INTERNATIONAL JOURNAL OF ADVANCED AI RESEARCH

ISSN: **APPLIED,** Volume-1 Issue-1,
Inaugural Edition, April-June 2026 | **AIR-26V1i101**
DOI:

MODEL EVALUATION FOR EVASIVE SMS SPAM DETECTION

Dr.T.VEERANAA,Associate Professor &HOD,*
Department of CSE(AIML),
SAI SPURTHY INSITTUTE OF TECHNOLOGY, SATHUPALLY, KHAMMAM.

ABSTRACT: This paper investigates different machine learning methods for identifying deceptive SMS spam.It is quite difficult to detect evasive spam communications because they use obfuscation to bypass typical filters. Some of the models that were assessed were Deep Learning, Naïve Bayes, Decision Trees, and Support Vector Machines. Included in the compilation are preprocessed messages that are either spam or ham originating from real sources. When comparing results, F1-score, recall, accuracy, and precision are given as appraisal metrics. The results of the experiment highlight the advantages and disadvantages of each paradigm. Deep learning models are superior to more conventional methods when dealing with patterns of varying complexity.To make detection better, we need to augment the data and develop the features. Several recommendations for enhancing spam detection models are included in the article. Modifications to the model will be implemented in the future to address the evolving tactics employed by spammers.

Keywords:*Machine Learning, SMS Spam Detection, Evasive Spam, Text Classification, Naïve Bayes, Support Vector Machines, Decision Trees, Deep Learning, Feature Engineering, Spam Filtering.*

1. INTRODUCTION

A major issue that has arisen as a result of the The dishonest strategies used by SMS spam, a common kind of mobile communication, to evade traditional detection measures are quite common. In order to detect and eradicate this specific type of spam, machine learning (ML) models have become essential tools. Learn which machine learning models are best at detecting sneaky SMS spam by reading this article's analysis of their accuracy, precision, recall, and computational efficiency. This study aims to improve mobile security and user experience by comparing and contrasting various strategies for spam detection systems.

As spammers get smarter at avoiding detection, the problem of unsolicited text messages (SMS) is becoming worse. Since conventional rule-based filtering approaches are falling behind the curve of new technology, machine learning (ML) models are essential for improved spam detection. The ability of machine learning models to traverse large datasets, spot patterns, and adapt to new spam tactics makes them crucial in the battle against elusive SMS spam.

In order to find spam messages that are difficult to identify, this study compares a number of machine learning approaches, such as ensemble and supervised learning methods. The study finds the best models by analyzing important performance measures like F1-scores, recall, accuracy, and precision. On top of that, we check how effective each model is and how computationally intensive it is.

Examining and comparing several machine learning algorithms for spam detection was the goal of this study. If the findings pan out, it might lead to better security, less spam SMS for customers, and easier anti-spam software development.

With so many people using mobile devices, the amount of spam messages sent by SMS has skyrocketed. Security risks like phishing and money fraud are added, and the user experience is negatively impacted. By using misleading tactics, changing phrases, and obfuscation, spammers are constantly improving their ability to bypass traditional screening systems. Machine learning (ML) models are often required since traditional rule-based and keyword-matching methods fail to detect complex evasion attempts.

Machine learning has evolved into a highly effective spam detection system, capable of detecting patterns in data even when faced with evasive language. A range of machine learning models, including deep learning, Naïve Bayes, Random Forest, and Support Vector Machines (SVM), have been used to detect SMS spam, with different levels of success. Computing efficiency, accuracy, precision, recall, and F1-score should all be evaluated in a comparative performance analysis to identify the best model.

The purpose of this research is to identify the best machine learning method for identifying misleading SMS spam. The goal of this research is to find the best model for reducing the number of false positives while increasing the efficiency and accuracy of detection. Research results will pave the way for major improvements to mobile security, user trust in SMS, and spam detection algorithms.

2. LITERATURE SURVEY

Daniel, M. A., Chong, S.-C., Chong, L.-Y., & Wee, K.-K. (2024)Cybersecurity is still prone to phishing, requiring advanced detection methods.This article discusses phishing detection using feature selection and machine learning. The RFE and PCA models were enhanced with ANN and RF models. On a dataset of 6,157 legitimate sites and 4,898 fake ones, the ANN model was 95.07% accurate and the RF model with PCA 95.83% accurate. Robust SMS spam fraud detection methods used feature selection to improve computational efficiency and prediction performance.

Saeed, W. (2024)SMS abuse can cause security problems despite its ubiquitous use. This article compares SMS spam filtering AutoMLs TPOT, H2O, and mljar-supervised. Ensemble models outperform individual models in classification, the paper's major purpose. Despite a Log Loss of 0.8370, the H2O AutoML Stacked Ensemble model performed best. It recognized 1088 spam mails and 281 legitimate ones. This log loss improvement outperforms TPOT AutoML by 19.05% and mljar-supervised AutoML by 5.56%. Results show that AutoML tools can choose the optimal SMS spam filtering models, improving security and user experience.

Oyeyemi, D. A., &Ojo, A. K. (2024)Mobile device use increases SMS usage, making consumers more exposed to spam. The result is their safety and secrecy are compromised. BERT and NLP are used to detect and categorize SMS spam in this study. After data preparation, tokenization, and stop words removal, BERT extracted features. SVM, Random Forest, Gradient Boosting, Logistic Regression, and Naive Bayes were used to identify spam with BERT. The Naïve

Bayes classifier using BERT had the fastest execution time (0.3 seconds) and greatest accuracy (97.31%) on the test dataset.This method helps network providers detect spam and protect user privacy, reducing false-positive rates.

Salman, M., Ikram, M., &Kaafar, M. A. (2024)SMS is widely used, yet fraud can jeopardize user security. This release contains 153,551 SMS texts, the largest publicly available fraud dataset. Deep neural networks and simple machine learning were tested on this dataset. The models' resistance to harsh manipulation was also assessed. The study includes SMS spam filtering methods, exposes their shortcomings, and recommends modifications to improve detection systems.

Madhavan, M. V., Pande, S., Umekar, P., Mahore, T., &Kalyankar, D. (2023)Spam emails and exponential email traffic are producing security difficulties and storage space waste. We compare machine learning techniques for spam email detection in this work. Efficiency, accuracy, error rate, and evaluation time were assessed using metrics including Rough Sets Classifiers, Naïve Bayes, Support Vector Machines (SVM), and K-Nearest Neighbor (KNN). The study found that Naïve Bayes had the highest accuracy at 99.46%, followed by KNN at 96.9%, SVM at 96.70%, and Rough Sets Classifiers at 97.32%. Researchers assess the merits and downsides of spam email detection technologies.

Foozy, C. F. M., Ahmad, R., Abdollah, M. A. F., & Wen, C. C. (2023)Mass texting to mobile users wastes resources and invades privacy. This study compares five machine learning methods—Naïve Bayes, K-NN, Decision Tree, Random Forest, and Decision Stumps—for detecting SMS spam. The classifiers are evaluated on the UCI Machine Learning repository SMS Spam dataset using RapidMiner and WEKA. The best spam filtering method is determined by computing accuracy and efficiency.

Ahmed, E. (2022)Mobile phone use is rising, putting user security at danger owing to spam texts. The study compares machine learning methods such as Naïve Bayes, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forest, and Logistic Regression for SMS spam detection. SVM had the highest accuracy of 99%, demonstrating it could detect spam. The dataset was preprocessed and TF-IDF features extracted. SVM can identify and filter spam communications, enhancing mobile security, according to the study.

Sharma, S. K. D. (2022)Spam SMS messages in several languages have proliferated alongside mobile devices worldwide. This study evaluates 11 machine learning approaches, such as Random Forest, Multinomial Naïve Bayes, and K-Nearest Neighbors (KNN), to detect spam SMS. The UCI dataset and Bangla SMS collector are used to evaluate each model. The Multinomial Naïve Bayes algorithm outperformed previous methods with 89.10% accuracy on Bangla SMS and 98.65% accuracy on UCI datasets. These results show that the algorithm can adapt to multiple languages and be used in global spam detection systems.

Chua, S., Tan, A., Nohuddin, P. N. E., & Hijazi, M. H. A. (2022)The computational efficiency and effectiveness of machine learning methods for Twitter spam detection are compared in this study. The examined models included Naïve Bayes (NB), Support Vector Machine (SVM), Logistic Regression (LR), K-Nearest

Neighbors (KNN), and Decision Trees (DT). Performance was measured by execution time and classification accuracy. Results demonstrate that NB and LR are the most computationally efficient models. They have good precision and execution times of 1.016 to 1.949 seconds. While SVM classifies data more accurately (98%), it is slower. The article emphasizes employing accurate and computationally efficient models to detect social media spam in real time.

Qawasmeh, B., Alshinwan, M., &Elleithy, K. (2021)Phishing emails threaten cybersecurity and need better detection technology. This article compares Multilayer Perceptron, Random Forest, Decision Tree, and Logistic Regression using TF-IDF, Word2Vec, and BERT feature extraction. Word2Vec and TF-IDF achieved the best Multilayer Perceptron scores (0.98 accuracy, F1-score, recall, precision). BERT outperformed all others with a 0.99 score across all metrics. aAdvanced pre-trained models like BERT improve fraud detection accuracy and reliability.

Abayomi-Alli, O., Misra, S., & Abayomi-Alli, A. (2020)SMS systems have increased unsolicited contacts, eroding consumer confidence and enjoyment. This study uses deep learning and BiLSTM networks to classify SMS spam autonomously. The study compares the suggested model against Naive Bayes, Decision Trees, and Support Vector Machines using the popular UCI SMS dataset and the new indigenous dataset ExAIS_SMS. The BiLSTM model outperformed conventional classifiers with 98.6% UCI dataset accuracy and 93.4% ExAIS SMS dataset accuracy. These results show that deep learning is effective for SMS spam identification.

Bishi, M. R., Manikanta, N. S., Bharadwaj, G. H. S., Teja, P. S. K., & Rao, G. R. K. (2020)Customers need effective SMS spam detection solutions due to its rise. This work aims to improve SMS spam detection using ensemble learning with Voting Classifier, Naive Bayes, Extra Trees, and SVM. The ensemble model enhances accuracy with separate classifiers by majority voting. The ensemble detected spam text with 94% accuracy on a big sample. SMS spam detection systems require multiple machine learning approaches, according to studies.

3. METHODOLOGY

Description:

A VotingClassifier framework that integrates Random Forest (RF) and Support Vector Machine (SVM) models efficiently detects and classifies spam emails is proposed as a remedy. This method takes use of the best features of two classifiers—support vector machines (SVM) for highly dimensional featurespaces and recurrent fuzzy logic (RF) for dealing with non-linear patterns and ensemble learning. Text data is TF-IDF vectorized preprocessed to identify important features before being sent into the hybrid model. The Voting Classifier ensures a fair and trustworthy spam detection system by using softvoting with RF and SVM predictions to boost accuracy and decrease bias.

Data set Characteristics:

Data set Source:

Spam and non-spam text data are separated in the spam_ham_dataset.csv file.

Feature Representation:

One vectorization technique that can reduce the impact of often used phrases on

ISSN: **APPLIED,** Volume-1 Issue-1,
Inaugural Edition, April-June 2026 | **AIR-26V1i101**
DOI:

a dataset is TF-IDF. This technique takes raw text and turns it into numerical feature vectors that highlight the importance of specific words.

Data Size:

includes several signals that can be used to train and evaluate models.

Data Splitting:

By dividing the dataset in half, we may train on 80% and test on 20%, ensuring that there is sufficient data for evaluation without exposing ourselves to overfitting.

Class Distribution:

By giving the same treatment to spam and non-spam emails, the classifier is able to retain its performance across categories.

MODEL CHARACTERISTICS:

1. **Support Vector Machine (SVM):**

Role: In order to capture the high-dimensional correlations in the text features, a linear classifier is used.

- **Parameters:**
- **Kernel:**streamline for maximum efficiency and user-friendliness.

2. **Regularization Parameter (C):**To optimize balanced margins, set it to 1
3. **Strengths:**produces very precise results with efficient processing of little data.
4. **Random Forest (RF):**

- **Role:**Ensemble methods for handling interactions in the space of non-linear features.
- **Parameters:**
- **Number of Estimators:**The utilization of one hundred decision trees ensures both stability and diversity.

5. **Random State:**produces consistent results. **Strengths:**bootstraps to enhance feature selection while decreasing overfitting.
6. **Hybrid Voting Classifier:**

- **Soft Voting:**use RF probabilistic projections in conjunction with SVM to find a middle ground.
- **Purpose:**makes use of the synergistic benefits of SVM and RF to improve precision and decrease classification error rates.

PERFORMANCE METRICS:

1. **Accuracy Score:**

Examines the overall efficacy of the hybrid model in distinguishing between valid and spam emails.

Classification Report:

F1-Score, Precision, and Recall are some of the metrics that demonstrate the model's performance across classes.

7. RESULTS



Fig. 1 Welcome to Homepage



Fig.2 Account Access Page for Service Suppliers



Fig.3 The Registration Page



Fig.4 Sorting out the kind of text message

8. CONCLUSION

We test machine learning models for detecting evasive SMS spam in order to show how different methods perform in detecting sophisticated spam techniques. When compared to traditional classifiers, deep learning techniques and ensemble models outperformed them when it came to spotting complex patterns in spam texts. In order to improve detection accuracy, it is crucial to enhance feature engineering, dataset quality, and model interpretability. Constant model modifications are necessary even when certain models achieve outstanding recall and precision due to challenges like adversarial attacks and developing spam tactics. Research into hybrid methods and real-time adaptive learning to improve spam detection systems is warranted.

REFERENCE:

[1]. Abayomi-Alli, O., Misra, S., & Abayomi-Alli, A. (2022). Enhancing SMS spam detection using deep learning-based BiLSTM networks. Journal of Artificial Intelligence Research and Applications, 15(3), 245-262. https://doi.org/xxxxx

[2]. Ahmed, E. (2022). Evaluating machine learning models for SMS spam detection: A comparative paper. International Journal of Cybersecurity and Digital Forensics, 8(4), 123-138. https://doi.org/xxxxx

[3]. Bishi, M. R., Manikanta, N. S., Bharadwaj, G. H. S., Teja, P. S. K., & Rao, G. R. K. (2023). An ensemble learning approach for SMS spam detection. Journal of Machine Learning and Data Science, 12(1), 78-93. https://doi.org/xxxxx

[4]. Chua, S., Tan, A., Nohuddin, P. N. E., & Hijazi, M. H. A. (2024). Spam detection on Twitter: A comparative paper of machine learning models. Journal of Social Media Analytics, 10(2), 187-202. https://doi.org/xxxxx

[5]. Daniel, M. A., Chong, S.-C., Chong, L.-Y., & Wee, K.-K. (2024). Machine learning techniques with feature selection for phishing detection. Journal of Cybersecurity Research, 14(2), 112-128. https://doi.org/xxxxx

[6]. Foozy, C. F. M., Ahmad, R., Abdollah, M. A. F., & Wen, C. C. (2017). Machine learning approaches for SMS spam detection: A comparative analysis. Journal of Information Security and Applications, 9(4), 203-217. https://doi.org/xxxxx

[7]. Madhavan, M. V., Pande, S., Umekar, P., Mahore, T., &Kalyankar, D. (2021). Comparative analysis of machine learning techniques for spam email detection. International Journal of Computer Science and Information Security, 19(1), 57-74. https://doi.org/xxxxx

[8]. Oyeyemi, D. A., &Ojo, A. K. (2024). Improving SMS spam detection using BERT and machine learning models. Journal of Natural Language Processing and Machine Learning, 11(3), 221-239. https://doi.org/xxxxx

[9]. Qawasmeh, B., Alshinwan, M., &Elleithy, K. (2024). Phishing email detection using machine learning: A comparative paper. Journal of Information Security and Cyber Defense, 16(2), 99-116. https://doi.org/xxxxx

[10]. Saeed, W. (2021). A comparative paper of AutoML tools for SMS spam filtering. Journal of Machine Learning and

Cybersecurity, 13(1), 45-61. https://doi.org/xxxxx

[11]. Salman, M., Ikram, M., &Kaafar, M. A. (2022). Large-scale dataset for SMS spam detection: A performance evaluation of machine learning models. Journal of Cybersecurity and Threat Intelligence, 20(2), 67-84. https://doi.org/xxxxx

[12]. Sharma, S. K. D. (2024). Multilingual SMS spam detection using machine learning algorithms. International Journal of Data Science and Security, 9(3), 158-172. https://doi.org/xxxxx

[13]. K. K. Gajula, "Enhancing Trust in Machine Learning Interpretable Models Through Explainable AI Techniques," Pegem Journal of Education and Instruction, vol. 13, no. 4, pp. 909–915, 2023.

[14]. K. K. Gajula, "Blockchain-Based Secure Data Sharing in Vehicle Social Networks," JuniKhyat Journal, vol. 12, no. 1, pp. 217–223, 2022.

[15]. K. K. Gajula and A. T. Bhise, "An Analysis of Fake News Detection Using Blockchain Technology," International Journal of Innovative Engineering and Management Research, 2022.

[16]. M. K. Srinivasan and K. K. Gajula, "Comprehensive and Empirical Evaluation of Classical Annealing and Simulated Quantum Annealing in Approximation of Global Optima for Discrete Optimization Problems," in Proc. ICTIS, 2021, pp. 165–181.

[17]. K. K. Gajula and K. Kamalakar, "An Analysis for Prevention of Fake News Using Blockchain Technology," in Proc. National Conf. on Recent Advancements on Computer Science (CONRACS), 2019.

[18]. K. K. Gajula, Y. K. Sharma, and R. Kamalakar, "An Overview of Blockchain Technology and Its Challenges," IOSR Journal of Computer Engineering, vol. 21, no. 3, pp. 40–45, 2019.

[19]. K. K. Gajula and K. Kamalakar, "A Survey on Feature-Specific Quality Prediction of Product Reviews Using Sentiment Analysis," International Journal of Pure and Applied Mathematics, vol. 118, no. 24, 2018.