The North African Journal of Scientific Publishing (NAJSP)
مجلة شمال إفريقيا للنشر العلمي (NAJSP)
P-ISSN:0000-0000, EISSN: 0000-0000
Volume 1, Issue 1, January-March 2023, Page No: 13-23
Website: https://najsp.com/index.php/home/index

Investigation of Three Machine Learning Models for the Detection of Emails Spam

Abdulsalam Ashour Almabrouk[1*], Salih Mousay Abraheem[2], Muftah Emtir Ali [3]
Mohamed Saleh Mansour [4]

[1] Department of Electrical Engineering, College of Science and Technology, Qaminis, Libya
[2,3] Department of Electrical-Electronics Engineering, Karabuk University, Karabuk, Turkey
[4] Department of Electrical Engineering, College of Science and Technology, Qaminis, Libya

[*]*Corresponding author: Ashourashour1942@yahoo.com*

Received: February 04, 2023	Accepted: March 01, 2023	Published: March 04, 2023

Abstract:
Recently, machine learning has been applied to different major areas such as text classification, machine translation, and spam detection. The great performance of machine learning algorithms in several fields provided humans with opportunities to tackle some of their hard jobs to be handled by machine learning systems. These tasks seem effortless for machines and need less time as the number of texts or spam that need to be classified is huge. Hence, in his paper, we propose three different machine-learning models for the task of email spam detection. The three models are trained and validated on a public spam dataset. Experimentally, the three models performed differently, and it was seen that the Naïve Bayes outperformed the other machine learning algorithms in terms of accuracy and other evaluation metrics**.**

Keywords: Machine learning, K-Nearest neighbor (KNN), Support vector machine

Cite this article as: A. A. Almabrouk, S. M. Abraheem, M. E. Ali, and M. S. Mansour and, "Investigation of Three Machine Learning Models for the Detection of Emails Spam," *The North African Journal of Scientific Publishing (NAJSP),* vol. 1, no. 1, pp. 13–23, January-March 2023.

Introduction

In the era of information technology, information sharing has become very easy and fast. Many platforms are available for users to share information anywhere across the world [1,2]. Among all information sharing mediums, email is the simplest, cheapest, and most rapid method of information sharing worldwide [3]. But, due to their simplicity, emails are vulnerable to different kinds of attacks, and the most common and dangerous one is spam. Email spam is a persistent problem that continues to plague email users around the world [4,5]. The proliferation of spam has made it increasingly difficult for people to manage their inboxes, leading to wasted time and reduced productivity.

However, machine learning offers a potential solution to this problem. By training models on large datasets of spam and non-spam emails, it is possible to build algorithms that can automatically detect and filter out spam messages [6,7]. In this direction, email exchange is a method of correspondence among individuals, and billions of cell and laptop phone users share various emails. In any case, such kind of correspondence is unsafe because of an absence of legitimate message-sifting techniques. Spam is the major reason for this weakness, as it threats the email exchange between users [8-12]. Spams are unsolicited junk emails or messages, which are undesirable for beneficiaries and are shipped off the users without their earlier consent. Moreover, email users invest most of their significant energy and time in arranging these spam sent emails [13,14]. Different duplicates of the same message are sent ordinarily which not just influences an association monetarily, yet in addition, aggravates the accepting users. Spam messages are not just encroaching on the user's messages, yet they are moreover delivering a huge measure of undesirable information and along these lines influencing the organization's ability and utilization [15-19].

Mostly, emails have a very similar structure which consists of the body of the email and its corresponding subject. A run of the mill spam mail can be ordered by filtering its substance. The cycle of spam mail detection depends on the supposition that the substance of the spam mail is not the same as the real or ham mail [20, 21]. For model words identified with the ad of any item, support of administrations, dating related content and so on. The cycle of spam email detection can be extensively sorted into two methodologies: machine learning and knowledge engineering approach [22,33]. The machine learning approach is to train machine learning algorithms to classify emails into spam or ham, and it showed more effective results than humans and knowledge engineering approaches [3,4]. Moreover, machine learning showed promising outcomes in different engineering areas such as image classification [24-30], prediction [31], and natural language processing [32].

Machine learning algorithms use statistical models to classify data. The detection of spam is largely based on the analysis of the content of the message. In the case of spam detection, a trained machine learning model can determine the order of words found in the e-mail that is closest to finding spam and which e-mail is the safest [33,34]. Spam detection is a highly monitored machine learning problem. Most email providers have their own huge records of labeled emails. This means that a machine learning model can be provided with several examples of spam and ham messages, and it will find relevant patterns that can be divided into two different categories [35,36].

Processing natural language has made many exciting advances in recent years, but artificial intelligence algorithms still don't understand language the way. One of the most important steps in developing a machine learning model for spam detectors is to prepare the data for statistical processing. The email spamming problem we are trying to solve is that spam data only accounts for 20% of our data [37-40]. If the algorithm can predict e-mails without spam, it can achieve an accuracy of 80%. Positive data can predict negative samples and give them an accuracy of up to 99%, but this type of model is useless in real-world scenarios [41]. Thus, in this article, we will train different machine learning filters using a collection of spam and non-spam emails (also known as ham). We will train the filters using text samples from a public dataset that is labeled as spam or ham. In contrast to regression, we do not output the categorized data [42].

The detection and filtering of email spam has become an increasingly important task in the modern era of information overload. Traditional rule-based systems for filtering spam have proven to be insufficient, as spammers have become adept at evading these filters by using more sophisticated techniques. Machine learning offers a promising solution to this problem, as it allows for the development of automated models that can learn from large datasets of spam and non-spam emails to accurately identify and filter spam messages. The paper contributes to the field by comparing the performance of three different machine learning models for the detection of email spam. The three models considered in this study are logistic regression, support vector machines, and random forests. These models were chosen due to their popularity and effectiveness in a wide range of classification tasks. This study explores three distinct machine learning techniques for classifying email messages as either spam or ham. The three algorithms leverage various email characteristics, including content and other parameters, to perform the classification task. Specifically, the models used in this study are Naïve Bayes (NB), K-nearest neighbor (KNN), and support vector machine (SVM). To train and evaluate

the performance of these models, the Spambase Dataset is utilized. The study comprehensively compares and discusses the effectiveness of the three models using a range of performances.

Related Works

The task of detecting email spam is critical in the field of cybersecurity, as it plays a vital role in preventing fraudulent activities. To this end, numerous spam detection methods have been developed, each based on different filtering categories [43]. In the following section, the Authors discussed some of the most advanced approaches that have been widely employed to address the challenge of email spam filtering.

- Content Based Filtering Technique: this approach depends on creating some automatic filtering rules and using machine learning models to classify emails into spams or non-spams [44]. These machine learning models can be Naïve Bayesian classification, Support Vector Machine, K-Nearest Neighbor, or Neural Networks. this method is used in this article and it is achieved by analyzing words in addition to their occurrences and distributions in the text. This analysis is then used to train the models and generate some rules to filter out incoming emails.

- Case Base Spam Filtering Method: this technique is one of the most common emails spam filtering methods [45]. It first extracts the spam and non-spam emails from the user's email using a collection model. These emails are then preprocessed and transformed using feature extraction techniques. Finally, the processed data is converted into vector sets. The blast stage is to classify these data into spam or non-spam using machine learning models.

- Previous Likeness Based Spam Filtering Technique: This is a memory-based approach in which a machine learning model is trained on a set of spam and non-spam samples and stores them [46,47]. A new incoming email will be then classified as a spam or not based on its resemblance to the stored training examples. k-nearest neighbor (kNN) is the most popular machine learning method used in this approach for filtering out spams.

Recent studies have been proposed to resent different email spam approaches and all have been successfully applied to classify data. These methods include probabilistic, decision tree [13], artificial immune system [48-54], support vector machine (SVM) [55], and artificial neural networks (ANN) [56]. Idris [17] proposed the detection of emails spam using a neural network and negative selection algorithm. In this work, the performance of the neural network for the such task was compared with the support vector machine and it was found the neural network using backpropagation achieved a higher accuracy than that reached by the SVM. Another research for spam classification using neural networks was proposed by Edstrom [57]. In their work, it was found that several hidden layers do not improve the accuracy of the network. A single hidden layer neural network was seen to achieve the best performance of 94,6% accuracy. Sharma and Bhardwaj [58] developed a hybrid bagged approach for the classification of emails spams. Their system is a combination of Naïve Bayes and J48 algorithms trained to classify emails into spam or non-spam. Their experiments showed that their proposed system achieved a classification accuracy of 87.5%, which is lower than the one our system has reached. Moreover, Pandey et al., [59] proposed a system based on SVM and NB to classify emails. In their paper, SVM achieved an accuracy of 91% while NB reached 92%.

Materials and Methods

This section discusses the dataset used for training and testing the three models in addition to the research methods and materials of the paper.

1. Dataset

The SpamBase dataset [60] is considered to be used for training and testing the three different employed models. This dataset consists of 4601 instances of both spam and non-spam emails. A learning scheme of 50:50 is considered for training and testing such models in which 50% is used for

training and 50% for testing. Table 1 shows a sample of some spam and non-spam instances of the dataset. Note that the two classes are labeled as “1” for spam and “0” for non-spam.

Table 1. A sample of spam and non-spam instances of the used dataset

Sample number	Class	Content
1	Spam	Sunshine Quiz! Win a super Sony DVD recorder if you can name the capital of Australia. Text MQUIZ to 82277. B
2	Ham	As I entered my cabin my PA said, " Happy B'day Boss !!". I felt special. She asked me 4 lunch. After lunch, she invited me to her apartment. We went there.
3	Spam	Today’s Voda numbers ending with 7634 are selected to receive a å£350 reward. If you have a match, please call 08712300220 quoting claim code 7684 standard rates apply.

2. Naïve Bayes (NB)

Naïve Bayes is a probabilistic machine learning algorithm for binary or multiclass classification tasks. Such an algorithm is based on Bayes’s theorem [61-63] and it works by assuming that the occurrence of a certain feature is independent of the occurrence of other features. Baye’s theorem is used to determine the probability of a hypothesis with prior knowledge. Baye's theorem is utilized to determine the likelihood of theory with earlier knowledge. The working formula of Baye’s theorem is:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \quad (1)$$

where $P(A \mid B)$ is the probability of hypothesis A, given that B is true. $P(B \mid A)$ is the likelihood hypothesis B, given that A is true? P(A) and P(B) are the probabilities of hypotheses A and B, independently.

3. K-Nearest Neighbor (KNN)

K-nearest neighbors are a basic and simple algorithm that stores every accessible instance and predicts classes of the new cases depending on a distance measure (e.g., Euclidean distance measures) [64-66]. For such an algorithm, a new case is classified by the majority voting of its neighbors. This case is then assigned to its k-nearest neighbors by measuring its corresponding distances to all its neighbors. Different distance measures can be used to compute distance, however, in this work, the Euclidean distance is used, and it is as follows:

$$E = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \quad (2)$$

Support Vector Machine (SVM)

SVM is a machine learning algorithm that can be for both classification and regression problems. This algorithm works mainly by finding a hyperplane in N-dimensional space to classify data points in different classes [67]. The idea is to find the best plane that has the maximum margin. In other words, the plane in which the distance between classes and data points is the maximum. It is then important to maximize this margin. SVM algorithm can maximize the margin between data points and hyperplane by computing the minimizing cost function, i.e., Hinge loss. Hinge loss, Exponential loss, Logit loss and many other types of loss can be used to train the SVM [68-71]. However, in this work, the hinge loss is used, and it is defined as follows:

$$h_\theta(x) = \begin{cases} 1 & if\ \theta^t x \geq 0 \\ 0 & else \end{cases} \quad (3)$$

Where θ is the angle between the two vectors, x and y.

Metrics for evaluation the model's performance

Several metrics are used in this work to evaluate the performance the three employed models for classifying email messages. These metrics include the accuracy, ROC, specificity, sensitivity [24].

$$Accuracy\ (Acc) = \frac{\text{TP+TN}}{\text{TP+ FN+TN+FP}} \quad (4)$$

$$Sensitivity\ (Sn) = \frac{\text{TP}}{\text{TP+ FN}} \quad (5)$$

$$Specificity\ (Sp) = \frac{\text{TN}}{\text{TN+ FP}} \quad (6)$$

where TP stands for true negative, and it indicates the number of correctly predicted positive classes. TN stands for true negative, and it indicates the number of correctly predicted negative classes. FP is the false positive, and it shows the incorrectly predicted positive data, while FN is the false-negative, and it indicates the number of incorrectly predicted negative data. AUC is the area under the Receiver Operating Characteristic curve (ROC), which is a graph that shows the performance of the network at thresholds. ROC plots the True positive rate versus the false positive rate

Results and discussion

In this section, the training and testing performance of the employed models are discussed. In terms of preprocessing, a machine learning inspired approach is proposed for spam mail detection. In spam mail identification framework, the first step is to collect the data which are unstructured in nature. Thus, preprocessing is required. Therefore, to decrease the calculations and obtain good results, email information should be pre-processed. Hence, data are first processed by deleting stop words. Moreover, word tokenization is likewise performed to secure important data. Finally, data are fed into the machine learning algorithms (NB, KNN, SVM) to be classified as spam or non-spam. Figure 1 shows spam filtering. Supervised learning uses labeled data for training, and then it can predict the new data. This type of learning can be used in solving various problems, i.e., advertisement popularity, spam classification, face recognition, and object classification. The process of supervised learning is illustrated in Figure 2.

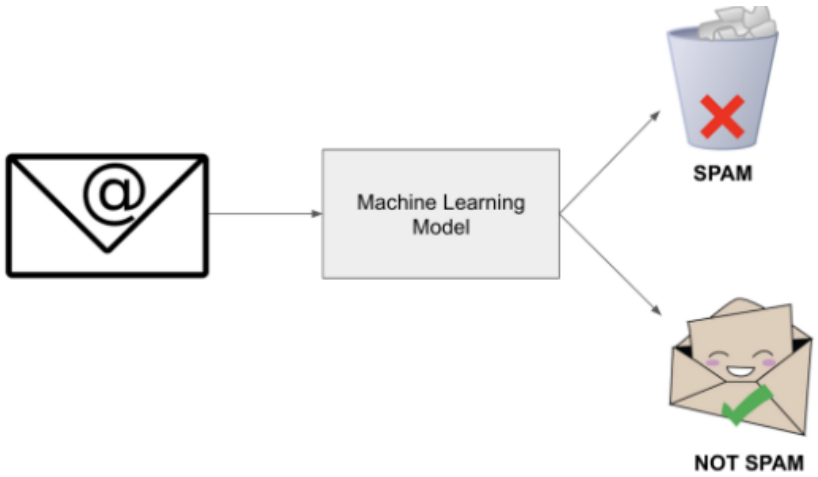

Figure 1: Spam filtering

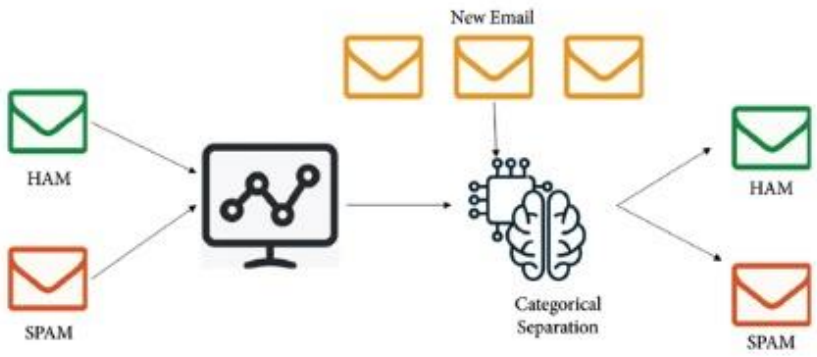

Figure 2: Emails data preprocessing

Due to classification results, the three employed models (BN, KNN, SVM) are trained and tested using data from the SpamBase dataset. The learning scheme used to train and test the models is 50:50 i.e., 50% of images are used for training, while the remaining 50% are used as a held-out test set for evaluating the network's performance. Note that the network is evaluated by calculating its training and testing accuracy and loss using the formulas in equations (3) and (4). Figure 3 presents the percentage of spam and non-spam (Ham) content in the used dataset. Table 2, illustrates the evaluation metrics of the models in classifying emails into messages or non-spam

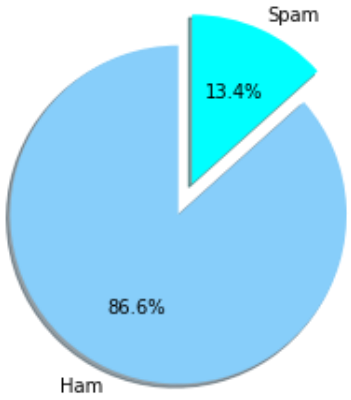

Figure 3: Ratio of spam and non-spam (Ham) content in the used dataset.

Table 3. The evaluation metrics of the models in classifying emails into messages or non-spam

	NB	KNN (K=1)	SVM
Accuracy	97%	89%	94%
Sensitivity	93%	80%	83%
Specificity	87%	71%	78%
AUC	93%	79%	81%
Processing time	0.494 (s)	0.630 (s)	0.86.032 (s)

In this context, Naive Bayes (NB), K nearest neighbors (KNN) and support vector machine (SVM) are the classifiers used to detect emails spams. The models are validated with 50% of the data and the experimental results. According to that ability, the KNN achieved the highest accuracy among them all. It achieved an accuracy of 97%, a sensitivity of 93%, a specificity of 87%, and an area under the curve (AUC) 0f 93%. In terms of computation time, Table 3 shows that the NB required a shorter time to achieve such accuracy than SVM and KNN.

On the hand, the other models achieved accuracies of 89% and 94% for KNN and SVM, respectively. Moreover, this ROC curve represents the measure of severability of the two classes: spam and non-spam. Hence, it is noted that the NB outperformed the other models in terms of AUC as it achieved a better degree of reparability between the two classes.

Table 4. Results comparison with earlier works [19,20].

	Accuracy
Hybrid bagged approach	87.5%
SVM	91%
NB	92%
Our work	97%

Comparison with other related works: Research has been extensively conducted to develop a machine learning-based detection system for emails spam. Sharma and Bhardwaj [19] proposed a hybrid bagged approach for the classification of emails spams. Their experimental results showed achieved a classification accuracy of 87.5%, which is lower than the one our system has reached. Moreover, Pandey et al., [20] applied SVM and NB to classify emails as spam. In their paper, it was seen that SVM achieved an accuracy of 91% while NB reached 92%. Table 3 shows the comparison of our work with the discussed works, and it is seen that our proposed NB outperformed the two other systems.

Challenges of spam detection

Spam detection is a critical task in the field of cybersecurity. However, there are several challenges associated with detecting email spam. One of the most significant challenges is the ever-evolving nature of spam. Spammers continually modify their tactics, making it difficult for detection algorithms to keep up. As a result, spam detection systems must be updated regularly to ensure that they remain effective. other challenge is the high rate of false positives and false negatives. False positives occur when a legitimate email is incorrectly flagged as spam, while false negatives occur when a spam email is not detected and allowed through. Both of these errors can be costly, as false positives can result in important emails being missed, while false negatives can result in unwanted emails reaching the user's inbox. Additionally, spammers often use sophisticated techniques to evade detection, such as disguising their messages or using botnets to send spam from multiple IP addresses. These tactics can make it challenging for spam detection systems to identify spam effectively. Moreover, email spam is often intertwined with other types of cyber threats, such as phishing, malware, and ransomware. Detecting these threats requires advanced techniques beyond standard spam detection methods, making the task even more challenging. Thus, privacy concerns can also pose challenges to spam detection. Spam detection systems must balance the need to detect spam with the user's privacy. Some users may be uncomfortable with the idea of having their emails scanned by automated systems, and as such, spam detection systems must be designed to respect the user's privacy while still being effective at detecting spam. Besides that, detecting email spam is a challenging task due to the constantly evolving nature of spam, the high rate of false positives and negatives, sophisticated evasion techniques, the presence of other cyber threats, and privacy concerns. Despite these challenges, continued research and development of advanced techniques can help improve the effectiveness of spam detection systems and enhance cybersecurity.

Conclusion

In recent years, email spam has become a significant problem for internet users, and detecting email spam has become an essential task in the field of cybersecurity. In this paper, a comparison of three different machine learning algorithms for detecting email spam is presented. The study demonstrates that different models can behave differently over the same spam dataset, and their performance can differ. To perform the study, the three models - Naive Bayes (NB), K-nearest neighbor

(KNN), and Support Vector Machine (SVM) - are trained and validated on a public dataset called the SpamBase dataset. The dataset consists of 4,601 emails with a binary classification of spam or ham. The study employs various evaluation metrics, such as accuracy, sensitivity, specificity, and area under the curve (AUC), to compare the performance of the models. The results of the study indicate that Naive Bayes outperformed all other models in terms of accuracy, sensitivity, specificity, and AUC. Furthermore, the processing time of each algorithm was computed, and SVM required the longest time for processing. These findings highlight the importance of carefully selecting the appropriate machine learning algorithm for detecting email spam. To validate the results, the study compares the performance of NB with related works and research. The study concludes that NB can achieve higher accuracy than other related models, such as the hybrid bagged approach, SVM, and NB. These results suggest that NB is a promising model for detecting email spam. In conclusion, this study presents a comprehensive comparison of three machine learning algorithms for detecting email spam. The results demonstrate that different models can perform differently over the same dataset, and NB outperformed all other models in terms of accuracy, sensitivity, specificity, and AUC. The findings of this study can be useful for researchers and practitioners in the field of cybersecurity in selecting the most appropriate machine learning algorithm for detecting email spam.

References

[1] H. Faris, A. M. Al-Zoubi, A. A. Heidari et al., “An intelligent system for spam detection and identification of the most relevant features based on evolutionary random weight networks,” *Information Fusion*, vol. 48, pp. 67–83, 2019.

[2] A. Alghoul, S. Al Ajrami, G. Al Jarousha, G. Harb, and S. S. Abu-Naser, “Email classification using artificial neural network,” *International Journal for Academic Development*, vol. 2, 2018.

[3] N. Udayakumar, S. Anandaselvi, and T. Subbulakshmi, “Dynamic malware analysis using machine learning algorithm,” in *Proceedings of the 2017 International Conference on Intelligent Sustainable Systems (ICISS)*, IEEE, Palladam, India, December 2017

[4] J. K. Kruschke and T. M. Liddell, “Bayesian data analysis for newcomers,” *Psychonomic Bulletin & Review*, vol. 25, no. 1, pp. 155–177, 2018.

[5] M. M. Khaleel, T. Mohamed Ghandoori, A. Ali Ahmed, A. Alsharif, A. J. Ahmed Alnagrat, and A. Ali Abulifa, “Impact of mechanical storage system technologies: A powerful combination to empowered the electrical grids application,” in 2022 IEEE 2nd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA), 2022.

[6] A. Barushka and P. Hájek, “Spam filtering using regularized neural networks with rectified linear units,” in *Proceedings of the Conference of the Italian Association for Artificial Intelligence*, Springer, Berlin, Germany, November 2016.

[7] K. S. Adewole, N. B. Anuar, A. Kamsin, K. D. Varathan, and S. A. Razak, “Malicious accounts: dark of the social networks,” *Journal of Network and Computer Applications*, vol. 79, pp. 41–67, 2017

[8] M. M. Khaleel, S. A. Abulifa, I. M. Abdaldeam, A. A. Abulifa, M. Amer, and T. M. Ghandoori, “A current assessment of the renewable energy industry,” AJAPAS, pp. 122–127, 2023.

[9] F. Jamil, H. K. Kahng, S. Kim, and D. H. Kim, “Towards secure fitness framework based on IoT-enabled blockchain network integrated with machine learning algorithms,” *Sensors*, vol. 21, no. 5, p. 1640, 2021.

[10] M. H. Arif, J. Li, M. Iqbal, and K. Liu, “Sentiment analysis and spam detection in short informal text using learning classifier systems,” *Soft Computing*, vol. 22, no. 21, pp. 7281–7291, 2018

[11] M. A. Ferrag, L. Maglaras, S. Moschoyiannis, and H. Janicke, “Deep learning for cyber security intrusion detection: approaches, datasets, and comparative study,” *Journal of Information Security and Applications*, vol. 50, Article ID 102419, 2020.

[12] X. Zheng, X. Zhang, Y. Yu, T. Kechadi, and C. Rong, “ELM-based spammer d S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, “Supervised machine learning: a review of classification techniques,” *Emerging artificial intelligence applications in computer engineering*, vol. 160, pp. 3–24, 2007.

[13] L. N. Petersen, “The ageing body in monty Python live (mostly),” *European Journal of Cultural Studies*, vol. 21, no. 3, pp. 382–394, 2018.

[14] D. Lee, M. J. Lee, and B. J. Kim, “Deviation-based spam-filtering method via stochastic approach,” *EPL (Europhysics Letters)*, vol. 121, no. 6, Article ID 68004, 2018.

[15] M. M. Khaleel, M. R. Adzman, S. M. Zali, M. M. Graisa, and A. A. Ahmed, “A review of fuel cell to distribution network interface using D-FACTS: Technical challenges and interconnection trends,” Int. J. Electr. Electron. Eng. Telecommun., pp. 319–332, 2021.

[16] A. K. Jain and B. B. Gupta, “Towards detection of phishing websites on client-side using machine learning based approach,” *Telecommunication Systems*, vol. 68, no. 4, pp. 687–700, 2018.

[17] M. F. N. K. Pathan and V. Kamble, “A review various techniques for content based spam filtering,” *Engineering and Technology*, vol. 4, 2018.

[18] M. Bassiouni, M. Ali, and E. A. El-Dahshan, “Ham and spam e-mails classification using machine learning techniques,” *Journal of Applied Security Research*, vol. 13, no. 3, pp. 315–331, 2018.

[19] A. d. A. Garcez, M. Gori, L. C. Lamb, L. Serafini, M. Spranger, and S. N. Tran, “Neural-symbolic computing: an effective methodology for principled integration of machine learning and reasoning,” *Journal of Applied Logic*, vol. 6, 2019.

[20] M. M. Khaleel, M. R. Adzman, and S. M. Zali, “An integrated of hydrogen fuel cell to distribution network system: Challenging and opportunity for D-STATCOM,” Energies, vol. 14, no. 21, p. 7073, 2021.

[21] J. Tanha, M. van Someren, and H. Afsarmanesh, “Semi-supervised self-training for decision tree classifiers,” *International Journal of Machine Learning and Cybernetics*, vol. 8, no. 1, pp. 355–370, 2017.

[22] H. Takhmiri and A. Haroonabadi, “Identifying valid email spam emails using decision tree,” *International Journal of Computer Applications Technology and Research*, vol. 5, 2016.

[23] W. Li, W. Meng, Z. Tan, and Y. Xiang, “Design of multi-view based email classification for IoT systems via semi-supervised learning,” *Journal of Network and Computer Applications*, vol. 128, pp. 56–63, 2019.

[24] K. Lei, Y. Liu, S. Zhong et al., “Understanding user behavior in Sina Weibo online social network: a community approach,” IEEE Access, vol. 6, pp. 13302–13316, 2018.

[25] S. O. Olatunji, “Improved email spam detection model based on support vector machines,” *Neural Computing & Applications*, vol. 31, no. 3, pp. 691–699, 2019.

[26] M. M. Khaleel, M. R. Adzman, and S. M. Zali, “An integrated of hydrogen fuel cell to distribution network system: Challenging and opportunity for D-STATCOM,” Energies, vol. 14, no. 21, p. 7073, 2021.

[27] N. F. Rusland, N. Wahid, S. Kasim, and H. Hafit, “Analysis of Naïve Bayes algorithm for email spam filtering across multiple datasets,” in *Proceedings of the IOP Conference Series: Materials Science and Engineering*, IOP Publishing, Busan, Republic of Korea, 2017.

[28] M. Dewis and T. Viana, “Phish Responder: A hybrid machine learning approach to detect phishing and spam emails,” *Appl. Syst. Innov.*, vol. 5, no. 4, p. 73, 2022.

[29] A. Alsharif, A. A. Ahmed, M. M. Khaleel, and M. A. Altayib, “Ancillary services and energy management for electric Vehicle: Mini-review,” *NAJSP*, pp. 9–12, 2023.

[30] A. Singh and S. Batra, “Ensemble based spam detection in social IoT using probabilistic data structures,” *Future Gener. Comput. Syst.*, vol. 81, pp. 359–371, 2018.

[31] M. A. Ghani, U. B. S. Informatika, H. Sulaiman, and U. B. S. Informatika, “Deteksi Spam Email dengan Metode Naive Bayes dan Particle Swarm Optimization (PSO),” *jit*, vol. 6, no. 1, pp. 11–20, 2023.

[32] F. Jáñez-Martino, R. Alaiz-Rodríguez, V. González-Castro, E. Fidalgo, and E. Alegre, “A review of spam email detection: analysis of spammer strategies and the dataset shift problem,” *Artif. Intell. Rev.*, vol. 56, no. 2, pp. 1145–1173, 2023.

[33] P. Sharma and U. Bhardwaj, “Email spam detection using bagging and boosting of machine learning classifiers,” *Int. J. Adv. Intell. Paradig.*, vol. 24, no. 1/2, p. 229, 2023.

[34] A. Alsharif , C. W. Tan, R. Ayop, A. A. Smin; A. A. Ahmed, F.H. Kuwil, Khaleel, M.M. ., “Impact of electric Vehicle on residential power distribution considering energy management strategy and stochastic Monte Carlo algorithm,” *Energies*, vol. 16, no. 3, p. 1358, 2023.

[35] D. Swain, N. Chillur, M. Kava, and S. Satapathy, “Intelligent system for detecting email spam messages using GRU,” in *Advances in Data and Information Sciences*, Singapore: Springer Nature Singapore, 2023, pp. 71–77.

[36] S. A. Chaturvedi and L. Purohit, “Feature selection-based spam detection system in SMS and email domain,” in *Advances in Intelligent Systems and Computing*, Singapore: Springer Nature Singapore, 2023, pp. 37–52.

[37] M. Salb, L. Jovanovic, M. Zivkovic, E. Tuba, A. Elsadai, and N. Bacanin, “Training logistic regression model by enhanced moth flame optimizer for spam email classification,” in *Computer Networks and Inventive Communication Technologies*, Singapore: Springer Nature Singapore, 2023, pp. 753–768.

[38] M. M. Khaleel, A. Alsharif, and I. I. K. Imbayah, “Renewable energy technologies: Recent advances and future predictions,” pp. 58–64, 2022.

[39] G. Revathi, K. N. Rao, and G. S. Ratnam, “Email Spam Detection using Naïve Bayes Algorithm,” *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 10, no. 9, pp. 653–655, 2022.

[40] A. Chakraborty, U. K. Das, J. Sikder, M. Maimuna, and K. I. Sarek, “Content based email spam classifier as a web application using naïve Bayes classifier,” in *Intelligent Computing & Optimization*, Cham: Springer International Publishing, 2023, pp. 389–398.

[41] S. M. Tamilarasan, M. Hithasri, and K. Pille, “Email spam detection using multilayer perceptron algorithm in deep learning model,” in *Information and Communication Technology for Competitive Strategies (ICTCS 2021)*, Singapore: Springer Nature Singapore, 2023, pp. 581–587.

[42] S. A. Khan, K. Iqbal, N. Mohammad, R. Akbar, S. S. A. Ali, and A. A. Siddiqui, “A novel fuzzy-logic-based multi-criteria metric for performance evaluation of spam email detection algorithms,” *Appl. Sci. (Basel)*, vol. 12, no. 14, p. 7043, 2022.

[43] A. A. Ahmed, A. Alsharif, T. Triwiyanto, M. M. Khaleel, C. W. Tan, and R. Ayop, “Using of neural network-based controller to obtain the effect of hub motors weight on electric vehicle ride comfort,” in *2022 IEEE 2nd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA)*, 2022, pp. 189–192.

[44] L. Á. Redondo-Gutierrez, F. Jáñez-Martino, E. Fidalgo, E. Alegre, V. González-Castro, and R. Alaiz-Rodríguez, “Detecting malware using text documents extracted from spam email through machine learning,” in *Proceedings of the 22nd ACM Symposium on Document Engineering*, 2022.

[45] W. G. Volante, C. Gendron, S. Sewell, and D. M. Sarno, “Do spam filters make us complacent?: Examining email legitimacy tags and phishing susceptibility,” *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 66, no. 1, pp. 736–736, 2022.

[46] S. A. Sheikh and M. T. Banday, “Mitigating BOT-based methods for email address harvesting and spamming,” *Research Square*, 2022.

[47] A. Alsharif, C. W. Tan, R. Ayop, A. A. A. Ahmed, A. Alanssari, and M. M. Khaleel, “Energy management strategy for Vehicle-to-grid technology integration with energy sources: Mini review,” pp. 12–16, 2022.

[48] U. Singh, V. Singh, M. K. Gourisaria, and H. Das, “Spam email assessment using machine learning and data mining approach,” in *2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT)*, 2022.

[49] A. Reddy, D. M. Umamaheswari, D. A. Viswanathan, G Vikram, and Mamatha K, “Using support vector machine for classification and feature extraction of spam in email,” *International Journal of Advanced Research in Science, Communication and Technology*, pp. 85–89, 2022.

[50] S. Zavrak and S. Yilmaz, “Email spam detection using hierarchical attention hybrid deep learning method,” *Research Square*, 2022.

[51] M. M. Khaleel, “Intelligent techniques for distribution static compensator using genetic algorithm, and fuzzy logic controller,” *Int. J. Comput. Commun. Instrum. Eng.*, vol. 2, no. 1, 2015.

[52] H. Mukhtar, J. Al Amien, and M. A. Rucyat, “Filtering Spam Email menggunakan Algoritma Naïve Bayes,” *CoSciTech*, vol. 3, no. 1, pp. 9–19, 2022.

[53] R. Natarajan *et al.*, “Hybrid big bang–big crunch with ant colony optimization for email spam detection,” *Int. J. Mod. Phys. C.*, vol. 33, no. 04, 2022.

[54] M. A. Nivedha and S. Raja, “Detection of email spam using Natural Language Processing based Random Forest approach,” *International Journal of Computer Science and Mobile Computing*, vol. 11, no. 2, pp. 7–22, 2022.

[55] N. Grewal, R. Nijhawan, and A. Mittal, “Email spam detection using machine learning and feature optimization method,” in *Lecture Notes in Electrical Engineering*, Singapore: Springer Nature Singapore, 2022, pp. 435–447.

[56] A. Harbani and A. Sidiyantoro, “Implementasi Simple Mail Transfer Protocol Relay Pada Mail Gateway Untuk Menentukan Konten Email Spam,” *teknois. jurnal. ilmiah. teknologi. informasi. dan. sains*, vol. 12, no. 1, pp. 57–66, 2022.

[57] M. Nicho, F. Majdani, and C. D. McDermott, “Replacing human input in spam email detection using deep learning,” in *Artificial Intelligence in HCI*, Cham: Springer International Publishing, 2022, pp. 387–404.

[58] M. M. Khaleel, “Enhancement power quality with Sugeno-type Fuzzy Logic and Mamdani-type Fuzzy Logic base on DVR,” *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, vol. 3, no. 4, pp. 8273–8283, 2014.

[59] S. Abhishek, H. Sathish, A. Kumar, and T. Anjali, “A Strategy for Detecting Malicious Spam Emails using various Classifiers,” in *2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA)*, 2022.

[60] S. Kumar, S. D. Vijayakumar, R. Praveenkumar, and V. Manimala, “Developing an efficient hybrid approach for email spam filtering: SPAM-NSGA-II-NvBys,” in *Techniques and Innovation in Engineering Research Vol. 1*, Book Publisher International (a part of SCIENCEDOMAIN International), 2022, pp. 164–179.

[61] A. Al-Ajeli, E. S. Al-Shamery, and R. Alubady, “An intelligent spam email filtering approach using a learning classifier system,” *Int. J. Fuzzy Log. Intell. Syst.*, vol. 22, no. 3, pp. 233–244, 2022.

[62] K. V. Samarthrao and V. M. Rohokale, “A hybrid meta-heuristic-based multi-objective feature selection with adaptive capsule network for automated email spam detection,” *Int. J. Intell. Robot. Appl.*, vol. 6, no. 3, pp. 497–521, 2022.

[63] A. Alsharif, C. W. Tan, R. Ayop, A. A. Ahmed, and M. M. Khaleel, “Electric vehicle integration with energy sources: Problem and solution review,” *African Journal of Advanced Pure and Applied Sciences (AJAPAS)*, pp. 17–20, 2022.

[64] Y. Dou, G. Ma, P. S. Yu, and S. Xie, “Robust spammer detection by Nash reinforcement learning,” in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020.

[65] S. Smadi, N. Aslam, and L. Zhang, “Detection of online phishing email using dynamic evolving neural network based on reinforcement learning,” *Decis. Support Syst.*, vol. 107, pp. 88–102, 2018.

[66] A. Alsharif, C. W. Tan, R. Ayop, A. Ali Ahmed, M. Mohamed Khaleel, and A. K. Abobaker, “Power management and sizing optimization for hybrid grid-dependent system considering photovoltaic wind battery electric vehicle,” in *2022 IEEE 2nd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA)*, 2022, pp. 645–649.

[67] Y. Cabrera-León, P. García Báez, and C. P. Suárez-Araujo, “Non-email spam and machine learning-based anti-spam filters: Trends and some remarks,” in *Computer Aided Systems Theory – EUROCAST 2017*, Cham: Springer International Publishing, 2018, pp. 245–253.

[68] M. Belrzaeg, A. A. Ahmed, M. M. Khaleel, A. Alsharif, M. M. Rahmah, and A. S. D. Alarga, “Suspension system control process for Buses with In-Wheel Motors,” *Engineering Proceedings*, vol. 29, no. 1, p. 4, 2023.

[69] M. Zavvar, Sama technical and vocational training college, Islamic Azad University, Gorgan Branch, Gorgan, Iran, M. Rezaei, and S. Garavand, “Email spam detection using combination of particle swarm optimization and artificial neural network and support vector machine,” *Int. J. Mod. Educ. Comput. Sci.*, vol. 8, no. 7, pp. 68–74, 2016.

[70] A. A. Ahmed, M. Almabrouk, Z. M. Sheggaf, M. M. Khaleel, M. Belrzaeg, “An investigation of the effect of the hub motor weight on vehicle suspension and passenger comfort,” *Int. J. Mech. Prod. Eng. Res. Dev.*, vol. 11, no. 5, pp. 51–64, 2021.

[71] A. K. Sharma and S. Sahni, “A comparative study of classification algorithms for spam email data analysis,” *International Journal on Computer Science and Engineering*, vol. 3, no. 5, pp. 1890–1895, 2011.