Capstone Project Final Report

Title: Sentiment Analysis for User Ratings in Product Reviews

Introduction:

In today's digital age, understanding customer sentiments plays a crucial role in business success. This capstone project aims to perform sentiment analysis on user ratings within product reviews to provide valuable insights for businesses.

Problem Statement:

The problem revolves around deciphering the sentiments expressed in user reviews and predicting their corresponding ratings. A five-point rating scale (1-5) is used to evaluate the sentiment associated with each review.

Approach:

Data Collection:

- Gathered a dataset containing user reviews and ratings.
- Explored the data structure and identified potential challenges.

Data Preprocessing:

- Handled missing values and outliers.
- Applied techniques like tokenization and stemming to process textual data.

Feature Engineering:

- Utilized VADER sentiment analysis to create a new feature representing the compound sentiment score.
- Employed TF-IDF vectorization to convert textual data into numerical format for modeling.

Modeling:

- Employed three different models: Logistic Regression, Random Forest, and Support Vector Machine (SVM).
- Applied Grid Search to fine-tune hyperparameters for Random Forest and SVM models.

Evaluation:

- Assessed model performance using accuracy, precision, recall, and confusion matrices.
- Executed Cross-Validation to ensure models generalize well.

Findings:

Logistic Regression:

- Achieved an accuracy of 95%.
- Provided detailed classification metrics and confusion matrix.

Random Forest:

- Demonstrated a high accuracy of 97%.
- Generated comprehensive classification metrics and confusion matrix.

Support Vector Machine (SVM):

- Surprisingly achieved near-perfect accuracy of 99.54%.
- Delivered precise classification metrics and confusion matrix.

Recommendations:

Enhanced Sentiment Analysis:

- Implement advanced sentiment analysis techniques like BERT or GPT models for a more nuanced understanding of user sentiments.

Regular Data Updates:

- Regularly update the dataset to adapt to changing user sentiments and language patterns.

Customer Feedback Loop:

- Establish a customer feedback loop by integrating the sentiment analysis model into the product's review platform. Use customer feedback for continuous improvement.

Conclusion:

This project successfully tackled the challenge of sentiment analysis in user ratings. The models demonstrated high accuracy, especially the SVM model, which achieved remarkable results. The recommendations provide avenues for further improvement and application of the sentiment analysis model in real-world scenarios.