

**T.C.
HARRAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

YÜKSEK LİSANS TEZİ

DUYGU ANALİZİ İÇİN YENİ BİR SÖZLÜK; NAYALex DUYGU SÖZLÜĞÜ

Yakup ATLI

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

**ŞANLIURFA
2022**

İÇİNDEKİLER

	Sayfa No
ÖZET	i
ABSTRACT	ii
TEŞEKKÜR	iii
ŞEKİLLER DİZİNİ	iv
ÇİZELGELER DİZİNİ	v
1. GİRİŞ	1
1.1. Sosyal Medya	1
1.2. Duygu Analizi	2
1.3. Duygu Teorisi	3
1.4. Duygu Sözlükleri	5
2. ÖNCEKİ ÇALIŞMALAR	7
3. MATERYAL ve YÖNTEM	12
3.1. Materyal	12
3.1.1. Test verisi	12
3.1.1.1. Büyük-küçük harf dönüşümü	13
3.1.1.2. Noktalama işaretlerinin temizlenmesi	14
3.1.1.3. Sayısal mesajların temizlenmesi	14
3.1.1.4. Etkisiz kelimelerin temizlenmesi	15
3.1.2. Doğrulama verisi	15
3.2. Yöntem	16
3.2.1. NAYALex duygu sözlüğü	16
3.2.2. NAYALex duygu sözlüğü oluşturma aşamaları	17
3.2.2.1. Ön işlem aşaması	18
3.2.2.2. Dönüşüm aşaması	19
3.2.2.3. Sınıflandırma-sözlük oluşturma aşaması	20
3.2.3. NAYALex kütüphanesi	22
3.2.4. Uygulanan sınıflandırma algoritmaları	24
3.2.4.1. Naive bayes	24
3.2.4.2. Destek vektör makinesi(SVM)	25
3.2.4.3. KNN(K-nearest neighbors)	25
3.2.5. Sınıflandırma algoritmaları performans değerlendirme ölçütleri	25
3.2.5.1. Confusion matrix(Karışıklık matrisi)	25
3.2.5.2. Doğruluk(Accuracy)	26
3.2.5.3. Duyarlılık(Recall)	26
3.2.5.4. Kesinlik(precision)	27
3.2.5.5. F-Ölçütü	27
4. ARAŞTIRMA BULGULARI ve TARTIŞMA	28
4.1. NAYALex Duygu Sözlüğü İle Sınıflandırma Ve Analiz	28
4.2. NAYALex Duygu Sözlüğünün Diğer Duygu Sözlükleri İle Karşılaştırılması	32
4.3. NAYALex Duygu Sözlüğünün Doğrulanması	34
4.4. Sınıflandırma Algoritmaları İle Değerlendirme	44
5. SONUÇLAR ve ÖNERİLER	48
KAYNAKLAR	52

ÖZET

Yüksek Lisans Tezi

DUYGU ANALİZİ İÇİN YENİ BİR SÖZLÜK; NAYALex DUYGU SÖZLÜĞÜ

Yakup ATLI

Harran Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Dr. Öğr. Üyesi Nagehan İLHAN
YIL: 2022, Sayfa: 56

İletişimde karşımıza farklı şekillerde (yüz ifadeleri, jestler, konuşma vb.) ortaya çıkan duygularda birçok bilgi gizlidir. Kişiler duygu ve düşüncelerini sosyal medya platformlarında çoğu zaman metinsel ifadeler ile ifade etmektedirler. Sosyal medya paylaşımları incelenerek insanların duygu durumları elde edilebilmektedir. Paylaşımında bulunan kişinin kişilik özelliklerinin ortaya çıkarılması önem kazanmıştır ve buna kaynaklık eden metinlerde gizlenen duyguların sıklığının tespiti ile ilgili birçok çalışma yapılmıştır. Kişilerin paylaşılan mesajlarında gizlenen duyguların ortaya çıkarılması ve tespiti önemlidir. Metinlerde gizlenen duyguların ortaya çıkarılması için kelime-duygu sözlükleri sıklıkla kullanılmaktadır. Kelime-duygu sözlükleri tespit edebildikleri duygu sayısı dikkate alınarak incelendiğinde, olumlu-olumsuz duygu ve 8 farklı duyguyla birlikte en fazla sayıda duygu çıkarımı yapabilen NRC Duygu Sözlüğüdür. Ancak, duygularını metinler aracılığı ile yansıtan sosyal medya kullanıcılarının duygularını kısıtlı sayıdaki duygu ile sınırlandırmak kişilik özelliklerinin ortaya çıkarılmasında yetersiz kalmaktadır. Kişilik özelliklerini ifade eden duyguların, kişiliğin tespitinde daha fazla temsil edilmesini sağlamak için metinlerden daha çok duyguyu ortaya çıkaran bir sözlüğe ihtiyaç olduğu açık bir şekilde görülmektedir. Bu çalışmada, literatürde bu alanda görülen eksikliğin giderilmesine katkı sunacak olan olumlu-olumsuz ile birlikte 38 farklı duyguyu(korku, kaygı, üzüntü, karamsarlık, iyimserlik, öfke, sevgi, umut vb.) ortaya çıkarabilen NAYALex Duygu Sözlüğü önerilmektedir. Plutchik'in Duygu Teorisine ve NRC Duygu Sözlüğüne dayandırılarak oluşturduğumuz NAYALex Duygu Sözlüğü, her bir kelimesi 38 farklı duygunun en az biri ile ilişkili 6469 İngilizce kelimedenden oluşan 245.822 duygu-kelime ilişkisi içermektedir. Bu çalışmamızda iki farklı veri üzerinde çalışmalar yapılmıştır. Birinci veri seti, Instagram kullanıcılarının paylaşımlarını içeren 11.880 farklı paylaşım verisini içermektedir. İkinci veri seti, doğrulama için kullanılan ve önerilen sözlükte yer alan 38 farklı duygunun her birini ayrı ayrı ifade edebilen İngilizce cümlelerden oluşmaktadır. Veri setindeki 380 farklı cümle, insan değerlendiricilere okutularak, okuyucuya göre(38 farklı duygu için) ifade ettiği duyguların belirtilmesi istenmiştir. İnsan değerlendiricilerden elde edilen doğrulama veri seti 26.600 cümle-duygu ilişkisi içermekte ve elde edilen veriler doğrultusunda önerilen NAYALex Duygu Sözlüğü'nün doğrulanması yapılmıştır. Instagram veri seti üzerinde yapılan deneysel çalışma ve analizlerle NAYALex sözlüğü'nün kişilik analizinde uygulanabilir ve kullanılabilir bir sözlük olduğu ispatlanmıştır. Doğrulama verisi kullanılarak insan değerlendiriciler ile yapılan çalışma sonucunda NAYALex Duygu Sözlüğü için %77 ortalama doğrulama oranı elde edilerek sözlüğün doğrulması sağlanmıştır. Ayrıca NAYALex doğrulaması için kullanılan doğrulama veri setine Naive Bayes, Destek Vektör Makineleri ve KNN(K-Nearest Neighbors) öğrenme algoritmaları uygulanarak performanslar incelenmiştir. Doğrulama veri seti üzerinde uygulanan sınıflandırma algoritmaları arasında %93 doğruluk oranı ile Destek Vektör Makineleri en yüksek değere sahip sınıflandırma algoritması olmuştur. NAYALex Duygu Sözlüğü; bugüne kadar yapılmış (LIWC, EmoSenticNet, NRC, Empath) duygu sözlükleriyle karşılaştırıldığında, Tiffany'nin belirttiği 154 duygu için kendisine en yakın rakibi olan %6.5 kapsam oranına sahip NRC(Emolex)'e göre yaklaşık dört kat daha fazla duygu çıkarımı yaparak %24.7 oranı ile bildiğimiz kadarıyla en kapsamlı duygu sözlüğü olmuştur.

ANAHTAR KELİMELELER: NAYALex, NRC, Duygu Analizi, Duygu Sözlüğü, Plutchik

ABSTRACT

MSc Thesis

A NEW DICTIONARY FOR SENTIMENT ANALYSIS; NAYALex EMOTION DICTIONARY

Yakup ATLI

Harran University
Graduate School of Natural and Applied Sciences
Department of Computer Engineering

Supervisor: Assist. Prof. Dr. Nagehan İLHAN
Year: 2022, Page: 56

A lot of information is hidden in the emotions that appear in different ways (facial expressions, gestures, speech, etc.) in communication. People often express their feelings and thoughts on social media platforms with textual expressions. By examining social media shares, people's emotional states can be obtained. It has gained importance to reveal the personality traits of the person sharing, and many studies have been carried out to determine the frequency of emotions hidden in the texts that are the source of this. It is important to reveal and detect the emotions hidden in the shared messages of the people. Word-emotion dictionaries are frequently used to reveal the emotions hidden in the texts. When the word-emotion dictionaries are examined considering the number of emotions they can detect, it is the NRC Sentiment Dictionary that can extract the highest number of emotions with positive-negative emotions and 8 different emotions. However, limiting the emotions of social media users who reflect their emotions through texts to a limited number of emotions is insufficient to reveal personality traits. It is clearly seen that there is a need for a dictionary that reveals more emotions than texts in order to ensure that emotions expressing personality traits are more represented in the determination of personality. In this study, the NAYALex Emotion Dictionary, which can reveal 38 different emotions (fear, anxiety, sadness, pessimism, optimism, anger, love, hope, etc.) together with positive-negative, which will contribute to the elimination of the deficiency seen in this field in the literature, is recommended. Based on Plutchik's Theory of Emotion and the NRC Emotion Dictionary, the NAYALex Emotion Dictionary contains 245,822 emotion-word relationships consisting of 6469 English words, each word associated with at least one of 38 different emotions. In this study, studies were carried out on two different data. The first data set contains 11,880 different sharing data, which includes the shares of Instagram users. The second data set consists of English sentences that can express each of the 38 different emotions used for verification and included in the proposed dictionary. 380 different sentences in the data set were read to human evaluators and they were asked to indicate the emotions they expressed according to the reader (for 38 different emotions). The validation data set obtained from human evaluators includes 26,600 sentence-emotion relationships, and the proposed NAYALex Emotion Dictionary was validated in line with the data obtained. With the experimental studies and analyzes on the Instagram dataset, it has been proven that the NAYALex dictionary is a viable and usable dictionary in personality analysis. As a result of the study with human evaluators using validation data, an average validation rate of 77% was obtained for the NAYALex Emotion Dictionary, and the dictionary was verified. In addition, Naive Bayes, Support Vector Machines and KNN (K-Nearest Neighbors) learning algorithms were applied to the validation data set used for NAYALex validation and the performances were examined. Among the classification algorithms applied on the validation dataset, Support Vector Machines has been the classification algorithm with the highest value with 93% accuracy. NAYALex Emotion Dictionary; Compared to emotion dictionaries (LIWC, EmoSentNet, NRC, Empath) made to date, Tiffany extracted 24.7% of emotions for 154 emotions, approximately four times more than its closest competitor, NRC (Emolex), which has 6.5% coverage. As far as we know, it has been the most comprehensive dictionary of emotions.

KEYWORDS: NAYALex, NRC, Emotion Analysis, Emotion Lexicon

TEŐEKKÜR

Lisansüstü eđitimim süresince bana birçok konuda bilgi ve birikimleri ile destek olan deđerli hocam Dr. Öğr. Üyesi Nagehan İLHAN'a ve ayrıca desteđini hiçbir zaman esirgemeyen eşim Yıldız'a teşekkür ederim



ŞEKİLLER DİZİNİ

Sayfa No

Şekil 1.1. 2022’de dünyada dijital ortam kullanımı	1
Şekil 1.2. Plutchik’in birincil, ikincil ve üçüncül ikililer duygu çiftleri.....	4
Şekil 2.1. Duygu analizinde kullanılan algoritmalar ve sınıflandırma yöntemleri	9
Şekil 3.1. Duygu sözlüğü oluşturma mimarisi	18
Şekil 3.2. NAYALex duygu sözlüğü python kütüphanesi	23
Şekil 3.3. NAYALex duygu sözlüğü python kütüphanesi kelime-duygu ilişkileri	24
Şekil 4.1. Kullanıcı bazlı duygu yoğunlukları.....	31
Şekil 4.2. 11 nolu instagram kullanıcıasına ait duygu yoğunlukları	31
Şekil 4.3. “5” nolu kullanıcıya ait duygu bulutu	32
Şekil 4.4. NAYALex ve Nrc(Emolex) ortak duygular için sınıflandırma oranları	34
Şekil 4.5. NAYALex 1. duygu grubu doğrulama başarı oranları.....	38
Şekil 4.6. NAYALex 2. duygu grubu doğrulama başarı oranları.....	39
Şekil 4.7. NAYALex 3. duygu grubu doğrulama başarı oranları.....	39
Şekil 4.8. NAYALex 4. duygu grubu doğrulama başarı oranları.....	40
Şekil 4.9. NAYALex 5. duygu grubu doğrulama başarı oranları.....	41
Şekil 4.10. NAYALex 6. duygu grubu doğrulama başarı oranları.....	41
Şekil 4.11. NAYALex 7. duygu grubu doğrulama başarı oranları.....	42
Şekil 4.12. NAYALex 8. duygu grubu doğrulama başarı oranları.....	43
Şekil 4.13. Tüm duygular için nayalex başarı oranları.....	43
Şekil 5.1. NAYALex doğrulama başarı oranları	49

ÇİZELGELER DİZİNİ

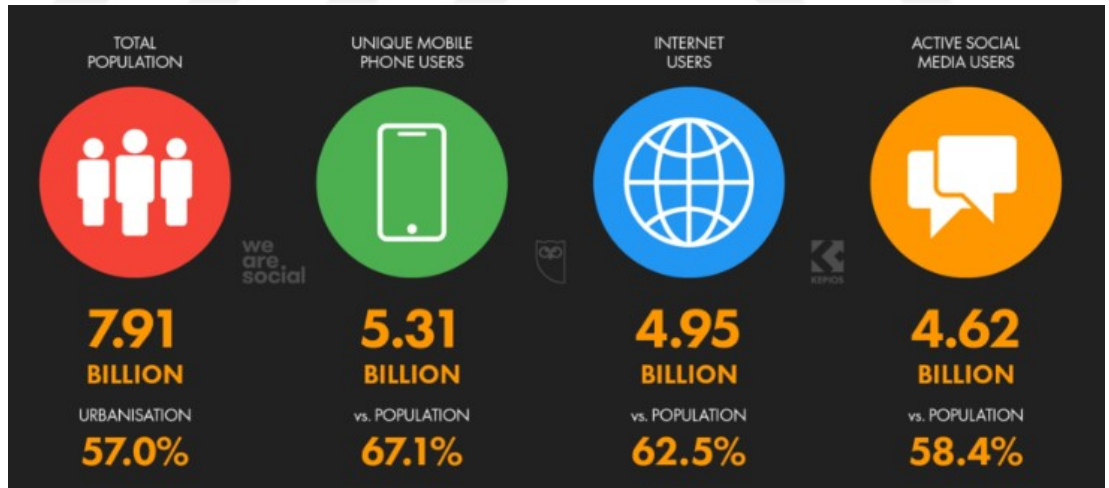
Sayfa No

Çizelge 3.1. Hesabı gizli olmayan Instagram kullanıcılarına ait işlenmemiş medya verileri	13
Çizelge 3.2. Büyük-küçük harf dönüşümü yapılmış medya verileri	13
Çizelge 3.3. Noktalama işaretleri temizlenmiş medya verileri	14
Çizelge 3.4. Sayısal mesaj içerikleri temizlenmiş medya verileri	14
Çizelge 3.5. Etkisiz kelimeleri temizlenmiş medya verileri	15
Çizelge 3.6. Duygu yüklü doğrulama veri seti	16
Çizelge 3.7. NRC duygu sözlüğü	18
Çizelge 3.8. Plutchik'in duygu çiftleri	19
Çizelge 3.9. Duygu dönüşümü ve yeniden sınıflandırma sembolik gösterimi	21
Çizelge 3.10. NRC ve NAYALex kelime-duygu eşleştirmesi	22
Çizelge 3.11. Karışıklık matrisi	26
Çizelge 4.1. Test verisi için kullanılacak Instagram kullanıcılarına ait işlenmemiş medya verisi	28
Çizelge 4.2. Instagram kullanıcılarına ait sınıflandırmaya uygun test veri seti	29
Çizelge 4.3. NAYALex duygu sözlüğü ile sınıflandırılmış veriler	29
Çizelge 4.4. Kullanıcı bazlı ortalama duygu değerleri	30
Çizelge 4.5. NAYALex'in diğer duygu sözlükleri ile duygu-kapsam karşılaştırması	33
Çizelge 4.6. Ortalama duygu yoğunluk oranları(0.0 ile 1.0 aralığında)	35
Çizelge 4.7. NAYALex ortalama duygu yoğunluk oranları(0.0 -1.0 aralığında)	37
Çizelge 4.8. Navi bayes sınıflandırma algoritması sonuçları	45
Çizelge 4.9. Destek vektör makinesi(SVM) sınıflandırma algoritması sonuçları	46
Çizelge 4.10. KNN(K-Nearest Neighbors) sınıflandırma algoritması sonuçları	47
Çizelge 5.1 Algoritmaları için performans değerlendirme ölçütlerinin karşılaştırılması	49
Çizelge 5.2. NAYALex ve duygu sözlüklerinin sınıflandırma oranlarının karşılaştırılması	50

1. GİRİŞ

1.1. Sosyal Medya

İnternet kullanımının artmasıyla birlikte, insanlar ve kuruluşlar arasındaki sanal işbirliği ve etkileşim büyük artış göstermektedir. Sosyal ağların temeli buradan oluşmaktadır. Bireyler arasındaki bu sanal etkileşim arttıkça, duygu analizi için bir temel teşkil eden verilerin elde edilebilirliği de artmaktadır. Şekil 1.1. de gösterildiği gibi, 2022 yılı başı itibari ile internet kullanıcılarına baktığımızda 4,95 milyara yükselmiş. Dünya nüfusunun yüzde 62,5'inin şuan internete bağlı olduğu ve milyarlarca kullanıcının çevrimiçi olduğunu bilinmektedir. Dünya çapında Ocak 2022 itibari ile 4,62 milyar sosyal medya kullanıcısı var. Ortalama bir internet kullanıcısı her gün 6,5 saatten fazla çevrimiçi olarak zaman harcıyor. Sosyal medya, insanların zamanını harcadığı alanlar içinde en büyük paya sahip olmaya devam ediyor(Datareportal, 2022).



Şekil 1.1. 2022'de dünyada dijital ortam kullanımı (Datareportal, 2022)

Dijital kullanımın bu derece artması ve buna paralel olarak sosyal medya kullanımının artması, sosyal medyayı araştırma noktası haline getirmiştir. “Sosyal medya kullanıcıları Facebook, Youtube, Instagram, Twitter, Google+, gibi birçok sosyal medya platformu aracılığı ile kişilik özelliklerini barındıran duyguları içeren

paylaşımlar yapmaktadırlar. Bu durum sosyal medya platformlarını kullanan kişileri yakından takip eden işletme ve araştırma kuruluşlarının bu alana yönelmesine sebep olmuştur” (Atlı ve İlhan, 2021). İşletme ve kuruluşlar sosyal medya kullanıcıların paylaşımlarında gizli olan kişiliklerine ait duyguları dikkate alarak farklı strateji ve uygulamalar geliştirmektedirler. Sosyal medyada insanlar duygularını, fikirlerini ve zevklerini barındıran birçok paylaşım yapmaktadır. Sosyal medya, bir araç olarak kullanıcıların duygularını yansıtan paylaşımlarla büyük bir veri kaynağı haline gelmiştir (Hidalgo ve ark.,2015; Atlı ve İlhan, 2021).

1.2. Duygu Analizi

Kişilik, insan yaşamı üzerinde önemi çok büyüktür ve çoğu zaman geleneksel yöntem olan anketler ile belirlenmektedir. Kişiliğin ortaya çıkarılması için geleneksel yöntem olan anketler yerine, fikirlerin ve duyguların serbestçe ifade edildiği metinlerden çıkarılması daha etkili ve daha çok tercih edilen bir durum olmaktadır. Kişilik, davranış ve duygularla çevresel ve biyolojik faktörlerden etkilenerek ortaya çıkmaktadır. Bazı sözcük gruplarının bazı duygularla ilişkili olduğu ve bazı duygularında kişilikle ilişkili olduğu gösterilmiştir (Mohammad and Kiritchenko, 2013).

Doğal dil işlemedeki belirsizlikler ve karmaşıklıklar, sosyal medya paylaşımlarından kişiliğin belirleyicisi olan duyguların ortaya çıkarılması çalışmalarını zorlaştırmaktadır (Hussein, 2018). “Kişilerin sosyal medya paylaşımlarındaki duygularını ortaya çıkarmaya odaklanan birçok çalışma yapılmıştır. Literatürde bir duygunun sadece olumlu-olumsuz olmasına dayandırılarak yapılan çalışmalar (örneğin, Alarid, 2016; Kuşen ve ark., 2017; Woolf, 2016;) daha çok iken, kişiliğe ait bireysel duyguları (üzüntü, öfke, neşe) ortaya çıkaran çalışmalar daha az olmuştur” (Atlı ve İlhan, 2021).

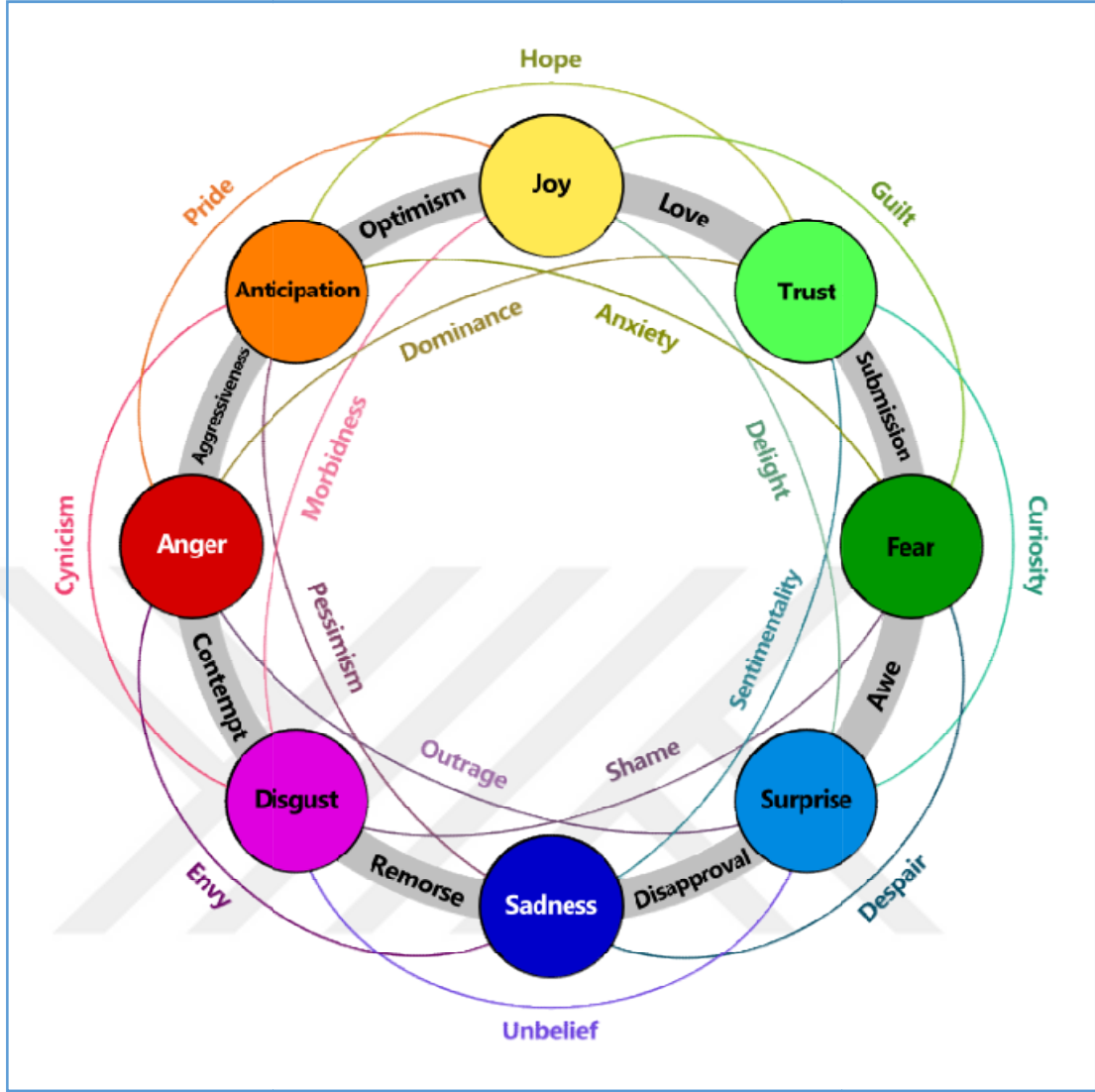
“Kişilik ile ilişkili olan duygular, kişilik tespitinde önemli yere sahiptirler. Kişilik tespiti için yapılan birçok çalışmada, duygular sadece olumlu- olumsuz olacak şekilde ele alınmakta ve bu şekilde sınırlı sayıdaki duygu ile yapılan

çalışmalar kişilik analizleri için yetersiz kalmaktadır” (Atlı ve İlhan, 2021). LIWC (Pennebaker ve ark., 2001), EmosenticNet (Poria ve ark., 2012), DepecheMood (Staiano and Guerini, 2014), Empath (Fast ve ark., 2016) ve NRC (Mohammad, 2016) gibi duygu sözlükleri kullanılan duygu sözlükleridir.

Kişiliğin ortaya çıkarılması üzerinde etkisi olan duygular, sabit olmayan davranışlarla ortaya çıkmaktadır.” Duyguların kişilik belirlemedeki etkisi, duyguların doğru belirlenmesini zorunlu kılmaktadır” (Atlı ve İlhan, 2021). Sosyal medya kullanıcılarının metinsel paylaşımları ile birlikte geniş veri havuzları ortaya çıkmakta ve duygularını paylaştıkları bu metinlere yansıtılmaktadırlar. Metinlerde saklı olan bu duyguların ortaya çıkarılması duygu sözlükleri ile yapılabilmektedir. . LIWC (Pennebaker ve ark., 2001), EmosenticNet (Poria ve ark., 2012), DepecheMood (Staiano and Guerini, 2014), Empath (Fast ve ark., 2016) ve NRC (Mohammad, 2016) gibi duygu sözlükleri için karşımıza çıkan bir dezavantaj vardır. “Bu sözlüklerin dezavantajı, anahtar kelimelere atanan duygu ifadelerinin sınırlı sayıda duyguyu kapsamakta ve bu şekilde ortaya çıkarılabilen sınırlı sayıdaki duygu çıkarımı, kişiliğin belirlenmesi üzerinde olumsuz etki oluşturmaktadır” (Atlı ve İlhan, 2021).

1.3. Duygu Teorisi

“Ayrık duygu teorisine göre, insanların doğuştan gelen bir dizi temel duyguya sahip olduğu düşünülmektedir (Colombetti, 2009). 1980'de Robert Plutchik, On Postulates'ten esinlenerek sekiz duygudan (neşe, güven, korku, şaşkınlık, üzüntü, tiksinti, öfke ve beklenti) oluşan bir duygu çarkı çizdi (Plutchik, 1980; Plutchik, 2001). Plutchik ayrıca Şekil 1.2. de iki duygudan oluşan ve "Birincil", "İkincil" ve "Üçüncül" ikililerden ve zıt duygulardan oluşan yirmisekiz farklı duyguyu ele alan temel duyguların psikoevrimsel teorisini ortaya attı (Dreus ve Krohn, 2007). Tiffany Watt Smith dünya çapında 154 farklı duygu olduğunu belirtmiştir” (Medhat ve ark., 2014; Atlı ve İlhan, 2021).



Şekil 1.2. Plutchik'in birincil, ikincil ve üçüncül ikililer duygu çiftleri (Drews and Krohn, 2007)

“Tiffany'nin 154 farklı duygunun varlığını belirtmesi, duygu çıkarımı yapan sözlüklerin sınırlı sayıdaki duygu ile analiz yapmaları duygu analizi için sınırlı olabileceğini ve daha fazla duygu çıkarımı yapabilen sözlüklere ihtiyaç duyulduğunu ortaya koymaktadır” (Medhat ve ark., 2014; Atlı ve İlhan, 2021).

1.4. Duygu Sözlükleri

“Duygu çıkarımı ve analizi için kullanılan önceki sözlüklerden, LIWC dört (Positive, Negative, Sadness, Anger), EmosenticNet altı (Disgust, Joy, Fear, Surprise, Anger, Sadness), Empath altı (Joy, Fear, Surprise, Anger, Sadness, Love), NRC sekiz (Trust, Anticipation, Disgust, Joy, Fear, Surprise, Anger, Sadness) farklı duygu için sınıflandırma yapabilmektedir” (Atlı ve İlhan, 2021).

“Bu çalışmada, literatürde sıklıkla kullanılan (NRC (Mohammad, 2016), EmosenticNet (Poria ve ark., 2012), DepecheMood (Staiano and Guerini, 2014), LIWC (Pennebaker ve ark., 2001) ve Empath (Fast ve ark., 2016)) sözlüklerin ortaya çıkarabildiği duyguları da kapsayan ve olumlu-olumsuz ile birlikte 38 farklı duygu için sınıflandırma yapabilen NAYALex Duygu Sözlüğü önerilmektedir. Plutchik’in Temel Duyguların Psikoevrimsel Teorisine (Drews and Krohn, 2007) ve NRC Duygu Sözlüğüne dayandırılarak oluşturduğumuz NAYALex Duygu Sözlüğü 6469 kelime ve her kelimesi 38 farklı duygu ile ilişkilendirilmiştir” (Atlı ve İlhan, 2021).

“NAYALex Duygu Sözlüğü ile Tiffany’nin belirttiği 154 farklı duygudan 38 tanesi için duygu sınıflandırması yapılabilmektedir. Bu çalışmada, kişiliğin belirlenmesinde büyük öneme sahip olan duygu sınıflandırmasının literatürde kısıtlı sayıda duygu çıkarımı yapan sözlüklerden(4 duygu(LIWC), 6 duygu(EmosenticNet, Empath) ve 8 duygu(NRC)) kaynaklanan dezavantajın ortadan kaldırılması hedeflenmiş ve bunun için 38 farklı duygu için sınıflandırma yapabilen NAYALex Duygu Sözlüğü oluşturulmuştur” (Atlı ve İlhan, 2021).

“Bu çalışma sonucunda, sözlük tabanlı duygu sınıflandırması için doğrudan 38 farklı duyguyu ortaya çıkarabilen ilk en kapsamlı duygu sözlüğü oluşturulmuş, test edilmiş ve doğrulanması yapılmıştır” (Atlı ve İlhan, 2021).

“Çalışmanın geri kalanı aşağıdaki gibi yapılandırılmıştır. Bölüm II’de önceki çalışmalardan bahsedilmiştir. Bölüm III’te materyal ve yöntemler açıkça belirtilmiş ve sözlüğün oluşturulması ayrıntılı olarak açıklanmıştır” (Atlı ve İlhan, 2021).

Bölüm IV'te araştırma ve bulgular detaylı olarak ele alınarak sözlüğün test edilmesi ve doğrulanması aşamaları ele alınmıştır. Bölüm V'de sonuçlar ve gelecekteki öneriler üzerinde durularak sözlüğün oluşturulması, test edilmesi ve doğrulanması ile ortaya çıkan durumların değerlendirilmesine ve literatüre olan katkısına yer verilmiştir.



2. ÖNCEKİ ÇALIŞMALAR

“Bir kişinin üzerinde yorum yaptığı faktörler kişiler, konular, nesnelere veya varlıklar olabilir. Duygu analizi ile herhangi bir kişi, konu, nesne veya varlık üzerinde görüş belirten kişinin duygusal durumunu tespit etmek amaçlanır (Cambria ve ark., 2013). Duygu analizi ile tutumların, değerlendirmelerin, görüşlerin, hislerin ve duyguların metinsel ifadelerden çıkarılması doğal dil işlemenin en önemli konularından biridir. Duygu analizi, bireylerin metinler aracılığı ile yansıttığı duygularının belirli yordamlarla davranış, görüş, tutum ve duyguların istatistiksel yöntemlere göre gerçekleştirilen hesaplamalı analizlerdir” (Medhat ve ark., 2014; Pang and Lee, 2009; Atlı ve İlhan, 2021).

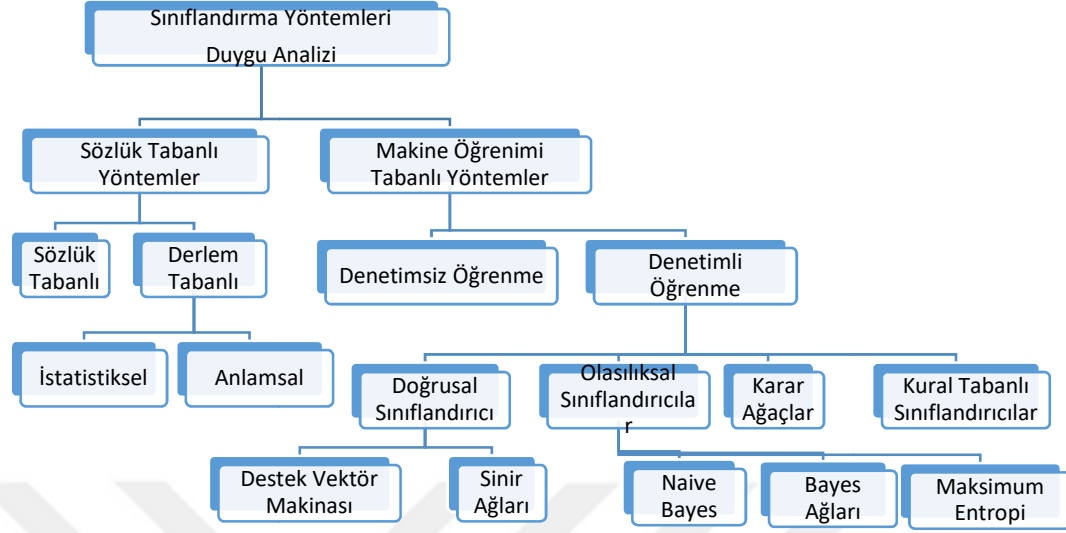
Konuşma, “etkileşimde etkili olan ve duyguların yansıtıldığı metinlerin yanında temel araçlardan biridir. Tanıma sorunlarına derin öğrenme modelleri kullanılarak çözümler üretilmektedir (Mittal ve ark., 2018; Bae ve ark., 2016; Malik ve ark., 2020). Konuşma duygularının ortaya çıkarılması ve sınıflandırılması için yapılan araştırmalar giderek ilgi görmeye başlamıştır. İnsanların akıllı sistemler ile etkileşimlerinin artması ve akıllı sistemlerin de veri işleme hızlarının ve performanslarının artmasıyla duyguların ortaya çıkarılmasında ilgiyi bu alana yöneltmiştir” (Atlı ve İlhan, 2021).

“Konuşma duygularının ortaya çıkarılması için uygulanan geleneksel yöntemlerde konuşmanın akustik içeriği ile ilgili özellikler çıkarılmaktadır. Öznitelikleri çıkarılmış konuşma duygularını sınıflandırmak için makine öğrenme teknikleri ve önceden belirlenmiş duygu etiketleri kullanılmaktadır. İnsan-bilgisayar etkileşimiyle konuşmadan duygu tanıma probleminde sınıflandırma doğruluğunu arttırmak için yeni bir hibrit mimari önermişlerdir (Er, 2020). Önerilen tekniğin konuşma duygularını doğru ve verimli bir şekilde sınıflandırabildiğini yapılan deneyler ile açık bir şekilde göstermişlerdir” (Atlı ve İlhan, 2021).

“Öznel bir olgu olan duygular, boyutsal ve kategorik olmak üzere iki grupta ele alınmaktadır. Kategorik modellerde tek kelime ya da kelime grupları kullanılarak duygular tanımlanırken, iki boyutlu modellerde uyarılma değerleri ve değerlilik ölçüleri ele alınmaktadır (Alarcao and Fonseca, 2017; Lang, 1995). Farklı önerilerde bulunan psikologlar ve bilim insanları duyguları öfke, korku, iğrenme, üzülmeye, mutluluk ve neşe olarak 6 temel sınıf olacak şekilde ayırmışlardır” (Picard, 2000: Atlı ve İlhan, 2021).

“Müziğin dinleyicileri üzerinde etkisi olduğu ve duyguları ortaya çıkarmada etkili olduğu kabul görmektedir. Dinleme esnasında beyinde oluşan elektriksel sinyallerin duyguların ortaya çıkarılmasında daha gerçekçi sonuçlar ortaya koyduğu ile ilgili birçok çalışma yapılmıştır. Müzik parçalarının dinlenmesi ile ortaya çıkan duyguların sınıflandırılması ve tanınması problemini ele almışlardır (Er and Çiğ, 2020). Bunun için Türk müziği parçalarını katılımcılara dinleterek ve bunun sonucunda katılımcıların beyinlerinde oluşan sinyalleri inceleyerek rahatlatıcı, gergin, hüznü ve mutlu duygularını tanımaya çalışmışlardır. Makine öğrenme algoritmaları kullanılarak sınıflandırmalar yapılmış ve elde edilen doğruluk oranları kullanılan yöntemlerinin performansını olumlu yönde etkilediği gözlemlenmiştir” (Atlı ve İlhan, 2021).

“Duygu sınıflandırması, duygu analizinin en önemli adımıdır ve son yıllarda en fazla ilgi gören çalışma alanlarından biri olmuştur (Lin ve ark., 2014). Duygu sınıflandırması, belge (Yessenalina ve ark., 2010), cümle (Pang ve ark., 2002) ve kelime (Breck ve ark., 2007) düzeyinde gerçekleştirilebilir. Duygu sınıflandırması için genellikle farklı yöntemler kullanılmaktadır. Duygu analizinde sözlüğe dayalı, makine öğrenmesine dayalı ve hibrit olmak üzere farklı yöntemler ele alınmaktadır (Medhat ve ark., 2014). Duygu analizinde kullanılan algoritmalar ve sınıflandırma yöntemleri Şekil 1.3. te gibi gösterilmiştir” (Medhat ve ark., 2014; Ravi and Ravi, 2015; Atlı ve İlhan, 2021).



Şekil 2.1. Duygu analizinde kullanılan algoritmalar ve sınıflandırma yöntemleri (Medhat ve ark., 2014; Ravi and Ravi, 2015)

“Makine öğrenmesi yaklaşımları dil yapılarının eğitilmesinde de kullanılmaktadır. Makine öğrenmesine dayalı yöntemlerin doğruluğu; uygun eğitilme kalitesine, kullanılan sınıflandırıcının uygunluğuna ve kullanılan eğitim setine bağlı olarak değişebilmektedir (Devika ve ark., 2016). Makine öğrenmesine dayalı yöntemlerin eğitim seti vb. alanlara bağımlı olmaları ve eğitim gerektirmeleri gibi dezavantajları bulunmaktadır. Ancak sözlük temelli yaklaşımlar herhangi bir eğitim seti gerektirmez ve eğitim gerektirmemesi sözlük tabanlı yaklaşımlara avantaj sağlamaktadır” (Atlı ve İlhan, 2021).

“Metin sınıflandırma için makine öğrenmesi yaklaşımları tercih edildiğinde denetimli ve denetimsiz öğrenme yöntemleri ele alınır. Denetimli öğrenmede, önceden etiketlenmiş bir veri kullanılarak bir sınıflandırma modeli oluşturulur. Daha sonra elde edilen sınıflandırma modeli, test verilerinin sınıf etiketi atamaları için kullanılır. Denetimli öğrenme yöntemi, makine öğrenmesi yaklaşımında daha fazla tercih edilmektedir. En yaygın olarak kullanılan denetimli öğrenme algoritmalarına baktığımızda, Olasılıksal Sınıflandırıcılar (Naive Bayes, Bayes Ağları, Maksimum-Entropi), Kural Tabanlı Sınıflandırıcılar, Doğrusal Sınıflandırıcılar (Destekçi Vektör

Makinesi, Sinir Ağları) ve Karar Ağaçları yöntemleridir” (Medhat ve ark., 2014) (Ravi and Ravi, 2015; Atlı ve İlhan, 2021).

“Etiketli eğitim verilerine ulaşmanın zor olduğu durumlarda denetimsiz öğrenme yöntemleri uygulanır. Önceden belirlenmiş duygu terimlerinin bir araya getirilmesi sözlük tabanlı yaklaşımlara kaynak sağlar. Sözlük tabanlı yaklaşımların doğrulukları, oluşturulan sözlükteki kelime-duygu ikililerinin bulunma derecelerinden etkilenebilmektedir (Koumpouri ve ark.,2015). Kullanılan duygu sözlüğünün ölçeği, metinden çıkarılacak duyguları doğrudan etkilemektedir. Sözlük tabanlı yaklaşımda, önce metinden duygu kelimeleri çıkarılır ve çıkarılan duygu kelimelerinin eş ve zıt anlamları sözlükte aranır (Medhat ve ark., 2014). Derlem tabanlı yöntemler ise, anlamsal ve istatistiksel yaklaşımlar olmak üzere iki farklı şekilde ele alınır ve kelime kökleri dikkate alınarak duygu içeren kelimelerin metin içerisinden aranması şeklinde gerçekleştirilir (Medhat ve ark., 2014; Pang and Lee, 2009). Hibrit yaklaşımlar, sözlük ve makine öğrenmesi yaklaşımlarının bir arada ele alınmasıyla oluşturulan yaklaşımlardır (Medhat ve ark., 2014; Liu, 2012). İstatistiksel yaklaşımda, istatistiksel yöntemler kullanılarak duygu ifade eden kelimelerin kökleri tespit edilir. Duygu, bir kelimenin pozitif-negatif cümle veya metin içinde bulunma sıklığı ile belirlenir. Kelimenin negatif veya pozitif cümlelerde bulunma frekansı, kelimenin pozitif, negatif veya nötr duygu bildiren bir kelime olması üzerinde etkilidir. Bir duyguyu ifade eden bir sözcükle bir kelimenin bir arada bulunma sıklığının daha yüksek olması, kelimenin benzer bir duyguya sahip olabileceği anlamına gelir(Medhat ve ark., 2014; Liu, 2012). Anlamsal yaklaşım, sözlük modeli oluşturma amacıyla benzer duygu içeren kelimelerin anlamsal olarak benzer olduğu temeline dayanır” (Liu, 2012; Atlı ve İlhan, 2021).

“Sözlük tabanlı yaklaşımda, duyguların çıkarılması için sözcük kaynakları kullanılır. Önceden elde edilmiş bir kelime havuzuna göre metinden duygu ifade eden kelimeler ve duyguların eşleştirilmesi şeklinde gerçekleştirilir. Sözlük tabanlı yaklaşımlarda, hem MPQA (Deng and Wiebe, 2015) ve SentiWordNet (Baccianella ve ark., 2010) gibi duygu sözlüklerinden hem de dil özelliklerinden faydalanırlar” (Atlı ve İlhan, 2021).

“Sözlük tabanlı yaklaşımları performans olarak değerlendirdiğimizde, sözlükteki duygu ifadelerinin kutupluluk derecesine bağlı olarak değiştiği görülmektedir (Turney, 2002). Sözlük tabanlı yaklaşımlarda, manuel olarak oluşturulan sözlüklerin yüksek maliyetinden kaçınmak için sözlüklerin otomatik bir şekilde oluşturulması önem kazanmıştır (Li ve ark., 2012; Hatzivassiloglou and McKeown, 1997). Duygu sözlüklerinin genişletilmesi için kelimelerin dilsel ilişkileri veya anlamsal sözlükleri (WordNet) kullanan birçok çalışma olmuştur (Rao and Ravichandran, 2009; Kamps ve ark., 2004). NRC Duygu Sözlüğü, kitle yaklaşımı ile oluşturulan bir sözlüktür (Mohammad ve ark., 2015). DepecheMood, kapsam açısından geniş ve yüksek hassasiyetli otomatik olarak oluşturulmuş bir sözlüktür (Staiano and Guerini, 2014). LIWC sözlüğü olarak bilinen Dilbilimsel Sorgulama ve Kelime Sayımı, geniş bir metin analiz yazılımıdır (Pennebaker ve ark., 2001). EmoSenticNet, SenticNet kavramlarına altı WordNetAffect duygu etiketi atayan sözcüksel bir kaynaktır (Poria ve ark., 2012). Empath, büyük duygusal kategori kümelerini analiz etmek için tasarlanmış kelime sınıflandırma yaklaşımıdır. Bununla birlikte, F. Koto ve M. Adrianitwitter verilerini kullanarak ve Plutchik’in duygu çarkına (Plutchik, 2001) atıfta bulunarak Hashtag Tabanlı Duygu (HBE) sözlüğünü oluşturmuşlardır (Koto and Adriani, 2015). Song ve arkadaşları 2016 yılında duygu-kelime sözlüğü kavramını, duyguların anlambilimini için içine katarak genişletmişlerdir” (Atlı ve İlhan, 2021).

Bu çalışmamızda diğer sözlüklere göre daha fazla sayıda duygu sınıflandırması yapabilen önerdiğimiz NAYALex Duygu Sözlüğü oluşturulmuş, test edilmiş, doğrulanmış ve kullanıma hazır hale getirilmiştir.

3. MATERYAL ve YÖNTEM

“Sosyal medyanın yaygın bir şekilde kullanımıyla paylaşılan içeriklerdeki duyguların ortaya çıkarılması giderek önem kazanmaktadır. Sosyal medya kullanıcıları duygularını, paylaştıkları metinlere yansıtarak bu alanda yapılacak çalışmalar için büyük veri kaynakları oluşturmaktadırlar. Farklı şekillerde ortaya çıkan duyguların, metinlerden çıkarılması duygu sözlükleri ile yapılabilmektedir. Ancak duygu sınıflandırması için kullanılan sözlüklere baktığımızda, anahtar kelimelere atanmış duygu ifadelerinin sınırlı sayıdaki duyguları içermesi dezavantaj olarak görülmektedir” (Atlı ve İlhan, 2021). Duygu sınıflandırması yapan sözlükler için az sayıdaki duygu çıkarımı dezavantajını ortadan kaldırmak için NAYALex Duygu Sözlüğünü geliştirip kütüphanesinin oluşturduk, test verisi ile uygulanabilirliğinin gösterdik ve doğrulama verisi ile güvenilirliğini gösterdik.

3.1. Materyal

3.1.1. Test verisi

“Deneysel işlemlerde kullanılan test veri seti, hesabı gizli olmayan Instagram kullanıcılarına ait” 11.880 medya paylaşımına ait metinsel yorum içeren İngilizce verilerden oluşmaktadır (Atlı ve İlhan, 2021). Elde edilen veriler ilk etapta üzerinde herhangi bir işlem yapılmamış doğrudan Instagram kullanıcılarının paylaşımlarına ait ham verilerdir.

Çizelge 3.1. de görülen veri kesitindeki her bir veri, farklı kullanıcılara ait farklı medya paylaşımlarına yapılmış İngilizce dilinde metinsel yorum bilgisi içermektedir.

Çizelge 0.1. Hesabı gizli olmayan Instagram kullanıcılarına ait işlenmemiş medya verileri

id	m_text
0	This kid. #brother #creatives
1	Feels good to be back in Chicago 88th. #redwi...
2	Always a delight to see this family. @danielap...
3	Instagram Inception @garrettcarter @beshark
4	Happy Vaisakhi
	...
6093	The #cabbageroses are my favorite design on th...
6094	March, it's you AGAIN??? Oh boy... I feel like...
6095	The samples of my new stationery line have arr...
6096	Calling all #sweetwine???? drinkers!\n\nThis b...

Veri seti Instagram API kullanılarak elde edilmiş hesabı gizli olmayan Instagram kullanıcılarına ait verileri içermektedir. Sadece İngilizce dilinde metinsel paylaşımlar dikkate alınarak elde edilen Instagram kullanıcılarına ait veriler test için kullanılmıştır. Veri seti üzerinde büyük-küçük harf dönüşümünün sağlanması, noktalama işaretlerinin temizlenmesi, sayısal mesaj içeriklerinin temizlenmesi ve etkisiz kelimelerin temizlenmesi gibi birtakım ön işlemler yapılmıştır.

3.1.1.1. Büyük-küçük harf dönüşümü

İşlenmemiş veri üzerinde büyük-küçük harf dönüşümü yapılarak tüm veriler küçük harfe dönüştürülmüştür. Veri setindeki her bir medya paylaşımına ait metinlerde yer alan büyük harfler, Çizelge 3.2. de gösterildiği gibi küçük harfe dönüşümü yapılmıştır.

Çizelge 0.2. Büyük-küçük harf dönüşümü yapılmış medya verileri

id	m_text
0	this kid. #brother #creatives
1	feels good to be back in chicago 88th. #redwin...
2	always a delight to see this family. @danielap...
3	instagram inception @garrettcarter @beshark
4	happy vaisakhi
5	little monster wants to steal my lunch
6	ray of sunshine on a cloudy week. *????
7	infinite possibilities
8	a night out in the land of or. @domiaru
9	finally some glorious sun here in vancouver.

3.1.1.2. Noktalama işaretlerinin temizlenmesi

Büyük-küçük harf dönüşümü tamamlanan Çizelge 3.2. deki medya verisinde bulunan noktalama işaretleri daha sonra Çizelge 3.3. te gösterildiği gibi temizlenmiştir.

Çizelge 0.3. Noktalama işaretleri temizlenmiş medya verileri

id	m_text
0	this kid brother creatives
1	feels good to be back in chicago 88th redwings...
2	always a delight to see this family danielapre...
3	instagram inception garrettcarter beshark
4	happy vaisakhi
5	little monster wants to steal my lunch
6	ray of sunshine on a cloudy week
7	infinite possibilities
8	a night out in the land of or domiaru
9	finally some glorious sun here in vancouver

3.1.1.3. Sayısal mesajların temizlenmesi

Noktalama işaretlerinin temizlenmesi aşamasından sonra veri setinde bulunan sayısal mesajlar Çizelge 3.4. te gösterildiği gibi temizlenmiştir.

Çizelge 0.4. Sayısal mesaj içerikleri temizlenmiş medya verileri

id	m_text
0	this kid brother creatives
1	feels good to be back in chicago th redwings b...
2	always a delight to see this family danielapre...
3	instagram inception garrettcarter beshark
4	happy vaisakhi
5	little monster wants to steal my lunch
6	ray of sunshine on a cloudy week
7	infinite possibilities
8	a night out in the land of or domiaru
9	finally some glorious sun here in vancouver

3.1.1.4. Etkisiz kelimelerin temizlenmesi

Sayısal mesajları temizlenmiş veri seti içinde bulunan etkisiz kelimeler(stopwords) Çizelge 3.5. te gösterildiği gibi veriden çıkarılmıştır.

Çizelge 0.5. Etkisiz kelimeleri temizlenmiş medya verileri

id	m_text
0	kid brother creatives
1	feels good back chicago th redwings blackhawks
2	always delight see family danielapresta rosell...
3	instagram inception garrettcarter beshark
4	happy vaisakhi
5	little monster wants steal lunch
6	ray sunshine cloudy week
7	infinite possibilities
8	night land domiaru
9	finally glorious sun vancouver

Elde edilen veri setinde bulunan kayıtlar içinde en az üç kelimedenden oluşmayan kayıtlar analizlerde göz ardı edilmek üzere veri setinden çıkarılmıştır. Test veri seti, NAYALex Duygu Sözlüğü için kullanılabilir bir test verisi haline getirilmiştir.

3.1.2. Doğrulama verisi

NAYALex Duygu Sözlüğü'nün doğrulaması için kullanılacak olan doğrulama veri seti üç farklı(imdb.com, amazon.com, yelp.com) web sitesinden alınmış duygu barındıran İngilizce cümleler içermektedir (Kotzias ve ark., 2015). İnsan değerlendiricilerle duyguların tespiti için kullanılacak olan doğrulama veri seti, üç farklı web sayfasından alınmış 3.000 farklı cümle içinden NAYALex Duygu Sözlüğü'nün 38 farklı duygusundan her biri için ilgili duyguyu açıkça ifade edebilen cümleler arasından rastgele 10'ar cümle olmak üzere 380 farklı cümle seçilmiştir. Çizelge 3.6. da örnek olarak gösterildiği gibi doğrulama veri setinin temelini oluşturan duygu yüklü cümleler 38 farklı duygu için toplamda 380 cümle içermektedir.

Çizelge 0.6. Duygu yüklü doğrulama veri seti(Kotzias ve ark., 2015)

S.No	Duygu Barındıran Cümle	Duygu
1	"So there is no way for me to plug it in here in the US unless I go by a converter."	positive
2	"At first glance it is a lovely bakery cafe - nice ambiance, clean, friendly staff."	sentimentality
3	"About half way through this film I started to dislike it."	contempt
4	"Then I exchanged for the same phone, even that had the same Problem"	despair
...
377	"If you see it, you should probably just leave it on the shelf."	disapproval
378	"After trying many many handsfree gadgets this is the one that finally works well."	unbelief
379	"I wish I could return the unit and get back my money."	outrage
380	"Very much disappointed with this company."	envy

Daha sonra duygu barındıran cümleler, değerlendirme ve sınıflandırmaya imkan verecek şekilde düzenlenerek değerlendiricilere sunulmuştur. Bir ay boyunca online bir şekilde insan değerlendiricilere sunulan çalışma için yaşları 25-40 arasında 70 farklı değerlendiriciye ulaşılmış ve duygu yüklü cümleler değerlendiricilere okutularak 38 farklı duygu için hangi duyguyu barındırdığını belirtmeleri istenmiştir. Çalışma sonunda 380 tane duygu yüklü cümle insan değerlendiriciler tarafından 38 farklı duygu için işaretlenmiştir. Bu şekilde elde edilen doğrulama veri seti, 26.600 farklı duygu işaretlemesi içermektedir. Doğrulama veri seti, NAYALex Duygu Sözlüğümüzün doğrulaması için kullanılmak üzere uygun hale getirilmiştir.

3.2. Yöntem

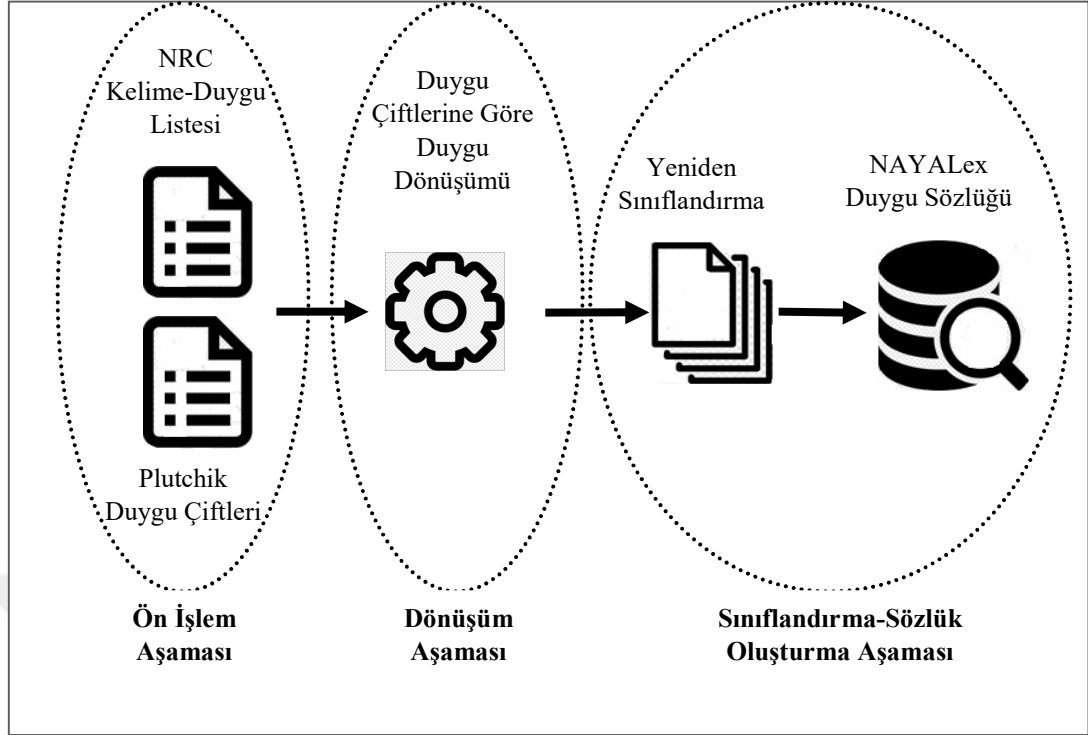
3.2.1. NAYALex duygu sözlüğü

Karşımıza farklı biçimlerde çıkan duyguların ortaya çıkarılması ve analiz edilmesi büyük önem taşımaktadır. Metinsel içeriklerden duyguların ortaya çıkarılması ve sınıflandırılması için birçok duygu sözlüğü kullanılmaktadır. Ancak mevcut duygu sözlükleri ele alındığında sınıflandırabildikleri duygu sayısı kişilik özelliklerinin ortaya çıkarılması için yeterli olmadığı görülmektedir. Sosyal medya aracılığı ile yapılan paylaşımlardan kişilik özelliklerinin ortaya çıkarılması birçok araştırma alanı için önem arz etmektedir. Bu özelliklerin daha kapsamlı bir şekilde

ortaya çıkarılması için de daha geniş kapsamlı duygu çıkarımlarına ihtiyaç duyulmaktadır. Kelime-duygu sözlüklerinin tespit edebildikleri duygu sayısı bakımından incelendiğinde, olumlu-olumsuz duygu ve 8 farklı duyguyla birlikte “en fazla sayıda duygu çıkarımı yapabilen NRC Duygu Sözlüğünden” daha kapsamlı bir duygu sözlüğüne ihtiyaç duyulduğu görülmektedir (Atlı ve İlhan, 2021). NAYALex Duygu Sözlüğü, kişilik özelliklerini ifade eden duyguların kişiliğin tespitinde daha fazla temsil edilmesini sağlayarak ve metinlerden daha çok duyguyu ortaya çıkararak bu alandaki eksikliğin ortadan kaldırılmaya kaynaklık edecektir. NAYALex Duygu Sözlüğü, literatürde bu alanda görülen eksikliğin giderilmesi için olumlu-olumsuz ile birlikte 38 farklı duyguyu(korku, kaygı, üzüntü, karamsarlık, iyimserlik, öfke, sevgi, umut vb.) ortaya çıkarabilen bir duygu sözlüğüdür. NRC Duygu Sözlüğü ve Plutchik’in Duygu Teorisi sentezlenerek elde edilmiştir. “Her bir kelimesi 38 farklı duygunun en az biri ile ilişkili 6469 İngilizce kelimedenden oluşan 245.822 duygu-kelime ilişkisi içermektedir” (Atlı ve İlhan, 2021).

3.2.2. NAYALex duygu sözlüğü oluşturma aşamaları

Kişiliğin belirlenmesinde büyük önemi olan duygu sınıflandırmalarının literatürdeki mevcut duygu sözlükleri ile kısıtlı bir şekilde yapılabilmesinden kaynaklanan dezavantajın ortadan kaldırılmasını sağlayan NAYALex Duygu Sözlüğü, Şekil 3.1. de gösterildiği gibi 3 ana aşamada oluşturulmuştur.



Şekil 3.1. Duygu sözlüğü oluşturma mimarisi

3.2.2.1 Ön işlem aşaması

“NRC Duygu Sözlüğündeki 14.182 kelimenin her birine duygu dönüşümü için bir ilişki derecesi atanarak NRC duygu kategorileri ile Çizelge 3.7. deki gibi eşleştirilmiştir” (Atlı ve İlhan, 2021).

Çizelge 0.7. NRC duygu sözlüğü (Mohammad, 2011; Atlı ve İlhan, 2021)

Sözlük	Kelime Sayısı	Duygu Kategorileri	İlişkilendirme Dereceleri
NRC Sözcük-Duygu Derneği Sözlüğü (NRC Duygu sözlüğü veya EmoLex olarak da adlandırılır)	14.182 unigram (kelime)	duygular: * olumlu * olumsuz	0 (ilişkili değil) veya 1 (ilişkili)
	~ 25.000 duygu	duygular: *öfke *beklenti *tiksinti *korku *sevinç *üzüntü *sürpriz *güven	İlişkili değil, zayıf, orta veya güçlü bir şekilde ilişkili

Aynı zamanda, Çizelge 3.8. de “NRC duygu kategorileri, Plutchik'in duygu çiftlerini oluşturmak için birleştirilir. Bir sonraki dönüşüm aşaması için NRC ikili Duygu kombinasyonları ve Plutchik'in duygu teorisindeki kategorileri oluşturmak için birleştirilir” (Atlı ve İlhan, 2021).

Çizelge 0.8. Plutchik'in duygu çiftleri (Drews, 2007)

#	İkili Duygu Kombinasyonları	Duygular	#	İkili Duygu Kombinasyonları	Duygular
1	Beklenti+Sevinç	İyimserlik, Cesaret	15	Sürpriz+Üzüntü	Onaylanmama, Hayal Kırıklığı
2	Beklenti+Güven	Umut, Kadercilik	16	Sürpriz+Tiksinme	İnançsızlık, Şok
3	Beklenti+Korku	Kaygı, Dehşet	17	Sürpriz+Öfke	Öfke, Nefret
4	Sevinç+Güven	Sevgi, Dostluk	18	Üzüntü+Tiksinme	Pişmanlık, Sefalet
5	Sevinç+Korku	Suçluluk, Heyecan	19	Üzüntü+Öfke	Kıskançlık, sıkıcılık
6	Sevinç+Sürpriz	Keyif, Kader	20	Üzüntü+Beklenti	Karamsarlık
7	Güven+Korku	Teslimiyet, Tevazu	21	Tiksinme+Öfke	Küçümseme
8	Güven+Sürpriz	Merak	22	Tiksinme+Beklenti	Alaycılık
9	Güven+Üzüntü	Duygusalılık,	23	Tiksinme Neşe	Hastalık, Kötülük
10	Korku+Sürpriz	Huşu, Telaş	24	Öfke+Beklenti	Saldırganlık, İntikam
11	Korku+Üzüntü	Umutsuzluk	25	Öfke+Sevinç	Gurur, Zafer
12	Korku+Tiksinme	Utanç, Küstahlık	26	Öfke+Güven	Hakimiyet
13	Sevinç+Üzüntü	Acı tatlılık	27	Korku+Öfke	Donukluk
14	Güven+Tiksinme	Kararsızlık	28	Sürpriz+Beklenti	Bilinç bul., şaşkınlık

3.2.2.2 Dönüşüm aşaması

“NRC Duygu sözlüğündeki her bir kelime, olumlu-olumsuz duygular ve 8 farklı duygu (neşe, güven, korku, şaşkınlık, üzüntü, tiksinişme, öfke ve beklenti) için 0 (ilişki yok) ve 1 (ilişki var) şeklinde derecelendirilmiştir” (Mohammad, 2016; Atlı ve İlhan, 2021). NAYALex Duygu Sözlüğü için “ön işlem aşamasında elde edilen duygu setleri referans alınarak, Plutchik duygu çiftlerinin etki ettiği ve ortaya çıkardığı yeni duygularla” eşleştirmeler yapılmıştır (Atlı ve İlhan, 2021). Bu eşleştirmeler sonucunda 28 duygu çiftinin ortaya çıkardığı yeni duygular belirlenmiştir.

3.2.2.3 Sınıflandırma-sözlük oluşturma aşaması

“NRC Duygu Sözlüğü’nde bulunan 8 duygu ile Plutchik’in Temel Duyguların Psikoevrimsel Teorisi’ne (Drews and Krohn, 2007) göre belirlenen 8 temel duygu aynıdır. Plutchik Duygu Çarkı’nda bulunan 8 temel duygu birbirleriyle birleşerek 28 duygu çifti oluşturmakta ve bu duygu çiftlerinin birleşimi daha farklı duyguları meydana getirmektedir. Çizelge 3.8. deki ikili Duygu Kombinasyonları dikkate alınarak oluşturulmuş duygu çiftleri için NRC Duygu Sözlüğü’nde bulunan her bir kelimenin ilişkili olduğu duygular, Plutchik’in duygu çiftlerine göre uyarlanmakta ve 28 farklı duygudaki her bir kelime Plutchik’in Çizelge 3.9. daki duygu kategorileriyle yeniden ilişkilendirilmektedir. Bu ilişkilendirme şu şekilde yapılmaktadır; 28 farklı duygu çifti için NRC Duygu Sözlüğü’ndeki her kelime aranmakta ve bulunan her duygu çiftine karşılık gelen duygu için 0 ve 1 şeklinde derecelendirilmektedir. Elde edilen kelime-duygu sözlüğündeki her kelime, ayrıca NRC Duygu sözlüğünde doğrudan karşılık bulan olumlu-olumsuz ve 8 farklı duygu için derecelendirilerek 38 farklı duygu içeren sözlük ortaya çıkarılmaktadır” (Atlı ve İlhan, 2021).

“NAYALex sözlüğünün oluşturulması Çizelge 3.9. da sembolik olarak ifade edilmiştir. Sembolik gösterimdeki (X,Y,Z) birer kelime, (A,B,C) NRC sözlüğünde bulunan ve (X,Y,Z) kelimeleriyle ilişkili olan duygular, (A+B, B+C, A+C) Plutchik’in teorisinde yer alan duygu çiftleri, (J,K,L) duygu dönüşümü sonucu ortaya çıkan ve NAYALex sözlüğünde bulunan duyguları belirtmektedir” (Atlı ve İlhan, 2021).

Çizelge 0.9. Duygu dönüşümü ve yeniden sınıflandırma sembolik gösterimi

Sözlük Kelimesi (X,Y,Z birer kelime)	NRC Duygu Etiketleri (A,B,C birer duygu)	Dönüşüme Etki Eden Plutchik Duygu Çifti (J,K,L oluşturulan yeni duygular)	NAYALex'te Oluşan Duygu Etiketleri 1→ilgili duygu sözlük kelimesi ile ilişkili 0→ ilgili duygu sözlük kelimesi ile ilişkili değil
X	A → 1 B → 1 C → 0	A + B → J	A → 1 B → 1 C → 0 J → 1 K → 0 L → 0
Y	A → 0 B → 1 C → 1	B + C → L	A → 0 B → 1 C → 1 J → 0 K → 0 L → 1
Z	A → 1 B → 1 C → 1	A + B → J A + C → K B + C → L	A → 1 B → 1 C → 1 J → 1 K → 1 L → 1

“NRC Duygu sözlüğünde bulunan her bir kelime (örneğin “X” kelimesi); NRC Duygu Sözlüğünde ilişkili olduğu duygular (A,B) için “1”, ilişkili olmadığı duygular (C) için “0” ile belirtilmiştir. “X” kelimesi için duygu dönüşümüne etki eden Plutchik duygu çifti/çiftleri A + B → J şeklinde ifade edilmiştir. Duygu dönüşümü sonrası oluşan NAYALex sözlüğümüzde “X” kelimesi için oluşan duygu ilişkileri A → 1, B → 1, C → 1, J → 1, K → 0, L → 0 şeklinde olmaktadır” (Atlı ve İlhan, 2021).

“NRC ve NAYALex sözlükleri için bazı kelimelerin geçtiği metinlerin ilişkili (1) etki gösterdiği duygular” Çizelge 3.10. da örnek olarak gösterilmiştir (Atlı ve İlhan, 2021).

Çizelge 0.10. NRC ve NAYALex kelime-duygu eşleştirmesi (Atlı ve İlhan, 2021)

Sözlükte Bulunan Kelime	İlişkili Olduğu Duygular	
	NRC Sözlüğü	NAYALex Sözlüğü
tribulation (sıkıntı)	Negatif, Korku, Üzüntü	Negatif, Korku, Üzüntü, Umutsuzluk
palsy (acizlik)	Negatif, Tikslenme, Korku, Üzüntü	Negatif, Tikslenme, Korku, Üzüntü, Utanç, Pişmanlık, Umutsuzluk
accolade (övgü)	Pozitif, Beklenti, Sevinç, Sürpriz, Güven	Pozitif, Beklenti, Sevinç, Sürpriz, Güven, İyimserlik, Şaşkınlık, Umut, Keyif, Sevgi, Merak
mighty (güçlü)	Pozitif, Öfke, Korku, Sevinç, Güven	Pozitif, Öfke, Korku, Sevinç, Güven, Donukluk, Gurur, Hakimiyet, Heyecan, Tevazu, Sevgi

3.2.3. NAYALex kütüphanesi

Sembolik olarak oluşturduğumuz NAYALex sözlüğümüzün veriler üzerinde sınıflandırma yapmasına olanak sağlayan Şekil 3.2. de Python dilinde nayalex.py kütüphanesini oluşturduk.

```

1  # -*- coding: utf-8 -*-
2  """Created on Fri Feb 12 23:53:36 2021
3  @author: YAKUP"""
4  from textblob import TextBlob
5  from collections import Counter
6  def build_word_affect(self):
7      # Build word affect function
8      affect_list = []
9      affect_dict = dict()
10     affect_frequencies = Counter()
11     lexicon_keys = self.lexicon.keys()
12     for word in self.words:
13         if word in lexicon_keys:
14             affect_list.extend(self.lexicon[word])
15             affect_dict.update({word: self.lexicon[word]})
16     for word in affect_list:
17         affect_frequencies[word] += 1
18     sum_values = sum(affect_frequencies.values())
19     affect_percent = {'positive': 0.0, 'negative': 0.0, 'optimism': 0.0,
20                     'hope': 0.0, 'anxiety': 0.0, 'love': 0.0,
21                     'guilt': 0.0, 'delight': 0.0, 'submission': 0.0,
22                     'curiosity': 0.0, 'sentimentality': 0.0,
23                     'awe': 0.0, 'despair': 0.0, 'shame': 0.0,
24                     'disapproval': 0.0, 'unbelief': 0.0,
25                     'outrage': 0.0, 'remorse': 0.0,
26                     'envy': 0.0, 'pessimism': 0.0, 'contempt': 0.0,
27                     'cynicism': 0.0, 'morbidness': 0.0,
28                     'aggressiveness': 0.0, 'pride': 0.0,
29                     'dominance': 0.0, 'bittersweetness': 0.0,
30                     'ambivalence': 0.0, 'frozenness': 0.0,
31                     'confusion': 0.0, 'fear': 0.0, 'anger': 0.0,
32                     'anticipation': 0.0, 'trust': 0.0,
33                     'surprise': 0.0, 'sadness': 0.0,
34                     'disgust': 0.0, 'joy': 0.0}
35     affect_percent2 = {}

```

Şekil 3.2. NAYALex duygu sözlüğü python kütüphanesi

Oluşturduğumuz Python Kütüphanesi, 6469 kelime ve bu kelimelerin 38 farklı duygu ile ilişkisini içermektedir. Oluşturulan sözlük yapısındaki 6469 kelimenin her biri için karşılık bulan duygu eşleştirmeleri Şekil 3.3. te gösterildiği gibi Python Kütüphanesine yansıtılacak şekilde düzenlenmiştir. NAYALex Duygu Sözlüğü yapısına uygun olarak oluşturduğumuz Python Kütüphanesi kullanılarak metinler üzerinde 38 farklı duygu için sınıflandırma yapılabilmektedir.

```

35     affect_percent2 = {}
36     for key in affect_frequencies.keys():
37         affect_percent.update({key: float(affect_frequencies[key]) / float
38     self.affect_list = affect_list
39     self.affect_dict = affect_dict
40     self.raw_emotion_scores = dict(affect_frequencies)
41     self.affect_frequencies = affect_percent
42
43
44     def top_emotions(self):
45         emo_dict = self.affect_frequencies
46         max_value = max(emo_dict.values())
47         top_emotions = []
48         for key in emo_dict.keys():
49             if emo_dict[key] == max_value:
50                 top_emotions.append((key, max_value))
51         self.top_emotions = top_emotions
52
53
54     class NAYALex:
55         #print('NAYALex çalışıyor.... anticipation düzeltmesi sonrası')
56         lexicon = {'abacus': ['trust'], 'abandon': ['fear', 'negative', 'sadness'],
57                 'abandoned': ['anger', 'fear', 'negative', 'sadness', 'depression'],
58                 'abandonment': ['anger', 'fear', 'negative', 'sadness', 'depression'],
59                 'abba': ['positive'], 'abbot': ['trust'], 'abduction': ['fear', 'negative', 'sadness'],
60                 'aberrant': ['negative'], 'aberration': ['disgust', 'negative'],
61                 'abhor': ['anger', 'disgust', 'fear', 'negative', 'shame'],
62                 'abhorrent': ['anger', 'disgust', 'fear', 'negative', 'shame'],
63                 'ability': ['positive'], 'abject': ['disgust', 'negative'],
64                 'abolish': ['anger', 'negative'], 'abolition': ['negative']

```

Şekil 3.3. NAYALex duygu sözlüğü python kütüphanesi kelime-duygu ilişkileri

3.2.4. Uygulanan sınıflandırma algoritmaları

3.2.4.1. Naive bayes

Naive Bayes sınıflandırma algoritması, bir sınıfı veya kategorisi bulunan önceden sınıflandırılmış veri üzerinde olasılıksal işlemler yaparak test verisinin kategorisinin/sınıfını tespit etmeyi amaçlar. Kullanımı kolay ve başarı oranı yüksek sınıflandırma algoritmalarından biridir (Aytuğ, 2017). Naive Bayes algoritması etkin hesaplama, sade, yüksek performans ve uygulama kolaylığı bakımından birçok uygulamada yaygın olarak sınıflandırma problemlerine çözüm bulmak için kullanılmaktadır (El Hindi, 2014).

3.2.4.2. Destek vektör makinesi(SVM)

Denetimli öğrenme algoritması olan Destek Vektör Makinesi(SVM), regresyon veya sınıflandırma problemlerinde kullanılır. Sınıflandırma problemlerinde yaygın olarak kullanılır. SVM(Destek vektör makinesi) 2 boyutlu, 3 boyutlu ve çok boyutlu uzayda doğrusal, düzlemsel ve hiperdüzlem olmak üzere farklı ayırma yeteneklerine sahip olduğundan veriyi iki veya daha çok sınıfa ayırabilmektedir (Güran ve ark., 2014).

3.2.4.3. KNN(K-nearest neighbors)

K-En Yakın Komşular (kNN), birçok durumda etkili olan parametrik olmayan bir sınıflandırma yöntemidir (Hand ve ark., 2001). Farklı algoritmalar için değerlendirme standartlarından biri performanslarıdır. kNN, basit ama etkili bir sınıflandırma yöntemidir ve metin sınıflandırmada en etkili yöntemlerden biri olarak bilinmektedir. kNN sınıflandırıcı, veriler üzerinde dikkate değer performans göstermektedir. Ancak, kNN sınıflandırmasının performansı, k değerinin seçimi, mesafe ölçülerinin seçimi vb. gibi bazı sorunlardan etkilenebilir. Bu sorunların üstesinden gelmek için birçok teknik geliştirilmiştir (Zhang, 2010)..

3.2.5. Sınıflandırma algoritmaları performans değerlendirme ölçütleri

3.2.5.1. Confusion matrix(Karışıklık matrisi)

Karışıklık matrisi, sınıflandırma problemleri için önemli olan performansların hesaplanması için kullanılır. Matris, sınıflandırılan verinin öngörülen değerlerinin ve gerçek değerlerinin farklı 4 kombinasyonunu Çizelge 3.11. de içermektedir.

Çizelge 0.11. Karışıklık matrisi(Han and Kamber, 2006)

Karışıklık Matrisi		Verinin Tahmin Edilen Sınıfı	
		Negatif (AA+) Müdahale Edilmeli	Pozitif (AA-) Gözlemlenmeli
Verinin Gerçek Sınıfı	Negatif (AA+) Müdahale Edilmeli	Doğru Negatif (DN)	Yanlış Pozitif (YP)
	Pozitif (AA-) Gözlemlenmeli	Yanlış Negatif (YN)	Doğru Pozitif (DP)

3.2.5.2. Doğruluk(Accuracy)

Doğruluk, modelin sınıflandırma başarısını göstermektedir. Sınıflandırma performansını görmek için kullanılan en yaygın ölçütlerden biridir. Doğru sınıflandırılmış örnekler ile toplam örnek sayısına oranı Denklem 3.1'deki gibi hesaplanmaktadır (Witten ve ark., 2016). Sıfır ve bir arasında değişen doğruluk oranının bire yaklaşması yüksek sınıf başarısı anlamına gelmektedir.

$$Doğruluk = \frac{DP+DN}{DP + DN + YP + YN} \quad (3.1)$$

3.2.5.3. Duyarlılık(Recall)

Pozitif olarak sınıflandırılmış bir örneğin tahmin ile de pozitif olma olasılığını ifade eder (Han and Kamber, 2006). Pozitif olarak tahmin edilmesi öngörülen değerlendirmelerin ne ölçüde pozitif tahmin edildiğini ifade eden bir metriktir. Duyarlılık Denklem 3.2'deki gibi hesaplanmaktadır. Duyarlılık değerinin bire yaklaşması sınıflandırmanın yüksek performansını ifade etmektedir.

$$Duyarlılık = \frac{DP}{DP + YN} \quad (3.2)$$

3.2.5.4. Kesinlik(precision)

Kesinlik(precision), pozitif olarak tahminlediğimiz değerlerin gerçekten kaç adedinin pozitif olduğunu göstermektedir. Sıfır ve bir arasında değişen kesinlik değerinin bire yaklaşması yüksek sınıflandırma performansı anlamına gelmektedir (Larose, 2014).

$$Kesinlik = \frac{DP}{DP + YP} \quad (3.3)$$

3.2.5.5. F-Ölçütü

Kesinlik ve duyarlılık ölçütlerinin ayrı ayrı değerlendirilmesinden kaynaklanan eksikliğin ortadan kaldırılması amacıyla hesaplanmaktadır (Powers, 2020). F-Ölçütü, duyarlılık ve kesinlik değerlerinin harmonik ortalaması alınarak oluşturulmaktadır. Sıfır ve bir arasında değişen F-Ölçüt değerinin bire yaklaşması yüksek sınıflandırma performansı anlamına gelmektedir. F-Ölçütü Denklem 3.4'teki gibi hesaplanır.

$$F - \text{Ölçütü} = 2 * \frac{\text{Duyarlılık} * \text{Kesinlik}}{\text{Duyarlılık} + \text{Kesinlik}} \quad (3.4)$$

4. ARAŞTIRMA BULGULARI ve TARTIŞMA

Bu bölüm üç aşamada ele alınacaktır. Birinci aşamada; “NRC Duygu Sözlüğü (Mohammad, 2016) ve Plutchik’in Temel Duyguların Psikoevrimsel Teorisine (Drews and Krohn, 2007) dayanarak melez bir şekilde oluşturulan NAYALex Duygu Sözlüğünün test veri seti” üzerindeki uygulamaları, analizleri ve değerlendirmeleri yapılacaktır (Atlı ve İlhan, 2021).

İkinci aşama; NAYALex Duygu Sözlüğünün diğer duygu sözlükleri ile duygu-kapsam bakımından karşılaştırılması ve değerlendirilmesi şeklinde gerçekleştirilecektir.

Üçüncü aşamada; insan değerlendiriciler ile yapılan çalışma sonucunda ortaya çıkan doğrulama veri seti üzerinde bir takım deneyler yapılarak NAYALex Duygu Sözlüğünün doğrulaması ile ilgili analizler ve değerlendirmeler yapılacaktır.

4.1. NAYALex Duygu Sözlüğü İle Sınıflandırma Ve Analiz

NAYALex Duygu Sözlüğü, verilen bir veri üzerinde doğrudan sınıflandırma yapabilme kabiliyetine sahip bir duygu sözlüğü olarak tasarlanmıştır. Sınıflandırma ve sınıflandırma sonucunda elde edilen değerler ile analiz yapılmasına olanak sunmaktadır.

Çizelge 4.1. Test verisi için kullanılacak instagram kullanıcılarına ait işlenmemiş medya verileri

id	m_text
0	This kid. #brother #creatives
1	Feels good to be back in Chicago 88th. #redwi...
2	Always a delight to see this family. @danielap...
3	Instagram Inception @garrettcarter @beshark
4	Happy Vaisakhi
	...
6093	The #cabbageroses are my favorite design on th...
6094	March, it's you AGAIN??? Oh boy... I feel like...
6095	The samples of my new stationery line have arr...
6096	Calling all #sweetwine???? drinkers!\n\nThis b...

NAYALex Duygu Sözlüğü'nün sınıflandırma ve analiz kabiliyetinin ortaya koymak için NAYALex Duygu Sözlüğü Çizelge 4.1. de örnek olarak verilen veri seti üzerinde çalıştırılmadan önce veri seti bir dizi ön işlemlere tabi tutularak Çizelge 4.2. deki sınıflandırmaya uygun test veri seti elde edilmiştir.

Çizelge 4.2. Instagram kullanıcılarına ait sınıflandırmaya uygun test veri seti

	m_id	k_id	m_text
0	4.305834e+17	1867426	kid brother creatives
1	4.328110e+17	1867426	feel good back chicago th redwing blackhawks
2	4.349117e+17	3534728	little monster want steal lunch
3	4.369365e+17	2274161	ray sunshine cloudy week
4	4.383627e+17	1867426	finally glorious sun van couver
...
2618	2.514123e+18	2430770	red one hope collection disappoint price point...

Çizelge 4.2. de belirtilen Instagram kullanıcılarına ait sınıflandırmaya uygun test veri seti için NAYALex sözlüğü çalıştırılarak medya verileri Çizelge 4.3. te 38 farklı duyguya göre sınıflandırılmıştır.

Çizelge 4.3. NAYALex duygu sözlüğü ile sınıflandırılmış veriler

	m_id	k_id	texti	positive	negative	optimism	hope	anxiety	love	guilt	...
0	430583350179523648	1867426	kid brother creatives	0.50	0.0	0.0	0.0	0.0	0.0	0.0	...
1	432810970224913600	1867426	feel good back chicago th redwing blackhawks	0.09	0.0	0.09	0.09	0.0	0.09	0.0	...
2	434911667177514560	3534728	little monster want steal lunch	0.00	0.22	0.0	0.0	0.0	0.0	0.0	...

Test veri setindeki her bir medya paylaşımı NAYALex Duygu Sözlüğü ile ayrı ayrı 38 farklı duygu için Çizelge 4.3. te gösterildiği gibi sınıflandırılmaktadır. Veri setinde sınıflandırılmış medya paylaşımının bazıları aynı kullanıcılara ait verilerdir. Söz konusu kullanıcıların her biri için NAYALex ile sınıflandırılmış medya verileri içinde farklı sayıda medya verileri filtrelenerek kullanıcı bazlı duygu

ortalamaları elde edilebilir. Elde edilen kullanıcı bazlı duygu ortalamaları ile bir kullanıcının kişiliği ile ilgili bilgi sahibi olunabilmektedir. Bir kullanıcının kişiliği, NAYALex ile sınıflandırılması sonucu elde edilen her bir medya paylaşımına göre değerlendirmek yerine Çizelge 4.4. te test veri setinde ilgili kullanıcı için tüm medya verilerinin sınıflandırma oranlarının ortalama değeri dikkate alınarak kişiliğin değerlendirilmesi daha kapsayıcı ve belirleyici sonuçlara ulaşmaya imkan vermektedir.

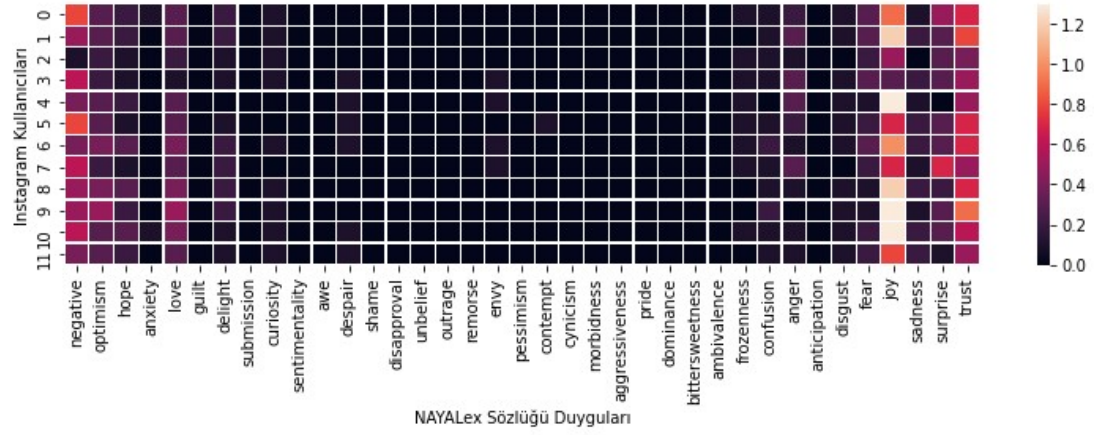
Çizelge 4.4. Kullanıcı bazlı ortalama duygu değerleri

	k_id	positive	negative	optimism	hope	anxiety	love	guilt	delight	submission	...
0	48863	0.15	0.08	0.03	0.02	0.01	0.03	0.0	0.02	0.0	...
1	341431	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
2	1046785	0.29	0.02	0.0	0.03	0.0	0.01	0.0	0.0	0.05	...
3	1188884	0.2	0.05	0.03	0.02	0.0	0.03	0.0	0.02	0.0	...
4	1742570	0.22	0.01	0.02	0.01	0.0	0.02	0.0	0.01	0.0	...
5	1867426	0.17	0.06	0.02	0.01	0.0	0.01	0.0	0.01	0.0	...
6	2199297	0.23	0.04	0.03	0.02	0.0	0.03	0.0	0.0	0.0	...
7	2274161	0.14	0.08	0.03	0.01	0.0	0.03	0.0	0.01	0.0	...
8	2430770	0.21	0.04	0.04	0.03	0.0	0.04	0.0	0.02	0.0	...
9	2512609	0.12	0.06	0.02	0.01	0.0	0.03	0.0	0.02	0.0	...
10	3056990	0.17	0.05	0.04	0.03	0.0	0.04	0.0	0.02	0.0	...

Çizelge 4.4. te ortalama duygu değerleri gösterilen “2430770” id nolu Instagram kullanıcısının test veri setinde 285 farklı medya paylaşımı bulunmaktadır. Test verisindeki ilgili kullanıcı hakkında bir değerlendirme yapmak için NAYALex tarafından sınıflandırılmış 285 medya verisini ayrı ayrı değerlendirmek yerine, Çizelge 4.4. te “2430770” id nolu kullanıcı için ortalama duygu değerlerini incelemek daha kapsayıcı ve belirleyici olmaktadır. Örneğin *anxiety(kaygılı)* bir kişiliğe sahip olmadığı ama *hope(umutlu)* olduğu ve daha çok *positive(olumlu)* söylenebilir.

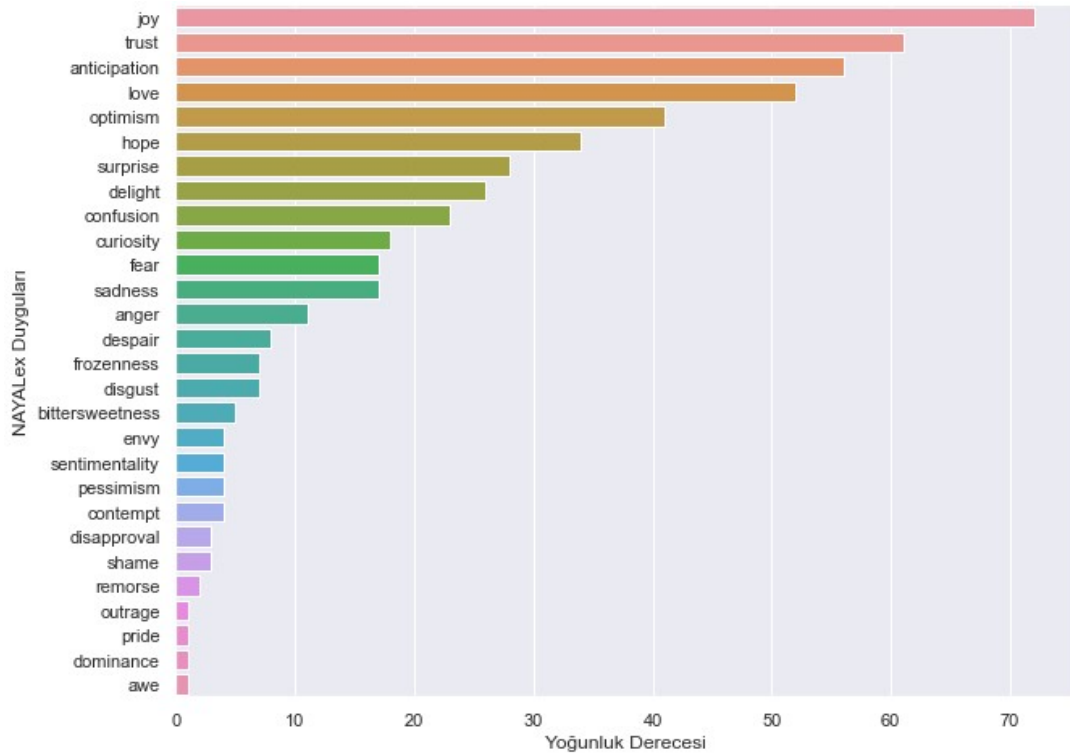
Veri setindeki tüm paylaşımlar her bir kullanıcı için filtrelenerek elde edilmiş ortalama duygu yoğunlukları Şekil 4.1. deki gibi ateş grafiği ile gösterilerek, veri setinde bulunan kişiler arasındaki benzer duygu yoğunlukları izlenebilmekte ve

araştırmaya tabi kullanıcıların yoğunlaştıkları duygu aralıkları görülebilmektedir.



Şekil 4.1. Kullanıcı bazlı duygu yoğunlukları

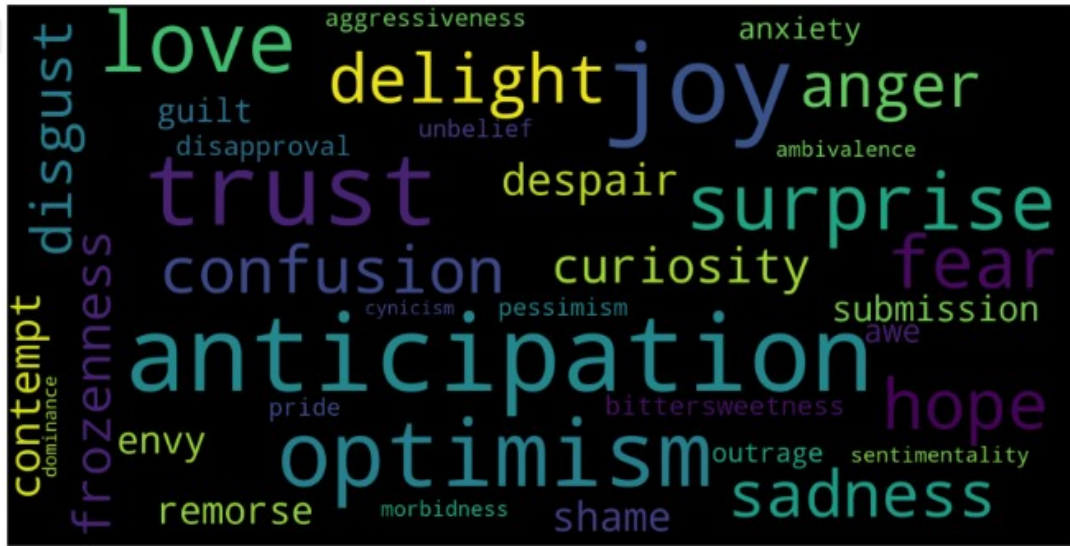
Örneğin 11 nolu Instagram “kullanıcısının veri setinde 232 farklı Instagram gönderisi-paylaşımı bulunmaktadır. Veri setinde bulunan 11 no’lu Instagram kullanıcısının 232 farklı paylaşımının duygu yoğunluğu bireysel olarak Şekil 4.2. de ele alınmıştır” (Atlı ve İlhan, 2021).



Şekil 4.2. 11 nolu instagram kullanıcısına ait duygu yoğunlukları

11 nolu kullanıcının tüm paylaşımlarını ayrı ayrı değerlendirmek yerine, “ortalama duygu yoğunluğu dikkate alınarak Şekil 4.2. de oluşturulan yoğunluk derecesi grafiğine göre en yoğun 3 duygusunun “*joy (sevinç)*”, “*trust (güven)*” ve “*anticipation (beklenti)*” olduğu görülmektedir” (Atlı ve İlhan, 2021).

“Alışılmış “kelime bulutu” kavramına farklı bir bakış katarak, başka bir kullanıcı olan 5 nolu kullanıcı için NAYALex Duygu Sözlüğü kullanılarak sınıflandırılmış 222 paylaşımı için örnek bir “duygu bulutu” Şekil 4.3. te gösterilmiştir” (Atlı ve İlhan, 2021).



Şekil 4.3. “5” nolu kullanıcıya ait duygu bulutu

Duygu bulutu ile kullanıcı kişiliğini yansıtan baskın duyguların ön plan çıktığı ve buna göre 5 no’lu kullanıcı kişisinin 11 no’lu kullanıcı kişisine göre daha çok “*anticipation (beklenti)*” içinde olan bir kişiliği yansıttığı görülmektedir.

4.2. NAYALex Duygu Sözlüğünün Diğer Duygu Sözlükleri İle Karşılaştırılması

“Duygu analizi için kullanılan sözlükler ile NAYALex duygu sözlüğü, duygu-kapsam bakımından Çizelge 4.5 teki gibi karşılaştırılmıştır” (Atlı ve İlhan, 2021).

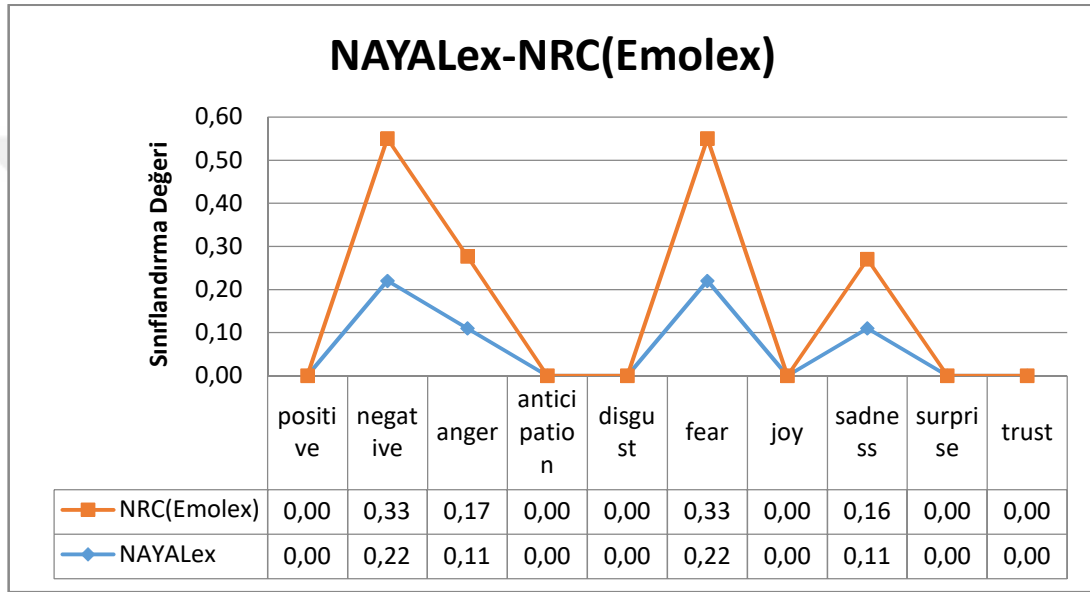
Çizelge 4.5. NAYALex'in diğer duygu sözlükleri ile duygu-kapsam karşılaştırması

Duygu	NAYALex	NRC (Emolex)	LIWC	EmoSenti c -Net	Empath
Positive (olumlu)	√	√	√	x	x
Negative (olumsuz)	√	√	√	x	x
Anger (öfke)	√	√	√	√	√
Sadness (üzüntü)	√	√	√	√	√
Joy (sevinç)	√	√	x	√	√
Fear (korku)	√	√	x	√	√
Surprise (sürpriz)	√	√	x	√	√
Disgust (tikinti)	√	√	x	√	x
Trust (güven)	√	√	x	x	x
Anticipation (beklenti)	√	√	x	x	x
Love (sevgi)	√	x	x	x	√
Optimism (iyimserlik), Hope (umut), Anxiety (kaygı), Guilt (suçluluk), Delight (kader), Submission (teslimiyet), Curiosity (merak), Sentimentality (duygusallık), Awe (huşu), Despair (umutsuzluk), Shame (utanç), Bittersweetness (acı tatlılık), Disapproval (onaylamama), Ambivalence (kararsızlık), Unbelief (inançsızlık), Outrage (öfke-nefret), Remorse (pişmanlık), Envy (kıskançlık), Pessimism (karamsarlık), Contempt (küçümseme), Cynicism (alaycılık), Morbidness (kötülük), Aggressiveness (saldırganlık), Pride (gurur), Dominance (hakimiyet), Frozenness (donukluk), Confusion (bilinç bulanıklığı)	√	x	x	x	x
<p>√ : Duygu, ilgili sözlük tarafından sınıflandırılabilir. x : Duygu, ilgili sözlük tarafından sınıflandırılmaz.</p>					

Çizelge 4.5. te ifade edilen duygular, “her bir sözlük için sınıflandırılabilir olup-olmadığı dikkate alınarak ifade edilmiştir” (Atlı ve İlhan, 2021). “Tablolaştırılmış duygular dikkate alındığında, kişiliğin belirlenmesi üzerinde etkili

olan duygular bakımından en fazla sayıda duyguyu sınıflandırabilen sözlüğün NAYALex Duygu Sözlüğü olduğu ve diğer duygu sözlüklerini kapsadığı görülmektedir” (Atlı ve İlhan, 2021).

Veri setinden rastgele seçilmiş bir cümle üzerinde NAYALex ve NRC duygu sözlükleri sınıflandırabildikleri ortak duygular için çalıştırılarak üretilen duygu yoğunluk değerleri Şekil 4.4. te gösterilmiştir.



Şekil 4.4. NAYALex ve NRC(Emolex) ortak duygular için sınıflandırma oranları

Şekil 4.4. te NAYALex-NRC(Emolex) duygu yoğunluk sınıflandırma değerleri incelendiğinde birbirine paralel sınıflandırma sonuçları verdikleri gözlemlenmektedir. Değerler arasındaki nicel farklılık, NAYALex Duygu Sözlüğünün sınıflandırabildiği fazla sayıdaki duygunun oransal olarak duygu yoğunluklarına olan etkisinden kaynaklanmaktadır.

4.3. NAYALex Duygu Sözlüğünün Doğrulaması

NAYALex Duygu Sözlüğü doğrulaması için insan değerlendiriciler ile bir çalışma yapılmıştır ve bunun sonucunda NAYALex Duygu Sözlüğünün doğrulamasına imkân veren doğrulama veri seti elde edilmiştir. Doğrulama veri seti, 380 farklı cümlenin 38 duyguya göre 70 farklı insan değerlendiricisi tarafından

duygu işaretlemesi yapılmış 26.600 cümle-duygu sınıflandırma ilişkisi içermektedir. Doğrulama veri setini incelediğimizde, NAYALex'in sınıflandırabildiği 38 farklı duygunun her biri için 10 farklı cümle içermekte ve bu cümleleri ifade etmesi beklenen duygu yoğunluk oranları 0-1 arasında bir değere sahiptir. İnsan değerlendiricilerle yapılan çalışma sonucunda her bir duygu için elde edilen 700 cümle-duygu sınıflandırma ilişkisi dikkate alınarak Çizelge 4.6. da gösterildiği gibi ortalama duygu yoğunluk oranları elde edilmiştir. Ortalama duygu yoğunluğu, ilgili duygu ile etiketlenen ilişki sayısının toplam ilişki sayısına oranı ile elde edilmektedir.

Çizelge 4.6. Ortalama duygu yoğunluk oranları (0.0 - 1.0 aralığında)

#	Duygu	İlgili Duygu İle Etiketlenen İlişki Sayısı	Toplam İlişki Sayısı	Ortalama Duygu Yoğunluğu
1	positive	660	700	0,94
2	negative	670	700	0,96
3	optimism	198	700	0,28
4	delight	126	700	0,18
5	disapproval	101	700	0,14
6	pessimism	47	700	0,07
7	hope	112	700	0,16
8	unbelief	48	700	0,07
9	curiosity	66	700	0,09
10	cynicism	48	700	0,07
11	anxiety	71	700	0,10
12	awe	114	700	0,16
13	outrage	152	700	0,22
14	aggressiveness	77	700	0,11
15	love	152	700	0,22
16	sentimentality	82	700	0,12
17	remorse	89	700	0,13
18	morbidness	61	700	0,09
19	guilt	92	700	0,13
20	despair	123	700	0,18
21	envy	82	700	0,12
22	pride	63	700	0,09
23	shame	90	700	0,13

Çizelge 4.6. (Devamı)

24	submission	50	700	0,07
25	contempt	174	700	0,25
26	dominance	31	700	0,04
27	bittersweetness	83	700	0,12
28	ambivalence	49	700	0,07
29	frozenness	55	700	0,08
30	confusion	118	700	0,17
31	anger	323	700	0,46
32	anticipation	375	700	0,54
33	disgust	299	700	0,43
34	fear	461	700	0,66
35	joy	240	700	0,34
36	sadness	372	700	0,53
37	surprise	579	700	0,76
38	trust	550	700	0,79

Örneğin; *positive(olumlu)* duygu ile sınıflandırılması beklenen 10 farklı cümle için 70 farklı değerlendiriciden elde edilen 700 cümle-duygu ilişkisinin 660 tanesi *positive(olumlu)* olarak işaretlenmiştir. *Positive(olumlu)* duygu için veri setinde yer alan, insan değerlendiriciler tarafından duygu sınıflandırma işaretlemesi yapılmış 10 farklı cümlenin her birinin duygu yoğunluk oranını elde etmek için *positive(olumlu)* olarak işaretlenmiş duygu sayısının cümledeki tüm duygulara oranı alınarak elde edilmektedir. Her bir cümle için elde edilen duygu yoğunluk oranları her duygu için seçilmiş 10'ar cümlenin de değerlerinin ortalaması alınarak Çizelge 4.6. daki ortalama duygu yoğunlukları elde edilmiştir. Her bir cümle için elde edilen ortalama duygu yoğunluk değeri 0-1 arasında değer almaktadır. Örneğin; Çizelge 4.6. da gösterildiği gibi *Submission(teslimiyet)* duygusu için elde edilmiş duygu yoğunluk ortalama değeri 0.07 ve *Contempt(küçümseme)* duygusu için duygu yoğunluk ortalama değeri 0.25 olarak elde edilmiştir. İki duygu arasındaki değer farkı bir duygunun cümlede var olan diğer duygulara göre varlığının ne derecede olduğunu göstermektedir. Yüksek değere sahip duyguların ilgili kişi veya cümlede daha baskın olduklarını ifade etmektedir. Buradaki duygu yoğunluk ortalama değerleri duygunun tespitinin ölçüsü olarak değil cümledeki varlığının etkisi olarak ifade edilmektedir.

Çizelge 4.7. de insan değerlendiricilerle yapılan çalışmada kullanılan veri setindeki her bir duygu için kullanılan cümleler üzerinde herhangi bir değişiklik yapılmadan NAYALex Duygu Sözlüğü ile sınıflandırılarak ortalama duygu yoğunluk oranları elde edilmiştir.

Çizelge 4.7. NAYALex ortalama duygu yoğunluk oranları (0.0 -1.0 aralığında)

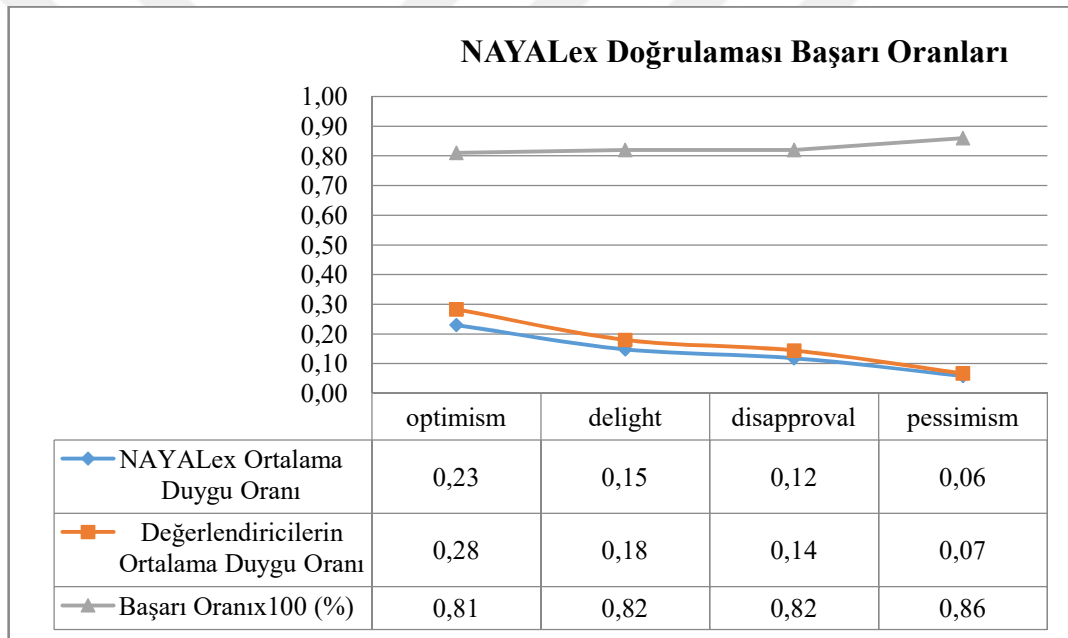
#	Duygu	NAYALex Ortalama Duygu Yoğunluğu	#	Duygu	NAYALex Ortalama Duygu Yoğunluğu
1	positive	0,98	20	despair	0,25
2	negative	0,97	21	envy	0,15
3	optimism	0,23	22	pride	0,05
4	delight	0,15	23	shame	0,14
5	disapproval	0,12	24	submission	0,04
6	pessimism	0,06	25	contempt	0,24
7	hope	0,14	26	dominance	0,05
8	unbelief	0,05	27	bittersweetness	0,20
9	curiosity	0,09	28	ambivalence	0,06
10	cynicism	0,04	29	frozenness	0,15
11	anxiety	0,14	30	confusion	0,10
12	awe	0,10	31	anger	0,50
13	outrage	0,10	32	anticipation	0,52
14	aggressiveness	0,07	33	disgust	0,50
15	love	0,25	34	fear	0,75
16	sentimentality	0,05	35	joy	0,50
17	remorse	0,22	36	sadness	0,55
18	morbidity	0,04	37	surprise	0,80
19	guilt	0,08	38	trust	0,87

Çizelge 4.7. de gösterilen ortalama duygu yoğunluk değerlerini elde etmek için, veri setindeki her bir duygunun sınıflandırılması için kullanılan 10'ar farklı cümlenin NAYALex Duygu Sözlüğü ile elde edilen duygu yoğunluk değerlerinin aritmetiksel ortalaması alınarak NAYALex ortalama duygu yoğunlukları elde edilmiştir.

Sözlüğün doğrulamasında kullanılan verinin güvenilirliğinin artırılması adına NAYALex ve insan değerlendiriciler için ortalama duygu yoğunlukları benzer

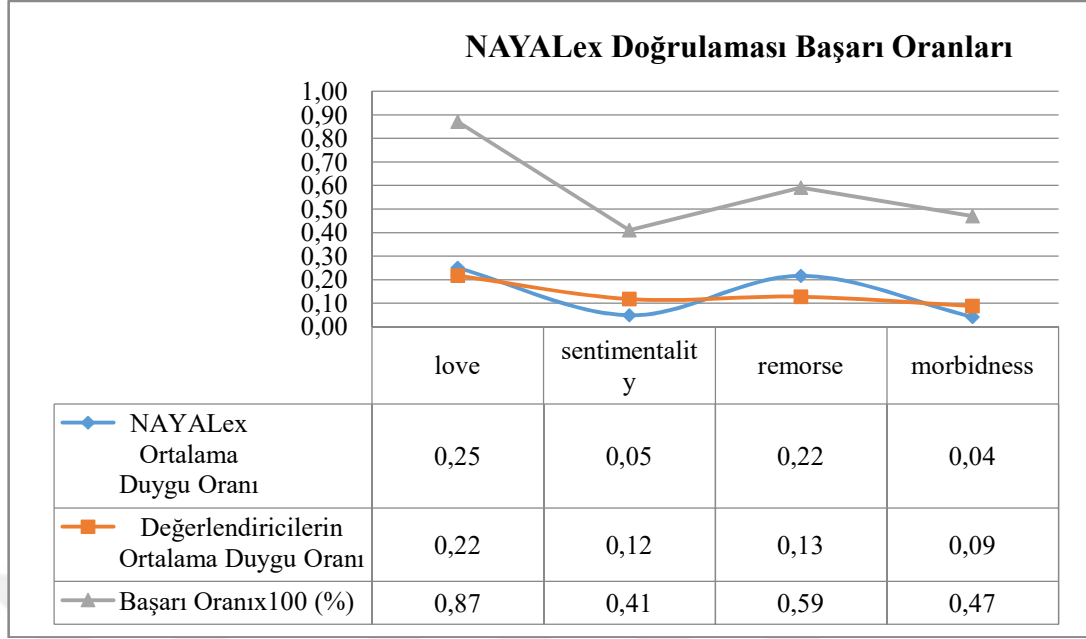
şekilde elde edilmiştir. NAYALex Duygu Sözlüğünün doğrulanması için elde edilen oranlar bu bölümün geri kalanında grafiklerle anlatılmaya çalışılmıştır. Duygular, NAYALex'in adım adım değerlendirilmesine imkan vermek için Plutchik'in duygu çarkındaki birincil, ikincil, üçüncül ve zıt duygu ikilileri dikkate alınarak sekiz farklı grupta ele alınacaktır(Plutchik, 1980; Plutchik, 2001).

Şekil 4.5. te birinci duygu grubu olarak ele aldığımız duygular için NAYALex başarı oranları *Optimism*(iyimserlik), *Delight*(kader), *Disapproval*(onaylamama) ve *Pessimism*(karamsarlık) duyguları için %81 ve üzeri olarak görülmüş ve her dört duygu için benzer sonuçlar elde edilmiştir.



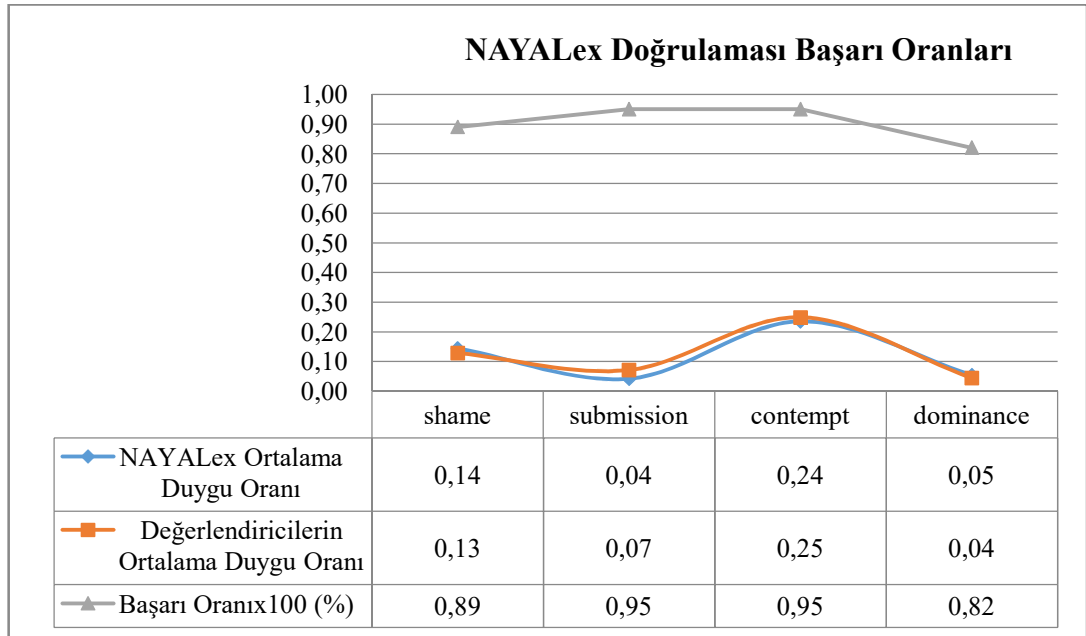
Şekil 4.5. NAYALex 1. duygu grubu doğrulama başarı oranları

Şekil 4.6. da ikinci duygu grubu olarak ele aldığımız duygular için NAYALex başarı oranları *Love*(sevgi), *Sentimentality*(duygusallık), *Remorse*(pişmanlık) ve *Morbidness*(kötülük) duyguları için farklı değerler göstermiş ve bu grup içinde en yüksek başarı oranına sahip duygu %87 ile *Love*(sevgi) duygusu olarak gözlemlenmiştir.



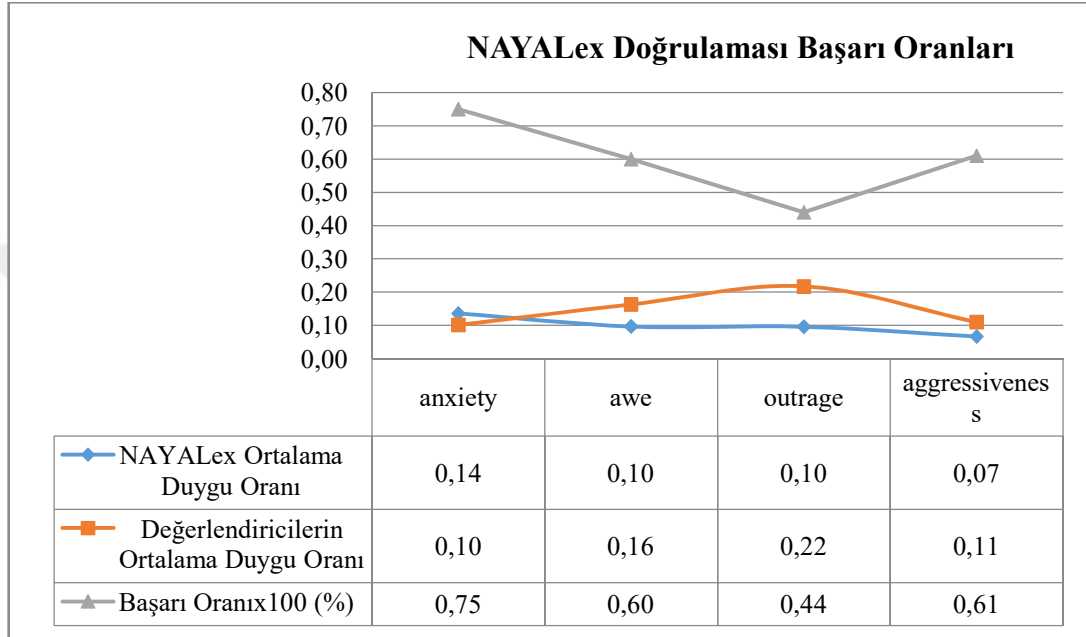
Şekil 4.6. NAYALex 2. duygu grubu doğrulama başarı oranları

Şekil 4.7. de üçüncü duygu grubu olarak ele aldığımız duygular için NAYALex başarı oranları *Shame(utanç)*, *Submission(teslimiyet)*, *Contempt(küçümseme)* ve *Dominance(hakimiyet)* duyguları için %82 - %95 arasında sonuçlar elde edilmiştir.



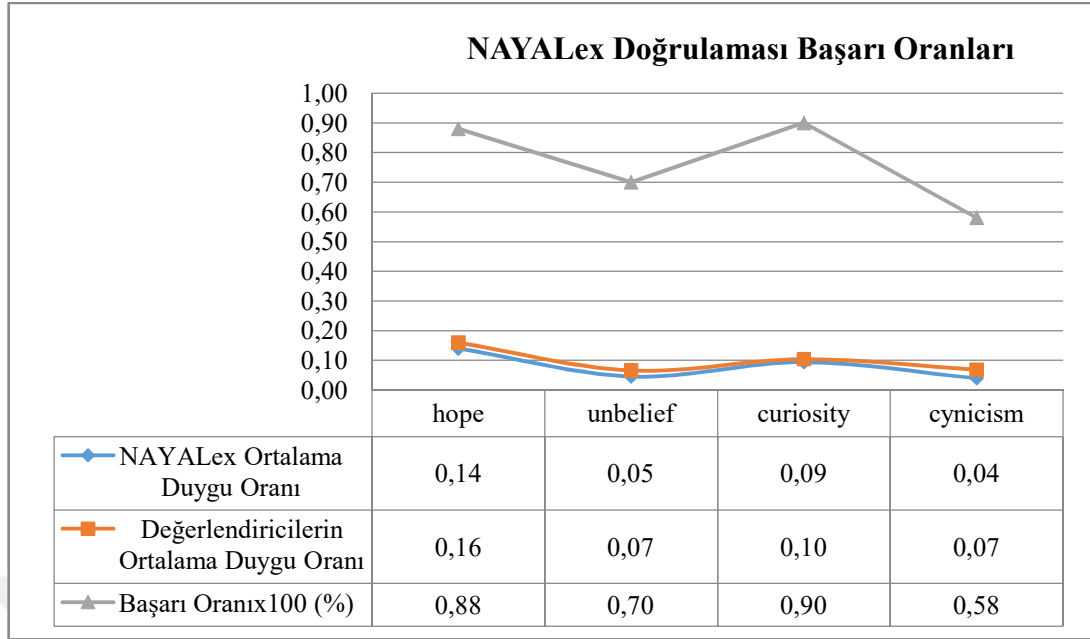
Şekil 4.7. NAYALex 3. duygu grubu doğrulama başarı oranları

Şekil 4.8. de dördüncü duygu grubu olarak ele aldığımız duygular için NAYALex başarı oranları *Anxiety(kaygı)*, *Awe(huşu)*, *Outrage(öfke-nefret)* ve *Aggressiveness(saldırganlık)* duyguları için %44 - %75 arasında sonuçlar elde edilmiştir ve dördüncü grupta en yüksek başarı oranına sahip duygu %75 ile *Anxiety(kaygı) duygusu olmuştur.*



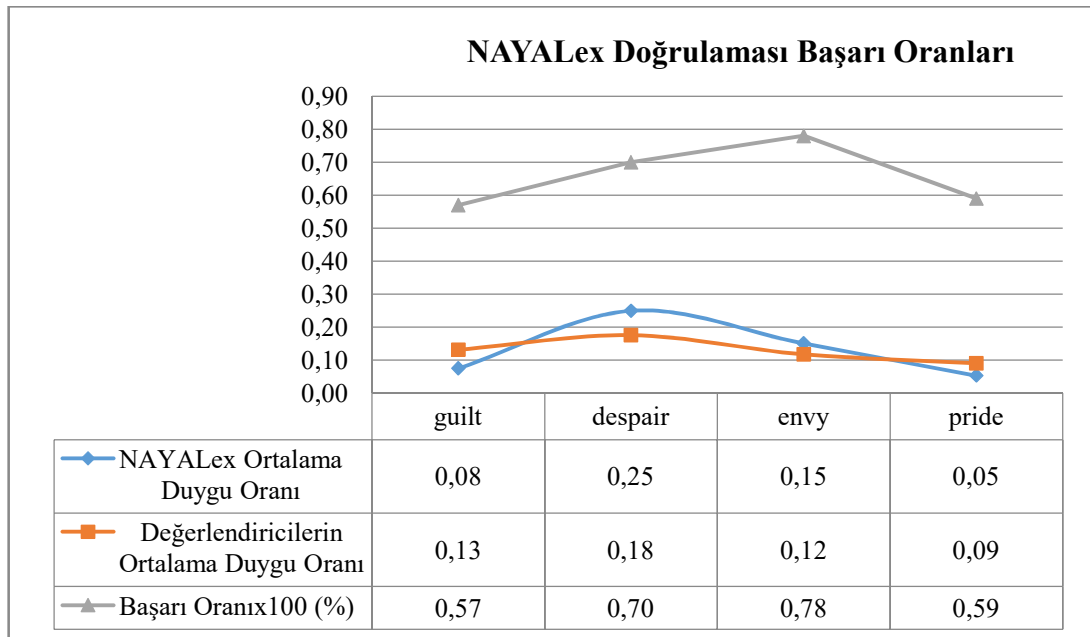
Şekil 4.8. NAYALex 4. duygu grubu doğrulama başarı oranları

Şekil 4.9. da beşinci duygu grubu olarak ele aldığımız duygular için NAYALex başarı oranları *Hope(umut)*, *Unbelief(inançsızlık)*, *Curiosity(merak)*, ve *Cynicism(alaycılık)* duyguları için %58 - %90 arasında sonuçlar elde edilmiştir ve beşinci grupta en yüksek başarı oranına sahip duygu %90 ile *Curiosity(merak) duygusu olmuştur.*



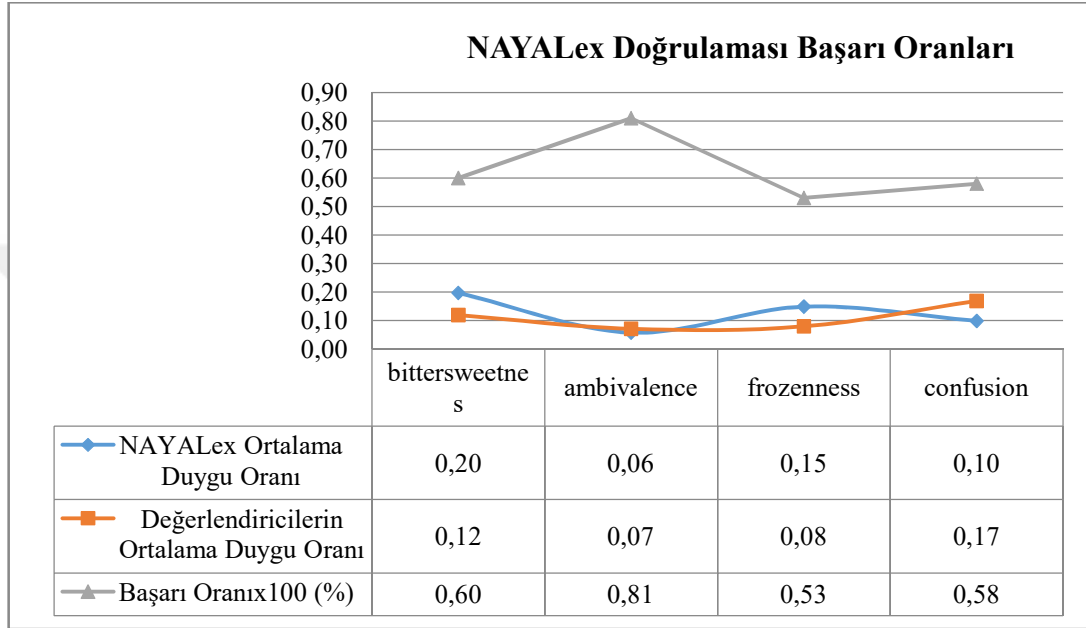
Şekil 4.9. NAYALex 5. duygu grubu doğrulama başarı oranları

Şekil 4.10. da altıncı duygu grubu olarak ele aldığımız duygular için NAYALex başarı oranları *Guilt(suçluluk)*, *Despair(umutsuzluk)*, *Envy(kıskançlık)*, ve *Pride(gurur)* duyguları için %57 - %78 arasında sonuçlar elde edilmiştir ve altıncı grupta en yüksek başarı oranına sahip duygu %78 ile *Envy(kıskançlık)* duygusu olmuştur.



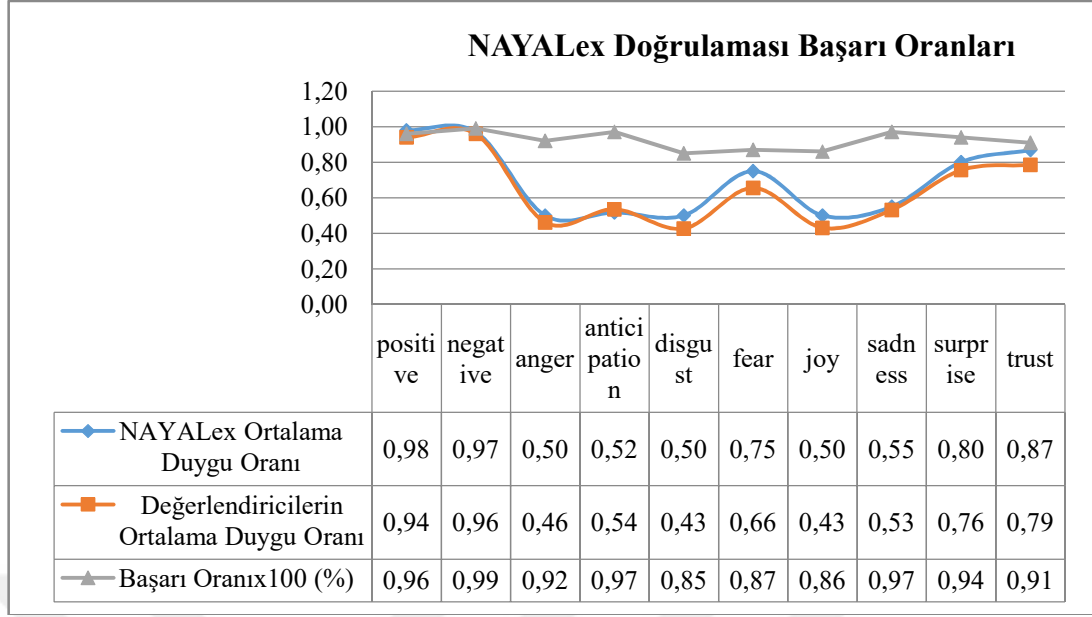
Şekil 4.10. NAYALex 6. duygu grubu doğrulama başarı oranları

Şekil 4.11. de yedinci duygu grubu olarak ele aldığımız duygular için NAYALex başarı oranları *Bittersweetness*(acı tatlılık), *Ambivalence*(kararsızlık), *Frozensness*(donukluk), ve *Confusion*(bilinç bulanıklığı) duyguları için %53 - %81 arasında sonuçlar elde edilmiştir ve yedinci grupta en yüksek başarı oranına sahip duygu %81 ile *Ambivalence*(kararsızlık) duygusu olmuştur.



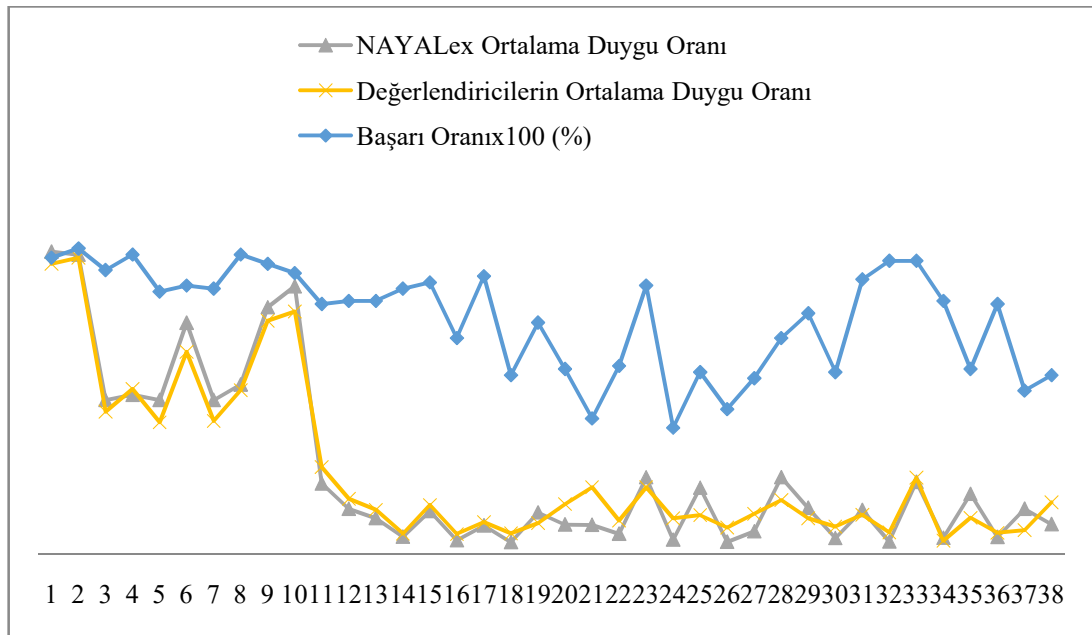
Şekil 4.11. NAYALex 7. duygu grubu doğrulama başarı oranları

Şekil 4.12. de sekizinci duygu grubu Plutchik'in duygu çarkının temelini oluşturmaktadır. NRC duygu sözlüğü tarafından da sınıflandırılabilen *Positive*(olumlu), *Negative*(olumsuz), *Anger*(öfke), *Sadness*(üzüntü), *Joy*(sevinç), *Fear*(korku), *Surprise*(sürpriz), *Disgust*(tikinti), *Trust*(güven), *Anticipation*(beklenti) ve *Love* (sevgi) duygularının NAYALex doğrulama başarı oranları %85 - %99 arasında sonuçlar elde edilmiştir ve yedinci gruptaki tüm duygular için benzer yüksek oranlar elde edilmiştir.



Şekil 4.12. NAYALex 8. duygu grubu doğrulama başarı oranları

NAYALex Duygu Sözlüğü sınıflandırabildiği tüm duygular için insan değerlendiricilerden elde edilen doğrulama veri seti üzerinde çalıştırılarak duygu yoğunluk oranları elde edilmiştir. NAYALex Duygu Sözlüğünden ve insan değerlendiricilerden elde edilen duygu yoğunluk oranları Şekil 4.13. te tüm duygular için karşılaştırılarak genel başarı oranı izlenmiştir.



Şekil 4.13. Tüm duygular için nayalex başarı oranları

Değerlendiricilerden elde edilen oranlar ile NAYALex'ten elde edilen oranlar birbirinden uzaklaştıkça başarı oranları düşmekte, oranlar birbirine yaklaştıkça da başarı oranları yükselmektedir. NAYALex tarafından sınıflandırılabilen 38 duygu Şekil 4.13.teki başarı oranları eğrilerine göre incelendiğinde tatmin edici sonuçlar verdiği ve NAYALex'ten elde edilen sınıflandırma oranı eğrileri ile insan değerlendiricilerden elde edilen sınıflandırma oranı eğrileri arasında paralel bir benzerlik olduğu görülmektedir.

4.4. Sınıflandırma Algoritmaları İle Değerlendirme

NAYALex sözlüğünün doğrulanması için kullandığımız doğrulama veri seti için Navie Bayes, Destek Vektör Makinesi(SVM) ve KNN(K-Nearest Neighbors) sınıflandırma algoritmaları kullanılarak doğruluk, duyarlılık, kesinlik ve F-Ölçütü değerleri hesaplanmıştır. Bu değerler ile sözlüğün doğrulanması için kullandığımız doğrulama veri setinin kullanılabilirliği, performansı ve geçerliği ile ortaya konulan sözlüğümüzün 38 farklı her bir duygusu için anlamlı sınıflandırmalar yaptığı açıklanmaya çalışılmıştır. Üç farklı sınıflandırma algoritması için aynı veri seti kullanılmıştır. Sınıflandırma algoritmaları için kullanılan eğitim ve test veri setleri doğrulama veri setinden random olarak seçilmiştir.

Navie Bayes sınıflandırma algoritması doğrulama veri seti üzerinde çalıştırılarak algoritmanın başarı oranları doğruluk, duyarlılık, kesinlik ve f-ölçütü değerleri Çizelge 4.8. deki gibi elde edilmiştir. Buna göre doğrulama veri seti için NAYALex'in sınıflandırabildiği 38 duyguya göre elde edilen değerler duyguların tamamına yakını için 1.00 değeri alarak yüksek performans göstermiştir. Doğrulama veri seti üzerinde çalıştırılan Navie Bayes için Doğruluk değeri 0.92 olmuştur. Aynı şekilde ortalama kesinlik, duyarlılık ve f1-score değerleri 0.93, 0.92 ve 0.91 olmuştur.

Çizelge 4.8. Navi bayes sınıflandırma algoritması sonuçları

	recision(kesinlik)	recall(duyarlılık)	f1-score
aggressiveness	1.00	1.00	1.00
ambivalence	0.50	0.60	0.55
anger	1.00	1.00	1.00
anticipation	1.00	1.00	1.00
anxiety	1.00	1.00	1.00
awe	1.00	1.00	1.00
bittersweetness	1.00	1.00	1.00
confusion	1.00	1.00	1.00
contempt	0.60	0.60	0.60
curiosity	1.00	1.00	1.00
cynicism	1.00	1.00	1.00
delight	1.00	1.00	1.00
despair	1.00	1.00	1.00
disapproval	1.00	1.00	1.00
disgust	1.00	1.00	1.00
dominance	0.50	1.00	0.67
envy	1.00	1.00	1.00
fear	1.00	1.00	1.00
frozenness	1.00	1.00	1.00
guilt	1.00	1.00	1.00
hope	1.00	1.00	1.00
joy	1.00	1.00	1.00
love	1.00	1.00	1.00
morbidness	1.00	1.00	1.00
negative	1.00	1.00	1.00
optimism	1.00	1.00	1.00
outrage	1.00	1.00	1.00
pessimism	1.00	1.00	1.00
positive	1.00	1.00	1.00
pride	0.60	0.60	0.60
remorse	1.00	0.40	0.57
sadness	1.00	1.00	1.00
sentimentality	0.62	1.00	0.77
shame	1.00	0.40	0.55
submission	0.67	0.40	0.50
surprise	1.00	1.00	1.00
trust	1.00	1.00	1.00
unbelief	1.00	1.00	1.00
accuracy(doğruluk)	0.92		
macro avg	0.93	0.92	0.91
weighted avg	0.93	0.92	0.91

Destek Vektör Makinesi(SVM) sınıflandırma algoritması doğrulama veri seti üzerinde çalıştırılarak algoritmanın başarımları oranları doğruluk, duyarlılık, kesinlik ve f-ölçütü değerleri Çizelge 4.9. daki gibi elde edilmiştir. Doğrulama veri seti üzerinde çalıştırılan Destek Vektör Makinesi(SVM) için doğruluk değeri 0.93 olmuştur.

Çizelge 4.9. Destek vektör makinesi(SVM) sınıflandırma algoritması sonuçları

	precision(kesinlik)	recall(duyarlılık)	f1-score
aggressiveness	1.00	1.00	1.00
ambivalence	0.67	0.40	0.50
anger	1.00	1.00	1.00
anticipation	1.00	1.00	1.00
anxiety	1.00	1.00	1.00
awe	1.00	1.00	1.00
bittersweetness	1.00	1.00	1.00
confusion	1.00	1.00	1.00
contempt	0.67	0.80	0.73
curiosity	1.00	1.00	1.00
cynicism	1.00	1.00	1.00
delight	1.00	1.00	1.00
despair	1.00	1.00	1.00
disapproval	1.00	1.00	1.00
disgust	1.00	1.00	1.00
dominance	0.56	1.00	0.71
envy	1.00	1.00	1.00
fear	1.00	1.00	1.00
frozensness	1.00	1.00	1.00
guilt	1.00	1.00	1.00
hope	1.00	1.00	1.00
joy	1.00	1.00	1.00
love	1.00	1.00	1.00
morbidity	1.00	1.00	1.00
negative	1.00	1.00	1.00
optimism	1.00	1.00	1.00
outrage	1.00	1.00	1.00
pessimism	1.00	1.00	1.00
positive	1.00	1.00	1.00
pride	0.60	0.60	0.60
remorse	1.00	0.40	0.57
sadness	1.00	1.00	1.00
sentimentality	1.00	0.80	0.89
shame	0.50	0.80	0.62
submission	0.67	0.40	0.50
surprise	1.00	1.00	1.00
trust	1.00	1.00	1.00
unbelief	1.00	1.00	1.00
accuracy(doğruluk)	0.93		
macro avg	0.94	0.93	0.92
weighted avg	0.94	0.93	0.92

KNN(K-Nearest Neighbors) sınıflandırma algoritması doğrulama veri seti üzerinde çalıştırılarak algoritmanın başarımları oranları doğruluk, duyarlılık, kesinlik ve f-ölçütü değerleri Çizelge 4.9. daki gibi elde edilmiştir. Buna göre doğrulama veri seti için elde edilen değerler 1.00 değeri olarak yüksek performans göstermiştir.

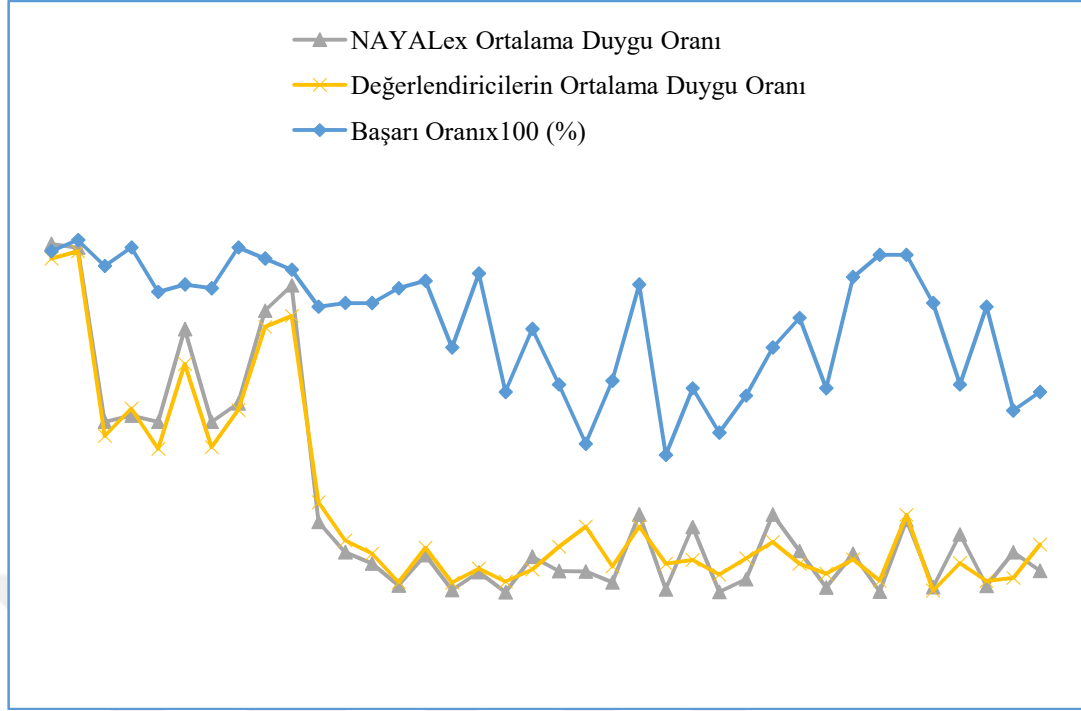
Çizelge 4.10. KNN(K-Nearest Neighbors) sınıflandırma algoritması sonuçları

	recision(kesinlik)	recall(duyarlılık)	f1-score
aggressiveness	1.00	1.00	1.00
ambivalence	0.48	1.00	0.56
anger	1.00	1.00	1.00
anticipation	1.00	1.00	1.00
anxiety	1.00	1.00	1.00
awe	1.00	1.00	1.00
bittersweetness	1.00	1.00	1.00
confusion	1.00	1.00	1.00
contempt	0.56	1.00	0.71
curiosity	1.00	1.00	1.00
cynicism	1.00	1.00	1.00
delight	1.00	1.00	1.00
despair	1.00	1.00	1.00
disapproval	1.00	1.00	1.00
disgust	1.00	1.00	1.00
dominance	0.56	1.00	0.71
envy	1.00	1.00	1.00
fear	1.00	1.00	1.00
frozenness	1.00	1.00	1.00
guilt	1.00	1.00	1.00
hope	1.00	1.00	1.00
joy	1.00	1.00	1.00
love	1.00	1.00	1.00
morbidness	1.00	1.00	1.00
negative	1.00	1.00	1.00
optimism	1.00	1.00	1.00
outrage	1.00	1.00	1.00
pessimism	1.00	1.00	1.00
positive	1.00	1.00	1.00
pride	0.50	0.40	0.44
remorse	1.00	0.40	0.57
sadness	1.00	1.00	1.00
sentimentality	1.00	0.60	0.75
shame	0.60	0.50	0.44
submission	0.50	0.52	0.60
surprise	1.00	1.00	1.00
trust	1.00	1.00	1.00
unbelief	1.00	1.00	1.00
accuracy(doğruluk)	0.91		
macro avg	0.89	0.91	0.89
weighted avg	0.89	0.91	0.89

5. SONUÇLAR ve ÖNERİLER

Duygu analizinde sözlük tabanlı yaklaşımları kullanarak yapılan çalışmaların “sınırlı sayıda duygu kullanımı, kişiliğin belirlenmesinde yetersiz kalabilmektedir” (Atlı ve İlhan, 2021). Çalışmamızda önerdiğimiz NAYALex Duygu Sözlüğü, sınırlı sayıdaki duygu ile kişiliğin belirlenmesi problemini iyileştirerek daha fazla sayıda duygu çıkarımına olanak sunmaktadır. Kişiliğin belirlenmesinde daha doğru sonuçlar elde etmek için metinsel ifadelerin daha geniş spektrumda duygularla etiketlenmesini zorunlu kılmaktadır. Bu çalışmada, daha fazla sayıdaki duygu ile kişiliğin belirlenmesine katkı sunacak olan “NAYALex Duygu Sözlüğü, NRC duygu Sözlüğüne (Mohammad, 2016) ve Plutchik’in Temel Duyguların Psikoevrimsel Teorisine (Drews and Krohn, 2007) dayandırılarak oluşturulmuştur” (Atlı ve İlhan, 2021). Oluşturulan NAYALex sözlüğü Instagramdan elde edilmiş veri seti üzerinde çalıştırılarak ve bir takım analizler yapılarak sözlüğün uygulanabilir olduğu ispatlanmıştır.

“Sözlük tabanlı yaklaşımlarda, manüel olarak oluşturulan sözlüklerin yüksek maliyetinden kaçınmak için sözlüklerin otomatik bir şekilde oluşturulması önemlidir” (Li ve ark., 2012; Atlı ve İlhan, 2021; Hatzivassiloglou and McKeown,1997). “NAYALex duygu sözlüğü, NRC duygu Sözlüğü (Mohammad, 2016) ve Plutchik’in Temel Duyguların Psikoevrimsel Teorisi (Drews and Krohn, 2007) kullanılarak otomatik olarak oluşturulmuş olup, yüksek maliyet ortadan kaldırılmıştır” (Atlı ve İlhan, 2021). Otomatik olarak oluşturulan NAYALex Duygu Sözlüğünü doğrulamak ve güvenilirliğini arttırmak için insan değerlendiricilerle ayrıca bir çalışma yapılmıştır. Çalışmada doğrulama veri setinde bulunan 380 cümle, 70 farklı insan değerlendirici tarafından 38 farklı NAYALex duygusu için duygu işaretlemesi yapılmıştır. NAYALex’in doğrulanmasına imkân vermek için insan değerlendiricilerle kullanılan aynı veri seti üzerinde NAYALex Duygu Sözlüğü de çalıştırılarak duygu sınıflandırması yapılmıştır. Şekil 5.1. de 38 farklı duygu için NAYALex’in insan değerlendiricilere göre başarı eğrisi verilmiştir. Şekil 5.1. incelendiğinde NAYALex’in 38 duygu için %77 olan ortalama başarı eğrisine yakın ve üzerinde başarı oranları gösterdiği görülmektedir.



Şekil 5.1. NAYALex doğrulama başarı oranları

Çalışmada kullanılan doğrulama veri seti NAYALex'in sınıflandırdığı duygular bakımından üç farklı sınıflandırma algoritmasına tabi tutularak performans değerlendirme ölçütleri elde edilmiştir. Çizelge 5.1 de üç farklı sınıflandırma algoritmasına göre doğruluk, duyarlılık, kesinlik ve f-ölçütü değerleri karşılaştırılmıştır.

Çizelge 5.1. Algoritmalar için performans değerlendirme ölçütlerinin karşılaştırılması

Sınıflandırma Algoritması	Performans Değerlendirme Ölçütleri			
	accuracy (doğruluk)	precision (kesinlik)	recall (duyarlılık)	f1-score
Navie Bayes	0.92	0.93	0.92	0.91
Destek Vektör Makinesi(SVM)	0.93	0.94	0.93	0.92
KNN(K-Nearest Neighbors)	0.91	0.89	0.91	0.89

NAYALex Duygu Sözlüğünün doğrulanması için kullanılan doğrulama veri seti, ayrıca makine öğrenmesine dayalı sınıflandırma algoritmaları için kullanılarak

değerlendirme ölçütleri elde edilmiştir. Bu değerlendirme ölçütleri dikkate alındığında kullanılan sınıflandırma algoritmalarının tamamı için yüksek oranlar elde edilmiştir. Kullanılan makine öğrenmesi algoritmalarından doğruluk, kesinlik, duyarlılık ve f1-score değerleri için en yüksek başarı oranlarına sahip algoritma Destek Vektör Makinesi(SVM) olduğu görülmüştür.

Bu çalışmamızdan önce “dünya çapında 154 farklı duygu olduğunu ifade eden Tiffany (Medhat ve ark., 2014; Atlı ve İlhan, 2021)’nin belirttiği duygular” dikkate alındığında bu duygular arasında en fazla sayıda duygu sınıflandırması yapabilen sözlük 8 duygu ile NRC(emolex) olarak bilinmekteydi. Bu çalışmamız ile birlikte önerdiğimiz NAYALex sözlüğüyle “Tiffany’nin belirttiği duygular dikkate alındığında; NAYALex bu duyguların %24,7’sini, NRC %6,5’ini, LIWC %2,6’sını, EmoSenticNet %3,9’unu ve Empath %3,9’unu oransal olarak temsil ettiği ve doğrudan sınıflandırabildiği Çizelge 5.2. de gösterilmiştir. Bildiğimiz kadarıyla bu çalışma, sözlük tabanlı duygu sınıflandırması yapan çalışmalar arasında, en yakın rakibi olan NRC(Emolex)’ye göre dört kat daha fazla ve %24,7 lik temsil oranıyla doğrudan 38 farklı duyguyu ortaya çıkarabilen ilk en kapsamlı sözlük olmuştur” (Atlı ve İlhan, 2021).

Çizelge 5.2. NAYALex ve duygu sözlüklerinin sınıflandırma oranlarının karşılaştırılması

#	NAYALex	NRC (Emolex)	EmoSentic-Net	Empath	LIWC
Sözlüğün Doğrudan Sınıflandırabildiği Duygu Sayısı	38	8	6	6	4
Tiffany’nin Belirttiği 154 Duyguyu Temsil Etme Oranı	%24,70	%6,50	%3,90	%3,90	%2,60

Oluşturduğumuz NAYALex Duygu Sözlüğü doğrudan bir çok alanda kullanıcı verileri üzerinde uygulanabilir. Duygu analiz sistemlerine katkı sunacak olan NAYALex müşteri memnuniyetinin değerlendirilmesi, sosyal medyada kriz yönetimi vb. alanlara geniş duygu kapasitesiyle katkı sunabilir. NAYALex bu

aşamda otomatik bir şekilde anlamlı bilgi çıkarımı yaparak duyguları 38 farklı sınıfta belirlemekte ve duygu analizine geniş bir alan sunmaktadır.

Gelecekteki çalışmamızda sözlüğümüz üzerinde farklı sınıflandırma algoritmaları kullanarak yeni modeller oluşturmaya odaklanacağız.



KAYNAKLAR

- ALARCAO, S. M., and FONSECA, M. J., 2017. Emotions recognition using EEG signals: A survey. *IEEE Transactions on Affective Computing*, 10(3): 374-393.
- ALARİD, M., 2016. Recruitment and radicalization: The role of social media and new technology. *Impunity: Countering illicit power in war and transition* 313-330.
- ATLI, Y. ve İLHAN, N., 2021. Duygu Analizi İçin Yeni Bir Sözlük; NAYALex Duygu Sözlüğü . *Avrupa Bilim ve Teknoloji Dergisi* , (27) , 1050-1060.
- AYTUĞ, O. N. A. N., 2017. Twitter mesajları üzerinde makine öğrenmesi yöntemlerine dayalı duygu analizi. *Yönetim Bilişim Sistemleri Dergisi*, 3(2): 1-14.
- BACCIANELLA, S., ESULİ, A., and SEBASTIANİ, F., 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *Lrec*, 10: 2200-2204.
- BAE, H. S., LEE, H. J., and LEE, S. G., 2016. Voice recognition based on adaptive MFCC and deep learning. In *2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA)* (pp. 1542-1546). IEEE.
- BRECK, E., CHOİ, Y., and CARDİE, C., 2007. Identifying expressions of opinion in context. In *IJCAI* (Vol. 7: pp. 2683-2688).
- CAMBRİA, E., SCHULLER, B., XİA, Y., and HAVASİ, C., 2013. New avenues in opinion mining and sentiment analysis. *IEEE Intelligent systems*, 28(2): 15-21.
- COLOMBETTİ, G., 2009. From affect programs to dynamical discrete emotions. *Philosophical Psychology*, 22(4): 407-425.
- DATAREPORTAL, 2022. Dünyada Dijital Ortam Kullanımı, erişim: 2 Mayıs 2022, <https://datareportal.com/>
- DENG, L., and WİEBE, J., 2015. Mpqa 3.0: An entity/event-level sentiment corpus. In *Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: human language technologies* (pp. 1323-1328).
- DEVİKA, M. D., SUNİTHA, C., and GANESH, A., 2016. Sentiment analysis: a comparative study on different approaches. *Procedia Computer Science*, 87: 44-49.
- DREWS, M., 2007. Robert Plutchik's Psychoevolutionary Theory Of Basic Emotions. (Erişim tarihi: 12.07.2021, <http://www.adliterate.com/archives/Plutchik.emotion.theorie.POSTER.pdf>)
- DREWS, M., and KROHN, M., 2007. Robert Plutchik's Psychoevolutionary theory of basic emotions. University of Applied Sciences Postdam, Germany. Retrieved from <http://www.markusdrews.de/Plutchiks.Emotionstheorie.PLAKAT.pdf>.
- EL HİNDİ, K., 2014. Fine tuning the Naïve Bayesian learning algorithm. *AI Communications*, 27(2): 133-141.
- ER, M. B., 2020. A Novel Approach for Classification of Speech Emotions Based on Deep and Acoustic Features. *IEEE Access*, 8: 221640-221653.

- ER, M. B., ve ÇİĞ, H., 2020. Türk Müziği Uyarınları Kullanılarak İnsan Duygularının Makine Öğrenmesi Yöntemi İle Tanınması. Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji, 8(2): 458-474.
- FAST, E., CHEN, B., and BERNSTEIN, M. S., 2016. Empath: Understanding topic signals in large-scale text. In Proceedings of the 2016 CHI conference on human factors in computing systems (pp. 4647-4657).
- GÜRAN, A., UYSAL, M., and DOĞRUSÖZ, Ö., 2014. Destek vektör makineleri parametre optimizasyonunun duygu analizi üzerindeki etkisi. Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi, 16(48): 86-93.
- HAN, J., and KAMBER, M., 2006. Data mining: concepts and techniques, 2nd. University of Illinois at Urbana Champaign: Morgan Kaufmann.
- HAND, D., MANNILA, H., and SMYTH, P., 2001. Principles of Data Mining". The MIT Press. In A comprehensive, highly technical look at the math and science behind extracting useful information from large databases (Vol. 546).
- HATZIVASSILOGLU, V., and MCKEOWN, K., 1997. Predicting the semantic orientation of adjectives. In 35th annual meeting of the association for computational linguistics and 8th conference of the european chapter of the association for computational linguistics (pp. 174-181).
- HIDALGO, C. R., TAN, E. S. H., and VERLEGH, P. W., 2015. The social sharing of emotion (SSE) in online social networks: A case study in Live Journal. Computers in Human Behavior, 52: 364-372.
- HUSSEIN, D. M. E. D. M., 2018. A survey on sentiment analysis challenges. Journal of King Saud University-Engineering Sciences, 30(4): 330-338.
- KAMPS, J., MARX, M., MOKKEN, R. J., and DE RIJKE, M., 2004. Using WordNet to measure semantic orientations of adjectives. In LREC (Vol. 4: 1115-1118).
- KOTO, F., and ADRIANI, M., 2015. HBE: Hashtag-based emotion lexicons for twitter sentiment analysis. In Proceedings of the 7th Forum for Information Retrieval Evaluation (pp. 31-34).
- KOTZIAS, D., DENIL, M., DE FREITAS, N., and SMYTH, P., 2015. From group to individual labels using deep features. In Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 597-606).
- KOUMPOURİ, A., MPORAS, I., and MEGALOOIKONOMOU, V., 2015. Evaluation of Four Approaches for " Sentiment Analysis on Movie Reviews" The Kaggle Competition. In Proceedings of the 16th International Conference on Engineering Applications of Neural Networks (INNS) (pp. 1-5).
- KUŠEN, E., CASCAVILLA, G., FİGL, K., CONTİ, M., and STREMBECK, M., 2017. Identifying emotions in social media: comparison of word-emotion lexicons. In 2017 5th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW) (pp. 132-137). IEEE.
- LANG, P. J., 1995. The emotion probe: Studies of motivation and attention. American psychologist, 50(5): 372.
- LAROSE, D. T., and LAROSE, C. D., 2014. Discovering knowledge in data: an introduction to data mining (Vol. 4). John Wiley and Sons.

- LǪ, F., PAN, S. J., JĪN, O., YANG, Q., and ZHU, X., 2012. Cross-domain co-Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 410-419).
- LĪN, C. K., LEE, Y. Y., Yu, C. H., and CHEN, H. H., 2014. Exploring ensemble of models in taxonomy-based cross-domain sentiment classification. In Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management (pp. 1279-1288).
- LĪU, B., 2012. Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1): 1-167.
- MALĪK, K. R., AHMAD, M., KHALĪD, S., AHMAD, H., AL-TURJMAN, F., and JABBAR, S., 2020. Image and command hybrid model for vehicle control using Internet of Vehicles. Transactions on Emerging Telecommunications Technologies, 31(5): e3774.
- MEDHAT, W., HASSAN, A., and KORASHY, H., 2014. Sentiment analysis algorithms and applications: A survey. Ain Shams engineering journal, 5(4): 1093-1113.
- MĪTTAL, S., AGARWAL, S., and NĪGAM, M. J., 2018. Real time multiple face recognition: A deep learning approach. In Proceedings of the 2018 International Conference on Digital Medicine and Image Processing (pp. 70-76).
- MOHAMMAD, S., 2011. Sentiment and Emotion Lexicons. (Eriřim tarihi: 12.07.2021, <http://saifmohammad.com/WebPages/lexicons.html>).
- MOHAMMAD, S. M., ZHU, X., KĪRĪTCHENKO, S., and MARTĪN, J., 2015. Sentiment, emotion, purpose, and style in electoral tweets. Information Processing and Management, 51(4): 480-499.
- MOHAMMAD, S., 2016. NRC Word-Emotion Association Lexicon. (Eriřim tarihi: 12.07.2021, <http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.html>).
- MOHAMMAD, S., and KĪRĪTCHENKO, S., 2013. Using nuances of emotion to identify personality. In Seventh International AAAI Conference on Weblogs and Social Media.
- PANG, B., and LEE, L., 2009. Opinion mining and sentiment analysis. Comput. Linguist, 35(2): 311-312.
- PANG, B., LEE, L., and VAĪTHYANATHAN, S., 2002. Thumbs up? Sentiment classification using machine learning techniques. arXiv preprint cs/0205070.
- PENNEBAKER, J. W., FRANCĪS, M. E., and BOOTH, R. J., 2001. Linguistic inquiry and word count: LIWC 2001. Mahway: Lawrence Erlbaum Associates, 71: 2001.
- PĪCARD, R. W., 2000. Affective computing. MIT press.
- PLUTCHĪK, R., 1980. A general psychoevolutionary theory of emotion. In Theories of emotion (pp. 3-33). Academic press.
- PLUTCHĪK, R., 2001. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. American scientist, 89(4): 344-350.
- PORĪA, S., GELBUKH, A., CAMBRĪA, E., YANG, P., HUSSAĪN, A., and DURRANĪ, T., 2012. Merging SenticNet and WordNet-Affect emotion lists for sentiment analysis. In 2012 IEEE 11th international conference on signal processing (Vol. 2: 1251-1255). IEEE.

- POWERS, D. M., 2020. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv preprint arXiv: 2010.16061.
- RAO, D., and RAVICHANDRAN, D., 2009. Semi-supervised polarity lexicon induction. In Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009): (pp. 675-682).
- RAVI, K., and RAVI, V., 2015. A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowledge-based systems, 89: 14-46.
- SONG, K., GAO, W., CHEN, L., FENG, S., WANG, D., and ZHANG, C., 2016. Build emotion lexicon from the mood of crowd via topic-assisted joint non-negative matrix factorization. In Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval (pp. 773-776).
- STAIANO, J., and GUERINI, M., 2014. Depechemood: a lexicon for emotion analysis from crowd-annotated news. arXiv preprint arXiv:1405.1605.
- TURNEY, P. D., 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. arXiv preprint cs/0212032.
- WITTEN, I. H., FRANK, E., HALL, M. A., PAL, C. J., and DATA, M., 2005. Practical machine learning tools and techniques. In Data Mining (Vol. 2: No. 4).
- WOOLF, N., 2016. As fake news takes over Facebook feeds, many are taking satire as fact. The Guardian. Accessed, 1: 04-18.
- YESSENALINA, A., YUE, Y., and CARDIE, C., 2010. Multi-level structured models for document-level sentiment classification. In Proceedings of the 2010 conference on empirical methods in natural language processing (pp. 1046-1056).
- ZHANG, S., 2010. KNN-CF approach: Incorporating certainty factor to knn classification. IEEE Intell. Informatics Bull., 11(1): 24-33.