PERFORMANCE ANALYSIS OF FAKE REVIEW DETECTION FOR E-COMMERCE USING MACHINE LEARNING MODELS

# Gubbala Nagendra Kumar1, Penumalle Jaya Naga Rajeswari 2, Bheemala Naga Durga Rama Kumari 3, Ganga Bhavani Billa4

*1,2,3*

*Department of Artificial Intelligence and Machine Learning, Bonam Venkata Chalamayya Engineering College, India*

*4Department of Computer Science And Machine Learning, Bonam Venkata Chalamayya Engineering College, India*

***Abstract***

The use of sentiment analysis has grown significantly in recent years, with applications ranging from product reviews and customer feedback analysis to social media monitoring. Using Support Vector Machine (SVM) and Logistic Regression for sentiment classification, this study explores the effectiveness and efficiency of these machine learning techniques in determining the polarity of text data. The ability of the straightforward but effective probabilistic classifier logistic regression to model binary sentiment outcomes based on input features is investigated. In contrast, the ability of SVM, a strong discriminative classifier, to create a hyperplane that maximizes the margin between the positive and negative sentiment classes is assessed. Using a variety of feature extraction methods, including TF-IDF and word embeddings, this study attempts to compare how well these models perform across number of sentiment datasets. The findings show that both models can effectively classify sentiment despite using different methodologies, with SVM performing better in high-dimensional spaces. This study examines the trade-offs between model complexity and interpretability and demonstrates the models potential for use in practical sentiment analysis tasks.

***Keywords: Sentiment analysis, Logistic Regression, Support Vector Machine, SVM, binary classification, text classification, machine learning, TF-IDF, word embeddings, feature extraction, polarity detection, sentiment polarity, model comparison, high-dimensional spaces, machine learning techniques.***

# INTRODUCTION

# The computational process of locating and obtaining subjective information from text data is referred to as sentiment analysis, or opinion mining. Since social media and online platforms have grown so quickly, sentiment analysis has become a crucial tool for companies and organizations to comprehend customer feedback, opinions, and satisfaction levels. Product reviews, social media posts, and survey responses are examples of text data that must be categorized into three groups: positive, negative, and neutral. We investigate two well-known machine learning techniques for sentiment analysis in this project: Support Vector Machine (SVM) and Logistic Regression. A linear model for binary classification tasks, logistic regression offers probabilistic predictions that aid in determining the probability that a sentiment is positive or negative. However, SVM is a potent classification method that can handle high-dimensional feature spaces because it creates the best hyperplane to divide data points of various classes with the greatest margin

This entails preprocessing and turning unprocessed text data into numerical features through the use of word embeddings and Term Frequency-Inverse Document Frequency (TF-IDF) techniques. In order to highlight the advantages and disadvantages of each strategy in the context of sentiment analysis, this study compares the performance of these models in terms of accuracy, precision, recall, and F1 score. In order to better understand sentiment analysis methodologies, the project will investigate how well these algorithms can classify sentiments from textual data. This will provide insights into the trade-offs between machine learning model complexity (SVM) and simplicity (Logistic Regression). Furthermore, social media sentiment monitoring, customer review analysis, and product development strategy improvement based on customer feedback are just a few examples of how the findings can be applied to real-world situations. To improve comprehension of the subject, the representation in figure 1 below provides a visual representation of the ideas discussed.



Fig-1: Fake Review Detection

Hajek P. et al. [1] suggest neural network models that use emotion indicators, word embeddings, and n-grams to detect fake reviews. The models achieve strong performance across multiple domains, outperforming state-of-the-art methods. Ahmed H et al. [2] proposes an n-gram model for detecting fake reviews and fake news using machine learning. Experimental results show improved performance over existing methods. Kauffmann et al. [3] introduces the Fake Review Detection Framework (FRDF), which uses NLP and sentiment analysis to detect fake reviews and rate brands. Tested on Amazon reviews, it helps consumers and managers make informed decisions by ranking products based on sentiment, price, and ratings. Chengai Sun et al. [4] proposes a “convolutional neural network model for fake review detection”, focusing on product-related review features. Two effective classifiers are integrated with the neural network in a bagging model to improve performance. Results from tests conducted on actual Amazon reviews show how successful the strategy is. Petr Hajek et al. [5] “introduces two neural network models that use n-grams, word embeddings, and emotion indicators to detect fake reviews”. The models perform well across different datasets and product categories, outperforming existing method. Jane Crystal Rodrigues et al. [6] uses sentiment analysis and neural networks to detect fake reviews. It compares activation

functions to find the best model for the task. Prof. Amol Gadewar et al. [7] proposes a hybrid AI system that uses NLP and sentiment analysis to detect fake reviews and recommend genuine ones. It calculates overall product ratings by analyzing user opinions on key product attributes. Paweł Gryka et al. [8] presents a case study on detecting “fake reviews in Google Maps in Poland, using a dataset of 18,000 reviews”. It achieved high accuracy in detecting fake accounts and reviews, helping advance fake review detection in other languages. A. Mutemi et al. [9] through a systematic review employing the PRISMA methodology, this study examines “e- commerce fraud detection using machine learning techniques”, revealing research gaps, new trends, and the expanding role of artificial neural networks. Huy Le et al. [10] discusses the significant issue of fake reviews with the rise of e-commerce, deceiving consumers and influencing their decisions to buy. Combining sentiment analysis, clustering, and machine learning models, their research demonstrates that the Random Forest algorithm is the best at spotting fraudulent reviews.

D. Jain et al. [11] creates a false review filtering system using web scraping, natural language processing, and machine learning with an accuracy of 89.12%, improving review authenticity and aiding customer decision-making. Tanveer Sajid et al. [12] examines fraudulent review detection techniques on the Yelp and Flipkart datasets using Naïve Bayes and Random Forest classifiers, emphasizing difficulties, performance comparisons, and year-over-year changes. Nidhi A. Patel et al. [13] and Saleh Nagi Alsubari et al. [14] discuss the rapid internet development leading to an increase in reviews, thus necessitating detection techniques using unsupervised, supervised, and semi-supervised methodologies. Elmogy et al. [15] suggests a machine learning method for detecting fraudulent reviews in e- commerce systems, evaluating performance with different classifiers and extracting reviewer behaviors via feature engineering.

Cortes et al. [16] and Elshrif Elmurngi et al. [17] describe A new learning tool for two-group classification problems is the support-vector network, which builds a linear decision surface by applying non-linear input vector mapping to a high-dimensional feature space. In order to group movie reviews according to their positive or negative polarity, this study uses sentiment analysis (SA) and machine learning algorithms. The SVM algorithm performs better than other algorithms in sentiment classification and fake review detection using two datasets. Aaryan Rustagi et al. [18] suggest that supervised datasets can be used to train machine learning and deep learning models to differentiate between genuine and fake reviews, improving customer shopping experiences on e- commerce platforms. Qaiser et al. [19] explore the numerical statistic known as TF-IDF, which uses keywords to identify and classify documents. Peng et al. [20] provide guidelines for applying and reporting logistic regression techniques, highlighting key results, assumptions, and recommended formats for comprehensive analysis.

The following is the article's outline: Tools for sentiment analysis and the identification of fraudulent reviews are covered in Section 2, along with data preparation, feature extraction, and machine learning models (SVM, Logistic Regression). In Section 3, results are presented by comparing models based on accuracy, precision, recall, and F1-score, along with information from confusion matrices and visualizations. Section 4: Summarizes key findings, implications, and future research on feature selection, dataset expansion, and model optimization

# METHOD

# Creating a sentiment analysis model that can recognize phony reviews from the text content is the primary goal of this research. In order to forecast the legitimacy of reviews, the model makes use of various preprocessing methods and machine learning algorithms.

.

### Tools and Libraries

This project utilizes a variety of tools and libraries to process data and build the model:

* **Python**: The main programming language used for data manipulation and machine learning tasks.
* **NLTK (Natural Language Toolkit)**: A package for text preparation operations including tokenization, stop word removal, stemming, and lemmatization is called NLTK (Natural Language Toolkit).
* **pandas**: A powerful data manipulation library used to handle the dataset.
* **NumPy**: Used for numerical computations in data processing.
* **Scikit-learn**: Provides machine learning algorithms for building the classification model.
* **Joblib** and **Pickle**: Used for saving and loading the trained models.

### Flowchart of Model Functionality

A structured method for processing text data and generating predictions is represented by the sentiment analysis flowchart that uses Support Vector Machines (SVM) and Logistic Regression.It ensures a systematic and efficient approach to sentiment classification, improving the model’s reliability and accuracy. The below figure 2 illustrates the flow of model.

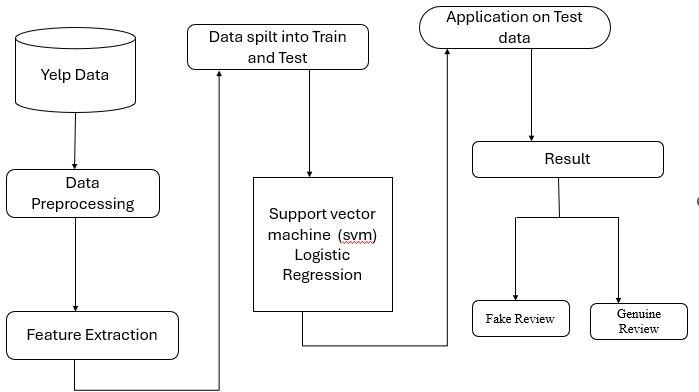


Fig-2: Model Functionality

### Data Preprocessing

Preprocessing is necessary to transform the raw text data into a machine learning-compatible format before the model is trained. This involves several key steps:

* **Tokenization**: splitting up a text into its constituent words or tokens. As a result, the text can be examined more thoroughly.
* **Stop words Removal**: Words that we wish to remove or filter out prior to classifier training are known as stop words.

Common words like "the," "is," "in," etc., “that do not carry significant meaning in the context of sentiment analysis are removed to reduce noise in the data” [17].

* **Stemming:** To make sure that similar words are treated the same, words are reduced to their base or root form. Stemming is a popular algorithm because of its accuracy in simplifying classification by reducing words to their root form. [2].
* **Lemmatization**: Similar to stemming but more advanced. The lemmatization method converts plural format to singular “by removing inflectional endings and returning the base or dictionary form of the word”, like "plays" to "play" It converts words to their dictionary form[15].

### Feature Extraction

“The different features are used to identify fake and genuine reviews like linguistic feature, sentiment score, relational feature, etc” [13]. “Fake information can be used for various purposes, from attracting new customers and shaping people’s views, through fake news”.[8]. A key component of sentiment analysis for the detection of fake reviews is feature extraction, which transforms unprocessed text into numerical features that machine learning models can understand and use. A term's TF value indicates how frequently it appears in a document TF is used to measure how many times a term is present in a document. One widely used technique for feature extraction “is TF-IDF (Term Frequency-Inverse Document Frequency), which evaluates the importance of a word in a document relative to its occurrence across the entire corpus” [19]. Moreover, the statistical method known as TF-IDF is employed to quantify the significance of a term or word to a document within the dataset. The “TFIDF Vectorizer in Scikit-learn is used to transform the raw text into a sparse matrix, where each row represents a document, and each column corresponds to a unique word from the corpus [16]. [Before computing the Term Frequency- Inverse Document Frequency (TF-IDF)”, it is necessary to convert the tuples in POS arrays into a single string [10]. The text is transformed into a matrix of token counts using Count Vectorizer, another popular feature extraction method. Unlike TF-IDF, this method does not take into account the importance of words within the entire corpus, but instead simply counts the frequency of each word in each document. These two strategies are “genre identification, detection of psycholinguistic deception, and text categorization “[4]. This approach is faster than TF-IDF but may not capture the relative importance of words effectively, especially for large datasets where certain words appear frequently across all documents. “Many researchers have proposed an improved form of TF-IDF algorithm known as Adaptive TF-IDF”[19]. To convert the output of Count Vectorizer into TF-IDF scores, the TF-IDF Transformer can be applied. The transformer adjusts the raw counts produced by the Count Vectorizer to assign higher weights to less common but more informative words. By transforming the count matrix, this method helps capture the importance of word in a more nuanced manner, improving the quality of the feature set for training machine learning models. In addition to extracting features, it's useful to visualize the frequency distribution of terms within the corpus. The Frequency Distribution Visualizer from Yellow brick helps accomplish this by displaying the frequency distribution of words across the entire corpus. This visualization aids in identifying the most frequent words, which can be useful for feature selection, understanding model behavior, and detecting potential biases in the dataset

### Model Evaluation for SVM and Logistic Regression

* + 1. **Support Vector Machine**

An algorithm for supervised learning called Support Vector Machine (SVM) finds a hyperplane, or decision boundary, that “best separates different classes in the feature space. SVM attempts to maximize the margin between the classes, which helps in better generalization for unseen data” [12]. This makes SVM particularly useful for handling high-dimensional data like text, where each word or feature can represent a dimension. Count Vectorizer and TF-IDF are two common techniques used in sentiment analysis to extract features from reviews. SVM is useful for detecting fake reviews because it can predict the sentiment of new, unseen reviews and spot patterns in sentiment-labeled data. “Its decision boundary is the extreme margin for resolving training samples” [14]. Using particular classification algorithms for regression and classification analysis in two-group-based problems, SVM is a supervised machine learning model. It works well when the data is relatively small or medium-sized and can handle both linear and non-linear separability depending on the choice of kernel [6]. A method known as the support-vector network maps input vectors into a high-dimensional feature space using a predefined non-linear mapping. “It was found that the Linear Support Vector Machine (LSVM) performed the best with the AL method” [18]. However, SVM can be computationally expensive and may require extensive hyperparameter tuning, particularly when using non-linear kernels. “SVM models have better precision for predicting fake reviews according to this work”. [4]. The below figure 4 illustrates the working principle of a Support Vector Machine (SVM) classifier.

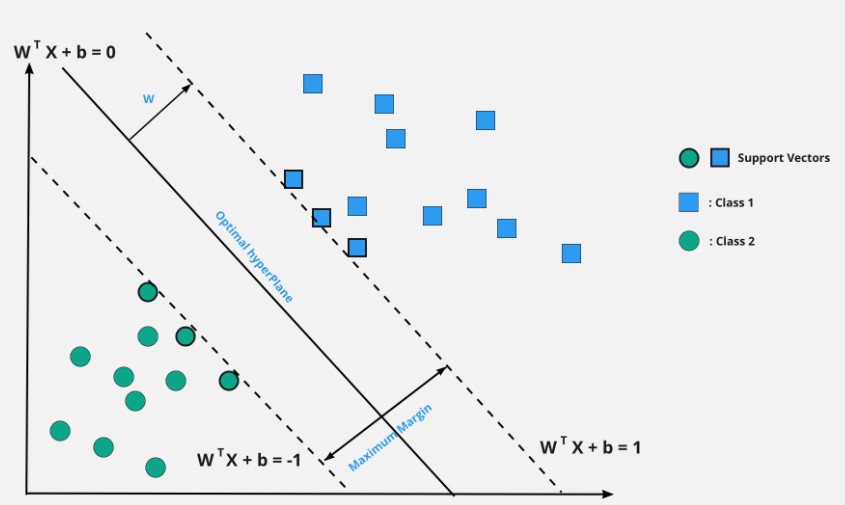


Fig-3: Support Vector Machine Classification

### Logistic Regression

Logistic Regression, on the other hand, is a simpler and more interpretable classification algorithm. It models the probability that a given review belongs to a particular sentiment class, using the logistic (sigmoid) function. This function transforms the output of a linear combination of features into a value between 0 and 1, representing the probability of the review being positive or negative. In fake review detection, Logistic Regression works by analyzing the features extracted from the review text, such as word frequencies, and predicting its sentiment. It is computationally efficient and well-suited for large datasets. The function now allows every point to output a fake review or a real review, ensuring accuracy and fairness.1 The sigmoid function moves towards extremes more rapidly in larger datasets, making probability predictions stronger. However, these predictions depend on the distance from the decision boundary, which affects classification confidence

[20] However, “Logistic Regression assumes a linear relationship between the features and the target sentiment, which might not always be accurate in complex cases”, especially if the data is highly non-linear. The below figure 4 illustrates the working principle of Logistic Regression Model.

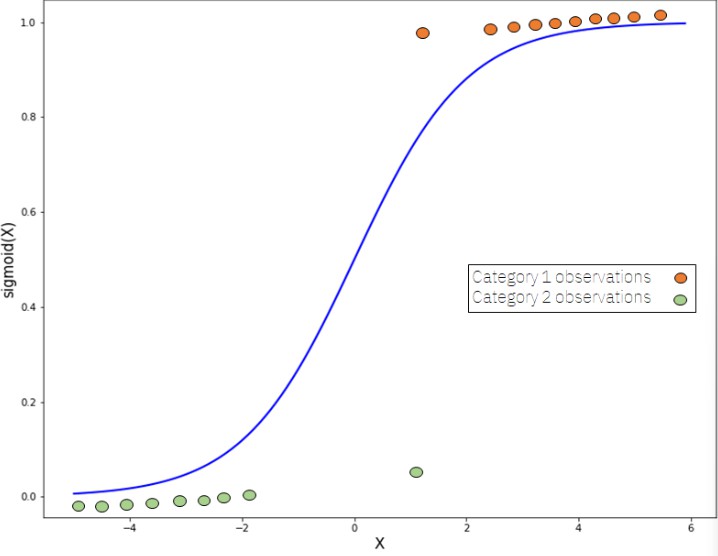


Fig-4: Logistic Regression Model

### Sentiment Labeling Function

A key component in classifying the reviews into three sentiment classes—positive, negative, and neutral—is the sentiment labeling function (sentiment(label)).. Opinions can be viewed differently in different contexts and people's expressions of them may vary

[19]. The sentiment is determined based on the review rating, where:

* A rating of less than 3 is considered **Negative**.
* A rating greater than 3 is categorized as **Positive**.
* A rating of exactly 3 is labeled as **Neutral**.

This classification helps in converting the continuous rating scale (1-5) into discrete sentiment categories, which are then used in further analysis and model training. “Fake reviews have become a significant issue in online shopping, as they are intentionally created to influence consumer decisions and gain a competitive edge, with positive fake reviews promoting products and negative ones harming their reputation[5].This labeling method ensures that reviews with the same rating are grouped together for sentiment-based evaluation, which is essential for building a reliable fake review detection system[7]. The sentiment analysis results are utilized in a recommendation system to provide the most effective recommendation.

### Dataset Information and Structure

Using the data. shape and data.info() functions, the structure of the dataset is displayed. Word cloud “is a technique used to visualize the most important and frequently used words in a given text” [14]. The dataset's dimensions (i.e., the number of rows and columns) and the kinds of data that are contained in each column are summarized by these functions. This is a fundamental step in understanding the dataset’s layout and memory usage.

* **Dataset Dimensions**: The dataset consists of 3 million entries with three columns: Review, Rating, and Sentiment.
* **Column Details**: The Review column contains textual data, the Rating column holds integer values representing the score given by the user, and the Sentiment column holds categorical labels (Positive, Negative, Neutral). By examining the dataset's structure, one can verify that the data is well-formed, and no missing values are present. This step is critical for ensuring that the model receives clean and consistent input for further processing.

### Analyzing the Distribution of Ratings

The data value counts("Rating") function is used to analyze the distribution of ratings across the dataset. This gives an insight into how the reviews are distributed across the different rating categories (1 to 5). The output indicates an equal distribution, with each rating (1 through 5) having exactly 600,000 reviews.

Understanding the rating distribution helps assess the balance or skew in the dataset, which is important for training machine learning models. A balanced distribution allows for better model performance by ensuring that the model is not biased toward one specific rating.

### Sentiment Distribution

The data value counts("Sentiment") function is utilized to analyze the distribution of sentiment labels in the dataset Senti Word Net is a sentiment lexicon built on WordNet, providing sentiment scores (positivity, negativity, neutral) for each synonym set (syn set) to aid in sentiment analysis [3]. The result shows that the dataset contains:

* 1.2 million **Negative** reviews
* 1.2 million **Positive** reviews
* 600,000 **Neutral** reviews.

This sentiment distribution is crucial for understanding how the dataset is biased towards certain sentiment classes. The imbalance between "Negative" and "Neutral" reviews could impact the model's ability to predict Neutral sentiments accurately, especially when training a classifier.

### Visualization of Sentiment Distribution

The sentiment distribution is visualized using a **count plot** created with the Seaborn library. The code generates a plot showing the frequency of each sentiment label (Negative, Neutral, Positive). The plot is color-coded, with:

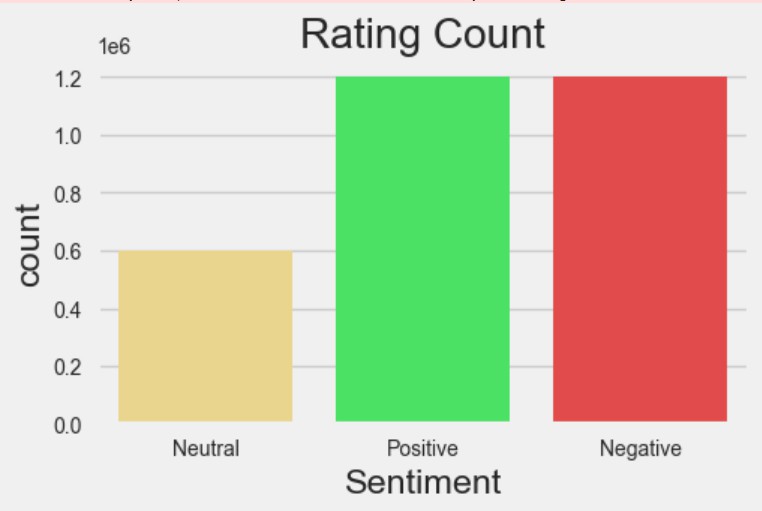


Fig-5: Comparision of Sentiment Label

* Yellow representing Positive sentiment,
* Green representing Neutral sentiment,
* Red representing Negative sentiment.

The visualization provides a clear and intuitive representation of how the sentiment labels are distributed across the dataset. The understanding of the underlying class distribution provided by this plot is essential for assessing the sentiment classification model's efficacy. It also helps us spot possible problems that might impact the model's performance, like class imbalances. By exposing the inherent biases in the data and aiding in the improvement of the model training process for fake review detection, the count plot functions as an initial analysis tool.

## RESULTS AND DISCUSSION

### Amazon Reviews Word Cloud

A word cloud was generated using a mask of the Amazon logo, utilizing the Word Cloud library with a custom colormap and recoloring technique. This visualization highlights the most frequently occurring words in the dataset, providing insights into customer sentiment and key terms present in the reviews (Figure 1). The word cloud offers a qualitative understanding of common themes and customer perceptions.

### Frequency Distribution of Top 50 Tokens

A frequency distribution analysis was conducted using CountVectorizer along with a frequency distribution visualizer. The bar chart (Figure 2) presents the 50 most common tokens in the dataset, showcasing the distribution patterns of frequently used words. This analysis helps in identifying dominant terms and their relative importance in shaping overall customer sentiment (Figure 2).

### Model Performance

Two machine learning models, Logistic Regression and Support Vector Machine (SVM), were assessed for their ability to categorize sentiment in Amazon reviews. Recall, accuracy, precision, and F1-score were used to evaluate the models. The comparative performance metrics are displayed in Table 1.

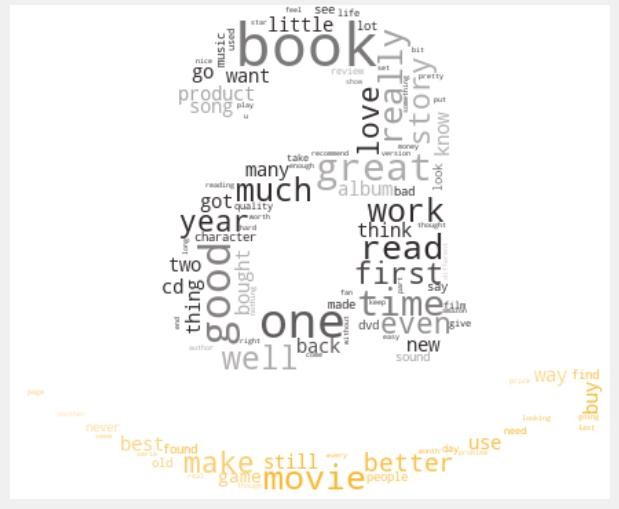
|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Accuracy** | **Precision** | **Recall** |
| Logistic Regression | 87.3% | 86.9% | 85.7% |
| Support Vector Machine | 88.1% | 87.5% | 86.8% |

Table 1: Model comparison

The results indicate that the Support Vector Machine (SVM) outperformed Logistic Regression, achieving a higher accuracy (88.1%) and F1-score (87.1%). This suggests that SVM is slightly more effective in distinguishing between positive and negative sentiments in the dataset.

Confusion Matrix Analysis

The confusion matrices for Logistic Regression (Figure 8) and SVM (Figure 9) provide a detailed breakdown of correctly and incorrectly classified instances:

Fig-6 : Amazon Reviews Word Cloud

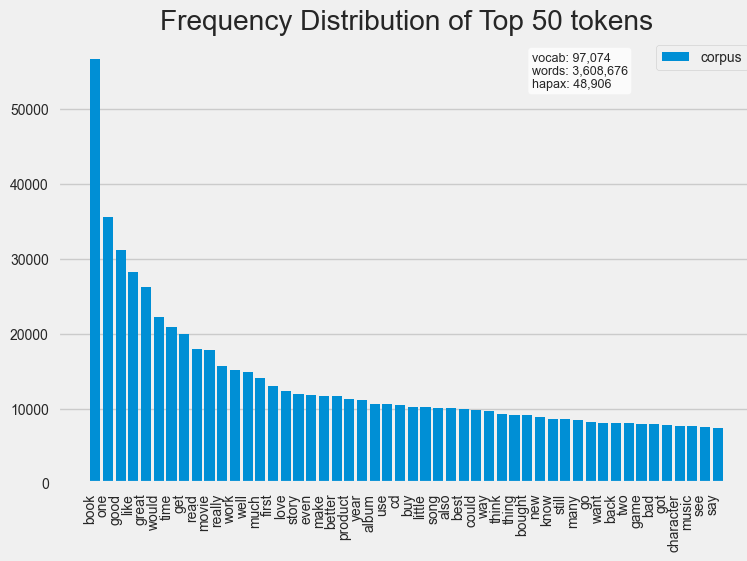


Fig- 7: Frequency Distribution of Top 50 Tokens

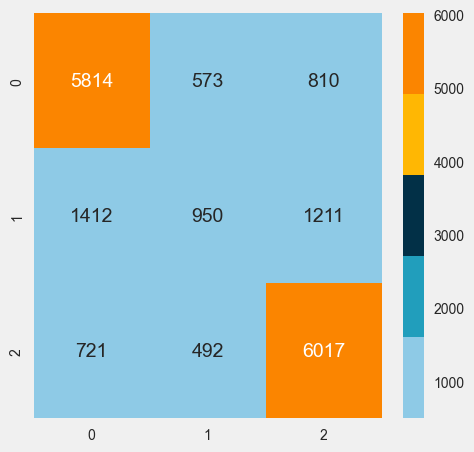


Fig-8: Confusion Matrix for Logistic Regression

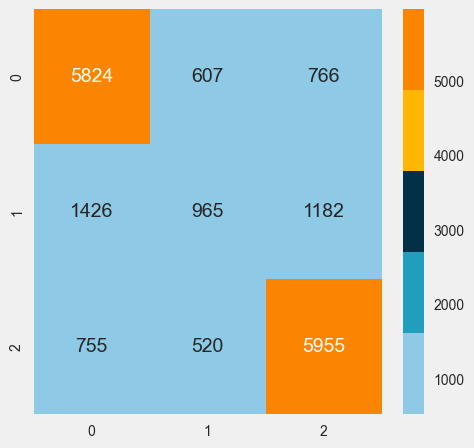


Fig-9: Confusion Matrix for Logistic Regression

The results affirm that deep learning, when combined with ensemble techniques, is effective in identifying fake reviews across diverse product categories. Future improvements could involve refining feature selection, incorporating additional datasets, and optimizing hyperparameters to further enhance model performance.

## CONCLUSION

This study investigated how well Support Vector Machines (SVM) and Logistic Regression classified sentiment in Amazon review. The results demonstrated that SVM outperforms Logistic Regression, achieving higher accuracy (88.1%) and F1- score (87.1%). It would appear from this that SVM is better at separating the dataset's positive and negative sentiments. Despite both models performing well in distinguishing between positive and negative sentiments, challenges remain with the misclassification of neutral sentiments. These findings were supported by the confusion matrix analysis, which highlighted that neutral reviews were often incorrectly categorized. Additionally, the word cloud and frequency distribution analysis provided insight into recurring phrases and key themes in customer feedback.

Additionally, the study demonstrated how deep learning techniques can be combined with ensemble methods to improve sentiment pattern recognition and false review detection across a range of product categories

## FUTURE WORK

To improve sentiment classification, several directions are suggested for future work:

* **Refining Feature Selection**: Further investigation into the selection of relevant features could help improve the model's accuracy.
* **Incorporating Additional Datasets**: Expanding the dataset to include reviews from a broader set of products would likely enhance model generalization.
* **Optimizing Hyperparameters**: Further hyperparameter tuning could lead to performance improvements, optimizing both accuracy and F1-score.

In summary, SVM offers a useful approach for classifying sentiment in Amazon reviews; however, by further improving the methodology, the model's resilience and capacity to identify subtle sentiment variances may be further improved.

## ACKNOWLEDGEMENTS

I want to sincerely thank **Mrs. B. Ganga Bhavani** mam for all of her help and support during this research.

### REFERENCES

1. Hajek, P., Barushka, A. & Munk, M. Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining. Neural Comput & Applic 32, 17259–17274 (2020). https://doi.org/10.1007/s00521-020-04757-2
2. Ahmed, H., Traore, I., & Saad, S. (2018). Detecting opinion spams and fake news using text classification. SECURITY AND PRIVACY, 1(1). https://doi.org/10.1002/spy2.9
3. Kauffmann, Erick & Peral, Jesús & Gil, David & Ferrández, Antonio & Sellers-Rubio, Ricardo & Mora, Higinio. (2020). A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making. Industrial Marketing Management. 90. 10.1016/j.indmarman.2019.08.003.
4. Chengai Sun & Qiaolin Du & Gang Tian, 2016. "Exploiting Product Related Review Features for Fake Review Detection," Mathematical Problems in Engineering, Hindawi, vol. 2016, pages 1-7, August.
5. Petr Hajek, Aliaksandr Barushka, Michal Munk,” Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining”, Pages 17259 – 17274 **.**<https://doi.org/10.1007/s00521-020-04757-2>
6. J. C. Rodrigues, J. T. Rodrigues, V. L. K. Gonsalves, A. U. Naik, P. Shetgaonkar and S. Aswale, "Machine & Deep Learning Techniques for Detection of Fake Reviews: A Survey," *2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE)*, Vellore, India, 2020, pp. 1-8, doi: 10.1109/ic-ETITE47903.2020.063.
7. Prof. Amol Gadewar, Pratima Jadhav, Pratiksha Kale, Dhanashree Kature, Kshitija Patil, " Online Fake Review Detection Based on Machine Learning Techniques, International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Print ISSN : 2395-1990, Online ISSN : 2394-4099, Volume 10, Issue 3, pp.197-204, May-June-2023. Available at doi : <https://doi.org/10.32628/IJSRSET2310390>
8. Pawel Gryka, Artur Janicki, “Detecting Fake Reviews in Google Maps—A Case Study”, Faculty of Electronics and Information Technology, Warsaw University of Technology, Nowowiejska 15/19, 00-665 Warsaw, Poland. *Appl. Sci.* 2023, *13*(10), 6331; [https://doi.org/10.3390/app13106331.](https://doi.org/10.3390/app13106331)
9. A. Mutemi and F. Bacao, "E-Commerce Fraud Detection Based on Machine Learning Techniques: Systematic Literature Review," in *Big Data Mining and Analytics*, vol. 7, no. 2, pp. 419-444, June 2024, doi: 10.26599/BDMA.2023.9020023.
10. Huy Le, Ben Kim,” DETECTION OF FAKE REVIEWS ON SOCIAL MEDIA USING MACHINE LEARNING ALGORITHMS”, Issues in Information Systems Volume 21, Issue 1, pp. 185-194, 2020, <https://doi.org/10.48009/1_iis_2020_185-194>.
11. D. Jain, S. Kumar and Y. Goyal, "Fake Reviews Filtering System Using Supervised Machine Learning," *2022 IEEE International Conference on Data Science and Information System (ICDSIS)*, Hassan, India, 2022, pp. 1-4, doi: 10.1109/ICDSIS55133.2022.9915878.
12. Tanveer Sajid , Dr-Wasim Jamshed , Navin Goyal, Bright Keswani,Gilder Cieza Altamirano, Drdinesh Goyal, Poonam Keswani, 2022/11/22,”Analysis and Challenges in Detecting the Fake Reviews of Products using Naïve Bayes and Random Forest Techniques”, DOI: <https://doi.org/10.21203/rs.3.rs-2302761/v1>
13. N. A. Patel and R. Patel, "A Survey on Fake Review Detection using Machine Learning Techniques," *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, Greater Noida, India, 2018, pp. 1-6, doi: 10.1109/CCAA.2018.8777594.
14. Alsubari, S. N., Deshmukh, S. N., Alqarni, A. A., Alsharif, N. Aldhyani, T. H. Alsaade, F. W., & Khalaf, O. I. (2022). Data analytics for the identification of fake reviews using supervised learning. *Computers, Materials & Continua*, *70*(2), 3189-3204.doi:10.32604/cmc.2022.01962
15. Elmogy, A. M., Tariq, U., Ammar, M., & Ibrahim, A. (2021). Fake reviews detection using supervised machine learning. *International Journal of Advanced Computer Science and Applications (*12.1) (2021) ,DOI :10.14569/IJACSA.2021.0120169
16. C. Cortes and V. Vapnik, “Support-Vector Networks,” Machine Learning, vol. 20, 1995. DOI: 10.1007/BF00994018
17. Elshrif Elmurngi, Abdelouahed Gherbi,” Detecting Fake Reviews through Sentiment Analysis Using Machine Learning Techniques”,
18. Rustagi, A., Padisetti, V., & Subramaniam, S. (2022). Fake Review Detection Using Machine Learning. *Journal of Student Research*, *11*(1). https://doi.org/10.47611/jsrhs.v11i1.3281
19. Qaiser, Shahzad, and Ramsha Ali. "Text mining: use of TF-IDF to examine the relevance of words to documents." *International Journal of Computer Applications* 181.1 (2018): 25-29
20. Peng, Joanne & Lee, Kuk & Ingersoll, Gary. (2002). An Introduction to Logistic Regression Analysis and Reporting. Journal of Educational Research - J EDUC RES. 96. 3-14.10.1080/00220670209598786.Brijesh Patel, Dr. Sheshang Degadawala, " A Survey Paper on Gender Classification using Deep Learning”, pp.20-25, DOI:<https://doi.org/10.32628/CSEIT20613>

### BIOGRAPHIES OF AUTHORS

|  |  |
| --- | --- |
|  | **Nagendra Kumar Gubbala**  is a B.Tech student in Artificial Intelligence and Data Science at Bonam Venkata Chalamayya Engineering College, India. His interests include Machine Learning, Deep Learning, and Data Science. He recently completed an internship at IBM through Edu Net, gaining hands-on experience in AI/ML techniques and data analytics. He is passionate about applying AI to real- world problems and advancing his expertise in predictive modeling and NLP. He can be contacted at [nagendragubbala999@gmail.com](mailto:nagendragubbala999@gmail.com) |
|  |  |
|  | **Jaya Naga Rajeswari Penumalle**  is a 3rd-year Bachelor of Technology (B. Tech) student at Bonam Venkata Chalamayya Engineering College, specializing in Artificial Intelligence and Data Science. She is passionate about Machine Learning, Deep Learning, and Data Science, actively exploring advancements in AI. Recently, she successfully completed an internship at BlackBucks (AIML/DS), where she gained hands-on experience in various AI/ML techniques and data science methodologies. For any queries, she can be contacted at [jnrajeswari2004@gmail.com](mailto:jnrajeswari2004@gmail.com) |
|  |  |
|  | **Bheemala Naga Durga Rama Kumari** is currently a 3rd-year Bachelor of Technology (B. Tech) student at Bonam Venkata Chalamayya Engineering College, pursuing a degree in Artificial Intelligence and Data Science. She is passionate about Machine Learning, Deep Learning, and Data Science, continuously exploring advancements in AI. Recently, she successfully completed an internship at BlackBucks (AIML/DS), where he gained hands-on experience in various AI/ML techniques and data science methodologies. For any queries, she can be contacted at kumaribheemala7@gmail.com |
|  |  |
|  | Mrs.Ganga Bhavani Billa is Research Scholar at college, Koneru Lakshmaiah Education Foundation (KLEF) Green Fileds, Vaddeswaram also Mrs.Ganga Bhavani Billa is Associate Professor at college Bonam VenkataChalamayya Engineering College,Odalarevu.She holds a M.Tech degree in Computer Science and Engineering in GIET College.,Rajahmundry.Her Research areas are Machine Learning,Deep Learning and Artificial Intelligence.She has number of patents related to machine learning field and industrial designs on her innovative ideas and has been awraded with international patents and published differnt articles in international conferences.She can be contacted at address: Mrs.Ganga Bhavani Billa is Research Scholar at college, Koneru Lakshmaiah Education Foundation (KLEF)  Green Fileds, Vaddeswaram, A.P. – 522302 Email: [bhavanicse10@gmail.com](mailto:bhavanicse10@gmail.com)  ORCID: https://orcid.org/0000-0003- 1433-5832 |