# SPAM MAIL DETECTION SYSTEM USING DEEP LEARNING

## SEMINAR REPORT

*Submitted by*

### Dushyant Krishna Sharma

*7323*

***in partial fulfillment for the award of the degree of***

## BACHELOR OF ENGINEERING

***in***

**THIRD YEAR COMPUTER ENGINEERING**

****

**ARMY INSTITUTE OF TECNOLOGY, PUNE 2025-26**



# ARMY INSTITUTE OF TECHNOLGY, PUNE

## DEPARTMENT OF COMPUTER ENGINEERING

CERTIFICATE

***This is to certify that the Seminar titled***

# SPAM MAIL DETECTION SYSTEM USING DEEP LEARNING

**has been prepared and presented by**

### Dushyant Krishna Sharma

**of Third Year (Computer Engineering)**

**in partial fulfillment of requirement for the award of**

**Degree of Bachelor of engineering in computer Engineering under the Savitribai Phule Pune University during the academic year 2025-26**

**SEMINAR GUIDE HEAD OF THE DEPARTMENT**

Prof. Vaishali Ganganwar Prof. Sunil Dhore

# ACKNOWLEDGEMENT

I would like to give heartful appreciation for my seminal guide, **Dr. Vaishali Ganganwar**, from computer engineering department. Her guidance provides me a definite path, which encouraged me for solving difficult problems, and helped me to learn step by step. They provide me their life learning suggestions and feedback which were very helpful for building this project, I would like to express my gratitude to her. She is one of my motivation and inspiration sources.

I give a sincere thanks to Head of department, **Dr. Sunil Dhore**, and all faculty to provide me all learning resources for this seminar of research. Their support provides me a path to succeed in this project.

At last, I am thankful to my friends and teachers, and my family. Their support guide me throughout my whole journey.

#### Dushyant Krishna Sharma (7323)

TE Computer-B

# ABSTRACT

The main mode of professional and personal communication in this world is Email, and main exploit ship of scammers is Email. Email works like transporter of data, which can be used by socialized or unsocialized people or organizations. It is very important to check whether email is safe or not. Spam mails are very notorious things which can degrade user experience, can damage user machines, any embedded thing in that email may harm user data privacy, or can be expose the user to scammers. It is very important to detect scams and protect user’s privacy. In previous years scams are significantly increasing day by day. Many organizations have developed email scam detection systems but are using classical and slower methods for users. This method can detect scams but not efficiently. Scammers easily bypass old detection systems. Scammers are evolving their techniques. It is important to develop with them. In this project I used modern deep learning techniques to stop and flag scam mails. I used transformers and neural networks to develop a defense which can flag potential emails to ham and spam.

I investigate spam emails using three different and powerful modern neural network architecture. The first model I used is Long-Short-Term Memory(LSTM) model, it captures long range dependencies in the text, also forget unnecessary things. The second one is Bidirectional Long Short Term Memory model which works same as LSTM but reads data bidirectionally. BiLSTM model is more advanced model as it detects data bidirectionally, it can reform and understand data patterns in both directions. The third model I used is DistilBERT transformer model, it is most advanced classifier model. This works on encoding and decoding data, by this it can understand data more efficiently, also it can understand basic word patterns of English language, it can be so accurate. Mostly deliver high accuracy among all models.

I used publicly available dataset named ( [mail\_data.csv](https://docs.google.com/spreadsheets/d/1KJIO9YikjzHQvE9mPUaw-ZBP96PgZpSQE4HkNf1Q_Xs/edit?gid=285209673&gid=285209673) ) to train these models. This data is used for preprocessing steps, to prepare it for the models. Each model was trained separately; I marked progress and result after training and used it for identification of best performed model. I classify models based on combined score which was calculated using four parameters : accuracy, precision, F1 score, and recall. The model which performed with high compound score is the best performed model. The outcome can represent advanced approach to detect and mitigate the risk bound with spam mail.

# TABLE OF CONTENTS

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. No.** | | **Title of Chapter** | **Page No.** |
| **\*\*** | | [**Certificate**](#_bookmark0) | **2** |
| **\*\*** | | [**Acknowledgment**](#_bookmark1) | **3** |
| **\*\*** | | [**Abstract**](#_bookmark2) | **4** |
| **1.** | | [**Introduction**](#_bookmark3) | **7 – 9** |
|  | 1.1 | Overview | 7 |
|  | 1.2 | Motivation | 7 |
|  | 1.3 | Problem definition | 8 |
|  | 1.4 | Approach | 9 |
| **2.** | | [**Literature Survey**](#_bookmark4) | **10 – 13** |
|  | 2.1 | Overview | 10 |
|  | 2.2 | Traditional Machine Learning Approaches | 10 |
|  | 2.3 | Feature Engineering and Text Representation | 11 |
|  | 2.4 | Deep Learning Based Approaches | 12 |
|  | 2.5 | Transformer Based Architecture | 12 |
|  | 2.6 | Comparative Analysis | 13 |
| **3.** | | [**Methodology**](#_bookmark5) | **14 – 18** |
|  | 3.1 | Software and hardware requirements | 14 |
|  | 3.2 | System design architecture | 15 |
|  | 3.3 | Overview of project models | 16 |
|  | 3.4 | Compound Score | 17 |
| **4.** | | [**Experiments And Results**](#_bookmark12) | **19 – 25** |
|  | 4.1 | Model Performance comparison | 19 |
|  | 4.2 | Error Analysis with confusion matrices | 20 |
|  | 4.3 | Resource Utilization | 23 |
|  | 4.4 | Testing and Result | 25 |
| **5.** | | [**Conclusion and future scope**](#_bookmark22) | **26** |
|  | 5.1 | Conclusion | 26 |
|  | 5.2 | Future scope | 26 |
| **6.** | | [**References**](#_bookmark23) | **27 – 30** |

## TABLES AND FIGURES

#### Dataset Description

* In spam mail detection system project, I used mainly spam/ham dataset, basically there are two datasets. They are as follows :

#### Dataset 1(5,573 samples) :

 [File size – 475 kb](https://docs.google.com/spreadsheets/d/1KJIO9YikjzHQvE9mPUaw-ZBP96PgZpSQE4HkNf1Q_Xs/edit?gid=285209673&gid=285209673)

#### Dataset 2(1,777 samples) :

 [File size – 2.59 mb](https://www.kaggle.com/datasets/shalmamuji/spam-email-classification)

* + Dataset 3(5,329 samples) :

 [File size – 24.53 mb](https://www.kaggle.com/datasets/ganiyuolalekan/spam-assassin-email-classification-dataset)

#### Dataset 2(2,14,845 samples) :

 [File size – 357.68 mb](https://www.kaggle.com/datasets/meruvulikith/190k-spam-ham-email-dataset-for-classification)

**List of Formulae**

* [**Compound Score**](#_bookmark7)
* [**Accuracy**](#_bookmark8)
* [**Precision**](#_bookmark9)
* [**Recall**](#_bookmark10)

**List of Figures**

* [**Score comparison of models(Lstm, BiLstm, DistilBERT)**](#_bookmark13)
* [**Confusion Matrix for LSTM**](#_bookmark14)
* [**Confusion Matrix for BiLSTM**](#_bookmark15)
* [**Confusion Matrix for DistilBERT**](#_bookmark16)
* [**GPU usage.**](#_bookmark19)
* [**CPU usage.**](#_bookmark17)
* [**Network traffic (bytes).**](#_bookmark20)
* [**Disk I/O utilization (mb).**](#_bookmark18)
* [**System Design Architecture**](#_bookmark6)
* [**Result**](#_bookmark21)

## CHAPTER – 1 INTRODUCTION

#### Overview

This project aims to identify spam mail in our mailbox by identifying and processing various Deep Learning models like LSTM, BiLSTM, DistilBERT. As we add and modify an Efficient testing model with these models it is easy to identify spam based on various datasets and unorganized data. As various datasets have multiple parameters depend on models, like Precision, Accuracy, F1 score etc.

This problem can be solved easily by classical methods of machine learning like logistic regression method or random forest method but, neural networks are some advance methods to identify and solve the related problems effectively. Especially DistilBERT model is very fast and easy to train, because only first layer of this model is used for training that’s why it is also known as frozen model.

In our day to day, it is very important to build an efficient method to solve such problems, like phishing attack, or spoofing etc.

#### Motivation

In this big word there are many good beings and many bad also, scammers are scamming good deeds, so here comes our small motivation. I shaped my imagination into this model and make an efficient Efficient approach model which can be used anywhere with any kind of data written in English language. As I was interested in machine learning I built that project with help of my teacher, guide **Dr. Vaishali Ganganwar .** Her guidance provides me a definite path, which encouraged me for solving difficult problems, and helped me to learn step by step. They provide me their life learning suggestions and feedback which were very helpful for building this project, I would like to express my gratitude to her. She is one of my motivation and inspiration sources.

In previous years scams are significantly increasing day by day. Many organizations have developed email scam detection systems but are using classical and slower methods for users. This method can detect scams but not efficiently.

Scammers easily bypass old detection systems. Scammers are evolving their techniques. It is important to develop with them. In this project I used modern deep learning techniques to stop and flag scam mails. I used transformers and neural networks to develop a defense which can flag potential emails to ham and spam.

It is very important to check whether email is safe or not. Spam mails are very notorious things which can degrade user experience, can damage user machines, any embedded thing in that email may harm user data privacy, or can be expose the user to scammers. It is very important to detect scams and protect user’s privacy.

#### Problem Definition

The specific goal of this project is to build an effective way to determine spam mail using a modern, fast, and efficient method.

The model should be optimal in four parameters : Accuracy, Precision, F1-Score, and recall. We can develop a compound score and simply compare these models. Basically, average of these four scores should be compound score.

This method can efficiently reduce harm caused by spam mail to users, as it can be detected easily and fast by my model. So, goal for using three models is to compare these models in real time, which is best for data, which is best for speed, which is best for memory, which is best for resources etc.

This can be used by a local or a public machine as their requirement. It can be used in real time detection simply after training. We can use reinforcement learning to train and test our models in real time. As there are only two classes for mail spam or ham, we can do our model train only in one class and if result is not identified, it is another class. This is called one class learning.

#### Approach

I solved this problem by train and comparing models step by step, after that I choose best of them to show which model is efficient for data. I solved this problem by performing following steps :

* + - I loaded the dataset (mail\_data.csv) having details of ham and spam messages into my code. Convert into binary format where I labeled ham = 0 and spam = 1. After that I split the training and testing sets into ratio (80:20).
    - Then I process the text for LSTM and BiLSTM into tokens using Keras Tokenizer. Padded sequence to a fixed length (max\_length = 100).
    - Train the data on LSTM and BiLSTM, after that evaluation of models on text data like accuracy, precision, recall, F1 score.
    - After that I use hugging face DistilBert Tokenizer Fast and DistilBert for Sequence Classification for fine tuning, encoding the training and testing data with tokenizer, after that I used hugging face trainer for training of DistilBERT with metrices accuracy, precision, recall, F1 score.
    - Finally, I conclude which model gives best performance by taking average of all four parameters and simply comparing them.

## CHAPTER – 2 LITERATURE SURVEY

#### Overview

Spam detection has been a wide domain of research for 20 years, because of exponential growth of spam. The main solution is advancement in Natural language processing (NLP). Transformers based advancement and use of neural networks is significantly increased day by day, this research is mainly based on deep learning and transformer based neural networks.

This chapter is for giving a brief discussion about work into four categories : traditional machine learning approaches, features engineering, deep learning approaches, transformer-based architecture.

#### Traditional Machine Learning Technologies

Traditional approaches are important to understand as they are foundations of spam detection research, these were very simple, effective, and direct for small datasets.

#### Naïve Bayes and Decision Trees

Naïve Bayes classifier is used earlier as an algorithm for spam detection; they are important for detecting conditional independence. It allows fast probability-based classification using word frequencies. It is simple and works as a baseline for spam detection, decision tree-based system can also identify spam by URLs, keywords, or unusual addresses. However, both techniques suffer from overfitting and poor handling.

#### Support Vector Machines (SVMs)

Support Vector Machines are very popular due to their ability to handle high-dimensional text data efficiently. Using kernels SVMs create a boundary between ham and spam. By research we can see that SVMs are more advanced than Naïve bayes in accuracy and precision. But unfortunately, SVMs are expensive for large datasets and do not capture information so efficiently.

#### Feature Engineering and Text Representation

As spammers are using advanced methods for phishing, researchers conclude that choice of features was as important as the algorithm itself.

#### Bag – of – Words and TF-IDF

It was a simple feature representation. It treats text as an unordered collection of words. These models are easy to implement, while Bag-of-word (BOW) ignored word order and contextual meaning. Term Frequency – Inverse Document Frequency (TF – IDF) are widely used to overcome frequency bias. TF – IDF emphasized terms that mostly appeared in spam but rarely in ham, it divides way better than its previous methods.

#### Word Embeddings

Word2vec and GloVe, like distributed introduction, represent a significant shift. For instances word like “offer”, “free”, and “discount” were placed close together in a vector space, making it more robust.

Researchers integrate it with classical classifiers, results in higher detection of accuracy from TF

– IDF.

#### Deep Learning Based Approaches

Because of success of Deep Learning in computer vision and NLP, neural networks are highly implemented in spam mail detection systems.

#### Recurrent Neural Networks (RNNs) and LSTM

RNN is limited due to its limitations of ability to learn long term dependencies, However Long Short – Term Memory (LSTM) network can remember sequence of words rather than just their frequencies. They reduced false positives compared to SVMs and Naïve Bayes.

#### Bidirectional LSTM (BiLSTM)

It is like LSTM, but the difference is that it processes text from both directions, left to right and right to left. It enables context from past and future tokens simultaneously. Research suggests that BiLSTM is more accurate and precise than LSTM.

#### Transformers – Based Architectures

This is advance in nature, Transformer based architectures have self-attention, natural language processing power and can enable parallel training on large corporations. I think that spam detection system will soon adapt transformer models for better text understanding.

#### BERT and Its Variants

Bidirectional encoder representation transformer (BERT) is pretrained on large scale data provided by corporations can be fine – tuned on given dataset. DistilBERT is a lightweight model that provides exceptional accuracy while training and tests, it is much more suitable for real time testing. These types of models can detect sarcasm strategies used by scammers.

#### Hybrid and Efficient Models

During my research I studied Efficient and hybrid models and studies suggested that using these models can increase model strength drastically. For instance, LSTM or BiLSTM + BERT gives high precision and recall. These types of approaches can generate efficient models having minimum errors, and making the system strong against adversarial spam designed to bypass filters.

#### Comparative Analysis

* **Traditional models (Naïve Bayes, SVM)** give strong foundation but are not so advance for modern spams.
* **Feature engineering** needs more improvement as it becomes old.
* **Deep Learning (LSTM, BiLSTM)** introduced sequence modelling and automate handling.
* **Transformers (BERT, DistilBERT)** are more accurate against modern spam mails and can handle new benchmarks for future technology.

## CHAPTER – 3

### Methodology

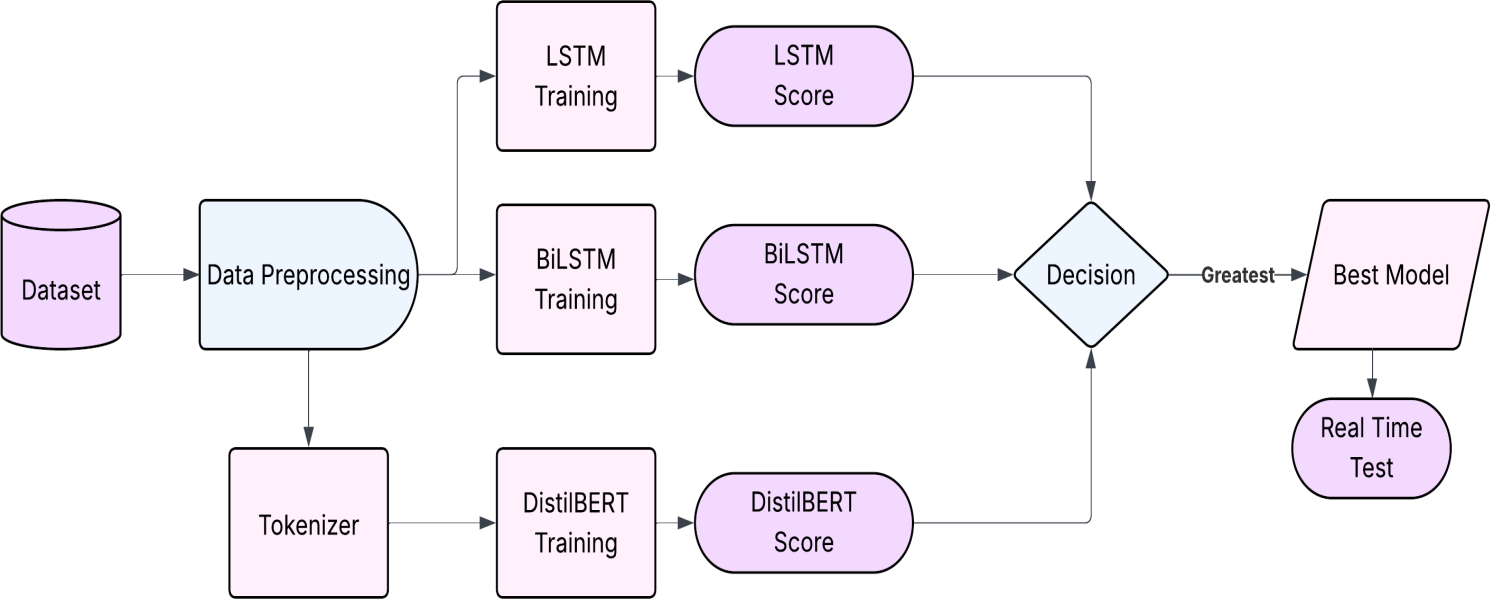
#### Software and Hardware Requirements

* + 1. **Software requirements**
* Operating system
  + Windows 10/11 (64 – bit) or Ubuntu Linux 20.04/22.04
* Programming Language
  + Python 3.8 or above
* Libraries and Frameworks
  + TensorFlow / Keras
  + Pytorch
  + Transformers (Hugging Face)
  + Scikit – Learn
  + NumPy and Pandas
  + Matplotlib and Seaborn
* Development Environment
  + Jupyter Notebook / Google Colab / PyCharm / Vscode

#### Hardware Requirements

* Minimum configuration
  + Processor : Intel i3 / AMD Ryzen 3 or equivalent (2.0 GHz or above)
  + RAM : 8 GB
  + Storage : 250 GB HDD / 128 GB SSD
  + GPU : Not mandatory, but training will be slow on CPU
* Recommended Configuration
  + Processor : Intel i5 / i7 or AMD Ryzen 5 / 7 (2.5 GHz or above, multi – core)
  + RAM : 16 GB or more
  + Storage : 512 GB SSD
  + GPU : NVIDIA GPU with CUDA support
  + Internet Connection

#### System Design Architecture

****

* 1. **Overview of Project Models**

For making an effective and efficient project I compared these three powerful models and select best model based on compound score, I know that every model has its own strengths for handling and managing data that’s why I calculated compound score by taking four parameters accuracy, precision, recall, and F1 score.

#### Long Short – Term Memory (LSTM) Model

* LSTM is a powerful type of **Recurrent Neural Network (RNN)** designed to eliminate vanishing gradient problem of traditional RNNs.
* This model has three gates input, output, and forget, it helps to understand text sequences and patterns and can remember long gaps in patterns. This type of architecture helps it to remember long – term dependencies in text sequences.
* In spam detection it learns similar patterns like “win it now” or “free offer”, which are

strongly represents spam like situations.

* This model is a baseline deep learning approach.

#### Bidirectional LSTM (BiLSTM) Model

* It is like LSTM model, but it processes the data from both directions, left to right and right to right, that’s why it increases the model’s understanding and intelligence.
* This type of reading allows you to learn past and future context simultaneously, it improves its understanding of word dependencies.
* Let’s take an example “get ready for a good news, claim your price” the words “claim” and “price” are spam words and can be understood fast if we read that text from right to left direction.
* From recent studies I found that BiLSTM provides much more accuracy than unidirectional LSTM.

#### DistilBERT (Transformer – Based Model)

* It is a transformer model based on encoding and decoding evaluation of text, it needs less resources from a BERT model as it is partially pretrained, it works on self – attention instead of recurrence.
* It is not like LSTM or BiLSTM, it does not read text sequentially, it processes it parallelly, can understand deeper contextual meaning.
* It is pre – trained on large corpora and then I fine tuned it on my publicly available dataset.
* It can balance efficiency and accuracy, that’s why it is suitable for real time spam detection.

#### Compound Score

I used compound score in this project for evaluating which model is best among LSTM, BiLSTM, DistilBERT. Basically, it is the average of four key evaluation metrics. It allows us to calculate a balance score between models as these all are positive feedback about models; we can take average of four and simply calculate best model.

𝑐𝑜𝑚𝑝𝑜𝑢𝑛𝑑 𝑆𝑐𝑜𝑟𝑒 =

𝐴𝑐𝑐𝑢𝑟𝑎𝑐𝑦 + 𝑃𝑟𝑒𝑐𝑖𝑠𝑖𝑜𝑛 + 𝑅𝑒𝑐𝑎𝑙𝑙 + 𝐹1 − 𝑆𝑐𝑜𝑟𝑒 4

(𝑒𝑞𝑢𝑎𝑡𝑖𝑜𝑛 − 1)

To understand compound score we need to understand about these four criteria named as accuracy, precision, recall, and F1 – score.

#### Evaluation Metrics

In such type of problem where we have binary yes or no types of classification, the performance of a model is determined by four key parameters : accuracy, precision, recall, and F1-score. These metrics are known as confusion metrics. These metrics have four classifications :

* True Positive (TP) : Spam correctly identified as spam.
* True Negative (TN) : Ham correctly identified as ham.
* False Positive (FP) : A ham incorrectly classified as spam.
* False Negative (FN) : A spam incorrectly classified as ham.

#### Accuracy

It is key parameters which determine totally correct classified instances out of all instances.

𝐴𝑐𝑐𝑢𝑟𝑎𝑐𝑦 =

𝑇𝑃 + 𝑇𝑁

𝑇𝑃 + 𝑇𝑁 + 𝐹𝑃 + 𝐹𝑁

(𝑒𝑞𝑢𝑎𝑡𝑖𝑜𝑛 − 2)

#### Precision

This key parameter is the ratio of how many emails predicted as spam are spam upon total emails predicted for spam.

𝑃𝑟𝑒𝑐𝑖𝑠𝑖𝑜𝑛 =

𝑇𝑃

𝑇𝑃 + 𝐹𝑃

(𝑒𝑞𝑢𝑎𝑡𝑖𝑜𝑛 − 3)

#### Recall

It determines how many of the actual spam mails are correctly detected by the model, it is cost of false negatives.

𝑅𝑒𝑐𝑎𝑙𝑙 =

𝑇𝑃

𝑇𝑃 + 𝐹𝑁

(𝑒𝑞𝑢𝑎𝑡𝑖𝑜𝑛 − 4)

#### F1 – Score

It is the harmonic mean of precision and recall, it provides balance measure when dataset is imbalanced.

𝐹1 − 𝑆𝑐𝑜𝑟𝑒 =

2 × 𝑃𝑟𝑒𝑐𝑖𝑠𝑖𝑜𝑛 × 𝑅𝑒𝑐𝑎𝑙𝑙

𝑃𝑟𝑒𝑐𝑖𝑠𝑖𝑜𝑛 + 𝑅𝑒𝑐𝑎𝑙𝑙

2𝑇𝑃

=

2𝑇𝑃 + 𝐹𝑃 + 𝐹𝑁

(𝑒𝑞𝑢𝑎𝑡𝑖𝑜𝑛 − 5)

So, using equations 1,2,3,4,5 we get

𝐶𝑜𝑚𝑝𝑜𝑢𝑛𝑑 𝑆𝑐𝑜𝑟𝑒 = (

1

𝑇𝑃 + 𝑇𝑁

4 𝑇𝑃 + 𝑇𝑁 + 𝐹𝑃 + 𝐹𝑁 𝑇𝑃 + 𝐹𝑃 𝑇𝑃 + 𝐹𝑁 2𝑇𝑃 + 𝐹𝑃 + 𝐹𝑁

+

𝑇𝑃

+

𝑇𝑃

+

2𝑇𝑃

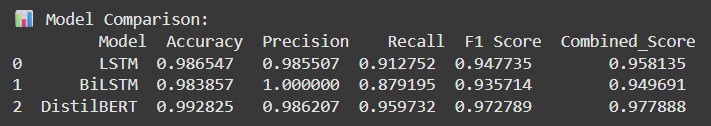
)

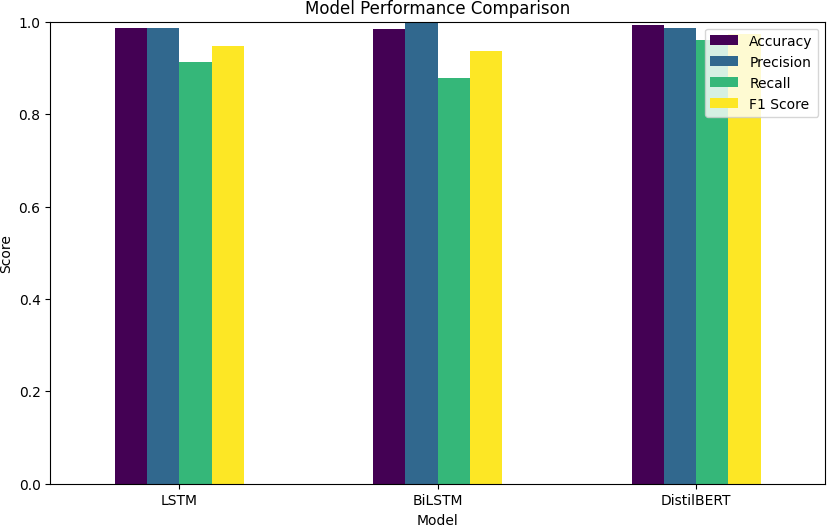
We can use this compound score to compare various models on same dataset.

## CHAPTER – 4

### Experiments and Results

#### Model Performance Comparison

****

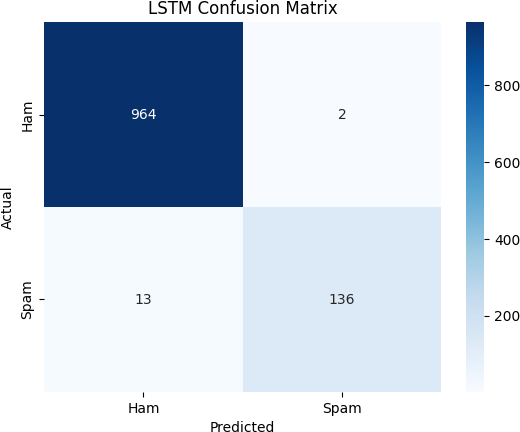
****

As we can see all three models are so powerful, but DistilBERT has more recall and F1 – Score that’s why it can achieve high performance. This represents its capability in correctly identifying spam and ham.

#### Error Analysis with Confusion Matrices

To understand each model capabilities, I analyzed their confusion matrix.

#### LSTM confusion Matrix

****

Here we can see that LSTM models can identify 964 messages correctly as Ham, and 136 messages as spam.

Here,

TP = 136, TN = 964, FP = 2, FN = 13.

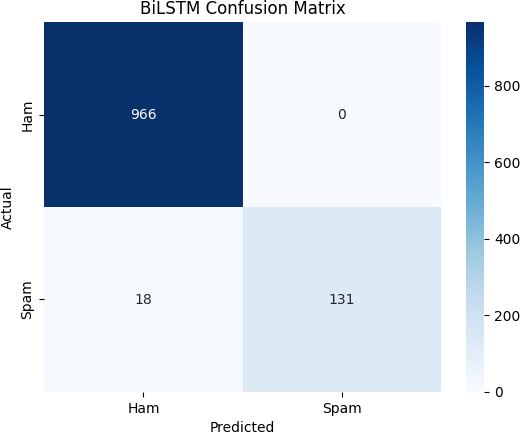
So,

* + - * Accuracy = 1100 / 1115 = 0.9865 (using --- [equation 2](#_bookmark8))
      * Precision = 136 / 138 = 0.9855 (using --- [equation 3](#_bookmark9))
      * Recall = 136 / 149 = 0.9127 (using --- [equation 4](#_bookmark10))
      * F1 – Score = 1.799 / 1.898 = 0.9478 (using --- [equation 5](#_bookmark11))

Hence ,

* + - * + Compound score = 0.9778 (using --- [equation 6](#_bookmark7))

#### BiLSTM confusion Matrix

****

The BiLSTM model can identify 966 Ham messages correctly as Ham, and 131 Spam messages Correctly as Spam.

Here,

TP = 131, TN = 966, FP = 0, FN = 18

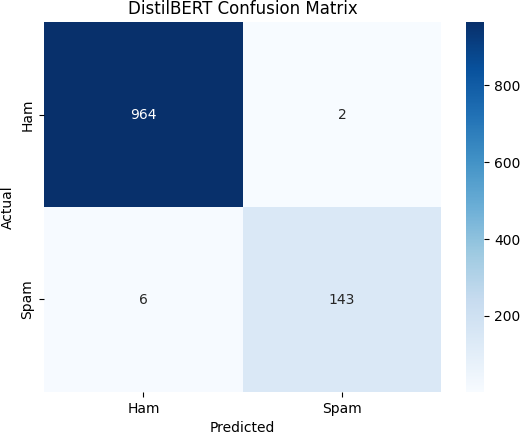
So,

* + - * Accuracy = 1097 / 1115 = 0.9838 (using --- [equation 2](#_bookmark8))
      * Precision = 131 / 131 = 1.0000 (using --- [equation 3](#_bookmark9))
      * Recall = 131 / 149 = 0.8791 (using --- [equation 4](#_bookmark10))
      * F1 – Score = 1.799 / 1.898 = 0.9478 (using --- [equation 5](#_bookmark11))

Hence,

* + - * + Compound Score = 0.9496 (using --- [equation 1](#_bookmark7))

#### DistilBERT confusion Matrix

****

The DistilBERT Model can identify 964 Ham messages as Ham, and 143 Spam messages as Spam.

Here,

TP = 143, TN = 964, FP = 2, FN = 6

So,

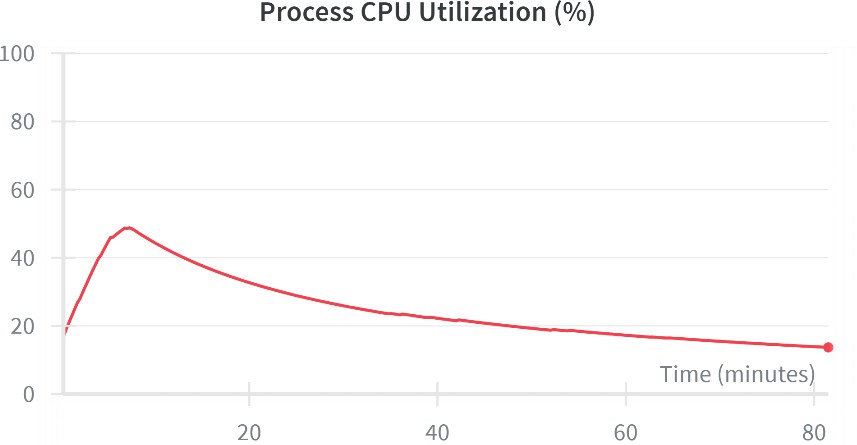
* + - * Accuracy = 1107 / 1115 = 0.9928 (using --- [equation 2](#_bookmark8))
      * Precision = 143 / 145 = 0.9862 (using --- [equation 3](#_bookmark9))
      * Recall = 143 / 149 = 0.9597 (using --- [equation 4](#_bookmark10))
      * F1 – Score = 1.893/1.946 = 0.9727 (using --- [equation 5](#_bookmark11))

Hence,

* + - * + Compound Score = 0.9778 (using --- [equation 1](#_bookmark7))

#### Resource utilization

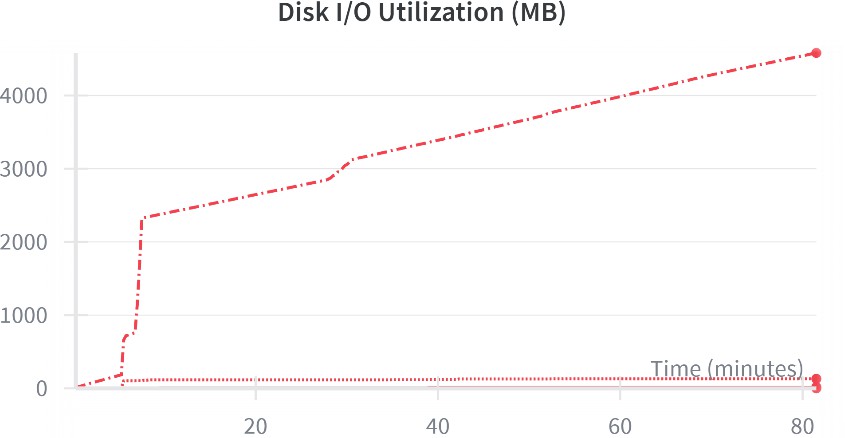
* + 1. **Process CPU utilization per minute**

****

Here we can see that

* Peak CPU utilization = 45 % (approx.)
* Minimum CPU utilization = 17% (approx.)
* Average CPU utilization = 31% (approx.)

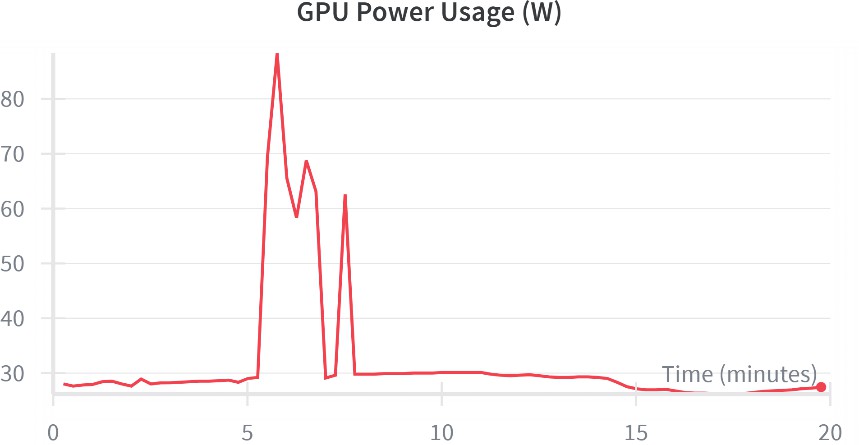
#### Disk I/O utilization

****

We can see that here disk utilization while I/O cycle is

* Maximum = 2.2 GB/minute (When process end)
* Minimum = 240 MB/minute (When process starts)

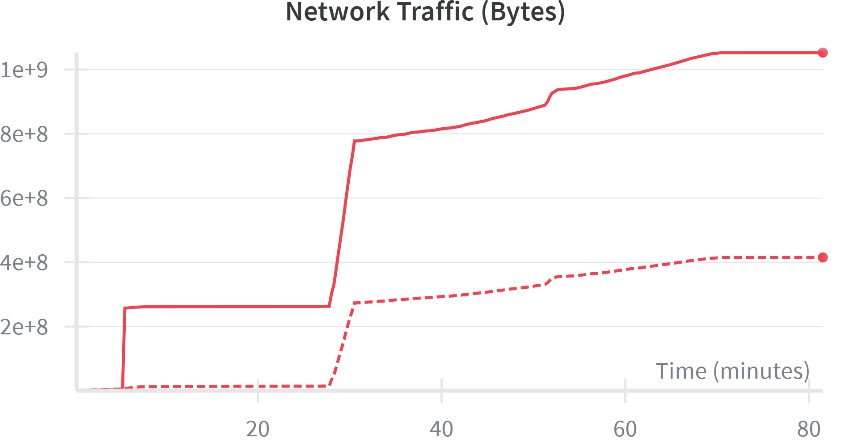
#### GPU power consumption (W)

****

Here we can see that,

* Maximum GPU usage = 87 W
* Minimum GPU usage = 28 W

#### Network Traffic utilization (bytes) in Google – Colab

****

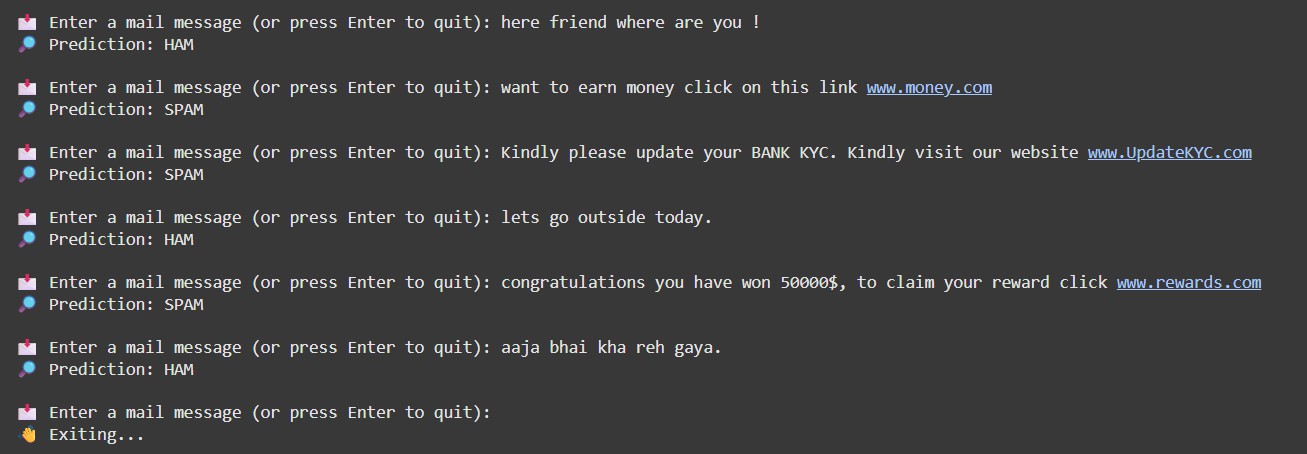
Here we can see that,

* Total received data = 1.05 GB or 1.05 × 109 bytes.
* Total uploaded data = 0.42 GB or 4.2 × 108 bytes.

#### Testing and Result

I write a python script to test this approach and DistilBERT is evaluated best, the script prompts

user to enter a message and model predicts it’s as a HAM or SPAM message.



This real time testing demonstrates the model’s ability to discern context, identify spam – related keywords, and even handle multilingual text.

Here we can see that our model works exceptionally well, correctly classify messages, and can stop phishing messages effectively.

## CHAPTER – 5

**CONCLUSION AND FUTURE SCOPE**

#### Conclusion

My project successfully compared three advanced neural network models : LSTM, BiLSTM, DistilBERT, and concluded that **DistilBERT** having compound score of 0.9778 and having phenomenal accuracy of 0.9928 is the most efficient and advanced model.

And practical real time testing by taking user input verified the powerful nature of DistilBERT. The result of this experiment demonstrates that transformer-based models are very fast and effective for text understanding. Also, DistilBERT model got highest F1 – Score among all models during my experiment, which was 0.9727. The resulting F1 – Score represents that it has highest balance between false positives and false negatives.

In conclusion, this project highlights the use of neural networks in NLP for Cybersecurity and justify the power of transformer – based models among other models.

#### Future Scope

While my project successfully compared three advanced neural networks models, several changes can be made during future work :

* + - Exploring advance models : we can use more transformer – based models such as RoBERTa, ALBERT, or the full BERT – based models to achieve more accuracy.
    - Dataset Expansion : we can train our models by using more extensive and diverse datasets, having multilingual characters and newer forms of spam.
    - Real – Time Implementation : we can make a lightweight API which can do real – time spam filtering on user’s device.
    - Hyperparameter optimization : we can use optimal hyperparameters using techniques like Grid – Search or Bayesian optimization to further fine – tune our model.

## CHAPTER – 6 REFERENCES

1. P. Jain, S. Singh and C. K. Saxena, "Detecting Email Spam with NLP: A Machine Learning Approach," 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT), Greater Noida, India, 2024, pp. 393-398, doi: 10.1109/IC2PCT60090.2024.10486769. keywords: {Training;Filtering;Unsolicited e- mail;Text categorization;Semantics;Machine learning;Vectors;Email spam detection;Machine learning;Natural Language Processing (NLP);Naive Bayes classification;Text vectorization;Text classification},
2. I. Vasudevan, M. M. Nasurudeen, M. S. Kumar, R. M. D. Charaan, S. A. Kumar and L. Jenefa, "Hierarchical Model for Email Fraud Detection Using Naïve Bayes and SVM," 2024 First International Conference for Women in Computing (InCoWoCo), Pune, India, 2024, pp. 1-6, doi: 10.1109/InCoWoCo64194.2024.10863458. keywords: {Support vector machines;Accuracy;Machine learning algorithms;Filtering;Unsolicited e- mail;Phishing;Vectors;Fraud;Naive Bayes methods;Classification algorithms;Support Vector Machine;Naive Bayes;Email Fraud Detection;Machine Learning;Spam Detection},
3. I. T. Almalkawi, M. F. Al-Hammouri, M. A. Mallouh and T. Barakat, "Improving Email Security Through Machine Learning-Based Phishing Attack Detection," 2024 International Jordanian Cybersecurity Conference (IJCC), Aِ mman, Jordan, 2024, pp. 180- 185, doi: 10.1109/IJCC64742.2024.10847270. keywords: {Deep

learning;Accuracy;Phishing;Prevention and mitigation;Supervised learning;Organizations;Predictive models;Feature extraction;Electronic mail;Protection;Email Security;Phishing Attacks;Artificial Intelligence;Machine Learning;Deep Learning},

1. Z. Aslanpençesi, M. Baykara and T. B. Alakuş, "SMS Phishing Detection with Hybrid CNN-GRU," 2024 7th World Symposium on Communication Engineering (WSCE), Tokyo, Japan, 2024, pp. 52-56, doi: 10.1109/WSCE65107.2024.00015. keywords:

{Training;Deep learning;Phishing;Banking;Message services;Tokenization;Mobile handsets;Data models;Electronic mail;Convolutional neural networks;smishing attacks;information security;deep learning},

1. M. RAZA, N. D. Jayasinghe and M. M. A. Muslam, "A Comprehensive Review on Email Spam Classification using Machine Learning Algorithms," 2021 International Conference on Information Networking (ICOIN), Jeju Island, Korea (South), 2021, pp. 327-332, doi: 10.1109/ICOIN50884.2021.9334020. keywords: {Support vector machines;Machine learning algorithms;Unsolicited e-mail;Machine learning;Feature extraction;Information filters;Classification algorithms;Spam Detection;Spam Classification;Spam Filter;E- mail;Supervised Learning;Machine Learning Algorithms;Email Classification;Spam Email Detection;Email Categorization;Email Feature Set Analysis;Spam Detection Using Machine Learning Algorithms},
2. J. Sultana and T. Siddiqui, "A Deep Learning Approach to Classify Spam URLs Data using MLP," 2025 8th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2025, pp. 767-773, doi: 10.1109/ICCMC65190.2025.11140908. keywords: {Deep

learning;Measurement;Accuracy;Unsolicited e-mail;Computational modeling;Robustness;Bayes methods;Decision trees;Spam URLs;Multi-Layer Perceptron (MLP);Decision Tree (DT);Logiboost (LB);Naive Bayes (NB)},

1. A. M. S, D. S, S. R, K. T, C. M and V. L, "An Improved Spam Detection Technique using Deep and Recurrent Neural Networks," 2025 8th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2025, pp. 1672-1678, doi: 10.1109/ICCMC65190.2025.11140957. keywords: {Deep learning;Recurrent neural networks;Phishing;Scalability;Artificial neural networks;Message services;Mobile communication;Electronic mail;Security;Postal services;Short Message Service spam detection;Recurrent Neural Network;Deep Neural Network},
2. F. Hossain, M. N. Uddin and R. K. Halder, "Analysis of Optimized Machine Learning and Deep Learning Techniques for Spam Detection," 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Toronto, ON, Canada, 2021, pp. 1-7, doi: 10.1109/IEMTRONICS52119.2021.9422508. keywords: {Deep learning;Training;Radio frequency;Heating systems;Recurrent neural networks;Mechatronics;Unsolicited e-mail},
3. G. Lakshmi and P. Swetha, "Advanced Phishing Website Detection Techniques in Internet of Things Using Machine Learning," 2024 4th International Conference on Mobile Networks and Wireless Communications (ICMNWC), Tumkuru, India, 2024, pp. 1-8, doi: 10.1109/ICMNWC63764.2024.10872066. keywords: {Surveys;Wireless communication;Uniform resource locators;Privacy;Reviews;Social networking (online);Phishing;Machine learning;Feature extraction;Internet of Things;phishing;URL;website phishing;phishing detection;machine learning;black lists;white lists;deep learning;heuristic based detection},
4. S. Juneja, R. Devi, B. S. Bhati and D. Tiwari, "An analytical review of Email Spam Recognition Using Deep Learning," 2024 4th International Conference on Advancement in Electronics & Communication Engineering (AECE), GHAZIABAD, India, 2024, pp. 87-91, doi: 10.1109/AECE62803.2024.10911051. keywords: {Deep learning;Adaptation models;Reviews;Unsolicited e-mail;Neural networks;Organizations;Position measurement;Motors;Security;Textiles;Neural networks;deep learning;machine learning;email spam recognition},
5. J. Vanitha, C. Mallika, A. Hema, K. Parkavi, K. K. Shandhosh Shree and S. Senthilkumar, "Detection System of Whaling Attack using Deep Learning Techniques," 2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS), Erode, India, 2024, pp. 493-498, doi: 10.1109/ICSSAS64001.2024.10760478.

keywords: {Training;Deep learning;Accuracy;Phishing;Unsolicited e-mail;Neural networks;Government;Fraud;Wire;Computer crime;Cyber Security Attacks;Whaling Attack;Deep Learning Techniques;Neural network;Three Data Sets},

1. G. N. K. Murthy, C. P. Nayana, P. Harshitha and M. Sandhya, "Email Spam Detection using Graph based Convolutional Neural Network with MALO," 2024 Second International Conference on Data Science and Information System (ICDSIS), Hassan, India, 2024, pp. 1-8, doi: 10.1109/ICDSIS61070.2024.10594655. keywords:

{Accuracy;Uncertainty;Unsolicited e-mail;Semantics;Knowledge graphs;Feature extraction;Transformers;Modified ant Lion Optimization;Graph convolutional neural network;Email Spam Detection;Semantic information;Knowledge Features},

1. T. Halder, A. Y. Srizon, N. T. Esha, S. M. Mahedy Hasan, M. F. Faruk and M. R. Hossain, "Enhancing Email Safety: Harnessing ML, DL, and LLM Models for Spam Detection," 2024 IEEE International Conference on Power, Electrical, Electronics and Industrial Applications (PEEIACON), Rajshahi, Bangladesh, 2024, pp. 1-6, doi: 10.1109/PEEIACON63629.2024.10800637. keywords: {Support vector machines;Deep learning;Training;Adaptation models;Logistic regression;Unsolicited e-mail;Large language models;Performance analysis;Convolutional neural networks;Security;Ham;Spam;Phishing;Large Language Model(LLM);Cyber Security;Artificial Intelligence;Fine Tuning;CNN;SVM;GNB;Logistic Regression},
2. N. A. Al-Dmour, S. Kousar, A. Khan, A. Ihsan, T. Abbas and A. Q. Saeed, "Enhancing Email Spam Detection Using Advanced AI Techniques," 2024 International Conference on Decision Aid Sciences and Applications (DASA), Manama, Bahrain, 2024,

pp. 1-6, doi: 10.1109/DASA63652.2024.10836525. keywords: {Support vector machines;Machine learning algorithms;Filtering;Reviews;Unsolicited e-mail;Biological system modeling;Ecosystems;Pressing;Telephone sets;Random forests;Deep learning;machine learning;SVC;naive-Bayes;Email spam detection},

1. N. Giridharan, A. Arif, P. Karmeshwaran and M. N. Ranjan, "Online Multilingual Spam Review Detection using Twin Support Vector Machine and Pre-Trained Word Embedding," 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2024, pp. 1097-1101, doi: 10.1109/ICAAIC60222.2024.10575013. keywords: {Support vector machines;Accuracy;Reviews;Metaheuristics;Optimization methods;Filtering algorithms;Artificial intelligence;Multilingual Spam Detection;Online Reviews;Review Filtering;Deceptive Reviews},
2. K. Debnath and N. Kar, "Email Spam Detection using Deep Learning Approach," 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), Faridabad, India, 2022, pp. 37-41, doi: 10.1109/COM-IT- CON54601.2022.9850588. keywords: {Deep learning;Support vector machines;Radio frequency;Unsolicited e-mail;Computational modeling;Bit error rate;Data preprocessing;Email Spam detection;Deep Learning;Machine Learning;LSTM;BERT},
3. R. Jain and N. Nihalani, "Botnet Detection in Distributed Network Using Machine Learning- A Detailed Review," 2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE), Gautam Buddha Nagar, India, 2024, pp. 888-895, doi: 10.1109/IC3SE62002.2024.10593476. keywords:

{Training;Reviews;Botnet;Unsolicited e-mail;Telecommunication traffic;Machine learning;Network security;Distributed Network;Security;Botnet;Detection;Machine Learning},

1. N. B. Chaudhuri and S. B. Chaudhuri, "Designing An Intelligent Architecture to Detect Malevolent E-Mail Using Machine Learning Approach," 2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU), Bhubaneswar, India, 2024, pp. 1-6, doi: 10.1109/IC-CGU58078.2024.10530788. keywords: {Unsolicited e-mail;Receivers;Ubiquitous computing;Electronic mail;Classification algorithms;Servers;Proposals;machine learning applications;spam detection;random forest classifier},
2. P. Jain, S. Singh and C. K. Saxena, "Detecting Email Spam with NLP: A Machine Learning Approach," 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT), Greater Noida, India, 2024, pp. 393-398, doi: 10.1109/IC2PCT60090.2024.10486769. keywords: {Training;Filtering;Unsolicited e- mail;Text categorization;Semantics;Machine learning;Vectors;Email spam detection;Machine learning;Natural Language Processing (NLP);Naive Bayes classification;Text vectorization;Text classification},
3. M. F. Rabbi, A. I. Champa and M. F. Zibran, "Phishy? Detecting Phishing Emails Using ML and NLP," 2023 IEEE/ACIS 21st International Conference on Software Engineering Research, Management and Applications (SERA), Orlando, FL, USA, 2023,

pp. 77-83, doi: 10.1109/SERA57763.2023.10197758. keywords: {Machine learning algorithms;Phishing;Machine learning;Natural language processing;Electronic mail;Cyberattack;Software engineering;phishing;spam;email;classification;detection;machine learning;natural language processing},

1. U. Bhadani, "Advanced Email Security with NLP and the Isolation Forest Algorithm," 2024 IEEE 12th International Conference on Information, Communication and Networks (ICICN), Guilin, China, 2024, pp. 497-503, doi: 10.1109/ICICN62625.2024.10761702. keywords: {Machine learning algorithms;Phishing;Scalability;Forestry;Transforms;Natural language processing;Vectors;Electronic mail;Numerical models;Security;Email Security;Spam;Phishing;DMARC;DKIM. SPF;Natural Language Processing},