Decision Making by Applying Machine Learning Techniques to Mitigate Spam SMS Attacks

Hisham AbouGrad[1], Salem Chakhar[2] and Ahmed Abubahia[3]

[1] School of Architecture, Computing and Engineering, CDT, University of East London (UEL), London, E16 2RD, UK
h.abougrad@uel.ac.uk
[2] Portsmouth Business School, CORL, University of Portsmouth, Portsmouth, PO1 3AH, UK
salem.chakhar@port.ac.uk
[3] School of Science, Technology & Health, York St John University, York, YO31 7EX, UK
a.abubahia@yorksj.ac.uk

Abstract. Due to exponential developments in communication networks and computer technologies, spammers have more options and tools to deliver their spam SMS attacks. This makes spam mitigation seen as one of the most active research areas in recent years. Spams also affect people's privacy and cause revenue loss. Thus, tools for making accurate decisions about whether spam or not are needed. In this paper, a spam mitigation model is proposed to find spam from non-spam and the different processes used to mitigate spam SMS attacks. Also, anti-spam measures are applied to classify spam with the aim to have high classification accuracy performance using different classification methods. This paper seeks to apply the most appropriate machine learning (ML) techniques using decision-making paradigms to produce an ML model for mitigating spam attacks. The proposed model combines ML techniques and the Delphi method along with Agile to formulate the solution model. Also, three ML classifiers were used to cluster the dataset, which are Naive Bayes, Random Forests, and Support Vector Machine. These ML techniques are renowned as easy to apply, efficient and more accurate in comparison with other classifiers. The findings indicated that the number of clusters combined with the number of attributes has revealed a significant influence on the classification accuracy performance.

Keywords: Machine Learning Algorithms • Feature Classification Algorithms • Decision-Making Method • Mitigating Spam Techniques • Spam Analytics Model • Mobile Network Security and Privacy Solution

1 Introduction

The levels of communication increased worldwide by using mobile devices to send and receive short text messages (SMS), which become critical for consumers because of spam messages. Indeed, spam SMS is used to access data. Thus, people have lots of concerns because of the disruptions and data loss. Also, digital service providers are affected by spam attacks, as these activities affect their reputation and growth.

Research around spam mitigation is growing as digital technologies advances that provide spammers with advanced tools to make spam attacks [1, 2]. Recent industrial research studies, in the United States, reveal that merchants will lose US$130 billion to fraud between 2018 and 2023, which comes from spam in many situations [3]. Although, financial technology (Fintech) innovations give attackers more options to collect money. Indeed, researchers are making efforts to detect spam SMS. Thus, ML techniques are considered the best way in resolving spam issues. Section 2 gives more details about the related research in this area.

In this paper, a proposed spam mitigation model is provided using the Delphi decision making method to bridge the gaps in the literature. This method is used to utilise ML techniques for better decisions regarding spam or non-spam [4]. ML algorithms contain three components, which are representation, evaluation, and optimization. First, the K-means clustering method is applied to group words based on their occurrence similarities. The text clustering implementation will be conducted with a variety of cluster numbers, which are 10, 20 and 30 clusters. Thereafter, three classification techniques were applied to the clustered data, which include: Naive Bayes; Support Vector Machine (SVM); and Random Forests. These classifiers are known for making accurate results and finding. The study indicates that the number of clusters compared to the number of attributes has a significant impact on the accuracy performance.

The paper is organised as follows: Section 2 introduces the research background and the related works that use machine learning techniques for mitigating unwanted spam messages; Section 3 describes both the study datasets and the experimental setup and presents in detail the proposed research method that is based on the Delphi decision-making method; Section 4 discusses the experimental results and interpret the main findings; and Section 5 provides a conclusion with further work.

2 Related work

In the domain of spam SMS mitigation to recognise attacks on mobile devices, many machine learning techniques are applied for decision-making to develop spam SMS recognition systems [5]. This section is a background of the most recognised previous work and makes a review of the previous ultimate algorithms applied in the study domain. This section has studies in this area to cover the following:

- Explore the usage of machine learning techniques for decision-making processes.
- What are the applied machine learning algorithms to mitigate spam?
- How machine learning techniques can be utilised to mitigate spam?

2.1 Background

People are affected by spam through their mobile devices and the development of digital technologies, especially text classification features and data mining [6, 7]. Certainly, machine learning algorithms have been found as an effective solution and most recent research has proven that these algorithms can mitigate spam to over 94%. Also, many studies proposed lots of approaches to spam classifications including

extracting specific features from text messages to generate anti-spam methods. Such a method with the use of ML classification algorithms can reach acceptable accuracy.

There are several renowned classic algorithms, which are applied for making text classifications, such as Random Forest, SVM and Multinomial Naive Bayes. These proved to be reliable as such algorithms achieved high-quality results [5]. Indeed, Random Forest can implement an in-depth classification to enhance performance by enhancing precision. Support vector machine is grouping binary classifier algorithms, which can achieve high accuracy using a hyperplane to train a judgement similar to an n-dimensional data presentation in two different places. The SVM classifier is reliable in categorising situations and indicates the presence of particular words, which are usually identified as spam [8]. Conversely, the Bayesian classification filter is a commonly utilised technology and ML technique [7, 9]. Bayesian methods like Naive Bayes become ML tools for information processing and retrieval [7]. Bayes' theorem applies the naïve assumptions by making all words unique and independent.

The frequency-inverse document frequency (TFIDF) sparse pattern is used to make a classification for the communications using SMS to be filtered to factual different classes or spam attacks according to feature classification algorithms. TFIDF is also commonly used and relies on the document, textual, or written text weighting. Thus, classification ML algorithms can be used along with TFIDF to filter SMS.

The development of machine learning analytical methods needs a decision-making framework to manage the classification process and measure the process performance to formulate the maturity model [10–12]. Thus, the Delphi method was found to be a suitable research framework to support the decision-making process to produce the study results and findings. Also, software development including programming and system analysis requires a methodology to collaborate, communicate, and work to make the implementation [13]. Hence, Agile software development (ASD) was adopted to conduct the system analysis, software development and programming [14]. Indeed, software development methods can produce the model to demote spam and prevent spammers from accessing people’s mobile devices and information [7].

2.2 Usage of Machine Learning Techniques for Decision-Making

The classification features of ML techniques can be applied to make-decisions due to their effective performance to compare and analyse to provide accurate numerical results and utilise accuracy metrics [5]. For instance, the key principle of SVM is minimizing the structural risk by finding a decision surface, which splits the instances into two classes. Also, the Random Forest algorithm is used as an ML classifier that provides high-accuracy results for predicting decisions on SMS spam. These high accuracy results can identify objects, such as SMS spam, to process them by finding the nature of such objects to accordingly process them based on requirements and decision rules.

According to Tejada et al. [6] techniques that applied SVM with a Radial Basis Function (RBF) non-linear kernel function are commonly utilised to assess activities such as daily reference crop evapotranspiration in a such specific region using limited meteorological datasets. The SVM compared to five other established empirical ML

techniques in terms of accuracy in daily activities estimation, and consequently, the results confirm that SVM made the best daily estimates. When SVM was compared to other ML techniques with analogous datasets, the SVM made high accuracies.

Improvement to make the best possible accuracies in the study results and findings may need different data mining methods for data analysis using several machine learning models [6, 9]. For example, the SVM algorithm is structured as a supervised ML technique by Vapnik [15] for data analysis and pattern recognition. SVM is commonly applied for forecasting, prediction and regression in many fields such as meteorology, agriculture, and environmental studies. In contrast, clustering methods such as ML unsupervised techniques can extract more accurate results and new knowledge using data mining methods [9]. Unsupervised descriptive data mining transfers clustering groups of data to sub-clustered datasets (sub-clusters).

Data mining statistical algorithms, such as K-means cluster analysis, have proven to achieve an outstanding performance compared to other ML algorithms, especially in providing accuracy performance [5, 9]. Of course, the accuracy of such algorithms with the use of decision-making methods rounds/steps provides significant excellent outcomes to mitigate SMS spam, secure information and protect people's privacy.

2.3 Applied Machine Learning Algorithms

The implementation of data mining's key steps, which include making data selection, pre-processing and transformation, running data mining methods, and finally, making an evaluation to produce accurate numerical metrics, results and findings [2, 9]. Based on the spam filtering study by Manaa et al. [9], there are four fundamental steps to follow using data mining methods to find spam. First, tokenise the incoming messages, so the dataset can be counted. Second, calculate the tokenised dataset after collecting it by selecting the key features from the data. Third, classify the data to prepare and make the feature vector to be ready for clustering. Fourth and final step is running the K-means model to recognise the spam cluster, and in-parallel, the Naïve Bayes model can be also applied to recognise the spam messages.

According to Pandya's [5] spam detection study, SVM efficiency can develop a spam recognition system that mitigates spam using classification and clustering-based SVM techniques. This proposed two algorithms to formulate a spam detection system, which combines clustering and classification to make what is called Clustering-based Support Vector Machine (CLSVM) system. This enhances a conventional system that uses the SVM in four steps with the use of training and testing datasets to produce a highly reliable classifier. Section 3.2 has more details.

Classification methods are supervised ML approaches, which are usually applied to make feature classification in data mining processes. Classification methods, such as the SVM classifier, eliminate the redundant features to enhance the classifiers in terms of runtime and prediction accuracy. Combining multiple methods, such as SVM and Naive Bayes, are able to find the optimum features [6, 16]. SVM can convert the data from low n-dimensions to a higher dimensional feature in such a tacit way [6]. SVM maps the relationship between input and output victor to transform data into features. Conversely, the Naive Bayes classifier uses a set of algorithms, which are

developed according to the Naive Bayes theorem [16, 17]. Thus, the Naive Bayes classification method has multiple algorithms, which work as a family to run feature classification. In general, such a family of algorithms have common rules, which are used to make each pair of features classified autonomously from each other.

2.4 Utilising Machine Learning Techniques to Mitigate Spams

Multiple ML techniques are commonly used to recognise spam messages to mitigate spam. This supports the usage of semi-supervised learning, which is learning from both, unsupervised learning and supervised learning [16, 17]. Although, Semi-supervised machine learning through different classification methods, such as SVM, Naive Bayes and Random Forests, can be applied to recognise spam messages as these methods can achieve goals by combining a small group of labelled data with a larger group of unlabelled data by the training steps. Indeed, achieving the targeted accuracy performance needs supervised learning. Thus, Random Forests can be also applied for classification. In this study, classification is used to find the number of trees in the forest, which indicates the proportion of the results, as the higher the number of trees will be achieved, the much better the accuracy of the results.

3 Methodology

The study seeks to support decision-making to mitigate spam using ML techniques, and therefore, the Delphi method rounds applied as a research framework for decision making. Also, the Delphi method is supported by Agile software development to make the core processes and programs for the experiments. ASD has the flexibility to apply the Delphi method for clustering and classification processes and measure them to make decisions on SMS messages whether they are spam or non-spam [12, 13].

The research framework used the collected datasets and ML techniques using three main stages to formulate accurate results and novel findings, which are discussed in detail in section 4. Also, the following sections explain the methodology and study framework to make the implementation toward accurate decisions and conclusions.

3.1 The Delphi Decision-Making Study Method

The Delphi method is renowned as a decision-making approach to recognise values and attributes, and therefore, it has been implemented in this study to measure Spam attacks, along with Agile methodology (Software Lifecycle) and its Scrum Sprints [12–14]. The Delphi method has three compulsory rounds, which when applied in such sequential steps, can make high quality decisions as illustrated in Table 1.

Table 1. The Delphi Method Rounds as described by Looy et al. and AbouGrad et al.[10–12].

Round	Input of the codification panel	Output of the expert panel
1	**Brainstorming** • Propose initial list of criteria • Request missing criteria	• Per initial criterion: — rate its importance — give open comments • For all criteria: — rate overall importance — give open comments • Propose missing criteria
2	**Narrowing down** • Consolidate criteria	• Per criterion: — rate its importance — give open comments • For all criteria: — rate overall importance — give open comments
3	**Weighing** • Determine final criteria • Request weightings	• For all criteria: — rate overall importance — give open comments • Weigh criteria and options

The Delphi method begins by Brainstorming to recognise the measurement indicators for the research framework [10–12]. This round is used to identify initial conditions, criteria, and classes. The second round is Narrowing Down, which is used to check and then approve the recognised indicators of the framework to obtain and be ready for the final rate of consensus to make the decision based on the framework. This is used to have the key identified values. The third and final round of the Delphi method is Weighing, which is used to make an overall evaluation.

3.2 Software Development Methodology and Algorithms

Agile methodology and the application of its resilience methods have improved data analytics processes by breaking the software development work to a set of iterations, also known as sprints, in order to coordinate and communicate between developers or development teams, which are mostly distributed in different locations [18]. Agile makes software development projects processed flexibly by supporting the project from different angles using an instrument based on process modelling to run all the project phases. Thus, the Delphi method rounds are applied to be implemented by an Agile instrument to develop the required algorithms and process them. In this study, ASD is an iterative process applied as a sprint in Scrum to produce each working piece of the software rapidly using the Scrum lifecycle [13]. Figure 1 illustrates the use of Scrum sprints in three phases where each phase implements a Delphi round.

Fig. 1. Agile methodology using Scrum sprints as described by Alsaqqa et al. [13].

The study Agile process starts from Sprint 1 to make the Brainstorming process, and then, Sprint 2 makes the Narrowing Down to check and approve the accuracy of the results, and lastly, Sprint 3 is used for Weighing to evaluate the results and validate the accuracy performance, as shown in Figure 1. The three Sprints do the clustering as

a pre-processing step followed by a classification step, which includes three different ML techniques using the proposed system by Pandya [5]. For the pre-processing step, the dataset is divided to two parts, the first part is the training dataset part, e.g. 80%, and the second part is the testing dataset part, e.g. 20% [4]. This uses the training data for feeding the algorithm to be trained where testing data is used for validation. The pre-processing uses clustering as a unique data cleaning feature for accuracy and to improve the data quality, but the disadvantage of this is time increase during data processing, which is overtaken by improvements in the levels of prediction (Figure 2).

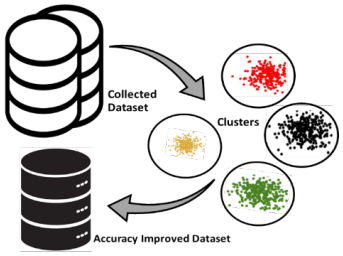

Fig. 2. The pre-processing step for clustering dataset.

When the pre-processing step is done and the collected dataset becomes clustered dataset, then the classification step can be implemented using the clustered dataset. The classification step process uses similar procedures, which are used in the previous process. Figure 3 illustrates the main concept of the classification step.

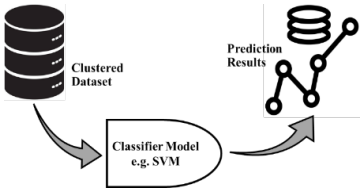

Fig. 3. The classification step to produce prediction results.

Clustering and classification algorithms produce 100% accuracy with high reliability outcomes [5]. For example, the clustering-based SVM system execution time was about 14.43 seconds for 5064 records, which is a great reliability number. Indeed, the CLSVM system has proven higher accuracy and better timing compared to SVM.

3.3 Data Collection and Processing

The study collected two datasets for the experimental setup in a comparable format. These were chosen from the Kaggle website as a trusted source. To find the dataset #1 URL link is https://www.kaggle.com/code/balaka18/email-spam-classification/data, and the dataset #2 URL link is https://www.kaggle.com/datasets/uciml/sms-spam-collection-dataset. The datasets are publicly available, and when the data is reviewed, both datasets are found to be reliable. Table 2 shows the description of the datasets.

Table 2. Description of the study datasets.

Dataset No.	Number of Rows	Number of Columns
1	5172	3002
2	5574	4

Table 2 depicts the used datasets in terms of the number of rows (records) and a number of columns (attributes). It is clear that dataset #2 has a greater number of attributes than dataset #1. Accordingly, more efforts in data reduction choices are required as this helps to focus on specific sets of attributes to be more efficient and effective.

The datasets reviewed by making data processes and TFIDF calculations feature to transform messages (text) into numerical vector data to be executed using algorithms. The dataset messages were reviewed first as these messages are the most important attribute of the research because spammers use messages for spamming. Thus, stop words removal and porter stemming are used to recognise terms (words), and then, the key spam terms are selected. Finally, the datasets were calculated using the TFIDF score for the recognised terms and label them as unique term features (UTF).

3.4 Experimental Study and Implementation

Three ML techniques were applied to conduct the experiment, which is discussed in the previous sections. These methods help to recognise SMS spam through all rounds of the study to mitigate spam attacks. Table 3 shows an example to demonstrate how to use the identified spam terms to count and recognise spam messages.

Table 3. example of message recognised terms profile matrix.

SMS No.	should	get	out	see	price	time	…	classification
SMS 1.	1	1	1	0	0	0	…	spam
SMS 2.	0	0	0	0	1	0	…	ham
SMS 3.	0	0	0	0	0	1	…	ham
SMS 4.	0	1	1	1	0	0	…	spam
SMS 5.	0	0	1	1	0	0	…	spam

The experiments were executed using computer devices, and the minimum device specifications are Intel (R) Core (TM) i5-4200U CPU@1.60GHz 2.30GHz Processor, 8.00GB RAM, and Windows 10 Pro 64-bit operating system for x64-based processor. During these experiments, three different numbers of clusters (i.e. 10, 20 and 30) were applied to perform data clustering. Thus, three algorithms are applied in combination for the execution of the proposed system. The first algorithm is the improved Naive Bayes, which is used to fix problems, such as the tendency to correct the word positioning by more than 10% in comparison to negative word accuracy [17, 19].

The study has applied K-means Clustering during the second round (phase) of the implementation using Algorithm 1. This method uses data from the first round for data pre-processing where the data is converted to vectors using TFIDF vectorizer as shown in Step-1 of Algorithm 2. After the data input, seven steps are followed, as shown in Algorithm 1 to conduct the dataset clustering in order to produce clustered dataset to be prepared for the data classification as shown in Step-3 of Algorithm 2.

Algorithm 1: K-means Clustering Method
Input: Vectorised SMS Dataset (see Algorithm 2, **Step-1**) **Output:** Clustered (grouped) Data **Step-1:** Select the number K to decide the number of clusters **Step-2:** Select random K points or centroids (it can be other from the input dataset) **Step-3:** Assign each data point to its closest centroid to make the predefined K clusters **Step-4:** Calculate the variance and place a new centroid of each cluster **Step-5:** Repeat **Step-3** for reassign to each datapoint to the new closest centroid of each cluster **Step-6:** If any reassignment occurs, then go to **Step-4** else go to **Step-7** **Step-7:** FINISH, the dataset is clustered (grouped) and ready for classification

The study experiment has three main steps to mitigate SMS spam using a K-means clustering based classification model, as shown in Algorithm 2. This uses a dataset, which is collected from SMS communications to provide high accurate and classified dataset. The first step is data Pre-processing with the use of TFIDF, which is used for removing the missing values, duplicated values, and stop words. The second step is data clustering (Algorithm 1), which consists of clustering the pre-processed data through 10, 20, and 30 clusters. The third step is conducting data classification through classifying the output dataset from Algorithm 2 Step-2 to make a comparison.

Algorithm 2: K-means Clustering based classification model
Input: Collected Communication Messages Dataset **Output:** High accurately classified dataset to mitigate spam attacks **Step-1:** *Data Pre-processing* by TFIDF Vectorizer for: Removing the missing values; Removing duplicated values; and Removing stop words **Step-2:** *Data Clustering* by using K-means Clustering Method (Algorithm 1). This includes clustering the pre-processed data into 10, 20, 30 clusters **Step-3:** *Data Classification* by classifying the output dataset from **Step-2** to compare the classification accuracy for: i. Naive Bayes; ii. Support Vector Machine (SVM); iii. Random Forests

For training and testing, the dataset has been split into two parts. The first part is used for training using 75% of the dataset. The second part is used for testing the dataset using 25% for validating the algorithm.

4 Discussions and Findings

The study results are explained here to discuss the key findings. The facts and outputs presented describe what the methodology and implementation produced.

4.1 Delphi's Reliability and Validity for Decision-Making

The study found that key objectives for decision-making studies can be achieved through the Delphi method, as consensus and stability have been notably experienced by processing spam datasets in three different rounds to reach consensus on several significant aspects. Also, the study approved that the Delphi method found practical issues and key indicators by weighing decision-making criteria using Delphi's rounds as a multiple criteria decision-making (MCDM) framework, which complies with other studies [11, 12]. The Delphi method confirms the requirements, and then, makes

decisions using Delphi's rounds as an MCDM process. According to AbouGrad et al. study [12], the Delphi method can identify, select, conceptualise, and validate factors. Hence, the Delphi method examines the algorithm's validity in the weighing round using a quantitative assessment of the reliability and validity of the model.

4.2 Clustering based classification for Decision-Making to Mitigate Spam

The K-means clustering is followed by three algorithms to formulate the study classification model, as shown in Algorithm 2. The main results and findings of the proposed work are exhibited in Tables 4 and 5. These tables demonstrate each number of clusters using ML algorithms and their results, which provide classification accuracy through the different steps using Algorithm 1 and Algorithm 2.

Table 4. Dataset #1 based Experimental Results.

Number of Clusters	Classification Algorithm	Classification Accuracy
10	Naive Bayes	0.75
	SVM	0.71
	Random Forests	0.77
20	Naive Bayes	0.84
	SVM	0.80
	Random Forests	0.87
30	Naive Bayes	0.94
	SVM	0.89
	Random Forests	0.97

According to dataset #1, as shown in Table 4, the Random Forests outperforms both classifiers of Naive Bayes and SVM. By using 10 clusters based on the experiment, Random Forests classifier scores an accuracy of approximately 77% while Naive Bayes scores 75% and SVM scores 71%. This means that the Naive Bayes classifier is still outperforming the SVM classifier. Looking at the observations from both 20 clusters and 30 clusters, the classification accuracy scores support the hypothesis for concluding that the Random Forests classifier gives better prediction than both Naive Bayes and SVM classifiers. Thus, the variations in the number of clusters make it clear that the more number of clusters, the higher the accuracy score can be achieved, which means 10 clusters increase with classification accuracy score by an extra 10%.

Table 5. Dataset #2 based Experimental Results.

Number of Clusters	Classification Algorithm	Classification Accuracy
10	Naive Bayes	0.91
	SVM	0.86
	Random Forests	0.94
20	Naive Bayes	0.92
	SVM	0.87
	Random Forests	0.95
30	Naive Bayes	0.93
	SVM	0.88
	Random Forests	0.96

According to dataset #2 (Table 5), the Random Forests classifier again outperforms Naive Bayes and SVM. In 10 clusters based on the study experiment, Random Forests

classifier scores an accuracy of about 94% while Naive Bayes scores 91% and SVM scores the lowest of 86%. This indicates that the Naive Bayes classifier is still outperforming. Looking at the observations from both 20 clusters and 30 clusters based on the experiment, the classification accuracy scores support the hypothesis for concluding that the Random Forests classifier produces better predictions than both Naive Bayes and SVM classifiers. Thus, a greater number of clusters lead to a higher classification accuracy score. This means 10 more clusters result in increasing the accuracy score by about 1%. The differences in incremental rates between dataset #1 (10%) and dataset #2 (1%) lead to conclude that the number of clusters in combination with the number of attributes have a significant influence on the classification accuracy. Also, an overall comparison of the classification accuracy has been implemented to identify how each classifier is performing (Figure 5). Indeed, the differences between classification algorithms illustrate that the Random Forests classifier first by 35%, and then the Naive Bayes by 33%, and the SVM classifier by 32%, as shown in Figure 5.

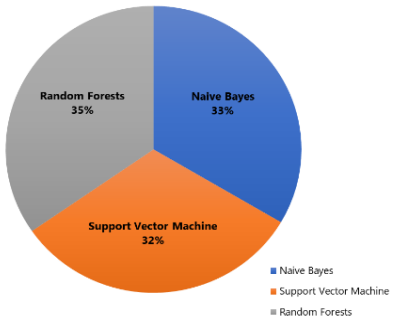

Fig. 5. Overall classification accuracy comparison between the selected classifiers.

5 Conclusion

The Delphi method with Agile software development as an iterative process has been applied. This led to utilising a sprint in Scrum to produce each piece of software in such a rapid approach using Scrum lifecycle. Indeed, this research work presented a machine learning based technique for the better decision of spam or non-spam to mitigate spam attacks. Also, the clustering-based classification algorithm applied K-means clustering to group words. Afterwards, three different classifiers were applied. The findings indicated that clusters number in combination with the attributes number produced a significant influence on the classification accuracy performance.

References

1. Aliza, H.Y., Nagary, K.A., Ahmed, E., Puspita, K.M., Rimi, K.A., Khater, A., Faisal, F.: A Comparative Analysis of SMS Spam Detection employing Machine Learning Methods. In: 2022 6th International Conference on Computing Methodologies and Communication (ICCMC). pp. 916–922. IEEE (2022).

2. Delen, D.: Predictive Analytics: Data Mining, Machine Learning and Data Science for Practitioners. Pearson Education, Inc., Old Tappan, New Jersey (2021).
3. King, S.T., Scaife, N., Traynor, P., Abi Din, Z., Peeters, C., Venugopala, H.: Credit Card Fraud Is a Computer Security Problem. IEEE Secur. Priv. 19, 65–69 (2021).
4. Achchab, S., Temsamani, Y.K.: Use of Artificial Intelligence in Human Resource Management: "Application of Machine Learning Algorithms to an Intelligent Recruitment System". Presented at the (2022).
5. Pandya, D.: Spam Detection Using Clustering-Based SVM. In: Proceedings of the 2019 2nd International Conference on Machine Learning and Machine Intelligence. pp. 12–15. ACM, New York, NY, USA (2019).
6. Tejada, A.T., Ella, V.B., Lampayan, R.M., Reaño, C.E.: Modeling Reference Crop Evapotranspiration Using Support Vector Machine (SVM) and Extreme Learning Machine (ELM) in Region IV-A, Philippines. Water. 14, 754 (2022).
7. Kim, S.-E., Jo, J.-T., Choi, S.-H.: SMS Spam Filterinig Using Keyword Frequency Ratio. Int. J. Secur. Its Appl. 9, 329–336 (2015).
8. Reaves, B., Vargas, L., Scaife, N., Tian, D., Blue, L., Traynor, P., Butler, K.R.B.: Characterizing the Security of the SMS Ecosystem with Public Gateways. ACM Trans. Priv. Secur. 22, 1–31 (2019).
9. Manaa, M., Obaid, A., Dosh, M.: Unsupervised Approach for Email Spam Filtering using Data Mining. EAI Endorsed Trans. Energy Web. 8, 162–168 (2021).
10. Looy, A., Poels, G., Snoeck, M.: Evaluating Business Process Maturity Models. J. Assoc. Inf. Syst. 18, 461–486 (2017).
11. AbouGrad, H., Warwick, J., Desta, A.: Developing the Business Process Management Performance of an Information System Using the Delphi Study Technique. In: Reyes-Munoz, A., Zheng, P., Crawford, D., and Callaghan, V. (eds.) EAI International Conference on Technology, Innovation, Entrepreneurship and Education, Lecture Notes in Electrical Engineering (LNEE) book series, Volume 532. pp. 195–210. Springer International Publishing, Cham (2019).
12. AbouGrad, H., Warwick, J.: Applying the Delphi Method to Measure Enterprise Content Management Workflow System Performance. In: Arai, K. (ed.) Intelligent Computing Proceedings of the 2022 Computing Conference, Volume 2. pp. 404–419. Springer International Publishing, Cham (2022).
13. Alsaqqa, S., Sawalha, S., Abdel-Nabi, H.: Agile Software Development: Methodologies and Trends. Int. J. Interact. Mob. Technol. 14, 246 (2020).
14. Martin, R.C.: Clean Agile: Back to Basics. Pearson, Boston (2020).
15. Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer New York, New York, NY (2000). https://doi.org/10.1007/978-1-4757-3264-1.
16. Bhattacharya Sohom, Bhattacharjee Shubham , Das Anup , Mitra Anirban , Bhattacharya Ishita, G.S., Bhattacharya, S., Bhattacharjee, S., Das, A., Mitra, A., Bhattacharya, I.: Machine learning-based Naive Bayes approach for divulgence of Spam Comment in Youtube station. Int. J. Eng. Appl. Phys. 1, 278~284 (2021).
17. Kamble, M., Dule, C.: Review Spam Detection Using Machine Learning: Comparative study of Naive Bayes, SVM, Logistic Regression and Random Forest Classifiers. Int. J. Adv. Res. Sci. Technol. 7, 292–294 (2020).
18. Biesialska, K., Franch, X., Muntés-Mulero, V.: Big Data analytics in Agile software development: A systematic mapping study. Inf. Softw. Technol. 132, 106448 (2021).
19. Khurshid, F., Zhu, Y., Xu, Z., Ahmad, M., Ahmad, M.: Enactment of ensemble learning for review spam detection on selected features. Int. J. Comput. Intell. Syst. 12, 387–394 (2018).