Spam Detection Using Quantum AI

Project Report

Team Members:
Soma Shekar Reddy (22BDS022)
Aravind (22BDS015)
Stephen (22BEC050)
Mahidar (22BEC022)
Sriram Bhargav (22BEC029)

1 Introduction

Spam messages have become a major concern in communication systems such as email, SMS, WhatsApp, and social media platforms. These messages often contain fraudulent content, phishing links, advertisements, or malicious attachments. Automated spam detection plays a crucial role in ensuring secure communication.

Machine Learning methods, especially Naive Bayes classifiers, have proven effective for text classification. Recently, quantum computing has introduced new ways to develop hybrid quantum-classical machine learning models. This project compares three approaches:

- Classical Naive Bayes
- Quantum Naive Bayes
- Advanced Quantum Naive Bayes

A Spam Detector Web App was also built using Gradio, enabling users to test messages in real-time.

2 Objectives

The project aims to:

- Implement Classical, Quantum, and Advanced Quantum Naive Bayes models.
- Compare their performance using a real-world spam dataset.
- Analyze accuracy, precision, recall, and F1-score.
- Build and deploy a spam detection web application.
- Understand the limitations of quantum models for text classification.

3 Dataset Description

The SMS Spam Collection Dataset is used in this project. It contains labeled text messages categorized as “spam” or “ham”.

Dataset Statistics

- Total messages: 1115
- Ham messages: 966
- Spam messages: 149

Preprocessing Steps

- Lowercasing text
- Removing stopwords
- Vectorizing text using CountVectorizer
- Splitting into training and testing sets

4 Methodology

4.1 Preprocessing

Text preprocessing included:

- Converting labels (ham = 0, spam = 1)
- Feature extraction using CountVectorizer (1000 features)
- 80:20 train-test split

4.2 Classical Naive Bayes

Implemented using MultinomialNB, known for efficient performance on text data.

4.3 Quantum Naive Bayes

This model uses:

- A 4-qubit quantum circuit
- AmplitudeEmbedding
- Rotational layers
- CNOT entanglement
- Hybrid probability mixing (70% classical, 30% quantum)

4.4 Advanced Quantum Naive Bayes

This approach includes:

- 3-qubit quantum circuit
- RY and RZ layers with CZ entanglement
- Quantum feature appended to classical features
- GaussianNB used as final classifier

5 Experimental Results

5.1 Classical Naive Bayes

Accuracy: 0.9839

```
Classical Naive Bayes Results:
Accuracy: 0.9839
              precision    recall  f1-score   support

           0       0.99      0.99      0.99       966
           1       0.95      0.93      0.94       149

    accuracy                           0.98      1115
   macro avg       0.97      0.96      0.96      1115
weighted avg       0.98      0.98      0.98      1115
```

5.2 Quantum Naive Bayes

Accuracy: 0.8969

```
Quantum Naive Bayes Results:
Accuracy: 0.8969
              precision    recall  f1-score   support

           0       0.90      0.99      0.94       966
           1       0.89      0.26      0.40       149

    accuracy                           0.90      1115
   macro avg       0.89      0.63      0.67      1115
weighted avg       0.90      0.90      0.87      1115
```

5.3 Advanced Quantum Naive Bayes

Accuracy: 0.1500

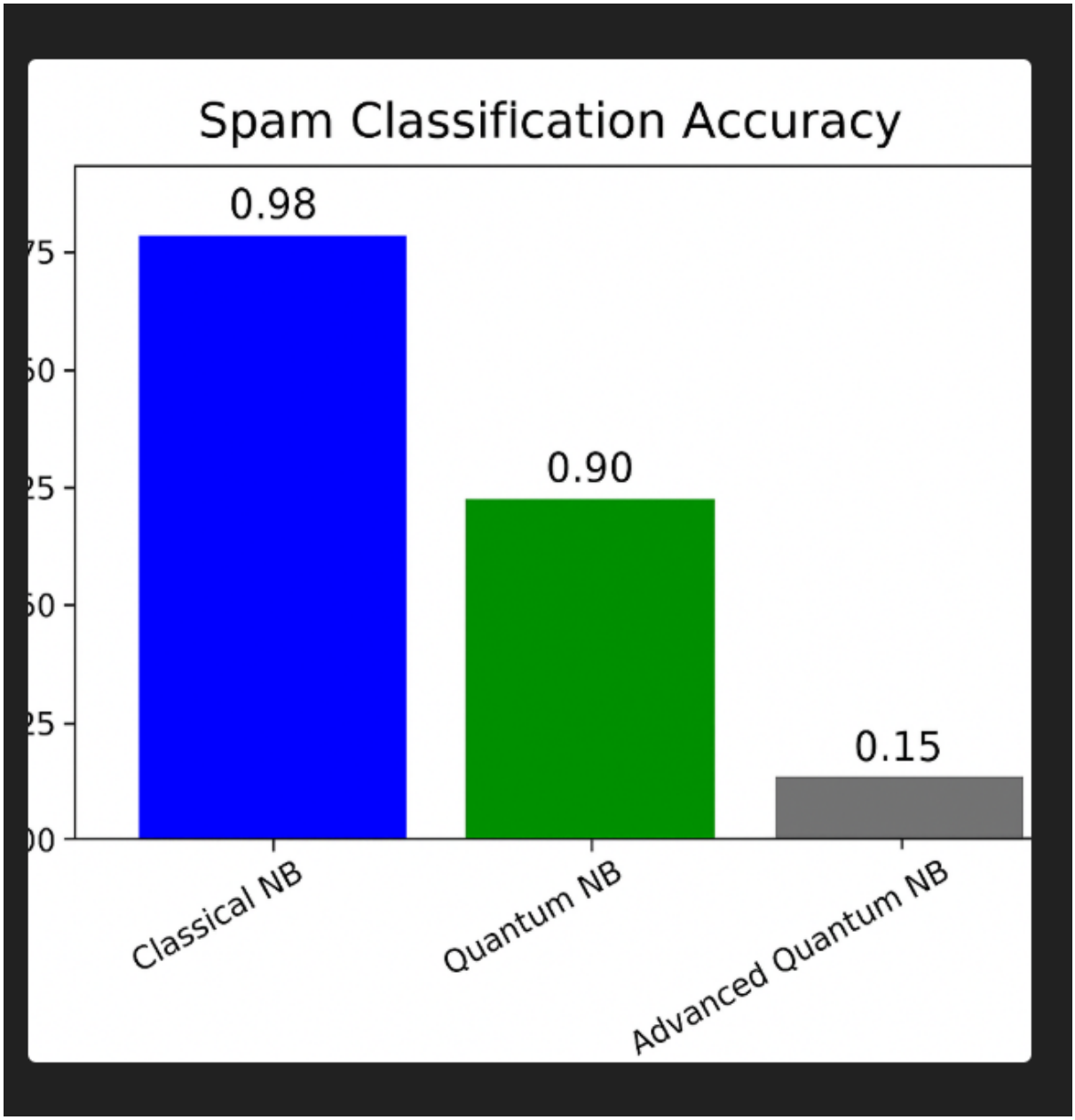
```
Advanced Quantum Naive Bayes Results:
Accuracy: 0.1500
              precision    recall  f1-score   support

           0       1.00      0.06      0.11        90
           1       0.11      1.00      0.19        10

    accuracy                           0.15       100
   macro avg       0.55      0.53      0.15       100
weighted avg       0.91      0.15      0.11       100
```

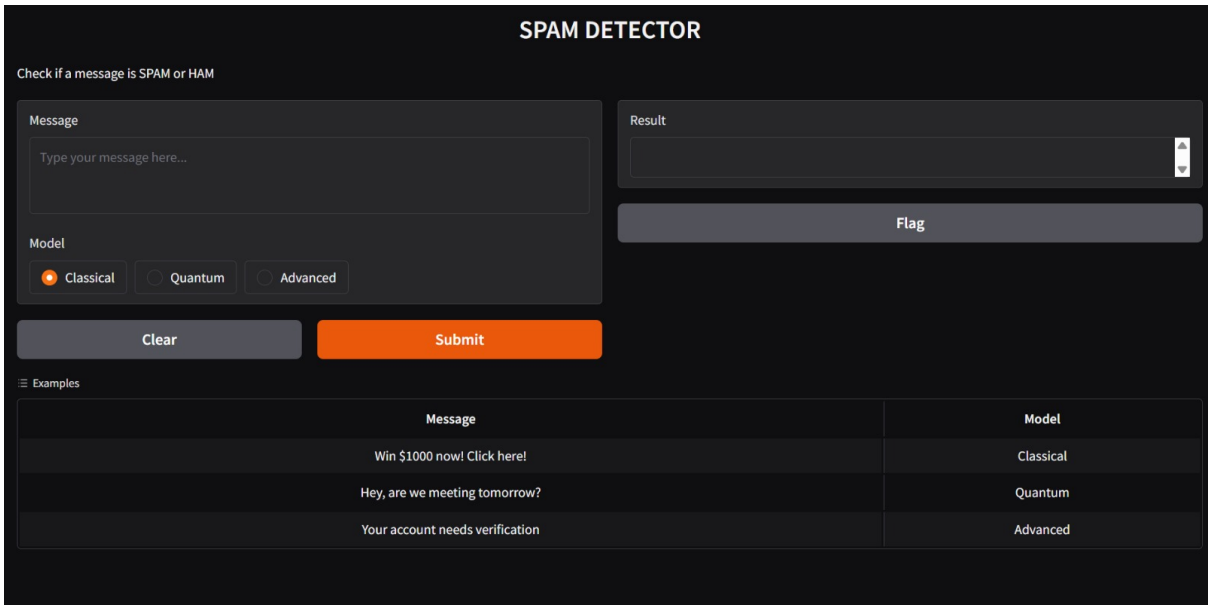
5.4 Accuracy Comparison

Model	Accuracy
Classical NB	0.9839
Quantum NB	0.8969
Advanced Quantum NB	0.1500

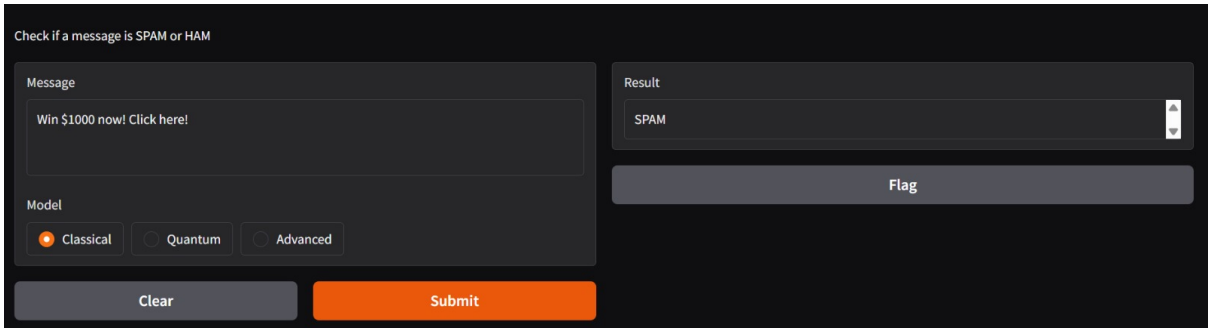

Spam Classification Accuracy
0.98
0.90
0.15
Classical NB
Quantum NB
Advanced Quantum NB

6 Spam Detector App Screenshots

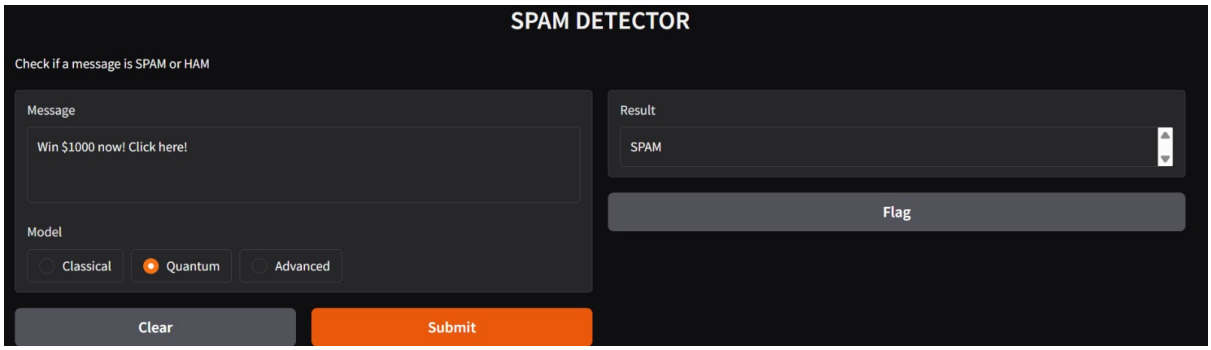
Home Screen



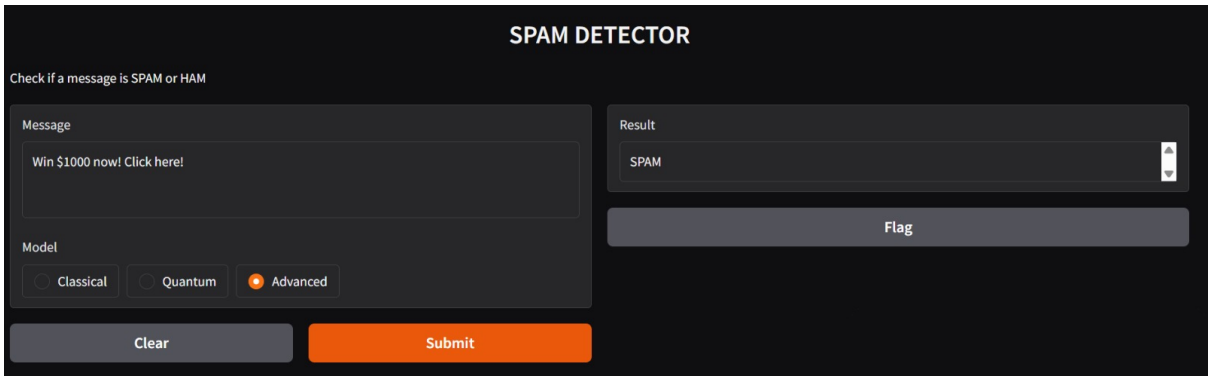
Classical NB Output



Quantum NB Output



Advanced NB Output



7 Discussion

The Classical Naive Bayes model achieved the highest accuracy (98.39%), demonstrating strong compatibility with text-based features.

Quantum Naive Bayes achieved 89.69% accuracy, but struggled to classify spam due to reduced feature dimensions and circuit limitations.

The Advanced Quantum Naive Bayes model performed poorly (15% accuracy) due to:

- Small subset of training samples
- Limited number of qubits
- Information loss in amplitude encoding

Quantum ML is promising, but currently limited for natural language processing tasks.

8 Conclusion

This project demonstrates the implementation and comparison of classical and quantum Naive Bayes models for spam detection. The following conclusions were drawn:

- Classical Naive Bayes is the most effective for text classification.
- Quantum models are experimental and need improvements in encoding and circuit design.
- The Spam Detector App successfully integrates all three models.
- Quantum ML is not yet suitable for real-world NLP tasks.

9 Future Enhancements

- Implement Quantum SVM and Quantum Neural Networks.
- Use angle embedding for better quantum feature representation.

- Train with larger datasets.
- Deploy the spam detection app online.
- Run quantum circuits on real IBM Quantum hardware.

10 References

- Qiskit Documentation: `https://qiskit.org/`
- Pennylane Documentation: `https://pennylane.ai/`
- Scikit-learn Documentation: `https://scikit-learn.org/`