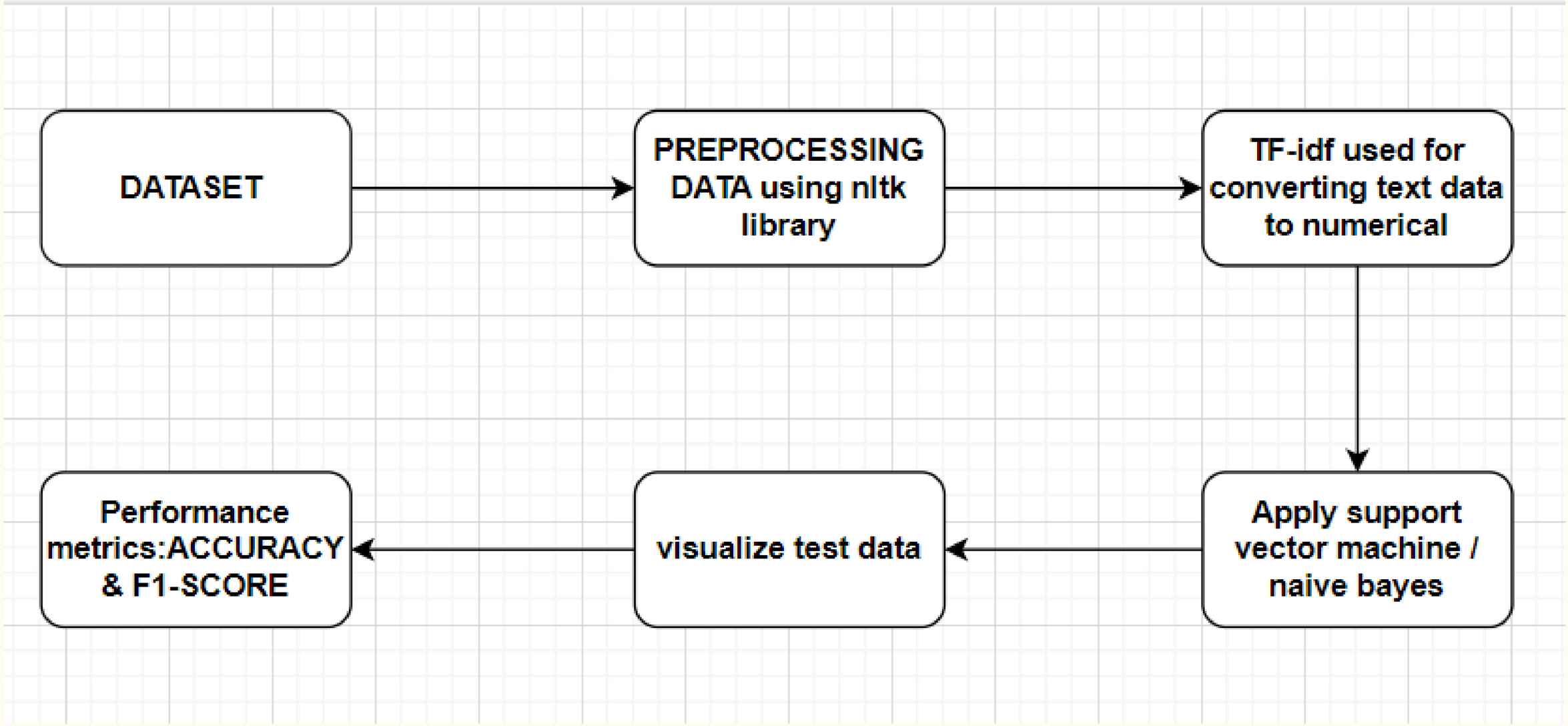
EMAIL CLASSIFICATION

TEAM : PIXELPIONEERS


DATASET
PREPROCESSING DATA using nltk library
TF-idf used for converting text data to numerical
Apply support vector machine / naive bayes
visualize test data
Performance metrics:ACCURACY & F1-SCORE

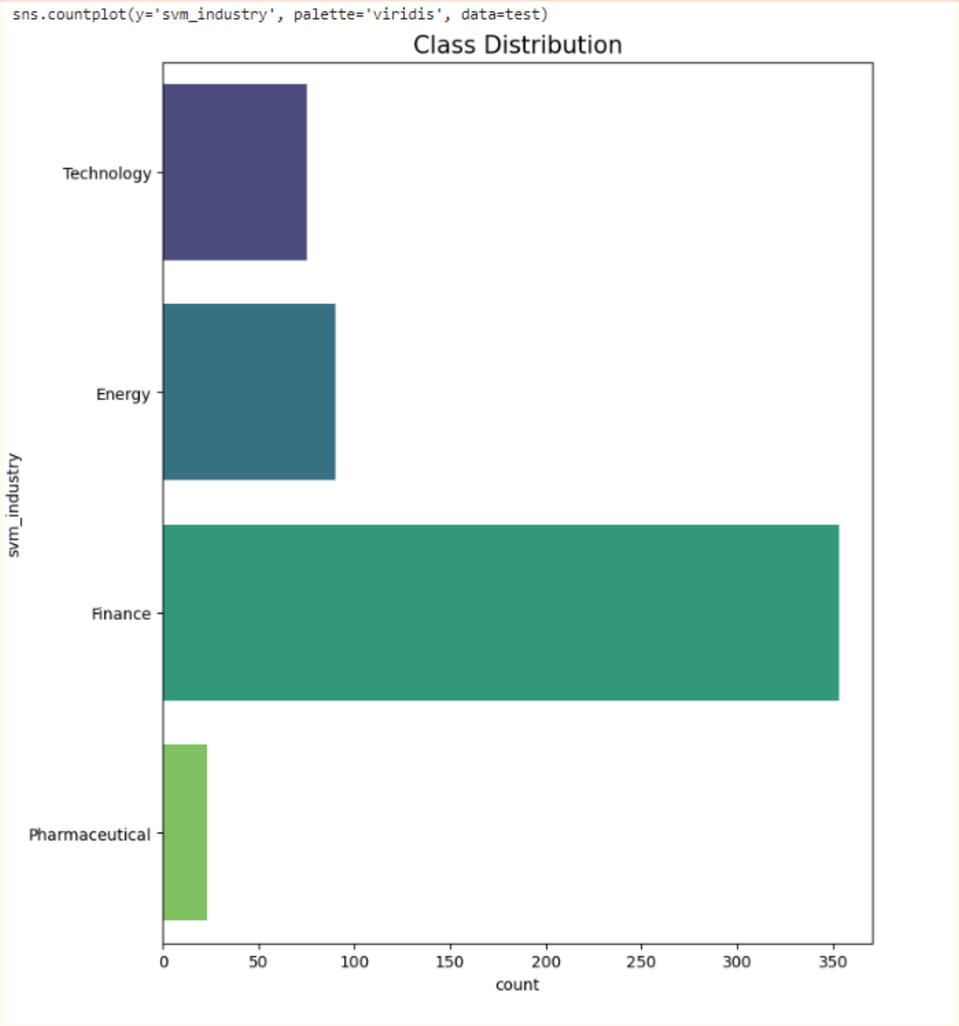
PREPROCESSING TECNIQUES

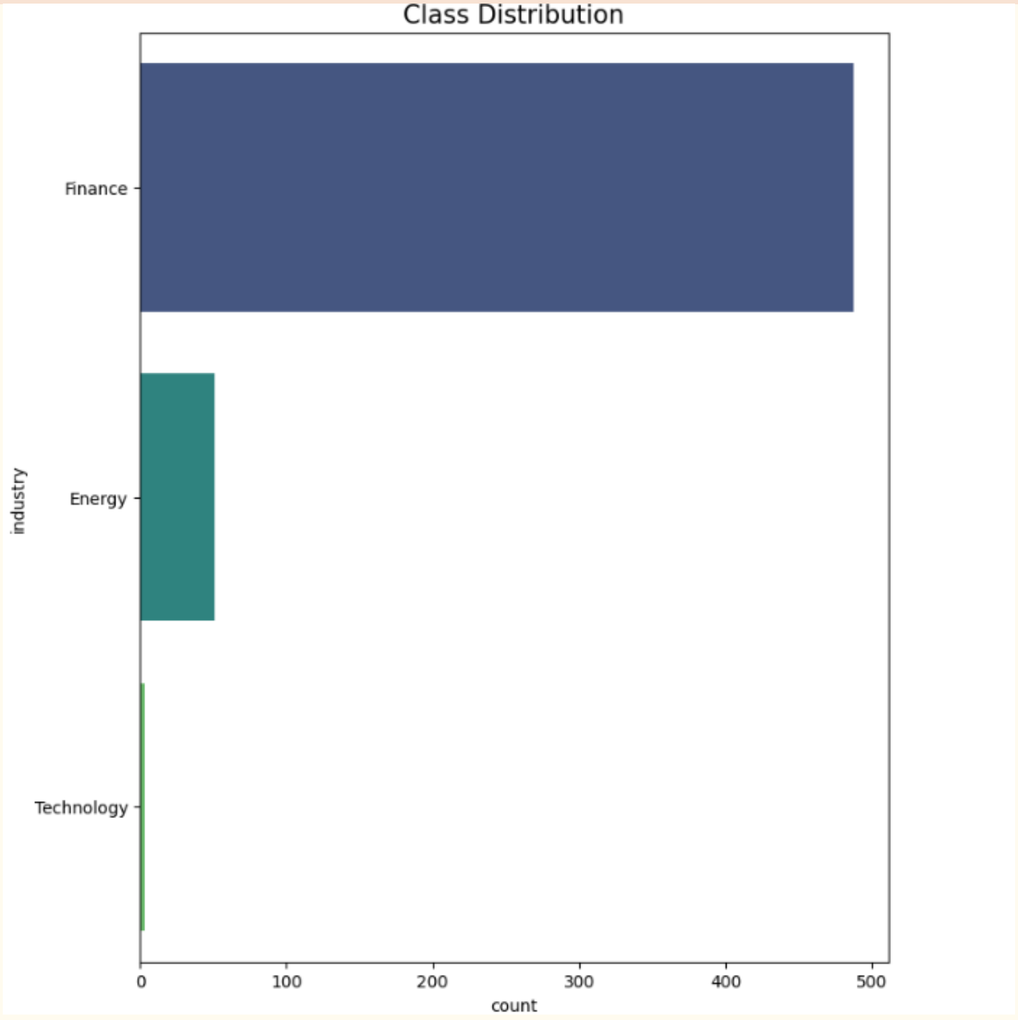
1.NLTK =>TOKENIZING , STOP WORDS REMOVAL, LEMMETIZATION

2.TFIDF => VECTORIZE STRING INTO NUMERICAL VALUES

SUPPORT VECTOR MACHINE VS NAIVE BAYES

```
sns.countplot(y='svm_industry', palette='viridis', data=test)
```





ACCURACY =0.9898580121703854

F1 score: [0.9921875 0.99365751 0.94915254 0.98550725 1.]

ACCURACY= 0.7606490872210954

F1 score: [0.90128755 0.80135823 0. 0.48351648 0.57142857]

FEATURES OF OUR SOLUTION

1] EMAIL CLASSIFICATION BASED ON DEPARTMENTS => SVM

2] ROUTING THE CLASSIFIED EMAILS INTO THE SPECIFIC DEPARTMENTS => SMTP

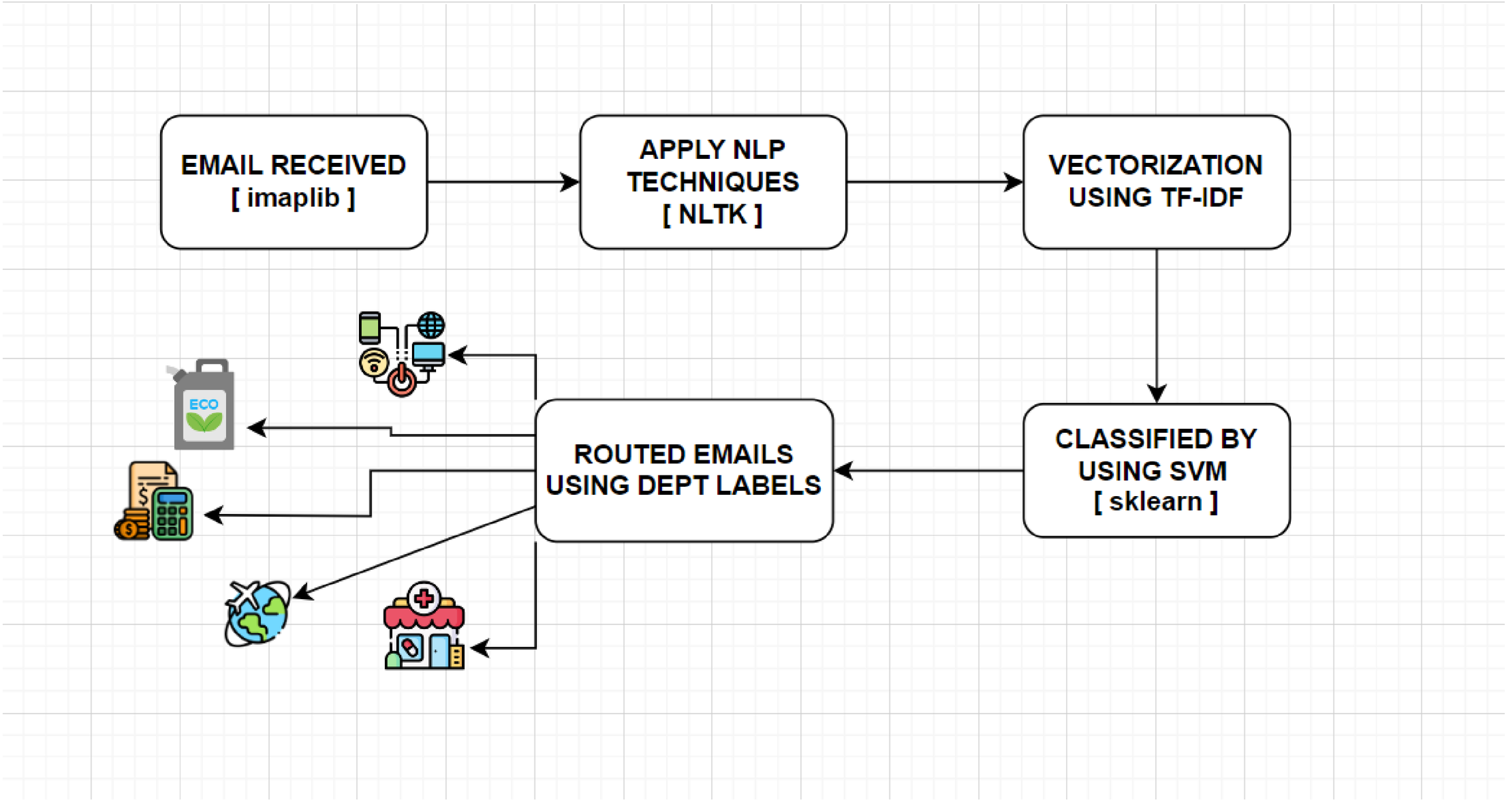
3] SENTIMENT ANALYSIS TO CLASSIFY INTO SATISFACTION ,CONCERN ,ENQUIRY ETC

OBSTACLES WE OVERCAME.....

1] MODEL SELECTION FOR BEST ACCURACY

2] DECIDING ON LIBRARIES FOR OUR DATASET

3] REAL-TIME INTEGRATION WITH EMAILS


EMAIL RECEIVED
[imaplib]
APPLY NLP
TECHNIQUES
[NLTK]
VECTORIZATION
USING TF-IDF
CLASSIFIED BY
USING SVM
[sklearn]
ROUTED EMAILS
USING DEPT LABELS
ECO

FUTURE SCOPE

1] EMAIL URGENCY NOTIFICATION

2] EASE OF DEPLOYMENT USING CONTAINRIZATION

3] TEXT SUMMARIZATION OF LARGE EMAILS FOR QUICK OVERVIEW

THANK YOU!!!