Implementation of Spam Classifier using Naïve Bayes Algorithm

Ajay Gangare[1], Jitesh Rathore[2], Akash Kumar Tadge[3], Abhinav Shrivastav[4], Rashmi Yadav[5], Pankaj Singh Sisodiya[6]

[1-6]Dept. of Computer Science and Engineering, Shri Balaji Institute of Technology and Management, Betul, M.P

---***---

Abstract - *Spam involves sending someone unwanted messages. Currently the internet is the biggest platform to get some information, also social media is going to be very popular nowadays. Because of that, many spammers will try to mislead users by sending lots of spam messages. And because of spam messages, there are lots of problems, fraud occurs. So we want to filter messages into spam or ham. To classify the messages as spam or not spam we are using machine learning (the multinomial naïve Bayes classifier algorithm) and CountVectorizer provided by Scikit-learn library in python programming. First, we collect the datasets and convert them into numerical data (matrix) by CountVectorizer and then we apply the naïve Bayes algorithm on datasets for classification purposes.*

***Key Words*: naive Bayes algorithm, CountVectorizer, machine learning, Bayesian classification.**

1.INTRODUCTION

In today's world, since social media and the internet is very popular, so there are lots of people using abusing messages or shady comments, also many spammers will send a bulk of spam messages like they send (malicious spam) fake link to us and when we click on the link, then they get the access of our information, because of that we may get into trouble. Many organizations and people could face financial loss. Some of spamming include unwanted advertisement of the product, sometimes it becomes very irritating for people. Some of the spammers are doing the work of spreading computer viruses.

1.1 Problem Statement

Every day, we receive tons of junk messages, emails, some spam messages are just marketing/advertising messages while some are fake spam messages. People are annoyed/irritated because of such spam messages they receive. The main objective of the problem is to classify the comments into spam or not spam. And in our project , we are designing 3 separate modules of spam classifier for social media spam, SMS spam, and email spam.

2. LITERATURE SURVEY

There are many types of spam, such as web spam, short message spam, email spam. social network spam and others. In this paper, we are focusing on social media, mobile SMS spam, and email spam.

2.1 Existing System

Data mining techniques are commonly used for spam classification. In our paper, we are using machine learning algorithm to automate the process of spam detection.

2.2 Proposed System

In our proposed system, count vectorizer is used for extracting features from a given dataset, and a design model is used for generating tests and training sets from given dataset. Then the naive bayes classifier is trained in training data. And the proposed system will say given data is spam or not.

3. Naïve Bayes

Naïve Bayes classifiers are a type of linear classifiers. Naïve Bayes algorithm is a very simple algorithm used for the classification purpose, and the naïve bayes classifier is based on a probabilistic model, the base of the naïve bayes algorithm is the Bayesian theorem. Naïve Bayes algorithm will calculate the probability of input word based on the probability of past (trained dataset) and classify input as spam or not spam. Naïve bayes algorithm uses the formula for classifying input message as spam or not spam is given below.

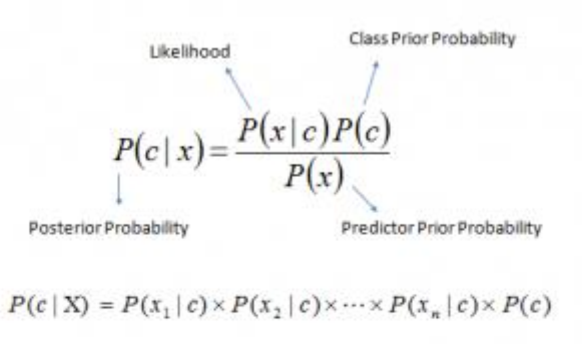
Likelihood
Class Prior Probability
$P(c|x)=\frac{P(x|c)P(c)}{P(x)}$
Posterior Probability
Predictor Prior Probability
$P(c|X)=P(x_1|c)\times P(x_2|c)\times\cdots\times P(x_n|c)\times P(c)$

Fig -3.1: Formula of Bayesian theorem

In this article [7] the author Vinoth introduced (in above picture) that

Above,

P(c|x) is the posterior probability of class (c, target) given predictor (x, attributes).

P(c) is the prior probability of class.

P(x|c) is the likelihood which is the probability of predictor given class.

P(x) is the prior probability of predictor.

4. Count Vectorizer

CountVectorizer is a tool in python programming. Since we can only apply our machine learning algorithm (naïve bayes algorithm) to numerical data, we need to convert our text message into numerical data, for that we use CountVectorizer. Now CountVectorizer will analyze the text data and create a matrix and in the matrix, each row represents the certain text data and each column represents

a unique word. The count of the word in certain text data is considered as the value of each cell. As a result, we get a vector that gives (specify) the length of the entire text data.

5. Flowchart

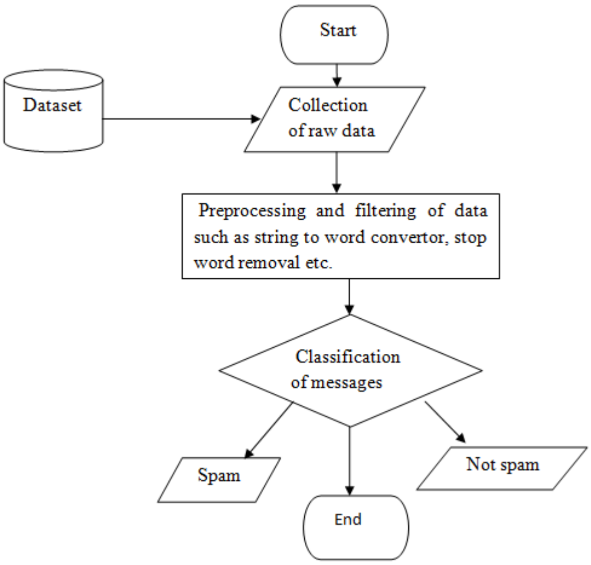
Start
Dataset
Collection of raw data
Preprocessing and filtering of data such as string to word convertor, stop word removal etc.
Classification of messages
Spam
Not spam
End

Fig 5.1: Flowchart of proposed system

6. METHODOLOGY AND IMPLEMENTATION

We are going to build our project (spam classifier application) in 4 phases

(i) Collection of Datasets (CSV files)

(ii) Pre-Processing of the dataset

(iii) Features Selection and Extraction

(iv) Classification phase

(i) Data Collection

In the dataset collection phase, the dataset that is going to be used for the training dataset we downloaded through API contained 9 CSV files of email spam and testing for machine learning algorithm is downloaded from the website UCI machine learning repository, The, mobile SMS spam, social media spam respectively.

(ii) Pre – processing

In the pre-processing phase, we take the collected dataset and we will preprocess it, preprocess means we delete some unwanted information (data) from the datasets like we delete some columns and rows that are not needed (important). If the collected dataset is not cleaned, then we have to perform certain operations to make it clean. In the pre-processing phase, if the same words are occurring multiple times then we consider it only once.

(iii) Feature Selection and Extraction

After preprocessing phase, we need to find the feature from the preprocessed dataset, so that the naïve bayes algorithm can be run on the feature. First, we analyze the preprocessed dataset and we use several feature selection methods to find the best feature that will be well suited to train data and give the best result.

(iv) Classification

In the classification phase we have to divide our data (dataset) into 2 phases – (a) Training Phase (b) Testing Phase. 60% of data is used for the training phase and 40% of data is used for the testing phase. After completing the step feature selection and extraction, we get the feature and the feature that is selected is considered spam. In this classification phase, our datasets will be trained based on the naïve bayes algorithm. For training the dataset we use a linear classification algorithm that is naïve bayes algorithm. The feature is selected in the feature selection and extraction phase and we use the naïve bayes algorithm to run on the features so that output is produced. There are many classifiers available for the naïve bayes algorithm, but in our article, we are using the naïve bayes multinomial classifier for classification purposes because the multinomial naïve bayes algorithm gives higher accuracy.

7. PROPOSED SYSTEM

There are many existing systems available that are based on data mining and machine learning techniques, but our proposed system will be based on a machine learning algorithm (naïve bayes algorithm) which is based on the Bayesian theorem. The proposed system will calculate the probabilities of spam and not spam messages, and according to the probabilities, it will produce the result.

8. RESULT

After Creating the model (completion of the project) we can predict whether the comment is SPAM or NOT We are able to classify the messages as spam or non-spam. After implementing algorithm (model) on dataset , after testing our model ,the efficiency of our model will be up to 94%.

9. CONCLUSION

In this article, we developed the spam classifier application (model) by using a machine learning algorithm. there is always the issue of security in the social media platform. The proposed system is used to identify the spam and classify the messages into spam and not spam categories using the Naıve Bayes technique. But if we use the Support Vector Machine for training dataset along with naïve bayes algorithm then, we can get better efficiency. After testing various test cases we conclude that the efficiency of our proposed system (model) will be up to 94%. We can improve the efficiency of our model by doing some improvements like using "tfidf vectorizer" etc.

REFERENCES

[1] Nurul Fitriah Rusland, Norfaradilla Wahid, Shahreen Kasim, Hanayanti Hafi 2017 IOP Conf. Ser.: Mater. Sci. Eng. 226 012091

[2] D. Sen, C. Das and S. Chakraborty, "A New Machine Learning based Approach for Text Spam Filtering Technique", Communications on Applied Electronics, Vol. 6, No. 10, pp. 28-34, 2017.

[3] H. Sajedi, G.Z. Parast and F. Akbari, "SMS Spam Filtering using Machine Learning Techniques: A Survey", Machine Learning Research, Vol. 1, No. 1, pp. 1-14, 2016.

[4] Rizky et al. "The Effect of Best First and Spread subsample on Selection of a Feature Wrapper with Naïve Bayes Classifier for Classification of Ratio of Inpatients". Scientific Journal of Informatics.

[5] Anjana Kumari," Study on Naive Bayesian Classifier and its relation to Information Gain," International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 2, Issue- 3, March 2014, pp.601 – 603.

[6] Rushdi Shams and Robert Mercer," Classifying Spam Emails using Text and Readability Features," IEEE 13th International Conference on Data Mining (ICDM), 2013, pp. 657-666.

[7] Naive Bayes Algorithm, 2019. https://huntdatascience.wordpress.com/2019/09/24/naive-bayes-algorithm/. Accessed September 24, 2019.