SMART INDIA HACKATHON 2024

- Problem Statement ID – 1681
- Problem Statement Title- Identification of algorithm from the given dataset using AI/ML Techniques
- Theme- Blockchain & Cybersecurity
- PS Category- Software
- Team ID - 20503
- Team Name – Cipher_XO

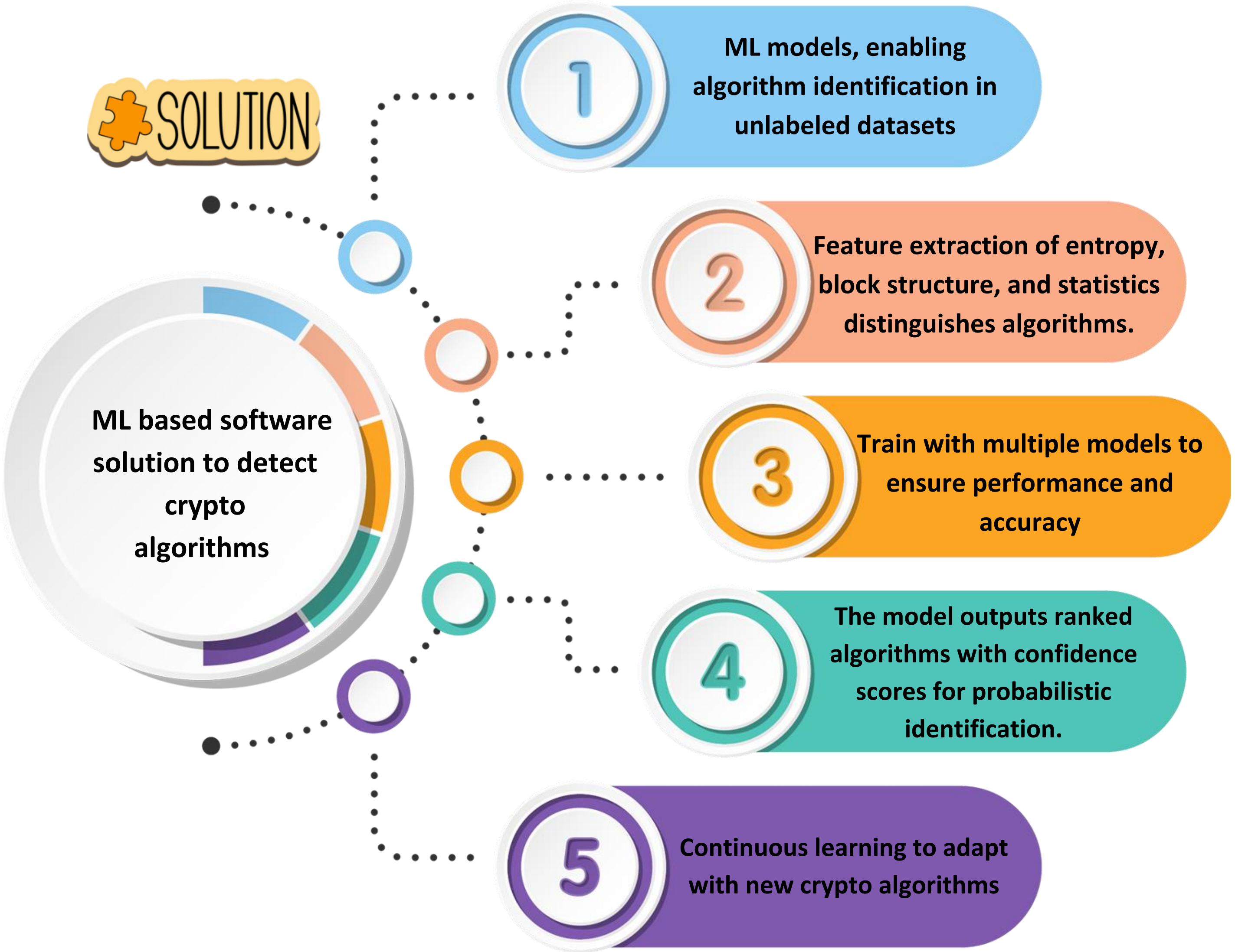
Cipher_XO

ALGO DECRYPT- AI/ML Based Cryptanalyzer

SMART INDIA HACKATHON 2024

ISSUE

- Unknown Security Strength
- Non-Compliance with Security Standards
- Lack of Data Integrity
- Over-reliance on Hidden Algorithms
- Inability to Optimize Performance
- Compatibility and Interoperability Issues



SOLUTION

ML based software solution to detect crypto algorithms

1. ML models, enabling algorithm identification in unlabeled datasets
2. Feature extraction of entropy, block structure, and statistics distinguishes algorithms.
3. Train with multiple models to ensure performance and accuracy
4. The model outputs ranked algorithms with confidence scores for probabilistic identification.
5. Continuous learning to adapt with new crypto algorithms

UNIQUE

Multi-model Approach: For better accuracy and robustness

Algorithm Rating: Rate algorithm on the basis of security strength

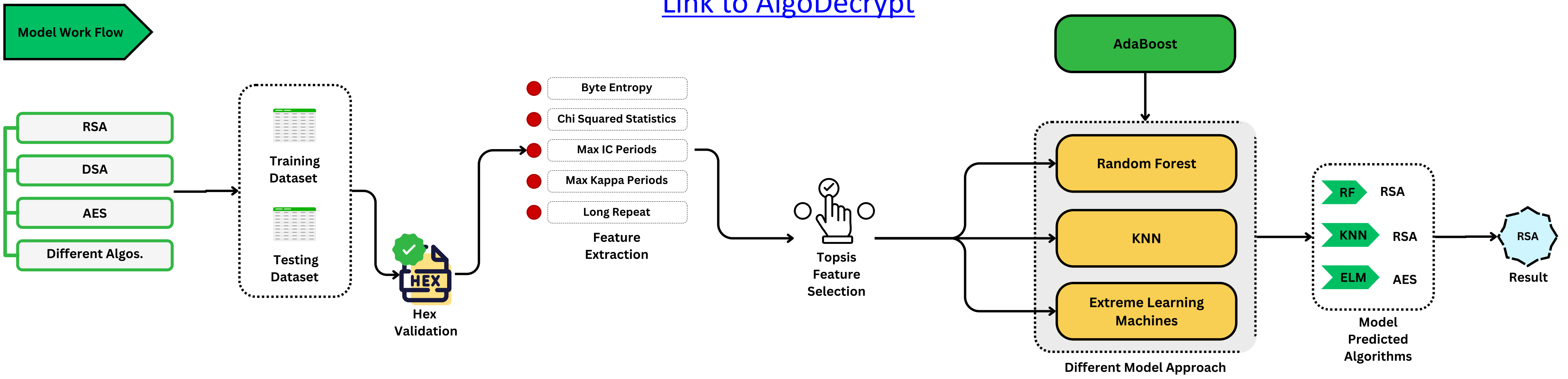
Checking compliance requirements: Checking against compliance requirements

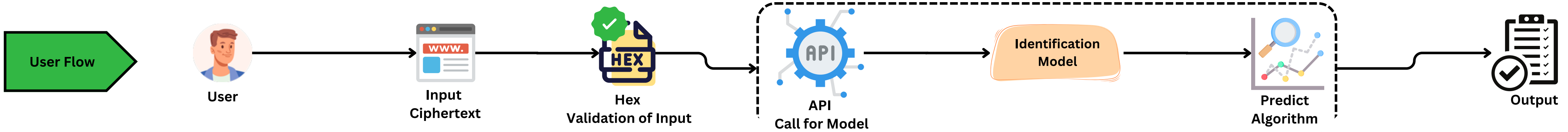
Advanced feature extraction: To ensure detailed approach for algorithm detection

Automation: For making it accessible and user efficient

TECHNICAL APPROACH

[Link to AlgoDecrypt]





Tech stack :

NEXT.JS

JS

pandas

seaborn

scikit learn

tailwindcss

Flask

Cipher_XO

FEASIBILITY AND VIABILITY

SMART INDIA HACKATHON 2024

Feature extraction from ciphertext is technically feasible using statistical and cryptanalysis techniques

High demand in industries like cybersecurity, forensics, and compliance ensures practical value

Encryption datasets can be synthetically generated using known algorithms

Ensuring the system is used for legitimate purposes prevents misuse

Growing demand for encryption auditing and security solutions across industries

Challenges:

- Similar ciphertext patterns
- System evolution needed for new algorithms
- Dataset quality
- High computational cost for large datasets
- Vulnerability to adversarial attacks

Overcoming Challenges:

- Robust feature engineering
- Availability of framework will make it easy
- Synthetic dataset generation and industry collaboration
- Using model optimization techniques
- Adversarial training to strengthen model robustness

IMPACT AND BENEFITS

Impact

- The solution can efficiently identify encryption algorithms, helping detect vulnerabilities in encrypted communications, thereby strengthening cybersecurity defenses
- The tool will help to enhance security by identifying cryptographic algorithms used in encryption, specifically in defense, medical and security service providers like McAfee
- Automating algorithm identification reduces reliance on manual cryptanalysis, lowering operational costs for organizations in need of cryptographic assessments
- By leveraging AI/ML techniques, our cryptanalyzer significantly reduces the time required for decryption and analysis, making it suitable for real-time threat detection
- The platform will help researchers to find cryptographic vulnerabilities and the development of more secure encryption techniques

Benefits

Benefits

- Identifying weak encryption prevents vulnerabilities, securing personal and financial data
- Early detection of weak encryption prevents costly breaches, fines, and damage.
- Companies providing encryption security solutions will likely see significant market growth
- Using energy-efficient encryption reduces data center, mobile, and IoT energy consumption
- Efficient encryption reduces computational load, extending hardware lifespan and minimizing e-waste.

Cipher_XO

RESEARCH AND REFERENCES

SMART INDIA HACKATHON 2024

Research

System Architecture: The research proposes an identification model using SVM classifiers to recognize cryptographic algorithms from ciphertext. However, the feature extraction process is vaguely defined, leading to uncertainty in the model's training and overall reliability

Feature Extraction: Cipher features are extracted from ciphertext, but the research fails to define a standard or effective feature set. This lack of clarity in feature extraction weakens the model's robustness, as the selection of irrelevant or inconsistent features can degrade performance

Evaluation Metrics: The system performs well with larger ciphertexts (100KB or more), achieving high identification accuracy. However, the model struggles with smaller or irregular ciphertext sizes, reducing its practicality in real-world cryptanalysis where ciphertext sizes vary significantly

References

An Approach to Identifying Cryptographic Algorithm from Ciphertext Link to paper

A block cipher algorithm identification scheme based on hybrid k-nearest neighbor and random forest algorithm Link to paper

Cryptographic algorithms: A review of the literature, weaknesses and open challenges Link to paper

Breaking an unknown cipher Link to paper

Machine Learning for Cryptographic Algorithm Identification Link to paper

Performance evaluation of cryptographic algorithms Link to paper