yapılmıştır. Ayrıca az veri ve çok veri arasında farklılıkları gözlemlemek amacıyla da çalışmalar iki ayrık veri kümesi üzerinden gerçekleştirilmiştir.

- i. Kısıtlı veri kümesiyle yapılan analiz çalışmalarında kullanılan etiketler: “hava, hava durumu, kış, bahar, yazgeldi” dir. Hava değişikliklerinin Türkçe Tweet paylaşan Twitter kullanıcıları üzerindeki etkisini gözlemleyebilmek için iki farklı deney yapılmıştır.

a.Kış aylarında (2017 Ocak-Şubat-Mart-Nisan) çekilen veri kümesi üzerinden gerçekleştirilmiştir. Bu deneyde amaçlanan Türkçe tweet atan kullanıcıların kış aylarında daha “olumsuz” ya da” karamsar” tepkiler verip vermedikleri görmektir. (Bkz. Şekil 4.1:1 Kış Aylarında Atılan Tweetlerin Duygu Yoğunluk Dağılımı)

b.Bahar ve yaz aylarında (2017 Mayıs-Haziran-Temmuz) çekilen veri kümesi üzerinden yapılmıştır. Bu deneyde amaçlanan Türkçe tweet atan kullanıcıların bahar aylarında “olumlu” ya da “pozitif” olup olmadıklarını görmektir. (Bkz. Şekil 4.1:2 Bahar Aylarında Atılan Tweetlerin Duygu Yoğunluk Dağılımı)

Bu iki deney sonrasında ise kullanıcıların genel olarak Kış aylarında daha karamsar oldukları gözlemlenmiştir. Fakat yaz aylarının gelmesiyle ve yazın başlamasıyla kullanıcıların tweetlerinin daha “olumlu”, daha “pozitif” olduğu gözlenmiştir.

ii. Büyük veri kümesiyle yapılan analiz çalışmalarında kullanılan etiketler: “ilkbahar, yaz, sonbahar, kış, spring, summer, fall, winter” dir. Hava değişiklerinin Türkçe Tweet paylaşan Twitter kullanıcıları üzerindeki etkisini gözlemleyebilmek için 4 mevsim için 4 deney yapılmıştır.

a. “Yaz” ve “summer” etiketleriyle toplanan verilerde (2017 Ağustos – Eylül - Ekim) yapılan çalışmalarda Türkçe tweet atan kullanıcıların çok fazla resim paylaşması nedeniyle genel olarak “nötr” oldukları görülmektedir. Fakat “olumsuz” olma oranı “olumlu” olma oranına göre düşüktür.

b.“Sonbahar” ve “fall” etiketleriyle toplanan verilerde (2017 Ağustos – Eylül - Ekim) yapılan çalışmalarda Türkçe tweet atan kullanıcıların “olumlu” ya da “pozitif” oldukları görülmektedir.

c. “İlkbahar” ve “spring” etiketleriyle toplanan verilerde (2017 Temmuz-Ağustos-Eylül) yapılan çalışmalarda Türkçe tweet atan kullanıcıların ağırlıklı olarak “olumlu” ve “nötr” oldukları görülmektedir.

d. “Kış” ve “winter” etiketleriyle toplanan verilerde (2017 Eylül-Ekim-Kasım) yapılan çalışmalarda Türkçe tweet atan kullanıcıların “nötr” ve “negatif” olma oranlarının yüksek olduğu görülmektedir.

Bu dört deney sonrasında ise kullanıcıların genel olarak Kış mevsiminde paylaştıkları tweetlerin daha karamsar oldukları gözlemlenmiştir. Fakat yaz mevsiminde kullanıcıların tweetlerinin daha olumlu, daha mutlu olduğu gözlemlenmiştir. Ayrıca sonbahar aylarında ise mevsim geçişinin kullanıcıları olumlu olarak etkiledikleri gözlemlenmiştir. Ek olarak ilkbahar mevsimine dair veri sayısının azdır bu tweetler incelendiğinde görülen sonuç ise kullanıcıların yaza özleminden bahsedilmektedir. Kullanıcıların bu özlemlerini belirtirken “olumsuz” kelimeler seçmeleri nedeniyle elimizde var olan verilerden “olumlu” ve “nötr” ağırlıklı bir duygu analizi sonucu çıkmıştır.

B. Haftanın Günleri Analiz Çalışması

Haftanın farklı günlerinde kişilerin paylaştıkları tweetlerin olumlu-olumsuz-nötr olma durumları ölçülmeye çalışılmıştır. Bu çalışmalar esnasında gün temelli kırılımlar oluşturulmuştur. Ayrıca 2 farklı veri kümesi üzerinden analiz çalışmaları yapılmıştır. Bu çalışmanın veri setinin oluşturulması esnasında “Pazartesi”, “Salı”, “Çarşamba”, “Perşembe”, “Cuma”, “Cumartesi” ve “Pazar” etiketleri kullanılmıştır. Ayrıca yapılan incelemelerde Twitter da Türkçe paylaşım yapan kullanıcıların İngilizce hashtaglerle paylaşım yapıldığının gözlemlenmesinin ardından “Monday”, “Tuesday”, “Wednesday”, “Thursday”, “Friday”, “Saturday”, “Sunday” etiketliyle de veriler toplanmıştır.

- i. Deneysel kısıtlı veri kümesi üzerinden yapılan analizle sonrasında kişilerin Pazartesi günleri “olumsuz” paylaşımlar yaptıkları görülmüştür. Cuma, Cumartesi ve Pazar günleri ise daha “olumlu”

paylaşımlar yaptıkları görüntülenmiştir. Haftanın tamamlanmasıyla beraber kişilerin daha “olumlu” paylaşımlar yapmaya başladıkları görülmüştür (Bkz. Şekil 4.2:1 Günlere Göre Duygu Dağılımı Grafiği).

- ii. Deneysel kapsamlı veri kümesi üzerinden yapılan analizler sonrasında kişilerin en çok tweet paylaştıkları günün Pazartesi olduğu görülmüştür. Ek olarak Cuma günleri ise en “olumlu” paylaşımların yapıldığı görülmüştür. Fakat tweetlerin içerikleri incelendiği zaman depresyon gibi ağır olumsuzluk ifade eden duygunun da en çok Pazartesi günlerinde paylaşıldığı görülmüştür. Ayrıca Pazar günleri de “olumlu” paylaşım oranının yüksek olduğu görülmüştür (Bkz. Şekil 4.2:2 Günlere Göre Duygu Dağılımı Grafiği).

C. Survivor Analiz Çalışması

Ek olarak yapılan Türkçe Tweet paylaşan kullanıcıların genel olarak olumsuz yorum yapma alışkanlıklarını ve bozuk Türkçe ile tweet attıklarını ispat edebilmek için Trend Topic listesinde günlerce üst sırada yer alan “survivor” etiketiyle atılan tweetler incelenmiştir. Burada beklenen sonuç nötr veya olumsuz gruplamaların olumlulardan daha fazla orana sahip olduğunun görüntülenmesiydi ve analizler tamamlandıktan sonra karşılaşılan tabloda da survivor etiketiyle paylaşılan tweetlerin genel olarak olumsuz olduğu gözlemlenmiştir (Bkz. Şekil 4.3:1 “survivor” Etiketine Göre Duygu Yoğunluk Dağılımı).

4.1 HAVA DEĞİŞİKLİĞİ VE MEVSİMLERİN KULLANICILAR ÜZERİNDEKİ ETKİSİNİN ANALİZİ

Hava durumunun ve mevsimsel değişimin Twitter da Türkçe tweet paylaşan kişilerin duygularında ki değişiklikleri ölçmek amacıyla analiz çalışmaları yapılmıştır. Bu çalışmalar için 2 ayrı veri kümesi kullanılmıştır. Bu iki ayrı veri kümesi için altında farklı deneyler yapılmıştır. Twitter ‘da yer alan kullanıcıların bozuk Türkçe kullanma alışkanlıkları nedeniyle, elde edilen verilerin bir kısmı işlenemmiştir. Toplamda ise 10.968 tweet ile işlem yapılmıştır.

- i. Kısıtlı veri kümesi: 2180 adet veriden oluşturulan birinci veri kümesi üzerinden yer alan Twitter kullanıcılarının duygularındaki etkileşimleri ölçmek adına yapılan analizler sonucunda kış ve bahar aylarında toplanan tweetler incelenmiştir. Bu verileri çekmek için “hava”, “havadurumu”, “bahar” ve “yazgelsin” etiketleri kullanılmıştır. Bu işlemler sonucunda kış aylarında ortaya çıkan veri kümesinde yer alan tweetlerin yüzde 29’unun Nötr, yüzde 34’ünün Olumlu ve geriye kalan yüzde 37’sinin ise Olumsuz olduğu tespit edilmiştir. Bahar aylarında ortaya çıkan veri kümesinde yer alan tweetlerin ise yüzde 60’ının Nötr, yüzde 15’inin Olumsuz ve geriye kalan yüzde 25’inin ise Olumlu olduğu tespit edilmiştir.
- ii. Büyük veri kümesi: 8788 adet veriden oluşturulan ikinci veri kümesinde yer alan Twitter kullanıcılarının duygularında ki mevsimlerle ilgili etkileşimi ölçmek adına analiz çalışmaları ilkbahar, yaz, sonbahar ve kış mevsimlerinde yapılmıştır. Bu deneyler sonucunda kış mevsiminde ve yaz mevsiminde yaklaşık olarak aynı sayılarda tweet atıldığı gözlenmiştir. Fakat verilerin toplanma tarihinin sonbahar aylarının tamamını kapsaması nedeniyle ve insanların yaz aylarından bunalıp bahara olan hasretlerini belirtmesi nedeniyle #sonbahar şeklinde çok fazla veri elde edilmiştir. Örneğin: Bu büyük veri kümesinde ki paylaşımlardan kış mevsiminde paylaşılan tweetlerin yüzde 27’si olumlu, yüzde 23’ü olumsuz ve yüzde 50’si nötrdür. Sonbahar mevsiminde paylaşılan tweetlerin ise yüzde 74’ü olumlu, yüzde 9’u olumsuz ve yüzde 17’si nötrdür. Yaz mevsiminde paylaşılan tweetlerin ise yüzde 38’i olumlu, yüzde 15’i olumsuz ve yüzde 47’si nötrdür. İlkbahar mevsiminde paylaşılan tweetlerin ise yüzde 36’sı olumlu, yüzde 24 olumsuz ve yüzde 40’ı nötrdür.

4.1.1 Kısıtlı Veri Kümesi Üzerinden Hava Durumuyla İlgili Analiz Çalışmaları

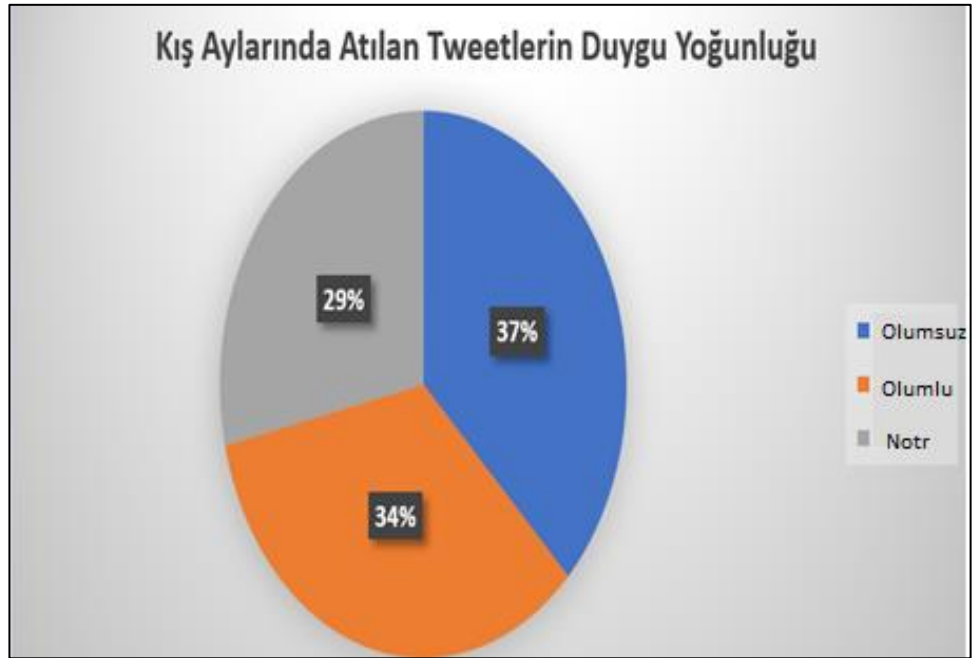
Bu kapsamda 2 farklı deney yapılmıştır. Bu deneyler kapsamında 2180 adet veriden oluşan kısıtlı veri kümesi olarak adlandırılan veri kümesi kullanılmıştır. Bu deneylerde amaçlanan hava durumunun Twitter üzerinden Türkçe tweet paylaşan kullanıcıların duygularını analiz etmektir. İlk olarak kullanıcıların soğuk havalarda olan duygu yoğunluklarını ölçmek için kış aylarında Türkçe tweet paylaşan kullanıcıların tweetleri

üzerinden analiz çalışmaları yapılmıştır. Ardından da kullanıcıların baharın gelmesiyle duygularında ki değişimi keşfedebilmek farklılıkları gözlemleyebilmek amacıyla yaz aylarında ve bahar aylarında paylaştıkları tweetler incelenmiştir. Analiz sonrasında karşılaşılan sonuçların beklenen değerlerle uyumlu olduğu gözlemlenmiştir.

Soğuk Havanın Kullanıcılar Üzerindeki Etkisinin Analizi

4.1.1.1
Yapılan Sentimental analizi çalışmaları sonrasında Türkçe hava durumuyla ilgili tweet atan insanların yüzde 34'ünün soğuk havadan olumsuz etkilendiği, yüzde 37'lik kesimin ise olumlu etkilendiği ölçümlenmiştir. Geriye kalan yüzde 29'luk dilim ise nötr olarak gruplandırılmıştır. Bu amaçla toplanan “hava” etiketli verilerin bulunduğu örneklerin yer aldığı tablo 4:1 aşağıda bulunmaktadır. Aşağıdaki Şekil 4.1:1 Kış Aylarında Atılan Tweetlerin Duygu Yoğunluk grafiği görüntülenmektedir.

Şekil 0:1 Kış Aylarında Atılan Tweetlerin Duygu Yoğunluk Dağılımı



Örnek1: myyamann kullanıcısı tarafından paylaşılan “hava” etiketiyle toplanan “Hava puslu, soguk Kirlar koyu kirmizi Saman sarisi ölü yesil Kis gelmek üzere oysa ki gönül kısa girmeye hazir degil” tweetin de yer alan duygu içeren birden fazla olumsuz kelime yer almaktadır. Bu kelimeler “kisa, soguk, degil” ‘dir. Burada yer alan kelimelerin duygu

değerleri sırasıyla “-4, -5, -5”. Tweet olumsuz olarak gruplandırılmıştır ve Toplam duygu değeri -14 olarak belirlenmiştir.

Ayrıca görüldüğü gibi kullanıcı bazı karakter Türkçe alfabe ile bazı kelimeleri ise İngilizce alfabeye yazmıştır. Fakat bu durum kelimelerin yakalanmasına engel olmamıştır.

Örnek2: arabel_la kullanıcı tarafından paylaşılan “hava” etiketiyle toplanan “Hava o kadar mükemmel ki birkaç gündür statik elektrikte yüklüyüm ve kedim benden nefret ediyor. (Bk.: kedi burnundan elektrik çarpması)” tweetinde bir olumlu bir olumsuz kelime bulunmaktadır. Bu kelimeler” mükemmel, nefret ediyor” ‘dur. Burada yer alan kelimelerin duygu değerleri sırasıyla “+5, -10”. Twit olumsuz olarak gruplandırılmıştır ve Toplam duygu değeri -5 olarak belirlenmiştir. Bu örnekte karşılaşılan kelime grubu (nefret ediyor) ve kelimeler duygu yoğunlukları fazla olan kelimelerdir. Bu nedenle tüm metnin değeri daha sıfıra yakın çıkmıştır.

Örnek3: vaasnihilus kullanıcısı tarafından paylaşılan “hava” etiketiyle toplanan “İnanılmaz güzel bir gün pırıl bir hava var bu var olus sancisini iliklerime kadar hissedebilirim demektir.” Tweetinde olumlu kelimeler bulunmaktadır. Burada yer alan duygu yüklü kelimeler “güzel, pırıl” ‘dır. Burada yer alan kelimelerin duygu değerleri sırasıyla “5, 5”. Tweet olumlu olarak gruplandırılmıştır ve Toplam duygu değeri 10 olarak belirlenmiştir.

Tablo 4:1: "hava" Kelimesine Göre Duygu Yoğunluğu Verileri

User_Name	Twit_Text	Impression	ResultValue	ContainingKeys	SearchKey	RT_Count
BuMuhamet	Kardesim kis mevsiminde oldugumuz için hava soguk anormal bisey yok	2	-15	anormal, soguk	hava	0
myyamann	Hava puslu, soguk Kirlar koyu kirmizi Saman sarisi ölü yesil Kis gelmek üzere	2	-14	kisa, soguk, degil	hava	0

	Oysa ki gönül kısa girmeye hazır değil. Nazım Hikmet					
skor_net1	RT @kabateyze: Hava buzz? bu saatte yatanlar sizden nefret ediyorum kalkın lan biz ise başladık ne uykusu bu ?? Günaydın i buraya bıraktım...	2	-11	lan, ?, günaydın, hava nefret ediyor	hava	81
skor_net1	RT @aytkncvk: Türküler siirler yasta, ben yorgunum eskisi gibi ve hava her zaman olduğu gibi yine soğuk...	2	-15	eski, soğuk, yorgun	hava	69
arabel_la	Hava o kadar mükemmel ki birkaç gündür statik elektrikle yüklüyüm ve kedim benden nefret ediyor. (Bkz: kedi burnundan elektrik çarpması)	2	-5	mükemmel, nefret ediyor	hava	0
Nunu9695	Hava ne biçim yaa açık güneşli ama kar yağıyor	2	-3	ne biçim, güneşli	hava	0
komikdeyill	Hava biraz daha ısınmazsa okulu bırakmak zorunda kalıcam	2	-5	zor	hava	0
vaasnihilus	İnanılmaz güzel bir gün pırıl pırıl bir hava var bu var oluş sancisini iliklerime kadar	1	10	güzel, pırıl	hava	0

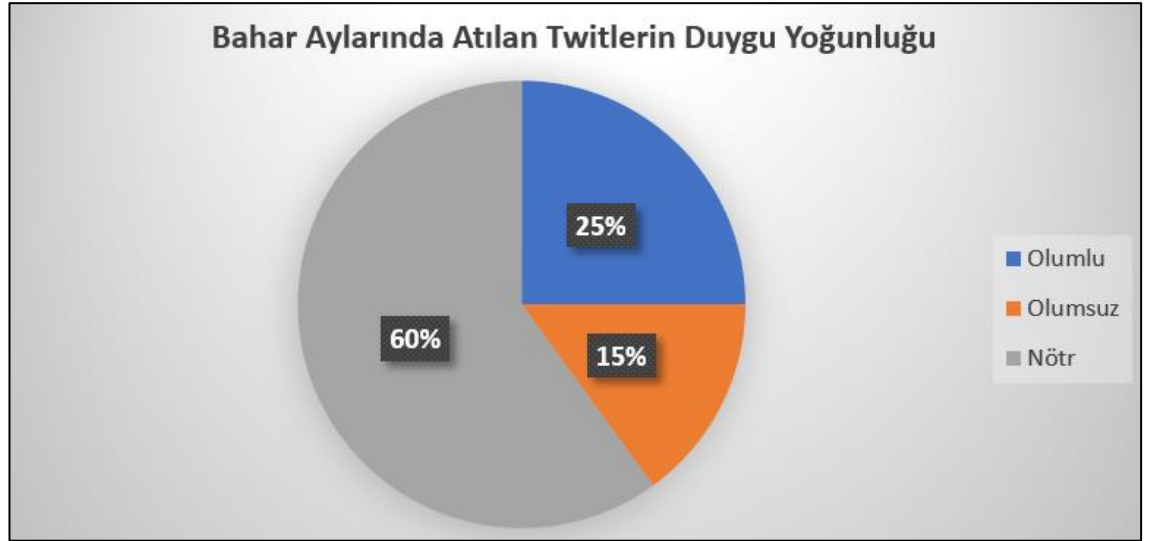
	hissedebilirim demektir.					
lavieestbella4	Ne kadar güzelsin öyle #istanbul piril sicacik günes, mis gibi hava... ????????	1	18	?, guzel, piril, mis gibi	hava	0

Sıcak Havanın Kullanıcılar Üzerindeki Etkisinin Analizi

4.1.1.2

Sentimental analiz çalışmaları sonrasında Türkçe hava durumuyla ilgili tweet atan insanların yüzde 15’inin sıcak havadan olumsuz etkilendiği, yüzde 25’lik kesmin olumlu etkilendiği ölçümlenmiştir. Geriye kalan yüzde 60’lık kesim ise nötr olarak gruplandırılmıştır. Ayrıca tweetlerin detayları ve metinleri Tablo 4:2 de veri tabanında yer aldığı şekilde bulunmaktadır. Aşağıdaki Şekil 4.1:2 Bahar Aylarında Atılan Tweetlerin Duygu Yoğunluk grafiğinde bu oranlar görüntülenmektedir

Şekil 0:2 Bahar Aylarında Atılan Tweetlerin Duygu Yoğunluk Dağılımı



Örnek1: Helin9977 kullanıcısı Retweet edilen “bahar” etiketiyle toplanan “RT @sunset_71: Dertleriniz kum tanesi kadar küçük, Sevinç'leriniz bahar yagmurlari kadar bol olsun, dualariniz kabul Cumaniz Hayirli olsun??G??...” tweetinde olumlu ve olumsuz duygu değeri içeren kelimeler bulunmaktadır. Bu kelimeler “dert, bol, hayirli” ‘dır.

Burada yer alan kelimelerin duygu deęerleri sırasıyla “-5, 5, 4”. Tweet olumlu olarak işaretlenmiştir. Toplam duygu deęeri ise 4 olarak belirlenmiştir.

Bu örnekte yer alan kullanıcının baharın gelmesiyle “pozitif” olarak etkilendięi görülmüştür. “Dert” kelimesi her ne kadar olumsuz olsa da “bol” ve “hayirli” kelimelerinin duygu deęer toplamaları daha büyük olması nedeniyle tweetin deęeri olumlu olarak hesaplanmıştır ve sonuç beklentiyi karşılamaktadır.

Örnek2: hemfikirli kullanıcısı tarafından paylaşılan “yazbaslasin” etiketiyle toplanan “#yazbaslasin etiketi yapin aninda algida kazan demistiniz hani nerde :)” tweetinde bir olumlu kelime ve birde olumlu olarak duygu deęerine sahip emoji bulunmaktadır. Bu kelimeler” kazan, :)” ‘dir. Burada yer alan kelimelerin duygu deęerleri sırasıyla “4, 2”. Tweet olumlu olarak gruplandırılmıştır ve toplam duygu deęeri 6 olarak belirlenmiştir. Bu örnek de yer alan metin olumlu anlam ifade etmektedir. Analiz sonucunda Duygu kütüphanesinde eşleştirilen kelimeler ve emojiler olmuştur. Bu metinde herhangi bir olumsuz kelime yakalanmamıştır. Bu nedenle de toplam duygu deęeri direkt “olumlu” olarak hesaplanmıştır ve sonuç beklentiyi karşılamaktadır.

Tablo 4:2: "bahar ve yazbaslasin" Kelimelerine Göre Duygu Yoęunluęu Verileri

User_Name	Twit_Text	Impression	ResultValue	ContainingKeys	SearchKey	RT_Count
hemfikirli	#yazbaslasin etiketi yapin aninda algida kazan demistiniz hani nerde :)	1	6	kazan, :)	yazbaslasin	0
demabaran	Bu kaçinci bahar kayip giden ellerimizden Kuslarin civiltisine, çiçeklerin gülümseyisine doyamadan, akasyali sokaklarda yürümeden kayip gdyr	2	-3	kayip, gulumse	bahar	0
mrskaysersoze	Bahar drake'e ilani ask etmis onu da instada gördüm	1	5	salak, ask	bahar	0

	direge çarpmismis yolda ne ariyon burda çocuklarimin babasi drake demis salakkk					
Helin9977	RT @sunset_71: Dertleriniz kum tanesi kadar küçük, Sevinç'leriniz bahar yagmurlari kadar bol olsun,dualariniz kabul Cumaniz Hayirli olsun??G??...	1	4	dert, bol, hayirli	bahar	47
demabaran	O kistan sonra açan ilk çiçek ve umut ve gözyasi ve insirah nasil unutulsun! Kis ne denli kara ise bahar o denli ebyaz olurmus..	2	7	gozyasi	bahar	0
7373Ako	RT @Acerakis: Izlerken insanin yüzüne bahar geliyor.. Ilaveten; Çok özledik be Selo baskan :(1	1	:(, cok	bahar	1421
Trgrpsh	Nisan'in 8'i olmus, bahar gelmis, sevgi, ask, kardeslik geleceğine yine terör geliyor. Bu dünyayi bu hale getirenler yok olsun..	1	7	ask, sevgi, teror	bahar	0
starboido	RT @iamimjae: Bahar geldigi için çiçek getiren Ahgase sayisi artiyor. Ayrica getirdiginiz çiçeklerin	1	15	?, muhtesem, cicek	bahar	1

	hepsi muhtesem kokuyordu. ?????					
--	------------------------------------	--	--	--	--	--

4.1.2 Büyük Veri Kümesi Üzerinden Mevsimsel Değişimlerle ilgili Analiz Çalışmaları

Bu kapsamda dört mevsim için incelemeler yapılmıştır. Bu deneyler kapsamında 8788 adet veriden oluşan büyük veri kümesi olarak adlandırılan veri kümesi kullanılmıştır. Bu deneylerde amaçlanan mevsim değişikliklerinin twitter üzerinden Türkçe tweet paylaşan kullanıcıların duygularında ki değişimi analiz etmektir. Kullanıcıların yaz aylarından bunalıp kışa olan özlemleri tweet sayısının yüksekliğinden görüntülenmiştir. Ayrıca kullanıcıların yaz mevsiminde paylaştıkları tweetlerin olumluluk oranının kış aylarında paylaştıkları tweetlerin olumluluk oranına göre yüksek olduğu görüntülenmiştir. Böylelikle SA sonrasında karşılaşılan sonuçların beklenen değerlerle uyumlu olduğu gözlemlenmiştir.

4.1.2.1 Mevsimlerin Kullanıcılar Üzerindeki Etkisinin Analizi

Türkçe olarak paylaşılan ve Sentimental Analiz çalışmalarımıza dahil edilen 8788 tekil tweet ile deneyler gerçekleştirilmiştir. Bu analiz çalışmaları için 2017 Mayıs, Haziran, Ağustos, Eylül, Ekim aylarında twitter dan veriler toplanmıştır. Bu toplanan veriler için mevsim isimleri kullanılmıştır. Kullanılan etiketler ilkbahar, yaz, sonbahar ve kıştır. Ayrıca veriler incelendikten sonra Türkçe tweet atan kullanıcıların İngilizce hashtag kullanma alışkanlıklarının görülmesinin ardından da İngilizce mevsim etiketleri üzerinden de çalışmalar yapılmıştır (spring, summer, fall, winter).

Bu veri setinde yer alan veriler incelendiğinde aşağıdaki sonuçlar görüntülenmektedir (Bkz. 4.1:3 Mevsimsel Olarak Paylaşılan Tweet Oranları):

- i. Mevsim geçişi olan sonbahar mevsiminde kullanıcıların daha çok tweet paylaştıkları görüntülenmiştir. Bu nedenle en çok dikkat çeken Sonbahar hashtag 'inin yüzde 63 oranda kullanılmasıdır.

- ii. İlkbahar aylarının verilerin toplanma tarihinden önceki aylarda olması nedeniyle ve verilerin çekildiği mevsimlerin yaz sonbahar aylarına denk gelmesi nedeniyle en az veri yüzde 2 ile İlkbahar'dır.
- iii. Yaz ve kış mevsimlerinde ise birbirine yakın oranlarda tweetler paylaşıldığı görülmüştür.

Sadece resim paylaşımı yapan ve bu resmi hashtag ile işaretleyen tweetler üzerinden duygu analizi yapılamamış bu veriler “nötr” olarak işaretlenmiştir.

Şekil 0:3 Mevsimsel Olarak Paylaşılan Tweet Oranları

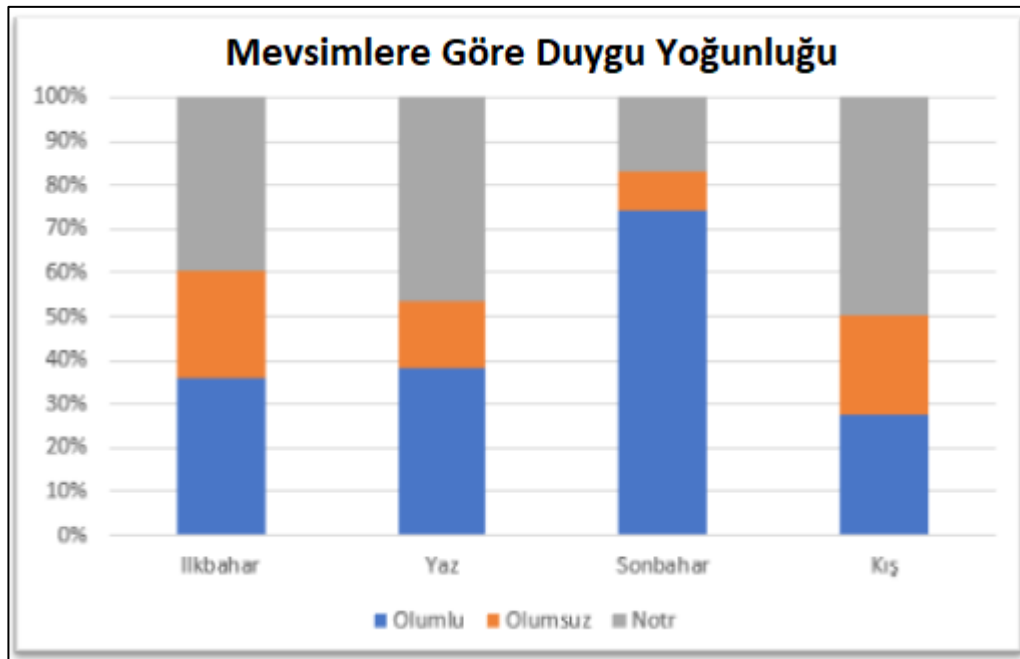


Türkçe tweet paylaşan kullanıcıların verileri üzerinden yapılan mevsimlerin kullanıcıların duygularını (olumlu, olumsuz, nötr) etkileme oranlarında beklendiği gibi yaz aylarında daha “olumlu” ve “pozitif” paylaşımlarda bulundukları gözlemlenmiştir. Ayrıca kullanıcıların çok fazla resim paylaştıkları da gözlemlenmiştir. Ek olarak sonbahar aylarında da kullanıcıların “olumlu” ve “pozitif” tweetlerinin oranının nötr ve olumsuzlara göre daha fazla olduğu gözlemlenmiştir.

Mevsimplere göre kullanıcıların paylaştığı tweetlerin oranları Şekil 4.1:4 de görüntülendiği gibidir. Tweetler incelendiğinde ise: Kış mevsiminde paylaşılan

tweetlerden yüzde 27'sinin olumlu, yüzde 23'ünün olumsuz ve yüzde 50'sinin nötr olduğu görülmektedir. Sonbahar mevsiminde paylaşılan tweetlerden yüzde 74'ü olumlu, yüzde 9'u olumsuz ve yüzde 17'si ise nötrdür. Yaz mevsiminde paylaşılan tweetlerin yüzde 38'inin olumlu, yüzde 15'inin olumsuz ve yüzde 47'sinin ise nötr olduğu görülmektedir. Ilkbahar mevsiminde paylaşılan tweetlerin ise yüzde 36'sının olumlu, yüzde 24'ünün olumsuz ve yüzde 40'ının nötr olarak görülmüştür.

Şekil 0:4 Mevsimlere Göre Duygu Yoğunluğu



4.2 HAFTANIN GÜNLERİNİN KULLANICILAR ÜZERİNDEKİ ETKİSİNİN ANALİZİ

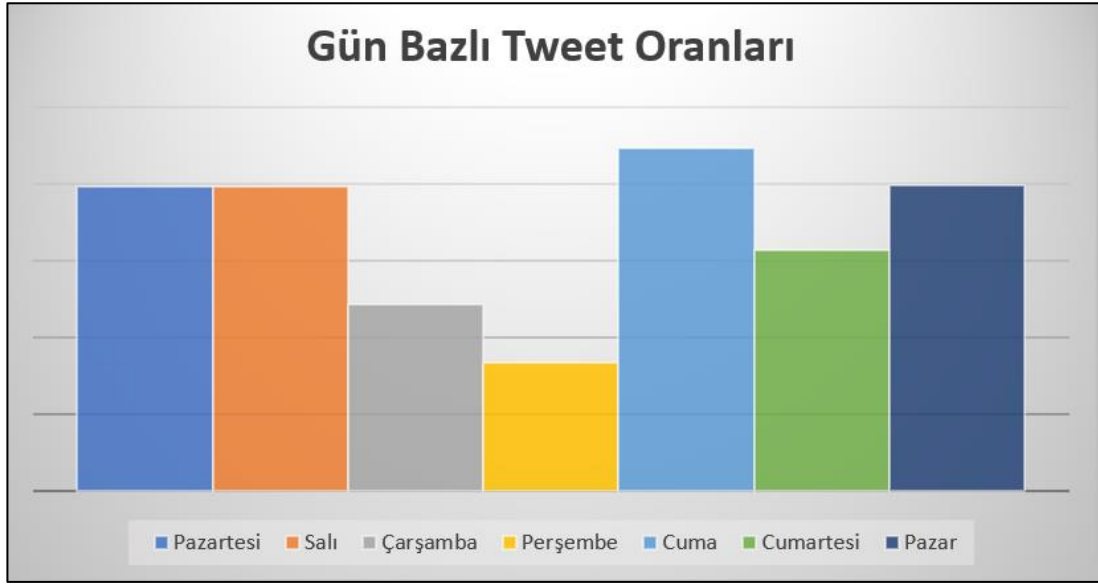
Haftanın günlerine göre Türkçe tweet paylaşan twitter kullanıcıların paylaştıkları tweetler üzerinden günlerin kullanıcıların duygularında ki etkileşimi ölçmek amacıyla iki farklı veri kümesiyle Sentimental Analiz çalışmaları yapılmıştır. Böylelikle kısıtlı veri kümesi ve kapsamlı veri kümesinin arasında ki fark gözlemlenmek istenmiştir. Sonuç beklendiği gibi farklılık göstermektedir. Örneğin az veriyle gerçekleştirilen analiz çalışmasında toplanan tweetlerde en çok tweetin Cuma günü paylaşıldığı görülmüştür fakat çok veriyle yapılan analiz çalışmalarında Pazartesi günü daha fazla tweetin paylaşıldığı görülmüştür.

4.2.1 Deneysel Kısıtlı Veri Kümesi Üzerinden Gün Bazlı Analiz Çalışmaları

Sentimental analiz çalışması için toplanan verilerde etiket olarak “pazartesi”, “salı”, “çarşamba”, “perşembe”, “cuma”, “cumartesi” ve “pazar” etiketleri kullanılmıştır. Bu deneyler kapsamında 2500 adet veriden oluşan kısıtlı veri kümesi olarak adlandırılan veri kümesi kullanılmıştır.

Toplanan veri sayısı 2500’dir. Bu verilerin yüzde 18’si Pazartesi, yüzde 16’sı Salı, yüzde 10’u Çarşamba, yüzde 7’si Perşembe, yüzde 19’u Cuma, yüzde 13’ü Cumartesi ve yüzde 17 ‘si Pazar gününe aittir. Aşağıda ki Şekil 4.2:1 Gün Bazlı Tweet Oranları grafik olarak görüntülenmiştir.

Şekil 0:1 Gün Bazlı Tweet Oranları Grafiği

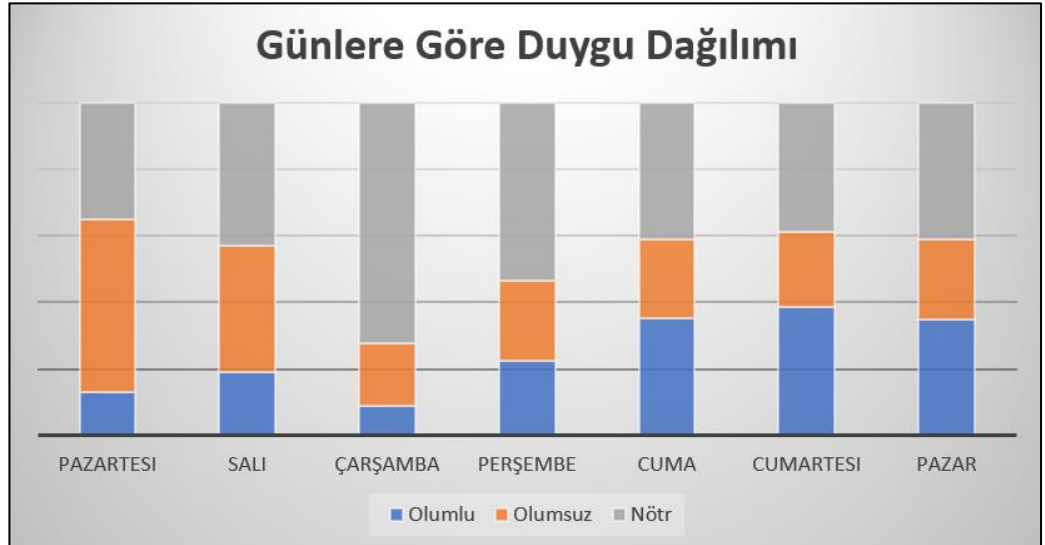


Kısıtlı veri kümesiyle yapılan incelemeler sonrasında, yukarıda belirtilen etiketlerle en çok cuma günü, en az ise perşembe günü tweet atılmıştır.

Haftanın değişik günlerinde atılan tweetlerin analiz çalışması yapıldığında ise Şekil 4.2:2 Günlere Göre Duygu Dağılımı Grafiğinde yer alan veriler elde edilmiştir. Bu veriler özet olarak aşağıdaki gibidir:

- i. Pazartesi günleri atılan tweetlerin yüzde on üçünün olumlu, yüzde elli ikisinin olumsuz ve yüzde otuz beşinin nötr olduğu,
- ii. Salı günleri atılan tweetlerin yüzde on dokuzunun olumlu, yüzde otuz sekizinin olumsuz ve yüzde kırk üçünün nötr olduğu,
- iii. Çarşamba günleri atılan tweetlerin yüzde dokuzunun olumlu, yüzde on dokuzunun olumsuz ve yüzde yetmiş ikisinin nötr olduğu,
- iv. Perşembe günleri atılan tweetlerin yüzde yirmi üçünün olumlu, yüzde yirmi dördünün olumsuz ve yüzde elli dördünün nötr olduğu,
- v. Cuma günleri atılan tweetlerin, yüzde otuz beşinin olumlu, yüzde yirmi dördünün olumsuz ve yüzde kırk birinin nötr olduğu,
- vi. Cumartesi günleri atılan tweetlerin yüzde otuz dokuzunun olumlu, yüzde yirmi üçünün olumsuz ve yüzde otuz dokuzunun nötr olduğu,
- vii. Pazar günleri atılan tweetlerin yüzde otuz beşinin olumlu, yüzde yirmi dördünün olumsuz ve yüzde kırk birinin nötr olduğu ölçümlenmiştir.

Şekil 0:2 Günlere Göre Duygu Dağılımı Grafiği



Genel olarak Türkçe atılan tweetler incelendiğinde olumlu tweet sayısının olumsuz tweet sayısından az olduğu gözlemlenmiştir.

Beklenen durum Cuma, Cumartesi günleri kullanıcıların mutluluk oranlarının pozitif yönde arttığını görmektir. Kısacası kullanıcıların daha olumlu metinlerle Türkçe tweet paylaştıklarını görmektir. Bu beklentiyle toplanan veri kümeleri üzerinde yapılan analiz

çalışmaları sonrasında kullanıcıların Cuma, Cumartesi ve Pazar günlerinde, haftanın diğer günlerine oranla daha olumlu tweetler paylaştıkları gözlemlenmiştir. Haftanın günlerine göre kullanıcıların paylaştıkları bazı tweet örnekleri Tablo 4.2:1 de bulunmaktadır.

Örnek1: oksurukluhoroz kullanıcı tarafından paylaşılan “pazartesi” etiketiyle toplanan “Pazartesi sabahi "bu ne lan dünün aynisi" karikatürünün altında bulusmak dilegiyle” tweetinde olumsuz duygu değerine sahip kelime bulunmaktadır. Bu kelime” lan”dır. Burada yer alan kelimenin duygu değeri “-3” dür. Tweet olumsuz olarak gruplandırılmıştır ve Toplam duygu değeri böylelikle -3 olarak belirlenmiştir. Bu örnekte yer alan metin genel olarak mutsuzluk ifade etmektedir. Bu nedenle Pazartesi etiketiyle paylaşılan bu Tweet’in sonucu “olumsuz” olarak işaretlenmiştir ve sonuç doğrudur.

Örnek2: serhatulkerr54 kullanıcı tarafından paylaşılan “cumartesi” etiketiyle toplanan “Muhtesem bi cumartesi oluyo arkadaşlar ” tweetinde olumlu duygu değerine sahip kelime bulunmaktadır. Bu kelime” muhteşem”dir. Burada yer alan kelimenin duygu değeri “11” dür. Tweet olumlu olarak gruplandırılmıştır ve Toplam duygu değeri böylelikle 11 olarak belirlenmiştir. Bu örnekte yer alan metin genel olarak mutluluk ifade etmektedir. Duygu kütüphanesinden de “muthesem” kelimesi yakalanmıştır ve duygu değeri “olumlu” olarak işaretlenmiştir.

Tablo 4:3: Haftanın Günlerine Göre Duygu Yoğunluğu Verileri

User_Name	Twit_Text	Impression	ResultValue	ContainingKeys	SearchKey	RT_Count
oksurukluhoroz	Pazartesi sabahi "bu ne lan dünün aynisi" karikatürünün altında bulusmak dilegiyle	2	-3	lan	Pazartesi	0
NuriyeYetis	DidemDidemdelen: Sali günü benim için çok önemli ben bu sevgiye önce Didem'le basladim sonra Addid olarak devam ettim	1	22	deva, sevgi, cok	Sali	6

filiz123ylmz	RT @olenekadartv: Kirli oyunlarin arasinda filizlenmeye çalisan tertemiz bir ask... #ÖleneKadar, yeni bölümüyle Perşembe 20:00'de @atvcomtr...	1	10	Kirli, tertemiz, ask	Perşembe	996
ceypaldi	RT @esocans: Birbirimizi Cuma günü kadar heyecanlandiramıyorsak Pazar günü kadar da sikmayalım çok rica edicem.	1	11	cok, heyecan	Cuma	69
canhalilcan22	Bundan yıllar yıllar sonra bir pazar kahvaltısında aklima düşeceksiniz ve ben anlamsızca gülüp hayatta olup olmadığınızı düşünüceğim.	2	-10	anlamsız, olu	Pazar	0
serhatulkerr54	Muhtesem bi cumartesi oluyo arkadaşlar	1	11	Muhtesem	cumartesi	0

4.2.2 Deneysel Kapsamlı Veri Kümesi Üzerinden Gün Bazlı Analiz Çalışmaları

Sentimental analiz çalışması için toplanan verilerde etiket olarak “pazartesi”, “salı”, “çarşamba”, “perşembe”, “cuma”, “cumartesi” ve “pazar” etiketleri kullanılmıştır. Bu deneyler kapsamında 14000 adet veriden oluşan Geniş kapsamlı veri kümesi olarak adlandırılan veri kümesi kullanılmıştır. Ayrıca İngilizce olarak da etiketler belirlenmiştir ve bu etiketlere ait veriler de toplanmıştır (Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday). Toplanan veri sayısı 14.000 civarındır. Bu verilerin yüzde 46’sı Pazartesi, yüzde 3’ü Salı, yüzde 3’ü Çarşamba, yüzde 4’ü Perşembe, yüzde 13’ü Cuma, yüzde 18’i Cumartesi ve yüzde 13’ü Pazar gününe aittir (Bkz. Şekil 4.2:3 Gün Bazlı Tweet Oranları)

Şekil 0:3 Gün Bazlı Tweet Oranları

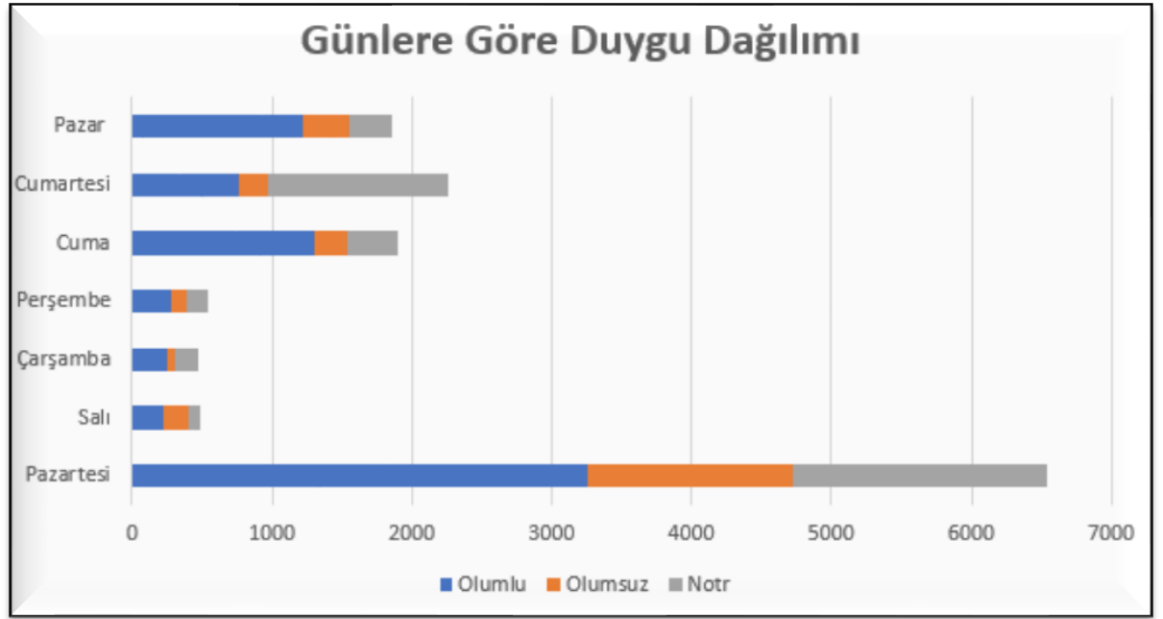


Deneyssel kapsamlı veri kümesi ile yapılan incelemeler sonrasında, yukarıda belirtilen etiketlerle en çok Pazartesi günü, en az ise Salı ve Çarşamba günleri tweet atılmıştır.

Haftanın farklı günlerine ait paylaşılan tweetlerin analiz çalışması yapıldığında özet olarak aşağıdaki sonuç elde edilmektedir (Bkz. 4.2:4 Günlere Gore Duygu Dağılımı):

- Pazartesi günleri atılan tweetlerin yüzde ellisinin olumlu, yüzde yirmi ikisinin olumsuz ve yüzde yirmi sekizinin nötr olduğu,
- Salı günleri atılan tweetlerin yüzde kırk beşinin olumlu, yüzde otuz altısının olumsuz ve yüzde on dokuzunun nötr olduğu,
- Çarşamba günleri atılan tweetlerin yüzde elli üçünün olumlu, yüzde on üçünün olumsuz ve yüzde otuz dördünün nötr olduğu,
- Perşembe günleri atılan tweetlerin yüzde elli ikisinin olumlu, yüzde yirmisinin olumsuz ve yüzde yirmi sekizinin nötr olduğu,
- Cuma günleri atılan tweetlerin, yüzde altmış dokuzunun olumlu, yüzde on ikisinin olumsuz ve yüzde on dokuzunun nötr olduğu,
- Cumartesi günleri atılan tweetlerin yüzde otuz dördünün olumlu, yüzde dokuzunun olumsuz ve yüzde elli yedisinin nötr olduğu,
- Pazar günleri atılan tweetlerin yüzde altmış altısının olumlu, yüzde on yedisinin olumsuz ve yüzde on yedisinin nötr olduğu ölçümlenmiştir.

Şekil 0:4 Günlere Göre Duygu Dağılımı

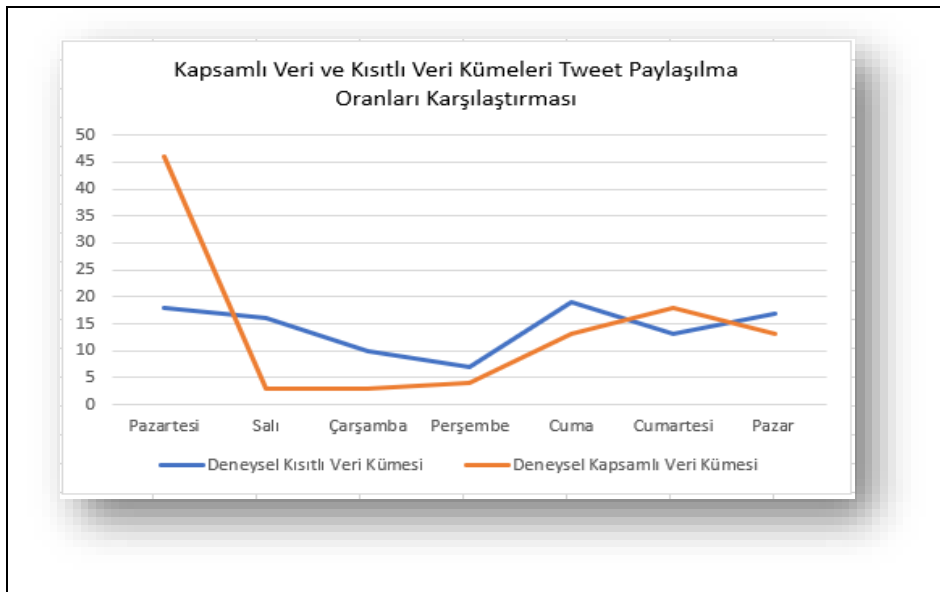


Deneysel kısıtlı veri kümesi ve deneysel kapsamlı veri kümeleriyle yapılan çalışmalar oransal bazda kıyaslandığında:

- i. Pazartesi hashtagiyle paylaşılan:
 - a. Toplam tweet oranlarının yüzde 28 yükseldiği,
 - b. olumluluk oranının yüzde 37 arttığı,
 - c. olumsuzluk oranının yüzde 30 azaldığı,
 - d. nötr olma oranının yüzde 7 azaldığı görülmüştür.
- ii. Salı hashtagiyle paylaşılan:
 - a. Toplam tweet oranlarının yüzde 13 düştüğü,
 - b. olumluluk oranının yüzde 26 arttığı,
 - c. olumsuzluk oranının yüzde 2 azaldığı,
 - d. nötr olma oranının yüzde 24 azaldığı görülmüştür.
- iii. Çarşamba hashtagiyle paylaşılan:
 - a. Toplam tweet oranlarının yüzde 7 düştüğü,
 - b. olumluluk oranının yüzde 44 arttığı,
 - c. olumsuzluk oranının yüzde 6 azaldığı,
 - d. nötr olma oranının yüzde 38 azaldığı görülmüştür.
- iv. Perşembe hashtagiyle paylaşılan:

- a. Toplam tweet oranlarının yüzde 3 düştüğü,
 - b. olumluluk oranının yüzde 29 arttığı,
 - c. olumsuzluk oranının yüzde 4 azaldığı,
 - d. nötr olma oranının yüzde 26 azaldığı görülmüştür.
- v. Cuma hashtagiyle paylaşılan:
- a. Toplam tweet oranlarının yüzde 6 düştüğü,
 - b. olumluluk oranının yüzde 34 arttığı,
 - c. olumsuzluk oranının yüzde 12 azaldığı,
 - d. nötr olma oranının yüzde 22 azaldığı görülmüştür.
- vi. Cumartesi hashtagiyle paylaşılan:
- a. Toplam tweet oranlarının yüzde 5 yükseldiği,
 - b. olumluluk oranının yüzde 5 azaldığı,
 - c. olumsuzluk oranının yüzde 14 azaldığı,
 - d. nötr olma oranının yüzde 18 arttığı görülmüştür.
- vii. Pazar hashtagiyle paylaşılan:
- a. Toplam tweet oranının da yüzde 4 düştüğü
 - b. olumluluk oranının yüzde 31 arttığı,
 - c. olumsuzluk oranının yüzde 7 azaldığı,
 - d. nötr olma oranının yüzde 24 azaldığı görülmüştür.

Şekil 0:5 Kapsamlı ve Kısıtlı Veri Kümeleri Tweet Paylaşım Oranları Kıyaslaması



Twitter üzerinden toplanan 4 aylık veri kümesinde görüldüğü üzere Pazartesi ve Cumartesi etiketleriyle paylaşılan tweetlerin toplam atılan tweetlerin yarısından fazla olduğunu gözlemlenmiştir. Ayrıca Tablo 4.4 de Kısıtlı ve Kapsamlı Veri kümelerinden çıkan sonuçların özeti çıkartılmıştır.

Tablo 4:4 Kısıtlı ve Kapsamlı Veri Kümelerinden Bazı Değerler

	Kısıtlı Veri Kümesi	Kapsamlı Veri Kümesi
En çok tweet atılan gün	Cuma ve Pazartesi	Pazartesi
En az tweet atılan gün	Çarşamba ve Perşembe	Salı ve Çarşamba
En olumlu tweet atılan gün	Cuma, Cumartesi ve Pazar	Cuma ve Pazar
En olumsuz tweet atılan gün	Pazartesi	Pazartesi ve Salı

Yapılan analiz çalışmaları sonrasında karşılaşılan az veriden oluşan veri kümesinde görüntülenen sonuçlarla kapsamlı veriden oluşan veri kümesinin sonuçlarının Şekil 4.2:5 de de görüntülediği gibi fark yaratması big data çalışmalarında ki veri setinin öneminin bir kere daha anlaşılmasını sağlamıştır.

4.3 SURVIVOR İZLEYEN KULLANICILARIN DUYGU ANALİZİ

Türk Televizyonlarında yer alan dünyaca ünlü yarışma programıdır. Trend Topic listesinde günlerce ilk sırada yer almıştır. Bu nedenle survivor etiketiyle paylaşılan tweetler arşivlenmiştir. Daha sonrasında ise veriler üzerinde gözle kontroller yapılmıştır. Farklı ve değişik yargılarla karşılaşılmıştır. Hep bahsedilen bozuk Türkçe kullanma alışkanlarına çok iyi örnek olması nedeniyle tez kapsamına alınmıştır. Yapılan bu kontroller sonrasında Duygu Analizi algoritmaları kullanılarak analiz işlemleri gerçekleştirilmiştir. Sentimental analiz çalışması sonrasında “Survivor” yarışmasıyla ilgili Türkçe paylaşılan tweetlerin yüzde 27’sinin olumlu, yüzde 34’ünün olumsuz etkilendiği görülmektedir. Ayrıca geriye kalan yüzde 39’luk kitle tarafından atılan tweetler ise bozuk Türkçe kullanımı ve görsel paylaşım yapmaları nedeniyle nötr olarak kaydedilmiştir ve bu veriler üzerinde duygu analizi çalışmaları yapılamamıştır. Yapılan bu analize ait grafik Şekil 4.3:1 de yer almaktadır. Bu çalışma için toplam 766 tweet

incelenmiştir. Sonuç olarak Türkçe tweet atan kullanıcıların survivor gibi bir eğlence programı için bile genel olarak olumsuz duyguları daha çok ifade etme alışkanlığına sahip olduğu gözlemlenmiştir. Beklenen sonuç ise bu kadar kişinin takip ettiği bir yarışma programı etketinin “olumlu ve mutlu” duygu değerine sahip sonuçlar üretmesidir. Fakat Türkçe tweet paylaşan kullanıcıların yorumlarının negatif ağırlıklı olması nedeniyle sonuç “olumsuz” ağırlıklı çıkmıştır bu durum esas ispat edilmek istenen durumu verilerle ispat etmiştir. Toplanan verilerden bir kaç örnek aşağıdaki Tablo 4:4 de bulunmaktadır.

Şekil 4.3:1 “survivor” Etiketine Göre Duygu Yoğunluk Dağılımı



Örnek1: leyla022002 kullanıcısı tarafından paylaşılan “survivor” etketiyle toplanan “Birisi, "Adamsan gel." diye tahrik eder. Birisi, "Delikanlı değilsin." der. Buram diskalifiye çabası kokuyor.” Tweetinde duygu içeren birden fazla olumsuz kelime yer almaktadır. Bu kelimeler “tahrik, değil, diskalifiye” ‘dir. Burada yer alan kelimelerin duygu değerleri sırasıyla “-3, -5, -4”. Tweet olumsuz olarak gruplandırılmıştır ve Toplam duygu değeri -12 olarak belirlenmiştir.

Bu örnekte yer alan kelimeler duygu kütüphanesinden kontrol edildiğinde eşleşen kelimeler sadece negatif duygu yoğunluk değerine sahip olanlardır. Bu nedenle sonuç otomatik olarak “olumsuz” işaretlenmiştir ve doğrudur.

Örnek2: strongday7 kullanıcı tarafından paylaşılan “survivor” etketiyle toplanan “survivor öyle bir adami destekliyorum ki bu kadar üne sahip olduğu halde öyle mütevazisinki hayran ben sana ilhan Mansız” tweetinde olumlu duygu değerine sahip

kelimeler bulunmaktadır. Bu kelimeler” mütevazı, destek, hayran “dır. Burada yer alan kelimelerin duygu değerleri sırasıyla “3, 2, 5”. Tweet olumlu olarak gruplandırılmıştır ve Toplam duygu değeri 14 olarak belirlenmiştir.

Bu örnekte yer alan kelimeler duygu kütüphanesinden kontrol edildiğinde eşleşen kelimeler sadece pozitif duygu yoğunluk değerine sahip olanlardır. Bu nedenle sonuç otomatik olarak “olumlu” işaretlenmiştir ve doğrudur.

Tablo 4:5: "survivor" Kelimelerine Göre Duygu Yoğunluğu Verileri

User_Name	Twit_Text	Impression	ResultValue	ContainingKeys	SearchKey	RT_Count
elifimmafc	RT @survivornotlari: Ilhan kim lan adami ne çok yücelttiniz ?? Survivor'dan sonra en fazla Pantene reklamında oynar saçlarından dolayı jsjsj...	1	15	lan, ?, yuce, cok	survivor	5
cansu_cansu1223	Fatih Terim'den Ilhan Mansiz'a "ne güzel yürekli adamsin keske çalışabilseymişim" demek için yanına gitti KARIZMA #survivor #i...	1	14	guzel, yurek, adamsin	survivor	20
leyla022002	Birisi, "Adamsan gel." diye tahrik eder. Birisi, "Delikanli degilsin." der. Buram buram diskalifiye çabasi kokuyor. #s...	2	-12	tahrik, degil, diskalifiye	survivor	79
ilfurhat	RT @ridiculous_1213: Kim ne derse desin bu aksam IlSemFurHat damgasi, efsanesi,	1	15	samimi, efsane	survivor	6

	ambargosu ve samimiyeti vardı. #survivor					
EnverMemmedli	Boksörlüğünden dolayı saldırganlık potansiyeli en yüksek olması beklenen Adem'in sesini dahi yükseltmemesi... Bu nasıl bir sa...	2	-10	saldırgan, bok, yüksek	survivor	12
Yigdim__x	ADAM tam anlamıyla kalite gerçekten. Su konuşmaya, karizmaya bak. Sen harikasin SAMPIYON ILHAN MANSIZ ?? #ilhanmansiz #...	1	19	?, harika, sampiyon	survivor	14
sedef92	Ilhan Mansiz en ufak bir elestiriyi kaldıramıyor maalesef. Kendisini elestirenin üzerine yürüyor. Adem haklı lafını esirgem...	2	-5	elestiri, maalesef, haklı	survivor	28
cinonil	Ademin bas parmağı kırılmış sakatlandı bu da mi ilahi adalet??Ama bence değil geçmiş olsun	1	6	sakat, adalet, ilahi, değil	survivor	9
strongday7	#survivor öyle bir adami destekliyorum ki bu kadar üne sahip olduğu halde öyle mütevazisinki hayran ben sana ilhan Mansiz	1	10	mutevazi, destek, hayran	survivor	0

yasyasdemir2	#Survivor ilhan destekledigim ilk günden bugüne iyiki seni tutmusum diyorum oz elestirine hayran kaldim her baba yigidin...	1	7	elestiri, destek, hayran	survivor	1
--------------	---	---	---	-----------------------------	----------	---

5. DEĞERLENDİRME

Bu tez kapsamında Düşünce Kütüphanesi ve kelime yakalama, sözlüksel benzeşim ve istatistiksel gruplama yöntemleri kullanılarak Duygu Analizi yapılmıştır. Buna ek olarak da Varlık seviyesinde ve Cümle seviyesinde Duygu Analizleri yapılmıştır.

Oluşturması oldukça zahmetli olan Düşünce Kütüphanesi birçok kelime, kelime grubu ve basit emojilerden oluşmaktadır. Ayrıca her kelimenin İngilizce klavye kullanılması durumunda oluşacak versiyonları da kütüphanede saklanmaktadır. Karşılaşılması beklenen kelime sayısı 2 katına çıkmaktadır. Türkçe tweet paylaşan kullanıcıların İngilizce klavye kullanma alışkanlıkları gözlenlenmiş ve bu nedenle İngilizce klavye ile Türkçe kelimelerin yazım şekilleri düzenlenmiştir. Ayrıca tweet üzerinden genel puanlama yapabilmek için çalışmalar esnasında karşılaşılan kullanıcıların bazı güzel yorumlarını paylaşmaları esnasında argo ya da küfür içeren kelime kullanmaları nedeniyle puanlama sistememinde argo kelimelere yüksek negatif (-) puan verildiği gibi argo kelimelerin zıt anlamlarına yüksek pozitif (+) puanlar verilmiştir. Yapılan incelemelerde ve deneylerde genelde kullanıcıların argo ve güzel kelimeleri bir arada kullanmasından oluşan anlam kargaşası kelimelere verilen puanlarla büyük ölçüde engellenmeye çalışılmıştır. Fakat bazı durumlarda bu durum engellenememiştir. Örneğin: Trgrpsh kullanıcısı tarafından paylaşılan “Nisan'ın 8'i olmus, bahar gelmiş, sevgi, ask, kardeslik geleceğine yine terör geliyor. Bu dünyayı bu hale getirenler yok olsun.” tweetin de yer alan duygu içeren birden fazla olumlu ve olumsuz kelime yer almaktadır. Bu kelimeler “ask, sevgi, terör” ‘dür. Bu örnek de yer alan metin olumsuz anlam ifade etmektedir fakat analiz sonucunda olumlu olarak işaretlenmiştir. Bunun sebebi metnin içinde yer alan ask ve sevgi kelimelerinin duygu yoğunluk değerlerinin toplamı terör kelimesinin duygu yoğunluk değerinden büyüktür. Bu durumda Türkçemizin esnekliğinden kaynaklanmaktadır.

Esnek bir dil olan Türkçe'nin Duygu Analizi çalışmaları sırasında karşılaşılan sorunları yönetebilmek için birçok algoritma geliştirilmiştir.

Bu tez kapsamında yapılan arařtırmalarda genel olarak Trk kullanıcıların oluřturduėu tweetlerin analiz alıřmaları esnasında karřılařılan sorunlar ařaėıdaki gibi listelenmektedir:

- i. Kullanıcıların hemen hepsinin Trke klavye kullanmamalarından kaynaklanan birok anlam bozukluėu ve eksikliėiyle karřılařılmıřtır. Kullanıcıların Trke metinler yazdıklarını dřnp Trke olmayan kelimelerle tweet atmaları tm kelimelerin ktphanede bulunamamasını saėlamıřtır. Bu nedenle iyileřtirme yapma gereksinimi ortaya ıkmıřtır.
- ii. Trke'nin cmle yapısındaki farklılıklar ve esnek bir dil olması da analiz alıřmaları esnasında byk bir problem yaratmıřtır.
- iii. Toplumumuzun olumsuz duyguları daha ok dile getirme alışkanlıėı nedeniyle yapılan arařtırmalarda genelde negatif sonular elde edilmiřtir. Ek olarak Trk toplumunun az kelime kullanarak ok fazla konuyu ifade etme alışkanlıėı sorunu ile de karřılařılmıřtır.
- iv. Trke 'de yer alan ikilemelerden dolayı da oluřan hesaplama iřlemlerindeki hataları ynetebilmek iinde algoritmalar retilmiřtir. Bir kelime kullanıldıktan sonra tekrar hesaplamaya dahil edilmeyecek řekilde algoritmalar yazılmıřtır.
- v. Twitter 'da yer alan kirli verinin ok olması ve Trklerin ok fazla kfr ierikli tweet atma alışkanlıklarının olması da bir bařka sorun olarak karřılařılmıřtır.

Twitter'a gnderilen bozuk verilerin anlamlandırılması iin NLP algoritmalarından bazıları kullanılmaya alıřılmıřtır. Elde edilen verilerin biroėu duygu iermemektedir ya da eksik harflerle paylařıldıėından analiz alıřmalarında ok veriyle az anlamlı sonular elde edilmiřtir. Bu nedenle veri toplama ve anlamlandırma ařamaları ok uzun zaman almıřtır. Fakat hali hazırda Trke iin istikrarlı olarak NLP veri tabanı olmaması nedeniyle alıřmalarda kullanılamamıřtır.

Geliřtirilen Trke Duygu Ktphanesinin performansı ve zenginleřtirilmesi alıřmaları esnasında yapılacak olan analizler iin toplanan tweetlere incelenmiřtir ve gereksinim dahilinde ktphaneye yeni eklemeler yapılmıřtır. Kapsamda gerekleřtirilen tm analiz alıřmaları iin tek bir Duygu Ktphanesi kullanılmıřtır. Bu analizler 3 bařlık altında

toplanmıştır. Ayrıca veri setinin kalitesi ve sayısının öneminin gör nt lenebilmesi amacıyla da  alıřmalar yapılmıştır.

Bu analizlerden birincisi olarak hava durumu ve mevsimlerin ge iřlerinin twitter  zerinden T rk e tweet paylařan kullanıcıların paylařımları  zerinden duygu analizi  alıřmaları ger ekleřtirilmiřtir. Duygu k t phanesini kullanarak elde ettiĐimiz sonu lar beklendiĐi  zere kiř aylarında insanların hava ile ilgili daha “olumsuz” tepkiler verdiklerini ve daha karamsar olduklarını g stermiřtir. Ayrıca mevsim ge iřlerinde kullanıcıların duygularında olan deĐiřiklikleri  l ebilmek i in yine aynı duygu k t phanesi  zerinden  alıřmalar yapılmıştır. Bu sayede sonbahar mevsiminde kullanıcıların diĐer mevsimlere g re daha fazla duygu aktaran tweet paylařtikları g r nt lenmiştir. Bu durumda hava durumunun kullanıcıları etkilediĐinin bir kanıtı olarak beklentiyi karřılamıştır. Duygu analizi  alıřmalarının sonu ları ger ekte beklenen deĐerler ile uyumlu olup geliřtirilen T rk e duygu k t phanesinde belirtilen kelimelerin duygu deĐerlerinin ger eĐi yansıttıĐını desteklemektedir.

Bu analizlerden ikincisi olarak T rk e tweet paylařan kullanıcıların haftanın g nlerine g re paylařtikları tweetler  zerinden duygu analizi  alıřmaları yapılmıştır. Bu kapsamda farklı tarihlerde oluřturan ve birbirine ait verileri kapsamayan 2 ayrı veri seti kullanılmıştır.

Bu  alıřmalarda haftanın her g n  i in T rk e tweet paylařan kiřilerin duygularında ki deĐiřiklikleri  l mek ve mutluluk oranlarını g r nt leyebilmek amacıyla toplanan veriler kıyaslanmıştır. Ger ekleřtirilen deneyler sonucunda veri setlerinin  nemi bir kez daha anlařılmıştır   nk  kısıtlı veri k mesinde paylařım yapan kullanıcıların Cuma, Cumartesi ve Pazar g nlerinde daha olumlu oldukları ortaya  ıkmıştır. Fakat kapsamlı(b y k) veri k mesinde paylařım yapan kullanıcıların Pazartesi ve Pazar g nleri daha fazla tweet paylařtikları g r nt lenmiştir. B ylelikle T rk e Duygu K t phanesinin de belirlenen kelimelerin duygu deĐerlerinin ger eĐi yansıttıĐı g r nt lenmiştir.

Ek olarak ise kullanıcıların en  ok Pazartesi g n  depresyon, bunalım tarzı kelimeleri i eren tweetler paylařtikları g r lm řt r. Kullanıcıların depresyon, bunalım gibi kelimelerin en az kullanıldıĐı g nler ise Cumartesi ve Pazar’dır. Bu

durumda kullanıcıların bunalımdan en uzak oldukları günlerin hafta sonları olduğunu göstermektedir.

Bu analizlerden üçüncüsü olarak ise geçtiğimiz aylarda reytinglerde çok üst sıralarda olan survivor ile ilgili analiz çalışması yapılmıştır. Bu çalışmayla Twitter üzerinden Türkçe paylaşım yapan kişilerin genelde olumsuz\negatif yorum yapma ve bozuk Türkçe alışkanlıklarını ispat etmek amaçlanmıştır. Twitter Türkiye Trend Topic Listesinde en üst sıralarda olan survivor etiketi ile incelemeler yapılmıştır. Bu televizyon programını seyreden kitlenin çok beğenip takip ettikleri bir yarışma programıyla ilgili olarak paylaşılan tweetler toplanmıştır ve analiz sonucunda “olumsuz” ve “nötr” olarak etiketlenen tweetlerin ağırlıklı oldukları görüntülenmiştir. Sonuçlarda en ağırlıklı ortalamaya sahip olan nötr olarak işaretlenen tweetler incelendiğinde ise çok fazla görsel paylaşıldığı gözlemlenmiştir. Survivor etiketiyle yapılan Türkçe paylaşımların incelendiğinde oranlardan görüldüğü gibi kullanıcılar ağırlıklı olarak negatif yorum yaptıkları görüntülenmiştir ve böylelikle de beklenti karşılanmıştır.

5.1 KISITLAMALAR VE EKSİKLİKLER

Hava durumu ile ilgili kısıtlı veri kümesinde yer alan, Trgrpsh kullanıcısı tarafından paylaşılan “bahar” etiketiyle çekilmiş “Nisan'ın 8'i olmus, bahar gelmiş, sevgi, ask, kardeslik geleceğine yine terör geliyor. Bu dünyayı bu hale getirenler yok olsun.” tweettin de yer alan duygu içeren birden fazla olumlu ve olumsuz kelime yer almaktadır. Bu kelimeler “ask, sevgi, terör” ‘dür. Burada yer alan kelimelerin duygu değerleri sırasıyla “10, 5, -8”. Tweet olumlu olarak gruplandırılmıştır ve Toplam duygu değeri 7 olarak belirlenmiştir. Burada ki sorun ise aşk ve sevgi kelimelerinin toplam duygu değerlerinin terör kelimesinin duygu değerinden daha yüksek çıkmasıdır. Bu örnek de yer alan metin olumsuz anlam ifade etmektedir fakat analiz sonucunda olumlu olarak işaretlenmiştir. Bunun sebebi metnin içinde yer alan ask ve sevgi kelimelerinin duygu yoğunluk değerlerinin toplamı terör kelimesinin duygu yoğunluk değerinden büyüktür. Kısaca işlem $10+5>-8$ ’dir. Bu nedenle sonuç “olumsuz” olarak işaretlenememiş olup “olumlu” olarak işaretlenmiştir ve sonuç yanlıştır.

Kütüphane zenginleştirmeleri yapılabilir. Örneğin rasgele okunan tweetlerde geçen bir kelime belirlenen bir sayı üzerinde kullanılmış ise direkt kütüphaneye ekle gibi yeni algoritmalar yazılarak otomatik bir şekilde geliştirilmesi sağlanabilir. Böylelikle elde edilen nötr verinin bir kısmı daha analiz kapsamına dahil edilerek daha farklı sonuçlar çıkartılabilir. Geliştirilmiş olan bu Türkçe Duygu Kütüphanesi iyi sonuçlar vermektedir ve bununla birlikte bütün kelimeleri içermediğini bilinmelidir. Kütüphanenin daha zenginleşmesi sonuçların mükemmel olmasını sağlayacaktır.

Günümüz teknolojisinde NLP algoritmalarıyla oluşturulmuş istikrarlı bir Türkçe Duygu kütüphanenin bulunmaması nedeniyle NLP çalışmaları çok kısıtlı kalmıştır.

Kullanıcıların birbirlerini etkileme oranları araştırılıp zaman bazlı etkileşim oranları ya da hangi kullanıcıların paylaştığı tweetlerin kaç kullanıcıyı etkilediği analiz edilebilir. Bu çalışmanın ardından da kullanıcıların güvenilirliği konusunda araştırmalar yapılabilir. Bu tez kapsamına ek olarak yapılması planlanabilir çünkü kapsam bazında kullanıcıların paylaştığı tweetlerin beğenilme(like) sayısı, tekrar paylaşılma(retweet) sayısı ve cevaplanma(reply) sayıları tweet bazlı olarak kaydedilmektedir. Etkileşim oranlarının da ilgili tablolara eklenmesinin ardından yapılacak zenginleştirmelerle bu Sentimental Analiz ürününe çok büyük değer kazandırılabilir.

6. SONUÇ

Bu tez kapsamında twitter üzerinden çekilen belli etiketli (hashtag) tweetlerle Duygu Analizi çalışması yapılmıştır. Duygu analiz yöntemi olarak varlık-görüş seviyesi ve cümle seviyesi kullanılmıştır. Ayrıca Düşünce Kütüphanesi geliştirilmiştir. Düşünce Kütüphanesine duygu yönelim gruplamaları ve duygu yoğunluğu parametreleri eklenerek ek özellikler eklenmiştir. Böylelikle cümle seviyesinde yapılan analiz çalışmalarında oluşan anlam kayıplarından çıkan hatalar en aza indirilmiş ve mantıklı sonuçlar üretilmiştir. Ek olarak Düşünce kütüphanesine eklenen emoji ler de duygu analizini destekleyecek şekilde puanlandırılmıştır ve böylelikle emoji lerden de kelimeler gibi faydalanılmıştır. Ayrıca Türkçe de yer alan kelime grupları da kapsama dahil edilmiştir ve duygu yoğunluk değeri olarak daha büyük puanlar verilerek etkime oranı arttırılmıştır.

Türkçe tweetlerde yaşanan sorunları egale edebilmek için kelimelerden Türkçe klavyede yer alan ve İngilizce klavyede yer almayan harflerin yerine neler kullanılabileceği tahminlenerek kütüphane zenginleştirilmiştir. Bu sayede de Twitter kullanan kişilerin Türkçe metinleri İngilizce alfabe kullanarak yazmaları sonrasında oluşan tweetleri de analiz edebilme kabiliyeti kazandırılmıştır. Ayrıca büyük küçük harf karmaşasından oluşabilecek sorunlar içinde çözümler üretilmiştir.

Twitter'dan veri çekme aksiyonu için etiketlerden (hashtag) faydalanılmıştır. Bu tez kapsamında 3 deney yapılmıştır.

Deneylerden birincisi olarak: Kişilerin hava durumuna göre değişen psikolojilerini ölçme amacıyla analiz çalışmaları yapılmıştır. Genel olarak alınan verilerden çıkan sonuç tahminlerle değil verilerle desteklenmiştir. Bu kapsamda iki farklı veri kümesi kullanılmıştır. İlk olarak kısıtlı veri kümesi üzerinden kış ve baharın gelmesi üzerine araştırmalar yapılmıştır. Bu araştırmalar kapsamında kış ve bahar-yaz için veriler toplanmıştır. Bu çalışmalar sonrasında kişilerin duygularının ve psikolojilerinin hava durumunun etkisinin olduğu sonucu ortaya çıkmıştır. Kış aylarında kişilerin mutsuz oldukları yazın gelmesiyle ise mutluluk oranlarının artması gözlemlenmiştir. İkinci olarak ise büyük veri üzerinden kullanıcıların duygularında ki değişimler ölçülmüştür. Mevsimlerin değişmesiyle kullanıcıların duygularının da değiştiği görülmüştür. Bu

nedenle en fazla verinin Sonbahar ayında paylaşıldığı görüntülenmiştir. Kişilerin yaz ve ilkbahar aylarında daha mutlu tweetler paylaştıkları gözlemlenmiştir.

Deneylerden ikincisi olarak: Kişilerin günlere göre değişen duygu yoğunluklarını ölçmek adına tweetler toplanmıştır. Bu deney kapsamında da iki farklı veri kümesi kullanılmıştır. Kısıtlı veri kümesi üzerinde yapılan analiz çalışmalarında kişilerin haftanın 2. Ve 3. Günlerinde (Salı ve Çarşamba) daha az tweet paylaştıkları görüntülenmiştir ayrıca kişilerin Cuma, Cumartesi, Pazar daha olumlu tweetler paylaştıkları görüntülenmiştir. Ek olarak pazartesi günleri de kişilerin olumsuz olarak tweet attıkları görüntülenmiştir. Bu durumda kullanıcıların hafta sonunun gelmesiyle duygularının olumlu olarak etkilendiğini göstermektedir. Ayrıca Pazartesi günlerinde tatil sonrası erken kalkma ve işe/okula başlama stresi nedeniyle olumsuz duygu paylaşımları yaptıklarını göstermektedir. Büyük veri üzerinden yapılan analiz çalışmalarında ise kullanıcıların genel ortalama olarak bakıldığında daha olumlu tweetler paylaştıkları görülmüştür. Kısıtlı veri kümesiyle büyük veri kümesinde yer alan tweetler kıyaslandığında Pazartesi günleri paylaşılan tweetlerin olumluluk oranının arttığı gözlemlenmiştir, Salı günleri paylaşılan tweetlerde ise nötr olma oranının da azaldığı gözlemlenmiştir fakat Pazar günü paylaşılan tweetler incelendiğinde ise hem tweet oranının düştüğü hem de negatif tweetlerin oranlarının arttığı gözlemlenmiştir. En çok duygu yüklü tweet paylaşılan günler ise sırasıyla: Pazartesi, Cumartesi, Pazar, Cuma'dır. Kısıtlı veride olduğu gibi en az tweet paylaşımı yine Salı günleri olmuştur.

Deneylerden üçüncüsü olarak da Türkçe tweet paylaşan Twitter kullanıcılarının bozuk dil kullanma, genelde olumsuz durumları daha çok ifade etme alışkanlıklarını ölçmek adına survivor etiketi ile atılan tweetler incelenmiştir. Survivor izleyen kitlenin genel olarak olumsuz ve negatif içerikli tweetler paylaştıkları sonucu çıkartılmıştır.

Gelecek Dönemde Yapılması Planlanan Detaylar

İlişkisel veri tabanı yerine daha esnek ve daha çok datayı sorunsuz bir şekilde saklayan NoSQL veri tabanları kullanılması planlanmaktadır. Böylelikle daha esnek ve daha hızlı bir yapı elde edilebilecektir.

Ara yüzde de düzenlemeler yapılması planlanmaktadır. Etiket(Keyword) bağımlılığı olmadan seçilen ülke üzerinden ya da tarih aralığı üzerinden verilerin NoSQL (Mongo, Hadoop vs.) veri tabanına kaydedilmesi planlanmaktadır. Bu durumda da kaydedilen tweetlerin içinde yer alan etiket bilgileri için veri tabanında ek alan açılma ihtiyacı duymadan ilişkisiz olarak saklanabilir ve bu veri kümesi üzerinden raporlar üretilebilir. Daha sonra da bu üretilen raporlar sayesinde yeni keşifler çıkartılabilir ve bu keşifler ışığında seçilen etiketlerden faydalanıp tekrardan veri toplama aksiyonları gerçekleştirilebilir.

Kütüphanede yer alan basit emojiilerin yanı sıra daha yeni versiyon olan Emoji listesi de eklenerek kütüphanede zenginleştirme çalışmaları yapılabilir ve böylelikle analiz sonuçları daha kapsamlı hale getirilebilir.

KAYNAKÇA

- [1] T. Nasukawa and J. Yi, "Sentiment Analysis: Capturing Favorability Using Natural Language Processing," in *Proceedings of the 2Nd International Conference on Knowledge Capture*, New York, NY, USA, 2003, pp. 70–77.
- [2] K. Dave, S. Lawrence, and D. M. Pennock, "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews," in *Proceedings of the 12th International Conference on World Wide Web*, New York, NY, USA, 2003, pp. 519–528.
- [3] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs Up?: Sentiment Classification Using Machine Learning Techniques," in *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10*, Stroudsburg, PA, USA, 2002, pp. 79–86.
- [4] A. Eyup Sercan, E. Caner, and D. Banu, "Twitter verileri ile duygu analiz." [Online]. Available: <http://docplayer.biz.tr/28952591-Pamukkale-universitesi-muhendislik-bilimleri-dergisi-pamukkale-university-journal-of-engineering-sciences.html>. [Accessed: 26-Sep-2017].
- [5] C. Anais, C. Crina, J. Damin, H. Omar, and B. Lionel, "A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation." .
- [6] Y. Choi and C. Cardie, "Learning with Compositional Semantics As Structural Inference for Subsentential Sentiment Analysis," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Stroudsburg, PA, USA, 2008, pp. 793–801.
- [7] S. Wang and C. D. Manning, "Baselines and Bigrams: Simple, Good Sentiment and Topic Classification," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers - Volume 2*, Stroudsburg, PA, USA, 2012, pp. 90–94.
- [8] A. Mudinas, D. Zhang, and M. Levene, "Combining Lexicon and Learning Based Approaches for Concept-level Sentiment Analysis," in *Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining*, New York, NY, USA, 2012, p. 5:1–5:8.
- [9] A. Yadollahi, A. G. Shahraki, and O. R. Zaiane, "Current State of Text Sentiment Analysis from Opinion to Emotion Mining," *ACM Comput Surv*, vol. 50, no. 2, p. 25:1–25:33, May 2017.
- [10] L. Bing, *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, 2012.
- [11] M. O. Shiha and S. Ayvaz, "The Effects of Emoji in Sentiment Analysis," 10.17706/ijcee.2017.9.1.360-369, Feb. 2017.
- [12] U. Alaettin, "AUTOMATIC SENTIMENT DICTIONARY TRANSLATION AND USING IN SENTIMENT ANALYSIS." 2014.
- [13] S. Mehmet Ulvi, "Sosyal Ağlarda Veri Madenciliği Üzerine Bir Uygulama," GAZİ ÜNİVERSİTESİ, FEN BİLİMLERİ ENSTİTÜSÜ, 2012.
- [14] W. Janyce M., B. Rebecca F., and O. Thomas P., "Development and Use of Gold Standard Data Set for Subjectivity Classifications," *Dep. Comput. Sci. Comput. Res. Lab. N. M. State Univ. Las Cruces NM 88003*.

- [29] W. Christine B. and G. Girish Jeff, *What is a Social Network Worth? Facebook and Vote Share in the 2008 Presidential Primaries*. Boston, MA: Department of International Studies and Government Bentley College, 2008.
- [30] C. Cihan, E. Mustafa, K. Evren, T. H. Kaan, and A. Duygu, "Sosyal Medyanın Politik Katılım ve Hareketlerdeki Rolü," *Bilkent Univ. Bilgi. Teknol. Ve Bilişim Sist. Bölümü Ank.*
- [31] J. Vargas, "Obama raised half a billion online." *Washington Post*, 20-Nov-2008.
- [32] Y. A. Aylin and S. Elif, "Social Media in Social Organization," *Eur. Sci. J.*, Jul. 2017.
- [33] M. Akgul, "Internet Yasaklari, Bilgi Toplumu ve Demokrasi," *blog.akgul.web.tr*, 19-Jul-2011. .
- [34] C. Duncombe, "The Twitter revolution? Social media, representation and crisis in Iran and Libya," presented at the Australian Political Science Association Conference (APSA) 2011, 2011, pp. 1–12.
- [35] P. D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," *arXiv:cs/0212032*, Dec. 2002.
- [36] R. M. Duwairi, "Sentiment analysis for dialectical Arabic," in *2015 6th International Conference on Information and Communication Systems (ICICS)*, 2015, pp. 166–170.
- [37] I. N. Nassar and F. T. Khamayseh, "Constructing activity diagrams from Arabic user requirements using Natural Language Processing tool," in *2015 6th International Conference on Information and Communication Systems (ICICS)*, 2015, pp. 50–54.
- [38] M. Wollmer *et al.*, "YouTube Movie Reviews: Sentiment Analysis in an Audio-Visual Context," *IEEE Intelligent Systems*, vol. 28, no. 3, pp. 46–53, 2013.
- [39] J. Steinberger *et al.*, "Creating Sentiment Dictionaries via Triangulation," in *Proceedings of the 2Nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, Stroudsburg, PA, USA, 2011, pp. 28–36.
- [40] S. Mohammad, C. Dunne, and B. Dorr, "Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus," in *Proceedings of the 2009 Conference on EMNLP*, 2009.
- [41] K. Jaap, M. MJ, M. Robert J., and R. Maarten de, *Using WordNet to measure semantic orientations of adjectives*. European Language Resources Association (ELRA), 2004.
- [42] "Esuli A, Sebastiani E (2006) SentiWordNet: a publicly available lexical resource for opinion mining. In: Proceedings of language resources and evaluation (LREC)."
- [43] "Statistic Brain. STATS | Twitter company statistics. Statistic Brain.," 30-Jan-2016. .
- [44] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a Social Network or a News Media?," in *Proceedings of the 19th International Conference on World Wide Web*, New York, NY, USA, 2010, pp. 591–600.
- [45] N. Dokoochaki, F. Zikou, D. Gillblad, and M. Matskin, "Predicting Swedish Elections with Twitter: A Case for Stochastic Link Structure Analysis," in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, New York, NY, USA, 2015, pp. 1269–1276.

- [46] K. O'Connor, P. Pimpalkhute, A. Nikfarjam, R. Ginn, K. L. Smith, and G. Gonzalez, "Pharmacovigilance on Twitter? Mining Tweets for Adverse Drug Reactions," *AMIA. Annu. Symp. Proc.*, vol. 2014, pp. 924–933, Nov. 2014.
- [47] J. Lardon *et al.*, "Adverse Drug Reaction Identification and Extraction in Social Media: A Scoping Review," *J. Med. Internet Res.*, vol. 17, no. 7, Jul. 2015.
- [48] L. Nguyen, P. Wu, W. Chain, W. Pang, and Y. Zhang, "Predicting Collective Sentiment Dynamics from Time-series Social Media," presented at the ACM SIGKDD WISDOM, 2012.
- [49] W. B. Claster, H. Dinh, and M. Cooper, "Naïve Bayes and unsupervised artificial neural nets for Cancun tourism social media data analysis," in *2010 Second World Congress on Nature and Biologically Inspired Computing (NaBIC)*, 2010, pp. 158–163.
- [50] E. Evirgen, "Türkçe Tweetlerin Duygu Analizi," Bacesehir Üniversitesi, 2016.
- [51] D. Kayahan, A. Sergin, and D. Banu, "Twitter ile TV Program Reytinglerinin Belirlenmesi," Bilgisayar Mühendisliği Bölümü Yıldız Teknik Üniversitesi.
- [52] O. Kaynar, Y. Görmez, M. Yıldız, and A. Albayrak, "Makine Öğrenmesi Yöntemleri ile Duygu Analizi - Sentiment Analysis with Machine Learning Techniques," 2016.
- [53] S. Demirci, "Emotion Analysis on Turkish Tweets," Middle East Technical University, 2014.
- [54] Ö. Çoban, G. Tumuklu Ozyer, and B. Ozyer, "Sentiment Analysis for Turkish Twitter Feeds," in *2015 23rd Signal Processing and Communications Applications Conference, SIU 2015 - Proceedings*, 2015.
- [55] N. Ozturk and S. Ayyaz, "Sentiment Analysis on Twitter: A Text Mining Approach to the Syrian Refugee Crisis. Telematics and Informatics," *Bahcesehir Univ.*, 2017.
- [56] D. M. Gordin, *Scientific Babel: How Science Was Done Before and After Global English*. Chicago, Illinois: University of Chicago Press, 2015.
- [57] S. E. Seker, "Metin Madenciliği (Text Mining)," *MISSözlük*, 15-Jun-2014. .
- [58] O. Dayıbaşı, "Doğal Dil İşleme (NLP)'nin Uygulamalarda Kullanım Alanları," *Medium*, 26-Mar-2016. .
- [59] K. Oflazer, "Türkçe ve Doğal Dil İşleme Turkish Natural Language Processing. Özet. Kemal Oflazer Carnegie Mellon Üniversitesi - Katar Doha, Katar - PDF," Carnegie Mellon Üniversitesi - Katar.
- [60] B. Liu, "Sentiment Analysis and Subjectivity," *Handb. Nat. Lang. Process.*, vol. 2, pp. 627–666, May 2010.
- [61] J. W. Pennebaker, M. R. Mehl, and K. G. Niederhoffer, "Psychological aspects of natural language. use: our words, our selves," *Annu. Rev. Psychol.*, vol. 54, pp. 547–577, 2003.
- [62] A. Garcia, S. Gaines, and M. T. Linaza, "A lexicon based sentiment analysis retrieval system for tourism domain," *Expert Syst Appl Int J*, vol. 39, 2012.
- [63] M. Ganapathibhotla and B. Liu, "Mining Opinions in Comparative Sentences," presented at the Proceedings of the 22nd International Conference on Computational Linguistics - Volume 1 (COLING '08), Vol. 1. Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 241–248.
- [64] P. D. Turney and S. Mohammad, "Crowdsourcing a Word–Emotion Association Lexicon," Institute for Information Technology, National Research Council Canada. Ottawa, Ontario, Canada, K1A 0R6, 2011.

EKLER

EK A1 Database Tabloları ve Detaylar

Database Detayları

Tablo Adı	Alan Adı	Alan Tipi	Açıklama
Main	Twit_Id	Bigint	Bu alana Tweetlerin Tweetlerde saklandığı gibi tekil (unique) ID'leri kaydedilir.
	Geo_Code	Integer	Bu alana kullanıcının ülke bilgisi kaydedilir.
	Creation_Time	Datetime	Bu alana Tweetlerin yaratılma zamanları kaydedilir.
UserData	Twit_Id	Bigint	Bu alana Tweetlerin Tweetlerde saklandığı gibi tekil (unique) ID'leri kaydedilir.
	User_Name	Varchar (50)	Bu alana Tweetlerin sahibinin kullanıcı adı kaydedilir.
	Follower_Count	Integer	Bu alana kullanıcının Twitterda ki takipçi sayısı kaydedilir.
	Following_Count	Integer	Bu alana kullanıcının Twitterda ki takip etmekte olduğu kişi sayısı kaydedilir.
Twits	Twit_Id	Bigint	Bu alana Tweetlerin Tweetlerde saklandığı gibi tekil (unique) ID'leri kaydedilir.
	Twit_Text	Varchar (200)	Bu alana paylaşılan Tweetin metni kaydedilir.
	Lang	Integer	Bu alana paylaşılan tweetin dili kaydedilir. Dil bilgisi sadece Türkçe olan Twitter üzerinden analiz yapılır. Değerler: 0 ve 1'dir. 0=TR, 1=EN
	RT_Count	Integer	Bu alana paylaşılan tweetin aldığı Retwit sayısı kaydedilir.
	Like_Count	Integer	Bu alana paylaşılan tweetin aldığı beğeni sayısı kaydedilir.
	Reply_Count	Integer	Bu alana paylaşılan tweetin aldığı Reply sayısı kaydedilir.

Result	Twit_Id	Bigint	Bu alana Tweetlerin Tweetlerde saklandığı gibi tekil (unique) ID'leri kaydedilir.
	Impression	Integer	Bu alana saveResult metodu çalıştırıldıktan sonra tweetle ilgili duygu kaydedilir. Değerler:0, 1, 2 0=Olumlu, 1=Olumsuz, 2=Nötr'dür.
	ResultValue	Integer	Tweet bazında cümlelerin içerisinde geçen kelimelerin toplam değeri bu alana kaydedilir.
	ContainingKeys	Varchar (200)	Tweet bazında cümlelerin içerisinde geçen ve Keywordlerle eşleşen kelimeler virgül (,) ile ayrılarak bu alana kaydedilir.
Keyword	Key_Id	Integer	Bu alana tekil ve artan (unique, incremental) ID değerleri kaydedilir.
	Impression	Integer	Duygu yönelim grupları bu alana kaydedilir. 0 = Nötr, 1 = Olumlu, 2 = Olumsuz olarak ifade edilir.
	Text	Varchar (50)	Farklı kaynaklardan ve gereksinimler doğrultusunda oluşturulan kütüphanede yer alan kelimeler bu alana Türkçe alfabe ile kaydedilir.
	KeyValue	Integer	Duygu yoğunluklarına göre Text'ler uygun olarak puanlandırılmıştır ve bu alana kaydedilmiştir. Puanlar, incelemeler ve araştırmalar sonrasında belirlenmiştir.
	KeyText	Varchar (50)	Farklı kaynaklardan ve gereksinimler doğrultusunda oluşturulan kütüphanede yer alan kelimeler bu alana İngilizce alfabe ile kaydedilir.

ÖZ GEÇMİŞ

Name & Surname: Semra Yıldırım

Permanent Address: 33 Shore Breeze Drive. Unit: 2305 M8V 0G1 Toronto, ON

Place and Year of Birth: Kuşadası, 1988

Foreign Languages: English, Russian

Secondary Education: Şehit Kayaaldoğan Anadolu Lisesi, 2005

Undergraduate: Applied Mathematics and Computer, Azerbaijan State Oil Academy, Azerbaijan, 2010

Postgraduate: Bahcesehir University, 2017

Name of Institute: Bahcesehir University, Istanbul.

Name of Master's Program: Engineering Management