Creating a URL Phishing Detector

by Aggrey Timbwa, Richard Macharia, Pamela Jepkorir Chebii, Cynthia Njambi, Omara Waldea

What is Phishing?

Deceptive Tactics

- Fraudsters use email, social media, and other online channels to trick users into revealing sensitive information.

Real-World Examples

- Emails impersonating banks, government agencies, and trusted brands are common phishing techniques.

Understanding Phishing Attacks

1

Business Risks

- Financial losses
- reputational damage.

2

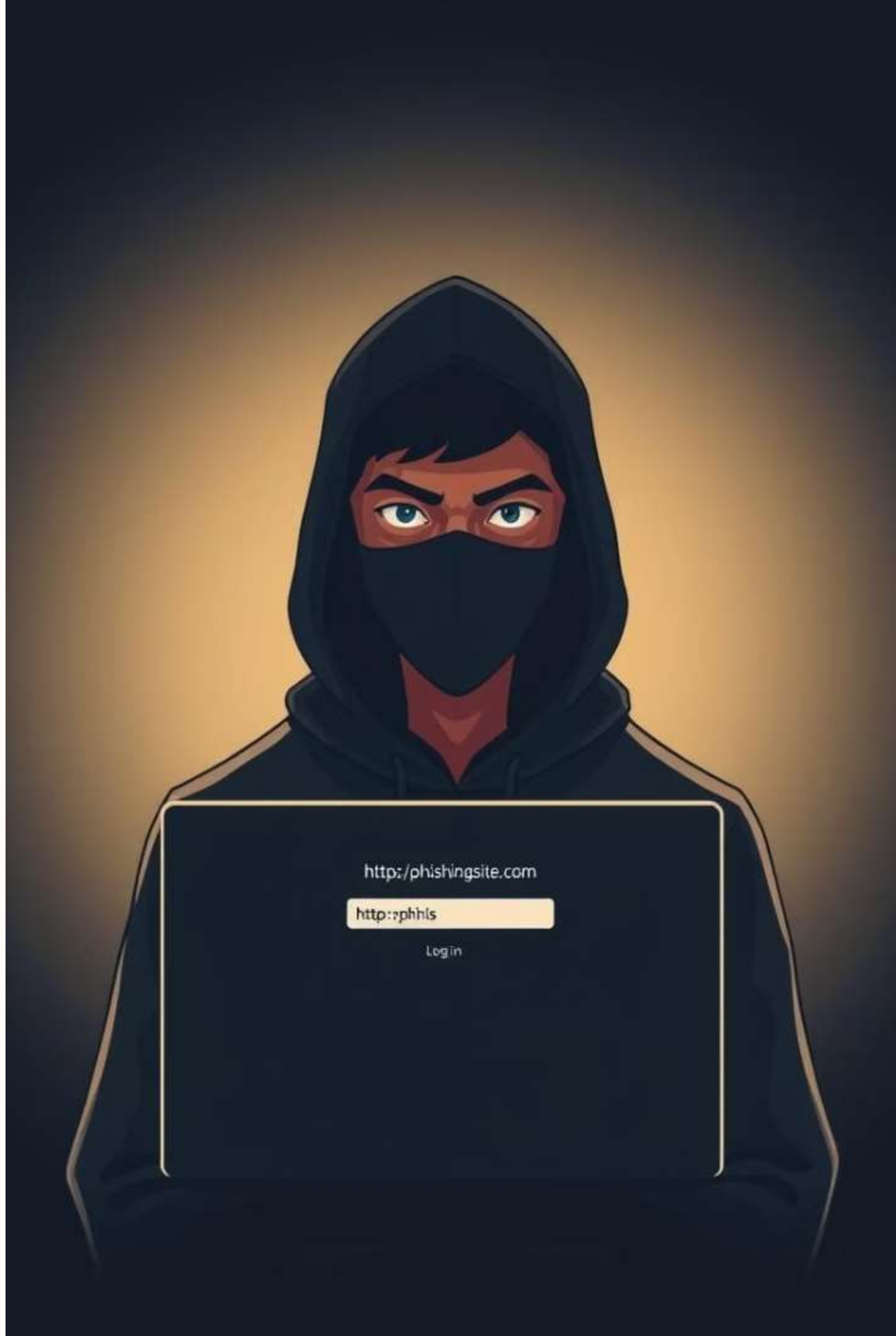
Cybercrime Sophistication

- Advanced tactics
- Social engineering.

3

Protection Necessity

- Tools to identify phishing URLs for real-time detection.



Business Objectives

High-Accuracy Detection

User-Friendly Deployment

Real-Time Classification

Expected Impact Globally

1 Financial Security

- Reduce financial losses due to phishing scams.

2 User Trust

- Build confidence in online interactions.

Preparing a Training Dataset

1

Data Collection

- Gather diverse labelled URLs from Mendely Dataset .

2

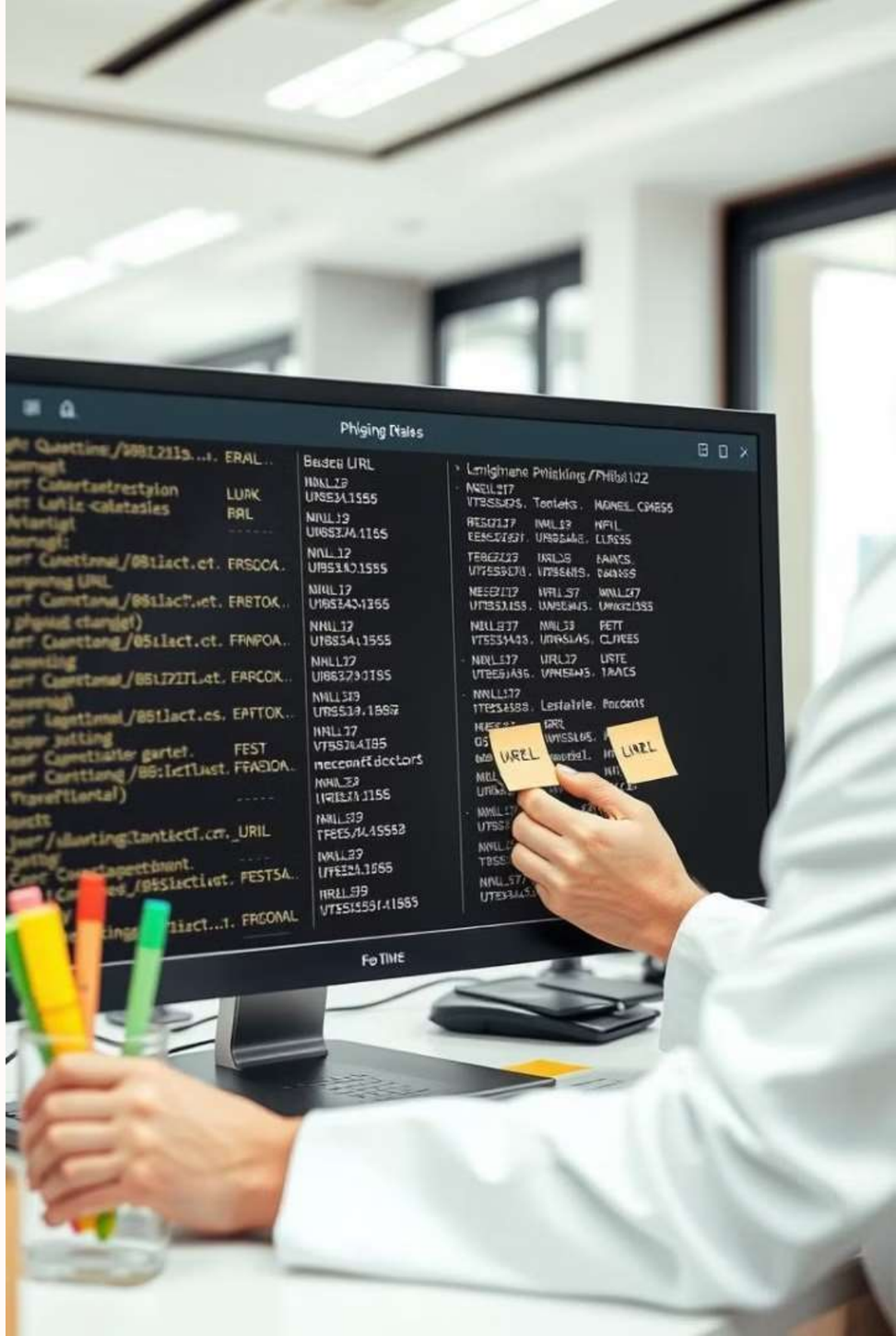
Data Preprocessing

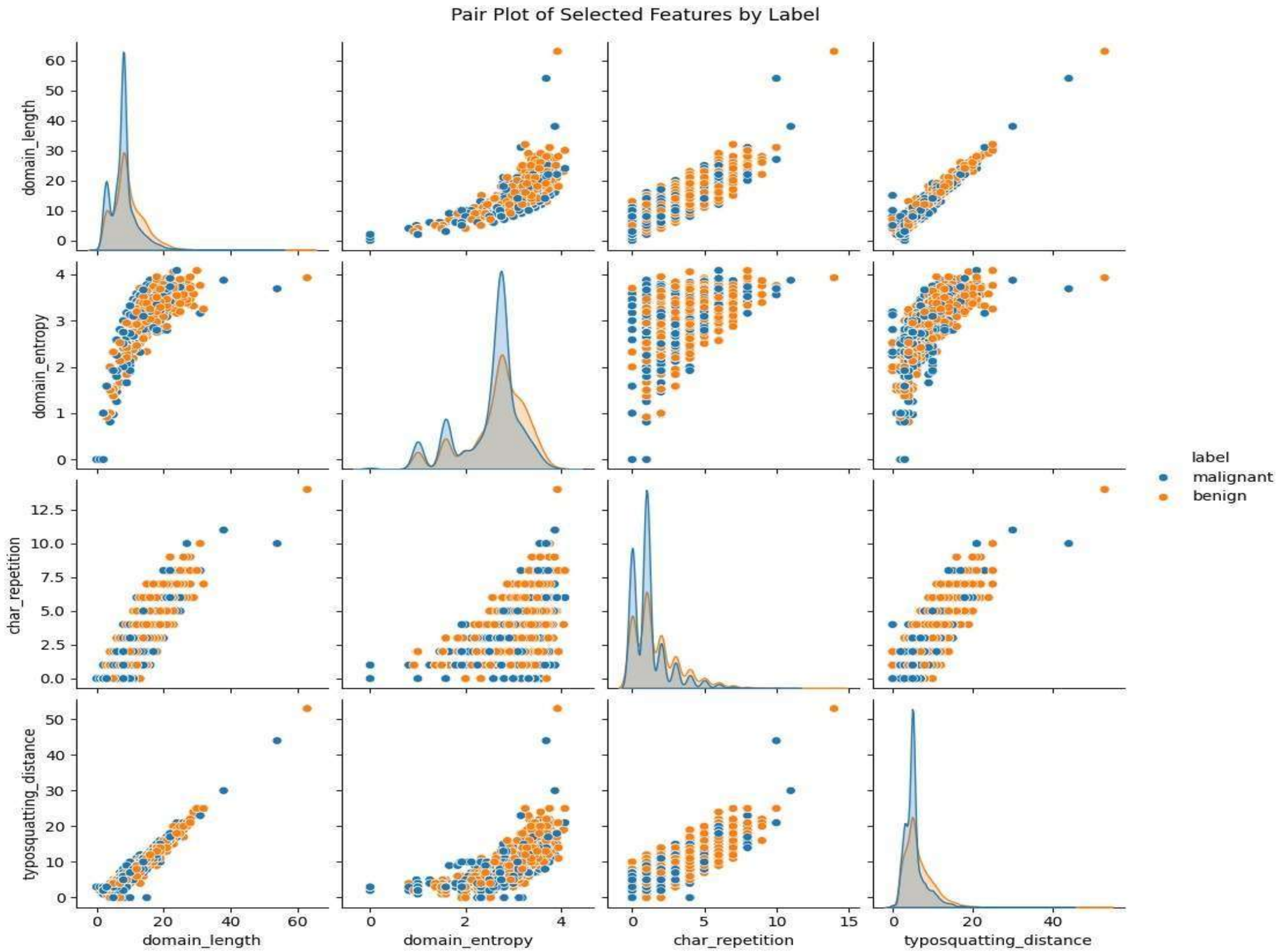
- Clean dataset by removing duplicates and invalid entries.

3

Feature Extraction and Engineering

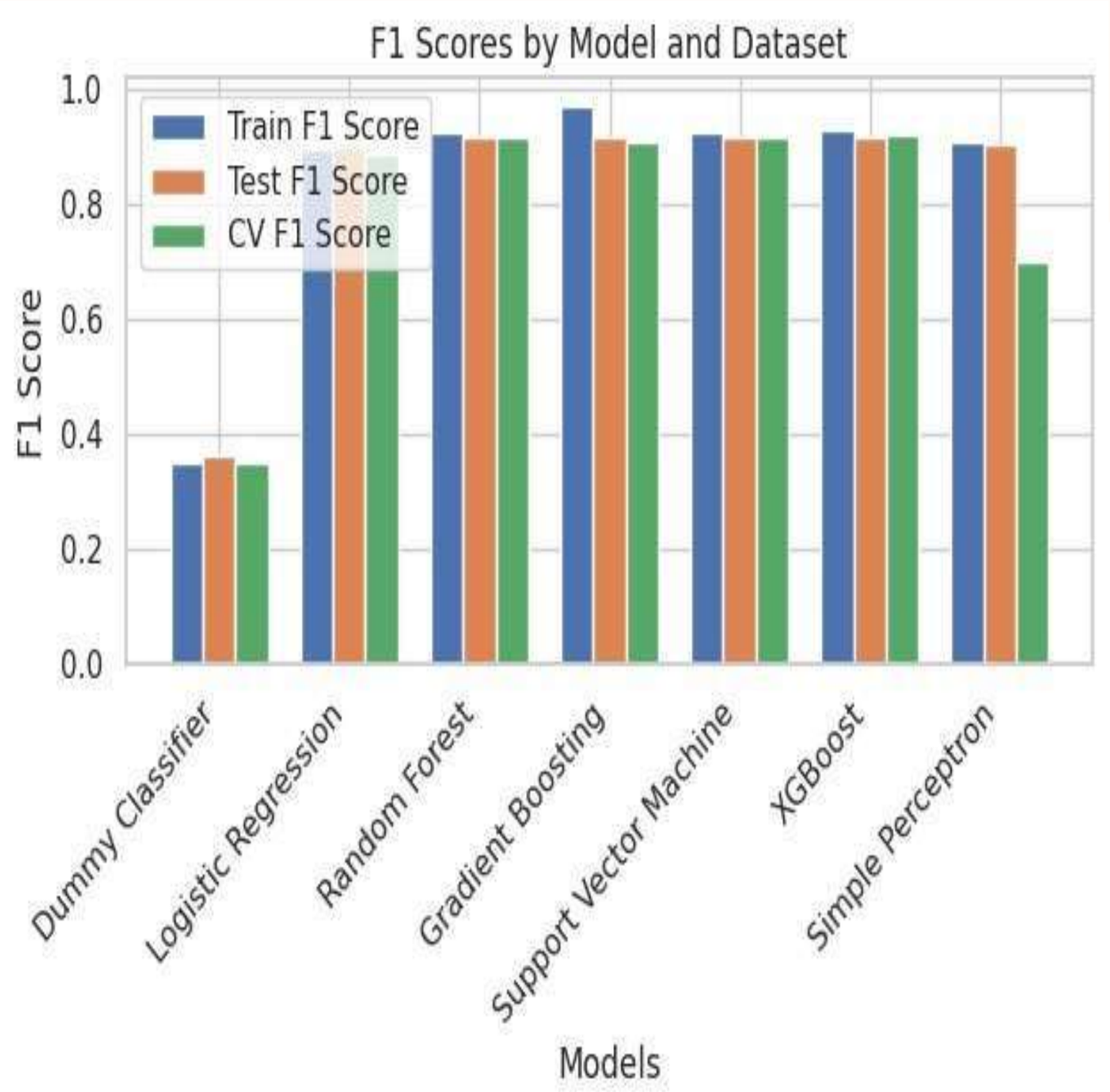
- Extract relevant URL attributes like domain age, URL length.





VISUALIZATION

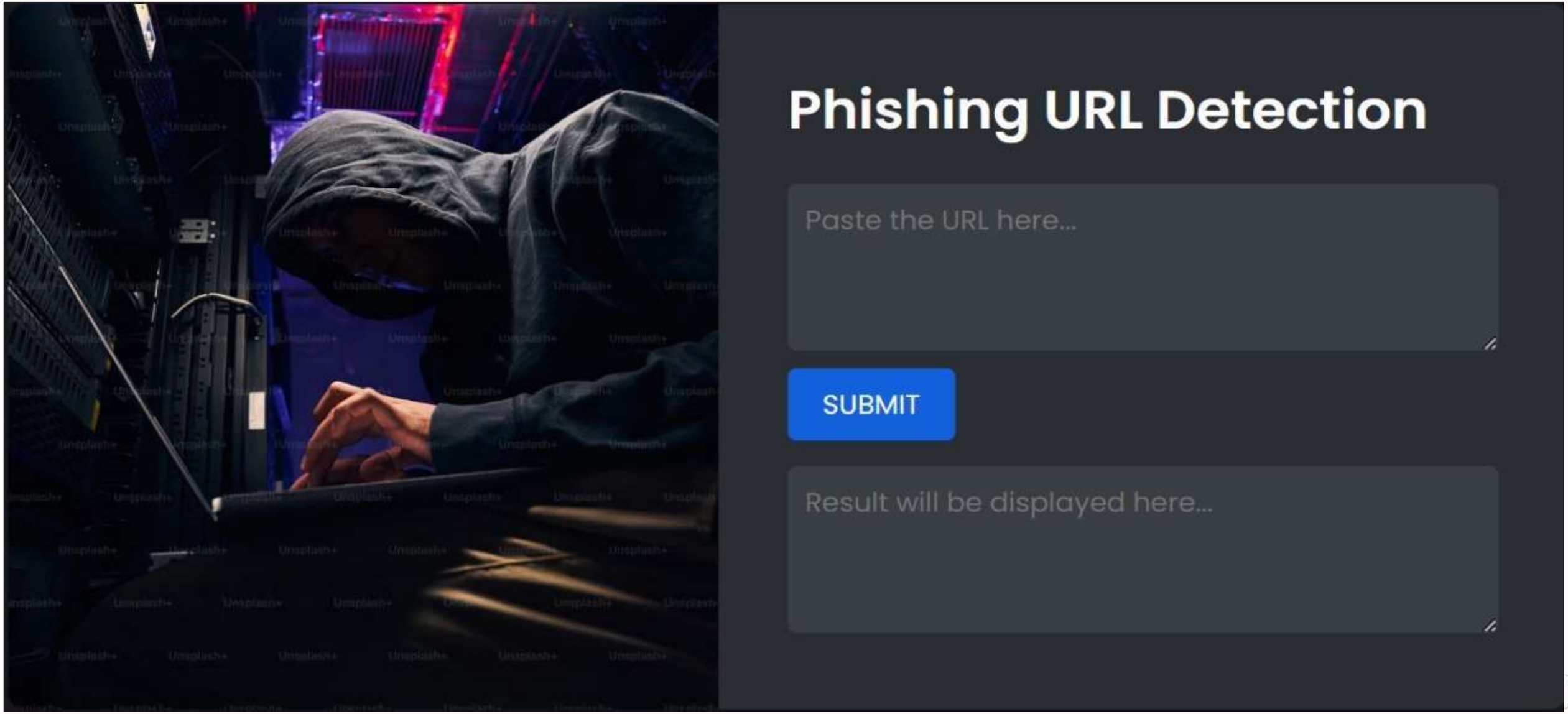
- The pair plot shows distinct feature distributions and relationships between benign and malignant labels, with noticeable patterns in domain length, entropy, character repetition, and typosquatting distance, indicating potential separation between the two classes based on these features.



Model Training and Evaluation

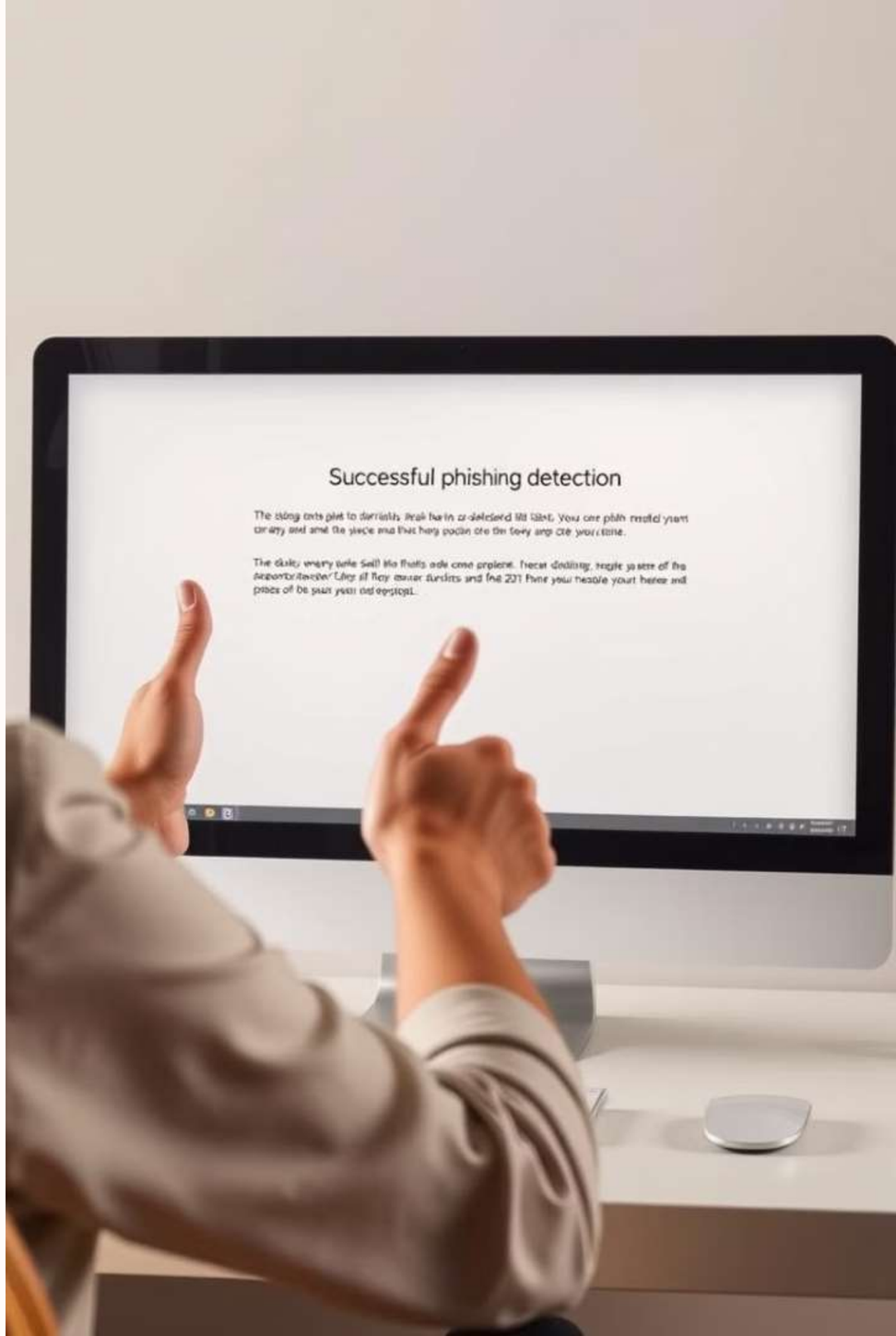
- The XGBoost and Gradient Boosting models exhibit the highest performance across all metrics, while Random Forest and Support Vector Machine also perform well;
- Logistic Regression and Simple Perceptron demonstrate good but slightly lower performance
- The Dummy Classifier shows the lowest scores, indicating it is a poor benchmark.

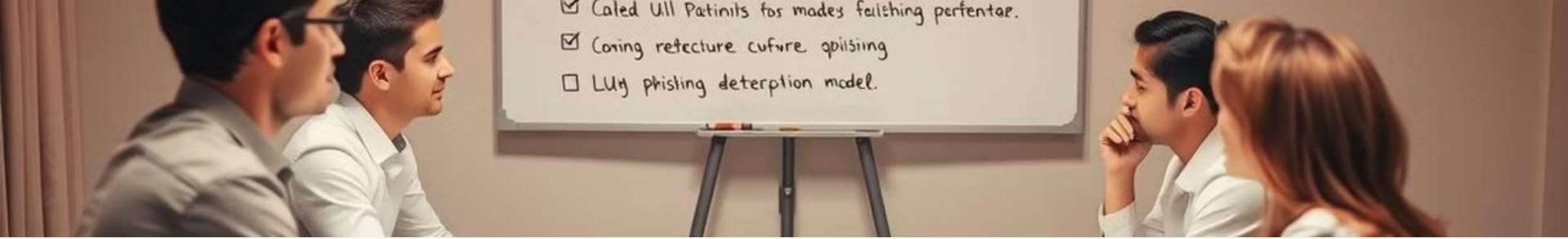
Deploying the Detector



Conclusion

- Achieved high-accuracy detection that ensures the solution is dependable and minimizes errors, meeting rigorous performance standards.
- Real-time classification reflects the goal of delivering immediate insights, allowing for quick decision-making and responsiveness in dynamic environments.
- A URL phishing detector is an effective tool for protecting individuals and organizations from phishing attacks.





Recommendations

1

Continuous Learning

- Regularly update the model with new phishing data to ensure its effectiveness.

2

User Feedback

- Gather user feedback to improve the application's usability and effectiveness.

3

Expansion

- Explore expanding the tool to detect other forms of online threats.

Q&A

Thank You