A lifelong spam emails classification model

Rami Mustafa A. Mohammad
*Department of Computer Information Systems,
College of Computer Science and Information Technology,
Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia*

Received 25 November 2019
Revised 14 January 2020
Accepted 20 January 2020

Abstract

Spam emails classification using data mining and machine learning approaches has enticed the researchers' attention duo to its obvious positive impact in protecting internet users. Several features can be used for creating data mining and machine learning based spam classification models. Yet, spammers know that the longer they will use the same set of features for tricking email users the more probably the anti-spam parties might develop tools for combating this kind of annoying email messages. Spammers, so, adapt by continuously reforming the group of features utilized for composing spam emails. For that reason, even though traditional classification methods possess sound classification results, they were ineffective for lifelong classification of spam emails duo to the fact that they might be prone to the so-called *"Concept Drift"*. In the current study, an enhanced model is proposed for ensuring lifelong spam classification model. For the evaluation purposes, the overall performance of the suggested model is contrasted against various other stream mining classification techniques. The results proved the success of the suggested model as a lifelong spam emails classification method.

Keywords Concept drift, Spam, Lifelong classification, Mining data streams

Paper type Original Article

1. Introduction and background

On 3-May-1978 and on the west coast of the United States, the first spam email was dispatched to 393 ARPANET users (all users at that time) by a "Digital Equipment Corp" marketing agent called Gary Thuerk [1]. Instead of sending a single email to each user, Gary Thuerk sent one email to all customers to inform them of the release of a new model of computers. Although the reply of the customers was considered very low, but the email had resulted in a lot of sales. Ten years later, the spam emails appeared again with a subject title "Make Money Fast" [1]. Thus, began the story of spam emails until they become a nagging problem that annoys email users in a daily base. Recently, a newly released statistical research [2] showed that in 2018 the amount of email beneficiaries reached more than four

The publisher wishes to inform readers that the article "A lifelong spam emails classification model" was originally published by the previous publisher of *Applied Computing and Informatics* and the pagination of this article has been subsequently changed. There has been no change to the content of the article. This change was necessary for the journal to transition from the previous publisher to the new one. The publisher sincerely apologises for any inconvenience caused. To access and cite this article, please use Mohammad, R. M. A. (2020), "A lifelong spam emails classification model", *Applied Computing and Informatics*. Vol. ahead-of-print No. ahead-of-print. https://10.1016/j.aci.2020.01.002. The original publication date for this paper was 23/01/2020.

Declaration of Competing Interest: The authors declare that there is no conflict of interest regarding the publication of this article.

Applied Computing and Informatics
Vol. 20 No. 1/2, 2024
pp. 35-54
Emerald Publishing Limited
e-ISSN: 2210-8327
p-ISSN: 2634-1964
DOI 10.1016/j.aci.2020.01.002

billion clients. This signifies a rise of one hundred million clients than the prior year. The same investigation pointed out that in 2009 the amount of email clients is projected at almost two billion clients. This tells that the amount of email clients has multiplied within just ten years. In fact, the projected worldwide population in 2018 is almost seven billion persons. Meaning that almost fifty percent of the entire globe benefits from email services. During 2022, the amount is believed to exceed a lot more than four billion. Actually, the amount of email accounts is greater than the amount of clients simply duo to the fact that a lot of clients may have several email accounts. Paid, domain based, and organizational email accounts constitute more than 25% of the total number of email accounts worldwide. 86% of industry experts picked email as their favored interaction medium. Almost a hundred and fifty billion email messages are sent daily. Such statistics reveal how critical the email is in peoples' professional and social activities endeavors. Unfortunately, the email merits seem to be devastated with the remarkable utilization of annoying, damaging, unethical and misleading email messages, which were frequently sent indiscriminately by dishonest persons who do not have any immediate relationships with the receiver. These kinds of email messages are typically denoted to as spam email messages. You may wonder as to why people open a message received from an unknown sender. Essentially, the "From" field as well as the "subject" field within an email message would be the primary promoters to make the decision whether to read a message or completely ignore it. Clients typically open a message if the "from" field implies that the message is received from a well-known person [3]. Nevertheless, in some cases clients lean towards reading and replying the inbox email messages regardless if such email messages appeared to be coming from anonymous sender because of the human's curiosity nature. Above all, it has been proven that 33% of clients open a message if it contains an attractive subject theme no matter whether the sender is well known or not [4]. Those who dispatch spam emails are usually called "Spammers". These people are intelligent enough to take advantage of each and every opportunity of sending alluring spam email messages. Spammers tend to combine social engineering and technical tricks when emailing spam email messages [5,6]. Let us remember how spammers could take benefit of "Hurricane Irma" for the purpose of distributing the spam email messages in 2017 [7]. Nepal earthquake is a second example which might be reported in this regard [8]. Such disasters have had calamitous results not merely on the citizens in which these disasters occurred but even on the global human society. However, these disasters were found to be like a gift to spammers. Sport events are actually an additional preferred subject to spammers. The impertinence of spam emails producers gone up to the level of harassing celebrities while not paying care about their sociable, economical, nor the political position. Among those people, are the spam emails which unfortunately used the identity of the 1st Lady of the UNITED STATES "Melania Trump" [9]. Marketing email messages are actually another way to allure people. In fact, it has been stressed that such type of email messages is considered the most widely used amongst others followed by adult-related advertisements. Spam email messages related to monetary matters rated in the third place [10]. Generally, spammers are intelligent people who take advantage of any sort of opportunities to deceive naive users. Therefore, these people are expected to benefit from any kind of news of interest to the community so as to spread spam email messages. It must be stated that a newly released survey revealed that most of the spam email messages are started from the United States, Russia, and China [11]. Spam email is frequently drafted in English language. Though recently the number of English language drafted spam emails reduced noticeably to ninety percent after hitting ninety six percent a few years earlier [12]. This drop is in fact offset in the increase of spam emails that are drafted by means of various languages. Consequently, we might assume that spam emails have become substantially more "international". The potential risks carried by spam email messages involve a number of other threats such as fraud, scam, as well as, phishing websites [13,6,14]. Belarus produces the highest spams per capita [12]. Spam emails counts more than 45 percent of emails traffic across the world [10]. Which indicate that approximately half of the email messages that are sent daily happen to be spam emails. Such a ratio appears to be horrifying, in particular when we can say that the quantity of delivered emails during 2017 is nearly 250 billion emails daily, a lot more than 100 billion of those happened to be spam emails. Which means that the world-wide per capita of spam email messages reaches approximately 16 spam email messages daily. The good thing is that spammers receive a single reply for almost 12 million spam email messages. Nevertheless, with so few people responding, you can ask what is the motivation behind spam email campaigns? Interestingly, it has been revealed that over one-year spam emails creators could gather almost $3000000 dollars even though barely a single respond every 12,000,000 emails emailed [10]. In early 2012, spam emails costed companies almost twenty billion dollars. This amount of money is likely to surpass 250 billion dollars in couple of years.

Considering the computerized innovation that we are witnessing nowadays, raw data tend to be easy to obtain, store, and analyze. The real benefits of such obtained data can be determined by their capacity to offer valuable details and facts that could assist in making decisions or perhaps comprehending the domains governing the source of the data. Typically, data analysis is a traditional manual task in which experts should familiarize themselves with the domains and by using different statistical methods they might develop summaries and reports to explain knowledge insight. Nevertheless, this method might suddenly deteriorate if the dimension and size of data expand. Consequently, if the data exploration goes beyond capabilities of regular manual investigation, experts start seeking a bit more dependable knowledge extraction solution. Data Mining (DM) is looked upon as a powerful approach to simplify generating probably valuable knowledge for making accurate decisions. DM or occasionally known as *"Knowledge Discovery"* [15–17] is a technique of examining datasets from diverse points of views and then offering them in useful and practical forms. Classification is a broadly researched technique in the DM. It is normally defined as the task of designing a classifier from the historical dataset to predict the value of a class variable associated with an unseen instance [82]. Classification is a supervised learning strategy since it is derived using a preliminary dataset in which the training instances are offered along with their related class labels. Commonly, a classification algorithm finds patterns and rules from a stationary dataset in which the set of input attributes tend not to frequently change. Such dataset is consequently thought to include all information needed to find out important knowledge regarding the underlying domain(s). This kind of learning methodology, however, has confirmed impractical for several real-world situations such as spam email classification where the training datasets are usually attained progressively in streams of examples rather than all training dataset examples are offered before start training the DM model. Further, this learning methodology will not be able to deal with new knowledge once new datasets become available. In general, growing datasets might cause a model working in dynamic situations, that might result in the what's normally known as *"Concept Drift"* issue [18,19]. Concept drift denotes the change in the relationships binding the input attributes (*"features"*) to the class variable(s). Normally, learning in the occurrence of a constantly evolving datasets demands a model that could be modified frequently to be able to leverage the newly gathered dataset, while at the same time preserving the abilities of the classification model on old datasets. One of the most well-known learning strategies which can efficiently deal with concept drift as well as catastrophic forgetting issues and produce a lifelong classification model is the ensemble learning strategy [20]. This strategy ensures that the classification algorithms learn new knowledge and maintain all historically learned knowledge.

Spam emails are deemed a common type of the dynamic classification problems in which the dataset instances are constantly arriving. Spam enticed researchers to deal with it from an intelligent perspective. Within the last decades, a lot of anti-spam solutions have been released. Yet, almost all of the state-of-the-art DM based anti-spam solutions consider the spam emails being a stationary problem in which the full datasets will be introduced to the DM algorithms to learn a new classification model. Nevertheless, spam emails classification is in fact an evolving classification issue where the group of characteristics which could be utilized to decide on the class of an email are actually continuously evolving. However, the concept drift occurring in spam emails is regarded as a cyclical concept drift because the list of characteristics used for forecasting spam emails may disappear at certain period of time, but they might return to reappear once again later on. Moreover, it could not be guaranteed that no spam email might come from past campaigns; but some are still existing. Therefore, learning a new model from the beginning could safeguard users from spam emails produced from fresh campaigns, however, it could leave them vulnerable to spam emails produced from past campaigns as an immediate consequence of the commonly termed "Catastrophic Forgetting" [21]. Catastrophic forgetting occurs if the previously acquired knowledge is lost once a new knowledge has been learned. Hence, Catastrophic forgetting is an additional concern which we try to resolve in the current research by making sure that generating a new model does not necessarily mean ignoring the already available model(s). Hence, handling catastrophic forgetting is an essential step towards producing a lifelong spam classification model. Generally speaking, creating a classification model which could be modified continually in order to keep up with any possible alterations which might impact the overall performance of the spam classification model is a vital and timely issue. The model is assumed to give a high true positive (TP) *("the proportion of genuine emails which were accurately labeled as genuine")* and true negative (TN) *(the proportion of spam email messages accurately classified as spam)* ratios. Overall, the motivation of doing the present investigation is to suggest an innovative model which can be used for creating a lifelong classification model that is capable to gain knowledge from growing datasets. The model will be after that applied to a critical internet security issue. Specifically, the spam email classification problem in an attempt to examine if the model is capable to cope with the concept drift which usually characterizes the spam email messages. An additional goal of applying the suggested model to spam emails is to determine whether it can deal with the catastrophic forgetting issue by guaranteeing that the model learns incrementally and on the top of that it maintains the formerly learnt knowledge. This article is divided into 5 sections. The-State-of-the-art in spam filtering domain is reviewed Section 2. The proposed algorithm is presented in Section 3. Section 4 is dedicated for assessing the performance of the suggested algorithm. Lastly, the obtained results are discussed in Section 5.

2. State of art in spam filtering domain

This section starts by exploring the cost of spam emails misclassification and them it sheds light on the contemporary spam classification approaches.

2.1 Cost of spam misclassification

Emails misclassification can be measured using either FN or FP rates. However, the cost of wrong classification depends mainly on the filtering context. To elaborate further, if a legitimate company sent a legitimate email and this email is classified as a spam one (*false positive*) then this might result in losses from both the company that sent the email, and from the customers that this email is supposed to reach. On the one hand, the company's reputation may be tarnished, or it may lose marketing opportunities that could generate profits. On the other hand, clients and customers may lose a profitable investment opportunity. In a completely contradictory scenario, if a spam email is considered as a genuine one (*false negative)*, this may result in financial losses from the customer who interacts with this email. There are several factors that may lead to email misclassification, amongst them we mention [22]:

- The utilization of various tactics in order to avoid the quick recognition of several keywords such as “legitimate”, “Iegitimate”, “1egitimate”, “le8itimate”, etc.
- Long or tiny URLs.
- IP-based URLs.
- Adding Prefix and Suffix to the domain part of the URL.
- Representing text using images.

Thanks to the regular expressions *(“regex”)* that facilitate identifying URL obfuscations [23]. Regular expressions proved their merits in reducing the misclassification of spam emails *(“especially the false positive emails”)* because of their implicit capacity to efficiently filter out and discover all URL obfuscations tactics employed by spammers [22]. Regular expressions can also be used for saving the computational cost if they are used as a pre-filter phase in the sense that no need for doing any additional analysis if an email message matches a regular expression pattern and is classified as spam [22].

2.2 Human based spam classification approach

Blocking spam emails prior to reaching the clients mail boxes is certainly the ultimate objective among all anti-spam tools. Ordinarily, the first line for defending against spam emails is the email clients. Several solutions rely on the honest and committed work of email clients in reporting malevolent, unsolicited, and annoying email messages [24,25]. These strategies are usually known as “Community Reporting Strategy”. The essential principle of these approaches is *“Users should honestly, assess and then report all spam emails they might came across in their mailboxes so as others could be made aware of such emails”*. Accordingly, this approach depends on the users’ knowledge in determining spam email messages. Nevertheless, this brings an exceptional burden on the email users to determine if a particular email message is actually a genuine or possibly a harmful one. Individuals are likewise evaluated based on their record of accurately revealing spam emails. The better the users’ record has become, the more reliable their reports is going to be. Nevertheless, this strategy can be described as a tedious method where a user has to dedicate a significant amount of his time in examining and then reporting spam emails found in their mailboxes. Principally, spammers realize that usually the human factor is the weakest chain in any protection system [21,26]. Additionally, end users might examine their mailbox to find spam email messages and if they come across some, they might delete them rather than reporting them. Therefore, a lot of useful information might be lost, although this information might be valuable to service providers when making a decision whether a particular email is spam and whether to immediately put it in the spam folder or not. In general, the more users report spam email messages, the better spam filtration a service provider may provide. Above all, users still didn’t do the complete jobs expected from them in the sense that they will traditionally report *“False Negative” (FN)* email messages (*“spam email messages which were incorrectly categorized as valid”*), but few of users might report *“False Positive” (FP)* emails (*“legitimate email messages which were incorrectly categorized as spam”*). Consequently, to do their task completely, end users need to examine not merely the inbox but also all the other folders including spam and junk folders. One more strategy to dealing with spam campaigns may be by the enactment of strict legal procedures that guarantee that perpetrators will be punished, and at the same time safeguarding potential users. The first law regulations that would criminalize spam email messages was enacted in Argentina in 2001 [27]. Several other countries have also enacted law regulations that criminalize spam mails [3]. Nevertheless, legal measures are actually hard to implement in practice mainly because spammers may start the spam campaigns and after that, they can easily disappear into the cyberspace.

A recent study [28], offer a “Credible and Personalized Spam Filtering Scheme” (CPSFS) depending on social reliability and interest likeness, in which individuals report his/her received spam email messages to their friends and followers in social media. CPSFS categorize spam into 2 classes: “complete-spam” and “semi-spam”. The first class is actually concerning every user. On the other hand, the second class are emails seen as spam by some people and as legitimate ones by others. The experiments revealed that the precision of the CPSFS surpasses the traditional Bayesian filtering algorithm and several various solutions.

2.3 Non-profit and for-profit based spam classification approaches

Recently, a lot of non-profit, for-profit, as well as empirical researches are carried out with the aim of overcoming spam email messages with very little end user participation. The *“Spamhaus project”* is a sort of non-profit enterprises that maintain a record of spam email messages as well as related risks including malware, botnet, and most importantly phishing [29]. SpamAssassin is another example of non-profit tools that filters spam emails [30]. SpamAssassin is a well-known filtering platform used to incorporate various anti-spam approaches. Since its establishment, SpamAssassin is considered the starting point for developing commercial anti-spam solutions such as “McAfee SpamKiller” [31] and “Symantec Brightmail” [32]. SpamAssassin enables the administrator to specify particular filters by developing filtration rules as necessary. In addition, carrying out a quick search through the internet, one will discover plenty of commercial anti-spam applications. Overall, the main success factor any of anti-spam tools relies on identifying spam email messages accurately before arriving at the users’ mailboxes.

2.4 Intelligent spam classification approaches

Besides various other elements, email messages are typically consisting of texts. Therefore, it does make sense to utilize text mining approaches to mitigate spam email messages. Several DM and ML approaches have been utilized for producing text mining based spam prediction systems such as Naïve Bayes [33,34], Neural Network [35], Support Vector Machine [36], and Nearest Neighbor [37]. Typically, text-mining based solutions begin by collating 2 datasets of labeled emails these are a set of spam emails and set of legitimate ones. Such emails happen to be deconstructed into several phrases and terms. Thereafter, the appearance percentage of every term/token is determined on every dataset. Once a new hidden email arrives, the set of terms/tokens which arise with greater frequency are believed as proof of the fact that an email is a spam or a legitimate one.

A lot of intelligent solutions have already been created with the aim of reducing the false negative ratios. Spam emails classification is frequently regarded as a binary classification problem considering that every email might be either spam or genuine. A lot of ML methods have proven to be appropriate for binary classification domains including spam email messages recognition. For instance, Logistic Regression [38], Neural Network [39], Support Vector Machine [40]. Naïve Bayes has also been effectively employed for detecting spam email messages [41]. Among the leading researches which employed Naïve Bayes is the research which tried to incorporate 3 Naïve Bayes classifiers to be able to increase the effectiveness of regular Naïve Bayes in recognizing spam emails [42]. One of the classifiers divides the training dataset into two possible class labels those are “spam” and “non-spam”. Another classifier utilizes linear programming towards enhancing the thresholds that determine the fine lines which usually isolate spam emails from legitimate ones. On the other hand, the last classifier combines the results of two other classifiers. The evaluation outcomes demonstrated that this approach outperforms the conventional Naïve Bayes. Naïve Bayes also has provided improvement once added to features extraction methods for example, “Cost-Sensitive Multi Objective Genetic Programming” [43]. Neural Network has positively contributed when it comes to discovering spam emails. The effort made by Ozgur et al. [44], confirmed that claim. Nonetheless, the experimental outcomes were not satisfactory. The authors in [45] introduced a rule-based system in which 23 traits are recognized out of a privately gathered spam datasets. All the features were after that designated a score. To be able to label the emails as being ham or spam, the gathered scores are compared to a predefined threshold value. Three different ML algorithms were used in the experiments those are Decision Tree, Naïve Bayes, and, Neural Network however the research is carried out using a small database of only 750 ham and spam email messages. The capacities of the type-2 fuzzy sets in forecasting spam emails have been evaluated [46]. This method offered the individuals a chance to decide on the type of the emails he wants to block by just using the thesaurus associated with that category. A server-side tool designed to blocking spam emails was developed and is known as SpamGuru [47]. SpamGuru offers a scores which usually varies from zero to one thousand for every email. The greater the scores, the more harmful the email can be. Clients could also report spam emails in order that the tool may increase its effectiveness through gaining knowledge constantly. The SpamGuru produces a directory containing a collection of emails which the tool could not provide a precise decision whether or not they happen to be spam emails or legitimate ones. Once a message is designated being a spam email, SpamGuru gives 4 plausible choices:

1- Deleting the email forever.

2- Archiving it.

3- Directing it to challenging queue that asks the sender to authenticate his ID.

4- Tagging the email as a probable spam and provide it to the end user in order to choose the proper action.

As stated previously, the existing study considers the spam emails as a supervised classification issue. However, many different studies handled it as an unsupervised classification problem in which a set of unlabeled instances are utilized for developing classification models. One example of the unsupervised classification systems is the research study prepared in [48] that primarily presumes that spam emails frequently are a part of a "spam campaign" and they're infrequently dispatched singularly. Consequently, they might be identified using "campaign signature". This method identifies spam campaigns as being a group of remarkably interrelated emails that happen to be reported by many users. Consequently, spam campaigns are commonly launched through setting up a single spam email and after that produce a lot of replicas of it by maintaining several parts unchanged and obfuscating the rest with the purpose of tempting the email clients. Identifying the unchanged and obfuscated sections is certainly an essential phase on the way to determining the campaign signature(s). The experimental analysis revealed competitive outcomes in comparison with various other supervised classification methods. A research study [49] was executed in 2007 to assess the performance of several DM and ML methods including *"Decision Trees", "Naïve Bayesian", "Support Vector Machine", and "Neural Network".* Many evaluation factors are generally employed for assessing the performance of the generated models process. The results revealed that the "Decision Tree" and "Naïve Bayesian" classifiers provided greater results when compared to "Neural Network" as well as "Support Vector Machine". Many researchers investigated the applicability of DM and ML strategies in predicting the non-English Language emails. For example, the work done in 2009 [50] was executed to measure the overall performance of 6 varied DM and ML methods in filtering spam email messages created in Arabic. The study finds out that little extra attributes need to be used to be able to enhance the accuracy of forecasting spam emails written and published in Arabic Language. One more research study has explored the

Turkish Language composed spam email messages [51]. The authors applied text elements as well as raw elements within an email message to determine if a message is actually a legitimate one or a malicious message. For the purpose of evaluating their model, the authors have used AdaBoost ensemble technique and the attained results were very encouraging. An Adaptive Neural Fuzzy Interference System was offered in 2012 for the purpose of classifying spam email messages [52] in which 5 various criteria are utilized for determining spam email messages these are: *"number of associated user pages, number of times marked as spam, text priority, presence of URL or Hyperlinks, and number of common timestamps"*.

In 2018, an innovative technique for detecting spam emails by transforming information from emails content into features by applying "WordNet ontology" [53]. This study concluded that concept drift should be taken into account when building any spam detection solution to ensure a lifelong spam emails classification system. Per the authors point of view, ensemble-based classification approach is a viable technique for building such lifelong anti-spam models. After applying WordNet Lexical Database, the study come up with several interesting finding, those are as follows:

- The occurrence of concept drift in legitimate email.
- Several subjects impacted by concept drift are overlapped in a single class, yet not in the other class.

These conclusions necessitate and motivate the need for building an anti-spam detection model that is able to contend with the dynamic nature of spam emails.

In 2006, A "Fuzzy Rule Induction Algorithm" (FURIA) has been proposed [54]. This method is an enhancement to the famous RIPPER algorithm. The empirical evaluation revealed that FURIA beats RIPPER and Decision Tree most of the time. An improved Bayesian algorithm for detecting spam email messages was offered in 2009 [55]. Through this research, boosting methods had been employed for making the Bayesian algorithm more powerful.

2.5 Viable techniques for creating lifelong spam classification models

Confronted with the ever-streaming datasets, the direction of lifelong DM methods *(also known as adaptable classification methods)* appear to be the best alternative to expose the new knowledge that could be contained in these datasets. Such knowledge if not learned might negatively change the overall performance of the classification model(s). Regularly updating the classification model as soon as a new dataset obtained ensures a long life for the classification models. All lifelong classification models are susceptible to progressive improvements in their structure(s). Usually, creating a lifelong classification models may be accomplished by making sure that it is capable to learn incrementally [56].

Generally speaking, any lifelong classification model needs to fulfill the following conditions:

1- Acquire new knowledge from any dataset that might be received at any time.

2- Acquiring new knowledge will not result in losing the formerly learned ones.

3- Acquiring new knowledge doesn't require using earlier dataset.

Incremental learning is usually a response to the existence of the concept drift. Incremental learning might be achieved in various methods, including online vs batch strategies dependent on the number of examples utilized at every training stage; or single-based vs ensemble-based approach dependent on how many classification model utilized to produce the final outcome. Many strategies were proposed in literature that facilitate producing lifelong classification systems that are characterized of being capable to learn new knowledge without losing the previously learned ones. Hereunder, we review these strategies:

- The easiest strategy is referred to as interleaved learning [57] in which the new dataset will be merged with all previously collected datasets. After that, we can choose to retrain the current model without changing its structure, or even remove the old model and create a new one from the beginning. By doing this, the interleaved learning strategy satisfies the first two previously mentioned conditions, however the third condition is obviously violated. In addition, this strategy could be memory consuming as well as computationally costly process [58]. Various other methods do the job by minimizing the dimensions of the entire dataset by using a subset from the new instances instead of utilizing the full dataset [59]. Nevertheless, this method often demands access to the previous datasets(s).

- Various strategies apply some sort of sliding windows on the newly arriving dataset instances. This method keeps only the most up-to-date instances in the FIFO ("first-in-first-out") data structure. The moment new set of examples are provided; they will be loaded at the start of the window. A similar number of instances will be deleted from the end of the window. This approach assumes that the knowledge obtained from the old examples are practically useless for classifying new instances and so these examples are removed from the dataset. The set of instances which fall within the window are thought new training dataset, and new a model is generated using these instances. Using this approach, the classification model might be created using the dataset instances inside fixed or dynamic sliding windows. The fixed sliding window approaches keep the newest x examples, whereby the model creator usually defines the size of window before building the model. On the other hand, in the dynamic sliding windows, the size of the window is autonomously modified with time [60]. The main concern in the dynamic sliding windows strategy is to choose a suitable size for the window. Small windows reflect the present distribution more accurately yet that could affect the classification model overall performance on the older instances [60] since the old cases could be removed from the window because there isn't space to maintain all of them inside the window. In contrast, a large window may maintain the model overall performance on the old cases, but it responds slowly to the concept drift. Several windows based algorithms devised in literature, among others we name: ADWIN [61], TWF [62] FLORA [18], and FRANN [19]. Nevertheless, the main issue with sliding windows-based approach is that they might result in the what's commonly called catastrophic forgetting [20], thus, violating the second condition of the lifelong classification models definition. Besides, this method might not succeed when the concept drift within the problem is a cyclical one. Therefore, although this approach might be useful for creating lifelong classification models it might not be the right choice for dealing with the cyclical concept drift the normally occurs in spam email classification problem.

- The utilization of ensemble strategies is actually preferred in lifelong learning cases due to some extent to their empirical performance [63,56,64–67]. Such technique combines a group of models whose individual prediction outputs will be combined collectively in certain way to facilitate making an enhanced composite classification system with superb overall classification performance. Normally, this approach discovers knowledge incrementally, doesn't forget the formerly learned knowledge and doesn't demand usage of the formerly used training datasets [64]. Ensemble based approaches happen to be the most frequently applied method for domains in which a cyclical concept drift might appear [68]. Ensemble based technique typically consists of classification models produced from previous dataset; these models might be reused for classifying newly arrived instances if they happen to be drawn from a reoccurring concept. Among others, the most commonly used ensemble algorithms we name: *"Accuracy Weighted Ensemble"* [65] (AWE), *"Multi-Partition Multi-Chunk Ensemble"* [69] (MPMC), and *"Accuracy Updated Ensemble"* [66] (AUE). The AWE uses weighted ensemble classification models to tackle concept drift. Each classifier in the ensemble is sensibly weighted according to its estimated classification accuracy on the recent testing dataset. Alike AWE, the MPC method creates an ensemble that contains a set of the most effective classifiers. Such an ensemble will then be modified from classifiers which were trained using the latest dataset. In contrast to AWE, the MPC approach splits the dataset into v even partitions which after that can be utilized to create a new group of possible classifiers. This method works as follows: Let *Ds* denotes the very last dataset. Calculate the error of every classifier on *Ds*. Let $D = Ds - m + 1, \cdots, Ds$, or m of the newest datasets. Split D arbitrarily into × equivalent partitions: $d1, \cdots, dx$ and train a new group of possible classifiers through going over the divisions. After that, find the related error of the classifiers. Finally, modify the ensemble by keeping the most effective Y classifiers. AUE creation is motivated by AWE and the weighting process it uses, yet, AUE enhances a number of defects that already exist in the AWE. For instance, AUE improves AWE by utilizing online learning based classifiers and adjusting these classifiers depending on recent distribution. Updating the classifiers according to the current distribution is considered the major improvement the AUE made over the AWE. One extra modification the AUE made over the AWE is the weighting process that resolves issues with unwanted classifier eliminating employed in AWE. Another difference among AWE and AUE lies in the fact that AWE was created for batch-based methods but the AUE was designed to work for online and incremental learners.

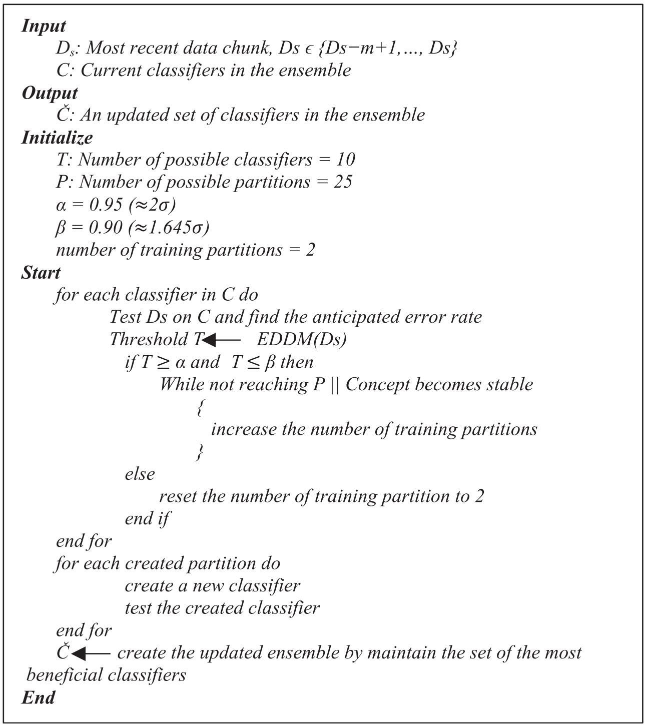
In the current study, the ensemble-based learning technique is utilized for producing a lifelong classification model. Such a model will be then applied to spam emails classification problem.

3. Ensemble based lifelong classification using adjustable dataset partitioning

This section intends to discuss the proposed *"Ensemble based Lifelong Classification using Adjustable Dataset Partitioning"* (ELCADP) (depicted in Figure 1), that aims to create robust lifelong classification models.

Such a method is talented enough to handle the concept drift and the catastrophic forgetting dilemmas that are considered the main challenges when creating lifelong classification models. Several domains might benefit from the ELCADP such as phishing websites classification [6,21], fraud classification, and spam emails classification. An important issue in order to mitigating concept drift is typically knowing the time, or even whether a concept drift has actually appeared. The ELCADP makes use of the information coming from some drift detection methods that confirm if a concept drift is coming up and after that, the ELCADP attempts to handle it in order to accommodate any change in the class distribution. A widely used concept drift recognition technique is the *"Early Drift Detection Method"* (EDDM) [70]. Such a technique utilizes a binomial distribution in order to compute the average distance amongst 2 successive classification error (*Ei*) and the standard deviation (*Si*) to figure out 2 threshold values those are:

- Warning Level α: Beyond this threshold, start gathering instances due to expectation of concept drift.



```
Input
    $D_s$: Most recent data chunk, Ds ϵ {Ds−m+1,…, Ds}
    C: Current classifiers in the ensemble
Output
    Č: An updated set of classifiers in the ensemble
Initialize
    T: Number of possible classifiers = 10
    P: Number of possible partitions = 25
    $\alpha = 0.95\ (\approx 2\sigma)$
    $\beta = 0.90\ (\approx 1.645\sigma)$
    number of training partitions = 2
Start
    for each classifier in C do
            Test Ds on C and find the anticipated error rate
            Threshold T ⟵ EDDM(Ds)
              if $T \geq \alpha$ and $T \leq \beta$ then
                  While not reaching P || Concept becomes stable
                        {
                          increase the number of training partitions
                        }
              else
                  reset the number of training partition to 2
              end if
    end for
    for each created partition do
              create a new classifier
              test the created classifier
    end for
    Č ⟵ create the updated ensemble by maintain the set of the most
 beneficial classifiers
End
```

Figure 1.
ELCADP pseudo code.

- Drift Level β: Beyond such threshold, concept drift is confirmed. A new classifier should be created using the gathered instances.

The EDDM after that benefits from such warning thresholds to decide when to start collecting new instances due to expectation of concept drift, and as soon as the concept drift is confirmed to be occurred, it modifies the classification model so as to deal with the concept drift. The ELCADP makes use of the MPC technique by adapting how many partitions to utilize within the ensemble but differs from MPC in using EDDM in order to recognize concept drift. This research study makes use of the EDDM capabilities to identify concept drift levels and employ these thresholds to consequently modify how many of partition(s) is required to create a new group of classification models. In [70], α and β were set to 0.95($\approx 2\sigma$), and 0.90($\approx 1.645\sigma$) respectively. The current research follows [70] in setting the value of α and β. Just like MPC, ELCADP preserves an ensemble of the top T classifiers. For each arriving data chunk, the ELCADP tests it and finds out its anticipated error. In contrast to MPC, the ELCADP after that finds out the drift threshold by using EDDM method over the arriving data chunk.

If the threshold is above α for a specific data chunk, but still below β, the data is thought to be drifting and the ELCADP increases the number of the training partition. However, as soon as the concept within the data stream becomes stable the ELCADP resets the size of the partition to the default value i.e. 2 to imitate MPC whilst not in concept drift detection. Actually, the proposed partitioning technique is inspired from the *"fuzzy c-mean"* technique [71]. To elaborate further, the ELCADP firstly assumes the stability of the concept distribution within the data streams. Hence, the ELCADP starts with 2 partitions and keeps splitting the data stream until the concept distribution within the data stream becomes stable.

At this point, the size of the partition is reset to 2 (“the default value”). The quality of the partition is assessed and the worst performing partition is divided into 2 different groups. The process is redone till the stopping criteria happen to be fulfilled. Considering that the number of partitions will not be known, following previous studies [71] and [72] the maximum number of partitions is fixed to 25. The partition process is stopped early whenever the partition quality is less than five times the top partition quality experienced. Such a number was empirically found to enhance the partitioning process because a bad data split might not be improved by consecutive partitions. In general, the ELCADP determines the number of partitions in accordance with the amount of drift experienced in the data chunk partition. Just like MPC, the ELCADP then trains T classifiers using the elected partitions. Each classifier is trained and tested independently. As soon as the training process is completed for all possible classifiers, each of which is tested and the expected error is calculated. After that, the set of the best performing classifiers are used for building the ensemble and the remaining classifiers are removed. In general, the major matter to notice is that the ELCADP dynamically tunes the number of partitions because of concept drift occurrence. It should be noted that the number of possible classification models is increased only during the class distribution instability. As a rule of thumb, if a classifier provides an error rate equals 50% that implies that it is randomly predicting the value of the class variable [73]. Nevertheless, when a classifier gives an error rate greater than 50% then the overall performance of the ensemble might be damaged [67,65,74]. In this study, when a classifier yields an error rate greater than or equals to 50%, it will be removed and prohibited from taking part in predicting the class values. Moreover, in case the number of classifiers is greater than a specific threshold, the least beneficial classifier will be deleted. However, there is no general guideline in literature regarding the maximum number of possible classifiers [21]. In this research, following previous studies [75,76,21] the ELCADP sets the maximum number of classification models within the ensemble to 10. In general, the ELCADP creates lifelong classification models using ensemble based approach. The algorithm firstly nominates the set of the most representative partitions of the current concept distribution from the whole dataset that is collected as soon as the concept drift is confirmed to be occurred. After that, the ELCADP creates a set of possible classifiers using the previously nominated datasets chunks and maintains the most effective ones, whereas the others will be deleted. The ELCADP has been implemented using WEKA [77] which is a well-known open source tool for implementing several DM and ML algorithms.

4. Evaluating the ELCADP

This section aims to assess the overall performance of the using several evaluation metrics ECLADP and contrast it to a number of other approaches. The obtained results will also be discussed thoroughly in this section.

4.1 Experimental setting

The performance of the ECLADP is assessed against 3 different stream mining approaches those are *“Simple Online Classifier” (SOC), “Batch Learner” (BL), and “Sliding Window” (SW).* The selection of these approaches is because they are commonly used for mining data streams. In addition, these methods utilize different techniques in selecting the training dataset and generating the classification model(s). Further, these methods have proved their capabilities in creating lifelong classification models due to their unique abilities in handling the concept drift and the catastrophic forgetting dilemmas. Moreover, these schemas cover the 3 possible approaches that are commonly used for creating lifelong classification models as discussed in Section. In the SOC, the classification model is updated whenever a new training dataset instance becomes available. In other words, the SOC firstly examines a new emerging example and then the classification model will be updated according to the new knowledge learnt from such an example. In contrast, the BL schemas retrain the whole classification model once a new dataset batch is obtained. Here lies the main difference between SOC and BL in the since that the former one updates the classification model as soon as a single instance becomes available, whereas the later one waits until collecting a specific number of examples before updating the classification model. Alike the BL, the SW waits until a set of examples are collected which will be then used for learning new knowledge so as to update the current classification model. Yet, the SW employ a simple forgetting procedure by assuming that the old examples become useless over time and then they should be replaced with new examples. All these techniques are implemented on WEKA [77]. All methods, apart from BL utilize an updated copy of the "Naive Bayes" algorithm as a base learner. This algorithm is selected due to the fact it is inherently learn incrementally. In addition, it has been frequently employed in creating classification models and particularly in classifying spam emails in particular [33,42,44] which is the domain under investigation through this article. Considering that BL may work with no incremental algorithm we also used "Support Vector Machine" (SVM) which has also been employed in predicting spam emails as well as several other domains [78,79,81]. We will use the default parameter settings of WEKA when creating the models. BL was carried out using 3 different batch sizes (100, 200, and 300). Moreover, the SW was created using 3 different window sizes (100, 200, and 300).

4.2 Evaluation metrics

Following earlier researches relevant to spam emails classification [37,38,50,35,79] we will calculate several evaluation metrics those are as follows:

1- Precision *(P)*: This can be calculated as per Eq. (1) and it denotes the percentage of accurately classified non-spam emails in relative to all cases which are classified as non-spam.

$$P = \frac{TP}{TP + FP} \tag{1}$$

2- Recall (R): The accurately labeled spam email messages in relation to the whole number of instances labeled as such.

$$R = \frac{TP}{TP + FN} \tag{2}$$

3- Accuracy (ACC): This can be calculated as per Eq. (2) and it denotes the percentage of accurately labeled instances with regards to the total numbers of instances.

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

4- Harmonic Mean (F1-Score): This can be calculated as per Eq. (4) and it weights the average of P and R. A model that produces good results in both P and R is preferred over a model that produces excellent results on one and bad results on the other.

$$F1 = \frac{2PR}{P + R} \tag{4}$$

All of the experiments are carried out in a computer system that has CPU Pentium Intel® CoreTM i7-8th Gen, RAM 8.00 GB. The operating system was Windows 7 64-bit.

4.3 Training dataset

The well-known "Enron-Spam" dataset [80] is utilized for evaluating the performance of ELCADP. Enron-Spam dataset contains email messages sent mainly by the senior persons in the "Enron Company". Such a dataset consists of 6 personal mailboxes which had been published following the Enron scandal. A processed version of the dataset mainly designed for "Spam" and "ham" emails classification is produced and used in our study. Such dataset involves 33,716 email messages, of which 16,545 email messages marked as "ham" and 17,171 email messages are marked as "spam". What makes this dataset suitable for doing our experiments is that it is time stamped and it is relatively big with compare to other datasets. All mails were text files written in English Language which include two main parts those are the header (*"which contains metadata about the email"*) and the email body. The set of input features were extracted from these two sources. It is worth mentioning that using all the input features in the training dataset might results in what is normally called "The Curse of Dimensionality" [15]. Therefore, with the aim of reducing the processing time as well as the storage requirements, "Information Gain" is used for selecting the most important set of features from the dataset. Information Gain is considered the most commonly used feature selection approach in spam classification because of its capacity to obtain a minimized number of features that has a lower impact in the overall performance of the classification model [21]. Additionally, several works revealed good results when incorporating Information Gain with DM and ML algorithms [15].

4.4 Results and discussion

The obtained results are depicted in Table 1.

From Table 1, we can obviously comprehend that the ELCADP outperforms all other contrasted stream mining algorithms in terms of Acc, P, R, and F1 at 95.80%, 94.40%, 95.80%, and 95.10% respectively. Such results confirm that the ELCADP is capable to adequately select the set of partitions from the original dataset that can be used for creating new classifier that are devised whenever a warning message is received about a possible occurrence of concept drift. SW produced the second-best classification accuracy at 92.00% when the batch was set to 200. One reason behind achieving such classification accuracy from the SW is that it typically uses the latest set of examples (window) for updating the classification model and therefore, despite the fact that SW do not incorporate any concept drift recognition mechanism, it still capable to cope with concept drift promptly. BL is not doing as good as SW and that could be because the classification model is updated whenever X set of instances are obtained regardless of whether such X examples represent the current concept distribution or not. The best classification accuracy obtained from BL was when

Method	Base Classifier	Acc%	P%	R%	F1%
SOC	NB	82.07%	85.80%	82.40%	82.60%
BL (100)	NB	86.09%	87.80%	86.70%	86.90%
BL (200)	NB	86.09%	87.80%	86.70%	86.90%
BL (300)	NB	85.52%	87.60%	86.30%	86.50%
BL (100)	SVM	89.98%	90.90%	90.70%	90.70%
BL (200)	SVM	87.24%	87.50%	87.20%	87.00%
BL (300)	SVM	87.22%	88.40%	87.90%	88.00%
SW (100)	NB	92.00%	90.20%	92.00%	92.00%
SW (200)	NB	89.81%	89.80%	89.80%	89.70%
SW (300)	NB	90.80%	90.80%	90.80%	90.70%
ELCADP	NB	95.80%	94.40%	95.80%	95.10%

Table 1.
Obtained results from ELCADP and other Techniques.

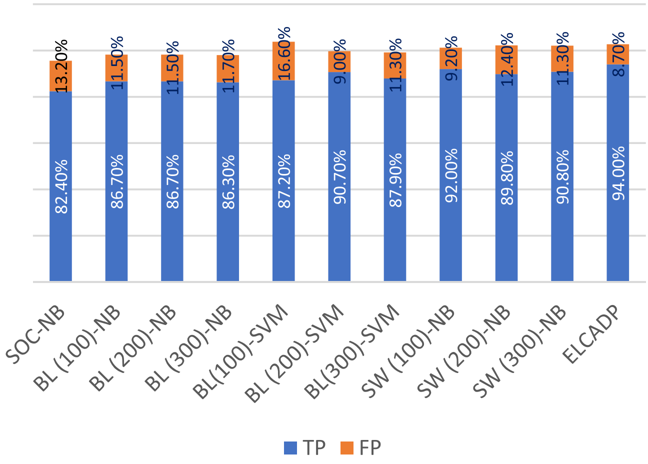


Figure 2. Shows a compromise between the ELCADP and the considered stream mining algorithms in terms of TP and FP.

using the SVM as a base classifier and the window size is set to 300 at 89.98%. To elaborate further, the concept distributions might overlap in the training dataset that is used in creating the classification model in the BL approach. SOC produced the worst classification accuracy among others. Nevertheless, the classification accuracy of the SOC is considered an achievement. Yet, the empirical experiments confirm that the online classification schema might not be the appropriate choice when creating a lifelong spam classification systems. The experimental results illustrated in Table 1 showed that the highest accuracy rate has been achieved when the size of the batch is set to 100 regardless of the utilized base classifiers. This size (i.e. 100) has also produced the highest classification accuracy in SW based classification models. One interesting observation from the obtained results comes from the SOC technique. The best F1-score obtained from ELCADP at 95.10%. Overall, the ELCADP was able to produce a lifelong spam classification model that is capable to handle the 2 major concerns of any classification model, i.e. concept drift and catastrophic forgetting. Furthermore, the ensemble-based classification approach adds the spam emails to its long list of domains where it achieved superb classification performance.

The results depicted in Figure 2 show that the ELCADP produced the highest TP rate at 94.00%. ELCADP also produced the lowest FP rate at 8.70%. Interestingly, the SOC produced the lowest TP rate and at the same time it produced the highest FP rate. These results confirm that the ensemble-based spam classification approach is more suitable for identifying spam emails. That can be attributed due to the fact the ensemble-based approach is capable to handle the concept drift issue and hence they can produce a lifelong spam classification system.

5. Conclusion

In this article, a lifelong spam emails classification model was created. Such a model is called *"Ensemble based Lifelong Classification using Adjustable Dataset Partitioning"*. The lifelong property of the model was achieved by ensuring that the model is capable to handle the concept drift and the catastrophic forgetting dilemmas that are usually presumed that main challenges when creating any DM and ML system. The ELCADP make use of the information obtained from the EDDM which is a well-known concept detection method. The information obtained from the EDDM provides information about the concept drift level. According to the concept drift level, the ELCADP is then decides on the number of partitions needed to for creating a new classification model. For the evaluation purposes, the well-known "Enron-Spam" dataset is used. The empirical evaluation showed that the ELCADP surpasses all the stream mining methods that are used for the comparison purposes. Four evaluation metrics were used for assessing the performance of the ELCADP those are: *Accuracy, Precision, Recall, and F1-Score.* The ELCADP showed sound results in all these evaluation metrics and that confirm that the ELCADP was able to create lifelong spam classification model. Overall, this research study showed that the traditional offline spam emails classification systems might not be the right choice for creating lifelong classification systems. However, it should be mentioned here that the performance ELCADP has not been examined in the case that the concept drift that might occur is a virtual concept drift where a new class value might appear while the input features are still unchanged. This in fact is left as a future work and the phishing websites classification is a possible domain for examining the ability of the ELCADP in handling the virtual concept drift the characterizes the phishing websites classification problem.

References

[1] S. Deffree, "1st spam email is sent, May 3, 1978", EDN, 3 May 2019. [Online]. Available: https://www.edn.com/1st-spam-email-is-sent-may-3-1978/. (accessed 1. 4. 2020).

[2] H. Tschabitscher, "Worldwide Email Statistics," Lifewire, 23 March 2018. [Online]. Available: https://www.lifewire.com/how-many-email-users-are-there-1171213. (accessed 1 May 2018).

[3] R.M. Mohammad, F. Thabtah, L. McCluskey, Tutorial and critical analysis of phishing websites methods, Comput. Sci. Rev. 17 (1) (2015) 1–24.

[4] H. Tschabitscher, "Fascinating email facts," Lifewire, 23 March 2018. [Online]. Available: https://www.lifewire.com/how-many-emails-are-sent-every-day-1171210. (accessed 1 May 2018 A).

[5] R.M. Mohammad, F. Thabtah, L. McCluskey, Intelligent rule based phishing websites classification, IET Inf. Secur. 8 (3) (2013) 153–160.

[6] R.M. Mohammad, F. Thabtah, L. McCluskey, Predicting phishing websites based on self-structuring neural network, Neural Comput. Appl. 25 (2) (2013) 443–458.

[7] N. Alfred, "Watch out for Hurricane Harvey phishing scams," CBS NEWS, 30 8 2017. [Online]. Available: https://www.cbsnews.com/news/hurricane-harvey-phishing-scams-cybercriminals/. (accessed 2. 5. 2018).

[8] F. Allego, "Threat Encyclopedia," TREND MICRO, 5 5 2015. [Online]. Available: https://www.trendmicro.com/vinfo/us/threat-encyclopedia/spam/608/spammers-use-earthquake-in-nepal-for-scam-donation-funds. (accessed 2. 5. 2018).

[9] A. Iovine, "Email scam claims that Melania Trump wants to give you 20 million dollars," AOL, 29 3 2017. [Online]. Available: https://www.aol.com/article/news/2017/03/29/email-scam-claims-that-melania-trump-wants-to-give-you-20-millio/22017185/. (accessed 6. 5. 2018).

[10] E. Bauer, "15 Outrageous Email Spam Statistics that Still Ring True in 2018," Propeller, 1 2 2018. [Online]. Available: https://www.propellercrm.com/blog/email-spam-statistics. (accessed 6. 5. 2018).

[11] G. Darya, V. Maria, S. Tatyana, "Spam and phishing in 2017," SecureList, 15 2 2018. [Online]. Available: https://securelist.com/spam-and-phishing-in-2017/83833/. (accessed 6. 5. 2018).

[12] S. Statistics, "Spam Statistics," Spam Statistics, 2018. [Online]. Available: https://antispamengine.com/spam-statistics/. (accessed 6. 5. 2018).

[13] R.M. Mohammad, F. Thabtah, L. McCluskey, An assessment of features related to phishing websites using an automated technique. International Conference for Internet Technology and Secured Transactions, 2012, London, 2012.

[14] R.M. Mohammad, F. Thabtah, L. McCluskey, An improved self-structuring neural network. Pacific Asia Knowledge Discovery and Data Mining Conference (PAKDD) 2016, Auckland, 2016.

[15] R.M. Mohammad, M. Alqahtani, A comparison of machine learning techniques for file system forensics analysis, J. Inf. Security Appl. 46 (1) (2019) 53–61.

[16] I.H. Witten, E. Frank, A.H. Mark, Data mining: practical machine learning tools and techniques with Java implementations, third ed., Morgan Kaufmann, 2011.

[17] R. Tadeusiewicz, Neural networks in mining sciences – general overview and some representative examples, Bull. Polish Acad. Sci.: Tech. Sci. 60 (4) (2015) 971–984.

[18] G. Widmer, M. Kubat, Learning in the presence of concept drift and hidden contexts, in: Machine Learning, Springer, 1996, pp. 69–101.

[19] M. Kubat, G. Widmer, Adapting to drift in continuous domains (Extended abstract). 8th European Conference on Machine Learning Heraclion, Crete, Greece, 1995.

[20] F.H. Hamker, Life-long learning cell structures–continuously learning without catastrophic interference, Neural Networks 14 (4–5) (2001) 551–573.

[21] R. Mohammad, An Ensemble Self-Structuring Neural Network Approach to Solving Classification Problems with Virtual Concept Drift and its Application to Phishing Websites, University of Huddersfield, Huddersfield, 2016.

[22] D. Ruano-Ordás, F. Fdez-Riverola, J.R. Méndez, Using evolutionary computation for discovering spam patterns from e-mail samples, Inf. Process. Manage. 54 (1) (2018) 303–317.

[23] Qualaroo, "Qualaroo," 2012. [Online]. Available: https://help.qualaroo.com/hc/en-us/sections/200344577-How-to-write-Regular-Expressions-for-URL-targeting, (accessed 22 December 2019).

[24] V.V. Prakash, A. O'Donnell, Fighting spam with reputation systems, in: Queue – Social Computing, 2005, p. 50.

[25] E. Zheleva, A. Kolcz, L. Getoor, Trusting spam reporters: A reporter-based reputation system for email filtering, ACM Trans. Inf. Syst. (TOIS) 27 (1) (2008).

[26] R.M. Mohammad, F. Thabtah, L. McCluskey, Predicting Phishing Websites using Neural Network trained with Back-Propagation, ICAI Las Vigas, 2013.

[27] M. D'Auro, I. de Achaval, "Data protection in Argentina: overview," 2014. [Online]. Available: http://www.ebv.com.ar/images/publicaciones/trdatap.pdf, (accessed 5. 11. 2018).

[28] X. Liu, P. Zou, W. Zhang, J. Zhou, C. Dai, F. Wang, X. Zhang, CPSFS: A credible personalized spam filtering scheme by crowdsourcing, Wireless Commun. Mobile Comput. (2017).

[29] Spamhaus, "Spamhaus," Spamhaus, 1998. [Online]. Available: https://www.spamhaus.org/, (accessed 12. 5. 2018).

[30] G. Bechis, K. Bräckelmann, A. Broens, B. Cole, J. Hardin, D. Jones, A. Katz, H. Krohns, S. Markowitz, M. Martinec, K. A. McGrail, M. Parker, J. Quinn, "Apache SpamAssassin," Apache SpamAssassin Project, 2003. [Online]. Available: https://spamassassin.apache.org/, (accessed 20 December 2019).

[31] McAfee, "SpamKiller," McAfee, 2005. [Online]. Available: https://download.mcafee.com/products/manuals/en-us/MSK_UserGuide_2006.pdf, (accessed 20 December 2019).

[32] Symantec, "Symantec Brightmail," Symantec Brightmail, 2007. [Online]. Available: https://www.symantec-norton.com/symantec_brightmail_p26.aspx, (accessed 20 December 2019).

[33] P. Graham, Better Bayesian filtering, Spam Conference, 2003.

[34] G.-C. J, "People and spam," in: The Spam Conference, 2005.

[35] Q. Ma, Z. Qin, F. Zhang, Q. Liu, Text spam neural network classification algorithm, International Conference on Communications, Circuits and Systems (ICCCAS), 2010, Chengdu, China, 2010.

[36] R. Shams, R.E. Mercer, Classifying spam emails using text and readability features, IEEE 13th International Conference on Data Mining (ICDM), 2013, Dallas, TX, USA, 2013.

[37] R.M. Aliguliyev, R.M. Aliguliyev, S. Nazirova, Classification of textual e-mail spam using data mining techniques, Appl. Comput. Intell. Soft Comput. (2011) 8.

[38] G.V. Cormack, M.D. Smucker, Efficient and effective spam filtering and re-ranking for large web datasets, Inf. Retrieval 14 (5) (2011) 441–465.

[39] S. Singh, A. Chand, S.L. Pranit, Improving Spam Detection Using Neural Networks Trained by Memetic Algorithm, Fifth International Conference on Computational Intelligence, Modelling and Simulation (CIMSim), 2013, Seoul, South Korea, 2013.

[40] S.O. Olatunji, Improved email spam detection model based on support vector machines, Neural Comput. Appl. (2017) 1–9.

[41] E.G. Dada, J.S. Bassi, H. Chiroma, S.M. Abdulhamid, A.O. Adetunmbi, O.E. Ajibuwa, Machine learning for email spam filtering: review, approaches and open research problems, Heliyon, 1(5), pp. 1–23, 2019.

[42] Y. Song, A. Kolcz, C. Lee Giles, Better naive bayes classification for high-precision spam detection, J. Software: Pract. Exp. (SPE) 39 (11) (2009) 1003–1024.

[43] Z. Yang, L. HongYu, M. Niranjan, P. Rockett, Applying cost-sensitive multiobjective genetic programming to feature extraction for spam e-mail filtering, European Conference on Genetic Programming, 2008.

[44] L. Özgür, T. Güngör, F. Gürgen, Spam mail detection using artificial neural network and bayesian filter, International Conference on Intelligent Data Engineering and Automated Learning, 2004.

[45] A.S. Aski, N.K. Sourati, Proposed efficient algorithm to filter spam using machine learning techniques, Pac. Sci. Rev. A: Nat. Sci. Eng 18 (2) (2019) 145–149.

[46] R. Ariaeinejad, A. Sadeghian, Spam detection system: A new approach based on interval type-2 fuzzy sets, Canadian Conference on Electrical and Computer Engineering (CCECE), 2011 24th, Niagara Falls, ON, Canada, 2011.

[47] R. Segal, J. Crawford, J. Kephart, B. Leiba, SpamGuru: An enterprise anti-spam filtering system, The First Conference on E-mail and Anti-Spam (CEAS 2004), Mountain View, CA, 2004.

[48] F. Qian, A. Pathak, Y. Charlie Hu, Z. Morley Mao, Y. Xie, A case for unsupervised-learning-based spam filtering, The ACM SIGMETRICS international conference on Measurement and modeling of computer systems, NEW YORK, 2010.

[49] S. Youn, D. McLeod, A comparative study for email classification, Advances and Innovations in Systems, Computing Sciences and Software Engineering, Dordrecht, 2007.

[50] A. El-Halees, Filtering spam e-mail from mixed arabic and english messages: a comparison of machine learning techniques, Int. Arab J. Inf. Technol. (IAJIT) 6 (1) (2009) 52–59.

[51] A. Çıltık, T. Güngör, Time-efficient spam e-mail filtering using n-gram models, Pattern Recogn. Lett. 29 (1) (2007) 19–33.

[52] D. Kalbande, H. Panchal, N. Swaminathan, P. Ramaraj, ANFIS based Spam filtering model for Social Networking Websites, Int. J. Comput. Appl. 44 (11) (2012) 0975–8887.

[53] D. Ruano-Ordás, F. Fdez-Riverola, J.R. Méndez, Concept drift in e-mail datasets: An empirical study with practical implications, Inf. Sci. 428 (1) (2018) 120–135.

[54] J. Huhn, E. Hullermeier, FURIA: an algorithm for unordered fuzzy rule induction, Data Min. Knowl. Disc. 19 (3) (2009) 293–319.

[55] P.-Y. Liu, L.-W. Zhang, Z.-F. Zhu, Research on e-mail filtering based on improved Bayesian, J. Comput. 4 (3) (2009) 271–275.

[56] D.M. Farid, Z. Li, A. Hossain, C.M. Rahman, R. Strachan, G. Sexton, K. Dahal, An adaptive ensemble classifier for mining concept drifting data streams, Expert Syst. Appl. 40 (15) (2013) 5895–5906.

[57] T. Seipone, J.A. Bullinaria, Evolving improved incremental learning schemes for neural network systems, The 2005 IEEE Congress on Evolutionary Computation, 2005., Edinburgh, Scotland, UK, 2005.

[58] D. Parikh, R. Polikar, An ensemble-based incremental learning approach to data fusion, IEEE Trans. Syst. Man Cybernet. 37 (2) (2007) 437–450.

[59] A.P. Engelbrecht, R. Brits, A clustering approach to incremental learning for feedforward neural networks, International Joint Conference on Neural Networks, 2001, Proceedings. IJCNN'01, Washington, DC, 2001.

[60] J. Gama, I. Žliobaitė, A. Bifet, M. Pechenizkiy, A. Bouchachia, A survey on concept drift adaptation, ACM Comput. Surveys (CSUR) 46 (4) (2014) pp.

[61] A. Bifet, R. Gavaldà, Kalman filters and adaptive windows for learning in data streams, 9th International Conference, DS 2006, Barcelona, Spain, 2006.

[62] M. Salganicoff, Tolerating concept and sampling shift in lazy learning using prediction error context switching, Artif. Intell. Rev. 11 (1–5) (1997) 133–155.

[63] A. Wiliński, S. Osowski, Ensemble of data mining methods for gene ranking, Bull. Polish Acad. Sci. Tech. Sci. 60 (3) (2012) 461–470.

[64] R. Polikar, L. Upda, S.S. Upda, V. Honavar, Learn++: an incremental learning algorithm for supervised neural networks, systems, man, and cybernetics, Part C: applications and reviews, IEEE Trans. 31 (4) (2001) 497–508.

[65] H. Wang, W. Fan, P.S. Yu, J. Han, Mining concept-drifting data streams using ensemble classifiers, The ninth ACM SIGKDD international conference on Knowledge discovery and data mining, Washington, D.C., 2003.

[66] D. Brzezinski, J. Stefanowski, Accuracy updated ensemble for data streams with concept drift, International Conference on Hybrid Artificial Intelligence Systems, Wroclaw, Poland, 2011.

[67] T.G. Dietterich, Ensemble methods in machine learning, The First International Workshop on Multiple Classifier Systems, London, 2000.

[68] R.T. Hoens, R. Polikar, N.V. Chawla, Learning from streaming data with concept drift and imbalance: an overview, Prog. Artificial Intell. 1 (1) (April 2012) 89–101.

[69] M.M. Masud, J. Gao, L. Khan, J. Han, B. Thuraisingham, A multi-partition multi-chunk ensemble technique to classify concept-drifting data streams, Pacific-Asia Conference on Knowledge Discovery and Data Mining, Bangkok, Thailand, 2009.

[70] M. Baena-Garca, J. Del Campo- Ávila, R. Fidalgo, A. Bifet, R. Gavaldà, R. Morales-bueno, Early drift detection method, Fourth International Workshop on Knowledge Discovery from Data Streams, 2006.

[71] J.C. Bezdek, P.R. Mikhil, J. Keller, R. Krisnapuram, Fuzzy Models and Algorithms for Pattern Recognition and Image Processing, Kluwer Academic, Norwell, MA, USA, 1999.

[72] N. Chawla, S. Eschrich, L.O. Hall, Creating ensembles of classifiers, First IEEE International Conference on Data Mining, 2000.

[73] T. Fawcett, An introduction to ROC analysis, Pattern Recogn. Lett. 27 (8) (2006) 861–874.

[74] L. Rokach, Ensemble-based classifiers, Artif. Intell. Rev. 33 (1) (2010) 1–39.

[75] N.W. Street, Y. Kim, A streaming ensemble Algorithm (SEA) for large-scale classification, The seventh ACM SIGKDD international conference on Knowledge discovery and data mining, New York, NY, USA, 2001.

[76] Z.J. Kolter, M.A. Maloof, Dynamic weighted majority: an ensemble method for drifting concepts, J. Mach. Learn. Res. 8 (12) (2007) 2755–2790.

[77] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, "Waikato Environment for Knowledge Analysis," University of Waikato, 2011. [Online]. Available: http://www.cs.waikato.ac.nz/ml/weka/. (accessed 20 December 2011).

[78] R.M. Mohammad, A neural network based digital forensics classification, 2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA), Aqaba, 2018.

[79] R.M. Mohammad, H.Y. AbuMansour, An intelligent model for trustworthiness evaluation in semantic web applications, 8th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 2017.

[80] V. Metsis, I. Androutsopoulos, G. Paliouras, “Enron-Spam datasets,” Enron-Spam datasets, 19 June 2006. [Online]. Available: http://nlp.cs.aueb.gr/software_and_datasets/Enron-Spam/index.html, (accessed 26 July 2015).

[81] R. Mohammad M., An Enhanced Multiclass Support Vector Machine Model and its Application to Classifying File Systems Affected by a Digital Crime, J. King Saud University - Comput. Inform. Sci. (2019), http://dx.doi.org/10.1016/j.jksuci.2019.10.010, In press.

[82] A. Gonsalves, H.F. Thabtah, R. Mohammad Mustafa, G. Singh, Prediction of Coronary Heart Disease using Machine Learning: An Experimental Analysis. ICDLT 2019: Proceedings of the 2019 3rd International Conference on Deep Learning Technologies, ACM, 2019, pp. 51–56.

Corresponding author
Rami Mustafa A. Mohammad can be contacted at: rmmohammad@iau.edu.sa