Journal of Physics: Conference
Series

An Efficient Spam SMS Analysis Model based on Multinomial Naïve Bayes model Using Passive Aggressive Algorithm

To cite this article: J. Shobana and D. Kanchana 2021 *J. Phys.: Conf. Ser.* **2007** 012047

An Efficient Spam SMS Analysis Model based on Multinomial Naïve Bayes model Using Passive Aggressive Algorithm

Mrs. J.Shobana[1], Mrs.D.Kanchana[2]

[1] Department of Computer Science & Applications, SRM Institute of Science &Technology, Chennai, India

[2]Department of Computer Science and Engineering, SRM Institute of Science &Technology, Chennai, India

e-mail: [1]shobanaj@srmist.edu.in, [2]kanchand@srmist.edu.in

Abstract. The social media can be a platform for information consumption nowadays. On the one hand, it's free of cost, easy access, and different data dissemination lead people to hunt out and consume social media news. On the contrary, it allows for the broad spread of "spams," i.e., inferiority news with deliberately false information. The widespread spread of spams has the potential for very negative impacts on people and society. Consequently, the detection of spam on social media has recently become an important research that draws tremendous attention. NLP, an artificial intelligence (AI) division, uses computers and human natural language to produce useful data. In text classification activities, such as spam detection and sentiment analysis, text generation, language translations and document classification, NLP is widely used

Keywords: SMS, Machine Learning, NLP

1. Introduction

Now a days in all the social media ready to involve in the problem of detecting the spams. In all the social media [1] like face book, twitter, bingo etc. spreading the news based on the customer analysis only. Many of the people search for the particular news that would be the hot news in the city. People does not know the news content real or fake , that only thing they did is keep on spreading the news [2]. The aim of this project is to analyze the news via machine learning algorithms and some artificial technique used to detect spams [3]. We are using different types of classifiers and algorithms to classify that is spam or ham [4]. That include in the attributes, missing values, article headlines and body and publisher all are data are useful in the detection.

Today, the quickest and easiest way to get information is through the internet and social media. Reviews, thoughts, reviews, messages and advice have become valuable sources of information in this period. Thanks to developments in technology, we are now able to use different Natural Language Processing (NLP) techniques [7] to extract useful information from such data. NLP, an artificial intelligence (AI) division [9], uses computers and human natural language to produce useful data.

Today world the fake content is spreading at fast and easy via social media platforms. Most of the people believes in their knowledge based on social media. Many of the people trust on the social medias to identify. Generally humans unable to recognize the news. The aim of this project is to analyze the news via machine learning algorithms [10] and some artificial technique used to detect fake news [4]. We are using different

types of classifiers and algorithms to classify that fake or real [5]. That include in the attributes, missing values, article headlines and body and publisher all are data are useful in the detection. We want to analyze different types of Classifiers like passive aggressive, naïve byes etc. After completing the analyzing of the classifiers we move on UI creation process. Fake news plays a major impact on social media, like US president election and Kerala flood news.

2. Proposed Architecture

The Proposed architecture is shown as in figure 1.The proposed model is build based on TD-IDF for feature extraction and multinomial Navie bayes [11] model Using Passive Aggressive Algorithm for classification.

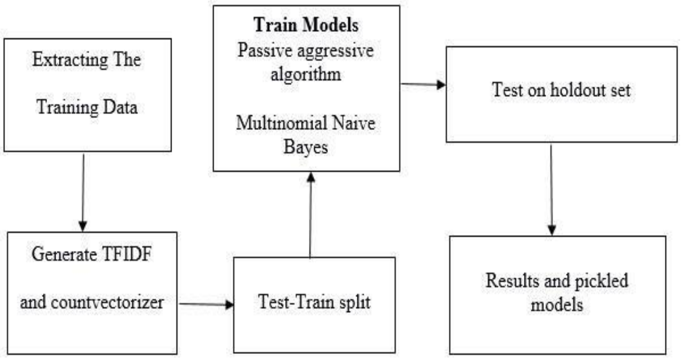


Figure 1. System structure of SMS Spam Analysis Modle

A proposed model is built based on the count vectorizer or a TD-IDF matrix for feature extraction process. Since this problem is a sort of text classification [8], it will be best to implement Passive aggressive classifier as this is standard for text-based processing. The benefit in our proposed system, the functionality of the Sci-kit Learns Grid Search is used to perform this task efficiently. For count vectorizer, the optimal parameters are normally highlighting, handle only two words, and use only words that appear in the corpus at only three times. The validated accuracy score for this Passive aggressive model is 87%, the true positive score is 91%. “Corresponding author.” This is the author to whom proofs of the paper will be sent. Proofs are sent to the corresponding author only.

2.1 Proposed Methodology

Spams are the irrelevant contents through the real life that would be different from the real life environment [15]. That spams are created for the different purposes like some political agendas, or for the popularity of the peoples or to make a controversy. Such the news are identify by different algorithms. Generally humans unable to recognize the news. The aim of this work is to analyze the machine learning algorithms [16] and some artificial technique [13] used to detect spams. We are using different types of classifiers and algorithms to classify spam or Ham [19]. That include in the attributes, missing values, article headlines and body and publisher all are data are useful in the detection. We want to analyze different types of classifiers like passive aggressive, naïve byes etc. After completing the analyzing of the classifiers we move on UI creation process. Fake news [18] plays a major impact on social media, like US president election and Kerala flood news.

2.2 Dataset Details

The dataset is collected from UCI repository which consists of massive amount of data. We have extracted mobile SMS messages dataset which consists of 5,574 short text messages. This is collected for the purpose of conducting Spam SMS analysis.

2.3 Preprocessing

Data collected from UCI dataset are need to be preprocessed before analysis. Preprocessing tasks involves tokenization, stemming, Lemmatization, Padding etc. Figure2 shows the preprocessing of mobile SMS text.

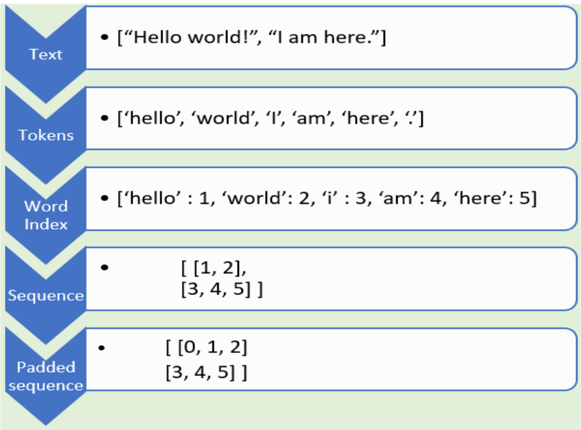

Text
• ["Hello world!", "I am here."]
Tokens
• ['hello', 'world', 'I', 'am', 'here', '.']
Word Index
• ['hello' : 1, 'world': 2, 'i' : 3, 'am': 4, 'here': 5]
Sequence
• [[1, 2], [3, 4, 5]]
Padded sequence
• [[0, 1, 2] [3, 4, 5]]

Figure 2. Preprocessing of SMS text

2.4 Feature Extraction

Feature Extraction phase starts after the pre-processing completed. Right features derived from the pre-processed data are feed into the classifier for classification. The widely used TD-IDF classical model is utilized in this work.

2.4.1 TF/IDF algorithm

The TF of a word is the frequency of a word in a text (i.e. the number of times it appears). The IDF of a word is the indicator in the entire corpus of how relevant that term is. If a word occurs in a document often, then it should be relevant and we should give a high score to that word. But it's probably not a unique identifier if a word appears in so many other papers, so we can give the word a lower ranking.

TF (Term Frequency)

In the term frequency is used to count the number of words appears in the document. If the result is given the higher value means the words comes frequently in the document.

IDF (Inverse Document Frequency)

It counts words that occur are irrelevant to the headlines. That measures the words and count that words in the document.

Building Counter Vectorizer

The number of terms appearing in the text is used in the term frequency to count. The higher value is given to the result, indicating that the words also appear in the text. Now considering the training data is nothing but we are already using data in the project. Trained data means frequently used data in our project. We are cutting the bag of words with a term frequency, IDF.

2.5 Proposed Algorithm Used For Classification

The proposed algorithm is discussed in this section . Figure 3. The system architecture of the model proposed is shown.

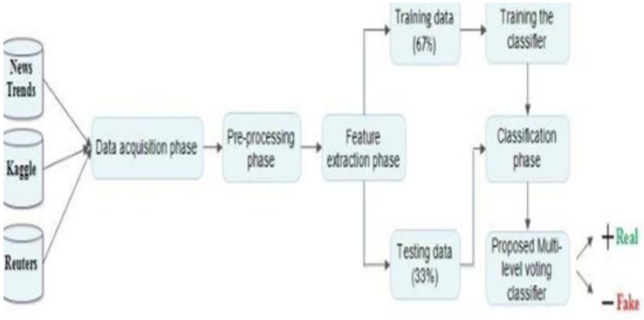


Figure 3. Proposed Model Architecture.

The proposed model composed of the following task:

- Load the data and Explore
- Prepare training data and testing data
- Use the proposed algorithm for classification
- Compare the results
- Identify the SPAM
-

2.5.1 Multinomial Naïve Bayes model with Passive-aggressive Algorithm:

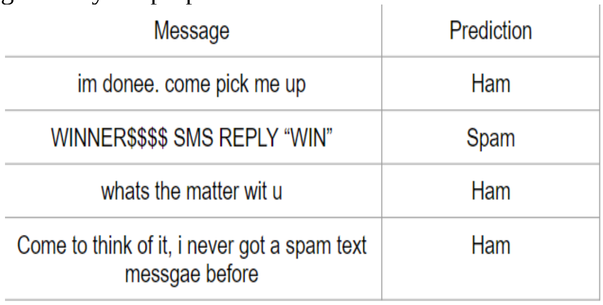
For classifying the SMS into actual or spam, the Naïve Bayes classification algorithm is used. A Naive Bayes classifier assumes that the presence in a class of a certain function is not related to the presence of any other function.

Passive-aggressive behavior is a pattern of indirectly expressing negative feelings instead of openly addressing them. ... For example, a passive-aggressive person might appear to agree — perhaps even enthusiastically — with another person's request. Passive aggressive classifier improves the accuracy and confusion matrix improvements. Passive aggressive algorithm for online learning process. In the algorithms remains for a correct classification process. It is not an easy job to detecting spam [17] with classifiers. That also given label to the given data as real or spam. This algorithm correct the classification result and miscalculation all other functions. Passive aggressive classifier improves the accuracy and confusion matrix improvements.

There exists an outsized body of research about machine learning methods for spam detection [20] are using only the classifiers, clusters and random forest methods. Current time all are using NLP (Natural language processing) for spam detection, using the NLP they are classifying the text to a Spam detection. First that classify the text or paragraph .Then process using algorithm to process the SMS with a comments and other article to finalize the results.

3. Results And Analysis

The results of the proposed system are tabulated and analyzed in this section. Fig.4 shows the spam or Ham analysis labelling done by the proposed model.



Message	Prediction
im donee. come pick me up	Ham
WINNER$$$$ SMS REPLY “WIN”	Spam
whats the matter wit u	Ham
Come to think of it, i never got a spam text messgae before	Ham

Figure 4. SPAM / HAM Labelling

3.1 Confusion Matrix

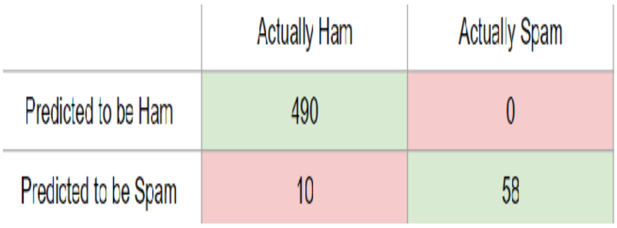
Confusion Matrices easier to match and skim, using the scikit-learn doc to create some readable confusion matrices. A confusion matrix shows the most diagonally of the right labels. The other cells display the incorrect labels, often called false positives or false negatives. One of those might be more significant depending on the problem.

The Confusion matrix consists of,

- ○ TRUE POSITIVE
- ○ FALSE NEGATIVE
- ○ TRUE NEGATIVE
- ○ FALSE POSITIVE

- True Positives: Positive values are reliably predicted.
- True Negatives: Negative values are reliably predicted.
- False Positives: Negative values were wrongly predicted as positive values.
- False Negatives: Negative values are wrongly predicted to be positive values.

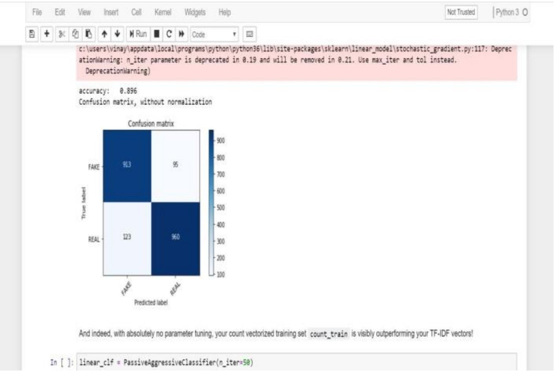
Results And Confusion Matrix



	Actually Ham	Actually Spam
Predicted to be Ham	490	0
Predicted to be Spam	10	58

Figure 5. Result and Confusion Matrix

- ❖ TRUE POSITIVE (4653)
- ❖ FALSE NEGATIVE (60)
- ❖ TRUE NEGATIVE (51)
- ❖ FALSE POSITIVE (4216



```
c:\users\vinay\appdata\local\programs\python\python36\lib\site-packages\sklearn\linear_model\stochastic_gradient.py:117: Deprec
ationWarning: n_iter parameter is deprecated in 0.19 and will be removed in 0.21. Use max_iter and tol instead.
  DeprecationWarning)

accuracy:   0.896
Confusion matrix, without normalization
```

And indeed, with absolutely no parameter tuning, your count vectorized training set `count_train` is visibly outperforming your TF-IDF vectors!

```
In [ ]: linear_clf = PassiveAggressiveClassifier(n_iter=50)
```

Figure 6. Results of Confusion Matrix

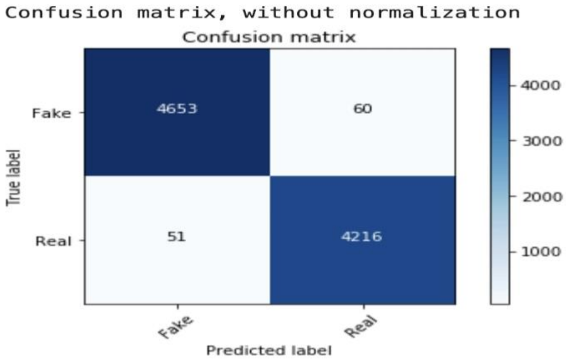


Figure 7. Confusion Matrix without Normalization

4. Conclusion

The size of the things / goods reviews is growing due to the rapid growth of the internet. On the internet, such enormous quantities of knowledge are produced, there is no examination of the quality of texts written by the user. SMS spam prediction analysis is one of the vital task in Natural Language Processing field. The proposed model Multinomial Naïve Bayes model Using Passive Aggressive Algorithm achieves 89% accuracy in prediction which is higher than state-of art methods.

References

[1] Shafi'i Muhammad Abdulhamid ,Muhammad Shafie ,Haruna Chiroma , Oluwafemi Osho , Gaddafi Abdul-," A Review on Mobile SMS Spam Filtering Techniques", IEEE Access @2017 ,Volume 5,PP :2169-3536 .

[2] Lutfun Naher Lota,"A Systematic Literature Review on SMS Spam Detection Techniques", July 2017,International Journal of Information Technology and Computer Science 9(7):42-50 ,DOI: 10.5815/ijitcs.2017.07.05

[3] S. Banerjee, A. Chua, J. Kim, "Using Supervised Learning to Classify Authentic and Fake Online Reviews ", Proceeding of the 9th International Conference on Ubiquitous Information Management and Communication", ACM, 2015.

[4] Wang, Z, T Hou, D Song, Z Li and T Kong, "Detecting review spammer groups via bipartite graph projection", The Computer Journal , 59(6), pp. 861–874, 2015

[5] S.L. Christopher and H. A. Rahulnath, "Review authenticity verification using supervised learning and reviewer personality traits,"

[6] Akoglu, L, R Chandy and C Faloutsos , "Opinion fraud detection in SMS by network effects". In Proceedings of the 7th AAAI International Conference on Weblogs and Social Media (ICWSM'13), pp. 2-11, 2013

[7] N. J. Conroy, "Automatic deception detection: Methods for finding fake news,," in in Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community, USA, 2015.

[8] J. D'Souza, "An Introduction to Bag-of-Words in NLP," 03 04 2018. [Online]. Available: https://medium.com/greyatom/an-introduction-tobag-of-words-in-nlp-ac967d43b428.

[9] G. Bonaccorso, "Artificial Intelligence – Machine Learning – Data Science," 10 06 2017. [Online]. Available: https://www.bonaccorso.eu/2017/10/06/mlalgorithms-addendum-passive-aggressivealgorithms/.

[10] Gilda, "Evaluating Machine Learning Algorithms for Fake News Detection," in IEEE 15th Student Conference on Research and Development (SCOReD), 2017.

[11] K. Daraje, M. Getachew and D. Jabesa, "Afaan Oromo Text Content-Based Fake News Detection using Multinomial Naive Bayes," International Journal of Innovations in Management, Science and Engineering (IJIMSE), vol. 01, no. 01, pp. 26-37, 01 March 2020.

[12] Suleiman, D. and Al-Naymat, G., "Sms spam detection", Procedia Computer Science, Vol. 113, (2017), 154-161.

[13] FakeNewsChallenge, "Exploring how artificial intelligence technologies could be leveraged to combat fake news.," 2019. [Online]. Available: http://www.fakenewschallenge.org/

[14] Olusola Abayomi-Alli, Sanjay Misra, Adebayo AbayomiAlli, Modupe Odusami. "A review of soft techniques for SMS spam classification: Methods, approaches and applications", Engineering Applications of Artificial Intelligence, 2019

[15] "Detecting Online Spams through Supervised Learning Techniques", International Journal of Innovative Technology and Exploring Engineering, 2019

[16] Nidhi A. Patel, Rakesh Patel. "A Survey on Spam Detection using Machine Learning Techniques", 2018 4th International Conference on Computing Communication and Automation (ICCCA), 2018

[17] Vijayasekaran G, Rosi S (2018) Spam and email detection in big data platform using naives bayesian classifier. IJCSMC 7(4):53–58

[18] Gupta A., Palwe S., Keskar D. (2020) Fake Email and Spam Detection: User Feedback with Naives Bayesian Approach. In: Bhalla S., Kwan P., Bedekar M., Phalnikar R., Sirsikar S. (eds) Proceeding of International Conference on Computational Science and Applications. Algorithms for Intelligent Systems. Springer, Singapore.https://doi.org/10.1007/978-981-15-0790-8_5

[19] Chan, P.P., Yang, C., Yeung, D.S. and Ng, W.W., "Spam filtering for messages", Neurocomputing, Vol. 155, (2015), 167-176

[20] Choudhary, N. and Jain, A.K., "Towards filtering of sms spam messages using machine learning based technique", in International Conference on Advanced Informatics for Computing Research, Springer., (2017), 18-30.