

Research Paper

**NEXTGEN OPINION INTELLIGENCE SYSTEM FOR E-COMMERCE FOOD
REVIEW ANALYTICS**

K. JAYA SRI¹, VENKATA SIVA NAGARAJU²

¹ Assistant Professor, Department of MCA, Chirala Engineering College, Chirala, Andhra Pradesh, India-523157.

Email: srij4620@gmail.com

² MCA Student, Department of MCA, Chirala Engineering College, Chirala, Andhra Pradesh, India-523157.

Email: nidubrolusiva6@gmail.com

ABSTRACT

The exponential growth of e-commerce platforms has led to a substantial increase in online customer reviews, making sentiment analysis an essential tool for understanding consumer opinions and product performance. Amazon food reviews, in particular, provide valuable insights into customer experiences regarding product quality, taste, packaging, and overall satisfaction. Traditional sentiment analysis approaches, such as manual review evaluation and rule-based techniques, often struggle with scalability, efficiency, and accuracy when processing large volumes of unstructured textual data. This research presents a Machine Learning-based framework for automated sentiment prediction of Amazon food reviews using textual features. The proposed system employs advanced Natural Language Processing (NLP) techniques, including text normalization, stopword removal, stemming, lemmatization, punctuation elimination, and noise reduction, to enhance data quality. The processed text data is transformed into numerical feature vectors using the Term Frequency–Inverse Document Frequency (TF-IDF) technique. To identify the most effective classification model, multiple machine learning algorithms are implemented and evaluated, including Multinomial Naïve Bayes (MNB), Support Vector Classifier (SVC), K-Nearest Neighbors (KNN), and Light Gradient Boosting Machine (LightGBM). Comparative performance analysis demonstrates that the LightGBM model achieves superior accuracy and classification performance in multi-class sentiment prediction, making it the proposed model for the system. Furthermore, an interactive Graphical User Interface (GUI) is developed to facilitate dataset uploading, text preprocessing, model training, performance evaluation, visualization, and sentiment prediction for unseen reviews. The proposed framework provides an efficient, accurate, and scalable solution for automated sentiment analysis, enabling businesses and consumers to derive meaningful insights from large-scale Amazon food review data.

Keywords: Natural Language Processing, Machine Learning, LightGBM, Term Frequency–Inverse Document Frequency, Graphical User Interface.

1. INTRODUCTION

In the digital age, the rapid growth of online platforms has generated an abundance of user-generated content, particularly in the form of reviews on e-commerce sites such as Amazon. These reviews provide a valuable resource for

understanding consumer sentiment, offering insights into customer satisfaction and product perception. For organizations seeking to improve customer satisfaction and refine their

offerings, analyzing the sentiment within this textual data is essential. Sentiment analysis, a specialized field within NLP, has become an essential tool for interpreting and quantifying these sentiments, enabling data-driven decision-making and customer-centric strategies.

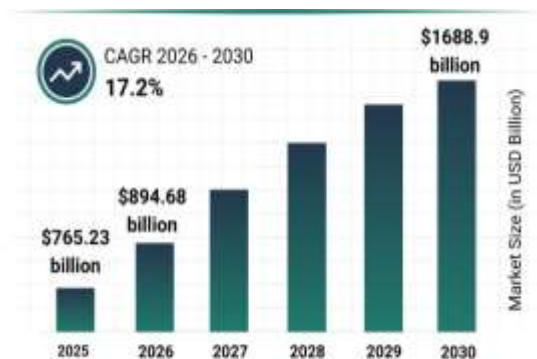


Fig. 1: Growth of online food services in India.

Since its inception, sentiment analysis has progressed from basic methods to highly sophisticated techniques. Originally, sentiment analysis relied on simple algorithms that quantified the frequency of positive and negative words in texts. However, these approaches were limited in their ability to capture the nuanced semantics and contextual subtleties of language [1]. The emergence of pre-trained vector representations marked a significant leap forward, as they enabled the encapsulation of semantic information within a continuous vector space. Techniques such as Word2Vec [2] and GloVe [3] were instrumental in improving sentiment analysis models by capturing semantic relations and word context. This development addressed previous limitations, particularly in recognizing contextual nuances, by focusing on the global co-occurrence statistics of words.

2. LITERATURE SURVEY

Mishra et al. [4] applied the TF-IDF method to hotel reviews to generate suitable inputs for

recommendations, highlighting the model's effectiveness in extracting meaningful features from textual data for sentiment-based filtering. Pimpalkar and Raj [5] investigated the impact of different preprocessing techniques on the performance of ML algorithms by applying TF-IDF and BoW to text analysis tasks. Their insights show that preprocessing steps can significantly affect model results, with the impact of TF-IDF being negligible, while the F1 score of the classifier is significantly improved when using BoW features.

Tripathy et al. [6] investigated the effectiveness of n-gram methods in sentiment classification of movie reviews from the IMDb dataset, comparing different ML algorithms and emphasizing the role of preprocessing and vectorization techniques. The study shows that their algorithm performs better when using unigrams or bigrams, and that accuracy decreases as n increases, likely due to the short length of the reviews in the dataset. To specifically compare these three methods, several studies have been conducted on different data sets. Hasan and Matin [7] compare the performance of TF-IDF with N-gram and BoW with N-gram for sentiment analysis of customer reviews. Hasan and Matin also confirm the results of Tripathy et al, which state that accuracy decreases significantly with higher values of n. Contrary to the findings of Pimpalkar and Raj, this study finds that TF-IDF performs better than BoW for most techniques, noting that this study combines them with the N-gram method.

While frequency-based methods have been effective, they process the text as a unified whole, often overlooking localised expressions of sentiment. To address this, Garapati and Chakraborty [8] propose the Review Text Granularity (RTG) model, which extracts and prioritises sentiment-rich phrases using linguistic patterns before applying traditional vectorisation. This selective emphasis improves feature quality for classifiers such as SVM and logistic regression, demonstrating how context-aware granularity can serve as a

valuable preprocessing step before conventional embedding techniques.

FastText [9], developed by Facebook's AI Research lab, extends Word2Vec by incorporating subword information. FastText considers not only whole words but also prefixes and suffixes, breaking words into subword units using n-grams. This allows the model to handle words outside its vocabulary, making it particularly effective for sentiment analysis of customer reviews, which often include slang ("tbh" for "to be honest"), misspellings ("amzn" instead of "Amazon"), and newly coined words ("tofurkey" for "vegan tofu turkey food") [10, 11]. By leveraging subword information, FastText ensures that the sentiment analysis model remains robust and accurate, even when faced with unconventional language.

Dharma et al. [12] conducted a study in which they apply these three methods to generate word embeddings for a dataset of 20 news topics. The authors found that FastText performed better than the others, but the difference in performance was not statistically significant. Kaibi et al. [13] conducted a study to evaluate the performance of different pre-trained word vectorization methods for sentiment analysis on Twitter. The results show that FastText performed slightly better than Word2Vec and GloVe. This is further confirmed by a study on sentiment analysis of hotel reviews. Khomsah et al. [14] compare Word2Vec and FastText. Similarly, FastText produced better results; however, the difference was negligible.

Nguyen et al. [15] addressed the complexity of designing hot-air drying systems by developing machine-learning models to predict two key performance metrics (air volume flow rate and heater power) from inputs including outdoor temperature, relative humidity, heater outlet temperature, exhaust humidity, and moisture evaporation. They compare six models (GLM, deep learning, decision tree, random forest, GBT, SVM) and

report that GBT best predicts ($R^2 = 0.994$, RE = 3.2%) and SVM excels for Q ($R^2 = 1$, RE = 0.4%). The authors note that while ML can efficiently model nonlinear, complex drying relationships, gaps remain in integrating such models with real-world dryer control systems and validating them under broader industrial conditions.

Thibault et al. [16] evaluated how different machine-learning model selection approaches affect the accuracy of predicting volume reduction (shrinkage) in plant-based dehydrated foods. They built models using techniques such as Extreme Learning Machine (ELM), Support Vector Machine (SVM), and Evolutionary Polynomial Regression (EPR) on experimental data, selecting optimal input combinations, and validated them with external data. Their results showed that SVM (with 9 inputs) gave the highest accuracy ($R = 0.895$, RMSE = 0.099), while a reduced-input ELM model was nearly as good and more parsimonious; they note that choosing the right model selection approach is critical, but a gap remains in generalizing these models across food types and drying conditions.

Frying is a critical food processing operation that enhances both texture and flavour, making it indispensable in the multi-billion-dollar snack food industry [17]. The design and optimization of large-scale industrial frying processes require an in-depth understanding of frying kinetics. Physics-based modelling has been widely employed to analyze frying kinetics, leading to the development of several mathematical models [18]. However, these models are computationally demanding and time intensive. As an alternative, ML-based predictive modelling presents a promising approach to improving energy efficiency and product quality in frying. Several researchers have successfully applied ML techniques to optimize frying kinetics, demonstrating their potential in process prediction.

3. PROPOSED SYSTEM

The proposed system is an automated ML-based sentiment prediction system designed to analyze Amazon food reviews and classify them into multiple sentiment categories. It replaces manual review analysis with a structured and intelligent workflow that improves accuracy, speed, and scalability. The complete working process of the proposed system is explained step by step below.

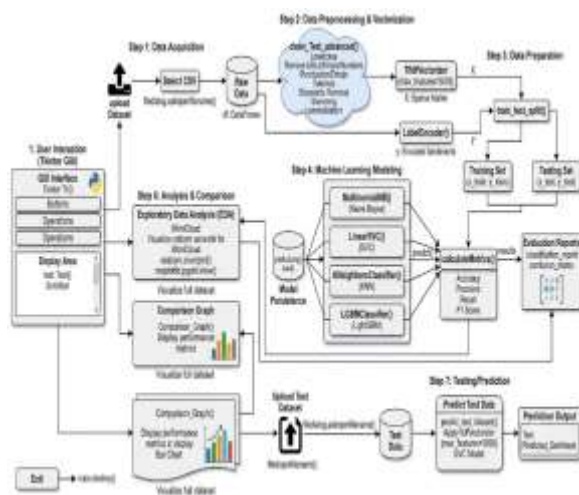


Fig. 2: Proposed system architecture.

Dataset Collection: The process begins with collecting customer review data in a structured format. The dataset contains textual reviews along with corresponding sentiment labels representing customer opinions. These reviews serve as the primary source of information for training and evaluating the prediction models. Proper dataset preparation ensures meaningful and reliable analysis.

Data Upload: The collected dataset is uploaded into the system through a user-friendly GUI. This allows users to easily load review data without requiring technical expertise. Once uploaded, the dataset is displayed for verification to ensure correct formatting. This step prepares the data for further preprocessing and analysis.

Text Preprocessing: Raw review text often contains noise such as special characters, extra spaces, and irrelevant words. In this step, NLP techniques are applied to clean and standardize the text data. The goal is to remove

unnecessary elements while preserving meaningful content. Effective preprocessing improves the quality of features extracted in later stages.

Feature Extraction: After cleaning, the textual data is transformed into numerical form so that ML models can process it. TF-IDF vectorization is used to convert words into weighted feature vectors based on importance. This helps the system capture meaningful patterns from textual data. Proper feature extraction plays a key role in improving model performance.

Label Encoding: Sentiment labels are converted into numerical values to make them compatible with ML algorithms. This step ensures that categorical sentiment classes can be processed mathematically. Each sentiment category is mapped to a unique numerical value. This enables smooth model training and evaluation.

Data Splitting: The processed dataset is divided into training and testing subsets. The training data is used to teach the models, while the testing data evaluates their performance. This separation helps in measuring how well the model generalizes to unseen data. It also prevents overfitting and ensures reliable results.

Model Training: Multiple ML models such as MNBC, SVC, KNN, and LGBM are trained using the processed dataset. Each model learns patterns and relationships between review text and sentiment labels. Training multiple models allows performance comparison and better understanding of strengths and weaknesses. This step builds the core predictive capability of the system.

Model Evaluation: After training, each model is evaluated using performance metrics like accuracy, precision, recall, and F1-score. Confusion matrices and comparison graphs are generated for better visualization. These evaluation methods help measure how well each model performs. Proper evaluation

ensures selection of a reliable prediction model.

Best Model Selection: Based on performance comparison, the model with the highest accuracy and consistent results is selected. In this project, LGBM demonstrates superior classification performance. It provides better handling of complex patterns in textual data. Therefore, it is chosen as the proposed model for the final prediction.

Prediction on New Data: The final trained model is used to predict sentiments for new unseen reviews. Users can upload new review datasets through the GUI and obtain instant predictions. This makes the system practical and usable in real-world scenarios. The prediction module enables automated sentiment classification with high efficiency.

LGBM Model

LGBM is an advanced ensemble ML algorithm based on the Gradient Boosting framework. It builds decision trees sequentially, where each tree tries to correct the errors made by previous trees. LGBM is optimized for speed and efficiency, handling large datasets and high-dimensional feature spaces like TF-IDF vectors in text classification. Its use of histogram-based algorithms and leaf-wise growth allows it to achieve high accuracy while consuming less memory and training time.

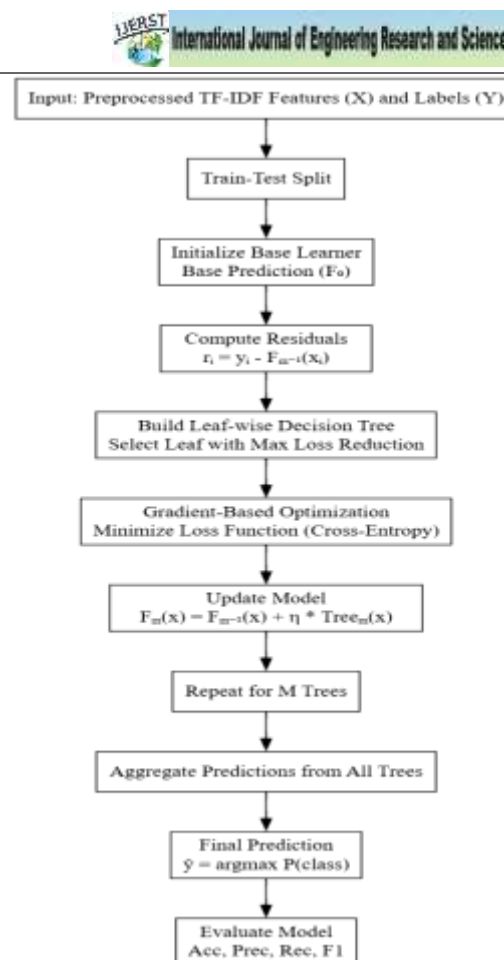


Fig. 3: Internal working of LGBM model.

LGBM constructs trees in a leaf-wise manner rather than level-wise. At each step, it chooses the leaf with the maximum reduction in loss to split, which improves predictive performance. Gradient Boosting allows the model to minimize a specified loss function by adding trees iteratively. Each tree focuses on residual errors from previous iterations. Additionally, LGBM uses techniques like Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) to reduce computation and handle sparse, high-dimensional data efficiently.

Input Pre-processed TF-IDF Features: The LGBM model receives the cleaned and vectorized dataset. Each Amazon food review is represented as a high-dimensional TF-IDF vector, and the corresponding sentiment label is included as the target. LGBM handles sparse matrices efficiently, which is ideal for text data.

Train-Test Split: The dataset is divided into training and testing sets. The training set is used to sequentially build trees, while the testing set is reserved for evaluating model performance. This ensures unbiased validation.

Initialize Base Learner: LGBM starts with a simple decision tree or a base prediction (e.g., predicting the mean of the target). This base learner serves as the starting point for iterative boosting.

Compute Residual Errors: For each training instance, LGBM calculates the residual error (difference between predicted and actual labels). These residuals represent the part of the data that the previous trees could not correctly predict.

Build Leaf-Wise Decision Trees: LGBM grows trees leaf-wise rather than level-wise, selecting the leaf with the maximum reduction in loss to split. This allows the model to focus on the most significant errors first, improving accuracy efficiently.

Gradient-Based Optimization: The model uses gradient descent to minimize the loss function (e.g., cross-entropy for classification). Each new tree is trained to correct the errors of all previous trees, gradually improving predictions.

Aggregate Tree Predictions: Predictions from all trees are combined (weighted sum) to generate the final output. The model uses a learning rate to control the contribution of each tree, preventing overfitting.

Prediction Phase: For test samples, LGBM aggregates predictions from all trained trees and determines the final sentiment label. The ensemble nature ensures robust and accurate classification.

Model Evaluation: Predicted labels are compared with actual labels from the test set. accuracy, precision, recall, and F1-score metrics are calculated to evaluate model performance. LGBM achieves superior results

compared to MNBC, SVC, and KNN in this project.

Advantages

- **High Accuracy:** LGBM builds multiple leaf-wise decision trees and focuses on reducing errors iteratively, which provides superior accuracy for sentiment classification of Amazon food reviews.
- **Handles High-Dimensional Data Efficiently:** TF-IDF features produce high-dimensional sparse data. LGBM's histogram-based and leaf-wise growth methods process such data faster than traditional boosting algorithms.
- **Fast Training and Low Memory Usage:** Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) reduce computation and memory consumption, making LGBM faster and more scalable for large datasets.
- **Robust to Overfitting:** The use of learning rate and regularization parameters prevents overfitting, ensuring that the model generalizes well to unseen reviews.
- **Handles Nonlinear Relationships:** LGBM captures complex patterns and nonlinear relationships in text features, unlike simpler models like MNBC or KNN.
- **Supports Sparse and Missing Data:** LGBM is naturally compatible with sparse TF-IDF matrices and can handle missing values without extra preprocessing.
- **Scalable for Large Datasets:** LGBM can process large datasets efficiently due to optimized tree-building

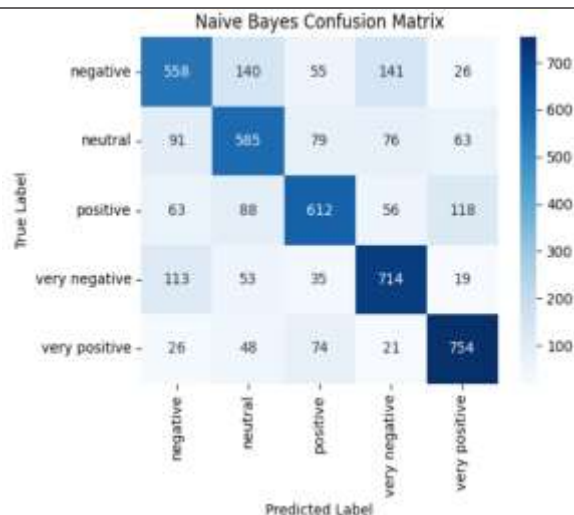


Fig. 7: Confusion matrix obtained using the MNBC model.

Fig. 7 displays the evaluation results of the MNBC model. Metrics such as accuracy, precision, recall, and F1-score are calculated, and the confusion matrix visualizes correct and incorrect predictions for each sentiment class. The figure demonstrates the MNBC model's ability to classify text-based reviews into five sentiment categories.

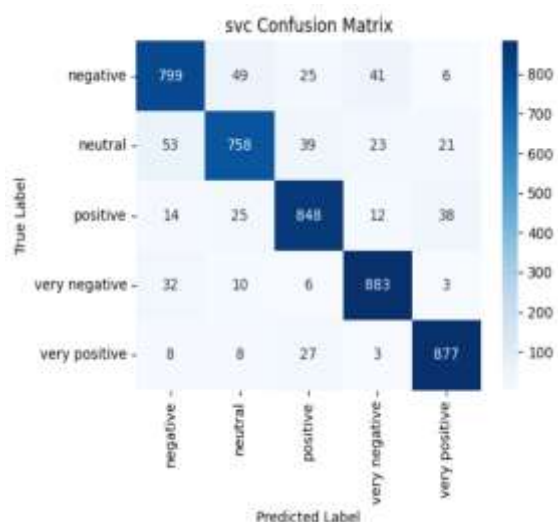


Fig. 8: Confusion matrix obtained using the SVC model.

Fig. 8 shows the evaluation of the SVC model. Similar metrics and a confusion matrix are presented. The SVC model achieves higher separation of classes by finding optimal hyperplanes in the feature space, providing

more precise predictions compared to MNBC in high-dimensional TF-IDF features.

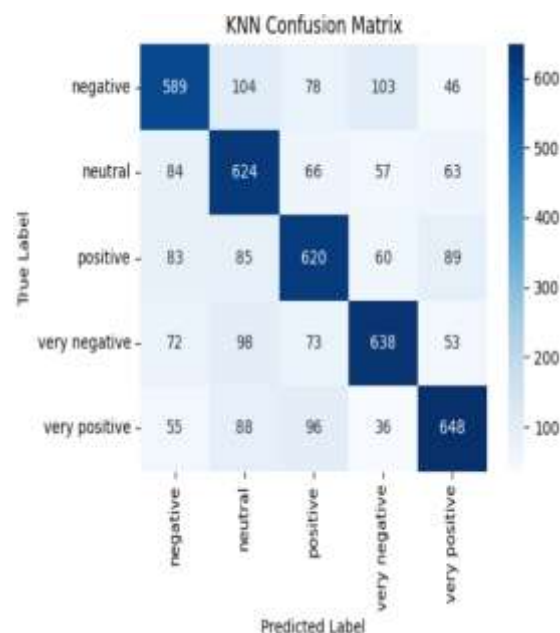


Fig. 9: Confusion matrix obtained using the KNN model.

Fig. 9 presents the results of the KNN model, including its accuracy and classification report. The KNN model predicts sentiments based on the closest neighbors in the vectorized feature space. The confusion matrix highlights the model's performance in distinguishing between similar sentiment classes.

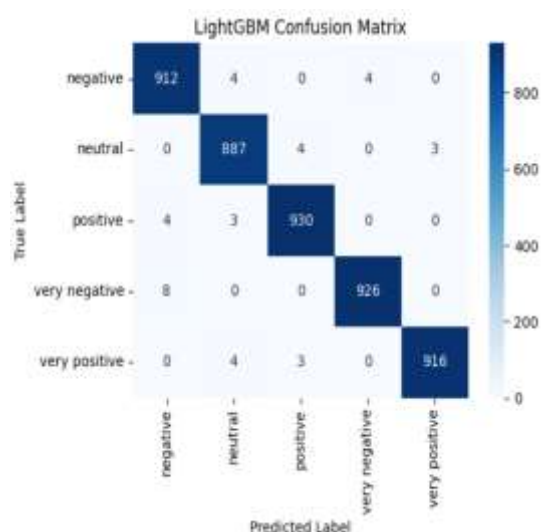


Fig. 10: Confusion matrix obtained using LGBM model.

Fig. 10 illustrates the evaluation of the LGBM model, which is the proposed model in the project. LGBM shows the highest Accuracy, Precision, Recall, and F1-score among all models. The figure includes a confusion matrix demonstrating precise classification of all five sentiment classes, confirming LGBM's effectiveness for high-dimensional textual data.

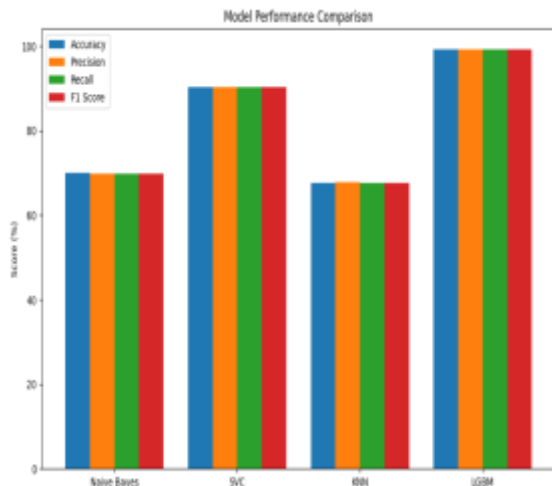


Fig. 11: Model performance comparison.

Fig. 11 compares the performance of MNBC, SVC, KNN, and LGBM models. A bar graph represents accuracy, precision, recall, and F1-score for each model, highlighting LGBM as the best-performing model. The comparison confirms that ensemble gradient boosting provides superior prediction for Amazon review sentiments.

Prediction completed!

Text Predicted Sentiment	
0 I do not know why this product is so hard to f...	positive
1 The first time I tried these, I thought, "WOW!..."	neutral
2 The five star system doesn't work well when th...	very negative
3 I agree with the previous review these were aw...	neutral
4 These treats sounded great, but unfortunately ...	neutral
5 I bought this when I saw that the price for 50...	positive
6 I agree with the previous reviewers - I should...	negative
7 The outside of the rice cake was thicker and a...	negative
8 If the world ever runs out of coffee I am in t...	neutral
9 I am a mother who had been successfully exclus...	negative

Fig. 12: Sample prediction on test data.

Fig 12 shows the GUI interface for predicting sentiments on a test dataset. Users upload new reviews, which are pre-processed and fed to the trained model. The predicted sentiment labels are displayed in the interface, allowing immediate analysis of new Amazon food reviews and validating the model's real-time usability.

Comparative Analysis

Table 1 presents the performance metrics of the four ML models trained on the Amazon food review dataset. MNBC achieved an accuracy of 69.94%, with corresponding precision, recall, and F1-score values close to 69–70%, indicating moderate performance in classifying sentiments. The SVC performed significantly better, achieving 90.39% accuracy, demonstrating strong classification ability across all sentiment classes. KNN showed lower performance with 67.69% accuracy, suggesting it is less effective for high-dimensional textual data. The LGBM model outperformed all other models, achieving an impressive 99.20% accuracy and equivalent precision, recall, and F1-score, confirming its capability to handle complex patterns in text data. These results highlight LGBM as the proposed model for this research due to its superior predictive performance.

Table 1: Performance comparison for the

Model	Accura cy (%)	Precisi on (%)	Recal l (%)	F1- Score (%)
NB Model	69.94	69.79	69.91	69.76
SVC Model	90.39	90.34	90.34	90.32
KNN Model	67.69	67.81	67.70	67.70
Proposed LGBM	99.20	99.19	99.20	99.20

MNBC, SVC, KNN, and LGBM models.

5. CONCLUSION

The research successfully implements an ML-based system to predict Amazon food review sentiments from textual and metadata features. The system handles the complete pipeline from dataset upload, preprocessing, EDA, model training, evaluation, and prediction using a GUI. Various models, including MNBC, SVC, KNN, and LGBM were trained and evaluated. Among these, LGBM demonstrated the highest performance, achieving superior accuracy, precision, recall, and F1-score, confirming its efficiency for high-dimensional textual data. The research emphasizes the importance of data preprocessing, including text cleaning, tokenization, stemming, lemmatization, and TF-IDF vectorization, to enhance model performance. EDA provided insights into class distribution and common textual patterns, guiding model selection and evaluation. The interactive GUI ensures ease of use for non-technical users, allowing smooth uploading of datasets, training models, and predicting sentiments in real-time. Overall, the research delivers an accurate, reliable, and user-friendly system for analyzing customer reviews, which helps businesses and researchers extract actionable insights from Amazon food reviews efficiently.

REFERENCES

- [1] Pang, B.; Lee, L. Opinion mining and sentiment analysis. *Found. Trends® Inf. Retr.*, 2, 1–135.
- [2] Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. *arXiv*, arXiv:1301.3781.
- [3] Pennington, J.; Socher, R.; Manning, C.D. Glove: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October; pp. 1532–1543.
- [4] Mishra, R.K.; Urolagin, S. A Sentiment analysis-based hotel recommendation using TF-IDF Approach. In Proceedings of the 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), Dubai, United Arab Emirates, 11–12 December 2019; pp. 811–815.
- [5] Pimpalkar, A.P.; Raj, R.J.R. Influence of pre-processing strategies on the performance of ML classifiers exploiting TF-IDF and BOW features. *Adcaij Adv. Distrib. Comput. Artif. Intell. J.* 2020, 9, 49–68.
- [6] Tripathy, A.; Agrawal, A.; Rath, S.K. Classification of sentiment reviews using n-gram machine learning approach. *Expert Syst. Appl.* 2016, 57, 117–126.
- [7] Hasan, T.; Matin, A. Extract Sentiment from Customer Reviews: A Better Approach of TF-IDF and BOW-Based Text Classification Using N-Gram Technique. In Proceedings of the International Joint Conference on Advances in Computational Intelligence: IJCACI, Daffodil International University, Birulia, Bangladesh, 20–21 November 2020; pp. 231–244.
- [8] Garapati, R.; Chakraborty, M. Enhancing Sentiment Analysis and Rating Prediction Using the Review Text Granularity (RTG) Model. *IEEE Access* 2025, 13, 20071–20100.
- [9] Joulin, A.; Grave, E.; Bojanowski, P.; Douze, M.; Jégou, H.; Mikolov, T. Fasttext. zip: Compressing text classification models. *arXiv* 2016, arXiv:1612.03651.

- [10] Edizel, B.; Piktus, A.; Bojanowski, P.;
Ferreira, R.; Grave, E.; Silvestri, F.
Misspelling oblivious word

embeddings. arXiv 2019,
arXiv:1905.09755.