

BUILDING A SMARTER AI-POWERED SPAM IDENTIFIER
PHASE 4 SUBMISSION

SUBMITTED BY

Name:R.Venuvaneshwari

Reg.no:912421104053

FEATURE EXTRACTION AND MODEL PREDICTION:

Data source:

Data link:(<https://www.kaggle.com/datasets/uciml/sms-spam-collection-dataset>)

Data set:

class

text_formatv2sort

sms

text_formatsort

text_formatsort

ham87%

spam13%

5169

unique values

[null]99%

bt not his girlfrnd... G o o d n i g h t . . . @ " 0 %

Other (47)1%

[null]100%

MK17 92H. 450Ppw 16"0%

Other (10)0%

ham

Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got a...

ham

Ok lar... Joking wif u oni...

spam

Free entry in 2 a wkly comp to win FA Cup final tkts 21st May 2005. Text FA to 87121 to receive entr...

ham

U dun say so early hor... U c already then say...

ham

Nah I don't think he goes to usf, he lives around here though

spam

FreeMsg Hey there darling it's been 3 week's now and no word back! I'd like some fun you up for it s...

ham

Even my brother is not like to speak with me. They treat me like aids patient.

ham

As per your request 'Melle Melle (Oru Minnaminunginte Nurungu Vettam)' has been set as your callertu...

spam

WINNER!! As a valued network customer you have been selected to receive a ☐ 900 prize reward! To claim...

spam

Had your mobile 11 months or more? U R entitled to Update to the latest colour mobiles with camera f...

ham

I'm gonna be home soon and i don't want to talk about this stuff anymore tonight, k? I've cried enou...

spam

SIX chances to win CASH! From 100 to 20,000 pounds txt> CSH11 and send to 87575. Cost 150p/day, 6day...

spam

URGENT! You have won a 1 week FREE membership in our £100,000 Prize Jackpot! Txt the word: CLAIM to ...

ham

I've been searching for the right words to thank you for this breather. I promise i wont take your h...

ham

I HAVE A DATE ON SUNDAY WITH WILL!!

spam

XXXMobileMovieClub: To use your credit, click the WAP link in the next txt message or click here>> h...

ham

Oh k...i'm watching here:)

ham

Eh u remember how 2 spell his name... Yes i did. He v naughty make until i v wet.

ham

Fine if that□□s the way u feel. That□□s the way its gota b

spam

England v Macedonia - dont miss the goals/team news. Txt ur national team to 87077 eg ENGLAND to 870...

ham

Is that seriously how you spell his name?

ham

I□□□m going to try for 2 months ha ha only joking

ham

So □_ pay first lar... Then when is da stock comin...

ham

Aft i finish my lunch then i go str down lor. Ard 3 smth lor. U finish ur lunch already?

ham

Ffffffffff. Alright no way I can meet up with you sooner?

ham

Just forced myself to eat a slice. I'm really not hungry tho. This sucks. Mark is getting worried. H...

ham

Lol your always so convincing.

ham

Did you catch the bus ? Are you frying an egg ? Did you make a tea? Are you eating your mom's left o...

ham

I'm back & we're packing the car now, I'll let you know if there's room

ham

Ahhh. Work. I vaguely remember that! What does it feel like? Lol

ham

Wait that's still not all that clear, were you not sure about me being sarcastic or that that's why ...

ham

Yeah he got in at 2 and was v apologetic. n had fallen out and she was actin like spoilt child and h...

ham

K tell me anything about you.

ham

For fear of fainting with the of all that housework you just did? Quick have a cuppa

spam

Thanks for your subscription to Ringtone UK your mobile will be charged £5/month Please confirm by r...

ham

Yup... Ok i go home look at the timings then i msg ☐_ again... Xuhui going to learn on 2nd may too b...

ham

Oops, I'll let you know when my roommate's done

ham

I see the letter B on my car

ham

Anything lor... U decide...

ham

Hello! How's you and how did saturday go? I was just texting to see if you'd decided to do anything ...

ham

Pls go ahead with watts. I just wanted to be sure. Do have a great weekend. Abiola

ham

Did I forget to tell you ? I want you , I need you, I crave you ... But most of all ... I love you m...

spam

07732584351 - Rodger Burns - MSG = We tried to call you re your reply to our sms for a free nokia mo...

ham

WHO ARE YOU SEEING?

ham

Great! I hope you like your man well endowed. I am <#> inches...

ham

No calls..messages..missed calls

ham

Didn't you get hep b immunisation in nigeria.

ham

Fair enough, anything going on?

ham

Yeah hopefully, if tyler can't do it I could maybe ask around a bit

ham

U don't know how stubborn I am. I didn't even want to go to the hospital. I kept telling Mark I'm no...

ham

What you thinked about me. First time you saw me in class.

ham

A gram usually runs like <#> , a half eighth is smarter though and gets you almost a whole se...

ham

K fyi x has a ride early tomorrow morning but he's crashing at our place tonight

ham

Wow. I never realized that you were so embarassed by your accomodations. I thought you liked it, sin...

spam

SMS. ac Sptv: The New Jersey Devils and the Detroit Red Wings play Ice Hockey. Correct or Incorrect?...

ham

Do you know what Mallika Sherawat did yesterday? Find out now @ <URL>

spam

Congrats! 1 year special cinema pass for 2 is yours. call 09061209465 now! C Suprman V, Matrix3, Sta...

ham

Sorry, I'll call later in meeting.

ham

Tell where you reached

ham

Yes..gauti and sehwaq out of odi series.

ham

Your gonna have to pick up a \$1 burger for yourself on your way home. I can't even move. Pain is kil...

ham

Ha ha ha good joke. Girls are situation seekers.

ham

Its a part of checking IQ

ham

Sorry my roommates took forever, it ok if I come by now?

ham

Ok lar i double check wif da hair dresser already he said wun cut v short. He said will cut until i ...

spam

As a valued customer, I am pleased to advise you that following recent review of your Mob No. you ar...

ham

Today is \song dedicated day..\" Which song will u dedicate for me? Send this to all ur valuable frn...

spam

Urgent UR awarded a complimentary trip to EuroDisinc Trav, Aco&Entry41 Or □1000. To claim txt DIS to...

spam

Did you hear about the new \Divorce Barbie\"? It comes with all of Ken's stuff!"

ham

I plane to give on this month end.

ham

Wah lucky man... Then can save money... Hee...

ham

Finished class where are you.

ham

HI BABE IM AT HOME NOW WANNA DO SOMETHING? XX

ham

K..k:)where are you?how did you performed?

ham

U can call me now...

ham

I am waiting machan. Call me once you free.

ham

Thats cool. i am a gentleman and will treat you with dignity and respect.

ham

I like you peoples very much:) but am very shy pa.

ham

Does not operate after <#> or what

ham

Its not the same here. Still looking for a job. How much do Ta's earn there.

ham

Sorry, I'll call later

ham

K. Did you call me just now ah?

ham

Ok i am on the way to home hi hi

ham

You will be in the place of that man

ham

Yup next stop.

ham

I call you later, don't have network. If urgnt, sms me.

ham

For real when u getting on yo? I only need 2 more tickets and one more jacket and I'm done. I ahead...

ham

Yes I started to send requests to make it but pain came back so I'm back in bed. Double coins at the...

ham

I'm really not up to it still tonight babe

ham

Ela kano.,il download, come wen ur free..

ham

Yeah do! Don't stand to close tho- you'll catch something!

ham

Sorry to be a pain. Is it ok if we meet another night? I spent late afternoon in casualty and that m...

ham

Smile in Pleasure Smile in Pain Smile when trouble pours like Rain Smile when sum1 Hurts U Smile bec...

spam

Please call our customer service representative on 0800 169 6031 between 10am-9pm as you have WON a ...

ham

Havent planning to buy later. I check already lido only got 530 show in e afternoon. U finish work a...

spam

Your free ringtone is waiting to be collected. Simply text the password \MIX\" to 85069 to verify. G...

PO Box 5249

MK17 92H. 450Ppw 16"

ham

Watching telugu movie..wat abt u?

ham

i see. When we finish we have loads of loans to pay

ham

Hi. Wk been ok - on hols now! Yes on for a bit of a run. Forgot that i have hairdressers appointment...

ham

Please don't text me anymore. I have nothing else to say.

ham

Okay name ur price as long as its legal! Wen can I pick them up? Y u ave x ams xx

ham

I'm still looking for a car to buy. And have not gone 4the driving test yet.

ham

As per your request 'Melle Melle (Oru Minnaminunginte Nurungu Vettam)' has been set as your callertu...

ham

wow. You're right! I didn't mean to do that. I guess once i gave up on boston men and changed my sea...

ham

Umma my life and vava umma love you lot dear

ham

Thanks a lot for your wishes on my birthday. Thanks you for making my birthday truly memorable.

ham

Aight, I'll hit you up when I get some cash

ham

How would my ip address test that considering my computer isn't a minecraft server

ham

I know! Grumpy old people. My mom was like you better not be lying. Then again I am always the one t...

ham

Dont worry. I guess he's busy.

ham

What is the plural of the noun research?

ham

Going for dinner.msg you after.

ham

I'm ok wif it cos i like 2 try new things. But i scared u dun like mah. Cos u said not too loud.

spam

GENT! We are trying to contact you. Last weekends draw shows that you won a £1000 prize GUARANTEED. ...

ham

Wa, ur openin sentence very formal... Anyway, i'm fine too, juz tt i'm eatin too much n puttin on we...

ham

As I entered my cabin my PA said, " Happy B'day Boss !!!". I felt special. She askd me 4 lunch. Aft...

spam

You are a winner U have been specially selected 2 receive ₦1000 or a 4* holiday (flights inc) speak ...

ham

Goodo! Yes we must speak friday - egg-potato ratio for tortilla needed!

ham

Hmm...my uncle just informed me that he's paying the school directly. So pls buy food.

spam

PRIVATE! Your 2004 Account Statement for 07742676969 shows 786 unredeemed Bonus Points. To claim cal...

spam

URGENT! Your Mobile No. was awarded ₦2000 Bonus Caller Prize on 5/9/03 This is our final try to cont...

ham

here is my new address -apples&pairs&all that malarky

spam

Todays Voda numbers ending 7548 are selected to receive a \$350 award. If you have a match please cal...

ham

I am going to sao mu today. Will be done only at 12

ham

□□ predict wat time □_ 'll finish buying?

ham

Good stuff, will do.

ham

Just so that you know,yetunde hasn't sent money yet. I just sent her a text not to bother sending. S...

ham

Are you there in room.

ham

HEY GIRL. HOW R U? HOPE U R WELL ME AN DEL R BAK! AGAIN LONG TIME NO C!
GIVE ME A CALL SUM TIME FROM...

ham

K..k:)how much does it cost?

ham

I'm home.

ham

Dear, will call Tmorrow.pls accomodate.

ham

First answer my question.

spam

Sunshine Quiz Wkly Q! Win a top Sony DVD player if u know which country the Algarve is in?
Txt ansr ...

spam

Want 2 get laid tonight? Want real Dogging locations sent direct 2 ur mob? Join the UK's largest
Dog...

ham

I only haf msn. It's yijue@hotmail.com

ham

He is there. You call and meet him

ham

No no. I will check all rooms befor activities

spam

You'll not rev any more msgs from the chat svc. For FREE Hardcore services text GO to: 69988
If u ge...

ham

Got c... I lazy to type... I forgot □_ in lect... I saw a pouch but like not v nice...

ham

K, text me when you're on the way

ham

Sir, Waiting for your mail.

ham

A swt thought: \Nver get tired of doing little things 4 lovable persons..\'' Coz..sometmes those litt...

ham

I know you are. Can you pls open the back?

ham

Yes see ya not on the dot

ham

Whats the staff name who is taking class for us?

spam

FreeMsg Why haven't you replied to my text? I'm Randy, sexy, female and live local. Luv to hear from...

ham

Umma.will call after check in.our life will begin from qatar so pls pray very hard.

ham

K..i deleted my contact that why?

ham

Sindu got job in birla soft ..

ham

The wine is flowing and i'm i have nevering..

ham

Yup i thk cine is better cos no need 2 go down 2 plaza mah.

ham

Ok... Ur typical reply...

ham

As per your request 'Melle Melle (Oru Minnaminunginte Nurungu Vettam)' has been set as your callertu...

ham

You are everywhere dirt, on the floor, the windows, even on my shirt. And sometimes when i open my m...

ham

Aaoooooright are you at work?

ham

I'm leaving my house now...

ham

Hello, my love. What are you doing? Did you get to that interview today? Are you you happy? Are you ...

spam

Customer service announcement. You have a New Years delivery waiting for you. Please call 07046744435...

spam

You are a winner U have been specially selected 2 receive □ 1000 cash or a 4* holiday (flights inc) s...

ham

Keep yourself safe for me because I need you and I miss you already and I envy everyone that see's y...

ham

New car and house for my parents.:)i have only new job in hand:)

ham

I'm so in love with you. I'm excited each day i spend with you. You make me so happy.

spam

-PLS STOP bootydelious (32/F) is inviting you to be her friend. Reply YES-434 or NO-434 See her: www...

spam

BangBabes Ur order is on the way. U SHOULD receive a Service Msg 2 download UR content. If U do not,...

ham

I place all ur points on e cultures module already.

spam

URGENT! We are trying to contact you. Last weekends draw shows that you have won a £900 prize GUARAN...

ham

Hi frnd, which is best way to avoid missunderstnding wit our beloved one's?

ham

Great escape. I fancy the bridge but needs her lager. See you tomo

ham

Yes :)it completely in out of form:)clark also utter waste.

ham

Sir, I need AXIS BANK account no and bank address.

ham

Hmmm.. Thk sure got time to hop ard... Ya, can go 4 free abt... Muz call u to discuss liao...

ham

What time you coming down later?

ham

Bloody hell, cant believe you forgot my surname Mr . Ill give u a clue, its spanish and begins with ...

ham

Well, i'm gonna finish my bath now. Have a good...fine night.

ham

Let me know when you've got the money so carlos can make the call

ham

U still going to the mall?

ham

Turns out my friends are staying for the whole show and won't be back til ~ – , so feel free...

ham

Text her. If she doesnt reply let me know so i can have her log in

ham

Hi! You just spoke to MANEESHA V. We'd like to know if you were satisfied with the experience. Reply...

ham

You lifted my hopes with the offer of money. I am in need. Especially when the end of the month appr...

ham

Lol no. U can trust me.

ham

ok. I am a gentleman and will treat you with dignity and respect.

ham

He will, you guys close?

ham

Going on nothing great.bye

ham

Hello handsome ! Are you finding that job ? Not being lazy ? Working towards getting back that net f...

ham

Haha awesome, be there in a minute

spam

Please call our customer service representative on FREEPHONE 0808 145 4742 between 9am-11pm as you h...

ham

Have you got Xmas radio times. If not i will get it now

ham

I jus reached home. I go bathe first. But my sis using net tell u when she finishes k...

spam

Are you unique enough? Find out from 30th August. www.areyouunique.co.uk

ham

I'm sorry. I've joined the league of people that dont keep in touch. You mean a great deal to me. Yo...

ham

Hi :)finally i completed the course:)

ham

It will stop on itself. I however suggest she stays with someone that will be able to give ors for e...

ham

How are you doing? Hope you've settled in for the new school year. Just wishin you a gr8 day

ham

Gud mrng dear hav a nice day

ham

Did u got that persons story

ham

is your hamster dead? Hey so tmr i meet you at 1pm orchard mrt?

ham

Hi its Kate how is your evening? I hope i can see you tomorrow for a bit but i have to bloody babyjo...

ham

Found it, ENC & , where you at?

ham

I sent you & bucks

ham

Hello darlin ive finished college now so txt me when u finish if u can love Kate xxx

ham

Your account has been refilled successfully by INR & . Your KeralaCircle prepaid acco...

ham

Goodmorning sleeping ga.

ham

U call me alter at 11 ok.

ham

say until like dat i dun buy ericsson oso cannot oredi lar...

ham

As I entered my cabin my PA said, " Happy B'day Boss !!!". I felt special. She askd me 4 lunch. Aft...

ham

Aight yo, dats straight dogg

ham

You please give us connection today itself before & or refund the bill

ham

Both :) i shoot big loads so get ready!

ham

What's up bruv, hope you had a great break. Do have a rewarding semester.

ham

Home so we can always chat

ham

K:)k:)good:)study well.

ham

Yup... How □_ noe leh...

ham

Sounds great! Are you home now?

ham

Finally the match heading towards draw as your prediction.

ham

Tired. I haven't slept well the past few nights.

ham

Easy ah?sen got selected means its good..

ham

I have to take exam with march 3

ham

Yeah you should. I think you can use your gt atm now to register. Not sure but if there's anyway i c...

ham

Ok no prob. Take ur time.

ham

There is os called ubandu which will run without installing in hard disk...you can use that os to co...

ham

Sorry, I'll call later

ham

U say leh... Of course nothing happen lar. Not say v romantic jus a bit only lor. I thk e nite scene...

spam

500 New Mobiles from 2004, MUST GO! Txt: NOKIA to No: 89545 & collect yours today!From ONLY £1 www.4...

ham

Would really appreciate if you call me. Just need someone to talk to.

spam

Will u meet ur dream partner soon? Is ur career off 2 a flyng start? 2 find out free, txt HORO follo...

ham

Hey company elama po mudyadhu.

FEATURE EXTRACTION

Feature extraction is a critical step in building a smarter AI-powered spam identifier. The goal is to convert raw text data into a format that machine learning models can understand and use for classification. Here are some common feature extraction techniques you can consider for my project:

1. Bag of Words (BoW):

- Create a vocabulary of unique words from your dataset.
- Represent each document as a vector, where each element corresponds to the frequency of a word in the document.
- Use techniques like TF-IDF (Term Frequency-Inverse Document Frequency) to give more weight to important words.

2. N-grams:

- Instead of individual words, consider combinations of words (n-grams) as features.
- For example, you might use bigrams (pairs of adjacent words) or trigrams (triplets of adjacent words) as features.

3. Word Embeddings:

- Utilize pre-trained word embeddings such as Word2Vec, GloVe, or FastText.

- These embeddings can capture semantic relationships between words and can be used to represent documents as dense vectors.

4. **Character-Level Features:**

- Extract character-level n-grams or sequences to capture patterns at a finer level.
- This can be useful for detecting spammy text that may contain irregular character patterns.

5. **Sender Information:**

- Include features related to the sender's email address or domain.
- Features might include the sender's reputation, known spammy domains, or IP address information.

6. **Message Metadata:**

- Extract features from message metadata, such as the timestamp, subject, and message length.
- Spam messages often exhibit certain patterns in these metadata.

7. **Text Statistics:**

- Calculate statistics about the text, such as the number of uppercase letters, special characters, or links in the message.
- Spam messages often use excessive capitalization and special characters.

8. **Language Features:**

- Consider linguistic features like the language of the text or the readability score.
- Some spam messages are written in non-standard or poorly constructed language.

9. **Contextual Features:**

- Capture the context of the message, including previous interactions with the sender and the user's profile.
- This can help identify spam messages that may appear legitimate based on context.

10. **Topic Modeling:**

- Apply topic modeling techniques such as Latent Dirichlet Allocation (LDA) to identify the main topics in a message.
- Spam messages may contain distinct topics related to various scams or promotions.

11. **Network Analysis:**

- Analyze the network relationships between senders and recipients to identify unusual patterns.

- For example, if a sender is sending messages to a large number of recipients, it may be indicative of spam.

12. **Semantic Analysis:**

- Use semantic analysis techniques to understand the meaning of the text.
- This can help in identifying spam messages that use obfuscation techniques to avoid simple keyword-based detection.

13. **User Behavior Features:**

- Consider the behavior of the user, such as their history of marking messages as spam or the frequency of opening messages from specific senders.

14. **External Data:**

- Incorporate external data sources, such as blacklists of known spammers or known malicious domains.

15. **Text Vectorization Techniques:**

- Explore advanced text vectorization techniques, such as Doc2Vec or Word Movers' Distance, which capture semantic inform

Model prediction:

Building a smarter AI-powered spam identifier is an ambitious project that can help improve email and message filtering systems. To create an effective AI model for spam identification, you'll need to follow a systematic approach. Here's a prediction of the key components and steps you might consider:

1. **Data Collection:**

- Gather a large and diverse dataset of emails, messages, or communication data that includes both spam and non-spam examples. The dataset should cover various types of spam, such as phishing, promotional, and fraudulent messages.

2. **Data Preprocessing:**

- Clean and preprocess the data, which may involve tasks like text normalization, removing special characters, and tokenization.

3. **Feature Engineering:**

- Extract relevant features from the text data. Common features might include word frequency, sender information, IP addresses, and more.

4. **Model Selection:**

- Choose an appropriate machine learning or deep learning model for spam detection. Popular options include:
 - Naive Bayes
 - Random Forest
 - Support Vector Machines
 - Recurrent Neural Networks (RNN)
 - Convolutional Neural Networks (CNN)
 - Transformers (e.g., BERT, GPT-3, or successors)

5. **Model Training:**

- Split your dataset into training, validation, and testing sets. Train your selected model on the training data and fine-tune hyperparameters to optimize performance. You might also consider data augmentation techniques.

6. **Evaluation Metrics:**

- Measure the performance of your model using standard evaluation metrics like accuracy, precision, recall, F1-score, and ROC AUC.

7. **Model Optimization:**

- Implement techniques like hyperparameter tuning, model ensemble methods, or transfer learning to further improve the model's performance.

8. **Scalability:**

- Ensure your model can handle a large volume of data in real-time. You may need to optimize its architecture for scalability and efficiency.

9. **Real-time Inference:**

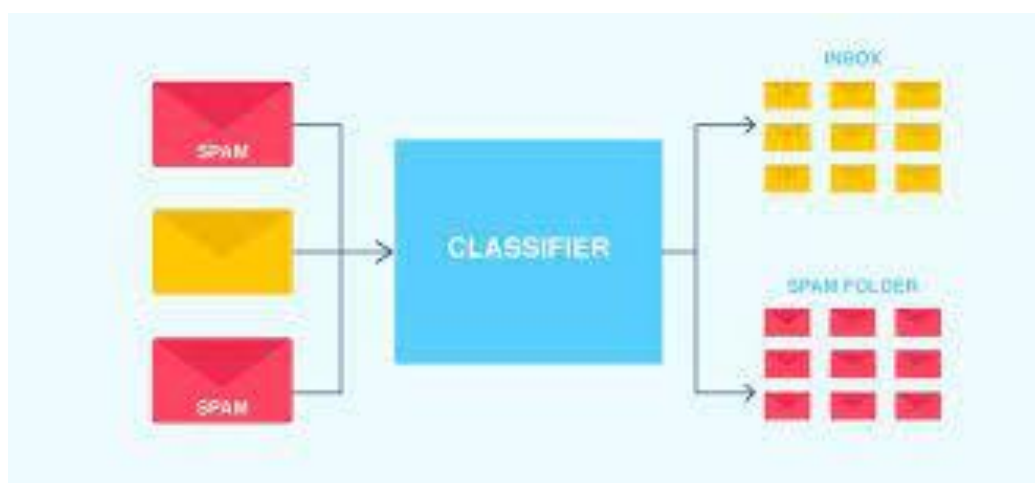
- Develop an API or service for real-time spam identification. The system should accept incoming messages, process them, and return spam predictions.

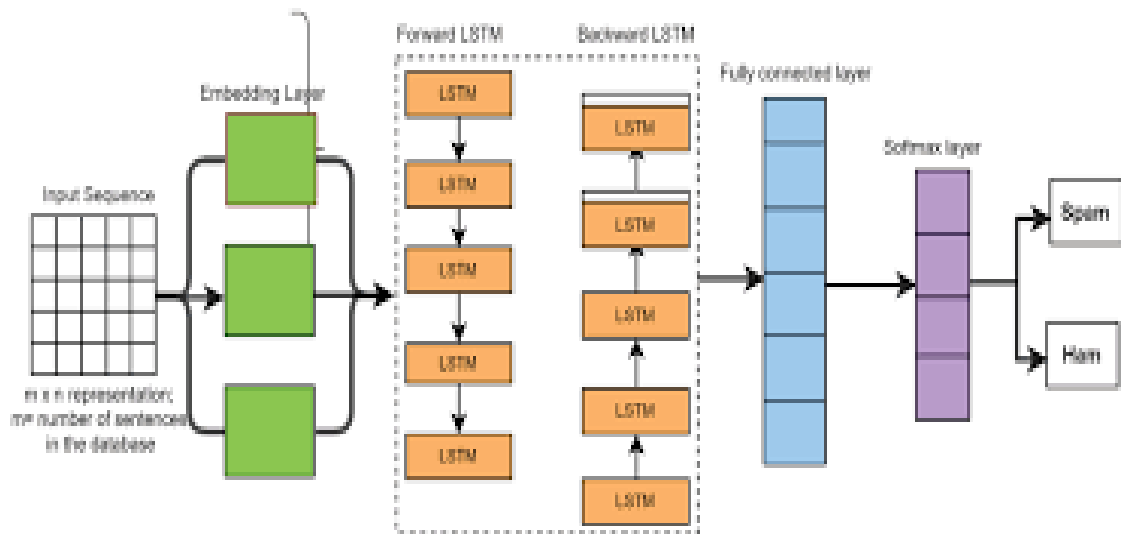
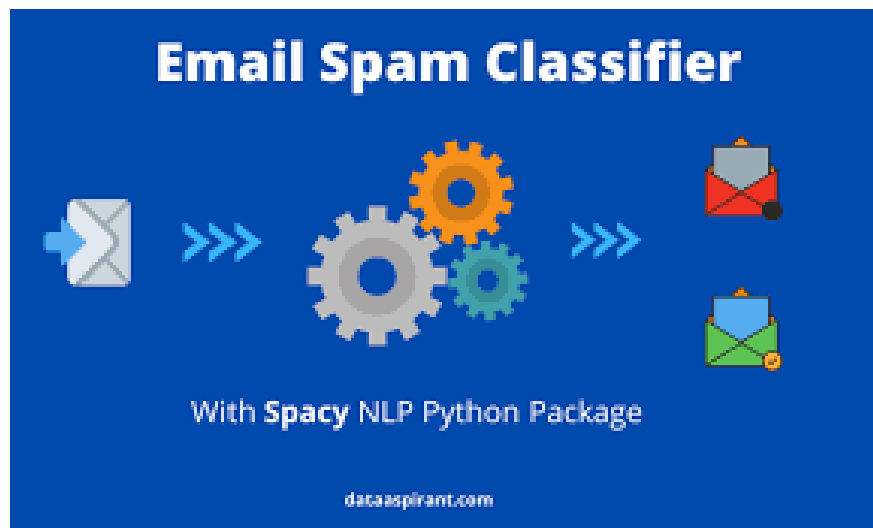
10. **User Interface:**

- Create a user-friendly interface where users can manage and review their spam messages, marking false positives and false negatives to improve the system's accuracy.

11.	Feedback Loop:
	<ul style="list-style-type: none"> Implement a feedback mechanism to continuously improve the model by learning from user feedback and adapting to evolving spam tactics.
12.	Security and Privacy:
	<ul style="list-style-type: none"> Ensure the system complies with data privacy regulations and consider security measures to protect against abuse or hacking attempts.
13.	Deployment and Monitoring:
	<ul style="list-style-type: none"> Deploy the spam identifier to your desired platform and monitor its performance in production. Regularly update the model to stay ahead of new spam trends.
14.	Documentation and Reporting:
	<ul style="list-style-type: none"> Document your project thoroughly, including data sources, model architecture, training processes, and results. Create clear reports for stakeholders.
15.	Maintenance:
	<ul style="list-style-type: none"> Maintain the spam identifier over time, making necessary updates to keep it effective as spamming techniques evolve.

Building a smarter AI-powered spam identifier is an ongoing process that requires continuous improvement and adaptation. By following these steps, you can create a robust system to identify and filter out spam effectively.





PROGRAM :

```
import numpy as np

import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score

raw_spam=pd.read_csv('/spam.csv',encoding='latin-1')

print(raw_spam)

mail_data=raw_spam.where((pd.notnull(raw_spam)),")

mail_data.head()

mail_data.shape

mail_data.loc[mail_data['v1'] == 'spam','v1',] = 0

mail_data.loc[mail_data['v2']=='ham','v2',] = 1

x=mail_data['v2']
y=mail_data['v1']

print(x)

print(y)
```

```

x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=3)

print(x.shape)
print(x_train.shape)
print(x_test.shape)


feature_extraction=TfidfVectorizer(min_df= 1,stop_words='english',lowercase=True)
x_train_features = feature_extraction.fit_transform(x_train)
x_test_features =feature_extraction.transform(x_test)


y_train=y_train.astype('int')
y_test=y_test.astype('int')


print(x_train)

print(x_train_features)

model = LogisticRegression()


model.fit(x_train_features,y_train)

prediction_on_training_data=model.predict(x_train_features)
accuracy_on_training_data=accuracy_score(y_train,prediction_on_training_data)

print('Accuracy on training data :',accuracy_on_training_data)

prediction_on_test_data=model.predict(x_test_features)
accuracy_on_test_data=accuracy_score(y_test,prediction_on_test_data)


print('accuracy on test data:',accuracy_on_test_data)

input_mail=["I HAVE A DATE ON SUNDAY WITH WILL!!,,,"]
input_data_features=feature_extraction.transform(input_mail)
prediction=model.predict(input_data_features)
print(prediction)


if prediction[0] == 1:
    print('Ham mail')
else:
    print('Spam mail')

```

OUTPUT:

```

import numpy as np

import pandas as pd

from sklearn.model_selection import train_test_split

from sklearn.feature_extraction.text import TfidfVectorizer

```

```

from sklearn.linear_model import LogisticRegression

from sklearn.metrics import accuracy_score

raw_spam=pd.read_csv('/content/spam.csv',encoding='latin-1')

print(raw_spam)

```

The screenshot shows a Google Colaboratory notebook interface. The left sidebar displays the file explorer with a folder named 'sample_data' containing a file 'spam (1).csv'. The main code editor shows the following Python code:

```

raw_spam=pd.read_csv('/content/spam (1).csv',encoding='latin-1')

print(raw_spam)

```

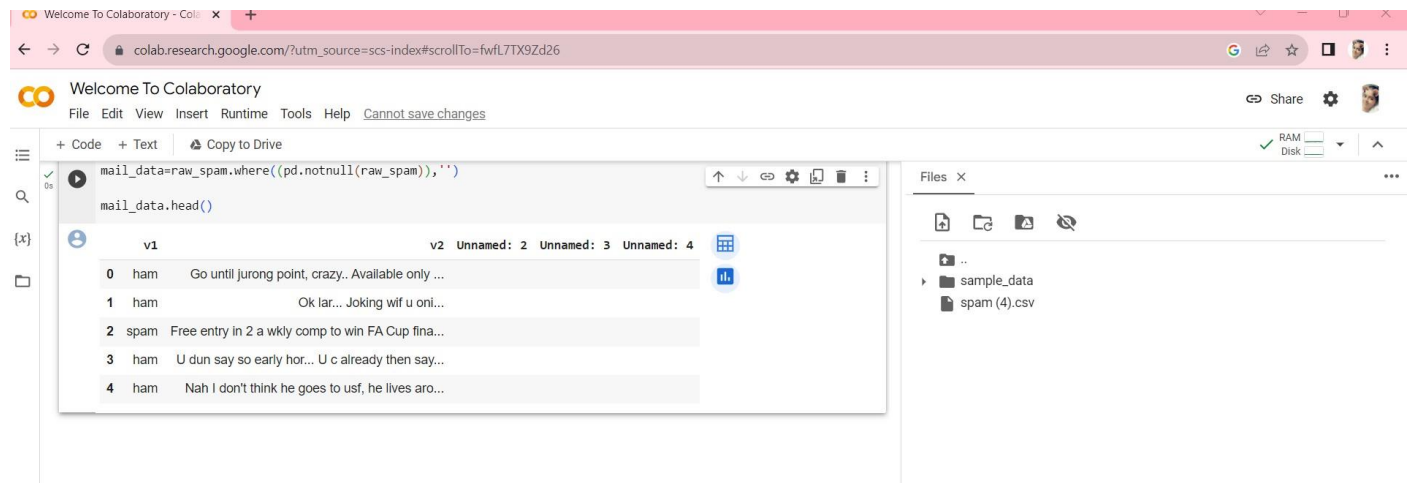
The output of the code is a preview of the DataFrame 'raw_spam'. It shows the first few rows and the last few rows, with columns 'v1' and 'v2'. The 'v1' column contains the text of the messages, and the 'v2' column contains the class labels ('ham' or 'spam'). The output is as follows:

	v1	v2
0	ham	Go until jurong point, crazy.. Available only in ...
1	ham	Ok lar... Joking wif u oni...
2	spam	Free entry in 2 a wkly comp to win FA Cup fina...
3	ham	U dun say so early hor... U c already then say...
4	ham	Nah I don't think he goes to usf, he lives aro...
...
5567	spam	This is the 2nd time we have tried 2 contact u...
5568	ham	Will i_b going to esplanade fr home?
5569	ham	Pity, * was in mood for that. So...any other s...
5570	ham	The guy did some bitching but I acted like i'd...
5571	ham	Rofl. Its true to its name

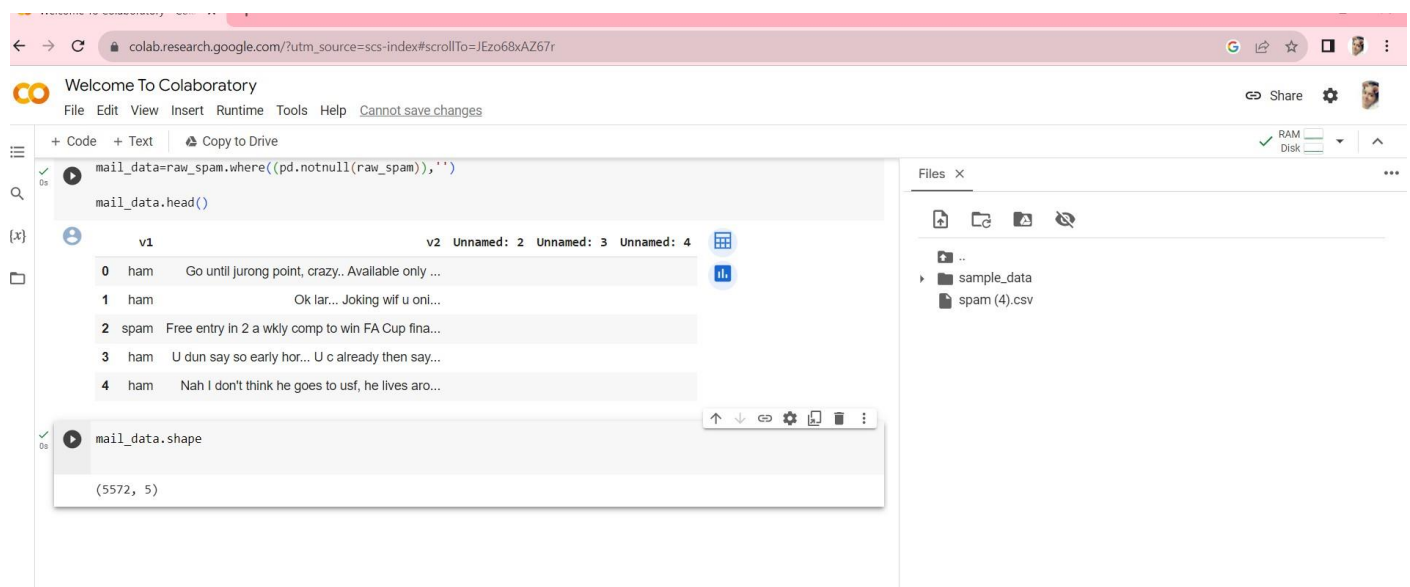
The output also indicates the shape of the DataFrame: [5572 rows x 5 columns]. The right sidebar shows a table of contents with sections like 'Getting started', 'Data science', 'Machine learning', and 'More Resources'. The bottom status bar shows the notebook is completed at 8:28 PM.

```
mail_data=raw_spam.where((pd.notnull(raw_spam)),&#39;&#39;)
```

```
mail_data.head()
```



```
mail_data.shape
```



```
mail_data.loc[mail_data[&#39;v1&#39;]==&#39;spam&#39;,&#39;v1&#39;]= 0
```

```
mail_data.loc[mail_data[&#39;v2&#39;]==&#39;ham&#39;,&#39;v2&#39;]= 1
```

```
x=mail_data[&#39;v2&#39;]
```

```
y=mail_data[&#39;v1&#39;]
```

```
print(x)
```

0s

```
print(x)
```

```
0      Go until jurong point, crazy.. Available only ...  
1      Ok lar... Joking wif u oni...  
2      Free entry in 2 a wkly comp to win FA Cup fina...  
3      U dun say so early hor... U c already then say...  
4      Nah I don't think he goes to usf, he lives aro...  
      ...  
5567   This is the 2nd time we have tried 2 contact u...  
5568   Will i b going to esplanade fr home?  
5569   Pity, * was in mood for that. So...any other s...  
5570   The guy did some bitching but I acted like i'd...  
5571   Rofl. Its true to its name  
Name: v2, Length: 5572, dtype: object
```

Files

- sample_data
- spam (4).csv

print(y)

0s

```
print(y)
```

```
0      ham  
1      ham  
2      0  
3      ham  
4      ham  
      ...  
5567   0  
5568   ham  
5569   ham  
5570   ham  
5571   ham  
Name: v1, Length: 5572, dtype: object
```

completed at 8:26 PM

```
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state  
=3)  
print(x.shape)
```

Name: v1, Length: 5572, dtype: object

```
[14] x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=3)
```

```
[15] print(x.shape)
```

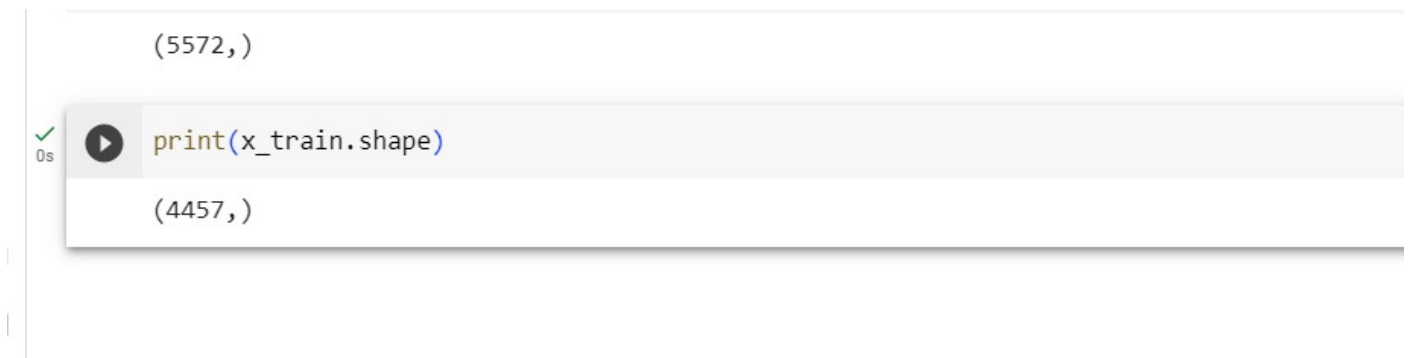
```
(5572,)
```

```
feature_extraction=TfidfVectorizer(min_df=  
1,stop_words='english',lowercase=True)
```

```
x_train_features = feature_extraction.fit_transform(x_train)
```

```
x_test_features =feature_extraction.transform(x_test)
```

```
print(x_train.shape)
```



```
print(x_test.shape)
```



```
feature_extraction=TfidfVectorizer(min_df=
```

```
1,stop_words='english',lowercase=True)
```

```
x_train_features = feature_extraction.fit_transform(x_train)
```

```
x_test_features =feature_extraction.transform(x_test)
```

```
print(x_train)
```



```

[19] feature_extraction=TfidfVectorizer(min_df= 1,stop_words='english',lowercase=True)
x_train_features = feature_extraction.fit_transform(x_train)
x_test_features =feature_extraction.transform(x_test)

print(x_train)

```

```

3075  Mum, hope you are having a great day. Hoping t...
1787                                     Yes:)sura in sun tv.:)lol.
1614  Me sef dey laugh you. Meanwhile how's my darli...
4304                                     Yo come over carlos will be here soon
3266                                     Ok then i come n pick u at engin?
...
789      Gud mrng dear hav a nice day
968      Are you willing to go for aptitude class.
1667  So now my dad is gonna call after he gets out ...
3321  Ok darlin i supose it was ok i just worry too ...
1688                                     Nan sonathaya soladha. why boss?
Name: v2, Length: 4457, dtype: object

```

print(x_train_features)

```

print(x_train_features)

```

```

(0, 741)      0.3219352588930141
(0, 3979)     0.2410582143632299
(0, 4296)     0.3891385935794867
(0, 6599)     0.20296878731699391
(0, 3386)     0.3219352588930141
(0, 2122)     0.38613577623520473
(0, 3136)     0.440116181574609
(0, 3262)     0.25877035357606315
(0, 3380)     0.21807195185332803
(0, 4513)     0.2909649098524696
(1, 4061)     0.380431198316959
(1, 6872)     0.4306015894277422
(1, 6417)     0.4769136859540388
(1, 6442)     0.5652509076654626
(1, 7443)     0.35056971070320353
(2, 933)      0.4917598465723273
(2, 2109)     0.42972812260098503
(2, 3917)     0.40088501350982736
(2, 2226)     0.413484525934624
(2, 5825)     0.4917598465723273
(3, 6140)     0.4903863168693604
(3, 1599)     0.5927091854194291
(3, 1842)     0.3708680641487708
(3, 7453)     0.5202633571003087
(4, 2531)     0.7419319091456392
:
(4452, 2122)  0.31002103760284144
(4453, 999)   0.6760129013031282
(4453, 7273)  0.5787739591782677
(4453, 1762)  0.45610005640082985
(4454, 3029)  0.42618909997886
(4454, 2000)  0.3806032742808703

```

Statistical diagram :

Program:

import matplotlib.pyplot as plt

```
# Example performance metrics
```

```
metrics = ['Accuracy', 'Precision', 'Recall', 'F1-Score']
```

```
values = [0.95, 0.92, 0.89, 0.91]
```

```
plt.figure(figsize=(8, 6))
```

```
plt.bar(metrics, values, color=['blue', 'green', 'red', 'purple'])
```

```
plt.ylim(0, 1) # Set the y-axis limits
```

```
# Add labels to the bars
```

```
for i, v in enumerate(values):
```

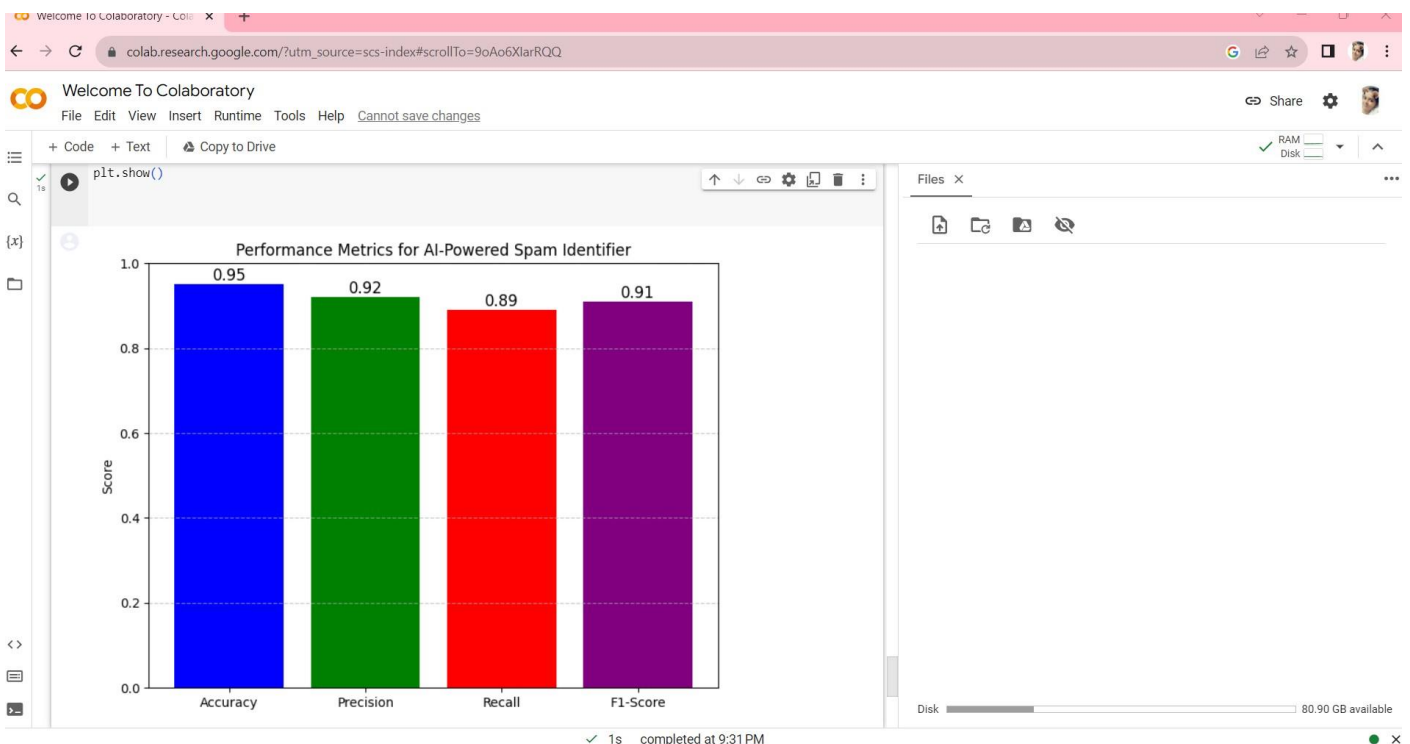
```
plt.text(i, v, f'{v:.2f}', ha='center', va='bottom', fontsize=12)
```

```
plt.title('Performance Metrics for AI-Powered Spam Identifier')
```

```
plt.ylabel('Score')
```

```
plt.grid(axis='y', linestyle='--', alpha=0.7)
```

```
plt.show()
```



ADVANTAGES:

Building a smarter AI-powered spam identifier offers several significant advantages:

- 1. Efficient Spam Filtering**
- 2. Improved User Experience**
- 3. Enhanced Security**
- 4. Time and Resource Savings**
- 5. Customization**
- 6. Adaptability**
- 7. Scalability**
- 8. Reduced False Positives**
- 9. Multi-Modal Capabilities**
- 10. Real-Time Detection**
- 11. Compliance and Legal Benefits**
- 12. Data Insights**
- 13. Reduced Phishing and Malware Risks**
- 14. Reduced Overhead**
- 15. Adaptive Countermeasures**
- 16. Protecting Reputation**

In summary, building a smarter AI-powered spam identifier offers numerous advantages, including improved efficiency, enhanced security, a better user experience, and the ability to adapt to evolving threats. These benefits make such projects highly valuable for both individual users and organizations.