

**OTOMATİK DUYGU SÖZLÜĞÜ GELİŞTİRİLMESİ VE
HABERLERİN DUYGU ANALİZİ**

**AUTOMATED SENTIMENT LEXICON GENERATION
AND SENTIMENT ANALYSIS OF NEWS**

FATİH SAĞLAM

DR. ÖĞR. ÜYESİ FUAT AKAL

Tez Danışmanı

PROF. DR. HAYRİ SEVER

Eş Danışman

Hacettepe Üniversitesi

Lisansüstü Eğitim - Öğretim ve Sınav Yönetmeliğinin

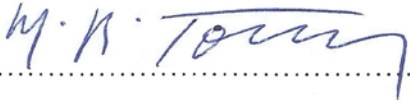
Bilgisayar Mühendisliği Anabilim Dalı İçin Öngördüğü

DOKTORA TEZİ olarak hazırlanmıştır.

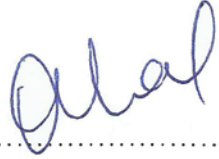
2019

FATİH SAĞLAM'ın hazırladığı “**Otomatik Duygu Sözlüğü Geliştirilmesi ve Haberlerin Duygu Analizi**” adlı bu çalışma aşağıdaki jüri tarafından **BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**’nda **DOKTORA TEZİ** olarak kabul edilmiştir.

Prof. Dr. Mehmet Reşit TOLUN
Başkan



Dr. Öğr. Üyesi Fuat AKAL
Danışman



Prof. Dr. İsmail Hakkı TOROSLU
Üye



Doç. Dr. Lale ÖZKAHYA
Üye



Dr. Öğr. Üyesi Burcak GENÇ
Üye



Bu tez Hacettepe Üniversitesi Fen Bilimleri Enstitüsü tarafından **DOKTORA TEZİ** olarak/...../..... tarihinde onaylanmıştır.

Prof. Dr. Menemşe GÜMÜŞDERELİOĞLU
Fen Bilimleri Enstitüsü Müdürü

YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izni alarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan *“Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge”* kapsamında tezim aşağıda belirtilen koşullar haricinde YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- ☐ Enstitü / Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren 2 yıl ertelenmiştir.
- ☐ Enstitü / Fakülte yönetim kurulu gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren ay ertelenmiştir.
- ☐ Tezim ile ilgili gizlilik kararı verilmiştir.

29/5/2019


FATİH SAĞLAM

Sevgili Eşim ve Oğluma...

ETİK

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada,

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- ve tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

29/5/2019



FATİH SAĞLAM

ÖZET

OTOMATİK DUYGU SÖZLÜĞÜ GELİŞTİRİLMESİ VE HABERLERİN DUYGU ANALİZİ

Fatih SAĞLAM

Doktora, Bilgisayar Mühendisliği Bölümü

Tez Danışmanı: Dr. Öğr. Üyesi Fuat AKAL

Eş Danışman: Prof. Dr. Hayri SEVER

Mayıs 2019, 141 sayfa

Kitle iletişim araçlarının toplumu bilgilendirme, farkındalık yaratma, etkileme ve hatta harekete geçirebilme konusunda günümüzde ulaştığı seviye oldukça yüksektir. Ana akım haber medyaları ön sıralarda yer alan kitle iletişim araçları olup, devasa boyutta ve çeşitlilikte içerik üretilmektedir. Araştırmacılar perspektifinden bakıldığında hazine niteliğindeki ve çoğunlukla metin türündeki bu içeriklerin analiz edilebilmesi, karar vericilerden politika üreticilerine kadar çok geniş bir yelpaze için oldukça önemlidir. Metinlerin analizi Doğal Dil İşleme disiplininin ilgi alanına girmekte olup, bu disiplinin bir alt dalı olan Duygu Analizi de, içeriğin duygu boyutuyla değerlendirilerek bakış açısının zenginleşmesine katkı sunmaktadır. İçeriğin analizinde temel gereksinim, dile özgü kaynakların varlığını gerekli kılarken, analizin doğruluğu ise bu kaynakların kapsam ve içerik olarak yeterliliğini gündeme taşımaktadır. İngilizce dili için mevcut olan kaynakların ve kütüphanelerin gelişmişlik düzeyi, Türkçe'nin de içinde bulunduğu

birçok diğ er dile g ore ileri seviyededir ve literat urdeki birçok  alıřma da dile  zg u kalmaktadır.

T rk e duygu analizi  alıřmalarında arařtırmacılar ağırlıklı olarak  eviri temelli yaklařımlar sergilemiřlerdir. Ancak diller arasındaki  eviri performansının genel bařarımı dođrudan etkileyeceđi a ıktır ve de T rk e gibi eklemeli dillerde problem biraz daha derinleřmektedir.  alıřmadaki temel hipotezimiz, dilin kaynaklarının dilin i erisinden  retilmesinin gerekliliđidir. Bu  er evede i erik olarak zengin ve dođru polaritelenmiř Genel Ama lı T rk e Duygu S zl đ  geliřtirilmesi tezin iki temel hedefinden birincisidir. B ylece T rk e  alıřma yapan b t n disiplinlere a ık eriřimli bir kaynak sunulmuř olacaktır. Tezde hedeflenen ikinci husus ise duygu analizini farklı bir perspektiften ele alan Duygu Haritası Modelinin geliřtirilmesidir.

Tez, temel olarak    fazda  alıřılmıřtır. Birinci fazda, ana akım haber medyalarındaki haber metinleri kullanarak T rk e Duygu S zl đ  geliřtirilmesi hedeflenmiřtir. Bu kapsamda, polaritesi bilinen haberlerden b y k bir derlem oluřturulmuřtur. Bu metinlerden elde edilen terimlerin ton ve polarite deđerleri belirlenerek, mevcut bir diğ er T rk e duygu s zl đ  ile birleřtirilmiř ve toplamda 37K hacmindeki, adını SWNetTR-PLUS olarak verdiđimiz T rk e duygu s zl đ n n ilk s r m  oluřturulmuřtur.

 ikinci fazda, SWNetTR-PLUS k t phanesinin zenginleřtirilmesi ile terimlerin polarite ve ton deđerlerinin daha kararlı hale getirilmesi hedeflenmiřtir. S zl đ m zdeki mevcut terimlerin negatif karřılıklarının tespiti i in ZIT anlamlılarının, s zl đ m z n kapsamının geniřletilmesine y nelik olarak da Eř anlamlı karřılıklarının kullanılmasına odaklanılmıřtır. Bu yaklařım kullanarak farklı kaynaklarla s zl k zenginleřtirilmiř ancak hatalı eřleřmelerin varlıđı ile s zl đ  yeni kazandırılan terimlerin polarite ve ton deđerlerinin tespitine y nelik sorunlar g ndeme gelmiřtir. Bu noktada problem, varlıklar arasındaki iliřkileri temsil g c  ve iliřkisel iřlemlerde sađladığı kolaylıklar sebebiyle  izge kullanarak modellenmiřtir. Bađ g c , ton yayılımı, yanlılık dengelemesi matematiksel modelleri geliřtirilerek bahsedilen problemlere uygulanmıř

ve sonuçlar değerlendirilmiştir. Bu fazın sonunda adını SWNetTR++ olarak isimlendirdiğimiz 49K kapasiteli Genel Amaçlı Türkçe Duygu Sözlüğü oluşturulmuştur.

Üçüncü fazda, duygu analizi literatürüne Duygu Haritası Modeli önerilmiştir. Bu modele götüren temel motivasyon özellikle doküman seviyesindeki duygu analizi çalışmalarının, metni pozitif/nötr/negatif şeklinde bir sınıflandırma ile sayısal bir skor atama şeklinde ele almasından kaynaklanmaktadır. Bu yaklaşıma getirdiğimiz eleştiri ise, metinlerdeki duygu dalgalanmalarının göz ardı ediliyor olması ve dokümanın salt bir polarite ile etiketlenmesinin her durumda doğru bir yaklaşım olamayacağıdır. Çözüm önerisi olarak analize tabi tutulan içeriğin bünyesindeki duygu dalgalanmalarının ortaya çıkarılmasına, görsel olarak sunulmasına ve yorumlanmasına imkan sağlayan Duygu Haritası Modeli geliştirilmiştir. Model, örnek içerikler üzerinde uygulanarak sonuçlar değerlendirilmiştir.

Anahtar Kelimeler: Türkçe Duygu Sözlüğü, Duygu Haritası Modeli, Çizge, Duygu Analizi, Doğal Dil İşleme.

ABSTRACT

AUTOMATED SENTIMENT LEXICON GENERATION AND SENTIMENT ANALYSIS OF NEWS

Fatih SAĞLAM

Doctor of Philosophy, Department of Computer Engineering

Supervisor: Dr. Öğr. Üyesi Fuat AKAL

Co-Supervisor: Prof. Dr. Hayri SEVER

May 2019, 141 pages

The level reached by mass media today, with respect to informing the society, raising awareness, affecting opinions, and even mobilizing masses, is impressive. Being mass communication tools, mainstream news media produces enormous amounts of content. Analysis of this content, which is mostly textual and treasured from an academic perspective, is also crucial for people in a large spectrum, from decision makers to policy makers. Text analysis is studied under the discipline of Natural Language Processing (NLP), and Sentiment Analysis, which is a subdiscipline of NLP, is focused on enriching this analysis by evaluating the content with respect to sentimentality. The main requirement in content analysis is the existence of necessary language resources, whereas the correctness of the analysis depends on the coverage and sufficiency of these resources. Availability and state of the resources in the English language far surpasses the resources in many other languages, including Turkish. Most studies in the literature are, hence, language dependent.

In Turkish sentiment analysis studies, researchers mostly followed a route through translation. However, it is obvious that the correctness of the translation between the languages will deeply impact the performance of the analysis, with the effect being amplified in agglomerative languages, such as Turkish. Our base hypothesis in this work is that the language resources need to be produced within the language. With respect to this philosophy, the first goal of this thesis is to produce a rich and correctly polarized General Purpose Turkish Sentiment Lexicon. Thus, we aim to provide an open resource to all disciplines working on the Turkish language. Our second goal, then, is to propose a Sentiment Map Model, which brings a fresh perspective to document sentiment analysis.

This thesis is mainly prepared in three phases. In the first phase, we aimed to produce a Turkish Sentiment Lexicon based on texts from main stream news media. To this end, a large corpus with known polarities was constructed. Once the tone and polarity values of the words from these texts were identified, they were merged with an existing Turkish Sentiment Lexicon, resulting in the first version of SWNetTR-PLUS, a 37K Turkish sentiment lexicon.

In the second phase, it was aimed to enrich SWNetTR-PLUS even further and stabilize the polarity and tone values of the words in the lexicon. To this end, synonyms and antonyms of the words in the lexicon were derived from different resources to extent the coverage of both positive and negative sentiments. At this point, it was chosen to model the whole lexicon with graphs to make it easier to explain and study the sentiment relations between the words. In order to compute missing polarities and tone values, and to stabilize the lexicon, we developed the concepts and methodologies of Tie Strength, Tone Propagation and Bias Balancing. At the end, we obtained SWNetTR++, which is a General Purpose Turkish Sentiment Lexicon with a capacity of 49K words.

In the third and the last phase, we proposed a new document sentiment analysis technique, namely Sentiment Map Model. Here, the motivation was the fact that in the literature document sentiment analysis is mostly based on assigning a

positive/neutral/negative sentiment value to a document, and this omits the sentiment fluctuations within the text, crippling the richness of the text. As a solution to that, we proposed Sentiment Map Model, which allows detecting, exposing and interpreting sentiment shifts within an analysed document. The proposed model was extensively tested on multiple texts to outline its applicability and sufficiency.

Keywords: Turkish Sentiment Lexicon, Sentiment Map Model, Graph, Sentiment Analysis, Natural Language Processing.



TEŞEKKÜR

Doktora programına kabulümden, tez sürecimin sonuna kadar desteğini ve yardımını her zaman yanımda hissettiğim, aynı zamanda tez eş danışmanlığımı da yapan, saygıdeğer Hocam Sayın Prof. Dr. Hayri SEVER'e,

Gerek tez izleme kurullarında, gerekse tezimin diğer süreçlerinde bana olan güvenini daima hissettiren ve desteğini hiçbir zaman esirgemeyen tez danışmanım Sayın Dr. Öğr. Üyesi Fuat AKAL'a,

İlk Tez İzleme Kurulunda eş danışmanlığımı da yapmış olan, ardından tezin bütün metodolojik süreçlerinde kapıları aralayan, güvenini sürekli vurgulayan, yüreklendiren ve desteğini her zaman hissettiğim Sayın Dr. Öğr. Üyesi Burak GENÇ'e,

Tez İzleme Kurullarında bana yorum ve önerileriyle çok değerli katkılar sunan Sayın Prof. Dr. Mehmet Reşit TOLUN'a, Sayın Doç. Dr. Lale ÖZKAHYA'ya, Sayın Dr. Öğr. Üyesi Engin DEMİR'e,

Hacettepe Üniversitesi Bilgisayar Mühendisliği Bölümünün çok kıymetli öğretim üyelerine ve çalışanlarına,

Doktora eğitimim boyunca, ihtiyaç duyduğum anlarda bana yardımlarıyla destek olan bütün çalışma arkadaşlarıma,

Özellikle dersler ve yeterlilik dönemlerinde, kendilerine ayıracağım vakitten çalmama rağmen desteklerini her zaman gönülden veren değerli eşim Feyza SAĞLAM ve oğlum Faruk Tuna SAĞLAM'a tüm içtenliğimle teşekkürlerimi sunarım.

İÇİNDEKİLER

ÖZET	i
ABSTRACT	iv
TEŞEKKÜR.....	vii
İÇİNDEKİLER	viii
ŞEKİLLER DİZİNİ	xi
ÇİZGELER DİZİNİ.....	xiii
SİMGELER VE KISALTMALAR.....	xv
1. GİRİŞ.....	1
1.1. Problemin Tanımı	1
1.2. Amaç ve Motivasyon	2
1.3. Özgün Değer.....	3
1.4. Tez Organizasyonu	4
2. ALAN BİLGİSİ VE ALANYAZIN ÖZETİ.....	5
2.1. Doğal Dil İşleme ve Duygu Analizi	5
2.1.1. Doğal Dil İşleme (DDİ).....	5
2.1.2. Duygu Analizi.....	8
2.1.3. Duygu Analizinin Uygulama Alanları	9
2.1.4. Duygu Analizi Seviyeleri	10
2.1.5. Duygu Analizinin Zorlukları	11
2.2. Duygu Analizi Yaklaşımları.....	14
2.2.1. Makine Öğrenmesi Temelli Duygu Analizi Yaklaşımları.....	15
2.2.2. Sözlük Temelli Duygu Analizi Yaklaşımları	17
2.3. Duygu Sözlüğü Oluşturmak	20
2.3.1. Manuel Yaklaşımla Duygu Sözlüğü Oluşturmak	20
2.3.2. Sözlük Temelli Yaklaşımla Duygu Sözlüğü Oluşturmak	22
2.3.3. Derlem Temelli Yaklaşımla Duygu Sözlüğü Oluşturmak.....	25
2.4. Duygu Sözlüklerinin Kapsamı	26

2.4.1. Genel Amaçlı Duygu Sözlüğü	27
2.4.2. Özel Amaçlı Duygu Sözlüğü	28
2.5. Türkçe Duygu Analizi ve Duygu Sözlüğü	30
2.5.1. Türkçe Dilinin Zorlukları	30
2.5.2. Türkçe Duygu Analizi ve Duygu Sözlüğü Geliştirme Çalışmaları	32
2.6. Ana Akım Medya ve Haber	36
2.7. Haberde Duygu	38
2.7.1. Haber Başlıklarına Yönelik Duygu Analizi Çalışmaları	39
2.7.2. Haber İçeriklerine Yönelik Duygu Analizi Çalışmaları	40
2.7.3. Haberlere Yapılan Okuyucu Yorumlarına Yönelik Duygu Analizi Çalışmaları	41
2.7.4. Haberlerden Duygu Sözlüğü Oluşturmaya Yönelik Çalışmalar	41
3. TÜRKÇE DUYGU SÖZLÜĞÜ	43
3.1. Terminoloji	43
3.2. Faz-1 : SWNetTR-PLUS	44
3.2.1. Veri Seti	44
3.2.2. Metodoloji	52
3.2.3. Performans Değerlendirmesi	56
3.3. Faz-2 : SWNetTR++	58
3.3.1. Veri Seti	59
3.3.2. Metodoloji	59
3.3.3. Performans Değerlendirmesi	75
4. DUYGU HARİTASI MODELİ	78
4.1. Shewhart Kontrol Diyagramı	85
4.2. Duygu Haritası Modeli ve Shewhart Kontrol Diyagramı	88
4.3. DHM Uygulama Örnekleri	92
4.3.1. Uygulama Örneği - 1	93
4.3.2. Uygulama Örneği - 2	101
4.3.3. Uygulama Örneği - 3	106
4.3.4. Uygulama Örneği - 4	108
4.3.5. Uygulama Örneği - 5	112
5. SONUÇLAR ve ÖNERİLER	116
5.1. Sonuçlar	116

5.2. Öneriler	119
6. KAYNAKLAR.....	121
EKLER	131
EK 1 - HABER-4 Haberinin Cümle Listesi.....	131
EK 2 - PAMUK PRENSES Kitabının Cümle Listesi.....	134
EK 3 - Tezden Türetilmiş Yayınlar	137
EK 4 - Tezden Türetilmiş Bildiriler	138
ÖZGEÇMİŞ	141



ŞEKİLLER DİZİNİ

Şekil 3.1 GDELT sürümleri ve veri tabloları.	46
Şekil 3.2 Örnek bir haber sayfası ve haber metni.	49
Şekil 3.3 SWNetTR-PLUS geliştirme süreci blok diyagramı.	52
Şekil 3.4 SWNetTR-PLUS Genel Amaçlı Türkçe Duygu Sözlüğünün yapısı.	56
Şekil 3.5 SWNetTR++ geliştirme süreci blok diyagramı.	60
Şekil 3.6 Eş ve zıt eşleşmelerdeki mantıksal hatalara ait birer örnek.	62
Şekil 3.7 Çizgedeki “abartı” düğümünün eş-zıt bağlantıları örneği.	63
Şekil 3.8 Bağ Gücü değeri hesaplanmış bir çizge örneği.	65
Şekil 3.9 “abartı” kelimesinin ilişki ağındaki bağlantılarının bağ gücü.	66
Şekil 3.10 “abartı” kelimesinin eş anlam ilişki ağı.	68
Şekil 3.11 “abartı” kelimesinin filtreleme sonrası eş ve zıt ilişkiler ağı.	69
Şekil 3.12 Örnek çizge: Ton Yayılımı öncesi.	70
Şekil 3.13 Örnek çizge: Ton Yayılımı sonrası.	73
Şekil 3.14 Ton kaydırma yaklaşımlarının sonucundaki değişim grafiği.	74
Şekil 4.1 Haber-1 cümle seviyesi duygu salınım grafiği.	80
Şekil 4.2 Haber-1 paragraf seviyesi duygu salınım grafiği.	81
Şekil 4.3 Haber-2 cümle seviyesi duygu salınım grafiği.	83
Şekil 4.4 Haber-3 cümle seviyesi duygu salınım grafiği.	85
Şekil 4.5 Shewhart kontrol diyagramı limitleri.	87
Şekil 4.6 Kontrol grafiği limit değerlerinin belirlenmesi süreci blok diyagramı.	88
Şekil 4.7 Shewhart kontrol diyagramı KURAL-1 ile duygu haritası analizi.	91
Şekil 4.8 Shewhart kontrol diyagramı KURAL-2 ile duygu haritası analizi.	91
Şekil 4.9 Shewhart kontrol diyagramı KURAL-3 ile duygu haritası analizi.	91
Şekil 4.10 Shewhart kontrol diyagramı KURAL-4 ile duygu haritası analizi.	92
Şekil 4.11 HABER-1 duygu haritası üzerinde uygulanan KURAL-1 ve sonuçları.	93
Şekil 4.12 HABER-1 metninin KURAL-1 sonuçlarındaki görseli.	94
Şekil 4.13 HABER-1 duygu haritası üzerinde uygulanan KURAL-2 ve sonuçları.	95
Şekil 4.14 HABER-1 metninin KURAL-2 sonuçlarındaki görseli.	96
Şekil 4.15 HABER-1 duygu haritası üzerinde uygulanan KURAL-3 ve sonuçları.	97
Şekil 4.16 HABER-1 metninin KURAL-3 sonuçlarındaki görseli.	98

Şekil 4.17 HABER-1 duygu haritası üzerinde uygulanan KURAL-4 ve sonuçları.	99
Şekil 4.18 HABER-1 metninin KURAL-4 sonuçlarındaki görseli.	99
Şekil 4.19 HABER-1 duygu haritası üzerinde uygulanan bütün kurallar ve sonuçları.	100
Şekil 4.20 HABER-1 metni üzerindeki bütün kuralların birleştirilmiş görseli.	100
Şekil 4.21 HABER-2 duygu haritası üzerinde uygulanan KURAL-1 ve sonuçları.	101
Şekil 4.22 HABER-2 metninin KURAL-1 sonuçlarındaki görseli.	102
Şekil 4.23 HABER-2 duygu haritası üzerinde uygulanan KURAL-2 ve sonuçları.	103
Şekil 4.24 HABER-2 metninin KURAL-2 sonuçlarındaki görseli.	103
Şekil 4.25 HABER-2 duygu haritası üzerinde uygulanan bütün kurallar ve sonuçları.	105
Şekil 4.26 HABER-2 metni üzerindeki bütün kuralların birleştirilmiş görseli.	105
Şekil 4.27 HABER-3 duygu haritası üzerinde uygulanan KURAL-1 ve sonuçları.	106
Şekil 4.28 HABER-3 metninin KURAL-1 sonuçlarındaki görseli.	107
Şekil 4.29 HABER-4 duygu haritası üzerinde uygulanan KURAL-1 ve sonuçları.	108
Şekil 4.30 HABER-4 duygu haritası üzerinde uygulanan KURAL-2 ve sonuçları.	109
Şekil 4.31 HABER-4 duygu haritası üzerinde uygulanan KURAL-3 ve sonuçları.	110
Şekil 4.32 HABER-4 duygu haritası üzerinde uygulanan KURAL-4 ve sonuçları.	111
Şekil 4.33 HABER-4 duygu haritası üzerinde uygulanan bütün kurallar ve sonuçları.	112
Şekil 4.34 PAMUK_PRENSES duygu haritası.	113
Şekil 4.35 PAMUK_PRENSES_EDIT duygu haritası.	115

ÇİZGELER DİZİNİ

Çizelge 2.1 Türkçe kök ve eklere yönelik bir örnek.	30
Çizelge 2.2 Türkçe olumsuzluk durumlarına ait bir örnek.	31
Çizelge 2.3 Türkçede zıt anlamlılık durumlarına ait bir örnek.	31
Çizelge 3.1 Faz-1’de kullanılan/geliştirilen veri setleri.	44
Çizelge 3.2 GDELT’in içerik analizinde kullandığı İngilizce dil kaynakları.	47
Çizelge 3.3 Etiketlenmiş Türkçe haberler derlemi.	50
Çizelge 3.4 MLTC-500 ve MLTC-353 Türkçe test derlemlerinin polarite dağılımları.	51
Çizelge 3.5 Fleiss Kappa istatistiğinin yorumlanması.	51
Çizelge 3.6 Kelimelerin kök formu örnekleri.	53
Çizelge 3.7 Haberlerden elde edilen en yüksek ve en düşük frekanslı kelimeler.	54
Çizelge 3.8 “intizar” teriminin ton vektörü.	55
Çizelge 3.9 “intizar” kelimesinin ton ve polarite değerleri.	56
Çizelge 3.10 SWNetTR-PLUS performans değerlendirme sonuçları.	57
Çizelge 3.11 Faz-2’de geliştirilen veri setleri.	59
Çizelge 3.12 EŞ-ZIT anlam karşılıkları için kullanılan kaynaklar ve kapasiteler.	60
Çizelge 3.13 Eş ve zıt anlam çiftlerine ait örnekler.	61
Çizelge 3.14 Çizgenin yapısal bilgileri.	69
Çizelge 3.15 <i>u</i> düğümü ve komşularına göre olası senaryolar ve katsayılar.	71
Çizelge 3.16 Pozitif ve negatif terim istatistiği.	74
Çizelge 3.17 SWNetTR++ performans değerlendirme sonuçları (MLTC-500).	76
Çizelge 3.18 SWNetTR++ performans değerlendirme sonuçları (MLTC-353).	77
Çizelge 4.1 Haber-1 içeriği.	79
Çizelge 4.2 Haber-1 cümle listesi.	79
Çizelge 4.3 Haber-1 paragraf listesi.	81
Çizelge 4.4 Haber-2 cümle listesi.	82
Çizelge 4.5 Haber-3 cümle listesi.	84
Çizelge 4.6 Shewhart kontrol diyagram analiz kuralları.	87
Çizelge 4.7 Türkçe cümlelerin Shewhart kontrol diyagramı limit değerleri.	89
Çizelge 4.8 DHM kapsamında Shewhart kurallarının değerlendirilmesi.	90
Çizelge 4.9 DHM uygulama örnekleri listesi.	93

Çizelge 4.10 HABER-4'ün ilk cümlesi.	111
Çizelge 4.11 Yoğun negatif tona sahip cümleler.	114
Çizelge 4.12 Yeniden düzenlenen ifadeler.	114



SİMGELER VE KISALTMALAR

Simgeler

μ	Ortalama
σ	Sigma (Standart Sapma)
α	Ton yayılımı yakınsama hız katsayısı
β	Ton yayılımı komşu düğümlerin yakınsama hız katsayısı

Kısaltmalar

DDİ	Doğal Dil İşleme
SVM	Destek Vektör Makinesi (<i>Support Vector Machine</i>)
NB	Naive Bayes
ME	Maksimum Entropi
PMI	Pointwise Mutual Information
GDELT	Global Database of Events, Language (Location) and Tone
BG	Bağ Gücü (<i>Tie Strength</i>)
DHM	Duygu Haritası Modeli
ÜKL	Üst Kontrol Limit
MÇ	Merkez Çizgi
AKL	Alt Kontrol Limit
TF-IDF	Term Frequency-Inverse Document Frequency

1. GİRİŞ

Tez çalışmasının giriş bölümü takip eden alt başlıklarda ele alınmıştır. İlk olarak problemin tanımı yapılmış, ardından amaç ve motivasyona yer verilmiştir. Çalışma kapsamında önerilen metodolojiler ve sergilenen yaklaşımlar doğrultusunda tezin özgün değeri hakkında değerlendirmelerde bulunulmuş ve son olarak da tezin organizasyonu sunulmuştur.

1.1. Problemin Tanımı

Çevrimiçi haber medyası ve sosyal medyada devasa boyutta ve çeşitlilikte veri üretilmektedir. Büyük veri konseptinin bütün karakteristiklerini gösteren bu içerikler görsel öğelerle desteklense de ağırlıklı olarak metin türündedir. Hazine niteliğindeki bu metinlerin analizi Doğal Dil İşleme (DDİ) disiplininin ilgi alanına girmektedir. DDİ'nin bir alt dalı olan Duygu Analizinde ise, içeriğin yansıttığı duygunun belirlenmesine çalışılır. Dile özgü kapsamlı ve gelişmiş bir duygu sözlüğüne olan ihtiyaç da bu noktada gündeme gelmektedir.

Dil bir ulusu oluşturan temel öğelerden birisidir. Doğal dil işleme alanındaki araştırmacılar, beklenileceği gibi ağırlıklı olarak kendi dillerinde çalışmalar yapagelmektedirler. Bunun sonucu olarak da doğal dil alanında geliştirilen metodolojiler, yaklaşımlar, kaynaklar ve özellikle de araçlar çoğunlukla dile özgü kalmakta, farklı dillere doğrudan transfer edilememektedir. Farklı dili konuşan doğal dil araştırmacılarının kendi dilleri dışında çalışmalar yaptıkları gözlemlenmektedir. Bu dil ise ağırlıklı olarak İngilizcedir. Bu ilgi sayesinde İngilizce dil kütüphaneleri ve araçlarının gelişmişlik düzeyi oldukça üst seviyededir. Türkçe'nin de içinde bulunduğu diğer birçok dilde, dil kütüphaneleri ve kaynakları henüz tam olarak gelişmemiştir. Literatürde bu ihtiyacın karşılanmasında İngilizce dil kaynaklarından çeviri yaklaşımının sergilendiği sıklıkla görülmektedir. Ancak çeviri sistematüğinden kaynaklanan anlam ve duygu kayıpları kaçınılmaz olacaktır. Çeviri yoluyla elde edilen dil kütüphanelerinin kullanımı, duygu analizinin doğruluğunu da olumsuz etkileyebilmektedir.

Türkçe perspektifinden bakıldığında, dilimiz için kapsamlı dil kütüphanelerine ihtiyaç olduğu değerlendirilmektedir. Yakın zamanda araştırmacıların Türkçe DDI çalışmalarına yöneldiklerini görmek memnuniyet verici olmakla beraber, kapsamlı ve doğru duygu skorlarına sahip Genel Amaçlı Türkçe Duygu Sözlüğü ihtiyacı, henüz tam anlamıyla karşılanamamıştır. Bu eksiklik, tez kapsamında ortaya konulan problemlerden birincisidir.

Dillerden bağımsız olarak duygu analizi dünyasındaki diğer bir problem ise, duygu analizinin bir metin sınıflandırma problemi olarak görülmesi ve bunun sonucunda içeriğin salt pozitif veya negatif gibi kesin bir sınıflandırmaya tabi tutulmasıdır. Bu yaklaşıma getirdiğimiz temel eleştiri, bir dokümanın içerisinde duygu salınımının mevcut olduğu ve bu karakteristiğin duygu analizinde göz ardı ediliyor olmasıdır. Bu eksiklik de, tez kapsamında ortaya konulan problemlerden ikincisidir.

1.2. Amaç ve Motivasyon

Yukarıda anılan birinci problem alanına yönelik olarak geniş kapsamlı, doğru polaritelenmiş ve erişime açık Genel Amaçlı Türkçe Duygu Sözlüğünün geliştirilmesi amaçlanmıştır. Böylece Türkçe içeriklerin duygu analizinin daha yüksek bir doğrulukla yapılması hedeflenmiştir. Çünkü, duygu analizi, karar vericilerden politika üreticilerine kadar geniş bir yelpazenin karar destek süreçlerine katkı sunabilecek kapasiteye sahiptir. Bu temel motivasyondan hareketle; her dilin kendi kuralları ve yapısı doğrultusunda o dile özgü doğal dil kütüphanelerinin, o dili konuşan araştırmacılar tarafından ve o dilin içerisinden üretilmesinin daha etkin ve verimli bir yaklaşım olduğu değerlendirilmektedir. Geliştirilecek Türkçe duygu sözlüğü, alanda Türkçe çalışma yapan birçok araştırmacının da temel başvuru kaynaklarından birisi olabilecektir. İlaveten, alana özel duygu sözlüklerinin, genel amaçlı duygu sözlüklerinden üretildiği göz önüne alındığında, Türkçe için de kapsamlı ve güvenilir Genel Amaçlı Türkçe Duygu Sözlüğünün varlığına olan ihtiyaç daha iyi anlaşılabilecektir.

Tez çalışmasında ortaya konulan ikinci problem alanına yönelik olarak, içeriklerdeki duygu salınımı görselleştirmek amacıyla Duygu Haritası Modeli (DHM) önerilmekte, duygu analizine keşifsel veri analizi penceresinden bakılmaktadır. Elde edilen duygu haritasının, Türkçe için belirlenen kontrol limitleri ve kontrol grafikleri dünyasının kurallarının uyarlanması ile analizi tez kapsamında ele alınmıştır.

DDİ disiplininin yelpazesi oldukça geniştir ve Türkçe için de çalışılmaya muhtaç olan çok fazla alt başlıklara sahiptir. Yukarıda belirtilen problemlere sunulacak çözüm önerilerinin, bu tez kapsamında ele alınmayan; duygu analizindeki iyileştirici yaklaşımlar, terimlerin birden fazla anlamının bulunması durumları ve hedef-tabanlı duygu analizi gibi DDİ çalışmalarına bir temel sunacağı öngörülmektedir.

1.3. Özgün Değer

Geliştirilen Türkçe duygu sözlüğüne ilaveten, duygu analizi literatürüne ve çizge dünyasına önerilen modellerle tezin aşağıdaki özgün değerlere sahip olduğu değerlendirilmektedir.

- En geniş hacimli ve açık erişimli SWNetTR++ Genel Amaçlı Türkçe Duygu Sözlüğünün, özelde Türkçe duygu analizi, genelde ise Türkçe DDİ çalışan araştırmacılara sunulması.
- Sözlüğün geliştirilmesi sürecindeki metodolojilerin tamamen otomatik olması ve kullanılan veri kaynaklarının genel erişime açık olması nedenleriyle farklı diller için de kullanılabilirliği.
- Çizge teorisi literatürüne Bağ Gücü ve Ton Yayılımı algoritmalarının önerilmesi.
- Alana özel duygu sözlüğü geliştirilmesinde SWNetTR++ dil kütüphanesinden faydalanılabileceği.
- Duygu Haritası Modeli ile duygu analizine farklı ve yenilikçi bir perspektiften bakış zenginliği kazandırılması.

1.4. Tez Organizasyonu

Tezin ikinci bölümü olan Alan Bilgisi ve Alanyazın Özeti başlığında DDİ, duygu analizi, duygu sözlüğü geliştirme yaklaşımları ve Türkçe dilinin bu çalışmalar kapsamındaki konumu incelenmiş, ana akım haber medyası ve haber içerikleri duygu perspektifinden ele alınmış ve literatürdeki ilgili çalışmalara yer verilmiştir.

Üçüncü bölümde, iki faz halinde çalışılan Türkçe Duygu Sözlüğü geliştirme çalışmasının adımları açıklanmıştır. Birinci faz kapsamında SWNetTR-PLUS ismini verdiğimiz Türkçe duygu sözlüğünün ilk sürümünün veri kaynakları ve geliştirilme süreçleri ele alınmıştır. İkinci faz kapsamında da, bu ilk sürümün kapasitesinin geliştirilerek, bünyesindeki terimlerin ton değerlerinin kararlı hale getirilmesinde kullanılan ve tez kapsamında geliştirilen çizge temelli metodolojiler hakkında bilgiler verilmiştir. Bu süreç sonunda oluşturulan kapsamlı ve kararlı genel amaçlı Türkçe duygu sözlüğü SWNetTR++'ın performans ölçümü sunulmuştur.

Son bölüm olan dördüncü bölümde, duygu analizine tabi tutulan içeriklerdeki duygu salınımını görselleştiren, yorumlamaya farklı bakış zenginliği kazandıran Duygu Haritası Modeli tanıtılmıştır. Önerilen model, toplamda beş farklı örnek içerik üzerinde uygulanarak, değerlendirmelere yer verilmiştir.

2. ALAN BİLGİSİ VE ALANYAZIN ÖZETİ

2.1. Doğal Dil İşleme ve Duygu Analizi

2.1.1. Doğal Dil İşleme (DDİ)

İnsan eliyle üretilmiş metinlerin bilgisayarlar tarafından ele alınmasını ifade eden doğal dil işleme (DDİ) kavramı, bilgisayar bilimlerinin ana disiplinlerinden biri olup, yapay zeka alanının bir alt dalı olarak da tanımlanmaktadır [1-4].

Dil, gelişimi çok uzun zaman alan ve karmaşık bir yapıya sahip olan bir olgudur. Geçen yıllar içerisinde kütüphanesine eklenen ve çıkan kelimelerle gelişime ve değişime uğramaktadır. Ancak dikkat çekici olan, içerikteki dinamiğe rağmen dilin kurallarında değişimin belirgin bir şekilde gözlemlenmemesidir [5]. Doğal dilin en temel özelliği, insanların birbirleriyle iletişim kurabilmelerini sağlayan ve bir ulusu oluşturan temel yapıtaşlarından birisi olmasıdır. Dolayısıyla her toplumun terimleri, dil kuralları ve yapısı diğer dillerden farklılık gösterir. Bunun sonucu olarak da doğal dil alanında geliştirilen metodolojiler, yaklaşımlar ve özellikle de araçlar çoğunlukla dile özgü kalmakta, her dilin kendi kuralları ve yapısı doğrultusunda o dile özgü çalışmaların araştırmacılar tarafından yapılmasını gerekli kılmaktadır.

Doğal dil işleme sürecinin iki temel girdisi mevcuttur. Bunlar metin ve ses öğeleridir. Akademik çalışmalarda girdi metin ise analiz üzerine, girdi türü ses ise sistemin reaksiyon göstermesi üzerine odaklanılmıştır.

DDİ'nin oldukça geniş bir yelpazede uygulama alanı mevcuttur [5-8]. DDİ'de, aşağıda listelenen problem ve ihtiyaç alanlarına çözüm üretilmeye çalışılmaktadır.

- Metnin yazıya dökülmesi,
- Metin seslendirme,

- Yazım yanlışlarının düzeltilmesi,
- Metinden bilgi çıkarımı,
- Metin özetleme,
- Diller arası çeviri,
- Duygu analizi,
- Konuşma tanıma,
- Konuşma üretme,
- Soru-cevap makineleri.

Doğal dil problemleri; biçimbilimsel (*morphology*), sözcüksel (*lexical*), gramer yapısal (*syntactic*), sözcük türü (*POS, part-of-speech*), anlamsal (*semantic*), içerik (*discourse*) ve fonetik (*phonology*) gibi aşağıda kısaca özetlenen farklı analiz seviyelerinde ele alınmaktadır [7, 9-11].

Biçimbilimsel analiz, kelimenin eklerinden arındırılarak kök/gövde seviyesinde çalışma sistematiğidir. Dilde tek başına anlamı olmayan, ancak birleşerek anlamlı kelimeler oluşturulabilen en küçük yapıtaşları (*morpheme*) ele alınmaktadır. Özellikle Türkçe'nin de içinde bulunduğu eklemeli dillerde biçimbilimsel çözümleme önem arz etmektedir.

Sözcüksel analiz, insanlarca yorumlanan ve anlaşılabilen dilin birimi olan kelimeler boyutundaki çalışmaları ifade etmektedir. Bu seviyedeki çalışmalarda özellikle sözlüklerden sıklıkla faydalanılmaktadır.

Gramer yapısal analiz, dilbilim kurallarına göre yürütülen çalışmalardır. Cümle seviyesinde dil kuralları işletilir. Kelimelerin konumları, diziliş sırası, kullanılan bağlaçlarla doğru anlamın elde edilmesine çalışılır. Özellikle sosyal medya paylaşımları ve kullanıcı yorumlarının resmi dil kurallarına tam olarak uymadığı göz önüne alındığında, formal metinlerin analizi gibi daha dar bir kapsamda uygulama alanı bulabilmektedir.

Sözcük türü (POS) analiz kelimelerin isim, sıfat, fiil, zarf ve edat gibi türlerini temel olarak yapılan çalışmalardır. Makine öğrenmesi temelli çalışmalarda sıklıkla kullanılmaktadır.

Anlamsal analiz, kelime anlamlarından yola çıkarak cümlelerin doğru anlaşılmasına yönelik çalışmalardır. Kelimelerin her dilde birden fazla anlamı olabilmektedir. Cümle içerisinde hangi anlamda kullanıldığını tespit etmek için diğer terimlerden, farklı dil kütüphanelerinden, dil kurallarından ve kelime türlerinden faydalanılmaktadır. Literatürün halihazırda sıcak başlıklarından olan bu problem, belirsizliğin giderilmesi (*word-sense disambiguation*) olarak isimlendirilmektedir. Bu alandaki gelişmeler sayesinde, bilgisayarların insanlar gibi dili anlayabileceğine çok yaklaşıldığı öngörülmektedir.

İçerik analizinde ise, girdi cümle seviyesinin de üzerinde metin seviyesinde ele alınmaktadır. Örneğin, zamirler ve kime atıf yaptıklarının tespiti, metne bütüncül bakmayı ve doğru anlamın yakalanmasını kolaylaştıracaktır. Özellikle gazete metinleri bu anlamda içerik bölümlerine ayrıştırılmakta ve başlık, haberdeki hikaye, önceki olaylar, alıntılar çıkarımsanmaktadır.

DDİ’de sadece metinler üzerinde değil, ses türü girdiler üzerinde de çalışılmakta olup, bu çalışmalarda fonetik analiz yapılmaktadır. Sesler, kelimelerdeki telaffuzlar ve seslendirmedeki dalgalanmalar ile vurgular analize konu olan alt fonetik başlıklardır.

DDİ disiplininde karşılaşılan zorluklar ise şu başlıklar altında özetlenebilir [5, 12-14].

- İçeriklerin dil kurallarına uygun olmaması.
- Terimlerin farklı anlamlara sahip olmasından dolayı, cümle içinde hangi anlamda kullanıldığının tespit edilerek belirsizliğin giderilmesi problemi.
- Deyim, atasözü gibi kalıpların tamamen farklı anlamlara sahip olabilmesi.
- Terimlerin argo, hiciv ve alay gibi farklı anlamlarda kullanılabilmesi.

- Dilin kurallarının her dile özgü olmasından dolayı ortak bir kurallar manzumesinin geliştirilememesi.
- İngilizce dışındaki birçok dilde gelişmiş dil kaynaklarının ve araçlarının olmaması/yetersizliği.
- Çeviri yoluyla elde edilen metinlerin/kaynakların, dilden dile ve kültürden kültüre farklılıklara sahip olması.
- Arka plan bilgisi (*background knowledge*) gerektiren metinlerin çözümlenmesi.
- Ses çözümlemesinde, farklı lehçelerin varlığı.

2.1.2. Duygu Analizi

Duygu, Türk Dil Kurumu sözlüğünde¹ “duyularla algılama, his”, “belirli bir nesne, olay veya bireylerin insanların iç dünyasında uyandırdığı izlenim” ve “nesneleri veya olayları ahlaki ve estetik yönden değerlendirme yeteneği” ifadeleriyle tanımlanmaktadır.

Web teknolojileri ve sosyal ağlar sayesinde üretilen içerik devasa boyutlara ulaşmıştır. Duygu analizi (*sentiment analysis*) DDİ’nin bir alt disiplini olup içeriklerdeki görüş, duygu, düşünce ve tutumları belirlemek için çalışan bir bilim dalıdır. Birbirlerinden küçük farklılıkları olsa da şu başlıklarla da literatürde ifade edilmektedir: “Görüş madenciliği” (*opinion mining*), “görüş çıkarımsama” (*opinion extraction*), “duygu madenciliği” (*sentiment mining*), “öznellik analizi” (*subjectivity analysis*), “etki analizi” (*affect analysis*), “duygulanım analizi” (*emotion analysis*), “görüş madenciliği” (*review mining*) [15-17]. Yakın zamanda tüm bu başlıklar duygu analizi şemsiyesi altında toplanmış ve sektörde “duygu analizi” başlığı benimsenmiştir. Akademik dünya ise “duygu analizi” tanımına ilaveten “görüş madenciliği” ifadesini de kullanmaktadır [15].

Duygu analizi birçok araştırmacı tarafından, bir sınıflandırma problemi olarak da tarif edilmektedir. Bu sınıflandırma ikili (pozitif, negatif) olabileceği gibi üçlü (pozitif, negatif, nötr) şeklinde de olabilmektedir [18-25]. Ayrıca dördü (pozitif, negatif, nötr,

¹ <http://www.tdk.gov.tr>

karişık) [26] ile kullanıcıların yıldız değerlendirme sistematüğinden esinlenilen beşli (çok negatif, negatif, nötr, pozitif, çok pozitif) [27] sınıflandırmalara da rastlanmaktadır. Bu sınıflandırmalara ilaveten, duygunun pozitiflik/negatiflik derecesini belirlemeye yönelik çalışmalar da duygu analizi kapsamındadır [28] ve bu yönüyle bir regresyon problemi olarak da algılanmaktadır [29]. Ancak duygu analizinde genellikle hedeflenen, içeriğın analiz edilerek yazarın duygusunun tespit edilmesi olmuştur [28, 30].

2.1.3. Duygu Analizinin Uygulama Alanları

Duygu analizi, karar vericilere ve her ölçekteki politika belirleyicilere öngörü imkanı sunabilecek ve bakış açılarının zenginleşmesine olanak sağlayabilecek kapasiteye sahiptir. Gelişime ve değışime açık bütün birimler için objektif bir geri besleme mekanizmasıdır. Bunun yanında, sosyal medya platformlarının ulaştığı seviye göz önüne alındığında, duygu analizi bireysel ölçekte de çok değerli veriler sunmaktadır. Genel hatlarıyla duygu analizinin uygulama alanları günümüzde ve gelecekte şu şekilde sıralanabilir:

- Ürün ve hizmete yönelik değerlendirmeler doğrultusunda en doğru reaksiyonun belirlenerek müşteri ilişkilerinin etkin yönetilebilmesi [31], ürün geliştirme ve AR-GE departmanlarına doğru ve sağlıklı geri bildirimler sunulabilmesi.
- Hedef varlıklara (politikacılar, filmler, ürünler, ülkeler vb.) yönelik bütün dijital platformlardaki duygu algılarının zaman eksenindeki değışiminin izlenebilmesi [20, 32].
- Kullanıcıların duygu durumlarına göre uygun diyalog sistemlerinin geliştirilebilmesi [33].
- Çevrimiçi eğitim platformlarında, eğitim alan bireyin duygu durumuna göre sistemin otomatik olarak eğitim içeriğini/akışını güncelleyebildiği akıllı sistemlerin geliştirilebilmesi [34].
- Duygu tonlaması da yapabilen gerçeğe daha yakın metin okuma sistemlerinin geliştirilebilmesi [35, 36], bilhassa görme engelli genç bireylerin eğitimine katkı sunulabilmesi.
- Ekonomi haberlerinin yarattığı etkinin doğru belirlenmesiyle, tüketici, üretici ve yatırımcılara daha sağlıklı bir öngörü imkanı sunulabilmesi.

Halihazırda araştırmacılar tarafından geliştirilen duygu analizi metodolojileri; ürün yorumları, film yorumları, otel yorumları, forumlar, bloglar, Facebook gibi sosyal paylaşım siteleri, haberler, haberlere yapılan yorumlar ve Twitter gibi mikroblog veri setleri üzerinde uygulanmaktadır [37].

Duygu analizi her ne kadar DDI'nin bir alt dalı olarak görülse de, bireylerin duygu ve düşünceleri çalışmanın öznesi olduğundan, yönetim bilişim sistemleri, sosyoloji, politika ve ekonomi/finans disiplinlerinin de ilgi alanına girmiş bulunmaktadır [38, 39].

2.1.4. Duygu Analizi Seviyeleri

Duygu analizi çalışmaları genel olarak 3 seviyede gerçekleştirilmektedir [38]. Bunlar: doküman seviyesi (*document-level*), cümle seviyesi (*sentence-level*) ve hedef tabanlı (*aspect-based*) duygu analizidir.

Doküman seviyesi duygu analizinde girdi olarak bir doküman verilir. Bu dokümanın tamamının analizi sonucunda, yansıttığı duygunun olumlu veya olumsuz olduğunun tespiti yapılır [40, 41]. Doküman içeriğinde farklı varlıklar, hatta farklı duygular bulunsa dahi, doküman neticede tek bir duygu (pozitif/negatif) ile etiketlenmektedir. Bu özelliği ile geleneksel metin sınıflandırma problemi gibi değerlendirilir. Duygu analizi çalışmalarının ilk dönemlerinde sıklıkla başvuru doküman seviyesindeki çalışmalar, bugüne gelindiğinde yetersiz kaldığı gerekçesiyle eleştirilmektedir. Ancak diğer taraftan da, çevrimiçi ürün yorumları gibi içerik hedefinin tek bir nesne olması durumlarında etkin olarak kullanılmaktadır.

Cümle seviyesi duygu analizinde ise doküman daha alt seviyede, cümleler bazında ele alınır. Her bir cümlelerin duygu polaritesi (pozitif/negatif/nötr) tespit edilmeye çalışılır. Bu karakteristiğinden dolayı Wiebe ve arkadaşları [42] tarafından öznellik analizi (*subjectivity analysis*) problemi olarak da nitelendirilmiştir. Öznellik analizi, duygu analizinin bir önceki adımıdır ve cümle öznellik içermiyorsa nötr olarak etiketlenir ve duygu analizi süreci işletilmez. Fakat, dilin esnekliğinden de kaynaklanan gerekçelerle

birçok metin görünürde nesnel olsa bile üstü kapalı olarak öznellik içermektedir [38, 43]. Bu durumda cümledeki duygu, pozitif veya negatif olarak etiketlenir. İkili veya üçlü sınıflandırma hususu, araştırmacıların doküman seviyesinde olduğu gibi cümle seviyesinde de hemfikir olamadıkları konulardan birisidir.

İçeriklerin birçoğunda tek bir özne ve bu özneye bağlı tek bir duygu bulunmamaktadır. Özellikle bir ürüne bağlı kullanıcı yorumlarında özne tek olmakla beraber, bu öznenin farklı nitelikleri hakkında farklı duygular ifade edilmektedir. Örneğin bir sinema filmine yapılan yorumda kullanıcı, filmin senaryosunu beğendiğini ancak süresini uzun bulduğunu ifade edebilir. Burada özne (sinema filmi) tek olmakla beraber, “senaryo” niteliği hakkında duygu pozitif iken, “süre” niteliği hakkındaki duygu negatiftir. Doküman ve cümle seviyesindeki analizler daha genel düzeyde kaldığından, bu problem, hedef tabanlı (*aspect-level*) duygu analizi ile adreslenmektedir [38]. Geçmiş çalışmalarda özellik seviyesi (*feature level*) [44] olarak isimlendirilmiş olan hedef tabanlı duygu analizi, yakın dönemde yoğun çalışılan duygu analizi konulardan birisidir.

2.1.5. Duygu Analizinin Zorlukları

Duygu analizi her ne kadar metinler üzerinde yürütülmüşse de, araştırmacılar paylaşımların sadece metin tabanlı değil, beraberinde video, görsel ve ses öğeleri ile desteklendiğini ve bu yönlerin de analiz sürecine dahil edilmesi gerektiğini yakın zamanda ileri sürmektedirler [45]. Ancak, metinler üzerindeki çalışmalar, İngilizce dili haricinde diğer dillerde henüz olgunlaşmamıştır ve bazı dillerde de neredeyse emekleme safhasındadır. Metin tabanlı içerikler üzerinde duygu analizi literatürde halen popüler olan bir konudur. Dolayısıyla duygu analizindeki zorluklar ifadesi, içeriğin metin olduğu durumları tarif etmektedir. Duygu analizinde ele alınması gereken ana hususlar literatürde ağırlıklı olarak şu başlıklar altında yer almaktadır [38, 46, 47].

- Olumsuzluk Durumları (*negation*): Dillerdeki bazı olumsuzluk ifadeleri, kendilerini takip eden terimlerin duygu değerini tersine çevirmektedir. İngilizcede “*not*” ifadesi, Türkçede “değil” kelimesi bu duruma örnektir. Bu konu Bölüm 2.2.2.’de daha detaylı açıklanmıştır.

- Şiddetlendirme (*intensification*): Kendilerini takip eden terimlerin duygu skorunu güçlendiren veya zayıflatan ifadelerle genel olarak sırasıyla kuvvetlendiriciler (*intensifiers*) ve zayıflatıcılar (*diminishers*) ismi verilmektedir. Bunlara örnek olarak İngilizcede “very”, Türkçede ise “çok” kelimeleri verilebilir. Bu konu da Bölüm 2.2.2.’de daha detaylı şekilde verilmiştir.
- Koşul Cümleleri (*conditional sentences*): Günlük yaşantıda, diyaloglarda ve metinlerde sıklıkla kullanılan koşul cümlelerinin duygu analizinde, terimlerin zaman ekleri ve kelime türleri önemli hale gelmektedir. Örneğin “Çok rahat olmasaydı beğenmezdim.” koşul cümlesi, normal cümlelerin dışında bir çözümleme sistematığına ihtiyaç duymaktadır. Narayanan ve arkadaşları [48] bu problem alanındaki öncü sayılan çalışmalarında, koşul cümlelerini zamanlarına göre 4 gruba ayırmışlar ve ardından duygu belirten terimleri, bu terimlerin cümledeki pozisyonlarını ve POS etiketlerini öznitelik olarak kullanmışlardır.
- Onaylayıcı Sorular (*rhetorical questions*): Onaylayıcı sorular, soru formatında gibi görünse de aslında bir mesaj içeren ve duygu aktaran yapılardır. Örnek olarak “Bu yemek beğenilmez mi?” cümlesi verilebilir. Bu cümlede aktarılan ama metnin çözülmesinde gizli kalan, yüksek pozitif değere sahip bir duygudur.
- Kinayeli İfadeler (*sarcastic phrases*): Çoğunlukla pozitif gibi görünen, ama iğneleyici bir duygu aktaran yapılardır. Örnek olarak “Ne demezsin harika olmuş.” ve “Renk seçimine BAYILDIM!” cümleleri verilebilir. Bu cümlelerdeki duygu analizinin doğruluğu, kaynak ve araç bakımından zengin İngilizce dilinde de oldukça düşüktür [15]. Literatürde ünlem işaretinin, duygu simgelerinin (*emoji*) ve tamamı büyük harfle yazılmış terimlerin varlığını kinayeli ifadelerin karakteristiği olarak kabul eden çalışmalar mevcuttur [49-51].
- Bağlaçlar (*conjunctions*): Bağlaçlar, bağladıkları terimlerin duygu polariteleri hakkında çok güçlü bilgi sunmaktadır. İngilizcedeki “and”, “but” ile Türkçedeki “ve”, “fakat” bağlaçları örnek olarak verilebilir. Duygu değeri bilinen bir terim ile bilinmeyen bir terim cümle içerisinde bağlaçla kullanılmışsa, bilinmeyen terimin duygu yönü kolaylıkla çıkarımsanabilir. Hatzivassiloglou ve McKeown [52] bu çözüm önerisinin ilk temsilcilerinden olup, duygu polaritesi bilinmeyen kelimelerin duygu yönünün belirlenmesinde bu yöntemden faydalanmışlardır.

- Deyim İfadeleri (*idiomatic issues*): Deyimler, doğal dildeki zenginliği ve esnekliği gösteren en önemli yapılardır. Çünkü bir deyimın anlamı, kendisini oluşturan kelimelerin anlamlarından oldukça uzaktır ve araştırmacılar tarafından ayrı bir problem alanı olarak tanımlanmaktadır. Örneğin “bıyık altından gülmek” deyimini negatif duygu yönüne sahip iken, geleneksel duygu analizinde “gülmek” deyiminin aşırı pozitif skoru nedeniyle pozitif yönlü duygu olarak tespit edilebilir. Literatürde, deyim ifadelerinin derlenmesi ve aktardıkları duyguların tespiti, manuel veya kitle kaynaklı (*crowdsourcing*) yöntemlerle yapılagelmiş ve ardından duygu analizi süreçlerine dahil edilmişlerdir. Williams ve arkadaşları [53], deyimlerin sentiment analizindeki rollerine yönelik yaptıkları çalışma kapsamında, İngilizcede 16 ana tema belirlemişler ve bu temalara yönelik toplamda 580 deyimini tespit ederek, kitle kaynaklı yöntemle de duygu kategorisini belirlemişlerdir. Çalışmanın sonucunda, deyimlerin ele alınmasıyla performansta ciddi bir artışın olduğunu ortaya koymuşlardır.
- Arka Plan Bilgisi (*background knowledge*): Bireyler iletişimde bulundukları süreçte aslında farkında olmadan oldukça yoğun arka plan bilgisi kullanmaktadır. Örneğin “2-1 yendik” ifadesi ile konunun bir müsabaka olduğu, hatta bu müsabakanın spor olduğu, bunun da ötesinde futbol maçından bahsedildiği bilgisi, yazan ve okuyan tarafların arka plan bilgisine dayanmaktadır. Bu özelliği nedeniyle arka plan bilgisi, DDİ dünyasının halihazırdaki en zor problemlerinden birisidir.

Bu ana hususların yanında duygu analizinde yaşanan diğer zorlukları şu şekilde listelemek mümkündür.

- Sözlük ihtiyacının bulunması, var olan sözlüklerin yetersizliği ve bunlara ilaveten alana özgü sözlüğe bazı çalışmalarda ihtiyaç duyulması.
- Birçok dilde, genel erişime açık ve etiketlenmiş bir derlemin bulunmaması.
- Sosyal platform içeriklerindeki dilin, dil kurallarına uymaması.
- İroni, hiciv, deyim ve atasözü metinleri, imalı ifadeler ile Twitter paylaşımları gibi kısa metinlerin çözümlenmesindeki zorluklar.

- Geliştirilen duygu analizi metodolojilerinin başarımının test edilmesinde, genel kabul görmüş derlemin eksikliği.
- Bir dil için geliştirilen metodoloji ve yaklaşımların, çoğunlukla diğer bir dile doğrudan uygulanamaması ve bunun sonucu olarak inovasyonun yaygınlaşamaması.
- Duygu analizi çalışan bazı araştırmacıların, dil olarak kendi dili dışında halihazırda gelişmiş araçları bulunan İngilizce dilini tercih etmesi.
- İngilizceden hedef dile çevrilen kaynakların, diller ve kültürler arasındaki farklılıklardan dolayı istenilen performansı karşılayamaması.

Türkçede de yukarıda anılan zorluklar mevcut olmakla beraber, kendine has özelliklerinden dolayı ilave zorluklar da söz konusudur. Türkçedeki zorluklar Bölüm 2.5.1’de ayrıca ele alınmıştır.

2.2. Duygu Analizi Yaklaşımları

Duygu analizinde kullanılan yaklaşımlar genel olarak iki ana gruba ayrılır: makine öğrenmesi temelli yaklaşımlar ve sözlük temelli yaklaşımlar. Makine öğrenmesi yaklaşımları, denetimli ve denetimsiz öğrenme metodolojileri ile duygu analizinde sıklıkla kullanılmaktadır [17, 20, 51]. Sözlük temelli yaklaşımlarda ise gereksinim duyulan sözlük duygu sözlüğü olup, Bölüm 2.2.2 ve Bölüm 2.4’de detaylarıyla ele alınmıştır.

Dilin kendine özgü kuralları, içeriğin türü, dildeki mevcut kaynak ve araçların durumu dikkate alındığında, bu iki yaklaşımın güçlü ve zayıf oldukları noktalar vardır. Farklı problemler ve çözüm arayışları sonucunda, iki yaklaşımı barındıran hibrit yöntemler de kullanılmaktadır. Genel olarak, sözlük temelli yaklaşımlar ölçeklenebilirliği ile ön plana çıkarken, makine öğrenmesi temelli yaklaşımlar ise alana özgü çalışmalarda yoğun olarak tercih edilmektedir. Literatürde üzerinde mutlak uzlaşa sağlanan bir yaklaşım mevcut değildir.

2.2.1. Makine Öğrenmesi Temelli Duygu Analizi Yaklaşımları

Duygu analizinde duygular genellikle pozitif ve negatif olarak sınıflandırılmaktadır. Makine öğrenmesi disiplininin en yoğun başlıklarından birisi de sınıflandırma problemleridir ve duygu sınıflandırma sürecinde makine öğrenmesi algoritmaları sıklıkla kullanılmaktadır. Makine öğrenmesinde ana belirleyici nokta, süreçte kullanılacak özniteliklerin doğru belirlenmesinde yatmaktadır. Geleneksel metin sınıflandırmada terimlerin sayıları, dokümanların kategorileri ve başlıklar nitelikleri belirlerken, duyguların sınıflandırılmasında ise genel olarak yazılımsal ve dilbilimsel özellikler ile duygu belirleyen terimler öznitelikleri oluşturmaktadır [17].

Makine öğrenmesi, denetimli ve denetimsiz olmak üzere iki genel kategoride ele alınır. Denetimli öğrenmede, eğitim ve test için kullanılacak etiketli veri setlerine ihtiyaç vardır. Farklı öğrenme algoritmaları eğitim veri seti kullanılarak eğitilir, performansları test veri seti üzerinde ölçülür ve polaritesi bilinmeyen içerik, öğrenilen modele tabi tutularak duygu sınıflandırması gerçekleştirilir.

Denetimli öğrenme ile duygu analizinde sıklıkla kullanılan sınıflandırıcılar Naive Bayes (NB), Maksimum Entropi (ME) ve Destek Vektör Makinesi (SVM: *Support Vector Machine*) algoritmaları olmuştur [54]. Bu alanın öncüleri olan Pang ve arkadaşları [40] tarafından film yorumları üzerinde yapılan çalışmada NB, ME ve SVM sınıflandırıcılar, n-gram ($n \leq 2$) ve sözcük türü etiketlerden oluşan özniteliklerle kullanılmıştır. Çalışmanın sonucunda, öznitelik olarak 1-gram ve sınıflandırıcı olarak da SVM ile en iyi performansın elde edildiği ifade edilmiştir. Yakın zamanda Mahendran ve Mekala [55] tarafından yapılan çalışmada ise, veri seti olarak ürün yorumları kullanılmış, öznitelik olarak n-gram ($n \leq 2$), sözcüklerin POS bilgileri ile sözcüklerin konum bilgileri seçilmiştir. NB, ME ve SVM sınıflandırıcılar kullanılarak yapılan çalışmada, SVM ile elde edilen sonucun daha yüksek olduğu belirtilmiştir. Parlar ve arkadaşları [56] ise, öznitelik seçimine yönelik yeni bir metodoloji önermiş, hem Türkçe hem de İngilizce film ve ürün yorumları üzerinde test etmişlerdir. Çalışma kapsamında NB'nin özel bir versiyonu olan çokterimli NB (NBM - *Naive Bayes Multinomial*) sınıflandırıcı ile SVM, Lojistik Regresyon (LR) ve Karar Ağaçları (J48) algoritmaları kullanılmıştır. Sonuçta; önerdikleri öznitelik seçim metodolojilerinin Ki-Kare (*Chi square*) ve Bilgi

Kazanımı (*Information Gain*) gibi diğer geleneksel öznitelik seçimi yaklaşımlarından daha iyi sonuç verdiğini, NBM sınıflandırıcının Türkçe beş farklı veri setinin hepsinde de en yüksek performansı gösterdiğini, İngilizce veri setleri üzerinde ise SVM ve LR sınıflandırıcıların başarımının yüksek olduğunu tespit etmişlerdir. Bu çalışmalar da göstermektedir ki, öznitelik seçimi ve sınıflandırma algoritması duygu analizinin performansının ana belirleyicileri olup, dillere göre de farklılık göstermektedir.

Denetimsiz öğrenmede, sistemin eğitiminde etiketsiz veriler kullanılır. Bir başka ifadeyle, elde bulunan veri seti etiketsiz ise, denetimsiz öğrenme yaklaşımları tercih edilir. Ancak buradaki adreslenen problem sınıflandırma değil, kümelemedir. Bu alana yönelik literatürde yer alan temel çalışmalardan birisi Turney [41] tarafından yapılmıştır. İki ana adıma sahip çalışmanın ilk adımında, kelime türlerinin ardışıklığına göre (örneğin: sıfat+isim, zarf+sıfat vb.) kurallar belirlenerek, bu kurala uyan terimler derlemde elde edilmiştir. Ardından elde edilen bu terimlerin, hem pozitif (örnek: *excellent*) hem de negatif (örnek: *poor*) referans terimlerle arasındaki ilişki ölçülmüştür. Bu ölçümde, istatistiksel olarak iki ardışık terimin birlikteliği hakkında bilgi veren bir metrik olan PMI (*Pointwise Mutual Information*) [57] kullanılmıştır. PMI değerlerinin elde edilmesinde de arama motorunda (*AltaVista*) terimlerin birlikteliğine yönelik elde edilen arama sonuç sayısı kullanılmış ve yorumların PMI ortalama değeri ile duygu polaritesi belirlenmiştir. Bu metodoloji ile araba yorumları veri setinde %84 doğruluk değeri elde edilmiştir. PMI gibi istatistiksel yaklaşıma sahip LDA (Latent Dirichlet Allocation) da denetimsiz öğrenmede kullanılan metotlardan birisidir. LDA, her bir dokümanda açıkça ifşa edilmeyen gizli konuların (topics) olduğunu varsayar ve bunların tespitine yönelik olasılıksal bir yaklaşım sunar. Örneğin, bir öğrenci kompozisyon yazma ödevini yerine getirirken, akıcı bir metin olması için öncelikle konunun ana hatlarını ifade eden terimler tespit eder. Ardından bu ana hatları (örnek: sağlığın önemi, sağlık hizmetleri, sağlık personeli vb.), uygun kelimelerle ifadeleştirir. Bu kompozisyona LDA gözüyle baktığımızda, gizli kalan ana hat ifadeleridir ve bunlar dokümandaki kelimeler kullanılarak çıkarılır. Xianghua ve arkadaşları [58], yorumlar ve blog sayfalarından elde ettikleri veri seti üzerinde LDA ile gizli kalan ana hatları elde etmişler, ardından bu ana hatların duygu polaritelerini belirlemek için de halihazırda bulunan Çince HowNet sözlüğünden faydalanmışlardır. Çalışmada sunulan yenilik olarak, adını kayan pencere (*sliding windows*) olarak verdikleri yaklaşımla yorumlar

üzerinde LDA'nın uygulanmasını ifade etmişler, çalışmanın sonucunda da duygu analizi başarımını %92.15 olarak hesaplamışlardır.

2.2.2. Sözlük Temelli Duygu Analizi Yaklaşımları

Sözlük temelli duygu analizindeki gereksinim duygu sözlüğüdür (*sentiment lexicon*). Literatürde görüş sözlüğü (*opinion lexicon*) olarak da geçen duygu sözlüğü, içeriğindeki terimlerin duygu yönünü (negatif/nötr/pozitif) ve duygu skorunu sayısal olarak tutan bir veritabanıdır [44, 59, 60].

Sözlük temelli duygu analizi yapılırken, analiz edilecek içerikte geçen her bir terim duygu sözlüğünde aranır. Sözlükte bulunursa, yön ve skor değerleri toplam skora eklenir. Sonuçta elde edilen toplam skor, sıfırdan küçükse negatif, büyükse pozitif, sıfıra eşit çıkarsa da metnin duygusu nötr olarak sınıflandırılır.

Literatürde, sözlükteki sayısal değerlere ve toplam skorun hesaplanmasına yönelik farklı yaklaşımlar da geliştirilmiştir. Dildeki bazı terimler, kullanıldığı yere göre duyguyu tersine çevirmektedir ve bu terimlere genel olarak “olumsuzluk” (*negation*) ifadeleri denir. İngilizce için “not” ve “never”, Türkçe için de “değil” terimleri örnek olarak verilebilir. Metinde geçecek “iyi değil” (“not good”) gibi bir ifade, normalde pozitif yön ve skora sahip “iyi” sözcüğünü olumsuzlaştırarak negatife dönüştürmektedir. Hu ve Liu [44] bu yaklaşımı da dahil ettikleri çalışmalarını beş farklı ürün yorumu veri seti üzerinde uygulayarak, performans artışını tespit etmişlerdir. Benzer şekilde Kim ve Hovy [61] de, farklı modellerle yaptıkları çalışmalarında bu yaklaşımla elde edilen genel performansın daha yüksek olduğunu ifade etmişlerdir. Türkçede, Dehkharghani [47] doktora tezi kapsamındaki çalışmasında olumsuzluk ifadelerini de göz önüne almıştır. Duygunun zıt yönüne geçişini sağlayan terim ve yapılar olarak: “değil” kelimesini, “ma” ve “me” eklerini alan fiilleri, “lu” ve “suz” eklerini alarak sıfat olan isim türündeki ifadeleri belirlemiştir. Bu yaklaşımının %2 civarında sınıflandırma başarısına katkı sunduğunu tespit etmiştir.

Olumsuzluk ifadelerinin, terimin duygu polaritesini tersine çevirmesine yönelik bir takım problemler de mevcuttur [38]. Örneğin bu yaklaşımla, “mutsuz değilim” ifadesi “mutluyum” şekline evrilmiştir. Ancak “mutsuz değilim” diyen birisi, aslında “mutlu” olduğunu da ifade etmemektedir. Polayni ve Zaenen [62] olumsuzluk ifadeleri dışında, duygu polaritesinin değiştiren farklı etkenlerin de olduğunu çalışmalarında belirtmişler ve bu etkenlere tutum değiştiriciler (*valence shifters*) ismini vermişlerdir. Bu kapsamda bir başka problem, karşılaştırmalı cümlelerde ortaya çıkmaktadır. “Çalışma temposu geçen haftadan iyi değil.” ifadesinden, çalışma temposunun geçen haftadan daha kötü olduğu gibi bir sonuç çıkarmak yanlış olacaktır. Çift olumsuzluk ifadesinin aynı cümlede geçtiği durumlar da, tutum değiştiriciler konusunda ayrıca ele alınması gereken istisnalardır. Bu duruma “güzel değil demek doğru değil” ifadesi örnek olarak verilebilir. Olumsuzluk ifadeleriyle duygu polaritesinin ters çevrilmesi, isim öbeklerinde de hataya sebep olabilmektedir. Örneğin “Ben okulu beğenmiyorum, bölümü değil.” ifadesinde bu durum görülmektedir.

Duygu polaritesini değiştiren etkenler kapsamında bir diğer çalışma Kennedy ve Inkpen [63] tarafından yapılan, kuvvetlendirici (*intensifier*) ve zayıflatıcı (*diminisher*) ifadelerin etkisidir. Kuvvetlendiriciler, pozitif ve negatif duyguyu artırırken, zayıflatıcılar pozitif ve negatif duygunun şiddetini azaltırlar. Örneğin, “iyi” pozitif terimi, “çok iyi” şeklinde kuvvetlendirici terimle beraber kullanıldığında daha pozitif bir duyguyu yansıtmaktadır. Benzer şekilde “nispeten iyi” gibi bir kullanımda ise zayıflatıcı etkinin sonucu olarak, daha az pozitif bir duygu vurgusu yapılmaktadır. Çalışmada bu etkilerin hesaplamaya dahil edilmesi için, duygu sözlüğündeki bütün pozitif terimler +2 olarak kullanılmış, kuvvetlendirici ile beraber kullanılanlara +3, zayıflatıcı ile kullanılan terimlere de +1 değeri atanmıştır. Benzer şekilde, bütün negatif terimler de -2, kuvvetlendirici ve zayıflatıcılarla kullanıldığından sırasıyla -3 ve -1 değerleri ile hesaplama sürecine sokulmuşlardır.

Kuvvetlendirici ve zayıflatıcıları farklı hesaplama yaklaşımıyla sürece dahil eden çalışmalar da mevcuttur. Taboada ve arkadaşları [28] tarafından yapılan çalışmada, terimlere duygu değeri -5 (en negatif) ile +5 (en pozitif) arasında olacak şekilde değer atanmış ve sıfır değeri kullanılmamıştır. Kuvvetlendirici ve zayıflatıcıların etkisi

yüzdelik oranlarla kullanılmıştır. Örneğin, *slightly* -%50, *somewhat* -%30, *pretty* -%10, *really* +%15, *very* +%25, *extraordinary* +%50 ve *most* +%100 oranlarında beraber geçtikleri terimlerin duygu skoruna etki etmektedir. Yine bu çalışmada, kuvvetlendirici ve zayıflatıcı etkisi, POS bilgileri (sıfat ve zarf), niteleyiciler (epey, pek çok vb.), büyük harf kullanımı, ünlem işareti ile bağlaçların (ve, fakat vb.) durumları göz önüne alınarak genişletilmiştir.

Türkçe yapılan çalışmada Dehkharghani [47], kuvvetlendirici ifadeler olarak “bayağı, gayet, çok, gerçekten, iyice, cidden”, zayıflatıcı ifadeler olarak da “biraz, azcık, yaklaşık” terimlerini belirlemiştir. Kuvvetlendirici ifadelerle birlikte geçen terimlerin duygu skorunun iki katını, zayıflatıcı ifadelerle birlikte geçenlerin de yarısını almıştır. Bu yaklaşımın sınıflandırma performansına %1 oranında katkı sunduğunu tespit etmiştir. Türkmengoğlu [9]’da Türkçe metinlerde duygu analizi tez çalışmasında, kuvvetlendirici ve zayıflatıcı olarak toplamda 20 terim belirlemiştir. Bu terimlere de çarpan katsayıları atamış (örnek: “acayip” +2, “cidden” +1, “azıcık” -1) ve kendilerinden sonra gelen terimin duygu skorunun sayısal çarpanı olarak hesaplamaya dahil etmiştir.

Yukarıda bahsedilen istisnai durumların daha yapısal bir yaklaşımla ele alınması gerekliliği Jia ve arkadaşları [64] tarafından önerilmiştir. Dilin kuralları kapsamında bağımlılık çözümleme ağacı (*dependency parse tree*) üzerinde bir takım kurallar geliştirmişlerdir. Bu metodolojide olumsuzluk ifadeleriyle ilişkili kelimelerin tespitine odaklanılmıştır. Türkçede web servisi üzerinden erişime sunulan bağımlılık çözümleme ağacı, Eryiğit [65] tarafından geliştirilmiştir. Dehkharghani ve arkadaşları [46] da, Türkçede farklı duygu analizi seviyelerini temel alan çalışmalarında bu araçtan faydalanmışlardır.

Sözlük tabanlı duygu analizindeki başarımların performansını etkileyen ana hususlardan bir diğeri de, genel amaçlı duygu sözlüğünün bazı özel alanlardaki performans kaybıdır. Doğal dillerde bir terim farklı anlamlara sahip olabilmektedir. Oldukça yaygın olan bu durum, cümle içindeki terimin hangi anlamda kullanıldığının göz ardı edilerek, sözlükte tek bir değerle karşılanması ile sonuçlanmaktadır. Bir diğer ifadeyle, genel amaçlı

duygu sözlüğü kullanıldığında terimler, bağlamına bakılmaksızın değerlendirilmektedir. Örneğin “örgüt” kelimesi, güvenlik içerikli bir metinde negatif duygu barındırırken, işletme gibi bir takım disiplinlerde nötr duygu tonunda ve “kurum” anlamında kullanılmaktadır. Çözüm olarak sunulan önerilerin başında, alana özel duygu sözlüğü kullanılması/geliştirilmesi gelmektedir ve bu konu üzerinde Bölüm 2.4.2’de durulmuştur.

2.3. Duygu Sözlüğü Oluşturmak

Duygu sözlüğü, içeriğindeki terimlerin duygu yönünü (negatif/nötr/pozitif) ve duygu skorunu sayısal olarak tutan bir sözlük türüdür. Türkçe bir duygu sözlüğünde “güzel”, “mükemmel”, “harika” kelimeleri pozitif duygu veren terimlere; “berbat”, “çirkin”, “nefret” gibi kelimeler de negatif duygu veren terimlere örnek olarak verilebilir. Duygu sözlüğünde tekil kelimelerin yanında kelime öbekleri ve deyimler de bulunabilir.

Duygu sözlüklerinin oluşturulmasında genel olarak 3 yaklaşım mevcuttur. Bunlar:

- Manuel yaklaşım,
- Sözlük temelli (*dictionary-based*) yaklaşım,
- Derlem temelli (*corpus-based*) yaklaşım.

2.3.1. Manuel Yaklaşımla Duygu Sözlüğü Oluşturmak

Duygu sözlüğü oluşturmanın en maliyetli yolu manuel yaklaşımdır. Sözlüğün kapsamı nispeten dar olmakla beraber, doğruluk derecesi diğer yaklaşımlardan daha yüksektir. Literatürde bu yaklaşımın, duygu sözlüğünün bulunmadığı veya edinilemediği durumlarda, ihtiyacı karşılayacak kadar dar kapsamlı bir duygu sözlüğünün geliştirilmesi amacıyla kullanıldığını da görmek mümkündür.

Nielsen [66], manuel olarak duygu polaritesini -5 ile +5 arasında bir tamsayı ile etiketlediği ve AFINN olarak isimlendirdiği İngilizce duygu sözlüğünü geliştirmiştir.

2009-2011 yılları arasını kapsayan çalışmanın son sürümü AFINN-111, toplamda 2477 terim içermektedir.

Abdul-Mageed ve arkadaşları [67], çalışmalarında Türkçe gibi biçimbilimsel açıdan zengin olan Arapça için duygu sözlüğünü geliştirmişlerdir. Bu doğrultuda, manuel olarak etiketlenmiş haber metinleri derleminden faydalanarak, 3892 sıfat türünde terimden oluşan duygu sözlüğü oluşturulmuştur.

Türkmenoğlu ve Tantuğ [68] çalışmalarında, Türkçede sözlük temelli ile makine öğrenmesi temelli yaklaşımlardan hangisinin daha iyi performans gösterdiğini tespit etmeye odaklanmışlardır. Çalışma kapsamında ihtiyaç duydukları Türkçe duygu sözlüğünü, İngilizcede kullanılan SentiStrength [51] isimli aracın sözlüğünü Türkçeye manuel olarak çevirerek elde etmişlerdir. İlk aşamada 2547 kelimedenden oluşan duygu sözlüğüne olumsuzluk ifadeleri ile gerekli gördükleri diğer kelimeleri de ekleyerek, toplamda 3657 kelimedenden oluşan Türkçe duygu sözlüğü oluşturulmuştur.

Günümüzde internet üzerinden kitle kaynağı (*crowdsourcing*) kullanılarak, manuel ama geniş katılımlı bir süreçle, duygu sözlüğü geliştirme çalışmaları da görülmektedir. Genellikle bireylerden hazırlanan sorulara cevap vermeleri veya bu amaç için geliştirilmiş oyunlar üzerinden veriler toplanarak sözlük oluşturma yoluna gidilmektedir.

Mohammad ve Turney [69], farklı sözlüklerden oluşturdukları veri setindeki terimlerin duygu polaritelerini, Amazon'un Mechanical Turk² isimli kitle kaynağı platformunu kullanarak çevrimiçi katılımcılardan aldıkları geri bildirimlerle belirlemişlerdir. Soru-cevap sistematığına sahip yapıda, soruların etkin seçimi ile hatalı cevapların elenmesine yönelik çözüm önerilerinin de sunulduğu çalışmanın sonucunda, toplamda 10170 terimden oluşan ve EmoLex ismi verilen İngilizce duygu sözlüğü geliştirilmiştir.

² <https://www.mturk.com/>

Korece dilinde An ve Kim [70] tarafından 35000 kolej öğrencisinin katılımıyla duygu sözlüğü geliştirme çalışması yapılmıştır. Folksonomi olarak da bilinen, öntanımlı ve merkezi bir sınıflandırma sisteminin yerine kullanıcılar tarafından konulan etiketler ile oluşan bir sınıflandırma sistemini ve oylama sürecini de çalışma kapsamında kullanmışlar, sonuçta 517K boyutunda Korece duygu sözlüğü geliştirmişlerdir.

Korece dilinde çalışma yapan bir diğer araştırmacılar da Hong ve arkadaşlarıdır [71]. Çalışmada duygu sözlüğü oluşturmak maksadıyla “Tower of Babel” isimli ve tetris benzeri çevrimiçi bir oyun geliştirilmiştir. Önerilen kitle katılımlı oyun sistematigi ile geleneksel manuel yaklaşımın performansı karşılaştırılmıştır. Sonuçta, önerilen sistematigin dil bağımsız olduğu ve geleneksel yaklaşımdan daha iyi sonuç verdiği vurgulanmıştır.

Lafourcade ve arkadaşları [72] tarafından duygu sözlüğü oluşturmak maksadıyla “LikeIt” isimli oyun geliştirilmiştir. Oyunda katılımcılardan, açık erişimli bir sözlükteki terimleri 3 sınıflı (negatif-nötr-pozitif) duygu ikonları ile etiketlemeleri beklenmektedir. Böylece mevcut terimlerin duygu polariteleri kitle katılımı ile tespit edilerek duygu sözlüğü oluşturulmuştur. 3 yıl gibi bir sürecin sonunda, Fransızca 385000 kelime, 100 milyon oylama ile polaritelenmiştir.

2.3.2. Sözlük Temelli Yaklaşımla Duygu Sözlüğü Oluşturmak

Sözlük temelli yaklaşımla duygu sözlüğü oluşturmak için iki temel sistematik ön plana çıkmaktadır. Bunlar:

- Tohum kelime (*seed words*) listesi, kelimelerin eş-zıt ilişkisi ve çizge (*graph*) yaklaşımının kullanımı,
- Çeviri sistematigi.

Birinci sistematikte, manuel hazırlanan bir tohum kelime listesi ile halihazırda var olan dil kaynaklarındaki eş-zıt ilişkilerden faydalanarak duygu sözlüğü oluşturulur. İngilizce

için geliştirilmiş kapsam ve içerik olarak da oldukça zengin olan WordNet [73] dil kaynağı, bu sistematikte sıklıkla kullanılmaktadır.

Hu ve Liu [44], müşteri yorumları üzerindeki çalışmaları için gerekli olan duygu sözlüğünü, sözlük temelli yaklaşımla oluşturmuşlardır. Sadece sıfat türündeki kelimelerden oluşan ve duygu polaritesi bilinen 30 kelimelik bir tohum kelime listesi hazırlanmış, ardından bu kelimelerin WordNet kütüphanesindeki eş ve zıt karşılıklarını bulan ve iteratif bir şekilde çalışan algoritma ile sıfatlardan oluşan İngilizce duygu sözlüğü geliştirilmiştir.

Bu tez kapsamındaki gibi çizge dünyasını kullanan çalışmalar da mevcuttur. Çünkü eş ve zıt ilişkilerin kullanımı, problemin çoğunlukla çizge ile modellenmesini beraberinde getirmektedir. Kamps ve arkadaşları [74] tarafından, Hu ve Liu [44]'nin çalışmasında olduğu gibi, sıfatlardan oluşan tohum listesindeki kelimelerin WordNet'teki eş ve zıt terimlere olan uzaklığını temel alan bir çalışma ortaya konulmuştur. Problemi çizge yaklaşımıyla ele alan metodolojileri, oluşan ağdaki tohum düğümler ile polaritesi bilinmeyen düğümlerin arasındaki uzaklığı temel metrik olarak almaktadır. Bu amaçla tohum düğümler ile arasındaki en kısa yolun uzunluğu, yeni terimin duygu skorunun ve yönünün belirlenmesinde kullanılmıştır.

Blair-Goldensohn ve arkadaşları [75] da, restoran ve otel sektörlerindeki müşteri yorumlarına yönelik duygu analizi çalışmaları için ihtiyaç duydukları duygu sözlüğünü, çizge yaklaşımı kullanarak elde etmişlerdir. İlk adımda, diğer çalışmalardan farklı olarak pozitif ve negatif kelimelerin yanında nötr kelimelerin de bulunduğu bir tohum kelime listesi manuel olarak oluşturulmuştur. Ardından çift yönlü ve öznitelikli çizge (*directed attributed graph*), pozitif ve negatif tohum kelimelerini temel alarak WordNet'in eş ve zıt ilişkileri ile genişletilmiştir. Nötr tohum kelimeler, ağdaki yayılımın durma noktaları olarak kullanılmıştır. Özelleştirilen etiket yayılım algoritması ile yeni kelimelerin skoru belirlenmiştir. Bu maksatla pozitif tohum kelimelere +1, negatif tohum kelimelere -1 ve diğer kelimelere de sıfır değeri başlangıçta atanmış ve algoritmanın yayılım sürecinde değerler güncellenmiştir. Belirli bir iterasyon sonucunda

sonlanan algoritma, ağıdaki bütün kelimelere bir skor değeri belirlemiş ve sonuçta duygu sözlüğü elde edilmiştir.

Çizge dünyasına taşınan probleme, etiket yayılım algoritmasının yanında minimum kesit (*mincut*) ve rastgele minimum kesit (*randomized mincut*) algoritmaları ile çözüm arayan çalışmalar da mevcuttur. Minimum kesit, bir çizgeyi iki parçaya en az kenar kullanarak kesilmesi için çalışan bir algoritmadır. Rao ve Ravichandran [76] yarı denetimli ve çizge temelli yaklaşımla duygu sözlüğü geliştirme çalışmalarında, 3 farklı çizge algoritmasının başarımını karşılaştırmışlardır. Diğer çalışmalarda olduğu gibi pozitif ve negatif tohum kelimelerini WordNet kütüphanesinden taramak suretiyle oluşturdukları ağda; etiket yayılım, minimum kesit ve rastgele minimum kesit algoritmalarını uygulamışlar, minimum kesit ve rastgele minimum kesit algoritmalarıyla oluşturulan duygu sözlüğünün daha iyi sonuç verdiğini tespit etmişlerdir.

Çeviri sistematüğinde ise, var olan kapsamlı bir duygu sözlüğünün hedef dile çevrilmesi ile duygu sözlüğü elde edilir. Bu sistematik çoğunlukla duygu sözlüğü olmayan ve dil kaynakları sınırlı olan İngilizce dışındaki diğer dillerde sıklıkla kullanılmaktadır. Hugo ve arkadaşları [77], AFINN [66] İngilizce duygu sözlüğünü Google çeviri (<https://translate.google.com>) ile Norveççe diline çevirmişlerdir. En gelişmiş genel amaçlı İngilizce duygu sözlüğü olan SentiWordNet'i [59], Basile [78] İtalyancaya, Perez-Rosas ve arkadaşları [79] da İspanyolcaya çevirerek hedef dilde duygu sözlükleri oluşturmuşlardır.

Uçan [80] da, tez çalışması kapsamında SentiWordNet [59] duygu sözlüğünü 3 farklı çeviri motoru kullanarak Türkçeye çevirmiştir. 27K boyutundaki bu Türkçe duygu sözlüğü, bu tez kapsamında da kullanılmıştır.

Doğrudan çevirinin yanında daha sistematik metotlar öneren çalışmalar da mevcuttur. Yao ve arkadaşları [81], Çince kelimelerin duygu yönünü İngilizce karşılıklarından faydalananarak tespit etmeyi ve böylece Çince duygu sözlüğü oluşturmayı amaçlamışlardır. Çince 3 değerlendirici tarafından tespit edilen 4120 duygu terimi, 10

farklı Çince-İngilizce sözlük kullanılarak İngilizceye çevrilmiştir. İngilizce kelimelerin duygu yönü bilinen bir kaynaktan elde edilerek eşleştirme yapılmıştır. Bütün diller arasında olduğu gibi, bir terimin diğer dilde birden fazla karşılığı olduğundan sonuçta bir kelime vektörü oluşturulmuştur. Makine öğrenmesi yaklaşımlarının da kullanılmasıyla çalışmanın sonucunda 1916 pozitif ve 2204 negatif Çince kelimedenden oluşan duygu sözlüğü elde edilmiştir.

2.3.3. Derlem Temelli Yaklaşımla Duygu Sözlüğü Oluşturmak

Derlem temelli yaklaşımın ana felsefesi, elde mevcut bulunan büyük etiketlenmiş (pozitif-negatif) bir derlemi kullanarak duygu sözlüğünün oluşturulmasıdır. Sözlük temelli yaklaşımda olduğu gibi, polaritelenmiş kelimelerden oluşan tohum listesi bu yaklaşımda da kullanılır. Polaritelenmiş derlemde, tohum kelimelerle ortak örüntüler gösteren terimler tespit edilerek sözlüğe eklenir. Bu süreçteki metodoloji iteratif olarak devam ettirilerek sözlüğün kapsamı ve hacmi genişletilir. Derlem temelli yaklaşımın en büyük güçlüğü, birçok dilde büyük hacimli, kapsamlı ve etiketlenmiş bir derlemin bulunmamasıdır. Genellikle var olan etiketlenmiş derlemler de, belirli bir alana özgüdür. Bu sebeple de, detaylarına Bölüm 2.4.2.'de yer verilen özel amaçlı duygu sözlüklerinin oluşturulmasında çoğunlukla derlem temelli yaklaşım kullanılmaktadır.

Hatzivassiloglou ve McKeown [52] çalışmalarında, duygu polariteleri bilinen ve sıfat türündeki kelimelerden oluşan tohum listesini kullanarak, mevcut derlemdeki diğer sıfat türlerini bulmaya ve duygu polaritelerini tespit etmeye çalışmışlardır. Temel felsefe, metinlerdeki bağlaçların varlığının göz önünde bulundurulmasına dayanmaktadır. Örneğin “AND” bağlacı ile bağlı iki sıfatın aynı duygu polaritesine sahip olacağından hareketle, hem yeni sıfatları elde etmişler hem de duygu polaritelerini belirlemişlerdir. Aynı şekilde “OR”, “BUT”, “EITHER-OR”, “NEITHER-NOR” bağlaçları için de benzer kurallar geliştirmişler ve bu yapıya duygusal tutarlılık (*sentiment consistency*) ismini vermişlerdir. Çizge dünyasında modellenen veriler üzerinde, kümeleme algoritmaları da kullanılarak iki sınıflı duygu sözlüğü oluşturulmuştur.

Kanayama ve Nasukawa [82]'da bağlaçları benzer şekilde kullanan yaklaşımı genişletmişlerdir. Bağlaçların sadece kelimeleri değil, cümleleri de benzer şekilde bağladığına dikkat çekmişler ve komşu cümlelerin bağlacın yönüne göre duygu değerine sahip olduğunu ortaya koymuşlardır. Önerdikleri sistematik, Japonca dilinde çalışılmış ve özel amaçlı duygu sözlüğü geliştirilmiştir.

Literatürde istatistiksel yaklaşımları kullanan çalışmalar da mevcuttur. Bu çalışmaların ana karakteristiği, bir tohum listesindeki terimlerle, derlemde kuralları belirlenen örüntülerin eşleştirilmesi şeklindedir. Bu doğrultuda Remus ve arkadaşları [83] tarafından, ürünlere yapılan müşteri yorumları veri setini kullanılarak duygu sözlüğü oluşturulmuştur. Başlangıçta tohum listesi ile 5100 farklı metinden oluşan pozitif yorumlar ve 5100 adet negatif yorum veri setleri bulunmaktadır. Çalışma kapsamında, PMI hesaplama tekniği ile Almanca duygu sözlüğü oluşturulmuştur.

Derlemin bulunduğu fakat etiketsiz olduğu durumda Kaity ve Balakrishnan [84] tarafından dilden bağımsız olarak duygu sözlüğü oluşturmaya yönelik bir çözüm önerisi sunulmuştur. Çalışma Arapça dili için uygulanmıştır. İhtiyaç duyulan veri setleri, İngilizce dilinde duygu yönü belirli olan bir tohum kelime listesi ile hedef dildeki etiketsiz derlemin varlığıdır. Etiketsiz derlem Facebook'tan elde edilmiştir. Sürecin ilk adımında, İngilizce tohum kelime listesi hedef dile çevrilmekte, derlemde aday kelimeler listesi elde edilmektedir. Ardından aday kelimeler listesi ve tohum kelimeler derlemde taranarak önerilen matematiksel teknikle aday kelimelerin duygu skoru çıkarılmaktadır. Pozitif ve negatif kelimelerin tespitine ilaveten, duygu skoru belirli sayısal aralıkta kalan kelimeler de nötr olarak sınıflandırılmış ve sonuçta Arapça duygu sözlüğü oluşturulmuştur.

2.4. Duygu Sözlüklerinin Kapsamı

Kapsam bakımından ele alındığında, duygu sözlükleri genel amaçlı (*domain-independent*) ve özel amaçlı (*domain-dependent*) olmak üzere iki kategoriye ayrılabilir.

2.4.1. Genel Amaçlı Duygu Sözlüğü

Genel amaçlı duygu sözlüklerinin kapsamı daha geniş olduğundan duygu analizi de yüksek doğrulukla yapılabilmektedir [15]. Bunun yanında içeriğin bağlamı ve kategorisi (eğitim, sağlık, spor vb.) göz ardı edildiğinden, duygu analizinde hassasiyet kaybına neden olduğu gerekçesi ile eleştiriler de söz konusudur. Fakat birçok çalışma için de temel bir kütüphanedir. Bu nedenle de her dil için böyle bir kaynağın varlığı oldukça önemsenmektedir.

General Inquirer [85], manuel yaklaşımla geliştirilmiş İngilizcedeki ilk duygu sözlüklerindendir. Terimleri kategorik bir yapıda tutan General Inquirer yaklaşık 11K terime sahiptir. Araştırmacılar tarafından tohum kelimelerin belirlenmesinde veya önerilen yöntemlerin sonuçlarının karşılaştırılmasında kullanılan temel standart sözlüklerden birisidir.

Genel amaçlı duygu sözlüğünün geliştirilmesinde, İngilizce dil kaynaklarından WordNet [86] ön plana çıkmaktadır. WordNet, bir duygu sözlüğü olmamasına rağmen duygu sözlüğü oluşturulmasında sıklıkla kullanılan bir kütüphanedir. Uzun bir zamana yayılan gelişim süreci, WordNet'i İngilizce için en temel dil kütüphanelerinden birisi haline getirmiştir. WordNet, adına "synset" denilen terimlerin anlamdaş kümeleri ile terimin sözlük anlamının örnek cümlelerle beraber yer aldığı hiyerarşik yapıya sahip bir dil kütüphanesidir. Bu hiyerarşik yapı ontoloji ağacı karakteristiğine sahiptir. Bir terimin, üst sınıf ve alt sınıfları ile terimin parçası olduğu bütün ile terimi oluşturan alt parçalar, ağaç sistematüğinde yapılandırılarak bütünsel yapıda bir WordNet ontoloji ağacı geliştirilmiştir. WordNet'te 117K anlamdaş kümesi mevcuttur [80]. Diğer dillerde de İngilizcedeki gibi gelişmiş olmasa da WordNet kütüphaneleri geliştirilmiştir. EuroWordNet [87] ve BalkaNet [88] gibi kütüphaneler, WordNet ile aynı veri yapısında ve bünyelerindeki terimlerin ILI (*Inter-Lingual-Index*) indeksi de WordNet ile eşleşecek şekilde geliştirilmişlerdir. BalkaNet projesi kapsamında geliştirilen Türkçe WordNet [89]'te aynı sistematüğe sahiptir.

Kapsamlı ve tutarlı içeriğe sahip olan İngilizce genel amaçlı duygu sözlüklerinin başında SentiWordNet [59, 60, 90] gelmektedir. SentiWordNet, WordNet kütüphanesi üzerinde derlem temelli yaklaşımla geliştirilmiş bir duygu sözlüğüdür. Süreç içerisinde geliştirilmesine devam edilen SentiWordNet'in bünyesindeki her bir eş anlamlılık kümesi, 0 ve 1 arasında değişen pozitif, negatif ve nötr değerlere sahip olup, bu değerlerin toplamı da 1'e eşittir. SentiWordNet'te toplam 117K anlamdaş kelimeler kümesi ve bu kümeler açıldığında da toplamda 206K kapasiteli farklı anlama sahip terim mevcuttur [91]. Bu terim zenginliğinin yanında, bir terimin WordNet'te farklı anlamdaş kümelerinde bulunabilmesi, bu terimin SentiWordNet'te de farklı duygu skorlarına ve yönüne sahip olabilmesini beraberinde getirmektedir. Bu da SentiWordNet'i güçlü kılan bir diğer özelliktir.

SentiWordNet dışında, WordNet'in sunduğu veriler ve veri yapısından faydalanarak geliştirilmiş ve kabul görmüş İngilizce duygu sözlükleri de bulunmaktadır. Bunlardan ön plana çıkanlar arasında WordNet-Affect [92], SenticNet [93], SentiSense [94], MicroWNOp [95], MPQA [96] ve Opinion Lexicon [44] sayılabilir.

2.4.2. Özel Amaçlı Duygu Sözlüğü

Genel amaçlı duygu sözlüklerine yöneltilen başlıca eleştiriler, kapsamının ve bu sözlüklerle yapılan analizin doğruluğunun eksikliğidir [97]. Kapsam eksikliği ile alana özel terimlerin sözlükte bulunmamasına vurgu yapılmaktadır. Diğer eleştiri de terimin bağlamına bakılmaksızın tek ve sabit bir değerle hesaplama sürecine dahil edilmesidir. Bu noksanlıklar, araştırmacıları alana özgü duygu sözlüğü geliştirilmesine yöneltmiştir.

Thelwall ve arkadaşları [51], çalışmalarında sosyal medya paylaşım içeriklerinin ve özellikle de kısa metinlerin duygu analizine odaklanmışlardır. Bu doğrultuda SentiStrength isimli algoritmik yaklaşımı önermişlerdir. Çalışma kapsamında ihtiyaç duyulan duygu sözlüğü, terim skorları -5 ile +5 arasında bir tamsayıyla etiketlenerek oluşturulmuştur. Ardından sosyal medya içerikleri kullanılarak makine öğrenmesi yaklaşımlarıyla sözlükteki terimlerin değerleri içeriğe göre güncellenmiştir. Geliştirilen SentiStrength bir dil aracı olmasının yanında, 2500 terimden oluşan duygu sözlüğü,

kuvvetlendirici ve zayıflatıcılar, deyimler, olumsuzluk ifadeleri ve duygu simgeleri (*emoji*) ile geniş bir dil kütüphanesine de sahiptir. Bu özelliklerinden dolayı İngilizce dışındaki diğer dillerde, SentiStrength dil kütüphanelerinin hedef dile çevrilerek kullanıldığı çalışmalar da mevcuttur.

Almatarnah ve Gamallo [98], derlem temelli ve makine öğrenmesi yaklaşımı ile manuel olarak etiketlenmiş film yorumları veri setini kullanarak, film yorumları alanına özel bir duygu sözlüğü geliştirmişlerdir. Genel amaçlı duygu sözlükleri ile performans karşılaştırmasında, geliştirilen duygu sözlüğünün daha başarılı olduğu belirtilmiştir.

Oliveira ve arkadaşları [99], borsa dünyasındaki mikroblog mesaj içeriklerinden istatistiksel yaklaşımlarla alana özel duygu sözlüğü geliştirmişlerdir. Veri kaynağı olarak StockTwits platformundan elde ettikleri 350K iletiyi ve Twitter'dan ilgili paylaşımları kullanmışlardır. PMI [57] ve iki yenilikçi istatistiksel yaklaşımla skorlarını tespit ettikleri terimlerden oluşan sözlüğün performansını, aralarında SentiWordNet [60], General Inquirer [85] ve MPQA [96] gibi bilinen genel amaçlı duygu sözlüklerinin performansı ile kıyaslamışlar ve geliştirdikleri borsa alanına özel duygu sözlüğünün daha yüksek performans gösterdiğini ifade etmişlerdir.

Asghar ve arkadaşları [100], araştırmalarında sağlık alanına özel duygu sözlüğü geliştirmeye odaklanmışlar, ilaç ve tedavi konulu paylaşımlardaki terimlerin ağırlıklı olarak genel amaçlı duygu sözlüklerinin kapsamı dışında olduğunu tespit etmişlerdir. Çalışmada kapsamında derlem temelli, PMI [57] istatistiksel tekniği ve SentiWordNet [60] kütüphanesinden faydalanılarak hibrit bir yaklaşım sergilenmiştir. Sonuçta, SentiHealth isimli ve 1520 terimli alana özgü duygu sözlüğü oluşturulmuş ve test edilmiştir. Yapılan testlerde başarımın SentiHealth ile daha yüksek olduğu vurgulanmıştır.

Sıfırdan alana özgü duygu sözlüğü oluşturmak yerine, var olan genel amaçlı duygu sözlüğü terimlerinin, polarite ve skorlarının güncellenerek alana duyarlı hale getirilmesine dayanan çalışmalar da mevcuttur. Vishnu ve arkadaşları [101]

çalışmalarında, SentiWordNet [60] ve 5 farklı kategorideki etiketlenmiş yorum derleminden faydalanarak alana özel duygu sözlüğü geliştirme yaklaşımını önermişlerdir. Derlem temelli bu yaklaşım; kitap, kamera, DVD, sağlık ve müzik alanları üzerinde uygulanmış ve performans değerlendirmesinde başta sağlık olmak üzere bütün alanlarda daha iyi bir sonuç elde edilmiştir.

2.5. Türkçe Duygu Analizi ve Duygu Sözlüğü

2.5.1. Türkçe Dilinin Zorlukları

Türkçe, Ural-Altay dil grubunda yer alan sondan eklemeli bir dildir. Biçimbilimsel olarak oldukça zengin olan Türkçede her kelime, bir kök ile bu kökün sonuna eklenen eklerden oluşmaktadır. Yapım ekleri ve çekim ekleri olmak üzere iki temel ek yapısı mevcuttur. Yapım ekleri sonlarına geldiği kelimenin anlamını ve türünü değiştirirken, çekim ekleri ise anlamını değiştirmemekte fakat türünü değiştirebilmektedir [102]. Çizelge 2.1’de kök ve eklere yönelik bir örnek görülmektedir.

Çizelge 2.1 Türkçe kök ve eklere yönelik bir örnek.

garip (sıfat)	garip-lik (isim)	garip-se-mek (fiil)
------------------	---------------------	------------------------

Bu örnekte, “garip” kökünden türeyen kelimeler, kelime türü bilgileriyle birlikte görülmektedir. Türemiş bir kelimenin kök formunun tespit edilmesi de ayrı bir problem alanıdır. Yukarıdaki örnekten hareketle “gariplik” kelimesinin kök formunun “gar” olarak değil “garip” olarak doğru bulunabilmesi, Türkçedeki zenginliği ve biçimbilimsel analizin önemini de göstermektedir.

Türkçede olumsuzluk ve zıt anlamlılık durumlarında da zenginlik söz konusudur. Çizelge 2.2’de de görüldüğü gibi, son eklerle terimler negatif hale getirilebildiği gibi, “değil” kelimesi ile de negatiflik söz konusudur. Bu örnekteki “tehlikeli” ve “tehlikesiz” terimleri olumsuzluk ilişkisinin yanında aynı zamanda zıt anlam ilişkisine de sahiptir.

Çizelge 2.2 Türkçe olumsuzluk durumlarına ait bir örnek.

tehlike (isim)	tehlike- li (sıfat) ➡	tehlikesiz değil
	tehlike- siz (sıfat) ➡	tehlikeli değil

Çizelge 2.3'teki örnek incelendiğinde ise, “büyümek” kelimesinin zıttı olan “büyümek” ile “küçülmek” kelimelerinin kendi aralarında eş anlamlı olması beklenebilir. Ancak “büyümek” ile “küçülmek” tam olarak eş anlamlı değildir.

Çizelge 2.3 Türkçede zıt anlamlılık durumlarına ait bir örnek.

	Sözdizimsel Zıt	Anlamsal Zıt
büyümek	büyümek	küçülmek

Sunulan bu birkaç örnek bile, istisnai durumların çok olduğu Türkçede doğal dil çalışmalarının zorluklarına dair ipuçları vermektedir.

Türkçe metinlerin duygu analizinde Bölüm 2.1.5’de sayılan zorluklar mevcut olmakla beraber, kendine has özelliklerinden dolayı ilave hususlar da vardır [25, 65, 103-105]. Bunlar:

- Eklemeli bir dil oluşu nedeniyle Türkçe doğal dil çalışmalarının ilk safhasında dilbilimsel analizin yapılması adeta bir zorunluluktur. İngilizce ve İngilizceye daha yakın olan dillerde böyle bir yapının olmaması, doğal dil çalışmalarında bu dilleri daha avantajlı hale getirmektedir.
- Her ne kadar Türkçe cümlelerin gramer ifadesinde doğal öge sıralaması Özne – Nesne – Yüklem şeklinde olsa da, çoğunlukla öğeler cümle içerisinde serbest dizilişe sahiptir. Bu karakteristiği, duygu analizinde zorluklara sebep olmakla beraber, özellikle hedef tabanlı duygu analizinde ciddi güçlükler neden olmaktadır.

- Özellikle sosyal paylaşım platformlarında paylaşılan içeriklerde, Türkçede bulunan ancak ASCII kod kümesinde olmayan harflerin yerine İngilizce eşdeğerlerinin kullanılması sıklıkla karşılaşılan bir durumdur ve terimlerin gerçek karşılıklarının tespitinde zorluklara sebep olmaktadır.
- Türkçede olumsuz ifadeleri oluşturmanın farklı yolları vardır. Örneğin fiiller “-me” ve “-ma” ekleri ile olumsuzlaştırılırken, isim türündekiler ise “-sız”, “-siz”, “-suz” ve “-süz” son ekleri ile olumsuz sıfatlar olmaktadır. Ayrıca “değil” ile “yok” kelimeleri de olumsuz ifadeler oluşturmak için kullanılmaktadır.
- Paralel derlem, makine öğrenmesiyle duygu analizi çalışmalarında performans ölçümündeki temel gereksinimlerdenidir. İngilizce-Çince ve İngilizce-Arapça gibi bir paralel derlem, İngilizce-Türkçe için mevcut değildir.

2.5.2. Türkçe Duygu Analizi ve Duygu Sözlüğü Geliştirme Çalışmaları

Türkçe dil kaynaklarının azlığı, var olanların kapsam ve içerik olarak sınırlılığı ile bazı kaynakların da açık erişime kapalı olmasından dolayı, Türkçe duygu analizi alanında çalışan araştırmacıların ağırlıklı olarak iki temel yaklaşımdan birisini sergilediği görülmektedir. Bunlar:

- Türkçe içeriklerin önce İngilizceye çevrilmesi ve ardından gelişmiş İngilizce dil kaynakları ile duygu analizinin yapılması.
- İngilizce dil kaynaklarının Türkçeye çevrilmesi ve ardından Türkçe içeriklerin duygu analizinde kullanılması.

Eroğul [106] yüksek lisans tezi çalışmasında, Türkçe duygu analizi için makine öğrenmesi temelli bir yaklaşım önermiştir. Türkçeye özgü dilbilimsel analizleri, olumsuzluk durumlarını, kelime türü bilgilerini, n-gram metotlarıyla da göz önüne alarak öznitelikler belirlemiş ve film yorumları üzerinde uygulamıştır. Çalışma sonucunda %85 düzeyinde bir başarının elde edildiğini, dilbilimsel özelliklerin, olumsuzluk durumlarının, n-gram yaklaşımının performans artışını desteklediğini, POS kapsamında “isim” türünün katkısının yüksek olduğunu, kelimelerin kök formlarının kullanılmasının ise sadece %1 gibi bir oranda düşüşe neden olduğunu vurgulamıştır.

Vural ve arkadaşları [107], Eroğul [106]'nın film yorumları derlemi üzerinde sözlük temelli duygu analizi yapmaya odaklanmışlar ve Bölüm 2.4.2'de de yer verilen SentiStrength [51] aracını kullanarak bir sistem geliştirmişlerdir. SentiStrength aracının İngilizce olan dil kaynağı kütüphanesi manuel olarak Türkçeye çevrilmiş ve eksik olan kelimelerin ilavesi yapılmıştır. Girilen Türkçe içerik üzerinde sırasıyla; cümlelerine ayırma, Türkçe karakter problemlerini ele alma, sözdizimsel analiz ile kelimelerin kök formlarının elde edilmesi, olumsuzluk eklerinin değerlendirilmesi ve son aşamada SentiStrength aracıyla duygu analizinin yapılması süreçleri uygulanmıştır. Film yorumları derlemi üzerinde yapılan performans testinde %76 başarı elde edilmiştir. Duygu sözlüğünün kapasitesi hakkında bilgi verilmemiştir.

Türkmenoğlu ve Tantuğ [68] çalışmalarında, sözlük temelli yaklaşım ile makine öğrenmesi yaklaşım performanslarını Twitter ve film yorumları derlemleri üzerinde karşılaştırmışlardır. Ön işlem aşamasında, Türkçe karakter problemlerinin giderilmesi, sözdizimsel analiz ve çok terimli ifadelerin ele alınması süreçleri gerçekleştirilmiştir. Sözlük temelli yaklaşım çerçevesinde, Vural ve arkadaşlarının [107] çalışmasında olduğu gibi SentiStrength [51] aracının İngilizce dil kütüphanesini Türkçeye manuel olarak çevirerek elde ettikleri 2547 terimli Türkçe duygu sözlüğünü, olumsuzluk ekleri ve kelime türlerini de dikkate alarak oluşturmuşlardır. Makine öğrenmesi temelli yaklaşım çerçevesinde ise, öznitelik vektörünü TF-IDF (*Term Frequency-Inverse Document Frequency*) modeli ile n-gram ($n \leq 2$) boyutlarında oluşturmuşlar ve SVM, NB ve Karar Ağacı (*J48*) sınıflandırma algoritmalarını kullanmışlardır. Çalışmanın sonucunda, makine öğrenmesi yaklaşımı ile daha iyi sonuç elde ettiklerini, ancak problem alanına ve analiz edilecek derleme göre birçok durumda da sözlük temelli yaklaşımın tercih edilebileceğini vurgulamışlardır.

Aytekin [108], Türkçe blog sayfalarındaki müşteri yorumlarını pozitif ve negatif olarak sınıflandıran yarı denetimli bir duygu analizi modeli önermiştir. Çalışma kapsamında 2178 sıfat ve zarf türünde kelimedenden oluşan İngilizce duygu terimleri kaynağı Türkçeye çevrilerek, toplamda 4744 Türkçe duygu teriminden oluşan duygu sözlüğü oluşturulmuştur. İngilizce kaynaktaki [109] terimler, sahip oldukları duygu tonlarına

göre renk skalası ile etiketlenmiş bir veri yapısındadır. Çalışmada, NB algoritması ve renk skalası verileri ile duygu sözlüğündeki terimler pozitif ve negatif olarak sınıflandırılmıştır. Duygu sözlüğü müşteri yorumları derlemi üzerinde test edilmiş ve %73 başarı performansı göstermiştir.

Kaya ve arkadaşları [110], politika konulu Türkçe gazete köşe yazıları üzerinde duygu analizini makine öğrenmesi temeli araştırmışlardır. Çalışma kapsamında dört farklı makine öğrenmesi algoritmasının performansı karşılaştırılmış, gereksinim duyulan duygu sözlüğü, güçlü duygu belirten ve manuel olarak seçilmiş 197 pozitif ve 300 negatif terimden oluşturulmuştur. Bu duygu sözlüğü ile yapılan sınıflandırma başarımı, makine öğrenmesi performanslarının kıyaslama noktası olarak kullanılmıştır. Makine öğrenmesi algoritmik süreçleri için geliştirilen farklı öznitelik vektörleri; karakter temelli n-gram dil modeli, SVM, ME ve NB sınıflandırma algoritmalarında kullanılmıştır. Politik köşe yazılarından oluşturdukları derlem üzerinde yapılan performans testi sonuçlarında; makine öğrenmesi yaklaşımların sözlük temelli yaklaşımdan daha iyi sonuç verdiği, karakter temelli n=8 dil modeli ve ME algoritmalarının NB ve SVM algoritmalarından daha iyi performans gösterdiği, haberler üzerinde duygu analizinin yorumlar üzerinde yapılagelen analizlerden daha zor olduğu tespitleri yapılmıştır.

Gökhan ve arkadaşları [111], Türkçe tweet içeriklerinin duygu sözlüğü temelli analizi için bir sistematik önermişlerdir. Çalışmada 1181 terim ve 398 deyim ve atasözünden oluşan bir duygu sözlüğü kullanılmıştır. Türkçe dilinin özel durumlarına vurgu yapılmış ve bu doğrultuda kelime, kelime grubu ve deyim/atasözleri seviyelerini de içeren ve olumsuzluk eklerini de dikkate alan kurallar geliştirilerek analiz sürecinde uygulanmıştır. Önerdikleri yaklaşımın performans testi sonuçlarına göre, kelime grubu seviyesinin dahil edildiği analizin daha iyi olduğunu ve sözlük temelli analizin Türkçe için esnek özelliklerinden dolayı çok daha iyi bir yaklaşım olduğunu vurgulamışlardır.

Türkçe genel amaçlı duygu sözlüğü oluşturulması konusunda, bizim tez hedefimize en yakın çalışma Dehkharghani [112] tarafından yapılan SentiTurkNet'tir. Manuel oluşturulan bir tohum listesinin, Türkçe WordNet [113] kullanılarak genişletilmesi

prensibine dayanmaktadır. Türkçe WordNet, 2001-2004 yılları arasında BalkaNet³ projesi kapsamında Sabancı Üniversitesi koordinatörlüğünde geliştirilen, 14K anlamdaş küme ve 20K terimden oluşan, açık erişime kapalı olan ilk ve tek Türkçe WordNet kaynağıdır. Ancak 117K anlamdaş terim kapasiteli İngilizce WordNet'e kıyasla oldukça düşük kapasiteye sahiptir. Geçen zaman içerisinde gelişimine devam ettirildiğine dair de literatürde herhangi bir çalışmaya rastlanılmamıştır. SentiTurkNet'in gelişimi sürecinin ilk aşamasında 3 farklı İngilizce dil kütüphanesinden faydalanılmıştır. Bunlar WordNet [86] , SentiWordNet [60] ve SenticNet [93] olup, SenticNet Türkçeye çevrilmiştir. Türkçe ve İngilizce WordNet kütüphaneleri, veri yapılarındaki ILI alanları üzerinden birebir eşleştirilmiş ve böylece Türkçe WordNet'teki anlamdaş kümelerin duygu skorları İngilizce SentiWordNet'ten elde edilmiştir. Elde edilen skorlar, çalışmanın bir sonraki aşaması olan makine öğrenmesi sürecinde öznitelikleri oluşturmuştur. Bu oluşturulan kaynağa ilaveten, iki kaynak daha geliştirilmiştir. Bunlar, manuel olarak oluşturulan tohum kelime listesinin eş ve zıt anlam yaklaşımıyla genişletilerek 1000 pozitif ve 1000 negatif kelime hacimli kütüphane ile PMI tekniği ile elde edilen ilave kaynaktır. Tüm bu verilerle yapılan sınıflandırma algoritmasının güvenilirlik skorları, anlamdaş kümelerin skorları olarak atanmıştır. WordNet ve WordNet'ten üretilen SentiWordNet gibi SentiTurkNet sözlüğünde de bir terim farklı anlamdaş kümelerinde bulunabilir ve doğal olarak da aynı terimin farklı skorları sözlükte yer alabilmektedir. Duygu analizinde, terimin hangi skorunun hesaplama sürecine dahil edileceğinin tespitinde, terimin bağlamının tespiti başarımlı performansını oldukça iyileştirecek bir yaklaşımdır. Bu amaçla WordNet kütüphanelerinin sözlük anlamlarını barındıran *Gloss* alanından faydalanılmaktadır. Ancak, Türkçe WordNet'teki terimlerin birçoğunda bu alanların bulunmadığı Gizem ve Yanıkoğlu [25] tarafından belirtilmiştir. Bu durumda, anlamdaş küme yapısının bağlam odaklı duygu analizine getirdiği fayda, SentiTurkNet için maalesef kaybolmaktadır.

WordNet kütüphanesinden faydalanarak duygu sözlüğü oluşturulmasına yönelik bir çalışma da Özsert ve Özgür [114] tarafından yapılmıştır. Dillere ait WordNet'lerin birbirleriyle uyumlu olan veri yapısından faydalanılan çalışmada, Türkçe ve İngilizce WordNet anlamdaş kümeleri ve ilişkileri çizge ortamında ele alınmıştır. Ardından 1730 pozitif ve 2085 negatif İngilizce tohum listesi manuel olarak belirlenmiş, bu liste ve

³ <http://www.dblab.upatras.gr/balkanet/>

WordNet veri yapısı kullanılarak 1398 pozitif ve 1414 negatif Türkçe tohum listesi yarı otomatik olarak elde edilmiştir. Rastgele yürüyüş tekniğinin önerilen modeli ile de bir terimin pozitif ve negatif tohum kümelerine uzaklığı bulunarak polaritesi tespit edilmiştir.

Uçan [80] yüksek lisans tezinde, çeviri sistematigiyle Türkçe duygu sözlüğü oluşturulmasını ve duygu analizinde kullanılmasını çalışmıştır. Temel varsayım duygu ifadelerinin evrenselliğine dayanmaktadır. En kapsamlı ve gelişmiş genel amaçlı duygu sözlüğü olan İngilizce SentiWordNet [60] kaynak olarak seçilmiş ve Türkçeye üç farklı sözlük kullanılarak önerilen yöntemlerle otomatik olarak çevrilmiştir. Sonuçta 27K boyutunda genel amaçlı bir Türkçe duygu sözlüğü oluşturulmuştur. Sözlüğün performansı otel ve film yorumları derlemleri üzerinde ölçülmüş ve sonuçta ortalama %80 başarı elde edilmiştir. Bu sonucun makine öğrenmesi yaklaşımlarıyla aynı düzeyde olduğu vurgulanmıştır. Bu çalışmada elde edilen duygu sözlüğü, tez çalışmamızda SWNetTR olarak isimlendirilerek kullanılmıştır.

2.6. Ana Akım Medya ve Haber

İletişim, kendisine çok sayıda anlam ve açıklama yüklenen bir kavram olmakla beraber [115], sosyal yapı içerisindeki bireylerin duygu, düşünce ve bilgileri birbirlerine aktarma sürecidir ve toplumun temelini oluşturan bir sistem olarak karşımıza çıkmaktadır. Günümüzde iletişim ağırlıklı olarak kitlesel iletişim araçları ile gerçekleştirilmekte olup, internet teknolojilerinin neredeyse tüm imkanları iletişimin etkinliğine ve hızına kanalize olmuş vaziyettedir. Çevrimiçi haber medyalarından, sosyal medya ortamlarına kadar geniş bir yelpazede sunulan içeriklerle hedeflenen, okuyucu/takipçi kitlesini bilgilendirmek, haberdar etmek ve etkilemektir. Gönenli ve Hürmeriç[116], çevrimiçi platformların bireylerin çevresiyle iletişime geçmesini sağlayan, aynı zamanda işbirliği ve ortaklık oluşturan, insanları etkileme gücünü arttıran ve toplumda haberlerin hızlı bir biçimde yayılmasını kolaylaştıran, günümüz insanı için olanak ve ayrıcalık alanı şeklinde tarif etmektedir.

Kitlesel iletiřimde n planda yer alan aralardan birisi de gazetelerdir. Teknolojik geliřmelerin toplumda kabul grmesi, geleneksel gazeteciliğın internet gazeteciliğine evrilmesini de beraberinde getirmiřtir. Bylece hem inovasyon hem de eriřilen kitlenin geniřletilmesi mmkn olabirmiřtir [117]. Deuze [118], internet gazeteciliğini ana akım, endeks/kategori, eleřtiri/yorum ve paylařım/tartıřma olmak zere drt kategoriye ayırmaktadır. Ana akım ile tarif edilen medya, geleneksel medya kuruluřlarının evrimii haber siteleridir.

Trkiye’de kitlesel iletiřim kategorisindeki internet gazeteciliğının geliřimi 2000 yılına dayanmaktadır. Bu dnemde, ana akım medya ile bağımsız gazetecilerin kurdukları haber siteleri grlmeye bařlanmıřtır. Ekonomik olarak geniř bir sahaya sahip olmayan bir alanda rekabet bařlamıř, ana akım medyalar ekonomik gleri ile bu rekabette ayakta kalmıřlar ve daha geniř okuyucu kitlesince kabul grmřlerdir [117]. Bu kabulde, ana akım medyanın toplumun geneline hitap eden, gndemi belirleyici ge sahip ve etik ilkelere azami zen gsteren medya olarak algılanmasının da katkısı byktr.

Haber, kitlesel iletiřimde topluma sunulan en yaygın ve nemli enformasyon olup, okuyucular tarafından ok hızlı tketilen bir medya ieriğidir. Haberi diğeri medya ieriklerinden belirgin řekilde ayıran bu zelliğı, onun toplum iinde ok gl bir etki alanına sahip olduğunun da gstergesidir. Ancak haberin etki řiddetinin belirlenmesi ok daha zor bir alandır. Yurdakul’a [119] gre, haberin topluma ulařtırılması ve herkese uygun dzeye getirilmesinde bir l var olabilir ama, herkesi aynı dzeyde etkileyebilecek bir lden bahsetmek olası değildir.

Bu noktada haber hakkında literatrde zerinde uzlařılmıř bir tanım mevcut değildir. Altun[120] haberi; “bilgidir”, “ğrendiğimize dir”, “okumadan, duymadan ya da izlemeden nce bilmediğimize dir”, “olayların bilmeyenlere aktarılmasıdır”, “bir olayın raporudur”, “evrede olup biten her řeydir ve bugnn olayıdır” nitelikleriyle tanımlamaktadır. Haber iin farklı tanımlamalar olsa da, genel anlamda haberde bulunması gereken zellikler konusunda oğulcu yaklařımda bir uzlařmadan sz edilebilir. Gz’e[121] gre oğulcu yaklařımda ne ıkan zelliklerden birincisi,

haberlin arpıtılmadan, olduĐu gibi ve dűřűncelerden baĐımsız verilmesi gerektiĐidir. DiĐeri ise haberde tarafsızlıĐın ve dengeliliĐin korunmasıdır.

2.7. Haberde Duygu

Bir habere, haberin editűrű perspektifinden bakıldıĐında bu haber nűtr kabul edilebilir. űnkű toplumsal sorumluluk gereĐi, haber yansız olarak aktarılmalıdır. Ancak habere konu olan olay/ierik perspektifinden bakıldıĐında, haber duygu barındırmaktadır. Bir bařka ifadeyle haberin okuyucuda yarattıĐı bir etki sűz konusudur. Namrata ve arkadaşları [122] da, bir haberin iyi veya kűtű olabileceĐini, ancak ok nadiren nűtr olabileceĐini ifade etmektedir.

Haber merkezleri, muhabirler ve editűrler bir haberi műmkűn olabilecek en objektif řekilde verseler dahi, o haberin okuyucuda bıraktıĐı etki/etkiler sűz konusudur ve bu hipotez bizi haber metninde duygunun var olduĐu sonucuna gűtűrmektedir. Bu doĐrultuda haber metinleri, duygu analizine aday verilerdir.

Arařtırmacılar da haberlerdeki duygunun keřfine duyarsız kalamamıřlar, bu tez kapsamında olduĐu gibi haber metinlerini gerek kaynak oluřturmak iin, gerekse de geliřtirdikleri űzűm űnerilerinin uygulama alanı olarak kullanmıřlardır. Veri kaynaĐı olarak haber metinlerinin tercih edilme nedenleri genel olarak řu řekilde űzetlenebilir:

- Veriye eriřim maliyeti dűřűktűr.
- Genel eriřime aıktır ve telif hakkı problemi yoktur.
- Editűrler tarafından tűm okuyucu kitlesi gűz űnűne alındıĐından gramere, yazım kurallarına uygudur ve sade bir dile sahiptir. Gűncel dilin terimlerini barındırır.
- Konu yelpazesi ekonomiden spora, saĐlıktan politikaya, magazinden eĐitime geniř bir eřitlilik gűsterir.
- Haber olmaya hak kazanmıř olayların oĐunlukla pozitif/negatif yűnű mevcuttur.

- Okuyucunun ilgisini çekmek için özellikle haber başlıklarında etki gücü yüksek terimler kullanılır.
- Duygu analizi ile dünyanın ilgilenilen bölgesindeki sosyal, politik, ekonomik iklimi öğrenmek mümkün olmakta, karar vericilerin ve politika belirleyicilerin küresel gündemden haber olmalarına olanak sunulmaktadır.
- Haberlere yapılan kullanıcı yorumlarıyla, haberin yaygın etkisi ve toplum üzerinde yarattığı etkiyi gözlemlemek mümkün olabilmektedir.

Literatürde veri kaynağı olarak haber metnlerinin şu çalışma alanlarında kullanıldığı görmek mümkündür:

- Haber başlıklarına yönelik duygu analizi çalışmaları,
- Haber içerikleri üzerinde duygu analizi çalışmaları,
- Haberlere yapılan kullanıcı yorumları üzerinde duygu analizi çalışmaları,
- Haber metinlerinden duygu sözlüğü oluşturma çalışmaları.

2.7.1. Haber Başlıklarına Yönelik Duygu Analizi Çalışmaları

Rieis ve arkadaşları [123], haber başlıklarındaki duygu ile haberin popüleritesi arasındaki ilişkiyi araştırmışlardır. 4 farklı ana akım medyadan derlenen toplam 69907 haber başlığı üzerinde negatif, nötr ve pozitif olmak üzere makine öğrenmesi yaklaşımlarıyla üçlü duygu sınıflandırması yapılmıştır. Haber başlığındaki duygunun hem pozitif hem de negatif yönde yüksek olması ile haberin popüleritesi arasında doğrusal bir ilişkinin varlığı ortaya konulmuştur. Çalışmada ayrıca, internet gazete haberlerinin Twitter mecrasındaki yayılım hızına ve haberin kalıcılığına da odaklanılmış ve sonuçta negatif haberlerin sosyal medyada çok daha hızlı yayıldığı, fakat pozitif haberlere göre daha kısa ömürlü olduğunu tespitlerine yer verilmiştir.

Strapparava ve Mihalcea [124], Semeval-2007 Task 14 takımlarının haber başlığı duygu sınıflandırmasına yönelik çalışmalarına yer vermişlerdir. Haber başlıklarının editörlerce özenle hazırlanan kısım olduğu, seçilen kısa ifadelerin çarpıcı etkiye ve okuma isteği

uyandırma gücüne sahip olması gerektiği tespitleri yapılarak, katılımcılardan haber başlıklarını duygu analizi ve duygusal durum kategorilerinde sınıflandırmaları istenmiştir. Duygu analizi ikili sınıflandırma (pozitif/negatif) şeklinde uygulanırken, duygusal durum da (*emotional state*) ise 6 farklı sınıflandırma (öfke / bıkkınlık / hüzn / korku / şaşkınlık / sevinç) ile çalışılmıştır. Takımların dilbilimsel, sözlüksel, istatistiksel, makine öğrenmesi gibi farklı yaklaşımlarla problemi ele almaya çalıştıkları, başarımlarının en yüksek %60'lar civarında olduğu, duygu analizinin oldukça zor bir görev olduğu çalışma sonucunda özetlenmiştir.

Salur [125] yüksek lisans tezi kapsamında Türkçe haber başlıklarının duygu analizini çalışmıştır. Makine öğrenmesi temelli yaklaşım için gerekli olan veri setleri, gazetelerin Twitter'da paylaşmış oldukları yaklaşık 105000 haber başlığından derlenmiştir. Haber başlıkları kullanıcılar tarafından pozitif/negatif/belirsiz şeklinde 3 farklı sınıfta etiketlenmiş ve büyük veri araçlarından Hadoop ekosisteminde yapılan sınıflandırmada %80 başarı seviyesinin elde edildiği belirtilmiştir.

2.7.2. Haber İçeriklerine Yönelik Duygu Analizi Çalışmaları

Haber içeriği üzerinde duygu analizi çalışmaları, araştırmacıların diğer bir ilgi alanıdır. Balahur ve arkadaşları [126], gazete haberlerinde duygu analizi üzerinde çalışmışlar ve bir haber metnine “yazar”, “metin” ve “okuyucu” olmak üzere 3 farklı perspektiften bakılması gerektiğini öne sürmüşlerdir. Yazar ve okuyucu tarafından bakıldığında, arka plan bilgilerinin (*background knowledge*), tecrübelerin, taraf olan siyasi ve politik duruşların vb. öznelliklerin farklılığına vurgu yapılarak, duygu analizinin metin perspektifinden yapılması gerektiği hipotezini savunmuşlardır. Çalışmadaki ana kısıt ise, aynı hedef varlık (özne) hakkındaki haber metinleri üzerinde analiz yapılmasıdır. Bu kapsamda, belirledikleri hedef varlığa yönelik haber metinlerinde geçen 1592 alıntı ifadeden oluşan bir veri seti oluşturulmuştur. Alıntı ifadeleri, iki değerlendirici tarafından iyi/kötü haber şeklinde manuel olarak etiketlenmiş, ardından İngilizce için var olan dil sözlüklerinden WordNet Affect [92] ve SentiWordNet [59]'in yanı sıra, kendi geliştirdikleri duygu sözlüğü ile haber metinleri üzerinde duygu analizi gerçekleştirilmiştir.

Türkçe haber içerikleri üzerinde duygu analizi çalışması kapsamında, Ertem ve arkadaşları[127] çevrimiçi gazetelerdeki hemşirelik mesleğine yönelik haberleri inceleyen çalışmalarında, 2006 yılı boyunca ana akım haber medyasındaki hemşirelik içerikleri haberleri taramışlardır. Geliştirdikleri 8 sorudan oluşan içerik çözümleme yönergesi ile hemşirelik temalı haberleri ele almışlar ve çalışma sonucunda, “Haberin Duygu Tonu” başlığı kapsamında, haberlerin yaklaşık %34’ünün “olumlu”, %36’sının “olumsuz”, %25’inin “karışık” ve %5’inin “yansız” olduğu tespitinde bulunmuşlardır.

Suat ve Yetkin [128] tarafından, Türkçe finans haberlerinde yansıtılan duygu ile borsada işlem gören şirketlerin finansal değerleri arasında bir ilişkinin olup olmadığı araştırılmıştır. Bu maksatla 2014 yılı analiz döneminde Borsa İstanbul (BIST 30)’daki firmalara yönelik basında çıkan haberler taranmış, toplamda 8434 haber metninden oluşan bir derlem elde edilmiştir. Haberlerin duygu analizi için gerekli olan duygu sözlüğü de İngilizceden Türkçeye çevrilerek oluşturulmuştur. Olumlu ve olumsuz şekilde iki sınıflı gerçekleştirilen duygu analizi sonucunda, haberler ile finansal değerler arasında yüksek bir korelasyonun var olduğu ortaya konulmuştur.

2.7.3. Haberlere Yapılan Okuyucu Yorumlarına Yönelik Duygu Analizi Çalışmaları

Haberlere yapılan okuyucu yorumları üzerinde duygu analizi çalışması, araştırmacıların ilgi gösterdikleri bir diğer alandır. Lin ve arkadaşları [129], haber metinlerine kullanıcı perspektifinden bakmaya çalışmışlar ve bu maksatla Çince haberlerden eğitim amaçlı 26K, test amaçlı da 11.4K boyutunda bir derlem oluşturmuşlardır. Bu derlemdeki her bir haber, 8 farklı duygu durumundan birisi ile okuyucular tarafından işaretlenmiştir. Çalışmada, en yüksek oylamaya sahip duygu durumu habere atanmış ve ardından farklı öznelilikler belirlenerek makine öğrenmesi temelli duygu sınıflandırması çalışılmıştır.

2.7.4. Haberlerden Duygu Sözlüğü Oluşturmaya Yönelik Çalışmalar

Haber metinlerini salt analiz amaçlı olmanın ötesinde, aynı zamanda duygu sözlüğü oluşturmak için kıymetli bir veri hazinesi olarak gören çalışmalar da mevcuttur. Xu ve arkadaşları [130], S-HAL (*Semantic-Hyperspace Analogue to Language*) ismini

verdikleri ve terimlerin hiper düzlemde bir nokta olarak temsil edildiği çalışmalarında, veri seti olarak Çince haberlerden elde ettikleri derlemi kullanmışlardır. Çalışmanın amacı, duygu analizi için terimlerin duygu yönünü belirlemek olarak ifade edilmiştir. Temel prensip, polaritesi bilinen terimlerle bilinmeyen terimlerin hiper düzlemdeki uzaklığına göre polarite tespitinin yapılmasıdır. 2008 yılının ilk altı aylık zaman diliminde toplamda 2.8 milyon haber sayfasından 530 milyon terim elde edilmiş ve ön işlem süreçleri sonucunda 116233 tekil kelimenin duygu yönü belirlenmiştir.

Bai [131] çalışmasında, Markov modeli temelinde kelimeler arasındaki ilişkisel bağımlılıkları tespit ederek bir sözlük oluşturmayı ve duygu analizinde kullanmayı amaçlamıştır. Makine öğrenmesini temel alan çalışmada, Google Haberler ve Reuters haber sitelerinden çalışmanın amacı doğrultusunda 600 haber belirlenmiş ve ardından her bir haber üç farklı değerlendirici tarafından manuel olarak üç sınıfta (negatif/nötr/pozitif) etiketlenmiştir. 3 farklı kategorideki (birleşmeler-satın almalar, finans ve karma) bu 600 haberden en yükseği 15685 terimden oluşan 3 farklı sözlük, önerilen makine öğrenmesi sürecinde kullanılmıştır.

Rao ve arkadaşları [132] çalışmalarında, dile bağımlı olmaksızın duygu sözlüğü geliştirilmesine yönelik bir metodoloji önermişlerdir. Bu hedef doğrultusunda Çince haberlerden elde ettikleri derlemdeki haberler, çevrimiçi 3064 kullanıcı tarafından 8 seviyedeki duygu durumu ile etiketlenmiş ve önerilen budama yaklaşımları sonucunda duygu sözlüğü oluşturulmuştur. İlaveten, haber metinlerinin başlıklarından, farklı bir başlık seviyesi duygu sözlüğü de geliştirilmiştir. Geliştirilen duygu sözlüklerinin performansı Çince ve İngilizce iki farklı derlem üzerinde ölçülmüştür. İlk derlem duygu etiketlemesi yapılmış 40897 Çince haberden oluşmakta iken, ikinci derlem SemEval 14. Uluslararası Duygu Değerlendirmeleri Çalıştayındaki 1246 haber başlığından oluşan İngilizce veri setidir (SemEval-2007) [133]. Kelime seviyesindeki sözlükle haber metinleri üzerinde yapılan duygu analizinde performansın oldukça başarılı olduğu tespit edilmiştir. Bunun yanında, haber başlığı seviyesindeki sözlük ile yapılan duygu analizi performansının, diğeri gibi çok üst seviyede olmadığı ve bunun gerekçesinin de haber başlıklarının kısa metinlerden oluşması gösterilmiştir.

3. TÜRKÇE DUYGU SÖZLÜĞÜ

Bölüm 1.2’de de belirtildiği gibi tez çalışmamızda iki temel hedef belirlenmiş olup, ilk hedef Genel Amaçlı Türkçe Duygu Sözlüğünün geliştirilmesidir. Bu bölümde söz konusu bu ilk hedefe yönelik yapılan çalışmalar sunulmuştur.

Duygu sözlüğünün geliştirilme süreci iki ana fazda yürütülmüştür. Bölüm 2.3’te belirtildiği gibi duygu sözlüğü oluşturmada manuel, sözlük temelli ve derlem temelli olmak üzere üç temel yaklaşım söz konusudur. Bizim yaklaşımımız ise hibrit bir karakteristiğe sahiptir. İlk fazda derlem+sözlük temelli olarak duygu sözlüğünün ilk sürümü SWNetTR-PLUS geliştirilmiştir. İkinci fazda ise bu ilk sürüm, önerilen metodolojilerle sözlük temelli olarak genişleterek kararlı hale getirilmiş ve SWNetTR++ olarak isimlendirilmiştir.

3.1. Terminoloji

Literatürde, duygu sözlüğündeki terimlerin duygu değerlerini tanımlamada duygu yönü, duygu skoru, duygu değeri, duygu polaritesi, duygu gücü gibi ifadeler kullanılmakta olup, üzerinde fikir birliğine varılmış kavram/kavramlar yoktur. Tezin geri kalan bölümünde, duygunun yönü ve değeri için aşağıdaki terminoloji kullanılmıştır:

Polarite : Duygunun yönünü ifade eder. -1 ve +1 tamsayı değerlerinden birisini alır. -1 negatif duyguyu, +1 pozitif duyguyu ifade etmektedir.

Ton : Duygunun değerini ifade eder. -1.0 ile +1.0 arasında reel bir değere sahiptir. -1.0 en negatif, +1.0 ise en pozitif duygu değeridir. Terimin pozitiflik/negatiflik derecesini belirtmektedir.

3.2. Faz-1 : SWNetTR-PLUS

Çalışmanın bu fazında, haber kaynaklarından duygu sözlüğünün elde edilmesine odaklanılmıştır. Takip eden alt bölümlerde sırasıyla kullanılan veri setleri, geliştirilen metodolojiler ve oluşturulan duygu sözlüğünün performans değerlendirme süreçlerine yer verilmiştir.

3.2.1. Veri Seti

Çizelge 3.1’de, bu fazda kullanılan/geliştirilen veri setleri kısa açıklamaları ile sunulmuştur. Detaylı açıklama gerektiren veri setlerine devam eden alt başlıklarda yer verilmiştir.

Çizelge 3.1 Faz-1’de kullanılan/geliştirilen veri setleri.

Veri Seti	Açıklama
SWNetTR	Uçan [80] tarafından; Bölüm 2.5.2’de de yer verildiği gibi yüksek lisans tezi kapsamında geliştirilen, sözlük temelli yaklaşımın çeviri tekniğiyle oluşturulan, 27K hacmindeki Türkçe Duygu Sözlüğü.
SWNetTR-GDELT	Detayları Bölüm 3.2.1.1’de verilen GDELT Projesi ⁴ verilerinden faydalanılarak derlem temeli yaklaşımı barındıran, Bölüm 3.2.2’deki metodoloji ile tarafımızca geliştirilen, 14K hacmindeki Türkçe Duygu Sözlüğü.
SWNetTR-PLUS	Birinci fazın nihai hedefi olan ve SWNetTR ile SWNetTR-GDELT sözlüklerinin birleşimi ile oluşturulan 37K hacmindeki Türkçe Duygu Sözlüğü.
MLTC	Türkçe Duygu Sözlüğü performansının test edilmesi için tarafımızca geliştirilen ve detaylarına Bölüm 3.2.1.3’te yer verilen, manuel etiketlenmiş Türkçe derlem.

⁴ <https://www.gdeltproject.org/>

3.2.1.1. GDELT Projesi

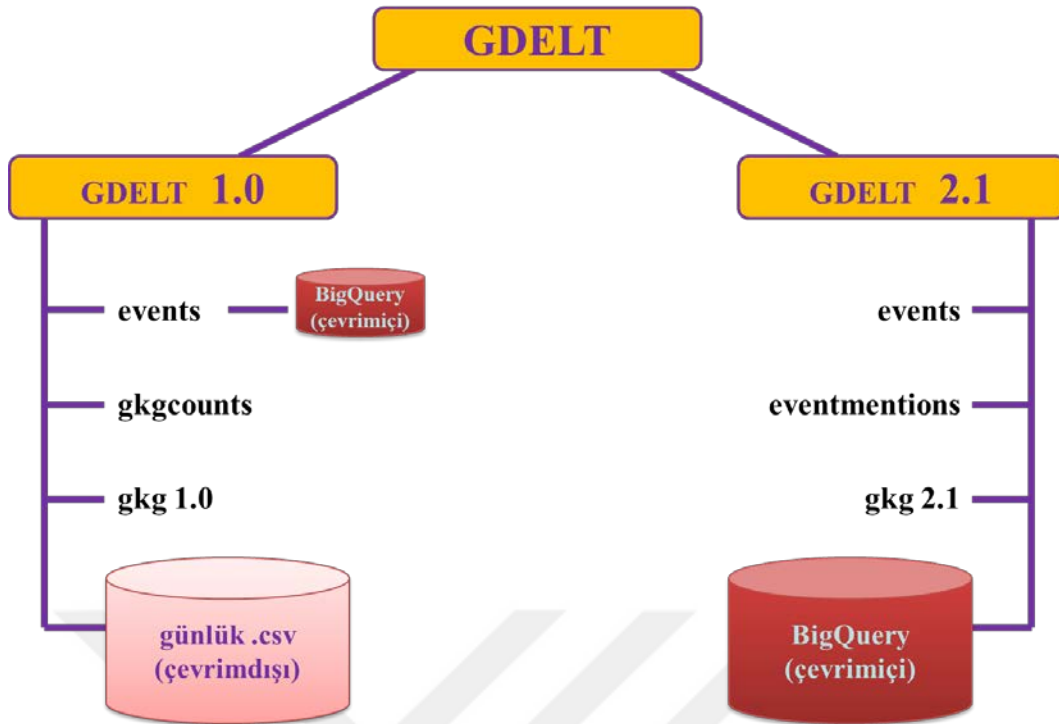
GDELT (*Global Database of Events, Language (Location) and Tone*) Projesi [134], toplumsal konularda araştırmacı olan Kalev H. Leetaru tarafından başlatılan ve geliştirilmesine devam edilen bir projedir. Amacı dünya üzerindeki bütün bilgi kaynaklarını tarayarak mantıksal formatta kodlanmış bir büyük veri oluşturmak olarak ifade edilen proje, 1.0 sürümüyle 2013 yılında hayata geçirilmiş ve açık kaynak olarak araştırmacılara sunulmuştur.

GDELT, küresel boyutta insanlığın sosyal hareketlerini 15 dakikalık zaman dilimlerinde kayıt altına alan bir mekânsal olay veritabanıdır. GDELT tarafından, bütün dünyadaki ana akım haber medyaları, diğer WEB kaynakları, TV yayınları, CORE, DTIC, JSTOR gibi dijital veri kaynakları ve haber videoları otomatik olarak taranmaktadır. Taranan kaynağın metin dili İngilizce dışında 65 farklı dilden birisine ait ise, GDELT kaynak dil tanımlayıcısı tetiklenir. Kaynak metin, gerçek zamanlı olarak İngilizceye betimlenir ve doğal dil işleme mekanizmaları çalıştırılarak, çıkarımsanan varlıklar ve her bir varlığa ilişkin etiketler ile metrikler veritabanına kaydedilir. Bu sistematik, her 15 dakikalık periyotlarda tekrarlanmaktadır.

2013 yılında ilk olarak 1.0 versiyonu ile hayata geçirilen GDELT projesi, zaman içinde ara bir sürüm olan 2.0'dan sonra 2015 yılında 2.1 sürümü ile kararlı hale getirilmiştir. 2018 yılı içerisinde duyurulan 3.0 alfa sürümüyle gelişimini sürdürmektedir. 2013 yılında hayata geçirilmiş olsa da, 1979 yılından günümüze kadar olan haberlerin analizleri veritabanına kazandırılmıştır. Geçmiş yıllara ait haberler için AFP, AP gibi büyük ajansların arşiv verileri kullanılmıştır.

1.0 sürümü “.csv” formatında ve günlük bazda GDELT'in dosya sunucuları üzerinden ücretsiz indirilebilme imkanı ile araştırmacıların hizmetine sunulurken, Şubat 2015'de hayata geçirilen 2.1 sürümünden itibaren bütün veri setleri bir bulut bilişim servisi olan Google BigQuery⁵ platformu üzerinden sunulmaktadır.

⁵ <https://bigquery.cloud.google.com/>



Şekil 3.1 GDELT sürümleri ve veri tabloları.

Şekil 3.1’de de görüldüğü üzere, GDELT projesinde farklı veri setleri bulunmakla beraber, temel olarak 3 adet veri seti (*events*, *eventmentions*, *gkg*) ön plana çıkmaktadır.

“*events*” veri seti yapısı, 1.0 ile çok büyük oranda aynıdır. *events* setindeki varlıklar Aktör1, Olay ve Aktör2’dir. Her varlığa ait konumsal veri bilgileri ile kategorik bilgiler ve olaya yönelik sayısallaştırılmış bilgiler tutulmaktadır. *events* tablosundaki kayıtlar 1979 yılına kadar bir geçmişi kapsamakta, yapısal olarak da her bir olay, bir kayıt satırına ve 61 alana sahiptir. Günlük bazda ortalama 150000 yeni olay kaydı girdisi yapılmaktadır. Her geçen gün taranan kaynak sayısının artırıldığı, alt yapının geliştirildiği ve akademik makaleler gibi ana akım haber kaynakları dışında da tarama yapıldığı bilgisi, proje geliştiricileri tarafından vurgulanmaktadır.

“*eventmentions*” veri seti ise GDELT 2.0 sürümü ile hayata geçirilmiştir. *events* tablosunda yer alan bir A olayından bahseden bütün haber kaynakları, sayısallaştırılmış bilgilerle birlikte toplamda 16 farklı alan bilgisi ile bu tabloda tutulmaktadır. Böylece A olayının bahsedilme sıklığına ve polaritesine dayalı olarak önemine ve yaygın etkisine

yönelik bir öngörü sunulmaktadır. Günlük ortalama bir milyon civarında yeni kayıt GDELT tarafından eklenmektedir.

“*gkg*” veri seti, GDELT 1.0 sürümüne oranla GDELT 2.1 sürümünde daha fazla gelişim kaydedilen GDELT derlemidir. Haber kaynaklarından, salt aktör-olay-aktör ilişkileri çıkarımı ile pozitif-negatif ton değerlendirilmesi yapılmasının yetersiz kalacağı, bağlam odaklı analizin de gerekli olduğu görüşü üzerine geliştirilmiştir. Bu doğrultuda her bir habere, 2300’ün üzerindeki etiket havuzu kullanılarak bağlam etiketlemesi yapılmaktadır. Ayrıca haber kaynağında yer alan multimedya öğeleri (resim, video), konuşmacıların alıntıları ve metnin çevirisinde kullanılan sözlük bilgileri *gkg* veri setinde sunulmaktadır. Günlük olarak yaklaşık 500000 satır yeni kayıt girişi yapılmaktadır.

GDELT’in 65 farklı dildeki haberleri ele alabilme yeteneği, bizim çalışmamızda ilgilendiğimiz ana özelliğidir. Bu amaçla Dil Çeviri Hattı (*Translingual Pipeline*) [135] ismi verilen sistemde, bir haber metninin öncelikle hangi dile ait olduğunun tespiti yapılmakta, ardından o dile ait dil kaynakları mevcut değilse İngilizceye anlık olarak çevrilerek, gelişmiş dil kütüphaneleri ile içerik analizine tabi tutulmaktadır. İngilizce dışında az sayıdaki bazı dillerde yeterli kaynağın bulunduğu göz önüne alındığında, Türkçe gibi birçok diğer dildeki ana akım haberleri anlık olarak İngilizceye çevrilmektedir. GDELT’in içerik analizinde kullandığı, gelişmiş İngilizce dil kaynaklarının listesi [136] Çizelge 3.2’de verilmiştir.

Çizelge 3.2 GDELT’in içerik analizinde kullandığı İngilizce dil kaynakları.

No	Dil Kaynağı
1	Forest Values
2	GDELT Global Knowledge Graph Themes
3	General Inquirer V1.02 (Harvard IV-4 Psychosocial Dictionary / NamenWirth & Weber’s Lasswell Dictionary)
4	Lexicoder Sentiment Dictionary
5	Lexicoder Topic Dictionaries
6	Linguistic Inquiry and Word Count (LIWC)

7	Loughran and McDonald Financial Sentiment Dictionaries
8	Opinion Observer
9	Regressive Imagery Dictionary
10	Roget's Thesaurus 1911 Edition
11	SentiWordNet 3.0
12	SentiWords
13	Subjectivity Lexicon
14	Body Boundary Dictionary
15	WordNet Affect 1.0
16	WordNet Affect 1.1
17	WordNet Domains 3.2
18	WordNet 3.1 Lexical Categories

3.2.1.2. Etiketlenmiş Türkçe Haberler Derlemi

GDELT Çizelge 3.2'deki gelişmiş dil kaynakları ile haber içeriğine yönelik yaptığı analizlerden birisi de duygu analizidir. Analiz sonucunda elde edilen duygu tonu -10 ile +10 arasında reel bir sayısal değerle, *gkg* veri tablosunun “*Tone*” alanında tutulmaktadır. Ayrıca haberin hangi dilde olduğu ve ilgili haberin URL adres bilgisi de veritabanında bulunmaktadır.

Çalışmamız kapsamında Mart 2016 tarihinde GDELT veritabanının *gkg* tablosundan, toplamda rastgele 100000 Türkçe haberin meta verileri Google BigQuery aracılığıyla sorgulanarak elde edilmiştir. Haberin URL bilgisi ile duygu tonu verisi çalışma kapsamında kullanılmıştır. Oluşacak derlemin dengeli olması için 50000 pozitif, 50000 de negatif ton değerine sahip olma kısıtı getirilmiştir. Bunun dışında haberin kaynağı, kategorisi (eğitim, sağlık, magazin vb.), uzunluğu gibi başka herhangi bir kriter verilmemiş, rastgelelik gözetilmiştir. Rastgelelik sayesinde oluşturulacak duygu sözlüğündeki sözcük hazinesinin zengin olması hedeflenmiştir. Sorgulamada haber tarihi başlangıcı 01 Ekim 2015 olarak belirtilmiş ve yaklaşık dört günlük bir zaman diliminde 100000 habere ulaşılmıştır. Kısa zaman diliminde bu kadar yüksek sayıda Türkçe haberin varlığı, GDELT'in dünya haber medyasını çok geniş bir yelpazede taradığının da bir göstergesidir.



Şekil 3.2 Örnek bir haber sayfası⁶ ve haber metni.

Gelinen aşamada, %50'si pozitif, %50'si de negatif polariteli, toplam 100000 Türkçe habere ait URL bilgisinden oluşan veri seti elde edilmiştir. Ardından, URL kaynakları teker teker ziyaret edilerek, haber metinlerinin elde edilmesi sürecine geçilmiştir. Bu amaç için, Python dilinde “*boilerpipe*” kütüphanesi kullanılarak yazılan kod parçası kullanılmıştır. Haber sayfalarında, Şekil 3.2'deki gibi haber metni dışında yer alan reklamlar, linkler, farklı haber başlıkları ve içerikleri gibi ekstra alanların filtrelenmesi “*DefaultExtractor*” parametresiyle sağlanmıştır. Mart 2016 tarihinde, 82912 haber sayfasına ulaşılmış, geri kalan 17088 linkin ölü olduğu tespit edilmiştir.

⁶ <https://www.ntv.com.tr/saglik/romatizma-hastalarini-bekleyen-tehlike,ofjhKRi020u0CB5URqBVg>

Çizelge 3.3 Etiketlenmiş Türkçe haberler derlemi.

	Pozitif	Negatif	Toplam
Erişilen Haber Metni	41769	41143	82912
	%50.38	%49.62	

Bu sürecin sonunda etiketli Türkçe haberler derlemi oluşturulmuştur. Çizelge 3.3'te de görüldüğü gibi derlem, pozitif ve negatif haberlerin sayısal eşitliği bakımından dengelidir.

3.2.1.3. MLTC: Manuel Etiketlenmiş Türkçe Test Derlemi

Duygu sözlüğünün performansının ölçülmesi amacıyla, etiketlenmiş Türkçe bir derleme ihtiyaç vardır. Bu ihtiyacı karşılamak için toplamda 500 haber, 3 değerlendirici tarafından pozitif ve negatif olmak üzere ikili sınıflandırmaya tabi tutulmuştur. Değerlendiricilerin bir habere yönelik yaptıkları değerlendirmede, oy çokluğu esasına göre haberin nihai polaritesi belirlenmiştir. Oluşturulan bu test derlemi MLTC-500 olarak isimlendirilmiştir.

500 haberden 353'ü tüm değerlendiriciler tarafından aynı polarite ile etiketlenmiştir, yani oy birliği söz konusudur. Bir başka ifadeyle, 353 haberin tamamı bütün değerlendiricilere göre pozitifdir (veya negatiftir). Performans değerlendirme sürecinde MLTC-500'e ilaveten, MLTC-353 olarak isimlendirdiğimiz bu derlem de kullanılmıştır.

Çizelge 3.4'de MLTC-500 ve MLTC-353 manuel etiketlenmiş Türkçe test derlemlerinin, değerlendirici tespitlerine göre polarite dağılımları sunulmuştur.

Çizelge 3.4 MLTC-500 ve MLTC-353 Türkçe test derlemelerinin polarite dağılımları.

Türkçe Test Derlemi	Pozitif Haber Sayısı	Negatif Haber Sayısı	Toplam Haber Sayısı
MLTC-500	265	235	500
	%53.00	%47.00	
MLTC-353	202	151	353
	%57.22	%42.78	

Değerlendiriciler arasındaki uyum, istatistiksel olarak da hesaplanmıştır. Literatürdeki Fleiss Kappa istatistiği [137], ikiden fazla değerlendirici arasındaki uyumu ölçen bir metriktir. Bu istatistiğin bir avantajı kolay hesaplanması ve yorumlanması iken, diğer avantajı da uyuşmanın şans eseri olma durumunun elimine edilmesini de sağlamasıdır. Fleiss Kappa istatistiği Eşitlik-1'deki gibi ifade edilmektedir.

$$\kappa = \frac{P_0 - P_e}{1 - P_e} \quad (1)$$

Eşitlik (1)'deki P_0 gözlenen uyumluluk oranını, P_e tesadüfi uyumluluk oranını ifade etmektedir. Kappa istatistiği -1 ile +1 arasında reel bir değere sahiptir. κ değerinin sıfır ve negatif olması iki değerlendirici arasında bir uyumun olmadığını gösterir. κ değeri +1'e yaklaştıkça değerlendiricilerin uyumunun şans eseri olmaktan uzaklaşmaya başladığı, yani homojenliğin sağlanmaya başladığı söylenebilir. +1 değeri tesadüften uzak mükemmel bir uyumu ifade eder. κ değer aralıklarını Çizelge 3.5'deki gibi yorumlamak mümkündür [138, 139].

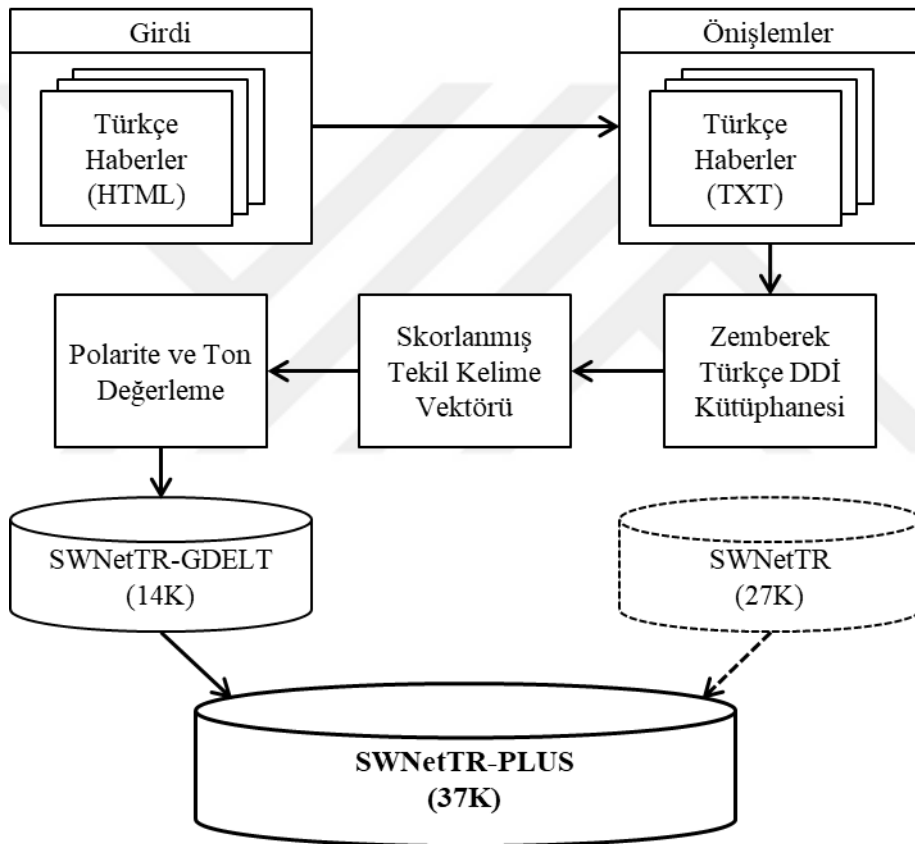
Çizelge 3.5 Fleiss Kappa istatistiğinin yorumlanması.

κ	Uyumun Gücü
< 0.00	Zayıf
0.00 - 0.20	Önemsiz
0.21 - 0.40	Düşük
0.41 - 0.60	Orta
0.61 - 0.80	Önemli
0.81 - 1.00	Çok Yüksek

MLTC-500 Türkçe test derleminde $\kappa = 0.61$ olarak hesaplanmış ve 3 değerlendiricinin “Önemli” derecede uyumlu olduğu tespit edilmiştir. Beklenileceği gibi MLTC-353 derlemi, oy birliği karakteristiğine sahip olduğundan $\kappa = 1.00$ ’dır.

3.2.2. Metodoloji

Şekil 3.3’te, bu fazda hedeflenen genel amaçlı Türkçe duygu sözlüğünün ilk aşaması olan SWNetTR-PLUS’ın geliştirilmesinde izlenen adımlar sunulmuştur.



Şekil 3.3 SWNetTR-PLUS geliştirme süreci blok diyagramı.

3.2.2.1 Haber Derleminin Oluşturulması

Detayları Bölüm 3.2.1.2’de ifade edildiği gibi sürecin ilk adımında, 100000 habere ait sayfalar ziyaret edilerek haber metinleri çekilmiş ve sonuçta 82912 adet ton ve polaritesi bilinen Türkçe haber derlemi oluşturulmuştur.

3.2.2.2 Haber Metinlerinde Önişlemler

Bazı haber sitelerinin, standartlar gözeterek iyi biçimlendirilmiş (*well-formed*) formatta web sayfalarını yapılandırdıkları gözlemlenirken, bu standartları göz ardı eden sayfaların da varlığı, önişlem sürecinde ele alınması gereken bir problem olarak karşımıza çıkmıştır. Şekil 3.2’de örnek bir haberin web sayfası ile bu sayfadan elde edilen metin görülmektedir. Önişlem sürecinde, ilk olarak metin formatındaki her bir haber, HTML ve javascript kod ifadelerinden, özel sembollerden ve sayısal ifadelerden filtrelenerek arındırılmıştır. Ardından, bazı haber sitelerinin web sayfalarında karakter seti bildirimi (örneğin utf-8) betimlenmediğinden, Türkçe ASCII karakter problemleri (ç, ğ, ı, ö, ü, ş, İ) ele alınmıştır. Çalışmanın bu aşamasındaki genel yaklaşım kelime-torbası (*bag-of-words*) olduğundan, noktalama işaretleri de filtrelenerek, önişlem süreci tamamlanmıştır.

Sıradaki aşamada, biçimbilimsel analiz yapabilen Zemberek [140] Türkçe DDİ kütüphanesi kullanılarak, sondan eklemeli Türkçe dili için oldukça önemli olan kelimelerin kök formları, haber içeriklerinden elde edilmiştir. Bir kelime için birden fazla kök adayı bulunduğu anda, bu adaylar Zemberek tarafından bünyesindeki kök aday seçici mekanizmasına göre sıralanarak verilmektedir. Bu durumda, Zemberek tarafından önerilen ilk aday, kök kelime olarak kullanılmıştır. Çizelge 3.6’da örnek kelimeler ile bu kelimelerin Zemberek tarafından önerilen kök formları sunulmuştur.

Çizelge 3.6 Kelimelerin kök formu örnekleri.

Kelime	Kök Formu
ok	ok
oktan	oktan / ok
okumak	oku
okuyorum	oku
okuldan	okul
okulundan	okul / oku

3.2.2.3 Ton ve Polarite Değerleri

Köklerin tespiti aşamasında aşağıdaki veriler de eş zamanlı olarak elde edilmiştir. Bu veriler:

- Kelimelerin haber metinlerinde geçme frekansı.
- Kelimelerin ait olduğu haber metnine GDELT tarafından atanmış ton skoru.

Çizelge 3.7’de frekansı en yüksek ilk 5 kelime ile, frekansı 1 olan 5 kelime görülmektedir. Altı çizili olan kelimeler, duygu analizi sürecinde durak kelimeler (*stop words*) grubuna girmekte ve genellikle hesaplama sürecine dahil edilmemektedir. Ancak, çalışmamızın bu aşamasında durak kelime filtrelemesi yapılmamıştır.

Çizelge 3.7 Haberlerden elde edilen en yüksek ve en düşük frekanslı kelimeler.

En Yüksek Frekanslı Kelimeler		En Düşük Frekanslı Kelimeler	
Kelime	Frekans	Kelime	Frekans
<u>ve</u>	72249	abajur	1
<u>ol</u>	66958	afetzedede	1
<u>et</u>	63915	ahize	1
<u>bir</u>	60825	alicensap	1
<u>bu</u>	58879	alafranga	1

Gelinen aşamada, kök formu yapısında toplamda 11.3 milyon kelime havuzu ve tekil olarak 14023 (14K) adet sözcüğe ait kelime ton vektörleri elde edilmiştir. 14K hacmindeki terimlerin ton vektörlerinden, her bir kelimenin ton skoru Eşitlik (2) de verilen matematiksel ifadeyle hesaplanmıştır.

$$S_T = \frac{\sum_{i=1}^n (d_i \cdot f_i)}{\sum_{i=1}^n f_i} \quad (2)$$

Eşitlikte n , T teriminin geçtiği toplam doküman sayısını; d_i , dokümanın tonunu; f_i , T teriminin dokümanda geçme frekansını; S_T , T teriminin ton skorunu ifade etmektedir.

Çizelge 3.8 “intizar” teriminin ton vektörü.

Doküman	Doküman Tonu	Terim Frekansı
Haber-1	-5.617978	3
Haber-2	-2.105263	3
Haber-3	-5.445545	4
Haber-4	-2.040816	3
Haber-5	-1.939655	2
Haber-6	6.666667	1
Haber-7	5.769231	3

Çizelge 3.8’de örnek olarak “intizar” kelimesinin ton vektörü sunulmuştur. “intizar” sözcüğü, toplamda 19 kez olmak üzere 7 farklı haber dokümanında yer almış ve bu haberlerden 2 tanesi pozitif tonda iken geri kalan 5 tanesi negatif tonda verilmiştir. Eşitlik (2)’deki matematiksel ifadeyle “intizar” kelimesinin ton değeri -1.63049 olarak hesaplanmıştır.

Bölüm 3.2.1.2’de de belirtildiği gibi, GDELT duygu analizinde ton değeri için -10 ile +10 arasında reel sayı sistemi kullanılmaktadır. Tez çalışmamızda ise, Bölüm 3.1’de de ifade edildiği gibi ton değeri için -1 ile +1 arasında reel sayı sistemi kullanılmıştır. Bu nedenle, Eşitlik (2) ile hesaplanan terimlerin ton değeri -1 ile +1 arasında olacak şekilde normalize edilmiştir. Yukarıdaki “intizar” kelimesi örneğinde, bu kelimeye atanan ton değeri normalize işleminden sonra -0.16305 olarak gerçekleşmiştir.

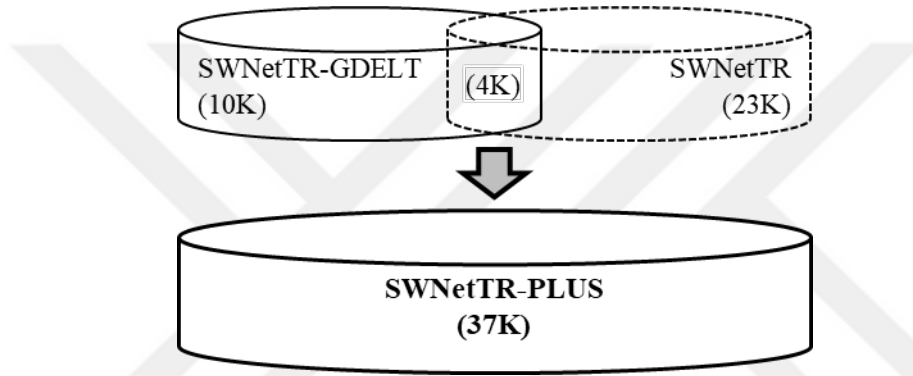
Polarite değeri bizim sistematiğimizde -1 veya +1 tam sayılarından oluşan ikili sistematiğidir. Yukarıdaki örnekte yer alan “intizar” kelimesinin polarite değeri, hesaplanan ton değeri doğrultusunda -1 olarak atanmıştır. “intizar” kelimesinin hesaplanan değerleri Çizelge 3.9’da sunulmuştur.

Çizelge 3.9 “intizar” kelimesinin ton ve polarite değerleri.

	Ton	Polarite
intizar	-0.16305	-1

Bütün kelimelere, kendileri için hesaplanan ton değerine göre polarite değeri atanarak, SWNetTR-GDELT olarak isimlendirdiğimiz, 14K hacminde, derlem temelli duygu sözlüğümüz bu aşamada oluşturulmuştur.

3.2.2.4 SWNetTR-PLUS Duygu Sözlüğünün Oluşturulması



Şekil 3.4 SWNetTR-PLUS Genel Amaçlı Türkçe Duygu Sözlüğünün yapısı.

Faz-1’in son aşamasında, Uçan [80] tarafından sözlük temelli ve çeviri sistemi ile geliştirilen ve bu tezde SWNetTR olarak isimlendirdiğimiz 27K hacmindeki duygu sözlüğü ile SWNetTR-GDELT sözlüğü, Şekil 3.4 gösterildiği gibi birleştirilmiştir. Böylece 37K hacmindeki SWNetTR-PLUS ismini verdiğimiz Genel Amaçlı Türkçe Duygu Sözlüğünün ilk sürümü oluşturulmuştur.

3.2.3. Performans Değerlendirmesi

Oluşturulan SWNetTR-PLUS duygu sözlüğünün başarımlı performansı, Bölüm 3.2.1.3’te belirtilen MLTC-500 ve alt kümesi olan MLTC-353 test veri setleri üzerinde ayrı ayrı ölçülmüştür. MLTC-500 setindeki haber metinleri, Bölüm 3.2.2.2’deki gibi ön işlem sürecinden geçirilmiş, ilave olarak da durak kelimelerinden arındırılmıştır. Durak kelimeleri, birçok metinde yer alan ama duygu değeri taşımayan, çoğunlukla işaret sıfatları ve bağlaçlardan oluşan terimlerdir.

Duygu sözlüğü ile duygu sınıflandırması, dokümanda geçen terimin sözlükteki polarite veya ton değerlerinin toplamına bakılarak yapılmaktadır. Polarite baz alındığında toplam değer sıfırdan küçükse negatif, büyükse pozitif olarak sınıflandırılır. Ton baz alındığında ise, dokümanın duygusunun ne derece pozitif ya da negatif olduğu bilgisi de, pozitif/negatif sınıflandırmayla birlikte edinilmiş olur. Yapılan testlerde, polarite bazlı değerlendirmede negatif çıkan bir analiz, ton bazlı değerlendirmede pozitif çıkabilmekte veya tam tersi olabilmektedir.

Mevcut durumda değerlendirmeye hazır hale getirilen MLTC-500 ve MLTC-353 test derlemleri, hem SWNetTR hem de bu fazda geliştirilen SWNetTR-PLUS duygu sözlükleri ile duygu analizine tabi tutulmuştur. Çizelge 3.10'da doğruluk değerlendirme sonuçları, GDEL'T'in de performansını gösterecek şekilde sunulmuştur.

Çizelge 3.10 SWNetTR-PLUS performans değerlendirme sonuçları.

	MLTC-500		MLTC-353	
	Polarite Bazlı	Ton Bazlı	Polarite Bazlı	Ton Bazlı
GDEL'T	%78.80		%84.99	
SWNetTR	%60.60	%64.20	%62.89	%67.99
SWNetTR-PLUS	%72.20	%75.20	%75.07	%78.75

Sonuçlar incelendiğinde;

- Bütün değerlendiricilerin üzerinde hem fikir oldukları haberlerden oluşan MLTC-353 test derlemindeki performansların, MLTC-500 derleminden daha yüksek olduğu,
- Ton bazlı değerlendirmelerin, Polarite bazlı değerlendirmelerden daha iyi performans gösterdiği,
- GDEL'T'in performansının her iki test derleminde de en yüksek değere sahip olduğu,

- Bu fazda geliřtirdiđimiz SWNetTR-PLUS duygu szlđnn, SWNetTR'den daha iyi olduđu grlmektedir.

Ancak, SWNetTR-PLUS ile belirli bir bařarı seviyesi (%75.20) yakalanmıř olsa da, alıřmamızdaki esas kıyas noktamız olan GDELT'in performansının (%78.80) gerisindedir. Bu nedenle de geliřtirilmesine ihtiya vardır. Bu sonular ve tespitler, alıřmamızın bir sonraki blm olan Faz-2'nin ana motivasyonunu oluřturmuřtur.

3.3. Faz-2 : SWNetTR++

Bir nceki fazda, 37K kapasiteli SWNetTR-PLUS duygu szlđnn performansı hedeflenen seviyenin altında gerekleřmiřtir. Buna gereke olarak;

- Duygu szlđnn kapasitesinin eksik olduđu, bu yzden bazı terimlerin szlkte bulunmaması,
- Ton ve polarite deđerlerinde de gncellemeye ihtiya olduđu tespitlerimiz vardır.

Bu ana eksikliklerle beraber, Trkede bir terimin anlamsal ve szdizimsel negatif karřılıkları mevcuttur. rneđin, “bymek” teriminin anlamsal negatifi “klmek” iken, szdizimsel negatifi ise “bymemek” terimidir. SWNetTR-PLUS szlđmz oluřturan, gerek SWNetTR szlđnde, gerekse SWNetTR-GDELT szlđnde negatif terimlerin azlıđı/olmaması da performansı etkileyen bir bařka noktadır.

Tm bu problemlerin ele alınarak zmler nerilmesi ve geliřmiř kaynaklara sahip diđer dillerin seviyesine Trkenin de eriřmesine katkı sunulabilmesi, Faz-2 kapsamında hedeflenmiřtir.

3.3.1. Veri Seti

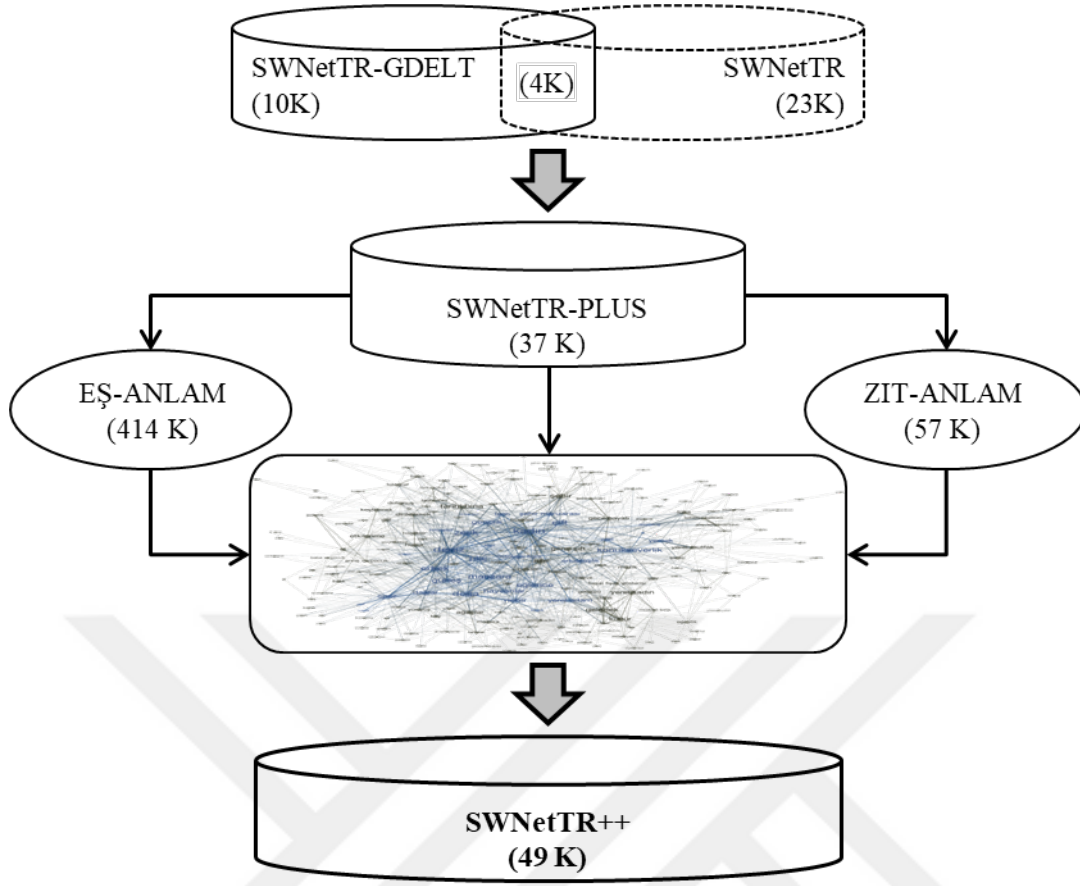
Bu fazda da, bir önceki fazda geliştirilen Çizelge 3.1’deki veri setleri kullanılmıştır. Bunlara ilaveten, bu fazda geliştirilen veri setleri kısa açıklamalarıyla beraber Çizelge 3.11’de sunulmuştur.

Çizelge 3.11 Faz-2’de geliştirilen veri setleri.

Veri Seti	Açıklama
EŞ-ZIT	SWNetTR-PLUS’ın kapasitesini geliştirmeye yönelik olarak internetteki açık kaynaklardan tarafımızca derlenen ve detaylarına Bölüm 3.3.2.1’de yer verilen eş ve zıt anlam karşılıkları veri seti.
SWNetTR++	İkinci fazın nihai hedefi olan ve SWNetTR-PLUS’ın geliştirilmesiyle oluşturulan 49K hacmindeki Türkçe Duygu Sözlüğü.

3.3.2. Metodoloji

Şekil 3.5’de, SWNetTR++’ın başlangıçtan itibaren gelişim süreci görülmektedir. Takip eden başlıklarda, süreçte uygulanan metodolojilere yer verilmiştir.



Şekil 3.5 SWNetTR++ geliştirme süreci blok diyagramı.

3.3.2.1. EŞ-ANLAM ve ZİT-ANLAM Derlemlerinin Oluşturulması

Genel amaçlı duygu sözlüğünün ilk sürümü olan SWNetTR-PLUS'ın kapasitesinin geliştirilerek iyileştirilmesine yönelik olarak eş ve zıt anlam karşılıklarından faydalanılması yaklaşımı benimsenmiştir. Sözlükteki terimlerin negatif anlam karşılıklarının belirlenmesi için zıt anlam, sözlüğün kapasitesinin artırılmasına yönelik olarak eş anlam karşılıklarının tespiti için dört farklı web kaynağından faydalanılmıştır. Çizelge 3.12'de edinilen eş ve zıt anlam sayıları ile kaynaklar sunulmuştur.

Çizelge 3.12 EŞ-ZİT anlam karşılıkları için kullanılan kaynaklar ve kapasiteler.

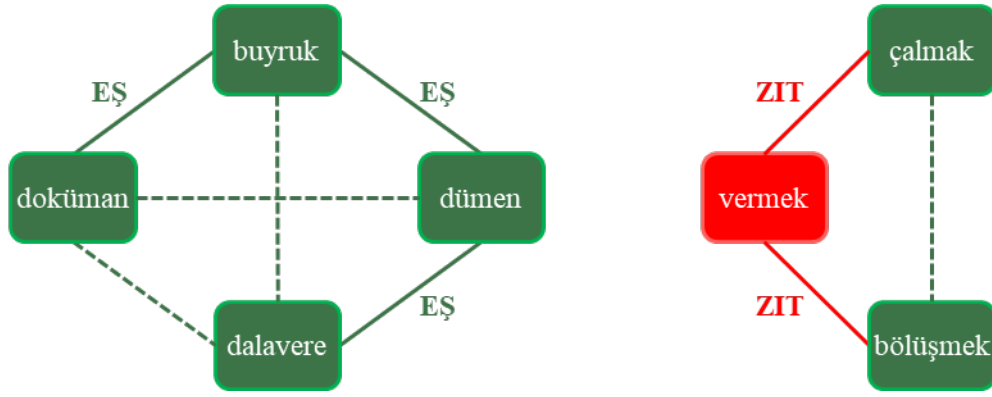
Kaynak	Eş Anlam Çifti	Zıt Anlam Çifti
www.dilbilgisi.net [141]	844	1447
www.es-anlam.com [142]	19816	1272
www.supersozluk.com [143]	371794	57521
YTU Kemik Dil Grubu [144]	59002	342

Manuel olarak oluşturulmuş olan www.dilbilgisi.net kaynağından en az veri elde edilirken, en fazla anlam karşılığı www.supersozluk.com'dan elde edilmiştir. Kaynaklardan elde edilen veriler birleştirilerek toplamda 414275 eş anlam çiftinden oluşan EŞ-ANLAM derlemi ile 57148 zıt anlam çiftinden oluşan ZIT-ANLAM derlemi elde edilmiştir. Çizelge 3.13'te elde edilen eş ve zıt anlam çiftlerine ait örnek sunulmuştur.

Çizelge 3.13 Eş ve zıt anlam çiftlerine ait örnekler.

Eş Anlam Çifti	Zıt Anlam Çifti
acemi - bilgisiz	acemi - erbap
acemi - kıdemsiz	acemi - usta
açgözlü - doymaz	açgözlü - eliaçık
açgözlü - ihtiraslı	açgözlü - tokgözlü

Kullanılan açık erişimli kaynaklar irdelendiğinde, hatalı eşlemelerin olduğu tespit edilmiştir. Örneğin “doküman-buyruk”, “buyruk-dümen”, “dümen-dalavere” eş çift olarak kaynaklardan edinilmiştir. Bu durumda “doküman” terimi ile “dalavere” terimi de eş anlamlıdır sonucu çıkmaktadır. Benzer şekilde “vermek-çalmak” ile “vermek-bölüşmek” zıt çiftleri kaynaklardan elde edilmiştir. Bu durumda da “çalmak” ile “bölüşmek” eş anlamlı durumuna düşmektedir. Şekil 3.6’da problem daha net görülmektedir. Şekilde, üzerinde EŞ ve ZIT ifadesi bulunan bağlantılar kaynaktan gelen eşleşmeleri ifade etmektedir. Bu eşleşmeler doğal olarak, kesikli çizgilerle gösterilen diğer eşleşmeleri oluşturmakta ve mantıksal hata da bu noktada gündeme gelmektedir.



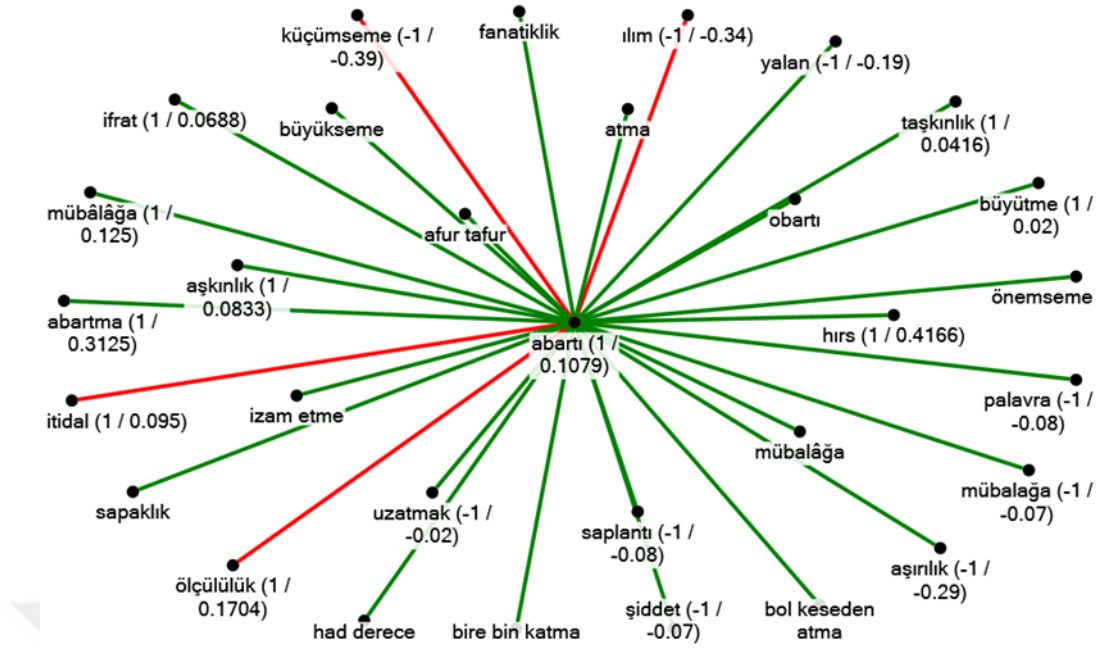
Şekil 3.6 Eş ve zıt eşleşmelerdeki mantıksal hatalara ait birer örnek.

Açık kaynaklardan elde edilen verilerdeki bir diğer hata da, yazım hatası olan terimlerin varlığıdır. Örneğin üç farklı şekilde yazılmış “mübalağa”, “mübalâğa” ve “mübâlâğa” kelimesi mevcuttur.

3.3.2.2. Problemin Çizge Dünyasına Taşınması

Hatalı eşleşmeler problemi, varlıklar arasındaki ilişkilerin en iyi temsil edildiği veri yapısı olan çizge ile modellenmiştir. Bu probleme ilaveten, bir diğer problem olan birden fazla yazılışa sahip terimlerin varlığı ile sözlüğe yeni kazandırılacak terimlerin ton ve polarite değerlerinin nasıl tespit edileceği problemleri de çizge veri yapısında ele alınarak çözüm aranmıştır.

Çizgede ilk olarak SWNetTR-PLUS sözlüğündeki terimlerle düğümler temsil edilmiş, ardından EŞ-ANLAM ve ZIT-ANLAM derlemlerindeki terimler ilgili oldukları terimlerle ilişkilendirilerek ağa dahil edilmiştir. Çizgenin kenarlarında anlam ilişkisi (eş/zıt) bilgisi bulunmakta iken, düğümlerde de ton ve polarite değerleri bulunmaktadır. Bu bilgileri de barındırdığı için oluşan çizge, Etiketli Yönsüz Çizge (*annotated undirected graph*) karakteristiğindedir. Şekil 3.7’de oluşan çizgenin bir alt çizgesi olarak “abartı” düğümünün bağlantılarına yer verilmiştir. Yeşil renkli kenarlarla eş anlam bağlantıları, kırmızı renkli kenarlarla da zıt anlam bağlantıları görselleştirilmiş, ilaveten polarite ve ton değerine sahip düğümlerin değerleri de parantez içerisinde sunulmuştur.



Şekil 3.7 Çizgedeki ‘abartı’ düğümünün eş-zıt bağlantıları örneği.

Büyük ölçekli bu ağda, toplam 79326 düğüm ve 471258 kenar mevcuttur. Ağda bağlı-bileşen (*connected-component*) sayısı 22087 iken, bu bağlı-bileşenlerin en büyüğünde (*giant connected-component*) 48861 düğüm mevcuttur. Açık kaynaklardan gelen hatalı eş-zıt bağlantı eşleşmeleri, çizgedeki zayıf bağlantıları artırmış, sonuçta fazla sayıda düğümün birbirine bağlı olmasına neden olmuştur. Bu üzerinde durulması ve çözülmesi gereken bir problemdir. Çözüm, düğümler arasındaki zayıf bağlantıların koparılmasından geçmektedir. Ancak, ağdaki hangi kenarların koparılacağını tespitine yönelik olarak çizge dünyasınca sunulan doğrudan bir çözüm metriği yoktur. Bu amaçla bir sonraki bölümde belirtilen Bağ Gücü (*Tie Strength*) yaklaşımı çözüm olarak önerilmiştir.

3.3.2.3. Bağ Gücü

Bağ Gücü (BG), iki düğüm arasındaki kenara atanan sayısal bir değer olup, terimler arasındaki zayıf ve güçlü bağlantıların belirlenmesine imkan vermek üzere geliştirilmiştir. BG, sayısal olarak 0 ile 1 arasında reel bir değere sahiptir. 1’e yaklaştıkça bağın güçlendiği, 1 değeri ile maksimum BG’nün varlığı ifade edilmektedir.

BG değeri, birbirine bağlı iki terimin eş ve zıt ilişkisine göre farklı matematiksel tanımlamaya sahiptir. A ve B birbirine bağlı iki düğüm olmak üzere aralarındaki ilişki türü eş anlam ise, bağ gücü $BG_{A-B}^{EŞ}$ Eşitlik (3) ile hesaplanmaktadır.

$$BG_A^{EŞ} = \frac{|A^{EŞ} \cap B^{EŞ}|}{|A^{EŞ}|} \quad BG_B^{EŞ} = \frac{|A^{EŞ} \cap B^{EŞ}|}{|B^{EŞ}|} \quad BG_{A-B}^{EŞ} = BG_A^{EŞ} \times BG_B^{EŞ} \quad (3)$$

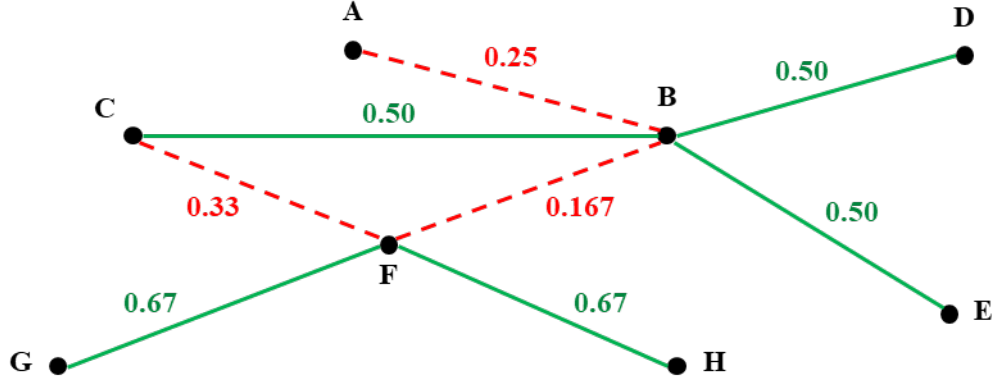
Eşitlik (3)'te $A^{EŞ}$, A düğümüne EŞ bağlantı türü ile bağlı olan komşular kümesini ifade etmektedir ve bu kümeye A düğümünün kendisi de dahildir. $B^{EŞ}$ de B düğümüne EŞ bağlantı türü ile bağlı olan ve B'nin kendisinin de dahil olduğu kümedir.

$BG_A^{EŞ}$, A ve B düğümünün arasındaki eş bağlantının A düğümü tarafından bakıldığındaki değerini ifade ederken, $BG_B^{EŞ}$ ise B düğümü tarafından bakıldığındaki değerini ifade etmektedir. Burada dikkat çekilen nokta, A ve B'nin farklı komşuları ve komşuluk ilişkilerinden dolayı bu iki değer her zaman eşit olmayacağıdır, bir diğer ifadeyle asimetri söz konusudur. Sonuçta A-B arasındaki bağın gücü $BG_{A-B}^{EŞ}$ ile tespit edilmekte ve ağ üzerinde bu kenara nitelik olarak atanmaktadır.

A ve B düğümleri arasındaki ilişki türü zıt anlam ise, BG_A^{ZIT} A düğümü tarafından bakıldığındaki zıt bağlantı değerini, BG_B^{ZIT} B düğümü tarafından bakıldığındaki zıt bağlantı değerini ifade etmekte olup, bağ gücü BG_{A-B}^{ZIT} Eşitlik (4) ile hesaplanmaktadır.

$$BG_A^{ZIT} = \frac{|A^{EŞ} \cap B^{ZIT}|}{|A^{EŞ}|} \quad BG_B^{ZIT} = \frac{|A^{ZIT} \cap B^{EŞ}|}{|B^{EŞ}|} \quad BG_{A-B}^{ZIT} = BG_A^{ZIT} \times BG_B^{ZIT} \quad (4)$$

Eşitlik (4)'te, $A^{EŞ}$ ve $B^{EŞ}$ sırasıyla A ve B düğümlerinin eş bağlantılı komşular kümesini, A^{ZIT} ve B^{ZIT} ise zıt bağlantılı komşular kümesini ifade etmektedir. Eş kümelerde düğümlerin kendileri de kümenin bir elemanıdır. Eşitlik (3) ile aynı mantıksal temele sahip olan bu notasyonda BG_A^{ZIT} , BG_B^{ZIT} ve BG_{A-B}^{ZIT} zıt bağlantıları betimleyen ifadelerdir.



Şekil 3.8 Bağ Gücü değeri hesaplanmış bir çizge örneği.

Şekil 3.8'deki örnek çizgede, düğümler terimleri, yeşil renkli düz çizgiler EŞ bağlantıları, kırmızı renkli kesikli çizgiler de ZIT bağlantıları temsil etmektedir. Kenarlardaki sayısal değerler de iki komşu düğümün arasındaki bağın gücünü ifade etmektedir. Örneğin, B ve C düğümleri arasındaki EŞ bağın gücü Eşitlik (3) ile hesaplanmaktadır. Buna göre;

$$B^{EŞ} = \{B, C, D, E\} \quad ve \quad C^{EŞ} = \{B, C\}$$

kümelerine göre;

$$BG_B^{EŞ} = 0.5 \quad ve \quad BG_C^{EŞ} = 1$$

olarak hesaplanacağından, B-C arasındaki bağ gücü $BG_{B-C}^{EŞ} = 0.5$ olacaktır.

Zıt bağlantı türüne örnek olarak B ve F düğümleri verilebilir ve bu düğümler arasındaki ZIT bağın gücü Eşitlik (4) ile hesaplanmaktadır. Buna göre;

$$B^{EŞ} = \{B, C, D, E\}, \quad B^{ZIT} = \{A, F\}, \quad F^{EŞ} = \{G, H, F\}, \quad F^{ZIT} = \{B, C\}$$

kümelerine göre;

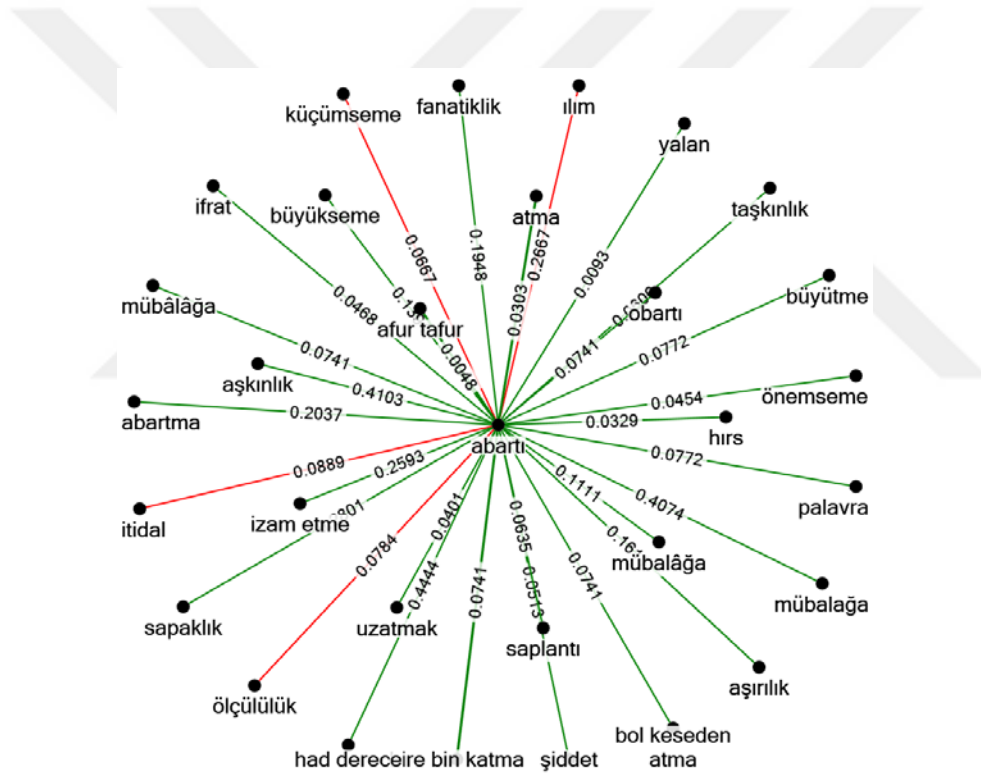
$$BG_B^{ZIT} = 0.5 \quad ve \quad BG_F^{ZIT} = 0.33$$

olarak hesaplanacağından, B-F arasındaki bağ gücü $BG_{B-F}^{ZIT} = 0.167$ olacaktır.

Düğümün EŞ ve ZIT kümelerine kendilerinin de dahil edilmesinin gerekçesi, iki düğümün sadece birbirine bağlı olduğu ve hiçbir komşularının bulunmadığı durum göz

önünde bulundurulduğunda daha net anlaşılabilecektir. Örneğin A düğümü ve B düğümünün eş anlamlı olduğu ve başka herhangi bir düğümle komşuluk bağlarının olmadığı bir senaryoda, A ve B düğümünün eş bağ gücünün en yüksek değere, yani 1.0'a sahip olması beklenir. Eşitlik (3) gereği, her iki düğümün eş kümelerine kendilerinin de dahil edilmesi ile bu sonuç elde edilebilecektir. Bu yaklaşım, benzer senaryoda A ve B'nin zıt anlamlı olduğu ve Eşitlik (4)'ün kullanıldığı durum için de geçerlidir.

Şekil 3.9'daki ağ yapısında, örnek olarak “abartı” kelimesinin komşuları ile arasındaki bağ gücünü gösteren çizge sunulmuştur.



Şekil 3.9 “abartı” kelimesinin ilişki ağındaki bağlantılarının bağ gücü.

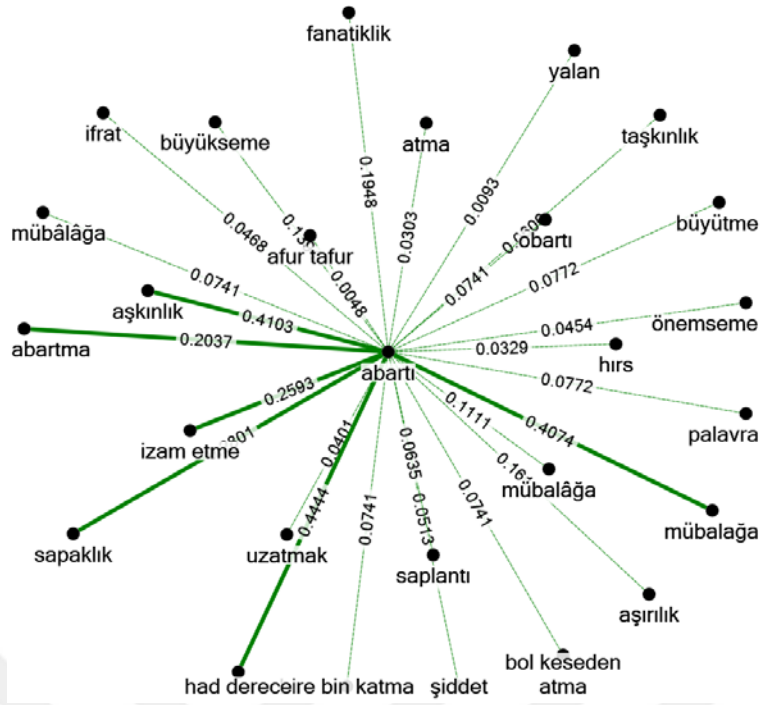
Sonuç olarak Eşitlik (3)'teki temel mantık, iki komşu düğümün eş kümeleri ne kadar çok örtüşürse, aralarındaki eş bağın gücünün de o nispette artacağıdır. Zıt kümelerde ise durum biraz daha farklıdır ve Eşitlik (4)'teki temel mantık ise, iki komşu düğümden birinin eş kümesi ile diğerinin zıt kümesi ne kadar çok örtüşürse, aralarındaki zıt bağın gücünün o oranda yüksek olacağıdır.

3.3.2.4. Ağdaki Zayıf Bağların Koparılması

Mevcut durumda, SWNetTR-PLUS sözlüğüne eş-zıt sistematigiyle yeni eklenen terimlerle oluşan ve terimler arasındaki bağın gücünün de bilindiği büyük bir çizge mevcuttur. Bu aşamada, ağdaki zayıf bağların tespit edilerek koparılması ve çizgenin daha kararlı bir hale getirilmesi süreci işletilecektir.

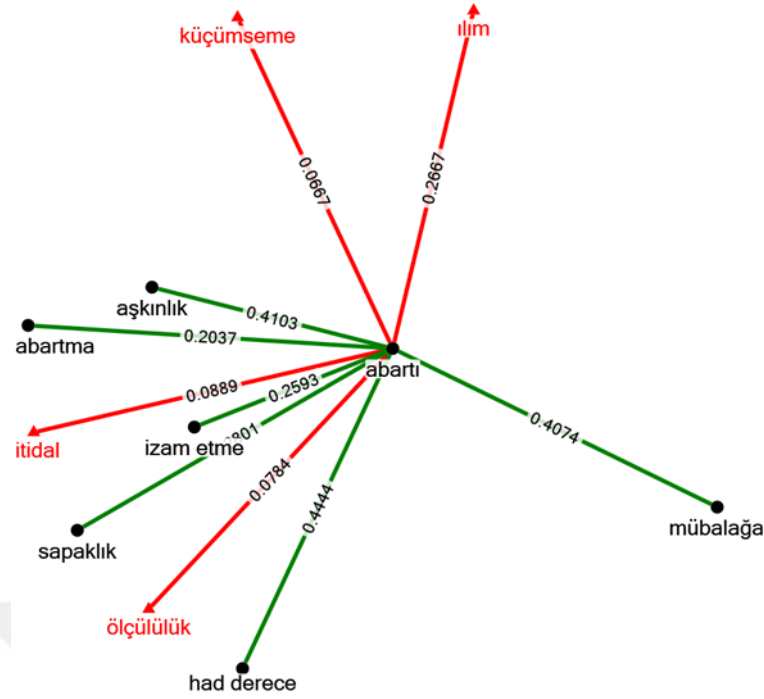
Örneğin Şekil 3.9’da verilen “abartı” kelimesi, “uzatmak”, “saplantı”, “şiddet” kelimeleri ile eş anlamlı olarak açık kaynaklardan elde edilmiştir. Ayrıca yazım yanlış bulunan “obartı” kelimesi ile de eş anlam ilişkisindedir. İlaveten “mübalağa” kelimesi Türkçede “abartı” kelimesinin eş anlamlısı olmakla beraber, “mübalâğa” ve “mübâlâğa” gibi hatalı yazımları da ağda eş anlam bağına sahiptir. Yapılan deneysel çalışmalar sonucunda zayıf bağlantıların eş bağlantı türlerinde olduğu, eş bağlantılarda da BG değeri 0.2’nin altındakilerde bu hataların yoğunlaştığı gözlemlenmiştir. Bu doğrultuda, eş bağlantılar için eşik değeri 0.2’nin altında olan kenarlar koparılmıştır. Diğer yandan, Bölüm 3.3.2.1’de de ifade edildiği gibi, ZIT-ANLAM derleminin daha düşük hacimli olmasının da etkisiyle, yaygın ve eşik değeri ile açıklanabilecek hatalı eşleşmeler zıt bağlantılarda gözlemlenmemiştir.

Şekil 3.10’da “abartı” kelimesinin eş anlam ilişki ağı görülmektedir. Düz ve kalın çizgiler BG değeri 0.2 ve üzerindeki güçlü bağları, kesikli çizgiler de 0.2’nin altında olan zayıf bağları temsil etmektedir.



Şekil 3.10 “abartı” kelimesinin eş anlam ilişkisi ağı.

Kesikli çizgilerle gösterilen zayıf kenarlar koparılmıştır. Bu işlemle hem “abartı” kelimesinin gerçek eş anlam karşılıkları edildiği gibi, yukarıda da bahsedilen yazım yanlışına sahip terimlerle olan bağlantıları da koparılmıştır. Sonuç olarak, “abartı” kelimesinin eş ve zıt ilişki ağı, bağ gücü değerleri ile birlikte Şekil 3.11’de sunulmuştur. Kırmızı renkle üçgen şekliyle temsil edilenler “abartı” kelimesinin zıt anlam karşılıklarını gösterirken, diğerleri eş anlamlı olduğu terimleri göstermektedir. Ayrıca kenarlar üzerinde de düğümler arasındaki bağın gücüne yer verilmiştir.



Şekil 3.11 “abartı” kelimesinin filtreleme sonrası eş ve zıt ilişkiler ağı.

Gelinen aşamada çizgede, SWNetTR-PLUS sözlüğünden gelen, ton ve polarite değeri belirli olan terimlerle hiçbir bağlantısı olmayan düğümler ortaya çıkmıştır. Bu terimler, bir sonraki bölümde ele alınan ton yayılımı ile bir değer alamayacaklarından dolayı ağdan filtrelenerek atılmışlardır. Sonuç olarak ağın başlangıcı ve bu süreç sonundaki yapısal bilgileri Çizelge 3.14’de sunulmuştur.

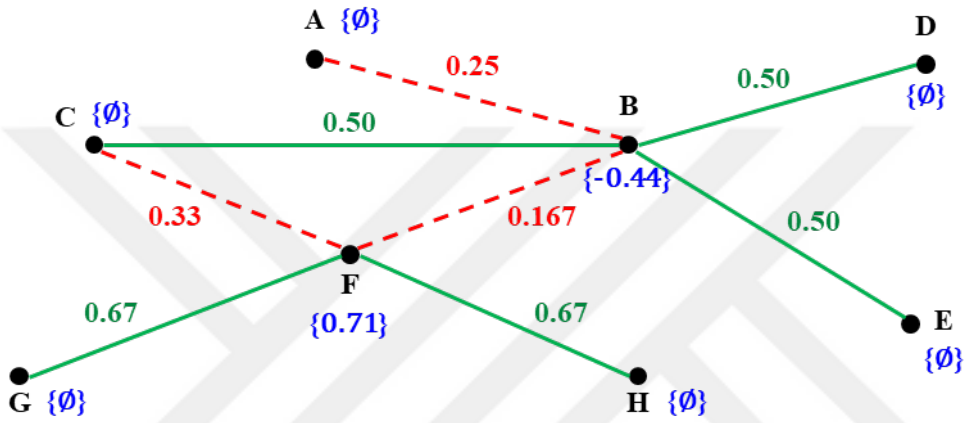
Çizelge 3.14 Çizgenin yapısal bilgileri.

	Başlangıç	Zayıf Bağların Koparılması Sonrası
Düğüm Sayısı	79326	63979
Kenar Sayısı	471258	210024

3.3.2.5. Ton Yayılımı

Çizgeye, dolayısı ile de duygu sözlüğüne yeni eklenen terimlerin ton ve polarite değerleri henüz bilinmemektedir. Şekil 3.12’de bu durum bir örnek üzerinde görselleştirilmiştir. Çizgenin kenarları üzerinde bağ gücü değerleri verilmiş ve eş anlam ilişkileri yeşil düz çizgiyle, zıt anlam ilişkileri de kırmızı kesik çizgiyle temsil

edilmiştir. B ve F düğümleri ton değerine sahip terimler olup, mavi renkle küme parantezi içerisinde gösterilmiştir. B düğümü negatif tonda, F düğümü ise pozitif tonda bir terim olup, B ve F düğümleri zıt anlam ilişkisine sahiptir. Bunun yanında, ağa yeni dahil olan A, C, D, E, G ve H terimlerinin ton değerleri yoktur ve mavi renkte küme parantezi içerisinde boş küme sembolü ile temsil edilmiştir. Sıradaki hedef, bu düğümlerin ton değerlerinin hesaplanmasıdır. Polarite değerleri ise, hesaplanan ton değerlerinin negatif/pozitif yönüne göre, -1 veya +1 şeklinde belirlenecektir.



Şekil 3.12 Örnek çizge: Ton Yayılımı öncesi.

Bu yeni terimler, ton değerine sahip olan komşularından bağ gücü ağırlıklarına göre ton değerlerine sahip olacaklardır. Aynı zamanda ton değerine zaten sahip olan düğümlerin de ton değerleri güncellenerek daha kararlı hale getirilecektir. Çalışmamız kapsamında geliştirilen bu yöntem “Ton Yayılımı” olarak isimlendirilmiş olup, Eşitlik (5)’deki matematiksel tanımlamaya sahiptir.

$$T'_u = \alpha \cdot T_u + \beta \cdot \sum_{v \in N_u} \left[S_{u-v} \frac{T_v \cdot BG_{u-v}}{\sum_{v \in N_u} BG_{u-v}} \right] \quad (5)$$

u : Üzerinde çalışılan düğüm

T_u : Hesaplama öncesi u teriminin ton değeri

T'_u : Hesaplama sonrası u teriminin bulunan/güncellenen değeri

N_u : u düğümünün ton değerine sahip olan komşular kümesi

T_v : u düğümüne komşu olan v düğümünün ton değeri

BG_{u-v} : u ile v düğümleri arasındaki kenarın Bağ Gücü değeri

$$S_{(u,v)} = \begin{cases} +1 & u \text{ ile } v \text{ düğümleri eş anlam ilişkisine sahip ise} \\ -1 & u \text{ ile } v \text{ düğümleri zıt anlam ilişkisine sahip ise} \end{cases}$$

İteratif olarak çalıştırılan Eşitlik (5), temel olarak iki ana bölümden oluşmaktadır. Alfa katsayısıyla çarpılan birinci kısım düğümün mevcut değerini, beta katsayısıyla çarpılan ikinci kısım ise düğümün komşularından hesaplanan ton değerini belirlemektedir. Alfa ve beta değerleri düğümlerin iterasyonlar esnasında ton değerlerinin yakınsama hızını kontrol etmektedir. Yapılan deneysel çalışmalarda en uygun değerlerin alfa=0.95 ve beta=0.05 olduğu tespit edilmiştir. Alfa değeri daha yüksek olduğunda, düğüm kendi ton değerine sıkı sıkıya bağlı kalmakta ve komşularından neredeyse hiç etkilenmemektedir. Alfa değeri daha düşük seçildiğinde ise, birkaç iterasyon sonrasında tüm komşu düğümler ortalama bir değerde buluşarak orijinal değerlerinden fazlasıyla uzaklaşabilmektedirler.

Ton değeri belirleme ve güncelleme sisteminde, u düğümü ve komşularının özelliklerinden dolayı 4 farklı senaryo ortaya çıkmıştır. Her bir senaryoya göre kullanılan alfa ve beta katsayıları Çizelge 3.15’de sunulmuştur.

Çizelge 3.15 u düğümü ve komşularına göre olası senaryolar ve katsayılar.

Senaryo	u Ton Değeri	N_u Kümesinde Eleman	Katsayılar
1	Yok	Yok	$\alpha = 0 \quad \beta = 0$
2	Yok	Var	$\alpha = 0 \quad \beta = 1$
3	Var	Yok	$\alpha = 1 \quad \beta = 0$
4	Var	Var	$\alpha = 0.95 \quad \beta = 0.05$

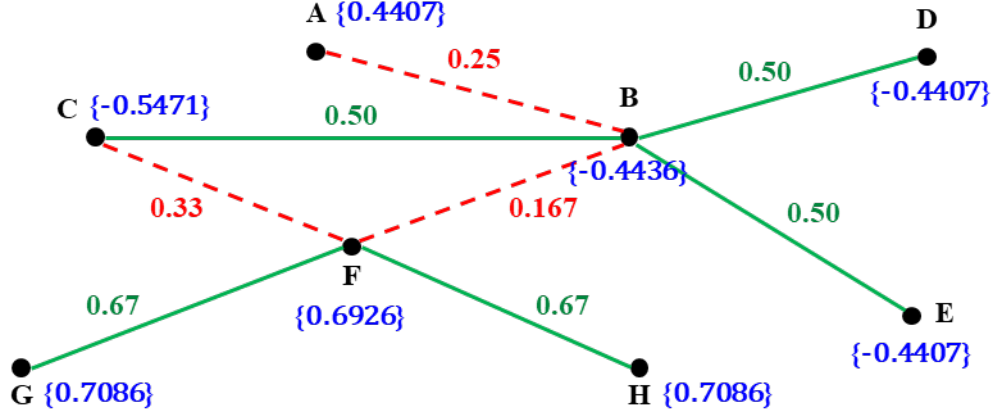
Senaryo-1’de, u düğümü ağa yeni dahil olmuş bir kelimedir ve ton değerine sahip komşu kümesi boş kümedir. Yani, u düğümünün değer alabileceği hiçbir komşusu yoktur. Bu nedenle alfa ve beta katsayılarına sıfır değeri verilerek, u düğümünün bir iterasyon sonucundaki yeni ton değeri sıfır olarak hesaplanır.

Senaryo-2’de, u düğümü Senaryo-1’deki gibi ağa yeni dahil olmuş, ton değeri bulunmayan bir kelimedir. Ancak ton değerine sahip komşusu/komşuları vardır. Bu durumda alfa katsayısına sıfır değeri verilirken, beta katsayısına 1 değeri verilerek, u düğümünün komşularından gelen ton değerini olduğu gibi kabul etmesi sağlanmaktadır.

Senaryo-3’te, u düğümünün ton değeri mevcuttur, ancak ton değerine sahip hiçbir komşusu yoktur. Bu durumda, alfa katsayısına 1 değeri verilerek, hesaplama sürecinde kendi değerini koruması sağlanmaktadır.

Senaryo-4’de, u düğümünün hem kendi ton değeri vardır, hem de ton değerine sahip komşular kümesinde kelime/kelimeler mevcuttur. Bu durumda da, u düğümü kendi ton değerini %95 oranında korurken, komşularından da %5 oranında etkilenmekte, yani değerini güncellemektedir.

Tüm bu hesaplama yaklaşımı, iteratif bir şekilde sürdürülmektedir. Geline nokta iterasyonun hangi noktada sonlanacağı önem arz etmektedir. Çok yüksek sayıda bir iterasyon, ağdaki bütün düğümlerin ortalama bir ton değerinde buluşmasına neden olmaktadır. Buna karşın, iterasyon sayısı çok küçük tutulduğunda da, ağda hesaplama sürecine dahil edilemeyen düğümler kalacaktır. Deneysel çalışmalar sonucunda, iterasyonun u düğümünün içinde bulunduğu bağlı-bileşenin (*connected-component*) çapı (en kısa yolların en uzununu) kadar yapılmasına karar verilmiştir. Böylece, bağlı-bileşen içindeki en uzak iki düğümün birbirlerinin ton değerine göreceli olarak etki etmesi garanti edilmiş, hem de aşırı iterasyonla düğümlerin gerçek ton değerlerinden uzaklaşması önlenmiştir. Şekil 3.12’de gösterilen örnek ağın, ton yayılımı uygulanması sonucundaki son durumu Şekil 3.13’te sunulmuştur.



Şekil 3.13 Örnek çizge: Ton Yayılımı sonrası.

Bu örnek çizgede, ağın çapı 3 olduğundan, Ton Yayılımı algoritması 3 iterasyon çalıştırılmıştır. Başlangıçta ton değerine sahip olan B ve F düğümlerinin ton değerleri güncellenirken, diğer düğümlere ise ton değeri hesaplanarak atanmıştır.

3.3.2.6. Ton Değeri Alamayanlar Problemi

Gelinen aşamada, sözlüğe yeni katılan terimlere ton değeri atanmış, sözlükte halihazırda mevcut olanların da ton değeri güncellenerek kararlı hale getirilmiştir. Ancak, ağdaki herhangi bir ton değerli düğümlle bağı olmayan bağlı-bileşenler herhangi bir polarite değerine sahip olamamışlardır. Her ne kadar bu terimler sözlüğe kazandırılmış yeni sözcükler gibi görünse de, duygu sözlüğünün yapısı gereği ton değerine sahip değildir. Ton değeri alamayanlar (*coldstart*) problemi olarak isimlendirdiğimiz bu sorun, bu bağlı-bileşenlerin filtrelenmesi ile çözülmüştür.

Bu aşamanın sonunda, duygu sözlüğü 49241 terimli, ton ve polarite değerine sahip bir yapıya kavuşturulmuştur.

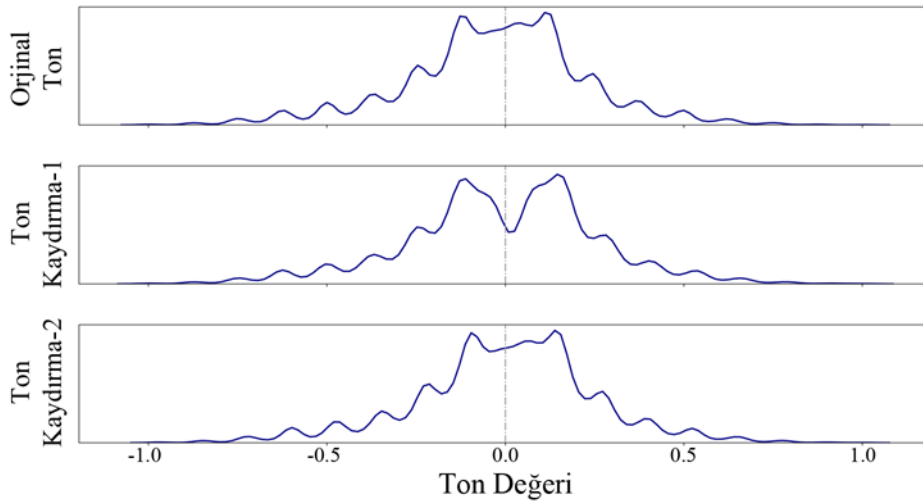
3.3.2.7. Negatif Yanlılık Problemi ve Ton Değeri Kaydırma

Mevcut durumda, geliştirilen sözlüğün pozitif ve negatif terim istatistiği Çizelge 3.16’da sunulmuştur.

Çizelge 3.16 Pozitif ve negatif terim istatistiği.

	Pozitif	Negatif	Toplam
Terim Sayısı	23603	25638	49241
Yüzde Oran	%48	%52	%100
Ton Değerleri Ortalaması	0.17	-0.21	-0.026

Çizelge 3.16 incelendiğinde, sözlükteki pozitif ve negatif terim sayılarında büyük oranda bir denge söz konusudur. Ancak ton değerleri ortalamasına göre, negatif terimlerin ton değerleri daha baskındır. Bu durum, duygu analizinde negatif yönde bir yanlılığa neden olacaktır. Çözüm önerisi olarak, ton kaydırma sistemi kullanılmış ve iki farklı yaklaşımla gerçekleştirilmiştir. Hem ton kaydırma öncesi, hem de ton kaydırma yaklaşımlarının sonuçları performans değerlendirme sürecinde karşılaştırılmıştır. Şekil 3.14’te, ton kaydırma yaklaşımları bir arada görselleştirilmiştir.



Şekil 3.14 Ton kaydırma yaklaşımlarının sonucundaki değişim grafiği.

Ton-Kaydırma-1 yaklaşımı, pozitif terimlerin ton ortalamasının, negatif terimlerin ton ortalamasına eşitlenmesidir. Böylece Çizelge 3.16’da görünen 0.17 pozitif ortalama değeri, 0.21’e çıkarılmıştır. Ancak negatif terimlerin ton değerlerinde herhangi bir değişiklik yapılmamıştır.

Ton-Kaydırma-2 yaklaşımı ise, sözlükteki bütün terimlerin ton değerlerinin ortalamasının sıfıra eşitlenecek şekilde ton değerlerinin güncellenmesidir. Bu sistematikte, ton değeri sıfıra yakın bazı negatif terimler, güncelleme sonrası pozitif ton değerine sahip olmaktadır. Her iki ton kaydırma işleminin sonrasındaki değişim Şekil 3.14’de de gözlemlenmektedir.

3.3.3. Performans Değerlendirmesi

Faz-2’nin sonunda, Bölüm 3.3.2’de sunulan metodolojik yaklaşımlarla 49K kapasiteye ulaşan ve SWNetTR++ olarak isimlendirdiğimiz Genel Amaçlı Türkçe Duygu Sözlüğü geliştirilmiştir. Geline aşamada terimler Ton Kaydırma Öncesi, Ton-Kaydırma-1 ve Ton-Kaydırma-2 süreçleri sonrasında olmak üzere 3 farklı ton değerine sahiptir. Bu bölümde amaçlanan, SWNetTR++’ın performansının bütün yönleriyle ölçülmesidir.

Performans ölçüm süreci Bölüm 3.2.3’te sunulan SWNetTR-PLUS’ın performans ölçümüyle aynı ilkeler doğrultusunda, MLTC-500 ve MLTC-353 test veri setleri üzerinde yapılmıştır. SWNetTR++’da yer alan terimlerin ton değerleri Ton Kaydırma Öncesi, Ton-Kaydırma-1 ve Ton-Kaydırma-2 süreçleri sonrasındaki değerler olmak üzere 3 farklı boyutta ele alınmış ve sonuçlar değerlendirilmiştir. Polarite değerleri, ton değerleri doğrultusunda tespit edilmiştir.

Çizelge 3.17 SWNetTR++ performans değerlendirme sonuçları (MLTC-500).

		MLTC-500	
		Polarite Bazlı	Ton Bazlı
GDELТ		%78.80	
SWNetTR		%60.60	%64.20
SWNetTR-PLUS		%72.20	%75.20
SWNetTR++	Ton Kaydırma Öncesi	%75.00	%73.80
	Ton-Kaydırma-1	%75.00	%80.40
	Ton-Kaydırma-2	%73.60	%77.80

Çizelge 3.17’de çalışmanın başlangıcından bu yana çalışma kapsamındaki mevcut bütün duygu sözlüklerinin performanslarının, MLTC-500 test veri seti üzerindeki ölçüm sonuçları sunulmuştur. Daha önce de belirtildiği gibi, GDELТ projesi kapsamında Türkçe haberler önce İngilizceye çevrilmekte ve ardından Çizelge 3.2’de sunulan gelişmiş İngilizce dil kaynakları kullanılarak analize tabi tutulmaktadır. GDELТ’in MLTC-500 Türkçe haber test veri seti üzerindeki başarısı %78.80 olarak ölçülmüştür. Bu bizim çalışmamızda ulaşmak ve geçmek istediğimiz ilk kontrol noktasıdır. Faz-1’de geliştirilen ve ilk sürüm olarak lanse ettiğimiz SWNetTR-PLUS Türkçe duygu sözlüğü ile %75.20 ile GDELТ’in başarısına yaklaşılmış olsa da geçilememiştir. Faz-2’deki yaklaşımlar sonucunda oluşturulan hem kapasitesi artırılan hem de ton değerleri daha kararlı hale getirilen SWNetTR++ Türkçe duygu sözlüğü ile başarımlar ton bazlı olarak %80.40 seviyesine çıkartılarak, GDELТ’in başarımlar performansı geçilmiştir. Bölüm 3.3.2.7’deki ton kaydırma yaklaşımlarından, pozitif terimlerin ton değerlerinin negatif terimlerin ton değerlerine eşitlenmesi ile gerçekleştirilen Ton-Kaydırma-1 yaklaşımı daha iyi sonuç vermiştir. İlaveten, terimlerin ton değerleri üzerinde duygu skoru belirlemeye dayalı ton bazlı performans değerlendirmesinin, terimlerin polarite değerleri ile duygu yönü belirlemeye dayalı polarite bazlı performans değerlendirmesinden daha iyi sonuçlar verdiği de tespit edilmiştir.

Çizelge 3.18’de ise performans ölçümünün, bütün değerlendiricilerin üzerinde hem fikir olduğu Türkçe haberlerden oluşan MLTC-353 test veri seti üzerindeki sonuçları sunulmuştur. Sonuçlar incelendiğinde, SWNetTR++ ile ulaşımlar başarımın GDELТ’i

yakaladığı ve SWNetTR-PLUS'ı geçtiği görülmektedir. Ton-Kaydırma-1 yaklaşımının ve ton bazlı değerlendirme için en iyi sonucu verdiği tespit edilmiştir.

Çizelge 3.18 SWNetTR++ performans değerlendirme sonuçları (MLTC-353).

		MLTC-353	
		Polarite Bazlı	Ton Bazlı
GDELT		%84.99	
SWNetTR		%62.89	%67.99
SWNetTR-PLUS		%75.07	%78.75
SWNetTR++	Ton Kaydırma Öncesi	%78.75	%77.05
	Ton-Kaydırma-1	%78.75	%84.42
	Ton-Kaydırma-2	%79.32	%83.85

İki farklı test derlemi üzerindeki performans analizleri kıyaslandığında ise, bütün değerlendiricilerin üzerinde hem fikir olduğu haberlerden oluşan MLTC-353 derlemi, bütün karşılaştırma kriterlerinde MLTC-500 derleminden daha yüksek performans göstermiştir.

4. DUYGU HARİTASI MODELİ

Bölüm 2.1.2’de de belirtildiği gibi, literatürde duygu analizi en geniş anlamda metin sınıflandırma ve/veya regresyon problemi olarak ele alınmaktadır. Doküman ve cümle seviyesindeki duygu analizlerinin sonucunda, ilgili içerik analiz edilerek yazarın duygusu genellikle ikili, üçlü veya duygu boyutları genişletilerek daha çok seviyeli sınıflandırmaya tabi tutulmaktadır. Bu çalışmaların bir ileri seviyesinde de, duygunun polaritesi ile beraber, duygunun şiddeti de sayısal olarak çıkarımsanmaktadır.

Yapılan çalışmalarda içerik analiz edilerek, yazarın duygu durumu mutlak bir yargıyla sunulmaktadır. Bu yaklaşımın tek başına yeterli olamayacağı, içerikteki duygu dalgalanmalarının göz ardı edildiği, bu bölüm kapsamındaki temel eleştirimizdir. Çalışmamızda bu problem alanına ve ihtiyaca yönelik olarak, duygu analizini keşifsel veri analizi bakışıyla ele alan ve görselleştiren bir yaklaşım önerilmektedir. Duygu Haritası Modeli (DHM) olarak isimlendirdiğimiz bu yaklaşımda içerik; cümle, paragraf veya zaman skalası bakış açılarıyla ele alınabilmekte olup, sözlük temelli duygu analizi sınıfına girmektedir.

Analize tabi tutulacak olan içerik metni, cümle seviyesinde veya paragraf seviyesinde ele alınabilir. Cümle bazında, cümlelerdeki terimlerin duygu sözlüğündeki ton değerleri toplamı, cümlenin duygu skorunu ifade eder. Benzer şekilde, paragraf seviyesinde duygu haritası oluşturmak için her bir paragrafın toplam duygu skoru sözlük sayesinde belirlenir. Seviyesine göre, duygu skorları bir çizgi grafiği üzerinde görselleştirilerek, dokümanın duygu salınım grafiği elde edilir.

Örnek olarak Çizelge 4.1’de yer alan Haber-1 [145] isimli içeriğin cümlelerine ayrılmış ve numaralandırılmış cümle listesi Çizelge 4.2’de, cümle seviyesi duygu salınım grafiği de Şekil 4.1’de sunulmuştur. Bu bölüm içerisinde verilen örneklerdeki cümlelerin ton değerlerinin tespitinde, geliştirdiğimiz SWNetTR++ Genel Amaçlı Türkçe Duygu Sözlüğü kullanılmıştır.

Çizelge 4.1 Haber-1 içeriği.

ATM'YE SAHTE PARA YATIRIP GERÇEK PARA ÇEKEN KALPAZANLAR YAKALANDI: 'ENFLASYONU DÜŞÜRECEKTİK, BIRAKMADINIZ'.

Mersin'de, küçük miktardaki paraların belli yerlerini 100 ve 200 TL'lik banknotlara yamalayıp ATM'ye yatıran ve sonra farklı ATM'den gerçek para çeken 4 kişi yakalandı. Sevk edildikleri adliyenin çıkışında görüntülenen şahıslardan bir tanesi, "Kolay değil 3 yılımızı verdik, enflasyonu düşürecektik bırakmadınız" dedi.

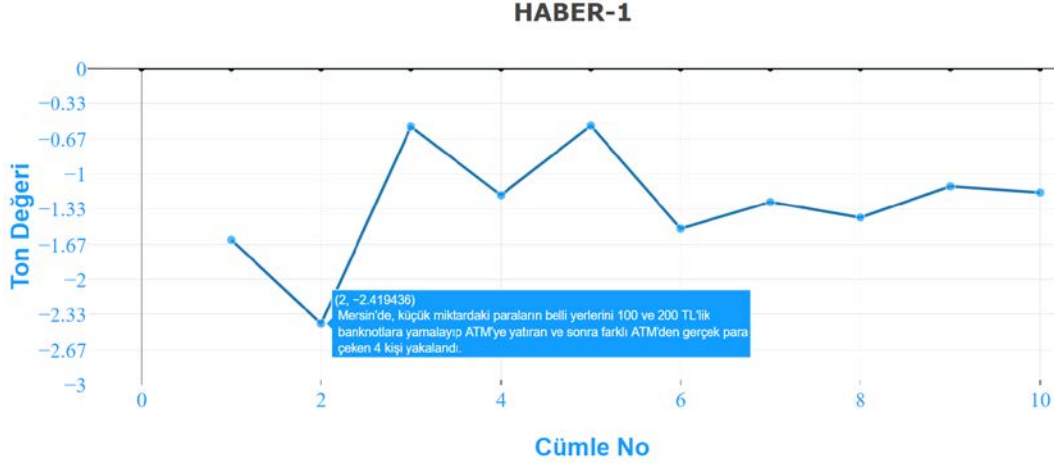
AA'nın aktardığı bilgiye göre, Mersin İl Emniyet Müdürlüğü Kaçakçılık ve Organize Suçlarla Mücadele (KOM) Şube Müdürlüğü ekipleri gelen bir ihbar üzerine K.G. isimli bir şahsı takibe aldı. 10 liralık banknotların parçalarını 200 liraya yamalamışlar.

Polis ekipleri, K.G.'nin 10 liralık banknotların belirli parçalarını kullanarak hazırladıkları sahte 200 liralı ATM'lere yatırıp bankaları dolandıran bir ekibin parçası olduğunu tespit etti. Yapılan operasyon sonucunda K.G. ile birlikte 4 kişi gözaltına alınırken, adreslerde 68 adet 100 liralık sahte para bulundu. ATM'lere sahte paraları yatıran kalpazanların Mersin, Gaziantep, Amasya ve Kayseri'deki tanıdıklarına üzerinden orijinal paraları çektiği ortaya çıktı.

K.G. ve S.O. tutuklanırken, diğer iki kişi ise adli kontrol şartıyla serbest bırakıldı. Çıkarıldığı mahkemece tutuklanan K.G., adliye çıkışında gazetecilere "Kolay değil ha 3 yılımızı verdik, enflasyonu düşürecektik bırakmadınız" dedi.

Çizelge 4.2 Haber-1 cümle listesi.

Cümle No	Cümle
1	ATM'YE SAHTE PARA YATIRIP GERÇEK PARA ÇEKEN KALPAZANLAR YAKALANDI: 'ENFLASYONU DÜŞÜRECEKTİK, BIRAKMADINIZ'.
2	Mersin'de, küçük miktardaki paraların belli yerlerini 100 ve 200 TL'lik banknotlara yamalayıp ATM'ye yatıran ve sonra farklı ATM'den gerçek para çeken 4 kişi yakalandı.
3	Sevk edildikleri adliyenin çıkışında görüntülenen şahıslardan bir tanesi, "Kolay değil 3 yılımızı verdik, enflasyonu düşürecektik bırakmadınız" dedi.
4	AA'nın aktardığı bilgiye göre, Mersin İl Emniyet Müdürlüğü Kaçakçılık ve Organize Suçlarla Mücadele (KOM) Şube Müdürlüğü ekipleri gelen bir ihbar üzerine K.G. isimli bir şahsı takibe aldı.
5	10 liralık banknotların parçalarını 200 liraya yamalamışlar.
6	Polis ekipleri, K.G.'nin 10 liralık banknotların belirli parçalarını kullanarak hazırladıkları sahte 200 liralı ATM'lere yatırıp bankaları dolandıran bir ekibin parçası olduğunu tespit etti.
7	Yapılan operasyon sonucunda K.G. ile birlikte 4 kişi gözaltına alınırken, adreslerde 68 adet 100 liralık sahte para bulundu.
8	ATM'lere sahte paraları yatıran kalpazanların Mersin, Gaziantep, Amasya ve Kayseri'deki tanıdıklarına üzerinden orijinal paraları çektiği ortaya çıktı.
9	K.G. ve S.O. tutuklanırken, diğer iki kişi ise adli kontrol şartıyla serbest bırakıldı.
10	Çıkarıldığı mahkemece tutuklanan K.G., adliye çıkışında gazetecilere "Kolay değil ha 3 yılımızı verdik, enflasyonu düşürecektik bırakmadınız" dedi.



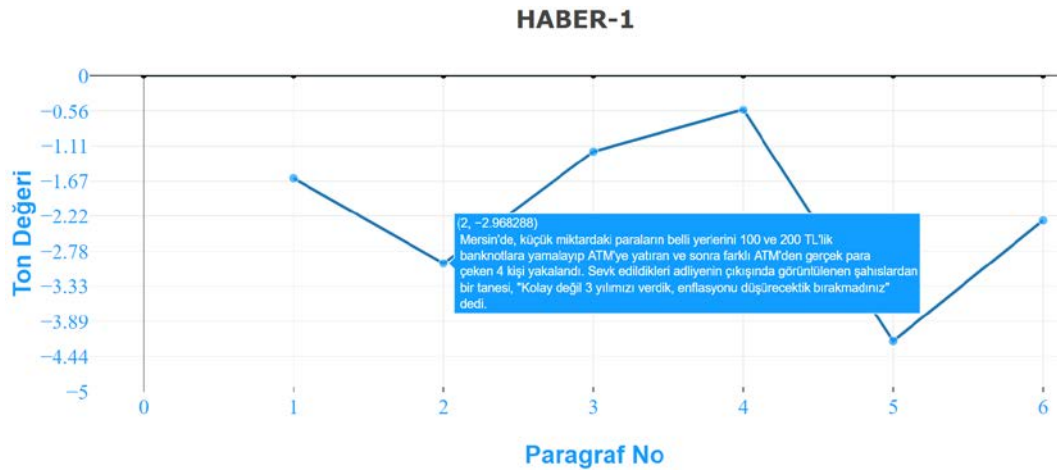
Şekil 4.1 Haber-1 cümle seviyesi duygu salınım grafiği.

Şekil 4.1’de, x eksenini içerikteki cümle sırasını gösterirken, y eksenini de ton değerini göstermektedir. İlave 2 numaralı cümleye de şekilde yer verilmiştir. Grafik incelendiğinde bütün cümlelerin duygu skorunun sıfır ekseninin altında salınım yaptığı, 2 numaralı cümlelerin en yüksek negatif ton değerine sahip olduğu gözlemlenmektedir. Her ne kadar dokümanın negatif polariteye sahip olduğu grafikten anlaşılıyor olsa da, DHM modellerinin amacı içeriği tek bir polariteyle etiketlemek olmadığından, metindeki duygu salınımı görselleştirilerek, analiz takip eden bölüm başlıklarında açıklanan kural tabanlı yaklaşımlarla ele alınacaktır.

Çizelge 4.1’deki Haber-1’in paragraflarına ayrılmış numaralı paragraf listesi Çizelge 4.3’te, paragraf seviyesi duygu salınım grafiği de Şekil 4.2’de sunulmuştur.

Çizelge 4.3 Haber-1 paragraf listesi.

Paragraf No	Paragraf
1	ATM'YE SAHTE PARA YATIRIP GERÇEK PARA ÇEKEN KALPAZANLAR YAKALANDI: 'ENFLASYONU DÜŞÜRECEKTİK, BIRAKMADINIZ'.
2	Mersin'de, küçük miktardaki paraların belli yerlerini 100 ve 200 TL'lik banknotlara yamalayıp ATM'ye yatan ve sonra farklı ATM'den gerçek para çeken 4 kişi yakalandı. Sevk edildikleri adliyenin çıkışında görüntülenen şahıslardan bir tanesi, "Kolay değil 3 yılımızı verdik, enflasyonu düşüreceklik bırakmadınız" dedi.
3	AA'nın aktardığı bilgiye göre, Mersin İl Emniyet Müdürlüğü Kaçakçılık ve Organize Suçlarla Mücadele (KOM) Şube Müdürlüğü ekipleri gelen bir ihbar üzerine K.G. isimli bir şahsı takibe aldı.
4	10 liralık banknotların parçalarını 200 liraya yamalamışlar.
5	Polis ekipleri, K.G.'nin 10 liralık banknotların belirli parçalarını kullanarak hazırladıkları sahte 200 liralı ATM'lere yatırıp bankaları dolandıran bir ekibin parçası olduğunu tespit etti. Yapılan operasyon sonucunda K.G. ile birlikte 4 kişi gözaltına alınırken, adreslerde 68 adet 100 liralık sahte para bulundu. ATM'lere sahte paraları yatan kalpazanların Mersin, Gaziantep, Amasya ve Kayseri'deki tanıdıklarına üzerinden orijinal paraları çektiği ortaya çıktı.
6	K.G. ve S.O. tutuklanırken, diğer iki kişi ise adli kontrol şartıyla serbest bırakıldı. Çıkarıldığı mahkemece tutuklanan K.G., adliye çıkışında gazetecilere "Kolay değil ha 3 yılımızı verdik, enflasyonu düşüreceklik bırakmadınız" dedi.



Şekil 4.2 Haber-1 paragraf seviyesi duygu salınım grafiği.

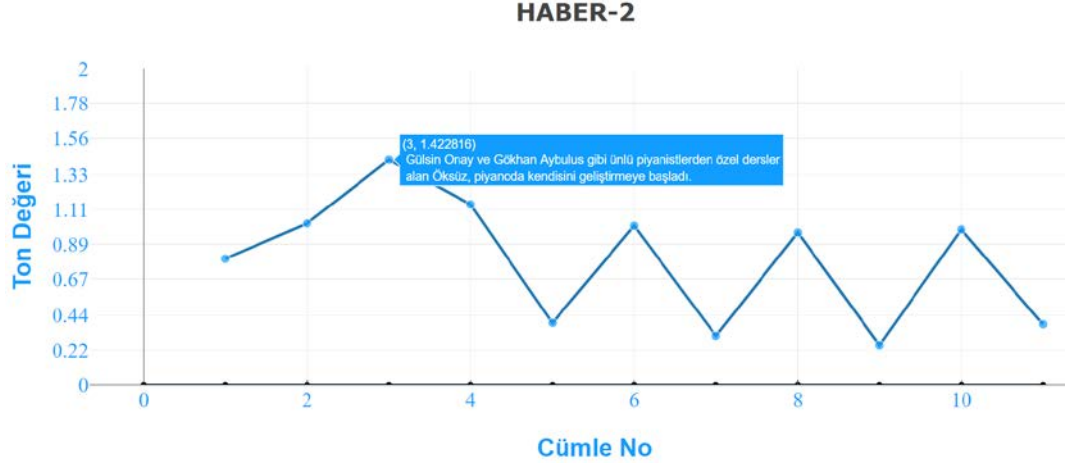
Şekil 4.2 paragraf seviyesi duygu salınım grafiği incelendiğinde, bütün paragrafların sıfır eksenin altında salınım yaptığı, 5 numaralı paragrafın duygu tonunun en yüksek negatif ton değerine sahip olduğu gözlemlenmektedir.

Çevrimiçi haber medyasında yer alan metinler incelendiğinde, paragraf ayrımının cümle ayrımına göre daha özensiz yapıldığı gözlemlenmektedir. Editörlerce düzenlenen içeriklerde, tek bir paragraftan oluşan metinler ile her bir cümlenin bir paragraf halinde düzenlendiği metinlere rastlanılmaktadır. Paragrafın yapılandırılma şekli, elde edilen duygu salınım grafiğini de doğrudan etkilediğinden gerek duygu salınım grafiği, gerekse sonraki bölümde açıklanan DHM’de metinler, cümle seviyesinde ele alınmıştır.

İkinci örnek olarak Çizelge 4.4’de Haber-2 [146] metninin cümle sırası, Şekil 4.3’te de cümle seviyesi duygu salınım grafiği sunulmuştur.

Çizelge 4.4 Haber-2 cümle listesi.

Cümle No	Cümle
1	Piyanonun Parlak Çocuğu: Dünya Birinciliği Gururu Yaşayan 9 Yaşındaki Nisan, İtalya'daki Yarışmaya Hazırlanıyor.
2	Müziğe, 6 yaşındayken org ile başlayan 9 yaşındaki Nisan Öksüz'ün yeteneği öğretmenleri tarafından keşfedilince, anne babası kızlarının müziğe olan ilgisinin sürmesi için ona piyano aldı.
3	Gülsin Onay ve Gökhan Aybulus gibi ünlü piyanistlerden özel dersler alan Öksüz, piyanoda kendisini geliştirmeye başladı.
4	Geçen sene İtalya'da düzenlenen yarışmada birincilik elde ederek Türkiye'yi gururlandıran Nisan, gelecek ay yine İtalya'da düzenlenecek Apulia Metropolitan Music Contest yarışmasına katılacak.
5	Nisan, bu yarışmada Stephen Heller, Frederic Chopin, Johann Sebastian Bach, Sergey Prokofyev gibi ünlü isimlerin eserlerini çalacak.
6	Genç piyanist ise, geçen sene birincilik elde ettiği yarışmaya giderken, yanında kendisine uğur getireceğine inandığı şeyleri götürceğini söyledi.
7	Anneannesinin yarışma için özel diktiği elbiseyi giyeceğini, kulağına sol anahtar şeklindeki küpeyi takacağını dile getiren Öksüz, yine öğretmenin verdiği anahtarlığı yanına alacağını kaydetti.
8	Anne Figen Öksüz, kızının başarısı ile ilgili olarak "Piyano onun arkadaşı, oyuncuğu, her şeyi.
9	Eve girer girmez piyanonun başına geçer.
10	Beste yapmak ister, sevdiği şarkıları piyanosunda çalmaya çalışır.
11	Benim en sevdiğim şarkıları açar" dedi.



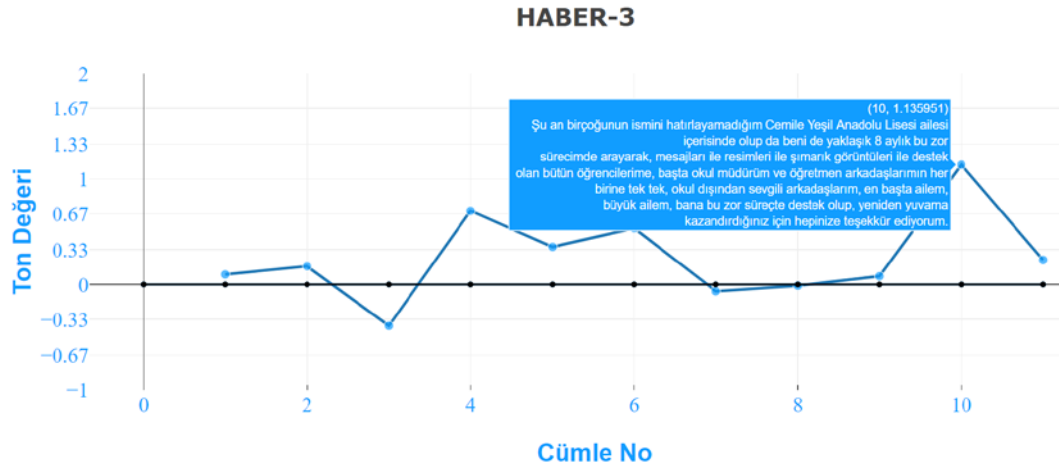
Şekil 4.3 Haber-2 cümle seviyesi duygu salınım grafiği.

Şekil 4.3 incelendiğinde, Haber-2'nin pozitif bölgede duygu salınımı yaptığı, en yüksek duygu skoruna 3 nolu cümlenin sahip olduğu görülmektedir.

Birinci örnekteki Haber-1 tamamen negatif bölgede, ikinci örnekteki Haber-2 ise tamamen pozitif bölgede duygu salınımına sahiptir. İçeriklerin doğası gereği, bu duruma her metinde rastlamak mümkün değildir. Üçüncü örnek olarak cümle listesi Çizelge 4.5'de verilen Haber-3'te [147], kanser hastalığını yenerek okuluna geri dönen bir öğretmen konu edilmiştir. Şekil 4.4'deki duygu salınım grafiğinden de görüleceği gibi haber içeriği, hem pozitif hem de negatif bölgede duygu salınımı göstermektedir.

Çizelge 4.5 Haber-3 cümle listesi.

Cümle No	Cümle
1	KANSERİ YENEN VE AYLAR SONRA OKULUNA DÖNEN ÖĞRETMEN SEVGİ SELİYLE KARŞILANDI.
2	Tekirdağ'ın Çorlu ilçesi Cemile Yeşil Anadolu Lisesi'nde görev yapan ve kanser hastalığını yenerek 8 ay sonra okuluna dönen coğrafya öğretmeni Muharrem Poyrazoğlu, öğrencilerinin ve arkadaşlarının sevgi gösterileriyle karşılandı.
3	DHA'nın haberine göre, Çorlu Cemile Yeşil Anadolu Lisesi'nde coğrafya öğretmeni olan Muharrem Poyrazoğlu, mide kanseri hastalığı nedeniyle okula ara verip, öğrencilerinden uzak kaldı.
4	Yaklaşık 8 ay tedavi gören Poyrazoğlu, hastalığını yenip, sağlığına kavuşunca bugün aylar sonra ilk kez okuluna geldi.
5	ÖĞRENCİLER VE ÖĞRETMENLER, POYRAZOĞLU'NA SÜRPRİZ BİR KARŞILAMA HAZIRLADI.
6	Okulun merdivenlerinde toplanan öğrenciler, Poyrazoğlu'nun girmesiyle hazırladıkları balon ve konfetileri üzerine atıp, alkışlar, şarkılar ve sloganlarla karşıladı.
7	ÖĞRENCİLERİN COŞKULU VE SEVGİ DOLU KARŞILAMASI SIRASINDA DUYGULANAN MUHARREM POYRAZOĞLU, ÖĞRENCİLERİNE SARILDI, GÖZYAŞLARI DÖKTÜ.
8	Dakikalar süren karşılama sırasında duygulu anlar yaşayan öğretmen Muharrem Poyrazoğlu, evine geldiği için çok mutlu olduğunu söyledi ve şöyle devam etti.
9	Hayatıma giren çok sayıda arkadaşım oldu.
10	Şu an birçoğunun ismini hatırlayamadığım Cemile Yeşil Anadolu Lisesi ailesi içerisinde olup da beni de yaklaşık 8 aylık bu zor sürecimde arayarak, mesajları ile resimleri ile şımarık görüntüleri ile destek olan bütün öğrencilerime, başta okul müdürüm ve öğretmen arkadaşlarımla her birine tek tek, okul dışından sevgili arkadaşlarım, en başta ailem, büyük ailem, bana bu zor süreçte destek olup, yeniden yuvama kazandırdığınız için hepinize teşekkür ediyorum.
11	Sizleri çok seviyorum.



Şekil 4.4 Haber-3 cümle seviyesi duygu salınım grafiği.

Şekil 4.4 incelendiğinde, tek yönde bir duygu mevcut değildir. 3. cümle en negatif cümle iken, 10. cümle de en yüksek duygu tonuna sahiptir.

Yukarıda sunulan haber örneklerinin duygu salınım grafikleri incelendiğinde, grafiklerin y eksenleri ilgili dokümana özgüdür, öznelidir. Bir başka ifadeyle, içeriklerin duygu salınım grafikleri ortak bir değerlendirme alt yapısına sahip değildir. Bu noktada, duygu salınım grafiklerini ortak bir değerlendirme skalasına oturtarak anlamlandıran ve analizini mümkün kılan Duygu Haritası Modeli (DHM) geliştirilmiştir. Bu modelde içerikteki duygu salınımı süreç bakış açısıyla ele alınmaktadır. Sürecin kontrolünde ve yorumlanmasında, detaylarına takip eden bölüm başlıklarında yer verilen Shewhart Kontrol Diyagramı [148] istatistiksel modeli kullanılmıştır.

4.1. Shewhart Kontrol Diyagramı

İstatistik disiplininde küçük harfle ifade edilen (σ) sigma, veri setindeki standart sapmayı tanımlamak için kullanılan bir ölçü birimidir. Özellikle de süreçlerin istatistiksel kontrolünde önemli bir yere sahiptir. Sigma'nın büyüklüğü, değerler arasındaki farklılaşmayı ve uzaklaşmayı anlatırken, küçüklüğü de homojenlik düzeyini işaret eder [148]. Bu özelliği nedeniyle standart sapma, literatürde ve uygulamada süreçlerin kalite kontrolünde, süreçlerin izlenmesinde ve geri besleme mekanizmalarının devreye alınmasında sıklıkla kullanılan ölçü birimlerindendir.

Shewhart kontrol grafikleri, 1930’lu yıllarda Walter A. Shewhart tarafından geliştirilmiştir. Üretim ve kalite kontrol odaklı geliştirilen kontrol grafikleri, sürecin ölçülmesinde ve değerlendirilmesinde ekonomik ve etkili bir araçtır. Bir sürecin kontrolünde değişim limitlerinin belirlenmesi gereklidir ve bu limitler istatistiksel anlamdaki güven sınırlarını belirler. Sınırlar içerisindeki değişimler istatistik kanunlara uygundur ve tesadüfi değişimler olarak kabul edilirken, sınırların dışındaki değişimler anlamlı bir kaymayı işaret etmektedir [149].

Değişim limitlerinin belirlenmesinde, μ veri setinin ortalamasını, σ standart sapmasını göstermek üzere Eşitlik (6)’daki matematiksel ifade kullanılmaktadır.

$$\text{ÜKL} = \mu + k \cdot \sigma \quad (6)$$

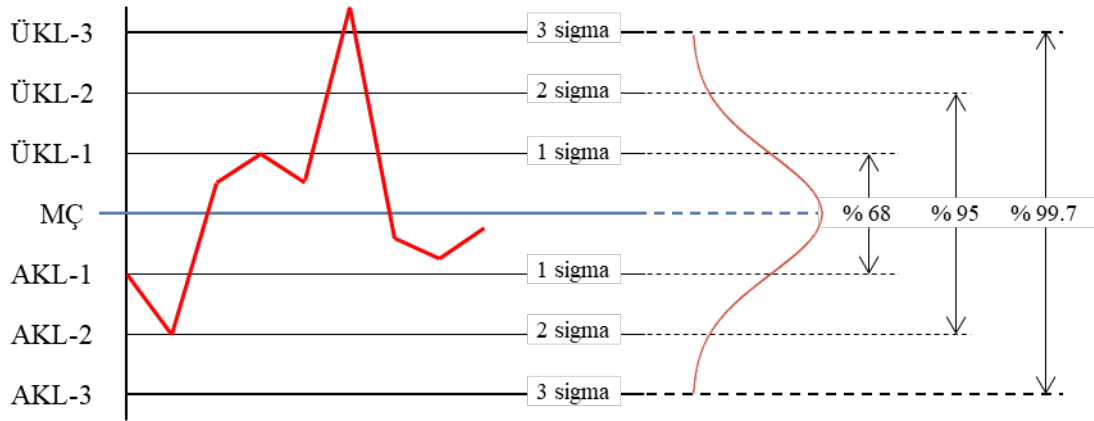
$$MÇ = \mu$$

$$AKL = \mu - k \cdot \sigma$$

Bu eşitlikte;

- ÜKL-k : Üst Kontrol Limit
- MÇ : Merkez Çizgi
- AKL-k : Alt Kontrol Limit
- k: Merkez çizgiden kaç sigma uzak olduğunu gösteren katsayı

Shewhart kontrol diyagramı limitleri, normal dağılım karşılıkları ile birlikte Şekil 4.5’de gösterilmiştir.



Şekil 4.5 Shewhart kontrol diyagramı limitleri.

Shewhart kontrol diyagramındaki örüntülerin analizine yönelik olarak The Western Electric [150] tarafından dört kural önerilmiştir. Bu kurallara ilave olarak araştırmacılar tarafından zaman içerisinde ilave kurallar da getirilmiştir [151]. Çalışmamız kapsamında da kullanılacak olan söz konusu dört kural Çizelge 4.6’da sunulmuştur.

Çizelge 4.6 Shewhart kontrol diyagram analiz kuralları.

	Kural
KURAL-1	9 ardışık noktadan en az 8 tanesinin (8/9), MÇ’nin bir tarafında bulunması.
KURAL-2	5 ardışık noktadan en az 4 tanesinin (4/5), MÇ’nin aynı tarafında ve 1 sigmadan daha uzakta bulunması.
KURAL-3	3 ardışık noktadan en az 2 tanesinin (2/3), MÇ’nin aynı tarafında ve 2 sigmadan daha uzakta bulunması.
KURAL-4	1 noktanın, merkez çizgiye uzaklığının 3 sigmadan büyük olması (üst ve alt limitlerin dışında olması).

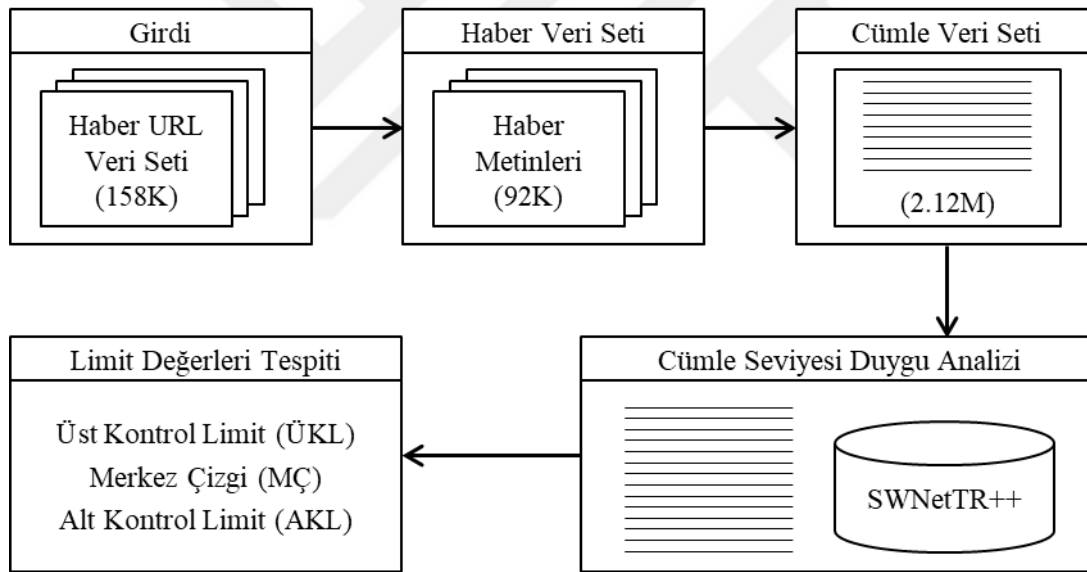
Kurallar kontrol grafiği üzerinde uygulanarak, kurala uyan noktalara Bildirim Noktası denilmektedir. Bu noktalardan hareketle değerlendirmeler şu şekilde yapılmaktadır:

- KURAL-1: Değişimin aynı bölgede ve istikrarlı olduğunu göstermektedir.
- KURAL-2: Tesadüften uzak küçük ölçekteki değişimi göstermektedir.
- KURAL-3: Tesadüften uzak orta ölçekteki değişimi göstermektedir.
- KURAL-4: Büyük bir sapmayı vurgulamaktadır.

İstatistiksel süreç analizini konu edinen Shewhart kontrol diyagramı ve analiz kuralları çerçevesinde geliştirilen Duygu Haritası Modeli, takip eden bölümde ele alınmıştır.

4.2. Duygu Haritası Modeli ve Shewhart Kontrol Diyagramı

Bir doküman, ilk cümlesinden son cümlesine kadar bir ifade aktarım sürecine sahiptir. Duygu analizinin cümleler bazında ele alınması ile doküman da bir süreç yapısına sahip olmaktadır. Cümleler bazında duygu skorları ile sayısallaştırılan içerik, Shewhart kontrol diyagramı ile ele alınarak, süreç izleme kapsamında DHM'nin geliştirilmesinde kullanılmıştır. Şekil 4.6'da DHM'de kullanılacak kontrol grafiğinin limit değerlerinin belirlenme süreci blok diyagramı sunulmuştur.



Şekil 4.6 Kontrol grafiği limit değerlerinin belirlenmesi süreci blok diyagramı.

Bölüm 4.1'de de belirtildiği gibi, kontrol grafikleri için ilk gereksinim kontrol limitlerinin belirlenmesidir. Sınır değerlerinin belirlenmesindeki veri seti ne kadar büyük olursa, standart sapma ve buna bağlı olarak elde edilen sınır değerleri de gerçek dünyayı o kadar iyi temsil edebilecektir. Bu doğrultuda, GDELT projesi veri kaynağından 158983 adet haberin URL bilgisi elde edilmiş ve bu haber kaynaklarına gidilerek haber metinleri elde edilmiştir. Ardından, bazı haber içeriklerinin çok kısa

olması nedeniyle, içeriğinde en az 5 cümle ve en az 5 paragraf içeren haber metinleri filtrelenerek seçilmiş, sonuçta 92150 adet haber metni derlemi oluşmuştur.

Sonraki aşamada, haber metinleri cümleler bazında ayrıştırılarak, her bir cümlenin duygu tonu SWNetTR++ Genel Amaçlı Türkçe Duygu Sözlüğü kullanılarak belirlenmiştir. Toplamda 2.12 milyon cümleden oluşan derlem, kontrol limitlerinin belirlenmesindeki veri setini oluşturmuştur. Veri setinin olabildiğince büyük olması ile belirlenecek limit değerlerinin gerçeğe yakın olması hedeflenmiştir. Eşitlik (6) kullanılarak Çizelge 4.7’de verilen limit değerleri elde edilmiştir.

Çizelge 4.7 Türkçe cümlelerin Shewhart kontrol diyagramı limit değerleri.

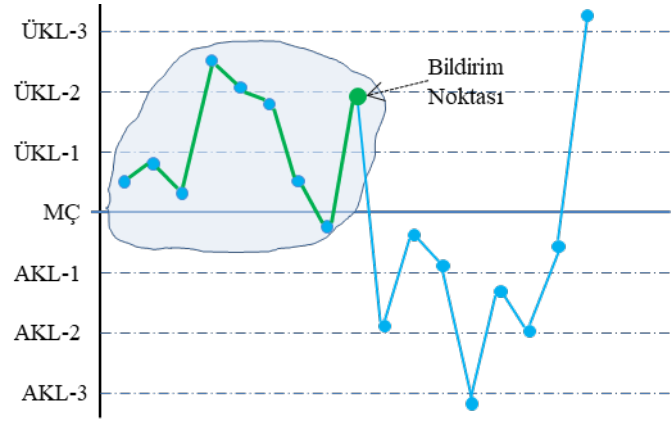
Limit	Değer
ÜKL-3	2.230570
ÜKL-2	1.462079
ÜKL-1	0.693588
MÇ	-0.074903
AKL-1	-0.843394
AKL-2	-1.611885
AKL-3	-2.380376

Shewhart kontrol grafiğinin analize yönelik Çizelge 4.6’da listelenen 4 temel kural, duygudaki salınımın analiz edilmesine imkan veren DHM’de uygulanmıştır. Her kural sonucunda elde edilen bilgilerin anlamlandırılması ve yorumlanabilmesi için, duygu analizi dünyasına özgü olarak Çizelge 4.7’deki kriterler kullanılmıştır. İlaveten, ilgili kuralı görsel olarak örneklendiren duygu haritalarına da yer verilmiştir. Duygu haritalarında, kuralın tespit edildiği noktalar “Bildirim Noktası” olarak isimlendirilmektedir. Bildirim noktaları, bu noktaların oluşumuna sebep olan cümlelerin bulunduğu bölgeye göre yeşil veya kırmızı renkle ve daha büyük olarak görselleştirilmiştir. İlaveten, bildirim noktalarının meydana gelmesine sebep olan cümleler arasındaki bağlantılar da polarite bölgesine göre yeşil veya kırmızı olarak renklendirilmiştir. Mavi renkteki çizgilerle, boyutu küçük olan düğüm noktaları

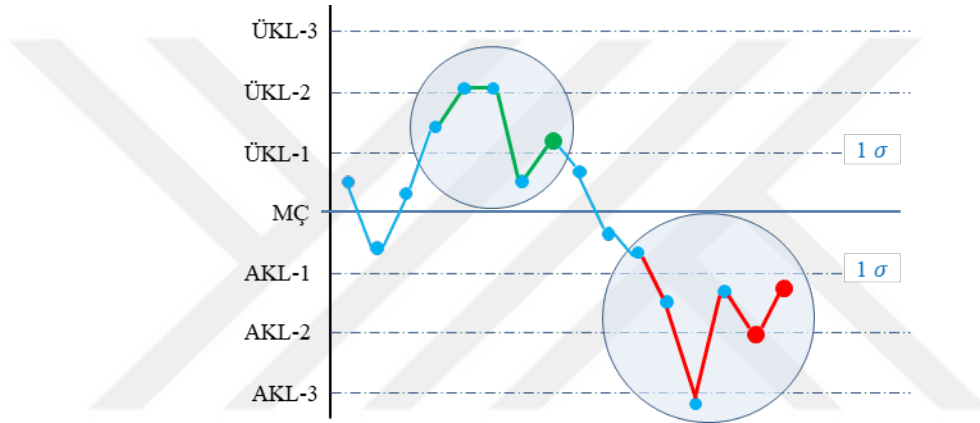
Shewhart kontrol diyagramı kurallarından etkilenmeyen, yani duygu salınımı itibariyle Shewhart perspektifinden bir tespiti uęramayan cümleleri tanımlamaktadır.

Çizelge 4.8 DHM kapsamında Shewhart kurallarının değeriendirilmesi.

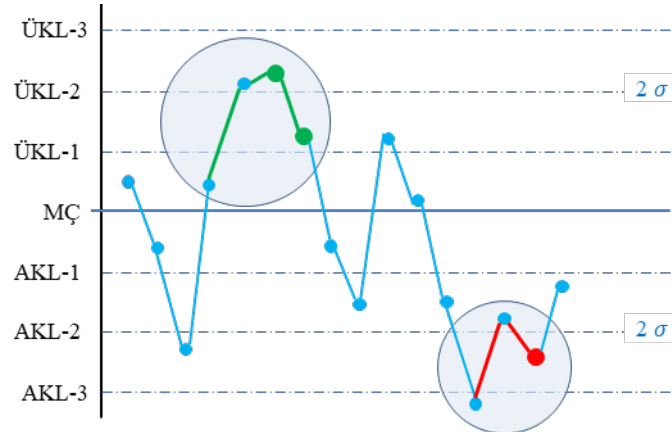
	Kural ve Deęerlendirme
KURAL-1	9 ardışık noktadan en az 8 tanesinin (8/9), MÇ'nin bir tarafında bulunması (Şekil 4.7). DEĞERLENDİRME: 9 ardışık cümleden en az 8 tanesinin pozitif veya negatif bölgede bulunması, metnin bu bölümüne yönelik olarak; duygunun istikrarlı şekilde ve aynı yönde verildiğini, yazarın pozitif ve negatif bölgeler arasında gezinmediğini, aynı polaritede ardışık cümleler kullanıldığını anlatmaktadır.
KURAL-2	5 ardışık noktadan en az 4 tanesinin (4/5), MÇ'nin aynı tarafında ve 1 sigmadan daha uzakta bulunması (Şekil 4.8). DEĞERLENDİRME: Ardışık 5 cümle özelinde; duygu tonlamasının aynı polaritede olmak üzere istikrarlı bir şekilde belirginleştğini göstermektedir.
KURAL-3	3 ardışık noktadan en az 2 tanesinin (2/3), MÇ'nin aynı tarafında ve 2 sigmadan daha uzakta bulunması (Şekil 4.9). DEĞERLENDİRME: Ardışık 3 cümle özelinde; duyguların aynı yönde, tesadüften uzak bir şekilde baskınlaştığını göstermektedir.
KURAL-4	1 noktanın, merkez çizgiye uzaklığının 3 sigmadan büyük olması (üst ve alt limitlerin dışında olması) (Şekil 4.10). DEĞERLENDİRME: ÜKL-3 ve AKL-3 sınırlarının dışına taşan bir cümle, metnin akışı içindeki ciddi anlamdaki duygu sapmasını/şiddetini göstermektedir. Yazar, tek bir cümle ile bütün içerięi baskılayabilecek çok güçlü bir duyguyu bu cümlede vermiştir.



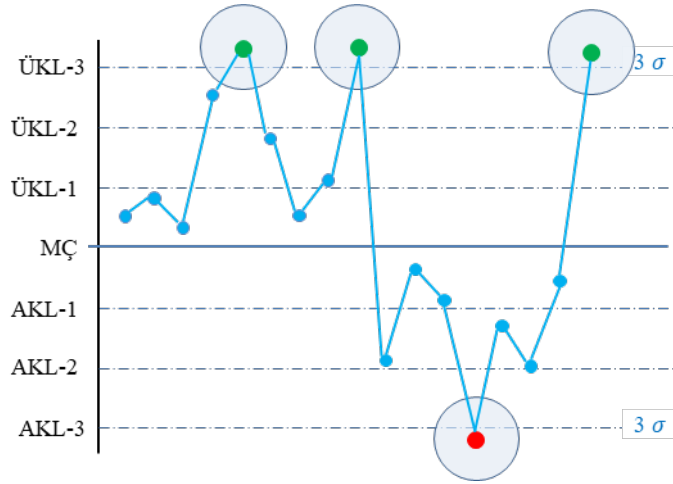
Şekil 4.7 Shewhart kontrol diyagramı KURAL-1 ile duygu haritası analizi.



Şekil 4.8 Shewhart kontrol diyagramı KURAL-2 ile duygu haritası analizi.



Şekil 4.9 Shewhart kontrol diyagramı KURAL-3 ile duygu haritası analizi.



Şekil 4.10 Shewhart kontrol diyagramı KURAL-4 ile duygu haritası analizi.

Özetle; ilk üç kural içerikteki duygu salınımının istikrarı ve gücü hakkında bilgi verirken, son kural duygudaki aşırı vurgunun/sapmanın olduğu cümleleri ortaya çıkarmaktadır.

4.3. DHM Uygulama Örnekleri

Bu bölümde DHM ile 5 farklı örnek üzerinde cümle seviyesinde dokümanların duygu haritaları çıkarılarak, haritalar üzerinde değerlendirmelere yer verilecektir. İlk 3 örnek Bölüm 4’de konu edilen haberlerden oluşmakta iken, 4. örnek ise duygudaki dalgalanmanın ve duygunun şiddetinin yoğun olduğu bir haberdir. 5. ve son örnek ise, ilköğretim 1. sınıf öğrencisi seviyesindeki bir okuma kitabını konu edinmektedir. Sunulacak örnekler ve konu içerikleri hakkındaki bilgiler Çizelge 4.9’da sunulmuştur.

Örnekler üzerinde 4 farklı kuralın ayrı ayrı görseline ilaveten bütün kuralların tek bir duygu haritası üzerinde görselleştirilmesine de yer verilmiştir. Duygu haritasına ilaveten her bir kural etkisinin renklendirilmiş metin sonuçları da ayrı bir şekil numarası ile farklı bir görsel olarak sunulmuştur.

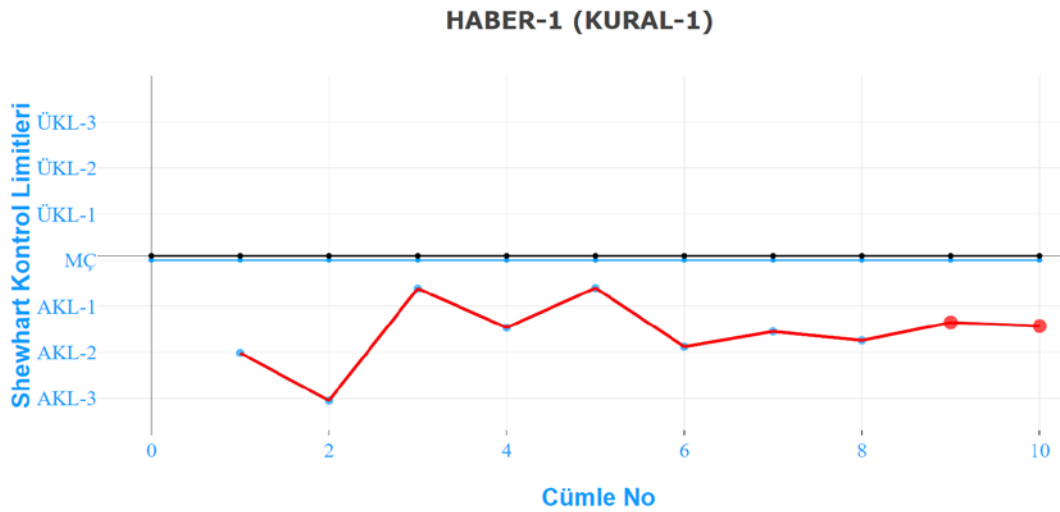
Çizelge 4.9 DHM uygulama örnekleri listesi.

	KONUSU
HABER-1	Çizelge 4.2’de cümle listesi verilen, ATM’lere yönelik bir dolandırıcılık haberi [145].
HABER-2	Çizelge 4.4’de cümle listesi verilen, küçük yaşta büyük başarı gösteren bir çocuğun başarı hikayesini konu eden haber [146].
HABER-3	Çizelge 4.5’de cümle listesi verilen, kansere yakalanan ve hastalıkla mücadelesini kazanarak okuluna geri dönen bir öğretmenin haberi [147].
HABER-4	Şırnak’ta şehit düşen bir polis memurunun ailesine yönelik yazdığı mektubu konu eden haber [152]. Haberin cümle listesi EK-1’de sunulmuştur.
PAMUK_PRENSES	İlköğretim 1. sınıf öğrenci seviyesindeki bir okuma kitabı [153]. Kitabın cümle listesi EK-2’de sunulmuştur.

4.3.1. Uygulama Örneği - 1

4.3.1.1. HABER-1 üzerinde KURAL-1

HABER-1 duygu haritası üzerinde KURAL-1’i uygulama sonuçları Şekil 4.11’de, renklendirilmiş haber metni de Şekil 4.12’de olduğu gibidir.



Şekil 4.11 HABER-1 duygu haritası üzerinde uygulanan KURAL-1 ve sonuçları.

ATM'YE SAHTE PARA YATIRIP GERÇEK PARA ÇEKEN KALPAZANLAR YAKALANDI: 'ENFLASYONU DÜŞÜRECEKTIK, BIRAKMADINIZ'.

Mersin'de, küçük miktardaki paraların belli yerlerini 100 ve 200 TL'lik banknotlara yamalayıp ATM'ye yatıran ve sonra farklı ATM'den gerçek para çeken 4 kişi yakalandı. Sevk edildikleri adliyenin çıkışında görüntülenen şahıslardan bir tanesi, "Kolay değil 3 yılımızı verdik, enflasyonu düşürecektik bırakmadınız" dedi.

AA'nın aktardığı bilgiye göre, Mersin İl Emniyet Müdürlüğü Kaçakçılık ve Organize Suçlarla Mücadele (KOM) Şube Müdürlüğü ekipleri gelen bir ihbar üzerine K.G. isimli bir şahsı takibe aldı.

10 liralık banknotların parçalarını 200 liraya yamalamışlar.

Polis ekipleri, K.G.'nin 10 liralık banknotların belirli parçalarını kullanarak hazırladıkları sahte 200 liralı ATM'lere yatırıp bankaları dolandıran bir ekibin parçası olduğunu tespit etti. Yapılan operasyon sonucunda K.G. ile birlikte 4 kişi gözaltına alınırken, adreslerde 68 adet 100 liralık sahte para bulundu. ATM'lere sahte paraları yatıran kalpazanların Mersin, Gaziantep, Amasya ve Kayseri'deki tanıdıklarına üzerinden orijinal paraları çektiği ortaya çıktı.

K.G. ve S.O. tutuklanırken, diğer iki kişi ise adli kontrol şartıyla serbest bırakıldı. Çıkarıldığı mahkemece tutuklanan K.G., adliye çıkışında gazetecilere "Kolay değil ha 3 yılımızı verdik, enflasyonu düşürecektik bırakmadınız" dedi.

Şekil 4.12 HABER-1 metninin KURAL-1 sonuçlarındaki görseli.

Çizelge 4.8'de de açıklandığı gibi KURAL-1, içerikteki duygunun tesadüften uzak, aynı yönde ve istikrarlı bir şekilde bulunduğu ardışık cümleleri tespit etmektedir. Bildirim noktaları bu tespitin teyit edildiği noktalardır. Ardışık 9 cümleden en az 8'inin aynı polaritede olması KURAL-1 tarafından aranan kriterdir.

Şekil 4.11'deki duygu haritasında KURAL-1 ile;

- 9. ve 10. cümlelerde bildirim noktası tespit edilmiştir.
- 9. cümledeki bildirim noktasına sebep olan, kendisi ile önceki sekiz cümledir (1-9 arası).
- Benzer durum 10. cümle için de söz konusudur (2-10 arası).
- Özetle; 1. cümleden itibaren tesadüften uzak, negatif yönde ve istikrarlı bir duygu salınımının varlığı betimlenmektedir.

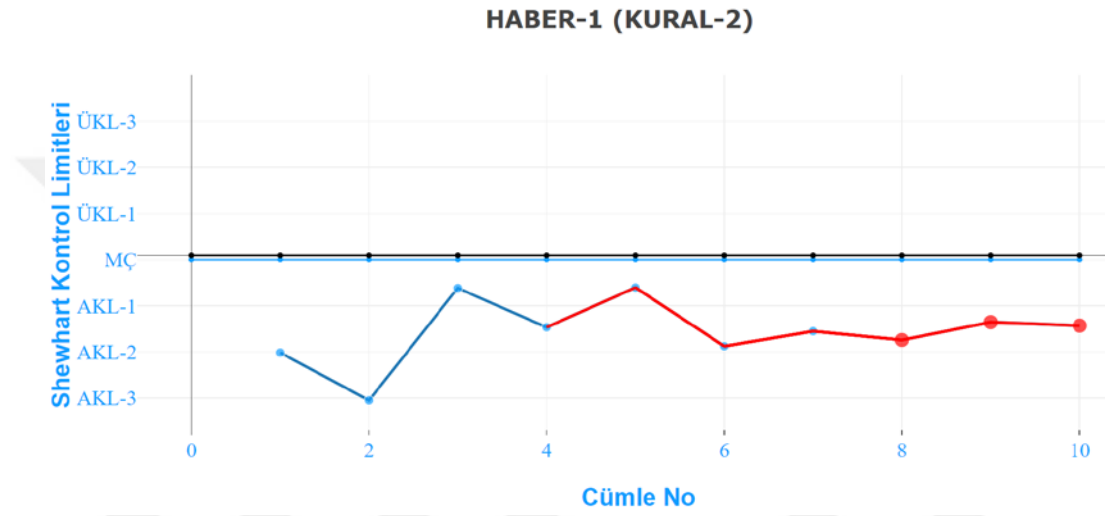
Şekil 4.12'deki metin görselinde;

- Bildirim noktası olarak tespit edilen 9. ve 10. cümleler kırmızı şeritle,

- Bu bildirim noktalarına sebep olan cümleler de kırmızı yazı rengiyle görselleştirilmiştir.

4.3.1.2. HABER-1 üzerinde KURAL-2

HABER-1 duygu haritası üzerinde KURAL-2'nin uygulama sonuçları Şekil 4.13'te, renklendirilmiş haber metni ise Şekil 4.14'de olduğu gibidir.



Şekil 4.13 HABER-1 duygu haritası üzerinde uygulanan KURAL-2 ve sonuçları.

Çizelge 4.8'de de açıklandığı gibi KURAL-2, tesadüften uzak ve istikrarlı şekilde belirgin duygu tonlamasının olduğu ardışık cümleleri tespit eder ve bildirim noktaları bu tespitin teyit edildiği noktalardır. Ardışık 5 cümleden en az 4'ünün 1 sigmadan büyük olması KURAL-2 tarafından aranan kriterdir.

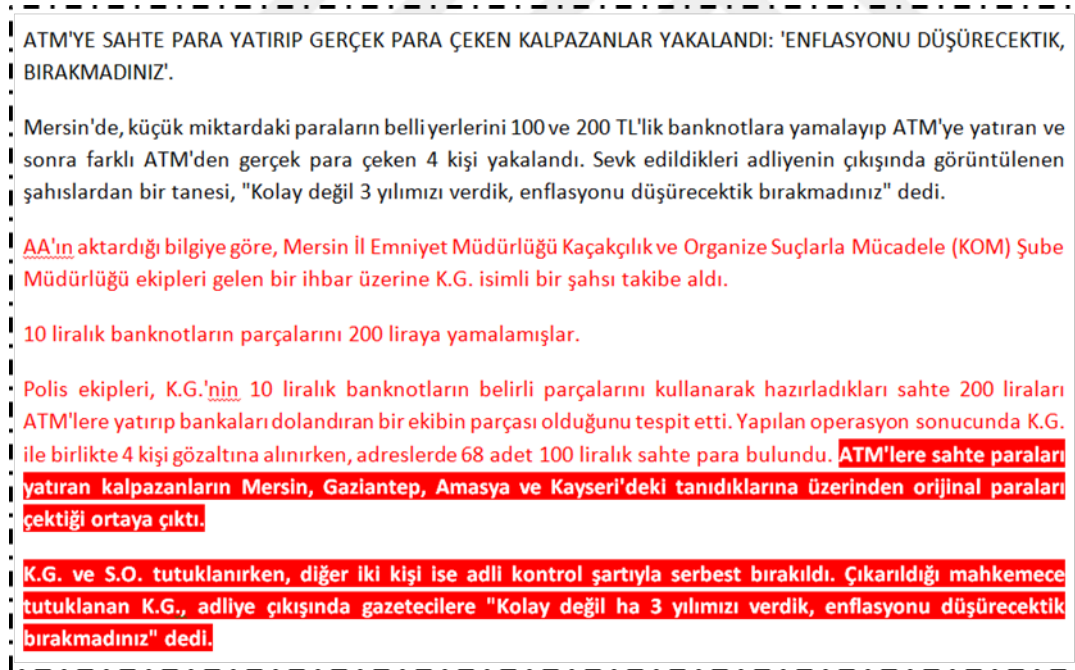
Şekil 4.13'teki duygu haritasına göre KURAL-2 ile;

- 8., 9. ve 10. cümlelerde bildirim noktası tespit edilmiştir.
- 8. cümlede bildirim noktasına sebep olan, bildirim noktası dahil son beş cümledir ve 4. cümleden başlamaktadır. Bu son beş cümle (4-8 arası) incelendiğinde; her ne kadar 5. cümle daha zayıf bir duygu tonu sahip olsa da, diğer dört cümlede duygu tonu 1 sigmadan büyüktür.

- Benzer tespitler, 9. ve 10. cümlelerdeki bildirim noktalarına bakıldığında da söz konusudur.
- Özetle; 4-10 cümleler boyunca (kırmızı çizgi ile gösterilenler) tesadüften uzak, istikrarlı ve negatif yönde belirgin bir duygu tonlamasının var olduğu, 5. cümlelerin daha az belirgin negatif tona sahip olduğu ancak genel istikrarı bozmadığı değerlendirilmelerini yapmak mümkündür.

Şekil 4.14'deki metin görselinde;

- Bildirim noktası olarak tespit edilen 8., 9. ve 10. cümleler kırmızı şeritle,
- Bu bildirim noktalarına sebep olan cümleler kırmızı yazı rengiyle,
- KURAL-2 perspektifinden bir tespite uğramayan cümleler siyah yazı rengiyle görselleştirilmiştir.

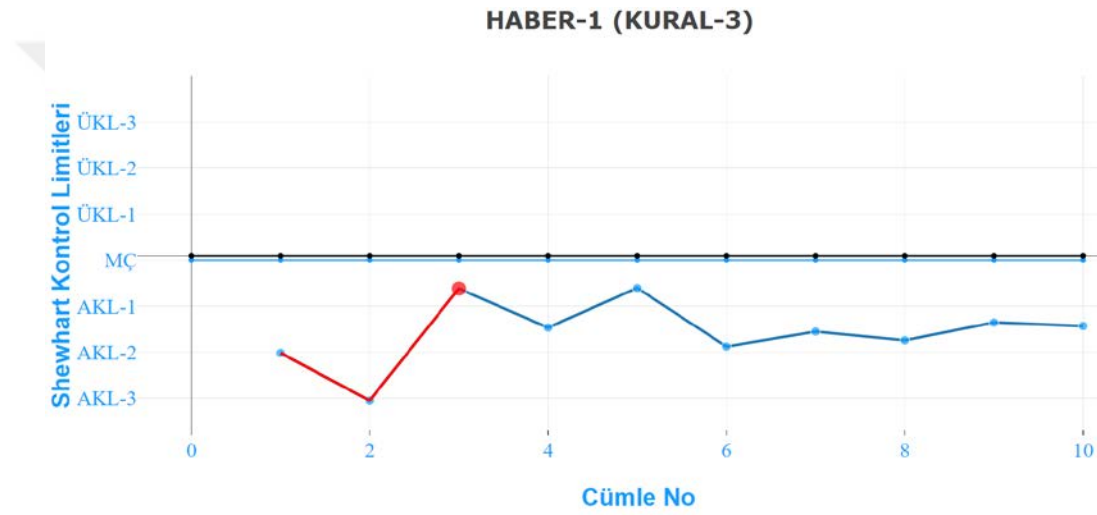


Şekil 4.14 HABER-1 metninin KURAL-2 sonuçlarındaki görseli.

4.3.1.3. HABER-1 üzerinde KURAL-3

HABER-1 üzerinde KURAL-3'ün uygulanması ile oluşan duygu haritası Şekil 4.15'de, renklendirilmiş haber metni Şekil 4.16'da olduğu gibidir.

Çizelge 4.8'de de açıklandığı gibi KURAL-3, tesadüften uzak ve aynı yönde duygunun baskın olduğu bölgeleri tespit eder ve bildirim noktaları bu tespitin teyit edildiği noktalardır. Ardışık 3 cümleden en az 2'sinin 2 sigmadan büyük olması KURAL-3 tarafından aranan kriterdir.



Şekil 4.15 HABER-1 duygu haritası üzerinde uygulanan KURAL-3 ve sonuçları.

Şekil 4.15'deki duygu haritasına göre KURAL-3 ile;

- 3. cümlede bildirim noktası tespit edilmiştir.
- Bu bildirim noktasına sebep olan 1. ve 2. cümlelerdir.
- Özetle; 1. ve 2. cümlede tesadüften uzak, negatif yönde baskın bir duygu tonlamasının var olduğu değerlendirilmesi yapılır.

ATM'YE SAHTE PARA YATIRIP GERÇEK PARA ÇEKEN KALPAZANLAR YAKALANDI: 'ENFLASYONU DÜŞÜRECEKTIK, BIRAKMADINIZ'.

Mersin'de, küçük miktardaki paraların belli yerlerini 100 ve 200 TL'lik banknotlara yamalıyıp ATM'ye yatıran ve sonra farklı ATM'den gerçek para çeken 4 kişi yakalandı. **Sevk edildikleri adliyenin çıkışında görüntülenen şahıslardan bir tanesi, "Kolay değil 3 yılımızı verdik, enflasyonu düşürecektik bırakmadınız" dedi.**

AA'nın aktardığı bilgiye göre, Mersin İl Emniyet Müdürlüğü Kaçakçılık ve Organize Suçlarla Mücadele (KOM) Şube Müdürlüğü ekipleri gelen bir ihbar üzerine K.G. isimli bir şahsı takibe aldı.

10 liralık banknotların parçalarını 200 liraya yamalamışlar.

Polis ekipleri, K.G.'nin 10 liralık banknotların belirli parçalarını kullanarak hazırladıkları sahte 200 liralı ATM'lere yatırıp bankaları dolandıran bir ekibin parçası olduğunu tespit etti. Yapılan operasyon sonucunda K.G. ile birlikte 4 kişi gözaltına alınırken, adreslerde 68 adet 100 liralık sahte para bulundu. ATM'lere sahte paraları yatıran kalpazanların Mersin, Gaziantep, Amasya ve Kayseri'deki tanıdıklarına üzerinden orijinal paraları çektiği ortaya çıktı.

K.G. ve S.O. tutuklanırken, diğer iki kişi ise adli kontrol şartıyla serbest bırakıldı. Çıkarıldığı mahkemece tutuklanan K.G., adliye çıkışında gazetecilere "Kolay değil ha 3 yılımızı verdik, enflasyonu düşürecektik bırakmadınız" dedi.

Şekil 4.16 HABER-1 metninin KURAL-3 sonuçlarındaki görseli.

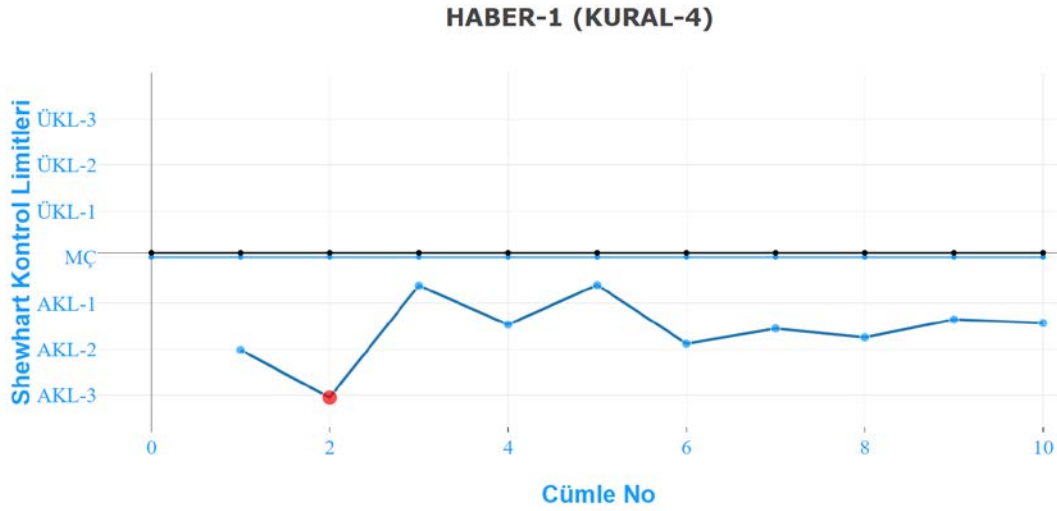
Şekil 4.16'deki metin görselinde;

- Bildirim noktası cümlesi kırmızı şeritle,
- Bildirim noktasına sebep olan cümleler kırmızı yazı rengiyle,
- KURAL-3'ün kapsamına girmeyen cümleler ise siyah yazı rengiyle görselleştirilmiştir.

4.3.1.4. HABER-1 üzerinde KURAL-4

Çizelge 4.8'de de açıklandığı gibi KURAL-4, metnin akışında ciddi anlamdaki duygu sapmasını/şiddetini göstermektedir. Duygu tonunun 3 sigmadan büyük olması, KURAL-4 tarafından aranan kriterdir.

HABER-1'in duygu haritası üzerinde KURAL-4'ün uygulanmasının sonucu Şekil 4.17'de, renklendirilmiş haber metni ise Şekil 4.18'de olduğu gibidir.



Şekil 4.17 HABER-1 duygu haritası üzerinde uygulanan KURAL-4 ve sonuçları.

ATM'YE SAHTE PARA YATIRIP GERÇEK PARA ÇEKEN KALPAZANLAR YAKALANDI: 'ENFLASYONU DÜŞÜRECEKTIK, BIRAKMADINIZ'.

Mersin'de, küçük miktardaki paraların belli yerlerini 100 ve 200 TL'lik banknotlara yamalayıp ATM'ye yatıran ve sonra farklı ATM'den gerçek para çeken 4 kişi yakalandı. Sevk edildikleri adliyenin çıkışında görüntülenen şahıslardan bir tanesi, "Kolay değil 3 yılımızı verdik, enflasyonu düşüreceklik bırakmadınız" dedi.

AA'n aktardığı bilgiye göre, Mersin İl Emniyet Müdürlüğü Kaçakçılık ve Organize Suçlarla Mücadele (KOM) Şube Müdürlüğü ekipleri gelen bir ihbar üzerine K.G. isimli bir şahsı takibe aldı.

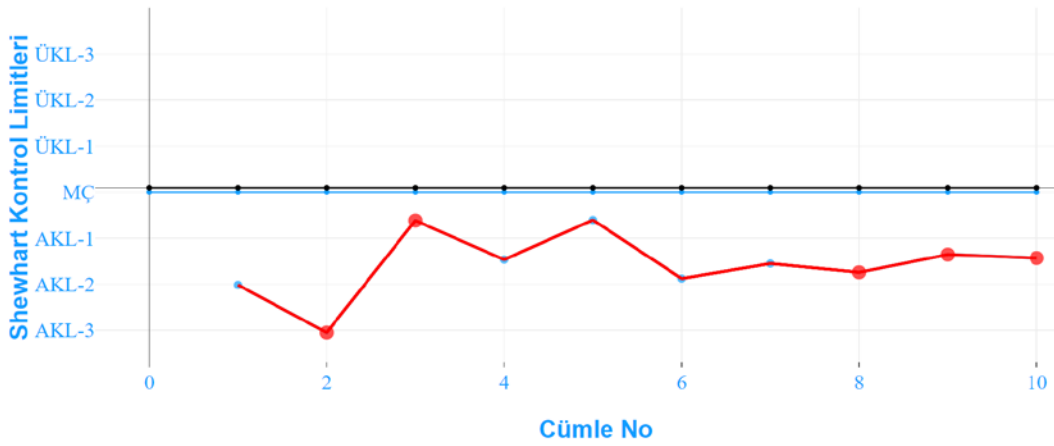
Şekil 4.18 HABER-1 metninin KURAL-4 sonuçlarındaki görseli.

HABER-1 duygu haritası üzerinde KURAL-4, Şekil 4.17'de de görüleceği gibi 2 numaralı cümlede çok büyük bir duygu şiddetini tespit etmiştir. İçerikteki tüm cümleler negatif bölgede duygu salınımı yaparken, 2 numaralı cümle ise bu salınımın en şiddetli olduğu yerdir. Söz konusu bu cümle, Şekil 4.18'de kırmızı şerit içerisinde gösterilmiştir.

4.3.1.5. HABER-1 üzerinde BÜTÜN KURALLAR

Bu aşamaya kadar analizler HABER-1 üzerinde her kural için ayrı ayrı görselleştirilerek verilmiştir. Şekil 4.19 ve Şekil 4.20'de ise bütün kurallar tek bir duygu haritası üzerinde birleştirilerek sunulmuştur.

HABER-1 (BÜTÜN KURALLAR)



Şekil 4.19 HABER-1 duygu haritası üzerinde uygulanan bütün kurallar ve sonuçları.

ATM'YE SAHTE PARA YATIRIP GERÇEK PARA ÇEKEN KALPAZANLAR YAKALANDI: 'ENFLASYONU DÜŞÜRECEKTIK, BIRAKMADINIZ'.

Mersin'de, küçük miktardaki paraların belli yerlerini 100 ve 200 TL'lik banknotlara yamalıyıp ATM'ye yatıran ve sonra farklı ATM'den gerçek para çeken 4 kişi yakalandı. Sevk edildikleri adliyenin çıkışında görüntülenen şahıslardan bir tanesi, "Kolay değil 3 yılımızı verdik, enflasyonu düşüreceklik bırakmadınız" dedi.

AA'nın aktardığı bilgiye göre, Mersin İl Emniyet Müdürlüğü Kaçakçılık ve Organize Suçlarla Mücadele (KOM) Şube Müdürlüğü ekipleri gelen bir ihbar üzerine K.G. isimli bir şahsı takibe aldı.

10 liralık banknotların parçalarını 200 liraya yamalamışlar.

Polis ekipleri, K.G.'nin 10 liralık banknotların belirli parçalarını kullanarak hazırladıkları sahte 200 liralı ATM'lere yatırıp bankaları dolandıran bir ekibin parçası olduğunu tespit etti. Yapılan operasyon sonucunda K.G. ile birlikte 4 kişi gözaltına alınırken, adreslerde 68 adet 100 liralık sahte para bulundu. ATM'lere sahte paraları yatıran kalpazanların Mersin, Gaziantep, Amasya ve Kayseri'deki tanıdıklarına üzerinden orijinal paraları çektiği ortaya çıktı.

K.G. ve S.O. tutuklanırken, diğer iki kişi ise adli kontrol şartıyla serbest bırakıldı. Çıkarıldığı mahkemece tutuklanan K.G., adliye çıkışında gazetecilere "Kolay değil ha 3 yılımızı verdik, enflasyonu düşüreceklik bırakmadınız" dedi.

Şekil 4.20 HABER-1 metni üzerindeki bütün kuralların birleştirilmiş görseli.

Duygu haritası üzerinde tüm kuralların uygulanması sonucunda:

- HABER-1'in negatif bölgede ve istikrarlı bir duygu salınımına sahip olduğu,
- Süregelen negatif istikrarın 8., 9. ve 10. cümlelerle teyit edildiği,
- 3. cümledeki bildirim noktası ile kendisinden önceki iki cümlede negatif tonlamanın baskın olduğu,

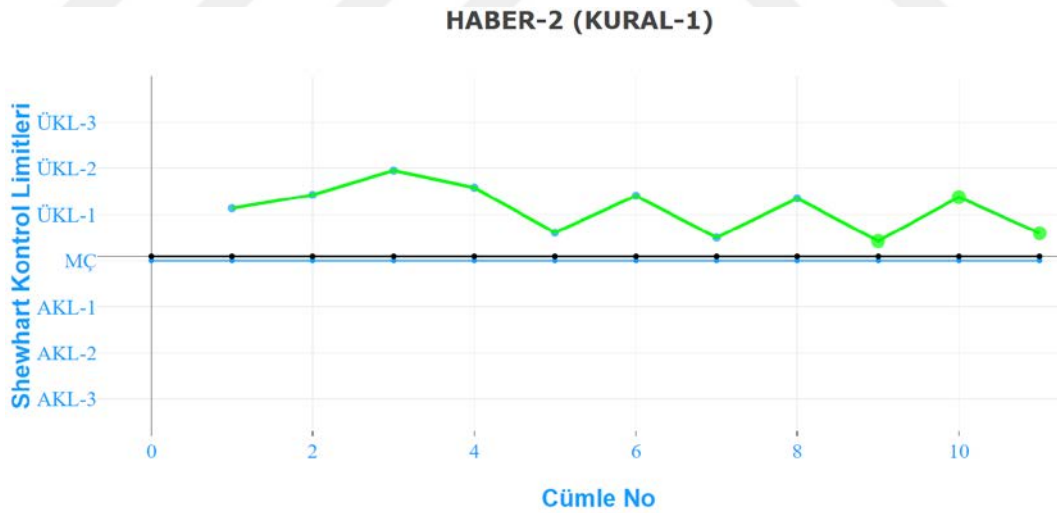
- 2. cümlede kontrol limitlerinin dışına taşan yani şiddetli bir duygu vurgulamasının var olduğu tespitlerini yapmak mümkündür.

Bu tespitler doğrultusunda, haberin negatif tona sahip olduğu görülmektedir. Editörlerce okuyucunun ilgisini çekmek için başlık ve giriş cümlelerinde genellikle baskın bir tonlamaya yer verildiği göz önüne alındığında, haberin giriş cümlelerinin baskın negatif tona sahip olduğu duygu haritası üzerinde de görülmektedir. Haberin ortasından sonuna kadar da olay belirgin ve istikrarlı bir negatif tonlamayla verilmiştir.

4.3.2. Uygulama Örneği - 2

4.3.2.1. HABER-2 üzerinde KURAL-1

HABER-2 üzerinde KURAL-1'in uygulanması ile oluşan duygu haritası Şekil 4.21'de, renklendirilmiş haber metni ise Şekil 4.22'de olduğu gibidir.



Şekil 4.21 HABER-2 duygu haritası üzerinde uygulanan KURAL-1 ve sonuçları.

Piyanonun Parlak Çocuğu: Dünya Birinciliğı Gururu Yaşayan 9 Yaşındaki Nisan, İtalya'daki Yarışmaya Hazırlanıyor.

Müziğe, 6 yaşındayken org ile başlayan 9 yaşındaki Nisan Öksüz'ün yeteneğı öğretmenleri tarafından keşfedilince, anne babası kızlarının müziğe olan ilgisinin sürmesi için ona piyano aldı.

Gülsin Onay ve Gökhan Aybulus gibi ünlü piyanistlerden özel dersler alan Öksüz, piyanoda kendisini geliştirmeye başladı. Geçen sene İtalya'da düzenlenen yarışmada birincilik elde ederek Türkiye'yi gururlandıran Nisan, gelecek ay yine İtalya'da düzenlenecek Apulia Metropolitan Music Contest yarışmasına katılacak.

Nisan, bu yarışmada Stephen Heller, Frederic Chopin, Johann Sebastian Bach, Sergey Prokofyev gibi ünlü isimlerin eserlerini çalacak. Genç piyanist ise, geçen sene birincilik elde ettiğı yarışmaya giderken, yanında kendisine uğur getireceğine inandığı şeyleri götürceğini söyledi.

Anneannesinin yarışma için özel diktiğı elbiseyi giyeceğini, kulağına sol anahtar şeklindeki küpeyi takacağını dile getiren Öksüz, yine öğretmenin verdiği anahtarlığı yanına alacağını kaydetti.

Anne Figen Öksüz, kızının başarısı ile ilgili olarak "Piyano onun arkadaşı, oyuncuğı, her şeyi. **Eve girer girmez piyanonun başına geçer. Beste yapmak ister, sevdiği şarkıları piyanosunda çalmaya çalışır. Benim en sevdiğim şarkıları açar**" dedi.

Şekil 4.22 HABER-2 metninin KURAL-1 sonuçlarındaki görseli.

Şekil 4.21'deki duygu haritasına göre KURAL-1 ile;

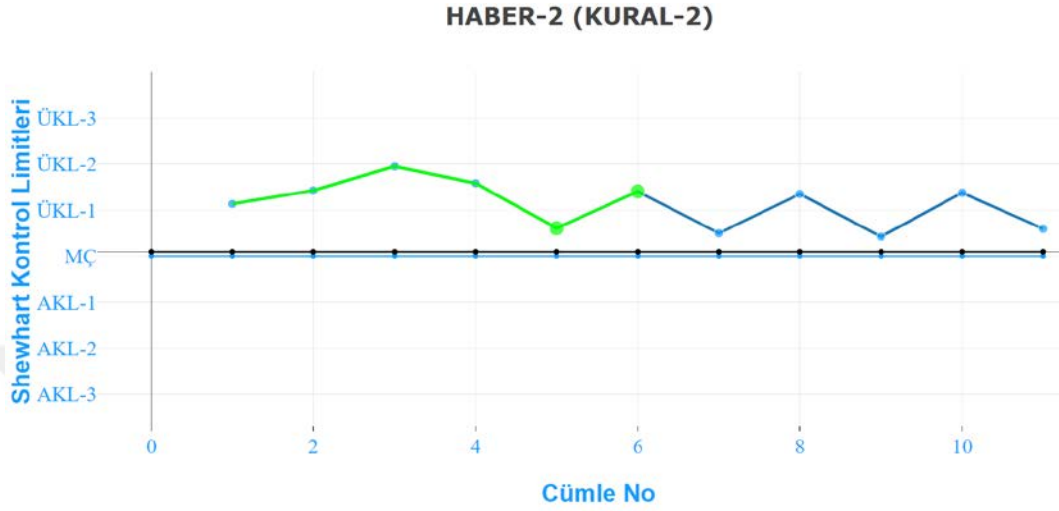
- 9., 10. ve 11. cümlelerde bildirim noktası tespit edilmiştir.
- 9. cümledeki bildirim noktasına sebep olan kendisi ve önceki sekiz cümledir (1-9 arası).
- Benzer durum 10. ve 11. cümle için de söz konusudur.
- Özetle; 1. cümleden itibaren tesadüften uzak ve pozitif yönde bir duygu istikrarı tespit edilmiştir.

Şekil 4.22'deki metin görselinde;

- Bildirim noktası olarak tespit edilen 9., 10. ve 11. cümleler yeşil şeritle,
- Bu bildirim noktalarına sebep olan cümleler de yeşil yazı rengiyle görselleştirilmiştir.

4.3.2.2. HABER-2 üzerinde KURAL-2

HABER-2 üzerinde KURAL-2'nin uygulanması ile oluşan duygu haritası Şekil 4.23'te, renklendirilmiş haber metni ise Şekil 4.24'de olduğu gibidir.



Şekil 4.23 HABER-2 duygu haritası üzerinde uygulanan KURAL-2 ve sonuçları.

Piyanonun Parlak Çocuğu: Dünya Birinciliğı Gururu Yaşayan 9 Yaşındaki Nisan, İtalya'daki Yarışmaya Hazırlanıyor.

Müziğe, 6 yaşındayken org ile başlayan 9 yaşındaki Nisan Öksüz'ün yeteneğı öğretmenleri tarafından keşfedilince, anne babası kızlarının müziğe olan ilgisinin sürmesi için ona piyano aldı.

Gülsin Onay ve Gökhan Aybulus gibi ünlü piyanistlerden özel dersler alan Öksüz, piyanoda kendisini geliştirmeye başladı. Geçen sene İtalya'da düzenlenen yarışmada birincilik elde ederek Türkiye'yi gururlandıran Nisan, gelecek ay yine İtalya'da düzenlenecek Apulia Metropolitan Music Contest yarışmasına katılacak.

Nisan, bu yarışmada Stephen Heller, Frederic Chopin, Johann Sebastian Bach, Sergey Prokofyev gibi ünlü isimlerin eserlerini çalacak. Genç piyanist ise, geçen sene birincilik elde ettiğı yarışmaya giderken, yanında kendisine uğur getireceğıne inandığı şeyleri götürceğini söyledi.

Anneannesinin yarışma için özel diktığı elbiseyi giyeceğini, kulağına sol anahtar şeklindeki küpeyi takacağını dile getiren Öksüz, yine öğretmenin verdiği anahtarlığı yanına alacağını kaydetti.

Anne Figen Öksüz, kızının başarısı ile ilgili olarak "Piyano onun arkadaşı, oyuncağı, her şeyi. Eve girer girmez piyanonun başına geçer. Beste yapmak ister, sevdiği şarkıları piyanosunda çalmaya çalışır. Benim en sevdiğim şarkıları açar" dedi.

Şekil 4.24 HABER-2 metninin KURAL-2 sonuçlarındaki görseli.

Şekil 4.23'teki duygu haritasına göre KURAL-2 ile;

- 5. ve 6. cümlelerde bildirim noktası tespit edilmiştir.

- İlk dört cümlenin duygu tonunun 1 sigmadan büyük olması, 5. cümlede bildirim noktasına sebep olmuştur.
- 2., 3., 4. ve 6. cümleler de 6. cümledeki bildirim noktasına sebep olmuştur.
- Özetle; 1-6 cümleler boyunca (yeşil çizgi ile gösterilenler) tesadüften uzak, istikrarlı ve pozitif yönde belirgin bir duygu tonlamasının bulunduğu, 5. cümlenin daha az belirgin pozitif tona sahip olduğu ancak genel istikrarı bozmadığı, 7. cümleden itibaren haberin sonuna kadar ise baskın olmayan pozitif bir tonda duygu salınımının var olduğu değerlendirilmeleri yapılır.

Şekil 4.24'deki metin görselinde;

- Bildirim noktası olarak tespit edilen 5. ve 6. cümleler yeşil şeritle,
- Bu bildirim noktalarına sebep olan cümleler yeşil yazı rengiyle,
- KURAL-2 perspektifinden bir tespiti uğramayan cümleler ise siyah yazı rengiyle görselleştirilmiştir.

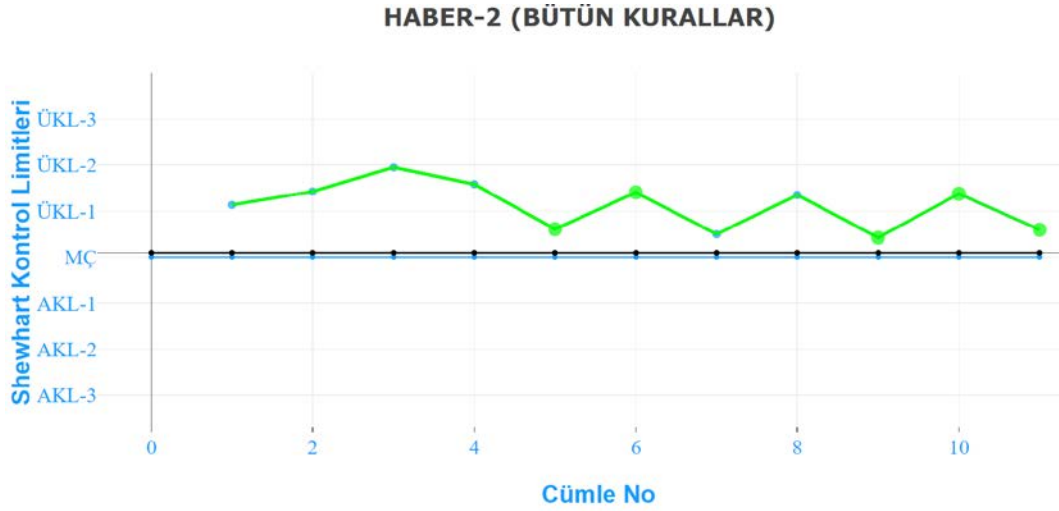
4.3.2.3. HABER-2 üzerinde KURAL-3 ve KURAL-4

HABER-2 duygu haritası üzerinde KURAL-3'ün uygulanması sonrasında herhangi bir bildirim noktası tespit edilmemiştir. Bu durumda, içerikte tesadüfilikten uzak bir şekilde baskın duyguya sahip bir bölümün olmadığı değerlendirilmesini yapmak mümkündür.

Benzer şekilde HABER-2 duygu haritası üzerinde KURAL-4 ile de herhangi bir bildirim noktası tespit edilmemiştir. Bir başka ifadeyle, haber içeriğinde kontrol limitlerini aşacak şekilde şiddetli duygu yansıtan bir cümlenin olmadığı sonucuna ulaşılmaktadır.

4.3.2.4. HABER-2 üzerinde BÜTÜN KURALLAR

Şekil 4.25 ve Şekil 4.26'da HABER-2 duygu haritası üzerinde bütün kural sonuçlarının birleştirilmesiyle oluşturulan duygu haritası ve renklendirilmiş metni sunulmuştur.



Şekil 4.25 HABER-2 duygu haritası üzerinde uygulanan bütün kurallar ve sonuçları.

Piyanonun Parlak Çocuğu: Dünya Birinciliğı Gururu Yaşayan 9 Yaşındaki Nisan, İtalya'daki Yarışmaya Hazırlanıyor.

Müziğe, 6 yaşındayken org ile başlayan 9 yaşındaki Nisan Öksüz'ün yeteneğı öğretmenleri tarafından keşfedilince, anne babası kızlarının müziğe olan ilgisinin sürmesi için ona piyano aldı.

Gülsin Onay ve Gökhan Aybulus gibi ünlü piyanistlerden özel dersler alan Öksüz, piyanoda kendisini geliştirmeye başladı. Geçen sene İtalya'da düzenlenen yarışmada birincilik elde ederek Türkiye'yi gururlandıran Nisan, gelecek ay yine İtalya'da düzenlenecek Apulia Metropolitan Music Contest yarışmasına katılacak.

Nisan, bu yarışmada Stephen Heller, Frederic Chopin, Johann Sebastian Bach, Sergey Prokofyev gibi ünlü isimlerin eserlerini çalacak. Genç piyanist ise, geçen sene birincilik elde ettiğı yarışmaya giderken, yanında kendisine uğur getireceğıne inandığı şeyleri götürceğini söyledi.

Anneannesinin yarışma için özel diktığı elbiseyi giyeceğini, kulağına sol anahtar şeklindeki küpeyi takacağını dile getiren Öksüz, yine öğretmenin verdiği anahtarlığı yanına alacağını kaydetti.

Anne Figen Öksüz, kızının başarısı ile ilgili olarak "Piyano onun arkadaşı, oyuncacı, her şeyi. **Eve girer girmez piyanonun başına geçer. Beste yapmak ister, sevdiği şarkıları piyanosunda çalmaya çalışır. Benim en sevdiğim şarkıları açar" dedi.**

Şekil 4.26 HABER-2 metni üzerindeki bütün kuralların birleştirilmiş görseli.

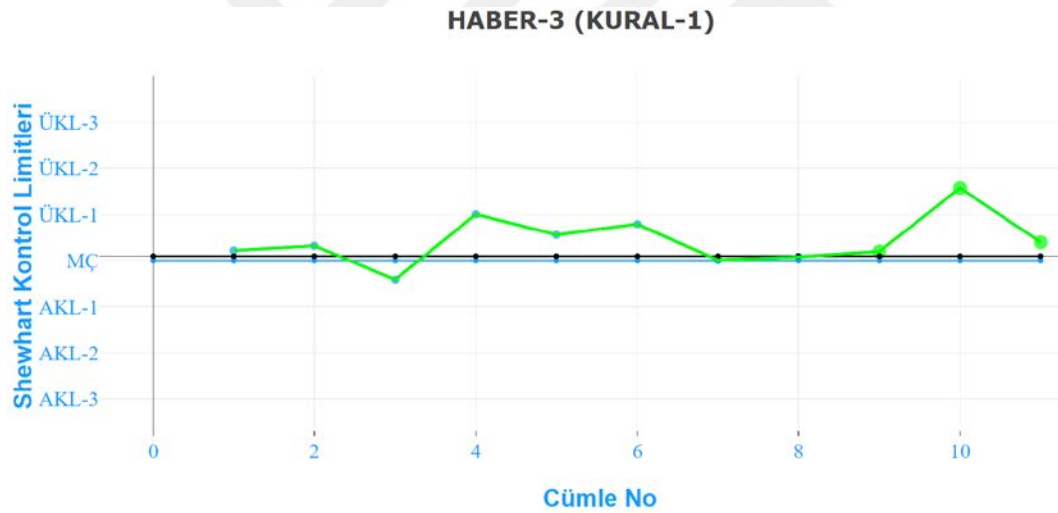
Duygu haritası üzerinde tüm kuralların uygulanması sonucunda:

- HABER-2'nin pozitif bölgede ve istikrarlı bir duygu salınımına sahip olduğu,
- Süregelen istikrarın 9., 10. ve 11. cümlelerdeki bildirim noktaları ile teyit edildiğı,

- 5. ve 6. cümlelerdeki bildirim noktasına bakarak, 1., 2., 3., 4. ve 6. cümleler boyunca (her ne kadar 5. cümlede istikrarı bozan küçük bir sapma olsa da) tesadüften uzak, belirgin ve pozitif bir tonlamanın olduğu,
- İçerikte belirginliğin ötesinde baskın diye nitelediğimiz herhangi bir cümle grubunun olmadığı,
- Haberde kontrol limitlerinin dışına taşan, şiddetli duygu tonlamasına sahip herhangi bir cümlenin olmadığı değerlendirmelerini yapmak mümkündür.

4.3.3. Uygulama Örneği - 3

HABER-3 duygu haritasında KURAL-1 uygulama sonuçları Şekil 4.27’de, renklendirilmiş haber metni ise Şekil 4.28’de olduğu gibidir.



Şekil 4.27 HABER-3 duygu haritası üzerinde uygulanan KURAL-1 ve sonuçları.

KANSERİ YENEN VE AYLAR SONRA OKULUNA DÖNEN ÖĞRETMEN SEVGİ SELİYLE KARŞILANDI.

Tekirdağ'ın Çorlu ilçesi Cemile Yeşil Anadolu Lisesi'nde görev yapan ve kanser hastalığını yenerek 8 ay sonra okuluna dönen coğrafya öğretmeni Muharrem Poyrazoğlu, öğrencilerinin ve arkadaşlarının sevgi gösterileriyle karşılandı.

DHA'nın haberine göre, Çorlu Cemile Yeşil Anadolu Lisesi'nde coğrafya öğretmeni olan Muharrem Poyrazoğlu, mide kanseri hastalığı nedeniyle okula ara verip, öğrencilerinden uzak kaldı. Yaklaşık 8 ay tedavi gören Poyrazoğlu, hastalığını yenip, sağlığına kavuşunca bugün aylar sonra ilk kez okuluna geldi.

ÖĞRENCİLER VE ÖĞRETMENLER, POYRAZOĞLU'NA SÜRPRİZ BİR KARŞILAMA HAZIRLADI.

Okulun merdivenlerinde toplanan öğrenciler, Poyrazoğlu'nun girmesiyle hazırladıkları balon ve konfetileri üzerine atıp, alkışlar, şarkılar ve sloganlarla karşıladı.

ÖĞRENCİLERİN COŞKULU VE SEVGİ DOLU KARŞILAMASI SIRASINDA DUYGULANAN MUHARREM POYRAZOĞLU, ÖĞRENCİLERİNE SARILDI, GÖZYAŞLARI DÖKTÜ.

Dakikalar süren karşılama sırasında duygulu anlar yaşayan öğretmen Muharrem Poyrazoğlu, evine geldiği için çok mutlu olduğunu söyledi ve şöyle devam etti. **Hayatıma giren çok sayıda arkadaşım oldu. Şu an birçoğunun ismini hatırlayamadığım Cemile Yeşil Anadolu Lisesi ailesi içerisinde olup da beni de yaklaşık 8 aylık bu zor sürecimde arayarak, mesajları ile resimleri ile şımarık görüntüleri ile destek olan bütün öğrencilerime, başta okul müdürüm ve öğretmen arkadaşlarımın her birine tek tek, okul dışından sevgili arkadaşlarım, en başta ailem, büyük ailem, bana bu zor süreçte destek olup, yeniden yuvama kazandırdığınız için hepinize teşekkür ediyorum. Sizleri çok seviyorum.**

Şekil 4.28 HABER-3 metninin KURAL-1 sonuçlarındaki görseli.

HABER-3 duygu haritası üzerinde, KURAL-1 dışındaki diğer kurallar tarafından herhangi bir bildirim noktası tespit edilmemiştir. Bu nedenle, HABER-3'ün duygu haritasının analizi sadece KURAL-1 çerçevesinde sunulmuştur. Şekil 4.27'de verilen duygu haritasının analizi sonucunda;

- İçeriğin pozitif ve negatif bölgelerde MÇ'nin yakın çevresinde salınım yaptığı,
- 3. cümlelerin negatif bölgede, diğerlerinin MÇ'ye yakın olmak üzere pozitif bölgede yer aldığı,
- 8., 9. ve 10. cümlelerde bildirim noktasının tespit edildiği, bir başka ifadeyle kendilerine kadar olan duygu salınımlarında tesadüften uzak pozitif yönde istikrar olduğu,
- Her ne kadar 3. cümlede negatif yönde bir sapma olsa da, süregelen pozitif istikrarın bozulmadığı tespitlerini yapmak mümkündür.

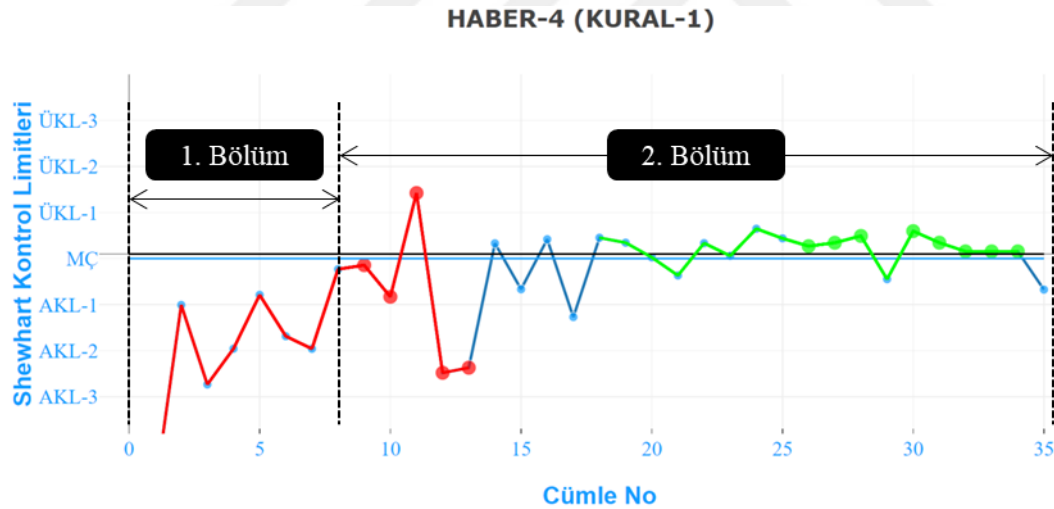
Negatif bölgedeki 3. cümle Şekil 4.28'de siyah yazı rengiyle gösterilmiştir.

4.3.4. Uygulama Örneği - 4

HABER-4'ün içeriği uzun olduğundan önceki örneklerden farklı olarak yalnızca duygu haritalarına yer verilmiş, metinlerin renklendirilmiş görselleri sunulmamıştır. Diğer örneklerden farklı olarak duygu salınımının çok dalgalı bir seyir izlediği haberde, bir polis memurumuzun Şırnak'ta şehit edilmesi ve şehidin sevdiklerine bıraktığı mektubun içeriği konu edilmiştir. Haber içeriği iki ana bölüme sahiptir. Birinci bölüm, haber editörlerince kaleme alınan haberin giriş kısmıdır. Diğer bölüm ise, şehidin mektubunun içeriğidir. Haberinin cümle listesi EK-1'de sunulmuştur.

4.3.4.1. HABER-4 üzerinde KURAL-1

HABER-4 duygu haritası üzerinde uygulanan KURAL-1, Şekil 4.29'da sunulduğu gibidir.



Şekil 4.29 HABER-4 duygu haritası üzerinde uygulanan KURAL-1 ve sonuçları.

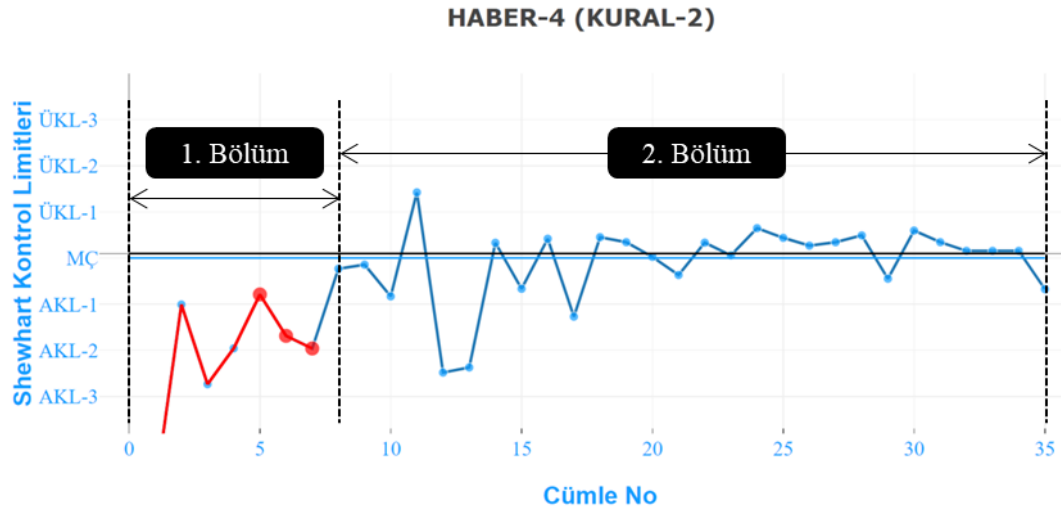
HABER-4'ün ilk bölümünde, şehidin nasıl ve nerede şehit olduğu, geride bıraktıklarının üzüntüleri anlatılmıştır. Bu duygu yoğunluğu, Şekil 4.29'daki duygu haritasına da yansımıştır. Haberin ikinci bölümünde ise, şehidin daha önceden yazdığı mektup aktarılmıştır. Mektubun giriş kısmında yoğun üzüntü ifade eden cümleler, mektubun sonuna doğru ise şehidin ailesi ve sevdiklerine yaptığı tavsiyeler, nasihatler ve iyi dilekler bulunmaktadır. Bu duygu akışı da duygu haritasında gözlemlenmektedir.

Şekil 4.29'daki duygu haritasında KURAL-1 ile;

- 9-13, 26-28 ve 30-34 arasındaki cümlelerde bildirim noktası tespit edilmiştir.
- Kırmızı ile gösterilen bölüm incelendiğinde; 11. cümlede pozitif bölgeye bir sapma olsa da tesadüften uzak, negatif yönde bir istikrarın olduğu,
- Mavi ile gösterilen bölüm incelendiğinde; KURAL-1 çerçevesinde tesadüfilik ile açıklanabilecek, her iki yönde de salınım yapan, belirli bir yönde istikrara sahip olmayan duygu değişiminin bulunduğu,
- Yeşil ile gösterilen bölüm incelendiğinde; her ne kadar 21. ve 29. cümlelerde negatif bölgeye bir sapma olsa da, 18. cümleden itibaren tesadüften uzak ve pozitif yönde bir istikrarın olduğu tespitlerini yapmak mümkündür.

4.3.4.2. HABER-4 üzerinde KURAL-2

HABER-4 duygu haritasının KURAL-2 tarafından değerlendirme sonuçları Şekil 4.30'da sunulmuştur.



Şekil 4.30 HABER-4 duygu haritası üzerinde uygulanan KURAL-2 ve sonuçları.

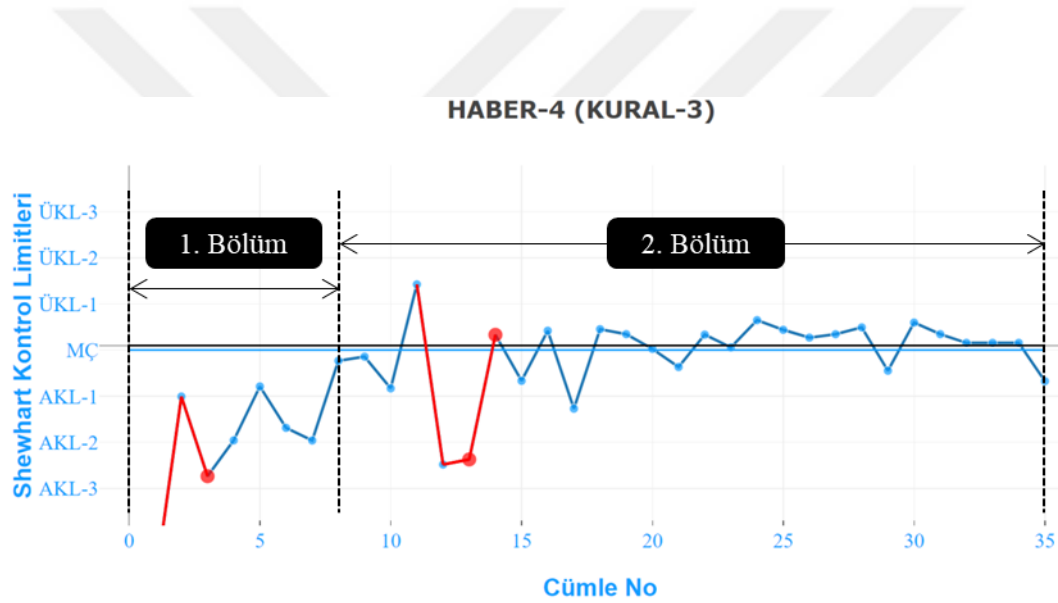
Şekil 4.30'daki duygu haritasında KURAL-2 ile;

- 5., 6. ve 7. cümlelerde bildirim noktası tespit edilmiştir.

- Her ne kadar 5. cümle 1 sigmadan daha zayıf negatif tonlamaya sahip olsa da, ilk yedi cümlede duygunun tesadüften uzak, istikrarlı ve belirgin negatif tonda verildiği,
- Mavi renkle gösterilen haberin geri kalan kısmında ise KURAL-2 çerçevesinde tesadüfle açıklanabilecek ve belirgin duygu kapsamına girmeyen duygu salınımının mevcut olduğu değerlendirilmeleri yapılabilir.

4.3.4.3. HABER-4 üzerinde KURAL-3

HABER-4 duygu haritasının KURAL-3 tarafından değerlendirme sonuçları da Şekil 4.31’de sunulmuştur.



Şekil 4.31 HABER-4 duygu haritası üzerinde uygulanan KURAL-3 ve sonuçları.

Şekil 4.31’deki duygu haritasında KURAL-3 ile;

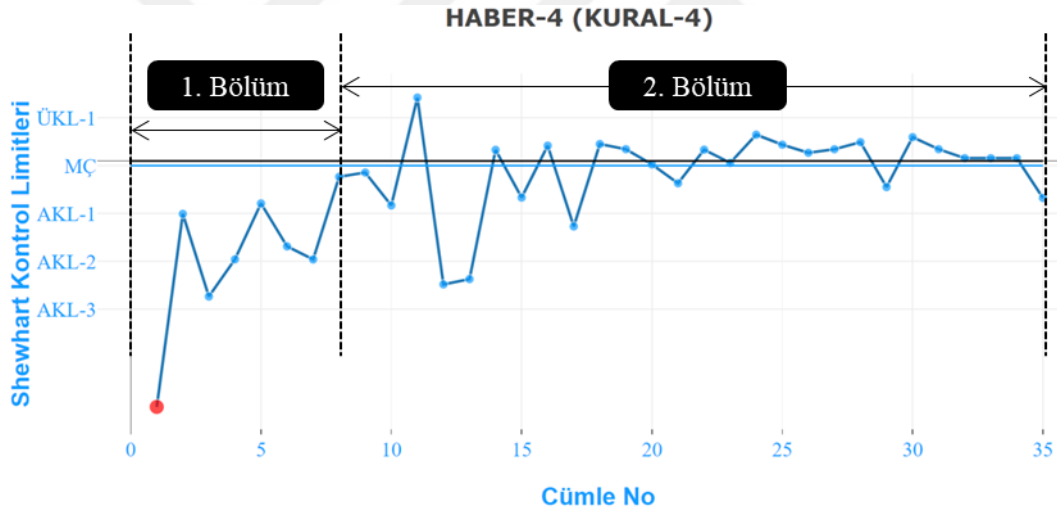
- 3. 13. ve 14. cümlelerde bildirim noktası tespit edilmiştir.
- 1. ve 3. cümleler ile 12. ve 13. cümlelerin bu bildirim noktalarının oluşmasına sebep olduğu ve bu cümlelerde tesadüften uzak ve baskın negatif bir tonlamanın var olduğu,
- Mavi renkli gösterilen bölümlerde ise KURAL-3 çerçevesinde, tesadüfle açıklanabilecek ve baskın duygu kapsamına girmeyen duygu salınımının mevcut olduğu değerlendirilmeleri yapılabilir.

4.3.4.4. HABER-4 üzerinde KURAL-4

HABER-4 duygu haritasının KURAL-4 tarafından değerlendirme sonuçları ise Şekil 4.32’de sunulmuştur. Sonuç incelendiğinde, haberin ilk cümlesinin kontrol limitlerinin çok ötesinde, yani şiddetli negatif duygu tonuna sahip olduğu KURAL-4 tarafından tespit edilmiştir. Haberin söz konusu bu ilk cümlesi Çizelge 4.10’da sunulmuştur.

Çizelge 4.10 HABER-4’ün ilk cümlesi.

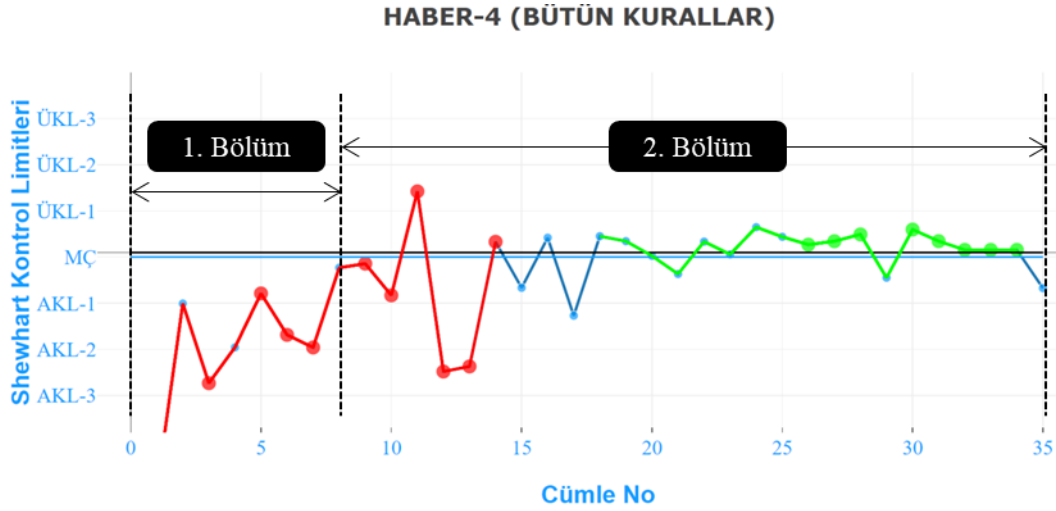
Cümle No	Cümle
1	Şırnak'ta PKK'lı teröristler tarafından polis aracına düzenlenen roketatarlı saldırı sonucu şehit olan polis memuru Mehmet Güngördü'nün şehadetinden önce yazdığı ve cüzdanında sakladığı mektup okuyanları gözyaşlarına boğdu.



Şekil 4.32 HABER-4 duygu haritası üzerinde uygulanan KURAL-4 ve sonuçları.

4.3.4.5. HABER-4 üzerinde BÜTÜN KURALLAR

Şekil 4.33’te HABER-4 duygu haritası üzerinde bütün kural sonuçlarının birleştirilmesiyle oluşturulan duygu haritası sunulmuştur.



Şekil 4.33 HABER-4 duygu haritası üzerinde uygulanan bütün kurallar ve sonuçları.

Şekil 4.33'te sunulan HABER-4 duygu haritası üzerinde bütün kuralların sonuçları analiz edildiğinde;

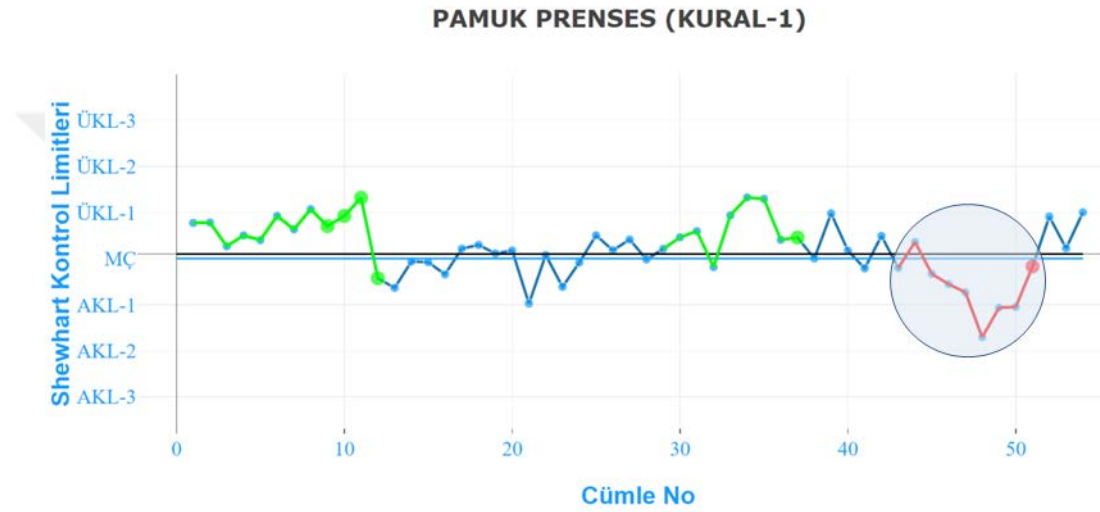
- Haberin ilk bölümünde istikrarlı ve yüksek seviyede negatif tona sahip cümlelerin kullanıldığı,
- Mektup içeriğinin yer aldığı ikinci bölümün giriş kısmında yoğun duygu bildirimlerinin olduğu,
- Mektup içeriğinin mavi renkte çizgilerle gösterilen 14-18 cümleleri arasında duygu salınımının pozitif ve negatif yönde salınım yaptığı ve kurallar kapsamının dışında kaldığı,
- Mektubun son bölümünde ise, istikrarlı ve pozitif duygu yansıtan cümlelerin kullanıldığı değerlendirmelerini yapmak mümkündür.

4.3.5. Uygulama Örneği - 5

Bu bölümde örnek olarak, ilköğretim 1. sınıf düzeyi okuma kitaplarından Pamuk Prenses'in duygu haritasının çıkarılması ve duygu haritası üzerindeki bütün kural sonuçlarının birleşimi ile oluşan sonuca yer verilmiştir. 6-8 yaş arasındaki çocukların duygusal gelişim süreci göz önüne alındığında, okuyacakları kitapların pedagojik olarak kendi seviyelerine uygunluğu oldukça önemlidir. Bu örneğimizin seçilmesindeki temel motivasyonu bu hassasiyet oluşturmuştur.

PAMUK_PRENSES duygu haritası üzerinde, KURAL-1 dışındaki diğer kurallar tarafından herhangi bir bildirim noktası tespiti yapılmamıştır. Bu nedenle duygu haritası analizi sadece KURAL-1 üzerinden yapılmıştır. Kitabın numaralı cümle listesi EK-2’de sunulmuştur.

Şekil 4.34’de PAMUK_PRENSES duygu haritası üzerinde KURAL-1 tarafından yapılan değerlendirme sonuçları görülmektedir.



Şekil 4.34 PAMUK_PRENSES duygu haritası.

Şekil 4.34’deki sonuçlar KURAL-1 perspektifinden analiz edildiğinde;

- Duygu salınımının ağırlıklı olarak MÇ etrafında ve ÜKL-1 ile AKL-1 limitleri arasında olduğu,
- 9-12 ve 36. cümlelerde KURAL-1 tarafından tespit edilen bildirim noktalarından hareketle 1-12 ve 29-37 arasında pozitif yönde istikrarlı bir duygu salınımının olduğu,
- 51. cümledeki bildirim noktasından hareketle de 43-51 arasında negatif yönde istikrarlı bir duygu salınımının bulunduğu,
- Hikayede en yüksek duygu tonunun 48. cümlede ve negatif yönde olduğu değerlendirilmeleri yapılabilir.

Kitabın hedef kitlesinin hassasiyeti göz önüne alındığında, Şekil 4.34’de daire içerisinde gösterilen, negatif bölgede istikrarlı salınım yapan ifadelerle yakından bakmak gerekmektedir. Bu kapsamda odaklanılan 47-51 arasındaki cümleler Çizelge 4.11’de verilmiştir.

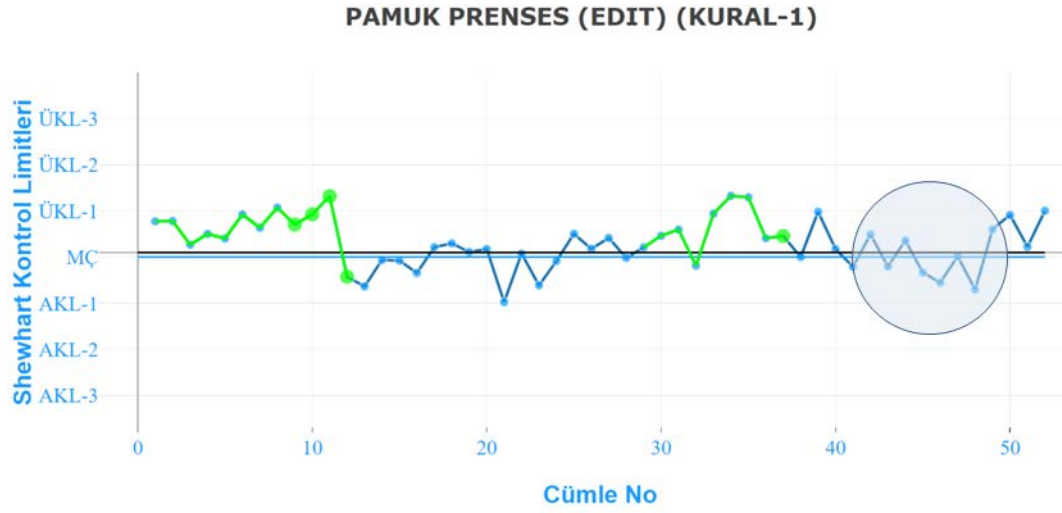
Çizelge 4.11 Yoğun negatif tona sahip cümleler.

Cümle No	Cümle
47	Kraliçe kaçarken bastığı taşa yıldırım düşmüş.
48	Aşağı yuvarlanan Kraliçe ölmüş.
49	Eve dönen cüceler Prensesin yerde yattığını görünce onun öldüğünü düşünmüşler.
50	Onu camdan bir tabuta koyup gece gündüz yanında beklemeye başlamışlar.
51	Bir gün ormandan geçen bir prens, Pamuk Prenses'in tabutunu görmüş.

Çizelge 4.11’deki cümleler incelendiğinde, ilkökul 1. sınıf düzeyine uygun olmayan ifadelerin varlığı dikkati çekmektedir. Hikaye orijinal dahi olsa, hedef kitleye göre düzenlemelerin yapılması gerekmektedir. Bu doğrultuda, hikaye bütünlüğünü bozmayacak şekilde, yukarıda listelenen cümleler daha yumuşak tonda yeniden tarafımızca hedef kitleye göre düzenlenmiş ve PAMUK_PRENSES_EDIT olarak isimlendirilmiştir. Yeniden düzenlenen ifadeler Çizelge 4.12’de verilmiştir. Düzenlenen PAMUK_PRENSES_EDIT hikaye kitabının duygu haritası çıkarılarak KURAL-1 ile değerlendirmeye tabi tutulmuş ve sonuçları Şekil 4.35’de sunulmuştur.

Çizelge 4.12 Yeniden düzenlenen ifadeler.

Cümle No	Cümle
47 - 48	Kraliçe kaçıp uzaklaşmış.
49	Eve dönen cüceler Prensesin uyuduğunu görmüşler ve gece gündüz yanında beklemeye başlamışlar.
50 - 51	Bir gün ormandan geçen bir prens, Pamuk Prenses'in uyuduğunu görmüş.



Şekil 4.35 PAMUK_PRENSES_EDIT duygu haritası.

Şekil 4.35’deki, hikayenin düzenlenmiş ifadelerden oluşan PAMUK_PRENSES_EDIT duygu haritasının daire içine alınan bölgesi incelendiğinde, KURAL-1 tarafından istikrarlı negatif duygu salınımının tespit edilmediği ve böylece hikayenin hedef kitlenin seviyesine uyarlanabildiği değerlendirilmektedir.

5. SONUÇLAR VE ÖNERİLER

5.1. Sonuçlar

Metin türü veya ses türü içeriklerin analizi doğal dil işleme (DDİ) disiplininin ilgi alanına girmekle beraber, dil kurallarının, kütüphanelerinin ve dil araçlarının dile özgü olması nedeniyle, küresel ölçekteki yenilikçi yaklaşımlar, geliştirilen kütüphaneler ve araçlar dil özelinde kalabilmekte ve diğer dillere transferi her zaman mümkün olamamaktadır. Bu tipik karakteristik DDİ'yi diğer bilgisayar bilimleri disiplininden belirgin şekilde ayırmakta ve o dili konuşan araştırmacılara sorumluluklar yüklemektedir. İngilizce için mevcut olan kapsamlı ve gelişmiş dil kaynaklarının Türkçe için de oluşturulması elzemdir. Genel amaçlı duygu sözlüğü de bu kaynaklardan birisi olup, duygu analizi için oldukça önemlidir. Çünkü duygu analizi çok sayıda alt probleme sahiptir ve bu problemlere çözüm önerileri çalışmaları için genel amaçlı duygu sözlüğü temel bir kaynaktır. Doğal dillerin kendilerine has niteliklerinden dolayı, İngilizce kütüphanelerinin çevrilmesi ile oluşturulan kaynaklarla yetinilemeyeceği açıktır. İngilizce dil kütüphanelerinin gelişim süreci incelendiğinde, uzun zamana yayıldığı ve araştırmacılarca kümülatif bir yaklaşımla gelişimine katkı sunulduğu görülmektedir. Türkçe için de araştırmacılar tarafından geliştirilmiş ve geliştirilmesine devam edilmesi gereken dil kaynakları mevcuttur. İngilizce dili için yapılan işbirlikçi yaklaşımın dilimiz için de sergilenmesi durumunda, kısa sürede içerik ve yapısal olarak zengin dilimiz için, gelecekte farklı araştırmalarda kullanılmak üzere temel dil kütüphaneleri kazandırılabilir. Bu temel felsefe doğrultusunda, tezimizin ilk hedefi Genel Amaçlı Türkçe Duygu Sözlüğü'nün oluşturulması şeklinde belirginleşmiştir.

Tez kapsamında duygu sözlüğü gelişim süreci 2 faz halinde çalışılmıştır. Birinci fazda, bütün dünyadaki ana akım haber medyalarını 15 dakikalık periyotlarla tarayarak, İngilizce diline çeviren ve ardından gelişmiş dil kaynakları ile duygu analizi de dahil olmak üzere olay analizi yaparak sonuçları araştırmacılara açık erişimle sunan GDELT veri setlerinden, polariteli Türkçe derlem oluşturulmuştur. Sözlük ve derlem tabanlı hibrit yaklaşımın sergilendiği ilk fazın sonucunda 37K hacminde SWNetTR-PLUS isimli genel amaçlı Türkçe duygu sözlüğü'nün ilk sürümü geliştirilmiştir. Sözlüğün performansı tez kapsamında oluşturduğumuz test derlemi üzerinde ölçülmüştür. Bu

noktada başarının azımsanmayacak bir seviyede olduğu, ancak hedeflediğimiz başarının da henüz yakalanmadığı tespit edilmiştir.

İlk sürüm duygu sözlüğünün kapsamının genişletilmesi ve terimlerin ton değerlerinin kararlı hale getirilmesi, ikinci faz kapsamında çalışılmıştır. Bu doğrultuda, sözlükteki terimlerin eş ve zıt anlam karşılıkları açık erişimli kaynaklardan derlenerek, ilişkiler dünyasını en iyi temsil eden veri yapısı olan çizge dünyasında yapılandırılmış ve problemlere çizge teorisi içinde çözüm aranmıştır. İlk olarak, açık kaynaklardan gelen kirliliğe eş-zıt bağlantılarının elimine edilmesi gündeme gelmiş ve bu doğrultuda kelimeler arasındaki bağın gücünü sayısallaştıran Bağ Gücü matematiksel modeli geliştirilmiştir. Ağ üzerindeki zayıf bağlar bağ gücü değerleri doğrultusunda kopararak ağdan filtrelenmiştir. Bir sonraki adımda, sözlüğe yeni kazandırılan terimlerin ton ve polarite değerlerinin tespitine yönelik olarak, Ton Yayılımı matematiksel modeli geliştirilmiştir. Böylece ağa yeni dahil olan terimlere ton değeri atanabildiği gibi, mevcut terimlerin de ton değerleri daha kararlı hale getirilmiştir. Sözlükteki negatif ve pozitif terimlerin ton değerlerinin dengelenmesi amacıyla Ton Kaydırma yaklaşımları sergilenmiş ve sonuçta 49 K kapasiteli Genel Amaçlı Türkçe Duygu Sözlüğü, SWNetTR++ geliştirilmiştir. Sözlüğün performansı, test derlemi üzerinde ölçülmüş ve performans başarımı önceki sürümün ve hedeflediğimiz noktanın üzerinde gerçekleşmiştir. Bu aşamada tez çalışması başlangıcında belirlediğimiz iki problemten ilkinde çözüm önerisi sunulmuştur.

SWNetTR++ duygu sözlüğü ile ana akım haber medyası üzerinde yapılan performans testleri, ham duygu analizi yaklaşımlarıyla doküman seviyesinde gerçekleştirilmiş sonuçları yansıtmaktadır. Bir başka ifadeyle, duygu analizinde iyileştirici çalışmalar ve alt problemler olarak kabul edilen; kuvvetlendirici ve zayıflatıcılar, kelime türü etkileri, olumsuzluk durumları, soru cümleleri ve kinayeli ifadeler gibi etkenler göz önünde bulundurulmamıştır. İlaveten, duygu analizinin güncel yönelimi olan hedef tabanlı duygu analizi de çalışmada ele alınmamıştır. Dolayısıyla SWNetTR++ duygu sözlüğü, anılan alt problemlere getirilecek çözümlerle beraber kullanıldığında, duygu analizi performansının çok daha da iyi olacağı değerlendirilmekte olup, gelecek çalışmalar kapsamında bu alt problemlerin ele alınması planlanmaktadır.

Birinci probleme yönelik geliştirilen çizge temelli matematiksel modellerin farklı problem alanlarında da uygulanabileceği ve bu sayede literatüre katkı sunulabildiği de değerlendirilmektedir. Bağ Gücü hesaplama modeli, eş ve zıt olarak ilişkilerle modellenebilen bütün problem alanlarına uygulanabilecek bir yaklaşımdır. Örneğin, tıp disiplininde proteinler arasındaki bağlantıların modellenmesi, ilaç sektöründe etken maddeler arası etkileşimin modellenmesi gibi, kenarların ağırlıklandırılmasına ihtiyaç duyulan bütün çizge veri yapılarında kullanılabileceği değerlendirilmektedir. Önerilen bir diğer çizge temelli matematiksel model olan Ton Yayılımı, ağdaki düğüm değerlerinin paylaşılmasına/kararlı hale getirilmesine ihtiyaç duyulan tüm problem sahalarında uygulama alanına sahiptir.

Tez çalışması kapsamında çözüm aradığımız ikinci problem ise, literatürde duygu analizi çalışmalarının ağırlıklı olarak metin sınıflandırma problemi olarak, kısmen de bir regresyon problemi olarak görülmesine yönelik eleştirimizdir. Halihazırdaki birçok duygu analizi çalışmasında, içeriğin duygusu pozitif veya negatif gibi mutlak bir yargıyla ifade edilmektedir. Bazı çalışmalarda ikili sınıflandırmanın yanında, üçlü veya daha çok duygu boyutlu sınıflandırma da yapılmaktadır. Ancak doküman içerisinde duygu, bir salınım göstermektedir ve bu değişimin göz ardı edilmemesi gerekmektedir. Bu motivasyon doğrultusunda, tez kapsamında Duygu Haritası Modeli (DHM) önerilmiştir. Model, cümle ve paragraf seviyesinde ele alınabileceği gibi, zaman skalası boyutunda da kullanılabilme esnekliğine sahiptir ve keşifsel veri analizi yaklaşımı sergilemektedir. İçeriklerdeki duygu salınımının ortak bir şablona yerleştirilmesi ile duygu haritası oluşturulur. Bu ortak şablonun özelliği, duygu skorlarını bölgelere ayırarak görselleştirmesidir. Kontrol limitleri olarak adlandırılan bu sınırların tespiti için 2.12 milyon cümleden oluşan duygu skoru havuzu oluşturulmuştur.

Duygu haritaları üzerindeki analiz için Shewhart Kontrol Diyagramı kuralları kullanılmıştır. 4 temel kural, duygu haritasındaki duygu salınımını; “istikrar”, “tesadüfi olmayan belirginlik”, “tesadüfi olmayan baskınlık” ve “şiddetli duygu sapması” şeklinde özetlenebilecek tespitlere imkan vermektedir. Önerilen bu metodoloji ile,

içeriklerdeki duyguyu farklı perspektiflerden değerlendirebilmek mümkün olabilmiş ve tezimizdeki ikinci probleme de çözüm olarak önerilmiştir.

DHM'nin uygulama alanının oldukça geniş olduğu değerlendirilmektedir. Bölüm 4.3.5'de verilen uygulama örneğinde olduğu gibi, çocuk kitaplarındaki duygu salınımının, okuyucu kitlesinin seviyesi göz önüne alınarak belirli limitler arasında olması pedagojik perspektifinden oldukça önemlidir. Okullar, kütüphaneler ve hatta Türkiye ölçeğinde Milli Eğitim Bakanlığı tarafından kitapların seviye uygunluğunun tespitinde önerilen modelin kullanılabileceği değerlendirilmektedir.

DHM ile duygu analizi, sosyal medya içeriklerinin farklı bakış açısı ile analizini de mümkün kılacak yeteneğe sahiptir. Belirli bir konu başlığı ve zaman periyodu içerisindeki paylaşımların duygu haritası, konuya yönelik toplumsal algının yönü, istikrarı ve şiddeti gibi çok değerli bilgiler sunacaktır.

Benzer bir yaklaşım ana akım haber medyasında yer alan sosyal bir olayın zaman içerisindeki duygusal seyrinin duygu haritaları ile yorumlanmasını, haberlerin içeriği yanında haberlere yapılan kullanıcı yorumlarının duygu haritası ile analiz edilebilmesini de mümkün kılmaktadır. Dünyanın bir bölgesindeki sosyal, politik, ekonomik iklimi, duygu haritası ile modellemek ve yorumlamak mümkündür. Bu sayede, karar vericilerden politika belirleyicilere kadar geniş bir yelpazenin karar destek süreçlerine katkı sunulabileceği de değerlendirilmektedir.

5.2. Öneriler

DDİ'nin dile özgü bir disiplin olması ve inovasyonun dillere özgü kalmasının da getirdiği zorluklarla beraber, çözülmesi gereken çok sayıda alt problem söz konusudur ve yapay zeka yaklaşımlarına muhtaç bir alandır. Araştırmacılar için ise, dil kaynaklarının varlığı ile zenginliği dil çalışmalarının temel başlangıç noktasıdır. İngilizce dilindeki kaynakların kapsam ve zenginliği göz önüne alındığında, Türkçe çalışan ve Türkçe konuşan araştırmacılara büyük sorumluluklar düşmektedir. Özellikle Türkçe WordNet dil kütüphanesinin gelişiminin tekrar canlandırılmasının ve

arařtırmacıların da artırımıyla katkı sunmasının oldukça önemli olduđu deęerlendirilmektedir. Türk Dil Kurumunun da dahil olduđu geniř katılımlı ve uzun soluklu alıřmalar ile etkin ve verimli dil kaynaklarının geliřtirilebileceęi öngörülmektedir.



6. KAYNAKLAR

- [1] D. Schlangen, T. Diekmann, N. Ilinykh, S. Zarrieß, slurk—A Lightweight Interaction Server For Dialogue Experiments and Data Collection, 22nd Workshop on the Semantics and Pragmatics of Dialogue (AixDial / semdial 2018), 8-10 November 2018, Aix-en-Provence, France, **2018**.
- [2] A. Danyluk, S. Buck, Artificial Intelligence Competencies for Data Science Undergraduate Curricula, The Ninth Symposium on Educational Advances in Artificial Intelligence (EAAI-19), 28-29 January 2019, Honolulu, Hawaii, USA, **2019**.
- [3] D. Küçük, N. Arıcı, Doğal Dil İşlemede Derin Öğrenme Uygulamaları Üzerine Bir Literatür Çalışması, Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi, 2 (**2018**) 76-86.
- [4] J. Allen, Natural Language Understanding, Benjamin/Cummings Publishing Company, USA, **1995**.
- [5] E. Adalı, Doğal Dil İşleme, Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 5 (**2012**) 1.
- [6] V. Jeyamani, J. Ashok, D. S. Suppiah, A review on significance of sub fields in artificial intelligence, International Journal of Latest Trends in Engineering and Technology, 6 (**2016**) 542-548.
- [7] S. E. ŞEKER, Doğal Dil İşleme (Natural Language Processing), YBS Ansiklopedi, 2 (**2015**) 14-22.
- [8] G. G. Chowdhury, Natural language processing, Annual Review of Information Science and Technology, 37 (**2005**) 51-89.
- [9] C. Türkmenoğlu, Türkçe Metinlerde Duygu Analizi, Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, **2015**.
- [10] M. Z. Kurdi, Natural Language Processing and Computational Linguistics 2: Semantics, Discourse and Applications, John Wiley & Sons, New Jersey, USA, **2017**.
- [11] E. D. Liddy, Natural Language Processing. Encyclopedia of Library and Information Science, M. Drake (Eds), Vol.2, New York, **2001**.
- [12] S. Hellmann, S. Auer, Towards Web-Scale Collaborative Knowledge Extraction. The People's Web Meets NLP, Springer, Berlin, 287-313, **2013**.
- [13] J. Hirschberg, C. D. Manning, Advances in Natural Language Processing, Science, 349 (**2015**) 261-266.
- [14] M. Bates, R. J. Bobrow, R. M. Weischedel, Critical Challenges for Natural Language Processing. Challenges in Natural Language Processing, M. Bates, R. M. Weischedel (Eds), Vol.1, Cambridge University Press, USA, 3-34, **1993**.
- [15] B. Liu, Sentiment analysis and opinion mining, Synthesis Lectures on Human Language Technologies, 5 (**2012**) 1-167.

- [16] A. Montoyo, P. MartíNez-Barco, A. Balahur, Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments, *Decision Support Systems*, 53 (2012) 675-679.
- [17] W. Medhat, A. Hassan, H. Korashy, Sentiment analysis algorithms and applications: A survey, *Ain Shams Engineering Journal*, 5 (2014) 1093-1113.
- [18] S. Sun, C. Luo, J. Chen, A review of natural language processing techniques for opinion mining systems, *Information Fusion*, 36 (2017) 10-25.
- [19] I. Chaturvedi, E. Cambria, R. E. Welsch, F. Herrera, Distinguishing between facts and opinions for sentiment analysis: Survey and challenges, *Information Fusion*, 44 (2018) 65-77.
- [20] B. Pang, L. Lee, Opinion mining and sentiment analysis, *Foundations and Trends® in Information Retrieval*, 2 (2008) 1-135.
- [21] M. Korayem, K. Aljadda, D. Crandall, Sentiment/subjectivity analysis survey for languages other than English, *Social Network Analysis and Mining*, 6 (2016) 75.
- [22] D. M. E. D. M. Hussein, A survey on sentiment analysis challenges, *Journal of King Saud University-Engineering Sciences*, 30 (2018) 330-338.
- [23] C. C. Aggarwal, C. Zhai, *Mining Text Data*, Springer Science & Business Media, NY, USA, 2012.
- [24] K. Ravi, V. Ravi, A survey on opinion mining and sentiment analysis: tasks, approaches and applications, *Knowl-Based Syst*, 89 (2015) 14-46.
- [25] G. Gezici, B. Yanıkoğlu, Sentiment analysis in Turkish. *Turkish Language Processing*, K. Oflazer, M. Saraçlar (Eds), Springer Publishing, Cham, Switzerland, 255-271, 2018.
- [26] D. M. El-Din, Enhancement bag-of-words model for solving the challenges of sentiment analysis, *International Journal of Advanced Computer Science and Applications*, 7 (2016) 244-247.
- [27] Z. Zhang, Weighing stars: Aggregating online product reviews for intelligent e-commerce applications, *Ieee Intell Syst*, 23 (2008) 42-49.
- [28] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, M. Stede, Lexicon-based methods for sentiment analysis, *Comput Linguist*, 37 (2011) 267-307.
- [29] B. Pang, L. Lee, in *Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating sales*, *ACL 43rd Annual Meeting*, Ann Arbor, Michigan, 25-30 July 2005, ACL, New Brunswick, 2005, p. 115-124.
- [30] R. Feldman, Techniques and applications for sentiment analysis, *Communications of the ACM*, 56 (2013) 82-89.
- [31] R. Bougie, R. Pieters, M. Zeelenberg, Angry customers don't come back, they get back: The experience and behavioral implications of anger and dissatisfaction in services, *Journal of the academy of marketing science*, 31 (2003) 377-393.
- [32] S. M. Mohammad, T. W. Yang, in A. Balahur, E. Boldrini, A. Montoyo, P. MartíNez-Barco (Eds), *Tracking Sentiment in Mail: How Genders Differ on Emotional Axes*, 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, Portland, OR, USA, 24 June 2011, Association for Computational Linguistics, 2011, p. 70-79.

- [33] N. Ravaja, T. Saari, M. Turpeinen, J. Laarni, M. Salminen, M. Kivikangas, Spatial presence and emotions during video game playing: Does it matter with whom you play?, *Presence: Teleoperators and Virtual Environments*, 15 (**2006**) 381-392.
- [34] D. J. Litman, K. Forbes-Riley, in D. Scott, W. Daelemans, M. A. Walker (Eds), *Predicting Student Emotions in Computer-Human Tutoring Dialogues*, 42nd Annual Meeting of the Association for Computational Linguistics, Barcelona, Spain, 21-26 July 2004, Association for Computational Linguistics, **2004**, p. 351.
- [35] V. Francisco, P. Gervás, in P. Sojka, I. Kopeček, K. Pala (Eds), *Automated mark up of affective information in english texts*, International Conference on Text, Speech and Dialogue, Brno, Czech Republic, 11-15 September 2006, Springer, **2006**, p. 375-382.
- [36] J. R. Bellegarda, in *Emotion analysis using latent affective folding and embedding*, NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, Los Angeles, California, USA, 5 June 2010, Association for Computational Linguistics, **2010**, p. 1-9.
- [37] M. R. Saleh, M. T. Martín-Valdivia, A. Montejo-Ráez, L. Ureña-López, Experiments with SVM to classify opinions in different domains, *Expert Systems with Applications*, 38 (**2011**) 14799-14804.
- [38] B. Liu, *Sentiment analysis: Mining opinions, sentiments, and emotions*, Cambridge University Press, USA, **2015**.
- [39] F. A. Pozzi, E. Fersini, E. Messina, B. Liu, *Challenges of sentiment analysis in social networks: An overview*. *Sentiment Analysis in Social Networks*, Elsevier, Cambridge, 1-11, **2017**.
- [40] B. Pang, L. Lee, S. Vaithyanathan, in *Thumbs up?: sentiment classification using machine learning techniques*, ACL-02 Conference on Empirical Methods in Natural Language Processing, PA, USA, 6-7 July 2002, Association for Computational Linguistics, **2002**, p. 79-86.
- [41] P. D. Turney, in *Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews*, 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, USA, 7-12 July 2002, Association for Computational Linguistics, **2002**, p. 417-424.
- [42] J. M. Wiebe, R. F. Bruce, T. P. O'Hara, in R. Dale, K. W. Church (Eds), *Development and Use of a Gold-Standard Data Set for Subjectivity Classifications*, 27th Annual Meeting of the Association for Computational Linguistics, College Park, Maryland, USA, 20-26 June 1999, Association for Computational Linguistics, **1999**, p. 246-253.
- [43] T. Wilson, J. Wiebe, R. Hwa, in *Just how mad are you? Finding strong and weak opinion clauses*, AAAI, California, 25-29 July 2004, AAAI, California, **2004**, p. 761-769.
- [44] M. Hu, B. Liu, in *Mining and summarizing customer reviews*, Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, USA, 22-25 August 2004, ACM, **2004**, p. 168-177.
- [45] S. Poria, E. Cambria, R. Bajpai, A. Hussain, *A review of affective computing: From unimodal analysis to multimodal fusion*, *Information Fusion*, 37 (**2017**) 98-125.

- [46] R. Dehkharghani, B. Yanikoglu, Y. Saygin, K. Oflazer, Sentiment analysis in Turkish at different granularity levels, *Natural Language Engineering*, 23 **(2017)** 535-559.
- [47] R. Dehkharghani, *Sentiment Analysis in Turkish: Resources and Techniques*, PhD Thesis, Sabancı University, İstanbul, **2015**.
- [48] R. Narayanan, B. Liu, A. Choudhary, in *Sentiment analysis of conditional sentences*, Conference on Empirical Methods in Natural Language Processing, Singapore, 6-7 August 2009, Association for Computational Linguistics, **2009**, p. 180-189.
- [49] A. Neviarouskaya, H. Prendinger, M. Ishizuka, in *Compositionality principle in recognition of fine-grained emotions from text*, Third International Conference on Weblogs and Social Media, San Jose, California, 17-20 May 2009, AAAI Press, California, **2009**, p. 278-281.
- [50] D. Davidov, O. Tsur, A. Rappoport, in *Semi-supervised recognition of sarcastic sentences in twitter and amazon*, Fourteenth Conference on Computational Natural Language Learning, Uppsala, Sweden, 15-16 July 2010, ACL, USA, **2010**, p. 107-116.
- [51] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, A. Kappas, Sentiment strength detection in short informal text, *Journal of the American Society for Information Science and Technology*, 61 **(2010)** 2544-2558.
- [52] V. Hatzivassiloglou, K. R. McKeown, in *Predicting the semantic orientation of adjectives*, 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics, Madrid, Spain, 7-12 July 1997, Morgan Kaufmann Publishers, **1997**, p. 174-181.
- [53] L. Williams, C. Bannister, M. Arribas-Ayllon, A. Preece, I. Spasić, The role of idioms in sentiment analysis, *Expert Systems with Applications*, 42 **(2015)** 7375-7385.
- [54] S. Vohra, J. Teraiya, A comparative study of sentiment analysis techniques, *Journal JIKRCE*, 2 **(2013)** 313-317.
- [55] N. Mahendran, T. Mekala, A Survey: Sentiment Analysis Using Machine Learning Techniques for Social Media Analytics, *International Journal of Pure and Applied Mathematics*, 118 **(2018)** 419-423.
- [56] T. Parlar, S. A. Özel, F. Song, QER: a new feature selection method for sentiment analysis, *Human-centric Computing Information Sciences*, 8 **(2018)** 10.
- [57] K. W. Church, P. Hanks, Word association norms, mutual information, and lexicography, *Computational Linguistics*, 16 **(1990)** 22-29.
- [58] F. Xianghua, L. Guo, G. Yanyan, W. Zhiqiang, Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon, *Knowledge Based Systems*, 37 **(2013)** 186-195.
- [59] A. Esuli, F. Sebastiani, in *Sentiwordnet: A publicly available lexical resource for opinion mining*, 5th International Conference on Language Resources and Evaluation, Genoa, Italy, 24-26 May 2006, Citeseer, **2006**, p. 417-422.
- [60] S. Baccianella, A. Esuli, F. Sebastiani, Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining, *The Seventh International Conference on Language Resources and Evaluation*, Valletta, Malta, **2010**.

- [61] S.-M. Kim, E. Hovy, in Determining the Sentiment of Opinions, 20th International Conference on Computational Linguistics, Geneva, Switzerland, 23-27 August 2004, ACM, **2004**, p. 1367.
- [62] L. Polanyi, A. Zaenen, Contextual valence shifters. Computing Attitude and Affect in Text: Theory and Applications, J. G. Shanahan, Y. Qu, J. Wiebe (Eds), Springer, Netherlands, 1-10, **2006**.
- [63] A. Kennedy, D. Inkpen, Sentiment Classification of Movie Reviews Using Contextual Valence Shifters, Computational Intelligence, 22 (**2006**) 110-125.
- [64] L. Jia, C. Yu, W. Meng, in The effect of negation on sentiment analysis and retrieval effectiveness, 18th ACM Conference on Information and Knowledge Management, Hong Kong, 2-6 November 2009, ACM, **2009**, p. 1827-1830.
- [65] G. Eryiğit, in G. Bouma, Y. Parmentier (Eds), ITU Turkish NLP Web Service, 14th Conference of the European Chapter of the Association for Computational Linguistics, Gothenburg, Sweden, 26-30 April 2014, ACL, **2014**, p. 1-4.
- [66] F. Å. Nielsen, A new ANEW: Evaluation of a word list for sentiment analysis in microblogs, arXiv preprint arXiv:1103.2903, (**2011**).
- [67] M. Abdul-Mageed, M. T. Diab, M. Korayem, in Subjectivity and Sentiment Analysis of Modern Standard Arabic, The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, Oregon, USA, 19-24 June 2011, ACL, **2011**, p. 587-591.
- [68] C. Türkmenoğlu, A. C. Tantıuş, in Sentiment analysis in Turkish media, International Conference on Machine Learning (ICML), Beijing, China, 22-24 June 2014, **2014**,
- [69] S. M. Mohammad, P. D. Turney, Crowdsourcing a word-emotion association lexicon, Computational Intelligence, 29 (**2013**) 436-465.
- [70] J. An, H.-W. Kim, Building a Korean sentiment lexicon using collective intelligence, Journal of Intelligence Information Systems, 21 (**2015**) 49-67.
- [71] Y. Hong, H. Kwak, Y. Baek, S. Moon, in Tower of Babel: a crowdsourcing game building sentiment lexicons for resource-scarce languages, 22nd International World Wide Web Conference, Rio de Janeiro, Brazil, 13-17 May 2013, ACM, **2013**, p. 549-556.
- [72] M. Lafourcade, A. Joubert, N. Le Brun, in R. Mitkov, G. Angelova, K. Bontcheva (Eds), Collecting and evaluating lexical polarity with a game with a purpose, Recent Advances in Natural Language Processing, Hissar, Bulgaria, 7-9 September 2015, INCOMA, **2015**, p. 329-337.
- [73] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, K. J. Miller, Introduction to WordNet: An on-line lexical database, International journal of lexicography, 3 (**1990**) 235-244.
- [74] J. Kamps, M. Marx, R. J. Mokken, M. De Rijke, in Using WordNet to Measure Semantic Orientations of Adjectives, LREC, Citeseer, **2004**, p. 1115-1118.
- [75] S. Blair-Goldensohn, K. Hannan, R. McDonald, T. Neylon, G. Reis, J. Reynar, Building a sentiment summarizer for local service reviews, NLP Challenges in the Information Explosion Era (NLPIX 2008), 22-24 April 2008, Beijing, China, **2008**.

- [76] D. Rao, D. Ravichandran, in Semi-supervised polarity lexicon induction, EACL 2009 : European Chapter of the ACL, Athens, Greece, 30 March - 3 April 2009, Association for Computational Linguistics, **2009**, p. 675-682.
- [77] H. Hammer, A. Bai, A. Yazidi, P. Engelstad, Building sentiment lexicons applying graph theory on information from three norwegian thesauruses, Norsk Informatikkonferanse (NIK 2014), 17-19 November 2014, Fredrikstad, Norway, **2014**.
- [78] V. Basile, Sentiment analysis on Italian tweets, 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 4-6 June 2013, Georgia, USA, **2013**.
- [79] V. Perez-Rosas, C. Banea, R. Mihalcea, Learning Sentiment Lexicons in Spanish, Eighth International Conference on Language Resources and Evaluation (LREC 2012), 23-25 May 2012, Istanbul, Turkey, **2012**.
- [80] A. Uçan, Automatic Sentiment Dictionary Translation and Using in Sentiment Analysis, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara, **2014**.
- [81] J. Yao, G. Wu, J. Liu, Y. Zheng, in Using bilingual lexicon to judge sentiment orientation of Chinese words, The Sixth IEEE International Conference on Computer and Information Technology (CIT'06), Seoul, Korea, 20-22 September 2006, IEEE, **2006**, p. 38-38.
- [82] H. Kanayama, T. Nasukawa, in Fully automatic lexicon expansion for domain-oriented sentiment analysis, 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP 2006), Sydney, Australia, 22-23 July 2006, Association for Computational Linguistics, **2006**, p. 355-363.
- [83] R. Remus, U. Quasthoff, G. Heyer, in N. Calzolari (Eds), SentiWS-A Publicly Available German-language Resource for Sentiment Analysis, International Conference on Language Resources and Evaluation (LREC 2010), Valletta, Malta, 17-23 May 2010, ELRA, **2010**, p. 1168-1171.
- [84] M. Kaity, V. Balakrishnan, An automatic non-English sentiment lexicon builder using unannotated corpus, The Journal of Supercomputing, 75 (**2019**) 2243-2268.
- [85] P. J. Stone, D. C. Dunphy, M. S. Smith, The General Inquirer: A computer Approach to Content Analysis, M.I.T. Press, Oxford, England, **1966**.
- [86] G. A. Miller, WordNet: a Lexical Database for English, Communications of the ACM, 38 (**1995**) 39-41.
- [87] P. Vossen, A Multilingual Database with Lexical Semantic Networks, Dordrecht: Kluwer Academic Publishers, 10 (**1998**) 978-994.
- [88] D. Tufis, D. Cristea, S. Stamou, BalkaNet: Aims, methods, results and perspectives. a general overview, Romanian Journal of Information Science and Technology, 7 (**2004**) 9-43.
- [89] O. Bilgin, Ö. Çetinoğlu, K. Oflazer, Building a wordnet for Turkish, Romanian Journal of Information Science and Technology, 7 (**2004**) 163-172.
- [90] A. Esuli, F. Sebastiani, SentiWordNet: a high-coverage lexical resource for opinion mining, Evaluation, 17 (**2007**) 1-26.

- [91] A. Ucan, B. Naderalvojud, E. A. Sezer, H. Sever, in SentiWordNet for new language: Automatic translation approach, 2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), IEEE, **2016**, p. 308-315.
- [92] C. Strapparava, A. Valitutti, in Wordnet affect: an affective extension of wordnet, Fourth International Conference on Language Resources and Evaluation, Lisbon, Portugal, 26-27-28 May 2004, Citeseer, France, **2004**, p. 1083-1086.
- [93] E. Cambria, R. Speer, C. Havasi, A. Hussain, in Senticnet: A publicly available semantic resource for opinion mining, AAAI Fall Symposium 2010 - Commonsense Knowledge, Virginia, USA, November 11-13, 2010, **2010**, p. 14-18.
- [94] J. C. de Albornoz, L. Plaza, P. Gervás, in SentiSense: An easily scalable concept-based affective lexicon for sentiment analysis, Eighth International Conference on Language Resources and Evaluation (LREC 2012), Istanbul, Turkey, 23-25 May 2012, **2012**, p. 3562-3567.
- [95] S. Cerini, V. Compagnoni, A. Demontis, M. Formentelli, G. Gandini, Micro-WNOp: A gold standard for the evaluation of automatically compiled lexical resources for opinion mining, Language Resources and Linguistic Theory, 1 (**2007**) 200-210.
- [96] T. Wilson, J. Wiebe, P. Hoffmann, in Recognizing contextual polarity in phrase-level sentiment analysis, HLT/EMNLP 2005, Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing, Vancouver, British Columbia, Canada, 6-8 October 2005, Association for Computational Linguistics, **2005**, p. 347-354.
- [97] S. Brody, N. Diakopoulos, in Cooooooooooooooooo!!!!!!!!!!!!!!!: using word lengthening to detect sentiment in microblogs, Conference on Empirical Methods in Natural Language Processing (EMNLP 2011), Edinburgh, UK, 27-31 July 2011, Association for Computational Linguistics, **2011**, p. 562-570.
- [98] S. Almatarneh, P. Gamallo, in Automatic Construction of Domain-Specific Sentiment Lexicons for Polarity Classification, 15th International Conference (PAAMS 2017), Porto, Portugal, 21-23 June 2017, Springer International Publishing, **2018**, p. 175-182.
- [99] N. Oliveira, P. Cortez, N. Areal, Stock market sentiment lexicon acquisition using microblogging data and statistical measures, Decision Support Systems, 85 (**2016**) 62-73.
- [100] M. Z. Asghar, S. Ahmad, M. Qasim, S. R. Zahra, F. M. Kundi, SentiHealth: creating health-related sentiment lexicon using hybrid approach, Springerplus, 5 (**2016**) 1139.
- [101] K. S. Vishnu, T. Apoorva, D. Gupta, in Learning domain-specific and domain-independent opinion oriented lexicons using multiple domain knowledge, Seventh International Conference on Contemporary Computing (IC3 2014), Noida, India, 7-9 August 2014, IEEE, **2014**, p. 318-323.
- [102] K. Ergün, Metin Madenciliği Yöntemleri ile Ürün Yorumlarının Otomatik Değerlendirilmesi, Doktora Tezi, Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Sakarya, **2012**.
- [103] K. Oflazer, M. Saraçlar, Turkish and Its Challenges for Language and Speech Processing. Turkish Natural Language Processing, K. Oflazer, M. Saraçlar (Eds), Springer, Cham, Switzerland, 1-19, **2018**.

- [104] K. Oflazer, Türkçe ve Doğal Dil İşleme, Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 5 (2012).
- [105] F. S. Çetin, G. Eryiğit, Türkçe Hedef Tabanlı Duygu Analizi İçin Alt Görevlerin İncelenmesi-Hedef Terim, Hedef Kategori ve Duygu Sınıfı Belirleme, Bilişim Teknolojileri Dergisi, 11 (2018) 43-56.
- [106] U. Eroğul, Sentiment analysis in Turkish, Ms Thesis, Middle East Technical University, Ankara, 2009.
- [107] A. G. Vural, B. B. Cambazoglu, P. Senkul, Z. O. Tokgoz, in A Framework for Sentiment Analysis in Turkish: Application to Polarity Detection of Movie Reviews in Turkish, Computer and Information Sciences III, London, Springer London, 2013, p. 437-445.
- [108] Ç. Aytekin, An opinion mining task in Turkish language: A model for assigning opinions in Turkish blogs to the polarities, Journalism and Mass Communication, 3 (2013) 179-198.
- [109] S. D. Kamvar, J. Harris, in We feel fine and searching the emotional web, Forth International Conference on Web Search and Web Data Mining (WSDM 2011), Hong Kong, 9-12 February 2011, ACM, 2011, p. 117-126.
- [110] M. Kaya, G. Fidan, I. H. Toroslu, in Sentiment analysis of turkish political news, International Conferences on Web Intelligence (WI 2012), Macau, China, 4-7 December 2012, IEEE Computer Society, 2012, p. 174-180.
- [111] G. Yurtalan, M. Koyuncu, Ç. Turhan, A polarity calculation approach for lexicon-based Turkish sentiment analysis, Turkish Journal of Electrical Engineering & Computer Sciences, 27 (2019) 1325-1339.
- [112] R. Dehkharghani, Y. Saygin, B. Yanikoglu, K. Oflazer, SentiTurkNet: a Turkish polarity lexicon for sentiment analysis, Language Resources and Evaluation, 50 (2015) 667-685.
- [113] Ö. Çetinoğlu, O. Bilgin, K. Oflazer, Turkish WordNet. Turkish Natural Language Processing, K. Oflazer, M. Saraçlar (Eds), Springer, Cham, Switzerland, 317-336, 2018.
- [114] C. M. Özsert, A. Özgür, in Word polarity detection using a multilingual approach, Computational Linguistics and Intelligent Text Processing - 14th International Conference (CICLing 2013), Samos, Greece, 24-30 March 2013, Springer, 2013, p. 75-82.
- [115] A. Mattelart, M. Mattelart, İletişim Kuramları Tarihi, 8 ed., (çev: M. Zıllıoğlu), İletişim Yayınları, İstanbul, 7, 2016.
- [116] G. Gönenli, P. Hürmeriç, Sosyal medya: Bir alan çalışması olarak facebook'un kullanımı. Sosyal Medya Akademi, T. Kara, E. Özgen (Eds), Beta Yayınları, İstanbul, 213-242, 2012.
- [117] K. Duman, Yeni Medya Çağında Haberleri Yeniden Düşünmek: Teknoloji ile İçeriğin Kesişmesi ve Yenilikçi Davranışlar, Uluslararası Sosyal ve Eğitim Bilimleri Dergisi, 4 (2017) 107-119.
- [118] M. Deuze, The web and its journalisms: considering the consequences of different types of newsmedia online, New Media & Society, 5 (2003) 203-230.
- [119] H. Yurdakul, Kitle iletişiminde haber ve etki: Liberal ve eleştirel yaklaşımlar, Nosyon: International Journal of Society and Culture Studies, (2018) 62-77.

- [120] S. Altun, Haber söylemi, deęişikleri: Cumhuriyet mitingleri ya da gelincik devrimi örnekleme, Yüksek Lisans Tezi, İstanbul Kültür Üniversitesi Sosyal Bilimler Enstitüsü, **2007**.
- [121] N. Güz, Haberde Yönlendirme ve Kamuoyu Araştırmaları, Nobel Yayınevi, Ankara, 60-63, **2005**.
- [122] N. Godbole, M. Srinivasaiah, S. Skiena, Large-Scale Sentiment Analysis for News and Blogs, ICWSM, 7 (**2007**) 219-222.
- [123] J. C. S. dos Rieis, F. B. de Souza, P. O. S. V. de Melo, R. O. Prates, H. Kwak, J. An, in Breaking the news: First impressions matter on online news, Ninth International AAAI Conference on Web and Social Media, Oxford, UK, 26-29 May 2015, AAAI Publications, **2015**, p. 357-366.
- [124] C. Strapparava, R. Mihalcea, in Semeval-2007 task 14: Affective text, 4th International Workshop on Semantic Evaluations, SemEval@ACL 2007, Prague, Czech Republic, 23-24 June 2007, ACL, **2007**, p. 70-74.
- [125] M. U. Salur, Büyük veri araçlarından hadoop kullanarak veri madencilięi, Yüksek Lisans Tezi, Pamukkale Üniversitesi Fen Bilimleri Enstitüsü, Denizli, **2016**.
- [126] A. Balahur, R. Steinberger, M. Kabadjov, V. Zavarella, E. Van Der Goot, M. Halkia, B. Pouliquen, J. Belyaeva, Sentiment analysis in the news, International Conference on Language Resources and Evaluation (LREC 2010), 17-23 May 2010, Valletta, Malta, **2013**.
- [127] G. Ertem, Y. C. Dönmez, E. Oksel, Günlük gazetelerde hemşirelik haberlerin incelenmesi, Maltepe Üniversitesi Hemşirelik Bilim ve Sanatı Dergisi, 2 (**2009**) 70-75.
- [128] S. Atan, Y. Çınar, Borsa İstanbul'da Finansal Haberler ile Piyasa Deęeri İlişkisinin Metin Madencilięi ve Duygu (Sentiment) Analizi ile İncelenmesi, Ankara Üniversitesi SBF Dergisi, 74(1) (**2019**) 1-34.
- [129] K. H.-Y. Lin, C. Yang, H.-H. Chen, in Emotion classification of online news articles from the reader's perspective, IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Sydney, 9-12 December 2008, IEEE Computer Society, **2008**, p. 220-226.
- [130] T. Xu, Q. Peng, Y. Cheng, Identifying the semantic orientation of terms using S-HAL for sentiment analysis, Knowledge-Based Systems, 35 (**2012**) 279-289.
- [131] X. Bai, Predicting consumer sentiments from online text, Decision Support Systems, 50 (**2011**) 732-742.
- [132] Y. Rao, J. Lei, L. Wenyan, Q. Li, M. Chen, Building emotional dictionary for sentiment analysis of online news, World Wide Web, 17 (**2014**) 723-742.
- [133] Michigan University, <http://web.eecs.umich.edu/~mihalcea/affectivetext/> (Erişim tarihi: 7 Nisan 2019).
- [134] K. Leetaru, P. A. Schrodt, in Gdelt: Global data on events, location, and tone, 1979–2012, ISA Annual Convention, San Francisco, California, USA, 3-6 April 2013, Citeseer, **2013**, p. 1-49.
- [135] GDELT, <https://blog.gdeltproject.org/gdelt-translingual-translating-the-planet/> (Erişim tarihi: 15 Nisan 2019).

- [136] GDELT, <https://blog.gdelproject.org/introducing-the-global-content-analysis-measures-gcam/> (Eriřim tarihi: **15 Nisan 2019**).
- [137] J. L. Fleiss, Measuring nominal scale agreement among many raters, *Psychological Bulletin*, 76 (1971) 378.
- [138] Ö. B. Bilgen, N. Doğan, Puanlayıcılar Arası Güvenirlik Belirleme Tekniklerinin Karşılaştırılması* The Comparison of Interrater Reliability Estimating Techniques, *Journal of Measurement Evaluation in Education Psychology*, 8 (2017) 63-78.
- [139] J. R. Landis, G. G. Koch, The measurement of observer agreement for categorical data, *Biometrics*, (1977) 159-174.
- [140] A. A. Akın, M. D. Akın, Zemberek, an open source nlp framework for Turkic languages, *Structure*, 10 (2007) 1-5.
- [141] <https://www.dilbilgisi.net> (Eriřim tarihi: **14 Aralık 2017**).
- [142] <http://www.es-anlam.com> (Eriřim tarihi: **25 Aralık 2017**).
- [143] supersozluk, www.supersozluk.com (Eriřim tarihi: **27 Aralık 2017**).
- [144] YTU Kemik Dil Grubu, <http://www.kemik.yildiz.edu.tr/?id=28> (Eriřim tarihi: **25 Aralık 2017**).
- [145] onedio, <https://onedio.com/haber/atm-ye-sahte-para-yatirip-gercek-para-ceken-kalpazanlar-yakalandi-enflasyonu-dusurecektik-birakmadiniz-869855> (Eriřim tarihi: **20 Nisan 2019**).
- [146] onedio, <https://onedio.com/haber/piyanonun-parlak-cocugu-dunya-birinciligi-gururu-yasayan-9-yasindaki-nisan-italya-daki-yarismaya-hazirlaniyor-869955> (Eriřim tarihi: **20 Nisan 2019**).
- [147] onedio, <https://onedio.com/haber/kanseri-yenen-ve-aylar-sonra-okuluna-donen-ogretmen-sevgi-seliyle-karsilandi-869678> (Eriřim tarihi: **20 Nisan 2019**).
- [148] A. Özçil, Shewart, Cusum ve Ewma Kontrol Grafiklerinin Bir Üretim İşletmesinde Uygulanması, Yüksek Lisans Tezi, Pamukkale Üniversitesi, Sosyal Bilimler Enstitüsü, Denizli, **2014**.
- [149] S. Patır, İstatistiksel proses kontrol teknikleri ve kontrol grafiklerinin Malatya'daki bir tekstil (iplik dokuma) işletmesinde bobin sarım kontrolüne uygulanması, *Sosyal Ekonomik Araştırmalar Dergisi*, 9 (2009) 231-250.
- [150] W. E. Company, Statistical quality control handbook, Western Electric Company, The Company, **1956**.
- [151] D. C. Montgomery, Introduction to statistical quality control, John Wiley & Sons, Pennsylvania State University, **2009**.
- [152] Habertürk, <https://www.haberturk.com/gundem/haber/1216333-sirnak-sehidi-mehmet-gungordunun-cebinden-cikan-mektup-yurekleri-dagladi> (Eriřim tarihi: **23 Nisan 2019**).
- [153] egitimhane, <https://www.egitimhane.com/1-sinif-masal-kitaplari-pdf-formatinda-d203605.html> (Eriřim tarihi: **28 Nisan 2019**).

EKLER

EK 1 - HABER-4 Haberinin Cümle Listesi

1. Şırnak'ta PKK'lı teröristler tarafından polis aracına düzenlenen roketatarlı saldırı sonucu şehit olan polis memuru Mehmet Güngördü'nün şehadetinden önce yazdığı ve cüzdanında sakladığı mektup okuyanları gözyaşlarına boğdu.
2. Şehit olacağını önceden hissedip bir mektup yazan şehit polis memuru Mehmet Güngördü'nün "Olur da bir gün şehit olursam" başlığı ile kaleme aldığı mektubu okuyanları ağlattı.
3. 9 Şubat tarihinde Şırnak merkezde teröristlerin roketli saldırısı sonucu şehit düşen polis memuru Mehmet Güngördü'nün cenazesi, Kayseri'nin Yeşilhisar ilçesi Güzelöz Mahallesi'nde düzenlenen törenle toprağa verilmişti.
4. Şehidin eşi Hülya Güngördü tarafından cüzdanında bulunan mektup ise okuyanları gözyaşlarına boğdu.
5. "SAKIN ÜZÜLMİYİN, AĞIT YAKIP İSYAN ETMEYİN"
6. Şehit polis memuru Mehmet Güngördü, şehadetinden önce kaleme aldığı mektubu katlayarak cüzdanında sakladı.
7. Eşi tarafından teslim alınan eşyaları arasındaki cüzdanda bulunan mektup okuyanları gözyaşlarına boğdu.
8. Şehit polisin mektubunda, "Olur da bir gün şehit olursam annem ve babam sakın üzülmeysin, eşime sahip çıkın.
9. Canım bitanecik eşim kendini sakın üzme, aileme, kardeşlerime ve kendine dikkat et. Sesimi duyar mısınız bilmiyoruz Cumhurbaşkanımız, Başbakanımız ve milletvekilleri polisine ve askerine sahip çık.
10. Ölmekten bir korkum yok, Allah'ıma binlerce şükür imanımız ve itikadımız yerinde, vatan sağ olsun.
11. Hakkı olan hakkını helal etsin, benim her kimde hakkım varsa helal olsun.
12. Ölmekten bir korkum yok demiştim ya, sadece kahpece ölmek zoruma gider.

13. Mayına basıp, şehadete ulaşmak zoruma gider, çünkü mayına bastığımızda devletimiz gerekli tedbiri almış olsaydı (Shortlantların altı zırhlı olsaydı) ben ve benim gibiler ölmemiş olurdu.
14. Yine de dediğim gibi devletimiz, milletimiz ve vatanımız sağ olsun.
15. Canımdan çok sevdiğim annem, babam, eşim, kardeşlerim ve yeğenlerim sizler sakın üzülmeyin, ağıtlar yakmayın, isyan etmeyin.
16. Allah'a daima şükredin.
17. Akrabalarım, dayılarım, amcam, teyzelerim, halam ve ailem üç günlük dünya için kimsenin kalbini kırmayın, hiçbir işe yaramaz ve değmez.
18. Polislik mesleğimde de o kadar ismini tek tek yazmaya sayfalar yetmeyecek kardeşlerim, ağabeylerim hepinizden Allah razı olsun" ifadelerini kullandı.
19. "ELVEDA TATLI BEBEĞİM"
20. Mektubunun bir kısmını eşi Hülya Güngördü'ye ayıran şehit polis memuru Mehmet Güngördü, eşini ne kadar çok sevdiğini yazarak, bundan sonraki hayatında mutlu olmasını diledi.
21. Şehidin mektubunda eşi için kaleme aldığı satırlar şöyle:
22. "Canım eşim seni çok seviyorum.
23. Acı tatlı günlerimiz oldu.
24. Allah'ım senden razı olsun inşallah.
25. Bensiz hayatına devam et. Hayatını akışına bırak.
26. Ben sensiz hayat düşünmedim.
27. Tanıdığım güne vesile olanlara ve Rabbime şükürler olsun.
28. Ailene ve aileme her zaman sahip çık.
29. Kimseyi üzme, kimsenin de seni üzmesine müsaade etme.
30. Sen zaten güçlü bir kişilik sahibisin.
31. Elveda tatlı bebeğim.
32. Vatan sağ olsun.
33. Vatan sağ olsun.

34. Vatan sađ olsun."

35. Őehit polis memuru Mehmet G ng rd , mektubunun son kısmına bir not d şerek mezarının annesi ve babasının uygun g rd đ  yere g m lmesi vasiyetinde de bulundu.



EK 2 - PAMUK PRENSES Kitabının Cümle Listesi

1. Uzun zaman önce, uzak bir ülkede yaşayan, genç ve güzel bir prenses varmış.
2. Onun adı Pamuk Prenses'miş.
3. Pamuk Prenses'in üvey annesi olan Kraliçe, acımasız ve kibirli biriymiş.
4. Her zaman en güzel olmak istermiş.
5. Kraliçenin sihirli bir aynası varmış.
6. Her gün aynaya sorarmış: Ayna, ayna, söyle bana en güzel kim bu ülkede?
7. Ayna da hep şöyle cevap verirmiş: En güzel sensin, Kraliçem.
8. Pamuk Prenses büyüdükçe daha da güzelleşiyormuş.
9. Ama kıskanç Kraliçe, Pamuk Prenses'e sürekli eski elbiseler giydiriyormuş ve onu sabahtan akşama kadar çalıştırıyormuş.
10. Kraliçe bir gün yine aynaya sormuş: Ayna, ayna en güzel kim? Söyle bana.
11. Ayna bu defa: En güzel Pamuk Prenses, deyivermiş.
12. Buna çok sinirlenen Kraliçe: O hemen yok edilmeli, diye bağırmış.
13. Kraliçe adamlarından birine Pamuk Prenses'i ormana götürüp orada öldürmesini emretmiş.
14. Ama adam prensesi öldürmeye kıyamamış.
15. Onu ormanda bırakıp gelmiş.
16. Ormanda yalnız kalan Pamuk Prenses çok korkmuş ve ağlamaya başlamış.
17. Onun bu halini gören sevimli hayvanlar hemen etrafına toplanmışlar.
18. Sevimli hayvanlar onu küçük şirin bir kulübeye götürmüşler.
19. Fakat kulübenin içindeki minik masa ve sandalyeler tozla kaplıymış.
20. Pamuk Prenses Belki burada bakıma ihtiyacı olan yetim çocuklar yaşıyordur.
21. Etrafı temizlersem burada kalmama izin verebilirler demiş ve her yeri süpürüp temizlemiş.
22. Bu kulübede Yedi Cüceler yaşarmış.

23. Onlar gündüzleri, çalışmak için madene gider akşam olunca evlerine dönerlermiş.
24. Yedi cüceler evlerini derli toplu ve tertemiz bulunca şaşırılmışlar.
25. Yukarı çıktıklarında Pamuk Prenses'i görünce şaşkınlıkları daha da artmış.
26. - Uyuyan bir kız, demiş Bilgin Cüce.
27. - O çok güzel, demiş Utangaç.
28. - Bu kız başımıza bela olacak.
29. Uyarmadı demeyin demiş Huysuz.
30. Sesleri duyan prenses uyanmış.
31. Pamuk Prenses başından geçenleri cücelere anlatmış.
32. Kızın haline üzülen cüceler, ona bizimle kalabilirsin demişler.
33. Pamuk Prenses buna çok sevinmiş.
34. Yemekten sonra Pamuk Prenses ve cüceler şarkı söyleyip eğlenmişler.
35. Pamuk Prenses yeni arkadaşlarını çok sevmiş.
36. Artık kendini güvende hissediyormuş.
37. Fakat Kraliçe, sihirli aynadan Pamuk Prenses'in hâlâ yaşadığını öğrenmiş.
38. Yaşlı bir seyyar satıcı kadın kılığına girmiş.
39. Pamuk Prenses'i sonsuza kadar uyutmak için büyülü bir elma hazırlamış.
40. Ertesi sabah cüceler madene gitmek için evden ayrılırken, Pamuk Prenses'i dikkatli olması için uyarmışlar.
41. Cüceler gidince kraliçe hemen eve gelmiş.
42. Zavallı Pamuk Prenses yaşlı kadın kılığındaki kraliçeye inanmış.
43. Kraliçenin uzattığı elmayı ısırınca olduğu yere yığılmış.
44. Sevimli hayvanlar bunu cücelere anlatmışlar.
45. Yedi cüceler hemen eve dönmüşler.
46. Kraliçe'nin kaçtığını görünce onu kovalamışlar.
47. Kraliçe kaçarken bastığı taşa yıldırım düşmüş.

48. Aşağı yuvarlanan Kraliçe ölmüş.
49. Eve dönen cüceler Prensesin yerde yattığını görünce onun öldüğünü düşünmüşler.
50. Onu camdan bir tabuta koyup gece gündüz yanında beklemeye başlamışlar.
51. Bir gün ormandan geçen bir prens, Pamuk Prenses'in tabutunu görmüş.
52. Prensesin güzelliğini görünce dayanamamış ve onu hafifçe öpmüş.
53. Bunun üzerine büyü bozulmuş ve Pamuk Prenses uyanmış.
54. Prens ve Prenses evlenip mutlu bir hayat sürmüşler.



EK 3 - Tezden Türetilmiş Yayınlar

F. Sağlam, B. Genç, H. Sever, Extending a Sentiment Lexicon with Synonym-Antonym Datasets: SWNetTR++, Turkish Journal of Electrical Engineering and Computer Sciences, 27 (2019) 1806-1820.



EK 4 - Tezden Türetilmiş Bildiriler

F. Sağlam, B. Genç, H. Sever, Developing Turkish Sentiment Lexicon for Sentiment Analysis Using Online News Media, IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA), Nov 29 – Dec 02 2016, Agadir, Morocco, 2016.





HACETTEPE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
YÜKSEK LİSANS/DOKTORA TEZ ÇALIŞMASI ORJİNALLİK RAPORU

HACETTEPE ÜNİVERSİTESİ
FEN BİLİMLER ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI BAŞKANLIĞI'NA

Tarih: 29/05/2019

Tez Başlığı / Konusu: **Otomatik Duygu Sözlüğü Geliştirilmesi ve Haberlerin Duygu Analizi**

Yukarıda başlığı/konusu gösterilen tez çalışmamın a) Kapak sayfası, b) Giriş, c) Ana bölümler d) Sonuç kısımlarından oluşan toplam 122 sayfalık kısmına ilişkin, 29/05/2019 tarihinde ~~şahsım~~/tez danışmanım tarafından *Turnitin* adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı % 1'dir.

Uygulanan filtrelemeler:

- 1- Kaynakça hariç
- 2- Alıntılar hariç/~~dehşil~~
- 3- 5 kelimeden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü Tez Çalışması Orjinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve bu Uygulama Esasları'nda belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.


29/05/2019

Adı Soyadı: Fatih SAĞLAM
Öğrenci No: N12161440
Anabilim Dalı: BİLGİSAYAR MÜHENDİSLİĞİ
Programı: BİLGİSAYAR MÜHENDİSLİĞİ
Statüsü: ☐ Y.Lisans ☐ Doktora ☒ Bütünleşik Dr.

DANIŞMAN ONAYI

UYGUNDUR.



Dr. Öğr. Üyesi Fuat AKAL



HACETTEPE UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING
THESIS/DISSERTATION ORIGINALITY REPORT

HACETTEPE UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING
TO THE DEPARTMENT OF COMPUTER ENGINEERING

Date: 29/05/2019

Thesis Title / Topic: **Automated Sentiment Lexicon Generation and Sentiment Analysis of News**

According to the originality report obtained by myself/my thesis advisor by using the *Turnitin* plagiarism detection software and by applying the filtering options stated below on 29/05/2019 for the total of 122 pages including the a) Title Page, b) Introduction, c) Main Chapters, d) Conclusion sections of my thesis entitled as above, the similarity index of my thesis is 1 %.

Filtering options applied:

1. Bibliography/Works Cited excluded
2. Quotes excluded / ~~included~~
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Science and Engineering Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.


29/05/2019

Name Surname: Fatih SAĞLAM
Student No: N12161440
Department: COMPUTER ENGINEERING
Program: COMPUTER ENGINEERING
Status: ☐ Masters ☐ Ph.D. ☒ Integrated Ph.D.

ADVISOR APPROVAL

APPROVED.



Asst. Prof. Dr. Fuat AKAL

ÖZGEÇMİŞ

Kimlik Bilgileri

Adı Soyadı : Fatih SAĞLAM
Doğum Yeri : Ankara
Medeni Hali : Evli
E-posta : fsaglam2002@yahoo.com
Adresi : Bağlıca Mah. Aymer Cad. 2/11 Ankara

Eğitim

Lisans : Gazi Ünv. Tek.Eğt.Fak. Bilgisayar Öğ.liği, Ankara (1990-1994)
Y. Lisans : Beykent Ünv. Eğitim Yönetim ve Denetimi, İstanbul (2004-2005)
B. Doktora : Hacettepe Ünv. Bilgisayar Müh.liği, Ankara (2012-2019)

Yabancı Dil

İngilizce : KPDS - 82.5 (2012)

İş Deneyimi

Bilgisayar Öğretmeni : Maltepe Askeri Lisesi (1994-2004)
Kuleli Askeri Lisesi (2004-2006)
Öğretim Görevlisi : Kara Harp Okulu (2006-2017)

Deneyim Alanları

Programlama Dilleri, Doğal Dil İşleme, Çizge Teorisi, Duygu Analizi, Makine Öğrenmesi.