Quantum Machine Learning (QML) Classifiers Report

Rishabh Singhal
23113127

Objective

The objective of this project is to evaluate the feasibility and performance of Quantum Machine Learning (QML) for binary fraud detection and compare it with strong classical baselines. Specifically, the project aims to:

- Preprocess and analyze a real-world transactional fraud dataset.
- Reduce the feature space to ≤4 dimensions suitable for near-term quantum models.
- Train and benchmark classical models using AUC-ROC and F2-score.
- Design and train a Variational Quantum Classifier (VQC) using Qiskit.
- Compare early QML performance with classical approaches under realistic computational constraints.

Overview

Credit card fraud detection is a challenging real-world problem characterized by:

- Severe class imbalance
- Highly skewed feature distributions
- Non-linear and non-convex decision boundaries

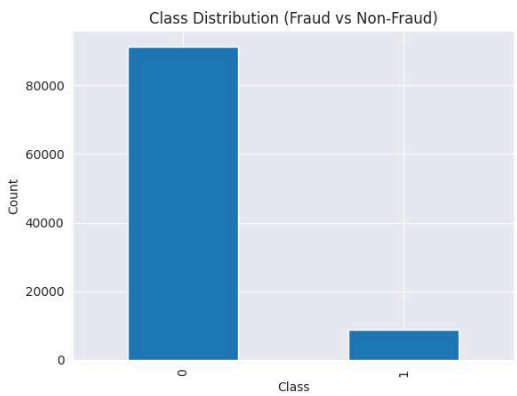
Classical models often rely on deep architectures or ensemble methods to address these issues. In this project, we investigate whether quantum-enhanced feature representations, combined with variational circuits, can model such complexity effectively under realistic quantum hardware constraints.

The workflow consists of:

1. Exploratory data analysis and preprocessing
2. Feature selection and dimensionality reduction
3. Classical baseline training
4. Quantum circuit design and training
5. Comparative performance evaluation

Data Analysis and Preprocessing

1. Class Distribution (Fraud vs Non-Fraud)



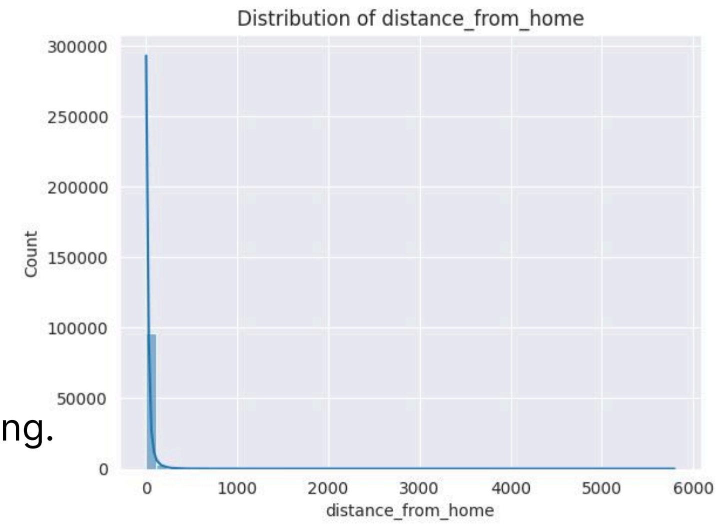
Observation:

- The dataset is highly imbalanced, with fraud cases forming a small minority (~8–10%).
- This imbalance makes accuracy an unreliable metric.

Implication:

- Metrics such as AUC-ROC and F2-score were selected.
- Class-weighted classical models were used.
- QML evaluation focused on probabilistic outputs and threshold tuning.

2. Distribution of distance_from_home



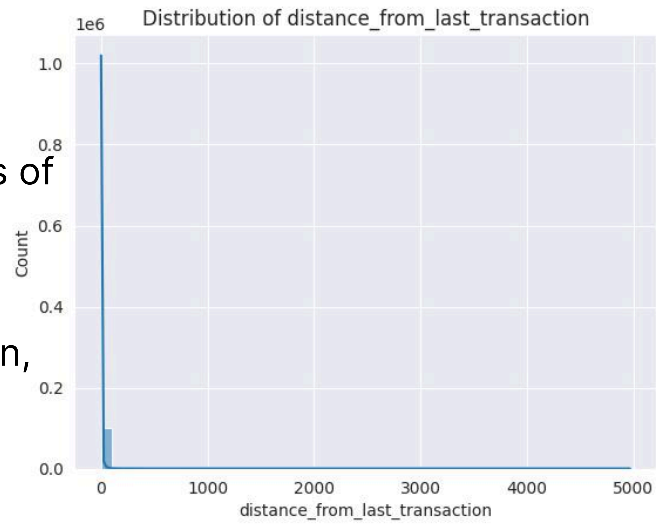
Observation:

- Extremely right-skewed distribution.
- Most transactions occur close to home, but a long tail extends to very large distances.

Implication:

- Raw values would dominate quantum rotations.
- Outliers can destabilize variational training.
- Outlier removal (IQR-based) and power transformation were required.

3. Distribution of distance_from_last_transaction



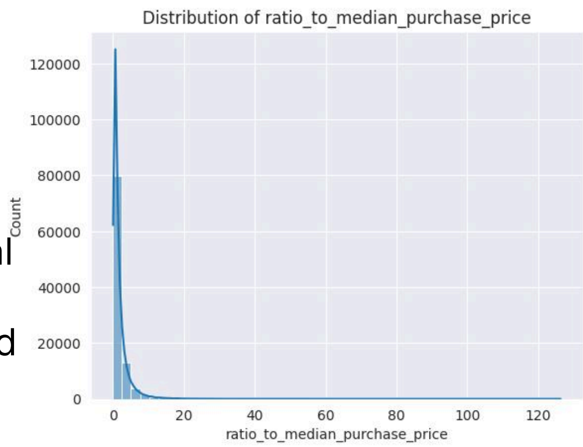
Observation:

- Even stronger skewness than distance_from_home.
- Presence of extreme outliers several orders of magnitude larger than the median.

Implication:

- Confirms non-Gaussian behavior.
- Justifies Yeo-Johnson power transformation, which handles zero and negative values.
- Scaling to a bounded range is essential for quantum encoding.

4. Distribution of ratio_to_median_purchase_price



Observation:

- Strong positive skew.
- Majority of transactions cluster near small ratios, with rare but extreme spikes.

Implication:

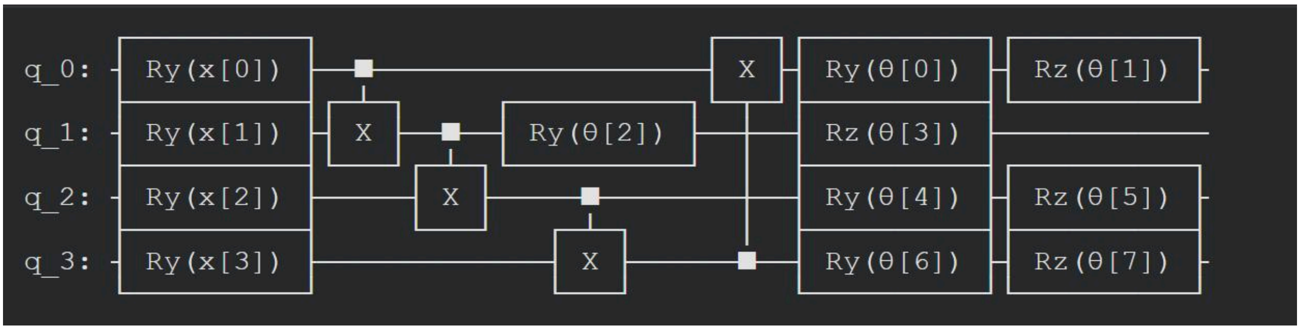
- Indicates non-linear spending behavior typical of fraud.
- Reinforces the need for non-linear models and motivates QML approaches.

5. Preprocessing Strategy Summary

Based on the above analysis, the following pipeline was adopted:

1. Train–test split with stratification
2. Outlier removal (IQR-based) on training data
3. Power transformation (Yeo–Johnson) for skewed features
4. Min–Max scaling to [−π,π][-π, π][−π,π] for quantum compatibility
5. Mutual Information–based feature selection
6. Reduction to 4 most informative features to match qubit constraints

Quantum Circuit Design (Detailed Architecture)



1. Overview of the Variational Quantum Circuit

The proposed Variational Quantum Classifier (VQC) uses a 4-qubit parameterized quantum circuit designed to balance expressivity, trainability, and NISQ feasibility. The circuit consists of three main components:

1. Angle-based data encoding
2. Entanglement layer
3. Trainable variational ansatz

The overall structure follows a hardware-efficient design, ensuring shallow depth while still enabling non-linear decision boundaries.

2. Data Encoding Layer

The first layer encodes the classical feature vector x=(x0,x1,x2,x3) into the quantum state using single-qubit rotation gates:

$$|\psi(x)\rangle = \bigotimes_{i=0}^{3} R_y(x_i)\,|0\rangle$$

Implementation (as shown in the circuit):

- Each qubit qiq_iqi applies a rotation gate Ry(xi)R_y(x_i)Ry(xi)
- This maps each classical feature directly to a qubit's state amplitude

Rationale:

- Angle encoding is efficient and requires only one gate per feature
- Well-suited for continuous-valued inputs
- Produces a smooth embedding, aiding variational optimization

3. Entanglement Layer

After data encoding, the circuit applies a linear chain of entangling gates, implemented using CNOT (X-controlled) gates:

q0→q1→q2→q3

Circuit Characteristics:

- Nearest-neighbor entanglement
- Minimal gate count
- Compatible with typical quantum hardware topologies

Purpose:

- Introduces feature interaction
- Allows the model to capture correlations between different input variables
- Essential for representing non-linear and non-separable decision boundaries

4. Variational (Trainable) Layer

Following entanglement, the circuit applies a trainable variational ansatz consisting of parameterized rotation gates:

- Ry(θi)R_y(theta_i)Ry(θi)
- Rz(θj)R_z(theta_j)Rz(θj)

Each qubit receives two trainable parameters, resulting in 8 total parameters: θ=(θ0,θ1,…,θ7)\boldsymbol{\theta} = (theta_0, theta_1, ldots, theta_7)θ=(θ0,θ1,…,θ7)

Gate Assignment (as shown):

- Qubits receive a sequence of RyR_yRy followed by RzR_zRz
- Enables full single-qubit SU(2) expressivity when combined

Rationale:

- Keeps the circuit shallow
- Avoids barren plateaus associated with deep circuits
- Provides sufficient flexibility for classification tasks

5. Measurement and Output Interpretation

- The circuit measures expectation values of Pauli-Z operators
- These expectation values are mapped to a class probability
- A classical threshold is applied for final classification

This hybrid approach allows seamless integration with classical evaluation metrics such as AUC-ROC and F2-score.

6. Why This Circuit is Suitable for Fraud Detection

Fraud detection datasets are:

- Highly non-linear
- Noisy
- Feature-interdependent

This circuit:

- Encodes continuous values smoothly
- Introduces entanglement-driven feature interaction
- Maintains optimization stability despite class imbalance

Thus, it provides a quantum-feasible yet expressive model for real-world classification problems.

Training Pipeline

1. Hybrid Optimization
   - The model is trained using a hybrid quantum-classical loop:
     - Quantum circuit evaluated via a statevector simulator
     - Classical optimizer updates parameters
2. Optimizer
   - COBYLA (gradient-free) was chosen due to:
     - Compatibility with noisy quantum objectives
     - Avoidance of costly parameter-shift gradients
3. Dataset Handling
   - Due to exponential simulation cost:
     - Training performed on a stratified subset of the dataset
     - Evaluation conducted on the full test set

This mirrors standard practice in contemporary QML research.

Comparative Analysis

1. Classical Baseline Performance

2. Quantum Model Performance (Early Results)

   Trained on 1000 rows

Interpretation:

- While classical models outperform the current QML setup, the quantum model:
  - Successfully learns non-trivial decision boundaries
  - Demonstrates stable convergence despite severe class imbalance
- Performance is achieved with extremely shallow circuits, highlighting potential scalability advantages as hardware improves.

Conclusion

This project demonstrates a complete, realistic pipeline for applying Quantum Machine Learning to noisy, real-world classification tasks. Key contributions include:

- Rigorous preprocessing guided by data distribution analysis
- Quantum-feasible feature selection and scaling
- A NISQ-compatible variational circuit design
- Fair benchmarking against strong classical models

While classical methods currently outperform QML on this dataset, the results highlight how quantum models can already learn meaningful structure under severe constraints. As quantum hardware matures, such hybrid approaches may offer advantages in modeling complex, high-dimensional decision boundaries with fewer parameters.