T.C
HARRAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

TWİTTER'DE HABER DİFÜZYON ANALİZİ

Duygu SAĞALTICI

ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

ŞANLIURFA
2021

Dr. Öğr. Üyesi Nagehan İLHAN danışmanlığında Duygu SAĞALTICI'nın hazırladığı "**Twitterde haber difüzyon analizi**" konulu bu çalışma 15/01/2021 tarihinde aşağıdaki jüri tarafından oy birliği ile Harran Üniversitesi Fen Bilimleri Enstitüsü Elektrik Elektronik Mühendisliği Anabilim Dalı'nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

İmza

Danışman : Dr. Öğr. Üyesi Nagehan İLHAN ……

Üye : Dr. Öğr. Üyesi Serdar ÇİFTÇİ ……

Üye : Dr. Öğr. Üyesi Onan GÜREN ……

Bu Tezin Elektrik Elektronik Mühendisliği Anabilim Dalında Yapıldığını ve Enstitümüz Kurallarına Göre Düzenlendiğini Onaylarım.

Doç. Dr. İsmail HİLALİ
Enstitü Müdürü

Bu çalışma HÜBAK Tarafından Desteklenmiştir.
Proje No: 19286

Not: Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

İÇİNDEKİLER

Sayfa No

ÖZET i
ABSTRACT ii
TEŞEKKÜR iii
ŞEKİLLER DİZİNİ iv
ÇİZELGELER DİZİNİ v
SİMGELER ve KISALTMALAR DİZİNİ vi
1. GİRİŞ 1
2. ÖNCEKİ ÇALIŞMALAR 3
3. MATERYAL ve YÖNTEM 7
3.1. Twitter 7
3.2. Veri Madenciliği 7
3.2.1. Veri Madenciliği Yöntemleri 9
3.3. Metin Madenciliği 9
3.3.1. Metin önişleme 11
3.3.2. Veri Temizleme 11
3.3.3. Veri Dönüştürme 12
3.3.4. Veri Bütünleştirme 12
3.3.5. Veri İndirgeme 13
3.4. Normalizasyon 13
3.5. Korelasyon 13
3.6. Doğal Dil İşleme (Natural Language Processing-NLP) 14
3.6.1. Türkçe WordNet 15
3.6.2. Zemberek 15
3.7. FastText 18
3.7.1. N-Gram 19
3.8. Sosyal Ağ Analizi (Network Analysis) 19
3.8.1. Çizge Teorisi (Graph Theory) 20
3.9. Veri Seti 23
4. ARAŞTIRMA BULGULARI ve TARTIŞMA 31
4.1. Tweetleri Haber Kategorilerine Göre Sınıflandırma 31
4.2. Tweetlerin Kategorilerine Göre Hayatta Kalma Süreleri 33
4.3. Sosyal Ağ Analizi ile Haber Yayılımı 38
5. SONUÇLAR ve ÖNERİLER 40
5.1. Sonuçlar 40
5.2. Öneriler 41
KAYNAKLAR 42
ÖZGEÇMİŞ 45

ÖZET

Yüksek Lisans Tezi

TWİTTER'DE HABER DİFÜZYON ANALİZİ

Duygu SAĞALTICI

Harran Üniversitesi
Fen Bilimleri Enstitüsü
Elektrik Elektronik Mühendisliği Anabilim Dalı

Danışman: Dr. Öğr. Üyesi Nagehan İLHAN
Yıl: 2021, Sayfa: 46

Teknolojinin hızla gelişmesi bilginin daha kolay elde edilmesini ve daha hızlı yayılmasını sağlamıştır. Bu konudaki en etkili kaynaklardan biri de sosyal medyadır. Sosyal medyada bilgi hızlı olmakla birlikte çok farklı yerlerde bulunan farklı kitlelere kolaylıkla ulaşabilir. Özellikle son yıllarda haberlerin yayılım hızı ve kullanıcıların verdiği tepkiler sosyal medyanın etkisinin ne kadar güçlü olduğunu kanıtlar niteliktedir. Bu nedenle haberlerin yayılma hızı ve etkenlerini inceleyen birçok çalışma yapılmaktadır. Sosyal ağların bilgi yayılımındaki rolü, özellikle haber paylaşımında yaygındır. Sosyal medya kullanıcıları birbirlerini kolaylıkla harekete geçirebiliyor, tartışabiliyor ve etkileyebiliyor, bu da dinamik bilgi akışı ve yayılımı sağlayan zengin bir sosyal ortam yaratıyor. Geçmiş çalışmalar, bilginin yayılmasının çeşitli faktörlere bağlı olarak değiştiğini teyit etmiştir. Bu faktörlerin içinde etkilisi, ağ komşularının paylaşım şekillerine bağlı olarak bir kullanıcının bu bilgiye maruz kalmasıdır. Özellikle Twitter, birçok kullanıcı tarafından bir haber kaynağı olarak kullanılan birincil bir sosyal medya platformudur.

Bu çalışmada Twitter'da haber difüzyonu incelenmiştir. Bir haberin yayılıp yayılmayacağı, ne kadar sürede yayılıp etkin kaldığı, olumlu ve olumsuz haberlerin yayılma hızındaki farklılıklarla ilgili yapılan çalışmalar incelenmiş ve yöntemler karşılaştırılmıştır. Twitter'den toplanan verilerle veri madenciliği ve makine öğrenmesi yöntemleri kullanılarak haber difüzyon analizi yapılmıştır.

Çalışmada ilk aşama olarak haber başlık etiketleri çekilmiş ve twitter aracılığı ile paylaşılan haberlerin verileri alınmıştır. Alınan veriler bir veri tabanına aktarılmış ve yöntemler uygulanmadan önce ön işleme aşamasından geçmiştir. Çalışmada amaç güncel haberlerin kategorilerine göre nasıl tepki aldıklarını, paylaşım ve etki durumlarını incelemektir. Haber başlıkları alınmış, sınıflandırılmış ve twitter aracılığı ile yayılım durumları incelenmiştir. Böylece bir haberin uyandırdığı yankı incelenmiş ve toplumun hangi konularda nasıl tepkiler verdiği gözlemlenmiştir.

ANAHTAR KELİMELER: haber difüzyonu, twitter, veri madenciliği, web madenciliği, haber sınıflandırma

ABSTRACT

MSc Thesis

NEWS DIFFUSION ANALYSIS ON TWITTER

Duygu SAĞALTICI

Harran University
Graduate School of Natural and Applied Sciences
Department Electrical and Electronic Engineering

Supervisor: Assist. Prof. Dr. Nagehan İLHAN
Year: 2021, Page: 46

Nowadays it is much easier to access and share the information with the help of rapid development in technology. Social media is one of the most effective sources in this regard. Information flow on social media is very dynamic and can easily reach to different audiences in different locations. Especially in recent years, the speed of news dissemination and the reactions of users prove how strong the influence of social media is. That's why many studies have been conducted on not only how fast the information spread among the people but also the factors affecting the speed of information flow. The role of social network in information flow is very common in news sharing. Social media users can easily mobilize, discuss and influence each other, creating a rich social environment that enables dynamic flow and dissemination of information. Past studies show that the diffusion of information varies depending on several factors. The most effective among these factors is the exposure of a user to this information, depending on how their network neighbors share. In particular, Twitter has emerged as the social media platform for the news source.

This study examined the diffusion of news in Twitter. Whether a news will be spread or not, how long it has spread and remained popular, differences in the speed of positive and negative news spreading are examined and methods are compared. News diffusion analysis performed using data mining and machine learning techniques on the data collected from Twitter.

As a first step of the study, title tags and data of the news shared via Twitter are retrieved. The retrieved data stored to a database and pre-processing stage implemented before the methods to be applied. The aim of the study is to examine the reaction, sharing and impact of current news according to their categories. Headlines of the news are collected, classified and how it spread via Twitter is studied. Thus, how a news create reactions is examined and how the society reacted on which kind of subjects is observed.

KEY WORDS: news diffusion, twitter, data mining, web mining, news classification

TEŞEKKÜR

Tezin konusunun seçiminde, uygulamasında ve çalışmamda yardımlarını esirgemeyen danışmanım Dr. Öğr. Üyesi Nagehan İLHAN'a, Ayrıca tez çalışmam boyunca bana maddi ve manevi destek veren aileme, arkadaşlarım Yasin ASLAN ve Cihan EFİL'e teşekkür ederim.

ŞEKİLLER DİZİNİ

Sayfa No

Şekil 1.1 Sosyal Medya Kanalları 1
Şekil 3.1 Twitter Application Management(APPS) sayfası 7
Şekil 3.2 Veri Madenciliği Bilgi Keşif Süreci 8
Şekil 3.3 Veri madenciliği yöntemleri 9
Şekil 3.4 Metin madenciliği adımları 10
Şekil 3.5 Zemberek ara yüzü 16
Şekil 3.6 Zemberek NLP mimarisi 17
Şekil 3.7 Kelime çantası ve skip gram yöntemleri 18
Şekil 3.8 Düğümler (node) ve kenarlar (edge) 21
Şekil 3.9 Kullanıcılar arasındaki ilişkinin sosyal ağlarla gösterimi 22
Şekil 3.10 Haber başlıkları ve linkleri 23
Şekil 3.11 Twitter ağından çekilen veriler 24
Şekil 3.12 Kullanılan önemli öznitelikler 25
Şekil 3.13 Çeşitli karakter ve simgelerden temizlenmiş tweetler 26
Şekil 3.14 Zemberek ile kelimelerin köklerine ayrılması 27
Şekil 3.15 Tweetlerde geçen kelimeler ve frekansları grafiği 28
Şekil 3.16 Sık geçen kelimeler (wordcloud) 29
Şekil 4.1 Sınıflandırma akış şeması 31
Şekil 4.2 Sınıflara Göre Aylık Hayatta Kalma Süresi 33
Şekil 4.3 MySQL ile oluşturulan tablolar 34
Şekil 4.4 FavoriteCount ve LifeSpan arasındaki bağıntı 36
Şekil 4.5 retweetCount ve LifeSpan arasındaki bağıntı 36
Şekil 4.6 retweetCount ve favoriteCount arasındaki bağıntı 37
Şekil 4.7 Alınan tweet sayısını içeren Twitter ağ analizi 38
Şekil 4.8 Dinamik ağ analiz grafiği 39

ÇİZELGELER DİZİNİ

Sayfa No

Çizelge 3.1 Sınıf (kategori) bilgileri 25
Çizelge 3.2 Sınıflara göre tweetlerin ortalama hayatta kalma süreleri (saat) 26
Çizelge 3.3 Silinen Gereksiz Karakterler 27
Çizelge 3.4 Sınıflara göre en sık geçen kelimeler 29
Çizelge 3.5 Haber başlığı tweetleri ve kategorileri 30
Çizelge 4.1 Sınıflandırma sonuçları 32
Çizelge 4.2 Sınıflara göre ortalama başarı oranı 32
Çizelge 4.3 Haber başlıkları ve hayatta kalma süreleri 34
Çizelge 4.4 Önemli Öznitelikler 35

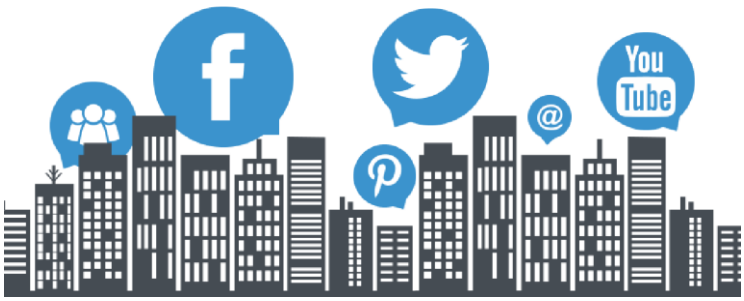
SİMGELER ve KISALTMALAR DİZİNİ

API	Application Programming Interface (Uygulama Programlama Arayüzü)
APPS	Twitter Application Management (Twitter Uygulama Yönetimi)
CBOW	Continuous Bag of Words (Kelime Çantası)
E	Edges (Kenarlar)
MySQL	My Structured Query Language (Yapısal Sorgu Dili)
NLTK	Natural Language Toolkit (Doğal Dil İşleme Kütüphanesi)
URL	Uniform Resource Loader (Tekdüzen Kaynak Bulucu)
V	Nodes (Düğümler)

1. GİRİŞ

Teknolojinin gelişmesiyle birlikte paylaşılan bilgiler artmış ve bilgi yayılımı çok hızlı bir hale gelmiştir. Bilgi yayılımı, son yıllarda birçok araştırmacının dikkatini çekmiştir. Birçok çalışma parçası, sosyal ağda bilgi yayılımının tahminini incelemek için çaba sarf etmiştir. Özellikle çevrim içi sosyal ağlar bu konuda zaman geçtikçe popüler olmaktadır (Hoang, 2018). Gerek kitle iletişim araçları ve gerekse internet, bilginin çok kısa bir zaman dilimi içerisinde, çok az maliyetle ve fazla çaba gerektirmeksizin, farklı yerlerde bulunan, çok farklı niteliklere sahip büyük kitlelere ulaşmasını sağlamaktadır. Tek yönlülük ve denetim eksikliği gibi dezavantajlarına rağmen sahip oldukları bu üstünlükler nedeniyle de bilgi toplumunun vazgeçilmezleri arasında yer almaktadır (Fang, 2017). Elbette yayılan bilginin doğru olması önem arz etmektedir ve buna gündemde yer alan haberler de dâhildir. Haber yayılımının analiz edildiği birçok çalışma yapılmıştır (Dilrukshi, 2013; Bhattacharya, 2012).

Teknolojinin gelişmesi ve bilgisayarların evlere girmesiyle iletişim araçlarının çeşitlenmeye başladığı günümüzde sosyal medya kullanımı, mobil teknolojinin ilerlemesi ve akıllı telefonların yaygınlaşmaya başlamasıyla her an gelişerek ve dönüşerek devam etmektedir. Öyle ki şuan dünya nüfusunun yaklaşık yarısına işaret eden 3.780 milyar kişi İnternet kullanırken ve bu kullanıcıların yaklaşık 2.850 milyarı da sosyal medyayı aktif olarak kullanmaktadır (Anonim, 2013).



Şekil 1.1 Sosyal Medya Kanalları (Anonim, 2013)

Şekil 1.1'de görüldüğü üzere bu kullanıcı kitlesi Facebook, Youtube, Instagram, Twitter, Google+, Linkedin ve daha birçok farklı sosyal mecraları kullanarak küresel çapta bir etkileşim ağı oluşturmaktadır. Böyle bir durumun sonucu olarak bu mecralar sadece kişilerin kullandıkları platformlar değil markaların, işletmelerin ve kurumlarında bulunmak zorunda olduğu yeni nesil bir yaşam ağı halini almıştır. Bu da milyonlarca kişiye erişmenin en kolay ve en ucuz yolu olan sosyal medyayı günümüzün en büyük bilgi paylaşım platformu ve reklam aracı yapmaktadır. Sosyal medyayı kısaca özetlemek gerekirse yeni nesil web teknolojilerinin getirdiği kullanıcı kolaylığı ve iletişim hızıyla yakalanan eş zamanlı bilgi paylaşımının takip edildiği dijital platformdur. Sosyal Medya içeriklerden oluşur, her bilgi bir içerik, her içerik bir kaynaktır. Sosyal Medya'da bilgi paylaşımı hiçbir maliyet gerektirmeden gerçekleşir ve anında çok geniş kitlelere ulaşır (Arklan, 2008).

2018 yılının İnternet, sosyal medya ve mobil kullanıcı istatistikleri

- 4.02 milyar internet kullanıcısı, dünya nüfusunun %53'ü
- 3.19 milyar sosyal medya kullanıcısı, dünya nüfusunun %42' si
- 5.13 milyar mobil kullanıcısı, dünya nüfusunun %68' i

2.95 milyar mobil sosyal medya kullanıcısı, dünya nüfusunun %39'unu ifade etmektedir.

Sosyal medya oldukça geniş bir bilgi paylaşım platformudur. Araştırmacılar bu platformdaki bilgi yayılımını incelemek adına Facebook, Twitter gibi oldukça popüler olan sosyal medya paylaşım sitelerinden yararlanmıştırlar (Arklan, 2008).

2. ÖNCEKİ ÇALIŞMALAR

Bilgi yayılımı uzun zamandır çalışılan bir konu ve hala çalışmaları devam etmektedir. Bilgi yayılımını anlamak ve daha iyi kontrol etmek için iletişim ortamını büyütmek çok önemlidir (Hoang, 2018). Gerçekten de, bazı araştırmalar sosyal medyanın yapılan seçimlerde hem ABD'de hem de Fransa'daki etkisini, çoğunlukla sahte haberlere ve bunların sosyal medyada yayılmasına odaklanmıştır. Seçmenlerin yüksek bir yüzdesi sosyal medyayı (18-29 yaşlarındaki insanların %35'i, Pew Research Center1'e göre) kullandığı için, paylaşımın asıl sayısı sahte hikâyeleri seçmenlere başarıyla ulaştıracaktır (Allcott, 2017). Bu, sosyal medya mesajlarının yayılmasını anlama ve tahmin etmenin önemini göstermektedir (Hoang, 2018).

Suh ve ark. (2010) belirli bir tweetin retweetlerinin sayısı ile ilişkili olabilecek bazı özellikleri tanımlamışlardır. 74 milyon tweet üzerinde geniş çaplı bir analiz düşünerek korelasyonu değerlendirdiler. Hesaptaki takipçilerin (followers), takiplerin (followees) ve yaşların sayısının retweet sayısıyla çok güçlü bir ilişkiye sahip olduğunu göstermişlerdir. Gönderenin takipçileri ve takipçilerinin sayısı ne kadar büyük olursa, tweetlerinin retweetlenmiş olması o kadar muhtemeldir. Ayrıca, yazmadan 300 gün önce kayıtlı olan "kıdemli kullanıcılar" tarafından yayınlanan tweet'ler, ortalamadan daha fazla sayıda retweet alır. Bir tweette hashtag veya URL varlığı, retweetlerin sayısıyla yüksek oranda ilişkili değildir. Suh ve diğ. retweetlerin %20.8'inin yalnızca hashtagler içerdiğini, retweetlerin %28.4'ünün URL içerdiğini bildirdiler. Ayrıca, geçmiş tweet'lerin sayısının, günlük tweet'lerin ortalama sayısıyla veya retweet oranıyla çok az veya hiç ilişkisi olmadığını da keşfettiler. Çalışmaları sonucunda kullanıcıların tercih ettiği tweetlerin sayısı retweetleri etkilemiyor gibi göründüğünü, çünkü retweetlerin sadece %8.7'si 100'den fazla favorili ürüne sahip yazarlar tarafından yazıldığını saptamışlardır.

Kwa ve ark. (2010) bir kullanıcının takipçilerinin sayısı ve tweetlerinin popülerliğiyle ilgili olarak 106 milyon tweet koleksiyonu arasındaki ilişkiyi incelemişlerdir. Yazarlar retweet ağaçlarını inşa ettiler ve ağaç zamansal ve mekansal

özelliklerini incelemiş, insanların sadece az sayıda insandan retweet aldıklarını ve sadece bir takipçinin bir alt kümesinin aslında retweet ettiğini göstermişlerdir. Ayrıca, 1000'den az takipçiye sahip kullanıcıların, yayınladıklarının ortalama aynı retweet sayısına sahip olma eğiliminde olduğu görülmüştür. Benzer şekilde, Remy ve ark. (2013) kullanıcıların takipçi sayısının mesajlarını yayma kapasitesi üzerindeki etkisini incelemiştir. İlginç bir şekilde, kullanıcıların çok sayıda takipçinin etkisinin, birkaç takipçisi olan kullanıcılardan istatistiksel olarak daha fazla olmadığını göstermişlerdir.

Thi Bich Ngoc Hoang ve Josiane Mothe (2018) yaptıkları çalışmalarda Twitter'de bilgi yayılımını incelemiş ve bir gönderinin yayılıp yayılmayacağını, ne kadar yayılacağını öngörmeyi amaçlamışlardır. Modellerini kullanıcı tabanlı (user-based), zaman tabanlı (time-based) ve içerik tabanlı (content-based) olmak üzere 3 tür özelliğe dayandırmışlardır. Kullandıkları tweet sayısı ise 16 milyona karşılık gelmektedir. Thi Bich Ngoc Hoang ve Josiane Mothe'un kullandıkları model makine öğrenmesine dayalıdır. Gelecekteki retweetlerin sayısı ile ölçülen mesajların popularitesini tahmin etmek için makine öğrenme tekniklerini kullanmışlardır. Yaptıkları çalışmada bir tweetin retweet alıp almayacağını tahmin ederken, modellerinin teknolojinin durumuna göre yaklaşık %5 F-ölçüsünü geliştirdiğini gösterip, ne kadar popüler olacağını tahmin etmişler. Bir kullanıcıya ait olan takipçi sayısı ve topluluk sayısı gibi bazı özelliklerin bir tweetin retweet alıp almayacağını öngörmek için çok önemli olduğunu da göstermektedirler. Bu çalışmalarında öncelikle olası propagasyonları tahmin etmek için tweetleri temsil edecek yeni özellikleri tanımlamış; sonra her iki özelliğin de üzerine kurdukları modeli üç koleksiyonda belirlemiş ve tanımladıkları kendi özelliklerinin tahmin aşamasında ne kadar yararlı olduğunu göstermişlerdir.

Inoshika Dilrukshi ve ark. (2013) ise haberleri farklı gruplara ayırmak ve böylece belirli bir ülkede belirli bir ülkede en popüler haber grubunu tanımlayabilmek için çalışmalar yapmışlardır. Twitter API kullanarak elde ettikleri veriler 12 gruba ayrılır. Bunlar; savaş-terörist-suç, ekonomi işleri, sağlık, spor, kalkınma-hükümet, siyaset, kaza, eğlence, afet-iklim, eğitim, toplum ve uluslararası. Bu verileri makine öğrenmesi yöntemlerinden biri olan destek vektör makinelerini kullanarak

sınıflandırmaktadırlar. Sınıflandırma için öncelikle her tweetteki kelimeyi bir özellik olarak ele almış ve örneklerin oluşturulması için bag-of-word yaklaşımı kullanılarak bir özellik vektörü oluşturulmuşlardır.

Fatma B. ve İlhan A. (2017) Yaptıkları çalışma ile dört farklı kategoriye ait (ekonomi, politika, spor ve sağlık) haber metinlerinin metin madenciliği teknikleri kullanılarak farklı öznitelik ve terim ağırlıklandırma yöntemleriyle sınıflandırılmasını, farklı yöntemlerin verimliliğinin ve başarısının test edilmesini amaçlamışlardır. Önerilen yöntem ile haber metinleri üzerinde etkin bir sınıflandırma başarımı yakalamışlar. Haber sitelerinden alınan, ekonomi, sağlık, politika ve spor başlıklı 4 kategori ve toplam 80 haber metnine ait sınıflandırma testleri sonucunda en iyi sonuçlara Kelime Kökü öznitelik yöntemi, Terim Frekansı terim ağırlıklandırma yöntemi ve Random Forest algoritması kullanılarak, %100'lük başarı oranıyla ulaşıldığı gözlemlenmiştir.

Cheong ve Lee (2009) Twitter'da popüler trend konuları ve ilgili kullanıcı demografilerini inceledi. Wu. ve ark. (2011), Twitter kullanıcılarının yüzde kaçının bilgi üreteci ve bilgi tüketicisi olduğunu inceleyerek bir Twitter kullanıcı analizi yaptı. Letierce ve ark. (2010) konferansın karma etiketini kullanarak katılımcıların bilimsel konferanslarda paylaştığı tweetleri gözlemleyerek Twitter'ın etkisine baktılar.

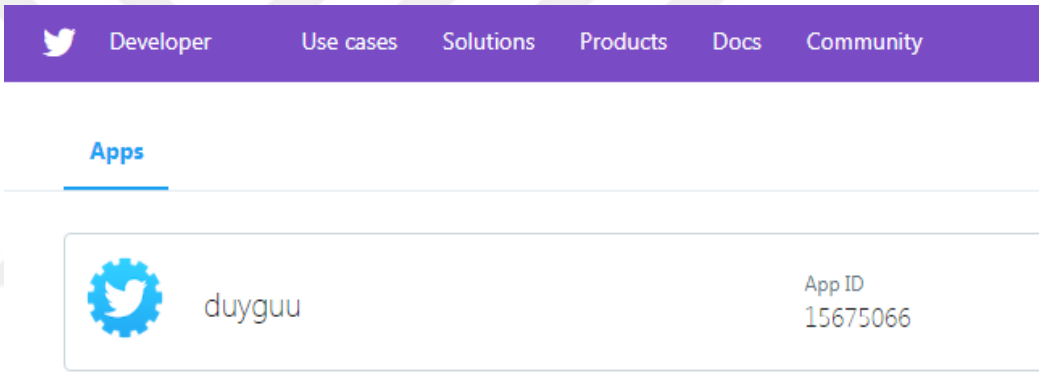
Anna Fang ve Zina Ben-Miled (2017) kendi çalışmalarında 'kötü haber daha hızlı yayılır mı?' sorusu temel alınarak, kullanıcı perspektifinin tartışmalı bir haber hikâyesi paylaşması üzerindeki etkisini incelemişlerdir. Twitter'da seçilen tweetleri takip ederek, kullanıcı paylaşma davranış modellerini oluşturmak için zaman içinde paylaşıldıkça haberlerin yayılmasını tanımlamışlardır. Birçok haber etkinliği olumlu veya olumsuz olarak görülebilir. Bu olaylarla ilgili tweet'leri genel kullanıcılar arasında karşılaştırmışlar. Aynı zamanda haber olaylarının kullanıcı ilgisine göre karşılaştırılabilir olmasını sağlamak için her olayın tweet frekansını izlemişler. Ayrıca, iki farklı kullanıcı grubu arasındaki tartışmalı bir olayın tweet'lerini izlemişlerdir (örneğin, etkinliği pozitif olarak gören ve bunu negatif olarak görenler).

Devipsita Bhattacharya ve Sudha Ram (2012) ağ ve zamansal analiz yöntemlerini birleştirerek, haber makalelerinin Twitter'da nasıl yayıldığını ve farklı yayılma mekanizmalarının haber makaleleri için farklı yaşam döngüleri ile nasıl sonuçlandığını incelemişlerdir. Farklı haber ajanslarının haber makaleleri hakkında farkındalık yaratma kapasitelerini ve Twitter kullanıcılarının onlar tarafından yayınlanan haber öğelerine ne kadar açık olduğunu anlamaya çalışmışlardır. Twitter'de çeşitli haber ajanslarının paylaştığı haber makalelerini izleyip ve ağ ve zamansal analizleri kullanarak tweet ve difüzyon modellerini incelemişlerdir.

3. MATERYAL ve YÖNTEM

3.1. Twitter

Twitter 335 milyon kullanıcıyla günde milyarlarca veriye ulaşabilen bir sosyal medya platformudur. Bu verilerin bir kısmının kullanılmasına izin veren bir uygulamaya sahip olması akademik çalışma alanında sıklıkla tercih edilmesini sağlamaktadır. Twitter Application Management (APPS) sitesinden bir twitter hesabı olması şartıyla API key alınabilmektedir. Bu anahtar şifre kullanılarak günde Twitter'in izin verdiği ölçüde kullanıcı verileri çekilebilmektedir.



Şekil 3.1 Twitter Application Management(APPS) sayfası

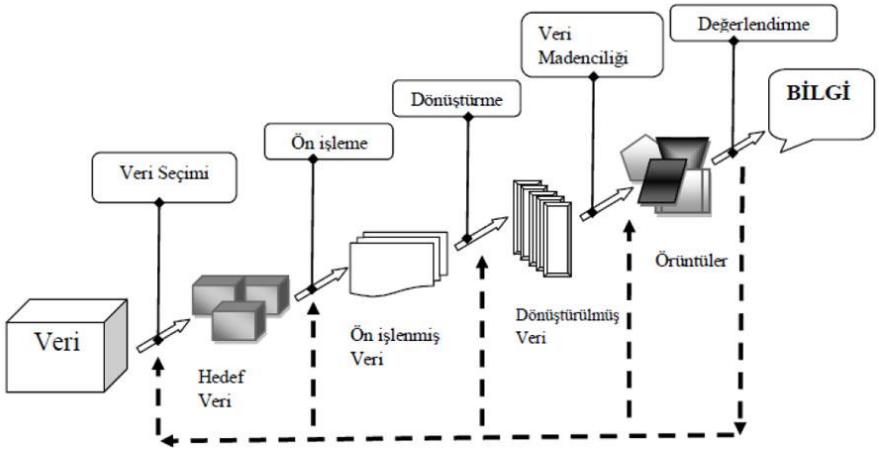
Şekil 3.1'de görüldüğü üzere Twitter'den veri çekmek için bir APPS oluşturulması gerekmektedir. APPS twitter tarafından onaylandıktan sonra API key, API secret key, Access Token ve Access token secret bilgileri hesaba tanımlanmaktadır. Java, R, Python, vb. birçok dil kullanılarak Twitter'den veri alınabilmektedir. Bu veriler web madenciliği, veri madenciliği, makine öğrenmesi yöntemleri ile analiz edilebilmektedir.

3.2. Veri Madenciliği

Veri madenciliği günümüzde teknolojinin hayatın her alanında yer almasıyla birlikte daha çok kullanılmaya başlayan en temel makine öğrenmesi yöntemlerinden

biri olarak karşımıza çıkmaktadır. Elde edilen milyarlarca bilgi yığını arasından anlamlı bilgilerin elde edilmesini sağlayan çözüm yollarından biridir. Veri madenciliğinde asıl amaç; bilgi yığınları içerisindeki gizli örüntü ve eğilimleri ortaya çıkarmaktır. Oldukça büyük kapasitedeki verilerden istenilen bilgiye ulaşmak uzun zaman almaktadır. Bu da veriye hızlı ulaşma yollarına sahip olan veri madenciliğinin önemini arttırmaktadır.

Veri madenciliğinde kullanılan verilerin, özellikle internet dosyalarında bulunanların, %80'inden fazlası yapısal olmayan verilerden oluşmaktadır (Shilakes, 1998). Anlamsız bilgi yığınlarından anlamlı bilgilerin çıkarılmasını sağlayan veri madenciliğinde en önemli nokta verilerin işlenmesi aşamasıdır. Verilerin analiz edilebilmesi için bazı önişlemlerden geçmesi gerekmektedir.

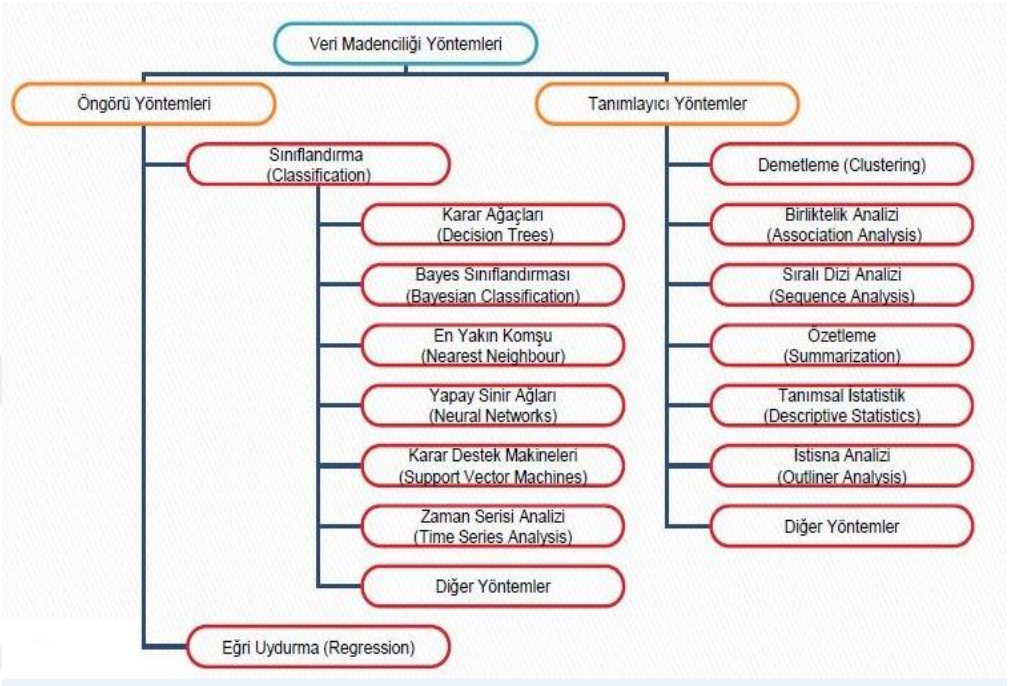


Şekil 3.2 Veri Madenciliği Bilgi Keşif Süreci (Savaş, 2012)

Şekil 3.2'de veri madenciliği yöntemleri kullanılarak asıl bilgiye ulaşma aşamaları yer almaktadır. Verilerin analize hazır hale gelecek şekilde işlenmesi ilgiye ulaşma sürecinin %80'lik kısmını oluşturmaktadır (Piramuthu, 2004). Verinin önişleme basamakları veri temizleme, veri bütünleştirme, veri indirgeme, veri dönüştürme ve veri ayrıklaştırma olarak beş maddeye ayrılabilir.

3.2.1. Veri Madenciliği Yöntemleri

Veri madenciliği yöntemleri öngörü ve tanımlayıcı yöntemler olmak üzere ikiye ayrılmaktadır. Şekil 3.3'te veri madenciliği yöntemleri yer almaktadır.



Şekil 3.3 Veri madenciliği yöntemleri (Akyokuş, 2018)

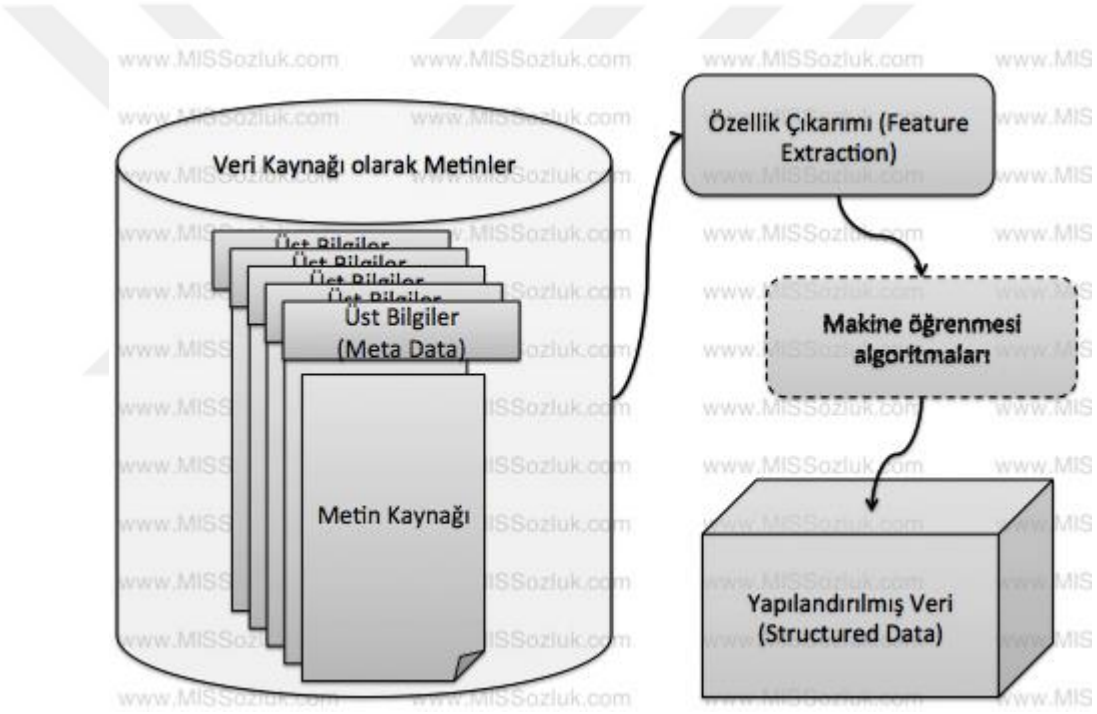
Veri madenciliğinde en sık kullanılan yöntemler sınıflandırma ve kümelemedir. Sınıflandırma yöntemlerindeki amaç gizli örüntüyü açığa çıkarmak ve yeni veriler ile ilgili öngörüye sahip olabilmektir. Verilerin sınıflandırılması ve belirli bir kural oluşturmak amacıyla gerekli ön işlemlerden geçen veri seti eğitim ve test olmak üzere ikiye ayrılır. Eğitim veri seti ile oluşturulan model test veri seti ile denenerek öngörü modelinin doğruluk oranı gözlemlenir. Ortaya çıkan sonuç ile birlikte karar verme sağlanmış olur (Özkan, 2016).

3.3. Metin Madenciliği

Son yıllarda özellikle de internetin yaygınlaşması ile birlikte yapısal olmayan verilerde artış gözlenmektedir. İnternet ortamlarında miyarlarca bilgi, doküman ve sosyal medya verileri yer almaktadır. Metin madenciliği de veri setini metin olarak

kabul eden bir veri madenciliği alanıdır. Çeşitli metinlerin sınıflandırılması, kümelere bölünmesi, metinlerde faydalı örüntülerin çıkarılması gibi birçok amaçla kullanılabilmektedir. Son yıllarda özellikle metinlerde duygu analizi ve sınıflandırma gibi çalışmalarda sıklıkla tercih edilen yöntemlerden biridir (Şeker, 2015).

Metin madenciliğinde amaç yapısal olmayan verilerin gerekli önişlemlerden geçirilerek veri tabanına atılacak duruma getirilmesidir (Karadağ, 2010). Veri tabanında yer alan veriler yapısal verilerdir. Yapısal veriler terim olarak işlem yapılmasını sağlayan ve veri madenciliği yöntemlerinin kullanılmasında kolaylık sağlayan verilerdir (Dolgun, 2009).



Şekil 3.4 Metin madenciliği adımları (Şeker, 2015)

Şekil 3.4'te görüldüğü üzere metin madenciliği adımları izlenerek metinlerden oluşan veri seti öncelikle işlenir, metinlerde bilgiye ulaşmayı sağlayan özellik çıkarımı yapılır, ardından çeşitli veri madenciliği, makine öğrenmesi, yöntemleri kullanılarak yapılandırılmış sonuca ulaşılır. Bu sonuçlara örnek olarak sınıflandırma, kümeleme ve tahmin modelleri gösterilebilmektedir.

Sosyal ağlardan elde edilen verilerin işlenmesi ve sonuç elde edilmesi aşamaları şu şekilde tanımlanabilmektedir:

- Verilerin toplanması ve seçilmesi
- Veri setindeki eksik ve hatalı verilerin temizlenmesi
- Gereksiz verilerin silinmesi (karakter, simge, noktalama işaretleri vb.)
- Veri setindeki metinlerin köklerine ayrılması ve nitelendirilmesi
- Özellik (nitelik) çıkarımı
- Veri ve metin madenciliği yöntemlerinin kullanılması
- Yapılandırılmış veri setinin tekrar kullanılabilecek duruma getirilmesi

Gerekli aşamalardan geçirilerek yapılandırılan veriler makine öğrenmesi yöntemleri kullanılarak anlamlı örüntü ve tahminler elde edilecek hale getirilmektedir (Weiss, 2010).

3.3.1. Metin önişleme

Çeşitli yöntemlerle elde edilen yapısal olmayan verilerin kullanılabilecek bir veri setine dönüştürülebilmesi için belirli ön işlemlerden geçirilmesi gerekmektedir. Bu ön işlemler veri madenciliği yöntemleri kullanıldığında hatalı sonuç oluşturabilecek gereksiz veriler, eksik veya hatalı bilgiler, bazı simge ve semboller, noktalama işaretleri ya da karakterleri içerebilmektedir. Önişlemlerden geçen veri seti ne kadar iyi oluşturulursa veri madenciliği yöntemlerinden elde edilen sonucun doğruluk payı o kadar iyi sonuç verecektir (Imberman, 2001).

3.3.2. Veri Temizleme

Veri tabanında yer alan gereksiz, eksik ya da hatalı verilerin veri setinden çıkarılması işlemidir. Veri madenciliği yöntemleri kullanılırken gereksiz yer kaplayan, kullanılması durumunda hatalı sonuçlara neden olabilecek verilere gürültü denir. Verilerin temizlenmesi işlem sırasında hatalarla karşılaşmamak açısından önem arz etmektedir. Örnek verilmesi gerekirse aynı anlamı ifade eden özellikler veri setinden

çıkarılabilir. Eksik ya da hatalı olan veriler yerine çeşitli yöntemlerle (missing value, karar ağaçları, regresyon, vb.) doğruluğu yüksek ortalama bilgiler eklenebilir ya da bu veriler silinebilir.

3.3.3. Veri Dönüştürme

Çeşitli yöntemlerle oluşturulan veri setlerinde uç değerlerde yer alan ve ayrık veri oluşumuna neden olan değerler bulunabilir. Özellikle sayısal veriler arasında bulunan ayrık değerler bulunan sonuçlarda sapmalara neden olabilmektedir. Sayılar arasındaki büyük farkların daha küçük bir aralığa dönüştürülmesi normalizasyon işlemleri olarak tanımlanır. En sık kullanılan veri dönüştürme yöntemleri:

- Min-Max
- Z-Score
- Ondalık Ölçekleme

Min-max dönüşümü verilerin daha küçük bir değer aralığında olacak şekilde dönüştürülmesi işlemidir. En çok tercih edilen aralık 0-1 aralığı olmaktadır. Burada veri setindeki en büyük değer 1'e eşitken en küçük değer ise 0'a eşit olmaktadır.

Z-score dönüşümünde ise verilerin ortalama ve standart sapma değeri kullanılarak z dönüşümü ile normalleştirilmesi işlemidir. Böylece veriler belirli bir aralık içinde sınırlandırılmaktadır.

Ondalık ölçekleme ise verilerin en büyük sayısına bağlı olarak ondalık kısımlarının değiştirilmesi (küçültülmesi) işlemidir (Oğuzlar, 2003).

3.3.4. Veri Bütünleştirme

Veri bütünleştirme, farklı veri tabanlarındaki verilerin birleştirilmesi işlemi olarak tanımlanabilmektedir. Bazı durumlarda ayrı veri tabanlarında bulunan veriler birleştirilirken aynı anlama gelen nitelikler farklı adlandırılmış olabilmektedir. Bu

durumda veri bütünleştirme işlemi sırasında hatalar oluşabilmektedir. Örnek verilecek olunursa, veri tabanında bulunan bir tabloda "erkek, kadın" gibi veriler tanımlanırken diğerinde "E,K" şeklinde tanımlanmış olabilmektedir. Farklı isimlerle nitelendirilmiş veriler aynı anlamı ifade etmektedir. Bu verilerin ortak ifade edilmesi işlemine veri bütünleştirme denilmektedir (Oğuzlar, 2003).

3.3.5. Veri İndirgeme

Veri madenciliği yöntemleri kullanırken bazı veri setlerinde gereksiz değişkenler, sonucu ulaşırken kullanılması bir değişiklik yaratmayan özellikler bulunabilmektedir. Bu tür değişkenler analiz işleminin uzamasına neden olabilmektedirler. Daha küçük boyutlu veri kümeleriyle sonuca daha hızlı ve sağlıklı bir şekilde ulaşmak mümkündür. Boyut indirgeme, veri sıkıştırma, veri birleştirme gibi veri indirgeme yöntemleri kısa sürede doğru sonuçlar elde etmek amacıyla kullanılmaktadır.

3.4. Normalizasyon

Bu yöntemde, bir grup verinin içerisindeki en büyük ve en küçük değerler ele alınır. Diğer bütün veriler, bu değerlere göre normalleştirilir. Buradaki amaç en küçük değeri 0 ve en büyük değeri 1 olacak şekilde normalleştirmek ve diğer bütün verileri bu 0-1 aralığına yaymaktır. Normalizasyon formülü (3.1) de yer almaktadır (Yavuz, 2012).

$$z_i = \frac{x_i - min(x)}{max(x) - min(x)} \tag{3.1}$$

3.5. Korelasyon

Korelasyon iki veya daha fazla değişken arasındaki ilişkinin gücü ve yönünü gösteren bir kavramdır. Korelasyonun negatif çıkması ters orantılı bir ilişkiyi gösterirken, pozitif olması doğru orantılı olan bir ilişkiyi gösterir. Korelasyon katsayısı -1 ile 1 değerleri arasında yer alır ve değişkenler arasında doğrusal bir ilişki olup

olmadığını belirlemeyi sağlar. Korelasyon (3.2) de yer alan formül ile hesaplanmaktadır (Alpar, 2003).

$$\rho_{x,y} = \frac{Cov(x,y)}{\sigma_x \sigma_y} = \frac{E((x-\mu_x)(y-\mu_y))}{\sigma_x \sigma_y} \quad (3.2)$$

3.6. Doğal Dil İşleme (Natural Language Processing-NLP)

Dil hayatımızın en önemli gereksinimlerinden biridir. Yaşantımızda oldukça yer kapsayan iletişim unsurları daha kolay ve konforlu bir hayat olanağı tanımaktadır. Bilgisayar bilimlerinin çokça kullanılmasıyla birlikte dil ve teknolojinin birleştirilmesi konusunda derin çalışmalar başlatılmıştır. Bu bağlamda doğal dil işleme bilgisayar yardımı ile dilin çözümlenmesini ve anlamlı hale gelmesini sağlama çalışması olarak tanımlanabilmektedir (Adalı, 2012).

Doğal dil günlük yaşantımızda iletişim kurmak için kullandığımız bir araçtır. Doğal dil işleme kullanmış olduğumuz dillerin makineler yardımı ile kurallı yapılarının çözümlenmesine ve yapılandırılmasına olanak sağlayan bir yapay zekâ kategorisidir. Doğal dil işleme çalışmaları yapılırken evrensel bir dil olma özelliğine sahip olan İngilizce dilinde çalışmalar yapılmıştır. Bu çalışmalar sırasında en çok dikkat edilen kategoriler dilin morfolojik yapısını oluşturan kelime dizimi, anlam, ses ve biçim bilimi olmuştur (Adalı, 2012).

Türkçe dili biçimsel olarak İngilizce ’den farklı yapı ve kurallar içerdiğinden Türkçe’ye özel doğal dil işleme algoritmaları oluşturulmuştur. Doğal dil işleme algoritmaları dile ait kurallar ve yapılara göre şekillendiğinden sabit değildir ve sürekli olarak geliştirilmektedir (Oflazer, 2016).

Doğal dil işleme kullanım alanlarına:

- Yazım denetimi
- Yazım hatalarının düzeltilmesi

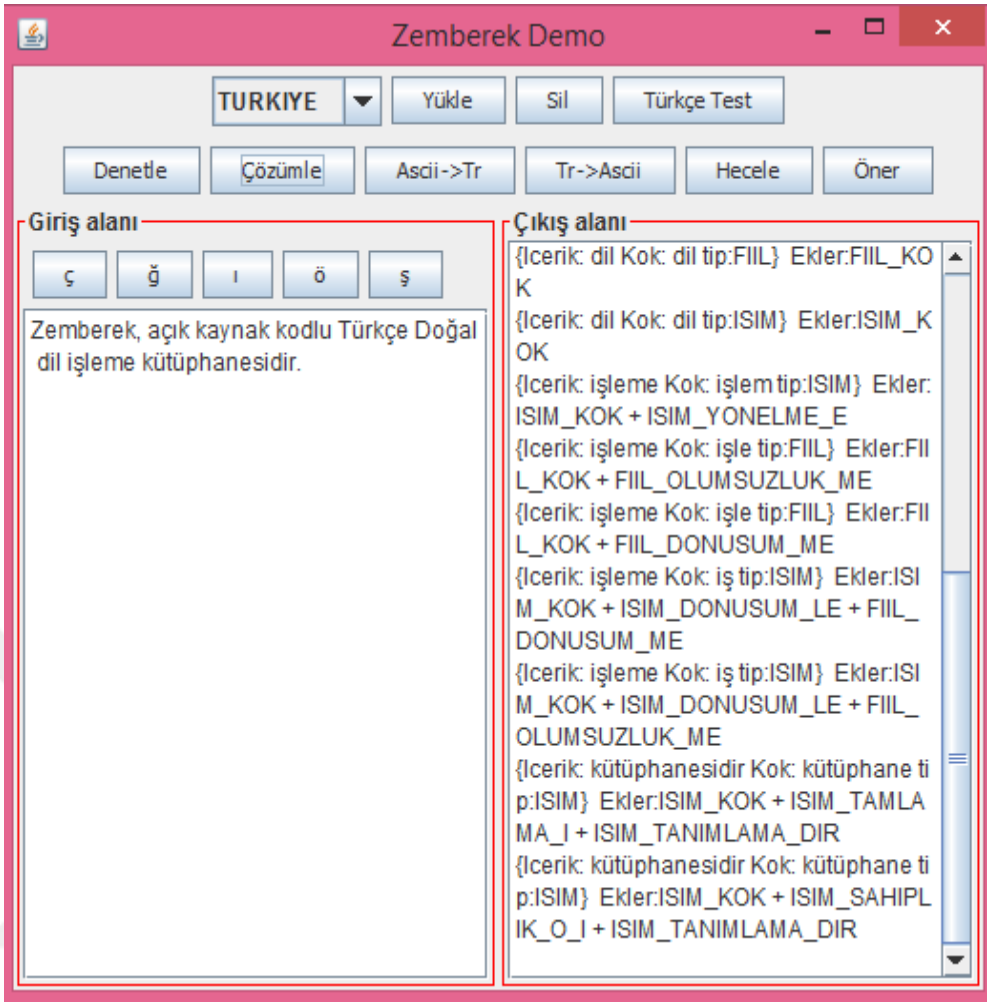
- Metin sınıflandırma
- Metin çevirisi
- Soru yanıtlama
- Ses ve karakter tanıma
- Konuşma bölümü etiketleme gibi örnekler verilebilir.

3.6.1. Türkçe WordNet

BalkaNet adındaki bir proje kapsamında geliştirilmiş Türkçe sözlüktür. Yaklaşık olarak 15 bin anlam içeren bir kümeden oluşmaktadır. Kelimelerin anlamları doğal dil işleme algoritmalarının temelini oluşturmaktadır. Dolayısıyla Türkçe metinlerde anlam analizi yapabilmek, bilgi çıkarımında bulunabilmek ve cümleleri ögelerine ayırabilmek oldukça önemlidir. Bu çalışmaların gerçekleştirilmesi amacıyla Türkçe biliminde uzmanlaşmış olan araştırmacılar bu projeyi geliştirmeye devam etmektedirler (Amasyalı, 2005).

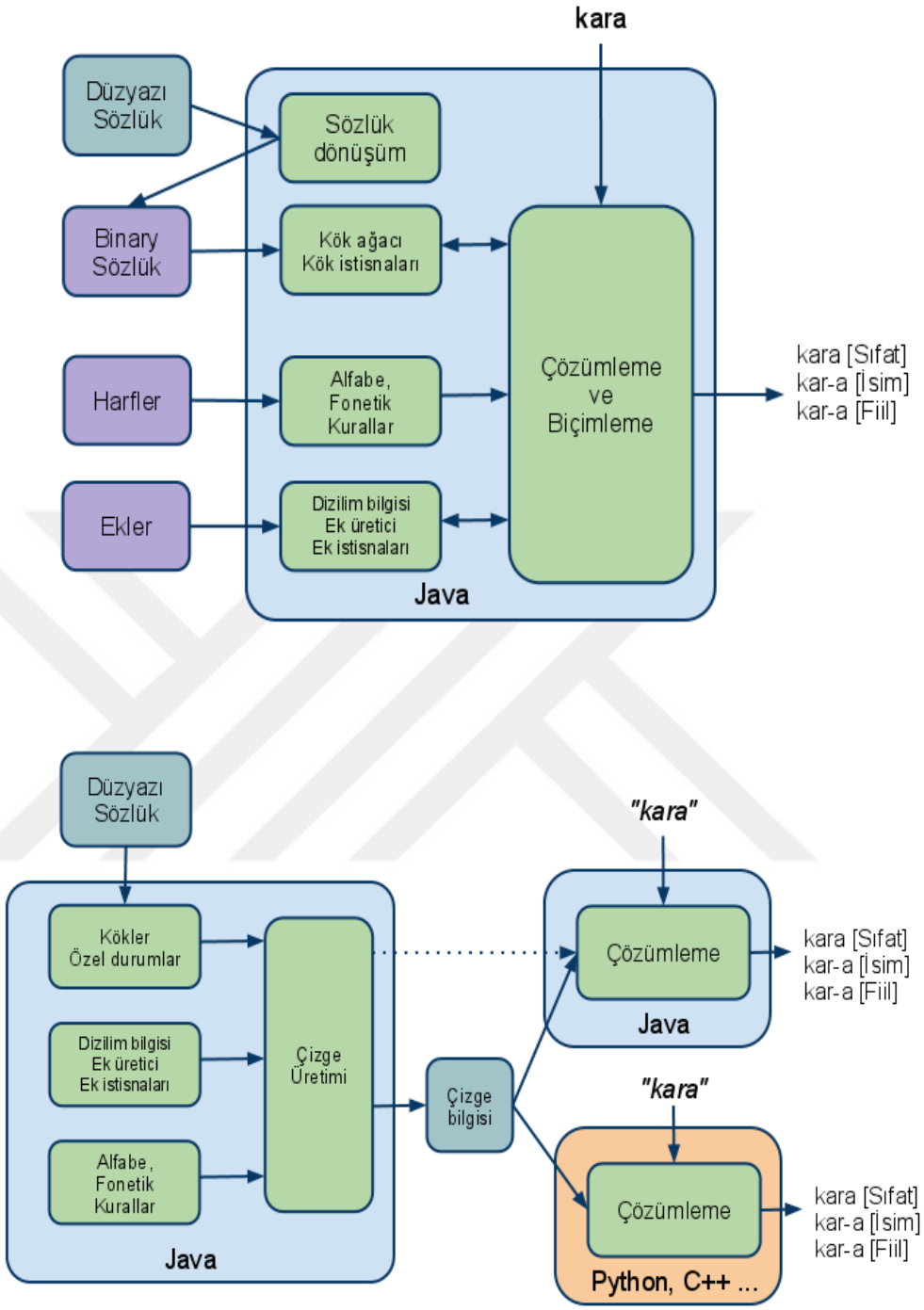
3.6.2. Zemberek

Doğal dil işleme alanında en önemli adımlardan biri kelime kökünün bulunmasıdır. İngilizce metinlerde kolaylıkla ulaşılan bu sonuç Türkçe gibi sondan eklemeli ve çok ek alan dillerde zorluk yaratmaktadır. Zemberek bu ihtiyacı karşılamak adına Ahmet A. Akın tarafından geliştirilmiş Java tabanlı bir doğal dil işleme kütüphanesidir. Açık kaynak kodlu bir kütüphane olup kelime köklerini bulmakla birlikte yazım hatalarını düzeltme, heceleme, kelimeleri etiketleme gibi özelliklere de sahiptir. Zemberek kütüphanesi bir kelimenin Türkçe olup olmadığını kontrol edebilme özelliğine sahiptir (Akın, 2007).



Şekil 3.5 Zemberek ara yüzü

Şekil 3.5'de görüldüğü üzere zemberek kütüphanesinin ara yüzünde Türkçe kelimelerden oluşan cümlelerin ek ve kök çözümlemelerini yapabilmektedir. Ek olarak Türkçe denetimi, heceleme, öneri gibi seçenekleri bulunmaktadır. Zemberek projesi günümüzde de geliştirilmeye devam etmektedir ve imla denetleme kütüphanesi olarak geliştirilmesine rağmen dilin morfolojik yapılarının eklenmesiyle birden fazla işleve sahip olmuştur.

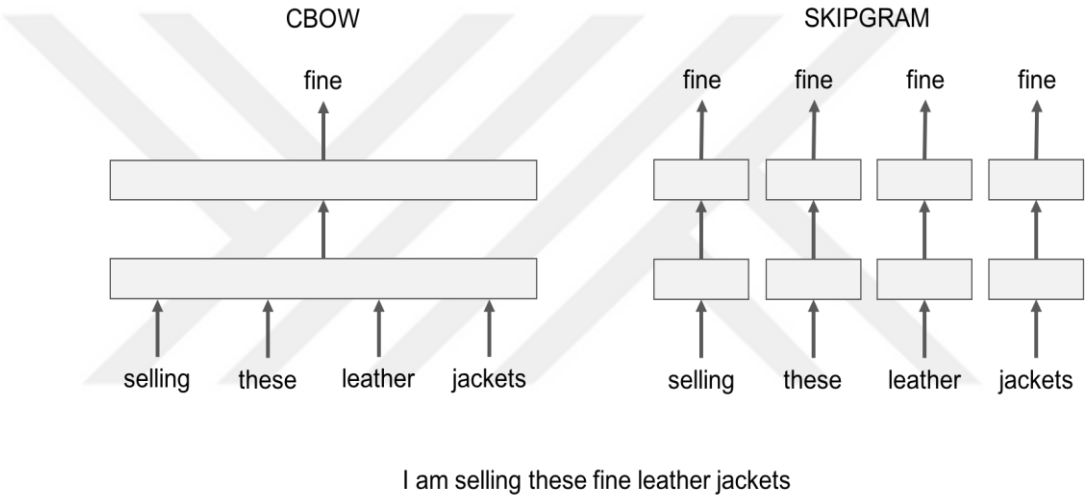


Şekil 3.6 Zemberek NLP mimarisi (Akın, 2011)

Şekil 3.6'da Zemberek kütüphanesinin çalışma prensibi ve mimarisi yer almaktadır. Zemberek 2 sözlüğü içerisinde kelimelerin her birinde türü ve özel durumları etiketlenip kodlanıyorken, Zemberek 3'te ise sözlük içerisinde çoğu bilgi otomatik olarak tespit edilebilmekte ve daha az etiketleme yapılmaktadır.

3.7. FastText

FastText Facebook Al Research (Joulin, 2016) ekibi tarafından geliştirilen ve farklı makine öğrenmesi yöntemleri kullanarak metin sınıflandırmasında kullanılan bir kütüphanedir. Facebook Açık Kaynak grubu tarafından geliştirilmiş olan bu kütüphane, metin ya da kelimeleri sürekli vektörlere dönüştürerek sınıflandırma kolaylığı sağlamaktadır. FastText Word2Vec yöntemi kullanabilme özelliğine sahiptir. Çalışma prensibinde Continuous Bag of Words (CBOW) ve Skip Gram kullanılmaktadır.



Şekil 3.7 Kelime çantası ve skip gram yöntemleri (Dede, 2018)

Şekil 3.7'de kelime çantası ve skip gram yöntemlerinin örneklemeleri görülmektedir. Kelime Çantası (Continuous Bag of Words) modelinde bütün kelimeleri ele alan bir bakış açısı ile vektör toplamı kullanılırken SKİPGRAM modelinde ise belirtilen kelimeye yakın sözcükler kullanılarak tahmin yapılır. Kelimeler arasındaki benzerlik açısal yöntemle hesaplanabilmektedir. Kosinüs benzerlik hesaplama (3.3) de görülmektedir (Joulin, 2016).

$$\cos\theta = \frac{A.B}{\|A\|_2\|B\|_2} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{3.3}$$

Metin sınıflandırma yöntemi olarak geliştirilen FastText N gram yöntemini kullanarak kelimelerin birbirleriyle olan anlam ilişkisini hesaplamaktadır. Modelde eğitim ve test aşamaları yer alarak kelime tahmin işlemleri yapılmaktadır.

3.7.1. N-Gram

N-gram dil modeli n-1 dereceden bir Markov Modeli sıralamasında bir sonraki elemanı tahmin eden istatistiksel bir dil modelidir (Akgül, 2016). N-gram analizi birçok çalışmada kullanılmış, uygulanması iyi anlaşılmıştır ve etkilidir. Bir dizi veriyi n-grama dönüştürerek, iki veya daha fazla veri akışını verimli bir şekilde karşılaştırmak için bir vektör boşluğuna gömülebilir. N gramlık metinler, metin madenciliği ve doğal dil işleme görevlerinde yaygın olarak kullanılmaktadır. Temel olarak belirli bir pencerede bir arada oluşan kelimeler kümesidir ve n-gramları hesaplarken tipik olarak bir kelime ileri taşınır (X kelimelerini daha gelişmiş senaryolarda ileri taşınabilir. Bir n-gram modeli, olasılığı bilinen kelime sırası temel alınarak sıradaki kelimenin ne olabileceğini tahmin etmeye çalışır. Doğal dil işleme yöntemlerinden biri olan n gram modeli belirlenmiş kelime sayısına bağlı olarak bir kelimeyi tahmin etmeye çalışır. N-gram modeli dilin yapısını öğrenme problemini basitliğe indirger.

3.8. Sosyal Ağ Analizi (Network Analysis)

Sosyal medyanın günlük hayatımızda büyük bir yer almasıyla birlikte sosyal ağ madenciliği adı altında ayrı bir araştırma alanı doğmuştur. Sosyal medya ile birlikte toplumsal algılar ve gündem sosyal ağ analizi ile belirlenebilmektedir (Demir, 2020). İnternete kolay erişimin gündelik hayatın bir parçası haline gelmesiyle sosyal medya ortamlarında veri paylaşımı hızla artmıştır. Böylece insanların içinde bulundukları sosyal ağlar genişlemiş ve insan ilişkileri ağlara yansımıştır. Sosyal medyadaki bilgi artışı geleneksel haber kaynaklarını ve gündemi etkilemeye başlamıştır (Sayre, 2010). Daha önce anket ve diğer kaynaklarla ulaşması zor olan veriler günümüzde sosyal medya platformlarından kolaylıkla elde edilebilmektedir. Veri madenciliğinde verilerden elde edilen örüntüler sosyal ağ madenciliğinde de ilişkileri görmemizi

sağlayan ağlardan elde edilebilmektedir (Knoke, 2019). Bu süreç sosyal medya platformlarından alınan yapısal olmayan verilerin işlenmesi ve kullanılabilir hale getirilmesiyle başlayıp işe yarar bilginin çıkarılmasına kadar devam etmektedir.

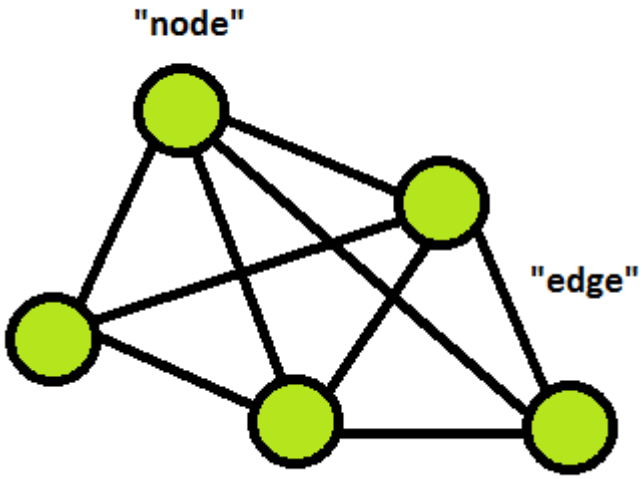
Sosyal ağ analizi ile gözle görülemeyen önemli bağıntılar elde edilebilmektedir. Ağdaki en etkili kişiler, gündemi etkileyen paylaşımlar, sosyal ağa katılım bilgileri, bilginin yayılımı ve davranış analizi gibi birçok bilgiye ulaşım imkânı yaratmaktadır. Sosyal ağları daha iyi anlayabilmek, gerekli hesaplamaları yapabilmek ve görselleştirmek için Çizge Teorisi (Graph theory) kullanılabilmektedir.

3.8.1. Çizge Teorisi (Graph Theory)

Sosyal ağlar genel olarak ağı meydana getiren kişilerden ve bu kişiler arasındaki ilişkilerden oluşmaktadır ve çizge kuramına göre kişiler arasında bir bağ var ya da yoktur (Wasserman, 1994). Çizge teorisi kişi ya da kullanıcıları düğüm (node), bu kişiler arasındaki bağlantıyı ise kenar (edge) olarak tanımlamaktadır. Çizge düğümler ve kenarlar kümesinden oluşmaktadır. Bu tanıma ait tanım (3.4) te gösterilmektedir.

$$G = (V, E) \tag{3.4}$$

(3.4) te yer alan tanımda V (nodes) düğümler kümesini belirtirken, E (edges) kenarlar kümesini belirtmektedir. Düğüm kullanıcıları, kenarlar ise bu kullanıcılar arasındaki ilişkiyi ifade etmektedir.



Şekil 3.8 Düğümler(node) ve kenarlar(edge)

Şekil 3.8'de 5 düğüm ve 9 kenarı temsil eden bir örnek görülmektedir. Çizge kuramına göre kenarlar yönlü (directed) ve yönsüz (undirected) olmak üzere ikiye ayrılmaktadır (Şeker, 2015). Twitter sosyal medya hesabında isteğe bağlı takibe dayalı olarak yönlü ve yönsüz graflar bulunabilmektedir. Örnek verilecek olunursa; Twitter hesabı ile kullanıcı takip ettiği kişiler tarafından takip edilmiyor olabilir. Bu durumda aradaki ilişki görüntülenirken kenar tek yönlü olacaktır. Karşılıklı takip durumunda ise yönsüz bir kenar ortaya çıkacaktır.

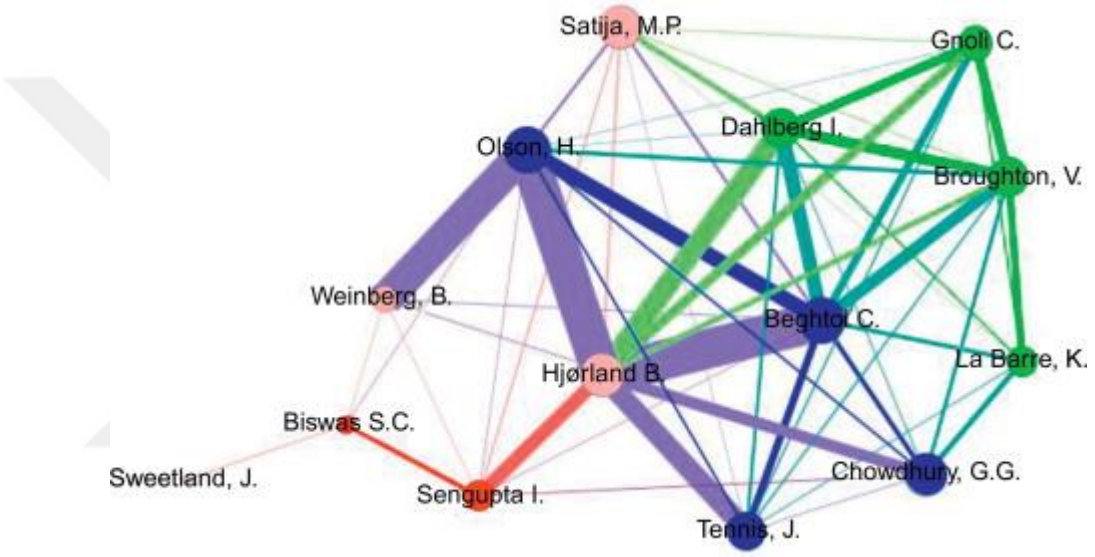
Sosyal ağ analizi yapabilmek için çizge teoreminden yararlanılarak çeşitli ölçütler geliştirilmiştir. Merkezilik (Centrality), Derece Merkeziliği (Degree) ve Özvektör Merkeziliği (Eigenvector Centrality) bu ölçütlere örnek olarak gösterilebilmektedir.

Merkezilik, en temel sosyal ağ analizi ölçütü olarak kabul edilebilir. Ağ içerisindeki düğüm ya da kullanıcının diğer düğümler arasında ne kadar merkezi olduğunu ifade etmektedir.

Derece Merkeziliği bir düğüm ya da kullanıcıya bağlanmış olan kenar sayısını ifade etmektedir (Knoke, 2019). Örneğin; Şekil 3.8'de bir düğümün 3, bir diğer

düğümün ise 4 kenara bağlı olduğu görülmektedir. Düğümlerin bağlı olduğu kenarlar onların derecesini ifade etmektedir.

Özvektör Merkeziliği ise derece merkeziliğinde geçerli olan kullanıcının kaç komşuya bağlı olduğunu gösteren derecesinin yanı sıra komşu olan düğümlerin dereceleri yani komşularını da ele alır. Bu duruma Twitter ile örnek gösterilecek olunursa bir kullanıcının yalnızca kendi arkadaşlarına değil aynı zamanda arkadaşlarının arkadaşlarına da bakılmaktadır.



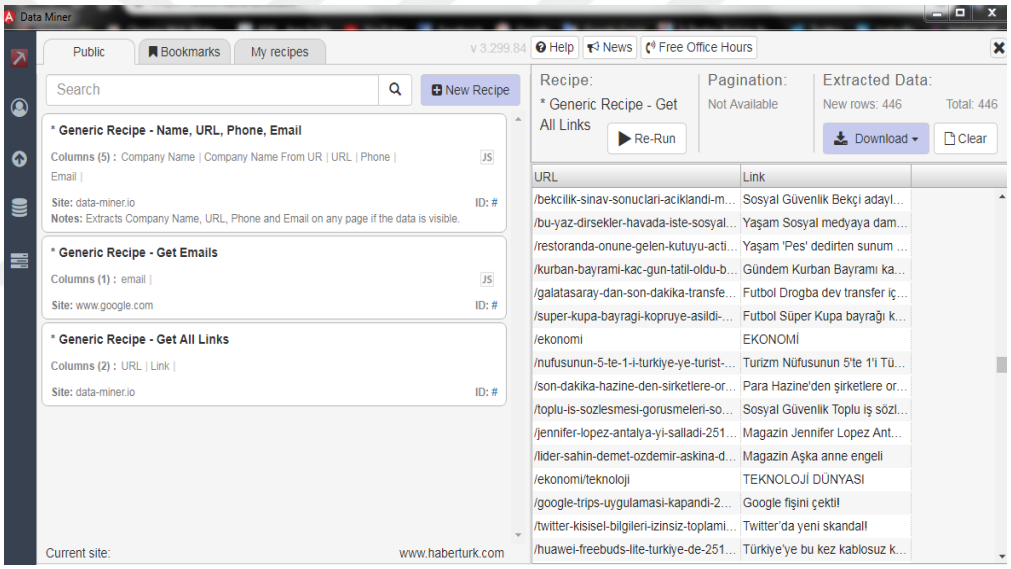
Şekil 3.9 Kullanıcılar arasındaki ilişkinin sosyal ağlarla gösterimi (Smiraglia, 2015)

Şekil 3.9'de görüldüğü üzere bir kullanıcının yalnız kendi arkadaşlarıyla kalmayıp arkadaşlarının arkadaşlarına da bakılması daha ayrıntılı bilgi akışı sağlayacaktır. Özellikle sosyal medya platformlarında bir kişinin arkadaşları ve tanıdık olarak nitelendirilebilecek arkadaşlarının arkadaşları sayısı kişinin popülerliğini belirtmektedir. Bu durumda popülaritesi yüksek olan kullanıcılar bilgi yayma ve hızlı bilgi akışı konusunda büyük öneme sahiptirler. Sosyal medya hesaplarında ayrıca bu bilgi tanıdık önerisi sunmak için de kullanılmaktadır.

3.9. Veri Seti

Haberlerin günlük yaşantımızda ne kadar yer aldığı ve önemli kararlarda etkisinin olup olmadığı incelenmek istenmektedir. Bilgi yayılımı konusunda literatürde pek çok çalışma yer almaktadır.

Bu çalışmada bir haber sitesinden çekilen tweetler üzerinde inceleme yapılarak hayatta kalma süreleri (lifeSpan), aldıkları beğeni sayısı ve retweet sayısına ulaşılmıştır. Bu bilgiler değerlendirilerek aralarındaki ilişkinin bulunması amaçlanmıştır. Şekil 3.10'da görüldüğü üzere haber sitelerinden haber başlıkları ve adresleri çekilmiştir.



Şekil 3.10 Haber başlıkları ve linkleri

Öncelikle twitter hesabı ile API key alınmıştır. Veriler Twitter sosyal paylaşım platformundan API key kullanılarak alınmıştır. Verilerin çekilmesi için R programlama dili kullanılmıştır. R studio programında bulunan TwitteR, ROAuth ve RCurl kütüphaneleri ile API key içeriğinde bulunan consumer_key, consumer_secret, access_token ve access_secret anahtarları kullanılarak her gün 100 tweet çekilmiştir. Çekilen haber başlığı tweetleri Şekil 3.11'deki gibi oluşturulmuştur.

no	text	favorited	favoriteCount	replyToSN	created	truncated	replyToSID	id	replyToUI	statusSou	screenNa	retweetCount	isRetweet	retweeted	longitude	latitude
1	Apartmandaki çat	FALSE	4	NA	03.05.2019 16:02	FALSE	NA	1.12E+18	NA	<a href="h	cnnturk	0	FALSE	FALSE	NA	NA
2	FETÖ kadın yapıla	FALSE	6	NA	03.05.2019 15:48	FALSE	NA	1.12E+18	NA	<a href="h	cnnturk	0	FALSE	FALSE	NA	NA
3	"Çocuğa cinsel ist	FALSE	9	NA	03.05.2019 15:35	TRUE	NA	1.12E+18	NA	<a href="h	cnnturk	1	FALSE	FALSE	NA	NA
4	'Sarallar'a operas	FALSE	3	NA	03.05.2019 15:11	FALSE	NA	1.12E+18	NA	<a href="h	cnnturk	0	FALSE	FALSE	NA	NA
5	#SONDAKİKA Yolc	FALSE	6	NA	03.05.2019 14:50	FALSE	NA	1.12E+18	NA	<a href="h	cnnturk	0	FALSE	FALSE	NA	NA
6	Damadı ve kardeş	FALSE	4	NA	03.05.2019 14:32	FALSE	NA	1.12E+18	NA	<a href="h	cnnturk	1	FALSE	FALSE	NA	NA
7	#CANLI Türkiye İn	FALSE	12	cnnturk	03.05.2019 14:21	TRUE	1.12E+18	1.12E+18	61168062	<a href="h	cnnturk	1	FALSE	FALSE	NA	NA
8	#CANLI Türkiye İn	FALSE	8	cnnturk	03.05.2019 14:19	TRUE	1.12E+18	1.12E+18	61168062	<a href="h	cnnturk	1	FALSE	FALSE	NA	NA
9	#CANLI Türkiye İn	FALSE	6	cnnturk	03.05.2019 14:11	TRUE	1.12E+18	1.12E+18	61168062	<a href="h	cnnturk	0	FALSE	FALSE	NA	NA
10	#CANLI Türkiye İn	FALSE	7	cnnturk	03.05.2019 14:04	FALSE	1.12E+18	1.12E+18	61168062	<a href="h	cnnturk	0	FALSE	FALSE	NA	NA
11	#CANLI Türkiye İn	FALSE	20	NA	03.05.2019 14:04	FALSE	NA	1.12E+18	NA	<a href="h	cnnturk	0	FALSE	FALSE	NA	NA
12	Kurumlar arası ge	FALSE	7	NA	03.05.2019 13:53	FALSE	NA	1.12E+18	NA	<a href="h	cnnturk	0	FALSE	FALSE	NA	NA
13	#SONDAKİKA Türl	FALSE	9	NA	03.05.2019 13:35	FALSE	NA	1.12E+18	NA	<a href="h	cnnturk	0	FALSE	FALSE	NA	NA
14	Muhtar olduğunu	FALSE	7	NA	03.05.2019 13:06	FALSE	NA	1.12E+18	NA	<a href="h	cnnturk	3	FALSE	FALSE	NA	NA
15	#CANLI Çamlıca Ci	FALSE	9	cnnturk	03.05.2019 13:02	TRUE	1.12E+18	1.12E+18	61168062	<a href="h	cnnturk	0	FALSE	FALSE	NA	NA
16	#CANLI Çamlıca Ci	FALSE	15	cnnturk	03.05.2019 12:57	TRUE	1.12E+18	1.12E+18	61168062	<a href="h	cnnturk	1	FALSE	FALSE	NA	NA
17	#SONDAKİKA Ayv	FALSE	2	NA	03.05.2019 12:52	FALSE	NA	1.12E+18	NA	<a href="h	cnnturk	0	FALSE	FALSE	NA	NA
18	#CANLI Çamlıca Ci	FALSE	8	cnnturk	03.05.2019 12:49	FALSE	1.12E+18	1.12E+18	61168062	<a href="h	cnnturk	0	FALSE	FALSE	NA	NA
19	#CANLI Çamlıca Ci	FALSE	13	cnnturk	03.05.2019 12:42	FALSE	1.12E+18	1.12E+18	61168062	<a href="h	cnnturk	0	FALSE	FALSE	NA	NA
20	#CANLI Çamlıca Ci	FALSE	47	NA	03.05.2019 12:39	FALSE	NA	1.12E+18	NA	<a href="h	cnnturk	7	FALSE	FALSE	NA	NA
21	AK Partili Özel: Ta	FALSE	12	NA	03.05.2019 12:38	FALSE	NA	1.12E+18	NA	<a href="h	cnnturk	0	FALSE	FALSE	NA	NA
22	63 bin kişinin ayn	FALSE	19	NA	03.05.2019 12:36	TRUE	NA	1.12E+18	NA	<a href="h	cnnturk	3	FALSE	FALSE	NA	NA

Şekil 3.11 Twitter ağından çekilen veriler

Veri setini oluşturmak için 6 ay boyunca bir haber kanalının paylaştığı 10,529 haber başlığı tweetleri çekilmiştir. Veriler no, text, favorited, favoriteCount, replyToSN, created, truncated, replyToSID, id, replyToUID, statusSource, screenName, retweetCount, isRetweet, retweeted, longitude, latitude ve LifeSpan olmak üzere 18 öznitelikten oluşmaktadır. Bu öznitelikler arasındaki favorited, replyToSN, truncated, replyToSID, replyToUID, statusSource, screenName, isRetweet, retweeted, longitude, latitude parametreleri aynı değer olmakla bilgi içermediğinden analiz sırasında kullanılmamıştır. Kullanılan önemli öznitelikler:

id: tweet numaraları

text: tweet haber metni

favoriteCount: tweet beğeni sayısı

created: tweetlerin oluşturulma tarihi

retweetCount: retweet sayısı

LifeSpan: tweetlerin hayatta kalma süreleri(saat, gün)

id	text	created	favoriteCount	retweetCount	LifeSpan
1,10E+23	Cumhurbaşkanı Erdoğan: Venezuela altını Çorum'da işlenecek https://t.co/1UodWt4Vb4 https://t.co/KpNZhVmZ	2019-02-12 14:12:12	48	12	3 gün
1,10E+23	Poşetlerle ilgili beyanda bulunmayana 18 bin 43 lira ceza https://t.co/UvsRKxup5W https://t.co/vDBrwNpEVD	2019-02-12 14:00:18	11	3	4 gün
1,10E+23	#CANLI Yerel seçimlere doğru https://t.co/ADcc7GLx9Y Cumhurbaşkanı Erdoğan 'İhtiyaç sahiplerine şehit yakınla	2019-02-12 13:54:47	11	4	1 gün
1,10E+23	Şehit Astsubay Üstçavuş Yakup Avşar Niğdede toprağa verildi https://t.co/l0r0mw4NgW https://t.co/etZLGPMDe	2019-02-12 13:51:42	28	2	1 gün
1,10E+23	Metrobüste 'biraz ilerle' tartışması kanlı bitti https://t.co/mcV5fqg2CB https://t.co/7Bd7yZ3hfD	2019-02-12 13:50:00	2	1	1 gün
1,10E+23	Pendik'te çok sayıda sahte içki ele geçirildi https://t.co/F3KTBEydV9 https://t.co/Uk8vKEZEMu	2019-02-12 13:40:01	7	1	1 gün
1,10E+23	Antalya'ya kuvvetli yağış uyarısı https://t.co/L0NR4CxJXZ https://t.co/ydYoItP0vv	2019-02-12 13:40:01	8	1	1 gün
1,10E+23	Sultangazi'de akraba kavgası: 2 yaralı https://t.co/yM9tlri4lt https://t.co/0aSYycDFWW	2019-02-12 13:36:12	4	0	1 gün
1,10E+23	#CANLI Yerel seçimlere doğru https://t.co/ADcc7GLx9YCumhurbaşkanı Erdoğan Çorum'da konuşuyorhttps://t.c	2019-02-12 13:33:34	11	4	1 gün
1,10E+23	#CANLI Yerel seçimlere doğru https://t.co/XPNd5YWcj4	2019-02-12 13:30:40	11	4	1 gün
1,10E+23	Makedonya'da NATO bayrağı göndere çekildi https://t.co/HXN62IWS2c https://t.co/PJoyuvZhZw	2019-02-12 13:30:25	8	2	1 gün
1,10E+23	BUDO'dan sefer iptalleri https://t.co/qyQc47DWVq https://t.co/xM57FTaJqk	2019-02-12 13:28:58	4	0	1 gün
1,10E+23	O doğdu vicdan öldü!Çuval içerisinde yeni doğmuş bebek bulundu https://t.co/ImQz4U4uXP https://t.co/b7B7pl	2019-02-12 13:24:18	10	3	1 gün
1,10E+23	Şanlıurfada PKK'nın gençlik yapılanmasına 5 tutuklama https://t.co/ih5YHCVf5C https://t.co/THThQAYUwB	2019-02-12 13:23:40	2	0	1 gün
1,10E+23	Cumhurbaşkanı Erdoğan: Fiyatlar yarıya indi https://t.co/fE51RgSkq8 https://t.co/D6naOrTfDb	2019-02-12 13:20:03	7	2	1 gün
1,10E+23	Antalyanın Kemer ilçesinde halk midibüsünün devrildiği anlar güvenlik kameraları tarafından saniye saniye kayde	2019-02-12 13:19:55	19	6	2 gün
1,10E+23	ASELSAN'dan 241 milyon dolarlık sözleşme https://t.co/KaspT8rzdN https://t.co/f7YbCPJrZ0	2019-02-12 13:10:04	9	1	1 gün
1,10E+23	Pompalı tüfekle soygun şüphelisi tutuklandı https://t.co/Si4jbllj4e https://t.co/XWPRine5NA	2019-02-12 13:00:58	2	0	1 gün
1,10E+23	Bıraktığı not tüyler ürpertti 'Kefen param hazır mezar tahtalarım evin altında' https://t.co/PuOtKlerM6 https://t.co	2019-02-12 13:00:02	8	1	1 gün

Şekil 3.12 Kullanılan önemli öznitelikler

Şekil 3.12'de görülen ve haber kanalından alınan tweetler aynı zamanda Türkiye, dünya, ekonomi, spor, magazin, yaşam, sağlık, kültür-sanat ve otomobil olmak üzere 9 sınıftan oluşmaktadır. Sınıflar ve açıklamalar Çizelge 3.1'de yer almaktadır.

Çizelge 3.1 Sınıf (kategori) bilgileri

Sınıf belirteci	Sınıf	Açıklama
1	Türkiye	Türkiye ile ilgili haber tweetlerini içerir.
2	Dünya	Dünya ile ilgili haber tweetlerini içerir.
3	Ekonomi	Ekonomi ile ilgili haber tweetlerini içerir.
4	Spor	Spor ile ilgili haber tweetlerini içerir.
5	Magazin	Magazin ile ilgili haber tweetlerini içerir.
6	Yaşam	Yaşam ile ilgili haber tweetlerini içerir.
7	Sağlık	Sağlık ile ilgili haber tweetlerini içerir.
8	Kültür-sanat	Kültür-sanat ile ilgili haber tweetlerini içerir.
9	Otomobil	Otomobil ile ilgili haber tweetlerini içerir.

Tweetlerin hayatta kalma süreleri için her bir tweet 10 gün(240 saat) boyunca takip edilmiştir. Hayatta kalma süreleri tweetlerin ilk atıldığı zaman ile son retweet veya son beğeni aldığı zaman arasındaki süreyi kapsamaktadır. Her tweet 240 saat izlenmiştir ve bunun sonucunda tweetlerin aylık ortalama kaç saat hayatta kaldığı Çizelge 3.2'de belirtilmiştir.

Çizelge 3.2 Sınıflara göre tweetlerin ortalama hayatta kalma süreleri(saat)

	Şubat	Mart	Nisan	Mayıs	Haziran	Temmuz	Ağustos
Türkiye	37	40	39	45	50	34	40
Ekonomi	43	41	48	52	45	39	42
Spor	78	61	72	80	85	77	76
Dünya	63	58	55	62	64	68	66
Yaşam	24	24	35	24	28	30	32
Sanat	24	24	24	20	28	26	24
Magazin	63	55	70	65	66	60	62
Otomobil	240	24	96	200	150	220	180

Verilerin kolay işlenmesi için MySQL veri tabanı kullanılmıştır. Veriler, sağlıklı işlem yapılabilmesi amacıyla, gerekli ön işlemlerden geçirilmiştir. Python doğal dil işleme kütüphanesi(NLTK) kullanılarak veriler Şekil 1.1'deki gibi arındırılmıştır.

```
text
Apartmandaki çatı yangınında 9 kişi dumandan etkilendi httpst.co7BOrfU7ChD httpst.coEDyWJo1IzT
FETÖ kadın yapılanmasının 2 üst düzey ismi kaçarken yakalandı httpst.copQQkDHBQXA httpst.cok2mf7F04uZ
Çocuğa cinsel istismarı öven kadının serbest bırakılmasıyla ilgili savcılığın itirazı reddedildi… httpst.copxQzDmELve
Sarallara operasyon Çok sayıda gözaltı var httpst.cov1eOImt4wB httpst.coGSGonwFYcx
SONDAKİKA Yolcu otobüsü devrildi Olay yerine ambulanslar sevk edildi httpst.coUZ0KS7XOPl httpst.coUNogOSsl4d
Damadı ve kardeşini yaralayan kayınpedere 19 yıl hapis cezası httpst.coxgcT2USkqp httpst.con56HHglEHq
CANLI Türkiye İnovasyon Haftası httpst.coDUksc8MyFmCumhurbaşkanı Erdoğan Türkiye kimsenin yutamayacağı kada… httpst.coHOTixigXG2
CANLI Türkiye İnovasyon Haftası httpst.coDUksc8MyFmCumhurbaşkanı Erdoğan Kafalarında bir Türk baharı var. O… httpst.coohyABp4BaS
CANLI Türkiye İnovasyon Haftası httpst.coDUksc8MyFmCumhurbaşkanı Erdoğan Kısa zamanda terörün belini bu ülk… httpst.coKzUXItfeFQ
CANLI Türkiye İnovasyon Haftası httpst.coDUksc8MyFmCumhurbaşkanı Erdoğan açıklamalarda bulunuyorhttpst.coPhO0MHh6Cc
CANLI Türkiye İnovasyon Haftası httpst.coQYPCO7KdBU
Kurumlar arası geçici görevlendirilmeye yeni düzenleme httpst.cob1X1CUB6zX httpst.coqXZdE9e4aD
SONDAKİKA Türkiyeden NATOya Güney Kıbrıs Rum Kesimi tepkisi Büyük bir gaflettir httpst.commleXaFcWM httpst.coAvpj2f2kbU
Muhtar olduğunu dava takibi için geldiği adliyede öğrendi httpst.coCpBMSMeT3E httpst.cocp00spWeCZ
CANLI Çamlıca Camii Açılıyor httpst.coOJJPDctknuCumhurbaşkanı Erdoğan Cihat yapıyorum diyerek kimse ibadeth… httpst.cobOzZE3mjFA
CANLI Çamlıca Camii Açılıyor Cumhurbaşkanı Erdoğan İnşallah Çamlıca Camii İstanbulun sembol eserlerinden biri o… httpst.coV2AzjmdHsL
SONDAKİKA Ayvalık'ta kaçak göçmen faciası 9 ölü httpst.coFWJvEnTcaW httpst.coDXz031Nm48
CANLI Çamlıca Camii Açılıyor Cumhurbaşkanı Erdoğan Burası sadece bir cami değil aynı zamanda bir külliyehttpst.coWKI4ZSnKI8
CANLI Çamlıca Camii Açılıyor Cumhurbaşkanı Erdoğan açıklamalarda bulunuyorhttpst.coWKI4ZSnKI8
CANLI Çamlıca Camii Açılıyor httpst.coFETa8ERHGK
AK Partili Özel Talebimiz İstanbulda seçimin yenilenmesi httpst.cosoUb4LcjIk httpst.coPR2XzrZQns
63 bin kişinin aynı anda ibadet edebileceği bir kompleks olarak tasarlanan Türkiyenin en büyük camisi Büyük Çamlı… httpst.coDzFcErRTRV
SONDAKİKA Atatürk Havalimanının işgal girişimine ilişkin davada 150 sanığa  müebbet istendi… httpst.cogAsxrn65sz
SONDAKİKA Devlet Bahçeliden Cumhur İttifakı mesajı İstikbalin mimarı istiklalin muhafızı olacaktır… httpst.conlV8f8Vap
Yalnız yaşayan kadınlarla duygusal yakınlık kurup 10 milyon TLlik vurgun yaptılar httpst.coVR7nN2Jwb7 httpst.conEW91k6616
Şehir merkezindeki plaj turistlerle doldu httpst.cofD9UB0n6Qf httpst.co8xEjaVF9vI
Gaziantepte tarım işçilerini taşıyan kamyonet devrildi 22 yaralı httpst.cotWjMxVeVli httpst.coEWqVmYEUF5
SONDAKİKA Cumhurbaşkanlığı İletişim Başkanı Altundan ABDye tepki httpst.coJOR2i7HaoD httpst.corE7ObZTPfV
Türk ve Azerbaycan Silahlı Kuvvetlerinin Azerbaycanda başlattığı ortak askeri tatbikatta temsili düşman hedefleri… httpst.coMxePUR8jYX
Ramazanda fırsatçılara göz açtırılmayacak httpst.co2wKq3o3T6U httpst.cofUZa877trT
Bir ayda 93 terörist etkisiz hale getirildi httpst.cokeHGYuTCfL httpst.coAkktzKhdoV
Adanada feci olay... Anne diyerek ambulansın arkasından koştu httpst.co04XiciVU31 httpst.coHZ6V8yDlTe
Batmanda fırtına çatıları uçurdu ağaçları devirdi httpst.coTprFRvPPc3 httpst.coIxfXws2rl5
FETÖ sınav sonuçları açıklanmadan tebrik etmiş httpst.conFNBeoYqNh httpst.co0d30SDYeC6
SONDAKİKA Bakan Albayraktan enflasyon açıklaması httpst.coSIgiv3JHrH httpst.coV9VYiaqi6w
Beyzbol sopalı akraba dehşeti httpst.co7MRzlgWETV httpst.coCvRMi2sxGX
```

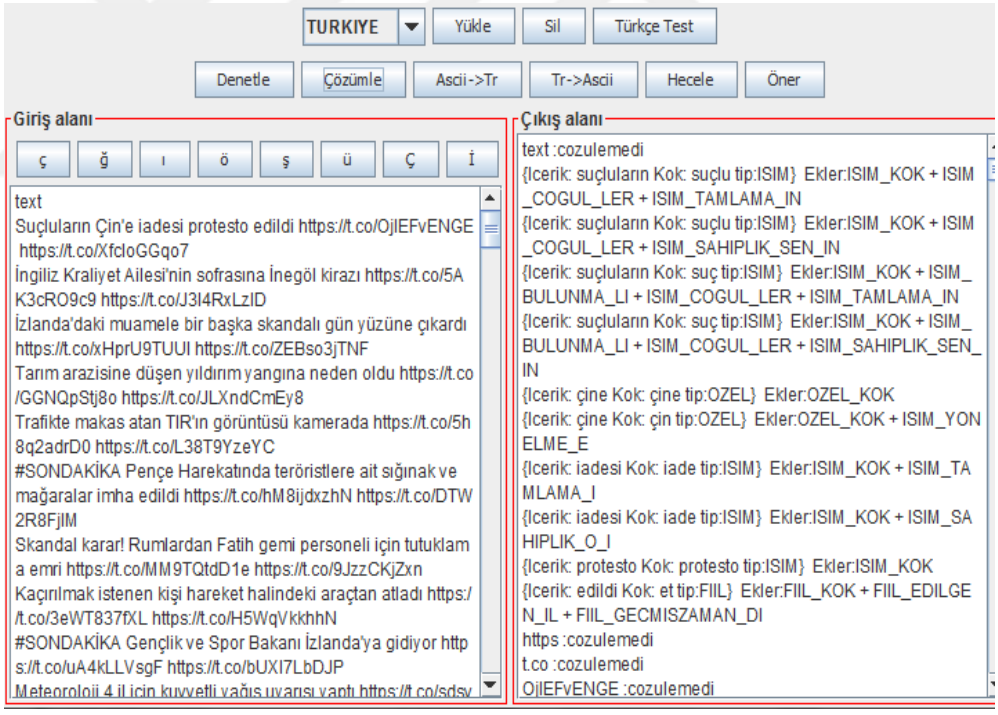
Şekil 3.13 Çeşitli karakter ve simgelerden temizlenmiş tweetler

Öncelikle verilerde yer alan simge ve noktalama işaretleri temizlenmiş, kelimelerin kökleri bulunmuştur. Temizlenen karakterler genel olarak Çizelge 3.3'de yer almaktadır.

Çizelge 3.3 Silinen Gereksiz Karakterler

Silinen Karakterler	Yerine Geçen	Silinen Karakter	Yerine Geçen
…	Boşluk	http	Silindi
……..	Boşluk	@	Silindi
:	Boşluk	(	Silindi
;	Boşluk	)	Silindi
,	Boşluk	_	Silindi
,	Boşluk	+	Silindi

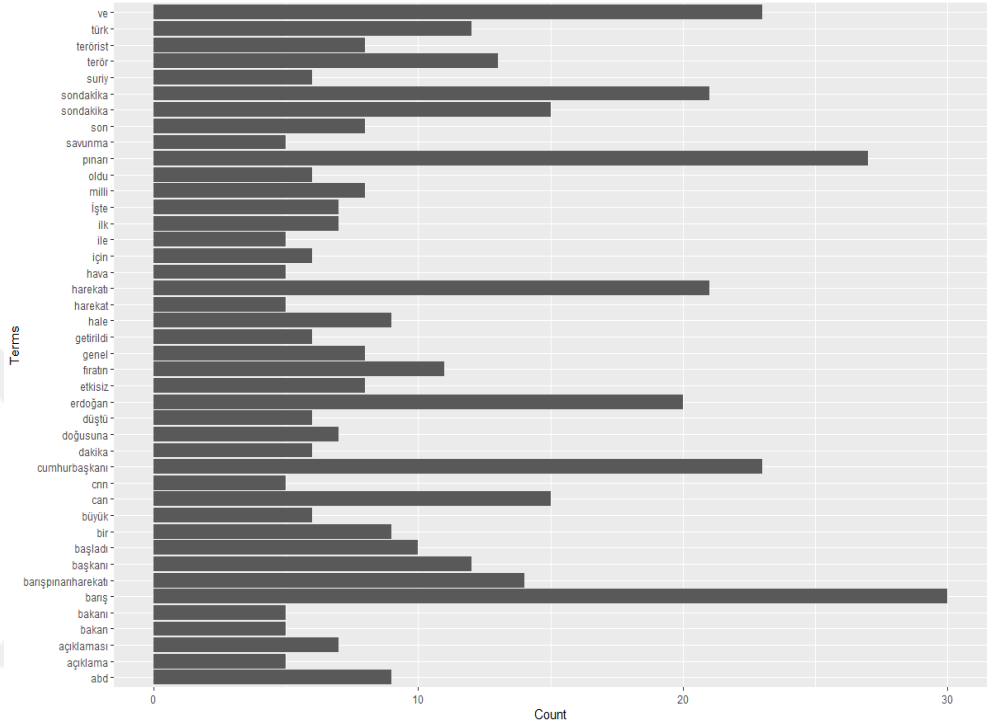
Büyük-küçük harf karmaşasını önlemek amacıyla tüm harfler küçük olacak şekilde işlemden geçirilmiştir. Veri setinde yer alan stopwords olarak adlandırılan gereksiz kelimeler silinmiştir. Zemberek kütüphanesi kullanılarak tweet içerisinde yer alan kelimeler köklerine ayrılmış ve kelime kökleri ile işlem yapılmıştır(Şekil 3.14).



Şekil 3.14 Zemberek ile kelimelerin köklerine ayrılması

Zemberek kütüphanesi kullanılarak elde edilen kökler ile temiz bir veri seti oluşturulmuştur. Bu veri seti kullanılarak kelimelerin frekansı R studio doğal dil işleme kütüphanesinden 'findFreqTerms' kullanılarak hesaplanmıştır. Şekil 3.15'de

kelimeler ve kelimelerin geçme sıklığı grafiği verilmiştir. Burada "Term" kelimeleri temsil ederken "Count" kelimelerin geçme sıklığını, diğer bir deyişle frekansını ifade etmektedir. Tüm parçada en sık geçen kelime "barış" kelimesidir ve 30 kez kullanılmıştır.



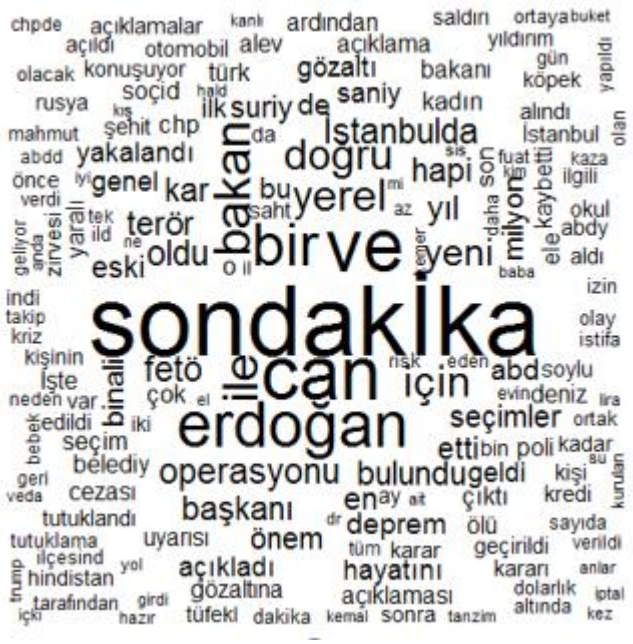
Şekil 3.15 Tweetlerde geçen kelimeler ve frekansları grafiği

En sık geçen kelimeler bulunduktan sonra haber sınıfları tek tek ele alınarak her bir sınıfa ait tweetlerden 10 gün konuşulan haberler seçilmiştir. Hayatta kalma süreleri 240 saat ve daha fazla olan bu haberler içerisinde en çok geçen 20 anahtar kelime (keywords) belirlenmiştir. Ardından her sınıf içinde en çok geçen 500 anahtar kelime belirlenmiştir. Çizelge 3.4'de sınıflara göre en sık geçen 9 kelime listelenmiştir.

Çizelge 3.4 Sınıflara göre en sık geçen kelimeler

TÜRKİYE	EKONOMİ	DÜNYA	MAGAZİN	SPOR
('cumhurbaşkanı', 49)	('tanzim', 2)	('suriye', 7)	('sondakika', 3)	('ümraniyespor', 3)
('erdoğan', 45)	('satış', 2)	('canlı', 7)	('efsane', 2)	('yarı', 3)
('canlı', 37)	('milyon', 2)	('zirve', 6)	('moda', 2)	('final', 3)
('ol', 32)	('lira', 2)	('cumhurbaşkanı', 3)	('hayat', 2)	('trabzonspor', 3)
('et', 31)	('mal', 2)	('açıklama', 3)	('veda', 1)	('şok', 3)
('yerel', 27)	('ol', 2)	('kriz', 3)	('usta', 1)	('ziraat', 3)
('bakan', 27)	('gümrük', 2)	('tırman', 3)	('oyuncu', 1)	('türkiye', 3)
('seçim', 26)	('kapı', 2)	('hindistan', 3)	('açıklama', 1)	('kupa', 3)
('doğru', 26)	('hizmet', 2)	('erdoğan', 2)	('önemli', 1)	('veda', 3)

R studioda veri görselleştirilmesinde kullanılan, 2 ve 3 boyutlu çizimler yapılmasını sağlayan Matplotlib kütüphanesi kullanılarak en sık kelimelerin yer aldığı bir şekil oluşturulmuştur. Şekil 3.16'da en sık kullanılan kelimeler görülmektedir.



Şekil 3.16 Sık geçen kelimeler(wordcloud)

Kategorilere göre sık geçen kelimeler bulunduktan sonra veri seti sınıflandırmaya hazır hale getirilmek için ön işlemlerden geçmiştir. FastText modeli ile sınıflandırılması amaçlanan veri Çizelge 3.5'deki gibi düzenlenmiştir.

Çizelge 3.5 Haber başlığı tweetleri ve kategorileri

Sınıf	Haber başlığı
türkiye	__label__türkiye şehit astsubay üstçavuş avşar niğde toprak ver
ekonomi	__label__ekonomi vergi konsey genel kurul bakan Albayrak önemli açıklama
magazin	__label__magazin usta oyuncu aytaç arman hayat kaybet
dünya	__label__dünya deprem kâhin frank yeni tahmin büyük
sağlık	__label__sağlık kanser neden art ümit aktaş
spor	__label__spor yarı final şok ziraat türkiye kupa veda et
otomobil	__label__otomobil tesla araç yeni özellik köpek mod
yaşam	__label__yaşam bisiklet tur tanış konvoy gör fotoğraf makine
KulturSanat	__label__KulturSanat istanbul sualtı zengin belgesel ol

Çizelge 3.5'de her haber kategorisinden birer örnek verilmiştir. Tweetler belirli ön işlemlerden sonra gereksiz kelime, karakter, sembollerden arındırılmış ve zemberek kütüphanesi kullanılarak köklerine ayrılmıştır.

.

4. ARAŞTIRMA BULGULARI ve TARTIŞMA

4.1. Tweetleri Haber Kategorilerine Göre Sınıflandırma

Zemberek kütüphanesinin kullanımıyla köklerine ayrılan ve analize hazır hale getirilen verilerden yeni dosyalar oluşturulmuştur. Model üretilmeden önce hazırlanan dosyalar:

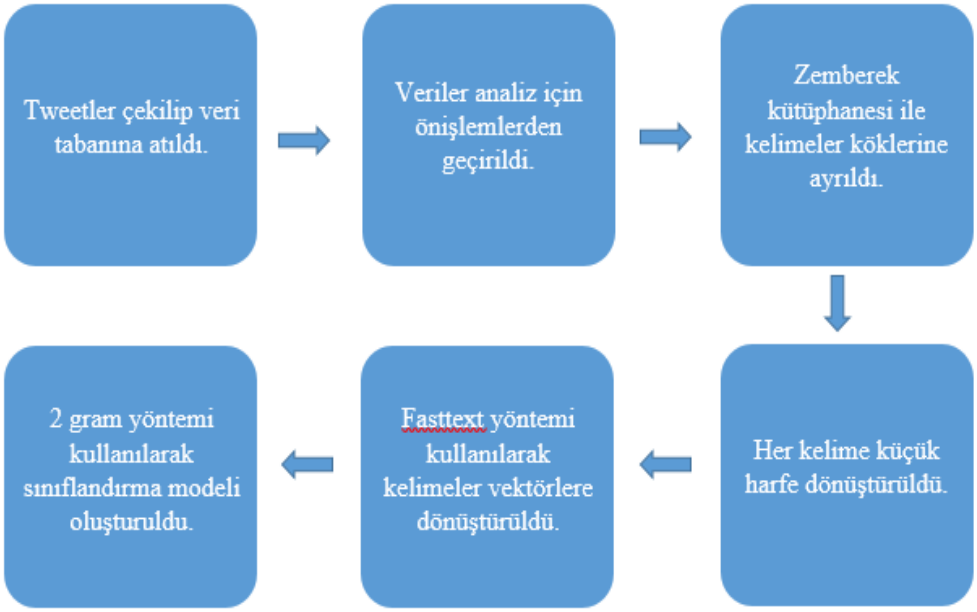
- **classification_result:** Proje çalıştırıldığında sonuçlar buraya gelecektir. Burada 2 sonuç görülmektedir:

i) **classification.txt:** Proje çalıştırıldıktan sonra sonuçlar bu dosyaya yazılmaktadır.

- **classification_percent.txt:** Bu txt dosyasında ise sonuçların yüzdesi yer almaktadır.
- **data:** Verilerin bulunduğu klasördür. Bu klasörde 2 txt dosyası bulunmaktadır:

i) **test.txt:** Test için hazırlanan veriler burada yer almaktadır.

ii) **train.txt:** Eğitim için hazırlanan veriler burada bulunmaktadır.



Şekil 4.1 Sınıflandırma akış şeması

Sınıflandırılma yapılırken kategorisi bilinen verilerin yer aldığı train.txt dosyası eğitime alınır ve bir model oluşturulmuştur. Model oluşturulduktan sonra kategorisi bilinmeyen tweetlerin oluşturduğu test.txt dosyasındaki veriler oluşturulan yeni modelle sınıflandırılmıştır. Modelde 2-Gram yöntemi ve zemberek ile köklerine ayrılmış veriler kullanılarak daha iyi sonuç elde edilmesi amaçlanmıştır. Modelin ortalama başarı oranı %78.9954225 olmaktadır. Çizelge 4.1'de kategorilere göre cümle başarı oranı yer almaktadır.

Çizelge 4.1 Sınıflandırma sonuçları

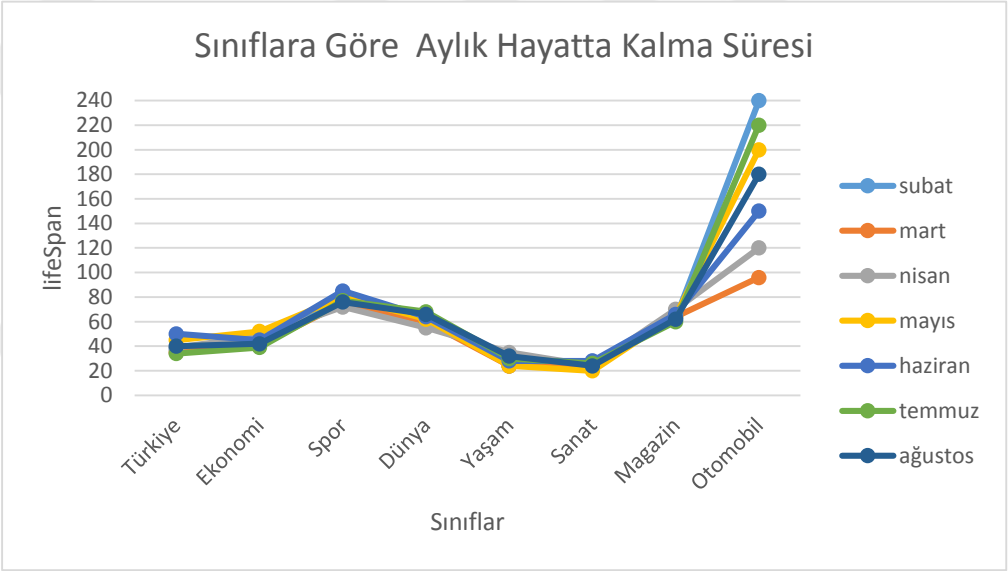
Veri	Sonuç(Tahmin)
fatih terim kosova maç ardından sessizlik boz	('spor', 0.966797)
ankara eski sevgili dehşet ölü	('türkiye', 0.878906)
putin hak yeni iddia	('dünya', 0.974609)
yılbaşı parti için konfetili balon	('kültürsanat', 0.265625)
altın ve gümüş yükseliş	('ekonomi', 0.994141)
emre mor şov yap	('magazin', 0.449219)
dünya en geliş selfie çubuğu	('yaşam', 0.386719)
yeditepe üniversite kellik çözüm krem	('sağlık', 0.226563)

Çizelge 4.2 Sınıflara göre ortalama başarı oranı

Sınıf	Ortalama Başarı Oranı
Türkiye	0.880845
Ekonomi	0.771432
Dünya	0.807948
Spor	0.8474
Yaşam	0.387039
KültürSanat	0.671118
Magazin	0.629583
Otomobil	0.413523
Sağlık	0.611895

4.2. Tweetlerin Kategorilerine Göre Hayatta Kalma Süreleri

Büyük kitleler söz konusu olduğunda sosyal medya ortamlarında bir haberin yaratmış olduğu tepki ve ilginin değerlendirilmesi o haberin almış olduğu beğeni sayısı, paylaşılma sayısı ve konuşulduğu gün diğer bir değişle hayatta kalma süresi dikkate alınarak gözlemlenebilmektedir. Bu çalışmada Twitter API kullanılarak her gün çekilen haber tweetleri 240 saat (10 gün) izlenerek hayatta kalma süreleri takip edilmiştir. Haber başlıklarının yer aldığı sınıflar bir haber kanalı olan kullanıcı tarafından kategorize edilmiştir.



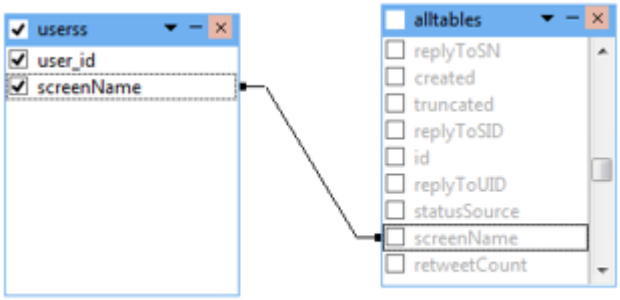
Şekil 4.2 Sınıflara Göre Aylık Hayatta Kalma Süresi

Şekil 4.2'de 6 ay boyunca çekilen tweetlerin sınıflarına göre aylık ortalama hayatta kalma sürelerinin karşılaştırılması yer almaktadır. Örneğin şubat ayında otomobil sınıfında yer alan tweetler 10 günden (240 saat) fazla gündemde kalırken mart ayında bu sürenin 4 güne (96 saat) düştüğü gözlemlenmiştir.

Çizelge 4.3 Haber başlıkları ve hayatta kalma süreleri

Tweet	LifeSpan (Gün)	Sınıf	Oluşturulma tarihi
Cumhurbaşkanı Erdoğan: Venezuela altını Çorum'da işlenecek	3	Türkiye	2019-2-12 14:12:12
Poşetlerle ilgili beyanda bulunmayana 18 bin 43 lira ceza	4	Ekonomi	2019-2-12 14:00:18
Makedonya'da NATO bayrağı göndere çekildi	1	Dünya	2019-2-12 13:30:25
SONDAKİKA Usta oyuncu Aytaç Arman hayatını kaybetti	2	Magazin	2019-2-26 14:27:01
Ümraniyespor yarı finalde! Trabzonspor'a şok	4	Spor	2019-2-26 19:03:00
Konvoyu gören fotoğraf makinesine sarıldı	1	Yaşam	2019-2-17 15:55:29
Dr. Ümit Aktaş "Bu kadar çok hasta olmamızın sebebi nedir?"	2	Sağlık	2019-2-20 18:06:47

Çizelge 4.3'de haber başlığını içeren tweetler, günü ifade eden hayatta kalma süreler, haber başlıklarının hangi kategoride olduğu ve tweetlerin oluşturulma zamanı yer almaktadır. Her tweetin hangi tarihte ve saatte paylaşıldığı bilgisi ile son ilgi aldığı zaman veri setinde yer almaktadır. Bu bilgiler kullanılarak sosyal medyada paylaşılan bir haberin ne kadar ilgi topladığının ve haberin yayılma hızının bulunması amaçlanmıştır. Her kategoride değişiklik gösteren hayatta kalma süreleri MySQL veri tabanına alınmış ve bağlantılı tablolar elde edilmiştir.



Şekil 4.3 MySQL ile oluşturulan tablolar

Şekil 4.3‘de yer alan tablolar analiz yapma kolaylığı sağlamak amacıyla oluşturulmuştur. Buna göre veri setinde öznitelik seçimi yapılarak birbiriyle bağlantısı olabilecek nitelikler tabloda birleştirilmiştir.

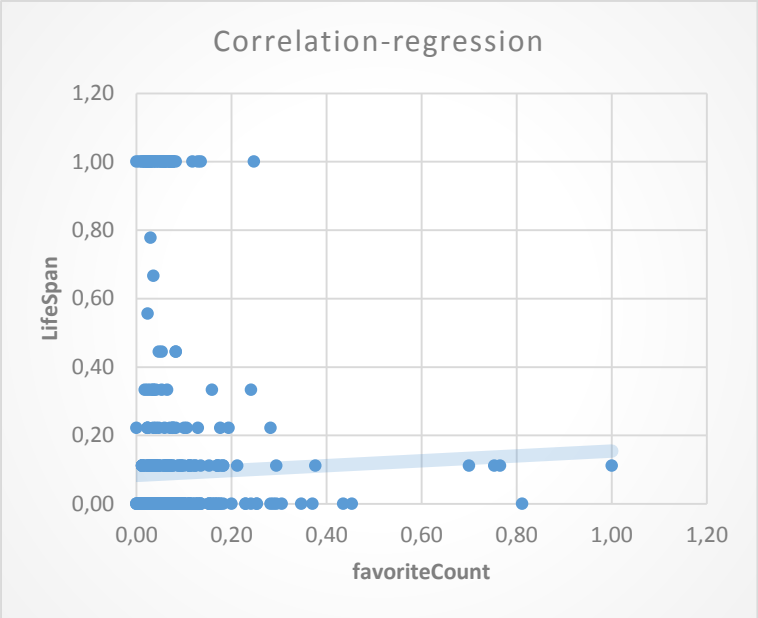
Çizelge 4.4‘de bazı tweetlere ait özniteliklerin oluşturduğu tablo örneği verilmiştir. Haber tweetlerinin aldıkları beğeni (favoriteCount) ve paylaşılma sayıları (retweetCount) da veri setinde yer almaktadır.

Çizelge 4.4 Önemli Öznitelikler

favoriteCount	LifeSpan	retweetCount	Class
48	3	12	Türkiye
11	4	3	Ekonomi
2	1	0	Otomobil
28	1	2	Türkiye
2	1	2	Sağlık
7	1	1	Türkiye
6	1	1	KültürSanat
4	1	0	Türkiye
13	3	2	Magazin
21	2	4	Spor
8	1	2	Dünya
13	2	0	Yaşam

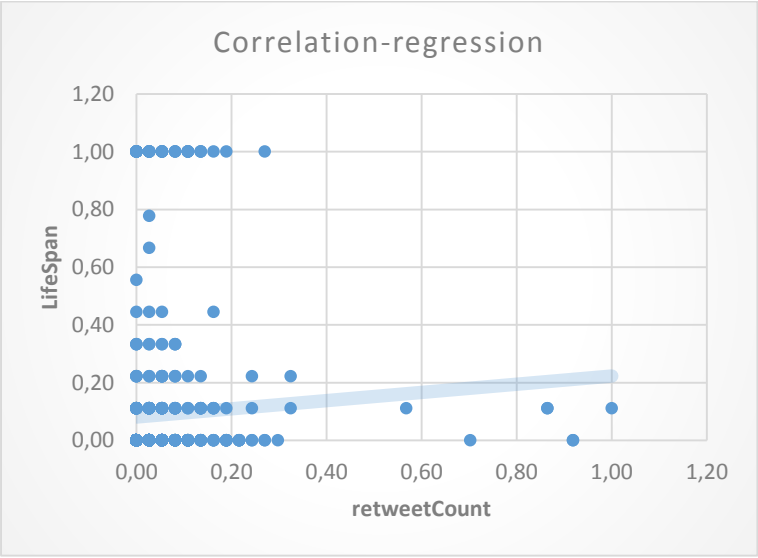
Bu çizelgenin oluşturulmasındaki amaç tweetlerin hayatta kalma süreleri ile beğeni ve retweet sayısı arasındaki bağlantıya ulaşmaktır. LifeSpan, favoriteCount ve retweetCount arasındaki bağlantının bulunabilmesi için korelasyon yöntemi kullanılmıştır.

Öncelikle veri setinde yer alan ayrık değerlerin sonucu yanlış etkilememesi amacıyla normalizasyon işlemi yapılıştır. Örnek olarak en yüksek beğeni sayısı 170 ve en düşük beğeni sayısı 0 arasındaki farkın indirgenmesi amacıyla (3.1)’de yer alan normalizasyon formülü kullanılmıştır. Veriler 0-1 aralığı arasına indirgenmiştir. Aynı şekilde LifeSpan için de 1 ile 10 gün arasında değişen değerler 0-1 aralığına indirgenmiştir. Analize uygun hale getirilen veri setinde öncelikle favoriteCount ile LifeSpan arasındaki bağıntı (3.2)’de yer alan korelasyon formülü ile hesaplanmıştır.



Şekil 4.4 FavoriteCount ve LifeSpan arasındaki bağıntı

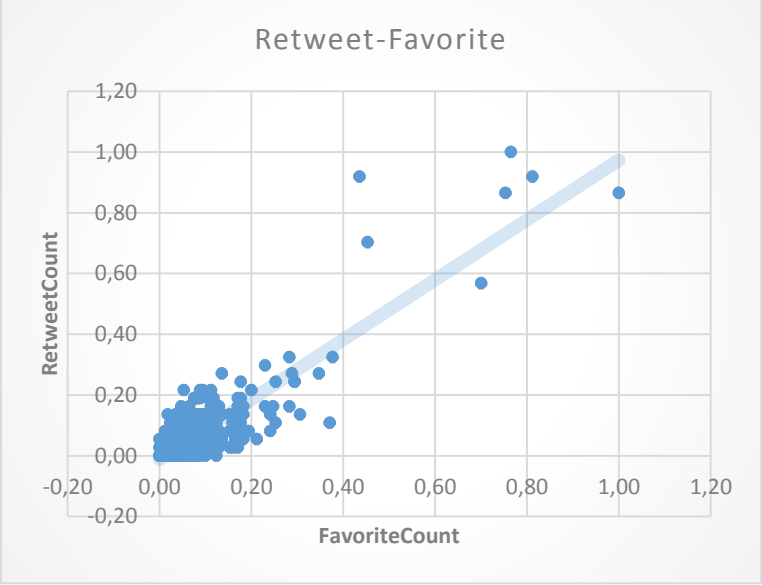
Şekil 4.4'de görüldüğü üzere haberlerin beğeni sayısı ile hayatta kalma süreleri arasında düşük pozitif korelasyon bulunmuştur. Korelasyon katsayısı 0.0244 olarak hesaplanmıştır. Hayatta kalma süreleri gün sayısı olarak ele alınmış bu da düşük de olsa beğeni sayısı ile aralarında pozitif bir bağlantı olduğunu göstermektedir.



Şekil 4.5 retweetCount ve LifeSpan arasındaki bağıntı

Şekil 4.5'de haber tweetlerinin retweet sayısı ile hayatta kalma süresi arasındaki bağıntı korelasyon yöntemi kullanılarak hesaplanmıştır. Burada yine ayrık değerlerin sonucu yanlış etkilememesi amacıyla normalizasyon yöntemi kullanılarak retweet için en büyük değer olan 37 ve en küçük değer olan 0 aralığı 0-1 aralığı arasına indirgenmiştir. Aralık içinde 1 en büyük değeri temsil etmektedir. Korelasyon sonucunda retweetCount ile lifeSpan arasında düşük pozitif korelasyona ulaşılmıştır. Korelasyon katsayısı 0.0557 olarak hesaplanmıştır. Bu sonuçlar göz önüne alınarak daha çok beğenilen ve paylaşılan haberler daha hızlı yayılım göstermekte ve daha çok yayılmaktadır yorumu yapılabilmektedir.

Normalizasyon yöntemi uygulanmış olan veri setinde korelasyon yöntemi kullanılarak tweetlerin beğeni sayısı ile paylaşılma sayısı arasındaki bağlantıya ulaşılmıştır.



Şekil 4.6 retweetCount ve favoriteCount arasındaki bağıntı

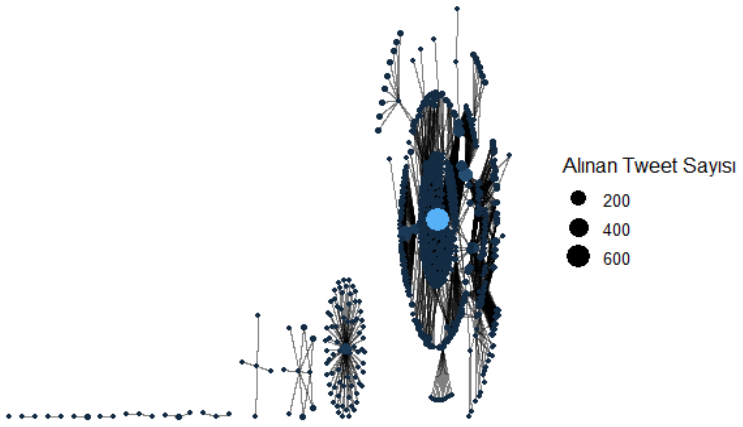
Yapılan analiz sonucunda favoriteCount ile retweetCount arasında yüksek pozitif korelasyona ulaşılmaktadır(Şekil 4.6). Korelasyon katsayısı ise 0.86 olarak hesaplanmıştır. Bu sonuca bakılarak bir tweet için çok beğeni alan haberlerin daha çok paylaşıldığı ve dolayısıyla daha çok yayıldığı söylenebilmektedir.

4.3. Sosyal Ağ Analizi ile Haber Yayılımı

Teknolojinin hızla geliştiği günümüzde bilgi yayılımı da oldukça hızlı gelişmektedir. Özellikle son yıllarda sosyal medya platformlarının sıklıkla kullanılmaya başlanması bilginin daha geniş çevrelere yayılmasına katkı sağlamaktadır. Bu çalışmada belirli bir kullanıcıdan çekilmiş olan tweetlerin yayılımı sosyal ağ analizi kullanılarak gözlemlenmeye çalışılmıştır.

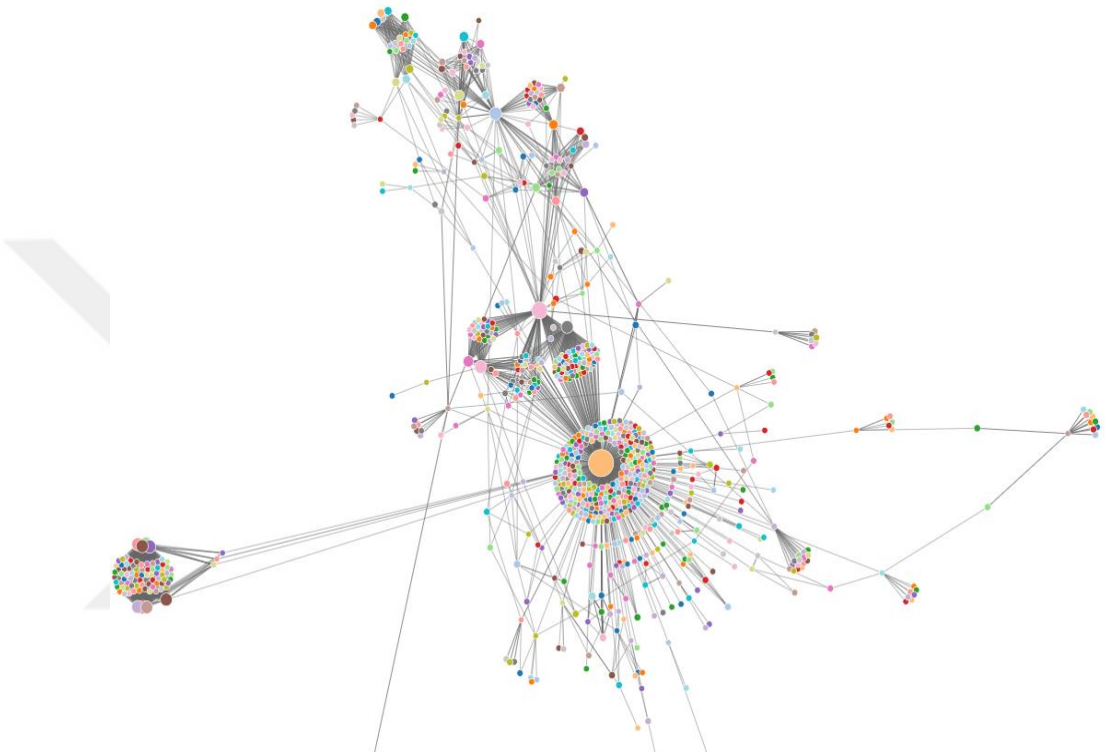
6 ay boyunca tweetlerini çekmiş olduğumuz kullanıcının arkadaşları ile arasındaki ilişki öncelikli olarak incelenmiştir. Bu durumda kullanıcının paylaşmış olduğu bir tweetin nasıl ve ne kadar hızlı yayıldığı ağ analizi görselleştirilmesi ile görülebilmektedir. Burada karşılıklı ilişkinin gözlemlenebilmesi için birbirleriyle ilişkili iki özelliğin olması gerekmektedir. Sosyal ağ analizinde bunlar düğüm (node) ve kenar (edge) olarak adlandırılmaktadır. Burada düğümler kullanıcının arkadaşları ve kenarlar ise paylaşmış oldukları tweetleri içermektedir.

Bu çalışmada sosyal ağ analizi R studio kullanılarak yapılmıştır. Haber tweetlerini çektiğimiz kullanıcının arkadaş bilgisini içeren veri setimiz yer almaktadır. Ağ görselleştirilmesi için R Studio da yer alan “ggraph” ve “networkD3” kütüphaneleri tercih edilmiştir.



Şekil 4.7 Alınan tweet sayısını içeren Twitter ağ analizi

Şekil 4.7‘da düğüm ve kenarlar yer almaktadır. Yuvarlak noktalar kullanıcıları içeren düğümleri temsil etmektedir. Çizgiler ise aralarındaki bağlantı olan tweetleri temsil etmektedir. Şekil 4.7’da da görüldüğü üzere paylaşılan tweet sayısına göre düğümlerin büyüklüğü değişmektedir. Burada bir kullanıcının ne kadar bilgi yaydığı gözlemlenebilmektedir.



Şekil 4.8 Dinamik ağ analiz grafiği

Şekil 4.8‘de dinamik ağ analiz grafiği yer almaktadır. Daha ayrıntılı gözlem yapılmasını sağlayan bu grafik R studio programının “networkD3” kütüphanesi kullanılarak elde edilebilmektedir. Grafikte haber tweetlerini en çok yayan kullanıcılar görülmektedir. Bu grafiklerden yola çıkılarak aktif olan ve arkadaş sayısı çok olan kullanıcıların haber yayılımında büyük rol oynadıkları ve bilgi yayılımını hızlandırdıkları söylenebilmektedir. Analitik olarak belirtilecek olunursa 1000’den fazla takipçi sayısına sahip kullanıcıların daha geniş kitlelere ulaştığı gözlemlenmiştir. Burada sözü geçen kullanıcı 5.2 milyon takipçi sayısına sahiptir.

5. SONUÇLAR ve ÖNERİLER

5.1. Sonuçlar

Çoğu araştırmada görülebileceği gibi, sosyal medyanın haberlerin yayılması üzerinde olumlu bir etkisi vardır. Bu etkiyi göz önünde bulundurarak çalışmamızda twitter kullandık çünkü twitter, haberlerin yayılmasında en etkili sosyal medya platformlarından biridir. Bu çalışmada, haberlerin yaşam süresinin beğeni ve retweet sayısı ile nasıl ilişkili olduğunun gözlemlenmesi, haber yayılımında etkili olan niteliklerin değerlendirilmesi amaçlanmıştır.

Proje kapsamında yapılmış olan çalışmada 6 ay boyunca her gün 100 tweet çekilerek oluşturulmuş veri seti gerekli ön işlemlerden geçirilerek analiz edilmiştir. Haber tweetlerinin %78.9954225 başarı oranı ile kategorize edilme işlemleri gerçekleştirilmiştir.

6 ay boyunca bir haber kanalı sayfasından toplanan tweet'ler ön işlemden geçirilerek uygun bir veri setine dönüştürüldü. Her tweet 10 gün boyunca takip edildi ve yaşam süresi gözlemlendi. Tweetlerde yer alan kelimeler zemberek kullanılarak köklendirildi, böylece kullanılacak yöntemler için sağlıklı değerler oluşturuldu. Veriler normalizasyon yöntemi ile 0-1 değerleri arasında dağıtıldı. Sonuç olarak LifeSpan-RetweetCount ile LifeSpan-FavoriteCount arasında düşük pozitif korelasyon varken, FavoriteCount-RetweetCount ile 0.86 korelasyon katsayısı arasında yüksek pozitif korelasyon olduğu görülmüştür.

Bununla birlikte sosyal ağ analizi yapılarak bir haber kanalı olan ve tweetlerini çekmiş olduğumuz twitter kullanıcısının arkadaşları ile ilişkisi gözlemlenmiştir. Ağ analizinden elde edilen bilgiler en az 100 tweet atan kullanıcıların haber yayılımında %17.1 değerle etkili olduğu ve 600 üzeri tweet paylaşan kullanıcıların bilgi yayılımında %69.7 gibi bir değerle daha etkili olarak büyük rol oynadığı tespit edilmiştir. Sonuç olarak Twitter'de haber difüzyonunun sağlanmasındaki en büyük

rolü aktif ve çevresi büyük olan kullanıcıların oynamış olduğu gözlemlenmektedir. Haberlerin hayatta kalma sürelerini ise beğeni ve retweet sayısı belirlerken geniş kitlelerin hangi türde haberleri daha çok takip ettiği ve ne kadar gündemde kaldığı gözlemlenebilmektedir.

5.2. Öneriler

Değerlendirmeler sonucunda, tweetlerin favori sayıları ile retweet sayıları arasındaki ilişkinin gelecekteki tweetlerin yayılma tahminini kolaylaştıracağı düşünülüyor. Ayrıca, retweet sayısı ve favori sayısını kullanarak haberlerin ömrünü tahmin edebileceğimizi düşünüyoruz.

İlerideki çalışmalarda sosyal ağ analizi ve haberlerin hayatta kalma süreleri kullanılarak gündemde en çok konuşulan ve en çok ilgi çeken haber türlerinin tahmin edilebileceği ve bir haberin ne kadar hayatta kalacağının hesaplanabileceği düşünülmektedir. Daha geniş çapta çalışmalar yapılarak bu bilgilere ulaşılabilir.

KAYNAKLAR

ADALI, E., 2012. Doğal Dil İşleme. Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 5(2): 1-19.

AKIN, A. A., 2011. Zemberek NLP. http://zembereknlp.blogspot.com/2011/04/zemberek-3-durum.html

AKIN, A. A., 2007. Zemberek, An Open Source Nlp Framework for Turkic Languages. Structure, 10: 1-5.

AKGÜL, E. S., ERTANO, C. ve DİRİ, B., 2016. Twitter Verileri ile Duygu Analizi. Pamukkale University Journal of Engineering Sciences, 22(2): 106-110.

AKYOKUŞ, S., 2018. TBD veri Madenciliği Günü. https://slideplayer.biz.tr/slide/2979950/

ALLCOTT, H. and GENTZKOW, M., 2017. Social Media and Fake News in The 2016 Election. Journal of Economic Perspectives, 31(2): 211-36.

ALPAR, R., 2003. Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş 1. Nobel Akademik Yayınları, Ankara, 404s.

AMASYALI, M. F., 2005. Türkçe Wordnet'in Otomatik Olarak Oluşturulması. IEEE 13. Sinyal İşleme ve İletişim Uygulamaları Kurultayı (SİU-2005), 16-18 Mayıs, Kayseri, s.1-14.

ANONİM., 2013. Sosyal Medya Uzmanı. http://www.sosyalmedyauzmani.com/

ARKLAN, Ü. T., 2008. Bilgi Toplumu Ve İletişim: Bilginin Yayılması Sürecinde Kitle İletişim Araçları Ve İnternet. Selçuk İletişim, 5(3): 67-80.

BAŞKAYA, F. A. ve Aydın İ., 2017. Haber Metinlerinin Farklı Metin Madenciliği Yöntemleriyle Sınıflandırılması. International Artificial Intelligence and Data Processing Symposium (IDAP), 16-17 Eylül, Malatya, s.5.

BHATTACHARYA, D. and RAM, S., 2012. Sharing News Articles Using 140 Characters: A Diffusion Analysis on Twitter. International Conference on Advances in Social Networks Analysis and Mining, 26-29 August, Istanbul, Turkey, s.966-971.

CHEONG, M. and LEE, V., 2009. Integrating Web-Based İntelligence Retrieval And Decision-Making From The Twitter Trends Knowledge Base. Proceedings of the 2nd ACM Workshop on Social Web Search and Mining, 2: 1-8.

DEDE, Y., 2018. FastText Nedir?. https://medium.com/@deduska/fasttext-nedir t%C3%BCrk%C3%A7e-1e10443e8567

DEMİR, Y. ve AYHAN, B., 2020. Sosyal Medyanın Gündem Belirleyicileri: Twitter'da Gündem Belirleme Süreci Üzerine Bir Sosyal Ağ Analizi. İletişim Kuram ve Araştırma Dergisi, 2020(51): 1-19.

DILRUKSHI, I., ZOYSA, K. D. And CALDERA, A., 2013. Twitter News Classification Using SVM. The 8th International Conference on Computer Science & Education (ICCSE), 26-28 April, Colombo, Sri Lanka, s.287-291.

DOLGUN, M. Ö., ÖZDEMİR, T. G. ve OĞUZ, D., 2009. Veri Madenciliği'nde Yapısal Olmayan Verinin Analizi: Metin Ve Web Madenciliği. İstatistikçiler Dergisi: İstatistik ve Aktüerya, 2(2): 48-58.

FANG, A. and MILED, Z. B., 2017. Does Bad News Spread Faster?. International Conference on Computing, 26-29 January, Santa Clara, USA, s.793-797.

HOANG, T. B. N. and MOTHE, J., 2018. Predicting information diffusion on Twitter – Analysis of predictive features. Journal of Computational Science, 28: 257-264.

IMBERMAN, S. P., 2001. Effective Use of The Kdd Process and Data Mining for Computer Performance Professionals. 27th International Computer Measurement Group Conference, 2-7 December, California, s.611-620.

JOULIN, A., GRAVE, E., BOJANOWSKİ, P. and MİKOLOV, T., 2016. Bag of Tricks for Efficient Text Classification. Computer Science arXiv preprint arXiv:1607.01759.

JOULIN, A., GRAVE, E., BOJANOWSKİ, P., DOUZE, M., JÉGOU, H. and MİKOLOV, T., 2016. Fasttext.Zip: Compressing Text Classification Models. Computation and Language arXiv:1612.03651.

KARADAĞ, A. ve TAKÇI, H. 2010. Metin Madenciliği ile Benzer Haber Tespiti. Akademik Bilişim.

KNOKE, D. and YANG, S., 2019. Social Network Analysis. Sage Publications, USA, 216s.

KWAK, H., LEE, C., PARK, H. and MOON, S., 2010. What is Twitter, A Social Network or A News Media?. Proceedings Of The 19th İnternational Conference On World Wide Web, 1 April, North Carolina, USA, s.591-600.

LETIERCE, J., PASSANT, A., DECKER, S. and BRESLİN, J. G., 2010. Understanding How Twitter is Used to Spread Scientific Messages. The Web Science Conference. 26-27 April, Raleigh, North Carolina, s.1-8.

OFLAZER, K., 2016. Türkçe ve Doğal Dil İşleme. Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 5(2): 1-12.

OĞUZLAR, A., 2003. Veri ön işleme. Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 21: 67-76.

ÖZKAN, Y., 2016. Veri Madenciliği Yöntemleri. Papatya Bilim Yayıncılık Eğitim, İstanbul, 240s.

PIRAMUTHU, S., 2004. Evaluating Feature Selection Methods for Learning in Data Mining Applications. European Journal of Operational Research, 156(2): 483-494.

REMY, C., PERVİN, N., TORİUMİ, F. and TAKEDA, H., 2013. Information Diffusion on Twitter: Everyone Has Its Chance, But All Chances Are Not Equal. 2013 International Conference on Signal-Image Technology & Internet-Based Systems, 2-5 December, Kyoto, Japan, s.483-490.

SAVAŞ, S., TOPALOĞLU, N. ve YILMAZ, M., 2012. Veri Madenciliği Ve Türkiye'deki Uygulama Örnekleri. İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 11(21): 1-23.

SAYRE, B., BODE, L., SHAH, D. and SHAH, C., 2010. Agenda Setting in a Digital Age: Tracking Attention to California Proposition 8 in Social Media, Online News and Conventional News. Policy & Internet, 2(2): 7-32.

SHILAKES, C. C., 1998. Enterprise information portals. Merrill Lynch in-depth report, 29: 47-52.

SMIRAGLIA, R. P., 2015. Empirical Techniques For Visualizing Domains. Domain Analysis for Knowledge Organization, s.51-89.

SUH, B., HONG, L., PİROLLİ, P. and CHİ, E. H., 2010. Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network. IEEE Second

International Conference on Social Computing, 20-22 August, Minneapolis, USA, s.177-184.

ŞEKER, S. E., 2015. Çizge Teorisi (Graph Theory). YBS ansiklopedi, *2*(2): 17-29.

ŞEKER, Ş. E., 2015. Metin Madenciliği (Text Mining). YBS Ansiklopedi, 2(3): 30-32.

WASSERMAN, S. and FAUST, K., 1994. Social Network Analysis: Methods and Applications. Cambridge university press, Cambridge, UK, 857s.

WEISS, S. M., INDURKHYA, N., ZHANG, T. and DAMERAU, F., 2010. Text Mining: Predictive Methods For Analyzing Unstructured İnformation. Springer Science & Business Media, USA, 249s.

WU, S., HOFMAN, J. M., MASON, W. A. and WATTS, D. J., 2011. Who Says What to Whom on Twitter. Proceedings of the 20th International Conference on World Wide Web, 1 March, New York, USA, s.705-714.

YAVUZ, S. ve DEVECİ, M., (2012). The Effect Of Statistical Normalization Techniques On The Performance Of Artificial Neural Network. Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 40: 167-187.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı : Duygu SAĞALTICI
Uyruğu : Türkiye Cumhuriyeti
Doğum Yeri ve Tarihi : Hatay, 20.03.1993
e-Mail : sagalticiduygu7@gmail.com

EĞİTİM

Derece		Adı, İlçe, İl	Bitirme Yılı
Lise	:	Necmi Asfuroğlu Anadolu Lisesi, HATAY	2011
Üniversite	:	Bursa Teknik Üniversitesi Mekatronik Mühendisliği, BURSA (Not: 2.77/4)	2017
Yüksek Lisans	:	Harran Üniversitesi Fen Bilimleri Enstitüsü Elektrik Elektronik Mühendisliği Ana Bilim Dalı, ŞANLIURFA (Not: 4.0/4.0)	2021

TEKNİK BİLGİLER

Bilgisayar ve Yazılım Bilgileri

- Bilgisayar destekli çizim programları (Autocad, 3DS Max, Blender, Solidworks, Matlab)
- C, C++, C# & WPF, Python dilleri ile uygulama geliştirme
- Web tabanlı tasarım ve programlama (PHP, MYSQL, CSS, Angular, JS)
- Sayısal mantık tasarımı (Verilog, VHDL, FPGA)

Elektrik Elektronik Bilgileri

- Baskı devre kartı tasarımı (PCB)
- Gömülü sistem tasarımı
- Analog ve dijital sistemler (Mixed signal)
- Endüstriyel PLC uygulamaları

YABANCI DİLLER

İngilizce (YÖKDİL 71.25)

YAYINLAR

SAĞALTICI, D., ALAY, F., D., EFİL, C., ve İLHAN, N., 2018. Veri Madenciliği Yöntemleri İle Meteorolojik Verilerden Kayıp Güneş Işınım Değerlerinin Tahmini Uluslararası GAP Yenilenebilir Enerji Ve Enerji Verimliliği Kongresi. 10-12 Mayıs 2018, Şanlıurfa, s.335-336

EFİL, C., SAĞALTICI, D., ve BEŞLİ, N., 2018. GAP Bölgesi için Güneş Takip Sistemi Yapılarının ve Kontrol Yöntemlerinin İncelenmesi. Uluslararası GAP Yenilenebilir Enerji Ve Enerji Verimliliği Kongresi. 10-12 Mayıs 2018, Şanlıurfa, s.356-357

İLHAN, N., SAĞALTICI, D., 2020. Twitter'da Duygu Analizi. Harran Üniversitesi Mühendislik Dergisi, 5(2): 146-156.