

Solution Document: Group 3 Project

Trading Strategy Using Machine Learning

Submission 3: Modeling and Strategy Development

Modeling and Strategy Development Solutions Document:

This document serves as a brief history and summary of the solutions that have been attempted to address the challenges laid out by Modeling and Strategy Development submission for the Group project, as well as the challenges met.

Data:

- End Date: 22nd November, 2018
- Data was provided from our submission two (feature engineering) which we import to provide useful features for our model.

Problem Statements:

Modeling

1. Decide on an algorithm or group of algorithms (for example, ensemble techniques).
2. Fit the model
3. Show that it works out of sample, and use appropriate cross-validation techniques.
4. Provide the following performance metrics:
 - (a) ROC curves,
 - (b) Confusion Matrix,
 - (c) Precision, Recall, F1-Score, Accuracy, and AUC.
5. Analysis of metrics and report.

Use Appropriate Cross-Validation Techniques and Ensemble techniques.

Challenge	Potential Solution/Improvement	Current Solution
Splitting the DataFrame (data) into training, cross validation and test set.	Generate NumPy arrays out of the original DataFrames by slicing the data into train and validation set. The training set taking twice the cross-validation set.	Group_3-G_Submission_3.ipynb

	Using sklearn.model_selection to import train_test_split in splitting the train set to training and test set.	
Model in training the training set.	Using XGBoost (extreme Gradient Boosting) which is comparatively better than other techniques	

The goal of cross-validation is to test the model's ability to predict new data that was not used in estimating it, in order to flag problems like overfitting or selection bias and to give an insight on how the model will generalize to an independent dataset (i.e., an unknown dataset, for instance from a real problem).

Goals:

- Complete project

Description:

- We reserve a sample data set (cross-validation set)
- Train the model using the remaining part of the dataset (training set)
- Use the reserve sample of the test (validation) set. This will help us in gauging the effectiveness of the model's performance. If the model delivers a positive result on validation data, we will then go ahead with the current model.

Analysis/Performance:

- Good attempt