SMART INDIA HACKATHON 2025

TITLE PAGE

- **Problem Statement ID – 25159**
- **Problem Statement Title-** Real-time AI/ML-Based Phishing Detection and Prevention System
- **Theme-** Blockchain & Cybersecurity
- **PS Category- Software**
- **Team ID-**
- **Team Name –** Veritas (The Latin word for "Truth")

Veritas

Aegis-AI: A Real-time, Multi-Modal Phishing Defense Framework

We propose **Aegis-AI**, a real-time, adaptive framework that integrates into browsers and email clients to detect and block sophisticated phishing attacks before they can cause harm.

Key Innovations:

- **Multi-Modal Analysis:** Goes beyond just links. Analyzes text (NLP), webpage visuals (CNNs), and domain relationships (GNNs) for a complete threat picture.
- **Real-time Edge Integration:** A lightweight browser extension provides instant, sub-50ms protection directly on the user's device, blocking threats before interaction.
- **Continuous Learning Pipeline:** Automatically learns from new threats and user feedback, ensuring it is always ready for zero-day attacks without manual updates.

Veritas

TECHNICAL APPROACH

SMART INDIA HACKATHON 2025

Technologies to be Used:

- **Backend:** Python (Flask/FastAPI), Docker, Kubernetes
- **ML/DL Frameworks:** PyTorch / TensorFlow, Hugging Face Transformers
- **Core Models:Text:** BERT / RoBERTa
- **Visual:** Convolutional Neural Networks (CNNs)
- **Domain:** Graph Neural Networks (GNNs)
- **Frontend:** JavaScript (for Browser Extension)

Methodology & Process Flow:

- **Input:** An email, SMS, or URL is received in real-time.
- **Parallel Analysis:**
- **Text Engine (NLP):** Scans for suspicious language and intent.
- **Link Engine (GNN):** Analyzes domain reputation, SSL history, and hidden connections.
- **Visual Engine (CNN):** Renders webpage in a sandbox to detect brand impersonation.
- **Risk Scoring:** An aggregator model combines signals to produce a final threat score.
- **Action:** If the score is high, the content is **instantly blocked**, and an alert is shown.
- **Feedback Loop:** Results are fed back into our continuous learning pipeline to improve the model.

Veritas

FEASIBILITY AND VIABILITY

Feasibility:
Our solution is highly feasible, built upon state-of-the-art yet proven AI models (Transformers, GNNs, CNNs) that are well-documented and have strong community support. The microservice architecture ensures scalability.

Potential Challenges	Our Mitigation Strategy
Performance Overhead	Use lightweight models (e.g., DistilBERT) & model quantization for fast edge inference.
Adversarial Attacks	Employ adversarial training to make models robust against evasion techniques.
False Positives	Implement a simple user feedback system to continuously fine-tune the model's accuracy.
Data Privacy	Prioritize on-device processing; use data anonymization for any cloud-based analysis.

Veritas

IMPACT AND BENEFITS

SMART INDIA HACKATHON 2025

Potential Impact:
Dramatically reduces the success rate of phishing attacks, safeguarding financial assets, personal credentials, and sensitive data for millions of users across education, government, and industry.

Benefits of the Solution:

- **Social:**
- Creates a safer and more trustworthy digital environment for all users.
- Protects vulnerable individuals from online fraud and scams.
- **Economic:**
- Prevents billions in losses from data breaches, ransomware, and credential theft.
- Reduces the high operational costs of cybersecurity incident response.
- **Technical:**
- Delivers proactive, real-time defense against zero-day threats.
- Vastly superior to traditional, slow, signature-based anti-phishing tools.

Veritas

RESEARCH AND REFERENCES

SMART INDIA HACKATHON 2025

Our approach is grounded in established academic and industry research:

1. Textual Analysis (Transformers):

1. "An Explainable Transformer-based Model for Phishing Email Detection: A Large Language Model Approach" (arXiv, 2024)
2. "URLTran: Improving Phishing URL Detection Using Transformers" (IEEE, 2021)

2. Graph-Based Analysis (GNNs):

1. "AGCN-Domain: Detecting Malicious Domains with Graph Convolutional Network and Attention Mechanism" (MDPI, 2024)

3. Visual Analysis (CNNs):

1. "Inferring Phishing Intention via Webpage Appearance and Dynamics: A Deep Vision Based Approach" (USENIX Security Symposium, 2022)
2. "Spotting brand impersonation with Swin transformers and Siamese neural networks" (Microsoft Security Blog, 2021)