**Capstone Three: Project Proposal**

# **Automated Customer Complaint Classification**

Customer complaints are very important for financial companies because they can reveal the shortcomings in their products and services. Resolving these complaints efficiently and in a timely manner can minimize customer dissatisfaction and increase customer loyalty. Moreover, it helps companies to continuously improve their services and attract more customers

Customer complaints are usually unstructured text data, which traditionally require companies to assign multiple support employees to evaluate and allocate each ticket to the relevant department. As the company grows and acquires a larger customer base, this process becomes tedious. Therefore, the company aims to automate its customer support ticket system. Being a financial company, the firm offers numerous products and services, including credit cards, banking, and mortgages/loans.

**Problem statement formation**

The task is to build a model that can efficiently classify customer complaints based on the products and services offered by the company. These complaints are unstructured text data, and the company wants to automate the process of evaluating and assigning each ticket to the relevant department.

* Ticket Classification: Develop a model that can accurately categorize customer complaints into one of the following five clusters:
  + **Credit card / Prepaid card**
  + **Bank account services**
  + **Theft/Dispute reporting**
  + **Mortgages/loans**
  + **Others**
* Quick Resolution: By segregating the clusters, the company aims to expedite the resolution process for customer issues. Efficiently handling complaints can lead to higher customer satisfaction and stronger loyalty.
* Unlabeled Data: The provided data is in JSON format and lacks accurate labels. We need to apply unsupervised techniques to analyze patterns and create meaningful clusters.

The solution should enable the company to streamline its customer support ticket system, improve service quality, and enhance customer experience.

**Criteria for success**

1. **High Accuracy**: The model should achieve high accuracy in classifying customer complaints into predefined clusters (credit card, bank account services, theft/dispute reporting, mortgages/loans, and others). Aim for an accuracy that minimizes misclassifications.
2. **Balanced Precision and Recall**: Strive for a balance between precision (correctly identifying relevant complaints) and recall (capturing all relevant complaints). High precision ensures accurate categorization, while high recall minimizes missed complaints.
3. **Robustness to Noise**: Account for noisy or misspelled text data in the complaints. The model should handle variations in language and formatting.
4. **Generalization**: The trained model should generalize well to new, unseen complaints. Validate its performance on a holdout dataset or through cross-validation.
5. **Scalability**: As the company’s customer base grows, ensure that the system scales efficiently. Consider the computational resources required for real-time processing.
6. **User Satisfaction**: Ultimately, success lies in improved customer satisfaction. Monitor user feedback and assess whether the system positively impacts customer support efficiency

**Scope of solution space**

To effectively analyze customer complaints, we recommend following these constructive steps:

* Gather unstructured data from various sources, then preprocess the text data by removing noise, tokenizing, and stemming.
* Extract relevant features from the complaint text and cluster similar complaints based on extracted features. After that, choose appropriate machine learning models for classification and evaluate model performance using various metrics. It's important to handle missing data and deploy the trained model in a production environment.
* We suggest integrating the system with the existing customer support ticketing system, monitoring model performance over time, and collecting user feedback to update the model periodically. Finally, visualizing topic clusters and trends can provide valuable insights for better decision-making.
* Evaluate model performance using metrics like accuracy, precision, recall, and F1-score.
* Visualize topic clusters and trends for better decision-making.

**Constraints**

* The text data in the complaints may be noisy and contain misspellings, symbols, or special characters. We’ll need to perform thorough text preprocessing, including tokenization, stemming, and removing stop words.
* Distribution of complaints across different product categories may be imbalanced. Some categories might have significantly more or fewer examples than others. Addressing class imbalance is crucial to ensure fair model performance.
* The model should generalize well to new, unseen complaints. Consider using techniques like cross-validation and hyperparameter tuning to optimize model performance.
* As the company grows and the number of customer complaints increases, the system should be scalable. Ensure that the chosen algorithms can handle large volumes of data efficiently.
* Consider computational resources, memory, and processing time when selecting algorithms. Some topic modeling techniques can be resource-intensive.

**Data sources**

* [Dataset.csv](https://www.kaggle.com/datasets/venkatasubramanian/automatic-ticket-classification)