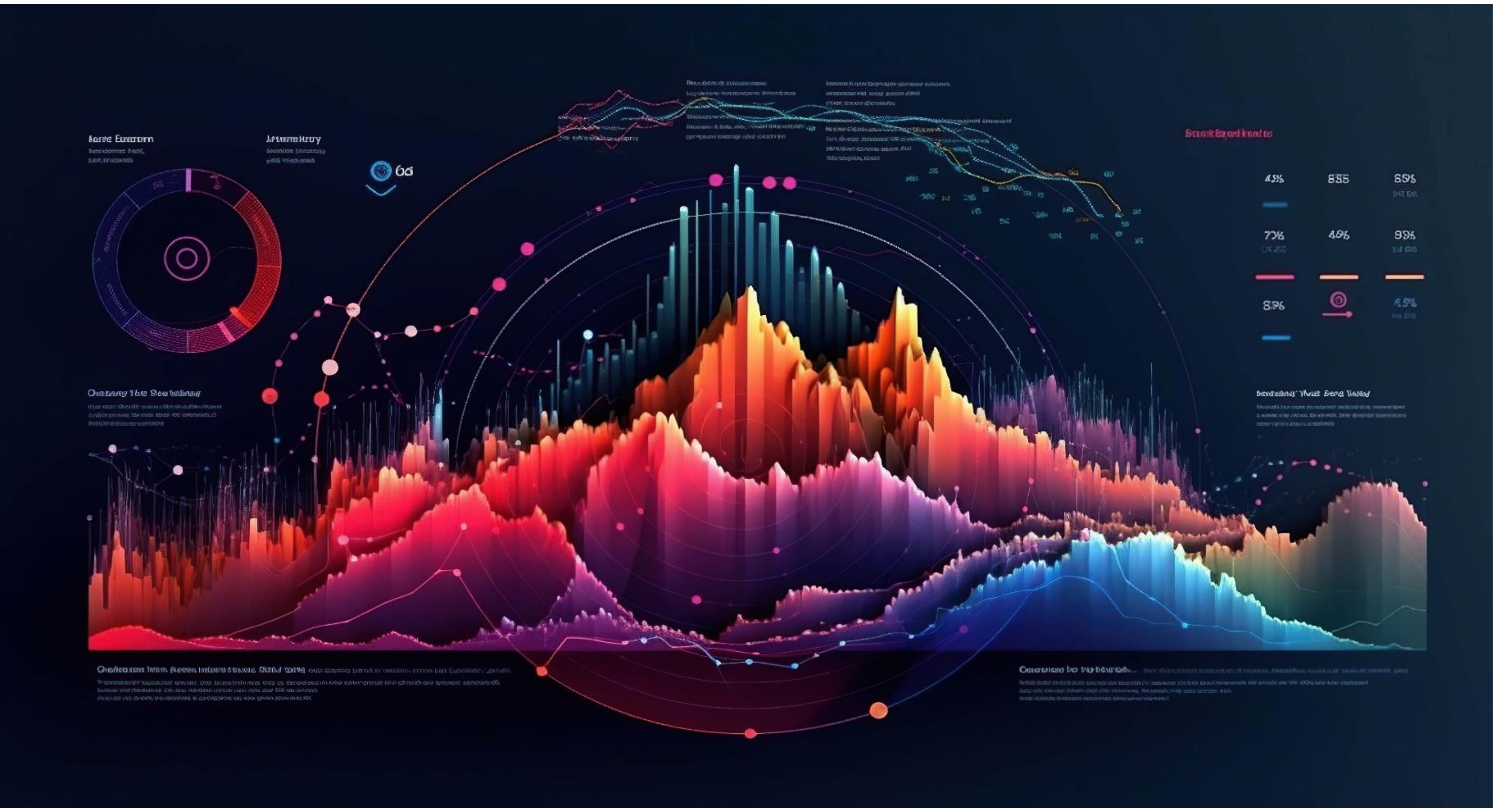
fermac risk

Credit Scoring, Artificial Intelligence, & Quantum Machine Learning

Fermac Risk White Paper

April 2024



Foreword

Our Fermac Risk Credit Scoring, Artificial Intelligence, and Quantum Machine Learning course can equip participants with cutting-edge skills to leverage quantum computing advancements and AI techniques in credit risk analysis. This knowledge enables more precise and efficient credit scoring models, leading to better risk assessment, optimized lending decisions, and enhanced financial inclusion.

In this white paper, we will highlight the practical advantages that you can acquire through specialized training. Specifically, we will discuss the importance of the ROC curve in credit scoring models, the advantages of Deep Learning and Machine Learning models over logistic regression, and the benefits of quantum machine learning models over classical machine learning models. Lastly, we will explore the economic impact of these models on banks.

---

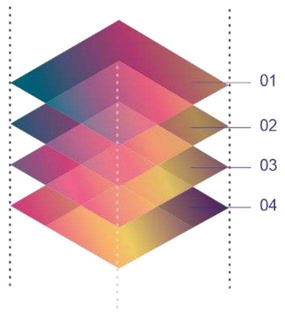
1. What is the importance of Receiver Operating Characteristic (ROC) in credit scoring?

In the context of scorecard development, the Receiver Operating Characteristic (ROC) curve is a crucial tool for evaluating the performance of a classification model. The area under the ROC curve (AUC) measures how well the scorecard distinguishes between the two classes (e.g., good and bad credit risks). Obtaining a higher ROC (or specifically, a higher AUC) offers several benefits:

1. **Better Risk Management**
   By accurately identifying the likelihood of default, lenders can tailor their credit products and interest rates to individual risk profiles, improving overall risk management. This can reduce default rates and potentially increase profitability.
2. **Increased Efficiency**
   A more accurate scorecard means fewer false positives and negatives this efficiency can improve customer satisfaction, reduce unnecessary credit losses, and optimize the allocation of resources like further credit checks or collections efforts.
3. **Competitive Advantage**
   In a competitive market, the ability to assess price risk more accurately can be a significant advantage. It can lead to better customer segmentation, more competitive pricing, and the development of innovative financial products tailored to different risk segments.
4. **Enhanced Strategic Decisions**
   With a reliable scorecard, financial institutions can make strategic decisions about market segments to target, products to offer, and credit policies to implement. This strategic advantage can lead to better portfolio diversification and risk-adjusted returns.

---

2. What are the advantages of using deep learning in contrast to logistic regression to development credit scoring models?



In the development of credit scoring models, transitioning from traditional methods like logistic regression to deep learning approaches can offer significant advantages, particularly as the complexity and volume of data increase. Here are some key benefits of using deep learning over logistic regression for credit scoring:

1. **Handling Non-linear Relationships**
   Logistic Regression is fundamentally linear, assuming a linear relationship between the independent variables and the dependent variable's log odds. However, it can struggle with complex, non-linear relationships inherent in real-world data. Deep Learning excels at identifying and modeling complex, non-linear interactions between variables

without needing explicit feature engineering. This can lead to more accurate predictions in complex scenarios.

2. **Feature Extraction and Selection**
   Logistic Regression requires careful feature selection and engineering to model relationships effectively. This process is often manual, based on domain knowledge and statistical tests.
   Deep Learning automatically detects relevant features and interactions during the training process. Layers within a neural network can learn to represent simple and complex data patterns, reducing the need for manual feature engineering.
3. **Handling Unstructured Data**
   Logistic Regression is primarily suited for structured data. Analyzing unstructured data such as text or images requires significant preprocessing to transform the data into a suitable format.
4. **Customization and Adaptability**
   Logistic Regression models are generally static, meaning they don't adapt well to changes in the underlying data distribution without manual reevaluation and adjustment. Deep Learning models can be continuously trained and adapted to new data, making them more resilient to market conditions or customer behavior changes over time.

While deep learning offers significant advantages in modeling complex patterns and relationships in data, it's important to balance these benefits with considerations like model interpretability, computational resources, and the availability of sufficient training data. Logistic regression, with its simplicity, transparency, and ease of implementation, remains valuable, especially in scenarios where explainability is crucial or data is limited.

3. What are the advantages of Quantum Machine Learning (QML) over traditional deep learning techniques for developing credit scoring models?

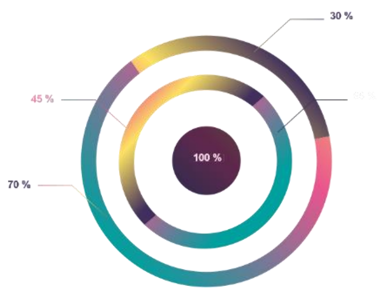
The use of quantum machine learning (QML) over traditional deep learning techniques, including feedforward neural networks and convolutional neural networks (CNNs), in credit scoring is an emerging area of interest, primarily due to the theoretical advantages quantum computing offers. While the practical application of quantum ML in credit scoring is still in its infancy, the theoretical benefits suggest a promising future. Here are some of the key benefits that quantum ML could offer over traditional deep learning models in the development of credit scoring systems:

1. **Speed and Efficiency**
   Due to quantum parallelism, quantum ML can theoretically process information much faster than classical computing methods. Quantum algorithms can significantly reduce computational complexity for some problems, offering exponential speedups for certain tasks. This could make the analysis of large datasets more efficient, potentially leading to quicker credit scoring processes.
2. **Complexity and Dimensionality**
   Quantum systems can naturally handle complex, high-dimensional data spaces due to the nature of quantum bits (qubits) that can represent multiple states simultaneously through superposition. This capability could allow for more nuanced and sophisticated modeling of credit risk, capturing complex patterns and relationships in the data that might be challenging for traditional models.
3. **Optimization**
   Quantum computing can enhance optimization processes, a key component in training machine learning models. Quantum optimization algorithms might find the global minimum of a loss function more efficiently than classical optimization techniques, which can sometimes get trapped in local minima, especially in complex, high-dimensional spaces typical of deep learning.
4. **Feature Selection and Data Compression**
   Quantum ML could potentially offer innovative ways to perform feature selection and data compression without losing critical information. Quantum techniques like quantum principal component analysis (qPCA) can process and reduce the dimensionality of large datasets more efficiently, helping in identifying the most informative features for credit scoring.
5. **Enhanced Pattern Recognition**
   Quantum ML models, particularly those that leverage quantum entanglement and interference, might be capable of recognizing patterns and correlations in data that are not apparent to classical algorithms. This could be particularly useful in identifying subtle signals of creditworthiness or default risk that traditional models might overlook.
6. **Handling Uncertainty**

Quantum computing naturally incorporates the concept of uncertainty through the probabilistic nature of quantum mechanics. This could be particularly advantageous in credit scoring, where uncertainty and probabilistic outcomes (such as the likelihood of default) are inherent to the models.

While the potential benefits of quantum ML in credit scoring are significant, it's important to note that quantum computing and quantum ML are still emerging fields. Many of the theoretical advantages have yet to be fully realized in practical applications due to current technological limitations, including the availability of sufficiently powerful and stable quantum computers. Additionally, issues related to model interpretability, data privacy, and the complexity of quantum algorithms also pose challenges. However, as quantum computing technology matures, it could offer substantial advancements in credit scoring models' accuracy, efficiency, and capabilities over traditional deep learning approaches.

4. What has been Fermac Risk's experience in developing credit scoring using deep learning and quantum computing?



At Fermac Risk, we started developing credit scoring models using traditional approaches like logistic regression, modern machine learning, and deep learning techniques.

In 2017, we worked with a high-quality database to develop credit scoring models. We used logistic regression with the normalized variables and obtained an ROC of 85%. With the variables discretized by WOEs, we had a ROC of 91%, which was a good result. However, we were eager to use machine learning techniques.

In the same year, we started using the R programming language and its powerful machine-learning libraries, such as the award-winning e1071 and caret. We applied SVM models, K-nearest neighbors, bagging, boosting, and random forest. The Random Forest model returned a ROC of 98%, which was seven percentage points higher than logistic regression. We were very happy with this result.

In 2018, regulators were concerned about measuring model risk due to economic losses caused by some models. To address this issue, we employed powerful validation techniques, such as k-fold cross-validation, bootstrapping, jackknifing, kappa, confusion matrices, and discriminant power statistics (Gini, ROC, KS, Kullback-Leibler, etc.). We measured model risk by removing the best variable and measuring the new ROC, which resulted in a predictable fall of the ROC. The logistic regression's ROC dropped significantly from 91% to 77%. This reduction in discriminant power later affected the accuracy of the PD.

In 2017, we built feed-forward deep learning models. The ROC was 80%, which was only 3 percentage points higher than logistic regression. The neural network architecture had 4 hidden layers and just over 1000 neurons. Despite the significant computational effort, it was not enough to improve the ROC.

So, in 2018, we worked hard to optimize the hyperparameters, including the activation functions, the number of hidden layers, the number of neurons, the learning rate, etc. The best neural network model had a ROC of 86%, improving by 9 percentage points compared to the logistic regression.

In 2019, we used a convolutional neural network to improve our model's ROC. We excluded the best variable and obtained an impressive ROC of 95%.

In 2020, we implemented SMOTE techniques for imbalanced data. We also used advanced dropout and early-stopping techniques. Additionally, at the request of our clients, we incorporated explainable artificial intelligence techniques, such as the shap library, to ensure that our models were interpretable.

In 2021, we began using probabilistic machine learning techniques such as Bayesian neural networks to measure the risk of the parameters in our models. We found that the convolved Bayesian neural network only slightly improved upon the normal convolutional neural network.

We started exploring the field of quantum computing and quantum mechanics in 2021. In 2022, we experimented with various Python libraries and utilized quantum machine learning techniques. Our team developed our first quantum model, the quantum support vector machine, which proved to be an improvement over the traditional SVC. Furthermore, we designed a hybrid model, a quantum convolutional neural network, which achieved a ROC of 97% on a sample excluding the best variable. This was by far our best model.

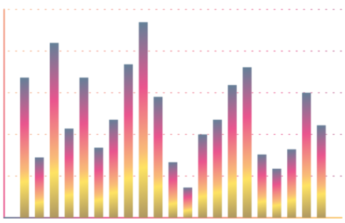
In 2023, we explored traditional probabilistic machine learning models such as Bayesian GAN neural networks for creating synthetic data. We are also testing variational classifiers, tensor networks, and quantum Bayesian neural networks in the quantum domain. We have used tensor networks in neural networks to reduce the number of parameters from 1,000,000 to just 1,000. This not only generates greater calculation speed but also reduces model risk.

Our models have been tested on both traditional computers and quantum computers in the cloud, though we have a limited number of qubits. However, we have worked to improve the speed of traditional computers.

The next quantum revolution is upon us, and banks must be prepared to manage the risks and benefits of the new quantum computers.

In 2024, we are working with Gen AI because can generate synthetic data that mimics real-world data distributions. This can augment the training dataset, improving the model's performance by providing a richer dataset that covers more potential scenarios. Gen AI can balance the dataset by generating synthetic samples of the underrepresented class, leading to more accurate and fair predictions. Gen AI can create realistic but synthetic datasets that do not map directly to real individuals, allowing for model development and testing in a way that preserves privacy and complies with regulations like GDPR in UE.

5. What is the economic impact of increasing the discriminative power in a bank's credit scoring model?



Suppose a bank, XYZ, uses a traditional model, such as logistic regression, to develop a credit scoring model for the retail portfolio. This credit scoring supports the Credit Decisioning step of a Credit Origination process. The bank needs to utilize the loss curve to determine the Acceptance Level, which is a numerical threshold used as the cut-off score. This curve is computed using a specific formula.

$Loss_i = (fp_i*LossC+fn_i*LossB)/(tn_i+fp_i+fn_i+tp_i)$

Where:

$Loss_i$ = loss in the i threshold,
fp_i=number of false positives in threshold i,
fn_i=number false negative in threshold i,
tn_i =number of true negatives in threshold i,
tp_i =true positives in threshold i.
LossC = Loss for accepting a false positive.
LossB = Loss for rejecting a false negative.

To estimate this, we need to estimate the ROC because it computes confusion matrices for various threshold values. A confusion matrix contains the number of instances of true positive (tp), false negative (fn), false positive (fp), and true negative (tn).

The bank has determined that the variable *LossC* represents the retail portfolio's average Loss Given Default LGD. This is because it represents the cost of accepting a false positive, which refers to a transaction that has been accepted by credit scoring but later defaulted. LGD is depicted as a percentage of total exposure at the time of default

LossB represents the opportunity cost, the potential forgone profit from a missed opportunity, of rejecting a transaction that was thought to be a default but turned out to be a good one.

The bank decided to use other machine learning and deep learning models to increase its discriminative power. Let's examine the results.

The value of *LossC* is 40%, which means the LGD=40%, and the *LossB* is 20%, the opportunity cost.

To begin with, the first step is to estimate the ROC (receiver operating characteristic) of each model.

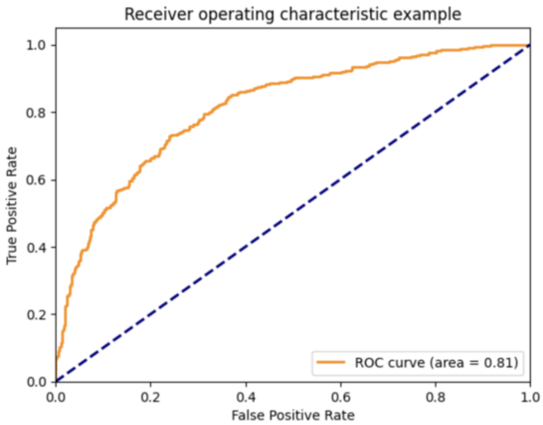


Exhibit 1: ROC Graph

The second step is to estimate *fn*, *fp*, *tp*, *tn*, and the loss curve, as shown in Table 1 below.

	Threshold	TN	FP	FN	TP	Specificity	Sensitivity	Client Loss
0	0.000000	0	511	0	489	0.000000	1.000000	0.2044
1	0.052632	0	511	0	489	0.000000	1.000000	0.2044
2	0.105263	9	502	0	489	0.017613	1.000000	0.2008
3	0.157895	105	406	12	477	0.205479	0.975460	0.1648
4	0.210526	180	331	32	457	0.352250	0.934560	0.1388
5	0.263158	234	277	47	442	0.457926	0.903885	0.1202
6	0.315789	278	233	56	433	0.544031	0.885481	0.1044
7	0.368421	319	192	73	416	0.624266	0.850716	0.0914
8	0.421053	340	171	93	396	0.665362	0.809816	0.0870
9	0.473684	367	144	121	368	0.718200	0.752556	0.0818
10	0.526316	394	117	145	344	0.771037	0.703476	0.0758
11	0.578947	421	90	181	308	0.823875	0.629857	0.0722
12	0.631579	439	72	209	280	0.859100	0.572597	0.0706
13	0.684211	457	54	237	252	0.894325	0.515337	0.0690
14	0.736842	473	38	276	213	0.925636	0.435583	0.0704
15	0.789474	491	20	326	163	0.960861	0.333333	0.0732
16	0.842105	501	10	382	107	0.980431	0.218814	0.0804
17	0.894737	511	0	489	0	1.000000	0.000000	0.0978
18	0.947368	511	0	489	0	1.000000	0.000000	0.0978
19	1.000000	511	0	489	0	1.000000	0.000000	0.0978

Table 1: Client loss for each threshold

The second graph (exhibit 2) shows that each curve represents a model with its respective ROC. The cut-off selection is where the loss curve has the minimum value. The cut-off is located at the score threshold of approximately 70%.

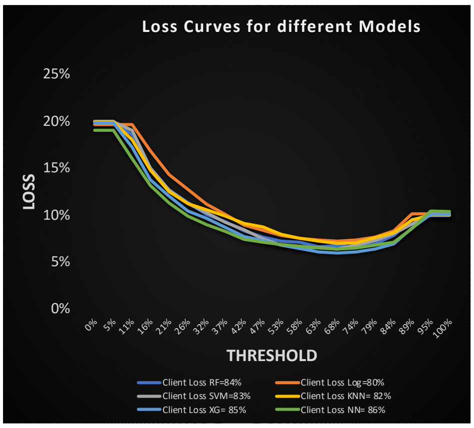


Exhibit 2: Loss Curves for different models and ROC

Let's compare the loss curves of the logistic regression model with those of the Machine Learning ML models in the cut-off threshold. The ROC for the logistic regression model is 80%; meanwhile, the ML models have greater values of ROC. The percentage differences for the K nearest neighbors, Support Vector Machine, Random Forest, Xgboosting, and Neural Network compared to the logistic regression model are 3.5%, 7.4%, 8.0%, 14.6%, and 15.4%, respectively, as we showed in Exhibit 3. This indicates that the Machine Learning models have reduced the loss by a significant margin, ranging from 3.5% to 15.4%.

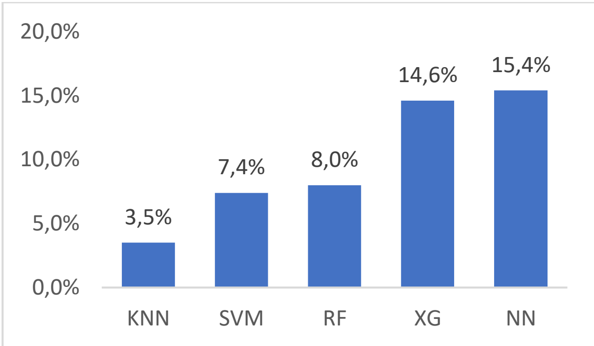


Exhibit 3: Comparative of Loss Reduction

The bank XYZ has strong incentives to implement machine learning (ML) models based on its parameters. Doing so could potentially reduce losses, improve risk management, and deliver tangible benefits. However, it is important to note that selecting a model based solely on its receiver operating characteristic (ROC) result is insufficient. The model's explainability and transparency, along with advanced validation tests, are also crucial factors in determining the parameters' uncertainty and the model's stability over time.