

Candidate Project Overviews

Please complete the following list for at least three projects to which you have made significant contributions. Please provide as much detail as you can in 2-3 sentences.

Project 1

Name: Momentum Factor Model for Point72 L/S Portfolios

Your role: team leader (3-people team)

Duration: 10 months

Problem: measure/classify equities' momentum

Action:

- Constructed data pipeline of day-trade data with engineer team and improved SQL database efficiency for future iterations; preprocessed data on database; setup threshold and cleaned out outliers by deviation.
- Tested training sets on Random Forest, Gradient Boosted Trees; conducted feature selection by RF feature importance.
- Improved test performance by ~12% via adding dropouts, regularization and gradient boosting for RNN model.
- Optimized the half-life parameter based on Barra model and other hyperparameters tuning through hold-out experiments.
- Maintained the platform by validating model performance; conducted weekly parameter tuning and model updating.

Result/Deliverable:

The momentum was classified in an 87% accuracy.

Tools Used/Methodologies:

Linear models; RF; GBM; Tensorflow and some clustering methods.

Your contributions:

Project 2

Name: Portfolio Manager Strategy Simulation

Your role: team member

Duration: 3 years

Problem: Simulate Portfolio trading strategy to predict the potential risk

Action:

- Collected, cleansed, and transformed data from vendors like Barra and Refinitiv by SQL and Pandas across multiple data sources including company performance metrics, market sentiments, geo-location and supply chain.
- Build regression model to predict the potential size of each equities in portfolio.
- Simulated portfolio loss through stress testing; executed risk management by estimating and monitoring 1-5% VaR; adjusted simulation by bootstrapping and optimized the loss estimation performance under skewed distribution assumption; conducted significance test on the conclusion.
- Explain the potential risk to Portfolio Managers by feature importance analysis

Result/Deliverable: Model achieved in low error metrics for backtracking trading data; successfully reduced the potential ~2% book loss by predicting the portfolio size

Tools Used/Methodologies: RF, Bayesian methods, GBM

Your contributions: worked on the feature engineering to feed model and optimized the parameter by backtracking data set testing experiments

Project 3

Name: Topic Modeling on Movie recommendation (school project)

Your role: Individual project

Duration: 1 months

Problem: Cluster unlabeled textual documents into groups and discover latent semantic structures to group movies

Action:

- Build data ETL pipeline to analyze movie rating dataset and conducted online analytical processing (OLAP) with Spark SQL.
- Preprocessed text by tokenizing, stemming and stop-words removing, and extracted features by term frequency-inverse document frequency(TF-IDF) approach.
- Trained unsupervised learning models of K-means clustering and Latent Dirichlet Allocation.
- Identified latent topics and keyword of each document for clustering and calculated document similarity.
- Visualized model training result by dimensionality reduction using PCA.

Result/Deliverable:

Grouped the movie by sentiments/theme topics

Tools Used/Methodologies:

Gram tokenizing, TF-IDF, LDA, clustering, PCA

Your contributions:

Completed the individual project in 1 months.