SMS SPAM DETECTION

UNDER THE GUIDANCE OF

B.MAGESH M.C.A.,M.Sc.,B.Ed.

SUBMITTED BY

RAGUL. S

Reg No:30222U18017

ARAKKONAM ARTS AND SCIENCE COLLEGE

ABSTRACT

- SMS spam is increasing and causes security and privacy issues.
- A machine learning model is used to detect spam messages.
- The model is trained using a dataset of spam and normal messages.
- The system accurately detects spam messages and improves security.

INTRODUCTION

- SMS is widely used for communication in daily life.
- Many unwanted messages (spam) cause inconvenience and security risks.
- A smart system that learns from data to detect spam.
- Build an accurate spam detection system to improve security.

Existing System

- Many spammers distribute messages that contain URLs for malicious purposes.
- Many studies have been conducted using machine learning to detect spam automatically.

Proposed System

- The prediction method will employ 3 machine learning algorithms which are boosting algorithms
- Those algorithms such as Adaboost, Catboost algorithms

TESTING

- Tested the model on unseen SMS messages.
- Confusion matrix helped analyze correct and incorrect prediction.
- Model correctly classified most spam and not spam messages.
- Future improvement:Reduce false positives and negatives.

MODULES

- COLLECTING THE RAW DATA
- PRE-PROCESSING THE DATA
- SPLITING THE DATA
- EVALUATING THE MODEL

COLLECTING THE RAW DATA

- **Data collection** involves gathering information on fraudulent transactions from various sources.
- This collected data is used to **create machine learning models**.
- The data used in this work is related to **cervical cancer**, with specific features.

PRE PROCESSING THE DATA

- **Data pre-processing** involves formatting, cleaning, and sampling to organize the data.

Formatting:

- The data may not be in a usable format.
- You might need to change the data from a proprietary file format to a relational database or text file.

SPLITING THE DATA

- **Feature extraction** is the process of reducing attributes by changing them, unlike feature selection,which ranks existing attributes based on their relevance.
- **Pre-processed data** is categorized using machine learning methods.
- **Random forest classifiers** were chosen for this task.

EVALUTING THE MODEL

- Using training data for evaluation can lead to overly optimistic result.
- Splits data into training and test sets for evaluation.
- Ensure the model represents data well and predicts future performance.

SOFTWARE HARDWARE REQUIREMET

Software Requirement:

- Operating System: Window 7,8 and 10(32 and 64 bit)
- Front End: Html, CSS and Java Script
- Pakages:numpy,pandas,sklearn,keras,tensorflow
- Back End:Python and Data Set

Hardware Requirement:

- Processor – Dual Core
- Speed – 3.1 GHz
- RAM – 4GB
- Hard Disk – 200 GB

Future Enhancement

•**Real-time Learning**: Implementing a real-time learning system where

the model continuously improves by learning from new spam patterns and user feedback.

• **Improved Feature Engineering**: Developing more sophisticated features that analyze message metadata

along with text content to enhance spam detection accuracy..

Conclusion

The research aims at predicting the messages whether the messages are spam or true one and it is runs on efficient machine learning algorithms and technologies having an good accuracy. The training datasets so obtained provide the enough insights for predicting the appropriate messages . Thus, the system helps the users in identification of their messages whether they are spam messages or true messages with certain accurate prediction .