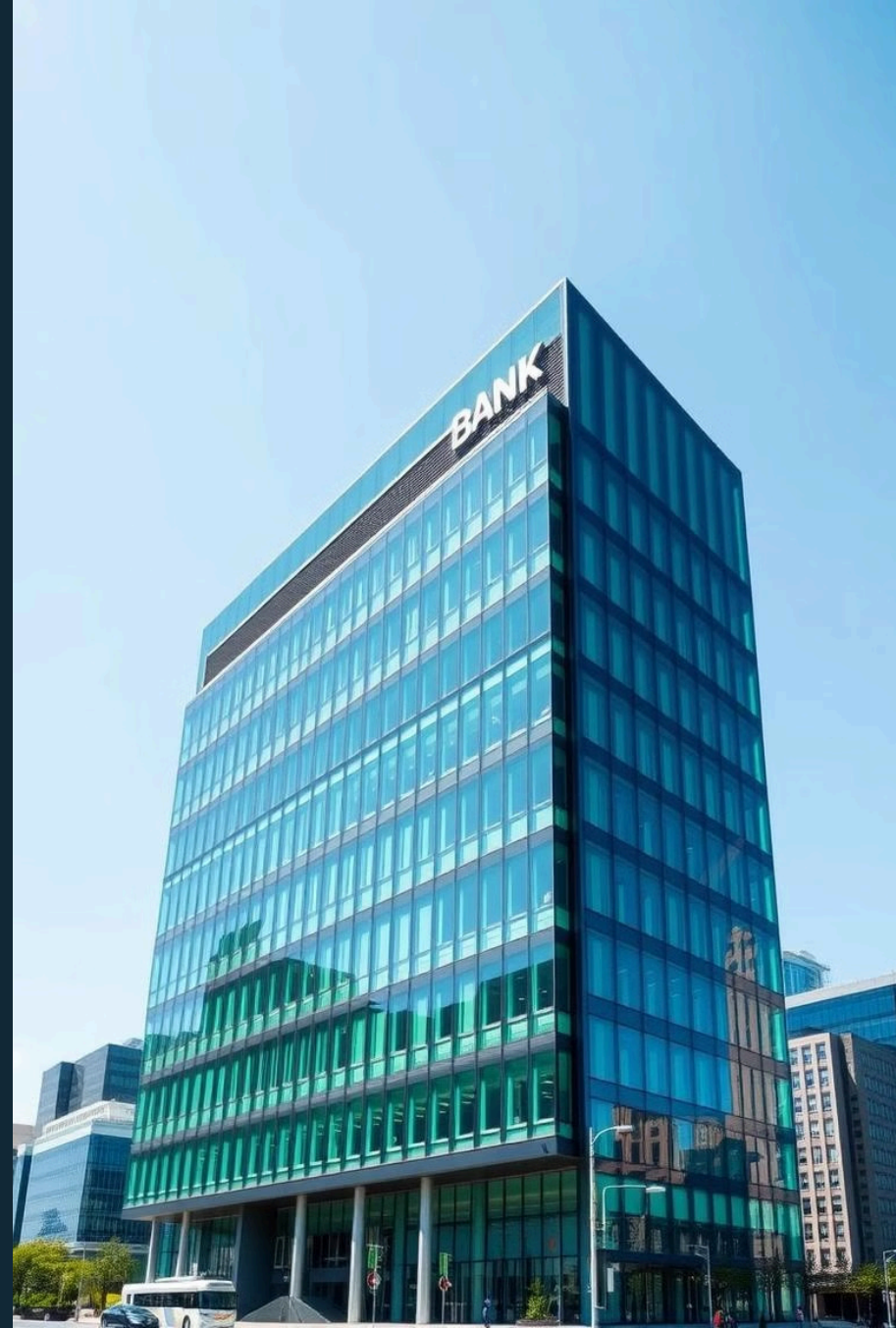


Bank Customer Complaint Classification

A Natural Language Processing (NLP) Project

November 2024

Authors: Mary Musyoka, Julian Kilyungi, Tabitha Kariuki, John Kul, Norah Oluoch and Joseph Ngige (Group 6)



Outline

1

Project Overview

2

Business
Understanding

3

Data Understanding &
Preparation

4

Modeling & Evaluation

5

Deployment &
Application

6

Conclusion &
Recommendations

Business Understanding

Problem Statement

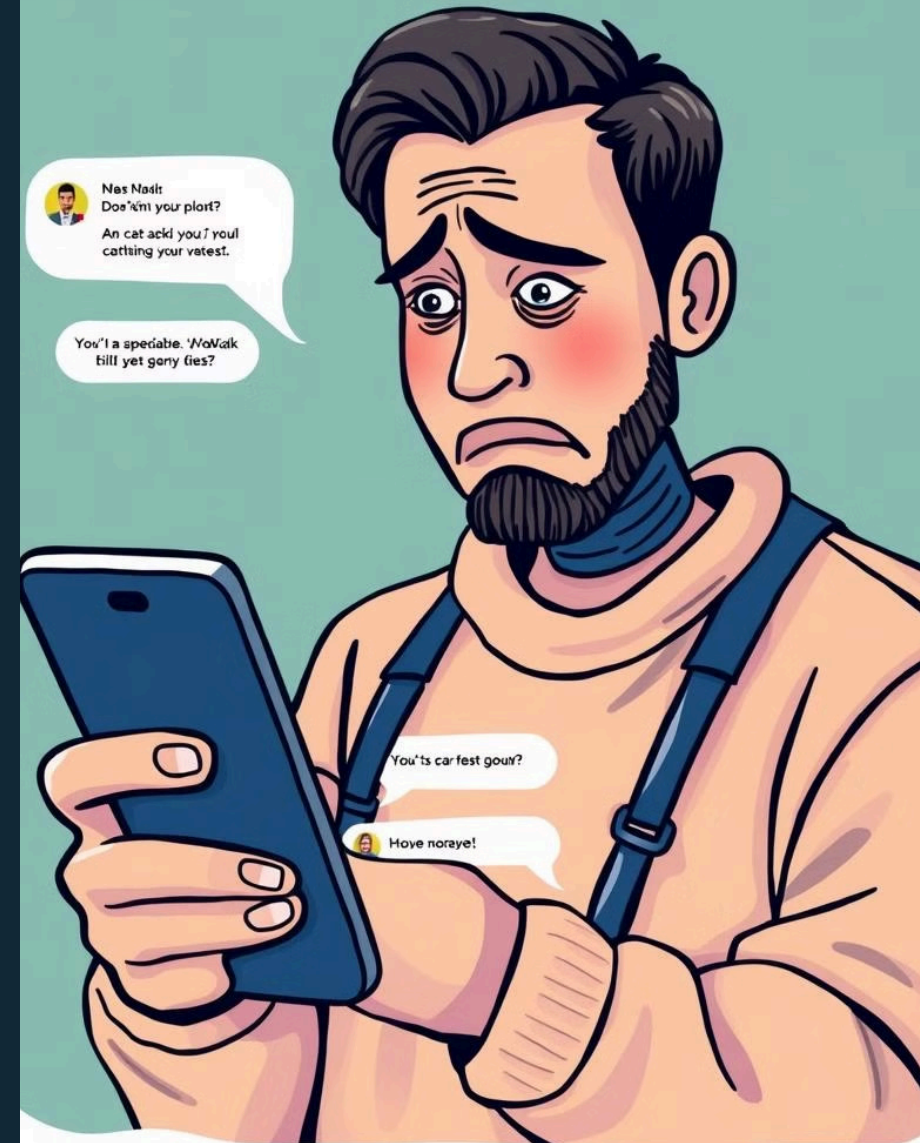
Customers face frustration due to complex complaint submission processes.

Root Causes

Complex navigation, inefficient chatbots, lack of personalization and limited data use.

Key Stakeholders

1. Customers and,
2. Customer support teams.



Data Understanding & Preparation

Dataset Source

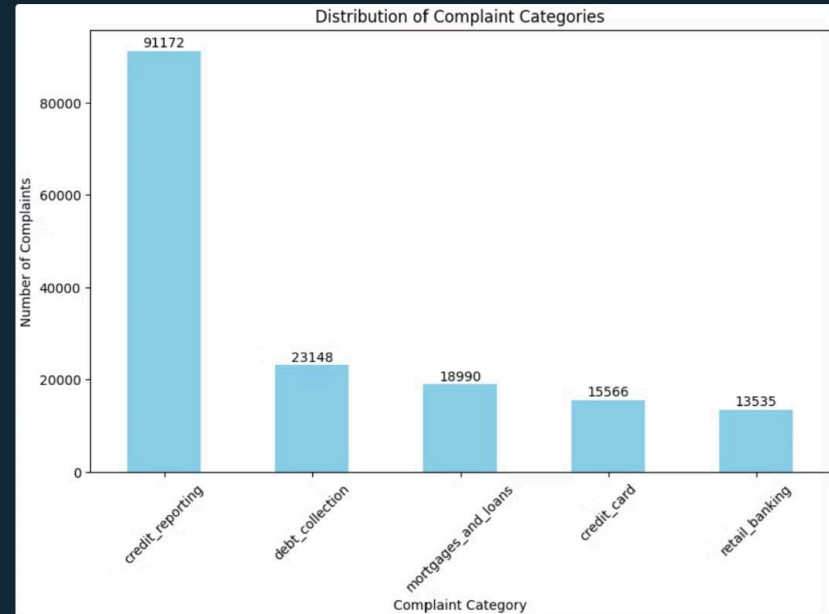
US Consumer Complaints
Dataset from [Kaggle](#).

Complaint Categories

Five (5): Credit reporting;
debt collection;
mortgages, credit cards;
and retail banking.

Key Details

162,400 records, imbalanced data, and varying narrative lengths.



The distribution of Complaint Categories shows a class imbalance, with the credit_reporting category significantly more represented than others.

Data Understanding & Preparation

Exploration

Handled missing values,
retained duplicates,
addressed class
imbalance.

Pre-processing

Standardized text,
tokenization, stop word
removal, and
lemmatization.

Data Transformation

Applied TF-IDF and MinMax scaling. SMOTE also applied later for model improvement.



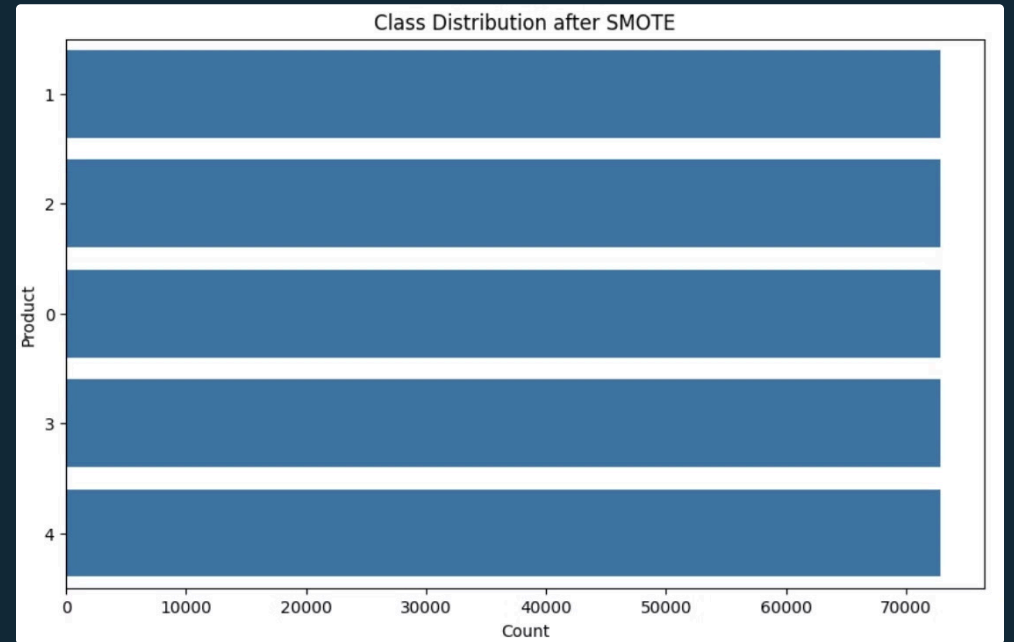
Word cloud highlighting specific terms associated with common issues for each product. Shows that customer complaints are centered around distinct themes based on the product type.

Data Understanding & Preparation

SMOTE Class Balancing

Synthetic Minority Oversampling Technique (SMOTE)

- Perfectly balanced class distribution after applying SMOTE.
- Each product category having an equal count of samples (approximately 73,000).
- This balanced distribution helps the model learn patterns more evenly across classes.



Modeling & Evaluation



Models

Baseline: MNB, SVM, Logistic Regression, Random Forest.

Improved: Tuned Random Forest, ExtraTrees, BERT.



Evaluation Metrics

Classification report, Confusion Matix, AUC-ROC curve with focus on **Macro F1 Score** supported by **Weighted F1-score**.



Model Selection

BERT seleted for its balanced performance and suitability for NLP tasks.



Modeling & Evaluation

BERT Model Performance

Metrics Report:

- Solid performance.
- The **Macro F1 of 0.85** shows consistent results across classes.
- **Weighted F1 0.89** and **Accuracy of 89%**.

Classification Report:

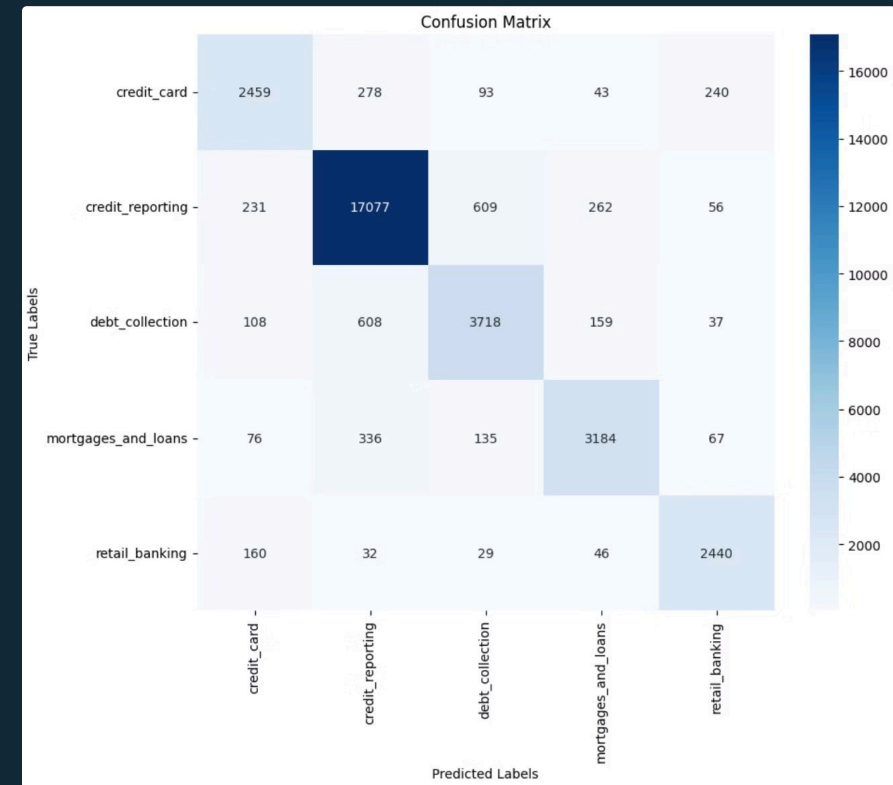
	precision	recall	f1-score	support
credit_card	0.81	0.79	0.80	3113
credit_reporting	0.93	0.94	0.93	18235
debt_collection	0.81	0.80	0.81	4630
mortgages_and_loans	0.86	0.84	0.85	3798
retail_banking	0.86	0.90	0.88	2707
accuracy			0.89	32483
macro avg	0.85	0.85	0.85	32483
weighted avg	0.89	0.89	0.89	32483

Modeling & Evaluation

BERT Model Performance

Confusion Matrix:

- Indicates strong classification performance.
- Some misclassification occurred between similar financial categories.
- Shows good overall differentiation.

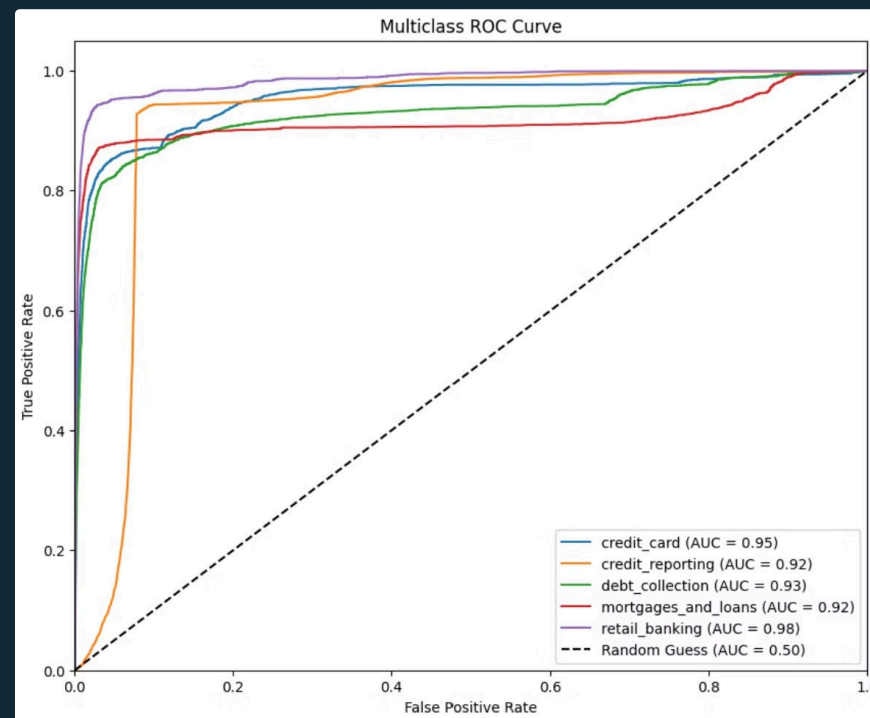


Modeling & Evaluation

BERT Model Performance

AUC-ROC Curve:

- Demonstrates high discriminative power across all complaint categories.
- AUC values ranging from 0.92 to 0.98.
- Overall reflects a strong ability of the model to distinguish between classes.



Deployment & Application

User-Friendly Interface

Built with Streamlit & deployed on [huggingface](https://huggingface.co).

Model Integration

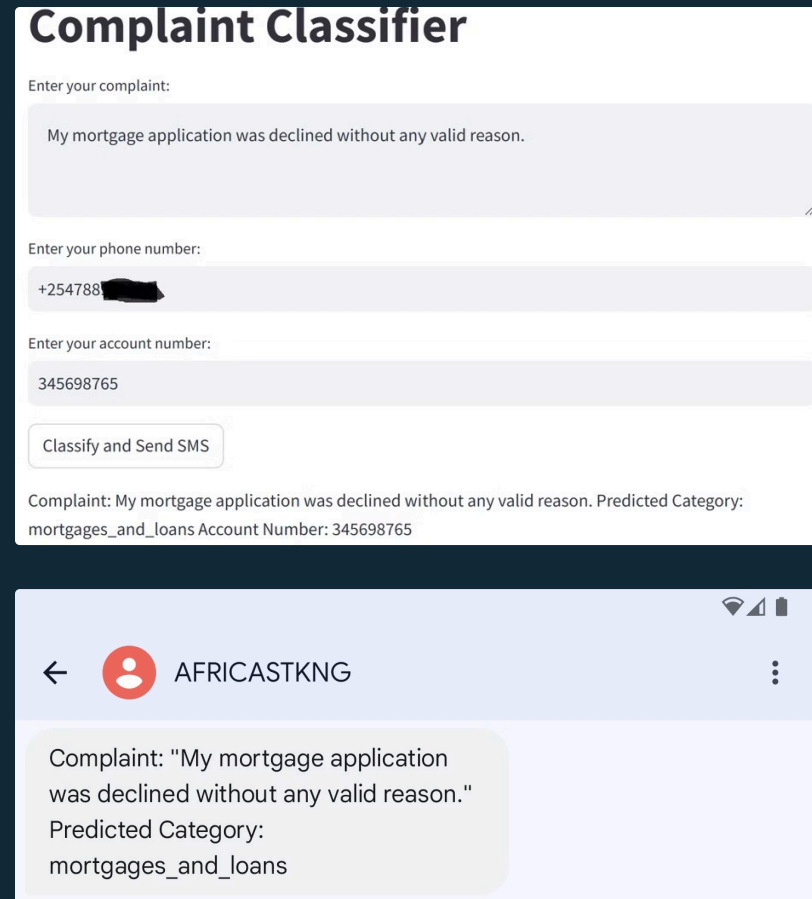
BERT classifies complaints in real-time.

Automated Notifications

Integrated with Africastalking's SMS API, sending classified notification.

Challenges

SMS limited to Airtel and Telkom; Email notification slow.



The image shows two screenshots. The top screenshot is a web interface titled "Complaint Classifier". It has three input fields: "Enter your complaint:" with the text "My mortgage application was declined without any valid reason.", "Enter your phone number:" with the text "+254788[REDACTED]", and "Enter your account number:" with the text "345698765". Below these is a button labeled "Classify and Send SMS". At the bottom, it displays the output: "Complaint: My mortgage application was declined without any valid reason. Predicted Category: mortgages_and_loans Account Number: 345698765". The bottom screenshot is an SMS notification from "AFRICASTKNG". It contains the text: "Complaint: 'My mortgage application was declined without any valid reason.'" and "Predicted Category: mortgages_and_loans".

Snapshot of the system simplified user interface and SMS notification received

Conclusion & Recommendations

Insight & Impact

1. BERT model with robust classification accuracy trained.
2. Simple complaint submission system deployed.
3. Prompt notification for enhanced responsiveness.

Future Improvements

- Feedback loops.
- Category optimization.
- Expanded notification channels.





Group Members:

Mary Musyoka: mary.musyoka@student.moringaschool.com

Julian Kilyungi: julian.kilyungi@student.moringaschool.com

Tabitha Kariuki: tabitha.kariuki1@student.moringaschool.com

John Kul: john.kul@student.moringaschool.com

Norah Oluoch: norah.oluoch@student.moringaschool.com]

Joseph Ngige: joseph.ngige@student.moringaschool.com

Thank you.