Detecting Spam Messages on Social Networks in Saudi Arabia

Ayman Alharbi[†] **Waad Algethami**[†] **Abeer Alghamdi**[†] **Marwan Albahar**[†] **Saleh Ibrahim**[††]
aarharbi@uqu.edu.sa s43980125@st.uqu.edu.sa s43980116@st.uqu.edu.sa mabahar@uqu.edu.sa, s.ibrahem@tu.edu.sa
[†] College of computes and information Systems, Umm Al-Qura University, Mecca , Saudi Arabia
[††] Electrical Engineering Department, College of Engineering, Taif University, Al-Hawiya, KSA
[††] Department of Computer Engineering, Faculty of Engineering, Cairo University, Giza, Egypt

Summary

The most remarkable feature of the Internet during the last few years has been the fast propagation of social networking platforms. These platforms allow users to communicate with each other and share information. Consequently, tens of thousands of messages are generated every second on social networks. Nevertheless, several security threats exist in these networks of which spam messages are considered the most prominent. Therefore, a great deal of research has been conducted to detect such messages. However, Arabic research is still limited. Thus, in this research, we proposed a new Arabic spam detection system that combines the Rule-Based scoring technique with the Naïve Bayesian classifier to detect spam messages in Arabic that is specifically targeting Saudi Arabia users of social networks. After gathering and analyzing the dataset, we chose three content-based features that can distinguish spam messages from legitimate messages. Based on our experimental results, we showed that the Rule-Based scoring technique achieved 52% accurate detection results, while the Naïve Bayesian classifier achieved 86% accurate detection results.

1. Introduction

Security has grown to be a central concern in computer technology since the advent of the Internet. According to Internet World Stats, the number of Internet users around the world exceeded 7,796,615,710 as of March 3, 2020 [1]. In line with this, the Kingdom of Saudi Arabia has seen a remarkable increase in the number of Internet users since it was introduced to the public in 1997. Based on the available statistics, there were about 26 million users in 2017 - some 82% of the total population [2]. This growth has encouraged various parties such as the government, the health, and the education sector to introduce their services online.

However, in the last decade, social networks have emerged and become increasingly popular, and have become an integral part of people's lives. The first appearance of social media was in the form of e-mail, where people could exchange information and files via e-mail addresses. After the emergence of smartphones, many applications such as Twitter, Facebook, WhatsApp, Snapchat, and Instagram came into existence. These have become a large part of people's social lives. According to a 2018 statistic, there are around 2 billion users of social networks around the world [3]. In Saudi Arabia, around 25 million people use social networks, presenting up to 75% of the total population [4].

Many social network users are unaware of the security threats that exist in these networks, including privacy violations, identity theft, spam messages, and phishing techniques, although these threats have been dealt with and handled previously. However, they have increased considerably as a result of the nature of social networks. Such attacks benefit from the wealth of personal information published in social media and can be used to attack users and their friends. These are known as social engineering attacks. For instance, an attacker can place malicious code within a spam message in order to alter personal information on a user's profile. Given the nature and spread of networks, the chance that a user will be attracted to such a message is large. However, these spam messages are a threat that can be found everywhere in social networks, either in direct messages, status updates, comments on videos, contact requests, etc. They usually target user resources such as their credit card number, secret account numbers, and even the user’s computer bandwidth, which is used for sending spam messages.

The number of spam messages on social media is increased rapidly. One message out of every 200 messages on social networks and a tweet out of every 21 tweets on twitter is classified as spam messages [5]. The need for identifying and filtering out these spam messages in social media is becoming urgent. Moreover, due to the political, economic and geographical position of the Kingdom of Saudi Arabia, it has become the most widely targeted country in terms of spam messages through social networks in the Middle East [6]. Hence, this research will focus on detecting Arabic spam messages that currently target different social networks in Saudi Arabia.

We organize the rest of this paper as follows: Section 2 presents the literature review. Section 3 describes our Arabic spam detection system and the test results obtained. Section 4 concludes the research.

2. Literature review

The research topics featured in the spam detection field include the detection of spam emails, spam webpages, spam

Manuscript received May 5, 2020
Manuscript revised May 20, 2020

instant messages, and social network spam messages [7]. Due to the fact that detection techniques for spam messages in social networks are slightly different from those used to detect spam on emails or web pages, we focus in this review on existing work which deals with social network spam messages. Moreover, social networks spam detection techniques usually depend on extracting and analyzing certain features, whether these are content-based features such as words, patterns, and URLs within a message, or user behavior features such as number of followers, number and type of sent messages, period between each message, etc. After the features are extracted, machine learning models are applied for classification purposes.

In this section, we discuss some of relevant research that deals with spam messages in social networks both for Arabic and English messages.

2.1. Research on Arabic spam messages in social networks

Most of the existing research with regard to spam detection focuses on English content. In fact, there is a limited number of research projects that have been carried out on Arabic content spams on social media [7]. In this subsection, we will discuss some of them.

In [6] , an empirical analysis has been performed on Twitter spam accounts that were targeting users of Saudi Arabia. They applied existing features from previous research in order to detect spam accounts, and they gathered over 2187 spam accounts in a two-month period. Not only did they analyze the content and profile characteristics, but also the network metrics of these accounts to better understand the behavior and patterns of these spam accounts. It was found that the Twitter spam accounts targeting users in the Kingdom of Saudi Arabia were still in their infancy and were controlled by a third party in terms of performing retweets, spreading duplicate content, and polluting the trendy hashtags with specific content, to frame public opinion. Moreover, in terms of the behavior of spam accounts it has been observed they were flooding hashtags with repetitive tweets about one particular idea, using different accounts. They usually tweeted multiple times in a short period on one day and then vanished for several days. Therefore, the study concluded that there were many Twitter accounts that had been created and customized to serve a particular purpose. However, the researcher attributed the reason for the establishment of such accounts as the recent unstable political situation in the Middle East, Ultimately, it will enable computer scientists to implement a spam detection system that helps to decrease the number of spam messages that target users in Saudi Arabia. That study has been considered the first research that analyzes the behavior of Arabic content of Twitter spam accounts in Saudi Arabia.

In [7] , a supervised spam detection approach has been developed to analyze Arabic content on social networks. The authors focused on spam comments such as those featuring repetitive content, advertising content, automated content, or inappropriate content. They collected posts and comments from Facebook social network for over 30 days by targeting the most commonly used Algerian pages. They chose Facebook because it is the most widely-used social network among Algerian users. From these selected Facebook pages, they collected 99 posts and almost 9697 related comments that contain 1112 spam comments and 8585 non-spam comments. This dataset was unbalanced since there has been a large difference in non-spam records and spam records. However, they balanced it by reducing the 7473 non-spam comments from the dataset. Moreover, they manually analyzed the content of the unbalanced dataset in order to identify features that distinguished spam content. They selected nine features comment size, number of hashtags, number of lines, number of diacritics, number of emoticons, existence of specific sequences, frequent repetitions of a comment, user publication frequency, and similarity between post and comment topics. For classification, seven machine learning algorithms Naive Bayes, Decision Table, J48, SMO, Logistic, Regression Classifier, and LWL has been tested. Their results showed that the J48 gave the best classification results with 91.73% of correctly classified instances for the unbalanced dataset and 76.57% for the balanced dataset.

In [8] authors conducted a comprehensive study of spam messages and the size of this problem in Saudi Arabia. Data has been collected/gathered from a wide range of relevant stakeholders via questionnaires, interviews, and meetings. The study covered spam messages on SMS, fax, and e-mails in different forms such as direct marketing, sports, phishing, etc. Also, spam countermeasures and awareness in different organizations were examined. They proposed a definition for spam since there was no formal and legal definition in the Kingdom. Their definition earned a 97.4% approval rate. Also, a framework to collect spam-related statistics has been developed that focused on three aspects:

- A comparison of spam rate, i.e. the fraction of messages that are spam,
- An identification of the spam sources, and
- A collection of spam statistics done by checking reliable data and by conducting a survey, interviews, and discussions with relevant personnel.

The results showed that the average number of spam e-mails in the Kingdom was 54%, 6% in faxes and between 1.25% and 1.75% in case of SMS, which indicates a serious problem. The main type of spam messages in the Kingdom of Saudi Arabia takes the form of direct marketing messages. Most of the organizations had not educated their employees on how to deal with these spams. The one exception was banks which provide educational programs to their employees and also to their customers. About 83% of stakeholders had a tool to fight against spams. They recommended to develop a national anti-spam policy framework to effectively fight spam in the Kingdom of Saudi Arabia by using this information.

In [9], they analyzed the content of Saudi tweets in order to detect spam tweets. They used two approaches: a Rule-Based approach and a Supervised Learning approach. In the Rule-Based approach, they worked on four content-based features: the words (if they are in the spam lexicon then they are spam), the presence of a phone number (either local or international), hashtags per tweet (if there are more than four non-sentiment hashtags, the tweet is spam), and URL presence. In Supervised Learning algorithm, they used two machine learning classifiers: Naïve-Bayes and Support Vector Machines (SVM). They performed experiments on a balanced dataset (containing 2500 spam tweets and 2500 non-spam tweets) and unbalanced datasets (contain 1054 spam tweets and 1992 non-spam tweets), with and without features. The Machine Learning approach gave better results. The Rule-Based approach was considered good if there were not enough training sets. The best features when it comes to classifying whether tweets are spam or not were the phone number & number of hashtags. The authors suggested as a future work to identify the existence of opinion spams (which are aimed to drive people to certain opinions).

2.2. Research on detecting English spam messages in social media.

In this subsection, we discuss the most common machine learning models for classification of spam messages in social networks, and the most prominent related work.

2.2.1. Naïve Bayesian classifier: The Naïve Bayes Classifier is based on Bayes' theorem and is considered to be one of the most efficient and effective classification algorithms. It can work on small sample sizes and produces an accurate classification result [10].

In [11], they worked on detecting spam on Twitter. They collected a real dataset from Twitter’s publicly-available information, then analyzed the dataset and extracted 6 features. The features are further divided into three content-based features and three user behavioral-based features. The content-based features are number of HTTP links, number of duplicate tweets, and number of replies and mentions, while number of followers, number of friends, and follower ratio are the behavioral-based features. They used and evaluated several machine learning approaches including Decision Tree (DT), Neural Network (NN), Support Vector Machine (SVM), and Naïve Bayesian classifier (NB). The results showed that the Naïve Bayesian classifier gives the best performance.

In [12], they introduced a new perspective for distinguishing between spam and legitimate content on two of the most popular social networks: Twitter and Facebook, due to their similar characteristics in terms of posts and user activities. Moreover, they collected two new datasets through their APIs. In the case of the Twitter dataset, they collected tweets from June to August 2015 and found 1937 spam tweets and 10942 legitimate tweets. In terms of the Facebook dataset, they collected data from July to August 2015. After that, they combine these datasets into one dataset that consisted of 1338 spam posts and 9285 legitimate posts. Weka tool was used for classification purposes. This tool contains multiple traditional classifiers such as Naïve Bayes, Logistic, Random Tree, J48, and Random Forest. From the results it was concluded that the detection of spam messages on Facebook decreases 50% of the spam on Twitter, while the detection of spam on Twitter decreases up to 71.2% of the spam on Facebook. This proves that the detection of spam in one social network can significantly help in the detection of spam in other social networks.

2.2.2. Support Vector Machine Classification: The purpose of the support vector machine algorithm is to find hyperplanes in an N-dimensional area (N-number of attributes) that clearly classify data points [13].

In [14], the authors proposed a machine learning-based spammer detection solution for social media. They collected a dataset of 16 million messages of 30,116 users from the Sine Weibo social network site. The dataset has been manually classified the users into non-spammers and spammers. After that, they extracted several features from the message content and the users' behavior and applied them into the SVM algorithm for classification purposes. Their solution is feasible and shows a better classification result.

2.2.3. Decision tree: A decision tree is a streamlined tree-like structure, where each internal node represents a test of an attribute, each branch being the result of the test, and the class label is represented by each node of the foliage (or node). Given group X, attribute values are tested in the group against the decision tree. The path from the root to the node of the sheet that carries the class prediction is tracked [15].

In [16], they proposed a fundamental evaluation of several machine learning algorithms on the detection of streaming spam tweets. To perform this evaluation, they gathered about 600 million public tweets, from which 6.5 million were spam tweets. For real-time spam detection, they extracted 12 features that can be used to differentiate spam tweets from non-spam tweets. They then leveraged these features to several machine learning-based spam classification algorithms, including decision tree, random forest, Naive Bayes, C4.5, Bayes network, support vector machine, and k-nearest neighbor. They found that the ability to detect spam tweets decreased when it was in a real-world scenario. Also, increasing training data wouldn't be more beneficial in terms of detecting spam tweets after a certain number of training samples.

2.2.4. Neural networks: Artificial neural networks are the modeling of the human brain. Each neural network consists of a large number of neurons that are connected together with certain coefficients. In the training process, the information is distributed in these connection points, thus that the network learns [17].

In [18], the authors presented a new approach for detecting spam tweets on Twitter. They claimed that the ways of detecting spam on Twitter that involved detecting and blocking spammers were not useful since spammers can create other accounts and post the spam tweets again. Therefore, they provided a new approach that can detect spam tweets at the tweet level and inhibit them from spreading in the network. The approach combines both traditional feature-based and deep learning methods using a multilayer neural network that acts as a meta-classifier. Moreover, they developed multiple deep learning models depending on convolutional neural networks (CNNs) that show good results in terms of multiple natural language processing tasks. They used one feature-based model and five CNN models. The feature-based model uses user-based, content-based, and n-gram features. They trained each CNN model using different word embedding's, and they used the tweets as the only input without any additional information. They evaluated the approach on two data sets, a balanced dataset (HSpam), and an unbalanced dataset (1KS10KN). The results showed that the developed approach outperforms the earlier methods.

In [19], the authors studied the social network from dissimilar directions and took into consideration the characteristics of spam posts and spammers. They presented an artificial intelligence-based spam detection solution for social networks. The solution is built on extracting a number of features from the message content and format, using them to train the Feed Forward Neural Network to classify the message as spam or not spam. The authors concluded that their solution is feasible and reliable when it comes to achieve an accurate detection result.

3. The proposed Arabic spam detection system for social networks in Saudi Arabia

In this work, we aim to identify and detect Arabic spam messages that target Saudi Arabia's social networks. To this end, our proposed Arabic spam detection system consists of six main stages as shown in Figure 1.

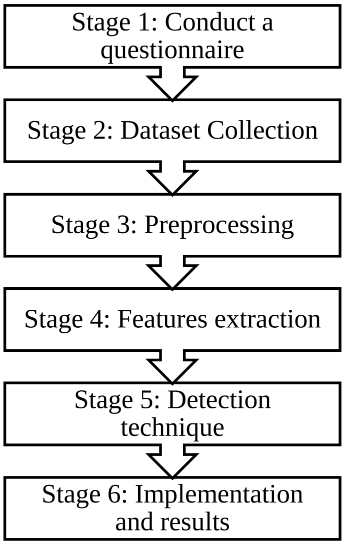
Stage 1: Conduct a questionnaire
Stage 2: Dataset Collection
Stage 3: Preprocessing
Stage 4: Features extraction
Stage 5: Detection technique
Stage 6: Implementation and results

Fig. 1 The proposed Arabic spam detection system stages.

Each stage will be discussed in detail in the following subsections.

3.1. Conduct a questionnaire

A survey has been conducted about spam messages in Saudi Arabia's social networks. The target group for the survey was social media users residents of Saudi Arabia. The questionnaire was divided into two parts. The first part consisted of questions intended to measure the awareness of Saudi society concerning spam messages and the users' perception of the spread and gravity of spam messages. The second part was a text box in which the targeted group could add spam messages that they had come across. This was helpful in terms of gathering the spam messages we used in this research.

The question part comprised the following questions:

1) How many messages have you received that were clearly spam messages?
2) Can you distinguish between regular and spam messages?
3) Do you want to make sure that the messages you want to resend are safe?
4) On which social networking sites did these messages appear?
5) If any of the following phrases are found in a message, you will not hesitate to copy and resend: The phrases are Urgent - Royal Orders - increased salary - subsidies - additional income - Congratulations! You have won - vitamins and tonics - study suspension - other.

The results of 257 replies revealed that 42% of people had received from 1 to 10 spam messages, while around 33% of the respondents did not count them. This is as shown in

Figure 2. However, 52.5% of people think, they can recognize spam messages, their result shown in Figure 3. Moreover, 89.9% of people want to be sure that the message is not spam before they send it. This is as shown in Figure 4. Noteworthy, the largest number of spam messages appeared in WhatsApp and SMS as shown in Figure5..

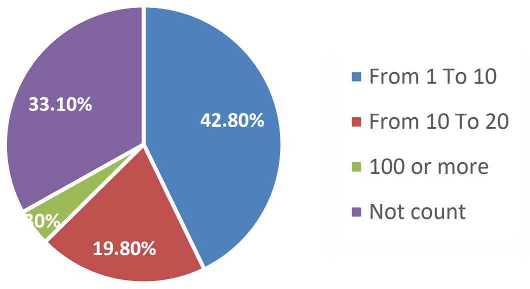

Fig. 2 The number of spamming messages received by social media users.

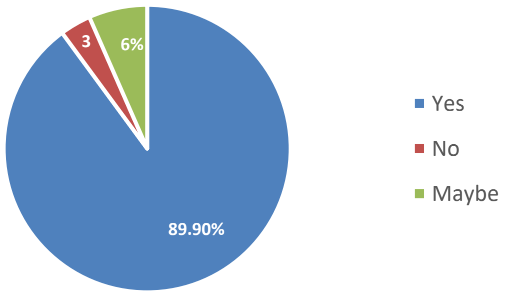

Fig. 3 percentage of users who can distinguish between regular and spam messages

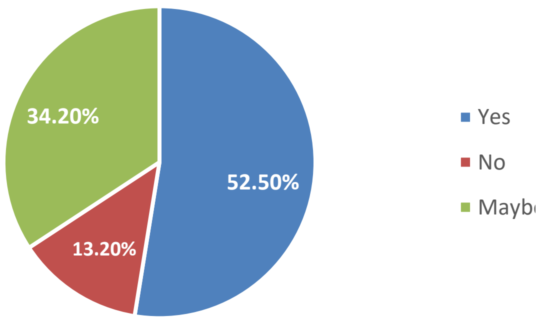

Fig. 4 Whether or not users want to resend safe messages.

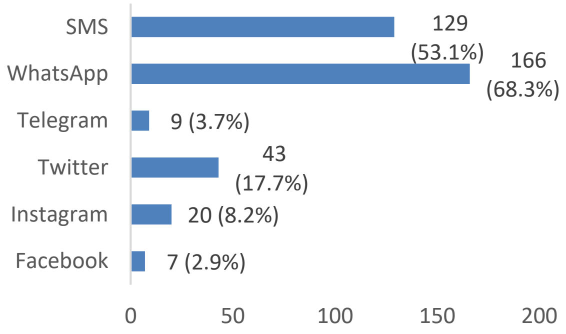

Fig. 5 The appearance of spam messages on different social networking sites.

Moreover, the phrases that raised the alarm were “ تعليق الدراسة”: “Study suspension” with (39.7%) and “أوامر ملكية”: “Royal Orders” with (30.7%) respectively. More details in Table 1.

Table 1: The phrases that raise the alarm.

English Translation	Percentage	Arabic Phrases
Study suspension	39.7%	تعليق الدراسة
Royal Orders	30.7%	أوامر ملكية
Urgent	26.8%	عاجل
Congratulations! You have won	17.5%	مبروك ! لقد فزت
Increased salary	16.3%	زيادة رواتب
Vitamins and tonics	11.7%	فيتامينات ومقويات
Financial subsidies	10.1%	إعانات مالية
Bonus	7.4%	دخل إضافي

3.2. Dataset Collection

We randomly collected messages from different sources such as WhatsApp and SMS. These messages were labeled manually into spam and non-spam messages to prepare the dataset for the algorithms. The collected dataset focused on the most common spam and non-spam messages on social networks in Saudi Arabia during the period from November to December 2018. The training dataset consisted of 100 messages (35 spam and 65 non-spam) and the testing dataset consisted of 50 messages (15 spam and 35 non-spam).

3.3. Preprocessing

The preprocessing stage was manually performed in order to prepare the messages for the analyzing stage, and to improve the accuracy of the proposed model. However, in the case of Arabic messages involved three steps as follows:

- Removing symbols and non-Arabic characters.
- Normalization: by transforming every letter to its standard form, for instance, the letters(إ ,آ ,أ) are converted into (ا), the letters (ؤ ,ئ) are converted into (ء) and the letter (ه) converted into (ة).
- Removing repeated letters and elongations: for instance "ابداااااااا " ,will become "ابدا" and "انـــــا" will become "انا".[9].

3.4. Features extraction

We analyzed the contents of the collected spam messages in the training dataset to understand the characteristics of those messages during the stage, we focused on finding the most common words, patterns, and URLs found in the spam messages to identify the features that will distinguish spam content from non-spam content in Saudi social networks.
After analysis, we chose three content-based features which are described as follows:

- The words in the message: there are certain words that frequently appear in spam messages. Therefore, they can be considered as an indicator of spam. These words were added to a blacklist entitled Arabic spam words blacklist which contained 47 words. Example of these words is shown in Table 2
- The URLs within the spam messages: they were added to the spam URLs blacklist.
- The phone numbers: some spam messages contained a phone number that was added to the spam phone number blacklist.

Table 2: Part of Arabic spam words blacklist.

English Translation	Arabic spam words
Achieve	حقق
Salary	راتب
Additional	اضافي
Double	مضاعف
Fortune	محظوظ
Investment	استثمر
Voucher	عرض
Free	مجاني

3.5. Detection technique

We have chosen two detection techniques to build our Arabic spam detection system:

3.5.1 The Rule-Based scoring technique:

This was chosen for its detection accuracy in real-time scenarios, its low processing time, and its suitability for use with small training sets [9].
Algorithm 1 presents the details of our Rule-Based scoring in pseudo-code:

Algorithm 1: Rule-based scoring

```
INPUT: Message M, Arabic spam words blacklist, spam URLs blacklist, spam phone number blacklist.
OUTPUT: P = (Spam, Non-Spam).
INITIALIZATION: Score: = 0, P: = Non-Spam,
and invoke preprocessing of message M.
FOR each $W_i \in$ M DO
    IF ($W_i \in$ spam URLs blacklist) THEN
        P:= Spam
    END IF
    IF ($W_i$ = international phone number OR
        $W_i \in$ spam phone number blacklist) THEN
        P := Spam
    END IF
    IF ($W_i \in$ Arabic spam words blacklist) THEN
        Score := Score + 1
    END IF
    IF Score = 3 THEN
        P := Spam
    END IF
END FOR
Write P // print the result
```

3.5.2 The Naïve Bayesian Classifier:

This was chosen for its efficiency and effectiveness in dealing with most text classification problems [9]. Moreover, it is a probabilistic classifier with an independence assumption between the features [20].

3.6. Implementation and results

We used the Java programming language to implement our Arabic spam detection system that incorporates two different detection techniques - the Rule-Based Scoring technique and the Naïve Bayesian Classifier. The user will copy and paste the message, chose the detection techniques listed in the program. After that, the detection result will be shown in the label on the interface, and the message will be classified as spam or non-spam.
However, we have used the testing dataset to test our proposed system. We used the following equation to estimate the accuracy of each detection technique.

$$accuracy = \frac{C}{N} \times 100\%$$

Where C is the number of samples that were detected correctly and N is total number of samples in the testing dataset.

Table 3 and Table 4 present the tested messages using the Rule-Based Scoring algorithm and the Naïve Bayesian Classifier respectively.

Table 2: Results of the Rule-Based Scoring algorithm

Testing dataset (msg)	Classified as spam	Classified as non-spam	Accuracy
Spam (15)	7 msg	8 msg	52%
Non-spam (35)	16 msg	19 msg	

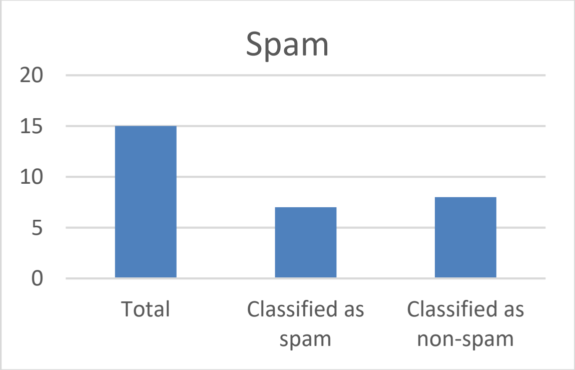

Fig. 6 Rule-Based spam message results

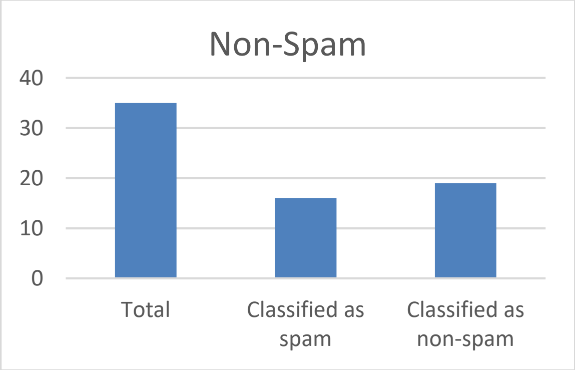

Fig. 7 Rule-Based Non-spam message results

Table 3: Results of the Naïve Bayesian Classifier.

Testing dataset (msg)	Classified as spam	Classified as non-spam	Accuracy
Spam (15)	10 msg	5 msg	$\frac{33+10}{50} \times 100$ $= 86\%$
Non-spam (35)	2 msg	33 msg	

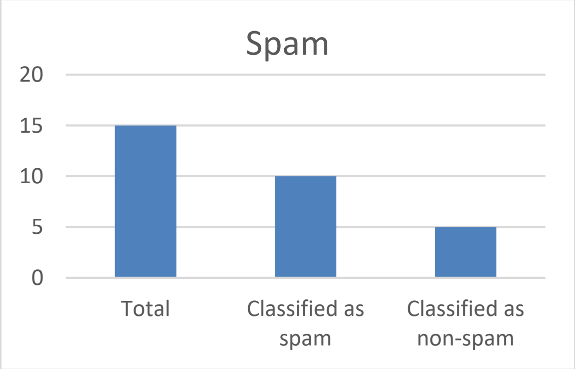

Fig. 8 Naïve Bayesian spam message results

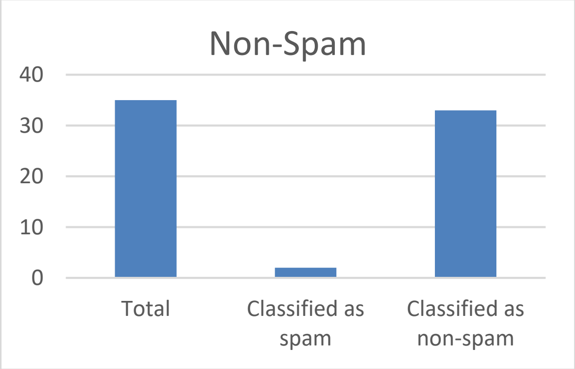

Fig. 9 Naïve Bayesian Non-spam message results

As can be depicted from figures 6,7,8 and 9, the Rule-Based Scoring technique gave 52% accurate detection results, while the Naïve Bayesian Classifier gave 86% accurate detection results.

4. Conclusion

In this research, we have proposed a spam detection system that aims to detect Arabic spam messages on Saudi Arabian social networks. We tested the proposed system on small datasets to demonstrate its effectiveness. Our results show that the Naïve Bayesian Classifier gave better detection rate than the Rule-Based Scoring technique with an accurate detection result of 86%. In the future, we will focus on real-time detection, including more detection techniques in our system, and use more features.

References

[1] Internet World Stats. (2020). Internet users in the world by regions. https://www.internetworldstats.com.

[2] Communications and information technology commission, Annual Report 2017, Saudi Arabia. https://www.citc.gov.sa/en/MediaCenter/Annualreport/Pages/default.aspx

[3] Kaur, R., Singh, S., & Kumar, H. (2018). Rise of spam and compromised accounts in online social networks: A state-of-the-art review of different combating approaches. Journal of Network and Computer Applications.

[4] Almgloth, A. (2018). congress of " Government communication with Snape Chat ". Alriyadh newspaper.

[5] Inuwa-Dutse, I., Liptrott, M., & Korkontzelos, I. (2018). Detection of spam-posting accounts on Twitter. ELSEVIER.

[6] Al-Khalifa, H. S. (2015). On the Analysis of Twitter Spam Accounts in Saudi Arabia. Riyadh : International Journal of Technology Diffusion.

[7] Mataoui, M., & Zelmati, O. (2017). A Proposed Spam Detection Approach for Arabic Social Networks Content. IEEE International Conference on Mathematics and Information Technology (ICMIT).

[8] Al-Kadhi, M. A. (2011). Assessment of the status of spam in the Kingdom of Saudi Arabia. Riyadh: Journal of King Saud University.

[9] Al Twairesh, N., Al Tuwaijri, M., Al Moammer, A., & Al Humoud, S. (2016). Arabic Spam Detection in Twitter. ResearchGate.

[10] Mansour, A. M. (2018). Texture Classification using Naïve Bayes Classifier. IJCSNS International Journal of Computer Science and Network Security,.

[11] Wang, A. (2010). Detecting spam bots in online social networking sites: a machine Learning Approach. Springer, Data and Applications Security and Privacy XXIV.

[12] Xu, H., Sun, W., & Javaid, A. (2016). Efficient spam detection across Online Social Networks. IEEE International Conference on Big Data Analysis (ICBDA).

[13] Gandhi, R. (2018). Support Vector Machine — Introduction to Machine Learning Algorithms. Towards Data science.

[14] Zheng, X., Zeng, Z., Chen, Z., Yu, Y., & Rong, C. . (2015). Detecting spammers on social networks. Neurocomputing

[15] Himani Sharma1, S. K. (2018). A Survey on Decision Tree Algorithms of Classification in Data Mining. International Journal of Science and Research (IJSR) , 2319-7064

[16] Chen, C., Zhang, J., Xie, Y., Xiang, Y., Zhou, W., Hassan, M., . . . Alrubaian, M. (2015). A performance evaluation of machine learning-based streaming spam tweets detection. IEEE Transactions on Computational social systems.

[17] Mijwel, M. M. (2018). Artificial Neural Networks Advantages and Disadvantages. ResearchGate.

[18] Madisetty, S., & Desarkar, M. S. (2018). A Neural Network-Based Ensemble Approach for Spam Detection in Twitter. IEEE Transactions On Computational Social Systems.

[19] Sabharwal, M., & Kaur, J. (2018). Spam Detection in Online Social Networks Using Feed Forward Neural Network. ResearchGate.

[20] wikipedia. (2019). Naive Bayes classifier. https://wikipedia.org.