JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

WhatsApp Chat Analysis and Spam Message Detection

Mothe Vikas Reddy
Undergraduate Scholar
Department of Information Technology
Anurag University

Arvapelly Ashrith
Undergraduate Scholar
Department of Information Technology
Anurag University

Vunnam Harish
Undergraduate Scholar
Department of Information Technology
Anurag University

Under the guidance of

Mr. B. Pruthviraj Goud
Assistant Professor
Department of Information Technology
Anurag University

***Abstract*— This paper presents Analysis of WhatsApp chat and detection of Spam and ham messages using various supervised machine learning algorithms like naïve Bayes Algorithm, support vector machines algorithm, and the maximum entropy algorithm and compares their performance in filtering the Ham and Spam messages. WhatsApp chats consists of various kinds of conversations held among group of people which consists of various topics. This web application aims to provide in depth analysis of the chat among users in particular group and two individuals. Irrespective of whichever topic the conversation is based this web application can be applied to detect spam messages as well.**

I.INTRODUCTION

In the developing period of the Internet, individuals are involving increasingly in free online services. Individuals tend to share their data on different sites, though that data is imparted to different organizations that spam individuals to offer their services.

WhatsApp chat Analyzer is an analyzing tool for the WhatsApp chats. The chat files can be exported from WhatsApp and it generates various plots and graphs showing, number of messages oremojis or images sent by a person, most active member in the group etc.,

SMS Spamming [2] [10] in extremely disappointing for the clients: numerous critical and valuable messages can get lost because of spam messages, Spam messages are additionally used to trap individuals, or bait them into purchasing services. As overall utilization of cell phones has grown, another road for e-junk mail has been opened for notorious advertisers. These publicists use instant messages (SMS) to target probable purchasers with undesirable publicizing known as SMS spam. This sort of spam is especially bothersome since, not at all like email spam, numerous PDA clients pay an expense for each SMS got.

Building up a classification algorithm [1] [11] that channels SMS spam would give a helpful apparatus for mobile phone suppliers.

Since naïve Bayes has been utilized effectively for email spam detection [9], it appears to be expected that it could likewise be used to build SMS spam classifier [7]. With respect to email spam [6][8], SMS spam represents extra difficulties for automated channels. SMS texts are regularly restricted to 160 characters, lessening the measure of content that can be utilized to distinguish whether a message is a ham or spam.

People have also regularly started using shorthand notations and slang which further makes it difficult to distinguish between ham and spam. We will test how well a simple naïve Bayes classifier [4] manages these difficulties.

We additionally fabricate models to group messages utilizing the SVM algorithm and the maximum entropy algorithm [3], and it is discovered that SVM gives us the most precise outcomes, with exactness up to 98 %, took after by Naïve bayes algorithm, followed by maximum entropy algorithm.

Spam messages can be classified as redundant messages sent to large number of people at once. The rise of spam messages are based on the following factors: 1) The accessibility to cheap bulk SMS-plans; 2) dependability (since the message comes to the cell phone client); 3) low possibility of accepting reactions from some unaware recipients; and 4) the message can be customized.5) Free services.

II.BACKGROUND STUDY

To construct the naïve Bayes classifier [4], we will use information and data collected from the SMS Spam collection which is available openly and consists of about 5574records
[5].

This dataset incorporates the content of SMS messages alongside a label signifying if the message is a ham or a spam. Junk messages are marked as spam, while true blue messages are marked as ham. A few cases of spam (Table 2) and ham (Table 1) are illustrated in the following illustration:

1. HAM MESSAGES

Draft a reasonable one. And I will see if something can happen.
Okay I can try, but cannot commit.
I am good too. Yes weekdays are busy, all thanks to office.

Table 1: Ham messages

As watched these messages are the everyday messages that individuals trade with each other, these are not junk messages and the client ought to get these messages with the spam filter
not separating them through.

2. SPAM MESSAGES

Post Diwali offer! Get 30% off + Free Cloudbar with select LED. Buy with your pre-approved loan.
Hi, good credit score makes you eligible for top loans & credit cards. Get your score in 3 minutes.
Want chocolate? Get a whole-some Chocolate Shake free on orders above Rs. 2000.

Table 2: Spam Messages

Taking a gander at the former specimen messages, we see some recognizing qualities or some repeated patterns of spam messages. One remarkable identification is that two of the three spam messages use the word "free", yet the same word (free) does not show up in any of the ham messages. Then again, two of the ham messages refer to particular days of week, at the point when contrasted with zero junk messages.
Our classifiers will exploit such examples in the word recurrence to decide if the SMS messages appear to better fit the profile of spam or ham. While it's not incomprehensible that "free" would show up outside of a spam SMS, a ham message is probably going to give extra words giving setting.

For example, a ham message may state "are you free on
Saturday?", while a spam message may utilize the expression "free melodies and ringtones." The classifier will figure the likelihood of spam and ham given the confirmation gave by every one of the words in the message.
We have a total of 5574 records, out of which 4827 messages are ham and 747 messages are spam (Chart 1).

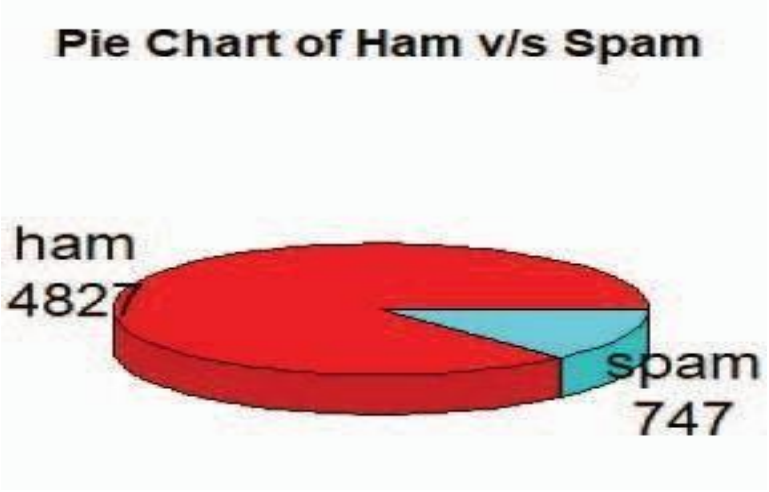
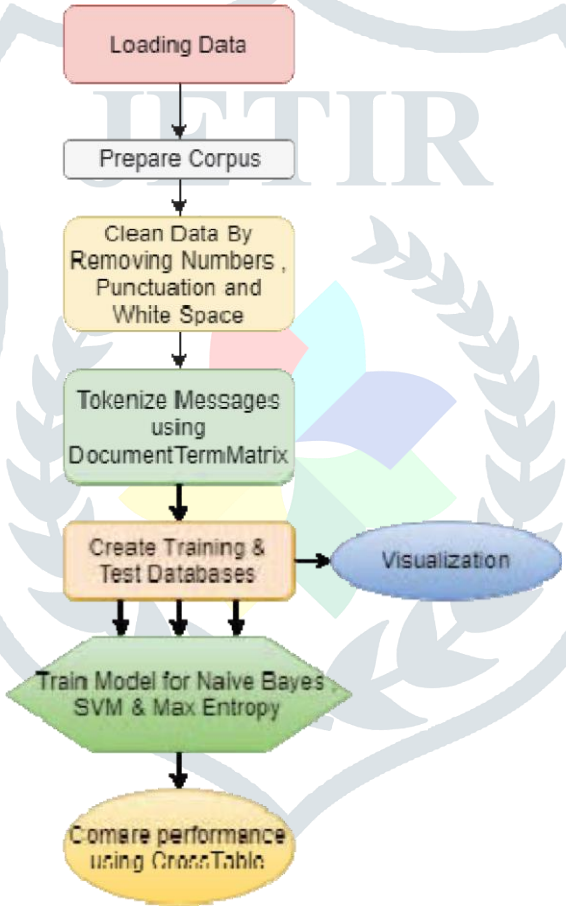


Chart 1: Ham v/s Spam

III. ARCHITECTURE OF THE CLASSIFIER



Flow-Diagram 1: Architecture of Spam Filter

As we have information in the crude shape in an excel record file, we initially import the information. We have two columns named "type" and "message". The message is the instant message while the type is the classifier of the message which is either ham or spam.

SMS messages are characters of content made out of words, punctuations, numbers, and breaks. Taking care of this kind of complex information takes a lot of attention and effort. We need to think how to evacuate punctuation, numbers, handle uninteresting words such as (and, or, but) which are called **stop words**, and how to break separated sentences into singular words. Gratefully, this utility has been given by individuals from the R group in a text mining bundle titled "tm".

The initial phase in preparing content information includes making a **corpus**, which alludes to an accumulation of text documents. For our understanding, a text document alludes to a solitary SMS message.

After removing the stop words, punctuations, numbers and blank spaces (Figure 1) we are ready to split the text messages.

WordCloud is an approach to outwardly delineate the recurrence at which words show up in information. The cloud is comprised Figure 1: Cleaning Before v/s after of words scattered fairly haphazardly around the figure.

Words seeming all the more regularly in the content are appeared in a bigger text style.Contrasting the spam wordcloud (Fig.3) and the ham wordcloud (Fig.4) will give us a thought regarding the catchphrases that will be utilized by our classifiers in separating ham and spam. On the off chance that words present in the spam cloud likewise show up as often as possible in the ham cloud, our classifier would not have solid watchwords for correlation, while if the outcomes are distinctive, the models will have the capacity to separate amongst ham and spam

well.sentences divided by total occurrences of word W1 (Spam + Ham). Similarly we can calculate for probability of ham, which will be given by the formula:

As we observe the most frequently occurring terms are completely different from the spam wordcloud, with the words occurring in the ham wordcloud being completely different from the spam wordcloud. This difference suggests that our classifiers will have strong keywords to differentiate between ham and spam.

V. NAÏVE BAYES CLASSIFIER We can characterize the issue as appeared in the accompanying formula, which catches the likelihood that a message is spam.

$$P(spam|W1 \wedge \sim W2 \wedge W3) = \frac{P(W1 \wedge \sim W2 \wedge W3|spam)P(spam)}{P(W1 \wedge \sim W2 \wedge W3)} \quad (1)$$

developed in fame as of late since it gives an approach to watch trending activities on social networking sites. We compare the wordclouds of ham and spam messages and see the difference between the frequently occurring terms in both the datasets.

$$P(ham|W1 \wedge \sim W2 \wedge W3) = \frac{P(W1 \wedge \sim W2 \wedge W3|ham)P(ham)}{P(W1 \wedge \sim W2 \wedge W3)} \quad (2)$$

For numerous reasons this equation (Eq. II) is computationally very hard to solve. As more features are added, large amount of memory is required to store the probabilities for the large part of the possible intersections.
A large number of training data would also be needed to make sure that sufficient information exists to cover all possible associations.

Our task becomes less tedious and memory efficient if we take advantage of the fact that the naïve bayes algorithm assumes independence between the events. Naïve bayes algorithm assumes **class-conditional independence,** which means that the events are not dependent upon each other as long as they are Suppose that there are total three words in the corpus , now if in a sentence word W1 and W3 appears but W2 does not appear , for finding the probability of spam , the naïve bayes algorithm takes the probability of word W1occuring in spam sentences. That is by dividing the total occurrences of word W1 in spam conditioned on similar class values. That this fact into consideration allows us to simplify the above formula using the probability rule for independent events, which is given by (Eq.3):

$$P(A \wedge B) = P(A) * P(B) \quad (3)$$

This result in a much simpler-to-compute equation, demonstrated below:

$$P(spam|W1 \wedge \sim W2) = \frac{P(W1|spam)P(\sim W2|spam)P(spam)}{P(W1)\,P(\sim W2)} \quad (4)$$

Similarly the equation for a ham message will be given by:

$$P(ham|W1 \wedge \sim W2) = \frac{P(W1|ham)P(\sim W2|ham)P(ham)}{P(W1)\,P(\sim W2)} \quad (5)$$

Therefore we can see that the naïve bayes is **98.2%** accurate in classifying a ham message and **90.9%** accurate in classifying a spam message. Therefore the naïve bayes algorithm gives an overall accuracy of **94.55%.** As we observe that the maximum entropy algorithm gives us the least accuracy in classifying the messages. The maximum entropy method gives an accuracy of **98%** in classifying ham messages and **85.9%** in classifying the spam messages. The overall accuracy given by the maximum entropy method is **91.95 %**(Table 5)**.**

VIII. REFERENCES

[1]Michael Crawford,Taghi M. Khoshgoftaar,Joseph D. Prusa, Aaron N. Richter andHamzah Al Najada, "Survey of Review spam detection using machine learning techniques", Journal of Big Data 2015.

[2] R Deepa Lakshmi , N. Radha , "Spam Classification using supervised learning techniques", A2CWiC'10 Proceedings of the 1st Amrita ACM-W Celebration of Women in Computing in India, Article No. 66.

[3] Anju Radhakrishnan et al, "Email Classification using Machine learning algorithms", International Journal of Engineering and technology(IJET).

[4] Dea Delvia Arifin ,Shaufiah , Moch. ArifBijaksana , "Enhancing Spam Detection on mobile phone short message service(SMS) performance using FP-Growth and naïve bayes classifier" , Wireless and Mobile .

[5] J.M. Gómez Hidalgo, T.A. Almeida, andA. Yamakamim " On the Validity of a New SMS Spam Collection" , Proceedings of the 11th IEEE InternationalConference on Machine Learning and Applications, (2012.)

[6] H. Kaur , "Survey on E-mail spam detection using supervised approach with feature selection" , International Journal of Engineering Sciences and Research Technology.

[7] Rekha and S. Negi, "A Review on Different Spam Detection Approaches", International Journal of Engineering Trends and Technology (IJETT), Vol.11, No.6, 2014.

[8] A. S. Aski and N. K. Sourati, "Proposed efficient algorithm to filter spam using machine learning techniques", Pacific Science Review- A Natural Science Engineering- Elsevier, Vol. 18, No. 2, Pp.145– 149, 2016.

[9] S. P. Teli and S. K. Biradar, "Effective Email Classification for Spam and Non- spam", International Journal of Advanced Research in Computer and software Engineering, Vol. 4, 2014

[10] Shafi'l Muhammad Abdulhamid , "A Review on Mobile SMS Spam Filtering Techniques", IEEE Access, 2017.

[11] Naresh Kumar Nagwani , Aakanksha Sharaff , "SMS Spam Filtering and thread identification using bi-level text classification and clustering techniques", Journal of Information Science , 2017.