

Building a Better Forecast Model for Apple's Stock Price Prediction

Problem Statement:

Currently, Apple is one of the top 5 Fortune 500 companies in the Nasdaq stock exchange. Millions of people around the world closely watch its stock price. Therefore, accurately forecasting Apple stock prices is of great interest to millions of individual investors, stakeholders, suppliers, manufacturers, and others. It's known that during economic downturns, stock market prediction becomes a more complex task. However, by capturing economic downturn events, historical stock data provide helpful information to build or improve models for better prediction coupled with relevant features such as dividends and earnings. This project aims to build a better forecast model to predict Apple stock prices.

Data Science Pipeline:

Data Acquisition

The stock price data is available on the Kaggle website. It was sourced from NASDAQ, Yahoo Finance, Zacks, and Alpha Vantage. The data information consisted of 4000+ companies over the years 1998 – 2021. The raw data consists of four dataset files, three files containing stock data (dividends, earnings, and stock prices), and one file containing a summary of the three stock datasets.

The Dividends dataset contains three variables: symbol, date, and dividend. The earning dataset contains six variables: symbol, date, qtr, eps_est, eps, and release_date. The stock prices dataset contains nine variables: symbol, date, open, high, low, close, close_adjusted, volume, and split_coefficient. As we can notice, the symbol and date variables are present in the three stocks datasets.

Data Wrangling

Data wrangling included removing and replacing missing values (NaN) and checking for duplicate rows. The NaN removal and replacing process was only performed in the dividends and earning datasets. The stock prices dataset doesn't have any missing values. Then, the Apple stock data was extracted, and the dividends, earnings, and stock prices datasets were merged into one set Apple and saved for further analysis.

Exploratory Data Analysis (EDA)

In the exploratory data analysis process, time-series of stock prices, dividends, and earnings data provide useful visualization to determine the daily, weekly, monthly, and quarterly stock price behavior. Using the Pearson correlation formula, the correlation between variables is also computed. Lastly, only the variables relevant for forecast modeling are saved for further modeling

Data Modeling

In this process, machine learning forecasting Arima and other models were examined to build the best forecasting model for this data.