International journal of application or innovation in engineering and management

Ijaiem.com/Mar 2025/ Volume 14/Issue 1/Article No-1/223-225

ISSN: 2319-4847

EVASION TECHNIQUES IN SMS SPAM FILTERING : A COMPARITIVE ANALYSIS USING ML

Bandi Indhu Goud[1], Perugu Vignyan[2] ,G Sathish [3]

[1,2]UG Scholar, Department of IT, St. Martin's Engineering College, Secunderabad, Telangana, India– 500100

[3]Assistant Professor, Department of IT,St. Martin's Engineering College, Secunderabad, Telangana ,India– 500100

indhugoudbandi@gmail.com

Article Info

Received: 28-03-2025 ***Revised: 05 -04-2025*** ***Accepted: 16-04-2025*** ***Published:27/04/2025***

Abstract:

The persistence of SMS spam remains a significant challenge, highlighting the need for research aimed at developing systems capable of effectively handling the evasive strategies used by spammers. Such research efforts are important for safeguarding the general public from the detrimental impact of SMS spam. In this study, we aim to highlight the challenges encountered in the current landscape of SMS spam detection and filtering. To address these challenges, we present a new SMS dataset comprising more than 68K SMS messages with 61% legitimate (ham) SMS and 39% spam messages. Notably, this dataset, we release for further research, represents the largest publicly available SMS spam dataset to date. To characterize the dataset, we perform a longitudinal analysis of spam evolution. We then extract semantic and syntactic features to evaluate and compare the performance of well-known machine learning based SMS spam detection methods, ranging from shallow machine learning approaches to advanced deep neural networks. We investigate the robustness of existing SMS spam detection models and popular anti-spam services against spammers' evasion techniques. Our findings reveal that the majority of shallow machine learning based techniques and anti-spam services exhibit inadequate performance when it comes to accurately classifying SMS spam messages. We observe that all of the machine learning approaches and anti-spam services are susceptible to various evasive strategies employed by spammers. To address the identified limitations, our study advocates for researchers to delve into these areas to advance the field of SMS spam detection and anti-spam services.

Keywords: Anti-spam services, evasive techniques, machine learning robustness analysis, SMS spam detection, spam dataset, SMS spam evolution.

1.INTRODUCTION

The rise of SMS spam poses not only financial risks but also threatens user trust in communication systems, making it a critical issue that demands immediate attention. As mobile phones are integral to daily life, the widespread nature of SMS spam affects individuals, businesses, and even governments. The deceptive tactics employed by scammers—ranging from phishing attempts, fraudulent offers, to impersonating trusted entities—have become increasingly sophisticated, making detection more difficult. The global nature of this problem, as evidenced by growing reports across multiple across countries, highlights the challenge of combating SMS spam in a rapidly evolving digital landscape.Despite advancements in detection techniques, spammers continually adapt, using methods such as number spoofing and automated systems to send out massive volumes of spam messages. The growing financial toll and the rapid evolution of scam tactics make it clear that SMS spam will continue to be a significant challenge.These tactics exploit vulnerabilities in telecommunication networks and often bypass existing security profound, with many victims losing large sums of money or compromising sensitive personal information. Moreover, the prevalence of SMS spam is not just a consumer issue; it also places a heavy burden on service providers, who must allocate significant resources to block and filter unwanted messages. Regulatory bodies in various countries, such as the ACCC in Australia and the Federal Trade Commission (FTC) in the US, have implemented strict rules and penalties, but enforcement alone has proven insufficient in curbing the threat. This underscores the need for a more holistic approach—one that includes technological innovation, international cooperation, and public awareness campaigns to reduce the impact of SMS spam.The growing financial toll and the rapid evolution of scam tactics make it clear that SMS spam will continue to be a significant challenge. Therefore, ongoing research and the development of more advanced machine learning models and detection systems are crucial. These efforts will help build more resilient defenses and ensure the safety of mobile communication networks in the face of this persistent and costly issue. In this work, we identify main challenges in SMS spam.The Availability of Data A major challenge in building SMS spam detection models is the scarcity of large, real-world, annotated datasets. Prior studies often rely on outdated and highly imbalanced datasets with only a few hundred spam messages. Lack of Benchmarks Datasets: Various methods have been proposed for SMS Spam detection however, the absence of standardised benchmark dataset for comprehensive comparisons has let to fragmented research in this SMS spam detection area.

2. LITERATURE SURVEY

The Investigating and Prosecuting Nigerian Fraud article by Buchanan and Grant discusses the various types of fraud schemes originating from Nigeria, with a particular focus on the infamous "419" advance-fee scam.It outlines the efforts of the Nigerian Crime Initiative (NCI) in combating these fraudulent activities and provides examples of successful prosecutions. The paper highlights challenges faced in prosecuting these cases, such as the need for international cooperation and collaboration between agencies, which are crucial in overcoming the complexities of transnational fraud. A critical point addressed is the need for international cooperation in combating such crimes, as fraud schemes often cross borders, making prosecution complicated. The challenges of interagency collaboration are explored, including differing legal frameworks, language barriers, and the complexities of gathering and sharing evidence across jurisdictions. By presenting successful case studies, the authors highlight the strides made in prosecuting fraudsters but also emphasize the ongoing need for more robust international frameworks and mechanisms for cross-border cooperation to effectively combat these frauds.The An Evaluation of Statistical Spam Filtering Techniques article by Zhang, Zhu, and Yao evaluate different statistical techniques used for filtering spam, emphasizing machine learning methods to effectively distinguish between spam and legitimate messages. The study particularly highlights the performance of classifiers like Naïve Bayes and Support Vector Machines (SVM), providing valuable insights into their accuracy and efficiency in various contexts. The paper serves as an in-depth analysis of how these statistical methods perform under different environments, aiding in the development of more effective spam filtering technologies.The paper examines the strengths and limitations of these classifiers. Naïve Bayes, for instance, is known for its simplicity and efficiency, but it may not always provide the best performance in the face of highly sophisticated spam tactics. On the other hand, SVMs, which are more computationally intensive, offer better accuracy in many cases, especially when the data is high-dimensional. This evaluation highlights the need for continuous improvement in spam filtering algorithms to keep up with increasingly sophisticated spam techniques, and provides a foundation for further research into combining multiple methods to enhance detection accuracy.

The On Attacking Statistical Spam Filters article by Wittel and Wu explore the vulnerabilities inherent in statistical spam filters, particularly how attackers can exploit these weaknesses, such as through the poisoning of training data. The authors evaluate the resilience of these filters when subjected to such attacks and contribute to a deeper understanding of how adversarial spam evolves. This research underscores the importance of enhancing spam filters' robustness against evolving and increasingly sophisticated spam techniques. The authors explore how these attacks can undermine the resilience of spam filters and emphasize the need for developing more robust systems capable of handling adversarial manipulations. Their research highlights an ongoing arms race between spam filter developers and spammers, where both sides continually adapt to outsmart each other. The paper provides recommendations for enhancing filter security, such as incorporating dynamic learning techniques and adversarial training to make spam filters more resilient to manipulation.

The Contributions to the Study of SMS Spam Filtering: New Collection and Results article by Almeida, Hidalgo, and Yamakami introduce a new SMS spam dataset and benchmark several spam filtering algorithms, including Naïve Bayes, SVM, and decision trees. The authors emphasize the rising need for effective mobile spam control due to the growing prevalence of SMS spam. The paper also discusses how machine learning techniques can be applied to improve SMS spam detection, providing insights into the challenges and opportunities for enhancing filtering accuracy on mobile platforms. The authors also explore the growing need for effective spam control on mobile platforms. As mobile devices become integral to personal communication and commerce, SMS spam poses a major threat to users' privacy and security. The paper discusses how machine learning can be leveraged to improve spam detection rates on mobile devices, making it easier for users to identify and block unwanted messages. By benchmarking different algorithms on this new dataset, the study provides a valuable resource for future developments in SMS spam filtering.

The Towards SMS Spam Filtering: Results Under a New Dataset article by Building on previous research, Almeida, Hidalgo, and Silva focus on improving SMS spam filters using a new dataset. The paper examines the practical aspects of deploying these filters, with a particular emphasis on increasing detection rates while minimizing false positives. It provides a detailed exploration of the methods and techniques that can be used to enhance the performance of SMS spam filtering systems, making them more efficient and reliable for real-world applications. The paper discusses several methods for improving the efficiency of SMS spam detection systems. One approach is to enhance the feature extraction process, which involves identifying key patterns in SMS content that are indicative of spam. The authors also explore how algorithms can be fine-tuned to adapt to the evolving nature of SMS spam. Through this research, the paper provides a roadmap for developing more reliable and user-friendly SMS spam filtering solutions, which is essential for ensuring the continued success of mobile communication systems.

The Comparative Study of Spam SMS Detection Using Machine Learning Classifiers by Gupta, Bakliwal, Agarwal, and Mehndiratta present a comparative study of various machine learning classifiers, such as Decision Trees, Random Forest, and SVM, for detecting SMS spam. The paper evaluates the performance of these classifiers across different datasets, highlighting key factors that affect their detection accuracy. It provides valuable insights into which classifiers are most effective in detecting spam and how different approaches can be optimized for better performance in SMS spam detection systems.

3. PROPOSED METHODOLOGY

The proposed system introduces several key improvements over existing SMS spam filtering methods, addressing many of the limitations found in current approaches. Robust detection is at the core of this system, as it leverages deep learning techniques combined with a comprehensive feature extraction process to improve the identification and classification of spam messages. By utilizing more sophisticated models that go beyond simple keyword matching, the system is capable of capturing complex patterns and behaviour exhibited by spam, leading to more accurate detection.A major strength of this system is its reliance on up-to-date data, the proposed system will continuously integrate a large and diverse dataset that reflects the latest spam trends. This will ensure the system can identify and respond to new types of spam as they emerge, significantly improving its adaptability and performance in dynamic environments. Staying ahead of evolving spam tactics is essential for maintaining high detection rates and minimizing false negatives.

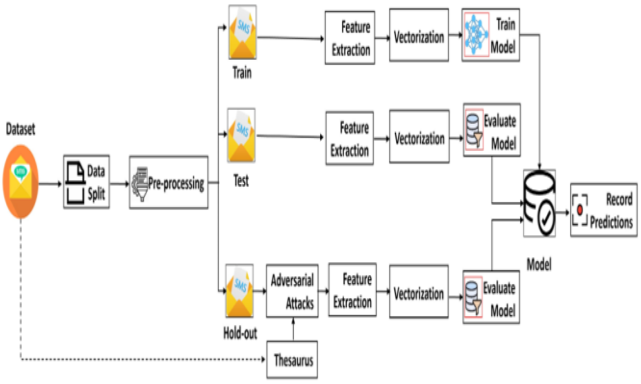


Figure 1: Proposed System Architecture.

The system is also designed to effectively counter advanced evasion detection techniques used by spammers. By incorporating evasion-aware strategies, the model will be equipped to detect and handle sophisticated spam methods such as character manipulation, paraphrasing, and the use of URL shorteners. These advanced techniques have proven to be effective in bypassing traditional filters, but the proposed system's deep learning algorithms will be trained to recognize such evasive manoeuvres, providing a much-needed layer of resilience.Additionally, the system will employ a hybrid approach, combining content-based analysis with metadata and network-level insights. This integration of multiple data sources will enable the system to evaluate both the message content and contextual information, such as sender behaviour, message frequency, and delivery patterns. By utilizing this multi-dimensional analysis, the system will improve its ability to detect spam while reducing the risk of misclassifying legitimate messages.

Furthermore, the proposed system will undergo real-world evaluation to ensure its effectiveness in actual deployment environments. Unlike many models that are limited to lab-based testing, this system will be rigorously tested in real-world scenarios, accounting for diverse messaging behaviour's and real-time spam trends. This comprehensive evaluation will help refine the system, ensuring that it can reliably detect and block spam under real-world conditions without introducing unnecessary friction for users.

Dataset:The foundation of any machine learning system lies in the dataset used for training and evaluation. In the context of spam detection, the dataset consists of messages that have been labeled as either legitimate or spam. A good dataset will have a diverse range of examples that reflect the varied types of messages encountered in real-world scenarios. It should include both text-based features (such as the content of the message) and metadata (such as sender information, time of sending, and frequency of messages). Proper dataset collection is crucial as it ensures that the model can generalize well to unseen data and can effectively handle various types of spam.

Data Split :Once a dataset has been collected, it is typically divided into three subsets: training, testing, and validation (or hold-out) data. The training data is used to train the machine learning model, allowing the system to learn patterns and relationships within the data. The testing data is used to evaluate the model's performance on previously unseen examples, providing an estimate of how well the model will perform in the real world. A common split is 70% for training, 15% for testing, and 15% for validation.

Pre-processing :Before feeding the data into a machine learning model, pre-processing is essential to clean and prepare the data. This may involve steps such as removing irrelevant information (e.g., special characters, stopwords), converting text to lowercase, normalizing text (e.g., stemming or lemmatization), and handling missing values. Pre-processing also includes handling outliers or noisy data that could disrupt the model's learning process. The goal is to ensure that the data is in a clean and usable format, optimizing the learning process.

Train : The training phase involves using the labeled dataset to train a machine learning model. The algorithm learns from the data by adjusting its internal parameters to minimize error. During training, the system is exposed to a large number of examples (both spam and legitimate messages) to help it distinguish between the two. This phase is crucial because the quality of training directly impacts the performance of the model. If the model is trained on a representative dataset, it is more likely to perform well on unseen data.

Test :Once the model has been trained, it must be evaluated on a separate test dataset to measure its performance. This phase tests the model's ability to generalize to new, unseen data. Performance metrics like accuracy, precision, recall, and F1 score are often used to evaluate how well the model is detecting spam while minimizing false positives and false negatives. The test set acts as a proxy for how the model will behave in real-world applications.

Hold-out :The hold-out set is used to ensure that the model's performance is evaluated on data that was not used during training. The hold-out set is typically kept aside during the training process, and once the model is trained and fine-tuned, the hold-out data is used for final evaluation. This step helps to prevent overfitting, where the model becomes too tailored to the training data and fails to generalize well to new data.

Adversarial Attacks :Adversarial attacks refer to techniques that attempt to mislead machine learning models by introducing slight perturbations to the input data. In the case of spam detection, attackers may manipulate the content of a message (e.g., changing characters, rewording, or using deceptive links) to evade detection by the model. It is important to test the model's robustness against adversarial attacks, ensuring it remains accurate and reliable even when faced with intentionally altered spam messages.

Applications :

Mobile Security: SMS spam filters are essential in protecting users from fraudulent schemes such as phishing, scams, and financial fraud.

User Experience: By filtering out spam messages, users are provided with a cleaner and more efficient messaging experience. This improves the usability of mobile devices and prevents the user interface from becoming cluttered with irrelevant content.

Prevention of Malware Spread: Spam messages often contain links to malicious websites or attachments that can lead to the installation of malware.

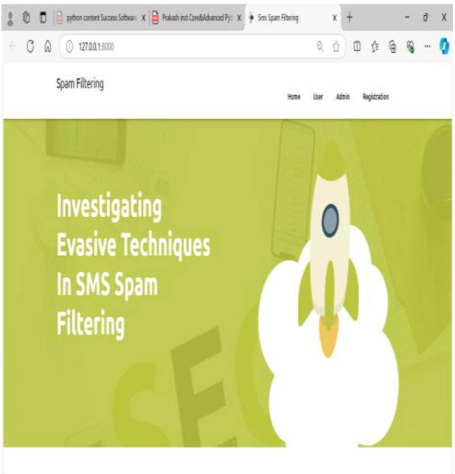
Advantages :

Adaptability to New Evasion Techniques: Machine learning models, especially those based on deep learning, can continuously evolve as new techniques for evading spam filters emerge.

High Accuracy and Precision: By analyzing a large volume of SMS data, machine learning models can provide high accuracy in detecting spam messages.

Real-time Detection: Machine learning models can be trained to filter SMS messages in real-time.

4. EXPERIMENTAL ANALYSIS



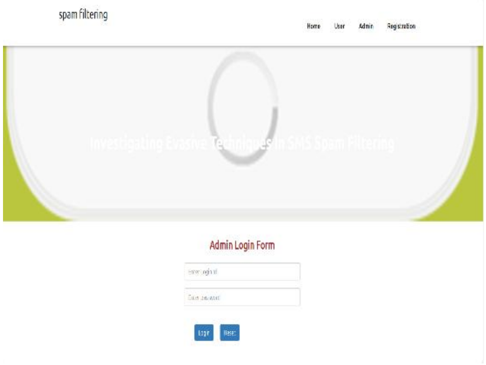
Investigating Evasive Techniques In SMS Spam Filtering

Figure 2: Home Page

Investigating Evasive Techniques In SMS Spam Filtering

User Register Form

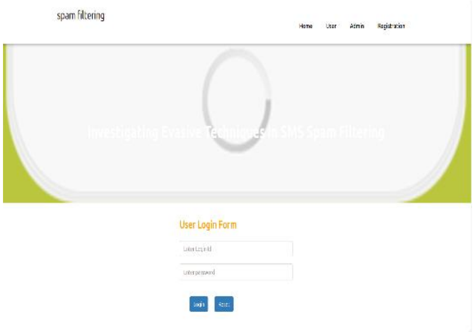
Figure3: Registration Page



Investigating Evasive Techniques In SMS Spam Filtering

Admin Login Form

Figure 4: Admin Page



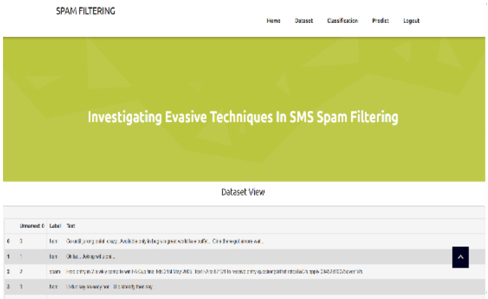
spam filtering

User Login Form

Enter Login Id

Enter password

Fig 5 : User Login Page



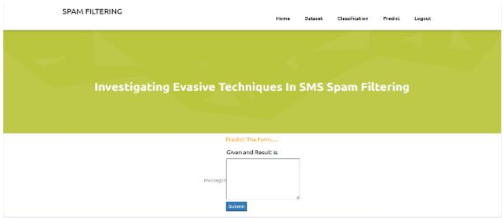
SPAM FILTERING

Investigating Evasive Techniques In SMS Spam Filtering

Dataset View

	Unnamed: 0	Label	Text
0	0	ham	Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got amore wat...
1	1	ham	Ok lar... Joking wif u oni...
2	2	spam	Free entry in 2 a wkly comp to win FA Cup final tkts 21st May 2005. Text FA to 87121 to receive entry question(std txt rate)T&C's apply 08452810075over18's
3	3	ham	U dun say so early hor... U c already then say...

Figure 6: User Home Page



SPAM FILTERING

Investigating Evasive Techniques In SMS Spam Filtering

Predict The form.....

Given and Result is

Messages

Figure 7: Final Prediction Page

5. CONCLUSION

To highlight the changing characteristics of SMS spam, we introduced and characterized a large new SMS dataset. We used the dataset to benchmark the performance and robustness of several machine-learning-based models and the anti-spam ecosystem for detecting SMS spam. The results showed that all of the machine learning based models efficiently identified legitimate SMS messages(ham SMSes). However, only a few deep learning models and anti-spam text apps performed well in classifying spam messages with a precision score above 90% and 80%, while the others failed to reach this benchmark. Our analysis of the machine learning model and SMS anti-spam ecosystem highlights the limitations of current anti-spam developments and potential research directions. We argue that SMS spam continues to pose a significant challenge, necessitating further research to develop systems capable of effectively addressing the evasion techniques employed by spammers and safeguarding the general public against SMS spam. serves to underscore the limitations of current anti-spam measures and advocate for additional research to develop enhanced detection methods.

FEATURE EXTRACTION MODELS

WORD2VEC : It uses a neural network to learn word associations from large corpora and generate a vector representation for words or tokens and captures semantic similarity between words. Word2Vec creates a vector space with each unique word in the corpus, often with several hundred dimensions, such that words with similar contexts in the corpus are close to one another in the space. Moreover, the vectors for unknown words are randomly initialized using a generic normal distribution.

GLOVE : In comparison to Word2Vec that utilises a window to establish local context, GloVe uses data from the whole text corpus to create an explicit word-context leveraging words' co-occurrence matrices. Each word in this method is represented by a high-dimensional vector and trained using the surrounding words throughout a large corpus.

FASTTEXT : It is an extension of Word2Vec with improved efficiency and effectiveness and addresses the out-of-vocabulary issues associated with Word2Vec feature model. Many other representations of word embeddings ignore the morphology of words by assigning a distinct vector to each word, which is addressed by fastText.

ELMO : It is a bidirectional Language Model (biLM) that extracts multi-layered word embeddings. ELMo word vectors address Polysemy, where a word has several meanings. These word vectors are pre-trained, deep bidirectional language model (biLM) functions. They can be easily integrated into current models and progress complex NLP tasks like question answering, textual entailment, and sentiment analysis.

BERT : It is a language model with a powerful transformer-based architecture and it's core principle- attention, that has demonstrated state-of-the-art performance on different NLP tasks. developed by Google. Generally, language models read the input sequence in one direction: either left to right or right to left. This kind of one-directional training works well when the aim is to predict or generate the next word. But in order to have a deeper sense of language context, BERT uses bidirectional training. BERT applies the bidirectional training of Transformer to language modeling and learns the text representations. BERT generates a language model depending on its context, capturing lexical, semantic and grammatical features.

MLCLASSIFIERS

PU IMPLEMENTATION There are two primary methods for implementing PU learning. These include PU bagging and a two-step process. The PU bagging is an ensemble of ensembles; training many ensemble classifiers in parallel. Each ensemble balances the classes according to the size of the positive class. On the other hand, the two-step method is a more complicated way to learn about PU. It uses ML techniques to change the labels of data while training and takes longer time to train. Therefore, we decided to opt the PU bagging approach (quicker than normal ensemble methods) and implemented it with random forest (RF) classifier.

LSTM : Due to the vanishing gradient issue, simple RNNs are incapable of representing longer contextual connections. They have mostly been supplanted by so-called long short term neural networks (LSTMs), which are similar to RNNs but can capture the lengthier context included in texts.

BIDIRECTIONAL LSTM (BILSTM) : LSTMs it can only process the unidirectional sequences. Consequently, state-of-the-art techniques

based on LSTMs evolved into so-called bidirectional LSTMs that can read the context left to right as well as right to left.

BIDIRECTIONAL GATED RECURRENT UNIT (BIGRU) : These are a type of bidirectional recurrent neural networks with only the input and forget gates. It enables the utilisation of knowledge from prior and subsequent time steps to create predictions about the present state.

DISTILBERT : The DistilBERT is a lighter, faster, and more affordable version of BERT. It learns a distilled (approximate) version of BERT that retains 95% of the performance but uses half the parameters. Specifically, it lacks token-type embeddings, a pooler, and preserves just half of Google's BERT's layers. DistilBERT makes use of a method called distillation to resemble Google's BERT, i.e. replacing the huge neural network with a smaller one. Following training a big neural network, the network's whole output distributions can be approximated using a smaller network.

ROBUSTLY OPTIMIZED BERT APPROACH (ROBERTA) : It is a retraining of BERT that incorporates an enhanced training process, 1,000% more data, and significantly more computational capacity. To improve the training approach, RoBERTA omits the Next Sentence Pre diction (NSP) task from BERT's pre-training and introduces dynamic masking, in which the masked token fluctuates between training epochs. Additionally, it was demonstrated that bigger batch sizes were more effective during the training process.

CROSS-LINGUAL LANGUAGE MODEL (XLM) : XLM is an enhanced version of BERT suggested by Facebook AI to deliver state-of-the-art results in a variety of natural language processing tasks, most notably cross-lingual classification, supervised and unsupervised machine translation. XLM employs a dual-language training mechanism in conjunction with BERT to acquire knowledge about the relationships between words in multiple languages. When a pre-trained model is used to initialize the translation model, the model out performs other models in a cross-lingual classification problem and considerably improves machine translation.

To highlight the changing characteristics of SMS spam, we introduced and characterized a large new SMS dataset. We used the dataset to benchmark the performance and robustness of several machine-learning-based models and the anti-spam ecosystem for detecting SMS spam. The results showed that all of the machine learning based models efficiently identified legitimate SMS messages(hams Mses). However, only a few deep learning models and anti-spam text apps performed well in classifying spam messages with a precision score above 90% and 80%, while the others failed to reach this benchmark. Our analysis of the machine learning model and SMS anti-spam ecosystem highlights the limitations of current anti-spam developments and potential research directions. We argue that SMS spam continues to pose a significant challenge, necessitating further research to develop systems capable of effectively addressing the evasion techniques employed by spammers and safeguarding the general public against SMS spam.Our dataset, available at https://github.com/smspamresearch/spstudy, and analysis serves to underscore the limitations of current anti-spam measures and advocate for additional research to develop enhanced detection methods.

Additionally, the integration of real-time feedback from users, the incorporation of adversarial training, and continual retraining of models will be crucial in maintaining the effectiveness of SMS spam filtering systems. Ultimately, the goal should not just be to create models that can detect spam, but to develop systems that can anticipate and adapt to new techniques as they emerge, providing users with a continuously improving defense against spam messages.

In conclusion, this study provides valuable insights into the ongoing challenge of SMS spam detection. While machine learning has proven to be a powerful tool in combating spam, the race between spammers and spam filters is far from over. As evasive techniques continue to evolve, machine learning models must adapt, improve, and remain for

even more robust and effective SMS spam detection systems, but it will require a concerted effort to stay one step ahead of those who seek to evade detection.

REFERENCES

[1] J. Buchanan and A. J. Grant, "Investigating and prosecuting Nigerian fraud," U.S. Att'ys Bull., vol. 49, pp. 39–47, Nov. 2001.

[2] L. Zhang, J. Zhu, and T. Yao, "An evaluation of statistical spam filtering techniques," ACM Trans. Asian Lang. Inf. Process., vol. 3, no. 4, pp. 243–269, Dec. 2004.

[3] G. L. Wittel and S. F. Wu, ‘‘On attacking statistical spam filters,’’ in Proc. CEAS, 2004.

[4] T. A. Almeida, J. M. G. Hidalgo, and A. Yamakami, ‘‘Contributions to the study of SMS spam filtering: New collection and results,’’ in Proc. 11th ACM Symp. Document Eng., Sep. 2011, pp. 259–262.

[5] T. Almeida, J. M. Hidalgo, and T. Silva, ‘‘Towards SMS spam filtering: Results under a new dataset,’’ JiSS, vol. 2, no. 1, pp. 1–18, 2013.

[6] M. Gupta, A. Bakliwal, S. Agarwal, and P. Mehndiratta, ‘‘A comparative study of spam SMS detection using machine learning classifiers,’’ in Proc. 11th Int. Conf. Contemp. Comput. (IC3), Aug. 2018, pp. 1–7.

[7] S. Rojas-Galeano, ‘‘Using BERT encoding to tackle the mad-lib attack in SMS spam detection,’’ 2021, arXiv:2107.06400.

[8] FCC. (2022). The Top Text Scams of 2022. Accessed: Oct. 8, 2023. [Online].

[9] ACCS. (2022). Accs Scam Statistics. [Online]. Available: https://www. scamwatch.gov.au/scam-statistics

[10] M. A. Abid, S. Ullah, M. A. Siddique, M. F. Mushtaq, W. Aljedaani, and F. Rustam, ‘‘Spam SMS filtering based on text features and supervised machine learning techniques,’’ Multimedia Tools Appl., vol. 81, no. 28, pp. 39853–39871, Nov. 2022.

[11] I. Ahmed, R. Ali, D. Guan, Y.-K. Lee, S. Lee, and T. Chung, ‘‘Semi_supervised learning using frequent itemset and ensemble learning for SMS classification,’’ Expert Syst. Appl., vol. 42, no. 3.

[12] C. Oswald, S. E. Simon, and A. Bhattacharya, ‘‘Spot Spam: Intentions analysis-driven SMS spam detection using BERT embeddings,’’ ACM Trans. Web, vol. 16, no. 3, pp. 1–27, Aug. 2022.

[13] S. Y. Yerima and A. Bashar, ‘‘Semi-supervised novelty detection with one class SVM for SMS spam detection,’’ in Proc. 29th Int. Conf. Syst., Signals Image Process. (IWSSIP), Jun. 2022, pp. 1–4.

[14] S. Tang, X. Mi, Y. Li, X. Wang, and K. Chen, ‘‘Clues in tweets: Twitter guided discovery and analysis of SMS spam,’’in Proc. AC MSI GSA CConf. Comput. Commun. Secur., Nov. 2022, pp. 2751–2764.

[15] A. van der Schaaf, C.-J. Xu, P. van Luijk, A. A. van't Veld, J. A. Langendijk, and C. Schilstra, ‘‘Multivariate modeling of complications with data driven variable selection: Guarding against overfitting and effects of data set size,’’ Radiotherapy Oncol., vol. 105, no. 1, pp. 115–121, Oct. 2012.

[16] T. Xia and X. Chen, ‘‘A discrete hidden Markov model for SMS spam detection,’’ Appl. Sci., vol. 10, no. 14, p. 5011, Jul. 2020.

[17] S. M. Abdulhamid, M. S. A. Latiff, H. Chiroma, O. Osho, G. Abdul-Salaam, A. I. Abubakar, and T. Herawan, ‘‘A review on mobile SMS spam filtering techniques,’’ IEEE Access, vol. 5, pp. 15650–15666, 2017.

[18] A. Narayan and P. Saxena, ‘‘The curse of 140 characters: Evaluating the efficacy of SMS spam detection on Android,’’ in Proc. 3rd ACM Workshop Security. Privacy Smartphones Mobile Devices, Nov. 2013, pp. 33–42.

[19] A. A. Al- Hasanand E.-S.-M . El – Alfy ,‘‘Dendritic cell algorithm
for mobile phone spam filtering,’’ Proc. Computer. Sci ,vol. 52,
pp. 244–251,Jan.2015.

[20] J. Li, S. Ji, T. Du, B .Li, and T. Wang ,‘‘Text Bugger: Generating
adversarial text against real-world applications,’’ 2018, arXiv:1812.05271.

[20] W. Wang, R. Wang, L.Wang,Z.Wang,andA.Ye,‘‘Towards a robust deep neural network in texts: A survey,’’ 2019, arXiv:1902.07285.

[21] A. Huq and M. Tasnim Pervin, ‘‘Adversarial attacks and defense on texts: A survey,’’ 2020, arXiv:2005.14108.

[22] J. Gao, J. Lanchantin, M. L. Soffa, and Y. Qi, ‘‘Black-box generation of adversarial text sequences to evade deep learning classifiers,’’ in Proc. IEEE Secur. Privacy Workshops (SPW), May 2018, pp. 50–56.

[23] (2023). Action Fraud. Accessed: Oct. 6, 2023. [Online]. Available: https://www.actionfraud.police.uk/

[24] (2012). UCI Machine Learning Repository—SMS Spam Collection Data Set. [Online]. Available: https://archive.ics.uci.edu/ml/datasets/sms

[25] J. Morris, E. Lifland, J. Y. Yoo, J. Grigsby, D. Jin, and Y. Qi, ‘‘TextAttack: A framework for adversarial attacks, data augmentation, and adversarial training in NLP,’’ in Proc. Conf. Empirical Methods Natural Lang. Processing: Syst. Demonstrations, 2020, pp. 119–126.

[26] F. Salahdine and N. Kaabouch, ‘‘Social engineering attacks: A survey,’’ Future Internet, vol. 11, no. 4, p. 89, Apr. 2019.

[27] A. Costello, Punycode: A Bootstring Encoding of Unicode for Internation alized Domain Names in Applications (IDNA), document RFC3492,2003.

[28] J. M. Gómez Hidalgo, G. C. Bringas, E. P. Sánz, and F. C. García, ‘‘Content based SMS spam filtering,’’in Proc. ACMSymp. Document Eng., Oct. 2006, pp. 107–114.

[29] T. Chen and M.-Y. Kan, ‘‘Creating a live, public short message service corpus: The NUSSMScorpus,’’Lang.Resour. Eval., vol. 47, pp. 299–335, Aug. 2012.

[30] A. Ghourabi and M. Alohaly, ‘‘Enhancing spam message classification and detection using transformer-based embedding and ensemble learning,’’ Sensors, vol. 23, no. 8, p. 3861, Apr. 2023. [33] N. A. M. Ariff, F. M. Husni, M. Z. Mas'ud, N. Bahaman, and E. Hamid, ‘‘A comparison of generative and discriminative classifiers in SMS spam classification,’’ in Proc. ICTEC, 2022.

[31] Z. Ilhan Taskin, K. Yildirak, and C. H. Aladag, ‘‘An enhanced random forest approach using Co Clust clustering: MIMIC-III and SMS spam collection application,’’ J. Big Data, vol. 10, no. 1, p. 38, Mar. 2023. [35] S. Mishra and D. Soni, ‘‘SMS phishing dataset for machine learning and pattern recognition,’’ in Proc. SoCPaR, 2023, pp. 597–604.

[32] Z. Liu, S. Ni, A. Aw, and N. Chen, ‘‘Singlish message para phrasing : A joint task of creole translation and text normalization,’’ in Proc. 29th Int. Conf. Comput. Linguistics, 2022, pp. 3924–3936.

[33] S. Y. Chow and F. Bond, ‘‘Singlish where got rules one? Constructing a computational grammar for Singlish,’’ in Proc. 13th Lang. Resour. Eval. Conf., 2022, pp. 5243–5250.

[34] Hudeček, L.-P. Schaub, D. Stancl, P. Paroubek, and O. Dušek, ‘‘A uni fying view on task-oriented dialogue annotation,’’ in Proc. LREC, 2022, pp. 1286–1296.

[35] S. A. Chaturvedi and L. Purohit, ‘‘Feature selection-based spam detection system in SMS and email domain,’’ in Proc. ICSADL, 2023, pp. 37–52.

[36] H. Nakano, D. Chiba ,T. Koide, N. Fukushi, T. Yagi, T. Hariu, K. Yoshioka, and T. Matsumoto, ‘‘Canary in Twitter mine: Collecting phishing reports from experts and non-experts,’’ 2023, arXiv:2303.15847.

[37] K. Mathew and B. Issac, ‘‘Intelligent spam classification for mobile text message,’’ in Proc. Int. Conf. Comput. Sci. Netw. Technol., vol. 1, Dec. 2011, pp. 101–105.

[38] P.K.Roy,J.P.Singh,andS.Banerjee,‘‘DeeplearningtofilterSMSspam,’’ Future Gener. Comput. Syst., vol. 102, pp. 524–533, Jan. 2020.

[39] G. Jain, M. Sharma, and B. Agarwal, ‘‘Optimizing semantic LSTM for spam detection,’’ Int. J. Inf. Technol., vol. 11, no. 2, pp. 239–250, Jun. 2019.

[40] N. H .Imamand V. G. Vassilakis,“A survey of attacks against Twitter spam detectors in an adversarial environment,” Robotics, vol. 8, no. 3, p. 50, Jul. 2019.

[41] S. Webb, S. Chitti, and C. Pu, “An experimental evaluation of spam filter performance and robustness against attack,” in Proc. Int. Conf. Collaborative Computing, Netw., Appl. Worksharing, 2005, p. 8.

[42] J. Graham-Cumming. (2004). How to Beat a Bayesian Spam Filter. [Online]. Available: http://www.spamconference.org.

[43] H. Li, S. Shan, E. Wenger, J. Zhang, H. Zheng, and B. Y. Zhao, “Blacklight: Scalable defense for neural networks against query-based black-box attacks,” 2020, arXiv:2006.14042.

[44] P. Faltstrom, P. Hoffman, and A. Costello, “Internationalizing domain names in applications (IDNA),” Tech. Rep. rfc3490, 2003.

[45] M. Danilk. (2023). Langdetect. [Online]. Available: https://pypi.org /project/langdetect/

[46] Google Inc. (2023). Google Language Detection API. [Online]. Available: https://cloud.google.com/translate/docs/basic/detecting-language

[47] A.Kay,“Tesseract: An open-source optical character recognition engine,” Linux J., vol. 2007, no. 159, p. 2, 2007.

[48] S. Bansal. (2023). GitHub—Sanshbansal/Spam Detection Analytics Tool. [Online]. Available: https://github.com/saranshbansal/spam-detection analytics-tool

[49] M. N. U. Hasan. (2023). SMS Spam Prediction. [Online]. Available: https://github.com/mohammadnoorulhasan/sms-spam-prediction

[50] A. Bhowmick and S. M. Hazarika, “Machine learning for e-mail spam filtering: Review, techniques and trends,” 2016, arXiv:1606.01042.

[51] (2023).Py spell checker. [Online]. Available: https: // pypi. org/project /py spell checker/

[52] D. Eleyan, A. Othman, and A. Eleyan, “Enhancing software comments readability using flesch reading ease score,” Information, vol. 11, no. 9, p. 430, Sep. 2020.

[53] S. Chhabra, A. Aggarwal, F. Benevenuto, and P. Kumaraguru, “Phi.sh/$oCiaL: The phishing landscape through short URLs,” in Proc. 8th Annu. Collaboration, Electron. Messaging, Anti-Abuse Spam Conf., Sep. 2011, pp. 92–101.

[54] E. Loper and S. Bird, “NLTK: The natural language toolkit,” 2002, .

[55] F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” Journal of Machine Learning Research, vol. 12, pp. 2825–2830, Nov. 2011

[56] Y. Zhang, R. Jin, and Z.-H. Zhou, “Understanding bag-of-words model: A statistical framework,” Int. J. Mach. Learn. Cybern., vol. 1, nos. 1–4, pp. 43–52, Dec. 2010.