CHAPTER 1 INTRODUCTION

1.1 Background

Basically, the stock market is an online buy-sell market where solo or institutional investors go to buy and sell shares or stocks. The shares that listed various companies on the stock market. The online stock exchange systems (such as the DSE, which is Dhaka stock exchange) essentially enable the purchase and sale of shares or stocks. These transactions must follow government regulations, which are designed to protect average investors from falling victim to scammers when buying or selling stocks. Let us now consider why a company distributes its stock to the public. A corporation frequently uses this to raise revenue or grow its operations. Investors earn from their purchases and sales of the corporation's shares. However, there is a danger of loss in such scenarios.

The increasing economic expansion of both developed and developing countries demonstrates the significance of the stock market in preserving stability and keeping up with the global economy. In Bangladesh, the stock market is equally important. However, in this sector, where "a little knowledge is a dangerous thing," ignorance and inexperience may be costly. Even if Bangladeshi general investors are more knowledgeable about the stock market than previously, their knowledge is still insufficient in real-world scenarios. According to studies, 75-85% of Bangladeshi investors do not know how to obtain stock market information or apply it to protect and increase their money. Even those with a thorough grasp of the market frequently lack the skills necessary to make wise decisions. Everyone, whether a small individual investor, a professional trader, a paid employee, or a little or big business owner, desires a secondary source of income to augment their primary earnings. The stock market provides such an opportunity, with sound judgment leading to significant returns on investment. However, this is an unpredictable topic that has greatly impacted the whole global economy system, a thorough understanding of stock market investing is required.

In the stock market concept, "fundamental analysis" and "technical analysis" are the two most important features of stock-market investing. They help to analyze a company's financial health and estimate future stock prices, offering critical information for making informed investment decisions. In the context of Bangladesh, making the right decision as an investor is relatively challenging due to the lack of sufficient resources. To solve this, we use a range of tactics, including machine learning to predict outcomes, which may help investors make better judgments.

A wide variety of technical and external factors impact stock market patterns, according to the research [1]. Although external implications are challenging, technical aspects are easy to identify.

Stock markets in developing countries like Bangladesh may be more sensitive to nuanced issues, such as public sentiment towards news stories. Consequently, you will get inaccurate results if you rely just on technical indicators to forecast the stock market. That being said, In the world of finance, stock trading is a big deal. One way to predict how much a stock or other financial instrument traded on a financial exchange will be worth in the future is via stock market forecasting.

Predictions of stock prices have always centered on the likelihood of substantial profit. In order to place wagers and foretell shifts in the market, investors have used a variety of strategies. Technical indicators like the Relative Strength Index (RSI), Money Flow Index (MFI), and Moving Average Convergence/Divergence (MACD) are used to analyze stock price patterns. Fundamental indicators like investor sentiment, news sentiment, EPS, and net asset value are also considered [3].

The value of stocks has been predicted using a variety of machine learning techniques. Stock prices may seem difficult to forecast due to their unpredictable nature. Due to the unpredictable nature of the stock market, simple regression or time series methods are not applicable. Gains in excess of average are still rare, despite the fact that financial institutions and traders have devised specialized methods to outperform the market. Gaining a little percentage point in stock forecasting might result in institutional gains of millions of dollars, making it an intriguing task. Time series linear models such as LR, SVR, LSTM, and ARIMA have traditionally served as the basis for prediction models. Stocks and assets are notoriously unpredictable, yet non-linear algorithms like XGBoost reduce forecasting errors compared to linear methods. This study forecasts a stock's value using machine learning.

1.2 Motivation of the Research

There have been very few research that use machine learning to anticipate stock values in the Bangladeshi stock market. Other tests on several stock markets to predict stock prices using machine learning approaches had disappointing results. We gathered stock price data from a variety of industries in Bangladesh, including basic commodities, consumer cyclicals, energy, and financials. We gathered information from 5 to 10 companies (such as Titas Gas, Apex etc) in each industry from numerous websites such as stock exchanges (investing.com, amarstock.com, etc.). The dataset has six columns: date, open, high, low, close, and volume. We also collected data on Bangladesh's news mood, DSE_index, Sector_index, and inflation rate. The High and Low represent the day's highest and lowest trading values, respectively, while the Date specifies a date and the Open displays a stock's opening price on that particular day. The Close indicates the day's closing price, whereas the Volume indicates the number of shares traded that day. We utilized the previous day's closing prices to anticipate our stock price for the next days. We tested the performance of SVR (Support Vector Machine-Regression) with Linear and Radial Basis Function (RBF) kernels, as well as Linear Regression, LSTM using historical pricing data. We aimed to predict stock prices more accurately than previous studies reported in the Literature Review section. One of the key motivations for doing his research was the scarcity of studies on Bangladeshi stock price forecast.

The motivation behind this study comes from the dynamic nature of Bangladesh's stock market, which is influenced by several political, economic, and company-specific factor. Investors frequently struggle to make sound judgments due to a lack of relevant industry-specific prediction tools. Current models either neglect Bangladesh's particular financial environment or are too broad for international markets.

1.3 Problem Statement

The use of powerful machine learning methods, such as neural networks, to forecast market values is not new. According to the claims, the models were inaccurate. Additionally, we discovered that all they do is gather technical data. We also gather information on the state of the news, the DSE index, inflation rate, and the sector index. In order to find out which model was the most effective at predicting stock prices for the Bangladeshi stock market, we compared several hybrid models that utilized SVR (Support Vector Machine-Regression) with Linear and Radial Basis Function (RBF) kernels and long short-term memory (LSTM) linear regression.

1.4 Objective of the Study

This research has as its goals:

- To identify the most effective algorithm.
- Enable traders to make short-term decisions (1-2 hours).
- Assist investors in planning long-term strategies (3-12 months).

CHAPTER 2 LITERATURE REVIEW

2.1 Introduction

Prediction on Stock market play an important role in investment of goods, economic stability and market development. By the use of prediction, it helps the investors to makes maximum profit, manage risks and make informed decisions. For policymakers, stock market trends serve as critical indicators of economic health, aiding in effective planning and crisis prevention. In a market like Bangladesh stock market predictions is more valuable due to market volatility and limited access to financial perception. Accurate and reliable predictions can attract the local and international investors to invest in Bangladeshi market which can help in economy growth and stable market.

2.2 Role of machine learning

Predictive analysis is greatly enhanced by large, dynamic datasets that enable machine learning to discover complex patterns and connections. Machine learning models outperform traditional statistical methods when it comes to dealing with non-linear data, adjusting for market swings, and taking into account variables such as historical prices, trading volume, and macroeconomic indicators. Short-Term Memory (LSTM) networks, Random Forest, and Support Vector Machines (SVM) are among the best ways to anticipate stock prices. These algorithms enhance the precision and dependability of stock market research by continuously refining forecasts, selecting features, and learning from past trends.

In emerging markets like Bangladesh, where data availability and market behavior present unique challenges, machine learning offers a flexible and scalable solution. It helps address these limitations by providing adaptive models that support better investment strategies, minimize risks, and contribute to a more resilient financial ecosystem.

2.3 Bangladesh’s Stock Market

The Dhaka Stock Exchange (DSE) and the Chittagong Stock Exchange (CSE) form the basis of Bangladesh's renownedly erratic and shallow stock market. Being a new and developing market, it has particular difficulties, such as insufficient liquidity, a large percentage of individual investors, and a restricted industry diversity. Furthermore, elements like interest rate fluctuations and political unpredictability have a significant influence on market behavior. These problems make predictive analysis more complex, highlighting the need for cutting-edge instruments like machine learning in effectively navigating and resolving them.

The Author [4] described this market. Their achieved highest accuracy of about 97.07%, followed by the SVR linear model and the SVR (radial basis function) model with the highest accuracy rates of about 97.06% and 96.82%.

2.4 Global Studies

The authors [8], explored to use the Autoregressive Integrated Moving Average (ARIMA) model is frequently applied in time-series analysis for finding linear relationships. The majority of academics, however, believe that the ARIMA model is incapable of detecting nonlinear patterns in data. As a consequence, SVM and ANN were utilized in the vast majority of techniques. The study proposed hybrid ARIMA and SVM models for stock price forecasting. Making use of data from many financial markets, the study [12], used an artificial neural network to forecast daily value of stocks. They focused on forecasting the market with artificial neural networks prices. ANNs are models that compute the data based on biological neural networks. SVM and textual analysis were employed in a study by [11] that examined the impact of news sentiment on the value of shares. Using a variety of textual representations, including named entities, noun phrases, and bags of words, the authors developed a forecasting machine learning method for financial news processing.

As per [5,7], research used Support Vector Regression (SVR) to predict stock prices. In the present work, parameter estimation is done manually or through trial and error. However, the hand calculation is inaccurate. To address this issue, the authors provided various kernel learning algorithms for improving SVR parameters. They found the accuracy with SVR to be 90%.

Using two different methods for feeding data into the models, the author [10] compares and contrasts four different stock market forecasting algorithms: ANN, SVM, Random Forest, and Naive Bayes. One uses trend-deterministic data to compute ten technical characteristics, while the other uses market transaction information (open, high, low, and closing prices) to reflect technical requirements. The accuracy of the prediction models is assessed by using the two input approaches.

Information gathered from the Dhaka Stock Exchange (DSE) between 2014 and 2021, according to studies done by [1]. Researchers used sentiment analysis and a multivariate long short-term memory (LSTM) neural network to forecast stock market movements.

While in a Predictions of stock market trends based on LSTM were 24% more accurate once external influences were included.

Finding expected relationships between different economic and financial indicators is one use case for gathering data using machine learning [18]. We determine the information gain for each model variable and then rank them based on the results. Establish a threshold to limit the inclusion of irrelevant variables in forecasting models. For level prediction and classification, we assess neural network models according to how well they foretell future values. Using cross-validation improves the models' generalizability. In order to test the theory, the models are fed S&P data spanning from March 1976 to December 1999.Outperforming buy-and-hold strategies, neural network models, and linear regression models in terms of risk-adjusted profitability are classification-guided trading techniques.

To better comprehend input model parameters and artificial neural network (ANN) stock market projections, the author [19] reviewed nine published papers. They look for the most important factors that improve the model's prediction accuracy. Their research shows that most ML methods base stock price predictions on technical factors instead of fundamental ones, even though microeconomic data is often used to determine the values of stock market indices. utilizing hybridized parameters is much better than utilizing just one kind of input variable.

The use of deep learning networks to the investigation and prediction of stock market behavior was the subject of a recent publication [20]. Because they can derive traits from massive volumes of raw data independently of the present knowledge of forecasters, deep learning networks are very useful for high-frequency stock market prediction. They lay forth the pros and cons of employing deep learning techniques to stock market research and forecasting in an unbiased manner. The authors study the effects of restricted Boltzmann machine, autoencoder, and principle component analysis, three unsupervised feature extraction methods, using high-frequency intraday stock returns as input data: Based on information gathered from 38 firms that were listed on the Korean stock exchange KOSPI between 2010 and 2014, the findings were presented.

In this paper, the authors address the difficulties of using kernel hyperparameters for stock market value estimation in support vector regression [21]. Learning could begin after a hyperparameter's value is known. In their pursuit of optimal system performance, they may tweak a number of hyperparameters. Their two-stage learning with multiple kernels system is the result of combining gradient projection with stepwise least optimization. Therefore, it is clear that up until this point conclusions drawn from the data collected Using the Taiwan Capitalization Weighted Stock Index as a benchmark, the superior method outperforms the competitors. The following study uses the most diverse set of machine learning approaches. Individual data mining approaches have successfully created accurate stock price movement projections, but traders have discovered that they must utilize numerous forecasting methods to have a greater understanding of the stock market's future.

We used daily stock closing price data from 2002-2005 for training, validation, and testing the model

A wide variety of machine learning methods are used in this study. Although certain data mining techniques have shown remarkable accuracy in predicting stock price changes, traders have discovered that a mix of methods is necessary to fully grasp the market's trajectory. A ten-method data mining strategy is used to forecast the next movement of the Hong Kong Hang Seng index [22]. Logit models, neural networks, K-nearest neighbor classification, LDA, QDA, support vector machines, LS-SVM, and naive Bayes with kernel estimate were all part of the group. Using data from January 3, 2000, to December 29, 2006, we examine the daily change in closing prices of the Hang Seng index using five predictors. Both LS-SVM and SVM perform better than competing models in experiments. When it comes to making predictions outside of a given sample, LS-SVM outperforms SVM in terms of hit rate and error rate, even if SVM is great when making predictions inside a given sample.

Research [23] provides a method for predicting the daily movement of the market. Using DNNs and ANNs, it trains on three datasets: the original, two variations generated by principal component analysis (PCA), and the third dataset that remains unchanged. According to the results, when overfitting is considered, the accuracy of DNNs is improved when the number of hidden layers is increased from 12 to 1000. When compared to other ML models and the original dataset, DNNs trained on PCA-transformed datasets achieve superior classification accuracy. Trading strategies that combine PCA data with DNN predictions also tend to be more successful. This dataset takes 60 economic and financial variables as inputs and returns the daily change (increase or decrease) in the closing price of the SPDR S&P 500 ETF as output. June saw a total of 25,181 trading days.

The nonlinearity of the data makes high accuracy in stock price forecasts difficult, despite several attempts to employ machine learning methods. There has been a lot of study on ANNs, or artificial neural networks. After looking at the Korean stock market, they found that, [13] The accuracy of the 30-day and 20-day moving averages was 83.11% and 81.34%, respectively. Another research looked at three stocks and found an accuracy percentage of 86.69% [14]. Additional research showed that ANN achieved the highest SSE scores (0.627 for Apple and 0.021 for LG), and that a 5-fold prediction model achieved an accuracy of 96.10%. It is also a good idea to use fewer square support vector machines.

The accuracy rate ranged from 90.5% to 93.3% when PSO was used with it. Adobe, HP, and American Express all showed MSE values of 0.5316, 0.7725, and 0.7905 when LS-SVM was paired with PSO, respectively, in the research of [23]. Positive outcomes were also produced using support vector machines (SVMs). Using a poisson kernel, SVM achieved an accuracy of 90.10 percent on the BSE-Sensex; using an RBF kernel, it achieved 88.088 percent. With a polynomial kernel, the accuracy for Infosys stock was 89.59%, and with an RBF kernel, it was 87.80%. The MAPE of the author's Linear Regression-Based Pattern Forecasting approach is 5.41% when applied to bank data and 5.42% when applied to all stock data. [16] Compared to Structural Support Vector Machines (SSVMs), which achieved 78% accuracy on training samples but fell short on testing data, K-Nearest Neighbors (KNN) achieved 83.52% accuracy. Since the majority of models continue to disregard the inherent non-linearity of stock market data, more investigation into this area is clearly required.

2.5 Research Scope

We discovered significant gaps in the research publications we analyzed. Few studies focused on projecting stock values using data from the Bangladeshi stock market, and the existing models were not very accurate. This is an excellent opportunity to deal with stock data from Bangladesh. Using powerful machine learning approaches such as Support Vector Regression (SVR), Linear Regression (LR), Long short-term memory (LSTM), XGBoost can result in increased accuracy. Our strategy for projecting future prices was based mostly on shares' closing prices. The Dhaka Stock Exchange (DSE) and Investing.com provided the data for our model.

CHAPTER 3 METHODOLOGY

3.1 Question Framing

Prior to doing any data science inquiry, it is essential to identify the appropriate questions to ask. By poring over the dataset, one could discover interesting hypotheses and queries. We found that the two most significant elements impacting stock prices, based on the data, were news sentiment and closing price. Therefore, in order to forecast the closing prices of the next days, we will mostly refer to the closing prices of the previous days.

3.2 The Model That Is Being Presented

The structure of our proposed model is represented in the diagram below.

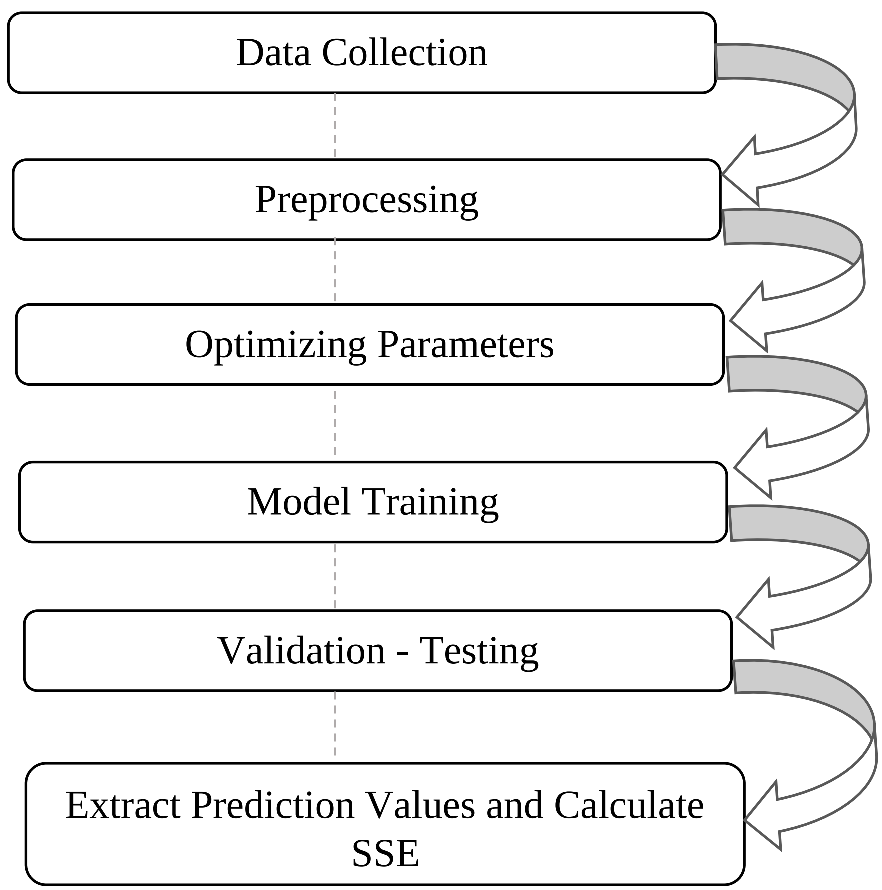
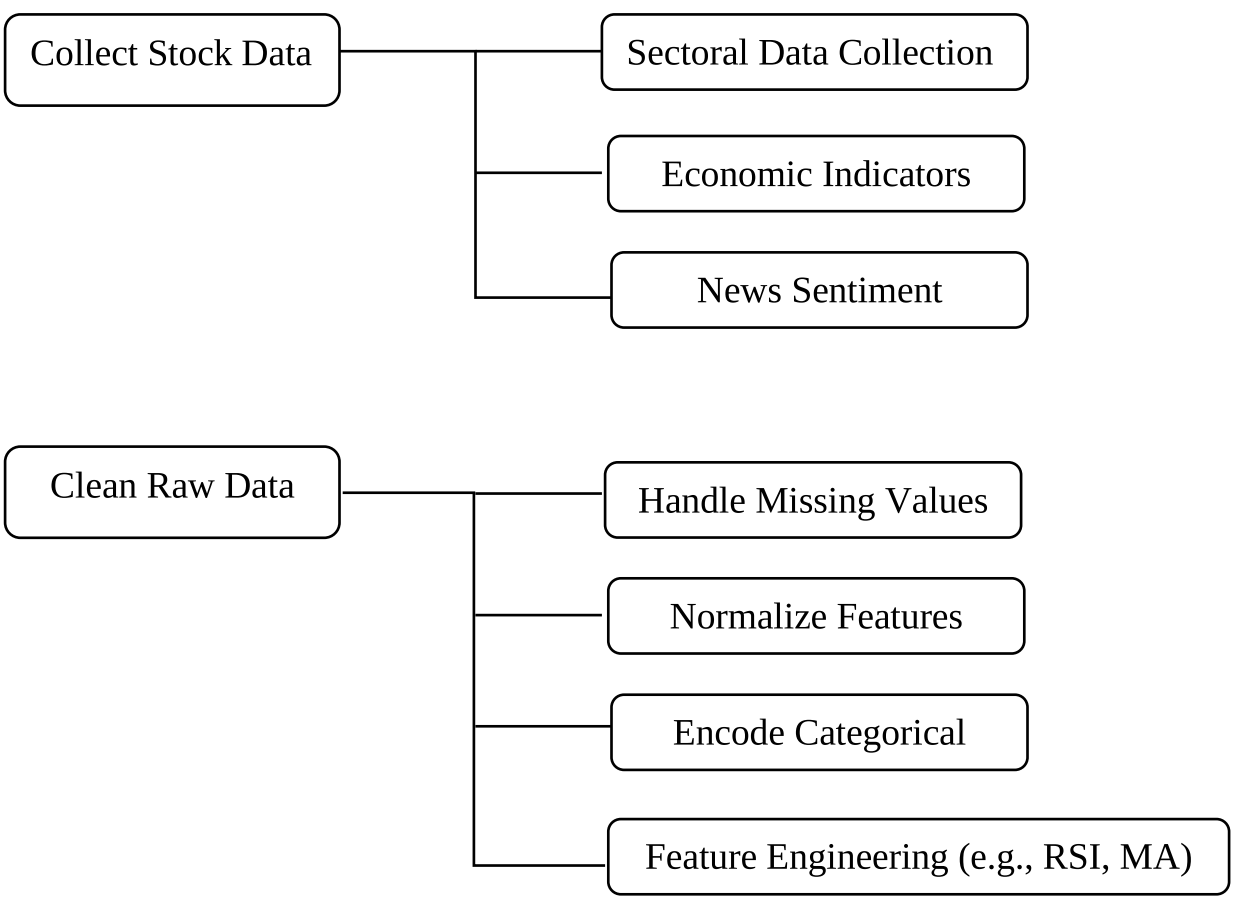


Fig 1: Proposed Model

Our proposed model, as shown in Fig 1, consists of some several steps. Initially, we collected and pre-processed the data. Next, we optimized the parameters required for the training algorithm. Using these optimized parameters, we tested the dataset in two ways. One is hybrid architecture (LSTM, XGBoost) and another one is SVR and Linear Regression (LR) Architecture. After that, we validated and test the dataset. Afterward, we extracted the output and evaluated the accuracy of the algorithms using SSE. Finally, we visualized a comparison between the actual values and the predicted values.

3.3 Gathering and Preprocessing Data

We gathered the information from a number of internet sources, including Investing.com, Amar Stock, and the Dhaka Stock Exchange (DSE). The Dhaka Stock Exchange have all of the companies historical prices available on this open-source website. We gathered information on several businesses across various industries between January 1, 2021, and December 30, 2024. We also gathered information on the inflation rate, DSE_index, Sector_index, and news mood in Bangladesh. From the most current price to the the older ones price form, we obtained the dataset. We had to reverse the data from the oldest price to the most current price order in order to process it in the order we wanted. We looked for any missing data. The RSI, MA_5 (Moving Average), MA_30, MA_60, and MA_90 (Long term) were then discovered. So, we used the dataset as it was.



1	Date	Price	Open	High	Low	Volume	News_Sentiment
2	12/26/2024	1,800.80	1,800.00	1,805.00	1,790.00	1030	Positive
3	12/24/2024	1,799.60	1,800.00	1,802.00	1,796.00	340	Positive
4	12/23/2024	1,800.10	1,791.00	1,802.00	1,791.00	580	Negative
5	12/22/2024	1,802.50	1,800.00	1,804.00	1,790.30	500	Neutral
6	12/19/2024	1,800.40	1,805.00	1,808.00	1,800.00	830	Positive
7	12/18/2024	1,809.30	1,780.00	1,810.00	1,780.00	330	Negative
8	12/17/2024	1,807.00	1,770.10	1,810.00	1,770.10	3320	Positive
9	12/15/2024	1,783.20	1,770.90	1,784.00	1,770.00	1810	Positive
10	12/12/2024	1,770.90	1,780.00	1,780.00	1,770.00	570	Positive
11	12/11/2024	1,780.40	1,779.90	1,782.90	1,775.00	116	Neutral
12	12/10/2024	1,770.50	1,770.50	1,777.00	1,768.50	1190	Neutral
13	12/9/2024	1,768.40	1,768.00	1,779.00	1,768.00	30	Negative
14	12/8/2024	1,768.00	1,770.00	1,770.00	1,762.50	30	Negative
15	12/5/2024	1,772.60	1,782.00	1,782.00	1,752.00	40	Positive
16	12/4/2024	1,771.60	1,710.00	1,782.30	1,710.00	230	Negative
17	12/3/2024	1,774.60	1,784.00	1,784.00	1,770.00	310	Neutral

Fig 2: Sample data set

3.4 Parameter Optimization from the Dataset

For predicting stock prices, we worked with our dataset's "Close" column as an input parameter. The closing prices from several dates were utilize as the dependent variable in prediction. The dataset was split into 2 parts: 80% training data and 20% testing data, which was stored in its own variable for use as parameters.

3.5 Model Training

3.5.1 Hybrid Architecture

We will be use hybrid architecture for data set. At first split the dataset then training the Long short-term model (LSTM). Then We will apply XGBoost on the dataset. To increase prediction accuracy, the outputs of LSTM and XGBoost are merged using weighted averages or stacking approaches. The hybrid model is fine-tuned using hyperparameter optimization to minimize prediction errors. The best-performing hybrid model is selected based on evaluation metrics.

```
Input Processed Data Set
  └── Split Data (70% Train, 15% Validation, 15% Test) ──────┐
          ├─ Train LSTM (Sequential Data)
          ├─ Train XGBoost (Structured Features)
          ├─ Hyperparameter Tuning
          └─ Combine Predictions (Hybrid Model)
```

3.5.2 SVR and Linear Regression (LR) Architecture

Also, we will use another model for more accurate accuracy. Support Vector Regression is used to capture nonlinear relationships in the data, while Linear Regression provides simple and interpretable predictions. The parameters for both models were carefully adjusted to improve accuracy. The predictions were then evaluated to compare performance and select the most effective model.

```
Input Processed Data Set
  └── Split Data (Train and Test Sets) ──────┐
          ├─ Train SVR (Support Vector Regression)
          ├─ Train LR (Linear Regression)
          ├─ Optimize Parameters
          └─ Compare Predictions with Metrics
```

3.6 Validation and Testing

Validation and testing are critical for ensuring that trained models work well on fresh data and generate correct predictions. Validation is the process of testing the model on a second dataset that was not used for training in order to modify hyperparameters and identify mistakes such as overfitting or underfitting. Metrics used to measure performance include mean absolute error (MAE), accuracy, precision, recall, and root mean squared error. Testing, on the other hand, evaluates the model's feasibility using previously unknown data. This phase demonstrates that the model can forecast stock changes on a regular basis, making it suitable for real-world applications.

3.7 Deriving the predicted value and computing the SSE

Once trained, the model may use either the training or testing datasets to provide predictions. These forecasts are compared to the real values in order to assess their precision and effectiveness. To measure the difference between predicted and observed values, one may use the following formula to get the Sum of Squared Errors (SSE):

$$\mathrm{SSE} = \sum_{i=1}^{n} (\mathrm{Actual}_i - \mathrm{Predicted}_i)^2$$

Here, \(n\) represents the total number of data points. A lower SSE indicates that the model has a better fit and smaller prediction errors. This step helps pinpoint areas where the model might need further refinement and ensures it performs well on unseen data. By minimizing errors, the model is optimized for real-world applications.

CHAPTER 4 CONCLUSION & FUTURE WORK

4.1 Summery

The thesis addresses the challenge of improving accuracy of forecasts in Bangladesh's volatile and dynamic stock market. This study's primary objective is to predict changes in stock prices by employing advanced machine learning algorithms to analyze historical market data. Thus far, there has been a substantial advancement in the collecting and the pre-processing of data, which serves as the foundation of the predictive modeling process. A thorough understanding of the market's past performance is ensured by the dataset's inclusion of important variables such stock prices, trading volumes, and macroeconomic factors. Our work also drew attention to practical insights for investors. Our results, highlighting data on the stock market, will give meaningful strategies for perhaps guiding the investors in picking profitable opportunities. It is, however, recommended that these be looked into in light of the limitations this study has imposed.

4.2 Limitation

The Bangladeshi stock market is particularly influenced by political and economic events, and unexpected occurrences, such as regulatory shifts or sudden market disruptions, might not be fully represented in historical data. This can reduce the model's overall accuracy. In addition, news sentiment analysis relies heavily on the effectiveness of natural language processing tools, which might introduce biases or fail to accurately capture the true sentiment of the market. Macro-economic factors, like inflation and exchange rates, often have delayed effects, which could cause short-term predictive accuracy to be slower in adjusting. Recognizing these challenges allows for further improvements in future iterations, aiming to enhance the reliability and effectiveness of stock market predictions within the context of Bangladesh.

4.3 Future Work

The first part of this study was primarily concerned with acquiring the essential data, but much more work still to be done before the predictive system can be completely developed. The following phases in the procedure will be:

Preparing the Data: We'll clean and arrange the data so it's ready for model training. This includes removing missing numbers, transforming categorical data into accessible representations, and creating new features such as moving averages and other pertinent indicators.

Developing New Features: The objective is to generate more sophisticated features that can better capture market behavior, sector performance, and investor mood, hence improving the model's forecasts.

Building the models together: Long Short-Term Memory (LSTM) networks will be used to evaluate time-series data, with XGBoost handling structural characteristics. We will also change the model's parameters to achieve the best results with LR, SVM (Support vector Machine)

Validation and Testing: Error rates (such as MAE and RMSE), recall, accuracy, and precision are among the measures used to assess the model's performance. This guarantees that when new, unidentified data is added to the model, it performs well.

Creating a Real-Time Prediction System: We aim to implement a live prediction system that will process real-time data and provide actionable insights.

Exploring Advanced Models: We will also experiment with more complex machine learning models, like transformers or ensemble techniques, to make the predictions more accurate and robust.

By following these steps, the research will ultimately create a practical and reliable stock market prediction system tailored to the unique characteristics of the Bangladeshi market.