Using machine learning to deal with Phishing and Spam Detection: An overview

Oumaima EL KOUARI
FSR, University Mohammed V
Rabat, Morocco
oumaima.elkouari@gmail.com

Hafssa Benaboud
IPSS, FSR, University Mohammed V
Rabat, Morocco
hafssa.benaboud@um5.ac.ma

Saiida Lazaar
ERMIA TEAM, ENSA of Tangier,
University Abdelmalek Essaadi
Tangier, Morocco
slazaar@uae.ac.ma

ABSTRACT

Cybersecurity is a growing field that requires a lot of attention due to the remarkable progress made in social networks, cloud and web technologies, online banking, mobile environment, smart networks, etc. Various approaches have been developed to solve many computer security problems, including those based on machine learning. This paper examines and highlights the various works using machine learning in network security. Two types of detection are discussed. Phishing detection and Spam detection. For each type, related work is presented and some proposed methods in the literature are compared taking into account their accuracy and other characteristics.

CCS CONCEPTS

• **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

KEYWORDS

Machine Learning; Cyber Security; Network Security; Phishing detection; Spam detection

ACM Reference Format:
Oumaima EL KOUARI, Hafssa Benaboud, and Saiida Lazaar. 2020. Using machine learning to deal with Phishing and Spam Detection: An overview. In *The 3rd International Conference on Networking, Information Systems & Security (NISS2020), March 31-April 2, 2020, Marrakech, Morocco.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3386723.3387891

1 INTRODUCTION

With the rapid development of web and mobile technologies and the integration of the Internet into our social life, attack techniques are also becoming more and more sophisticated and advanced, which exposes us to more and more serious security threats, then the cybersecurity field is in need of attention. Various approaches have been developed to solve many computer security problems, including those based on machine learning.

NISS2020, March 31-April 2, 2020, Marrakech, Morocco

ACM ISBN 978-1-4503-7634-1/20/03…$15.00
https://doi.org/10.1145/3386723.3387891

The trend of Machine Learning affects several scientific fields due to their unique properties such as adaptability, scalability and the potential for rapid adaptation to new and unknown challenges.. Today, many people interact daily with systems based on machine learning, for example in image recognition systems, such as those used on social media; voice recognition systems, used by virtual personal assistants; and recommendation systems, such as those used by online retailers.
With the rapid evolution of threats, the security of communications has become an important concern for users and companies, which leads researchers to adopt machine learning methods in their work for better network security and to help predict new threats.

In this paper, we are interested in studying and comparing a number of works on network security using various machine learning techniques. There are several types of network attacks. In this state of the art, we are interested in two types of attacks which are, phishing and Spam. Our paper is organized as follows. Section 2 gives some methods of machine learning proposed in the literature on phishing detection and, summarizes and compares these methods. Section 3 gives machine learning based methods to detect spams. Each section summarizes performances into tables and gives a comparison of the presented methods in a histogram. We conclude our paper with perspectives in section 4.

2 PHISHING DETECTION

Phishing is a special type of network attack in which the attacker aims to steal users' personal, financial, transaction or password data or to install malware on the victim's machine. Phishing attacks can be carried out by malware or social engineering, etc. Attackers use fake web pages or emails. Phishing is one of the major security challenges right now. Because of this type of attacks, many businesses and individuals have lost billions of dollars.

Nowadays, many solutions based on machine learning [1–13] have been proposed to detect various types of phishing. We present in this section works presented in [1], [2] and [3].

2.1 New Rule-Based Phishing Detection Method [1]

In [1], the authors devised a new rule-based method for classifying web pages using two new sets of features proposed to detect phishing web pages in internet banking, four characteristics to assess the identity of the resources of the web page and four characteristics to identify the protocol for accessing the resource elements of the web page and nine characteristics from related work ([9], [10] and [11]). The proposed model is based on the following steps:

- Characteristics extraction by examining the content of the page and the DOM (Document Object Model).
- Classification of web pages using the SVM (Support Vector Machine) algorithm with a three degree polynomial kernel and cross-validation by 10 folds.
- Rules extraction from the model using the SVM and DT (Decision tree) algorithms.
- The implementation of an extension to the Google Chrome browser.

The used data sets come from two different sources: $http://dir.yahoo.com$ and $http://www.phishtank.com$. In order to prepare the data for classification, the authors applied the preprocessing phase which consists in deleting irrelevant data, eliminating data with errors and losses, and removing redundant data.

The experiments carried out by the authors have shown that:

- The accuracy of the global detection of the model has been increased and the error rate has been reduced using the 17 characteristics taken into account.
- The positive effect of six characteristics on the eight proposed according to the sensitivity analysis of the proposed characteristics.
- The method is more precise when using the rules extracted from their model than when using those extracted directly from training data.

Hidden knowledge extraction from the model has been proposed for integration into a PhishDetector extension for Google Chrome in order to make the proposed solution more functional and easy to use. The evaluation of the browser extension implemented indicates that it can detect phishing attacks in Internet banking with great accuracy.

2.2 A ML based approach for phishing detection using hyperlinks information [2]

In [2], the authors presented another method of detecting phishing web pages by analyzing hyperlinks in the website's HTML source code. They incorporated various new characteristics specific to hyperlinks, so they divided them into 12 categories: total hyperlinks, no hyperlink, internal hyperlinks, external hyperlinks, internal error, external error, internal redirect, external redirect, null hyperlink, link to the form connection, internal/external CSS and internal/external favicon.

The performance of the proposed solution was evaluated with different algorithms: SMO (Sequential Minimal Optimization), NB (Naive Bayes), RF (Random Forest), SVM, Adaboost, ANN (Artificial Neural Networks), C4.5 and RL (Logistic Regression) using 10-fold cross-validation and the following datasets: Phish tank data set (2018), Stuffgate Free Online Website Analyzer and Alexa top websites (2018), list of online payment service providers (2018). They worked on live sites because the life of phishing sites is very short.

The results showed that the designed work is effective for the classification of phishing websites.

2.3 Intelligent rule-based phishing websites classification [3]

In paper [3], an intelligent rule-based phishing websites classification was proposed. In this method, the authors used 17 different features automatically extracted to detect phishing web pages.

The characteristics were categorized into 4 groups. Characteristics based on the address bar, Abnormal characteristics, Characteristics based on HTML/Javascript and Characteristics based on the domain. The characteristics are extracted automatically using Javascript programs and PHP scripts. In order to determine the most used features for the design of Phishing websites, the authors calculated the frequencies of each feature and the results showed that: Request URL, Age of Domain, HTTPS and SSL obtained the highest rate.

The experiments were carried out using different algorithms: C4.5, PRISM, CBA (Classification Based Association), RIPPER (Repeated Incremental Pruning to Produce Error Reduction) for each category of characteristics in order to extract new hidden knowledge.

One of the experiments was carried out before selecting the characteristics and which showed that the algorithm C4.5 surpassed RIPPER, PRISM and CBA in terms of accuracy and error rate. Another experiment is carried out after selecting the characteristics with the chi-square method. It has shown that the accuracy of the forecast has improved and the error rate has decreased for all algorithms specifically the CBA algorithm.

The selected characteristics are: request URL, age of domain, HTTPS and SSL, website traffic, long URL, subdomain/multi subdomain, adding prefix / suffix separated by (-) to domain, URL of anchor and using the IP address.

2.4 Summary and comparison

Table 1 summarizes the above presented ML-based methods to detect phishing, and table 2 gives a comparison between these methods. Figure 1 shows the histogram which compares the accuracy of each algorithm.

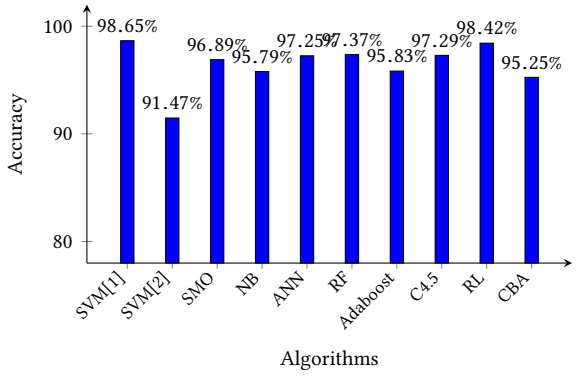


Figure 1: Comparison of Anti-Phishing Approaches Accuracy

Table 1: Summary of ML-based methods for Phishing Detection

Paper	Method	Contribution	Limits	future work
[1] (2016)	A method to detect phishing web pages in internet banking.	• Proposal for two new feature sets to increase the accuracy of web pages' classification. • Hidden knowledge extraction from the model for integration into a PhishDetector browser extension for Google Chrome.	• The model depends on the page's content. • No-HTML pages • Banking services only.	• Reducing dependency. • Identification of all types of web page. • Proposal of new methods and new systems to identify the different attacks on mobile devices.
[2] (2018)	A Method for detecting phishing attacks by analyzing hyperlinks in the website HTML source code.	• Proposal of a new set of functionalities based on the Hyperlinks of the web page.	• The model depends on the page's content. • No-HTML pages	• Reducing dependency. • Identification of all types of web page. • Proposal of new methods and new systems to identify the different attacks on mobile devices. • Integration of a tool in the web browser.
[3] (2014)	Method for detecting phishing web pages using classification.	• New categorization of characteristics.		• Hidden knowledge extraction from the model for integration into a browser extension.

2.5 Discussion

After analyzing the table 2 and the figure 1, we notice that :

- The size of the data set in the 3 works is different.
- Solutions proposed in [1], [2] and [3]:
  - have used features that are not based on search engines, because only popular sites appear in the first search results.
  - used independent language characteristics.
  - can detect the "Zero Hour" attack unlike most approaches that cannot detect it because they are designed to detect a particular type of phishing website, like in : [5] and [6].
- Solutions [2], [3] are in real time, fast, intelligent and essential because most of the existing methods based on machine learning extract "3rd party" characteristics, such as: WHOIS lookup, DNS, etc.
- In [2], the RL algorithm outperformed other algorithms with accuracy up to 98.42% with larger data size than that used in [1] which obtained an accuracy of 98,65%. SVM dans [1] outperformed SVM in [2] and RL. In [3], authors didn't give the exact figures but they mentioned that the accuracy of all the algorithms was increased and the CBA algorithm outperformed the other used algorithms.

  It can be seen that the work [3] marked the highest error rate and the lowest accuracy compared to the other presented works.

3 SPAM DETECTION

Spam has become a real economic scourge. Spam can be defined as an electronic communication sent automatically, unwanted, unsolicited, with unwanted content, received without the full consent of the recipient. Spam can take many forms; Spam in email, spam in SMS, spam in reviews, etc.

Currently, researchers are mainly focused on the use of machine learning techniques to detect spam, among these works we find: [14], [15] [16] and [17] in which the authors have proposed approaches to detect spam in different forms. In this section we mainly expose the works [14],[15] and [16].

3.1 An Approach for Detecting Spam in Arabic Opinion Reviews [14]

In [14], the authors have proposed a method for detecting opinion spam only in Arabic by combining the methods "data mining" and "text mining". Data were collected from online economic websites in

Table 2: Comparative Table of the Anti-Phishing Approaches

Paper	Data	Algorihms	Error Rate (%)	Accuracy (%)	Search engine independence	Language independence	"Zero Hour" Detection	"3rd Party Services" independence
[1] (2016)	Yahoo directory service Phish Tank No. Of legt.P 549 No. Of Phi.P : 1158 Total :1707	SVM	1.35	98.65	Yes	Yes	Yes	Yes
[2] (2018)	Phishtank dataset (2018). Alexa top websites (2018) Stuffgate Free Online Website Analyzer (2018) Liste des prestataires de services de paiement en ligne(2018) No. Of legt.P 1116 No. Of Phi.P : 1428 Total :2544	SMO NB RF SVM Adaboost ANN C4.5 RL	3.11 4.21 2.63 8.53 4.17 2.75 2.71 1.58	96.89 95.79 97.37 91.47 95.83 97.25 97.29 98.42	Yes	Yes	Yes	Yes
[3] (2014)	Phishtank Millersmiles yahoo directory No. Of legt.P 450 No. Of Phi.P : 450 Total :900	C4.5 RIPPER PRISM CBA	- - - 4.75	- - - 95.25	Yes	Yes	Yes	No

Arabic; tripadvisor.com.eg, booking.com and agoda.ae and analyzed manually. They used a combination of features regarding review content, reviewer, and hotels.

To evaluate the method, the authors carried out three experiments: Classify the reviews using only the data mining techniques on all attributes except the review content attribute, classify the reviews using only the text mining techniques on the review content and classify the review by combining the two.

The method has been evaluated with three classifiers: NB, K-NN (K-Nearest Neighbors) [**?**] and SVM using 10-fold cross-validation, the experimental results have shown that the combination of data mining and text mining methods is very useful.

3.2 Optimizing Semantic LSTM for Spam detection [15]

In [15], the authors implemented a special architecture using the LSTM (Long Short Term Memory) [**?**] for detecting spam in the form of short texts.

The proposed approach is LSTM with a semantic layer added, it is a layer for preprocessing before using the classifier whose text is converted into semantic word vectors using word2vec, WordNet and ConceptNet, after they have introduces the vector as input into the LSTM unit with the output of the previous LSTM unit. This is repeated with each input sentence and in this way the LSTM units continue to record the important characteristics. Consequently, each LSTM layer will cause the output sequence on each layer s_1, s_2,... s_n. Finally, the softmax function is applied to s_t, the output of the last LSTM unit to obtain a class label.

Two sets of data were used to assess the results: SMS Spam Collection dataset and Twitter dataset. The classification results obtained are compared with those obtained by SVM, NB, ANN, k-NN and RF. The results of the experiments have shown that the precision of SLSTM is the highest compared to other classifiers thanks to the semantic representation and the sequential processing of the text using LSTM.

The approach proposed in [17] is based on the same principle as [15], the difference is that LSTM is able of capturing a long-term dependence in the characteristics of the text, while the SCNN (Semantic Convolutional Neural Network) is able of detect short-term correlations and temporal characteristics in texts.

3.3 A Support Vector Machine based Naive Bayes algorithm for Spam Filtering [16]

In [16], the authors proposed a spam filtering system, they took the advantages of the NB and SVM algorithms and made a combination to propose the SVM-NB system. The NB classifier is widely used in the spam detection field because of its classification speed, but its performance is limited by strong independence, which is not the case in real messaging systems. For the improvement of spam detection systems it is necessary to remove the dependency, then the authors have gone through 4 steps:

- Step 1: the NB Classifier has classified the training set into two categories: spam and non-spam.
- Step 2: SVM has built a hyperplane to divide the two categories.

- Step 3: Verification of the results with elimination of any sample whose result from its nearest neighbor is different.
- Step 4: the NB Classifier is used for the second time to classify the new training set and detect spam.

To validate and evaluate SVM-NB, experiments were carried out with the DATAMALL data set. The SVM-NB system offers better results compared to the NB and SVM used individually.

3.4 Summary and Comparison

Table 3 summarizes the work done for spam detection using machine learning, and table 4 summarizes the comparison between these works.

3.5 Discussion

After analysing the table 4 and figures 2 and 3, we note that:

- The size of the data set used by the cited works is different.
- In [14], the data between the two classes (HAM and SPAM) are balanced with the Over-Sampling method (Duplicate samples from the minority class and add them to the data set), as already mentioned in the article. The distribution of data between classes should be balanced or almost balanced for better classification because otherwise such a classifier predicts that the samples have the majority class and completely ignore the minority class. The NB algorithm gave the best Accuracy and F1 score.
- In [15] and [17], SLSTM et SCNN have surpassed traditional techniques on both data sets so the semantic layer is an added value. The SLSTM algorithm has surpassed SCNN therefore the sequential processing of the text using LSTM helps to improve the results and is effective for data linked to social networks because the sequential processing can include slang expressions and new words. SLSTM and SCNN performed better in the SMS dataset than the Twitter dataset. The algorithms: SVM, NB, K-NN, RF and ANN obtained similar results in the two works [15] and [17].
- Algorithms: SVM, NB and K-NN used in [14] show more results than in [15].
- In [16], the authors did not mention the exact figures but in general the work scored the lowest accuracy and F1 score compared to the other works presented.

4 CONCLUSION

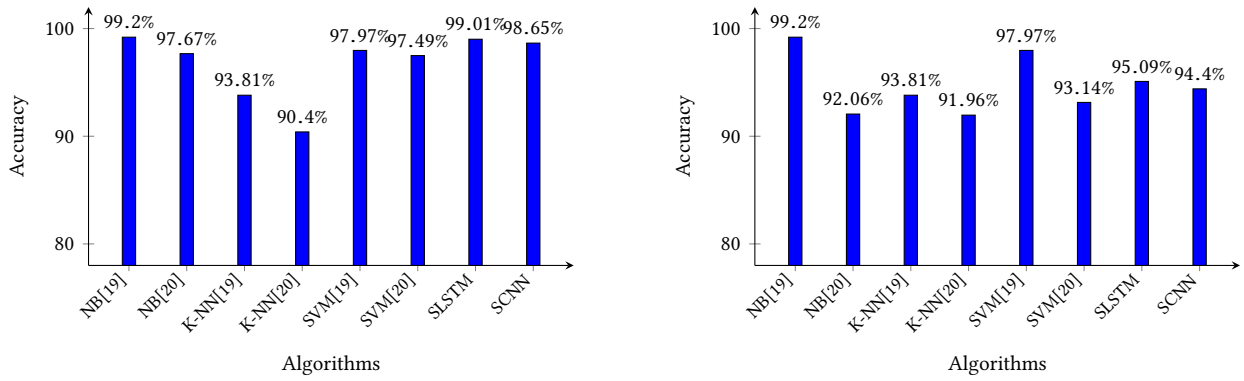
In this paper we have summarized and compared the work based on the use of machine learning to detect Phishing and Spam. The results of the comparative approaches demonstrate the effectiveness of machine learning in network security. In addition, there is still open research and future work on the use of machine learning to improve network security. We have presented work related to phishing and spam only, and there are other types of attacks that we will study in a future work such as machine learning used to detect malware attacks.

REFERENCES

[1] M. Mahmood and V. A. Yazdian, "New Rule-Based Phishing Detection Method," Expert Systems With Applications, 2016.
[2] A. K. Jain and B. B. Gupta, "A machine learning based approach for phishing detection using hyperlinks information," Ambient Intelligence and Humanized Computing, 2018.
[3] M. R. M., T. Fadi, and M. Lee, "Intelligent rule-based phishing websites classification," IET Information Security, 2014.
[4] A. A. Akinyelu and A. O. Adewumi, "Classification of phishing email using random forest machine learning technique," Journal of Applied Mathematics, vol. 2014, no. 425731, 2014.
[5] W. Zhang1, Q. Jiang1, L. Chen, and C. Li, "Two-stage elm for phishing web pages detection using hybrid features,"World Wide Web,vol.20(4),no.425731,p.797–813, 2016.
[6] M. Aburrousa, M.A.Hossaina, K. Dahala, and F. Thabtahb, "Intelligent phishing detection system for e-banking using fuzzy data mining," Expert Systems with Applications, vol. 37, pp. 7913–7921, 2010.
[7] G. Ali, M. Sara, and A. Yarmohammadib, "Detection of phishing attacks in Iranian e-banking using a fuzzy–rough hybrid system," Applied Soft Computing, vol. 35, pp. 482–492, 2015.
[8] S.L.VandV.MS,"Efficient prediction of phishing websites using supervised learning algorithms," Procedia Engineering,, vol. 30, pp. 798–805, 2012.
[9] M. He, S.-J. Horng, P. Fan, M. K. Khan, R.-S. Run, J.-L. Lai, R.-J. Chen, and A. Sutanto, "An efficient phishing web page detector," Expert Systems with Applications, p. 12018–12027, 2011.
[10] M.Aburrous, M.A.Hossain, K.Dahal, and F.Thabatah,"Intelligent phishing detection system fore-banking using fuzzy data mining,"Expert Systems with Applications, vol. 20(4), no. 425731, pp. 7913–7921, 2010.
[11] G. XIANG, J. HONG, C. P. ROSE, and L. CRANOR, "Cantina+ : A feature-rich machine learning framework for detecting phishing web sites," ACM Transactions on Information and System Security, 2011.
[12] C. Whittaker, B. Ryner, and M. Nazif, "Large-scale automatic classification of phishing pages," In :Proceedings of the network and distributed system security symposium, San Diego, p. 1–14, 2010.
[13] R. M. Mohammad, F. Thabtah, and L. McCluskey, "Predicting phishing websites based on self-structuring neural network," Neural Comput and Applic, p. 443–458, 2014.
[14] H. A. Abu and E.-H. Alaa, "An Approach for Detecting Spam in Arabic," The International Arab Journal of Information Technology, 2014.
[15] J. Gauri, S. Manisha, and A. Basant, "Optimizing semantic LSTM for spam detection," International Journal of Information Technology, 2018.
[16] F. Weimiao, S. Jianguo, Z. Liguo, C. Cuiling, and Y. Qing, "A Support Vector Machine based Naive Bayes Algorithm for Spam Filtering," IEEE 35th International Performance Computing and Communications Conference (IPCCC), 2016.
[17] G. Jain, M. Sharma, and B. Agarwal, "Spam detection on social media using semantic convolutional neural network," International Journal of Knowledge Discovery in Bioinformatics, vol. 8(1), p. 443–458, 2018.

Table 3: Summary of ML-based methods for SPAM Detection

Paper	Method	Contribution	Limits	future work
[14] (2014)	Method for detecting opinion spam combining two techniques : « data / text mining » in a single classification method.	• First solution for detecting notice spam in Arabic	• Spam detection in Arabic only	• Use extended data with more distinctive attributes. • Detect other types of spam (web spam, Email spam, etc.) and in different languages. • Filter other content submitted by users.
[15] (2018)	LSTM with a semantic layer to detect short texts.	• Able to learn the abstract features and choose the best information to move it to the next layer and store the information for a long time. • Sequential processing of input words while correlating with past words. • The semantic representation.	• Spam with different data formats	• Increase the volume of data for training. • Find optimized LSTM settings automatically. • Detect other types of spam. • Use pre-trained word vectors like Word2Vec.
[16] (2016)	Method for detecting spam by combining SVM and NB	• Reduced dependence between samples.	• Spam with different data formats	• Detection of spam emails containing various data formats.

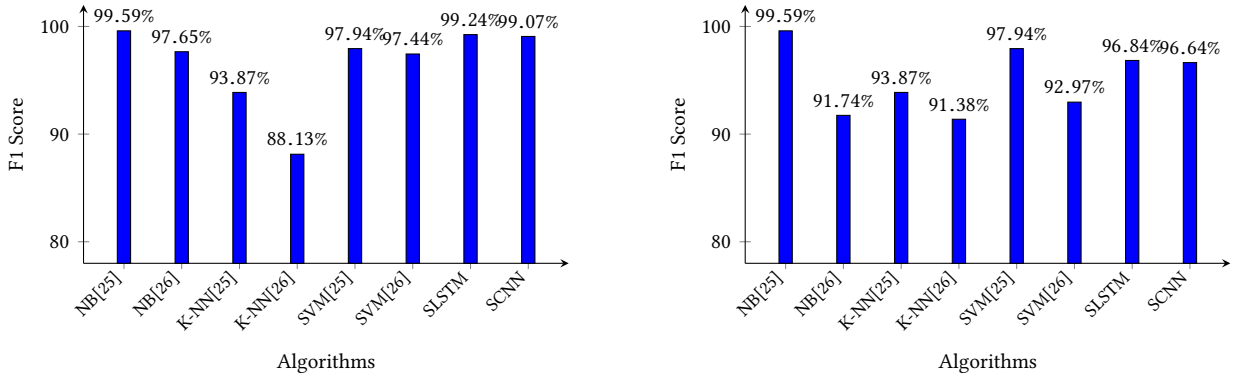


(a) with SMS Dataset **(b) with Twitter Dataset**

Figure 2: Accuracy Comparison of Anti-Spam Approaches

Table 4: Comparative Table of the Anti-Spam Approaches

Paper	features selection	Data	Algorithms	F1 Score (%)	Accuracy (%)
[14] (2014)	Yes	Collected from TripAdvisor/Booking/Agoda **Ham :** 2469 **Spam:** 379 **Total:** 2848	NB K-NN SVM	99.59 93.87 97.94	99.20 93.81 97.97
[15] (2018)	No	SMS spam Collection Twitter dataset **Ham SMS :** 4827 **Spam sms :** 747 **Ham twitter :** 4231 **Spam twitter :** 865 **Total :** 10607	SLSTM SVM NB ANN k-NN RF	SMS: 99.24 Twitter: 96.84 SMS: 97.44 Twitter: 92.97 SMS: 97.65 Twitter: 91.74 SMS: 97.40 Twitter: 91.41 SMS: 88.13 Twitter: 91.38 SMS: 97.77 Twitter: 93.04	SMS: 99.01 Twitter: 95.09 SMS: 97.49 Twitter: 93.14 SMS: 97.67 Twitter: 92.06 SMS: 97.40 Twitter : 91.18 SMS: 90.40 Twitter : 91.96 SMS: 97.85 Twitter: 93.43
[17] (2018)	No	SMS spam Collection Twitter dataset **Ham SMS :** 4827 **Spam sms :** 747 **Ham twitter :** 4231 **Spam twitter :** 865 **Total :** 10607	SCNN SVM NB ANN k-NN RF	SMS: 99.07 Twitter: 96.64 SMS: 97.44 Twitter: 92.97 SMS: 97.65 Twitter: 91.74 SMS: 97.40 Twitter: 91.41 SMS: 88.13 Twitter: 91.38 SMS: 97.77 Twitter: 93.04	SMS: 98.65 Twitter : 94.40 SMS: 97.49 Twitter : 93.14 SMS: 97.67 Twitter: 92.06 SMS: 97.40 Twitter: 91.18 SMS: 90.40 Twitter: 91.96 SMS: 97.85 Twitter: 93.43
[16] (2016)	Yes	DATAMALL **Ham:** 4000 **Spam:** 4000 **Total :** 8000	NB SVM NB-SVM	- - almost 89	- - almost 93



(a) with SMS Dataset **(b) with Twitter Dataset**

Figure 3: F1 Score Comparison of Anti-Spam Approaches