Group Name: Stock Analysts

Member: Matthew Zhou, [mattjzhou888@gmail.com](mailto:mattjzhou888@gmail.com), USA, Georgia Institute of Technology, Data Science

Problem Description: Predict future stock prices using time-series data

Business understanding: Having an accurate stock price predictor will allow business such as banks and investment funds to make better decisions. We can sell this product to them as SaaS.

Project lifecycle: Understand dataset, create models, test/validate models, create web api, create data pipeline to stream new data, test final product.

Dataset: <https://www.kaggle.com/datasets/debashis74017/stock-market-data-nifty-50-stocks-1-min-data>

Data understanding: Stock data of last price, current ask price and volume, current bid price and volume every 10 seconds (timestamped in the dataset).

Problems in the data: There were no NA values in the data. However, the timestamps were sometimes delayed (could be more than 10 seconds between data points).

Approaches to overcome problems: Use the closest datapoint to the 10 second interval. Modify the streaming source to constantly refresh when data is delayed.

Data cleansing: No NA values to handle. Worked on improving stream to reduce gaps in time.

Github Repo Link: https://github.com/mattjzhou888/dg\_project.git