An Application of Sentiment Analysis Based on Hybrid Database of Movie Ratings

Aanchal Badgaiyya[1], Prachi Shankarpale[2], Rutuja Wankhade[3], Utkarsha Shetye[4], Prof. Kalyani Gholap[5], Prof. Sagar Pande[6]

[1,2,3,4]Student's Dept. of Computer Science & Engineering , Dhamangaon Education Society's College of Engineering and Technology, Dhamangaon Rly. , Maharashtra, India.
[5]Assistant Professor, Dept. of Computer Science & Engineering , Dhamangaon Education Society's College of Engineering and Technology, Dhamangaon Rly. , Maharashtra, India.
[6]Assistant Professor, Dept. of Computer Science & Engineering, Lovely Professional University. , Phagwara, India.

---***---

Abstract - *Sentiment analysis is one of the popular methodologies for the identification of the sentiments or emotions or feelings of the people based on the data collected digitally. It is playing a major role in various fields and particularly in E-commerce and entertainment-related aspects. This particular aspect is considered for the basis of our proposed framework. The major aspect of our proposed framework is generating a database from two various sources such as using already existing data and the data extracted from Twitter and the data considered was movie ratings. For extracting data from Twitter we utilized the natural language processing concepts using python. Once the complete data generated then the data passed on to the BERT model to identify the various features which play a crucial role in binary classification as per the ratings as good or bad. Through this model, we able to attain an accuracy of 95.9% and its consistency will be supported by other evaluation metrics such as precision, recall, and F1-score.*

***Key Words*: Multi-variate correlation, KNN, correlation detection of anomalies, intruder identification, error detection in the channel.**

1. INTRODUCTION

There has been a massive increase in the usage of microblogging sites like Twitter in the past few years. Businesses and media organizations are increasingly trying to figure out ways to explore Twitter for knowledge about what individuals think and feel about their goods and services, driven by that development. Companies such as Twitratr, tweetfeel, and Social Note are only a few that as one of their offerings promote Twitter sentiment analysis. There exist a decent content of research for identifying the various sentiments that are communicated across various domains like online reviews and news stories. Yet, a few contents of research has been conducted for identifying the various sentiments are communicated considering the informal speech and message-length restrictions of online media. Apps like automated element tags and techniques like sentimental lexicons have proven effective in many other fields for sentiment analysis, and yet can they still prove beneficial on Twitter for sentiment analysis? The enormous range of subject matter discussed is yet another microblogging issue. Saying that users tweet on everything and anything is not a misconception. Therefore one needs a framework for easily defining details that could be utilized for learning is required to be allowed to develop a framework for mining Twitter's feelings on any given subject. Thus, this report discusses one strategy for constructing such information utilizing hashtags that are utilized in Twitter to classify various sentiments like positive, negative, and neutral and utilize various-way emotion classifiers for learning. Networking sites offer an incentive for companies by offering a forum for ads to communicate with their clients. For decisionmaking, users often rely to a large degree on customer-based content online. For example, if some user wants to purchase an item or wishes to utilize some utility, the user would be looking for the corresponding ratings online, which follows chat about that on networking sites before making a decision. For a typical user to examine, the volume of information created by users is too large. There is a big urge to simplify this since different types of sentiment analysis are commonly utilized. Analyzing various sentiments advises the customer whether the goods or the services obtained would be satisfactory not in advance of buying those products. This type of research information is used by advertisers and companies to recognize their goods or services according to the needs of customers that can be delivered.

Methodologies for the extraction of textual data primarily concentrate on extracting, scanning, or evaluating the current solid evidence. The information has an analytical dimension, but contextual aspects are reflected in certain other textual material. These aspects form the heart of the Sentiment Analysis, primarily thoughts, feelings, judgments, behaviors, and emotions. It provides many daunting possibilities for emerging innovations to be created, primarily due to the tremendous development of inaccessible knowledge from online outlets such as blogs and social networks. For an instance, it is possible to forecast suggestions for products suggested by a recommendation framework by considering factors like positive or negative perceptions on those products by using the sentimental analysis.

1.1 Sentiment Analysis

The automatic method of processing textual information and organizing it into positive, negative, or neutral sentiments is considered as the sentiment analysis. The use of various tools for the implementation of the sentiment analysis based on social networking site information to interpret views may allow enterprises to learn how customers speak about their products or services. Proud networking site Twitter has 330 million active users on an average monthly, allowing marketers to access a wide range of audiences without mediators to communicate with consumers. There is so much data on the underside that it is difficult for marketers to spot unfavorable social mentions easily that might damage their reputation. That provides the reason for the necessity of social listening has become a core technique in social media marketing, which entails tracking interactions on social networking sites. Listening to various users of social networking sites helps businesses to consider their following, catch up with the intention of the users and their intention towards the company and their competition, and uncover emerging market trends.

1.2 The essentiality of sentiment analysis

Assume having just introduced a new brand item feature and seeing a sudden spike in trend in social networking site mentions. This raises a big question that, are consumers tweeting hugely on this item as the latest feature delights them? Or are they moaning about the feature? It will take entirely too much money to move into any of these remarks individually. One would miss out on useful reviews which might help them boost the satisfaction of a client with the new feature immediately. One can easily grasp the tone and meaning of social citations on social networking sites by running sentiment analysis along with various techniques of machine learning.

The various advantages of twitter-based sentiment analysis can be mentioned as follows:

Scalability: The particular brands related to a large number of posts on networking sites can be analyzed automatically instead of manual analysis. As networking site information increasing and acquires useful insights by easy scaling of sentiment analysis tools.

Real-Time Analysis: Analysis of various posts in social networking site-based sentiments are critical for tracking abrupt changes in consumer feelings, predicting when concerns are on the increase, and taking steps before issues worsen. In realtime, one can track brand references on social networking site-based user posts with sentiment analysis and obtain valuable apprehensions.

Consistent Criteria: Stop discrepancies resulting from human mistakes. From each piece of information, customer representatives may not always settle about which tag to utilize, so one can finish up with incorrect results. Rather than utilizing one system of regulations, machine learning techniques conduct sentiment analysis, so one can guarantee that all their information is reliably tagged.

The essentiality of sentiment analysis

The twitter-based sentiment analysis consists of various crucial steps to attain the sentiments that hidden the text format. Those steps can be represented as follows:

Gathering the information from the social networking site, Twitter using keywords as well as the tags.

Process the obtained data into a structural format so that it would be ready for further analysis.

Create a model that withdraws the hidden sentiments in the information obtained from Twitter.

Visualization of the results based on the classification of sentiment aspects is necessary for the obtained results.

The proposed framework is based on Twitter sentiment analysis. In this aspect, we considered the Movie revies as the basis. By using various keywords and tags able to extract the information. Such information is necessary to market the film based on the sentiment in the market. Particularly, social media playing a vital role in affecting the interest of the audience towards a specific movie. In this aspect, social media can bring a major section of the audience to that specific movie and at the same time the same social media can throw out the major section of the audience not to come for that specific movie. So, the sentiment analysis able to identify the sentiments that are floating in the social media, and accordingly remedies can be planned before reaching the worst scenario. The article is organized into various sections. Section 2 discusses the literature related to sentimental analysis in various aspects, section 3 discusses the overview of the implemented proposed framework and related methodology, section 4 discusses the dataset and related software requirements along with the various libraries utilized for the implementation of the proposed framework, section 5 discusses results obtained from the proposed framework and section 6 discusses the conclusion future work as well as the conclusion based on the implementation of the proposed framework.

2. LITERATURE REVIEW

Abdullah Alsaeedi and Mohammad Zubair Khan 2019[1] discuss various Twitter-sentiment techniques and methods of analysis including machine learning have been discussed, dictionary (lexicon) focused on approaches. Furthermore, hybrid and Twitter ensemble techniques in emotion analysis have been investigated. The results showed the techniques of machine learning; for instance, the SVM and MNB have been the most accurate, especially if there have been several features. SVM facilities traditional learning techniques for classifiers can be considered and the strategies focused on dictionary (lexicon) are very viable some days, no effort in the archive marked with human beings. Algorithms for machine learning like Naive Bayes, Entropy, and SVM have been reached accuracy around 80 percent when the model was n-gram and bigram used. Hybrid-based Twitter sentiment analysis and ensemble algorithms were more successful than the supervised machine learning techniques so they achieve the accuracy of the description is around 85%. Twitter Ensemble was a commonly predicted method of sensation processing, it would be better than supervised algorithms for machine learning, as mixed many classifiers and often separate model functions. But the hybrid approaches were still popular and successful fair accuracy ratings so they must profit from all classifiers in machine learning and sentiment analytics approaches for lexicon-based Twitter.Chen, X. et al.2018[2] proposes a new approach to extract the SA feature from product feedback is suggested in the report. Initially, the widespread TF – IDF function vectors are obtained from diversified product analysis expression types and implementation of semantic synonym. Then the local because of the numerous product reviews with OPSM biclustering algorithm detects function vector patterns.

Finally, we boost the algorithm of PrefixSpan detection of the strong discriminatory and repeated pseudo-consecutive phrases, Contain word order information. Besides, other essential considerations are differentiation and the ability to differentiate words, are often used to increase the polarity of sentiment's discrimination ability. The text is based on the above measures vectors of operation are eliminated. A variety of experimental and comparative findings show SA's performance on the review of the commodity is improved substantially.Gui, L. et al.2017[3] use a heterogeneous network to design the shared polarity in product evaluations and learn consumers' representations, goods simultaneously reflected on and use terms. The fundamental concept is to create a heterogeneous network that connects consumers, goods, and words in the product feedback and the term polarity. Based on the network installed, node representations are taught employing a network embedding process that is subsequently integrated for sentiment analysis into a coevolutionary neural network. Consumer analysis assessments including IMDB, Yelp 2013, and Yelp 2014 data sets, suggest that the methodology being proposed achieves state-of-the-art efficiency.

Haque, T. U. et al.2018[4] discuss the world that we see today is growing more digitalized. E-commerce takes place in this digitalized environment by having links to products for customers who don't have to leave their house. As we now trust online goods, so examination of the product increases. For a product range, a customer must move through thousands of feedbacks of a product. But on this flourishing day of automatic learning, it would be much better if a model were a thousand ratings by using to polarise and benefit about these reviews. We have been using the method of supervised learning on a broad amazon dataset to polarize and achieve sufficient precision. 90% off-score, precision, and accuracy are achieved. Ireland, Robert, and Ang Liu 2018[5] Advanced data analysis are one of the most technological advancements in the 21st century, which makes the discovery of patterns by sophisticated computer methods. Millions of online product reviews, e-commerce, and social media have been released by clients and they can offer designers invaluable insights into product design. A summary of this article provides an online product evaluation research concept framework. The goal is to use this data generated by the computer to identify a variety of consumer criteria. The system is designed to transmit vast amounts of quality data into quantitative insights into product characteristics for manufacturers to make informed choices. Any case shall be used to affirm the effectiveness of the proposal process is used for a product report on the E-Commerce website and Amazon. The system reveals a statistical approach to online product feedback research. The system behaves like an interface between quantitative success and the qualitative and innovative design process. Additional performance review identifies much of the rational, algorithm approaches mixed into the exceedingly subjective and imaginative design process.

Walaa Medhat et al 2014[6] consider some papers refer to several fields connected with SA which is used for different fields in the modern world. Afterward, it is clear from the review of these papers that the algorithms of SC and FS are still an open research area. Naive Bayes and support vector machines are the most popularly used ML algorithms to solve the problem of SC. It is a reference model in which several proposed algorithms are considered and compared to the proposed algorithm. The emphasis in this area is on languages other than English, as the tools and study related to these languages are not yet accessible. The most important source of the lexicon is WordNet, which is used in other than English languages. Many people do require building tools, used for SA tasks and natural languages. The information from the media is an important aspect of communicating the mood of people or opinions on a particular subject or product. Usage of social network websites and microblogging websites as a database that includes a more detailed review. Kiran, V. K. M et al.2017[7] discuss the suggestion of the products to target people who meet your needs are very relevant to vendors specifications in the international economy to survive. The suggested solution is novel and a better solution in this article to score a commodity-based on its requirements analysis of a vast number of customer ratings that are derived from many leading e-commerce websites. This prevents the user from browsing for online views and suggestions before buying. The method suggested for this analysis extracts battery, CPU, specifier list camera and customer reviews mentioned for users specific product from numerous websites and recognizes vital items terms of the technological characteristics of the product for determining the polarity of the product and classification under the list of requirements. A score based on polarity is given to each specification i.e. reviews positive/negative feedback. The overall rating of the product is calculated by applying the relevant score to a single attribute. This is a useful strategy for customers who target particular characteristics in a product. Efthymios Kouloumpis et al.2011 [8] discuss the usefulness of language components for Twitter tweets sentiment detection. We test the accessibility and quality of current lexical tools and capturing knowledge on microblogging's casual and imaginative vocabulary. We follow a controlled strategy to the problem, but to use existing Twitter hashtags for building data instruction. Current features of sentiment lexicon are mildly useful for microblogging functions, however, the features of microblogging (i.e. positive/negative/neutral) were most useful, undoubtedly. It was useful to use hashtags to gather training data, as data from positive and negative emotions are gathered. However what technique yields better information on training and whether the two training sources are the type of features used can depend on the complement. Our results demonstrate that the advantage of emotion training data is reduced when microblogging characteristics are included.

Ruxia Liang and Jian-Qiang Wang 2019[9] adopted method contains three modules: a collection of knowledge, information integration, and transformation. We use the knowledge collection module to collect intuitional language in each analysis, fuzzy knowledge is evaluated by emotions. The Info transformation Module is also used to convert intuitionist knowledge from language natural insight (Lincs) clouds. The integration module is used to get each product to have Lincs. Alternate rating products are determined. A Taobao.com case study is to demonstrate the performance and effectiveness then sensitivity and contrast of the proposal analyzes, to check its stability and superiority. Mohamed M. Mostafa 2018[10] complete this analysis in this reporting gap via a 3,919 halal-food tweets random sample study. The study was carried out using a 6,800 seed adjective lexicon specialist. A usually good mood towards halal food was identified by descriptive statistics while dividing the clustering of medoids has shown that halal food consumers can be clustered in four separate segments. Halal food consumers thus tend to be a highly heterogeneous community, divisible by religiosity degree, self-identity, animal care, and food consideration authenticities. Mubarok.

3. METHODOLOGY

The proposed framework is intended to classify the movie based reviews into two categories like positive and negative sentiments. The dataset considered was a combination of the existing data as well as the data extracted from Twitter. This section divided into various subsections such as subsection – A discusses the process followed for the gathering of the data, subsection – B discusses the model utilized for sentimental analysis and the usage of the BERT model, subsection – C discusses the advantages and disadvantages related to sentiment analysis, and subsection – D discusses the flowchart and algorithm of the proposed framework related to sentiment analysis.

3.1 Gathering Dataset

The dataset considered is a combination of the information extracted from Twitter and the existing data based on movie reviews. For the extraction of information from Twitter, various steps need to be followed.

Step 1: Set up the Twitter application either in a web browser or on a smartphone and generate an account. Usually, a web browser is highly recommended.

Step 2: Get the developer access to the Twitter application by creating an application on Twitter. Once the user gets the developer access, then Twitter provides a consumer key, consumer secret key, access token key, and access token secret key which are crucial and private for a user for extraction of the information from Twitter.

Step 3: By importing the library tweepy we can use various functionalities to attain the authentication.

Step 4: Search the tweets using various search words that are related to movies and the date as well to obtain the data from that specific date to the date and time of the data extraction.

Step 5: The data obtained considered to be raw data combined with the existing data along with ratings in a structured manner as per the existing data.

Step 6: The complete data need to be processed using the nltk library by removing the common words, punctuations, and stop words.

Step 7: Thus, the processed data considered for further analysis to classify the sentiments that are hidden in the reviews. The discussed data gathering is the major step in sentiment analysis. Once data gathered and processed then the remaining part is easier to analyze.

3.2 Dataset and System Description

The dataset comprises two sections of datasets, of the first dataset is gathered from Twitter data and the other dataset was already a build-in dataset obtained from open source, popularly known as Kaggle. Thus, the obtained two sections of datasets combined to form a single dataset which consists of 25000 instances. It is implemented on Windows 10 operating system with RAM of 18 GB and Intel® Core™ i3-8130U CPU @ 2.20GHz – 2.21GHz. The programming language utilized is Python with various libraries. The libraries utilized and their functions are mentioned as follows:

Tensorflow 2.3.1: It is an open-source library-based tool for numerical computing with high efficiency. Its modular platform provides computing to be efficiently distributed across a range of platforms like CPUs, GPUs, TPUs, from desktop computers and database clusters to smartphone and edge computers.

Django 3.1.2: With the aid of this library one can launch various applications based on the web that can be constructed very easily from the stage of a very basic concept to launch within hours of duration easily. Django keeps track of most of the Web creation difficulty, meaning one can concentrate on developing the application without reinventing the machine. On a final note, it is an open-source and free tool.

NLTK 3.5: The full form for NLTK is Natural Language Toolkit and it will be working on the concepts of natural language processing i.e, human languages and it is a pack of libraries related to text processing like tokenization, stemming, wrapping, parsing, categorization, and, etc. It is an open-source and free tool.

Tweepy 3.9.0: It is a python environment-based library that is utilized for retrieving the Twitter API which is responsible for the extraction of the Twitter data by using keywords, and tags. It is an open-source and free tool.

Ktrain 0.22.3: It is a very useful library that is utilized for the building of the neural network, training the neural network, and deploying the neural network as well as certain machine learning methodologies by wrapping with the library TensorFlow Keras.

Pandas 1.1.4: It is useful in attaining the data in a structured format as if in a tabular structure. It is the basic library essential for any analysis dealt with structured data. It stores the data in a data structure format popularly known as Dataframe. This library also an open-source and free tool.

Keras 2.4.3: It is an efficient API used for the deployment of Artificial Neural Network (ANN) based on Python. By using this library, a very high complex neural network can also build and deployed very easily and it works along with the TensorFlow library. It is an open-source and free tool.

Shutil: The shutil library contains a variety of high-level file operations as well as file arrays. Functions that enable file copy and deletion are offered in general. It is an open-source and free tool.

Os: This library presents a compact manner of utilizing features based on the operating system. It includes various operations like open, manipulation of the existing paths, or reading all the lines. It is an open-source and free tool.

3.3 Model Utilized for Sentiment Analysis

The neural network model is a pre-trained based on natural language processing utilized for sentiment analysis and that modeis popularly known as BERT. The full form of BERT is "Bi-directional Encoder Representation from Transformers". Even Google also utilizes this BERT model for the understanding of various natural languages. In simple terms, it is a combination of a neural network as well as natural language processing. Neural networks will mimic the human neural system and be developed for the identification of patterns, classifications based on both texts, and images. Nowadays these networks are popularly used in financial projects for the prediction aspects. Natural language processing can also be considered as NLP corresponds to a branch of AI that interacts with linguistics to enable computers to learn and understand the communications that humans naturally considered. The major applications of NLP are social listening, sentiment analysis, and word suggestions, chatbot, and so on. The way of training the models conventionally got broken by the BERT model. The training in this particular model will be done in a to and forth way (from first to last and again from last to first). A complete sequence of words will be passed onto the model and the model will get trained as mentioned previously. Instead of only the word that primarily precedes or follows it, BERT enables the language framework to understand word meaning based on contextual terms. Google refers BERT model as "deeply bi-directional" as from the very bottom of a deep neural network," the contextual interpretations of words start.

The usage of the BERT model:

It is relatively easy to use BERT for a particular operation such as natural language processing. For a wide range of language operations, BERT could be utilized by only injecting a single layer to the core framework.

1. Comparably to the following Sentence categorization, categorization tasks like sentiment analysis are performed by injecting a categorization layer for the [CLS] token on top of the Transformer output.

2. The program receives a query about a text sequence in query answering activities and is expected to label the response in the sequence. A query and answering model can be learned utilizing BERT by training two additional vectors that label the start and the finish of the response.

3. The program attains a linguistic sequence in the Named Entity Identification (NEI) and is expected to label the different types of entities that appear in the textual considerations. By loading the resultant vector of each token into a categorization layer that forecasts the NEI label, an NEI model can be learned utilizing BERT.

Certain takeaways of the BERT model can be considered as discussed as follows:

1. Even on a large scale, the framework size is important. The highest framework of its kind is BERT-large, with a huge number of parameters is 345 million. It is a higher-level to BERT-base, which utilizes a similar structure with a small number of parameters is 110 million on small-scale aspects.

2. A greater number of training steps indicates attaining a higher precision with sufficient training data.

3. This bidirectional methodology converges very slower than the left-to-right methodologies.

3.4 Advantages and Disadvantages of Sentiment Analysis

One can measure how clients perceive various aspects of one's organization by utilizing sentiment analysis without reading thousands of consumer posts at once. It is difficult for an individual to read all of those comments if one had thousands of reviews every month. One can easily mine into various consumer segments of one's market and get a deeper knowledge of sentiment in these segments by utilizing sentiment analysis and streamlining this procedure. Analysis of sentiment is a valuable resource for any entity or organization for which popular opinion or behavior concerning them is essential for their achievement, regardless of the concept of success. The outcomes of sentiment analysis aid companies identify the discussions and interactions that occur about them and allow them to respond and act accordingly. To evaluate student responses obtained either out of their polls or from online outlets such as social media, educational institutions may utilize sentiment analysis. To recognize and resolve any instances of student discontent, they will then utilize the findings, as well as recognize and expand on those situations where students share positive opinions. Sentiment analysis is not a complete substitute for analyzing survey responses, although it is beneficial. Sometimes in the posts themselves, there are valuable variations. Where one can additionally assist with sentiment analysis is by defining which of those posts one can read. Computer programs have difficulties comprehending details such as sarcastic humor, denials, comedy, and exaggerations - the kinds of details an individual will have little difficulty in recognizing. And the findings can be distorted by failing to consider these. In certain aspects, the word 'disappointment' might be considered as a negative aspect but in the cases of 'wasn't disappointed' needs to be recognized as positive but it will identify as negative instead of positive. So, it is very essential to understand the scenario and various aspects around it need to be considered before launching sentimental analysis otherwise the efficiency and precision of the model greatly affect.

3.5 Flowchart and Algorithm of the Proposed Framework

The complete implementation of the proposed framework can be implemented in various steps as mentioned in fig 1. These steps were mentioned as follows:

1. Gathered Data: The data is a mixture of Twitter data as well as the existing data for attaining the larger data. Due to this, the size of the data reached 25,000 instances.

2. Data pre-processing: Once the data gathered, it needs to be pre-processed to attain uniformity in the form of data that the analyzer looking for. Particularly, natural language processing involving project will have these three steps for the preprocessing aspects as mentioned as follows:

   a. Tokenization: It is the process of dividing the meaningful words, characters, subwords, or some symbols from the existing statements.

   b. Stop Words Removal: The words, for an instance, is, an, a, on, in, and so on, considered as stopwords. These words will not give a meaning separately but these are used to enhance the structure of the statement. So, these are useless for our analysis. Hence, we will remove these stop words.

   c. Stemming and Lemming: Stemming and lemming are almost one another represents similar working but they differ minutely. Stemming as well as lemming is getting the basic word from the existing word, but stemming won't consider the meaning of the word obtained whereas lemming will consider the meaning of the word obtained. So it is always essential that lemming should be implemented along with stemming.

3. Extraction of features: The processed data need to pass into the proposed model to get identify the features and their mapping for the identification of various sentiments depending on the dataset.

4. Identifying the keywords and classifying the movie ratings according to the recognized sentiments.

5. The training dataset and testing dataset need to pass into the proposed model and evaluate the model.

Now, the validation dataset will be passed on to the trained and tested model to attain the prediction or identification of polarized sentiments as positive or negative

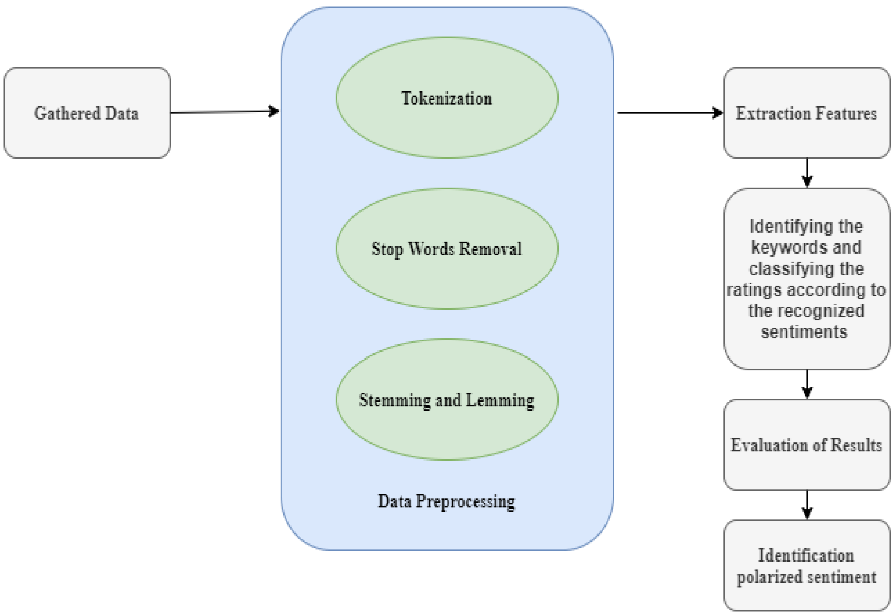

Fig 1 Flowchart representing the proposed framework

4. RESULT ANALYSIS

The proposed framework was implemented and attained the confusion matrix as well as the accuracy, precision, recall, and F1score. The confusion matrix can be represented as mentioned in figure-2. Depending on this confusion matrix, the accuracy, precision, recall, and F1-score can be evaluated and can be compared with the obtained values. The various terms that are used in the confusion matrix can be defined as follows:

1. TP: The full form of TP is true positive which refers to the number of cases that are classified as positive in the actual scenario also represents positive.

2. FN: The full form of FN is false negative which refers to the number of cases that are classified as positive in the actual scenario represents negative.

3. FP: The full form of FP is false positive which refers to the number of cases that are classified as negative in the actual scenario represents positive.

4. TN: The full form of TN is true negative which refers to the number of cases that are classified as negative in the actual scenario also represents negative.

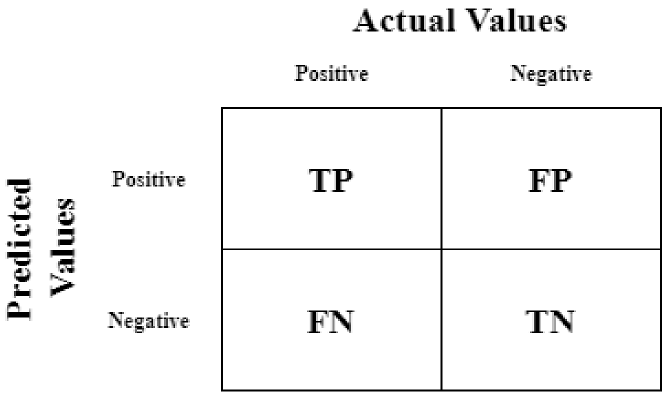

Fig 2. Representation of the confusion matrix

From this confusion matrix, we can evaluate the various metrics as mentioned as follows:

1. $$Accuracy = \frac{TP + TN}{TP+FP+FN+TN} \quad (1)$$

2. $$Precision = \frac{TP}{TP+FP} \quad (2)$$

3. $$Recall = \frac{TP}{TP+FN} \quad (3)$$

4. $$F1-score = \frac{2*Recall*precision}{Recall+Precision} \quad (4)$$

The obtained confusion matrix can be represented as mentioned in figure-3 with consideration of 1000 instances for the validation aspects.

Predicted Values \ Actual Values	Positive	Negative
Positive	TP 499	FP 23
Negative	FN 18	TN 460

Fig 3. Obtained confusion matric from the proposed framework

The obtained evaluation metrics obtained from evaluating the proposed model with 1000 validation instances as follows:

1. Accuracy = 0.959
2. Precision = 0.965
3. Recall = 0.956
4. F1-score = 0.960

From the above metrics, one can conclude that the efficiency of the model is better and in the majority of the cases, we able to identify the accurate sentiment based on movie reviews. From this framework, there are possibilities of about 4.4% for the false negatives and 3.5% for the false positives. If we comprise both the scenarios, there is a possibility of a misclassification rate of 4.1% which is very nominal in the case of sentimental analysis.

5. CONCLUSION

Sentiment analysis paly a vital role across many fields and these provide inherent feelings and emotions of the customers or clients through social networking sites. From the sentimental analysis, one can able to identify the sentiment or emotion as positive, negative. In this framework, we considered only two polarized sentiments like positive and negative. It can be further extended to attain more classes like positive, neutral, and negative or worst, bad, average, good, and very good. In this as0pect multi-class classification of sentiments needs to be identified. Besides, identifying and classifying these sentiments, the proposed framework can further extendable to the recommendation system. The recommendation system provides better options for the customers depending on their interests based on their historical data.

REFERENCES

[1] Alsaeedi, Abdullah, and Mohammad Zubair Khan. "A study on sentiment analysis techniques of Twitter data." International Journal of Advanced Computer Science and Applications 10.2 (2019): 361-374.

[2] Chen, Xin, et al. "A novel feature extraction methodology for sentiment analysis of product reviews." Neural Computing and Applications 31.10 (2019): 6625-6642.

[3] Gui, Lin, et al. "Learning representations from the heterogeneous network for sentiment classification of product reviews." Knowledge-Based Systems 124 (2017): 34-45.

[4] Haque, Tanjim Ul, Nudrat Nawal Saber, and Faisal Muhammad Shah. "Sentiment analysis on large scale Amazon product reviews." 2018 IEEE International Conference on Innovative Research and Development (ICIRD). IEEE, 2018.

[5] Ireland, Robert, and Ang Liu. "Application of data analytics for product design: Sentiment analysis of online product reviews." CIRP Journal of Manufacturing Science and Technology 23 (2018): 128-144.

[6] W. Medhat, A. Hassan, and H. Korashy, “Sentiment analysis algorithms and applications: A survey,” Ain Shams Eng. J., vol. 5, no. 4, pp. 1093–1113, 2014.

[7] Kiran, M. Vamsee Krishna, et al. "User-specific product recommendation and rating system by performing sentiment analysis on product reviews." 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS). IEEE, 2017.

[8] Kouloumpis, Efthymios, Theresa Wilson, and Johanna Moore. "Twitter sentiment analysis: The good the bad and the omg!." Fifth International AAAI Conference on weblogs and social media. 2011.

[9] Liang, Ruxia, and Jian-Qiang Wang. "A linguistic intuitionistic cloud decision support model with sentiment analysis for product selection in E-commerce." International Journal of Fuzzy Systems 21.3 (2019): 963-977.

[10] M. M. Mostafa, "Clustering halal food consumers: A Twitter sentiment analysis," Int. J. Mark. Res., vol. 61, no. 3, pp. 320–337, 2019.

[11] Mubarok, Mohamad Syahrul, Adiwijaya, and Muhammad Dwi Aldhi. "Aspect-based sentiment analysis to review products using Naïve Bayes." AIP Conference Proceedings. Vol. 1867. No. 1. AIP Publishing LLC, 2017.

[12] Ramadhani, A. M. and H. S. Goo, "Twitter sentiment analysis using deep learning methods," Proc. - 2017 7th Int. Annu. Eng. Semin. Ina. 2017, 2017.

[13] Raj, Vasundhara. "Sentiment Analysis on Product Reviews." 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS). IEEE, 2019.

[14] Shivaprasad, T. K., and Jyothi Shetty. "Sentiment analysis of product reviews: a review." 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT). IEEE, 2017.

[15] S. Symeonidis, D. Effrosynidis, and A. Arampatzis, "A comparative evaluation of pre-processing techniques and their interactions for Twitter sentiment analysis," Expert Syst. Appl., vol. 110, pp. 298–310, 2018.

[16] Sagar Pande and Ajay B. Gadicha, " Prevention Mechanism on DDOS Attacks by using Multilevel Filtering of Distributed Firewalls" International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 3 Issue: 3.

[17] Sagar Dhanraj Pande Aditya Khamparia, "A Review on Detection of DDOS Attack Using Machine Learning and Deep Learning Techniques", THINK INDIA JOURNAL, 2035-2043,2019.

[18] Aditya Khamparia, Sagar Pande, Deepak, Gupta and Ashish Khanna and Arun Kumar , " Multi-l evel framework for anomaly detection in social networking".

[19] Ganorkar, S. S., Vishwakarma, S. U., & Pande, S. D. (2014). An information security scheme for cloud based environment using 3DES encryption algorithm. International Journal of Recent Development in Engineering and Technology, 2(4).

[20] Divya K., Sirohi A., Pande S., Malik R. (2021) An IoMT Assisted Heart Disease Diagnostic System Using Machine Learning Techniques. In: Hassanien A.E., Khamparia A., Gupta D., Shankar K., Slowik A. (eds) Cognitive Internet of Medical Things for Smart Healthcare. Studies in Systems, Decision and Control, vol 311. Springer, Cham.

[21] Pande S., Khamparia A., Gupta D., Thanh D.N.H. (2021) DDOS Detection Using Machine Learning Technique. In: Khanna A., Singh A.K., Swaroop A. (eds) Recent Studies on Computational Intelligence. Studies in Computational Intelligence, vol 921. Springer, Singapore.

[22] Yang, Li, et al. "Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning." IEEE Access 8 (2020): 23522-23530.

[23] Zhao Jianqiang and G. Xiaolin, "Deep Convolution Neural Networks for Twitter Sentiment Analysis," vol. 3536, no. c, 2017.

BIOGRAPHIES

Ms. Aanchal Badgaiyya- Btech Student
Dept. CSE, DES's College of Engineering and Technology, Dhamangaon Rly. Maharashtra, India.
Domain: Deep Learning

Ms. Prachi Shankarpale- Btech Student
Dept. CSE, DES's College of Engineering and Technology, Dhamangaon Rly. Maharashtra, India.
Domain: Deep Learning

Ms. Rutuja Wankhade- Btech Student, Dept. CSE, DES's College of Engineering and Technology, Dhamangaon Rly. Maharashtra, India.
Domain: Machine Learning

Ms. Utkarsha Shetye - B.Tech Student
Dept. CSE, DES's College of Engineering and Technology, Dhamangaon Rly. Maharashtra, India.
Domain : Deep Learning

Prof. Kalyani Gholap – Assistant Professor
Dept. CSE, DES's College of Engineering and Technology, Dhamangaon Rly. Maharashtra, India.
Domain : Deep Learning

Prof. Sagar Pande –
Assistant Professor, Dept. CSE, Lovely Professional University, Phagwara, India.
Domain : Deep Learning