# Corporate Credit Rating Classification using Quantum KNN Algorithm

Rohit Patnaik
*Department of Physics, IISER Bhopal*
*Supervisor: Dr. Kuntal Roy*
*EECS Department, IISER Bhopal*

## Abstract

*Corporate credit ratings are essential indicators of a company's financial health and creditworthiness. Traditional credit classification models often face limitations in handling large datasets and complex patterns. This report explores the application of the Quantum K-nearest neighbor (KNN) algorithm for corporate credit classification.*

*By using the corporateCreditRating [1] dataset, we selected 150 companies and four key financial metrics. These metrics were normalized and encoded into quantum states using amplitude encoding. The Quantum KNN algorithm [2], implemented with the SWAP test for distance measurement, achieved an accuracy of 80% in distinguishing between investment-grade and junk-grade companies. The report also discusses future work, including the development of a multi-class classification model and the exploration of other quantum algorithms to enhance predictive accuracy.*

## 1. Introduction

In the dynamic field of corporate finance, evaluating the creditworthiness of companies is a critical task that influences investment decisions, risk management, and regulatory compliance. Traditional methods for credit classification, while foundational, often face significant challenges when dealing with large and complex datasets inherent in modern financial systems. The K-nearest neighbor (KNN) algorithm, a popular method in pattern recognition and data classification, offers a non-parametric approach that is simple yet effective in various applications. However, its classical implementation can struggle with scalability and computational efficiency as data volumes grow.

This report explores an innovative approach by applying the Quantum K-nearest neighbor (QKNN) algorithm to the domain of corporate credit ratings. By leveraging quantum computing principles, QKNN aims to enhance the computational capabilities and accuracy of credit classification models. Through the examination of the *corporateCreditRating* dataset, this project highlights the potential of quantum algorithms to address the limitations of traditional models, paving the way for more robust and efficient financial analytics.

The rest of this report is organized as follows: Section 2 reviews the fundamental concepts of finance and the classical K-Nearest Neighbors (KNN) classifier, along with an explanation of the quantum KNN sub-algorithm, the SWAP test, which uses fidelity as a measure of similarity. Section 3 presents the complete quantum framework of the KNN classification algorithm. Conclusions are given in the last section.

## 2. Background

In this section, we briefly review the fundamental concepts related to credit rating and the classical K-Nearest Neighbors (KNN) algorithm. We will then explain our main quantum sub-algorithm, the swap test, which will be used to determine the similarity between states.

### 2.1. Credit Rating

Corporate credit ratings, issued by specialist agencies, provide an assessment of the creditworthiness of a company and act as a pivotal financial indication to potential investors. These ratings help investors understand the risk associated with the company's credit investment returns. Companies strive to achieve good credit ratings to attract more investment and secure lower debt interest rates.

Most credit rating agencies use unique discrete ordinal rating scales. For instance, the rating scale of Standard & Poor's (S&P) includes 22 grades ordered from AAA (the most promising) to D (the most risky). S&P broadly classifies companies with a rating higher than BB+ as investment-grade and those below as junk-grade companies.

Credit ratings are primarily determined by analyzing financial ratios derived from balance sheets, income statements, and cash flow statements. Balance sheets provide a snapshot of assets, liabilities, and equity at a specific point in time. Income statements detail revenue, expenses, and profit over a period. Cash flow statements track cash movements from operations, investments, and financing activities, revealing liquidity. Key financial metrics, such as net profit margin and long-term debt to equity, play a crucial role in the evaluation process.

### 2.2. Classical K-Nearest Neighbors (KNN) Algorithm

The classical k-nearest neighbors (KNN) algorithm is a straightforward yet effective machine learning approach used for both classification and regression tasks. During training, the K-NN model memorizes all available labeled data points. To classify a new data point, it calculates the distance (typically Euclidean) between the new point and all points in the training set. The class of the new data point is determined by majority voting among its k nearest neighbors. The parameter k specifies the number of neighbors considered and influences the algorithm's decision-making process.

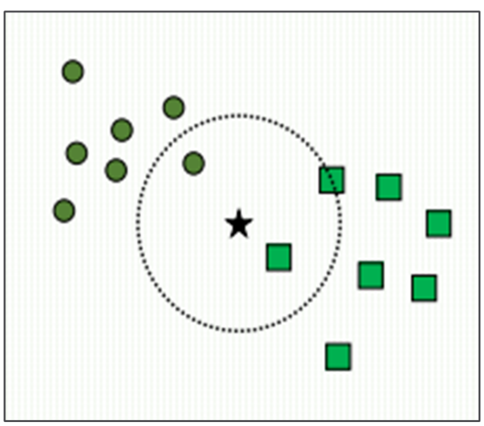

**Figure 1. Illustration of KNN algorithm.**

While K-NN is easy to understand and implement, it can be computationally intensive for large datasets due to its reliance on calculating distances to all training examples. Therefore, we will use a Quantum KNN Model, which offers significant speedup over Classical KNN by leveraging quantum principles to simultaneously evaluate distances from multiple quantum states. This enhances computational efficiency and potentially provides advantages in processing large datasets.

### 2.3. Swap Test

The SWAP test is a straightforward yet powerful quantum algorithm used to assess the similarity between two quantum states. This algorithm is particularly useful for estimating the fidelity $F(\psi,\phi) = |\langle\psi|\phi\rangle|^2$ between two arbitrary n-qubit pure states $|\psi\rangle$ and $|\phi\rangle$.

To carry out the SWAP test between states $|\psi\rangle$ and $|\phi\rangle$, we prepare three registers in the states $|0\rangle$, $|\psi\rangle$, and $|\phi\rangle$, respectively. The initial combined state of these three registers is $|0\rangle|\psi\rangle|\phi\rangle$. The CSWAP gate acts on these registers according to the following rules:

$$\mathrm{CSWAP}|0\rangle|\psi\rangle|\phi\rangle = |0\rangle|\psi\rangle|\phi\rangle$$

$$\mathrm{CSWAP}|1\rangle|\psi\rangle|\phi\rangle = |1\rangle|\phi\rangle|\psi\rangle$$

The implementation of the SWAP test involves three main steps. First, a Hadamard gate is applied to an ancillary qubit. Next, controlled-SWAP (CSWAP) gates are iteratively applied to the corresponding training and test features, with the ancillary qubit acting as the control. Finally, another Hadamard gate is applied to the ancillary qubit to complete the circuit.

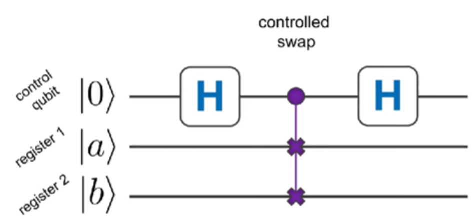


**Figure 2. Illustration of Swap Test circuit.**

Upon measurement of the circuit, the probabilities of the first register are given by:

$$\Pr(0) = \frac{1}{2} + \frac{1}{2}|\langle\psi|\phi\rangle|^2$$

$$\Pr(1) = \frac{1}{2} - \frac{1}{2}|\langle\psi|\phi\rangle|^2$$

The difference between these probabilities, $\Pr(0) - \Pr(1)$, provides the desired fidelity, indicating the similarity between the two quantum states. This inner product serves as a crucial metric in determining how closely the states $|\psi\rangle$ and $|\phi\rangle$ resemble each other.

## 3. Complete Quantum Framework for KNN Classifier

In this section, we present a detailed quantum framework for the implementation of the K-Nearest Neighbors (KNN) classifier, outlined step by step.

### 3.1. Quantum State Encoding

First, we encode the test sample features into a quantum state using Amplitude Encoding. Amplitude Encoding is a technique in quantum computing used to represent classical data within a quantum state. In this method, the features of a data sample are encoded into the amplitudes of a quantum state. Each feature value corresponds to the amplitude of a specific basis state, hence for $n$ features, we use $n$ qubits to encode the feature values.

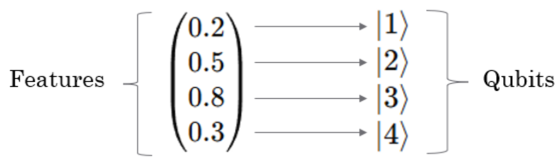


**Figure 3. Amplitude Encoding.**

The combined feature vector encoding is obtained by taking the Kronecker product (tensor product) of the individual qubit states. We sequentially combine the encoded states of the features until we obtain the final state. In a similar manner, we encode the entire training set along with the indices of each vector into a single superposition state.

We then verify the validity of these state vectors and initialize these quantum states into quantum registers using Qiskit's '`initialize`' method. This process ensures that the quantum states accurately represent the features and indices of the training set, ready for further quantum processing and analysis.

### 3.2. Building the Quantum Circuit

Now, we create an additional qubit to store the similarity information and integrate the swap test circuit into our main quantum circuit. The encoded training state qubits and encoded test state qubits are designated as state $a$ and state $b$, respectively, for similarity measurement. The newly created qubit is used as the ancillary qubit to store the similarity information.

By appending the swap test circuit, we can assess the similarity between the encoded training and test states. This setup enables the measurement and storage of similarity data within the quantum circuit, which is essential for the subsequent classification process.

### 3.3. State Measurement

Upon completing the quantum circuit, we introduce a classical measurement register to complete the process. The qubits representing the index of the training vectors and the qubit representing the similarity are measured. During this measurement, the index qubits, which were initially in a superposition state encompassing all training vectors, will collapse to reveal a single training vector index and its associated features.

Simultaneously, we measure the similarity between the features of this specific training vector and those of the test vector using the swap test. The similarity result is obtained from the measurement of the similarity register.

We perform this measurement procedure multiple times to evaluate the similarity between the test state and all the training states. To ensure accuracy, we repeat this measurement procedure several times to obtain precise similarity scores for each training vector.

### 3.4. Classical Post-Processing

From the measurement count results, we first calculate the probability of the similarity register (ancillary qubit) being in state $|1\rangle$. Using this probability, we determine the similarity or inner product with the formula:

$$\text{Similarity} = 1 - 2 \cdot P(1)$$

Once we have the similarity scores for all training vectors, we sort these indices in descending order based on their similarity scores. We then select the top $k$ indices corresponding to the highest similarity scores.

Next, we perform a majority vote among the $k$ selected training data points to determine the most common label. This majority label is then assigned to the test data point, completing the classification process.

### 3.5. Model Evaluation

We repeat this procedure by running the circuit for each test data point individually. By encoding and processing each test data point one by one, we determine the $k$ nearest neighbors for each test data point and assign labels accordingly.

Finally, we evaluate the model's performance using the accuracy score, defined as:

$$\text{Accuracy} = \frac{\text{Number of correctly classified test data points}}{\text{Total number of test points}}$$

## 4. Conclusions

In this project, we employed the corporateCreditRating [1] dataset, selecting 150 companies and four key features to train our quantum K-Nearest Neighbors (KNN) model utilizing the swap test. The model achieved an accuracy of 80%, demonstrating the promising potential of quantum techniques in credit risk assessment.

For future work, we propose developing a multi-class classification model to provide precise credit ratings for companies using quantum machine learning methods. This advancement aims to enhance the accuracy and detail of credit assessments by leveraging advanced quantum techniques. Additionally, another approach could involve using an algorithm based on Hamming distance [3] to develop a QKNN classifier. This method would offer a different perspective and potentially improve the classification performance by leveraging the distinct properties of Hamming distance in the quantum domain.

## References


[1] https://ssrn.com/abstract=4163283
or https://www.kaggle.com/datasets/kirtandelwadia/corporate-credit-rating-with-financial-ratios

[2] Basheer, A., Afham, A., & Goyal, S. K. (2020). *Quantum k-nearest neighbors algorithm*. arXiv preprint arXiv:2003.09187.

[3] Li, J., Lin, S., Yu, K. et al. *Quantum K-nearest neighbor classification algorithm based on Hamming distance*. Quantum Inf Process 21, 18 (2022). https://doi.org/10.1007/s11128-021-03361-0