Mühendislik Bilimleri ve Tasarım Dergisi
8(2), 366 – 374, 2020
e-ISSN: 1308-6693

Journal of Engineering Sciences and Design
DOI: 10.21923/jesd.546224

Araştırma Makalesi Research Article

SENTIMENT ANALYSIS FROM SOCIAL MEDIA COMMENTS

Özer ÇELİK[1*], Usame Ömer OSMANOĞLU[2], Büşra ÇANAKCI[3]

[1] Eskisehir Osmangazi University, Faculty of Science, Department of Mathematics-Computer, Eskisehir, Turkey
[2] Eskisehir Osmangazi University, Faculty of Medicine, Department of Biostatistics, Eskisehir, Turkey
[3] Eskisehir Osmangazi University, Faculty of Science, Department of Mathematics-Computer, Eskisehir, Turkey

Keywords

Sentiment Analysis, Machine Learning, Social Media, Classification Algorithms, Data Set.

Abstract

Nowadays, many firms and companies are curious about what people think and want and they are working in this direction. For this reason, it is tried to learn the ideas and emotions of people in various ways. However, as it is impossible to process and analyze a large number of emotions and thoughts with human hands, emotion analysis gain more importance. The emotions and thoughts of the people are analyzed and acted according to these requests through the emotion analysis which is quite functional in social networks. The aim of this study is to realize the learning with the data sets obtained from the interpretations made to the social platforms of the determined brands and to transfer the subject of the emotion analysis to the researchers in the best way. The range of accuracy rates reached is wide because of the disadvantages such as not paying attention to the rules of writing on social media or other digital platforms. In our study, a accuracy rate of 70% was achieved. This demonstrates the usefulness of machine learning in interpretation classification and emotion analysis.

SOSYAL MEDYA YORUMLARINDAN DUYGU ANALİZİ

Anahtar Kelimeler

Duygu Analizi, Makine Öğrenmesi, Sosyal Medya, Sınıflandırma Algoritmaları, Veri Seti.

Öz

Günümüzde birçok firma ve şirket insanların ne düşündüğü ve istediği konusunu merak etmekte ve bu doğrultuda çalışmalar yapmaktadır. Bu nedenle çeşitli yollarla insanların fikirleri ve duyguları öğrenilmeye çalışılmaktadır. Ancak çok sayıda duygu ve düşüncenin insan eliyle işlenip analiz edilmesi imkânsız olduğundan dolayı devreye 'Duygu Analizi' girmektedir. Sosyal ağlarda oldukça işlevsel olan duygu analizi sayesinde insanların duygu ve düşünceleri analiz edilip bu isteklere göre hareket edilmektedir. Bu çalışmanın amacı belirlenen markaların sosyal platformlarına yapılan yorumlardan elde edilen veri setleri ile öğrenmeyi gerçekleştirmek ve araştırmacılara duygu analizi konusunu en iyi şekilde aktarmaktır. Ulaşılan doğruluk oranları, sosyal medyada veya diğer dijital platformlarda yazım kurallarına dikkat edilmemesi gibi dezavantajları nedeniyle geniştir. Çalışmamızda %70'lik bir doğruluk oranı elde edilmiştir. Bu, makine öğrenmenin yorum sınıflandırma ve duygu analizinde kullanılabilir olduğunu göstermektedir.

Alıntı / Cite

Celik, O., Osmanoglu, U. O., Canakci, B., (2020). Sentiment Analysis From Social Media Comments, Journal of Engineering Sciences and Design, 8(2), 366-374.

Yazar Kimliği / Author ID (ORCID Number)

Ö. Çelik, 0000-0002-4409-3101
U.Ö. Osmanoğlu, 0000-0002-1198-2447
B. Çanakcı, 0000-0002-9641-0383

Makale Süreci / Article Process	
Başvuru Tarihi / Submission Date	28.03.2019
Revizyon Tarihi / Revision Date	24.02.2020
Kabul Tarihi / Accepted Date	25.02.2020
Yayım Tarihi / Published Date	25.06.2020

* İlgili yazar / Corresponding author: ozer@ogu.edu.tr, +90-222- 239-3750 / 2234

1. Introduction

Due to the widespread use of the Internet, the multiplicity of social networks, the ease of expressing thoughts and thoughts, access to brands, and rapid interaction, people spend most of their time on internet. As a result of this activity on social networks, meaningful and meaningless data are formed. The development of technology and the growth of data, the speed and cost to obtain the data provide the opportunity to work in this field (Baykara et. al., 2017).

Emotional analysis is a research area in which artificial intelligence techniques such as natural language processing, text mining are used to extract subjective information such as opinions, feelings and attitudes within the text. Social networking reviews are a great source of emotion analysis (Aytug & Korukoglu, 2016).

In this study, the comments made to the brands determined via Facebook for the emotion analysis were used as the data set. In addition, accurate estimation rates were compared by using TF-IDF and vector analysis.

TF-IDF (Term Frequency-Reverse Text Frequency):

$$W_{i,d} = tf_{i,d} \times \log(n/df_i) \qquad (1)$$

calculates the term of i and text of d. TF is simply defined as the ratio of the term to the most used term.

$$tf_{i,d} = aza(fr_{i,d}/df_i) \qquad (2)$$

The term frequency in the text d for the term i is the ratio of the term i to the number of repetitions of the term with the highest number of repetitions in the text d.

In this formula, n means the total number of documents, and df the document frequency (the number of how many different documents i term has passed) (Calis et. al., 2013).

1.1. n-Gram:

It is the method for finding the repetition rate in a given sequence. n is the value at which the repeat is checked. Gram is used to express the weight of this repeat in the sequence.
1-gram= unigram
2-gram=bigram
3-gram=trigram

These values are specifically named in the literature and values greater than three are generally expressed in n-grams (Cavnar & Trenkle, 1994).

2. Literature Survey

Boynukalin (2012) developed a framework with Turkish text for the analysis of emotions and given informations about Weka, Zemberek, Wllr ranking and n-gram approaches. In the study, two types of data sources, the international survey data set and the Turkish tale data set, were used. The first data set was translated into Turkish and the spelling errors were corrected using the zemberek library. Twenty-five tales were used for the other data set and divided because sentence and paragraph level emotion analysis was created. Classification was made as joy-anger-fear-sad. As a result of the study, different classification methods and different weightings were used and the accuracy rates varied between 42% and 85% (Boynukalin, 2012).

Akbas (2012), in her thesis paper, categorized the positive and negative thoughts on Turkish tweets. This classification is composed of 5-point scales. A data set was created in which emotion classification was performed manually (Akbas, 2012).

Garcia et al. (2015), mentioned positive-negative classification and 1 or 5-star classification in their study. Methods include clustering, model sweeping, test error, etc. In the classification, tree classifier, Naive bayesian classifiers and model extraction are mentioned. As a result, they prepared a classifier for positive-negative sentence estimation (Garcia & Yin, 2015).

Akgul et al. (2016) created four separate data sets with tweets made in Turkish according to a specific query word and classified the results as positive-negative and neutral in their study. They have done the Turkish character conversions by clearing unnecessary characters and words. They used a dictionary and n-gram model in their study and observed a 5% to 10% increase in the three data sets in the dictionary method and scoring. The N-gram study showed an accuracy rate of between 4% and 7% in neutral tweets. As a result, the dictionary and character-based n-gram methods used were approximately 70% and 69% respectively (Akgul et. al., 2016).

Kaynar et al. (2016) preferred film interpretations in their study. They created a positive and negative data set consisting of 200 data and performed their experiments using Matlab software. They made researches with classification algorithms and used these researches in their experiments. 75% of the data set was used for the training and the remaining 25% was reserved as test data. As a result, 75% correct classification was performed for SVM and ANN test data sets (Kaynar, 2016).

Gozukara et al. (2016) have made a positive-negative classification of 50 e-commerce sites in their study by using their own data sets. According to normalization n-gram accuracy rate ranged from 84% to 91% by performing TF-IDF experiments (Gozukara, 2016).

Durahim et al. (2018) made music classification. Predefined categories such as music genres and moods were created and 45 Turkish popular artists were selected. For the classification, labeling was carried out if 2 out of 3 persons agreed. The data set was prepared to be 75 songs in each of the four sensory categories. As a result of the trainings, the most successful classification algorithm is Multinomial NB with a success rate of 46% (Durahim et al., 2018).

Baykara et al. (2017) in their study, they mentioned social media analysis, studies on this subject and how to classify data in emotion analysis. In their work, they used PHP language to make emotion analysis and content classification according to the shares about a particular Twitter user. The classification was made by scoring. By scoring positive, negative and neutral word, calculations were done manually (Baykara et al., 2017).

Nalcakan et al. (2015) chose Twitter as the source of their study. The results of emotional analysis were classified as positive, negative and neutral. In the studies, three companies were identified from the technology sector and a separate data set was created for each company. NB, RF, LibSVM, J48 and Kstar classification algorithms were used in the experiments. The training set has been converted to Weka. When all the results were evaluated, it was seen that there were no significant differences between the classification results obtained from the data in the first set of training sets and the results of the education sets which were created by correcting all words by 75% (Nalcakan et al., 2015).

In the study conducted by Yigit, negative/positive percentage, mean negative/positive score, total negative/positive score calculations were made from call center text mining, converting calls from voice call centers to text and positive and negative classification. In the experiment, algorithms such as decision tree, KNN, SVM have been tried. According to the results of the experiment, the most successful classification with 82% accuracy was SVM algorithm (Yigit, 2017).

3. Material and Method

All experiments were performed using Jupyter Notebook – Python language. Random Forest, Logistic Regression, Multinomial Naive Bayes, Support Vector classifier algorithms are used on the data set. In order to be used in machine learning, the comments in the data set have been converted to numerical data. CountVectorizer and TF-IDF methods were used for this purpose. The data set was taken from the relevant social media platform by writing code with the Ruby programming language.

3.1. Naive Bayes Classification Algorithm

In the Naive Bayes classification, data is presented to the system at a given rate in the class. The new test data presented to the system with the probabilistic operations performed with the learned data are processed according to the previously obtained probability values and it is tried to determine the category of the test data given. The higher the number of data learned, it can be the more precise to determine the categories of test data. The Naive Bayes classification is a classification technique based on Bayes' Theorem (Andrade et al, 2019).

Bayes' Rule;
$p(x|Cj)$: Probability of being x of an example from class j
$P(Cj)$: First probability of class j
$p(x)$: Probability of being x of any example

$P(Cj|x)$: Probability of being from a class j of an x example (last probability)

$$P(Cj|x) = \frac{p(x|Cj)P(Cj)}{p(x)} = \frac{p(x|Cj)P(Cj)}{\sum_k p(x|Ck)P(Ck)} \qquad (3)$$

Naive Bayes Classifier, Each sample found in the T learning set get defined in n-dimensional space, X= $(x_1, x_2, .. x_n)$, m is the number of class in the data set, $C_1, C_2, \ldots, C_m$. In the classification, maximum last probability is sought (the maximal $P(C_i|X)$). Derived from Bayes Theorem, P(C_i│X)=(P(X│C_i)P(C_i))/(P(X)). Since the probability of P(X) is constant for all classes, only the maximum value is searched for the probability of $P(C_i|X) = P(X|C_i)P(C_i)$. $P(C_i|X) = P(X|C_i)P(C_i)$, if all the features are independent in this simplified expression, $P(X|C_i)$ can be written as follows:

$$P(X|C_i) = \prod_{k=1}^{n} P(x_k|C_i) = P(x_1|C_i) \times P(x_2|C_i) \times \ldots \times P(x_n|C_i) \qquad (4)$$

Thus the complexity of the account is greatly reduced (Kaynar et. al., 2016).

3.2. Multinomial Naive Bayes

Naive Bayes classification is one of the first preferred methods in many fields due to its easy adaptation, fast and consistent. Multinomial Naive Bayes can examine the model distribution of the words as multicategorical (Ardil, 2009). The purpose of the multi-term model is to determine the frequency of the terms in the text. This method comes to the fore as the importance of the words in the classification of the text is high. The frequency of use of the word in the text also allows testing whether there is any effect on the classification of the text.

3.3. Random Forest Classification Algorithm

It is an algorithm that aims to increase the classification value by producing more than one decision tree for classification. It is a combination of many tree decisions that are trained in different training clusters for each tree (Aksu & Karaman, 2017).

$$Gini(T) = 1 - \sum_{j=1}^{n} (p_j)^2 \qquad (5)$$

T: All data set
P_j: The square of division of the number of object smaller and greater than itself of each data in the data set
n: Selected data
After the Gini index is determined, the test data set classes are determined according to the gini index.

3.4. Support Vector Machines

Support vector machines is the classification algorithm based on statistical learning theories Support vector machines originally designed for the classification of two-class linear data are then generalized to the classification of multi-class and non-linear data. The principle of operation is to estimate the optimal decision function that can distinguish between two classes (Pham et al, 2019).

3.4.1. For Linear Separable Data

In the support vector machines classification, it is aimed to separate the samples belonging to the two classes, which are usually represented by the class labels {1, +1}, by means of a decision function obtained by the training data. By using the decision function, there is a hyper plane which can best distinguish the training data.

In the case of a linear classification of two classes of classification problems, the inequalities of the optimal hyperplane are as follows if it is assumed that the training data consisting of a number of samples for training of the support vector machines is $x_i, y_i\}$, i = 1, ..., k;

$$w.x_i + b \geq +1 \;\; \text{for each } y = +1 \qquad (6)$$
$$w.x_i + b \leq +1 \;\; \text{for each } y = -1 \qquad (7)$$

Here; $x \in R^N$ shows an N-dimensional space, y∈ {-1, +1} shows class labels, w shows weight vector (normal of hyper plane) and b shows trend value. In order to determine the optimal hyperplane, two hyperplane should be determined which will form parallel and boundary to this plane (Figure 1). The points forming this hyper plane are called support vectors.

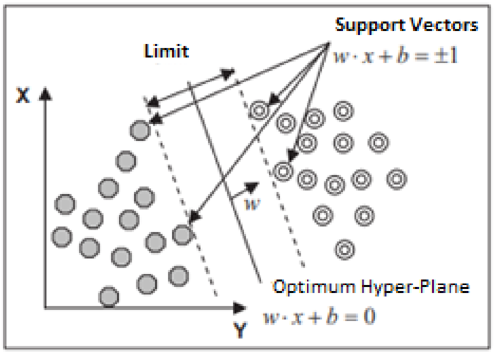

Figure 1. Determination of the hyperplane for linearly separable data sets (García-Gonzalo et al, 2016)

By minimizing ||w|| the optimum hyperplane limit maximizes. In this case, the solution of the limited optimization problem is required to solve the optimal hyperspace.

$\min[\frac{1}{2}||w||^2]$ Limitations related to this are;

It is expressed as $y_i(w.x_i + b) - 1 \geq 0\ ve\ y_i \in \{1, -1\}$. This can be solved using the Lagrange Equations.

$$L(w, b, \alpha) = \frac{1}{2}||w||^2 - \sum_{i=1}^{k} a_i y_i (w.x_i + b) + \sum_{i=1}^{k} \alpha_i \qquad (8)$$

The equality is obtained. As a result, the decision function can be written in $f(x) = sign(\sum_{i=k}^{k} \delta_i y_i (x.x_i) + b$ for a linearly separable problem (Kavzoglu & Colkesen, 2010).

3.4.2. For Linear Non-Separable Data

The problem that arises due to the fact that part of the training data remains on the other side of the optimal hyperplane is solved by defining a positive artificial variable (ε_i) (Figure 2). The equilibrium between the maximum of the limit and the minimization of errors can be controlled by defining an adjustment parameter indicated by C, which takes positive values.

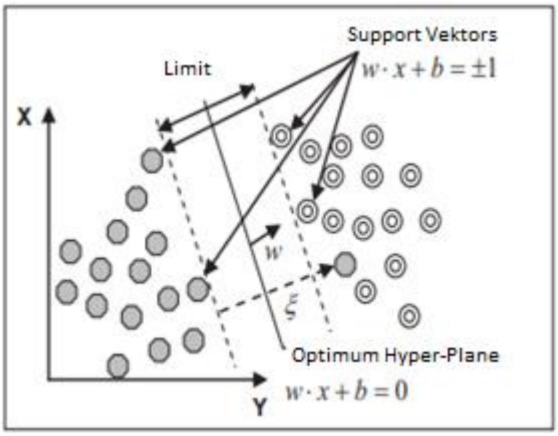

Figure 2. Optimization problem for non-discriminatory data takes the form (García-Gonzalo et al, 2016);

$\min[\frac{||w||^2}{2} + C.\sum_{i=1}^{r} \varepsilon_i]$. Sequences related to this take the shape of $y_i(w.\varphi(x_i) + b) - 1 \geq 1 - \varepsilon_i$, $\varepsilon_i \geq 0\ ve\ i = 1, .., N$.

Support vector machines can be mathematically transformed into non-linear transformations by means of a kernel function, which is expressed in the form of $K(x_i, x_j) = \varphi(x).\varphi(x_j)$ and allows the separation of data in high dimensions. The solution of a two-class problem that cannot be separated linearly using the kernel function can be solved by the rule $f(x) = sign(\sum_i \alpha_i y_i \varphi(x).\varphi(x_i) + b)$. It is essential to determine the optimal parameters of the kernel function and function to be used for a classification process to be performed with support vector machines (Kavzoglu & Colkesen, 2010).

3.5. Logistic Regression

The logistic regression predicts the probability of a result that can only have two values. Linear regression is not appropriate for the values that can be expressed in binary system as yes/no, exist/not. Because it can estimate the value outside the range 0 and 1. Logistic regression produces a limited logistic curve with values between 0 and 1. Logistic regression is similar to linear regression but is created by using natural logarithm of probabilities of target variable instead of curve probability (Manogaran, & Lopez, 2018).

Linear regression;

$$y = b_0 + b_1 X \qquad (9)$$

$$\text{Logit(p)=log(p/(1-p))} \qquad (10)$$

In logistic regression b_0 moves the slope to the right and left, while b_1 defines the slope of the curve. The logistic regression equation can be written with the probability ratio (logit (p)) (Figure 3) (Akin, 2018).

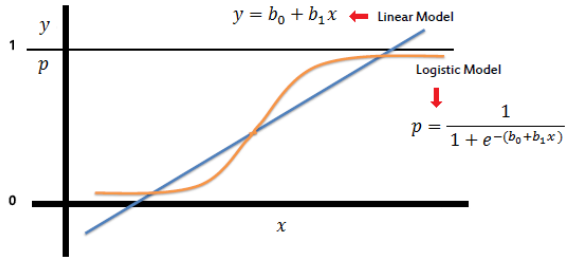

Figure 3. Logistic Regression (Akin, 2018)

Learning has been described by Simon as the process of improving behavior through the discovery of new information over time. The learning is called Machine learning when perform by a machine. The concept of improvement is the status of finding the best solution for future problems by gaining experience from the existing examples in the process of machine learning (Celik, 2018). With the development of information technologies over time, the concept of big data has emerged. The concept of big data is defined as very large and raw data sets that limitless and continue to accumulate, which cannot be solved by traditional databases methods (Altunisik, 2015).

The operations performed on the computer using the algorithm are performed according to a certain order without any margin of error. However, unlike the commands created to obtain the output from the data entered in this way, there are also cases where the decision making process takes place based on the sample data already available. In such cases, computers can make the wrong decisions such as mistakes that people can make in the decision-making process. In other words, machine learning is to gain a learning ability similar to human brain to computer by taking advantage of data and experience (Celik & Aslan, 2019).

The primary aim of machine learning is to develop models that can train to develop themselves and by detecting complex patterns and to create models to solve new problems based on historical data (Turkmenoglu, 2016).

With the advancement of technology, machine learning approaches increase its importance in many areas. For example, smart spam classifiers classify emails from large amounts of spam data and user feedback. Advertising systems try to match ads and content accurately; fraud detection systems protect structures such as banks from malicious acts.

3.6. Linear Regression

Linear regression is a statistical approach used to model the relationship between the dependent variable and one or more independent variables. If the number of independent variables is one, it is called simple linear regression, and if there are more than one independent variable, it is called multiple linear regression (David, 2009). When the number of dependent variables is more than one, it is called multivariate linear regression (Rencher & Christensen, 2012).

3.7. Confusion Matrix

The mess matrix shows the correct class of data and the number of classes estimated (Table 1).

Table 1. Confusion Matrix (Celik, & Osmanoglu, 2019)

		Actual Values	
		Positive (1)	**Negative (0)**
Predicted Values	**Positive (1)**	TP	FP
	Negative (0)	FP	TN

TP: True Positive FP: False Positive TP: True Positive FP: False Positive

Accuracy of the model; is the ratio of the number of accurately classified samples (TP + TN) to the total number of samples (TP + TN + FP + FN). The error rate is calculated by subtracting the accuracy rate from 1. In other words, it is the ratio of the number of misclassified samples (FP + FN) to the total number of samples (TP + TN + FP + FN) (2011Celik, & Osmanoglu, 2019).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (11)$$

4. Experimental Results

4.1. Data Set

In this section, a data set consisting of Facebook comments about different companies is used. The comments in the data set were divided into three classes as positive, negative and neutral by using the manual procedure and converted into a data set training set. Firstly, 5000 data from the tagged data was used for training. However, since the results obtained for this model were below the expectations, the data set was improved. For this, 1004 positive, 1000 and neutral 1000 in total 3004 comments were taken. 75% of these data were used for training and 25% for testing.

4.1.1. Data Set Properties

In this section, all comments are converted to lowercase letters and Turkish characters have been converted to Latin alphabet letters (ç-ı-ö-ı-u letters to c-g-i-o-u). Numbers, special characters, emojis, unnecessary words (with, one, -s, etc.) and non-Latin interpretations have been extracted.

The reason why Turkish characters are changed in Latin alphabet is because social media expresses a universal integrity. Furthermore, it is seen that the Turkish alphabet was not used correctly in the comments made about the firms.

Table 2. Examples and Classes from Data Set

CLASS	**SAMPLE DATA**
Negative	telefon ekrani soguk havada calismiyormus
Positive	goruntu guzel
Neutral	Lazer yazici nekadar fiyati ogrenebilir miyim

5. Result and Discussion

Interpretations in the data set were modeled using Multinomial Naive Bayes, Random Forest, Support Vector Machines and Logistic Regression classification algorithms using the Python programming language in Jupyter Notebook using the machine learning methods supervised learning approach.

The data set consisting of 1004 positive, 1000 negative and 1000 neutral were used 75% of them for the training, the remaining (25%) for the test.

It would be more appropriate to compare the studies conducted in similar regions to alleviate the impact of regional differences on emotion analysis. The line with this opinion is taken into account in studies in Turkey. The data set resources used in these studies are surveys, tweets, film reviews, comments on e-commerce sites, comments on facebook and comments in the call center. It is observed that the success rate of the studies has ranged from 42 to 91%. One of the biggest constraints of emotion analysis through interpretation is the non-observance of the grammar rules. Due to similar reasons, the accuracy rate range remains wide.

As a results of our study were examined, the best result with 0.7 accuracy of the model test of the Logistic Regression algorithm and with the accuracy of 0.56 of the KNN algorithm produced the lowest result. No significant differences were observed in TF-IDF and Count Vectorizer + TF-IDF comparison (Table 3).

Table 3. TF-IDF and Vector Analysis + TF-IDF Model Achievements and Comparison

Algorithms	TF-IDF	Count Vectorizer+TF-IDF
Multinomial NB	0.65	0.65
Random Forest	0.63	0.65
Logistic Regression	0.7	0.7
KNN	0.56	0.56
Support Vector Machines	0.67	0.67

Conflict of Interest

No conflict of interest was declared by the authors.

References

Akbas, E. (2012). Aspect based opinion mining on Turkish tweets. The Graduate School of Engineering and Science of Bilkent University, Ankara.

Akgul, E. S., Ertano, C., & Diri, B. (2016). Twitter verileri ile duygu analizi.

Akin, C. E. (2018). "Logistic Regression", http://cagriemreakin.com/veri-bilimi/logistic-regression-classification-8.html,

Aksu, M. C., & Karaman, E. (2017). Karar Ağaçları ile Bir Web Sitesinde Link Analizi ve Tespiti. Acta INFOLOGICA, 1(2), 84-91.

Altunisik, R. (2015). Büyük Veri: Fırsatlar Kaynağı mı Yoksa Yeni Sorunlar Yumağı mı?. Yildiz Social Science Review, 1(1).

Andrade, D., Tamura, A., & Tsuchida, M. (2019). Analysis of the Use of Background Distribution for Naive Bayes Classifiers. Journal of Intelligent Systems, 28(2), 259-273.

Ardil, E. (2009). Esnek hesaplama yaklaşımı ile yazılım hata kestrimi (Master's thesis).

Aytug, O., & Korukoglu, S. (2016). Makine öğrenmesi yöntemlerinin görüş madenciliğinde kullanılması üzerine bir literatür araştırması. Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 22(2), 111-122.

Baykara, M., Gurturk, U., & Teknolojileri, N. B. (2017). Sosyal Medya Paylaşımlarının Duygu Analizi Yöntemiyle Sınıflandırılması, 2. International Conferance on Computer Science and Engineering, 911-916.

Boynukalin, Z. (2012). Emotion analysis of Turkish texts by using machine learning methods. Middle East Technical University.

Calis, K., Gazdagi, O., & Yildiz, O. (2013). Reklam İçerikli Epostaların Metin Madenciliği Yöntemleri ile Otomatik Tespiti. Bilişim Teknolojileri Dergisi, 6(1), 1-7.

Cavnar, W. B., & Trenkle, J. M. (1994). N-gram-based text categorization. In Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval (Vol. 161175).

Celik, O., & Aslan, A. F. (2019). Gender Prediction from Social Media Comments with Artificial Intelligence. Sakarya Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 23(6), 1256-1264.

Celik, O., & Osmanoglu, U. O. (2019). Comparing to Techniques Used in Customer Churn Analysis. Journal of Multidisciplinary Developments, [S.l.], v. 4, n. 1, p. 30-38. ISSN 2564-6095.

Celik, O. (2018). A Research on Machine Learning Methods and Its Applications. Journal of Educational Technology and Online Learning, 1(3), 25-40.

David A. F. (2009). Statistical Models: Theory and Practice. Cambridge University Press. p. 26. A simple regression equation has on the right hand side an intercept and an explanatory variable with a slope coefficient. A multiple regression equation has two or more explanatory variables on the right hand side, each with its own slope coefficient

Durahim, A. O., Coskun, S. A., Basarir, O, B., & Kebapci, H. (2018). Music emotion classification for Turkish songs using lyrics. Pamukkale University Journal of Engineering Sciences, 24(2).

Garcia, S., & Yin, P. (2015). User Review Sentiment Classification and Aggregation.

García-Gonzalo, E., Fernández-Muñiz, Z., García Nieto, P., Bernardo Sánchez, A., & Menéndez Fernández, M. (2016). Hard-rock stability analysis for span design in entry-type excavations with learning classifiers. Materials, 9(7), 531.

Gozukara, F., & Ozel, S. A. (2016). An Experimental Investigation of Document Vector Computation Methods for Sentiment Analysis of Turkish and English Reviews.

Kavzoglu, T., & Colkesen, I. (2010). Destek vektör makineleri ile uydu görüntülerinin sınıflandırılmasında kernel fonksiyonlarının etkilerinin incelenmesi. Harita Dergisi, 144(7), 73-82.

Kaynar, O., Gormez, Y., Yildiz, M., & Albayrak, A. (2016). Makine öğrenmesi yöntemleri ile Duygu Analizi. In International Artificial Intelligence and Data Processing Symposium (IDAP'16), September (pp. 17-18).

Manogaran, G., & Lopez, D. (2018). Health data analytics using scalable logistic regression with stochastic gradient descent. International Journal of Advanced Intelligence Paradigms, 10(1-2), 118-132.

Nalcakan, Y., Bayramoglu, S. S., & Tuna, S. (2015). Sosyal Medya Verileri Üzerinde Yapay Öğrenme ile Duygu Analizi Çalışması. Technical Report.

Pham, B. T., Jaafari, A., Prakash, I., & Bui, D. T. (2019). A novel hybrid intelligent model of support vector machines and the MultiBoost ensemble for landslide susceptibility modeling. Bulletin of Engineering Geology and the Environment, 78(4), 2865-2886.

Rencher, A. C., Christensen, W. F. (2012), "Chapter 10, Multivariate regression – Section 10.1, Introduction", Methods of Multivariate Analysis, Wiley Series in Probability and Statistics, 709 (3rd ed.), John Wiley & Sons, p. 19, ISBN 9781118391679.

Sebastian R. (2015). Python Machine Learning. Birmingham.UK: Packt Publishing, 2015. ISBN: 978-1783555130.

Turkmenoglu, C. (2016). Türkçe Metinlerde Duygu Analizi (Doctoral dissertation, Fen Bilimleri Enstitüsü).

Yigit, I. O. (2017). Çağrı Merkezi Metin Madenciliği Yazılım Çerçevesi.