Predicting Stock Price by Using Linear Regression Model

Abstract

The use of data science in the stock exchange is not new, but this does not apply to the Saudi Stock Exchange (Tadawul), it needs to be explored and studied in depth. This project aims to analyze the Saudi stock market (Tadawul) data, as it answers some questions related to the data, such as: What is the day that achieves the highest trades? What companies have the highest stock price? What companies have the lowest stock price? Then I built a model that predicts stock closing prices for these companies(TAPRCO and STC) which are two oldest companies in the communication service sector by using the (LR) model.

Design:

The data is provided by the Saudi Stock Exchange (Tadawul), we built a model using a linear regression model to predict the stock closing price of the two oldest companies in the communication service sector (TAPRCO and STC).

Dataset:

Dataset: The data of Saudi stock market companies since 2000-01-01, collected from Saudi Stock Exchange (Tadawul). This data set contains 593819 rows, and 14 columns. In this project I try to focus on companies that have high stock prices and a large number of deals.

Regarding the characteristics I expect to work on that will help me come up with a certain idea are:

- name(String) Name of the company
- trading_name (String): The trading name of the company
- sectoer (String): The sector in which the company operates
- date (Date): The date of the stock price
- open (Decimal): The opening price
- high (Decimal): The highest price of the stock at that day
- low (Decimal): The lowest price of the stock at that day

- change (Decimal): The change in price from the last day
- perc_Change (Decimal): The percentage of the change
- volume_traded (Decimal): The volume of the trades for the day
- value_traded (Decimal): The value of the trades for the day

Algorithms

- Data cleaning by deleting null values and duplicate rows, and deleting some features that will not add anything to the model (eg (symbol, name, sector).
- Converting some attribute types to suitable types to facilitate data analysis. We converted the feature type (date), which was an object type, to type (datetime64[ns]).

- Adding a new column for the day for each row, as we extracted the day from the (date) feature.

Model:

In this project, I used a linear regression model to predict stock prices for both companies (TAPRCO and STC).

Model Splitting:

A data set was first divided into two main sets, the first is for (TAPRCO) company data, and the second is for (STC) company data.Then, each data set from these two was divided into two sets, one for training, representing 80% of the data, and the other for testing, representing 20% of the data.

Tools

I worked on this project using python language and using the following libraries(Pandas, Numpy, Matplotlip, seaborn, sklearn).

Communication

The two graphs represent the linear regression models for both companies (TAPRCO and STC).