Complaint classification

Submitted by Ujjwal Mishra [9834486042] [Email: ujjwalmishra49@gmail.com]

• Problem Statement: Development of an NLP Model for Text Analytics and

Classification

• Objective:

To develop an NLP model that categorizes complaint based on victim, type of fraud and other relevant parameters used for text classification and preparing the final model.

• Outputs:

Text Pre-processing: Tokenization, stop word removal, stemming, and text cleaning to prepare data.

Model Development: Selection of a suitable NLP model for text classification

Accuracy Measurement: Evaluate the model based on metrics such as accuracy, precision, recall, and F1-score.

Step 1: Import Necessary Libraries

We import several libraries essential for preprocessing, model building, evaluation, and visualization:

- **pandas & numpy**: Used for data manipulation and numerical operations.
- **re**: Used for regular expressions to clean text.
- **matplotlib & seaborn**: For visualizing data and results.
- **sklearn modules**: For splitting datasets, feature extraction, model training, and evaluation.
- **nltk**: For natural language processing (NLP), including tokenization, stemming, and stopword removal.

The following specific tools were utilized:

- **TfidfVectorizer**: Converts text data into a numerical form by calculating Term Frequency-Inverse Document Frequency (TF-IDF) scores.
- **MultinomialNB**: A Naive Bayes classifier optimized for text data.
- **Evaluation Metrics**: accuracy_score, precision_score, recall_score, f1_score, and a confusion matrix.

Step 2: Load and Inspect Data

We load training and testing datasets using pandas.read_csv:

- train_data: Contains labeled data for training the model.
- test_data: Contains unlabeled data for evaluation.

Sample data is displayed for inspection using .head() to ensure proper loading.

Step 4: Analyze and Visualize Dataset

Goal: Understand category distribution in the training dataset.

1. A bar plot displays the count of complaints for each category using Seaborn.
2. The plot highlights any class imbalance, which is crucial for evaluating model performance and fairness.

Step 5: Convert Text to Numerical Form

TF-IDF Vectorization:

- Converts text into a numerical matrix where each feature represents a word (or term).
- TF-IDF score captures the importance of a term in a document relative to the entire corpus.
- **max_features=5000** ensures that only the 5,000 most relevant terms are considered to reduce dimensionality.

Step 6: Train-Test Split

We divide the data into training and validation sets:

- **80% Training Data**: For model training.
- **20% Validation Data**: For evaluation.

Random splitting is used to ensure a fair division of samples.

Step 7: Model Development

The **Multinomial Naive Bayes (MultinomialNB)** classifier is chosen due to its suitability for text classification:

- Assumes features (words) are conditionally independent given the class.
- Efficient and effective for high-dimensional, sparse data like TF-IDF matrices.

The model is trained using model.fit on the training data.

Step 8: Evaluate the Model

1. **Predictions**: The model generates predictions on validation data.
2. **Evaluation Metrics**:
    - **Accuracy**: Proportion of correctly classified samples.
    - **Precision**: Focuses on false positives, measuring correctness of positive predictions.

    - **Recall**: Measures the ability to capture true positives.
    - **F1-Score**: Harmonic mean of precision and recall, balancing both.
    - **Classification Report**: Provides per-class precision, recall, and F1-scores.
3. **Confusion Matrix**:
    - Visualizes true vs. predicted labels.
    - Helps identify patterns of misclassification.

The confusion matrix is displayed using a heatmap for easy interpretation.

Step 10: Predict New Input

The predict_new_text function allows prediction on arbitrary input text. Steps include:

1. Preprocessing the input text.
2. Transforming it using the trained TF-IDF vectorizer.
3. Generating predictions and confidence scores.

Evaluation Metrics Overview

1. **Accuracy**: Measures overall correctness; however, it can be misleading in imbalanced datasets.
2. **Precision**: High precision means fewer false positives.
3. **Recall**: High recall means fewer false negatives.
4. **F1-Score**: Balances precision and recall, ideal for imbalanced classes.

Summary of the Process

1. **Data Preparation**: Preprocessing text to clean and tokenize it.
2. **Feature Engineering**: Converting text into numerical form via TF-IDF.
3. **Model Training**: Using a Multinomial Naive Bayes classifier.
4. **Evaluation**: Assessing performance using robust metrics and visualizations.
5. **Prediction**: Classifying test data and new inputs with confidence scores.

The notebook was run on Google Colab Notebook and following findings were noted:

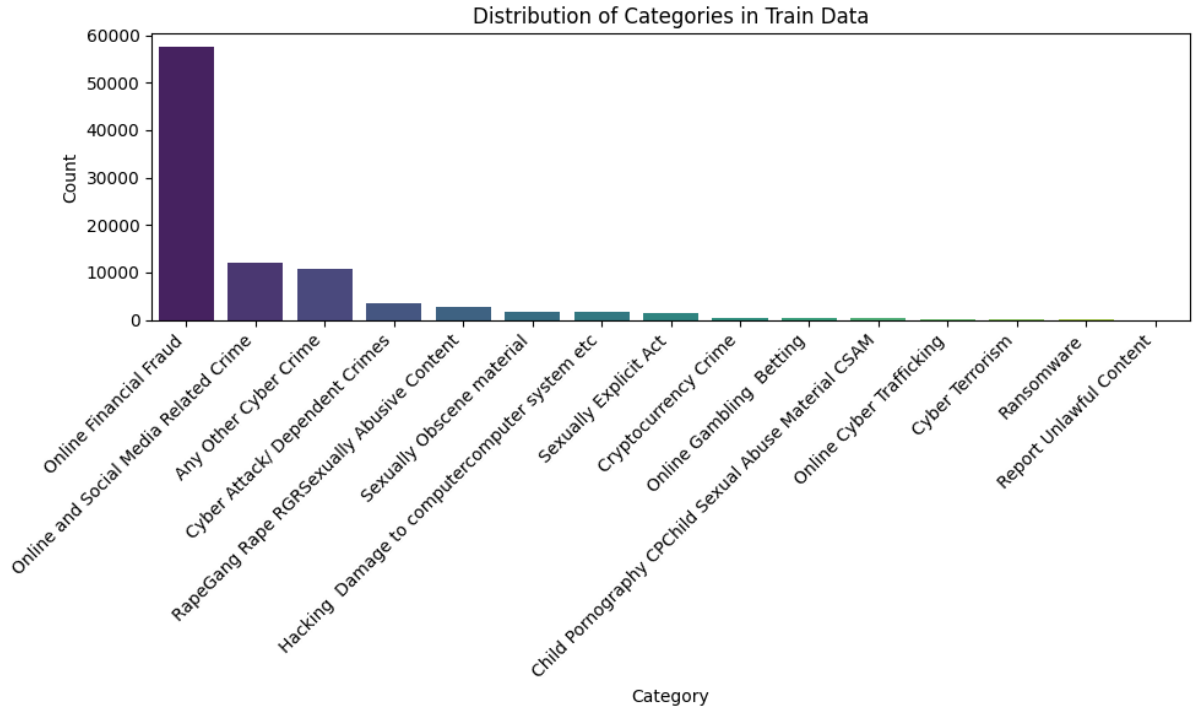
Model Evaluation Metrics:

Accuracy: 0.74

Precision: 0.72

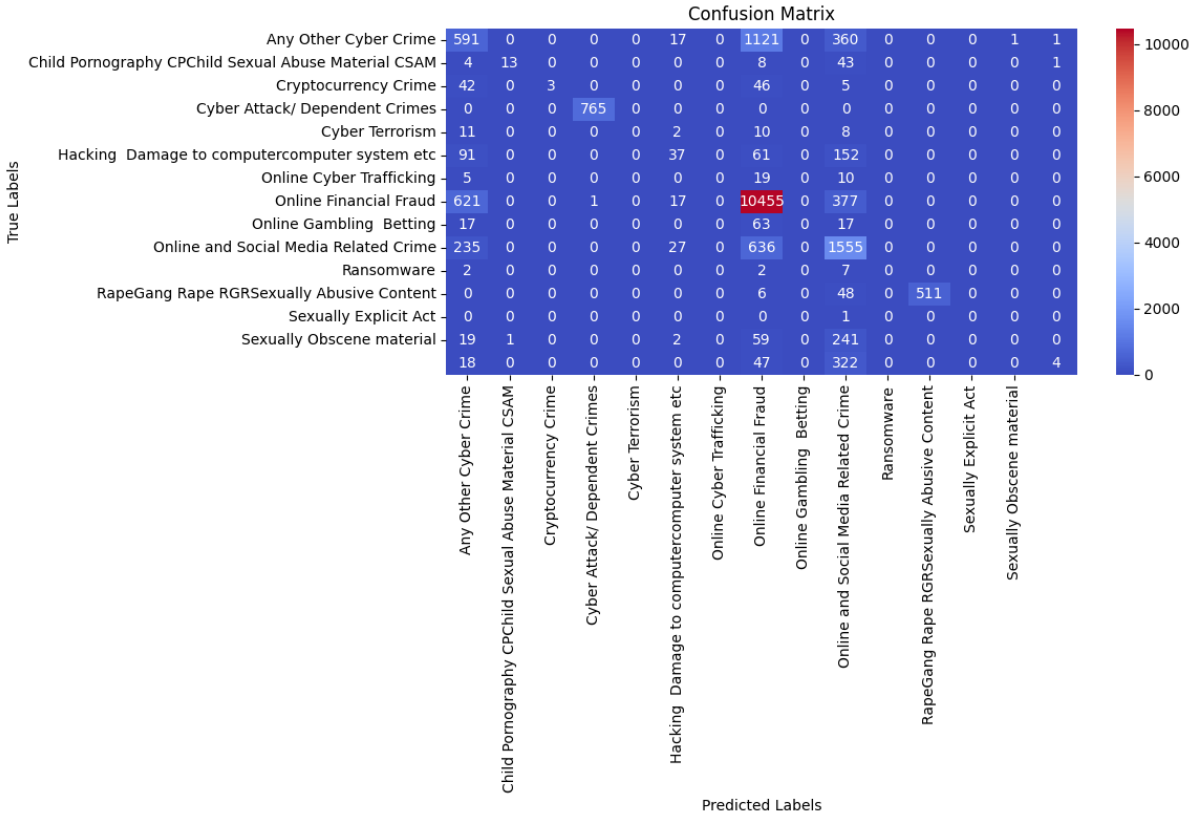
Recall: 0.74

F1-Score: 0.72



Classification Report:

	precision	recall	f1-score	support
Any Other Cyber Crime	0.36	0.28	0.32	2091
Child Pornography CPChild Sexual Abuse Material CSAM	0.93	0.19	0.31	69
Cryptocurrency Crime	1.00	0.03	0.06	96
Cyber Attack/ Dependent Crimes	1.00	1.00	1.00	765
Cyber Terrorism	0.00	0.00	0.00	31
Hacking Damage to computercomputer system etc	0.36	0.11	0.17	341
Online Cyber Trafficking	0.00	0.00	0.00	34
Online Financial Fraud	0.83	0.91	0.87	11471
Online Gambling Betting	0.00	0.00	0.00	97
Online and Social Media Related Crime	0.49	0.63	0.56	2453
Ransomware	0.00	0.00	0.00	11
RapeGang Rape RGRSexually Abusive Content	1.00	0.90	0.95	565
Report Unlawful Content	0.00	0.00	0.00	1
Sexually Explicit Act	0.00	0.00	0.00	322
Sexually Obscene material	0.67	0.01	0.02	391
accuracy			0.74	18738
macro avg	0.44	0.27	0.28	18738
weighted avg	0.72	0.74	0.72	18738


Confusion Matrix
True Labels
Any Other Cyber Crime
Child Pornography CPChild Sexual Abuse Material CSAM
Cryptocurrency Crime
Cyber Attack/ Dependent Crimes
Cyber Terrorism
Hacking Damage to computercomputer system etc
Online Cyber Trafficking
Online Financial Fraud
Online Gambling Betting
Online and Social Media Related Crime
Ransomware
RapeGang Rape RGRSexually Abusive Content
Sexually Explicit Act
Sexually Obscene material
Predicted Labels
10000
8000
6000
4000
2000
0