OPTIMIZING SPAM FILTERING WITH MACHINE LEARNING

INTRODUCTION

OVERVIEW

A spam filter is a program used to detect unsolicited, unwanted and virus-infected emails and prevent those messages from getting to a user's inbox.

Like other types of filtering programs, a spam filter looks for specific criteria on which to base its judgments.

An email spam filter is a tool used in email hosting software that churns out unsolicited, unwanted, and virus-infested emails and keeps such emails off of the user's inbox.

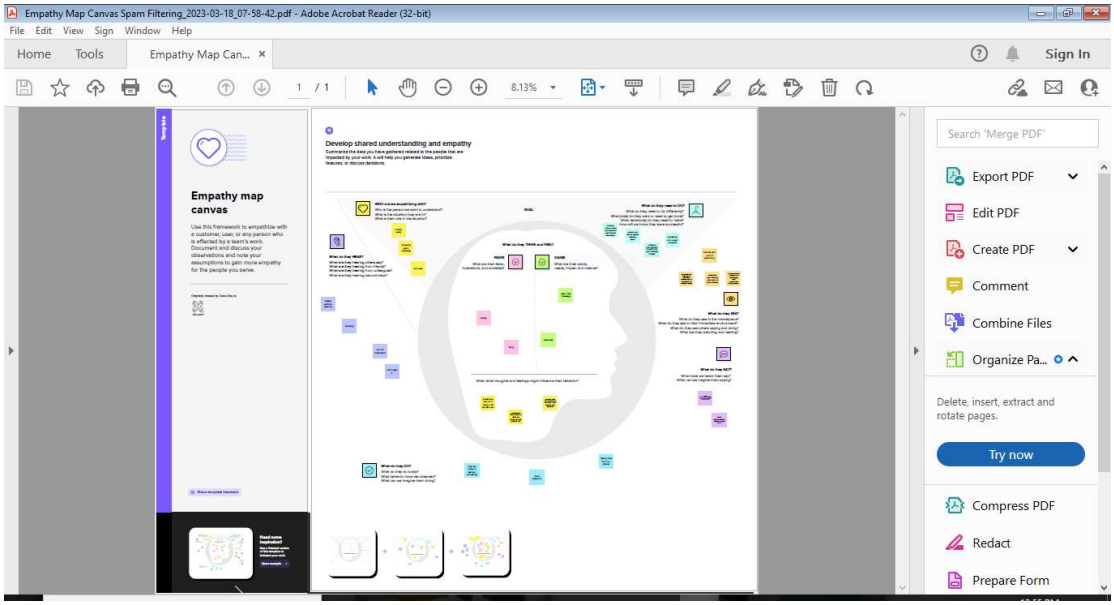
This protects the user from any potential cyber threat and facilitates smooth communications and workflow.

PURPOSE

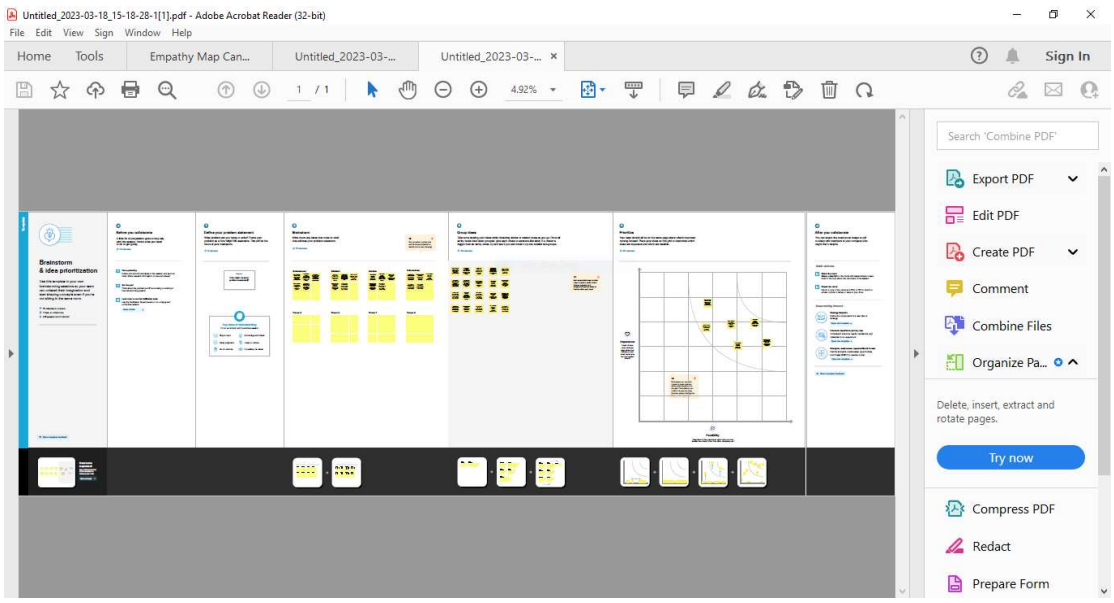
An email spam filter is a tool used in email hosting software that churns out unsolicited, unwanted, and virus-infested emails and keeps such emails off of the user's inbox. This protects the user from any potential cyber threat and facilitates smooth communications and workflow.

A spam filter is a program used to detect unsolicited, unwanted and virus-infected emails and prevent those messages from getting to a user's inbox. Like other types of filtering programs, a spam filter looks for specific criteria on which to base its judgments.

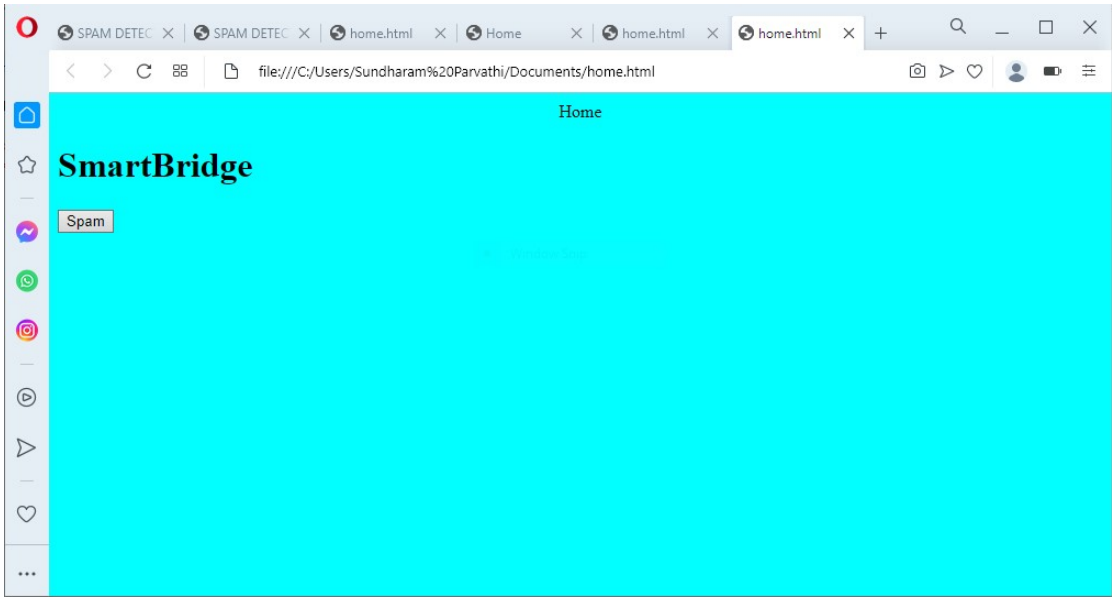
EMPATHY MAP

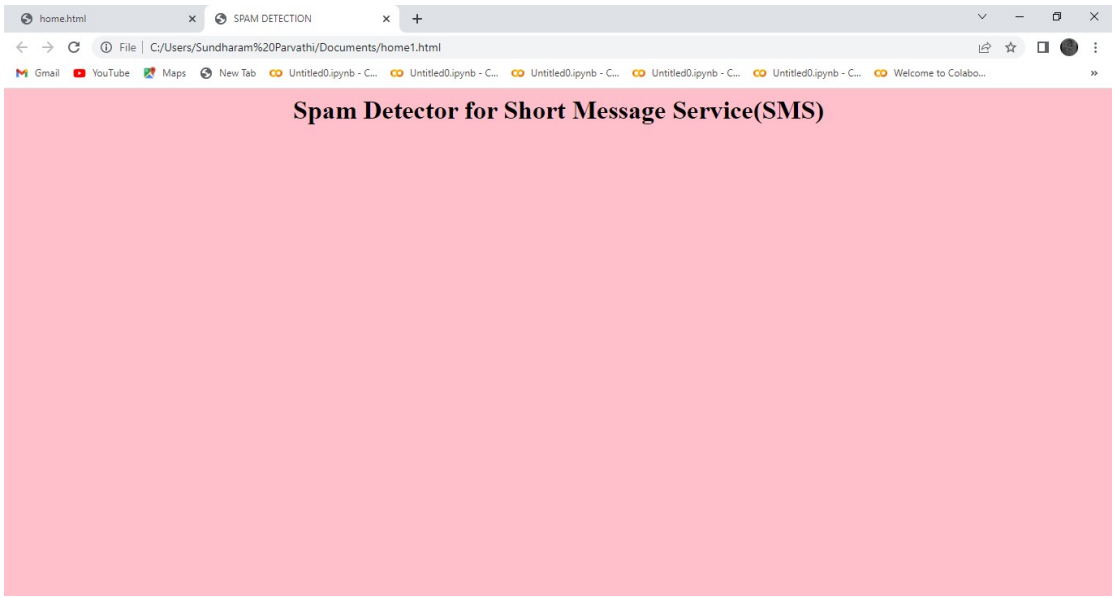


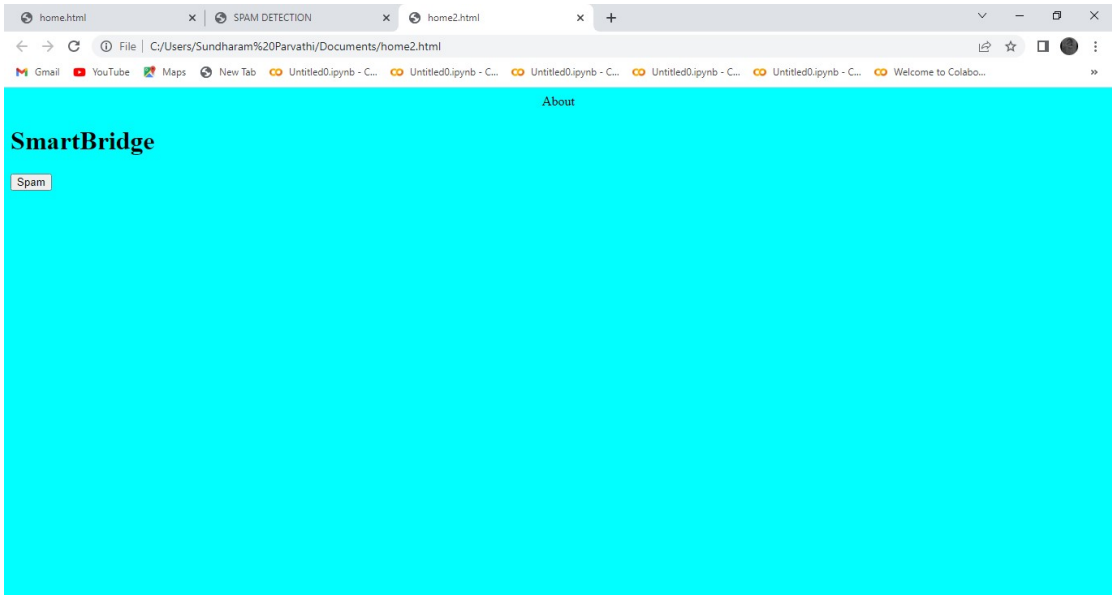
BRAINSTORMING MAP



RESULT


SPAM DETEC
SPAM DETEC
home.html
Home
home.html
home.html
file:///C:/Users/Sundharam%20Parvathi/Documents/home.html
Home
SmartBridge
Spam


home.html
SPAM DETECTION
File | C:/Users/Sundharam%20Parvathi/Documents/home1.html
Gmail
YouTube
Maps
New Tab
Untitled0.ipynb - C...
Welcome to Colabo...
Spam Detector for Short Message Service(SMS)



ADVANTAGES OF SPAM FILTERING

Spam is most well-known for spreading viruses and scams to unwitting people across the internet, but it can actually cause plenty of problems for the modern business.

This is why effective spam filtering, like Securence spam filtering, is an important part of running a successful business in

the 21st century.

Here are just a few reasons why spam filtering is important for not only keeping you safe from viruses, but also for helping your company be more effective and successful.

Spam filtering is also incredibly affordable, making it a cheap but extremely effective way to keep yourself safe.

DISADVANTAGES OF SPAM FILTERING

Thousands of spam emails may reach Inboxes before a spammer's email address, IP or domain is blacklisted.

Spam filtering is machine-based so there is a room for mistakes called “false positives.” Bayesian filters may be fooled by spammers, e.g. in a case of using large blocks of legitimate text.

Spam is unethical. Spammers don't only violate

laws and people's privacy but they also steal their money.

Most email users dial into Internet providers who charge on an hourly or even minute basis. Spammers take hold of valuable Internet resources by sending junk mail but sending it "postage due".

APPLICATION

A spam filter is a program used to detect unsolicted, unwanted and virus-infected emails and prevent those message from getting to a user's inbox.

CONCLUSION

Bayesian spam filtering is an incredibly powerful statistical technique—with acceptable computational complexity—for identifying spam messages.

Bayesian techniques address many

weaknesses of other methodologies: The entire message can be examined, not just special parts.

present a review on supervised machine learning strategies for filtering spam emails.

They concluded that the Naïve Bayes method provides faster results and decent precision over all other methods (except SVM and ID3) from all the techniques discussed.

FUTURE SCOPE

The algorithms developed so far have not been able to remove the requirement of manual checking of the reviews.

Hence there is scope for complete automation of spam filtering systems with maximum efficiency.

A spam filter is a program used to detect unsolicited, unwanted and virus-infected emails and prevent those messages from getting to a user's inbox.

Like other types of filtering programs, a spam

filter looks for specific criteria on which to base its judgments.

It has a broadcasted, rather than targeted, message.

It suits the purposes of the sender rather than the receiver.

Most important, the message is distributed without the explicit permission of the recipients.

APPENDIX

Source code

spam_filter.ipynb

 # -*- coding: utf-8 -*-

"""Untitled0.ipynb

Automatically generated by Colaboratory.

Original file is located at

https://colab.research.google.com/drive/1YsI9jqi5cBjbhgpqF-y7NbbewYV__Y0o

"""

#df["lable"].value_counts().plot(kind="bar",figsize=)(12,6)

#plt.xticks(np.arange(2),('Non spam','spam'),rotation=0);

#perfroming feature scaling op[eration using standard scaller on x part of the dataset because]

#there different type of values in the columns

#sc=StandardScaler()

#x_bal=sc.transform(x_bal)

#x_bal = pd.DataFrame(x_bal,columns=names)

#Splitting data into train and validation sets using train_test_split

```
#from sklearn.model_selection import
train_test_split

#X_train, X_test, Y_train, Y_test =
train_test_split(X,Y, test_size =
0.20,random_state = 0)

##train size 80% and test size 20%


from sklearn.tree import
DecisionTreeClassifier

model=DecisionTreeClassifier()

#model.fit(X_train_res, Y_train_res)


#from sklearn.ensemble import
RandomForestClassifier

#model1=RandomForestclassifier()

#model1.fit(X_train_res, Y_train_res)


#from sklearn.naive_bayes import
MultinomialNB

#model = mulitinominalNB()
```

#Fitting the model to the training sets

#model.fit(X_train_res, Y_train_res)

#from tensorflow.keras.model import sequential

#from tensorflow.keras.layers import Dense

#Fitting the model to the training sets

#model=sequential()

#X_train.shape

#model.add(Dense(units=x_train_res.shape[1],activation="relu",kernal_initializer="random_uniform"))

#model.add(Dense(units= 100,activation="relu",kernal_initializer="random_uniform"))

#model.add(Dense(units=
100,activation="relu",kernal_initializer="rando
m_uniform"))

#model.add(Dense(units=
1,activation="sigmoid"))

#model.compile(optimizer="adam",loss="binary
_crossentropy"),metrics=['accuracy']

#generator=model.fit(X_train_res,Y_train_res,ep
ochs=10,steps_per_epoch=len(X_train_res)//64)

#generator=model.fit(X_train_res,Y_train_res,ep
ochs=10,steps_per_epoch=len(X_train_res)//64)

#Y_pred=model.predict(X_test)

#Y_pred

```
#Y_pr=np.where(Y_pred>0.5,1,0)


#y_test


#from sklearn.metrices import
confusion_matrix,accuarcy_score

#cm=confusion_matrix(y_test,y_pr)

#score=acccuracy_score(y_test,y_pr)

#print(cm)

#print('Accuracy score Is:-',score*100)


def new_review(new_review):

  new_review = new_review

 # new_review = re.sub('[^a-zA-Z]',)'
',new_review)

 #new_review = new_review.lower()

 #new_review = new review.split()

 #ps=porterstemmer()

 #all_stopwords=stopwords.words('english')
```

#all_stopwords.remove('not')

#new_review=[ps.stem(word) for word in new_review if not word in set(all_stopwords)]

#new_review=' '.join(new_review)

#new_corpus=[new_review]

#new_x_test=cv.transform(new_corpus).yoarray()

#print(new_x_test)

#new_y_pred=loaded_model.predict(new_X_test)

#print(new_y_pred)

#new_x_pred=np.where(new_y_pred>0.5,1,0)

#return new_y_pred

new_review=new_review(str(input("Enter new review...")))

from sklearn.metrics import confusion_matrix,accuracy_score, classification_report

#cm=confusion_matrix(y_test, y_pred)

#score = accracy_score(y_test,Y_pred)

#print(cm)

#print('accuracy score is Naive Bayes:-',score* 100)

#cm = confusion_matrix(y_test, y_pred)

#score = accuracy Score(y_test,y_pred)

#print(cm)

#print('Accuracy Score Is;-' ,score*100)

#cm1 = confusion_matrix(y_test, y_pred1)

#score1 = accuracy_score(y_test,y_pred1)

#print(cm1)

#print('Accuracy Score Is;- ' ,score1*100)

#from sklearn.metrics import confusion_matrix,accuracy_score

#cm = confusion_matrix(y_test, y_pr)

#score = accuracy_score(y_test,y_pr)

```
#print(cm)

#print('Accuracy Score Is;- ' ,score*100)


#from sklearn.metrics import
confusion_matrix,acuuracy_score

#cm = confusion_matrix(y_test, y_pr)

#score = accuracy_score(y_test,y_pr)

#print(cm)

#print('Accuracy Score Is;- ' ,score*100)
```