

DOCUMENTATION

Name: Vaishnavi Kashyap

Objective: The objective of this assignment is the analysis and forecasting of historical stock prices for a dataset comprising 500 different stocks. The primary objectives are to identify patterns in stock price movements, select a subset of five stocks for in-depth forecasting, and implement time series forecasting models to predict future stock price trends.

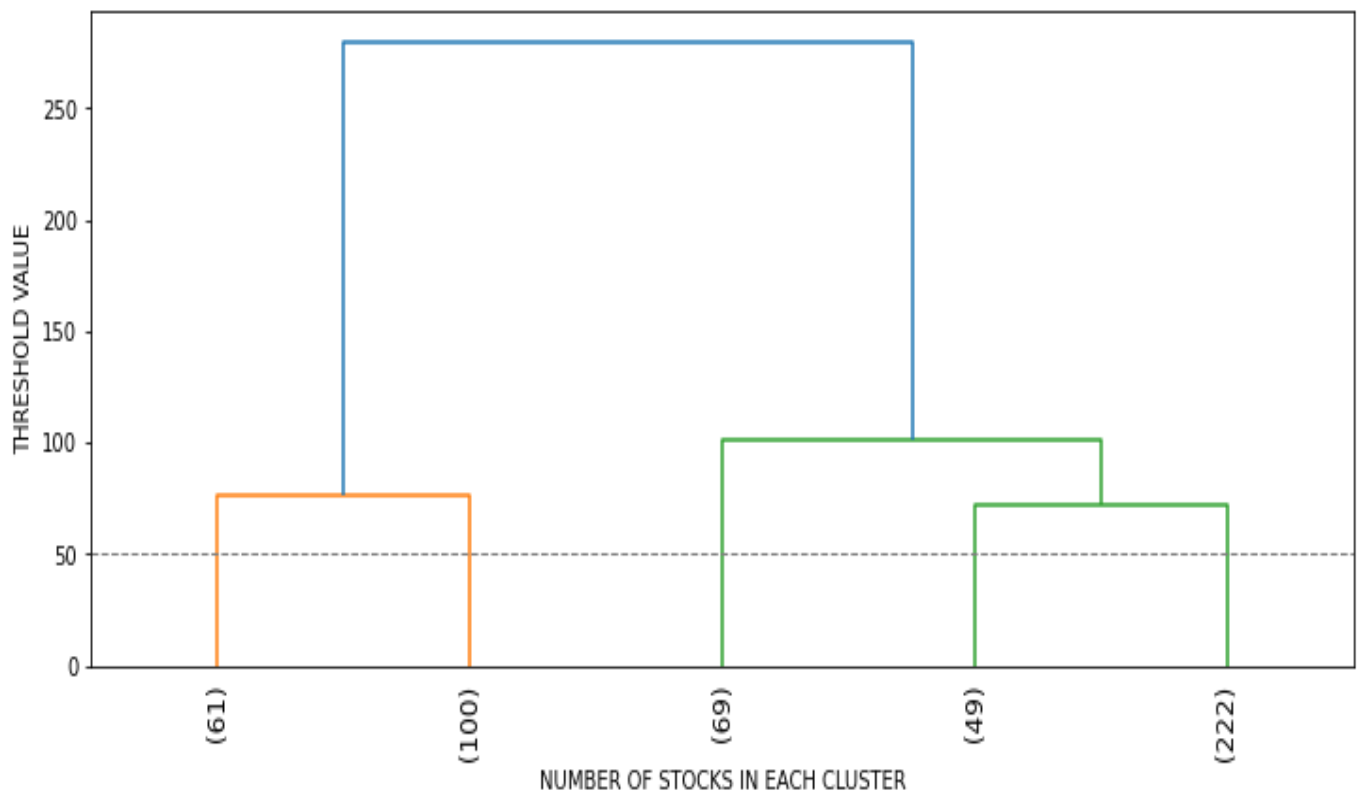
❖ DATA PREPROCESSING -

- The first step in the data preprocessing was to convert the date column from Unix timestamps to datetime objects.
- The dataset initially contained minute-by-minute stock price data. To obtain a more manageable and informative view of the data, the data was resampled to hourly intervals, taking the mean of prices within each hour. Resampling to hourly intervals helps in simplifying the dataset while retaining the overall trend and variation in stock prices.
- Any columns with NaN (Not a Number) values were dropped from the dataset.

❖ ANALYSIS OF PRICE MOVEMENTS OF STOCKS –

Hierarchal clustering was performed on the 500 stocks to derive a total of 5 clusters. Selecting a lower threshold value would lead to identification of more clusters with increased similarity in each cluster. To aid visibility, a cluster count of 5 was decided.

- *Dendrogram* –



A matrix of correlation between the stocks was created. Creating a dendrogram based on the correlation between different stocks can help understand how similar or dissimilar their price movements are. Each of these clusters is composed of stocks that have exhibited similar price movements during the analyzed time frame.

Stocks within the same cluster may be influenced by similar market factors, economic conditions, or industry trends. Investors can use this information for portfolio diversification or risk management by avoiding overexposure to stocks within the same cluster.

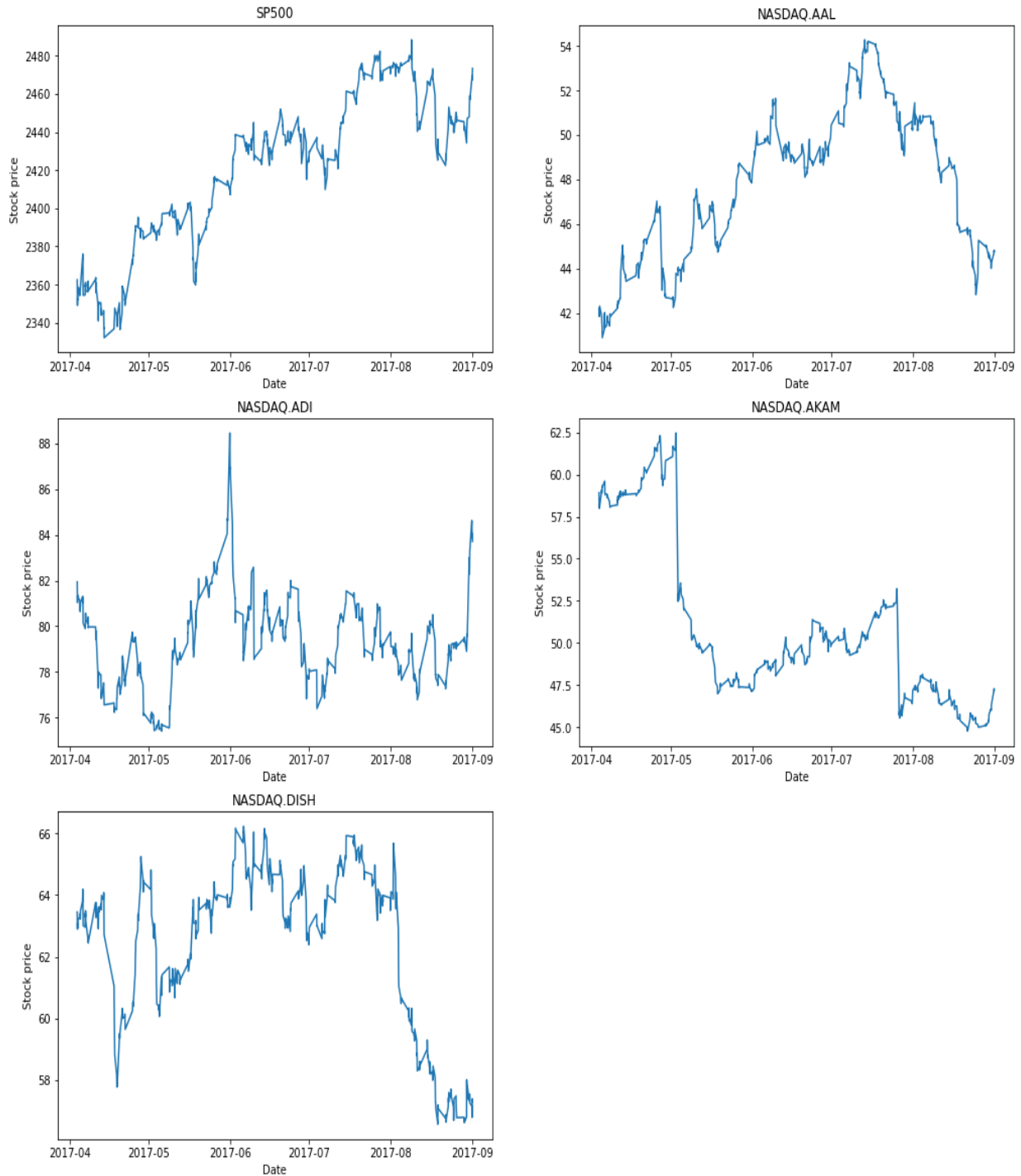
• *Stocks in every cluster*

Cluster no.	Stocks
5	['NASDAQ.DISH', 'NASDAQ.ESRX', 'NASDAQ.GT', 'NASDAQ.HOLX', 'NASDAQ.IDXX', 'NASDAQ.KLAC', 'NASDAQ.MDLZ', 'NASDAQ.MYL', 'NASDAQ.NAVI', 'NASDAQ.PAYX', 'NASDAQ.PBCT', 'NASDAQ.PDCO', 'NASDAQ.QCOM', 'NASDAQ.SBUX', 'NASDAQ.XRAY', 'NYSE.ABC', 'NYSE.ADS', 'NYSE.AES', 'NYSE.ALK', 'NYSE.AVB', 'NYSE.CAH', 'NYSE.CFG', 'NYSE.CL', 'NYSE.CVS', 'NYSE.DHR', 'NYSE.ETN', 'NYSE.EVHC', 'NYSE.F', 'NYSE.GGP', 'NYSE.HCA', 'NYSE.HCP', 'NYSE.HD', 'NYSE.HST', 'NYSE.HSY', 'NYSE.IPG', 'NYSE.JCI', 'NYSE.JEC', 'NYSE.JNPR', 'NYSE.LB', 'NYSE.LEG', 'NYSE.LLY', 'NYSE.LUK', 'NYSE.LVLT', 'NYSE.MET', 'NYSE.MLM', 'NYSE.MO', 'NYSE.MRK', 'NYSE.NUE', 'NYSE.PNR', 'NYSE.PPG', 'NYSE.RHI', 'NYSE.SCG', 'NYSE.TIF', 'NYSE.UA', 'NYSE.UAA', 'NYSE.UAL', 'NYSE.UHS', 'NYSE.UNP', 'NYSE.VMC', 'NYSE.WFC', 'NYSE.ZBH']
4	['NASDAQ.AKAM', 'NASDAQ.BBBY', 'NASDAQ.CHRW', 'NASDAQ.COST', 'NASDAQ.CSCO', 'NASDAQ.CTXS', 'NASDAQ.DISCA', 'NASDAQ.DISCK', 'NASDAQ.DLTR', 'NASDAQ.FAST', 'NASDAQ.FFIV', 'NASDAQ.FOX', 'NASDAQ.FOXA', 'NASDAQ.INFO', 'NASDAQ.INTC', 'NASDAQ.KHC', 'NASDAQ.MAT', 'NASDAQ.ONLY', 'NASDAQ.ROST', 'NASDAQ.SRCL', 'NASDAQ.STX', 'NASDAQ.SYMC', 'NASDAQ.TRIP', 'NASDAQ.TSCO', 'NASDAQ.ULTA', 'NASDAQ.VIAB', 'NASDAQ.WBA', 'NYSE.AAP', 'NYSE.ADM', 'NYSE.APA', 'NYSE.APC', 'NYSE.ARNC', 'NYSE.AZO', 'NYSE.BHI', 'NYSE.BLL', 'NYSE.BXP', 'NYSE.CAG', 'NYSE.CHK', 'NYSE.CMG', 'NYSE.COP', 'NYSE.CPB', 'NYSE.CTL', 'NYSE.CXO', 'NYSE.DFS', 'NYSE.DIS', 'NYSE.DPS', 'NYSE.DVA', 'NYSE.DVN', 'NYSE.EOG', 'NYSE.FL', 'NYSE.FLR', 'NYSE.FLS', 'NYSE.FTI', 'NYSE.GE', 'NYSE.GIS', 'NYSE.GPC', 'NYSE.GPS', 'NYSE.GWW', 'NYSE.HAL', 'NYSE.HES', 'NYSE.HOG', 'NYSE.HP', 'NYSE.HPE', 'NYSE.HRL', 'NYSE.IBM', 'NYSE.K', 'NYSE.KMB', 'NYSE.KMI', 'NYSE.KR', 'NYSE.LOW', 'NYSE.M', 'NYSE.MAC', 'NYSE.MKC', 'NYSE.MOS', 'NYSE.MRO', 'NYSE.NBL', 'NYSE.NFX', 'NYSE.NOV', 'NYSE.OMC', 'NYSE.OXY', 'NYSE.PSA', 'NYSE.PXD', 'NYSE.RIG', 'NYSE.RRC', 'NYSE.SIG', 'NYSE.SJM', 'NYSE.SLB', 'NYSE.SLG', 'NYSE.SNA', 'NYSE.SO', 'NYSE.SPG', 'NYSE.SYY', 'NYSE.T', 'NYSE.TAP', 'NYSE.TJX', 'NYSE.VNO', 'NYSE.WU', 'NYSE.WY', 'NYSE.XEC', 'NYSE.XOM']
3	['NASDAQ.ADI', 'NASDAQ.AMD', 'NASDAQ.CMCSA', 'NASDAQ.EXPD', 'NASDAQ.FLIR', 'NASDAQ.GRMN', 'NASDAQ.HBAN', 'NASDAQ.ILMN', 'NASDAQ.INCY', 'NASDAQ.NTAP', 'NASDAQ.PCAR', 'NASDAQ.QRVO', 'NASDAQ.SNI', 'NASDAQ.SWKS', 'NASDAQ.TXN', 'NYSE.AGN', 'NYSE.AIV', 'NYSE.AN', 'NYSE.BF.B', 'NYSE.CBS', 'NYSE.CLX', 'NYSE.CNP', 'NYSE.COF', 'NYSE.COG', 'NYSE.CVX', 'NYSE.D', 'NYSE.DLR', 'NYSE.EIX', 'NYSE.EMR', 'NYSE.EQT', 'NYSE.ETR', 'NYSE.EXR', 'NYSE.FCX', 'NYSE.FE', 'NYSE.FRT', 'NYSE.GS', 'NYSE.HPQ', 'NYSE.IFF', 'NYSE.IRM', 'NYSE.JWN', 'NYSE.KIM', 'NYSE.KORS', 'NYSE.KSS', 'NYSE.LYB', 'NYSE.MAA', 'NYSE.MNK', 'NYSE.MUR', 'NYSE.NEM', 'NYSE.NKE', 'NYSE.NLSN', 'NYSE.O', 'NYSE.OKE', 'NYSE.PCG', 'NYSE.PEG', 'NYSE.PFE', 'NYSE.PG', 'NYSE.PH', 'NYSE.PPL', 'NYSE.PRU', 'NYSE.PWR', 'NYSE.RAI', 'NYSE.REG', 'NYSE.RL', 'NYSE.SYF', 'NYSE.TGT', 'NYSE.TSN', 'NYSE.URI', 'NYSE.VZ', 'NYSE.WMB']
2	['NASDAQ.AAL', 'NASDAQ.CA', 'NASDAQ.CERN', 'NASDAQ.CSX', 'NASDAQ.GOOG', 'NASDAQ.GOOGLE', 'NASDAQ.HAS', 'NASDAQ.HSIC', 'NASDAQ.MAR', 'NASDAQ.MU',

	'NASDAQ.NTRS', 'NASDAQ.VRSK', 'NASDAQ.WDC', 'NASDAQ.XLNX', 'NYSE.AIG', 'NYSE.AIZ', 'NYSE.ALLE', 'NYSE.BEN', 'NYSE.CHD', 'NYSE.CMA', 'NYSE.CMI', 'NYSE.CO', 'NYSE.COTY', 'NYSE.DAL', 'NYSE.DRI', 'NYSE.FBHS', 'NYSE.GLW', 'NYSE.HCN', 'NYSE.IP', 'NYSE.IR', 'NYSE.ITW', 'NYSE.IVZ', 'NYSE.KEY', 'NYSE.LUV', 'NYSE.MCK', 'NYSE.MDT', 'NYSE.MON', 'NYSE.MPC', 'NYSE.MTB', 'NYSE.NSC', 'NYSE.NWL', 'NYSE.PM', 'NYSE.SCHW', 'NYSE.SEE', 'NYSE.SHW', 'NYSE.SWK', 'NYSE.UTX', 'NYSE.VTR', 'NYSE.WHR']
1	['SP500', 'NASDAQ.AAPL', 'NASDAQ.ADBE', 'NASDAQ.ADP', 'NASDAQ.ADSK', 'NASDAQ.ALXN', 'NASDAQ.AMAT', 'NASDAQ.AMGN', 'NASDAQ.AMZN', 'NASDAQ.ATVI', 'NASDAQ.AVGO', 'NASDAQ.BIIB', 'NASDAQ.CBOE', 'NASDAQ.CELG', 'NASDAQ.CHTR', 'NASDAQ.CINF', 'NASDAQ.CME', 'NASDAQ.CTAS', 'NASDAQ.CTSH', 'NASDAQ.EA', 'NASDAQ.EBAY', 'NASDAQ.EQIX', 'NASDAQ.ETFC', 'NASDAQ.EXPE', 'NASDAQ.FB', 'NASDAQ.FISV', 'NASDAQ.FITB', 'NASDAQ.GILD', 'NASDAQ.INTU', 'NASDAQ.ISRG', 'NASDAQ.JBHT', 'NASDAQ.LKQ', 'NASDAQ.LRCX', 'NASDAQ.MCHP', 'NASDAQ.MNST', 'NASDAQ.MSFT', 'NASDAQ.NDAQ', 'NASDAQ.NFLX', 'NASDAQ.NVDA', 'NASDAQ.NWS', 'NASDAQ.NWSA', 'NASDAQ.PCLN', 'NASDAQ.PYPL', 'NASDAQ.REGN', 'NASDAQ.SNPS', 'NASDAQ.SPLS', 'NASDAQ.TROW', 'NASDAQ.VRSN', 'NASDAQ.VRTX', 'NASDAQ.WFM', 'NASDAQ.WLTW', 'NASDAQ.WYNN', 'NASDAQ.ZION', 'NYSE.A', 'NYSE.ABBV', 'NYSE.ABT', 'NYSE.ACN', 'NYSE.AEE', 'NYSE.AEP', 'NYSE.AET', 'NYSE.AFL', 'NYSE.AJG', 'NYSE.ALB', 'NYSE.ALL', 'NYSE.AME', 'NYSE.AMG', 'NYSE.AMP', 'NYSE.AMT', 'NYSE.ANTM', 'NYSE.AON', 'NYSE.APD', 'NYSE.APH', 'NYSE.ARE', 'NYSE.AVY', 'NYSE.AWK', 'NYSE.AXP', 'NYSE.AYT', 'NYSE.BA', 'NYSE.BAC', 'NYSE.BAX', 'NYSE.BBT', 'NYSE.BBY', 'NYSE.BCR', 'NYSE.BDX', 'NYSE.BK', 'NYSE.BLK', 'NYSE.BMY', 'NYSE.BRK.B', 'NYSE.BSX', 'NYSE.BWA', 'NYSE.C', 'NYSE.CAT', 'NYSE.CB', 'NYSE.CBG', 'NYSE.CCI', 'NYSE.CCL', 'NYSE.CF', 'NYSE.CI', 'NYSE.CMS', 'NYSE.CNC', 'NYSE.COL', 'NYSE.COO', 'NYSE.CRM', 'NYSE.CSRA', 'NYSE.DD', 'NYSE.DE', 'NYSE.DG', 'NYSE.DGX', 'NYSE.DHI', 'NYSE.DLPH', 'NYSE.DOV', 'NYSE.DOW', 'NYSE.DTE', 'NYSE.DUK', 'NYSE.DXC', 'NYSE.ECL', 'NYSE.ED', 'NYSE.EFX', 'NYSE.EL', 'NYSE.EMN', 'NYSE.EQR', 'NYSE.ES', 'NYSE.ESS', 'NYSE.EW', 'NYSE.EXC', 'NYSE.FDX', 'NYSE.FIS', 'NYSE.FMC', 'NYSE.FTV', 'NYSE.GD', 'NYSE.GM', 'NYSE.GPN', 'NYSE.HBI', 'NYSE.HIG', 'NYSE.HON', 'NYSE.HRB', 'NYSE.HRS', 'NYSE.HUM', 'NYSE.ICE', 'NYSE.IT', 'NYSE.JNJ', 'NYSE.JPM', 'NYSE.KMX', 'NYSE.KO', 'NYSE.KSU', 'NYSE.L', 'NYSE.LEN', 'NYSE.LH', 'NYSE.LLL', 'NYSE.LMT', 'NYSE.LNC', 'NYSE.LNT', 'NYSE.MA', 'NYSE.MAS', 'NYSE.MCD', 'NYSE.MCO', 'NYSE.MHK', 'NYSE.MMC', 'NYSE.MMM', 'NYSE.MS', 'NYSE.MSI', 'NYSE.NEE', 'NYSE.NI', 'NYSE.NOC', 'NYSE.NRG', 'NYSE.ORCL', 'NYSE.PEP', 'NYSE.PFG', 'NYSE.PGR', 'NYSE.PHM', 'NYSE.PKI', 'NYSE.PLD', 'NYSE.PNC', 'NYSE.PNW', 'NYSE.PRGO', 'NYSE.PSX', 'NYSE.PVH', 'NYSE.PX', 'NYSE.RCL', 'NYSE.RF', 'NYSE.RHT', 'NYSE.RJF', 'NYSE.ROK', 'NYSE.ROP', 'NYSE.RSG', 'NYSE.RTN', 'NYSE.SPGI', 'NYSE.SRE', 'NYSE.STI', 'NYSE.STT', 'NYSE.STZ', 'NYSE.SYK', 'NYSE.TDG', 'NYSE.TEL', 'NYSE.TMK', 'NYSE.TMO', 'NYSE.TRV', 'NYSE.TSO', 'NYSE.TSS', 'NYSE.TWX', 'NYSE.TXT', 'NYSE.UDR', 'NYSE.UNH', 'NYSE.UNM', 'NYSE.UPS', 'NYSE.USB', 'NYSE.V', 'NYSE.VAR', 'NYSE.VFC', 'NYSE.VLO', 'NYSE.WAT', 'NYSE.WEC', 'NYSE.WM', 'NYSE.WMT', 'NYSE.WRK', 'NYSE.WYN', 'NYSE.XEL', 'NYSE.XL', 'NYSE.XRX', 'NYSE.XYL', 'NYSE.YUM', 'NYSE.ZTS']

❖ STOCK PRICE FORECASTING -

- The five stocks selected are – SP500, NASDAQ.AAL, NASDAQ.ADI, NASDAQ.AKAM, NASDAQ.DISH. All stocks selected for forecast belong to different clusters.



The above diagram shows the stock prices for the five selected stocks based on the hourly resampling of data.

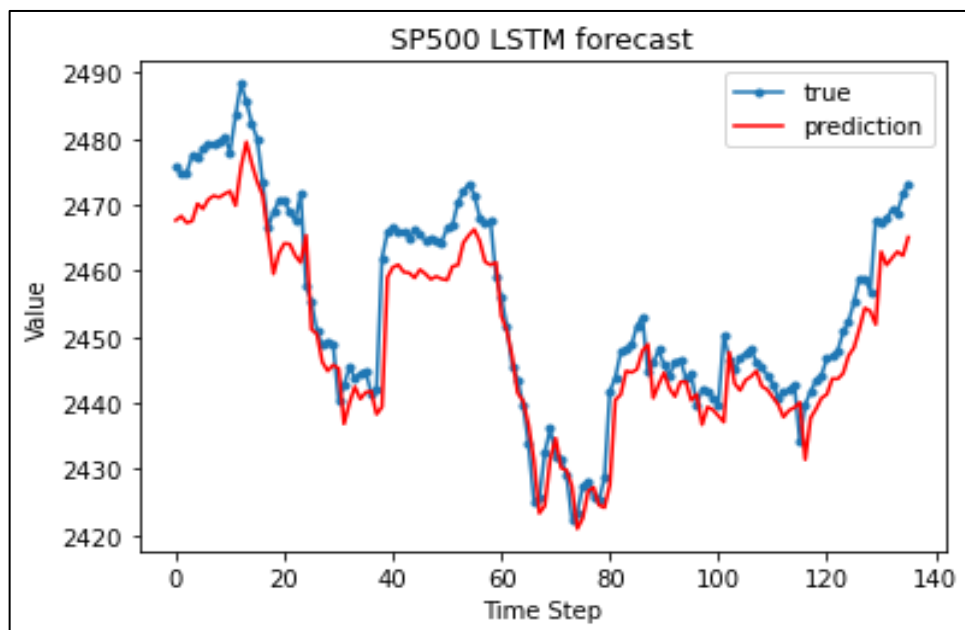
A. LSTM MODEL

Data preprocessing

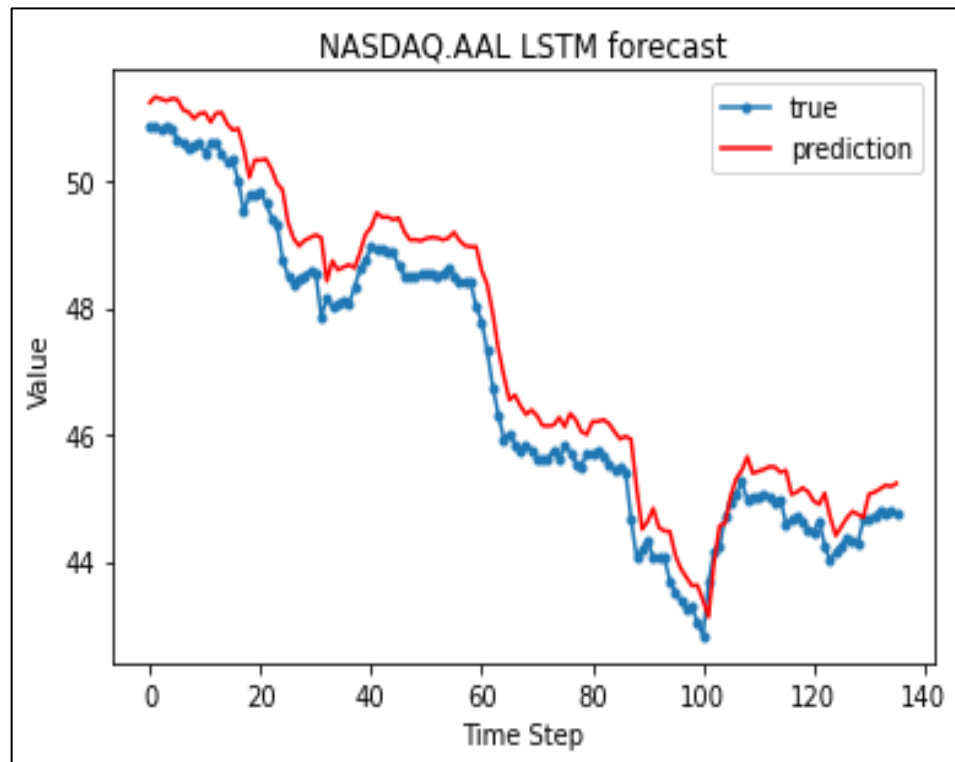
- Applied Min-Max scaling to rescale the stock prices, ensuring they fall within the range of 0 to 1.
- Given that this is a time series dataset, it is crucial to maintain the temporal order when splitting. Used the first 80% of the data as the training set and the remaining 20% as the test set.
- To create training sequences, used a step size of 12. This means that for each data point, used the previous 12 data points to predict the 13th data point.
- Reshaped the training data to prepare it for input into the LSTM model. The shape of the input data should be in the format (samples, time steps, features) for LSTM. (time steps = 12, features = 1).

Visualization for forecasted stock prices

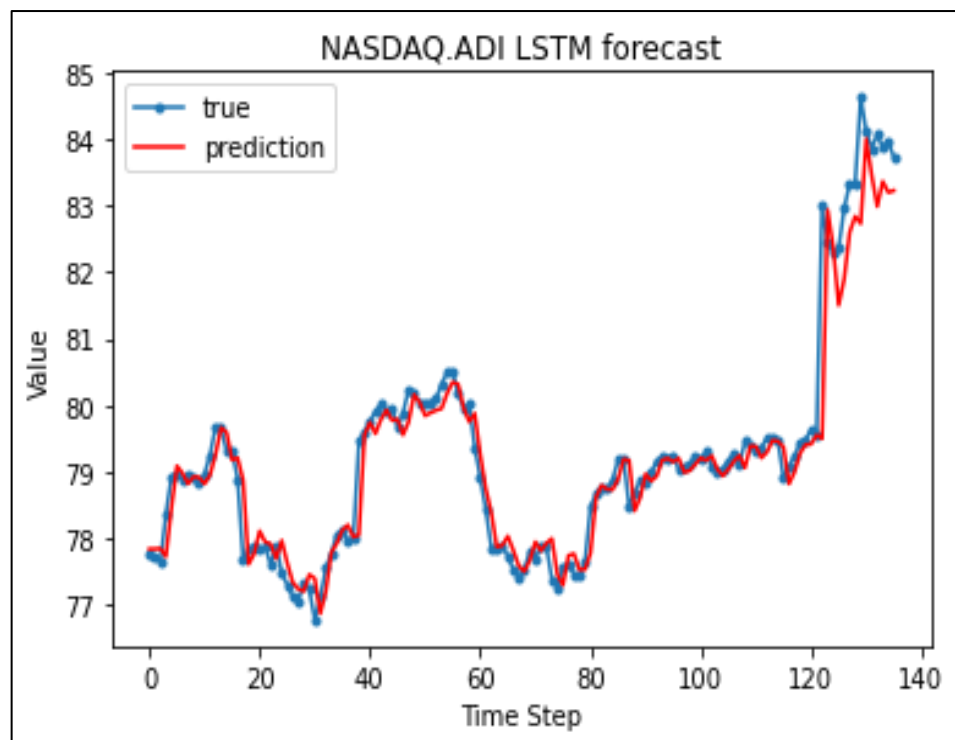
- SP500 -



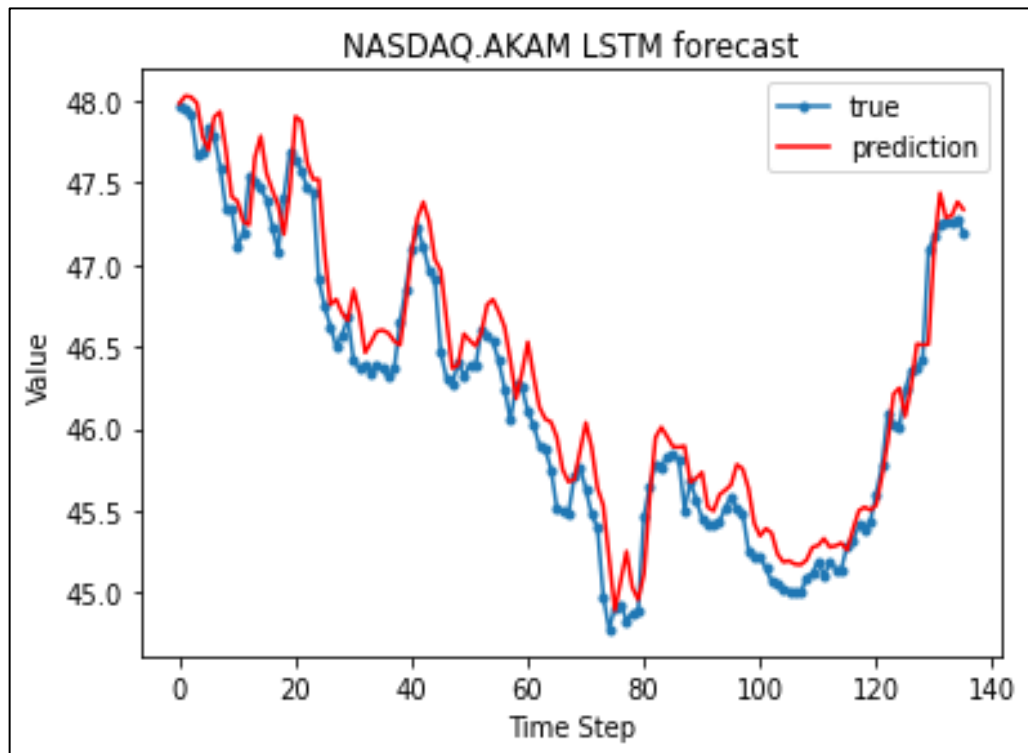
- NASDAQ.AAL -



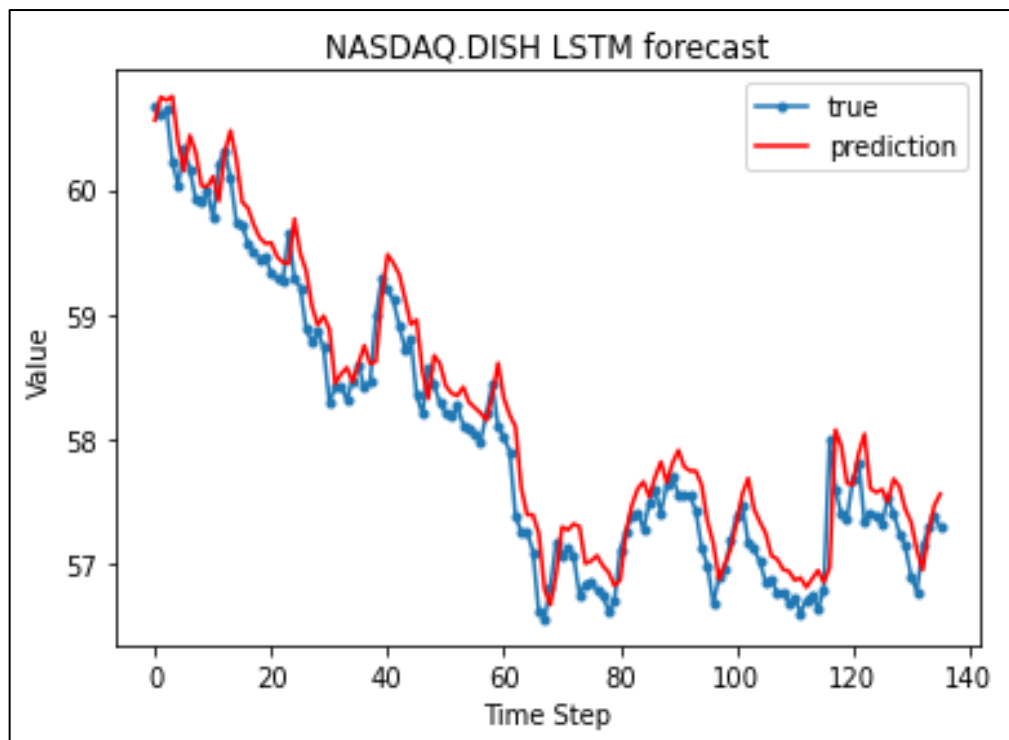
- NASDAQ.ADI –



- NASDAQ.AKAM -



- NASDAQ.DISH –



B. ARIMA MODEL

Data preprocessing -

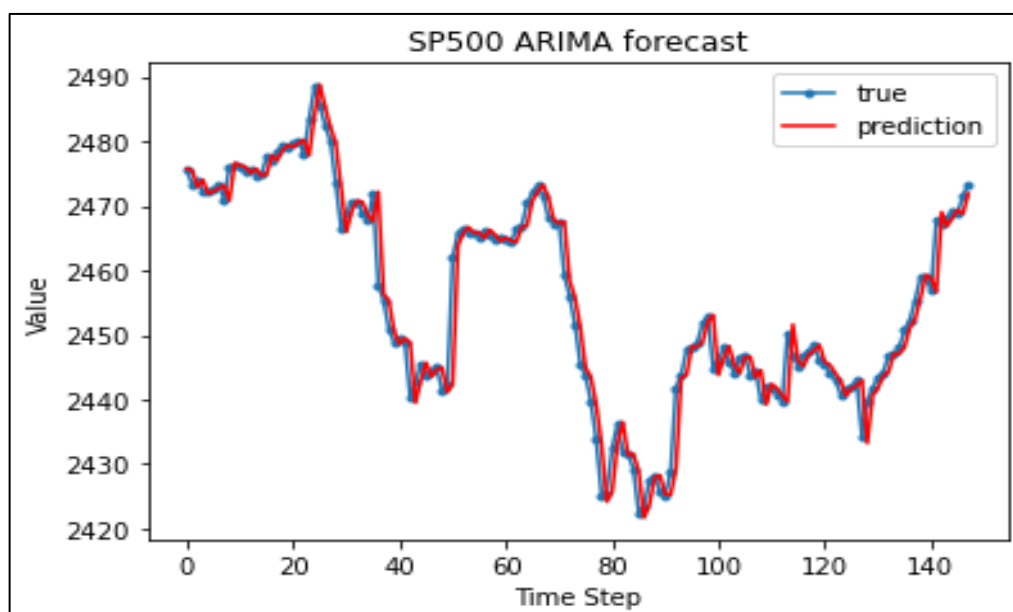
- Given that this is a time series dataset, it is crucial to maintain the temporal order when splitting. Used the first 80% of the data as the training set and the remaining 20% as the test set.

Rolling forecast –

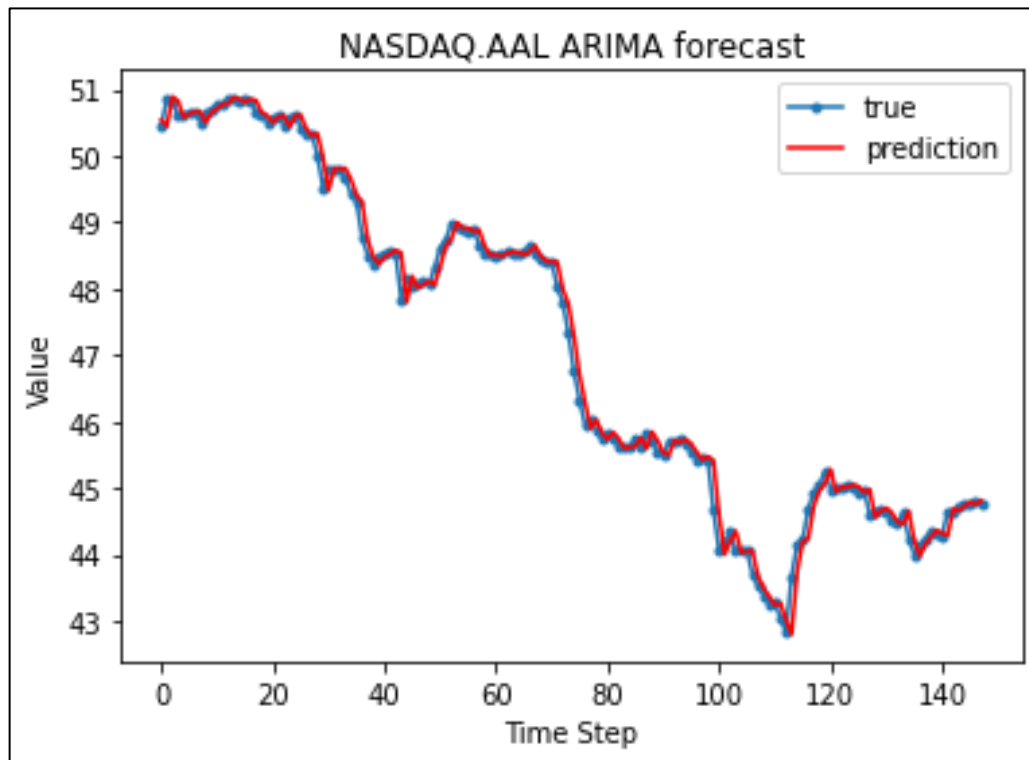
- Incorporated the actual stock price for the current time step into the training set. Refit the ARIMA model using the updated training data
- Generated a new forecast for the next time step using the retrained model. This forecast becomes the prediction for that time point.
- Continued the process iteratively. After each prediction, updated the training data, retrained the model, and made the next forecast.
- This rolling forecast procedure is repeated until the entire test set is covered.

Visualization for forecasted stock prices

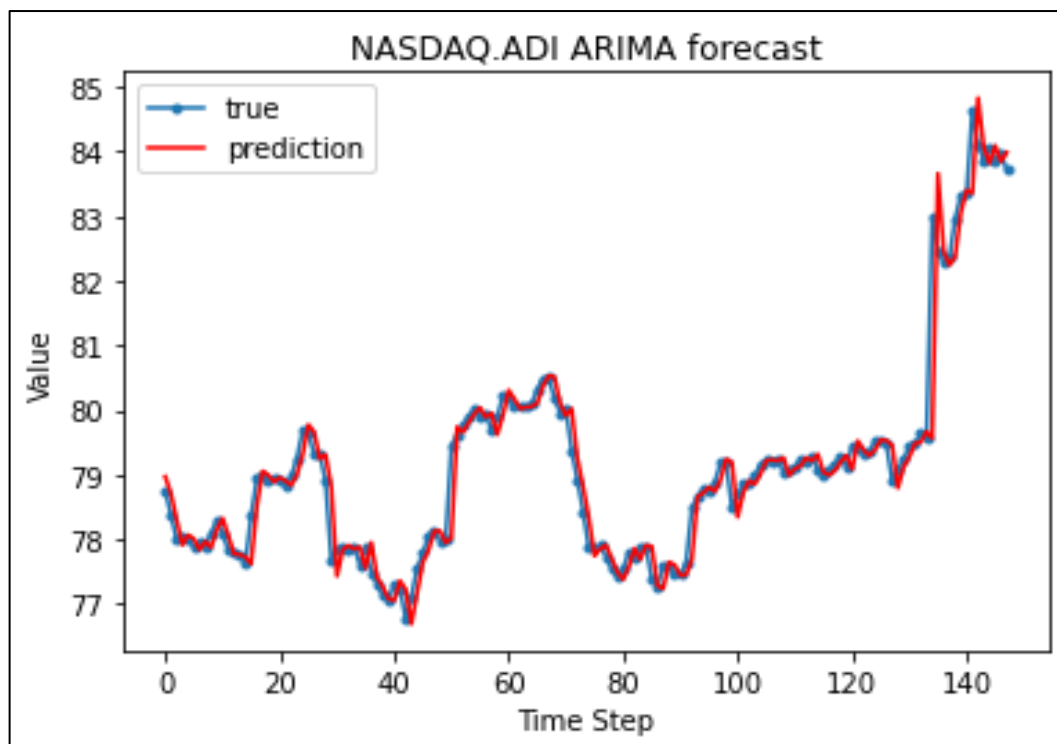
- SP500 –



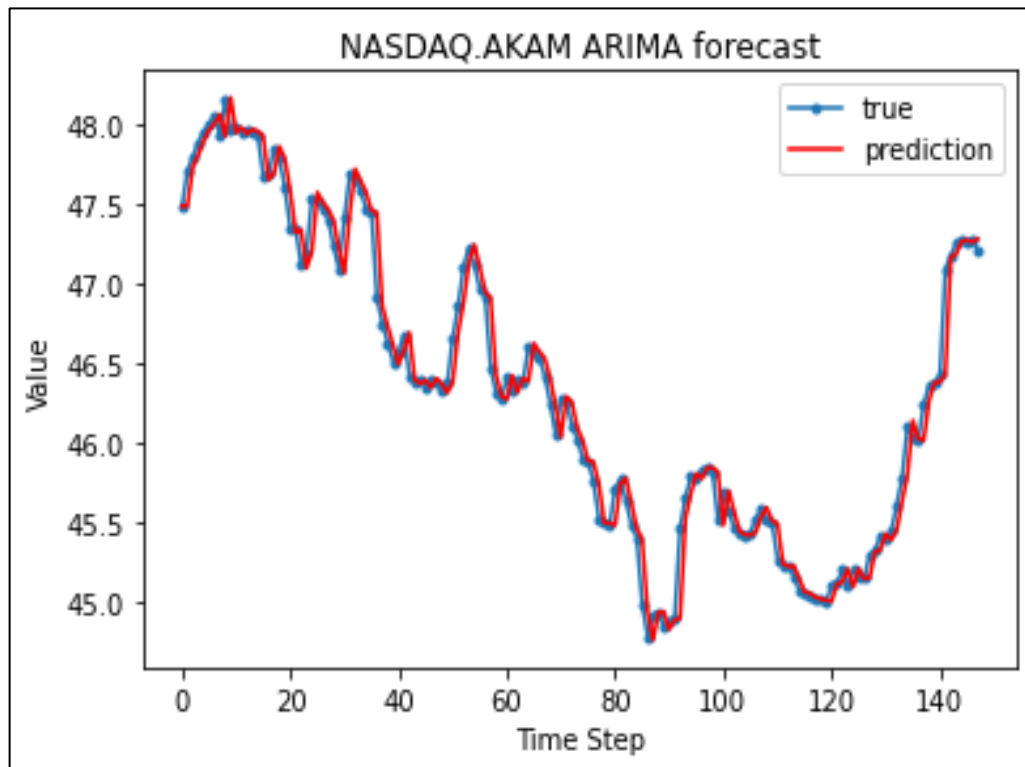
- NASDAQ.AAL –



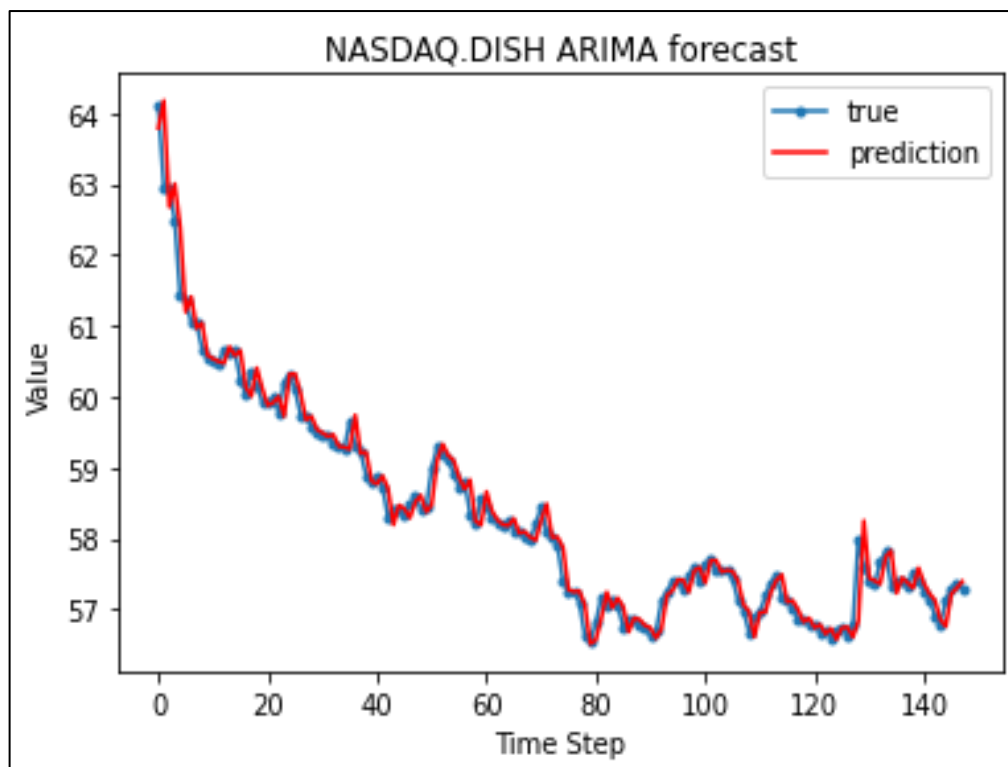
- NASDAQ.ADI –



- NASDAQ.AKAM –



- NASDAQ.DISH –



C. Performance metrics – RMSE

- RMSE gives higher weight to large prediction errors. In stocks, a large error could result in significant financial losses, so it is essential to pay attention to such errors.
- RMSE is naturally scaled to the range of the target variable. Since our stocks have different ranges, it can also help compare the errors across the stocks. It is also in the same unit as the target variable, making it more interpretable.
- A lower RMSE indicates that the model's predictions are closer to the actual stock prices. This means that the model is more accurate in forecasting future stock movements, which can be crucial for investors and traders.
- In a business context, deploying a model with a lower RMSE score in real-time trading or investment systems can lead to more successful execution of trading strategies and better returns on investments. Reducing errors in stock price predictions can lead to cost savings. Fewer trading losses and more successful investment decisions mean lower expenses and higher profits.

	LSTM	ARIMA
SP500	6.141523	3.836510
NASDAQ.AAL	0.590920	0.213643
NASDAQ.ADI	0.485054	0.427133
NASDAQ.AKAM	0.231004	0.155847
NASDAQ.DISH	0.303877	0.263540

- As per the RSME values, the ARIMA model is better at predicting the future stock prices across all five stocks. Visual analysis of the prediction line charts also leads to the same conclusion where the prediction lines given by the ARIMA model are much closely mapped as compared to the LSTM model.
- Both models are good at predicting the trends in stock prices increasing or decreasing and general shape of the stock data.