DEVELOPMENT AND DEPLOYMENT OF SENTIMENT ANALYZER: A CASE STUDY OF ALIEXPRESS ELECTRONIC PRODUCTS

A PRESENTATION BY TEAM BRAINIAC

TABLE OF CONTENT

PROBLEM STATEMENT

In today’s digital age, e-commerce platforms face the challenge of analyzing vast amounts of customer feedback to understand product sentiment accurately. Understanding customer sentiment is crucial for businesses to make informed decisions about product improvements, marketing strategies, and customer satisfaction. However, manually analyzing thousands of product reviews is time-consuming and inefficient. Therefore, there is a need for an automated sentiment analysis solution to process and interpret these reviews efficiently.

PROJECT OBJECTIVES

The aim of this project is to develop and deploy a comprehensive sentiment analysis system for e-commerce product reviews, specifically focusing on the Electronics category on Aliexpress by utilizing advanced Data Science and Data Engineering techniques to create an end-to-end solution capable of extracting, preprocessing, analyzing, and visualizing customer sentiment from textual reviews.

DATA ARCHITECTURE

EXTRACT

TRANSFORM

LOAD

Airflow used to orchestrate the ETL process

AliExpress

Aliexpress website

Extraction(Web Scraping)

Python Script
(script to extract
and load data)

Google Drive
(store raw file)

Transformation

Python Script
(script to
transform data)

Loading

Google Drive
(store processed
file)

ETL PROCESS

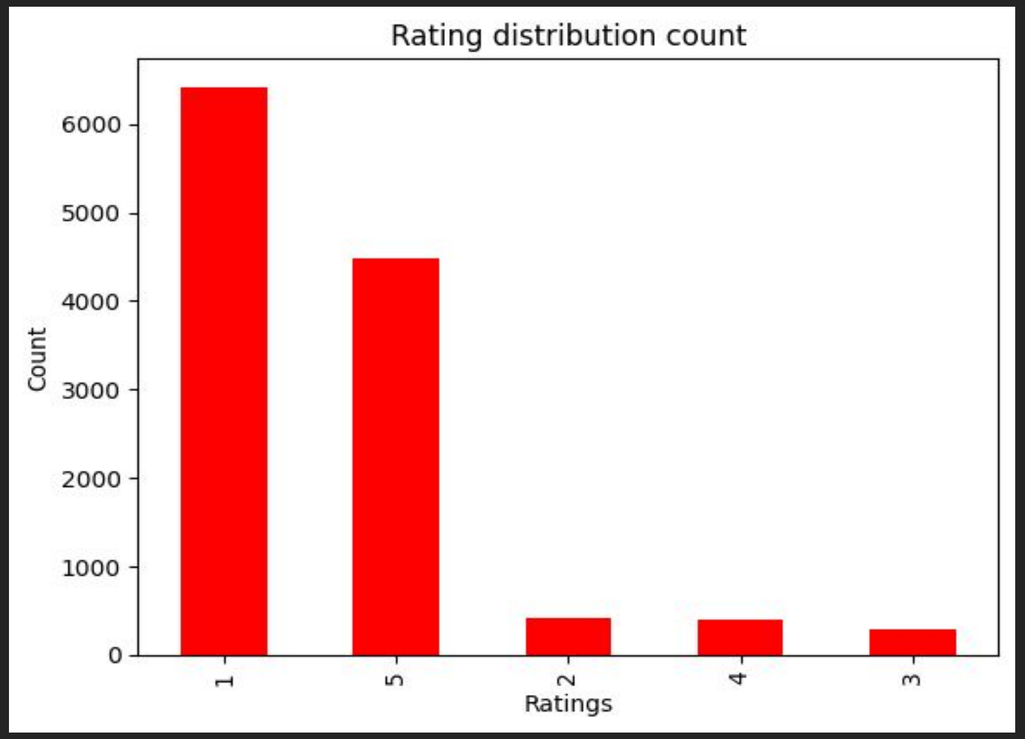
1. **DATA EXTRACTION:**
   - Webscraping: The electronics data was extracted from the Aliexpress website using selenium and beautiful soup. The data was then stored in a folder on a google drive called “raw”. The following data was extracted:
     - Reviewer
     - Country
     - Review
     - Subject
     - Star rating
     - Date

2. **DATA TRANSFORMATION:**
   - The raw data was extracted from the google drive and transformed. Duplicates were removed, the column renamed, null values dropped and transformed data was written in a csv format and stored in a dataframe

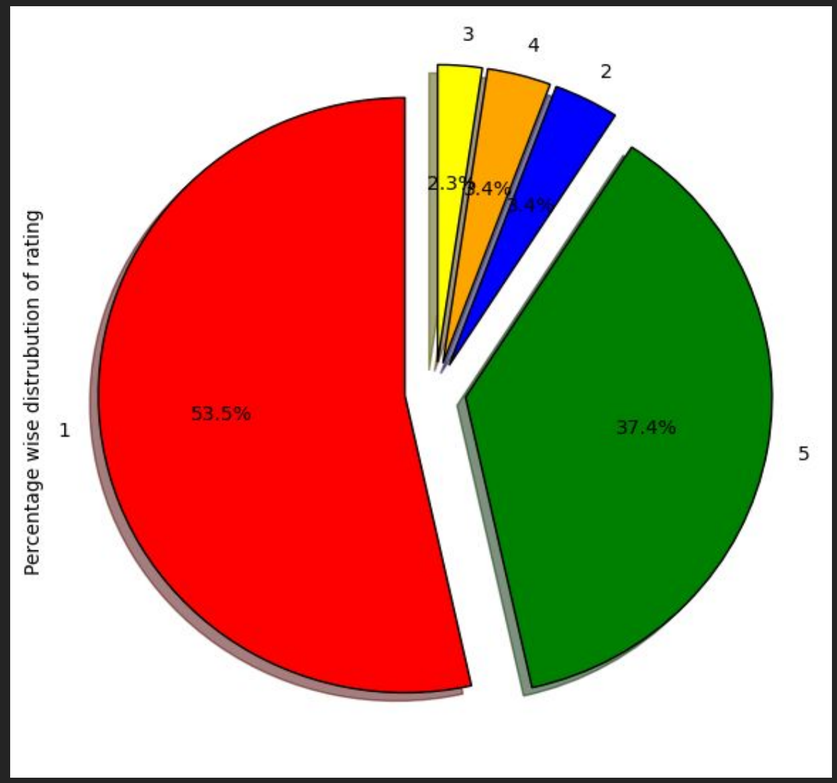
3. **DATA LOADING:**
   - The transformed data was loaded into a folder called “processed” on a google drive.

EXPLORATORY DATA ANALYSIS (EDA)

- 12,000 electronic product reviews were analyzed
- Star Rating: 1 as Very bad, 2 as Bad , 3 as Good, 4 as Very Good, 5 as Excellent
- Highest star rating is 1 and the lowest is 2

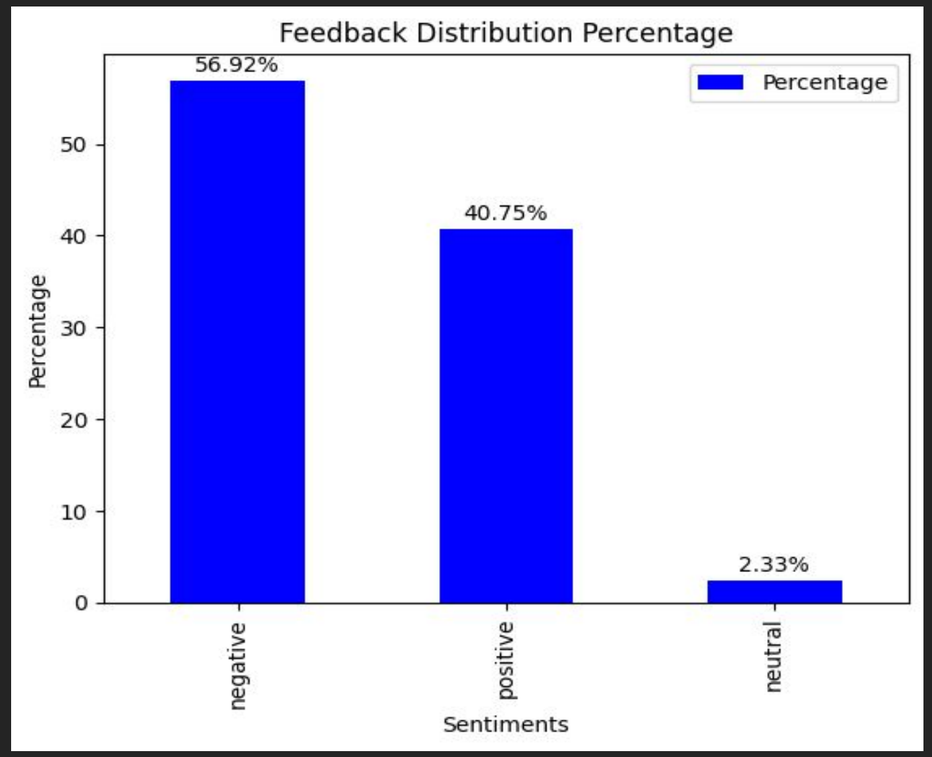


Over 6000 product had rating of 1, about 4500 had a rating of 5 while 2, 3 and 4 had ratings below 500

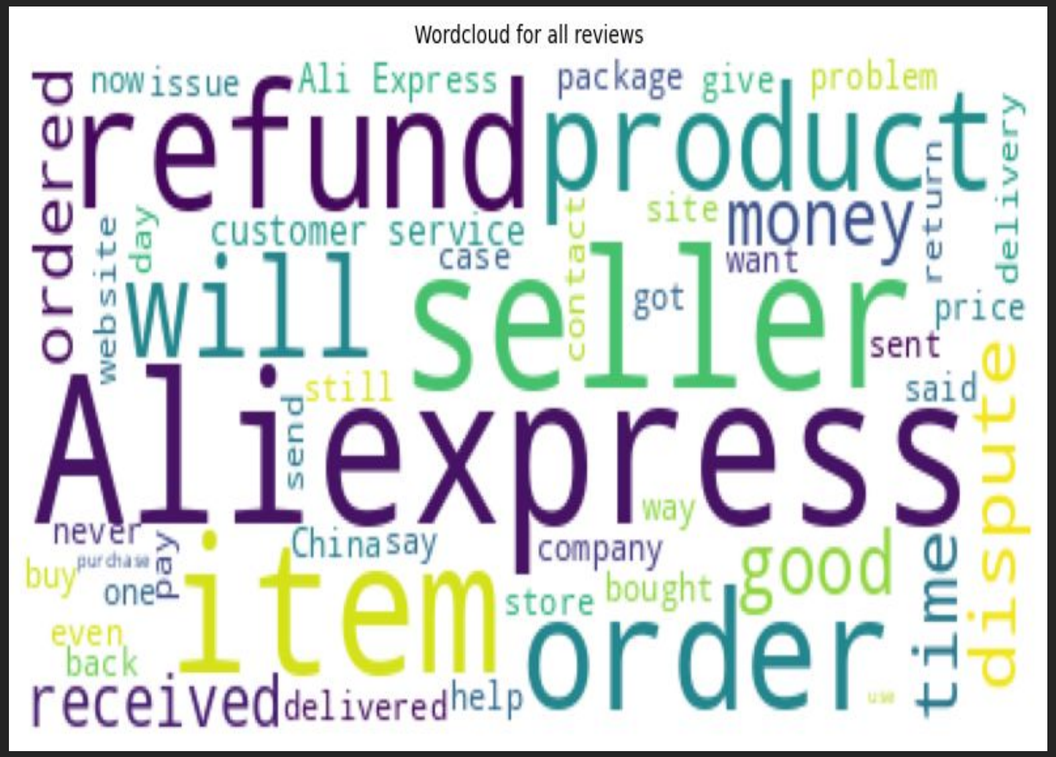


53.5% of customers rates the product as Very Bad
37.4% of customers rates the product as Excellent
3.4% of customers rates at Very Good and Bad, while 2.3% rates at Good

EXPLORATORY DATA ANALYSIS



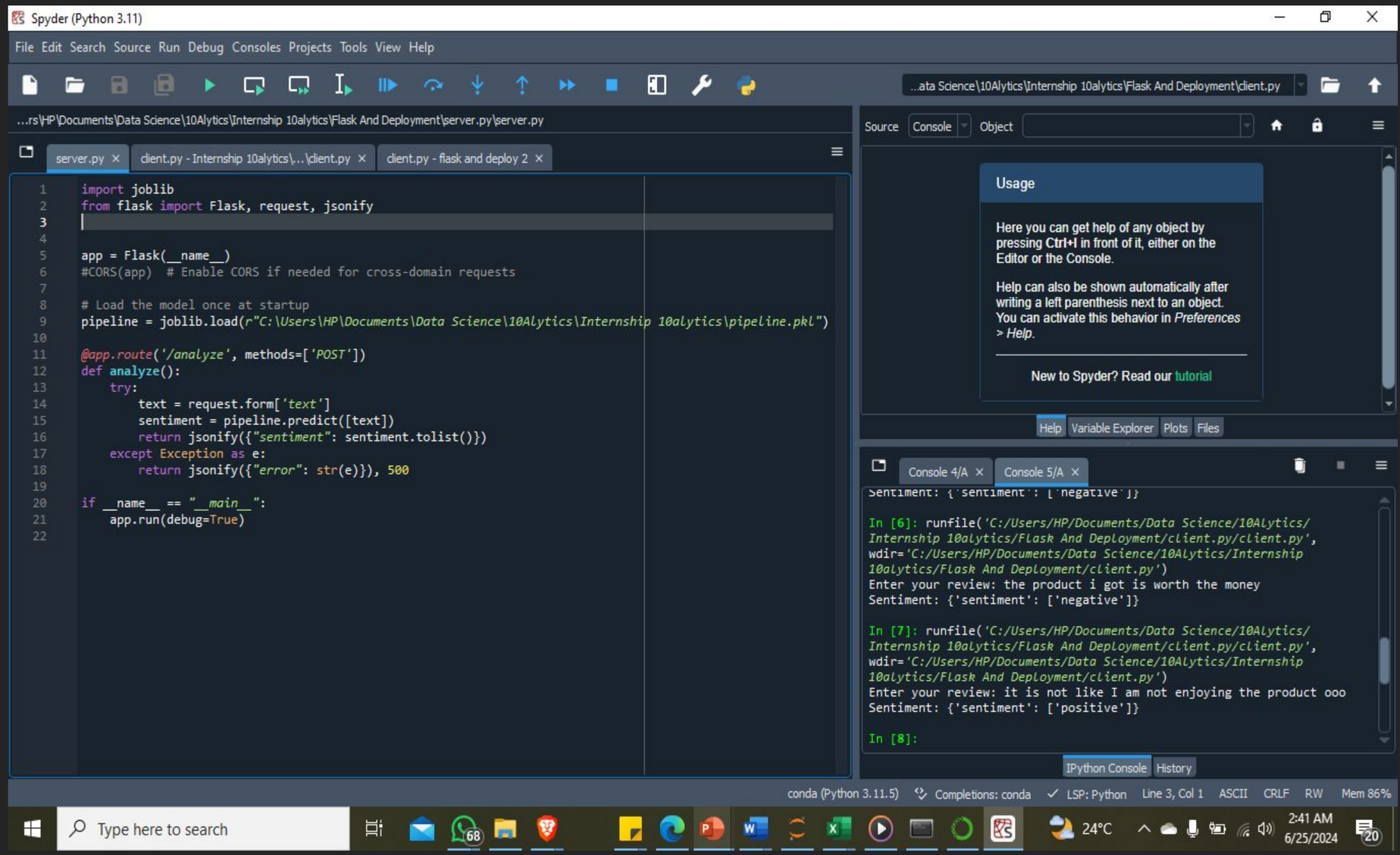
56.92% have a negative feedback on the product
40.75% have a positive feedback while 2.33% of customers were neutral on their feedback.



Most of the commonly used words in the feedback were Refund, Money, Want, Dispute, Return, Product. This indicates displeasure or dissatisfaction of purchase.

PERFORMANCE OF THE DEVELOPED MODELS

model	accuracy	precision	recall	f1-score	time	average sentiment
Naives Bayes Model	0.89505	0.90405	0.89505	0.89259	9.64 um	
Random Forest Pipeline Model	0.90316	0.90469	0.90316	0.90221	6.2 um	
Logistic Regression Pipeline	0.93686	0.93693	0.93686	0.93667	6.2 um	
VADER Sentiment Analyzer					2.00187	0.05266
TextBlob Sentiment Analyzer					0.74726	0.09334
RoBERTa Sentiment Analyzer					277.08	-0.3561


Spyder (Python 3.11)
File Edit Search Source Run Debug Consoles Projects Tools View Help
...rs\HP\Documents\Data Science\10Alytics\Internship 10alytics\Flask And Deployment\server.py\server.py
server.py
client.py - Internship 10alytics\...\client.py
client.py - flask and deploy 2
import joblib
from flask import Flask, request, jsonify
app = Flask(__name__)
#CORS(app) # Enable CORS if needed for cross-domain requests
# Load the model once at startup
pipeline = joblib.load(r"C:\Users\HP\Documents\Data Science\10Alytics\Internship 10alytics\pipeline.pkl")
@app.route('/analyze', methods=['POST'])
def analyze():
try:
text = request.form['text']
sentiment = pipeline.predict([text])
return jsonify({"sentiment": sentiment.tolist()})
except Exception as e:
return jsonify({"error": str(e)}), 500
if __name__ == "__main__":
app.run(debug=True)
Usage
Here you can get help of any object by pressing Ctrl+I in front of it, either on the Editor or the Console.
Help can also be shown automatically after writing a left parenthesis next to an object. You can activate this behavior in Preferences > Help.
New to Spyder? Read our tutorial
Help
Variable Explorer
Plots
Files
Console 4/A
Console 5/A
In [6]: runfile('C:/Users/HP/Documents/Data Science/10Alytics/Internship 10alytics/Flask And Deployment/client.py/client.py', wdir='C:/Users/HP/Documents/Data Science/10Alytics/Internship 10alytics/Flask And Deployment/client.py')
Enter your review: the product i got is worth the money
Sentiment: {'sentiment': ['negative']}
In [7]: runfile('C:/Users/HP/Documents/Data Science/10Alytics/Internship 10alytics/Flask And Deployment/client.py/client.py', wdir='C:/Users/HP/Documents/Data Science/10Alytics/Internship 10alytics/Flask And Deployment/client.py')
Enter your review: it is not like I am not enjoying the product ooo
Sentiment: {'sentiment': ['positive']}
In [8]:
IPython Console
History
conda (Python 3.11.5)
Completions: conda
LSP: Python
Line 3, Col 1
ASCII
CRLF
RW
Mem 86%
Type here to search
24°C
2:41 AM
6/25/2024

BUSINESS IMPACTS OF THE MODEL

Customer Insights and Preferences

Understanding Customer Sentiment: Sentiment analysis can help businesses understand how customers feel about their electronic products, identifying common praises and complaints.
Product Improvement: By analyzing customer feedback, businesses can identify specific features or aspects of their products that are appreciated or criticized, guiding product development and improvements.
Targeted Marketing: Insights from sentiment analysis can be used to tailor marketing strategies to highlight popular features and address common concerns, making marketing efforts more effective.

Customer Service and Support

Enhancing Customer Support: Sentiment analysis can highlight recurring issues that customers face, allowing businesses to improve their customer service and support systems to address these issues more effectively.
Proactive Issue Resolution: Businesses can identify and resolve issues before they escalate, improving customer satisfaction and loyalty.

Sales and Revenue Growth

Optimizing Product Listings: Positive sentiments can be highlighted in product descriptions and reviews, enhancing the attractiveness of listings and potentially increasing sales.
Predicting Sales Trends: Sentiment analysis can help predict sales trends by correlating positive sentiments with sales spikes, enabling better inventory and supply chain management.

Customer Loyalty and Retention

Building Trust and Loyalty: By addressing the feedback and improving products based on customer sentiments, businesses can build stronger customer trust and loyalty.
Personalized Experiences: Understanding customer sentiment allows for more personalized interactions and recommendations, enhancing the overall customer experience.

CONCLUSION

1. **Sentiment Analysis Effectiveness:**
   - The Logistic Regression model is the most effective for sentiment analysis of electronic product reviews on Aliexpress, demonstrating superior performance metrics.
2. **Rating Distribution Insights:**
   - The rating distribution indicates a significant proportion of extremely negative reviews (53.5% for rating 1) and a substantial portion of extremely positive reviews (37.4% for rating 5). This polarized distribution highlights the need for further analysis to understand the drivers behind these ratings.
3. **Sentiment Distribution:**
   - The sentiment distribution reveals that 56.92% of the feedback is negative, 40.75% is positive, and 2.33% is neutral. This indicates a generally negative customer perception, suggesting that targeted strategies are necessary to address customer concerns.
4. **Automated System Benefits:**
   - The automated sentiment analysis system significantly reduces the time and effort required to process large volumes of reviews, providing timely insights for business decisions. Our analyzer can process 1 million reviews in about 6 secs or less.
5. **Strategic Actions:**
   - Leverage insights from sentiment analysis to improve product quality and customer satisfaction. Addressing common negative feedback themes can lead to improved overall ratings and customer loyalty.
6. **API Implementation:**
   - The provided Flask-based API allows for efficient deployment and integration of the sentiment analysis system into the existing e-commerce platform, facilitating real-time sentiment analysis and continuous monitoring.

RECOMMENDATION

Model Implementation:

- Deploy the **Logistic Regression Pipeline** for sentiment analysis, as it shows the highest performance metrics in terms of accuracy (0.93686), precision (0.93693), recall (0.93686), and F1-score (0.93667).

Real-Time Analysis Consideration:

- For scenarios where rapid response is critical, consider using the **VADER Sentiment Analyzer** due to its low processing time (2.00187 um).

Integration with E-commerce Platform:

- Integrate the chosen sentiment analysis model into the Aliexpress review system to automatically process and categorize reviews. The provided Python script, which uses Flask for creating an API endpoint, can be utilized for this purpose.

API Deployment:

- This API will allow seamless integration with the e-commerce platform, enabling real-time sentiment analysis.

Visualization and Reporting:

- Use the provided visualizations to present sentiment distribution and ratings to stakeholders. This will assist in making informed decisions about product improvements and marketing strategies.

Continuous Monitoring:

- Continuously monitor sentiment distribution and feedback. Investigate products with high negative sentiment and take necessary actions to improve customer satisfaction.

TEAM MEMBERS

NAME	LINKEDIN
Ifeyinwa Ibekwe	https://www.linkedin.com/in/ifeyinwa-ibekwe-625923297
Dr. Solomon Omale	https://www.linkedin.com/solomon-okpanachi-omale-26a911109/
Usman Sampras	https://www.linkedin.com/in/usmansampras/
Olushola Kareem	http://linkedin.com/in/olushola-kareem-msc-043a62237
Stellamaris Enwerem	http://linkedin.com/in/stellamaris-enwerem-1a032216a
Stephen Ogodo	https://www.linkedin.com/in/stephen-ogodo-a270016b/
Unuero Thompson	https://www.linkedin.com/in/thompson-unuero-291715201
Joel Madumere	https://www.linkedin.com/in/joel-madumere
Chika Onyenenue	https://www.linkedin.com/in/chika-onyenenue-19a8a71b7

Question and Answer

Thank you