Investigating Evasive Techniques in SMS Spam Filtering: A Comparative Analysis of Machine Learning Models

Neha Khare, Prof. Arpana Jaiswal

Abstract:

The proliferation of SMS spam poses a growing threat in the domain of cybersecurity, often resulting in financial scams, privacy breaches, and negative user experiences. Although machine learning algorithms have become central to spam detection, adversaries continue to refine their evasion tactics—employing methods like character obfuscation, altered vocabulary, and adversarial perturbations. These evolving techniques challenge the effectiveness of traditional rule-based and standard machine learning systems, highlighting the demand for more resilient and adaptive spam-filtering frameworks.

This review delivers a thorough exploration of modern evasion strategies in SMS spam and critically evaluates the performance of machine learning and deep learning models in detecting such threats. We investigate conventional classifiers—such as Naïve Bayes, Support Vector Machines (SVM), and Decision Trees—alongside cutting-edge deep learning models like Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Transformer architectures including BERT and GPT. Additionally, ensemble learning approaches that combine the strengths of multiple models are discussed to assess their contribution to improved detection accuracy.

To support ongoing research, we spotlight key challenges that remain unaddressed in SMS spam filtering, such as the demand for real-time and privacy-preserving solutions, and language-independent filtering mechanisms. We also emphasize the necessity for ethically grounded AI systems that can minimize inherent biases in spam classification.

Keywords: SMS spam, evasion strategies, machine learning, deep learning, adversarial examples, NLP, robust classifiers, spam filtering models

1. Introduction

The task of filtering SMS spam remains a persistent issue within the cybersecurity landscape, primarily due to the adaptive nature of spammer techniques. As adversaries continuously refine their methods to bypass detection, the reliability of existing filtering mechanisms steadily declines. This paper reviews contemporary research focused on evasion strategies used in SMS spam and examines their implications for detection systems, with the aim of identifying knowledge gaps and proposing avenues for future investigation.

With billions of messages exchanged globally each day, SMS continues to be a prevalent and accessible form of communication. Unfortunately, this ubiquity also makes it a preferred channel for cybercriminals, who exploit it to disseminate deceptive messages ranging from phishing attacks to fake promotions. Such spam not only facilitates identity theft and financial fraud but also contributes to user frustration and puts additional load on network infrastructure.

Modern spam detection efforts have largely transitioned from static rule-based systems to machine learning-based solutions. These models have demonstrated notable success; however, spammers actively adopt novel tactics to circumvent detection. Common evasion techniques include:

- **Character obfuscation:** Replacing letters with symbols or digits (e.g., “Fr33” instead of “Free”)
- **Word substitution:** Using phonetically similar alternatives (e.g., “pr1ze” for “prize”)
- **Text distortion:** Inserting random spaces or characters (e.g., “W i n a c a r”)
- **URL manipulation:** Shortening or redirecting malicious URLs to mask intent

Such adaptive strategies require spam detection frameworks to be equally dynamic and capable of identifying nuanced manipulations.

Interestingly, this pattern of adaptation is also evident across platforms such as Twitter, where spammers continuously evolve to evade detection algorithms. Findings from social media spam detection can therefore offer valuable insights for enhancing SMS spam filtering models. As noted by Yang et al. (2011), spammer behavior is highly responsive to advances in filtering technologies, reinforcing the need for continuous innovation in detection methods.

2. Overview of SMS Spam and Evasion Techniques

SMS spam typically involves unsolicited and often deceptive messages sent to individuals with malicious or promotional intent. These messages can be used to execute financial fraud, harvest personal data, or promote products without consent.

Unlike email spam, SMS-based spam is particularly invasive due to the personalized nature of mobile communication and limited filtering capabilities on mobile platforms.

Several researchers have highlighted the inherent difficulties in effectively filtering SMS spam. Among the challenges are the brevity of SMS messages, which reduces the availability of contextual cues, and the limited size and diversity of publicly available datasets (Almeida et al., 2011). The introduction of newer, more diverse datasets has allowed for better benchmarking of models, especially in assessing the resilience of classifiers such as SVMs against manipulated or adversarial text.

Another dimension of complexity arises from the susceptibility of machine learning models to adversarial exploitation. For example, data poisoning—where spam messages are subtly embedded within training data to mislead the model—can significantly compromise classifier performance (Biggio et al., 2014). Defensive techniques such as outlier detection and noise filtering (Delany et al., 2012) have shown promise in mitigating these effects, but ongoing research is essential to maintain robustness.

2.1 Categories of SMS Spam

Spam messages transmitted via SMS can take multiple forms, each designed to mislead recipients and extract personal or financial information. Common types include:

1. Smishing (SMS Phishing): Deceptive messages crafted to obtain sensitive details such as banking information or login credentials.
2. Fake Lottery and Prize Scams: Messages falsely claiming the recipient has won a reward, often requesting payment or account verification.
3. Unsolicited Promotions: Bulk advertising messages sent without user permission, typically to promote commercial services or products.
4. Fraudulent Loans and Investment Offers: Claims of quick loans, lucrative stock tips, or crypto opportunities aimed at exploiting users.
5. Malware Links: Messages containing links that redirect to harmful websites capable of installing malware on the device.

2.2 Evasion Strategies Employed by Spammers

Modern spammers employ a variety of tactics to evade detection by spam filters. These techniques exploit weaknesses in language models, keyword filters, and URL analysis:

- Character Substitution: Using special characters or numbers in place of common letters to avoid keyword detection.
- Lexical Alteration: Replacing high-risk words with phonetically similar or synonymous alternatives.
- Intentional Misspellings and Spacing: Adding spaces or typographical errors to disrupt text matching algorithms.
- Masked URLs: Utilizing link shorteners or redirection to disguise the final destination of malicious URLs.
- Vague or Ambiguous Phrasing: Crafting messages that lack context, making it harder for classifiers to identify spam intent.
- Template Variation: Generating multiple message versions using templates to avoid pattern-based detection.
- Adversarial Perturbations: Creating AI-generated adversarial texts that are specifically designed to confuse and bypass spam detection algorithms.

2.3 Key Challenges in Evasive Spam Detection

Even with the advances in artificial intelligence, spam detection faces several obstacles:

- Rapid Evolution of Attacks: Spammers constantly adapt, outpacing traditional detection mechanisms.
- Data Scarcity: Existing datasets often lack adversarial examples, which limits model generalization.
- Linguistic Manipulation: Sophisticated language distortions are difficult to detect using standard NLP tools.
- Need for Real-Time Filtering: Effective spam detection must operate in real time without causing message delivery delays.

3. Machine Learning-Based Techniques for SMS Spam Detection

With the ever-changing landscape of SMS spam techniques, there arises a pressing need for intelligent and adaptable spam filtering mechanisms. Conventional rule-based filters and simple keyword-matching strategies are proving to be increasingly ineffective in the face of modern evasion methods employed by spammers. In this context, machine learning (ML) approaches have gained considerable attention, offering data-driven and evolving solutions for identifying spam content with greater accuracy.

This section delves into the various machine learning methodologies employed for SMS spam detection. It highlights their working principles, key advantages, known limitations, and their overall suitability in tackling the dynamic nature of spam evasion tactics.

3.1 Role of Machine Learning in SMS Spam Filtering

Machine learning enables systems to make informed decisions by learning from data, rather than relying on rigid rules. In the case of SMS spam detection, ML models are trained using labelled datasets, where each message is categorized as either spam or ham (legitimate). The standard procedure followed in implementing ML-based spam filters typically involves the following stages:

1. Data Acquisition: Collection of datasets containing classified SMS messages, clearly labelled as spam or non-spam.
2. Preprocessing of Data: This involves cleaning the text, removing noise, and converting it into a suitable format for analysis.
3. Feature Extraction: Deriving meaningful indicators such as term frequencies, character-level patterns, and TF-IDF scores.
4. Model Development: Training classification algorithms on preprocessed data.
5. Prediction and Performance Evaluation: Testing the model on unseen data to evaluate its accuracy, precision, and generalisation ability.

This structured pipeline ensures that the models evolve with time and can adapt to emerging patterns in spam messaging.

3.2 Broad Categories of Machine Learning Models

The models used for detecting SMS spam through machine learning can be classified under three broad categories:

1. Conventional Machine Learning Models: These are statistical models that work well with structured features and are computationally efficient.
2. Deep Learning Models: These models, based on neural networks, are capable of capturing complex patterns in unstructured text.
3. Ensemble and Hybrid Models: These involve a combination of multiple algorithms, enhancing the predictive strength and robustness of the overall system.

Each of these categories contributes uniquely to the task of spam detection, and their application depends on the nature of the dataset and the desired level of accuracy.

3.3 Traditional Machine Learning Models

Conventional machine learning classifiers have long been the foundation of spam filtering systems. Their simplicity, ease of interpretation, and low resource requirements make them suitable for real-time deployment, particularly on mobile devices.

3.3.1 Naïve Bayes Classifier (NBC)

The Naïve Bayes classifier is one of the most commonly applied algorithms in text classification, especially in spam detection tasks. It operates on the principles of **Bayes' theorem**, with an assumption that the features used for classification are mutually independent—hence the term 'naïve'.

Working Principle:

- The classifier calculates the **posterior probability** of a message being spam based on the frequency and presence of specific words.
- The mathematical formulation is as follows:

$$P(spam|message) = \frac{P(Message|spam).P(Spam)}{P(message)}$$

- Despite the simplifying assumption of feature independence, Naïve Bayes performs remarkably well for short text messages, where limited contextual information is available.

Merits:

- Fast and efficient even with large datasets.
- Requires minimal computational resources.
- Offers good performance on balanced datasets.

Limitations:

- Assumption of independence among features may not hold true in all cases.
- May underperform on datasets with complex linguistic variations or adversarial perturbations.

3.3.2 Support Vector Machines (SVM)

Support Vector Machines (SVM) are among the most widely used supervised learning algorithms for binary classification tasks, such as distinguishing between spam and legitimate (ham) messages. The core idea behind SVM is to find the most optimal separating boundary between the two classes in a high-dimensional feature space.

Working Principle:

- SVM constructs a decision boundary, also known as a hyperplane, that ensures the maximum possible margin between the spam and ham classes.
- It leverages kernel functions—such as the linear kernel and the Radial Basis Function (RBF)—to project the input text data into a higher-dimensional space, where complex patterns can be more easily separated.

Advantages:

- Offers high classification accuracy, especially in binary settings.
- Performs well on text datasets with high dimensionality.

- Suitable for moderate-sized datasets, where computational efficiency is still manageable.

Limitations:

- Can become computationally intensive for very large datasets.
- The performance may deteriorate in the presence of noisy or adversarially perturbed data.

3.3.3 Decision Trees and Random Forests

Decision Trees are rule-based models that make classifications based on a hierarchical structure of conditions (if-else rules). On the other hand, Random Forest is an ensemble technique that combines the outputs of multiple decision trees to achieve better performance and generalization.

Key Characteristics:

- Decision Trees are intuitive and easy to interpret, where each internal node represents a feature-based decision.
- Random Forest reduces variance by aggregating the predictions of several decision trees, thereby mitigating overfitting.

Advantages:

- Capable of capturing **non-linear patterns** in the data.
- Random Forests provide **improved accuracy** and are **less prone to overfitting** due to ensemble averaging.

Limitations:

- Individual decision trees may overfit the training data.
- For larger datasets, especially with many features, the computational burden increases significantly.

3.4 Deep Learning Approaches for SMS Spam Detection

Deep learning models have revolutionized spam detection by enabling automated feature extraction and the ability to capture intricate patterns in the data. These models, inspired by the structure of the human brain, are particularly effective in handling evasive spam techniques that manipulate text in subtle and complex ways.

3.4.1 Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM)

Recurrent Neural Networks (RNNs) and their improved variant Long Short-Term Memory (LSTM) networks are designed to handle **sequential data**, making them suitable for processing natural language text such as SMS messages.

Working Mechanism:

- RNNs analyze text word by word, maintaining a contextual memory of previous tokens to inform future predictions.
- LSTMs overcome the limitations of standard RNNs—particularly the vanishing gradient problem—by incorporating memory cells that selectively retain or discard information over long sequences.

Advantages:

- Effectively captures temporal dependencies and sequential structure in spam messages.
- Performs better in identifying obfuscated or modified words compared to traditional ML models.

Limitations:

- Demands significant computational resources for training.
- Requires a large volume of labelled data to achieve high accuracy.

3.4.2 Transformer-Based Models (e.g., BERT, GPT)

Recent advancements in natural language processing have led to the development of Transformer-based architectures, such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer). These models rely on self-attention mechanisms to understand contextual relationships between words in a sentence.

Advantages:

- Extremely effective in detecting adversarial or syntactically complex messages.
- Can model bidirectional context, making them superior to traditional RNNs in understanding nuances in language.

Limitations:

- Require high-end computational infrastructure (e.g., GPUs or TPUs).
- Need to be fine-tuned on domain-specific datasets (like SMS) to achieve optimal spam filtering performance.

3.5 Hybrid and Ensemble Learning Models

To enhance accuracy and adaptability in spam detection, researchers often integrate multiple learning models through hybrid or ensemble techniques. These methods harness the complementary strengths of individual algorithms, leading to improved overall performance.

3.5.1 Stacking and Boosting Techniques

- Stacking involves training several base classifiers and then using a meta-classifier to combine their outputs. This approach benefits from the diversity of the base learners.
- Boosting, as in methods like XGBoost or AdaBoost, improves performance by focusing on examples that are difficult to classify correctly, thereby incrementally enhancing weaker models.

3.5.2 Hybrid NLP Techniques

By integrating traditional feature extraction methods like TF-IDF with deep learning models, a robust pipeline can be created. This hybrid strategy improves resilience against evasive spam tactics, ensuring that both surface-level and contextual patterns are captured effectively.

3.6 Comparison of Machine Learning Models for SMS Spam Detection

Model	Strengths	Weaknesses
Naïve Bayes	Simple, fast, good for small datasets	Struggles with adversarial text
SVM	High accuracy, effective in high-dimensional space	Computationally expensive
Random Forest	Handles nonlinear relationships, reduces overfitting	Slower for large datasets
LSTM	Captures text sequence and context	Requires large labeled data
BERT	Handles adversarial text, state-of-the-art accuracy	High computational power required
Ensemble Models	Improves accuracy by combining models	More complex to implement

4. Adversarial Attacks in SMS Spam Detection

With the advancement of machine learning (ML)-based spam detection systems, spammers have resorted to more deceptive strategies, often referred to as adversarial attacks. These are designed to mislead detection models by subtly altering the content of spam messages using obfuscation, character replacements, misspellings, or even semantically modified phrases.

To counter such strategies, researchers have proposed innovative techniques aimed at making spam filters more resistant to adversarial manipulation.

4.1 Understanding Adversarial Attacks in SMS Filtering

Adversarial attacks refer to the deliberate modification of spam messages with the intention of deceiving machine learning classifiers. These techniques exploit the vulnerabilities of spam detection algorithms, causing them to misidentify spam as legitimate messages (ham).

4.1.1 Common Traits of Adversarial SMS Spam

- Text Distortion: Introduction of typographical changes such as letter substitution, character insertion, or misspellings.
- Obfuscation Techniques: Replacement of common words with lookalike characters (e.g., “fr33” instead of “free”).
- Semantic Rewriting: Using synonyms, paraphrasing, or inserting irrelevant words to deceive keyword-based filters.
- Invisible Characters: Use of hidden Unicode characters to disrupt tokenization.
- Contextual Mixing: Blending spam-related words with legitimate text to lower the chances of detection.

4.2 Types of Adversarial Attacks

Studies in adversarial machine learning have revealed various methods by which spammers subvert ML classifiers. Notably, frameworks such as **DISP** (Zhou et al., 2019) provide means to identify malicious perturbations without altering the core architecture of NLP models.

Additionally, the work of Biggio et al. (2014) has been instrumental in proposing structured evaluation frameworks to assess the security and robustness of classifiers under adversarial stress.

4.3 Machine Learning Vulnerabilities in SMS Spam Detection

The susceptibility of spam detection models arises from certain inherent limitations:

- Overdependence on Keywords: Many models rely heavily on keyword matching, making them vulnerable to variations in spelling or vocabulary.
- Sensitivity to Specific Features: ML algorithms trained on static datasets struggle to generalise against unseen evasion techniques.
- Lack of Adaptability: Most spam filters are not designed to learn or evolve in real-time.
- Insufficient Contextual Understanding: Traditional models often fail to comprehend the semantic meaning behind obfuscated or paraphrased text.

4.4 Countermeasures Against Adversarial Attacks

To enhance model resilience, several countermeasures have been proposed:

4.4.1 Adversarial Training

- **Approach:** Include adversarially modified spam messages in the training dataset.
- **Benefit:** Improves the model's ability to detect manipulated or deceptive text inputs.

4.4.2 Character-Level and Subword Tokenization

- **Approach:** Use tokenization methods like Byte Pair Encoding (BPE) to capture subword units.
- **Benefit:** Effective in handling misspellings, character substitutions, and informal language.

4.4.3 Context-Aware Embeddings (BERT, GPT)

- **Approach:** Employ pretrained transformer-based models that understand deeper contextual semantics.
- **Benefit:** Capable of identifying spam even when traditional keywords are absent or altered.

4.4.4 Anomaly Detection Techniques

- **Approach:** Apply unsupervised models to detect rare patterns in incoming messages.
- **Benefit:** Useful for flagging previously unseen adversarial spam variants.

4.4.5 Ensemble Learning Models

- Approach: Use a combination of classifiers (e.g., boosting, bagging, stacking) to improve robustness.
- Benefit: Minimises the chance of a single point of failure and improves generalisation.

5. Dataset Challenges and Benchmarking

The effectiveness of any spam detection model significantly hinges on the **quality, size, and diversity** of the dataset it is trained on. Issues like **data scarcity, class imbalance, language limitations**, and lack of updated samples hamper the practical deployment of such models.

5.1 Challenges in Dataset Development and Preprocessing

5.1.1 Limited Public Access to Real-World Datasets

- Regulatory frameworks like GDPR and CCPA restrict access to actual SMS communications.
- Datasets from telecom operators are often confidential and not shared publicly.
- Many open-source datasets are outdated and fail to reflect current spam trends.

5.1.2 Data Noise and Preprocessing Hurdles

- Real-world SMS includes emojis, URLs, abbreviations, and non-standard formats.
- Inconsistent labelling and incomplete samples introduce noise into the training process.

5.1.3 Linguistic Diversity and Code-Mixing

- Most available datasets are in English, whereas spam messages occur in multiple regional languages and dialects.
- The increasing use of code-mixed languages (like Hinglish) complicates spam detection.

5.2 Class Imbalance and Evolutionary Nature of Spam

5.2.1 Imbalanced Dataset Distribution

- A large gap between spam and ham message counts skews model training.
- Models tend to be biased toward the majority class, resulting in false negatives.

5.2.2 Concept Drift

- Spamming strategies evolve rapidly, rendering old datasets obsolete.
- Continuous updating of datasets is necessary to maintain model performance.

5.2.3 Limited Spam Categories

- Datasets are often confined to a few spam types such as lottery messages or advertisements.
- Modern spam involves phishing, fraud, and malware-requiring broader coverage.

5.3 Popular Benchmark Datasets

5.3.1 UCI SMS Spam Collection

- Messages: 5,574 (747 spam)
- Limitations: Dated and lacks language diversity.

5.3.2 SpamAssassin Corpus

- Messages: 9,324 (email-based)
- Limitations: Requires significant preprocessing to suit SMS format.

5.3.3 NUS SMS Dataset

- Messages: 33,952 (Singlish and English)
- Limitations: Limited to a specific linguistic demographic.

5.3.4 Kaggle SMS Spam Dataset

- Messages: 5,572
- Limitations: Small size and potential overlap with UCI dataset.

5.3.5 Proprietary Telecom Datasets

- Advantages: Real-world and large-scale.
- Limitations: Not accessible due to data privacy norms.

6. Future Research Directions

Looking ahead, there is a critical need for multilingual and region-specific SMS spam datasets that can capture evolving spam strategies. Research can also explore hybrid approaches, combining rule-based, content-based, and behavioural analysis models for holistic spam detection.

Furthermore, real-time learning mechanisms and adversarially robust models will play a key role in ensuring continued effectiveness against emerging threats. Collaborative efforts between academia, industry, and telecom providers can also accelerate dataset availability and model deployment.

7. Conclusion

The review of evasive SMS spam strategies and associated detection techniques illustrates that spam filtering is a continuously evolving challenge. While machine learning has enabled significant progress, adversarial spam tactics demand adaptive and intelligent countermeasures. Strengthening dataset diversity, improving model robustness, and ensuring real-world applicability are essential for future advances in this field.

By synthesising existing literature and highlighting key challenges, this paper aims to guide ongoing efforts in developing effective and context-aware SMS spam detection systems.

References:

1. Muñoz-González, Luis., Biggio, B.., Demontis, Ambra., Paudice, Andrea., Wongrassamee, Vasin., Lupu, Emil C.., & Roli, F.. (2017). Towards Poisoning of Deep Learning Algorithms with Back-gradient Optimization. Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security . http://doi.org/10.1145/3128572.3140451
2. Zhou, Yan., Kantarcioglu, Murat., Thuraisingham, B.., & Xi, B.. (2012). Adversarial support vector machine learning. , 1059-1067 . http://doi.org/10.1145/2339530.2339697
3. Yang, Chao., Harkreader, R.., & Gu, G.. (2011). Empirical Evaluation and New Design for Fighting Evolving Twitter Spammers. IEEE Transactions on Information Forensics and Security , 8 , 1280-1293 . http://doi.org/10.1109/TIFS.2013.2267732
4. Almeida, Tiago A.., Hidalgo, J. M. G.., & Yamakami, A.. (2011). Contributions to the study of SMS spam filtering: new collection and results. , 259-262 . http://doi.org/10.1145/2034691.2034742
5. Biggio, B.., Corona, Igino., Nelson, B.., Rubinstein, Benjamin I. P.., Maiorca, Davide., Fumera, G.., Giacinto, G.., & Roli, F.. (2014). Security Evaluation of Support Vector Machines in Adversarial Environments. ArXiv , abs/1401.7727 . http://doi.org/10.1007/978-3-319-02300-74
6. https://www.semanticscholar.org/paper/80a560f8e3a6bba85051ff8a418ed80f5cabd33f
7. Zhou, Yichao., Jiang, Jyun-Yu., Chang, Kai-Wei., & Wang, Wei. (2019). Learning to Discriminate Perturbations for Blocking Adversarial Attacks in Text Classification. ArXiv, abs/1909.03084. http://doi.org/10.18653/v1/D19-1496
8. Delany, Sarah Jane., Buckley, Mark., & Greene, Derek. (2012). SMS spam filtering: Methods and data. Expert Syst. Appl. , 39 , 9899-9908 . http://doi.org/10.1016/J.ESWA.2012.02.053
9. Yadav, Kuldeep., Kumaraguru, P.., Goyal, A.., Gupta, Ashish., & Naik, Vinayak. (2011). SMSAssassin: crowdsourcing driven mobile-based system for SMS spam filtering. , 1-6 . http://doi.org/10.1145/2184489.2184491
10. Dada, E.., Bassi, Joseph Stephen., Chiroma, H.., Abdulhamid, S.., Adetunmbi, A.., & Ajibuwa, O.. (2019). Machine learning for email spam filtering: review, approaches and open research problems. Heliyon , 5 . http://doi.org/10.1016/j.heliyon.2019.e01802
11. https://www.semanticscholar.org/paper/1eb1c5f369da4f90c8f763f778a21c49cc605117
12. Biggio, B.., Corona, Igino., Fumera, G.., Giacinto, G.., & Roli, F.. (2011). Bagging Classifiers for Fighting Poisoning Attacks in Adversarial Classification Tasks. , 350-359 . http://doi.org/10.1007/978-3-642-21557-537
13. Almeida, Tiago A.., Silva, Tiago P.., Santos, Igor., & Hidalgo, J. M. G.. (2016). Text normalization and semantic indexing to enhance Instant Messaging and SMS spam filtering. Knowl. Based Syst. , 108 , 25-32 . http://doi.org/10.1016/j.knosys.2016.05.001
14. Biggio, B.., Fumera, G.., & Roli, F.. (2014). Security Evaluation of Pattern Classifiers under Attack. IEEE

Transactions on Knowledge and Data Engineering , 26 , 984-996 . http://doi.org/10.1109/TKDE.2013.57

15. Adewole, K.., Han, Tao., Wu, Wanqing., Song, Houbing., & Sangaiah, A. K.. (2018). Twitter spam account detection based on clustering and classification methods. The Journal of Supercomputing, 76, 4802 - 4837. http://doi.org/10.1007/s11227-018-2641-x

16. Ibitoye, Olakunle., Abou-Khamis, Rana., elShehaby, Mohamed., Matrawy, A.., & Shafiq, M. O.. (2019). The Threat of Adversarial Attacks against Machine Learning in Network Security: A Survey. Journal of Electronics and Electrical Engineering. http://doi.org/10.37256/jeee.4120255738

17. https://www.semanticscholar.org/paper/eb774433ca5daaaa9f8c28530f7299411b2ea11e

18. Abdulhamid, S.., Latiff, Muhammad Shafie Abd., Chiroma, H.., Osho, Oluwafemi., Abdul-Salaam, Gaddafi., Abubakar, Adamu I.., & Herawan, T.. (2017). A Review on Mobile SMS Spam Filtering Techniques. IEEE Access , 5 , 15650-15666. http://doi.org/10.1109/ACCESS.2017.2666785