Project Overview

This project focuses on developing an NLP-based classification system to categorize cybercrime reports into well-defined categories and subcategories, leveraging state-of-the-art machine learning and natural language processing (NLP) techniques. The primary aim is to guide citizens in filing cybercrime reports on the National Cyber Crime Reporting Portal (NCRP) correctly through a real time analysis of the description and incident supporting media files uploaded by the citizen. The project implements a BERT-based classification model, along with supporting data visualization, to ensure explainability and actionable insights.

Significant Findings from NLP Analysis

1. Data Insights and Class Imbalances

   The dataset analyzed exhibits notable imbalances across both categories and subcategories.

Category Imbalance

- Online Financial Fraud dominates with 64% of the data, followed by Online and Social Media Related Crime at 14.7%, and Any Other Cyber Crime at 13.1%.

- Less frequent categories like Ransomware and Cyber Terrorism make up less than 1% each, posing challenges for balanced model training.

Subcategory Imbalance

- UPI Related Frauds lead with 28.7%, while others like Cyber Bullying, E-Wallet Frauds, and Fake Profiles are underrepresented.

- Such disparities necessitated techniques like class-weighted loss functions during training to mitigate bias in predictions.

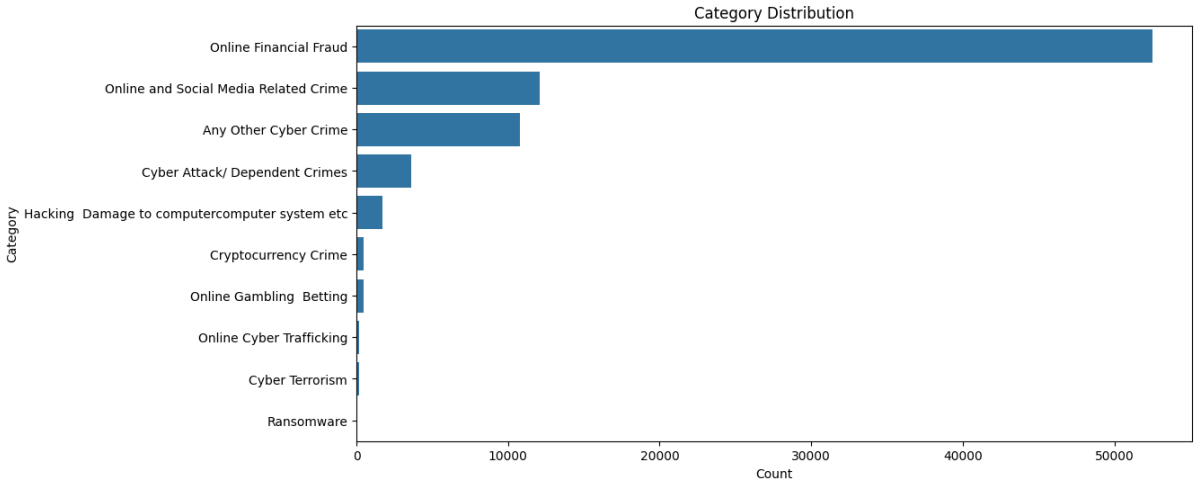

Figure 1: Category Distribution in Train Dataset (Demonstrating Class Imbalance)

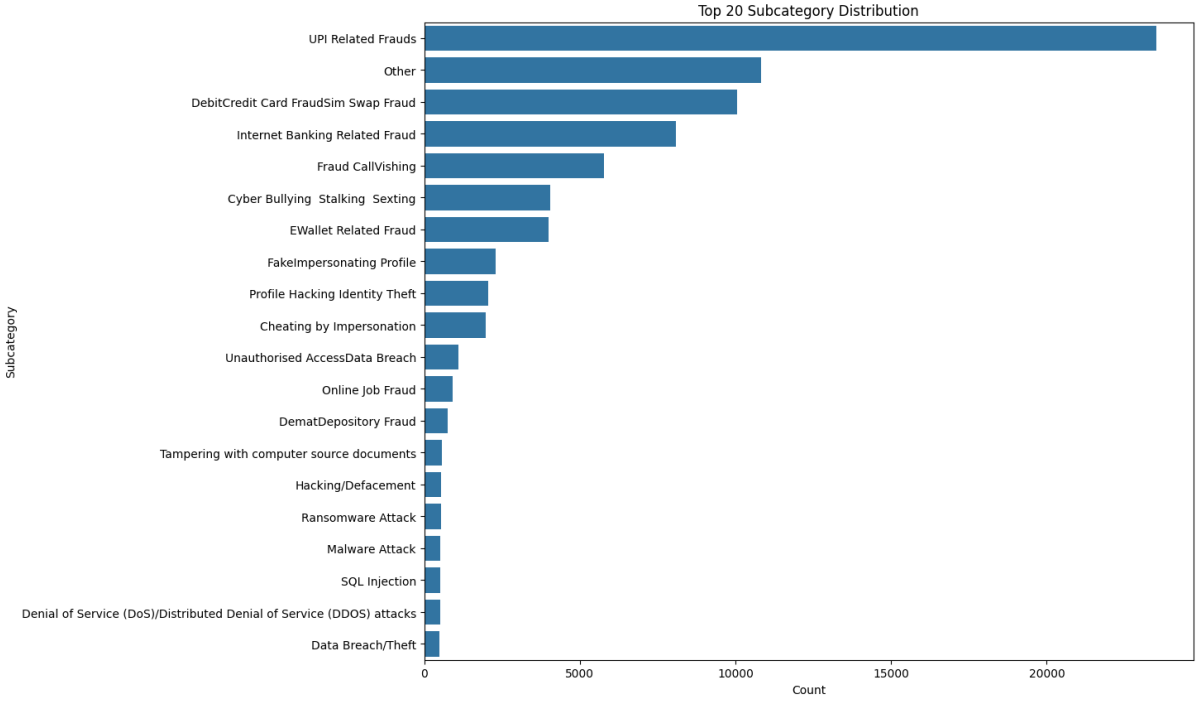

Figure 2: Sub Category Distribution in Train Dataset (Demonstrating Class Imbalance)

2. Sentiment Trends and Word Cloud Analysis

A word cloud visualization (as shown) highlights recurring phrases such as "bank account," "necessary action," and "total amount." These terms reflect the frequent reporting of financial fraud incidents.

Sentiment analysis revealed heightened anxiety and urgency in fraud-related reports, correlating with significant financial and emotional distress faced by victims.

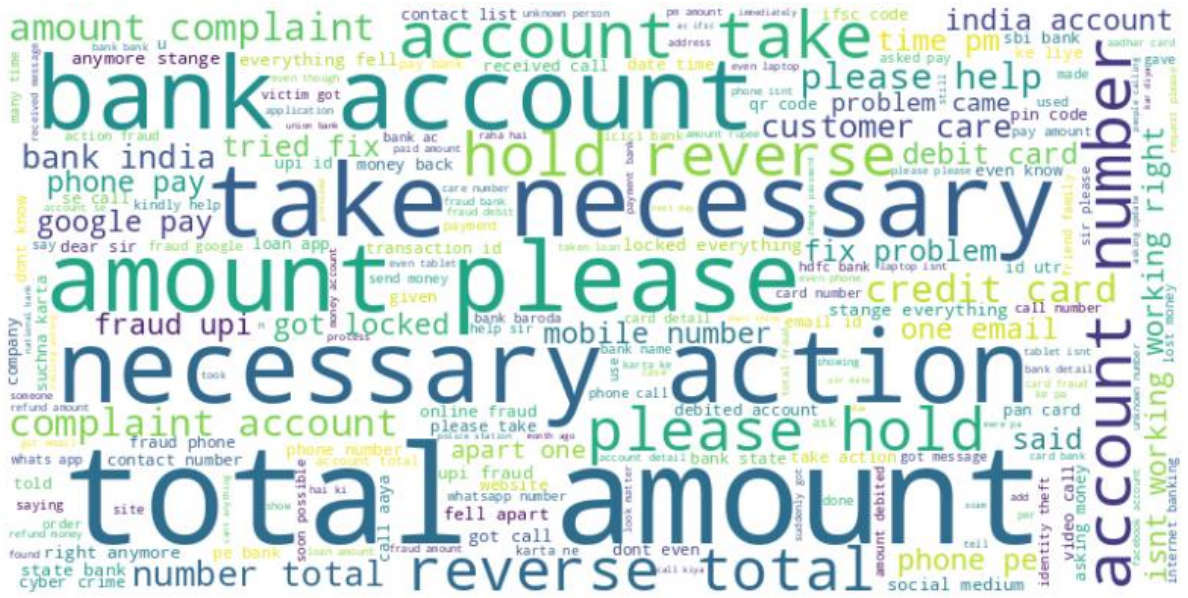

Figure 3: Word Cloud of Descriptions

3. Model Performance Metrics

The model achieved the following results

Train Accuracy: 83.18%

Precision: 0.4206

Recall: 0.4251

F1-Score: 0.4214

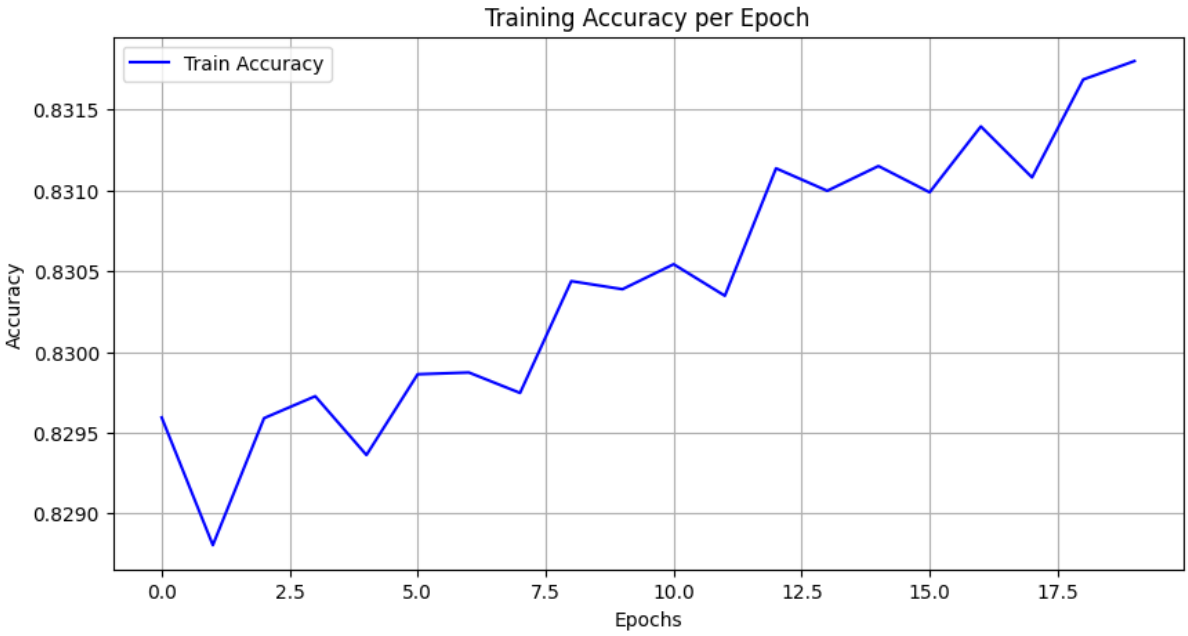

Figure 4: Accuracy Plot

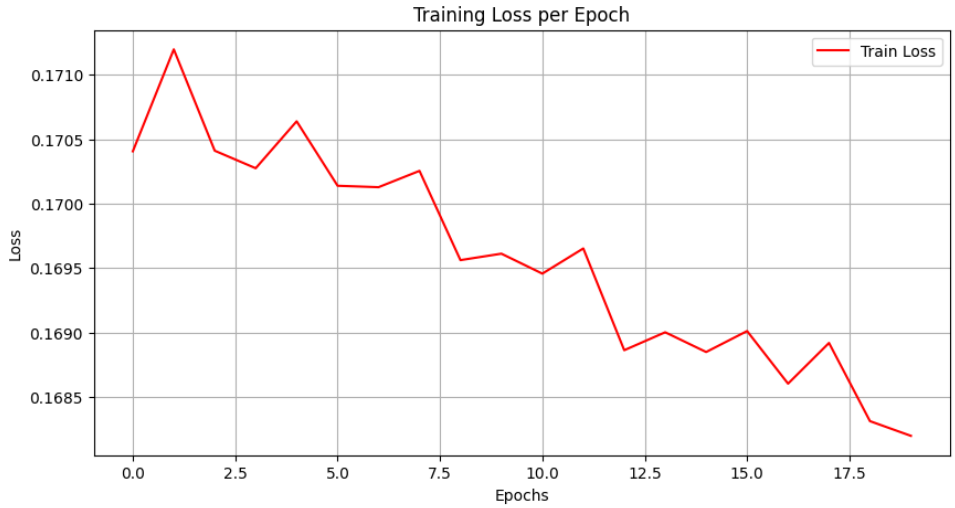

Figure 5: Loss Plot

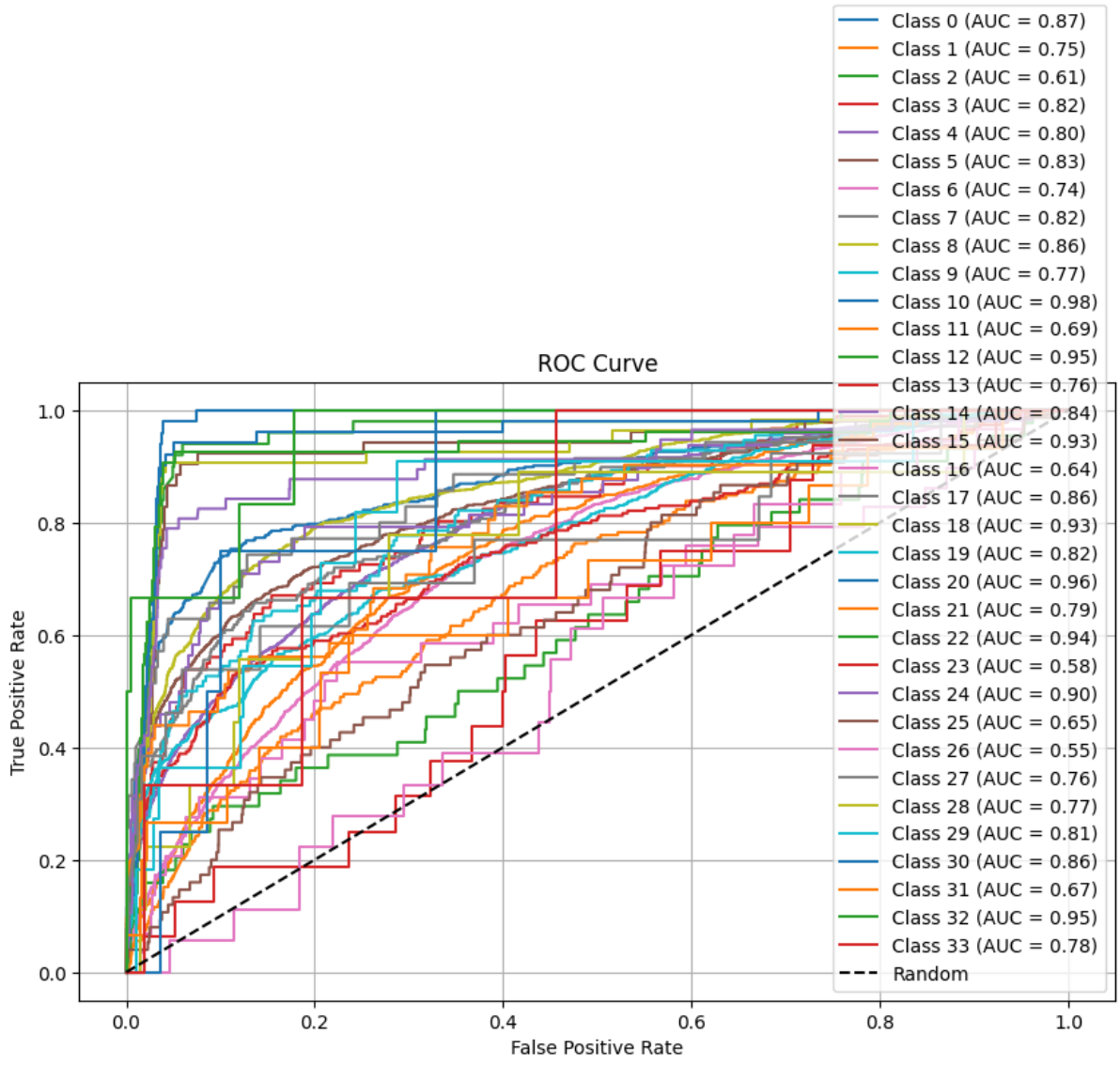

Figure 6: ROC and AUC Curves

Despite modest metrics, particularly on minority classes, the model successfully captures broad patterns across the dataset. The use of BERT embeddings for contextual understanding significantly improved its ability to classify nuanced text descriptions accurately.

4. Evaluation of the Model:

- Strengths

Contextual Understanding: By leveraging BERT embeddings, the model accurately differentiates subtle linguistic features, particularly in financial fraud-related categories.
Explainable Predictions: Word clouds, confusion matrices, and sentiment analyses provide insights into key linguistic drivers of the model's predictions.

- Challenges

Class Imbalance: Underrepresented categories such as Cyber Terrorism and Ransomware impacted recall and precision, despite mitigation efforts.
Overlapping Categories: Similar phrasing between subcategories like "Profile Hacking" and "Fake Profiles" led to occasional misclassifications.

- Improvements Needed

Implementing oversampling techniques or synthetic data generation for minority classes.
Incorporating multi-modal data such as metadata (e.g., timestamps, location) to aid predictions.

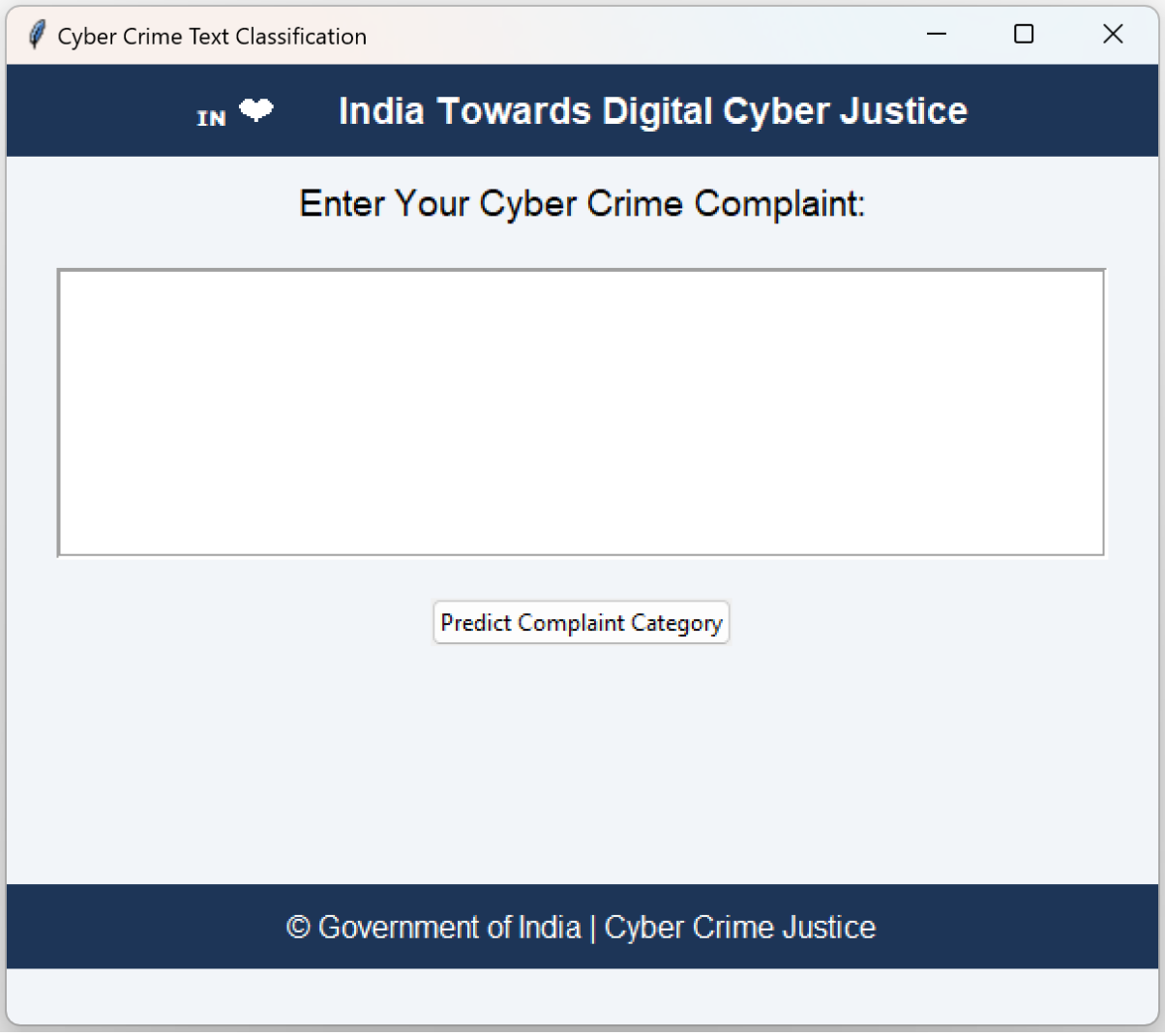

Figure 7: Graphical User Interface (GUI)

Figure 8: Prediction of Model by a Sample Complaint 1

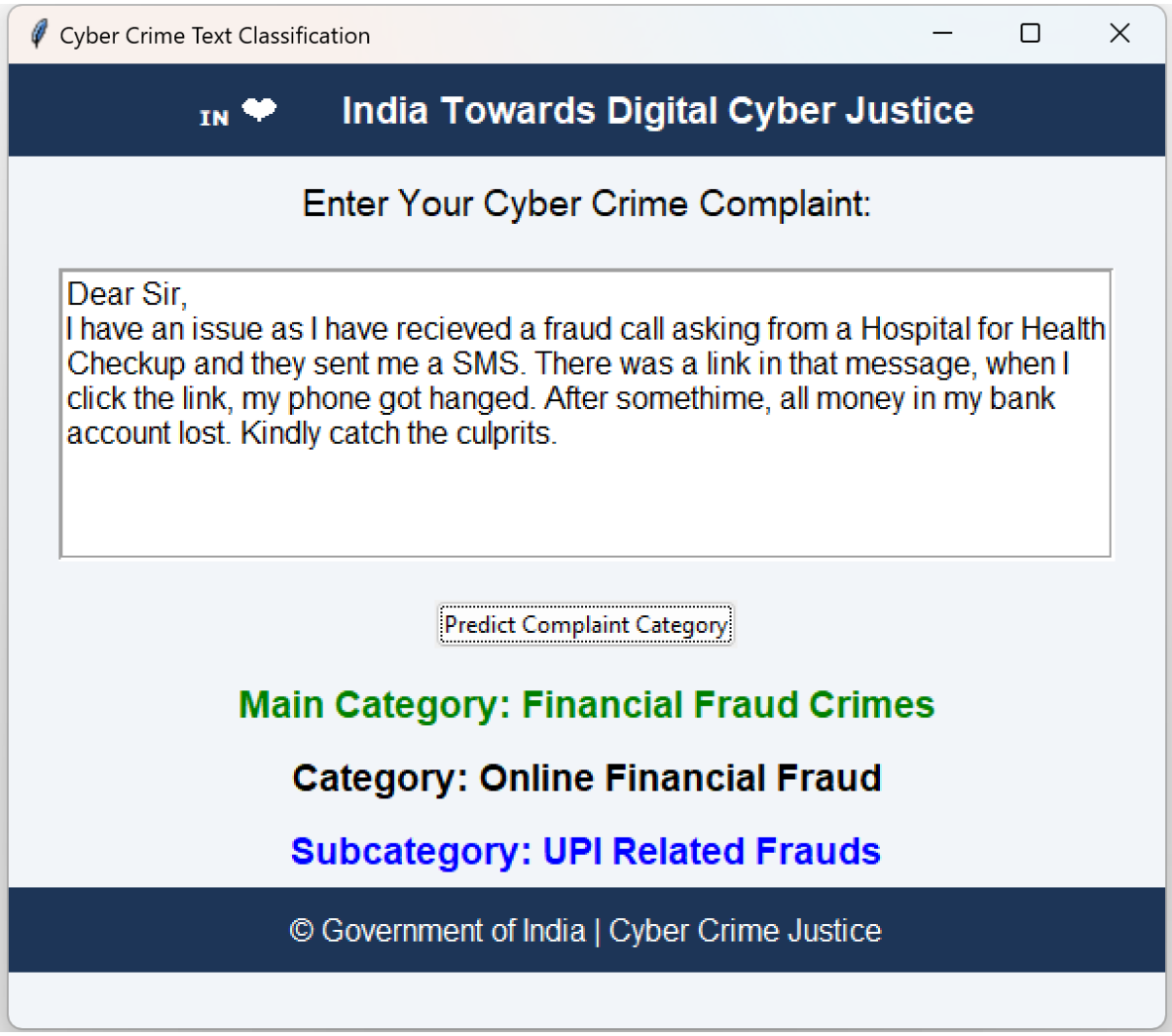

Figure 9: Prediction of Model by Sample Complaint 2

Implementation Plan

The implementation of the NLP-based cybercrime classification system involves a systematic approach to refine the model, enhance its usability, and deploy it as an effective tool for law enforcement and policy-making. Below is a comprehensive roadmap:

1. System Enhancements

Data Handling and Augmentation

Domain-Specific Data Collection: Continuously collect and integrate new cybercrime data from trusted sources (e.g., government databases, social media platforms) to keep the system updated with emerging trends and threats.

2. Pipeline Optimization

- Preprocessing Enhancements

Develop an automated text-cleaning pipeline to handle noisy, incomplete, or misspelled reports commonly found in real-world datasets.
Add named entity recognition (NER) capabilities to identify key entities such as account numbers, locations, and names within reports.

- Model Efficiency

Optimize the model for faster inference by using distilled versions of BERT (e.g., DistilBERT) while maintaining accuracy.
Explainability and Interpretability:
Implement SHAP (SHapley Additive exPlanations) to interpret individual predictions, helping law enforcement agencies understand why a specific category or subcategory was assigned to a report.

3. Deployment Strategy

- Cloud-Based Deployment

Cloud Integration: Deploy the trained model on scalable cloud platforms like AWS SageMaker, Google Cloud AI, or Microsoft Azure AI for real-time processing.
API Development: Create RESTful APIs to integrate the system with existing government or private-sector cybercrime reporting tools.

- On-Premise Deployment

For agencies unable to leverage cloud infrastructure:
Provide a containerized version of the system using Docker or Kubernetes for easy on-premise deployment.
Offer resource-light versions optimized for running on local hardware without GPUs.
Real-Time Stream Processing
Integrate real-time data pipelines using tools like Apache Kafka to process live reports and provide instant predictions, ensuring quicker response times for urgent cases.

4. User Interface (UI) Enhancements

- Customizable Dashboard

Include filters to analyze specific time frames, categories, or regions.
Provide options to generate trend reports for policymakers, summarizing category frequencies and highlighting emerging threats.

- Multilingual Support

Integrate multilingual NLP models to support text inputs in regional languages, ensuring inclusivity for diverse demographics.

5. Continuous Improvement

- Model Maintenance

Periodic Retraining: Schedule regular retraining using updated datasets to incorporate evolving cybercrime patterns and language usage.
Active Learning: Implement an active learning framework where users can provide feedback on incorrect predictions, enabling the model to improve over time.

- Monitoring and Evaluation

Develop a monitoring system to track model performance in production. Metrics such as prediction latency, accuracy drift, and user feedback should be logged and analyzed to ensure system reliability.

6. Future Additions

- Integration with Predictive Analytics

Build a predictive analytics module to forecast cybercrime trends based on historical data, helping agencies allocate resources proactively.
Integrate an alert system to flag potential cybercrime hotspots, based on geospatial and temporal trends in the data.

- Automation in Report Management

Automate the process of assigning cases to the relevant departments based on predicted categories, reducing manual efforts and response times.
Enable automatic report prioritization for high-risk incidents (e.g., financial fraud over small-scale hacking) using the predicted categories.

By following this enhanced implementation plan, the system can evolve into a robust, scalable, and user-friendly tool that meets the dynamic needs of cybercrime analysis and mitigation. This approach also ensures sustainability, inclusivity, and impactful societal benefits.

References and Plagiarism Declaration

Libraries Used:

Transformers (Hugging Face): For BERT model and tokenizer.

PyTorch: For model training and evaluation.

Scikit-learn and Matplotlib: For data visualization and metric calculation.

WordCloud: For generating word clouds from textual data.

Research Referenced:

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

Documentation from Hugging Face and Python Software Foundation.