VISVESVARAYA TECHNOLOGICAL UNIVERSITY
Belagavi - 590 018

BCD586 – Mini Project

PHISHGAURD:AI-Powered Phishing Email Detection

Presented By:
Akhina(1AY23CD007)
Jhashank(1AY23CD027)
Kavana(1AY23CD030)
Liya(1AY23CD033)

Under the Guidance of
Prof. Mohammad Tahir Mirji
Assistant Professor

Content

Abstract

- Traditional email filters fail against evolving phishing attacks that exploit human trust
- AI-powered detection using BiLSTM with attention mechanism and feature engineering via Flask web interface
- Provides protection with high accuracy, safeguarding users from data breaches and financial loss

Problem statement

- Email phishing is a widespread cyber threat impacting millions globally
- Traditional rule-based filters often fail to detect advanced phishing tactics
- Need for intelligent, adaptable detection solutions

Introduction

- Phishing attacks exploit user trust via deceptive emails
- Increasing sophistication of attacks requires advanced defense
- Motivation: Protect users from financial and data loss
- Valid and urgent in the current digital era Scope includes individuals, businesses, and organizations

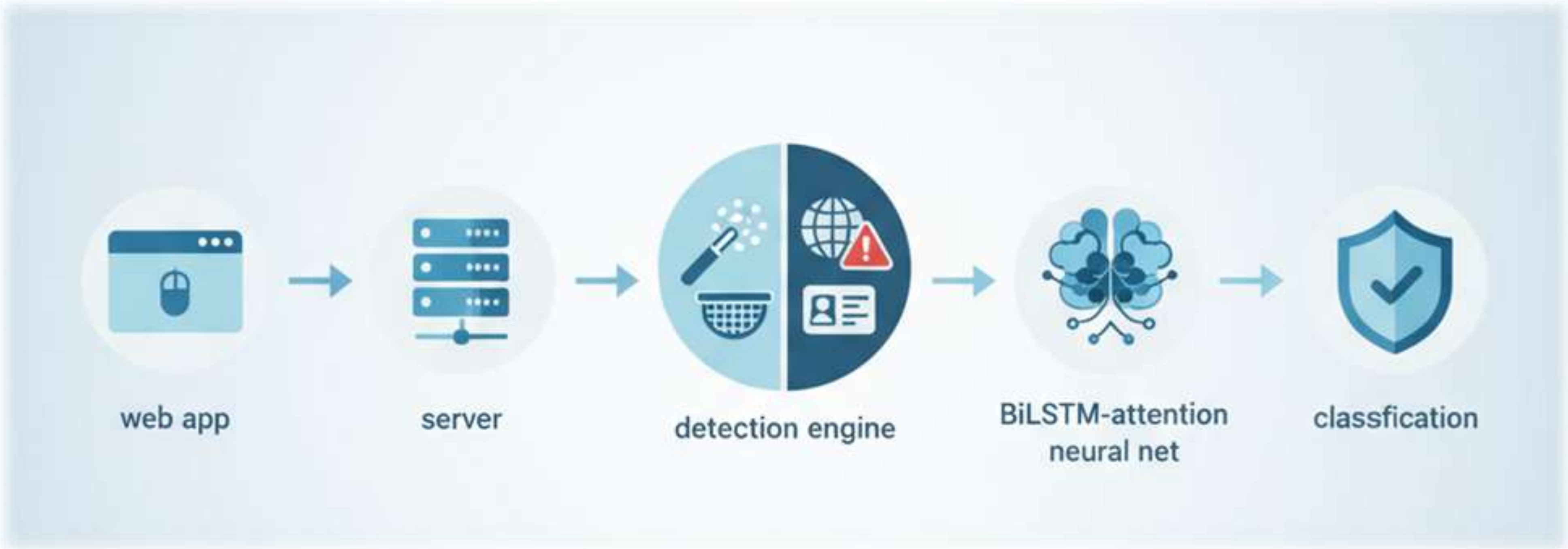
Objective

- Build an effective, AI-powered email phishing detection system
- Ensure high accuracy and low false positives in scam detection

Literature Survey

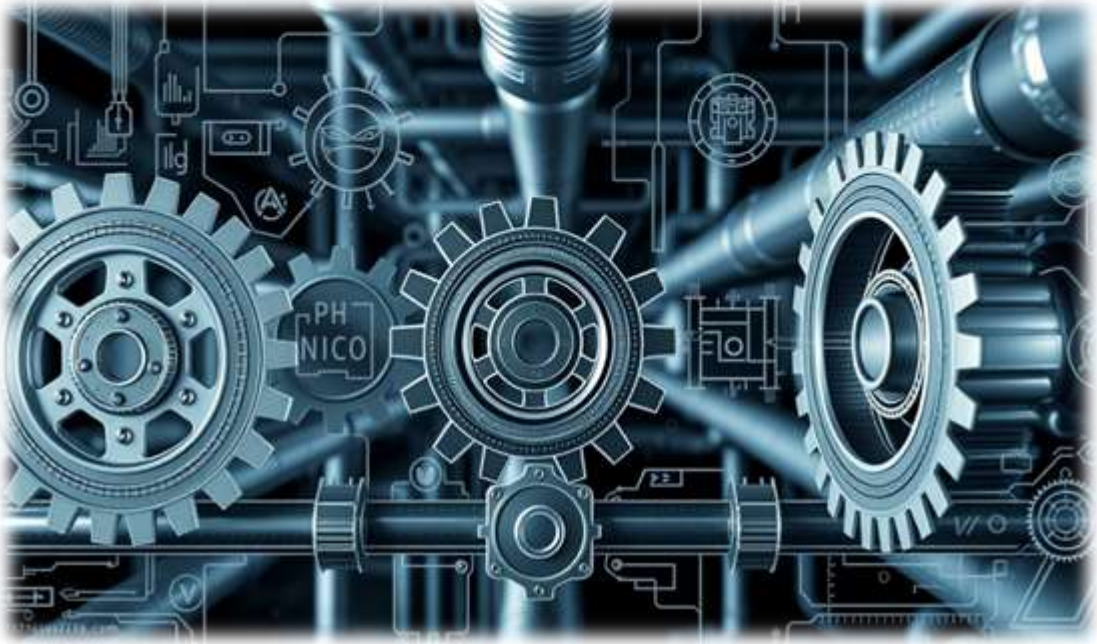
Sl No.	Citations	Methodologies	Research Gaps
01	Literature Survey — “Phishing Detection: A Literature Survey” (survey/review).	Systematic literature review of detection approaches (rule/header checks, heuristics, ML, DL, user studies, system mitigations)	“No standardized datasets, cross-domain benchmarks, or defenses for AI-generated phishing.”
02	AdaPhish: AI-Powered Adaptive Defense and Education Resource Against Deceptive Emails — Meguro & Chong.	System/platform design: LLM-based automated anonymization + vector DB for phishing “phish-bowl”; real-time detection + adaptive reporting; integrates education modules	“Needs large-scale validation, stronger privacy controls, and resistance to evolving LLM phishing.”
03	AI-Powered Phishing Detection: A Data-Driven Cyber-security Approach (conference paper).	Data-driven detection — likely ML/DL classifiers and empirical evaluation; (conference paper format)	“Lacks cross-dataset testing, strong generalization, and protection from advanced AI phishing.”

System Architecture/ Flow Diagram



Methodology

- Collect Enron fraud dataset for training
- Preprocess email text, build vocabulary
- Train deep learning model with class balancing, regularization
- Deploy Flask-based web interface for real-time detection
- Use thresholds to optimize F1-score

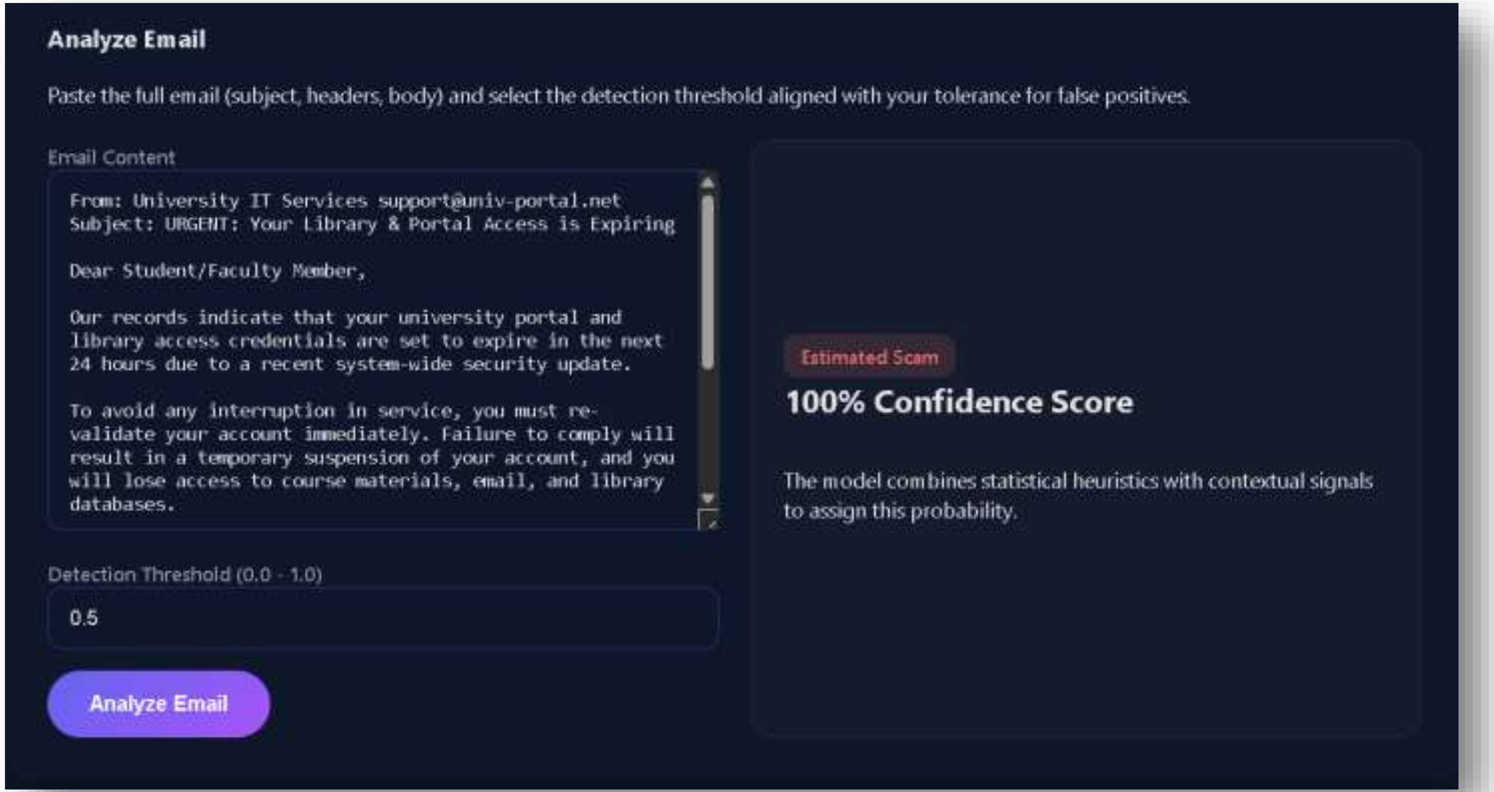


System Testing

Sample Email Input	Classification	Confidence Score
"Urgent! Your account will be suspended. Click here to verify your password immediately."	Phishing	94.7%
"Hi team, please find attached the quarterly report for review. Let me know if you have questions."	Legitimate	21.2%
"Congratulations! You've won $1,000,000. Send your bank details to claim your prize now!"	Phishing	98.3%
"Meeting scheduled for tomorrow at 3 PM. Agenda attached. See you there."	Legitimate	49.5%
"Your payment is overdue. Click this link to avoid legal action within 24 hours."	Phishing	96.1%

System Testing

- Example test cases for email classification



Analyze Email

Paste the full email (subject, headers, body) and select the detection threshold aligned with your tolerance for false positives.

Email Content

```
From: University IT Services support@univ-portal.net
Subject: URGENT: Your Library & Portal Access is Expiring

Dear Student/Faculty Member,

Our records indicate that your university portal and
library access credentials are set to expire in the next
24 hours due to a recent system-wide security update.

To avoid any interruption in service, you must re-
validate your account immediately. Failure to comply will
result in a temporary suspension of your account, and you
will lose access to course materials, email, and library
databases.
```

Detection Threshold (0.0 - 1.0)

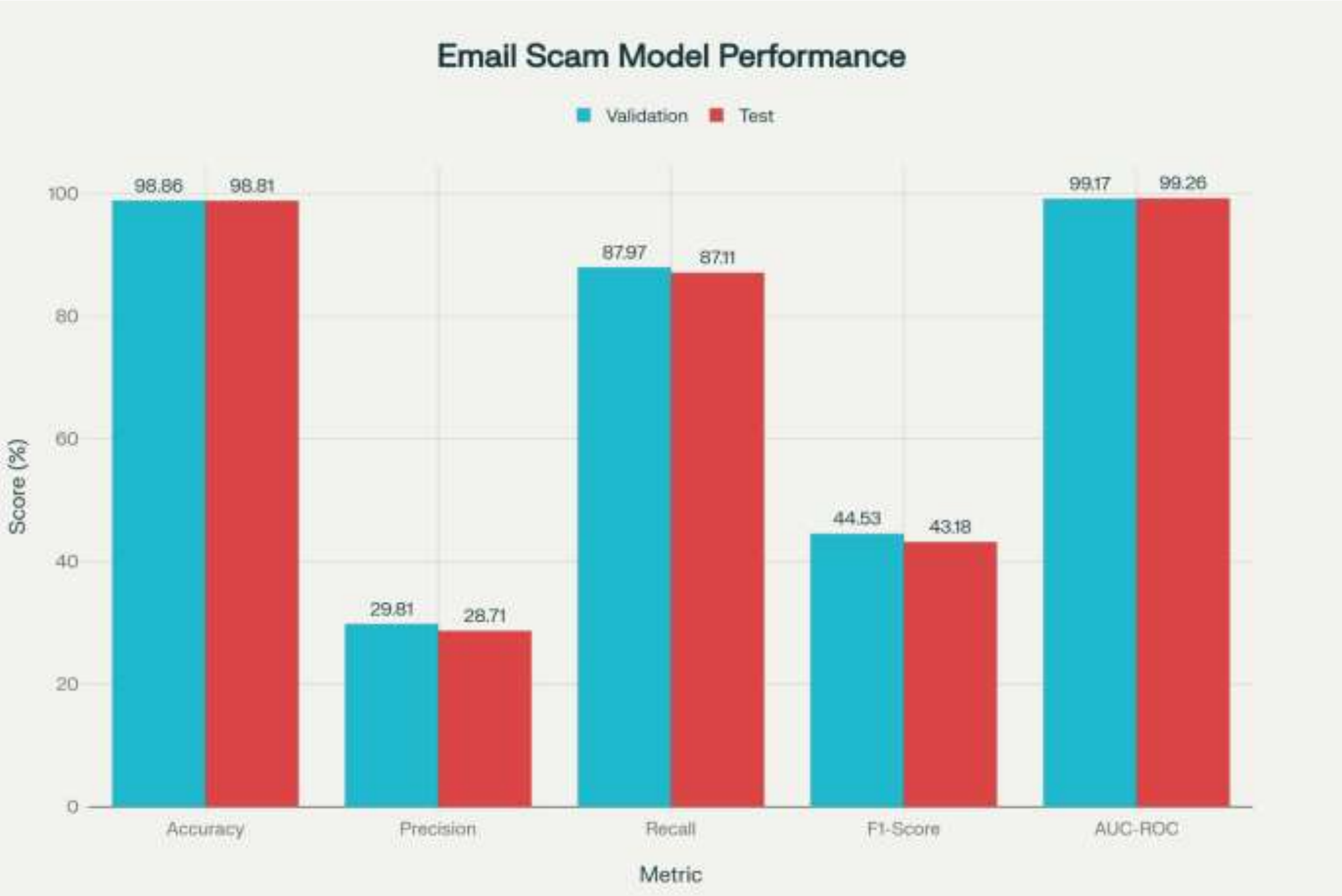
0.5

Estimated Scam

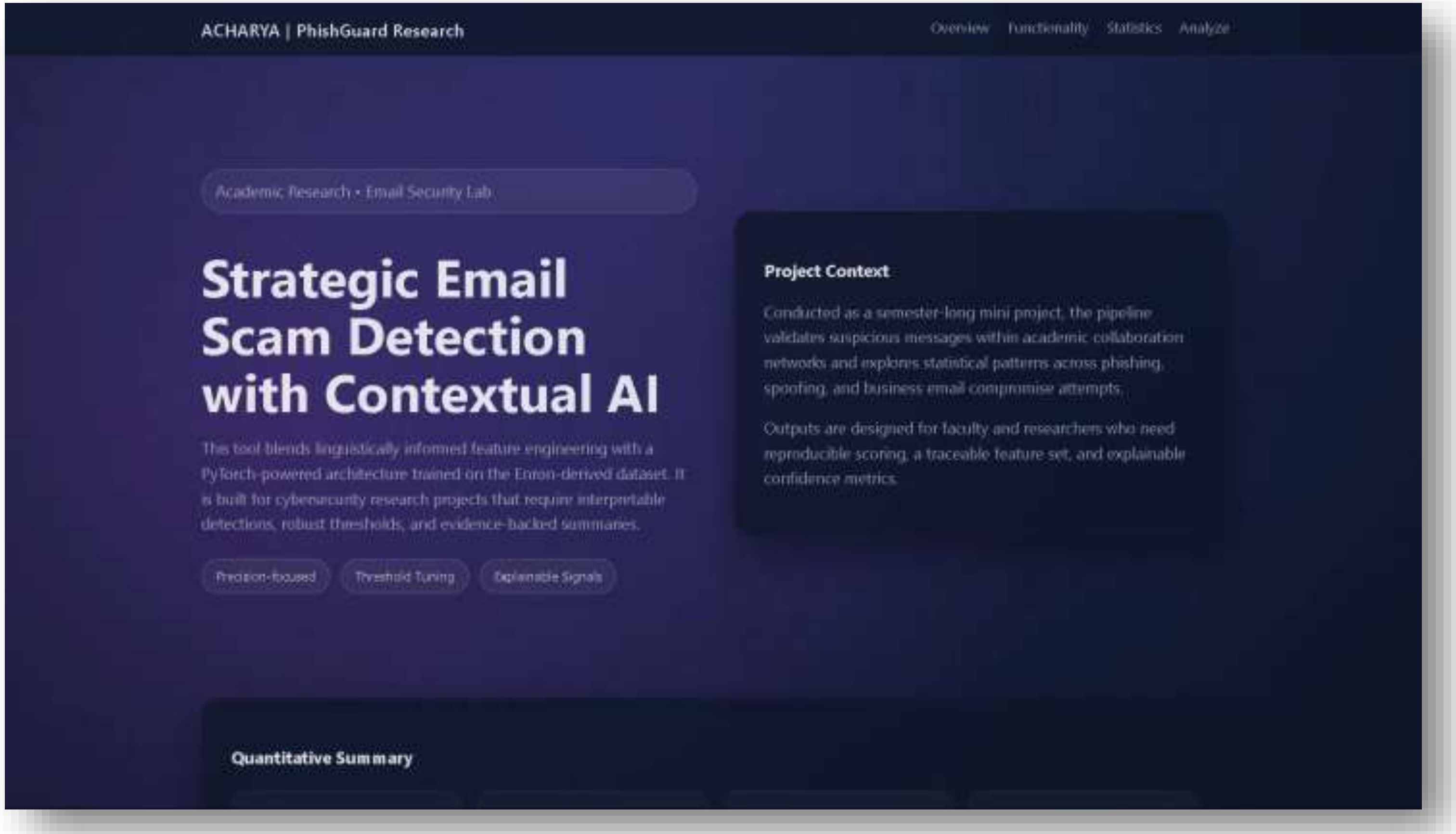
100% Confidence Score

The model combines statistical heuristics with contextual signals to assign this probability.

Results & Discussion



Demonstration of Prototype


ACHARYA | PhishGuard Research
Overview
Functionality
Statistics
Analyze
Academic Research • Email Security Lab

Strategic Email Scam Detection with Contextual AI

This tool blends linguistically informed feature engineering with a PyTorch-powered architecture trained on the Enron-derived dataset. It is built for cybersecurity research projects that require interpretable detections, robust thresholds, and evidence-backed summaries.

Project Context

Conducted as a semester-long mini project, the pipeline validates suspicious messages within academic collaboration networks and explores statistical patterns across phishing, spoofing, and business email compromise attempts.

Outputs are designed for faculty and researchers who need reproducible scoring, a traceable feature set, and explainable confidence metrics.

Quantitative Summary

Future Works

- Expand dataset, multi-language support
- Integrate with email clients, mobile app development
- Improve model with transformer-based architectures

Conclusion

- PhishGuard provides robust, real-time email phishing detection
- Harnesses advanced AI to safeguard users effectively

References

1. R. Roman, J. Zhou, and J. Lopez, "On the features and challenges of security and privacy in distributed internet of things,*" Computer Networks*, vol. 57, no. 10, pp. 2266-2279, 2013.
2. AdaPhish: AI-Powered Adaptive Defense and Education Resource Against Deceptive Emails by Rei Meguro & Ng S.T. Chong.
3. AI-Powered Threat Intelligence: Enhancing Real-Time Cyber Threat Detection and Response by Sumita Mukherjee, Kavita Thapliyal, Utpal Paul, Ravneet Singh Bhandari, Aditya Sinha, Yogesh Kumar

ACHARYA