PROPOSAL

Bank Customer Complaints Classification using Natural Language Processing (NLP)

Business Understanding:

Problem Statement

Customers of financial institutions often experience significant frustration and dissatisfaction when lodging complaints related to financial services. The existing complaint submission processes typically involve navigating through multiple selection options or answering a series of questions generated by chatbots. This complexity not only prolongs the time required to submit a complaint but also adds to customer frustration, especially when timely resolution is critical. There is a pressing need for a faster and more efficient method for customers to lodge complaints without the burden of excessive questioning.

Root Causes

- **Complex Navigation:** Complaint submission systems require customers to navigate through multiple menus, making the process time-consuming.
- **Inefficient Chatbot Interactions:** Chatbots often ask redundant or irrelevant questions, prolonging the interaction and causing frustration.
- **Lack of Personalization:** The absence of complaint-specific pathways leads to a generic experience that doesn't meet diverse customer needs.
- **Underutilized Data:** Systems do not fully leverage customer data, resulting in repetitive information requests that could otherwise be streamlined.

Key Stakeholders

1. **Customers:** Seek a quick, intuitive way to submit complaints and receive prompt resolutions. They prefer a minimal-step interface that reduces complexity and reassures them of efficient issue handling.
2. **Customer Service Support Teams:** Aim to manage and categorize complaints efficiently for faster resolutions. They need streamlined categorization processes, access to accurate data, and tools for tracking complaint statuses to maintain high service quality.

Proposed Solution

This project proposes leveraging Natural Language Processing (NLP) to create an automated system for classifying customer complaints. By training an NLP model to interpret complaint content and integrate it in an appropriate, simple application interface,

the system can streamline the submission process, reducing the number of questions and options customers must navigate. This approach aims to minimize submission time and enhance the customer experience by providing an easy-to-use platform for logging complaints.

Objectives

1. **Model Training:** Develop and train an NLP model that can automatically categorize consumer complaints into predefined classes based on their textual content.
2. **Improving User Experience:** Design a streamlined submission interface that reduces the need for lengthy navigation, allowing customers to submit complaints quickly and efficiently.
3. **Enhancing Responsiveness:** Enable financial institutions to address complaints faster by categorizing them accurately and routing them to the appropriate support teams.

Data Understanding:

We plan to use the publicly available Consumer Complaints Dataset from the Consumer Financial Protection Bureau (CFPB), sourced from Kaggle, containing approximately 162,400 records of complaint narratives on various financial products. The dataset includes detailed complaint narratives across five categories: Credit Reporting; Data Collection; Mortgages and Loans; Credit Cards and Retail Banking.

This dataset provides a strong foundation for training an NLP model to accurately categorize complaints.

Data Preparation Approach

The project will start with data exploration to check missing values, duplicates, class imbalances etc. to inform data preprocessing, transformation and modeling steps. A thorough preprocessing pipeline will be established, involving text normalization steps such as lowercasing, special character removal, tokenization, stop word filtering, and lemmatization. After cleaning, the data will be vectorized using techniques like TF-IDF and scaled with MinMax Scaling.

Model Training, Evaluation, and Selection Approach:

We then plan to explore multiple machine learning algorithms for initial model development, including Multinomial Naive Bayes, Support Vector Machine (SVM), Logistic Regression, and Random Forest. Following initial evaluations, we will assess the potential of an ExtraTrees ensemble model and a BERT-based transformer model, given BERT's effectiveness in NLP tasks. Model evaluation will focus on accuracy, precision, recall, and F1-score, with emphasis on Macro F1 and Weighted F1 to ensure balanced classification across complaint categories and account for class imbalances.

Deployment and Application Approach:

Upon selecting the best-performing model, we will deploy an application with a simple user interface built with Streamlit. This interface will allow customers to enter their complaint details along with identifying information. The system will use the best performing model to categorize complaints automatically, and through an integrated SMS notification system, the complaint will be routed to the designated support team and a copy sent to the customer for their records. The application will be deployed on Hugging Face, allowing bank customers to access the interface via a direct link and minimizing the need for complex navigation.

Expected Business Impact:

By leveraging NLP, the solution aims to simplify the complaint submission process, enhance customer satisfaction, and improve the bank's responsiveness to complaints. The proposed solution promises a robust system for automated complaint management, with the flexibility to evolve based on real-time customer needs and institutional requirements.