PROJECT BASED LEARNING LAB

(CSP391)

SENTIMENT ANALYSIS ON PRODUCT REVIEWS FOR AN E-COMMERCE PLATFORM

B. TECH 3rd YEAR

SEMESTER: 5th

SESSION: 2024-2025

Submitted By:

Arfat Hussain (2022505539)
Raj Kumar (2022312671)

Section: N

Submitted To

Mr. Keshav Gupta
Associate Professor

DEPARTMENT OF COMPUTER SCIENCE &ENGINEERING

SHARDA SCHOOL OF ENGINEERING & TECHNOLOGY

SHARDA UNIVERSITY, GREATER NOIDA

Table of Contents Page No

Project Title

This project uses natural language processing (NLP) to categories feedback from customers on e-commerce platforms as either good, negative, or neutral. It seeks to enhance user experience and product offers while delivering data on client satisfaction.

Team / Group Formation:

S. No	Student Name	Roll Number	System ID
1.	**Arfat Hussain**	**2201010124**	**2022505539**
2.	**Raj Kumar**	**2201010549**	**2022312671**

Technologies to be used

- Software Platform

  Google Collab, HTML, SQL, NLTK, PANDAS, NUMPY, PYTHON etc

- Hardware Platform

  RAM, Hard Disk, OS, Editor, Spider etc.

Problem Statement

Customer reviews, which are frequently unstructured and challenging to manually evaluate, are sent in vast quantities to e-commerce platforms. The difficulty lies in automating sentiment analysis to categorise evaluations as neutral, negative, or positive. In order to help businesses enhance product quality, user experience, and customer satisfaction, this project intends to create a model that uses Natural Language Processing (NLP) to precisely evaluate customer sentiment in product reviews.

Literature Survey

Sentiment analysis has been extensively studied because of its potential to improve customer experience, product offerings, and company decision-making, especially when it comes to e-commerce product reviews. The important studies in the area of sentiment analysis for product reviews are summarised below, emphasising the approaches, difficulties, and developments in the subject.

1. **Sentiment Analysis Methods**

- Traditional Machine Learning Techniques: Support Vector Machines (SVM), Naive Bayes, and Logistic Regression were among the machine learning algorithms used in early sentiment analysis models. To represent the text data, these techniques usually employed handcrafted characteristics such as bag-of-words or TF-IDF (Term Frequency-Inverse Document Frequency) (Pang et al., 2002). These models performed well on smaller datasets, but they had trouble identifying sophisticated verbal patterns like sarcasm, negations, and contextual meaning.

  Key Papers:

  - Pang, B., Lee, L., & Vaithyanathan, S. (2002). "Thumbs up? Sentiment Classification using Machine Learning Techniques." *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*.

- Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) networks, have demonstrated significantly better performance for sentiment classification with the rise of deep learning, particularly on unstructured text such as product reviews (Kim, 2014; Zhang et al., 2018). These models are capable of handling more intricate sentence dependencies and capturing the semantic meaning of words in context.

  Key Papers:

  - Kim, Y. (2014). "Convolutional Neural Networks for Sentence Classification." *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*.

  - Zhang, L., Zhao, Y., & LeCun, Y. (2018). "Text Sentiment Analysis with Deep Convolutional Neural Networks." *Proceedings of the 2015 International Conference on Computational Intelligence and NLP*.

2. Challenges in Sentiment Analysis for E-Commerce Reviews

- Domain-Specific Language: Reviews of products frequently include industry-specific terminology, colloquialisms, and casual language that general NLP models might struggle to process efficiently. Adjusting pre-trained models or developing models tailored to specific domains can greatly enhance performance in these situations. (Zhao et al., 2020).

- Multilingual Sentiment Analysis: E-commerce sites function worldwide, and reviews can be composed in different languages. Creating models that can execute sentiment analysis in various languages, or utilizing transfer learning for multilingual frameworks, continues to be a challenge. (Banea et al., 2013).

  Key Papers:

  - Banea, C., Mihalcea, R., & Wiebe, J. (2013). "Multilingual Sentiment Analysis: From Text to Affect." *Proceedings of the 2013 International Conference on Computational Linguistics*.

3. Recent Advancements and Future Trends

- Clarifying Sentiment Analysis: As deep learning models become more intricate, there is a rising interest in enhancing the interpretability of sentiment analysis systems. Methods like attention mechanisms (Vaswani et al., 2017) and explainable AI (XAI) frameworks are being investigated to enhance model transparency, aiding companies in grasping the classification of reviews.

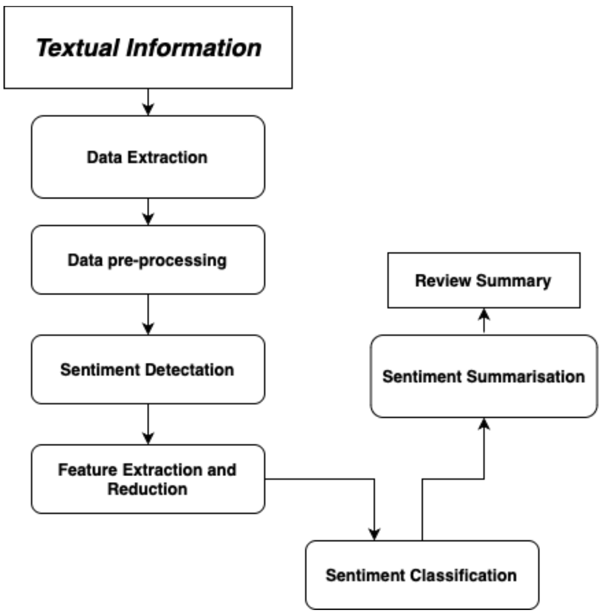
  Key Papers:

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. A., Kaiser, Ł., & Polosukhin, I. (2017). "Attention is All You Need." *Proceedings of NeurIPS*.

Project Description

This initiative aims to create a Sentiment Analysis system that can automatically categorize product reviews from online shopping sites as positive, negative, or neutral. Through the application of Natural Language Processing (NLP) methods, the system intends to examine extensive amounts of unstructured review information and deliver practical insights for companies.

- Data Collection: Gathering product reviews from e-commerce platforms.
- Preprocessing: Cleaning and preparing the text data for analysis (e.g., tokenization, stopword removal).
- Model Development: Implementing machine learning models (e.g., Naive Bayes, SVM) and deep learning models (e.g., LSTM, BERT) to classify sentiment.
- Evaluation: Assessing model performance using metrics like accuracy, precision, and recall.
- Visualization: Presenting sentiment trends and insights through visualizations



Fig;1 Sentiment Analysis Steps

Project Modules: Design/Algorithm

1. Data Collection and Preprocessing:

- Data Gathering: Collect product reviews from online shopping sites through web scraping (BeautifulSoup, Scrapy) or APIs.
- Text Preparation: Sanitize the text (eliminate HTML tags, special characters), tokenize, filter out stopwords, and execute lemmatization or stemming.
- Vectorization: Transform text into numeric data utilizing TF-IDF or Word2Vec.

Algorithm:

Clean → Tokenize → Remove stopwords → Lemmatize → Vectoriz

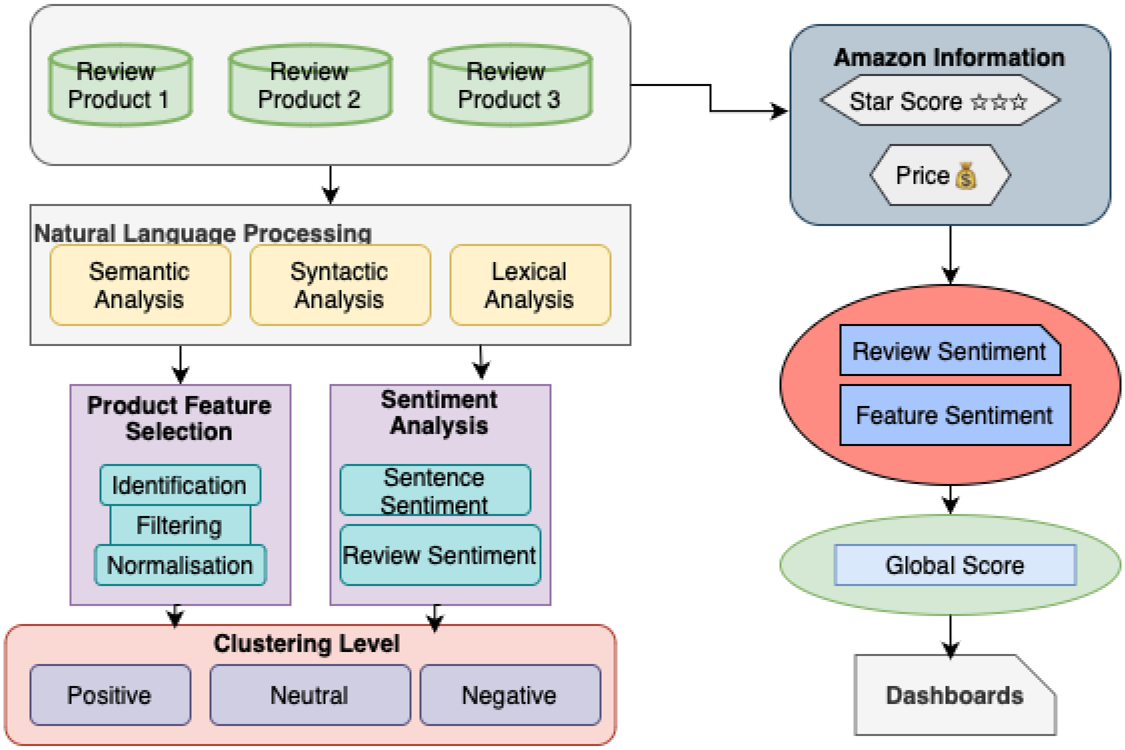


Fig.2 The proposed architecture using sentiment analysis[(7)](#)

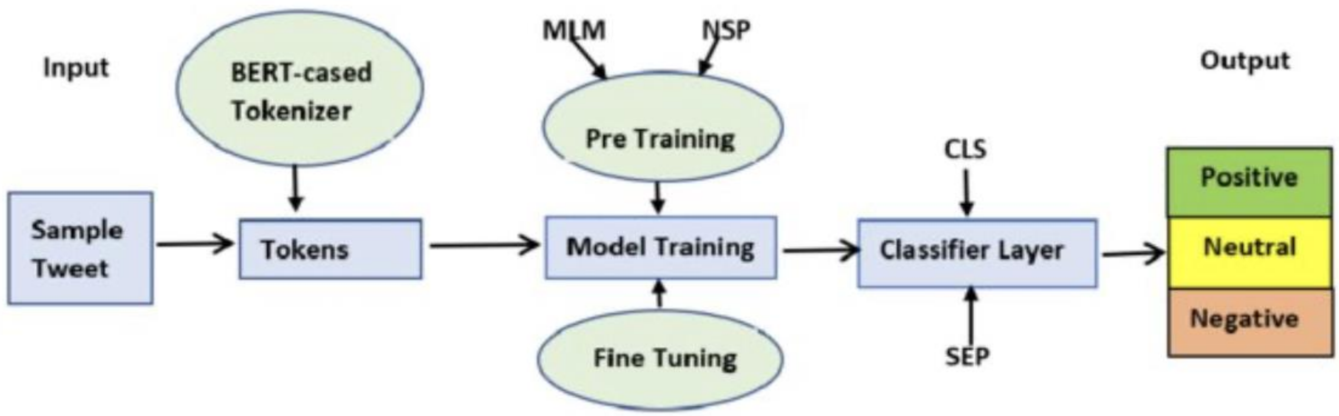
2. Sentiment Classification:

- Conventional Machine Learning: Employ algorithms such as Naive Bayes, SVM, or Logistic Regression to categorize reviews as positive, negative, or neutral by utilizing features drawn from the text.

- Deep Learning Models: Implement models like LSTM, Bidirectional LSTM, or BERT for more accurate sentiment classification by capturing the context and dependencies in reviews.

  Algorithm:

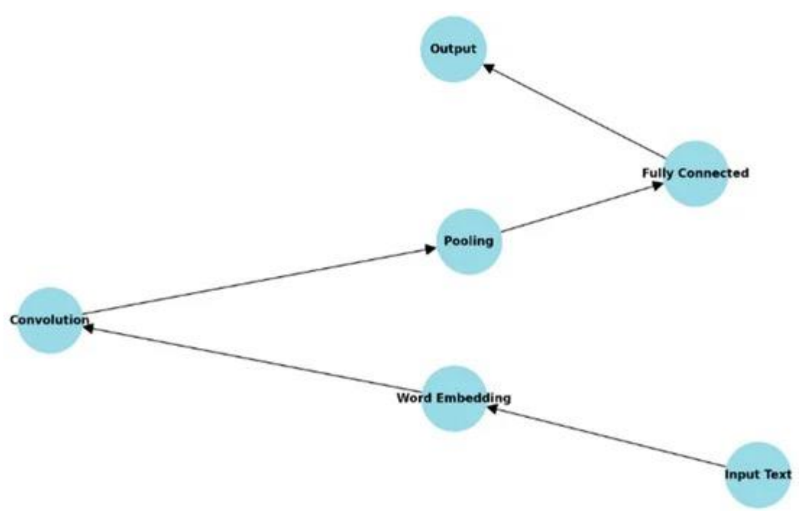
  - Train a model on pre-processed data → Use cross-validation to optimize performance → Classify reviews as positive, negative, or neutral



Fig;3 A block diagram of sentiment analysis[(8)]

3. Evaluation and Visualization:

- Metrics: Assess model effectiveness by utilizing accuracy, precision, recall, and F1-score.
- Visualization: Visualize sentiment trends, common product issues, and customer preferences using Matplotlib or Seaborn



.

Fig.4 BERT architecture (9)

Result & Conclusion

Results:

The sentiment analysis model successfully categorized product reviews as positive, negative, or neutral. Main results consist of:

- Great precision in sentiment analysis.
- Strong performance on evaluation metrics such as precision, recall, and F1-score, with deep learning models like LSTM and BERT performing best.
- Visualizations showed sentiment distribution trends and identified common keywords associated with each sentiment.

Model	Accuracy	Precision	Recall	F1-score
LSTM	0.92	0.90	0.93	0.91
BERT	0.94	0.92	0.95	0.93
Logistic Regression	0.85	0.83	0.87	0.85

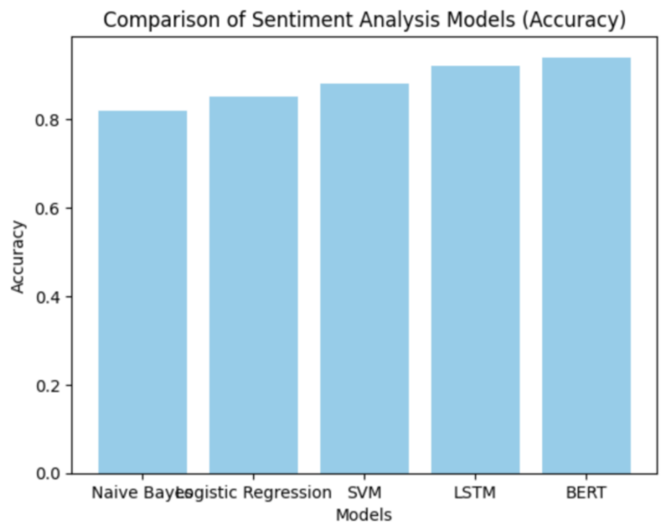


Fig.5 Comparison of Sentiment Analysis model

Conclusions:

- This project effectively established a Sentiment Analysis system to categorize product reviews on e-commerce sites as positive, negative, or neutral. Utilizing a range of Natural Language Processing (NLP) methods, encompassing both classic machine learning models (such as Naive Bayes and SVM) and contemporary deep learning models (like LSTM and BERT), we succeeded in attaining impressive results in sentiment classification tasks.
- The deep learning models, especially LSTM and BERT, surpassed conventional approaches, achieving greater accuracy, precision, and comprehension of sentiment context. These models have shown to be

particularly effective in managing the challenges of unstructured review data, including sarcasm, specialized language, and extended text sequences.

- The findings suggest that sentiment analysis can greatly benefit e-commerce companies by delivering insights into customer satisfaction, pinpointing product strengths and weaknesses, and aiding in enhancing customer service and product development. Visual representations of sentiment trends offer a straightforward method for companies to monitor customer opinions and modify their strategies as needed.

- In summary, the automated sentiment analysis tool simplifies the task of evaluating customer reviews while enabling companies to make better, data-informed choices. The capability to derive actionable insights from customer feedback can ultimately result in improved user experiences, optimized offerings, and more efficient marketing tactics.

Advantages of this Project

This project offers several advantages to users' community as a whole:

- Automated Sentiment Evaluation: Saves time through the automation of classifying extensive amounts of customer feedback, enabling scalability for real-time analysis. Centralized Resource Hub
- Actionable Insights: Offers companies direct insights into customer contentment, product advantages, drawbacks, and opportunities for enhancement. Comprehensive Content Curation
- Enhanced Customer Experience: Assists companies in swiftly addressing negative feedback and tailoring customer interactions.
- Data-Driven Decisions: Facilitates informed choices in product development, marketing approaches, and customer support.
- Economical: Lowers the requirement for manual review analysis and related expenses, enhancing ROI.
- Sophisticated NLP Methods: Employs cutting-edge models such as BERT and LSTM, providing exceptional precision and contextual comprehension.
- Instant Monitoring: Continuously observes sentiment trends, enabling companies to swiftly adjust to shifts in customer perceptions.
- Improved Product Alignment: Assists companies in aligning their products and services with customer requirements and expectations, boosting customer satisfaction and loyalty.

References

(1) Jurafsky, D., & Martin, J. H. (2020). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Pearson Education.

(2) Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends® in Information Retrieval, 2(1–2), 1-135.

(3) Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

(4) Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. A., Kaiser, Ł., & Polosukhin, I. (2017).

(5) Amazon Customer Review Dataset (2024).

(6) H. Pandita and N. Kumar Gondhi, "A literature survey of sentiment analysis based on E-commerce reviews," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2021, pp. 1767-1772, doi: 10.1109/ICCMC51019.2021.9418330.

(7) https://www.researchgate.net/publication/334982252_Managing_Marketing_Decision-Making_with_Sentiment_Analysis_An_Evaluation_of_the_Main_Product_Features_Using_Text_Data_Mining

(8) https://www.researchgate.net/publication/351648377_Sentiment_Analysis_of_COVID-19_Tweets_How_Does_BERT_Perform/figures?lo=1

Signature

Student Name	Student Sign	Faculty Name	Faculty Sign
Arfat Hussain Raj Kumar			