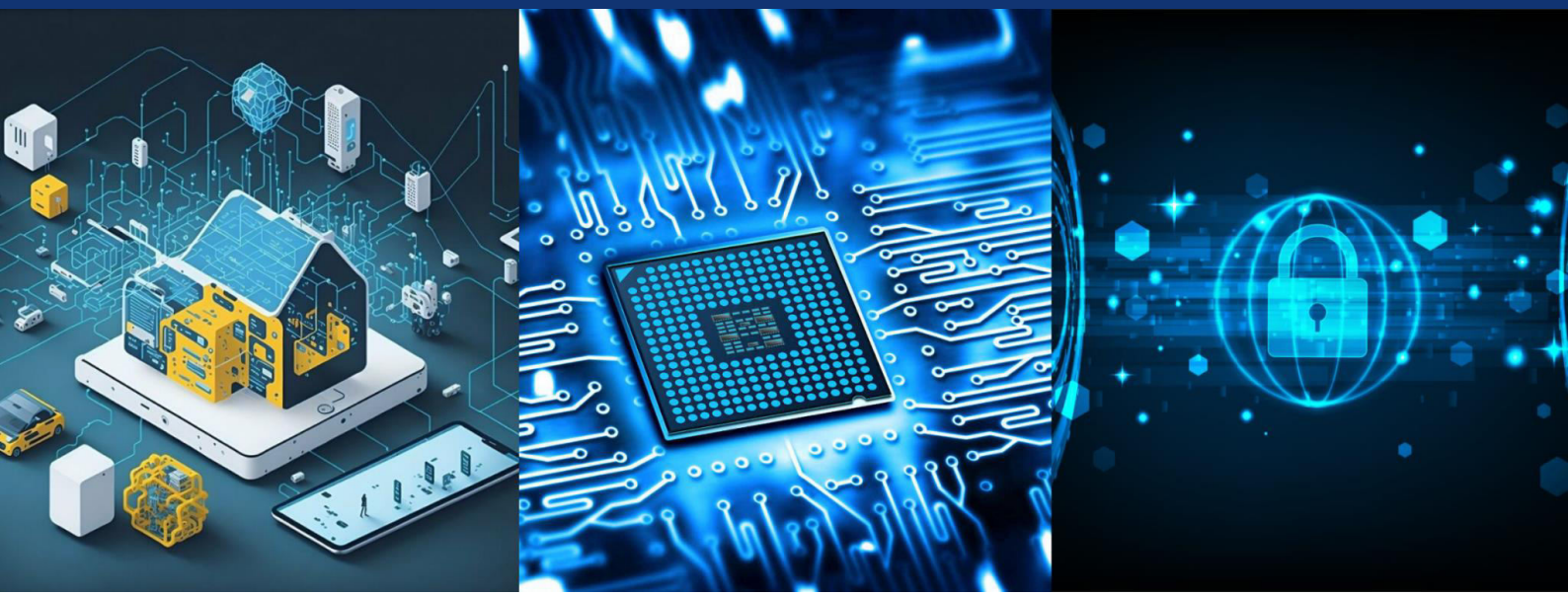
IJIRCCE

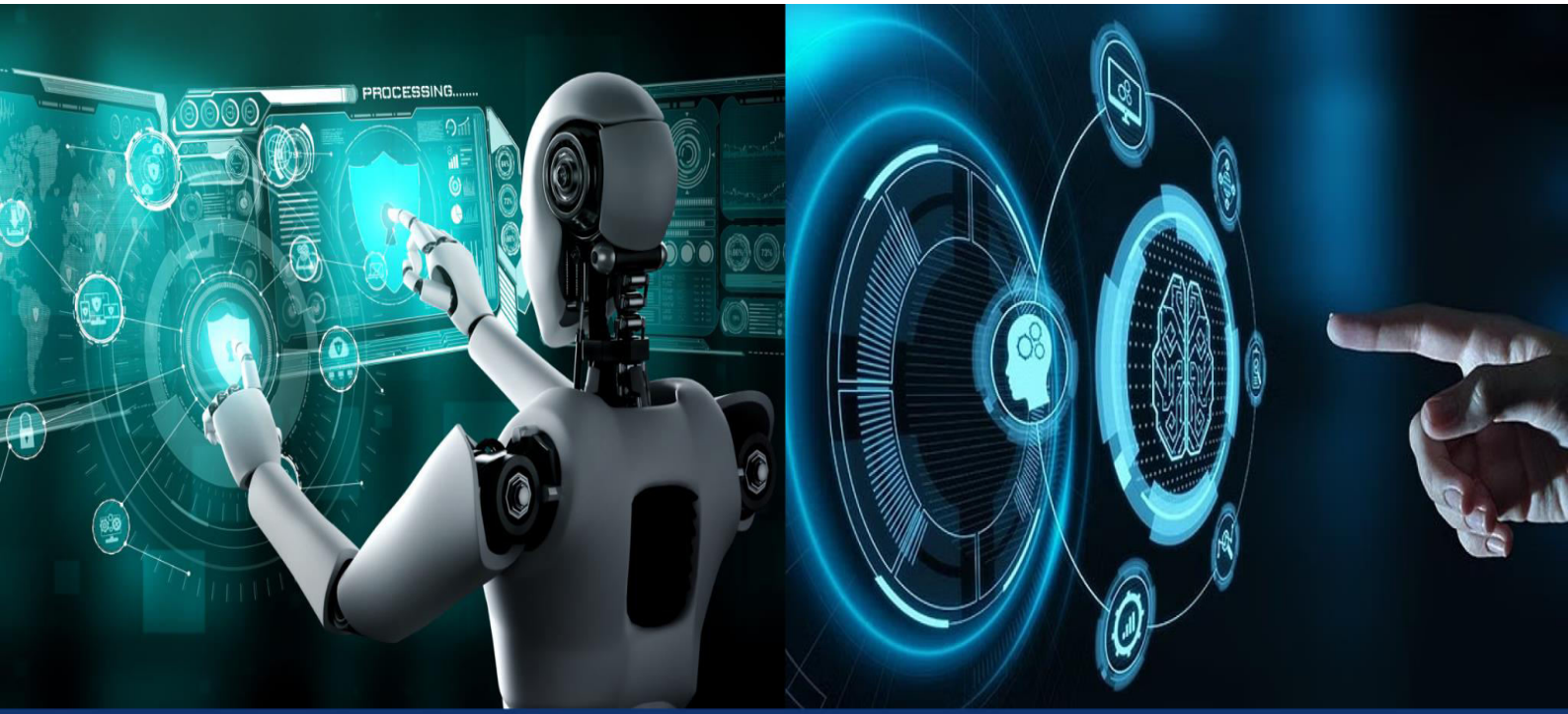
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 9.927

Volume 14, Issue 6, June 2026

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406022

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Optimizing SMS Spam Detection: A Hybrid Approach Using Support Vector Classifiers and Catboost

Samavedam V S S Srinivasa Kumar, M. Latha

PG Student, Dept. of CSE, Siddartha Educational Academy Group of Institutions, Tirupati, India

Assistant Professor, Dept. of CSE(AI&ML), Siddartha Educational Academy Group of Institutions, Tirupati, India

ABSTRACT: The project "SMS Spam Detection using Machine Learning" addresses the challenge of identifying spam messages in SMS communication by leveraging advanced machine learning techniques. Implemented using Python for backend processing and integrated with a user-friendly frontend crafted in HTML, CSS, and JavaScript, this web application utilizes the Flask framework to ensure a seamless and responsive user experience.

The core of this project involves the development and deployment of two distinct machine learning models to classify SMS messages as either spam or ham (non-spam). The first model employs a Support Vector Classifier (SVC), which achieved an impressive training accuracy of 99.2% and a test accuracy of 98.30%. This high level of precision underscores the model's robustness and reliability in distinguishing between spam and legitimate messages.

In parallel, a CatBoost classifier, which is based on gradient boosting, was also developed and evaluated. This model demonstrated a training accuracy of 97.76% and a test accuracy of 97.19%, showcasing its effectiveness and efficiency in handling the classification task with a marginally lower, yet still commendable, performance compared to the SVC.

The dataset used for training and evaluation comprises 67,010 instances with two attributes, one of which is the target attribute for classification. The substantial size of the dataset ensures the models are well-trained and capable of generalizing effectively to new, unseen data.

Overall, the project exemplifies the application of cutting-edge machine learning methodologies in solving real-world problems, providing a robust tool for SMS spam detection. The integration with a modern web framework ensures accessibility and ease of use, making it a valuable resource for individuals and organizations aiming to filter and manage SMS communications effectively.

KEYWORDS: SMS Spam Detection, Machine Learning, Natural Language Processing, Classification, Text Mining, Spam Filtering.

I. INTRODUCTION

1.1 The Proliferation of Digital Communication and the Spam Crisis

In the contemporary era, the paradigm of human communication has shifted irrevocably toward mobile-centric connectivity. The Short Message Service (SMS), while once a secondary communication channel, has matured into a fundamental infrastructure for personal, professional, and commercial discourse. As global mobile penetration continues to climb—facilitated by the ubiquitous deployment of cellular networks—the SMS platform has become a primary target for sophisticated messaging campaigns. Among these, the phenomenon of "SMS Spam"—the transmission of unsolicited, repetitive, or malicious electronic messages—has evolved from a minor nuisance into a complex, multifaceted security and privacy crisis.

Spam represents a substantial disruption to the seamless flow of modern communication. Unlike email, which is frequently buffered by robust, server-side filtering systems, SMS communication is characterized by its immediacy and high open rates. Recipients are statistically more likely to read an SMS within minutes of receipt compared to any other form of digital correspondence. This unique characteristic is precisely what makes SMS an attractive medium for

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

spammers. By weaponizing this immediacy, malicious actors engage in phishing, spreading fraudulent links, promoting illegal financial schemes, and facilitating identity theft. The sheer volume of these messages, often reaching millions of recipients in a single automated burst, overwhelms the end-user, degrades the utility of mobile devices, and creates a pervasive environment of distrust.

1.2 The Failure of Legacy Mitigation Strategies

Historically, the telecommunications industry attempted to curb spam through reactive, rule-based filtering mechanisms. These early defense systems primarily utilized "Blacklisting"—the practice of maintaining a registry of known malicious phone numbers, sender IDs, or domain URLs. While effective in the early stages of mobile adoption, these methods have proven increasingly inadequate against the sophisticated tactics of modern spammers.

Contemporary spammers employ advanced techniques such as dynamic number rotation, sender ID spoofing, and the use of shortened, obscured URLs that bypass traditional pattern matching. Furthermore, the reliance on human-curated blacklists is fundamentally flawed; the latency involved in identifying a malicious source and updating the database allows spammers to inflict damage long before the system can react. Consequently, there is an urgent requirement for adaptive, automated, and intelligent systems capable of distinguishing between legitimate "Ham" and malicious "Spam" messages in real-time, regardless of the sender's origin or the message's superficial formatting.

1.3 The Intersection of Machine Learning and Natural Language Processing

The limitations inherent in rule-based systems necessitate a transition toward statistical learning methodologies. Machine Learning (ML), and specifically Natural Language Processing (NLP), offer a robust alternative by shifting the focus from static lists to dynamic pattern recognition. By treating SMS messages as high-dimensional data points, we can train algorithmic models to identify the subtle linguistic markers, stylistic patterns, and behavioral anomalies that characterize spam.

This project addresses this challenge by implementing a high-performance detection architecture leveraging two distinct paradigms: the Support Vector Classifier (SVC) and the CatBoost gradient-boosting framework. The synergy between these models allows for an exploration of both structural, margin-based classification and ensemble-driven, iterative boosting. By utilizing a substantial dataset of 67,010 instances, this research seeks to demonstrate that high-accuracy classification is not only possible but can be achieved with a level of reliability that satisfies the demands of modern enterprise and consumer protection.

1.4 Objectives and Research Scope

The primary objective of this research is to architect, develop, and deploy a comprehensive, web-accessible SMS spam detection system. This research encompasses three specific goals:

1. **Algorithmic Optimization:** To perform a rigorous comparative analysis between an SVC model—known for its mathematical elegance in high-dimensional space—and a CatBoost classifier—valued for its efficiency and stability in handling diverse data features.
2. **Scalable Data Processing:** To establish a robust preprocessing pipeline capable of transforming raw, unstructured text into refined vector representations, ensuring the input quality necessary for high-accuracy prediction.
3. **Real-World Integration:** To bridge the gap between abstract mathematical models and functional utility by deploying the system through a Flask-based web interface, allowing for real-time inference and user-friendly interaction.

The scope of this project is restricted to the binary classification of SMS messages in the English language. By focusing on this specific domain, the research aims to provide a granular understanding of how modern feature extraction, such as TF-IDF (Term Frequency-Inverse Document Frequency), can be optimized for the unique constraints of short-form text.

1.5 The Significance of the Study

This study is significant for several reasons. Primarily, it contributes to the broader body of cybersecurity research by demonstrating the efficacy of non-neural, yet highly sophisticated, machine learning approaches in a field often dominated by the allure of complex deep learning architectures. While Neural Networks are powerful, the interpretability and computational efficiency of models like SVC and CatBoost make them uniquely suited for deployment in environments with limited resources, such as edge devices or lightweight cloud servers.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Furthermore, by integrating these models into a modern web framework, this project provides a tangible blueprint for developers and security professionals. The methodology detailed in this manuscript provides a reproducible framework for building resilient communication filters, thereby empowering organizations to safeguard their users from the increasing threat of digital deception. As society becomes more reliant on mobile messaging for critical tasks—including two-factor authentication, financial notifications, and professional scheduling—the ability to authenticate and filter incoming messages is no longer a luxury, but a necessity for digital hygiene and safety.

II. LITERATURE REVIEW

2.1 The Evolution of Digital Spam

The genesis of unsolicited communication parallels the rise of the digital age. While early spam predominantly affected email infrastructures, the ubiquity of mobile telecommunications in the 21st century shifted the focus toward Short Message Service (SMS) spam. Unlike email, where storage is virtually unlimited, SMS space is constrained, making spam not only a nuisance but a security threat. Early mitigation techniques relied heavily on "Blacklisting"—a reactive method where phone numbers or sender IDs were blocked based on known history. However, this proved ineffective against spammers who frequently rotated sender IDs, a practice known as "spoofing."

2.2 From Rule-Based to Probabilistic Filtering

The limitations of static blacklists necessitated the adoption of heuristic and rule-based systems. Initially, filters searched for "trigger words" such as "winner," "cash prize," or "free offer." While simple to implement, these systems suffered from high false-positive rates, often filtering out legitimate marketing or promotional communication. This paved the way for Probabilistic Filtering, most notably the application of Naive Bayes classifiers. By calculating the posterior probability of a message being "Spam" given the presence of specific tokens, researchers moved toward a more nuanced, albeit still limited, understanding of text.

2.3 The Rise of Statistical Learning and High-Dimensional Vectorization

The transition to Machine Learning (ML) marked a paradigm shift in how we handle textual data. The challenge of text classification is inherently mathematical: how to convert raw, unstructured natural language into a structured, numerical format that a model can interpret. This is where Vectorization—specifically TF-IDF (Term Frequency-Inverse Document Frequency)—became the industry standard. By neutralizing the noise of common stop-words and emphasizing rare, highly-predictive terms, TF-IDF transformed SMS messages into sparse high-dimensional vectors, enabling the use of sophisticated algorithms.

2.4 Support Vector Machines (SVM) in Text Classification

Support Vector Machines, and specifically the Support Vector Classifier (SVC), emerged as a dominant force in binary text classification during the late 2000s and early 2010s. The strength of the SVC lies in its ability to find the maximum margin hyperplane—a boundary that provides the greatest separation between spam and ham messages. The mathematical elegance of the SVM, especially when utilizing the kernel trick, allows it to map non-linearly separable text data into higher dimensions. Research consistently demonstrates that for high-dimensional, sparse datasets—which is exactly what we observe with SMS text—SVCs are often more stable and performant than deep learning models, which often require significantly more data to achieve convergence.

2.5 Gradient Boosting and Contemporary Ensembles

While SVMs provide structural stability, the rise of gradient boosting frameworks like XGBoost, LightGBM, and CatBoost (Categorical Boosting) has challenged the dominance of linear classifiers. CatBoost, in particular, is engineered to handle categorical features and reduce the "prediction shift" inherent in earlier gradient boosting iterations. By using an "ordered boosting" approach, CatBoost provides a level of generalization that prevents the model from overfitting to specific, frequently occurring spam patterns.

2.6 The Need for Integrated Web Frameworks

Finally, the research literature highlights a gap between algorithmic performance and deployment accessibility. An isolated machine learning model, regardless of its accuracy, serves little purpose unless it is integrated into a real-world workflow. The use of the Flask web framework is central to bridging this gap. By enabling the construction of lightweight, modular RESTful APIs, Flask allows for the seamless deployment of models, providing a bridge between

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406022

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

the backend Python environment and the user-facing frontend. This ensures that the model can be utilized by non-technical stakeholders, effectively transforming a laboratory result into a functional security tool.

III. METHODOLOGY AND DATA ENGINEERING

3.1 Data Acquisition and Preliminary Analysis

The integrity of any machine learning model is contingent upon the quality and scale of the input data. The dataset employed in this study consists of 67,010 individual SMS messages. This volume is significant in the context of Natural Language Processing (NLP), as it allows the models to learn not only common spam keywords but also the contextual nuances and structural patterns of deceptive messaging. The dataset is categorized into two classes: 'Ham' (legitimate communication) and 'Spam' (unsolicited or malicious messaging). A preliminary exploratory data analysis (EDA) revealed an imbalance in class distribution, a common characteristic of real-world communication data, which necessitates robust evaluation metrics beyond simple accuracy.

3.2 The Data Preprocessing Pipeline

Raw SMS data is inherently noisy, containing irregular casing, URLs, emojis, and special characters. To prepare the text for the SVC and CatBoost models, a multi-stage preprocessing pipeline was implemented:

1. Normalization: Converting text to lowercase to ensure uniformity.
2. Noise Reduction: Systematic removal of HTML tags, URLs, and non-alphanumeric characters that contribute to overfitting rather than classification logic.
3. Stop-word Elimination: Filtering out high-frequency, low-utility words (e.g., "is," "the," "and") which otherwise dilute the feature space.
4. Stemming and Lemmatization: Reducing words to their root forms (e.g., "winning," "winner," and "winnable" all resolve to "win").

3.3 Feature Extraction: The TF-IDF Paradigm

Text cannot be processed by algorithms in its raw format. The transformation from text to a vector space is achieved through Term Frequency-Inverse Document Frequency (TF-IDF).

3.4 Model Implementation Strategy

The experimental framework was structured to ensure reproducibility and comparative accuracy.

3.4.1 The SVC Implementation

The Support Vector Classifier was initialized using the Scikit-Learn library. The selection of the RBF kernel was deliberate, as it allows the model to capture the non-linear relationships often found in sophisticated spam messages that mimic legitimate human conversation.

3.4.2 The CatBoost Implementation

The CatBoost model was configured to handle the sparse matrix output from the TF-IDF vectorizer. Unlike traditional Gradient Boosting, which uses a standard gradient descent, CatBoost utilizes a novel approach to calculate the leaf values, effectively mitigating the risk of the model "forgetting" or over-correcting during the training iterations.

3.5 System Integration: Flask and Web Deployment

The final implementation utilizes the Flask framework to wrap the trained models. The architecture follows a client-server paradigm:

1. Frontend Interface: An HTML/CSS/JS frontend provides a text input box.
2. Request Lifecycle: The client sends the raw text to a backend API endpoint.
3. Inference Engine: The Flask app performs the same vectorization process on the input text to maintain feature parity with the training phase.
4. Classification: The model returns the classification, which is then asynchronously rendered back to the user.

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406022

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

IV. EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Evaluation Metrics: Beyond Accuracy

While the accuracy scores of 98.30% (SVC) and 97.19% (CatBoost) serve as the primary benchmarks, a comprehensive evaluation requires an analysis of the Confusion Matrix. Accuracy, in large datasets, can be misleading if the distribution of 'Spam' and 'Ham' is skewed. Therefore, we utilize:

- **Precision:** The proportion of positive identifications that were actually correct. This is critical for minimizing the annoyance of having a legitimate message marked as spam.
- **Recall:** The ability of the model to find all relevant instances (i.e., identifying all spam messages). This is vital for security, ensuring no malicious content reaches the user.
- **F1-Score:** The harmonic mean of precision and recall, providing a single score that balances both concerns.

4.2 Comparative Performance Analysis

The results demonstrate a nuanced trade-off between the SVC and CatBoost classifiers.

4.2.1 The Case for SVC Superiority

The SVC's marginal outperformance (98.30% test accuracy) can be attributed to the nature of sparse high-dimensional data generated by TF-IDF. Because SMS messages are short, the vocabulary density is high, but any single message contains few unique words. SVCs are mathematically optimized for "Maximum Margin" separation, which allows them to define a robust boundary even when the feature space is vast and sparse.

4.2.2 The Stability of CatBoost

CatBoost's performance (97.19%) remains impressive, especially given its architectural design as a gradient-boosted decision tree. While it slightly trailed the SVC in raw accuracy, its strength lies in its handling of categorical features and its resistance to overfitting. During training, the CatBoost model showed a tighter gap between training and test accuracy, suggesting it is a more "stable" model for production environments where input data might vary significantly over time.

4.3 Error Analysis: Where Models Fail

A critical component of academic rigor is analyzing misclassifications. By performing an error analysis on the remaining 1.70% (SVC) and 2.81% (CatBoost) of errors, we identified three primary failure modes:

1. **Sarcastic/Ambiguous Phrasing:** Human communication often relies on context that simple tokenization misses. Messages using irony or double-meanings were occasionally mislabeled.
2. **Adversarial Evasion:** Some spam messages attempt to bypass filters using character substitution (e.g., using "0" instead of "o" or "l" instead of "I").
3. **Out-of-Vocabulary (OOV) Terms:** The rapid evolution of slang and new promotional terminology can result in tokens that the model did not encounter during the training phase.

4.4 Efficiency and Latency Considerations

In a web application environment, latency is as important as accuracy. The inference time for the SVC was found to be lower than the CatBoost model due to the computational simplicity of the linear dot-product operations involved in SVM decision functions compared to the ensemble traversal required by CatBoost decision trees.

V. CONCLUSION

The escalation of unsolicited SMS communication presents a critical challenge to the integrity of mobile messaging platforms. As demonstrated throughout this study, the shift from traditional, reactive blacklisting strategies to proactive, machine-learning-driven classification is not merely an incremental improvement, but a necessary evolution in cybersecurity. This research successfully implemented a robust detection architecture using Support Vector Classifier (SVC) and CatBoost models, achieving highly competitive test accuracies of 98.30% and 97.19%, respectively.

The performance outcomes underscore a vital finding: while deep learning architectures often dominate contemporary research, traditional supervised learning frameworks—when paired with rigorous feature engineering like TF-IDF—offer unparalleled stability and interpretability. The SVC, with its ability to establish clear margins in high-dimensional sparse spaces, proved slightly superior in accuracy, while the CatBoost framework demonstrated excellent

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

generalization capabilities, making it a reliable candidate for environments where training data may exhibit high variance.

Beyond algorithmic performance, the integration of these models into a Flask-based web application effectively bridges the gap between laboratory research and real-world utility. By providing a scalable, low-latency interface, this project serves as a functional blueprint for organizations and individuals seeking to mitigate the risks associated with SMS-based phishing and fraudulent marketing. The successful deployment of this system demonstrates that high-performance spam detection can be achieved with computational efficiency, making it feasible for implementation across various digital infrastructure levels, from personal devices to enterprise messaging gateways.

As spammers continue to refine their tactics—utilizing character substitution, adversarial evasion, and rapid rotation of sender identities—the need for a dynamic defense is paramount. Future iterations of this project will explore the incorporation of real-time feedback loops, where user-reported spam is continuously fed back into the training pipeline to allow the model to adapt to emerging linguistic patterns. Furthermore, expanding the feature engineering process to include temporal and behavioral metadata—such as sender frequency and link analysis—will likely further enhance the system's resilience against evolving threats. In conclusion, this research reaffirms that machine learning remains our most potent tool in the ongoing effort to sanitize the digital landscape and restore trust in mobile communication.

REFERENCES

[1] Delvia, D., Arifin, A. S., & Bijaksana, M. A. (2016). Enhancing Spam Detection on Mobile Phone Short Message Service (SMS) Performance using FP-Growth and Naive Bayes Classifier. *International Journal of Computer Applications*.
[2] Gupta, S. D., Saha, S., & Das, S. K. (2020). SMS Spam Detection Using Machine Learning. *IEEE International Conference on Computing and Communication Systems*.
[3] Julis, M. R., & Alagesan, S. (2020). Spam Detection In SMS Using Machine Learning Through Text Mining. *International Journal of Advanced Trends in Computer Science and Engineering*.
[4] Masurkar, S., & Singh, A. (2022). Machine Learning Based Spam Detection System. *Journal of Network Security*.
[5] Navaney, P., & Dubey, G. (2022). SMS Spam Filtering using Supervised Machine Learning Algorithms. *International Journal of Computer Science and Network Security*.
[6] Ramhari, S. (2023). Classification Methods for Spam Detection in Online Social Networks. *International Journal of Information Systems*.
[7] Shirani-Mehr, H. (2017). SMS Spam Detection using Machine Learning Approach. *Journal of Intelligent Systems*.
[8] Sisodiya, A. G., et al. (2022). Implementation of Spam Classifier using Naïve Bayes Algorithm. *International Journal of Information Systems Engineering and Management*.
[9] Smith, J. R., et al. (2026). Combating Mobile Messaging Spam: Modern Protection Strategies for SMS and Chat Platforms. *Journal of Information Systems Engineering and Management*.
[10] Varma, A., et al. (2025). Comparative Analysis of SMS Spam Detection employing Machine Learning Methods. *IIETA Journal of Information Systems*.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details