

As the global financial market is generating mass data of different types every day, it is becoming more crucial and more difficult to effectively extract and use these data to predict the trend of stocks. The short term timing strategy has a few difficulties as follows:

1. Market sentiments strongly influence the short-term market trend;
2. How to extract effective factors;
3. How to build nonlinear factors;
4. How to solve colinearity among factors.

Fortunately, the development of machine learning offers us a possible approach. In this project, we recognize the future rise or fall as a classification problem and implement several machine learning algorithms to predict the future rise or fall of Wind All A Index, an index indicating the trend of Chinese A Share stocks, to build a short-term timing investment strategy.

We implement a feature selection to choose 18 features (factors) out of 31 daily factor data and 8 alpha factors from WorldQuant101 to establish classifiers using logistic regression, naive Bayes, KNN, perceptron, decision tree, SVM, XGBoost and a Sequential neural network model in Keras to predict the rise or fall of Wind All A Index the next day. The time period of these data is from April 1, 2008 to March 6, 2020. The whole work flow is shown in Figure 1.

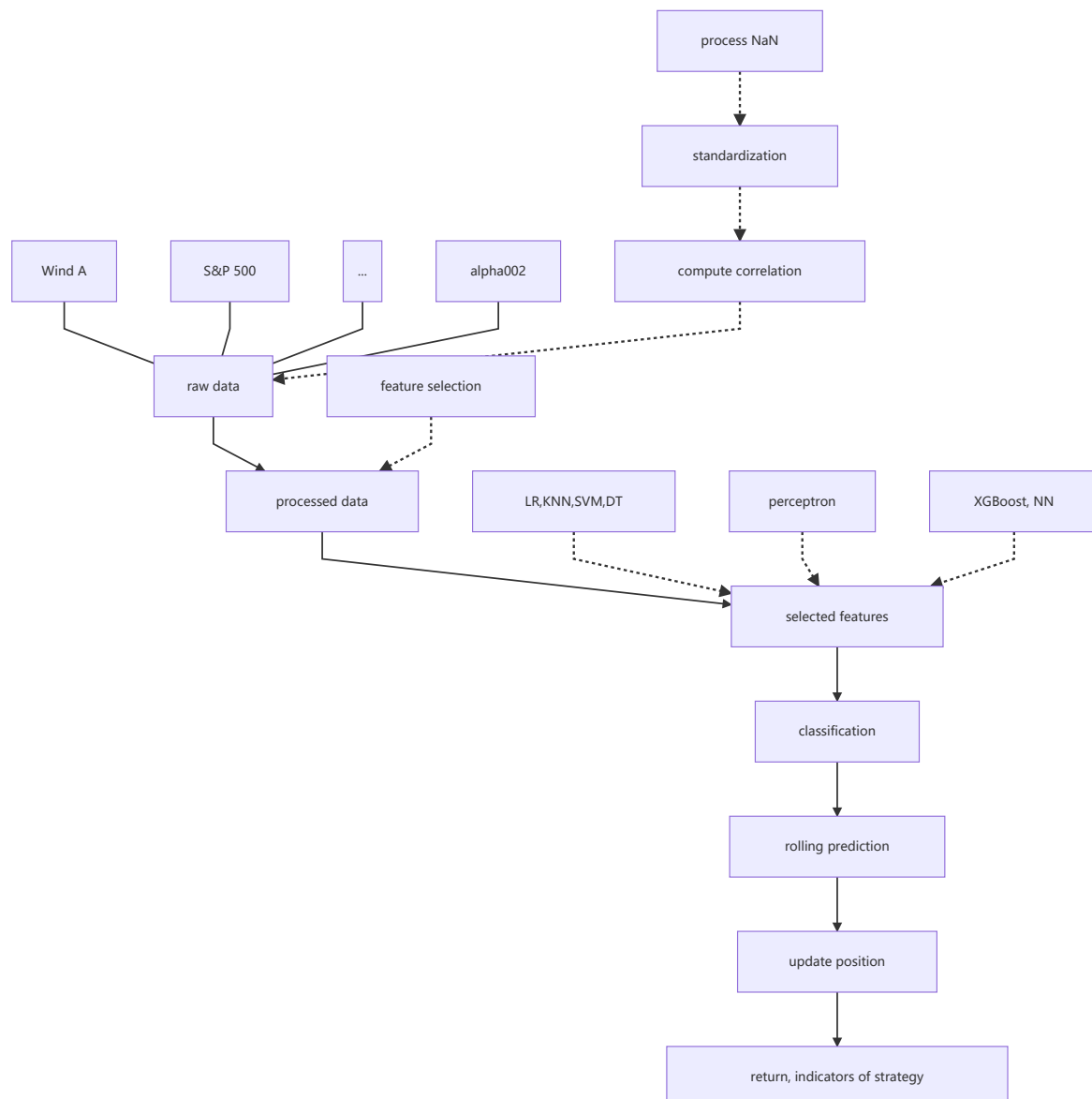


Figure 1. Work flow of the project

As the financial data are time series data, implementing a random splitting of training and testing data sets will lead to the use of future data, which makes the results of the models inaccurate. Therefore, we implement an expanding window backtest procedure as follows:

1. We get at least 1800 days' data as the training dataset and implement different k-fold cross validation methods to tune the hyperparameters for the best model, so the first signal we can get in on February 24, 2015.
2. We implement the best model in the previous step to predict the rise or fall of Wind All A Index the next day. If the predicted signal is rise, then we buy Wind All A Index at the close of the day. If it is predicted to fall, then we short it at the close of the day.
3. We use the best model in Step 2 for 20 consecutive trading days and then add the 20 days' data into the training set in Step 1 to enter Step 1 again.

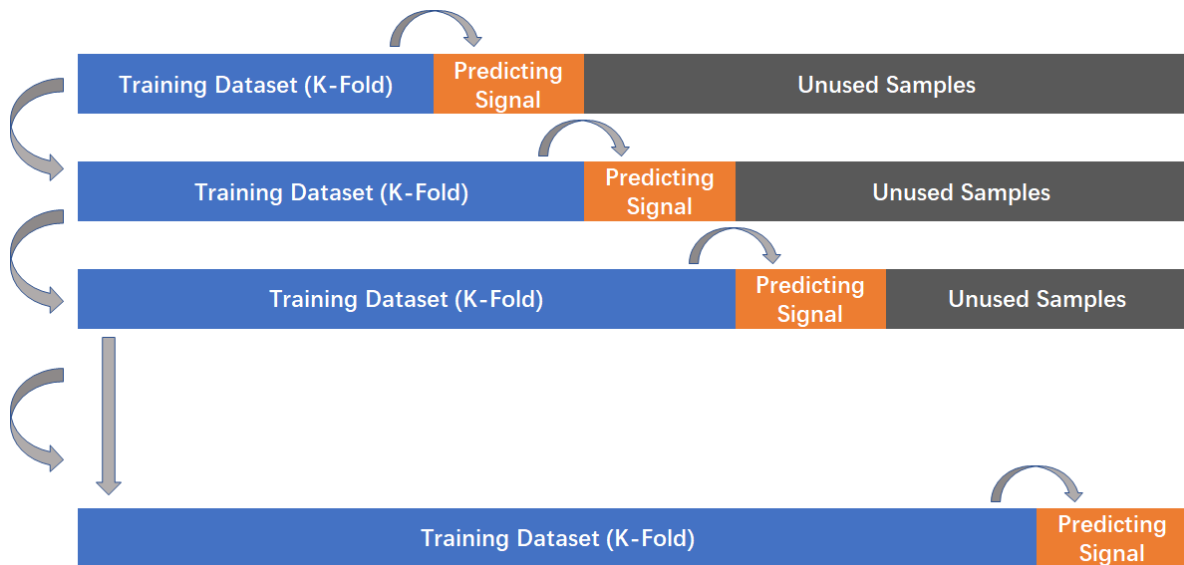


Figure 2. Flow chart of the expanding window backtest model for short-term timing strategy

As we can see from Figure 2, every 20 consecutive trading days the training dataset will expand 20 more days' data.