Comparative Analysis of ARIMA, Deep Learning, and Lasso Regression Models for Time Series Forecasting: Assessing Accuracy, Robustness, and Computational Efficiency

Sanjay Kumar[1], Meenakshi Srivastava[1], Vijay Prakash[2]

[1] *Amity University Lucknow, India*
[2] *Babu Banarsi Das University Lucknow, India*

Abstract

This paper provides a comprehensive review of time-series forecasting models for forecasting the performance of Indian mutual funds. Specifically, we evaluate the effectiveness of three popular approaches: ARIMA, deep learning, and Lasso regression. Using a dataset of historical mutual fund data from the Indian market, we compare the predictive accuracy of these models using various evaluation metrics. Our findings indicate that Lasso regression outperforms both ARIMA and Deep Learning (LSTM) models in capturing the complex patterns and dynamics of mutual fund data. These findings offer valuable insights for investors and financial practitioners, shedding light on the most effective modeling approaches for predicting Indian mutual fund performance. This study contributes to the field of time series forecasting by providing a comprehensive comparison of ARIMA, Deep Learning, and Lasso Regression models. The findings can guide researchers and practitioners in selecting the most suitable model for specific forecasting tasks based on the desired balance between accuracy, robustness, and computational efficiency. The proposed research focuses on providing sustainability in investment domain. Lasso Regression models exhibit superior accuracy and competitive performance with a lower computational cost. The popular methods MAE, RMSE, MAE, R2 Score, MAPE, and MPE are used to measure the accuracy of the models.

Keywords

Time-series forecasting, performance analysis, ARIMA, deep learning, Lasso , regression, predictive models.

1. Introduction

Autoregressive Integrated Moving Average (ARIMA), deep learning, and Lasso regression. Each of these models presents distinctive benefits and methodologies for extracting valuable insights from time-series data. ARIMA, a traditional statistical model, has been widely employed in the domain of time-series prediction. It captures the linear dependencies and trends present in the data by incorporating parameters such as autoregressive (AR), differencing (I), and moving average (MA). The interpretability and simplicity of ARIMA make it a popular choice for forecasting in various domains. On the other hand, deep learning models have received considerable attention in recent years due to their ability to model complex non-linear relationships and time dependencies Long Short-Term Memory (LSTM) networks, which belong to the category of recurrent neural networks (RNNs), are particularly adept at capturing long-term dependencies and patterns in sequential data. Deep learning models have demonstrated

AISD 2023: 1st International Workshop on "Artificial Intelligence: Empowering Sustainable Development co-located with International Conference on Artificial Intelligence: Towards Sustainable Intelligence, September 4-5, 2023, Pune,India
k.sanjay123@gmail.com (S. Kumar); msrivastava@lko.amity.edu (M. Srivastava); vijaylko@gmail.com (V. Prakash)
https://dblp.org/pid/340/0794.html (M. Srivastava)

encouraging outcomes when applied to financial time-series data, showcasing their potential in forecasting stock prices, identifying market trends, and predicting various financial indicators. Sustainability in investment has gained significant traction in recent years as more investors recognize the importance of long-term sustainability for both financial returns and broader societal well-being. Various investment products, such as sustainable mutual funds, exchange-traded funds (ETFs), and green bonds, cater to investors looking to align their financial goals with their values. Ensuring sustainability in investment involves a combination of research, analysis, due diligence, and ongoing monitoring.

2. Review of Literature

In recent years, various time-series forecasting models have gained prominence in the financial domain for their potential to capture the complex dynamics of financial data.

Zeroual, A et al. [1] studies five deep learning models to forecast the new and recovered cases of COVID-19. VAE (Variational Autoencoder) algorithm shows superior performance among all. Benevento, E et al. [2] evaluate the predictive performance of lasso regression, random forest, support vector regression, artificial neural networks, and ensemble methods using a range of error metrics and computation time measurements. The results reveal that the ensemble method surpasses other approaches in accurately predicting latency. Zhang, L et al. [3] introduced, where the stochastic trend data is eliminated from the SSE Composite Index to obtain de-noised training data for the SVM (Support Vector Machine). Subsequently, the SVM is trained using this de-noised data to make predictions on the test data. The SVM achieves a 25% hit rate when predicting with the noisy training data. Guo, K et al. [4] applying the ARIMA model, both the original data series and the logarithmic series of the S&P 500 Exponential Weekly Data Series and found that model predicts accurate stock price. Pandey, A et al. [5] developed a model for investors to accurately forecast prices, regardless of the employed strategy. The primary objective of this research is to analyze and predict changes in the stock market. By examining past historical trends, the model aims to identify and forecast emerging patterns that will manifest in the upcoming days. Xu, Y et al. [6] presents a predictive analysis conducted across various economic cycles, uncovering that the social media sentiment index demonstrates the strongest predictive ability during periods of economic expansion. Dai, Z et al. [7] predicts stock earnings volatility by utilizing the partially least squares technique, which identifies crucial predictors from a data-rich context. The research findings illustrate the efficacy of the partial least squares approach in improving the accuracy of stock return volatility predictions in data-rich environments. This approach surpasses alternative models and exhibits a significant advancement over benchmark models .Ma, F et al. [8] proposes the use of dimensionality reduction and contraction techniques to forecast stock market returns. This research provides fresh insights into stock market return projections by considering macroeconomic fundamentals as a basis for analysis. Li, X et al. [9] proposes a MS-MIDAS-LASSO model that shows superior predictive accuracy compared to both the conventional LASSO strategy and its regime-switching extension. Notably, the outstanding predictive performance of this model remains unchanged even in the face of the onset of the COVID-19 pandemic. Ren, X et al. [10] identify that the Fourier transform-based LSTM method enhances the prediction accuracy of stock price fluctuation dynamics. This improvement is observed from both statistical and economic standpoints, as we exploit the role of oil shocks in the analysis. Zhu, Z and He, K [11] Finding the best models to predict stock price trends has always been a topic of great interest and is closely related to investor investment behavior. However, LSTM models still need to be improved in terms of performance to reduce distortion. We expect to discover more models for predicting stock prices in the future. Lee, H. Y et al. [12] purpose of this study was to extract valuable outlier information from the residuals of ARIMA modeling using the Continuous Wavelet Transform (CWT). The obtained CWT information was then incorporated into the ARIMA forecasts, resulting in the creation of long-term heterogeneous forecasts. Liu, T et al. [13] suggests a new stock price forecast model named VML with the aim of enhancing forecast accuracy and achieving improved forecast results. The proposed approach involves splitting the decomposed subseries into multiple tasks using the MAML algorithm. This facilitates the training of the LSTM model with initial parameters that possess strong generalization capabilities. Experimental outcomes obtained from Chinese and American stock market datasets demonstrate that the proposed method significantly enhances prediction accuracy. Nair, A. V and Narayanan, J [14] suggest a stock market forecasting model was suggested to anticipate the future performance of a company's stock. The incorporation of machine learning techniques represents the latest advancement in market analysis technology, enabling the determination of current stock index values by leveraging past values. Zeng, L et al. [15] proposes an optimal combinatorial framework for agricultural commodity price forecasting was introduced. This framework integrates a decomposition-reconstruction ensemble technique and an enhanced global optimization algorithm, inspired by natural processes. Wu, D et al. [16] introduces a hybrid stock market forecasting model that merges a multilayer artificial neural perceptron network (MLP-ANN) with the conventional Altman Z-score model. Empirical analysis demonstrates that the hybrid neural network model achieves a notable average correct classification rate. Isabona, J et al. [17] study indicate that the prediction errors of the suggested MLP model, when compared to the measured data, are highly favorable and surpass those obtained through the conventional logarithmic distance-based path loss model. Li, G et al. [18] proposes a technique called the PCC-BLS framework was suggested to choose multi-indicator functions for predicting stock prices. This approach utilizes the Pearson's correlation coefficient (PCC) and the broad learning system (BLS). Initially, PCC was employed to select input features from a pool of 35 options, which encompassed original stock prices, technical indicators, and financial indicators. Banerjee, S and Mukherjee, D [19] emphasis his study on the utilization of nonparametric approaches like stacked multilayer perceptions (MLP), long short-term memory (LSTM), and gated recurrent units (GRU). Specifically, long-term short-term bidirectional memory (BLSTM) and gated bidirectional recurrent units (BGRU) were employed to forecast short-term stock prices for three NSE-listed banks. The performance of these models was then compared against a flat neural network benchmark. Ji, X et al. [20] proposes a novel forecasting approach was introduced, which combines conventional financial indicators with social media text features as inputs for predictive models. Additionally, a unique stock price prediction model incorporating both traditional financial variables and social media text features extracted through deep learning methods was suggested in this study. Kumar, D [21] proposes that stock market prediction is a cohesive process, implying the need for a closer examination of specific parameters relevant to stock market forecasting. Tanwar, R et al. [22] proposed a hybrid deep learning approach, specifically a model combining Convolutional Neural Network and Long Short-Term Memory (CNN-LSTM), designed for the identification of stress. Tanwar, R et al.[23] introduced a hybrid deep learning model that incorporates an attention mechanism. This allows for thorough feature extraction and dynamic prioritization of information. Makwana, Y et al.[24] Conducts a comparative analysis of different methods and technologies, with a particular focus on the effectiveness of Convolutional Neural Network (CNN) in food recognition. The research reveals insights into various CNN models, showcasing their accuracy and outcomes in the context of food recognition.

3. Problem Statement

The problem at hand is the lack of a comprehensive assessment of time-series forecasting models for predicting the performance of Indian mutual funds. Although various approaches, such as ARIMA, deep learning (LSTM), and Lasso regression, have shown promise in other domains, their effectiveness and comparative performance in the context of Indian mutual funds remain unclear. The evaluation seeks to address this research gap by conducting a comprehensive assessment of the ARIMA, deep learning, and Lasso regression approaches.

i. This analysis will provide insights into the models' ability to accurately predict mutual fund performance.

ii. This evaluation will help determine the models' ability to adapt and provide reliable forecasts under different circumstances.

iii. This analysis will provide insights into how well the models can generalize their predictions beyond the training data and make accurate forecasts for unseen mutual fund performance.

4. Data for proposed model

This paper focuses on analyzing historical mutual fund data of TATAPOWER. The data, which can be obtained from the yahoo finance site, encompasses the period from January 1, 2011, to April 28, 2023. To facilitate analysis, the data is divided into training and testing segments, with 80% allocated for training and 20% for testing. Prediction tasks are then carried out on this dataset using ARIMA (0, 1, 0), Deep Learning (LSTM), and Lasso Regression models.

Table 1

Sample Dataset (TATAPOWER)

Date	Open	High	Low	Close	Adj Close	Volume
2011-01-03	133.558380	133.558380	132.014343	132.665741	102.871704	1747585
2011-01-04	132.506500	133.558380	131.584915	133.235092	103.313179	2267182
2011-01-05	132.979370	135.777908	132.120499	135.189255	104.828468	3228574
2011-01-06	134.619888	136.163925	133.321945	135.034851	104.708755	2761494
2011-01-07	133.881653	135.763443	132.796005	134.065002	103.956696	3027490
...	...	...	...	...	...	...
2023-04-24	196.500000	196.699997	194.800003	195.850006	194.042862	5017631
2023-04-25	195.850006	198.800003	195.350006	197.649994	195.826233	5957551
2023-04-26	197.649994	198.949997	196.149994	198.199997	196.371170	4910837
2023-04-27	198.449997	199.949997	197.649994	198.500000	196.668396	5215692
2023-04-28	199.500000	201.550003	199.000000	201.100006	199.244415	7951645

Dataset contains 3038 rows × 6 columns from TATAPOWER mutual fund from dated 201-01-03 to 2023-04-28.

5. Research methodologies

5.1. ARIMA (Autoregressive Integrated Moving Average)

The Autoregressive Integrated Moving Average (ARIMA) model is a commonly employed technique for time-series forecasting. It incorporates three essential components: auto regression (AR), differencing (I), and moving average (MA). The ARIMA model is defined by the order assigned to each component, denoted as ARIMA (p, d, q). In this notation, 'p' represents the autoregressive order, ’d’ represents the differencing order, and 'q' represents the moving average order.

5.1.1. Autoregressive Component (AR)

The autoregressive component of the model captures the linear association between the present observation and its previous values. The AR component of order p is represented by the equation:

$$AR(p): Xt = c + \Sigma(\phi i * Xt - i) + \varepsilon t \qquad (1)$$

Here, Xt represents the current observation, c is a constant term, ϕi represents the autoregressive coefficients for lagged values X t-i, and εt is the error term at time t.

5.1.2. Moving Average Component (MA)

The moving average component addresses the interdependence between the current observation and the error terms within the model. It acknowledges the relationship between them. The MA component of order q is represented by the equation:

$$MA\ (q): Xt = c + \varepsilon t + \Sigma(\theta i * \varepsilon t - i) \qquad (2)$$

Here, θi represents the moving average coefficients for the lagged error terms εt-i. Combining the three components, the ARIMA (p, d, and q) model is given by:

$$ARIMA\ (p, d, q): Xt = c + \Sigma(\phi i * Xt - i) + \varepsilon t + \Sigma(\theta i * \varepsilon t - i) \qquad (3)$$

The ARIMA model aims to estimate the optimal values of the parameters (p, d, q) that minimizes the disparity between the observed values and the predicted values. This estimation is commonly accomplished through techniques like maximum likelihood estimation.

6. Deep learning

Deep learning, a branch of machine learning, concentrates on training artificial neural networks with multiple layers to acquire knowledge and make predictions based on intricate data. At the heart of deep learning lies artificial neural networks, consisting of interconnected layers of artificial neurons (also referred to as nodes or units). Each neuron conducts a weighted summation of its inputs, applies an activation function, and generates an output.

The mathematical representation of the output of a neuron can be expressed as:

$$z = w_1x_1 + w_2x_2 + \ldots + w_nx_n + b \qquad (4)$$

In this context, x_1, x_2, ..., x_n denote the input values or activations from the preceding layer, w_1, w_2, ..., w_n refer to the respective weights, b represents the bias term, and z denotes the weighted sum of inputs.

To train a deep learning model, a loss or cost function is necessary, which measures the disparity between the predicted output and the true output. The objective is to minimize this difference using an optimization algorithm called backpropagation. Backpropagation calculates the gradient of the loss function concerning the weights and biases in the network, enabling their adjustment in a manner that reduces the error. The gradient descent algorithm is commonly employed for this purpose. The process of updating the weights and biases is governed by the following equations:

$$w_i(new) = w_i(old) - \text{learning rate} * \partial loss/\partial w_i \qquad (5)$$

$$b(new) = b(old) - \text{learning rate} * \partial loss/\partial b \qquad (6)$$

Here, w_i(new) and b(new) represent the updated weights and biases, w_i(old) and b(old) are the current weights and biases, learning rate is a hyper parameter that determines the step size of the update, and $\partial loss/\partial w_i$ and $\partial loss/\partial b$ represent the derivatives of the loss function with respect to the weights and biases.

7. Lasso Regression

Lasso Regression, which stands for Least Absolute Shrinkage and Selection Operator, is a linear regression technique that integrates regularization to enhance model performance and select relevant features. Given a dataset with n observations and p features, let X be an n x p matrix representing the predictor variables, y is an n-dimensional vector representing the response variable, and β be a p-dimensional vector representing the coefficients to be estimated.

The formulation of the Lasso Regression model can be expressed as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon \quad (7)$$

where ε is the error term.

The primary goal of Lasso Regression is to minimize the total of squared residuals while adhering to a constraint on the absolute sum of the coefficients:

$$\text{minimize: } (1/2n) * \Sigma(y_i - (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi}))^2 \quad (8)$$

subject to: $\Sigma|\beta_j| \leq t$,

where i ranges from 1 to n, j ranges from 1 to p, and t is a tuning parameter that controls the level of regularization.

The constraint $\Sigma|\beta_j| \leq t$ encourages sparsity in the model, meaning it promotes the selection of a subset of relevant features by driving some coefficients to zero. The characteristic of Lasso Regression makes it valuable for the purpose of feature selection since it automatically conducts variable selection by reducing the coefficients of irrelevant features towards zero.

8. Findings and Discussions

8.1. ARIMA Model (Result analysis)

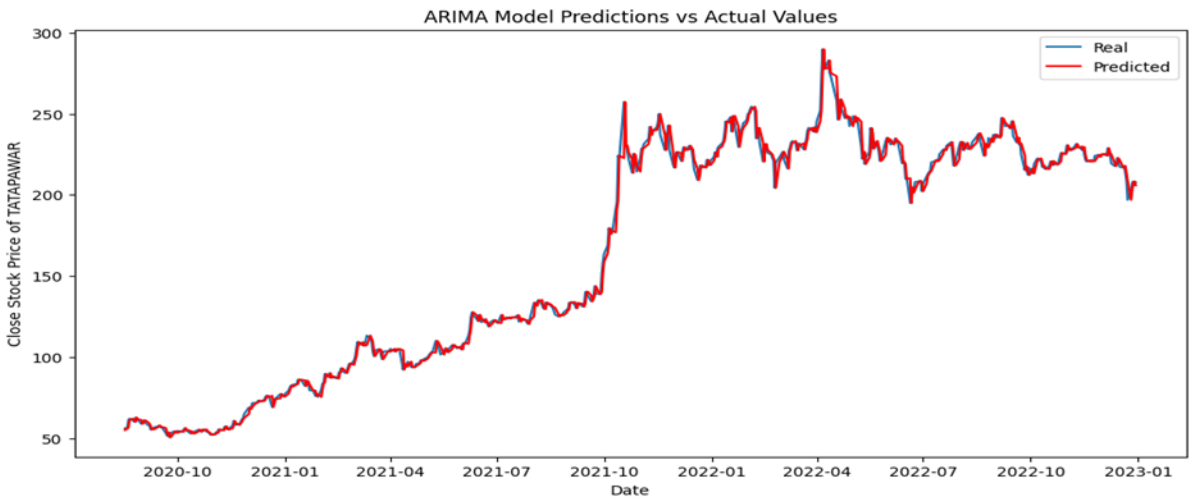

Figure 1: ARIMA model predictions vs. Actual values

In Fig. actual closing price of TATAPOWER mutual fund and predicted closing price of TATAPOWER mutual fund taken into consideration. The fig. shows forecasted and actual closing price of mutual fund results i.e. very closed to each other. So we can say that performance of model is very adequate. The MAE value is 3.020% and RMSE is 4.764% also shows the accuracy of model.

Table 2
ARIMA Model Predicted Result

Predicted	Values	Actual	Values	Difference
Predicted	55.59999847	Actual	55.25000000	-0.34999847
Predicted	55.25000000	Actual	55.95000076	0.70000076
Predicted	55.95000076	Actual	57.65000153	1.70000076
Predicted	57.65000153	Actual	60.59999847	2.94999695
Predicted	60.59999847	Actual	59.15000153	-1.44999695
Predicted	59.15000153	Actual	58.09999847	-1.05000305
Predicted	58.09999847	Actual	58.84999847	0.75000000
Predicted	58.84999847	Actual	59.95000076	1.10000229
Predicted	59.95000076	Actual	61.50000000	1.54999924
Predicted	61.50000000	Actual	62.34999847	0.84999847
Predicted	62.34999847	Actual	64.90000153	2.55000305
Predicted	64.90000153	Actual	68.84999847	3.94999695
Predicted	68.84999847	Actual	67.94999695	-0.90000153
Predicted	67.94999695	Actual	69.25000000	1.30000305
Predicted	69.25000000	Actual	71.65000153	2.40000153
Predicted	71.65000153	Actual	71.65000153	0.00000000
Predicted	71.65000153	Actual	72.00000000	0.34999847
Predicted	72.00000000	Actual	73.15000153	1.15000153
Predicted	73.15000153	Actual	72.65000153	-0.50000000
Predicted	72.65000153	Actual	72.80000305	0.15000153

8.2. Deep Learning Model (Result analysis)

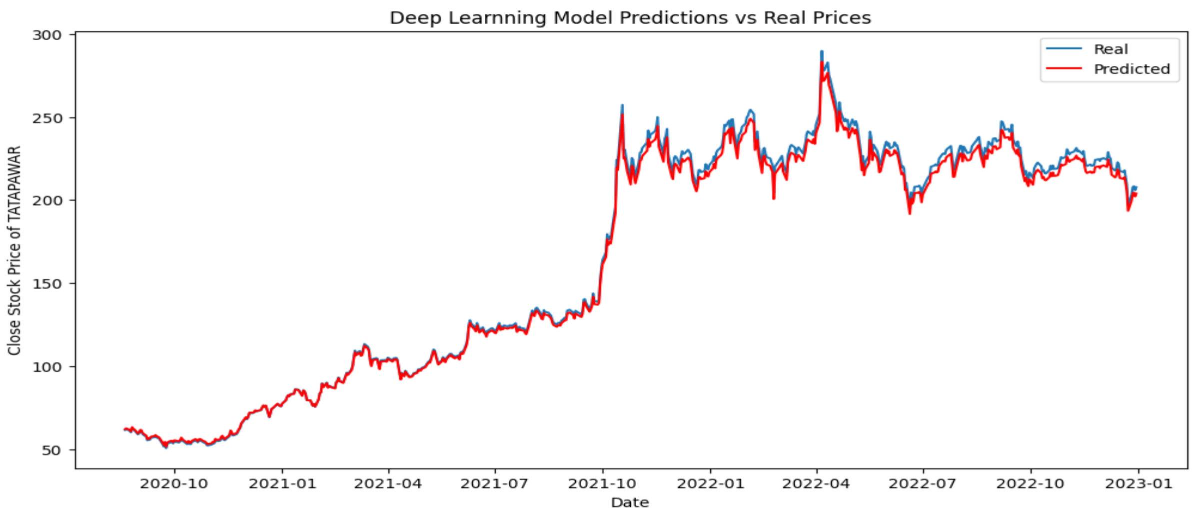

Figure 2: Deep learning model predictions vs. Real prices

Fig- shows the real closing price of TATAPOWER mutual fund and predicted closing price of TATAPOWER mutual fund. The graph shows that proposed model the actual value and predicted value of this mutual fund is very close to each other. Forecasting analysis also proves the accuracy of model with MAE value is 03.3140% and RMSE is 04.7740% these values slightly differ from ARIMA model.

Table 3

LSTM (Deep Learning) Model Predicted Result

Predicted	Values	Actual	Values	Difference
Predicted	84.04821000	Actual	84.09999847	0.05178847
Predicted	79.80599000	Actual	79.84999847	0.04400847
Predicted	76.41883000	Actual	76.44999695	0.03116695
Predicted	86.94885000	Actual	87.00000000	0.05115000
Predicted	101.78369000	Actual	101.76106262	-0.02262738
Predicted	52.88502000	Actual	52.79999924	-0.08502076
Predicted	103.62716000	Actual	103.58976746	-0.03739254
Predicted	96.58548000	Actual	96.59999847	0.01451847
Predicted	82.10055000	Actual	82.15000153	0.04945153
Predicted	84.34804500	Actual	84.40000153	0.05195653
Predicted	91.97608000	Actual	92.01438904	0.03830904
Predicted	208.20403000	Actual	208.55000305	0.34597305
Predicted	46.76868400	Actual	46.70000076	-0.06868324
Predicted	53.48473700	Actual	53.40000153	-0.08473547
Predicted	78.12807500	Actual	78.16638947	0.03831447
Predicted	68.91384000	Actual	68.90222168	-0.01161832
Predicted	72.31850400	Actual	72.32803345	0.00952945
Predicted	212.88274000	Actual	212.89999390	0.01725390
Predicted	80.15499000	Actual	80.19999695	0.04500695
Predicted	77.46432000	Actual	77.50000000	0.03568000

8.3. Lasso Regression Model(Result Analysis)

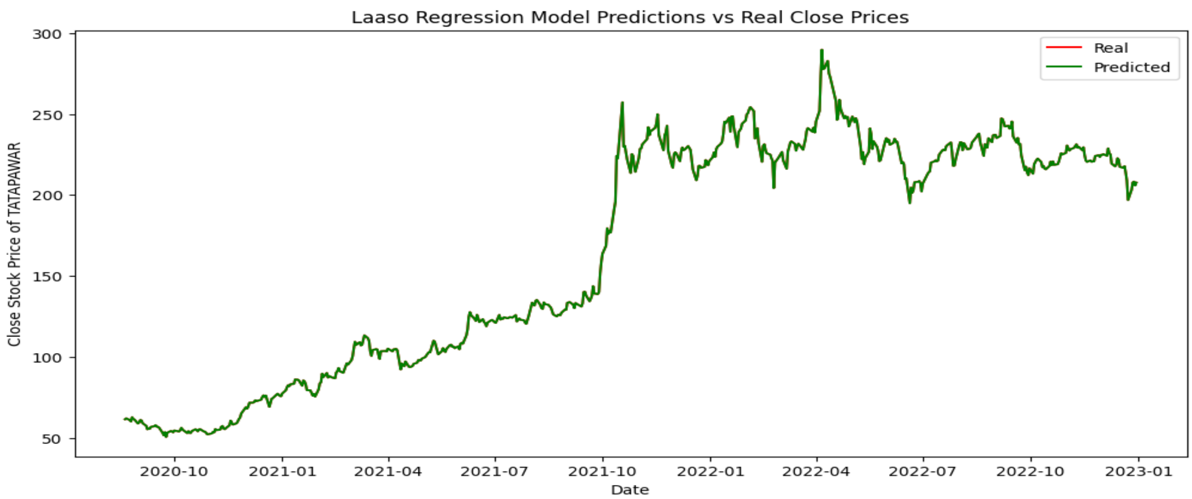

Figure 3: Lasso regression model predictions vs. Real close prices

Fig- shows the real closing price of TATAPOWER mutual fund and predicted closing price of TATAPOWER mutual fund. The graph shows that proposed Lasso Regression model's actual closing price and predicted value of this mutual fund is very close to each other. Forecasting analysis also proves the accuracy of model with MAE value is 0. 0.0274% and RMSE is 0.0333% these values slightly differ from ARIMA model. This model performs more actuate than both above models.

Table 4
Lasso Regression Model Predicted Result.

Predicted	Values	Actual	Values	Difference
Predicted	55.32085481	Actual	55.25000000	-0.07085481
Predicted	56.03800361	Actual	55.95000076	-0.08800285
Predicted	57.62731604	Actual	57.65000153	0.02268548
Predicted	60.47315409	Actual	60.59999847	0.12684439
Predicted	59.26959854	Actual	59.15000153	-0.11959702
Predicted	58.13499009	Actual	58.09999847	-0.03499162
Predicted	58.88828272	Actual	58.84999847	-0.03828425
Predicted	59.92623244	Actual	59.95000076	0.02376832
Predicted	61.52778718	Actual	61.50000000	-0.02778718
Predicted	62.35728638	Actual	62.34999847	-0.00728791
Predicted	64.79042694	Actual	64.90000153	0.10957459
Predicted	68.62412994	Actual	68.84999847	0.22586853
Predicted	67.96343689	Actual	67.94999695	-0.01343994
Predicted	69.20334094	Actual	69.25000000	0.04665906
Predicted	71.55890072	Actual	71.65000153	0.09110080
Predicted	71.71626831	Actual	71.65000153	-0.06626679
Predicted	71.95281789	Actual	72.00000000	0.04718211
Predicted	73.15079437	Actual	73.15000153	-0.00079285
Predicted	72.60795916	Actual	72.65000153	0.04204236
Predicted	72.84362851	Actual	72.80000305	-0.04362545

9. Model Evaluation Criteria

9.1. Mean Squared Error (MSE)

MSE is another way to calculate the accuracy and error of the forecast model used:

$$\text{MSE} = \frac{1}{n}\sum_{n=1}^{n}(\text{Yi} - \hat{\text{Y}}\text{i})2 \qquad (9)$$

Ŷi is the predicted ith value and Yi is the actual/ observed value.

9.2. Root-mean-square deviation (RMSE)

RMSE is another way to calculate the accuracy of proposed model but it considers the error calculation based on standard deviation. The final output is one standard deviation of the magnitude of the error, and the individual calculations are reported as residuals:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{n=1}^{n}}(\text{Yi} - \hat{\text{Y}}\text{i})2 \qquad (10)$$

Ŷ i is the predicted ith value and Yi is the actual / observed value.

9.3. Mean absolute percentage error (MAPE)

MAPE may be a formula for calculating the precision of estimates. The calculation is done by taking the contrast between the real value and the anticipated esteem and separating the distinction by the actual value.

$$\text{MAPE} = \frac{100}{n} \sum_{n=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \quad (11)$$

F_t is the predicted value and At is the actual / observed value

Table 5
Performance of ARIMA, Deep Learning and LASSO Regression

	ARIMA	DEEP LEARNNING	LAASO REGRESSION
Mean Squared Error (MSE)	22.696882585088	22.791609209855	0.001109710175
Root Mean Squared Error (RMSE)	4.764124535010	4.774055844861	0.033312312668
Mean Absolute Error (MAE)	3.020861425915	3.314005833935	0.027450366388
R2 Score	0.995494145023	0.995444658959	0.999999778203
Explained Variance Score	0.995507361935	0.996379227349	0.999999908015
Mean Absolute Percentage Error (MAPE)	68.207455468409	68.739239429167	0.014450309727
Mean Percentage Error (MPE)	-30.388460164209	-31.916412239513	0.011001326314

i. All three methods (ARIMA, Deep Learning, and LASSO Regression) seem to perform well, as indicated by high R2 scores and Explained Variance Scores. They explain a significant portion of the variance in the data.
ii. The LASSO Regression method has extremely low MSE, RMSE, and MAE values, indicating very accurate predictions.
iii. ARIMA and Deep Learning have similar performance metrics, with ARIMA having a slightly lower RMSE and MAE.
iv. The Mean Absolute Percentage Error (MAPE) for ARIMA and Deep Learning is relatively high, suggesting that the percentage errors can be significant.

In contrast, LASSO Regression has an exceptionally low MAPE and MPE, indicating very accurate percentage error estimates.

Conclusion

In conclusion, this study aimed to perform a Comparative Analysis of ARIMA, Deep Learning, and Lasso Regression Models for Time Series Forecasting on an Indian mutual fund dataset. Through a comprehensive evaluation and comparison of these models, several significant findings have emerged. Firstly, the ARIMA model exhibited robust performance in capturing the temporal patterns and trends in the mutual fund data. Secondly, the deep learning models, particularly the long short-term memory (LSTM) networks, demonstrated comparable predictive capabilities to ARIMA. Lastly, the Lasso regression approach, which leverages regularization techniques, offered a unique perspective by incorporating variable selection and regularization into the forecasting process. It proved to be effective in handling multicollinearity and identifying significant predictors for mutual fund performance.Table-5 shows the accuracy results of different models Lasso Regression Model outperforms over Deep Learning and ARIMA model. Sustainability in investment refers to the practice of considering environmental, social, and governance (ESG) factors when making investment decisions. It goes beyond traditional financial analysis by evaluating how a company's operations and practices impact the planet, society, and its long-term performance. The goal of sustainable investing is to generate positive financial returns while also promoting positive outcomes for the environment and society. It is crucial to acknowledge that the choice of an appropriate forecasting model should consider multiple factors, such as the specific objectives, characteristics of the data, and the desired balance between accuracy and interpretability. Researchers and practitioners can leverage the insights gained from this study to make informed decisions when selecting a time-series forecasting model for Indian mutual fund performance analysis. Additionally, further research could explore ensemble techniques that combine the strengths of different models to enhance forecasting accuracy and robustness.

References

[1] Zeroual, A., Harrou, F., Dairi, A., & Sun, Y. (2020). Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study. Chaos, Solitons & Fractals, 140, 110121.

[2] Benevento, E., Aloini, D., & Squicciarini, N. (2023). Towards a real-time prediction of waiting times in emergency departments: A comparative analysis of machine learning techniques. International Journal of Forecasting, 39(1), 192-208.

[3] Zhang, L., Li, C., Chen, L., Chen, D., Xiang, Z., & Pan, B. (2023). A Hybrid Forecasting Method for Anticipating Stock Market Trends via a Soft-Thresholding De-noise Model and Support Vector Machine (SVM). World Basic and Applied Sciences Journal, 13(2023), 597-602.

[4] Guo, K., Jiang, Z., & Zhang, Y. (2023). Prediction of S&P500 Stock Index Using ARIM and Linear Regression. Highlights in Science, Engineering and Technology, 38, 399-407.

[5] Pandey, A., Singh, G., Hadiyuono, H., Mourya, K., & Rasool, M. J. (2023, January). Using ARIMA and LSTM to Implement Stock Market Analysis. In 2023 International Conference on Artificial Intelligence and Smart Communication (AISC) (pp. 935-940). IEEE.

[6] Xu, Y., Wang, J., Chen, Z., & Liang, C. (2023). Sentiment indices and stock returns: Evidence from China. International Journal of Finance & Economics, 28(1), 1063-1080.

[7] Dai, Z., Zhang, X., & Li, T. (2023). Forecasting stock return volatility in data-rich environment: A new powerful predictor. The North American Journal of Economics and Finance, 64, 101845.

[8] Ma, F., Lu, X., Liu, J., & Huang, D. (2022). Macroeconomic attention and stock market return predictability. Journal of International Financial Markets, Institutions and Money, 79, 101603.

[9] Li, X., Liang, C., & Ma, F. (2022). Forecasting stock market volatility with a large number of predictors: New evidence from the MS-MIDAS-LASSO model. Annals of Operations Research, 1-40.

[10] Ren, X., Xu, W., & Duan, K. (2022). Fourier transform based LSTM stock prediction model under oil shocks. Quantitative Finance and Economics, 6(2), 342-358.

[11] Zhu, Z., & He, K. (2022). Prediction of Amazon's Stock Price Based on ARIMA, XGBoost, and LSTM Models. Proceedings of Business and Economic Studies, 5(5), 127-136.

[12] Lee, H. Y., Beh, W. L., & Lem, K. H. (2023). Forecasting with information extracted from the residuals of ARIMA in financial time series using continuous wavelet transform. International Journal of Business Intelligence and Data Mining, 22(1-2), 70-99.

[13] Liu, T., Ma, X., Li, S., Li, X., & Zhang, C. (2022). A stock price prediction method based on meta-learning and variational mode decomposition. Knowledge-Based Systems, 252, 109324.

[14] Nair, A. V., & Narayanan, J. (2022, August). Indian Stock Market Forecasting using Prophet Model. In 2022 International Conference on Connected Systems & Intelligence (CSI) (pp. 1-7). IEEE.

[15] Zeng, L., Ling, L., Zhang, D., & Jiang, W. (2023). Optimal forecast combination based on PSO-CS approach for daily agricultural future prices forecasting. Applied Soft Computing, 132, 109833.

[16] Wu, D., Ma, X., & Olson, D. L. (2022). Financial distress prediction using integrated Z-score and multilayer perceptron neural networks. Decision Support Systems, 159, 113814.
[17] Isabona, J., Imoize, A. L., Ojo, S., Karunwi, O., Kim, Y., Lee, C. C., & Li, C. T. (2022). Development of a multilayer perceptron neural network for optimal predictive modeling in urban microcellular radio environments. Applied Sciences, 12(11), 5713.
[18] Li, G., Zhang, A., Zhang, Q., Wu, D., & Zhan, C. (2022). Pearson correlation coefficient-based performance enhancement of Broad Learning System for stock price prediction. IEEE Transactions on Circuits and Systems II: Express Briefs, 69(5), 2413-2417.
[19] Banerjee, S., & Mukherjee, D. (2022). Short Term Stock Price Prediction in Indian Market: A Neural Network Perspective. Studies in Microeconomics, 10(1), 23-49.
[20] Ji, X., Wang, J., & Yan, Z. (2021). A stock price prediction method based on deep learning technology. International Journal of Crowd Science, 5(1), 55-72.
[21] Kumar, D., Sarangi, P. K., & Verma, R. (2022). A systematic review of stock market prediction using machine learning and statistical techniques. Materials Today: Proceedings, 49, 3187-3191.
[22] Tanwar, R., Phukan, O. C., Singh, G., & Tiwari, S. (2022). CNN-LSTM Based Stress Recognition Using Wearables.
[23] Tanwar, R., Phukan, O. C., Singh, G., Pal, P. K., & Tiwari, S. (2024). Attention based hybrid deep learning model for wearable based stress recognition. Engineering Applications of Artificial Intelligence, 127, 107391.
[24] Makwana, Y., Iyer, S. S., & Tiwari, S. (2022). The food recognition and nutrition assessment from images using artificial intelligence: a survey. ECS Transactions, 107(1), 3547.