

Stock Prediction Capstone – Sprint 2

By: Will Lin

Problem Statement

How can I use data science to predict future stock prices?



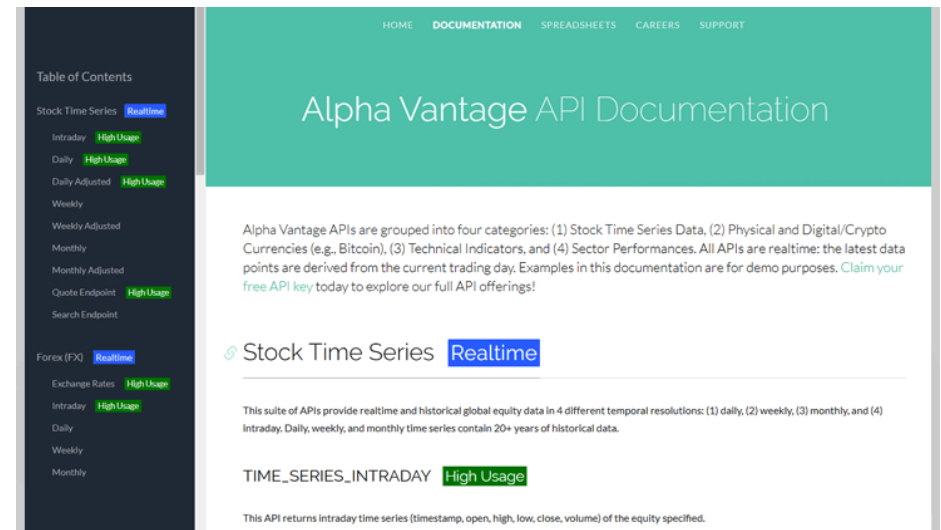
Solution:

Use machine learning models that cater to regression problems.

Start off with baseline models and add two more complex models

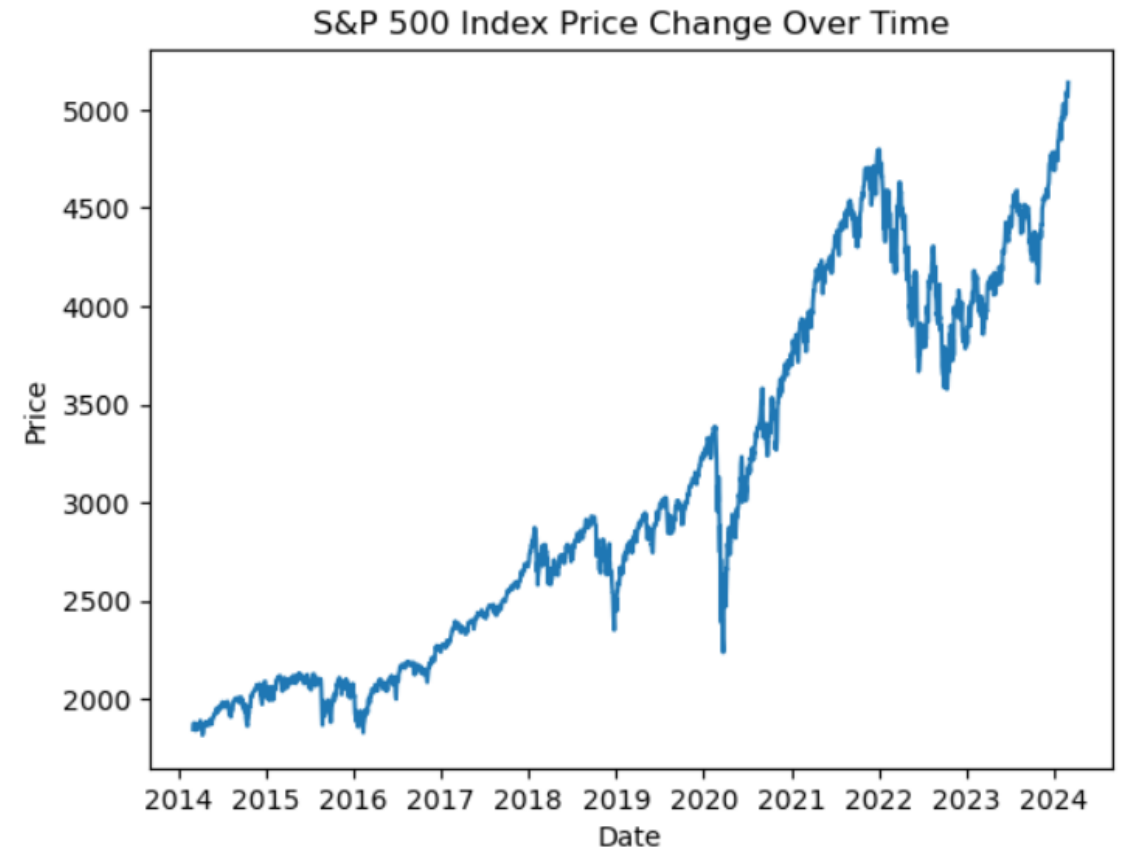
Overview of Datasets

- Obtained premium access to a new API called Alpha Vantage
- Obtained the adjusted price datasets that accounts for price splits
- Able to pull Technical Indicators which I can add to potentially improve modelling

A screenshot of the Alpha Vantage API Documentation website. The page has a green header with navigation links: HOME, DOCUMENTATION, SPREADSHEETS, CAREERS, and SUPPORT. The main title 'Alpha Vantage API Documentation' is centered in white. Below the title, a paragraph explains that APIs are grouped into four categories: (1) Stock Time Series Data, (2) Physical and Digital/Crypto Currencies, (3) Technical Indicators, and (4) Sector Performances. It notes that all APIs are realtime and data points are derived from the current trading day. A link to 'Claim your free API key' is provided. On the left, a dark sidebar contains a 'Table of Contents' with links to 'Stock Time Series' (Realtime), 'Forex (FX)' (Realtime), and 'Exchange Rates' (High Usage). Under 'Stock Time Series', there are links for 'Intraday' (High Usage), 'Daily' (High Usage), 'Daily Adjusted' (High Usage), 'Weekly', 'Weekly Adjusted', 'Monthly', 'Monthly Adjusted', 'Quote Endpoint' (High Usage), and 'Search Endpoint'. The main content area shows the 'Stock Time Series' section with a 'Realtime' tag. It describes the suite of APIs providing realtime and historical global equity data in 4 different temporal resolutions: (1) daily, (2) weekly, (3) monthly, and (4) intraday. It notes that daily, weekly, and monthly time series contain 20+ years of historical data. The 'TIME_SERIES_INTRADAY' endpoint is highlighted with a 'High Usage' tag. A brief description states: 'This API returns intraday time series (timestamp, open, high, low, close, volume) of the equity specified.'

EDA

- Time is your friend in investing
- Over a long period of time, your money will grow



Baseline Models

- Linear Regression
- Random Forest Regressor

	Model	Metrics	Result
0	Linear Regression	MAE	1.691710
1	Linear Regression	MSE	5.481171
2	Random Forest Regressor	MAE	138.618410
3	Random Forest Regressor	MSE	19229.054351

Plan and Next Steps

Identify Area of Interest and
Problem Statement

Acquire a dataset

Basic EDA/Preprocessing

Modelling

Model Tuning & Adding
features for better predictions