Machine Learning-Based Phishing Email Detection

High-Level Design (HLD)

System Overview

- **Data Layer**: Cleaned text dataset (17,000+ emails) with lexical and metadata features
- **Processing Layer**: Text cleaning, tokenization, feature encoding, ML and DL training, evaluation
- **Application Layer**: API or Streamlit app for real-time email checking

Detailed Design (DLD)

Input Data

17,000+ email records with raw text and engineered NLP features, labeled as **1 = Phishing**, **0 = Safe.**

Preprocessing

- Remove null and duplicate emails
- Encode labels numerically
- Clean text by removing links, punctuation, and converting to lowercase
- Tokenize and pad text sequences for deep learning models

Feature Set

- **Lexical**: Word count, special character frequency, email length
- **Metadata**: Presence of suspicious terms, unusual formatting
- **Semantic**: NLP embeddings for contextual analysis

Model Pipeline

1. Load and clean dataset; train-test split (80/20)
2. Train models: Naive Bayes, Logistic Regression, Random Forest, XGBoost
3. Build and train LSTM deep learning model
4. Evaluate using accuracy, precision, recall, F1, ROC-AUC
5. Save best performing models

Flowcharts / Diagrams (Text Format)

Workflow Diagram

```
User Email → Preprocessing → Feature Extraction → ML/DL Model → Prediction → Result
```

Data Flow Diagram (DFD)

```
Level 0: User ↔ Phishing Detection System ↔ Email Dataset

Level 1:
Email Input → Text Cleaner → Feature Generator → Classifier → Output
```

Process Flow (Training Phase)

```
Start
↓
Load Dataset
↓
Clean & Preprocess Data
↓
Train Models (NB, LR, RF, LSTM)
↓
Evaluate Models
↓
Save Best Model
↓
End
```

Process Flow (Prediction Phase)

```
Start
↓
User Inputs Email Text
↓
Preprocess
↓
Load Model
↓
Predict Phishing/Safe
↓
Show Result
↓
```

End

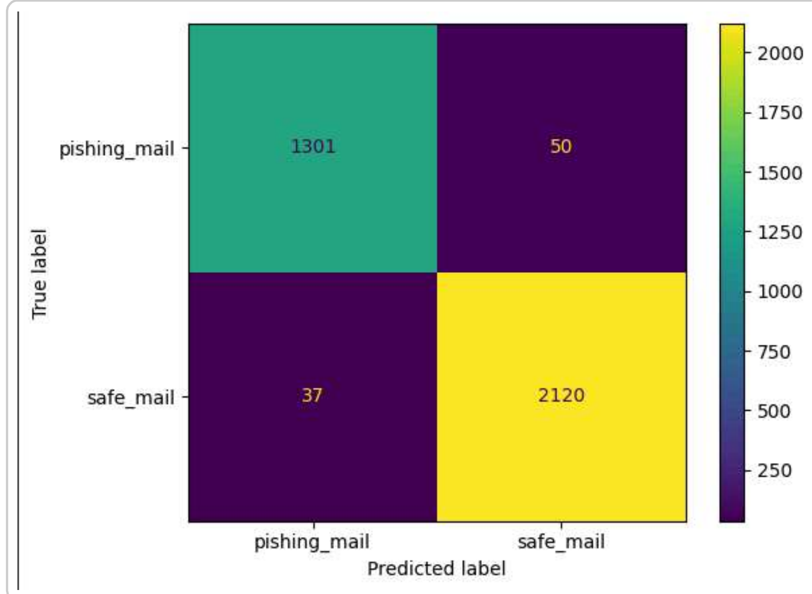
Results

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Naive Bayes	97.5%	0.98	0.98	0.98	0.99
Logistic Regression	98.2%	0.98	0.98	0.99	0.99
LSTM (Deep Learning)	96.5%	0.97	0.97	0.97	0.99

Some Visual Results per Model

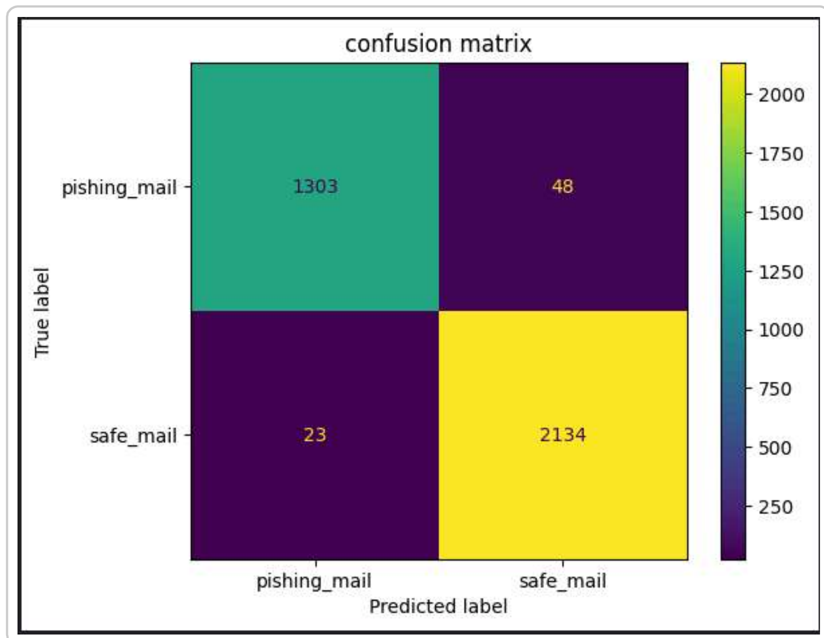
Naive Bayes

```
accuracy from native bayes: 97.52 %
f1 score from naive bayes: 97.99 %
classification report :

              precision    recall  f1-score   support

           0       0.97      0.96      0.97      1351
           1       0.98      0.98      0.98      2157

    accuracy                           0.98      3508
   macro avg       0.97      0.97      0.97      3508
weighted avg       0.98      0.98      0.98      3508
```

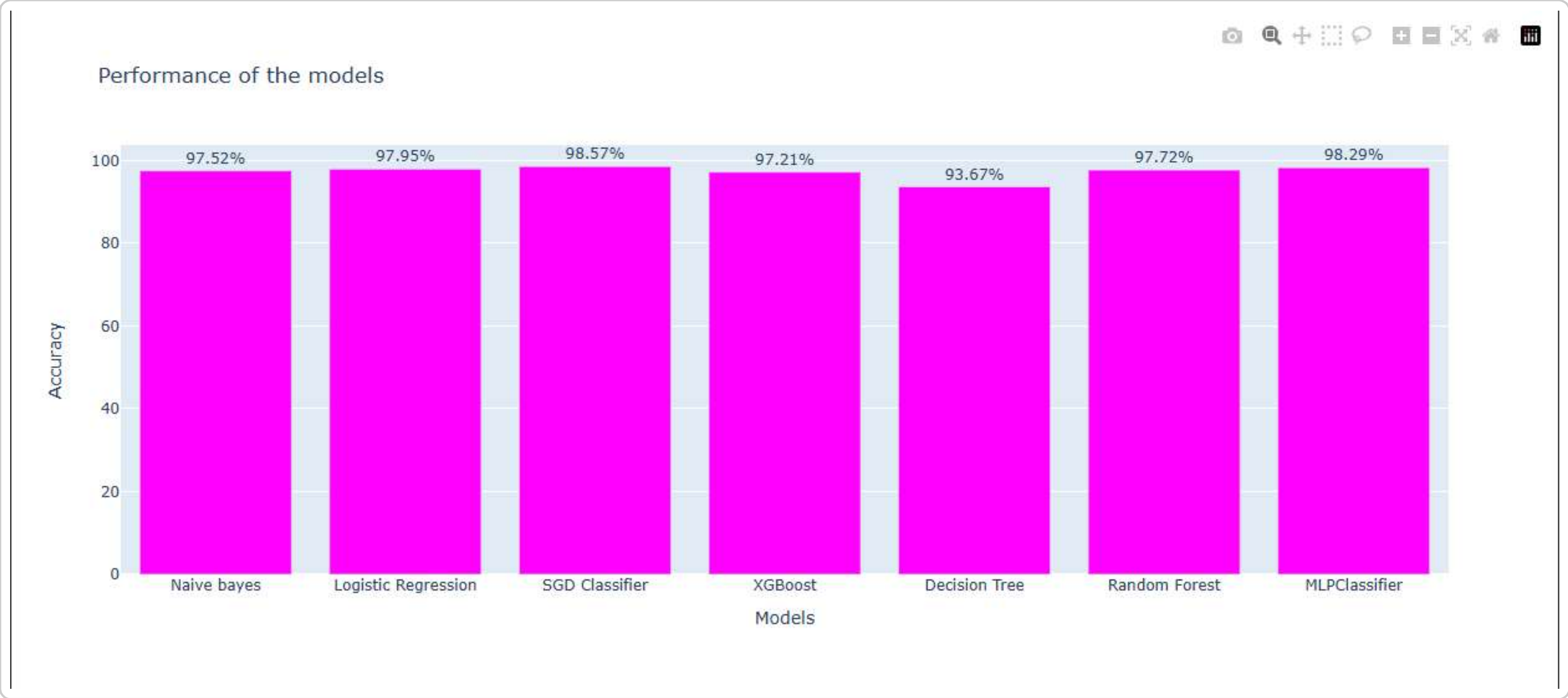


Logistic Regression

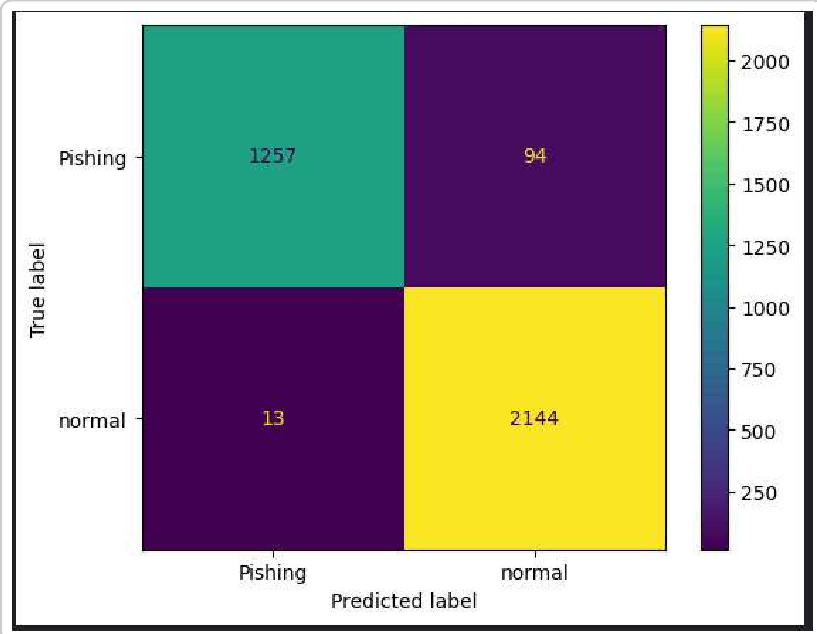
```
accuracy from logistic regression:97.98 %
f1 score from logistic regression: 98.36 %
classification report :
              precision    recall  f1-score   support

           0       0.98      0.96      0.97      1351
           1       0.98      0.99      0.98      2157

    accuracy                           0.98      3508
   macro avg       0.98      0.98      0.98      3508
weighted avg       0.98      0.98      0.98      3508
```

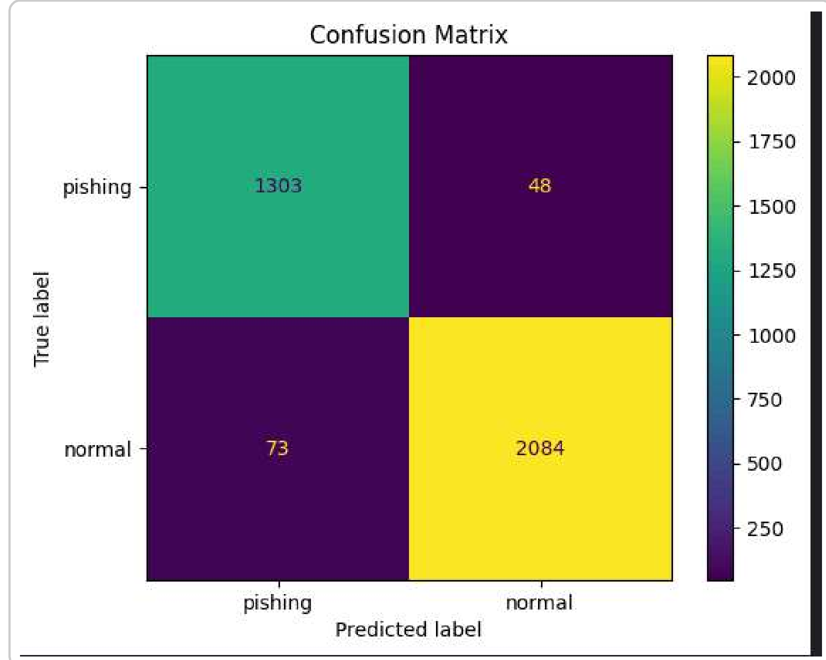


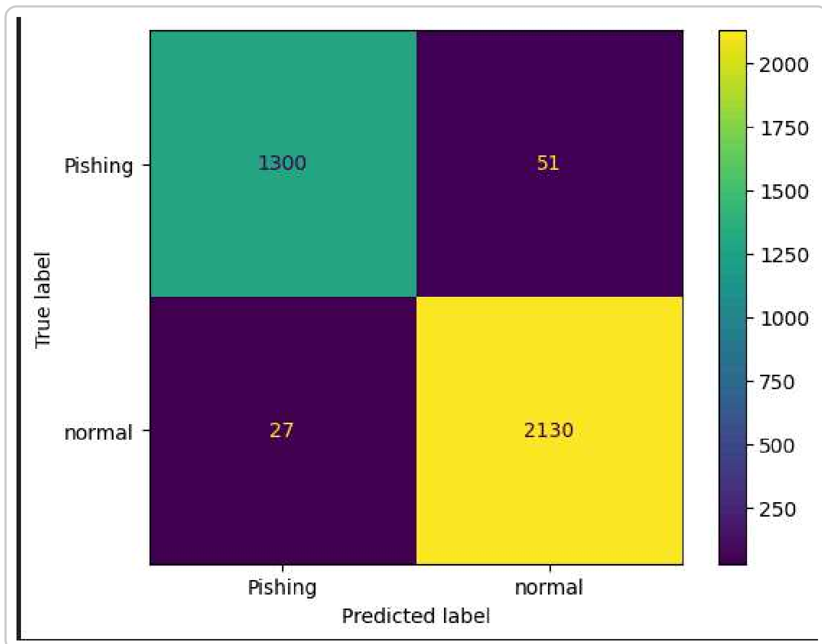
Comparison of Models



LSTM (Deep Learning)


Pishing
normal
1257
94
13
2144
Pishing
normal
Predicted label
True label
2000
1750
1500
1250
1000
750
500
250


Confusion Matrix
pishing
normal
1303
48
73
2084
pishing
normal
Predicted label
True label
2000
1750
1500
1250
1000
750
500
250


Pishing
normal
1300
51
27
2130
Pishing
normal
Predicted label
True label
2000
1750
1500
1250
1000
750
500
250

References (IEEE Format)

1. N. Abdelhamid, A. Ayesh, and F. Thabtah, "Phishing detection based associative classification data mining," *Expert Systems with Applications*, vol. 41, no. 13, pp. 5948–5959, 2014.
2. Y. Zhang, J. Hong, and L. F. Cranor, "CANTINA: A content-based approach to detecting phishing websites," in *Proc. 16th Int. Conf. World Wide Web (WWW)*, 2007, pp. 639–648.
3. D. Miyamoto, et al., "An evaluation of machine learning-based methods for detection of phishing sites," in *Proc. APWG eCrime Researchers Summit*, 2008.
4. S. Marchal, G. Armano, et al., "PhishStorm: Detecting phishing with streaming analytics," *IEEE Trans. Comput.*, vol. 65, no. 5, pp. 1352–1365, 2016.
5. Kaggle, "Phishing Email Dataset," [Online]. Available: https://www.kaggle.com/