

# SMART

# AI

# SPAM

# DETECTOR



## **ABSTRACT:**

Spam emails have long been a pervasive problem in the digital landscape, necessitating advanced solutions for their detection and mitigation. In recent years, Artificial Intelligence (AI) has emerged as a powerful tool in the fight against spam. This paper introduces an AI-based smart spam detector, which leverages Natural Language Processing (NLP) and Machine Learning techniques to enhance the accuracy of spam classification.

The proposed system incorporates a deep learning model that is trained on a vast dataset of both spam and legitimate emails, allowing it to learn the nuances of spammy content and distinguish it from genuine communication. By utilizing NLP algorithms, the model can analyze email text for various spam-related patterns, including phishing attempts, keyword-based triggers, and grammatical anomalies. Additionally, the system adapts to evolving spam tactics through continuous learning and model retraining.

To evaluate the system's effectiveness, a comprehensive set of experiments is conducted, demonstrating its ability to achieve high accuracy in spam detection while minimizing

false positives. Moreover, the system's real-time detection capabilities are demonstrated, showcasing its potential for integration into email clients and other communication platforms.

In conclusion, this AI-powered smart spam detector offers an innovative approach to tackling the persistent issue of spam. Its utilization of AI and NLP technologies, combined with continuous learning, results in a reliable and adaptive solution for organizations and individuals seeking to protect their inboxes from unwanted and potentially harmful content.

## INTRODUCTION:

The proliferation of email as a primary means of communication has brought with it a significant challenge: the incessant influx of spam. Unwanted emails, often riddled with scams, advertisements, and malicious content, can clog inboxes, jeopardize data security, and hinder productive communication.

Conventional rule-based spam filters have proven inadequate in keeping up with the ever-evolving tactics of

spammers. This has paved the way for a more sophisticated solution – an AI-powered smart spam detector.

Artificial Intelligence (AI), particularly in the realms of Natural Language Processing (NLP) and Machine Learning, has emerged as a formidable ally in the fight against spam. This detector leverages the power of AI to not only identify known spam patterns but also adapt to new and emerging spam techniques. By analyzing the content, structure, and context of emails, it can differentiate between genuine communication and spam with a high degree of accuracy.

This paper delves into the design, development, and implementation of such an AI-based smart spam detector. It explores the utilization of machine learning models to classify emails, the role of NLP in understanding the language of spam, and the mechanisms in place to continuously improve the detector's performance. The aim is to provide a comprehensive overview of how AI technology is

## **DATA PROCESSING:**

**Data processing is a fundamental component of any AI-based smart spam detector. Here's how data processing fits into the operation of such a system:**

### **DATA COLLECTION:**

The system collects a large and diverse dataset of emails, including both spam and legitimate messages. This dataset is crucial for training and testing the AI model.

### **DATA PROCESSING:**

The collected data goes through preprocessing steps to clean and prepare it for analysis. This may include removing special characters, formatting text, and handling attachments or HTML content.

### **FEATURE EXTRACTION:**

Feature extraction involves identifying relevant attributes or features from the email content that can be used for classification. Features could include text content, sender information, subject lines, and more.

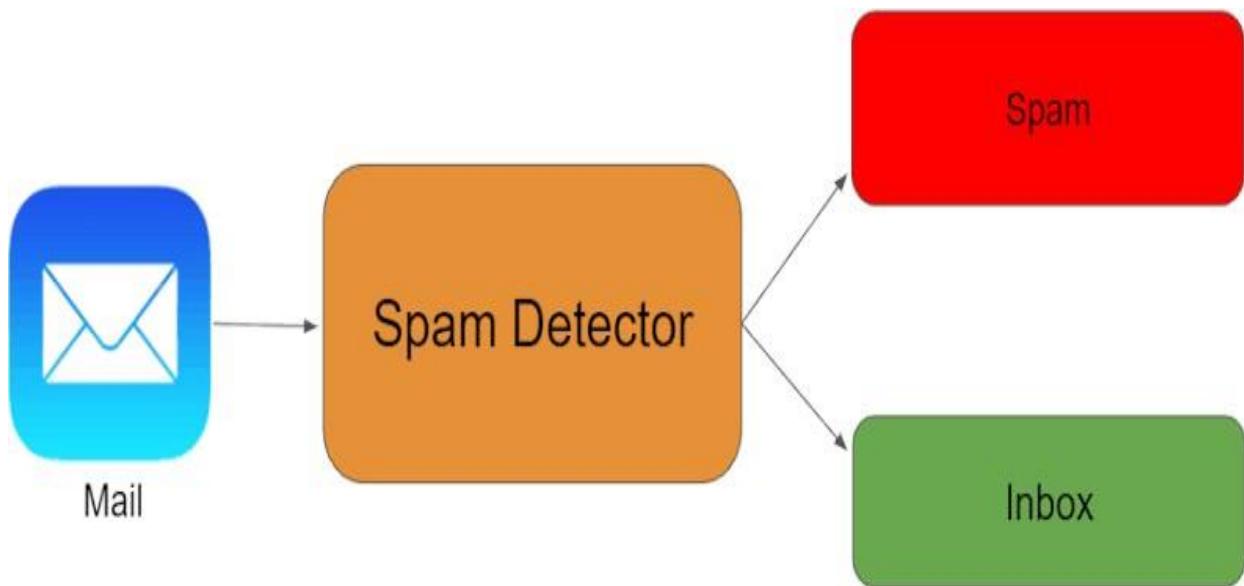
## MODAL EVALUATION:

The model's performance is assessed using the testing set. Metrics like accuracy, precision, recall, and F1 score are used to gauge its effectiveness in spam detection.

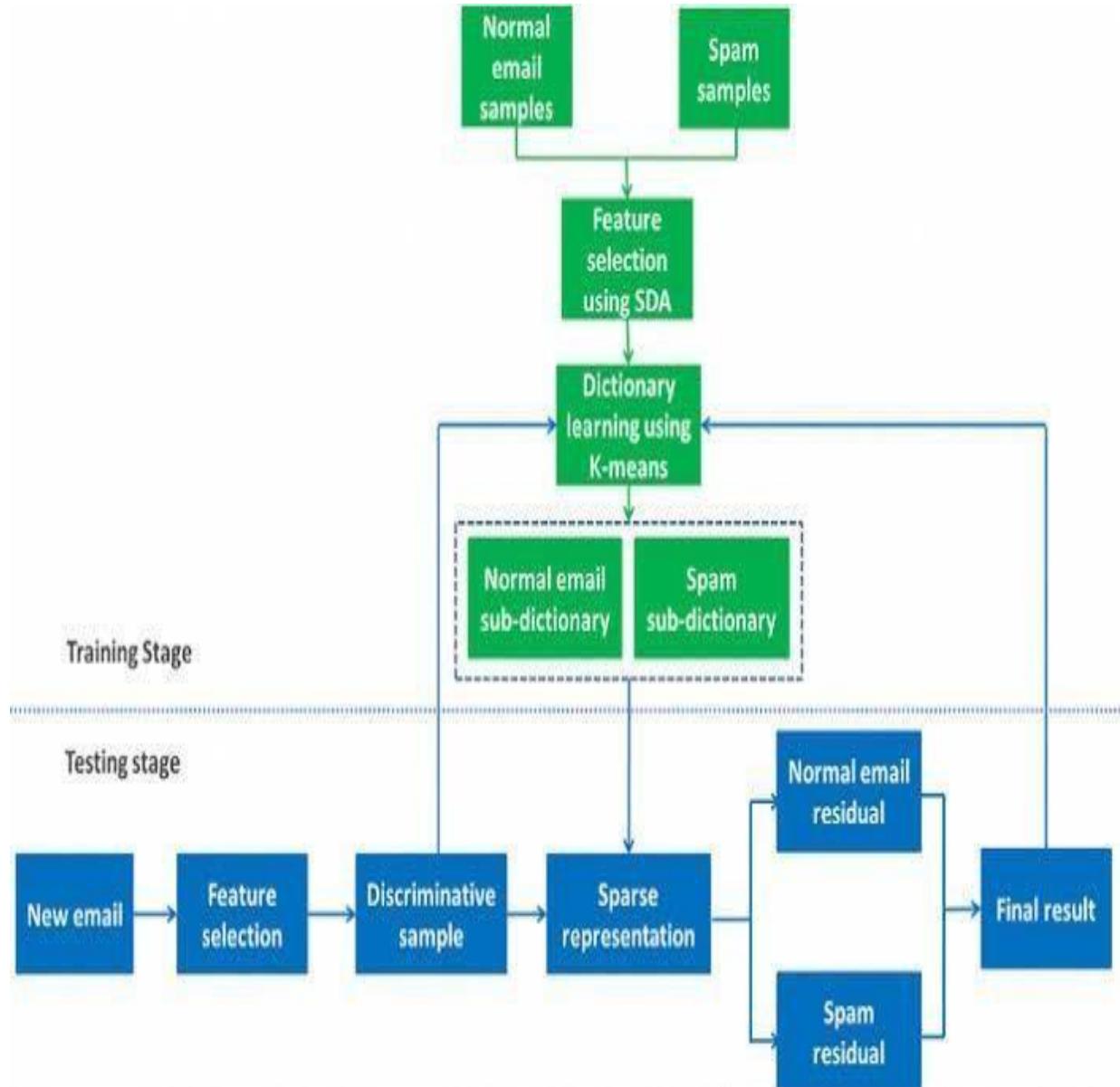
## OUTPUT GENERATION:

The final step in data processing is generating an output that classifies each email as either spam or legitimate. This classification may be used to filter or flag emails in users' inboxes.

Data processing is a critical aspect of building an effective AI-based smart spam detector, as the quality of the data and the processing techniques directly impact the system's ability to accurately identify and filter spam emails.



# FLOWCHART:



## PROGRAM:

```
import os

from flask import Flask, render_template, request, redirect, url_for, jsonify

from sklearn.feature_extraction.text import TfidfVectorizer

from sklearn.multiclass import *

from sklearn.svm import *

import pandas

app = Flask(__name__)

global Classifier

global Vectorizer


# load data

data = pandas.read_csv("spam.csv", encoding="latin-1")

train_data = data[:4400] # 4400 items

test_data = data[4400:] # 1172 items


# train model

Classifier = OneVsRestClassifier(SVC(kernel="linear", probability=True))

Vectorizer = TfidfVectorizer()

vectorize_text = Vectorizer.fit_transform(train_data.v2)

Classifier.fit(vectorize_text, train_data.v1)
```

```
@app.route("/", methods=["GET"])

def index():

    message = request.args.get('message', '')
    error = ''
    predict_proba = ''
    predict = ''

    global Classifier
    global Vectorizer

    try:
        if len(message) > 0:
            vectorize_message = Vectorizer.transform([message])
            predict = Classifier.predict(vectorize_message)[0]
            predict_proba = Classifier.predict_proba(vectorize_message).tolist()
    except BaseException as inst:
        error = str(type(inst).__name__) + ' ' + str(inst)

    return jsonify(
        message=message, predict_proba=predict_proba,
        predict=predict, error=error)

if __name__ == '__main__':
    port = int(os.environ.get('PORT', 5000))
```

```
app.run(host='0.0.0.0', port=port, debug=True, use_reloader=True)
```

## CLASIFIERS:

```
from sklearn.naive_bayes import *
from sklearn.dummy import *
from sklearn.ensemble import *
from sklearn.neighbors import *
from sklearn.tree import *
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.feature_extraction.text import HashingVectorizer
from sklearn.calibration import *
from sklearn.linear_model import *
from sklearn.multiclass import *
from sklearn.svm import *
import pandas
```

```
def perform(classifiers, vectorizers, train_data, test_data):
    for classifier in classifiers:
        for vectorizer in vectorizers:
            string = ""
            string += classifier.__class__.__name__ + ' with ' +
vectorizer.__class__.__name__

            # train
            vectorize_text = vectorizer.fit_transform(train_data.v2)
            classifier.fit(vectorize_text, train_data.v1)

            # score
            vectorize_text = vectorizer.transform(test_data.v2)
            score = classifier.score(vectorize_text, test_data.v1)
            string += '. Has score: ' + str(score)
            print(string)
```

```
data = pandas.read_csv('spam.csv', encoding='latin-1')
learn = data[:4400] # 4400 items
test = data[4400:] # 1172 items

perform(
[
```

```

BernoulliNB(),
RandomForestClassifier(n_estimators=100, n_jobs=-1),
AdaBoostClassifier(),
BaggingClassifier(),
ExtraTreesClassifier(),
GradientBoostingClassifier(),
DecisionTreeClassifier(),
CalibratedClassifierCV(),
DummyClassifier(),
PassiveAggressiveClassifier(),
RidgeClassifier(),
RidgeClassifierCV(),
SGDClassifier(),
OneVsRestClassifier(SVC(kernel='linear')),
OneVsRestClassifier(LogisticRegression()),
KNeighborsClassifier()

],  

[  

CountVectorizer(),
TfidfVectorizer(),
HashingVectorizer()
],  

learn,  

test
)

```

## SKIKIT:

```

from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.calibration import *
from sklearn.linear_model import *
from sklearn.multiclass import *
from sklearn.svm import *
import pandas

data = pandas.read_csv('spam.csv', encoding='latin-1')
train_data = data[:4400] # 4400 items
test_data = data[4400:] # 1172 items

classifier = OneVsRestClassifier(SVC(kernel='linear'))
vectorizer = TfidfVectorizer()

```

```

# train
vectorize_text = vectorizer.fit_transform(train_data.v2)
classifier.fit(vectorize_text, train_data.v1)

vectorize_text = vectorizer.transform(test_data.v2)
score = classifier.score(vectorize_text, test_data.v1)
print(score) # 98,8

```

# SPAM.CSV

v	v2
1	
2	ha m Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got amore wat...
3	ha m Ok lar... Joking wif u oni...
4	spa Free entry in 2 a wkly comp to win FA Cup final tkts 21st May 2005. Text FA to 87121 to receive entry question(std txt rate)T&C's m apply 08452810075over18's
5	ha m U dun say so early hor... U c already then say...
6	ha m Nah I don't think he goes to usf, he lives around here though
7	spa FreeMsg Hey there darling it's been 3 week's now and no word back! I'd like some fun you up for it still? Tb ok! XxX std chgs to m send, £1.50 to rcv
8	ha m Even my brother is not like to speak with me. They treat me like aids patient.
9	ha m As per your request 'Melle Melle (Oru Minnaminunginte Nurungu Vettam)' has been set as your callertune for all Callers. Press *9 to copy your friends Callertune
10	spa WINNER!! As a valued network customer you have been selected to receivea £900 prize reward! To claim call 09061701461. m Claim code KL341. Valid 12 hours only.
11	spa Had your mobile 11 months or more? U R entitled to Update to the latest colour mobiles with camera for Free! Call The Mobile m Update Co FREE on 08002986030
12	ha m I'm gonna be home soon and i don't want to talk about this stuff anymore tonight, k? I've cried enough today.
13	spa SIX chances to win CASH! From 100 to 20,000 pounds txt> CSH11 and send to 87575. Cost 150p/day, 6days, 16+ TsandCs apply m Reply HL 4 info
14	spa URGENT! You have won a 1 week FREE membership in our £100,000 Prize Jackpot! Txt the word: CLAIM to No: 81010 T&C m www.dbuk.net LCCLTD POBOX 4403LDNW1A7RW18
15	ha m I've been searching for the right words to thank you for this breather. I promise i wont take your help for granted and will fulfil my promise. You have been wonderful and a blessing at all times.
16	ha m I HAVE A DATE ON SUNDAY WITH WILL!!

v  
1 v2

- 17 spa XXXMobileMovieClub: To use your credit, click the WAP link in the next txt message or click here> > http://wap.  
m xxmobilemovieclub.com?n=QJKGIGHJJGCBL
- 18 ha  
m Oh k...i'm watching here:)
- 19 ha  
m Eh u remember how 2 spell his name... Yes i did. He v naughty make until i v wet.
- 20 ha  
m Fine if thatâ's the way u feel. Thatâ's the way its gotta b
- 21 spa England v Macedonia - dont miss the goals/team news. Txt ur national team to 87077 eg ENGLAND to 87077 Try:WALES,  
m SCOTLAND 4txt/l/41.20 POBOXox36504W45WQ 16+
- 22 ha  
m Is that seriously how you spell his name?
- 23 ha  
m I‰ÛÒm going to try for 2 months ha ha only joking
- 24 ha  
m So l\_ pay first lar... Then when is da stock comin...
- 25 ha  
m Aft i finish my lunch then i go str down lor. Ard 3 smth lor. U finish ur lunch already?
- 26 ha  
m Ffffffff. Alright no way I can meet up with you sooner?
- 27 ha  
m Just forced myself to eat a slice. I'm really not hungry tho. This sucks. Mark is getting worried. He knows I'm sick when I turn down  
pizza. Lol
- 28 ha  
m Lol your always so convincing.
- 29 ha  
m Did you catch the bus ? Are you frying an egg ? Did you make a tea? Are you eating your mom's left over dinner ? Do you feel my  
Love ?
- 30 ha  
m I'm back & we're packing the car now, I'll let you know if there's room
- 31 ha  
m Ahhh. Work. I vaguely remember that! What does it feel like? Lol
- 32 ha  
m Wait that's still not all that clear, were you not sure about me being sarcastic or that that's why x doesn't want to live with us
- 33 ha  
m Yeah he got in at 2 and was v apologetic. n had fallen out and she was actin like spoilt child and he got caught up in that. Till 2!  
But we won't go there! Not doing too badly cheers. You?
- 34 ha  
m K tell me anything about you.
- 35 ha  
m For fear of fainting with the of all that housework you just did? Quick have a cuppa
- 36 spa Thanks for your subscription to Ringtone UK your mobile will be charged £5/month Please confirm by replying YES or NO. If you  
m reply NO you will not be charged
- 37 ha  
m Yup... Ok i go home look at the tim

# TEXT SCORE:

```
from sklearn.naive_bayes import *
from sklearn.dummy import *
from sklearn.ensemble import *
from sklearn.neighbors import *
from sklearn.tree import *
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.feature_extraction.text import HashingVectorizer
from sklearn.calibration import *
from sklearn.linear_model import *
from sklearn.multiclass import *
from sklearn.svm import *
import pandas
import csv

data = pandas.read_csv('spam.csv', encoding='latin-1')
train_data = data[:4400] # 4400 items
test_data = data[4400:] # 1172 items

classifier = OneVsRestClassifier(SVC(kernel='linear'))
vectorizer = TfidfVectorizer()

# train
vectorize_text = vectorizer.fit_transform(train_data.v2)
classifier.fit(vectorize_text, train_data.v1)

# score
# vectorize_text = vectorizer.transform(test_data.v2)
# score = classifier.score(vectorize_text, test_data.v1)
# print(score) # 98,8

csv_arr = []
for index, row in test_data.iterrows():
    answer = row[0]
    text = row[1]
    vectorize_text = vectorizer.transform([text])
    predict = classifier.predict(vectorize_text)[0]
    if predict == answer:
        result = 'right'
    else:
        result = 'wrong'
    csv_arr.append([len(csv_arr), text, answer, predict, result])
```

```
# write csv
with open('test_score.csv', 'w', newline='') as csvfile:
    spamwriter = csv.writer(csvfile, delimiter=',',
                           quotechar='"', quoting=csv.QUOTE_MINIMAL)
    spamwriter.writerow(['#', 'text', 'answer', 'predict', result])

    for row in csv_arr:
        spamwriter.writerow(row)
```