

SMS BLACK LIST DETECTION USING MACHINE LEARNING

RAVULAPALLI CHANDANA¹, MD.KAUSAR FATHIMA², A.RISHITHA³, SK.BABLU⁴

Student, Department of CSE, NRI INSTITUTE OF TECHNOLOGY, Vijayawada, A.P., India.

Mr.K.CHANDRAMOULI Associate professor, Department of CSE, NRI INSTITUTE OF TECHNOLOGY, Vijayawada, A.P.

ABSTRACT

Nowadays spam messages is the biggest issue in SMS message communication. This spam messages causes danger to our personal data .In order to solve this problem, an accurate solution is needed to detect the spam in mobile message communication. So, we are using machine learning algorithms to solve this problem. The algorithms we using are Logistic regression (LR), K-nearest neighbour (K-NN), and decision tree (DT) to detect ham and spam messages in SMS messages. This spam classification has become more challenging due to complexities of the messages imposed by spammers.

Even with the filtering mechanisms the number of spams are increasing day by day. This is the major problem which are facing by the people. Spams are defined as unsolicited bulk messages in various forms such as unwanted advertisements, credit opportunities or fake lottery winner notifications. Spam classification has

become more challenging due to complexities of the messages imposed by spammers.

INTRODUCTION

The SMS spam collection data set is used for testing the method. The dataset is split into two categories for training data and testing data. when you separate a data set into a training set and testing set, most of the data is used for training, and a smaller portion of the data is used for testing data. The training data is considered as 80% and testing data considered as 20%. After training and testing the data, we have to apply pre-processing techniques. Some of the pre-processing techniques used are data cleaning and removing unnecessary words in the dataset. After all these, the models are applied to these dataset to get the accurate result. i.e., spam or ham message.

- **HAM:**

The term 'ham' was originally coined by SpamBayes sometime around 2001 and is currently defined and understood to be "E-

mail that is generally desired and isn't considered spam.”

- **SPAM**

Mobile msg spam is a form of spam (unsolicited messages, especially advertising), directed at the text messaging or other communications services of mobile phones or smart phones. As the popularity of mobile phones surged in the early 2000s, frequent users of text messaging began to see an increase in the number of unsolicited (and generally unwanted) commercial advertisements being sent to their telephones through text messaging. This can be particularly annoying for the recipient because, unlike in email, some recipients may be charged a fee for every message received, including spam. Mobile phone spam is generally less pervasive than email spam, where in 2010 around 90% of email is spam. The amount of mobile spam varies widely from region to region. In North America, mobile spam steadily increased after 2008 and accounted in terms of keeping alltypes of information. The system was verified. The form was checked out and the errors for half of all mobile phone traffic by 2019. In parts of Asia up to 30% of messages were spam in 2012.Candidatemanagement:

EXISTING SYSTEM

Theexistingssystem is a manual system which consists of different sub

modules. The dataset is readily available for this project which consists of more than 10,000 records. Now perform data preprocessing techniques to clean the dataset also removes the unnecessary words. Then applying machine learning models called KNN, LR and DT for this dataset to get the accurate result i.e., the message obtained is spam or ham.

PROPOSE SYSTEM

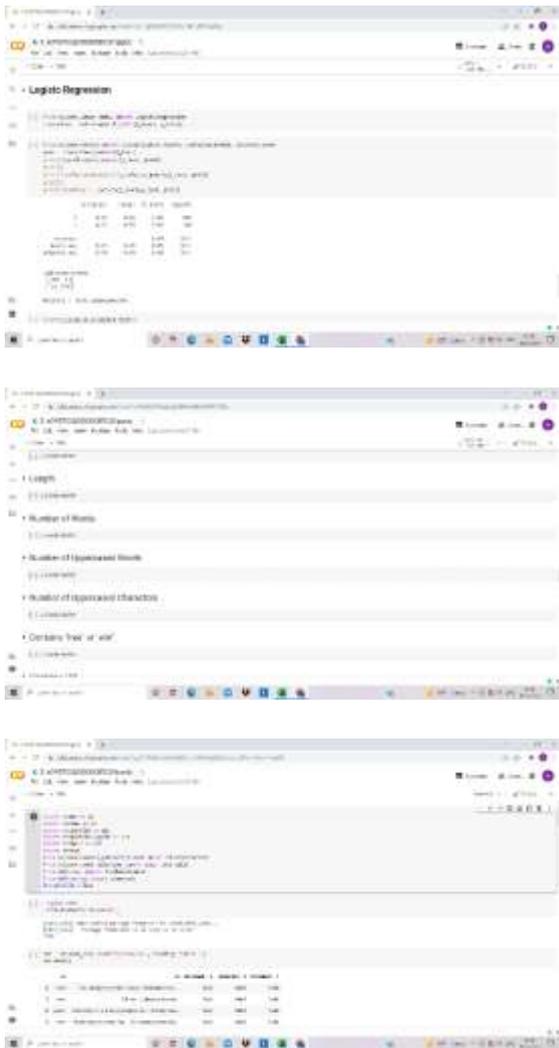
In this project, the useful information is available. So, we are adding auto-delete option to our project to delete the spam messages directly when the user wants to delete the spam message directly without seeing spam messages.

IMPLEMENTATION (modules)

Implementing through Machine learning classifiers such as machine learning classifiers such as Logistic regression (LR), K-nearest neighbor (K-NN), and decision tree (DT) are used for classification of ham and spam messages in mobile device communication. The SMS spam collection data set is used for testing the method. The dataset is split into two categories for training and testing the research. The results of the experiments demonstrated that the classification performance of LR is high as compared with K-NN and DT, and the

LR achieved a high accuracy of 99%. Additionally, the proposed method performance is good as compared with the existing state-of-the-art methods

SAMPLE SCREENS



CONCLUSION

Detection of spam is important for securing message communication. The accurate detection of spam is a big issue, and many detection methods have been proposed by various researchers. However,

these methods have a lack of capability to detect the spam accurately and efficiently. To solve this issue, we have proposed a method for spam detection using machine learning predictive models. The method is applied for the purpose of detection of spam. The experimental results obtained show that the proposed method has a high capability to detect spam. The proposed method achieved 99% accuracy which is high as compared with the other existing methods. Thus, the results suggest that the proposed method is more reliable for accurate and on-time detection of spam, and it will secure the communication systems of messages.

FUTURE SCOPE FOR FURTHER DEVELOPMENT

As of now there is no spam folder present. When the spam message came it will be mixed with the normal messages format. There is no separate folder for the spam messages in mobile message communication. So, now separate folder is developed in future for spam messages. So, users can identify the spam and ham messages easily in future.

REFERENCES

- [1]<https://iopscience.iop.org/article/10.1088/1742-6596/1797/1/012017/pdf>
- [2]<https://www.hindawi.com/journals/scn/2020/8873639/>

[3]<https://towardsdatascience.com/nlp-spam-detection-in-sms-text-data-using-deep-learning-b8632db85cc8>

[4]<http://cs229.stanford.edu/proj2013/ShiraniMehr-SMSSpamDetectionUsingMachineLearningApproach.pdf>

[5]<https://learn.vonage.com/blog/2020/11/19/sms-spam-detection-with-machine-learning-in-python/>

[6]<https://ieeexplore.ieee.org/document/9441783>

[7]<https://www.analyticsvidhya.com/blog/2021/05/sms-spam-detection-using-lstm-a-hands-on-guide/>

[8]https://www.researchgate.net/publication/327194370_SMS_Spam_Filtering_Using_Supervised_Machine_Learning_Algorithms

[9]https://www.researchgate.net/publication/352127853_SMS_Spam_Detection_using_Machine_Learning_and_Deep_Learning_Techniques

[10]www.wikipedia.com