OTOMATİK DUYGU SÖZLÜĞÜ ÇEVİRİMİ VE DUYGU ANALİZİNDE KULLANIMI

AUTOMATIC SENTIMENT DICTIONARY TRANSLATION AND USING IN SENTIMENT ANALYSIS

Alaettin UÇAN

Prof. Dr. Hayri SEVER
Tez Danışmanı

Doç. Dr. Ebru AKÇAPINAR SEZER
İkinci Danışman

Hacettepe Üniversitesi
Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin
Bilgisayar Mühendisliği Ana Bilim Dalı için Öngördüğü
YÜKSEK LİSANS TEZİ olarak hazırlanmıştır

2014

Alaettin UÇAN'ın hazırladığı "**Otomatik Duygu Sözlüğü Çevirimi ve Duygu Analizinde Kullanımı**" adlı bu çalışma aşağıdaki jüri tarafından **Bilgisayar Mühendisliği Ana Bilim Dalı**'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Yrd. Doç. Dr. Sevil ŞEN AKAGÜNDÜZ
Başkan ……...

Prof. Dr. Hayri SEVER
Danışman ……...

Yrd. Doç. Dr. Ahmet Burak CAN
Üye ……...

Yrd. Doç. Dr. Erhan MENGÜSOĞLU
Üye ……...

Öğretim Görevlisi Dr. Fuat AKAL
Üye ……...

Bu tez Hacettepe Üniversitesi Fen Bilimleri Enstitüsü tarafından **YÜKSEK LİSANS TEZİ** olarak onaylanmıştır.

Prof. Dr. Fatma SEVİN DÜZ

Fen Bilimleri Enstitüsü Müdürü

Sevgili Eşime ve Kızıma

ETİK

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

- 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 bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

15/12/2014

Alaettin UÇAN

ÖZET

OTOMATİK DUYGU SÖZLÜĞÜ ÇEVİRİMİ VE DUYGU ANALİZİNDE KULLANIMI

Alaettin UÇAN

Yüksek Lisans, Bilgisayar Mühendisliği Bölümü

Tez Danışmanı: Prof. Dr. Hayri SEVER

İkinci Danışman: Doç. Dr. Ebru AKÇAPINAR SEZER

Aralık 2014, 84 sayfa

İnsanlar günlük ya da dönemlik eylemlerinde karar vermek için, diğer insanların duygularını, deneyimlerini veya görüşlerini öğrenerek hareket etmek isterler. Bu eğilim, aynı zamanda tecrübe aktarımı ve toplumsal hafızanın genişlemesi anlamına gelir. Örneğin satın alınılacak bir ürüne ya da hizmete karar vermek, izlemek için film seçimi yapmak ve hatta seçimlerde oy verilecek adaya karar vermek bu tür kararlar arasında yer alır. Diğer insanların bir konu hakkında ne düşündükleri çoğu zaman önemlidir ve bu düşünceleri öğrenebilmek için o kişiyle konuşmak ya da o kişinin yazdıklarını okumak gerekir. Aslında, yazılara ulaşmak görece daha kolaydır. Ve günümüzün etkileşimli web uygulamaları ve sosyal medya ortamları sayesinde büyük bir kitlenin ürettiği büyük bir içeriğe erişmek daha da kolay hale gelmiştir. Ancak erişilen içeriğin hacmi arttıkça, tek tek okunup anlaşılması hatta istenen içeriğe ulaşılması bile insan eforu ve zamanı içinde gerçekleştirilemeyecek

boyutlardadır. Bu nokta, genel çerçevede içerik tabanlı erişim sistemlerinin, metin madenciliğinin ve doğal dil işleme konularının, daha özelde duygu analizi çalışmalarının ana motivasyonunu oluşturur.

Duygu Analizi, yazılı halde verilen bir ifadenin (belge, yorum, eposta vb.) yansıttığı duygunun bilgisayar yardımıyla otomatik olarak tespit edilmesi işidir. Bu çalışmaların Düşünce Madenciliği veya Duygu Sınıflama olarak adlandırıldığı da görülür. Tespit edilmesi hedeflenen duygu, yazarın ruh halini, konu hakkındaki düşüncesini, yapmak istediği vurguyu ya da yaratmak istediği etkiyi içerebilir.

Son 15 yıldır yapılmış olan Duygu Analizi çalışmaları uyguladıkları yaklaşıma göre, “sözlük veya derlem”, “istatistiksel veya makine öğrenmesi” olmak üzere iki ana gruba ayrılabilir. Türkçe için, birçok kez denenmiş olan makine öğrenmesi tekniklerinin aksine, duygu sözlüğü kullanarak duygu analizi çalışması yapılmadığı gözlenmiş ve bu eksikliği gidermek için doğrudan sözlük kullanarak Duygu Analizi gerçekleştirimi hedeflenmiştir.

Türkçe Dili’ne özgü hazırlanmış ve anlamlar arası ilişkileri ifade etmesi sağlanmış bir sözlük bulunmadığından ayrıca duygu ifadelerinin evrensel olabileceğinden yola çıkılarak İngilizce için hazırlanmış olan duygu sözlüğünün Türkçe’ye otomatik çevirisi yapılarak Türkçe Duygu Sözlüğü oluşturulmuştur. Bu çalışmanın önemi kullanımı için önceden veri üretimi ve eğitimi gerekli olmadan, alan bağımlılığı taşımadan ve karmaşıklık düzeyini alt seviyede tutarak duygu analizi gerçekleştirimini sağlayabilmektedir. Elde edilen sözlüğün tamlığı tamamen bu tezin hedefleri dışında tutulmuş, yapılan çevirinin doğruluğu ve sözlüğün Duygu Analizinde kullanabilir olduğu düzenlenen deneyler ile sınanmıştır.

Elde edilen sonuçların karşılaştırılması ve yorumlanabilmesi için, aynı deneyler makine öğrenimi yöntemleri ile de yinelenmiştir. Duygu sözlüğü kullanılarak elde edilen sonuçlar hem makine öğrenmesi yöntemlerinin sonuçlarıyla hem de diğer dillerdeki sözlük çalışmalarının sonuçlarıyla karşılaştırılmış ve geliştirilen yöntemin başarılı olduğu gözlenmiştir.

Anahtar Kelimeler: Türkçe Duygu Sözlüğü, Otomatik Çevirim, Duygu Analizi, Düşünce Madenciliği, Duygu, Tutum, Sosyal Medya, Doğal Dil İşleme, Metin Madenciliği

ABSTRACT

AUTOMATIC SENTIMENT DICTIONARY TRANSLATION AND USING IN SENTIMENT ANALYSIS

Alaettin UÇAN

Master of Science, Department of Computer Engineering

Supervisor: Prof. Dr. Hayri SEVER

Co-Supervisor: Assoc. Dr. Ebru AKÇAPINAR SEZER

December 2014, 84 pages

People want to decide in their daily and seasonal activities by referring other people's emotions, experiences or opinions. This tendency also means transfer of experience and expansion of communal memory. For instance, making a decision about a product or service to buy, a movie to watch or a candidate to elect can be counted in this manner. Other people's opinions on a particular subject is usually important and there is a necessity of having conversation with that person or reading his/her writings to learn it. On the other hand, accessing to writings is relatively easier. Moreover, accessing to massive content of crowds has become much more easier via todays web applications and social media. However, as the volume of this content rises, extraction and assessment of focused content becomes impossible by human effort in plausible time. Thus, this point particularly constitutes the main motivation of sentiment analysis besides content retrieval systems, text mining and natural language processing.

Sentiment analysis is the automatic detection of emotion in textual contents (e.g. document, comment, e-mail) by utilization of computer. This is sometimes named as opinion mining or sentiment classification. The emotion which is being detected can contain author's mood and his/her ideas about subject. Furthermore, it can involve the emphasis points or the effect he wants to create.

The sentiment analysis studies in last 15 years can be classified into two main groups: (1) "dictionary and collection", (2) "statistical or machine learning".

For Turkish, in contrast to well known machine learning methods, it has been observed that there exists no study via sentiment dictionary. Thus, it is targeted to make Sentiment Analysis by directly utilizing dictionary.

As there exist no Turkish dictionary which is designed for constituting inter relations between meanings and starting with the idea of "sentimental expressions are universal", an English Sentimental Dictionary has been automatically translated to Turkish. The main contribution of this study is to provide sentiment analysis by no use of prior training data and field dependence. Furthermore it requires low level complexity. The completeness of the dictionary is kept out of this study. The success of translations and the feasibility of created dictionary were evaluated with the experiments.

The same experiments were also conducted with machine learning methods in order to compare results. The results which were generated by use of sentimental dictionary were compared with both machine learning based methods and the results of other dictionary studies using other languages. As a result, the proposed method is assessed as successful.

Keywords: Turkish Sentiment Dictionary, Automatic Translation, Sentiment Analysis, Opinion Mining, Emotion, Attitude, Mood, Social Media, Natural Language Progressing, Text Mining

TEŞEKKÜR

Danışmanım Prof. Dr. Hayri SEVER' e ve desteğini esirgemeyen ikinci danışmanım Doç. Dr. Ebru AKÇAPINAR SEZER' e kıymetli vakitlerini ayırdıkları ve sabırla yol gösterdikleri için çok teşekkür ederim.

Çalışmalarımı destekleyen jüri üyeleri; Yrd. Doç. Dr. Sevil ŞEN AKAGÜNDÜZ'e, Yrd. Doç. Dr. Ahmet Burak CAN'a, Yrd. Doç. Dr. Erhan MENGÜSOĞLU'na ve Dr. Fuat AKAL'a değerli katkıları ve önerileri için teşekkür ederim.

Tecrübelerinden faydalandığım mesai arkadaşım Arş. Gör. A. Selman BOZKIR' a teşekkür ederim.

Ayrıca desteği ve hoşgörüsü için sevgili eşim Zeynep'e ve beni hayata daha sıkı bağlanmaya teşvik eden güleç kızım Asel'e çok teşekkür ederim.

İÇİNDEKİLER

Sayfa

ÇİZELGELER

Sayfa

ŞEKİLLER

Sayfa

EŞİTLİKLER

SİMGELER VE KISALTMALAR

ML	Machine Learning (Makine Öğrenimi)
ME	Maximum Entropy
NB	Naïve Bayes
OM	Opinion Mining (Düşünce Madenciliği)
SA	Sentiment Analysis (Duygu Analizi)
SVM	Support Vector Machine (Destek Vektör Makinesi)
Random Forest	Rastgele Orman
Synset	Synonym Set (Anlamdaş kümesi)
Hyponym	Dilbilimde altanlamlık ilişkisi
Meronymy	Dilbilimde parça bütün ilişkisi
Troponymy	Dilbilimde alt eylem ilişkisi örneğin “yeme – atıştırma” arasındaki ilişki
POS	Part Of Speech – Kelimenin dilbilgisi açısından türü
TDSs	Seri Türkçe Duygu Sözlüğü,
TDSp	Paralel Türkçe Duygu Sözlüğü
TDSv1	Türkçe Duygu Sözlüğü sürüm 1
Prior Polarity	Öncül Yön Puanı
DDİ	Doğal dil işleme
TP	True Positive (Doğru Pozitif)
TN	True Negative (Doğru Negatif)
FP	False Positive (Yanlış Pozitif)
FN	False Negative (Yanlış Negatif)
WEKA	Waikato Environment for Knowledge Analysis

1.GİRİŞ

"Sen anılması güzel bir söz ol. Çünkü insan kendisi hakkında söylenen güzel sözlerden ibarettir."

-Hz. Mevlana

Duygu, bireyin ruh halinde biyokimyasal (içsel) ve çevresel tesirlerle etkileşiminden doğan kompleks psikofizyolojik bir değişimdir[1]. Geçmişten günümüze insanoğlu duygularını yazılı ve sözlü olarak ifade etmiştir. Edebi eserlerde rastlanan duygu içeren ifadeler okuyucuya yazarın ruh halini yansıtmaktadır. Okuyucu yazarın ifadelerinden konu ya da kişi hakkında hissettiklerini anlayabilir.

Eskiden sadece yazar ve edebiyatçı gibi işi yazmak olan profesyonellerin yazılarını okuma fırsatımız varken günümüzde internet ile yaratılan etkileşimli ortamda neredeyse herkesin yazdıklarını okuyabilir hale geldiğimizi tespit edebiliriz. Günümüzde mikro-blog sitelerinin ve sosyal paylaşım ortamlarının artan sayısı, erişim hızı ve kolaylığı sayesinde, kişilerin birbirleriyle olan iletişimlerinin büyük çoğunluğu sanal ortamlarda gerçekleşmektedir. Kişi hayata bakış açısını, durum karşısında hissettiklerini, yaşantısındaki yenilikleri ve benzeri duygularını sanal ortamlarda paylaşmaktadır.

Kişiler çoğu zaman olumlu duyguları hissetmek ya da olumsuz duyguları hissetmemek için başkalarının benzer durumda ne hissettiği ile ilgilenir. Örneğin;

- Bir ürünü ya da hizmeti satın alan kişilerin yorumlarını okumak o ürünü ya da hizmeti satın aldığınızda size neyle karşılaşacağınıza dair ipuçları verir.
- Hizmeti ya da ürünü sunan kurum, müşteri yorumlarından müşteri memnuniyeti hakkında fikir sahibi olabilir.

[1] http://tr.wikipedia.org/wiki/Duygu

- Firmalar müşteri yorumları ile pazarlama stratejileri oluşturabilir ve geliştirebilir.
- Bir siyasi parti, adaylarını belirlerken sosyal medyada aday adayları hakkında yapılmış yorumları dikkate alabilir.
- Kişinin yazıları kişisel duygu durumunu gösterirken, toplumun yazıları ülkenin duygu haritasını gösterebilir.
- Halkın herhangi bir konuda memnuniyeti hakkında fikir sahibi olunabilir.
- Sosyal psikoloji ile ilgili araştırmalar yapılabilir.
- Bir firmanın hem basında hem sosyal medyadaki itibarı ölçümlenebilir.
- Kurumun ya da devletin yönetimi ile ilgili anketlerle elde edilebilecek birçok bilgi elde edilebilir.

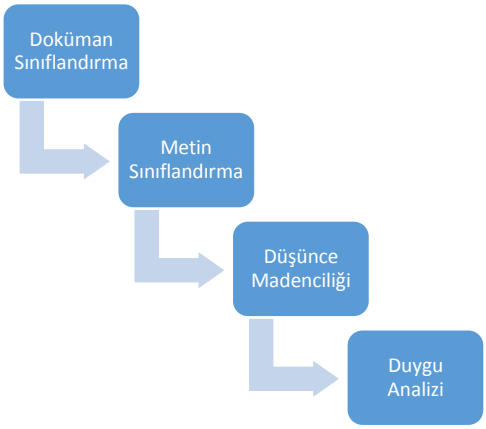
Sonuç olarak, kişilerin ürün, hizmet, kurum veya şahıs hakkında yazdıkları yorumlar hissettikleri duyguları ya da genel kanaatlerini tasvir eder ve bu bilgiyi kullanarak yeni bilgilere ulaşmak mümkündür.

1.1. Problem Tanımı

İlk kez 19. yüzyılın son çeyreğinde anılmaya başlanan doküman sınıflama yöntemleri sayesinde belge, metin ve mektup gibi benzer içerikli dokümanlar sınıflara ve kategorilere otomatik veya algoritmik olarak ayrılabilir hale gelmiştir. Bilgiye erişim konusunda ilk ve en çok işlenen konulardan metin sınıflama problemi zamanla doküman içerisindeki ana fikri tanımlamaya doğru yönelmiştir. Belge içerisinden fikri ayıklayarak sınıflayabilen yöntemler geliştikçe belgenin boyutu küçülmüş ve kısa metinlere dönüşmeye başlamıştır. Zamanla iş cümlenin veya bir paragrafın ana fikrini ayıklamaya kadar indirgenmiştir.

Milenyumla birlikte araştırmacılar metin içerisinde sadece fikir değil duyguların da barındığı düşüncesine kapılmış ve belgenin öznel bilgiler içerip içermediği sorusuna cevap aramışlardır. *“Oltayı atıp saatlerce beklemek bana zaman kaybı gibi geliyor.”* Cümlesindeki gibi öznel ifadelerin tespit edilmesinin mümkün olmasıyla birlikte o ifadelerin barındırdığı duygunun içeriği ve belgeye kattığı anlam değerlendirilmeye başlanmıştır. Araştırmacılar büyük belgelerden başlayarak kısa metinlerin, paragrafların hatta ifadelerin duygularını belirlemeye çalışmışlardır.

Duygu Analizi (*Sentiment Analysis*), başka bir deyişle Düşünce Madenciliği (*Opinion Mining*), otomatik ya da yarı otomatik olarak, cümlenin, belgenin veya ifadenin yansıttığı duyguyu tespit edip, yorumlama işlemine verilen addır. Şekil 1 ile duygu analizinin kronolojik gelişimi gösterilmektedir.



Şekil 1. Duygu Analizinin Evrimi

Günümüzde tek başına 800 milyon aktif kullanıcıya ve yaklaşık 300 Petabayt (2^{50} bayt) veriye sahip olan Facebook gibi dev sosyal paylaşım siteleri, 150 milyon aktif kullanıcıya ve 700 milyon ürüne sahip eBay gibi dev e-pazarlama siteleri internette hizmet sunmaktadır. Bunca veri içerisinden gerekli olan ürün ya da hizmet hakkında, güvenilir ve gerçek kişiler tarafından yapılmış yorumlara ulaşmak ve bu yorumları analiz etmek insan yetenekleri ve limitleri ile neredeyse imkânsızken Duygu Analizi yöntemleri ile mümkündür.

Duygu Analizi, doğal dil ile yazılmış belgelerin bilgisayar yazılımları tarafından işlenerek istatistiksel ve algoritmik çözümler ile analiz yapılmasına dayanmaktadır. Doğal dilde ifadenin yeri, zamanı, oluşum süreci, kim tarafından yapıldığı gibi bilgiler yer almaktayken, bilgisayar ile bu bilgilerin kısıtlı bir bölümünü hesaplayıp analiz edebildiğimiz için bilginin bir kısmı kaybolmaya mahkûmdur. İncelenecek belgenin boyutu küçüldükçe bilgi kaybı artmakta, analiz yapmak oldukça zor olmaktadır. Duygu Analizi yaparken en sık karşılaşılan problemleri sıralanacak olursa;

- Günümüz sanal dünyasında sıkça karşılaşılan kısa cümlelerin analiz edilmesi,

- Yanlış yazılmış ya da kısaltılmış ifadelerin tanınıp analiz sürecine dâhil edilmesi,
- Olumsuz ifadelerin tespit edilerek analiz yapılabilmesi,
- Asıl anlamından uzak imalı cümlelerin gerçek anlamlarına kavuşturulup analiz edilmesi,
- Gerçek kişiler veya bilgisayarlar tarafından yazılmış reklam ve yanıltmaya yönelik cümlelerin ihmal edilerek analiz edilmesi.

İngilizce ve akraba dilleri için yapılan Duygu Analizi çalışmalarında karşılaşılan tüm bu zorlukların yanı sıra dilimize özel zorluklar da vardır.

- Türkçe'nin cümle yapısındaki farklar,
- Türkçe karakterlerde yapılan yazım yanlışları,
- Türklerin az kelimeyle çok şeyi ifade etme alışkanlığı,
- Toplumumuzun olumsuz duyguları daha çok dile getirme anlayışı.

Duygu analizi yaparken karşılaşılan en büyük problem ise analizin klasik konu tabanlı metin sınıflama problemine öykünerek yapılmasından kaynaklanan her bağlam için farklı öğrenme kümesi veya bir değerlendirme sözlüğü gereksinimidir. Çünkü metin sınıflama yöntemlerinde bir ifadenin, bir bağlamda metne kattığı anlam güçlü ve ayrıştırıcı özelliği yüksekken başka bir bağlamda bunun tam tersi olabilmektedir.

Örnek cümlelerden de anlaşılacağı üzere problemin çözüm tekniğinin metin sınıflama yöntemlerine benzerliğinden dolayı Duygu Analizi yöntemlerinde bir ifade, bir bağlamda olumlu duygu barındırırken başka bir bağlamda olumsuz duygulara karşılık gelebilmektedir. Bu nedenle bir bağlam içerisinde oluşturulmuş öğrenme kümesi ile üretilen model ile başka bir bağlamda Duygu Analizi yapmak çok da doğru sonuçlar elde edilemeyeceği anlamına gelmektedir.

- Taşınabilir diskin kapasitesinin aksine boyutu oldukça küçük.$^{+}$
- Aracın bagajı bir ailenin ihtiyaçlarını karşılayamayacak kadar küçük.$^{-}$

Araştırmacılar bahsedilen bağlama bağlı Duygu Analizi probleminin çözümü için iki farklı çözüm önermişlerdir. İlk çözümde her analiz kendi bağlamında değerlendirilerek yapılırken diğer çözümdeyse birçok bağlamı kapsayacak büyüklükte bir derlem oluşturulup derlemdeki tüm ifadeler için bir duygu puanı hesaplanmış ve analiz bu bağlam bağımsız sözlük vasıtasıyla yapılmıştır. Her derlem için bir öğrenme kümesi veya bir sözlüğe ihtiyaç duyulması Duygu Analizinin

kendi bağlamında yapılmasını güçleştirdiği için bağlam bağımsız sözlükler yoluyla Duygu Analizi yapmak daha avantajlıdır.

Duygu analizi için İngilizce'ye ait sözlükler WordNet [1] yapısı üzerine kurulmuştur. WordNet tanım itibarı ile İngilizce için yapılmış sözlüksel veritabanıdır. WordNet içerisinde isimler, fiiller, sıfatlar ve zarflar her biri farklı bir kavramı temsil edecek biçimde bilişsel eşanlamlı gruplara bölünmüştür. Sözlük içerisinde anlamdaş kümeleri birbiri ile ilişkilendirilmiştir. Duygu Analizi sözlüğü WordNet içerisindeki ilişkilerden faydalanarak genişletilmiş ve genel geçer bir Duygu Sözlüğü haline gelmiştir.

Çizelge 1. Yapılan WordNet çalışmaları

Language	WordNet
Albanian	Albanet
Arabic	Arabic Wordnet
Chinese	Chinese Wordnet (Taiwan)
Danish	DanNet
English	Princeton WordNet
Farsi/Persian	Persian Wordnet
Finnish	FinnWordNet
French	WOLF (Wordnet Libre du Francais)
Hebrew	Hebrew Wordnet
Italian	MultiWordNet
Japanese	Japanese Wordnet
Catalan	Multilingual Central Repository
Galecian	Multilingual Central Repository
Spanish	Multilingual Central Repository
Indonesian	Wordnet Bahasa
Malaysian	Wordnet Bahasa
Norwegian	Norwegian Wordnet
Polish	plWordNet
Portuguese	OpenWN-PT
Thai	Thai Wordnet

Çizelge 1'de görüldüğü üzere WordNet zamanla İngilizceye yapısal olarak yakın akraba dillere de uyarlanmıştır. Bu uyarlama yıllar sürmüştür. Bazı diller için başlatılan WordNet çalışmaları henüz tamamlanamamıştır. Türkçe için birkaç deneme yapılmış olsa da sonuç elde edilememiştir. Türkçe WordNet gibi bilişsel bir dilbilim sözlüğü ilişkisel veritabanı oluşturmak için büyük bir ekibin yıllarca

uğraşması gerekmektedir. Günümüz şartlarında Türk Dil Kurumu seviyesinde yapılması gereken, disiplinler arası bir iştir.

Her bağlam için ayrı bir işlem gerekmeksizin Duygu Analizi yapabilmek oldukça büyük bir derlemde oluşturulmuş bir Duygu Sözlüğü ile mümkündür. Türkçe için bu anlamda oluşturulmuş geniş ölçekli bir derlem bulunmamakla birlikte ifadeler arası ilişkileri değerlendirebileceğimiz bir WordNet yapısı da henüz mevcut değildir.

1.2. Çalışma Konusu ve Kapsamı

2000'li yılların başlarında, İnternetin popüler olmasıyla beraber Duygu Analizi araştırmaları başlamıştır. Doğal dil işleme (DDİ), makine öğrenmesi, veri madenciliği gibi disiplinlerin birleştirilmesi ve farklı bakış açısından değerlendirilmesi ile ortaya çıkan duygu analizi yöntemleri ile ilgili 16.000'e yakın çalışma yapılmıştır.

Var olan yaklaşımları; kelime yakalama, sözlüksel benzeşim ve istatistiksel yöntemler olmak üzere üç ana yaklaşımda gruplayabiliriz. Kelime yakalama yaklaşımı, metni etkili ve benzersiz kelimelerin varlığına göre sınıflamaktır. Sözlüksel benzeşim yaklaşımı, hazır bir duygu sözlüğü yardımıyla duygu puanlaması yapıp eşik değere göre sınıflama yapmaktır. Naïve Bayes (NB), Support Vector Machine (SVM) (Destek Vektör Makinesi) gibi makine öğrenmesi yöntemleri ile duygusal sınıflama yapmak ise istatistiksel yöntemlere örnektir. Bir başka yaklaşım olan karma tekniklerde ise kelime yakalama ile istatistiksel yöntemler, kelime yakalama ile sözlüksel benzeşim ya da hepsini bir arada kullanılarak duygu analizi yapılır.

Diğer yöntemlerin aksine sözlüksel benzeşim yönteminde duygu analizi yapmanın konu bağımsız, hızlı ve kolay analiz yapabilme imkânları vardır. Yine gözetimsiz makine öğrenmesi yöntemleri kullanarak duygusal sınıflama yapabilmek için hazır bir duygu sözlüğüne ihtiyaç vardır. İngilizce için oluşturulmuş ve akraba bazı dillere de uyarlanmış bir takım duygu sözlükleri bulunmakla birlikte Türkçe için hâlihazırda yapılmış bir duygu sözlüğü bulunmamaktadır.

Duygu sözlükleri genelde konuya özel hazırlanmış veri kümesi ile yarı gözetimli makine öğrenmesi teknikleri ve el yordamıyla oluşturulmaktadır. Bu süreç için oldukça fazla işgücü-zaman gerekmektedir. Oysa günümüzde İngilizce bir kelimenin Türkçe karşılığını bulmak nispeten kolaydır. Duyguların kültür ve dil bağımsız olduğunu varsayarak İngilizce için hazırlanmış bir duygu sözlüğünü bilgisayar

yardımıyla otomatik olarak Türkçeye çevirmek Türkçe Duygu Sözlüğü elde etmenin oldukça kolay bir yolu olabilir.

1.3. Amaç

Çalışma kapsamında daha önce oluşturulmamış olan Türkçe Duygu Sözlüğünü hazır olan başka bir dildeki duygu sözlüğünden otomatik tercüme ederek hazırlamak ve bu sayede sözlüksel benzeşim yöntemiyle duygu analizi altyapısını oluşturarak konu bağımsız, kolay ve hızlı bir şekilde duygu analizi yapmak hedeflenmiştir.

Yine çalışma kapsamında Türkçe için hazır olmayan duygu analizi veri kümesi temin altyapısını oluşturma ve ölçüm amaçlı birden fazla duygu veri kümesi oluşturma alt görevleri hedeflenmektedir.

1.4. Motivasyon ve Özgün Değer

İletişimin, eğlencenin, pazarlamanın, siyasetin ya da eğitimin yani neredeyse her faaliyetin bir şekilde internet yoluyla yapıldığı günümüzde duygu analizine gerçekten ihtiyaç vardır. Bir alıcı ya da satıcı için ürün veya hizmet hakkında yapılan yorumların ne derece “olumlu” olduğu nasıl önemli ise, bir siyasetçi için seçmenlerin kendi hakkında ne derece “olumlu” düşündükleri de bir o kadar önemlidir. Türkçe Duygu Sözlüğü ile yapılacak olan duygu analizleri insanların benzeri birçok ihtiyacını karşılayacaktır.

Bu çalışma duygu sözlüğü kullanma tecrübesini arttırıp, sonuçlarını duyurarak makine öğrenmesi ve istatistiksel yöntemler ile bir kez daha kıyaslayabilme imkânı sunmaktadır. Ayrıca ilk kez oluşturulan Türkçe Duygu Sözlüğü ve duygu analizi veri kümeleri Türk bilim dünyasına önemli bir katkı yapmaktadır.

Hazırlanan duygu analizi ortamı gerçek hayatta önemli bazı işlevleri yerine getirerek kritik uyarılarda bulunabilir. Örneğin sosyal medyayı takip edip duygu analizi yaparak yorumları toplayan bir marka analizi sistemi oluşturulabilir ve takip edilen kurum hakkında sosyal medyada bir anda olumsuz yorumlar çoğalmaya başladığında o kurum uyarılabilir. Siyasi bir partinin milletvekili adaylarını, ankete gerek kalmadan, sosyal medyada hakkında seçmenlerin daha çok olumlu yorumlar yaptığı aday adayları arasından belirleyebileceği bir itibar yönetimi sistemi oluşturulabilir.

2. LİTERATÜR ÖZETİ

2000'li yılların öncesinde metaforların yorumlanması, duygu sıfatları, öznellik, bakış açısı ve etkileri konularında bazı çalışmalar [2] yapılmış olsa da asıl çalışmalar sonrasında yapılmaya başlamıştır. Duygu Analizi [3] ve Fikir Madenciliği [4] terimleri ilk kez 2003 yılında ortaya çıkmıştır.

Etkili kelimelerin varlığına göre analiz yapan Elliott [5], Ortony ve ekibi [6] ve son olarak da Stevenson ve ekibi [7] ilkel duygu analizi yöntemleri ile fikir madenciliği yapmışlardır.

Duygu analizi, yöntemleri birleştirilerek yapılan hibrit çalışmalar olsa da makine öğrenmesi yöntemleri ve sözlüksel benzeşim yöntemleri olmak üzere ikiye ayrılmaktadır. Bu bağlamda yapılan çalışmalar makine öğrenmesi yöntemlerini kullanan çalışmalar, sözlük benzeşimi yöntemlerini kullanan çalışmalar ve Türkçe için yapılmış çalışmalar olmak üzere üç başlık altında incelenmiştir.

2.1. Makine Öğrenimi Yöntemleri ile Yapılan Çalışmalar

İstatistiksel yöntemler ve sınıflayıcılar ile farklı diller ve farklı veri kümeleri için yapılmış çalışmalar mevcuttur. Pang ve arkadaşları [8] Naïve Bayes (NB), Maksimum Entropy (ME) ve Support Vector Machine (SVM) (Destek Vektör Makinesi) algoritmaları yardımıyla, elle seçtikleri öznitelikler ve 1.400 yorumlu veri kümleri ile duygu analizi yapmışlar ve ortalama %82 başarı elde etmişlerdir.

Pang ve Lee [9] çoklu sınıf SVM regresyon algoritmaları ile %60 civarında başarı elde etmişlerdir. Çalışmada yıldız ve sembollerle etiketlenmiş verilerde sınıflama yapma problemi irdelenmiştir.

Hu ve Liu [10] ticari ürünler hakkında yapılan yorumları derlemiş olumlu ve olumsuz olmak üzere sınıflamışlardır. Yaptıkları sınıflama sonucunda %70 civarı başarı elde etmişlerdir.

Turney [11] ise web üzerinden topladığı 4 farklı konu hakkındaki 410 yorumu cümlelerin polaritesine göre PMI-IR algoritması ile sınıflamıştır. Turney çalışma sonucunda ortalama %74 başarı elde etmiştir.

Pang ve Lee [12] önce veriden öznel ifadeleri çıkarmış sonra bu ifadelerin polaritelerini hesaplamış ve SVM, NB sınıflayıcılar ile ortalama %85 başarı elde etmişlerdir.

Dave ve arkadaşları [13] veri toplama ve kırpma sistemlerine yerleştirmek üzere ürün yorumlarını otomatik olarak sentezleyen ve duygu analizi yapan bir araç tasarlamışlardır. Tasarladıkları araçta kendi sınıflayıcı algoritmaları ile %85 civarında sonuç elde etmişlerdir.

Çalışmaların birçoğu, her sınıftan tohum kelime listesi ile başlamakta ve sınıflama algoritmaları yardımıyla ilişkili cümleleri ve ya belgeleri sınıflayarak duygu analizi yapmaktadır. Başta verilen küçük tohum liste genelde tam kapsamı yakalayamamakta ve sınıflayıcı konu bağımlı bir sınıflama yapmaktadır. Başarıların değişkenlik göstermesinin sebebi sınıflamakta kullanılan özniteliklerin ve veri kümelerinin farklılıklarıdır.

2.2. Sözlüksel Benzeşim Yöntemleri ile Yapılan Çalışmalar

Sözlüksel benzeşim yöntemlerinde ise iki farklı bakış açısı vardır; duygu sözlüğü ve duygu derlemi. İstatistiksel yöntemlerden makine öğrenmesi ile duygu analizi yapılabilmesi için gerekli olan öznitelikleri elde etmek için bir derleme sahip olmak gerekmektedir. Bu derlemi oluştururken her sınıfa ait tohum kelimeler, her bağlam için yeniden oluşturulmaktadır. Duygu derlemi bakış açısı ile sadece bir bağlam için duygu içeren kelime listesi oluşturulmaktadır. Oluşturulan listelerin konu bağımlı olması, tohum listenin el yordamıyla hazırlanmasından dolayı için eksik ya da hatalı olma riski, vakit ve iş gücü kısıtlarından ötürü; kapsamlı ve doğru bir duygu sözlüğü oluşturma ihtiyacı ortaya çıkmıştır.

Hu ve ekibinin %84 başarı elde ettikleri çalışmalarında [10] el yordamıyla toplanmış bir miktar olumlu ve olumsuz tohum kelimeler listelenir. Zıt ve eş anlamları gösteren WordNet [14] gibi bir sözlük alınır ve liste tohum kelimelerin eş ya da zıt anlamlıları bulunarak genişletilir. Sonra yine el yordamıyla liste arındırılır. Buna benzer bazı araştırmalarda [15] [16] ise, son aşamada listedeki hatalı elemanları temizleme işlemi, kelimelerin temsil yeteneğini tespit eden istatistiksel yöntemler yardımıyla yapılmaya çalışılmıştır. Kelimelere gelen eklerle zıt anlam tespit etmeyi deneyen Mohamed ve ekibi [17] bazı zıt anlamları listeye eklemeyi başarmıştır. Yine başka bir teknikte ise Kamps ve ekibi [18] tıpkı Williams ve ekibinin [19] çalışmalarında olduğu gibi mesafe tabanlı algoritmalarıyla her kelimenin duygu yönünü, olumlu ve olumsuz elemanları bulunan bir tohum listesine olan mutlak uzaklıklarına göre tespit etmeyi denemiştir.

Steinberger ve arkadaşları [20] çalışmalarında birçok dilde duygu sözlüğü oluşturmak için yarı otomatik bir yöntem sunmaktadır. İki farklı dil için oluşturulmuş olan sözlükler otomatik yöntemlerle üçüncü dillere çeviri yapılmıştır. Çalışmada İngilizce, İspanyolca, Arapça, Çekçe, Fransızca, Almanca, İtalyanca ve Rusça üzerine çalışmalar yapılmıştır. Türkçe için denenmiş tamamlanamamıştır. Yaklaşık 2.000 civarı kelimelik sözlükler insan eliyle kontrol edilmiş ve ortalama %76 başarı sağlanmıştır.

Esuli ve Sebastiani [21] çalışmalarında Turney [11] ve Kamp'ın [18] tohum kelimelerini eğitim kümesi olarak kullanmış ve makine öğrenmesi yöntemleri ile sözlükteki kelimeleri negatif ve pozitif olarak sınıflamışlardır. Elde ettikleri kelimeleri tohum sözlere eklemiş tekrar bu işlemi yinelemeli olarak yapmışlardır. Sonunda oldukça geniş bir ikili vektör elde etmişlerdir. Sonrasında çalışmalarını genişletmişler [22], her anlamın negatif ve pozitif puanlarının yanı sıra nesnel puanlarını da hesaplamışlardır. Bunu yaparken kelimelerin eş anlam, zıt anlam ve altanlamlık ilişkilerinden faydalanmışlardır. Makine öğrenmesi teknikleri ile WordNet sözlüğündeki tüm kelimelere ait negatif, pozitif ve objektif puanları hesaplamışlar ve genel bir duygu sözlüğü olan SentiWordNet [23] sözlüğünü yayınlamışlardır. Yayınladıkları sözlüğü güncelleyerek günümüzde üçüncü sürümünü araştırmacılar ile paylaşmaktadırlar.

SentiWordNet sözlüğü yardımıyla yapılan çalışmalardan; Taboada ve ekibinin [24] çalışmasında Pang ve Lee'nin [8] veri kümesi kullanılarak film yorumlarının %66'sı doğru etiketlenebilmiştir. Ohana ve Tierney [25] ise 1000 film yorumu üzerinde yaptıkları deneyde %65,85 doğruluk elde etmişlerdir. Hamouda ve Rohaim [26] çalışmalarında ise %67 başarı elde etmişlerdir.

2.3. Türkçe için Yapılmış Duygu Analizi Çalışmaları

Türkçe duygu analizi tüm bu gelişmelerden birkaç yıl sonra gündeme gelmiş hazır sözlükler ve derlemler olmadığı için önce makine öğrenmesi yöntemleri ile yapılmış ve başarılı olmuştur. Alandaki ilk çalışmada Eroğul [27] makine öğrenmesi algoritmalarından SVM [28] ve doğal dil işleme teknikleri ile film yorumlarını N-gram model kullanarak olumlu ve olumsuz kategorilerde sınıflamış ve %85 başarı elde etmiştir.

Nizam ve arkadaşları [29] çalışmalarında gözetimsiz makine öğrenmesi yöntemleri ile duygu sınıflaması yapılmaktadır. Twitter verisi üzerinde çalışılmış tüm kelimeler öznitelik olarak seçilmiştir. ngram özellikleri kullanılarak yapılan sınıflama sonucunda eşit dağılımlı veri kümesi ile yapılan deney dengesiz veri kümesiyle yapılan deneyden daha başarılı olmuş ve en fazla %72 başarı elde edilmiştir.

Boynukalın [30] İngilizce bir veri kümesini insan eliyle Türkçeye çevirip ve bir takım Türkçe verileri de elle etiketleyerek elde ettiği veriler üzerinde makine öğrenmesi teknikleri ile sınıflamalar yapmış ortalama %80 civarı başarı elde etmiştir.

Vural ve ekibi [31] yaptıkları çalışmada İngilizce için yapılmış ve birçok dile çevrilmiş olan SentiStrength [32] altyapısını kullanarak duygu analizi yapmıştır. Uygulamanın kullandığı yaklaşık 1000 kelimelik sözlüğü elle çevirmiş ve makine öğrenimi yöntemleri ile sınıflama yaparak duygu analizini gerçekleştirmiştir.

Meral ve Diri [33] yaptıkları çalışmada 8.500 civarı Twiti elle etiketlemiş, kelime yakalama bakış açısı ve makine öğrenmesi yöntemlerinden Rastgele Orman (*Random Forest*), Destek Vektör Makinesi (SVM) ve Naïve Bayes (NB) sınıflayıcılarıyla Duygu Analizi yapmışlardır. Korelasyon tabanlı öznitelik seçimi yaparak %90 civarında başarı elde etmişlerdir.

Mayda ve Aytekin [34] çalışmalarında sosyal medyada rekabet analizi için karşılaştırma görevine yönelik bir fikir madenciliği modeli geliştirmiştir. Bu amaçla karşılaştırma siteleri, YouTube ve teknoloji forumlarından iz sürme tekniği ile karşılaştırma ifadesi içeren 100 yorum manuel olarak derlenmiş ve bu yolla bir test veri tabanı oluşturulmuştur. Sadece anma metriği ile sunulan sonuçlara göre sistem %70 civarı başarılı olmuştur.

Çakmak ve arkadaşları [35] çalışmalarında Türk dili için kelime kökleri ve cümle bazında duygu ilişkilerini incelemiştir. Bulanık mantık kullanılmış ve hazır bazı kelime listeleri Türkçeye çevrilmiştir. Bazı istisnalar dışında kelime kökleri ve cümleler arasında yüksek ilişki olduğu tespit edilmiştir. Çakmak ve arkadaşları [36] yine bir başka çalışmada deneyleri bulanık mantık tip 2 ye göre yeniden yapmışlar ve aynı sonuçları almışlardır.

Kaya ve arkadaşları [37] çalışmalarında farklı kaynaklardan topladıkları haber yorumlarını birden çok makine öğrenimi yöntemiyle duygu analizi yapmış ve ortalama %77 başarı sağlamışlardır. Politika haber yorumları bağlamında duygu

analizi yaparken yaşanan özel sorunları ele almışlar ve bazı çözümler üretmişlerdir. Akba ve ekibi [38] ise öznitelik seçme algoritmaları ve SVM yardımıyla film veri kümesi üzerinde tamamen gözetimsiz makine öğrenmesi teknikleri ile duygu sınıflaması yapmış ve yaklaşık %84 başarı elde etmişlerdir.

Balahur ve arkadaşları [39] çalışmalarında analiz yapılacak veriler makine çevirisi ile İngilizceye çevrilmiş ve İngilizce için hazır olan makine öğrenmesi yöntemleri ve öznitelik seçme yöntemleriyle analiz yapılmıştır. Verilerin bir kısmı ana dili Türkçe olan katılımcılar tarafından İngilizceye çevrilmiş ve aynı sonuçlar alınmıştır. İster makine çevirisi isterse insan çevirisi ile İngilizceye çevrilen verilerin barındırdıkları duyguyu koruduğu tespit edilmiştir. Türkçe duygu analizi sonucunda elde edilen başarı %60'a yakındır.

Akbaş [40] yaptığı çalışmada Twitter üzerinden topladığı verilerle makine öğrenimi yöntemleri kullanarak duygu analizi yapmıştır. Öznitelik seçiminde, oluşturduğu duygusal kelime listesini kullanarak hibrit bir yöntem kullanmıştır. Çalışma sonucunda pozitif ve negatif duygu sınıflamasında %85 civarı başarı elde etmiştir.

Çetin ve Amasyalı [41] yaptıkları çalışmada Türkçe Twitter verisi üzerinde birçok deney gerçekleştirmiş ve deney sonuçlarını karşılaştırmışlardır. Sonuç olarak eğiticili yöntemlerin daha başarılı olduğunu tespit etmişlerdir. Ortalama %60 civarı başarı elde etmişlerdir. Çetin ve Amasyalı [42] yine başka bir çalışmalarında ise makine öğrenimi yöntemlerinden NB ile sınıflama esnasında eğitim kümesinin sayısını %50 azaltıp aktif öğrenme algoritmaları uygulamıştır. Tüm eğitim kümesine göre daha başarılı olmuşlar ve %64 başarı elde etmişlerdir.

Özsert ve Özgür [43] yaptıkları çalışmada duygu analizinde çok önemli olan kelime polaritelerini belirlemek üzere bir takım deneyler yapmıştır. Kelime polaritelerini belirlemek için yarı otomatik bir yöntem geliştirmişlerdir. İngilizce için oluşturulmuş olan tohum kelimeleri Türkçeye çevirmiş adım adım öğrenme metodu ile kelimelerin polaritelerini tespit etmişlerdir. Tespit ederken elde ettikleri başarı yaklaşık %90 civarındadır.

Tüm bu bilgiler ışığında hala Türkçe için açık kaynak bir duygu sözlüğü ve deneysel bir veri kümesi yoktur.

3. DUYGU SÖZLÜĞÜ

Duygu Analizi yapmaya yarayan ve duygu bildiren ifadelerden oluşan sözlüklere Duygu Sözlüğü (*Sentiment Lexicon*) adı verilmektedir. Yapılan çalışmalar genellikle bağlam bağımsız ve geniş ölçekli olmakla birlikte hala gelişmektedir. Çizelge 2 incelendiğinde şimdiye kadar yapılan çalışmaların birbirlerini kapsama oranlarından anlaşılacağı üzere, bir duygu sözlüğü oluşturulurken genellikle önce oluşturulmuş başka sözlüklerden örnekler alınmaktadır.

Çizelge 2. Duygu Sözlükleri Arası Benzerlik

	MPQA [44]	Hu ve Liu [10]	Stone ve ekibi [45]	SentiWordNet [23]
MPQA	–	33/5402 (0.6%)	49/2867 (2%)	1127/4214 (27%)
Opinion Lexicon		–	32/2411 (1%)	1004/3994 (25%)
Inquirer			–	520/2306 (23%)
SentiWordNet				–

Stone ve ekibi [45] 1966 yılında bilgisayarlı içerik analizi için “Harvard General Inquirer” adlı bir sözlük oluşturmuşlardır. Sözlük içerisinde pozitif, negatif ve 180 civarı sınıf etiketi bulunan 11.800’e yakın terim bulunmaktadır. İlkel bir sözlük olsa da dönemine göre çok kapsamlıdır. Kendinden sonraki girişimlere taban oluşturduğu için oldukça önemlidir.

2004 yılında Hu ve Liu [10] tarafından düz metin şeklinde bir Duygu Sözlüğü oluşturulmuştur. Sözlük içerisinde 4.783 negatif ve 2.006 pozitif kelime bulunmaktadır. İçerisinde yanlış yazılmış kelimeleri, kelimelerin morfolojik çeşitlerini ve argo kelimeleri barındıran sözlük oldukça işlevseldir.

Wiebe ve ekibi [44] tarafından 2005 yılında oluşturulan 8.000 kelimelik Öznel Sözlük içerisinde kelimenin:

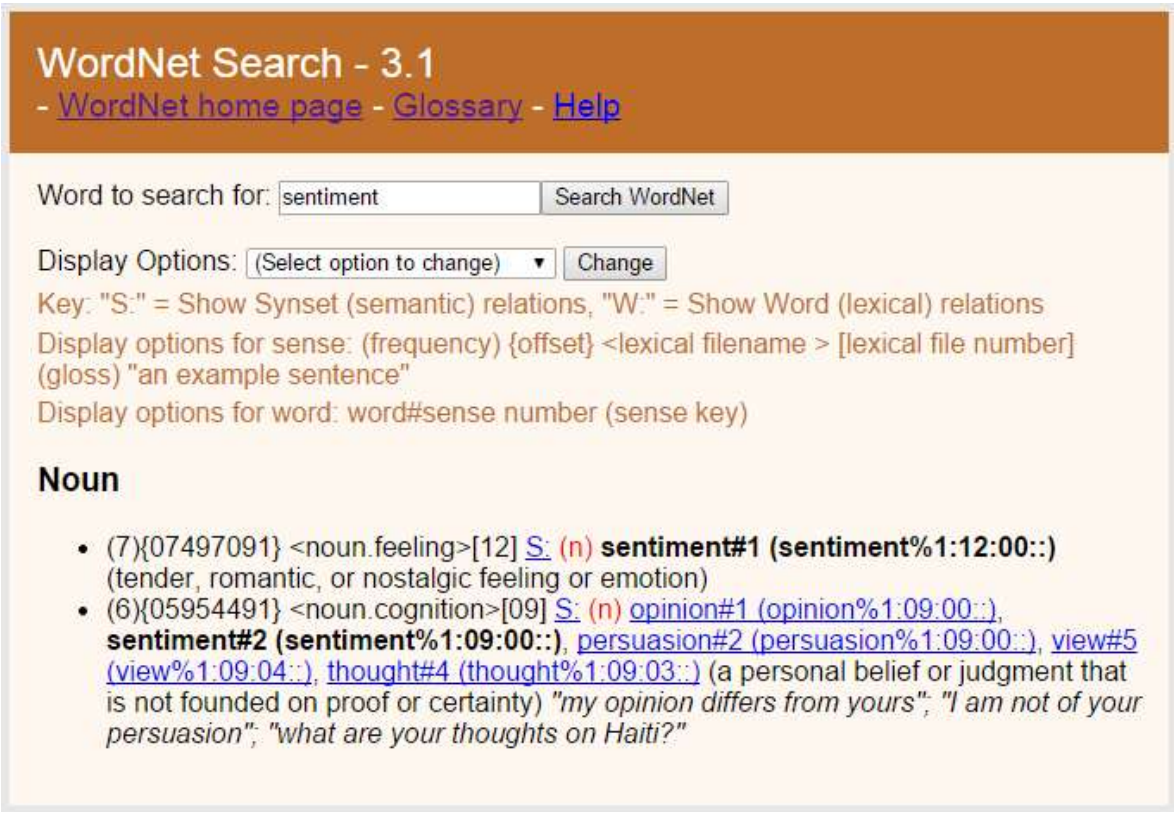
- Öznel kuvveti: zayıf-güçlü,
- Uzunluğu: kaç kelimeden oluştuğu,
- Kelimenin kendisi,
- POS (*Part Of Speech*) etiketi: isim, sıfat, fiil, zarf,
- Ek alıp almadığı,

- Öncül yön puanı (*prior polarity*) özelliklerini barındırmaktadır.

Bu çalışmaların sonrasında daha geniş kapsamlı bir sözlüğe ihtiyaç duyulmuş ve SentiWordNet [23] araştırmaları başlamıştır. SentiWordNet çalışmalarını anlayabilmek için öncelikle WordNet yapısını anlamak gerekmektedir.

3.1. WordNet

WordNet [14] , bilişsel dilbilim, psikoloji, doğal dil işleme ve İngiliz dili gibi bilim dallarının birleşerek yaklaşık 30 yılda tamamladığı yüzlerce gönüllünün görev aldığı dev bir sözlük veritabanıdır. İsimler, fiiller, sıfatlar ve zarflardan eş anlamlı olanlar anlam kümelerinde gruplanmıştır. Barındırdığı 117.000 anlamdaş kümesinin (*synonym set*, kısaca *synset*) tamamı kavramsal olarak ilişkili anlamdaş kümeleriyle bağlanmıştır. Şekil 2'deki WordNet arama ekran görüntüsünde görüleceği üzere sözlükteki her terim bir açıklamaya ve çoğu zaman örnek cümlelere sahiptir.



Şekil 2. WordNet arama altyapısından bir görüntü

WordNet'i çalışma açısından önemli kılan ise anlam kümeleri arası ilişkilerdir.

- Sözlükte en çok rastlanan ilişki altanlamlıktır (*hyponym*). Örneğin "gül ve çiçek" arasında "şahin ve kuş" arasındaki gibi altanlamlık ilişkisi vardır.
- Parça bütün ilişkisi (*meronymy*) ise yine karşımıza sıkça çıkmaktadır. Örneğin "parmak - el"in parçasıdır aralarında parça bütün ilişkisi vardır.
- Alt eylem ilişkisi (*troponymy*) ise "atıştırmak - yeme"nin alt eylemidir şeklinde açıklanabilir.
- Sözlükte, bakmak-görmek gibi sebep sonuç ilişkileri hareket-gezmek-koşmak gibi hiyerarşik ilişkiler de mevcuttur.
- Sözlükte sıfatlar doğrudan veya dolaylı karşıt anlamları ile ilişkilidir.
- Sözlükte "suç – suçlu" gibi türemiş sıfatlar arasındaki ilişkiler de ele alınmıştır.

Türkçe için kelimeleri bu şekilde ilişkileri ile birlikte inceleyen bir sözlük maalesef yoktur. Bununla birlikte açık erişime sahip olmayan, eş anlamlılar, zıt anlamlılar gibi farklı sözlükler bulunmaktadır. WordNet'i Türkçeleştirmeye çalışan Bilgin ve ekibi [46] proje kapsamında 11.628 anlamdaş kümesinin ve 16.095 terimin bir kısmını el yordamı ile ve çoğunluğunu İngilizce Türkçe sözlük vasıtası ile çevirmiştir. Ekibin %66 oranında ilişkileri yansıttıklarını iddia ettikleri çalışmaya erişim sağlanamamıştır. Yine aynı dönem WordNet'i otomatik yöntemler ile Türkçeye çevirmeye çalışan Amasyalı [47] ise diğer çalışmayla aynı sonuçları üretmiştir. Bu çalışmanın kaynaklarına da erişim sağlanamamıştır.

3.2. SentiWordNet

SentiWordNet Duygu Analizi için hazırlanmış sözlüksel bir kaynaktır. Esuli ve Sebastiani daha önce yaptıkları üçlü sınıflama (olumlu, olumsuz ve nesnel) altyapılarını [21] [22] anlam kümelerini sınıflamak için uyarlamışlardır. Üçlü sınıflayıcının ürettiği değerler her bir anlamdaş kümesinin olumlu, olumsuz ve nesnel değerlerine işaret etmektedir. Daha önce yayınlanmış kısmi duygu sözlükleri ve elleriyle işaretledikleri listeleri tohum sözcükler olarak kabul etmişler, öğrenme algoritması ile bu sözlüğü genişleterek işlemi yinelemeli olarak tüm anlam kümeleri değerlendirilene kadar tekrar etmişlerdir. Bu işlem sırasında sınıflayamadıkları anlam kümelerini WordNet içerisindeki eş anlamlılık ve karşıt anlamlılık gibi ilişkilerinden faydalanarak sınıflamayı başarmışlardır.

SentiWordNet içerisinde 117.659 adet anlamdaş kümesi barındırmaktadır. Bu anlam kümelerinin %69,79'u isim, %15,43'ü sıfat, %3,07'si zarf ve %11,7'si fiildir. SentiWordNet içerisinde Şekil 3'te görüleceği üzere her anlamdaş kümesinin pozitif, negatif ve objektif olmak üzere 3 puanı vardır.

```
n   07481951    0.125   0.25    sentiment#1 tender, romantic, or nostalgic feeling or emotion
n   07482128    0.25    0.625   sentimentality#2    extravagant or affected feeling or emotion
n   07482267    0.125   0.5 mawkishness#1 bathos#2  insincere pathos
n   07482368    0.125   0.375   razbliuto#1 the sentimental feeling you have about someone you on
n   07482521    0   0.125   complex#3   (psychoanalysis) a combination of emotions and impulses t
n   07482782    0   0   oedipus_complex#1 oedipal_complex#1 a complex of males; desire to possess
n   07483005    0   0   electra_complex#1   a complex of females; sexual attraction to the father
n   07483120    0.375   0   inferiority_complex#1   a sense of personal inferiority arising from
n   07483305    0.25    0.375   ambivalency#1 ambivalence#1 mixed feelings or emotions
n   07483439    0.125   0.125   conflict#2  opposition between two simultaneous but incompatible
n   07483622    0   0.625   apathy#1    an absence of emotion or enthusiasm
n   07483782    0.25    0.5 unemotionality#1 stolidity#1 phlegm#1 indifference#2 impassivity#1 im
n   07484109    0.5 0   listlessness#1 lassitude#2 languor#2    a feeling of lack of interest or
n   07484265    0   0.125   desire#1    the feeling that accompanies an unsatisfied state
n   07484547    0.5 0.125   dream#3 aspiration#2 ambition#1 a cherished desire; "his ambition is
n   07484792    0   0   american_dream#1    the widespread aspiration of Americans to live better
n   07484929    0.25    0   emulation#1 ambition to equal or excel
```

Şekil 3. SentiWordNet içerisinden bir görüntü.

Bu üç puanın toplamı en fazla 1 olabilir. Bu durumda anlamdaş kümesi bir miktar olumlu, bir miktar olumsuz ve bir miktar nötr olabilir. En olumlu anlamdaş kümesinin pozitif puanı 1, negatif puanı 0 ve objektif puanı 0 olurken; en olumsuz anlamdaş kümesinin pozitif puanı 0, negatif puanı 1 ve objektif puanı 0 olmaktadır. Şekil 3'te görüldüğü üzere sözlük içerisinde her bir anlamdaş kümesi için pozitif ve negatif olmak üzere iki puan yer almaktadır ve *objektif puan= 1-(pozitif puan + negatif puan)* formülü ile bulunabilir.

Anlam kümelerinin 18.028 adet kümenin negatif puanı, 17.015 adet kümenin pozitif puanı sıfırdan büyüktür. Yine anlam kümelerinin pozitif puanından negatif puanını çıkardıktan sonra puanı sıfırdan büyük olan 13.128 adet anlamdaş kümesi kalmaktadır.

Anlamdaş kümeleri aynı anlama gelen kelimelerin temsil ettiği anlam gibi düşünülebilir. Örneğin göz anlamına gelen "eye" için bulunan ilk anlama karşılık gelen anlam kümesi şu şekildedir:

-optic, oculus, eye.

SentiWordNet içerisindeki anlam kümelerini terimlere böldükten sonra terim sayısı 206.941, ayrık terim sayısı 147.306 olmuştur. Bu ayrık terimlerden 24.650 tanesinin pozitif puanı, 26.113 tanesinin negatif puanı birleştirerek düşünürsek 19.320 tanesinin ise pozitif puanından negatif puanını çıkardıktan sonra puanı sıfırdan büyüktür.

SentiWordNet duygu sözlüğü Şekil 4'te görüleceği gibi İngilizce için yapılmış ve WordNet içerisinde ilişkileri ve karşılıkları tanımlanmış İtalyanca, Fransızca, Almanca gibi Latin Avrupa dillerine otomatik olarak çevrilmiştir. Türk dili için WordNet yapısı bulunmaması ve terim karşılıklarının sözlüklere göre farklılık göstermesi nedeniyle otomatik çeviri yapılması Avrupa dillerine göre daha zordur.

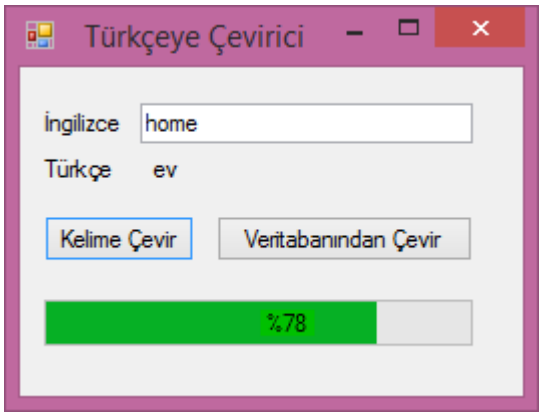
Çalışma kapsamında SentiWordNet Duygu Sözlüğü birkaç farklı yöntemle otomatik olarak Türkçeye çevrilmiştir. Bu çevirim esnasında karşılaşılan zorluklar 4. bölümde anlatılmıştır.

Şekil 4. SentiWordNet arama altyapısından bir görünüm.

4. OTOMATİK DUYGU SÖZLÜĞÜ ÇEVİRİMİ VE DUYGU ANALİZİNDE KULLANIMI

Akba ve arkadaşları [38] çalışmalarında makine öğrenmesi yöntemleri ve öznitelik seçme algoritmalarıyla yaptıkları uygulamanın sadece ilgili bağlamda geçerli olduğu tespit edilmiştir. Bu bilgiler ışığında daha genel bir yöntem geliştirme çabası içerisine girilmiştir. İnsan faktörünü kullanmadan tam otomatik olarak duygu analizi yapma ihtiyacını gidermek için literatüre göre bir duygu sözlüğü gereksinimi bulunmaktadır. Geniş kapsamlı duygu sözlüklerinden SentiWordNet Duygu sözlüğünü hızlı ve insan eli değmeden oluşturmak hedeflenmiştir.

Herhangi bir Türkçe kelimeyi veya metni internet üzerinde açık ve ücretsiz sözlüklerden [48] [49] [50] veya Google Translate API [51] hizmetinden faydalanarak doğrudan çeviri yapılabilir. Kelime öbeklerini ve cümleleri tercüme edebilmesinden ötürü Google Translate API öncelikli olmak üzere diğer üç sözlüğü kullanarak çeviri yapabileceğimiz bir ortam oluşturulmuştur. (Bkz. Şekil 5)



Şekil 5. Türkçeye Çevirici yazılımın ekran görüntüsü

4.1. Duygu Sözlüğü Hesaplama ve Tekilleştirme İşlemleri

SentiWordNet'in yapısı ["ID", "POS", "SynsetTerms", "PosScore", "NegScore", "Gloss"] şeklindedir. Burada *ID*: benzersiz bir anahtar, *SynsetTerms*: Anlamdaş kümesi elemanları kaçıncı anlamda oldukları bilgisiyle beraber yer almaktadır, *POS*: (*Part-Of-Speech*) kelimenin cümle öğelerinden hangisi olduğu, *PosScore*: Pozitif puanı, *NegScore*: Negatif puanı, *Gloss* ise örnek cümleleri temsil eder.

ID	POS	PosScore	NegScore	SynsetTerms	Gloss
6775812	n	0,125	0	sentimentalism#1	the excessive expression of te...
4628080	n	0,5	0,125	sentimentalism#2	a predilection for sentimentality
10579676	n	0,375	0,25	sentimentalist#1 romanticist#1	someone who indulges in exce...
7482128	n	0,25	0,625	sentimentality#2	extravagant or affected feeling ...
1219306	n	0	0	sentimentalization#1 sentimentalisation#1...	the act of indulging in sentiment
533185	v	0	0,125	sentimentalize#1 sentimentalise#2	look at with sentimentality or tu...
579105	v	0,125	0	sentimentalize#2 sentimentalise#1	make (someone or something) ...
449609	r	0,25	0	sentimentally#1	in a sentimental manner; "I mi...
12944	v	0	0	sentimentize#1 sentimentise#1 sentiment...	act in a sentimental way or ind...

Şekil 6. SentiWordNet içeriğinden bir örnek

SentiWordNet içerisinde yer alan 117.659 adet anlamdaş kümesi Şekil 6'daki gibi küme halinde çeviri yapılamayacağından terimlere bölünmüştür. SentiWordNet terimlere bölündükten sonra terim sayısı 206.941 olmuştur. Sözlük terimlere bölünürken, tüm terimler ait oldukları anlamdaş kümelerinin sahip olduğu negatif ve pozitif puan kaydedilmiştir. Terimin yansıttığı duygu puanı Eşitlik 1'de görüldüğü üzere sahip olduğu pozitif puanından negatif puanını çıkararak hesaplanmaktadır. [52] Bu şekilde tüm terimlerin duygu puanı hesaplanmıştır.

$$Skor_{Terim} = Skor_{Pozitif_{Terim}} - Skor_{Negatif_{Terim}}$$

Eşitlik 1. Terim Duygu Puanı Hesaplama

Sözlüğün yapısı Şekil 7'de görüldüğü üzere ["TerimNo", "Terim", "POS", "Skor"] şekline dönüştürülmüştür. Buradaki TerimNo: terimin kaçıncı anlamı olduğunu, Terim: terimi, POS: kelimenin cümle öğelerinden hangisi olduğunu, Skor ise Eşitlik 1'e göre hesaplanmış puanı ifade eder.

Elde edilen terimlerin Şekil 7'de görüldüğü üzere birden fazla anlamı sıralarına göre kaydedilmiştir. Mevcut çeviri altyapısı ile bir kelimenin kaçıncı anlamının Türkçe hangi kelimeye karşılık geldiği bulunamadığından terimlerin tüm anlamları POS etiketlerine göre birleştirilmiştir.

TermNo	Term	POS	Score
1	sentiment	n	-0,125
2	sentiment	n	-0,3125
1	sentimental	a	0,5
2	sentimental	a	0,0625
1	sentimentalisation	n	0
1	sentimentalise	v	0,125
2	sentimentalise	v	-0,0625
3	sentimentalise	v	0
1	sentimentalism	n	0,125
2	sentimentalism	n	0,1875
1	sentimentalist	n	0,125
1	sentimentality	n	-0,125
2	sentimentality	n	-0,1875
1	sentimentalization	n	0
1	sentimentalize	v	-0,125
2	sentimentalize	v	0,0625

Şekil 7. SentiWordNet tüm terimler

Bu birleşik terimin puanı hesaplanırken birleşime katılan her anlamın puanı anlam sırasına göre ağırlıklandırılır ve birleştirilen terimin puanına eklenir. Bu sayede birleşik terimin puanı, ilk anlamın ağırlığı en çok, son anlamın ağırlığı en az olacak şekilde Eşitlik 2'ye göre hesaplanmıştır.

$$S_t = \frac{\sum_{i=1}^{n} \frac{1}{i+1} s_i}{\sum_{i=1}^{n} \frac{1}{i}}$$

Eşitlik 2. Ağırlıklarına göre terimlerin toplam puanları

Bu eşitlik SentiWordNet kaynak sitesindeki [53] örnek uygulamadan alınmıştır. İlgili sitedeki, İngilizce için yazılmış örnek uygulamanın çıktılarıyla oluşturulan yeni listeden örnekler karşılaştırılmış ve hesaplamanın doğruluğu teyit edilmiştir. Şekil 8'te görüleceği üzere terimlerin tüm anlamları POS etiketlerine göre birleştirilmiş ve puanları hesaplanarak kaydedilmiştir.

Term	POS	CalScore
sentiment	n	-0.292
sentimental	a	0.375
sentimentalisation	n	0.000
sentimentalise	v	0.034
sentimentalism	n	0.208
sentimentalist	n	0.125
sentimentality	n	-0.208
sentimentalization	n	0.000
sentimentalize	v	-0.034
sentimentally	r	0.250
sentimentise	v	0.000
sentimentize	v	0.000

Şekil 8. Örnek ayrık terim-POS ikilileri ve hesaplanmış puanları

4.2. Duygu Sözlüğünü Türkçeye Çevirme İşlemleri

POS etiketi ve terim bazında ayrık terimlerden oluşan ve her terimin tüm anlamlarına ait puanları Eşitlik 2'ye göre birleştirilerek elde edilmiş yeni bir liste ortaya konulmuştur (Bkz. Şekil 8). Listede hesaplanmış puanı sıfırdan farklı olan 39.965 terim bulunmakta iken 1.743 terim ise tekrar etmektedir. Terimlerin tekrar etmesinin nedeni bazı kelimelerin yazılışlarının aynı, POS etiketlerinin farklı olmasındandır.

Her terim teker teker sözlükler vasıtası ile çevrilmeye çalışılmıştır. Çeviri yapılırken POS etiketleri göz ardı edilmiş ve sadece kelime bazında çeviri işlemi gerçekleştirilmiştir. Çeviri çalışmaları sonucunda puanı 0'dan farklı 39.965 terimin %97,5'i başarı ile çevrilmiştir.

Tercih edilen çeviri altyapısında öncelikli olarak Google Translate API [51] servisi kullanılmıştır. Kelime öbeklerini ve deyimleri çoğu zaman doğru çevirdiği gözlemlenmiş ve bu hizmete öncelik verilmiştir. Google servisi 80 dilde (2014 yılı itibarı ile) çeviri yapabilmektedir. Servis makine çevirisi ve istatistiksel çeviri algoritmalarıyla çalıştığı için çeviri kalitesini belirleyen asıl unsur o dilde insan eliyle çevrilmiş kaynakların çokluğudur.

Diğer kullanılan sözlükler ise sırasıyla

- sözlük 1 (tureng.com) [48],

- sözlük 2 (zargan.com) [49]
- sözlük 3 (tr.bab.la) [50] çevrimiçi sözlükleridir.

Çevrimiçi sözlüklerin hizmet verdiği web sayfaları, uygulama ortamında yapısal olarak çözümlenip, veri kümesi oluşturma kısmında ayrıntılı şekilde anlatıldığı üzere, ayrıştırılarak (*parsing*) otomatik çeviri altyapısı oluşturulmuştur.

Çizelge 3. Kaynaklara göre çeviri oranları

Kaynak	Çeviri Miktarı	Çeviri Oranları
Google	29.786	
Sözlük 1	29.944	
Sözlük 2	18.157	
Sözlük 3	13.938	
Toplam sözlüklerden	36.077	%94,39
Eşanlamdan bulunan	1.186	% 3,10
Çevrilemeyen	959	% 2,51
Ayrık Terim Sayısı	38.222	
POS etiketi farklı terim aynı	1.743	
Toplam	39.965	

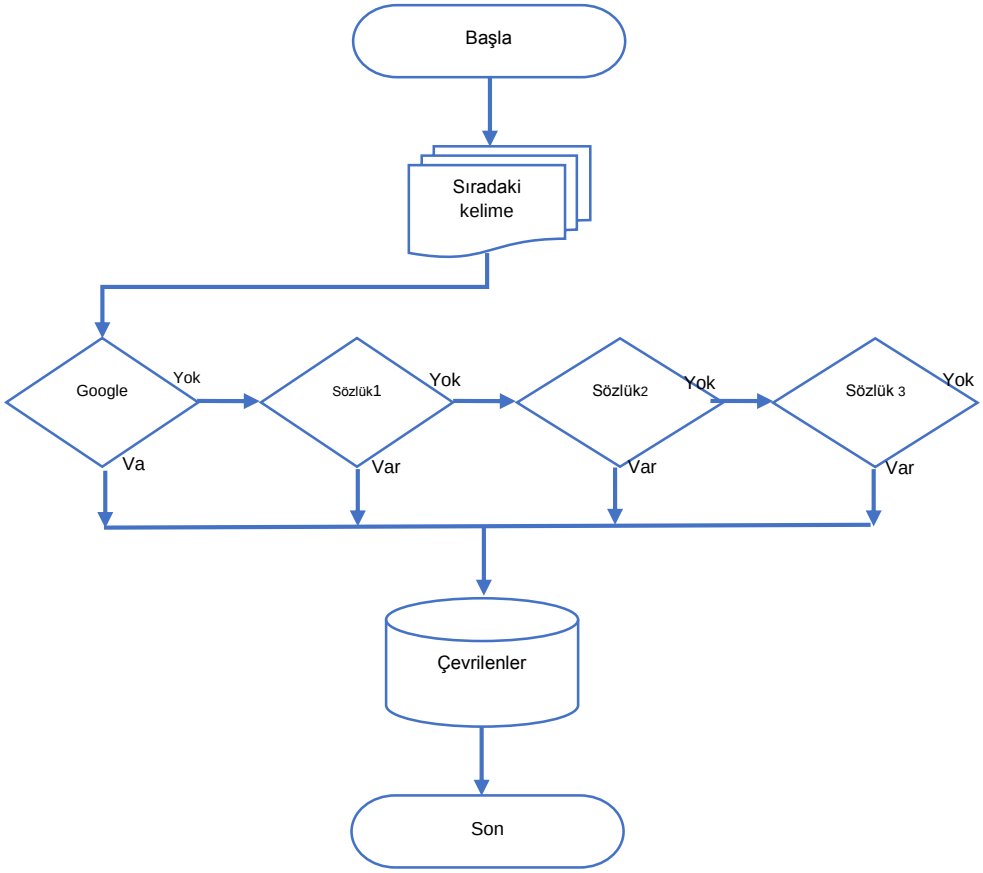
Çeviri altyapısı

- seri bağlanmış sözlükler,
- paralel bağlanmış sözlükler
- anlamdaş anlamını kullanmak şeklinde üç farklı yaklaşımda tasarlanmıştır.

Tüm çeviri istatistikleri Çizelge 3. Kaynaklara göre çeviri oranları

4.2.1. Seri Bağlanmış Sözlükler ile Çevirme İşlemleri

Bu yaklaşımda sözlükler birbirine seri olarak bağlanmıştır (Bkz. Şekil 9). Terim önce Google sözlüğünde aranmıştır. Bulunan çeviri kaydedilmiş, bulunamayan terimler ise sırasıyla sözlük 1 [48], sözlük 2 [49] ve sözlük 3 [50] kaynaklarından aranmış ve bulunan çeviri kaydedilmiştir. Bu yaklaşımda sözcüğe hızlıca, en az bir anlam bulabilmek amaçlanmıştır. Oldukça hızlı ve kapsamlı çeviri elde edilmiştir. Çevrilmesi planlanan 38.222 terimin %94,39'una karşılık bulunmuş ve 36.077 adet terim Türkçeye çevrilmiştir. Bununla beraber 2.145 terimin karşılığı hiçbir sözlükte bulunamamıştır.

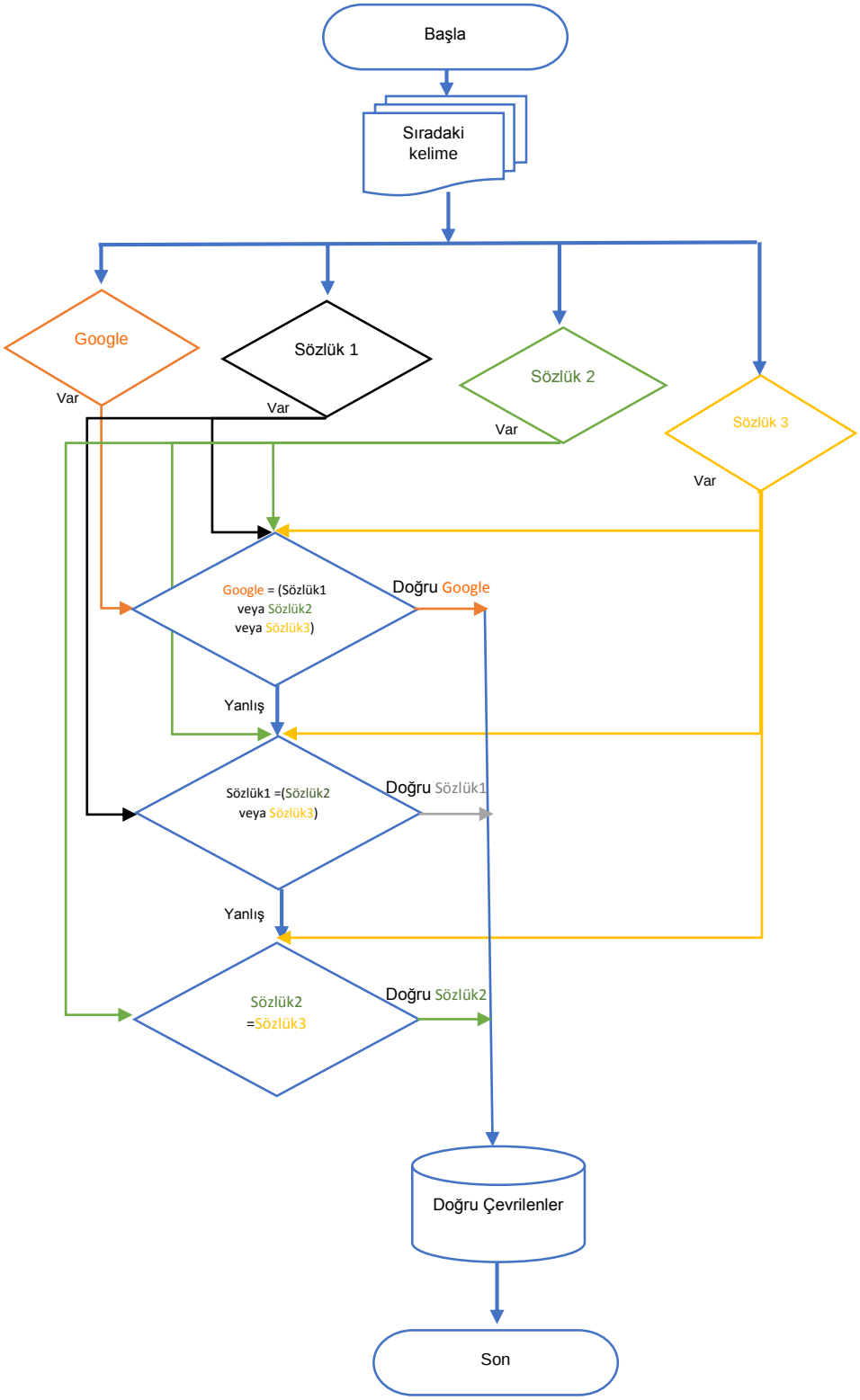


Şekil 9. Seri Çeviri algoritması

Karşılığı bulunan 36.077 adet terimin %82,5'i olan 29.786 adet terim Google sözlüğünden, %10,8'i olan 3.918 terim sözlük 1'den, %5,3'ü olan 1.922 terim sözlük 2'den ve %1'i olan 451 terim ise sözlük 3'ten faydalanarak Türkçeye çeviri yapılmıştır. Seri çeviri algoritması yöntemiyle çeviri yapılan ve Türkçe karşılıkları aynı olan 8.826 adet terim tekrar etmektedir. Yüksek oranda çeviri yapılmasına rağmen çevrilen terimlerin Türkçe karşılıklarından emin olunamamıştır. Terimlerin otomatik olarak doğru çevrildiğinden emin olabilmek adına, aynı kelimelerin bir başka sözlük vasıtasıyla çevirisi yapılıp karşılıklarının kontrol edilmesi yöntemi tercih edilmiş ve paralel çeviri algoritması geliştirilmiştir.

4.2.2. Paralel Bağlanmış Sözlükler ile Çevirme İşlemleri

İkinci yaklaşımda ise sözlükler bir birine paralel olarak bağlanmıştır. Terim Google, sözlük 1 [48], sözlük 2 [49] ve sözlük 3 [50] olmak üzere tüm sözlüklerde aynı anda aranmıştır. Şekil 10'da görüldüğü üzere herhangi iki sözlük çevirisi birbiri ile aynı ise, doğru olarak çeviri yapıldığı kabul edilmiş ve terim kaydedilmiştir.



Şekil 10. Paralel Çeviri Algoritması

Paralel çeviri algoritması (Bkz. Şekil 10) ile çevirisi yapılan terim sayısı azalmakla birlikte terimlerin doğru çevrilme olasılıkları artmıştır. Çevrilmesi planlanan 38.222 terimden sadece 11.369 tanesinin karşılığı bulunabilmiştir. Bu kapsamda çevrilen

sözcüklerin %92'si olan 8.691 terim Google kaynağından, %18'i olan 2.063 terim Sözlük 1 kaynağından ve %5'i olan 615 terim ise Sözlük 2 kaynağından doğru bir şekilde çeviri yapılmıştır. Paralel çeviri algoritması yöntemiyle çevrilen ve Türkçe karşılıkları aynı olan 1.957 adet terim tekrar etmektedir.

4.2.3. Anlamdaş İlişkisi Yoluyla Çevirme İşlemleri

Üçüncü yaklaşımda ise, hem seri çeviri algoritması hem de paralel çeviri algoritması yöntemleri ile Türkçe karşılığı bulunamayan terimlere, SentiWordNet içerisindeki anlamdaş ilişkisinden faydalanarak bir karşılık bulunmaya çalışılmıştır.

Öncelikli olarak anlamdaş ilişkisi olan terim tespit edilmiştir. Bulunan anlamdaş terimin Türkçe karşılığı alınmıştır. Anlamdaşın varsa paralel çeviri algoritması ile bulunmuş olan karşılığı, yoksa seri çeviri algoritması ile bulunmuş olan karşılığı kaydedilmiştir. Çevrilmesi planlanan 38.222 terimin %3,10'u yani 2.145 adet terimin ne seri ne de paralel çeviri algoritmaları vasıtasıyla Türkçe karşılığı bulunamamıştır. Bu terimlerin %55'i olan 1.186 adet terim anlamdaş ilişkisi yoluyla başarıyla Türkçeye çevrilmiştir.

4.3. Türkçe Duygu Sözlüğü Oluşturulması

Çevrilmesi planlanan 38.222 adet terimin %97,5'i tüm çeviri algoritmaları kullanılarak Türkçeye çevrilmiş ve takip eden tekilleştirme işlemleri sonucu oluşturulan sözlüğe Türkçe Duygu Sözlüğü ismi verilmiştir.

Çevrim algoritmalarının başarılarını karşılaştırmak amaçlı, Seri Türkçe Duygu Sözlüğü (TDSs), Paralel Türkçe Duygu Sözlüğü (TDSp) ve her ikisinin birleşiminden oluşan Türkçe Duygu Sözlüğü sürüm 1 (TDSv1) olmak üzere üç farklı sözlük oluşturulmuştur.

Seri Türkçe Duygu Sözlüğü, sözlükler vasıtasıyla en az bir Türkçe karşılığı bulunmuş olan 37.817 terimin yanı sıra, sözlükte bulunamayan ve anlamdaş ilişkisi yoluyla bulunan 1.186 terimin alınmasıyla oluşturulmuştur. Paralel Türkçe Duygu Sözlüğü ise sözlüklerin en az ikisinde geçtiği için paralel anlamı bulunan 12.357 terim ve anlamdaş ilişkisiyle bulunan 1.186 terim birleştirilerek oluşturulmuştur. Türkçe Duygu Sözlüğü sürüm 1 ise paralel sözlük terimlerine öncelik verilmek üzere seri sözlük terimlerinin birleştirilmesi ile oluşturulmuştur. Oluşturulan sözlüklerde

tekrar eden terimler bir sonraki bölümde anlatıldığı üzere birleştirilmiş ve ayrık terimlerden oluşan sözlükler oluşturulmuştur.

4.4. Türkçe Duygu Sözlüğü Hesaplama Yöntemleri

Sözlükler içerisinde, İngilizcede birden fazla anlama karşılık gelen Türkçe kelimeler bulunmaktadır. Bazı terimlerin karşılığı kaynaklarda bulunamadığı için anlamdaş ilişkisi yoluyla çevirisi yapılmıştır. Bu sebeplerden dolayı Çizelge 4'te tekrar eden terimler gösterilmiştir. Sözlüklerde bunların dışında POS etiketleri farklı ama yazılışları aynı 1.743 kelime bulunmakta ve bu terimlerinde Türkçe karşılıkları tekrar eden terimlerin sayısını arttırmaktadır.

Çizelge 4. Çeviri sonrası tekrar eden sözcükler

	SentiWordNet	TDSs	TDSp	TDSv1
Terim Sayısı	39.965	39.004	13.544	39.004
Çeviri sonrası tekrar eden terim sayısı	-	8.596	2.893	8.802
POS etiketi farklı ama yazılışları aynı	1.743	3.157	1.239	3.061
Ayrık terim sayısı	38.222	27.251	9.412	27.141

Sözlüğümüzü ["Terim", "Puan"] şeklinde kullanabilmek için tekrar eden sesteş terimlerden ayıklamak gerekmiştir. Bu ayıklama işlemi esnasında bilgi kaybını en aza indirgemek için, birleştirilen her terimin puanını birleşime aktarmak hedeflenmiştir. Bu aktarım üç yaklaşımda ele alınmıştır. İlk yaklaşımda tüm sesteşlerin aynı oranda etkisi olması için puanlarının aritmetik ortalamaları alınmıştır.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Eşitlik 3. Aritmetik Ortalama

İkinci yaklaşımda ise sesteşlerin sahip oldukları puanların en etkili olanı yani mutlak değeri en büyük olanı alınmıştır.

$$x_{maks} = \max |x_i|$$

Eşitlik 4. Mutlak maksimum eleman

Üçüncü ve son yaklaşımda ise değerine bakılmaksızın terimin temsil ettiği yön alınmıştır. Birleştirilen terim puanlarının aritmetik ortalama değeri sıfırdan büyükse Yön Puanı 1, değer sıfırdan küçükse Yön Puanı -1 olacak şekilde hesaplanmıştır.

$$x_{yön} = \begin{cases} -1, & \bar{x} < 0 \\ 1, & \bar{x} > 0 \end{cases}$$

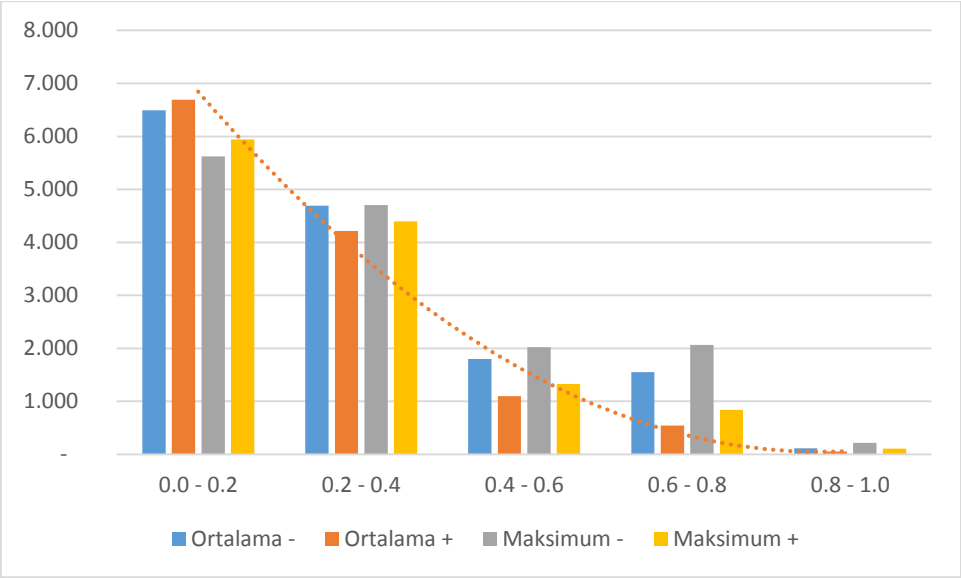
Eşitlik 5. Aritmetik ortalamanın değerine göre temsil edilen yön değeri

Bu yaklaşımlar neticesinde yapılan hesaplamalar ile elde edilen değerlerle [“Terim”,”OrtSkor”,”MaksSkor”,”YönSkor”] düzeninde üç farklı Türkçe Duygu Sözlüğü oluşturulmuştur.

Çizelge 5. Seri Türkçe Duygu Sözlüğü Terim Dağılımı

Aralık	Ortalama -	Ortalama +	Maksimum -	Maksimum +	Yön -	Yön +
0.0 - 0.2	6.494	6.693	5.624	5.946	-	-
0.2 - 0.4	4.695	4.213	4.707	4.394	-	-
0.4 - 0.6	1.799	1.096	2.024	1.329	-	-
0.6 - 0.8	1.549	542	2.067	837	-	-
0.8 - 1.0	115	55	215	108	14.652	12.599
Toplam	14.652	12.599	14.637	12.614	14.652	12.599

Elde edilen terimlerin sözlük içerisindeki dağılımları, Seri Türkçe duygu Sözlüğü (TDSs) için Çizelge 5’te görüldüğü üzere olmuştur. Sözlüğün dağılımına bakıldığında puanı yüksek-etkisi güçlü olan negatif terimlerin sayısı pozitif terimlerin sayısından fazla olduğu görülmektedir.



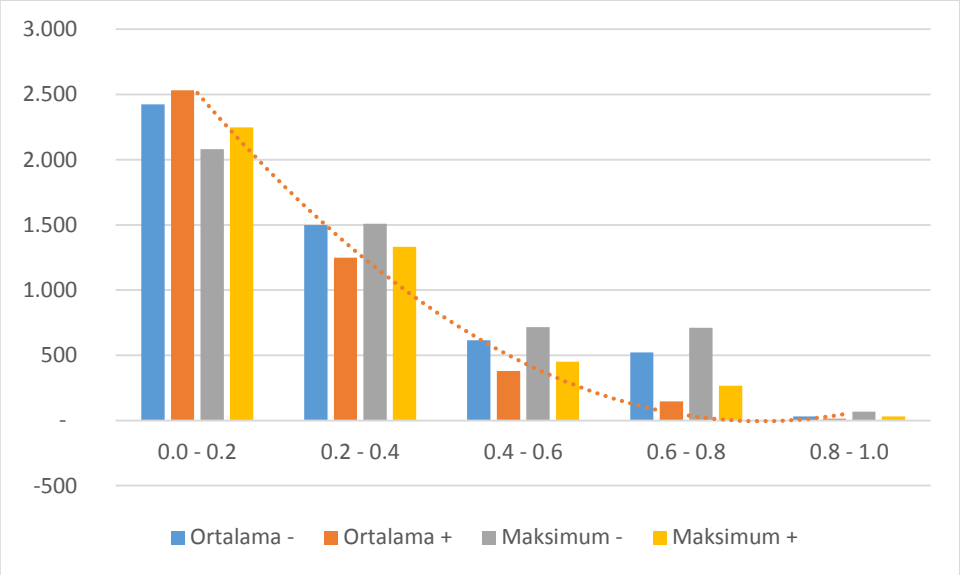
Şekil 11. Seri Türkçe Duygu Sözlüğü Terim Dağılımı

Seri Türkçe Duygu Sözlüğü (TDSs) ortalama puanlar için 0,4 – 0,6 aralığındaki pozitif terimlerin sayısı 1.096 iken negatif terimlerin sayısı önemli ölçüde artarak 1.799 olmuştur. Yine 0,6 – 0,8 aralığındaki pozitif değerler 542 iken negatif terimlerin sayısı neredeyse üç katına çıkarak 1.549 olmuştur. 0,8-1,0 aralığında ise pozitif terim sayısı 55 iken negatif terim sayısı yaklaşık iki katı 115 adettir. Aynı yönelim maksimum puanlar içinde geçerli olmaktadır.

Çizelge 6. Paralel Türkçe Duygu Sözlüğü Terim Dağılımı

Aralık	Ortalama -	Ortalama +	Maksimum -	Maksimum +	Yön -	Yön +
0.0 - 0.2	2.424	2.533	2.080	2.248	-	-
0.2 - 0.4	1.498	1.249	1.510	1.332	-	-
0.4 - 0.6	616	379	715	451	-	-
0.6 - 0.8	522	147	710	267	-	-
0.8 - 1.0	31	13	67	32	5.091	4.321
Toplam	5.091	4.321	5.082	4.330	5.091	4.321

Elde edilen terimlerin sözlük içerisindeki dağılımları, Paralel Türkçe Duygu Sözlüğü (TDSp) için Çizelge 6'da görüldüğü üzere olmuştur. Aynı TDSs'de olduğu gibi TDSp'de de pozitif ve negatif terimler dengesiz bir şekilde dağılmıştır.

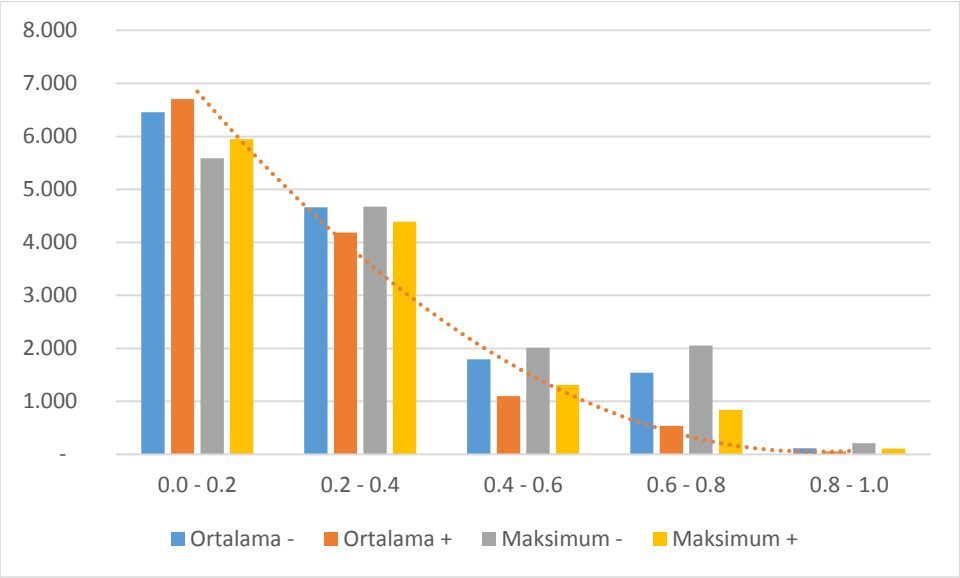


Şekil 12. Paralel Türkçe Duygu Sözlüğü Terim Dağılımı

Çizelge 7. Türkçe Duygu Sözlüğü Sürüm 1 Terim Dağılımı

Aralık	Ortalama -	Ortalama +	Maksimum -	Maksimum +	Yön -	Yön +
0.0 - 0.2	6.457	6.702	5.587	5.950	-	-
0.2 - 0.4	4.663	4.185	4.676	4.392	-	-
0.4 - 0.6	1.794	1.101	2.010	1.310	-	-
0.6 - 0.8	1.538	537	2.056	840	-	-
0.8 - 1.0	113	51	212	108	14.565	12.576
Toplam	14.565	12.576	14.541	12.600	14.565	12.576

Elde edilen terimlerin sözlük içerisindeki dağılımları, Seri ve Paralel sözlüklerin birleşiminden oluşan Türkçe Duygu Sözlüğü sürüm 1 (TDSv1) için Çizelge 7'de görüldüğü üzere olmuştur.



Şekil 13. Türkçe Duygu Sözlüğü Sürüm 1 Terim Dağılımı

Elde edilen her üç sözlük içerisinde gözlemlenen terim dağılımlarındaki dengesizlik, yapılacak deneylerin temsil ağırlığı çok olan yöne yani negatife doğru baskın çıkmasını sağlayacağından deney başarısını olumsuz etkileyebilir niteliktedir.

4.5. Cümle Polaritesi Hesaplama Yöntemleri

Cümle içerisinde öznel kelime geçiyorsa ve bu öznel kelime oluşturulan sözlük içerisinde yer alıyorsa o cümlenin polarite puanı hesaplanabilir. Hesaplanan puan Eşitlik 6'da görüleceği gibi değerlendirilip cümlenin polaritesine karar verilir.

$$Polarite = \begin{cases} Negatif, & \sum_{n}^{i=1} T_i < 0 \\ Pozitif, & \sum_{n}^{i=1} T_i > 0 \end{cases}$$

Eşitlik 6. Polarite Karar işlevi

Sözlüklerin doğal yapısı gereği negatif terimlerin puanı sıfırdan küçük, pozitif terimlerin puanı sıfırdan büyüktür. Sözlükten yararlanılarak veri kümesindeki cümlelerin her terimine karşılık gelen puanları toplanıp bir polarite skoru elde edilmiştir. Elde edilen polarite skoru, doğal eşik değer olan sıfırdan büyükse pozitif, sıfırdan küçükse negatif olarak değerlendirilmiştir.

5. DENEYLER VE BULGULAR

Kullanıma açık hazır Türkçe veri kümeleri olmadığından ilgili kaynaklarla uyumlu olarak film ve otel yorumlarından yeni iki veri kümesi oluşturulmuştur. Oluşturulan veri kümelerinin dengeli dağılımda olması sağlanmıştır. Elde edilen sözlük vasıtasıyla iki farklı veri kümesinde duygu analizi yapılmıştır. Literatürle paralel sonuçlar alınmıştır. Veri kümlerinin elde edilme aşaması, deney aşaması ve sonuçların karşılaştırılması ayrıntılı bir şekilde irdelenmiştir.

5.1. Deney Hazırlıkları

Tez kapsamında bir dizi deneyler tasarlanmıştır. Yapılan deneyler yoluyla oluşturulan Türkçe Duygu Sözlüğü hesaplama yöntemlerinin başarısı, çeviri yöntemlerinin başarısı, gövdelemenin başarısı gibi birçok başarım gözlemlenmiştir.

Öncelikle doğal web kaynaklarından veri kümeleri oluşturulmuştur. Oluşturulan veri kümeleri gereksiz karakterlerden temizlenip, gövdelenmiştir. Veri kümeleri oluşturma ve kelimeleri gövdeleme esnasında kullanılan kütüphane ve ortamlar anlatılmıştır.

5.2. Veri Kümeleri

Birçok çalışmada veri kümeleri film ve otel yorumları toplanarak oluşturulmuştur [38] [27] [31] [11]. İster sözlük tabanlı isterse istatistiksel yöntemlerle olsun tüm çalışmalarda deneylerin doğruluğunu tespit edebilmek adına etiketli veri kümesine ihtiyaç duyulmaktadır.

Yorumu yapan kişinin kendi yorumunu etiketlediği doğal etiketli veri kümeleri, yorumun başka kişi veya uygulamalar tarafından etiketlendiği veri kümelerinden daha tutarlıdır. Genelde otel, film veya ürün yorumları toplayan web siteleri, kişilerden yorumun yanı sıra puan da istemektedir. Bu puanlama sistemi bazen yüzlük bazen beşlik sistemde, bazense yıldız ya da sembol seçtirerek yapılmaktadır.

Yorumun etiketi bulunamazsa devreye uygulamalar ve insan gücü girmektedir. Örneğin sosyal medyada yapılan yorumlar veya haber yorumları etiketsizdir. Bu etiketsiz yorumlar bazı çalışmalarda elle, bazılarında daha önce yapılmış olan duygu analizi uygulamaları ile bazen de her iki yöntem bir arada kullanılarak etiketlenmektedir.

Türkçe Duygu Analizi için oluşturulmuş ve kullanıma açılmış bir veri kümesi bulunmamaktadır. Öncelikle deneylerde kullanmak üzere veri kümeleri oluşturulmuştur. Veri kümeleri oluşturulurken başka diller için oluşturulmuş örneklerden yola çıkılmıştır.

Öncelikle ne tür kaynaklardan veri toplanması gerektiği tespit edilmiş. İlgili kaynaklardan çok kullanılan ve bol miktarda yorum barındıran örnekler tespit edilmiştir.

Veri toplama altyapısı için Microsoft Visual Studio 2013 ortamında C# diliyle bir proje oluşturulmuştur. İlgili veri kaynakları web ortamında olduğundan html ile kodlanmış web sayfaları öncelikle ayrıştırılmıştır. Bu ayrıştırma işlemi için açık kaynaklı Html Agility Pack [54] kütüphanesi kullanılmıştır. Web sitelerinde yorumlar birbirini tekrar eden yapılar içerisinde düzenli olarak tutulmaktadır. Bir yorum sitesinde yorumların nasıl bir yapı içerisinde saklandığı tespit edildikten sonra yinelemeli olarak tüm sayfalar gezilmiş ve tüm yorumlar bulunmuştur. Bulunan yorumlar ilgili kaynaktan alınıp yerel veri tabanına kaydedilmiştir. Projede veri tabanı olarak Microsoft SQL Server 2012 kullanılmıştır. Her bir yorum web ortamına özel karakter kodlamasından kurtarılıp Unicode karakter kodlamasına dönüştürülerek kaydedilmiştir.

Çizelge 8. Film Yorumları Veri tabanı istatistikleri

Puan	Sayı	Toplam
0,5	5.554	
1	16.829	26.740
1,5	1.758	
2	2.599	
2,5	42.694	
3	7.000	62.253
3,5	12.559	
4	84.038	
4,5	15.679	130.210
5	30.493	

Çok kullanılan ve ünlü olan film yorumu sitelerinden beyazperde.com [55] üzerindeki 5.660 civarındaki film için yapılan yorumlar toplanarak bir film yorumları veri tabanı oluşturulmuştur (Bkz. Çizelge 8). Oluşturulan veri tabanı içerisinde 219.203 yorum

bulunmaktadır. Yorumlar “0,5”, ”1”, ”1,5” ,…, ”5” aralığında yorumu yapan kullanıcı tarafından puanlanmış durumdadır. Olumsuz yorum sayısı 26.740 iken olumlu yorum sayısı 130.210’dur.

Dengeli bir veri kümesi oluşturmak adına sayısı az olan olumsuz yorumların tamamı ve olumlu yorumların ise olumsuzlar kadarını alarak bir film veri kümesi oluşturulmuştur. Oluşturulan veri kümesi içerisinde 26.700 olumlu ve 26700 olumsuz yorum bulunmaktadır. Yapılan yorumlardan en uzunu 1.566 kelimeye en kısası ise 1 kelimeye sahiptir. Yorumlarda kullanılan ortalama kelime sayısı 33’tür.

Çizelge 9. Otel yorumları veri tabanı istatistikleri

Puan	Sayı	Toplam
0	4.593	
1 - 10	-	
11 - 20	-	5.802
21 - 30	999	
31 - 40	210	
41 - 50	1.094	
51 - 60	643	6.177
61 - 70	878	
71 - 80	3.562	
81 - 90	1.832	
91 - 99	1.164	6.499
100	3.503	

Yine çok kullanılan otel yorum sitelerinden otelpuan.com [56] üzerindeki 550 civarı otel hakkındaki yorumlar toplanarak bir otel yorumları veri tabanı oluşturulmuştur (Bkz. Çizelge 9). Toplam yorum sayısı 18.478’dir. Yorumlar 100 üzerinden yorumu yapan kullanıcı tarafından puanlanmıştır. Olumsuz yorum sayısı 5.802 iken olumlu yorum sayısı 6.499 olmuştur.

Dengeli bir veri kümesi oluşması için olumlu ve olumsuz kümelerden eşit miktarda alınmıştır. Oluşturulan veri kümesinde 5.800 olumlu ve 5.800 olumsuz yorum kullanılmıştır. Yapılan en uzun yorumda 2.304 kelime varken en kısa yorumda 1 kelime vardır. Yorumlar ortalama 74 kelimeden oluşmaktadır.

Çizelge 10. Derlem İstatistikleri

	Maksimum		Minimum		Ortalama	
Veri kümesi	Varlık	Frekans	Varlık	Frekans	Varlık	Frekans
Otel Yorumları	1308	2304	1	1	63	74
Film Yorumları	1060	1566	1	1	29	33

İki veri kümesi de kelimelere bölünmüş içerisindeki her kelime frekansı ile beraber kaydedilmiştir. Bazı kelimeler yorumlarda sıkça tekrar etmektedir. Çizelge 10'da görüldüğü üzere bazı kelimeler o kadar tekrar etmiştir ki kelimelerin varlıklarına (*occurrence*) göre yorumda 1.308 kelime varken frekanslarına göre sayıldığında yorum aslında 2.304 kelime barındırmaktadır.

5.2.1. Veri Kümesi Temizleme ve Gövdeleme İşlemleri

Veri kümeleri içerisinden alınan dengeli dağılmış veriler öncelikle kelimelere bölünmüştür. Yorum içerisinde geçen kelimeler önce yazım kontrolünden geçirilmiştir. Yazım kontrolü neticesinde yanlış veya eksik yazıldığı tespit edilen kelimeler düzeltilmeye çalışılmıştır. Düzeltilen kelimelerin ekleri silinerek gövdeleri alınmıştır. Tüm bu Doğal Dil İşleme (DDİ) işlemleri bir sonraki alt başlıkta ayrıntılı şekilde anlatılmış olan Zemberek Kütüphanesi [57] kullanılarak gerçekleştirilmiştir. Gövdeler ise yorum içerisinde kaç kez geçtiği bilgisi saklanarak kaydedilmiştir.

Çizelge 11. Veri Kümelerinde Kelime ve Gövde Dağılımı

Veri Kümesi	Etiket	Yorum Sayısı	Kelime	Ayrık Kelime	Ayrık Gövde	Gövde Miktarı
Otel Yorumları	Negatif	5.800	542.868	79.275	7.711	651.082
	Pozitif	5.800	188.768	30.997	4.938	218.668
Film Yorumları	Negatif	26.700	828.430	96.956	8.958	935.757
	Pozitif	26.700	739.333	90.188	8.359	839.392

Veri kümeleri incelendiğinde her veri kümesinde negatif ve pozitif eşit sayıda yorum alınmıştır. Çizelge 11'de görüldüğü üzere kelime sayıları incelendiğinde ise Film Yorumları veri kümesinin aksine Otel Yorumları veri kümesinin pozitif işaretli yorumlarının daha az kelime içerdiği ortaya çıkmaktadır.

Kelimeler eklerinden arındırılıp gövdelere dönüştürüldükten sonra ayrık gövde sayıları kelime sayılarına paralel olarak azalmıştır. Çizelge 11'de Pozitif Otel Yorumlarında görüldüğü üzere 7.711 gövdeden 79.275 kelime üretilmiş tüm veri kümelerinde ortalama bir gövdeden 10 farklı kelime üretilmiştir. Yorumlarda gövde kaç kez geçtiyse kaydedilmiştir. Tüm veri kümelerinde %15 ila %20 bir kelime birden fazla kez kullanılmıştır.

5.2.2. Zemberek Kütüphanesi

Zemberek [57], açık kaynak kodlu, bir Türk Lehçeleri Doğal Dil İşleme kütüphanesidir. Java dilinde yazılmıştır. İçerisinde Türk Lehçelerinin kullanımına uygun şekilde kodlanmış işlevler mevcuttur. TÜBİTAK Bilgem tarafından desteklenmektedir. Türkçe Doğal dil işlemede yaygın olarak kullanılmaktadır. Akademik çalışmaların yanı sıra Open Office ve Libre Office içerisinde eklenti olarak sunulmaktadır.

Lehçe yapısı bilgisi ve DDİ uygulamaları olmak üzere iki ana bölümden oluşmaktadır. Türk Lehçelerinin geneli için programlandığı için öncelikle işlemlerin yapılacağı Lehçenin dilbilgisi özellikleri ayarlanmalıdır.

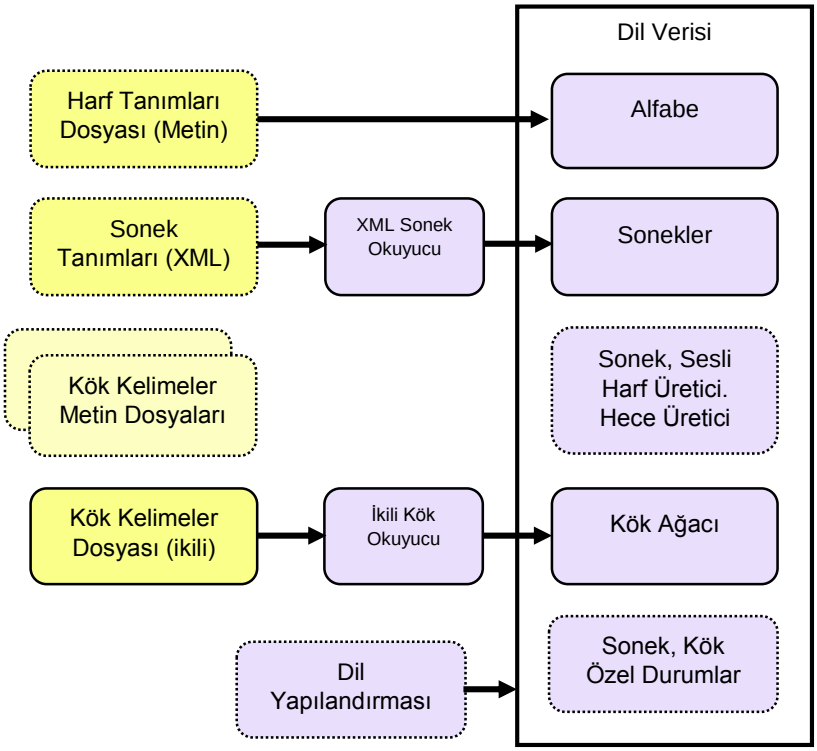
Zemberek Kütüphanesini Türkçe için kullanıma hazırlama aşamaları şu şekildedir:

- Veri kümesinde kullanılan harfleri içeren harfler dosyası hazırlanır.
- Soneklerin yer aldığı sonekler dosyası hazırlanır.
- Kök kelimeleri ve özel durumları içeren kök kelimeler dosyası hazırlanır.
- İsteğe göre yazım denetimi için sık kullanılan kelimelerden oluşan önbellek dosyası hazırlanır.

Hazırlık aşaması tamamlanan Zemberek Kütüphanesinin kullanımı ise şu şekildedir:

- Hece üretici sınıf ile kelime hecelere bölünebilir.
- Sonek üretici sınıf ile verilen kelimeye uygun sonekler üretilebilir.
- Özel durumlu kelimeler için istisnai sonekler üretilebilir.
- Kelime kökleri, gövdeleri ve özel durumları tespit edilebilir.
- Yazım denetimi yapılıp yanlış veya eksik yazılmış kelimeler düzeltilebilir.

- Kelimelerin morfolojik analiz yapılabilir. (Bkz. Şekil 14)



Şekil 14. Zemberek Kütüphanesi Çalışma Şeması

Yapılan çalışmada yorumlar şu sırada işlenmiştir:

I. Kelime içerisindeki sayılar, noktalama işaretleri ve özel karakterler temizlenmiştir.

II. Kelime Zemberek içerisindeki "*kelimeDenetle*" işlevi ile denetlenmiş, yazım yanlışı yoksa V. adıma geçilmiştir.

III. Yazım yanlışının karakter kodlamasından kaynaklanma ihtimali için Zemberek Kütüphanesi içerisindeki "*asciidenTurkceye*" işlevinden faydalanarak kelime önerisi istenmiş eğer bir öneri gelmişse V. adıma geçilmiştir.

IV. Düzeltilemeyen yazım yanlışı için Zemberek Kütüphanesi içerisindeki "*oner*" işlevinden faydalanarak yazım önerisi istenmiştir. Eğer bir öneri gelmemişse kelime işlenememiştir.

V. Yazım denetimi sonucunda doğru yazıldığı kabul edilen kelimelerin gövdeleri Zemberek Kütüphanesi içerisindeki "*kelimeCozumle*" işlevi sayesinde bulunmuştur.

5.2.3. Performans Metrikleri

Duygu Analizi çalışmaları analitikten ziyade deneysel olduğundan yapılan deneyin kapsamını ve doğruluğunu ölçümlemeye ihtiyaç vardır. Etiketli verilerle yapılan Duygu Analizi sonucunda yorumların polaritesi hesaplanmaktadır. Analiz istatistiklerinde pozitif etiketli yorum polaritesi pozitif hesaplanmışsa Doğru Pozitif (*True Positive* kısaca TP) olarak, negatif hesaplanmışsa Yanlış Pozitif (*False Positive* kısaca FP) olarak gösterilir. Yine analiz istatistiklerinde negatif etiketli yorum polaritesi negatif hesaplanmışsa Doğru Negatif (*True Negative* kısaca TN) olarak, pozitif hesaplanmışsa Yanlış Negatif (*False Negative* kısaca FN) olarak gösterilir.(Bkz. Çizelge 12. *Çelişki Matrisi*)

Çizelge 12. Çelişki Matrisi

		Hesaplanan Polarite	
		Pozitif	Negatif
Etiket	Pozitif	TP	FP
	Negatif	FN	TN

Sınıflandırma başarısı genelde Duyarlık (π - *Presicion*) ve Anma (ρ - *Recall*) ile ölçümlenir. Her bir kategori için hesaplanması gereken duyarlık ve anma duygu analizinde olumlu ve olumsuz sınıflamalar için ayrı ayrı hesaplanmaktadır.

$$\pi = \frac{TP}{TP + FP}$$

Eşitlik 7. Duyarlık (*Presicion*)

Duyarlık (π) polaritenin ne ölçüde doğru tahmin edildiğini ölçümlemek için hesaplanmaktadır. Duyarlık polaritesi doğru tahmin edilen cümlelerin sayısının tüm tahmin edilen cümlelerin sayısına olan oranıdır. Başka bir deyişle doğru pozitif (TP) sayısının doğru ve yanlış pozitiflerin (TP + FP) toplamına olan oranıdır.

Olumlu cümlelerin duyarlığı, doğru tahmin edilen olumlu cümle sayısının toplam olumlu cümle sayısına oranı iken, olumsuz cümlelerin duyarlığı ise doğru tahmin edilen negatif cümle sayısının tüm negatif cümlelerin sayısına oranı olarak hesaplanmaktadır.

$$\rho = \frac{TP}{TP + FN}$$

Eşitlik 8. Anma (*Recall*)

Anma (ρ) ise kategoriye ait elemanların ne kadarının tahmin edildiğini ölçümlemek için hesaplanmaktadır. Anma polaritesi doğru tahmin edilen cümlelerin sayısının o kutuptaki tüm cümlelerin sayısına olan oranıdır. Başka bir deyişle doğru pozitif (TP) sayısının doğru pozitif ve yanlış negatiflerin (TP + FP) toplamına olan oranıdır.

$$F_1 = 2 \times \frac{\pi \times \rho}{\pi + \rho} = \frac{2TP}{2TP + FP + FN}$$

Eşitlik 9. F1 Skoru

Duyarlık ve anmanın harmonik ortalamasından oluşan F1 skoru ise hem duyarlığın hem de anmanın beraber ölçümlenebilmesini sağlamaktadır. F1 skoru duyarlık ile anmanın çarpımlarının toplamlarına oranının iki katıdır.

$$ACC = \frac{TP + TN}{TP + FP + TN + FN}$$

Eşitlik 10. Doğruluk

Yapılan tahminlemenin doğruluğu ise doğru tahmin edilen olumlu ve olumsuz cümlelerin toplam sayısının tüm cümlelerin toplam sayısına olan oranıdır.

Bilgiye Erişim konularında yapılan deneyler açıklanan performans metrikleri vasıtası ile karşılaştırılırlar. Bir çalışmanın performans metriklerinin diğerlerinden fazla olması daha doğru ve kapsamlı sonuçlar ortaya koymakta olduğunun göstergesidir.

5.3. Deneyler

Duygu analizi çalışmaları, sözlüksel ve istatistiksel olmak üzere iki ana eksende yapılmaktadır. Çalışma kapsamında hazırlanan veri kümeleri ve Türkçe Duygu Sözlüğü altyapısı ile bir takım deneyler yapılmış ve daha sonra aynı veri kümeleri ile Makine Öğrenmesi yoluyla yapılan deneylerle karşılaştırılmıştır. Yapılan deneyler bazı performans metrikleri ile karşılaştırılmıştır.

Hazırlanan veri kümeleri önceki bölümlerde anlatıldığı üzere temizlenip, yazım hatalarından arındırılıp, gövde haline getirildikten sonra her iki deneyde de kullanılmıştır.

5.3.1. Türkçe Duygu Sözlüğü İle Yapılan Deneyler

Türkçe Duygu Sözlükleri hazırlanırken İngilizce Duygu Analizi çalışmaları dikkate alınmıştır. Öncelikle İngilizce bir Duygu Sözlüğü olan SentiWordNet [23] önceki bölümlerde anlatılan yöntemlerle “Terim-Puan” şekline dönüştürülmüş ve elde edilen İngilizce Duygu Sözlüğü oluşturulan çeviri algoritması ile Türkçeye çeviri yapılmıştır. Çeviri sonrasında bir Türkçe kelimenin birden fazla İngilizce kelimeye karşılık gelmesi ve POS etiketlerinin çeviride ele alınamamasından kaynaklı tekrar eden terimlerin ortalama puanları alınarak Türkçe Duygu Sözlüğü elde edilmiştir. Otomatik iki farklı çeviri metodu benimsenmiş ve her iki metottan da birer sözlük oluşturulmuştur. Daha sonra oluşturulan sözlüklerin birleşiminden üçüncü bir sözlük elde edilmiştir.

Oluşturulan sözlüklerin veri kümelerini oldukça kapsayıcı nitelikte olduğu görülmektedir. Her iki veri kümesi birlikte ele alındığında 65.000 yorumun sadece 96 tanesi sözlükte hiç yer almayan kelimelerden oluşmaktadır.

Negatif etiketli 5800 Otel yorumunun 33 tanesinde sadece pozitif puanı olan kelime yer alırken, 20 tanesinde ise sadece negatif puanı olan kelime yer almış 4 yorum ise sözlükte hiç geçmeyen terimlerden oluşmaktadır. Pozitif etiketli 5800 otel yorumunun 740 tanesi sadece pozitif terimlerden oluşurken, 46 tanesi sadece negatif terimlerden ve 14 tanesi de sözlükte hiç geçmeyen terimlerden oluşmaktadır. (Bkz. Çizelge 13)

Çizelge 13. Duygu Analizine Katılan Ayrık Terim Sayıları

Veri Kümesi	Etiket	Sadece+	Sadece-	Katılmayan	Her İkisi	Toplam
Otel Yorumları	Negatif	33	20	4	5.743	5.800
	Pozitif	740	46	14	5.000	5.800
Film Yorumları	Negatif	498	377	19	25.806	26.700
	Pozitif	1.924	260	59	24.457	26.700

Pozitif etiketli 26.700 film yorumunun 1.924 tanesi sadece pozitif puanı olan terimlerden, 260 tanesi sadece negatif puanı olan terimlerden ve 59 tanesi de sözlükte geçmeyen terimlerden oluşmaktadır. Negatif etiketli 26.700 film yorumunun 498 tanesi sadece pozitif puanı olan terimlerden oluşurken 377 tanesi sadece negatif puanı olan terimlerden oluşmakta ve 19 tanesi ise sözlükte yer almayan terimlerden oluşmaktadır (Bkz. Çizelge 14).

Çizelge 14. Veri Kümesi İçerisindeki Türkçe Duygu Sözlükleri Terim Sayıları

Sözlük	Toplam	Film Yorumları	Otel Yorumları	Her İki Veri Kümesi
TDSp	9.412	2.298	1.950	2.437
TDSs	27.251	3.161	2.627	3.399
TDSv1	27.141	3.187	2.651	3.417

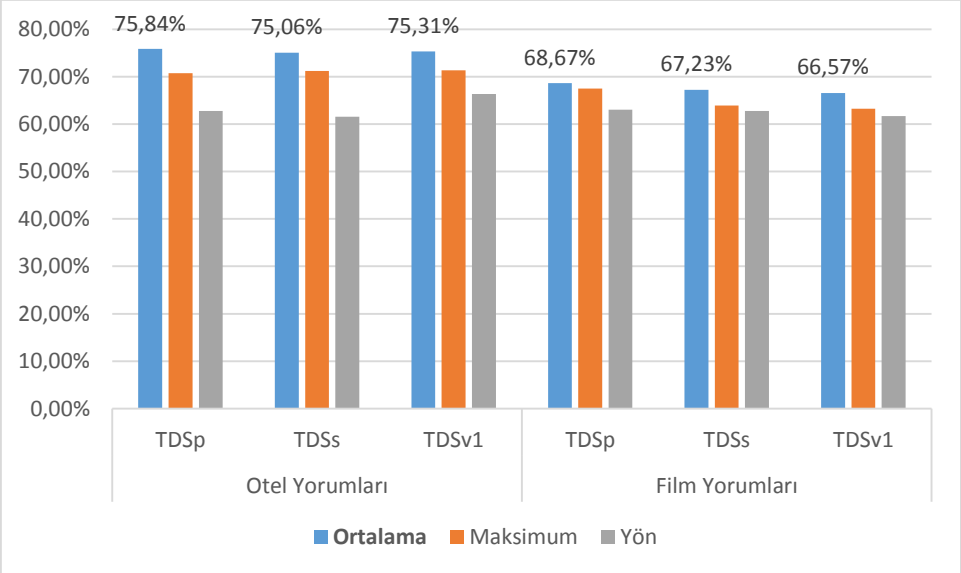
Oluşturulan sözlüklerin veri kümeleri içerisinde karşılaşılan terim sayıları ise veri kümesine göre değişiklik göstermektedir. Örneğin Film Yorumları veri kümesinde ortalama 2.882 kelime karşılık bulurken Otel Yorumları veri kümesinde ortalama 2.409 kelime karşılık bulmuştur. Her iki veri kümesi beraber ele alındığında ise sözlükler içerisindeki kelimelerin ortalama 3.084 tanesi veri kümesi içerisinde yer almaktadır.

Türkçeye çeviri işlemleri sonrasında sözlük içerisinde tekrar eden terimlerin duygu puanları önceki bölümlerde anlatıldığı gibi; ortalama, maksimum ve yön olmak üzere üç farklı şekilde birleştirilmiştir. Sözlük hesaplama yöntemlerinin başarılarını ölçümlemek amaçlı sözlükler ve veri kümeleri ile bazı deneyler gerçekleştirilmiştir.

Çizelge 15. Sözlük Hesaplama Yöntemleri Doğruluk Tablosu

	Sözlük	Ortalama	Maksimum	Yön
Otel Yorumları	TDSp	75,84%	70,73%	62,77%
	TDSs	75,06%	71,23%	61,57%
	TDSv1	75,31%	71,35%	66,36%
Film Yorumları	TDSp	68,67%	67,51%	63,05%
	TDSs	67,23%	63,90%	62,77%
	TDSv1	66,57%	63,24%	61,68%

Çeviri sonrası tekrar eden terimleri birleştirmekte kullanılan yöntemleri (Bkz. Bölüm 4.4) karşılaştırmak için yapılmış olan deneylerin, performans metriklerinden Doğruluk (*Accuracy*), sonuçları karşılaştırılmıştır. Çizelge 15'te görüldüğü gibi, deney sonuçlarına göre terimler "Ortalama" alınarak birleştirildiğinde, her iki veri kümesinde ve oluşturulan her sözlükte yüksek performans sergilemiştir.



Şekil 15. Sözlük Hesaplama Yöntemleri Doğruluk Grafiği

Veri kümeleri içerisindeki yorumlarda tekrar eden kelimeler bulunmaktadır. Tekrar eden kelimelerin yorumun polaritesine sağladıkları katkıyı gözlemlemek için bir takım deneyler yapılmıştır.

Çizelge 16. Veri Kümelerinde Yorum Başına Düşen Kelime Sayıları

	Maksimum		Minimum		Ortalama	
Veri Kümesi	Varlık	Frekans	Varlık	Frekans	Varlık	Frekans
Otel Yorumları	1.308	2.304	1	1	63	74
Film Yorumları	1.060	1.566	1	1	29	33

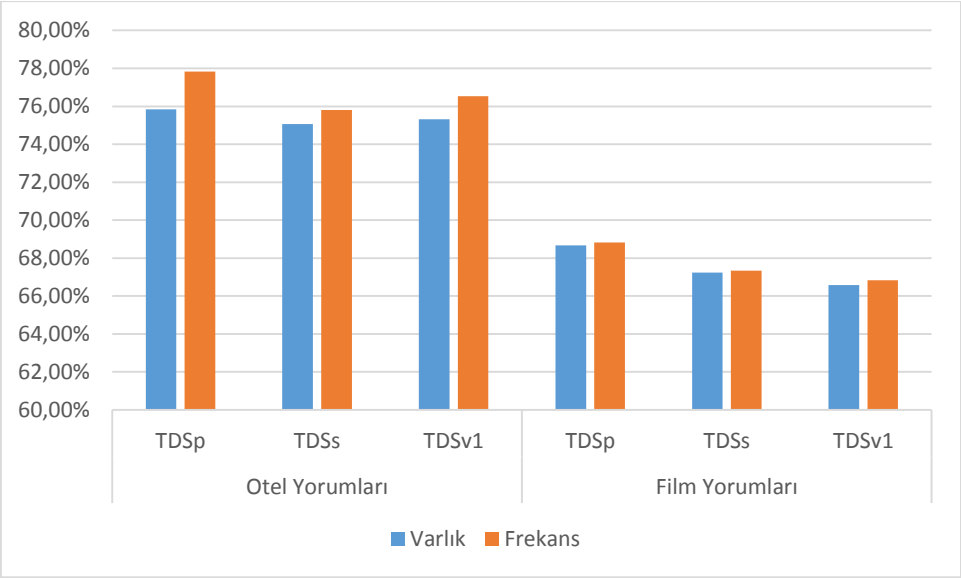
Çizelge 16'da varlık (*occurrence*) ayrık kelime sayısını, frekans ise yorum içerisindeki kelime sayısını ifade etmektedir. Çizelge incelendiğinde kelimelerin çok fazla tekrar ettiği gözlemlenebilir. Doğal dilde tekrar eden kelimeler ifadenin anlamını kuvvetlendirmektedir. Tekrar eden kelimelerin Duygu Analizine etkisini gözlemlemek için deneyler yapılmıştır.

Yapılan deneyler sonucunda, Çizelge 17'de görüldüğü üzere, Doğruluk (*accuracy*) performans metriğine göre yorum içindeki terimlerin frekanslarıyla yapılan hesaplamalar terimlerin varlıklarıyla yapılan hesaplamalara göre daha yüksek performans sağlamıştır.

Çizelge 17. Terim Varlık Frekans Doğruluk Sonuçları

	Sözlük	Varlık	Frekans
Otel Yorumları	TDSp	75,84%	77,82%
	TDSs	75,06%	75,81%
	TDSv1	75,31%	76,54%
Film Yorumları	TDSp	68,67%	68,82%
	TDSs	67,23%	67,34%
	TDSv1	66,57%	66,82%

Veri kümeleri içerisinde yer alan yorum cümlelerinde kelimeler farklı çekimlerde kullanılmaktadır. Aynı zamanda Türkçe Duygu Sözlüğü içerisinde yer alan bazı terimler ekli gövdelerden oluşmaktadır. Sözlüklerde yer alan bu durumun analiz sonuçlarını nasıl etkileyeceğini değerlendirmek için gövdelerine ayırarak Duygu Analizi hesaplamalarına katılan terimlerin gövdelerine ayrılmadan doğrudan analiz yapılması planlanmıştır.



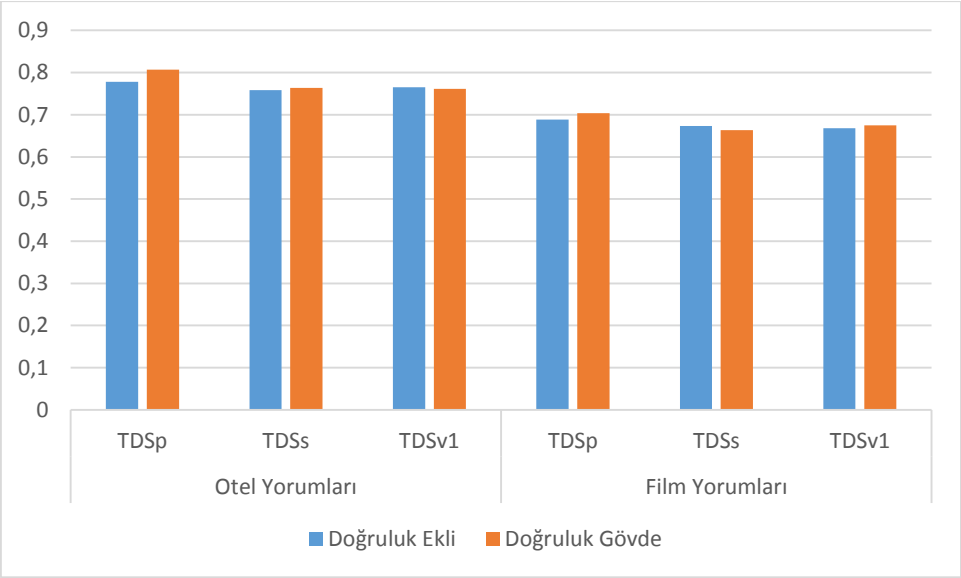
Şekil 16. Terim Varlık Frekans Doğruluk Sonuçları

Yapılan deneylerden elde edilen sonuçlar incelendiğinde yapılan 6 deneyin dördünde kelimenin ”gövde” hali daha başarılı olmuştur. Çizelge 18’de görüleceği üzere deney sonuçları arasındaki farkların az olduğu gözlemlenmiştir.

Çizelge 18. Ekli Kelime ve Gövde Hali Karşılaştırması

Veri Kümesi	Sözlük	Doğruluk	
		Ekli	Gövde
Otel Yorumları	TDSp	0.778242	0.806811
	TDSs	0.758098	0.763339
	TDSv1	0.765384	0.761353
Film Yorumları	TDSp	0.688178	0.703470
	TDSs	0.673386	0.663782
	TDSv1	0.668228	0.674912

Gövdeleme işlemi için harcanan CPU zamanı göz önüne alındığında daha hızlı ve az işlem yaparak Duygu Analizi yapmak istenirse, Türkçe Duygu Sözlüğü ile kelimeleri gövdelerine ayırmadan çok az veri kaybıyla Duygu Analizi yapılabilir sonucu çıkmaktadır.



Şekil 17. Ekli Kelime ve Gövde Hali Karşılaştırması

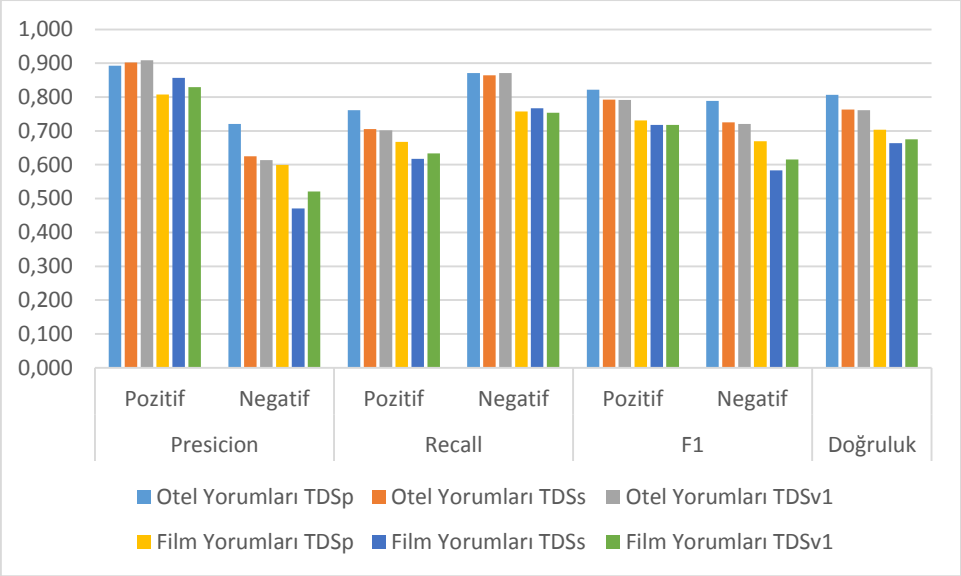
Oluşturulan sözlüklerin birbirleri karşısında başarımlarını ölçümlemek amacıyla bir takım deneyler gerçekleştirilmiştir. Deneylerde Paralel Türkçe Duygu Sözlüğü (TDSp), Seri Türkçe Duygu Sözlüğü (TDSs), Türkçe Duygu Sözlüğü Sürüm 1 (TDSv1) olmak üzere oluşturulan üç farklı sözlük değerlendirilmiştir.

Çizelge 19. Sözlük Metrikleri Karşılaştırması

Veri Kümesi	Sözlük	Duyarlık		Anma		F1		Doğruluk
		Pozitif	Negatif	Pozitif	Negatif	Pozitif	Negatif	
Otel Yorumları	TDSp	0,893	0,721	0,761	0,871	0,822	0,789	0,807
	TDSs	0,902	0,625	0,706	0,865	0,792	0,725	0,763
	TDSv1	0,909	0,614	0,701	0,871	0,792	0,720	0,761
Film Yorumları	TDSp	0,808	0,600	0,668	0,758	0,731	0,669	0,703
	TDSs	0,857	0,471	0,618	0,767	0,718	0,584	0,664
	TDSv1	0,829	0,521	0,633	0,753	0,718	0,616	0,675

Çizelge 19'daki duyarlık metriği ele alındığında pozitifleri bulmakta öne çıkan Türkçe Duygu Sözlüğü Sürüm 1 (TDSv1) negatifleri bulmakta ise Paralel Türkçe Duygu Sözlüğü (TDSp) ile üretilen sonuçların gerisinde kalmıştır. Öte yandan anma metriği ele alındığında ise TDSp hem pozitiflerin hem negatiflerin bulunmasında daha başarılı olmuştur. Duyarlık ve duyarlığın harmonik ortalaması olan F1 metriğinde ise hem pozitiflerde hem negatiflerde yine TDSp başarılı olmuştur. Doğal olarak F1

metriklerinin ortalaması olan Doğruluk (*Accuracy*) metriğinde de Paralel Türkçe Duygu Sözlüğü (TDSp) en başarılı sözlük olmuştur.



Şekil 18. Sözlük Metrikleri Karşılaştırması

İngilizce terimlerin iki veya daha fazla sözlükten aynı karşılığa sahip olanlarından oluşturulan Paralel Türkçe Duygu Sözlüğü (TDSp) diğer iki sözlükten daha başarılı sonuçlar üretmiştir.

5.3.2. Alternatif Yöntemler İle Yapılan Deneyler

Duygu Analizi araştırmaları iki farklı bakış açısı ile yapılmaktadır. Analiz ya bu çalışmada yapıldığı gibi sözlük veya derlem vasıtasıyla ya da makine öğrenmesi veya istatistiksel yöntemler yoluyla yapılmaktadır. Her iki bakış açısının birleşiminden oluşan yöntemler olsa da araştırmalar bu iki ana eksende yoğunlaşmaktadır.

Hem veri kümesinin tutarlılığını gözlemlemek hem de oluşturulan Türkçe Duygu Sözlükleri ile yapılan duygu analizi sonuçlarını başka yöntemler ile karşılaştırmak için aynı veri kümeleriyle makine öğrenmesi yöntemleri kullanılarak bazı deneyler gerçekleştirilmiştir.

Makine öğrenmesi ilgi alanı veri üzerinden öğrenme olan bir bilim dalıdır. Makine öğrenmesi algoritmalarıyla, girdilere göre oluşan model yardımıyla öngörü yapılabilir

veya karar verilebilir. Makine öğrenmesi programlanmış içerikten daha iyi sonuçlar üretebilmektedir.

Tüm veriyi olduğu gibi makine öğrenmesi algoritmasına dâhil etmek deney sonuçlarını olumsuz etkilemektedir. Makine öğrenmesinde tüm veri yerine öznitelik (*feature*) olarak adlandırılan, veri kümesini temsil edecek ve ayrıştırıcı özellikler taşıyan nitelikler kullanılmaktadır. Bu temsil yeteneği yüksek öznitelikleri seçmek ayrı bir araştırma konusudur.

Veri kümesi ile yapılan öğrenme işlemi sonucunda bir model ortaya çıkar. Model vasıtasıyla yeni girdiler hakkında öngörüler yapılabilir. Makine öğrenmesi teknikleri gözetimli, gözetimsiz ve yarı gözetimli olmak üzere üç farklı şekilde yapılmaktadır. Veri kümesi içerisinde tüm veriler etiketliyse gözetimli makine öğrenmesi, veri kümesi içerisinde bazı veriler etiketli bazıları etiketsiz ise yarı gözetimli makine öğrenmesi, veri kümesi içerisinde hiç etiketli veri yoksa gözetimsiz makine öğrenmesi teknikleri uygulanmaktadır. Yürütülen çalışmada kullanılan veri kümeleri içerisindeki tüm veriler etiketli olduğu için deneylerde gözetimli makine öğrenmesi teknikleri kullanılmıştır.

Gözetimli makine öğrenmesi teknikleri veriyi sınıflamak için kullanılmaktadır. Genelde doküman sınıflama için kullanılan gözetimli makine öğrenmesi teknikleri, duygu sınıflama için de kullanılmış ve kötü olmayan sonuçlar elde edilmiştir.

Duygu sınıflama işlemi ile veri kümesi içerisindeki olumlu ve olumsuz etiketli veriler yardımıyla bir model oluşturulur. Oluşturulan model üzerinden duygu sınıflaması gerçekleştirir. Duygu sınıflama bir tür Duygu Analizi yöntemidir.

Son zamanlarda Türkçe için yapılmış ve karşılaştırmaya uygun olan gözetimli makine öğrenmesi çalışmalarından Akba ve ekibinin [38] deney ortamı karşılaştırma yapmak üzere tekrar oluşturulmuştur.

Deneyde Waikato Üniversitesinde geliştirilmiş olan WEKA (*Waikato Environment for Knowledge Analysis*) [58] makine öğrenmesi algoritmaları koleksiyonu ile çalışılmıştır. Çalışmaya göre daha başarılı sonuçlar verdiği için gözetimli makine öğrenmesi algoritmalarından SVM(*Support Vector Machine*) sınıflayıcı olarak seçilmiştir. Öznitelik seçme yöntemi olarak ise x^2 (*Chi-Square*) tercih edilmiş ve deneyde en etkili 375 öznitelik kullanılmıştır.

```
@relation Otel Yorumları
@ATTRIBUTE Öznitelik1 NUMERIC
@ATTRIBUTE Öznitelik2 NUMERIC
@ATTRIBUTE Öznitelik3 NUMERIC
.
.
300 Öznitelik
.
.
@ATTRIBUTE Öznitelik299 NUMERIC
@ATTRIBUTE Öznitelik300 NUMERIC
@ATTRIBUTE etiket {Positive,Negative}
@data
0,2,5,2,1,1,2,3,2,.. 300 Öznitelik ...,0,0,0,1,Negative
0,0,0,0,0,0,0,0,1,.. 300 Öznitelik ...,0,0,1,0,Negative
0,1,8,0,1,0,1,1,0,.. 300 Öznitelik ...,0,1,0,0,Negative
.
.
5800 adet yorum
.
.
2,1,3,0,1,1,0,2,1,.. 300 Öznitelik ...,0,0,3,0,Negative
0,0,0,0,0,0,0,0,0,.. 300 Öznitelik ...,2,0,0,0,Positive
0,0,2,0,0,0,0,0,0,.. 300 Öznitelik ...,0,0,1,0,Positive
```

Şekil 19. Makine Öğrenmesi Örnek Girdi Dosyası

Her iki veri kümesi içerisinde bulunan toplam 65.000 yorumun yarısı 32.500'ü her veri kümesi için ayrı olmak üzere öğrenme kümesi (*train set*) kalan 32.500'ü deney kümesi (*test set*) olarak deney verisi hazırlanmıştır. Sözlüklü yöntemdeki gibi kelimeler önce Zemberek Kütüphanesi [57] ile düzeltilmiş, gövde haline getirilmiş ve yorum içerisindeki frekanslarıyla beraber kaydedilmiştir.

Deney için kelimelerin gövde halleri alınmış ve kelimelere birer benzersiz kimlik verilmiştir. Kelimeler kimlikleriyle ve yorum içerisindeki frekanslarıyla birlikte kaydedilmiştir. Seçilen çalışmadakiyle aynı öznitelik seçme yöntemleri kullanılarak yapılan öznitelik seçimi sonucunda Otel yorumları öğrenme kümesi içerisinden en fazla 300 öznitelik çıkarılabilmiştir. Diğer öğrenme kümesi içinde aynı öznitelik sayısı tercih edilmiştir. Seçilen öznitelikler ile temsil edilemeyen toplam 831 yorum vardır.

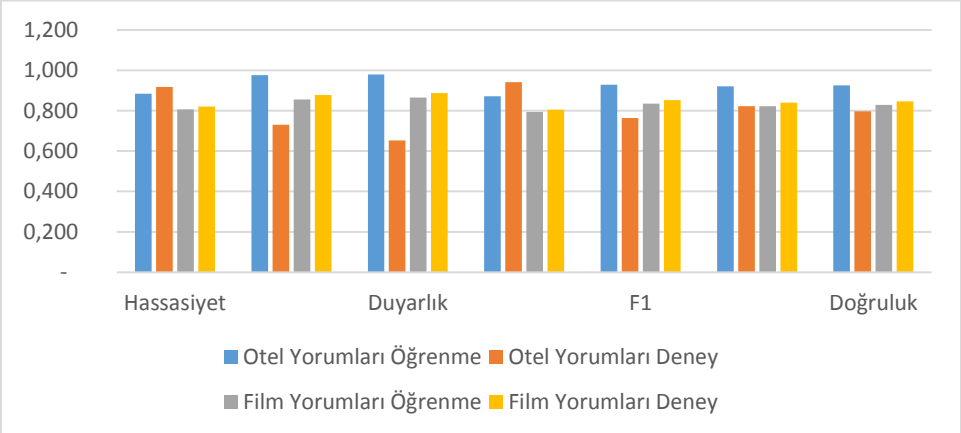
Öğrenme ve deney kümeleri girdi dosyaları hazırlanırken Şekil 19'da görüldüğü gibi her bir yorum için, seçilen özniteliklerin yani ayrıştırıcı kelimelerin yorumdaki

frekansları yazılmış ve yorumun etiketi eklenmiştir. Elde edilen öğrenme kümesi WEKA ortamına yüklenmiş, “-S 0 -K 2 -D 3 -G 0.0 -R 0.0 -N 0.5 -M 40.0 -C 1.0 -E 0.001 -P 0.1” parametreleri ve LibSVM [59] algoritması yardımıyla sınıflama işlemi gerçekleştirilmiştir.

Çizelge 20. Makine Öğrenmesi (SVM) Deney Sonuçları

Veri Kümesi	Küme	Toplam Yorum Sayısı	Temsil Edilemeyen Yorum Sayısı	Duyarlık		Anma		F1		Doğruluk
				Pozitif	Negatif	Pozitif	Negatif	Pozitif	Negatif	
Otel Yorumları	Öğrenme	5.800	133	0,884	0,976	0,979	0,872	0,929	0,921	0,925
	Deney	5.800	43	0,917	0,731	0,653	0,941	0,763	0,823	0,797
Film Yorumları	Öğrenme	26.700	341	0,807	0,855	0,865	0,793	0,835	0,823	0,829
	Deney	26.700	314	0,820	0,878	0,888	0,805	0,852	0,840	0,846

Eğitilen sistemle bir model oluşturulmuş ve sistem yeni elemanlar için öngörü yapılabilir hale gelmiştir. Oluşturulan model ile deney kümesi için yapılan öngörüler veri kümesi etiketleriyle karşılaştırıldığında %80 ila %85 aralığında başarı elde edildiği tespit edilmiştir.



Şekil 20. Makine Öğrenmesi (SVM) Deney Sonuçları

Yapılan deneyler neticesinde Paralel Türkçe Duygu Sözlüğü ile alınan sonuçlarının başarısı Makine öğrenmesi teknikleriyle alınan sonuçlar ile karşılaştırıldığında

üretilen sözlüğün alternatif yöntemler ile rekabet edebilecek olduğu gözler önüne serilmektedir.

Çizelge 21. TDSp ve SVM Deney İle Duygu Analizi Sonuçları

Veri Kümesi	Duyarlık		Anma		F1		Doğruluk
	Pozitif	Negatif	Pozitif	Negatif	Pozitif	Negatif	
Otel Yorumları (TDSp)	0,893	0,721	0,761	0,871	0,822	0,789	0,807
Otel Yorumları (SVM deney)	0,917	0,731	0,653	0,941	0,763	0,823	0,797
Film Yorumları (TDSp)	0,808	0,600	0,668	0,758	0,731	0,669	0,703
Film Yorumları (SVM deney)	0,820	0,878	0,888	0,805	0,852	0,840	0,846

6. SONUÇ

Bu tezde sözlük kullanarak Duygu Analizi yapılması amaçlanmıştır. Dilimiz için daha önce hazırlanmamış ve hazırlanması oldukça zahmetli olan Türkçe Duygu Sözlüğü, İngilizce için hazırlanmış bir duygu sözlüğünü otomatik tercüme yöntemiyle Türkçeleştirerek oluşturulmuştur.

Farklı tercüme algoritmalarıyla daha doğru çeviri yapılmaya çalışılmıştır. Sözcüklerin karşılığı birden fazla sözlükte aranmış, bu sayede kontrollü çeviri yapılması hedeflenmiştir.

Hem çeviri yapılarak duygu analizi yapılabilirliğinin hem de farklı çeviri algoritmalarıyla oluşturulan sözlüklerin doğruluğunu kontrol etmek için bir takım deneyler yapılmıştır. Deneyleri gerçekleştirmek için öncelikle doğal veriden iki farklı veri kümesi oluşturulmuştur. Elde edilen veri kümeleri ile oluşturulan sözlük denenmiş ve başarılı sonuçlar elde edilmiştir.

Oluşturulan sözlük her iki veri kümesini de oldukça kapsayıcı niteliktedir. Sözlük içerisindeki 3.400 civarı sözcük veri kümelerinde geçmektedir. Her iki veri kümesinde neredeyse her yorum içerisinden en az bir kelime geçtiği tespit edilmiştir.

Karşılaştırılma yapılabilmesi için bir makine öğrenmesi yöntemiyle aynı deneyler tekrar yapılmıştır. Sözlüklü yöntemle elde edilen sonuçlar makine öğrenmesi yöntemiyle elde edilen sonuçlarla karşılaştırılmıştır.

Çeviri yöntemlerinden bir ya da daha fazla İngilizce Türkçe sözlükte karşılığının aynı olması durumunu temsil eden Paralel Türkçe Duygu Sözlüğü ile İngilizce Türkçe sözlüklerin her hangi birinde karşılığı bulunma durumunu temsil eden Seri Türkçe Duygu Sözlüğü karşılaştırılmış ve Paralel Türkçe Duygu Sözlüğünün başarısının yüksek olduğu gözlemlenmiştir. Türkçe Duygu Sözlüğü oluşturulurken paralel anlamı olanlara öncelik verilmiştir.

Çeviri sonrası tekrar eden terimlerin puanları birkaç farklı yöntemle birleştirilmiş ve bu yöntemler karşılaştırıldığında ortalama alınarak birleştirmenin diğerlerinden daha başarılı sonuçlar ürettiği gözlemlenmiştir.

Deneyler esnasında veri kümelerinde bulunan yorumlar kelimelere bölünmüş, her kelimenin gövde hali bulunmuştur. Kelimelerin hem ekli hem de gövde halleriyle

yapılan deneylerde gövde halinin %2 daha yüksek başarı ürettiği bilgisine ulaşılmıştır.

Sözcük gövdelerinin yorum içerisinde bulunma durumu temsil eden varlık bilgisi ile yorum içerisinde kaç kez geçtiğini temsil eden frekans bilgisi karşılaştırılmış ve frekans bilgisi ile yapılan deneylerin az da olsa başarılı olduğu tespit edilmiştir.

Oluşturulan Türkçe Duygu Sözlüğü ile yapılan duygu analizi sonucunda %80 civarı başarı elde edilmiştir. Hem İngilizce için yapılan deneylerin sonuçları ile hem de tez kapsamında yapılmış olan makine öğrenmesi deneylerinin sonuçları ile karşılaştırıldığında Türkçe Duygu Sözlüğünün başarısının yaklaşık olarak aynı düzeyde olduğu ortaya konulmuştur.

Gelecek çalışmalarda tez kapsamına alınmayan nötr ifadeleri tespit etmek için İngilizce duygu sözlüğünün tamamının Türkçeye çevrilmesi planlanmaktadır. Metinlerin içerdikleri olumsuzluk ifadelerini tespit etme işlevi deney başarısını arttırmak için kullanılabilir. Metinlerde geçen abartma, hiciv ve alay gibi anlamsal ifadelerin tespit edilmesi başarıyı arttırmak için denenebilir. İfade içerisinde geçen gülen yüzlerin(*smiley*) olumlu ve olumsuz katkısı sözlüğe eklenebilir. Oluşturulan sözlük neticesinde ortaya çıkan ifade puanlarının makine öğrenmesi yöntemlerinde öznitelik olarak kullanılması ve hibrit yöntemler geliştirilmesi planlanmaktadır.

KAYNAKLAR

[1] C. Fellbaum and R. Tengi, "About WordNet," Princeton University, 7 November 2013. [Online]. Available: http://wordnet.princeton.edu/wordnet/. [Accessed 20 August 2014].

[2] L. Bing, "Sentiment analysis and opinion mining," *Synthesis Lectures on Human Language Technologies,* vol. 5, no. 1, pp. 1-167, 2012.

[3] T. Nasukawa and J. Yi, "Sentiment analysis: Capturing favorability using natural language processing," in *Proceedings of the 2nd international conference on Knowledge capture*, Sanibel Island, FL, USA , 2003.

[4] K. Dave, S. Lawrence and M. P. David, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews," in *Proceedings of the 12th international conference on World Wide Web*, ACM, 2003.

[5] C. D. Elliott, "The Affective Reasoner: A process model of emotions in a multi-agent system," Northwestern University, Evanston, IL, USA, 1992.

[6] A. Ortony, The cognitive structure of emotions, Cambridge university press, 1990.

[7] R. A. Stevenson, J. A. Mikels and T. W. Jam, "Characterization of the affective norms for English words by discrete emotional categories," *Behavior Research Methods,* vol. 39, no. 4, pp. 1020-1024, 2007.

[8] B. Pang, L. Lee and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, Association for Computational Linguistics, 2002.

[9] B. Pang and L. Lillian, "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales.," in *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics.*, Association for Computational Linguistics, 2005.

[10] M. Hu and L. Bing, "Mining and summarizing customer reviews.," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining.*, ACM, 2004.

[11] P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL, 2002.

[12] B. Pang and L. Lee, "A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts," in *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, ACL, 2004.

[13] K. Dave, S. Lawrence and D. M. Pennock, "Mining the peanut gallery: opinion extraction and semantic classification of product reviews," in *Proceedings of the 12th international conference on World Wide Web*, ACM, 2003.

[14] C. Fellbaum, "Introduction to WordNet: An On-line Lexical Database," *International Journal of Lexicography,* vol. 3, no. 4, pp. 235-244, 1990.

[15] A. Valitutti, C. Strapparava and S. Oliviero, "Developing Affective Lexical Resources," *PsychNology Journal,* vol. 2, no. 1, pp. 61-83, 2004.

[16] S. M. Kim and E. Hovy, "Determining the sentiment of opinions," in *Proceedings of the 20th international conference on Computational Linguistics*, ACL, 2004.

[17] S. Mohammad, C. Dunne and B. Dorr, "Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus," in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, ACL, 2009.

[18] J. Kamps, M. J. Marx, R. J. Mokken and M. De Rijke, "Using wordnet to measure semantic orientations of adjectives," in *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC)*, Paris, 2004.

[19] G. K. Williams and S. S. Anand, "Predicting the Polarity Strength of Adjectives Using WordNet," in *3rd Int'l AAAI Conference on Weblogs and Social Media*, California, 2009.

[20] J. Steinberger, M. Ebrahim, M. Ehrmann, A. Hurriyetoglu, M. Kabadjov, P. Lenkova, R. Steinberger, H. Tanev, S. Vazquez and V. Zavarella, "Creating sentiment dictionaries via triangulation," *Decision Support Systems,* vol. 53, no. 4, pp. 689-694, 2012.

[21] A. Esuli and F. Sebastiani, "Determining the semantic orientation of terms through gloss classification," in *Proceedings of the 14th ACM international conference on Information and knowledge management*, ACM, 2005.

[22] A. Esuli and F. Sebastiani, "Determining Term Subjectivity and Term Orientation for Opinion Mining.," in *Proceedings of EACL-06, 11th Conference of the European Chapter of the Association for Computational Linguistics*, Trento, 2006.

[23] A. Esuli and F. Sebastiani, "SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining," in *Proceedings of LREC-06, 5th Conference on Language Resources and Evaluation*, Genova, 2006.

[24] M. Taboada, J. Brooke, M. Tofiloski, K. Voll and M. Stede, "Lexicon-Based Methods for Sentiment Analysis," *Computational Linguistics,* vol. 37, no. 2, pp. 267-307, 2011.

[25] B. Ohana and T. Brendan, "Sentiment classification of reviews using SentiWordNet," in *9th. IT & T Conference*, Dublin, 2009.

[26] A. Hamouda and M. Rohaim, "Reviews classification using sentiwordnet lexicon," *The Online Journal on Computer Science and Information Technology,* vol. 2, no. 1, pp. 120-123, 2011.

[27] U. Eroğul, "Sentiment analysis in Turkish. master's thesis," Middle East Technical University, Ankara, 2009.

[28] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning,* vol. 3, no. 20, p. 273297, 1995.

[29] H. Nizam and S. S. Akın, "Sosyal Medyada Makine Öğrenmesi ile Duygu Analizinde Dengeli ve Dengesiz Veri Setlerinin Performanslarının Karşılaştırılması," in *XIX. Türkiye'de İnternet Konferansı*, İzmir, 2014.

[30] Z. Boynukalın, "Makine öğrenimi teknikleriyle Türkçe metinlerde duygu analizi. master's thesis," Midde, Middle East Technical University, 2012.

[31] A. G. Vural, B. B. Cambazoglu, P. Senkul and Z. O. Tokgoz, "A framework for sentiment analysis in Turkish: Application to polarity detection of movie reviews in Turkish," in *Computer and Information Sciences III*, London, Springer, 2013, pp. 437-445.

[32] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai and A. Kappas, "Sentiment strength detection in short informal text," *Journal of the American Society for Information Science and Technology,* vol. 61, no. 12, pp. 2544-2558, 2010.

[33] M. Meral ve B. Diri, «Twitter Üzerinde Duygu Analizi,» %1 içinde *Signal Processing and Communications Applications Conference (SIU)*, Trabzon, 2014.

[34] İ. Mayda and Ç. Aytekin, "SOSYAL MEDYADA REKABET ANALİZİ İÇİN KARŞILAŞTIRMA GÖREVİNE YÖNELİK FİKİR MADENCİLİĞİ MODELİ," *Journal Academic Marketing Mysticism Online,* vol. 7, no. 27, pp. 414-425, 2013.

[35] O. Cakmak, A. Kazemzadeh , D. Can, S. Yildirim and S. Narayanan, "Root-word analysis of Turkish emotional language," in *Corpora for Research on Emotion Sentiment & Social Signals*, 2012.

[36] O. Cakmak, A. Kazemzadeh, S. Yildirim and S. Narayanan, "Using interval type-2 fuzzy logic to analyze turkish emotion words," in *Signal Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Hollywood, California, 2012.

[37] M. Kaya, "Sentiment analysis of Turkish political columns with transfer learning. master's thesis," Middle East Technical University, Ankara, 2013.

[38] F. Akba, A. Uçan, E. Akçapınar Sezer and H. Sever, "Assessment of feature selection metrics for sentiment analyses: Turkish movie reviews," in *8th European Conference on Data Mining 2014*, Lizbon, 2014.

[39] A. Balahur, M. Turchi, R. Steinberger, J.-M. Perea-Ortega, G. Jacquet, D. Küçük, V. Zavarella and A. El Ghali, "Resource Creation and Evaluation for Multilingual Sentiment Analysis in Social Media Texts," in *The 9th edition of the Language Resources and Evaluation Conference*, Reykjavik, 2014.

[40] E. Akbaş, "ASPECT BASED OPINION MINING ON TURKISH TWEETS," Bilkent University, Ankara, 2012.

[41] F. Çetin and M. Amasyalı, "Supervised and traditional term weighting methods for sentiment analysis," in *Signal Processing and Communications Applications Conference (SIU)*, Girne, 2013.

[42] M. Çetin and F. Amasyalı, "Active learning for Turkish sentiment analysis," in *Innovations in Intelligent Systems and Applications (INISTA)*, Albena, Bulgaria, 2013.

[43] C. Özsert and A. Özgür, "Word polarity detection using a multilingual approach," in *Computational Linguistics and Intelligent Text Processing*, Springer, 2013, pp. 75-82.

[44] J. Wiebe, T. Wilson and C. Cardie, "Annotating expressions of opinions and emotions in language," *Language Resources and Evaluation,* vol. 39, no. 2-3, pp. 165-210, 2005.

[45] P. Stone, D. Dunphy and M. Smith, The General Inquirer: A Computer Approach to Content Analysis, MIT press, 1966.

[46] O. Bilgin, Ö. Çetinoğlu and K. Oflazer, "Building a wordnet for Turkish," *Romanian Journal of Information Science and Technology,* Vols. 1-2, no. 7, pp. 163-172, 2004.

[47] M. F. Amasyalı, "Türkçe Wordnet'in Otomatik Olarak Oluşturulması," in *IEEE 13. Sinyal İşleme ve İletişim Uygulamaları Kurultayı (SİU-2005)* , Kayseri, 2005.

[48] "tureng.com," [Online]. Available: http://tureng.com. [Accessed 20 August 2014].

[49] "www.zargan.com," [Online]. Available: http://www.zargan.com/. [Accessed 20 August 2014].

[50] "tr.bab.la," [Online]. Available: http://tr.bab.la/. [Accessed 20 August 2014].

[51] "Google Translate API," Google Inc., [Online]. Available: http://developers.google.com/translate/. [Accessed 20 August 2014].

[52] M. Guerini, L. Gatti and M. Turchi, "Sentiment analysis: How to derive prior polarities from SentiWordNet," *arXiv preprint arXiv:1309.5843,* 2013.

[53] P. Törnberg, "SentiWordNet Sample code," 2013. [Online]. Available: http://sentiwordnet.isti.cnr.it/. [Accessed July 2014].

[54] S. Mourier, "Html Agility Pack agile HTML parser," [Online]. Available: http://htmlagilitypack.codeplex.com/. [Accessed 22 January 2014].

[55] "beyazperde.com," [Online]. Available: http://www.beyazperde.com/. [Accessed 10 April 2014].

[56] "otelpuan.com," [Online]. Available: http://www.otelpuan.com/. [Accessed 15 April 2014].

[57] A. A. Akın and M. D. Akın, "Zemberek, an open source NLP framework for Turkic Languages," 2007. [Online]. Available: http://zemberek.googlecode.com/files/zemberek_makale.pdf. [Accessed 3 8 2014].

[58] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten, "The WEKA Data Mining Software: An Update," *SIGKDD Explorations,* vol. 1, no. 11, 2009.

[59] Y. El-Manzalawy and H. Vasant, "WLSVM: Integrating libsvm into WEKA environment," 2005. [Online]. Available: http://www. cs. iastate. edu/yasser/wlsvm. [Accessed 10 8 2014].

EKLER

Ek 1. Paralel Türkçe Duygu Sözlüğü Örnek Terimler

Terim	Ortalama	Maksimum	Yön	İngilizce terimler
abanoz	-0.068	-0.068	-1.0	ebony
abartı	0.108	0.125	1.0	overstatement,exaggeration
abartılı	0.021	0.125	1.0	exaggerated,exaggeratedly
abartılmış	0.125	0.125	1.0	overstated
abartmak	0.521	0.625	1.0	overstate,exaggerate
abla	0.375	0.375	1.0	big_sister
abluka	-0.090	-0.090	-1.0	blockade
abone	0.068	0.068	1.0	subscriber
aç	-0.125	-0.125	-1.0	hungry
acayip	-0.108	-0.333	-1.0	queer,queer,queer,weird,kinky
aceleci	0.042	0.625	1.0	impetuous,pushy
acelecilik	0.125	0.125	1.0	hurrying,rashness
acemice	0.167	0.167	1.0	ineptly
acemilik	0.125	0.125	1.0	inexperience
açgözlü	-0.017	-0.068	-1.0	rapacious,greedy
açgözlülük	-0.101	-0.136	-1.0	rapacity,covetousness,greediness
acı	-0.348	-0.625	-1.0	anguish,anguish,bitter,bitter,bitter
açı	-0.069	-0.069	-1.0	angle
açık	0.312	0.625	1.0	open,on,obvious,luculent,ruttish,explicit
açıkça	0.175	0.250	1.0	openly,bluntly,clearly,explicitly
açıkçası	0.438	0.500	1.0	frankly,obviously
açıkgözlük	0.188	0.250	1.0	incisiveness,astuteness
açıklama	0.041	0.125	1.0	description,explanation,remark,statement
açıklamak	0.136	0.136	1.0	explain
açıklanabilir	0.625	0.625	1.0	explicable,explainable
açıklanamaz	-0.375	-0.375	-1.0	inexplicable
açıklanmamış	-0.375	-0.375	-1.0	unexpressed
açıklayıcı	0.125	0.125	1.0	elucidative
acıklı	-0.458	-0.458	-1.0	tragic
açıklık	0.102	0.375	1.0	openness,explicitness
acılık	-0.296	-0.300	-1.0	acridness,bitterness
açılır	-0.250	-0.250	-1.0	pop_fly

açılış	0.014	0.014	1.0	opening
açılmamış	-0.500	-0.500	-1.0	unopened
açılmış	-0.136	-0.136	-1.0	opened
acıma	-0.417	-0.417	-1.0	commiseration
acımak	-0.125	-0.125	-1.0	deplore
acımasız	-0.335	-0.667	-1.0	pitiless,relentless,brutal,merciless,bowelless
acımasızca	0.042	-0.500	1.0	remorselessly,pitilessly,atrociously,cruelly,mercilessly
acımasızlık	-0.319	-0.542	-1.0	pitilessness,ruthlessness,mercilessness
acımış	-0.542	-0.542	-1.0	rancid
acınacak	-0.605	-0.750	-1.0	pitiable,piteous,pitiful
açısal	0.125	0.125	1.0	angular
aciz	-0.648	-0.648	-1.0	helpless
acizlik	-0.271	-0.375	-1.0	helplessness,impotency
açlık	0.021	0.083	1.0	hunger,hunger,starvation
açma	-0.250	-0.250	-1.0	swingy
açmak	0.067	0.125	1.0	turn-on,turn_on
ad	0.062	0.108	1.0	name,name
adaçayı	0.159	0.250	1.0	sage,sage
adalet	0.248	0.375	1.0	justice,fairness
adaletsizce	0.250	0.250	1.0	iniquitously
adaletsizlik	0.042	0.042	1.0	injustice
adam	0.007	0.007	1.0	man
adama	0.383	0.383	1.0	dedication
adamak	0.074	0.080	1.0	devote,dedicate
adenokarsinom	-0.125	-0.125	-1.0	adenocarcinoma
adenopati	0.125	0.125	1.0	adenopathy
adî	-0.318	-0.318	-1.0	sleazy
adil	0.500	0.500	1.0	equitable
adilane	0.250	0.250	1.0	equitably
adileştirme	-0.083	-0.083	-1.0	vulgarization
adileştirmek	0.034	0.034	1.0	vulgarize
adım	-0.064	-0.064	-1.0	step
adına	0.042	0.042	1.0	behalf
adinamik	-0.583	-0.583	-1.0	adynamic
adlandırma	0.500	0.500	1.0	naming
adli	-0.250	-0.250	-1.0	forensic

adrenalin	0.125	0.125	1.0	adrenaline
adrenerjik	0.250	0.250	1.0	adrenergic
adres	0.017	0.017	1.0	address
adressiz	0.125	0.375	1.0	undirected,unaddressed
adsız	-0.750	-0.750	-1.0	untitled
adventif	-0.750	-0.750	-1.0	adventive
adyabatik	-0.625	-0.625	-1.0	adiabatic
aerobik	0.167	0.167	1.0	aerobic
aerodinamik	0.042	0.042	1.0	aerodynamic
af	0.080	0.125	1.0	amnesty,forgiveness
afacan	-0.250	-0.250	-1.0	impish
afaki	-0.625	-0.625	-1.0	aphakia
affedici	0.250	0.250	1.0	absolvitory
affedilebilir	0.458	0.542	1.0	excusable,forgivable
affedilmez	-0.375	-0.625	-1.0	unforgivable,unpardonable
affetmek	0.165	0.333	1.0	remit,remit,forgive
affetmez	-0.750	-0.750	-1.0	unforgiving
afgan	0.500	0.500	1.0	loverlike
aforoz	-0.042	-0.042	-1.0	anathema
ağ	0.121	0.121	1.0	web
ağaçlık	-0.250	-0.250	-1.0	wooded
ağaçlılık	0.250	0.250	1.0	woodsiness
ağaçsız	-0.250	-0.250	-1.0	treeless
ağarık	-0.333	-0.333	-1.0	bleached
ağarmış	0.250	0.250	1.0	hoary
agav	-0.042	-0.042	-1.0	cantala
ağıl	-0.125	-0.125	-1.0	pinfold
ağır	-0.093	-0.250	-1.0	onerous,weighty,heavy
ağırbaşlı	0.122	0.500	1.0	overmodest,solemn,earnest,imperturbable
ağırbaşlılık	0.073	0.375	1.0	sedateness,solemness,soberness,equanimity
ağırkanlı	0.625	0.625	1.0	phlegmatical,phlegmatic
ağırlaştırıcı	-0.250	-0.250	-1.0	aggravating

Ek 2. Seri Türkçe Duygu Sözlüğü Örnek Terimler

Terim	Ortalama	Maksimum	Yön	İngilizce terimler
abanoz	-0.068	-0.068	-1.0	ebony

abartı	0.108	0.125	1.0	overstatement,exaggeration
abartılı	-0.100	-0.875	-1.0	overact,overdone,overacting,exaggerated,exaggeratedly
abartılması	0.375	0.375	1.0	overestimation
abartılmış	0.125	0.125	1.0	overstated
abartırız	0.625	0.625	1.0	hyperbolise
abartısız	0.750	0.750	1.0	understated
abartma	0.313	0.375	1.0	overestimate,overcharge
abartmak	0.556	0.625	1.0	hyperbolize,overstate,exaggerate
abartmalı	-0.542	-0.542	-1.0	flatulent
abazanlık	0.250	0.250	1.0	horniness
abdias	-0.167	-0.167	-1.0	abdias
abes	0.250	0.250	1.0	nugatory
abetalipoproteinemi	-0.750	-0.750	-1.0	abetalipoproteinemia
abidance	0.341	0.341	1.0	abidance
abiyotrofi	-0.250	-0.250	-1.0	abiotrophy
abla	0.375	0.375	1.0	big_sister
ablasyon	-0.125	-0.125	-1.0	ablated
ablatif	0.208	0.208	1.0	ablative
ablefaron	0.500	0.500	1.0	ablepharia
abluka	-0.090	-0.090	-1.0	blockade
abone	0.068	0.068	1.0	subscriber
abreaksiyon	0.375	0.375	1.0	abreaction
abselenme	-0.083	-0.083	-1.0	suppuration
absorbanlar	0.500	0.500	1.0	absorbate
absorpsiyon	-0.077	-0.077	-1.0	absorption
abuk	-0.250	-0.250	-1.0	honky
aç	-0.125	-0.125	-1.0	hungry
acaba	0.415	0.602	1.0	wonder,wonder
acardia	-0.500	-0.500	-1.0	acardia
acayip	-0.126	-0.375	-1.0	kooky,peculiarly,queer,queer,queer
acayiplik	-0.179	-0.625	-1.0	strangeness,whimsicality,freakishness
accurse	-0.125	-0.125	-1.0	accurse
acele	-0.011	-0.136	-1.0	in_haste,haste,scurrying,hurry,hurry,rush,rush
aceleci	-0.521	-0.542	-1.0	impetuous,overhasty
aceleciliği	-0.125	-0.125	-1.0	impetuousness

acelecilik	-0.188	-0.750	-1.0	hurrying,pushiness,rashness,hastiness
aceleyle	0.125	0.125	1.0	hurriedly,hastily
acemi	-0.148	-0.341	-1.0	neophyte,unfledged
acemice	0.167	0.167	1.0	ineptly
açgözlü	-0.041	-0.333	-1.0	rapacious,overgreedy,rapaciously,all-devouring,covetously,covetous,grasping,greedy
açgözlülüğüyle	0.125	0.125	1.0	voraciously
açgözlülük	-0.127	-0.375	-1.0	voracity,avaritia,rapacity,covetousness,greed,greediness
açgözlülükle	0.125	0.125	1.0	greedily
acı	-0.358	-0.725	-1.0	painful,kibe,painfully,sting,sting,suffering,suffering,suffer,anguish,anguish,bitterly,bitter,bitter,bitter,brackish,gip
açı	-0.097	-0.125	-1.0	sorrower,angle
açık	0.186	0.875	1.0	incumbent_on,on_the_loose,open,on_the_offensive,overt,on_the_hook,on_the_go,on_hand,outdoor,receptive,obvious,on_the_nose,turned_on,unopen,exterior,luculent,on_the_button,ruttish,clear,clear,clear,clear,explicit,express
açıkça	0.228	0.500	1.0	plainly,manifestly,perspicuously,openly,patently,unmistakably,clearly,explicitly
açıkçası	0.438	0.500	1.0	frankly,obviously
açıkgöz	0.226	0.500	1.0	hardheaded,heady,canny
açıkgözlük	0.188	0.250	1.0	incisiveness,astuteness
açıkgozlükle	0.125	0.125	1.0	enterprisingly
açıkladı	0.125	0.125	1.0	explicate
açıklama	-0.023	-0.250	-1.0	description,explanation,professing,remark
açıklamak	0.136	0.136	1.0	explain
açıklanabilecek	0.625	0.625	1.0	explicable
açıklanabilir	0.625	0.625	1.0	explainable
açıklanamayan	-0.208	-0.500	-1.0	uncomprehended,unexplained
açıklanamaz	-0.375	-0.375	-1.0	inexplicable
açıklanan	0.250	0.250	1.0	disclosed
açıklanmamış	-0.375	-0.375	-1.0	unexpressed
açıklanmayan	-0.250	-0.250	-1.0	undisclosed
açıklayıcı	-0.083	-0.250	-1.0	paraphrastic,illustrative
acıklı	-0.288	-0.614	-1.0	pathetic,tearjerker,distressfully
açıklık	-0.001	0.250	-1.0	perspicuousness,openness,patency
açıklıklı	0.153	0.153	1.0	span
açık-ve-kapalı	0.500	0.500	1.0	open-and-shut
acil	-0.178	-0.273	-1.0	emergent,emergency
açılacağız	-0.125	-0.125	-1.0	wuss

acılaştırmak	-0.208	-0.208	-1.0	acerbate
açıldı	-0.136	-0.136	-1.0	opened
acilen	0.250	0.250	1.0	importunately
acılığın	-0.500	-0.500	-1.0	bitter_principle
acılık	-0.249	-0.352	-1.0	piquancy,acerbity,acridness,acridity,bitterness
acılır	-0.375	-0.375	-1.0	pull_away
açılır	-0.208	-0.250	-1.0	pop_fly,pop-up
açılış	0.014	0.014	1.0	opening
acillik	0.083	0.083	1.0	instancy
açılmak	0.100	0.100	1.0	unfold
açılmamış	-0.500	-0.500	-1.0	unopened
acımasız	-0.310	-0.750	-1.0	pitiless,ruthless,relentless,slashing,brutal,merciless,bowelless
acımasızca	0.088	-0.500	1.0	remorselessly,ruthlessly,pitilessly,unfeelingly,atrociously,cruelly,despitefully,inexorably,mercilessly
acımasızlaştırılmasına	0.068	0.068	1.0	brutalization
acımasızlık	-0.472	-0.625	-1.0	pitilessness,truculency,relentlessness,ruthlessness,inexorableness,mercilessness
acımış	-0.542	-0.542	-1.0	rancid
acınacak	-0.412	-0.750	-1.0	pitiable,pathetically,piteous,pitiful
acınası	0.500	0.500	1.0	pitiably
açısal	0.125	0.125	1.0	angular
acısız	-0.375	-0.375	-1.0	painlessly
acıyan	-0.750	-0.750	-1.0	achy
acıyı	-0.625	-0.625	-1.0	poignance
aciz	-0.400	-0.500	-1.0	incapacitated,incapable
acizane	0.125	0.125	1.0	unworthily
acizlik	-0.292	-0.417	-1.0	impotency,incapableness
açkılı	0.250	0.250	1.0	burnished
açkısızdır	-0.500	-0.500	-1.0	unburnished
açlık	0.004	-0.170	1.0	hunger,hunger,starvation,famishment,esurient
açlıktan	-0.283	-0.375	-1.0	starve,starving,famished
açma	-0.188	-0.250	-1.0	swingy,tripping
açmak	-0.085	-0.625	-1.0	unlace,turn_up,turn-on,unlaced,denudate,turn_on

Ek 3. Türkçe Duygu Sözlüğü Sürüm 1 Örnek Terimler

Terim	Ortalama	Maksimum	Yön	İngilizce terimler
abanoz	-0.068	-0.068	-1.0	ebony
abartı	0.108	0.125	1.0	overstatement,exaggeration
abartılı	0.042	0.125	1.0	overdone,exaggerated,exaggeratedly
abartılması	0.375	0.375	1.0	overestimation
abartılmış	0.125	0.125	1.0	overstated
abartırız	0.625	0.625	1.0	hyperbolise
abartısız	0.750	0.750	1.0	understated
abartma	0.313	0.375	1.0	overestimate,overcharge
abartmak	0.556	0.625	1.0	hyperbolize,overstate,exaggerate
abartmalı	-0.542	-0.542	-1.0	flatulent
abazanlık	0.250	0.250	1.0	horniness
abdias	-0.167	-0.167	-1.0	abdias
abes	0.250	0.250	1.0	nugatory
abetalipoproteinemi	-0.750	-0.750	-1.0	abetalipoproteinemia
abidance	0.341	0.341	1.0	abidance
abiyotrofi	-0.250	-0.250	-1.0	abiotrophy
abla	0.375	0.375	1.0	big_sister
ablasyon	-0.125	-0.125	-1.0	ablated
ablatif	0.208	0.208	1.0	ablative
ablefaron	0.500	0.500	1.0	ablepharia
abluka	-0.090	-0.090	-1.0	blockade
abone	0.068	0.068	1.0	subscriber
abreaksiyon	0.375	0.375	1.0	abreaction
abselenme	-0.083	-0.083	-1.0	suppuration
absorbanlar	0.500	0.500	1.0	absorbate
absorpsiyon	-0.077	-0.077	-1.0	absorption
aç	-0.125	-0.125	-1.0	hungry
acardia	-0.500	-0.500	-1.0	acardia
acayip	-0.102	-0.375	-1.0	kooky,peculiarly,queer,queer,queer,weird,kinky
acayiplik	-0.179	-0.625	-1.0	strangeness,whimsicality,freakishness
accurse	-0.125	-0.125	-1.0	accurse
acele	-0.027	-0.136	-1.0	haste,scurrying,rush,rush
aceleci	-0.139	0.625	-1.0	impetuous,pushy,overhasty
aceleciliği	-0.125	-0.125	-1.0	impetuousness
acelecilik	-0.188	-0.750	-1.0	hurrying,pushiness,rashness,hastiness

aceleyle	0.125	0.125	1.0	hurriedly,hastily
acemi	-0.148	-0.341	-1.0	neophyte,unfledged
acemice	0.167	0.167	1.0	ineptly
acemilik	0.125	0.125	1.0	inexperience
açgözlü	-0.013	-0.250	-1.0	rapacious,overgreedy,rapaciously,all-devouring,covetously,greedy
açgözlülüğüyle	0.125	0.125	1.0	voraciously
açgözlülük	-0.170	-0.375	-1.0	avaritia,rapacity,covetousness,greediness
açgözlülükle	0.125	0.125	1.0	greedily
acı	-0.361	-0.661	-1.0	kibe,painfully,sting,sting,suffering,suffering,anguish,anguish,bitter,bitter,bitter,brackish,gip
açı	-0.097	-0.125	-1.0	sorrower,angle
açık	0.184	0.875	1.0	incumbent_on,on_the_loose,open,on_the_offensive,overt,on_the_hook,on_the_go,on,on_hand,receptive,obvious,on_the_nose,turned_on,unopen,exterior,luculent,on_the_button,ruttish,clear,clear,clear,clear,explicit,express
açıkça	0.231	0.500	1.0	plainly,manifestly,perspicuously,openly,patently,unmistakably,bluntly,clearly,explicitly
açıkçası	0.438	0.500	1.0	frankly,obviously
açıkgöz	0.089	0.261	1.0	hardheaded,heady
açıkgözlük	0.188	0.250	1.0	incisiveness,astuteness
açıkgozlükle	0.125	0.125	1.0	enterprisingly
açıkladı	0.125	0.125	1.0	explicate
açıklama	-0.017	-0.250	-1.0	description,explanation,professing,remark,statement
açıklamak	0.136	0.136	1.0	explain
açıklanabilir	0.625	0.625	1.0	explicable,explainable
açıklanamayan	-0.208	-0.500	-1.0	uncomprehended,unexplained
açıklanamaz	-0.375	-0.375	-1.0	inexplicable
açıklanan	0.250	0.250	1.0	disclosed
açıklanmamış	-0.375	-0.375	-1.0	unexpressed
açıklayıcı	-0.014	-0.250	-1.0	paraphrastic,elucidative,illustrative
acıklı	-0.236	-0.458	-1.0	tragic,tearjerker,distressfully
açıklık	0.093	0.375	1.0	perspicuousness,openness,patency,explicitness
açık-ve-kapalı	0.500	0.500	1.0	open-and-shut
acil	-0.178	-0.273	-1.0	emergent,emergency

açılacağız	-0.125	-0.125	-1.0	wuss
acılaştırmak	-0.208	-0.208	-1.0	acerbate
acilen	0.250	0.250	1.0	importunately
acılığın	-0.500	-0.500	-1.0	bitter_principle
acılık	-0.226	-0.352	-1.0	piquancy,acridness,acridity,bitterness
acılır	-0.375	-0.375	-1.0	pull_away
açılır	-0.208	-0.250	-1.0	pop_fly,pop-up
açılış	0.014	0.014	1.0	opening
acillik	0.083	0.083	1.0	instancy
açılmak	0.100	0.100	1.0	unfold
açılmamış	-0.500	-0.500	-1.0	unopened
açılmış	-0.136	-0.136	-1.0	opened
acıma	-0.417	-0.417	-1.0	commiseration
acımak	-0.125	-0.125	-1.0	deplore
acımasız	-0.237	-0.667	-1.0	pitiless,relentless,slashing,brutal,merciless,bowelless
acımasızca	0.088	-0.500	1.0	remorselessly,ruthlessly,pitilessly,unfeelingly,atrociously,cruelly,despitefully,inexorably,mercilessly
acımasızlaştırılmasına	0.068	0.068	1.0	brutalization
acımasızlık	-0.472	-0.625	-1.0	pitilessness,truculency,relentlessness,ruthlessness,inexorableness,mercilessness
acımış	-0.542	-0.542	-1.0	rancid
acınacak	-0.412	-0.750	-1.0	pitiable,pathetically,piteous,pitiful
acınası	0.500	0.500	1.0	pitiably
açısal	0.125	0.125	1.0	angular
acısız	-0.375	-0.375	-1.0	painlessly
acıyan	-0.750	-0.750	-1.0	achy
acıyı	-0.625	-0.625	-1.0	poignance
aciz	-0.483	-0.648	-1.0	incapacitated,helpless,incapable
acizane	0.125	0.125	1.0	unworthily
acizlik	-0.319	-0.417	-1.0	helplessness,impotency,incapableness
açkılı	0.250	0.250	1.0	burnished
açkısızdır	-0.500	-0.500	-1.0	unburnished
açlık	0.004	-0.170	1.0	hunger,hunger,starvation,famishment,esurient
açlıktan	-0.375	-0.375	-1.0	starving,famished
açma	-0.188	-0.250	-1.0	swingy,tripping

açmak	-0.085	-0.625	-1.0	unlace,turn_up,turn-on,unlaced,denudate,turn_on
açmaz	-0.333	-0.333	-1.0	toughie
actinomyces	-0.125	-0.125	-1.0	genus_actinomyces

Ek 4. Makine Öğrenmesi, Otel Yorumları Veri Kümesi Öznitelik Örnekleri

	Kimlik	Kelime	x^2		Kimlik	Kelime	x^2
1	11750	yok	1.376.992.479	51	8898	sabah	259.610.505
2	1664	bu	1.201.599.370	52	10903	tur	252.463.024
3	8091	otel	948.296.758	53	7057	memnun	249.797.174
4	1289	berbat	892.580.103	54	260	akşam	248.759.417
5	1404	bile	855.475.294	55	3454	fakat	247.243.520
6	3931	gibi	829.993.532	56	11422	ya	246.256.226
7	5354	kadar	776.401.205	57	8458	pislik	246.204.436
8	11305	ve	742.406.573	58	903	az	242.610.739
9	6316	kötü	726.839.532	59	4508	her	240.720.586
10	2743	diye	723.769.533	60	9505	şey	239.692.652
11	412	ama	713.702.240	61	729	asla	236.656.764
12	7660	ne	698.505.560	62	7968	önce	235.864.813
13	4145	gün	652.466.540	63	8780	resmen	229.075.200
14	4537	hiç	645.688.005	64	9557	sıfır	216.226.821
15	11264	var	645.556.906	65	11623	yazık	214.834.487
16	1451	bir	637.293.031	66	5085	ise	213.421.695
17	11658	yemek	588.714.798	67	478	ancak	205.215.019
18	2354	daha	570.929.700	68	7639	nasıl	200.901.824
19	2451	de	534.063.329	69	6007	kişi	200.653.896
20	9873	sonra	528.814.030	70	11674	yer	197.640.348
21	9999	su	519.225.852	71	4810	iğrenç	197.504.011
22	2477	değil	513.067.877	72	4848	iki	193.186.471
23	2340	da	504.990.658	73	2321	çünkü	191.097.734
24	4763	için	466.064.248	74	2358	dahil	190.621.147
25	1264	ben	451.813.179	75	4504	hep	190.620.572
26	8445	pis	448.850.235	76	9589	şikayet	188.052.143
27	7839	oda	443.865.817	77	4429	hayal	185.567.098
28	8896	saat	401.928.545	78	9861	son	182.504.502
29	8803	rezalet	398.636.447	79	8139	öyle	176.650.937
30	8233	para	395.803.179	80	5835	kendi	174.710.942
31	8930	sadece	377.673.623	81	8807	rezervasyon	168.971.389
32	5904	ki	367.630.218	82	10379	tavuk	166.440.899
33	878	aynı	358.425.390	83	3899	geri	165.697.904
34	10433	tek	355.215.713	84	3958	giriş	162.211.228
35	10361	tatil	353.356.657	85	4494	hemen	156.363.025
36	3190	en	352.292.815	86	5943	kimse	151.676.738
37	1126	başka	346.836.009	87	7194	mi	144.291.191
38	1630	böyle	331.516.952	88	9124	sanki	143.281.539

39	1518	biz	331.322.577	89	80	adam	140.442.576
40	4915	ilk	320.129.870	90	7944	olur	140.002.662
41	10256	tam	315.543.026	91	3509	fazla	138.682.455
42	3822	gece	310.353.624	92	2825	dolu	137.923.599
43	11903	zaten	304.189.164	93	2042	çeşit	137.746.163
44	8809	rezil	303.047.895	94	885	ayrıca	134.709.833
45	4881	ile	299.572.006	95	8459	pişman	132.475.852
46	9022	sakın	296.411.743	96	11353	veya	130.330.619
47	4424	havuz	280.368.080	97	1066	banyo	129.624.263
48	11508	yani	277.656.465	98	1646	bozuk	126.708.949
49	4764	içinde	270.044.969	99	12008	zor	126.289.018
50	9765	siz	267.446.597	100	4160	günlük	123.462.468

ÖZGEÇMİŞ

Kimlik Bilgileri

Adı Soyadı	:	Alaettin UÇAN
Doğum Yeri	:	Kahramanmaraş
Medeni Hali	:	Evli
E-posta	:	aucan@hacettepe.edu.tr
Adresi	:	Hacettepe Üniversitesi Bilgisayar Mühendisliği Bölümü Beytepe/Çankaya/ANKARA

Eğitim

Lise	:	Afşin Anadolu Lisesi
Lisans	:	Kırgızistan Türkiye Manas Üniversitesi
Yüksek Lisans	:	Hacettepe Üniversitesi

Yabancı Dil ve Düzeyi

İngilizce:	İyi
Rusça:	Orta

İş Deneyimi

Hacettepe Üniversitesi
Bilgisayar Mühendisliği Bölümü
Araştırma Görevlisi (ÖYP) (2012 – Halen)

Osmaniye Korkut Ata Üniversitesi
Yönetim Bilişim Sistemleri Bölümü
Araştırma Görevlisi (33. Madde) (2010 – Halen)

Kardelen Yazılım Ltd.
Yazılım Uzmanı (2008 – 2010)

Deneyim Alanları

Tezden Üretilmiş Projeler ve Bütçesi

Tezden Üretilmiş Yayınlar

- F. Akba, A. Uçan, E. Akçapınar Sezer and H. Sever, "Assessment of feature selection metrics for sentiment analyses: Turkish movie reviews," in 8th European Conference on Data Mining 2014, Lizbon, 2014.

Tezden Üretilmiş Tebliği ve/veya Poster Sunumu ile Katıldığı Toplantılar

- The European Conference on Data Mining (ECDM'14)