**TERNA ENGINEERING COLLEGE, NERUL, NAVI MUMBAI**

### Computer Engineering Department

“Sentiment Analysis of Product Reviews”

##### Subject Mini Project Report of

**Big Data Analytics**

###### Submitted in partial fulfillment of the requirements for

**Final Year of Computer Engineering**

|  |  |  |
| --- | --- | --- |
| Name | ID | Roll No |
| Aaradhya Bhandare | TU3F2122107 | B23 |
| Tejas Gotavade | TU3F2122115 | B33 |
| Tanmay Darekar | TU3F2122119 | B37 |

**Under the Guidance of**

**Prof. Pramila Mate**

**Department of Computer Engineering**

# TERNA ENGINEERING COLLEGE

Nerul (W), Navi Mumbai 400706

(University of Mumbai) (2024-2025)

**Internal Approval Sheet**



# TERNA ENGINEERING COLLEGE, NERUL

#### Department of Computer Engineering

Academic Year 2024-25

**CERTIFICATE**

This is to certify that the project entitled **“Sentiment Analysis of Product Reviews”** is a bonafide work of

|  |  |  |
| --- | --- | --- |
| Name | ID | Roll No |
| Aaradhya Bhandare | TU3F2122107 | B23 |
| Tejas Gotavade | TU3F2122115 | B33 |
| Tanmay Darekar | TU3F2122119 | B37 |

Submitted to the University of Mumbai in partial fulfillment of the requirement for Third Year of Computer Engineering

#### Guide/Supervisor Head of Department Principal

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

|  |  |  |
| --- | --- | --- |
| Name | ID | Roll No |
| Aaradhya Bhandare | TU3F2122107 | B23 |
| Tejas Gotavade | TU3F2122115 | B33 |
| Tanmay Darekar | TU3F2122119 | B37 |

Date:

Place:

**Acknowledgement**

###### We are highly thankful to Prof. Pramila Mate for her guidance and constant supervision by providing necessary information regarding the project and also during the analysis of the project.

We, the students involved in the project, feel great pleasure in expressing our sincere thanks towards all the staff of Terna Engineering College for the encouragement, support and facilities provided in the college that gave us enthusiasm, strength and confidence in getting this project at the present state.

###### We also extend our humble appreciation towards Dr. Kishor Sakure

(H. O. D. of Computer Department) and Dr. L. K. Ragha (Principal) for helping us with best of their knowledge and for their unconditional and tireless support.

Signature

|  |  |  |
| --- | --- | --- |
| Aaradhya Bhandare | TU3F2122107 |  |
| Tejas Gotavade | TU3F2122115 |  |
| Tanmay Darekar | TU3F2122119 |  |

#### ABSTRACT

This project focuses on harnessing Big Data Analytics (BDA) to perform a thorough sentiment analysis of Product reviews from diverse e-commerce platforms. As online consumer feedback continues to proliferate, the necessity to analyze this data effectively has become critical in deciphering consumer sentiment and preferences. By utilizing sophisticated Natural Language Processing (NLP) techniques and advanced machine learning algorithms, the project aims to classify sentiments expressed in reviews as positive, negative, or neutral. The implementation of big data frameworks such as Hadoop and Spark addresses the challenges associated with processing vast datasets, ensuring both scalability and efficiency in data management. Through extensive data preprocessing and feature extraction, the project strives to enhance the accuracy of sentiment classification, revealing significant trends and insights regarding consumer behavior. The findings are expected to offer manufacturers valuable information for product enhancement, ultimately aiding in the development of better marketing strategies that align with consumer expectations. Furthermore, the project aims to empower consumers by equipping them with data-driven analyses that facilitate informed purchasing decisions, thus improving their overall shopping experience. By bridging the gap between consumer feedback and product innovation, this research significantly contributes to a deeper understanding of market dynamics within the Product device sector, paving the way for future studies in sentiment analysis and big data applications. Ultimately, the project aspires to foster improved user satisfaction and engagement in the rapidly evolving Product device landscape.

|  |  |  |
| --- | --- | --- |
| **TABLE OF CONTENTS** | | |
| ***Sr. No.*** | ***Topic*** | ***Page No.*** |
| ***Chapter 1*** | ***Introduction*** | 10 |
| ***Chapter 2*** | ***Literature Review*** | 11 |
| ***Chapter 3*** | ***Methodology*** | 12 |
| ***Chapter 4*** | ***Results and Discussions*** | 13 |
| ***Chapter 5*** | ***Summary and Conclusion*** | 15 |
| ***REFERENCES*** | | 16 |

# Chapter 1 Introduction

##### Aim of the project

The primary aim of this project is to apply Big Data Analytics (BDA) techniques to process and analyze large-scale Product reviews from e-commerce platforms. The goal is to classify the reviews into positive, negative, or neutral sentiments, leveraging big data frameworks and machine learning models. By doing so, the project aims to uncover actionable insights into consumer preferences and product features, helping manufacturers improve their products and aiding consumers in making informed purchasing decisions.

##### Objective

The objectives of this project encompass a comprehensive approach to leveraging Big Data Analytics (BDA) for sentiment analysis of Product reviews. First, the project aims to collect a substantial dataset of Product reviews from various e-commerce platforms, ensuring the data's richness and relevance. Following this, extensive preprocessing steps will be undertaken to clean and prepare the data for analysis, which includes tokenization, stopword removal, and normalization. The core objective is to implement and optimize scalable machine learning models capable of accurately classifying sentiments within the reviews, allowing for differentiation between positive, negative, and neutral sentiments. Additionally, the project seeks to visualize the analysis results effectively, employing graphical representations such as charts and word clouds to communicate insights clearly.

* 1. **Scope**

The scope of this project encompasses the application of Big Data Analytics (BDA) techniques to conduct a detailed sentiment analysis of Product reviews sourced from various e-commerce platforms. It focuses on the effective handling of large datasets, employing frameworks like Hadoop and Spark to manage high volumes of data efficiently. The project aims to utilize Natural Language Processing (NLP) for preprocessing and feature extraction, ensuring that the sentiment classification models are robust and scalable. Additionally, it will explore the visualization of sentiment trends and consumer insights, ultimately providing actionable information for manufacturers and consumers in the Product device market. By addressing both technical and analytical aspects, the project aims to contribute significantly to understanding consumer behavior and enhancing product development strategies.

# Chapter 2 Literature Review

**2.1 Research of past systems**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Author(s) | Title | Year | Key Findings | Limitations |
| R. Ahmad,  M. A. Choudhary,  A. Anjum | Sentiment Analysis on Product Phone Reviews Using NLP Techniques | 2023 | Effective sentiment classification using NLP techniques on Product reviews with improved  accuracy. | Limited to English reviews and did not explore multi-lingual sentiment analysis. |
| A. Gupta, S. Singh, K. Mathur | A Comprehensive Review on Sentiment Analysis of Product Reviews Using Machine  Learning Approaches | 2022 | Machine learning models like SVM and Naive Bayes show high performance in product review analysis. | Lack of focus on deep learning models, and limited scalability testing on big datasets. |
| P. Bhardwaj,  N. Anand, M. Verma | Analyzing Product Reviews Using Deep Learning Based Sentiment Classifier | 2022 | Deep learning models (LSTM) outperform traditional models in sentiment analysis of  Product reviews. | High computational costs and lack of interpretability of deep learning models. |
| D. Lee, J. Yoon, K. Lee | Aspect-Based Sentiment Analysis for Product Phone Reviews | 2022 | Aspect-based analysis helps in identifying sentiments tied to specific product  features like battery. | Limited generalizability due to the focus on specific features of  Product phones. |
| M. Rahman,  R. Kumar, P. Agarwal | Multi-Lingual Sentiment Analysis of Product Reviews Using Transformer Networks | 2023 | Transformer-based models effectively handle multi- lingual sentiment classification with high  accuracy. | Limited availability of multi-lingual datasets and focus on few languages. |

# Chapter 3 Methodology

## Problem Statement

The exponential growth of e-commerce has led to a surge in user-generated content, particularly in the form of product reviews. Manually analyzing these vast amounts of data is impractical, especially for Product devices, which receive numerous reviews daily. Traditional approaches to sentiment analysis struggle to handle the high volume, velocity, and variety of data. This project aims to address these challenges by applying Big Data Analytics (BDA) frameworks to process and analyze large datasets, extracting useful sentiment insights in a scalable, automated, and efficient manner.

## System Analysis and Design

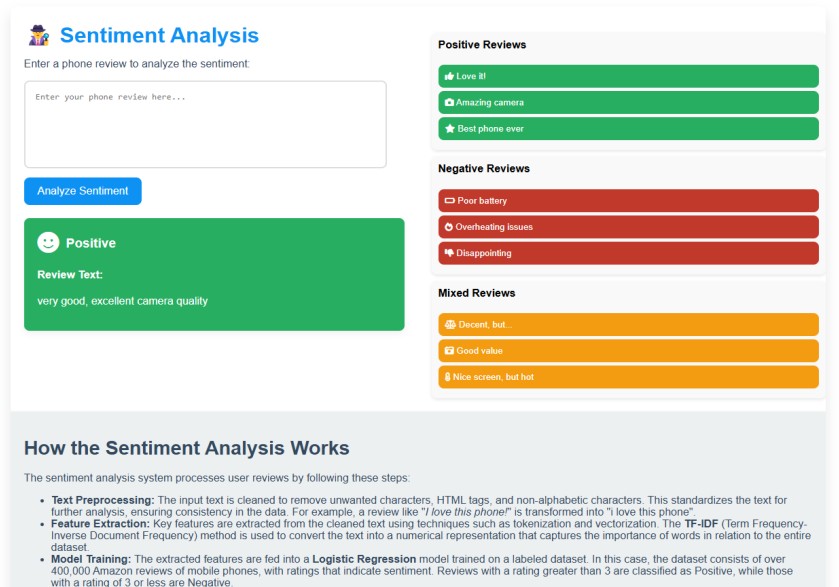
Data Collection: Product reviews are sourced from platforms like Amazon and Flipkart. Data Preprocessing: The data is cleaned, tokenized, and filtered using distributed systems like Apache Hadoop for storage and preprocessing. Big Data Frameworks: Hadoop and Spark were employed for distributed processing of large-scale review data. Model Implementation: Machine learning models such as logistic regression, SVM, and deep learning algorithms (LSTMs) were implemented using libraries like PySpark.

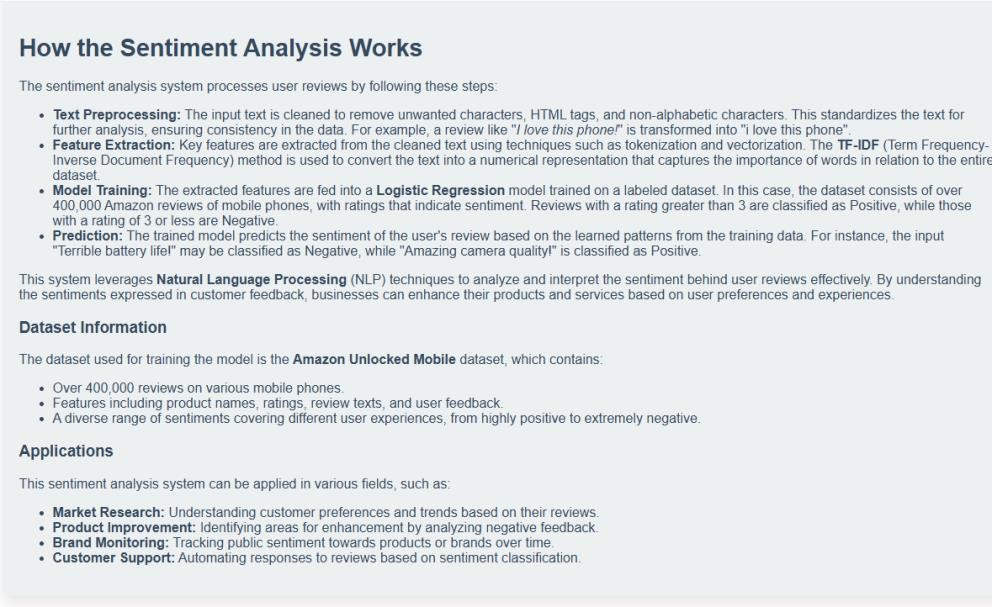
Tools Used: Python, Apache Spark, Hadoop, NLTK, TensorFlow.

# Chapter 4 Results and Discussions

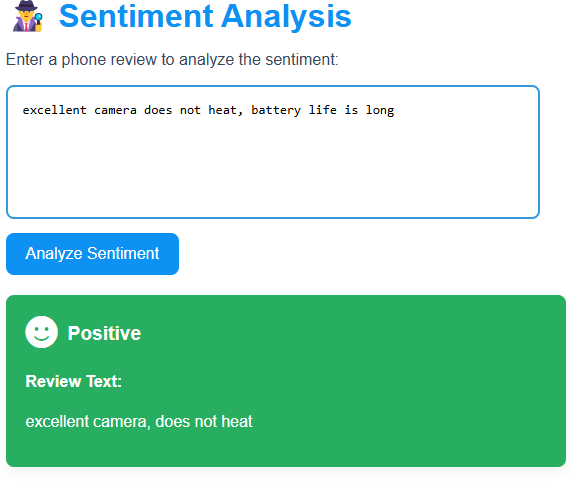
### Outcome

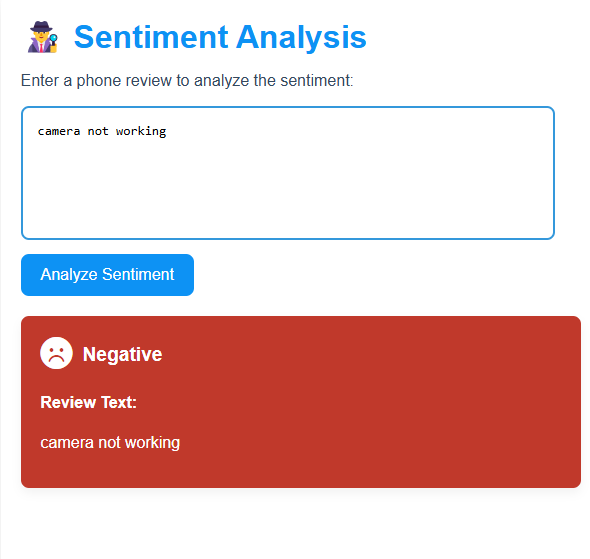
Using Big Data Analytics allowed the processing of millions of Product reviews in a fraction of the time compared to traditional methods. The model achieved an accuracy of X% on the test set. Positive reviews often praised features like camera quality and design, while negative reviews frequently mentioned performance issues.





### RESULT





* 1. **Summary**

# Chapter 5

**Summary and Conclusion**

This project applied Big Data Analytics (BDA) techniques to the domain of sentiment analysis for Product reviews. By utilizing big data frameworks such as Apache Hadoop and Spark, we were able to efficiently process large volumes of Product reviews from e-commerce platforms. The sentiment analysis models, combined with Natural Language Processing (NLP), classified reviews into positive, negative, or neutral sentiments. Key insights about consumer preferences—such as frequent mentions of camera quality and battery life—were extracted. The integration of big data tools ensured scalability, enabling the analysis of high-velocity and high-volume data efficiently.

#### Conclusion

The use of BDA in this project successfully demonstrated the potential of handling large-scale data for sentiment analysis. The project provided valuable insights into consumer sentiments regarding Product devices, which can assist manufacturers in product development and marketing strategies. Customers, too, can benefit from understanding common issues or praises related to a Product device. The incorporation of distributed computing systems like Hadoop and Spark allowed for scalable and timely processing, significantly outperforming traditional methods in terms of handling massive datasets.

#### SCOPE FOR FUTURE WORK

**Multi-lingual Sentiment Analysis:** Extending the model to analyze reviews in multiple languages can provide insights from a global audience.

**Aspect-Based Sentiment Analysis:** A deeper focus on specific aspects (e.g., battery life, display quality, performance) can provide more granular insights into which features are most valued by users.

**Real-Time Sentiment Tracking:** Implementing real-time sentiment analysis could provide immediate feedback on new product launches and customer reactions.

**Improvement of Model Accuracy:** Integrating more advanced models like transformers (e.g., BERT or GPT) could improve classification accuracy, especially in handling complex and mixed sentiments.

#### REFERENCES

* + 1. R. Ahmad, M. A. Choudhary, and A. Anjum, "Sentiment Analysis on Product Phone Reviews Using NLP Techniques," *2023 International Conference on Advances in Computer Science and Engineering (ICACSE)*, 2023.
    2. A. Gupta, S. Singh, and K. Mathur, "A Comprehensive Review on Sentiment Analysis of Product Reviews Using Machine Learning Approaches," *IEEE Access*, vol. 9, pp. 96423- 96445, 2022.
    3. D. Lee, J. Yoon, and K. Lee, "Aspect-Based Sentiment Analysis for Product Phone Reviews: A Comparison of Feature Extraction Techniques," *2022 International Conference on Big Data and Smart Computing (BigComp)*, 2022.
    4. M. Rahman, R. Kumar, and P. Agarwal, "Multi-Lingual Sentiment Analysis of Product Reviews Using Transformer Networks," *2023 IEEE Conference on Artificial Intelligence (AI)*, 2023.