TCS iON RIO 125 - Automate sentiment analysis of textual comments and feedback

Internship report

Submitted to

Mathematics & Statistics

Faculty of Science & Technology

Under TCS ION INDUSTRY HONOUR PROGRAM

Vishwakarma University, Pune (Maharashtra)

By

Prasad Chandrakant Kabade

Under the supervision of

INDUSTRY MENTOR	FACULTY MENTOR
TATA CONSULTANCY SERVICES	VISHWAKARMA UNIVERSITY, PUNE

Internship Project Title	TCS iON RIO 125 - Automate sentiment analysis of textual comments and feedback
Project Title	Sentiment Analysis for Amazon Product Reviews
Name of the Company	TCS iON
Name of the Industry Mentor	Debashis Roy
Name of the Institute	Vishwakarma University, Pune

Start Date	End Date	Total Effort (hrs.)	Project Environment	Tools used
07/12/2023	05/01/2024	126	Jupyter Notebook	Python (Pandas, seaborn, matplotlib, tensorflow, torch, Numpy, nrclex, TextBlob, Transformer).

Project Synopsis:

Objective: The objective of sentiment analysis on the Amazon Fine Food Reviews dataset is to classify reviews as positive, negative, or neutral, compare sentiment analysis techniques/models, identify factors influencing sentiment, analyze sentiment trends over time, and generate actionable insights for business improvement.

Key Components:

Data Preparation: This involves loading the dataset, understanding its structure, and preprocessing the text data. Preprocessing steps may include removing special characters, tokenization, converting text to lowercase, and removing stop words.

Sentiment Analysis Techniques: Implement different sentiment analysis techniques such as lexicon-based methods (e.g., VADER) and machine learning/deep learning models (e.g., RoBERTa). Each technique may require specific libraries, configurations, and training data.

Evaluation Metrics: Define evaluation metrics to assess the performance of sentiment analysis models. Common metrics include accuracy, precision, recall, F1-score, and ROC-AUC for binary classification tasks.

Visualization: Visualize the results of sentiment analysis using plots and graphs. This could include bar plots showing the distribution of sentiment scores, scatter plots for comparing sentiment scores between models, and line plots for analyzing sentiment trends over time.

Interpretation: Interpret the results of sentiment analysis to gain insights into customer opinions and sentiments. This involves analyzing the relationship between sentiment scores and other variables (e.g., product features, review length, reviewer demographics) to understand factors influencing sentiment.

Actionable Insights:Use the insights gained from sentiment analysis to make data-driven decisions for business improvement. This could involve addressing areas of concern highlighted by negative sentiment, optimizing marketing strategies based on positive sentiment, or enhancing product features to meet customer preferences.

Key Technologies Used:

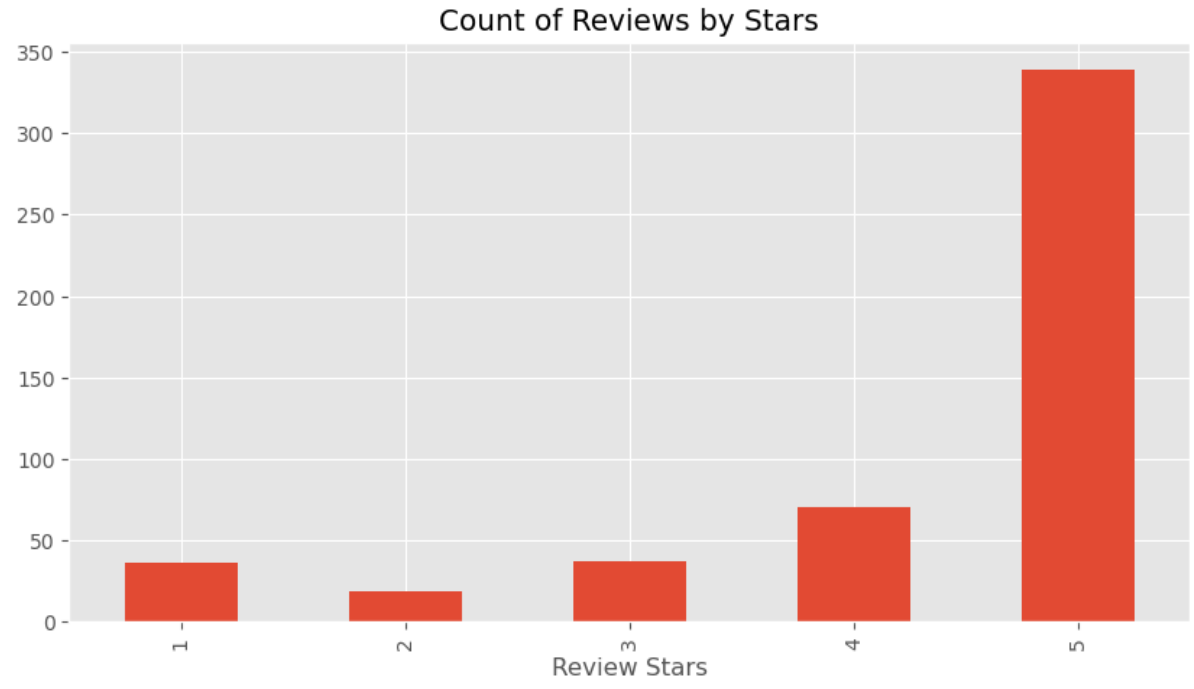
Python

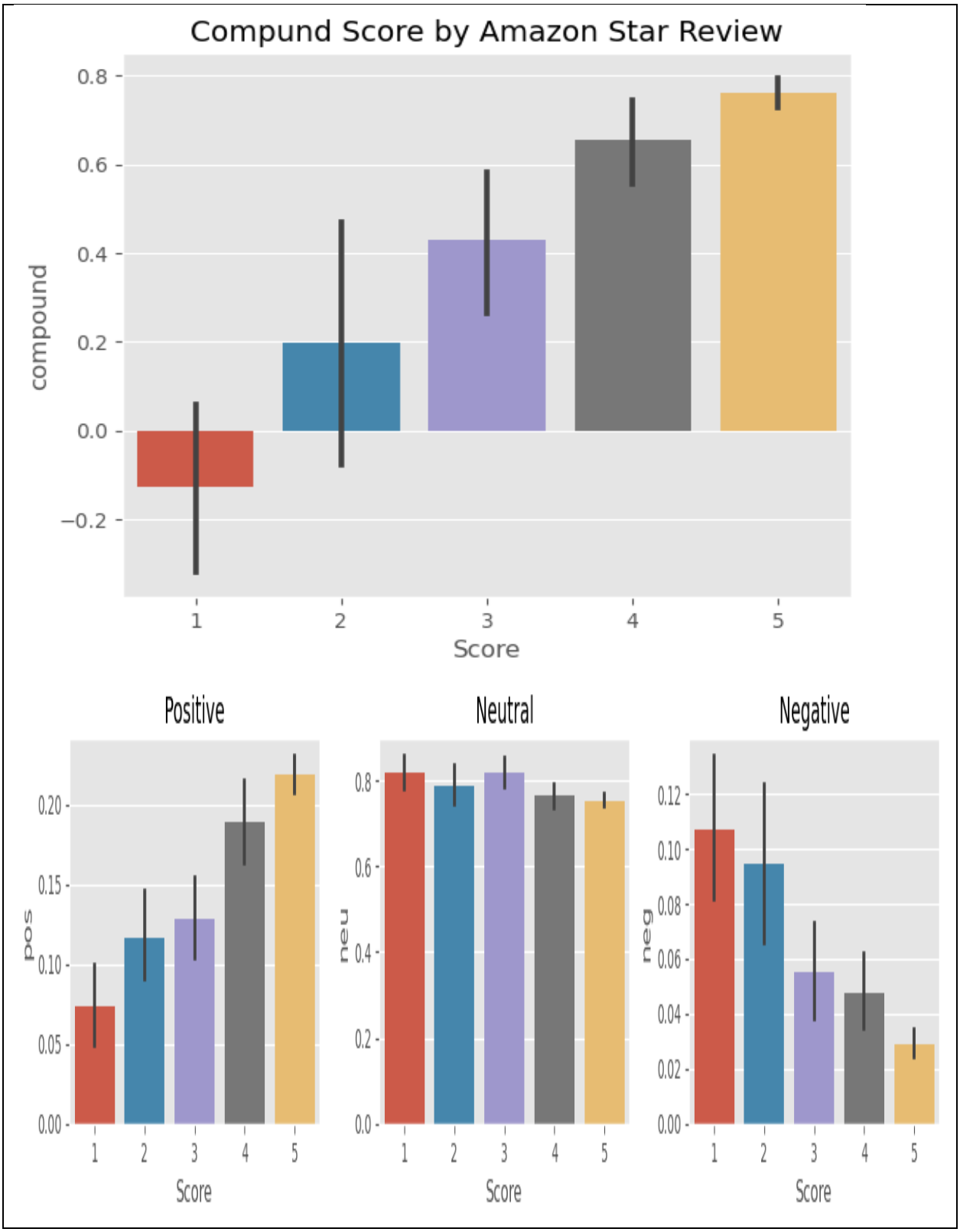
Pandas, NLTK, VEDAR, Transformer library,
Scikit learn,matplotlib and seaborn

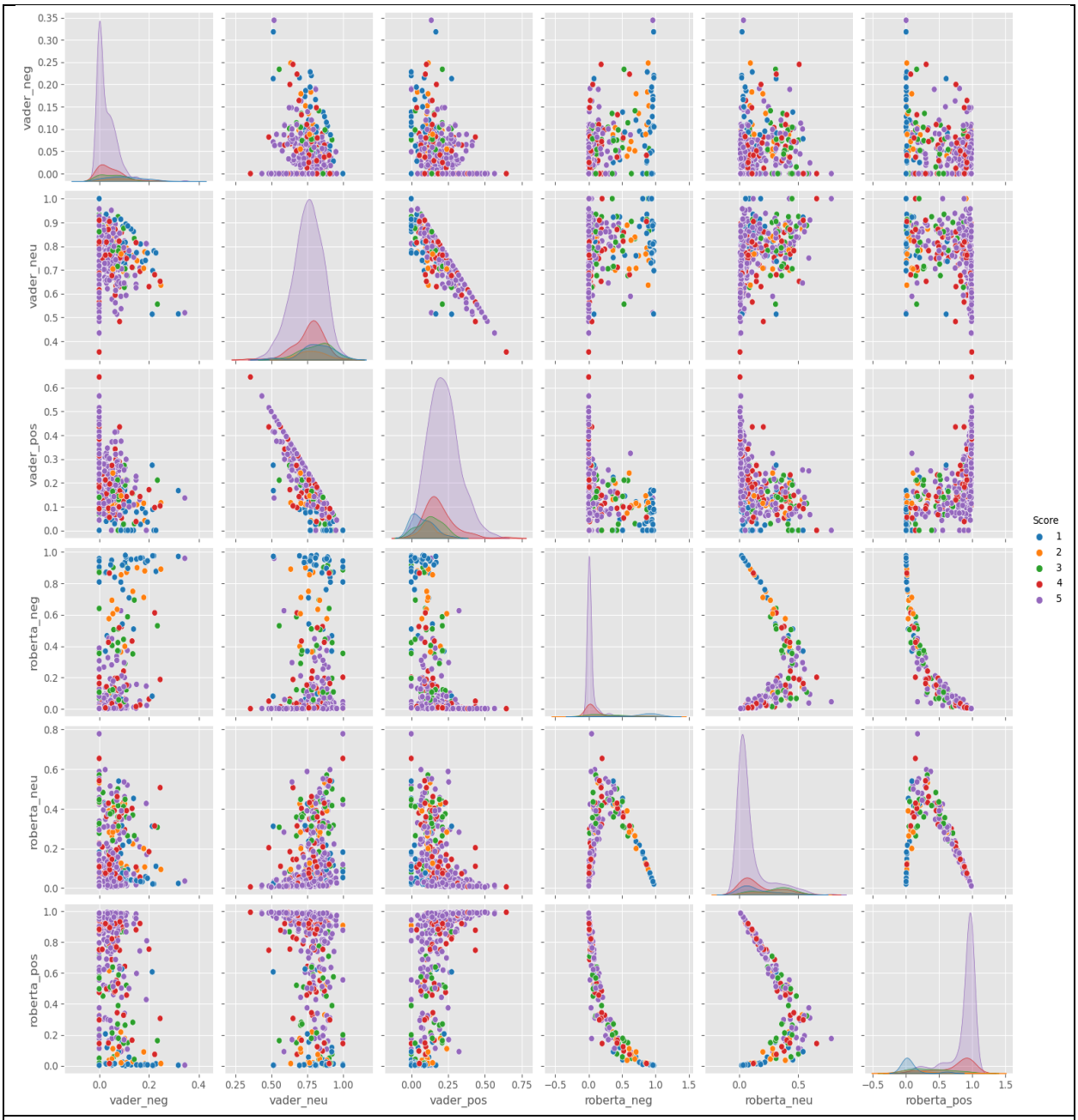
Solution Approach: The solution approach for the sentiment analysis project on the Amazon Fine Food Reviews dataset involves acquiring the dataset and understanding its structure, followed by preprocessing the text data to clean and normalize it. After conducting exploratory data analysis to gain insights into the distribution of review scores and other relevant features, sentiment analysis techniques such as VADER and machine learning models like RoBERTa are implemented to classify the reviews into positive, negative, or neutral sentiments. Comparative analysis is then conducted to evaluate the performance of these techniques, identifying key factors influencing sentiment and deriving actionable insights to improve customer satisfaction. Finally, documentation and presentation of the findings are prepared to effectively communicate the insights and recommendations to stakeholders.

Assumptions: For the sentiment analysis project on the Amazon Fine Food Reviews dataset, assumptions include the assumption that the reviews accurately represent customers' opinions and experiences, reflecting the broader sentiment towards the products. It is assumed that the language in the reviews is clear and consistent, facilitating accurate sentiment analysis, and that the dataset is of high quality, with minimal errors or biases. The findings are assumed to be generalizable to similar food products and online retail platforms, and ethical considerations regarding data privacy and usage are assumed to be addressed in compliance with relevant guidelines and regulations.

Project Diagrams:




Compund Score by Amazon Star Review
compound
0.8
0.6
0.4
0.2
0.0
−0.2
1
2
3
4
5
Score
Positive
pos
0.20
0.15
0.10
0.05
0.00
Score
Neutral
neu
0.8
0.6
0.4
0.2
0.0
Score
Negative
neg
0.12
0.10
0.08
0.06
0.04
0.02
0.00
Score



Algorithms: For sentiment analysis on the Amazon Fine Food Reviews dataset, a variety of algorithms can be employed. This includes lexicon-based methods like VADER, which assigns sentiment scores to words and aggregates them to compute an overall sentiment. Machine learning algorithms such as logistic regression, naive Bayes, support vector machines (SVM), random forests, gradient boosting machines (GBM), and neural networks can be utilized for classification tasks. Deep learning models like BERT and RoBERTa, which leverage transformer architectures, offer state-of-the-art performance and can be fine-tuned for sentiment analysis. Ensemble methods and hybrid approaches, combining the strengths of multiple algorithms, are also effective strategies to improve sentiment analysis accuracy. By exploring these diverse algorithms, analysts can determine the most suitable approach for extracting sentiment insights from the Amazon Fine Food Reviews dataset

Outcome: The outcome of the sentiment analysis project on the Amazon Fine Food Reviews dataset is multifaceted:

Insights into Customer Sentiment: The project provides valuable insights into the sentiments expressed by customers towards the food products sold on Amazon. This includes understanding the overall distribution of positive, negative, and neutral sentiments, as well as identifying specific product attributes or experiences that contribute to varying sentiment levels.

Evaluation of Sentiment Analysis Techniques: The project evaluates the performance of different sentiment analysis techniques and algorithms, allowing for comparisons between lexicon-based methods, machine learning models, and deep learning architectures. This evaluation helps in determining the most effective approach for sentiment analysis in the context of food reviews.

Identification of Improvement Opportunities: By analyzing the factors influencing sentiment and correlating them with product features or customer experiences, the project identifies opportunities for product improvement, marketing strategies, or customer service enhancements. For example, it may highlight areas where product quality can be enhanced or customer concerns addressed to improve overall satisfaction.

Actionable Insights for Business Decision-Making: The project generates actionable insights that can inform business decision-making processes. These insights may include recommendations for product development, marketing campaigns, customer engagement strategies, or operational improvements aimed at enhancing customer satisfaction, driving sales, and strengthening brand reputation.

Documentation and Reporting: The project outcomes are documented and presented in a clear, concise manner, often accompanied by visualizations and reports. This documentation serves as a valuable resource for stakeholders, enabling them to understand the findings, recommendations, and next steps for implementation.

Overall, the outcome of the sentiment analysis project empowers businesses with actionable insights derived from customer feedback, enabling them to make informed decisions that positively impact product offerings, customer experiences, and business performance.

Exceptions considered: In the sentiment analysis project on the Amazon Fine Food Reviews dataset, several exceptions are considered to ensure the accuracy and reliability of the analysis. These include handling data quality issues such as missing values and outliers, standardizing language variations including slang and misspellings, capturing the contextual understanding of reviews to account for nuances like sarcasm, adhering to privacy and ethical guidelines regarding user data, acknowledging the limitations of sentiment analysis models including biases and generalization errors, and optimizing computational resources for efficient processing. By addressing these exceptions comprehensively, the project aims to produce trustworthy insights into customer sentiments towards food products on Amazon while maintaining the integrity of the analysis process.

Enhancement Scope: The sentiment analysis project on the Amazon Fine Food Reviews dataset presents numerous opportunities for enhancement. These include fine-tuning sentiment analysis models through domain-specific training or leveraging transfer learning techniques, expanding the analysis to perform aspect-based sentiment analysis for more granular insights, and incorporating multimodal data such as images or videos for a comprehensive understanding of customer sentiments. Additionally, temporal analysis can reveal sentiment trends over time, while sentiment propagation techniques can uncover patterns of sentiment diffusion. Interactive visualization tools can facilitate dynamic exploration of sentiment analysis results, and correlation analysis can unveil relationships between sentiment and other metadata variables. Cross-domain analysis offers the chance to compare sentiment patterns across different product categories, and integrating sentiment prediction models enables proactive decision-making. Finally, feedback integration ensures that sentiment analysis insights contribute to continuous improvement in customer satisfaction and product quality.
Link to Code and executable file: