CHAPTER III

RESEARCH METHODOLOGY

Predicting unknown events and producing facts for the future are among the most crucial research tasks in commercial applications, and knowledge is produced by combining several disciplines. In business, the data is utilized to develop cutting-edge technology, identify potential risk factors, and examine behavioral patterns to make dynamic decisions that ultimately lead to opportunities.

Predictions are derived from the analysis of the variables included in transactional and historical database sets by identifying the relationships between them, the changes that impact the other variables, and the nature of those changes. Time is one such variable, and it's a significant one that has recently gained importance. Nowadays, a lot of businesses use time series analysis and forecasting to create and improve their business plans.

3.1 Basics of Time series analysis

Time series data, where the variable Time is observed as a collection of data sequentially and indexed by a timestamp or date, are used in time-dependent applications. Time series data and time-dependent variables are used by businesses for forecasting, analysis, and system control optimization. Time series data analysis produces patterns, and analysts use these patterns to infer important information that might guide businesses in taking preventative action. On the other hand, time series forecasting projects future events and can offer data scientists insights that help guide decisions that could potentially alter patterns.

Time series data are being used more often in a variety of applications, particularly with the development of ANN approaches. Numerous opportunities have been found, and potential research issues need to be investigated. Time series forecasting is frequently used in applications such as meteorology, sales, economics, finance & budgeting, stock market, utility studies, astronomy, supply & demand, inventories, and many more. It helps with management, minimization, optimization, and the development of resources and services.

Extrapolating historical data or projecting the future concerning time is the first step in identifying trends and patterns, as many time-dependent applications like finance and stock research require. For the straightforward reason that many variables are time-dependent and exhibit behavior changes over time, the relationship between them can be inferred using a straightforward deterministic linear regression technique.

Other modeling techniques used in time series forecasting include the more recent vector autoregressions, neural networks, bootstrapping, and bagging techniques, as well as the classical methods of "autoregression, autoregression, autoregressive Integrated Moving Average (ARIMA), • GARCH (Generalized Autoregressive Conditional Heteroskedasticity) and exponential smoothing". When selecting a technique or model, one should take into account the application's criteria as well as the amount of data, resources available, accuracy rate attained, and model usage strategy.

3.2 Prediction methods

Statistical prediction methods play an important role in the prediction of stock. Here are some commonly used statistical prediction methods for stock market prediction:

- **Linear regression:** The dependent variable and one or more independent variables (economic indicators, historical stock prices, etc.) have a linear connection established by linear regression models. Based on prior performance and pertinent variables, they can be used to forecast future stock values.

- **ARIMA:** Time series forecasting, particularly stock market prediction, makes extensive use of ARIMA models. To account for trends, seasonality, and autocorrelation, they model the link between a set of observations and the lags of the series. After differencing, stationary time series data can be subjected to ARIMA models to eliminate seasonality and trends.

- **GARCH:** GARCH models are used to model and anticipate volatility in financial time series, such as stock prices. They can record the fluctuations in stock returns over time or the way they cluster together. GARCH models are particularly useful for estimating and forecasting the conditional variance of asset returns.
- **Vector Auto Regression (VAR):** VAR models are used to extend autoregressive models to multivariate time series data. They capture the way that numerous elements, such as interest rates, stock prices, and economic indicators, interact dynamically with one another. The interdependencies and causal relationships between different financial variables can be investigated using VAR models.

- **ML Algorithms:** The use of ML techniques for stock market forecasting is expanding. These algorithms can capture nonlinear interactions and complex patterns in the data. The processes of feature engineering, regularization, and hyperparameter tuning are required to apply ML algorithms to stock market prediction in an efficient manner.
- **Ensemble approaches:** Numerous prediction models are combined in ensemble methods to increase prediction flexibility and correctness. To combine the predictions of various models, such as regression models, ARIMA models, or ML algorithms, techniques.
- **Bayesian methods:** In Bayesian statistical procedures, posterior distributions are updated based on observed data, and prior knowledge is taken into consideration. Bayesian approaches can be used for model selection, parameter estimations, and uncertainty quantification in stock market prediction.
- **Time series decomposition:** Time series decomposition techniques break down the time series into its constituent parts, such as seasonal decomposition of time series (STL) or single spectrum analysis (SSA). By using these methods, prediction models become more accurate and help uncover and simulate the underlying patterns in stock prices.

3.3 Experimental Setup and Dataset Description

Validating models, comprehending their performance, and implementing data-driven enhancements all depend on the analysis and findings of experiments. The present research work focused on making a more efficient stock market prediction model based on optimized ELM with investor sentiment analysis to make the right decision at the right time. The analysis results are executed using MATLAB. The present sub-section discussed the dataset collection and its details as follows,

3.3.1 Datasets

Six datasets are collected from the Yahoo financial section [64]. Two stock indices datasets such as Nifty 50 and S & P BSE Sensex, three bank stock equity datasets are considered such as SBIN, ICICI, and HDFC, and one software company equity data such as Microsoft (MSFT). These datasets are collected over the 10 financial year datasets between March 2013 to March 2023 (Total, 3247 trading days samples). The following subsection discusses the datasets and their history. Samples datasets such as S & P BSE and Nifty 50 index, SBIN, ICICI, HDFC bank, and MSFT datasets are shown in Appendix-1.

i) S & P BSE Sensex:

The free-float market-weighted BSE SENSEX, also referred to as the S&P, is composed of thirty respectable, financially sound companies that are listed on BSE. The 30 constituent companies represent various industrial sectors of the Indian economy and include some of the major and most dynamically traded stocks.

ii) Nifty 50 :

The benchmark NIFTY 50 index is the weighted average of fifty main Indian companies. With an ecosystem made up of NSE and SGX futures and options, exchange-traded funds (both onshore and offshore), and exchange-traded funds, it has grown to become the largest financial product in India. The greatest exported agreement internationally is the NIFTY 50. NSE is regarded as a frontrunner by surveys from WFE, IOM, and FIA.

The NIFTY 50 provides investment managers with exposure to the market confidential a single portfolio by encompassing thirteen sectors of the Indian economy. Financial services, including banking, are given 36.81% weighting in the NIFTY 50 as of January 2023. It is given 14.70%, oil and gas are given 12.17%, consumer goods are given 9.02%, and autos are given 5.84%.

iii)SBIN:

SBI is a global public sector bank and financial services regulatory body in India, with its main office situated in Mumbai, Maharashtra. SBI, ranked 48th internationally in terms of total assets, is the only Indian bank named on the Fortune Global five hundred list of the world's major companies. This public sector bank is the largest in India, with a 25 percent market share in loans and payments. It is the tenth-largest employer in India with roughly 250,000 employees. SBI became the third lender and seventh largest market capital on Indian stock exchanges when it crossed the ₹ 5 trillion benchmark for the first time. The biggest public lender in the country reached a noteworthy milestone on February 7, 2024, when its market value surpassed ₹ 6 lakh crore. This made it, after Life Insurance Corporation, the second public sector endeavor to accomplish this achievement.

iv) ICICI bank :

ICICI is a global bank in India, with its headquarters located in Mumbai and its registered office situated in Vadodara. This development finance company operates in 17 countries and has 16,650 ATMs and 5,900 locations across India. The largest firms in India are represented in significant indices such as the Nifty 50 and S&P BSE Sensex, which include ICICI Bank's stock. Several significant indices, including the BSE Sensex and Nifty 50, which monitor the performance of the leading Indian corporations, include ICICI Bank.

v) HDFC bank :

HDFC is one of the biggest private-sector banks in India, and makes up the HDFC dataset utilized for stock market prediction. The dataset facilitates the analysis of volatility, sentiment-driven price changes, and stock patterns. News, analyst reports, and conversations on social media all have a significant impact on HDFC stock.

vi) MSFT :

One of the biggest publicly traded IT firms is Microsoft Corporation (MSFT), which is listed on the NASDAQ Stock Exchange. It is a crucial gauge of market performance since it has a significant impact on the Dow Jones Industrial Average (DJIA), NASDAQ-100, and S&P 500.

3.4 Consequence of ANNs in predictive study

ANNs have a big impact on predictive analytics. These tools solve complicated problems that are hard and wasteful for traditional statistical approaches to solve, but they also produce far more reliable findings. Its methods for providing answers to dynamic data and associated challenges are reliable, and it comes with supplementary tools for investigating novel situations. ANNs are sought after by many creativities due to their capability to monitor and analyze large volumes of data and offer highly likely answers.

More consistent information makes it relaxed to make informed decisions. Using the application of ANN approaches, insights that were not visible using conventional methods can be found. This can highlight imaginable risks and problems and expand the quality of the application. It has been used in situations when prediction proved difficult, and the methods' accuracy has shown promise. To tackle bottlenecks that emerge in classical approaches with many constraints, there exist ML events with distinct and varied factors [65].

3.4.1 ANNs

ANNs are one of the most powerful tools in artificial intelligence and it is constructed as layers of nodes with input, hidden, and output layers that are related to each other, weights, and bias that act as the threshold to operate on the data. Its nodes are based on mathematical and computational concepts. The nodes' architecture and connections are inspired by the connections seen in human brain cells. Weights and thresholds are initially set to random values when ANNs are trained using training data. After that, the data is processed through several levels, opening at the input layer, calculated, and then meaningfully changed at the output layer.

ANNs can learn intricate patterns from historical data and produce forecasts based on those learned patterns, ANNs have become more and more prominent in the stock market prediction space. An ANN is a kind of ML algorithm that is made up of linked nodes, or neurons, arranged in layers and is demonstrated after the structure and operations of the human brain. Various neural network types include:

- Back propagation neural network (BPNN)
- Radial basis function (RBF)
- ELM

ANNs are used in stock market prediction in the following ways:

- Pattern Recognition: ANNs are a feasible option for analyzing historical data because of their excellent efficacy in finding relations and patterns in large datasets. Using past price data, trading volumes, technical indicators, and other relevant information, an ANN can be trained to identify patterns that predict future price changes.
- Non-linear Relationships: Predictive statistical approaches may find it difficult to explain the compound and non-linear relationships that are often seen in stock market data. Since ANNs are so good at modeling these non-linear interactions, they can recognize subtle patterns and trends in the data that old analysis techniques would miss.
- Feature extraction: ANNs may extract applicable features from untreated data, removing the need for human feature engineering. This is especially supportive for predicting the stock market, where pertinent features could include different technical indicators, macroeconomic variables, news mood, and market emotion.
- Flexibility: ANNs are extremely adaptable and may be used for a wide range of stock market forecast tasks, such as volatility prediction, trend prediction, price prediction, and portfolio optimization. To increase forecast accuracy even more, they can be used in conjunction with other ML strategies, such as ensemble methods.

- Time-series Forecasting: ANNs are a respectable fit for applications involving time series forecasting, including proud future returns or stock prices. ANNs can be trained to precisely anticipate forthcoming price movements across a variety of periods by participating in past price data and other pertinent time-series variables.
- Risk management: When forecasting the stock market, ANNs can also be exploited for risk management. They can be used, for instance, to calculate Value at Risk (VaR), spot anomalies or outliers in trading data, and maximize trading methods to reduce negative returns.
- Constant Learning: ANNs may be trained on fresh data regularly, which enables them to adjust to shifting market conditions and take in new information as it becomes available. This flexibility is essential in the volatile and dynamic financial markets.

While ANNs offer several advantages in stock prediction, it's crucial to distinguish that they also have convinced disadvantages. ANNs are particularly prone to overfitting when trained on sparse or noisy data, and fluctuations in the input data or model design may have an impact on the predictions made by the model. Moreover, it may be challenging to understand the underlying principles of ANNs, which complicates understanding the logic behind specific forecasts. Thus, careful validation, calibration, and interpretation are required when using ANNs for stock market prediction.

3.5 Research framework

The present research work focused on three phases of contributions as follows,

3.5.1 Phase -1

The ELM approach computes the output weights analytically and first initializes hidden node values at random. The main advantage of ELM is that the buried layer does not need to be tuned. For the randomly chosen input weights and hidden layer biases, ELM will produce the least squares solution of a system of line equations for the unknown output weights with the lowest norm property. For numerous real-world applications, ELM has shown good generalization results with a very fast learning rate [17-20]. Nevertheless, there are still a lot of issues with the actual application of the ELM. The most important one is the selection of the optimal hyperparameters such as the number of hidden nodes, connection weights and biases, and learning rate which are usually done by trial and error. Generally speaking, the ELM requires more hidden neurons than more predictable tuning-based learning methods.

As a result, a lot of studies have been done to determine the best weights and biases for ELM to improve prediction outcomes. However, there are numerous flaws in the standard methods, including local optima, a lengthy computation time, and a low convergence rate. Therefore, the weight and bias of the ELM known as DELM, which was used in the stock price forecast, were adjusted using DWM. When the DELM was compared to some conventional and variant ELMs, the accuracy result was greater. Further, the historical datasets are considered such as open, high, low, closing, and adjusted close in this phase for stock price prediction.

3.5.2 Phase -2

Traders risk making the mistake of overfitting their trading methods to historical data, which results in tactics that perform well on historical data but are not suitable for upcoming market conditions. In real-world trading, overfitting can lead to subpar results. It is frequently difficult to discern important patterns from random variations in historical pricing data due to the high noise content. Erroneous trading decisions might result from traders misinterpreting noise as noteworthy trends or patterns. Technical indicators offer valuable information about market patterns, price movements, and possible future price directions, which is useful when examining stock market data. Technical indicators are useful for verifying market trends or price changes.

Price and indicator moves that diverge could indicate future market reversals or continuations. It helps with risk management by offering information about possible gains or losses. Technical indicators are useful instruments for analyzing the market, but to make informed trading decisions, they should be combined with other types of analysis, such as fundamental analysis. Hence, the second phase of research work uses the ten different technical indicators such as Simple Moving Average (SMV), 10-day Moving Average, Momentum, Stochastic (K%), Stochastic (D%), Relative Strength Index (RSI), Williams (%R), Moving Average Convergence Divergence (MACD), Commodity Channel Index (CCI), Price Oscillator (PO) are used for predicting stock price using DWM-ELM.

3.5.3 Phase - 3

Conversely, the developed stock market prediction algorithms ignore other factors that affect stocks and their complex internal workings in favor of primarily using historical data as their input. Because people are inconsistent, the stock market does not always operate according to systematic rules. Their behavioral, psychological, and emotional characteristics are vital in the economic system. Furthermore, new research has shown that investor attitude may have an important effect on stock market returns. There is a strong hint that investors are not irrational, and as social networks become more important in people's lives, shareholder connections are getting easier and more common.

Thus, an investor's perspective and decision-making procedures may be influenced by the feelings and opinions of other investors as well as those shared on social media. Next, the CNN technique is used to extract the user comments. The suggested plan considers investor sentiment by computing binary sentiment indices for optimistic and negative sentiments. An algorithm for sentiment analysis based on CNN that divides stock market remarks into bullish and negative perspectives. Next, the stock price that is improved by DWM is predicted using enhanced ELM. Finally, technical indicators and user comments datasets are considered as inputs to the prediction method.

The developed sentiment-based DWM-ELM (S-DELM) is working to forecast the stock price. The proposed architecture is shown in Figure 2. Further, some of the following remarkable points are short definitions of the contribution as follows,

- The sentiment-based optimized ELM is applied to predict the stock price.
- To determine user sentiment on the stock, a CNN algorithm is used to calculate the sentiment index (SI).
- To identify factors affecting the share market with stockholder sentiment analysis or classification of opinion sentiment into bullish or bearish.
- To produce sentiment analysis in real-time that can forecast the value of the stock market and, using historical data, estimate the approximate share price.
- The DWM method is used to optimize the weights and bias of ELM to enhance the performance ELM method.
- Six datasets are used to investigate the performance of S-DELM

3.6 Summary

Finally, the use of technical indicators, optimal hyperparameters optimization, and customer sentiment are considered in the proposed research work for achieving the following remarkable objectives such as reducing prediction error, enhancing prediction accuracy, and making fast learning time and fast convergence rate.

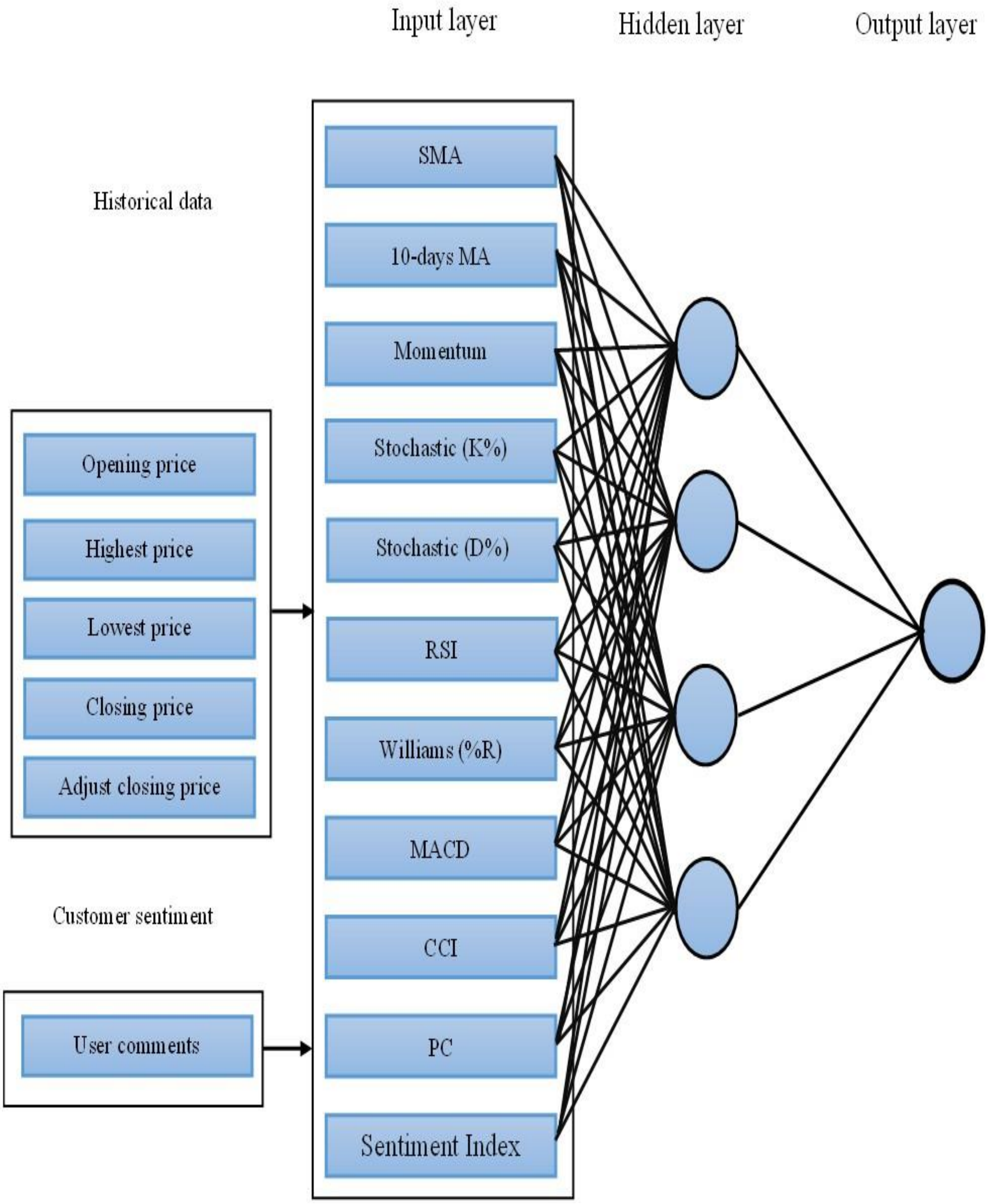

Figure 2: Architecture of proposed method