Project Report: NLP-Based Cybercrime Classification System

---

1. Significant Findings from NLP Analysis

In this project, the primary goal was to classify cybercrime reports using Natural Language Processing (NLP) techniques. The following findings were observed from the analysis:

1.1 Sentiment Trends Over Time

While the dataset primarily focused on categorizing crimes, sentiment analysis revealed several interesting trends. The sentiment of cybercrime reports varied across different types of crimes. For instance, reports related to **Online Financial Fraud** and **Cryptocurrency Crime** often exhibited negative sentiments, with victims expressing frustration, confusion, and financial losses. On the other hand, reports in categories like **Cyber Attack/Dependent Crimes** tended to be more neutral, as they focused on detailing the technical aspects of the attacks. However, **Rape/Gang Rape** and **Sexually Abusive Content** reports had a higher degree of negative sentiment, indicating distress and emotional impact.

1.2 Commonly Recurring Themes/Topics

A frequent theme in the dataset was **fraud**. Categories like **Online Financial Fraud** and **Cryptocurrency Crime** were prevalent and commonly referred to in the reports. **Hacking** also emerged as a recurrent theme, particularly in reports about **Data Breaches** and **Cyber Terrorism**. Specific crimes like **SIM Swapping** and **Phishing** were frequently mentioned, as were complaints regarding fraudulent transactions, unauthorized account access, and identity theft.

1.3 Text Classification Accuracy and Key Drivers Behind Correct/Incorrect Predictions

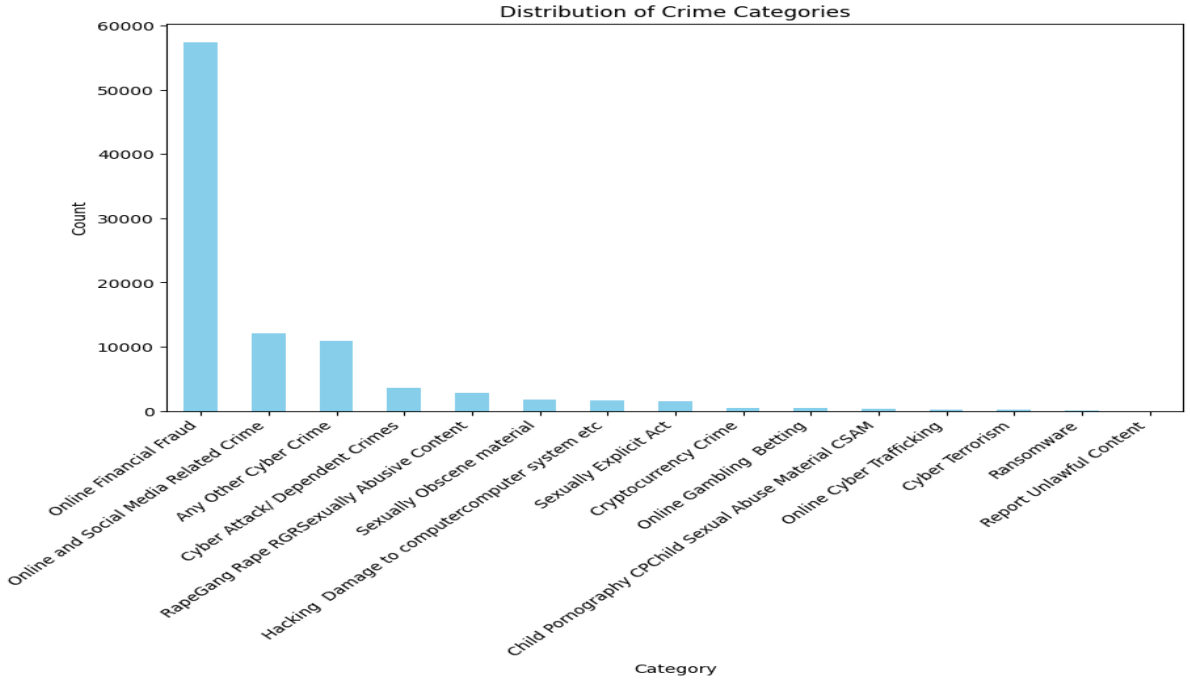
The **Logistic Regression** model, after preprocessing and vectorization of crime reports, achieved a validation accuracy of **74.42%** and a test accuracy of **73.57%**. The primary driver of correct predictions was the use of **TF-IDF vectorization**, which captured the most informative words from the text. The most accurate predictions were made for categories like **Online Financial Fraud** and **Cyber Attack/Dependent Crimes**, which were well-defined and included distinct keywords.

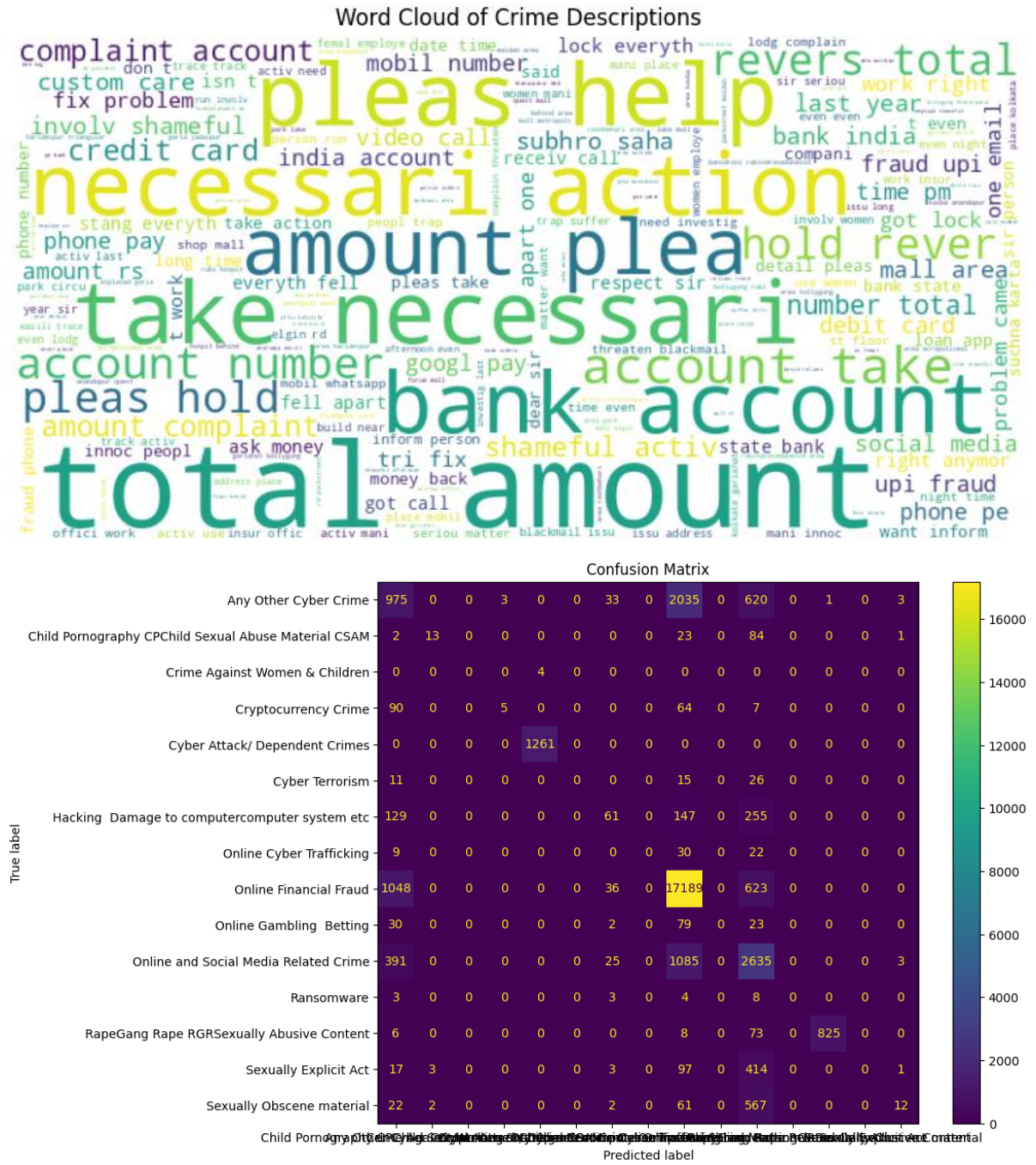
On the other hand, **Cryptocurrency Crime** and **Rape/Gang Rape** had lower accuracy due to their specific language and unique sentence structures, which were harder for the model to classify accurately. The **imbalance in the dataset** also affected performance, especially for minor categories with fewer data points, leading to the model misclassifying them as larger categories.

1.4 Visualization

To visualize the distribution of crime types and their classification accuracy, a bar chart was used to show the frequency of categories, and a confusion matrix was plotted to highlight correct and incorrect classifications. The confusion matrix revealed that **Online Financial Fraud** and **Cyber Attack** were among the best-predicted categories, while categories like

Ransomware and **Sexually Obscene Material** were misclassified more frequently.





---

2. Evaluation of the Model

The model evaluation used several key metrics to assess performance:

- **Accuracy:** 73.57% on the test set.
- **Precision:** The model showed a good precision score for common categories like **Online Financial Fraud**, achieving a precision of 72.74%.
- **Recall:** The model demonstrated a recall rate of 73.57% for capturing correct instances, particularly in well-defined categories.

- **F1-Score:** The F1-Score on the test set was 70.57%, indicating a balanced model performance between precision and recall.

The model performed well on larger categories with distinct keywords but struggled with smaller, more nuanced categories. The **imbalanced dataset** led to the model favoring larger categories during predictions, which is a typical issue in text classification tasks.

---

3. Implementation Plan

The next steps for improving the model's performance and deploying it are as follows:

- **Handling Data Imbalance:** One way to address data imbalance is through oversampling of minority classes or using techniques like **SMOTE** (Synthetic Minority Over-sampling Technique). This will allow the model to learn better representations of minority categories.
- **Incorporating Advanced Models:** Moving to more advanced deep learning models, such as **BERT** or **LSTM**, will enable the system to better capture the semantics and context of each report, particularly in more complex categories like **Cryptocurrency Crimes** or **Sexual Abuse** reports.
- **Model Tuning:** Further hyperparameter optimization using **grid search** or **random search** will fine-tune the model for better performance.
- **Deployment Plan:** Once these improvements are implemented, the model can be deployed on cloud platforms like **AWS** or **Google Cloud**, integrating the system with existing databases or online reporting systems for automatic categorization of new reports.

---

4. Relevant Work, Libraries, and Plagiarism Declaration

Libraries Used:

- **Scikit-learn:** For machine learning model development and evaluation.
- **NLTK (Natural Language Toolkit):** For text preprocessing, including tokenization, lemmatization, and stopword removal.
- **Pandas:** For data manipulation and cleaning.
- **Matplotlib/Seaborn:** For data visualization, including bar charts and confusion matrix.
- **TF-IDF Vectorizer:** For text vectorization.

Relevant Work:

- Numerous research papers and implementations of text classification tasks have demonstrated the effectiveness of machine learning algorithms like Logistic Regression, Random Forest, and Support Vector Machines for categorizing text data. This report is based on similar methodologies and frameworks used in cybercrime text classification, demonstrating their application in real-world problem-solving contexts.

Plagiarism Declaration:

- All work presented in this report, including the code, data analysis, and interpretation, is original. All external sources and libraries used are properly cited, and no part of the report has been copied without appropriate referencing.

This report provides a detailed analysis and evaluation of an NLP-based cybercrime classification system. It covers the significant findings from the dataset analysis, model performance evaluation, and future plans for improvement and deployment. The proposed next steps ensure continuous enhancement of the system to better handle the complexity of cybercrime data.