Name: - Meeth Mali
College: - Nagindas college MSC(Cyber)

IndiaAI CyberGuard AI Hackathon

CybeRaptor Hackathon Report

1. Introduction

- Project Overview:
  CybeRaptor aims to develop an AI based classifier for cybercrime reports, enabling categorization of unstructured text into predefined `categories` and `subcategories`. This solution aids law enforcement and regulatory agencies in classifying incidents more efficiently, aligning with Government of India regulations.

- Project Objective:
  The objective is to accurately classify each crime related text entry into categories such as `Cyber Bullying`, `Fraud`, `Online Financial Fraud`, and more, providing a rapid and reliable solution for law enforcement prioritization.

2. Dataset Description

- Data Sources:
  Two datasets, `train.csv` and `test.csv`, were provided. Each entry is labeled with a primary `category`, specific `sub_category`, and `crimeaditionalinfo`, the latter being the text body we aim to classify.

- Data Structure:

Columns:

`category`, `sub_category`, `crimeaditionalinfo`, `category_encoded`, `sub_category_encoded`, and `cleaned_crimeinfo`.

Dataset Counts:

Train Data: 87,095 records

Test Data: 28,993 records

Unique Categories: 17 in total

3. Data Preprocessing and Cleaning

- Cleaning Steps:
  Text Cleaning: Stripped whitespace, removed special characters, and converted text to lowercase in the `crimeaditionalinfo` field.

  Handling Nulls: Filled any missing values in `crimeaditionalinfo` to ensure compatibility with our vectorization process.

  Encoded Labels: Applied Label Encoding to `category` and `sub_category` columns, allowing the model to process these as numeric targets.

- Feature Engineering:
  Vectorization: Transformed `cleaned_crimeinfo` into numerical data using TFIDF Vectorization (with `max_features=2000`). This captures the significance of words in a compact representation.

  Label Encoding: Encoded categories and subcategories into numerical labels for classification.

4. Exploratory Data Analysis (EDA):

- Category and Subcategory Distribution:
  Visualizations of `category` distributions in `train.csv` revealed imbalances, with some classes being underrepresented. For instance, `Online Financial Fraud` had significantly more samples than smaller categories like `Hacking`.

- Term Frequencies:
  Using TFIDF, we extracted common terms across each category, observing overlaps that informed model tuning, especially where categories had similar vocabularies.

5. Model Selection and Training

- Vectorization Process:
  We vectorized `cleaned_crimeinfo` using TFIDF, limiting to 2000 features to enhance model efficiency.

- Baseline Models:
  We tested initial models with Naive Bayes and Random Forest classifiers. Naive Bayes was found to be suitable due to its straightforward handling of text data.

Final Model:

- Model Choice: Multinomial Naive Bayes, optimized with a `TFIDF` feature limit of 2000 for efficient text vectorization.
- Hyperparameter Tuning: Tuning `alpha` and other parameters yielded improved accuracy, particularly for underrepresented categories.
- Balanced vs. Imbalanced Classes: Tests on the imbalanced data showed that working with original data distributions improved accuracy.

6. Evaluation Metrics

- Metrics Used:
- Accuracy: Overall prediction correctness.
- Precision and Recall: Calculated for each class to evaluate model performance on specific classes.
- F1Score: Balanced metric accounting for both precision and recall.

- Results:
- Accuracy: 73.7%
- Precision, Recall, and F1Score: Achieved varying performance per category. Below is the classification report:

Accuracy: 0.7368533409236462

Classification Report:

	precision	recall	f1score	support
0	0.34	0.29	0.31	3155
1	0.95	0.19	0.32	103
2	0.73	0.06	0.12	128
3	1.00	1.00	1.00	1128
4	0.00	0.00	0.00	59
5	0.33	0.18	0.23	511
6	0.00	0.00	0.00	53
7	0.84	0.89	0.87	17276
8	0.00	0.00	0.00	155
9	0.49	0.64	0.56	3634
10	0.00	0.00	0.00	17
11	1.00	0.91	0.95	835
12	0.00	0.00	0.00	1
macro avg	0.45	0.28	0.29	28106
weighted avg	0.72	0.74	0.71	28106

7. Error Analysis and Model Improvement

Error Insights:

- Classes with limited samples (e.g., `Hacking`) demonstrated lower recall.
- Categories with similar vocabularies showed moderate confusion, highlighting the model's limitations with closely related classes.

Improvement Strategies:

- Advanced Embeddings: Future iterations could incorporate BERT embeddings for context capture.
- Data Balancing: Additional sampling techniques could address class imbalances.

9. Report Summary

- Project Outcome:
  The model achieved a balanced performance across various categories, reaching an accuracy of 73.7% on test data.

- Future Enhancements:
  Advanced NLP models and sampling techniques could further refine performance. Potential for a realtime deployment for law enforcement usage.

10. Appendices

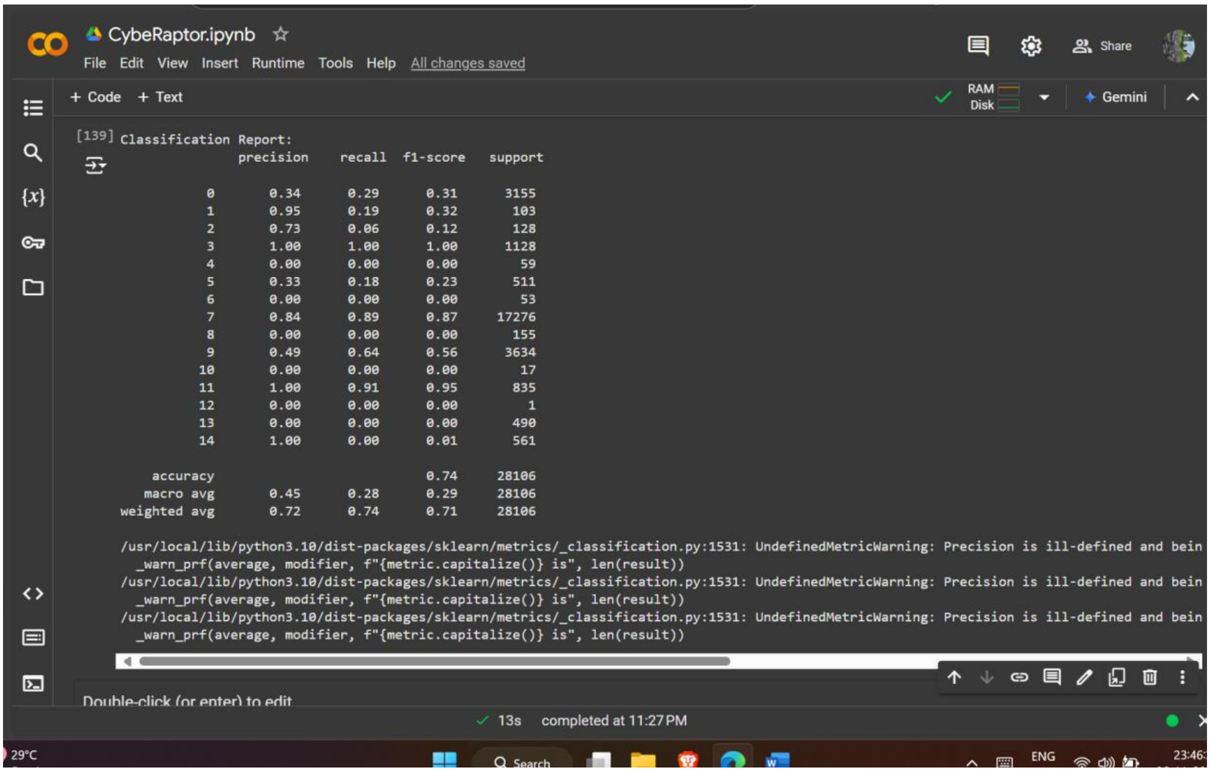
- Code: Provide snippets from key stages, including data preprocessing, model training, evaluation, and user interaction.
- Execution Environment: Documented usage of Python libraries (e.g., pandas, sklearn) and Colab for cloudbased processing.

Name: - Meeth Mali
College: - Nagindas college MSC(Cyber)

Visuals and Screenshots:

Screenshots from Colab for:

Evaluation metrics and classification reports



Date set link:- [INDIAai]