Rajasthan Police Hackathon 1.0

AUTOMATED AI/ML SYSTEM FOR DETECTING AND MITIGATING ONLINE FRAUD

Sumitra Sharma, Payal Kaur, Soumya Pandey

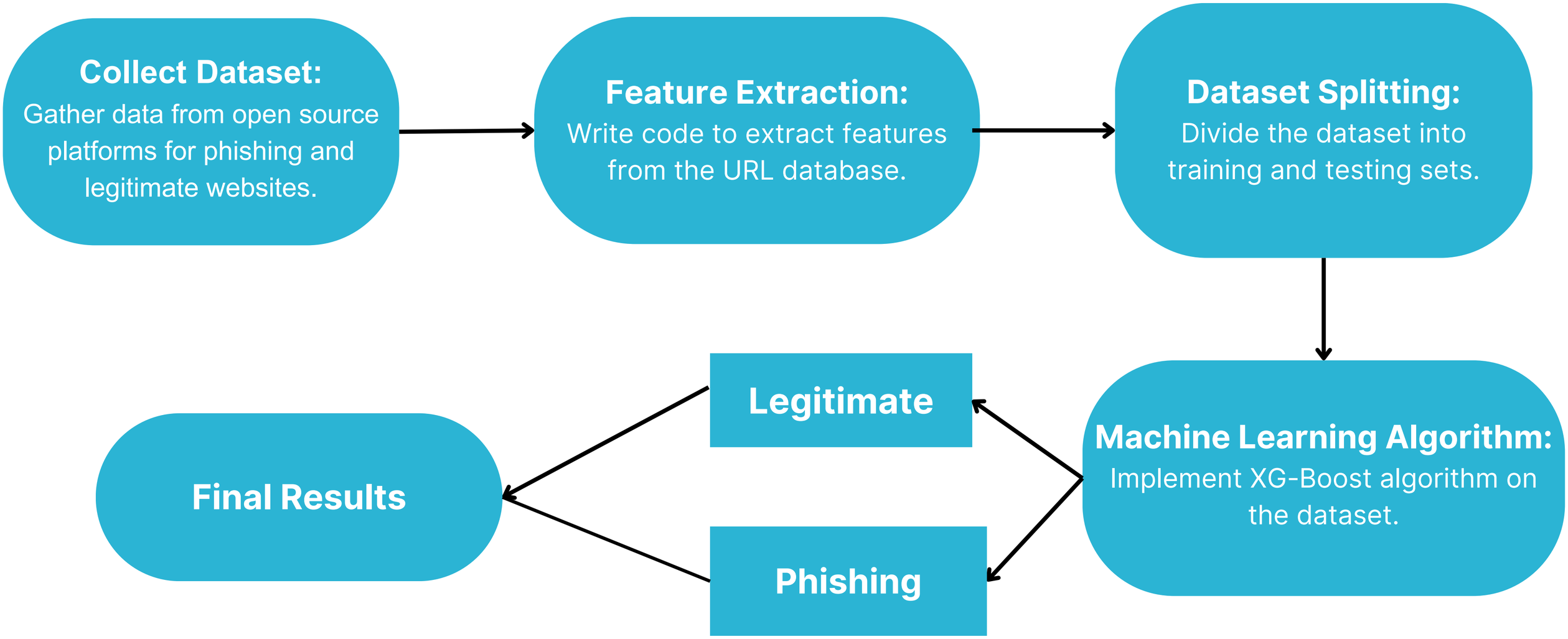
Team: CyberZen

OBJECTIVE

Create and implement an AI/ML-based system that can autonomously analyze and categorize online content, distinguishing between authentic and fake/fraudulent websites, advertisements, and customer care numbers. The system aims to achieve the following:

1. Website Authentication
2. Ad Content Analysis
3. Customer Care Number Verification
4. Real-time Detection.
5. User Feedback Integration

Approach for Website Authentication



FEATURE EXTRACTION

Address Bar based Features: “Domain of URL”, “Redirection ‘//’ in URL”, ”IP Address in URL”, ” ‘http/https’ in Domain name”, ” ‘@’ Symbol in URL”, ”Using URL Shortening Service”, “Length of URL”, “Prefix or Suffix "-" in Domain”, “Depth of URL”

Domain based Features: “DNS Record”, “Age of Domain”, “Website Traffic”, “End DNS Period of Domain”

HTML & Javascript based Features: “Iframe Redirection”, “Disabling Right Click”, “Status Bar Customization”, “Website Forwarding”

ADDITIONAL FUNCTIONS FOR WEBSITE AUTHENTICATION

Is Domain Legitimate

Is Certificate Valid

Is HTTPS

WEB SCRAPING FOR HYPERLINKS EXTRACTION

Selenium (Framework)

BeautifulSoup (Python library)

- Img
- amp-img
- amp-ad
- iframe
- a tag
- Style
- script

AD AND IMAGE CONTENT ANALYSIS THROUGH NLP

EasyOCR (library for python)

A technology that extracts text from images or scanned documents, making it possible to convert images containing text into machine-readable text data.

Dataset

Cross experiment by merging sms dataset and comments dataset found on Kaggle, in which statements are classified as (1) spam or (0) ham.

AD AND IMAGE CONTENT ANALYSIS THROUGH NLP

Data Splitting:
The dataset is divided into two parts: a training set and a validation set.

Tokenization:
converting words/characters to numbers, essential for embeddings.

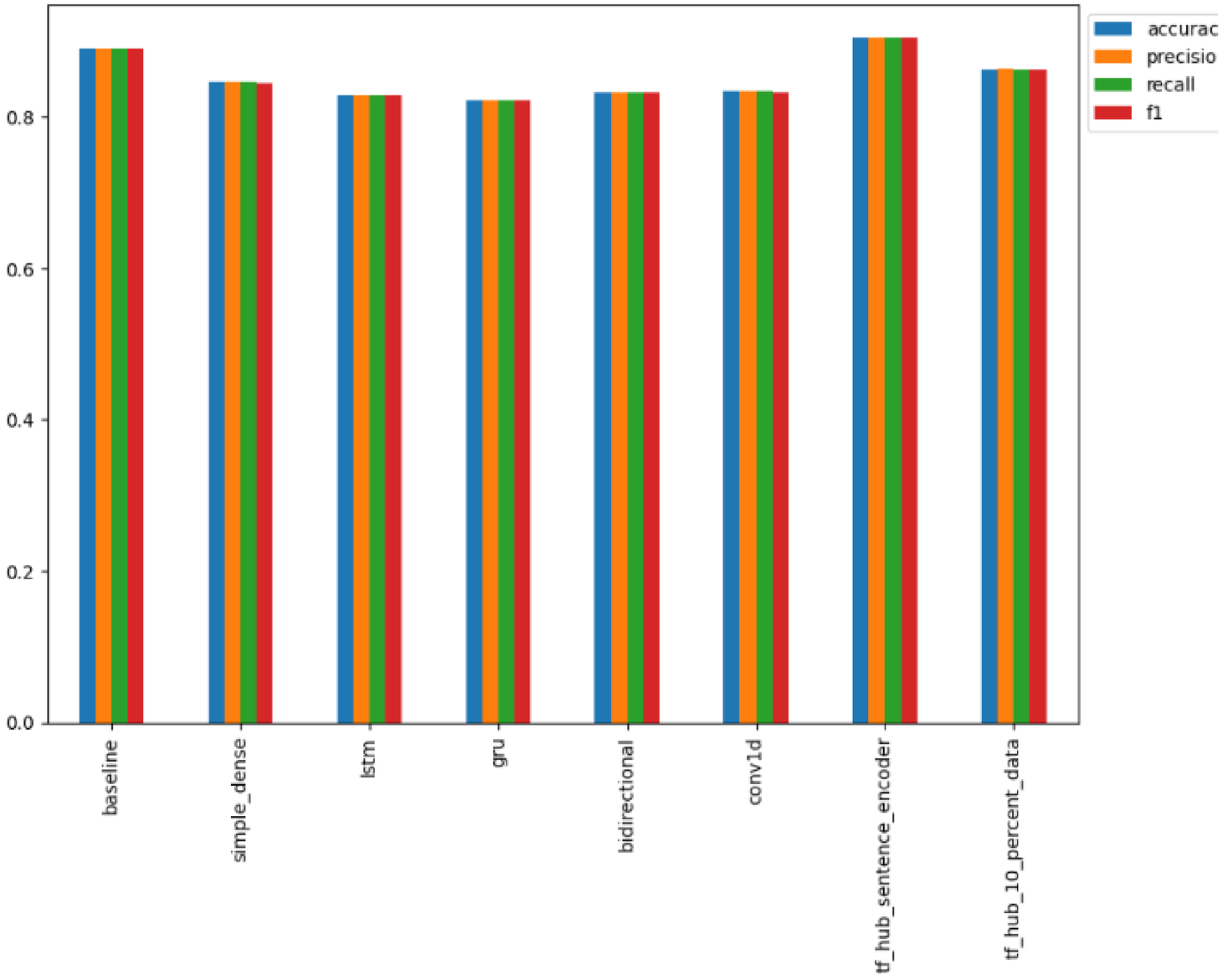
Pretrained Embeddings and Transfer Learning:
Utilizing pretrained embeddings and transfer learning leverages knowledge from a larger model for a different task.

Universal Sentence Encoder:
A chosen pretrained embedding from TensorFlow Hub converting entire sentences into numerical representations.

Model : TensorFlow Hub Pretrained Sentence Encoder:
The new model uses USE as its embedding layer.

Saving and Loading Models:
Two common formats for saving models in TensorFlow are mentioned: HDF5 and SavedModel.

Comparison of different NLP models for Ad and Image content analysis



CUSTOMER CARE NUMBER VERIFICATION

Model deployed on Google Cloud.

Using Truecaller’s API for customer care verification

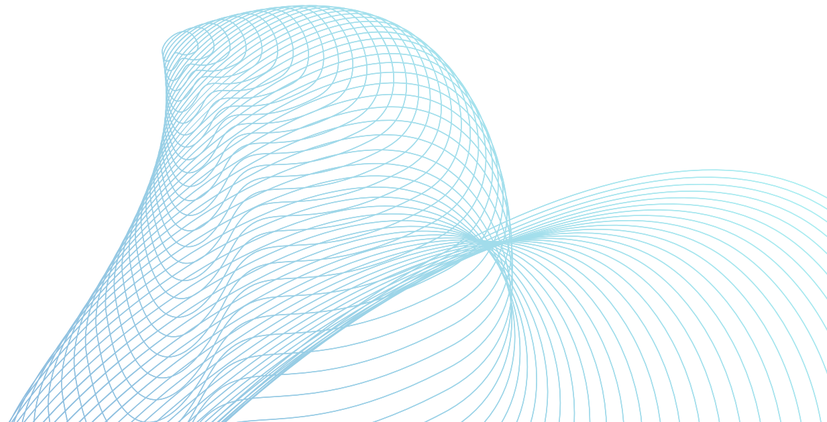
LIMITATIONS AND FUTURE WORKS

Future Work

1. Develop browser extension integrating URL analyzer for real-time fraudulent URL detection for end user.
2. Scale web analyzer system by implementing recursive hyperlink analysis for input URLs.
3. Advance to multilingual OCRs, followed by NLP models.

Limitations

1. Utilizing TensorFlow in NLP model with GPU preference, currently restricted to CPUs due to resource limitations.
2. Achieved an accuracy of 77%.
3. Ongoing efforts to improve efficiency in OCR processing.



TECH STACK USED

THANK YOU