Journal of Educational Technology & Online Learning
Volume 3 | Issue 1 | 2020
http://dergipark.gov.tr/jetol

Sentiment Analysis for Distance Education Course Materials: A Machine Learning Approach

Usame Omer OSMANOGLU[a]
Osman Nuri ATAK[b]
Kerim CAGLAR[c]
Hüseyin KAYHAN[d]
Talat Cemre CAN[e]

[a] usa.osm@gmail.com; Osmangazi University, Eskisehir, Turkey; ORCID: 0000-0002-1198-2447
[b] osmannuriatak@anadolu.edu.tr, Anadolu University, Eskisehir, Turkey; ORCID: 0000-0002-1198-2447
[c] kerimcaglar@anadolu.edu.tr, Anadolu University, Eskisehir, Turkey; ORCID: 0000-0001-7675-4552
[d] huseyinkayhan@anadolu.edu.tr, Anadolu University, Eskisehir, Turkey, Eskisehir, Turkey; ORCID: 0000-0001-9036-3312
[e] talatcemrecan@gmail.com, Eskisehir Osmangazi University, Eskisehir, Turkey; ORCID: 0000-0002-0220-3878

Doi: 10.31681/jetol.663733
Suggested citation: Osmanoglu,U.O., Atak,O.N., Caglar,K., Kayhan, H. &Can, T.C. (2020). Sentiment Analysis for Distance Education Course Materials: A Machine Learning Approach. *Journal of Educational Technology & Online Learning*, *3*(1), 31-48.

Article Info

Received : 23.10.2019
Revised : 17.11.2019
Accepted : 11.12.2019

Research Article

Abstract

Nowadays many companies and institutions are interested in learning what do people think and want. Many studies are conducted to answer these questions. That’s why, emotions of people are significant in terms of instructional design. However, processing and analysis of many people's ideas and emotions is a challenging task. That is where the 'sentiment analysis' through machine learning techniques steps in. Recently a fast digitalization process is witnessed. Anadolu university, that serves 1 million distant students, is trying to find its place in this digital era. A learning management system (LMS) that distant students of the Open Education Faculty (Açıköğretim Fakültesi) is developed at the Anadolu University. Interaction with students is the clear advantage of LMS's when compared to the hard copy materials. Book, audio book (mp3), video and interactive tests are examples of these materials. 6059 feedbacks for those online materials was scaled using the triple Likert method and using machine learning techniques sentiment analysis was performed in this study. 0.775 correctness ratio was achieved via the Logistic regression algorithm. The research concludes that machine learning techniques can be used to better understand learners and how they feel.

Keywords: Sentiment Analysis, Machine Learning, Application Feedbacks, Deep Learning, Distance Education.

1. INTRODUCTION

Many students of the open universities are employed, already working individuals. Many of them hardly find time to study and, therefore, looking for learning materials that they can access anytime anywhere. Thus, digital learning materials become even more important for distance learners than ever. Besides, tailoring these materials is possible by benefiting from sentiment analysis even at systems where learning at scale occurs.

At Anadolu University, which is a Giga University with more than one million students (Bozkurt, 2019a), students in eCampus Learning Management system can reach all the materials of their active semester courses (Büyük et al., 2018; Düzenli et al., 2019) and they are able to give 250 characters limited feedbacks for these materials. In this regard, analyzing feedbacks collected from students have become an important data source to better understand what has been happening in eCampus learning ecology.

Distance education can be defined as "any learning activities within formal, informal, and non-formal domains that are facilitated by information and communication technologies to lessen distance, both physically and psychologically, and to increase interactivity and communication among learners, learning sources and facilitators" (Bozkurt, 2019b, p. 267) and the field of distance education has changed dramatically after 2000s when a paradigm shift observed due to capacity increase derived from ICT and online networked technologies (Bozkurt, 2019c). While it was once considered a special form of education using nontraditional delivery systems, it is now becoming an important concept in mainstream education. Concepts such as networked learning, connected learning spaces, flexible learning and hybrid learning systems have enlarged the scope and transformed the nature of earlier distance education models (Gunawardena, & McIsaac, 2013). In this transformation process, information and communication technologies played a vital role (Bozkurt, Zawacki-Richter, & Aydin, 2019). Considering interaction and communication is an important element of distance education processes, analyzing, identifying and understanding learners' feelings through sentiment analysis is considered significant.

In this study, researchers will analyze feedbacks gathered from eCampus system by using machine learning techniques. After analyzing feedbacks about a material, we expect to have an idea about sentimental value of the material such as positive, negative or neutral. By doing so, materials that has mostly negative feedbacks can have a priority over the others to be improved before the next semester.

1.1.Machine Learning

Machine learning first appeared in 1950's as a sub-branch of artificial intelligence. Till 1990's, there was no important progress at machine learning. However, the studies on machine learning restarted in 1990's and machine learning has a continuous progress till today. There is no doubt that it will progress even more in the future. Machine learning is based on the idea of finding the best model for new data by using the previous gathered data. That is why machine learning studies will continue as more data gathered (Celik, & Altunaydın, 2018) and will be significant to better understand changing perspectives of distance education landscape (Bozkurt, 2019d; Sharma, Kawachi, & Bozkurt, 2019a; 2019b).

Learning has been described as the process of improving behavior through the discovery of new information over time. Machine learning provides effective solutions for educational processes and 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 (Altunisik, 2015). 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 (Bozkurt, 2017; Sirmacek, 2007). 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 like human brain to computer by taking advantage of data and experience (Gor, 2014) 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). Machine learning and data-driven approaches are becoming very important in many areas. For example, smart spam classifiers protect our e-mails by learning from large amounts of spam data and user feedback. Ad systems learn to match the right ads with the right content; fraud detection systems protect banks from malicious attackers; Anomaly event detection systems help experimental physicists to find events that lead to new physics.

2. LITERATURE

Boynukalın (2012) developed at framework for Turkish text at a study conducted in 2012 for analyzing emotions. This study gives information about weka, zemberek, Wllr ordering and n-gram approaches. An international questionnaire dataset and Turkish tales' dataset are used in the study. First dataset was translated to Turkish, and typos were corrected using the zemberek library. For the second dataset, 25 tales were used and tales were divided to paragraphs and sentences because those contained emotions. Emotions were classified as happiness-angerfear-sadness. Different methods and weighting were used and success ratios between %42 and %85 obtained. Guran, Uysal, & Dogrusoz (2014) condcuted a study about the feedbacks on the internet for the products people bought. They used support vector machine (SVM) for classifying feedbacks and got successful results. They evaluated the results by analyzing the SVM parameters. Similarly, Turkmenoglu and Tantug (2014) studied sentiment analysis on Turkish texts in 2015. They used two different sentiment analysis methods and divided texts in to two datasets one containing long texts and the other containing long texts. In their study, Garcia and Yin (2015) mentioned positive-negative classification and 1 or 5 star classification. Clustering, model sweep and test error are mentioned in the methodology. In the classification, tree classifier, Naive Bayesian classifiers and model inference are used. As a result, they prepared a classifier for the prediction of positive-negative sentences. Akgul et al, (2016), in their study, formed four separate data sets by using a specific query word in Turkish in Twitter environment and classified the results as positive-negative and neutral. They have made Turkish character transformations by removing unnecessary characters and words. They used dictionary and n-gram model in their studies and observed an increase of 5% and 10% in three data sets in dictionary method and scoring. The N-gram study yielded a 4% to 7% increase in success in neutral tweets. As a result, they achieved approximately 70% and 69% success in dictionary and character-based n-gram methods, respectively. Kaynar et al. (2016) implemented a study based on the comments made for the movies on Twitter. According to the content of the comments, Naive Bayes conducted emotion and thought analysis using classification algorithms such as Center Based Classifier, Multilayer Artificial Neural Networks (MLP) and Support Vector Machines (SVM). They found that artificial neural networks and support vector machines gave better results in both training and test data. Baykara and Gurturk (2017) analyzed the comments of a specific twitter user in their work in 2017. They used Bayesian algorithm in their studies and classified them according to their contents. Not only positive, negative or neutral but also categorized message content (news, politics, culture) successfully. Parlar et al. (2017), in their study, conducted sentiment analysis from the shares made on Twitter. Using the Entropy Modeling classification algorithm on data sets, they compared the performances of 4 feature selection methods, Chi-square, information gain, query expansion ranking and ant colony optimization. Query Expansion Sort sensitivity analysis on the performance of Ant Colony Optimization on Turkeys' Twitter data from other traditional methods of feature selection methods have been observed to exhibit better performance. Gazioğlu and Seker (2017) conducted emotion analysis on English tweets in their study in 2017. Unlike other studies, they used emojis instead of classifying them as positive, negative and neutral. They created 15 different emoji groups and divided the tweets into these emoji classes. Durahim et al. (2018) conducted music classification studies in 2017. Predefined categories such as music genres and moods were created, and 45 Turkish popular artists were selected and labeling was done for the classification in 2 of 3 people if consensus was reached. The data set was prepared with 75 songs in each of the four sensory categories. As a result of the training of a successful model, the most successful classification algorithm is found to be Multinomial Naive Bayes which has a success rate of 46%. In the study conducted by Yigit (2017), call center data for text mining was used to convert the calls received from call centers to voice-to-text. Also, positive-negative classification, negative / positive percentage, average negative / positive score, total negative / positive score have been calculated. In experiments, decision tree, KNN, SVM, etc. algorithms were used. According to the results of the experiment, the most successful classification was SVM algorithm with 82% accuracy. In their study, Celik et al (2017) aimed to estimate the gender of the commentators through machine learning techniques by analyzing the comments of the companies registering on Facebook. In the study, the gender of the commentators was labeled according to the names in the comments collected from Facebook. The data set is divided into 70-30% of training and test data. As a result of the study, it was seen that machine learning methods were estimated with similar accuracy rates and the highest accuracy rate (74.13%) was obtained by logistic regression method. Celik and Osmanoglu (2020) further aimed to realize the learning with the data sets obtained from the comments made on the social platforms of the identified brands and to give the researchers the best way to convey emotion analysis. Achieved accuracy rates are wide due to disadvantages such as lack of attention to spelling rules on social media or other digital platforms. In the study, an accuracy of 70% was obtained. This demonstrates that machine learning can be used in review classification and emotion analysis. As explained above, though there are a vide arrange of studies on text mining and sentiment analysis, there is fewer research in the context of distance education. In this regard, this study intends to contribute to related literature by focusing on distance education processes and textual data in LMSs.

3. METHODOLOGY

3.1. Classification Algorithms Used for Research Model

This study benefits from data mining and analysis approaches. All analyzes were performed using Jupyter Notebook-Python. DecisionTreeClassifier, MLPClassifier, XGBClassifier, Support Vector Classifier (SVC), Multinomial Logistic Regression, GaussianNB and KNeighborsClassifier algorithms were used on the data set. In order to be used in machine learning, Clean Text, Spell Checker and Stop Words pretreatment processes were applied on the feedbacks gathered in the data set.

3.2. Logistic Regression

Logistic regression predicts the likelihood of a result having only two values. Linear regression is not suitable for values that can be expressed in binary system such as yes/no, true/false. Logistic regression produces a logistic curve limited to values between 0 and 1. Logistic regression is similar to linear regression, but it is generated using the natural logarithm of the probabilities of the target variable instead of the curve probability. Linear regression formula can be explained as followings;

$$y = b_0 + b_1 X$$

Logit(p)=log(p/(1-p))

In logistic regression b_0 moves the slope to the right and left, b_1 defines the slope of the curve. Logistic regression equation can be written with probability ratio (logit (p)) as a result (Figure 1) (Sebastian, 2015).

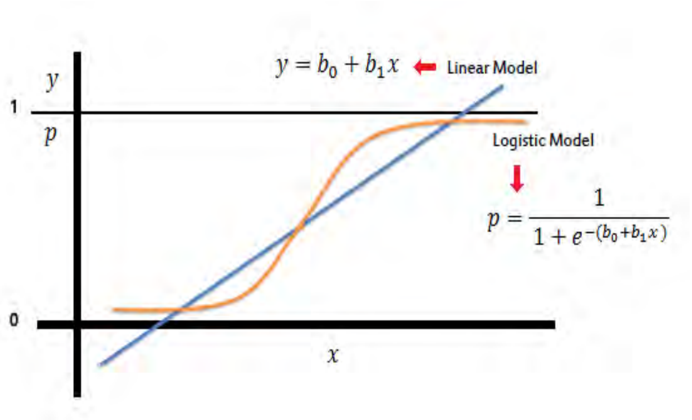


Figure 1. Logistic Regression Model Graph.

3.3. Criteria Used in Comparison of Classification Algorithms

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

Table 1

Confusion Matrix

		Predict	
		Class 1	Class 0
Actual	Class 1	TP	FP
	Class 0	FN	TN

*TP: True Positive; FP: False Positive; FN: False Negative; TN: True Negative

The accuracy rate of the model; is the ratio of the number of correctly classified samples (TP + TN) to the total number of samples (TP + TN + FP + FN). The error rate is the one that completes the accuracy rate to 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) (Celik & Osmanoglu, 2019).

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}$$

3.4. Data Set

In this study, automatic 3-point Likert-type scaling was performed according to the words in the comments written by the users for the e-campus application. For example; if there are words such as ‘bad’, ‘less’, ‘low’, ‘inadequate’, ‘low’, ‘far’, ‘away’, ‘long’, ‘short’, ‘disliked’ in the content of the comment, they are labeled as 1. In the content of the comment, the words ‘should’, ‘if there were’,‘ more’, ‘would be’, ‘should show’, ‘should put’, ‘should write’ etc. are labeled as 2. If there are words like ’good’, ‘liked’, ‘super’, ‘useful’, ‘beautiful’, ‘successful’ in the comment content, they are labeled as 3. With this process, it was observed that the data set was unbalanced (Figure 2). After the labeling process, the data set distribution was as follows: 4438, 815, 806.

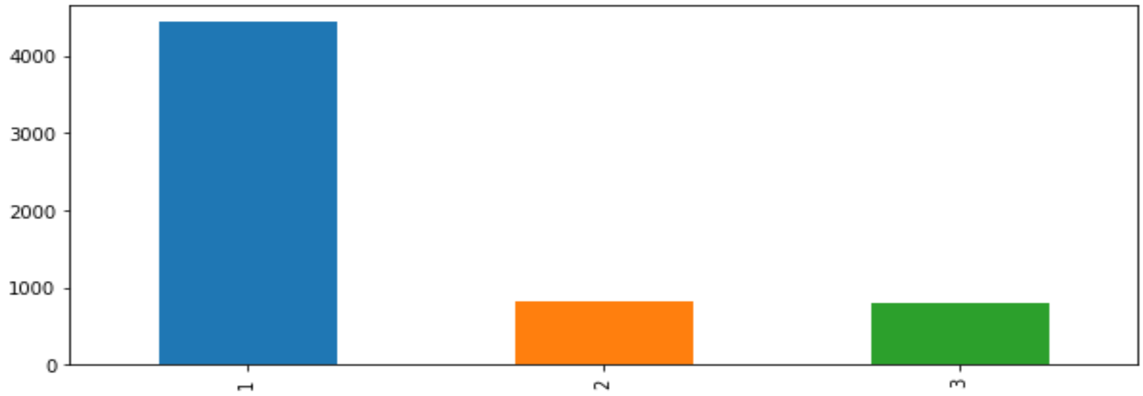


Figure 2. Tag class and number of comments in unbalanced data set.

The unbalanced data set was balanced by random sampling technique and the following distribution was reached: 800, 815, 806 (Figure 3).

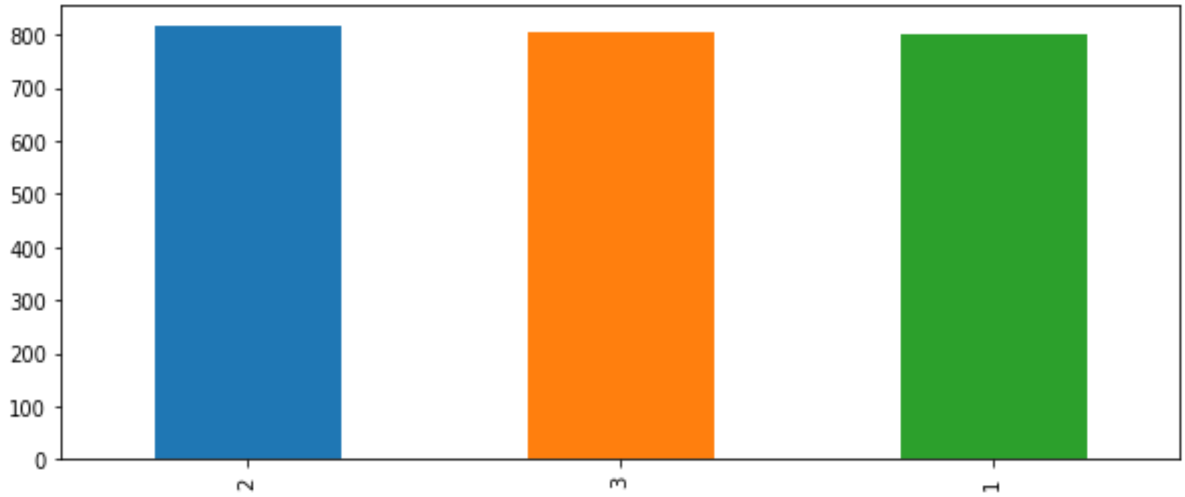


Figure 3. Label class and number of comments in balanced data set

3.5. Data Preprocessing

In this section, all comments are converted into lowercase letters and Turkish characters are converted into Latin alphabet letters (ç-ğ-ı-ö-ş-ü letters to c-g-i-o-u letters). Numbers, special characters, emojis, unnecessary words (with, one, -s, etc.) and non-Latin interpretations were omitted. The following techniques were applied to the balanced data:

1- Clean Text (CT); The application is used to achieve a general standard by performing the cleaning process on the comments. With this technique, all comments are converted to lowercase letters, and numeric expressions and punctuation are deleted.

2- Spell Checker (SC); to correct the misspelled words, it is applied on the comments
3- Stop Words (SW); The application is applied on comments to clear special characters, emojis, unnecessary (with, one, s, etc.), irrelevant and general words.

Eight different data sets obtained by applying these three techniques respectively were analyzed and the results are given in the table (Table 2). In addition, sample comments from the highest achievement data set are shown in Table 2.

Table 2

Results of the Analysis

		CT				SC				SW				CT + SC				CT + SW				SC + SW				CT + SC + SW				No Adj.			
		1	2	3	ACC	1	2	3	ACC	1	2	3	ACC	1	2	3	ACC	1	2	3	ACC	1	2	3	ACC	1	2	3	ACC	1	2	3	ACC
DT	1	184	72	19	0.67	187	71	17	0.68	185	62	28	0.67	207	49	19	0.75	96	166	13	0.35	196	49	30	0.71	228	50	18	0.77	168	79	13	0.65
	2	72	147	10	0.64	63	151	15	0.66	124	83	22	0.36	75	143	11	0.62	57	164	8	0.72	137	76	16	0.33	171	99	31	0.33	49	171	14	0.73
	3	17	8	213	0.89	17	19	202	0.85	81	19	138	0.58	9	11	218	0.92	32	84	122	0.51	80	17	141	0.59	104	26	164	0.56	16	11	206	0.88
	ACC	0.733				0.728				0.547				0.765				0.515				0.557				0.551				0.750			
MLP	1	179	67	29	0.65	195	63	17	0.71	113	134	28	0.41	200	55	20	0.73	92	163	20	0.33	196	48	31	0.71	221	52	23	0.75	184	61	15	0.71
	2	50	159	20	0.69	56	150	23	0.66	49	157	23	0.69	58	151	20	0.66	51	163	15	0.71	130	83	16	0.36	159	109	33	0.36	58	151	25	0.65
	3	19	21	198	0.83	20	29	189	0.79	22	80	136	0.57	15	25	198	0.83	29	83	126	0.53	79	17	142	0.60	98	28	168	0.57	16	27	190	0.82
	ACC	0.722				0.720				0.547				0.740				0.513				**0.567**				**0.559**				0.722			
XGB	1	188	57	30	0.68	176	73	26	0.64	88	171	16	0.32	159	88	28	0.58	53	207	15	0.19	95	160	20	0.35	96	194	6	0.32	131	112	17	0.50
	2	87	130	12	0.57	70	144	15	0.63	39	180	10	0.79	32	187	10	0.82	25	199	5	0.87	59	164	6	0.72	43	237	21	0.79	20	201	13	0.86
	3	3	8	227	0.95	13	6	219	0.92	11	102	125	0.53	2	14	222	0.93	17	117	104	0.44	20	99	119	0.50	17	144	133	0.45	3	19	211	0.91
	ACC	0.735				0.726				0.530				0.765				0.480				0.509				0.523				0.747			
SVC	1	169	85	21	0.61	174	81	20	0.63	111	139	25	0.40	190	71	14	0.69	80	179	16	0.29	99	148	28	0.36	224	55	17	0.76	170	77	13	0.65
	2	51	165	13	0.72	38	176	15	0.77	46	165	18	0.72	51	166	12	0.72	38	184	7	0.80	57	166	6	0.72	168	103	30	0.34	39	183	12	0.78
	3	7	12	219	0.92	12	18	208	0.87	15	73	150	0.63	10	12	216	0.91	27	89	122	0.51	24	80	134	0.56	98	30	166	0.56	12	17	204	0.88
	ACC	0.745				0.752				0.574				0.771				**0.520**				0.538				0.553				0.766			
LR	1	180	73	22	0.65	184	71	20	0.67	113	136	26	0.41	193	66	16	0.70	81	177	17	0.29	100	144	31	0.36	222	51	23	0.75	170	77	13	0.65
	2	41	171	17	0.75	44	169	16	0.74	44	165	20	0.72	52	167	10	0.73	34	176	19	0.77	58	161	10	0.70	160	102	39	0.34	38	181	15	0.77
	3	7	16	215	0.90	14	15	209	0.88	16	68	154	0.65	12	11	215	0.90	25	85	128	0.54	21	75	142	0.60	96	25	173	0.59	9	14	210	0.90
	ACC	**0.763**				**0.757**				**0.582**				**0.775**				0.519				0.543				0.558				**0.772**			
GNB	1	139	111	25	0.51	146	106	23	0.53	105	45	125	0.38	197	54	24	0.72	50	198	27	0.18	86	36	153	0.31	258	29	9	0.87	162	70	28	0.62
	2	37	171	21	0.75	29	168	32	0.73	58	74	97	0.32	107	96	26	0.42	15	200	14	0.87	45	64	120	0.28	227	61	13	0.20	52	164	18	0.70
	3	24	28	186	0.78	11	30	197	0.83	15	16	207	0.87	22	19	197	0.83	12	96	130	0.55	11	10	217	0.91	147	38	109	0.37	26	31	176	0.76
	ACC	0.668				0.689				0.520				0.660				0.512				0.495				0.480				0.691			
KNN	1	195	76	4	0.71	166	105	4	0.60	174	87	14	0.63	171	84	20	0.62	84	178	13	0.31	100	49	126	0.36	99	180	17	0.33	172	79	9	0.66
	2	108	114	7	0.50	72	150	7	0.66	113	99	17	0.43	78	138	13	0.60	46	170	13	0.74	63	79	87	0.34	59	220	22	0.73	92	129	13	0.55
	3	60	28	150	0.63	56	48	134	0.56	78	33	127	0.53	45	35	158	0.66	37	84	117	0.49	35	22	181	0.76	23	114	157	0.53	55	22	156	0.67
	ACC	0.619				0.606				0.539				0.629				0.500				0.485				0.534				0.629			

Table 3

Sample Comments and Classes from the Data Set

Class	Sample Comments
3	I usually like the expression of my teacher Mutlu. The video is not fragmented, but it was nice to have it presented at once. Thanks for the video.
2	MP3 WILL BE GOOD.
1	why there is no lecture video about this course

4. FINDINGS AND DISCUSSION

The interpretations used in the data set were modeled with the algorithm of Decision Tree Classifier, MLP Classifier, XGB Classifier, Support Vector Classifier, Multinomial Logistic Regression, Gaussian NB and KNeighbors Classifier with Python programming language in Jupyter Notebook by using supervised learning approach of machine learning method.

First, 6059 tagged comments were used for training. However, since the results obtained from this model were poor due to the unbalanced data set, the data set was improved. For this purpose, a total of 2421 comments were analyzed from 800, 815 and 806 of the three classes, respectively. Around 70% of these data were used for training and 30% for testing. When the results of the study were examined, the best results were obtained with 0.775 accuracy of the model test of Logistic Regression algorithm (Table 4).

Table 4

Confusion Matrix for Logistic Regression Model After CT + SC Operations

		Prediction			
		1	2	3	ACC (%)
Actual	1	193	66	16	0.70
	2	52	167	10	0.73
	3	12	11	215	0.90
	ACC (%)				0.775

Table 4 shows that there is no significant difference between the success rate after the CT + SC corrections and the success rate of the data without any corrections. However, since the correction process takes time, it may be preferable to establish a model without performing correction process and to make the analysis according to this model and data.

Table 5

No Adj. Case for Confusion Matrix for Logistic Regression Model

		Prediction			
		1	2	3	ACC (%)
Actual	1	170	77	13	0.65
	2	38	181	15	0.77
	3	9	14	210	0.90
	ACC (%)				0.772

Such an approach can be used in a wide arrange of applications. For instance, there are some compulsory common courses in Turkish Higher Education System that are delivered through distance education (Durak et al., 2017) and analysis of discussions forums of these courses can provide interesting insights regarding their effectiveness and efficiency of the educational processes. Similarly, such analysis can helpful to improve social dimension of LMSs. For instance, social LMSs like Edmodo (Durak, 2017) or Massive Open Online Courses (Artsın, 2018) can provide better learning experiences if sentiment of the learners identified which would eventually increase the motivational aspects of learning (Şenocak, 2019; Uçar and Kumtepe, 2018).

In addition to above thoughts, sentiment analysis can be used in online networked learning spaces. Due to capacity in online networks, online networked societies and networked individuals are the reality of digital knowledge age (Castells, 2004; Chatti, Jarke, & Quix, 2010; Rennie, &Wellman, 2012) and learning occurs in these networked knowledge ecologies (Bozkurt, & Keefer, 2017; Bozkurt, & Hilbelink, 2019; Siemens, 2006). The literature suggests that text-mining and sentiment analysis are very promising (Siemens, 2012; Shen, & Kuo, 2015) and much can be learnt about the learning environments and learners through sentiment analysis (Oliveiar, & Figueira, 2017).

5. CONCLUSION AND SUGGESTIONS

The data set resources used in this study was feedbacks of the distance learners. Accordingly, in distance education systems, where learning at scale occurs, such machine learning approaches can be used and that would enable to get insights how learners in these systems feel about.

For future research directions, researchers advise following suggestion. Accordingly, it would be more appropriate to compare the studies conducted in similar regions to alleviate the impact of regional differences on sentiment analysis. However, it should be noted that the success rate of the studies has ranged from 42% to 85%. One of the biggest constraints of sentiment analysis through interpretation is the non-observance of the grammar rules. Due to similar reasons, the accuracy rate range remains wide.

Uzaktan Eğitim Ders Materyalleri için Duygu Analizi: Bir Makine Öğrenme Yaklaşımı

Özet
Günümüzde birçok şirket ve kurum insanların ne düşündüğünü ve ne istediğini öğrenmek istemektedir. Bu soruları cevaplamak için birçok çalışma yapılmıştır. Bu yüzden, insanların duyguları öğretim tasarımı açısından önemlidir. Bununla birlikte, birçok insanın fikir ve duygularının işlenmesi ve analizi zor bir iştir. Makine öğrenme teknikleri ile 'duygu analizi' devreye giriyor. Son zamanlarda hızlı bir dijitalleşme süreci yaşanıyor. 1 milyondan fazla uzaktan eğitim öğrencisine hizmet veren Anadolu Üniversitesi, bu dijital çağdaki yerini bulmaya çalışıyor. Bu amaç doğrultusunda Anadolu Üniversitesi'nde Açıköğretim Fakültesi'nin (Açıköğretim Fakültesi) uzak öğrencilerini kapsayan bir öğrenme yönetim sistemi (LMS) geliştirilmiştir. Öğrencilerle etkileşim, basılı materyallerine kıyasla LMS'lerin açık avantajıdır. Kitap, sesli kitap (mp3), video ve interaktif testler bu materyallere örnektir. Bu çevrimiçi materyaller için 6059 geri bildirim üçlü Likert yöntemi kullanılarak ölçeklendirilmiş ve bu çalışmada makine öğrenme teknikleri duygu analizi kullanılarak yapılmıştır. Lojistik regresyon algoritması ile 0.775 doğruluk oranı elde edilmiştir. Araştırma, makine öğrenme tekniklerinin öğrencileri ve nasıl hissettiklerini daha iyi anlamak için kullanılabileceği sonucuna varıyor.

Anahtar Kelimeler: Duygu Analizi, Makine Öğrenmesi, Uygulama Geri Bildirimleri, Derin Öğrenme, Uzaktan Eğitim.

About the Author(s)

Usame Omer OSMANOGLU

Usame Omer Osmanoglu has got bachelor's degree from Eskisehir Osmangazi University, Department of Statistics in 2012. He has got master's degree from Eskisehir Osmangazi University, Department of Biostatistics in 2016. And, he has been continuing his academic life as a doctoral student in the same department in Eskisehir Osmangazi University, while he is working as Data Scientist in a software company. His interest areas are biostatistics, machine learning, artificial intelligence and traditional archery.

Mailing Address: Eskisehir Osmangazi University, Department of Biostatistics, Eskisehir, TURKEY, 26040,
E-mail: usa.osm@gmail.com

Osman Nuri ATAK

Osman Nuri Atak graduated from Yıldız Technical University, Department of Computer Engineering in 2005 and got master's degree from Osmangazi University, Department of Computer Engineering in 2019. He continues his academic career as a doctoral student in the Department of Mathematics and Computer Science of Osmangazi University. In the meantime, ATAK conducts his work activities and is interested in suggestion systems, machine learning, learning management systems, learning analytics, software architecture.

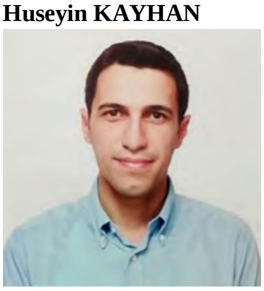
Mailing Address: Anadolu University, Learning Technology Research and Development Department, Eskisehir, TURKEY, 26210,
E-mail: osmannuriatak@anadolu.edu.tr

Kerim CAGLAR

Kerim Caglar has got bachelor's degree from Eskisehir Osmangazi University, Deparment of Mathematics and double major with Electrical electronics enginnering. He has got master's degree from Eskisehir Osmangazi University, Department of Mathematics-Computer in 2018. And, he has been continuing his academic life as a doctoral student in the same department in Eskisehir Osmangazi University, while he is working as software developer in Anadolu university. His interest areas are machine learning, artificial intelligence, internet of things, robotics and mobile programming.

Mailing Address: Anadolu University, Learning Technology Research and Development Department, Eskisehir, TURKEY, 26210,
E-mail: kerimcaglar@anadolu.edu.tr

Huseyin KAYHAN



Huseyin Kayhan received his bachelor's degree in Electrical and Electronics Engineering from Eskişehir Osmangazi University. He is a graduate student in Computer Engineering at Bilecik Şeyh Edebali University. He works as a software developer. His research interests include embedded systems, machine learning and artificial intelligence.

Mailing Address: Anadolu University, Learning Technology Research and Development Department, Eskisehir, TURKEY, 26210,
E-mail: huseyinkayhan@anadolu.edu.tr

Talat Cemre CAN

Talat Cemre CAN has got bachelor's degree from Eskisehir Osmangazi University, Department of Mathematics and Computer Science in 2018. He is currently studying for a master's degree at Eskisehir Osmangazi University, Department of Computer Science. In the same time, he is working as a data scientist at the software company. He is involved in python open source community and passionate about generative adverisial networks.

Mailing Address: Eskisehir Osmangazi University, Department of Biostatistics, Eskisehir, TURKEY, 26040,
E-mail: talatcemrecan@gmail.com

REFERENCES

Akgul, E. S., Ertano, C., & Diri, B. (2016). Twitter verileri ile duygu analizi. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 22*(2), 106-110.

Altunisik, R. (2015). Big Data: Is it a Source of Opportunities or New Problems? *Yildiz Social Science Review, 1*(1), 45-76.

Artsın. M. (2018). *Kitlesel Açık Çevrimiçi Derslerde Öğrenenlerin Öz-Yönetimli Öğrenme Becerilerinin İncelenmesi.* Anadolu Üniversitesi, Eskişehir.

Baykara, M., & Gurturk, U. (2017). Sosyal Medya Paylaşımlarının Duygu Analizi Yöntemiyle Sınıflandırılması. In proceedings of *2. International Conferance on Computer Science and Engineering* (pp. 911-916). Retrieved from http://web.firat.edu.tr/mbaykara/ubmk3.pdf

Boynukalın, Z. (2012). Emotion Analysis of Turkish Texts by Using Machine Learning Methods. Middle East Technical University. Retrieved from http://etd.lib.metu.edu.tr/upload/12618821/index.pdf

Bozkurt, A. (2016). Öğrenme analitiği: e-öğrenme, büyük veri ve bireyselleştirilmiş öğrenme. *Açık Öğretim Uygulamaları ve Araştırmaları Dergisi (AUAd), 2*(4), 55-81.

Bozkurt, A. (2019a). The historical development and adaptation of open universities in Turkish context: case of Anadolu university as a giga university. *International Review of Research in Open and Distributed Learning, 20*(4), 36-59. DOI: https://doi.org/10.19173/irrodl.v20i4.4086

Bozkurt, A. (2019b). From Distance Education to Open and Distance Learning: A Holistic Evaluation of History, Definitions, and Theories. In S. Sisman-Ugur, & G. Kurubacak (Eds.), *Handbook of Research on Learning in the Age of Transhumanism* (pp. 252-273). Hershey, PA: IGI Global. doi: https://doi.org/10.4018/978-1-5225-8431-5.ch016

Bozkurt, A. (2019c). Intellectual roots of distance education: a progressive knowledge domain analysis. *Distance Education, 40*(4), 497-514. DOI: https://doi.org/10.1080/01587919.2019.1681894

Bozkurt, A. (2019d). Vizyon 2023: Türkiye'de açık ve uzaktan öğrenme alanında somut ve soyut teknolojiler bağlamında eğilimler. *Açık Öğretim Uygulamaları ve Araştırmaları Dergisi (AUAd), 5*(4), 43-64. http://auad.anadolu.edu.tr/yonetim/icerik/makaleler/479-published.pdf

Bozkurt, A., & Hilbelink, A. (2019). Paradigm Shifts in Global Higher Education and e-learning: An ecological perspective. *eLearn Magazine, 2019*(5). DOI: https://doi.org/10.1145/3329488.3329487

Bozkurt, A., Keefer, J. (2017). Book Review: Knowing Knowledge. *The European Journal of Open, Distance and E-Learning (EURODL).* Retrieved from http://www.eurodl.org/materials/review/2017/Bozkurt_Keefer.pdf

Bozkurt, A., Zawacki-Richter, O., & Aydin, C. H. (2019). Using social network analysis to review the research in open and distance learning. In Proceedings of *The Association for Educational Communications and Technology (AECT) 2019 International Convention* (pp. 38-44). 21-25 October 2019, Las Vegas, NV. USA. Retrieved from https://members.aect.org/pdf/Proceedings/proceedings19/2019/19_06.pdf

Büyük, K., Kumtepe, A. T., Uça Güneş, E. P., Koçdar, S., Karadeniz, A., Özkeskin, E. … Öztürk A. (2018). Uzaktan öğrenenler ve öğrenme malzemesi tercihleri [Distance learners and their learning material preferences]. Eskişehir: Anadolu Üniversitesi. Retrieved from https://ekitap.anadolu.edu.tr/#bookdetail172074

Castells, M. (2004). *The network society: A cross cultural perspective*. MA, Northampton: Edward Elgar Publishing Limited.

Celik, O., & Altunaydın, S. S. (2018). A Research on Machine Learning Methods and Its Applications. *Journal of Educational Technology and Online Learning*, *1*(3), 25-40. DOI: 10.31681\jetol.457046

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*, *4*(1), 30-38.

Celik, O., & Osmanoglu, U. O. (2020). Sentiment Analysis from Social Media Comments. *Mühendislik Bilimleri ve Tasarım Dergisi*. *8*(1).

Chatti, M. A., Jarke, M., & Quix, C. (2010). Connectivism: The network metaphor of learning. *International Journal of Learning Technology*, *5*(1), 80-99.

Durahim, A. O., Coşkun Setırek, A., Başarır Özel, B., & Kebapci, H. (2018). Music emotion classification for Turkish songs using lyrics. *Pamukkale University Journal of Engineering Sciences*, *24*(2).

Durak, G. (2017). Using social learning networks (SLNs) in higher education: Edmodo through the lenses of academics. *The International Review of Research in Open and Distributed Learning, 18*(1).

Durak, G., Çankaya, S., Yünkül, E., & Bozkurt, A. (2017). 5İ Derslerini Uzaktan Eğitimle Alan Öğrencilerin Görüşleri. *VII. Uluslararası Eğitimde Araştırmalar Kongresi* (s.89). 27-29 Nisan 2017, Çanakkale, Türkiye.

Düzenli, H., Özdamar, N., & Bozkurt, A. (2019). Examination of a distance education course through the lens of activity theory. In Proceedings of *International Open & Distance Learning Conference (IODL19)* (pp. 275-282). Anadolu University, Eskişehir, Turkey.

Garcia, S., & Yin, P. (2015). User Review Sentiment Classification and Aggregation. Retrieved from http://cs229.stanford.edu/proj2015/048_report.pdf

Gazioglu, K., & Seker, S. E. (2017). Veri Madenciliği Yöntemleri ile Twitter Üzerinden Girişimcilik Analizi. *YBS Sözlük, 4*(4), 1-6.

Gor, I. (2014). A Desing and Implementation of Geometrical Learning Algorithm for Vector Quantization. Adnan Menderes University, Department of Mathematics, Aydin.

Gunawardena, C. N., & McIsaac, M. S. (2013). Distance education. In *Handbook of research on educational communications and technology* (pp. 361-401). Routledge.

Guran, A., Uysal, M., & Dogrusoz, O. (2014). The Effect of Parameter Optimization on Support Vector Machines on Emotion Analysis. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi, 16*(48), 86-93. Retrieved from https://dergipark.org.tr/tr/pub/deumffmd/issue/40797/492168

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

Oliveiar, L., & Figueira, A. (2017, April). Visualization of sentiment spread on social networked content: learning analytics for integrated learning environments. In *2017 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1290-1298). IEEE.

Parlar, T., Sarac, E., & Ozel, S. A. (2017, May). Comparison of feature selection methods for sentiment analysis on Turkish Twitter data. In *25th Signal Processing and Communications Applications Conference (SIU)* (pp. 1-4). IEEE. doi: 10.1109/SIU.2017.7960388

Rainie, L., & Wellman, B. (2012). *Networked: The new social operating system.* MIT Press.

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

Şenocak, D. (2019). *Açık ve uzaktan öğrenmede oyuncu tiplerinin motivasyon ve akademik başarı bağlamında incelenmesi.* Yüksek lisans tezi. Anadolu Üniversitesi, Sosyal Bilimler Enstitüsü, Uzaktan Eğitim Anabilim Dalı. Eskişehir.

Sharma, R. C., Kawachi, P., & Bozkurt, A. (2019a). Exploring Changing Perspectives in Distance Education. *Asian Journal of Distance Education*, *14*(1),1-6.

Sharma, R. C., Kawachi, P., & Bozkurt, A. (2019b). The landscape of artificial intelligence in open, online and distance education: Promises and concerns. *Asian Journal of Distance Education, 14*(2),1-2.

Shen, C. W., & Kuo, C. J. (2015). Learning in massive open online courses: Evidence from social media mining. *Computers in Human Behavior*, *51*, 568-577.

Siemens, G. (2006). *Knowing knowledge.* Vancouver, BC, Canada: Lulu Press.

Siemens, G. (2012, April). Learning analytics: envisioning a research discipline and a domain of practice. In Proceedings of *the 2nd international conference on learning analytics and knowledge* (pp. 4-8). ACM.

Sirmacek, B. (2007). *Modelling a learning algorithm for a mobile robot with FPGA.* Yildiz Technical University, Graduate School of Natural and Applied Sciences, Istanbul.

Turkmenoglu, C. (2016). *Sentiment Analysis in Turkish Texts.* Istanbul Technical University, Instıtute of Science and Technology, Istanbul.

Turkmenoglu, C., & Tantug, A. C. (2014, June). Sentiment analysis in Turkish media. In Proceedings of *International Conference on Machine Learning (ICML)*. Retrieved from https://sentic.net/wisdom2014turkmenoglu.pdf

Uçar, H. & Kumtepe, A. T. (2018). Integrating Motivational Strategies into Massive Open Online Courses (MOOCs): The Application and Administration of the Motivation Design Model. In *Administrative Leadership in Open and Distance Learning Programs* (pp. 213-235). IGI Global.

Yigit, I. O. (2017). Çağrı Merkezi Metin Madenciliği Yazılım Çerçevesi. Retrieved from http://ceur-ws.org/Vol-1980/UYMS17_paper_3.pdf