

B5W6: Intelligent Complaint Analysis for Financial Services

Interim Report

1. Project Understanding

1.1 Business Objective

The primary objective of this project is to develop an intelligent system for analyzing customer complaints in the financial services sector. The system aims to: - Automate the categorization of financial complaints - Identify emerging issues and trends - Improve response times and resolution rates - Enhance customer satisfaction and regulatory compliance

1.2 Key Components

- 1. Data Collection & Processing**
 - Gathering complaint data from multiple sources
 - Cleaning and normalizing text data
 - Handling missing values and inconsistencies
- 2. Exploratory Data Analysis**
 - Analyzing complaint patterns and trends
 - Identifying most common issues by product category
 - Temporal analysis of complaint volumes
- 3. Machine Learning Pipeline**
 - Text preprocessing and feature extraction
 - Model development for complaint classification
 - Sentiment analysis of complaint narratives
- 4. Visualization & Reporting**
 - Interactive dashboards for monitoring
 - Automated report generation
 - Alert system for emerging issues

2. Project Progression

2.1 Completed Work

- Successfully acquired and loaded the CFPB complaint dataset (~9.6M records)
- Performed initial data exploration and quality assessment
- Implemented data cleaning and preprocessing pipeline
 - Handled missing values in key columns
 - Processed text data (narratives)
 - Filtered and prepared dataset for analysis
- Developed initial visualizations for:
 - Missing value patterns
 - Complaint distribution by product category
 - Temporal trends in complaints

2.2 Current Challenges

- 1. Data Quality Issues**
 - High percentage of missing values in key columns

- Inconsistent product categorization
- Noisy text data in complaint narratives

2. Technical Challenges

- Large dataset size requiring efficient processing
- Class imbalance in product categories
- Need for advanced NLP techniques for text analysis

2.3 Strategy Moving Forward

1. Immediate Next Steps

- Complete comprehensive EDA
- Implement text preprocessing pipeline
- Develop baseline classification models

2. Risk Mitigation

- Implement robust error handling for data processing
- Create data validation checks
- Set up version control for models

3. Areas of Focus

- Feature engineering for text data
- Model selection and evaluation
- Performance optimization for large-scale processing

2.4 Timeline

- **Week 1-2:** Data preparation and EDA
- **Week 3-4:** Feature engineering and model development
- **Week 5:** Model evaluation and refinement
- **Week 6:** Dashboard development and final reporting

3. Preliminary Findings

3.1 Data Overview

- Total complaints: 9,609,797
- Key product categories:
 - Credit reporting: 72.8%
 - Debt collection: 8.3%
 - Mortgages: 4.4%
 - Checking/savings: 3.0%

3.2 Data Quality

- 69% of complaints lack narrative text
- 93% missing values in 'Tags' column
- 92% missing values in 'Consumer disputed?' field

3.3 Initial Insights

- Clear seasonality in complaint volumes
- Significant variation in complaint types by product category
- Opportunities for automated categorization of common issues

4. Conclusion

The project is progressing according to plan, with the foundation for data processing and analysis now in place. The next phase will focus on developing and refining the machine learning models for complaint analysis. The insights gained will help financial institutions improve their products and services while enhancing customer satisfaction.