IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

COMPARATIVE ANALYSIS OF VARIOUS HYBRID MODELS OVER STOCK MARKET DATASET

[1]Rishabh Saxena, [2]Sandeep Kumar

[1]Student, [2]Associate Professor

[1]Department of Computer Science and Engineering

[1]Sharda University, Greater Noida, India

Abstract : The world of financial markets faces a formidable challenge when it comes to accurately anticipating stock market movements. Conventional prediction techniques have struggled to contend with the intricate and uncertain nature of market dynamics, often yielding suboptimal outcomes. Inaccurate forecasts can have far-reaching implications, impacting investment strategies, financial choices, and overall economic stability. Consequently, there is an urgent demand for the exploration of fresh and inventive methods that can bolster our capacity to forecast stock prices with increased accuracy and dependability. This study investigates the potential of hybrid Machine-Learning (ML) models as a promising remedy to this persistent issue. This research presents a comparative analysis between multiple hybrid models applied to stock market datasets. These models were assessed using three distinct datasets spanning the years 2022-2023, 2021-2023, and 2018-2023 for five major stocks: RELIANCE, TCS, HDFC, ITC, and INFOSYS. Result dictate that ARIMA-HMM and RDAWA are the two models from the chosen ones that provide good results. Out of the two, ARIMA gives the best perform with an accuracy of 80% and metrics sitting under 0.3 for all datasets. Following that, RDAWA gives a good and robust perform with an accuracy of 70% to 75% with metrics sitting under 0.3 for RELIANCE and TCS.

Index Terms - Stock market, forecasting models, predicting model, market closing price, hidden Markov model, ARIMA, MLP, Random Forest, RDA.

I. INTRODUCTION

Accurately predicting the stock market remains a great hurdle in the world of financial market. Traditional forecasting methods have often grappled with the intricacies and uncertainties of market dynamics, frequently resulting in less-than-optimal results. The consequences of inaccurate predictions can be significant in areas like investment strategies, financial decisions and economic stability. And so due to that, there is a pressing need to explore innovative approaches that can enhance our ability to predict stock prices with higher precision and reliability. This research delves into the concept of hybrid Machine-Learning (ML) models as a potential solution to this enduring problem.

The Main purpose of this study can be tracked down to the shortcomings of traditional models and techniques. formal methods such as Autoregressive Integrated Moving Average (ARIMA) may prove insufficient in capturing the intricate patterns and non-linear relationships that are present in the financial markets. To address these challenges, we turn our attention to hybrids models, which fuse together diverse ML models and approaches. These hybrid models have the possibility to finally close the gap between the complexities of financial markets & limitations of conventional forecasting methodologies.

Insights from recent researches offer evidence of the potential of hybrid models when tested within European financial and cryptocurrency markets. Exponential Smoothing Artificial Neural Network (ETS-ANN) and ARIMA show great promise in enhancing investment decisions despite the subpar accuracy metrics [1]. Furthermore, another comprehensive overview identifies the emergence of hybrid artificial intelligence (AI) approaches as one of the most promising directions. The emphasis on the importance of the findings, insights and future research makes the analysis critical for advancing stock market forecasting using ML [2]. We can also look over to novel models which aim to overcome the limitations mentioned before working within a unique situation [3]. The potential of all these approaches lies in guiding investment decision and influencing the stock market so to overall contribute to the efficiency and stability of the financial sector.

II. LITERATURE REVIEW

In the work done for stock market forecasting and prediction, the mission for enhanced accuracy and reliability has led researchers to explore a multitude of innovative models. One such model, a PCA-based hybrid, offers an improved approach to predicting the patterns of the Nifty 50 index. Combining technical indicators with support vector classification (SVC) and random forest (RF), this approach demonstrates high accuracy, F1-score, and AUC scores when applied to the Nifty 50 index data [4]. An extensive review of stock price forecasting literature reveals a diverse landscape of prediction methods categorized by model & feature perspectives. The models extend pre-established methods like statistical analysis, traditional machine learning, and deep learning, while the features encompass combinations of numeric and lexical data. The collective findings suggest a mosaic of methodologies that necessitate deeper exploration and examination to address stock price prediction challenges [5]. Moving further into the world of stock prediction, a comprehensive review unfolds the latest developments in hybrid deep learning approaches. These approaches, including ARIMA, Long-Short Term Memory(LSTM) and various Hybrid Convolutional Neural Network(CNN) models, aim to optimize short-term trading returns. A detailed examination of these models reveals their potential for both accurate stock rate prediction and trend identification, underlining the advantage of hybrid models in precise stock forecasting [6]. Exploring alternative avenues, a novel stock price prediction approach is introduced, based on the concept of a continuous Time Hidden Markov Model (CT-HMM). This model, initially a first-order HMM, evolves into a second-order CT-HMM, accommodating price fluctuations effectively. The model's performance is validated against the Hang Seng Index (HSI), consistently outperforming three benchmark models based on RMSE, MAE, and R2 evaluation metrics [7].

The adoption of Hidden Markov Models (HMM) emerges as a promising strategy for predicting stock prices, particularly for selected manufacturing companies on the Nigerian Stock Exchange. The HMM is shown to accurately forecast closing prices for a 60-day period, with superior prediction accuracy compared to ARIMA and neural network (NN) models. This research suggests the potential adoption of HMM for more accurate stock price forecasting [8]. From these diverse approaches and insights into the complex world of stock price prediction, the path forward for enhancing accuracy and precision in this domain becomes increasingly promising. Diving deeper into the landscape of stock price prediction, another study centers on high-frequency financial data, proposing two innovative models. The Fast Recurrent Neural Network (Fast RNN) stands as the first, specifically designed for stock price forecasting and prediction. The second model, a hybrid, combines Fast RNNs, CNNs, and Bidirectional LSTM architectures, aiming to predict sudden alteration in stock prices. When evaluated using 1-minute interval data from four companies, these models demonstrate superior performance with lower Root Mean Squared Error (RMSE) and computational complexity as compared to conventional models like ARIMA and LSTM, making them suitable for live predictions of stock values [9]. Amid the rising significance of deep learning approaches in stock and forex prediction, a comprehensive survey highlights the trend of combining LSTM with other methods, showcasing the surge in importance of DL in financial modeling [10]. The significance of stock price prediction extends beyond the research community, as investors seek higher profits and deploy various strategies. A bibliometric study delves into the methods used for stock price prediction, revealing opportunities for further research and exploring both statistical and artificial intelligence-based approaches [11]. As technological advancements continue to shape the field of financial market forecasting, novel hybrid models come to the forefront. In this context, CNN-LSTM, Gated Recurrent Unit combined with CNN, and ensemble models using recurrent neural RNN-based architectures play a central role in forecasting both one-time-step and multi-time-step of the closing prices of various indices. These models outperform traditional machine-learning counterparts, showcasing improvements with respect to mean squared error (MSE) and mean absolute error (MAE) [12].

A multi-parameter forecasting model focusing on stock price forecasting introduces a model for predicting close and high prices, with the help of a single Layer RNN model. The model achieves superior performance over existing methods such as CNN-RNN & CNN-LSTM, demonstrating its effectiveness in aiding investment decisions [13]. In the pursuit of accurate price prediction, a hybrid CNN-LSTM model is also shown which leverages convolutional layers for feature extraction and LSTM for long-term dependency learning. This model outperforms SOTA methods and underscores the superiority of deep learning techniques in financial modeling [14]. An innovative approach that combines BiLSTM with an improved Transformer model advances stock price prediction. Known as BiLSTM-MTRAN-TCN, this model achieves remarkable performance improvements in terms of RMSE and R2, demonstrating its robustness and accuracy in predicting stock prices across various time periods [15]. Adding to the growing body of research, a hybrid information mixing module is introduced to predict stock price movement by combining stock and news data. Multimodal interactions between time-series price data and semantic text data are extracted using a multilayer perceptron-based model. Experiments conducted on extremely volatile stock markets are used to confirm this method's accuracy, F1 score for performance evaluation, and Matthew's correlation coefficient (MCC). [16]. In the quest for efficient stock price trend prediction, a hybrid machine learning approach known as LT-SMF is introduced. By using brown planthopper and butterfly optimization techniques for data preprocessing and feature selection. This method significantly outperforms SOTA models and achieves higher accuracy for social media data evaluation. [17].

The investigation of hybrid models includes the incorporation of Hidden Markov Model (HMM) and RNN for stock price prediction. By addressing the sensitivity of the initial parameters, this method improves the classification and state detection of HMMs, which leads to increased accuracy in stock price prediction. [18]. Efficiency and speed are critical for high-frequency trading (HFT). An efficient online time series forecasting method, O-LGT, is introduced, integrating LSTM, GRU, and transformer models. O-LGT utilizes output from previous trading data and current data for immediate future forecasts, ensuring rapid updates.

Evaluation using Chinese market high-frequency limit order book (LOB) data demonstrates O-LGT's comparable speed with significantly enhanced accuracy, making it suitable for HFT [19]. Another great model for stock price prediction is I-NSGA-II-RF. This model combines a three-stage feature engineering process with the random forest algorithm to address challenges posed by complex stock price data. It outperforms other algorithms in terms of accuracy and runtime [20]. hybrid Recurrent Neural-Network (HyRNN) significantly improves predictive performance by integrating stock features and financial news sentiment, it, outperforming earlier models like GRU and LSTM [21]. ANN's stock price parameters can be optimized using the Barnacles Mating Optimizer (BMO). Significant superiority is shown by the BMO-ANN hybrid model, demonstrating its usefulness in parameter tuning for increased predictive accuracy. [22].

Through a hybrid model that combines LSTM, Gated Recurrent Units (GRU), and 1D Convolutional Neural Networks (1D-CNN). What is obtained is an effective stock price forecasting with superior prediction accuracy compared to LSTM and other models [23]. The Malaysia Counsellor Performance Indicator (M-CPI) aims to identify key components for measuring counselor performance. Employing a mixed-method approach, it gathers quantitative data from 102 respondents and qualitative insights from interviews with eight counselors [24]. Getting back to stock prediction, a hybrid framework known as DAELM (DWT-AE-ELM) is a framework that combines Discrete Wavelet Transformation (DWT) with Extreme Learning Machines (ELM), and Auto Encoder (AE) with feature penalty. It denoises raw data using DWT and trains an AE with feature penalty, using its encoder for feature extraction in training the ELM model. Results from a dataset of 400 stocks show improved prediction accuracy and a profitable investment strategy compared to buy-and-hold strategies, highlighting the potential of this hybrid framework in stock market prediction [25]. To address the challenges thown forward by the concentrated nature of volatility data distribution, VU-GARCH-LSTM is a model that combines the filtering approach with LSTM to predict realized volatility. It outperforms existing hybrid models, achieving a significant improvement in root mean square error (RMSE) and enhancing accuracy, especially in the domain region near zero volatility [28]. A hidden Markov model can also be employed to forecast price movements in diverse liquidity markets, such as Thailand and Korea exchanges. Results suggest that market efficiency creation occurs over a short period, with order imbalance indicators predicting price movements. This approach excels in lower liquidity markets, achieving an impressive 0.83 average hit ratio for positive price movements at a 5-minute frequency [27].

Stock price forecasting can also be a challenging task due to the intricate nature of stock data. This challenge can be addressed by introducing a novel hybrid deep learning model that integrates an attention mechanism, multi-layer perceptron, and bidirectional long-short term memory neural network. The model enhances forecasting performance by extracting valuable information from diverse datasets, improving focus on critical temporal data, and outperforming seven baseline models, contributing to investment and national policy research [28]. Shifting gears, a hybrid model integrating fractional order derivatives and deep learning, specifically long-short term memory (LSTM) networks, is proposed for predicting fast-fluctuated and high-frequency financial data. The model combines ARFIMA model-based filters with LSTM networks to capture both linear tendencies and nonlinearity, addressing volatility and overfitting issues. Evaluations using PSX company data demonstrate significant improvements in forecasting accuracy, highlighting the potential of this hybrid model [29]. Deep learning has becoming more popular in the big data era for predicting stock market trends and prices. A study presents a comprehensive method that combines a specialized deep learning model with customized feature engineering. The extensive feature engineering of the suggested solution sets it apart and yields greater accuracy for stock market trend prediction. By providing thorough insights into prediction term lengths, feature engineering, and data pre-processing approaches, it benefits both the financial and technological domains. [30]. Another study addresses stock price volatility prediction by introducing a novel Hybrid Time-Series Predictive Neural Network (HTPNN) that incorporates news impact. HTPNN captures stock price trends by learning from news and time series, enhancing accuracy while maintaining computational efficiency [31]. An extended hidden semi-Markov model is also introduced for chart pattern matching (HSMM-CP) in financial time series. This model is able to utilize a simplified training process and the Viterbi algorithm for pattern detection, achieving higher accuracy and recall on a set of 53 chart patterns, particularly excelling in distinguishing similar shapes like "Head-and-Shoulders Tops" and "Triple Tops" [32].

Getting back to novel models, another study employs LSTM recurrent neural networks (RNN) in conjunction with the ARIMA model to predict the correlation coefficient of stock prices for portfolio optimization. This hybrid approach effectively captures both temporal dependencies and linear tendencies in the data, outperforming traditional models in predictive accuracy. The ARIMA-LSTM model offers promising potential for enhancing correlation forecasting and optimizing investment portfolios [33]. In another novel hybrid model, PCA, ARIMA, and BP neural network are combined for predicting the Shanghai Composite Index. This model's effectiveness is demonstrated through improved prediction precision, providing guidance for investors to mitigate risks and enhance benefits [34]. A combined DT-ANN model achieves high accuracy, surpassing the individual performance of both single ANN and DT models, particularly in the electronic industry [35]. The variety of data types makes it difficult to predict data flow accurately when it comes to automation. A study presents REMD-LSTM, an innovative time-series forecasting approach, to improve model versatility and precision. This method combines long short-term memory (LSTM) with recursive empirical mode decomposition (REMD). By addressing issues with conventional decomposition techniques, REMD makes accurate predictions for a range of data kinds possible. The suggested model outperforms state-of-the-art models in terms of accuracy and versatility across a variety of datasets [36]. A novel hybrid model is put forth for stock price forecasting, combining long-short term memory (LSTM) and improved particle swarm optimization (IPSO). This novel model prevents premature convergence to local optima by introducing an adaptive mutation factor into the IPSO optimization process [37].

Coming to the end of the review, a novel stock market prediction model is introduced which encompasses feature extraction, optimal selection, and prediction phases. The model's performance surpasses conventional methods through precise stock movement prediction [38]. In the face of stock market volatility and non-linearity, precise prediction of returns is formidable. However, a study employs Artificial Neural Network and Random Forest methods for predicting next-day closing prices of diverse sector companies. Novel variables derived from financial data are used as model inputs, and evaluation based on RMSE and MAPE underscores the models' efficacy in accurate stock closing price prediction [39]. Finally, a novel Cyclic Attribution Technique (CAT) for feature selection in Human Activity Recognition (HAR) which leverages group theory and cyclic group properties to effectively reduce the feature set from 561 to 63. Tested on the UCI-HAR dataset, the CAT-enhanced model achieves a remarkable overall accuracy of 96.7%, addressing overfitting and reducing training time [40].

III. METHODOLOGY

3.1 Dataset Preparation

In this research, the datasets play a pivotal role in training and evaluating various stock price prediction models. The datasets were meticulously selected from finance.yahoo.com, a reliable source that provides real-time stock data for numerous companies. These datasets are publicly available for download at finance.yahoo.com in Comma Separated Values (CSV) format, ensuring their suitability for analysis. The research focuses on companies primarily from the Information Technology (IT) sector in India, including "HDB" (HDFC), "INFY_NS" (Infosys), "ITC_NS" (ITC), "RELIANCE_NS" (Reliance), and "TCS_NS" (Tata Consultancy Service). The choice of companies from the IT sector is driven by the high volatility and significance of this sector in the Indian stock market.

To ensure a comprehensive analysis and gain insights into model performance, three variations of each dataset were chosen. These variations encompass stock data from different timeframes:

- Datasets representing stock data from 2022 to 2023.
- Datasets representing stock data from 2021 to 2023.
- Datasets representing stock data from 2018 to 2023.

This diversity in dataset variations allows for a robust evaluation of the models, taking into account different market conditions and trends over the specified time periods.

Each dataset includes the following attributes:

Table 1: Description of dataset attributes

S. No	Attribute	Description
1	Date	Date that each entry's time label is derived from when the stock data instances were recorded.
2	Opening Price	The price of the stock at the start of the trading day.
3	Highest Price	The highest price at which the stock closed on a given trading day.
4	Lowest Price	The stock's lowest price during a trading day.
5	Closing Price	The finalized price of a stock at the end of the trading day.
6	Adjusted Closing Price	accurate representation of the stock's value by taking corporate actions into account.
7	Volume	Amount the quantity of shares or contracts exchanged in a given trading session, signifying the activity and liquidity of the market.

The "Adjusted Closing Price" emerged as a central feature for stock price prediction models. This attribute is particularly valuable because it encapsulates the corrected closing price of a stock, addressing some of the complexities and nuances inherent in stock market data. The "Adjusted Closing Price" involves various corporate decisions, such as stock splits, dividends, and new stock offerings. These actions can significantly impact a stock's closing price, and ignoring these adjustments could lead to inaccurate predictions. The "Adjusted Closing Price" allows the models to capture the true market value of a stock at the end of a trading day, irrespective of corporate actions. This ensures that the models are working with consistent and reliable data. It also aligns with the needs of investors and analysts who rely on accurate historical prices to make informed decisions. Moreover, stock prices are subject to various external factors and market dynamics, and the "Adjusted Closing Price" plays a crucial role in reflecting these influences. For instance, changes in interest rates, geopolitical events, economic indicators, and company-specific news can affect stock prices. By using this attribute, the models can better account for these complex factors. While the primary focus of the research was on predicting stock prices, the "Volume" attribute was included selectively in certain models. The "Volume" represents the number of shares or contracts traded during a particular instance, typically in a trading day. It is an essential indicator of market activity and liquidity, as higher trading volumes often signify greater market interest and participation. The inclusion of "Volume" was especially relevant for models that aimed to understand trading patterns and dynamics in the stock market. Analyzing trading volumes can reveal insights into investor sentiment, market sentiment, and potential buying or selling pressure. Models that used the "Volume" attribute focused on predicting the level of trading activity or other trading-related metrics, rather than directly forecasting stock prices. In summary, the choice of features for this research was guided by the need for accurate and comprehensive stock price prediction. The "Adjusted Closing Price" was central to account for corporate actions and accurately reflect the market value of a stock, while the "Volume" was selectively used to provide additional insights into trading patterns and market dynamics.

This feature selection strategy aimed to equip the models with the most relevant and informative data for their respective forecasting tasks, contributing to the overall effectiveness of the research.

Selecting which models to use and which models are capable of hybridization comes after choosing our datasets and the attributes that each model will be working on. It should be noted that some models, like support vector machines (SVM) and linear regression, are not very effective or have extremely complicated architectures when it comes to hybridization. That's not to say that these models cannot hybridize; in fact, as was previously indicated, many of these models have been applied to produce fresh, original solutions to problems. The models that are being used in this specific study are:

3.2 ARIMA-HMM

ARIMA is a well-known and reputable time series forecasting model. It is made up of three primary parts: Autoregressive (AR) component and time series' prior observations. The AR component models the relationship between the current observation and the observations of the time series makes predictions based on the its historical values. Integrated (I), which creates stationarity in the time series data by differentiating it. Finally, the relationship between the current observation and previous prediction errors is modelled by the Moving Average (MA). It facilitates the time series' ability to capture transient dependencies. The forecasting equation is constructed as follows. First, let y denote the d^{th} difference of Y, which can be defined as,

$$\text{If } d=0: \quad y_t = Y_t \qquad (3.2.1)$$

$$\text{If } d=1: \quad y_t = Y_t - Y_{t-1} \qquad (3.2.2)$$

$$\text{If } d=2: \quad y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2} \qquad (3.2.3)$$

Using equation (3.2.3) the general equation constructed for Y is,

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \ldots - \theta_q e_{t-q} \qquad (3.2.4)$$

Determining the (p, d, q) values for an ARIMA model involves a systematic process. Start by visually inspecting the time series data for trends, seasonality, and stationarity. Ensure stationarity by differencing (d) the data as needed. Next, analyse the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots to estimate the autoregressive (p) and moving average (q) orders. Conduct a grid search, fitting ARIMA models with various (p, d, q) combinations, and select the best-fitting model based on evaluation metrics like AIC and BIC. Verify the model by examining residuals for autocorrelation and normality. Fine-tune the model as needed, and validate its performance using out-of-sample data or cross-validation techniques. Domain knowledge and iterative refinement are essential in this process, and automated selection methods can aid in simplifying the model parameter selection.

HMM is a probabilistic model used for modelling sequences of data, such as time series, where there may be underlying hidden states. An HMM consists of two main components –

- Hidden States: These are unobservable states that evolve over time. In the context of time series, hidden states could represent different regimes or patterns in the data.
- Observable Emissions: These are the observed data points associated with each hidden state. In the context of time series, emissions represent the actual data values at each time step.

3.3 Multilayer Perceptron (MLP-HMM)

MLP is an artificial neural network model that serves as the foundational architecture for a number of applications, such as pattern recognition, regression, and classification.

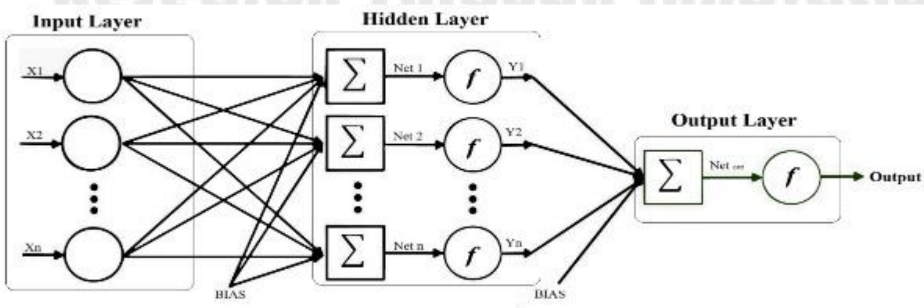

Figure 1: Multi-layer perceptron architecture

3.3.1 Input Layer

In the input layer, individual input features or variables from the dataset are represented by neurons. The dimensionality of the input data determines how many neurons are present in the input layer. The input layer merely forwards the input values to the subsequent layer without performing any calculations.

3.3.2 Hidden Layers

Hidden layers process and modify the incoming data. The neurons in the layer above provide weighted inputs to every neuron in a hidden layer. The result of the neuron is the sum of these weighted inputs multiplied by an activation function. The architecture can change depending on the particular problem, as can the number of hidden layers and neurons in each layer. Many hidden layers, each responsible for hierarchical feature extraction, make up deep multilayer parsers.

3.3.3 Output Layer

The MLP's output layer, which generates the network's output, is the last layer. The type of task determines how many neurons are in this particular layer:

- One neuron with a sigmoid function is utilized for binary classification, resulting in a probability between 0 and 1.
- One neuron is utilized for each class in multi-class classification, and class probabilities that add up to one are generated via the application of the softmax activation function.
- For regression, the output layer may contain multiple neurons for different output values or a single neuron with a linear activation function.

3.3.4 Weights and Connections

Each connection between neurons in adjacent layers is associated with a weight. Weights determine the strength of the connection, and they are learned during training. The weighted sum of inputs from the previous layer is computed for each neuron in the current layer, and this sum is used to compute the neuron's output. Weight updates through backpropagation are crucial for the network to learn to make accurate predictions.

3.3.5 Activation Functions

The network can learn intricate relationships in the data by introducing non-linearity through activation functions. Typical activation functions consist of:

Table 2: Common Activation Functions

S. No	Function Name	Range	Usage
1	Sigmoid	(0, 1)	Binary classification
2	Hyperbolic Tangent	(-1, 1)	Vanishing gradient
3	ReLU	(0, x)	Hidden Layer for efficiency
4	Softmax	(0, 1)	Multiclass classification
5	Linear	$(-\infty,+\infty)$	Regression model

3.4 Random Forest (RF-HMM)

Random Forest (RF) is an ensemble learning model which is applied to both classification and regression problems. Its foundation is the notion that by combining several decision trees, overfitting can be decreased and prediction accuracy can be increased.

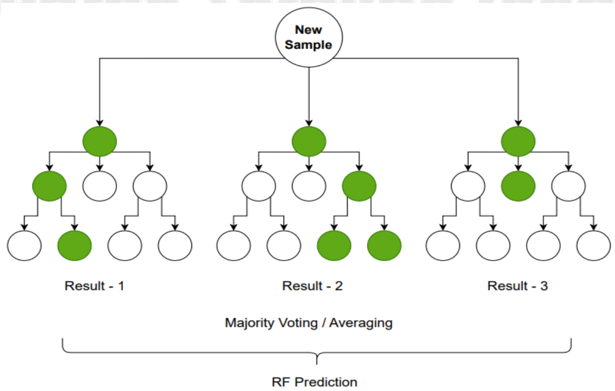

Figure 2: Random Forest Architecture

The three main hyperparameters of RF algorithms have to be set prior to training. These consist of the node size, number of trees and the number of features sampled. Regression and classification issues can then be resolved using the random forest classifier.

All sub trees are made up of a bootstrap sample, which is a sample data from the training set and its replacement. 1/3 of this sample is designated as test data; this is referred to as the out-of-bag (oob) sample. Feature bagging is then used to introduce yet another randomness, increasing dataset diversity and decreasing decision tree correlation. The prediction's determination will change depending on the kind of problem. The individual decision trees in a regression task will be averaged, and in a classification task, the predicted class will be determined by a majority vote, or the most frequent categorical variable. Lastly, that prediction is confirmed through cross-validation using the oob sample.

3.5 The Red Deer Adopted Wolf Algorithm (RDAWA)

The Red Deer Adopted Wolf Algorithm (RDAWA) hybrid model is a combination of two different optimization techniques: Red Deer Algorithm (RDA) and Grey Wolf Optimization (GWO). Both RDA and GWO are metaheuristic optimization algorithms used to find solutions to complex optimization problems. When they are combined into a hybrid model, it aims to harness the strengths of both algorithms to potentially improve optimization performance [41].

3.5.1 Red Deer Algorithm (RDA)

The Red Deer Algorithm makes its roots by taking inspiration from the behaviour of red deer stags during the rutting season. During this season, red deer stags use different strategies to optimize their reproductive success. The model recreates these strategies as optimization algorithms, including aggressive exploration and exploitation phases. It is often used for solving global optimization problems and has been shown to have good exploration and exploitation capabilities.

3.5.2 Grey Wolf Optimization (GWO)

Grey Wolf Optimization (GWO) is another metaheuristic algorithm that draws inspiration from the social structure and hunting techniques of grey wolves. With alpha, beta, and delta wolves, the algorithm here resembles the hierarchical structure of a wolf pack. The most well-known solutions are embodied by these wolves. The model looks for the best answers in a problem space based on the positions and motions of the wolves. It is renowned for being effective and straightforward and is frequently used for optimization tasks.

3.6 Recursive Empirical Mode Decomposition (REMD)

Recursive Empirical Mode Decomposition (REMD) is an extension of the Empirical Mode Decomposition (EMD) method, which is a signal processing technique used for analysing and decomposing non-linear and non-stationary time series data. REMD was developed to overcome some of the limitations of the original EMD method [36]. EMD main objective is to break down the complex signals into centralized Functions called Intrinsic mode Functions (IMFs), but it faces challenges like mode mixing and boundary effects. REMD addresses these issues by applying EMD iteratively, refining the decomposition with each step. This recursive nature enhances stability and consistency, particularly in noisy data. REMD is applied in fields such as signal processing, finance, and biomedical signal analysis, where non-linear and hidden patterns are sought in complex time series data, although it may come at a higher computational cost. Careful parameter selection and stopping criteria are essential for optimal results.

3.7 Hidden Semi-Markov Models and Chart Pattern Prediction (SAHM)

Combining Hidden Semi-Markov Models (HSMMs) with chart pattern recognition, this model specialized in predicting stock price volatility by identifying chart patterns in financial time series data.

3.7.1 Hidden Semi-Markov Models (HSMMs)

Hidden Semi-Markov Models (HSMMs) are probabilistic models that extend Hidden Markov Models (HMMs) by allowing states to persist for variable durations. Unlike HMMs, which assume fixed-duration states, HSMMs enable each state to emit multiple observations before transitioning, making them particularly useful in fields like speech recognition, natural language processing, bioinformatics, and finance. HSMMs incorporate state duration modelling parameters, such as duration probabilities and distributions, which enable the modelling of sequences with flexible state durations. Their applications range from speech phonemes and linguistic units to biological sequences with variable-length motifs and market regimes in financial time series data. Training involves parameter estimation, and inference often includes decoding to find the most likely sequence of hidden states given observed data [32].

3.7.2 Chart Pattern Recognition

Chart pattern recognition is a technique used in technical analysis of financial markets to identify recurring patterns in price charts, such as head and shoulders, triangles, flags, and many others. Traders and analysts use chart patterns as visual tools to make trading decisions, as these patterns often provide insights into potential future price movements. By recognizing these patterns, traders seek to predict market trends, identify entry and exit points, and manage risk. Chart pattern recognition involves pattern identification, confirmation through volume and other indicators, and the application of various trading strategies based on the specific patterns observed, ultimately aiding in the development of informed trading decisions in financial markets.

Majority of the models chosen have the implementation of HMM as the other half of the hybrid model. HMMs are well-suited for creating hybrid models in the context of machine learning and statistical modelling due to their specific characteristics and versatility. Their ability to represent hidden states, model state transitions, and work with uncertainty makes them an excellent choice for combining with other models, such as neural networks or Bayesian networks, to create hybrid models that benefit from the strengths of multiple approaches. By incorporating specialized domain knowledge and features, these hybrid models can improve interpretability, predict sequential data more effectively, and handle a wide range of applications, including speech recognition, sensor fusion, and anomaly detection [42].

IV. RESULTS

The provided data shows how well different hybrid models performed in forecasting the stock prices of five significant corporations. (RELIANCE, TCS, HDFC, ITC, and INFOSYS) over three different time periods: 2022-2023, 2021-2023, and 2018-2023. These models, which combine traditional time series methods and machine learning techniques, have been evaluated using different metrics to assess their accuracy, robustness, and flexibility.

Table 3: Error Metrics (2022 – 23)

ARIMA-HMM			
	MAE	**MSE**	**RMSE**
Reliance	0.24	0.13	0.25
TCS	0.25	0.13	0.25
HDFC	0.25	0.12	0.25
ITC	0.25	0.12	0.25
Infosys	0.25	0.13	0.26
MLP-HMM			
	MAE	**MSE**	**RMSE**
Reliance	0.17	0.69	0.18
TCS	0.22	0.10	0.23
HDFC	0.29	0.18	0.30
ITC	0.43	0.55	0.52
Infosys	0.05	0.78	0.06
RF-HMM			
	MAE	**MSE**	**RMSE**
Reliance	0.20	0.15	0.27
TCS	0.15	0.10	0.23
HDFC	0.48	0.87	0.66
ITC	0.72	0.17	0.92
Infosys	0.09	0.37	0.13
RDAWA			
	MAE	**MSE**	**RMSE**
Reliance	0.22	0.16	0.22
TCS	0.22	0.15	0.22
HDFC	0.64	0.13	0.64
ITC	0.28	0.28	0.28
Infosys	0.14	0.69	0.14
REMD			
	MAE	**MSE**	**RMSE**
Reliance	0.12	0.31	0.12
TCS	0.16	0.52	0.16
HDFC	0.32	0.21	0.32
ITC	0.19	0.74	0.19
Infosys	0.70	0.98	0.70
SAHM			
	MAE	**MSE**	**RMSE**
Reliance	0.12	0.31	0.12
TCS	0.16	0.52	0.16
HDFC	0.32	0.21	0.32
ITC	0.19	0.75	0.19
Infosys	0.70	0.98	0.70

Table 4: Error Metrics (2021 – 23)

	ARIMA-HMM		
	MAE	**MSE**	**RMSE**
Reliance	0.25	0.12	0.25
TCS	0.25	0.13	0.25
HDFC	0.25	0.14	0.26
ITC	0.23	0.17	0.29
Infosys	0.24	0.13	0.25
	MLP-HMM		
	MAE	**MSE**	**RMSE**
Reliance	0.23	0.21	0.32
TCS	0.17	1.02	0.22
HDFC	0.81	0.19	0.99
ITC	0.36	0.40	0.44
Infosys	0.12	0.50	0.15
	RF-HMM		
	MAE	**MSE**	**RMSE**
Reliance	0.14	0.67	0.18
TCS	0.13	0.70	0.18
HDFC	0.46	0.81	0.64
ITC	0.27	0.18	0.30
Infosys	0.30	0.31	0.39
	RDAWA		
	MAE	**MSE**	**RMSE**
Reliance	0.25	0.20	0.25
TCS	0.27	0.23	0.27
HDFC	0.62	0.13	0.62
ITC	0.27	0.24	0.27
Infosys	0.15	0.85	0.15
	REMD		
	MAE	**MSE**	**RMSE**
Reliance	0.70	0.98	0.70
TCS	0.16	0.55	0.16
HDFC	0.32	0.21	0.32
ITC	0.15	0.51	0.15
Infosys	0.74	0.11	0.74
	SAHM		
	MAE	**MSE**	**RMSE**
Reliance	0.12	0.30	0.12
TCS	0.16	0.54	0.16
HDFC	0.32	0.21	0.32
ITC	0.15	0.52	0.15
Infosys	0.74	0.11	0.74

Table 5: Error Metrics (2028 – 23)

	ARIMA-HMM		
	MAE	**MSE**	**RMSE**
Reliance	0.24	0.12	0.25
TCS	0.25	0.12	0.25
HDFC	0.25	0.15	0.27
ITC	0.23	0.15	0.28
Infosys	0.25	0.12	0.25
	MLP-HMM		
	MAE	**MSE**	**RMSE**
Reliance	0.17	0.11	0.24
TCS	0.20	0.14	0.26
HDFC	0.61	0.12	0.79
ITC	0.48	0.92	0.68
Infosys	0.12	0.52	0.16
	RF-HMM		
	MAE	**MSE**	**RMSE**
Reliance	0.14	0.56	0.18
TCS	0.15	0.79	0.19
HDFC	0.45	0.67	0.58
ITC	0.33	0.36	0.42
Infosys	0.89	0.30	0.12
	RDAWA		
	MAE	**MSE**	**RMSE**
Reliance	0.21	0.16	0.21

TCS	0.24	0.19	0.24
HDFC	0.67	0.16	0.67
ITC	0.22	0.17	0.22
Infosys	0.11	0.52	0.11
REMD			
	MAE	**MSE**	**RMSE**
Reliance	0.97	0.20	0.97
TCS	0.13	0.37	0.13
HDFC	0.29	0.18	0.29
ITC	0.12	0.33	0.12
Infosys	0.55	0.70	0.55
SAHM			
	MAE	**MSE**	**RMSE**
Reliance	0.97	0.20	0.97
TCS	0.13	0.37	0.13
HDFC	0.29	0.18	0.29
ITC	0.12	0.33	0.12
Infosys	0.55	0.70	0.55

The ARIMA-HMM model generally demonstrates an average accuracy 80% with relatively low MAE, MSE, and RMSE values across all time periods. This suggests that it provides accurate predictions for the stock prices of the five companies as compared with other models. MLP-HMM performs well in terms of accuracy, especially in the 2018-2023 dataset, where it has the lowest RMSE as given in the table. However, it is less accurate in the 2022-2023 dataset, as indicated by higher MAE and RMSE values. This model exhibits varying accuracy levels depending on the dataset. The RF-HMM model has mixed accuracy results. It performs well in the 2018-2023 dataset but struggles in the 2022-2023 dataset, with the highest RMSE. Hence similar to the previously mentioned model, its accuracy may not be consistent across different time periods. RDAWA consistently provides and average accuracy of 70% to 75% predictions with low MAE, MSE, and RMSE values. It appears to be a robust and reliable model in terms of accuracy. The REMD model's accuracy varies significantly, with high RMSE values in the 2018-2023 dataset, indicating less accurate predictions. It is, however, more accurate in the other two datasets, as reflected in lower RMSE values. SAHM exhibits relatively good accuracy. especially in the 2018-2023 dataset. However, like REMD, its accuracy varies across time periods.

ARIMA-HMM RDAWA

Figure 3: Graph between predicted values and actual values for ARIMA-HMM and REMD

MLP-HMM's robustness is variable. It performs well in the 2018-2023 dataset but struggles in the 2022-2023 dataset, indicating that it may not be as robust in adapting to changing conditions. RF-HMM exhibits variable robustness, with strong performance in the 2018-2023 dataset but weaker performance in the 2022-2023 dataset. RDAWA is consistently robust, with accurate and reliable performance across all three time periods. Both REMD's and SAHM demonstrates variable robustness. RDAWA performs well in the 2022-2023 and 2021-2023 datasets, whereas on the other hand SAHM struggles in the 2018-2023 dataset. It shows strong performance in the 2018-2023 dataset but weaker performance in the 2022-2023 dataset.

ARIMA-HMM is relatively flexible, as it consistently provides accurate predictions across different time periods, indicating its adaptability to changing conditions. RDAWA too can be considered flexible as it consistently provides accurate predictions across all time periods, indicating its adaptability.

V. CONCLUSION AND FUTURE WORK

The ARIMA-HMM model is a dependable option for predicting throughout a variety of time periods due to its consistent accuracy and robustness, as demonstrated by the examination of numerous stock price prediction models. On the other hand, the accuracy and robustness of the MLP-HMM, RF-HMM, and SAHM models fluctuate, indicating a limited ability to adjust to changing situations. RDAWA constantly proves to be an accurate and versatile model, showcasing its dependability and versatility. The performance of the REMD model changes with time; it is inconsistently flexible, performing well in some instances but falling short in others. These results offer insightful information about the advantages and disadvantages of each model, assisting in the choice of the best strategy for stock price prediction.

Looking ahead, the promising results obtained from the ARIMA-HMM, RDAWA, and potentially other successful models in this study provide a solid foundation for future research in the field of stock price prediction. To further enhance the accuracy and reliability of forecasting models, it is imperative to explore more complex models and methodologies. Future research may involve the integration of advanced machine learning techniques, deep learning algorithms, and the incorporation of additional financial indicators and external data sources. Comparative analyses will be conducted to benchmark the performance of the established models against these more intricate approaches, with a keen focus on identifying the optimal trade-offs between complexity and predictive power.

REFERENCES

[1] Ampountolas, A. (2023). Comparative Analysis of Machine Learning, Hybrid, and Deep Learning Forecasting Models: Evidence from European Financial Markets and Bitcoins. Forecasting, 5(2), 472-486.

[2] Strader, T. J., Rozycki, J. J., Root, T. H., & Huang, Y. H. J. (2020). Machine learning stock market prediction studies: review and research directions. Journal of International Technology and Information Management, 28(4), 63-83.

[3] Chong, W. U., Peng, L. U. O., Yongli, L. I., & Chen, K. (2015). Stock price forecasting: hybrid model of artificial intelligent methods. Engineering Economics, 26(1), 40-48.

[4] Manjunath, C., Marimuthu, B., & Ghosh, B. (2023). Analysis of Nifty 50 index stock market trends using hybrid machine learning model in quantum finance. International Journal of Electrical and Computer Engineering (IJECE), 13(3), 3549-3560.

[5] Zhao, C., Wu, M., Liu, J., Duan, Z., Shen, L., Shangguan, X., ... & Wang, Y. (2023). Progress and Prospects of Data-Driven Stock Price Forecasting Research. International Journal of Cognitive Computing in Engineering.

[6] Shah, J., Vaidya, D., & Shah, M. (2022). A comprehensive review on multiple hybrid deep learning approaches for stock prediction. Intelligent Systems with Applications, 200111.

[7] Su, Z., & Yi, B. (2022). Research on HMM-Based Efficient Stock Price Prediction. Mobile Information Systems, 2022.

[8] Amiens, E. O., & Osamwonyi, I. O. (2022). Stock price forecasting using hidden Markov model. International Journal of Information and Decision Sciences, 14(1), 39-59.

[9] Yadav, K., Yadav, M., & Saini, S. (2022). Stock values predictions using deep learning based hybrid models. CAAI Transactions on Intelligence Technology, 7(1), 107-116.

[10] Hu, Z., Zhao, Y., & Khushi, M. (2021). A survey of forex and stock price prediction using deep learning. Applied System Innovation, 4(1), 9.

[11] Tupe-Waghmare, P. (2021). Prediction of Stocks and Stock Price using Artificial Intelligence: A Bibliometric Study using Scopus Database.

[12] Song, H., & Choi, H. (2023). Forecasting Stock Market Indices Using the Recurrent Neural Network Based Hybrid Models: CNN-LSTM, GRU-CNN, and Ensemble Models. Applied Sciences, 13(7), 4644.

[13] Zaheer, S., Anjum, N., Hussain, S., Algarni, A. D., Iqbal, J., Bourouis, S., & Ullah, S. S. (2023). A Multi Parameter Forecasting for Stock Time Series Data Using LSTM and Deep Learning Model. Mathematics, 11(3), 590.

[14] Abdullah W, Salah A. A novel hybrid deep learning model for price prediction. International Journal of Power Electronics and Drive Systems. 2023;13(3):3420. doi:10.11591/ijece.v13i3.pp3420-3431

[15] Wang, S. (2023). A Stock Price Prediction Method Based on BiLSTM and Improved Transformer. IEEE Access.

[16] Choi, J., Yoo, S., Zhou, X., & Kim, Y. (2023). Hybrid Information Mixing Module for Stock Movement Prediction. IEEE Access, 11, 28781-28790.

[17] Venkateswararao, K., & Reddy, B. V. R. (2023). LT-SMF: long term stock market price trend prediction using optimal hybrid machine learning technique. Artificial Intelligence Review, 56(6), 5365-5402.

[18] Aydogan-Kilic, D., & Selcuk-Kestel, A. S. (2023). Modification of hybrid RNN-HMM model in asset pricing: univariate and multivariate cases. Applied Intelligence, 1-22.

[19] Li, C., Shen, L., & Qian, G. (2023). Online Hybrid Neural Network for Stock Price Prediction: A Case Study of High-Frequency Stock Trading in the Chinese Market. Econometrics, 11(2), 13.

[20] Zeng, X., Cai, J., Liang, C., & Yuan, C. (2023). Prediction of stock price movement using an improved NSGA-II-RF algorithm with a three-stage feature engineering process. Plos one, 18(6), e0287754.

[21] John, A., & Latha, T. (2023). Stock market prediction based on deep hybrid RNN model and sentiment analysis. Automatika, 64(4), 981-995.

[22] Mustaffa, Z., & Sulaiman, M. H. (2023). Stock price predictive analysis: An application of hybrid Barnacles Mating Optimizer with Artificial Neural Network. International Journal of Cognitive Computing in Engineering, 4, 109-117.

[23] Naufal GR, Wibowo A. Time series forecasting based on deep learning CNN-LSTM-GRU model on stock prices. International Journal of Engineering Trends and Technology. 2023;71(6):126-133. doi:10.14445/22315381/ijett-v71i6p215

[24] Ghazali, N. M., Yusoff, A., Wan Jaafar, W. M., Amat, S., Aden, E., & Anuar, A. (2023). Unique components of Malaysia Counsellor Performance Indicator (M-CPI): A tool to measure the performance of counsellors in Malaysia. International Journal of Industrial Engineering & Production Research, 34(2), 1-13.

[25] Wu, D., Wang, X., & Wu, S. (2022). A hybrid framework based on extreme learning machine, discrete wavelet transform, and autoencoder with feature penalty for stock prediction. Expert Systems with Applications, 207, 118006.

[26] Koo, E., & Kim, G. (2022). A hybrid prediction model integrating garch models with a distribution manipulation strategy based on lstm networks for stock market volatility. IEEE Access, 10, 34743-34754.

[27] Wu, P., & Siwasarit, W. (2020). Capturing the order imbalance with hidden markov model: A Case of SET50 and KOSPI50. Asia-Pacific Financial Markets, 27, 115-144.

[28] Chen, Q., Zhang, W., & Lou, Y. (2020). Forecasting stock prices using a hybrid deep learning model integrating attention mechanism, multi-layer perceptron, and bidirectional long-short term memory neural network. IEEE Access, 8, 117365-117376.

[29] Bukhari, A. H., Raja, M. A. Z., Sulaiman, M., Islam, S., Shoaib, M., & Kumam, P. (2020). Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting. Ieee Access, 8, 71326-71338.

[30] Shen J, Shafiq O. Short-term stock market price trend prediction using a comprehensive deep learning system. Journal of Big Data. 2020;7(1). doi:10.1186/s40537-020-00333-6

[31] Wang, Y., Liu, H., Guo, Q., Xie, S., & Zhang, X. (2019). Stock volatility prediction by hybrid neural network. IEEE Access, 7, 154524-154534.

[32] Wan, Y., & Si, Y. W. (2018). A hidden semi-Markov model for chart pattern matching in financial time series. Soft Computing, 22(19), 6525-6544.

[33] Choi, H. K. (2018). Stock price correlation coefficient prediction with ARIMA-LSTM hybrid model. arXiv preprint arXiv:1808.01560.

[34] Luo, H., & Wang, S. (2016). Based on the PCA-ARIMA-BP hybrid model of stock price prediction research. ANZIAM Journal, 58, E162-E178.

[35] Tsai, C. F., & Wang, S. P. (2009, March). Stock price forecasting by hybrid machine learning techniques. In Proceedings of the international multiconference of engineers and computer scientists (Vol. 1, No. 755, p. 60).

[36] Yang, Y., Fan, C., & Xiong, H. (2022). A novel general-purpose hybrid model for time series forecasting. Applied Intelligence, 52(2), 2212-2223.

[37] Ji, Y., Liew, A. W. C., & Yang, L. (2021). A novel improved particle swarm optimization with long-short term memory hybrid model for stock indices forecast. Ieee Access, 9, 23660-23671.

[38] Alotaibi, S. S. (2021). Ensemble technique with optimal feature selection for Saudi stock market prediction: a novel hybrid red deer-grey algorithm. IEEE Access, 9, 64929-64944.

[39] Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock closing price prediction using machine learning techniques. Procedia computer science, 167, 599-606.

[40] Myo, W. W., Wettayaprasit, W., & Aiyarak, P. (2019). A cyclic attribution technique feature selection method for human activity recognition. International Journal of Intelligent Systems and Applications, 11(10), 25.

[41] Zitar, R. A., Abualigah, L., & Al-Dmour, N. A. (2021). Review and analysis for the Red Deer Algorithm. Journal of Ambient Intelligence and Humanized Computing, 1-11.

[42] Altarawneh, G. A. (2019). Prediction of stock price using a hybrid technical analysis method. Science International Lahore, 31, 391-96.