University of South Wales

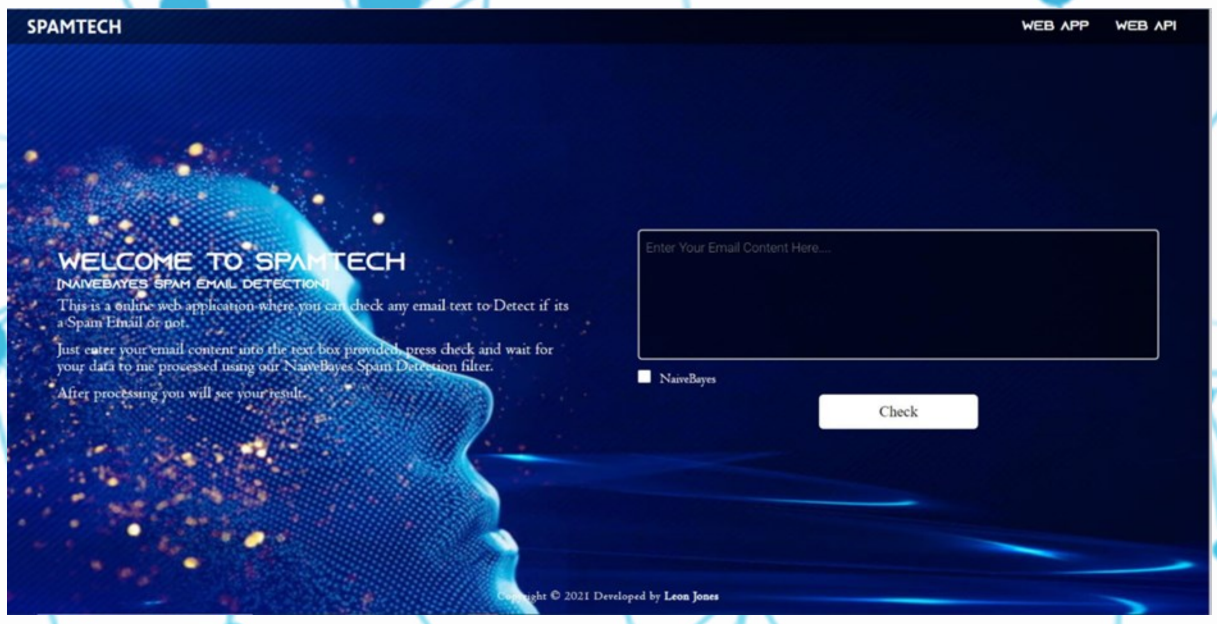
Machine Learning to Detect Spam Emails

By Leon Jones
Student Number - 18048528

Abstract

Spam emails have been increasing at a rapid rate since the early 1990's and accounts for 77.2% of all the worlds email traffic (Insu Park, 2016). As a result the there has been a need for constant development of spam email development software.

Spam email classification is a traditional problem in Nature Language Process (NLP). This dissertation contains extensive research to gain an understanding of machine learning algorithms, modern techniques and tools available. While providing a solution to the problem using the Naïve Baye approach and numerous pre-processing methods to develop a model capable of accurately predicting spam emails. Ultimately showing the feasibility of using Naïve Baye classification.



WELCOME TO SPAMTECH
[NAIVEBAYES SPAM EMAIL DETECTION]
This is a online web application where you can check any email text to Detect if its a Spam Email or not.
Just enter your email content into the text box provided, press check and wait for your data to me processed using our NaiveBayes Spam Detection filter.
After processing you will see your result.
Copyright © 2021 Developed by Leon Jones

Figure 1. Final Spam Detection Prototype

Introduction & Research

Spam email detection is the process of detecting unwanted and malicious junk emails (spam) to prevent them getting into the users email inbox. Although spam email detection software are not 100% accurate they are the only solution to solve this rapidly growing problem. A 2012 report by Kaspersky found that 2.2% of all emails contained malicious files and email attacks such as Identity theft, Phishing Attacks and Viruses (Namestnikova, 2012).

There are many different spam detection algorithms all using different machine learning techniques to provide accurate predictions. This dissertation outlines five machine learning algorithms as shown below.

- Naïve Bayes Classifier - Naïve Bayes is a probabilistic machine learning model base from figure 2. Bayes theorem of probability Wei, 2018) .
- Decision Tree - Decision tree classifier is a machine learning algorithm that follows a tree like structure to create a model that can decide the value of the sample based on attributes from existing data (Kiamarzpour, 2013).
- Clustering - Clustering method is a way to divide a group of objects into similar groups of objects known as clusters (a group of patterns).
- Support Vector Machines - Supervised machine learning algorithm that is used for classification and regression.
- Artificial Neural Network - Machine learning algorithm that are design to mimic the way that a human's brain works to recognize relationships between sets of data (Berndt Müller, 1995) .

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Figure 2. Theorem of Probability

Implementation

The dissertation used the Naïve Bayes classification approach to spam email filtering as from analysis of the research Naïve Bayes offered the best prediction accurately. The implementation was split into two prototypes, the first prototype was the creation of the machine learning spam email detection model. The machine learning model was supplied with a training dataset, this dataset was cleaned and the data visualised ensuring the model was supplied accurate training data. The text processing and vectorization was done using CountVectorizer, bag of words and TF-IDF to assign the weighting of each unique word. The vectorised data could then be used to train the Naive Bayes model. The model was tested using a 80/20 training test split and produced 97% accurate spam email predictions.

Figure 3 is a flow diagram representing each process within the final prototype. The final prototype was the implementation of the Naïve Bayes prediction model created into a django web framework allowing for the model to make prediction on individual emails entered and to be displayed in a more user friendly manner. This involve the designing and creation of the web interface and the retrieval of information entered for the model to make a prediction. Figure 1 shows the final prototype after all development was completed.

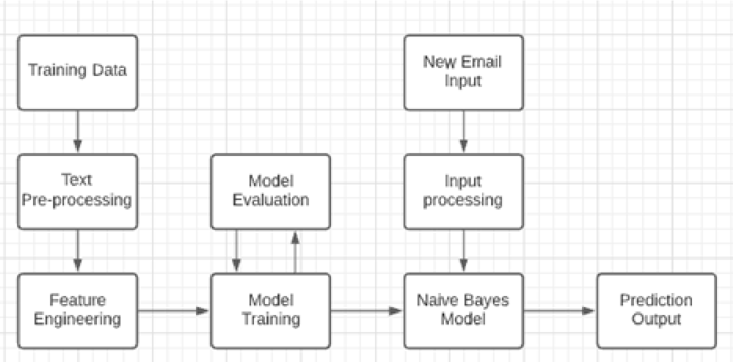


Figure 3. Final Prototype Flow Diagram

Testing & Review

After the successful creation of the final prototype extensive testing was conducted to ensure the functionality and prediction accuracy. The testing was split into 3 testing section. The first section contain 3 email messages used to test the accuracy of the predictions produced, this was a success as all prediction were accurate, showing the model prediction power within a controlled environment. Section 2 was a testing plan containing 10 tests to test each feature of both the model and interface this was also successful as all tested were passed. The last section was user feedback, 5 peers were given the final prototype and asked to test the accuracy of the model, the result were recorded using questionnaire which could be used to for future development of the project.

Conclusion

The main goal of this dissertation project was to create a machine learning model for spam email detection. This goal was met, as the spam filtering model developed can accurately predict spam emails successfully and was integrated into a Django web framework for individual email predictions. The Naïve Baye classification approach was chosen as extensive research was carried out indicating that this was the most efficient and accurate algorithm giving the small dataset used for training. The biggest issue was the limited data contained within the dataset if the dataset was large this would drastically improve the prediction power of the model. If future development of the project was to continue a number a factures would be added such as the ability to make predation for a dataset of emails, option to use other algorithms to compare their accuracy and further development of normalisation techniques such as stemming.

References

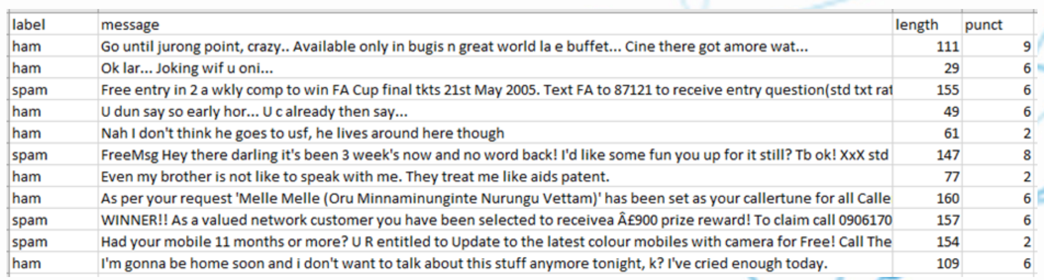
Berndt Müller, J. R. (1995). Neural Networks. Springer-Verlag Berlin Heidelberg. 52-68.

Insu Park, R. S. (2016). The Effect of Spam and PrivacyConcerns on E-mail Users'Behavior. The Effect of Spam and PrivacyConcerns on E-mail Users'Behavior, 61-64.

Kiamarzpour, F. (2013). J48 Decision Trees. Improving the methods of email classification based on words ontology, 262-266.

Namestnikova, M. (2012). Spam Report: April 2012. Kaspersky Lab.

Wei, Q. (2018). AIP Conference Proceedings. Understanding of the naive Bayes classifier in spam filtering, 1-7.



label	message	length	punct
ham	Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got amore wat...	111	9
ham	Ok lar... Joking wif u oni...	29	6
spam	Free entry in 2 a wkly comp to win FA Cup final tkts 21st May 2005. Text FA to 87121 to receive entry question(std txt rat	155	6
ham	U dun say so early hor... U c already then say...	49	6
ham	Nah I don't think he goes to usf, he lives around here though	61	2
spam	FreeMsg Hey there darling it's been 3 week's now and no word back! I'd like some fun you up for it still? Tb ok! XxX std	147	8
ham	Even my brother is not like to speak with me. They treat me like aids patent.	77	2
ham	As per your request 'Melle Melle (Oru Minnaminunginte Nurungu Vettam)' has been set as your callertune for all Calle	160	6
spam	WINNER!! As a valued network customer you have been selected to receivea Â£900 prize reward! To claim call 0906170	157	6
spam	Had your mobile 11 months or more? U R entitled to Update to the latest colour mobiles with camera for Free! Call The	154	2
ham	I'm gonna be home soon and i don't want to talk about this stuff anymore tonight, k? I've cried enough today.	109	6

Figure 4. Spam Collection Dataset