Materials Today: Proceedings xxx (xxxx) xxx

Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

An effective approach to protect social media account from spam mail – A machine learning approach

Vishnu Dutt Sharma [a,*], Santosh Kumar Yadav [a], Sumit Kumar Yadav [b], Kamakhya Narain Singh [c,*], Suraj Sharma [d]

[a] Department of Computer Science, JJT University, Rajasthan, Jaypur 333001, India
[b] Department of Computer Science, IGDTUW, Kashmere Gate, Delhi 110006, India
[c] School of Computer Applications , KIIT Deemed University, Bhubaneswar, Odisha 751024, India
[d] International Institute of Information Technology, Khurda, Odisha 751003, India

ARTICLE INFO

Article history:
Received 28 November 2020
Accepted 9 December 2020
Available online xxxx

Keywords:
Spam classifier
Decision tree
KNN classifier
Feature selection
Feature standardization
Social media profile

ABSTRACT

The internet users in the world are rising rapidly. This technology given the opportunities to grow in the field of business, education, sports, entertainment and social media. With a huge amount of person getting connected to Internet, the security threats which leads massive harms are increasing also. In the last few years, we have been observing a quick development in information being produced and shared in social media. The usage of social media increased exponentially, and rapid growth of users are unexpected in the history of technology insurgency. The uninterrupted growth in social media and connected technologies has managed to an ecosystem where users with different culture and geolocation are connected on various social media platforms. While the explosive growth of internet users in Social Media results a high risk for user's data due to cyber security breaches and data theft. Now a days, hackers use spam emails to data theft of social media's profile. They send the spam mails with malicious links, after clicking on the link, details of the profiles are sent to hackers. In this Paper, we discuss the various steps taken to protect the social media profile and propose spam filter using decision tree model. We compare the testing and training accuracy of various algorithms on the emails_small and emails_full datasets and try to explain which one is better with different dataset. We evaluate the performance of Decision Tree (DT) based classifier with Information gain criterion which achieve accuracy 90% and Gini impurity criterion which achieve accuracy 89%. Evaluate the ROC curves for DT and K-Nearest Neighbors (KNN) classifier over emails_full dataset to test the idea of the classifier quality and can be used to quantify the area under the curve.

Selection and peer-review under responsibility of the scientific committee of the Emerging Trends in Materials Science, Technology and Engineering.

1. Introduction

Social media profile's security is always a challenge as hackers & attackers always discover the new ways to perform cybercrime & cyber-attack. Although researchers have done many researches in this domain but still, there is a requirement to research and innovate more effective solutions to countermeasure new and emerging vulnerabilities & threats. Even after many years of solutions & research towards the social media security, there are still some flaws and loop hole exist which clauses security threats including social engineering attacks, data theft, social engineering attacks, etc. New technology domains like Industry 4.0, Cloud Computing, Blockchain, Edge computing & Internet of Things (IoT) have carried new security solutions, issues and challenges. In addition to that, the evolution in Artificial Intelligence and Machine Learning combined with Big-Data allows improved vulnerability & real-time threat analysis (Table 1).

The rapid growth of the internet & other related technology, give rise to new breakthrough in every many fields such as business, education, sports entertainment and so on [1]. With an increase amount of people getting connected to Internet, the security threats which leads to massive harms are also increasing.

* Corresponding author.
E-mail addresses: vashistha31@gmail.com (V.D. Sharma), kamakhya.vphcu@gmail.com (K.N. Singh), suraj@iiit-bh.ac.in (S. Sharma).

https://doi.org/10.1016/j.matpr.2020.12.377
2214-7853/
Selection and peer-review under responsibility of the scientific committee of the Emerging Trends in Materials Science, Technology and Engineering.

Table 1
Diagnosing Classification Predictions.

	Predicted: Spam Email	Predicted: Real Email
Actual: Spam Email	True Positive	False Negative
Actual: Real Email	False Positive	True Negative

Cybercrime and Cyberattack are the global problems which drawn the attention of many researchers [19]. It possesses the risk or threat to individual data security and privacy. Even though the bigger MNCs', Governments' and Banks' data are at risk. In today's scenario organized cybercrimes are becoming the new threats. This type of cybercrime is performed by the group of highly trained developers, hackers and network professionals who are continuously performing cyberattacks, and exploit so much data. It makes extremely challenging to provide Cybersecurity measures. Information security & Cyber security are the countermeasure techniques against the cyber-attacks, unauthorized access, data theft. Information security & Cyber security includes actions for providing security, privacy, reliability, integrity users' data [21].

To secure the System from Cybercrime, Cybersecurity is the protection of internet-connected systems, including hardware, software and data, from cyber-attacks. Internet also has its own disadvantages like illegal activity or Crime committed on the internet. So, in order to secure the data from these type of problems, Cyber Security is required [2]. There are various major problems that cause the cyber security - spam emails, virus, hackers, malware, phishing etc. A hacker is an individual who uses computer, networking or other skills to create a technical problem on targeted system. The hacker is an individual who practices his or her technical expertise to get unauthorized access to the network or systems in order gain data or information. Generally, three types of hackers are there – black hat hackers, white hat hackers and grey hat hackers. The white hat hackers work as an ethical hackers or plain old network security specialists used to work as a full-time technician. Linus Benedict Torvalds creator and principal developer or Linux kernel which is the core of secure open source Linux operating system is a great example of white hat hacker. For developing this he did the security analysis on exiting operating system by applying different attack and security measures. On other hand, black hat hackers use their skills to harm the system. Kevin Mitnick was the example of black hat hacker who is most infamous in the world. While Gray hat hackers is blend of both black hat & white hat hacking. Whereas Marcus Hutchins is famous gray hat hacker. China is the maximum cyber attacked country; 41% of the globe and Singapore is the best prepared for the cyberattack.

Cyberattacks include widely believe un-hackable organizations like Apple's server and US white house. A 16 years old, high school student from Australia, hacked the Apple servers and he was able to downloaded aver 90 GB of secured files, including the authorization keys of users for login and end up to access multiple accounts. There are a large number of black hat hackers who habitually hacks the government websites. Some decade back an Indian hacker hacked the official website of white house USA. When widely heavy tightened secure organization is hackable then just amazing about other organization and common internet users.

Cyberattacks are malicious activities to damage or access the user data and computing system. Cybersecurity includes detecting, responding and preventing to the cyberattacks which effects on any community the organizations, individual, or the government organization. Protection from attacks have become a crucial problem to be deliberated for an individual user, any organization or any national level agency.

On 18th July 2019, Lt. Gen. (Dr) Rajesh Pant, National Cyber Security Coordinator, declared that today's cybersecurity threat is extremely scary. Due to cybercrimes, we are downing almost 2.5% of the nation GDP.

According to national cyber security coordinator report [23], 232 records were exposed per second, 4 billion passwords have been stolen in recent years, one in two persons use same password for multiple accounts, 30 percent users use special dates like birthday for their password and most used password in recent years is 123456. Spam emails are used as one of the most significant form of cybercrime. They may contain malicious programs as attachments or have links to malicious websites full of malware and scams. So, there is a need to develop spam classifier in terms of interpretability, complexity and performance etc. Filtering spam from relevant emails is a typical machine learning task. Some spam classifiers are already existing using various machine learning based algorithms. The existing classifiers have some strengths and drawbacks also. But, spam filter using Support Vector Machine (SVM) is the best filter among existing spam filters and it achieves accuracy of 98% [20], while most of the existing spam filters achieve accuracy up to 85% [24]. However, For sizable datasets SVM would be not appropriate due to its excessive training time. Compared to Naïve Bayes, DT and KNN; SVM took additional time in training. It is sensitive to the type of kernel used so it works weakly with overlapping classes. In cybersecurity, for Spam detection and filtering SVM is a great option to explore but due to its dependency on type of kernel used, which distress the performance of SVM.

In this Paper, various steps taken to protect the social media profile and propose spam filter using decision tree model is discussed. Information such as word frequency, character frequency and the amount of capital letters can indicate whether an email is spam or not. We have used various Machine Learning based algorithms to classify spam. We compare the testing and training accuracy of various algorithms on the emails_small and emails_full datasets and try to clarify which one is performing better with which type of dataset. We also evaluated the performance of Decision Tree (DT) based classifier with Information gain criterion which achieve accuracy 90% and Gini impurity criterion which achieve accuracy 89%. Furthermore, we draw the ROC curves for DT and K-Nearest Neighbors (KNN) classifier over emails_full dataset to test the idea of the classifier quality and can be used to quantify the area under the curve.

The rest of the sections are organized as follows. Section 2 presents the literature survey, the proposed model and working mechanism of proposed model are presented in Section 3. In Section 4, the simulation results are discussed. In Section 5, future work is proposed and in the last section, conclusions have been provided.

2. Related studies

The Attackers and Hackers always try to find the new ways to launch Cybercrime and Cyber-attacks. Hence, social media profile's security has always been a challenge. While researchers have done many researches in this domain but still, researchers need to explore the flaws and fins the solution to countermeasure the new threats and weaknesses of the system. Even with so many years of effort, research and solutions contributed with the new emerging technology and techniques, there is still some chance to increase the threats to data security and privacy. In the existence of new technology such as Blockchain, IoT, Cloud computing, edge and Fog computing security is the major issue. Whereas to countermeasure these security threats, the advancement in Machine Learning allows new techniques and approaches are there to improve security mechanism.

In the literature survey, different related work which deal with prevention or detection using machine learning techniques for cyber security and exiting spam classifiers. The three major classes of social media security attacks to users shared data in social media [3]. The first class of attacks covers the conventional attacks like Sybil attack, malware, phishing, spamming, clickjacking, inference, deanonymization and profile cloning attack etc. The second class of attacks contain multimedia content and related threats. The third class of attack includes social media related threats and risks. In this category cyber bullying and grooming, corporate espionage, and cyber stalking. Moreover, attackers can steal bank account details by observing individual users' personal data and bank fraud [4]. Social media attacks can be user account hacking, impersonation attacks, malware distribution by analysis and fraud. A refined social media attack can compromise the business networks and organization system. Social media such as Facebook produce more personal data by like, share, and click. These data are used by hacker to target users. As per current statistics of Social media, average sharing and viewing of video and records on Facebook is rapidly growing. In every minute approximately 137,000 pics are uploaded and around 8 billion videos are viewed every day on Facebook, this much amount is double than year 2015. Because of huge volume of multimedia data present on social media, security risk is very high [7]. User identity and location can be tracked through malicious information shared on SM [8]. Even in twitter does not ask users to expose secret details, but hackers understand the sequence of the user's posted messages and reveal their undisclosed secret data. Sammy worm had hacked the My Space in 2005, he explores the weaknesses and communicated very fast. He had not stolen user's private details but proved that My Space is not safe. Mikeyy worm had hacked twitter in April 2009, he had not also stolen the user's private data, but exchanged their data with unusable data. Koobface worm had hacked Facebook in May 2009 and stolen private data [10]. Hence, The Internet Security Threat Report suggested to be careful about SMS from hackers. In 2018, hackers have hacked Mark Zuckerberg's, Facebook CEO Twitter and Pinterest accounts by using "dadada." his LinkedIn password.

Similarly, Newsweek and Delta Air Lines Inc. Social Media accounts were infiltrated by attackers and sent the fake messages. After analysis of aforesaid attack data, we felt that Social media account is not safe from hackers. Many resolutions have been proposed to alleviate these attacks by security corporations and researches. Such solutions include watermarking [12] stag analysis, spam detection, phishing detection and digital forgetfulness for protecting Social media users against attack towards multimedia data. There are several inherent security solutions are applicable like privacy setting, and authentication mechanisms [13]. Some commercial solutions are also available such as social protection application and insignificant monitor which protect from all types of threats. Researches have studied about security challenges in SNs. There are four types of threat issues: (a) Malware attacks, (b) Network structural-based attacks, (c) Privacy issues and (d) Viral marketing. They have explained in detail about issues and their protect mechanism.

Details about link prediction, user attributes and location hubs have been discussed by Novak [16]. Jin have discussed about user behavior such as I. connection and interaction, II. traffic activity, III. mobile social behavior and IV. malicious behavior.

In the paper [17], Author presented a review of several Security and Privacy threats about cyber security. In that the threats are categorized into four cases: I. modern threats, II. classic threats, III. threats targeting children and IV. combination threats. They have also proposed a classification of existing countermeasure for preventing cyber security. A possible of traditional security attacks in social media security based on social network stakeholders. Authors introduced various threats on Social Network Security (SNS) structure [22]. Authors also introduced various security mechanisms for mitigating these set of security attacks. But many challenges that occurs are using this mechanism in real world applications. They also described various similar attacks for cyber security, including phishing, identity theft, sybil, and other malware. Security exercise which involves five phases: preparation, dry run, execution, evaluation and repetition. Cyber range system [14] demonstrated and created a taxonomy on the basis of considering the various aspects. Intrusion detection system [11] utilized several machine learning algorithms on the base of deep learning approaches and they utilized thirty-five cyber datasets and classified into seven categories.

An E-Mail Spam Detection and Classification [15] demonstrated using SVM and Feature Extraction and attaining accuracy of 98%. Effects of performance discussed [5] using SVM for spam filtering due to choose of the kernel. The descriptions of six prevalent machine learning algorithms such as KNN, SVM, Bayesian classification, ANN, SVMs, Rough sets and Artificial immune system [16] and comparing their performance on the Spam Assassin spam corpus. SMD system to detect spam email [6] using hybrid bagged approach and achieved accuracy of 87.5% and implemented total of three experiments using individual Naïve Bayes, J48 algorithms and hybrid bagged and results are compared in terms of recall, precision, accuracy, true negative rate, f-measure, false negative rate and false positive rate. Adaptive approach and fast classification of email [18] using machine learning and cluster computing by generating new rules and using cluster approach increasing computing speed, whereas extracted features do not match with the existing rules then new rule is generated. Common Vector Approach [9] method is used to select features for classification of email, which is based on subspace pattern classifier for decreasing the number of features without affecting recognition rates. Support vector machine and Logistic regression-based system developed [25] for using to analyses different attack techniques for social media profile security and its subsequence and they focus on various security issues arise during tagging, commenting, sharing, liking and blogging on social media.

Predict the future attacks model developed legitimate and illegitimate access points to connected and remote network, which is to identify a rouge access points in Wi-Fi using decision tree and multi-layer perceptron algorithm and achieved higher accuracy than existing methods. Automated detection of malicious URLs [5] developed to utilizes two methodologies to identify malevolent URLs with two unusual datasets by applying Naïve Bayes and support vector machines (SVM) respectively.

3. Methodology

The evolution of learning and Intelligence initialized by observations of data such as finding the patterns in data, understanding the organization of data, get trained from the data and make the conclusion for future decision. In the Machine Learning several approaches are used for analysing and classifying the data; it mainly grouped into two categories supervised and unsupervised learning.

We filter spam from relevant emails as spam emails. Information such as word frequency, character frequency and the amount of capital letters can indicate whether an email is spam or not. We have developed Spam classifier to predict the spam email and applied on emails_small dataset, which is only a small fraction of the entire dataset emails_full. Now, this classifier is applied on emails_full dataset, and to calculate better accuracy, we generate the confusion matrix.

To increase the interpretability, the classifier is modified to filter spam based on one threshold and applied on emails_small as well as emails_full dataset. Now, dataset we added some relevant data such as word and character frequencies to filter the spam for every email. This time, dataset is spited into training and test set. We trained the data and applied prediction on test data. Here, we build two decision trees based on different splitting criteria like Information gain criterion and Gini impurity criterion. Finally, we compared DT model with k-Nearest Neighbors model through ROC curves to get the idea of the classifier's quality and can be used to quantify the area under the curve. Fig. 1 illustrated the proposed workflow architecture Fig. 2 represents the result of spam classifier using decision tree model and Fig. 3 shows the accuracy of KNN classifier. ROC curve of DT and KNN model is depicted in Fig. 4 and Fig. 5 presents the comparison between two classifiers.

3.1. Pre-Processing

The pre-processing phase is applied to eliminate the irrelevant data from email address

3.2. Feature selection and extraction

The Feature selection and extraction is applied to select and extract the features from the emails. Here, we have selected word's and characters frequency from emails and created a vector matrix. For training the classifier, word count vector of 3000 size for each email can be used and extract the feature and found most of them are zero.

There are 500 words in the dictionary. Each word count vector covers the frequency of 500 words in the training classifier. Suppose text in training file is "Get the work done, work done" so it can be written in encoded form as [0,0,0,0,0,. . .. . .0.0,0,2,0,0,0,. . . . . .,0,0,1,0,0,. . .0,0,1,0,0,. . .. . .2,0,0,0,0,0]. In this example all the word counts would be located at 296th, 359th, 415th, 495th index of 500 length word count vector and the rest are marked as zero. Here, the value at index 'ij' will be the number of incidences of j^{th} word of dictionary in i^{th} file.

3.3. Model training

For the model training, email spams are used. The spam content is used as the dataset for training the system and classifier is also trained by spam content.

3.4. Machine learning model

After training the system, the classifier (Model) would be prepared to classify the spam emails and will be applied on Testing dataset to classify the same.

3.5. Classifier testing

The classifier model would be tested with various training dataset to test the accuracy of the classifier.

3.6. Classification results

It produces result with the Boolean value i.e. 0 and 1 where, 1 means TRUE which means spam email and 0 means FALSE which means that is not a spam email, after supplying the sample email.

3.7. Performance evaluation

Estimate the performance analysis of proposed method with other existing methods. The purposed solution achieves up to 90% accuracy for Information gain criterion technique and 89% accuracy for Gini impurity criterion technique.

4. Experimental analysis

For implementation purpose, utilize spam emails dataset taking from UCI machine learning data. The attributes are spam (1) or not spam (0) and the results are found in the column spam. The considered feature in emails to predict whether it was spam or not is avg_capital_seq. It is the average amount of sequential capital letters found in each email. Here, we have developed a spam filter name as spam_classifier () using avg_capital_seq to predict whether an email is spam or not. In the function definition, it's important to realize that × refers to avg_capital_seq. So, where the avg_capital_seq is greater than 4, spam_classifier () predicts the email is spam (1), if avg_capital_seq is inclusively between 3 and 4, it predicts not spam (0), and so on. This classifier's methodology of predicting whether an email is spam or not seems pretty random. We have inspected the emails dataset, apply spam_classifier to it and compare the predicted labels with the true labels.

The concept of "SPAM" is diverse that is most of the time the email we get related to fast money making schemes, advertisements for product promotion or websites, pornography links, chain

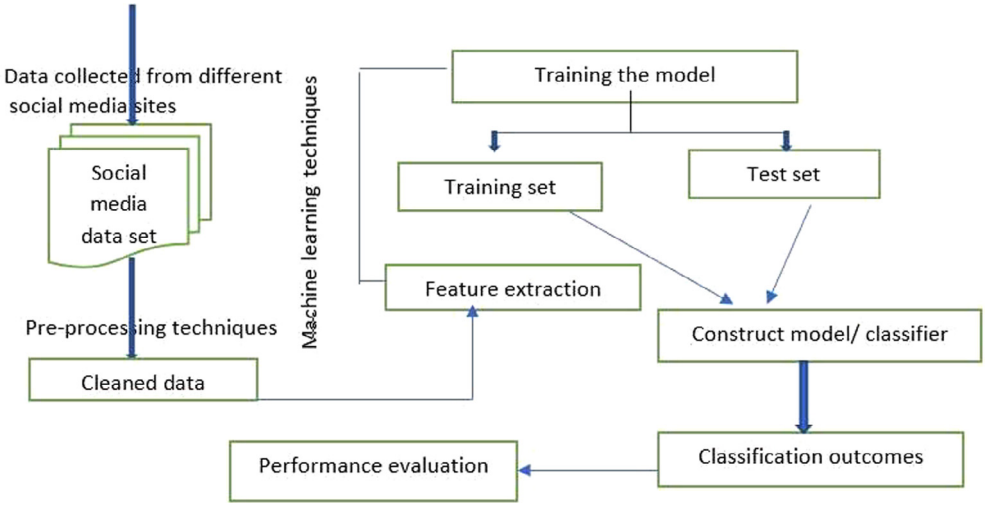

Fig. 1. Proposed Workflow Architecture.

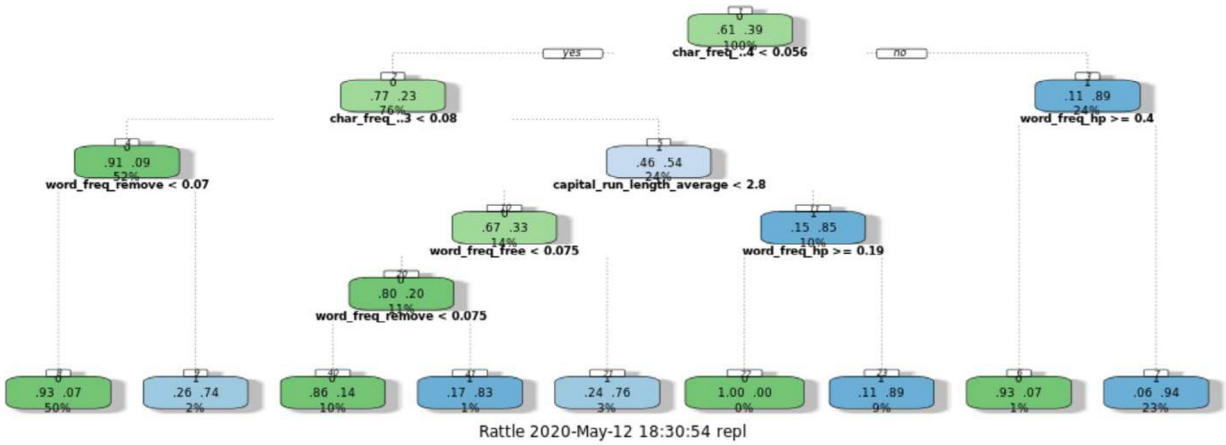

Fig. 2. Spam_Classifier using DT Model.

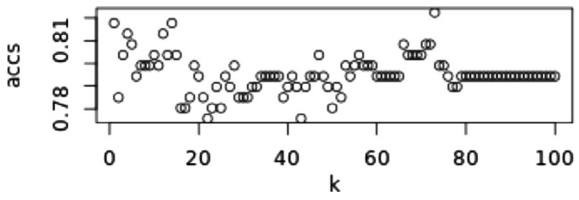

Fig. 3. The accuracy of KNN Classifier.

letters etc. The collection of spam emails used in this paper are collected from postmaster and individuals which had field spam. The non-spam emails are collected from personal emails and field work and therefore, the word 'George' and the area code '650' are indicators of non-spam content. All these are suitable for creating a personalized spam filter. With the help of those collected non-spam indicators, the spam contents are generated with the help of general-purpose spam filter.

It perfectly filtered the spam. However, the set (emails_small) we used to test our classifier is only a small fraction of the entire dataset emails_full. The accuracy for the set emails_small is equal to 1, but, the accuracy for the entire set emails_full is substantially lower. So, this spam_classifier () is bogus. It simply overfits on the emails_small set and, as a result, doesn't generalize to larger datasets such as emails_full. For applying this classifier on full dataset to calculate better accuracy, we generate confusion matrix. Confusion matrix is used to visualize the performance of a classifier.

A classifier predicts spam email correctly is True Positive (TP) and real email correctly is True Negative (TN) but, if a classifier does not predict correctly real email is True Negative (TN) and not predicts correctly spam email is False Positive (FP). If not, many real emails are predicted as a spam, is known as high precision and predicted most spam emails correctly is high recall.

Confusion Matrix of Spam Classifier:

```
      1    0
1  1123  690
0   892 1896
```

This classifier gave an accuracy of 65% on the full dataset, which is way worse than the 100% on the small dataset. Hence this classifier cannot be generalized also. Hence, above classifier is bogus and overfits on the small dataset such as emails_small set and as a result, does not generalize to larger datasets such as emails_full. So, we have modified the classifier and simply filtered spam based on one threshold for avg_capital_seq greater than 4 as spam. By doing this, it increases the interpretability of the classifier and restrict its complexity. However, this increases the bias, i.e. the error due to restricting its model.

Now we have simplified the rules of the spam_classifier. Emails with an avg_capital_seq strictly longer than 4 are spam (labeled with 1), all others are seen as no spam (0). Applied this spam_classifier on both datasets such as emails_small and emails_full. Calculated confusion matrix for both and accuracy of the both datasets respectively.

Here accuracy of the small dataset and big dataset is 77% and 73% respectively. Hence this model is no longer fits the small dataset perfectly but it fit the big dataset better. This increase the bias

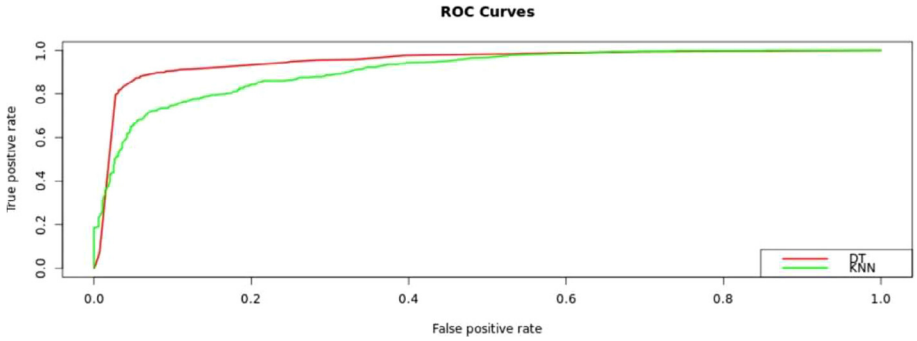

Fig. 4. ROC curves of DT Model & KNN Model.

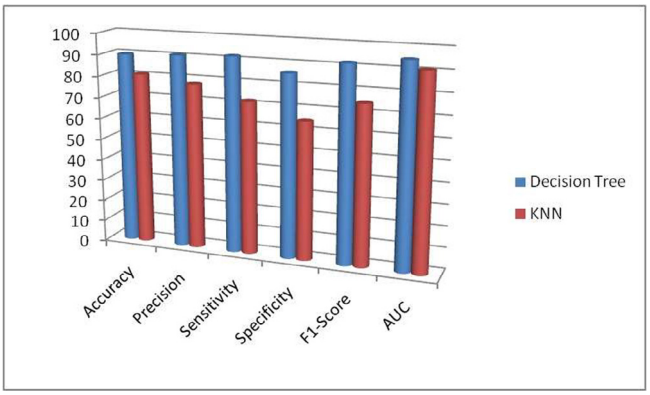

Fig. 5. Comparison between classifiers.

on the model and caused it to generalize better over the complete dataset. An accuracy of 73% is far from satisfying for a spam filter. So, this model also cannot be generalized.

So, again we modified the model to make generalize. We added some relevant data for every email that will help filter the spam, such as word and character frequencies. This time, we applied splitting criteria on dataset. Dataset is splitted into training and test set. We trained the data and applied prediction on test data. Here, we build two decision trees based on different splitting criteria like Information gain criterion and Gini impurity criterion. One decision tree is trained (tree_g) with the Gini impurity criterion, which rpart () uses by default, tested (pred_g), computed confusion matrix (conf_g) and calculate accuracy (acc_g). Second decision tree is trained (tree_i) with information gain criterion. To split the tree using the information gain criterion, changed the arguments in the rpart () function in the next block of code to test (pred_i), compute confusion matrix (conf_i) and calculate accuracy (acc_i). Here, we draw a fancy plot of tree_g and tree_i using fancyRpartPlot () function and calculate the accuracy of both the first and second models.

To get higher the gain, dataset should split the better. However, the standard splitting criterion of rpart () is the Gini impurity. Hence, using different splitting criterion can be influenced the resulting model using this algorithm. However, the resulting tree is quite similar and accuracy on the test set is comparable: 89% and 90%. It fits the big dataset better so; this model can be generalized. Confusion matrix of tree_i of DT model is below Where accuracy, precision, recall and F1 score is used to be calculated using following equations.

```
     0   1
0  771  65
1   78 466
```

Accuracy = (TP + TN) / (TP + FP + TN + FN) (i)
Classification accuracy: 90%
Precision = TP/(TP + FP) (ii)
Precision accurateness: 90.8%
Sensitivity = TP /(TP + FN) (iii)
Sensitivity 92.2%
Specificity = TN/(FP + TN) (iv)
Specificity 85.6%
F1score = 2. (Precision. Recall)/ (Precision + Recall) (v)
F1 score: 91.5%

Precision, sensitivity, specificity and F1Score is 90.8%, 92.2%, 85.6% and 91.5% respectively.

Confusion matrix of tree_g of DT model is below.

Precision, sensitivity, specificity and F1Score is 88.2%, 94.4%, 80.6% and 91.2% respectively.

Now, we compared this model with k-Nearest Neighbours (KNN) model in terms of accuracy, AUC value and ROC curves.

Accuracy of KNN model is 80.8% and AUC value of DT and KNN model is 95% and 90.7% respectively. Here, DT model is better than KNN because AUC value of DT is 5% greater than KNN.

To compare ROC curves, we access two models. This time, we have some predictions from two spam filters. These spam filters calculated the probabilities of unseen observations in the test set being spam. The real spam labels of the test set can be found in test$spam. The assigned probabilities for the observations in the test set are: probs_t for the decision tree model, probs_k for k-Nearest Neighbors. probs_t and probs_k are the probabilities of being spam, predicted by the two classifiers. Using prediction (), creating prediction objects called pred_t and pred_k for probs_t and probs_k respectively. Using performance () function, creating performance objects called perf_t and perf_k for prediction objects pred_t and pred_k. Now using predefined function draw_roc_lines (tree = perf_t, knn = perf_k) where the first argument is the performance object of the tree model, and the second one is the performance object of the k-Nearest Neighbor model. After executing the program, two curves are given below.

In the above plot you can see two ROC curves. The red one belongs to a decision tree model, DT, and the green one belongs to K- Nearest Neighbours classifier model, KNN. These curves are formed on the test set used in the previous model, i.e. a test set of the Emails_full dataset. ROC curve gives the idea of the classifier quality and can be used to quantify the area under the curve. Here curves of DT are close to upper left corner, its pretty good and better than KNN.

These results indicate that the Decision Tree classifier is most prominent in this dataset. Included accuracy, precision, sensitivity, specificity, F1-score, AUC parameters to compare the classifiers. As can see, all the performance parameters such as accuracy, precision, sensitivity, specificity along with F1-score and AUC value of the decision tree is higher than KNN classifier. So, decision tree is most promising classifier for this dataset.

5. Conclusion

In this Paper, we discussed the various steps taken to protect the social media profile and propose spam filter using decision tree model to protect from malicious URL's based spam email. Comparison of the performance of the various spam classifier on emails_small and emails_full dataset are presented. The experiments show a very promising result for the spam email classification. Two model of DT using Information gain criterion (tree_i) and Gini impurity criterion (tree_g) are presented to classify the spam emails. The resulting tree is quite similar and accuracy on the test set is comparable, while in terms of accuracy, precision, specificity, sensitivity and F1 score, We can find DT model using Information gain criterion (tree_i) is producing satisfying performance over DT using Gini impurity criterion (tree_g), more research need to be done to escalate the performance of tree_i. KNN classifier is also presented on emails_full to compare the performance with DT model. The resulting ROC curves of DT and KNN are quite similar and AUC value of DT and KNN is comparable, while in terms of accuracy, AUC value and ROC curve, we can find DT based model's result is better than KNN classifier-based spam classifier. ROC curve of DT is bigger and closer to upper left corner which justify the good model and better than KNN. Finally spam classifier based on DT model is most efficient way to generalize a spam classifier to filter malicious URL's base email.

6. Future scope

In the future, implement the proposed spam classifier in real environment to validate the classifier in real environment and address the challenges during implementation. Our proposed classifier can improve by considering more data samples to discover more deceases with better accuracy.

CRediT authorship contribution statement

Vishnu Dutt Sharma: Conceptualization, Methodology, Software. **Santosh Kumar Yadav:** Methodology, Software. **Sumit Kumar Yadav:** Software. **Kamakhya Narain Singh:** Writing - original draft. **Suraj Sharma:** Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Vishnu Dutt Sharma, Santosh Kumar Yadav, Sumit Kumar Yadav, Kamakhya Narain Singh, "Social Media Ecosystem: Review on Social Media Profile's Security and Introduce a New Approach", Advances in Computational Intelligence and Informatics, ICACII 2019. Lecture Notes in Networks and Systems, Springer, Volume 119, pp. 229-235.

[2] Verma Toran Shradhanjali, E-Mail Spam Detection and Classification Using SVM and Feature Extraction, Int. J. Adv. Res., Ideas Innovations Technol. IJARIIT 2 (3) (2017) 1491–1495.

[3] Manmohan Singh, Rajendra Pamula, Shudhanshu Kumar Shekhar, Email Spam Classification by Support Vector Machine, the proceeding of IEEE International Conference on Computing, Power and Communication Technologies (GUCON), 2019.

[4] M. Singh, R. Pamula and S. k. shekhar, "Email Spam Classification by Support Vector Machine," 2018 International Conference on Computing, Power and Communication Technologies (GUCON), pp. 878-882.

[5] W.A. Awad, S.M. ELseuofi, Machine Learning Methods for SPAM E-mail Classification, Int. J. Computer Sci. Inform. Technol. (IJCSIT) 3 (1) (2011) 173–184.

[6] Priti Sharma Uma Bhardwaj, Machine Learning based Spam E-Mail Detection, Int. J. Intelligent Eng. Syst. 11 (03) (2018) 1–10.

[7] Amandeep Singh Rajput, J.S. Sohal, Vijay Athavale, Email Header Feature Extraction using Adaptive and Collaborative approach for Email Classification, Int. J. Innovative Technol. Exploring Eng. (IJITEE) 8 (7S) (2019) 158–164.

[8] Serkan Gunal, Semih Ergin, Mehmet Bilginer Gulmezoglu, Omer N. Gerek, On Feature Extraction for Spam E-Mail Detection, in: Proceeding Multimedia Content Representation, Classification and Security, MRCS 2006, 2006, pp. 635–642.

[9] A. Swetha, K. Shailaja, An Effective Approach for Security Attacks Based on Machine Learning Algorithms, in: Proceeding Advances in Computational Intelligence and Informatics, ICACII 2020, Springer Lecture Notes in Networks and Systems, 2020, pp. 293–299.

[10] Srinivasu Badugu, Ramakrishna Kolikipogu, Supervised Machine Learning Approach for identification of malicious URLs, Advances in Computational Intelligence and Informatics. Springer Lecture Notes in Networks and Systems 119 (2020) 187–197.

[11] Zahra S. Torabi, Mohammad H. Nadimi-Shahraki, Akbar Nabiollahi, Efficient Support Vector Machines for Spam Detection: A Survey, (IJCSIS) Int. J. Computer Sci. Inform. Security 13 (1) (2015) 11–28.

[12] Amr E. Mohamed, Comparative Study of Four Supervised Machine Learning Techniques for Classification, Int. J. Appl. Sci. Technol. 7 (2) (2017) 5–18.

[13] IdamAlhamib IzzatAlsmadia, Clustering and classification of email contents, J. King Saud University Computer Inform. Sci. 27 (1) (2015) 46–57.

[14] Emmanuel Gbenga Dada, Joseph Stephen Bassi, Haruna Chiroma, Shafi'i Muhammad Abdulhamid, Adebayo Olusola Adetunmbi, Opeyemi Emmanuel Ajibuwa, "Machine learning for email spam filtering: review, approaches and open research problems", Heliyon, Elsevier, Volume 5, Issue 6, 2019.

[15] Zahra Razi, Seyyed Amir Asghari, Providing an Improved Feature Extraction Method for Spam Detection Based on Genetic Algorithm in an Immune System, J. Knowledge-Based Eng. Innovation 3 (8) (2017) 569–605.

[16] Mehdi Babagoli, Mohammad Pourmahmood Aghababa, Vahid Solouk, Heuristic nonlinear regression strategy for detecting phishing websites, in: Soft Computing A fusion of Foundations Methodologies and Applications, Springer, 2018, pp. 1–13.

[17] Uma Bhardwaj, Priti Sharma, Email Spam Detection using Ensemble Methods, Int. J. Recent Technol. Eng. (IJRTE) 8 (3) (2019) 4148–4153.

[18] M. Bassiouni, M. Ali, E.A. El-Dahshan, Ham and Spam E-Mails Classification Using Machine Learning Techniques, J. Appl. Security Res., Taylor & Francis 13 (3) (2018) 315–333.

[19] N.A. Ghani, S. Hamid, I.A.T. Hashem, E. Ahmed, Social media big data analytics: A survey, Comput. Hum. Behav. 101 (2019) 417–428.

[20] GuangJun, L., Nazir, S., Khan, H. U., & Haq, A. U. (2020). Spam Detection Approach for Secure Mobile Message Communication Using Machine Learning Algorithms. Security and Communication Networks, 2020.

[21] T. Gangavarapu, C.D. Jaidhar, B. Chanduka, Applicability of machine learning in spam and phishing email filtering: review and approaches, Artif. Intell. Rev. (2020) 1–63.

[22] M. Singh, Classification of spam email using intelligent water drops algorithm with naive bayes classifier, in: Progress in Advanced Computing and Intelligent Engineering, Springer, Singapore, 2019, pp. 133–138.

[23] G. Jain, M. Sharma, B. Agarwal, Optimizing semantic LSTM for spam detection, Int. J. Inform. Technol. 11 (2) (2019) 239–250.

[24] B.K. Dedeturk, B. Akay, Spam filtering using a logistic regression model trained by an artificial bee colony algorithm, Appl. Soft Comput. 106229 (2020).

[25] M. Diale, T. Celik, C. Van Der Walt, Unsupervised feature learning for spam email filtering, Comput. Electr. Eng. 74 (2019) 89–104.