SPAM DETECTION

Spam messages are unwanted, unsolicited digital communications sent out in bulk. They can come in various forms, including:

* Emails
* Text messages
* Phone
* Social media

CHARACTERISTICS OF SPAM:

1.Unsolocited

2.Sent in bulk

3.Harmful

4.Deceptive

PREVENTIONS:

1.Never click on links or download attachments from unknown senders.

2.Use spam filters and firewalls to block unwanted messages

3.Don't provide your personal information in response to spam messages

4.Report spam to the appropriate authorities

SPAM DETECTION:

Spam detection refers to detecting spam messages or mails. It is crucial aspect for online safety. There are many techniques to detect the spam. In this, Machine learning technique is used.

NAÏVE BAYES CLASSIFIER:

Naïve Bayes is one of the machine learning technique used to detect spam messages. It is a probabilistic method based on bayes theorem. In this project, this algorithm to split into training and testing data then finds the spams.

PROCESS:

* First, import all the required libraries.
* Import the dataset
* Focus on only message and class which it belongs, neglect all other fields.
* Use CountVectorizer to convert text messages into numerical feature vectors.
* Fit the data which means it counts the occurrence of words and finds relation between unique words
* Split the dataset into training data and testing data
* Use Naïve bayes classifier
* Fit the training data using this algorithm to detect new messages.

CODE:

import pandas as pd

import numpy as np

from sklearn.feature\_extraction.text import CountVectorizer

from sklearn.model\_selection import train\_test\_split

from sklearn.naive\_bayes import MultinomialNB

data = pd.read\_csv("https://raw.githubusercontent.com/amankharwal/SMS-Spam-Detection/master/spam.csv", encoding= 'latin-1')

data.head()

x = np.array(data["message"])

y = np.array(data["class"])

cv = CountVectorizer()

X = cv.fit\_transform(x) # Fit the Data

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.33, random\_state=42)

clf = MultinomialNB()

clf.fit(X\_train,y\_train)

sample = input('Enter a message:')

data = cv.transform([sample]).toarray()

answer=clf.predict(data)

if answer=="spam":

  print("SPAM ALERT,SPAM ALERT!!!")

else:

  print("No worry,It is safe to reply!")