

**E-MAIL SPAM CLASSIFIER**

**BY:- RISHABH JOHRI**

**DATA SCIENTIST INTERN**

**ACKNOWLEDGEMENT**

It is a great pleasure to express my gratitude to Team Flip Robo, for giving me the opportunity to work on a interesting project, which helped me in improving my knowledge, coding skills and my analysation skills.

Team Flip Robo also gave me opportunity to build PowerPoint Presentation and Project Report, which will help me to share steps taken while building the entire model. It has helped me in deciding about the future prospects of various Data Science fields. Now, I will explain the understanding of the project through this report.

**CONTENT:-**

Spamming is one of the major attacks that accumulate many compromised machines by sending unwanted messages, viruses, and phishing through E-Mails. It conveys the principal aim behind choosing this project under with. There are many people who try to fool you, just by sending you fake e-mails! Some of them have the messages like - You have won 1000 dollars, or we have deposited this much amount to your account. Once you open this link, then they will track you and try to hack your information.

**SCOPE OF PROJECT:-**

Spamming is a big problem! Anyone who is connected to internet faces this problem now and then, but with Corporates, it becomes a serious problem. Phishing attacks, greedy traps and many create a havoc if not handled properly. Our product E-Mail Spam Classifier is a solution for these Spams. Through our Intelligent Web- Based Product, you can find Spam E-Mails and can label them as Spam or ham for future reference.

The context of this product is as described above, and our team’s goal is to make a safer and crime-free CyberWorld. This is the first time we are working on this product, and it is something new for us. We are fully aware of the fact that there will be several challenges to make this product efficient and trustworthy. As spammers are getting smarter each day, we also need highly efficient Machine Learning Algorithm to tackle and knock out these Spammers.

Major highlights of our product’s functionality are:

▪ This product will help identify Spam E-Mails similar to the Spam encountered earlier, which are stored in a vast library of Spam E-Mails.

▪ This product will also help in identifying new Potential Spam E-Mails from known & unknown sources. This what is going to be a speciality of our product.

**Hardware Requirements:-**

**The Hardware required to run this Product is very minimal and is mentioned below:**

**▪ Operating System – Win/Mac/Linux (any distribution)**

**▪ Web Browser–Chrome/Mozilla/Opera (Updated with latest FTP/HTTP support)**

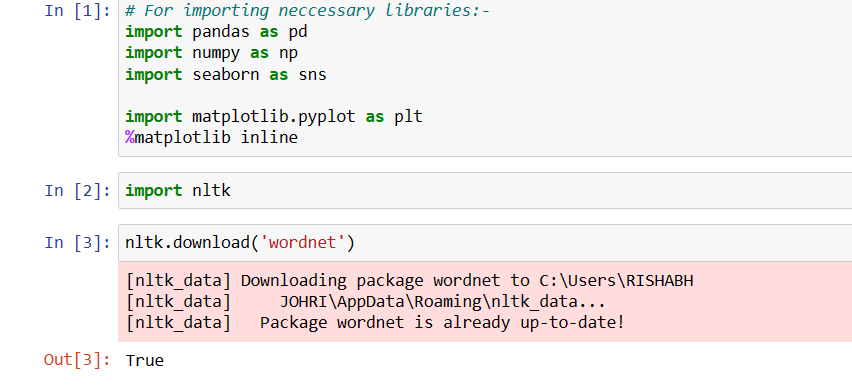
**▪ Proper Internet Connection (above 1 Mbps).**

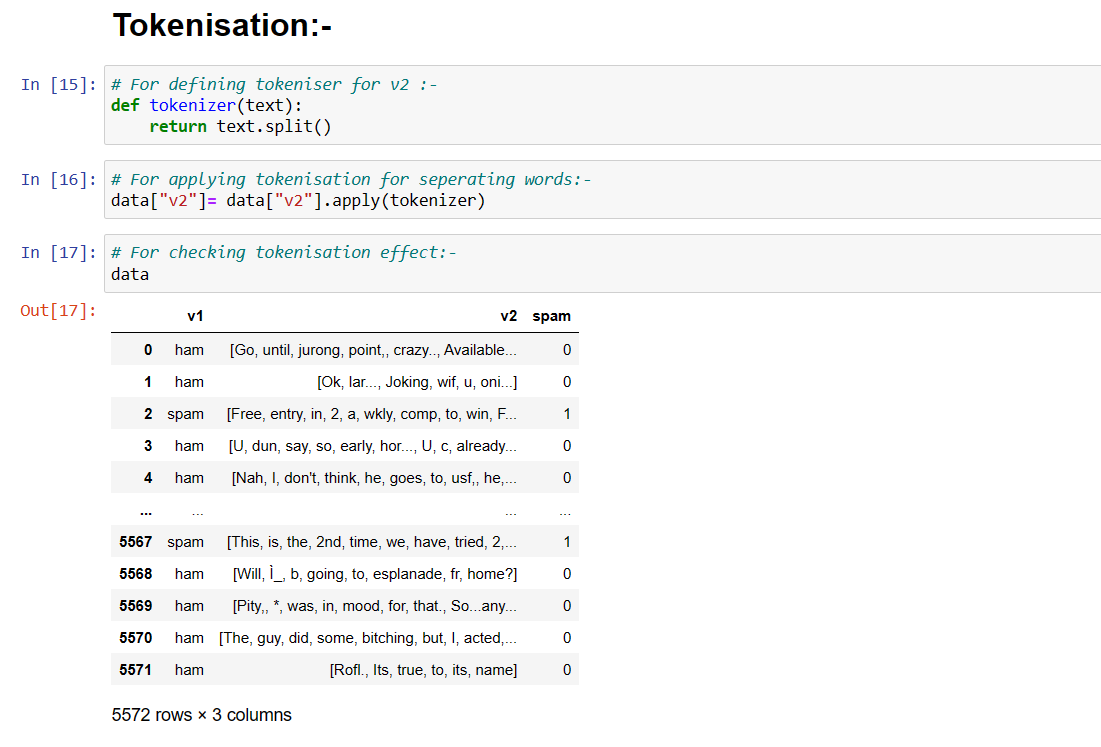
**Software Interfaces:-**

There are tons of software which are available online which can be used to make a Spam Classifier. But we have mentioned some softwares which we have used and will use in our Development of this Product, which are mentioned below:

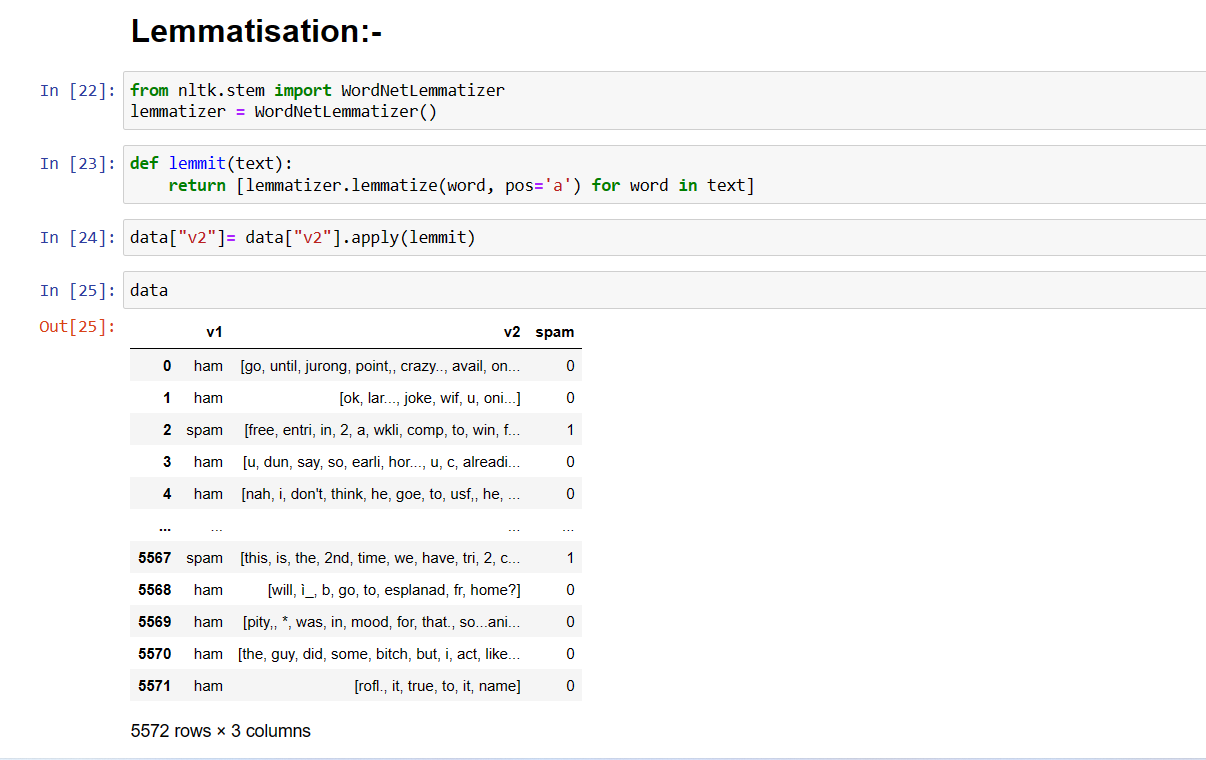
▪ Jupyter Notebook | Google Collab.

**Building Model:-**

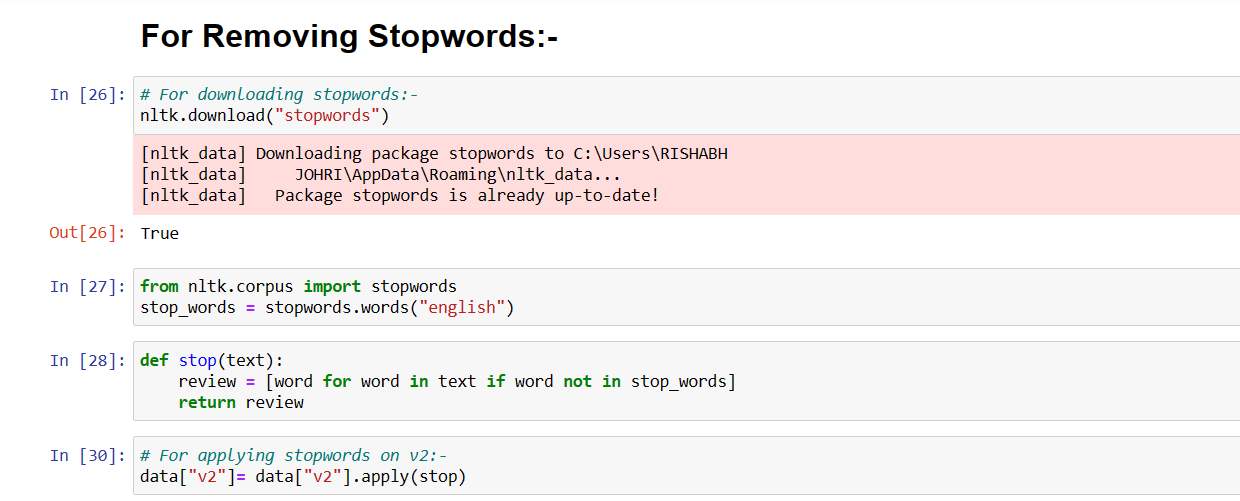




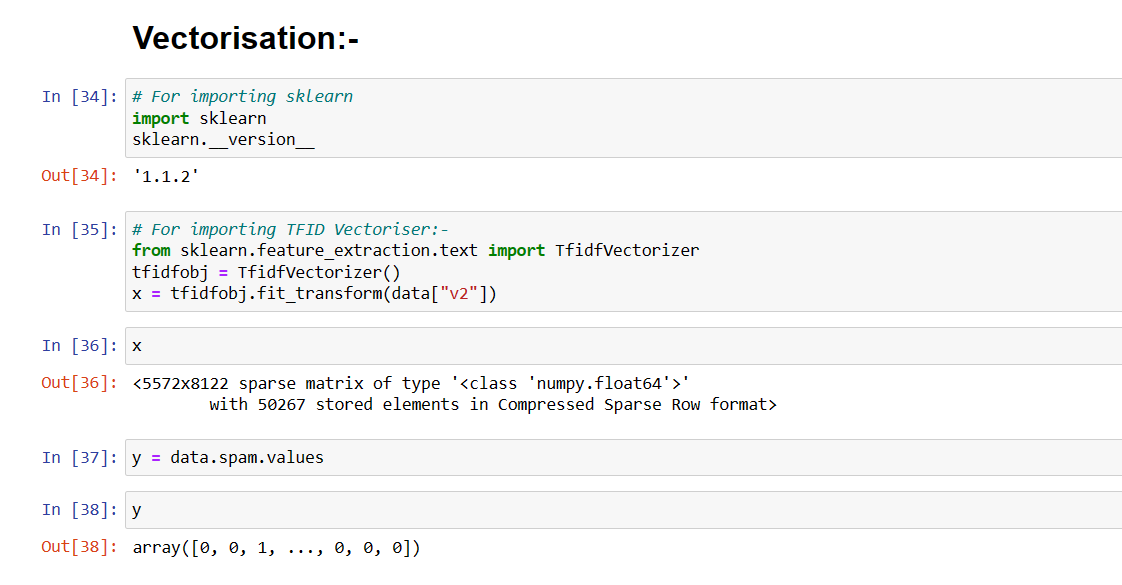




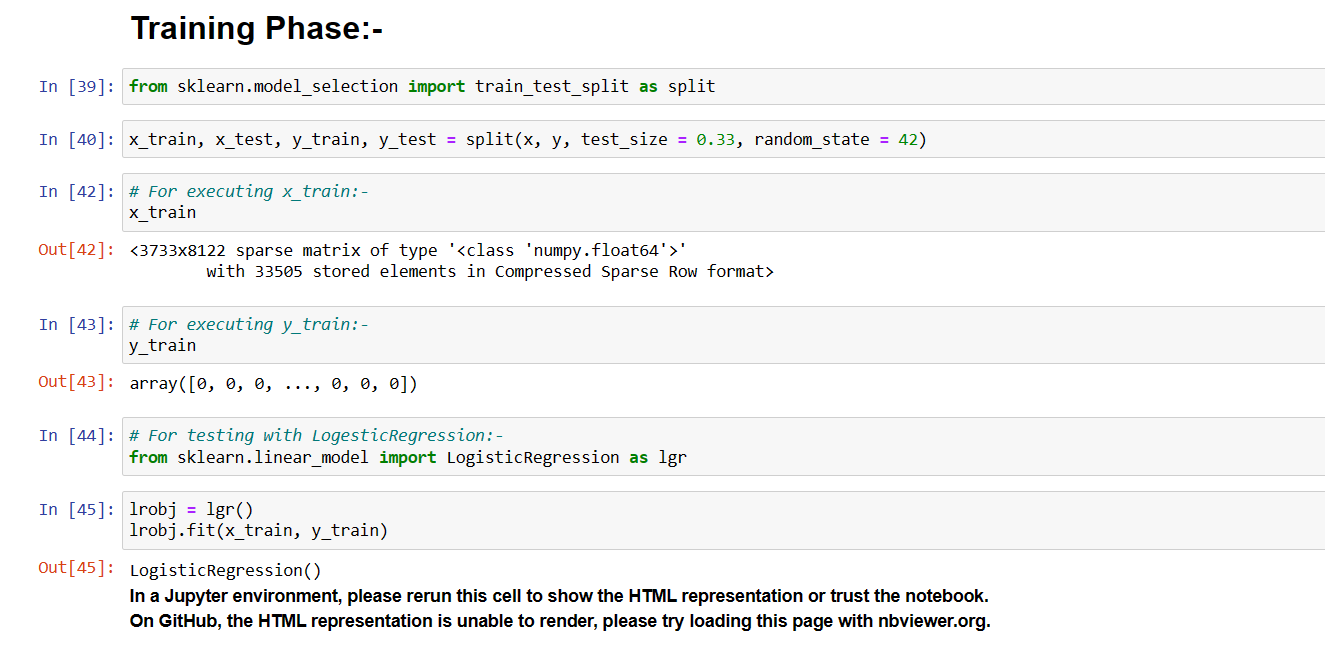
**For Removing Stop words from the dataset:-**



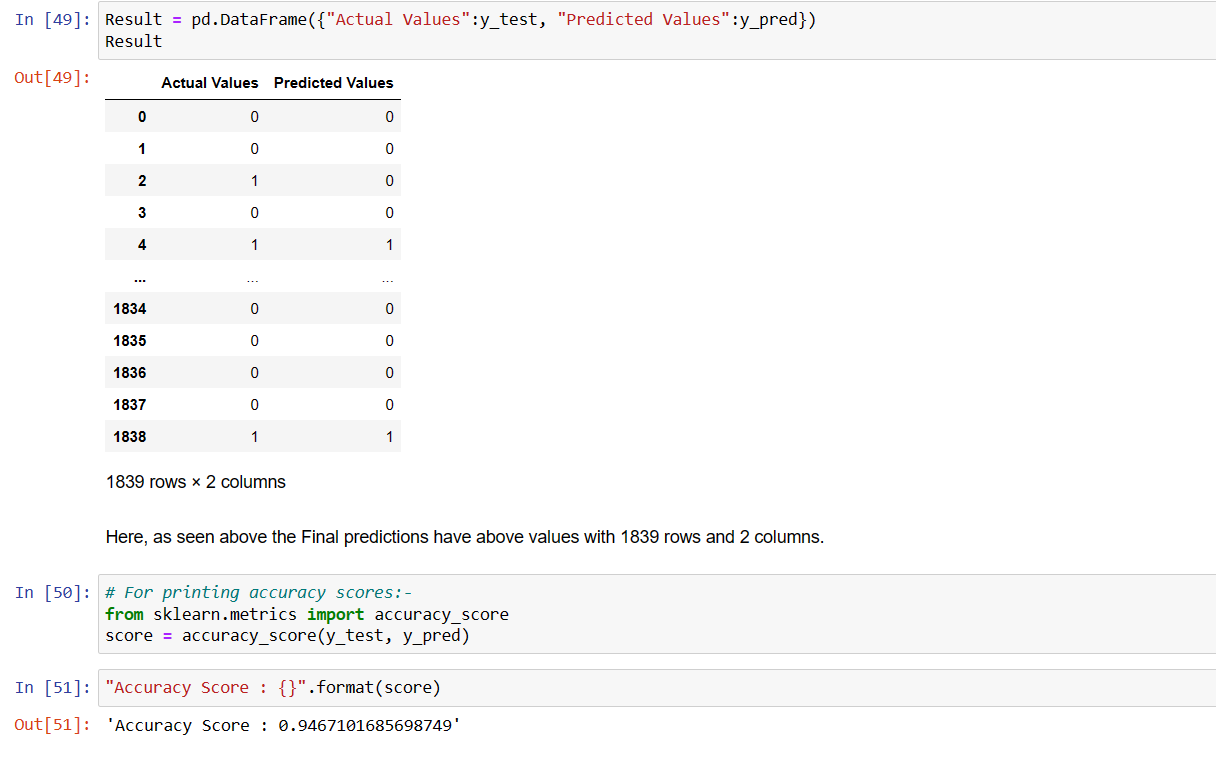
**VECTORISATION:-**

****

**Training the data:-**

****

**PREDICTIONS AND ACCURACY SCORES:-**

****

**CONCLUSION:-**

In this project report, I have used Machine Learning algorithms to predict the spammers as it is a spam or Ham. We have used proper procedure to analyse the dataset and finding the correlation between the features.

Here we have selected the features which are correlated to each other and are independent in nature.

Data cleaning is one of the most important steps to remove 0 values and columns which are having more than 90% 0 values.

Using these feature, I have deployed Logestic Regression algorithm to find the best model and we have achieved the accuracy score of 94.6 % which is very good to proceed.

Then we saved the best model and predicted the label. Our model performance was good when we saw the predicted and actual values were almost same it felt really good and feels good performance by our model.

To conclude, the project Email Spam Classifier , We hope this study will move a small step ahead in providing some methodological and empirical contributions to crediting institutes, and presenting an alternative approach to the valuation of defaulters.