Credit Card Fraud Detection

Real-world application of advanced Machine Learning Classification Techniques and Quantum Simulation

Michael Tarino

Background

Classical Classification ML Techniques:

- Feature Selection with Cross Validation [AND ___]
- Basic: Trees/RF/Boosting, QDA, Logistic Regression, SVM, Naive Bayes
- Advanced: Resampling Imbalanced Data, Neural Nets

Quantum-Computing Enhanced ML:

- ??
- ??

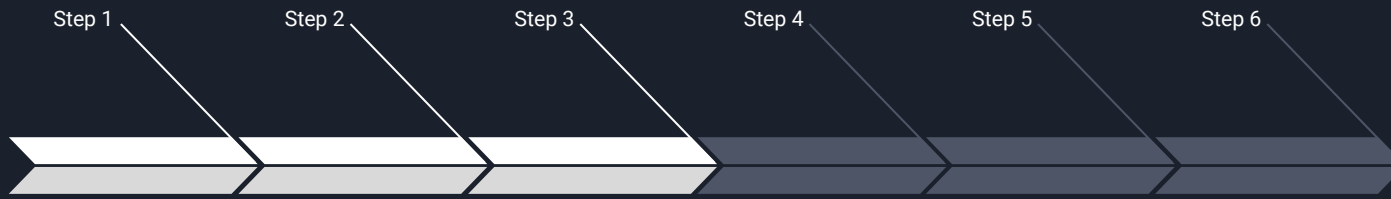
Project Objective

To use an ensemble of machine learning techniques and methods to solve a real-world problem, credit card fraud detection. A key component of this project was the exploration of QC-enhanced ML.

Target Audience

Prospective employers, fellow classmates, & NYC Data Science Academy educators

Project Timeline



Research

Extensive reading into various Quantum software platforms (Xanadu's SF, Rigetti Forests, IBM QISKIT)

Preprocessing

EDA

Training Data

Training along....

Feature Selection*

Resampling

This was an iterative process

Classification

Trees/RF/Boosting, QDA, Logistic Regression, SVM, Naive Bayes

Neural Nets

QML

Strawberry Fields / Xanadu.ai

Research

Credit Card Data & Quantum Computing

- Kaggle competition
-
-
-

State of Quantum Technology

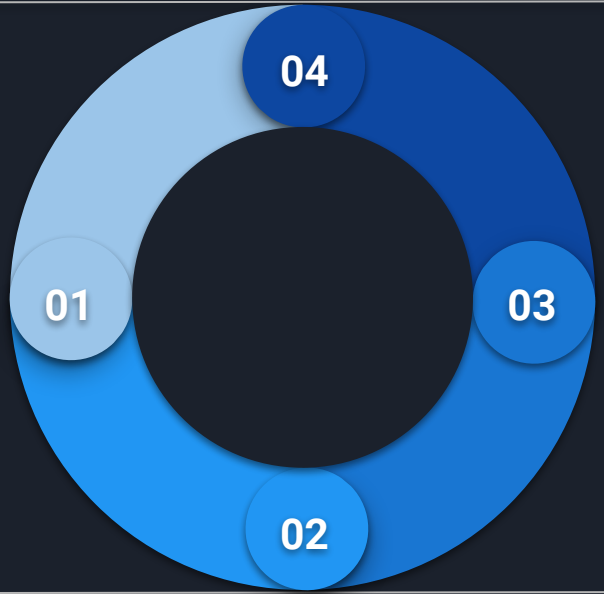
01 Hardware - Adiabatic (annealing), Gate Model (circuit model)

02 Software - quantum cloud platforms; optimization, machine learning & sampling problems

Preprocessing & Exploratory Data Analysis

Data tidying

Data Manipulation



Analysis

Data Visualization

Preprocessing & Exploratory Data Analysis

01 Data Tidying - PCA Transform already done;

02 Data Manipulation -

03 Data Visualization - normalize confusion matrix

04 Analysis - ROC AUC; CV vs Public LB

Exploratory Data Analysis (EDA)

Training the Data

Feature Selection

Resampling Imbalanced Data

Modelling and Predictive Analysis

Why we explored the following models:

d

Classification Prediction Model

Logistic Regression

Also known as a Generalized Linear Model (GLM)

d

Classification Prediction Model

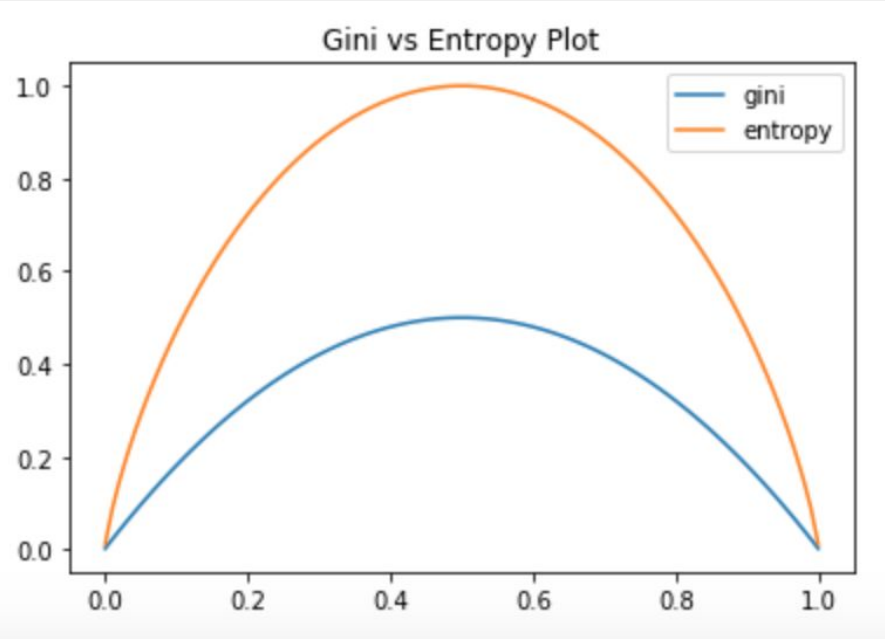
Gaussian Naive Bayes

Classification Prediction Model

Support Vector Machines (SVM)

Classification Prediction Model

Decision Trees



Trees: Gini Coefficient and Information Gain

Classification Prediction Model

Decision Trees

Trees: Bagging / RF

- While Decision Trees are not always an optimal approach, they are useful for their interpretability. By aggregating different approaches, we were able to improve our model's accuracy.
- RF approach averages over a collection of de-correlated Trees. With a big enough # of trees, we do not have to worry about overfitting. RMSE 0.1562
-

Classification Prediction Model

Decision Trees

Trees: Boosting

- Boosting with Cross Validation
- Boosting
  - Trees are grown sequentially; each tree grown based on information from previously grown trees
  - Tuning Parameters
  - XGBoost (high accuracy, scalability, faster) RMSE 0.1294

Classification Prediction Model

Bayesian Optimization

Classification Prediction Model

Neural Nets

Classification Prediction Model

Quantum Algorithms

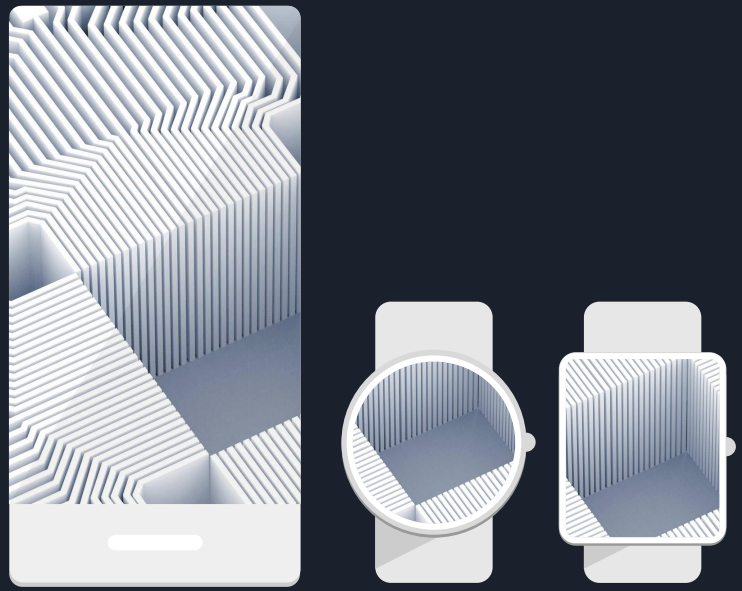
Ensemble Procedure

Voting vs. Average vs. Stacking

Procedures

Feature Engineering most essential

Resampling key for this kind of problem. Not enough time to play around with sequencing..



Ensemble Methods

Stacking/averaging a diverse ensemble of ML algorithms to improve model's predictive accuracy.

Conclusions and Moving Forward

01 [Takeaways from the process, specifically:]

02 [Takeaways from classifical ML algorithms, generally:]

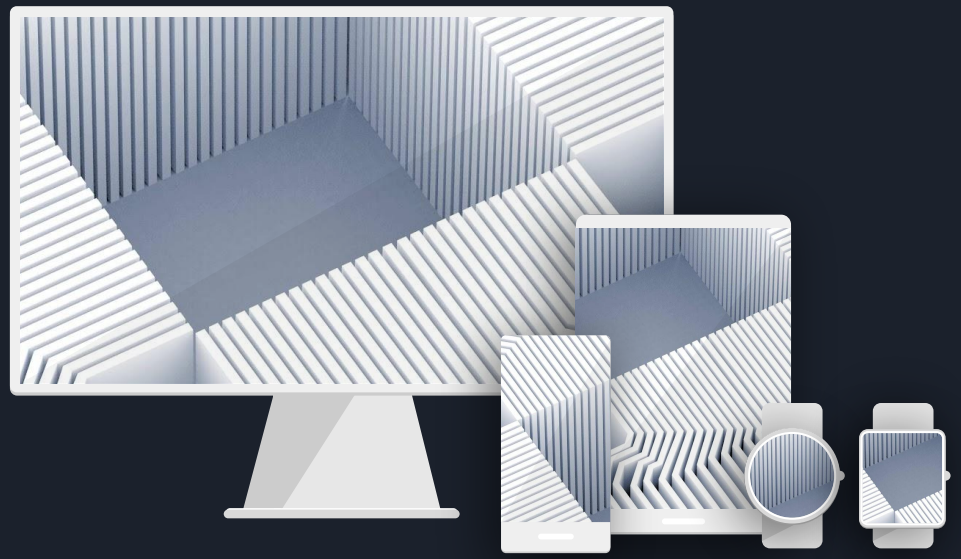
03 With more time, I would explore other software platforms in detail. Quantum algorithms are still in their nascent phase. Bridging the gap between quantum computing and ML will be a major challenge for this industry.

04 Future Applications: Quantum Machine Learning algorithms for financial services, big pharma, & logistical problem-solving are among the major applications of quantum technology so keep these developments in mind as you continue your data science journey.

Thank you!

“Anyone who is not shocked by quantum theory has not understood it.”

- Niels Bohr



License

-