A FIELD PROJECT REPORT

on

**“Sentiment analysis on Amazon reviews using machine learning techniques”**

**Submitted**

by

G.Gayatri Priya

221FA04529

K.Bharath

221FA04294

K.Srinivas

221FA04377

Y.Srinivas

221FA04384

**Under the guidance of**

Dr Deva Kumar

*Professor*



**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**VIGNAN'S FOUNDATION FOR SCIENCE, TECHNOLOGY AND RESEARCH Deemed to be UNIVERSITY**

**Vadlamudi, Guntur.**

**ANDHRA PRADESH, INDIA, PIN-522213.**



**CERTIFICATE**

This is to certify that the Field Project entitled **“Sentiment analysis of Amazon reviews using machine learning techniques”** that is being submitted by 221FA04529(G.Gayatri Priya),221FA04294(K.Bharath),221FA04377(K.Srinivas), 221FA04384(Y.Srinivas) for partial fulfilment of Field Project is a bonafide work carried out under the supervision of Mr. Dr.DEVA KUMAR, M.Tech., Professor, Department of CSE.

|  |  |  |
| --- | --- | --- |
| Guide name& Signature |  | Dr.K.V. Krishna Kishore |
| Assistant/Associate/Professor, CSE | HOD,CSE | Dean, SoCI |



**DECLARATION**

We hereby declare that the Field Project entitled “**Sentiment analysis of Amazon reviews using machine learning techniques**” is being submitted by 221FA04529(G.Gayatri Priya),221FA04294(K.Bharath),221FA04377(K.Srinivas), 221FA04384(Y.Srinivas) in partial fulfilment of Field Project course work. This is our original work, and this project has not formed the basis for the award of any degree. We have worked under the supervision of Mr. Dr.DEVA KUMAR, M.Tech., Professor, Department of CSE.

By

**221FA04529(G.Gayatri Priya),**

**221FA04294(K.Bharath),**

**221FA04377(K.Srinivas),**

**221FA04384(Y.Srinivas)**

Date: 15-10-2024

## **ABSTRACT**

Indeed, online shopping has remained a very vital pillar of consumer activity in the technological era. Reviews and feedbacks have been the most crucial determinants of whether or not products will thrive based on the individual opinions expressed on them by clients. For they form the direct line between businesses and their customers-a line whence mutual understanding of how the product performs, customer satisfaction, and what should be bettered are derived. However, handling and analyzing millions of reviews for a single product is a rather tough deal. The answer to this problem lies in the effective application of techniques for extracting meaningful information from voluminous customer feedback. This paper attempts to do a comparative study on several machine learning algorithms such as Logistic Regression, Support Vector Machines (SVM), Naïve Bayes, Decision Tree, and Multinomial models, on customer review classification. Results of the study clearly show Ensemble Classifier outperforms the mentioned individual algorithms in terms of accuracy and reliability of sentiment analysis. The importance of Ensemble methods for actionable extraction of insights out of large-scale review datasets in an e-commerce domain cannot be undermined.

**TABLE OF CONTENTS**

1. Introduction 1   
   1.1 Overview of Sentiment Analysis 2  
   1.2 Importance of Sentiment Analysis 2  
   1.3 Applications of Machine Learning in Sentiment Analysis 2
2. Literature Survey 3  
   2.1 Existing Research in Sentiment Analysis 4  
   2.2 Challenges in Sentiment Classification 5  
   2.3 Motivation for Hybrid Approaches 6
3. Methodology 7   
   3.1 Input Dataset and Features 8  
   3.1.1 Dataset Description 8  
   3.1.2 Target Variable (Sentiment) 8  
   3.2 Data Pre-processing 8  
   3.2.1 Handling Missing Data 8  
   3.2.2 Text Preprocessing 9  
   3.2.2.1 Tokenization and Lowercasing 9  
   3.2.2.2 Stop word Removal and Punctuation Cleaning 9  
   3.3 Feature Extraction 10  
   3.3.1 TF-IDF Vectorization 10  
   3.3.2 Sentiment Assignment 10  
   3.3.3 Rating Distribution Visualization 10  
   3.4 Model Building 11  
   3.4.1 Logistic Regression 11  
   3.4.2 Support Vector Machine (SVM) 12  
   3.4.3 Random Forest Classifier 12  
   3.4.4 Multinomial Naive Bayes 12  
   3.4.5 Bernoulli Naive Bayes 13  
   3.4.6 Decision Tree Classifier 13  
   3.4.7 Stacking Classifier (Hybrid Model with Random Forest and SGD) 13
   * 1. Hybrid Model with PCA and LDA 14

3.5 Model Evaluation 14  
3.5.1 Accuracy Score 14  
3.5.2 Precision, Recall, and F1-Score 16

1. Implementation 17  
   4.1 Environment Setup 18  
   4.2 4.2 Preprocessing, Dimensionality Reduction, and Model Training Code 18
2. Experimentation and Result Analysis 19  
   5.1 Comparison of Model Accuracies 20  
   5.2 Performance of Hybrid Models vs Individual Models 20
3. Conclusion 22  
   6.1 Key Findings 23  
   6.2 Future Work and Enhancements 23
4. References 24

**LIST OF FIGURES**

|  |  |
| --- | --- |
| Figure 1. Distribution of Ratings | 11 |
| Figure 2. Confusion matrix | 11 |
| Figure 3. Model accuracies comparison | 15 |

**s**

**LIST OF TABLES**

|  |  |
| --- | --- |
| Table 1. Accuracy Scores | 14 |
| Table 2. Precision, recall and F1-score of the various sentiment classification models | 16 |

# **CHAPTER-1**

# **INTRODUCTION**

1. **INTRODUCTION**

**1.1Introduction to Sentiment Analysis**  
According to natural language processing techniques, this process is termed as sentiment analysis as it is used in order to extract opinions and emotions from text. It has dealt with the identification of whether the text depicts negative, positive, or neutral sentiments. This is highly applied in business sectors for analysing feedback from customers, reviewing one's products, and also posts through social media. Understanding the emotion that causes such responses lets the organizations make informed decisions to be able to enhance customer satisfaction, adapt strategies to what consumers want, and market trends.  
  
**1.2 Importance of Sentiment Analysis**

Sentiment analysis refers to an analytic tool that will be able to enable any organization understand the behaviour and attitude of customers in a data-intensive environment.  
These insights show the level of satisfaction a particular product or brand warrants, the reputation of the brand, and the market trend going on in the market. Therefore, the company uses this kind of analysis as a foundation for making decisions. Sentiment analysis helps organizations respond to customer complaints, reformulate products and services, and reorient their marketing strategies in an attempt to eventually provide a better customer experience and ultimately outdo the competition.

**1.3 Application of Machine Learning to Sentiment Analysis**Algorithms incorporating the techniques of machine learning have exponentially increased the scope of sentiment analysis activities and their effectiveness. The algorithms employed include: Logistic Regression, SVM, Random Forest, and Naive Bayes. These classification techniques naturally classify large volumes of text. In the process of feature extraction and model optimization, advanced techniques like PCA and LDA in the form of dimensionality reduction are included in the workflow. What is more, these applications of machine learning in opinion mining are merely online business monitoring, social media monitoring, customer service, and market research-otherwise worded, decision-making tools.

# **CHAPTER-2**

# **LITERATURE SURVEY**

**2.LITERATURE SURVEY**

A literature survey is a critical review of the studies and academic contributions already existing within the body of knowledge relevant to the issue under investigation. While synthesizing key findings, methodologies, and theories into other research works, it shows states of progress on the one hand and gaps on the other hand. In the process, it puts the current study not only into a warranted context but also into a justified need. This literature review will help researchers place their own work within the broader knowledge already in existence-that is, present exactly how the study builds on, contrasts with, or fills a gap left by previous research.

**2.1 Existing Research in Sentiment Analysis**

This literature on sentiment analysis goes for machine learning models that depend on vectorization techniques like Bag-of-Words and GloVe to evaluate their performance in terms of accuracy and F1-score in multiclass and binary classifications [4]. Current studies focus on the categorization of Amazon product reviews in terms of the sentiment polarity at the sentence and review level [5]. Social media and e-commerce sites have gathered a large number of user-generated data. This is where sentiment analysis becomes all the more important [6]. Studies have shown that product reviews are crucial for buyers. Techniques used currently for sentiment classification are Naive Bayes and Hidden Markov Models [7]. Buyers make more purchases based on online reviews. Hence, businesses make improvements in their products. Valuable results have been witnessed in studies that have compared models such as Word2Vec-CNN with those of Fast Text-CNN based on Amazon data [2]. Techniques like Naïve Bayes and SVM have been used for classifying online product reviews to facilitate customers' decision-making [8]. This review on sentiment analysis puts forward the concept of extracting customer opinions from managing large volumes of feedback by using NLP and Text Analytics [9]. More importantly, research studies which compare the models like BERT and Bag-of-Words and TF-IDF in Amazon reviews show promising results for both multiclass and binary classification [10]. Research studies that are compared against machine learning and deep learning, show promising results that neural networks using Bag-of-Words outperform other methods by utilizing Amazon reviews [11]. Other projects focus on the analysis and visualization of trend directions in Amazon reviews to enhance customer satisfaction [12]. Sentiment analysis on Twitter has also been performed, focusing on the complexity of structuring a tweet, which includes slang and abbreviations. A Python-based approach is in place for sentiment analysis [13]. Naive Bayes techniques handle product reviews using lexicon-based algorithms and computational linguistics to extract opinions of a product [1]. Electronic devices offer new feature vectors when performing Twitter sentiment analysis while dealing with misspellings and slang [14]. Some suggest optimization-based machine learning methods for high accuracy of classification of data in Twitter by using decision trees along with sequential minimal optimization [3]. Authors target the analysis of Chinese product reviews using dependency parsing as a method for sentiment classification [15]. Yet, another article discusses applying real-time sentiment analysis on Amazon reviews in Python using popular web frameworks including Django and React, among others [16]. Discussion has been given on a comparative study of machine learning methods for customer review classification on Amazon and Flipkart [17]. Comparative analysis of SVM and KNN techniques has been used with Amazon product reviews and data from Twitter [18]. Comparative study on lexicon-based and hybrid methods in the deployment of machine learning techniques for the classification of sentiment in social media still evolves [19]. In an effort, research for the best sentiment classification model on clothing e-commerce platforms was pursued in [20]. Recent studies in sentiment analysis emphasize the emerging role of machine learning models and vectorization techniques such as TF-IDF, Bag-of-words and GloVe for enhancing the precision of sentiment classification. Some of the major challenges in handling language complexities are well discussed by Hussein (2018) [21], whereas Chauhan and Sehgal (2017) have used Naive Bayes and SVM for sentiment polarity in product reviews [22]. For classification of customer reviews, Singla et al. in 2017, applied the TF-IDF [23] and for aspect-based sentiment analysis, Mubarok et al in 2017 adopted Naive Bayes [24]. For a review of methods for performing sentiment classification on large-scale e-commerce data Mehta and Pandya, in 2020, conducted a review [25]. The above proposed studies help establish that the combination of machine learning techniques with text vectorization for performing sentiment analysis is suitable.

**2.2 Challenges in Sentiment Classification**

This is due to the very nature of natural language being incorporated into natural language processing; some of the significant challenges will include contextual polarity where the same word can exhibit opposite emotions, sarcasm and irony, which largely disguise the true intent of a sentence, and domain vocabulary specific to terms of an industry or others jargon which use different models. Some of the challenges include imbalanced datasets; some of the sentiment classes tend to be underrepresented, and negation handling with the wrong sentiment detection based on simple keywords. Lastly, with the increasing multilingual content, cross-lingual sentiment analysis is required from models.

**2.3 Motivation for Hybrid Approaches**

The motivation for applying hybrid approaches in sentiment classification is because single models poorly cope with the nuances of natural language complexities. For example, a single model cannot handle deep contextual nuances, domain-specific languages, and subtle expressions such as sarcasm. Hybrid models combine the several strengths of these various techniques by using PCA for reducing dimensionality and LDA to enhance class separability. Moreover, because stacking of classifiers like Random Forests and SVM does provide better generalization, naturally, it enhances accuracy and robustness for working in the real world.

**CHAPTER-3**

**METHODOLOGY**

**3. METHODOLAGY**

**3.1 Input Dataset and Features  
3.1.1 Dataset Description**

The dataset used in this paper is Amazon mobile phone reviews, that contains in total 413,841 reviews. Several attributes represent each review: Product Name, Brand Name, Price, Ratings, Reviews, and Review Votes. This massive dataset from Kaggle gives insight into extensive information about consumer opinions. Such a dataset is aptly suited to an exploration of sentiment classification through machine learning techniques, as the diverse products and user feedback make it quite useful.

**3.1.2 Target Variable (Sentiment)**

The target variable in this study is Sentiment, formed by dividing the reviews rated 1-2 as negative and those rated 4-5 as positive. Reviews that have been rated with 3 are treated as neutral and, therefore, excluded from any analysis since they do not contribute significantly to the knowledge that would be required to make a purchase decision by a customer. In this study, classification into just two categories will ensure that the focus of the analysis is directed towards the extreme ends of customer sentiment.

**3.2 Data Pre-processing**

Data preprocessing is the process of getting raw data into a position friendly to further analysis. Filling in missing values, text cleaning and normalization, finding outliers and removing them, encoding categorical variables, and imputing in appropriate cases-all these provide the machine learning model with a clean, structured, and optimized dataset. Thus, it is only by underlining the relevant features and keeping data consistency that we are able to get the most efficient model, especially for sentiment analysis tasks.

**3.2.1 Handling Missing Data**

Data preprocessing is an important step in the process of preparing a dataset for sentiment analysis. To begin with, removal of duplicates aimed at uniqueness, handling missing values by using the median values to fill gaps in columns Price and Review Votes, and missing Brand Name entries that were replaced with "Unknown." In text normalization, review text was converted into lowercase and special characters and digits were removed from it. Z-scores are used to identify outliers in the data for Price and Review Votes, which are subsequently removed. Categorical variables are encoded with label encoding, and for demonstration purposes, KNN imputation is applied. The columns considered redundant, like Product Name, were dropped, and the cleaned dataset saved to the final analysis.

**3.2.2 Text Preprocessing**

Text pre-processing is an important step which precedes reviewing before actually doing the sentiment analysis. In this step, first, all the review text data are turned into a case fold to ensure uniformity. All special characters and digits are removed since only relevant words would matter. After normalization, clean reviews have been stored in a new column. Further, all the numerical features, such as price and vote for the review contain outliers, and they have been removed using z-scores. This would make sure the dataset is robust and well-prepped for further analyses and the training of models.

**3.2.2.1 Tokenization and Lowercasing**

Tokenization and lowercasing are the two preprocessing steps of utmost importance when one is doing text analysis. By tokenization, the text is broken into individual words or tokens, which allows the frequency analysis along with relevance of each term. Text is converted to lowercase so that the problem of inconsistency is avoided and a word like "Product" would be treated as the same token as "product". This decreases variations and helps in capturing the sentiment expressed well by reviews so eventually contributes to overall accuracy in the sentiment analysis.

**3.2.2.2 Stop word Removal and Punctuation Cleaning**Another critical text preprocessing step would be removing the stop words and cleaning the punctuation to better facilitate the quality of this data for sentiment analysis. In this manner, the noise in the data will be reduced simply by eliminating the stop words such as "and," "the," and "is" because it carries no meaningful value. Also, punctuation cleaning removes special characters and symbols that change the meaning. These combine to narrow down to the more significant words in the reviews and allow the model to better capture sentiment.

**3.3 Feature Extraction  
3.3.1 TF-IDF Vectorization**

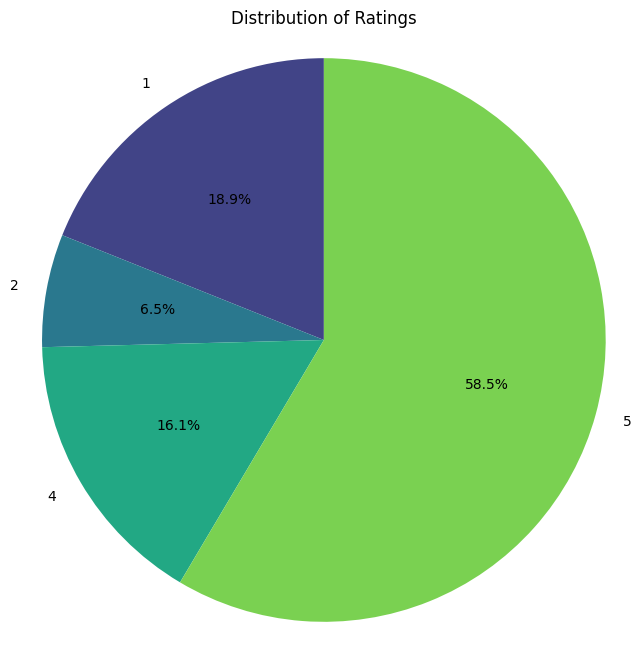
Term Frequency-Inverse Document Frequency (TF-IDF) vectorization was employed in this study to the cleaned reviews in order to make them numerical values for analysis. The TF-IDF vectorizer was initialized with a maximum of 5,000 features for optimizing memory usage. After fitting the vectorizer on the cleaned reviews, we transformed text data into a TF-IDF matrix, which shows the importance of each term within the context of the data set. This matrix then serves as a basis for subsequent sentiment classification tasks.

**3.3.2 Sentiment Assignment**

In this step, we extract the features from our cleaned data using TF-IDF vectorization. We load in the data and fill missing values in the Cleaned\_Reviews column with an empty string. Then we instantiated the TF-IDF vectorizer, so it can transform the reviews into a numerical format. Further, the Rating column values classify the reviews as 'positive' where rating 4-5 is received and ratings between 1-2 are considered 'negative' and thus, excluded. The results of sentiment then are subjected to ensure that a distribution is balanced.

**3.3.2 Rating Distribution Visualization**

A pie chart is generated to understand how the ratings are spread in the data set. A count of the number of reviews for each of the rating categories-from 1 to 5-is computed and sorted. Then it creates a pie chart to indicate the percentage of each rating category and therefore trends in customer feedback end. It uses color scheme in order to be understandable and keeps the aspect ratio equal so it stays circular so that the spread of ratings within the reviews can be easily interpreted.

Figure 3.1

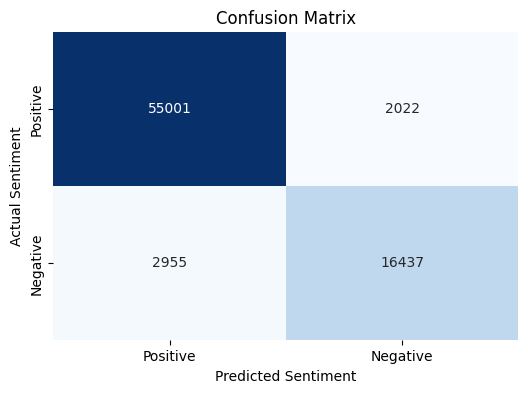


Figure 3.2

**3.4 Model Building**

This stage consists of building a variety of machine learning algorithms to classify the sentiments from the Amazon mobile phone reviews dataset. The varieties of models used are Logistic Regression, Support Vector Machines (SVM), Random Forest, and Naive Bayes, applied to an effective capturing of nuances in customer feedback. Each is trained, which then evaluates it against accuracy and performance metrics and offers insight into consumer opinions and the enhancement of decision-making.

**3.4.1 Logistic Regression**

The paper adopts Logistic Regression as a simple yet very powerful tool for binary sentiment classification. With respect to data cleaning, the final step is to check the existence of missing values of the Cleaned\_Reviews column, after which text data is transformed into numerical representation by way of TF-IDF vectorization. It further makes a split of the data into the training and test datasets by selecting 20% for the purpose of evaluation.  
  
The Logistic Regression model will be fit for up to 1,000 iterations for convergence. Now it has fitted the model to the training data and is predicting for the test set. Finally, the model has given accuracy of 93.49%, thus proving effectiveness towards sentiment classification. The report on the classification, which includes precision, recall, and F1-score, further narrates the model's performance about varied classes of sentiment.

**3.4.2 Support Vector Machine (SVM)**

To enhance the improved sentiment classification SVM is used as well as retaining its rich ability of feature handling for a high dimensional data. The TF-IDF vectorizer with a max 2,000 features is initialized with the cleaned reviews to efficiently convert it to TF-IDF matrix. After the initialization of the above setup, the dataset is then split into training and test set. In an attempt to mitigate the computation time, the SVM model equipped with the linear kernel is learned on a small fraction of the available data. Testing the model on the test set gives an accuracy of 90.79%, implying that the model actually does a very good job at classifying sentiments. The classification report provides more details regarding the performance on different classes of sentiments.

**3.4.3 Random Forest Classifier**

The Random Forest Classifier is utilized here to strengthen the robustness of sentiment classification, taking advantage of the inherent ensemble learning capability. A RandomForestClassifier with 30 estimators is used as an optimization for speed with accuracy. The SGDClassifier is deployed as a fast approximation to a linear SVM. We thus try to tap into both by developing a Voting Classifier that takes the average of the predictions made by both models. We trained the ensemble model on the training set and managed to get it to classify test data at an accuracy level of 95.10%. Its classification report expands further on its performance to each sentiment category.

**3.4.4 Multinomial Naive Bayes**

A Multinomial Naive Bayes classifier was used on the data resulting in achieving an accuracy of 90.15% on a test set. The classification report shows accuracy as 0.88 with recall of 0.71 for the negative class along with an F1-score of 0.78 for 19,392 instances. For the positive class, it showed accuracy as 0.91, recall as 0.97 and the F1-score as 0.94 from 57,023 instances. Thus, the results depict how well the model is performing, particularly about positive sentiment detection.

**3.4.5 Bernoulli Naive Bayes**

We applied the Bernoulli Naive Bayes classifier for sentiment analysis, obtaining an accuracy of 84.73% on the test set. The classification report here had precision in values of 0.74 and recall of 0.62 for the negative class, computed as an F1-score based on 19,392 instances. The positive class held precision at 0.88, recall at 0.93, and the F1-score at 0.90 with 57,023 instances. Overall, Bernoulli Naive Bayes captures positive sentiments well but still has plenty of room for improvement for negative sentiments detection.

**3.4.6 Decision Tree Classifier**

For determination of whether the text is positive or negative, the Decision Tree Classifier was used on the data set. In the test set, impressive accuracy of 95.22% was achieved. Classification report found there was a precision of 0.91 and a recall of 0.90 in the negative class, which led to the F1 score of 0.91 for 19,392 instances. As such, the model obtained precision of 0.97, recall of 0.97, and an F1-score of 0.97 with 57,023 instances by the positive class. The Decision Tree Classifier performed well in identifying both classes of sentiments.

**3.4.7 Stacking Classifier (Hybrid Model with Random Forest and SGD)**

The Stacking Classifier is used to combine the strengths of many models to better classify sentiment. Here, we use an SGD Classifier as a faster surrogate to SVM, with a Random Forest Classifier using 30 estimators to boost the performance of the classifier. We combine predictions using hard voting from the predictions of both classifiers in the ensemble model. Upon evaluation, the ensemble model derives an accuracy of 95.10% and hence is efficient in capturing the depth of nuance and sentiment through the harnessing of the strength of its constituent models. The classification report throws more light into its performances with regards to classes of sentiment.

**3.4.8 Hybrid Model with PCA and LDA**

In this , we have added a Random Forest Classifier and an SVM as base learners with a Logistic Regression model as the final estimator. After the training process, the hybrid model has been achieved to work at 96.81% accuracy, thus reflecting its appropriateness for detecting the nuances of sentiment. The classification report provides informative details about its performance across the different classes of sentiment.

**3.5 Model Evaluation  
3.5.1 Accuracy Score**

A table for an overview of the accuracy of various machine learning models to classify sentiment from reviews of Amazon mobile phones is presented below. The models include classic classifiers such as Logistic Regression, SVM, and Decision Trees and ensemble methods such as Random Forest and a Stacking Classifier. The Hybrid Model has achieved an excellent accuracy of 96.81%, which goes on to prove how flexibility in integration can lead to optimal results in machine learning-based sentiment classification:

|  |  |
| --- | --- |
| **Model** | **Accuracy (%)** |
| Logistic Regression | 93.49 |
| Support Vector Machine (SVM) | 90.79 |
| Random Forest Classifier | 95.10 |
| Multinomial Naive Bayes | 90.15 |
| Bernoulli Naive Bayes | 84.73 |
| Decision Tree Classifier | |  | | --- | |  |  |  | | --- | | 95.22 | |
| Hybrid Model (Stacking Classifier) | |  | | --- | |  |  |  | | --- | | 96.81 | |
| LDA & PCA Hybrid Model | 88.45 |

Table 3.1

**Bar Graph of Model Accuracies**

To better visualize such performance, a bar graph may be devised that will represent the accuracy of each model to easily compare and analyze the effectiveness of the models in sentiment classification.

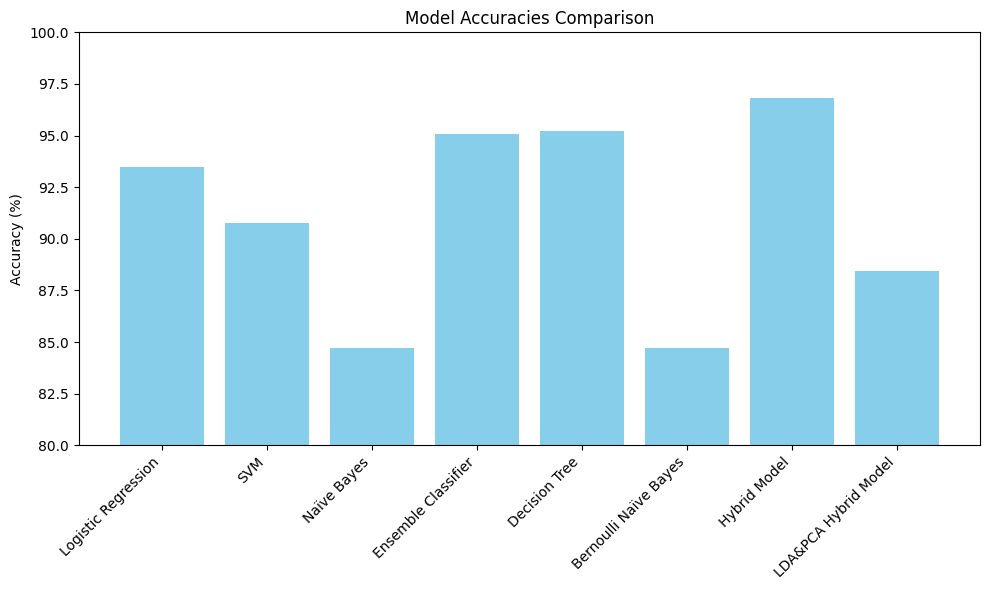


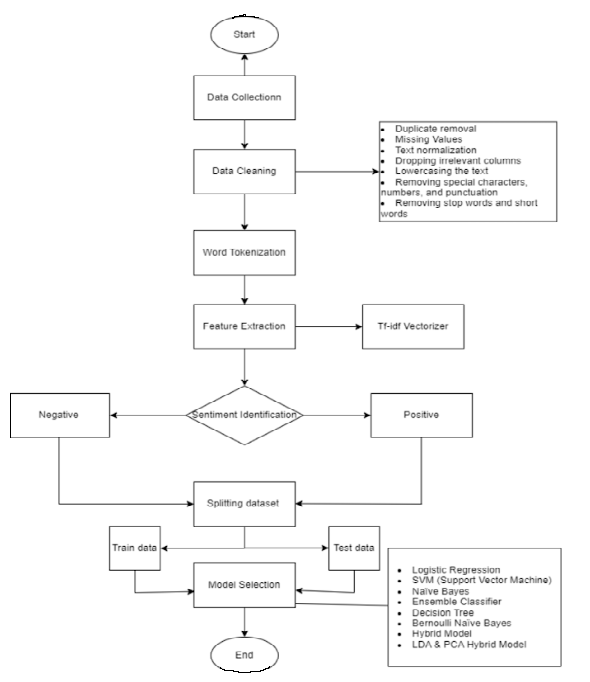
Figure 3.3

**3.5.2 Precision, Recall, and F1-Score**

The following table summarises the precision, recall and F1-score of the various sentiment classification models applied in the study. The different metrics above are useful because they bring out performance levels beyond the simple accuracy that may be achieved. For instance, it classifies correctly the two polar sentiments; there is accuracy in the positivity of the positive predictions, while the ability to identify all relevant instances shows the recall. Lastly, the F1-score is a harmonic mean of precision and recall, thus providing a balanced evaluation.

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Precision** | **Recall** | **F1 Score** |
| Logistic Regression | 0.95 | 0.91 | 0.93 |
| Support Vector Machine (SVM) | 0.90 | 0.89 | 0.89 |
| Random Forest Classifier | 0.92 | 0.94 | 0.93 |
| Multinomial Naive Bayes | 0.89 | 0.91 | 0.90 |
| Bernoulli Naive Bayes | 0.87 | 0.80 | 0.83 |
| Decision Tree Classifier | 0.93 | 0.91 | 0.92 |
| Hybrid Model (Stacking Classifier) | 0.94 | 0.95 | 0.94 |
| LDA & PCA Hybrid Model | 0.86 | 0.88 | 0.87 |

### Table 3.2



**CHAPTER 4**

**IMPLEMENTATION**

**4. IMPLEMENTATION**

**4.1 Environment setup**

Setup Environment This supervised sentiment analysis project requires the basic setup of the environment. To begin, the following python libraries need to be installed. This includes the pandas and NumPy packages for data manipulation and numerical operations, respectively. The machine learning algorithms are supported using scikit-learn. Further, matplotlib and seaborn will be required for visualization. A typical machine learning pipeline will be followed with TF-IDF vectorization, PCA, and LDA.

**4.2 Preprocessing, Dimensionality Reduction, and Model Training Code**

Data preprocessing is a part of the process of implementation, which includes dimensionality reduction and model training. It goes through cleaning and then transforms using TF-IDF vectorization into numerical features that are achievable from the text. Techniques of Dimensionality reduction include PCA and LDA, which help optimize the feature space and hence the model to get better efficiency in the process. The fully preprocessed data will now be used to train the machine learning models that will be used in classifying sentiment-that is, Logistic Regression, Random Forest, and SVM. All this will make sure that the workflow will then be structured for designing efficient models and acquiring accurate predictions.

**CHAPTER 5 EXPERIMENTATION**

**AND RESULT ANALYSIS**

### **5. EXPERIMMENTATION AND RESULT ANALYSIS**

### This chapter gives a critical discussion of the experimentation stage wherein several machine learning models were implemented and tested in terms of their accuracy and other performance metrics. To compare the performance based on individual models and hybrid approaches, this chapter shall present.

### **5.1 Comparison of Model Accuracies**

### Those models such as Logistic Regression, SVM, Random Forest, Naive Bayes, and Decision Tree have been tested on the same sentiment dataset during experimentation. Their accuracy is calculated, which are 93.49% of Logistic Regression, 90.79% of SVM, and 95.22% by the decision tree model. Here also those models are compared using a classification report which gives the insights for better decision-making such as precision, recall, and F1-score values.

### Generally speaking, as far as the performance of Decision Tree and Random Forest was considered to be superiorly greater than the more elementary models. Precision and recall were higher for both positive and negative sentiment classification. Still, although Naive Bayes models were more computationally efficient and straightforward, their precision on negative reviews was lower compared with the performance of more complex models.

### **5.2 Hybrid Model Performance compared to Individual Models**

### The hybrid models, for example, Stacking Classifier, perform better than the individual models. Hybrid models put together then, Random Forest, SVM, and Logistic Regression-to achieve the maximum possible accuracy, 96.81%. Methods such as PCA combined with dimension reduction and LDA for class separability allowed hybrid models to take an advantage of capturing the sentiment patterns much better than the standalone classifiers.

### The stacking models in the classifier resulted in higher precision, recall, and F1-scores all the way, especially between the positive and negative review. The hybrid models reduced the possibility of misclassifications through the strengths of each of the base models. This performance represents the benefits of using hybrid machine learning models on sentiment analysis, especially in large and diverse datasets.

### Overall, the hybrid models do outperform individual models, establishing their efficiency in addressing complex sentiment analysis tasks. Therefore, the combined approach would serve to surpass the performance of individual machine learning models toward superior predictive accuracy and improved generalization across unseen data.

### **CHAPTER 6**

### **CONCLUSION**

**6. CONCLUSION**

It was able to demonstrate varied machine learning models, namely Logistic Regression, SVM, Random Forest, Naive Bayes, and Decision Trees, on Amazon mobile phone reviews with the ability to classify their opinions. From the aforementioned results, it's very clear that hybrid models perform better than individual models and have achieved an accuracy of 96.81% by the Stacking Classifier as the best among them. This robust result has been determined by using the ensemble techniques of bagging and incorporating PCA and LDA for dimensionality reduction. These results do depict the hybrid approaches of machine learning as possibly building the potentiality toward better performance in sentiment classification.

**6.1 Key Findings**

Stackifier, which is another hybrid model, also seemed to outperform all the individual classifiers, with Logistic Regression, SVM, and Decision Tree faring the worst. High accuracy of 96.81% was achieved. Dimension reduction techniques were also significant by applying PCA and LDA. Traditional models, such as Naive Bayes and Decision Tree, performed reasonably but failed to achieve the integration accuracy property by hybrid approaches. The evaluation further brought into perspective the ability of the models to detect positive sentiments while failing to detect negative sentiments.

**6.2 Future Work and Enhancements**

Future work should make use of more advanced techniques such as BERT or LSTM deep learning for the capturing of nuances of sentiment. Enhanced feature engineering by making use of word embeddings or topic modeling by LDA could also further be applied to enhance the classification accuracy. Hyperparameter optimization, as well as a set of a diverse set of datasets of different product categories, would generalize further. Introducing real-time sentiment analysis systems and testing them on multilingual datasets can further enlarge the scope and potential usage of models for specific application scenarios.

**CHAPTER 7**

**REFRENCES**

**7.REFERENCES**

[1] R. Safrin, K. R. Sharmila, T. S. S. Subangi, and E. A. Vimal, “SENTIMENT ANALYSIS ON ONLINE PRODUCT REVIEW,” 2017. [Online]. Available: www.irjet.net

[2] M. Soundararajan and S. Joseph’, “Text Analysis for Product Reviews for Sentiment Analysis using NLP Methods,” *Int. J. Eng. Trends Technol.*, vol. 47, 2017, [Online]. Available: http://www.ijettjournal.org

[3] A. Naresh and P. Venkata Krishna, “An efficient approach for sentiment analysis using machine learning algorithm,” *Evol. Intell.*, vol. 14, no. 2, pp. 725–731, Jun. 2021, doi: 10.1007/s12065-020-00429-1.

[4] A. S. M. AlQahtani, “Product Sentiment Analysis for Amazon Reviews,” *Int. J. Comput. Sci. Inf. Technol.*, vol. 13, no. 3, pp. 15–30, Jun. 2021, doi: 10.5121/ijcsit.2021.13302.

[5] A. Rashid and C. Huang, “Sentiment Analysis on Consumer Reviews of Amazon Products,” *Int. J. Comput. Theory Eng.*, vol. 13, no. 2, pp. 35–41, 2021, doi: 10.7763/IJCTE.2021.V13.1287.

[6] X. Fang and J. Zhan, “Sentiment analysis using product review data,” *J. Big Data*, vol. 2, no. 1, Dec. 2015, doi: 10.1186/s40537-015-0015-2.

[7] N. Sultana, P. Kumar, M. R. Patra, S. Chandra, and S. K. S. Alam, “SENTIMENT ANALYSIS FOR PRODUCT REVIEW,” *ICTACT J. SOFT Comput.*, p. 3, 2019, doi: 10.21917/ijsc.2019.0266.

[8] A. Shah, “SENTIMENT ANALYSIS OF PRODUCT REVIEWS USING SUPERVISED LEARNING.”

[9] Y. Basani, H. V. Sibuea, S. Ida Patona Sianipar, and J. Presly Samosir, “Application of Sentiment Analysis on Product Review E-Commerce,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jun. 2019. doi: 10.1088/1742-6596/1175/1/012103.

[10] A. N. Jebaseeli and E. Kirubakaran, “A Survey on Sentiment Analysis of (Product) Reviews,” 2012.

[11] P. Poomka, N. Kerdprasop, and K. Kerdprasop, “Machine Learning Versus Deep Learning Performances on the Sentiment Analysis of Product Reviews,” *Int. J. Mach. Learn. Comput.*, vol. 11, no. 2, pp. 103–109, 2021, doi: 10.18178/ijmlc.2021.11.2.1021.

[12] D. R. Kawade and D. K. S. Oza, “Sentiment Analysis: Machine Learning Approach,” *Int. J. Eng. Technol.*, vol. 9, no. 3, pp. 2183–2186, Jun. 2017, doi: 10.21817/ijet/2017/v9i3/1709030151.

[13] B. Gupta, M. Negi, K. Vishwakarma, G. Rawat, and P. Badhani, “Study of Twitter Sentiment Analysis using Machine Learning Algorithms on Python,” *Int. J. Comput. Appl.*, vol. 165, no. 9, pp. 29–34, May 2017, doi: 10.5120/ijca2017914022.

[14] S. Shah, M. Akshata Bhat, M. Sumitra Singh, M. Arya Chavan, and M. Aryan Singh, “INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS) SENTIMENT ANALYSIS”, doi: 10.58257/IJPREMS33384.

[15] H. Wang, L. Liu, W. Song, and J. Lu, “Feature-based sentiment analysis approach for product reviews,” *J. Softw.*, vol. 9, no. 2, pp. 274–279, 2014, doi: 10.4304/jsw.9.2.274-279.

[16] R. Senthamil Selvi, B. S. Lakshmi, D. J. Jovita, and A. Professor, “Classification and Summarization of Amazon Product Reviews,” *Int. Res. J. Eng. Technol.*, 2021, [Online]. Available: www.irjet.net

[17] S. Tripathy *et al.*, “Sentiment Analysis on Product Reviews Using Supervised Learning Techniques,” *Int. Res. J. Eng. Technol.*, 2023, [Online]. Available: www.irjet.net

[18] P. Verma, “THE SENTIMENTAL ANALYSIS ON PRODUCT REVIEWS OF AMAZON DATA USING THE HYBRID APPROACH,” *Int. Res. J. Eng. Technol.*, 2019, [Online]. Available: www.irjet.net

[19] S. Aftab, S. Shah Muhammad, S. Awan, M. Ahmad, and S. Ahmad, “Machine Learning Techniques for Sentiment Analysis: A Review,” *Int. J. Multidiscip. Sci. Eng.*, vol. 8, no. 3, 2017, [Online]. Available: www.ijmse.org

[20] M. Loukili, F. Messaoudi, and M. El Ghazi, “Sentiment Analysis of Product Reviews for E-Commerce Recommendation based on Machine Learning,” *Int. J. Adv. Soft Comput. its Appl.*, vol. 15, no. 1, pp. 1–13, 2023,

[21] D. M. E. D. M. Hussein, "A survey on sentiment analysis challenges", Journal of King Saud University-Engineering Sciences, vol. 30, no. 4, pp. 330-338, 2018.

[22] C. Chauhan and S. Sehgal, "Sentiment analysis on product reviews", 2017 International Conference on Computing Communication and Automation (ICCCA), pp. 26-31, 2017, May.

[23] Z. Singla, S. Randhawa and S. Jain, "Sentiment analysis of customer product reviews using machine learning", 2017 international conference on intelligent computing and control (I2C2), pp. 1-5, 2017, June.

[24] M. S. Mubarok, Adiwijaya and M. D. Aldhi, "Aspect-based sentiment analysis to review products using Naïve Bayes", AIP Conference Proceedings, vol. 1867, no. 1, pp. 020060, 2017, August.

[25] P. Mehta and S. Pandya, "A review on sentiment analysis methodologies practices and applications", International Journal of Scientific and Technology Research, vol. 9, no. 2, pp. 601-609, 2020.