**Project Title**

Analysing Customer Complaints to Improve Service Quality: An AI/ML Approach

**Author**: Varsha Srivastava

**Executive summary**

This project focuses on analysing customer complaints to identify patterns and trends using natural language processing (NLP) and machine learning techniques. By understanding the underlying reasons for customer dissatisfaction, businesses can implement targeted improvements to enhance customer satisfaction and loyalty.

**Rationale**

Understanding and addressing customer complaints is crucial for businesses to retain customers and improve overall service quality. Analysing complaints helps identify recurring issues and pain points, enabling proactive measures to resolve them and prevent customer churn.

**Research Question**

Can we effectively categorize and analyse customer complaints to uncover insights that can drive improvements in service quality?

**Data Sources**

The project utilizes a dataset containing customer complaints from various sources, including:

* Complaint narratives
* Product or service categories

Link to dataset on Kaggle

**Methodology**

1. **Data Cleaning and Preprocessing**: Preprocessed the complaint data by cleaning text, removing stop words, and tokenizing text for analysis.
2. **Exploratory Data Analysis (EDA)**: Analyzed complaint frequencies, distribution across categories, and identified common themes or keywords.
3. **Text Vectorization**: Used techniques like TF-IDF (Term Frequency-Inverse Document Frequency) to convert text data into numerical vectors suitable for machine learning models.
4. **Model Building**: Employed classification algorithms such as MultinomialNB to categorize complaints into predefined categories based on their textual content.
5. **Model Evaluation**: Evaluated model performance using metrics such as accuracy, precision, recall, and F1-score. Visualized results using confusion matrices and classification reports.

**Results**

1. Model Accuracy Before and After Tuning:
   * Before hyperparameter tuning: The model accuracy was 79.00%. This means that, initially, the model correctly predicted the outcome for 79.00% of the samples in the test set.
   * After hyperparameter tuning: The model accuracy improved to 86.00%. This indicates that after adjusting the hyperparameters (parameters that are set before the learning process begins), the model's performance improved, with correct predictions increasing to 86.00% of the test set samples.
2. Error Analysis:
   * Before tuning: There were 1034 misclassified samples. This tells us that before adjusting the hyperparameters, the model made incorrect predictions for 1034 samples in the test set.
   * After tuning: The number of misclassified samples decreased to 712. This indicates that after hyperparameter tuning, the model's ability to correctly classify samples improved, resulting in fewer errors.
3. Recommendations:
   * Use Bag of Words (BOW) with the best hyperparameters for optimal performance: This suggests that using the Bag of Words (BOW) approach, along with the best-performing hyperparameters identified through tuning, is recommended to achieve the highest possible accuracy and minimize misclassifications.

Interpretation:

* Accuracy Improvement: Hyperparameter tuning significantly improved the model's accuracy from 79% to 86%. This is a substantial improvement, indicating that the chosen hyperparameters better suited the data and problem at hand.
* Error Reduction: The reduction in misclassified samples (from 1034 to 712) further underscores the effectiveness of hyperparameter tuning in enhancing the model's predictive capabilities.
* Actionable Recommendation: Using Bag of Words (BOW) along with optimized hyperparameters is advised based on the analysis. This recommendation aims to maintain or further improve the model's performance in future applications or deployments.

**Next steps**

1. Enhance model performance by fine-tuning hyperparameters and exploring advanced NLP techniques.
2. Implement sentiment analysis to understand the emotional context of complaints.
3. Integrate real-time data streams to enable proactive complaint handling and resolution.
4. Collaborate with customer service teams to implement actionable insights and monitor impact on service quality.

**Outline of project**

* Link to notebook 1 - Data Cleaning and EDA
* Link to notebook 2 - Text Vectorization and Model Building

**Contact and Further Information**

For additional details or inquiries, please contact [srivastavavarsha1811@gmail.com](mailto:srivastavavarsha1811@gmail.com)

Further information can be found at [VarshaSrivastava1811/Capstone-Project (github.com)](https://github.com/VarshaSrivastava1811/Capstone-Project)