My Final Year Project

**### Project Title:**

\*\*ML-Based Phishing Website Feature Classification\*\*

**### Project Overview:**

For my MTech final year project (Sep 2022 – May 2023), I'm working on a machine learning-based project focused on classifying features of phishing websites. The primary goal is to develop a robust classification system that can distinguish between legitimate and phishing websites based on various features.

**### Project Objectives:**

1. **\*\*Dataset Collection: \*\***

- Gather a diverse dataset of labelled examples, including both legitimate and phishing websites.

- Include features such as URL structure, SSL certificate details, HTML and JavaScript analysis, etc.

2. **\*\*Feature Extraction: \*\***

- Pre-process the collected data and extract relevant features from the websites.

- Features could include URL length, presence of HTTPS, domain age, use of pop-ups, and other indicators of phishing behaviour.

3. **\*\*Data Labelling: \*\***

- Manually label the dataset for training the machine learning model.

- Ensure a balanced representation of phishing and legitimate websites in the labelled dataset.

4. **\*\*Model Selection: \*\***

- Choose appropriate machine learning algorithms for classification.

- Consider algorithms such as Decision Trees, Random Forest, Support Vector Machines, or Neural Networks based on the nature of your dataset.

5. **\*\*Model Training: \*\***

- Split the dataset into training and testing sets.

- Train the selected machine learning model using the training set.

6. **\*\*Evaluation: \*\***

- Evaluate the model's performance using the testing set.

- Metrics may include accuracy, precision, recall, F1-score, and ROC curves.

7. **\*\*Feature Importance Analysis: \*\***

- Analyse the importance of different features in the classification process.

- This analysis can provide insights into the indicators of phishing websites.

8. **\*\*Model Optimization: \*\***

- Fine-tune the model for better performance.

- Consider techniques such as hyperparameter tuning and cross-validation.

9. **\*\*Deployment: \*\***

- Integrate the trained model into a practical tool or application.

- Consider building a web browser extension or an API for real-time website classification.

10. **\*\*Documentation and Reporting: \*\***

- Document the entire project, including dataset details, methodology, and results.

- Prepare a final report and presentation for the MTech project defence.

**### Challenges and Future Work:**

- Address imbalanced datasets, choose features wisely, and ensure model interpretability.

- Future work may involve improving the model with more diverse datasets and exploring additional features.

**### Conclusion:**

This project aims to contribute to the field of cybersecurity by developing an effective machine learning model for classifying phishing websites, ultimately enhancing web security.