Capstone Project Proposal

Lidan An

Problem statement:

The goal of this project is to analyze the google company historical stock data in Python, build a machine learning model to predict the google stock buying signal. The model can be an indicator for the investor to decide whether to buy or not. It is beneficial for companies and individuals to make proper investment decisions.

Context:

The Stock market process is full of uncertainty. The price is affected by many factors. However, the Stock market prediction is one of the important factors in finance and business. There are two types of analysis possible for prediction, technical and fundamental. In this project, only technical analysis is considered. The method involves collecting data sets, cleaning data, evaluating different algorithms, analyzing the output and making conclusions.

1.We first need to collect our data. Before collecting data, we need to understand our project. Then decide what kind of data we need, what kind of software and tools we require and how we are going to achieve our outcomes. After enough stock market research, we collected our data from a website called Kaggle. Besides everything, It is important to define possible risks and solutions.

2. To use the python programming language, we first need to install a new version of Python, the software and import some libraries. The software we need to install includes sciPy. The libraries we need to import include pandas, numpy, matplotlib and seaborn.

3. The data collection process can have numerous loopholes. There are different factors which promote inaccuracy in data, for example, errors in readings of machines, useless rows and columns etc. To handle these issues, we need to clean the data to make it easier to read. The data cleaning steps include sorting data, editing data and statistical adjustment etc.

4. Once we prepare the data, the next step is to analyze that data using different algorithms then find the best one with the highest prediction rate.

5. The last step is to analyze the outcome from the algorithms. From the outcome, the investor can make better decisions, the company can adjust their marketing strategy.

Criteria for Success:

Define the tasks that we are going to achieve then work on those tasks step by step. Then collect the information related to topics, gather all the information together and create a final draft of the project.

Prepare the clean data for analyzing, discover the pattern and trend from the data visualization. Find the proper features and target value to build a machine learning model.

Scope of solution space:

1. The data that we get from websites is usually unstructured data, so our first task is going to be to create structured data using Python programming language.
2. We will use Notebook in Anaconda for data analysis, it works with python programming language.

3. When working with a big data file, data loss is the most common problem. To deal with this, we will always store backups for our data file. If there is any data loss, we can use our backup file.

4. There are a number of different algorithms to analyze data files. First, we will evaluate three different algorithms to find the optimal one. Then use the chosen one for the final analysis.

Constraints:

1. Find structured stock market data.

2. Software error.

3. Data loss.

4. Comparing different algorithms.

Stakeholders:

Individual businesses

Consumers

Investor

Key Data Sources:

https://www.kaggle.com/datasets/dgawlik/nyse