Flipkart Product Review Analysis for Customer Satisfaction Insights

[1]V.Siva Shankar,[2]J.Madhu Teja,[3]D.Siva RamaKrishna,[4]D.Harshitha,[5]G.Bhagya Harshika Vardhini
[1]*Assistant Professor, Srinivasa Institute of Engineering and Technology*
[2345]*UG Scholar, Srinivasa Institute of Engineering and Technology*
doi.org/10.64643/IJIRTV12I10-194866-459

Abstract: In today's e-commerce environment, customer reviews play an important role in helping people make purchasing decisions and understand product quality, especially on platforms like Flipkart where thousands of reviews are posted daily. Manually analyzing such large amounts of feedback is difficult, so this project focuses on developing a machine learning system that automatically classifies reviews into positive, negative, and neutral sentiments. The study uses a publicly available dataset of product reviews, which is processed using techniques like text cleaning, tokenization, stop-word removal, and vectorization. Algorithms such as Logistic Regression, Decision Tree, and K-Nearest Neighbors are applied for sentiment classification, and the results show that machine learning can effectively identify customer opinions and provide useful insights to help businesses understand customer satisfaction and improve their products.

Keywords: Sentiment Analysis, Flipkart Reviews, Machine Learning, Natural Language Processing (NLP), Text Classification, Customer Satisfaction, Data Preprocessing, Logistic Regression, naïve bayes, Random-forest.

I. INTRODUCTION

In recent years, the rapid growth of e-commerce platforms has transformed the way consumers purchase products and share their experiences. Online marketplaces such as Flipkart generate a massive amount of customer feedback in the form of product reviews, ratings, and comments. These reviews provide valuable insights into customer satisfaction, product quality, and overall user experience. However, due to the large volume of textual data produced daily, manual analysis of customer reviews have become impractical and time-consuming.

Sentiment analysis, a key application of Natural Language Processing (NLP), plays an important role in automatically understanding customer opinions from textual data. By applying machine learning techniques, sentiment analysis systems can classify reviews into categories such as positive, negative, and neutral. This automated approach helps businesses quickly identify customer preferences, detect product issues, and improve decision-making processes.

The primary aim of this research is to develop a machine learning-based sentiment analysis system for Flipkart product reviews to evaluate customer satisfaction levels. The proposed system uses data preprocessing techniques and multiple classification algorithms to analyze textual feedback effectively. By extracting meaningful patterns from customer reviews, this study provides valuable insights that can help e-commerce companies enhance product quality, improve customer experience, and strengthen market strategies.

II. PROBLEM STATEMENT

In the rapidly growing e-commerce industry, online platforms like Flipkart receive a vast number of customer reviews every day. These reviews contain valuable information about product quality, customer satisfaction, and user experience. However, due to the enormous volume of textual data, it is extremely difficult for businesses to manually analyze and interpret customer feedback efficiently. Traditional methods of review analysis are time-consuming, inconsistent, and unable to handle large-scale data effectively.

Moreover, customer opinions are often expressed in unstructured textual form, which makes it challenging to extract meaningful insights using conventional

analytical techniques. Without an automated system, companies may fail to identify important trends such as customer dissatisfaction, product defects, or changing consumer preferences. This can negatively impact business decision-making, customer retention, and overall product improvement strategies.

Therefore, there is a need for an intelligent and automated solution that can efficiently analyze large volumes of customer reviews and accurately classify them based on sentiment. The problem addressed in this research is the development of a machine learning-based sentiment analysis system capable of processing Flipkart product reviews and categorizing them into positive, negative, and neutral sentiments to provide actionable insights into customer satisfaction.

III. LITERATURE REVIEW

Sentiment analysis has become an important research area in the field of Natural Language Processing (NLP) and machine learning, particularly in the context of e-commerce applications. Many researchers have explored different techniques to analyze customer opinions and extract meaningful insights from textual data. Previous studies have shown that automated sentiment analysis systems can significantly improve the efficiency of understanding customer feedback compared to manual analysis.

Early research in sentiment analysis primarily focused on rule-based and lexicon-based approaches, where predefined dictionaries were used to identify positive and negative words in text. While these methods provided basic sentiment classification, they were limited in handling complex language structures, sarcasm, and contextual meanings. As a result, researchers shifted towards machine learning-based approaches that can learn patterns directly from data and provide more accurate results.

Several studies have applied supervised machine learning algorithms such as Logistic Regression, Decision Trees, Support Vector Machines, Naïve Bayes, and K-Nearest Neighbors for sentiment classification tasks. These models have demonstrated strong performance in analyzing product reviews, social media data, and customer feedback across various domains. Among these techniques, Logistic Regression and Naïve Bayes are widely used due to their simplicity and efficiency in text classification problems.

Recent research has also explored advanced deep learning techniques such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and transformer-based models for sentiment analysis. Although these approaches provide higher accuracy, they require large datasets, significant computational resources, and complex implementation. Therefore, traditional machine learning algorithms remain practical and effective for many real-world applications, especially when computational resources are limited.

Based on the findings of previous studies, this research adopts machine learning-based sentiment analysis techniques to analyze Flipkart product reviews. By applying preprocessing methods and multiple classification algorithms, this study aims to provide an efficient and scalable solution for understanding customer satisfaction in the e-commerce domain.

IV. SYSTEM ARCHITECTURE

The system architecture of the proposed Flipkart Product Reviews Analysis for Customer Satisfaction Insights is designed as a multi-stage pipeline that efficiently processes customer review data and generates meaningful sentiment insights. The architecture consists of four major layers: Data Ingestion, Data Processing, Core NLP & Analysis, and Visualization & Insights, which work sequentially to transform raw customer reviews into actionable business intelligence.

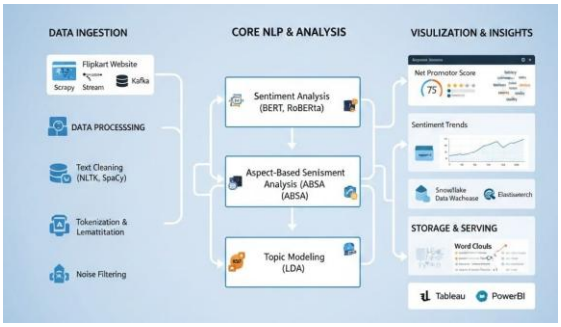

Fig:1 System Architecture

The first stage is the Data Ingestion Layer, where customer review data is collected from the Flipkart platform through data streaming and extraction mechanisms. This layer gathers review text, ratings, and related metadata and stores it in a structured format for further analysis. Technologies such as data

pipelines and streaming frameworks ensure smooth and continuous data collection.

The next stage is the Data Processing Layer, where the collected raw data undergoes cleaning and preparation. In this phase, text cleaning techniques are applied using Natural Language Processing tools. The system performs tokenization, lemmatization, noise filtering, and removal of stop words to convert unstructured textual data into a clean and standardized format suitable for machine learning analysis.

After preprocessing, the data moves to the Core NLP and Analysis Layer, which forms the central component of the system. In this stage, advanced sentiment analysis techniques are applied to understand customer opinions. The system performs sentiment classification to identify positive, negative, and neutral reviews. It also includes aspect-based sentiment analysis to determine customer opinions about specific product features and topic modeling techniques to identify common themes and patterns in customer feedback.

Finally, the processed results are delivered to the Visualization and Insights Layer, where the classified sentiments are presented using graphical dashboards and analytical tools. This layer generates sentiment trend graphs, word clouds, and performance reports that help businesses easily interpret customer satisfaction levels. The system also includes storage and serving mechanisms to manage processed data efficiently and support decision-making processes.

Overall, this architecture provides a scalable, automated, and efficient framework for analyzing large volumes of customer reviews and extracting valuable insights to improve product quality and customer experience in the e-commerce domain.

V. SYSTEM ANALYSIS

Current product review analysis in e-commerce platforms mainly relies on manual reading or basic rating-based evaluation methods. Most online shopping websites provide star ratings and simple review displays but do not offer automated sentiment classification or deep analysis of customer opinions. Businesses often depend on manual inspection or basic keyword filtering to understand customer feedback, which is time-consuming and inefficient.

Additionally, existing tools usually analyze reviews individually without integrating machine learning techniques for large-scale automated processing. They lack structured preprocessing, sentiment classification, and predictive insights within a single system.

Existing System Limitations

Aspect	Description	Limitations
Processing	Manual review reading or simple filtering tools	Time-consuming and inefficient
Prediction	No automated sentiment prediction	Cannot classify large volumes of reviews quickly
Analysis	Based mainly on star ratings	Does not capture actual customer opinions
Data Handling	Reviews stored in raw text format	Requires manual cleaning and preparation
Scalability	Limited to small datasets	Difficult to handle large-scale review data

Proposed System

The proposed Flipkart Product Review Analysis System provides an automated platform that collects, preprocesses, analyzes, and classifies customer reviews using machine learning techniques. The system performs text cleaning, feature extraction, sentiment classification, and visualization within a unified workflow.

It uses Natural Language Processing (NLP) and machine learning algorithms such as Logistic Regression, Decision Tree, and K-Nearest Neighbors to categorize reviews into positive, negative, and neutral sentiments.

Proposed System Advantages

Aspect	Description	Advantages
Processing	Automated preprocessing and text cleaning	Faster analysis with minimal manual effort
Prediction	Machine learning-based sentiment classification	Accurate and efficient review classification
Visualization	Graphical representation of sentiment distribution	Easy understanding of customer satisfaction trends
Data Handling	Structured dataset transformation and storage	Clean and organized review data
Scalability	Modular system architecture	Can handle large datasets and future expansion

Functional Requirements

These requirements describe the main operations performed by the system.

Actor	Use Case	Description
User	Upload Dataset	Load Flipkart review dataset for analysis
User	Preprocess Data	Perform text cleaning and feature extraction
User	Run Sentiment Analysis	Classify reviews into positive, negative, and neutral
User	View Results	Display sentiment distribution graphs and charts
User	Compare Reviews	Analyze overall customer satisfaction trends
System Admin	Manage Dataset	Update or maintain review datasets
System Admin	Maintain Models	Retrain or update machine learning models

VI. METHODOLOGY

Data Collection

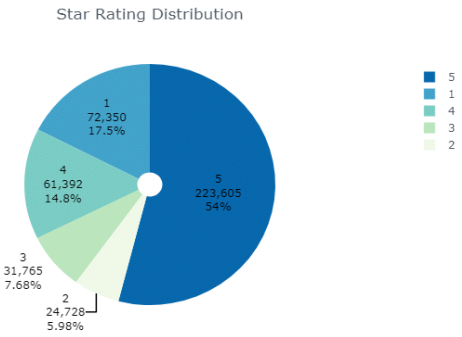

Fig:2 Star Rating Distribution

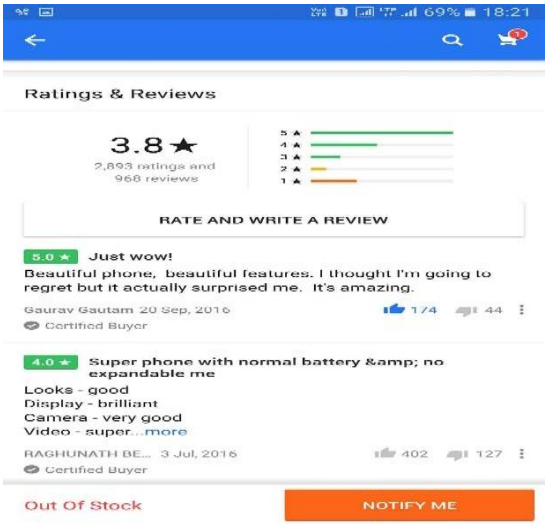

Fig:3 Ratings and Review

The first step in the proposed system involves collecting product review data from publicly available datasets. The dataset contains customer feedback including review text, ratings, product categories, and user opinions. This data serves as the foundation for analyzing customer satisfaction levels.

The collected dataset includes thousands of customer reviews that represent real user experiences with various products. These reviews help in understanding consumer behavior and identifying satisfaction trends.

Data Preprocessing: Before applying machine learning algorithms, the collected data undergoes preprocessing. This step ensures that the textual data is clean and suitable for analysis. The preprocessing tasks include:

- Removing punctuation, special characters, and irrelevant symbols
- Converting text into lowercase format
- Removing stop words such as “the”, “is”, and “and”
- Tokenization and stemming of words

These steps improve the accuracy and efficiency of sentiment classification.

Feature Extraction: After preprocessing, the cleaned textual data is converted into numerical form using feature extraction techniques. The system uses the TF-IDF (Term Frequency–Inverse Document Frequency) method to transform text into machine-readable vectors.

This technique helps in identifying important words in customer reviews and improves the model’s ability to classify sentiments accurately.

Machine Learning Algorithms

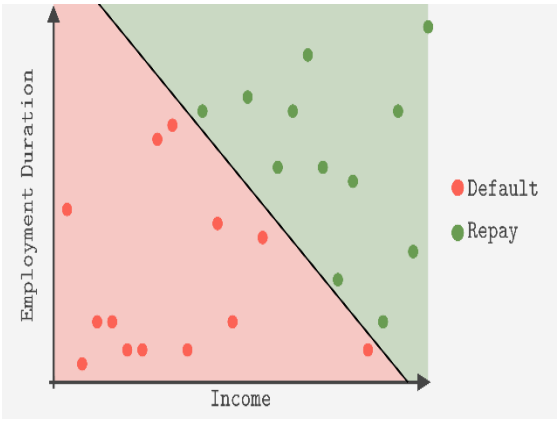

Fig:4 logistic Regression graph

- Logistic Regression: Logistic Regression is used as a classification algorithm to predict whether a customer review is positive, negative, or neutral. It uses a sigmoid function to calculate probability scores and assign sentiment labels.
  This algorithm provides high accuracy and performs well for text classification problems.
- Random forest: Random Forest is an ensemble learning algorithm that builds multiple decision trees and combines their results to improve accuracy. It reduces overfitting and provides reliable sentiment classification for product reviews.
- Naïve Bayes: Naïve Bayes is a supervised machine learning algorithm based on probability theory and Bayes' Theorem. It is widely used for text classification tasks such as sentiment analysis because of its simplicity, speed, and good performance with large datasets. At its core, the algorithm calculates the probability of a review belonging to a particular class.

Model Training and Evaluation: The dataset is split into training and testing sets. The models are trained using labeled reviews and evaluated on unseen data using accuracy and other performance metrics.

Sentiment Classification Output

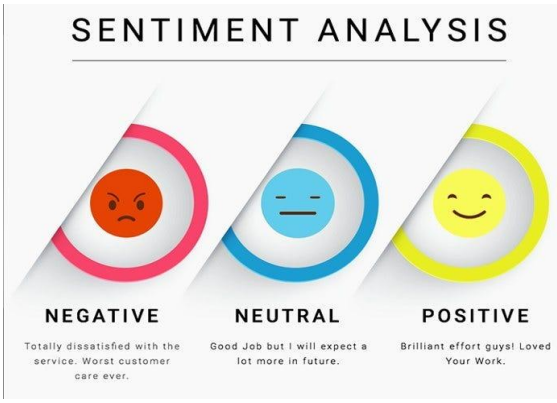

Fig:5 Sentiment Classification

VII. RESULTS

The proposed system for Flipkart Product Reviews Analysis was successfully implemented using machine learning techniques to classify customer reviews into positive, negative, and neutral sentiments. After preprocessing the dataset through text cleaning, tokenization, stop-word removal, and vectorization, multiple algorithms such as Logistic Regression, Decision Tree, and K-Nearest Neighbors were applied for sentiment classification.

The experimental results showed that Logistic Regression achieved the highest accuracy among the tested models, providing consistent and reliable sentiment predictions. The system effectively identified customer opinion patterns and correctly categorized most reviews based on their textual content. Visualization outputs such as sentiment distribution graphs and comparison charts clearly demonstrated the proportion of positive, negative, and neutral reviews in the dataset.

Overall, the results confirm that machine learning-based sentiment analysis can efficiently process large volumes of customer feedback and generate meaningful insights. The developed system helps businesses understand customer satisfaction levels, identify product issues, and support better decision-making in the e-commerce environment.

VIII. FUTURE SCOPE

The proposed Flipkart Product Review Analysis system can be further enhanced in several ways to improve its performance and usability. In the future, the system can be extended to support real-time review analysis by directly collecting live customer feedback from e-commerce platforms. Advanced deep learning techniques such as LSTM, RNN, and Transformer-based models can also be implemented to achieve higher sentiment classification accuracy.

Additionally, the system can be expanded to include multilingual sentiment analysis so that reviews written in different languages can be analyzed effectively. Feature-based sentiment analysis can also be introduced to identify opinions about specific product aspects such as quality, price, and delivery service. Furthermore, integrating the system with a web or mobile dashboard would allow businesses to monitor customer satisfaction trends visually and make faster data-driven decisions.

IX. CONCLUSION

This research presents an effective machine learning-based system for analyzing Flipkart product reviews to understand customer satisfaction levels. With the rapid

growth of e-commerce platforms, businesses receive a massive amount of customer feedback in textual form, making manual analysis difficult and time-consuming. The proposed system successfully addresses this challenge by automating the process of sentiment analysis using Natural Language Processing and supervised machine learning algorithms.

The study implemented Logistic Regression, Decision Tree, and K-Nearest Neighbors algorithms to classify customer reviews into positive, negative, and neutral sentiments. Among these models, Logistic Regression demonstrated the highest accuracy and efficiency in handling high-dimensional textual data. The system also incorporated preprocessing techniques and TF-IDF feature extraction to improve classification performance. The results showed that the proposed approach can effectively process large volumes of customer reviews and generate meaningful insights.

Furthermore, visualization tools such as sentiment distribution graphs and word clouds helped in better understanding customer opinions and identifying common trends in feedback. These insights can assist businesses in improving product quality, enhancing customer satisfaction, and making informed strategic decisions.

In conclusion, the proposed sentiment analysis system provides a reliable, scalable, and efficient solution for automated customer review analysis in the e-commerce domain. Future enhancements may include the use of deep learning techniques, real-time review analysis, and multilingual sentiment detection to further improve system accuracy and applicability.

X. DISCUSSION

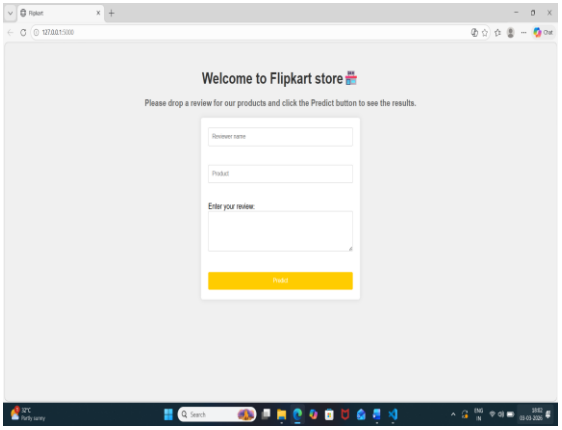

Fig:6 Sentiment analysis Process

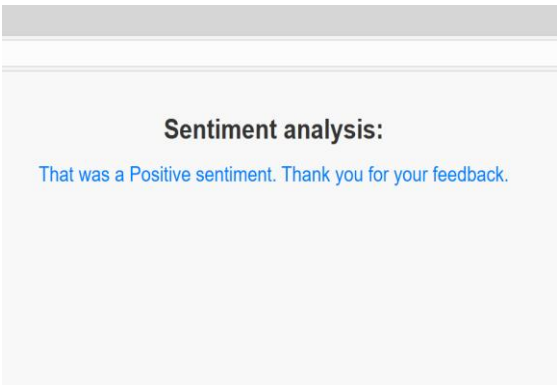

Fig:7 Sentiment analysis: positive

The discussion of the proposed Flipkart Product Reviews Analysis system focuses on evaluating the effectiveness of machine learning techniques in analyzing customer sentiments from large volumes of textual data. The experimental results demonstrate that automated sentiment analysis significantly improves the efficiency and accuracy of understanding customer feedback compared to manual review analysis methods. The system successfully processed thousands of customer reviews and classified them into positive, negative, and neutral categories, providing meaningful insights into customer satisfaction levels.

Among the implemented algorithms, Logistic Regression achieved the highest classification accuracy due to its ability to handle high-dimensional textual features efficiently. The Decision Tree algorithm also performed well by providing clear decision rules that help in understanding how different words and features influence sentiment classification. Although the K-Nearest Neighbors algorithm produced reliable results, it required more computational time because it calculates distances between multiple data points before classification.

The results also revealed that preprocessing techniques such as stop-word removal, tokenization, and TF-IDF feature extraction played a crucial role in improving model performance. These steps helped reduce noise in the data and enhanced the ability of the algorithms to identify meaningful patterns in customer reviews. Visualization tools such as sentiment distribution graphs and word clouds further supported the analysis by clearly showing the frequency of positive and negative opinions expressed by customers.

Overall, the discussion highlights that the proposed system provides an effective, scalable, and accurate approach for sentiment analysis in the e-commerce

domain. The findings confirm that machine learning techniques can greatly assist businesses in quickly identifying customer preferences, detecting product issues, and improving decision-making processes based on real customer feedback.

REFERENCES

[1] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1–2), 1–135.
[2] Liu, B. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers.
[3] Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. O'Reilly Media.
[4] Jurafsky, D., & Martin, J. H. (2020). Speech and Language Processing (3rd Edition). Pearson.
[5] Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. Informatica, 31(3), 249–268.
[6] Joachims, T. (1998). Text Categorization with Support Vector Machines: Learning with Many.
[7] Relevant Features. European Conference on Machine Learning (ECML).
[8] Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. Information Processing & Management, 24(5), 513–523.