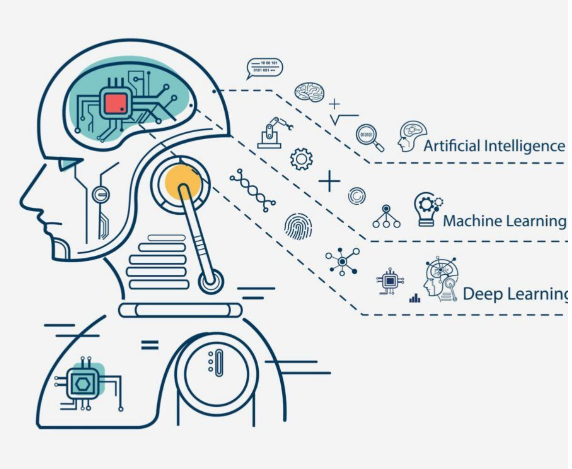


A Machine Learning-Integrated Dashboard for DNS Traffic Monitoring and Anomaly Detection

Ali Khan
Sebastian Skubisz

CS 656
Professor Moshiur Rahman

INTRODUCTION

- Evolving cyber threats exploit DNS for malicious activities such as phishing, malware, and command-and-control
- Traditional DNS monitoring methods lack sophistication for modern cyber attacks.
- Objective: Develop a data-driven dashboard integrating ML-based anomaly detection.

Reason For Choosing This Topic

- Anomaly detection is crucial across diverse fields like finance, networking, and healthcare, where identifying irregularities can prevent fraud, optimize systems, and save lives.
- Manual analysis and static rules are not sufficient or sophisticated enough to analyse the ever evolving landscape of cyber attacks
- We believe using machine learning is a more sophisticated approach to monitoring DNS traffic as well as mitigating cyber attacks
- Machine learning can help us look for anomalies in a more efficient manner

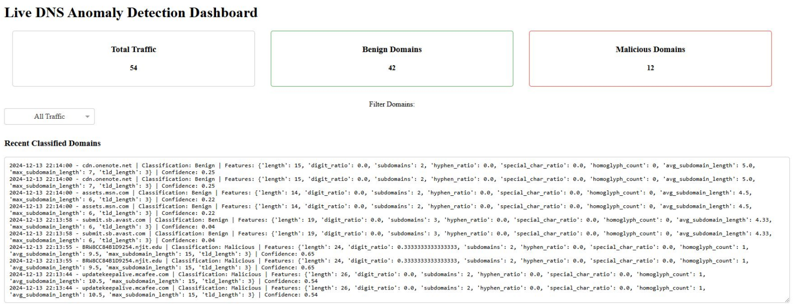
Research and Methodology

Dashboard: Python-based dashboard for real-time DNS monitoring.

- Filters by benign/malicious traffic with dynamic updates.

ML Models:

- Supervised:
  - Neural Network: High accuracy, trained on balanced datasets with RFE optimization.
  - Random Forest: Reliable with feature importance insights.
  - XGBoost: Focused on improving performance for imbalanced data.
  - KNN: Baseline model with limited scalability for high-dimensional data.
- **Unsupervised**
  - Isolation Forest: Flags anomalies by identifying outliers in feature space.
  - Local Outlier Factor (LOF): Detects density-based irregularities.
  - Autoencoder: Identifies anomalies with high reconstruction errors.



Live DNS Anomaly Detection Dashboard

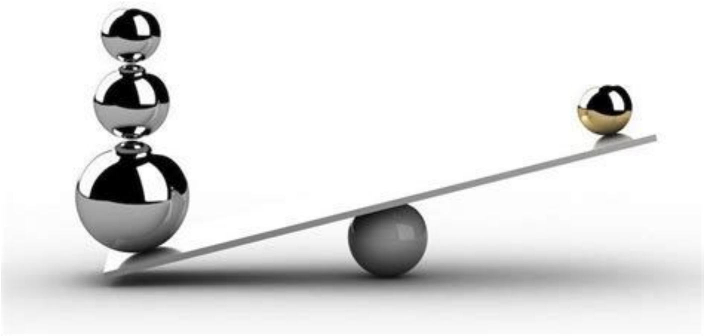
Total Traffic	Benign Domains	Malicious Domains
54	42	12

Recent Classified Domains

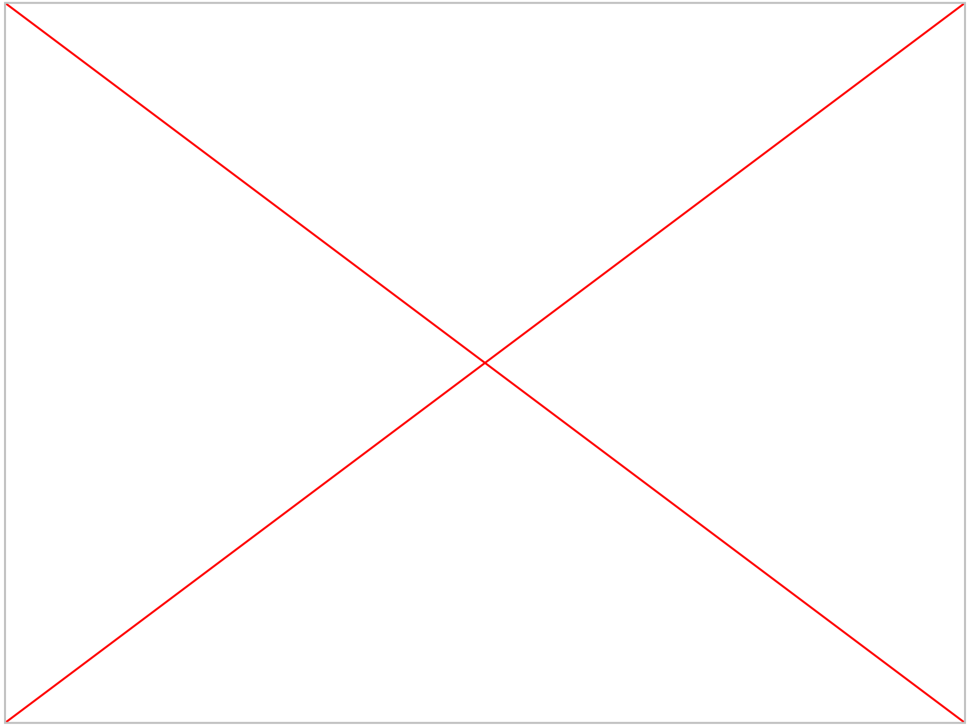
```
2024-12-13 22:14:00 - cdn.onenote.net | Classification: Benign | Features: {'length': 15, 'digit_ratio': 0.0, 'subdomains': 2, 'hyphen_ratio': 0.0, 'special_char_ratio': 0.0, 'homoglyph_count': 0, 'avg_subdomain_length': 5.0,
'max_subdomain_length': 7, 'tld_length': 3} | Confidence: 0.25
2024-12-13 22:14:00 - cdn.onenote.net | Classification: Benign | Features: {'length': 15, 'digit_ratio': 0.0, 'subdomains': 2, 'hyphen_ratio': 0.0, 'special_char_ratio': 0.0, 'homoglyph_count': 0, 'avg_subdomain_length': 5.0,
'max_subdomain_length': 7, 'tld_length': 3} | Confidence: 0.25
2024-12-13 22:14:00 - assets.msn.com | Classification: Benign | Features: {'length': 14, 'digit_ratio': 0.0, 'subdomains': 2, 'hyphen_ratio': 0.0, 'special_char_ratio': 0.0, 'homoglyph_count': 0, 'avg_subdomain_length': 4.5,
'max_subdomain_length': 6, 'tld_length': 3} | Confidence: 0.22
2024-12-13 22:14:00 - assets.msn.com | Classification: Benign | Features: {'length': 14, 'digit_ratio': 0.0, 'subdomains': 2, 'hyphen_ratio': 0.0, 'special_char_ratio': 0.0, 'homoglyph_count': 0, 'avg_subdomain_length': 4.5,
'max_subdomain_length': 6, 'tld_length': 3} | Confidence: 0.22
2024-12-13 22:13:58 - submit.sb.avast.com | Classification: Benign | Features: {'length': 19, 'digit_ratio': 0.0, 'subdomains': 3, 'hyphen_ratio': 0.0, 'special_char_ratio': 0.0, 'homoglyph_count': 0, 'avg_subdomain_length': 4.33,
'max_subdomain_length': 6, 'tld_length': 3} | Confidence: 0.04
2024-12-13 22:13:58 - submit.sb.avast.com | Classification: Benign | Features: {'length': 19, 'digit_ratio': 0.0, 'subdomains': 3, 'hyphen_ratio': 0.0, 'special_char_ratio': 0.0, 'homoglyph_count': 0, 'avg_subdomain_length': 4.33,
'max_subdomain_length': 6, 'tld_length': 3} | Confidence: 0.04
2024-12-13 22:13:55 - BRW8CC84B1D9254.njit.edu | Classification: Malicious | Features: {'length': 24, 'digit_ratio': 0.3333333333333333, 'subdomains': 2, 'hyphen_ratio': 0.0, 'special_char_ratio': 0.0, 'homoglyph_count': 1,
'avg_subdomain_length': 9.5, 'max_subdomain_length': 15, 'tld_length': 3} | Confidence: 0.65
2024-12-13 22:13:55 - BRW8CC84B1D9254.njit.edu | Classification: Malicious | Features: {'length': 24, 'digit_ratio': 0.3333333333333333, 'subdomains': 2, 'hyphen_ratio': 0.0, 'special_char_ratio': 0.0, 'homoglyph_count': 1,
'avg_subdomain_length': 9.5, 'max_subdomain_length': 15, 'tld_length': 3} | Confidence: 0.65
2024-12-13 22:13:44 - updatekeepalive.mcafee.com | Classification: Malicious | Features: {'length': 26, 'digit_ratio': 0.0, 'subdomains': 2, 'hyphen_ratio': 0.0, 'special_char_ratio': 0.0, 'homoglyph_count': 1,
'avg_subdomain_length': 10.5, 'max_subdomain_length': 15, 'tld_length': 3} | Confidence: 0.54
2024-12-13 22:13:44 - updatekeepalive.mcafee.com | Classification: Malicious | Features: {'length': 26, 'digit_ratio': 0.0, 'subdomains': 2, 'hyphen_ratio': 0.0, 'special_char_ratio': 0.0, 'homoglyph_count': 1,
'avg_subdomain_length': 10.5, 'max_subdomain_length': 15, 'tld_length': 3} | Confidence: 0.54
```

Dataset Explanation

- **Benign Domains**: Public datasets like Alexa's top websites.
- **Malicious Domains**: Sources like ViriBack and PhishTank.
- **Features Extracted**: Domain length, subdomains,special characters, and digit ratios.
- Dataset imbalance addressed using SMOTE for balanced training.



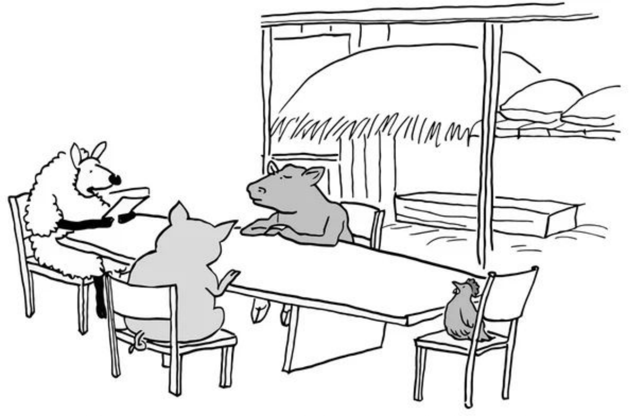
Recorded Demonstration



Future Scope

- Continuously train models on new malicious data to stay updated and accurate.
- Try to handle zero-day threats with real-time data processing.
- Use adaptive models to manage evolving traffic patterns.
- Enable continuous learning to tackle emerging cyber threats.
- Combat malicious domain generation algorithms effectively.
- Enhance visuals, like threat maps, for clearer insights and rapid response.

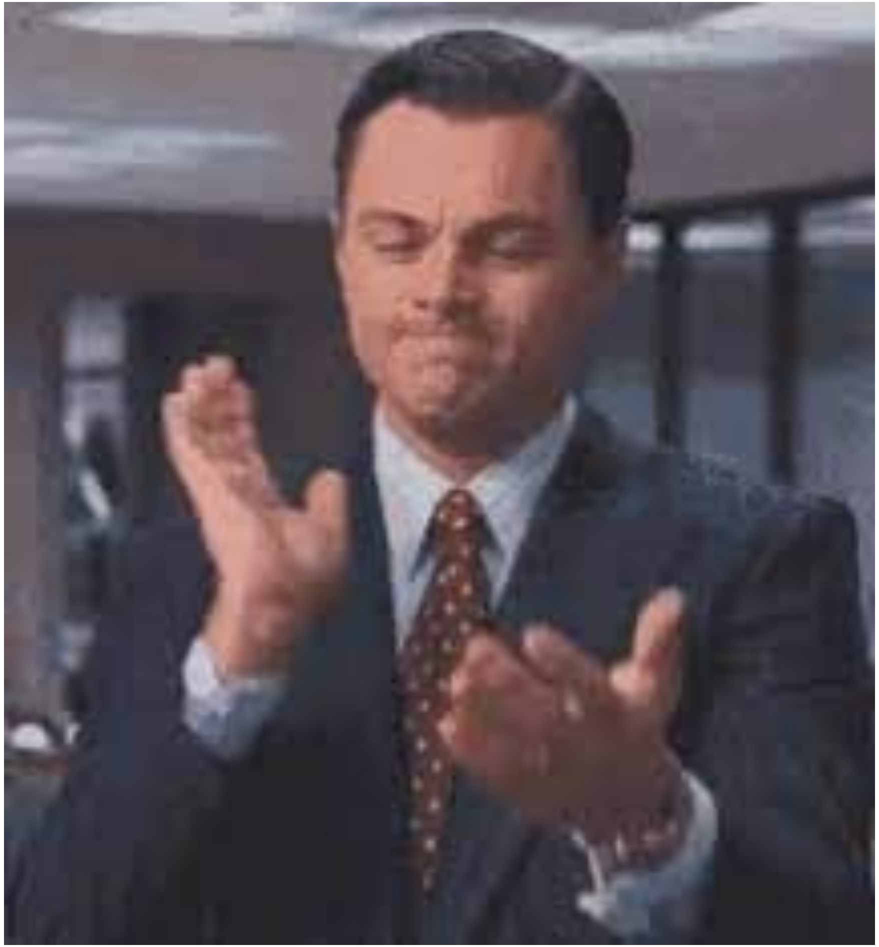
Conclusion



“The cow mooed, the pig oinked, the chicken clucked, I baaed and then we adjourned.”

- Machine learning enhances DNS monitoring and anomaly detection.
- Anomaly detection is essential for identifying irregularities across industries, enhancing security, decision-making, and efficiency.
- While challenges like imbalanced data, false alarms, and scalability remain, advancements in deep learning, Explainable AI, and edge computing promise to improve adaptability and real-time detection.
- Mastering these techniques ensures robust data analysis and proactive problem-solving.

THANK YOU FOR LISTENING



References

[1] DataCamp, "An Introduction to Anomaly Detection," [Online]. Available: https://www.datacamp.com/tutorial/introduction-to-anomaly-detection. [Accessed: Dec. 13, 2024].

[2] GeeksforGeeks, "Anomaly Detection using R," [Online]. Available: https://www.geeksforgeeks.org/anomaly-detection-using-r/. [Accessed: Dec. 13, 2024].

[3] Y. Jin, K. Kakoi, N. Yamai, N. Kitagawa, and M. Tomoishi, "A Client Based Anomaly Traffic Detection and Blocking Mechanism by Monitoring DNS Name Resolution with User Alerting Feature," *2018 International Conference on Cyberworlds (CW)*, Singapore, 2018, pp. 351-356, doi: 10.1109/CW.2018.00070.

[4] Z. Wang and M. Zhang, "The Research of DNS Anomaly Detection Based on the Method of Similarity and Entropy," *2010 International Conference on Intelligent Computation Technology and Automation*, Changsha, China, 2010, pp. 905-909, doi: 10.1109/ICICTA.2010.442.