Phishing Detection Using Neural Network

Student Information:

- **Name:** Habiba Tarek Nassar
- **ID:** 2205188
- **Course**: AI Security Issues

1. Introduction

Phishing attacks are a prevalent cybersecurity threat where attackers impersonate legitimate entities to steal sensitive information. This report presents a deep learning-based approach to detect phishing attacks using two models:

- Multi-Layer Perceptron (MLP) (Section 4)
- Logistic Regression Model (Section 5)

We compare their performance, analyze MITRE ATT&CK techniques, and propose mitigation strategies.

---

2. Dataset Description

- **Dataset**: dataset_phishing.csv
- **Features**: 87 attributes (e.g., URL structure, domain characteristics)
- **Target**: Binary classification (status):
    - 0 = Legitimate
    - 1 = Phishing

Preprocessing Steps:

- Removed URL column (non-numeric data).
- Normalized features using MinMaxScaler.
- Split data into training (80%) and testing (20%) sets.

---

3. Model Implementations

3.1 Multi-Layer Perceptron (MLP)

Training Parameters:

- Optimizer: Adam (lr=0.001)
- Loss Function: Binary Cross-Entropy (BCELoss)

- Batch Size: 32
- Epochs: 100

Results:

Epoch	Train Accuracy	Train Loss	Val Accuracy	Val Loss
1	0.91	0.21	0.94	0.15
100	0.98	0.05	0.95	0.18

- **Final Test Accuracy**: 98.1862%

3.2 Logistic Regression Model

Training Parameters:

- Optimizer: Adam (lr=0.001)
- Loss Function: Binary Cross-Entropy (BCELoss)
- Batch Size: 32
- Epochs: 100

Results:

Epoch	Train Accuracy	Train Loss	Val Accuracy	Val Loss
1	0.92	0.19	0.91	0.20
100	0.94	0.15	0.93	0.16

- **Final Test Accuracy**: 94.3291%

---

4. Model Comparison

Metric	MLP (Section 4)	Logistic Regression (Section 5)
Accuracy	98.1862%	94.3291%
Loss	0.0987	0.2012
Complexity	High	Low
Training Time	Longer	Faster
Overfitting	Controlled	Minimal

Conclusion:

- MLP performs better (~2.3% higher accuracy).
- Logistic Regression is simpler but less accurate.

---

5. MITRE ATT&CK Framework Analysis

Phishing Techniques Identified:

Technique ID	Name	Indicators	Mitigations
T1566.001	Spearphishing Attachment	having_IP_Address, URL_of_Anchor	User Training (M1017)
T1566.002	Spearphishing Link	Shortining_Service, having_At_Symbol	Restrict Web-Based Content (M1021)
T1566.003	Spearphishing via Service	SFH, web_traffic	Software Configuration (M1054)

Mitigation Strategies Applied:

- Blocking detected phishing URLs (prediction > 0.5)
- User training (M1017): Educate users on phishing
- Restrict suspicious web content (M1021)

MITRE Attack Framework:

The MITRE Attack Framework class is a custom Python module that maps phishing-related activities to specific techniques and sub techniques within the **MITRE ATT&CK** framework. It acts as a bridge between observed phishing features in a dataset and known attack patterns, helping analysts or detection systems interpret feature-based phishing indicators through a structured cybersecurity lens.

Purpose

This framework was developed to:

- **Identify** phishing tactics and sub-techniques using dataset features (e.g., presence of IP addresses, shortened links).
- **Assess** the likelihood (confidence) that a phishing technique was used.
- **Recommend** appropriate **mitigations** aligned with MITRE's knowledge base (e.g., user training or restricting web content).

MITIGATIONS Dictionary:

Maps mitigation IDs to descriptive strategies:

- M1017: User Training

- M1021: Restrict Web-Based Content
- M1054: Software Configuration

These mitigations are essential controls to reduce the success rate of phishing attacks.

How It Works :

- You feed the method a dictionary like:
  {'having_IP_Address': 1, 'Shortining_Service': 0.6, 'SFH': 0.2}
- The method checks:
  - Which technique each feature relates to.
  - Whether the feature value is significant (i.e., above 0.5).
  - Counts how many indicators are "on".
  - Produces a confidence score for each match.
  - Recommends which attack techniques may be happening and how to mitigate them.

---

6. Discussion & Future Work

- MLP is superior for phishing detection but requires more computational power.
- Logistic Regression is suitable for lightweight applications.

---

7. Conclusion

This project successfully implemented two deep learning models for phishing detection:

- MLP: 98.1862% accuracy
- Logistic Regression: 94.3291% accuracy

The MITRE ATT&CK framework helped map attack techniques to mitigation strategies.

Recommendation:

- Use MLP for high-security environments.
- Use Logistic Regression for lightweight systems.