Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

A Hybrid Nowcasting Approach for Stock Price Prediction

Shubham Maurya, B.Tech.

A Dissertation

Presented to the University of Dublin, Trinity College

in partial fulfilment of the requirements for the degree of

Master of Science in Computer Science (Intelligent Systems)

Supervisor: Dr. John Mc Donagh

August 2022

Declaration

I, the undersigned, declare that this work has not previously been submitted as an exercise for a degree at this, or any other University, and that unless otherwise stated, is my own work.

Shubham Maurya

August 19, 2022

Permission to Lend and/or Copy

I, the undersigned, agree that Trinity College Library may lend or copy this thesis upon request.

Shubham Maurya

August 19, 2022

A Hybrid Nowcasting Approach for Stock Price Prediction

Shubham Maurya, Master of Science in Computer Science
University of Dublin, Trinity College, 2022

Supervisor: Dr. John Mc Donagh

Rapid advancements in financial services technologies have democratized participation in the stock markets around the world. With the increasing number of success stories, individuals' interest in stock market investment is increasing at an exponential rate. To bring some objectivity and certainty to decision-making, investors seek to predict stock prices. However, stock prices are influenced by a multitude of factors that make it very difficult to predict stock prices. Researchers have built numerous techniques and strategies that make use of various analyses and market indicators to predict stock prices. Market indicators are derived by quantifying crucial factors that are expected to affect stock prices, per fundamental and technical analyses. However, they are not very responsive to price changes in real-time. Researchers have used real-time indicators such as Google trends and sentiment analysis to solve this problem. Real-time indicators can capture real-time changes but are unable to forecast changes in future. None of the prediction techniques used hitherto combined fundamental analysis, technical analysis, and real-time indicators simultaneously to predict stock prices. Further, very few studies have been found useful to predict stock prices using high-frequency data. This study aims to fill that lacuna by combining technical analysis, fundamental analysis, and real-time indicators to predict stock prices with higher precision. The proposed Hybrid Nowcasting Model (HNM) has two main components: the predictor model and the Nowcasting model. The Predictor model is used to generate a leading indicator which is further used by the Nowcasting model to nowcast stock prices. The HNM model generates all the indicators internally and uses the prediction algorithm at its core. After rigorous testing and comparison of 5 state-of-the-art algorithms, Ridge regression was selected as the prediction algorithm for HNM model. The study also compares prediction performance using technical indicators, real-time indicators, and the proposed HNM model individually on a standalone basis. It is found that the proposed HNM Model has the lowest RMSE and MAPE. To test the robustness of the HNM model, it is evaluated on high-frequency (hourly) data, the results show better performance than state-of-the-art-model such as ARIMA and Prophet. The study was further generalized by testing on FTSE and S&P500 indices. The findings demonstrate that the HNM model outperforms ARIMA and Prophet model which means the study is not limited to the selected stocks but can also be used on any stock exchange, stocks from any industry, and for both high and medium frequency prediction.

Acknowledgments

I would like to express my sincere gratitude and thanks to my supervisors Dr. John McDonagh and second reader Dr. Susan Connolly from School of Computer Science and Statistics, Trinity College Dublin for their guidance, moral support and valuable feedback. I am thankful to Dr. Mohit Jain, Associate professor at School of Automation Banasthali Vidyapeeth, India for their constant encouragement and guidance. I am thankful to Mr. Prabhakar Yadav, MBA Student, Trinity Business School, for providing guidance on technical indicators, trading strategies, and general information on trading in stock markets.

Shubham Maurya

University of Dublin, Trinity College
August 2022

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Over the decade, people have seen a lot of potentials to profit in the equity or stock market. The Investors in the stock market are increasing at an exponential rate. Investing in stocks was always seen as a risky proposition but in recent years, the equities market's popularity has grown to the point that investors now view it as a reliable alternative. Investors attempt to forecast the desired stock's starting price, closing price, highest price, lowest price. Therefore, predictability of stock prices is a hot topic of discussion among researchers. It is believe that the market is stochastic and can not be predictable which leads to two very famous hypotheses: the random walk hypothesis and the efficient market hypothesis (1) (2). Despite the hypothesis, researchers believe that the stock price prediction will produce better results than random selection.

Everyone anticipates a rise in the stock market but stock prices are volatile and do not have constant variance (3). Stock prices depend on many factors such as market news, economic environment, industrial development, political policy and natural factors. Thus, It is quite difficult to estimate the stock price accurately. Numerous techniques for predicting stock prices have been developed by scientists and researchers. These approaches include fundamental analysis, technical analysis, and prediction algorithms, constitute the foundation of these methodologies (4).

Fundamental analysis considers various parameters such as company performance, dividends, earnings, top management changes, financial statements and economic factors. These parameters may influence the stock price and return (5). Returns are also determined by other factors that might be deduced from a financial statement. The stock market turnover ratio, price/earnings ratio, earnings per share, liquidation value, and market value ratio are a few of these measures (6) (4). Since the aforementioned metrics are designed to assess a company's overall performance and are released annually or quarterly, thus the users willing to make short-term investments may not be able to benefit from fundamental analysis' ability to forecast stock prices.

The technical analysis identifies short trends in stock price movement that may be used to forecast future stock prices. It is a group of indicators used to research stock price patterns using historical stock data (7) (8) (9). Some of the most commonly used indicators in technical analysis are Simple Moving Average (SMA), Exponential Moving Average (EMA), ATR, Keltner Channel (KC), Moving Average Convergence Divergence (MACD), and Relative Strength Index (4) (10). These indicators are widely utilized for predicting short-term trends. However, the quantitative statistics cannot fully reflect the diverse financial situation of enterprises (4). Technical indicators are lagging indicators; they are unable to reflect changes in the stock price that occur in real-time. Additionally, these indications are helpful for short-term investments, however, they may result in inaccurate predictions for long-term investments.

Researchers have created a number of methods that integrate technical and fundamental analysis, but these methods still lag to capture current trends. To solve this problem, researchers merged technical analysis with news sentiment to improve the accuracy of prediction model. With the advent of text mining and sentiment analysis techniques, these techniques gained popularity in recent years. However, there are several problems with text processing techniques that have an impact on the efficacy and effectiveness of the predictions. Another challenge with these techniques is handling multilingual text and there is a scarcity of tools that supports multiple languages (11).There is a need for an indicator that can foretell changes in stock prices because all of the approaches presented so far use lagging or real-time indicators.

Recent studies have developed various intelligent techniques and models such as fuzzy logic, Artificial Neural Networks (ANN), Machine learning models such as Linear regression, Support Vector Machines (SVM), the autoregressive integrated moving model (ARIMA), Prophet etc. These models use historical training data to find patterns and trends in the data. ANN is widely utilised by researchers for stock price prediction (12) (13). Since every stock has a different pattern, while training the ANN model it becomes very difficult to find an optimal value of the parameter. To optimize these parameters, various optimization algorithms such as the Genetic Algorithm (GA) (14), Particle Swarm Optimization (15), Owl Search Algorithm (16), Sine Cosine Algorithm (17) etc. have been proposed. These methods are combined to create new models for prediction (18).

All the above-discussed methods are based on historical analysis and lag in capturing the real-time changes in the stock price. Despite getting all the positive signals for all the analysis, the entire analysis may fail in unseen unpredictable conditions such as Covid 19 pandemic, war, recession etc. Thus real-time indicators play an important role in stock price prediction. In order to get real-time insights, the researchers have developed

Nowcasting techniques using google trends. The publicly available online search tool "Google trends" enables the user to see how frequently particular keywords have been searched within a specified period. Scholars have also found patterns and correlations between search trends and a stock price that can be used to find future trends (19) (20).

There are various state-of-the-art machine learning models and techniques that are used to forecast stock prices. The machine learning algorithm tries to find the patterns in the input data and draws the best fitting curve to forecast the stock price thus, it becomes very important to use an appropriate set of indicators for the analysis. Researchers have used various combinations of indicators real-time indicators or technical indicators but as per the best knowledge of the author, there was no evidence that all technical indicators, news sentiment analysis and google trends were used simultaneously to predict the stock price. Moreover, there is a scarcity of machine learning techniques for predicting the stock price on high-frequency (hourly) data.

In this work, a novel technique is proposed that utilizes Google Trends, and News Sentiments along with technical indicators to forecast the stock price. Furthermore, the leading indicator is generated using the predictor model of the proposed hybrid Nowcasting model. The proposed HNM architecture for Nowcasting the stock price uses all three types of indicators leading, lagging and real-time. It works on both low frequency (daily) and high frequency (hourly) data.

1.0.1 Research Question

- Does the proposed Nowcasting methods outperform conventional statistical forecasting methods for stock price prediction?
- Will the proposed combination of the indicators provide better results than individual indicators?
- Will the hybrid Nowcasting model (HNM) provide better performance for high-frequency (hourly) data?

To address the aforementioned research problems, the proposed HNM model is evaluated on 10 National Stock Exchange listed stocks in India. A rigorous comparative study is on HNM and the state-of-the-art models such as ARIMA and Prophet on all the considered 10 stocks. Additionally, a comparison between individuals and a proposed combination of indicators is performed. The results reveals that the suggested combinations of indicators have a lower root mean square and mean absolute percentage errors than the individual indicators. The Financial Times Stock Exchange (FSTE) and the Standard and Poor's 500 (S&P 500) index are also used to evaluate the reliability and effectiveness of proposed HNM. The model is further evaluated using high-frequency data and the comparative analysis of standard metrics depicts that the HNM model outperforms ARIMA and Prophet in terms of error.

The rest of the work is further structured as follows: Chapter 2 elaborates the literature review, the techniques employed in this research work are discussed in Chapter 3, Chapter 4 covers the detailed implementation of HNM, followed by result and discussion. finally the conclusion of the work is summarised in Chapter 6

Chapter 2

Related Work

There is a rich literature on time series analysis and its application in stock price prediction. There are various classical models utilized by scholars, some of these models are autoregressive conditional heteroscedasticity (ARCH), ARIMA, and GARCH models. The machine learning models Ridge Regression, Support Vector Machine, Support Vector Regressor, and KNN are widely utilized for stock movement and trend prediction. Recently, Researchers have proposed various methods based on Artificial Neural Network (ANN)(21) (20).

The author presented an automated investment tool based on Simple Moving Average (EMA) and Exponential Moving Average (EMA). The author demonstrates that the proposed system gives strong returns from the Indian and Russian markets using a buy and hold strategy (22).

In the paper (23), three models were used to forecast the direction of stock movement. They used a Random forest, Decision tree and Naive Bayesian Classifier with three macroeconomics and ten microeconomics variables. The microeconomic indicators are Weighted Moving Average (WMA), MACD, Accumulator/Distribution Oscillator Simple Moving Average (SMA), stochastic %D & %K , Larry Williams's % R and CCI. The findings of the experiment demonstrate that the decision tree model outperforms the random forest model and Naive Bayesian Classifier.

Stock prices are highly volatile in nature and it is difficult to predict the exact price of the stocks. Researchers & Scholars try to develop new tools and strategies to maximize the return using various combinations of technical Indicators and Intelligent techniques. A hybrid algorithm for forecasting stock prices based on Neural Network and Technical indicators is proposed. This amalgamation of technical analysis and Machine learning yields a better result than benchmark approaches do (24).

With the advancement in text mining techniques, scholars started to explore the effect of sentiment analysis of financial news on stock prices. Researchers used the frequency of and overall sentiments associated with the News feed along with technical analysis as an input feature for the ML algorithm (25). The results showed that the frequency and sentiment of news improved the performance. Recently, a hybrid model of was developed using a combination of sentiment analysis and deep learning neural networks to predict stock prices. The model is used to classify the investor's sentiment and performs better with sentiment analysis (26).

Nowcasting is useful in economics because the statistical analysis lags to capture the current situation of economic activity (27). Researchers have proposed various algorithms using real-time indicators. The author recently proposed an algorithm called Improved Sine Cosine Algorithm (ISCA) that uses google trends as a real-time indicator for forecasting the stock price movement. A comparative study was done with and without google trends. The results demonstrate that Google trends can help in Nowcasting the movement of the stock market (21).

The table 2.1 presents the recent work on stock price prediction using various combinations of intelligent techniques and technical analysis.

From the above literature review, we can conclude that sentiment analysis, google trends and technical indicators play a vital role in Nowcasting the stock prices. As per the best knowledge of the author and according to (28) there is a limited amount of research done on real-time data. Furthermore, there is no technique to date that uses technical indicators, sentiment analysis and google trends simultaneously for Nowcasting stock prices. This motivated us to build a hybrid Nowcasting technique that can be useful for both real-time as well as future stock price prediction

Table 2.1: Detailed Literature Survey

Author	Data	Technical Analysis	Prediction techniques	SA	GT	performance measure
Deng S. et al (25)	US stock	Rate of Change (ROC) MACD, BAIS	Multiple Kernel Learning Regression SVR	Yes	No	RMSE, MAE and MAPE
de Souza et al. (22)	BRICS markets	SMA, EMA	Fixed buy, sell strategy	No	No	percentage return
Jing N. et al. (26)	Shanghai Stock Exchange	SMA, EMA, MACD, Relative Strength Index Williams' %R, Momentum index Chande Momentum Oscillator Ultimate Oscillator,Volume indicators	CNN, LSTM	yes	No	Precision, Recall, F-measure
Hu H. et al. (21)	S&P 500 and Dow Jones	Opening price, Closing price, Highest price Lowest price, Trading volume	Improved Sine Cosine Algorithm ANN	No	Yes	Normalized mean squared error Root mean squared error Mean absolute error (MAE) Mutual information (MI)
Chavarnakul T. (10)	S&P 500	Equivolume charting Volume adjusted moving average	Generalized regression NN Fuzzy Logic Genetic Algorithm	No	No	RMSE, simulation testing
Vaiz JS. et al. (29)	NSE India	MACD, ADX, TDI, Aroon, VHF, SMA EMA, WMA, VWMA, DEMA, Stochastic RSI, SMI, WPR, CMO,CCI, Bbands ATR, Dochain Channel, OBV, MFI, CMF	Classification and regression trees Iterative Dichotomizer (ID3)	No	No	ROC curve, AUC
Norinder N. et al. (30)	US stocks (Google)	Highest price	MultiLayer Perceptron Random Forest Classifier	Yes	Yes	Mean precision Median precision
Mohanty S. et al.(31)	NYSE	Closing price	LSTM	No	Yes	RMSE
Wu S. et al.(32)	CITIS, BOC,ICBC,CCB ABC	Stochastic oscillator index (%K) William index (%R) Relative strength index (RSI)	CNN, LSTM	Yes	No	MAE, MSE, RMSE
Khan MA. et al.(28)	National Stock Exchange Bombay Stock Exchange	Price	Lasso Regression (LR) Ridge Regression (RR) Generalised Linear Model (GLM) Gradient Boosting Tree (GBT) Random Forest (RF)	No	No	Symmetric MAPE Directional Symmetry (DS), Theil's U Coefficien Diebold Mariano (DM)

Chapter 3

Methodology

We have collected the data from the National Stock Exchange (NSE) of India using the yfinance library, based on Yahoo finance API, in python. The methodology is divided into data gathering & processing, Feature Engineering & Market indicators, Prediction models and the architecture of the proposed Nowcasting Model.

3.1 Data Collection & Preprocessing

The lifecycle of a firm is broadly classified into three stages growing, mature and declining. These classifications are used to choose the top 10 NSE stocks. To generalize the investigation, hourly data of Adani power stock, the S&P 500 and FTSE 100 indexes are used. Table 3.1 shows the stock's name, scrip name, stage, date range and frequency of collected data.

Table 3.1: Stock dataset details

Stock Name	**SCRIP**	**STAGE**	**Frequency**	**Date Range**
Tata Consultancy services	TCS	Mature	Daily	01-12-2021 10-08-2022
HDFC bank	HDFCBANK	Mature	Daily	01-12-2021 10-08-2022
Bajaj Finance	BAJFINANCE	Growing	Daily	01-12-2021 10-08-2022
Vedanta	VEDL	Mature	Daily	01-12-2021 10-08-2022
Adani Total Gas	ATGL	Growing	Daily	01-12-2021 10-08-2022
Hindustan Unilever	HINDUNILVR	Mature	Daily	01-12-2021 10-08-2022
Engineers India	ENGINERSIN	Declining	Daily	01-12-2021 10-08-2022
Zomato	ZOMATO	Growing	Daily	01-12-2021 10-08-2022
Adani power	ADANIPOWER	Mature	Daily	01-12-2021 10-08-2022
Adani wilmare	AWL	Growing	Daily	01-12-2021 10-08-2022
Adani power	ADANIPOWER	Mature	Houly	01-07-2021 10-08-2022
FTSE 100 index	ˆFTSE	-	Daily	01-12-2021 10-08-2022
S&P500 index	ˆGSPC	-	Daily	01-12-2021 10-08-2022

Datasets are extracted using the yfinace library that is stored in the panda's dataframe.

The datasets are processed for discontinuities, breaks and missing values. The dataset consists of eight columns: "Date", "Dividends", "Stock Split", "High", "Low", "Close", "Open" and "Volume". The dataset is represented by a candle chart shown in Figure 3.1

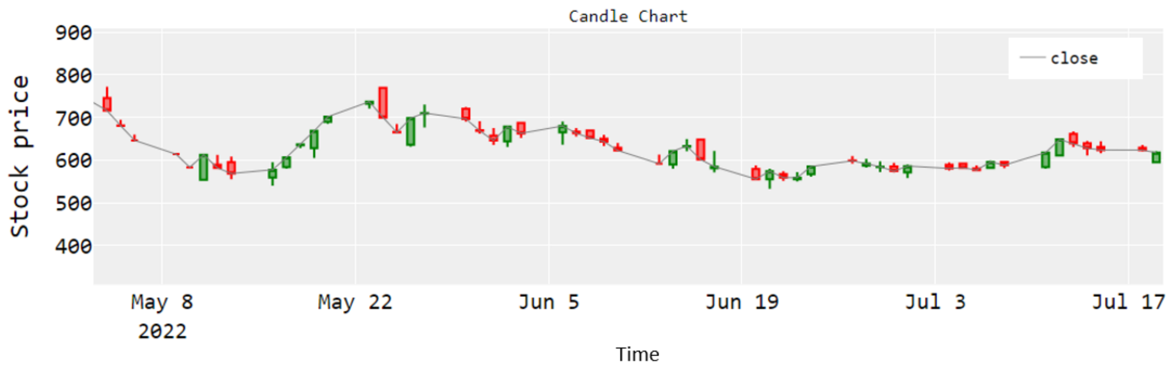


Figure 3.1: Candle chart of the dataset with closing line

Each candle depicts price action for the chosen time period; in our case, each candle represents price movements for one day. The highest price, lowest price, open price, and closing price are the four prices that make up a candle. The green candle indicates that the closing price was higher than the day's opening price, while the red candle indicates that the closing price was lower than the day's opening price.

The percentage return is calculated using the equation (3.1). The data at y(t) represents the current closing price and y(t-1) represents the previous closing price.

$$\text{return}(t) = \frac{y(t) - y(t-1)}{y(t-1)} * 100 \tag{3.1}$$

It is very important to find input data, also called as 'feature' in machine learning parlance, for prediction. The proper selection of input features can improve the prediction accuracy. Researchers have developed various ways to find these input features (33). The process of extracting features from input data is known as feature engineering.

3.2 Feature Engineering

Feature engineering is the practice of extracting data from datasets that helps in better prediction accuracy. One of the most often used feature types for time-series forecasting are Lag and rolling window features. These features try to capture the trend in the data. Depending on how frequently the datasets are updated, the study uses either an hourly lag or a daily lag. The hourly lag feature will be utilized for analysis if the dataset has hourly data, and the daily lag feature will be used if the dataset contains daily data.

Hourly lag features are used to capture the short trends in the data. hourly lag is taken at an interval of every 1 hour i.e. 'T-1Hour', 'T-2Hour', 'T-3Hour'. For Nowcasting the price of the stock at time y(T), the closing price of the stock at y(T-1Hour), y(T-2Hour) and y(T-3Hour) is considered.

Days lags are used to capture daily trends and patterns. The stock price at a time 'T' is given by the stock price at y(T-1Day), y(T-2Day) and y(T-3Day)

Rolling window features are generated using a 'K' size window, meaning 'k' number of previous consecutive time periods to determine the current value. For example, a 14-day Simple Moving Average (SMA), generally referred to as SMA(14), is calculated using 14 candlesticks immediately preceding the current day. Moving Averages are one of the most utilised rolling windows features.

Other factors like weekly and yearly features are not taken into account in this study. Since Nowcasting uses current or recent past data to forecast the near future and present. Thus, monthly and annual aspects are not taken into account.

3.3 Technical Analysis & Market Indicators

Technical analysis is used to analyze the trend, momentum and volatility of the stock price movement. To get these analyses SMA, EMA, MACD, Keltners Channel and ADX indicators are used as technical input features.

3.3.1 Simple Moving Average (SMA)

Moving averages are a crucial analytical technique for determining current price trends and the likelihood of change in trend. SMA apply equal weights to all data points and is calculated using equation (3.2). In this study, 5SMA and 22SMA are considered for the predictions which represent the moving average of no. of trading days in a week and in a month respectively.

$$\text{SMA} = \frac{y(t) + y(t-1) + ... + y(t - Days)}{Days} \tag{3.2}$$

3.3.2 Exponential Moving Average (EMA)

Exponential Moving averages are widely used by researchers for technical analysis. The most often referenced and studied short-term averages are the 12-period and 26-period EMAs. EMA uses a weighted mean that gives greater significance to recent points and is calculated using equation (3.3).

$$\text{EMA}(t) = y(t) * \frac{smoothing}{1 + Days} + EMA(t-1) * (1 - \frac{smoothing}{1 + Days}) \tag{3.3}$$

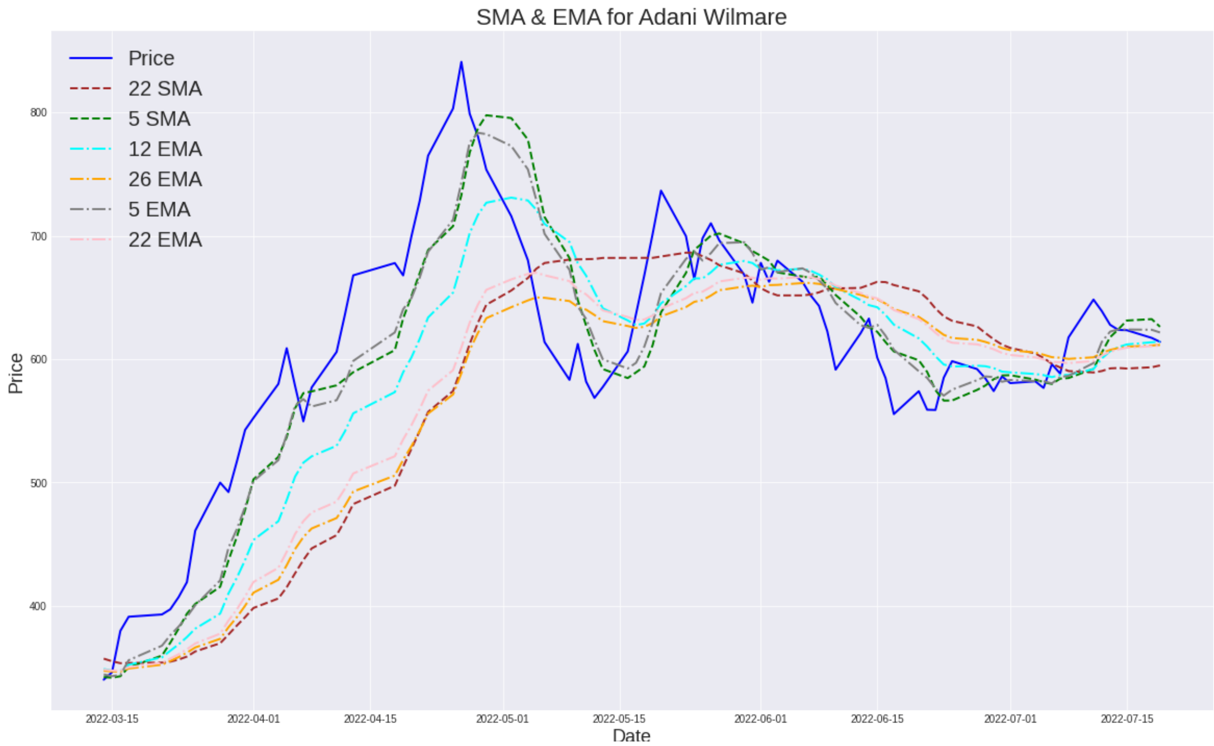


Figure 3.2: SMA & EMA of Adani Wilmar

In the above Figure 3.2, the blue colour line represents the closing price of the stock. The dash lines and dot lines represent SAM and EMA respectively. Both SMA and EMA are lagging behind the price line thus, these indicators are known as lagging indicators. The 5-period SMA captures small changes in price while the 22-period SMA smoothens the changes. For comparative analysis of EMA and SMA, the 5-period and 22-period EMA are plotted. It is observed that EMA are more sensitive to price and are quicker to reflect the changes in the stock price whereas the SMA are slow to respond.

3.3.3 Keltner's Channel

Keltner's Channel is volatility-based channel that is positioned on each side of the exponential moving average of stock price and can help determine the trend's direction. The Keltner channel has three parts upper band (kc_upper), EMA (kc) and lower band (kc_lower). These are calculated using the below equations.

$$\text{kc} = EMA \tag{3.4}$$

$$kc\_upper = EMA + multiplier * ATR \tag{3.5}$$

$$kc\_lower = EMA - multiplier * ATR \tag{3.6}$$

Where EMA is exponential moving average and ATR is Average True Range.

Nowadays, users have started using algorithmic trading (a.k.a Algo trading). Investors build various strategies to buy and sell stocks. The Keltner channel can be used to decide on entry and exit points. The upper band is used as an exit point and the lower band can be used as an entry point. The width of the channel depends on the value of the multiplier. For example, choosing 3 as the multiplier will place the upper band as well as the lower band 3x ATR away from the EMA line. Choosing too small a multiplier would generate too many signals compromising the trust in signals, while choosing too large a multiplier will impair the responsiveness of the tool, i.e. the signal will be generated too late for the user to open a trade. It is, therefore, recommended to choose a medium size multiplier, 3 being the most used value. In this study, 2 has been used as the multiplier because responsive is the most critical feature in Nowcasting. Any value lower than 2 would generate too many signals or noise, while values greater than 2 will not be responsive enough to be useful in Nowcasting. The impact of choosing the right multiplier on number of signals generated, and on the trade outcome of the user, is illustrated in Figure 3.3,3.4, &3.5.

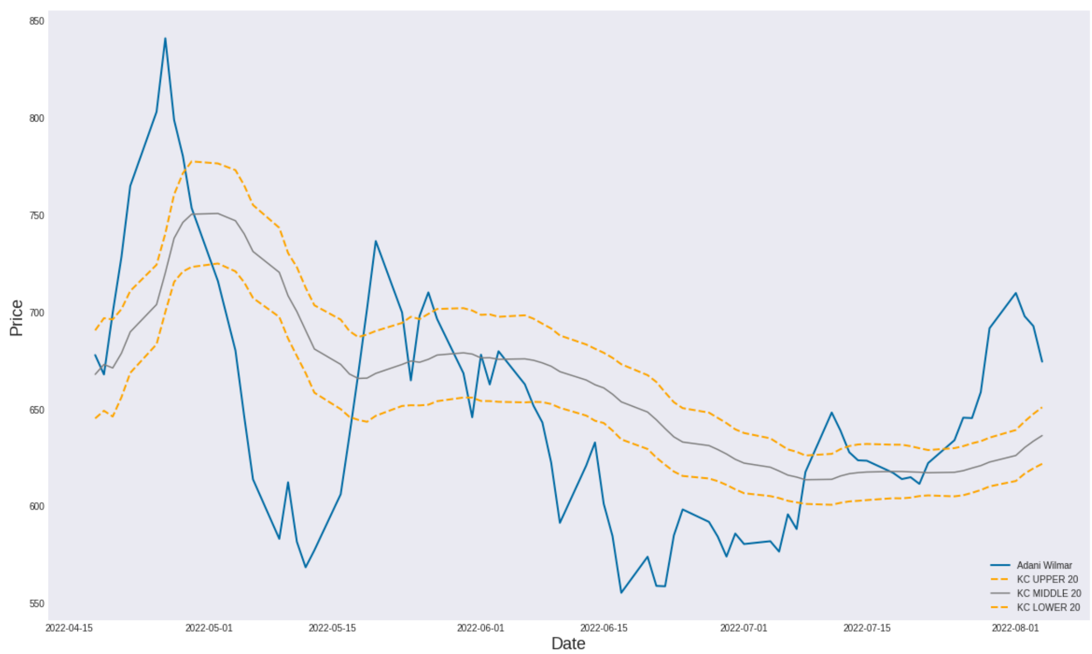


Figure 3.3: Keltner Channel with multiplier equal to 0.5 for Adani Wilmar

In the above Figure 3.3, there are too many entry and exit signals. Since every transaction has a trading cost therefore there will be a higher transaction cost because of too many transactions.

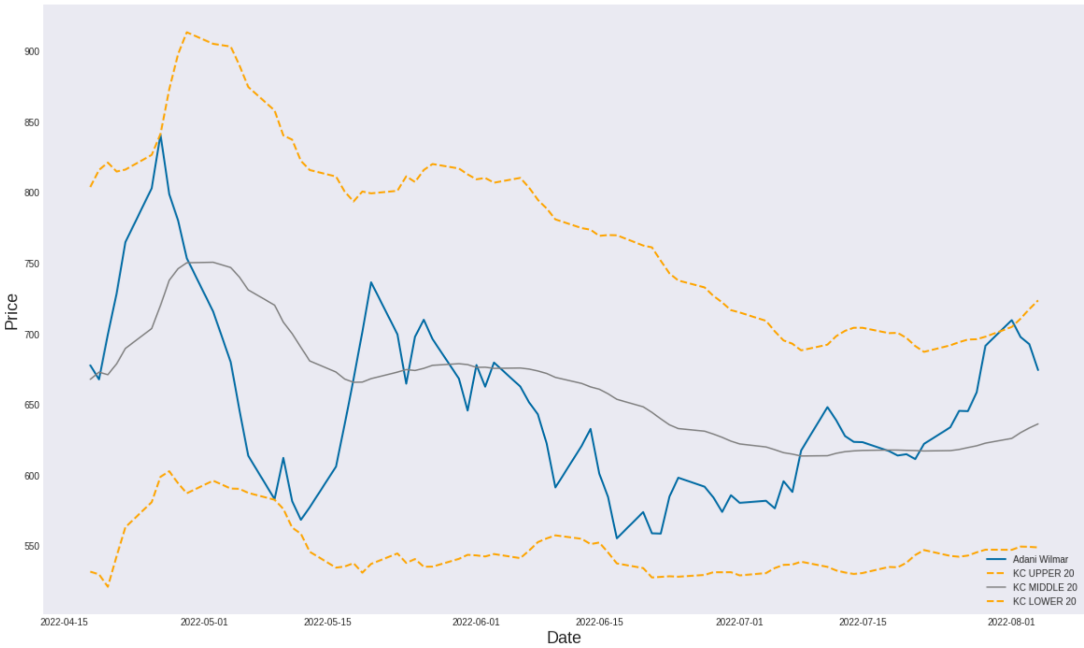


Figure 3.4: Keltner Channel with multiplier equal to 3 for Adani Wilmar

The higher value of multiple may return no entry or exit position. In this condition, the investment remains untouched. In Figure 3.4, there are no entry or exit signals due to the wider band. Thus the selection of the optimal value of the multiplier plays an important role in the overall entry and exit strategy.

The selection of multiplier value is an individual choice. If the optimum value of multiplier is used then the user can get proper entry or exit signals. The Figure 3.5 shows the Keltner's channel for Adani Wilmar with a multiplier value of 2

3.3.4 Moving Average Convergence Divergence (MACD)

Moving Average Convergence Divergence is used to find the points where trends might be accelerating. MACD signal is calculated by taking the difference between fast and slow EMA. The most common value of EMA used are 12-period EMA and 26-period EMA. MACD helps investors to understand whether the price movement is getting stronger or weaker. It is represented by two moving averages and the difference between these two lines is represented by the histogram.

In the Figure 3.6, there are two sub-graphs, first presents a candle chart of stock price. The second graph show MACD (macd), Signal (macd_s) and histogram (macd_h). The

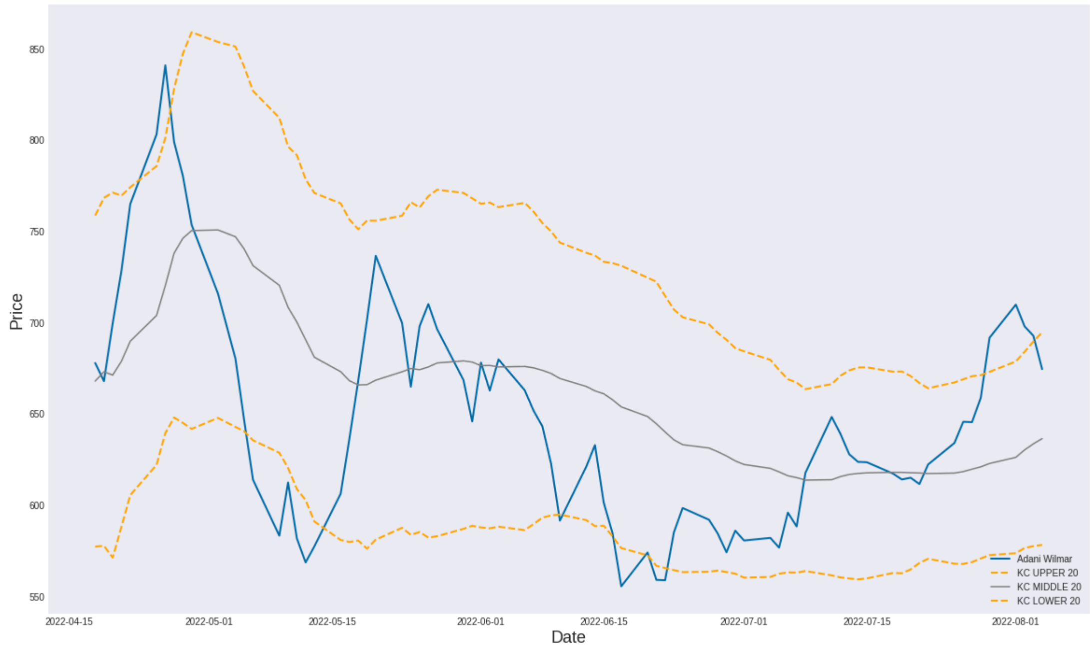


Figure 3.5: Keltner Channel with multiplier equal to 2 for Adani Wilmar

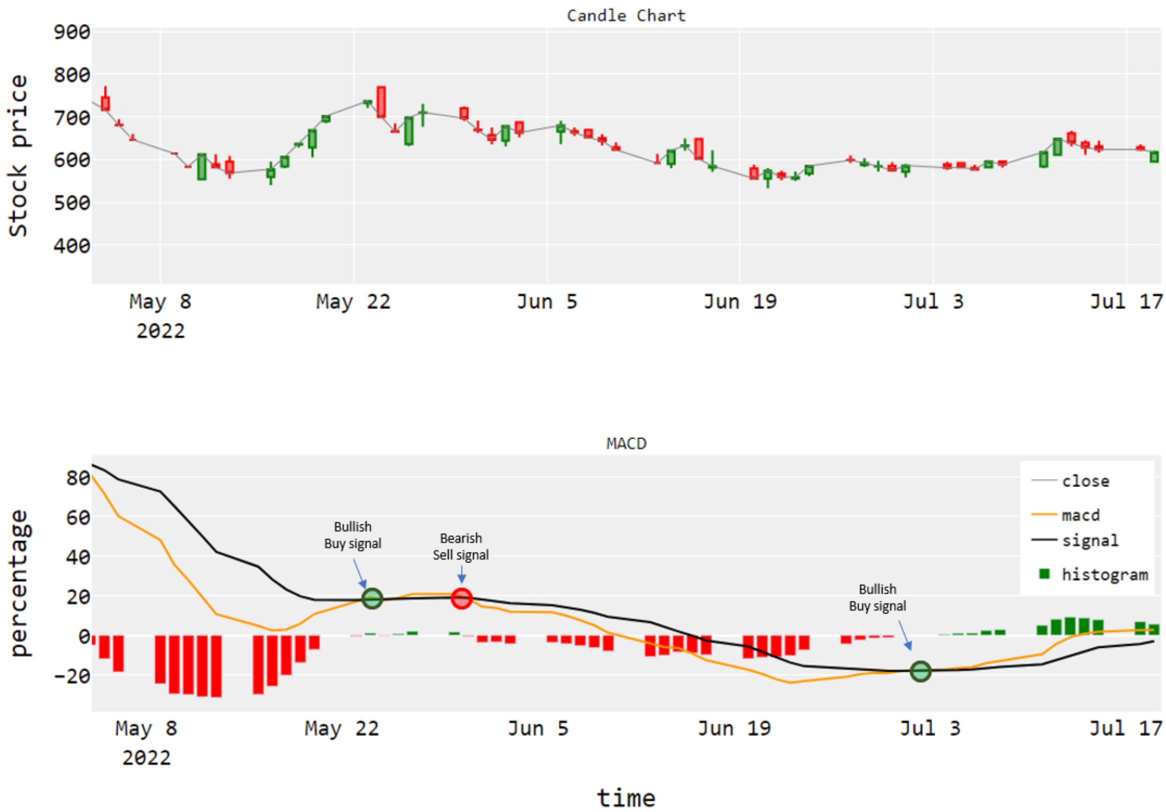


Figure 3.6: MACD Indicator for Adani Wilmar

candle chart is prepared by using open price, close price, and maximum & minimum value of the stock. The grey line in the candle graph represents the closing price of the stock.

In the MACD graph, the orange line represents "macd" which is calculated using the difference between 12 and 26-periods EMA. The black Signal line is 9-period EMA that is used to calculate the convergence and divergence.

The sentiments in the stock market analysis are referred to as 'bullish' and 'bearing'. Bullish sentiment indicates optimism i.e. expectation that the prices will go up whereas bearish sentiment signifies pessimism i.e. expectations of market downturn. Generally, the bullish signal means to buy the stock or enter the market and the bearish means it is the time to sell the stock or exit the market. In the above MACD chart, the signal line crosses the MACD line at two points. When the Signal line rises above the MACD line, it generates a bullish signal. Conversely, when it falls below the MACD line, it generates a bearish signal. Examples of both bullish and bearish signals are represented by green and red circles respectively in the Figure 3.6.

3.3.5 Average Directional Index (ADX)

The average directional index is used to find the strength of a trend. It has three components ADX, negative directional indicator (-DI) and positive directional indicator (+DI). The ADX, positive direction indicator and negative direction indicator can be calculated using the below equations. The ADX helps determine the strength of the trend, while +DI and -DI represent the positive and negative trend direction respectively. These components are calculated using the following equations.

$$\text{+DI} = \left(\frac{Smoothed + DM}{ATR}\right) * 100 \tag{3.7}$$

$$\text{-DI} = \left(\frac{Smoothed - DM}{ATR}\right) * 100 \tag{3.8}$$

$$\text{DX} = \left(\frac{|+DI - -DI|}{|+DI + -DI|}\right) * 100 \tag{3.9}$$

$$\text{ADX} = \frac{(PriorADX * 13) + CurrentADX}{14} \tag{3.10}$$

where +DM is the difference of current high and previous high, -DM is the difference of previous low and current low and ATR is Average True Range.

Various researchers and scholars have observed that when ADX crosses 25 and moves in an upward direction then the current market trend, whether bullish or bearish, has gathered strength and may be expected to continue in the short term. Conversely when ADX falls below 20, it is construed to indicate consolidation in the market(34). ADX is not a directional indicator which means it does not provide any information about the direction of stock price movement but it tells us about the strength of the trend. Thus, If ADX is growing up does not mean that market is going up conversely if ADX is falling that does not mean the market is falling.

Various strategies are built around ADX using +DI and -DI to get a buy or sell signal. Investors get a buy signal if the +DI line crosses the -DI line and the ADX is above 25. On the other side, there is a chance to make a potential short trade if the +DI crosses below the -DI and the ADX line is above 25. The ADX for Adani wilmar is shown in Figure 3.7

Figure 3.7: ADX Indicator for Adani Wilmar

All the above-discussed indicators are lagging indicators which means they have some delay to capture the changes in stock price. Therefore, employing these indicators to forecast stock prices carries the possibility of producing inaccurate results, as well as financial loss. To overcome this issue, real-time indicators such as google trends and news sentiments are used for Nowcasting the stock prices.

3.3.6 Google Trends (GT)

Google Trends is one of the real-time indicators utilized by various researchers for Nowcasting applications. Google Trends is a percentage volumetric indicator of queries entered by users on google. In order to get daily or weekly search trends, user types their query in the search bar of the Google Trends dashboard shown below 3.8.

Google trends provide multiple options for customization such as country, duration, category and search type. The country menu is used to select a specific country, for instance, if the user wants to do the analysis for a specific country he/she can select it from the country drop-down menu. Similarly, the duration menu is used to customize the timeframe of the analysis. Users can choose the duration from the list of options such as past hour, past 4 hours etc. or they can enter a custom time range. Google Trends also

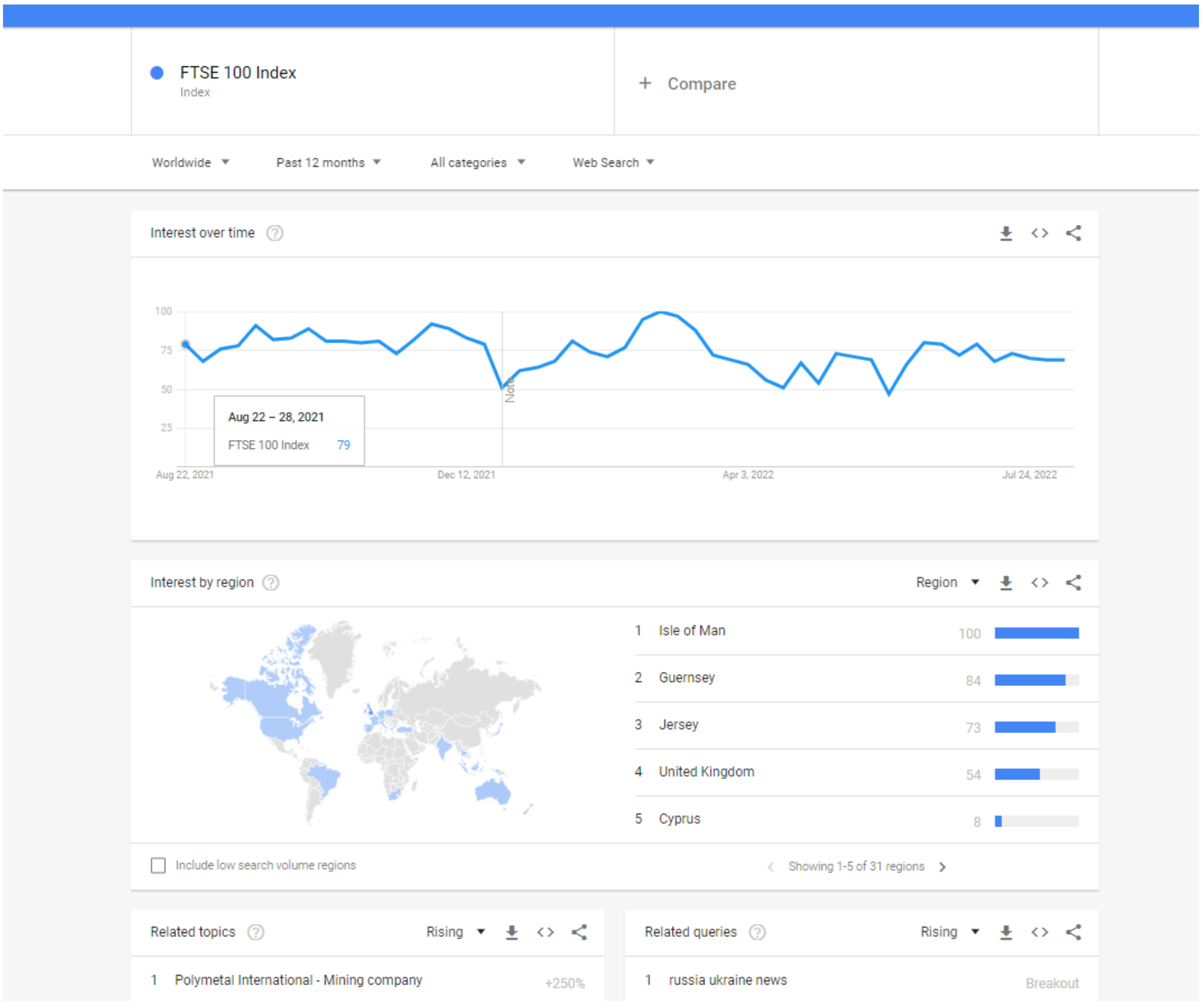


Figure 3.8: Google Trends Dashboard

provides an option to select a search category from a list of categories such as finance, entertainment, News etc. The search type is used to select the type of the search for example Image search, YouTube search, news search, web search and google shopping.

Google Trends provide an Interest over time graph, Interest by region, Related topics and Related queries. Interest over time is a line chart where the x-axis represents the timestamp and the y-axis represent search interest relative to the highest point on the chart for the given region and time. Interest by region provides a map of search Interest and the list of top five regions. Further, the Related topics and Related queries provide a piece of information about related topics and queries. There are other Analytics such as Trending searches, and real-time search trends that may be useful for Nowcasting applications.

Researchers and scholars have studied these trends with the stock price. Google Trends have a correlation with the stock prices and it may be used for Nowcasting the stock market (19).

In this study, pytrends, a python library, has been used to fetch the search trends. The query is made using Stock Name for instance, if the name of the stock is google then the query is "Google". The output of this query is a dataframe, that has two columns timestamp and percentage search. Since the Stock market remains close on Saturday and Sunday thus dataset needs to be preprocessed before using it as a real-time indicator for Nowcasting the stock prices.

To ensure that all the indicators have the same timestamp, the weekdays and list of all holidays are deleted from the dataset during preparation. Additionally, hourly data for a specified date range is the maximum frequency of data that can be obtained using the Pytrends module in python.

3.3.7 Sentiment Analysis (SA)

Stock prices depend on various factors such as a change in top management, financial statements, recession, pandemic situation, natural disasters etc. The historical data or technical indicator will not be able to capture these unseen conditions. It is likely that despite getting all the positive signals for a bullish market from indicators market may crash or fall. News can be used as a real-time indicator that captures all the information about the stock market. Since this news will be in the form of text, in order to understand the sentiment and context of the text Sentiment Analysis is performed on the text.

Sentiment analysis is the process of identifying the sentiment of a text and broadly categorizing them into positive, negative and neutral categories (35). Sentiment Analysis is a part of Natural Language processing. The basics of Sentiment Analysis make use of word corpus, a list of words, and every word has some sentiment. Based on the frequency of occurrence of the word, the overall sentiment is estimated. With the advancement in text mining techniques and deep learning networks, new models were built that not only take the frequency of words into account but also try to get the context of a sentence.

In this work, The NewsAPI Client is used to get the English news of the last 60 days. The API takes query, language and API key as input and returns a dictionary of status, total no of results and articles. The results contain various information such as the author, content, description, publishedAt, source, title and URL. For Sentiment Analysis, publishedAT, author and title of the news are considered Figure 3.9

The researchers have developed various tools for sentiment analysis. One of them is the distilRoberta-financial-sentiment transformer-based model provided by hugging face. It is fine-tuned model of distilroberta-base trained on the financial_phrasebank dataset. The model has an accuracy of 98.23%. The model can be used using a few lines of code

index	publishedAt	Author	title
0	2022-07-18T14:03:00Z	Business Desk	Fortune Oil To Be Cheaper; Adani Wilmar Cuts Edible Oil Prices by Up to Rs 30 Per Litre - News18
1	2022-08-03T03:59:30Z	Rakesh Patil	Market LIVE Updates: Sensex, Nifty trade flat; Voltas, Zomato, eClerx Services in focus - Moneycontrol
2	2022-08-03T12:18:14Z	Vikas Srivastava	Adani Power Q1 Results: Revenue Doubles, Higher Imported Coal Prices Impact Operational Performance ...
3	2022-08-04T04:53:21Z	Rakesh Patil	Market LIVE Updates: Indices trade higher with Nifty above 17,400; IT, metals shine; Zomato most active ...
4	2022-07-18T08:05:10Z	PTI	Adani Wilmar cuts prices of edible oil by up to Rs 30 per litre - Moneycontrol
5	2022-07-31T03:46:27Z	Sheetal Bhalerao	Adani Group Plans This New IPO; Plans To Raise $188 Million
6	2022-07-18T10:33:00Z	Research and Markets	Castor Oil Global Market Outlook Report 2022-2026: Mounting Utilization of Biodiesel Feedstock
7	2022-08-02T15:21:39Z	None	India Edible Rice Bran Oil, Mustard Oil & Groundnut Oil Market Outlook Report 2022-2025 - ResearchAndMarkets.com
8	2022-07-19T10:52:02Z	Moneycontrol News	These stocks will be either winners or losers due to rupee weakening - Moneycontrol
9	2022-08-03T10:53:55Z	Nikhil Agarwal	Adani Wilmar Q1 Results: Profit rises 10% YoY to Rs 194 crore; revenue up 30%
10	2022-07-18T08:23:23Z	PTI	Prices of Adani Wilmar's Fortune edible oil cut by Rs 30 per litre

Figure 3.9: News Dataset for Sentiment Analysis

in python. It returns the sentiment score for positive, negative and neutral sentiment. Further, The overall sentiment of the news is given the emotion with the highest value. The emotion is categorized into three groups: neutral (value = 0), positive (value = +1), and negative (value = -1).

The real-time indicators can capture real-time changes in the stock market but to maximize the gain from the stock market, there is a need for a leading indicator that can forecast the changes in stock prices. The output of the best prediction algorithm is used as a leading indicator. To find a leading indicator, five cutting-edge prediction algorithms are utilized in the study.

3.4 Prediction Algorithm

Machine learning (ML) is a subset of Artificial Intelligence that uses historical data and provides predictions. These algorithms try to find the pattern in the data and draw a correlation to find the best fitting line, curve or plane. Researchers and Scholars have developed various algorithms for Predictions. In this thesis, Ridge Regression, K-Nearest Neighbour (KNN), Support Vector Regression (SVR), ARIMA and Prophet have been used for Nowcasting the stock price.

3.4.1 Ridge Regression

Model over-fitting is one of the frequent issues with linear regression. Ridge regression is used to apply a penalty to the cost function and penalize the excessive coefficient values. It uses linear least squares with L2 Regularization. The objective function is given by the equation (3.11)

$$\mathrm{L} = ||y - Xw||_2^2 + alpha * ||w||_2^2 \tag{3.11}$$

where alpha is the hyper parameter that is used to handle the penalty. The optimal value of alpha is estimated using Sklearn's GridSerachCV over a range of alpha values from 0.001 to 1000. It uses cross-validation to find the optimal value of alpha.

3.4.2 K-Nearest Neighbours (KNN)

Non-linearity is one of the most significant issues with time-series data. To use Linear models one of the assumptions is that there should be no non-linearity. A KNN is a non-linear classifier that captures the non-linearity contained in the dataset. KNN is one of the popular algorithms used in regression and it is applied in various time-series analyses (36).

KNN uses the k neighbours to predict the stock price, these neighbours are selected using euclidean distance with uniform weights. The K is the hyper parameter, and the optimal value of K is obtained using sklearn's GridSearchCV over a range of values from 2 to 25.

3.4.3 Support Vector Regressor (SVR)

A Support Vector Regressor (SVR) works on the same principle as the Support Vector Machine (SVM). The SVM finds the best fit line or plane that best classifies the data. Figure 3.10 A shows the two labelled classes. These two classes can be separated by a line shown in Figure 3.10 B.

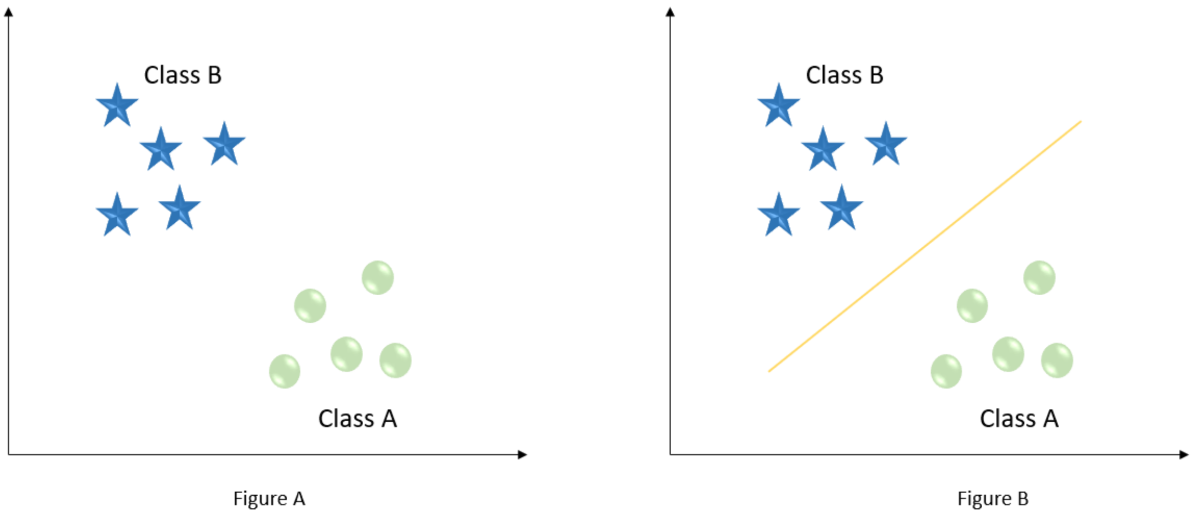


Figure 3.10: Data classification with SVR

Now consider another case shown in Figure 3.11 A, now these two classes cannot be separated by a straight line. A new dimension can be introduced; for instance, z-axis, now it can be separated by a line shown in Figure 3.11 B. While transforming back to the initial plane we get a circular boundary this is how SVM works. It finds the best fitting line, curve or plane. Similarly, SVR has two decision boundaries and tries to find the best fitting line on the hyperplane that has the maximum number of data points. The distance between the decision boundary and hyperplane is given by epsilon. It is a hyper-parameter whose optimum value is calculated using 5-fold cross-validation and Root Mean Square Error (RMSE). The SVR model can be imported from the sklearn package in python. It can detect the non-linearity in data and may provide a proficient Nowcasting model.

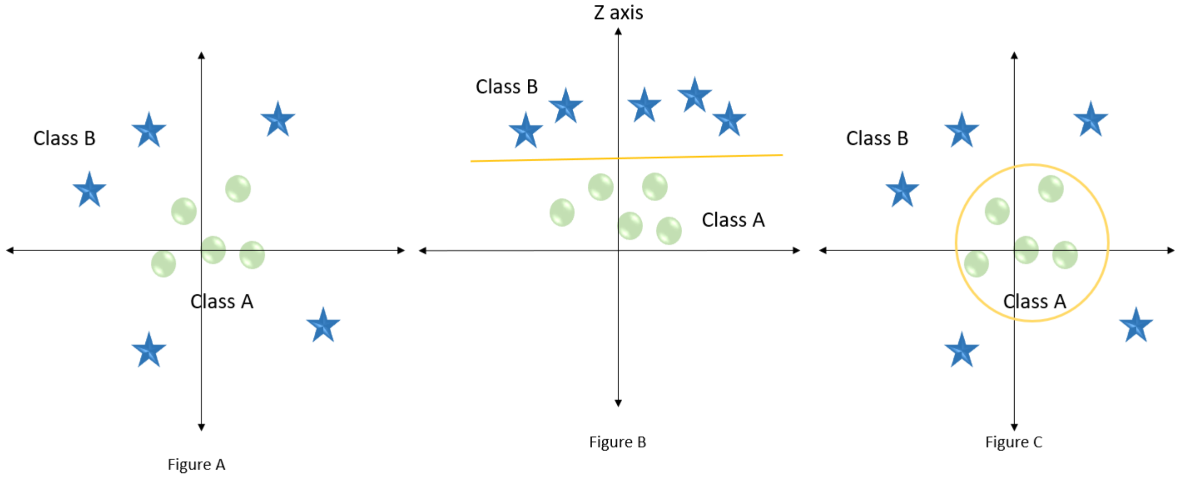


Figure 3.11: Data classification with SVR

3.4.4 Autoregressive Integrated Moving Average (ARIMA)

It is one of the state-of-the-art statistical models for forecasting time series data. ARIMA stands for Autoregressive Integrated Moving Averages. It is an amalgamation of Autoregression (AR) and Moving Average (MA). AR means that the model uses the relationship between current and past or lagged values. Integrated is used to make the data stationary by taking the difference between current observation and previous observation. The moving average model uses a relationship between a lagged observation and residual error.

The widely utilised notation ARIMA(p,d,q) is used in python. Where 'p' is lag order or a number of lag observations, 'd' is the degree of differencing and 'q' is the order of MA. These parameters can take integer values depending on the values ARIMA model behaviour will change. For example, if d=0 is zero the model will become Autoregressive Moving Average (ARMA). If the value of d and q is zero that means the ARIMA model will simply act as an Autoregressive model. Similarly, if p and d are zero then the ARIMA model will act as the Moving Average Model.

There are other names of the ARIMA model based on the values of p,q and d. The ARIMA(1,0,0) is also known as the first-order AR model, where the order is represented by the value of 'q'. ARIMA(0,1,0) is known as the random walk model. ARIMA(1,1,0) is known by the name differenced first-order autoregressive model similarly the name will change with values of p,d and d.

The selection of optimum values p,q, and d can enhance the accuracy of the model. In the current work, to get the optimum value of p,q, and d the auto ARIMA model is used. The 'pmdarima' library is used to get the auto ARIMA model in python. Auto ARIMA uses various combinations of values for p,q, and d to estimate the AIC score. The model with minimum AIC is selected as shown in Figure 3.12.

```
Performing stepwise search to minimize aic
 ARIMA(0,1,0)(0,0,0)[0] intercept   : AIC=2638.512, Time=0.11 sec
 ARIMA(1,1,0)(0,0,0)[0] intercept   : AIC=2637.518, Time=0.07 sec
 ARIMA(0,1,1)(0,0,0)[0] intercept   : AIC=2637.301, Time=0.14 sec
 ARIMA(0,1,0)(0,0,0)[0]             : AIC=2638.175, Time=0.04 sec
 ARIMA(1,1,1)(0,0,0)[0] intercept   : AIC=2638.036, Time=0.18 sec
 ARIMA(0,1,2)(0,0,0)[0] intercept   : AIC=2638.003, Time=0.18 sec
 ARIMA(1,1,2)(0,0,0)[0] intercept   : AIC=2640.007, Time=0.22 sec
 ARIMA(0,1,1)(0,0,0)[0]             : AIC=2637.157, Time=0.05 sec
 ARIMA(1,1,1)(0,0,0)[0]             : AIC=2638.078, Time=0.12 sec
 ARIMA(0,1,2)(0,0,0)[0]             : AIC=2637.988, Time=0.07 sec
 ARIMA(1,1,0)(0,0,0)[0]             : AIC=2637.351, Time=0.04 sec
 ARIMA(1,1,2)(0,0,0)[0]             : AIC=2639.942, Time=0.34 sec

Best model:  ARIMA(0,1,1)(0,0,0)[0]
Total fit time: 1.591 seconds
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                 1158
Model:               SARIMAX(0, 1, 1)   Log Likelihood               -1316.578
Date:                Tue, 16 Aug 2022   AIC                           2637.157
Time:                        12:59:04   BIC                           2647.264
Sample:                             0   HQIC                          2640.971
                               - 1158
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ma.L1         -0.0528      0.020     -2.626      0.009      -0.092      -0.013
sigma2         0.5700      0.010     54.657      0.000       0.550       0.590
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):              3760.36
Prob(Q):                              1.00   Prob(JB):                         0.00
Heteroskedasticity (H):               0.71   Skew:                             1.28
Prob(H) (two-sided):                  0.00   Kurtosis:                        11.45
===================================================================================
```

Figure 3.12: The output of Auto ARIMA

The outcome of the auto ARIMA model is used in the ARIMA model to forecast the stock price.

3.4.5 Prophet

The Prophet is an open-source forecasting tool provided by Facebook. It can be used in python or R. It is an automatic tool that can be used for forecasting. The Prophet model can be used to identify and analyze hourly, daily, and weakly trends, seasonality, holidays, missing values, historical trends, outliers, non-linear growth etc.

Prophet is designed in such a way that a person with zero knowledge of time series methods will not only utilize the tool but also can tune the hyper-parameters of the model using the easy-to-understand parameters. Prophet internally uses various state-of-the-art models such as ARIMA, exponential smoothing etc and finds the optimum value for the hyper-parameters thus it selects the optimal model for the given dataset.

Prophet uses a non-parametric regression method also known as an additive model (AM) with four main components: linear or logistic growth trend, Fourier series for modelling yearly seasonality, Dummy variables for weakly seasonality, and a list of holidays.

The Figure 3.13 is automatically generated that shows trends and seasonality in the data. The 'prophet' package is used to access the Prophet model in Python. The model takes two inputs ds and y. In this study, dates are mapped to ds and the closing price of the Adani power stock is mapped to y. The following graphs are automatically generated by the model.

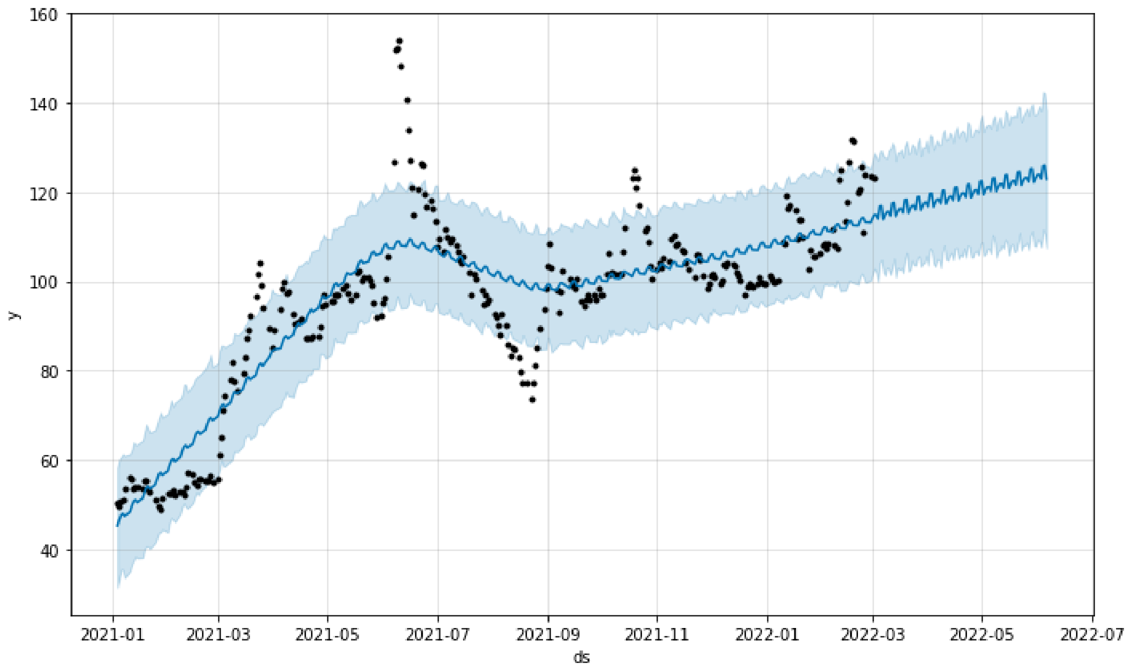


Figure 3.13: Plots generated by Prophet model

In the Figure 3.13, the x-axis shows the date and the y-axis shows the closing price of the stock. The black dot shows the actual data points i.e the closing price of the stock on that day. The blue line is the prediction that is represented by the 'yhat' in the data. Prophet model also returns various parameters such as ds, trends, yhat_lower, yhat_upper, trends_upper, additive_terms, additive_terms_lower, additive_terms_upper, weekly, weekly_lower, weekly_upper, multiplicative_terms, multiplicative_terms_lower, multiplicative_terms_upper and yhat. The shaded region in the Figure3.13 is plotted using yhat_upper and yhat_lower. yhat_upper is the upper bound of the prediction whereas the yhat_lower is the lower bound of the prediction.

The Figure 3.14 shows the day-wise seasonality in data. It can be observed that the stock price remains unchanged during weekends since the stock market does not operate on weekends. from Figure 3.14 it is observed that the stock price starts to increase from Monday to Wednesday and achieves its maximum on Wednesday than the price start to drop.

As shown in the literature review in Table 2.1, researchers and scholars have used various combinations of features such as technical indicators, sentiment analysis or google

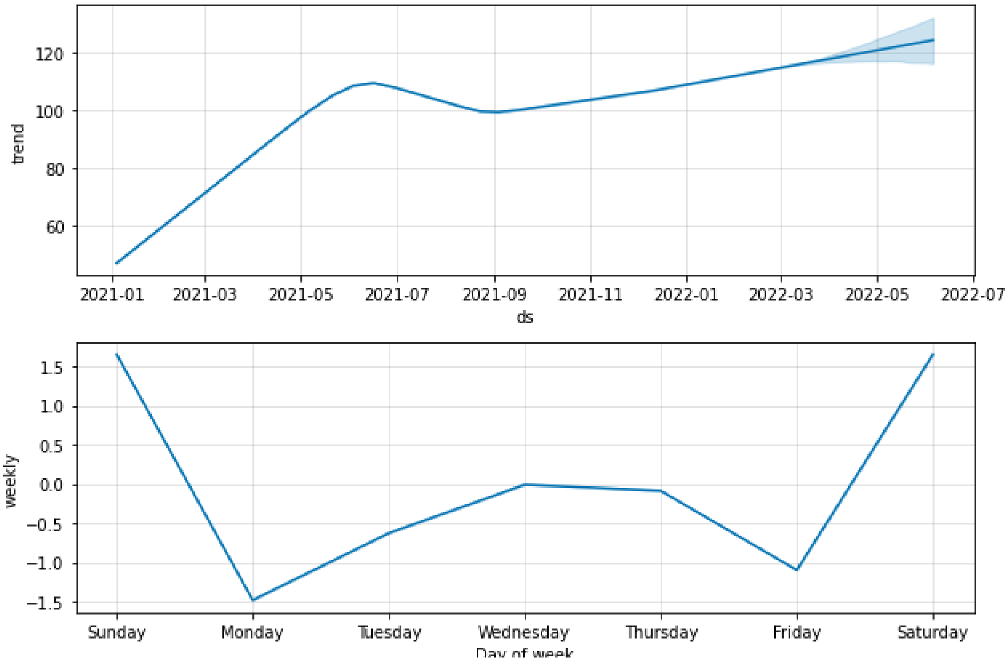


Figure 3.14: Trends and Seasonality graph - Prophet

trends but as per the best knowledge of the author, there is no evidence that all the discussed parameters are used into the same model simultaneously. Furthermore, there is a scarcity of techniques that can be used for Nowcasting or real-time prediction of stock prices. Thus, A hybrid model for Nowcasting the stock price is proposed.

3.4.6 A hybrid model for Nowcasting the stock prices (HNM)

As per the literature review, the combination of all features discussed in the previous section is never used for Nowcasting the stock price. Since indicators can be classified into three types: lagging indicators, real-time indicators, and leading indicators. The proposed Architecture (HNM) calculates all the indicator internally using stock data, google trends, and news data. The HNM model has two parts, the predictor and the Nowcasting model as shown in Figure 3.15

The predictor model is used to forecast the stock price based on lag features, moving averages and the closing price of the stock. The output of the predictor (leading indicator) is fed to the Nowcasting model along with all the indicators discussed in the previous section.

The prediction model shown in architecture uses Ridge Regression, KNN, SVR, ARIMA and Prophet model discussed in the previous section. The accuracy of these models is evaluated using the Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The model with the least RMSE and MAPE is used for prediction. Further, the results of these predictions will be taken as input to the Nowcasting model shown in Figure 3.15.

The Nowcasting model takes various types of input features such as lagging indicators, real-time indicators and a leading indicator (predictions). The lagging indicators are simple moving average, closing price, exponential moving average, Keltner channel, ADX and MACD. Real time indicators are Google Trends and News Sentiments Score.

Keltner channel has three components kc_middle, kc_upper and kc_lower. Similarly, MACD has three components macd, macd_h (Histogram) and macd_s (Signal). ADX has plus_di, adx and minus_di that are used by Nowcasting Model. These are lagging indicators that are used in the Nowcasting model.

Real-time indicators such as google trends and news sentiment are calculated based on the name of the stock. Google trends will get the data based on the query "Name of the stock" and the sentiment of the financial news is used as an input to the Nowcasting model.

The Nowcasting model uses Ridge regression, KNN, Support Vector Regressor, ARIMA and Prophet model. The dataset is divided into the test and train datasets in order to assess these models. These models are tested on the test dataset and their performance is evaluated by RMSE and MAPE.

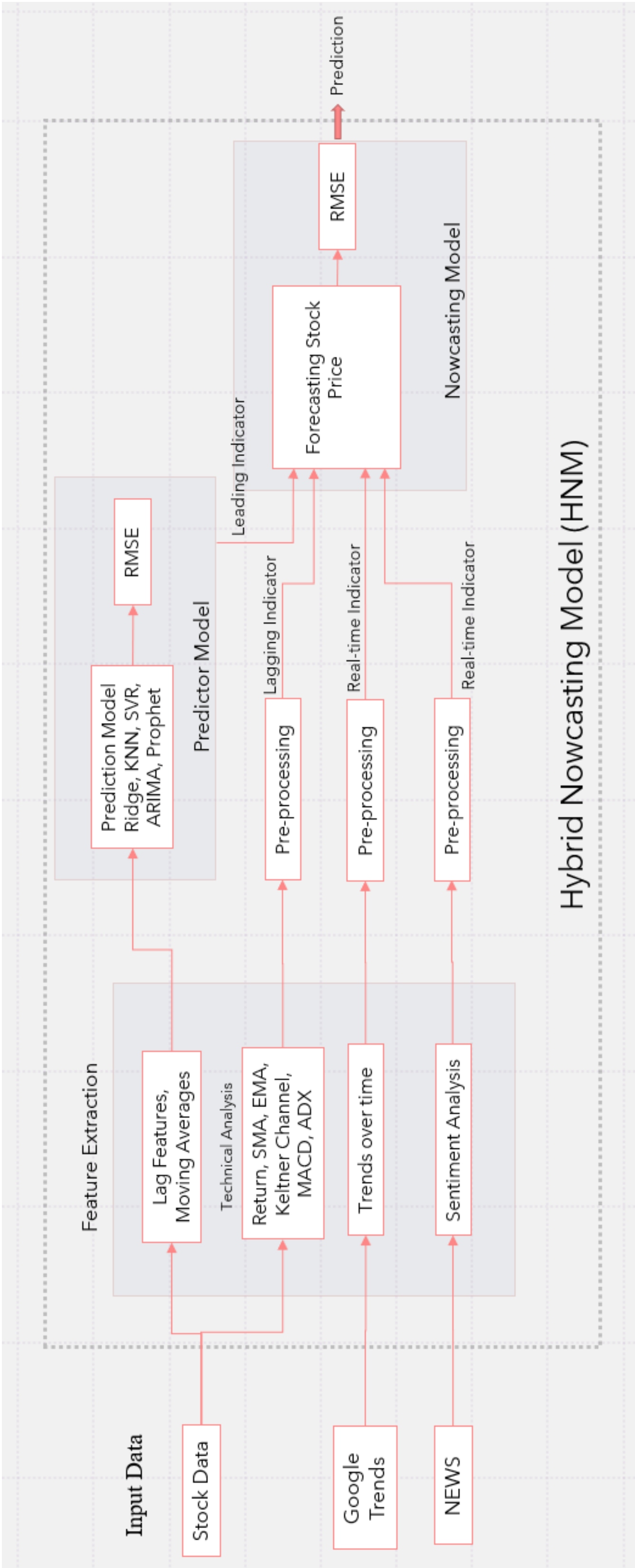


Figure 3.15: Architecture of Hybrid Nowcasting Model (HNM)

Chapter 4

Implementation

The dataset is collected using the yfinacne library in python. The dataset have eight columns: "Date", "Open", "High", "Low", "Close", "Volume", "Dividends" and "Stock Split". These columns are used to create the features discussed in the Chapter 3. All the features are clubbed with the original dataset.

The datasets are divided into train and test datasets. The typical test-train split does not apply since it randomly divides the data and can alter the time sequence. In this study, the datasets are divided based on length to preserve the time sequence. The split ratio for the study is 75:25, meaning that 75% of the datasets will be utilized to train the model and the remaining 25% to evaluate the data.

The hybrid Nowcasting model (HNM) has two parts: predictor model and the Nowcasting model. The predictor model takes the list of Close price, Volume, 5MVA, 22MVA, T-1Days, T-2Days, and T-3Days as an input feature to generate a leading indicator. The input to these models has two components: input features and target stock price. The target stock price at the time 't' is given by the stock price at the time (t+1). This can be achieved by shifting the close by 1. The input features are shown in Figure 4.1.

The second part of the HNM is the Nowcasting model. It takes leading, lagging and real-time indicators. The lagging indicators are calculated using the equation discussed in previous section and the outcome of the predictor model will be used as a leading indicator. All these will be clubbed with the original dataset.

The real-time indicators are obtained from Google Trends and News Sentiments. The Google trends are fetched using the pytrends library in python. The pytrends library takes search string and timeframe as input and returns a dataframe of search trends. The search string is the name of the stock. The trends can be utilized in various ways such as trends over time, trends over the region etc. In this study, trends over time are considered. This will return the trends for every day, Since the stock market works only on weekdays,

Date	Close	Volume	5MVA	22MVA	T-1Days	T-2Days	T-3Days	prediction
2022-03-11	344.200012	10801113	342.319995	356.397723	343.399994	340.399994	342.799988	341.100006
2022-03-14	341.100006	3889204	342.379999	357.270451	344.200012	343.399994	340.399994	340.250000
2022-03-15	340.250000	8759343	341.870001	355.179542	341.100006	344.200012	343.399994	346.250000
2022-03-16	346.250000	4512997	343.040002	353.599996	340.250000	341.100006	344.200012	379.799988
2022-03-17	379.799988	31782591	350.320001	353.759087	346.250000	340.250000	341.100006	391.250000
2022-03-21	391.250000	19955446	359.729999	354.095451	379.799988	346.250000	340.250000	393.100006
2022-03-22	393.100006	9561082	370.129999	354.824997	391.250000	379.799988	346.250000	397.250000
2022-03-23	397.250000	30739072	381.529999	356.527270	393.100006	391.250000	379.799988	407.100006
2022-03-24	407.100006	11671061	393.700000	358.927271	397.250000	393.100006	391.250000	419.250000
2022-03-25	419.250000	17327956	401.590002	363.093180	407.100006	397.250000	393.100006	461.000000

Figure 4.1: Input data for predictor model

the trends data needs to be processed. The weekdays and holidays are removed from the dataset and clubbed with the original dataset.

In the case of high-frequency data, hourly data for a specified date range is the maximum frequency of data that we can obtain using the Pytrends module in python.

Similarly, the News Data is collected using the newsapi package in python. The news API takes a query, api_key and language as input. The query is the name of the stock and the English language is considered for the analysis. The result of the query is a dictionary that contains the publish date, author, summary title etc. The publish date and title are used for sentiment analysis. The news API offers data for the last 60 days only. Furthermore, the data has to be preprocessed because there are multiple dates with no news. The dates with no news were given neutral sentiment (Value =0).

As discussed in feature engineering section, the sentiment analysis is done using a transformer-based model. The outcome of the model is the list of the sentiment score for negative, neutral and positive sentiment. The sentiment that has the maximum value is assigned to the overall sentiment of the news. The sentiments can be negative, neutral, or positive, with values of -1, 0, or +1 respectively. Moreover, the data is grouped on time since it has multiple news on the same date, in this case, the overall sentiment is calculated using the sum of the sentiments. If the aggregate sentiment is positive, the sentiment is assigned a score of +1. Similarly, scores of -1 or 0 are assigned where the aggregate sentiment is negative or neutral respectively.

All these indicators are clubbed to the original dataset. The final dataset has 21 columns: 'Date', 'Close', 'Volume', 'T-1Days', 'T-2Days', 'T-3Days', '5SMA', '22SMA', '12EMA', 'Ridge', '26EMA', 'kc_middle', 'kc_upper', 'kc_lower', 'macd', 'macd_h', 'macd_s', 'plus_di', 'minus_di', 'adx', 'trends', 'sentiment'. The model uses five different state-of-the-art prediction algorithms, discussed in Chapter 3, to predict the stock prices. The outcome of these algorithms is evaluated using the evaluation matrix (RMSE & MAPE). The best predictive algorithm is selected for the Nowcasting model.

The performance of the overall Hybrid Nowcasting Model (HNM) is tested on 10 NSE stocks, one high-frequency data, S&P 500 index and the FTSE index.

Chapter 5

Results & Discussion

The hybrid model of Nowcasting has two major sub-parts: the predictor model for a leading indicator and the Nowcasting model for overall predictions. For the analysis, the test set of 5 state-of-the-art prediction models is used in the proposed Hybrid Nowcasting model. To evaluate the prediction model the following evaluation matrix is used.

5.0.1 Evaluation Matrix

The evaluation matrix plays an important role in evaluating any model, analysing the results, and performing the comparative study. In this study, the RMSE and MAPE are used for evaluating the models.

Root Mean Square Error (RMSE)

The root of the mean square error, or RMSE, is determined by Equation (5.1), where e is the error term and n is the total number of terms. Error is the difference between the predicted stock price and the actual price of the stock.

$$\text{RMSE} = \sqrt{\frac{\sum e_t^2}{n}} \tag{5.1}$$

Mean Absolute Percentage Error

One of the most used methods of measuring accuracy for time-series data is the MAPE, which is used to determine the absolute error indicated by the Equation (5.2). where 'n' is the total number of words and 'e' is the error term. Error is the difference between the predicted stock price and the actual price of the stock.

$$\text{MAPE} = \frac{100}{n} * \sum_{i=0}^{n} * |\frac{y_{actual} - y_{predicted}}{y_{actual}}| \tag{5.2}$$

The performance of the predictor model is evaluated using RMSE and MAPE for three stocks: Adani Power, Tata Consultancy Services and Adani Wilmar. The comparison of five predictive algorithms for Adani Wilmar, Adani Power, and Tata Consultancy Services is shown in the figure below.

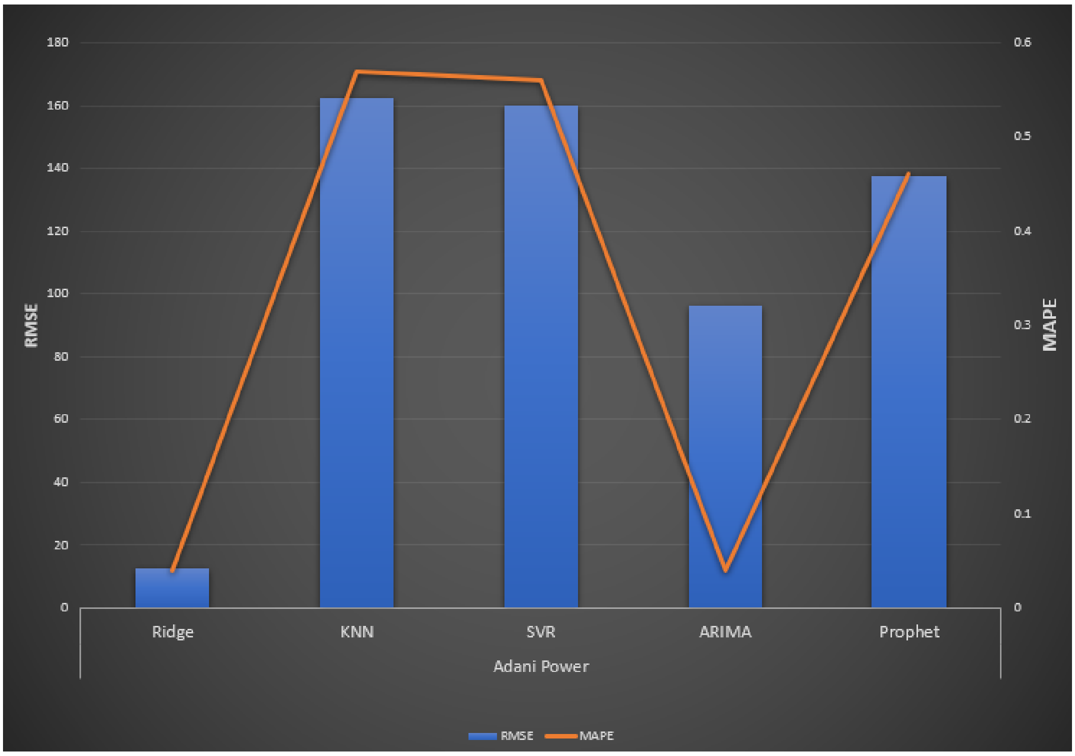


Figure 5.1: Comparison of RMSE and MAPE for Adani Power

In the Figure 5.1, 5.2, andA.5, the blue bar and the orange line represents the RMSE and MAPE respectively. There are two y-axes in the graph, the left y-axis is used for RMSE and the right y-axis is used for MAPE. The plots show that, across all three graphs, the Ridge Regression has the lowest RMSE and MAPE values. Thus, Ridge Regression is used for the prediction model. Ridge regression's hyperparameter alpha is calculated using GridSerchCV. An alpha value of 1000 is used for the analysis.

It is observed that the RMSE and MAPE value for Ridge is minimum in all three stocks. Thus, Ridge Regression is employed in HNM's predictor model. Similar to the approach used for determining the right algorithm for Predictor model, algorithm for Nowcasting model is also decided on the basis of performance of 5 different algorithms in forecasting the stock price of 10 NSE stocks. The outcome of each algorithm is evaluated using RMSE and MAPE. A detailed comparison of these results is shown in Figure 5.4&5.5.

In Figure 5.4&5.5, the green colour shows the minimum error value for the stock and the blue colour shows the second lowest error of the stock. For all 10 stocks, it has been found that ridge regression has the least error. ARIMA, on the other hand, is the second-

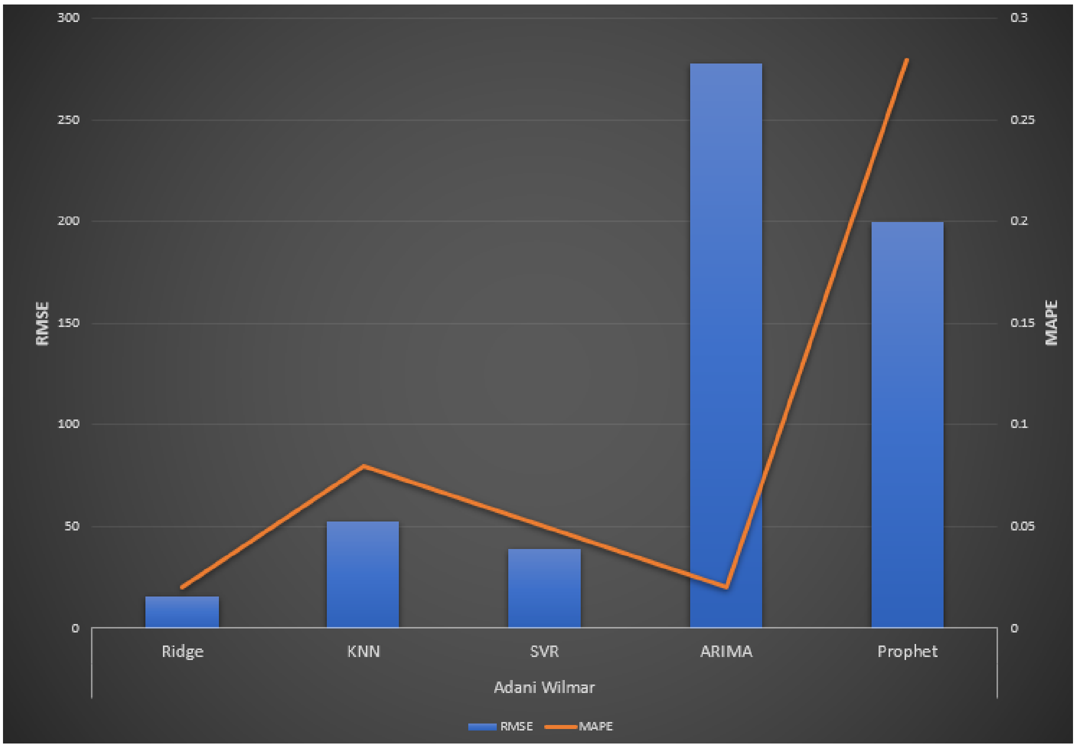


Figure 5.2: Comparison of RMSE and MAPE for Adani Wilmar

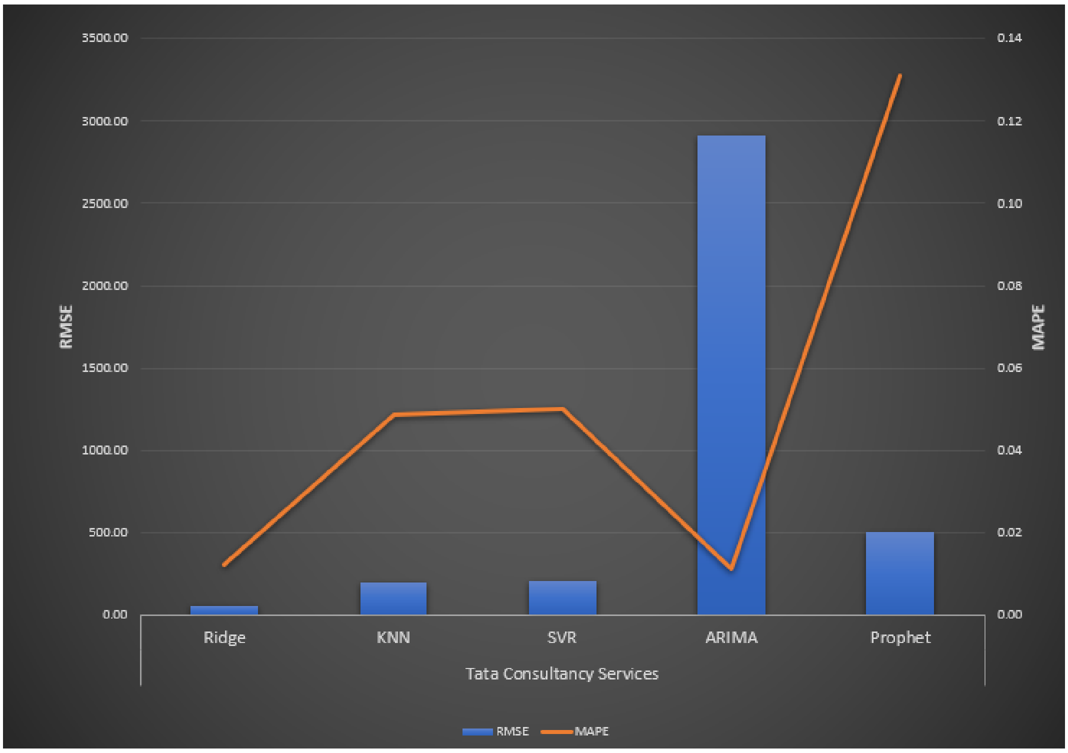
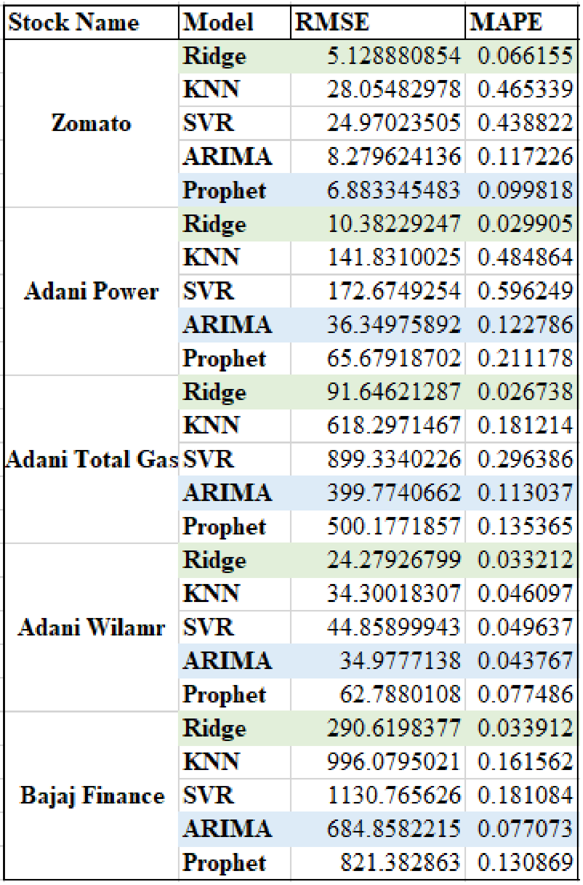


Figure 5.3: Comparison of RMSE and MAPE for TCS



Stock Name	Model	RMSE	MAPE
Zomato	Ridge	5.128880854	0.066155
	KNN	28.05482978	0.465339
	SVR	24.97023505	0.438822
	ARIMA	8.279624136	0.117226
	Prophet	6.883345483	0.099818
Adani Power	Ridge	10.38229247	0.029905
	KNN	141.8310025	0.484864
	SVR	172.6749254	0.596249
	ARIMA	36.34975892	0.122786
	Prophet	65.67918702	0.211178
Adani Total Gas	Ridge	91.64621287	0.026738
	KNN	618.2971467	0.181214
	SVR	899.3340226	0.296386
	ARIMA	399.7740662	0.113037
	Prophet	500.1771857	0.135365
Adani Wilamr	Ridge	24.27926799	0.033212
	KNN	34.30018307	0.046097
	SVR	44.85899943	0.049637
	ARIMA	34.9777138	0.043767
	Prophet	62.7880108	0.077486
Bajaj Finance	Ridge	290.6198377	0.033912
	KNN	996.0795021	0.161562
	SVR	1130.765626	0.181084
	ARIMA	684.8582215	0.077073
	Prophet	821.382863	0.130869

Figure 5.4: Detailed Comparative Analysis of Nowcasting model

best algorithm for the analysis. For nine stocks, ARIMA has the second-lowest value. Thus, ridge regression is utilized for the HNM's Nowcasting model. Figure 5.6 shows the RSME and MAPE of ridge regression for all 10 stocks. The range of MAPE is 0.066 to 0.009, where Zomato has the maximum MAPE of 6.6% and HDFC BANK has the lowest MAPE of 0.9%. Similarly, the RMSE ranges from 1.98 to 290.61 for Zomato and Bajaj Finance respectively. Further, the prediction results for Zomato and HDFC are shown in Figure 5.7 & 5.8 respectively. The predictions results for rest of the stocks are attached in Appendix.

The performance analysis of the proposed hybrid Nowcasting model (HNM) on individual Indicators is separately performed using Technical indicators, Real time indicators, and the HNM architecture. The comparison between all three is shown in Table 5.1. It is evident from the result that the proposed hybrid Nowcasting Model has the minimum MAPE value and RMSE value. Thus it can be concluded that the combination of these indicators provides better accuracy for stock price prediction.

To ascertain the robustness of the HNM model, the proposed architecture is also evaluated using high-frequency data (Hourly). The results are compared with the baseline

Stock Name	Model	RMSE	MAPE
Engineers Indai	Ridge	1.983705	0.028561
	KNN	4.644045	0.067485
	SVR	5.327555	0.077199
	ARIMA	4.726175	0.0568
	Prophet	5.123974	0.077119
HDFC BANK	Ridge	17.4708	0.009982
	KNN	101.7141	0.065304
	SVR	90.26213	0.059412
	ARIMA	43.16315	0.024888
	Prophet	62.41881	0.039237
Hindustan Uniliver	Ridge	52.46003	0.017644
	KNN	307.3157	0.102856
	SVR	270.0723	0.089769
	ARIMA	312.8935	0.103506
	Prophet	322.9484	0.105982
Tata Consultancy Services	Ridge	97.60394	0.02733
	KNN	407.4741	0.123213
	SVR	400.1102	0.12157
	ARIMA	174.9308	0.045848
	Prophet	419.9505	0.127599
Vedanta	Ridge	8.700048	0.025638
	KNN	71.36989	0.307111
	SVR	68.29296	0.292802
	ARIMA	65.8691	0.281667
	Prophet	100.0007	0.437244

Figure 5.5: Detailed Comparative Analysis of Nowcasting model

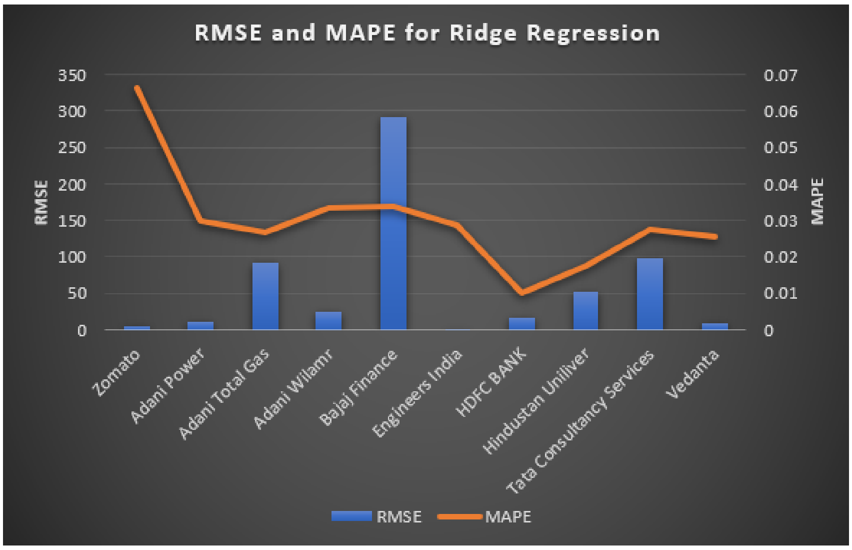


Figure 5.6: RMSE and MAPE for Ridge Regression

models i.e. ARIMA and Prophet. Table 5.2 demonstrates that the RMSE value for the hybrid Nowcasting model is 8.4532, which is lower than that of the ARIMA Prophet. Additionally, the mean absolute percentage error is 0.023, which is the minimum among

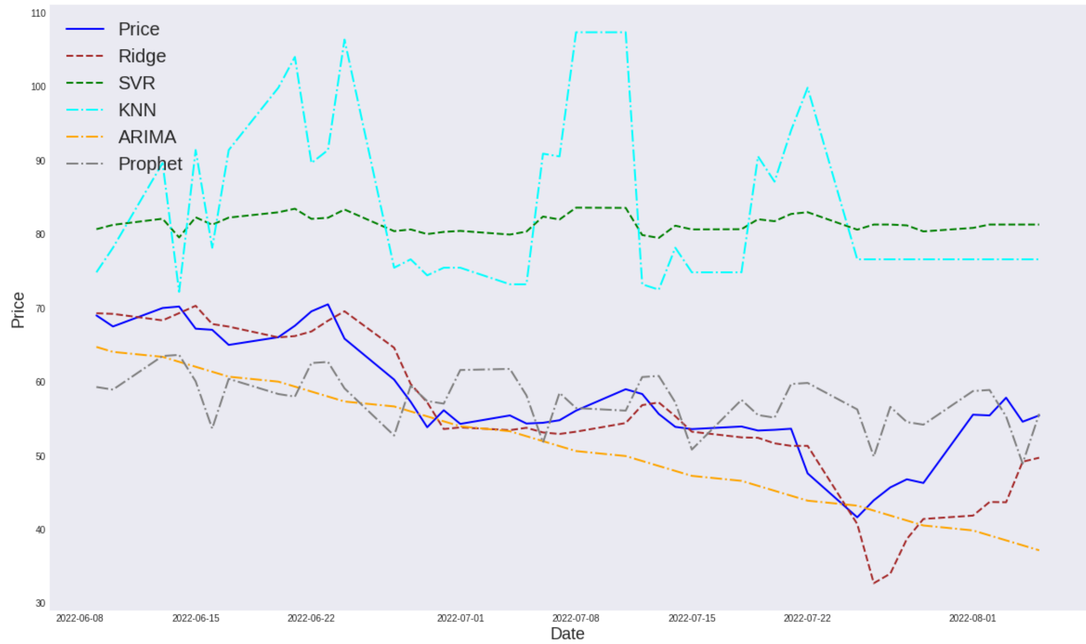


Figure 5.7: Predicted and Actual Stock price of Zomato

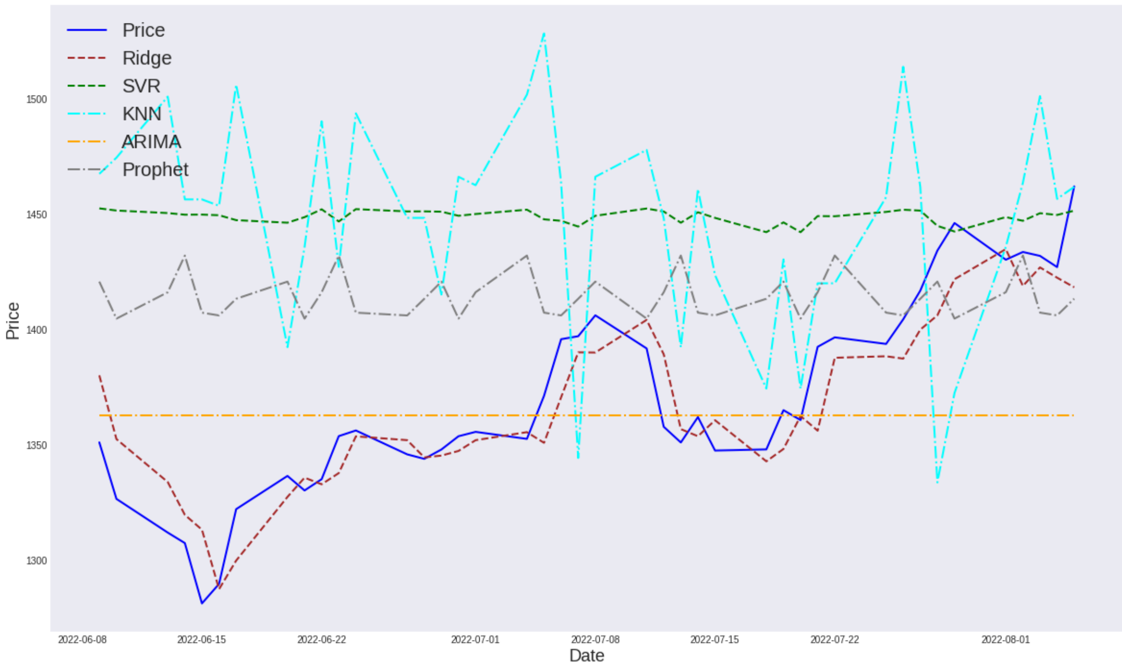


Figure 5.8: Predicted and Actual Stock price of HDFC

Table 5.1: Comparative study with individual indicator and HNM architecture

Error	**Technical Indicator**	**Real-time Indicator**	**HNM Architecture**
RMSE	16.03893888	128.5971168	10.38229247
MAPE	0.048104424	0.436242935	0.029905248

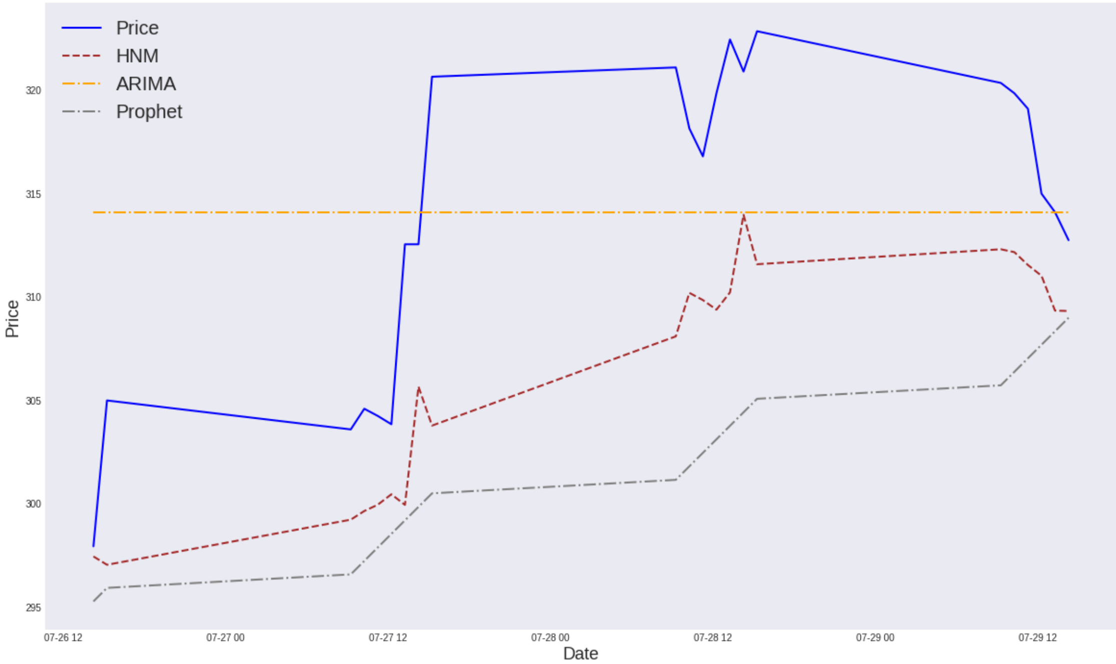


Figure 5.9: prediction for high frequency (Hourly) data of Adani Power

all three models. The prediction for Adani Power hourly data is shown in Figure 5.9. The red, orange, and grey dotted lines represent the forecasts made by HNM, ARIMA, and Prophet, respectively. The blue line represents the stock price. The ARIMA and Prophet findings lag in reflecting changes in the stock price, but the HNM predictions more closely follow the actual stock prices.

Table 5.2: Comparative study on S&P500 index, FTSE index, Adani power (Daily) and Adani Power (Hourly) data.

Stock Proce	**Model**	**RMSE**	**MAPE**
Adani Power (Hourly)	HNM	8.4532394	0.02383575
	ARIMA	14.861879	0.04154317
	PROPHET	12.431446	0.03541025
Adani Power (Daily)	HNM	6.8833455	0.09981827
	ARIMA	36.349759	0.12278568
	PROPHET	65.679187	0.21117804
S&P500	HNM	71.566933	0.01485454
	ARIMA	251.93774	0.05650521
	PROPHET	394.64107	0.09558031
FTSE	HNM	92.603158	0.00928192
	ARIMA	271.42275	0.03326455
	PROPHET	223.57461	0.02623717

To make the study even more rubust as well as to generalize the findings, performance

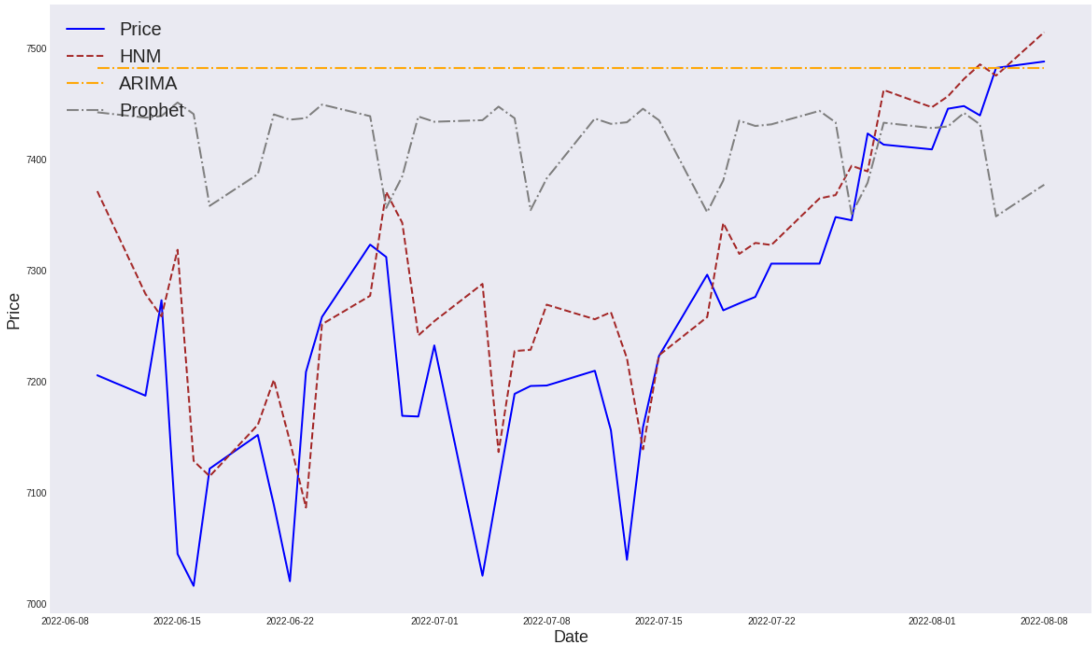


Figure 5.10: Stock Index prediction for FTSE

of HNM architecture vis-a-vis ARIMA and Prophet models is also evaluated on The FTSE 100 and S&P 500 indexes.The FTSE 100, also known as Financial Times Stock Exchange, is a share index of the 100 highest capitalisation companies. FTSE is a subsidiary of the London Stock Exchange Group. In a similar vein, the S&P 500 is an index of 500 large US-registered corporations. The comparison of the proposed HNM model, ARIMA, and Prophet on the FTSE 100 and S&P 500 indices is shown in Table 5.2. The results show that the HNM performs better and records the lowest value of RMSE and MAPE among all three models.

The prediction of FTSE and SP 500 is shown in Figure 5.10 & 5.11. The blue line represents the index value and the dotted lines represent the predicted value of the index. The HNM predictions are very close to the actual index prices for both FTSE and SP 500. The results show that the HNM not only performs better for the Indian stock market but it can be generalised to any stock exchange. The results show that the HNM model performs better than the state-of-the-art models for high-frequency data as well as low-frequency data.

The FTSE and S&P 500 index values predicted by HNM are fairly close to the actual index values. The findings demonstrate that the HNM performs better in the Indian stock market as well as on other stock exchanges. The outcomes demonstrate that the HNM model outperforms cutting-edge models ARIMA and Prophet for both high-frequency and low-frequency data.

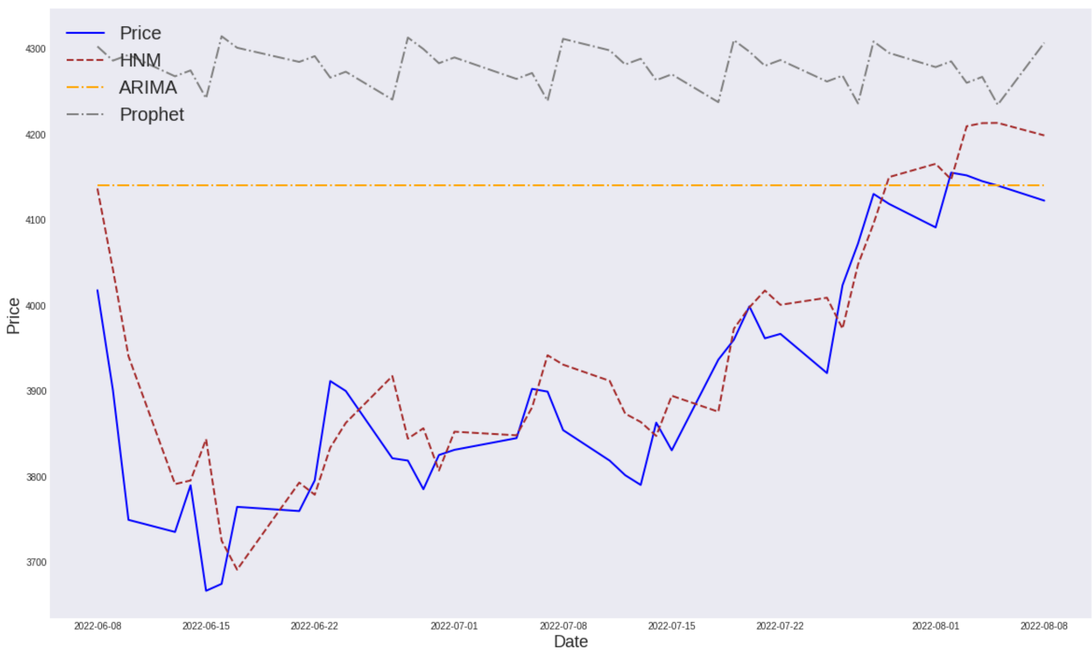


Figure 5.11: Stock Index prediction for S&P500

Chapter 6

Conclusion

This work is focused on precise stock price prediction using Nowcasting techniques. The literature survey motivated us to combine the sentiment analysis, google trends and technical indicators which play a vital role in Nowcasting the stock prices. Therefore, a novel hybrid Nowcasting technique is proposed that may be useful for both real-time as well as future stock price prediction. The proposed HNM model has two components: The predictor model and the Nowcasting model. The best algorithm for the predictor model is selected from five state-of-the-art prediction algorithms such as Ridge Regression, KNN, SVR, ARIMA and Prophet. The predictor model uses basic stock information, lag features, and moving averages as input features to generate a leading indicator. A comparative study on three stocks: Adani Power, Adani Wilmar, and Tata Consultancy Services is performed and the results are evaluated using RMSE and MAPE metrics. The result demonstrates that the Ridge Regression has the lowest RMSE and MAPE than the other employed models. Similar to the predictor model, the best algorithm for the Nowcasting model is selected among the same set of prediction algorithms, based on RMSE and MAPE score on 10 NSE Indian stock. The results show that the Ridge regression has the minimum RMSE and MAPE whereas the second best performer is ARIMA model. Thus Ridge Regression is considered for the HNM model while the ARIMA and Prophet are used as baseline models.

The MAPE of 4.8% and 43.6% are recorded when the prediction is done by only utilizing the technical and real-time indicators respectively, whereas the proposed HNM model provides MAPE of 2.9%. This confirms that the proposed combination of indicators provides better accuracy as compared to individual indicators. The robustness of the proposed HNM model is also evaluated on hourly high-frequency data for ADANIPOWER stock, and its results are compared with baseline models i.e. ARIMA and Prophet. The proposed model offers MAPE of 2.3% whereas the ARIMA and Prophet models provide MAPE of 4.1% and 3.5% respectively. To make the analysis even more robust and generalised, predictions on FTSE100 and S&P500 indices are made using the proposed model. The results demonstrate that the HNM outperforms than statistical models such as ARIMA and Prophet. Additionally, the proposed HNM is also found efficient for Dublin bike availability using a different set of features however, the results are not included due to space constraints. Thus, it is concluded that the proposed hybrid Nowcasting model is robust enough to be tested for any kind time-series forecasting application.

Limitations The HNM uses ridge regression internally for the predictor model and the Nowcasting model thus, the cascading of predictor model with Nowcasting model may amplify the error produced by predictor model. Further, limitation of the model is to analyse the sentiment of Multi-lingual News.

Future Scope The present work only considers English news for the sentiment analysis. Thus, the work may be extended to multi-lingual News. Apart from this, the technical indicators such as SMA, EMA, Keltner's Channel, MACD and ADX are used however other combination of technical indicators may also be analysed for the proposed HNM. The market indicators have their own positive and negative impact and thus, a research study may be conducted for assigning the weights to the input indicators of the HNM with the help of a suitable optimization algorithm.

Bibliography

[1] E. F. Fama, "Efficient capital markets: A review of theory and empirical work," *The journal of Finance*, vol. 25, no. 2, pp. 383–417, 1970.

[2] B. G. Malkiel, *A random walk down Wall Street: including a life-cycle guide to personal investing.* WW Norton & Company, 1999.

[3] G. W. Schwert, "Why does stock market volatility change over time?" *The journal of finance*, vol. 44, no. 5, pp. 1115–1153, 1989.

[4] I. K. Nti, A. F. Adekoya, and B. A. Weyori, "A systematic review of fundamental and technical analysis of stock market predictions," *Artificial Intelligence Review*, vol. 53, no. 4, pp. 3007–3057, 2020.

[5] N. Molodovsky, "A theory of price–earnings ratios," *Financial Analysts Journal*, vol. 51, no. 1, pp. 29–43, 1995.

[6] S. Ozlen, "The effect of company fundamentals on stock values," *European Researcher*, no. 3-2, pp. 595–602, 2014.

[7] T. Anbalagan and S. U. Maheswari, "Classification and prediction of stock market index based on fuzzy metagraph," *Procedia Computer Science*, vol. 47, pp. 214–221, 2015.

[8] E. Ahmadi, M. Jasemi, L. Monplaisir, M. A. Nabavi, A. Mahmoodi, and P. A. Jam, "New efficient hybrid candlestick technical analysis model for stock market timing on the basis of the support vector machine and heuristic algorithms of imperialist competition and genetic," *Expert Systems with Applications*, vol. 94, pp. 21–31, 2018.

[9] C.-H. Park and S. H. Irwin, "What do we know about the profitability of technical analysis?" *Journal of Economic surveys*, vol. 21, no. 4, pp. 786–826, 2007.

[10] T. Chavarnakul and D. Enke, "A hybrid stock trading system for intelligent technical analysis-based equivolume charting," *Neurocomputing*, vol. 72, no. 16-18, pp. 3517–3528, 2009.

[11] H. Solanki, "Comparative study of data mining tools and analysis with unified data mining theory," *International Journal of Computer Applications*, vol. 75, no. 16, pp. 23–28, 2013.

[12] M. Qiu and Y. Song, "Predicting the direction of stock market index movement using an optimized artificial neural network model," *PloS one*, vol. 11, no. 5, p. e0155133, 2016.

[13] A. A. Adebiyi, A. O. Adewumi, and C. K. Ayo, "Comparison of arima and artificial neural networks models for stock price prediction," *Journal of Applied Mathematics*, vol. 2014, 2014.

[14] D. Whitley, "A genetic algorithm tutorial," *Statistics and computing*, vol. 4, no. 2, pp. 65–85, 1994.

[15] R. Poli, J. Kennedy, and T. Blackwell, "Particle swarm optimization," *Swarm intelligence*, vol. 1, no. 1, pp. 33–57, 2007.

[16] M. Jain, S. Maurya, A. Rani, and V. Singh, "Owl search algorithm: a novel nature-inspired heuristic paradigm for global optimization," *Journal of Intelligent & Fuzzy Systems*, vol. 34, no. 3, pp. 1573–1582, 2018.

[17] S. Maurya, M. Jain, and N. Pachauri, "Development of sine cosine toolbox for labview," in *Intelligent Communication, Control and Devices*. Springer, 2020, pp. 747–753.

[18] S. K. Chandar, M. Sumathi, and S. Sivanandam, "Prediction of stock market price using hybrid of wavelet transform and artificial neural network," *Indian journal of Science and Technology*, vol. 9, no. 8, pp. 1–5, 2016.

[19] T. Preis, H. S. Moat, and H. E. Stanley, "Quantifying trading behavior in financial markets using google trends," *Scientific reports*, vol. 3, no. 1, pp. 1–6, 2013.

[20] A. Hamid and M. Heiden, "Forecasting volatility with empirical similarity and google trends," *Journal of Economic Behavior & Organization*, vol. 117, pp. 62–81, 2015.

[21] H. Hu, L. Tang, S. Zhang, and H. Wang, "Predicting the direction of stock markets using optimized neural networks with google trends," *Neurocomputing*, vol. 285, pp. 188–195, 2018.

[22] M. J. S. de Souza, D. G. F. Ramos, M. G. Pena, V. A. Sobreiro, and H. Kimura, "Examination of the profitability of technical analysis based on moving average strategies in brics," *Financial Innovation*, vol. 4, no. 1, pp. 1–18, 2018.

[23] S. B. Imandoust and M. Bolandraftar, "Forecasting the direction of stock market index movement using three data mining techniques: the case of tehran stock exchange," *International Journal of Engineering Research and Applications*, vol. 4, no. 6, pp. 106–117, 2014.

[24] M. Agrawal, A. U. Khan, and P. K. Shukla, "Stock price prediction using technical indicators: a predictive model using optimal deep learning," *Learning*, vol. 6, no. 2, p. 7, 2019.

[25] S. Deng, T. Mitsubuchi, K. Shioda, T. Shimada, and A. Sakurai, "Combining technical analysis with sentiment analysis for stock price prediction," in *2011 IEEE ninth international conference on dependable, autonomic and secure computing.* IEEE, 2011, pp. 800–807.

[26] N. Jing, Z. Wu, and H. Wang, "A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction," *Expert Systems with Applications*, vol. 178, p. 115019, 2021.

[27] M. Banbura, D. Giannone, and L. Reichlin, "Nowcasting with daily data," *European Central Bank, Working Paper*, p. 18, 2011.

[28] M. A. A. Khan, C. Bhushan, V. Ravi, V. S. Rao, and S. S. Orsu, "Nowcasting the financial time series with streaming data analytics under apache spark," *arXiv preprint arXiv:2202.11820*, 2022.

[29] J. S. Vaiz and M. Ramaswami, "A study on technical indicators in stock price movement prediction using decision tree algorithms," *American Journal of Engineering Research (AJER)*, vol. 5, no. 12, pp. 207–212, 2016.

[30] N. Norinder, "Predicting stock market movement using machine learning: Through r/wallstreetbets sentiment & google trends, herding versus wisdom of crowds," 2022.

[31] S. Mohanty, A. Vijay, and N. Gopakumar, "Stockbot: Using lstms to predict stock prices," *arXiv preprint arXiv:2207.06605*, 2022.

[32] S. Wu, Y. Liu, Z. Zou, and T.-H. Weng, "S_i_lstm: stock price prediction based on multiple data sources and sentiment analysis," *Connection Science*, vol. 34, no. 1, pp. 44–62, 2022.

[33] S. Bouktif, A. Fiaz, A. Ouni, and M. A. Serhani, "Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches," *Energies*, vol. 11, no. 7, p. 1636, 2018.

[34] M. A. Ruggiero, *Cybernetic Trading Strategies: developing a profitable trading system with state-of-the-art technologies.* John Wiley & Sons, 1997, vol. 68.

[35] H. Kaur, V. Mangat *et al.*, "A survey of sentiment analysis techniques," in *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC).* IEEE, 2017, pp. 921–925.

[36] S. Tajmouati, B. E. Wahbi, A. Bedoui, A. Abarda, and M. Dakkoun, "Applying k-nearest neighbors to time series forecasting: two new approaches," *arXiv preprint arXiv:2103.14200*, 2021.

Appendix A

Appendix

The prediction results of HNM's predictor model for remaining 8 NSE stocks discussed in the results & discussion section are attached below.

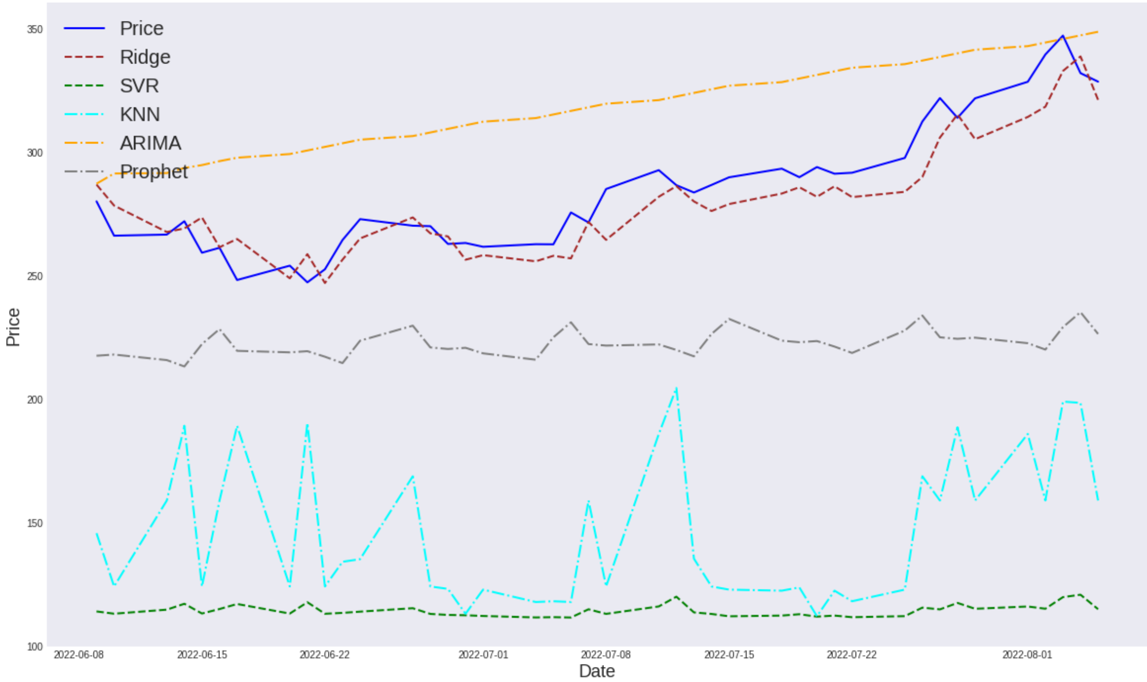


Figure A.1: Predicted and Actual Stock price of Adani Power

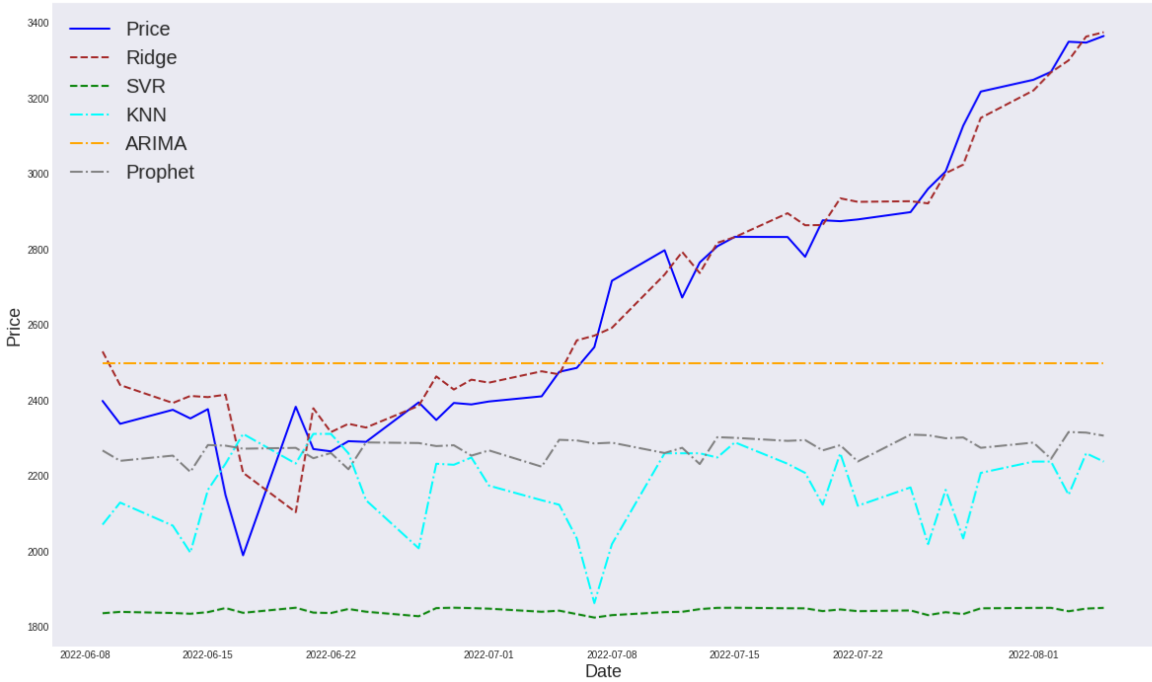


Figure A.2: Predicted and Actual Stock price of Adani Total GAS

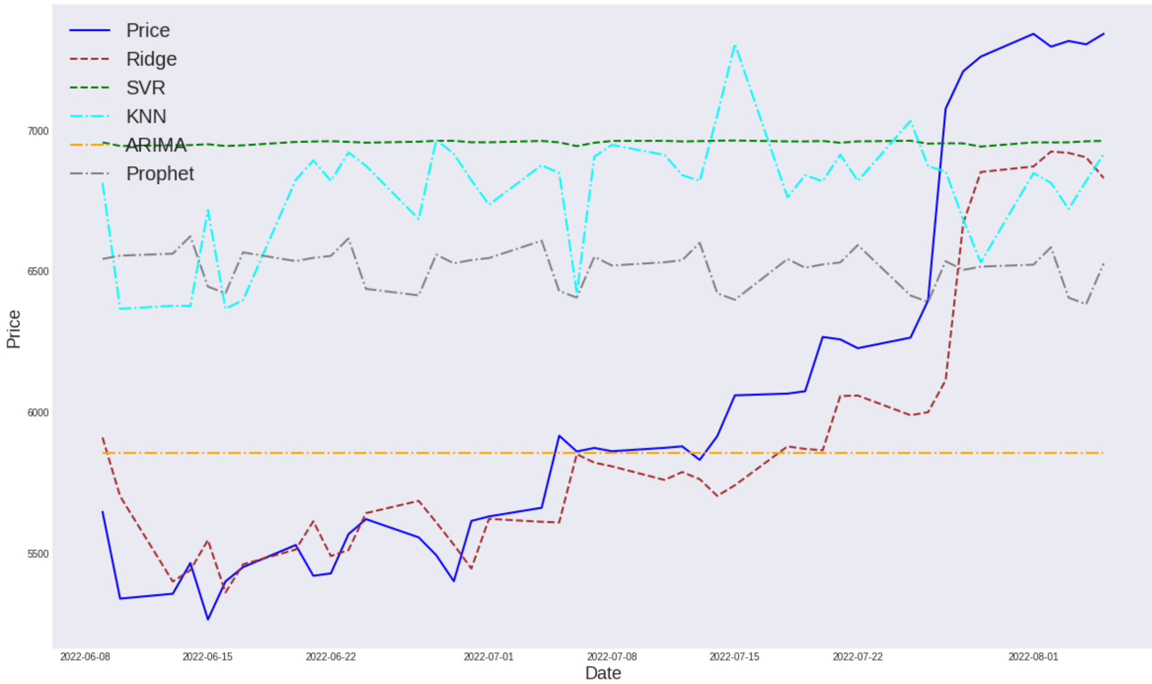


Figure A.3: Predicted and Actual Stock price of Bajaj Finance

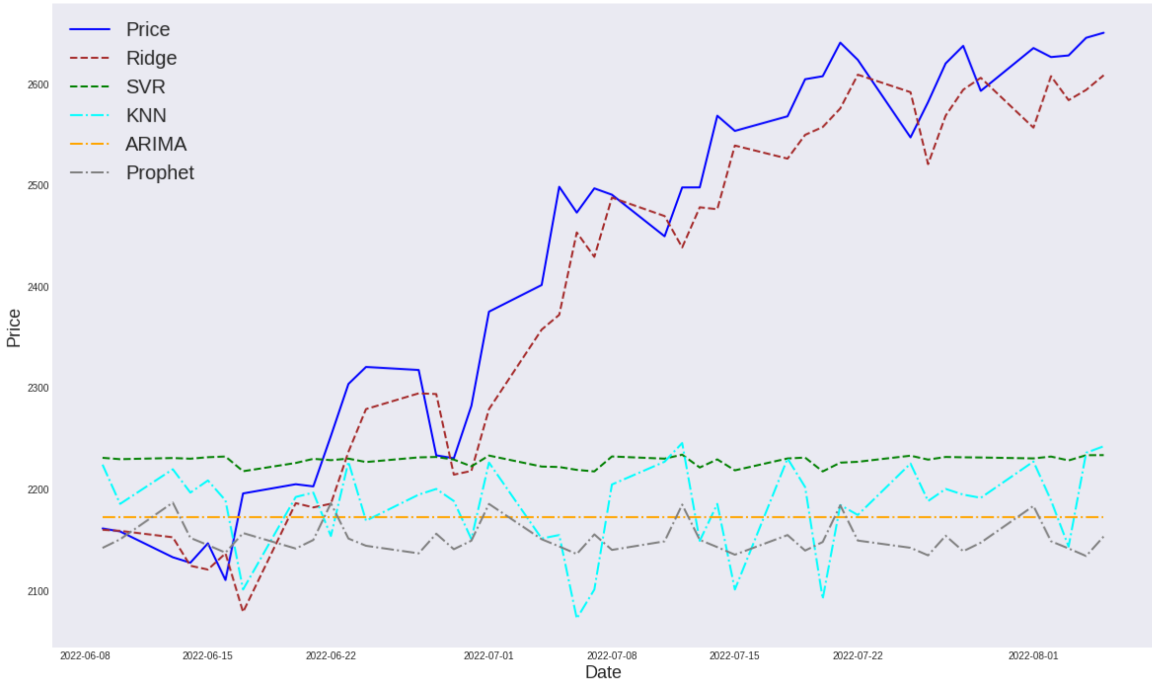


Figure A.4: Predicted and Actual Stock price of Hindustan Unilever

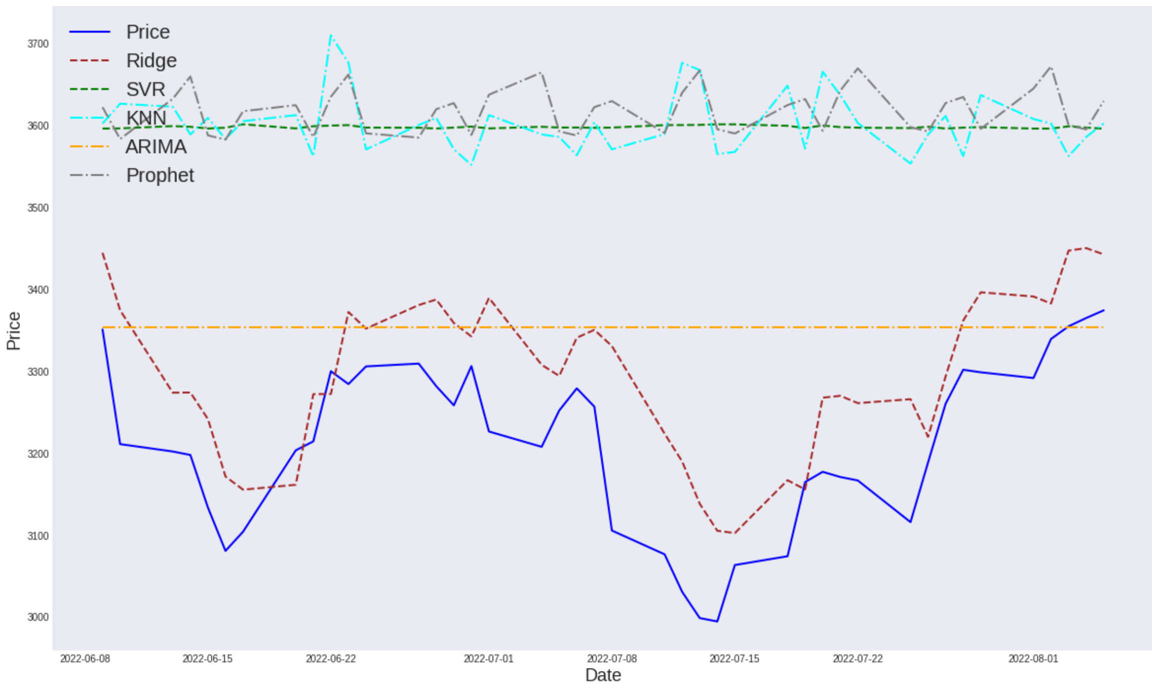


Figure A.5: Predicted and Actual Stock price of TCS

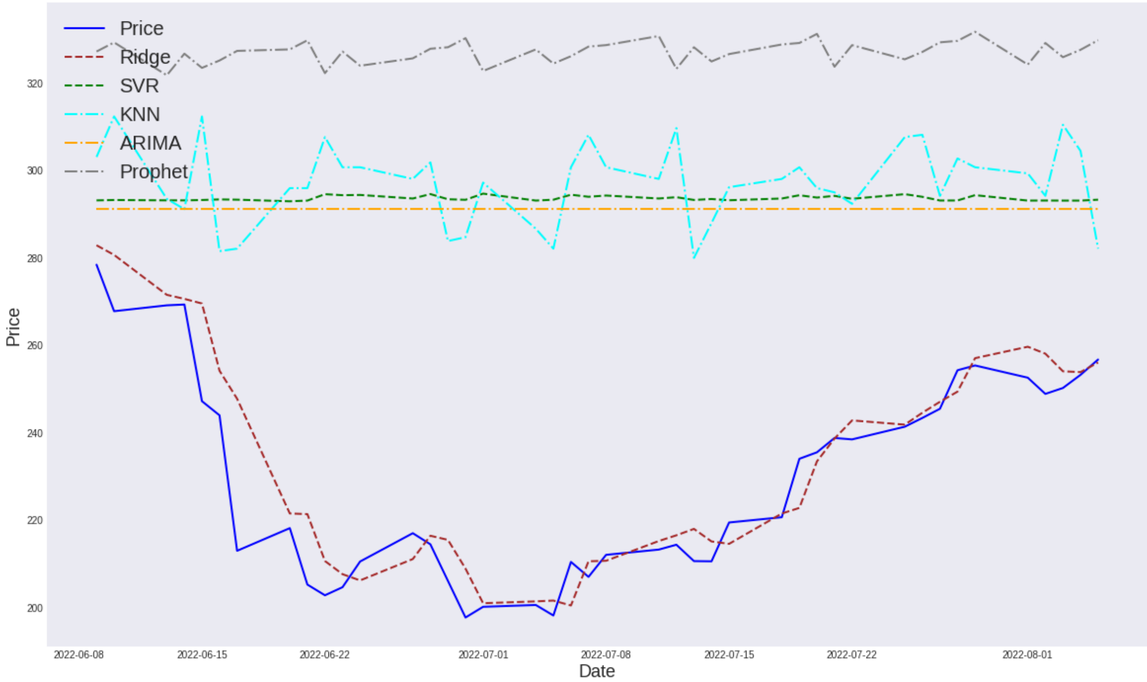


Figure A.6: Predicted and Actual Stock price of Vedanta

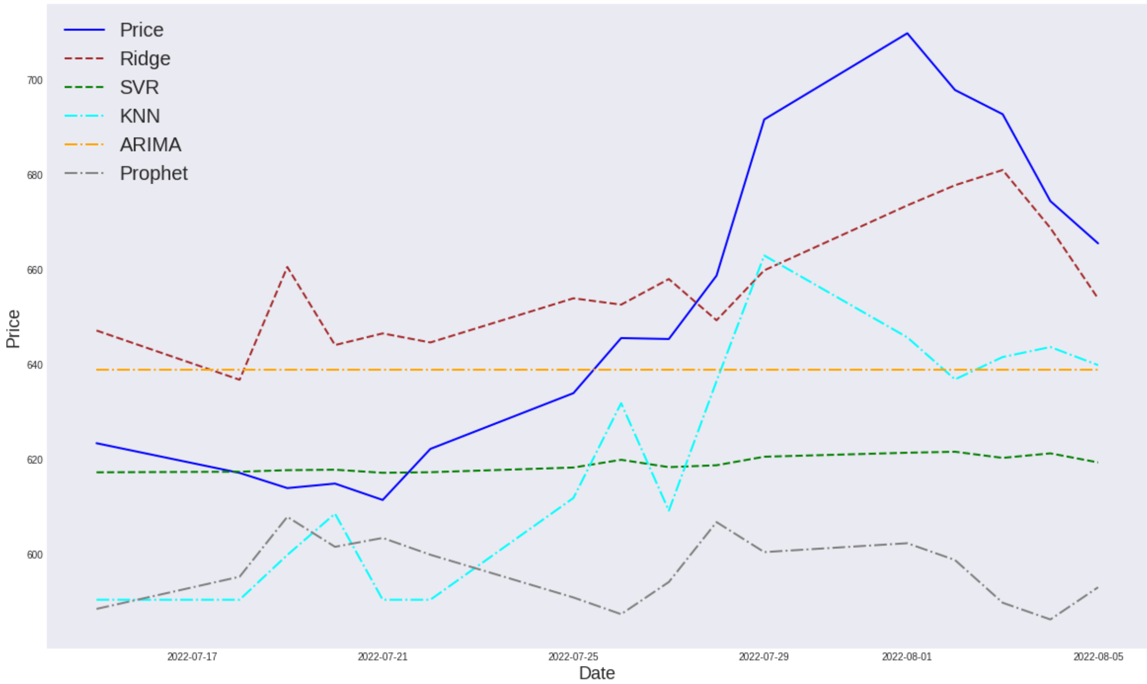


Figure A.7: Predicted and Actual Stock price of Adani Wilmar