As per UGC guidelines an electronic bar code is provided to seure your paper

International Journal for Modern Trends in Science and Technology, 7(01): 05-09, 2021

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0701002
Available online at: http://www.ijmtst.com/vol7issue01.html

Analysis of Naıve Bayes Algorithm for Email Spam Filtering

RajKishore Sahni

Information Technology, Maharaja Agrasen Institute of Technology, Rohini, Delhi

To Cite this Article
RajKishore Sahni, "Analysis of Naıve Bayes Algorithm for Email Spam Filtering", *International Journal for Modern Trends in Science and Technology*, Vol. 07, Issue 01, January 2021, pp.-05-09.

Article Info
Received on 22-November-2020, Revised on 18-December-2020, Accepted on 22-December-2020, Published on 28-December-2020.

ABSTRACT

The upsurge in the volume of unwanted emails called spam has created an intense need for the development of more dependable and robust antispam filters. Machine learning methods of recent are being used to successfully detect and filter spam emails. We present a systematic review of some of the popular machine learning based email spam filtering approaches. Our review covers survey of the important concepts, attempts, efficiency, and the research trend in spam filtering. The preliminary discussion in the study background examines the applications of machine learning techniques to the email spam filtering process of the leading internet service providers (ISPs) like Gmail, Yahoo and Outlook emails spam filters. Discussion on general email spam filtering process, and the various efforts by different researchers in combating spam through the use machine learning techniques was done. Our review compares the strengths and drawbacks of existing machine learning approaches and the open research problems in spam filtering. We recommended deep learning and deep adversarial learning as the future techniques that can effectively handle the menace of spam emails

KEYWORDS: *spam filtering,spam detection ,naïve bayes classifier,spam,ham*

I. INTRODUCTION

Nowadays, e-mail provides many ways to send millions of advertisement at no cost to sender. As a result, many unsolicited bulk e-mail, also known as spam e-mail spread widely and become serious threat to not only the Internet but also to society. For example, when user received large amount of email spam, the chance of the user forgot to read a non-spam message increase. As a result, many e-mail readers have to spend their time removing unwanted messages. E-mail spam also may cost money to users with dial-up connections, waste bandwidth, and may expose minors to unsuitable content. Over the past many years, many approaches have been provided to block e-mail spam

For filtering, some email spam are not being labelled as spam because the e-mail filtering does not detect that email as spam. Some existing problems are regarding accuracy for email spam filtering that might introduce some error. Several machine learning algorithms have been used in spam e-mail filtering, but Na¨ıve Bayes algorithm is particularly popular in commercial and open-source spam filters . This is because of its simplicity, which make them easy to implement and just need short training time or fast evaluation to filter email spam. The filter requires training that can be provided by a previous set of spam and

non-spam messages. It keeps track of each word that occurs only in spam, in non-spam messages, and in both. Naıve Bayes can be used in different datasets where each of them has different features and attribute.

The research objectives are: (i) to implement the Na¨ıve Bayes algorithm for e-mail spam filtering on two datasets, (ii) to evaluate the performance of Na¨ıve Bayes algorithm for e-mail spam filtering on the chosen dataset.

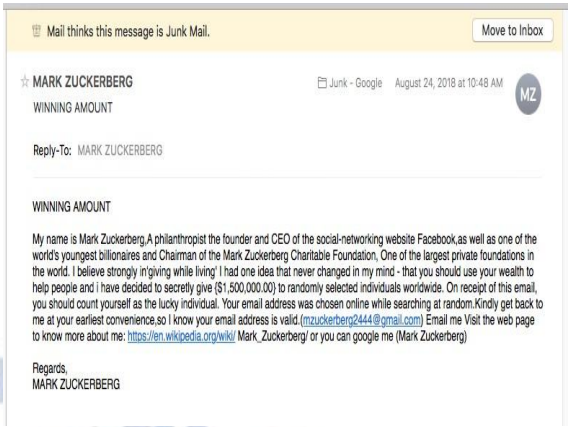
The rest of the paper is organized as follows: Section II describes the related work on Na¨ıve Bayes algorithm for e-mail spam filtering. Section III presents the methodology process of e-mail spam Section IV presents the experimental setup. Section V shows the result and analysis on two datasets. Finally, Section VI concludes the work and highlights the direction for future research. .

OBJECTIVES

A spam filter is a program that is used to detect unsolicited and unwanted email and prevent those messages from getting to a user's inbox. Like other types of filtering programs, a spam filter looks for certain criteria on which it bases judgments. For example, the simplest and earliest versions (such as the one available with Microsoft's Hotmail) can be set to watch for particular words in the subject line of messages and to exclude these from the user's inbox. This method is not especially effective, too often omitting perfectly legitimate messages (these are called *false positives*) and letting actual spam through. More sophisticated programs, such as Bayesian filters or other heuristic filters, attempt to identify spam through suspicious word patterns or word frequency.

II. RELATED WORK

Spammers are now able to launch large scale spam campaigns, malware and botnets helped spammers to spread spam widely. Upon receiving and opening a spam email, Internet users is exposed to security issues as spams are normally broadcast for bad intention. One of the common email spam example received by users are an email requesting for IDs and passwords(Refer to


Mail thinks this message is Junk Mail.

Move to Inbox

MARK ZUCKERBERG

Junk - Google August 24, 2018 at 10:48 AM

MZ

WINNING AMOUNT

Reply-To: MARK ZUCKERBERG

WINNING AMOUNT

My name is Mark Zuckerberg,A philanthropist the founder and CEO of the social-networking website Facebook,as well as one of the world's youngest billionaires and Chairman of the Mark Zuckerberg Charitable Foundation, One of the largest private foundations in the world. I believe strongly in'giving while living' I had one idea that never changed in my mind - that you should use your wealth to help people and i have decided to secretly give ($1,500,000.00) to randomly selected individuals worldwide. On receipt of this email, you should count yourself as the lucky individual. Your email address was chosen online while searching at random.Kindly get back to me at your earliest convenience,so I know your email address is valid.(mzuckerberg2444@gmail.com) Email me Visit the web page to know more about me: https://en.wikipedia.org/wiki/ Mark_Zuckerberg/ or you can google me (Mark Zuckerberg)

Regards,
MARK ZUCKERBERG

Figure 1. Sample of spam data requesting for ID and password

Several machine learning algorithms have been employed in anti-spam e-mail spam filtering, including algorithms that are considered top-performers in Text Classification like Boosting algorithm, Support Vector Machines (SVM) algorithm and Na¨ıve Bayes algorithm

There is a rapid increase in the interest being shown by the global research community on email spam filtering. In this section, we present similar reviews that have been presented in the literature in this domain. This method is followed so as to articulate the issues that are yet to be addressed and to highlight the differences with our current review. Lueg presented a brief survey to explore the gaps in whether information filtering and information retrieval technology can be applied to postulate Email spam detection in a logical, theoretically grounded manner, in order to facilitate the introduction of spam filtering technique that could be operational in an efficient way. However, the survey did not present the details of the Machine learning algorithms

III. METHODOLOGY

This section describes the methodology that is used for the research. The methodology that is used for the filtering method is machine learning techniques that divide by three phases.The methodology is used for the process of e-mail spam filtering based on Na¨ıve Bayes algorithm.

A.3.1. Na¨ıve Bayes classifier

The Na¨ıve Bayes algorithm is a simple probabilistic classifier that calculates a set of probabilities by counting the frequency and combination of values in a given dataset [4]. In this research, Na¨ıve Bayes classifier use bag of words features to identify spam e-mail and a text is

representing as the bag of its word. The bag of words is always used in methods of document classification, where the frequency of occurrence of each word is used as a feature for training classifier. This bag of words features are included in the chosen datasets.

Na¨ıve Bayes technique used Bayes theorem to determine that probabilities spam e-mail. Some words have particular probabilities of occurring in spam e-mail or non-spam e-mail. Example, suppose that we know exactly, that the word Free could never occur in a non-spam e-mail. Then, when we saw a message containing this word, we could tell for sure that were spam email. Bayesian spam filters have learned a very high spam probability for the words such as Free and Viagra, but a very low spam probability for words seen in non-spam e-mail, such as the names of friend and family member. So, to calculate the probability that e-mail is spam or non-spam Na¨ıve Bayes technique used Bayes theorem as shown in formula below.

$$P(spam \mid word) = \frac{P(\text{spam}).P(\text{word}|\text{spam})}{P(\text{spam}).P(\text{word}|\text{spam}) + P(\text{non}-\text{spam}).P(\text{word}|\text{non}-\text{spam})}$$

Where:

(i) *P*(*spamword*) is probability that an e-mail has particular word given the e-mail is spam.

(ii) *P*(*spam*) is probability that any given message is spam.

(iii) *P*(*wordspam*) is probability that the particular word appears in spam message.

(iv) *P*(*non − spam*) is the probability that any particular word is not spam.

(v) *P*(*wordnon − spam*) is the probability that the particular word appears in non-spam message.

To achieve the objective, the research and procedure is conducted in three phases. The phases involved are as follows:

(i) Phase 1: Pre-processing

(ii) Phase 2: Feature Selection

(iii) Phase 3: Na¨ıve Bayes Classifier

The following sections will explain the activities that involve in each phases in order to develop this project. Figure 2 shows the process for e-mail spam filtering based on Na¨ıve Bayes algorithm.

B. 3.2. Pre-processing

Today, most of the data in the real world are incomplete containing aggregate, noisy and missing values [9]. Pre-processing of e-mails in next step of training filter, some words like conjunction words, articles are removed from email body because those words are not useful in classification.

(Refer to Figure 3 for sample of data).

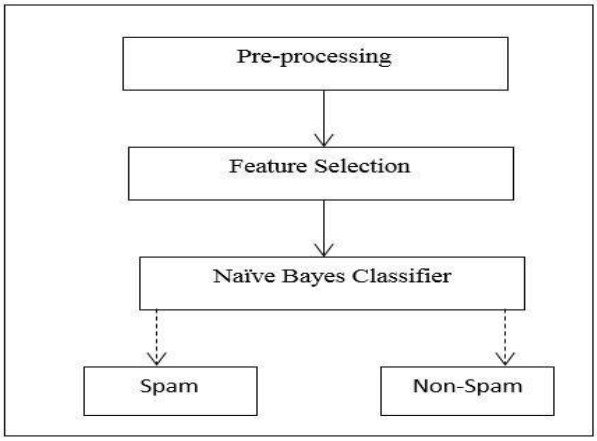


Figure 2. Process of E-mail spam filtering based on Na¨ıve Bayes Algorithm

```
examples[0]

'I am Dr. Bakare Tunde, the cousin of Nigerian Astronaut, Air Force Major Abacha Tunde. He was the first African in space when
he made a secret flight to the Salyut 6 space station in 1979. He was on a later Soviet spaceflight, Soyuz T-16Z to the secret
Soviet military space station Salyut 8T in 1989. He was stranded there in 1990 when the Soviet Union was dissolved. His other S
oviet crew members returned to earth on the Soyuz T-16Z, but his place was taken up by return cargo. There have been occasional
Progrez supply flights to keep him going since that time. He is in good humor, but wants to come home. In the 14-years since he
has been on the station, he has accumulated flight pay and interest amounting to almost $ 15,000,000 American Dollars. This is
held in a trust at the Lagos National Savings and Trust Association. If we can obtain access to this money, we can place a down
payment with the Russian Space Authorities for a Soyuz return flight to bring him back to Earth. I am told this will cost $ 3,0
00,000 American Dollars. In order to access the his trust fund we need your assistance. Consequently, my colleagues and I are w
illing to transfer the total amount to your account or subsequent disbursement, since we as civil servants are prohibited by th
e Code of Conduct Bureau (Civil Service Laws) from opening and/ or operating foreign accounts in our names. Needless to say, th
e trust reposed on you at this juncture is enormous. In return, we have agreed to offer you 20 percent of the transferred sum,
while 10 percent shall be set aside for incidental expenses (internal and external) between the parties in the course of the tr
ansaction. You will be mandated to remit the balance 70 percent to other accounts in due course.'

examples[1]

'Hi Bob, lets have our meeting discussing Knicks unfortunate draft lottery'
```

Figure 3. Sample of spam datafragment

A full list of the attributes in this data set appears in the ”Attributes” frame as shown in Figure 4. Random selection of attribute are performed for the further process.

Attributes *capital run length _ average*,*capital run length longest* and *capital run length total* are removed from the list by checking the box to their left and hitting the Remove button.

C. 3.3. Feature Selection

After the pre-processing step, we apply the feature selection algorithm, the algorithm which deploy here is Best First Feature Selection algorithm

IV. EXPERIMENTAL SETUP

The experimental setting of the research is like follows:

D. 4.1. The Evaluation Metric

Evaluation metrics are used to evaluate the performance based on datasetth at had been chosen. The most simple measure is filtering accuracy namely percentage of

```
[4] df.head()
```

	text	spam
0	Subject: naturally irresistible your corporate...	1
1	Subject: the stock trading gunslinger fanny i...	1
2	Subject: unbelievable new homes made easy im ...	1
3	Subject: 4 color printing special request add...	1
4	Subject: do not have money , get software cds ...	1

Figure 4. Sample list of the attributes in the "Attributes" frame

Evaluation measures for spam filters

Evaluation Measure	Evaluation Function
Accuracy	$Acc=\frac{TP+FN}{TN+FP+TP+TN}$
Recall	$r=\frac{TP}{TP+FN}$
Precision	$P=\frac{TP}{TP+FP}$
F-measure	$F=\frac{p^2}{pr+r}$

Where accuracy, recall, precision, F-measure, FP, FN, TP and TN are defined as follows:

(i) Accuracy: Percentage of correctly identified spam and not spam message

(ii) Recall: Percentage spam message manage to block

(iii) Precision: Percentage of correct message for spam e-mail

(iv) F-measure: Weighted average of precision and recall

(v) False Positive Rate (FP): The number of misclassified non spam emails

(vi) False Negative Rate (FN): The number of misclassified spam emails

(vii) True Positive (TP): The number of spam messages are correctly classified as spam

(viii) True Negative (TN): The number of non-spam e-mail that is correctly classified as non-spam

E.4.2. Dataset

Dataset is a collection of data or related information that is composed for separate elements. A collection of dataset for e-mail spam contains spam and non-spam messages. In this research, two datasets are be used to evaluate the performance of Na¨ıve Bayes algorithm to filter e-mail spam.

SPAMBASE was taken from kaggle This dataset contains 5728 email messages and 2 attributes. This dataset collection of non-spam email came from filled work, personal e-mail and single e-mail account. This dataset is composed of spam and non spam emails .this dataset is not that large but tis works fine on this to

V. RESULT AND DISCUSSIONS

This section discussed the experimental result by utilising google colab tool using Na¨ıve Bayes algorithm. The datasets are compared based on the percentage of correctly identified spam and nonspam message, percentage of spam message manage to block, percentage of correct message for spam e-mail and weighted average of precision and recall.

Below is the result we got after using baysian classifier on Kaggle dataset let for running multiple times and below is the avegare value

For FP, FN, TP and TN,

- FP: Total 1 number of misclassified non spam emails
- FN: 12 Total number of misclassified spam emails
- TP: Total 3445 number of spam messages are correctly classified as spam
- TN: 1098 Total number of non-spam e-mail that is correctly classified as non-spam

```
              precision    recall  f1-score   support

           0       1.00      0.99      0.99       870
           1       0.97      1.00      0.98       269

    accuracy                           0.99      1139
   macro avg       0.98      0.99      0.99      1139
weighted avg       0.99      0.99      0.99      1139


confusion matrix:
 [[862   8]
 [  1 268]]

Accuracy:
 0.9920983318700615
```

```
[ ]
```

In order to use your classifier, you must vectorize the example emails. Finally, you can classify the emails. For our examples above, we got the

following results with the first email classified as 'spam' and the second as 'not spam'.
The classifier was successful. In order to get a better understanding of the performance of the model, the accuracy and F1 score was measured. There were also a total of only 22 false positives and false negatives for a testing set with 1293 emails
The average results we got after multiple test run is 98 percent on the same dataset

```
Spam Filter Accuracy Score:  0.982985305491106
Spam Filter F1 Score:  0.9705093833780161
```

Figure 8. Average F-measure result for 10 runs of experiment
dataset only has 4601 instance e-mails and 58 attributes, but Na¨ıve Bayes classifier manage to get good result from this dataset. This is because Na¨ıve Bayes classifier no needs many instances of e-mails and attributes to train the classifier for e-mail spam filtering.

SPAMBASE dataset is multivariate dataset contain data from a single e-mail account while Spam Data dataset collect from many e-mail account. From this we can know, Na¨ıve Bayes classifier can perform well with dataset that come from a single e-mail account than many email account. This is because Na¨ıve Bayes classifier can focus train with various type of e-mail spam that come from single e-mail account located on same e-mail servers.
Besides that, we also test if these datasets have same total of attribute and minimizing the attribute with six attribute, wether it gives different result. From what we get, the result not given different result because Na¨ıve Bayes classifier that used SPAMBASE dataset still has the best performance than Spam Data dataset. But the percentage for accuracy, precision, recall and F-measure drop a little bit. This is proving that total of attribute give important role to Na¨ıve Bayes classifier for filtering e-mail spam.

Although minimizing of attribute decreased the performance of Na¨ıve Bayes classifier, but in the other hand it improved time complexity. The time taken to build model is faster with less attribute and the best performance Na¨ıve Bayes classifier that used SPAMBASE dataset need 0.14 second to build the model. However, it is not important because e-mail spam filtering must have highest accuracy, precision, recall and F-measure to filter that e-mail spam or non-spam.

VI. CONCLUSION

In this study, we reviewed machine learning approaches and their application to the field of spam filtering. A review of the state of the art algorithms been applied for classification of messages as either spam or ham is provided. The attempts made by different researchers to solving the problem of spam through the use of machine learning classifiers was discussed. The evolution of spam messages over the years to evade filters was examined.

E-mail spam filtering is an important issue in the network security and machine learning techniques; Na¨ıve Bayes classifier that used has a very important role in this process of filtering e-mail spam. The quality of performance Na¨ıve Bayes classifier isalso based on datasets that used. As can see, dataset that have fewer instances of e-mails and attributes can give good performance for Na¨ıve Bayes classifier. Na¨ıve Bayes classifier also can get highest precision that give highest percentage spam message manage to block if the dataset collect from single email accounts. Naive Bayes is powerful yet simple algorithm that is especially useful when filtering out spam emails

REFERENCES

[1] Rushdi, S. and Robet, M, "Classification spam emails using text and readability features", *IEEE 13th International Conference on Data Mining*, 2013.

[2] Androutsopoulos, I., Paliouras, G., and Michelakis, "E. Learning to filter unsolicited commercial e-mail", *Technical report NCSR Demokritos*, 2011.

[3] Rathi, M. and Pareek, V. "Spam Mail Detection through Data Mining A Comparative Performance Analysis", *I.J. Modern Education and Computer Science*, 2013, 12, 31-39.

[4] Patil, T. and Sherekar, S. "Performance Analysis of Na¨ıve Bayes and Classification Algorithm for Data Classification", *International Journal Of Computer Science And Applications*, 2013.

[5] Kumar, S., Gao, X., Welch, I. and Mansoori, M., "A Machine Learning Based Web Spam Filtering Approach", IEEE 30th International Conference on Advanced Information Networking and Applications (AINA), Crans-Montana, 2016, pp. 973-980.

[6] Tariq, M., B., Jameel A. Tariq, Q., Jan, R. Nisar, A. S., "Detecting Threat E-mails using Bayesian Approach", *IJSDIA International Journal of Secure Digital Information Age*, Vol. 1. No. 2, December 2009.

[7]. www.wikepedia.org

[8]. D.M. Fonseca, O.H. Fazzion, E. Cunha,I. Las-Casas, P.D. Guedes, W. Meira, M. ChavesMeasuring characterizing, and avoiding spam traffic costs

[9]. M. Awad, M. FoqahaEmail spam classification using hybrid approach of RBF neural network and particle swarm optimization

[10]. www.researchgate.net