ISSN: 2278-4632
Vol-11 Issue-02 2021

TWEETS CLASSIFICATION FOR EMOTION RECOGNIZATION USING DEEP LEARNING

1. Prof.R.Yadagiri Rao,professor,Head,H&S,Sri Indu Institute ofEngineering&TechnologySIIET, Sheriguda,Ibrahimpatnam,Hydarabad,
2.A.Vijay Kumar,Assistant Professor,CSE,SIIET,Sheriguda,Ibrahimpatnam,Hydarabad
emailid:swarna.rekha1991.18@gmail.com
3.P.H.Swarnarekaha,Assistant Professor,CSE,SIIET,Sheriguda,Ibrahimpatnam,Hydarabad,
4.Sai kumar,Student,CSE,SIIET,Sheriguda,Ibrahimpatnam,Hydarabad
5.Praveen Kumar,Student,CSE,SIIET,Sheriguda,Ibrahimpatnam,Hydarabad
6.Praveen reddy ,Student,CSE,SIIET,Sheriguda,Ibrahimpatnam,Hydarabad

ABSTRACT:

Social networking platforms have become an essential means for communicating feelings to the entire world due to rapid expansion in the Internet era. Several people use textual content, pictures, audio, and video to express their feelings or viewpoints recognizes polarity. Opinion mining has become difficult due to the growing number of user generated content on social media. Sentiment analysis is a technique for analyzing people's attitudes, emotions, and opinions. It assesses whether the author has a negative, positive, or neutral attitude The analysis of social media posts is a challenging task, particularly the recognition of user emotions. Text is one of the most common mediums used by humans to express emotion, particularly on social media platforms. As emotions play a pivotal role in human interaction, the ability to recognize them by analyzing textual content has various applications in human-computer interaction (HCI) and natural language processing (NLP). Emotion analysis aims to automatically extract user emotional states from their social network text activity. Mining can be improved by utilizing Machine Learning models which can be carried out without manually reading tweets. For emotion, seven Machine Learning models have been implemented and Recognition is achieved by categorized tweets as positive and negative

.

INTRODUCTION

Automatic emotion recognition, pattern recognition and computer vision have become significantly important in Artificial Intelligence lately with applications is a wide range of areas. Recently, social media platforms such as Twitter have generated enormous amounts of structured, unstructured and semi-structured data. One of the most recent example is COVID-19 infodemic that shows

ISSN: 2278-4632
Vol-11 Issue-02 2021

misinformation in social media can be far more important and devastating than a disaster such as a
pandemic. There is a need to analyze to accurately assign sentiment classes on a large scale. To perform such tasks, accurate NLP techniques and machine learning (ML) models for text

classification are required. Twitter provides an opportunity to its users to analyze its data on a large and broader point of view. Efficient methods are important to automatically label text data due to its noisy nature. In the past
many studies have been performed on Twitter sentiment classification. As Twitter is very fast and an efficient micro-blogging examination that facilitates the end users to transmit small posts are said to be tweets. Twitter is a highly demanding app in the world and is a successful platform in social media.

Free account can be created by using Twitter that can provide an enormous audience potential. With the purpose of business and marketing, Twitter can be proved as the best platform, through
which one can get in touch with very rich and famous personalities like stars and celebrities, so their purchasing can be very charming for them as well as for advertisers. But, it has a note range;

up to 280 letters for each post and it can type a post or link on the website since it has no cost and
also open as the advertisements as well. There is no problem with clusters of personal ads which are similar to other social networking sites. It is quick because as a tweet is posted on Twitter, the public who is subsequent to respective business will get it without delay. Companies and advertisers can compose utilization of this source to check the diverse operational point of views which are veryconsiderable. With help of this, they will obtain an immediate response from their followers. Remarkably, a lot of businesses with the intention of purchase,Twitter followers increase their deals. Twitter facilitates the followers by making them identify
regarding fresh business, products, services, websites, blogs, eBooks etc. Consequently, Twitter clients might tick lying on link and also optimistically endow in a manufactured goods or examine the products presented and to get share in profit. It is extremely effortless to utilize as people can
follow to get the news and updates, as organizations can tweet or re-tweet, they can mark favorite or selected people to send the tweets, also know how to propel the posts plus to be able to endow
their money and instance through it. Academy, Industry, super bowls and Grammy Awards of such major Sports and Entertainment events

ISSN: 2278-4632
Vol-11 Issue-02 2021

generate a lot of buzz in the global world by using it.

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. Competition is rising among different products on Twitter. People love to express their feelings

about a particular product on social networks like twitter. Product owners are ready to spend more money on social media platforms to better advertise their products and to generate more revenue.

When a person shares experience about a product, it helps the owner to change their market strategy, selling schemes, and improving the quality. Customer reviews serve as a feedback to the

owners or manufacturers too. The data generated in such a way is of large amount and requires an analysis expert team to classify the customer sentiment from the reviews. Experts can make a

human error in sentiment analysis, therefore it requires machine learning and ensemble learning classifiers to accurately classify the sentiment of the customers. This study compares various machine learning models for emotion recognition by tweet classification using Tf and TF-IDF. This research aims to estimate the performance of famous ML

classifiers on twitter datasets. The key contributions are as follows: • Machine learning-based classifiers including support vector machine (SVM), Decision Tree Classifier (DTC), Naive Bayes (NB), Random Forest (RF), Gradient Boosting Machine (GBM) and Logistic Regression (LR) trained on Twitter dataset are compared for

emotion recognition.

LITERATURE SURVEY:

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. 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. In another

research Alsaeedi and Zubair Khan analyzed various kinds of sentiment analysis that is

ISSN: 2278-4632
Vol-11 Issue-02 2021

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, lexiconbased, 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. performed emotionbased 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

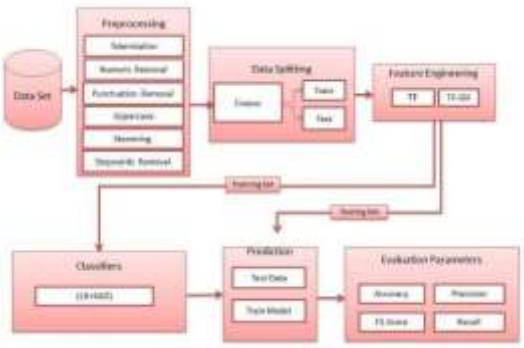
outperformed other state-of-the art approaches such as Latent Dirichlet Allocation and Point wise Mutual Information. Event related tweets are

identified by researchers on geo related tweets. They used specific tweets of local festivities in one year. They also identified different parameters that helped in event discovery. Alsinet et al. analyzed tweets from political domains. They claimed accepted tweets are stronger as compared to the rejected tweets. Rumor detection in tweets is performed by using an encoder to analyze human behavior in comments.

Pre-processing steps 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

System Architecture:



EXISTING SYSTEM:

Deep learning has been utilized by many researchers for image classification and tweet classification. The researcher applied pre-processing on the dataset. The influence about

ISSN: 2278-4632
Vol-11 Issue-02 2021

feature extraction methods, together with TF, TF-IDF, along with word2vec, proceeding the classification accuracy has been examined. The experiment dataset incorporates in excess of 151,000 unique

assessments which marked into two classes, positive and Negative. 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. 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 Senti Strength 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. The dataset is utilized to recognize slant and feeling from tweets and their answers and estimated

the impact scores of clients dependent on different Tweet based and client based parameters

PROPOSED SYSTEM:

In this proposed system different techniques have been used for methodology in ML for its objectives. Versatile experiments were examined using different methods and techniques. Multiple classifiers applied on the dataset, but the Voting classifier is an ensemble of Logistic Regression

and Stochastic Gradient Descent outperforms than all other ML models in terms of accuracy, recall, precision and F1-score. Twitter dataset used in this experiment is scrapped from Kaggle repository. First the dataset is pre-processed by removing unwanted data. Then, the data was split into two sets:

training set and testing set. The training set was given the percentage of 70% while the test set portion is 30%. After that feature engineering techniques are applied on the training set. Multiple

machine learning classifiers are trained on the training set and tested using the test set. The evaluation parameters used in this experiment are: (a) Accuracy (b) Recall (c) Precision (d) F1- score. In this section classifiers utilized for tweet classification will be discussed. This work

ISSN: 2278-4632
Vol-11 Issue-02 2021

utilized five supervised machine learning algorithms: Support Vector Machines (SVM), Naive Bayes (NB),

Random Forest (RF), Decision Tree (DT), Gradient Boosting model (GBM), Logistic Regression (LR) and Voting Classifier (Logistic Regression + Stochastic Gradient Descent classifier).

IMPLEMENTATION

ALGORITHMS:

Lexical Based Approach:

The Lexicon-based approach uses pre-prepared sentiment lexicon to score a document by aggregating the sentiment scores of all the words in the document. The pre-prepared sentiment lexicon should contain a word andcorresponding sentiment score to it. The negation form of vocabulary words should be added to the lexicon as individual entries, and they should be given higher precedence over the corresponding non negation terms. Simple rules can also be used to

handle negation terms.

Sentiment Analysis:

Sentiment analysis, also referred to as opinion mining, is an approach to natural language processing (NLP) that identifies the emotional tone behind a body of text. This is a popular way

for organizations to determine and categorize opinions about a product, service, or idea. It involves the use of data mining, machine learning (ML) and artificial intelligence (AI) to mine text for

sentiment and subjective information.

Vader Sentiment Analysis:

VADER (Valence Aware Dictionary and sentiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media.It is used for

sentiment analysis of text which has both the polarities i.e. positive/negative. VADER is used to quantify how much of positive or negative emotion the text has and also the intensity of emotion.

Random Forest:

RF is a tree based classifier in which input vector generated trees randomly. RF uses random features, to create multiple decision trees, to make a forest. Then class labels of test data are

predicted by aggregating voting of all trees. Higher weights are assigned to the decision treeswith low value error. Overall prediction accuracy is improved by considering trees with low error

rate.

Support Vector Machine

ISSN: 2278-4632
Vol-11 Issue-02 2021

The Support vector machine (SVM) is understood that executes properly as sentiment analysis. SVM typifies preference, confines and makes usage of the mechanisms for the assessment and

examines records, which are attained within the index area. Arrangements of vectors for every magnitude embody crucial details. Information (shown in form of vector) has been arranged in type to achieve this target. Next, the border is categorized in two training sets by stratagem. This is a long way from any area in thetrainingsamples .Support-vector machines in machine learning

includes focused learning models connected to learning evaluations which inspect material that is exploited to categorize, also revert inspection .

Naive Bayes

Ordering approach, Naïve Bayes(NB), with sturdy (naive) independent assumptions among stabilities, depends on Bayes' Theorem. NB classifier anticipates that the proximity of a specific

element of class that is confined to the closeness of a couple of different variables. For instance, a natural organic product is presumably viewed as an apple, if its shading is dark red, if type of it is round and it is roughly 3 creeps in expansiveness. In machine learning, Naive Bayes classifiers are a gathering of essential "probabilistic classifiers" considering applying Bayes' speculation with gullible opportunity assumptions between the features. They are considered as the minimum problematic Bayesian network models.

Decision Tree

Algorithm is the category of supervised ML and is being widely used in regression and classification tasks. Selection of root node of a tree of each level is its main challenge which is called as attribute selection Gini index and information gain are most commonly used methods for attribute selection. In this study, gini index is used to find probability of root node by calculating sum of squares of attribute values and then subtracted by 1

Gradient Boosting Machine

It is a ML based boosting model and is widely being used for regression and classification tasks, which works by a model formed by ensemble of weak prediction models, commonly decision trees. In boosting, weak learners are converted to strong learners. Every new generated tree is a modified form of previous one and use gradient as loss function. Loss calculate the efficiency of model coefficients fitting over

underlying data. Logically loss function is used for model optimization.

Logistic Regression

In LR class probabilities are estimated on the basis of output such as they predict if the input is from class X with probability x and from class Y with probability y. If x is greater than y, then

ISSN: 2278-4632
Vol-11 Issue-02 2021

predicted output class is X, otherwise Y. Insight, a logistic approach used for demonstrating the probability of a precise group or else, occurrence is obtainable, e.g., top/bottom, white/black,

up/down, positive/negative or happy/unhappy. This is able to stretch out and to show a small number of classes about events, for example, to make a decision if a image includes a snake, hound, deer, etc., every article being famous in the image would be appointed a probability wherever in

the series of 0 and 1 with whole addition to one.

Stochastic Gradient Descent

Gradient Descent'stypes include Stochastic Gradient Descent (SGD). SDGD is an iterative strategy for advancing a target work with appropriate perfection properties (for example differentiable or sub differentiable) .Degree of advancement is calculated by it in light ofdevelopment of alternative variables. It is very well, may be viewed as a stochastic guess of inclination plummet advancement,

since it replaces the genuine angle (determined from the whole

informational index) by a gauge

thereof (determined from an arbitrarily chosen subset of the information).

RESULT ANALYSIS

Fig. Run Server

Fig.URL Generation

Fig. Output Screen

Fig.Input text given for analysis

ISSN: 2278-4632
Vol-11 Issue-02 2021

Fig. Result Screen

This is the final result of the given text. It shows that the given text is positive. Here 1.0 describes 100 percent and 0.0 describes 0 percent.

CONCLUSION AND FUTURE SCOPE

The classifying tweets as positive or negative. Our experiments showed that one can improve the performance of models by recognizing patterns efficiently and through effective averaging combination of models. The use of LR and SGD as a classifier for emotion recognition by categorizing tweets as negative, positive, happy, or unhappy. Our experiments demonstrated that by recognizing patterns efficiently and combining models effectively, one can improve the performance of models. A review of the existing techniques for both emotion and sentiment detection is presented. As per the paper's review, it has been analyzed that the lexicon-based technique performs well in both sentiment and emotion analysis. The performance of machine learning algorithms and deep learning algorithms depends on the pre-processing and size of the dataset. In situations where the dataset is vast, the deep learning approach performs better than machine learning.

As future work we see a need for combining machine learning classifiers learned from crisis domain data with more general affective lexicons. In this way we think that better classification performance can be achieved than using the methods individually. Moreover, we suggest extending the used feature set with extracted part-of-speech tags since such information most likely will help determine if it is the author of a tweet who is having a certain emotion, or if it is someone else. Other areas to look into is how to deal with the use of sarcasm and slang in the user generated content.

From a crisis management perspective, it will also be necessary to investigate to what extent the used methodology and the developed classifiers are capable of coping with more generic

ISSN: 2278-4632
Vol-11 Issue-02 2021

situations. That is, we hope to have developed classifiers that to at least some significant extent classify based on hurricane and crises behavior in general, rather than solely being able to classifying Sandy
specific data. Investigating this requires that one retrieves and tags new
datasets to test the classifiers on. Doing this for several different crisis types and then applying the same classifiers, should make it possible to quantify how capable the developed classifiers are.

REFERENCES

[1] N. F. F. da Silva, E. R. Hruschka, and E. R. Hruschka, ‘‘Tweet sentiment analysis with classifier ensembles,’’ Decis. Support Syst., vol. 66, pp. 170–179, Oct. 2014.

[2] C. Kariya and P. Khodke, ‘‘Twitter sentiment analysis,’’ in Proc. Int. Conf. Emerg. Technol. (INCET), Jun. 2020, pp. 212–216.

[3] A. Alsaeedi and M. Zubair, ‘‘A study on sentiment analysis techniques of Twitter data,’’ Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 2, pp. 361–374, 2019.

[4] A. Bandhakavi, N. Wiratunga, D. Padmanabhan, and S. Massie, ‘‘Lexicon based feature extraction for emotion text classification,’’ Pattern Recognit. Lett., vol. 93, pp. 133–142, Jul. 2017.

[5] T. Alsinet, J. Argelich, R. Béjar, C. Fernández, C. Mateu, and J. Planes, ‘‘An argumentative approach for discovering relevant opinions in Twitter with probabilistic valued relationships,’’
Pattern Recognit. Lett., vol. 105, pp. 191–199, Apr. 2018.

[6] W. Chen, Y. Zhang, C. K. Yeo, C. T. Lau, and B. S. Lee, ‘‘Unsupervised rumor detection based on users’ behaviors using neural networks,’’ Pattern Recognit. Lett., vol. 105, pp. 226–233, Apr. 2018.

[7] H. Hakh, I. Aljarah, and B. Al-Shboul, ‘‘Online social media-based sentiment analysis for us airline companies,’’ in New Trends in Information Technology. Amman, Jordan: Univ. of Jordan, Apr. 2017.

[8] R. Xia, C. Zong, and S. Li, ‘‘Ensemble of feature sets and classification algorithms for sentiment classification,’’ Inf. Sci., vol. 181, no. 6, pp. 1138–1152, Mar. 2011.