

Türkçe Twitter Verilerinde Duygu Analizi için Nitelik Seçim Yöntemlerinin Karşılaştırılması

Comparison of Feature Selection Methods for Sentiment Analysis on Turkish Twitter Data

Tuba PARLAR ve Esra SARAÇ

Mustafa Kemal Üniversitesi, Matematik Bölümü
Hatay, Türkiye
Adana Bilim ve Teknoloji Üniversitesi, Bilgisayar
Mühendisliği Bölümü, Adana, Türkiye
tparlar@mku.edu.tr, esrasarac@adanabtu.edu.tr

Selma Ayşe ÖZEL

Çukurova Üniversitesi
Bilgisayar Mühendisliği Bölümü
Adana, Türkiye
saozel@cu.edu.tr

Özetçe—Günümüzde internet ve sosyal medya insanların görüşleri hakkındaki önemli bir bilgi kaynağı haline gelmiştir. Ancak hızla artan miktardaki online belgeler, arzulanan görüş bilgilerini toplamayı ve analiz etmeyi oldukça zaman alıcı ve zor bir iş haline getirmiştir. Duygu analizi belgelerde ifade edilen duygunun sınıflandırılmasıdır. Genellikle, sınıflandırma işleminin performansını arttırmak amacıyla en değerli niteliklerin belirlenmesini sağlayan nitelik seçim yöntemleri uygulanır. Bu çalışmada, Türkçe Twitter verikümesi üzerinde Ki-kare, Bilgi Kazanımı, Sorgu Genişletme Sıralaması ve Karınca Kolonisi Optimizasyonu olmak üzere dört farklı nitelik seçim yönteminin performansı, Maksimum Entropi Modelleme sınıflandırma algoritması kullanılarak karşılaştırılmıştır. Böylelikle nitelik seçim yöntemlerinin Türkçe Twitter verileri üzerindeki duygu analizi işleminin performansı üzerindeki etkisi değerlendirilmiştir. Deney sonuçları, duygu analizinde Sorgu Genişletme Sıralaması ve Karınca Kolonisi Optimizasyon yöntemlerinin geleneksel nitelik seçim yöntemlerinden daha başarılı olduğunu göstermiştir.

Anahtar Kelimeler — *duygu analizi; nitelik seçimi; metin sınıflandırma.*

Abstract—The Internet and social media provide a major source of information about people's opinions. Due to the rapidly growing number of online documents, it becomes both time-consuming and hard task to obtain and analyze the desired opinionated information. Sentiment analysis is the classification of sentiments expressed in documents. To improve classification performance feature selection methods which help to identify the most valuable features are generally applied. In this paper, we compare the performance of four feature selection methods namely Chi-square, Information Gain, Query Expansion Ranking, and Ant Colony Optimization using Maximum Entropy Modeling classification algorithm over Turkish Twitter dataset. Therefore, the effects of feature selection methods over the performance of sentiment analysis of Turkish Twitter data are evaluated. Experimental results show that Query Expansion Ranking and Ant Colony Optimization methods outperform other traditional feature selection methods for sentiment analysis.

Keywords — *sentiment analysis; feature selection; text classification.*

I. GİRİŞ

İnternetin hızla yaygınlaşmasıyla web üzerindeki bilgi de büyük miktarlarda artmaktadır. Günümüz insanları artık duygu ve görüşlerini bloglar, forumlar ve sosyal ağlar üzerinden kolayca paylaşmaktadırlar. Paylaşılan duygu ve görüşler, iş dünyasının ürünler ve hizmetler konusundaki kararlarında önemli rol oynamaktadır. İnternetin bu denli yaygın kullanımından önce halkın genel görüşü ve tüketici bakış açısı ile ilgili bilgiler genellikle pahalı ve uzun soluklu araştırmalar sonucu elde edilirdi. Sosyal medya kullanımı toplumun her kesimini temsil etmemesine karşın giderek temsil gücü artmaktadır. Sosyal medya üzerinden marka, şirket ve kişi analizleri yaparak itibar ölçümleri yapılabilmekte, pozitif veya negatif yönde eğilimler duygu analizi ile belirlenebilmektedir. Bu analizler için makine öğrenmesi algoritmaları kullanılarak üst seviyede anlamlı bilgiler üretilebilmektedir. E-ticaret piyasası 2000 yılında 7,4 milyon dolar iken 2012'de 1 trilyon dolara ulaşmıştır¹. Aynı zamanda internet kullanıcılarının %81'inin bir ürünü almadan önce online araştırma yaptığı ve bunların %73 ile %87'sinin ürün hakkındaki yorumlardan etkilendikleri belirtilmiştir.

Duygu analizi için çoğunlukla makine öğrenmesi yaklaşımları kullanılmaktadır. Makine öğrenmesi ile yapılan duygu analizinde, önce analizi yapılacak yorum metinleri için nitelikler kümesi oluşturulur; sonra, elde edilen niteliklere göre yorum metinlerinin bir bölümü kullanılarak bir sınıflandırıcı eğitilir ve geri kalan yorumlar üzerinde test edilerek başarıları hesaplanır [1]. Sınıflandırmada kullanılan niteliklerin seçimi, eğitim verilerinden elde edilen terimler arasından yapılır. Nitelik seçimi, gereksiz niteliklerin elenmesiyle sınıflandırmanın doğruluğunu artırır aynı zamanda sınıflama işleminin süresini de azaltır. Makine öğrenme algoritması olarak Naive Bayes, Maksimum Entropi, Karar Ağaçları, Destek Vektör Makineleri gibi yöntemler kullanılabilir.

¹ <https://www.emarketer.com/Article/Ecommerce-Sales-Topped-1-Trillion-First-Time-2012/1009649>

Bu çalışmada metin sınıflandırma algoritması olarak Maksimum Entropi Modelleme (MEM) kullanılmıştır. MEM sınıflayıcısı [1, 2], Naive Bayes sınıflayıcıya benzer, olasılıksal bir modeldir. Verilen bağlamda bir sınıfa bağlı olan belge sınıflandırma sisteminin entropisini maksimize etmelidir. Maksimum Entropi sınıflayıcı bir çok Doğal Dil İşleme alanında başarıyla uygulanmıştır [1,3].

Duygu analizi problemlerinde makine öğrenme yöntemleri yaygın olarak kullanılmaktadır. Pang ve ark. [4] duyguların nitelikleri ve konular arasındaki farklılıkları belirlemek için birçok makine öğrenme yöntemini kullanarak karşılaştırmıştır. Mullen ve Collier [5] yorum belgelerini sınıflandırmak için destek vektör makinelerini kullanmışlardır. Sınıflandırma süreci genellikle iki aşamadan oluşur: subjektif ifadeleri belirlemek ve bunları sınıflandırıcıya girdi olarak vermek. Pang ve Lee [6] film yorumlarından çıkardıkları subjektif cümlelere dayanarak belgeleri sınıflandırmışlardır. Duric ve Song [1] film yorumları veri setinden [6] belge sıklığı ile nitelik seçimi gerçekleştirerek ve Maximum Entropi Modelleme ile bu yorumları sınıflandırmıştır. Stylios ve ark. [7] ürün yorumları üzerinden duygu analizi yapmak için Karınca Kolonisi (KKO) ve Parçacık Sürü Optimizasyonu (PSO) yöntemlerini nitelik seçimi için kullanarak %83.66 olan doğruluk değerini %90.59'a çıkarmışlardır. Benzer şekilde, Saraswathi ve Tamilarasi [8] KKO yöntemini klinik dökümanları üzerinde nitelik seçim yöntemi olarak kullanmıştır. Kullanılan KKO nitelik seçim yöntemi ile seçilen niteliklerin görüş madenciliği için kullanılan sınıflayıcının performansını artırdığı gözlemlenmiştir. Kaur ve ark. [9] ise KKO yöntemini destek vektör makineleri (SVM) sınıflayıcısı ile hibrid bir yöntem olarak kullanarak İngilizce Twitter verileri üzerinde duygu analizi yapmışlardır. Yalnız SVM ile %75.54 olarak aldıkları doğruluk değerini, hibrid yöntem ile (SVM+KKO) %86.74'e çıkarmışlardır.

Önceki çalışmalar incelendiğinde genellikle İngilizce Twitter yorumları üzerinde çalışmaların yapıldığı görülmektedir. Bu çalışmada ise, Türkçe Twitter yorumları üzerinde yapılan duygu analizinde farklı nitelik seçimi yöntemlerinin sınıflandırma performansına etkisinin karşılaştırılması hedeflenmiştir. II. bölümde kullanılan nitelik seçim yöntemleri özetlenmekte, III. bölümde kullanılan veri seti, performans değerlendirme kriteri ve deneysel sonuçlar yer almaktadır. IV. bölümde ise genel sonuçlar ve öneriler tartışılmaktadır.

II. KULLANILAN YÖNTEMLER

Bu çalışmada amaç Türkçe Twitter verilerinin duygu analizinde farklı nitelik seçim yöntemlerinin etkilerini karşılaştırmaktır. Bu amaçla yaygın kullanılan nitelik seçim yöntemlerinden Ki-kare (χ^2) ve Bilgi Kazanımı (IG) yöntemleri ile Sorgu Genişletme Sıralaması (QER) [10] ve sezgisel bir yaklaşım olan Karınca Kolonisi Optimizasyonu (KKO) yöntemleri kullanılmıştır.

A. Bilgi Kazanımı (IG)

Bilgi kazanımı yöntemi entropi değerini kullanarak her niteliğin karar verilecek sınıf için ayırt ediciliğini hesaplamaya çalışır. Bir niteliğin bilgi kazancını hesaplamak için öncelikle niteliğin her sınıf için ayrı ayrı entropisi hesaplanır. Bu değer

beklenen bilgi değerinden çıkarılır. Herhangi bir A niteliğinin bilgi kazanımı, niteliğin gözlemlendikten önceki ve sonraki entropi değerleri arasındaki fark olarak Denklem (3)'teki gibi hesaplanır. Bilgi kazanımı en yüksek olan nitelikler sınıflandırma işleminde kullanılır.

$$H(C) = -\sum_{i=1}^m P(c_i) \log_2 P(c_i) \quad (1)$$

$$H(C | A) = \sum_{j=1}^n \left(-P(a_j) \sum_{i=1}^m P(c_i | a_j) \log_2 P(c_i | a_j) \right) \quad (2)$$

$$IG(A) = H(C) - H(C | A) \quad (3)$$

B. Ki-kare (χ^2)

Ki-kare bir nitelik ve bir sınıf arasındaki bağımsızlığın derecesini ölçer. Eğer ilgili c sınıfındaki bir f niteliği düşük bir skora sahipse bu nitelik daha az bilgi içerdiği için göz ardı edilebilir [11]. Bir f niteliğinin ki-kare değeri Denklem (5)'e göre hesaplanır.

$$\chi^2(f, c) = \frac{N(AD - CB)^2}{(A + C)(B + D)(A + B)(C + D)} \quad (4)$$

$$\chi^2(f) = \sum_{i=1}^c P(c_i) \chi^2(f, c_i) \quad (5)$$

C. Sorgu Genişletme Sıralaması (QER)

Sorgu genişletme, bilgi erişimi için kullanılan önemli yöntemlerden biridir. Araştırmacılar, bilgi erişiminde ilgili belgeleri bulma işlevini geliştirmek için önemli oranda sorgu genişletme tekniklerini kullanmışlardır [12]. QER yöntemi, sorgu genişletme yöntemlerinden esinlenilerek geliştirilmiştir [10]. Denklem (6)'e göre hesaplanan terim ağırlıklarına göre nitelikler sıralanır. df_+^f , pozitif sınıfta f niteliğini içeren belge sayısını, df_-^f negatif sınıfta f niteliğini içeren belge sayısını, n^+ pozitif sınıftaki belge sayısını, n^- ise negatif sınıftaki belge sayısını ifade eder.

$$score_f = \frac{p_f + q_f}{|p_f - q_f|} \quad (6)$$

$$p_f = \frac{df_+^f + 0.5}{n^+ - 1.0} \quad (7)$$

$$q_f = \frac{df_-^f + 0.5}{n^- - 0.5} \quad (8)$$

Böylece, bu hesaplamalar ile hangi niteliklerin daha ayırt edici olduğu belirlenir.

D. Karınca Kolonisi Optimizasyonu

Çalışmada kullanılan bir diğer nitelik seçim yöntemi ise sezgisel bir yöntem olan Karınca Kolonisi Optimizasyon (KKO) yöntemidir. Kullanılan KKO yönteminde, her bir nitelik bir düğümü ifade etmektedir ve tüm düğümler bir birlerine tam olarak bağlı ve bağımsız düğümlerdir. Düğümler (nitelikler) kendi $P_k(i)$ seçim olasılıklarına göre seçilmektedir. Bu değer Denklem (9)'da verilmiştir.

$$P_k(i) = \frac{[\tau(i)]^{\alpha} * [\eta(i)]^{\beta}}{\sum_{l \in N_k^i} [\tau(l)]^{\alpha} * [\eta(l)]^{\beta}} \quad (9)$$

Denklem (9)'da, $\eta(i)=df(i)$, öyle ki $df(i)$ nitelik i için doküman frekansdır ve karıncalar için sezgisel bilgiyi ifade eder. $\tau(i)$ ise nitelik i için feromon değerini ifade eder. N_k^i değeri k karıncasının henüz ziyaret etmediği görülebilir düğümlerin listesini tutar. [13]' e göre başlangıç feromon değeri 10 olarak, α ve β parametreleri ise 1 olarak belirlenmiştir. α parametresi sezgisel değerlerin ağırlığını belirlerken, β parametresi ise feromon bilgisinin ağırlığını belirlemektedir. Tüm karıncalar kendi turlarını tamamladıktan sonra, feromon değerleri Denklem (10)'da verilen global güncelleme kuralına göre güncellenmektedir.

$$\tau(i) = \rho * \tau(i) + \sum_{k=1}^m \Delta \tau_k(i) \quad (10)$$

Denklem (10)'da ρ feromon buharlaşma katsayısıdır ve o oranda feromonu halihazırda bulunan feromondan buharlaştırarak azaltır, m ise karınca sayısıdır. Her bir karıncanın i düğümüne katkıda bulunduğu feromon miktarı $\Delta \tau_k(i)$ ise Denklem (11)'deki gibi hesaplanır.

$$\Delta \tau_k(i) = \begin{cases} B_k * 2 * L_k & \text{eger dugum } i \text{ karınca } k \text{ tarafından kullanılmış ve } k \text{ elit ise} \\ B_k * L_k & \text{eger dugum } i \text{ karınca } k \text{ tarafından kullanılmış} \\ 0 & \text{diğer durumlarda} \end{cases} \quad (11)$$

Denklem (11)'de, L_k k karıncasının oluşturduğu altkümenin (çözümün) F-skor değeridir ve B_k da birim feromon değeridir. Bunun anlamı, F-skor değeri yüksek olan karıncaların seçtiği düğümlere (niteliklere) daha fazla feromon katkısı olacaktır böylece daha sonraki iterasyonlarda iyi çözümler (nitelik alt kümeleri) içerisinde yer alan düğümlerin seçim olasılığı daha yüksek olacaktır.

Her karınca belirli sayıda nitelik seçene kadar, seçilmemiş niteliklerin seçim olasılıkları Denklem (9)'daki gibi hesaplanır. Olasılık hesabından sonra, sıradaki niteliği seçmek için rulet tekeri seçim algoritması kullanılmıştır[14].

Tüm karıncalar alt küme seçim işlemlerini tamamladıktan sonra seçilen niteliklerin sınıflandırma performanslarını ölçmek için, seçilen nitelikler ile öğrenme ve test veri kümeleri kullanılarak sınıflandırmanın doğruluğu hesaplanır. Elde edilen F-skor değerlerine göre seçilen niteliklerin feromon değerleri güncellenir. Eğer bir karınca iyi bir çözüm üretiyorsa, sınıflandırma performansı yüksek olacaktır. Feromon değerleri güncellendiğinde, iyi sonuç üreten karınca tarafından seçilen niteliklerin feromon değerleri daha yüksek olur. Bir sonraki

nitelik seçim aşamasında, bu feromon değerleri karıncayı daha iyi çözümler bulma konusunda yönlendirir. İterasyonlar sonucunda, seçilen en iyi nitelik alt kümesi daha önce görülmemiş test verisinde denenmek üzere döndürülür. Yapılan çalışmada deneysel olarak 30 karınca kullanılmıştır. İterasyon sayısı ise 250 olarak belirlenmiştir. Daha önceki çalışmalarda 30 karıncanın, 250 iterasyonda iyi sonuçlar veridiği gözlenmiştir[15].

III. KULLANILAN VERİ SETİ VE DENEYLER

A. Veri seti

Önerilen çalışmayı değerlendirmek için deneyler telekomunikasyon sektöründeki özel bir şirkete ait Türkçe Twitter verileri [16] üzerinde yapılmıştır. Veri seti üç farklı sınıfa ait (pozitif, negatif, nötr) toplam 3000 tweet içermektedir². Çalışmada pozitif ve negatif olarak toplam 2043 twitter verisi kullanılmıştır. Nitelikler bag-of-words yöntemi ile yalnızca alfabetik karakterler içerecek şekilde çıkarılmış ve terim frekansına göre nitelik vektörleri oluşturulmuştur. Twitter verileri, sınıf dağılımları dengeli olacak biçimde beş katlı çapraz doğrulama yapmak üzere yaklaşık olarak %80'i eğitim ve %20'si test verileri olarak ayrılmıştır.

B. Performans Değerlendirme

Sınıflandırma deneylerinin performansı F-Skor (F) ile değerlendirilmiştir. F-Skor, sınıflandırıcının hem kesinliğini hem de duyarlılığını göz önünde bulunduran bir değerlendirme ölçütüdür. Kesinlik (P), sınıflandırıcının bir sınıf için yaptığı sınıflandırmalarda ne kadar hassas olduğuyla ilgili bir değerlendirme ölçütüdür. Duyarlılık (R) ise örneklerin düzgün sınıflandırılmasını değerlendiren ölçüttür. F-Skor ise kesinlik ve duyarlılık değerlerinin harmonik ortalamasıdır.

$$F = 2 \frac{P \times R}{P + R} \quad (12)$$

C. Metin Sınıflandırma

Bu çalışmada Weka³ veri madenciliği yazılım paketi içerisindeki Logistic Regression (LR) sınıflayıcısı kullanılmıştır. Weka'nın kullandığı bu sınıflayıcı multinomial bir lojistik regresyon modelini tanımlar. Multinomial lojistik regresyon modelinin Maksimum Entropi Modellemesine (MEM) eşdeğer olduğu bilinmektedir[17].

Python NLTK⁴ kütüphanesinden yararlanılarak kodlar geliştirilmiş olup, beş katlı çapraz doğrulama için eğitim ve test verileri hazırlanmıştır.

D. Deneysel Sonuçlar

Nitelik seçim yöntemlerinin performanslarını karşılaştırmak amacıyla öncelikle seçim yapılmadan Türkçe Twitter veri kümesi üzerinde Maksimum Entropi Sınıflayıcı ile beş katlı çapraz doğrulama yapılarak sınıflandırma yapılmıştır. Daha sonra azaltılmış nitelik sayıları ile geleneksel yöntemlerden ki-kare (χ^2) ve bilgi kazanımı (IG) yöntemleri ile Sorgu

² <http://www.kemik.yildiz.edu.tr/?id=28>

³ <http://www.cs.waikato.ac.nz/ml/weka>

⁴ <http://nltk.org>

Genişletme Sıralaması (QER) ve Karınca Kolonisi Optimizasyonu (KKO) yöntemleri Türkçe Twitter verileri üzerinde test edilmiştir. Deneyler 10, 50, 100, 250 ve 500 nitelik sayıları ile Maksimum Entropi Sınıflayıcı ile beş katlı çapraz doğrulama yaparak gerçekleştirilmiştir.

Nitelik seçimi yöntemleri uygulanmadan tüm nitelikler için sınıflandırma performansı ortalama F-Skor ile 0.6913 olarak elde edilmiştir. Tablo 1’de seçilen dört farklı nitelik seçim yönteminin farklı nitelik sayıları ile sınıflandırma sonuçları verilmiştir. Tablo 1’de görüldüğü gibi, ki-kare ve bilgi kazanımı yöntemlerinde seçilen ilk 100 ve 250 nitelik sayılarında, nitelik seçimi uygulanmadan gerçekleştirilen sınıflandırma sonucuna göre daha iyi sonuç vermiştir. Bununla beraber, QER ve KKO yöntemlerinde ise sınıflandırma başarılarının 100 ve 250 nitelik sayılarında geleneksel yöntemleri geçtiğini ve 250 ve 500 nitelik sayılarında ise 0.1’e yakın bir farkla sınıflandırma performansını artırdığı görülmektedir. Tablo 1’de gözlemlenebileceği gibi ki-kare ve bilgi kazanımı nitelik seçim yöntemlerinin sınıflandırma F-Skor değerleri birbirine oldukça yakındır.

TABLO I. AZALTILMIŞ NİTELİK SAYILARI İLE NİTELİK SEÇİM YÖNTEMLERİNİN SINIFLANDIRMA SONUÇLARI

Nitelik sayıları	Nitelik Seçim Yöntemleri			
	χ^2	IG	QER	KKO
10	0.6280	0.6149	0.5580	0.6988
50	0.6888	0.6876	0.6300	0.7507
100	0.6972	0.6949	0.7145	0.7609
250	0.6936	0.6962	0.7699	0.7888
500	0.6963	0.6953	0.7793	0.7656
Seçimsiz Tüm Nitelikler	0.6913			

IV. SONUÇ VE ÖNERİLER

Yapılan çalışmanın amacı nitelik seçimi yöntemlerinin Türkçe Twitter verilerinin duygu analizindeki sınıflandırma performanslarına etkilerini karşılaştırmaktır. Sınıflandırma sürecinde, önceki bölümlerde açıklanan Maksimum Entropi sınıflayıcı ile sonuçlar değerlendirilmiştir. Nitelik seçimi yapılmadan ve her bir nitelik seçim yöntemine göre seçim yapıldıktan sonra yapılacak sınıflandırma deneylerinin performansı F-skor değerlendirme ölçütlerine göre karşılaştırılmıştır. Ayrıca kullanılan nitelik seçim yöntemlerinin performansı ki-kare ve bilgi kazanımı yöntemleri gibi yaygın kullanılan nitelik seçim yöntemlerinin performansları ile karşılaştırılmıştır.

Tablo 1’de de görüldüğü üzere, nitelik seçiminin sınıflandırma performansı üzerine oldukça olumlu etkileri vardır. Nitelik sayılarını azaltmak sınıflandırma performansını artırırken sınıflandırma süresini de keskin bir şekilde azaltmıştır. Bu da nitelik seçiminin sistemin genel performansı için kullanılabilir bir yöntem olduğunu göstermektedir. Nitelik seçimi yöntemleri uygulanmadan tüm nitelikler için sınıflandırma performansı ortalama F-Skor ile 0.6913 olarak elde edilmiştir. Elde edilen tüm sonuçlar Tablo 1’de verilmiştir. Önerilen her iki yöntemin de geleneksel yöntemlere göre daha

iyi sonuçlar verdiği gözlemlenmiştir. Tüm yöntemler karşılaştırıldığında Karınca Kolonisi Optimizasyonunun (KKO) en yüksek F-Skor değerine sahip olduğu görülmüştür. Bu sonucun elde edilmesi doğaldır, çünkü KKO nitelikleri seçerken sınıflandırıcının performansını dikkate alarak seçim yapmıştır. QER sınıflayıcı bağımsız bir yöntem olmasına karşın, niteliklerin farklı sınıflardaki olasılıklarını dikkate alarak seçim yaptığından başarılı olmuştur. Bu nedenle sonraki çalışmamızda en başarılı iki yöntem olan Sorgu Genişletme Sıralaması (QER) ve Karınca Kolonisi Optimizasyonu (KKO) nitelik seçim yöntemlerini farklı sınıflayıcılar ile işlem zamanını göz önünde bulundurarak performans analizlerini yapmayı planlamaktayız.

KAYNAKLAR

- [1] Duric, A. and Song, F., “Feature selection for sentiment analysis based on content and syntax models”, Proc. of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, ACL-HLT, Oregon, USA, pp. 96–103, 2011.
- [2] Manning, C.D. and Schütze, H., *Foundations of Statistical Natural Language Processing*. Cambridge, MA, USA: MIT Press, 1999.
- [3] Liu, B. and Zhang, L., “A survey of opinion mining and sentiment analysis. In C.C. Agarwal, and C. Zhai (Eds.), Mining text data (pp. 415-463) Springer, 2009.
- [4] Pang, B., Lee, L., and Vaithyanathan, V., “Thumbs up? Sentiment classification using machine learning techniques”, In: Proceedings of the conference on empirical methods in natural language processing”, Morristown, pp. 79–86, 2002.
- [5] Mullen, T. and Collier, N., “Sentiment Analysis using Support Vector Machines with Diverse Information Sources”, Proceedings of EMNLP 2004, pp. 412-418, 2004.
- [6] Pang, B. and Lee, L., “A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts”, Proc. of the annual meeting for the Association of Computational Linguists, 2004.
- [7] Stylios, G., Katsis, C.D., ve Christodoulakis, D., “Using Bio-inspired Intelligence for Web Opinion Mining”. *International Journal of Computer Applications*, 2014; 87(5).
- [8] Saraswathi, K. Ve Tamilarasi, A., “Ant Colony Optimization Based Feature Selection for Opinion Mining Classification”, *Journal of Medical Imaging and Health Informatics* 6(7):1594-1599, 2016.
- [9] Kaur, J., Sehra, S.S. ve Sehra, S.K., “Sentiment Analysis of Twitter data using Hybrid Method of Support Vector Machine and Ant Colony Optimization”, *International Journal of Computer Science and Information Security (IJCSIS)*, Vol. 14, No. 7, July 2016.
- [10] Parlar, T. and Özel, S.A., “A New Feature Selection Method for Sentiment Analysis of Turkish Reviews”, Proceedings of the International Symposium on INnovations in Intelligent SysTems and Applications, Sinaia, Romania, 2016, pp. 1-6.
- [11] Nicholls, C. and Song, F., “Comparison of feature selection methods for sentiment analysis” In: Proceedings of the 23rd Canadian conference on Advances in Artificial Intelligence (AI’10), pp. 286-289, 2010.
- [12] Robertson, S.E. and Sparck Jones K., “Relevance Weighting of Search Terms”, *JASIS* 27, pp. 129-146, 1976.
- [13] Dorigo, M., and Stützle, T., *Ant Colony Optimization*. MIT Press, Cambridge, MA., 2004.
- [14] Bäck, T., *Evolutionary Algorithms in Theory and Practice*, Oxford University Press. New York, USA 1996.
- [15] Saraç, E. ve Özel, S.A., An Ant Colony Optimization Based Feature Selection for Web Page Classification. *The Scientific World Journal Volume 2014*.
- [16] Çetin, M. and Amasyalı, M.F., “Supervised and traditional term weighting methods for sentiment analysis”, In proceedings of the Signal Processing and Communications Applications Conference (SIU), 2013.
- [17] Mount J. The equivalence of logistic regression and maximum entropy models, <http://www.winvector.com/dfiles/LogisticRegressionMaxEnt.pdf> (2011, accessed August 2015).