Machine Learning Engineer Nanodegree

Capstone Proposal

Adelmo M. A. Filho
January 1st, 2020

Predicting the Ibovespa through recurrent neural networks

Proposal

Domain Background

Bovespa Index (Ibovespa) is one of the most important benchmark index traded on the B3, stock exchange located in São Paulo, Brazil. Ibovespa takes into account around 80 stocks that comprehend brazilian commpanies from multiple sectors (financial, mining, oil & gas, electric utilities) and, every four months, it is revised when participation of stocks on the index can be changed (FINKLER, 2017).

Described as an indicator of the average performance of the most tradable and representative assets of the Brazilian stock market (FARIA, 2012), Ibovespa fluatuations and trends tend to represent important aspects of brazilian economy, such as foreign investments, monetary policy decisions and political issues.

Problem Statement

For this project, a time series regression to predict the closing value for Bovespa index for the next trading day is proposed.

In general terms, Bovespa index closing value can be defined as a function of its previous values (endogenous varibles) and independent (exogenous) variables, for example, calendar variables (weekday, month), stock values, dolar exchange.

The use of such kind of model provides useful support to decisions, simulate different scenarios and understand variables importance for the Bovespa index closing value.

Datasets and Inputs

The datasets are provided by the python package Yahooquery which works as a wrapper for an unofficial Yahoo Finance API. Data used on this project was obtained for free, there was no need for a Yahoo Finance premium subscription.

The `history` method from `Ticker` class of Yahooquery package allows to retrive daily data about stock markets. The following table shows a sample of historical data for the Bovespa Index.

symbol	date	open	close	low	high	volume
^BVSP	2020-04-08	76335.0	78625.0	76115.0	79058.0	10206300.0
^BVSP	2020-04-07	74078.0	76358.0	74078.0	79855.0	11286500.0
^BVSP	2020-04-06	69556.0	74073.0	69556.0	75260.0	9685400.0
^BVSP	2020-04-03	72241.0	69538.0	67802.0	72241.0	10411300.0
^BVSP	2020-04-02	70969.0	72253.0	70957.0	73861.0	10540200.0

Not only Bovespa Index data is expected to be used on this project, but also historical data from the main stocks that represents its portfolio. The following table presents the main stocks that compose Bovespa Index and their global participation on the portfolio.

Ticker	Company	IBOVESPA Participation
ITUB4	Itaú Unibanco Holding S.A.	10,50%
BBDC4	Banco Bradesco S.A.	9,12%
VALE3	Vale S.A.	8,59%
PETR4	Petróleo Brasileiro S.A. - Petrobras	7,06%
PETR3	Petróleo Brasileiro S.A. - Petrobras	5,14%
ABEV3	Ambev S.A.	5,14%
BBAS3	Banco do Brasil S.A.	4,47%
B3SA3	B3 S.A. - Brasil, Bolsa, Balcão	4,15%
ITSA4	Itaúsa - Investimentos Itaú S.A.	3,86%

Solution Statement

(approx. 1 paragraph)

In this section, clearly describe a solution to the problem. The solution should be applicable to the project domain and appropriate for the dataset(s) or input(s) given. Additionally, describe the solution thoroughly such that it is clear that the solution is quantifiable (the solution can be expressed in mathematical or logical terms) , measurable (the solution can be measured by some metric and clearly observed), and replicable (the solution can be reproduced and occurs more than once).

Benchmark Model

(approximately 1-2 paragraphs)

In this section, provide the details for a benchmark model or result that relates to the domain, problem statement, and intended solution. Ideally, the benchmark model or result contextualizes existing methods or known information in the domain and problem given, which could then be objectively compared to the solution. Describe how the benchmark model or result is measurable (can be

measured by some metric and clearly observed) with thorough detail.

Evaluation Metrics

(approx. 1-2 paragraphs)

In this section, propose at least one evaluation metric that can be used to quantify the performance of both the benchmark model and the solution model. The evaluation metric(s) you propose should be appropriate given the context of the data, the problem statement, and the intended solution. Describe how the evaluation metric(s) are derived and provide an example of their mathematical representations (if applicable). Complex evaluation metrics should be clearly defined and quantifiable (can be expressed in mathematical or logical terms).

Project Design

(approx. 1 page)

In this final section, summarize a theoretical workflow for approaching a solution given the problem. Provide thorough discussion for what strategies you may consider employing, what analysis of the data might be required before being used, or which algorithms will be considered for your implementation. The workflow and discussion that you provide should align with the qualities of the previous sections. Additionally, you are encouraged to include small visualizations, pseudocode, or diagrams to aid in describing the project design, but it is not required. The discussion should clearly outline your intended workflow of the capstone project.

References

Before submitting your proposal, ask yourself. . .

- Does the proposal you have written follow a well-organized structure similar to that of the project template?
- Is each section (particularly **Solution Statement** and **Project Design**) written in a clear, concise and specific fashion? Are there any ambiguous terms or phrases that need clarification?
- Would the intended audience of your project be able to understand your proposal?
- Have you properly proofread your proposal to assure there are minimal grammatical and spelling mistakes?
- Are all the resources used for this project correctly cited and referenced?