A COMPARISON OF MACHINE LEARNING ALGORITHMS ON DETECTION OF PHISHING WEBSITES

TEAM

Madhuri Vedulla - 700741625

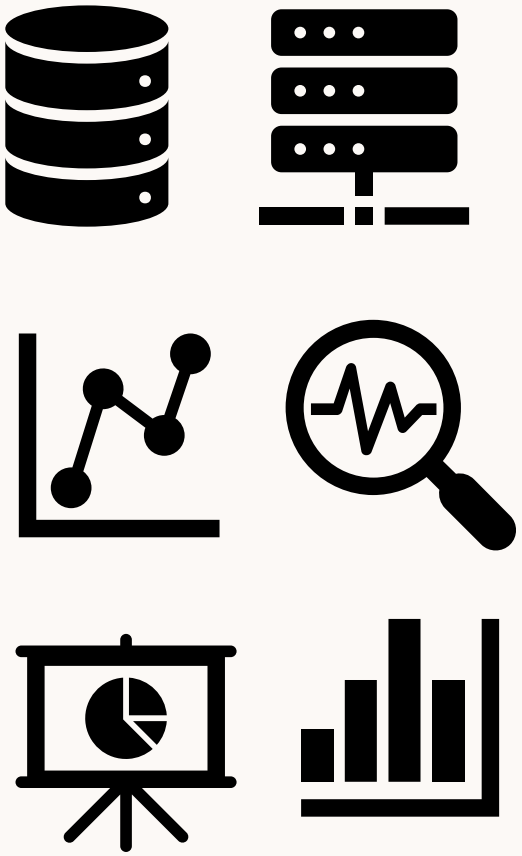
Kamala Ramesh - 700745451

Priyanka Bojja - 700739528

Sai Priyanka Narra - 700741613

CONTRIBUTION

Person	Task
Madhuri Vedulla (700741625)	Data Collection + Documentation
Kamala Ramesh (700745451)	Feature Extraction + Documentation
Priyanka Bojja (700739528)	Model Implementation (3) + Documentation
Sai Priyanka Narra (700741613)	Model Implementation (3) + Documentation



MOTIVATION

Phishing – most frequent cyber attacks

Includes pretending to be trustworthy websites to steal sensitive data like,

- Login and Password Details
- Credit Card Credentials
- Personal Information

These information are then used for identity theft and financial fraud.

Hence Phishing website detection is critical to protect personal information and prevent financial losses.

OBJECTIVE

To compare the performance of different machine learning algorithms for detecting phishing websites and evaluating them based on metrics.

Proposed Algorithms:

1. Logistic Regression
2. K-Nearest Neighbor classifier
3. Naive Bayes Classifiers
4. Random Forest classifier
5. Decision tree classifier
6. Support Vector Machine

RELATED WORK

There are several related works that discuss on the evolution of Phishing attack in the recent times

The discussions include,

- Types of Phishing
- Methods used to identify Phishing
- Lexical Structure
- Features Extraction Techniques
- Model Development
- Model Testing and Evaluation
- Results and Comparison

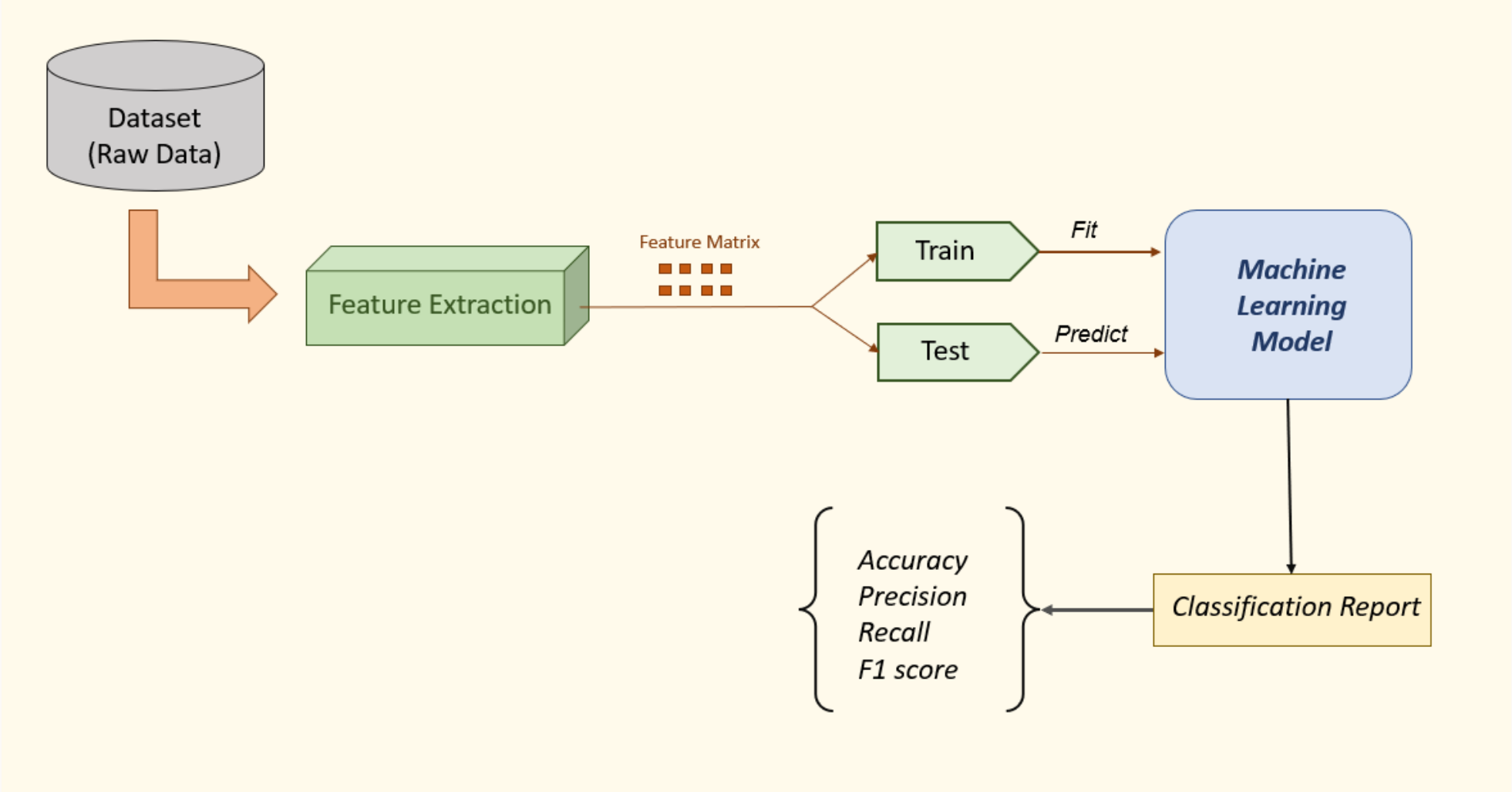
PROBLEM STATEMENT

In order to avoid getting phished,

- Users should have awareness of phishing websites.
- Have a blacklist of phishing websites which requires the knowledge of website being detected as phishing.
- Detect them in their early appearance, using machine learning and deep neural network algorithms.

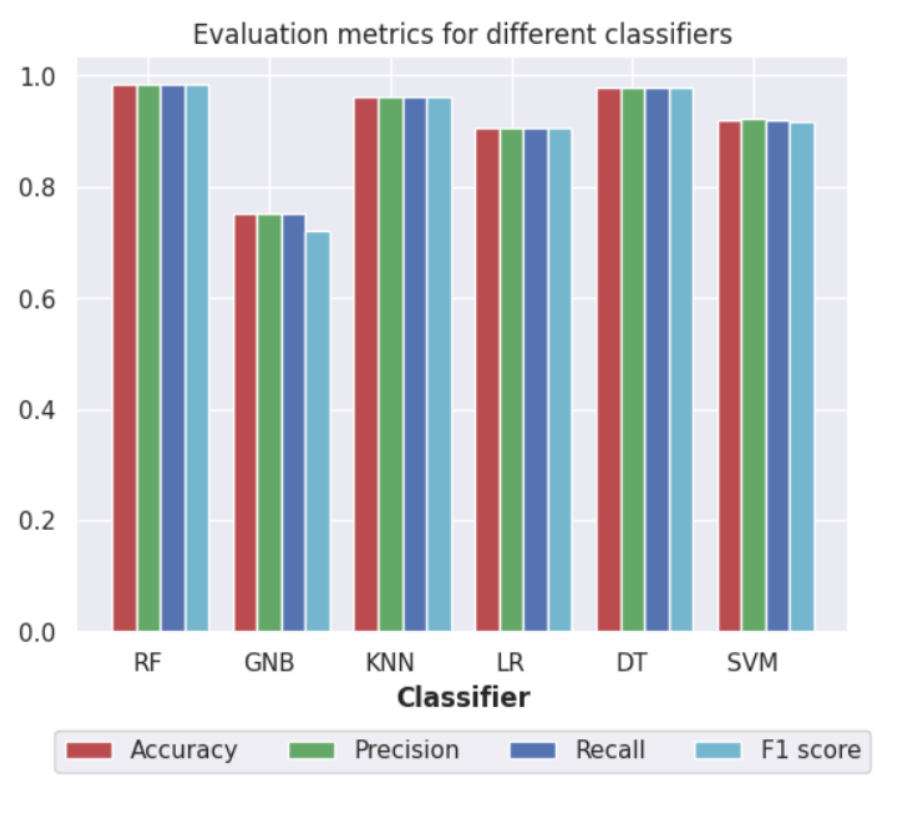
Of the above methods, the machine learning models can be trained to identify and classify phishing websites more accurately than traditional rule-based approaches

PROPOSED SOLUTION



RESULTS & SIMULATIONS

Models/Metrics	Accuracy (%)	Precision	Recall	f1-score
Random Forest	98.5	0.985	0.985	0.985
Gaussian Naïve Bayes	75.2	0.751	0.752	0.720
K Nearest Neighbor	96.1	0.961	0.961	0.961
Logistic Regression	90.7	0.907	0.907	0.906
Decision Tree	97.8	0.978	0.978	0.978
Support Vector Machines	92.0	0.922	0.920	0.918



REFERENCES

[1] M. Korkmaz, O. K. Sahingoz and B. Diri, "Detection of Phishing Websites by Using Machine Learning-Based URL Analysis," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-7, doi: 10.1109/ICCCNT49239.2020.9225561.

[2] M. H. Alkawaz, S. J. Steven, A. I. Hajamydeen and R. Ramli, "A Comprehensive Survey on Identification and Analysis of Phishing Website based on Machine Learning Methods," 2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), Penang, Malaysia, 2021, pp. 82-87, doi: 10.1109/ISCAIE51753.2021.9431794.

[3] J. Kumar, A. Santhanavijayan, B. Janet, B. Rajendran and B. S. Bindhumadhava, "Phishing Website Classification and Detection Using Machine Learning," 2020 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2020, pp. 1-6, doi: 10.1109/ICCCI48352.2020.9104161.

[4] S. Sindhu, S. P. Patil, A. Sreevalsan, F. Rahman and M. S. A. N., "Phishing Detection using Random Forest, SVM and Neural Network with Backpropagation," 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), Bengaluru, India, 2020, pp. 391-394, doi: 10.1109/ICSTCEE49637.2020.9277256.

[5] PhishTank-Friends of PhishTank," PhishTank. [Online]. Available: https://www.phishtank.com/friends.php. [Accessed: 09-Mar-2020].

[6] Marchal, S. (Creator) (2014). PhishStorm - phishing / legitimate URL dataset. Aalto University. urlset(v.zip). 10.24342/f49465b2-c68a-4182- 9171-075f0ed797d5

[7] S. Zaman, S. M. Uddin Deep, Z. Kawsar, M. Ashaduzzaman and A. I. Pritom, "Phishing Website Detection Using Effective Classifiers and Feature Selection Techniques," 2019 2nd International Conference on Innovation in Engineering and Technology (ICIET), Dhaka, Bangladesh, 2019, pp. 1-6, doi: 10.1109/ICIET48527.2019.9290554.

THANK YOU