iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** V **Month of publication:** May 2021

DOI: https://doi.org/10.22214/ijraset.2021.34737

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: 7.429
Volume 9 Issue V May 2021- Available at www.ijraset.com

Sentiment Analysis on Restaurant Review using Hybrid Approach

Komal Yadav[1], Dr. Miral Patel[2]
[1, 2]*Department of Information Technology, G. H. Patel College of engineering and technology, Gujarat, India*

Abstract: ***Sentiment analysis describes the overall sentiment behind the document or Sentence. SA is a field of NLP to analyze insights from social media mostly from Twitter, Amazon, Zomato, Trip Advisor, etc. Sentiment analysis of document or sentence generally categories as Positive, Negative, or Neutral. There are some drawbacks in machine learning and dictionary-based sentiment analysis, to overcome these issues hybrid approach works. In this research work, we discover a hybrid model for classification which is a combination of both machine learning and dictionary-based approach to enhance overall performance and accuracy and reducing the time and human effort required to manually labelling. In this paper, we performed labelling of the dataset according to their sentiment score using VADER lexicon-based technique and also evaluated multiclass classification using machine learning on restaurant reviews. This prediction can help to improve business service. This article also presents the study of several sentiment analysis classifier, tools and techniques and to perform sentiment analysis and provide future scope.***

Keywords: Sentiment analysis, Machine learning, Hybrid approach, VEDAR, Natural language processing (NLP).

I. INTRODUCTION

Sentiment analysis is part of NLP which evaluates human-written texts. Sentiment analysis is used for identifying and analyzing the opinion of areas such as political reviews, spam email filtering, movie reviews, stock market prediction, blog posts, recommendation systems, comments, reviews of products, tweets, advertisements, etc. The success of a business depends on its customer's likes and dislikes so, sentiment analysis is used to improve business. The result of sentiment analysis using machine learning depends on the chosen algorithm and libraries and the size or complexity of the dataset. Sometimes positive text also analyzed as negative or neutral and to classify reviews into multiple sentiment categories such as happy, sad, angry, Thrilled, depressed, frustrated, satisfied, confused, etc are challenging tasks. This research aims to study several tools and techniques to perform sentiment analysis. The processed restaurant review dataset contains comments, Likes, positive scores, negative scores, neutral scores, and compound scores. This research paper focuses on analyzing comments using a hybrid approach. VADER Rule-based lexicon and multi-class classification using machine learning approach applied to restaurant reviews for building a model that makes predictions about the probabilities that review belongs to which particular class from the three classes.

Sentiment analysis classified into three categories based on the technique used to perform the classification.

A. Lexicon Based Method
B. Machine learning Based Method
C. Hybrid Method

1) *Machine Learning (ML) Based Approach:* ML is an automatic classification technique that performs classification by applying training and testing to make predictions.ML based classification technique categorized into supervised, unsupervised and semi-supervised approach. In a supervised learning-based approach applied dataset should be labeled, and the unsupervised approach takes unlabeled data as input. Quality of Sentiment analysis using ML depends on chosen algorithm and methods such as countvectorizer, vectorization, data preprocessing, tokenization, IF-IDF model, a bag of words model, etc are further discussed in this paper.
2) *Lexicon Based Approach:* The lexicon-based model involves a rules-based and dictionary-based approach which is a collection of words such as WordNet, SentiWordNet, etc, and a corpus-based approach. This method can be evaluated by calculating the score of text with emoticons. In this approach, compound scores can calculate by an average of positive, negative, and neutral scores of the particular review.
3) *Hybrid Approach:* The hybrid model for sentiment analysis includes a rule-based lexicon and machine learning approach. A rule-based approach is used to calculate a score according to a sentiment of a particular word or emoticon and based on that score we can label the reviews as positive, negative, or neutral.

II. LITERATURE REVIEW

Table 1: Literature Summary

Sr .	Author Name	Publication	Ref. No.	Features
1	Annet John, AniceJohn,Prof. Reshma Sheik	2019 IEEE	[1]	sentiment classification is performed using: (1)SentiWordNet (SWN)(2) hybrid lexicon (HYBRID) and (3)hybrid lexicon followed by Sentiment Adjustment Factors (H+S ADF).
2	1Siva Kumar Pathuri 2Dr.N.Anbazhagan 3Dr.G.Balaji Prakash	2020 IEEE	[3]	SVM, LR and with combined Bag-Boosting approach is used. - Achieved 94% accuracy.
3	Thanh Le	2019 IEEE	[2]	Liu and SentiWordNetlexicons is applied for sentiment analysis of 1) Amazon dataset 2) IMDb dataset 3) Yelp dataset.
4	F.M Takbir Hossain Md. Ismail HossainMs. SamiaNawshin	2017 IEEE	[15]	-Used Natural language Toolkit (NLTK) for data preprocessing. -used Sci-kit learn library for building classification model. -Multinomial Naïve Bayes (MNB), Support Vector Machine (SVM) K nearest neighbor (KNN) and Linguistic Regression (LR) algorithms applied.
5	Rachana Bandana	2018 IEEE	[30]	Python 3.4 and Natural Language Processing Toolkit (NLTK) are used.
6	Rahul Vasundhara Raj Monika	2019 IEEE	[27]	A literature survey shows that Support Vector Machine and Naïve Bayes technique has provided good and adept results on their respective datasets.
7	NishaYadav, Rajeev Kumar, Dr.BhupeshGourDr.AsifUllah Khan	2019 IEEE	[8]	The hybrid approach is applied using Arcing Classifier and the NB(Naïve Bayes), SVM, GA(Genetic algorithm) models are built as base classifiers to predict classification scores.
8	K. Kaviya, C. Roshini, V. Vaidhehi and J. DhaliaSweetlin	2017 IEEE	[18]	Lexalytic tool is applied which consists of a large dictionary of phrases that has sentiment values in different languages. -It has a sentiment score for each phrase for the analysis.
9	Suvarna G Kanakaraddi Ashok K ChikaraddiKaruna C. Gull	2020 IEEE	[6]	Machine learning algorithms: Support Vector Machine, Navie Bayes, Max Entropy, LSTM, CNN, Random forest are involved.
10	Hemalatha, Ramathmika	2019 IEEE	[4]	Classification of Yelp reviews is performed using supervised learning:1. Naive Bayes 2. Multinomial Naive Bayes 3. Bernoulli Naive Bayes 4. Logistic Regression 5. Linear SVC (Support Vector Clustering).
11	Meng Zhang	2020 IEEE	[20]	Sentiment classification model is applied based on sentiment lexicon and CNN algorithm.
12	DimosthenisBeleveslis, Christos Tjortjis , DimitrisPsaradelis, DimitrisNikoglou	2019 IEEE	[7]	-A Probabilistic classification method is applied (Random Forest,Decision Tree and XGBoost), supervised learning methods, a hashtag-based filtering is applied to each tweet in Greek regarding the recent European Elections.
13	Bhavitha B K Anisha P Rodrigues Dr.Niranjan N Chiplunkar	2017 IEEE	[25]	Supervised Machine learning techniques are applied.
14	R.AkilaS.Revathi	2020 IEEE	[21]	Naïve Bayes algorithms applied on McDonald's data set.
15	TarunAnand Vikrant Singh Bharat Bali Biswa Mohan SahooBasuDevShivhare Amar Deep Gupta	2020 IEEE	[22]	Tweepy ,Matplotlib tools are used which Portrays advancement of technology.
16	Aishwarya, ParthWadhwa, Prabhishek Singh	2020 IEEE	[5]	-Positive and negative analysis of reviews. -Data fetched from twitter through API keys and twitter access

				tokens.
17	RachmawanAdiLaksono,KellyRossaSungkono,RiyanartoSarno,CahyaningtyasSekarWahyuni	2019 IEEE	[28]	-Naive Bayes method on WEKA and a comparison with the TextBlob sentiment analysis is performed. -Naive Bayes method has 72.06% accuracy.
18	Dimple TiwariBhartinagpal	2020 IEEE	[29]	-Summary of Articles provides contributions for SA ensemble techniques. -Still a lack of resources for other languages rather than English in the field of SA.
19	Rathan M Vishwanath R HulipalledMurugeshwari P Sushmitha H.M	2017 IEEE	[10]	-Survey on Sentiment Analysis.
20	I. K. C. U. Perera H.A. Caldera	2017 IEEE	[14]	-"Dependency Parser" and "SentiWordNet" are applied.
21	Chaithra V D	2019 IJECE	[36]	Naive bayes and sentiment VADER for analyzing sentiment of mobile unboxing video comments

III.METHODOLOGY

This section will go over the steps to perform sentiment analysis for the dataset, and issues with the dataset in addition to, data cleaning, text preprocessing and feature extraction using methods such as countvectorizer and Tfidf. This section mainly focuses on tools and techniques for sentiment analysis and performing several multiclass classifiers and rule based lexicon approach.

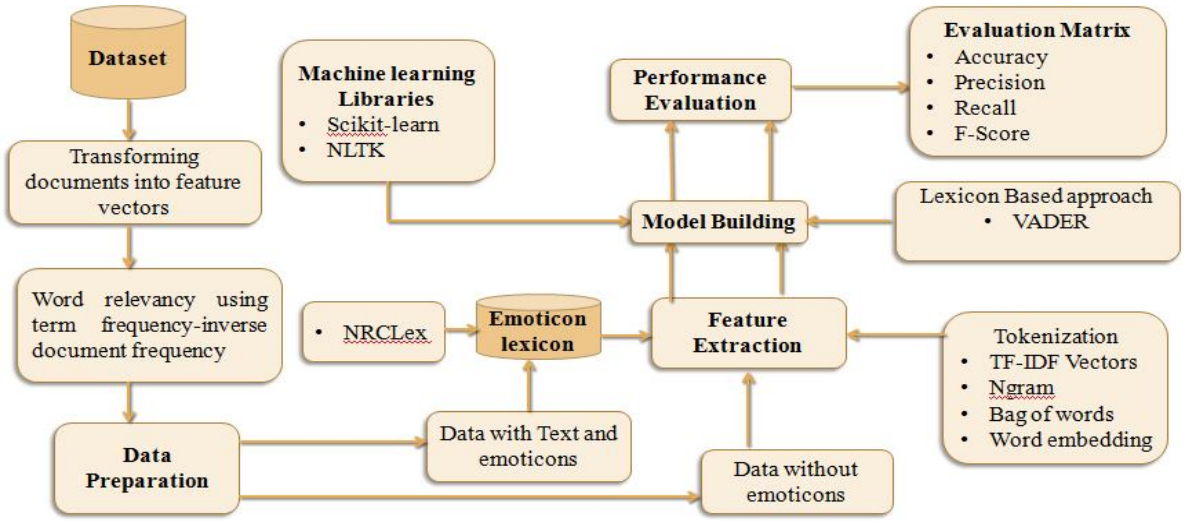


Fig. 1: Proposed work flow diagram

A. Data Preparation

Restaurant Review is a dataset from Kaggle. The Dataset consists of total 1000 reviews.

	Review
0	Wow... Loved this place.
1	Crust is not good.
2	Not tasty and the texture was just nasty.
3	Stopped by during the late May bank holiday of...
4	The selection on the menu was great and so wer...

Fig. 2: Comments

B. Data Labeling using VADER

VADER (Valence Aware Dictionary and sEntiment Reasoner is the rule-based lexicon dictionary that uses a sentiment intensity score for a particular word. VADER is generally used for social media texts. As shown in fig. 2 the dataset was labeled as scores according to the polarity of reviews and then an average score is calculated to get a compound score. In the next step based on compound score polarity of review is decided.

	Review	Liked	scores	compound
0	Wow... Loved this place.	1	{'neg': 0.0, 'neu': 0.435, 'pos': 0.565, 'comp...	0.5994
1	Crust is not good.	0	{'neg': 0.445, 'neu': 0.555, 'pos': 0.0, 'comp...	-0.3412
2	Not tasty and the texture was just nasty.	0	{'neg': 0.34, 'neu': 0.66, 'pos': 0.0, 'compou...	-0.5574
3	Stopped by during the late May bank holiday of...	1	{'neg': 0.093, 'neu': 0.585, 'pos': 0.322, 'co...	0.6908
4	The selection on the menu was great and so wer...	1	{'neg': 0.0, 'neu': 0.728, 'pos': 0.272, 'comp...	0.6249

Fig. 3: Scoring of reviews

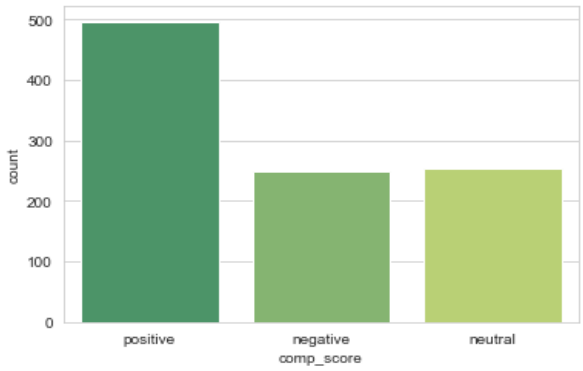


Fig. 4. Classification of reviews based on polarity

C. Preprocessing Data

Preprocessing step includes removal of stop words and also remove URLs, Letters, Special Symbols, prefixes, suffixes, plurals, tenses such that only root word remains to get the real sense. This step also includes Lemmatization and Stemming.

1) *Stemming:* It is a procedure of normalizing text into its base form. Porter Stemmer algorithm is applied for stemming.
2) *Lemmatization:* Lemmatization is a procedure of mapping different forms of words into their root word.
3) *Tokenization:* Creating a matrix for the classification model and this process is known as tokenization.

D. Bag of the Word Model

We cannot use the text data directly into machine learning algorithms. To apply ML algorithms, we must have to transform the text to numeric feature vectors. TF-IDF and Countvectorizer applied for this. Since we are about to use the bag of words model, which creates a sparse matrix for words. We can count the frequency of a word that seems inside the dataset with Countvectorizer. TF-IDF uses a statistical method to do this.

E. TF-IDF Model

Term Frequency (TF) is the calculation of how many times the same word appeared in the document. Sometimes if the size of a document is large some words may appear frequently.

$$TF = \frac{\text{Number of time term appear in the document}}{\text{Total Number of words in the document}}$$

Inverse Document Frequency (IDF) measures the importance of the term. Certain words such as “food”, “and”, “this”, etc may appear many times but have less importance. Thus TDF is used to weigh down the terms which are frequently appeared.

Scikit-learn’s TF-IDF Transformer library is used to transform a bag of words into a sparse vector of TF-IDF values.

The equations for the IDF that are implemented in scikit-learn are:

$$\text{idf}(t, d) = \log\frac{n_d}{1 + \text{df}(d, t)},$$

Where *nd* is a total number of words that appeared in the documents, and df(d, t) is the total number of document d which contain the token t.

The TF-IDF equation implemented by scikit-learn:

$$\text{tf-idf}(t, d) = \text{tf}(t, d) \times (\text{idf}(t, d) + 1)$$

F. Sentiment Classification Using Multiclass Classifiers

We employed cleaning and preprocessing of text data, performed feature extraction using methods such as count vectorizer and TF-IDF (Term frequencies and inverse document frequencies).

We performed Multiclass classification using:

1) Logistic Regression
2) SVC
3) LinearSVC
4) MultinomialNB
5) XGBoost

IV. RESULT AND ANALYSIS

Table- 2: Results Based on Classification Model

Classification Algorithm	Results				
	Confusion Matrix	Accuracy (in %)	Precision	Recall	F1 Score
Logistic Regression	[[19 12 14] [8 32 14] [10 14 77]] 0.64	64%	0.65	0.64	0.64
SVC	[[15 11 19] [4 26 24] [5 9 87]] 0.64	64%	0.67	0.67	0.66
Linear SVC	[[20 14 11] [8 36 10] [9 18 74]] 0.65	65%	0.66	0.65	0.65
MultinomialNB	[[20 9 16] [6 26 22] [11 8 82]] 0.64	64%	0.67	0.64	0.65
XGBoost	[[18 19 8] [6 43 5] [9 18 74]] 0.675	67.5%	0.70	0.68	0.68

1) *Explanation:* The Restaurant review dataset is split into two parts in which 20% of data is used in the testing set and 80% data is applied in the training part of the model. After training the machine learning classification model they predict the sentiment of a particular comment. A classification result also depends upon Training and Testing split values. LinearSVC, XG Boost ,and Logistic Regression perform better than the other classifiers, with XG Boost having a slight advantage with an accuracy of 67.5%.

a) *Confusion-Matrix:* Confusion Matrix is the used to measure performance of a classification model where output can be two or more classes. Tabular method for visualize the performance of test data.

b) *Accuracy:* The fraction of the total samples that were correctly predicted by the classification model. Formula used to calculate Accuracy is: *(TP+TN)/(TP+TN+FP+FN).*

c) *Precision:* Out of all prediction it shows correctly classified true positive predictions Formula used to calculate precision is: *TP/(TP+FP).*

d) Recall: It shows fraction of total samples and correctly predicted samples. Formula used to calculate recall is: *TP/(TP+FN).*

e) F1-score: F1 score gives overall performance measure for each class of testing dataset.

It is the harmonic mean of precision and recall. Formula used to calculate F1 score is:

F1 = 2 * (precision*recall) / (precision+recall)

2) NRC Lexicon: It is a dictionary that incorporates human-provided phrases with their sentiment tags. NRCLex computes emotion from data. It has approximately 27,000 words and it belongs to National Research Council Canada (NRC) which affects the lexicon and the NLTK's WordNet synonym [37]. We performed NRCLex to our dataset which gives the result as shown below:

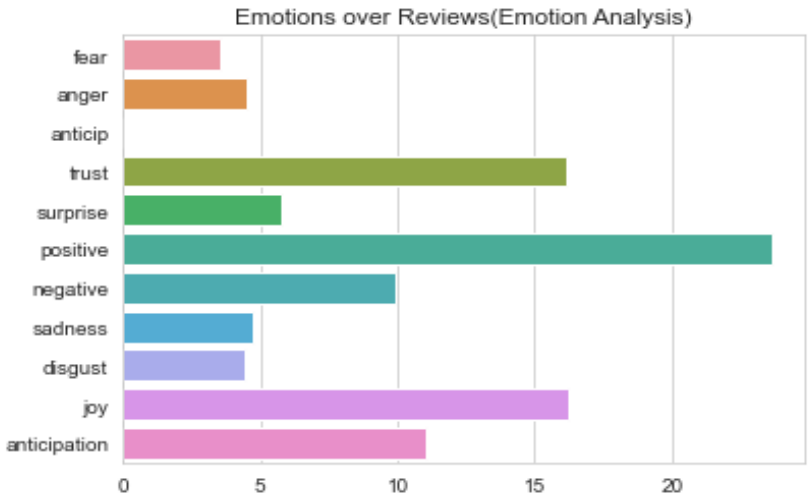


Fig. 5: Emotion analysis

V. OVERVIEW OF CLASSIFICATION MODELS

A. *Logistic Regression*

Logistic Regression is suitable for binary classification problems with a large dataset. This algorithm works based on the mathematical sigmoid function to find out the relation between a dependent variable and an independent variable by approximating probability. The independent variables are features that are used to guess the target class. The dependent variable is the target class to be predicted.

Logistic Regression classification is given by

$$log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m$$

Here p indicates the occurrence probability, βirepresents regression coefficients and xi indicates explanatory variables.

B. *SVC*

SVC is SVM based classifier using libsvm. SVC multi-class mode is implemented using one vs one scheme while LinearSVC uses one vs the rest.

It is possible to implement one vs the rest with SVC by using the OneVsRestClassifier wrapper. SVC uses C regulation parameter for optimizing margin of hyperplane [39].

C. *Extreme Gradient Boosting (XGBoost)*

XGBoost is an implementation of Gradient Boosting which uses more accurate predictions to find the best model.

XGBoost works faster and efficiently than Gradient Boosting because it contains tree learning and linear model solver and it provides parallel computing on a single machine.

D. LinearSVC

Works based on one-vs-the-rest approach. It supports multiclass classification of sparse datasets and the kernel is linear for classification.

LinearSVC is more flexible due to liblinear. Provides additional functionality such as penalty L1 and L2 normalization and loss function[39].

E. MultinomialNB

MuntinomialNB is an implementation of the naive Bayes algorithm for the dataset that is discrete. Classification of text performs using the two methods in which data represented as count vector or tf-idf vectors.

The distribution is parameterized by vectors θy=(θy1,…,θyn) for each class y,

where n = Total features and θyi = probability P(xi|y) of feature i appearing in a document which belongs to the class y [38].

The parameters θy is calculated by a smoothed version of maximum probability:

$$\hat{\theta}_{yi} = \frac{N_{yi} + \alpha}{N_y + \alpha n}$$

Where $N_{yi} = \sum_{x \in T} x_i$ is the occurrence of feature i in the set of training T,

And $N_y = \sum_{i=1}^{n} N_{yi}$ are the total features for class y [38].

VI. CONCLUSIONS AND FUTURE SCOPE

In our research work, we proposed a hybrid approach for performing sentiment analysis of restaurant reviews. VADER (Rule-based lexicon) is applied to the labeled dataset. We performed five Machine learning-based classification models. The sentiment classification was performed via jupyter notebook platform and using Python's Scikit-Learn package and NLTK. We also performed emotion analysis on reviews using NRC lexicon. This paper also shows literature reviews of models which are used to perform sentiment analysis. The accuracy and performance of classification model is depends on size of dataset we taken to train the model.

In the future, more work is needed on a real-time large dataset to perform further improving measures and sometimes positive text also analyzed as negative or neutral and to classify reviews into multiple sentiment categories such as happy, sad, angry, Thrilled, depressed, frustrated, satisfied, confused, etc. This paper could further studied for other factors such as the use of other languages, dealing with negation expressions, including emojis, Sarcasm detection, and the complexity of sentence/document.

REFERENCES

[1] John and R. Sheik, "Context Deployed Sentiment Analysis Using Hybrid Lexicon," 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), CHENNAI, India, 2019, pp. 1-5, doi: 10.1109/ICIICT1.2019.8741413.

[2] T. Le, "A Hybrid Method for Text-Based Sentiment Analysis," 2019 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 2019, pp. 1392-1397, doi: 10.1109/CSCI49370.2019.00260.

[3] S. K. Pathuri, N. Anbazhagan and G. B. Prakash, "Feature Based Sentimental Analysis for Prediction of Mobile Reviews Using Hybrid Bag-Boost algorithm," 2020 7th International Conference on Smart Structures and Systems (ICSSS), Chennai, India, 2020, pp. 1-5, doi: 10.1109/ICSSS49621.2020.9201990.

[4] H. S. and R. Ramathmika, "Sentiment Analysis of Yelp Reviews by Machine Learning," 2019 International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India, 2019, pp. 700-704, doi: 10.1109/ICCS45141.2019.9065812.

[5] Aishwarya, P. Wadhwa and P. Singh, "A New Sentiment Analysis based Application for Analyzing Reviews of Web Series and Movies of Different Genres," 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2020, pp. 390-396, doi: 10.1109/Confluence47617.2020.9058137.

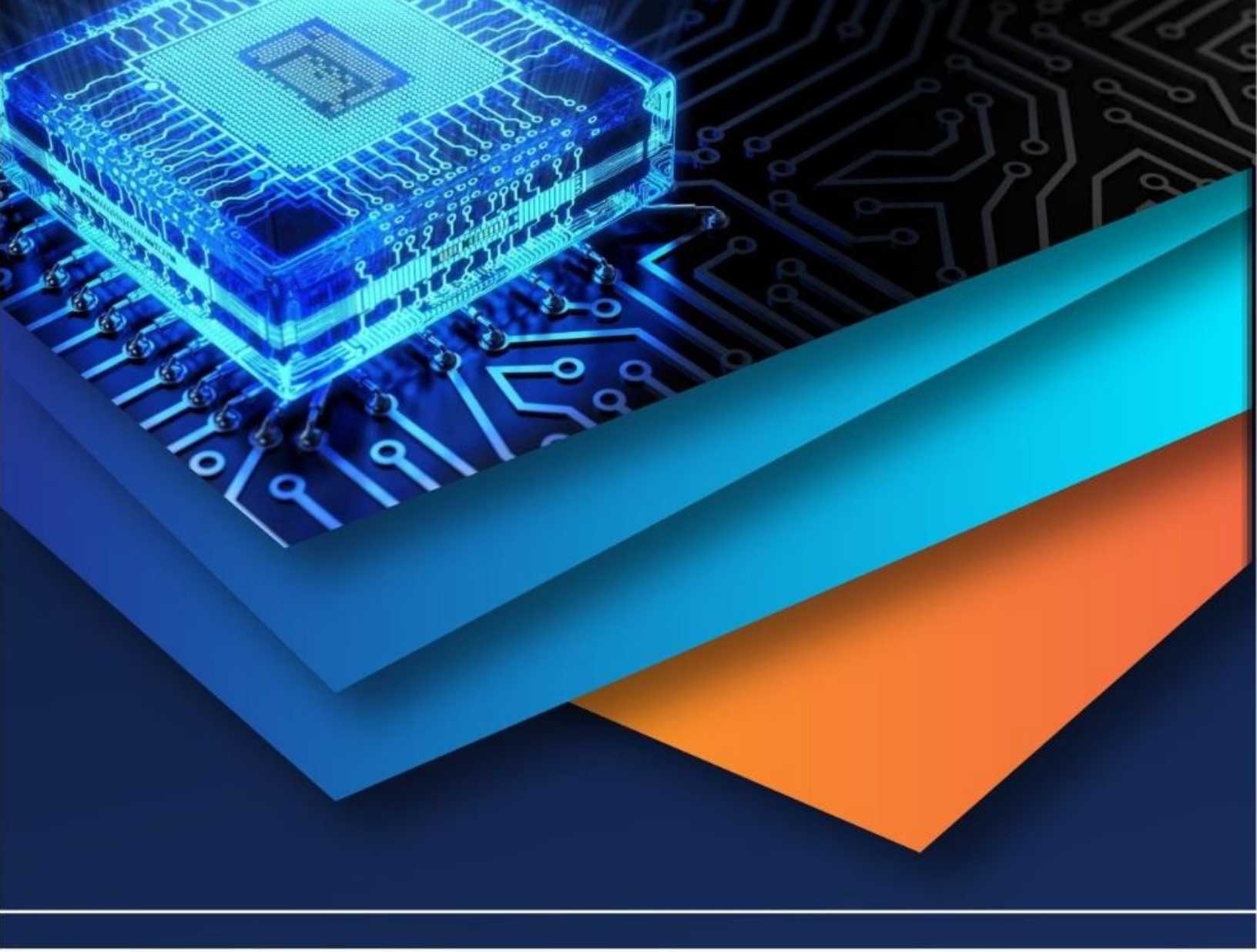
[6] S. G. Kanakaraddi, A. K. Chikaraddi, K. C. Gull and P. S. Hiremath, "Comparison Study of Sentiment Analysis of Tweets using Various Machine Learning Algorithms," 2020 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2020, pp. 287-292, doi: 10.1109/ICICT48043.2020.9112546.

[7] D. Beleveslis, C. Tjortjis, D. Psaradelis and D. Nikoglou, "A Hybrid Method for Sentiment Analysis of Election Related Tweets," 2019 4th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM), Piraeus, Greece, 2019, pp. 1-6, doi: 10.1109/SEEDA-CECNSM.2019.8908289.

[8] N. Yadav, R. Kumar, B. Gour and A. U. Khan, "Extraction-Based Text Summarization and Sentiment Analysis of Online Reviews Using Hybrid Classification Method," 2019 Sixteenth International Conference on Wireless and Optical Communication Networks (WOCN), Bhopal, India, 2019, pp. 1-6, doi: 10.1109/WOCN45266.2019.8995164.

[9] V. Singh, G. Singh, P. Rastogi and D. Deswal, "Sentiment Analysis Using Lexicon Based Approach," 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC), Solan Himachal Pradesh, India, 2018, pp. 13-18, doi: 10.1109/PDGC.2018.8745971.

[10] M. Rathan, V. R. Hulipalled, P. Murugeshwari and H. M. Sushmitha, "Every post matters: A survey on applications of sentiment analysis in social media," 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon), Bangalore, 2017, pp. 709-714, doi: 10.1109/SmartTechCon.2017.8358463.
[11] R. Baldania, "Sentiment analysis approaches for movie reviews forecasting: A survey," 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, 2017, pp. 1-6, doi: 10.1109/ICIIECS.2017.8275938.
[12] S. Bhardwaj and J. Pant, "A Survey of Approaches for Sentiment Analysis on Social Media," 2019 6th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2019, pp. 1020-1024.
[13] M. Ezhilarasan, V. Govindasamy, V. Akila and K. Vadivelan, "Sentiment Analysis On Product Review: A Survey," 2019 International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC), Melmaruvathur, Chennai, India, 2019, pp. 180-192, doi: 10.1109/ICCPEIC45300.2019.9082346.
[14] I. K. C. U. Perera and H. A. Caldera, "Aspect based opinion mining on restaurant reviews," 2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA), Beijing, 2017, pp. 542-546, doi: 10.1109/CIAPP.2017.8167276.
[15] F. M. T. Hossain, M. I. Hossain and S. Nawshin, "Machine learning based class level prediction of restaurant reviews," 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), Dhaka, 2017, pp. 420-423, doi: 10.1109/R10-HTC.2017.8288989.
[16] N. Rathee, N. Joshi and J. Kaur, "Sentiment Analysis Using Machine Learning Techniques on Python," 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2018, pp. 779-785, doi: 10.1109/ICCONS.2018.8663224.
[17] R. R. Nair, J. Mathew, V. Muraleedharan and S. D. Kanmani, "Study of Machine Learning Techniques for Sentiment Analysis," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 978-984, doi: 10.1109/ICCMC.2019.8819763.
[18] K. Kaviya, C. Roshini, V. Vaidhehi and J. D. Sweetlin, "Sentiment analysis for restaurant rating," 2017 IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), Chennai, 2017, pp. 140-145, doi: 10.1109/ICSTM.2017.8089140.
[19] S. Shayaa et al., "Sentiment Analysis of Big Data: Methods, Applications, and Open Challenges," in IEEE Access, vol. 6, pp. 37807-37827, 2018, doi: 10.1109/ACCESS.2018.2851311.
[20] M. Zhang, "E-Commerce Comment Sentiment Classification Based on Deep Learning," 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), Chengdu, China, 2020, pp. 184-187, doi: 10.1109/ICCCBDA49378.2020.9095734.
[21] R. Akila, S. Revathi and G. Shreedevi, "Opinion Mining on Food Services using Topic Modeling and Machine Learning Algorithms," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020, pp. 1071-1076, doi: 10.1109/ICACCS48705.2020.9074428.
[22] T. Anand, V. Singh, B. Bali, B. M. Sahoo, B. D. Shivhare and A. D. Gupta, "Survey Paper: Sentiment Analysis for Major Government Decisions," 2020 International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2020, pp. 104-109, doi: 10.1109/ICIEM48762.2020.9160252.
[23] N. Arunachalam, S. J. Sneka and G. MadhuMathi, "A survey on text classification techniques for sentiment polarity detection," 2017 Innovations in Power and Advanced Computing Technologies (i-PACT), Vellore, 2017, pp. 1-5, doi: 10.1109/IPACT.2017.8245127.
[24] P. Yang and Y. Chen, "A survey on sentiment analysis by using machine learning methods," 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, 2017, pp. 117-121, doi: 10.1109/ITNEC.2017.8284920.
[25] B. K. Bhavitha, A. P. Rodrigues and N. N. Chiplunkar, "Comparative study of machine learning techniques in sentimental analysis," 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, 2017, pp. 216-221, doi: 10.1109/ICICCT.2017.7975191.
[26] S. K. Jain and P. Singh, "Systematic Survey on Sentiment Analysis," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 2018, pp. 561-565, doi: 10.1109/ICSCCC.2018.8703370.
[27] Rahul, V. Raj and Monika, "Sentiment Analysis on Product Reviews," 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2019, pp. 5-9, doi: 10.1109/ICCCIS48478.2019.8974527.
[28] R. A. Laksono, K. R. Sungkono, R. Sarno and C. S. Wahyuni, "Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naïve Bayes," 2019 12th International Conference on Information & Communication Technology and System (ICTS), Surabaya, Indonesia, 2019, pp. 49-54, doi: 10.1109/ICTS.2019.8850982.
[29] D. Tiwari and B. Nagpal, "Ensemble Methodsof Sentiment Analysis: A Survey," 2020 7th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2020, pp. 150-155, doi: 10.23919/INDIACom49435.2020.9083693.
[30] R. Bandana, "Sentiment Analysis of Movie Reviews Using Heterogeneous Features," 2018 2nd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech), Kolkata, 2018, pp. 1-4, doi: 10.1109/IEMENTECH.2018.8465346.
[31] Yuki Yamamoto, Tadahiko Kumamoto, and Akiyo Nadamoto.“Role of Emoticons for Multidimensional Sentiment Analysis of Twitter”. In: Proceedings of the 16th International Conference on Information Integration and Web-based Applications & Services. ACM. 2014, pp. 107–115.
[32] Vohra, S. M., and J. B. Teraiya. "A comparative Study of Sentiment Analysis Techniques." Journal of Information, Knowledge and Research in Computer Engineering 2.2 (2012):313-317.
[33] ChauhanAshish P and Dr. K. M. Patel , “Sentiment Analysis Using Hybrid Approach: A Survey”, Int. Journal of Engineering Research and Applications www.ijera.com ISSN : 2248-9622, Vol. 5, Issue 1(Part 5), January 2015, pp.73-77
[34] O. Chapelle, B. Schölkopf, and A. Zien, “Semi-supervised learning,” Cambridge, Mass.: MIT Press, 2006.
[35] Z. Chen, A. Mukherjee, and B. Liu, "Aspect extraction with automated prior knowledge learning," in Proceedings of ACL, 2014, pp. 347-358.
[36] Sivakumar R, Chaithra V D, “Hybrid approach: naive bayes and sentiment VADER for analyzing sentiment of mobile unboxing video comments,” 2019, International Journal of Electrical and Computer Engineering (IJECE) Vol. 9, No. 5, October 2019, pp. 4452~4459 ISSN: 2088-8708, DOI: 10.11591/ijece.v9i5.pp4452-459.
[37] https://pypi.org/project/NRCLex/
[38] https://scikit-learn.org/stable/modules/naive_bayes.html#multinomial-naive-bayes
[39] https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY