Quantopian and Machine Learning

Andrés Fernando Román Arévalo, Samael de Jesus Salcedo Amortegui y Germán David García Nieves
afromana@unal.edu.co, sjsalcedoa@unal.edu.co,gdgarcian@unal.edu.co
National University of Colombia

CONTENTS

***Abstract*—Quantopian is a platform where you can create an algorithm strategy for trading. This paper has the objectives to do a brief introduction to quantopian and the trading, also we are going to present some strategies that use machine learning and compare with others that don't**

I. KEY WORDS:

Quantopian, machine learning, trading, stock,invest, supervised learning, unsupervised learning

II. BASIC CONCEPTS:

A. Quantopian:

Quantopian provides free education, data, and tools so anyone can pursue quantitative finance. Aims to create a crowd-sourced hedge fund by letting freelance quantitative analysts develop, test, and use trading algorithms to buy and sell securities.

B. Trading:

It is a profession that consists in the study of the markets through technical analysis and the fundamental analysis to invest in different financial instruments with the objective of obtaining a profit

C. Stock:

All of the shares into which ownership of the corporation is divided. A single share of the stock represents fractional ownership of the corporation in proportion to the total number of shares. Here, these are a set of merchandise or products that are stored pending sale or marketing.

D. Return Of Investment (ROI):

The ROI measures the gain or loss generated on an investment relative to the amount of money invested. ROI is usually expressed as a percentage and is typically used for personal financial decisions, to compare a company's profitability or to compare the efficiency of different investments.

E. Machine Learning (ML):

.

Is the scientific study of algorithms and statistical models that computer systems use in order to perform a specific task effectively without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task.

F. ML Strategies:

1) Supervized Learning: Is the machine learning task of learning a function that maps an input to an output based on example input-output pairs.[1] It infers a function from labeled training data consisting of a set of training examples.

a) SVM: A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. In two dimentional space this hyperplane is a line dividing a plane in two parts where in each class lay in either side.

b) Linear Regression: Simple linear regression is a statistical method that allows us to summarize and study relationships between two continuous (quantitative) variables:

One variable, denoted x, is regarded as the predictor, explanatory, or independent variable. The other variable, denoted y, is regarded as the response, outcome, or dependent variable.

c) Naive Bayes: Naive Bayes classifiers are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naive) independence assumptions between the features.

d) Neural Networks: An ANN is based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit a signal from one artificial neuron to another. An artificial neuron that receives a signal can process it and then signal additional artificial neurons connected to it. The original goal of the ANN approach was to solve problems in the same way that a human brain would.

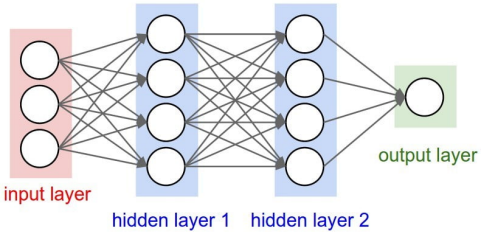

Figure 1: Basic Network

2) Unsupervized Learning: Is a type of learning that helps find previously unknown patterns in data set without pre-existing labels. It is also known as self-organization and allows modeling probability densities of given inputs.

a) K-means: Is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.

III. Introduction

Currently, one of the most popular alternatives to regular jobs is trading. Each day more and more people are trying to earn some money doing this, but unfortunately, this activity requires a lot of discipline and study of the financial market and usually, at the beginning people lose money. For that reason, they believe that trading is a fraud.

Generally trading is an activity that people do by themselves without the help of an automatic algorithm.In other words the decision to invest in certain stocks is taken by the person or trader.

But an alternative is use some bots or algorithms that take the decision of invest in certain stocks by themselves. This algorithms can use different strategies like math, machine learning, neural networks, etc.. and in this way we can improve the profits without needing to be pending all the time of the market

For that reason,we consider that an algorithm trading is a very good alternative to traditional trading because in this way people can use their time doing some others activities while the algorithm make money by itself.

IV. Quantopian

Quantopian is a platform that allows anyone to create an algorithm and invest in different stocks like apple, google, microsoft, etc...The best part is that you don't need spend real money like in traditional trading.

For make a valid algorithm that allows you to participate in the contest and make some money , you need to follow some basics rules:

1) Structural constraints met:
   - **Tradable Universe:** The algorithm is trading within the Quantopian Tradable Universe.
   - **Optimize API:** The algorithm is ordering using the Optimize API.

2) Risk Constraint met:
   - **Sector Exposures**
     Exposure to various economic sectors. The values displayed are the rolling 63-day mean. A list of sectors' numeric codes, as well as their constituent industries
   - **Style Exposures**
     Exposure to various investing styles. The values displayed are the rolling 63-day mean. The relevant styles are:
     a) **Momentum:** The difference in return between assets on an upswing and a downswing over 11 months.
     b) **Size:** The difference in returns between large- and small-cap assets.
     c) **Value:** The difference in returns between expensive and inexpensive assets (as measured by Price/Book ratio).
     d) **Short Term Reversal:** The difference in returns between assets with strong losses to reverse, and

strong gains to reverse, over a short time period.

e) **Volatility:** The difference in return between high- and low-volatility assets.

- **Leverage**
  Contest entries must have between 0.8 and 1.1 of their capital invested in US equities.

- **Turnover**
  Contest entries must have a mean daily turnover between 0.05 - 0.65 measured over a 63-trading-day rolling window.

- **Low Beta-To SPY**
  Contest entries must have an absolute beta-to-SPY below 0.3 (low correlation to the market).

- **Low Position Concentration** Contest entries cannot have more than 0.05 of their capital invested in any one asset.

- **Net Dollar Exposure**
  The net dollar exposure constraint allows a small outlier tolerance. Net dollar exposure can go up to 0.2 at the 100th percentile, and must be under 0.1 at the 98th percentile.

3) Total Returns:
   The total profits of the algorithm must be positive.

V. Strategies

A. Traiding: buy the Stocks

In this strategy a certain amount of a number of shares are purchased. They can be purchased from the previously defined shares and in an order according to priority. The main task is to obtain a profit when it is bought, therefore it is always important at what time it should be invented in one action rather than in another

B. Traiding using Linear Regression

In comparison to the above, in this strategy a linear regression model provided by the python library called "sklearn" will be used, which is commonly used for machine learning. The main task is to obtain a profit when investing in a given action, or when it is decided that the model will not be trained using 7 characteristics collected for 30 days prior to the day in which the prediction was to be made.

C. Traiding using Random forest

Finally, after knowing the performance of the previous strategies, we proceed to train a random forest model using the "sklearn" library. The main task is to obtain a profit when investing in a certain action, or when it is decided that the model will not be trained using 7 characteristics collected for 40 days prior to the day in which the prediction was to be made.

VI. Results

A. Traiding: buy the Stocks

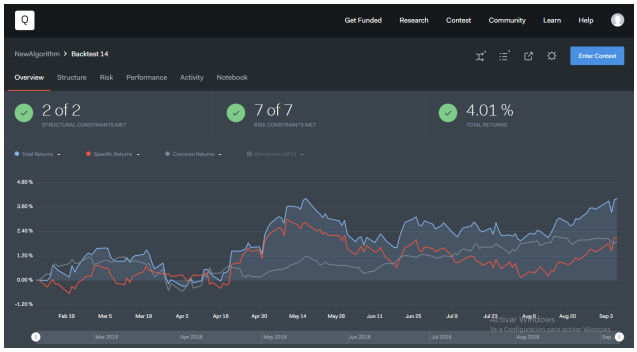

Figure 2: results trainding

The previous graph shows the performance of the algorithm in which the main strategy is to buy shares. This algorithm has a positive gain and complies with the restrictions imposed by quantopian.

B. Traiding using Linear Regression

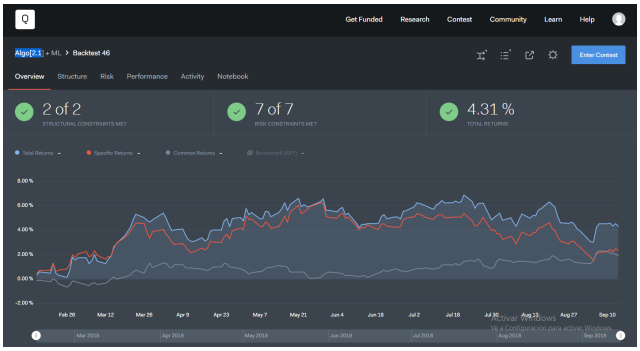

Figure 3: results trainding

The previous graph shows the performance of the algorithm in which the main strategy is to train a linear regression model and from it decide whether to buy a particular share. This algorithm has a positive gain and complies with the restrictions imposed by quantopian.

C. Traiding using Random forest

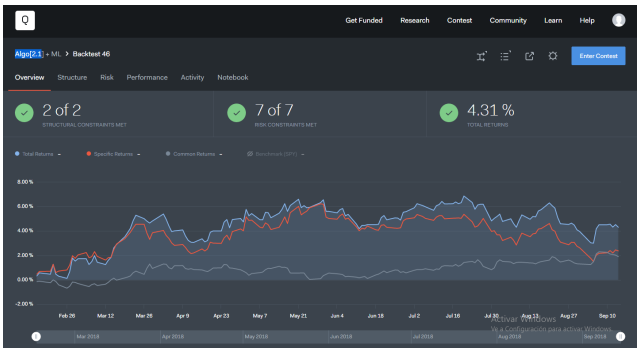

Figure 4: results traiding

The previous graph shows the performance of the algorithm in which the main strategy is to train a random forest model and from the same decide whether to buy a particular share. This algorithm has a positive gain and complies with the restrictions imposed by quantopian.

VII. Conclusions

- The results of perform trading using a random forest model, have a positive tendency in comparison to do trading with the linear regression model or even without having used machine learning
- Through the analysis of Machine Learning, it is possible to take advantage of market behaviors that cannot be seen with the naked eye.
- With the use of certain Machine Learning algorithms, greater profitability is extracted for certain stocks.

References

[1] Thomas Wiecky. (2019). Machine Learning on Quantopian. Basic parts. Website: https://www.quantopian.com/posts/machine-learning-on-quantopian

[2] Rudolph Rusell. (2018). MACHINE LEARNING: Guia Paso a Paso Para Implementar Algoritmos De Machine Learning Con Python.

[3] Oliver Velez, Greg Capra, Isabel Murillo. (2011). Day Trading: Negociación intradía: estrategias y tácticas. Técnicas rentables para todo tipo de estrategias de trading (FINANZAS Y CONTABILIDAD).