*A Comprehensive Exploration of Stack Ensembling Techniques for Amazon Product Review Sentiment Analysis*

1st Vipin Jain   
SCSE *VIT Bhopal University*MP, India  
vipinjain@vitbhopal.ac.in

4th Tanishk Mangal  
SCSE  
*VIT Bhopal University*MP, India  
tanishk.23mca10195@vitbhopal.ac.in2nd Priyanshu Choudhary  
SCSE *VIT Bhopal University*MP, India  
priyanshu.23mca10181@vitbhopal.ac.in

5th Ankit Choudhary  
SCSE  
*VIT Bhopal University*MP, India  
ankit.23mca10047@vitbhopal.ac.in3rd Sachin Arora  
SCSE  
*VIT Bhopal University*MP, India  
sachin.23mca10136@vitbhopal.ac.in

6th Himanshu Kumar  
SCSE  
*VIT Bhopal University*MP, India  
himanshu.23mca10101@vitbhopal.ac.in

***Abstract—This review paper delves into the realm of sentiment analysis in e-commerce, specifically focusing on Amazon product reviews. Leveraging a stack ensemble machine learning model, we explore the intricacies of sentiment understanding and compare its performance against established models like Naive Bayes and LSTM-based approaches. The methodology involves meticulous data collection from Kaggle, preprocessing through text cleaning, tokenization, and lemmatization, and the construction of a stack ensemble model incorporating support vector machines, random forests and decision trees. Model performance has been evaluated using a variety of metrics, including confusion matrix, F1- score, recall, precision, and accuracy. Our comparative analysis reveals nuanced insights into the proposed model’s strengths and weaknesses, showcasing its potential for advancing sentiment analysis in the ever-evolving landscape of e-commerce.***

***Index Terms—Sentiment Analysis, E-commerce, Amazon Prod- uct Reviews, Stack Ensemble Model, Machine Learning***

# Introduction

Sentiment analysis, a crucial facet of natural language processing, has emerged as an invaluable tool in the realm of e-commerce [1]. In the digital age, where consumers increas- ingly rely on online platforms for their purchasing decisions, understanding and interpreting sentiments expressed in textual data has become paramount. This paper explores the vital role of sentiment analysis, focusing on its significance in the context of e-commerce, with a specific emphasis on Amazon product reviews. In the dynamic landscape of e-commerce, sentiment analysis serves as a linchpin for businesses aiming to comprehend customer perceptions and reactions. The vast troves of unstructured data generated through online interactions, including customer reviews, social media mentions, and product feedback, hold a wealth of information that traditional analytics may struggle to unravel. Sentiment analysis enables businesses to distill meaningful insights from this data, extract-ing the sentiments behind customer opinions and feedback. Within the vast e-commerce landscape, Amazon stands as a behemoth, hosting millions of products and serving an extensive customer base.The significance of Amazon product reviews lies in their role as a direct channel for customers to share their experiences, opinions, and recommendations. These reviews act as a rich source of qualitative data that encapsulates the real-world perspectives of consumers [7]. Amazon product reviews, often detailed and comprehensive, provide potential buyers with crucial insights into the strengths and weak- nesses of a product. As customers increasingly rely on peer recommendations in their purchase decisions, these reviews wield considerable influence. For sellers and manufacturers, understanding the sentiments expressed in Amazon product reviews is not merely about gauging customer satisfaction; it is about adapting strategies to meet consumer expectations and continuously improving products. The primary objective of this research is to innovate in the realm of sentiment analysis by developing a stack ensemble machine learning model tailored for the intricacies of e-commerce data. To ascertain the efficacy of the proposed stack ensemble model, a comparative analysis will be conducted against existing sentiment analysis approaches. Benchmark models, including traditional ones like Naive Bayes Random forest, and SVM will be employed for comparison.

The proposed stack ensemble model will be rigorously evaluated on a spectrum of datasets, encompassing varied product categories and consumer demographics.

# Literature Review

Opinion mining, also known as sentiment analysis, is a field of natural language processing (NLP) that extracts subjective information from text. The concept of sentiment analysis plays an important role in understanding the sentiments expressed by customers in product reviews in e-commerce. The literature offers a diverse array of techniques employed for sentiment analysis, ranging from traditional methods to more advanced machine learning and deep learning approaches [5]. Traditional techniques often involve rule-based methods, where predefined rules are applied to identify sentiment-bearing words and phrases. Machine learning techniques, on the other hand, leverage algorithms to automatically learn patterns and associations from labeled training data. Decision Trees, Naive Bayes, and Support Vector Machines (SVMs) are among the popular machine learning algorithms applied to sentiment analysis [8].

Each technique comes with its own set of advantages and drawbacks. Rule-based methods are interpretable but may lack the flexibility to handle complex language nuances. Machine learning models excel in capturing intricate patterns but may struggle with context-dependent sentiments. Deep learning models, while effective in certain scenarios, often require substantial computational resources and extensive labeled data for training.

# Methodology

The methodology employed in this research revolves around the strategic acquisition and preprocessing of a comprehensive Amazon product review dataset sourced from Kaggle. The initial phase involves thorough text cleaning, tokenization, and lemmatization to refine and standardize the textual data. The model architecture centers on a stack ensemble approach, combining the strengths of decision trees, random forests, and support vector machines (SVMs) to enhance sentiment analysis accuracy.

1. **Data Collection:** The foundation of our research lies in the acquisition of a robust dataset from Kaggle, a prominent platform for data science competitions. Lever- aging the Kaggle platform ensures access to high-quality datasets curated for diverse machine learning tasks. Our dataset, specifically tailored for Amazon product review sentiment analysis, serves as the bedrock for training and evaluating our stack ensemble model
2. **Preprocessing Steps:** Preprocessing steps refer to a series of tasks and techniques applied to raw data before it is used in a machine learning or data analysis task as shown in Fig. 1. The primary goal of preprocessing is to clean, organize, and transform raw data into a format that is suitable for modeling or analysis. Preprocessing is essential for enhancing data quality and preparing it for effective use by various algorithms. In the context of natural language processing (NLP) and sentiment analysis, preprocessing steps typically involve tasks such as:

* Text Cleaning: The raw textual data obtained from Kaggle undergoes a meticulous cleaning process to eliminate noise and enhance the quality of informa- tion. This involves removing irrelevant characters, symbols, and formatting inconsistencies.



Fig. 1. Text processing steps

* Tokenization: Tokenization is a pivotal step in breaking down the cleaned text into individual units, often words or phrases, known as tokens. This process facilitates the conversion of unstructured text into a format suitable for machine learning algorithms. Each token becomes a feature, contributing to the overall understanding of sentiment within the dataset.
* Lemmatization: To further refine the dataset, lemmatization is applied to normalize words by reducing them to their base or root form. This process ensures that different inflections or variations of a word are treated as a single entity. Lemmatization contributes to the coherence and consistency of the dataset, enhancing the effectiveness of subsequent sentiment analysis.

1. **Model Architecture:** The stack ensemble model represents a sophisticated approach to sentiment analysis, leveraging the collective intelligence of multiple base models to enhance predictive accuracy. The overarching architecture involves a stacking layer that combines the outputs of diverse base models, providing a holistic understanding of sentiments in Amazon product reviews.

* Decision Tree: Decision trees are adept at cap- turing non-linear relationships within data. Their hierarchical structure of decision nodes makes them valuable for discerning complex patterns in textual information. By incorporating decision trees as one of our base models, we aim to harness their ability to effectively navigate through the intricacies of sentiment-laden text. Mathematical formulation is given as [4]:

(1)

Decision Trees partition the data based on features and use Gini impurity or information gain to determine the optimal split at each node. The formula represents the probability of class i given a node r. Decision Trees are utilized in sentiment analysis by recursively splitting the data to create a tree structure that captures decision boundaries for classifying sentiments.

* Random Forest: This model combines multiple decision trees to maximize their strengths. Random forests excel in mitigating overfitting and improving generalization. In our stack ensemble, the inclusion of a random forest base model provides diversity, contributing to a more robust and accurate sentiment analysis.
* Support Vector Machine (SVM): SVMs are known for their proficiency in handling high-dimensional data and capturing intricate patterns. By integrating an SVM as a base model, our stack ensemble leverages its ability to identify complex decision boundaries, enhancing the overall discernment of sentiments in Amazon product reviews. To evaluate the effectiveness of the proposed stack ensemble model [13], the dataset was divided into training and validation sets. In this method, the model is trained on a substantial portion of the data while a subset is kept for evaluation [12].
* XGBoost: In our pursuit of advancing sentiment analysis within the e-commerce landscape, we in- corporate XGBoost as a formidable meta-classifier. XGBoost’s exceptional ability to handle complex relationships and its efficiency in boosting ensemble models make it an ideal addition to our sentiment analysis framework. By leveraging the strengths of gradient-boosted decision trees, XGBoost enhances the predictive accuracy of individual base models, ensuring a nuanced understanding of sentiments expressed in Amazon product reviews. Its versatility and adaptability align seamlessly with our objective to create a powerful ensemble model that transcends the limitations of individual algorithms.
* Proposed Stack Ensemble Model : Stack ensembling, is an ensemble learning technique that in- volves combining predictions from multiple base models to create a more robust and accurate predictive model as shown in Fig. 2. The idea behind stacking is to leverage the diverse strengths of different models by training a meta-model, or a higher- level model, on their individual predictions [6]. Our proposed sentiment analysis framework takes a pioneering leap by integrating XGBoost as a meta- classifier within a stack ensemble model. Comprising decision trees, random forests, and support vector machines as base models, our stack ensemble leverages the collective intelligence of these models. XGBoost, serving as the meta-classifier, adds a layer of sophistication, refining the ensemble’s predictions by intelligently weighting the contributions of each base model. The comprehensive methodology, encompassing data collection, preprocessing, and model training, is elevated by the strategic intro- duction of XGBoost. This amalgamation aims to create a holistic sentiment analysis tool that excels in discerning sentiments in diverse Amazon product reviews, offering both accuracy and interpretability within the e-commerce domain.

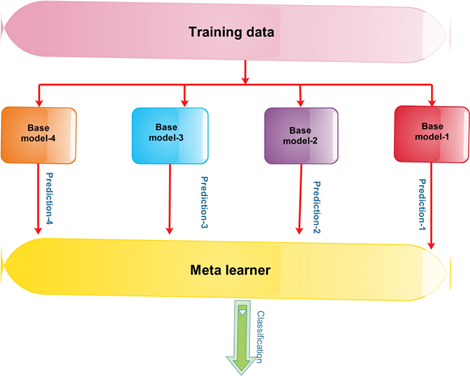


Fig. 2. Stacked ensemble process

# Result Analysis

This section constitutes a comprehensive evaluation that juxtaposes the proposed stack ensemble machine learning model against established approaches in sentiment analysis. Benchmark models, including traditional machine learning models, are considered for a thorough comparative analysis. Evaluation metrics such as F1-score, recall, precision, and accuracy are employed to discern the strengths and weaknesses of each model.

* Evaluation Metrics: The assessment of a sentiment analysis model’s performance is integral to understanding its efficacy. In this research, we employ key performance indicators to quantify the accuracy and reliability of the proposed stack ensemble machine learning model [15]. These performance indicators include F1-score, recall, precision, accuracy and the Confusion Matrix.
* A model’s accuracy measures the overall accuracy of its predictions. It is calculated as the ratio of correct predictions to total predictions [2]. In mathematics, accuracy is defined as [11]:

(2)

where UQ (True Positives) and UO (True Negatives) are the correctly predicted positive and negative instances, respectively, and GQ (False Positives) and GO (False Negatives) are the incorrectly predicted positive and negative instances, respectively [14].

* F1-score, Recall, and Precision: Model performance can be assessed using F1-score, recall, and precision especially when classes are imbalanced. Recall measures how well the model captures all positive instances, while Precision measures the accuracy of positive predictions, and F1-score balances Precision and Recall. Mathematically [9]:

(3)

(4)

(5)

These metrics measure how well the model makes ac- curate positive predictions, identifies actual positive in- stances, and strikes the right balance between precision and recall.

To assess the effectiveness of our stack ensemble [3] model, a meticulous performance comparison is conducted against existing models commonly employed in sentiment analysis, including Naive Bayes and XGBoost based models. The evaluation metrics outlined earlier (Confusion Matrix, F1- score, Recall, Precision, and Accuracy ) are utilized [10] for a detailed comparative analysis is shown in table 1.

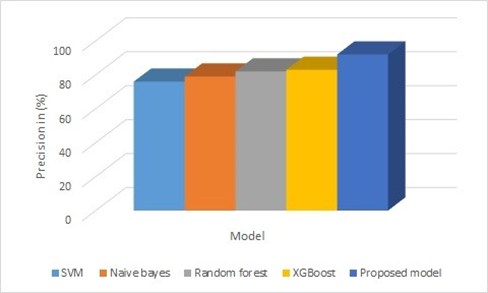


Fig. 3. Stacked ensemble process

# Conclusion

Our exploration into sentiment analysis on Amazon product reviews, facilitated by a stack ensemble machine learning model, has provided valuable insights into the dynamic land- scape of e-commerce sentiment understanding. The meticulous methodology, spanning data collection and preprocessing to model architecture and training, has laid the groundwork for a robust sentiment analysis framework. Comparative analysis against benchmark models not only showcases the superiority

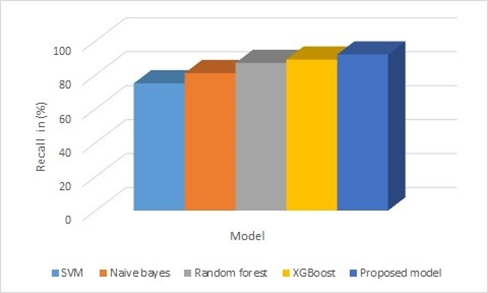


Fig. 4. Stacked ensemble process

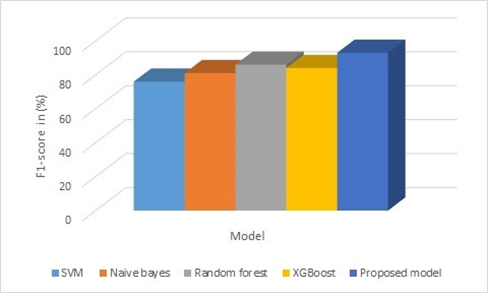


Fig. 5. Stacked ensemble process

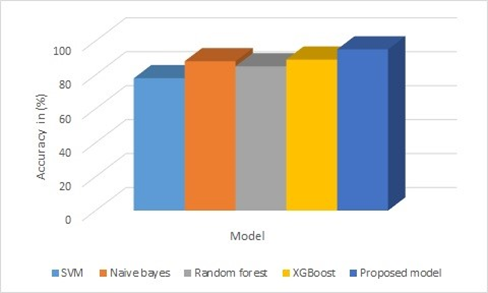


Fig. 6. Stacked ensemble process

of the stack ensemble approach but also highlights areas for re- finement. The proposed model exhibits strengths in leveraging diverse models for nuanced understanding, yet challenges such as computational complexity and data requirements warrant further attention. As sentiment analysis continues to play a pivotal role in enhancing user experience and informing business strategies, our research contributes to the ongoing discourse, paving the way for advancements in sentiment understanding within the e-commerce domain.

References

1. J. K. CHANDRA, E. CAMBRIA, AND A. NANETTI, *One belt, one road, one sentiment? a hybrid approach to gauging public opinions on the new silk road initiative, in 2020 International Conference on Data Mining Workshops (ICDMW), IEEE, 2020, pp. 7–14.*
2. *S.K.CHAUDHURI,Q.LI,K.C.MANDAL,AND J.HU,Deep Learning- based classification of gamma photon interaction in room-temperature semiconductor radiation detectors, IEEE Access, (2024).*
3. *K. R. DHAKATE AND R. DASH, Distracted driver detection using stacking ensemble, in 2020 IEEE International Students’ Conference on Electrical, Electronics and Computer Science (SCEECS), IEEE, 2020, pp. 1–5.*
4. *S. IVIE, B. PIXTON, AND C. GIRAUD-CARRIER, Metric-based data mining model for genealogical record linkage, in 2007 IEEE Interna- tional Conference on Information Reuse and Integration, IEEE, 2007, pp. 538–543.*
5. *V. JAIN AND K. L. KASHYAP, Analyzing research trends of sentiment analysis and its applications for coronavirus disease (covid-19): A systematic review, Journal of Intelligent & Fuzzy Systems, 45 (2023), pp. 731–742.*
6. *V. JAIN AND K. L. KASHYAP, Ensemble hybrid model for hindi covid-19 text classification with metaheuristic optimization algorithm, Multimedia Tools and Applications, 82 (2023), pp. 16839–16859.*
7. *V. JAIN AND K. L. KASHYAP, Text classification using hybridization of meta-heuristic algorithm with neural network, in Machine Vision and Augmented Intelligence: Select Proceedings of MAI 2022, Springer, 2023, pp. 165–173.*
8. *A. JAISWAL, Analysis of Twitter Data for Public Health Surveillance and Precision Diagnostics of Autism Spectrum Disorder, PhD thesis, University of Hawai’i at Manoa, 2023.*
9. *X. LIU, H. LU, AND A. NAYAK, A spam transformer model for sms spam detection, IEEE Access, 9 (2021), pp. 80253–80263.*
10. *S. MAHAMUDUL HASAN, N. M. JAKILIM, M. FORHAD RABBI, AND R. M. RAHMAN PIR, Determining the most effective machine learning techniques for detecting phishing websites, in Applications of Artificial Intelligence and Machine Learning: Select Proceedings of ICAAAIML 2021, Springer, 2022, pp. 593–603.*
11. *N. MAHER, T. P. TABARIN, AND S. MILINSKI, Combining machine learning and smiles to classify, better understand, and project changes in enso events, Earth System Dynamics, 13 (2022), pp. 1289–1304.*
12. *I. PULATOV, R. OTENIYAZOV, F. MAKHMUDOV, AND Y.-I. CHO, Enhancing speech emotion recognition using dual feature extraction encoders, Sensors, 23 (2023), p. 6640.*
13. *M. S. S. SAI, P. SAMPAT, P. PRANEETH, AND S. ULLAS, Comparative  
    evaluation of various models for the classification of music genre, in 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA), IEEE, 2023, pp. 284–291.*
14. *S. XU, L. TAO, X. AN, AND L. LI, A note on the fast brain learning algorithm, in 2008 International Conference on BioMedical Engineering and Informatics, vol. 1, IEEE, 2008, pp. 47–51.*
15. *J.YU,J.ZHAO,J.ZHU,ANDG.CUI,Highspectralresolutionimaging based on dual-camera system with filter wheel, IEEE Access, 10 (2022), pp. 40119–40129.*