Spam Filtering in the Modern Era: A Review of Machine Learning, Deep Learning, and System Comparisons

Xinyi Wang
School of Electrical Engineering and Artificial Intelligence, Xiamen University Malaysia,
Jalan Sunsuria, Bandar Sunsuria, Malaysia

Keywords: Spam Filtering, Machine Learning, Deep Learning, Navie Bayes, Support Vector Machines, Random Forest, BERT, Convolutional Neural Networks, Concept Drift, LLM-Modified Spam.

Abstract: Spam poses a great challenge to Internet users, threatening their productivity, data security, and overall user experience. This review examines the current state of spam filtering systems, particularly those using Machine Learning (ML) and Deep Learning (DL) models. A comparative analysis of models such as Naive Bayes, Support Vector Machines (SVM), Random Forest (RF), Bidirectional Encoder Representations from Transformers (BERT), and Convolutional Neural Networks (CNNs) is conducted to assess their effectiveness and limitations in distinguishing spam from legitimate (ham) emails. Additionally, this study highlights emerging challenges, including Concept Drift and Large Language Model (LLM)-modified Spam, which necessitate the development of more adaptive and intelligent filtering solutions. The future trends indicate that the spam filtering system will gradually shift to Artificial Intelligence (AI)-driven automation and personalized, with lifelong learning models that continuously learn from new data leading the way. Such advancements are essential to counter sophisticated, hard-to-detect spam, like sophisticated LLM-crafted spam, like those crafted by large language models, ensuring a safer and more personalized email experience for users.

1 INTRODUCTION

In the digital age, spam has become a common problem, filling the inbox and posing a major threat to the safety and productivity of users. Spam usually contains phishing links, malware or misleading advertisements aimed at deceiving recipients or promoting unsolicited products. And as spammers continue to enhance their tactics, it is becoming more difficult for traditional filtering methods to keep up with the changing pattern of spam.

This article explores the development of spam filtering technology, focusing on the transition from early rule-based systems, such as SpamAssassin, to contemporary machine learning and deep learning methods. By comparing models, such as Navie Bayes, Support Vector Machine (SVM), Random Forest and Bidirectional Encoder Representations from Transformers (BERT), a comprehensive overview of the advantages and disadvantages of current spam filtering technologies is provided. In addition, this study addresses emerging challenges and future trends in spam detection, underscoring the necessity of adaptability and innovation in combating spam.

2 DEFINITION AND CHARACTERISTICS OF SPAM EMAILS

Spam emails are the unsolicited messages which are typically sent in bulk to a wide range of users. These emails generally contain phishing links, malware or unwanted advertisements, aiming to mislead recipients or promote products (Aragão Ferreira, Oliveira, Kuehne, Moreira, & Carpinteiro, 2021; Ghosh & Senthilrajan, 2023). Spam emails not only result in the loss of productivity but also leads to the exposure of inappropriate content to inappropriate audiences (Aragão et al., 2021).

There are three main types of spam emails. The first category comprises commercial advertisement emails, which are typically filled with promotional messages designed to sell products or services. The

Wang and X.
Spam Filtering in the Modern Era: A Review of Machine Learning, Deep Learning, and System Comparisons.
DOI: 10.5220/0013526000004619
In *Proceedings of the 2nd International Conference on Data Analysis and Machine Learning (DAML 2024)*, pages 451-458
ISBN: 978-989-758-754-2

second, phishing emails, impersonates reliable sources in an attempt to obtain credit card numbers or passwords. Finally, malware emails usually carry viruses or other harmful programs that can damage your device and compromise your data security once opened. (Kaddoura, Alfandi, & Dahmani, 2020).

Moreover, these emails often exhibit some common notable characteristics. First of all, they frequently use some specific keywords and phrases, which are eye-catching and likely related to promotions, investment opportunities or impractical promises, such as “free”, “urgent”, and “limited time offer”. Secondly, many spam emails contain images, and such images may hide harmful content or induce the user to click on the link. Besides, some spam emails will also contain suspicious links leading to fake websites or downloads. And additionally, attachments is also a common feature of spam, which may contain viruses or other malware.

3 DEVELOPMENT OF SPAM FILTERING TECHNIQUES

Spam filtering technologies have experienced a remarkable development in the past few decades. However, with the popularity of internet and the widespread application of emails, the quantity and types of spam emails are continuously increasing. As a consequence, developing effective spam filtering technology is becoming more and more important.

3.1 Early Filtering Techniques (Rule-Based Filtering)

The earliest spam filtering method was mainly a rule-based system, such as SpamAssassin. These systems applied more than 700 tests and assign the scores of spam emails based on a series of predefined rules. For example, the system would check for specific keywords in emails, the sender's reputation, and header information. Specifically, they combined technologies, including Bayesian filtering, blacklisting, and Domain Keys Identified Mail (DKIM) verification. This method can identify spam by analyzing the content and source of the email, so as to effectively filter out most of the obvious spam (Elliott, 2023). However, with the continuous evolution of spam methods, the traditional rule-based filtering technology has gradually become less flexible and accurate.

Although these early filtering systems were effective at the time, traditional rule-based filtering techniques have become increasingly limited in flexibility and accuracy. This is because spammers began using more complex strategies, such as disguising email content and altering sender addresses, which made it difficult for many filtering systems to effectively identify spam. Consequently, relying solely on rule-based methods has proven inadequate for addressing the increasingly sophisticated challenges posed by spam (Mohammad, 2020).

3.2 Transition to Modern Filtering Methods

As spammers are increasingly better at evading simple filtering rules, researchers began to turn to study more complex modern filtering methods. The emergence of machine learning has provided a new solution for spam filtering. Machine learning algorithms, such as Naive Bayes and Support Vector Machines (SVM), can train models on large data sets and identify complicated characteristics of spam emails. Besides, these algorithms can learn how to distinguish between spam and ham emails by analyzing historical mail data, in order to improve the accuracy of filtering. For instance, the Naive Bayes algorithm uses Bayes' theorem to analyze the frequency and probability of keywords in historical email data, learning how to effectively distinguish between spam and ham emails (Feng, Sun, Zhang, Cao, & Yang, 2016). And on the other hand, Support Vector Machines work by finding the optimal hyperplane in feature space to maximally separate spam from ham emails (Ghosh & Senthilrajan, 2023). These algorithms can continuously learn and optimize through historical data analysis, thereby improving filtering accuracy.

3.3 Current Research Trends (Machine Learning and Deep Learning Approaches)

In recent years, deep learning approaches have gained wide attention on detecting spam emails. Such deep learning models, like BERT (Bidirectional Encoder Representations from Transformers), can greatly improve the accuracy of spam classification through using attention mechanism and context understanding. And these models can effectively capture the subtle nuances in emails, enabling them to deal with diverse spam characteristics (Devlin, Chang, Lee, & Toutanova, 2019). The current research trend focuses on optimizing these algorithms to improve the efficiency and accuracy of filtering, and respond to constantly changing spam strategies. For example, researchers are exploring how to combine the advantages of multiple models to enhance overall detection capability. Additionally, leveraging natural language processing techniques to further understand the context and semantics of email content has also emerged as a new research direction (Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser, & Polosukhin, 2017).

In conclusion, the development process of spam filtering technology illustrates the transformation from simple rule-based systems to complex intelligent algorithms. With ongoing technological advancements, spam filtering systems are progressing toward higher efficiency and intelligence, ultimately aiming to provide users with a safer and cleaner email environment.

4 MAIN METHODS OF SPAM FILTERING

4.1 Content-Based Filtering

In the content-based filtering methods, the core is the comprehensive analysis of email content, so as to capture the potential characteristics of spam emails. This method usually evaluate the text, links and attachments in emails to determine whether they meet the specific spam characteristics. Through the integrated application of technologies, such as Natural Language Processing (NLP), keyword detection, pattern matching and context analysis, this method can effectively identify and filter spam emails (Saidani, Adi, & Allili, 2020). Outlined below are several key technologies and their implementations within this approach.

4.1.1 Keyword Detection

NLP technology is the basis of keyword detection. It can analyze words and phrases in email text, identifying the content that is highly relevant to spam emails. For example, promotional and eyecatching words, such as “free”, “discount” and “winning” are often used in spam emails. To support detection, the filtering system establishes a dataset containing commonly used spam words (Junnarkar, Adhikari, Fagania, Chimurkar, & Karia, 2021). Consequently, the system would analyze the keywords in the mail to determine whether it is likely to be a spam. Besides, this process can be enhanced with machine learning algorithms, which means that the system could learn and update emerging spam words and expressions to maintain the efficiency of filter (Roumeliotis, Tselikas, & Nasiopoulos, 2024).

4.1.2 Pattern Matching Technology

Pattern matching technology is mainly used to identify the unusual formats or structures in emails. Spammers typically adopt tactics which can evade conventional detection, such as using irregular type setting, abnormal coding or special characters. But by analyzing the metadata and header information of emails, pattern matching technology can detect improperly formatted email headers, non-standard character encodings and suspicious embedded links (Beaman & Isah, 2021). In addition, pattern matching can also recognize specific HyperText Markup Language (HTML) and Cascading Style Sheets (CSS) codes, which are often used by spammers to cover up malicious links or hide actual content (Zych, 2020).

4.1.3 Contextual Analysis and Semantic Understanding

In content-based filtering, simple keyword detection is insufficient to cope with all spam. Contextual analysis will further reduce misjudgment by evaluating the overall semantics of the email and considering the meaning of words in different contexts. For example, the word "free" may be a sign of spam in advertising emails, but in customer support emails it is reasonable. Supported by NLP technology, the filtering system can deeply understand the grammar and semantic structure of the message and decide whether it is spam according to the context (Si, Wu, Tang, Zhang, & Wosik, 2024). Thus, this semantic analysis technology effectively improves the accuracy and intelligence of the system, minimizing the risk of blocking legitimate emails.

4.1.4 Multi-layer Test Mechanism of SpamAssassin

SpamAssassin is a classic representative of content-based spam filters. It uses a multi-level testing mechanism to identify spam emails by scanning various suspicious features in emails. SpamAssassin will firstly detect common spam keywords, and also analyze the format and structure of the email to identify non-standard mail headers or abnormal character encodings (Chu, Hsu, Sheu, Yang, & Lee, 2020). In addition, SpamAssassin will detect the links in the email to identify potential malicious or suspicious links. And if certain feature in the email match known spam patterns, then the system will score the email. When the cumulative score exceeds the preset threshold, the email will be labelled as spam. This scoring mechanism enables SpamAssassin to handle different types of emails more flexibly, reducing the possibility of misjudgment (Elliott, 2023).

4.2 Behavior-Based Filtering

Behavior-based spam filtering not only relies on the analysis of email content, but also provides more accurate filtering by monitoring and evaluating email-related behavior. This method identifies potential spam emails by observing how users interact with emails and the sender's behavior history. Additionally, combined with sender reputation management, mail source verification and other technologies, behavior-based filtering can effectively improve the accuracy of detection and prevent false alarms and misjudgments. Outlined below are several key methods and their specific implementations.

4.2.1 User Behavior Monitoring

The behavior-based filtering method will firstly determine whether the email is spam by monitoring how user interacts with it. For example, the system will record whether the user opens the email, clicks on the link in it, or marks it as spam. Such behavior data helps the system identify the types of mail that users are not interested in or suspect that it is spam. And if certain emails are often labelled as spam by a large number of users, the system will raise the level of suspicion of the sender's email, so as to treat emails from the same source more strictly in future filtering (Beckel, 2024).

4.2.2 Sender's Reputation Management

Filtering methods based on the sender's reputation is a core component of behavior filtering. The spam filtering system will continuously monitor and evaluate the sender's history, mail delivery behavior and interaction with the recipients. And by using techniques such as blacklist and whitelist, the filter can make a preliminary judgment based on the sender's reputation. The blacklist has recorded the senders who are identified as sending spam, and their emails will be blocked directly; while the whitelist has listed reliable senders whose emails will be filtered first. And besides, the behavior filtering system will also refer to the sender's historical delivery rate, opening rate, complaint rate and other data. If emails from a particular sender are frequently ignored or reported as spam, the system reduces the sender's reputation score, potentially affecting their future delivery rates (Beckel, 2024; Blanchard, 2024).

4.2.3 Interaction Rate Analysis

The interaction rate is an important measurement based on behavior filtering. Through analyzing the opening rate, click rate, reply rate and other key indicators of the email, the system would evaluate whether the email meets the user's expectations. And if the interaction rate of an email is significantly lower than that of a normal email, and a large number of users choose to ignore or delete the email, then the email may be labelled as spam by the system. Emails with high interaction rates usually come from trusted senders, while emails with low interaction rates may be spam, especially when other emails from the sender also show similar low interaction characteristics (Beckel, 2024).

4.2.4 Mail Source Verification Technology (SPF and DKIM)

Email source verification is an important step to prevent spam in the hybrid filtering method. And the Sender Policy Framework (SPF) and Domain Key Identification Mail (DKIM) are two commonly used technologies. SPF confirms whether the sending server of the mail has the right to send mail on behalf of the domain to prevent the email address from being forged. And DKIM verifies whether the mail is tampered during transmission through digital signature. Through these mail source verification technologies, the system can determine whether the mail comes from a legal sender, thus improving the accuracy of filtering and reducing spam from fake domain names. These technologies are widely used in systems like SpamAssassin to verify mail through SPF and DKIM records (Blanchard, 2024; Van Impe, 2016).

4.3 Machine Learning-Based Filtering

In modern spam filtering systems, machine learning technologies play a crucial role. By analyzing and classifying large amounts of email data, these technologies could effectively distinguish between spam and ham emails. Outlined below are several commonly used machine learning methods and their applications in spam filtering.

4.3.1 Naive Bayes

Naive Bayes is a probability-based classification algorithm which is widely used in text classification tasks. This model assumes that each word in an email appears independently and predicts the category of the email based on the conditional probability of these words. And Naive Bayes will learn from past emails and calculate the possibility of each word appearing in spam or ham emails. Thus, when a new email arrives, the system would calculate the possibility of it being spam based on the combination of words.The classifier model uses a function that assigns the object (such as an email) to a class ˆC = C_f from the existing classes {C_1, . . . , C_f, . . . , C_k}, as described in Equation (1) (Aragão et al., 2021).

$$\hat{C} = arg \max_{k\in\{1,..,K\}} p(C_k) \prod_{i=1}^{n} p(x_i \mid C_k) \quad (1)$$

Fast computing speed and easy implementation are the main strength of Naive Bayes, especially when handling large amounts of text (Ghosh & Senthilrajan, 2023).

As shown in Figure 1, the Naive Bayes model illustrates the class probabilities P(c) directly depending on the features x_1, x_2, . . . , x_n, and no probability between the features is taken into account.

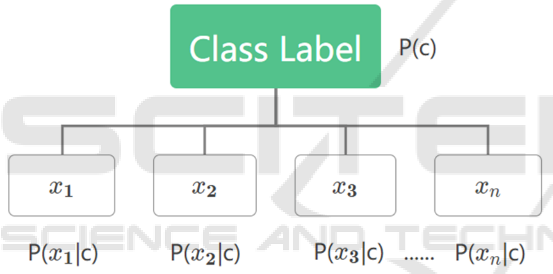

Figure 1: Naive Bayes Model Visualization of Class Probabilities model (based on Aragão et al., 2021)

4.3.2 Support Vector Machines (SVM)

SVM distinguishes between spam and ham emails by building a hyperplane. The fundamental goal of SVM is to identify the best decision-making boundary to keep distinct mail categories as separate as possible (Ghosh & Senthilrajan, 2023). Equation (2) can be used to describe the hyperplane.

$$H = VX + c \quad (2)$$

where c is a constant and V is the vector (Gibson, Issac, Zhang, & Jacob, 2020). And SVM is suitable for processing high-dimensional data, such as word vectors in emails, and can still perform well even with few training samples (Ghosh & Senthilrajan, 2023). In spam filtering, the strength of SVM is that it can effectively handle emails with blurred boundaries and minimizes the risk of misjudgment.

4.3.3 Random Forest

Random Forest is an integrated learning algorithm that creates numerous decision trees and combines their results to make classifications (Gibson et al., 2020). And each decision tree makes binary decisions about the characteristics in the content of emails to narrow the scope of categories layer by layer. Having many decision trees makes the random forest algorithm have high robustness and high accuracy compare to other machine learning algorithms. And in spam filtering, the random selection of features in random forests can reduce over-fitting and improve adaptability to new spam (Ghosh & Senthilrajan, 2023).

4.3.4 Ensemble Learning

Ensemble learning methods can combine multiple classifiers to improve the accuracy of spam detection. The typical ensemble algorithms include Adaptive Boosting (AdaBoost) and Bootstrap Aggregating (Bagging). AdaBoost gradually trains multiple weak classifiers by increasing the weight of misidentified samples in previous classification and finally forms a powerful classifier. In spam email classification, Bagging is used to count the frequency of spam-related words in the training dataset. It performs multiple classifications based on bootstrap samples of the training data. Finally, it will combine all the results to make a final prediction (Ghosh & Senthilrajan, 2023). Research has showed that ensemble learning methods can significantly improve spam filtering accuracy by combining the strengths of multiple classifiers (Kumar, Sonowal, & Nishant, 2020).

4.4 Deep Learning-Based Filtering

Recently, deep learning models have made significant progress in the field of spam filtering, surpassing the traditional machine learning methods. Unlike traditional models that rely on feature engineering, deep learning models can automatically extract complex features, especially when handling long text sequences and complex spam behaviors. Outlined below are several widely used deep learning models and their specific applications in spam filtering.

4.4.1 Convolutional Neural Networks (CNNs)

Although CNNs were initially performed well in image recognition, they are also suitable for text classification tasks. By treating email text as a onedimensional data sequence, CNNs can extract local text patterns through multiple convolutional layers. For example, CNNs can capture common words, phrases, or specific sentence structures which are often found in spam emails. And with the convolution operation and max pooling layers, CNNs could identify representative features from the emails and then use them to classify spam or ham emails (Roy, Singh, & Banerjee, 2020). CNNs ' high computational efficiency allows for the rapid processing of large-scale email data, particularly useful for recognizing short text segments and fixed patterns.

4.4.2 Recurrent Neural Networks (RNNs) and Its Variants

RNNs perform well in natural language processing tasks because of their ability to process serial data. In spam filtering, RNNs can effectively capture the timing dependency relationship in emails, helping to understand the contextual meaning of the email content. However, standard RNNs easily encounter gradient disappearance problems when handling long sequences, and this is the reason why Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) are widely used to replace RNNs (Roy et al., 2020). These variants can retain contextual information over longer time periods and have a better ability to capture the long text structure of the email. For example, LSTMs can remember the key phrases that appear throughout an email, helping distinguish between emails that seem similar but have different meanings. Besides, GRU could simplify structure so that it can improve computing efficiency while maintaining excellent performance, which makes it suitable for application environments with limited resources (Regina, Gopu, & Govindasamy, 2020).

4.4.3 Transformer-Based Models

Recently, Transformer-based models have shown excellent performance in text processing tasks, and BERT is a representative of it. BERT is pre-trained on large datasets such as English Wikipedia, containing 2.5 billion words, and BookCorpus, with 800 million words (Kaddoura, Alfandi, & Dahmani, 2020). Unlike traditional RNNs, Transformers can use self-attention mechanisms to consider the relationships between all words in a text at the same time, thus capturing deeper semantic meaning in emails (AbdulNabi & Yaseen, 2021). And BERT further improves this process by using a bidirectional encoder to understand the context of words, which allows it to interpret word meanings more accurately in different contexts. This makes it perform outstanding in spam filtering, especially when dealing with complex spam strategies and ambiguous words. And additionally, through the pre-trained BERT model, the spam filtering system can make use of a large amount of knowledge from the external dataset to improve the accuracy of recognition (Tida & Hsu, 2020).

4.4.4 Hybrid Models and Ensemble Learning

In practice, spam filtering systems often combine multiple deep learning models to improve robustness and accuracy. Combining CNNs with RNNs (such as LSTM), Hybrid models can not only capture the local features in emails but also retain context information in the long sequences. Additionally, ensemble learning methods are also widely applied in deep learning models, using predictions from multiple models to make a final decision. For example, multi-layer classifiers can gradually improve detection accuracy and reduce the occurrence of false alarms and omissions (Ablel-Rheem, Ibrahim, Kasim, Almazroi, & Ismail, 2020)

5 COMPARISON OF EXISTING SPAM FILTERING SYSTEMS

Spam filtering algorithms are commonly assessed using metrics such as accuracy, precision, recall, and F1 score.

Accuracy: It is defined as the percentage of all emails (true positives, false positives, true negatives, and false negatives) that are correctly identified, including both true positives (T_p) and true negatives (T_n). The formula for calculating accuracy is represented in Equation (3) (Ablel-Rheem et al., 2020; Adnan, Imam, Javed, & Murtza, 2024).

$$\text{Accuracy} = \frac{T_p + T_n}{T_p + F_p + T_n + F_n} \tag{3}$$

Precision: Precision refers to the percentage of emails that are successfully identified as spam (true positives, T_p) out of all emails that are projected to be spam (true positives and false positives, $T_p + F_p$)， as calculated by Equation (4) (Ablel-Rheem et al., 2020; Adnan et al., 2024).

$$\text{Precision} = \frac{T_p}{T_p + F_p} \tag{4}$$

Recall: Recall quantifies the percentage of real spam emails that the model successfully classified, calculated as shown in Equation (5). True positives and false negatives ($T_p + F_n$) are both taken into account (Ablel-Rheem et al., 2020; Adnan et al., 2024).

$$\text{Recall} = \frac{T_p}{T_p + F_n} \tag{5}$$

F1-Score (F1): The F1-score is calculated as the precision and recall harmonic means, as shown in Equation (6) (Ablel-Rheem et al., 2020; Adnan et al., 2024).

$$\text{F1-Score} = 2* \frac{\mathrm{P} * \mathrm{R}}{\mathrm{P} + \mathrm{R}} \tag{6}$$

Table 1 presents performance metrics of five key models across datasets, highlighting accuracy, precision, recall, and F1 scores (Ghosh & Senthilrajan, 2023).

Table 1: Performance Metrics of Various Spam Filtering Models (Ghosh & Senthilrajan, 2023).

Model	Accuracy (%)	Precision	Recall	F1 Score
Naive Bayes	87.63	0.82	0.85	0.83
Support Vector Machines	95.10	0.93	0.94	0.93
Random Forest	99.91	0.99	0.99	0.99
BERT	98.67	0.98	0.99	0.98
CNN	98.50	0.97	0.98	0.98

Random Forest and BERT consistently demonstrate outstanding performance. Random Forest achieves near-perfect accuracy in multiple datasets, largely due to its ability to handle diverse and complex data structures. On the other hand, BERT's advantage lies in its contextual understanding of email content, enabling it to analyze more intricate relationships within the text. Naive Bayes, while widely used for its simplicity and efficiency, tends to perform less effectively compared to more advanced models like SVM and CNN. The support vector machines offer strong boundary definitions between spam and legitimate emails, contributing to their high recall rates. CNN, particularly useful in analyzing data with spatial hierarchies such as email text, also performs well but falls slightly behind transformer models like BERT in terms of accuracy (Ghosh & Senthilrajan, 2023).

6 CHALLENGES AND FUTURE TRENDS IN SPAM FILTERING

One of the critical challenges in spam filtering is the issue of concept drift, which refers to the gradual evolution of spam characteristics over time. As spammers continually refine their techniques, traditional models struggle to adapt to new spam patterns. This results in the need for more dynamic, adaptable spam detection systems.

The rise of Large Language Model (LLM)-modified spam, where Large Language Models (LLMs) are used to craft highly sophisticated emails that bypass conventional filters, presents another growing threat. Existing models are vulnerable to these new types of spam that mimic legitimate communication patterns with high accuracy.

To address these challenges, researchers are exploring lifelong learning models like Error Learning Continuous Adaptive Detection Process (ELCADP), which are capable of continuously updating their knowledge base and adapting to new spam trends. This approach ensures that the model can evolve alongside the changing landscape of spam emails.

In the future, Artificial Intelligence (AI)-driven automation and personalized filtering systems will become key trends. These systems will rely on individual user behavior and preferences to fine-tune the filtering process, thereby minimizing false positives and false negatives. Models like BERT and transformers are expected to remain at the forefront of these innovations, leveraging their ability to process complex contextual information to improve filtering accuracy.

7 CONCLUSION

Spam filtering technology has developed from a basic rule-based system to a field dominated by advanced machine learning and deep learning models. Models such as Random Forest and BERT perform well in spam detection, especially when dealing with complex and diverse mail patterns. However, the continuous evolution of spam policies, such as the emergence of LLM-modified spam, continues to challenge existing filters.

The future of spam filtering lies in building a dynamic system that can adapt to the changing characteristics of spam. Lifelong learning model and AI-driven automation will play a key role in maintaining the accuracy and reliability of the spam detection system. Looking to the future, integrating these advanced technologies is crucial to ensure that users have a safer and more efficient email environment.

REFERENCES

Aragão, M. V. C., Ferreira, I. C., Oliveira, E. M., Kuehne, B. T., Moreira, E. M., & Carpinteiro, O. A. S., 2021. A study and evaluation of classifiers for anti-spam systems. IEEE Access, 9, pp. 157482-157496.

AbdulNabi, I., & Yaseen, Q., 2021. Spam email detection using deep learning techniques. Procedia Computer Science, 184, 853-858.

Ablel-Rheem, D. M., Ibrahim, A. O., Kasim, S., Almazroi, A. A., & Ismail, M. A., 2020. Hybrid feature selection and ensemble learning method for spam email classification. International Journal of Advanced Trends in Computer Science and Engineering, 9(1.4), 217-223.

Adnan, M., Imam, M. O., Javed, M. F., & Murtza, I., 2024. Improving spam email classification accuracy using ensemble techniques: A stacking approach. International Journal of Information Security, 23, 505–517.

Beaman, C., & Isah, H., 2021. Anomaly detection in emails using machine learning and header information. Canadian Institute for Cybersecurity, University of New Brunswick.

Beckel, Z., 2024. How email spam filters work: Understanding your email filter system. Qudni.

Blanchard, T., 2024. An in-depth guide to how email spam filter works. Sterling Technology Solutions.

Chu, K.-T., Hsu, H.-T., Sheu, J.-J., Yang, W.-P., & Lee, C.-C., 2020. Effective spam filter based on a hybrid method of header checking and content parsing. IET Networks, 9(6), 338-347.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K., 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv.

Elliott, A., 2023. SpamAssassin scores explained: What email marketers need to know. MailerCheck.

Feng, W., Sun, J., Zhang, L., Cao, C., & Yang, Q., 2016. A support vector machine based naive Bayes algorithm for spam filtering. 2016 IEEE 35th International Performance Computing and Communications Conference (IPCCC), 1-8.

Ghosh, A., & Senthilrajan, A., 2023. Comparison of machine learning techniques for spam detection. Multimedia Tools and Applications.

Gibson, S., Issac, B., Zhang, L., & Jacob, S. M., 2020. Detecting spam email with machine learning optimized with bio-inspired metaheuristic algorithms. IEEE Access, 8, 187914-187932.

Junnarkar, A., Adhikari, S., Fagania, J., Chimurkar, P., & Karia, D., 2021. E-Mail spam classification via machine learning and natural language processing. 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 693-699.

Kaddoura, S., Alfandi, O., & Dahmani, N., 2020. A spam email detection mechanism for English language text emails using deep learning approach. 2020 IEEE 29th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 193-198.

Kumar, N., Sonowal, S., & Nishant., 2020. Email spam detection using machine learning algorithms. 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), 108-113.

Kaddoura, S., Alfandi, O., & Dahmani, N., 2020. A spam email detection mechanism for English language text emails using deep learning approach. 2020 IEEE 29th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 193-198.

Mohammad, R. M. A., 2020. A lifelong spam emails classification model. Applied Computing and Informatics, 20(1/2), 35-54.

Roumeliotis, K. I., Tselikas, N. D., & Nasiopoulos, D. K., 2024. Next-generation spam filtering: Comparative fine-tuning of LLMs, NLPs, and CNN models for email spam classification. Electronics, 13(2034).

Roy, P. K., Singh, J. P., & Banerjee, S., 2020. Deep learning to filter SMS spam. Future Generation Computer Systems, 102, 524-533.

Regina, K., Gopu, A., & Govindasamy, V., 2020. Spam detection using recurrent neural networks. International Journal for Research in Engineering Application & Management, 6(1), 313-318.

Saidani, N., Adi, K., & Allili, M. S., 2020. A semantic-based classification approach for an enhanced spam detection. Computers & Security, 94, 101716.

Si, S., Wu, Y., Tang, L., Zhang, Y., & Wosik, J., 2024. Evaluating the performance of ChatGPT for spam email detection. arXiv.

Tida, V. S., & Hsu, S., 2020. Universal spam detection using transfer learning of BERT model. Center of Advanced Computer Studies, University of Louisiana at Lafayette.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I., 2017. Attention is all you need. In 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

Van Impe, K., 2016. Understanding the SPF and DKIM spam filtering mechanisms. Security Intelligence.

Zych, S., 2020. Patterns are power: A beginner’s guide to spam detection. Medium.