



T.C

**BANDIRMA ONYEDİ EYLÜL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
EKONOMETRİ ANABİLİM DALI**

YÜKSEK LİSANS TEZİ

**METİN MADENCİLİĞİ YÖNTEMİ İLE AKILLI
ULAŞIM SİSTEMLERİ ALGISININ DUYGU
ANALİZİ İLE ÖLÇÜLMESİ**

Mübin ÖZKUL

175010011

Tez Danışmanı:

Prof. Dr. Mehmet TEKTAŞ

BANDIRMA 2019

T.C

BANDIRMA ONYEDİ EYLÜL ÜNİVERSİTESİ

SOSYAL BİLİMLER ENSTİTÜSÜ

EKONOMETRİ ANABİLİM DALI

YÜKSEK LİSANS TEZİ

**METİN MADENCİLİĞİ YÖNTEMİ İLE AKILLI
ULAŞIM SİSTEMLERİ ALGISININ DUYGU
ANALİZİ İLE ÖLÇÜLMESİ**

Mübin ÖZKUL

175010011

Tez Danışmanı:

Prof. Dr. Mehmet TEKTAŞ

BANDIRMA 2019

YÜKSEK LİSANS TEZ ONAYI

Bandırma Onyedi Eylül Üniversitesi Sosyal Bilimler Enstitüsü EKONOMETRİ Anabilim Dalı, Ekonometri Programında Yüksek Lisans öğrencisi Mübin ÖZKUL tarafından Mehmet TEKTAŞ'ın danışmanlığında hazırlanan “ Metin Madenciliği Yöntemi İle Akıllı Ulaşım Sistemleri Algısının Duygu Analizi İle Ölçülmesi” başlıklı tez aşağıdaki jüri üyeleri tarafından 11/07/2019 tarihinde yapılan Tez Savunma Sınavında oybirliği ile başarılı bulunmuş ve Yüksek Lisans Tezi olarak kabul edilmiştir.

Jüri Başkanı

Jüri-Danışmanı

Jüri Üyesi

Jüri Üyesi

Jüri Üyesi

ETİK BEYAN

TÜRKİYE CUMHURİYETİ

BANDIRMA ONYEDİ EYLÜL ÜNİVERSİTESİ

SOSYAL BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Bu belge ile bu tezdeki bütün bilgilerin akademik kurallara ve etik ilkelere uygun olarak toplanıp sunulduğunu beyan ederim. Bu kural ve ilkelerin gereği olarak çalışmada bana ait olmayan tüm veri, düşünce ve sonuçları andığımı ve kaynağını gösterdiğimi ayrıca beyan ederim.(../.../20..)

Mübin ÖZKUL

ÖZET

METİN MADENCİLİĞİ YÖNTEMİ İLE AKILLI ULAŞIM SİSTEMLERİ ALGISININ DUYGU ANALİZİ İLE ÖLÇÜLMESİ

Mübin ÖZKUL

Çalışmamızda günlük hayatımızda sıklıkla karşılaştığımız ve gündemde bir konu olan akıllı ulaşım sistemlerini inceledik. Çalışmamızdaki amaç günlük hayatımızda her an karşılaştığımız akıllı ulaşım sistemleri hakkında bireylerin duygu ve düşüncelerinin hangi yönde olduğunu belirlemektir. Çalışmanın ilk bölümünde giriş kısmına ve literatür taramasına yer verilmiştir. İkinci bölümde ise akıllı ulaşım sistemleri tarihçesi ve uygulama alanları hakkında açıklamalara yer verildi. Bir sonraki bölümde ise analiz aşamasında kullandığımız yöntem olan metin madenciliği hakkında tanımlamalarda bulunuldu. Dördüncü bölümde ise metin madenciliği tekniği olan duygu analizi hakkında bilgi verildi. Daha sonra analiz aşamasına geçerek anahtar kelimeler hakkında atılan tweetler üzerinde duygu analizi tekniği uygulayarak bireylerin konular hakkında duygularının ne yönde olduğu belirlendi. Analiz aşamasını gerçekleştirirken RapidMiner programını kullanarak bireylerin önceden belirlenen konular hakkında attıkları tweetlere ulaşıldı ve bu tweetler üzerinde duygu analizi tekniği uygulandı. Son bölümde ise çıkan sonuçları değerlendirildi ve sonuç olarak bireylerin akıllı ulaşım sistemi uygulamaları üzerinde yüzde 62 oranında olumlu bir görüşe sahip olduğu ve bu uygulamaların günlük hayatımızda daha fazla artırılması yorumunda bulunuldu.

Anahtar kelimeler: Akıllı Ulaşım Sistemleri, Otonom Araçlar, Trafik Yoğunluğu, Metin Madenciliği, Duygu Analizi

ABSTRACT

MEASUREMENT OF INTELLIGENT TRANSPORT SYSTEMS PERCEPTION BY TEXT MINING METHOD WITH SENTIMENT ANALYSIS

Mübin ÖZKUL

In our study, we examined the intelligent transport systems that we frequently encounter in our daily lives and are on the agenda. The aim of our study is to determine the direction of the feelings and thoughts of individuals about intelligent transportation systems that we encounter every moment in our daily lives. In the first part of the study, introduction and literature review are given. In the second part, we gave explanations about the history and application areas of intelligent transportation systems. In the next section, we have identified text mining, which is the method we use in the analysis phase. In the fourth part we gave information about emotion analysis, which is text mining technique. Then we went through the analysis phase and applied the emotion analysis technique on tweets about keywords to determine the direction of the individual's feelings about the subject. When we were performing the analysis phase, we used the RapidMiner program to reach the tweets that individuals have already tweeted on the topics that were identified, and we applied the emotion analysis technique on these tweets. In the last section, we evaluated the results and as a result, we found that 62 percent of individuals have a positive view on Intelligent Transportation System applications and commented on increasing these practices in our daily lives.

Keywords: Intelligent Transportation Systems, Autonomous Vehicles, Traffic intensity, Text Mining, Sentiment Analysis

ÖNSÖZ

Yüksek lisans tez çalışmamda yardımlarını hiçbir zaman esirgemeyen ve beni sabırla karşılayarak çalışmamda ilerlememi sağlayan tez danışmanım Sn. Prof. Dr. Mehmet TEKTAŞ hocama sonsuz teşekkürlerimi sunarım.

Tez çalışmamın her aşamasında yardımlarını benden esirgemeyen Sn. Dr. Öğretim Üyesi Ufuk ÇELİK hocama sonsuz teşekkürlerimi sunarım.

Tez çalışmamda ve hayatımın her anında benden desteklerini hiç esirgemeyen ilk başta annem Nursel ÖZKUL ve babam Ali ÖZKUL olmak üzere tüm ÖZKUL ve EZEN ailesine en içten duygularıyla teşekkürlerimi sunarım.

Ayrıca tez çalışmam süresince beni yalnız bırakmayan başta Aykut DAĞLI olmak üzere maddi manevi emeği geçen bütün arkadaşlarıma teşekkürü bir borç bilirim.

Son olarak, bu tez çalışmasını yaşamı boyunca gülüşü yüzünden hiç eksik olmayan ve benimle her zaman gurur duyan amcam merhum İsmail ÖZKUL' a atfediyorum.

Mübin ÖZKUL

Bandırma Onyedli Eylül Üniversitesi

11/07/2019

İÇİNDEKİLER

YÜKSEK LİSANS TEZ ONAYI.....	iii
ETİK BEYAN.....	iv
ÖZET.....	v
ABSTRACT.....	vi
ÖNSÖZ.....	vii
ŞEKİLLER.....	xii
KISALTMALAR.....	xiii
1. BİRİNCİ BÖLÜM.....	1
1.1 GİRİŞ.....	1
1.2 LİTERATÜR.....	3
1.3 ÇALIŞMANIN KONUSU VE İŞLEYİŞİ.....	6
1.4 ÇALIŞMANIN AMACI.....	6
2. İKİNCİ BÖLÜM.....	7
2.1 AKILLI ULAŞIM SİSTEMLERİ(AUS).....	7
2.2 UYGULAMA ALANLARI.....	8
2.2.1 Trafik Yönetim Sistemleri.....	8
2.2.2 Kavşak Kontrol Cihazı (KKC).....	10
2.2.2.1 Sabit Zamanlı Çalışma.....	10
2.2.2.2 Çoklu Zamanlı Çalışma.....	10
2.2.2.3 Yarı Trafik Uyarmalı Çalışma.....	10
2.2.2.4 Tam Trafik Uyarmalı Çalışma.....	11
2.2.2.5 Arter Bazlı Merkezi Çalışma.....	11
2.2.2.6 Yeşil Dalga Çalışma.....	11
2.2.3 Trafik Kontrol Merkezi.....	12
2.2.3.1 Sürücü ve Yolcu Bilgilendirme.....	13
2.2.3.2 Yol Üstü Bilgilendirme Sistemleri.....	13
2.2.3.3 Web ve mobil uygulamalar.....	13
2.2.3.4 Radyo ve Televizyonlar.....	13
2.2.4 Toplu Ulaşım Sistemleri.....	14
2.2.5 Otopark Yönetim Sistemleri.....	15
2.2.6 Kural İhlali Denetim Sistemleri.....	16

2.3	DÜNYADAKİ BAZI UYGULAMALAR	16
2.4	TÜRKİYE'DE AUS'UN DURUMU	17
3	ÜÇÜNCÜ BÖLÜM.....	18
3.1	METİN MADENCİLİĞİ.....	18
3.2	VERİ MADENCİLİĞİ	19
3.3	METİN MADENCİLİĞİ TANIMI.....	22
3.4	METİN MADENCİLİĞİ İLE İLGİLİ KAVRAMLAR.....	23
3.5	METİN MADENCİLİĞİ ÇALIŞMALARİ	25
3.6	METİN MADENCİLİĞİ ÇALIŞMALARINDA KULLANILAN YÖNTEMLER	26
3.6.1	BİLGİ GETİRİMİ	26
3.6.2	BİLGİ ÇIKARIMI	27
3.6.3	HECE ANALİZİ	27
3.6.4	ETİKETLEME.....	28
3.6.5	ÖRÜNTÜ TANIMA.....	28
3.7	METİN MADENCİLİĞİ ANALİZ AŞAMALARI	28
3.8	METİN MADENCİLİĞİNİN İLGİLİ ALANINA GİREN SORUNLAR	29
3.9	METİN MADENCİLİĞİ KULLANIM ALANLARI.....	30
4	DÖRDÜNCÜ BÖLÜM.....	31
4.1	DUYGU ANALİZİ.....	31
4.2	DUYGU ANALİZİ ÇALIŞMALARİ HANGİ AMAÇLARLA KULLANILIR?.....	32
4.3	DUYGU ANALİZİ KELİME SINIFLANDIRMA AŞAMALARI.....	33
4.4	ÖZİNİTELİK SEÇİMİ.....	34
4.5	DUYGU ANALİZİNDE KULLANILAN YÖNTEMLER	35
4.5.1	MAKİNE ÖĞRENMESİ YÖNTEMLERİ	35
4.5.1.1	Dil Etkileşim Sözlüğü(DAL).....	35
4.5.1.2	Naive Bayes	36
4.5.1.3	k-NN Algoritması	37
4.5.1.3.1	k-NN Algoritması Uzaklık Hesaplama Yöntemleri.....	38
4.5.1.4	Bulanık Mantık(Fuzzy Logic).....	39
4.5.1.5	Destek Vektör Makineleri(SVM)	39
4.5.1.5.1	Doğrusal Destek Vektör Makineleri	39
4.5.1.5.2	Doğrusal Olmayan Destek Vektör Makineleri.....	41

4.5.1.6	Vektör Uzay Modeli (VSM).....	41
4.5.2	MAKİNE ÖĞRENİMİ YÖNTEMLERİNİN BAŞARISI ARTTIRMAYA YÖNELİK UYGULAMALAR.....	43
4.5.2.1	Kelime Öbekleri(Bag of Words).....	43
4.5.2.2	TF-IDF(Term frequency – Inverse document frequency) AĞIRLIKLANDIRMA YÖNTEMİ.....	43
4.5.2.3	POS Tagger (Metin Parçası Etiketleme) Yöntemi.....	44
4.5.2.4	SeedWords(Tohum Kelimeler) Yöntemi.....	45
4.5.3	SÖZLÜK TABANLI YÖNTEM.....	46
4.5.3.1	Öncül Yön Puanlandırma (Prior Polarity).....	46
4.6	SINIFLAMA.....	46
4.7	KÜMELEME.....	47
4.8	İLİŞKİLENDİRME.....	47
4.9	MODEL BAŞARISINI ÖLÇME YÖNTEMLERİ.....	47
4.9.1	Kesinlik Ölçütü.....	47
4.9.2	Duyarlılık Ölçütü.....	48
4.9.3	Doğruluk Ölçütü.....	48
4.9.4	F Ölçütü.....	48
5	BEŞİNCİ BÖLÜM.....	49
5.1	RAPİDMİNER.....	49
5.2	METODOLOJİ.....	50
5.2.1	Verilerin Toplanması.....	50
5.2.2	Veri Toplama Süreci.....	51
5.3	AUS Hakkında Yapılan Yorumların Duygu Analizi Sonuçları.....	54
5.4	Trafik Yoğunluğu Hakkında Paylaşılan Tweetlerin Duygu Analizi Sonuçları.....	57
5.5	Otonom Araçlar Hakkında Paylaşılan Tweetlerin Duygu Analizi Sonuçları.....	60
6	ALTINCI BÖLÜM.....	62
6.1	BULGULAR.....	62
7	SONUÇ, DEĞERLENDİRME VE ÖNERİ.....	67
7.1	SONUÇ.....	68
7.2	DEĞERLENDİRME.....	70
7.3	ÖNERİLER.....	72
	KAYNAKÇA.....	74



ŞEKİLLER

Şekil 1 : Trafik yönetim sistemi şeması	9
Şekil 2 : Kavşak kontrol cihazı	10
Şekil 3 : Yeşil Dalga Uygulaması	12
Şekil 4 : Trafik Kontrol Merkezi.....	12
Şekil 5 : Akıllı Durak ve MOBİETT Uygulaması	14
Şekil 6 : Otopark Yönetim Sistemleri	15
Şekil 7 : Kural İhlali Denetim Sistemleri.....	16
Şekil 8 : Veri madenciliğinde değerli olan veriyi elde etme aşamaları.....	21
Şekil 9 : Metin madenciliği kavramı ilişkili kavramlar	24
Şekil 10 : Metin madenciliği alanında gerçekleştirilen çalışmalar	25
Şekil 11 : Metin madenciliği tekniğinde kullanılan yöntemler.....	26
Şekil 12 : Metin madenciliği tekniğinin çalışma aşamaları	29
Şekil 13 : Duygu analizi kelime sınıflandırma şeması.....	33
Şekil 14 : Duygu analizinde tekniğinde kullanılan yöntemler	35
Şekil 15 : İki boyutlu düzlemde sınıflandırma işlemi	40
Şekil 16 : Vektör uzay modeli düzlem üzerinde gösterimi.....	42
Şekil 17 : Twitter ve AYLIEN operatörleri bağlantı platformları	51
Şekil 18 : Twitter sosyal ağı üzerinde tarama yapılarak elde edilmiş tweet örneği.....	52
Şekil 19 : Duygu analizi ve twitter parametrelerinin özellikleri	52
Şekil 20 : Akıllı ulaşım sistemleri hakkında atılan tweet'lerin duygu analizi sonuçları.....	54
Şekil 21 : Akıllı ulaşım sistemleri duygu analizi sonuç grafiği	55
Şekil 22 : Akıllı ulaşım sistemleri öznellik sonuçları	56
Şekil 23 : Trafik yoğunluğu hakkında atılan tweet'lerin duygu analizi sonuçları	57
Şekil 24 : Trafik yoğunluğu duygu analizi sonuç grafiği.....	58
Şekil 25 : Trafik yoğunluğu öznellik sonuçları.....	59
Şekil 26 : Otonom araçlar hakkında atılan tweet'lerin duygu analizi sonuçları	60
Şekil 27 : Otonom araç duygu analizi sonuç grafiği	61
Şekil 28 : Otonom araç öznellik sonuçları	61
Şekil 29 : Akıllı ulaşım sistemleri Duygu ve Öznellik analizi sonuçları	62
Şekil 30 : Trafik yoğunluğu sistemleri duygu ve öznellik analizi sonuçları.....	63
Şekil 31 : Otonom araç duygu ve öznellik analizi sonuçları.....	64
Şekil 32 : Akıllı Ulaşım Sistemleri 2010-2019 yılları arasında paylaşılan tweet sayıları.....	65
Şekil 33 : Trafik Yoğunluğu 2010-2019 yılları arasında paylaşılan tweet sayıları.....	66
Şekil 34 : Otonom Araç 2010-2019 yılları arasında paylaşılan tweet sayıları.....	67

KISALTMALAR

AUS :Akıllı Ulaşım Sistemleri

SA :Sentiment Analysis(Duygu Analizi)

TF : Terim Frekansı

IDF : Ters Belge Frekansı

VSM : Vektör Uzay Modeli

SVM : Destek Vektör Makineleri

DAL : Dil Etkileşim Sözlüğü

TP :Hem Pozitif Olan Hem de Pozitif Sınıflandırılmış Veriler

TN :Gerçekte Negatif Olan Ve Negatif Sınıflandırılmış Veriler

FP :Gerçekte Pozitif Olan Ancak Pozitif Sınıflandırılmamış Veriler

FN :Gerçekte Negatif Olan Ancak Negatif Sınıflandırılmamış Veriler

KKC : Kavşak Kontrol Cihazı

1. BİRİNCİ BÖLÜM

1.1 GİRİŞ

Geçmişten bu zamana kadar teknolojinin gelişmesiyle birlikte her geçen gün insanların günlük yaşantıları değişmekte, teknolojik gelişmeler ve ürünler bireylerin günlük hayatında vazgeçilmez bir unsur haline gelmektedir. Geçmişten günümüze her geçen gün insanlar için zaman kullanımının önemi artmakta ve bireyler bu geçen zamanı daha tasarruflu kullanabilmek için teknolojik anlamda gelişmelere ihtiyaç duymaktadırlar. Teknolojinin gelişmesiyle birlikte bireylerin zamanı tasarruflu kullanımı her geçen gün artmakta ve bu zaman daha verimli kullanılmaktadır. Yaşanan bu teknolojik gelişmelere paralel olarak bilgisayar ve internet teknolojileri de günlük hayatımıza dahil olmuş ve günlük hayatımızda büyük değişimlerin temeli atılmıştır. Özellikle son yıllarda bilgisayar teknolojilerindeki kullanımının inanılmaz artışı ve buna paralel olarak internetin dünya çapında yaygınlaşması ile birlikte bilgi paylaşımı da aynı oranda artmıştır.

Bilgi kaynaklarının artması ve bilgi paylaşım ağının gelişmesiyle birlikte bilginin depolanması daha önemli hale gelmiştir. Bu da bilgisayar donanımlarının daha fazla geliştirilmesine sebep olmuştur. Bilginin depolanması ile birlikte depolanan bu verilerin gelecek zamanlarda kullanımı önemli hale gelmiştir. Yazılım sektöründeki gelişmeler ile birlikte depolanan veriler üzerinde analizler yapılmaya başlamış ve bireyler bu analizleri kullanarak günlük hayatta ihtiyaç duyulan işlerinde kullanmışlardır. Örneğin, muhasebe alanından örnek verecek olursak; sayısal verilerin depolanması bu alan için büyük önem arz etmekte ve bu depolanan verilerin analiz edilebilmesi için muhasebe otomasyon yazılımları geliştirilmiş ve işlemler daha kolay hale gelmiştir.

İnternetin dünya çapında yaygınlaşması yeni siteler, haber kanalları ve sosyal haberleşme ağlarının ortaya çıkması ile bilgi paylaşımının inanılmaz artışını sağlamıştır. Özellikle günümüzde sosyal haberleşme ağlarının kullanımı insanların hayatının önemli bir parçası haline gelmiştir. İnsanlar, sosyal medya ağları üzerinden kişisel bilgilerini, ilgi alanlarını ve günlük hayatta yaşadığı olayları paylaşmaktadır, bunun sonucunda

sosyal medya kullanımı gün geçtikçe artmaktadır. Sosyal medya kullanımının artması ile farklı kültür ve seviyedeki insanların duygu ve düşünceleri hakkında bilgi edinebilmektedir. Sosyal medya ortamı bu anlamda önemli bir bilgi merkezi olarak görülebilmektedir.

İnternet ve sosyal medya alanında yaşanan gelişmeler üzerine bu ortamlar firmalar için pazarlama faaliyetlerini gerçekleştirebileceği, reklam verebilecekleri yani ticari anlamda kazançlarını arttırabileceği bir ortam haline gelmektedir. Bunun yanı sıra bireylerin yaptıkları paylaşımlar sayesinde firmalar kendileri hakkında olumlu ve olumsuz düşünceleri görebilmekte ve buna göre pazarlama-satış politikalarını geliştirmektedirler. Firmalar veya kurumlar bireylerin duygu ve düşüncelerini ölçebilmek için birebir anket yöntemi uygulamakta fakat bu yöntem hem birebir çalışmadan dolayı hem de alınan anketlerin sisteme el yordamıyla girilmesinden ötürü zaman kaybına sebep olmaktadır. Günümüzde ise anket yöntemi internet ortamında gerçekleşmekte veya firmalar internet sitesi, sosyal medya üzerinde şikâyet ve yorum kısımlarındaki verileri inceleyerek bireylerin firmalar üzerindeki olumlu ve olumsuz düşüncelerini görmekte, bireylerin firmalara olan güven düzeyini ölçebilmektedir.

Tez çalışmamızdaki temel amaç, son yıllarda hayatımızda önemli bir yere sahip olan ulaşımın akıllı uygulamalarını içeren ve çok yaygın olarak kullanılmaya başlayan akıllı ulaşım sistemleri hakkında görüşlerinin ne yönde olduğunu incelemektir. Bu incelemeyi yaparken bireylerin konu hakkında twitter sosyal ağı üzerinden atılan tweetleri veri kaynağımız olarak ele aldık. Sosyal medya ağları içerisinde twitter uygulamasının seçilme sebeplerinden birisi, bu uygulamanın yaygın kullanımınıdır. Diğer bir sebep, twitter sosyal ağının farklı kültür seviyelerinde bulunan bireylerin herhangi bir konu hakkında duygu ve düşüncelerini bu sosyal ağ yardımıyla daha çok paylaşmasından dolayıdır.

1.2 LİTERATÜR

Türkçe Twitter Mesajlarının Duygu Analizi isimli çalışmada ilk önce twitter'dan alınan Türkçe veri seti oluşturulmuştur. Twitter'dan çekilen bu veri seti için bireylerin mesajlarının negatif mi yoksa pozitif mi olduğunu anlamak için öznellik çıkarım modelinin gelişmesi sağlanmış ve bu modelin sınıflandırma yapabilme başarısına etkisi ölçülmüştür. Bu çalışmada k-NN, SVM, Naive Bayes gibi algoritmalar kullanılarak Türkçe tweetler sınıflandırılmıştır. Sonuç olarak Türkçe tweetlerin duygu analizi makine öğrenmesi yöntemleriyle sınıflandırılabilceği doğrulanmıştır.(ÇOBAN, ÖZYER, & TÜMÜKLÜ ÖZYER, Türkçe Twitter Mesajlarının Duygu Analizi, 2015)

Metin Madenciliği Yöntemleri ile Twitter Duygu Analizi isimli projede metin madenciliği ve duygu analizi yöntemleri uygulanarak metnin asıl verilmek istenen duygusunu emoji şeklinde gösterilmesi hedeflenmiştir. Projede 14200 tweet alınmış ve emoji bulunduran veriler kullanılarak duygu analizi yapılabilme olasılığının var olduğu anlaşılmıştır. Proje çerçevesinde değişik makine öğrenme algoritmalar uygulanarak başarının seviyesinin artırılması amaçlanmış ve sonunda ise başarının en iyi olduğu oran %52.09 ile Naive Bayes(Kernel) algoritmasında ulaşılmıştır.(YEŞİLYURT & ŞEKER, 2017)

Çalışmada twitter sosyal ağı üzerinde belli bir kullanıcıya ait paylaşımlar haber, siyaset gibi kategorilere ayrılarak bu metinler üzerine duygu analizi uygulanmıştır. Duygu analizi çalışması gerçekleştirilirken Bayes algoritması kullanılmıştır. Gelecek çalışmalarda ise daha farklı algoritmalar kullanarak bu çalışmaların geliştirilmesi ve daha başarılı sonuçlara ulaşmak amaçlanmıştır. (BAYKARA & GÜRTÜRK, 2017, s. 957)

Sosyal Medya Üzerinde Veri Analizi: Twitter 2017 isimli projede sosyal medya ortamında tartışılan güncel bir konu başlığının verileri toplanmış ve duygu analizi aşamaları gerçekleştirilmiştir. Ardından elde edilen analiz sonuçları görselleştirilmiştir. Sonuç olarak, karar alıcı mevkilerin sergilediği yaklaşımı sosyal medya ortamında oluşan tepkiler şekillendirebilecek ve kitleler sesini üst mercilere çok daha hızlı bir şekilde duyurabileceklerdir.(ALBAYRAK, TOPAL, & ALTINTAŞ, 2017)

Duygu Analizi ve Fikir Madenciliği Algoritmalarının İncelenmesi isimli projede duygu analizi genel anlamda anlatılmaya çalışılmış ve bunun için yapılan projelerde sunulan duygu analizi algoritmaları ve onun uygulamaları sunulmuştur. Güncel bir sonuç ortaya koymak için son zamanlarda yazılmış projeler dikkate alınmıştır. Bu aşamadan sonra projeler kategorize olacak şekilde yapılmış ve kısa biçimde özeti çıkarılmıştır. Buna ek olarak duygu sınıflandırma sorunlarının çözümü için Naive Bayes ve destek vektör makinelerinin kullanımı en fazla olan makine öğrenme algoritmalarının bulunduğu sonucu karşımıza çıkmıştır. Uygulanan analizlerin sonucu olarak İngilizce dili için birçok kaynağın var olduğu ve İngilizceden farklı dillerde kaynağın az olduğu göz önüne çıkmıştır.(CAN & ALATAŞ, 2017)

Türkçe Metinlerde Duygu Analizi isimli projede, İngilizce ve Türkçe dili için gerçekleştirilen çalışmalarda kullanılan Makine öğrenmesi ve sözlük tabanlı duygu analizi yöntemleri farklı özellikler eklenerek oluşturulmuş, iki farklı veri seti üzerinde incelenmiştir. Analiz sonucu değerlendirildiğinde film yorumları veri seti en etkili ikinci model olarak belirlenmiş, fakat toplam duygu düzeyine bir etki sağlayamamıştır. twitter veri seti incelendiğinde ise metnin toplam duygu düzeyinde değişim yapabilecek bir etkinin olduğu sonuca ulaşılmıştır.(TÜRKMENOĞLU, 2015)

Duygu Analizi (Santimantal Analysis) isimli projede duygu analizi kavramını detaylı bir şekilde açıklamaktır. Genel anlamda duygu analizi yönteminde kullanılan terimler, özellik çıkarım teknikleri, duygu analizi alanında karşılaşılan problemler ve büyük boyutlu veri setleri hakkında çözüm önerileri sunulmaya çalışılmıştır. Yazıda büyük veri dünyasında kullanılan Spark isimli program üzerinde çalışan MLlib isimli makine öğrenme kütüphanesinden örnek kodlar gösterilerek kullanılan algoritmalar açıklanmıştır.(ŞEKER S. E., Duygu Analizi(Sentiment Analysis), 2016)

Çalışmada belirlenen anahtar kelimeler ile twitter üzerinden tweetler çekilmiş ve bu tweetler olumlu, olumsuz ve nötr olmak üzere üç kategoride etiketlenmiş. Çalışmada sözlük ve n-gram olarak iki model kullanılmış ve sonuç olarak sözlük yönteminin daha başarılı olduğu kanısına varılmıştır.(AKGÜL, ERTANO, & DİRİ, 2015)

Çalışmada metin sınıflandırma yöntemlerinden olan kelime torbası yöntemi ve sinir ağı temelli yaklaşımlar karşılaştırılmış ve bu sınıflandırma işleminin Türkçe dili

açısından nasıl geliştirilebileceği konusunda tartışılmıştır. Sonuç olarak geleneksel yöntemlerin kelime gömme gibi yeni yaklaşımlarla yarışabileceği kararına varılmıştır.(YILDIRIM & YILDIZ, 2018)

Türkçe film yorumları üzerine duygu analizi gerçekleştirilen bu çalışmada ise duygu analizi çalışmalarında en çok kullanılan makine öğrenmesi ve sözlük tabanlı yaklaşım yöntemleri hakkında bilgi verilmiş ve bu iki yöntem film yorumları üzerinde uygulanmıştır. Sonuç olarak ise makine öğrenmesi yöntemlerinin sözlük tabanlı yöntemle göre daha başarılı bir sonuç verdiğini ve bunun nedeni olarak da sözlük tabanlı yöntemde kullanılan sözlük içerisinde bulunan pozitif ve negatif kelimelerin sinema alanı için uygun olmadığı ve makine öğrenmesi yöntemlerinde kullanılan modellerin bu alana daha uygun olduğu belirtilmiştir. (SEVİNDİ, 2013)

Çalışmada IMDB platformunda yer alan 2000 adet film yorumları alınmış ve bunların üzerine duygu analizi yöntemi uygulanmış. Analiz gerçekleştirilirken makine öğrenmesi yöntemlerinden olan Yapay Sinir Ağları, Destek Vektör Makineleri, Naive Bayes ve Merkez Tabanlı Sınıflayıcı algoritmaları kullanılmış ve bu yöntemler model başarı ölçütlerine göre karşılaştırılmıştır. Bu karşılaştırma sonucunda en iyi sonucu %89.73 ile yapay sinir ağları ve %84.07 ile destek vektör makineleri vererek doğru sınıflandırma oranı en yüksek iki yöntem olarak seçilmiştir. (Kaynar, Yıldız, Albayrak, & Görmez, 2016)

Twitter sosyal ağı üzerinden atılan mesajlara uygulanan bu duygu analizi çalışmada Naive Bayes, Destek Vektör Makineleri gibi sınıflandırma yöntemleri kullanılarak bir sistem oluşturulmuş ve sonuçlar karşılaştırılmıştır. Sonuç olarak %90'a yaklaşık bir başarı oranı gösteren bir sistem oluşturulmuştur. (Meral & Diri, 2014)

Çalışmada duygu analizinde kullanılan yöntemlerin sonuçları karşılaştırılmış ve deney sonuçlarına göre kısa zamanda geri dönüş yapabilecek bir sistem oluşturulmaya çalışılmıştır. Analiz aşamasında veri olarak Türkçe film yorumları ele alınmış ve öznitelik seçme yönteminin destek vektör makinesi yöntemi üzerinde etkisi tespit edilmiştir. Çalışmada veriler sadece pozitif ve negatif olarak sınıflandırıldığında %83.9 pozitif, negatif ve nötr olarak sınıflandırıldığında ise %63.3 başarı elde edilmiştir.

Çalışmada önerilen sistem öznitelik metriklerinin duygu analizinde kullanılabilir bir yöntem olduğunu kanıtlamıştır. (AKBA, 2014)

1.3 ÇALIŞMANIN KONUSU VE İŞLEYİŞİ

Bu tez çalışmasının ilk bölümünde, metin madenciliği tekniği ve metin madenciliğinin uygulama alanlarından olan duygu analizi yöntemi ele alınmıştır. Çalışmanın ikinci bölümünde ise akıllı ulaşım sistemlerinin tanımı, ilgili temel kavramlar ve uygulama alanları hakkında bilgilere yer verilmiştir. Çalışmanın üçüncü bölümünde, veri madenciliği analizlerinde kullanılan sınıflama ve kümeleme gibi terimlerin tanımı yapılmış ve sonrasında model başarısını ölçmede kullanılan ölçütler hakkında bilgilere yer verilmiştir. Ayrıca literatür incelenerek duygu analizi çalışmalarında kullanılan makine öğrenmesi ve sözlük tabanlı yöntem ile ilgili çalışmalara yer verilmiştir. Makine öğrenmesi yöntemleri arasında yer alan Naive Bayes, k-NN, Destek Vektör Makineleri ve Vektör Uzay Modeli gibi algoritmalar ile ilgili tanımlamalara yer verilmiş ve algoritmalarda kullanılan formüller açıklanmıştır. Analiz aşamasında ilk olarak twitter üzerinden belirlenen konularda tweetlerin çekilebilmesi için “Twitter Application” anahtarları alınarak tweetler çekilmeye başlanmıştır. RapidMiner analiz programı kullanılarak çekilen tweetler ilk önce veri temizleme işlemine tabi tutularak @, # (hashtag), http gibi simgeler tweetlerin içerisinden arındırılmış ve ardından duygu analizi gerçekleştirilmiştir. Duygu analizi sonucunda polarity(duygu düzeyi) ve subjectivity(öznellik) katsayıları gerekli formüller uygulanarak hesaplanmıştır. Ardından çıkan sonuçlar yüzdelik olarak belirlenip grafik ortamına aktarılarak yorumlanmıştır.

1.4 ÇALIŞMANIN AMACI

Tez çalışmamızdaki amaç, akıllı ulaşım sistemlerinin günlük hayatımızda kullanılan uygulamaları hakkında bireylerin görüşlerinin ve duygu düzeylerinin ne yönde olduğunu belirlemektir. Bu amacı gerçekleştirebilmek için, analiz aşamasında veri madenciliğinin bir alt seviyesi olan ve metinler üzerinden istatistiksel çıkarımlarda bulunan metin madenciliği yöntemi kullanılmıştır. Bu yöntem gerçekleştirilirken duygu analizi tekniğinden faydalanılmıştır. Analiz aşamasında ilk olarak, önceden belirlediğimiz anahtar kelimeleri RapidMiner programı yardımıyla twitter sosyal

ağından belirlenen konular hakkında tweetler çekilmiştir. Ardından RapidMiner programının içerisine eklenen “AYLIEN Text Analysis” paket programı yardımıyla konuların duygu düzeyleri belirlenmiştir. Sonuç olarak, akıllı ulaşım sistemleri ile ilgili paylaşılan yorumların %62’sinin pozitif yönde olduğu ve bireylerin olumlu bir yaklaşım sergilediği görülmüştür. Ayrıca, trafik yoğunluğu hakkında atılan tweetlerin %93’ünün de olumsuz yönde bir yaklaşım sergilediği görülmüştür. Akıllı ulaştırma sistemleri uygulamalarının yaygınlaştırılmasının trafik yoğunluğu, trafik güvenliği gibi sorunları azaltarak bireylerin memnuniyet düzeylerinin arttıracacağı öngörüsüne ulaşılmıştır.

2 İKİNCİ BÖLÜM

2.1 AKILLI ULAŞIM SİSTEMLERİ(AUS)

Günümüzde teknolojinin de gelişmesi ile birlikte bireylerin yaşamını kolaylaştıracak ve zaman tasarrufu sağlayacak birçok gelişme olmuştur. Geçmiş göz önüne aldığımızda bireyler ulaşmak istedikleri noktalara daha kısa sürede varmak için çeşitli çabalarda bulunmuşlardır. Bunun sonucunda bilim ve teknolojiye yaşanan gelişmelerle birlikte hava, kara ve deniz yolu olarak farklı ulaşım seçenekleri oluşmuştur. Gün geçtikçe nüfusun artmasıyla birlikte şehirlerdeki nüfus yoğunluğu artmakta ve bunun sonucunda ulaşım ağları yetersiz kalmaktadır. Artan nüfus ve gelişen teknoloji ile birlikte özellikle Avrupa’da karayolu ağlarında büyük gelişim sağlanmıştır. Karayolu ağlarında yaşanan bu gelişme otomobil sektörünün de doğuşuna temel hazırlamıştır. Özellikle otomobil sektöründe yaşanan gelişmeler sayesinde bireyler ulaşmak istediklere noktalara daha kısa sürede ulaşmakta bu durumda otomobile olan talebi arttırmaktadır. Geçmişten günümüze nüfusunda artmasıyla birlikte karayolu ulaşımına olan talep artmakta ve bunun sonucunda trafik yoğunluğu, kazalar ve emisyon artışı gibi problemler ortaya çıkmaktadır. Ortaya çıkan bu problemlerin gün geçtikçe günlük hayatımızı daha fazla zorlaştırması ile birlikte ulaşımında yaşanan problemlerin çözülebilmesi için belli başlı çalışmalar yapılmıştır. Bilgi ve teknolojinin ulaşım sektörüne senkronize edilmesiyle birlikte “Akıllı Ulaşım Sistemleri(AUS)” uygulamalarını ortaya çıkarmıştır.

Tanımsal olarak AUS; trafik güvenliğinin artırılması, yolculuk sürelerinin kısaltılması ve enerji verimliliği sağlanarak hem çevreye hem de ülke ekonomisine katkı

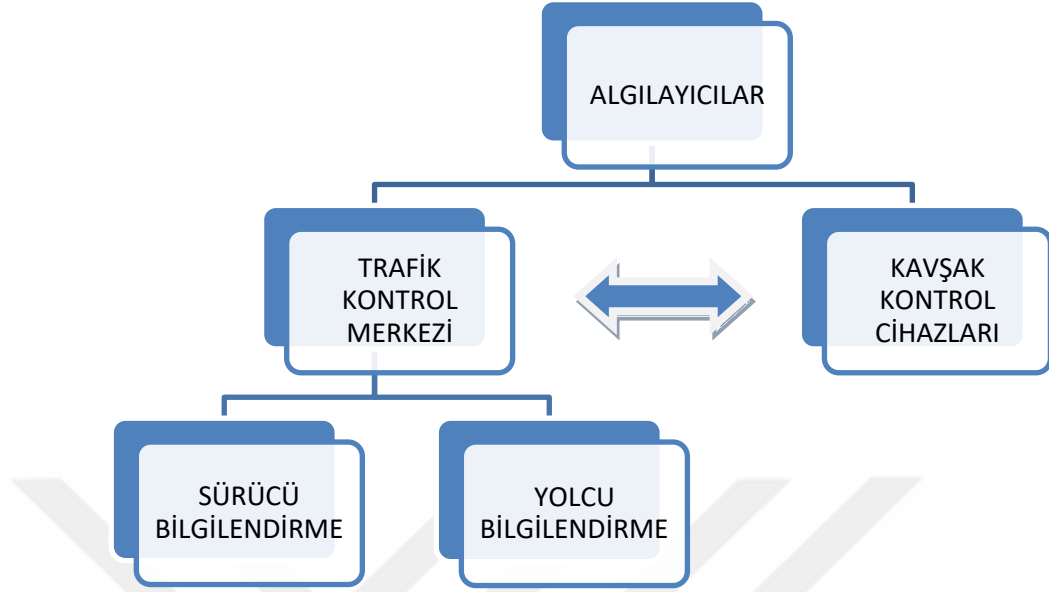
sağlamak gibi amaçlarla geliştirilen kullanıcı-araç-altyapı-merkez arasında çok yönlü veri alışverişi ile izleme, ölçme, analiz ve kontrol içeren sistemlerdir. (Tektaş, Korkmaz, & Erdal, 2016) Akıllı Ulaşım Sistemleri ilk olarak 1960'lı yıllarda ortaya çıkmış, 1980'li yıllarda internet ve bilgisayar teknolojilerinin gelişmesiyle birlikte yaygınlaşmıştır. AUS uygulamaları, 1990'lı yıllarda tüm dünyada rağbet görmüş ve bu konuda bilgi paylaşımı ve iş birliği sağlayabilmek amacıyla akıllı ulaşım sistemi birlikleri kurulmuştur. Kurulan bu birlikleri kısaca inceleyecek olursak, ilk olarak 1990 yılında Avrupa'da akıllı ulaşım sistemi standartlarının belirleme ve yapılacak çalışmalara kaynak sağlamak amacıyla ERTICO-ITS Europe kurulmuştur. Yine aynı yılda Amerika'da ITS America kurulmuştur. Ülkemizde ise 1992 yılında ISO TC204 AUS teknik komitesi kurularak AUS'nin dünya çapındaki standartlarını belirlemek için çalışmalar başlamıştır. 1994 yılında AUS Dünya Kongresi toplantıları başlamış aynı yıl VERTIS adında Japonya AUS birliği kurulmuş fakat 2001 yılında ismi ITS Japan olarak değişmiştir. Günümüzde AUS Dünya Kongreleri ERTICO, ITSA VE ITS Japan birliklerinin işbirliği ile düzenlenmektedir.

Genel anlamda her geçen yıl nüfusun artışı ile doğrudan artış gösteren araç sahipliliği şehir içerisindeki trafiği olumsuz etkilemektedir ki büyük şehirlerin en önemli sorunlarından biri trafik problemidir. Türkiye'de trafikte dolaşan araç sayısı son 15 yıllık süreçte üç katına çıkmıştır. Türkiye'de 1994 yılında sadece 4.584.717 araç varken 2019 yılında trafiğe kayıtlı araç sayısı 22.922.164'e ulaşmıştır ki bu sayının sadece 4 milyon 170 bini aşkını İstanbul'da bulunmaktadır ki bu bile metropol olan şehrimizdeki trafiği göz önüne sermektedir.(Türkiye İstatistik Kurumu İl Göstergeleri) Trafikteki akışkanlık günümüzdeki önemli sorunlardandır, bu sorunların giderilmesi için akıllı ulaşım sistemlerine ihtiyaç duyulmaktadır.

2.2 UYGULAMA ALANLARI

2.2.1 Trafik Yönetim Sistemleri

Trafik yönetim sisteminin temelinde aşağıdaki aşamalar yer almaktadır;



Şekil 1: Trafik yönetim sistemi şeması

Sistemin asıl amacı, şehir içindeki trafik sorununu gidermek ve trafiğin akışını düzenlemektir. Trafik yönetim sisteminde algılayıcılara gelen bilgi kontrol cihazları ve trafik kontrol merkezine aktarılır ve kontrol merkezi de gelen bu bilgiyi sürücü ve yolculara iletmektedir.

2.2.2 Kavşak Kontrol Cihazı (KKC)

Kavşak kontrol cihazları algılayıcıları kodlayarak trafik ışıklarının kontrolünü gerçekleştirir ve bu sayede trafik akışının düzenli bir şekilde seyretmesini sağlamaktadır.



Şekil 2: Kavşak kontrol cihazı

Kavşak kontrol cihazları farklı modlarda çalışabilmektedir, bu 6 farklı mod hakkındaki tanımlamalar aşağıda kısaca açıklanmaktadır.

2.2.2.1 Sabit Zamanlı Çalışma

Trafik yoğunluğunun daha az olduğu yerlerde maliyetinin ucuz oluşu nedeniyle kullanılan yöntemdir ancak trafiğin yoğun olduğu yerlerde ve genel olarak tercih edilmez.

2.2.2.2 Çoklu Zamanlı Çalışma

En çok kullanılan kavşak kontrol cihazıdır. Semtte pazarın kurulması, okul çıkış saatleri vb. alanlarda kullanıldığı gibi herhangi bir algılayıcıya ihtiyaç duymaz.

2.2.2.3 Yarı Trafik Uyarmalı Çalışma

Genel anlamda T tipi yollarda kullanılan bir kavşak kontrol cihazıdır. Asıl amaç ana yoldaki trafik akışını sağlamaktır ancak yan yollardan yaya ve araç gelişi olmadığı sürece.

2.2.2.4 Tam Trafik Uyarmalı Çalışma

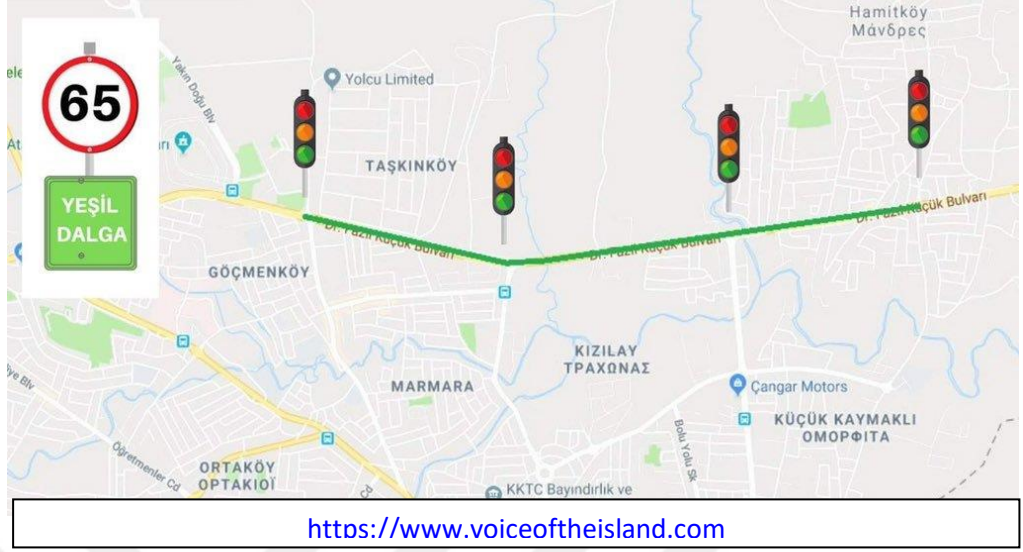
Kavşaktaki anlık durumlar için geçiş sürelerinin belirlenmesine tam zamanlı trafik uyarmalı çalışma denir. Çalışma prensibinde iki çeşit yöntem vardır, bunlar; basit tam zamanlı trafik uyarmalı ve kompleks tam zamanlı trafik uyarmalı yöntemleridir. Basit yöntemde, her kavşaktaki yol bitiminin belli bir mesafe gerisine loplar yerleştirilmektedir. Kavşaktaki tüm yönlerde yeşil zamanları için artırım, minimum, maksimum sabit süreleri vardır. Bir yöndeki yeşil yanma sırası geldiğinde, yanan yeşil lamba en kısa süreye kadar yanmaktadır. Eğer algılayıcı tarafından herhangi bir bildirim gelmiyorsa artırım sabiti uygulanır ve bunun sonucunda maksimum süreye uzamaktadır. Ancak bu durum olmaz ve algılayıcı bildirim gönderirse yeşil yanma süresi minimuma çekilerek kırmızı ışığa döner ve trafiği boşu boşuna beklenmesi engellenir. Kompleks çalışmada ise, her yön için birden fazla lop var olmakta veya radar, video algılayıcı vb. gibi araç sayısını, hızları ölçebilecek algılayıcılar yerleştirilmektedir. Bu sayede kavşağın kullanımını daha etkin olacaktır.

2.2.2.5 Arter Bazlı Merkezi Çalışma

Herhangi bir arterde bulunan kavşaklardaki ani yoğunluk belirtisi önce kavşak kontrol cihazı daha sonra da merkez aracılığıyla algılanır ve bu sayede arterdeki yoğunluk üzerine kavşaklar planlı bir şekilde çalışmaktadır.

2.2.2.6 Yeşil Dalga Çalışma

Bu yöntem diğer yöntemlere ek bir özelliktir. Kavşaklardaki yeşil yanma zamanını, arter araç hızına göre kontrolüdür. Buradaki amaç ise aracın kırmızı ışığa denk gelmeden belirli bir hız ile arterde ilerlemesidir.



Şekil 3: Yeşil Dalga Uygulaması

2.2.3 Trafik Kontrol Merkezi



Şekil 4: Trafik Kontrol Merkezi

Trafik kontrol merkezi ulaşım anında trafik ile ilgili verileri bir kontrol merkezinde toplayarak sürücü ve yolculara bilgilendirilmenin yapıldığı bir ortamdır. Trafik kontrol merkezinin özellikleri aşağıdaki şekilde sıralanabilir;

- Trafikte olan yoğunluğun saniyesi saniyesine almak
- Şehir trafiğini izlemek
- Trafikteki araç sahiplerinin anlık değişimlerden haberdar olması
- Sinyalize kavşakların yönetimi

- Trafiğin ve yolların yoğunluk durumlarının internet veya telefon ile bilgilendirilmesi

Kentlerin yapı değişikliği, trafik kontrol merkezinin değişiklik göstererek Ulaşım Kontrol Merkezi olmasına sebep olmaktadır.

Ulaşım Kontrol Merkezleri, Trafik Kontrol Merkezlerinden farklı olarak metropollerde gerçekleşen ulaşımdaki her bir aktörün kontrolünün sağlandığı merkezler olacaktır. Bunun sonucunda araçların yolculuk süreleri ve yol haritaları seçenekleri kişiler için anlık ulaşım verileri üretilebilecektir. Pek çok ülke bu konu üzerine plan ve uygulama yapmaktadır.

2.2.3.1 Sürücü ve Yolcu Bilgilendirme

Sürücü ve yolcuları hava, yol ve trafik durumlarının hakkında bilgi edinebilmeleri için şu unsurlar kullanılmaktadır.

2.2.3.2 Yol Üstü Bilgilendirme Sistemleri

Ortana ve Intetra gibi şirketlerin başı çektikleri bu sistemde yol üzerinde sürücüleri bilgilendirmek amacıyla kurulmuş sistemlerdir.

2.2.3.3 Web ve mobil uygulamalar

Gelecekte daha yaygın olacağı düşünülen bu sistem sürücü ve yolcu bilgilendirmeye yaramaktadır.

2.2.3.4 Radyo ve Televizyonlar

Günümüzde aktif olarak radyo ve televizyonlardan trafik durum bilgisi tam zamanlı veya kısmi zamanlı olarak yayınlanmaktadır.

2.2.4 Toplu Ulaşım Sistemleri



Şekil 5: Akıllı Durak ve MOBİETT Uygulaması

Akıllı Ulaşım Sisteminin temel amaçlarından biri ise zaman ve enerjinin verimsiz kullanımının önüne geçmektir. Bu sebeple toplu taşımanın geliştirilmesi önemli öncelik haline gelmiştir. Yolcu bilgilendirme sistemlerinin amacı budur. Bugün ileri yolcu bilgilendirme sistemlerinin de içerdiği sıradaki toplu taşıma aracının ne zaman geleceği yolculara aktaran “NextBus” gibi sistemler akıllı duraklar olarak bulunur.

Elektronik bilet kavramı toplu taşımada öne çıkan diğer bir konudur. Birçok ülkede farklı amaçlarla ve çeşitli teknolojiler kullanıldığının örnekleri bulunur. Gelişmiş ülkelerde ulaşım hizmetleri genelde özel sektörün kontrolü altında olduğundan, ulaşım için gerekli temas gerektirmeyen akıllı kartlar genelde bir bölge ve bir şirket üzerinde toplanmaktadır. Ancak bu genel durum dışında Londra’da her ulaşım için tek bir kart ile tüm ulaşımı sağlanabilmektedir. Hong Kong’da hem ulaşım için hem de alışverişte de geçerli olan bir kart sistemi bulunmaktadır.

2.2.5 Otopark Yönetim Sistemleri



Şekil 6: Otopark Yönetim Sistemleri

Her yıl 1 milyon satılan araç için, 10 milyon m² yeni otoparka gereksinim duyulmasına rağmen, ihtiyaç duyulanın alanın çok azı inşa edilebiliyor. İstanbul'da hemen hemen bir otopark için sekiz araç düşmektedir.

Teknolojinin gelişiyle paralel olarak otopark sistemleri tam otomatik hale dönüşmektedirler. Bu sistemler içinde pek çok teknik unsur kullanılmakta ve tamamen makine kontrollü otoparklar kurulmaktadır. Teknoloji ile birlikte otopark sistemleri tam otomatik duruma dönüşmektedir.

2.2.6 Kural İhlali Denetim Sistemleri



<https://www.tgrthaber.com.tr>

Şekil 7: Kural İhlali Denetim Sistemleri

Şehir içi ve dışındaki trafik kurallarının ihlal tespiti için kurulan sistemler şunlardır:

- Otobüs yolu ihlalleri
- Hız koridoru ihlalleri
- Ters yön ihlalleri
- Kırmızı ışık ihlalleri
- Emniyet şeridi ihlalleri
- Yaya yolu ihlalleri
- Park ihlalleri için sistemler geliştirilmiştir.

Hem ülkemizde hem de dünyada bu tür sistemler yaygınlaşmaktadır. Özellikle batı ülkelerinde ceza kesmek yerine kaza önleme mantığı ile elektronik denetim sistemleri konusunda yeni teknolojiler araştırılıp üretilmektedir.

2.3 DÜNYADAKİ BAZI UYGULAMALAR

ABD'den 12 kat daha fazla nüfus yoğunluğu olan Japonya'da ulaşım problemleri için AUS'nin geliştirilmesi konusunda önemli adımlar atılmıştır. 1995 yılında bu proje için 670 milyon dolar bütçe ayrılmıştır. 1996 yılından sonra ise 40 farklı araç üzerinde araç

içi navigasyon 25 şirket tarafından pazarlanmaya başlanmıştır. Bir milyondan fazla satışın yapıldığı bilinmektedir. Aynı zamanda çalışmaların Avrupa ve Amerika’da da sürmesi ve bu rekabetin bulunması AUS için çok önemli bir gelişme olmasını sağlamıştır.(YARDIM & AKYILDIZ, 2005, s. 411)

2.4 TÜRKİYE’DE AUS’UN DURUMU

Türkiye için ilk adımların 1980’li yıllarda yapılan Gebze, İzmit Ekspresyolu gösterilebilirken somut olarak 1992 yılında uygulamaya başlanan “Otoyol Ücret Toplama Sistemi”(OGS) ile başladığını söylemek gerekir. Sistemde araçların aldıkları mesafe ve sınıflarına göre ücretlendirilmesi için geliştirilmiştir.(YARDIM & AKYILDIZ, 2005, s. 410-411)

Son yıllardaki ani nüfus ve araç sahipliliğinin artışı sonucunda özellikle büyük kentlerde bir trafik sorunu meydana gelmiştir. Bu sorunun giderilmesi için yapılan uygulamalardan biri 1999 yılında Fatih Sultan Mehmet köprüsünde “Otoyol Ücret Toplama” (OGS) uygulamasıyla zaman kaybını önlemek ve hızlı geçişi sağlamak amaçlanmıştır. Otoyol Geçiş Sisteminin yanında, buna ek olarak “Kartlı Geçiş Sistemi” (KGS) yerini almıştır. Bu sistemlerde kart okuyucular sayesinde gişe memuruna ihtiyaç duymadan ve araç kullanıcıları zaman kaybetmeden yollarına devam edebilmektedirler.(YARDIM & AKYILDIZ, 2005, s. 411)

Son yıllarda özel sektör bu konuya ilgi göstermektedir. Bu konuda 1999 yılında Bolu Dağı Trafik Bilgi Sistemi Projesi, 2000 yılında Ankara Trafik Bilgi Sistemi Projesi ve Aydın-İzmir Otoyolu Selatin Tüneli Projesi, 2001 yılında ise İstanbul Trafik Bilgi Sistemi Projesi uygulanmaya başlanmıştır.

İstanbul Büyükşehir Belediyesi 1995 yılında geçiş sistemini daha verimli hale getirmek amacı ile yolcu sayısını ve dağılımını verilere dayandırmak amaçlı Elektronik Bilet Sistemi olan “Akbil “ kullanılmaya başlanmış ve günümüzde gelişmeye devam etmektedir.(YARDIM & AKYILDIZ, 2005, s. 411-412)

3 ÜÇÜNCÜ BÖLÜM

3.1 METİN MADENCİLİĞİ

Geçmişten günümüze kadar geçen zamanı incelediğimizde insanoğlunun farklı arayışları, günlük hayatını kolaylaştırmakla birlikte zamanı daha tasarruflu kullanma talepleri bilimsel çalışmalara zemin hazırlamıştır. Yaşanan bu gelişmeler üzerine bilim insanların yeni teoriler ortaya atması teknolojik gelişmeleri de hızlandırmış ve bu durum günümüze kadar devam etmiştir. Teknolojik gelişmelerle birlikte internet ve bilgisayar kavramları ortaya çıkmış ve bilgi kavramının önemi daha da benimsenmiştir. İnternetin gelişimi ile birlikte bilgi üretimi daha fazla artmış ve bunun sonucunda bilgiye ulaşım daha da kolaylaşmıştır. Geçmişten günümüze internet ortamındaki veri sayısı yani bilgi kapasitesi genişledikçe, gelecek dönemler için bilginin depolanması daha önemli bir hal almakta ve bu duruma çözüm olarak bilgisayar donanım ve yazılımlarında da gelişme göstermektedir.

İnternet ağındaki bilgi kapasitesinin geçmişten günümüze bu kadar gelişim göstermesi ile birlikte internet siteleri, sosyal medya gibi kavramları ortaya çıkartmıştır. İnternet ortamında bu gibi gelişmelerin ortaya çıkmasından sonra bireyler internet ortamında daha fazla zaman geçirmekte ve interneti hayatının önemli bir parçası haline getirmektedir. Özellikle günümüzde sosyal medya uygulamaları bireyler tarafından çok talep görmekte ve bireyler, herhangi bir konu hakkında duygu ve düşüncelerini bu sosyal medya uygulamaları üzerinden paylaşabilmektedir. Sosyal medyanın günlük hayatta bu kadar yer alması ve bilginin bu ortamlar üzerinden hızlı bir şekilde yayılabilmesi ticari anlamda firmaların da dikkatini çekmektedir. Firmalar günümüzde sosyal medya uygulamaları üzerinde reklam çalışmaları yaparak satışlarını arttırabilmektedir. Sosyal medyanın firmalar için dikkat çeken başka bir özelliği de müşterilerin şikâyet ve yorumlarını bu uygulamalar sayesinde takip edebilmekte ve müşterilerin firmalar üzerindeki memnuniyet düzeylerini de belirleyebilmektedir. Bu gelişmeler üzerine internet ve sosyal medya uygulamaları hem bireyler hem de ticari kuruluşlar için önem arz etmektedir.

İnternet ve sosyal medyanın bu denli gelişim göstermesi veri madenciliği, metin madenciliği gibi kavramların ortaya çıkmasına yol açmıştır. İnternet ve sosyal medya

ortamında bilgi kaynakları genellikle yazılı kaynaklardır. Örneğin, herhangi bir ürün hakkında sosyal medya veya firmaya ait site üzerinden atılan yorumlar, internet ortamında oluşturulmuş siteler içerisindeki metinler, bunların hepsi yapılandırılmamış veri kaynaklarıdır. Geçmiş dönemlerde yani metin madenciliği kavramı ortaya çıkmadan önce gerçekleştirilen analizler, yapısal veriler yani sayısal veriler üzerinden gerçekleştirilirdi. Veri madenciliği alanı da bu dönemlerde ve günümüzde yapılandırılmış veriler ile çalışmalarını gerçekleştirmektedir. Geçmiş dönemlerde internet ve sosyal medya ortamında yapılandırılmamış kaynaklar yani metinler üzerinde analiz işlemi gerçekleştirilemezken bu soruna çözüm olarak veri madenciliğinin bir alt kolu olarak kabul gören metin madenciliği kavramı oluşturulmuştur.

Metin madenciliği kavramının temeli, 1980'li yıllara dayanmaktadır fakat 2000'li yıllardan itibaren bu konuda yapılan çalışmalar yaygınlaşmıştır.(MELEK, 2012, s. 28-29) Metin madenciliği, müşteri yorumları, web sayfaları gibi metin kaynaklarını girdi olarak kullanarak bu metinler üzerinden kullanışlı ve anlamlı bilgiyi keşfetmeyi amaçlar. Günümüzde internet kullanımının yaygınlaşmasıyla insanların sosyal mecralarda olan paylaşımları oldukça artmakta ve metin madenciliği için internet ulaşılması çok kolay bir veri kaynağı olmaktadır. Bu da internet dünyasını metin madenciliğinin merkezine almasına neden olmaktadır. Dönem şartlarından kaynaklı bir ürünün reklamının en iyi yapıldığı yerin internet olması ve internetin maliyet açısından daha uygun olmasından dolayı sanal ticarete de oldukça kullanıldığı görülmektedir.

Metin madenciliği kavramı, veri madenciliği alanının bir alt dalı olarak görülmektedir. Metin Madenciliği kavramına değinmeden önce veri madenciliği alanı hakkında kısa bir bilgilendirme yapalım.

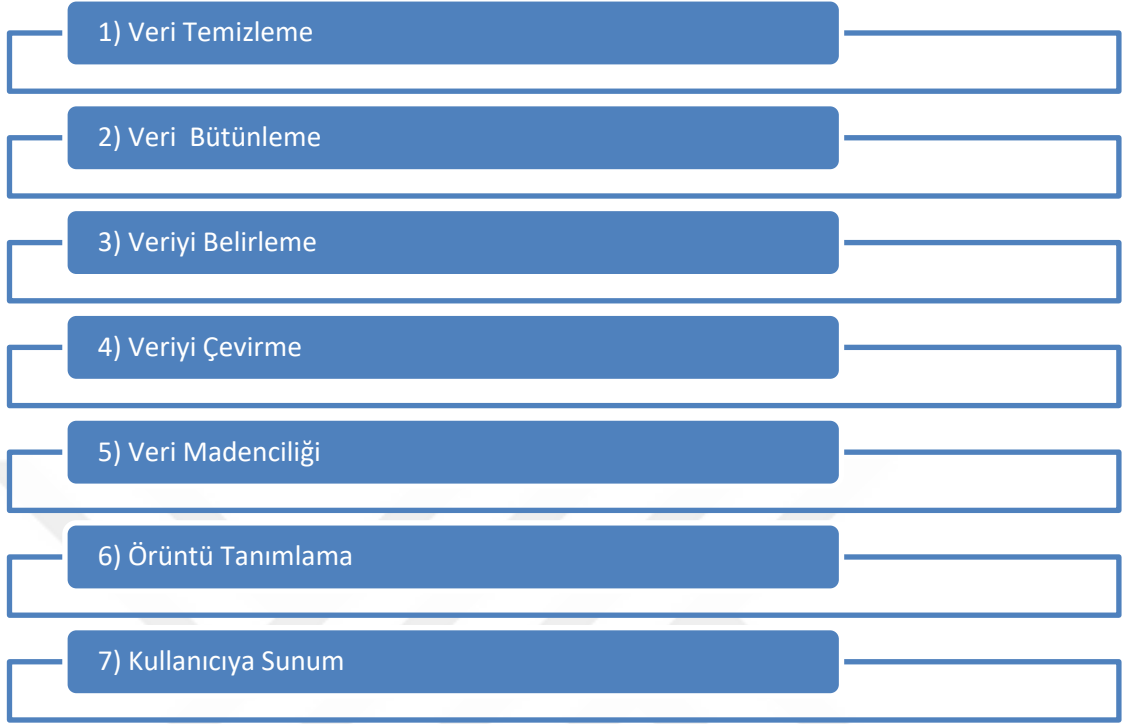
3.2 VERİ MADENCİLİĞİ

Veri madenciliği kavramını ilk bakışta çıkan haberlere ve yüzeysel olarak yapılan değerlendirmelere göre yakın dönemde ortaya çıkmış bir kavram olarak düşünülmektedir. Aslında veri madenciliği kavramı, geçmiş dönemlerde teknolojinin de gelişimiyle birlikte bilginin üretimi ve sonucunda veri setlerinin ortaya çıkmasına kadar dayanmaktadır. Bilgi üretiminin artması ve internet ortamı sayesinde bilgi paylaşımının

da artmasıyla birlikte bazı bilimsel teoriler ortaya çıkmaya başlamıştır. Bazı kaynaklara göre veri madenciliğinin temeli 1763 yılında ortaya atılan Bayes teoremine dayandığı söylenmektedir. Veri madenciliği kavramını biraz daha yakın tarihte incelediğimizde 1960'lı yıllarda bilgisayar ortamında veri taramalarında kullanılmakta fakat veri madenciliği yerine veri taraması olarak anılmaktadır. 1989 yılında KDD Türkçesi veri tabanlarında bilgiyi keşfetmek anlamına gelen uygulama ile bu kavram daha fazla tanınmaktaydı. Ardından 1990'lı yıllardan sonra veri madenciliği kavrama istatistik alanıyla da ortak hareket ederek makine öğrenmesi yöntemlerini de içinde barındırarak günümüze kadar gelmektedir.(EXASTAX BLOG, 2018)

Literatürde farklı tanımlamalar incelendiğinde tanımsal olarak veri madenciliği büyük boyutlu veri setleri içerisinde bizim için değerli olan gizli veriyi keşfetme işlemi olarak yorumlanabilir. (KOYUCUGİL & ÖZGÜLBAŞ, 2009, s. 24)Veri madenciliği bu işlemi gerçekleştirirken bazı istatistiksel yöntemler, matematiksel formüller ve yapay zeka alanından faydalanmakta ve bu alanlarda kullanılan algoritmaları içerisinde barındırmaktadır.

Veri madenciliği büyük veri setleri içerisinde değerli olan veriyi elde ederken aşağıdaki şekilde ifade edilen aşamalar gerçekleştirilmektedir. (Şekil 8)



Şekil 8: Veri madenciliğinde değerli olan veriyi elde etme aşamaları

İlk aşamada büyük çaplı veri seti içerisinde kullanışlı olmayan veriler temizlenmektedir. İkinci aşamada, diğer veri kaynakları birleştirilir ve sonrasında da analiz için gerekli olan veri belirlenmektedir. Sonrasında seçilen veriler, veri madenciliği yöntemlerinin uygulanabilecek hale getirilerek dönüştürme işlemi uygulanır ve sonraki aşamada ise veri madenciliği yöntemi uygulanır. Bundan sonraki aşama, analiz edilen verideki örüntüler tanımlanarak elde edilen sonuç kullanıcıya ulaştırılır.(IŞIKLI, 2009)

Veri madenciliği, verinin özellikle büyük veri olan her uygulamada yaygın olarak kullanılan bir konudur. Örneğin, astronomi, psikoloji, istatistik, yapay zeka vb. alanlarındaki bilimsel çalışmalarda, ulaşım sistemleri uygulamalarında(EDS, HGS, OGS, Trafik yoğunluğu, Trafik yönetimi vb.) sektörler gere göre incelendiğinde reklamcılık, pazarlama, finans ve sigorta vb. sektörlerde çok sayıda uygulamaları görülmektedir.

Veri madenciliği alanında uygulanan model oluşturma yöntemlerini incelediğimizde sınıflama, kümeleme ve ilişkilendirme gibi 3 grupta tanımlandığı

görülmektedir.(MELEK, 2012, s. 12) Bu model oluşturma yöntemlerinin metin madenciliği tekniğinde duygu analizi gibi yöntemlerde de kullanıldığını bilmekteyiz. Sınıflandırma işlemi, sisteme dahil olan verinin mevcut diğer verilerin özellikleri ile karşılaştırılarak daha önceden belirlenen sınıflara atanmasına sınıflandırma işlemi adı verilmektedir. Sınıflandırma işlemi, denetimli ve denetimsiz olarak iki gruba ayrılmaktadır. Denetimli sınıflandırma, mevcut verilere ve sınıflara ait özelliklerin bilinerek sınıflandırma işleminin gerçekleşmesi anlamına gelmektedir. Denetimli sınıflandırma işlemi uygulanırken makine öğrenmesi algoritmalarından faydalanılmaktadır. Denetimsiz sınıflandırma ise sınıflandırma işlemi uygulanmamış verinin yapısını tespit etmek için kullanılmaktadır. Kümeleme yöntemi ise veriye ve sınıfa ait özelliklerin bilinmediği durumlarda uygulanır yani gözetimsiz sınıflandırma işlemi olarak değerlendirilir. Kümeleme yöntemi, benzer özellikte olan verilerin belirli uzaklık hesaplama yöntemleri kullanılarak bir araya getirilip kümelere ayrılma işlemi olarak tanımlanmaktadır. Uzaklık hesaplama yöntemlerine gaussian, euclidyen, Manhattan ve Minkowski gibi örnekler verilebilir. İlişkilendirme yönteminde de veri ve sınıflara ait özellikler bilinmektedir. İlişkilendirme yöntemi ise özellikleri belirlenmiş veri ve sınıflar ile test ve eğitim veri seti grupları arasında ilişkilendirme işlemleri uygulanarak sınıflandırılmasına ilişkilendirme yöntemi adı verilmektedir.(AKBA, 2014, s. 9-10)

3.3 METİN MADENCİLİĞİ TANIMI

Metin madenciliği, rastgele bir yazılı kaynağı veri kaynağı gibi düşünerek bu metinler üzerinden yapılandırılmış veri elde etme sürecidir. Oxford sözlüğü' ne göre metin analizi; büyük boyutlu yazılı kaynakların, özellikle programlama dilleri kullanılarak yeni bir bilgi üretmek amacıyla incelenmesi veya uygulama yapılması olarak tanımlar.(TETİK, 2017, s. 3-4)Örneğin, metin üzerinde duygu analizi, metnin özetlenmesi ve metnin yazarının belirlenmesi vb. gibi çalışmalar yapılabilir. Metin madenciliği kavramı veri madenciliğinin alt kolu olarak düşünülebilir ve bu iki kavramı birbirinden ayıran özellik ise veri madenciliğinde kalıplar düzgün veri tabanlarından çıkartılırken metin madenciliğinde ise doğal dil metinlerinden çıkarılmasıdır. (ŞEKER S. E., sadievrenseker.com, 2014)Yani köşe yazıları, kitap, makale veya internet üzerindeki web siteleri olabilir. Bu yazıların, metin madenciliği açısından önemli bir de

üst bilgileri olması söz konusudur. Örneğin, yazının tarihi, yazının yayınlandığı web sitesi, yazar bilgisi gibi, yazının içinde bulunmayan ama yazıyla ilgili metin madenciliğinde kullanılabilir önemli üst bilgiler bulunabilir. Özellik çıkarımı aşamasında, metinlerin doğrudan içeriğinden veya üst bilgilerinden istenilen özellikler elde edilebilir. Bu bir yazarın hangi kelimeyi hangi sıklıkta kullandığından çıkarılmaktadır.

Metin madenciliğinin çerçevesi, dağınık diyebileceğimiz metin belgelerinden sayısal veriler elde ederek bu verilere gerekli aşamaları uygulayarak anlamlı bilgiye ulaşmak esasına dayanır. Bu çerçeveyi 3(üç) aşamada ele alabiliriz.

- İlk aşamada aynı konudaki belgeler belirlenir, bulunur.
- İkinci aşamada birbirine yakın anlamdaki, birbiriyle ilişkisi bulunan belgeler belirlenir.
- Üçüncü ve son aşamada bulunan belgeler sıralanarak sistematik bir düzen oluşturulur.

Metin madenciliği çalışmaları özellikle reklamcılık, müşteri profili inceleme ve sosyal ağ gibi alanlarında kullanılmaktadır. Metin Madenciliği aşağıdaki sorunlar ile ilgilenmektedir; Enformasyon Getirimi, Adlandırılmış Varlık Tanıma, Örüntüsü Tanımlı Varlıkların Bulunması, İlişki, Kural, Olay Çıkarımları, Duygu Analizi(ŞEKER S. E., sadievrenseker.com, 2014)

3.4 METİN MADENCİLİĞİ İLE İLGİLİ KAVRAMLAR

Metin madenciliği kavramı ilişkili olduğu kavramlar aşağıdaki şekilde verilmiştir. (Şekil 9)

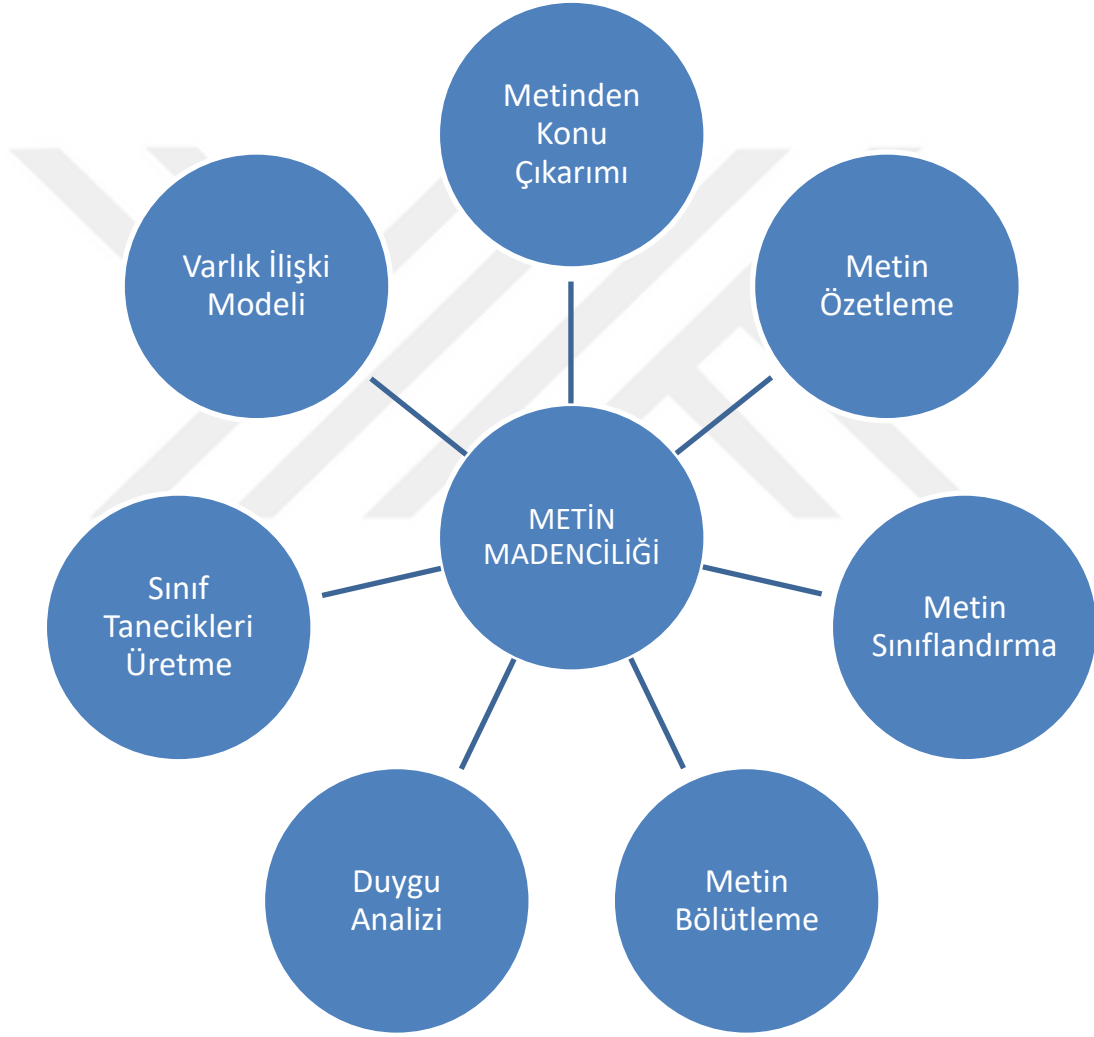


Şekil 9: Metin madenciliği kavramı ilişkili kavramlar

Burada metinler üzerinden analiz gerçekleştirilirken doğal dil işleme kavramından faydalanır. Metin madenciliği çalışmalarında doğal dil işleme yöntemi, özellik çıkartımı çalışmalarında kullanılmaktadır. Doğal dil işleme yöntemi, dil bilimi çalışmalarında yapay zeka alanı ile birlikte çalışmaktadır. Metin sınıflandırma ve metin kümeleme çalışmaları metin madenciliği alanında gerçekleştirilen yöntemlerden olduğu gibi aynı şekilde veri madenciliği alanında kullanılmaktadır. Burada açıklanan kavramlar birbiriyle iç içe ve bağlantılı kavramlardır. Örneğin istatistik alanında metin madenciliği, veri madenciliği, makine öğrenmesi, doğal dil işleme, kavram çıkarımı ve bilgiye erişim gibi konular yapay zeka ile bağlantılıdır.(DELEN, Thomas, Robert, John, & Gary, 2012, s. 31-32)

3.5 METİN MADENCİLİĞİ ÇALIŞMALARI

Metin madenciliği alanında gerçekleştirilen çalışmaların amacı, yazılı kaynaklar üzerinden yapılandırılmış veriyi elde etmektir. Bu işlemi gerçekleştirirken bazı çalışmaları gerçekleştirmeyi amaçlar. Bu çalışmalar aşağıdaki şekilde verilmiştir.(Şekil10)



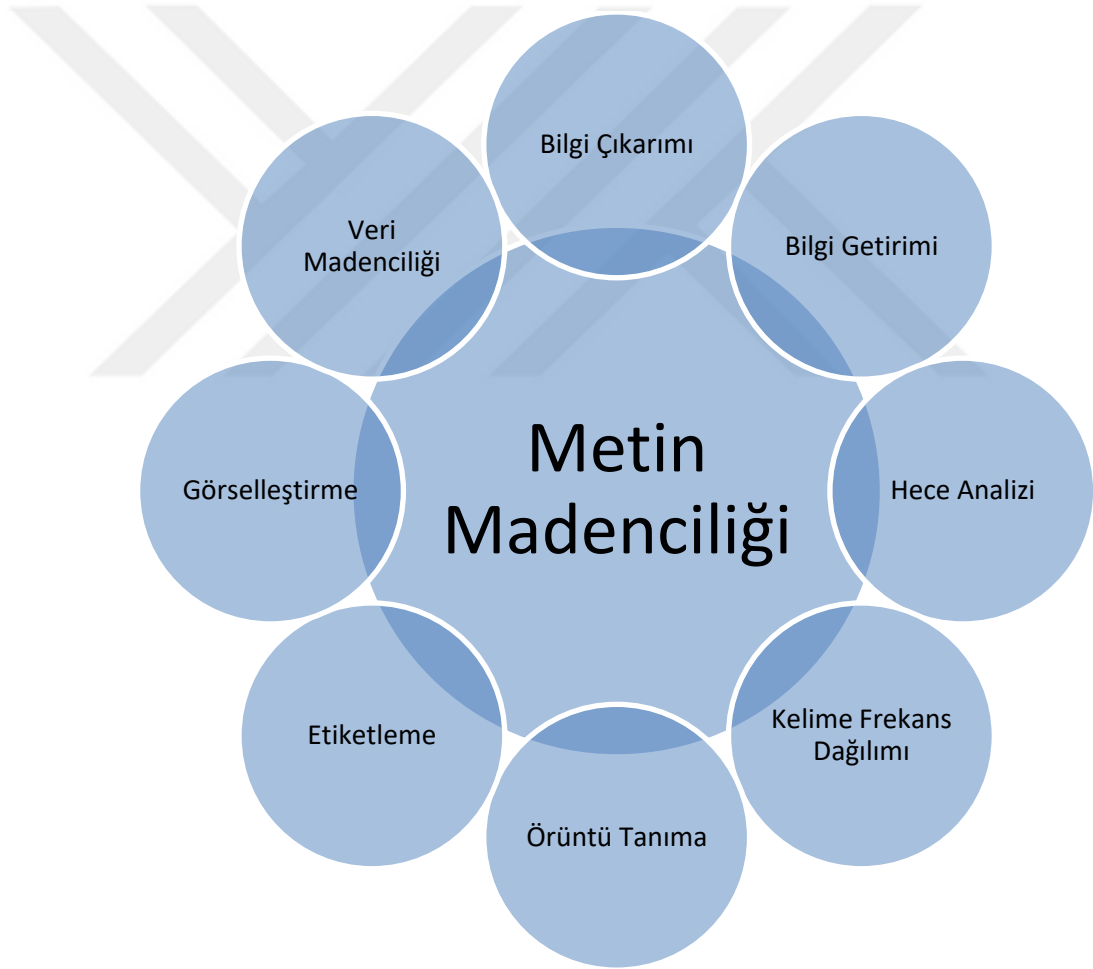
Şekil 10: Metin madenciliği alanında gerçekleştirilen çalışmalar

Metin Madenciliği yöntemi ile herhangi bir metinden konu çıkarımı yapılabilir ya da bir metnin hangi yazara ait olduğu belirlenebilir. Ayrıca metinler üzerinde kategorilere ayırma işlemi metin sınıflandırma ve metin kümeleme gibi çalışmalar ile gerçekleştirilebilir. Metin madenciliği tekniğinde en fazla kullanılan özelliklerden biri

de duygu analizidir. Bu yöntem ile herhangi bir yazılı kaynağın veya bir yorumun hangi duygu düzeyinde olduğunu olumlu, olumsuz ve tarafsız olarak belirleyebiliriz.(ŞEKER S. E., sadievrenseker.com, 2014)

3.6 METİN MADENCİLİĞİ ÇALIŞMALARINDA KULLANILAN YÖNTEMLER

Metin madenciliği tekniğinde yukarıda belirtilen çalışmaların gerçekleştirilmesi için belli başlı yöntemler kullanılmaktadır. Bu yöntemler aşağıdaki şekilde gösterilmiştir.(Şekil.11)



Şekil 11: Metin madenciliği tekniğinde kullanılan yöntemler

3.6.1 BİLGİ GETİRİMİ

Bilgi getirimi kavramı büyük boyutlu veri setleri içerisinde istenilen tarzda veriye ulaşma aşması olarak yorumlanabilir. Örneği bir yazarın yazdığı kitabın hangi

konuda olduğunun belirlenmesi bir bilgi getirme işlemidir. Örneğin internet ortamında sizin ihtiyaç duyduğunuz bilgiyi taratmanız da bir bilgi getirmedir.(ŞEKER S. E., sadievrenseker.com, 2014)

Bilgi getirme işlemi gerçekleştirilirken iki ölçüt kullanılmaktadır. Bunlar;

- 1) Kesinlik= Herhangi bir bilgi için tarama yaptığımızda elimize ulaşan verinin ne kadarı istediğimiz bilgi ile alakalı onu belirlemek amaçlanmaktadır.

$$Kesinlik(Precision) = \frac{\{İlgiliGetirim\} \cap \{BütünVeriÇıkarımı\}}{\{BütünVeriÇıkarımı\}}$$

Buradaki formülün açıklaması, tarama yapıldığında elimize ulaşan doğru veri sonuçlarının elimize ulaşan verinin tamamına oranı olarak yorumlanır.

- 2) Hassasiyet= Hassasiyet ölçütünün anlamı ise bir bilgiyi elde etmek için yapılan taramada elimize ulaşması gereken doğru verinin ne kadarı bize ulaşmış bunu belirlemektir.

$$Hassasiyet (Recall) = \frac{\{İlgiliGetirim\} \cap \{BütünVeriÇıkarımı\}}{\{İlgiliVeriÇıkarımı\}}$$

Bu ölçüt ise tarama sonucunda elimize ulaşan doğru sonucun bize ulaşması gereken doğru sonuca oranı olarak yorumlanabilir.

3.6.2 BİLGİ ÇIKARIMI

Bilgi çıkarımı, bilgi getirme kısmının bir aşaması olarak düşünülebilir. Bilgi çıkarımı bilgi getirme aşaması gerçekleştirilmeden önce uygulanan bir işlemdir ve bilgi getirme aşamasına ortam hazırlar. Bilgi çıkarımında amaç bilgi getirme işlemi uygulanmadan önce mevcut verilerin sınıflandırılması olarak düşünülebilir.(ŞEKER S. E., sadievrenseker.com, 2014)

3.6.3 HECE ANALİZİ

Metin madenciliği kavramında, doğal dil işleme yönteminin önemli bir yeri bulunmaktadır. Doğal dil işleme yönteminde zaman kavramı önemlidir. Bir fiilin zamanın belirlenmesi için sınıflandırma yöntemleri belirlenmiştir. Bu yöntemlere örnek verecek olursak Reichenbach zaman analizi yöntemi gösterilebilir. Hece analizinde de

kelime bilimsel açıdan eylemlerin zaman tasnifleri incelenmektedir. Kelime bilimsel bakış açısının mantığı ise aynı zaman dilimini belirten eylemlerin bakış açıları da aynıdır. Tam tersi durumda farklı zaman dilimlerini belirten bir durumda farklı bakış açılarında olmaktadır. Bu işlemlerin sonucunda ise kelimeleri veya eylemleri sıralayabilmektedirler.(ŞEKER S. E., sadievrenseker.com, 2014)

3.6.4 ETİKETLEME

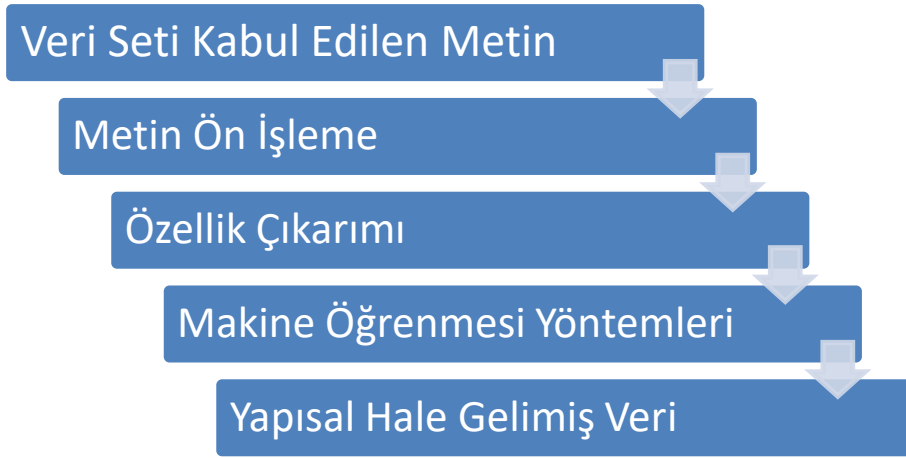
Etiketleme işlemi, bir nevi sınıflara ayırma gibi düşünülebilir. Etiketleme işleminde uygulama bazında örnek verecek olursak “Pos Tagging” yöntemi örnek gösterilebilir. Bu yöntemde etiketleme işlemi olarak kelimeler dil yapılarına göre isim, sıfat gibi hangi sınıfa ait ise o sınıfa etiketlenilip yerleştirilir. Kısaca etiketle bu anlama gelmektedir.

3.6.5 ÖRÜNTÜ TANIMA

Örüntü tanıma işlemi, bir nevi veriler arasında belli özelliklere göre değerlendirerek sınıflandırma işlevi olarak düşünülebilir. Tanımsal olarak örüntülü tanıma, mevcut veriler arasında benzerlik bulunan veya ortak bir ilişki tespit edilen verileri önceden belirlenen özelliklere göre sınıflara ayırma işlemi olarak tanımlanabilir. Örüntü tanıma işlemleri üç ayrı grupta değerlendirilmektedir. Bunlar; istatistiksel, akıllı ve yapısal örüntü tanıma işlemleri olarak ayrılmaktadır.(ÖZKAYNAK, 2011)

3.7 METİN MADENCİLİĞİ ANALİZ AŞAMALARI

Metin madenciliği tekniğinin çalışma aşamalarını aşağıdaki şekilde özetleyebiliriz.(Şekil.12)



Şekil 12: Metin madenciliği tekniğinin çalışma aşamaları

Şekil 12’den de görüldüğü gibi metin madenciliği çalışmalarının ilk aşamasında sisteme dahil edilen yapılandırılmamış verinin yani metnin üzerinde ön işleme uygulanarak gerekli olmayan kelimelerin ayıklanması ve kelimeler arasında mevcut hataları düzeltilmesi gibi işlemler uygulanır. Ardından özellik çıkarımı işlemine tabi tutulur.(ŞEKER Ş. E., Metin Madenciliği (Text Mining), 2015, s. 31) Özellik çıkarımı işlemi uygulanırken Bag of Words ve Pos Tagging gibi yöntemler uygulanır. Özellik çıkarımı uygulanan veri üzerinde makine öğrenmesi yöntemleri uygulanır. Bu yöntemler; Naive Bayes, Destek Vektör Makineleri, k-NN algoritması gibi yöntemler olup sınıflandırma, tahmin vb. işlemler yapılarak sonuçta yapısal hale gelmiş veri elde edilir.

3.8 METİN MADENCİLİĞİNİN İLGİ ALANINA GİREN SORUNLAR

- a) Enformasyon Getirimi=İlgilenilen konu hakkında ön bilginin toplandığı aşamadır.
- b) Doğal Dil İşleme Aşaması= Yazılı kaynak içerisinde özellik belirleme ve bazı verilerin elde edilmesi aşamasıdır.
- c) Adlandırılmış Varlık Tanıma=Metnin analiz aşamasında istatistiksel özellikleri belirlemek için kullanılır.
- d) Örüntüsü Tanımlı Varlıkların Bulunması=Belirli koşullarda, yazılı kaynağın içerisindeki özel veriler metin madenciliğine konu olabilir. Örneğin mail adresleri, telefon numaraları, tarih gibi bazı bilgileri özel olarak almak isteyebiliriz.

- e) İlişki, Kural, Olay Çıkarımları=Çeşitli hedefler doğrultusunda yazılı kaynağın içerisinden gerekli verinin çıkarımı
- f) Duygu Analizi(Sentiment Analysis)=Yazılı kaynaklarda geçen duygusal ifadeleri tespit etmeyi sağlar.(ŞEKER Ş. E., Metin Madenciliği (Text Mining), 2015, s. 32)

3.9 METİN MADENCİLİĞİ KULLANIM ALANLARI

Metin madenciliği ile çok farklı alanlarda çalışmalar yapılabilmektedir. Örneğin, herhangi bir kitabın konunun veya romanın özetini çıkartabilmektedir. Ya da bir metinde veya kitapta yazılan metin içerisindeki konuyu belirleyebilir. Başka bir çalışma alanı ise farklı metinler incelenip analiz edilerek o metnin hangi yazar tarafından yazıldığına dair tahminde bulunabilir. Örneğin, metin sınıflandırma yöntemi kullanılarak herhangi bir metni konu bazında veya herhangi bir ölçüte göre sınıflandırabiliriz. İnternet ortamında müşterilerin şikayet ve yorumları incelenip bunlar üzerinde duygu analizi gerçekleştirilebilir ve bu sonuçlar gerekli birimlere yönlendirilebilir. Metin madenciliği alanında yapılabilecek farklı bir çalışma ise kümeleme yönteminden faydalanılarak benzer tarzda metinler bir araya getirilebilir. Bu çalışma benzer tarzda haberlerin bir araya toplanması olarak örneklendirilebilir.

Burada metin madenciliğinin kullanım alanları ve yapılabilecek çalışmalar daha fazla da örneklendirilebilir. Görüldüğü üzere metin madenciliği yöntemi ile çok farklı çalışmalar yapılmakta ve bu çalışmalar çeşitli sektörlerde kullanılabilmektedir. Muhasebe, E-Ticaret ve özel sektördeki çeşitli alanlardaki firmalar tarafından metin madenciliği teknikleri kullanılmaktadır.

Muhasebe alanında metin madenciliği yönteminin kullanımını incelediğimizde finansal tabloların analizinde kullanıldığını görmekteyiz. Finansal tablo analizlerinde yapısal olmayan verilerin yani metin analizinde kullanılmaktadır. Bu sayede muhasebe alanında metin madenciliği yönteminin, hileli işlemlerin tespiti ve firmaların performanslarının ölçülmesi gibi faydaları bulunmaktadır.

E-Ticaret alanında ise metin madenciliği yöntemi, mevcut internet siteleri içerisinde elektronik ticaret sitelerinin belirlenmesinde kullanılmaktadır. Bu sayede

kullanıcıların internet ortamında bulunan birçok internet sitesi arasından E-Ticaret sitelerine daha kolay bir şekilde ulaşması sağlanmaktadır.

4 DÖRDÜNCÜ BÖLÜM

4.1 DUYGU ANALİZİ

İnsanoğlu, yaşamı boyunca her geçen gün bilgi ve birikimini arttırmakta, farklı fikirler ve duygulara sahip olmaktadır. Tabii bu fikir ve duygular bazen kişiden kişiye bazense konudan konuya çeşitlilik gösterebilir. Örneğin, internette film izlemek istiyorsunuz ve izleyeceğiniz filme karar verirken o filme ait yapılan yorumlara göz gezdiriyorsunuz ancak görüyoruz ki yorumların bazıları film hakkında olumlu görüşler belirtirken bazıları olumsuz hatta bazıları tarafsız görüş bile sergileyebiliyor. İşte Duygu Analizi(SA) de aynı örnekteki gibi bireylerin herhangi bir konu hakkındaki görüşlerinin olumlu, olumsuz ve tarafsız olarak sınıflandırılmasında yardımcı olmakta ve o konu hakkındaki genel görüşün ne yönde olduğunu göstermektedir. Duygu analizi, bir metin analizi işlemi olup yazılı kaynağın duygu düzeyi sınıfını belirlemeyi amaçlar. Literatürde duygu analizi çalışmaları duygusal kutupsallık (sentimental polarity) olarak geçmektedir ve verilen metni positive, negative ve nötr olarak sınıflandırmayı amaçlamaktadır.(ŞEKER S. E., 2016, s. 21-36) Duygu analizi, her bir verinin bir kategoriye temsil ettiği bir veri sınıflandırma işlemi olarak düşünülebilir.(Prabowo, Rudy, & Thelwall, 2009, s. 143-157). Duygu analizi yapılırken analiz sonucunda sağlıklı ve mantıklı sonuçlar elde edebilmek için incelenen metnin belli bir dilbilgisi kuralına uygun olması gerekmektedir. Duygu analizi gerçekleştirilirken metnin içerisinde duygu belirten kelimelerin düzgün bir şekilde tespit edilmesi gerekir. Bunu tespit edebilmek için önceden belirlenmiş kelime listeleri kullanılarak ya da öznitelik seçimi yöntemi kullanılarak metnin içerisindeki kelimelerin hangi duyguyu içerdiğini tespit edebiliriz.

Çalışmada, duygu analizi gerçekleştirirken kullanacağımız verileri yani metinler twitter sosyal ağından analiz programı yardımıyla elde edilmiştir. Twitter sosyal ağından tarama yapmadan önce, belirli konu başlıkları belirlenip

bu konular ile alakalı tarama yapılmıştır. Analizde kullanacağımız metinleri elde ederken Twitter sosyal ağını tercih etmemizin sebebi, twitter farklı kültür ve özellikten milyonlarca kullanıcıya hitap ettiği için farklı dillerde ve içeriklerde veriler toplamak mümkün (ÇOBAN, ÖZYER, & TÜRÜKLÜ ÖZYER, Türkçe Twitter Mesajlarının Duygu Analizi, 2010, s. 2-3) olduğundan dolayı twitter sosyal ağını tercih edilmiştir. Gün geçtikçe bu tarz çalışmaların sayısı artmakta ve ilgi alanı genişlemektedir. Özellikle, twitter mesajları olmak üzere çeşitli sosyal medya ağlarından elde edilen veriler üzerinde duygu analiz tekniği kullanılarak yapılan çalışmalar firmaların ilgisini çekmektedir.(Tang, Huifeng, Tan, & Cheng, 2009, s. 10769)

4.2 DUYGU ANALİZİ ÇALIŞMALARI HANGİ AMAÇLARLA KULLANILIR?

Duygu analizi çalışmaları bireylerin duygu ve düşüncelerinin ne yönde olduğunu belirlemeyi amaçlamaktadır. Gün geçtikçe internetin sosyal hayatımızda daha fazla yer edinmesi firmaların ticari anlamda bu alanlara yönelmesine yol açmıştır. İnternet ortamındaki bilgi düzeyinin artması bununla birlikte sosyal medya ağlarının bireyler tarafından daha fazla talep görmesi, firmaların bu ortamlarda reklamcılık faaliyetlerine başlamasına ortam hazırlamıştır. İnternet ve sosyal medya alanındaki yaşanan bu hızlı gelişim üzerine firmaların reklamcılık, satış ve pazarlama gibi alanlarda birbirleriyle olan rekabeti giderek artmıştır.

Duygu analizi kavramı bu kısımda özellikle firmalar için büyük bir önem arz etmektedir. Firmalar satış ve pazarlama alanında veya reklamcılık alanında duygu analizi tekniğinden faydalanmaktadır. Duygu analizi tekniği, özellikle firmalar açısından bakıldığında büyük avantajlar sağlamaktadır. Örneğin, firmaların satış ve pazarlama anlamında, ürünlerinin müşteriler üzerindeki etkisini, müşterilerin ürüne olan memnuniyet düzeyini ve firmanın piyasadaki marka değerini belirleyebilmesi gibi konularda büyük fayda sağlamaktadır. Reklamcılık sektöründe duygu analizi tekniğinin önemi incelendiğinde ise özellikle sosyal medya ağlarının gelişimi ve bireylerin bu sosyal ağlara olan talebinin artmasıyla birlikte internet ağları ve sosyal medya uygulamaları firmaların reklamcılık anlamında göz bebeği haline gelmiştir. Bu durum,

firmaların sanal ortamda birbirleriyle olan rekabetini gün geçtikçe arttırmakta ve firmalar için duygu analizi kavramını da o denli önemli kılmaktadır.

Duygu analizi yöntemi gün geçtikçe hayatımızdaki yerini ve gerekliliğini arttırmakta ve buna istinaden twitter, instagram ve facebook gibi sosyal medya uygulamaları günümüzde firmalar tarafından reklamcılık anlamında daha fazla tercih edilmektedir. Özellikle, twitter sosyal ağı üzerinde milyonlarca kullanıcı bulunmakta ve gün içerisinde milyonlarca tweet atılmaktadır. İnternet ve sosyal medya üzerindeki gerçekleştirilen şikayet ve yorumlardan oluşan bu tarzda büyük veri boyutlarının el yordamıyla firmalar tarafından takip edilmesi zor olacağından duygu analizi yönteminin gün geçtikçe firmalar için önemi daha fazla artmakta ve talep görmektedir.

4.3 DUYGU ANALİZİ KELİME SINIFLANDIRMA AŞAMALARI

Görselde duygu analizi yönteminde kelimelerin nasıl sınıflandırıldığına dair aşamalı anlatım yer almaktadır.



Şekil 13: Duygu analizi kelime sınıflandırma şeması

Burada ilk aşamada, sisteme dahil edilen metin önceden belirlenmiş bir sözlük tarafından okunur. Bu aşamada, sözlük içerisinde daha önceden olumlu ve olumsuz kelimelerin yer aldığı bunların da bir puanlama cetvelinin olduğu bilgiler bulunmaktadır. Ardından okunan bu kelimeler pozitif ve negatif olma durumlarına göre

etiketlenerek iki ayrı liste haline getirilir. Burada, etiketlenen kelimeler içerisinde etiketlenmemiş olanlar üzerinde bir ön işleme yapılarak temizleme işlemi yapılır ve geriye, doğru olarak etiketlenmiş pozitif ve negatif kelimeler kalmaktadır. Burada, kelimelerin ve cümlelerin duygu düzeyleri belirlenirken bazı hesaplamalar yapılmaktadır. Bu hesaplamalar;

$$\text{Duygu Değeri} = \frac{\text{Etiketlenmiş Pozitif Kelime} - \text{Etiketlenmiş Negatif Kelime}}{\text{Etiketlenmiş Toplam Kelime Sayısı}}$$

Burada, duygu değeri sonucu sıfırdan büyük çıkarsa (Duygu Değeri > 0) kelime pozitif eğer duygu değeri sonucu sıfıra eşit ise (Duygu Değeri = 0) kelime nötr ve duygu değeri sonucu sıfırdan küçük çıkarsa (Duygu Değeri < 0) kelime negatif olarak belirlenmiş olacaktır.

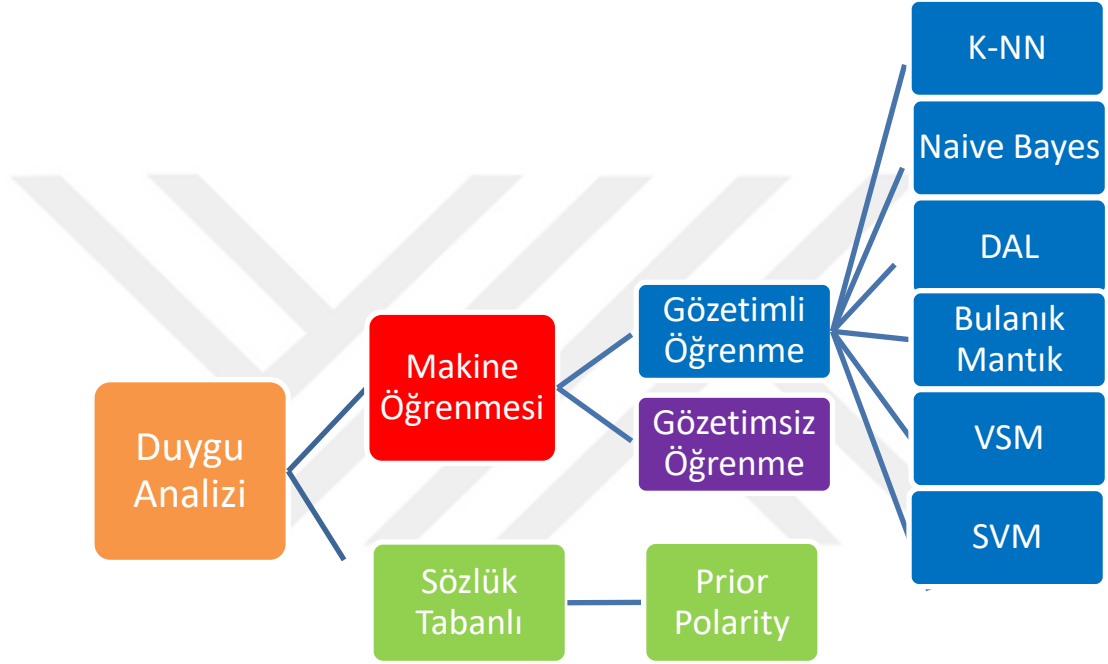
4.4 ÖZİNİTELİK SEÇİMİ

Öznitelik seçimi, var olan veri setimizi temsil edebilecek en iyi alt kümenin seçimi olarak tanımlanmaktadır. Bu tanımı biraz daha açacak olursak elimizdeki veri setinin içerisindeki işimize daha az yarayan veya hiç yaramayan özniteliklerin çıkarılması anlamına gelmektedir. Yani, veri setimizin içerisindeki X tane nitelik içersinden en iyi Y tane niteliğin seçilme işlemidir. Öznitelik seçimi, istatistik ve veri madenciliği gibi alanlarda verileri sınıflandırma veya veriler arasında eleme yapma gibi işlemlerde sıkça kullanılmaktadır. Öznitelik seçiminin bize faydalarını inceleyecek olursak, büyük boyutlu verilerle çalışırken öznitelik işlemi sonucunda veri boyutu küçüleceği için bize hem zaman tasarrufu sağlarken hem de algoritmaların çalışma hızını arttırmaktadır. Gerçekleştirilen bu işlem bize ekstra veri depolama alanı oluştururken ayrıca hem verilerimizin kalitesini artırır ve analiz sonuçlarımızın da daha başarılı ve sağlıklı olmasını sağlamaktadır.

Öznitelik seçiminin çalışma mantığını inceleyecek olursak, bu seçim yapılırken çeşitli algoritmalar kullanılmaktadır. Bu algoritmalardan biri arama stratejisidir ve öznitelik havuzunu, rastgele sezgisel bir şekilde arama yöntemidir ve çok fazla

kullanılan bir yöntem değildir. Diğeri ise öznitelik kalitesini değerlendirmedir ve içerisinde zarflayıcı, wrapper ve filtreleme yöntemleri kullanılmaktadır. Filtreler hızlı çalışmalarından dolayı büyük öznitelik uzaylarında çalışma açısından bize fayda sağlamaktadır. (AKBA, 2014, s. 4)

4.5 DUYGU ANALİZİNDE KULLANILAN YÖNTEMLER



Şekil 14: Duygu analizinde tekniğinde kullanılan yöntemler

Çalışmada analiz aşamasında sözlük tabanlı yöntem ile öncül yön puanlandırma sistemi kullanılmıştır. Sözlük tabanlı yöntem ile elde edilen metinler fiil köklerine ayrılarak bir sözlük ile karşılaştırılır ve her kelimenin frekansı sözlük sayesinde belirlenerek kelimenin ifade ettiği duygu düzeyi belirlenir.

4.5.1 MAKİNE ÖĞRENMESİ YÖNTEMLERİ

4.5.1.1 Dil Etkileşim Sözlüğü(DAL)

Dil etkileşim sözlüğü, ilk olarak 1989 yılında ortaya çıkmış ve kelimelerde bulunan duygusal ifadeleri puanlamak amacıyla ortaya çıkmıştır. DAL yöntemi başlangıçta sadece İngilizce dili için geliştirilmiş fakat sonradan Fransızca ve Almanca gibi diler içinde geliştirilmiş ve başarılı bir şekilde uygulanmıştır. DAL yöntemi halen daha duygu analizi çalışmalarında kullanılan bir yöntemdir.

Dil etkileşim yöntemi ilk çalışmaları günümüz koşullarına göre biraz daha zorlayıcı şekilde olmuştur. O dönemin de şartları göz önüne alındığında mevcut bir analiz programı olmadığı için bireyler işlemleri el yordamıyla gerçekleştirmişlerdir. DAL yöntemi, ilk uygulamalarında bireyler herhangi bir sözlük içerisindeki duygu içeren kelimeleri el yordamıyla çıkararak sınıflandırmışlardır. Sınıflandırma işleminde her bir gruba 1-5 arası puan vererek etiketlemişlerdir. Bu puanlama işlemine karar verirken 1-2 aralığını olumsuz duygu içeren kelimeler 4-5 aralığını olumlu duygu içeren kelimeler ve geri kalan kısmı da ne olumlu ne de olumsuz duygu içeren yani nötr kelime olarak gruplandırılmıştır. Yapılan bu işlemlerin ardından bazı cümleler test cümlesi olarak seçilir ve bir kök eleme işlemine tabi tutulur. Ardından test cümlesinde bulunan kök halinde kelimeler sistem içinde bulunuyorsa bu köklerin sonuçlarının yani değerlerinin aritmetik ortalaması alınır ve hangi duyguyu gösterdiği tespit edilir.(AKBA, 2014, s. 12-13)

4.5.1.2 Naive Bayes

Naive Bayes yöntemi Bayes teoreminden ilham alınarak oluşturulmuş, makine öğrenme yöntemleri arasında temel olarak kabul edilen ve kolaylıkla uygulanabilen ve anlaşılabilen bir sınıflandırma yöntemidir. Naive Bayes yönteminin diğer yöntemlerden ayıran en önemli özelliği, durumları olasılık hesaplamaları yaparak değerlendirmesidir. Bu değerlendirmenin yapılabilmesi için sistem içerisinde daha önceden etiketlenmiş bir eğitim veri seti kullanılır. Naive Bayes yöntemi kullanılarak test edilecek bir durumda önceden belirlenmiş olan eğitim veri seti kullanılarak bir olasılık değeri hesaplanır ve hesaplanan bu sonuca göre test işlemine tabi tutulacak verinin hangi sınıfa ait olduğu belirlenir.(KAYNAR, GÖRMEZ, YILDIZ, & ALBAYRAK, 2016, s. 237)

$$P(A/B) = \frac{P(B/A)*P(A)}{P(B)} \quad (8.1)$$

Bu formülde terimleri ve anlamlarını yorumlayacak olursak;

$P(A)$ = A durumunun gerçekleşme olasılığı ya da A terimini bir öznitelik vektörü olarak düşünüp herhangi bir sınıfa ait olma olasılığı olarak yorumlayabiliriz.

$P(B)$ = B durumunun gerçekleşme olasılığı ya da B öznitelik vektörü olarak tanımlayabiliriz.

$P(A/B)$ = B öznitelik vektörünün oluşması durumunda A vektörünün X sınıfına ait olma olasılığı olarak yorumlayabiliriz.

Naive Bayes yöntemi ile uygulanan bu formülde, pay kısmındaki ifadenin sonucunun sıfır çıkması gibi bir durumla karşılaşılabilir. Bu sorunun çözümü için eğer $P(B/A)*P(A)$ kısmının sıfır çıkması durumunda sıfır yerine eşik değer denilen uygun bir değer yerleştirilir ve işlem devam eder.(KAYNAR, GÖRMEZ, YILDIZ, & ALBAYRAK, 2016, s. 237-238)

4.5.1.3 k-NN Algoritması

Makine öğrenme yöntemleri arasında etkin kullanılan yöntemlerden biri olan k-NN algoritmasını aslında kısaca en yakın komşu olarak da tanımlayabiliriz. Naive Bayes yönteminde olduğu gibi k-NN yönteminin de doğru bir şekilde çalışabilmesi için bir eğitim setine ihtiyaç vardır. K-NN algoritması gereksiz bilgilerin fazla olduğu veri setlerinde sınıflandırma işleminde başarılı olan bir yöntemdir. K-NN algoritması ile k-means yöntemi arasında birçok benzerlik bulunmaktadır fakat iki yöntem arasında bazı farklılıklar da mevcuttur. Kısaca bu farklılıklara değinecek olursak k-NN algoritmasında yöntem gerçekleştirilirken bir eğitim seti kullanılır fakat k-means yönteminde bu tarz bir eğitim seti kullanılmamaktadır.

Daha önceden de belirttiğimiz gibi k-NN algoritması en yakın komşu mantığına dayandığı için nesnelere arası uzaklığı hesaplar ve bu hesaplama işlemini gerçekleştirirken “Manhattan”, “Euclidean”, “Minkowski” yöntemleri ve k-means yönteminde de kullanılan “Öklid Uzaklığı” gibi yöntemler kullanılır. K-NN algoritmasının olumsuz yanı ise sisteme eklenen her veride ve her sorguda hesaplamaların tekrardan yapılmasından dolayı işlem fazlalığı ortaya çıkmaktadır(AKBA, 2014, s. 14).

Bu algoritmanın çalışma mantığını inceleyecek olursak mevcut sistemin içerisine yeni bir veri eklendiğinde bu verinin hangi sınıfa ait olduğunu belirlemek için gerçekleştirilecek ilk aşama k değerini belirlemektir. Ardından, sistemde mevcut olan

nesneler ile hedef nesne arasındaki uzaklık hesaplanır. Daha sonra hesaplanan bu uzaklıklar sıralanır ve en yakın komşu uygun k değeri kadar seçilir. Örneğin işlemin en başında belirttiğimiz k değeri 3 olsun en yakın ilk 3 uzaklık seçilir. Bu işlemin ardından belirlenen bu 3 uzaklık içerisinde hangi sınıfın sayısı daha fazla ise(örneğin 2. Kategoriden 1 tane, 3. Kategoriden ise 2 tane sınıf var sonuç olarak 3. Kategori seçilir) o sınıf tercih edilir ve sisteme eklenen yeni veri o sınıfa dahil edilir. Ayrıca sisteme eklenen her yeni veride bu işlem tekrarlanır.(UZUN, 2016)

4.5.1.3.1 k-NN Algoritması Uzaklık Hesaplama Yöntemleri

Öklid Uzaklık Hesaplama Yöntemi;

$$d(A, B) = \sqrt{(X1 - X2)^2 + (Y1 - Y2)^2}$$

Manhattan Hesaplama Yöntemi (8.2)

$$\text{Manhattan} = \sum_{n=1}^k (|Xi - Yi|)$$

Euclidean Hesaplama Yöntemi (8.3)

$$\sqrt{\sum_{i=1}^n (Xi - Yi)^2}$$

Minkowski Hesaplama Yöntemi (8.4)

$$\left(\sum_{n=1}^k (|Xi - Yi|)^q \right)^{1/q}$$

4.5.1.4 Bulanık Mantık(Fuzzy Logic)

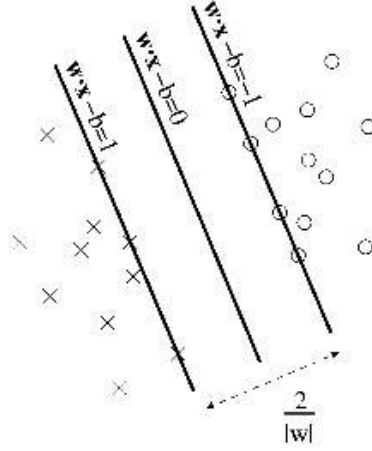
Daha önceden tanıtılan sınıflandırma yöntemlerinde çalışma sistemi doğru ya da yanlış şeklinde seçme aşaması vardı. Fakat bulanık mantık yönteminde ise kesin yargılar bulunmamakta ve niteliksel tanımlara daha fazla yer verilmektedir. Bulanık mantığın çalışma sistemi sistem içerisindeki sorguya doğru ya da yanlış şeklinde cevap vermek yerine Az çok doğru gibi cevaplar vermesidir. Buradaki amaç kesin yargılardan doğacak olan riskten kaçmaktır. Bulanık mantık yönteminin de çalışma yapısı biraz k-NN algoritmasına benzemektedir. Bulanık mantık yöntemi de verilerin birbirine olan yakınlığını inceler yani bir veriyi incelerken diğer verilere olan uzaklığa bakarak inceleme işlemini gerçekleştirir.(AKBA, 2014, s. 18)

4.5.1.5 Destek Vektör Makineleri(SVM)

1963 yılında Vapnik ve Chervonenkis tarafından temelleri atılan bu yöntem sınıflandırma işleminde oldukça etkili olan parametrik olmayan gözetimli bir makine öğrenme yöntemidir. Destek Vektör Makineleri uzaktan algılama sistemi, yüz tanıma sistemi ve ses analizi gibi alanlarda kullanılmaktadır. Bu yöntemin en önemli avantajları yüksek boyutlu uzayda etkili olmasıdır. Bunun yanı sıra bellek yönetiminde verimli olmakla birlikte doğrusal olmayan vektör makinelerinde uygulanan kernel fonksiyonları gibi yöntemler kullanılmaktadır. Destek Vektör Makine yöntemi temel olarak bir düzlem üzerindeki iki sınıfın arasına bir sınır çizerek ya da diğer adı hiper düzlem çizerek ayırma işlemine dayanmaktadır. İlk aşamada iki grupta doğrusal verilerin sınıflandırılması için oluşturulan bu yöntem daha sonra ikiden fazla grupta ve doğrusal olmayan veriler içinde geliştirilmiştir. (Üstüner, Balık Sanli, Bektas Balcik, & Esetlili, 2013) Bu yöntem iki grup arasına çizilecek sınırı belirlerken grup üyelerine yakın iki adet sınır çizgisi belirler ve bu sınırları birbirine yaklaştırarak ortak bir sınır çizgisi oluşturur.

4.5.1.5.1 Doğrusal Destek Vektör Makineleri

Aşağıdaki şekilde iki farklı grubun iki boyutlu bir düzlemde sınıflandırılması yer almaktadır.



Şekil 15: İki boyutlu düzlemde sınıflandırma işlemi

Şekil üzerindeki her bir boyutu ve düzlemi birer özellik olarak düşünürsek sisteme giren her bir verinin özellik çıkarımı yapılmış ve düzlemde her veriyi gösteren farklı bir nokta elde edilmiştir. Bu noktaların sınıflandırılması çıkarılan özelliklere göre sisteme giren verinin sınıflandırılması demektir.(ŞEKER S. E., SVM (Support Vector Machine, Destekçi Vektör Makinesi), 2008)

$$\mathcal{D} = \{(\mathbf{x}_i, c_i) | \mathbf{x}_i \in \mathbb{R}^P, c_i \in \{-1, 1\}\}_{i=1}^n$$

Yukarıdaki düzlemdeki her bir noktanın tanımını bu denklem ile açıklamak mümkündür. Denklem ne anlatmak istediğini kısaca açıklayacak olursak her bir x, c ilişkisi vektör uzayımızda bir nokta ise ve bu noktanın -1 ile $+1$ arasında olduğunu gösteren ise c simgesidir. Vektör uzayındaki noktalar $i=1$ den n ye kadar gitmektedir. Gerçekleşen bu olayın bir aşırı düzlem(hyperplane) üzerinde gerçekleştiğini varsayalım. Bu varsayımı açıklamadan önce aşırı düzlemi(hyperplane) kısaca açıklayacak olursak, hyperplane bir kesişim olarak tanımlanır ve 3 boyutun üzerindeki boyutlarda aşırı düzlemler vardır, bu o boyutlardaki kesişimleri ifade eder. Şimdi aşırı düzlemlerdeki bu olayın nasıl gerçekleştiğini açıklayacak olursak buradaki her noktanın $wx-b=0$ denklemi ile tanımlayabiliriz. Bu denklem $ax+b$ denklemine benzetilmekle birlikte denklemdeki w hyperplane dik olan vektörü x ise noktanın parametresini ve b ise kayma oranını göstermektedir. Yukarıdaki denkleme göre iki sınıf arasındaki mesafeyi hesaplamak

istersek formülümüz $b/[w]$ şeklinde olacaktır. Ayrıca yukarıda ilk şekilde gösterilen iki denklem mevcuttur. Bu denklemler;

$$wx-b=1$$

$$wx-b=-1$$

olarak bulunmuştur. Bu iki denklem doğruların kaydırılması sonucu elde edilen en yüksek değerlerin bulunması işleminin sonucudur. (ŞEKER S. E., SVM (Support Vector Machine, Destekçi Vektör Makinesi), 2008)

Destek Vektör Makineleri yöntemi verilerin doğrusal olup olmama durumuna göre doğrusal destek vektörleri ve doğrusal olmayan destek vektörleri olarak ikiye ayrılmaktadır.

4.5.1.5.2 Doğrusal Olmayan Destek Vektör Makineleri

Doğrusal olmayan destek vektörler yöntemi doğrusal destek vektörler gibi çalışmazlar. Bunun nedeni ise doğrusal olmayan bir veri setinde doğrusal bir hiper düzlem çizememeleridir. Bu sorunun çözümü için çekirdek yöntemi ya da çekirdek numarası olarak da adlandırılan kerneltrick'ler kullanılmaktadır. Bu alanda en çok kullanılan çekirdek yöntemlerinden bir tanesi Polynomial Kernel yöntemi iken diğeri ise Gaussian RBF (Radial Basis Function) Kernel yöntemidir. (ÜLGEN, 2017)

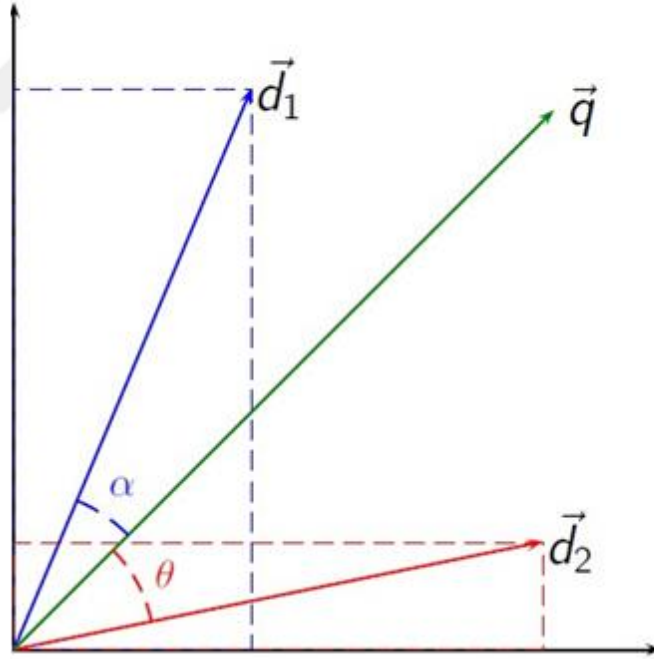
4.5.1.6 Vektör Uzay Modeli (VSM)

Özellikle doğal dil işleme ve metin madenciliği alanlarında uygulanan bir yöntemdir. Vektör uzay modelini kısaca tanımlayacak olursak herhangi bir metni veya kelimeyi bir vektör olarak göstererek uzay içerisinde tanımlama işlemine denir. Vektör uzay modeli yöntemi ile metinleri belli kategoriler altında sınıflayabilir ya da elimizdeki mevcut metne benzeyen diğer metinleri listeleyebilir gibi işlemleri gerçekleştirebiliriz. Bu işlemleri gerçekleştirebilmek için matematiksel yöntemler kullanmamız bunların analizini gerçekleştirmemiz gerekir. Bu matematiksel işlemleri gerçekleştirebilmek için ilk aşamada elimizdeki metni vektör olarak ifade etmemiz gerekmektedir. Bu sayede elimizdeki metin üzerinde matematiksel işlemler yapıp metinler arasındaki ilişkiyi de inceleyebiliriz. Burada metinler arasındaki ilişkiyi belirlerken kosinüs benzerliği yöntemi kullanılmaktadır. Vektör uzay modelini gösterecek olursak;

$$q = (w_{1,q}, w_{2,q}, \dots, w_{n,q}) \quad (8.5)$$

$$d_j = (w_{1,j}, w_{2,j}, \dots, w_{t,j})$$

Burada d arama yaptığımız metni ve q ise metin vektörlerimizi göstermektedir. Her vektör içerisinde metinlere ait sayısal değerler mevcuttur. Örneğin, bir kelime metin içerisinde geçiyorsa 1 eğer metin içerisinde geçmiyorsa 0 değerine alacaktır ve (0,1) şeklinde ifade edilir. Bu sayısal değerleri belirlemenin yani metnin vektörler şeklinde ifade edilmesinin birçok yöntemi vardır fakat en çok kullanılan yöntem TF-IDF yöntemidir. Bunun yanı sıra Metin etiketleme(POS Tagger) ve N-Gram gibi yöntemlerde kullanılabilir. Metinler arasındaki ilişkiyi kısa bir örnek ile tanımlayacak olursak;



Şekil 16: Vektör uzay modeli düzlem üzerinde gösterimi

Grafikte gördüğümüz gibi yukarıdaki q ve d_1, d_2 simgeleri modeldeki anlamlarına karşılık gelmektedir. Buradan Öklid geometrisi ve vektörlerin iç çarpımı şu şekilde ifade edilir;

$$\vec{q} \cdot \vec{d}_2 = ||d_2|| \cdot ||q|| \cdot \cos \theta \quad (8.6)$$

Bu formül içerisinde kosinüs ifadesini yalnız bırakırsak metinler arası ilişkiyi yani kosinüs benzerliğini bulabiliriz.

$$\cos \theta = \frac{\mathbf{d}_2 \cdot \mathbf{q}}{||\mathbf{d}_2|| ||\mathbf{q}||} \quad (8.7)$$

Bu formül üzerinden çıkan sonucu nasıl yorumlayabiliriz diye bakacak olursak buradaki θ açısının değeri azaldıkça metinler arasındaki benzerlik de artacaktır. Örneğin, $\cos(90)=0$ yani metin veya kelimeler arasında hiç benzerlik yok anlamı taşırken $\cos(0)=1$ metinler arasında yüzde yüz benzerlik olduğunu ifade etmektedir. (AKÇORA, 2016)

4.5.2 MAKİNE ÖĞRENİMİ YÖNTEMLERİNİN BAŞARISI ARTTIRMAYA YÖNELİK UYGULAMALAR

4.5.2.1 Kelime Öbekleri(Bag of Words)

Genellikle makine öğrenimi yönteminde önce terim çıkarımlarında kullanılan bu yöntem doğal dil işleme gibi alanlarda sıklıkla kullanılmaktadır. Bag of Words yöntemi herhangi bir cümle içerisinde geçen kelimeleri tek bir kelime gibi görüp o şekilde değerlendirmemizi sağlayan bir yöntemdir. Bu işlemin gerçekleşebilmesi için analize tabi tutulacak kelime öbeklerinin sisteme dahil edilmesi yani öğretilmesi gerekmektedir. Varsayalım ki analize tabi tutacağımız sorgu önceden eğitilmiş olan veri setindeki kelime öbekleri ile eşleşiyorsa bu kelime öbekleri vektörler halinde test haznesine kaydedilir ve sonrasında bu kelimelerin geçme sıklıklarına göre hesaplamalar yapılır ve sonuca ulaşılır.

4.5.2.2 TF-IDF(Term frequency – Inverse document frequency)

AĞIRLIKLANDIRMA YÖNTEMİ

Günümüzde yazılmış bazı dokümanlarda belirli terimlerin çok sık kullanıldığı ve sürekli bir şekilde tekrarlandığı görülmekte ve sanki mevcut dokümanı temsil eder hale

gelmişlerdir. Türkçe karşılığı Terim Frekansı - Ters Belge Frekansı anlamına gelen TF-IDF modeli metin madenciliği gibi alanlarda kullanılan ve dokümanların içerisindeki fazla tekrarlanan veya terim kirliliğine sebep olan durumları azaltmak için kullanılan bir terim çıkarma yöntemidir. TF-IDF modelinin çalışma mantığı cümlenin içerisinde tekrarlanan kelimelerin ne kadar sıklıkla kullanıldığına yani frekansına ve diğer dokümanlarda geçme sıklığına birlikte bakarak hesaplamaları yapmaktadır. Böylelikle gruplar için önemli olan kelimeleri belirlemiş olur. Bu modelin analiz aşamasında bize artılarını inceleyecek olursak analize tabi tutulacak kelime sayısını düşürmemize ve işlem süresini azaltmamıza yardımcı olur. Ayrıca cümlelerin içerisinde sıklıkla rastladığımız ve analiz sonucuna etki etmeyecek zamir gibi yapıların sonuca etki etmesini önlemektedir.

TF-IDF modelinde hesaplama işlemlerinde iki ölçüt kullanılmaktadır. Bunlar Terim Sıklığı(TF) ve Ters Belge Sıklığı (IDF) ölçütleridir. Burada cümlelerin içerisinde tekrarlanan kelimelerin sıklığı arttıkça TF-IDF hesap değeri artacaktır. Şimdi TF-IDF modelinin formülünü inceleyelim;

TF(i, j) = (Dokümandaki i. terimin sıklığı) / (Dokümandaki toplam terim sayısı)

$$TF = \frac{f}{df} \quad (8.8)$$

IDF(i) = $\log_2(\text{Toplam doküman sayısı} / \text{i. terimi içeren doküman sayısı})$

$$IDF = \frac{n}{df} \quad (8.9)$$

Burada TF ve IDF değerleri hesaplandıktan sonra TF*IDF işlemi uygulanarak değer hesaplanmış olacaktır.(AKBA, 2014, s. 16-17)

Kısaca TF-IDF modelinde hesaplama işlemi bu formül üzerinden yapılmaktadır.

4.5.2.3 POS Tagger (Metin Parçası Etiketleme) Yöntemi

Pos Tagger yöntemi cümle içerisindeki kelimelerin dil yapısına göre etiketlenilme işlemidir. Bu tanımdaki dil yapısından kasıt isim, sıfat, bağlaç, zarf gibi terimlerdir ve işlemin devamında ise kelimeler bu dil yapılarından hangisinin sınıfına ait olduğu belirlenir ve o sınıfa etiket olarak konulur. (AKBA, 2014, s. 16)

Gerçekleştirilen bu işlem ilk bakışta basit olarak görülebilir fakat bazı zorlukları da vardır. Örneğin, koyun kelimesi hayvan olarak koyun anlamına gelirken koymak fiilinin 2.çoğul şahıs çekimi olan koyun fiili anlamına da gelmektedir. Bu kelimenin cümle içinde hangi anlamda kullanıldığını anlayabilmek için cümle içerisinde kelimedenden önceki ve sonraki kelimelere bakarak anlayabiliriz. Örneğin, kitabı masanın üzerine koyun ve çoban koyun otlatıyor, cümlelerini ele alacak olursak masanın üzerine koyun derken koymak fiili olduğu ve çobanın koyun otlatması cümlesinde hayvan olan koyundan bahsettiği kelimedenden önceki ve sonraki kelimelerden anlaşılmaktadır. Bu gibi durumların çözümü için Hidden Markov modeli örnek gösterilebilir ve bu modelin çalışma mantığında ise olasılık hesaplamaları kullanılmaktadır. İngilizce dil yapısına göre değerlendirecek olursak bu dil üzerine yapılan çalışmalarda gramer yapısında belirteçten sonra yüzde 40 ihtimalle sıfat ve isim gelirken yüzde 20 ihtimal ile sayı gelmektedir. Bu sonuçlar bize gösteriyor ki bir belirtecin ardından fiil gelme ihtimali olmadığıdır. Örneğin, A lion has escaped from its cage. (Aslan kafesinden kaçtı.) cümlesinde escaped(kaçmak) fiil olarak kullanılırken Thees cape was planned very carefully. (Kaçış çok dikkatlice planlandı.) cümlesinde kaçmak fiili thees cape şeklinde bir belirteçten sonra isim halini almıştır.(ŞEKER Ş. E., POS Tagger (Metin Parçası Etiketleme), 2012)

4.5.2.4 SeedWords(Tohum Kelimeler) Yöntemi

SeedWords yöntemi aslında kısaca dışarıdan kelime ekleme yöntemidir. Tohum kelimeler yönteminin çalışma mantığını özetleyecek olursak duygu ifade etmek için sıkça kullanılan kelimelerin analizin daha düzgün sonuçlar vermesi için sisteme eklenmesine denmektedir. Yani sistemin analizi yapılacak bir konu hakkında gerekli bir kelimeyi öğrenememiş olduğunu varsayalım ve bu kelimenin öğrenilememesi analiz sonucuna yani yapılan test sonuçlarına etki etmeyecek böylelikle bu önemli kelimeler olmaması gereken bir sınıfta yer alacaktır. Bu da bizim test başarı oranımızın düşmesine yol açacaktır. Bu durumu önlemek için aynı toprağa bir tohum eker gibi kişinin kendi müdahalesi ile teste tabi tutulması gereken kelimeyi sisteme ekleyerek bu sorunu çözmüş ve daha doğru sonuçlar elde etmiş olacağız.(AKBA, 2014, s. 18-19)

4.5.3 SÖZLÜK TABANLI YÖNTEM

4.5.3.1 Öncül Yön Puanlandırma (Prior Polarity)

İnsanoğlu yaşamını sürdürebilmek için bir iletişim yöntemi bulmuş ve zamanla her topluluk kendi iletişim dilini oluşturmuştur. Günümüzde de ülkelerin kendi bir dili ve bu her dilin belli bir yapısı ve kuralı mevcuttur. İşte duygu analizi yöntemlerinden bir diğeri olan öncül yön puanlandırma sistemi de kelimelerin dil bilgisi kurallarına göre veya duygu içeriklerine göre puanlama yaparak sınıflandırmayı amaçlamaktadır. Bu puanlamayı gerçekleştirirken Pos-Tagger yöntemini kullanmaktadır. Bu yöntem kelimeleri fiil, isim, sıfat vb. türüne göre ve kelimelerin duygu içeriğine göre etiketlendirip hepsini tek tek puanlamaktadır. Bunun sonucunda her kelime ilişkilendirilip duygu puanları belirlendikten sonra her cümledeki kelimelerin sahip olduğu puan hesaplanır ve sonucunda cümlenin hangi duyguyu ifade ettiği belirlenir.

Öncül yön puanlandırma sisteminin başarılı bir sonuç vermesi için en başta geniş kapsamlı bir sözlüğe ihtiyaç vardır. İngilizce dili için geliştirilen ilk sözlük SentiWordNet sözlüğüdür. Bu sözlük Türkçe metinler üzerinde yapılacak analizler için de oluşturulmuştur ve örnek olarak da SentiTurkNet gösterilebilir. Bu yöntemde kullanılan kelimeleri temsil eden puanlama cetveli ne kadar geniş kapsamlı oluşturulursa bu yöntem o kadar fazla başarılı olacaktır.(AKBA, 2014, s. 33-34)

4.6 SINIFLAMA

Tanımsal olarak sınıflama elimizde bulunan verinin özelliğinin diğer verilerin nitelikleri ile karşılaştırma yapılarak belirlenmesine sınıflama işlemi denir. Sınıflama işlemini uygulamaya yönelik tanımlayacak olursak veriler arasında sınıflama işlemi yapılırken verilerin belli bir kısmı eğitim veri seti olarak kullanılırken diğer kısmı da test için kullanılmaktadır. Veri sınıflama işlemi iki yol ile gerçekleştirilir. Bunlar gözetimli ve gözetimsiz olarak ikiye ayrılır.

Gözetimli sınıflama daha önceden verilere ait değerlerin ve sınıfların bilindiği durumda Naive Bayes, k-NN vb. gibi algoritmalar ile yapılan analizin başarısının tespit edildiği durumdur. Gözetimsiz sınıflama ise mevcut veri üzerinde daha önceden sınıflandırma işleminin yapılmadığı durumlarda verinin yapısının tespit edilmesi için kullanılır.

4.7 KÜMELEME

Veri sınıflama yöntemlerinde biri olan kümeleme, sınıf sayısı ve veri türünün bilinmediği durumlarda yani gözetimsiz sınıflandırma için kullanılır. Kümeleme işlemi, elimizde mevcut bulunan bir grup değişkenin daha önceden belirlenen özelliklere göre kümelere ayrılma işlemidir. Bu sayede birbirine benzer niteliklere sahip değişkenler bir araya gelerek kümelenir ve bu da değişken sayımızı azaltarak veri setimizi daha iyi anlamamızı sağlar. Kümeleme işleminin sınıflandırmadan ayıran özelliği ise kümeleme gözetimsiz yani daha önceden verilerin hangi sınıfa ait olduğu bilinmezken sınıflandırmada ise tam tersi gözetimli sınıflama mevcuttur. Kümeleme işlemi yapılırken veriler arasındaki korelasyona yani uzaklığına bakılır ve birbirine daha yakın noktalarda biriken verilere sınırlar verilir ve kümeleme işlemi gerçekleştirilir. Veriler arasındaki uzaklık belirlenirken gaussian gibi uzaklık ölçüm yöntemleri kullanılır.

4.8 İLİŞKİLENDİRME

Veri sınıflama yöntemlerinden bir tanesi de ilişkilendirme yöntemidir. İlişkilendirme yöntemi, gözetimli olarak sınıflandırma yapar yani ilişkilendirmede de veriler hakkında önceden sınıf ve veri özellikleri gibi bilgiler mevcuttur ve ona göre sınıflandırma işlemi gerçekleştirilir. İlişkilendirme yönteminde çoğunlukla etiketli verilerde test ve eğitim veri setleri arasında inceleme yapılmaktadır.(AKBA, 2014, s. 10-11)

4.9 MODEL BAŞARISINI ÖLÇME YÖNTEMLERİ

4.9.1 Kesinlik Ölçütü

Deney sonucunda yapılan ölçümlerin birbirini ne derece destekler nitelikte olduğunu göstermektedir. Kesinlik değeri hesaplanırken doğru etiketlenmiş pozitif veri sayısının toplam pozitif veri sayısına bölümü ile elde edilir. Kesinlik değeri her zaman sıfır ile bir arasında sonuç vermelidir.(KAYNAR, GÖRMEZ, YILDIZ, & ALBAYRAK, 2016)

$$\text{Kesinlik} = \frac{TP}{TP+FP} \quad (8.10)$$

4.9.2 Duyarlılık Ölçütü

Duyarlılık ölçütü analiz sonucunda gerçekte ulaşmak istediğimiz sonuca ne kadar yaklaşabildiğimizi, hedefimizi ne kadar tutturabildiğimizi göstermektedir. Duyarlılık değeri hesaplanırken doğru etiketlenmiş pozitif veri sayısının gerçekten pozitif olan veri sayısına bölünmesi ile oluşur.

$$\text{Duyarlılık} = \frac{TP}{TP+FN} \quad (8.11)$$

4.9.3 Doğruluk Ölçütü

Doğruluk ölçütü, yapılan analiz sonucunda ortaya çıkan sonucun gerçek yani asıl sonuca ne kadar yakın olup olmadığını gösteren bir ölçüttür. Ayrıca doğruluk ölçütü sınıflandırma yönteminde en çok kullanılan ölçüttür. Doğruluk ölçütü hesaplanırken doğru olarak sınıflandırılmış örneklerin toplam örnek sayısına bölünmesi ile elde edilir.

$$\text{Doğruluk} = \frac{TP+TN}{TP+TN+FP+FN} \quad (8.12)$$

4.9.4 F Ölçütü

F ölçümü yapılmış olan testin anlamlılığı hakkında bize bilgi vermektedir. F ölçütü kesinlik ve duyarlılık ölçütlerinin harmonik ortalaması alınarak hesaplanan bir ölçüttür.

$$\text{F Ölçütü} = \frac{2 \cdot \text{Duyarlılık} \cdot \text{Kesinlik}}{\text{Kesinlik} + \text{Duyarlılık}} \quad (8.13)$$

Yukarıda formülüzasyonda kullanılan TP, TN, FP, FN ile ilgili açıklamaları şu şekildedir;

TP= Hem pozitif olan hem de pozitif sınıflandırılmış veriler

TN= Gerçekte negatif olan ve negatif sınıflandırılmış veriler

FP=Gerçekte pozitif olan ancak pozitif olarak sınıflandırılmamış veriler

FN= Gerçekte negatif olan ancak negatif olarak sınıflandırılmamış veriler

5 BEŞİNCİ BÖLÜM

5.1 RAPİDMİNER

Program 2001 yılında YALE ismi ile yapay zeka birimleri tarafından oluşturulmuştur. RapidMiner 2006 yılında Rapid-I adlı bir şirket tarafından geliştirilmiş ve 2013 yılında ise programın adı YALE yerine RapidMiner olarak değiştirilmiştir. RapidMiner analiz programı Java programlama dili ile yazılmış açık kaynak kodlu bir analiz platformudur. RapidMiner programı, veri madenciliği yöntemlerinin tamamını desteklemesinden dolayı bu ve benzeri alanlarda yapılacak uygulamalarda genellikle tercih edilmektedir. RapidMiner programı, R ve Python gibi kod yazma işlemlerine gerek duymamasından dolayı kullanıcılar tarafından daha fazla tercih edilmektedir.(ALTINOK, 2017) RapidMiner analiz programı gibi işlemleri gerçekleştiren açık kaynak kodlu farklı platformlar da mevcuttur. Bu platformlar weka, knime ve R gibi programlar ile örneklendirilebilir. RapidMiner makine öğrenmesi, metin madenciliği ve iş analizi gibi hedeflere yönelik geliştirilmiş bir yazılım uygulamasıdır.

Literatürde veri madenciliği çalışmalarında en çok tercih edilen programlar RapidMiner, Weka ve Knime analiz programlarıdır. Bu programlar arasında bir karşılaştırma yapmak gerekirse RapidMiner ile wekayı karşılaştırdığımızda RapidMiner, 22 civarında dosya formatı desteklerken Weka ise 4 farklı dosya formatı desteklemektedir. RapidMiner programının wekadan bir diğer avantajı ise RapidMiner, Weka da mevcut olan modeller ve algoritmaları içerisinde barındırmaktadır. Ayrıca RapidMiner, veri görselleştirme açısından Weka programına göre daha gelişmiş bir ara yüze sahiptir. Knime programının diğer iki programdan farkı ise metin madenciliği işlemlerini bir modül kullanarak yapmasıdır. RapidMiner analiz programı toplamda 129 algoritmayı içerisinde bulundururken, Knime 102 algoritma ve Weka ise 76 algoritmayı içerisinde barındırmaktadır. RapidMiner, Knime ve Weka programlarının ortak

özellikleri ise hepsinin Java yazılım dili ile yazılması ve veri analizi, veri görselleştirme, veri ön işleme özelliklerinin bu üç programda da mevcut olmasıdır.(ŞEKER & YILDIZ, Veri Madenciliği Araçları(Data Mining Tools), 2016, s. 18-19)

5.2 METODOLOJİ

Çalışmanın analiz aşamasında ilk olarak daha önceden belirlenen akıllı ulaşım sistemleri, trafik yoğunluğu, otonom araç gibi anahtar kelimeler Twitter sosyal ağı üzerinde RapidMiner programı ile taratılmıştır. Bu taramanın yapılabilmesi için Twitter Application adı verilen anahtar koda ihtiyaç duyulmaktadır. Bu kodun alınmasında sonra RapidMiner programı ile belirlenen konular üzerinde tarama yapılmıştır. Çıkan sonuçlar ilk önce veri temizleme işleminden geçirilerek sonuç üzerinde etkili olmayacak verilere temizleme işlemi uygulanmıştır. Daha sonra duygu analizi işlemi uygulanmış ve çıkan sonuçlar yorumlanmıştır.

5.2.1 Verilerin Toplanması

Veriler RapidMiner programı kullanılarak twitter sosyal ağı üzerinden elde edilmiştir. Veriler elde edilirken twitter sosyal ağının tercih edilme sebepleri çok fazla kullanıcıyı içerisinde barındırması, çeşitli kültürel özelliklere ve duygulara sahip bireyleri bir arada bulundurması ve gündemin twitter sosyal ağı tarafından ölçülmesinin kolay olması tercih edilme sebebi olarak sıralanabilir. Twitter sosyal medya uygulaması üzerinden veri elde edilirken önceden belirlenmiş anahtar kelimeler kullanılmış ve bu kelimeleri içeren tweetler RapidMiner programı yardımıyla taratılmıştır. Tarama işleminden sonra elde edilen sonuçlara duygu analizi tekniği uygulanmıştır.

5.2.2 Veri Toplama Süreci



Şekil 17: Twitter ve AYLIEN operatörleri bağlantı platformları

Veri toplama sürecinin aşamaları yukarıda yüzeysel olarak nasıl gerçekleştiği anlatılmıştır. Bu aşamada ise RapidMiner programı üzerinde veri toplama sürecinin nasıl gerçekleştirildiği programın içerisinden alınan görseller ile uygulamalı olarak anlatılmaktadır. Uygulama üzerinde veri toplama süreci üç aşamada gerçekleşmektedir. İlk aşamada programın ara yüzüne gerekli parametreler eklenir. Bu aşamada twitter üzerinde veri taraması yapabilmek “Twitter Search” parametresi eklenmektedir. İkinci aşamada ise eklenen parametrelerin twitter ile bağlantı kurabilmesi için herhangi bir twitter kullanıcı hesabı ile Twitter Application üyeliği oluşturularak program ile twitter uygulaması arasında bağlantı kurulur. Son aşamada ise oluşturulan üyelik üzerinden “Application ID” ve “Application Key” anahtar şifreleri alınarak programın ara yüzünde bulunan parametrelere bu şifreler girilir ve twitter üzerinde veri tarama işlemi başlatılmaktadır.

Text
"Architects can address all sorts of challenges in an intelligent, holistic way, from transportation issues to homelessness, to environmental issues, education problem...
THINK BIGGER!!! The New Intelligent Transportation System Will Offer Smart Roads for the Smart Cars: "A Road So Smart It Will Be Able to Charge Your Car"
RT @PIARC_Roads: [#ITS] #PIARC Technical Committee B.1 on #Road Network Operations/Intelligent Transportation Systems meets in #Vienna (Aus...
Smart road: China is building serious momentum in intelligent transportation https://t.co/VWLkbJy0Ps
RT @FinancialXpress: Smart road: China is building serious momentum in intelligent transportation
Smart road: China is building serious momentum in intelligent transportation
Field Engineer - Intelligent Transportation System (ITS) - Houston, TX https://t.co/9oN1Ka3LLN
I could not be more excited about this year's #Plugfest! Monday, May 7, features a Connected Vehicle Workshop; keynotes include Kirk Stuedle, Director, Michigan Dep...
RT @_skconnor: AI is revolutionizing nearly every industry - from transportation, to finance, to healthcare. Read how small businesses are...
RT @ajitbirsingh: "The road to China's autonomous-driving future is paved with solar panels, mapping sensors and electric-battery recharger..."
RT @PIARC_Roads: [#ITS] #PIARC Technical Committee B.1 on #Road Network Operations/Intelligent Transportation Systems meets in #Vienna (Aus...
RT @Cuesta_Campos: From Vienna, lunch meeting at Kapsch TrafficCom HQ with Alexandra & Philipp from Legal. Our client is a renowned provide...
From Vienna, lunch meeting at Kapsch TrafficCom HQ with Alexandra & Philipp from Legal. Our client is a renowned provider of intelligent transportation systems, wit...
"The road to China's autonomous-driving future is paved with solar panels, mapping sensors and electric-battery rechargers as the nation tests an 'intelligent highwa...
The road to China's autonomous-driving future is paved with #solar panels, mapping sensors & electric-battery rechargers as the nation tests an "intelligent highway"
The road to China's autonomous-driving future is paved with #solar panels as the nation tests an "intelligent highway" that could speed the transformation of the glob...

Şekil 18: Twitter sosyal ağı üzerinde tarama yapılarak elde edilmiş tweet örneği

Veri tarama işlemi sonucunda örnekte görüldüğü üzere belirlenen konu ile ilgili tweetler sosyal medya üzerinde tarama işlemi gerçekleştirildikten sonra bir araya getirilmiştir. Twitter üzerinde yapılan bu tarama işlemi haftalık olarak tekrarlanmış ve konuyla ilgili atılan tweetler bir araya toplanarak duygu analizi evreleri uygulanmıştır.

Parameters ✕

Search Twitter

connection ?

query

result type ?

limit ?

Parameters ✕

Analyze Sentiment

connection ?

input attribute ?

sentiment mode ?

Is input URL ?

Şekil 19: Duygu analizi ve twitter parametrelerinin özellikleri

Şekil(19)'de RapidMiner programı ara yüzüne eklenen “Twitter Search” parametresi ve duygu analizi işlemini gerçekleştiren “Sentiment Analyze” parametresi yer almaktadır. Twitter Search parametresinin içerisinde bulunan özellikleri kısaca açıklayacak olursak “Connection” kısmında “Twitter Application” uygulaması tarafından elde edilen şifreler ile kurulan bağlantı gösterilmektedir. Özellikler arasındaki “Query” seçeneği ise twitter üzerinde taratılmak istenen konunun belirtildiği alandır. “Result Type” anlamı sonuç türü olan özellik içerisinde ise üç farklı seçenek bulunmaktadır. Bu seçenekler ise “Recent”, “Popular” ve “Recent or Popular” seçenekleridir. Recent seçeneği ile konu ile alakalı yakın zamanda atılan tweetler taratılırken, “Popular” seçeneği ise konu ile alakalı atılan popüler tweetleri, “Recent or Popular” seçeneği ile hem popüler olan hem de yakın tarihte atılmış tweetler taratılmaktadır. Son olarak limit kısmında ise twitter üzerinde konu ile alakalı taraması yapılacak maximum tweet sayısı belirlenmektedir.

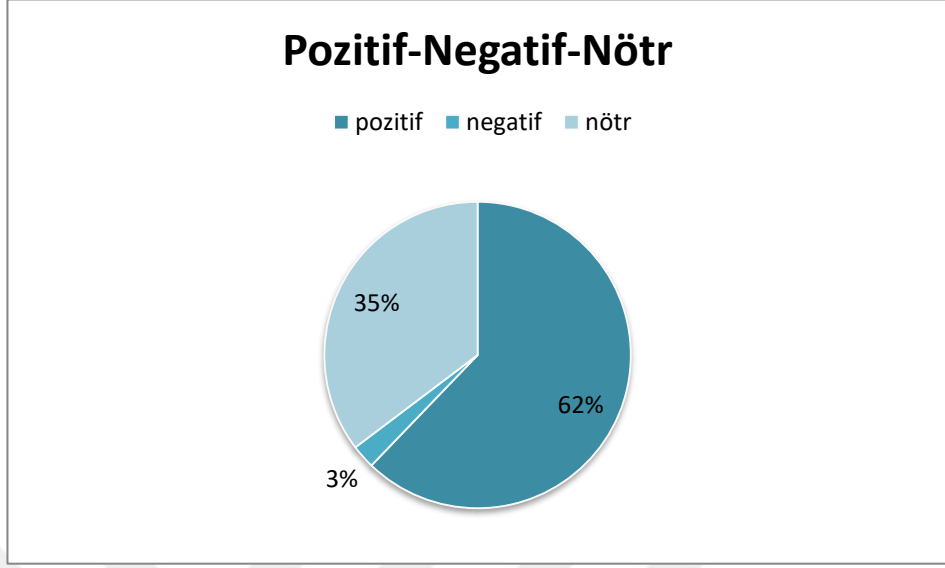
“Sentiment Analyze” operatörü ile duygu analizi işlemleri gerçekleştirilmektedir. “Sentiment Analyze” operatörü RapidMiner programı içerisinde kurulu olmayan bir operatördür. Fakat RapidMiner programı içerisine AYLIEN Text Analysis paket programı indirilerek duygu analizi operatörü ara yüze eklenebilir. “Sentiment Analyze” operatörü içerisindeki özellikleri inceleyecek olursak “Connection” kısmında bağlantı sağlanabilmesi için AYLIEN’in sitesinden üyelik oluşturularak uygulama ismi ve uygulama şifreleri elde edilir ve program ile bağlantı kurulur. “Input Attribute” ve “Sentiment Mode” kısımlarında ise duygu analizi yapılacak kaynak seçilmektedir.

5.3 AUS Hakkında Yapılan Yorumların Duygu Analizi Sonuçları

Row No.	Id	polarity_confidence	subjectivity_confidence	polarity	subjectivity	Created-At	From-User
1	9852016517...	0.653	0.575	positive	subjective	Apr 14, 2018 ...	riyanto
2	9851719810...	0.653	0.575	positive	subjective	Apr 14, 2018 ...	Mae_denz
3	9851186741...	0.653	0.575	positive	subjective	Apr 14, 2018 ...	NarGanie
4	9851020859...	0.653	0.575	positive	subjective	Apr 14, 2018 ...	Almubarraq
5	9851009330...	0.551	0.982	neutral	objective	Apr 14, 2018 ...	Svetlana
6	9850828241...	0.653	0.575	positive	subjective	Apr 14, 2018 ...	Deckstablest..
7	9850685050...	0.653	0.575	positive	subjective	Apr 14, 2018 ...	анастасия
8	9850552917...	0.653	0.575	positive	subjective	Apr 14, 2018 ...	Sayang
9	9850534293...	0.653	0.575	positive	subjective	Apr 14, 2018 ...	Ekram Hasan
10	9850517592...	0.653	0.575	positive	subjective	Apr 14, 2018 ...	COSMOT90#...
11	9849919263...	0.653	0.575	positive	subjective	Apr 14, 2018 ...	maya
12	9849735259...	0.653	0.575	positive	subjective	Apr 14, 2018 ...	fara_buduk
13	9848713664...	0.653	0.575	positive	subjective	Apr 13, 2018 ...	Anna Pushkar
14	9848707182...	0.653	0.575	positive	subjective	Apr 13, 2018 ...	Angelica
15	9848560456...	0.653	0.575	positive	subjective	Apr 13, 2018 ...	Denis Serafi...
16	9848463358...	0.653	0.575	positive	subjective	Apr 13, 2018 ...	Chiki Mya

Şekil 20: Akıllı ulaşım sistemleri hakkında atılan tweet'lerin duygu analizi sonuçları

Görselde her bir kullanıcıya ait twitter sosyal ağı üzerinde akıllı ulaşım sistemleri ile ilgili atılan tweetler ve bu metinlere ait duygu analizi sonuçları yer almaktadır. Akıllı ulaşım sistemleri hakkında atılan tweetler önceden veri temizleme işlemine tabi tutularak sonuca etkisi olmayan “@, http, #” gibi simgeler metinlerin içerisinden çıkartılmış ve duygu analizi tekniği uygulanmıştır. Duygu analizi sonucunda metinlerin duygu düzeyleri, yorumların öznel ve nesnel olma durumları, atılan tweetin tarihi ve kullanıcılar ile alakalı bilgiler bulunmaktadır. Analiz sonucunu incelediğimizde polarity kısmı tweetlerin pozitif, negatif ve nötr olma durumlarını gösterirken, subjectivity ise öznellik durumu hakkında sonuçları göstermektedir.



Şekil 21: Akıllı ulaşım sistemleri duygu analizi sonuç grafiği

Bu görselde akıllı ulaşım sistemleri hakkında tüm kullanıcılara ait atılan tweetlerin duygu analizi sonuçları yer almaktadır. AUS hakkında atılan tweetlerin duygu analizi sonuçları yüzdeler olarak hesaplanmış ve grafik üzerinde gösterilmiştir. Analiz sonuçlarını incelediğimizde olumlu tweet sayısı 74, olumsuz tweet sayısı 3, tarafsız tweet sayısı 42 olmak üzere toplamda 119 tweet'e ulaştık. Sonuçlarımızı grafik içerisine aktardığımızda AUS hakkında atılan tweetlerin %62'si pozitif, %3'ü negatif ve %35'inin ise nötr bir yaklaşım sergilediğini görmekteyiz. Twitter kullanıcılarının akıllı ulaşım sistemleri hakkında attıkları tweetlere ait analiz sonuçlarını yorumladığımızda kullanıcıların genelinin bu konu hakkında olumlu bir görüşe sahip olduğunu görmekteyiz.



Şekil 22: Akıllı ulaşım sistemleri öznellik sonuçları

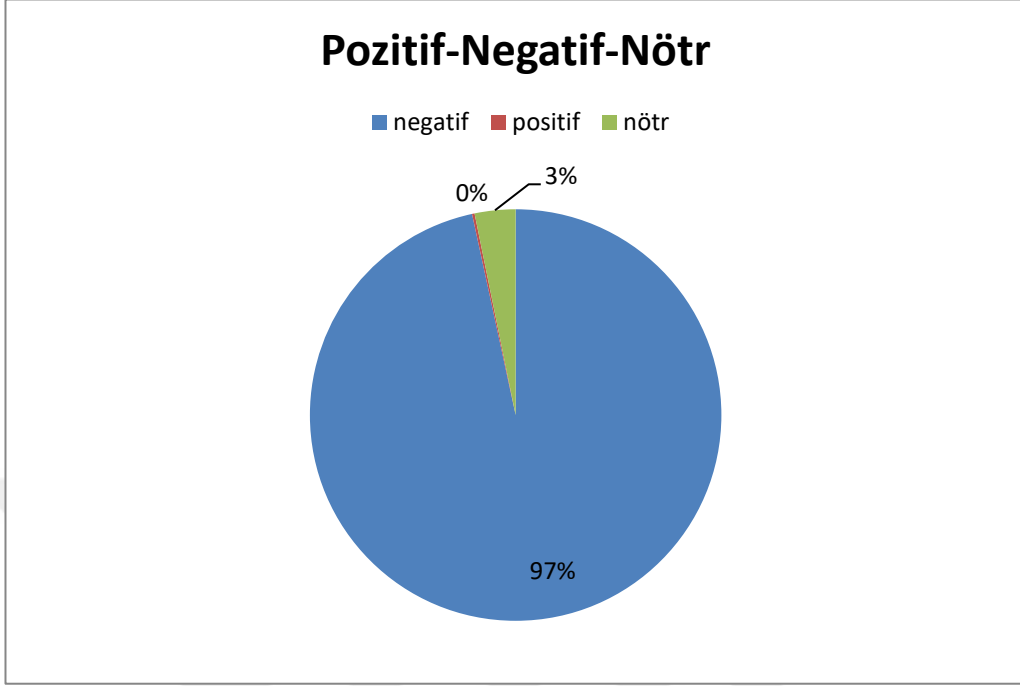
Grafikte kullanıcıların konu ile alakalı atılan tweetlerin duygu düzeyleri değil sadece yaptıkları yorumların öznel ve nesnel olma durumları incelenmektedir. Twitter kullanıcılarının AUS ile ilgili paylaştıkları tweetleri öznel veya nesnel olma durumlarına göre incelediğimizde paylaşılan tweetlerin 83 adetinin öznel yorumlardan ve 36 adetinin ise nesnel yorumlardan oluştuğunu görmekteyiz. Duygu analizi sonuçlarını yüzdeler olarak hesapladığımızda %70'inin öznel yorumlardan oluştuğunu ve %30'unun ise nesnel yorumlardan oluştuğunu görmekteyiz.

5.4 Trafik Yoğunluğu Hakkında Paylaşılan Tweetlerin Duygu Analizi Sonuçları

Row No.	Id	polarity_con...	subjectivity_...	polarity	subjectivity	Created-At	From-User	From-User-Id
1	9847619099...	0.979	1	negative	subjective	Apr 13, 2018 ...	Hassaan Niazi	131094666
2	9844788544...	0.748	1.000	negative	objective	Apr 12, 2018 ...	Mehr Tarar	588587397
3	9818711294...	0.491	1	neutral	subjective	Apr 5, 2018 3:...	Ali-A	155620461
4	9852316203...	0.976	1.000	neutral	objective	Apr 14, 2018 ...	Wahono Kolo...	165109538
5	9852313273...	0.952	1.000	neutral	objective	Apr 14, 2018 ...	Old Man Stuart	52622109
6	9852312602...	0.929	1.000	neutral	objective	Apr 14, 2018 ...	Beat Squad ...	3301910092
7	9852309643...	0.729	1	neutral	subjective	Apr 14, 2018 ...	Piehole Need...	16167428
8	9852303860...	0.945	1.000	neutral	objective	Apr 14, 2018 ...	Jonathan Hall	43435534
9	9852298281...	0.581	0.796	neutral	objective	Apr 14, 2018 ...	ꠘꠘꠘꠘꠘꠘ ꠘꠘ...	509103872
10	9852298068...	0.525	1	positive	subjective	Apr 14, 2018 ...	Chandler	1029321258
11	9852297598...	0.930	1	neutral	subjective	Apr 14, 2018 ...	Jill Ragan	3125515909
12	9852295922...	0.570	1	positive	subjective	Apr 14, 2018 ...	Suresh Kumar	8704824944...
13	9852288597...	0.721	1.000	neutral	objective	Apr 14, 2018 ...	ꠘꠘꠘꠘꠘꠘ	113629111
14	9852286082...	0.493	1.000	negative	objective	Apr 14, 2018 ...	Jacques Oeuf	505746509
15	9852282659...	0.974	1.000	neutral	subjective	Apr 14, 2018 ...	Marcelo Inácio	1036277450

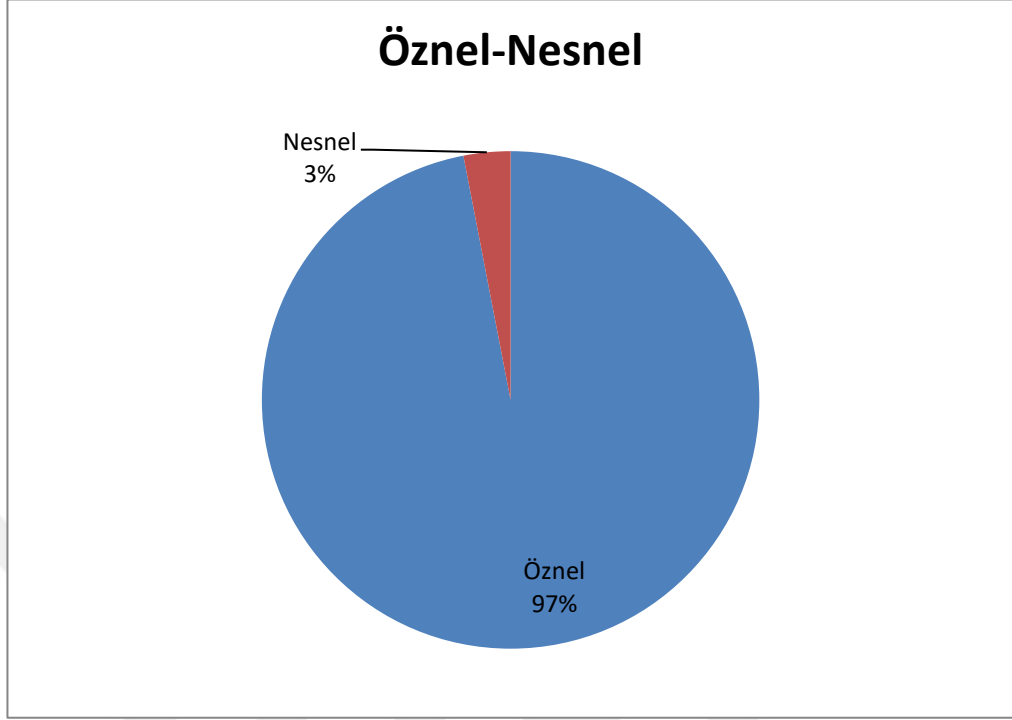
Şekil 23: Trafik yoğunluğu hakkında atılan tweet'lerin duygu analizi sonuçları

Trafik yoğunluğunu konusunu anahtar kelime olarak tercih etmemizin ve analize dahil etmemizin sebebi, bu konu hakkında bireylerin düşüncelerinin ne yönde olduğunu ve bireylerin günlük hayatına nasıl bir etkisi olduğunu ölçmek istememizdir. Analiz sonucunda kullanıcılarının trafik yoğunluğu hakkında paylaştıkları tweetlerin duygu analizi katsayıları, hesaplanan her bir katsayının olumlu, olumsuz ve tarafsız olmak üzere duygu düzeyleri, öznel ve nesnel olma durumları ile alakalı öznellik katsayıları ve bunların düzeyleri görülmektedir. Ayrıca konu ile alakalı yorumda bulunan kullanıcıların isimleri ve tweet attıkları tarih gibi kullanıcılara ait bilgileri de görmekteyiz.



Şekil 24: Trafik yoğunluğu duygu analizi sonuç grafiği

Twitter üzerinde yapılan tarama sonucunda trafik yoğunluğu konusu ile ilgili atılan toplam 2499 adet tweet'e ulaştık. Konu ile alakalı atılan tweetlerin 2414 tanesinin negatif yorumlardan oluştuğunu, 5 tanesinin pozitif yorumlardan oluştuğunu ve 80 tanesinin de nötr(tarafsız) yorumlardan oluştuğunu görmekteyiz. Analiz sonucunda trafik yoğunluğu hakkında paylaşılan tweetlerin %97'sinin negatif, %3'ünün nötr bir tutum sergilediğini ve trafik yoğunluğu hakkında neredeyse olumlu anlamda bir görüşün bulunmadığını görmekteyiz. Sonuçları yorumladığımızda kullanıcıların büyük bir kısmının trafik yoğunluğu hakkında olumsuz bir görüşe sahip olduğunu istatistiksel anlamda da kanıtlamış bulunmaktayız.



Şekil 25: Trafik yoğunluğu öznel sonuçları

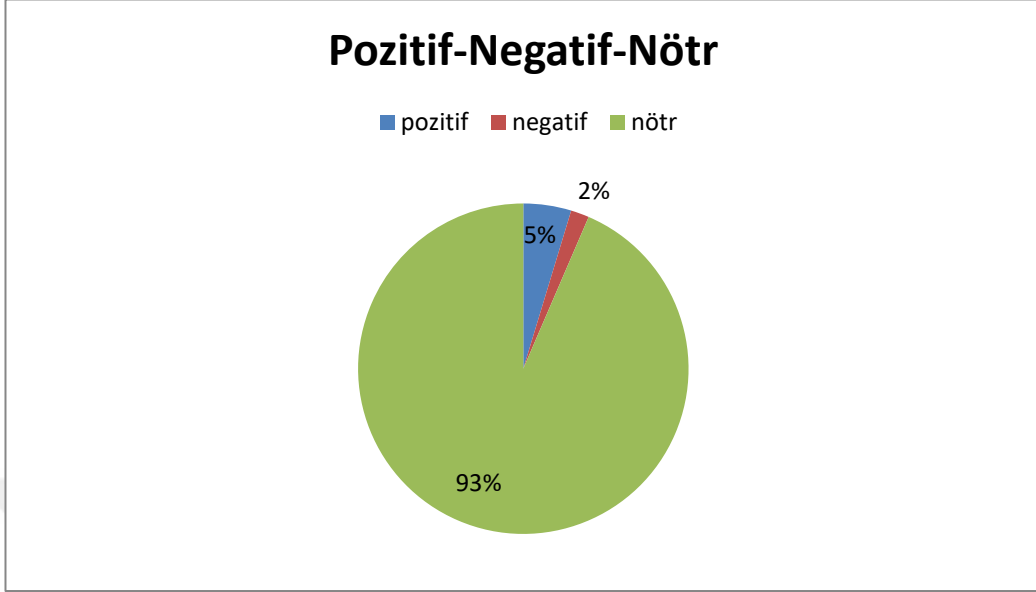
Trafik yoğunluğu hakkında öznel katsayılarının hesaplanarak belirlenmiş olduğu tweetleri öznel veya nesnel olma durumlarına göre yorumladığımızda atılan tweetlerin 2423 adetinin öznel yorumlardan oluştuğunu 76 adetinin ise nesnel yorumlardan oluştuğunu görmekteyiz. Analiz sonuçlarını yüzdeler olarak yorumladığımızda %97'sinin öznel ve %3'ünün ise nesnel yorumlardan oluştuğunu görmekteyiz. Analiz sonucunda ulaştığımız bu verileri yorumladığımızda bireylerin trafik yoğunluğu hakkında yaptıkları yorumların herkes tarafından bilinen görüşlerden ziyade genellikle kişisel düşüncelerinden oluştuğunu görmekteyiz.

5.5 Otonom Araçlar Hakkında Paylaşılan Tweetlerin Duygu Analizi Sonuçları

Row No.	Id	polarity_con...	subjectivity_...	polarity	subjectivity	Created-At	From-User	From-User-Id
1	9858126390...	0.968	0.998	neutral	objective	Apr 16, 2018 ...	Channel New...	38400130
2	9857798336...	0.849	0.971	neutral	objective	Apr 16, 2018 ...	Global Times	49616273
3	9858423862...	0.857	1.000	neutral	objective	Apr 16, 2018 ...	dena	25086173
4	9858683081...	0.862	0.991	neutral	subjective	Apr 16, 2018 ...	Idea Quarter	3335825441
5	9858679730...	0.992	1.000	neutral	objective	Apr 16, 2018 ...	New Gizmo B...	1357334300
6	9858669653...	0.960	1.000	neutral	objective	Apr 16, 2018 ...	Jim Campbell	19958841
7	9858662700...	0.726	1.000	neutral	objective	Apr 16, 2018 ...	EUREF-Cam...	1978920655
8	9858655887...	0.950	1.000	neutral	subjective	Apr 16, 2018 ...	OCE	38190216
9	9858654952...	0.929	1.000	neutral	objective	Apr 16, 2018 ...	Carthage Trib...	422215906
10	9858649786...	0.933	0.895	neutral	objective	Apr 16, 2018 ...	Pam Gordon ...	2884564738
11	9858649697...	0.901	0.859	negative	subjective	Apr 16, 2018 ...	B.J. Wade	2409371190
12	9858640216...	0.901	0.981	negative	subjective	Apr 16, 2018 ...	Sean Herndon	20793622
13	9858625557...	0.984	1	neutral	subjective	Apr 16, 2018 ...	Olivier Soula	219379601
14	9858616684...	0.872	1.000	neutral	objective	Apr 16, 2018 ...	GX Group	1118765467
15	9858610888...	0.984	1	neutral	subjective	Apr 16, 2018 ...	Lucy Killip	4561882462
16	9858606649...	0.902	1.000	neutral	subjective	Apr 16, 2018 ...	Two Ex-Goog...	9087324910...

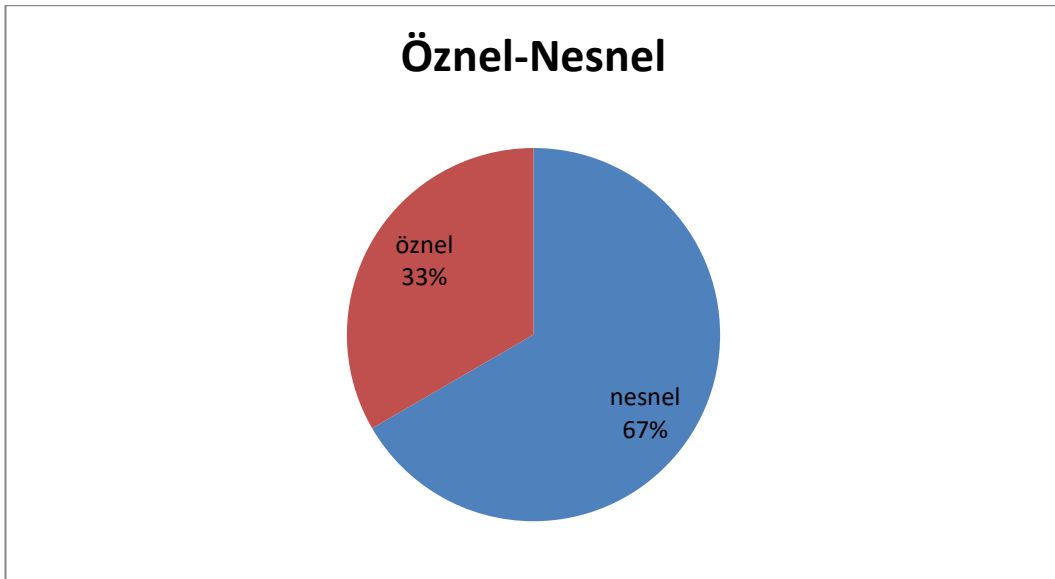
Şekil 26: Otonom araçlar hakkında atılan tweet'lerin duygu analizi sonuçları

Otonom araç teknolojisi akıllı ulaşım sistemleri ile bağlantılı olan ve akıllı ulaşım sistemlerine entegre edilebilecek bir teknolojidir. Otonom araç konusunu araştırmamızdaki sebep bireylerin bu konu ile alakalı görüşlerinin ne yönde olduğunu görmek ve bu konu üzerinden bireylerin akıllı ulaşımına olan ilgisini tespit etmektir. Görselde twitter kullanıcılarının otonom araçlar hakkında paylaştıkları tweetlerin duygu analizi katsayıları, duygu düzeyleri, öznellik katsayıları ve düzeyleri ve kullanıcılara ait bilgiler yer almaktadır.



Şekil 27: Otonom araç duygu analizi sonuç grafiği

Twitter sosyal ağı üzerinden otonom araçlar ile ilgili atılan toplam 1004 adet tweete ulaşılmıştır. Atılan bu tweetlerin 47 tanesinin pozitif yorumlardan oluştuğu, 18 adetinin olumsuz yorumlardan oluştuğu ve 939 adetinin ise tarafsız yorumlardan oluştuğunu tespit ettik. Analiz sonuçları grafik üzerinde değerlendirdiğimizde paylaşılan tweetlerin %5'inin pozitif, %2'sinin negatif ve %93'ünün ise nötr olduğu tespit ettik.



Şekil 28: Otonom araç öznellik sonuçları

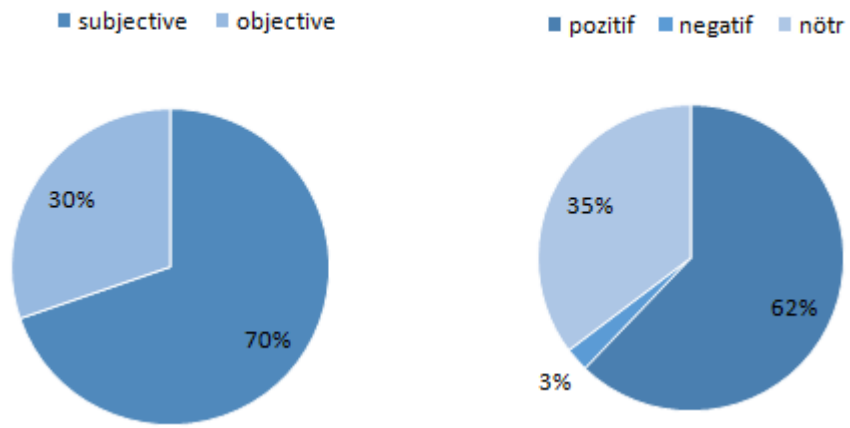
Grafikte subjectivity katsayı hesaplanarak bireylerin otonom araç hakkındaki yorumlarının öznel ve nesnel olma durumları incelenmiştir. Kullanıcıların otonom araç ile ilgili paylaştıkları tweetleri öznel veya nesnel olma durumlarına göre değerlendirdiğimizde paylaşılan tweetlerin 335 adetinin öznel yorumlardan oluştuğunu 669 adetinin ise nesnel yorumlardan oluştuğunu görmekteyiz. Analiz sonuçlarını yüzdeler olarak incelediğimizde %33'ünün öznel yorumlardan ve %67'sinin ise nesnel yorumlardan oluştuğunu görmekteyiz. Ulaştığımız bu sonuçlara göre bireylerin kişisel yorumlardan ziyade genel kitle tarafından kabul görmüş yorumlarda bulunduğunu görmekteyiz.

6 ALTINCI BÖLÜM

6.1 BULGULAR

Çalışmada kullanıcıların önceden belirlenen konular hakkındaki düşüncelerinin ne yönde olduğunu inceledik. Analiz aşamasında twitter üzerinde AUS ile ilgili paylaşılan 119 tweete trafik yoğunluğu ile ilgili paylaşılan 2499 tweete ve otonom araçlar ile ilgili paylaşılan 1004 adet tweete ulaştık. İlk olarak akıllı ulaşım sistemleri hakkında atılan tweetlerin duygu analizi sonuçlarını inceledik.

Öznel-Nesnel Duygu Analizi Pozitif-Negatif-Nötr

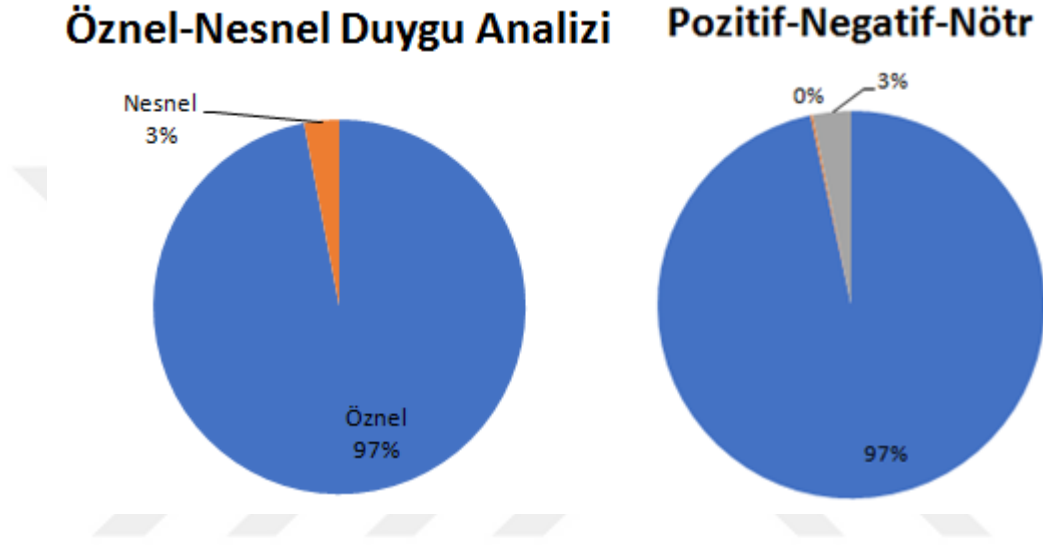


Şekil 29: Akıllı ulaşım sistemleri Duygu ve Öznellik analizi sonuçları

Duygu analizi sonuçlarını değerlendirdiğimizde kullanıcıların “Akıllı Ulaşım Sistemleri” ile ilgili paylaştıkları tweetlerin %62'sinin pozitif, %3'ünün negatif ve

%35'inin ise nötr bir yaklaşım sergilediğini tespit ettik. Twitter üzerinde AUS ile ilgili paylaşılan yorumları öznel ve nesnel olma durumlarına göre incelendiğinde %70'inin öznel ve %30'unun nesnel yorumlardan oluştuğu görülmektedir.

Twitter üzerinde trafik yoğunluğu hakkında atılan tweetleri duygu analizine tabi tuttuk.

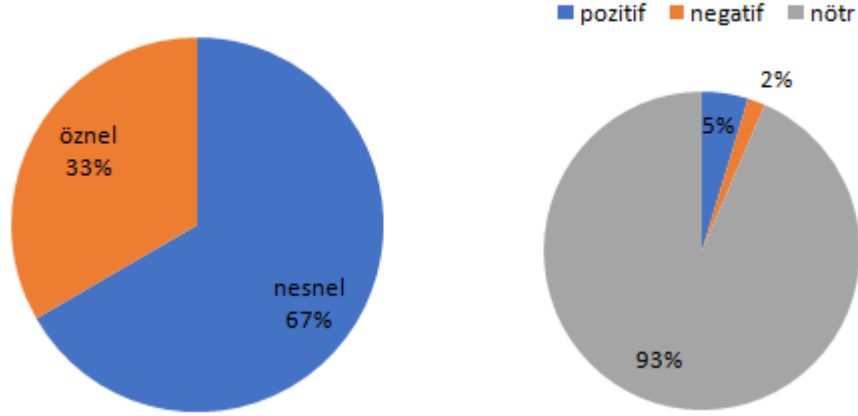


Şekil 30: Trafik yoğunluğu sistemleri duygu ve öznellik analizi sonuçları

Trafik yoğunluğu ile ilgili paylaşılan tweetleri incelediğimizde ise %97'sinin negatif, %3'ünün nötr bir yaklaşım sergilediğini ve bunun sonucunda trafik yoğunluğu ile ilgili hiç olumlu bir görüşün bulunmadığı tespit ettik. Trafik yoğunluğu ile ilgili paylaşılan yorumları öznel ve nesnel olma durumlarına göre incelendiğinde %97'sinin öznel ve %3'ünün ise nesnel yorumlardan oluştuğu sonucuna ulaşılmıştır.

Son aşamada twitter üzerinde otonom araçlar hakkında atılan yorumları inceledik.

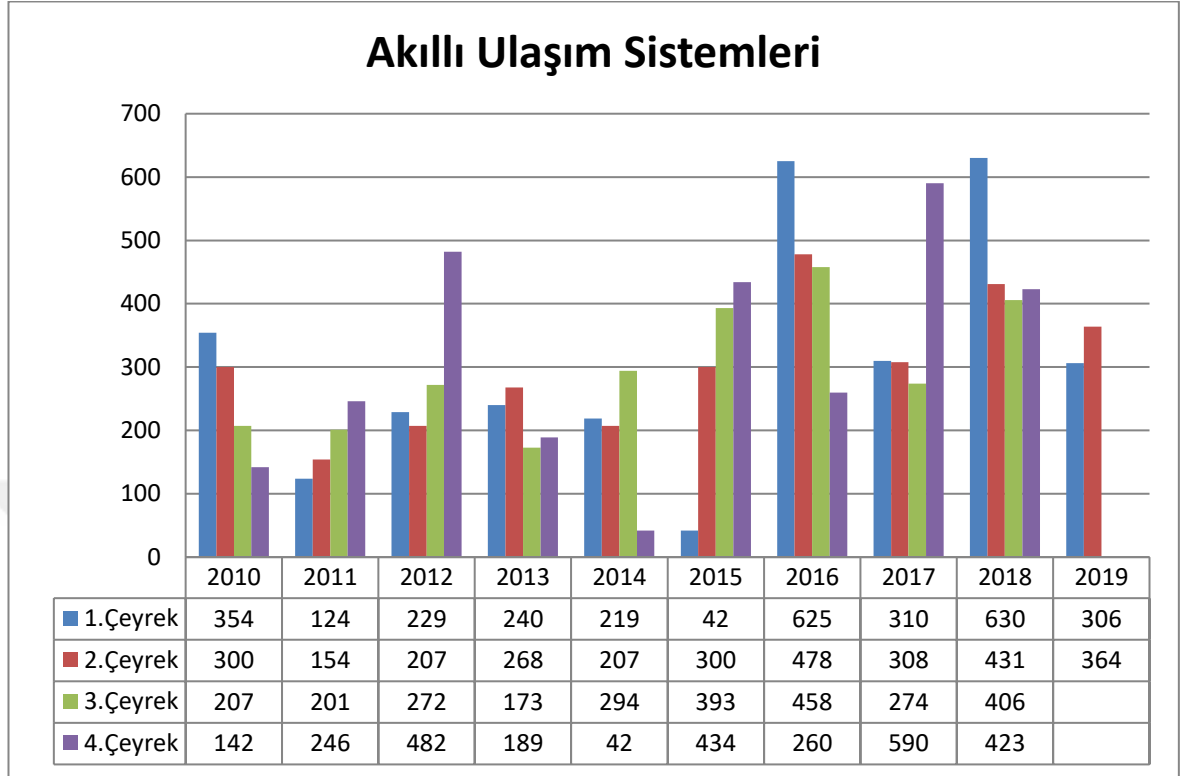
Öznel-Nesnel Duygu Analizi Pozitif-Negatif-Nötr



Şekil 31: Otonom araç duygu ve öznellik analizi sonuçları

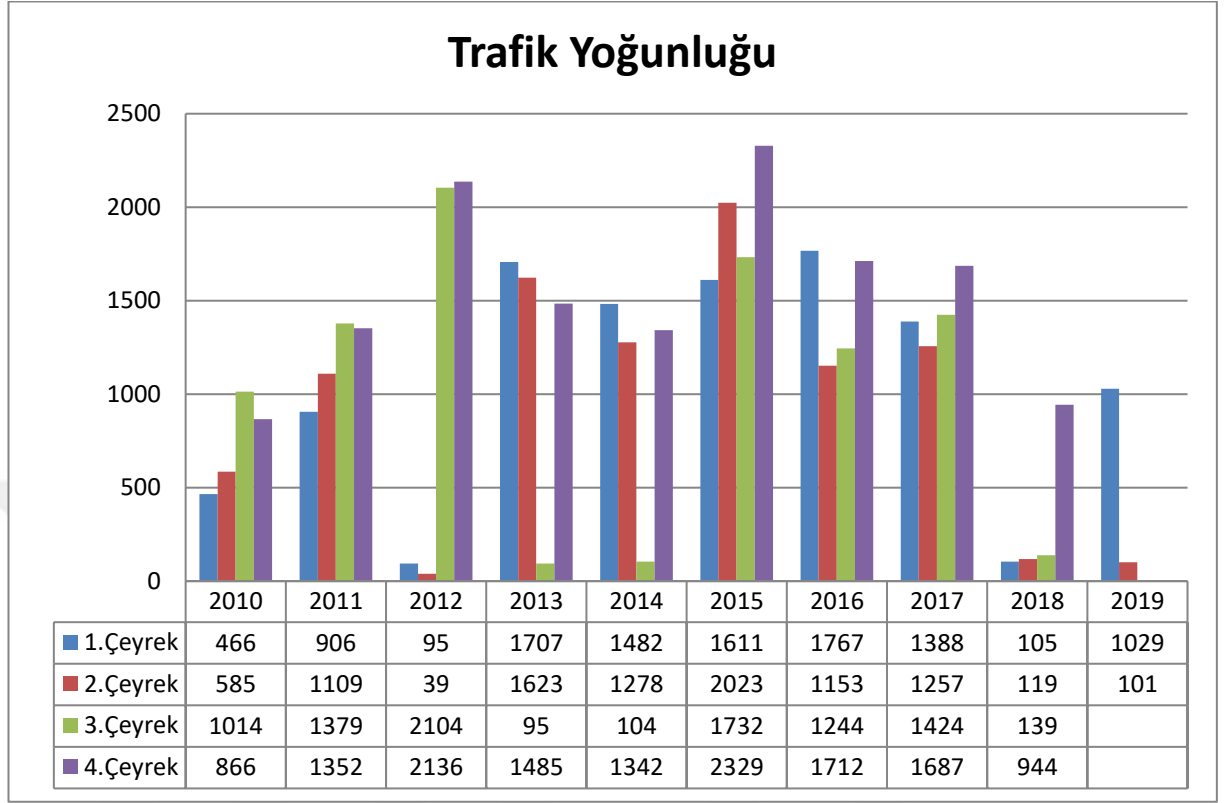
Otonom araçlar ile ilgili paylaşılan tweetlerin %5'inin pozitif, %2'sinin negatif ve %93'ünün ise nötr bir yaklaşım sergilediği sonucuna ulaşılmıştır. Otonom Araç ile ilgili paylaşılan yorumları öznel ve nesnel olma durumlarına göre incelendiğinde %33'ünün öznel ve %67'sinin ise nesnel yorumlardan oluştuğu tespit edilmiştir.

Çalışmada akıllı ulaşım sistemleri, trafik yoğunluğu ve otonom araç anahtar kelimeleri için twitter üzerinde 2010-2019 yılları arasında atılmış bütün tweetlerin taraması gerçekleştirildi. Twitter üzerinde veri tarama işlemi gerçekleştirilirken veriler üç aylık yani çeyrek dönemlik halinde elde edilmiştir. Veri taraması, analiz aşamasında kullandığımız RapidMiner programı yerine "Google Chrome" tarayıcısı üzerine "Web Scraper" uzantısı eklenerek gerçekleştirilmiştir. Tarama işleminde "Web Scraper" uzantısını kullanmamızın sebebi twitter uygulamasının geçmişe dönük en fazla 14 günlük veri ulaşımına izin vermesidir. Bundan dolayı tarama işlemi analiz aşamasında kullandığımız RapidMiner programı yerine "Web Scraper" uzantısı kullanılarak gerçekleştirilmiştir.



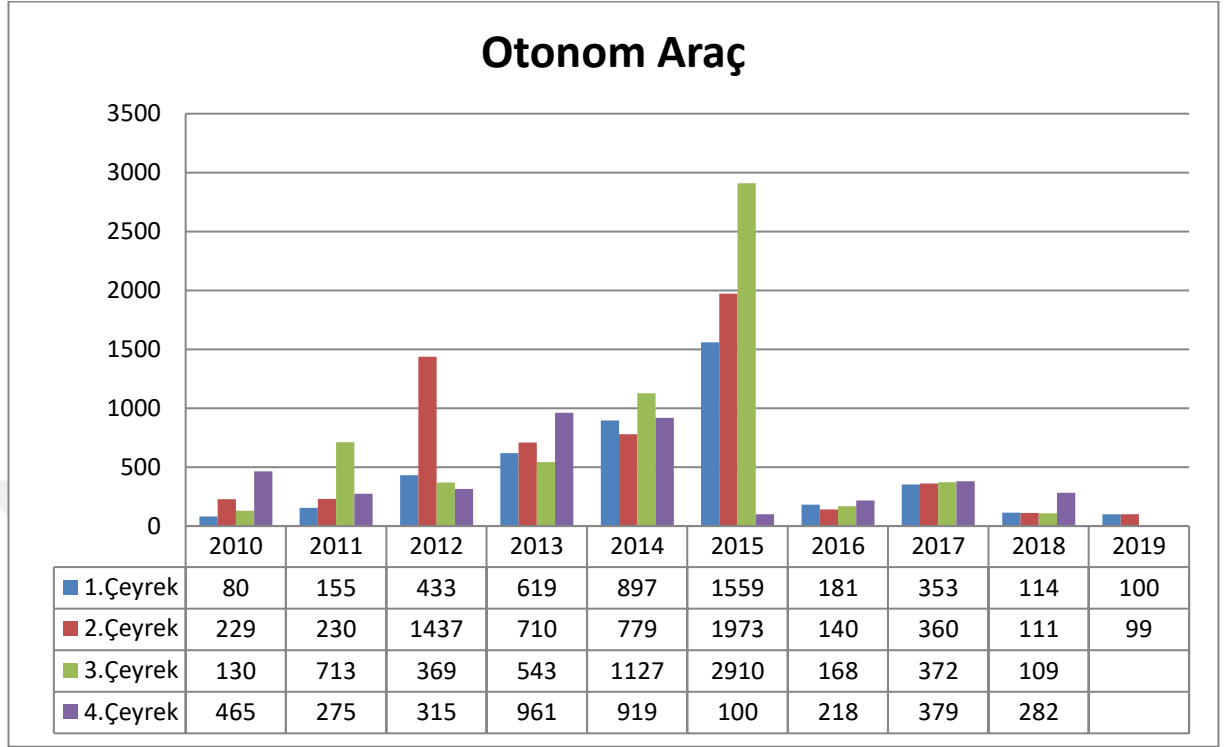
Şekil 32: Akıllı Ulaşım Sistemleri 2010-2019 yılları arasında paylaşılan tweet sayıları

Akıllı ulaşım sistemleri üzerinde atılan tweetlerin yıllara ve dönemlere göre dağılımı grafik üzerinde görülmektedir. Grafikte 1.Çeyrek; Ocak, Şubat ve Mart, 2.Çeyrek; Nisan, Mayıs ve Haziran, 3.Çeyrek; Temmuz, Ağustos ve Eylül, 4.Çeyrek; Ekim, Kasım ve Aralık aylarını içermektedir. Grafiği yıllara göre incelediğimizde akıllı ulaşım sistemleri hakkında atılan tweetlerin her geçen yıl daha da artış gösterdiğini ve bunun sonucu olarak da bireylerin her geçen yıl akıllı ulaşım sistemleri uygulamaları hakkında daha fazla bilgi edindiği ve bu uygulamalara her geçen yıl daha fazla ilgi duyduğu yorumunda bulunulabilir.



Şekil 33: Trafik Yoğunluğu 2010-2019 yılları arasında paylaşılan tweet sayıları

Grafikte 2010-2019 yılları arasında trafik yoğunluğu hakkında atılan tweetlerin yıllara ve dönemlere göre dağılımı görülmektedir. Grafiği incelediğimizde 2010-2017 yılları arasında trafik yoğunluğu hakkında atılan tweetlerin daha fazla olduğu ve son iki yılda trafik yoğunluğu hakkında paylaşılan yorumların azaldığı görülmektedir. Bireylerin akıllı ulaşım istemleri hakkındaki yorumlarının son yıllarda artış göstermesi ve duygu analizi sonucuna göre bireylerin akıllı ulaşım sistemleri hakkında genellikle olumlu bir görüşe sahip olması, son iki yılda trafik yoğunluğu hakkında paylaşılan yorumların azalmasının sebebi olarak görülebilir.



Şekil 34: Otonom Araç 2010-2019 yılları arasında paylaşılan tweet sayıları

Grafikte otonom(sürücüsüz) araç hakkında paylaşılan yorumların yıllara ve dönemlere göre dağılımı görülmektedir. Grafiği incelediğimizde 2015 yılında otonom araç hakkında paylaşılan yorumların fazlalığı göze çarpmaktadır. Tesla firmasının ikinci seviye oto pilotunu 2015 yılında tanıtan ilk firma olması ve ikinci seviye oto pilotun ilk kez 2015 yılında piyasaya tanıtılması(SÜRÜCÜSÜZ ARAÇLAR VE TÜRKİYE ARALIK 2017 ARAŞTIRMA RAPORU, 2017), otonom araca olan ilginin artmasına ve bu sayede 2015 yılında paylaşılan yorumların artış göstermesine sebep olarak görülebilir.

7 SONUÇ, DEĞERLENDİRME VE ÖNERİ

Bu bölümde, çalışmamızda “Akıllı Ulaşım Sistemleri” kavramı hakkındaki bireylerin duygu ve düşüncelerinin altıncı bölümdeki bulgulara dayanarak “AUS, Trafik Yoğunluğu ve Otonom Araç” ile ilgili elde edilen sonuçlar, bu sonuçlara bağlı değerlendirmeler ve öneriler sırasıyla aşağıda verilmiştir.

7.1 SONUÇ

Yaşanan nüfus artışı, araç sahipliliği ve buna bağlı artan araç sayısı ile birlikte trafik yoğunluğu, çevre kirliliği ve trafik kazaları gibi ulaşım sorunları tüm ülkelerin ortak meselesi olmuştur. Bu sorunların çözümü için 1990'lı yıllarda ortaya çıkan "Akıllı Ulaşım Sistemleri" uygulamaları yaygın etkisini tüm ulaşım sistemlerinde gittikçe arttırmaktadır. Geçmişten günümüze kadar dünyada ve ülkemizde bu alanda çok çeşitli çalışmalar yapılmıştır. Özellikle, Japonya, Kore, Almanya, Amerika bu sistemlerin en yoğun kullanıldığı ve aynı zamanda büyük otomotiv üreticisi ülkelerdir.

Ülkemizde ulaşım sektöründe yaşanan trafik sorunları gün geçtikçe akıllı ulaşım sistemlerinin önemini ve kullanımını arttırmaktadır. Günlük hayatımızda her yerde böyle uygulamalar ve planlamalar ile görsel veya sosyal medya ortamında karşılaşırız. Özellikle, Akıllı Ulaşım Sistemleri'nin uygulandığı pilot bölgelerde ölümlü ve yaralanmalı kazalarda yaklaşık % 20 azalma sağlanması toplumun ilgisini Akıllı Ulaşım Sistemleri kavramına yöneltmiştir.

Bu nedenle, çalışmamızda Akıllı Ulaşım Sistemleri kavramı hakkındaki bireylerin duygu ve düşüncelerinin ne yönde olduğu incelenmiştir. Çalışmada, bireylerin Akıllı Ulaşım Sistemleri hakkındaki düşünceleri incelenirken metin madenciliği tekniklerinden duygu analizi yöntemi kullanılmıştır.

İnternet ve sosyal medya alanında yaşanan gelişmelere paralel olarak özellikle reklamcılık ve pazarlama sektöründeki firmalar kendilerini tanıtmak, geliştirmek ve rekabet düzeyini arttırmak için bu ortamları etkin kullanmaya başlamışlardır. Sosyal medya uygulamalarına olan talebin artması bu alanda firmaların reklamcılık faaliyetlerini olumlu yönde arttırmıştır. Firmalar müşterilerin sosyal medya ve internet ortamındaki yorumlarını, şikâyetlerini inceleyerek müşteri profilleri hakkında bilgi edinmekte ve müşterinin memnuniyet düzeyini ölçebilmektedir.

Yaşanan bu gelişmeler duygu analizi kavramının aktif kullanımını ve önemini gün geçtikçe arttırmaktadır. Çalışmamızda da bireylerin önceden belirlenen anahtar kelimeler üzerinden twitter sosyal medya ağı üzerinden attıkları yorumlar incelenirken Akıllı Ulaşım Sistemleri üzerindeki duygu düzeyleri belirlenmiştir.

Altıncı bölümdeki bulgulara dayanarak “AUS, Trafik Yoğunluğu ve Otonom Araç” ile ilgili elde edilen sonuçlar sırasıyla aşağıdaki gibidir;

i)Altıncı bölümdeki bulgulara dayanarak AUS ile ilgili elde edilen sonuçlar aşağıdadır;

- Analiz aşamasında twitter üzerinde Akıllı Ulaşım Sistemleri(AUS) ile ilgili paylaşılan 119 adet tweete ulaşılmıştır.
- Duygu analizi sonuçlarını değerlendirdiğimizde kullanıcıların “Akıllı Ulaşım Sistemleri” ile ilgili paylaştıkları tweetlerin %62’sinin pozitif, %3’ünün negatif ve %35’inin ise nötr bir yaklaşım sergilediğini tespit edilmiştir.
- AUS ile ilgili paylaşılan yorumlar öznel ve nesnel olma olmalarına göre incelendiğinde %70’inin öznel ve %30’unun nesnel yorumlardan oluştuğu görülmektedir.
- AUS ile ilgili 2010-2019 yılları arasında atılan tweetlerin yıllara göre zaman serisini incelediğimizde her geçen yıl AUS ile ilgili paylaşılan tweet sayısının arttığı görülmektedir. 2010 yılında toplam 1003 adet tweet paylaşılırken, 2017 yılında 1455 tweet ve 2018 yılı içerisinde toplamda 1890 adet tweet paylaşılmıştır.

ii)Altıncı bölümdeki bulgulara dayanarak Trafik Yoğunluğu ile ilgili ulaştığımız sonuçlar aşağıdadır;

- Analiz aşamasında twitter üzerinde trafik yoğunluğu ile ilgili paylaşılan 2499 adet tweete ulaşılmıştır.
- Duygu analizi sonuçlarını değerlendirdiğimizde kullanıcıların “Trafik Yoğunluğu” ile ilgili paylaştıkları tweetlerin %97’sinin negatif, %3’ünün ise nötr bir yaklaşım sergilediğini ve trafik yoğunluğu ile ilgili hiç olumlu bir görüşün bulunmadığı sonucuna ulaşılmıştır.
- Trafik yoğunluğu ile ilgili paylaşılan yorumları öznel ve nesnel olma durumlarına göre incelediğimizde %97’sinin öznel ve %3’ünün ise nesnel yorumlardan oluştuğu tespit edilmiştir.
- Trafik yoğunluğu ile ilgili 2010-2019 yılları arasında atılan tweetlerin yıllara göre zaman serisini incelediğimizde 2010-2017 yılları arasında trafik yoğunluğu

hakkında atılan tweetlerin daha fazla olduğu ve son iki yılda trafik yoğunluğu hakkında paylaşılan yorumların azaldığı görülmektedir. Örneğin son dört yılda trafik yoğunluğu ile ilgili paylaşılan tweet sayılarını incelediğimizde, 2016 yılında toplam 5876 adet tweet paylaşılırken, 2017 yılında 5756 tweet, 2018 yılında 1307 tweet ve 2019 yılı içerisinde toplamda 1130 adet tweet paylaşıldığı tespit edilmiştir.

iii)Altıncı bölümdeki bulgulara dayanarak Otonom Araç ile ilgili elde edilen sonuçlar aşağıdadır.

- Analiz aşamasında twitter üzerinde otonom araçlar ile ilgili paylaşılan 1004 adet tweete ulaşılmıştır.
- Duygu analizi sonuçlarını değerlendirdiğimizde kullanıcıların “Otonom Araç” ile ilgili paylaştıkları tweetlerin %5’inin pozitif, %2’sinin negatif ve %93’ünün ise nötr bir yaklaşım sergilediği sonucuna ulaşılmıştır.
- Otonom Araç ile ilgili paylaşılan yorumları öznel ve nesnel olma durumlarına göre incelediğimizde %33’ünün öznel ve %67’sinin ise nesnel yorumlardan oluştuğu tespit edilmiştir.
- Otonom Araç ile ilgili 2010-2019 yılları arasında atılan tweetlerin yıllara göre zaman serisini incelediğimizde özellikle 2015 yılında atılan tweet sayısının diğer yıllara göre daha fazla olduğu görülmektedir. 2010 yılında toplamda 904 adet tweet paylaşılırken, 2011 yılında 1373 tweet, 2012 yılında 2554 tweet, 2013 yılında 2833 tweet, 2014 yılında 3722 tweet, 2015 yılında 6542 tweet, 2016 yılında 707 tweet, 2017 yılında 1464 tweet, 2018 yılında 616 tweet ve 2019 yılı haziran ayı sonuna kadar ise toplamda 199 adet tweet paylaşıldığı sonucuna ulaşılmıştır.

7.2 DEĞERLENDİRME

7.1’de elde edilen sonuçlara göre AUS ile ilgili değerlendirmeler aşağıda yer almaktadır;

- Akıllı Ulaşım Sistemleri ile ilgili duygu analizi sonucunu incelediğimizde, bireylerin %62 oranında AUS ile alakalı olumlu yönde bir görüşe sahip olduğunu

yani AUS uygulamalarından memnun oldukları değerlendirmesinde bulunabiliriz. Ayrıca, öznellik sonuçları incelendiğinde ise bireylerin %70 oranında öznel yorumlarda bulunduğunu yani genel geçer bilgiler yerine kendi düşüncelerini paylaştıkları yorumunda bulunabiliriz. AUS ile ilgili paylaşılan yorumların her geçen yıl arttığını da göz önüne alırsak bireylerin Akıllı Ulaşım Sistemleri uygulamalarının trafik yoğunluğunu azaltması, trafik güvenliğini artırması, enerji verimliliği sağlayarak hem ülke ekonomisine katkı sağlaması hem de çevre kirliliğini azaltmasından haberdar olduğu ve bundan dolayı AUS uygulamalarının günlük hayatımızda daha fazla yer alması gerektiği yorumunda bulunabiliriz. Yani, AUS farkındalığı ve bilincinin gelişimine sosyal medya olumlu katkı sağlamıştır.

7.1’de elde edilen sonuçlara göre “Trafik Yoğunluğu” ile ilgili değerlendirmeler şunlardır;

- Trafik yoğunluğu ile ilgili duygu analizi sonuçlarını incelediğimizde bireylerin %97 oranında olumsuz görüşe sahip olduğunu ve olumlu bir görüşün bulunmadığını görmekteyiz. Trafik yoğunluğu ile ilgili 2010-2019 yılları arasında atılan tweetleri incelediğimizde bu konu ile ilgili paylaşılan yorumların son iki yılda daha da azaldığını bilmekteyiz. Tüm bu sonuçları değerlendirdiğimizde, her geçen yıl Akıllı Ulaşım Sistemleri uygulamalarına olan ilginin ve memnuniyetin artmasında göz önüne alırsak, AUS uygulamalarının trafik yoğunluğu ile ilgili olumsuz görüşlerin ve paylaşımların azalmasında etkili olduğu ve bu uygulamaların günlük hayatımızı olumsuz etkileyen trafik sorununa çözüm olacağı inancının oluşmaya başladığı yorumunda bulunabiliriz.

7.1’de elde edilen sonuçlara göre “Otonom Araç” ile ilgili değerlendirmeler şunlardır;

- Otonom araçlar ile ilgili duygu analizini sonuçları incelediğimizde, bireylerin %93 oranında tarafsız bir görüşe sahip olduğunu ve öznellik analizine göre ise nesnel yorumlarda bulunduğunu biliyoruz. Otonom araç ile ilgili 2010-2019 yılları arasında paylaşılan tweet sayılarını incelediğimizde ise özellikle 2015 yılında paylaşılan yorumların diğer yıllara göre daha fazla olduğu ve bunun nedeni olarak da Tesla firmasının ikinci seviye oto pilotunu 2015 yılında ilk kez piyasaya tanıtması olarak değerlendirilebilir. 2010 yılından itibaren otonom araç ile ilgili paylaşımların her

geçen yıl arttığını da göz önüne alırsak AUS'nin en güncel bileşenlerinden olan otonom araçların, hem sosyal ve görsel medya katkısıyla hem de trafik sorunlarının çözümünde sağladığı avantajlar ile Akıllı Ulaşım Sistemleri üzerindeki farkındalığın artmasına katkı sağlayacağı açıktır.

Ulaşılan tüm sonuçlar üzerinde genel bir değerlendirme yapacak olursak; subjectivity analizi sonuçlarına göre kullanıcıların genelinin kişisel yani öznel yorumlarda bulunduğunu görmekteyiz.

Duygu analizi sonuçları hakkında genel bir değerlendirme yapacak olursak; Akıllı Ulaşım Sistemleri'nin kullanıcılar üzerinde olumlu bir etkiye sahip olduğunu söyleyebiliriz. Kullanıcıların trafik yoğunluğu ile ilgili %97 oranında olumsuz bir görüşe sahip olduğunu da göz önüne alırsak; yolculuk sürelerini kısaltması, zaman tasarrufu sağlaması, hareketliliği arttırması ve trafik yoğunluğunu azaltması gibi özelliklerinden dolayı akıllı ulaşım sistemleri uygulamalarının yaygınlaşması kullanıcılar üzerinde daha olumlu bir etki oluşturacağı açıktır.

7.3 ÖNERİLER

Literatür incelendiğinde duygu analizi alanında yapılan çalışmalarda genellikle makine öğrenmesi yönteminin kullanıldığı görülmektedir. Duygu analizinde sözlük tabanlı yöntem ile yapılan çalışmalarda veriler genellikle İngilizce dilinde kullanılmaktadır. Çalışmamız da analiz aşamasında duygu analizi yöntemlerinden olan sözlük tabanlı yöntem kullanılmış ve veriler İngilizce dilinde elde edilmiş, analizde bu dili destekleyecek RapidMiner tarafından sağlanan sözlük kullanılmıştır. RapidMiner analiz programının Türkçe tabanlı sözlükleri desteklememesi bizim çalışmamızda Türkçe tabanlı kelimeleri kullanmamıza engel olmuştur. Bu durum, Türkiye ve dünyada Akıllı Ulaşım Sistemleri veya başka bir konu için karşılaştırma yapmamıza ve yorumlara imkan vermemiştir.

Ayrıca, çalışmamız da twitter uygulamasının uzun dönem aralıklarında veri ulaşımına izin vermemesinden dolayı kısıtlı sayıda ve kısıtlı zaman aralığında (14 gün) tweet' e ulaşılmıştır. Bu çalışmamız açısından geçmiş dönemlerle karşılaştırma yapma imkanını ve atılan tweetlerin RapidMiner programı üzerinden bölgelere göre incelenmesi durumunu ortadan kaldırmıştır.

Ülkemizde “Akıllı Ulaşım Sistemleri” konusunda yapılan ilk ulusal metin madenciliği yüksek lisans tezi olan çalışmamız yukarıda açıkladığımız bütün dezavantajlara rağmen benzer çalışmalara akademik öncülük etmesi açısından önemlidir ve bu önemin artarak devam etmesi için aşağıda sunulan başlıkları katkı olarak öneriyoruz.

- Sonuç ve değerlendirme kısımlarında açıkladığımız Akıllı Ulaşım Sistemleri algısının ve farkındalığının artırılması için sosyal medya ortamları, haber sayfaları ve internet siteleri üzerinden yapılan paylaşımların sayısı daha da arttırılabilir.
- Resmi ve yerel otoriteler, Akıllı Ulaşım Sistemleri ile ilgili yapılacak çalışmalara destek vererek bu farkındalığın artmasına katkı sağlayabilirler.
- Üniversitelerimiz bünyesinde yüksek lisans ve doktora programları açılarak Akıllı Ulaşım Sistemleri algısı ve farkındalığı daha da arttırılabilir.
- Gelecek çalışmalarda, duygu analizi yöntemiyle Türkçe tweetler üzerinde gerçekleştirilen çalışmaların sayısının az olmasından dolayı bu alandaki duygu analizi çalışmaları daha da geliştirilebilir.
- Türkçe tweetler üzerinde yapılacak çalışmalarda “Zemberek” ve “SentiTurkNet” gibi sözlükler kullanılarak duygu analizi ile Türkçe tweetler üzerinde yapılan çalışmaların sayısı daha da arttırılabilir.
- Gelecek çalışmalarda uzun dönem aralıklarında tweet tarama işlemi, RapidMiner analiz programının Türkçe tabanlı sözlükleri desteklemesi ile daha fazla veriye ulaşılabilir ve duygu analizi çalışması daha da geliştirilebilir.
- Akıllı Ulaşım Sistemleri uygulamalarının olumlu etkilerinin Kamu Spotu olarak görsel medyada yayınlanması ve bu şekilde AUS algısının yaygınlaştırılması.
- AUS ile ilgili toplantı, çalıştay, konferans, sempozyum, kongre ve benzeri akademik etkinliklerin arttırılması.
- Lisans düzeyinde tüm fakültelerde Akıllı Ulaşım Sistemleri ile ilgili kavram ve uygulamaları içeren 2 saatlik bir dersin zorunlu olarak verilmesi.
- Mobil uygulamalar ile AUS uygulamaları tüm kullanıcılarla paylaşılabilir.

Tez çalışmasının metin madenciliği ile AUS kavramı üzerine yapılan ilk çalışma olmasından dolayı daha detaylı benzer çalışmalara zemin hazırlaması trafik

kazalarının, trafik yoğunluğun ve karbon salınımı gibi trafikten kaynaklı tüm sorunların minimum olduğu “TÜRKİYE” dileğiyle.

KAYNAKÇA

AKBA, F. (2014, Kasım 27). DUYGU ANALİZİNDE ÖZNETELİK SEÇME METRİKLERİNİN DEĞERLENDİRİLMESİ: TÜRKÇE FİLM ELEŞTİRİLERİ. 33-34.

AKÇORA, B. (2016, Mart 13). *Vektör Uzay Modeli* . Bahadır Akçora : <https://bahadirakcora.wordpress.com> adresinden alınmıştır

AKGÜL, E. S., ERTANO, C., & DİRİ, B. (2015, Temmuz). Twitter verileri ile duygu analizi.

ALBAYRAK, M., TOPAL, K., & ALTINTAŞ, V. (2017). Sosyal Medya Üzerindeki Veri Analizi:Twitter.

ALTINOK, B. (2017). *En İyi Açık Kaynak 6 Veri Madenciliği Aracı*. Webmaster Kitchen: <https://webmaster.kitchen> adresinden alınmıştır

BAYKARA, M., & GÜRTÜRK, U. (2017). Sosyal Medya Paylaşımının Duygu Analizi Yöntemiyle Sınıflandırılması. 2. *Uluslararası Bilgisayar Bilimleri ve Mühendisliği Konferansı*, (s. 957).

CAN, Ü., & ALATAŞ, B. (2017). Duygu Analizi ve Fikir Madenciliği Algoritmalarının İncelenmesi.

ÇOBAN, Ö., ÖZYER, B., & TÜRÜKLÜ ÖZYER, G. (2010). Türkçe Twitter Mesajlarının Duygu Analizi. 2-3.

ÇOBAN, Ö., ÖZYER, B., & TÜRÜKLÜ ÖZYER, G. (2015). Türkçe Twitter Mesajlarının Duygu Analizi.

DELEN, D., Thomas, H., Robert, N., John, E., & Gary, M. (2012). The Seven Practice Areas of Text Analytics. <https://www.amazon.com> , 31-32.

EXASTAX BLOG. (2018, OCAK 31). *VERİ MADENCİLİĞİ TARİHİ VERİ ANALİTİĞİ*. EXASTAX: <https://www.exastax.com.tr> adresinden alınmıştır

İsPark Android Uygulaması. (tarih yok). <http://www.scroll.com.tr>. adresinden alınmıştır

İŞIKLI, B. (2009, Şubat 15). *Veri Madenciliği(Data Mining) Nedir ve Nerelerde Kullanılır.* wordpress: <https://burakisikli.wordpress.com> adresinden alınmıştır

KAYNAR, O., GÖRMEZ, Y., YILDIZ, M., & ALBAYRAK, A. (2016). Makine Öğrenmesi Yöntemleri ile Duygu Analizi. *ResearchGate* , 237.

Kaynar, O., Yıldız, M., Albayrak, A., & Görmez, Y. (2016, Eylül 17-18). Makine Öğrenmesi Yöntemleri ile Duygu Analizi.

KOYUCUGİL, A. S., & ÖZGÜLBAŞ, N. (2009). Veri Madenciliği: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları. *BİLİŞİM TEKNOLOJİLERİ DERGİSİ* , 23-24.

MELEK, C. (2012). METİN MADENCİLİĞİ TEKNİKLERİ İLE ŞİRKETLERİN VİZYON İFADELERİNİN ANALİZİ. s. 28-29.

Meral, M., & Diri, B. (2014). Twitter Üzerinde Duygu Analizi . *2014 IEEE 22nd Signal Processing and Communications Applications Conference (SIU 2014)*. Trabzon.

ÖZKAYNAK, E. (2011). *Örüntü Tanıma* . Blogcu: <http://emrahozkaynak.blogcu.com> adresinden alınmıştır

Prabowo, Rudy, & Thelwall, M. (2009). Sentiment Analysis:A combined approach. 143-157.

SEVİNDİ, B. İ. (2013, Ocak). TÜRKÇE METİNLERDE DENETİMLİ VE SÖZLÜK TABANLI DUYGU ANALİZİ YAKLAŞIMLARININ KARŞILAŞTIRILMASI.

(2017). *SÜRÜCÜSÜZ ARAÇLAR VE TÜRKİYE ARALIK 2017 ARAŞTIRMA RAPORU*. thinktech.

ŞEKER, S. E. (2014). [sadievrenseker.com](http://bilgisayarkavramlari.sadievrenseker.com):
<http://bilgisayarkavramlari.sadievrenseker.com> adresinden alınmıştır

ŞEKER, S. E. (2016). Duygu Analizi(Sentiment Analysis). 21-36.

ŞEKER, S. E. (2008, 12 01). *SVM (Support Vector Machine, Destekçi Vektör Makinesi)*. Bilgisayar Kavramları : <http://bilgisayarkavramlari.sadievrenseker.com> adresinden alınmıştır

ŞEKER, Ş. E. (2015). Metin Madenciliği (Text Mining). *Ybs Ansiklopedi* , 31-32.

ŞEKER, Ş. E. (2012, Kasım 21). *POS Tagger (Metin Parçası Etiketleme)*. Bilgisayar Kavramları: <http://bilgisayarkavramlari.sadievrenseker.com> adresinden alınmıştır

ŞEKER, Ş. E., & YILDIZ, M. (2016). Veri Madenciliği Araçları(Data Mining Tools). *YBS Ansiklopedi* , 18-19.

Tang, Huifeng, Tan, S., & Cheng, X. (2009). A survey on sentiment detection of reviews. 10769.

Tektaş, M., Korkmaz, K., & Erdal, H. (2016). AKILLI ULAŞIM SİSTEMLERİNİN GELECEĞİ. *ICOMEPEP 2016 | International Congress of Management Economy and Policy | Proceedings Book* , (s. 1).

TETİK, S. (2017). Text Mining Application with Twitter and R. *ACADEMIA* , 3-4.

Türkiye İstatistik Kurumu İl Göstergeleri. (tarih yok). Türkiye İstatistik Kurumu: <http://www.tuik.gov.tr> adresinden alınmıştır

TÜRKMENOĞLU, C. (2015). Türkçe Metinlerde Duygu Analizi.

UZUN, E. (2016, Ocak 17). *KNN ALGORİTMASI*. Makine Öğrenmesi Knn Algoritması: <https://www.e-adys.com> adresinden alınmıştır

ÜLGEN, E. K. (2017, Ekim 16). *Makine Öğrenimi k- En Yakın Komşuluk*. Medium: <https://medium.com> adresinden alınmıştır

Üstüner, M., Balık Sanlı, F., Bektas Balcık, F., & Esetlili, M. T. (2013). DESTEK VEKTOR MAKİNELERİ TEKNİĞİ İLE SINIFLANDIRMA: RAPİDEYE ORNEĞİ. *Türkiye Ulusal Fotogrametri ve Uzaktan Algılama Birliği VII. Teknik Sempozyumu*, (s. 1-2).

YARDIM, M. S., & AKYILDIZ, G. (2005). AKILLI ULAŞTIRMA SİSTEMLERİ VE TÜRKİYE'DEKİ UYGULAMALAR. 410-411.

YEŞİLYURT, A., & ŞEKER, Ş. E. (2017). Metin Madenciliği Yöntemleri ile Twitter Duygu Analizi.

YILDIRIM, S., & YILDIZ, T. (2018, Şubat 23). Türkçe için karşılaştırmalı metin sınıflandırma analizi.