iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** IV **Month of publication:** April 2021

DOI: https://doi.org/10.22214/ijraset.2021.33644

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 IV Apr 2021- Available at www.ijraset.com

A Review Paper on Sentiments Analysis using Social Media

Pranit Malviya[1], Shubham Kumar[2], Sunidhi Raj Kohli[3]
[1, 2, 3]*Computer Engineering Department, NMIMS University, Mumbai.*

Abstract: ***In an era of the developing world, social sites that are used as one of the platforms that most people connect. Everyone has a direct or indirect connection to social networking sites. People tend to take updates on anything before they do- like taking movie reviews, online shopping, restaurants and much more. Getting updates means knowing the ideas on things. In one way it can be described as an emotional analysis. Here, the current work comes with the concept of Social Media (Twitter) Analysis Sentiment, where the process is designed to know a person's thoughts on a particular tweet they have done. After knowing the views of people on any issue, anyone can come to a conclusion. To carry out this type of analysis, the current work has taken machine learning methods with Natural Language Processing (NLP), in addition to a hybrid system that uses Logistic Regression, Support Vector Machines and neural networks for sensory segregation. Therefore, research paper that we are using describes the types of machine learning strategies used. The revised articles were 21 articles from 2016 to 2020. type, and basic findings. The results of this review show the positive effects of using a variety of electronic learning techniques on social media to analyze emotions. Overall, this review provides an appropriate guide to ongoing research by identifying research needs in the field of emotional communication.***

Keywords: Sentiment Analysis, Machine learning, Support Vectors Machine, social media, Multinomial Naive Bayes, Logistic regression.

I. INTRODUCTION

Critical analysis is the process of obtaining information from a consumer's perspective about a product, service, or product. How they react to a product or event. Today, social networking sites such as Instagram, YouTube, Facebook, Twitter, and more are on the way. The only purpose of the present work is to penetrate one's behavior, feelings,

Vision, experience from text data. Personal opinion about any outcome or decision making is very important because people's opinions are based on their previous knowledge. In sentiment analysis, this method helps to find conflict between different emotions of a person who loves happiness, sadness, boredom, hatred, love, excitement, worry, relief, happiness, anger, nothingness, surprise, and neutrality. It is not only based on negative or constructive propagation for words and concepts, but also on the combined tree of descriptive sentences.

The program attempts to read between lines, find confusing or coherent expressions, provide interpretation of contradictory words, adjust word size based on adjectives, adjectives, conjunctions, or related verbs, noting certain abortions.

The emotional analysis is the process of determining the emotional tone behind a series of words, which are used to gain an understanding of the attitudes, thoughts, and feelings expressed in this text. It is a method of checking written or spoken language to determine whether a statement indicates, or is neutral.[1]

A. Motivation And Need For Sentiment Analysis

With the vast amount of data amassed on the internet, it is past time that social media and its data be given top priority. What some people think must be followed by all is a commonly held belief in the general public. This article provides a nearly chronological order in the creation of social networks, data collection and interpretation on social media, and ultimately predicting feelings within these data.[16]

B. Search Criteria

Over well-known literature, a systematic study of Machine Learning Techniques was conducted. The following search string terms were used: sentiment analysis, social media, and machine learning techniques. It also includes deep learning techniques, random forest, support vector machine, logistic regression and so on. The search engine list is listed in Table 1 of this study.

Table 1 Finding Engine Choose

FINDING ENGINE	SOURCE ADDRESSS
IEEE Xplore	https://ieeexplore.ieee.org
Springer	https://www.springer.com
Science Direct	https://www.sciencedirect.com
Semantic Scholar	https://www.semanticscholar.org
Academia	https://www.academia.edu

C. Data Sources

Various data sources were considered for this survey. The research paper used Scopus, journals, Google Scholar, books, and journals as databases to extract similar research papers from conferences and journal research papers. There are following databases that we are using in our search papers.

D. The Exploration Criteria

The study was carried out between 2016 and January 2020. Included among the scanned research papers are those that pass the consistency test. This includes peer-reviewed Articles, Google Scholar, books, blogs, and white papers.

E. Quality Assessment

The background of the study is one of the possible ways of conducting a variety of quantitatively specific research. Quality evaluation criteria were used to include and exclude research papers from search results. After the initial review of the abstraction, some academic papers were finished. Those research papers were extensively reviewed, and depending on the criteria, they were either dropped or included for examination. The following are the primary admittance criteria for research papers:

1) The major focus on the Machine Learning and its Techniques.
2) Second, it involves selecting the research papers that define the various analyses of social media essentially Twitter data by various machine learning and deep learning techniques.

F. Context And Design Of Research

Traditional emotional analysis involves the use of dictionaries to determine the accuracy of those ideas and then to measure these values in order to determine the meaning of the text. The partition was then performed using a simple ML model. This can only be achieved by removal of the "features" that are been using in the text and can use these features to predict the “label” as well.

Traditional ML methods can get the right results, but they are problematic because they involve manual labor to create features. They no longer have a clear answer to the word order. Those problems are solved with a section of machine learning techniques called "Deep Learning."

The first method is based on rules and uses a dictionary of terms attached to the label with a sense test. Emotional standards are often paired with additional rules for reducing sentences that include arguments, sarcasm, or contradictory clauses.

1) *Rule-based Sentiment Analysis:* This first method is basically on rules and use a dictionary of terms attached to the label with a sense test. Emotional standards are often paired with additional rules for reducing sentences that include arguments, sarcasm, or contradictory clauses.
2) *Machine Learning (ML) Based Sentiment Analysis:* Using a training label with an emotional label, we train the ML model to identify emotions based on words and their order. This method is highly dependent on the type of algorithm and the accuracy of the training data used.

II. OBJECTIVE

The goal of Sentiment Analysis is to identify automated methods for the extraction from text, such as thoughts and emotions, of subjective information in the natural language so as to establish organized and workable knowledge for either decision support system or decision maker. With the emergence and development of social networking, sentimental analysis gained much more importance.

The model approaches a variety of computer domains, including the processing of natural languages and statistics.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IV Apr 2021- Available at www.ijraset.com

III. HOW DATA IS PROCESSED FOR ANALYSIS

A. Data Collection By Web Scraping

Critical analysis is a subset of machine learning that performs the process of identifying and extracting confidential basic information. It can be an idea, a judgment, or a feeling of some problem. With web removal technology used to retrieve large amounts of data from the web. After data collection, we used a variety of methods, including the use of the Twitter Scraper module, to evaluate the positive, negative and neutral responses of users according to their message.[4][15]

B. Data Preparation

Data processing involves the preparation of data from multiple sources. Data processing. In essence, we obtained raw, unwashed data. We must organize the data into a particular structure for further analysis. In order to prepare the model, we must complete a particular task. Data preparation is referred to as the steps to prepare the information in a certain form that can be further analyzed. We can then start the model preparation process. After it is obtained from the Twitter API, the Data Frame Pandas was designed to save data.[4][5][13]

C. Data Visualization

Data visualization is the graphical information representation technique. Data visualization requires certain methods for properly visualizing the specific data. Toevaluate the design from visualization the data visualization technique is used. Currently, tools and technologies for data visualization are widely used in the industry.[4][5][7][12]

D. NLP and Text Processing

NLP has been used for search engines, feed news analysis, and voice-enabled applications such as Siri and Alexa recently. A collection of libraries can be used to create such text processing systems, Natural Language Toolksit (NLTK) via Python. Indigenous Language Processing refers to any software or computer fraud or text or speech detection. The analogy is that people work together, understand and respond appropriately to each other's opinions. This touch, understanding and response is not produced by anyone in NLP but on a computer.[4]

IV. RESEARCH METHOD OVERVIEW

The review process began with the selection of a subject and gathering basic information, followed by a list of questions to be reviewed, a search for relevant literature, an analysis of the selected literature, a record of the findings, and an exploration of the research challenges.

A. Decision Tree

It is used in the decision tree learning as a predictive model. The branches reflect the various findings, and the tree's leaves represent the conclusions. The primary objective of using a decision tree is to evaluate and forecast the class or value of target variables based on training data. The decision tree classifier has many advantages over other methods, including ease of analysis and relatively little data planning. However, it can often result in complex trees that are difficult to generalize. This approach is widely used in fields such as data mining and statistics.[2][12]

B. Logistic Regression

Logistic regression in situations where one or more independent variable occurs to determine the results is used for statistical analysis. For outcomes estimation a dichotomous variable is used. Even a special case of linear regression can be taken into account logistic Regression.[17]

C. Support Vector Machines

SVMs are a linear, discriminatory classification. SVMs are linear. The theory of maximization of margins is mainly focused on them. It is used to minimize structural risks. The basis of support of vector machinery is margin maximization (SVMs). The complexity of the classifier is greatly improved. To solve them, only a small part of the training data is required. This is because hyper aircraft are easily measured for SVMs. SVM strives for exceptional results for generalization.[2]. A single kernel SVM is commonly used to process data in a variety of contexts, including social media and linear regression. SVM is considered one of the most effective methods of textual distribution. Each data item is sorted as a point in the n-dimensional space, with the value of each element equal to the value of each link, where n is called the number of the elements used.[14][17]

D. Artificial Neural Network

An Artificial Neural Network (ANN) is a data-processing computing system made up of several interconnected components. These components are referred to as neurons because they work together to solve a problem. The biological nervous system inspired the processing model in ANNs. ANNs have a range of benefits over other classifiers, including increased multilayer functionality and back-propagation. The preprocessing and evaluation of the necessary dataset for analysis are discussed in the following section.

E. Random Forest

Random Forest is a supervised learning algorithm that can be used to solve problems like regression and classification. A Random Forest is simply a set of trees, each of which is distinct from the others.

It creates several decision trees and then merges them to obtain an absolute and stable value, which is primarily used for training and class performance.[13]

1) The Random Forest algorithm chooses N records at random from a dataset. The tree is built based on the N documents.
2) The N records are used to construct the decision tree.
3) The number of trees was chosen based on the data available.
4) In the case of a regression problem, each tree in the forest predicts a new record's Y value.
5) To determine the final value, all of the trees in the forest estimated the sum of all values.
6) In the case of classification issues, each tree in the forest predicts which group the new record belongs to.
7) The new record was then allocated to the appropriate group.

F. Adaboost Decision Tree

ith programs defined as 1 to t, the AdaBoost algorithm repeats the basic learning process. The main purpose of the promotion algorithm is to spread the weight set across the training database. If Dt I is distributed weight of training database I use in the first step, all the instruments are evenly distributed, and the weight of the wrongly divided wolves gradually increases with each iteration, and the weaker students are more focused on training examples. Training data in the form of t (x1, y1), (xm, ym), where xi represents the fast space X and i i represent the class symbol set Y, is included in the reinforcement algorithm. Let's assume Y = 0 and 1. [12]

G. Multinomial Naïve Bayes Algorithm

Multinomial bayes have the potential for the distribution of P by a piece of text for each class, with the "ignorant" assumption that the work is conditionally independent. e. Judgment consistency should also be limited by MultinomialNB () in terms of the conditions of the conditional input period. s. Multinomial Naive Bays calculate the probability of a potential P (c | y) activity where c is a potential outcome and y is a given example, representing a specific factor, to be categorized; it

$$P(c|y) = \frac{P(y|c) * P(c)}{P(y)}$$

Multinomial Naïve Bayes can be used to classify tweets as positive and negative in input [14]

H. Maximum Entropy Classifier

No assumptions have been made in the Maximum Entropy separator regarding the relationship between the data extracted from the database. By limiting the conditional distribution of the category marker, this category of editing attempts to maximize system penetration at all times.

The multiplicity of entropy is also considerate of fragmentation and is similar to the systematic approach to determining the distribution of groups. In contrast to the Naive Bayes, the conditional distribution is described as MaxEnt not making sense of independence with its features.[19]

I. Lexicon-Based Approaches

To test cohesion, a dictionary-based approach [20] uses an emotional dictionary with theoretical concepts and compares it with data. They give emotional statistics in word ideas, which shows how good, contradictory, or intentional words in the dictionary are.[19]

J. Optical Sentiment Analysis

Human emotion analysis is carried out not only through records, but also through photographs. An optical sentiment analysis categorization method based on deep convolution neural networks was used on over a million labeled images obtained from Flicker. Using adjective noun expressions extracted involuntarily from the pictures, this technique, which was applied on the novel deep learning system Caffe, assisted in evaluating the emotions depicted in the images. Traditional methods, such as support vector machine (SVM) categorization techniques, were outperformed by this method.

V. DISCUSSION

A. Comparison to Various Machine Learning Techniques

Authors and Paper Title	Accuracy
PAPER 1	SVM - 91% Naive Bayes-83% Maximum Entropy-80%
PAPER 2	Maximum Entropy-80%
PAPER 5	Neural Network-74.5%
PAPER 12	Decision Tree-84% SVM-82%
PAPER 13	Random Forest-97% SVM-92%
PAPER 14	Logistic Regression-86.23% SVM-85.69% Naive Bayes-83.54%
PAPER 17	Naïve Bayes-80.62% SVM - 92.42% Random Forest-83.08% Logistic Regression -89.24%
PAPER 19	SVM-86.40% Deep Learning-80.70%

B. Constraints

The difficulties in implementing automated analysis are because the complexity of natural language as well as characteristics of the posted material are drawbacks. The study of tweets is an example of this, since they are often accompanied by hashtags, emoticons, and links, making identifying the expressed sentiment difficult.

Sentiment analysis is much more difficult with social media monitoring because there is no fixed contextualization mechanism. Humans rate sentiment correctly 74.5 percent of the time, even though they can sense irony or sarcasm. Not only for a computer, but even for a person, it can be difficult to comprehend.

C. Applications Of Social Media Sentiment Analysis

Sentiment analysis's development has been aided by the wide variety of applications available. People have been able to make sense of large quantities of social media data using sentiment analysis tools in order to make informed decisions and consider social trends, product marketing strategies, and political events. The four papers chosen for this Special Issue address using sentiment analysis to improve health care, better understanding Helps patients, e-commerce consumer profiling, and identifying cyber-crime.[7][21]

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IV Apr 2021- Available at www.ijraset.com

D. The Logical Components

The components of the Sentiment and Knowledge Mining scheme [8] used in this analysis are as follows:

1) Crawler, a flexible and some sort of selective tool which collects docs that is consider the Database sources, internet, or Intranet.
2) A Semantic engine that finds semantic connections and facts in texts to identify specific details.
3) The search engine query in Natural Language, Semantic, and Semantic-Role.
4) Machine Translation Engine that allows search results to be automatically translated.
5) Geo-refermentation Engine, which make its possible for each user to view documents in map-like manner.
6) The Separation Engine that collects search results into clusters and subgroups, highlights meaningful connections between them, or provides documentation to predefined groups.

VI. FUTURE DIRECTIONS

We will do the same research on Facebook and Instagram as well. So, we can get a good picture of how much people enjoy expressing their feelings. We can list emotions in this way. To maximize project performance, a more sophisticated algorithm such as the Naive Bayes, used for text classification, should be applied to it. With a clear understanding, many mathematical parameters should be applied to the research project.

If the device uses social media sites like Twitter, YouTube, and Online-News to collect the requisite data from customers/people. These are useful for streaming public data on their own. Various pre-processing operations are performed on this data to produce reports such as histograms, graphs, bar plots, pie-charts, and so on. Real-time analysis is a powerful tool for industry, marketing, and politics that extracts the outcome of current trends through feedback and aids in decision-making. This model can also be made faster by using parallel distributed databases. Many high-end report representations can also be used for much deeper research. Future work involves:

A. Mining fine-grained opinions on closed captions of YouTube.
B. Videos Deep emotion intensities.
C. Fine-grained emotions analysis.

Sentiment analysis has a wide range of applications, including product reviews, movie reviews, determining consumer needs, gaining access to online news data, spam detection, and more.

It is possible to achieve high accuracy by adjusting the parameters of the machine learning algorithms. The release of three machine learning methods, including SVM, Multinomial Naive Bayes, and Logistic regression, was compared in this study. When using bigram models, Logistic Regression was 86 percent accurate.

When it comes to Twitter's emotional analysis, asset order surpasses other learning algorithms for monitored machines. Future studies may include evaluating the effectiveness of emotion analysis algorithms when considering specific factors.

To identify Twitter feelings and increase decision-making confidence, other effective learning strategies such as error reduction, pool-based sampling, uncertainty sample, and more will be used.[14]

VII. CONCLUSION

It's easy to see how critical public opinion is for a product-based business or government these days. The definition of fine-grained sentiment analysis was described in this paper using a widely used social networking site called Twitter. The assessment of the feature extraction is from twitter data and has been completed and can be used as a further research and decision making that can be order to mine opinion and sentiments.

In current study, it was also discovered that decision trees and logistic regression techniques are more accurate than SVM. Due to the sheer amount of data on the Internet, it also offers applications and evolving problems. Finally, sentiment analysis is a new field of Decision Support System science. This one-of-a-kind paper examines social networks and related concepts in great detail. The work in the cluster, community, and social sectors has been completed. Networks have been established in their depth. This article's main goal is to point out the flaws in a variety of papers that help researchers use sentiment analysis methods after collecting data from social media.

REFERENCES

[1] U. Hassan, J. Hussain, M. Hussain, M. Sadiq and S. Lee, "Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression," 2017 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea (South), 2017, pp. 138-140, doi: 10.1109/ICTC.2017.8190959.

[2] A. Gupta, A. Singh, I. Pandita and H. Parashar, "Sentiment Analysis of Twitter Posts using Machine Learning Algorithms," 2019 6th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2019, pp. 980-983.

[3] R. Meena and V. T. Bai, "Study on Machine learning based Social Media and Sentiment analysis for medical data applications," 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 2019, pp. 603-607, doi: 10.1109/I-SMAC47947.2019.9032580.

[4] D. Dansana, J. D. Adhikari, M. Mohapatra and S. Sahoo, "An approach to analyse and Forecast Social media data using Machine Learning and Data Analysis," 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA), Gunupur, India, 2020, pp. 1-5, doi: 10.1109/ICCSEA49143.2020.9132895.

[5] A. K. Goel and K. Batra, "A Deep Learning Classification Approach for Short Messages Sentiment Analysis," 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India, 2020, pp. 1-3, doi: 10.1109/ICSCAN49426.2020.9262430.

[6] Ahmad, S., Asghar, M.Z., Alotaibi, F.M. et al. Detection and classification of social media-based extremist affiliations using sentiment analysis techniques. Hum. Cent. Comput. Inf. Sci. 9, 24 (2019). https://doi.org/10.1186/s13673-019-0185-6

[7] Iglesias, Carlos & Moreno, Antonio. (2019). Sentiment Analysis for Social Media. Applied Sciences. 9. 5037. 10.3390/app9235037.

[8] Federico Neri, Carlo Aliprandi, Federico Capeci, Montserrat Cuadros, and Tomas By. 2012. Sentiment Analysis on Social Media. In Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012) (ASONAM '12). IEEE Computer Society, USA, 919–926. DOI:https://doi.org/10.1109/ASONAM.2012.164

[9] A. Manjula1, A. Rama Mohan Reddy, "Sentiment Analysis on Social media", International Journal of Computer Engineering In Research Trends, 6(11):pp1-6, November 2019. DOI: https://doi.org/10.22362/ijcert/2019/v6/i11/v6i1101

[10] Bharat Naiknaware , Bindesh Kushwaha , Seema Kawathekar, "Social Media Sentiment Analysis using Machine Learning Classifiers".

[11] Abdallah Yousif1 · Zhendong Niu1,2 · John K. Tarus1 · Arshad Ahmad1,"A survey on sentiment analysis of scientific citations". Published online: 5 December 2017 © Springer Science+Business Media B.V., part of Springer Nature 2017

[12] M. Rathi, A. Malik, D. Varshney, R. Sharma and S. Mendiratta, "Sentiment Analysis of Tweets Using Machine Learning Approach," 2018 Eleventh International Conference on Contemporary Computing (IC3), Noida, India, 2018, pp. 1-3, doi: 10.1109/IC3.2018.8530517.

[13] P. Karthika, R. Murugeswari and R. Manoranjithem, "Sentiment Analysis of Social Media Network Using Random Forest Algorithm," 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), Tamilnadu, India, 2019, pp. 1-5, doi: 10.1109/INCOS45849.2019.8951367.

[14] A. Poornima and K. S. Priya, "A Comparative Sentiment Analysis Of Sentence Embedding Using Machine Learning Techniques," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020, pp. 493-496, doi: 10.1109/ICACCS48705.2020.9074312.

[15] M. H. Abd El-Jawad, R. Hodhod and Y. M. K. Omar, "Sentiment Analysis of Social Media Networks Using Machine Learning," 2018 14th International Computer Engineering Conference (ICENCO), Cairo, Egypt, 2018, pp. 174-176, doi: 10.1109/ICENCO.2018.8636124.

[16] K. Chakraborty, S. Bhattacharyya and R. Bag, "A Survey of Sentiment Analysis from Social Media Data," in IEEE Transactions on Computational Social Systems, vol. 7, no. 2, pp. 450-464, April 2020, doi: 10.1109/TCSS.2019.2956957.

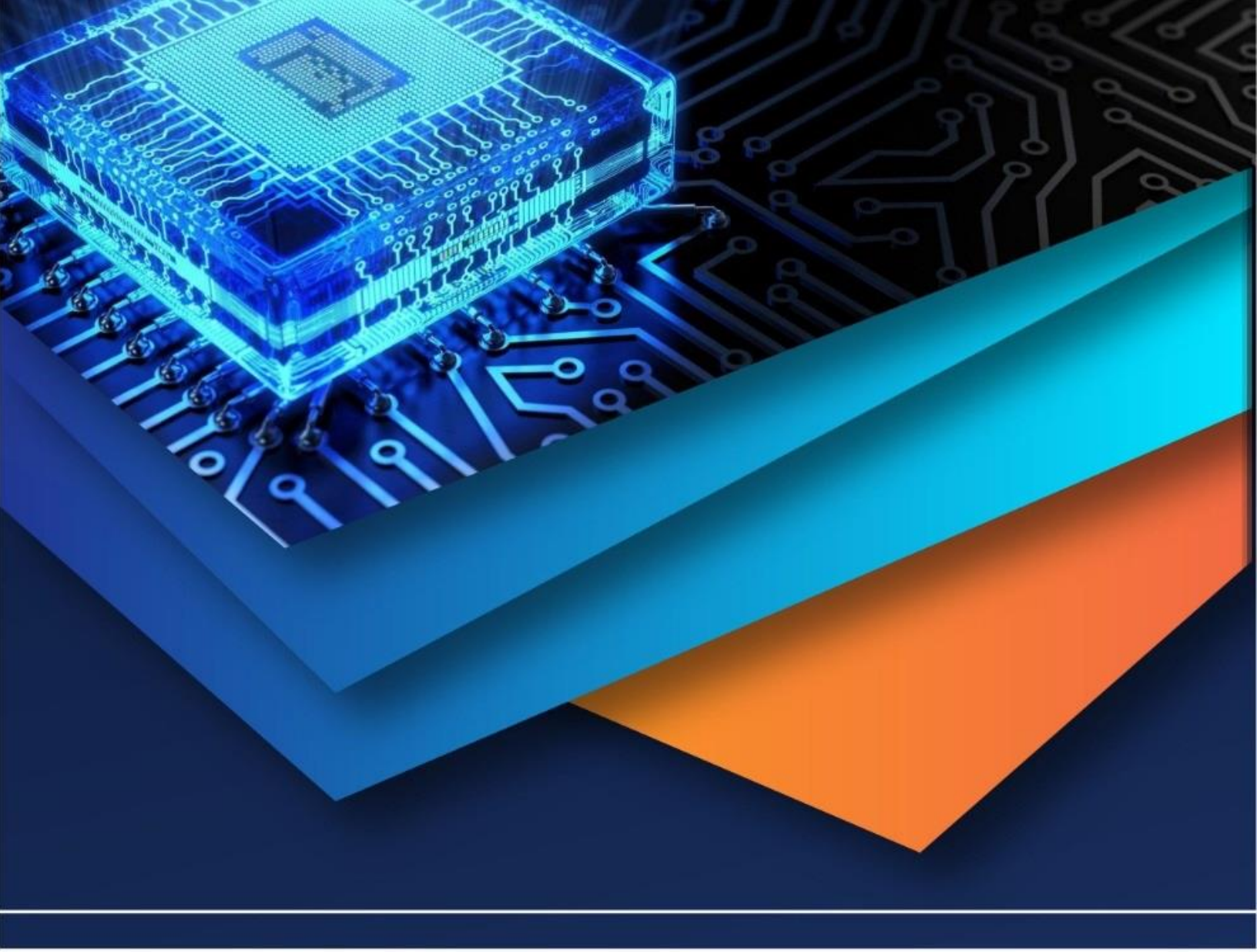
[17] C. Kaur and A. Sharma, "Social Issues Sentiment Analysis using Python," 2020 5th International Conference on Computing, Communication and Security (ICCCS), Patna, India, 2020, pp. 1-6, doi: 10.1109/ICCCS49678.2020.9277251.

[18] Balahur Dobrescu A. Sentiment Analysis in Social Media Texts. In Conference Proceedings: Alexandra Balahur, Erik van der Goot, Andres Montoyo, editors. Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. Stroudsburg, PA (USA): Association for Computational Linguistics; 2013. p. 120-128. JRC82012

[19] A., Vishal, and S.S. Sonawane. "Sentiment Analysis of Twitter Data: A Survey of Techniques." International Journal of Computer Applications 139.11 (2016): 5–15. Crossref. Web.

[20] T. Wang, K. Lu, K. P. Chow and Q. Zhu, "COVID-19 Sensing: Negative Sentiment Analysis on Social Media in China via BERT Model," in IEEE Access, vol. 8, pp. 138162-138169, 2020, doi: 10.1109/ACCESS.2020.3012595.

[21] M. Lovelin Ponn Felciah and R. Anbuselvi, "A study on sentiment analysis of social media reviews," 2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, India, 2015, pp. 1-3, doi: 10.1109/ICIIECS.2015.7192949.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY