SIR M.VISVESVARAYA INSTITUTE OF TECHNOLOGY

International Airport road, Husnasamaranahalli,
Yelahanka, Sir M Visvesvaraya Inst Road, Yelahanka,
Bengaluru, Karnataka 562157
Phone :080 2846 7248

DATA ANALYCTICS MINI-PROJECT

TOPIC: Product Review Analysis

Name : Amuru Motheesh, Supriya Khorvi
USN : 1MV23MC005, 1MV23MC047
Semester : 3rd
Department : MCA
Subject :Data Analytics Lab with Mini Project
Subject code :22MCAL36
Faculty in charge:Dr.Ch.Vanipriya
Head of the Department(HoD)
Dept. of MCA
Date:
Signature:

TITLE

PRODUCT REVIEW SENTIMENT ANALYSIS AND VISUALIZATION USING NLTK & WORDCLOUD

ABSTRACT

This project aims to analyze customer sentiment in Amazon product reviews using Natural Language Processing (NLP) techniques. The dataset, sourced from Kaggle's Amazon Review Dataset, undergoes text preprocessing steps such as stopword removal, tokenization, and lemmatization. Sentiment analysis is performed using NLTK’s VADER and TextBlob, classifying reviews into positive, negative, and neutral sentiments. To better understand customer feedback trends, visualizations like word clouds and sentiment distribution graphs are generated.

The results provide insights into common themes in customer reviews, enabling businesses to identify strengths and areas for improvement in their products. This project demonstrates a scalable approach to automated sentiment analysis, which can be further enhanced using machine learning models or deep learning techniques like BERT. Future improvements include deploying the analysis as a web application for real-time sentiment tracking.

INTRODUCTION

Online product reviews serve as a valuable source of information for both consumers and businesses. Customers rely on these reviews to make informed purchasing decisions, while businesses use them to gauge product performance, identify areas for improvement, and enhance customer satisfaction. However, manually analyzing thousands of reviews is time-consuming and inefficient.

This project leverages Natural Language Processing (NLP) techniques to automate sentiment analysis on Amazon product reviews. By classifying reviews into positive, negative, and neutral categories, businesses can gain actionable insights into customer opinions. Additionally, visualization techniques such as word clouds and sentiment distribution graphs help in identifying frequently used words associated with different sentiments, offering a deeper understanding of customer feedback trends.

OBJECTIVES

- **Extract and preprocess** product reviews from the Amazon dataset.
- **Analyze sentiments** (Positive, Negative, Neutral) using **NLTK VADER** and **TextBlob**.
- **Visualize** the most common words in different sentiment categories using **WordCloud**.
- **Generate insights** about customer feedback trends.

PROBLEM STATEMENT

Understanding customer sentiment from textual reviews is challenging due to:

1. **Unstructured data** – Reviews contain informal language, typos, and abbreviations.
2. **Subjectivity** – The same word may have different meanings in different contexts.
3. **Scalability** – Analyzing thousands of reviews manually is impractical.

By automating sentiment classification and text visualization, this project provides a scalable solution for deriving insights from large datasets.

METHODOLOGY

The project follows these key steps:

Step 1: Data Collection

- The dataset (train.csv) is sourced from Kaggle's Amazon Product Reviews.
- The dataset contains product reviews and their sentiment labels.

Step 2: Data Preprocessing

- Convert text to lowercase.
- Remove punctuation, numbers, and special characters.
- Remove stopwords (common words like *the, is, an* that do not add meaning).
- Apply tokenization and lemmatization to standardize words.

Step 3: Sentiment Analysis

- **Using NLTK VADER**: Classifies reviews as Positive, Negative, or Neutral based on polarity scores.
- **Using TextBlob**: Analyzes the overall sentiment polarity of the text.

Step 4: Data Visualization

- **Sentiment Distribution Graph**: Displays the proportion of positive, negative, and neutral reviews.
- **Word Cloud**: Highlights frequently used words in each sentiment category.
- **Bar Charts**: Show insights into the most common sentiment words.

Step 5: Result Interpretation

- Identify common words in positive and negative reviews.
- Understand customer pain points and preferences.
- Generate actionable insights into product improvements.

PROJECT WORKFLOW

1. **Preprocess reviews** → Clean and standardize text.
2. **Perform sentiment analysis** → Assign sentiment labels.
3. **Generate visualizations** → Word clouds and bar charts.
4. **Interpret insights** → Identify key trends in customer feedback.

TECHNOLOGIES USED

- **Programming Language**: Python
- **Libraries**:
    - NLTK → Text cleaning, sentiment analysis (VADER)
    - TextBlob → Additional sentiment analysis
    - WordCloud → Word frequency visualization
    - Pandas → Data Processing
    - Matplotlib & Seaborn → Graphs and plots

EXPECTED OUTCOMES

- A cleaned dataset with labeled sentiments.
- Accurate sentiment classification of product reviews.
- Interactive visualizations showing customer sentiment trends.
- Insights into customer feedback to help businesses improve products.

CONCLUSION

This project provides a systematic approach to sentiment analysis, helping businesses extract valuable insights from product reviews. By leveraging NLTK, TextBlob, and WordCloud, it delivers efficient sentiment classification and intuitive visualizations, enabling data-driven decision-making.