Building a Smarter AI-Powered Spam Classifier: Designing a Web Application to Classify Spam Messages Using TF-IDF, Multinomial Naive Bayes, and Other NLTK Libraries with Iterative Improvement to Enhance Accuracy, Precision, Recall, and F1-score.

Outline of problem statement, design thinking process, and the phases of development:

Problem Statement:

Building a Smarter AI-Powered Spam Classifier

The problem statement is to create an AI-powered spam classifier that can identify spam messages in a dataset. The goal is to design a system that not only performs initial spam classification but also continuously learns from new data to improve its accuracy, precision, recall, and F1-score over time.

Design Thinking Process:

The design thinking process involves the following phases:

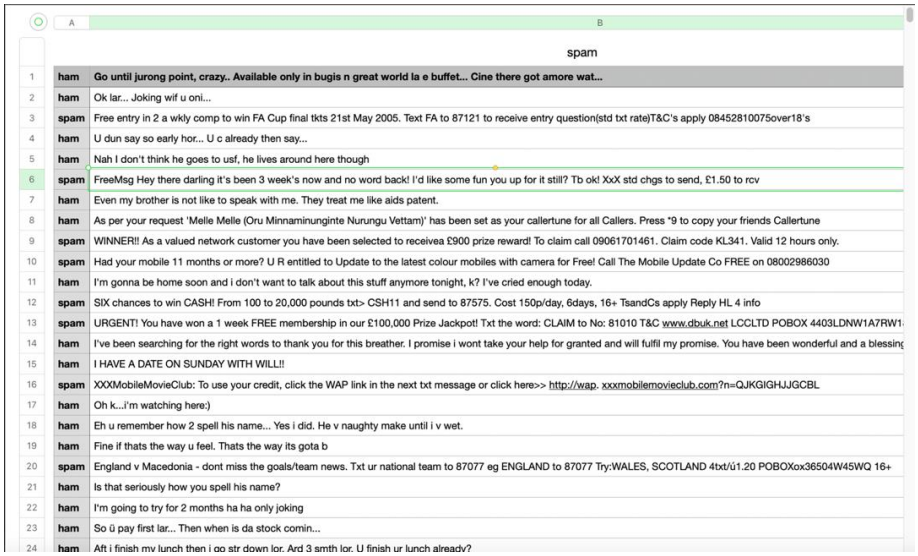
1. **Data Collection:** Acquiring a dataset containing "label" (spam or ham) and "message" (text messages) columns, which serves as the foundation for training and improving the model.
2. **Data Preprocessing**: Cleaning the text data by removing special characters, converting text to lowercase, tokenizing, removing stopwords, and performing stemming to prepare the data for analysis.
3. **Feature Extraction:** Using TF-IDF to convert tokenized words into numerical features, allowing the model to work with the data more effectively.
4. **Model Selection:** Choosing the Multinomial Naive Bayes algorithm as the initial machine learning model and training it on the TF-IDF-transformed data.
5. **Evaluation:** Assessing the model's performance with metrics such as accuracy, precision, recall, and F1-score.
6. **Iterative Improvement:** Incorporating a mechanism to update the model with new data, enabling it to learn and improve its performance continuously.
7. **Deployment:** Preparing the model for deployment, developing a user interface, and hosting it on a server or cloud platform to provide real-time spam classification.

Description of the dataset used, data preprocessing steps, and feature extraction Techniques:

Dataset Description:

The document mentions the use of the "spam.csv" dataset from Kaggle's SMS Spam Collection Dataset, which contains two columns: "label" and "message." The "label" column indicates whether a message is spam or not, and the "message" column contains the text of the messages.

SAMPLE:



	A	B
		spam
1	ham	Go until jurong point, crazy.. Available only in bugis n great world la e buffet... Cine there got amore wat...
2	ham	Ok lar... Joking wif u oni...
3	spam	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 apply 08452810075over18's
4	ham	U dun say so early hor... U c already then say...
5	ham	Nah I don't think he goes to usf, he lives around here though
6	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 still? Tb ok! XxX std chgs to send, £1.50 to rcv
7	ham	Even my brother is not like to speak with me. They treat me like aids patent.
8	ham	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
9	spam	WINNER!! As a valued network customer you have been selected to receivea £900 prize reward! To claim call 09061701461. Claim code KL341. Valid 12 hours only.
10	spam	Had your mobile 11 months or more? U R entitled to Update to the latest colour mobiles with camera for Free! Call The Mobile Update Co FREE on 08002986030
11	ham	I'm gonna be home soon and i don't want to talk about this stuff anymore tonight, k? I've cried enough today.
12	spam	SIX chances to win CASH! From 100 to 20,000 pounds txt> CSH11 and send to 87575. Cost 150p/day, 6days, 16+ TsandCs apply Reply HL 4 info
13	spam	URGENT! You have won a 1 week FREE membership in our £100,000 Prize Jackpot! Txt the word: CLAIM to No: 81010 T&C www.dbuk.net LCCLTD POBOX 4403LDNW1A7RW1
14	ham	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
15	ham	I HAVE A DATE ON SUNDAY WITH WILL!!
16	spam	XXXMobileMovieClub: To use your credit, click the WAP link in the next txt message or click here>> http://wap. xxxmobilemovieclub.com?n=QJKGIGHJJGCBL
17	ham	Oh k...i'm watching here:)
18	ham	Eh u remember how 2 spell his name... Yes i did. He v naughty make until i v wet.
19	ham	Fine if thats the way u feel. Thats the way its gota b
20	spam	England v Macedonia - dont miss the goals/team news. Txt ur national team to 87077 eg ENGLAND to 87077 Try:WALES, SCOTLAND 4txt/ú1.20 POBOXox36504W45WQ 16+
21	ham	Is that seriously how you spell his name?
22	ham	I'm going to try for 2 months ha ha only joking
23	ham	So ü pay first lar... Then when is da stock comin...
24	ham	Aft i finish my lunch then i go str down lor. Ard 3 smth lor. U finish ur lunch already?

Data Preprocessing:

Data preprocessing involves several steps, including removing special characters and punctuation, converting text to lowercase, tokenization, removing stopwords, and stemming using the Porter Stemmer. These steps ensure that the text data is in a suitable format for feature extraction and model training.

Feature Extraction:

Feature extraction is performed using TF-IDF (Term Frequency-Inverse Document Frequency) to represent the text data as numerical features. The maximum number of features is set to 2500, meaning that the top 2500 most important words are selected as features for the model.

Explaination of the choice of machine learning algorithm, model training, and evaluation metrics:

Machine Learning Algorithm:

- **Naïve Bayes Algorithm:**

Naive Bayes is a popular and simple machine learning algorithm often used in text classification, spam detection, and various other applications. Here are some key points

Model Training:

The model is trained on the training dataset, and its performance is evaluated using metrics like accuracy, precision, recall, and F1-score

Model Selection:

The document selects the Multinomial Naive Bayes algorithm for initial model training. The model is trained on the TF-IDF-transformed training data.

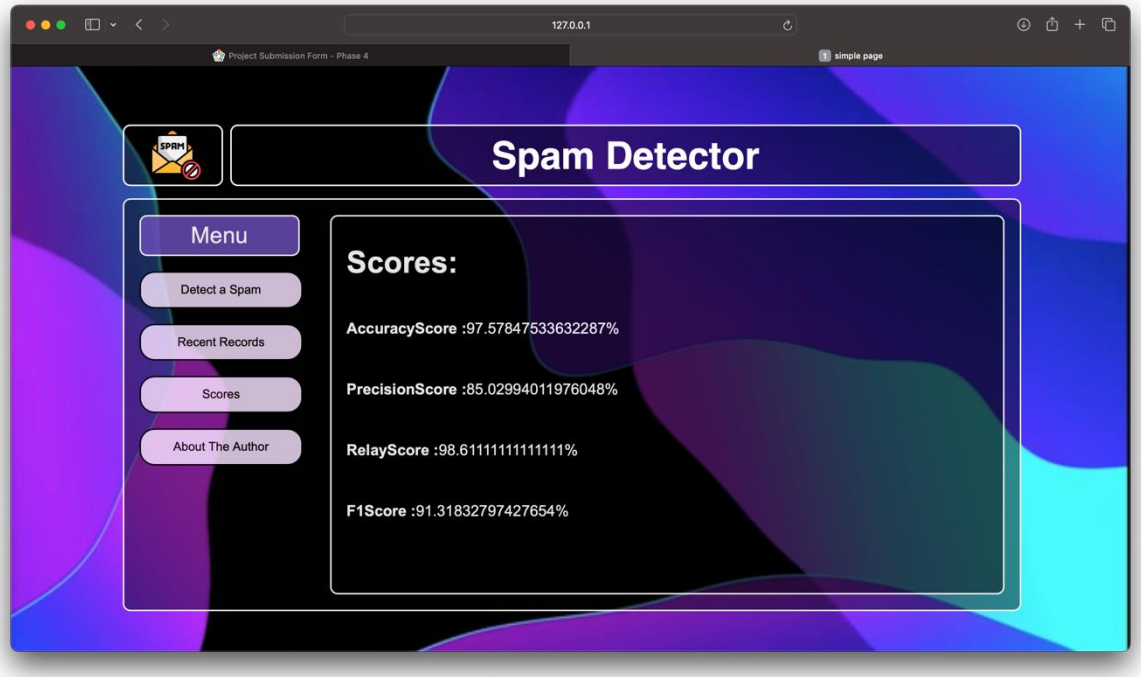
Performance Evaluation Metrics:

The document mentions several performance metrics for evaluating the model:

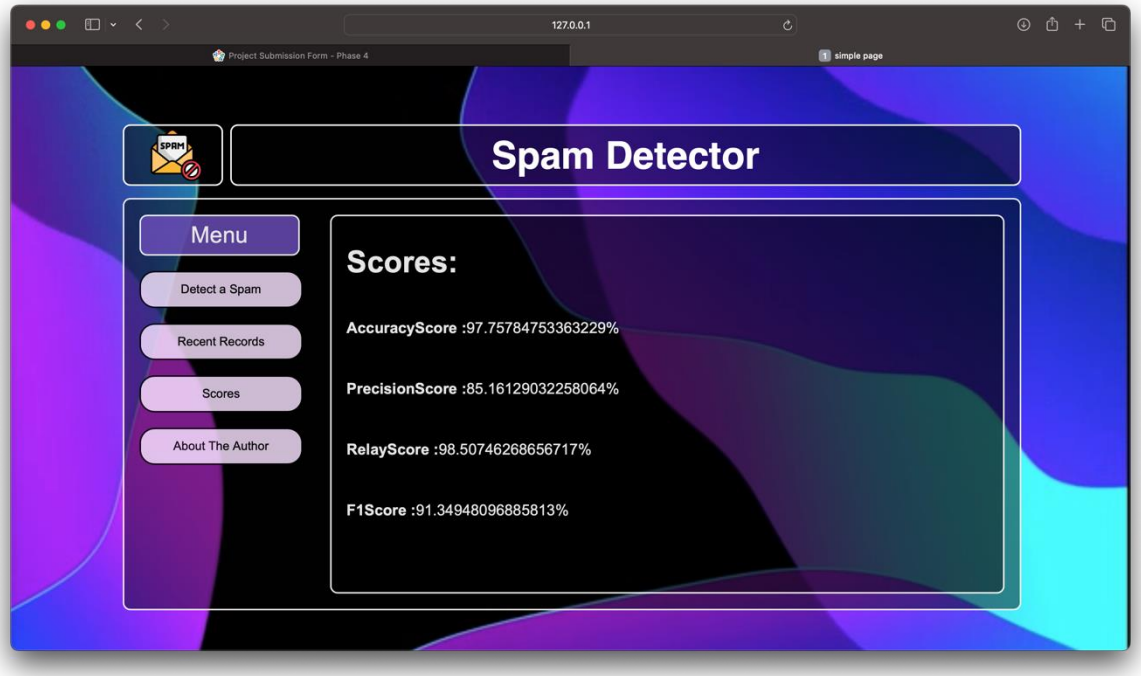
- **Accuracy**: Measures the proportion of correctly classified messages.
- **Precision**: Measures the proportion of true spam messages among the messages classified as spam.
- **Recall**: Measures the proportion of true spam messages correctly classified as spam.
- **F1-score**: A harmonic mean of precision and recall, balancing both metrics.

SAMPLE:

BEFORE EVALUATION:



AFTER EVALUATION:



Approaches used during the development:

Innovative Approaches:

The document introduces the concept of iterative improvement, where the model learns from new examples and potentially improves over time. This approach allows the model to adapt to changing spam patterns and maintain its accuracy and effectiveness.
Overall, the document provides a structured and detailed plan for building and improving an AI-powered spam classifier, from data collection to continuous learning and deployment.

Sample:

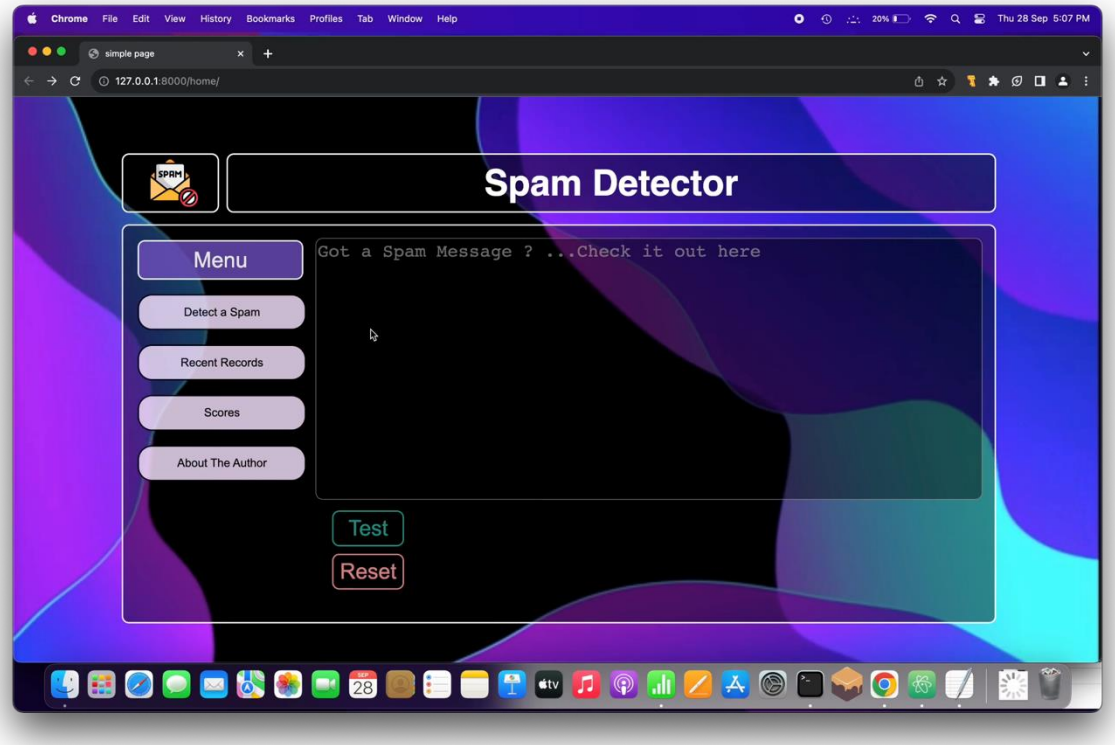
```
{'label': 'spam', 'message': 'congratulations free cash click the link'}
     label                                                message
0      ham  Go until jurong point, crazy.. Available only ...
1      ham                      Ok lar... Joking wif u oni...
2     spam  Free entry in 2 a wkly 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...
...    ...                                                ...
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
5572   ham                                hello how are you
5573  spam          congratulations free cash click the link

[5574 rows x 2 columns]
```

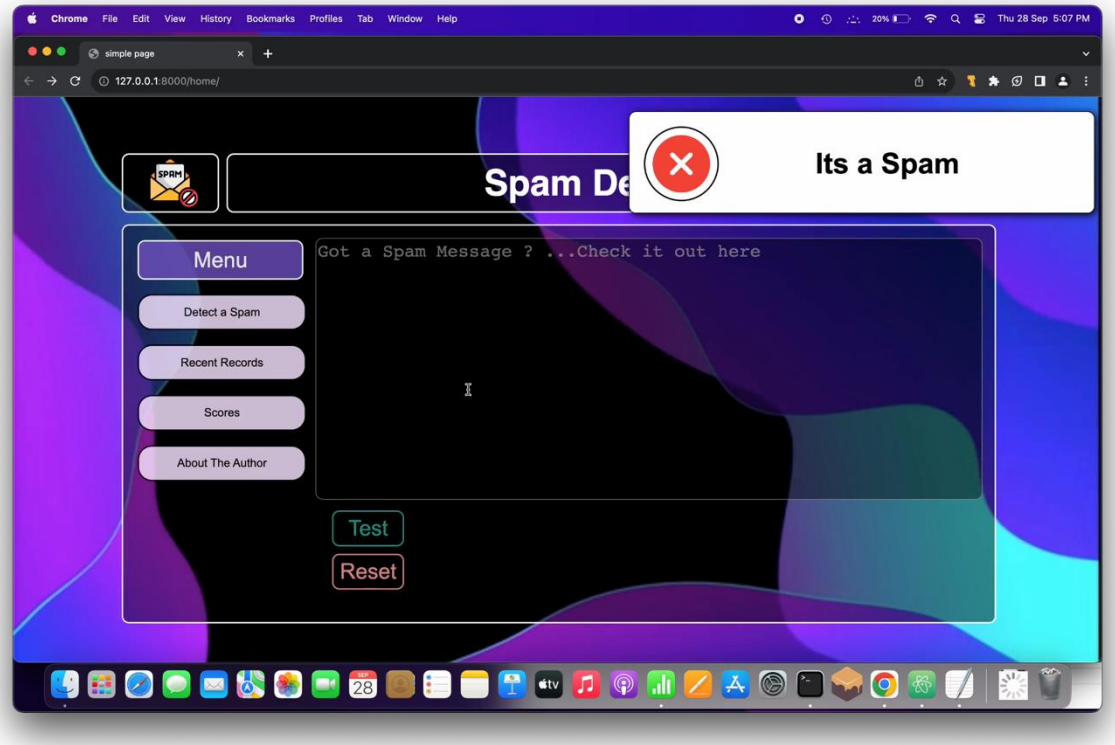
At First Iteration
At Second Iteration

PROJECT DEMO IMAGES:

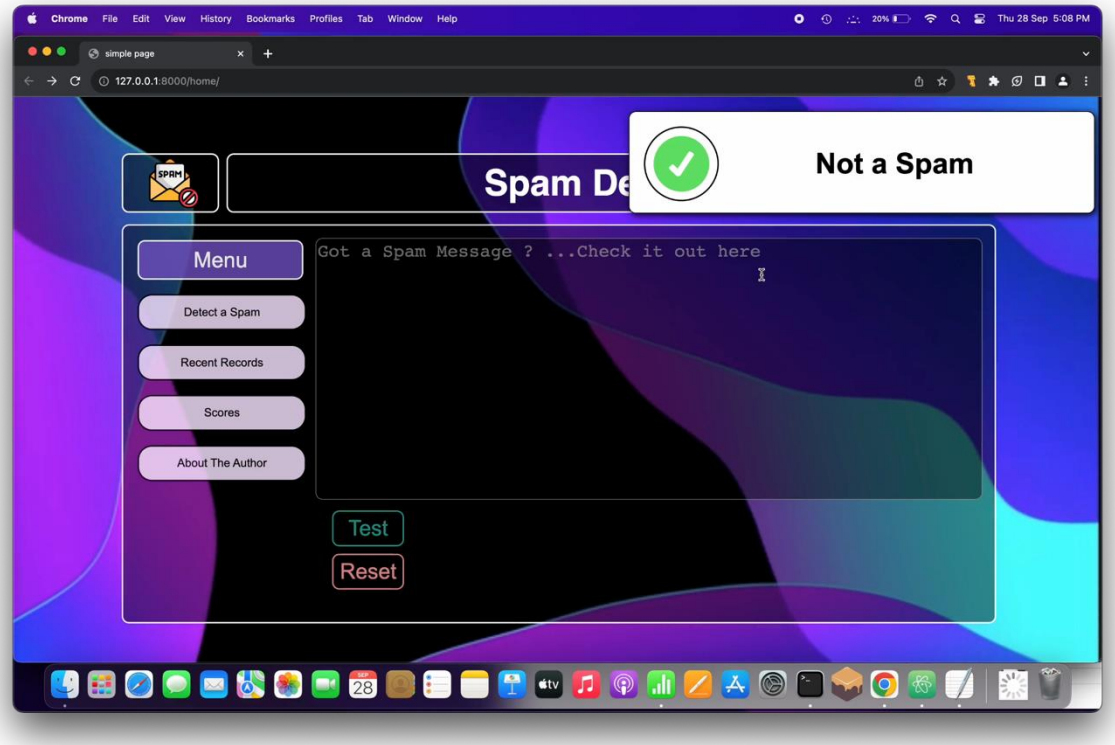
HOME PAGE:



ON SPAM ALERT:



ON NON-SPAM ALERT:



SPAM AND NON-SPAM HISTORY: