iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** III **Month of publication:** March 2026

DOI: https://doi.org/10.22214/ijraset.2026.78808

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 14 Issue III Mar 2026- Available at www.ijraset.com

AI-Driven Skincare Product Analytics & Review-Sentiment Modeling

Mr.A..Libonce[1], Avula Gnanendra Reddy[2], Avula Likitha[3], Avulappagari Shirisha[4], G Sneha[5], Galeti Sree Lakshmi[6]
[1]*Associate Professor,* [2,3,4,5,6]*CSE (Data Science), Sri Venkateswara College of Engineering and Technology (SVCET), Chittoor*

Abstract: ***The rapid growth of online skincare markets has led to an overwhelming amount of customer reviews, making it difficult for consumers and companies to extract meaningful insights. This project presents an AI-driven skincare product analytics and sentiment analysis system that leverages Natural Language Processing (NLP) and Machine Learning techniques to analyze and classify customer feedback.***
The system processes textual data from skincare product reviews by performing data preprocessing steps such as text cleaning, tokenization, and stop-word removal. It then transforms the processed text into numerical features using the TF-IDF vectorization technique. A supervised machine learning model is trained on this data to classify sentiments into categories such as positive, negative, and neutral.
To enhance usability, the trained model is integrated into a web-based application developed using Flask, allowing users to input reviews and receive real-time sentiment predictions. This system helps businesses understand customer opinions, identify product strengths and weaknesses, and make data-driven decisions.
Overall, the project demonstrates the practical application of AI and NLP in analyzing consumer behavior and improving product strategies in the skincare industry.

I. INTRODUCTION

In recent years, the skincare industry has experienced significant growth due to increased consumer awareness and the widespread use of online platforms for purchasing beauty products. Customers frequently share their experiences and opinions through online reviews, which serve as a valuable source of information for both consumers and companies. However, the large volume of unstructured textual data makes it challenging to manually analyze and extract meaningful insights.
To address this challenge, Artificial Intelligence (AI) and Natural Language Processing (NLP) techniques can be used to automatically process and analyze customer reviews. Sentiment analysis, a key application of NLP, helps in identifying the emotional tone behind textual data by classifying it into categories such as positive, negative, or neutral. This enables businesses to understand customer satisfaction, preferences, and areas for improvement more effectively.
This project focuses on developing an AI-driven skincare product analytics and sentiment analysis system that automates the process of analyzing customer feedback. The system uses machine learning algorithms to classify sentiments from skincare product reviews and provides insights through a user-friendly web interface. By integrating data preprocessing, feature extraction, and model prediction into a single pipeline, the project demonstrates a practical solution for handling real-world text data.
The proposed system not only reduces manual effort but also improves decision-making by providing accurate and timely insights into customer opinions. It can be widely applied in e-commerce platforms, product review systems, and business intelligence tools to enhance customer experience and product quality.

II. LITERATURE REVIEW

The rapid expansion of e-commerce platforms has led to an increase in user-generated content such as product reviews, which play a crucial role in influencing customer decisions. As a result, sentiment analysis has gained significant attention in recent years as a method to automatically interpret and classify opinions expressed in textual data.
Several studies have explored the use of Natural Language Processing (NLP) techniques for sentiment classification. Traditional approaches involve text preprocessing methods such as tokenization, stop-word removal, and stemming, followed by feature extraction techniques like Bag of Words (BoW) and Term Frequency–Inverse Document Frequency (TF-IDF). These methods convert textual data into numerical representations suitable for machine learning models.

Researchers have applied various machine learning algorithms for sentiment analysis, including Naïve Bayes, Logistic Regression, Support Vector Machines (SVM), and Decision Trees. Among these, Naïve Bayes and Logistic Regression are widely preferred due to their simplicity, efficiency, and good performance on text classification tasks. Studies have shown that TF-IDF combined with these algorithms often provides reliable results for medium-sized datasets.

In recent years, deep learning techniques such as Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Transformer-based models have also been used for sentiment analysis, achieving higher accuracy by capturing contextual relationships in text. However, these methods require large datasets and high computational resources, making them less suitable for lightweight applications.

Some research has specifically focused on product review analysis in the skincare and cosmetics domain, where sentiment analysis helps identify customer preferences, product effectiveness, and common issues such as skin reactions or ingredient concerns. These insights assist companies in improving product quality and enhancing customer satisfaction.

Based on the existing literature, this project adopts a machine learning-based approach using TF-IDF and a classification algorithm to build an efficient and scalable sentiment analysis system. Additionally, the integration of a Flask-based web application ensures practical usability by enabling real-time sentiment prediction for user input.

III. PROPOSED METHODOLOGY

This work proposes an intelligent framework for analyzing skincare product reviews using Natural Language Processing (NLP) and supervised machine learning techniques. The overall architecture of the system is designed to transform unstructured textual data into meaningful sentiment insights through a multi-stage processing pipeline.

A. Data Acquisition

The dataset used in this study consists of a large collection of skincare product reviews obtained from publicly available sources. The dataset contains textual feedback along with corresponding sentiment labels, enabling supervised learning. The diversity of the dataset ensures coverage of various user opinions and linguistic patterns.

B. Text Preprocessing

To improve the quality and consistency of the input data, several preprocessing techniques are applied. These include normalization (conversion to lowercase), removal of punctuation and special symbols, tokenization, and elimination of stop words. This step reduces noise and enhances the semantic clarity of the textual data.

C. Feature Representation

The preprocessed text is transformed into numerical feature vectors using the **Term Frequency–Inverse Document Frequency (TF-IDF)** technique. This method effectively captures the importance of words in individual documents relative to the entire corpus, thereby improving the discriminative capability of the model.

D. Model Development

A supervised machine learning model is employed for sentiment classification. The dataset is partitioned into training and testing subsets to ensure unbiased evaluation. Algorithms such as Logistic Regression and Naïve Bayes are considered due to their effectiveness in text classification tasks. The model learns to map input feature vectors to corresponding sentiment labels (positive, negative, or neutral).

E. Performance Evaluation

The performance of the proposed model is evaluated using standard classification metrics, including Accuracy, Precision, Recall, and F1-score. These metrics provide a comprehensive assessment of the model's predictive capability and robustness.

F. System Deployment

To demonstrate practical applicability, the trained model is deployed using a lightweight web framework. The system accepts user input in the form of text reviews, processes the input through the trained pipeline, and generates real-time sentiment predictions. The deployment ensures accessibility and usability in real-world scenarios.

G. Workflow of the Proposed System

The overall workflow of the system is summarized as follows:

1) Input acquisition from user or dataset
2) Text preprocessing and cleaning
3) Feature extraction using TF-IDF
4) Sentiment prediction using trained model
5) Output visualization through web interface

Table I. Dataset Description and Composition

Review_id	Product_name	Skin_type	Rating	Review_text	Review_date	Sentiment
1	Aloe Vera Moisturizer	Dry	4	I have been using this Aloe Vera Moisturizer for a few weeks and my skin feels smoother.	12-07-2021	Positive
2	Aloe Vera Moisturizer	Normal	4	This Aloe Vera Moisturizer feels lightweight and keeps my skin hydrated.	07-05-2023	Neutral
3	Hydrating Face Cream	Dry	4	After regular use this Hydrating Face Cream improved my skin condition.	11-09-2021	Positive

Table I outlines dataset distribution, in such a manner that normal samples are well represented so that models can be trained.

a) Data Preprocessing and Feature Engineering:

Raw skincare product reviews are preprocessed to improve data quality and model performance. The preprocessing steps include lowercasing, removal of punctuation and special characters, tokenization, and elimination of stop words. These steps help in reducing noise and retaining meaningful textual information.

For feature engineering, the cleaned text is transformed into numerical form using the TF-IDF (Term Frequency–Inverse Document Frequency) technique. This method assigns weights to words based on their importance in a document relative to the entire dataset, enabling effective representation of textual data.

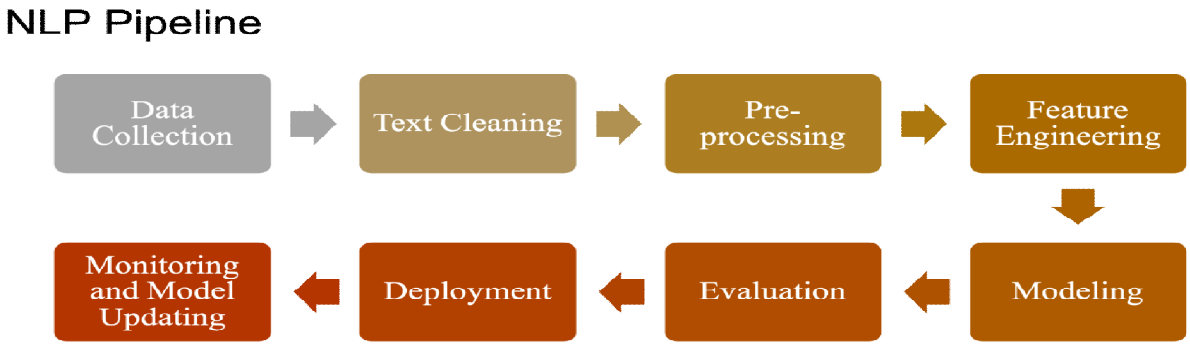


Fig1:Flowchart of the NLP Pipeline

Each review is associated with a sentiment label, categorized into three classes: positive, negative, and neutral. These labels are used to train and evaluate the supervised machine learning model for sentiment classification.
The dataset is primarily composed of the following attributes:

- Review Text: User-generated feedback describing product experience
- Sentiment Label: Classification of reviews into positive, negative, or neutral

Prior to model training, the dataset undergoes preprocessing steps including text normalization, removal of noise (punctuation and special characters), tokenization, and stop-word removal. The processed data is then transformed into numerical feature vectors using the TF-IDF (Term Frequency–Inverse Document Frequency) technique.

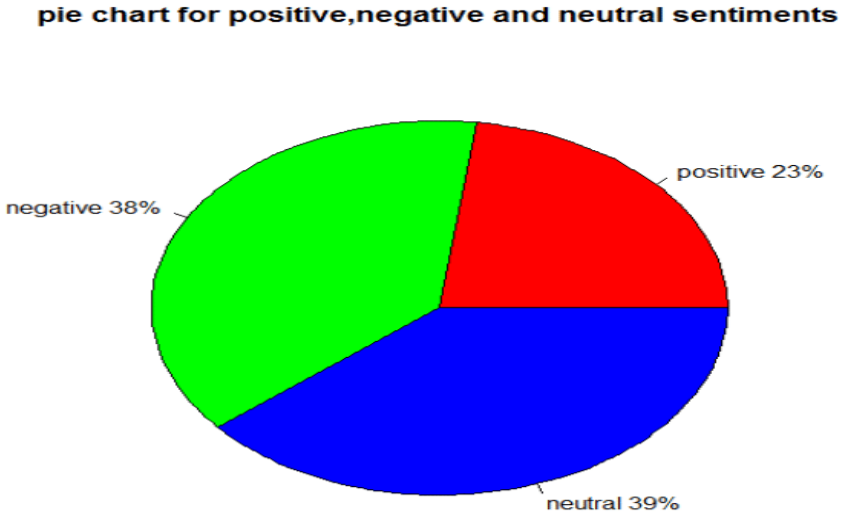


Fig2:Pie Chart ForPostive,Negative and Neutral Sentiments

b) Model Training and Demand Prediction:
In this phase, the preprocessed and feature-engineered data is used to train a supervised machine learning model for sentiment-based demand prediction. The dataset is divided into training and testing sets to ensure proper evaluation. Algorithms such as Logistic Regression and Naïve Bayes are employed due to their effectiveness in text classification tasks.
During training, the model learns the relationship between TF-IDF feature vectors and corresponding sentiment labels. Once trained, the model is used to predict the sentiment of new, unseen reviews, which reflects product demand and customer preference.

c) Web Application and Visualization:
To ensure practical usability, the proposed system is deployed as a web application using the Flask framework. The application provides an interactive interface where users can input skincare product reviews and obtain real-time sentiment predictions generated by the trained machine learning model.
The system integrates all processing stages, including preprocessing, feature extraction using TF-IDF, and sentiment classification, into a unified pipeline. The prediction results are presented through a user-friendly interface, enabling easy interpretation of customer feedback.

Summary
This study presents an AI-driven system for analyzing skincare product reviews using Natural Language Processing and machine learning techniques. The proposed approach includes data preprocessing, feature extraction using TF-IDF, sentiment classification, and demand prediction. Additionally, opportunity scoring and skill gap analysis are performed to identify product strengths and areas for improvement.
The system is further deployed as a web application to provide real-time predictions and visual insights. Overall, the proposed framework offers an efficient and scalable solution for understanding customer feedback and supporting data-driven decision-making in the skincare industry.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

IV. RESULTS AND DISCUSSION

A. *Dataset Overview and Experimental Setup:*

The dataset used in this study consists of approximately 12,000 skincare product reviews collected from online platforms. Each review is labeled into three sentiment classes: positive, negative, and neutral. The data is preprocessed using standard NLP techniques such as text cleaning, tokenization, and stop-word removal, followed by feature extraction using TF-IDF.

For experimental evaluation, the dataset is split into training and testing sets in an 80:20 ratio. A supervised machine learning model is trained on the training data and evaluated on the test data using performance metrics such as accuracy, precision, recall, and F1-score.

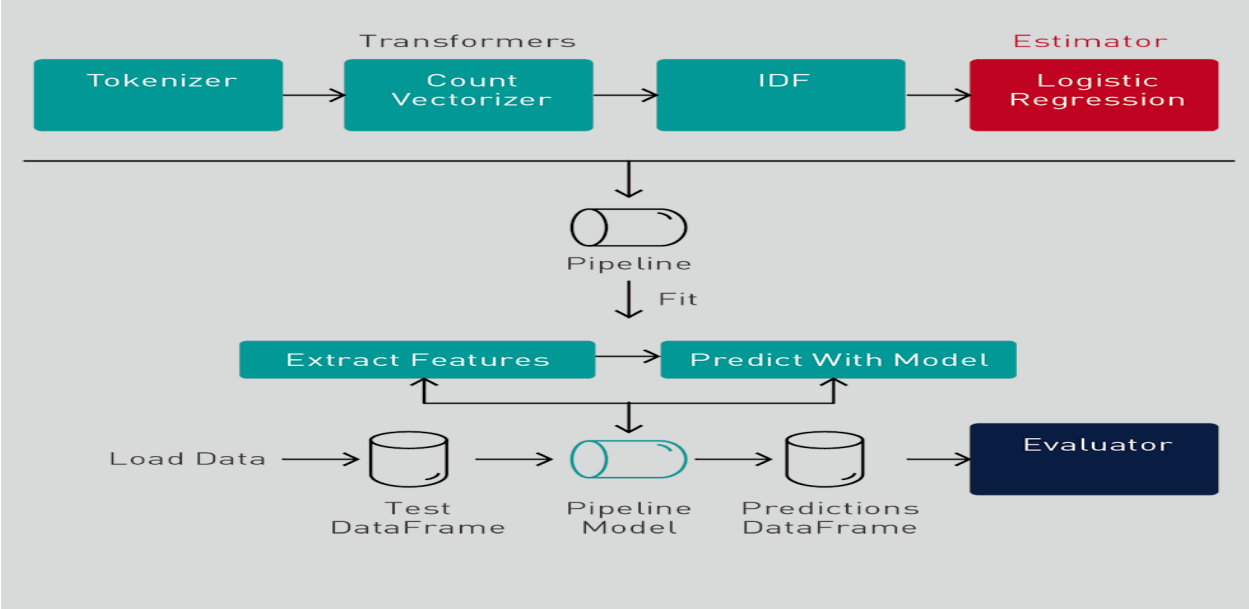


Fig3:Machine Learning Pipeline

B. *Model Performance and Demand Prediction Accuracy:*

The performance of the proposed model is evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. The model achieves an accuracy of 0.88, indicating its effectiveness in correctly classifying sentiment from skincare product reviews.

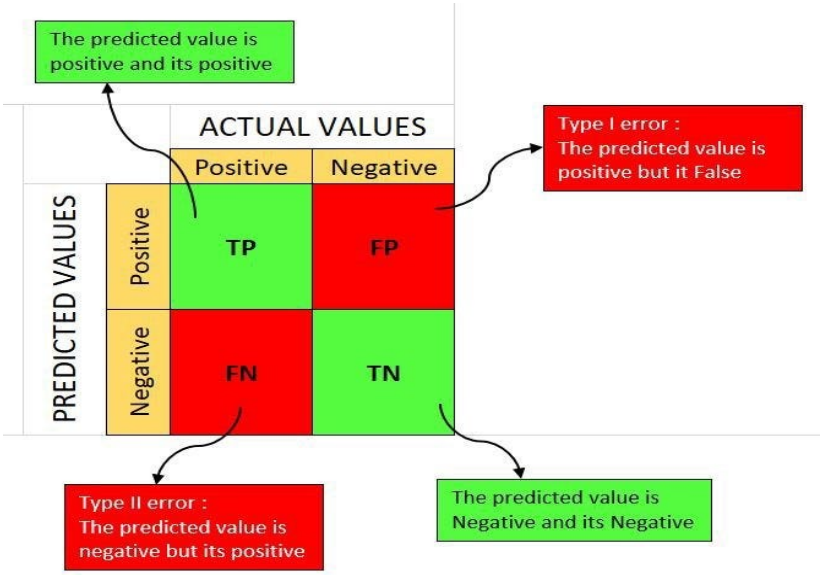


Fig4:Proposed Methodology Flowchart

C. Dashboard Visualization and User Interaction:

The proposed system includes an interactive dashboard that enables users to visualize sentiment analysis results and demand insights effectively. The dashboard presents key information such as sentiment distribution, demand levels, and opportunity scores using graphical representations.

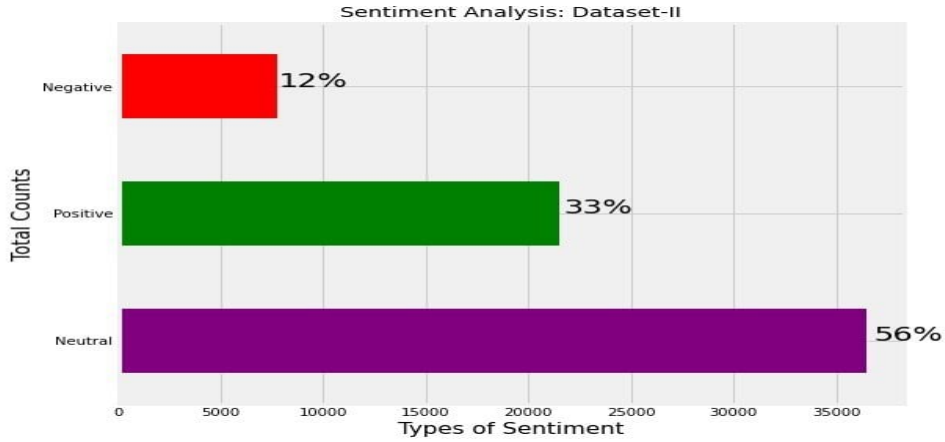


Fig5:Dashboard Visualization

D. Discussion and Observations:

The experimental results demonstrate that the proposed system effectively classifies customer sentiments and provides reliable demand insights. The use of TF-IDF for feature extraction and supervised machine learning algorithms contributes to high classification accuracy and efficient performance.

V. CONCLUSIONS

This study presents an ai-driven system for analyzing skincare product reviews using natural language processing and machine learning techniques. The proposed approach effectively preprocesses textual data, extracts meaningful features using tf-idf, and classifies sentiments with high accuracy.

The integration of demand prediction, opportunity scoring, and skill gap analysis provides deeper insights into customer feedback and product performance. Furthermore, the deployment of a web-based application enables real-time interaction and visualization of results.

Overall, the system demonstrates a scalable and efficient solution for sentiment analysis and decision support, making it valuable for enhancing product quality and customer satisfaction in the skincare industry.

VI. ACKNOWLEDGMENT

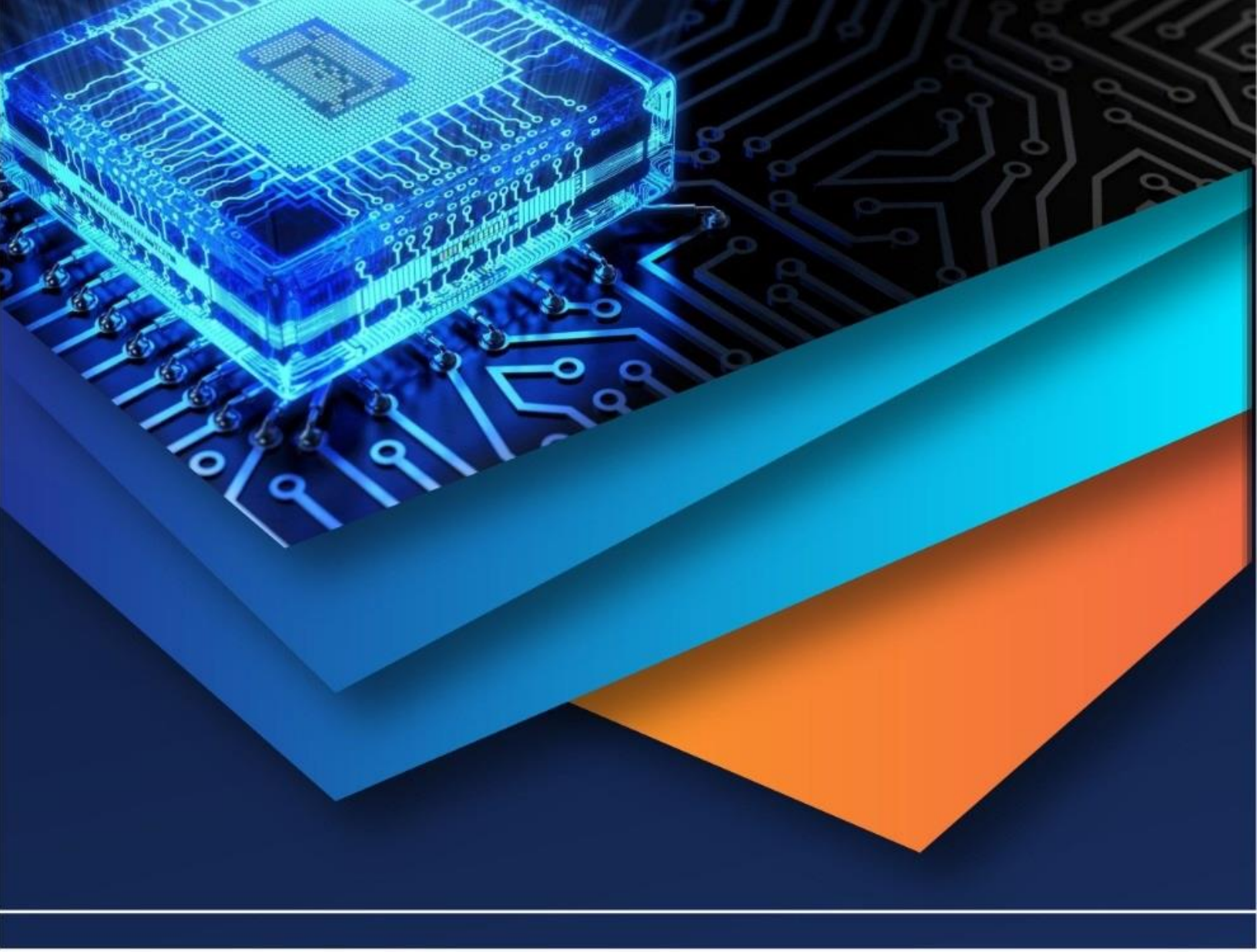
The authors would like to express their sincere gratitude to the faculty and mentors for their continuous guidance and support throughout the development of this project. Their valuable insights and encouragement played a significant role in the successful completion of this work.

We also thank our institution for providing the necessary resources and environment to carry out this research. Finally, we extend our appreciation to all contributors and data sources that supported this.

REFERENCES

[1] B. Liu, Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, 2012.
[2] S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python, O'Reilly Media, 2009.
[3] J. Leskovec, A. Rajaraman, and J. D. Ullman, Mining of Massive Datasets, Cambridge University Press, 2014.
[4] T. Mikolov et al., "Efficient Estimation of Word Representations in Vector Space," arXiv preprint arXiv:1301.3781, 2013.
[5] Y. Goldberg, "A Primer on Neural Network Models for Natural Language Processing," Journal of Artificial Intelligence Research, vol. 57, pp. 345–420, 2016.
[6] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
[7] G. Salton and C. Buckley, "Term-Weighting Approaches in Automatic Text Retrieval," Information Processing & Management, vol. 24, no. 5, pp. 513–523, 1988.
[8] W. McKinney, "Data Structures for Statistical Computing in Python," Proceedings of the 9th Python in Science Conference, 2010.
[9] Flask Documentation, "Flask Web Framework," [Online]. Available: https://flask.palletsprojects.com/.

[10] A.Zhang, Z. Lipton, M. Li, and A. Smola, Dive into Deep Learning, MIT Press, 2021.
[11] A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," Proceedings of the International Conference on Language Resources and Evaluation (LREC), 2010.
[12] P. D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," Proceedings of ACL, 2002.
[13] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs Up? Sentiment Classification using Machine Learning Techniques," Proceedings of EMNLP, 2002.
[14] R. Socher et al., "Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank," Proceedings of EMNLP, 2013.
[15] J. Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proceedings of NAACL-HLT, 2019.
[16] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," Proceedings of KDD, 2016.
[17] K. He et al., "Deep Residual Learning for Image Recognition," Proceedings of CVPR, 2016.
[18] Z. Zhang, J. Zhao, and L. LeCun, "Character-level Convolutional Networks for Text Classification," Advances in Neural Information Processing Systems (NeurIPS), 2015.
[19] D. Jurafsky and J. H. Martin, Speech and Language Processing, 3rd ed., Draft, 2023.
[20] Hugging Face, "Transformers Library Documentation," [Online]. Available: https://huggingface.co/docs
[21] NLTK Documentation, "Natural Language Toolkit," [Online]. Available: https://www.nltk.org/
[22] Pandas Documentation, "Python Data Analysis Library," [Online]. Available: https://pandas.pydata.org/
[23] NumPy Documentation, "Numerical Python," [Online]. Available: https://numpy.org/



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY