Secure Content Filtering in Document Summarization

Team Members

Akshat Gangrade (axg210048)

Ankit Sahu (axs210226)

Sarthak Gupta (sxg200139)

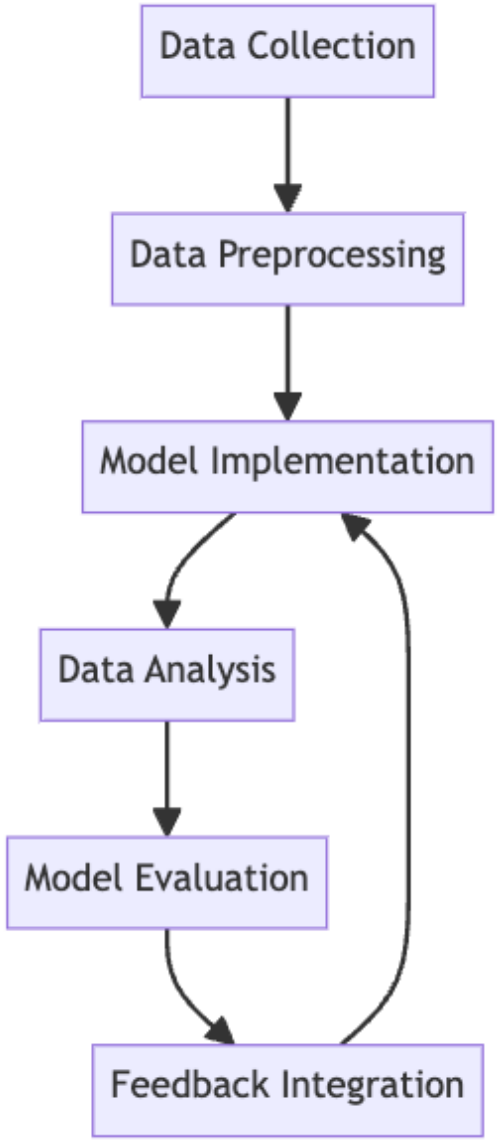
Yash Shah (yxs210015)

Problem Statement

- Integration of Security Protocols in Summarization
- Handling Classified Information
- Innovative Use of NLP Techniques
- Dynamic Clearance Check Function
- Continuous Improvement and Adaptation
- Beyond Traditional Summarization Methods

System Design

- Data Collection and Preprocessing
- Model Implementation with NLP and Security Layers
- Data Analysis for Security Assessment
- Model Evaluation: Accuracy and Sensitivity Analysis



System Implementation

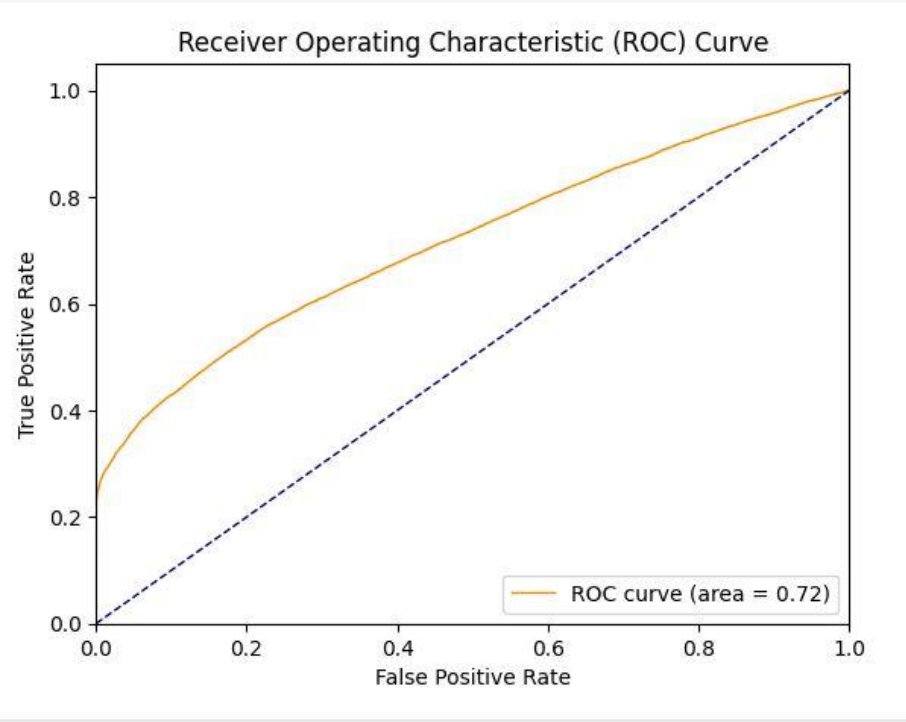
- Feature Analysis with NLP Tools
- Data Visualization for Preprocessing Insights
- Model Training with Machine Learning Algorithms
- Implementing Security Layer in Summarization
- Evaluation on Test Dataset
- Practical Application for Controlled Access

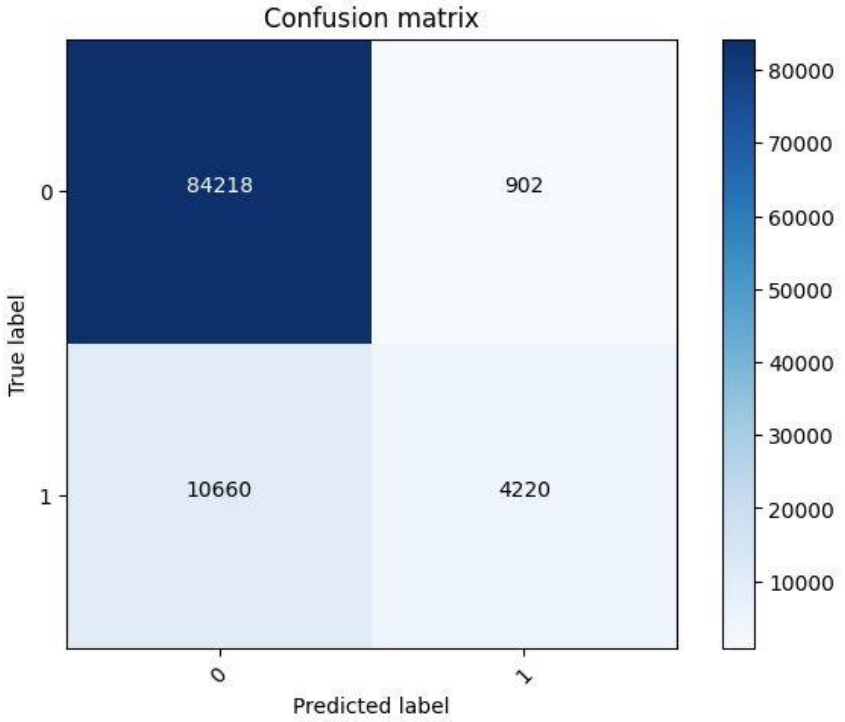
Experiment Setup

- **Platform:** Google Colab for cloud computation.
- **Programming:** Python for its extensive data science and ML libraries.
- **Visualization Tools:** Data visualization with Matplotlib and Seaborn.
- **Libraries:** Pandas for data manipulation, SkLearn for ML, NumPy for numerical tasks, and NLP tools (NLTK, spaCy).
- **Data Splitting:** Dataset divided into 80% training and 20% testing.
- **Evaluation Metrics:** Accuracy, precision, recall, F1 score, ROC curve.
- **Model Comparison:** Baseline (Logistic Regression, Random Forest) vs advanced NLP models for content filtering and summarization.
- **Testing & Optimization:** Model testing on test dataset, fine-tuning with cross-validation.
- **Documentation:** Comprehensive documentation for reproducibility.

Results

Confusion Matrix and ROC curve Diagrams





Results

Different models – Random Forest Classifier & Logistic Regression

Classification Report

	precision	recall	f1-score	support
0	0.89	0.99	0.93	85030
1	0.81	0.28	0.42	14970
accuracy			0.88	100000
macro avg	0.85	0.63	0.68	100000
weighted avg	0.88	0.88	0.86	100000

	precision	recall	f1-score	support
0	0.88	0.99	0.93	84816
1	0.82	0.28	0.41	15184
accuracy			0.88	100000
macro avg	0.85	0.63	0.67	100000
weighted avg	0.87	0.88	0.86	100000

Confusion Matrix

```
[[84218   902]
 [10660  4220]]
```

```
[[84197   968]
 [10673  4162]]
```

Challenges Encountered

- Difficulty finding an appropriate dataset
- Limited features
- Ethical and Privacy Concerns
- Real-time Data Processing
- Language and Textual Data Challenges

Future Directions

- The model's objective is to automatically identify and replace sensitive information like Zip Codes, SSNs, Addresses, and Passwords in uploaded PDFs with an 'X'.
- Replace the sensitive information with ‘X’ when sending the information from one person to another person.
- Currently the model is built on the English language and has limited functionality in other languages.
- The model is built on sensitive filtration but in future we can add more features like toxic, inserting, obscene and threat.

Thank You