International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 2, pp 24-30, February-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.402005
ISSN (Online): 2583-7257

Recommendation System for Marketing with Sentimental Analysis Based on Customer Product Reviews Using ML Algorithms

[1]S.M.P Qubeb, [2]V. Yuvarajachari, [3]B. Fayal Khan, [4]J. Srividya, [5]M. Sameena Begum, [6]S. Venkatesh, [7]G. Vamsidhar Reddy, [8]S. Venkatesh

[1,2,3,4,5,6,7,8]Department of Computer Science Engineering (Artificial Intelligence), Gates Institute of Technology, Gooty, Andhra Pradesh, India

E-mail: [1]qubeb.s@gatesit.ac.in, [2]vadlayuvaraj8247@gmail.com, [3]b.fayazkhan006@gmail.com, [4]jayanthisrividya@gmail.com, [5]msameena08@gmail.com, [6]sandhanavenkatesh8@gmail.com, [7]gaddamvamshi.798@gmail.com, [8]suguruvenkatesh460@gmail.com

***Abstract* - In the modern digital era, online reviews significantly shape consumer opinions and purchasing behavior. E-commerce platforms such as Amazon and Flipkart enable users to share their experiences, offering future buyers valuable insights into product performance. To effectively analyze the vast number of reviews, it is essential to categorize them based on sentiment—positive or negative. This study focuses on applying sentiment analysis techniques to classify over 400,000 mobile phone reviews into two sentiment categories. Machine learning models, including Naïve Bayes, Support Vector Machine (SVM), and Decision Tree, were implemented for this purpose. The models' effectiveness was assessed through 10-fold cross-validation to identify the most accurate classifier.**

Keywords: Data processing, natural language processing (NLP), opinion mining, textual data categorization, artificial intelligence.

I. INTRODUCTION

With the growing demand for smartphones, the mobile phone market is expanding rapidly. As the smartphone industry experiences this boom, there is a need for comprehensive reviews of both brands and specific phone models. Numerous brands compete in the market, with leading names like Samsung and Apple dominating a significant share of the industry.

E-commerce platforms play an essential role in boosting mobile phone sales and shaping consumer purchasing behaviors. Reviews available on these platforms serve as valuable tools, helping consumers make informed decisions when choosing products.

Retail websites such as Amazon offer various options for customers to leave feedback, including numerical ratings (1 to 5 stars) and written comments about the product.

Given the vast number of products from different brands, providing relevant and meaningful reviews is essential. The number of reviews linked to a product or brand has been rising at an exponential rate, creating challenges akin to managing big data. By classifying reviews based on customer sentiment—positive or negative—it becomes easier to understand the sentiment of each review, leading to better purchasing decisions.

Organizing reviews by sentiment helps potential buyers distinguish between positive and negative feedback, aiding them in making more informed choices based on their needs. This process increases the credibility of reviews, helping users better assess the smartphones' features and performance.

In this research, mobile phone reviews were collected from Amazon.com as unstructured data. The data was then filtered to eliminate irrelevant content and pre-processed for sentiment analysis using supervised learning techniques. The reviews were classified using various machine learning models, including Naïve Bayes, Support Vector Machine (SVM), and Decision Tree. These models were cross-validated to identify the most effective classifier for sentiment analysis.

II. RELATED WORK

Data analytics has become an essential tool in discovering hidden trends and patterns within vast datasets. The concept of Big Data is primarily characterized by the three Vs—volume, velocity, and variety. Additionally, veracity and value have emerged as crucial attributes that define the effectiveness and impact of Big Data. The enormous amount and rapid pace at which data is generated

International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 2, pp 24-30, February-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.402005
ISSN (Online): 2583-7257

daily often exceed the processing abilities of many IT infrastructures. Platforms like e-commerce websites gather extensive and varied product reviews, which, if analyzed effectively, can offer valuable insights into consumer behavior and support decision-making processes. These reviews may exist in structured or unstructured formats, requiring advanced techniques to extract meaningful business intelligence while filtering out irrelevant or noisy data.

Leveraging Big Data allows businesses to transition from intuition-based strategies to evidence-driven decision-making. It facilitates several operations, including but not limited to: identifying marketing opportunities, segmenting customers more efficiently, improving social media targeting, forecasting demand, recognizing fraudulent patterns, and gaining an in-depth understanding of consumer preferences and actions.

Sentiment analysis is a subfield of data analysis that aims to determine the emotional tone embedded in textual content. It can categorize sentiments as positive, negative, or neutral, and is commonly implemented at three analytical levels: the document level, sentence level, and phrase level. Several studies have pre-classified words and phrases with sentiment polarity to streamline the analysis. While this is helpful, the context of a word often changes its sentiment, a phenomenon known as contextual polarity. In many cases, the same word may convey different emotions based on its surrounding context. To overcome such ambiguity, modern approaches incorporate subjective detection techniques, which allow for accurate sentiment interpretation without losing essential information.

There has been extensive research in this domain, covering various datasets such as movie reviews, tweets, and product feedback. Notably, sentiment analysis has expanded to support multilingual applications, moving beyond English. For example, datasets collected from Twitter in different languages have been used to train sentiment classifiers that can detect emotional tones across cultures. These classifiers are designed to accurately identify positive, negative, or neutral opinions.

A number of advanced machine learning algorithms have been utilized to enhance sentiment classification. For instance, a support vector machine (SVM) model was developed that takes into account sarcasm, grammatical inconsistencies, and spam detection. Another study introduced an enhanced Naïve Bayes model by integrating features such as negation handling, n-grams, and effective feature selection techniques. Furthermore, sentiment classification for Chinese text was achieved using a combination of five classifiers—including Naïve Bayes, SVM, Centroid, Window, and K-nearest Neighbor (KNN)—alongside four distinct feature selection methods. Among these, SVM consistently demonstrated superior performance. Similar outcomes were observed in experiments involving travel review datasets, where character-based N-gram models and SVMs outperformed traditional Naïve Bayes algorithms.

Beyond e-commerce and social media platforms, the utility of sentiment analysis extends to stock market predictions, news analytics, and even political discourse evaluation. By applying rule-based sentiment analysis, businesses can identify and respond to customer dissatisfaction effectively. For example, when a product receives negative feedback, the system can suggest competitive alternatives to users. This approach not only improves customer experience but also increases brand competitiveness.

Sentiment analysis is also used in targeted advertising. By understanding a user's sentiment preferences and behavior, advertisements tailored to individual interests can be displayed. A notable development in this area is the Blogger-Centric Contextual Advertising Framework, which analyzes bloggers' content to match and display relevant ads that align with their personal interests and writing patterns.

The structure of this paper is organized as follows: Section 3 presents the dataset used and the methodology adopted for analysis. Section 4 focuses on sentiment detection and classification. Section 5 explains the machine learning classifiers implemented for the task. Section 6 highlights the data balancing techniques used, followed by Section 7, which discusses experimental outcomes and the cross-validation of model accuracy. Finally, Section 8 concludes the research and outlines possible directions for future work.

III. FRAMEWORK

The proposed framework of the research work is conducted in three different modules as shown in Fig. 1.

A. Dataset and its Features

The first module includes data collection and pre-processing of data. A large sample of online reviews is collected from the e-commerce giant Amazon.com. The data set consists of over 400,000 reviews for approximately 4500 mobile phones. It includes six features as explained in table 1.

B. Approach

The approach followed by the proposed framework is described in Fig. 1. Initially, the experimental data is collected from an e-commerce website Amazon.com. Each data set is in

International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 2, pp 24-30, February-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.402005
ISSN (Online): 2583-7257

the Comma Separated Values (CSV) file format and available as supplement. In the second step, data are pre-processed to remove stop words, punctuation marks, whitespaces, digits and special symbols. 'tm' package [19] is employed for text mining. In the third step, feature selection is performed to extract relevant features from the data set. In the given data set out of the six features, only three features, i.e., Product Name, Brand Name and Reviews have been considered. In the fourth step, sentiment orientation of the reviews is determined. In the fifth step, 'Pos/Neg' tags are appended to the dataset to corresponding to each review to conduct supervised learning. The sixth step involves training and testing the classified data using Naïve Bayes, SVM and Decision Tree models. The accuracy so obtained is validated using 10-fold cross validation.

IV. SENTIMENT ANALYSIS

An internal package called 'Syuzhet' [20] has been utilized to carry out Sentiment Analysis. This package includes three distinct sentiment lexicons. One of them, the NRC sentiment lexicon, is employed to identify eight core emotions and their respective intensities within the textual data derived from user reviews. The ten emotional categories extracted through this method include anger, anticipation, disgust, fear, joy, sadness, surprise, trust, positivity, and negativity. This approach helps in understanding the emotional tone conveyed in the reviews more comprehensively.

Sentiment analysis, when combined with emotion detection frameworks like Syuzhet, enables researchers and businesses to extract actionable insights from unstructured textual content such as customer reviews, social media posts, and feedback surveys.

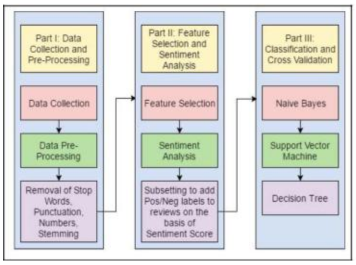
Table 1: Features Included In the Data Set

Feature	Description
Product Name	Model name of mobile phone
Brand Name	Manufacturing brand
Price	Price of the mobile in dollars
Rating	User rating between 1 to 5
Reviews	User reviews provided for every mobile phone
Review Votes	Number of people who found the review helpful

V. PROPOSED SYSTEM

To overcome the limitations of traditional recommendation systems, this project introduces a Knowledge-Based Recommendation System that integrates Sentiment Analysis and Deep Learning to enhance the relevance and accuracy of recommendations. Instead of relying solely on past user interactions or predefined item attributes, the proposed system analyzes customer reviews and feedback to extract sentiment-based insights, ensuring that recommendations align with user preferences and emotions. The proposed system aims to develop a Knowledge-Based Recommendation System that enhances traditional recommendation methods by integrating Sentiment Analysis and Deep Learning. Unlike conventional recommendation systems that rely solely on user behavior and item features, this system incorporates textual reviews and sentiment insights to provide more personalized and accurate recommendations.

The architecture of the proposed system involves several key components including data preprocessing, sentiment classification, feature extraction, and recommendation generation. Initially, customer reviews are collected and preprocessed to remove noise and irrelevant content. Sentiment analysis is then performed using machine learning and deep learning models to determine the emotional polarity of each review.



text1	positive	negative	polarity
I feel so LUCKY to have found this used (p	5	1	4
nice phone, nice up grade from my panta	2	0	2
Very pleased	1	0	1
It works good but it goes slow sometimes	2	0	2
Great phone to replace my lost phone. Th	2	1	1
I already had a phone with problems... I k	3	2	1
The charging port was loose. I got that sol	1	1	0
Phone looks good but wouldn't stay charg	2	2	0
I originally was using the Samsung S2 Gal	3	1	2
It's battery life is great. It's very responsiv	3	2	1
My fiance had this phone previously, but	1	0	1
This is a great product it came after two d	2	1	1
These guys are the best! I had a little situ	1	0	1
I'm really disappointed about my phone a	2	1	1
Ordered this phone as a replacement for	4	0	4

In this research, only positive and negative sentiment orientation has been considered for classification of reviews. Figure 2 depicts the percentage of positive and negative reviews present in the dataset using bar graph

VI. ADVANTAGES OF PROPOSED SYSTEM

Improved Recommendation Accuracy – Integrating Sentiment Analysis and Deep Learning ensures that recommendations are more relevant and context-aware.

Emotion-Aware Suggestions – Unlike traditional systems, this approach considers user opinions and emotions from reviews, leading to more personalized recommendations.

Solves the Cold-Start Problem – The system can recommend items even for new users or new products by analyzing textual reviews instead of relying solely on historical data.

Hybrid Approach for Better Results – Combines collaborative filtering, content-based filtering, and sentiment analysis, improving the diversity and quality of recommendations.

International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 2, pp 24-30, February-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.402005
ISSN (Online): 2583-7257

Real-Time Adaptability – The system continuously learns from new reviews and feedback, making it dynamic and adaptive to changing user preferences.

Enhanced User Satisfaction – By understanding customer sentiments, the system suggests products that align better with user expectations, leading to higher engagement and retention.

Enhanced User Satisfaction – By understanding customer sentiments, the system suggests products that align better with user expectations, leading to higher engagement and retention.

Scalability and Efficiency – Deep learning techniques allow the system to handle large datasets efficiently, making it suitable for real-world applications.

VII. CLASSIFICATION

Classification is the process of classifying reviews on the basis of their sentiment into two classes: Positive and Negative. The sentiment score established for each positive and negative polarity using NRC sentiment dictionary is then added to the reviews dataset. This individual score is used to calculate the overall polarity as given by Eq. 1 of the sentiment as shown in Fig. 3.

Polarity = Positive Score í Negative Score

After calculating the polarity corresponding to each review, different subsets of positive and negative sentiment are formed having polarity 0 to 10 and -1 to -10 respectively. The subsets so obtained are combined together in a CSV file along with the Pos/Neg tags. The classified data is shown in Fig. 4.

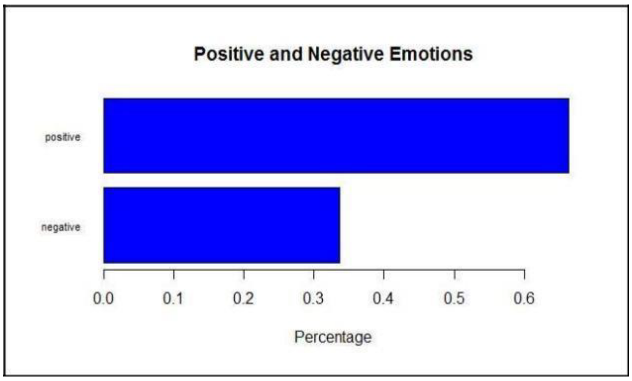


Fig. 3: Overall Polarity of the Review

A. Classifiers

The classification models selected for categorization of text are: Naïve Bayesian, Support Vector Machine and Decision Tree.

1) Naïve Bayesian Classifier

A statistical classifier that maps input feature vectors to output class labels [21]. For a set of training data D, each row is represented by an n-dimensional feature vector, X =x1, x2, …, xn.

There are K classes, K1, K2, ..., Km in the output class label. For every tuple X, the classifier will predict 2 as given by Eq. 2 that X belongs to Ki if and only if: P (Ki|X) > P (Kj |X), where i, j Ì [1,m] and i j.

P(Ki|X) = Ên P(xk|Ki)

2) Support Vector Machine (SVM)

SVM is used for a labelled training data that categorizes testing dataset using an optimal hyperplane [22]. A hyperplane is separates data of one class from another which is defined as given in Eq. 3.

3) Decision Tree

A hierarchical tree structure encompassing decision nodes for representing attributes and edges for denoting attribute values. This representation in the form of a tree allows to construct decision rules that classify new instances of the data [23].

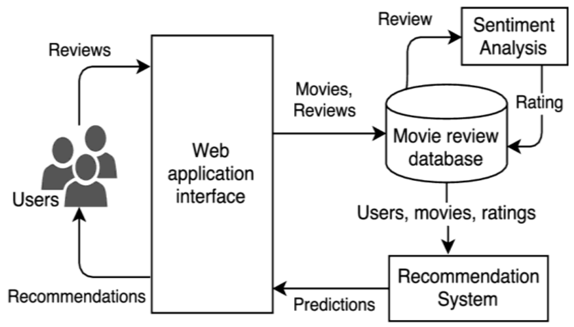
VIII. DATA BALANCING

Figure 2 represents that the number of positive reviews is more than double the negative reviews. This implies that the data are imbalanced as the target variable has imbalanced proportion of classes. Therefore, running machine learning models for classification would yield biased predictions and misleading accuracies. To avoid such scenarios, data balancing is employed. There are different methods that can be used to transform imbalanced data into balanced data like under sampling and oversampling [26].

For treating imbalanced data, under sampling technique has been used. Under sampling means to reduce the number of observations from majority class to balance the data set. The balanced data so obtained has almost equal number of positive and negative reviews.

International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 2, pp 24-30, February-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.402005
ISSN (Online): 2583-7257

IX. ARCHITECTURE



System architecture is a crucial aspect of designing and implementing a recommendation system, as it defines how various components interact to deliver personalized suggestions efficiently. The architecture of a knowledge-based recommendation system involves several stages, including data collection, preprocessing, feature extraction, model processing, and user interaction. Each of these stages plays a significant role in ensuring accurate and effective recommendations for users.

X. MODEL EVALUATION AND RESULTS

After appending the data with a class having positive or negative tags and removing imbalanced proportion of classes, a random sample of 3000 reviews is taken to train and test the dataset on three classifiers.

A. Cross Validation

Cross Validation is a model evaluation parameter that demonstrates the ability of the system to make new predictions accurately. In the proposed work, K- fold cross validation has been implemented to determine the efficiency of the models. In K-fold cross validation, the dataset is divided into k subsets which is repeated k times. For every iteration, k subset is used as the training sample and k-1 subsets are used for testing. The three models are cross validated 10 times. Table 2 shows the cross validation of the three models for ten runs. It is observed that the accuracy of all the three models in all the iterations varies in the range of ±10.

The scatter plot in Fig. 5 elucidates that the SVM model reaches the highest accuracy mark of 81.75 among all the models for a number of iterations. Naïve Bayes model has the lowest accuracy of 64.57 among the three models. The graph clearly depicts that SVM model has the best accuracy out of the three models and Naïve Bayes model has the least predictive accuracy.

Table 2: Cross Validation

Runs	Naïve Bayes	SVM	Decision Tree
1	67.11	80.12	74.31
2	64.57	79.44	77.13
3	67.57	78.96	70.80
4	66.77	82.25	78.00
5	67.57	80.52	72.63
6	64.57	78.92	76.22
7	67.77	77.70	71.84
8	66.71	79.72	77.62
9	68.31	78.00	67.68
10	68.57	81.75	81.25

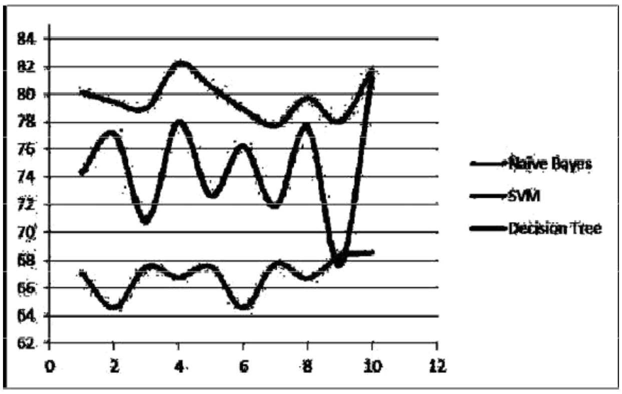


Fig. 5: Scatter Plot

B. Performance Comparison

After appending the data with a class having positive or negative tags and removing imbalanced proportion of classes, a random sample of 3000 reviews is taken to train and test the dataset on three classifiers.

Table 3: Predictive Accuracy of Models

Model Name	Accuracy
Naïve Bayes	66.95
SVM	81.77
Decision Tree	74.75

XI. CONCLUSION AND FUTURE SCOPE

An evolutionary shift from offline markets to digital markets has increased the dependency of customers on online reviews to a great extent. Online reviews have become a platform for building trust and influencing consumer buying patterns. With such dependency there is a need to handle such large volume of reviews and present credible reviews before the consumer. Our research is aiming to achieve this by conducting sentiment analysis of mobile phone reviews and classifying the reviews into positive and negative sentiment.

Article DOI: https://doi.org/10.47001/ICJES/2025.402005
ISSN (Online): 2583-7257

After balancing the data with almost equal ratio of positive and negative reviews, three classification models have been used to classify reviews. Out of the three classifiers, i.e., Naïve Bayes, SVM and Decision Tree, predictive accuracy of SVM is found to be the best. The accuracy results have been cross validated and the highest value of accuracy achieved was 81.75% for SVM among the three models.

In future, the work can be extended to perform multi-class classification of reviews which will provide delineated nature of review to the consumer, hence better judgement of the product. It can also be used to predict rating of a product from the review. This will provide users with reliable rating because sometimes the rating received by the product and the sentiment of the review do not provide justice to each other. The proposed extension of work will be very beneficial for the e-commerce industry as it will augment user satisfaction and trust.

REFERENCES

[1] S. Erevelles, N. Fukawa, and L. Swayne, “Big data consumer analytics and the transformation of marketing,” Journal of Business Research, vol. 69, no. 2, pp. 897–904, 2022.

[2] P. Russom et al., “Big data analytics,” TDWI best practices report, fourth quarter, pp. 1–35, 2022.

[3] S. Erevelles, N. Fukawa, and L. Swayne, “Big data consumer analytics and the transformation of marketing,” Journal of Business Research, vol. 69, no. 2, pp. 897–904, 2022.

[4] V. Hatzivassiloglou and K. R. McKeown, “Predicting the se- mantic orientation of adjectives,” in Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 2021, pp. 174–181.

[5] T. Wilson, J. Wiebe, and P. Hoffmann, “Recognizing contextual polarity in phrase-level sentiment analysis,” in Proceedings of the conference on human language technology and empirical methods in natural language processing. Association for Computational Linguistics, 2021, pp. 347–354.

[6] B. Pang and L. Lee, “A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts,” in Proceedings of the 42nd annual meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2022, p. 271.

[7] A.Pak and P. Paroubek, “Twitter as a corpus for sentiment analysis and opinion mining.” in LREc, vol. 10, no. 2023, 2023.

[8] M. WAHYUDI and D. A. KRISTIYANTI, “Sentiment analysis of smartphone product review using support vector machine algorithm-based particle swarm optimization.” Journal of The- oretical & Applied Information Technology, vol. 91, no. 1, 2023.

[9] D. N. Devi, C. K. Kumar, and S. Prasad, “A feature based approach for sentiment analysis by using support vector ma- chine,” in Advanced Computing (IACC), 2023 IEEE 6th International Conference on. IEEE, 2016, pp. 3–8.

[10] V. Narayanan, I. Arora, and A. Bhatia, “Fast and accurate sentiment classification using an enhanced naive bayes model,” in International Conference on Intelligent Data Engineering and Automated Learning. Springer, 2023, pp. 194–201.

[11] S. Tan and J. Zhang, “An empirical study of sentiment analysis for chinese documents,” Expert Systems with applications, vol. 34, no. 4, pp. 2622–2629, 2023.

[12] Q. Ye, Z. Zhang, and R. Law, “Sentiment classification of online reviews to travel destinations by supervised machine learning approaches,” Expert Systems with Applications, vol. 36, no. 3, pp. 6527–6535, 2023.

[13] L.-C. Yu, J.-L. Wu, P.-C. Chang, and H.-S. Chu, “Using a contextual entropy model to expand emotion words and their intensity for the sentiment classification of stock market news,” Knowledge-Based Systems, vol. 41, pp. 89–97, 2023.

[14] M. Hagenau, M. Liebmann, and D. Neumann, “Automated news reading: Stock price prediction based on financial news using context-capturing features,” Decision Support Systems, vol. 55, no. 3, pp. 685–697, 2023.

[15] T. Xu, Q. Peng, and Y. Cheng, “Identifying the semantic orientation of terms using s-hal for sentiment analysis,” Knowledge- Based Systems, vol. 35, pp. 279–289, 2023.

[16] I.Maks and P. Vossen, “A lexicon model for deep sentiment analysis and opinion mining applications,” Decision Support Systems, vol. 53, no. 4, pp. 680–688, 2023.

[17] G. Qiu, X. He, F. Zhang, Y. Shi, J. Bu, and C. Chen, “Dasa: dissatisfaction-oriented advertising based on sentiment analysis,” Expert Systems with Applications, vol. 37, no. 9, pp. 6182– 6191, 2010.

[18] T.-K. Fan and C.-H. Chang, “Blogger-centric contextual ad- vertising,” Expert systems with applications, vol. 38, no. 3, pp. 1777–1788, 2021.

[19] I.Feinerer, “Introduction to the tm package text mining in r,” 2015.

[20] M. Jockers, “Package âAŸsyuzhetâAZ,” 2022.

International Current Journal of Engineering and Science (ICJES)
Volume 4, Issue 2, pp 24-30, February-2025
Article DOI: https://doi.org/10.47001/ICJES/2025.402005
ISSN (Online): 2583-7257

[21] K. M. Leung, "Naive bayesian classifier," Polytechnic University Department of Computer Science/Finance and Risk Engineering, 2020.

[22] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," Journal of machine learning research, vol. 2, no. Nov, pp. 45–66, 2021.

[23] I.Jenhani, N. B. Amor, and Z. Elouedi, "Decision Trees as possibilistic classifiers," International Journal of Approximate Reasoning, vol. 48, no.3, pp. 784-807, 2022.

Citation of this Article:

S.M.P Qubeb, V. Yuvarajachari, B. Fayal Khan, J. Srividya, M. Sameena Begum, S. Venkatesh, G. Vamsidhar Reddy, & S. Venkatesh. (2025). Recommendation System for Marketing with Sentimental Analysis Based on Customer Product Reviews Using ML Algorithms. *International Current Journal of Engineering and Science - ICJES*, 4(2), 24-30. Article DOI: https://doi.org/10.47001/ICJES/2025.402005

****