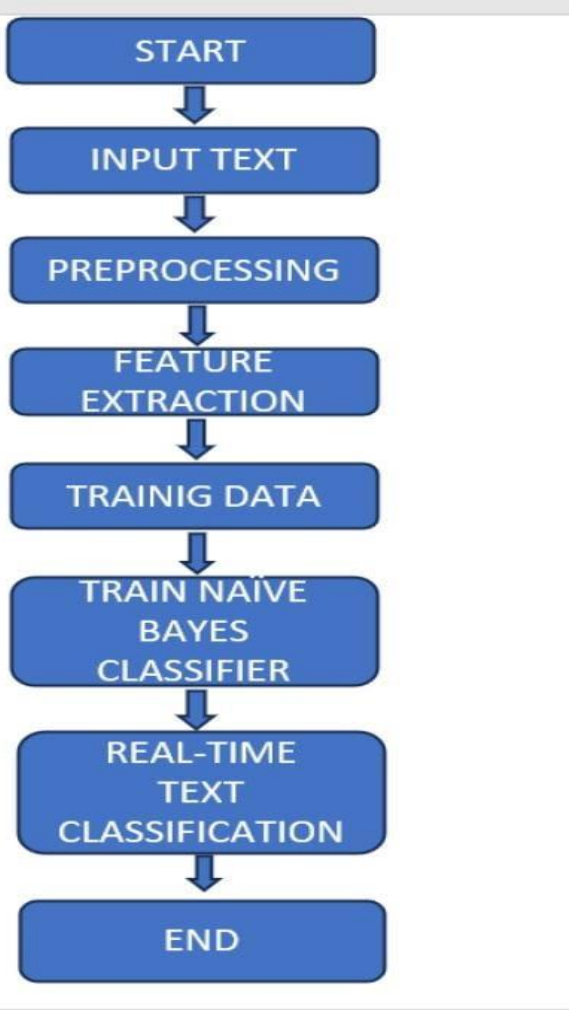
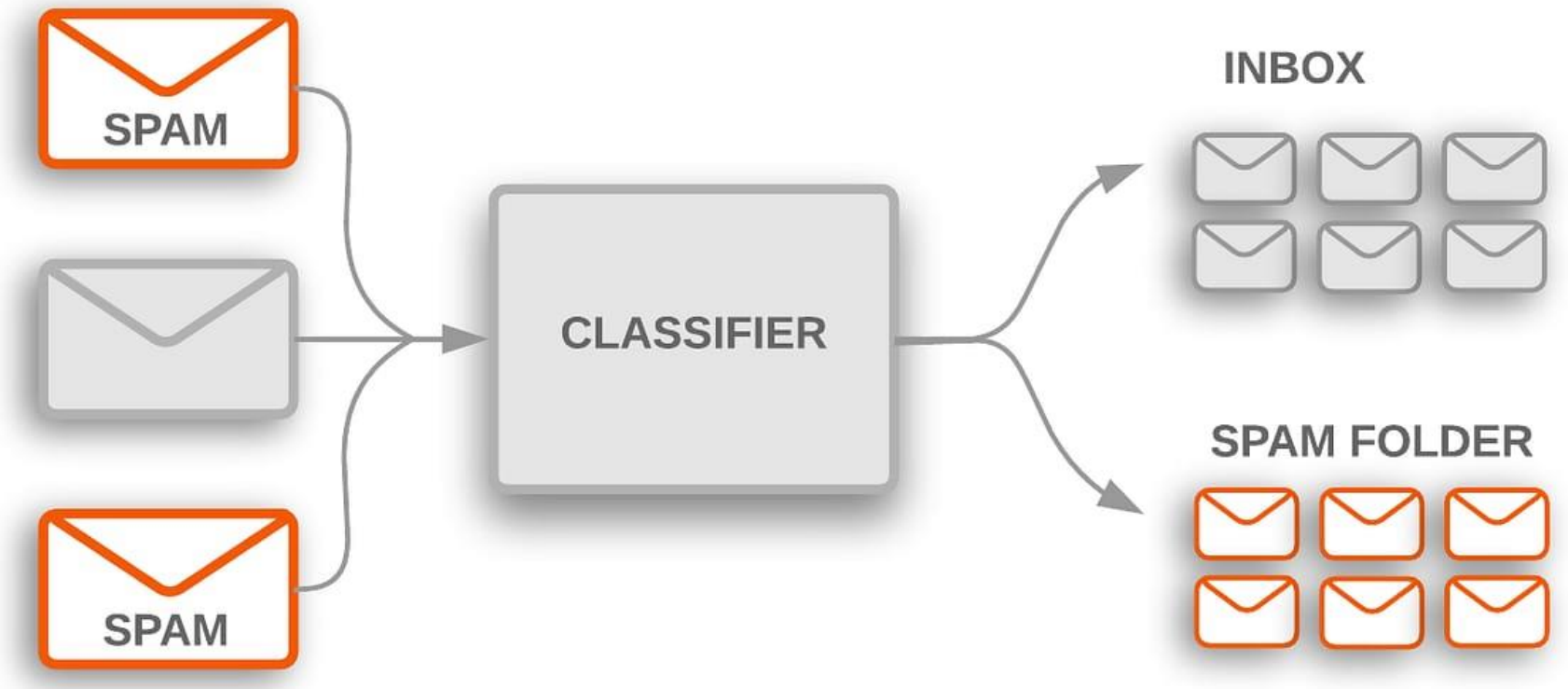
Naïve Bayes Text Classifier : Unmasking Spam in Real - Time

Flow Chart


START
INPUT TEXT
PREPROCESSING
FEATURE EXTRACTION
TRAINIG DATA
TRAIN NAÏVE BAYES CLASSIFIER
REAL-TIME TEXT CLASSIFICATION
END

Architecture Diagram



Coding

In [5]:

```
import pandas as pd
import nltk
import string
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.probability import FreqDist
from nltk.classify import NaiveBayesClassifier

# Download the stopwords corpus and punkt tokenizer if you haven't already
nltk.download('stopwords')
nltk.download('punkt')

# Sample labeled dataset for spam detection (you can replace this with your own dataset)
dataset =  [
    ("Get a new iPhone for free! Limited time offer!", "spam"),
    ("Hey, how are you doing?", "ham"),
    ("Claim your prize now! Click the link to win $1000.", "spam"),
    ("Can we reschedule our meeting for tomorrow?", "ham"),
    ("Congratulations! You've won a trip to Hawaii.", "spam")
]


def preprocess_text(text):
    # Remove punctuation
    text = text.translate(str.maketrans('', '', string.punctuation))

    # Tokenize the text into words
    words = word_tokenize(text.lower())

    # Remove stopwords
    stop_words = set(stopwords.words("english"))
```

```
    return classifier

def predict_spam(text, classifier):
    # Preprocess the input text
    preprocessed_text = preprocess_text(text)

    # Extract features from the preprocessed text
    features = extract_features(preprocessed_text)

    # Predict if the message is spam or ham using the trained classifier
    predicted_label = classifier.classify(features)

    return predicted_label

# Train the classifier
classifier = train_naive_bayes_classifier(dataset)

# Example usage: Predict if a new message is spam or ham
new_message = "Win a free vacation now!"
predicted_label = predict_spam(new_message, classifier)
print("Predicted label:", predicted_label)
```

Output

```
Classifier accuracy: 1.0
Predicted label: spam
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\lenovo\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\lenovo\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
```