

Word2Vec Modelini Kullanarak Türkçe ve İngilizce Twitter Mesajlarının Duygu Analizi Sentiment Analysis of Turkish and English Twitter Feeds Using Word2Vec Model

Abdullah Ammar KARCIOĞLU

Bilgisayar Mühendisliği Bölümü
Ege Üniversitesi
İzmir, Türkiye

E-mail: abduallah.ammar.karcioglu@ege.edu.tr

Tolga AYDIN

Bilgisayar Mühendisliği Bölümü
Atatürk Üniversitesi
Erzurum, Türkiye

E-mail: atolga@atauni.edu.tr

Özetçe— Sosyal medya günlük hayatın önemli bir parçası haline gelmiştir. Sosyal medya servislerinin en popülerlerinden birisi olan Twitter ile kullanıcılar Twitter mesajlarını kullanarak duygu ve düşüncelerini tüm dünyaya ifade etmektedirler. Bu nedenle Twitter mesajları duygu analizi çalışmalarında önemli bir kaynak haline gelmiştir. Bu çalışmada, İngilizce ve Türkçe Twitter mesajlarındaki etiketli verilerin sınıflandırılmasında Word2Vec modelinin uygulanması ve mesajlar üzerinde kök alma işleminin Word2Vec modeline olan etkisi araştırılmaktadır. Çalışmamızda İngilizce ve Türkçe olmak üzere iki farklı veri kümesi bulunmaktadır. Her bir veri setine Twitter mesajlarının kökleri alınmamış hali ve kökleri alınmış hallerine öznitelik çıkarma yöntemlerinden kelime torbası (BOW) ve Word2Vec modelleri uygulanmıştır. Python programlama dilinde gerçekleştirilen bu çalışmada başarı yüzdeleri, scikit-learning sınıflandırma algoritmalarından Lineer SVM ve Lojistik Regresyon uygulanarak karşılaştırılmış ve duygu analizi sınıflandırmalarının sonuçları gösterilmiştir.

Anahtar Kelimeler — *Word2Vec, Duygu Analizi, Metin Sınıflandırılması, Kelime Yerleştirmeleri, Makine Öğrenmesi.*

Abstract— Social media has become an important part of daily life. With twitter, one of the most popular social media services, users express their feelings and thoughts to the whole world using twitter posts. For this reason, twitter feeds have become an important source of sentiment analysis. In this study, the apply of Word2Vec model in the classification of labeled data in English and Turkish Twitter feeds and the effect of getting root on feeds to Word2Vec model are investigated. Our study has two different data sets, English and Turkish. BOW and Word2Vec models were applied to each data set in the case where twitter feeds were not get roots and get roots were extracted. In this study, which is implemented in the Python programming language, the success percentages are compared by applying the scikit-learn classification algorithms, Linear SVM and Logistic Regression.

Keywords — *Word2Vec, Sentiment Analysis, Text Classification, Word Embeddings, Machine Learning.*

I. GİRİŞ

Twitter, ayda 330 milyon aktif kullanıcısıyla ve günde 500 milyondan fazla tweet mesajıyla kullanıcıların duygu ve düşüncelerini diğer kullanıcılarla kolay, hızlı ve güvenilir bir

şekilde paylaşmalarını sağlayan sosyal medya platformlarından biridir [1]. Uzun ve yapılandırılmış içeriğe sahip geleneksel belgelerden farklı olarak, Twitter mesajları 140 karakterle sınırlıdır [2]. Bu nedenle mikro-blog servisi olarak adlandırılmıştır. Twitter bir blog servisine benzemektedir, fakat paylaşılan karakter sayısı sınırlı olduğundan bir blog servisi gibi değildir [3]. Duygu analizi üzerine literatürde Twitter üzerinden belirli markalar için yazılanların, o marka için iyi mi, kötü mü veya duygu belirtmeyen bir cümle mi sorularından, makine öğrenmesi yöntemlerini kullanarak, geri bildirim elde etme üzerine çalışmalar yapılmıştır [7]. Ayrıca makine öğrenmesi yöntemlerinden denetimli öğrenme yaklaşımı kullanılarak sosyal medyada Twitter üzerine duygu analizi çalışması yapılmıştır ve bazı gıda firmalarının çeşitli ürünlerine ait yapılan yorumlardan oluşturulan veri setlerine makine öğrenmesi algoritmaları uygulanarak dengeli ve dengesiz veri setlerinin başarı yüzdeleri kıyaslanmıştır [8].

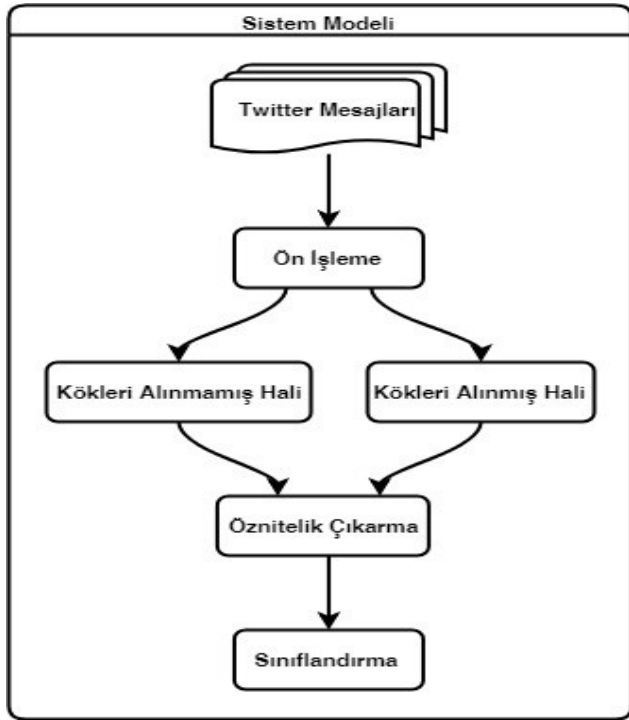
Literatürde konu modelleme üzerine LSA(Latent Semantic Analysis), PLSA(Probabilistic Latent Semantic Analysis), LDA(Latent Dirichlet Allocation) gibi birçok algoritma ve yöntem geliştirilmiştir [4]. Ancak Mikolov ve arkadaşları tarafından geliştirilen Word2Vec modeline ilgi zaman içinde artmıştır ve bu modelde sözcükler arasında anlamsal ilişkinin olduğu ve doğal dil işleme çalışmalarında çok kullanışlı olduğu gösterilmiştir [5]. Bu nedenle Word2Vec modeli bize Twitter mesajlarının sınıflandırılmasında daha doğru sonuçlar üretecektir. Ayrıca bu çalışmada, veri setlerinin her birinin köklerini alarak bu veri setlerini kendi içlerinde kıyasladığımızda kök alma işleminin kelimeler arasındaki anlamsal ilişkiye olan etkisini incelemek amaçlanmaktadır.

Word2Vec modelinde, sabit bir bağlam penceresi içinde bir kelimenin görülme ihtimalini en üst düzeye çıkararak sözlükteki her bir kelime için sığ sinir ağından yoğun gerçek değerli bir vektör öğrenmenin mümkün olduğu gösterilmiştir [6]. Bu çalışmada metinlerin basit temsillerini kullanarak seçmiş olduğumuz yöntemlerin etiketli mesajların sınıflandırılmasında iyi sonuçlar verebileceği gösterilmiştir. Önerilen sistem her biri belirli bir metin temsiline (BOW veya Word2Vec) sahip üç temel modelden oluşmaktadır. Bu modellerimiz, verilerimizin kökleri alınmış ve kökleri alınmamış hallerine uygulanmıştır.

Temel sınıflandırıcı olarak makine öğrenmesi algoritmalarından Linear SVM ve Lojistik Regresyon kullanılmıştır.

II. SİSTEM MODELİ

Çalışmamızda farklı metin temsili (öznelik çıkarma) teknikleriyle üç farklı modeli içeren bir sistem sunulmuştur. Tüm temel sınıflandırıcılar aynı veri kümesinde eğitim almıştır, ancak farklı bir temsil veya öznelik çıkarma yöntemiyle gerçekleştirilmiştir. Tüm modellerde sınıflandırıcılarımız Linear SVM ve Lojistik Regresyon 'dur. Birinci modelde Twitter mesajları kelime torbası (BOW) kullanılarak temsil edilmiştir. İkinci modelde Twitter mesajları kelime yerleştirmelerinin (Word2Vec) ortalamasıyla temsil edilmiştir. Üçüncü modelde Twitter mesajlarının ağırlıklı kelime yerleştirmelerinin (Word2Vec) ortalamasıyla temsil edilmiştir.



Şekil 1: Sistem Modeli

A. Metin Temsilleri

Bu çalışmada Twitter mesajları, metin temsillerinin üç çeşidi kullanılarak modellenmiştir. Birincisi, terim frekansı - ters metin frekansı (TF-IDF) tarafından ağırlıklı bir kelime torbası (BOW) modelidir. İkincisi, cümledeki tüm kelime yerleştirmelerinin (Word2Vec) ortalamasını alarak bir cümleyi temsil eder. Üçüncüsü, tüm kelimelerin ağırlıklı kelime yerleştirmelerinin (Word2Vec) ortalamasını alarak bir cümleyi temsil eder.

1. Kelime Torbası

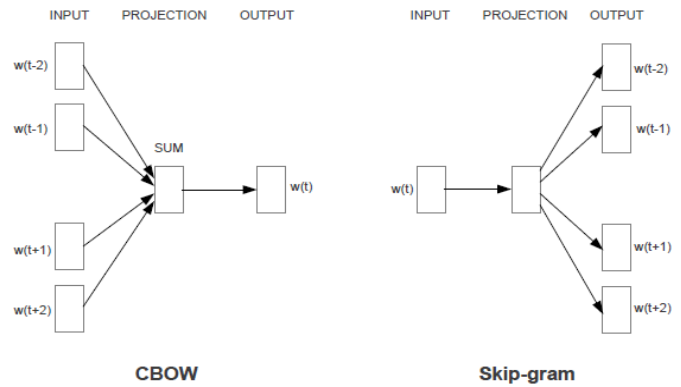
BOW modelinde, önceden hazırlanmış bir sözlük göz önüne alındığında, bir metin bir dizi kelimeyle temsil edilir, bu gösterim ikili olabilir. Bir kelime, metinde geçiyor ise 1 değerini, aksi halde 0 değerini alır [9]. TF (Term Frequency), bir terimin bir metindeki toplam terimlere oranı iken, IDF (Inverse Document Frequency) ise tüm metinlerin kaç tanesinde seçili terimimizin bulunduğudır. Kısaca TF-IDF hesabı sırasında iki kritik sayı bulunmaktadır. Bunlardan birincisi o anda ele alınan

dokümandaki terimin sayısı diğeri ise bu terimi derlemde içeren toplam doküman sayısıdır [10]. Pang ve arkadaşları her bir kelimenin birer vektör olarak temsil edildiği BOW modeli alanında öncülük etmişlerdir [14]. BOW modelinin sorunlarından birincisi, kelime gösteriminin yüksek boyutta olması ve yedek bilgi sınıflandırma performansının sınırlı olmasıdır. İkincisi, bu modelin kelimeler arasındaki anlamsal ilişkiyi yakalayamamasıdır [13]. Ancak Word2Vec modeli $\text{vec}(\text{Berlin}) - \text{vec}(\text{Germany}) + \text{vec}(\text{France}) = \text{vec}(\text{Paris})$ gibi kelimeler arasındaki anlamsal ilişkiyi yakalayabildiği için [14] çalışmamızda Word2Vec modeli kullanılmıştır.

2. Word2Vec

Mikolov ve arkadaşları giriş katmanı, projeksiyon katmanı ve yakındaki kelimeleri tahmin etmek için bir çıktı katmanı içeren sığ bir sinir ağı mimarisi kullanarak her kelimenin vektör gösterimini öğrenmek için geliştirilen 'Skip-Gram' ve 'CBOW' modellerine sahip Word2Vec modelini tanıtmışlardır [11,12]. Word2Vec modeli doğrusal olmayan dönüşümlerden kaçındığı için eğitimi son derece verimli hale getirir [14]. Word2Vec sözlükteki milyonlarca kelimedenden ve milyarlarca kelimeye sahip çok büyük veri setlerinden gömülü kelime vektörlerini öğrenmeyi sağlar [13] ve "(1)"de gösterildiği gibi hesaplanmaktadır. Bu çalışmada, vektörlerin içerik (skip-gram) veya kelime (cbow) tahminini yapmak için sinir ağı eğitilerek öğrenildiği Word2Vec modeli kullanılmıştır. Ayrıca, cbow modeli eğitim için skip-gram modelinden çok daha hızlı olmasına rağmen çalışmamızda veri seti küçük olan sistemlerde daha başarılı sonuçlar üreten skip-gram modeli [11] kullanılmıştır.

$$\frac{1}{T} \sum_{t=1}^T \sum_{j=-k}^{j=k} \log_p(w_{t+j} | w_t) \quad (1)$$



Şekil 2: Word2Vec Model Mimarisi [12]

B. Sınıflandırıcılar

Python programlama dilinde gerçekleştirilen çalışmamızda her iki temel sınıflandırıcı için de varsayılan parametre değerleriyle scikit-learn kütüphanesi [15] kullanılmıştır.

1. Lojistik Regresyon

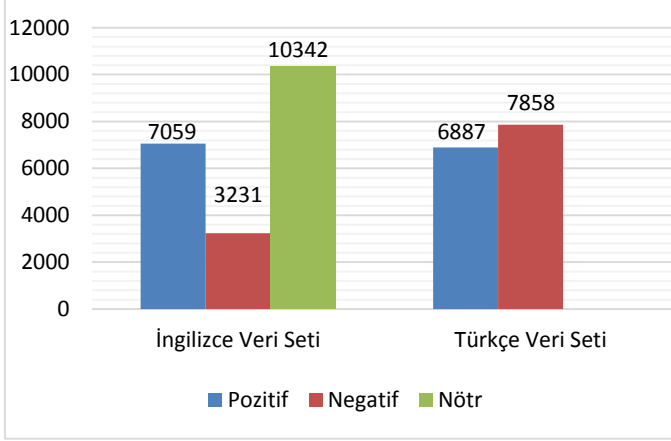
Gerçek değerli giriş vektörüne sahip bir olasılık sınıflandırma modelidir. Giriş vektörünün boyutları özellikler olarak adlandırılır ve ilişkili özellikler için herhangi bir sınırlama getirilmez. Lojistik fonksiyonu, özelliklerin doğrusal bir birleşimidir. Çıktı genellikle ikili formattadır, ancak lojistik regresyon çok sınıflı sınıflandırma problemlerine de uygulanabilir [16].

2. Lineer SVM

Sınıflandırma için Lineer SVM fikri verinin n boyutlu özellik uzayında veriyi iki sınıfa ayıran en uygun alt düzlemi (hat/alan sınırlayıcı) bulmayı hedefler. Bu sistem ile SVM üzerindeki en uygun hiper düzlem çözümü yerel bir optimuma sahip değildir ve bunun sonucunda bir tek çözüm olur. SVM kolayca uygulanabilir yöntemdir ve mevcut veri örneğinin sınırlamaları içinde, yüksek boyut problemini çözmek için kullanılan uygun yöntemlerden biridir [16].

III. VERİ SETLERİ

Çalışmamızda duygu analizi sınıflandırılmasında İngilizce [9] (twitter-2016test/20632 adet tweet) Türkçe [19] (14657 adet tweet) olmak üzere bağımsız iki veri seti kullanılmıştır.



Şekil 3: Etiketli Veri Setlerinin Dağılımı

1. Veri Ön İşleme

Twitter mesajlarındaki @ karakteriyle başlayan kullanıcı adı, (#) diyez karakteriyle başlayan hashtag'ler, URL adresleri, tüm rakamlar, noktalama ve imge işaretleri kaldırılmıştır. Daha sonra bu veri setini dizgi parçalamadan geçirip durak kelimelerini kaldırıp temizlenmiş veri seti elde edilmiştir.

2. Kök Alma

İngilizce veri setinde kök alma işlemi Python programlama dilindeki en çok kullanılan kök alma kütüphanelerinden 'Porter Stemmer' ile gerçekleştirilmiştir. Türkçe veri setinde kök alma işlemi için de Python programlama dilinde kök alma kütüphanesi olan 'Turkish Stemmer' ile gerçekleştirilmiştir.

3. K - Çapraz Doğrulama

Model performansı tahmini için geliştirilen K-çapraz doğrulamada veri seti, analiz için rastgele alt kümelere bölünür. K-çapraz doğrulamada, veri seti rastgele olarak "(2)"de gösterildiği gibi k tane alt kümelere bölünür. Her alt küme, test verileri olarak kabul edilirken, kalanlar eğitim verileri olarak kullanılır. Bu süreç, k kere tekrarlanır. Son olarak, tüm işlemlerin geçerlilik sonuçlarının tek ve son bir tahmin için ortalaması alınır [18]. Çalışmamızda k değerimizi 10 seçerek tüm veri setimizin %90'ı eğitim veri seti, %10'u ise test veri seti olarak kullanılmıştır.

$$Sonuç = \frac{\sum_{i=0}^k SF(t_i, VK - t_i)}{k} \quad (2)$$

4. Kelime Yerleştirmeleri

İngilizce veri setinin eğitilmesi için Word2Vec modeliyle eğitilerek oluşturulmuş 3 milyon kelimeye sahip 300 boyutlu GoogleNews-vectors-negative300 derlemi [20] kullanılmıştır. Türkçe veri setinin eğitilmesi için de Boğaziçi Üniversitesi tarafından Word2Vec modeliyle eğitilerek oluşturulmuş olan 416051 kelimeye sahip 300 boyutlu wiki-tr derlemi [21] kullanılmıştır. Çalışmamızda varsayılan değerleriyle uygulanan Word2Vec modelinde, bir Twitter mesajındaki tüm kelimelerin önceden eğitilmiş vektörleri olmadığından, rastgele başlatılan boyut uzunluğu 300 olan vektör tayin edilmiştir.

IV. DENEYSEL SONUÇLAR

TABLO I. İNGİLİZCE VERİ SETİ SONUÇLARI

Sistem Modeli	Kökleri Alınmamış Hali		Kökleri Alınmış Hali	
	Ortalama Anma %	Doğruluk %	Ortalama Anma %	Doğruluk %
1.Model+LineerSVM	57,91	59,84	56,66	59,25
1.Model+LojistikReg	52,20	54,72	52,06	54,48
2.Model+LineerSVM	54,27	59,76	53,35	57,68
2.Model+LojistikReg	54,95	59,87	52,64	57,31
3.Model+LineerSVM	51,68	57,44	50,11	55,21
3.Model+LojistikReg	52,78	57,63	50,26	55,48

Tablo 1'de gösterildiği gibi, veri setlerimizi değerlendirmek, karşılaştırmak ve kıyaslamak için ortalama anma (average recall) ve doğruluk (accuracy) metrikleri kullanılmıştır. Yaptığımız çalışma sonucu, İngilizce Twitter mesajlarında kök alma işlemi gerçekleştirildiğinde üç modelde de doğruluk ve anma yüzdelerinin düştüğü gözlemlenmiştir. En yüksek doğruluk yüzdesi %59,87 ile 2. modelde elde edilmiştir. Kök alma işlemi uygulandığında anlamsal ilişkiyi kaybettiği için Word2Vec modelinde doğruluk değerleri azalarak daha düşük doğruluk ve anma yüzdeleri elde edilmiştir.

TABLO II. TÜRKÇE VERİ SETİ SONUÇLARI

Sistem Modeli	Kökleri Alınmamış Hali		Kökleri Alınmış Hali	
	Ortalama Anma %	Doğruluk %	Ortalama Anma %	Doğruluk %
1.Model+LineerSVM	60,47	65,62	60,14	65,18
1.Model+LojistikReg	60,99	64,50	59,56	64,10
2.Model+LineerSVM	50,14	51,23	50,10	51,19
2.Model+LojistikReg	50,21	51,02	50,13	50,78
3.Model+LineerSVM	50,07	50,92	50,19	51,17
3.Model+LojistikReg	50,28	51,07	62,86	63,51

Tablo 2'de gösterildiği gibi, Türkçe Twitter mesajlarında kök alma işlemi gerçekleştirildikten sonra başarı yüzdelerinin sadece ilk iki modelde düştüğü, fakat son modelde %62,86 ile anma yüzdesinin ve %63,51 ile doğruluk yüzdesinin arttığı gözlemlenmiştir. İngilizce veri setinde aynı veri setini kullanan başka bir çalışmada [20] elde edilen anma yüzdesinin %52,70'den %57,91'e, doğruluk yüzdesinin ise %54,20'den %65,62'ye yükseldiği gözlemlenmiştir. Türkçe veri setinde elde edilen doğruluk yüzdesi, aynı veri setini kullanan başka bir çalışmada [20] BOW modeliyle elde edilen başarı yüzdesinin %62,48'den %65,62'ye yükseldiği gözlemlenmiştir. Derlem boyutlarının farklı olması nedeniyle İngilizce veri setlerinin köklerinin alınmamış halinde Word2Vec modeliyle elde edilen doğruluk yüzdesi BOW modeliyle elde edilen doğruluk

yüzdesinden daha yüksek elde edilirken Türkçe veri setinde daha düşük elde edilmiştir.

Ancak Lojistik Regresyon çok sınıflı problemlerle başa çıkmak için, eğitim algoritmasında tekli ve iki taraflı yaklaşımlar kullandığı için üç etikete sahip İngilizce veri setinde elde ettiğimiz sonuçlarda bu sınıflandırma algoritmasıyla daha yüksek başarı yüzdesi elde edilmiştir. Lineer SVM ise iki sınıf arasındaki doğruluğu maksimize eden karar sınırını bulduğu için bu sınıflandırma algoritmasıyla iki etikete sahip Türkçe veri setinde elde ettiğimiz sonuçlarda daha yüksek başarı yüzdesi elde edilmiştir.

V. SONUÇ VE GELECEK ÇALIŞMALAR

Bu çalışmada, sosyal medya üzerine doğal dil işleme çalışmalarını yapmak hedeflenmiştir. Twitter mesajlarında duygu analizini gerçekleştirmek için kelimelerin temsil edilme yöntemlerini kullanarak, kök alma işleminin Word2Vec modeline olan etkisi incelenmiştir.

Yaptığımız çalışma sonucu kök alma işleminin İngilizce metinler üzerinde kelime yerleştirmeleri uygulandığında başarıyı düşürdüğü gözlemlenmiştir. Diğer bir yandan, Türkçe metinler için sunmuş olduğumuz ilk iki modelde kök alma işleminin başarıyı düşürdüğü, ancak Türkçe'nin sondan eklemeli (agglutinative) bir dil olmasına rağmen kök alma işleminin son modelde başarıyı artırdığı tespit edilmiştir. Böylece kök alma işlemi uygulanarak Word2Vec modeliyle İngilizce metinlerde başarının azalabileceğini, Türkçe metinlerde ise uygulanan modele göre başarının artırılabilirliği gösterilmiştir. Sonuç olarak, doğal dil işleme çalışmalarında kök alma işleminin duygu analizi çalışmalarında uygulanmaması gerektiği anlaşılmıştır ve bu işlemin başarıyı düşürdüğü gözlemlenmiştir.

Gelecekteki çalışmalarımızda daha başarılı sonuçlar elde etmek için amacımız etiket sayıları eşit olan farklı veri setlerini kullanmak, verilerin sayısını artırmak ve kök alma kütüphanelerinin geliştirilmesine katkı sağlamaktır. Ek olarak Türkçe metin çalışmalarında daha yüksek boyutlu bir derlem geliştirmek ve Pennington ve arkadaşlarının geliştirmiş olduğu Glob2Vec modelini [17] uygulayarak sistemimizi geliştirmek hedeflenmektedir.

BİLGİLENDİRME

Türkçe veri setinin temin edilmesinden ötürü ve doğal dil işleme çalışmalarında aktarmış olduğu bilgilerden dolayı Dr. Öğr. Üyesi Gülşah TÜMÜKLÜ ÖZYER'e teşekkür ederiz.

KAYNAKLAR

- [1] Xie, W., Zhu, F., Jiang, J., Lim, E. P., & Wang, K. (2016). Topicsketch: Real-time bursty topic detection from twitter. *IEEE Transactions on Knowledge and Data Engineering*, 28(8), 2216-2229.
- [2] Nugroho, R., Molla-Aliod, D., Yang, J., Zhong, Y., Paris, C., & Nepal, S. (2015, May). Incorporating tweet relationships into topic derivation. In *International Conference of the Pacific Association for Computational Linguistics* (pp. 177-190). Springer, Singapore
- [3] T. Y. Kim, M. Min, T. Yoon, J. H. Lee, "Semantic Analysis of Twitter contents using PLSA and LDA", *SCIS & ISIS*, vol. 2010, pp. 189-192, 2010
- [4] Alghamdi, R., & Alfalqi, K. (2015). A survey of topic modeling in text mining. *Int. J. Adv. Comput. Sci. Appl.(IJACSA)*, 6(1).
- [5] Rong, X. (2014). Word2vec parameter learning explained. In arXiv preprint arXiv:1411.2738.
- [6] Xio Yang, Craig Macdonald, and Iadh Ounis. 2016. Using word embeddings in twitter election classification. arXiv preprint arXiv:1606.07006 (2016).
- [7] Nalçakan, Y., Bayramoğlu, Ş. S., & Tuna, S. (2015). Sosyal Medya Verileri Üzerinde Yapay Öğrenme ile Duygu Analizi Çalışması. Technical Report.
- [8] H. Nizam, S. S. Akin, "Sosyal Medyada Makine Öğrenmesi ile Duygu Analizinde Dengeli ve Dengesiz Veri Setlerinin Performanslarının Karşılaştırılması", 19. Türkiye' İnternet Konferansı İzmir, 2014.
- [9] E. A. Correa, V. Q. Marinho, L. B. d. Santos, "Nilc-usp at semeval-2017 task 4: A multi-view ensemble for twitter sentiment analysis", *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval'17)*, 2017.
- [10] Amine YEŞİLYURT, Şadi Evren ŞEKER, "Metin Madenciliği Yöntemleri ile Twitter Duygu Analizi", *YBS Ansiklopedi*, v. 4, is. 2, pp. 26- 36, Haziran 2017.
- [11] Mikolov, Tomas, Chen, Kai, Corrado, Greg, and Dean, Jeffrey. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781, 2013a.
- [12] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111-3119).
- [13] Qiang, J., Chen, P., Wang, T., Wu, X.: Topic modeling over short texts by incorporating word embeddings. In: Kim, J., Shim, K., Cao, L., Lee, J.-G., Lin, X., Moon, Y.-S. (eds.) *PAKDD 2017. LNCS (LNAI)*, vol. 10235, pp. 363–374. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-57529-2_29
- [14] Bo Pang , Lillian Lee , Shivakumar Vaithyanathan, Thumbs up?: sentiment classification using machine learning techniques, *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, p.79-86, July 06, 2002 [doi>10.3115/1118693.1118704]
- [15] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research*, 12(Oct), 2825-2830.
- [16] Setiawan, E. B., Widyantoro, D. H., & Surendro, K. (2016, October). Feature expansion using word embedding for tweet topic classification. In *Telecommunication Systems Services and Applications (TSSA), 2016 10th International Conference on* (pp. 1-5). IEEE.
- [17] Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).
- [18] Çelik, H. (2013). *Sentiment Analysis For Turkish Language* (Doctoral dissertation, İstanbul Kültür Üniversitesi/Fen Bilimleri Enstitüsü/Bilgisayar Mühendisliği Anabilim Dalı).
- [19] Çoban, Ö., Özyer, G. T., "Sentiment classification for Turkish Twitter feeds using LDA.", 24th. *Signal Processing and Communication Application Conference (SIU)*, 2016.
- [20] Miháltz, M. 12 Mayıs 2016. word2vec-GoogleNews-vectors, <https://github.com/mmihaltz/word2vec-GoogleNews-vectors>
- [21] Grave, E, 3 May 2017. Pre-trained Word Vectors. <https://github.com/facebookresearch/fastText/blob/master/pretrained-vectors.md>