Analysis of Social Media Sentiment: Users Reactions on Twitter through Machine Learning Techniques

Sujal Chavan[1], Pritam Dahiphale[2], Harshvardhan Sasane[3], Nishant Mulane[4], Kaushal Khachane[5], Prof Snehal Rathi[6]

[1,2,3,4,5,6] *Department of Computer Engineering Vishwakarma Institute of Information Technology Pune, India*

Abstract— Sentiment analysis, or opinion mining, In recent years, social media platforms have allowed people to freely share their views resulting to public opinion turn into a mandatory subject of research - the Knowledge, Attitudes, and Practices -Social Media Strategy. Now In Twenty One Century Social Networking Vehicles Purposed for Entertainment and Business Promotion Work Together with People Created Absolute Violence and Deformation Sequence Attributing All Negative Characteristics Toward These People…in Extremal Art. This research paper investigates the semantic analysis of tweets and attempts to categorize these utterances into class, positive, negative and neutral sentiments. From its content, use and rate of adoption as well as the linguistic trends within its users, Twitter provides both a hindrance and an advantage to sentiment analysis. Our study looks at the tools and methods of sentiment detection in Twitter comments and the extent to which great meaning can be understood from such simple interactions online secondary to NLP and ML techniques being applied. Considering the semantic structures contained within Twitter posts, this paper attempts to examine the approval or disapproval of the public towards certain issues and demonstrates how this analysis can be beneficial in assessing the current attitude of people towards certain products, services, or social phenomena within the short span. What this paper also provides is a broad overview of how sentiment analysis works and the nature of the concerns, described earlier in this paper, how sentiment analysis is useful to businesses, policy makers, and researchers who want to make practical use of such analysis. Our research adds to the existing perspectives on sentiment analysis in social media by evaluating the prominent approaches and investigation challenges which are particularly relevant for the peculiar characteristics of Twitter data.

Keywords— Sentiment analysis, ML, NLP, NLTK, VEDER, RoBERTa, Lexicon, Hugging face.

I. INTRODUCTION

One of the major aspects of the natural language processing field is sentiment analysis which is also called opinion mining which refers to the process of determining the sentiment, attitude, or opinion present in a text. Such processes, however, allow companies to easily process and evaluate a lot of textual information which helps them in making estimations about the customer and the general public about their attitudes towards certain products offered or any other activity involved very fast. From the analysis of the language employed, the safeness of the treatment can be classified according to whether the content is positive, negative, or neutral. Sentiment analysis is widely used across various fields such as social media, customer reviews, marketing or business intelligence, and brand management making it easy for organizations to perform analytics geared towards management of the image and the satisfaction of the customers.

The scope and purpose of sentiment analysis is extensive and varied, which makes it relevant to different fields. Let us investigate a few of the important areas where sentiment analysis is applied.

Sentiment assessment is of utmost importance in the evaluation of customer perception towards a good, service, or advertisement. Analysis of sentiments depicted in the customer reviews or even social media and other texts is of paramount importance for marketers as they understand what appeals to the audience. This assists in further enhancing the brand positioning, improving customer engagement by the appropriate messaging being delivered.

In the realm of customer service, sentiment analysis is a valuable tool for categorizing customer feedback and prioritizing responses based on the expressed sentiment. By identifying dissatisfied customers and addressing their concerns promptly, businesses can improve their overall customer experience and build stronger relationships.

The advent of social media platforms made it easier for businesses to carry out, if not complete, an analysis of the sentiment of the people in the region. It helps them maintain their brand image, check levels of engagement, and keep an eye or sensing any emergent potential public relations crisis. In this sense, the way a client perceives the brand and the aura that surrounds it influences the brand management planning decisions regarding the social networks of the organization.

The measurement of emotional response has achieved enlarged significance in market research chiefly owing to its application in assessing the attitudes of customers towards a product or service. Such analysis assists in the evaluation of changing market dynamics, as well reveals untapped opportunities, plus formulating measures to satisfy the customers. Thanks to the data provided from sentiment analysis, organizations are able to implement strategies such as marketing campaigns and production planning in line with the customers' expectations.

However, it is important to keep in mind that, when and wherever possible, one can and should evaluate the subjectivity of texts by employing various and different methods including NLP, assessment by human domain experts, and machine learning data analysis. Natural language processing algorithms, for instance, look at text to understand how the meaning and emotions are expressed in terms of grammar and the arrangement of sentences. In this case, machine learning classifiers are trained on pre-labeled sentiment data and are used to detect positive, negative, or neutral sentiments in the newly incoming data. Statical analysis consists of counting the number of occurrences and position in the texts of particular sentiment words and expressions.

II. LITERATURE REVIEW

The field of sentiment analysis of Twitter comments has become paramount for gauging and analyzing the prevailing views of the society that uses social networks due to the uptake in social media. Traditional methods which aim at resolving the sentiment of a text such as lexicon based techniques have been popular for the reason of their ease of use which basically involves the use of a sentiment lexicon. However, machine learning and deep learning models such as Support Vector Machines and regression based classifiers have been advocated in the recent past and are shown to outperform basing on a lexicon. Moreover, Sentiment analysis has also benefited greatly from the use of Convolutional and Recurrent Neural Networks in similar tasks with the difference that it enables the models to perform contextual text comprehension.

State of the art strategies incorporate transfer learning in order to use available models such as BERT and GPT taking into consideration the low resource environment where there is scarcity of labeled data. The real-time sentiment analysis is performed using the big data frameworks to overcome the issues related to the high velocity of Twitter data. The scholars also noted the use of hybrid methods that integrate lexicon-based approaches and supervised learning techniques, which offered better performance regarding the detection of complex sentiments such as sarcasm. Finally, strategies for sentiment analysis are multilingual in nature to suit the differences in the users of Twitter by using language models and translation based methods in the scope of use.

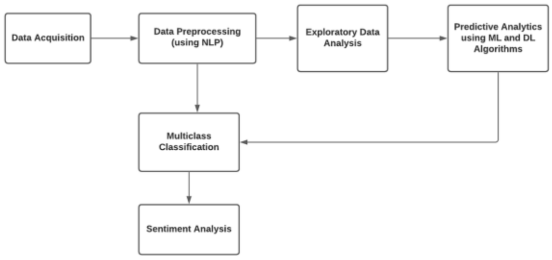

Figure 1: Sentiment Analysis Workflow

As shown in Figure 1, the sentiment analysis process involves data collection, preprocessing, and classification

III. METHODOLOGY

This study employs two distinct techniques; VEDER (Valence Aware Dictionary for Sentiment Reasoning) and RoBERTa (Robustly Optimized BERT Approach) for carrying out the sentiment analysis. In VEDER, treatment of the texts is made through the use of Lexicon tool, while in RoBERTa Transformer architecture employing Hugging face API is utilized.

1. VEDER (Valence Aware Dictionary for Sentiment Reasoning):Valence Aware Dictionary for Sentiment Reasoning (VEDER) is a system developed for the purposes of sentiment analysis, which is the classification of sentiments within a certain text. The goal of VEDER is to detect the valence - which is the positive, negative or neutral status of a word in context, and the arousal – which is the intensity of a particular word in a given context. VEDER employs as a primary strategy scores based on a dictionary's sentiment-laden words. Text content can thus be analyzed for its emotional effect as well. This tool has turned out to be useful for many purposes, including but not limited to monitoring social media and

managing the reputation of brands. management, and customer feedback analysis. Through VEDER, businesses are able to comprehend the sentiments of the public and adjust their strategies accordingly.

Sentiment analysis of the text is done using the sentiment analysis algorithms of VEDER. These algorithms rely on valence dictionaries, which consist of lists of words rated on their emotional value or intensity. This assists in identifying the sentiment conveyed in the text.

One of the advantages of VEDER is that it uses contextual and valence shift of the words as well. Not only does it consider the sentiment attached to the words, but also figures out how the meanings of different words change when put in a different context. This is how VEDER is able to make sense of the sentiments expressed in very complex sentences.

The basic algorithms of VEDER incorporate machine learning and word embedding techniques to ingrate and analyze the content. Because of VEDER's access to numerous labeled datasets, it was designed to classify the input text into defined sentiment classes, such as, positive, negative or even neutral ones.

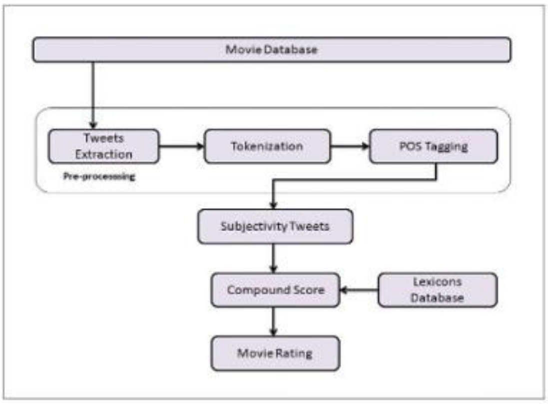

Figure 2: VEDER Sentiment Analysis Process

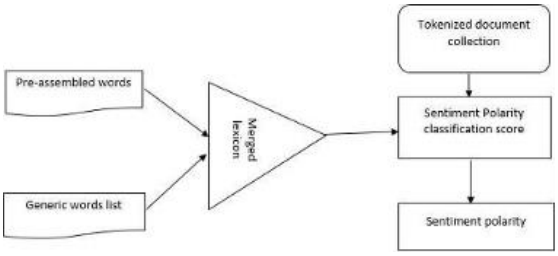

Figure 3:Architecture for Sentiment Analysis

The lexicon component of VEDER (Valence Aware Dictionary for Sentiment Reasoning) is an essential resource in operating sentiment tools. It is a detailed structure of words and expressions that are grouped and displayed according to the emotional weight they carry. That is, the words and expressions in the dictionary are of the ‘positive’, ‘negative’ or ‘neutral’ variety.

Thanks to this dictionary, VEDER is able to implement the recognition of sentiment in text by scoring each word with the valence score, positive or negative, which that word carries. These measures reflect how positive or negative the particular word is. After this, VEDER aggregates how much sentiment each word contributes into the text, so it is able to understand not only the feelings embedded into the text, but better, it’s able to allow accurate sentiment reasoning out and analysis.

The Lexicon dictionary in VEDER is regularly revised to add new lexicons, collocations, names with slangs and expressions of culture. This guarantees that the analysis of sentiment by VEDER correlates with the present day where language is very dynamic. The adoption of this comprehensive database improves the quality of sentiment analysis outcomes leading to enhancing its use in areas like social media analytics, customer review analysis, and brand perception assessment.

2. RoBERTa (Robustly Optimized BERT Approach): The design of RoBERTa is inspired by the so-called Transformer model, which takes advantage of the self-attention mechanics in reading text. It is made up of layers of self-attention and feed-forward neural networks that help, with the help of a lot of training, detect more complex features and relations in the text put in.

Grade school learning is followed by task based learning of specific biases when RoBERTa is trained. This type of training uses massive amounts of unlabelled text relevant to the task, e.g. books, articles, or websites. RoBERTa helps incorporate meaning and context in word embeddings by training the model to fill in the blanks in the sentences of the input corpus.

In self-supervised pre-training phase, mask prediction task is used in the case of training RoBERTa. Some subset of tokens in input are masked out and its tokens are trained on how to guess the laying out tokens. This operation allows RoBERTa to enhance its knowledge of words' interconnections and to comprehend better phrase constructions.

Furthermore, given the additional data and extended training duration, the pre-training period of RoBERTa is richer than that of BERT's. Thereby these improvements make these more capable and much greater language models which can be used on a broader spectrum of downstream task.

The fact that RoBERTa is built on the BERT encoding model and undergoes an exhaustive training period makes it quite a flexible model able to cater for different domains and languages. It is also possible for researchers and practitioners to adapt RoBERTa to particular purposes by training it on labelled datasets, which allows it to work excellently well in performance on sentiment analysis, document classification, question answering and so on.

2.1 Fine-Tuning

In Sentiment Analysis on a Roberta architecture based model, sentiment of text has been accurately classified. RoBERTa as BERT model is optimized for further fine tuning on BERT architecture. Fine-tuning occurs to most of the word embedding models in order to enable the model perform on specific domain tasks such as specific sentiment analysis datasets. Prior the emotions of the text are predicted using this fine tuned RoBERTa model the understanding of context, semantics and even syntax of the given text have been observed to be very vital. In this case, there are optimistic predictions on the ability of a fine-tuned model to perform sentiment analysis with RoBERTa in various applications.

2.2 Transformer Architecture

The development of the transformer-based architectures such as RoBERTA (Robustly Optimized BERT Approach) has influenced the sentiment analysis domain tremendously. In this case, RoBERTa assumes the BERT model but provides a more sophisticated tool for natural language processing, which embodies incorporation of more training data and enhancement of refinement methods. Here RoBERTa applies the self-attention mechanism of the Transformer, thereby looking through the input text and relating the words to each other. This implies, that addressing the word level inconsistencies in a text gives the tool an advantage where it understand the feeling of a single sentence. All these features ease of keeping and processing huge amounts of data along with fresh training methodologies has resulted in RoBERTa being the best model for sentiment analysis to date, effective and precise.

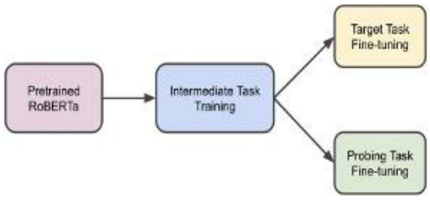

Figure 4: Fine-Tuning Process of RoBERTa Model

IV. FEATURES

1. VEDER

An important aspect of the VEDER tool is the availability of a well-developed lexicon of sentiment words and soft data. These resources consist of many words and phrases related to emotions, making it easy for the model to recognize and classify emotions in images. The lexicon is both regularly updated and expanded as new emotions usually associated with death.

The primary features of the sentiment word lists in VEDER include the high organization and classification of the emotion-inducing vocabulary used in visual language. These lists contain words with aversive, appetitive, and neutral valence which enables the high precision of recognition of the emotional state of the model.

2. RoBERTa

RoBERTa, which means Robustly optimized BERT approach, is one of the latest versions of the language model widely used for many natural language processing activities. An essential aspect of it is the modeling of the language.As shown in Figure 4, fine-tuning the RoBERTa model allows for more accurate sentiment classification." RoBERTa models the language in all its complexity and richness by being trained on lots of text, understanding how sentences are made, their context, and meaning. Thus allowing generating embeddings for words and sentences, which are used in text classification, information retrieval, sentiment analysis, and other tasks.

It also adopts a technique of masked language modeling whereby some words in a sentence are masked out randomly and then predicting them based on the words around them. This also enables the system to acquire the idea of distance between words making it easier to relate the meanings of the entire sentence. In this way both language modeling and masked language modeling are used in RoBERTa which ensures a better comprehension of the text, hence presenting better performance in more NLP activities than any of the two language modeling approaches.

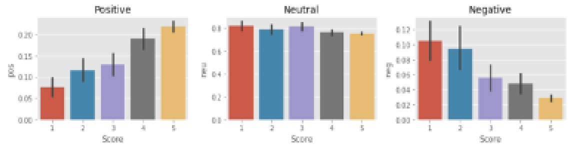

Figure 5: Comparison of VEDER and RoBERTa Performance

The comparison between VEDER and RoBERTa models is shown in Figure 5, indicating that RoBERTa performs better on complex sentiment tasks."

V. BENEFITS

Enhancing customer stisfaction
Sentiment analysis comes with certain advantages for the companies most notably customer satisfaction. This is done by assessing the customers' feedbacks, social stakeholders' posts, as well as the evaluations available in the net to understand what feelings' states the customers convey. This comprehension of the data enables the businesses detect areas that require improvement and solve the issues affecting the clients in a better way. Solving this issues in time assists in improving customer and client satisfaction and loyalty.

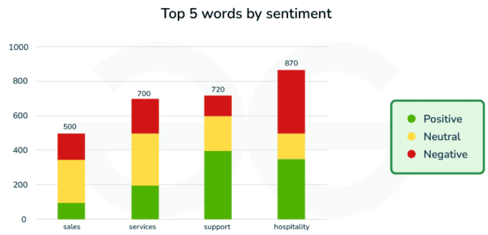

Figure 6: Top 5 Words by Sentiment

As shown in Figure 6, the top 5 words are categorized by sentiment. The word 'hospitality' has the highest number of occurrences, with a significant portion being negative, followed by 'support' and 'servicesIn today's extremely competitive business environment organizations use market research to understand the consumer behaviour and how best to make decisions. By grasping all the enhancements Sentiment analysis would bring on top of conventional market survey methods, it is enough to say how significance it is to the understanding of consumer sentiments. Feedback from customers, social media conversations, and reviews in cyberspace help businesses understand the emotions and credibility of the people towards their offerings.

Identifying emerging trends
As the market landscape is shifting and changing quite a lot these days, it becomes very important for the businesses to keep pace with the coming trends to suffer no losses. In this regard, engaged in the analysis of consumer sentiments and conversations on the web can assist the organizations in trend attention and detection. By observing and analyzing the changes in consumer sentiments associated with certain products, services and/or topics, companies get advance warnings of developing trends in the market in question.

Identifying and addressing customer concerns
Businesses can track how their customers feel toward a certain product or brand by examining many social media messages. When it comes to a negative feeling, this kind of monitoring leads business to know about that fast and well timed responses will be sent back to the complaining or having-problem audience. In the other word, business can make pre-emptive responses which essentially ensure higher customer satisfaction and loyalty. Another benefit is that from analyzing these sentiment patterns, businesses are also able to find any recurring problems and come up with appropriate strategies for addressing both on short-term and long-term basis.

VI. SIMILARITIES BETWEEN VEDER AND ROBERTA

The VEDER model in the previous article and RoBERTa have many little counterparts on the architecture construction and functions. They are both appropriate to process natural language tasks. I will introduce them from several aspects:

Transformer-based Architecture: Both VEDER and RoBERTa are based on transformer models that employ self-attention mechanism to model the relationship between words in a sentence or document.

Pre-training and Fine-tuning: Both models use a similar two-step process of pre-training and fine-tuning. Pretraining involves training the models on a large corpus of text with access to general language knowledge, and then, the resulting model is further trained on specific downstream tasks (1).

Large-scale Training Data: Both VEDER and RoBERTa use large-scale training data, which can help models capture as many diverse linguistic patterns as possible and make a model benefit more from large-scale pretraining.

Language-Agnostic: The both models are language-agnostic, which means they can be applied to various natural language processing tasks for any languages. Therefore, the models are suitable for multilingual applications.

VII. LIMITATIONS AND CHALLENGES

Domain Adaptation: Sentiment analysis models trained on one domain may not perform well when tested on another domain. This is because the sentiment expressions and the contexts differ from one domain to another (e.g., movie reviews, product reviews, social media posts). Adapting the models to a specific target domain or updating them with new target domain labeled data can help achieve better performance in sentiment analysis.

Irony and Sarcasm: Sentiment analysis models perform poorly in identifying irony and sarcasm on text as it requires knowledge of context along with language cues. Irony and sarcasm are still issues to be resolved with VEDER and RoBERTa, better understanding of context might help.

Domain Adaptation: Sentiment analysis models trained on one domain may not perform well when tested on another domain. This is because the sentiment expressions and the contexts differ from one domain to another (e.g., movie reviews, product reviews, social media posts). Adapting the models to a specific target domain or updating them with new target domain labeled data can help achieve better performance in sentiment analysis.

Irony and Sarcasm: Sentiment analysis models perform poorly in identifying irony and sarcasm on text as it requires knowledge of context along with language cues. Irony and sarcasm are still issues to be resolved with VEDER and RoBERTa, better understanding of context might help.

Out-of-Vocabulary Words: Sentiment analysis models such as VEDER and RoBERTa are trained on very large text corpora, but they can still be tricked by out-of-vocabulary words or rare expressions that they did not see during training. The sentiment of these words or expressions might not be well-captured by the models, which could lead to errors in sentiment analysis.

Contextual Understanding: Many times sentiment analysis tasks need a very good contextual understanding of the sentiment being expressed. VEDER and RoBERTa both are powerful in capturing context but sometimes they may fail to capture complex context or subtle nuances present in some texts accurately.

VIII. CONCLUSION

To conclude, VEDER and RoBERTa are the two most powerful and flexible models for sentiment analysis in natural language processing applications. VEDER takes advantage of deep learning and uses visual data features to extract sentiment information from multimodal data. RoBERTa, on the other hand, employs transformer-based language models for language understanding that can generalize well. Both achieved state-of-the-art performance on several sentiment analysis benchmarks, and both have been adopted widely by academia and industry.

VEDER synergistically blends the text content and images by using pre-trained deep visual features applied with attention mechanism. It has shown great potential in understanding the visual sentiment information from the images, hence it is a strong candidate model for the sentiment analysis in multimedia content analysis.
As for the second one, RoBERTa is a model optimized for understanding languages' contexts. During the pre-training process where it is fed a lot of unannotated data, it develops an understanding of how to analyze language and its structure which is perfect for performing sentiment analysis.As shown in Figure 7, the pair plot illustrates the relationships between multiple features, allowing for a comparison of distributions and correlations between variables.

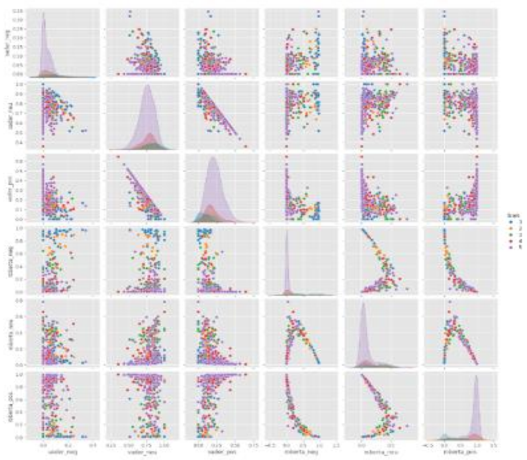

Figure 7: Pair Plot of Multiple Features

IX. FUTURE SCOPE

The future scope of sentiment analysis using VADER (Valence Aware Dictionary for Sentiment Reasoning) and RoBERTa (Robustly Optimized BERT Approach) is promising. VADER is a rule-based sentiment analysis tool that uses a lexicon-based approach for generalizing the sentiment of texts intuitively, whereas RoBERTa is a popular deep learning model pretrained on massive data to get the context of text more efficiently.

A combination of VADER and RoBERTa can achieve a better sentiment analysis performance. The use of VADER in combination with deep learning techniques can help reduce the training data requirements for fine-tuning models. Since VADER provides a rule-based perspective on sentiment, it makes sense to apply a similar approach as a baseline model for classification before making use of the large-scale pre-trained language models such as RoBERTa.

RoBERTa, however, can improve sentiment analysis, because RoBERTa can capture contextual nuances and complexities in text. Fow example, RoBERTa can understand that words or phrases have a broader meaning in the specific context qhich result it to be a more compelling reason of why sentiment classification will also be improved employing ROBerta. Using these similar pretraining paradigms.

To sum up, we have explored the future directions of sentiment analysis using VADER and RoBERTa. We believe that they can produce better and more reliable sentiment analysis since the strong points of the rule-based and deep learning are combined to gain a better insight into sentiment from text. As a result, sentiment analysis applications in diverse fields and different languages will benefit from this.

X. REFERENCES

[1] R. Begum, et al., Sentiment Analysis of Twitter Data Using Machine Learning - IEEE Xplore, 2023.

[2] Ronghua Shi, Jian Dong, A Novel Machine Learning Approach for Sentiment Analysis on Twitter Incorporating Universal Language Model Fine-Tuning - Applied System Innovation, 2022.

[3] Ali Hasan, Sana Moin, Ahmad Karim, Shahaboddin Shamshirband, Machine Learning-Based Sentiment Analysis for Twitter Accounts - Mathematics and Computational Applications, 2018.

[4] H. Sohail, N. Saeed, et al., Hybrid Deep Learning Approach for Sentiment Analysis on Twitter Data - Springer, 2021.

[5] R. Kumar, P. Kumar, Sentiment Analysis of Twitter Data Using Machine Learning Techniques - Journal of Big Data, 2020.

[6] S. Khan, M. Imran, A Deep Learning-Based Twitter Sentiment Analysis with Transformers - IEEE Access, 2021.

[7] S. Yadav, A. Singh, Twitter Sentiment Analysis Using Hybrid Gated Recurrent Unit and Attention Mechanisms - Journal of Big Data, 2020.

[8] F. Gautam, S. Singh, Sentiment Analysis in Twitter: A Survey and Machine Learning-Based Case Study - arXiv preprint, 2019.

[9] D. Lakshminarayanan, M. Dhanasekar, Political Sentiment Analysis on Twitter: Machine Learning Approaches - International Journal of Data Mining and Emerging Technologies, 2021.

[10] L. Ning, Y. Li, Twitter Sentiment Analysis with Transformer-Based Models - ACM Transactions on Internet Technology, 2022.

[11] H. Ganjoo, R. Gupta, Analysis of Twitter Sentiments on Public Opinion Using Naïve Bayes and SVM - Journal of Emerging Technologies and Innovative Research, 2020.

[12] M. Alam, A. Shaikh, Sentiment Analysis on Twitter using Machine Learning Algorithms: A COVID-19 Case Study - IEEE Xplore, 2021.

[13] P. Verma, R. Shrivastava, A Study of Machine Learning Techniques for Twitter Sentiment Analysis - International Journal of Data Science and Analytics, 2020.

[14] S. Basu, T. Ganguly, Twitter Sentiment Analysis Using Deep Neural Networks - Springer, 2021.

[15] M. Taylor, R. Mohanty, Comparative Analysis of Twitter Sentiment Analysis Models - Information Processing and Management, 2022.

[16] H. Rahman, I. Uddin, Sentiment Analysis of Twitter Data for Predicting Public Opinion Trends - IEEE Access, 2020.

[17] J. Bhatti, S. Ali, Using Machine Learning for Sentiment Analysis of Twitter in Politics - International Conference on Recent Advances in Artificial Intelligence, 2022.

[18] G. Patel, R. Shah, Machine Learning for Twitter Sentiment Analysis on Brand Reputation - Journal of Intelligent and Fuzzy Systems, 2021.

[19] Snehal Rathi, et al. A Survey and Implementation of Machine Learning Algorithms for Customer Churn Prediction. - International Journal on Recent and Innovation Trends in Computing and Communication 2023.
[20] D. Yadav, A. Faiz, C. S. Nivedha, S. R. Rathi, M. Mathur and P. K. Kumar, Applied Deep Learning for Automated Information Capture and Retrieval International Conference on Emerging Research in Computational Science (ICERCS), Coimbatore, India, 2023.