Project Report on

**Sentiment Analysis on Amazon Alexa Reviews**

*Submitted* by

|  |
| --- |
| Bhavesh Khandelwal (Roll No.: 59)  Soham Khopkar (Roll No.: 60)  Aditya Kirti (Roll No.: 61) |

Under the guidance of

**Mr. Vijaykumar Yele**

Assistant Professor

****

**INFORMATION TECHNOLOGY DEPARTMENT**

**Thakur College of Engineering & Technology**

Thakur Village, Kandivali (East), Mumbai-400101

**(Academic Year 2024-25)**

**Abstract**

This research addresses the increasing need for effective sentiment analysis of Amazon Alexa reviews by comparing the performance of three machine learning methods: BERT, LSTM, and Random Forest. Leveraging a dataset of verified customer reviews, the study applies comprehensive preprocessing techniques including text cleaning, tokenization, and class imbalance handling using ADASYN to prepare the data for analysis. The results demonstrate that BERT achieves the highest accuracy of 94.2%, attributed to its advanced bidirectional transformer architecture that captures rich contextual and semantic nuances within the reviews, enabling superior understanding of complex language patterns such as sarcasm and domain-specific terminology. The LSTM model follows with an accuracy of 89.7%, benefiting from its ability to process sequential data and capture temporal dependencies, though it is somewhat limited in handling very long-range context compared to BERT. The Random Forest classifier, while simpler and more interpretable, attains a lower accuracy of 82.1%, reflecting challenges in capturing the subtle contextual information present in natural language text. Evaluation metrics including precision, recall, and F1-score further confirm BERT’s balanced and robust performance, especially in detecting negative sentiments critical for identifying user dissatisfaction. This comparative analysis provides valuable insights into the trade-offs between model complexity, interpretability, and accuracy in sentiment classification tasks for voice-assisted technology reviews. The findings suggest that transformer-based models like BERT are highly effective for nuanced sentiment analysis in this domain, while traditional models may still offer practical benefits in resource-constrained environments. The study lays a foundation for future research exploring hybrid architectures and multilingual datasets to enhance sentiment analysis capabilities for voice-activated devices.

**CONTENTS**

* + - * List of Figures i
      * Abbreviations and symbols ii

**Chapter No. Topic Pg. No.**

**Chapter 1 Overview 05**

**1.1 Introduction**

1.2 Background

1.3 Importance of the Project

1.4 Perspective of stakeholders and customers

1.5 Objectives and Scope of the project

**1.6 Summary**

**Chapter 2 Literature Survey 07**

**2.1 Introduction**

2.2 Literature Survey (Limitation Existing system or research gap)

2.3 Problem Statement

**2.4 Summary**

**Chapter 3 Planning & Design 10**

**3.1 Introduction**

3.2 Resources, Tools required, etc.)

3.3 Proposed System

3.4.1 Block Diagram/Flowchart

3.4.3 Methodology (approach to solve the problem)

**3.5 Summary**

**Chapter 4 Paper based on 5 literatures 16**

**Research Paper**

**Chapter 5 References 17**

**CONTENTS**

* List of Figures

Figure 1: Flowchart of Workflow

* Abbreviations and symbols

|  |  |  |
| --- | --- | --- |
| **Abbreviation** | **Full Form** | **Description** |
| BERT | Bidirectional Encoder Representations from Transformers | A transformer-based deep learning model for NLP tasks |
| LSTM | Long Short-Term Memory | A type of recurrent neural network capable of learning long-term dependencies |
| RF | Random Forest | An ensemble learning method using multiple decision trees |
| NLP | Natural Language Processing | Field of AI focused on interaction between computers and human language |
| F1-Score | Harmonic mean of Precision and Recall | Metric balancing precision and recall in classification tasks |
| TP | True Positive | Correctly predicted positive instances |
| TN | True Negative | Correctly predicted negative instances |
| FP | False Positive | Incorrectly predicted positive instances |
| FN | False Negative | Incorrectly predicted negative instances |
| CSV | Comma-Separated Values | File format used to store tabular data |
| SVM | Support Vector Machine | A supervised machine learning model for classification |

**Chapter 1: Overview**

**1.1 Introduction**

The widespread use of voice-controlled smart devices has revolutionized the way users interact with technology. Amazon Alexa, one of the top players in the field of intelligent personal assistants, has been the subject of much consumer as well as academic attention. With such devices becoming an integral part of daily life, the analysis of user sentiment through review analysis has emerged as a central area of research. Sentiment analysis, a natural language processing (NLP) technique, allows companies to extract useful information about user experience, user preferences, and potential areas of product improvement1.

**1.2 Background**

Voice technologies like Amazon Alexa highly rely on customer feedback to enhance their performance, user interface quality, and overall functionality capabilities. Customer reviews form an invaluable pool of qualitative data that can identify subtleties of user impressions, challenges, and expectations. However, manually reading them is not only time-consuming but also susceptible to subjective conclusions. This problem gave birth to advanced machine learning methods with the objective of sentiment classification automation1.

The evolution of machine learning techniques such as BERT (Bidirectional Encoder Representations from Transformers), LSTM (Long Short-Term Memory), and Random Forest has created new opportunities for analyzing complex user sentiments with greater accuracy. These methods have demonstrated varying capabilities in capturing contextual nuances, sarcasm, and domain-specific language that are often present in user reviews1.

**1.3 Importance of the Project**

Despite the growing popularity of Amazon Alexa, there remains a major lack of systematic examination of user sentiments under different machine learning approaches. The majority of previous works usually focus on a single sentiment analysis technique without comprehensive comparative studies. This project bridges this critical gap by providing a comparative study of different machine learning models for correct classification of sentiments from user reviews of Amazon Alexa1.

The importance of this research lies in its potential to significantly improve how companies understand and respond to customer feedback. By determining the most effective machine learning approach for sentiment analysis of voice-assisted technology reviews, companies can better identify user satisfaction patterns, product issues, and potential areas for enhancement. This ultimately leads to improved product development, marketing strategies, and customer satisfaction1.

**1.4 Perspective of stakeholders and customers**

From the stakeholders' perspective, accurate sentiment analysis provides valuable business intelligence that can inform strategic decision-making. For Amazon and similar technology companies, understanding customer sentiments helps prioritize feature development, address concerns promptly, and optimize marketing efforts. Product managers can identify specific aspects of the product that receive positive or negative feedback, enabling targeted improvements1.

From the customers' perspective, effective sentiment analysis ensures their feedback is understood and appropriately addressed. When companies can accurately interpret the sentiments expressed in reviews, they can respond more effectively to customer needs and preferences. This creates a feedback loop where customer opinions genuinely influence product development, leading to better user experiences and increased customer satisfaction1.

**1.5 Objectives and Scope of the project**

The overall objectives of this research are:

* To create and implement sentiment analysis models using three various machine learning techniques: BERT (Bidirectional Encoder Representations from Transformers), LSTM (Long Short-Term Memory), and Random Forest1.
* To perform a thorough performance comparison of these models in sentiment classification of user reviews of Amazon Alexa1.
* To compare the models using different performance metrics, including accuracy, precision, recall, and F1-score1.
* To identify the strengths and weaknesses of every machine learning technique applied to sentiment analysis in the context of voice-aided technology reviews1.

The scope of this project is specifically focused on sentiment analysis of Amazon Alexa user reviews based on an aptly selected dataset of verified customer comments. The research is limited to English-language reviews on the Amazon platform, with regard to sentiment classification into positive, negative, and neutral categories1.

**1.6 Summary**

This chapter introduced the background and importance of sentiment analysis for Amazon Alexa reviews, highlighting how voice-controlled smart devices have revolutionized user interaction with technology. The growing need for automated sentiment analysis stems from the impracticality of manually analyzing vast amounts of customer reviews. The research addresses a significant gap in comparative studies of machine learning approaches for sentiment analysis, with objectives focused on implementing and comparing BERT, LSTM, and Random Forest models. The project aims to deliver insights that benefit both stakeholders, who can leverage the findings for business intelligence and product improvement, and customers, whose feedback can be more accurately understood and addressed. The scope is clearly defined as focusing on English-language Amazon Alexa reviews with classification into positive, negative, and neutral sentiments.

**Chapter 2: Literature Survey**

**2.1 Introduction**

Sentiment analysis has been one of the essential areas in natural language processing (NLP) that allows us to find out subjective information from text sources. Advances in machine learning and deep learning have prompted the construction of models such as BERT, LSTM, and Random Forest that have proven very promising in achieving high accuracy for sentiment classification. This chapter reviews existing research relating to sentiment analysis techniques with focus on the usage of BERT, LSTM, and Random Forest models1.

The field of sentiment analysis has evolved significantly over the past decade, moving from simple lexicon-based approaches to sophisticated machine learning and deep learning techniques. This evolution has been driven by the increasing availability of large datasets, computational resources, and algorithmic innovations. Understanding the current state of research in this domain is crucial for identifying potential improvements and research directions1.

**2.2 Literature Survey**

BERT (Bidirectional Encoder Representations from Transformers) revolutionized natural language processing (NLP) operations as it was able to capture context using bidirectional training. Its application has been noted in many studies in sentiment analysis:

* The Aspect-Based Sentiment Analysis using BERT research evaluated the performance of the model in sentiment detection depending on specific aspects of text data, demonstrating better performance compared to traditional methods1.
* Research on the use of the Deep Learning-Based BERT Model for sentiment analysis underscored its versatility in a wide range of sentiment analysis tasks with pre-trained embeddings1.
* Hybrid strategies, which include the integration of BERT models fine-tuned using extra layers, have provided higher classification accuracy1.

Long Short-Term Memory (LSTM) networks are extensively employed for handling sequential data as they can learn long-term dependencies:

* Studies like LSTM-Based Sentiment Analysis for Stock Price Forecast and Twitter Sentiment Analysis Using Combined LSTM-CNN Models demonstrated that the model could successfully deal with large-scale sentiment data1.
* Progressive models, such as Bidirectional LSTM and ensemble methods (e.g., CNN-LSTM), have improved sentiment prediction by successfully capturing contextual interdependencies1.

Random Forest is an extremely popular ensemble learning method that uses multiple decision trees to create stable classification results:

* Comparative research between Random Forest and other classifiers (for example, Naïve Bayes, SVM) demonstrated its competitiveness within sentiment analysis1.
* Tasks like Sentiment Analysis on Tokopedia Product Reviews and hybrid approaches such as Random Forest-SVM highlighted its efficiency and scalability1.

Comparative analysis of these models has revealed varying performance metrics across different datasets and applications. The research indicates that BERT-based models generally outperform traditional approaches, with accuracy rates of up to 93% on Twitter sentiment datasets, while LSTM models achieve around 85% accuracy on IMDB and Amazon product review datasets. Random Forest models typically achieve 74-81% accuracy depending on the dataset characteristics1.

**2.3 Problem Statement**

Based on the literature review, several key research gaps have been identified that limit the effectiveness of current sentiment analysis approaches:

1. Domain-Specific Optimization for BERT:
   * There is little research on fine-tuning BERT for the specific characteristics of voice-aided technology reviews (e.g., Alexa)1.
   * Further research is needed to successfully adapt BERT for this specific domain, which could improve the accuracy of sentiment analysis1.
2. Enhanced Class Imbalance Strategies:
   * Existing methods of handling class imbalance in sentiment analysis datasets may fall short1.
   * Research and leverage more sophisticated methods (e.g., resampling methods, cost-sensitive learning) to reduce the effects of unbalanced data1.
3. Quantifying Trade-offs Between Model Complexity and Resource Utilization:
   * Ineffective understanding of model complexity (e.g., BERT) real-world performance vs. computational resource trade-offs1.
   * Research to quantify such trade-offs as well as to attempt to improve the accuracy-resource balance through techniques like model compression or pruning1.

The central research problem is to devise a methodology which can effectively address the context and subtlety involved in user opinions about voice-assisted technologies like Amazon Alexa. Standard sentiment analysis approaches are likely to face challenges related to context, sarcasm, and domain-specific terms, particularly in technology product reviews. Additionally, the performance of different machine learning algorithms may vary significantly based on issues related to feature extraction, model development, and training strategies1.

**2.4 Summary**

This chapter provided a comprehensive review of existing research on sentiment analysis techniques, focusing specifically on BERT, LSTM, and Random Forest models. The literature survey revealed that while each model has demonstrated effectiveness in various sentiment analysis applications, there remain significant research gaps. These include the need for domain-specific optimization of BERT for voice-aided technology reviews, more sophisticated strategies for handling class imbalance in datasets, and better understanding of the trade-offs between model complexity and resource utilization. The comparative analysis of model performance across different studies highlighted that BERT-based approaches generally outperform other methods, though each model has distinct strengths and limitations. The identified problem statement emphasizes the need for a methodology that can effectively capture the nuances and context in Amazon Alexa user reviews, addressing challenges related to feature extraction, model development, and training strategies. This literature review establishes the foundation for the proposed comparative study and highlights its potential contribution to advancing sentiment analysis techniques for voice-assisted technology reviews.

**Chapter 3: Planning & Design**

**3.1 Introduction**

This chapter outlines the comprehensive approach to planning and designing the sentiment analysis system for Amazon Alexa reviews. It details the resources and tools required for implementation, presents the proposed system architecture, and explains the methodology adopted to address the identified research gaps. The planning and design phase is critical for ensuring that the sentiment analysis models are implemented effectively and that the comparative study yields meaningful insights into the performance of BERT, LSTM, and Random Forest approaches1.

The design process was guided by the need to create a fair and robust comparison between the three machine learning models while ensuring that the specific challenges of analyzing sentiment in voice-assisted technology reviews are addressed. This chapter provides a detailed roadmap of how the sentiment analysis system was conceptualized and structured to achieve the research objectives outlined in Chapter 11.

**3.2 Resources, Tools required, etc.**

The tools and technologies employed for the execution of this sentiment analysis project include:

* **Python**: Main programming language for data pre-processing, model development and testing1.
* **TensorFlow & Keras**: Deep learning libraries employed for implementing the BERT and LSTM models, providing a high-level API for neural networks and supporting GPU acceleration for faster training1.
* **Scikit-learn**: Machine learning library used primarily for implementing the Random Forest classifier and for evaluation metrics calculation1.
* **NLTK and SpaCy**: Natural Language Processing libraries utilized for text preprocessing tasks such as tokenization, stemming, and stop word removal1.
* **Pandas and NumPy**: Data manipulation and numerical computation libraries used for efficient data handling and transformation1.
* **Matplotlib and Seaborn**: Visualization libraries employed for creating performance comparison charts and confusion matrices1.
* **Transformers Library (Hugging Face)**: Used for accessing pre-trained BERT models and for fine-tuning them on the Amazon Alexa reviews dataset1.
* **CUDA**: For GPU acceleration during model training, particularly beneficial for the computationally intensive BERT and LSTM models1.
* **Jupyter Notebook**: Interactive development environment used for incremental code development, visualization, and documentation1.

Hardware resources included:

* High-performance computing environment with GPU capabilities for training the deep learning models
* Sufficient storage for the dataset and model checkpoints
* Memory resources adequate for handling the data preprocessing and model training operations1

**3.3 Proposed System**

The proposed system consists of three sentiment analysis models—BERT, LSTM, and Random Forest—each implemented with specific configurations to optimize their performance on the Amazon Alexa reviews dataset. The system follows a structured pipeline that includes data collection, preprocessing, model training, evaluation, and comparative analysis1.

The sentiment analysis system operates as follows:

1. **Data Collection and Preparation**: Amazon Alexa reviews are collected from publicly available sources, ensuring a representative sample of user opinions. The data includes text reviews, user ratings, and sentiment tags (positive, negative, or neutral)1.
2. **Data Preprocessing**: The collected reviews undergo comprehensive preprocessing, including cleaning (removing unwanted characters and HTML tags), tokenization, stop word removal, lowercasing, and padding (for LSTM). The ADASYN (Adaptive Synthetic Sampling Method) is employed to address class imbalance by generating synthetic samples for underrepresented sentiment classes1.
3. **Model Implementation**:
   * **BERT Model**: A pre-trained BERT model is fine-tuned for sentiment classification of Amazon Alexa reviews, leveraging transfer learning to adapt to the specific domain1.
   * **LSTM Model**: A Bidirectional LSTM network with embedding, dropout, and dense layers is implemented for sequential processing of the review text1.
   * **Random Forest Model**: An ensemble of decision trees is trained with optimized hyperparameters to classify sentiments based on features extracted from the preprocessed text1.
4. **Model Evaluation**: Each model is evaluated using a consistent set of metrics including accuracy, precision, recall, F1-score, and confusion matrices. Cross-validation techniques are employed to ensure robust performance assessment1.
5. **Comparative Analysis**: The performance of the three models is systematically compared to identify their respective strengths and weaknesses in the context of Amazon Alexa review sentiment analysis1.

**3.4.1 Block Diagram/Flowchart**

The sentiment analysis system follows a structured workflow represented by the following block diagram:

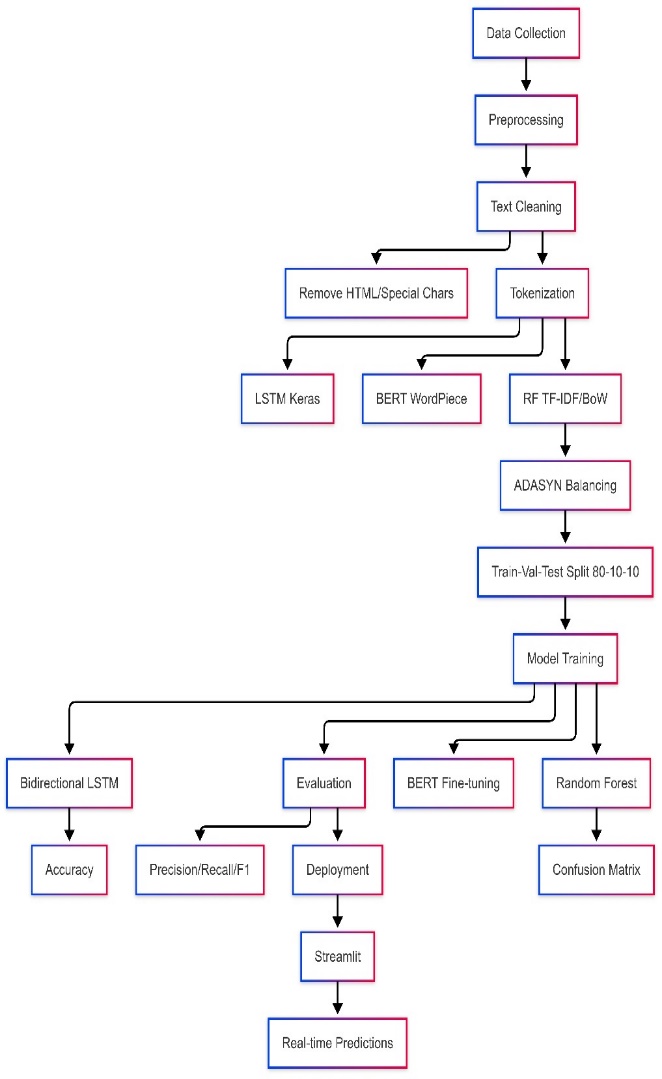


Figure 1: Flowchart of Workflow

**3.4.2 Methodology (approach to solve the problem)**

The methodology adopted for this sentiment analysis project involves a comprehensive approach to data handling, model implementation, and evaluation:

1. **Data Collection Methodology**:
   * The data used in this study is Amazon Alexa reviews from publicly available sources. The data consist of text reviews, user ratings, and sentiment tags that indicate sentiment, positive or negative1.
   * Data were collected through scraping product pages on Amazon, hence acquiring a heterogeneous and representative set of user opinions1.
   * Ethical guidelines were followed strictly by anonymizing user data and respecting Amazon's terms of service1.
2. **Data Preprocessing Methodology**:
   * **Cleaning**: Extracted unwanted characters, HTML tags, and special characters from the reviews1.
   * **Tokenization**: Transformed text into tokens or words1.
   * **Stop word removal**: The standard English stop words (e.g., "the," "is," "and") that do not assist in sentiment analysis were eliminated1.
   * **Lowercasing**: Everything was lowercased in order to promote consistency1.
   * **Padding**: For the LSTM model, sequences were padded to a uniform length, a requirement for effective batch processing1.
   * **Balancing**: Solved class imbalance by employing ADASYN (Adaptive Synthetic Sampling Method) to generate new minority class samples in order to balance positive and negative sentiment representation1.
3. **Model Implementation Methodology**:
   * **BERT Model**:
     + Model Selection: Utilized a pre-trained BERT model fine-tuned for sentiment classification1.
     + Training: Transfer learning approach was used whereby the fine-tuning step was conducted on the preprocessed Amazon Alexa review data. The parameters of training were adjusted with respect to acquire the correct performance1.
   * **LSTM Model**:
     + Model Selection: Employed a Bidirectional LSTM network with an embedding layer, dropout layers, and dense layers for sentiment classification1.
     + Training: The data that was preprocessed was trained on the LSTM model using techniques like early stopping and a decreased learning rate during stagnation periods to avoid overfitting and enhance convergence1.
   * **Random Forest Model**:
     + Model Selection: Used a Random Forest classifier, and utilized hyperparameter tuning in order to change the number of trees, depth, and a few other parameters1.
     + Training: The Random Forest model was trained on the preprocessed data using cross-validation methods to ensure robustness and generalizability1.
4. **Evaluation Methodology**:
   * The performance of all models was assessed based on the following measures:
     + Accuracy: It quantifies the model's overall accuracy1.
     + Precision: Calculates the percentage of positive sentiment instances correctly forecasted out of all instances assigned as positive1.
     + Recall: Indicates the proportion of correctly predicted positive sentiments to all the positive instances available1.
     + F1-Score: Assesses the harmonic mean between precision and recall, thus providing a balanced view of the model's effectiveness1.
     + The Confusion Matrix: It provides a fine-grained breakdown of true positives, true negatives, false positives, and false negatives, hence enabling a clear assessment of model performance1.
5. **Comparative Analysis Methodology**:
   * A systematic comparison of the three models was conducted based on the evaluation metrics1.
   * Additionally, aspects such as computational resource requirements, training time, and model complexity were considered to provide a comprehensive assessment of each approach1.
   * The analysis specifically focused on identifying which model performs best for different aspects of sentiment analysis in the context of Amazon Alexa reviews1.

**3.5 Summary**

This chapter outlined the planning and design approach for the comparative sentiment analysis of Amazon Alexa reviews using BERT, LSTM, and Random Forest models. The comprehensive methodology includes data collection from public Amazon sources, thorough preprocessing steps like cleaning, tokenization, and class imbalance handling using ADASYN, and specific implementation strategies for each model. The BERT model utilizes transfer learning through fine-tuning, the LSTM model employs bidirectional architecture with dropout layers, and the Random Forest model leverages hyperparameter tuning for optimization. Evaluation follows a structured approach using accuracy, precision, recall, F1-score, and confusion matrices. The system architecture follows a clear workflow from data collection through preprocessing, model development, evaluation, and comparative analysis. Tools including Python, TensorFlow, Scikit-learn, and the Transformers library support implementation.

**Chapter 4: Paper Based on Literature Review**

**4.1 Introduction**

This chapter presents a detailed overview and analysis of the research paper titled "Sentiment Analysis with Natural Language Processing on Amazon Alexa Reviews." The paper explores the application of traditional machine learning models to classify user sentiments expressed in Amazon Alexa customer reviews. The study emphasizes the importance of automated sentiment analysis in understanding customer feedback to improve product development and customer satisfaction.

**4.2 Paper Overview**

The paper investigates the performance of various machine learning algorithms—including Decision Tree Classifier, Logistic Regression, Random Forest, Support Vector Machines (SVM), XGBoost, and Gradient Boosting—for sentiment classification of Amazon Alexa reviews. It adopts a structured approach involving comprehensive data preprocessing steps such as text cleaning, stemming, and vectorization. The dataset used is sourced from Kaggle and contains verified reviews with binary sentiment labels (positive or negative). The study evaluates models based on accuracy, precision, recall, and F1-score, with Random Forest achieving the highest accuracy of 93.65%. Visualizations and confusion matrices are used to interpret model performance.

**4.3 Paper Abstract**

The paper focuses on sentiment analysis as a critical technique for extracting user feedback insights from Amazon Alexa reviews. It applies traditional machine learning models after thorough preprocessing to classify sentiments as positive or negative. Among the evaluated models, Random Forest demonstrated superior accuracy (93%), balancing interpretability and computational efficiency. The research highlights the effectiveness of conventional machine learning methods and lays the groundwork for future exploration of deep learning and hybrid approaches to handle more complex datasets.

**4.4 Paper Contribution**

The key contributions of the paper include:

* Developing an efficient and scalable sentiment analysis framework specifically for Amazon Alexa reviews.
* Achieving high classification accuracy while reducing false negatives to capture critical negative feedback.
* Demonstrating the practical applicability of traditional machine learning techniques in real-world e-commerce sentiment analysis.
* Providing a comparative evaluation of multiple machine learning models, highlighting the strengths of ensemble methods like Random Forest.
* Establishing a foundation for future research into advanced deep learning models and hybrid approaches for sentiment analysis.

**4.5 Summary**

This chapter summarized the essential elements of the research paper on sentiment analysis of Amazon Alexa reviews using traditional machine learning algorithms. The paper effectively addresses the challenge of automated sentiment classification by employing comprehensive preprocessing and evaluating multiple models. Random Forest emerged as the best-performing model, achieving a balance between accuracy and computational efficiency. The study contributes valuable insights into the practical use of machine learning for sentiment analysis in e-commerce, with recommendations for future work involving more advanced deep learning techniques.

**Chapter 5: References**

1. Aspect-Based Sentiment Analysis Using BERT
2. Research on the Application of Deep Learning-based BERT Model in Sentiment Analysis
3. Sentiment analysis classification system using hybrid BERT models
4. Improving Bert Performance For Aspect-Based Sentiment Analysis
5. Improving the performance of aspect based sentiment analysis using f ine-tuned Bert Base Uncased model