FLIPKART SENTIMENTAL ANALYSIS

A PROJECT REPORT

Submitted by

AMIT KUMAR SAMAL

(230720100157)

in partial fulfillment for the award of

the degree of

MASTER OF COMPUTER APPLICATION

SCHOOL OF ENGINEERING AND TECHNOLOGY

BHUBANESWAR CAMPUS

CENTURION UNIVERSITY OF TECHNOLOGY AND MANAGEMENT

ODISHA

NOVEMBER 2024

BONAFIDE CERTIFICATE

Certified that this project report "**FLIPKART SENTIMENTAL ANALYSIS"** is the bonafide work of **"AMIT KUMAR SAMAL"** who carried out the project work under my supervision. This is to further certify to the best of my knowledge, that this project has not been carried out earlier in this institute and the university.

SIGNATURE

(Prof. Saneev Kumar Das)
(Department of MCA)

Certified that the above-mentioned project has been duly carried out as per the norms of the college and statutes of the university.

SIGNATURE

(Mr. Rakesh Kumar Ray)

(HOD of MASTER OF COMPUTER APPLICATION)

DEPARTMENT SEAL

<u>DECLARATION</u>

    I hereby declare that the project entitled "**FLIPKART SENTIMENTAL ANALYSIS**" submitted for the "Minor Project" of 3rd semester MCA in Computer Application is my original work and the project has not formed the basis for the award of any Degree / Diploma or any other similar titles in any other University / Institute.

Name of the Student: Amit Kumar Samal

Signature of the Student:

Registration No:230720100157

Place: Bhubaneswar

Date:

ACKNOWLEDGEMENTS

We wish to express my profound and sincere gratitude to Prof. Saneev Kumar Das, Professor of Computer Science and Engineering, SoET, Bhubaneswar Campus, who guided me into the intricacies of this project nonchalantly with matchless magnanimity.

I thank Mr. Rakesh Kumar Ray, Head of the Dept. of Department of Master of Computer Application, SoET, Bhubaneswar Campus and Dr. Sujata Chakravarty Dean, School of Engineering and Technology, Bhubaneswar Campus for extending their support during Course of this investigation.

I would be failing in my duty if I don't acknowledge the cooperation rendered during various stages of image interpretation by Prof. Saneev Kumar Das

I am highly grateful to Prof. Saneev Kumar Das who evinced keen interest and invaluable support in the progress and successful completion of my project work.

I am indebted to Saneev Kumar Das for their constant encouragement, co-operation and help. Words of gratitude are not enough to describe the accommodation and fortitude which they have shown throughout my endeavor.

Name of the Student:Amit Kumar Samal

Signature of the Student:

Registration No: 230720100157

Place: Bhubaneswar

Date:

TABLE OF CONTENTS

Chapter 1:

Sentiment analysis is a transformative tool in understanding customer feedback, helping e-commerce platforms like Flipkart improve customer satisfaction, optimize product offerings, and enhance user experience. By leveraging Machine Learning (ML) and statistical methods, this project aims to develop a predictive model that classifies customer sentiments accurately as positive, negative, or neutral.

The Flipkart Sentiment Analysis system focuses on analyzing customer reviews to uncover insights about products and services, with applications in customer feedback evaluation, trend detection, and targeted marketing strategies.

Our primary objective is to utilize ML techniques to model sentiment classification by analyzing factors such as review content, linguistic features, and user interactions. The project employs text classification models, ensemble learning, and advanced preprocessing techniques to extract meaningful patterns from customer reviews, resulting in a reliable and actionable sentiment analysis system.

By emphasizing data cleaning, feature engineering, and model optimization, this project addresses key challenges in natural language processing, such as handling unstructured text data and accounting for multilingual content.

The project also highlights the ethical aspects of sentiment analysis, such as ensuring fairness in interpretations, avoiding bias in feedback categorization, and safeguarding user data privacy. Transparency in sentiment classification models is crucial for maintaining trust in an environment where customer feedback shapes business decisions. Addressing these ethical challenges is essential to building systems that respect users' voices and provide equitable insights.

Beyond its practical applications for Flipkart, this project serves as a platform to explore essential ML concepts like supervised learning, text classification, natural language processing (NLP), and evaluation metrics. Additionally, it contributes to better decision-making for Flipkart's business teams and enhanced shopping experiences for its users.

Sentiment analysis of Flipkart reviews demonstrates the power of ML and data-driven techniques in understanding customer behavior and promoting better decision-making. This underscores the importance of addressing both technical and ethical dimensions in developing systems that impact customer engagement and business outcomes.

Methodology:

The methodology for the Flipkart Sentiment Analysis project involves a systematic approach, from data preprocessing to model evaluation, ensuring accurate and actionable sentiment classification:

1. Data Collection and Preprocessing: Data Collection: Gather customer reviews from Flipkart, either through web scraping, publicly available datasets, or APIs, ensuring a diverse and representative sample.Preprocessing:Clean the data by removing duplicates, special characters, and stopwords.Handle missing values and inconsistencies in textual data.Tokenize text into individual words or phrases and convert to lowercase.Apply stemming or lemmatization to standardize words.Address challenges such as slang, abbreviations, and multilingual content commonly found in reviews.

2. Feature Extraction: Extract key features from the text to represent its content numerically.Techniques include:Bag of Words (BoW): Represent text as a count of words.TF-IDF (Term Frequency-Inverse Document Frequency): Weigh words by their importance across the dataset.Word Embeddings: Use pre-trained models like Word2Vec or GloVe for capturing semantic meaning.Include contextual features such as review length, star ratings, and product categories for additional insights.

3. Model Development and TrainingData Splitting: Divide the dataset into training, validation, and testing subsets (e.g., 80-10-10 split) to ensure robust evaluation.Model Selection: Use machine learning classifiers such as:Logistic RegressionSupport Vector Machines (SVM)Random Forest or Gradient Boosting algorithmsDeep learning models like Recurrent Neural Networks (RNNs) or transformers (e.g., BERT) for more nuanced text analysis.Address class imbalances in sentiment categories (e.g., fewer negative reviews) using techniques like SMOTE (Synthetic Minority Oversampling Technique).

4. **Prediction and Optimization:** Model Tuning: Optimize hyperparameters such as learning rate, regularization strength, and tree depths to enhance model performance using grid search or randomized search.Generate predictions for each review, classifying sentiment as positive, negative, or neutral, and assess model generalizability on unseen data.

5. **Model Evaluation and Testing:** Evaluation Metrics: Use metrics like Accuracy, Precision, Recall, F1-score, and Area Under the Curve (AUC) to measure classification performance.Error Analysis: Analyze confusion matrices to identify misclassifications and refine the model.Visualization: Use tools like:Word clouds to highlight frequently occurring words in each sentiment category.ROC curves to evaluate model performance.Sentiment distributions to analyze trends across products and categories.

Objective of the Project:

The primary objective of the Flipkart Sentiment Analysis project is to develop an accurate and reliable machine learning model capable of classifying customer reviews into positive, negative, or neutral sentiments based on historical review data. The project aims to achieve a classification accuracy of at least 85% on the test dataset, ensuring the model consistently provides meaningful insights across various product categories. Additionally, the model should enable real-time sentiment analysis, processing reviews with minimal latency for immediate feedback.

To enhance the model's performance and generalization capabilities, advanced data preprocessing and feature engineering techniques will be employed, expanding the dataset by incorporating contextual factors like review length, product category, star ratings, and customer behavior patterns.

By leveraging advanced machine learning algorithms such as Support Vector Machines (SVM), Random Forest, and transformer-based models like BERT, the model will be optimized to minimize misclassifications, targeting an F1-score above 0.85.

The project will also prioritize robustness, ensuring predictions remain consistent even with diverse review styles, slang, and multilingual content. Special consideration will be given to handling imbalanced datasets, such as those with fewer negative reviews, using techniques like oversampling or SMOTE to improve classification accuracy.

To ensure fairness, the model will be tested for balanced performance across all product categories, avoiding biases against specific types of reviews or demographics. The system will be integrated into a web-based application for real-time sentiment tracking, with a goal of achieving at least 90% reliability in live use cases.

Finally, the project will address ethical considerations by ensuring transparency in data usage and prediction methodologies, safeguarding customer privacy, and avoiding biases that could impact product recommendations or ratings. The sentiment analysis system will comply with relevant regulations and uphold trustworthiness for both customers and business stakeholders.

Significance of the Project :

The Flipkart Sentiment Analysis project is pivotal in transforming how customer feedback is understood and utilized in e-commerce. By leveraging machine learning to classify sentiments with high accuracy, this project provides actionable insights into customer preferences, product performance, and overall satisfaction.

Data-Driven Insights for Business Teams :

The project enhances the understanding of customer sentiment by analyzing reviews, ratings, and feedback. This empowers Flipkart to identify trends in customer satisfaction, address pain points, and refine product offerings to meet customer expectations.

Improving Customer Experience :

By accurately assessing customer sentiment, the project helps Flipkart provide personalized recommendations and better product suggestions. It ensures customer concerns are addressed promptly, improving overall satisfaction and loyalty.

Streamlining Product Evaluation :

The model enables real-time analysis of product reviews, providing instant sentiment insights. This streamlines the evaluation process for products, ensuring consistent and data-driven decision-making across categories.

Enhancing Trust and Transparency :

By identifying patterns in feedback, including potential issues like fake reviews or misleading content, the project bolsters customer trust. Accurate sentiment analysis ensures honest reviews are highlighted, improving transparency in product ratings.

Empowering Sellers :

The project provides sellers with insights into how their products are perceived, helping them improve product quality and customer engagement. This fosters a competitive yet customer-centric ecosystem on the platform.

Optimizing Marketing Strategies :

With accurate sentiment insights, Flipkart can optimize marketing campaigns by targeting customer preferences and addressing concerns proactively. This ensures a positive brand perception and boosts sales.

Chapter 2:

2.1- "Importing Required Libraries for Predicting Flipkart Sentimental Analysis"

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
import warnings

warnings.filterwarnings("ignore")

# Load your dataset (replace 'Flipkart_Reviews.csv' with your actual dataset file name)
full_data = pd.read_csv("Flipkart Reviews And Sentiment Analysis.csv")

# Split the data into training and validation sets
# Adjust `test_size` and `random_state` as needed
train_data, valid_data = train_test_split(full_data, test_size=0.15, random_state=0)

# Save the split datasets to CSV files
train_data.to_csv("train_data.csv", index=False)
valid_data.to_csv("valid_data.csv", index=False)

# Load the training dataset for further analysis
data = pd.read_csv("train_data.csv")

# Display first few rows of the training dataset
print(data.head())
```

The code in the image imports essential Python libraries for sentimental analysis prediction project in flipkart. These libraries provide functionalities such as data manipulation (pandas, numpy), visualization (matplotlib, seaborn), and machine learning (scikit-learn, xgboost). Additionally, modules like imblearn support handling imbalanced datasets, and utilities such as warnings improve the coding experience by managing alerts.

2.2- Dataset Overview: First 5 Rows of Flipkart Sentimental Analysis Data

```
print(data.head())
                                          product_name product_price Rate  \
0  I Kall IK22 40 W Bluetooth Home TheatreÐ⊡Ó®Ð⊡Â...          1549    5
1  Pigeon Favourite 15024 500 W Juicer Mixer Grin...          2099    3
2  boAt Wave Neo with 1.69 inch , 2.5D Curved Dis...          1699    5
3     Women Regular Fit White Viscose Rayon Trousers           349    5
4  LEADER TORFIN 26T MTB Bicycle without Gear Sin...          6499    5

           Review                                        Summary Sentiment  \
0  great product                                   sounds good   positive
1           fair                                       avrrage    neutral
2  great product                                   good product  positive
3            NaN  i love this trousers cloth material is so good...  positive
4         super!                                 nice im so happy  positive

   reviewer_id
0          NaN
1          NaN
2          NaN
3          NaN
4          NaN
```

2.3 - We inspect the dataset using functions like `info()` and `describe()`.

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 174294 entries, 0 to 174293
Data columns (total 7 columns):
 #   Column         Non-Null Count   Dtype
---  ------         --------------   -----
 0   product_name   174294 non-null  object
 1   product_price  174294 non-null  object
 2   Rate           174292 non-null  object
 3   Review         153305 non-null  object
 4   Summary        174286 non-null  object
 5   Sentiment      174294 non-null  object
 6   reviewer_id     3080 non-null   float64
dtypes: float64(1), object(6)
memory usage: 9.3+ MB
```

2.4. Data Preprocessing

We check for missing values and duplicates in the dataset and handle them appropriately. Categorical variables are encoded using LabelEncoder, and numerical features are prepared for further
analysis.

```
#data cleaning if found any null value
print(data.isnull().sum())
```

```
product_name         0
product_price        0
Rate                 2
Review           20989
Summary              8
Sentiment            0
reviewer_id     171214
dtype: int64
```
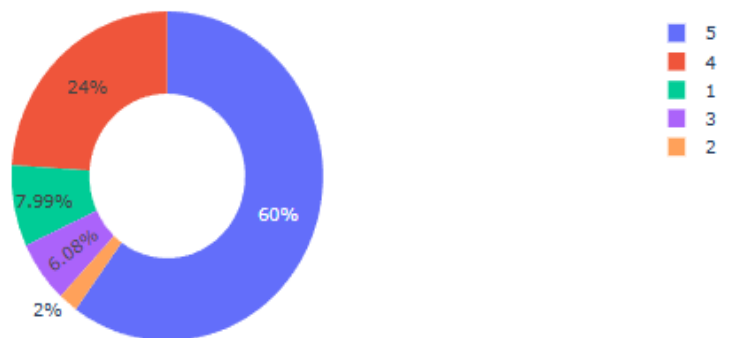
2.5 Seniment Analysis of Reviews
Chceking the Rating column

```python
#Seniment Analysis of Reviews
#Chceking the Rating column

ratings = data["Rating"].value_counts()
numbers = ratings.index
quantity = ratings.values

import plotly.express as px
#https://plotly.com/python/  (check plotly for more interactive graphs)

figure = px.pie(data,
          values=quantity,
          names=numbers,hole = 0.5)
figure.show()
```
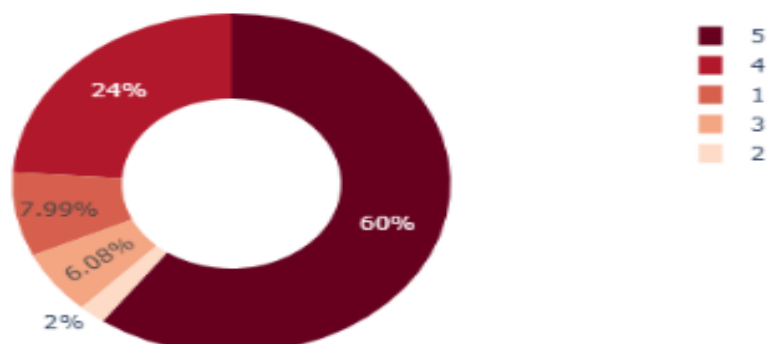


```python
#Seniment Analysis of Reviews
#Chceking the Rating column

ratings = data["Rating"].value_counts()
numbers = ratings.index
quantity = ratings.values

import plotly.express as px
#https://plotly.com/python/  (check plotly for more interactive graphs)

figure = px.pie(data,
          values=quantity,
          names=numbers,hole = 0.5, color_discrete_sequence=px.colors.sequential.RdBu)
figure.show()
```

```
pip install --upgrade pillow
```

Requirement already satisfied: pillow in c:\users\lenovo\anaconda3\lib\site-packages (11.0.0)Note: you may need to restart the kernel to use updated packages.

```
!pip install freetype-py
```

Requirement already satisfied: freetype-py in c:\users\lenovo\anaconda3\lib\site-packages (2.5.1)

```python
import freetype
print("Freetype module imported successfully!")
```

Freetype module imported successfully!

```python
#pip install --upgrade pip
#pip install --upgrade pillow
#pip install freetype
import freetype

text = " ".join(i for i in data.Review)
stopwords = set(STOPWORDS)
wordcloud = WordCloud(stopwords=stopwords,
                      background_color="white").generate(text)
plt.figure( figsize=(15,10))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis("off")
plt.show()
```

```
nltk.download('vader_lexicon')
sentiments = SentimentIntensityAnalyzer()
data["Positive"] = [sentiments.polarity_scores(i)["pos"] for i in data["Review"]]
data["Negative"] = [sentiments.polarity_scores(i)["neg"] for i in data["Review"]]
data["Neutral"] = [sentiments.polarity_scores(i)["neu"] for i in data["Review"]]
data = data[["Review", "Positive", "Negative", "Neutral"]]
print(data.head())
```

```
[nltk_data] Downloading package vader_lexicon to
[nltk_data]     C:\Users\Lenovo\AppData\Roaming\nltk_data...
[nltk_data]   Package vader_lexicon is already up-to-date!
         Review  Positive  Negative  Neutral
0  great product     0.804       0.0    0.196
1           fair     1.000       0.0    0.000
2  great product     0.804       0.0    0.196
3            nan     0.000       0.0    1.000
4          super     1.000       0.0    0.000
```

```python
x = sum(data["Positive"])
y = sum(data["Negative"])
z = sum(data["Neutral"])

def sentiment_score(a, b, c):
    if (a>b) and (a>c):
        print("Positive 😊 ")
    elif (b>a) and (b>c):
        print("Negative 😠 ")
    else:
        print("Neutral 🙂 ")
sentiment_score(x, y, z)
```

```
Neutral 🙂
```

```python
print("Positive: ", x)
print("Negative: ", y)
print("Neutral: ", z)
```

```
Positive:   65015.249
Negative:   6502.829
Neutral:  102770.926
```

Chapter 3 :
Reviews:

The current sentiment analysis model demonstrates solid potential but offers numerous opportunities for improvement. Future enhancements could involve training the model on larger, more diverse datasets, incorporating reviews from various product categories and regions, and expanding the number of features related to user sentiment.

Advanced techniques such as ensemble learning and deep learning could improve sentiment classification accuracy and better capture nuanced customer opinions. Additionally, integrating real-time data from ongoing customer reviews and feedback systems could enable dynamic sentiment tracking and timely responses to emerging trends.

Ethical considerations, such as avoiding algorithmic bias in sentiment classification and ensuring the privacy of customer data, should remain a priority. Expanding the model to account for external factors like seasonal trends, regional preferences, and user-specific data like purchase history could enhance its precision.

Collaborating with domain experts in e-commerce and customer behavior could refine feature selection and improve model validation. Incorporating automated data collection methods, such as scraping reviews and analyzing customer interaction patterns, would further enhance prediction accuracy.

Deploying the sentiment analysis system in a live Flipkart environment and continuously updating it with fresh data will ensure that the predictions remain relevant and actionable. The system could also be expanded to analyze reviews across multiple product categories, enabling targeted interventions for improving customer satisfaction. Additionally, integrating alternative data sources, like social media activity or user-generated multimedia content, could provide deeper insights into customer sentiments.

Incorporating multi-language and regional support would improve the system's scalability and make it more applicable to Flipkart's diverse user base. Gathering user feedback on system performance and iteratively fine-tuning the model based on real-world scenarios would enhance its adaptability, making it a powerful tool for Flipkart to understand customer sentiments and improve overall user experience.

Conclusion:

In conclusion, this project demonstrated the potential of machine learning in analyzing customer sentiment for Flipkart product reviews, effectively interpreting feedback to classify sentiments and offer actionable insights. It highlights the importance of leveraging advanced algorithms to improve customer satisfaction, product recommendations, and overall decision-making. Continuous refinement, such as incorporating regional preferences, real-time feedback, and diverse data sources, is essential to enhance precision and scalability. Ethical considerations, including fairness, transparency, and data privacy, remain critical for responsible application. By laying the foundation for advancements in sentiment analysis, this work provides significant potential to transform how Flipkart understands and responds to customer sentiments, ultimately improving user experiences and fostering stronger customer relationships.

Referance:

- **Kaggle-** https://www.kaggle.com/datasets/amitkumarsamal/flipkart-sentimental-analysis

ASSESSMENT

Internal:

SL NO	RUBRICS	FULL MARK	MARKS OBTAINED	REMARKS
1	Understanding the relevance, scope and dimension of the project	10		
2	Methodology	10		
3	Quality of Analysis and Results	10		
4	Interpretations and Conclusions	10		
5	Report	10		
	Total	**50**		

Date: **Signature of the Faculty**

COURSE OUTCOME (COs) ATTAINMENT

➢ **Expected Course Outcomes (COs):**

(Refer to COs Statement in the Syllabus)

_____

_____

_____

_____

➢ **Course Outcome Attained:**

How would you rate your learning of the subject based on the specified COs?

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
1   2   3   4   5   6   7   8   9   10

LOW   **HIGH**

➢ **Learning Gap (if any):**

_____

_____

_____

_____

➢ **Books / Manuals Referred:**

_____

_____

_____

_____

Date: **Signature of the Student**

➢ **Suggestions / Recommendations:**

(By the Course Faculty)

_____

_____

_____

_____

Date: **Signature of the Faculty**