iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** III **Month of publication:** March 2023

DOI: https://doi.org/10.22214/ijraset.2023.49535

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue III Mar 2023- Available at www.ijraset.com

Opinion Mining For E-Commerce in Social Media Accounts

Badri Narayanan S[1], Chandra Moulee K V[2], Desvar K J[3], Vimal V R[4]
[1, 2, 3, 4]*Department of Information Technology, Sri Venkateswara College of Engineering - Sriperumbudur, Tamil Nadu, India*

Abstract: ***The usage of the Internet has been growing rapidly, and in the field of natural language processing (NLP), sentiment analysis has become one of the most popular techniques. Using sentiment analysis, the emotional tone behind a body of text can be mined effectively for different occasions. Instagram has become a popular platform for people to buy and sell online. Nonetheless, some studies have found that frauds occurred as a result of buying and selling on the platform, such as when product quantity and specifications differed from what was claimed, a defective product was received, and other frauds involved Instagram sellers. Hence, trust is vital when customers are engaged in S-Commerce activities on Instagram, and there is a need for a trust model to evaluate the trustworthiness of sellers. Nowadays, people provide their feedback after a long time of usage only on social media platforms like Instagram. Therefore, these comments play a major role in helping people decide if the product will be beneficial for them over a long period of time. Mining such content to evaluate people's sentiments can play a critical role in making decisions to keep the situation under control. {objective} Six different classifiers have been used to classify the data. The experiment achieved the highest accuracy of 76.7% with logistic regression. This study can be useful to determine the trustworthiness of the company.***

Index Terms: ***Instagram, E-commerce, Marketing, machine learning, natural language processing (NLP), opinion mining, sentiment analysis, social media.***

I. INTRODUCTION.

In marketing, Instagram is now getting popular because it is very easy to promote the product. Therefore, companies are investing a lot in Instagram to promote their newly developed products before launch. Since the product is easily accessible, people want to know if the product as well as the company are reliable. Nowadays, people provide their feedback after a long time of usage only on social media platforms like Instagram. Therefore, these comments also play a major role in helping people decide if the product will be beneficial for them over a long period of time. Positive "word of mouth" is the key to successful innovation diffusion. Innovation managers pay close attention to customer sentiment. Online reviews are often the most accessible sources of customer feedback. Online review ratings and online review volume are the two most common metrics to interpret customer sentiment.

Sentiment analysis is the process of diagnosing the emotions behind the words. It can generally be depicted as the assignment of opinion categories and scores according to keyword and phrase, which are matched with sentiment score lexicons and customised dictionaries. It is a subsection of text mining that includes the processing of natural language to automate the process of extracting and classifying the emotions in the written text.

This process helps in specifying the writer's attitude toward any entity, topic, product, etc. India is one of the largest markets for digital consumers, with 565 million Internet users. So, the best possible way to get genuine texts written by common people is through social media. Instagram has now grown into a platform where people can easily start small scale businesses and where large-scale industries can expand their marketing strategies because advertising on Instagram is much easier than on other platforms. Because people express their true feelings on social media platforms such as Instagram, it is appropriate to analyse that content in order to effectively judge sentiment scores based on the context of comments.

Also, there is limited knowledge of the general public's sentiment about the main topics that have been discussed over time.

Various approaches can be used to implement sentiment categorization. These approaches can be broadly classified into the three types listed below: 1) a lexicon-based approach; 2) a machine learning or deep learning approach; and 3) a hybrid approach. In this work, a machine learning approach is used to classify customer opinions. To use machine learning for sentiment analysis, data pre-processing on raw data is a prerequisite since the efficiency of the algorithm used is directly proportional to the quality of the training and testing datasets. Section II presents a literature review related to sentiment analysis in social media. The methodology used to execute the experiment and the experimental setup used in this work are discussed in Section III.

©IJRASET: All Rights are Reserved

Results and analysis have been explained in Section IV. Section V contains the most frequent challenges faced and possible limitations of this work. The conclusion of the research and future work are mentioned in Section VI.

II. LITERATURE REVIEW

Sentiment analysis from social media data is one of the highly emerging research fields. It could play a critical role in analysing the trustworthiness of e-commerce companies, and hence it is more crucial. Though a lot of research from various angles on sentiment classification and NLP is still in progress, some of the completed works are as follows.

Brito et al. [1] described the way politicians communicate with the electorate and run electoral campaigns was reshaped by the emergence and popularization of contemporary social media (SM), such as Facebook, Twitter, and Instagram social networks (SNs). Due to the inherent capabilities of SM, such as the large amount of available data accessed in real time, a new research subject has emerged, focusing on using the SM data to predict election outcomes. Despite many studies conducted in the last decade, results are very controversial and many times challenged. In this context, this article aims to investigate and summarize how research on predicting elections based on the SM data has evolved since its beginning, to outline the state of both the art and the practice, and to identify research opportunities within this field. In terms of method, they performed a systematic literature review analysing the quantity and quality of publications, the electoral context of studies, the main approaches to and characteristics of the successful studies, as well as their main strengths and challenges and compared their results with previous reviews. They identified and analysed 83 relevant studies, and the challenges were identified in many areas such as process, sampling, modelling, performance evaluation, and scientific rigor. Main findings include the low success of the most-used approach, namely volume and sentiment analysis on Twitter, and the better results with new approaches, such as regression methods trained with traditional polls. Finally, a vision of future research on integrating advances in process definitions, modelling, and evaluation is also discussed, pointing out, among others, the need for better investigating the application of state-of-the-art machine learning approaches.

Chandra et al. [2] mentioned that social media plays a crucial role in shaping the worldview during election campaigns. Social media has been used as a medium for political campaigns and a tool for organizing protests; some of which have been peaceful, while others have led to riots. Previous research indicates that understanding user behavior, particularly in terms of sentiments expressed during elections can give an indication of the election outcome. Recently, there has been tremendous progress in language modelling with deep learning via long short-term memory (LSTM) models and variants known as bidirectional encoder representations from transformers (BERT). Motivated by these innovations, they developed a framework to model the US general elections. They investigated if sentiment analysis can provide a means to predict election outcomes. They used the LSTM and BERT language models for Twitter sentiment analysis leading to the US 2020 presidential elections. Their results indicate that sentiment analysis can provide a general basis for modelling election outcomes where the BERT model indicates Biden winning the elections.

Es-Sabery et al. [3] used a Fuzzy Deep Learning Classifier (FDLC) approach to design a Sentence level classification. Owing to the rise in the number of social platforms and their extensive use by people, enormous amounts of data are produced hourly. However, sentiment analysis or opinion mining is considered as a useful tool that aims to extract the emotion and attitude from the user-posted data on social media platforms by using different computational methods to linguistic terms and various Natural Language Processing (NLP). Therefore, enhancing text sentiment classification accuracy has become feasible, and an interesting research area for many community researchers. In this study, a new Fuzzy Deep Learning Classifier (FDLC) is suggested for improving the performance of data-sentiment classification. Our proposed FDLC integrates Convolutional Neural Network (CNN) to build an effective automatic process for extracting the features from collected unstructured data and Feedforward Neural Network (FFNN) to compute both positive and negative sentimental scores. Then, Mamdani Fuzzy System (MFS) is used as a fuzzy classifier to classify the outcomes of the two used deep (CNN+FFNN) learning models in three classes, which are: Neutral, Negative, and Positive. Also, to prevent the long execution time taking by our hybrid proposed FDLC, is implemented our proposal under the Hadoop cluster. An experimental comparative study between our FDLC and some other suggestions from the literature is performed to demonstrate our offered classifier's effectiveness. The empirical result proved that our FDLC performs better than other classifiers in terms of true positive rate, true negative rate, false positive rate, false negative rate, error rate, precision, classification rate, kappa statistic, F1-score and time consumption, complexity, convergence, and stability.

Gupta et al. [4] implied emotion in the text can be mined effectively for different occasions. People are using social media to receive and communicate different types of information on a massive scale during COVID-19 outburst. Mining such content to evaluate people's sentiments can play a critical role in making decisions to keep the situation under control.

The objective of this study is to mine the sentiments of Indian citizens regarding the nationwide lockdown enforced by the Indian government to reduce the rate of spreading of Coronavirus. In this work, the sentiment analysis of tweets posted by Indian citizens has been performed using NLP and machine learning classifiers. From April 5, 2020 to April 17, 2020, a total of 12 741 tweets having the keywords "Indialockdown" are extracted. Data have been extracted from Twitter using Tweepy API, annotated using TextBlob and VADER lexicons, and pre-processed using the natural language tool kit provided by the Python. Eight different classifiers have been used to classify the data. The experiment achieved the highest accuracy of 84.4% with LinearSVC classifier and unigrams. This study concludes that the majority of Indian citizens are supporting the decision of the lockdown implemented by the Indian government during corona outburst.

Kim et al. [5] alluded that positive "word of mouth" is the key to successful innovation diffusion. Innovation managers pay close attention to examine customer sentiment. Online reviews are often the most accessible sources of customer feedback. Online review ratings and online review volume, the two most common metrics to interpret customer sentiment from online reviews, have some critical limitations. Online review ratings are prone to extremity bias since extremely satisfied or dissatisfied customers are most likely to leave online reviews. Online review volume can increase for any reason, and it does not necessarily indicate positive or negative customer feedback. This article explores text mining methods and proposes some alternative metrics to interpret customer sentiment. The findings show that sentiment scores might be less prone to extremity bias compared to online review ratings. Sentiment scores tended to fit a normal distribution while online review ratings were skewed to extreme values. Sentiment scores and review lengths, when combined, can provide a new angle to observe enthusiasm.

On et al. [6] defined the sentiment-aware web crawling problem and then propose two hash-based methods for the implementation. One is based on hash join and the other is bucket-sorted hash join. In particular, it is proposed as a novel bucket-sorted hash join for the efficient sentiment-aware web crawling method. Our experimental results show that the proposed web crawling method using bucket-sorted hash join outperforms existing web crawling methods by significantly reducing the running time and storage space. In the proposed method, the time taken to execute the sentiment-aware task per web page is 0.016 seconds and the database space can be saved by 59% compared to the existing web crawling methods.

Rodriguez et al. [7] focused on the analysis of sentiment dynamics and their characterization from statistical and mathematical perspectives. Here, a set of basic methods are applied to analyze the statistical and temporal dynamics of sentiment analysis on political campaigns and assess their scope and limitations. To this end thousands of Twitter messages mentioning political parties and their leaders posted several weeks before and after the 2019 Spanish presidential election. A twofold analysis strategy: statistical characterization using indices derived from well-known temporal and information metrics and methods –including entropy, mutual information, and the Compounded Aggregated Positivity Index– allowing the estimation of changes in the density function of sentiment data; and feature extraction from nonlinear intrinsic patterns in terms of manifold learning using autoencoders and stochastic embeddings. The results show that both the indices and the manifold features provide an informative characterization of the sentiment dynamics throughout the election period. Measurable variations in sentiment behaviour and polarity across the political parties and their leaders and observed different dynamics depending on the parties' positions on the political spectrum, their presence at the regional or national levels, and their nationalist or globalist aspirations.

Saad et al. [8] manifested by earlier studies and concluded sentiment analysis of drug reviews has a large potential for providing valuable insights to assist healthcare professionals and companies for evaluating the safety of drugs after it has been marketed. Such insights help safeguard patients and increase their trust in medical companies. The existing systems either follow a lexicon-based approach or a learning-based approach for sentiment analysis in the medical domain. Learning-based techniques require annotated data while lexicon-based techniques tend to be domain-specific which restricts their wide use. This research embarks on a hybrid technique that utilizes both learning-based and lexicon-based approaches to achieve better results. General-purpose sentiment lexicons, such as AFFIN, TextBlob, and VADER, are used for annotating the reviews.

Silva et al. [9] have demonstrated a study of the topicregarding the mention of controversial and non-controversial words related to COVID-19 on analyse the sentiments shown by Brazilian Twitter users about SUS before and during COVID-19 pandemic. To reach this goal, a database of Portuguese tweets regarding SUS posted between December 2019 and October 2020 was created. The tweets were pre-processed, classified and then analysed. The results show that, in most cases, users are in favour of SUS.

Wang et al. [10] considering the emoticon symbols and punctuation symbols in web social media text. Similar to language symbols, emoticons' symbols and punctuation symbols in web social media text also contain certain sentiment information. In order to make full use of sentiment information contained in web social media to solve the problem of text context semantics missing, they propose a sentiment classification method of web social media based on multidimensional and multilevel modelling.

By modelling web social media text from three dimensions (language symbols, emoticons' symbols, and punctuation symbols) and three levels (words, sentences, and documents) based on a deep learning framework, in this article, they attempt to solve text context semantics missing faced by the sentiment classification of web social media and improve the accuracy of sentiment classification of web social media. The experimental results on Sina Weibo and Twitter datasets show that the average accuracy of their method is 0.9479, which achieves more than 5.86% performance compared with the existing sentiment classification methods.

From the above discussion, it is clear that sentiment analysis is one of the prominent decision-making methods. In such a marketing age sentiment analysis may play a critical role in determining the validity of a product and transparency of the company. Very few studies have been reported on sentiment analysis of companies in social media. Therefore, the main objective of this research is to get the sentiment of the companies by performing deep analysis of the comments and providing a conclusion to influence the decision of the customers.

The key objectives of this article are as follows.

1) To mine the sentiments of customers regarding the company from the comment sections of the posts.

2) The sentiment analysis of comments posted by customers has been performed using NLP and machine learning classifiers.

3) To classify the data accurately, six different classifiersare used.

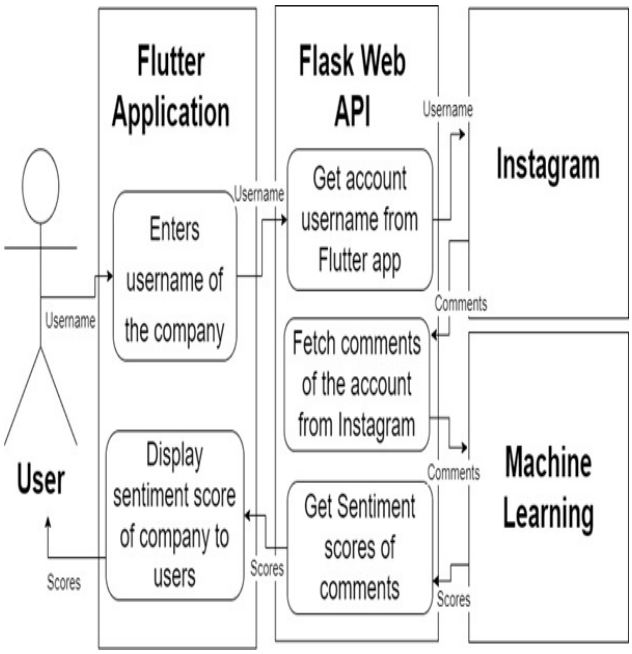


Fig. 1. Methodology used.

III. PROPOSED APPROACH

In this section, a framework is proposed for sentiment analysis as shown in Fig. 1. The Proposed work comprises of building an application that will provide a solution for finding credibility of Instagram businesses. This application uses a Machine Learning NLP approach to analyze comments and provide positive, negative scores as the output. The various phases of the framework to perform the sentiment analysis of Instagram comments are as follows.

A. Data Extraction

Instagram is an American photo and video sharing social networking service founded in 2010 by Kevin Systrom and Mike Krieger, and later acquired by Facebook Inc. The app allows users to upload media that can be edited with filters and organized by hashtags and geographical tagging. Posts can be shared publicly or with preapproved followers. Users can browse other users' content by tag and location, view trending content, like photos, and follow other users to add their content to a personal feed.

Instagram is targeted by most of the businesses to promote their products. They post content (posts) regularly in the platform, talking about their products, while the users and audience comment under the comments sections about their views and user experience of the product. But it will be so hard for the user to go through each and every comment to know if the product is good or if the company is trustable.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue III Mar 2023- Available at www.ijraset.com

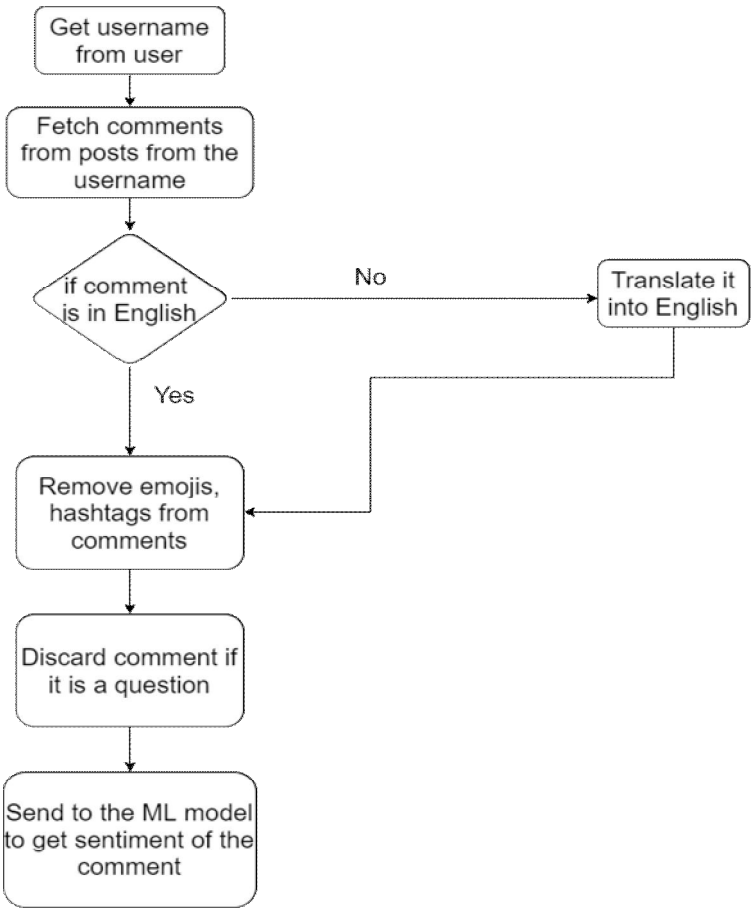


Fig. 2. Data extraction process

This module will be focusing on how to fetch the user comments from Instagram, based on the user entered account name. Using the comments fetched the analysis is performed. Fig. 2 depicts a flow diagram on how the comments are fetched from Instagram. It is also made sure that whether the username exists or not. Then it is verified whether there are any comments or not. Only if there are comments it is possible to calculate the sentiments. Then emojis and hashtags are removed from comments. Question comments should also be discarded.

B. Computing Sentiment Scores

The Natural Language Toolkit, or more commonly NLTK, is a suite of libraries and programs for symbolic and statistical natural language processing (NLP) for English written in the Python programming language. It was developed by Steven Bird and Edward Loper in the Department of Computer and Information Science at the University of Pennsylvania. NLTK includes graphical demonstrations and sample data. It is accompanied by a book that explains the underlying concepts behind the language processing tasks supported by the toolkit, plus a cookbook.

NLTK is intended to support research and teaching in NLP or closely related areas, including empirical linguistics, cognitive science, artificial intelligence, information retrieval, and machine learning. NLTK has been used successfully as a teaching tool, as an individual study tool, and as a platform for prototyping and building research systems.

A technique is needed to identify whether the given comment from the user is positive or not. Only then the end user the info can provide whether the company is trustable or not. For that, Machine Learning with NLP approach can be used.

This module explains about how to use an NLP based model to fetch the sentiment data of a comment. A dataset is prepared which has product data along with the sentiment. Fig. 3 depicts how the feature set is derived from the dataset of reviews. Various classifiers are used to create an ensemble model, which can provide the sentiment as the output. Fig. 4 depicts how an ensemble model is created with the help of the feature set.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

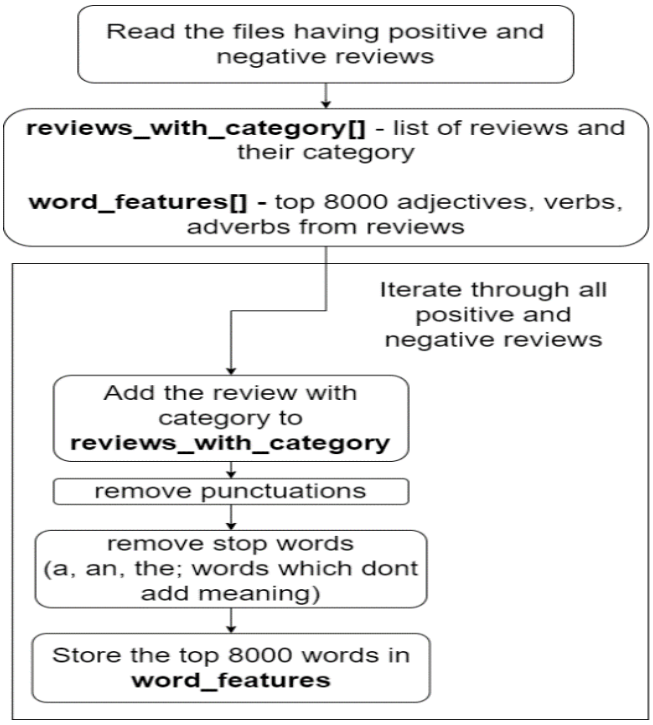


Fig. 3. Feature set creation.

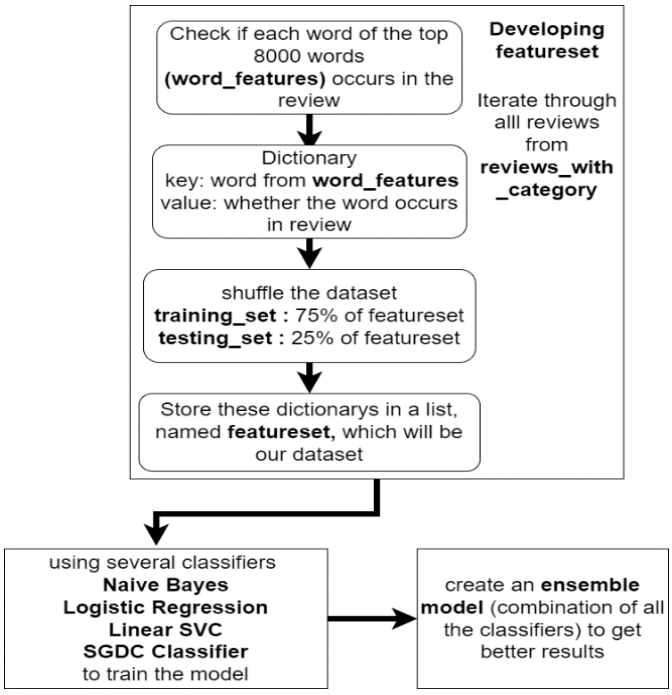


Fig. 4. Machine Learning Model Creation

C. Enabling WEB services

Flask is a small and lightweight Python web framework that provides useful tools and features that make creating web applications in Python easier. It gives developers flexibility and is a more accessible framework for new developers since you can build a web application quickly using only a single Python file. Flask is also extensible and doesn't force a particular directory structure or require complicated boilerplate code before getting started.

Flask is a micro web framework written in python. It is being used it to create an API. The flask API will process the requests received from users. For input, it receives a POST request which includes the username of the Instagram business as a part of a JSON. Once the input is received, it establishes a connection with Instagram, and fetches the comments of the username.
The comments are then sent to the classifier which predicts the sentiment of the comment. Based on the sentiments, percentage values of positive and negative comments are calculated. Then, the computed sentiment scores are sent as a JSON response. In case the username doesn't exist, or there are no comments for the user, or there is some other error, it returns an error response telling Failure occurred.

D. Application Creation

Flutter is Google's portable UI toolkit for crafting beautiful, natively compiled applications for mobile, web, and desktop from a single codebase. Flutter works with existing code, is used by developers and organizations around the world, and is free and open source. Flutter is designed to support mobile apps that run on both Android and iOS, as well as interactive apps that you want to run on your web pages or on the desktop. Since the technique involves Machine Learning and various complex techniques, which can't be easily done by a user who doesn't have any computer expertise, a Front-end application is developed which will receive a company's Instagram username as input. Then the app will communicate with the backend through an API to receive the sentiment scores and display them as output in the app. If the entered username is invalid it will display Invalid username entered.

IV. RESULTS AND ANALYSIS

After observing the accuracies, precision, scores of six different classifiers we have got the best results with logistic regression with an accuracy of over 76.7% as depicted in the Table 1. However, to improve accuracy even further, we prepared an ensemble model which will take into all the models into consideration and based on votes, gives the prediction. So, we have predicted the sentiment analysis of multiple accounts using this combination and we have found that, for company Oneplus, 65% of people are talking positive about the company one plus and 35% of the people are feeling negative due to some reason as depicted in Fig. 5.
The reviews for the company Oppo are very welcoming from the people's perspective because of their good quality and budget phones and the result was according to our expectations as 95% of the people provided positive feedbacks about the oppo company while only 5% gave negative feedback as depicted in Fig.6

TABLE I
CLASSIFIERS WITH ACCURACY

Classifier	Accuracy
Naïve Bayes	67.9
Multinomial Naïve Bayes	74.6
Bernoulli Naïve Bayes	67.7
Logistic Regression	76.7
Linear Support Vector (SVC)	75.4
Stochastic Gradient Descent (SGDC)	76
Overall (Ensemble Model)	78

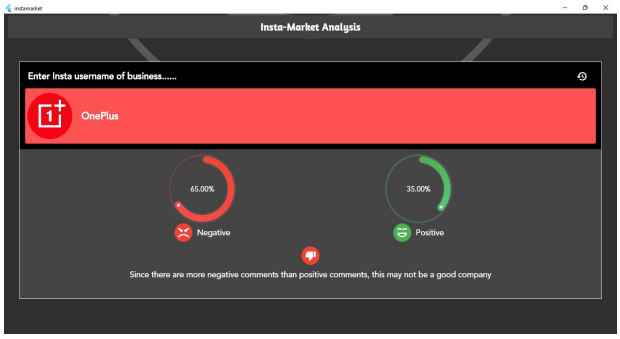


Fig.5 Sentiment analysis for the company OnePlus

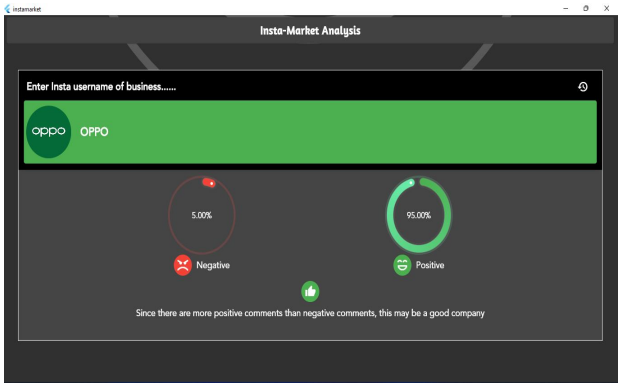


Fig.6 Sentiment analysis for the company Oppo

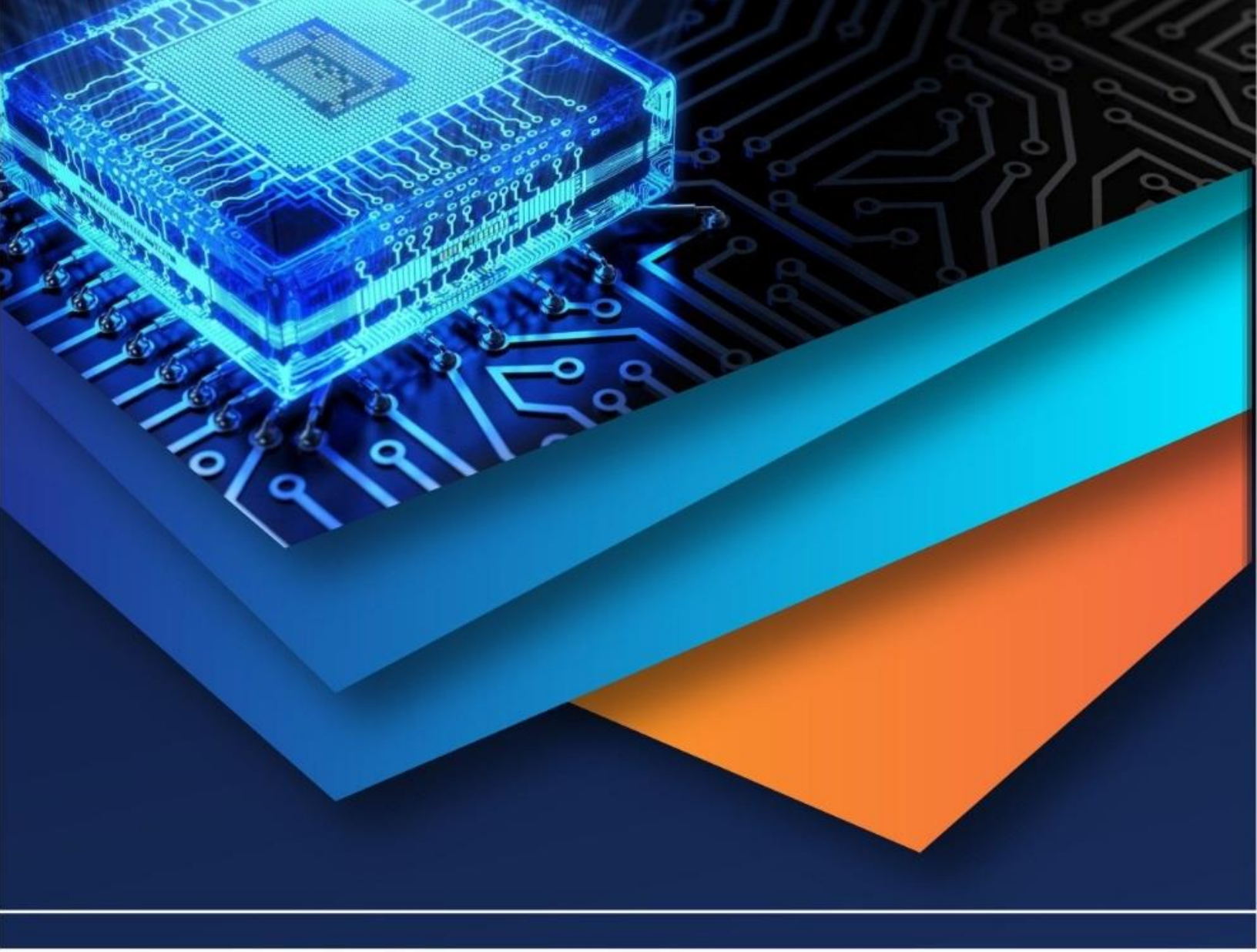
V. CONCLUSION AND FUTURE WORK

Social media is witnessing a massive increase in the number of users per day. People prefer to share their honest opinions on social media instead of sharing with someone in person. Using the comments from Instagram, common public's aggregate reaction toward the companies and the products they launch is examined. Six supervised machine learning techniques are used with different grams of text after annotation and pre-processing. Best performance is observed with the Logistic Regression classifier. The ensemble model classifier gives us an accuracy of seventy-seven percentage, which is best in all the classifier which are executed on the data set. It is followed by Stochastic Gradient Descent Classifier which gives an accuracy of seventy-six percentage. An ensemble model is further created which is based on all the six models used and gives the output based on vote received from the six classifiers, which has an accuracy of seventy-eight percentage. The concluded result provided by the application will greatly influence the buying decision of the customers. It will prevent the customers to lose money or succumb to fraud in social media.

Sentiment analysis of natural languages itself contained a vast scope to work on, and due to marketing scope, this work is also manifested with a wide range of future scopes. Future studies can consider spam comments and how they can be removed from the data which will be reviewed. Also, sentiments can be tracked with the help of the emojis by preparing a separate data set for it. Comparison between multiple companies at the same time can also be done. The overall speed in which the comments are fetched from the social media can be improved. Instead of analysing companies as a whole, the individual products from the company can be analysed and comparison can be made to conclude which is the best product of the company. From the technical point of view, future studies can look to improve the accuracy of the model and can experiment on a large corpus.

REFERENCES

[1] Brito K. D. S., Filho R. L. C. S. and Adeodato P. J. L. (2021) "A Systematic Review of Predicting Elections Based on Social Media Data: Research Challenges and Future Directions," in IEEE Transactions on Computational Social Systems, Aug., vol. 8, no. 4, pp. 819-843.
[2] Chandra R. and Saini R. (2021) "Biden vs Trump: Modeling US General Elections Using BERT Language Model," in IEEE Access, vol. 9, pp. 128494-128505.
[3] Es-Sabery F., Hair A., Qadir J., Sainz-De-Abajo B., García-Zapirain B. and Torre-Díez I. D. L. (2021) "Sentence-Level Classification Using Parallel Fuzzy Deep Learning Classifier," in IEEE Access, vol. 9, pp. 17943-17985.
[4] Gupta P., Kumar S., Suman R. R. and Kumar V. (2021) "Sentiment Analysis of Lockdown in India During COVID-19: A Case Study on Twitter," in IEEE Transactions on Computational Social Systems, Aug., vol. 8, no. 4, pp. 992-1002.
[5] Kim R.Y. (2021) "Using Online Reviews for Customer Sentiment Analysis," in IEEE Engineering Management Review, Dec., vol. 49, no. 4, pp. 162-168, 1 Fourth quarter.
[6] On B. W., Jo J. Y., Shin H., Gim J., Choi G. S. and Jung S. M. (2021) "Efficient Sentiment-Aware Web Crawling Methods for Constructing Sentiment Dictionary," in IEEE Access, vol. 9, pp. 161208-161223.
[7] Rodríguez-Ibáñez M., Gimeno-Blanes F. J., Cuenca-Jiménez P. M., Soguero-Ruiz C. and Rojo-Álvarez J. L. (2021) "Sentiment Analysis of Political Tweets From the 2019 Spanish Elections," in IEEE Access, vol. 9, pp. 101847-101862.
[8] Saad E. and Din S. (2021) "Determining the Efficiency of Drugs Under Special Conditions from Users' Reviews on Healthcare Web Forums," in IEEE Access, vol. 9, pp. 85721-85737.
[9] Silva H., Andrade E., Araújo D. and Dantas J. (2022) "Sentiment Analysis of Tweets Related to SUS Before and During COVID-19 pandemic," in IEEE Latin America Transactions, Jan., vol. 20, no. 1, pp. 6-13.
[10]Wang B., Shan D., Fan A., Liu L. and Gao J. (2022) "A Sentiment Classification Method of Web Social Media Based on Multidimensional and Multilevel Modeling," in IEEE Transactions on Industrial Informatics, Feb., vol. 18, no. 2, pp. 1240-1249.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)