International Journal of Advanced Multidisciplinary Research and Educational Development
Volume 2 - Issue 6 | November - December 2026 | www.ijamred.com
ISSN: 3107-6513

An AI-Based Spam Email and SMS Detection System Using NLP and Forensic Signal Analysis

Abiramii R K[1], Deeksha M M[2], Gayathri S[3], Kousikkrishna R[4], Dhanya K R[5]
Student, I Year, Department of Computer Science with Artificial Intelligence and Data Science,
KPR College of Arts, Science and Research, Coimbatore, Tamil Nadu, India
Assistant Professor, Department of Computer Science with Artificial Intelligence and Data Science,
KPR College of Arts, Science and Research, Coimbatore, Tamil Nadu, India
abiramiiramesh08@gmail.com, deekshamm2009@gmail.com, gayathriselvaraj627@gmail.com, rkousikkrishna@gmail.com, dhanyafab@gmail.com

Abstract
'Spam emails and SMS messages have become a common cybersecurity problem in digital communication. Attackers and spammers use deceptive language, urgency, financial pressure, fake rewards, impersonation and suspicious links to influence users. Traditional spam detection methods that depend only on individual keywords may fail when messages use modified spelling, indirect language or technically suspicious URLs. To address this problem, VOID // Shield is proposed as a browser-based spam email and SMS analysis system that combines Natural Language Processing (NLP) heuristics with URL, domain and forensic signal analysis. The system analyzes the content of an email or SMS and identifies security-related linguistic and technical indicators. The analysis includes suspicious phrase detection, urgency analysis, financial scam indicators, URL extraction, domain examination, homograph-style detection, raw IP detection, URL-shortener analysis and entropy-oriented character analysis. The system also provides an explainable result so that users can understand the signals contributing to the final assessment.
The proposed system provides a practical foundation for spam and phishing-oriented message analysis. Rather than depending on a single indicator, it combines multiple signals to produce a risk-oriented assessment. A formal research evaluation can be performed using a labelled dataset containing legitimate and spam messages and standard metrics such as accuracy, precision, recall and F1-score. The project can be further enhanced through supervised machine learning, transformer-based NLP, multilingual support and threat-intelligence integration.

Keywords— Spam Detection, Email Security, SMS Security, Natural Language Processing, Phishing Detection, URL Analysis, Domain Analysis, Cybersecurity,
Forensic Analysis.

1. INTRODUCTION

To develop an intelligent message-analysis system capable of identifying suspicious email and SMS messages.

3.2 Specific Objectives

Electronic communication has become an essential part of everyday life. Email and SMS are used for education, banking notifications, shopping, business communication, authentication and personal communication.
Along with these benefits, unwanted and deceptive messages have increased considerably.

- To analyze both email and SMS messages.
- To identify suspicious words and phrases.
- To detect urgency-based language.
- To identify financial scam indicators.
- To detect prize and reward-related scam patterns.
- To identify suspicious URLs.

'Spam messages are not limited to advertisements. Modern spam can contain phishing links, fake account warnings, fraudulent payment requests, prize claims, impersonation attempts and other forms of social engineering.
Such messages attempt to make users click links, provide information, transfer money or respond quickly without carefully examining the content.
'A major challenge in spam detection is that suspicious messages can change their appearance frequently, Attackers can modify words, use unusual characters, insert urgency-related language or hide destinations
behind shortened URLs. Therefore, a detection system should consider multiple types of evidence.
VOID J Shield is designed as a browser-based message security analysis system. It provides analysis for both email and SMS and combines NLP-oriented heuristics with technical URL and domain analysis. The
system is intended to provide an interpretable security assessment rather than simply returning a spam or non-spam label.
The project focuses on the following major areas:

1. Email and SMS message analysis.
2. NLP-based suspicious phrase detection.

ISSN: 3107-6513

3. Urgency and social-engineering signal detection.
4. URL extraction and structural analysis.
5. Domain and homograph-style analysis.
6. Entropy-oriented character analysis.
7. Explainable security signals.
8. Risk-oriented classification.

2. PROBLEM STATEMENT

Users receive a large number of digital messages every day. Some messages are legitimate while others are spam, phishing attempts or fraudulent communications.

Simple spam filters often depend on fixed keywords or previously identified senders. Such methods may not be sufficient when attackers modify the wording of messages or use new domains.

The problem addressed by this project is:

«To develop a browser-based system that analyzes email and SMS messages using linguistic and technical security indicators and provides an explainable spam-risk assessment.»

The system should identify suspicious characteristics without depending entirely on a single keyword or URL property.

3. OBJECTIVES

The main objectives of VOID J Shield are:

3.1 Primary Objective

• Email or SMS type

6.2 Preprocessing Layer

The message is prepared for analysis. This stage can include:

• To analyze domain structures.
• To identify possible homograph-style domain manipulation.
• To detect raw IP-based URLs.
• To identify shortened or redirected URLs.
• To calculate supporting forensic indicators.
• To provide an explainable result.

To create a foundation for future machine-learning integration.

• Text normalization
• Tokenization
• URL extraction
• Character-level processing
• Preservation of security-relevant symbols

4. EXISTING SYSTEM

Traditional spam detection systems generally use one or more of the following approaches:

6.3 Analysis Layer

The analysis layer performs several independent checks:

• NLP signal detection
• Urgency analysis
• Financial scam analysis
• URL analysis
• Domain analysis
• Homograph detection
• Entropy analysis

6.4 Decision Layer

The detected signals are aggregated into a risk-oriented assessment.

• Keyword matching
• Sender blocking
• Blacklists
• Rule-based filtering
• Statistical classification
• URL reputation
• Machine-learning classification

Keyword-based systems search for words commonly associated with spam. Although this approach is simple, it can produce false positives because legitimate messages may also contain words such as “urgent”, “payment” or “account”.

6.5 Presentation Layer

The result is presented through the browser interface using:

Blacklist-based systems depend on previously identified malicious senders or domains. Their limitation is that new domains and previously unseen messages may not be included.

• Classification result
• Threat level
• Risk-oriented information
• Detected signals

Some systems analyze only the message text, while others concentrate primarily on URLs. A more comprehensive system can combine both.

5. PROPOSED SYSTEM

VOID // Shield combines multiple evidence categories into one analysis workflow.

The proposed architecture consists of the following stages:

Input → Preprocessing → NLP Analysis → URL Extraction →
Domain Analysis → Forensic Analysis → Risk Aggregation →
Explainable Result

The system accepts a message and processes it through several analysis stages.

5.1 Major Modules

Module
Function
Input Module
Accepts email/SMS information
Preprocessing

ISSN: 3107-6513

Normalizes and prepares the message.

7. MESSAGE PREPROCESSING

NLP Module
Detects suspicious linguistic patterns
URL Module
Extracts and examines URLs
Domain Module
Examines domain structure
Preprocessing is an important stage because the same message may contain normal text, URLs, numbers, special characters and encoded information.
Urgency Module
Detects pressure-oriented language
Entropy Module
Provides character-distribution analysis
Decision Module
Aggregates detected signals

7.1 Text Normalization

Normalization can include converting text into a consistent representation for matching.
Explainability Module
Displays evidence and reasoning

6. SYSTEM ARCHITECTURE

Example
Original: “URGENT! Your account needs verification NOW.”
A normalized representation can help identify the words:
The architecture of VOID // Shield is designed as a multi-stage security analysis pipeline.

6.1 Input Layer

The user provides message information such as:

- urgent
- account
- verification
- now

However, the original representation should also be retained because capitalization and punctuation can provide useful forensic information.

- Sender
- Email subject
- Message body

7.2 Tokenization

Tokenization divides the message into smaller units called tokens.
Example
Message:
«Verify your account immediately.»
Tokens
- Verify
= your
- account
- Query parameters,
- Fragment

9.1 Suspicious URL Indicators

The system can examine:
- Very complex URLs.
- Unusual domain structures,
- Raw IP addresses.
- Encoded components.
- Multiple subdomains.
- immediately

7.3 URL Extraction

URLs should be extracted before aggressive text cleaning because URL structure contains
important security information.

8 NLP-BASED SPAM ANALYSIS

Natural Language Processing allows a computer system to analyze human language.
VOID 1! Shield uses security-oriented linguistic signals to identify potentially suspicious messages

8.1 Suspicious Phrase Detection

The system can examine phrases associated with:
- Urgency.
= Account warnings.
- Financial requests.
- Prize claims.
- Fee requests.
- Suspicious calls to action.

8.2 Urgency Detection

Attackers may attempt to force users to act quickly.
Examples of concepts that can indicate urgency include
- Immediate action.
- Account suspension warnings.
~ Limited-time claims.
- Emergency payment requests,
Urgency alone should not determine whether a message is malicious because legitimate services can also send urgent notifications.

8.3 Financial Scam Detection

Financial-related indicators can include:
- Payment requests,
- Unexpected fees
- Bank-related claims.
- Transfer requests.
- Requests for financial information

8.4 Prize and Reward Detection

‘Some spam messages claim that the recipient has won’ - Money.

10. DOMAIN ANALYSIS

Domain analysis examines the host portion of a URL.
‘A domain may be suspicious because of:
- Unusual spelling,

ISSN: 3107-6513

- Excessive subdomains,
- Character substitutions.
- Unexpected domain structure.
- Domain impersonation.

10.1 Homograph-Style Detection

‘A homograph attack can use visually similar characters to imitate another domain.

- Gifts
- Vouchers.
- Rewards
- Lottery prizes.

‘Such messages can be examined as one category of suspicious behaviour.

9. URL ANALYSIS

For example, a deceptive domain may use characters that visually resemble characters from a trusted domain

URLs are an important component of phishing and spam messages.

AURL can contain information about:

‘The purpose of the detector is to identify such structural warning signs and present them to the user.

11, FORENSIC SIGNAL ANALYSIS

- Scheme.
- Domain
- Subdomain.
- Path

The project includes forensic-style indicators that provide additional information beyond simple keyword detection.

10. Analyze domain structures.

11, Check for possible homograph-style pattems.

12. Calculate supporting entropy information

11.4 Urgency Vector

‘An urgency-oriented score can represent the concentration of urgency-related signals in the message

13. Aggregate detected signals.

14. Generate an explainable assessment.

16. Display the result.

16. Stop.

14, USER INTERFACE AND EXPLAINABILITY

A major feature of VOID i Shield is explainabilty. A higher value indicates stronger urgency-related language, but it should not be interpreted as proof of spam.

11.2 Character Entropy

Entropy describes the uncertainty or distribution of characters

within a string

Shannon entropy can be expressed as:

Instead of displaying only:

«SPAM»

the system can show the evidence associated with the assessment.

Possible result components include.

H(X) = ~E p09) log: p(x)

where

= H(X) = entropy.

- p(X) = probability of character x.

Higher entropy may occur in encoded or random-looking strings.

However, legitimate technical content can also have high entropy. Therefore, entropy should be used as a supporting signal.

12. RISK AGGREGATION

The system combines several indicators.

Acconceptual model can be represented as:

Risk = WsNLP + WURL + WsDomain + WiUrgency + WuEntropy

+ Signal breakdown,

- Forensic metrics,

This makes the system easier to understand and useful for academic demonstration.

15. DATASET DESIGN

where

- NLP = linguistic security signals.
- URL = URL-telated signals.
- Domain = domain-related signals.
- Urgency = urgency-related signals.
- Entropy = character-distribution signal
- W= corresponding weight.

The exact production weights should be determined from the implemented system and validated experimentally rather than assumed For scientific evaluation, a labelled dataset should be prepared A suitable dataset can contain

Field] Description

ID] Message identifier

Type| Email or SMS

Sender| Sender information

‘Subject Email subject

Body] Message content

Label| Spam or legitimate

URL| URL presence

Source| Dataset source

The dataset should contain different types of legitimate and spam messages

Examples include:

- Normal personal messages.
- Promotional messages.
- Bank notifications.
- Delivery notifications.
- Account notifications.
- Phishing messages.

Input: Email or SMS message

Output: Risk-oriented security assessment

1, Start

ISSN: 3107-6513

2. Accept email/SMS message.
3. Read sender, subject and message body.
4. Normalize the text
5, Extract URLs.
6. Scan the message for security-risk phrases.
7. Analyze urgency indicators.
8. Analyze financial and reward-related indicators.
9. Inspect URL structures,

should not be presented as independently validated accuracy or benchmark performance without a controlled dataset.

- Prize scams.
- Financial scams.
- Suspicious SMS messages.

16. EXPERIMENTAL METHODOLOGY

A formal experiment should divide the dataset into appropriate subsets.

For example:

+ Training set.
- Validation set
- Testing set

The testing data should not be used while developing the detector.

20. ADVANTAGES

The proposed system provides several advantages:

1, Supports both email and SMS.
2. Uses multiple security signals,
3. Provides NLP-oriented analysis.
4. Performs URL analysis.
5. Provides domain-oriented analysis
6. Supports suspicious phrase detection,
7. Includes urgency analysis.
8. Provides explainable output.
9. Can be extended with machine leaming. The system should generate predictions for the test messages and compare those predictions with the actual labels.

10, Provides a browser-based security workflow.

24. LIMITATIONS

Despite its useful features, the system has limitations.

21.1 Heuristic Limitation

Fixed rules can become outdated when attackers change their techniques

21.2 False Positives

Legitimate messages may contain words such as “urgent”, "payment" or "verify"

21.3 False Negatives

Sophisticated spam may avoid common suspicious words.

21.4 URL Ambiguity

A complex URL is not automatically malicious.

21.5 Entropy Ambiguity

High entropy may occur in legitimate technical strings.

21.6 Dataset Requirement

A formal evaluation requires a reliable labelled dataset.

17. EVALUATION METRICS

A.confusion matrix can be used to evaluate the detector.

| Predicted Legitimate| Predicted Spam
Actual Legitimate] TN| FP
‘Actual Spam| FN] TP

17.1 Accuracy

Accuracy = (TP + TN) / (TP + TN + FP + FN)

17.2 Precision

Precision = TP / (TP + FP)

17.3 Recall

Recall = TP / (TP + FN)

17.4 F1-Score

F1 = 2x Precision x Recall / (Precision + Recall)

These metrics should be calculated from actual experimental results. Numerical performance values should not be claimed unless they have been obtained from a documented test dataset.

18, TEST CASES

Test Case| Input| Expected Result

22, PRIVACY AND SECURITY

Message content can contain private information. Therefore, privacy should be considered during system development.

TCO1| Normal message| Low suspicious activity
TC02| Urgent financial message| Financial and urgency indicators

A privacy-oriented implementation should

- Minimize unnecessary storage.
- Avoid collecting message content without permission.
- Avoid unnecessary transmission to external services

C03} Suspicious URL| URL warning
TC04| Raw IP URL| IP-based indicator
TCO5| Homograph-style domain] Domain warning
TC06| Prize-retated message| Reward/prize signal
TC07| Suspicious SMS| NLP-based waming
TC08| Legitimate notification] Should avoid relying on one
keyword,
COS} Complex legitimate URL| Should be reviewed using multiple signals

- Clearly document third-party APIs.
- Protect user inputs.
- Safely render analyzed content.
- Avoid storing sensitive messages permanently.

The system should clearly communicate its actual data-handling behaviour to users.

23. FUTURE ENHANCEMENTS

ISSN: 3107-6513

19. RESULTS AND DISCUSSION

The main contribution of VOID // Shield is its multi-signal analysis approach.
Several improvements can be added in future versions.

23.1 Machine Learning

'Supervised learning can be introduced using labelled spam
and legitimate messages.
Possible models include:
'A message is not evaluated using only one word. Instead, the
system can consider:
Message Text + NLP Signals + URL + Domain + Urgency +
Forensic Indicators
This provides a broader analysis framework.
The explainability layer is particularly important because users can inspect the signals that contributed to the assessment

- Logistic Regression.
- Naive Bayes.
- Support Vector Machine,
- Random Forest.
- Gradient Boosting.

23.2 Transformer-Based NLP

The displayed values of a demonstration interface should be treated as interface outputs. They Transformer models can improve contextual language understanding and detect indirect or sophisticated suspicious language.
VOID // Shield — Spam Email and SMS Detection project website, supplied by the author.

23.3 Multilingual Detection

Future versions can support:

- Tamil

~ English

- Tamil-English code-mixed messages.
- Other Indian languages.

23.4 Threat Intelligence

Domain and URL reputation information can be integrated into the system.

23.5 Continuous Learning

The detector can be periodically evaluated using newly collected and labelled examples.

23.6 Explainable Al

More advanced explanations can show why particular signals influenced the final assessment.

24, CONCLUSION

VOID I Shield is a browser-based spam email and SMS detection system that combines Natural Language Processing-oriented heuristics with URL, domain and forensic signal analysis. The system focuses on identifying suspicious communication patterns such as urgency, financial pressure, reward-related claims, suspicious URLs and domain manipulation The major advantage of the proposed architecture is its multi-signal approach. Instead of depending entirely on a single keyword or technical property, the system combines different forms of evidence and presents them through an explainable interface. 'The project provides a strong foundation for an academic cybersecurity application. However, formal claims about detection performance require evaluation using a labelled dataset and standard metrics such as accuracy, precision, recall and F1-score.
Future development can transform the heuristic system into a more advanced intelligent security platform by integrating machine learning, transformer-based NLP, multilingual support, threat intelligence and continuously updated datasets.

REFERENCES.

1. Manning, C. D., Raghavan, P., & Schii tze, H. Introduction to Information Retrieval
2. Cambridge University Press, 2008
3. Sebastiani, F. "Machine Learning in Automated Text Categorization." ACM Computing Surveys, 34(1), 2002.
4. Guvella, T. S., & Caminhas, W.M. "A Review of Machine Learning Approaches to Spam Filtering." Expert Systems with Applications, 36(7), 2009.
5. Bird, S., Klein, E., & Loper, B. Natural Language Processing with Python. O'Reilly Media, 2009.
6. Bishop, C. M. Pattern Recognition and Machine Learning. Springer, 2006.
7. Goodfellow, I., Bengio, Y., & Courville, A. Deep Learning. MIT Press, 2016.
8. Shannon, C. E. "A Mathematical Theory of Communication." Bell System Technical Journal, 1948.