Project Proposal

1 Project Title

Machine Learning Stock Price Prediction

2 Team Members

- Augusto Menezes Savaris
- Ian Gonzalez-Alfonzo

3 Project Details

3.1 Project Objective

Please describe your project objective in this part.

- What is the project Objective? Our objective is to analyze the feasibility of using Machine Learning to predict future stock prices. We will compare and evaluate how different regression models perform under this task.
- What problem to solve? The problem to solve is obtaining an accurate prediction for a stock based on past stock performance.
- Why the problem is important? Predicting future stock price movement given the past is important for giving investors an additional tool for analyzing the market. Many people, nowadays, invest in the stock market and giving them a more educated prediction based on different Machine Learning algorithms can be useful.
- Why machine learning can help to solve the problem? Since machine learning can accurately predict values along a range of possible values in a continuous way, it is a good way to approach the problem of stock price prediction. Furthermore, Neural Networks (one of the selected algorithms) can learn very complex functions to approximate the underlying distributions. Since stock price movement is complex and non-linear, ML is a good tool for this problem.

3.2 Datasets

Please describe your dataset in this section.

- What is the data and where you obtain it? We will obtain the data from Kaggle (). We downloaded a .csv file containing the price for stocks of the Brazilian exchange from 1994 up until 2020. It contains the ticker (name of the stock), date for each row, and the prices and volume for that day (open, high, low, close, volume).
- How the data is collected? The data was already collected and is available to be downloaded on Kaggle. As mentioned before, we just downloaded the dataset.
- What will be the features and labels you will use? The features will be the past prices: open, low, high, and close for each trading day. We will also use volume as a feature. The labels will be the open price for the next day.

- How many examples for training, validation, and testing? Our dataset has 1.8 million rows. We will use 0.7 of the dataset for training, 0.2 for validation, and 0.1 for testing.

3.3 Machine Learning Algorithm

Please describe the machine learning algorithm you want to use for your project.
Please justify your selection.
Since we want to compare different algorithms for this regression problem, we will fit, evaluate and compare a linear regression algorithm, a support vector machine algorithm, and a neural network regression algorithm. We expect the data to have a non-linear relationship with the labels. Hence, we are choosing SVM and Neural Networks because they do a good job at approximating complex functions. Furthermore, Neural Networks usually perform well with large datasets, which is the case for this problem. We will also study how linear regression performs so that we have a comparison and a baseline for the other two algorithms.

3.4 Expected Outcomes

What is your expected outcome for this project?
Our expected outcome is to have a comprehensive view of how the three selected algorithms perform predicting future stock price movement. The problem of predicting future stock prices is complex as it deals with very noisy data. Thus, good results are not an expected outcome - instead, the expected outcome is to be able to conclude if predicting future stock prices using only past price movement is possible by using the selected algorithms.