Machine Learning Engineer Nanodegree

Capstone project proposal

Bruno Kiyoshi Ynumaru
Florianópolis, Brazil. October 12th, 2020
ynumaru@gmail.com
+55 (48) 9 9861-6921

1. Domain Background

Student briefly details background information of the domain from which the project is proposed. Historical information relevant to the project should be included. It should be clear how or why a problem in the domain can or should be solved. Related academic research should be appropriately cited. A discussion of the student's personal motivation for investigating a particular problem in the domain is encouraged but not required.

Time-series prediction of asset prices or asset price movements is one the most studied Deep-Learning applications in Finances.

Technical Analysis is a field in Finances in which an analyst makes predictions on future asset prices based on price charts. Technical Analysis fundamentals were postulated by American journalist Charles Dow, in the late XIX century, in what is known as the Dow Theory.

There are six main components to the Dow Theory, namely:

1. The Market Discounts Everything
2. There Are Three Primary Kinds of Market Trends
3. Primary Trends Have Three Phases
4. Indices Must Confirm Each Other
5. Volume Must Confirm the Trend
6. Trends Persist Until a Clear Reversal Occurs

The detailed explanation of each of these components is beyond the scope of this document, but they will be further explained in the Jupyter Notebook if/when the concepts are used.

The most used type of chart in this field is the candlestick chart. The horizontal axis represents time, and the vertical axis represents asset price.

Candlestick chart of EUR/USD currency pair on daily timeframe in MetaTrader 5 trading platform. (CC BY-SA 4.0 Available at https://upload.wikimedia.org/wikipedia/commons/3/3c/Candlestick_Chart_in_MetaTrader_5.png)

Each time-step in a candlestick chart is called a candle, and it indicates four basic values: the open price, the closing price, and the maximum and minimum values a given asset is negotiated at in each timeframe.

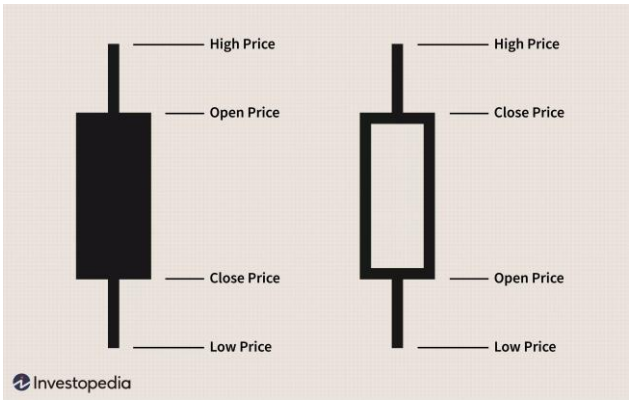


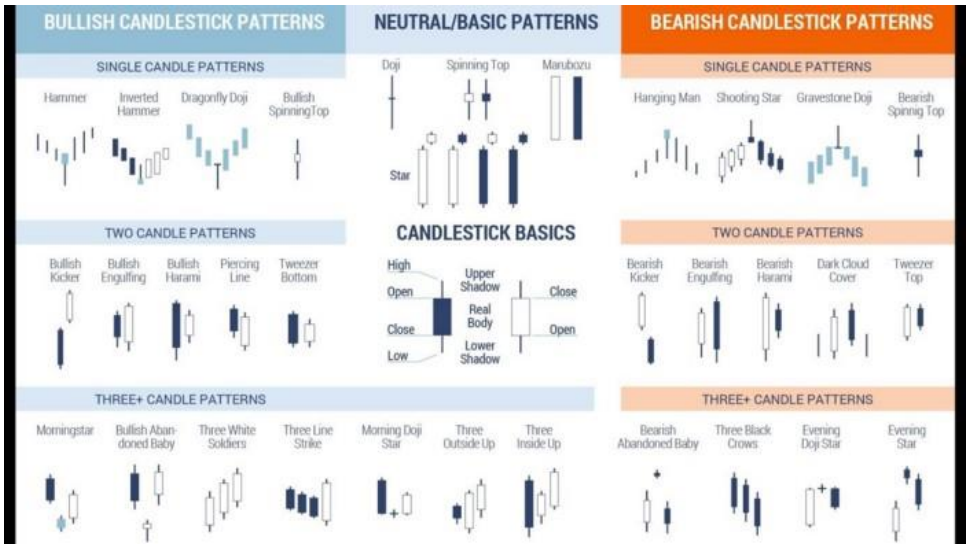
Image by Julie Bang © Investopedia 2019 (available at https://www.investopedia.com/thmb/9nwzVl-16xZlP9epM3b2H7kdBSc=/6250x0/filters:no_upscale():max_bytes(150000):strip_icc():format(webp)/UnderstandingBasicCandlestickCharts-01_2-7114a9af472f4a2cb5cbe4878c1767da.png)

Classically, white or green candles indicate that the closing price is higher than the opening price, and the opposite is indicated by black or red color. Historically, some candlestick patterns have shown to indicate future movements of asset prices (they can be both indications of reversal and continuation movements). Some of these patterns are:

1. Bullish Engulfing
2. Hammer
3. Bullish Harami
4. Bearish Engulfing
5. Falling Star
6. Bearish Harami
7. Doji
8. Three Line Strike
9. Two Black Gapping

10. Three Black Crows
11. Evening Star
12. Abandoned Baby

These patterns can be visualized in the following figure:



Visualization of some of the main candlestick patterns used in Technical Analysis. (Available at https://miro.medium.com/max/700/1*OpahFyAd6nHkdVtO59DX9w.jpeg)

Detailed explanation of how these patterns work and what they predict is also beyond the scope of this document.

2. Problem Statement

Student clearly describes the problem that is to be solved. The problem is well defined and has at least one relevant potential solution. Additionally, the problem is quantifiable, measurable, and replicable.

Financial market forecasting can be both a regression and a classification problem. If one tries to predict the price a given asset will have in the future, it's a regression problem. Simply trying to foresee whether a price is going to rise, fall, or continue the same is a classification problem. Similar methods can be used to predict indexes such as the S&P 500 – and their related ETF's if one exists.

3. Datasets and inputs

The dataset(s) and/or input(s) to be used in the project are thoroughly described. Information such as how the dataset or input is (was) obtained, and the characteristics of the dataset or input, should be included. It should be clear how the dataset(s) or input(s) will be used in the project and whether their use is appropriate given the context of the problem.

Stock prices available at the Brazilian Stock Exchange, B3, for one or more assets still to be selected. Data should contain daily prices for this stock, including open, close, maximum and minimum values for a period of at least a one year.

From this data, some technical analysis features shall be extracted, such as moving averages, Bollinger bands, and fitness to candlestick patterns.

4. Solution statement

Student clearly describes a solution to the problem. The solution is applicable to the project domain and appropriate for the dataset(s) or input(s) given. Additionally, the solution is quantifiable, measurable, and replicable.

Feature Engineering: Candlestick patterns will be identified and labeled inside the input time-series data. Also moving averages and Bolinger Bands boundary values will be added to the dataset.

More than one ML model could be tried, but primarily LSTM should be applied for this project.

5. Benchmark model

A benchmark model is provided that relates to the domain, problem statement, and intended solution. Ideally, the student's benchmark model provides context for existing methods or known information in the domain and problem given, which can then be objectively compared to the student's solution. The benchmark model is clearly defined and measurable.

There are many academic studies in the price-prediction field, as can be assessed in [1], but [3] also uses LSTM and was applied to the Brazilian stock index IBOVESPA, which should be one of the study objects of the project.

6. Evaluation metrics

Student proposes at least one evaluation metric that can be used to quantify the performance of both the benchmark model and the solution model presented. The evaluation metric(s) proposed are appropriate given the context of the data, the problem statement, and the intended solution.

We should evaluate whether the model correctly predicts an average upward or downward on the asset price on a determined number of candlesticks after a candlestick pattern is detected. Also, we can try to determine if the model can set a target price for the given asset and compare it to the value reached by the asset.

7. Project design

Student summarizes a theoretical workflow for approaching a solution given the problem. Discussion is made as to what strategies may be employed, what analysis of the data might be required, or which algorithms will be considered. The workflow and discussion provided align with the qualities of the project. Small visualizations, pseudocode, or diagrams are encouraged but not required.

First step will be doing some exploration on the data, check perceivable trends and see how the market behaved in different times, including financial crisis days, such as the ones in March 2020.

Second step will be to feature engineer the time-series data, and label time frames where a technical analyst could have pointed a candlestick pattern. Bollinger bands, moving averages and other features can be engineered as well.

Data shall then be split into training and test data, models shall be created and trained. The endpoint will be created and, if enough time is available, a web app could be created, using API Gateway and Lambda, as learned during the Nanodegree.

8. References

[1] "Financial time series forecasting with deep learning : A ...." 01 Mai. 2020, https://www.sciencedirect.com/science/article/pii/S1568494620301216.
[2] " On stock return prediction with LSTM networks" 10 Jul. 2017, https://lup.lub.lu.se/student-papers/search/publication/8911069.