A Review on Emotion Recognition by Textual Tweets

[1]B.Prasanna Kumar, [2]T.Bhavani Shanvitha, [3]S.Jyothi Sri, [4]Y.Prathyusha, [5]S.Keerthana

[1]Assistant Professor, Department of CSE, KKR & KSR Institute of Technology and Sciences (A)

[2,3,4,5] B.Tech Students, Department of CSE, KKR & KSR Institute of Technology and Sciences (A)

Abstract: The proliferation of user-generated content on social media has made opinion mining an arduous job. As a micro-blogging platform, Twitter is being used to collect views about products, trends, and politics. Sentiment analysis is a technique used to analyse the attitude, emotions, and opinions of different people towards anything, and it can be carried out on tweets to analyse public opinion on news, policies, social movements, and personalities. By employing Machine Learning models, opinion mining can be performed without reading tweets manually. Their results could assist governments and businesses in rolling out policies, products, and events. Seven Machine Learning models are implemented for emotion recognition by classifying tweets as happy or unhappy. With an in-depth comparative performance analysis, it was observed that the proposed voting classifier (LR-SGD) with TF-IDF produces the most optimal result with 79% accuracy and 81% F1 score. To further validate the stability of the proposed approach on two more datasets, one binary, and another multi-class dataset, and achieved robust results.

I. INTRODUCTION

In recent times, social media sites like Twitter have produced vast volumes of organized, unstructured, and semi-structured data. One of the most recent examples is the COVID-19 infodemic, which demonstrates how false information spread through social media can have a significantly greater impact and be far more catastrophic than a natural event like a pandemic. To correctly assign sentiment classifications on a broad scale, analysis is required. Accurate NLP methods and machine learning (ML) models for text classification are needed to complete these tasks.

Twitter gives its users the chance to examine its data from a bigger and more comprehensive perspective. Due to the noisy nature of text input, efficient algorithms are crucial for automated labelling. Numerous research on the categorization of Twitter sentiment has been conducted in the past. Tweets are tiny posts that are sent over Twitter, which is a quick and effective micro-blogging platform that enables users to do so. Twitter is a popular social networking platform that is in high demand worldwide. Twitter allows users to create free accounts, which have the potential to reach a huge audience.

Twitter may be proven to be the ideal platform for business and marketing since it allows one to connect with extremely wealthy and well-known people, such as actors and celebrities, making their purchases appealing to both them and advertising. It can type a post or a website link but only with a small note range (140 characters for each post). Free and available to the public, including adverts. It is rapid because the public who is following the relevant business will immediately get a tweet once it is posted on Twitter. Utilizing this source allows businesses and advertising to examine a variety of operational viewpoints that are quite important. They will get a quick reaction from their followers as a result of this. Surprisingly, a large number of firms enhance their transactions by increasing their Twitter following.

Twitter helps its users by enabling them to discover new businesses, goods, and services, as well as websites, blogs, eBooks, and other things. Super Bowls, Grammy Awards, Academy Awards, and other significant sporting and entertainment events use it to create a lot of talk across the world. On Twitter, competition between various products is increasing. On social networks like Twitter, people love to share their opinions about a certain product. Product owners are prepared to spend more money on social media platforms in order to promote their goods more effectively and increase sales.

A person's experience with a product can assist the owner alter their marketing tactics, sales tactics, and product quality. Customer feedback is also given to owners or manufacturers through reviews. In order to categorise the consumer sentiment from the reviews given the volume of data created in this manner, a team of analysis experts is needed. It takes machine learning and ensemble learning classifiers to effectively classify the sentiment of the consumers because

experts might make human errors in sentiment analysis. By classifying tweets using TF and TF-IDF, this study evaluates several machine learning models for emotion identification.

Sentiment analysis
Sentiment analysis is a machine learning tool that analyzes texts for polarity, from positive to negative. By training machine learning tools with examples of emotions in text, machines automatically learn how to detect sentiment without human input.

To put it simply, machine learning allows computers to learn new tasks without being expressly programmed to perform them. Sentiment analysis models can be trained to read beyond more definitions, to understand things like, context, sarcasm, and misapplied words. For example:

"Super user-friendly interface. Yeah right. An engineering degree would be helpful."
Out of context, the words 'super user-friendly' and 'helpful' could be read as positive, but this is clearly a negative comment. Using sentiment analysis, computers can automatically process text data and understand it just as a human would, saving hundreds of employee hours.

Twitter
Twitter was created by Jack Dorsey, Noah Glass, Biz Stone, and Evan Williams in March 2006 and launched in July of that year. Twitter, Inc. is based in San Francisco, California and
has more than 25 offices around the world.

 Twitter is a social networking website and became very popular because of the lack of restrictions and frequencies of posts. Posts are called tweets and the maximum length of a tweet is only 140 characters. Twitter data is a popular choice for text analysis tasks because of the limited amount characters allowed and the global use of Twitter to express opinions on different issues among people of all ages.

II. LITERATURE REVIEW

Sentiment analysis inspires corporations to define clients' preferences about products, services, and brands. Further, it plays an important role in interpreting information about industries and corporations to reserve them in making entity review.

Sarlan et al. [1] established a sentiment analysis through extracting number of tweets with the help of prototyping and the results organized customers' views via tweets into positive and negative. Their research divided into two phrases. The first part is based on literature study which involves the Sentiment analysis techniques and methods that nowadays are used. In the second part, the application necessities and operations are described preceding to its development.

Alsaeedi and Zubair Khan [2] analyzed various kinds of sentiment analysis that is applied on to Twitter dataset and its conclusions. The distinct approaches and conclusions of algorithm performance were compared. Methods were used which were supervised ML based,, lexicon-
based, ensemble methods. Authors used four methods that were Twitter sentiment Analysis using Supervised ML Approaches; Twitter sentiment Analysis using Ensemble Approaches. Twitter sentiment Analysis is using lexicon based Approaches.

Lexicon based approaches have been explored by many researchers for emotion classification. Bandhakavi et al. [3] performed emotion-based feature extraction using domain specific lexicon generation. They captured association of words and emotions using a unigram mixture model. They used tweets that are weakly labelled to classify emotions. Their proposed architecture outperformed other state-of-theart approaches such as Latent Dirichlet Allocation and Point wise Mutual Information.

"Alsinet et al.[4]analysed tweets from political domains". They claimed accepted tweets are stronger as compared to rejected tweets. Rumor detection in tweets is performed by using an encoder to analyse human behavior in comments.

Hakh et al. [5] used SMOTE method to remove excessive challenges of Twitter dataset. In addition, they applied different feature selections for rapidity of sentiment analysis method. Authors projected methodology that was estimated beside the dataset application decision, squashy favorable results on all operated evaluation metrics. Pre-processingsteps were applied on their dataset after that they used TF-IDF features that were used to measure important weight of terms. Then classification methods were used (i.e. AdaBoost, Linear SVM, Kernel SVM, Random Forest, Decision Tree, Naïve Bayes and K-NN) and at last to relate classification's effectiveness: Accuracy and F1-score measures were used.

In [6], Xia et al. created the proportional training of the efficiency about collaborative method on behalf of Sentiment's arrangement. They set two types of feature in the context of sentiment analysis. Firstly, the feature set was totally depend on the part of speech and word relation was depending on the feature set. Secondly, the following familiar text classification algorithms that were maximum entropy,support vector machines and naive Bayes. Thirdly, the following ensemble strategies,that was the fixed combination, meta-classifier combination and weighted combination. They used 5 document-level datasets broadly utilized along with arena of Sentiment's arrangement. Experiments shown in this research the ensemble techniques are more effective than rest of the classifier which is also shown in our search that ensemble of two classifiers that are Logistics regression and stochastic gradient decent classifiers ensemble and give better result than other classifiers.

Rustam et al. [7] presented a Tweets Classification for US Airline Companies Sentiments. The researcher applied pre-processing on the dataset. The influence about feature extraction methods, together with TF, TF-IDF, along with word2vec, proceeding the classification accuracy has been examined. In addition, execution about the long short-term memory (LSTM) was studied in certain dataset. Paper of researcher proposes a Voting Classifier (VC) who helps to process similar administrations. Voting Classifier must dependent the Spatial Estimation (SE), Stochastic Gradient Descent classifier (SGDC) along with simple ensemble method for concluding results. Various types of ML classifiers tested with the use of precision, accuracy, recall and F1-score by way of working metrics. Results indicate that proposed VC is more efficient than one of the phase actors. The experiment also demonstrated the efficiency of machine learning students improved while TF-IDF utilizes a feature input.

Santos and Bayser [8] examined a sentiment analysis of short texts. In the experiment, researchers suggest a first-hand profound convolution neural network that achieve from character to sentence level material to accomplish sentiment analysis of little texts.

Mohamed [9] evaluated a sentiment analysis of mining halal food consumers. This examination fills this gap through the investigation of an irregular example of 100,000 tweets managing halal food. To lead the examination, a specialist predefined dictionary of seed descriptors was utilized. By investigating halal food feelings communicated via web-based networking media, this examination adds expansiveness and profundity to the discussion over such an underrepresented region. Distinct investigation recognized for the most part positive estimation toward halal food, while geo-found Twitter maps indicated that "strict diaspora" broadly utilizes computerized presents on impart about halal food.

Parveen and Pandey [10] studied sentiment analysis on Twitter dataset that uses NB algorithm. Analyst use Hadoop Framework for preparing film informational collection which is reachable on Twitter site as reviews, input and opinions. Sentiment analysis on Twitter data is explored in three classes that are positive, negative and neutral.

Alomari et al. [11] analyzed SVM utilizing TF-IDF. The study presented the Arabic Jordanian Twitter corpus where Tweets are explained seeing that any positive or negative. It researched distinctive directed machine learning opinion examination classifiers when applied to Arabic client's online life of general subjects that are found in either Modern Standard Arabic (MSA) or Jordanian tongue. Analyses were conducted to assess the utilization of various weight plans, stemming and N-grams terms strategies and situations.

Gamal et al. [12] built Twitter benchmark dataset for Arabic Sentiment Analysis. A benchmark Arabic dataset suggested in experiment for estimation investigation demonstrating social event strategy about the latest tweets in various Arabic vernaculars. The experiment dataset incorporates in excess of 151,000 unique assessments which marked into two classes, negative and positive. ML algorithms are functioned in SC; ML algorithm attached through learning arrangements. Sentiment analysis ordinarily executed using one fundamental methodology from a ML(lexicon-based approach) based approach. The calculations functioned via SC on the dataset accomplished 99.90% precision utilizing TF-IDF.

Kumar and Garg [13] explored the sentiment analysis of multimodal Twitter data. The experiment utilized a multi-method feeling examination approach to decide slant extremity mark for approaching tweet that is printed picture information realistic. Picture estimation marking was accompanied by utilizing SentiBank along with SentiStrength marking for Regions with convolution neural network (R-CNN). For a picture posted in Twitter, the picture module is executed which utilizes a current module of SentiBank along with R-CNN that decide the feeling estimation mark of the picture. After pre-processing, the content module utilizes an AI-based troupe strategy gradient boosting to characterize tweets into extremity classifications, to be specific, positive, negative or neutral High execution exactness of 91.32% is watched on behalf of arbitrary multi method tweet dataset utilize assess the planned model.

Sailunaz [14] investigated the feeling through the dataset that analyzed by a sentiment analysis from Twitter texts. The objective this work was to recognize and investigate assessment and feeling communicated by individuals from content in their Twitter posts and to use them for creating suggestions.

Mohri [15] Regarding unsupervised learning approaches, these algorithms try to find hidden structure in unlabeled data in order to build models for emotion classification. As occurs in supervised learning, among unsupervised learning proposals also it can be found systems based on categorical and dimensional emotion models.

III. REVIEW FINDINGS

On social networks like twitter, people love to share their opinions about a certain product. Product owners are prepared to spend more money on social media platforms in order to promote their goods more effectively and increase sales. A person's experience with a product can assist the owner alter their marketing tactics, sales tactics, and product quality.

Customer feedback is also given to owners or manufacturers through reviews. In order to categorize the consumer sentiment from the reviews given the volume of data created in this manner, a team of analysis experts is needed. It takes machine learning and ensemble learning classifiers to effectively classify the sentiment of the consumers because experts might make human errors in sentiment analysis. By classifying tweets using TF and TF-IDF, this study

evaluates several machine learning models for emotion identification. This study introduces a voting classifier (LR-SGD) and seeks to gauge how well known ML classifiers perform on twitter datasets.

IV. REFERENCES

[1] A. Sarlan, Chayanit Nadam, and S. Basri. Twitter sentiment analysis. Proceedings of the 6th International Conference on Information Technology and Multimedia, pages 212–216, 2014.

[2] Abdullah Alsaeedi and Mohammad Zubair Khan. A study on sentiment analysis techniques of twitter data. International Journal of Advanced Computer Science and Applications, 10(2):361–374, 2019.

[3] Anil Bandhakavi, Nirmalie Wiratunga, Deepak Padmanabhan, and Stewart Massie. Lexicon based feature extraction for emotion text classification. Pattern Recognition Letters, 93:133–142, jul 2017.

[4] Teresa Alsinet, Josep Argelich, Ramón Béjar, Cèsar Fernández, Carles Mateu, and Jordi Planes. An argumentative approach for discovering relevant opinions in twitter with probabilistic valued relationships. Pattern Recognition Letters, 105:191–199, apr 2018.

[5] Heba Hakh, Ibrahim Aljarah, and Bashar Al-Shboul. Online social mediabased sentiment analysis for us airline companies. 04 2017.

[6] Rui Xia, Chengqing Zong, and Shoushan Li. Ensemble of feature sets and classification algorithms for sentiment classification. Inf. Sci., 181:1138– 1152, 03 2011.

[7] Furqan Rustam, Imran Ashraf, Arif Mehmood, Dr. Saleem Ullah, Gyu Sang Choi, and Khan Yar. Tweets classification on the base of sentiments for us airline companies. Entropy, 21, 11 2019.

[8] Cicero Dos Santos and Maira Gatti de Bayser. Deep convolutional neural networks for sentiment analysis of short texts. 08 2014.

[9] M. Mohamed, ''Mining and mapping halal food consumers: A geo-located Twitter opinion polarity analysis,'' J. Food Products Marketing, vol. 24, pp. 1–22, Dec. 2017.

[10] H. Parveen and S. Pandey, ''Sentiment analysis on Twitter data-set using naive Bayes algorithm,'' in Proc. Int. Conf. Appl. Theor. Comput. Commun. Technol., Jan. 2016, pp. 416–419.

[11] K. M. Alomari, H. M. Elsherif, and K. Shaalan, ''Arabic tweets sentimental analysis using machine learning,'' in Proc. Int. Conf. Ind., Eng. Appl. Appl. Intell. Syst., Jun. 2017, pp. 602–610.

[12] Donia Gamal, Marco Alfonse, El-Sayed El-Horbarty, and Abdel-Badeeh M.Salem. Twitter benchmark dataset for arabic sentiment analysis. International Journal of Modern Education and Computer Science, 11:33 – 38, 01 2019.

[13] Akshi Kumar and Geetanjali Garg. Sentiment analysis of multimodal twitter data. Multimedia Tools and Applications, 78, 03 2019.

[14] Kashfia Sailunaz. Emotion and sentiment analysis from twitter text. Journal of Computational Science, 36, 07 2019.

[15]Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. 2012. Foundations of Machine Learning. MIT Press.