Evaluation of Customer Satisfaction about Telecom Operators in Turkey by Analyzing Sentiments of Customers through Twitter

Doğukan Kündüm[1], Zeynep Hilal Kilimci [2,*], Mitat Uysal [3], Ozan Uysal [3]
[1]*Department of Computer Engineering, Doğuş University, İstanbul, Turkey*
[2]*Department of Information System Engineering, Kocaeli University, Kocaeli, Turkey*
[3]*Department of Software Engineering, Doğuş University, İstanbul, Turkey*

***Abstract*—Sentiment analysis is the procedure of determinating whether a bit of a text is negative, neutral, or positive. Text analysis based sentiment analysis consolidates natural language processing models and machine learning algorithms to determine sentiment scores to the entities, topics, themes and categories within a phrase or, sentence. Furthermore, customer satisfaction is an evaluation of how products and services supplied by a company satisfy or exceed customer expectation. In this work, we propose to analyze customer satisfaction of three big telecommunication operators which are Turkcell, Turk Telekom, and Vodafone in Turkey by utilizing sentiment analysis of customers of them. For this purpose, Twitter social media platform is used for the purpose of gathering the related tweets that are mentioned with hashtags by the customers of operators. To improve system performance, various pre-processing models are used such as removing punctuation marks, stop-words elimination, removing tags, URLs filter, stemming. Finally, sentiment of users is evaluated through machine learning algorithms namely, random forest, support vector machine (SVM), multilayer perceptron (MLP), k-nearest neighbors (k-NN), naive Bayes (NB), and decision tree. The experiment results present remarkable classification performance with accuracy of over 80 percent for all telecom operators. Thus, this study can inspire telecommunications companies to analyze customer satisfaction through the social media platform.**

***Keywords*—Sentiment Analysis, Customer Satisfaction, Random Forest, Support Vector Machines, Multilayer Perceptron, Telecom Operators.**

I. INTRODUCTION

Sentiment analysis aims to define text data is positive, negative or neutral. Sentiment analysis is generally employed for this reason to evaluate customer satisfaction. Sentiment analysis is utilized in social media posts, tweets, and online product reviews. In terms of customer service and experience, the significance of sentiment analysis cannot be ignored. Because it is a great way as market research, brand or product reviews and customer experience analysis. Accordingly, the accuracy of sentiment analysis and predictions can be obtained by behavioral analysis based on social media. In this study, we focus to reveal the views of the customer of three big telecommunication operators in Turkey, datasets are collected from the user accounts of Twitter, and according to this on the measurement of customer satisfaction with sentiment analysis.

Machine learning is a subfield of artificial intelligence and big data science [1]. The main applications of machine learning include various fields such as computer vision, natural language processing, image processing, speech recognition, etc. In machine learning, learning methodology is divided into four main branches namely, reinforcement learning, supervised, semi-supervised, and unsupervised learning. Random Forest, Multilayer Perceptron (MLP), Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), Naive Bayes classifier (NB), Decision Tree are commonly employed in the literature as machine learning models. In this study, we concentrate on supervised learning technique by evaluating SVM, DT, RF, NB, MLP and k-NN.

In this paper, we collect and check out the sentiment analysis of tweets related to three big companies of Turkey telecommunication companies that are Turk Telekom (TT), Turkcell, and Vodafone Turkey. In particular, tweets in Turkish have been selected, also, 1 October 2019 and 1 December 2019. During this period, we meticulously collected a Turkish tweet belonging to Turkcell, Vodafone and Turk Telekom companies. Idea in the proposed work is to expose the sentiment, which could be either customer experience. It covers the classification of three different vectors containing Turkish tweets from three major telecom operators for sentiment analysis and interpretation of customer satisfaction. The aim of the study, enabling telecom companies in Turkey to evaluate customer reviews by using text mining on social media platforms and how maximum customer satisfaction can be reached through this evaluation ultimately. For this purpose, Turkish texts are gathered from Twitter by employing web-scraper. After getting textual data, various pre-processing techniques are implemented to remove the influence of dirty data and remove stop-words to terminology of Turkish language. The study is asked to be expressed and classified as machine learning algorithms, which is widely used in sentiment analysis. Machine learning techniques employed in this work

Manuscript received July 22, 2020; accepted December 31, 2020.
*Corresponding author: zeynep.kilimci@kocaeli.edu.tr

are Random Forest, Support Vector Machine (SVM), Multilayer Perceptron (MLP), K-Nearest Neighbors (k-NN), Naive Bayes (NB), Decision Tree.

The rest of the article is organized as follows: Section II of previous studies have been focused. Section III of datasets, methods, tools, sentiment analysis system components and classification. Section IV of the tests is described. Experimental results and conclusions are given in Section V.

II. Related Work

In this section, a brief summary of the literature review of the studies' effort on customer reviews.

Humera Shaziya et al. [2] classified movie reviews for sentiment analysis using popular one of software instrument WEKA. This work is a sentiment classification that analyzes views that convey positive or negative emotions. They report that accuracy has been of 85.1% for NB.

ther research has reviewed and classified the latest technology established on popular methods used for about personal point of view and sentiment analysis in Arabic: supervised learning employing machine learning models [3, 4], unsupervised learning using sensitivity dictionaries [5, 6], and a hybrid approach that consolidates the two techniques [6, 7].

Reviews of some studies on sentiment analysis with machine learning technologies have been in this section about telecom companies. Qamar et. al [8] focus on Saudi Telecom Companies in Saudi Arabia. This work is shown that machine learning methods used to guide brand and product management by providing information about customer experience. Their study made to tweets published in English, belonging to the various local telecom companies. In [9], The performances of k-NN and NB as machine learning models are evaluated. In [9], the proposed model can achieve 80.1% accuracy with an asymmetric k-NN variant while using the cosine similarity. The other work [10] compares NB and SVM methods can be used to applied given from Saudi Telecom Companies tweets. In [10], they present this content in August 2022 (since 2015).

Amolik et. al. in [11], created datasets using comments of movie reviews shared by Twitter and tweets about these movies. They use someone machine learning that classifiers such as NB, SVM, ensemble classifier, k-NN, and Artificial Neural Networks, for the purpose of classifying tweets as positive, negative, and neutral classes. Kaya et al. al [12] presents sensitivity classification techniques used in the field of political news from columns in diverse Turkish news sources on the sites. They evaluate the results of NB, SVM, and Maximum Entropy classification by developing with different features of each classifier. They report the classification accuracy boosts from 65% to 77%.

Result of literature researches, there have not found to use machine learning at Turkish in the telecom sector customer satisfaction. As a result, in this study, we propose to provide a valuable resource using machine learning methods for the customer satisfaction in telecom sector by analyzing sentiment of their customers-site.

III. Proposed Model

In this section, a summary of the methods, materials and proposed framework are presented.

A. Data Collection

In this study, in order to estimate the satisfaction of the Turkish Telecommunication Corporates, comments in Twitter are collected between 1 December 2019 and 1 September 2019. Thus, we propose to estimate the satisfaction of customers in their praetors of Turkcell, Turk Telekom (TT) and Vodafone by analyzing customer comments in Twitter between aforementioned date interval.

Turkish customer comments from Twitter are collected with the label of each operator of names. These are mentions of #vodafone, #turktelekom and #turkcell. Selenium browser is used to collect as many tweets as we like without worrying about the limit by the Twitter APIs. In addition, there are using many filters to Twitter with Selenium browser. For example, basically only text tweets between two dates are used as a source in our study.

B. Machine Learning Algorithms

In this study, we utilize commonly employed machine learning techniques such as Random Forest, Support Vector Machine (SVM), Multilayer Perceptron (MLP), K-Nearest Neighbors (k-NN), Naive Bayes (NB), Decision Tree.

Support Vector Machine (SVM)

Support Vector Machine (SVM), image classification, sound classification etc. Besides, it is also used in the field of text categorization. SVM was chosen for classification in experiments. SVM is a supervised learning model that can use to resolve regression and classification problems. In [13], it has been used to classify both linear and nonlinear data. SVM is created dimensional space to solve the classification or regression problem. This plane aims to obtain the best solution separation by taking the greatest distance from the closest training data points of the categories (which can be called functional boundary).

K-Nearest Neighbours (k-NN)

k-NN is a kind of lazy learning technique that is employed for both classification and regression problems. K is a simple algorithm that can be easily implemented with the help of the nearest neighbor, distance functions. This technique stores all present cases and categorizes novel states based on a similarity measure such as Euclidean, Manhattan, and Minkowski [14, 15]. In both classification and regression problems, the input depends on the k closest training instances in the feature space while the output consists of whether k-NN is utilized for categorization or regression.

In this work, k-NN method is used for the classification task by implementing Euclidean model as a distance function and adjusting k as 3.

Naïve Bayes (NB)

Naïve Bayes is a member of a classifier based on realizing Bayes' theorem that assumes independence between properties [16]. The Naïve Bayes classifier is especially considered when the dimensionless of input properties is high. In [17], the Naïve Bayes classifier is considered to be a good algorithm in terms of speed performance and accuracy success. Short calculation time for training is one of the biggest advantages of a pure Bayesian classifier.

Multilayer Perceptron (MLP)

The multilayer perceptron (MLP) is a popular supervised learning technique among other models of artificial neural networks that is implemented for estimation problems in the literature. The multilayer perceptron consists of basically three layers; the input layer, hidden layer, and the output layer. In additional, the hidden layer may have one or more than of more activation defined. Multilayer perceptron is a feed-forward model that maps data onto a set of related outputs [18]. As the number of hidden layers' increases can be lead to performance issues. The multilayer perceptron has robust computing competency due to that enable the solution of non-linear sectional problems.

Decision Tree (DT)

Decision tree is one of the most used prediction models especially used in statistical interpretation, data mining and machine learning. The main purpose of a decision tree is based on research about an item shown in sub-units up to the conclusions about the target value of an item. The Decision Tree or classification tree uses a tree representation to learn that forebodes the value of the variable dedicated to the values of parameter input. The Decision Tree use into divided sub-trees the smaller datasets ensure the questions it. In [19], decision tree has been considered as one of the most practical and simple to classification method.

Random Forest (RF)

Random Forest is performed both regression and classification tasks. It is a kind of learning method that assembles a few weak learning models to form a strong model [20]. Random Forest is one of the most reliable model in the classification tasks because of consisting of decision consensus of each models. Its sensibility and sensitivity depend a lot on overfitting.

C. Proposed Framework

In this study, user comments from Twitter are gathered using with the mentions: #vodafone: a total of 32,167 tweets, #turktelekom: a total of 96,061 tweets, #turkcell: a total of 52,182 tweets. A total of 180,410 Turkish tweets are collected with the three different labels using the Selenium crawler to collect user comments from Twitter. First, the collected raw data set is cleaned with different pre-processing techniques. In this study, removing punctuation marks, stop-word elimination, removing Twitter hashtags, removing special characters, removing URLs with Twitter search filters are applied. Then, all characters in all words are transformed to lower-case. Stemmer words using a Turkish stemmer named "Zemberek Library" [21] methods are applied. Zemberek is the known and applied NLP tool in Turkish that is used for morphological analysis.

Zemberek has functions that can be used for stemming. In "Fig. 1", text pre-processing flow diagram is demonstrated.

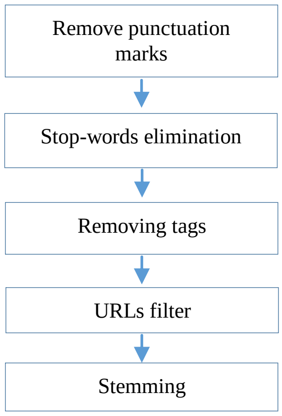

Fig 1. Pre-processing flow diagram of text in documents

TextBlob is a Python library for processing textual data. TexBlob [22] is utilized, to label the preprocessed datasets. However, because of the lack of Turkish pre-trained datasets in the TextBlob, we use Turkish pre-trained dataset by employing Turkcell customer reviews from the website of Yıldız University Kemik Labs (http://www.kemik.yildiz.edu.tr/). With the usage of pre-trained dataset, each tweet that are released from customers of telecom operators are labelled through TextBlob. Machine learning algorithms used in this work evaluate model and each comment/tweet is labelled as 1 and 0. Tweets that are positive are labeled as 1 and negative tweets are tagged as 0. These labelled datasets are total of size 14,725 by positive 9,205 and negative 5,520.

After acquiring labelled dataset, it is splitted into training and test sets, 80% and 20%, separately. After arranging training set percentage of the dataset, we feed the fetched tweets (customer satisfaction reviews tweets) into the machine learning classifier aforementioned. These machine learning techniques are support vector machine, k-nearest neighbors, naive Bayes, decision tree, random forest as an ensemble learning method, and multilayer perceptron as an artificial neural network. Thus, we also measure the impact of traditional machine learning methods, ensemble learning methodology, and artificial neural networks by comparing the classification performance of them in datasets of telecom operators.

In "Fig. 2", The flowchart of proposed model is presented. In order demonstrate the classification performance of the system, precision, recall, f1-measure, and accuracy are employed as evaluation metrics.

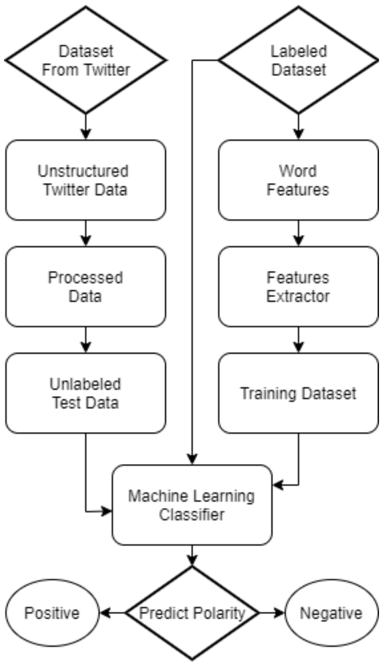

Fig 2. The Flowchart of Proposed Model

IV. Tables, Figures

In this section, the result of the classification of the sentiment analysis based upon customer reviews are evaluated on three big operators in Turkey, that are, Turk Telekom, Vodafone, and Turkcell. The classification of customer reviews as shown in matrixes for each operators separately represents the comparison between accuracy scores that are SVM, k-NN, DT, RF, MLP and NB.

The result of the sentiment analysis is presented in the form of table total accuracy for all telecommunications operators. According to these accuracies, final aggregate table of these operators is formed in order to sentiment analysis.

A. *Customer Satisfaction Reviews of Turk Telekom Operator*

For Turk Telekom (TT), reviews are classified into each review by dividing into two sentiments as positive and negative which means there are two classes to be assigned for Turk Telekom (TT) dataset.

Table I represents the evaluation between SVM, k-NN, DT, MLP, RF, and NB classifier algorithm for the classification process of reviews of Turk Telekom (TT). From the accuracy of the text classification, it is shown that RF has better classification performance than other machine learning algorithms. RF exhibits the best classification performance with 89.33% of accuracy. Table I represents the results of each evaluation metric by analyzing six machine learning classifier algorithms for classification process of reviews of Turk Telekom (TT).

TABLE I. CLASSIFICATION PERFORMANCE OF EACH ALGORITHM IN TERMS OF DIFFERENT EVALUATION METRICS IN TURK TELEKOM

Algorithm	Precision	Recall	F1-score	Accuracy
SVM	0.87	0.89	0.87	89.29
k-NN	0.82	0.81	0.81	81.15
RF	0.89	0.89	0.86	89.33
MLP	0.86	0.87	0.87	86.76
DT	0.86	0.88	0.84	88.44
NB	0.81	0.80	0.80	79.92

In “Fig 3.” shows the result of sentiment analysis for customer reviews of Turk Telekom (TT). There are negative reviews than positive ones. The number of positive and negative sentiments of customer reviews in Turk Telekom (TT) dataset. There are totally 96k customer reviews that of 11,627 customer reviews are positive and 84,434 are negative.

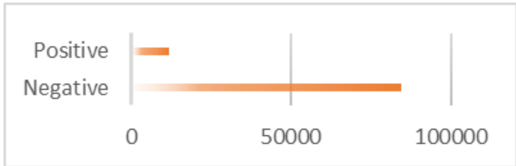

Fig 3. The Number of Positive and Negative Sentiments of Customer Reviews in Turk Telekom (TT).

Line interruptions are observed to be the most problematic issues based on customers' comments. Complaints are therefore increasing, as it has a wider audience in terms of the range of customers. In addition, customers are observed to be satisfied because the invoices are affordable. It is observed that the problems like other operators are experienced during the earthquake that occurred in Istanbul during our research period, and the longer problems continued in for comparison to other operators negatively affected customer satisfaction.

B. *Customer Satisfaction Reviews of Vodafone Operator*

For Vodafone, reviews are classified into each review by dividing as positive and negative sentiments which means there are two cases to be assigned for Vodafone. Table II represents the evaluation between SVM, k-NN, DT, MLP, RF, and Naïve Bayes classifier algorithms for the text classification of customer reviews of Vodafone. The result of accuracy from the sentiment analysis, the RF algorithm has been shown to be more successful than other algorithms. RF reached the 80.66% accuracy number. Table II exhibits the evaluation performance of each evaluation metrics in terms of six machine learning classifier algorithms for classification process of reviews of Vodafone.

TABLE II. CLASSIFICATION PERFORMANCE OF EACH ALGORITHM IN TERMS OF DIFFERENT EVALUATION METRICS IN VODAFONE

Algorithm	Precision	Recall	F1-score	Accuracy
SVM	0.78	0.80	0.77	80.44
k-NN	0.76	0.78	0.76	78.22
RF	0.80	0.81	0.76	80.66
MLP	0.77	0.78	0.78	77.92
DT	0.76	0.79	0.72	78.90
NB	0.73	0.68	0.70	67.96

In “Fig 4”. shows the result of sentiment analysis for customer reviews of Vodafone. There are negative reviews than positive ones. The number of positive and negative comments of Vodafone is as follows: “Fig 4.” shows the result of sentiment analysis for customer reviews of

Vodafone. There are totally 32k customer reviews that of 7,090 customer reviews are positive and 25,077 are negative.

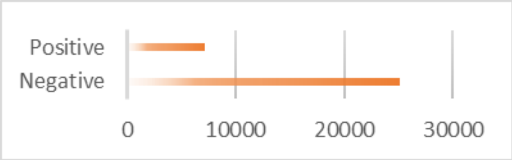

Fig 4. The Number of Positive and Negative Sentiments of Customer Reviews in Vodafone

The operator is complained of short-term disruptions in line cuts. Likewise, problems occur when an earthquake occurs. However, it is observed that users express their satisfaction as they offer faster solutions than Turk Telekom (TT). For Vodafone, the most negative review is that they were not less campaign and financially customer-centric. In addition, the customer is dissatisfied with the expensive phone invoice based on performance.

C. *Customer Satisfaction Reviews of Turkcell Operator*

For Turkcell, reviews are classified into each review by dividing into two sentiments as positive and negative which means there are two classes to be assigned for Turkcell. Table III represents the evaluation between SVM, k-NN, DT, MLP, RF, and NB classifier algorithm for the classification process of reviews of Turkcell. From the accuracy of the text classification, it is shown that RF has better performance than other algorithms. RF reached the average 79% accuracy number. Table III shows the evaluation performance of each evaluation metrics in terms of six machine learning classifier algorithms for classification process of reviews of Turkcell.

TABLE III. CLASSIFICATION PERFORMANCE OF EACH ALGORITM IN TERMS OF DIFFERENT EVALUATION METRICS IN TURKCELL

Algorithm	Precision	Recall	F1-score	Accuracy
SVM	0.77	0.79	0.76	78.74
k-NN	0.74	0.77	0.70	76.65
RF	0.78	0.80	0.75	79.03
MLP	0.75	0.76	0.76	76.20
DT	0.74	0.77	0.70	76.65
NB	0.71	0.70	0.70	70.38

In “Fig 5.” shows the result of sentiment analysis for customer reviews of Turkcell. There are negative reviews than positive ones. It is seen that Turkcell is at a point between two other operators in a similar date range. These are also reflected in the number of positive and negative reviews. There are totally 53k customer reviews that of 13,159 customer reviews are positive and 40,023 are negative.

Similarly, in operator line interrupts, most of customers complain about the long waiting time needed for communication. For operator Turkcell based on a high-speed packet access aspect, the most negative review is being exceed the connection problems of an operator, especially on special days. However, it is observed that these complaints do not take too long. The success of Turkcell in providing quick solutions can be derived from this. Customer satisfaction oriented studies have been observed. In addition, moreover, the customer is not satisfied with the expensive phone invoice by performance. It is similar to Vodafone in this regard.

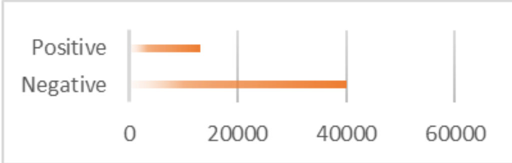

Fig 5. The Number of Positive and Negative Sentiments of Customer Reviews in Turkcell

D. *Customer Satisfaction Reviews of All Operators*

In this study, operators have problems due to the earthquake occurred in time interval aforementioned before. The earthquake special day is also evaluated in terms of customer satisfaction number between 1 September 2019 and 1 December 2019. In Table IV, this date is also included in the summary of the results table. Table IV also demonstrates the evaluation between SVM, k-NN, DT, MLP, RF, and NB classifier algorithm for the classification process of reviews of all operator.

TABLE IV. CLASSIFICATION ACCURACIES OF SIX DATASETS IN TERMS OF EACH ALGORITHM AT TRAINING SET 80%

Operator	SVM	k-NN	RF	MLP	DT	NB
TT	89.29	81.15	89.33	86.76	88.44	79.92
Vodafone	80.44	78.22	80.66	77.92	78.90	67.96
Turkcell	78.74	76.65	79.03	76.20	76.65	70.38
TT[a]	90.40	89.91	90.68	88.74	89.98	76.18
Vodafone[a]	82.36	78.97	82.72	81.76	82.24	69.22
Turkcell[a]	74.57	72.06	75.23	71.74	72.56	66.87

[a.] Special Date of Earthquake

The following is the result of a comparison of the value of accuracy made on the RF method, other machine learning algorithms which can be seen in Table IV, RF outperforms other methods in all datasets in terms of accuracy results. In addition, NB compared to the other machine learning methods is which have an accuracy rate the lowest accuracy.

V. DISCUSSION AND CONCLUSIONS

Customer satisfaction is an important factor that determines an operator's business success. Therefore, evaluation of high-speed services and non-disruption of service based on customer reviews is critical as it provides constructive comments, so the telecommunications operator can measure and improve the quality of service to ensure customer satisfaction. This study used six machine learning algorithms, such as decision tree, multilayer perceptron, support vector machine, random forest, k-nearest neighbors, and Naïve Bayes classifier, using sensitivity analysis text mining for three major operators to extract information from customer reviews. shows techniques.

In this way, we propose a customer satisfaction model for three big telecommunication operators which are Turkcell, Turk Telekom, and Vodafone in Turkey by analyzing sentiment analysis of customers of them. For this purpose, Twitter as a social media platform is employed for gathering the related comments that are mentioned with hashtags by the customers of operators. To enhance the system performance, pre-processing models are utilized such as removing punctuation marks, stop-words elimination, removing tags, URLs filter, and stemming. Eventually, sentiment of users is appraised through machine learning algorithms namely, random forest, support vector machine (SVM), multilayer perceptron (MLP), k-nearest neighbors (k-NN), Naive Bayes (NB), and decision tree. The experiment results demonstrate considerable classification success with accuracy of over 80 percent for all telecom operators. Thus, this study can inspire telecommunications companies to analyze customer satisfaction through the social media platform. In details, the best classification performance is observed with the usage of random forest model for all telecom operator datasets while Naive Bayes techniques exhibits the poorest performance among others. Meanwhile, in Table IV it is seen that the classification accuracy of RF outperforms all methods with 89.38% of accuracy. It is followed by SVM, DT, MLP, k-NN, and NB with 89.29%, 88.44%, 86.76%, 81.15%, and 79.92% of accuracies, respectively in Turk Telekom dataset. The classification accuracies of all machine learning algorithms exhibit similar attitude in remaining datasets. Thus, the classification success of the system can be generalized as following performance order: RF > SVM > DT > k-NN > MLP > NB.

In conclusion, it has been found that the Vodafone telecom operator data set exhibits better quality services and facilities compared to the other two terminals, due to the predominance of negative reviews in some respects. The finding of this study suggests that telecommunications operators should prioritize improvement in areas where there is a majority of negative criticism. In the future, we plan to evaluate the impact of deep learning methodologies on this telecom operator to improve the system's classification performance.

ACKNOWLEDGMENT

Part of this study was presented orally in the Third International Conference on Data Science and Applications 2020 (ICONDATA'20).

REFERENCES

[1] The definition "without being explicitly programmed" is often attributed to Arthur Samuel, who coined the term "machine learning" in 1959, but the phrase is not found verbatim in this publication, and may be a paraphrase that appeared later. Confer "Paraphrasing Arthur Samuel (1959), the question is: How can computers learn to solve problems without being explicitly programmed?" in Koza, John R.; Bennett, Forrest H.; Andre, David; Keane, Martin A. (1996). Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming. Artificial Intelligence in Design '96. Springer, Dordrecht. pp. 151–170. doi:10.1007/978-94-009-0279-4_9.

[2] Humera Shaziya, G.Kavitha, Raniah Zaheer, 2015, Text Categorization of Movie Reviews for Sentiment Analysis, International Journal of Innovative Research in Science, Engineering and Technology, Vol. 4, Issue11.

[3] Ahmed, S., Pasquier, M. and Qadah, G. 2013. Key issues in conducting sentiment analysis on Arabic social media text. In 9th International Conference on Innovations in Information Technolog (IIT), (Abu Dhabi, UAE, March 17-19, 2013). IEEE, 72-77. DOI= 10.1109/Innovations.2013.6544396.

[4] Mountassir, A., Benbrahim, H. and Berrada, I., 2012. In Bramer, M., and Petridis, M, A cross-study of Sentiment Classification on Arabic corpora. Research and Development in Intelligent Systems XXIX, Springer London ,2012, 259-272.

[5] Abdulla, N., Majdalawi, R., Mohammed, S., AlAyyoub, M. and Al-Kabi, M. 2014. Automatic lexicon construction for Arabic sentiment analysis. In Proceedings of the Future Internet of Things and Cloud (FiCloud), (Barcelona,Spain, 2014). IEEE , 547-552. DOI= 10.1109/FiCloud.2014.95.

[6] Ohana, B. and Tierney, B. 2009. Sentiment classification of reviews using SentiWordNet. In Proceedinf of the 9th. IT & T Conference, (Dublin, Ireland, October 22-23,2009). Dublin Institute of Technology, 13.

[7] Abdul-Mageed,M., Kübler, S. and Diab, M. 2014. SAMAR: A system for subjectivity and sentiment analysis of Arabic social media. Computer Speech and Language. 28, 1 (January. 2014),20-37.

[8] Ali Mustafa Qamar, Suliman A. Alsuhibany and Syed Sohail Ahmed, "*Sentiment Classification of Twitter Data Belonging to Saudi Arabian Telecommunication Companies*" in International Journal of Advanced Computer Science and Applications, Vol. 8, No. 1, 2017, pp. 395-401.

[9] D. Zimbra, M. Ghiassi, and S. Lee, "*Brand-related Twitter sentiment analysis using feature engineering and the dynamic architecture for artificial neural networks*" in 49th Hawaii International Conference on System Sciences, HICSS 2016, Koloa, HI, USA, January 5-8, 2016, 2016, pp. 1930–1938.

[10] Latifah Almuqren and Alexandra I. Cristea, "Twitter Analysis to Predict the Satisfaction of Telecom Company Customers", Journal of Big Data, 2019, Springer.

[11] Akshay Amolik, Niketan Jivane, Mahavir Bhandari, Dr .M. Venkatesan, Twitter Sentiment Analysis of Movie Reviews using Machine Learning Techniques, School of Computer Science and Engineering, VIT University, Vellore.

[12] Mesut Kaya, Guven Fidan, Ismail Toroslu, Sentiment Analysis of Turkish Political News, DOI: 10.1109/WI-IAT.2012.115, Conference: Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology - Volume 01.

[13] A.Onan, "Sarcasm identification on twitter: a machine learning approach", in Proceedings of CSOC 2017, Germany, 2017, pp.374-383.

[14] Tsigkritis, T. , Groumas, G. and Schneider, M. (2018) On the Use of k-NN in Anomaly Detection. Journal of Information Security, 9, 70-84. doi: 10.4236/jis.2018.91006.

[15] Martín-Valdivia M T, Rushdi Saleh M, Ureña-López L A, MontejoRáez A, "Experiments with SVM to classify opinions in different domains", Expert Systems with Applications, 38(12), 14799- 14804, 2011.

[16] J.Ren, S.D.Lee, X.Chen, B.Kao, R.Cheng, D.Cheung, "*Naive Bayes Classification of Uncertain Data*", Ninth IEEE International Conference on Data Mining, 2009. ICDM '09, pp. 944 – 949

[17] C. Horn, "Analysis and Classification," Graz University of Technology, Graz, Austria, 2010.

[18] Nehal Mamgai, Ekta Meht, Ankush Mittal, Gaurav Bhatt, "Sentiment analysis of top colleges in India using Twitter data", International Conference on Computational Techniques in Information and Communication Technologies, 2016.

[19] Breiman, L., Friedman, J.H., Olshen, R.A. and Stone C.J. (1984) Classification and Regression Trees. Wadsworth, Belmont, CA.

[20] Ho, Tin Kam (1995). *Random Decision Forests* (PDF). Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995. pp. 278–282.

[21] (July 23th 2019) A. A. Akın, Zemberek NLP Available url: https://github.com/ahmetaa/zembereknlp

[22] Oscar Araque, Ignacio Corcuera-Platas, J. Fernando Sánchez-Rada and Carlos A. Iglesias, *"Enhancing deep learning sentiment analysis with ensemble techniques in social applications"*, Expert Systems with Applications, vol. 77, pp. 236-246, 2017.