Deep Learning Based Detection of Phishing Websites

David Toluwani Ehigie

Week 1 Progress

Motivation

- Phishing is a major cybersecurity threat
- Manual detection is difficult
- Goal: Build AI-based phishing detection system

Problem Statement

Binary Classification Problem:

- Class 0: Legitimate Website
- Class 1: Phishing Website
- Supervised Learning Approach

Dataset

Web Page Phishing Dataset

- 11,000+ website samples
- URL and domain-based features
- Structured tabular dataset

Data Preprocessing

- Data cleaning
- Feature-label separation
- StandardScaler normalization
- Train-test split (80/20)

Baseline Model

Random Forest Classifier

- Ensemble learning method
- Handles tabular data well
- Reduces overfitting

Evaluation Metrics

- Accuracy
- Precision
- Recall
- F1-score

Next Steps

Week 2:

- Deep Learning model
- URL tokenization
- LSTM/CNN training