

Duygu Analizi için Çoklu Popülasyon Tabanlı Parçacık Sürü Optimizasyonu

Multiple Population based Particle Swarm Optimization for Sentiment Analysis

İlhan AYDIN¹, Mehmet Umut SALUR², Fatma BAŞKAYA³

^{1,3}Bilgisayar Mühendisliği Bölümü, Fırat Üniversitesi, Elazığ, Türkiye
iaydin@firat.edu.tr, fatma_sdu_@hotmail.com

²Bilgisayar Mühendisliği Bölümü, Harran Üniversitesi, Şanlıurfa, Türkiye
umutsalur@harran.edu.tr

Öz

Metin tabanlı içerikler arasında bulunan duyguların tespit edilmesi duygu analizi olarak ifade edilir. İnternet altyapısının dünya genelinde güçlenmesi, insanlara bir konudaki düşüncelerini çevrimiçi ifade etme imkânı sağlamıştır. İnternet ortamında toplanan bu verilerden önemli bilgilerin çıkarılması hemen hemen her alan için önem arz etmektedir. Bu çalışmada da çokça kullanılan Twitter veri kümeleri üzerinde duygu analizi işlemi gerçekleştirilmiştir. Metinlerdeki duygular olumlu, olumsuz veya belirsiz olarak sınıflandırılmıştır. Sınıflandırma işleminden önce veri kümeleri üzerinde metin madenciliği ön işlemleri uygulanmış ve sonrasında özellik çıkarımı yapılmıştır. Bu çalışmada sınıflandırma işlemi için optimizasyon tabanlı yeni bir yöntem önerilmiştir. Bu yöntemle elde edilen sınıflandırma başarımının literatürdeki çalışmalardan daha başarılı olduğu deneylerle tespit edilmiştir.

Anahtar Sözcükler — Duygu Analizi, Parçacık Sürü Optimizasyon, Metin Ön işleme, Özellik Çıkarımı.

Abstract

The determination of emotions among text-based content is expressed as sentiment analysis. The strengthening of the Internet infrastructure around the world has allowed people to express their thoughts online. Extraction of important information from these data collected on the Internet is important for almost every field. In this study, sentiment analysis processing was performed on the frequently used Twitter data sets. Sentiments in the texts are classified as positive, negative or uncertain. Text

mining preprocesses were applied on the data sets prior to the classification, and feature extraction was performed afterwards. In this study, a new optimization-based method for classification is proposed. Experiments have shown that the classification performance obtained by this method is more successful than the studies in the literature.

Keywords — Sentiment Analysis, Particle Swarm Optimization, Text Preprocessing, Feature Extraction.

1. Giriş

İnternet kullanımının yaygınlaşması, internet okur-yazarlığının artması toplumun iletişim yöntemlerinde değişikliklere neden olmuştur. İnternet altyapısını kullanan herkes aynı zamanda bir dijital veri üreticisi konumundadır. Bu bakış açısıyla insanlar tarafından her gün boyutu terabaytlara ulaşan veri kümeleri üretilmektedir. Veriler içeriklerine göre günlük rutin işlerden kaynaklı oluşan veriler (banka işlemleri, alış-veriş vb.), sosyal medya verileri (Twitter, Facebook, Instagram vb.), üretim sahası verileri (fabrika, atölye, imalathane vb.), kişisel veriler (ad, soyad, medeni durumu, çocuk sayısı vb.) olarak sınıflandırılabilir.

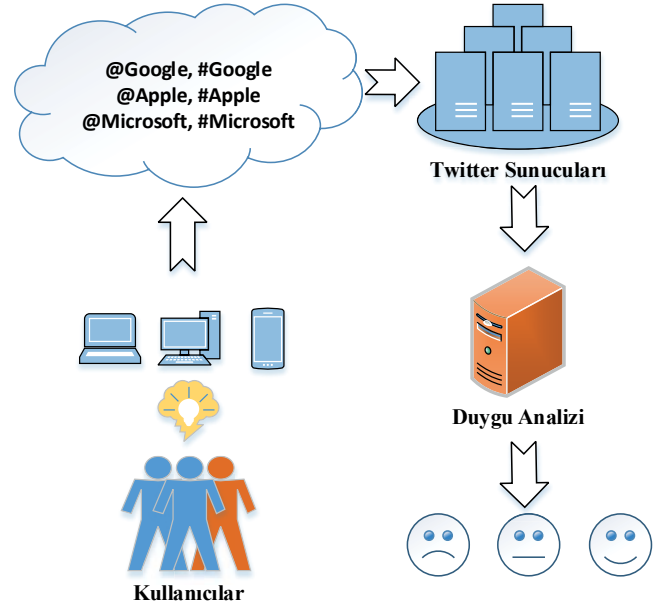
Web teknolojilerinin gelişmesi sonucunda kullanıcılar arasında karşılıklı etkileşimli küresel bir sanal ortam meydana gelmiştir. Web 2.0 olarak isimlendirilen bu web versiyonu kullanıcılar arasında hızlı ve çift yönlü etkileşim sağlayan bir ortam sunmuştur [1]. Web 2.0'ın en önemli bileşeni ise sosyal medya uygulamaları olmuştur. Sosyal medya uygulamaları kullanıcılara geniş içerikte özgür bir paylaşım ortamı sağlamaktadır [1][2]. Sosyal medya uygulamalarından çokça tercih edilenlerden biri olan Twitter, kullanıcılara temel de metin tabanlı paylaşım ortamı sağlamaktadır. Twitter uygulamasını kullanan

Gönderme ve kabul tarihi: 02.02.2018-03.05.2018

kullanıcılar, internet erişiminin olduğu her yerden uygulama yardımıyla sosyal içerikler (ne hissettiği, ne yaptığı, ne düşündüğü gibi) paylaşmaktadırlar.

Dünya genelinde Twitter uygulamasına aylık yaklaşık 313 milyon kişi giriş yapmaktadır. Bunu yanında Twitter üzerinden günlük ortalama 500 milyon içerik paylaşılmaktadır [3]. Bu paylaşımlar genellikle kullanıcıların bir konu, durum veya olay hakkındaki duygu ve düşüncelerinden oluşmaktadır. Araştırmacılar Twitter'ın kullanıcılara sunmuş olduğu açık erişimli servisler yardımıyla Twitter veri tabanından belli parametreler doğrultusunda ücretsiz bir şekilde veri alabilmektedirler. Bu avantajlı yönüyle Twitter verileri birçok araştırmacının materyali olma niteliği taşımaktadır. Metin tabanlı içeriklerin araştırılması ve bu içeriklerden kullanıcıların duygularının tespit edilmesi duygu analizi olarak ifade edilmektedir [4][5]. Twitter tabanlı duygu analizi çalışmaları genellikle bir konu hakkındaki duyguların tespit edilmesi, müşterilerin bir hizmet hakkındaki duyguların tespit edilmesi ve bir topluluğun duyguların tespit edilmesi şeklinde olmaktadır [1][6][7][8].

Duygu analizinde bir metni olumlu, olumsuz veya belirsiz olarak sınıflandırmak için metin üzerinde bir dizi işlem (önişlemler, doğal dil işleme, özellik çıkarımı vb.) gerçekleştirilir. Bu işlemlerin öncelikli amacı veri kümesinin sınıflandırma başarımını arttırmaktır. Bu amaç kapsamında literatürde başta Twitter veri kümeleri olmak üzere çeşitli veri kümeleri üzerinde duygu analizi/sınıflandırması çalışması yapılmıştır. Twitter üzerinden yapılan duygu analizi çalışmalarının genel yöntemi Şekil-1'deki gibi gösterilebilir. Kullanıcılar tarafından paylaşılan mesajlar Twitter sunucularında saklanır, araştırmacının amacı doğrultusunda kullanılacak veri kümeleri belirlendikten sonra Twitter sunucularında bulunan servisler yardımıyla veriler toplanır. Toplanan bu veriler araştırmacının kişisel bilgisayarında veya başka sunuculara işlenir.



Şekil 1. Twitter üzerinden duygu analizi genel gösterimi

Duygu analizinin sıklıkla yapıldığı çalışmalardan biri, web veya sosyal medya ortamında bir film hakkında yapılan yorumların olumlu veya olumsuz olarak değerlendirilmesidir. Çek dili üzerine yapılan film yorumlarını sınıflandırma çalışmasında, yorumlar pozitif, negatif ve belirsiz olarak üç sınıfa ayrılmıştır [9]. Çalışmada özellik çıkarımı ve metinlerin sayısallaştırılması aşamasında N-gram, PoS (Part of Speech) Tags ve TF-IDF (Term Frequency- Inverse Document Frequency) gibi yöntemlerin başarımları karşılaştırılmıştır. Film yorumlarının sınıflandırıldığı bir diğer çalışmada n-gram yöntemi referans alınarak duygu sınıflandırma işlemi gerçekleştirilmiştir [7].

Film yorumlarının yanı sıra duygu analizi otel yorumları üzerinde de gerçekleştirilmektedir. Seyahat edecek kişiler için konaklamayı planladıkları otellerin kalitesi ve daha önce otelde konaklayan müşterilerin memnuniyeti önemli olmaktadır. Bu kişiler otel rezervasyon uygulamaları (booking.com, hotels.com vb.) yardımıyla kalacakları otelere karar verebilmektedirler. Bu uygulamalar kullanıcılara otel hakkında yorum yapma ve hizmet değerlendirme puanı vermek gibi özellikler sunmaktadır. Bu uygulamalara ek olarak seyahat için özelleşmiş rehber niteliğinde olan tripadvisor.com, foursquare.com gibi başka uygulamalar da bulunmaktadır. Belli bir otelde konaklamayı düşünen bir kişi otelle ilgili bu yorumları okuyarak otelin kalitesi hakkında fikir edinmektedir. Bu nedenle otel yorumlarının bilgisayarlar tarafından değerlendirilmesi önemlidir. Otel yorumlarının sınıflandırılması hususunda [10]'daki çalışmada otel yorumları farklı yöntemlerle sınıflandırılmıştır. Sınıflandırma başarımları değerlendirilmiştir. [1]'de

yapılan çalışmada ise otel yorumlarından özet çıkarma üzerine bir yöntem önerilmiştir.

Film ve otel yorumları haricinde Twitter üzerinden paylaşılan bir şirketle ilgili, bir hizmetle ilgili veya bir ürünle ilgili mesajlardan insanların duyguları tespit edilmeye çalışılmaktadır. Bu amaçlardan herhangi birisiyle ilgili veri kümelerinin sınıflandırılmasında makine öğrenmesi algoritmalarına ek olarak günümüzde çeşitli optimizasyon yöntemleri de kullanılmaktadır. Twitter verilerini kullanan ve bu çalışmanın da fikir kaynağı olan [11]'de; Twitter'dan alınan Apple, Google, Twitter, Microsoft vb. konularda yapılan paylaşımlardan oluşan veri kümeleri bir optimizasyon yöntemi olan Guguk Kuşu Algoritması (GKA) ve K-Ortalama Kümeleme yönteminin melez kullanımı ile sınıflandırılmıştır. Sınıflandırma işleminde metin içindeki duygular olumlu, olumsuz veya belirsiz olacak şekilde duygu sınıflandırması yapılmıştır. Çalışmada küme merkezlerinin belirlenmesinde sezgisel bir algoritma olan GKA kullanılmıştır. Önerilen melez yöntemle yapılan sınıflandırma işleminde elde edilen başarımların sonuçları Karar Destek Vektörleri, Naive Bayes gibi sınıflandırma yöntemlerinden daha iyi olduğu gösterilmiştir. Bir başka melez çalışma olan [6]'da film yorumlarının olumlu veya olumsuz sınıflandırılması için Destek Vektör Makinalar (DVM) algoritmasıyla PSO birlikte kullanılmıştır. Bu melez yöntem sınıflandırma başarımını yaklaşık %5 oranında arttırmıştır.

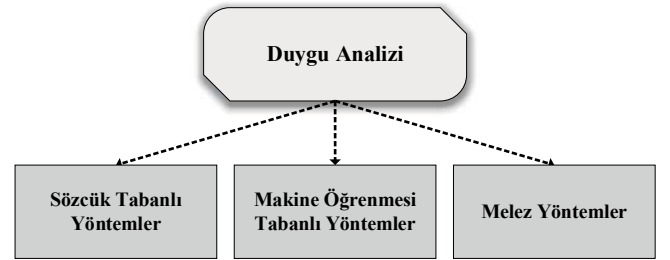
Bu çalışmada ise Apple, Google, Microsoft gibi firmalarla ilgili Twitter'da paylaşılan üç veri kümesi olumlu, olumsuz veya belirsiz olarak PSO ve K-EYK algoritmalarıyla sınıflandırılmıştır. Öncelikle veri kümeleri üzerinde metin madenciliği ön işlemleri gerçekleştirilmiş, sonrasında veri kümelerinden 10 adet öznitelik çıkarımı yapılmıştır. Elde edilen öznitelikler önerilen melez yöntem ile sınıflandırılmış ve [11]'deki çalışmadan daha iyi sınıflandırma başarımı elde edilmiştir. Önerilen yöntemin temel katkısı her bir sınıf için sınıflandırma amacıyla bir küme merkezi bulmak yerine çoklu küme merkezlerinin bulunmasıdır. Her bir sınıfı temsil eden küme merkezleri o sınıf için yarışan parçacıklardan oluşmaktadır. Aynı sınıftaki parçacıklar birbirlerine doğru hareket ederken farklı sınıfta olan parçacıklardan uzaklaşmaktadırlar. Bu özellik ile K-EYK algoritmasının performansı arttırılmıştır. Herhangi bir test verisi o küme içinde en yakın parçacığa göre sınıflandırılmaktadır. Ayrıca özellik çıkarımı aşamasında elde edilen özellikler duygu sınıflandırma için ayırt edici olup n-gram ve TF-IDF gibi hesaplama yükü fazla olan yöntemlere göre daha avantajlıdır.

Bu makalenin ikinci bölümünde duygu analizinden bahsedilmiş ve duygu analizinde kullanılan yöntemlere değinilmiştir. Üçüncü bölümünde yapılan çalışmanın basamakları anlatılmıştır. Dördüncü bölümde ise çalışmanın sonuçları değerlendirilmiştir.

2. Duygu Analizi

İnsanlar web ortamında herhangi bir konu hakkındaki düşüncelerini ifade eden yorumlar, açıklamalar paylaşabilmektedir. Paylaşılan bu içeriklerden kişilerin düşüncelerinin belli sınıflar kapsamında belirlenmesi işlemleri duygu analizi olarak ifade edilmektedir [12]. Duygu analizinde amaç algoritmalar tarafından metin tabanlı içeriğinin ikili veya çoklu sınıflar arasında ayrıştırma işlemlerinin otomatik olarak yapılmasının sağlanmasıdır. Bu şekilde üretim yapan şirketler ürünleri ve hizmet veren kurumlar hizmetleri hakkında müşteri memnuniyetini kolaylıkla belirleyebilir. Diğer taraftan kişilerin paylaşımlarından suç işleme potansiyeline sahip kişiler tespit edilebilir [13]. Bu kişilerin tespit edilmesi olası felaket veya kazaların önüne geçilmesi açısından önemli olmaktadır. İnternet ortamında haber yapan sitelerin paylaştıkları haberlerin içeriklerine göre de duygu analizi yapılmaktadır [14][15]. Bu şekilde bir ülkede yaşanan olaylarla ilgili insanların genel duygu haritaları çıkarılabilmektedir. Bu da ülkelerin mutluluk veya mutsuzluk düzeylerini belirlemede kullanılabilir.

Metin tabanlı içeriklerden duyguların çıkarılması işlemi Şekil 2'de gösterilen yöntemler yardımıyla yapılmaktadır.



Şekil 2. Duygu analizi yöntemleri

2.1. Sözlük Tabanlı Yöntemler

Dünya üzerinde konuşulan veya yazılan her dilin kendisine göre kuralları ve yapıları vardır. Diller kelime tabanlı veya ek tabanlı olabilmektedirler. Bir dildeki kelimelerin en az bir anlamı vardır. Tüm kelimelerin anlamlarını olumlu, olumsuz veya belirsiz anlamlı olarak sınıflandırmak mümkündür.

Duygu analizinde kullanılan sözlük tabanlı yaklaşımlarda her kelimenin anlamları referans alınarak tüm metnin duygusu sınıflandırılmaktadır [16]. Her dil için oluşturulmuş olan duygu

sözlüklerinde her kelimenin olumluluk veya olumsuzluk derecesi bulunmaktadır. Bu derecelerin farklı bağlamlar altında birleştirilmesiyle tüm metnin duygusu ortaya çıkarılmaktadır [17]. Kelime tabanlı dillerde bu yöntemlerin kullanılması duyguların tespit edilmesinde başarılı sonuçlar verirken Türkçe gibi sondan eklemeli dillerde sözlük tabanlı duygu analizi çok iyi sonuçlar vermemektedir. Bunun temel nedeni ise kelimenin köküne gelen her yapım ekinin kelimenin duygu özelliği üzerinde büyük değişikliklere neden olmasıdır. Örneğin; Türkçe bir kelime birden fazla yapım veya çekim eki alabilmektedir. Her ek için kelimelerin duygularının derecelendirilmesi mümkün görülmemektedir.

Birçok dil için duygu analizi sözlüğü geliştirilmiştir. Üzerinde duygu analizinin en çok yapıldığı dillerden biri olan İngilizce için SentiWordNet [18], SenticTweety [19], MPQA Opinion Corpus [20] gibi birçok duygu sözlüğü oluşturulmuştur. Bu sözlüklerdeki kelimelerin duygu dereceleri farklı yöntemlerle birleştirilerek tüm metnin duygusu ortaya çıkarılıp sınıflandırılmaktadır.

2.2. Makine Öğrenmesi Tabanlı Yöntemler

Makine öğrenmesi yöntemleri denetimli ve denetimsiz olmak üzere iki temel kısma ayrılmaktadır. Denetimli yöntemlerde algoritmalar öncelikle etiketlenmiş bir eğitim verisi ile eğitilir, bu eğitimle algoritmaya öğrenme yetisi kazandırılmış olur. Öğrenen algoritma yeni gelen sınıf etiketi belli olmayan verileri karar mekanizmasını kullanarak sınıflandırır. Denetimli öğrenmede genellikle Yapay Sinir Ağları, Destek Vektör Makinesi, Naive Bayes, Lojistik Regresyon, Rastgele Orman, Karar Ağaçları ve K-EYK algoritmaları kullanılmaktadır.

Denetimsiz yöntemlerde algoritma eğitilmez, algoritmanın kendisinin öğrenmesi beklenir. Bu öğrenme veriler arasındaki benzerlik veya farklılıklar baz alınarak gerçekleştirilir. Denetimsiz öğrenme yöntemleri genellikle kümeleme işlemi için kullanılmaktadır. Denetimsiz öğrenmede K-Ortalamlar, K-Medoids veya CLARANS gibi yöntemler kullanılmaktadır.

Duygu analizinde genellikle denetimli makine öğrenmesi algoritmaları kullanılmaktadır. El ile veya kelime tabanlı olarak işaretlen eğitim verisiyle algoritmanın eğitimi gerçekleştirilir.

2.3. Melez Yöntemler

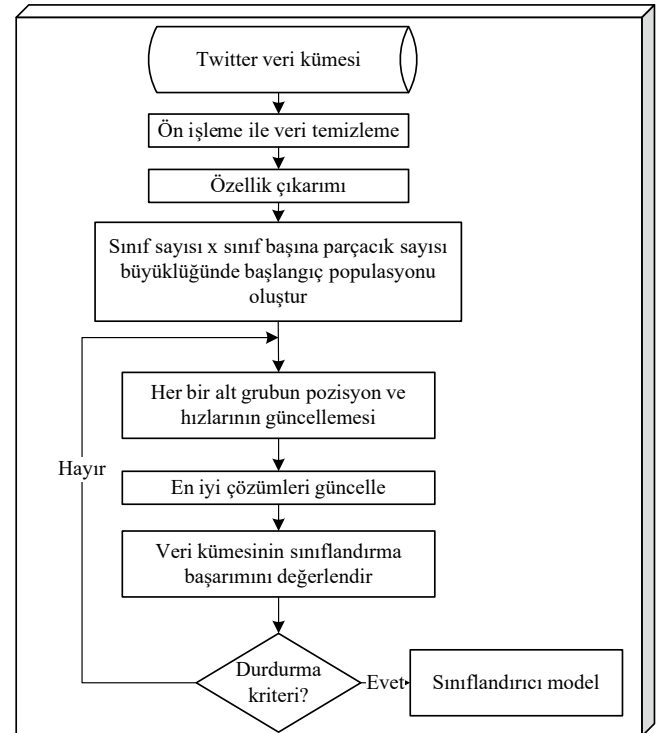
Duygu analizinde farklı yöntemlerin güçlü yanları birleştirilerek oluşturulan melez yöntemlerle de duygu çıkarımı yapılmaktadır. Melez yöntemler genellikle

sözlük tabanlı yöntemler, kural tabanlı yöntemler ile makine öğrenmesi yöntemlerinin birlikte kullanıldığı uygulamalardır [21]. Bu uygulamaların yanında günümüzde ayrıca makine öğrenmesi yöntemleri optimizasyon yöntemleriyle birlikte kullanılmaktadır.

Optimizasyon algoritmalarının kullanıldığı yöntemlerde sınıflandırma başarımı yalın makine öğrenmesi yöntemlerine göre daha yüksek olmaktadır [6][11]. Genellikle optimizasyon yöntemleri ile duygu küme merkezleri için en uygun nokta belirlenmektedir.

3. Önerilen Yöntem

Bu çalışma kapsamında duygu analizi için optimizasyon tabanlı melez bir yöntem önerilmiştir. Önerilen yöntemin etkinliğinin belirlenmesi için birçok çalışmada kullanılan Twitter veri kümeleri kullanılmıştır. Bu veri kümelerinin makine öğrenmesi algoritması tarafından işlenebilmesi için ön işleme, özellik çıkarma gibi bir dizi işlemlerden geçirilmiştir. Şekil 3'te önerilen yöntemin öbek şeması verilmiştir.



Şekil 3. Önerilen yöntemin öbek şeması

Şekil 3'te verilen öbek şemasında Twitter veri kümesinden ilk olarak ön işlemede ilgisiz içerikler silinmektedir. Daha sonra her Twitter mesajından özellikler çıkarılmaktadır. Elde edilen özellikler çoklu PSO tabanlı sınıflandırıcı ile sınıflandırılmaktadır.

3.1. Veri Önışleme

Genellikle web ortamında paylaşılan metinlerin içerisinde kelime kökleri ve ekleri haricinde birçok içerik bulunabilmektedir. Bu ek içeriklerin başında noktalama işaretleri, rakamlar, linkler, #, @, ♦, %, &, [, *,], \$, {, } vb. semboller gelmektedir. Bu içerikler metin içinde barındırılan duyguya herhangi bir katkıda bulunmamaktadır [9][16]. Metin tabanlı içeriklerin yüksek doğruluk değerleriyle sınıflandırma gerçekleştirmesi açısından, metnin içerisinde olan ve metnin barındırdığı duygulara herhangi bir katkısı olmayan içeriklerden arındırılması gerekmektedir. Bu içerikler literatürde gürültü olarak isimlendirilmektedirler.

Metin önışleminin sınıflandırma başarımı üzerindeki etkisi [3]'te incelenmiştir. Yapılan bu çalışmada iyi bir sınıflandırma başarımının sağlanması açısından önışlemin gerekli olduğu sonucuna varılmıştır.

Çalışmanın bu aşamasında kullanılan veri kümeleri Java programlama diliyle yazılmış olan önışlem-özellik çıkarımı uygulaması yardımıyla gürültülerden arındırılmıştır. Gerçekleştirilen önışlem adımlarının özeti Şekil 4'te görüldüğü gibidir. Şekil 4'te verilen önışlemler adımlardan 12'inci adımda; herhangi bir kelimenin yazımında eğer kullanıcı birden fazla karakterle vurgu yapmak istemişse bu tür kelimeler düzeltilmiştir. Örneğin; çalışmada kullanılan üçüncü veri kümesi içerisinde bulunan "@thesexiestchic jokes! Now why I got to be alllllll that!" bu cümlede Twitter kullanıcısı İngilizcedeki "all" kelimesine vurgu yapmak istemiştir. Bu tür yazım hataları yine önışlem aşamasında düzeltilmiştir.

Benzer şekilde Twitter'dan mesaj paylaşan kullanıcılar kelimeler arasındaki boşluklara dikkat etmezler. Hızlı yazmaya çalışan kullanıcılar kelimeler arasında birden fazla boşluk olacak şekilde mesaj paylaşmaktadırlar. Yine 7'nci önışlem adımında bu tür yazım yanlışları düzeltilmiştir. Her kelime arasında bir boşluk olacak şekilde her mesaj yeniden düzenlenmiştir.

Çalışmanın özellik çıkarım aşamasında metin içerisinden birçok özellik çıkarımı yapılmıştır. Veri kümelerinden çıkarılan özelliklerden biri metin içerisindeki olumlu, olumsuz veya belirsiz emojilerdir. Çıkarılan özniteliklerden sadece emojiler her mesaj için önışlem aşamasından önce veri kümelerinden elde edilmiştir. Çünkü emojiler (, : ; , ,) , = gibi karakter harici sembollerden oluşmaktadır. Önışlem aşamasında bu semboller silinmektedir.

Şekil 4'te verilmeyen bir diğer önışlem ise sonlandırıcı kelimelerin (stop words) metin içerisinden temizlenmesidir. Örneğin; "and, any, are, as, at, be, because ..." gibi kelimeler metin içerisinden temizlenmiştir. Bu kelimelerin metin içerisindeki duygu ifadesi üzerinde olumlu veya olumsuz bir etkisi yoktur.

1	Tüm karakterleri küçük harflere dönüştür
2	Linkleri url diye değiştir
3	@username'i usr diye değiştir
4	Kelimeler arası çoklu boşlukları sil
5	(, [, /, & gibi karakterleri sil
6	#hashtag'lerdeki #'i sil
7	Tweet'in başındaki ve sonundaki fazla boşlukları sil
8	Noktalama işaretlerini sil
9	Diğer tanınmayan karakterleri sil
10	Tüm rakamları sil
11	Alfabe ile başlamayan kelimeleri sil
12	Çoklu karakter tekrarlarını ingilizceye göre düzelt

Şekil 4. Metin önışlem adımları

3.2. Özellik Çıkarımı

Metin formatındaki verilerin makine öğrenmesi algoritmaları tarafından işlenmesi/sınıflandırılması için bu verilerin sayısal formata dönüştürülmesi gerekmektedir. Literatürde birçok yöntemle metin formatındaki veriler sayısallaştırmaktadır. Bu yöntemlerin başında vektör uzay modeli gelmektedir [25]. Bu yöntemde temel felsefe her metni sayısal bir vektör olarak ifade etmektir. Bu yöntemlerin yanı sıra metin içerisinde metni ifade eden özellikler çıkarılarak da metinler sayısallaştırılabilmektedir. Bu

çalışmada da kullanılan üç veri kümesi için aynı özellik çıkarımları yapılmıştır. Toplamda her metin için 10 adet özellik çıkarılmıştır. Özellik çıkarımından sonra her Twitter mesajı 10 özellik ile ifade edilmiştir. Tüm veri kümelerinden çıkarılan özellikler küçük örneklerle Şekil 5'te verilmiştir.

➤	Toplam Karakter Sayısı
➤	Olumlu Emoji Sayısı [=), :, =o: , :D]
➤	Olumsuz Emoji Sayısı [= (, =' (, =([
➤	Belirsiz Emoji Sayısı [: = , :-}X , *<<<<+]
➤	Olumlu Kelime Sayısı [amazed, renewed, selective]
➤	Olumsuz Kelime Sayısı [aborts, awfully, blame]
➤	Belirsiz Kelime Sayısı [drawer, underdrawers, drawing]
➤	Olumlu Ünlem Sayısı [aha, ooh-la-la, yeeeeeah]
➤	Olumsuz Ünlem Sayısı [aaaahh, augh, oww]
➤	Olumsuzluk Anlamı Veren Kelime Sayısı [doesn't, isn't, wasn't, shouldn't]

Şekil 3. Çıkarılan özellikler

Çıkarılan özellikler aşağıdaki gibi açıklanabilir;

- Toplam karakter sayısı; veri kümelerindeki her mesaj için toplam karakter sayısı bir öznelik olarak çıkarılmıştır.
- Olumlu emoji sayısı; her Twitter mesajının içerisindeki olumlu emoji sayısıdır.
- Olumsuz emoji sayısı; her Twitter mesajının içerisindeki olumsuz emoji sayısıdır.
- Belirsiz emoji sayısı; her Twitter mesajının içerisindeki belirsiz emoji sayısıdır.
- Olumlu kelime sayısı; her Twitter mesajının içerisinde bulunan olumlu kelimelerin sayısıdır.
- Olumsuz kelime sayısı; her Twitter mesajının içerisinde bulunan olumsuz kelimelerin sayısıdır.
- Belirsiz kelime sayısı; her Twitter mesajının içerisinde bulunan olumlu veya olumsuz anlam ifade etmeyen kelimelerin sayısıdır.
- Olumlu ünlem sayısı; her Twitter mesajının içerisinde bulunan olumlu ünlemlerin sayısıdır.

- Olumsuz ünlem sayısı; her Twitter mesajının içerisinde bulunan olumsuz ünlemlerin sayısıdır.
- Olumsuzluk anlamı veren kelimelerin sayısı; her Twitter mesajının içerisinde bulunan ve İngilizce de kullanılan olumsuzluk anlamı veren dilin yapısal kelimelerinin sayısıdır.

Çıkarılan özelliklerle her veri kümesindeki Twitter mesajları sayısal olarak ifade edilmiştir. Elde edilen özellikler önerilen melez yöntem ve başka makine öğrenmesi yöntemleriyle sınıflandırılmıştır.

3.3. Çoklu Popülasyon Tabanlı Parçacık Sürü Optimizasyonu ile Duygu Analizi

Parçacık sürü optimizasyonu kuş ve balık sürülerinin yiyecek arama davranışlarının benzetimi ile ortaya çıkan bir optimizasyon yöntemidir. Parçacık sürü optimizasyonunda sürüdeki her birey aday bir çözümdür. Her parçacık aday bir çözüm olup hız ve pozisyon bilgisine sahiptir. Parçacıklar kendilerinin daha önceki en iyi pozisyonu ve popülasyondaki en iyi bireyin pozisyonuna göre güncellenir. Standart bir PSO algoritmasının güncelleme denklemleri denklem (1) ve (2)'de verilmiştir.

$$V_{id} = WxV_{id} + c_1xr_1x(P_{id} - X_{id}) + c_2xr_2x(P_{gd} - X_{id}) \quad (1)$$

$$X_{id} = X_{id} + V_{id} \quad (2)$$

Denklem (1) ve (2)'de X_{id} pozisyon ve V_{id} hız değerini göstermektedir. Diğer parametrelerden r_1 ve r_2 [0,1] aralığında üretilen rastgele sayıları, c_1 ve c_2 öğrenme katsayılarını ve W atalet ağırlık değerini ifade eder. Parçacık sürü optimizasyonu verilen bir amaç fonksiyonuna göre bir optimizasyon problemi için en uygun parametreleri bulur. Parametreler bulunurken bütün parçacıklar tek bir amaca doğru yaklaşıma çalışır.

Bu çalışmada en yakın komşu sınıflandırma için çoklu parçacık optimizasyonu ile duygu analizi önerilmektedir [26]. Önerilen yöntemde her bir parçacık en yakın komşu kuralına göre örüntüleri sınıflandırmak için bir çözümdür. Parçacıkların yer değiştirme kuralı değiştirilerek bütün parçacıklar için ortak bir çözüm bulmak yerine yerel uygunluklar bulunmaktadır. Her parçacık aynı zamanda bir sınıf etiketine sahip olup, eğitim süresi boyunca değişmemektedir. Her bir parçacığın boyutu özellik

sayısını göstermektedir. K adet sınıf için N adet parçacığın gösterimi Çizelge 1’de verilmiştir.

Çizelge 1. Yapılandırma dosyaları ve görevleri

Parçacık	Pozisyon	Sınıf
Parçacık 1	X _{1,1} X _{1,2} ... X _{1,D}	0
Parçacık 2	X _{2,1} X _{2,2} ... X _{2,D}	1
...
Parçacık K	X _{K,1} X _{K,2} ... X _{K,D}	K-1
Parçacık K+1	X _{K+1,1} X _{K+1,2} ... X _{K+1,D}	0
...
Parçacık NxK	X _{NxK,1} X _{NxK,2} ... X _{NxK,D}	K-1

Çizelge 1’de D her parçacığın boyutunu gösterir. Eğer elimizde 3 sınıflı bir sınıflandırma problemi var ise ve her bir sınıf için 10 adet parçacık üretilecek ise toplam popülasyon boyutu 30 olacaktır. Önerilen algoritmanın akış şeması Şekil 6’da verilmiştir.

Algoritma: Çoklu Popülasyon Tabanlı PSO

```

Veri kümesini yükle
Özellik sayısı boyutunda parçacıkları oluştur
Her sınıf için N tane parçacık oluştur
While(Maksimum adım sayısı kadar) do
Her sınıf için yarışan/yarışmayan parçacıkları bul
  for her parçacık için do
    Lokal uygunluğu hesapla
    Sosyal adaptasyon parametresini hesapla
    Parçacık sınıfı için yarışmayan sınıftaki en yakın parçacığı bul (Çekim merkezi)
    Parçacık sınıfı için yarışmayan kümedeki en yakın parçacığı bul
    Parçacığın sonraki pozisyonunu bul
  endfor
Parçacıkları taşı
En yakın parçacığı kullanarak eğitim kümesindeki örüntülere sınıfları ata
Sürü sınıflandırma başarımını hesapla
endwhile

```

Şekil 6. Çoklu popülasyon tabanlı PSO ile duygu analizi

Şekil 6’da verilen çoklu popülasyon tabanlı PSO algoritmasının her adımda hız güncelleme denklemi denklem (3)’te verilmiştir [26].

$$V_{i+1,d} = \chi (w \times v_{i,d} + c_1 \times r_1 \times (p_{i,d} - x_{i,d}) + c_2 \times r_2 \times \text{sign}(a_{i,d} - x_{i,d}) \times S f_i + c_3 \times r_3 \times \text{sign}(x_{i,d} - r_{i,d}) \times S f_i) \quad (3)$$

Denklem (3)’te kullanılan parametreler ve açıklamaları Çizelge 2’de verilmiştir.

Çizelge 2. Hız güncelleme parametreleri

Parametre	Açıklama
$v_{i,d}$	Parçacık i’nin hızının d. bileşeni
$x_{i,d}$	Parçacık i’nin konumunun d. bileşeni
c_1, c_2, c_3	Sabit ağırlık faktörleri
r_1, r_2, r_3	[0,1] aralığında rastgele değişkenler
p_i	Parçacık i’nin mevcut en iyi pozisyonu
w	Moment ağırlığı
χ	Ölçekleme faktörü
a_i, r_i	Parçacık i’nin çekim ve itme merkezi
$S f_i$	Sosyal adaptasyon faktörü
sign	Etki yönünü belirten signum fonksiyonu

Denklem (3)’teki ilk terim moment ağırlığı ile önceki hızın çarpımı olup standart PSO’daki ile aynıdır. İkinci terim ise standart PSO’da da olan parçacığın o ana kadar ki en iyi pozisyonu ile arasındaki farkı almaktadır. Üçüncü terim bir parçacığın kendisine en yakın ve aynı sınıfta olan parçacık ile arasındaki farkın signum fonksiyonudur. Son parametrede signum fonksiyonu kullanılmış olup hız değişiminin yönü itme merkezine göre belirlenmektedir. Şekil 7’de verilen bir i parçacığına en yakın yarışan parçacığı bulma algoritması verilmiştir.

```

y=yarisanparcacik(pop, i)
for j=1:popboyut do
  if(j!=i) then
    if (pop.yarisan(j))==1 &&
      pop.C(j)==pop.C(i) then
        mesafe=oklid(pop, i,j);
        mesafe=[mesafe;mes]
        indis=[indis;j];
      endif
    endif
  endif
endfor
ind=min(mesafe);
y=pop.pos(indis(ind));

```

Şekil 7. Sınıf için yarışan en yakın parçacık bulma algoritması

Her adımda parçacıkların güncellenmesi için iki komşuluk durumu tanımlanmıştır. Bunlar yarışan ve

yarışmayan parçacıklardır. Bir parçacık ile aynı sınıfta olmayan ve bu sınıftan en az bir tane örüntüyü sınıflandıran parçacıklar yarışmayan parçacıklardır. Yarışan parçacıklar ise mevcut parçacık ile aynı sınıfta olup bu sınıftan en az bir tane örüntüyü sınıflandıran parçacıklardır.

Şekil 7’de verilen algoritma ile popülasyondaki i . parçacık ile aynı sınıf etiketine sahip olup bu sınıf için yarışan ve i . parçacığa en yakın olan parçacık bulunmaktadır. Bu parçacık i . parçacığın aynı zamanda çekim merkezini (a_i) oluşturur. Yarışmayan bu en yakın parçacığa ise itme merkezi (r_i) denilmektedir. Önerilen yöntemde her bir parçacık yerel bir sınıflandırıcı olarak kullanıldığı için bir amaç fonksiyonuna sahiptir. Herhangi bir örüntünün sınıfı en yakın olan parçacığa göre belirlenir. Dolayısıyla bir parçacığın doğru sınıflandırdığı ve yanlış sınıflandırdığı örüntü sayısı kullanılmaktadır. Bu iki özellik denklem (4)’e göre hesaplanmaktadır.

$$D = \sum_{j \in (d)} \frac{1}{1 + m_{ij}} \quad (4)$$

$$Y = \sum_{j \in (y)} \frac{1}{1 + m_{ij}}$$

Denklemden (4)’te (D) kümesi i . parçacık tarafından doğru sınıflandırılan örüntü kümesini (Y) ise yanlış sınıflandırılan örüntü kümesidir. Denklemden m_{ij} ise i . Parçacık ile j . örüntü arasındaki öklid mesafesini gösterir. Her parçacığın uygunluk fonksiyonu denklem (5)’e göre hesaplanmaktadır.

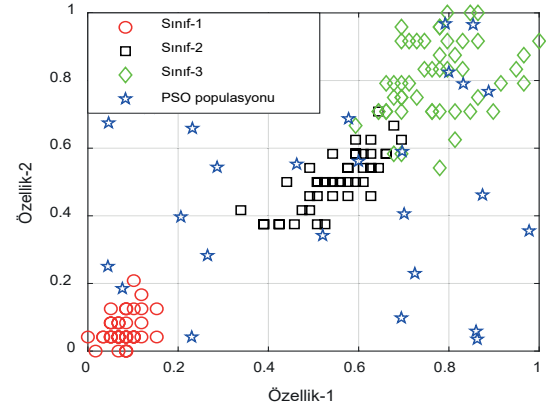
$$Uygunluk_i = \begin{cases} \frac{D}{toplam} + 2.0 & \text{Eger } (d) \neq 0, \\ & (y) = 0 \\ \frac{D-Y}{D+Y} + 1.0 & \text{Eger } (y) \neq 0, \end{cases} \quad (5)$$

Denklemden (5)’te eğer bir parçacık sadece kendi sınıfından örüntüleri sınıflandırıyorsa uygunluk değeri 2.0 ‘dan büyük olur. Eğer bir parçacık ait olmadığı sınıftan da örüntü sınıflandırıyorsa uygunluk $[0, 2]$ arasında çıkar.

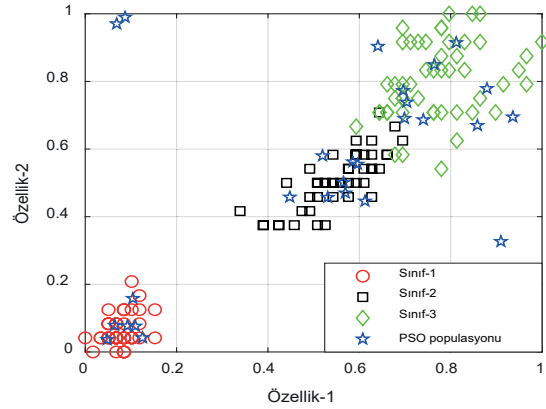
Hız güncellemesinde kullanılan sosyal adaptasyon parametresi parçacıkların yarışan ve yarışmayan komşularından etkilenme oranını belirler. Bu parametre ile parçacıkların yarışan parçacıklardan uzak ve yarışmayanlara doğru hareketini sağlar. Bu parametre denklem (6)’da verilmiştir.

$$Sf_i = \frac{1}{(En\ iyi\ uygunluk_i + 1)} \quad (6)$$

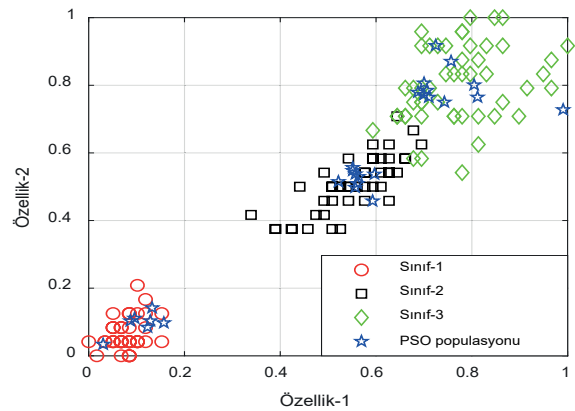
Denklemden (6)’da eğer i . parçacığın o ana kadarki en iyi uygunluğu düşük ise Sf_i 1’e yakın çıkar ve etkilenme oranı yüksek çıkar. Eğer en iyi uygunluk yüksek çıkarsa (parçacığın uygunluğu 2’nin üzerinde ise) bu değer küçük çıktığından etkilenme düşük olur. Önerilen çoklu PSO sınıflandırıcısının üç sınıflı bir sınıflandırma problemi üzerinde çalıştırılma adımları Şekil 8’de verilmiştir.



(a) Başlangıç popülasyonu



(b) 15 adım sonra



(c) 50 adım sonra

Şekil 8. Çoklu PSO sınıflandırıcısının örnek veri setinde çalışmasının gösterimi

Şekil 8 (a)'da rastgele üretilen başlangıç popülasyonunun uzayın tamamını kapsadığı görülmektedir. Şekil 8 (b)'de ise 15 adım sonra her parçacık grubunun kendi sınıfına doğru gittiği ve halen bazı parçacıkların kendi sınıflarına yaklaşmadığı görülmektedir. Şekil 8 (c)'de ise her grubun kendi sınıfına doğru yaklaştığı görülmektedir. Dikkat edilmesi gereken bir nokta ise parçacıkların tek bir hedef yerine sınıf içinde dağılım göstermesidir. Bu özellik sosyal adaptasyon parametresi, çekim ve itme merkezlerinden kaynaklanmaktadır. Bu aşamadan sonra her bir örüntü en yakın olduğu parçacığın sınıfına göre sınıflandırılmaktadır.

3.4. Uygulama Sonuçları

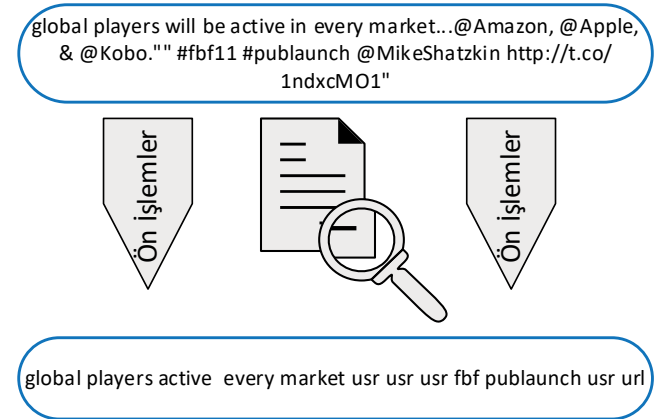
Twitter'ın hemen hemen herkes tarafından kullanılıyor olması Twitter'ın iyi bir veri kümesi niteliği taşımasına neden olmaktadır. Dünya genelinden bu verilere Twitter'ın sunmuş olduğu servisler yardımıyla erişim sağlanabilmektedir. Bu çalışma kapsamında önerilen yöntem üç ayrı Twitter veri kümesi üzerinde test edilmiştir. Kullanılan Veri Kümesi-1 (VK-1), Veri Kümesi-2 (VK-2) ve Veri Kümesi-3 (VK-3)'ün özellikleri Tablo 3'te verilmiştir. Bu veriler veri kümesi hazırlayıcılar tarafından manuel olarak olumlu, olumsuz veya belirsiz (nötr) olarak etiketlenmiştir [11].

Çizelge 3. Kullanılan veri kümelerinin özellikleri

Özellikler	VK-1	VK-2	VK-3
Adı	Twitter-sanders-apple2	Twitter-sanders-apple3	Twitter dataset
Toplam Örnek Sayısı	479	988	2000
Sınıf Sayısı	2	3	2
Pozitif Örnek Sayısı	163	163	1000
Negatif Örnek Sayısı	316	316	1000
Belirsiz Örnek Sayısı	-	509	-
Alınma Tarihi	15.10.2011 20.05.2011		17.11.2014 10.12.2014
Konu	Apple, Google, Microsoft, Twitter		Spor, Öğrenciler, Komik Resimler vb.

- *Veri Kümesi-1:* Sanders Analytics tarafından Apple firması için toplanan bir veri kümesidir. Apple, Google, Microsoft ve Twitter hakkında paylaşılan mesajlardan oluşmaktadır. Veri kümesinde 471 adet mesaj bulunmaktadır.
- *Veri Kümesi-2:* Veri kümesi-1 ile aynı özelliklere sahiptir. Sadece örnek sayısı daha fazla olmakla birlikte, veri kümesinde üç ayrı duyguya ait mesajlar bulunmaktadır. Veri Kümesi-1 ile Veri Kümesi-2'ye [22]'de verilen linkten ulaşılabilir.
- *Veri Kümesi-3:* Twitter duygu analizi bağlamı içerisinde alınan bir kısım veriden oluşan bir alt kümedir [23]. Veri kümesi [24]'te verilen linkten elde edilebilir. Bu veri kümesinde 1000 adet olumlu ve 1000 adet olumsuz mesaj bulunmaktadır.

Veri kümesi alındıktan sonra bir ön işleme adımı ile düzeltilmesi ve gereksiz bilgilerin çıkarılması gerekmektedir. Şekil 9'da VK-2'de bulunan örnek bir cümle için ön işlemlerden önceki ve sonraki hali verilmiştir.



Şekil 9. Ön işlem örneği

Ön işleme adımından sonra çoklu PSO sınıflandırıcısı için özellikler elde edilmektedir. Her bir Twitter mesajı için toplam 10 özellik elde edilmiştir. Elde edilen özellikler çoklu PSO sınıflandırıcısına verilerek sınıflandırma işlemi yapılmaktadır. Çoklu PSO tabanlı sınıflandırıcı her bir sınıf için 10 parçacık kullanılmaktadır. Dolayısıyla sadece pozitif ve negatif etiketlerin olduğu veri kümelerinde popülasyon boyutu 20, belirsiz etiketinin de olduğu veri kümesinde ise 30 olarak seçilmiştir. Parçacık sürü optimizasyonunun özellikleri Çizelge 4'te verilmiştir.

Çizelge 4. Önerilen PSO'nun parametreleri

Parametre	Değer
χ	0.5
w	0.1
c_1	1.0
c_2	1.0
c_3	0.25

Parçacık sürü optimizasyonu her bir durum için 10 kez çalıştırılmış ve 10 katlı çapraz doğrulama yöntemi ile doğrulanmıştır.

Çoklu PSO tabanlı duygu sınıflandırıcının başarımlarını değerlendirmesinde gerçek sınıfı ve tahmin edilen sınıfı içeren karışıklık matrisi kullanılmıştır. Üç sınıflı bir duygu sınıflandırma için karmaşıklık matrisinin gösterimi Çizelge 5'te verilmiştir.

Çizelge 5. Üç sınıflı eğitim için karışıklık matrisi

Tahmin etiket	Gerçek Etiket			
	1	2	3	Toplam
1	DP ₁₁	YP ₁₂	YP ₁₃	TT ₁
2	YP ₂₁	DP ₂₂	YP ₂₃	TT ₂
3	YP ₃₁	YP ₃₂	DP ₃₃	TT ₃
Toplam	DT ₁	DT ₂	DT ₃	

Karışıklık matrisi kullanılarak sınıflandırma başarımlarını belirlemek için hassasiyet, keskinlik, doğruluk oranı ve F1 ölçütü kullanılmıştır. Keskinlik, her sınıf için doğru şekilde tahmin edilen örneklerin bu sınıfta tahmin edilen toplam örnek sayısına oranıdır. Karışıklık matrisine göre bu değer denklem (7)'ye göre hesaplanmaktadır.

$$Keskinlik = \frac{1}{S} \sum_{i=1}^S \frac{DP_{ii}}{TT_i} \quad (7)$$

Denklem (7)'de DP_{ii} etiketi i olan sınıf için doğru şekilde sınıflandırılan test örneklerinin sayısıdır. TT_i parametresi ise sınıf etiketi i olarak sınıflandırılan örnek sayısıdır. Denklemde S sınıf sayısını göstermektedir.

Hassasiyet, her sınıf için doğru bir şekilde sınıflandırılan örneklerin olması gereken toplam örnek sayısına bölümüdür. Bu ölçüm denklem (8)'e göre elde edilmektedir.

$$Hassasiyet = \frac{1}{S} \sum_{i=1}^S \frac{DP_{ii}}{DT_i} \quad (8)$$

Denklemdaki DT_i parametresi sınıf etiketi i olan sınıftaki test örneklerinin sayısını gösterir ve ilgili sütunun toplamı ile hesaplanır.

F1 başarımlar ölçütü ise keskinlik ve hassasiyet parametrelerinin birleştirilen bir ölçüt olup denklem (9)'a göre hesaplanmaktadır.

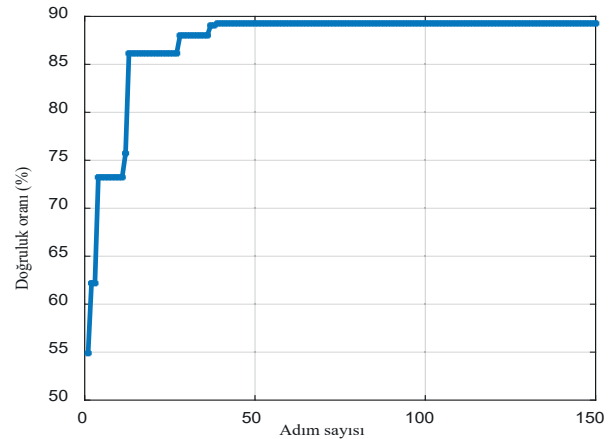
$$F1 = \frac{2 \times Keskinlik \times Hassasiyet}{Keskinlik + Hassasiyet} \quad (9)$$

Denklem (9)'daki F1 parametresi 0 ile 1 arasında değer alır ve 1 olması durumunda tüm örneklerin doğru sınıflandırıldığını ifade eder.

Son başarımlar ölçütü ise doğruluk olup doğru şekilde sınıflandırılan örneklerin oranını gösterir. Karışıklık matrisine göre doğruluk oranı denklem (10)'a göre hesaplanmaktadır.

$$Dogruluk = \frac{\sum_{i=1}^S DP_{ii}}{toplamlar} \quad (10)$$

Veri kümesi-1 için çoklu PSO tabanlı sınıflandırıcının ortalama başarımlar oranı Şekil 10'da verilmiştir.



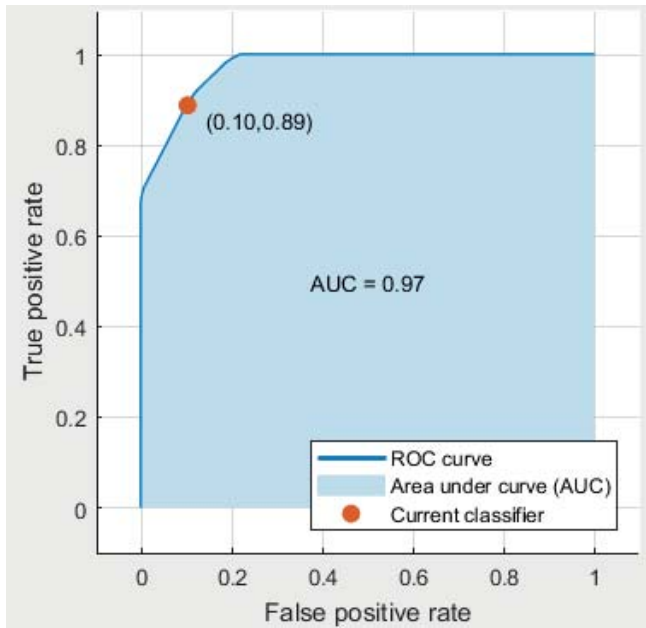
Şekil 10. Çoklu PSO tabanlı duygu sınıflandırma başarımlarını

Şekil 10'da çoklu PSO tabanlı duygu sınıflandırıcı yönteminin 50 adımda maksimum başarımlara ulaştığı görülmektedir. Çizelge 6'da ise bu veri kümesi için sınıflandırma sonucunda elde edilen karmaşıklık matrisi verilmiştir.

Çizelge 6. VK-1 için karışıklık matrisi

Tahmin edilen sınıf	Doğru sınıf	
	Pozitif	Negatif
Pozitif	145	18
Negatif	32	285

Tablo 6’da verilen karışıklık matrisine göre doğruluk oranı %89.6 olarak elde edilmiştir. Şekil 11’de bu veri kümesi için ROC eğrisi verilmiştir.



Şekil 11.Veri kümesi-1 için ROC eğrisi

Şekil 11’deki ROC eğrisinde gösterildiği gibi VK-1 veri kümesinin sınıflandırma başarımı yüksektir. Veri kümesi-2’de aynı veri kümesine nötr etiketini içeren veriler eklenmiştir. Veri-kümesi 2 için karışıklık matrisi Çizelge 7’de verilmiştir.

Çizelge 7. VK-2 için karışıklık matrisi

Tahmin edilen sınıf	Doğru sınıf		
	Pozitif	Negatif	Nötr
Pozitif	120	13	31
Negatif	21	257	38
Nötr	29	49	445

Çizelge 7’de çoklu PSO algoritmasının veri kümesi-2 üzerinde sınıflandırma başarımı %83.19 olarak elde edilmiştir. Uygulama sonuçları nötr sınıf

eklendiğinde de başarımın çok düşmediğini göstermektedir.

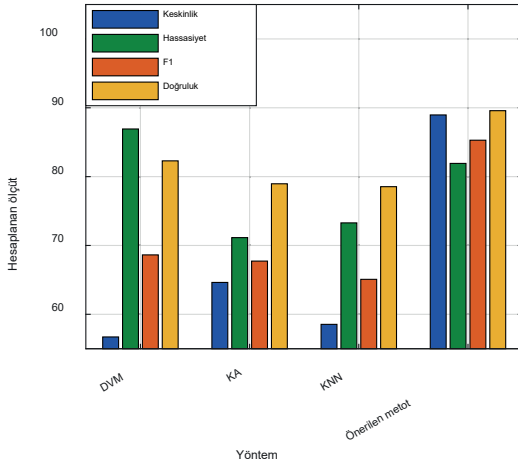
Çoklu PSO tabanlı duygu sınıflandırıcı ayrıca üçüncü bir veri kümesi üzerinde de test edilmiştir. Bu veri kümesi diğerlerine göre daha büyük olup sadece pozitif ve negatif görüşler içermektedir. Ayrıca her bir sınıf için örnek sayısı eşittir. Veri kümesi-3 için karışıklık matrisi Çizelge 8’de verilmiştir.

Çizelge 8. VK-3 için karışıklık matrisi

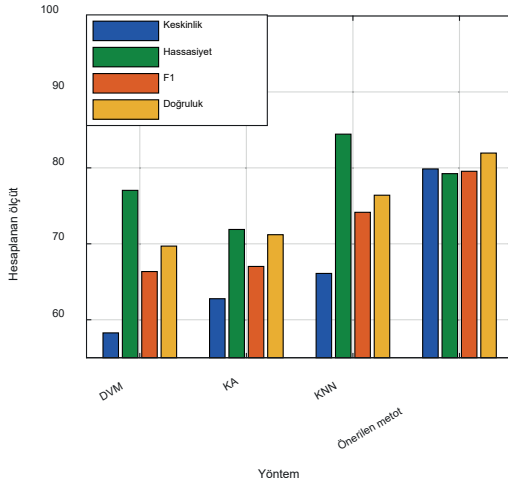
Tahmin edilen sınıf	Doğru sınıf	
	Pozitif	Negatif
Pozitif	747	253
Negatif	334	666

Çizelge 8’den görüldüğü gibi üçüncü sınıf üzerinde de %69.95’lik bir sınıflandırma başarımı elde edilmiştir. Önerilen çoklu PSO tabanlı duygu sınıflandırma yöntemi bilinen makine öğrenmesi yöntemleri ile karşılaştırılmıştır. Sınıflandırma başarımlarını karşılaştırmak için Keskinlik, Hassasiyet, F1 skoru ve Doğruluk Oranı kullanılmıştır. Önerilen yöntem sınıflandırma başarımı yüksek olan karar ağacı (KA), k-en yakın komşu algoritması (KNN) ve destek vektör makinalar (DVM) ile karşılaştırılmıştır. Karşılaştırma yapmak amacıyla MATLAB’da bulunan Classifier Learner aracı kullanılmıştır. Bu araçta birçok makine öğrenmesi tabanlı sınıflandırma algoritması olmasına rağmen bu çalışmada en yüksek başarımı veren üç algoritma seçilmiştir. Önerilen yöntem ve bu üç algoritmanın sınıflandırma başarımları Şekil 12’de verilmiştir.

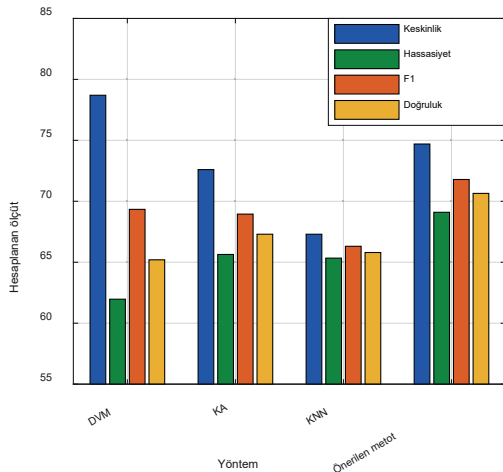
Şekil 12’de elde edilen sonuçlara göre önerilen yöntem birinci veri kümesi için bütün ölçütlerde en başarılı sonuçları vermektedir. İkinci veri kümesinde sadece KNN’nin hassasiyeti daha iyi çıkmıştır. Diğer ölçütlerde önerilen yöntem daha iyi sonuçlar vermiştir. Üçüncü veri kümesinde ise sadece DVM’nin keskinlik değeri önerilen yöntemle göre daha yüksek çıkmıştır. Bu sonuçlar önerilen yöntemin bilinen makine öğrenmesi yöntemleri ile yarışabileceğini göstermektedir.



(a) VK-1 için karşılaştırma sonuçları



(b) VK-2 için karşılaştırma sonuçları



(c) VK-3 için karşılaştırma sonuçları

Şekil 12. Önerilen yöntemin farklı makine öğrenme yöntemleri ile karşılaştırılması

4. Sonuç

Bu çalışmada duygu analizi ve sınıflandırma için çoklu PSO tabanlı bir sınıflandırma yöntemi sunulmuştur. Önerilen yöntem standart PSO algoritmasından farklı olarak birden çok hedefe giden parçacık kümelerinden oluşmaktadır. Aynı sınıfta olan parçacıklar birbirlerini çekerken farklı grupta olanları itmekteler. Önerilen yöntem üç farklı Twitter veri kümesi üzerinde test edilmiştir. Bu amaçla öncelikle alınan Twitter mesajları üzerinde ön işlem uygulanarak verilerin hazırlanması sağlanmıştır. Daha sonra özellik çıkarımı aşamasında 10 özellik çıkarılmıştır. Önerilen yöntem sınıflandırma başarımı iyi olan makine öğrenmesi yöntemleri ile karşılaştırılmış ve daha iyi sonuçlar elde edilmiştir. Önerilen yöntem ile optimizasyon sonucunda parçacıklar veri kümeleri üzerinde en uygun pozisyonda bulunarak sınıflandırma işlemini yapmaktadır. Sınıflandırma aşamasında k-en yakın komşu algoritmasına göre bir örüntünün en yakın olduğu parçacığa göre yapılmaktadır.

KAYNAKÇA

- [1] Y.-H. Hu, Y.-L. Chen, and H.-L. Chou, "Opinion mining from online hotel reviews--A text summarization approach," *Inf. Process. Manag.*, vol. 53, no. 2, pp. 436–449, 2017.
- [2] D. Stojanovski, "Twitter Sentiment Analysis using Deep CNN," vol. 9121, no. JUNE, 2015.
- [3] Z. Jianqiang and G. Xiaolin, "Comparison research on text pre-processing methods on twitter sentiment analysis," *IEEE Access*, vol. 5, pp. 2870–2879, 2017.
- [4] M. Bouazizi and T. Otsuki, "A Pattern-Based Approach for Sarcasm Detection on Twitter," *IEEE Access*, vol. 4, pp. 5477–5488, 2016.
- [5] F. Wu, Z. Yuan, and Y. Huang, "Collaboratively Training Sentiment Classifiers for Multiple Domains," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 7, pp. 1370–1383, 2017.
- [6] A. S. H. Basari, B. Hussin, I. G. P. Ananta, and J. Zeniarja, "Opinion mining of movie review using hybrid method of support vector machine and particle swarm optimization," *Procedia Eng.*, vol. 53, pp. 453–462, 2013.
- [7] A. Tripathy, A. Agrawal, and S. Kumar, "Classification of Sentimental Reviews Using Machine Learning Techniques," vol. 0, no. November, pp. 117–126, 2014.
- [8] D. Jiang, X. Luo, J. Xuan, and Z. Xu, "Sentiment Computing for the News Event Based on the Big Social Media Data," *IEEE*

- Access*, vol. 3536, no. c, pp. 1–1, 2016.
- [9] I. Habernal, T. Ptáček, and J. Steinberger, “Reprint of ‘supervised sentiment analysis in Czech social media,’” *Inf. Process. Manag.*, vol. 51, no. 4, pp. 532–546, 2015.
- [10] Y. H. Hu, K. Chen, and P. J. Lee, “The effect of user-controllable filters on the prediction of online hotel reviews,” *Inf. Manag.*, vol. 54, no. 6, pp. 728–744, 2017.
- [11] A. Chandra Pandey, D. Singh Rajpoot, and M. Saraswat, “Twitter sentiment analysis using hybrid cuckoo search method,” *Inf. Process. Manag.*, vol. 53, no. 4, pp. 764–779, 2017.
- [12] Aydın, I., Başkaya, F., & Salur, M. U. (2017, October). Sentiment classification with PSO based weighted K-NN. In *Computer Science and Engineering (UBMK), 2017 International Conference on* (pp. 739-744). IEEE.
- [13] J. Azeez and D. J. Aravindhar, “Hybrid approach to crime prediction using deep learning,” *2015 Int. Conf. Adv. Comput. Commun. Informatics*, pp. 1701–1710, 2015.
- [14] Salur, M. U., Gündüz G., “Mahout ile Veri Madenciliği Uygulaması Gazete Tweetleri Üzerinde Duygu Analizi,” *Uluslararası Bilgi sayar Bilimleri ve Mühendisliği Konferansı (UBMK 2016)*, Tekirdağ, 2016.
- [15] V. Bobichev, O. Kanishcheva, and O. Cherednichenko, “Sentiment Analysis in the Ukrainian and Russian News,” pp. 1050–1055, 2017.
- [16] R. Bhonde, B. Bhagwat, S. Ingulkar, and A. Pande, “Sentiment Analysis Based on Dictionary Approach,” *Int. J. Emerg. Eng. Res. Technol.*, vol. 3, no. 1, pp. 51–55, 2015.
- [17] F. K. Chopra, “Sentiment Analyzing by Dictionary based Approach,” vol. 152, no. 5, pp. 32–34, 2016.
- [18] “SentiWordNet.” [Çevrimiçi]. URL: <http://sentiwordnet.isti.cnr.it/>. [Erişim: 29-Kasım-2017].
- [19] “SenticTweety.” [Çevrimiçi]. URL: <http://tweety.sentic.net/>. [Erişim: 29-Kasım-2017].
- [20] “MPQA Opinion Corpus.” [Çevrimiçi]. URL: http://mpqa.cs.pitt.edu/#subj_lexicon. [Erişim: 29-Kasım-2017].
- [21] O. Appel, F. Chiclana, J. Carter, and H. Fujita, “A hybrid approach to sentiment analysis,” *2016 IEEE Congr. Evol. Comput.*, no. Cci, pp. 4950–4957, 2016.
- [22] “Twitter-sanders-apple.” [Çevrimiçi]. URL: <http://boston.lti.cs.cmu.edu/classes/95-865-K/HW/HW3/>. [Erişim: 24-Ekim-2017].
- [23] “Twitter Sentiment Corpus.” [Çevrimiçi]. URL: <http://www.sananalytics.com/lab/twitter-sentiment/>. [Erişim: 11-Ekim-2017].
- [24] “Twitter Dataset.” [Çevrimiçi]. URL: https://drive.google.com/file/d/0BwPSGZHAP_yoN2pZcVl1Qmp1OEU/view. [Erişim: 12-Ekim-2017].
- [25] B. Altunel and M. C. Ganiz, “A new hybrid semi-supervised algorithm for text classification with class-based semantics,” *Knowledge-Based Syst.*, vol. 108, pp. 50–64, 2016.
- [26] A. Cervantes, I.M. Galvan, P.Isasi, “AMPSO: A new Particle swarm method for nearest neighborhood classification,” *IEEE T [1]rans. On Systems, Man, and Cybernetics-Part B*, vol. 39, pp. 1082-1091, March 2009.