

Respected Sir,

As per the instruction, below are the ways that I approached to solve the given Problem.

After getting the Problem statement, Since I knew both the labels and features in the given data set, I tried with the logistic regression and got ended up with many errors. Then I Started Stalking through the several websites and YouTube videos to understand which algorithm is best suitable for Spam detection and after spending some hours, all I got to know is naive bayes algorithm is best suited and its one of the trending algorithms which have been used for the spam detection till today and started implementing it.

In the given data set naive bayes algorithm calculates the probability of the words one by one which are classified as spam and ham as returns back the given sentence is Spam or ham based on the accuracy of the Spam and ham probability.

- 1) Used pandas library to read the data set
- 2) Used sklearn library which is the most Useful and robust library for ML in python which includes all the efficient tools
- 3) Set all the Spam variable to 1 and Ham variable to 0 for converting the frequency of the word to array
- 4) Categorized sentences as labels and the results (spam and ham) as features
- 5) Used CountVectorizer function to convert the word to vector
- 6) Split the test train datasets using test_train_split
- 7) Used Naïve Bayes Classifiers to train the model by passing the test train data set
- 8) Predict the test data through train data
- 9) Comparing the desired Sentence with predicted test data finally it gives the result as Spam or Ham

Further Improvements:

We can use stemming and lemmatization methods of Natural language tool kit to get the root word of the words in the sentences and only convert those words to the vector and use it to compare with test sets

I am learning these concepts, and I am more interested towards the python since I am working with Python from the first year if I get a chance of working with you from your guidance and training, I can build up my skills and abilities.

Looking forward for you reply.

Regards,

Chetan C Mundargi,

PES University.