



Türkçe Metinlerde Duygu Analizi İçin Farklı Makine Öğrenmesi Yöntemlerinin Karşılaştırılması

Comparison of Different Machine Learning Approaches for Emotion Analysis in Turkish

Mansur Alp Toçoğlu^{1*}, Azer Çelikten², İrfan Aygün³, Adil Alpkoçak⁴

^{1,2,3} Manisa Celal Bayar Üniversitesi, Hasan Ferdi Turgutlu Teknoloji Fakültesi, Yazılım Mühendisliği Bölümü,
Turgutlu, Manisa, TÜRKİYE

⁴ Dokuz Eylül Üniversitesi Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Buca, İzmir, TÜRKİYE

Sorumlu Yazar / Corresponding Author *: alpkocak@ceng.deu.edu.tr

Geliş Tarihi / Received: 21.02.2019

DOI:10.21205/deufmd.2019216303

Kabul Tarihi / Accepted: 08.03.2019

Araştırma Makalesi/Research Article

Atıf şekli/ How to cite: TOÇOĞLU, M., A., ÇELİKTEN, A., AYGÜN, İ., ALPKOÇAK, A. (2019). Türkçe Metinlerde Duygu Analizi İçin Farklı Makine Öğrenmesi Yöntemlerinin Karşılaştırılması. DEUFMD, 21(63), 719-725.

Öz

Bu çalışmada, Türkçe metinlerden duygu çıkarımı alanında kullanılan TREMO veri seti üzerinde farklı makine öğrenmesi algoritmalarının sınıflandırma sonuçları karşılaştırılmıştır. Duygu analizi bir metin sınıflandırma problemi olarak ele alınmış ve Yapay Sinir Ağları (YSA), Destek Vektör Makineleri (DVM), Random Forest (RF) ve K-En Yakın Komşu (KEYK) algoritmaları olmak üzere dört yaklaşım incelenmiştir. İncelenen duygu kategorileri olarak veri setinin sağladığı, mutluluk, korku, öfke, üzüntü, tıksınme ve şaşırma kategorileri kullanılmıştır. Veri ön işleme bölümünde, veri setini oluşturan kelimelerin kökleri ilk beş karakter (F5) yöntemi kullanılarak tespit edilmiştir. Kelimeler kök haline getirildikten sonra Vektör Uzay Modeli ile veri seti modellenmiş ve her duyu için en önemli ilk 500 kelime Karşılıklı Bilgi (Mutual Information-MI) yöntemi ile tespit edilmiştir. Sınıflandırma sonuçlarının karşılaştırılmasında doğruluk metriği esas alınmıştır. Deneysel çalışma sonucuna göre, YSA algoritması en iyi sonucu vermiştir. DVM, RF ve KEYK algoritmaları ise bu sırada azalan başarıyı göstermişlerdir.

Anahtar Kelimeler: TREMO, Duygu Analizi, Makine Öğrenmesi, Metin Madenciliği

Abstract

In this research, the classification results of different Machine Learning Algorithms were compared on the validated TREMO data set used in the field of emotion extraction from Turkish texts. Emotion analysis was considered as text classification problem and four different machine algorithms, Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forest (RF) and K-Nearest Neighbor (KNN) have been investigated. The categories provided by the data set, which are happiness, fear, anger, sadness, disgust and surprise, were used as emotion categories. In the preprocessing phase, stemming process was performed using the truncate at five (F5) method. After stemming process, the data set was modeled using the Vector Space Model. After that, the first 500 words for each emotion in the data set were identified by the Mutual Information (MI) formula. The comparison of classification results was based on accuracy metric. According to experimental study results, the ANN classifier was performed best, and SVM, RF and KNN performed, in descending order.

Keywords: TREMO, Emotion Analysis, Machine Learning, Text Mining

1. Giriş

Sosyal medya uygulamalarının kullanımındaki artış, beraberinde oldukça büyük miktarda işlenmemiş metin verilerine ulaşma imkanı sağlamıştır. Bu imkan doğrultusunda, bahsedilen büyük miktardaki ham veriden anlamlı verilerin çıkarılması gündeme gelmektedir. Fakat bu tür yapılandırılmış verilerden anlamlı bilgilerin çıkarılması çok karmaşık ve pahalı süreçlerin uygulanmasını gerektirmektedir. Literatürde bu sorunun üstesinden gelmek için birçok sınıflandırma algoritması geliştirilmiştir. Bu algoritmaların temel amacı, farklı kategoriler oluşturmak için metin verilerini benzer yapılara ve anlamlara göre sınıflandırmaktır. Yeni oluşturululan bu kategori grupları sayesinde yapısal olmayan metin dosyalarını sınıflandırmak mümkün hale gelecektir.

Twitter ve Facebook gibi sosyal medya araçları, herhangi bir metinden bilgi çıkarma sürecinde büyük veri kaynakları olarak önemli role sahiptir. Bunun en önemli nedeni, bu uygulamalar sayesinde üretilen metin verilerinin her geçen gün önemli oranda artmasıdır. Fakat bu kaynaklar kategorize edilmediği için sınıflandırma algoritmalarının gereksinimlerini karşılamamaktadır. Literatürde bu sorunu çözmek için oluşturulan birçok veri seti bulunmaktadır. Örneklerden birisi, Türkçe metin sınıflandırması için oluşturulan TTC-3600 Benchmark veri setidir [1].

Literatürde İngilizce için duyu sınıflandırması yapmak amacıyla oluşturulan birçok veri seti bulunmaktadır. En sık kullanılanlardan birisi ve İngiliz merkezli Uluslararası Duygu Öncüler ve Tepkiler Anketi (ISEAR) veri setidir [2]. 37 farklı ülkeden üç bin gönüllü, bu projeye mutluluk, korku, öfke, üzüntü, iğrenme, utanç ve suçluluk duyguları ile ilgili yaşam deneyimlerini ve tepkilerini yazarak katılmışlardır. Bu veri seti duyu sınıflandırması için yapılan bazı çalışmalarda kullanılmaktadır [3]. Giachanou ve Crestani yaptıkları çalışmada Twitter'da belirli bir konuya ilgili yapılan paylaşımlarından fikir çıkarımı konusunu ele almıştır [4]. Bu hedefe ulaşmak için kullandıkları veri setlerinden biri 50 başlık ve 5.000 tweetten oluşmaktadır [5]. İlkinci veri seti ise fikirlerle ilgili terimlerin tanımlanmasında kullanılacak 2.000'den fazla kelimedenden oluşan AFINN sözlüğüdür [6]. Go, Bhayani ve Huang, tweet'leri negatif ya da pozitif olarak sınıflandırmayı amaçlamışlardır. Olumlu ve olumsuz duygular arasındaki farkları

gösteren ifadelere göre toplanan tweetler ile veri setleri oluşturulmuştur. Mohammad [7], kelime-duyu birliği sözlüklerinin, n-gram özelliklerini kullanmaktan daha iyi sonuçlar verip vermeyeceğini belirlemeye odaklanmıştır. Sonuç olarak, duyu sözlüğü özelliklerinin yeni alanlarda n-gram özelliklerini kullanmaktan daha iyi sonuçlar verdienenini gözlemlemiştir. Strapparava ve Valitutti [9], bu adımları başarmak için Ekman [8] tarafından belirlenen altı duyu için oluşturulan WordNet Affect sözlüğünü kullanmıştır. Eğitim veri seti için SemEval-2007 Affective Text Corpus'u kullanılmıştır [10]. Chaffer ve Inkpen [11] tarafından yapılan çalışmada; haber başlıklar, peri masalları ve bloglar gibi kaynaklardan heterojen bir veri seti kullanılarak altı farklı duyu çıkarılmıştır. Kouloumpis, Wilson ve Moore [12], denetimli öğrenme yaklaşımı ile öznitelikleri kullanmanın Twitter duyarlılık analizi üzerindeki etkilerini değerlendirmiştir. Bu hedefe ulaşmak için üç farklı Twitter mesaj yapısı kullanılmıştır. Bunlardan ikisi olan etiketlenmiş kelimeler ve ifadelerden oluşan veri setleri, eğitim veri setleri olarak kullanılmıştır. Modelleri test edebilmek için ise açıklamalı bir veri setinden faydalانılmıştır. Yang, Lin ve Chen [13] çalışmalarında dört duyu kategorisi; neşe, mutluluk, üzüntü ve korku için duyu sınıflandırma problemleri üzerine odaklanmıştır. Blog yazılarını ve ifadelerini eğitim veri setleri şeklinde kullanarak Destek Vektör Makineleri ile koşullu rastgele alan sınıflandırıcılarının elde ettiği sonuçlar karşılaştırılmıştır.

Türkçe metinlerden duyu çıkarımı yapabilmek için, bu alanda sınıflandırılmış bir veri setine ihtiyaç duyulmaktadır. Bu çalışmada Türkçe duyu analizi için hazırlanmış olan TREMO veri seti kullanılmıştır [14]. Çalışmada faydalanan veri setinin duyu analizi konusunda Türkçe dili için farklı bir alternatif bulunmamaktadır. Bu çalışmada 25.989 belge ile veri setinin doğrulanmış hali kullanılmıştır. TREMO veri setine yönelik analizlerin yapıldığı çalışmada en yüksek sınıflandırma sonuçlarının Destek Vektör Makineleri (DVM, Ing., Support Vector Machines) ile elde edildiği saptanmıştır. Bu çalışmada ise Yapay Sinir Ağları (YSA, Ing., Artificial Neural Network), DVM, Random Forest (RF) ve k-En Yakın Komşu (KEYK, Ing., k-Nearest Neighbor) karşılaştırılmıştır. Sonuçlar incelendiğinde YSA kullanılarak geliştirilmiş sınıflandırma modelinin daha yüksek sonuçlar verdiği gözlemlenmiştir.

Çalışmanın kalanı ise aşağıdaki şekilde düzenlenmiştir. 2. Bölümde makalenin içerisinde kullanılmış olan materyaller ve metodlara yer verilmiştir. 3. Bölümde çalışmada gerçekleştirilen deneylerin sonuçları paylaşılıp değerlendirilmiştir. Son bölümde elde edilen sonuçların kısa bir özeti paylaşılıp makalenin literatüre katkıları paylaşılmıştır.

2. Materyal ve Yöntem

Bu bölümde çalışmanın içerisinde kullanılan veri seti ve yöntemler hakkında bilgi verilmiştir.

2.1. TREMO Veri Seti

Veri setinin hazırlanması için farklı yaş gruplarından ve farklı bölgelerde yaşayan 5.000 katılımcı ile bir anket çalışması yapılmıştır. Bu ankette katılımcılardan Ekman'in tarif ettiği altı duygusal kategorisi için anılarını ve yaşadıkları deneyimlerini metin olarak paylaşmaları istenmiştir. Anket sonucunda, 4.709 adet katılımcının dokümanları onaylanmış ve 27.350 adet belge toplanmıştır. Belgelerin duygusal kategorilerine göre dağılım sayısı Tablo 1'de yer almaktadır.

Tablo 1. Belgelerin duygusal kategorilerine göre dağılımı.

Mutluluk	Korku	Öfke	Üzüntü	Tıksınme	Şaşırma
4.700	4.616	4.636	4.664	4.522	4.212

TREMO veri seti sınıflandırma algoritmalarından eğitim modeli oluşturulurken belirsiz ve sahte belgelerin olumsuz etkilerinin azaltılması amacıyla doğrulama sürecinden geçirilmiştir. Doğrulama işlemi için her belge en az üç, en fazla 5 farklı kullanıcıya sunularak oy birliği veya oy çokluğuyla belgenin duygusal kategorisine karar verilmiştir. Doğrulama işleminde 48 gönüllü kullanıcısı tüm belgeler için 92.986 oy vermiştir. Duygusal kategorileri belirsiz olan ve veri setinin %4,98'ini oluşturan 1.361 belge oylama sonucunda sistemden çıkarılarak doğrulanmış veri seti elde edilmiştir. Tablo 2'de belgelerin duygusal kategorilerine göre dağılımları yer almaktadır.

Tablo 2. Doğrulama işleminden sonra belgelerin duygusal sınıflarına göre dağılımı

Duygu sınıfı	Orijinal belge sayısı	Doğrulama işleminden sonraki belge sayısı
Mutluluk	4.700	5.229
Korku	4.616	4.393
Öfke	4.636	4.723
Üzüntü	4.664	5.021
Tıksınme	4.522	3.620
Şaşırma	4.212	3.003
Toplam	27.350	25.989

2.2. Yöntem

Bu bölümde doğrulanmış veri seti üzerinde duygusal kategorileri analiz etmek için kullanılan yöntemlerin adımları açıklanmıştır. Sınıflandırma işlemine geçmeden önce veri ön işleme ve öznitelik seçimi gerçekleştirilmiştir. Veri ön işleme adımında gereksiz terimler ve sayısal değerler silindikten sonra belge metinlerindeki kelimelerin kök bulma işlemi yapılmıştır. Kök bulma işlemi için sabit önek (fixed prefix stemming) yöntemi, öznitelik seçimi için ise mutual information (MI) yöntemi kullanılmıştır [15]. Veri ön işleme ve öznitelik seçimi aşamalarından sonra veri seti TFxIDF ağırlıklandırma yöntemi kullanılarak vektör uzay modeli oluşturulmuştur. Son olarak YSA kullanılarak sınıflandırma işlemi gerçekleştirilmiş ve sonuçlar açıklanmıştır.

2.2.1 Veri Ön İşleme

Veri setini sınıflandırma işlemine hazırlamak için veri ön işlemeye ihtiyaç vardır. İlk olarak metin içerisindeki kelimelerin köklerini bulmak amacıyla sabit önek yöntemi kullanılmıştır. Bu yöntemde kelimelerin ilk n karakteri kök olarak alınarak geri kalan kısımları kesilir. Bu çalışmada n değeri 5 olarak kabul edilmiştir (F5). Bunun nedeni, F5 yönteminin F4 ve F7 yöntemlerine göre optimum performans gösterdiğini tespit edilmesidir [16]. Böylece, veri setinde kök bulma işlemi için F5 metodu uygulanmıştır. Daha sonra noktalama işaretleri, sayısal değerler, fazla boşluklar ve gereksiz terimler silinerek veri ön işleme adımı tamamlanmıştır. Tablo 3'te Ön işlemeden geçirilmiş veri setinin sayısal değerleri paylaşılmıştır.

Tablo 3. Ön işlemeden geçirilmiş veri setinin sayısal değerleri

TREMO Versiyonu	Toplam Belge	Toplam Kelime	Birincik Kelime
F5_V	25.989	126.593	6.280

2.2.2 Öznitelik Seçimi

Veri setindeki önemsiz değerleri eleyerek önemli öznitelikleri tespit etmek ve sınıflandırma işlemi sırasında performans artışı sağlamak amacıyla öznitelik seçimi yapılmıştır. Duygu kategorileri için en önemli öznitelikleri belirlemek için karşılıklı bilgi yöntemi kullanılmıştır. Her duyguya sınıfı için öznitelikler önem derecesine göre tekrar sıralanarak, duyguya sınıfları için en önemli öznitelikler seçilmiştir. Önem değerine göre her duyguya sınıfı için ilk 500 öznitelik seçilmiştir.

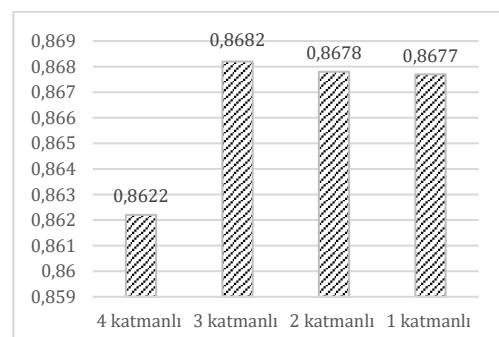
Veri ön işleme ve öznitelik seçim aşamaları tamamlandıktan sonra, veri seti vektör uzay modeline dönüştürülmüştür. Bu modelde her belge Document Term Matrisi (DTM)'de bir vektör olarak temsil edilir ve her satır terimlerini sütunların oluşturduğu bir vektörden oluşur. DTM'de sadece belgelere karşılık gelen hücrelerde veri bulunmaktadır, diğer hücrelerin değeri 0'dır. Vektör uzay modelinde TFxIDF ağırlıklandırma yöntemi kullanılmıştır [15].

3. Bulgular

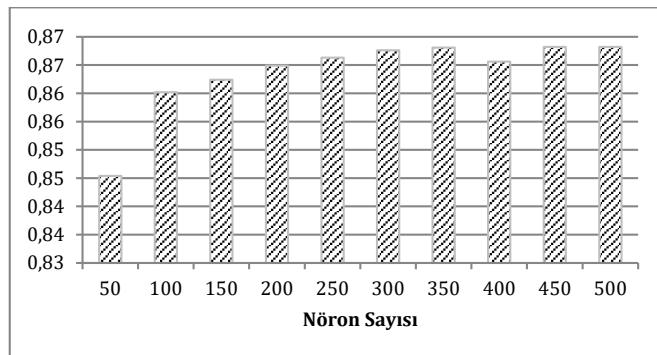
Bu bölümde, iki farklı makine öğrenme algoritmasının TREMO veri seti üzerinde uygulanmasıyla elde edilen sonuçlar karşılaştırılmıştır. Buradaki hedef, [14] çalışmada en yüksek doğruluk değerini veren DVM algoritmasından daha yüksek doğruluk değerini elde edebilecek bir algoritma tespit etmektedir. Bu amaç doğrultusunda, çok iyi bilinen bir makine öğrenme algoritması olan YSA kullanılmıştır.

Çalışmada kullanılan YSA, 5 katmandan oluşan Multi Layer Perceptron(MLP) ağ mimarisini modeli kullanılarak oluşturulmuştur. Bu katmanlar sırasıyla 1.397 adet nöron

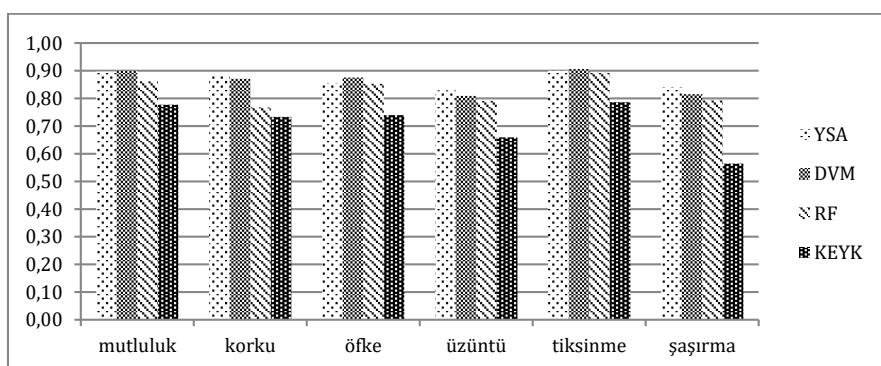
bulunduran giriş katmanı, her bir katmanda 450 adet nöron bulunduran 3 adet ara katman ve 6 adet nöron bulunduran çıktı katmanıdır. 3 adet ara katman kullanılmasının nedeni 3 katmanlı bir yapının test edilen diğer katman sayılarına sahip sinir ağlarına oranla daha iyi performansa sahip olmasıdır. Şekil 1'de katman sayılarına göre doğruluk değerleri yer almaktadır. YSA modelinin katmanlarında bulunan nöron sayıları belirli kurallara göre belirlenmiştir. Giriş katmanında bulunan 1.397 adet nöron TREMO veri seti üzerinde ilk 500 öznitelik seçimi yöntemi ile elde edilmiş benzersiz kelimeleri temsil etmektedir. Ara katmanlarda 450 adet nöron kullanılması ile daha yüksek doğruluk değeri elde edilmiştir. Şekil 2'de modelin farklı nöron sayıları ile test edilmesi sonucu elde edilen doğruluk değerleri gösterilmektedir. Çıkış katmanında bulunan 6 adet nöron sınıflandırırmada kullanılan duyguya kategorilerini ifade etmektedir. Bunlar sırasıyla; mutluluk, korku, öfke, üzüntü, tıksınme ve şaşırma duygularıdır. YSA modelinde gizli katmanlarda rectifier aktivasyon fonksiyonunu, çıktı katmanında ise softmax aktivasyon fonksiyonu kullanılmıştır. Ayrıca, sinir ağındaki en optimal ağırlıkları bulmada kullanılan optimizasyon algoritması Olasılıksal Dereceli Azalma (Stochastic Gradient Descent) algoritmasıdır. Bununla beraber modelde çıktı katmanı 6 adet nöronundanoluştuğu için logaritmik kayıp fonksiyonu olarak kategorik çapraz entropi kullanılmıştır.



Şekil 1. Ara katman farklılıklarına göre YSA modelinin verdiği genel doğruluk değerleri.



Şekil 2. Ara katmanı oluşturan nöron sayılarının farklılıklarına göre YSA modelinin verdiği genel doğruluk değerleri

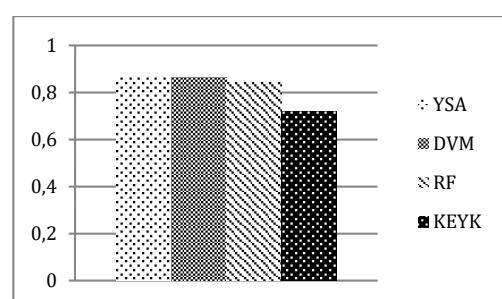


Şekil 3 YSA, DVM, RF ve KEYK algoritmalarının duygusal kategorilerde doğruluk değerlerinin karşılaştırılması

Bu çalışmada, YSA, DVM, RF ve KEYK algoritmalarının ürettiği sınıflandırma sonuçları 10 Katmanlı Çarpraz Doğrulama yöntemi kullanılarak değerlendirilmiştir. Şekil 4'de her iki algoritmalarının genel doğruluk değerleri karşılaştırılmıştır. Sonuçlar incelenince, YSA algoritmasının 0,0045 gibi az bir farkla da olsa DVM algoritmasına üstünlüğü gözükmemektedir.

Genel doğruluk değerlerinin dışında bir diğer incelenmesi gereken sonuçlar ise her bir duygusal kategorisi için hesaplanan doğruluk değerleridir. Böylece, her bir duygusal bireysel sınıflandırma performansı sergilenebilir. Şekil 3'de YSA, DVM, RF ve KEYK algoritmalarının duygusal kategorilerde doğruluk değerlerinin karşılaştırılması gösterilmiştir. Öncelikle, her iki algoritma için de üzüntü kategorisi en düşük, tıksinme kategorisi en yüksek doğruluk değerlerini vermiştir. Bunun yanında, YSA algoritması korku, üzüntü ve şaşırma duygularında DVM algoritmasına üstünlük

kurarken, mutluluk, öfke ve tıksinme kategorilerinde DVM algoritmasının YSA algoritmasına üstünlüğü mevcuttur.



Şekil 4. YSA, DVM, RF ve KEYK algoritmalarının genel doğruluk değerlerinin karşılaştırılması

Tablo 5, 6, 7 ve 8'de sırasıyla YSA, DVM, RF ve KEYK algoritmaları için elde edilen karışıklık matrisleri gösterilmektedir. Bu matrislerin

paylaşılmasındaki temel amaç, hangi duygunun hangi duyguya daha çok ya da daha az karıştığını tespit etmektir. Her iki algoritmanın sonuçlarına bakıldığından, aynı duyguya kategorilerinin birbirleriyle en çok ve en az miktarlarda

karışıklık gösterdikleri tespit edilebilmektedir. Örneğin Tablo 5'de şaşırma duygusunun 247 belge sayısıyla en çok mutluluk duygusuyla, 16 belge sayısıyla da en az tıksınme duygusuyla karıştığı gözlemlenebilmektedir.

Tablo 5. YSA algoritmasının kullanımı sonucu elde edilen karışıklık matrisi

	Mutluluk	Korku	Öfke	Üzüntü	Tıksınme	Şaşırma	Doğruluk
Mutluluk	4.682	89	189	160	0	100	0,8969
Korku	106	3.901	104	209	68	32	0,8826
Öfke	157	112	3.798	216	91	66	0,8554
Üzüntü	369	118	219	3.980	9	105	0,8292
Tıksınme	22	82	214	20	3.509	53	0,8997
Şaşırma	247	55	112	85	16	2.695	0,8396

Tablo 6. DVM algoritmasının kullanımı sonucu elde edilen karışıklık matrisi

	Mutluluk	Korku	Öfke	Üzüntü	Tıksınme	Şaşırma	Doğruluk
Mutluluk	4.701	95	128	146	15	144	0,8990
Korku	171	3.826	100	210	59	27	0,8709
Öfke	210	81	4.134	200	52	46	0,8753
Üzüntü	480	129	234	4.060	24	94	0,8086
Tıksınme	91	65	143	28	3.279	14	0,9058
Şaşırma	296	48	106	91	14	2.448	0,8152

Tablo 7. Random Forest (RF) algoritmasının kullanımı sonucu elde edilen karışıklık matrisi

	Mutluluk	Korku	Öfke	Üzüntü	Tıksınme	Şaşırma	Doğruluk
Mutluluk	4.504	89	152	306	38	140	0,8614
Korku	725	3.830	134	215	64	25	0,7671
Öfke	217	81	4.029	230	118	48	0,8531
Üzüntü	439	176	295	3.971	45	95	0,7909
Tıksınme	69	82	164	52	3.229	24	0,8920
Şaşırma	282	61	103	150	26	2.381	0,7929

Tablo 8. KNN algoritmasının kullanımı sonucu elde edilen karışıklık matrisi

	Mutluluk	Korku	Öfke	Üzüntü	Tıksınme	Şaşırma	Doğruluk
Mutluluk	4.064	179	245	368	65	308	0,7772
Korku	274	3.367	211	333	97	311	0,7331
Öfke	363	138	3.419	325	207	171	0,7396
Üzüntü	625	380	387	3.309	84	236	0,6590
Tıksınme	155	174	274	94	2.846	77	0,7862
Şaşırma	561	161	255	248	83	1.695	0,5644

4. Tartışma ve Sonuç

Bu çalışmada, doğrulanmış TREMO veri seti üzerinde literatürde yaygınla kullanılan YSA, DVM, RF ve KEYK makine öğrenme algoritmaları doğruluk metriği kullanılarak duyu analizi sonuçları karşılaştırılmıştır. Bunun için öncelikle TREMO veri seti ön işleme bölümünden geçirilmiştir ve F5 yöntemi ile TREMO'da bulunan bütün kelimelerin kökleri bulunmuştur. Sonrasında, veri setindeki her bir duyu için en önemli ilk 500 kelime MI formülü kullanılarak tespit edilerek özellik seçimi yapılmıştır ve veri seti uzay vektör modeli kullanılarak modellenmiştir. Elde edilen bu veri modeli YSA, DVM, RF ve KEYK algoritmalarını eğitmek ve test etmek amacıyla kullanılmıştır. Değerlendirme metriği olarak doğruluk değerinin kullanıldığı sınıflandırma sonuçlarına bakıldığından, genel doğruluk değerlerinde YSA algoritmasının diğer üç algoritmala göre üstünlük kurdugu gözlemlenmiştir. Sonuçlar duyu kategorileri incelendiğinde YSA ve DVM algoritmaları birbirile yakını yakın sonuçlar üretmişlerdir.

Gelecekte yapılması planlanan işler arasında, TREMO veri setinin YSA ağlarının temelini oluşturuğu derin öğrenme yöntemleri kullanılarak duyu analizi yapılması vardır.

- [8] Ekman, P. 1992. An Argument for Basic Emotions, Cognition & Emotion. Cilt. 6, s. 169-200.
- [9] Strapparava, C., Valitutti, A. 2004. Wordnet Affect: an Affective Extension of Wordnet. International Conference on Language Resources and Evaluation, 26-28 Mayıs, Lizbon, 1083-1086.
- [10] Strapparava, C., Mihalcea, R. 2007. SemEval-2007 task 14: Affective Text. 4. International Workshop on Semantic Evaluations, 23-24 Haziran, Prag, 70-74.
- [11] Chaffar, S., Inkpen, D. 2011. Using a Heterogeneous Dataset for Emotion Analysis in Text. Conference on Artificial Intelligence, 25-27 Mayıs, St. John's, 62-67.
- [12] Kouloumpis, E., Wilson, T., Moore, J.D. 2011. Twitter Sentiment Analysis: The Good the Bad and the Omg!. 5. International Conference on Weblogs and Social Media, 17-21 Temmuz, Barselona, 538-541.
- [13] Yang, C., Lin, K. H.Y., Chen, H.H. 2007. Emotion Classification Using Web Blog Corpora. Web Intelligence Conference, 2-5 Kasım, Washington DC, 275-278.
- [14] Tocoglu, M.A., Alpkocak, A. 2018. TREMO: A Dataset for Emotion Analysis in Turkish. Journal of InformationScience. DOI: <https://doi.org/10.1177/0165551518761014>.
- [15] Manning, C.D., Raghavan, P., Schütze, H. 2008. Boolean Retrieval. Introduction to Information Retrieval, 1-18.
- [16] Can, F., Kocherber, S., Balcik, E., Kaynak, C., Ocalan, H. C., Vursavas, O. M. 2008. Information Retrieval on Turkish Texts. Journal of the American Society for Information Science and Technology, Cilt. 59, s. 407-421. DOI: <https://doi.org/10.1002/asi.20750>

Kaynakça

- [1] Kılınç, D., Özçift, A., Bozyigit, F., Yıldırım, P., Yüçalar, F., & Borançag, E. 2017. TTC-3600: A New Benchmark Dataset for Turkish Text Categorization. Journal of Information Science, Cilt. 43, s. 174-185. DOI: <https://doi.org/10.1177/0165551515620551>
- [2] Scherer, K.R., Wallbott, H.G. 1994. Evidence for Universality and Cultural Variation of Differential Emotion Response Patterning. Journal of Personality and Social Psychology, Cilt. 67, s. 55. DOI: <http://dx.doi.org/10.1037/0022-3514.67.1.55>
- [3] Danisman, T., Alpkocak, A. 2008. Feeler: Emotion Classification of Text Using Vector Space Model. AISB 2008 Convention Communication, Interaction and Social Intelligence, 1-4 Nisan, Aberdeen, 53-59.
- [4] Giachanou, A., Crestani, F. 2016. Opinion Retrieval in Twitter Using Stylistic Variations. 31. Annual ACM Symposium on Applied Computing, 4-8 Nisan, Pisa, 1077-1079.
- [5] Luo, Z., Osborne, M., Wang, T. 2015. An Effective Approach to Tweets Opinion Retrieval. World Wide Web, Cilt. 18, s. 545-566.
- [6] Go, A., Bhayani, R., Huang, L. 2009. Twitter Sentiment Classification Using Distant Supervision. CS224N Project Report, Stanford, 1(12).
- [7] Mohammad, S. 2012. Portable Features for Classifying Emotional Text. North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 3-8 Haziran, Montreal, 587-591.