IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

AI Framework for Real-Time Spam Detection in Twitter

Mrs. T. Nandhini Raju[1], Balaji K[2] , Alangara Meervin A[3]
Assistant Professor, Department of Computer Science and Engineering[1]
Students, Department of Information Technology[2, 3,]
Dhanalakshmi Srinivasan University, Trichy, Tamil Nadu, India

Abstract: *Social media platforms like Twitter face a persistent threat from spam accounts and malicious users who degrade the user experience and compromise platform integrity. This paper presents a real-time spam detection framework using artificial intelligence techniques tailored for the Twitter ecosystem. The proposed framework utilizes layered system architecture comprising a presentation interface, backend logic, data storage, AI processing, and third-party integrations. It incorporates machine learning algorithms for classifying tweets as spam or legitimate in real-time using a combination of textual features, user behavior, and network patterns. The framework demonstrates improved accuracy and performance in identifying and mitigating spam compared to traditional methods.*

Keywords: Spam Detection, Artificial Intelligence, Twitter, Real-Time Processing, Machine Learning, System Architecture

I. INTRODUCTION

Twitter, as one of the most influential social media platforms, enables rapid dissemination of information but also attracts a high volume of spam and malicious content. Spammers exploit the platform's openness to spread advertisements, phishing links, and misinformation, which degrades the user experience and platform credibility. Traditional spam filtering methods struggle to keep up with the evolving tactics of spammers and often lack real-time capabilities. This paper introduces an AI-driven framework designed to detect spam on Twitter in real-time using machine learning and natural language processing (NLP) techniques. The system focuses on analyzing tweet content and user behavior to classify messages as spam or legitimate, offering a scalable and responsive solution to enhance platform security and trustworthiness.

II. SYSTEM OVERVIEW

The proposed system is a real-time, AI-powered spam detection framework designed specifically for Twitter. It utilizes a multi-layered architecture comprising a user-friendly interface, backend processing logic, data management systems, AI analytics, and integration modules. Tweets are ingested via the Twitter API and processed through various natural language processing (NLP) techniques such as tokenization, stop-word removal, and TF-IDF vectorization. The processed data is then classified using machine learning algorithms like Naive Bayes, SVM, Random Forest, or LSTM. The system is built on the Django framework and supports real-time detection with immediate feedback to the user. Designed with scalability, accuracy, and ease of use in mind, the system not only delivers high detection performance but also provides an interactive experience suitable for both casual users and analysts.

2.1 System Architecture

The proposed framework adopts a five-layered architecture designed for modularity, scalability, and real-time responsiveness. These layers are: the Presentation Layer, Application Layer, Data Layer, AI & Analytics Layer, and Integration Layer. Each layer performs a distinct function, collectively enabling efficient spam detection and seamless

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

user interaction. The layered approach facilitates maintainability and future scalability, allowing enhancements such as support for deep learning models or real-time multilingual processing.

2.2 Presentation Layer (User Interface)

The Presentation Layer provides the front-end interface that allows users, administrators, and analysts to interact with the system. Developed using modern web technologies (HTML5, CSS3, JavaScript, and Bootstrap), the interface includes dynamic elements such as neon effects, profile image support, and real-time feedback. Users can input tweets, select the preferred classification algorithm, and receive immediate predictions. The responsive design ensures accessibility across devices, enhancing usability for both technical and non-technical users.

2.3 Application Layer (Backend Logic)

The backend consists of RESTful services developed.

2.4 Data Layer (Database and Storage)

Data is stored in a hybrid schema.

2.5 AI & Analytics Layer

This is the core of the spam detection engine.

2.6 Integration Layer

The system integrates with the Twitter API.

III. HARDWARE COMPONENTS

The framework is designed to be deployed on scalable hardware.

IV. SYSTEM OPERATION

1. Tweets are continuously ingested using Twitter's Streaming API.

Table 1: Hardware Components Used

Component	Specification
CPU	Intel Xeon Gold 6226R (16 cores, 2.9GHz)
GPU	NVIDIA Tesla T4 (16 GB)
RAM	64 GB DDR4
Storage	1 TB NVMe SSD
Network Interface	1 Gbps Ethernet
Cloud Infrastructure	AWS EC2 (t3.large, G4 instances)

V. RESULTS

The system was evaluated on a dataset of 50,000 labeled tweets, demonstrating strong performance across different models. Naive Bayes achieved around 78% accuracy with fast prediction speeds, while Support Vector Machine (SVM) and Random Forest models delivered over 85% accuracy, offering improved precision and recall. The LSTM model performed best, reaching approximately 90% accuracy by capturing deeper language patterns, though it required more processing time. Despite this, all models provided results in under 2 seconds, supporting real-time detection. Feature extraction techniques like TF-IDF and Word2Vec significantly contributed to model effectiveness.

VI. CONCLUSION

In conclusion, the proposed AI framework for real-time spam detection on Twitter effectively combines machine learning techniques with a scalable system architecture to address the growing issue of spam on social media. The

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

integration of models like Naive Bayes, SVM, Random Forest, and LSTM allows for flexible and accurate detection based on user needs and resource availability. Real-time processing capabilities, supported by efficient data handling and feature extraction methods, ensure timely and reliable results. The system's user-friendly interface and modular design make it suitable for practical deployment, with future potential for multilingual support, deep learning enhancements, and cross-platform spam detection.

REFERENCES

[1]. A. Gupta and R. Kumar, "Real-time spam detection on Twitter using machine learning," IEEE Access, vol. 8, pp. 146303–146311, 2020.

[2]. Twitter Developer Documentation. [Online]. Available: https://developer.twitter.com/en/docs

[3]. M. S. Islam, F. A. Barbhuiya, and S. Das, "Spam detection in social media using semi-supervised learning," Proc. IEEE ICC, 2021.

[4]. H. Gao et al., "A survey of machine learning techniques for spam detection in social networks," Neural Computing and Applications, vol. 33, pp. 14701–14728, 2021.

[5]. TensorFlow Serving. [Online]. Available: https://www.tensorflow.org/tfx/guide/serving