ST ALOYSIUS
(DEEMED TO BE UNIVERSITY)
MANGALURU 575003 - INDIA
Institute of Management and IT - AIMIT
ESTD : 1880

International Conference On
Advanced IT, Engineering and Management

EXPLORING STOCK MARKET PREDICTIONS WITH ML : A DEEP DIVE INTO INDIA'S LARGEST COMPANIES

Alphy Joby, Alister Wilbur Dsouza, Mrs Vanitha T

Department of IT, AIMIT, St. Aloysius (Deemed to be University),
Mangalore, Karnataka, India

OVERVIEW

- INTRODUCTION
- RESEARCH SIGNIFICANCE
- DATASET OVERVIEW
- ALGORITHMS
- METHODOLOGY
- RESULTS
- KEY FINDINGS
- FUTURE DECISIONS
- CONCLUSION
- REFERENCES

INTRODUCTION

INTRODUCTION

Stock Market Volatility:

- The stock market is inherently volatile and unpredictable, influenced by global events and economic changes.

Need for Prediction:

- Accurate stock market forecasting helps minimize losses and maximize profits for investors.

Objective of the Study:

- Evaluate five algorithms—K-Nearest Neighbors (KNN), Linear Regression, Support Vector Regression (SVR), Decision Tree Regression (DTR), and Long Short-Term Memory (LSTM)—for predicting stock prices.

Scope:

- Focus on stock price data from 10 leading Indian companies (2017–2024).

Evaluation Metrics:

- Performance analyzed using SMAPE, R^2, and RMSE to assess the effectiveness of ML and DL algorithms.

RESEARCH SIGNIFICANCE

RESEARCH SIGNIFICANCE

Challenges in Prediction:

- Stock prices are affected by complex and unpredictable factors such as global events, economic shifts, and investor behavior.

Advancements in Technology:

- Machine learning (ML) and deep learning (DL) offer new possibilities to improve accuracy in financial forecasting.

Importance of Model Selection:

- The study underscores how selecting the right algorithm is crucial for addressing dataset-specific characteristics and prediction goals.

Practical Applications:

- Insights from this research can aid investors, financial analysts, and policymakers in making informed decisions.

Contribution to Literature:

- Highlights the strengths and limitations of ML and DL models in stock market prediction, providing a foundation for future hybrid approaches.

DATASET OVERVIEW

DATASET OVERVIEW

Data Source:

- Historical stock data collected using the yfinance Python package from Yahoo Finance.

Timeframe:

- Dataset spans from 2017 to October 2024, capturing diverse market conditions.

Top 10 Indian Companies:

- Focus on the largest firms by market capitalization to ensure robust insights into market trends.

Key Financial Indicators:

- Includes trade volumes, daily closing prices, and adjusted closing prices (factoring in splits, dividends, etc.).

Market Conditions Captured:

- Reflects the effects of events like the COVID-19 pandemic, economic recovery periods, and other macroeconomic factors.

ALGORITHMS

ALGORITHMS

K-NEAREST NEIGHBOUR (KNN)

Predicts based on the average of the closest data points using distance measures like Euclidean or Manhattan.
Strength: Works well for small to medium datasets.
Limitation: Not scalable for large or high-dimensional data.

DECISION TREE REGRESSION

Splits data into nodes based on feature values.
Strength: Works with any dataset size; no need for feature scaling.
Limitation: Can overfit when there are no limits on depth.

LONG SHORT-TERM MEMORY (LSTM)

A neural network that captures temporal patterns using memory cells with gates.
Strength: Great for time-series and sequential data.
Limitation: Needs lots of computation and well-prepared data.

LINEAR REGRESSION

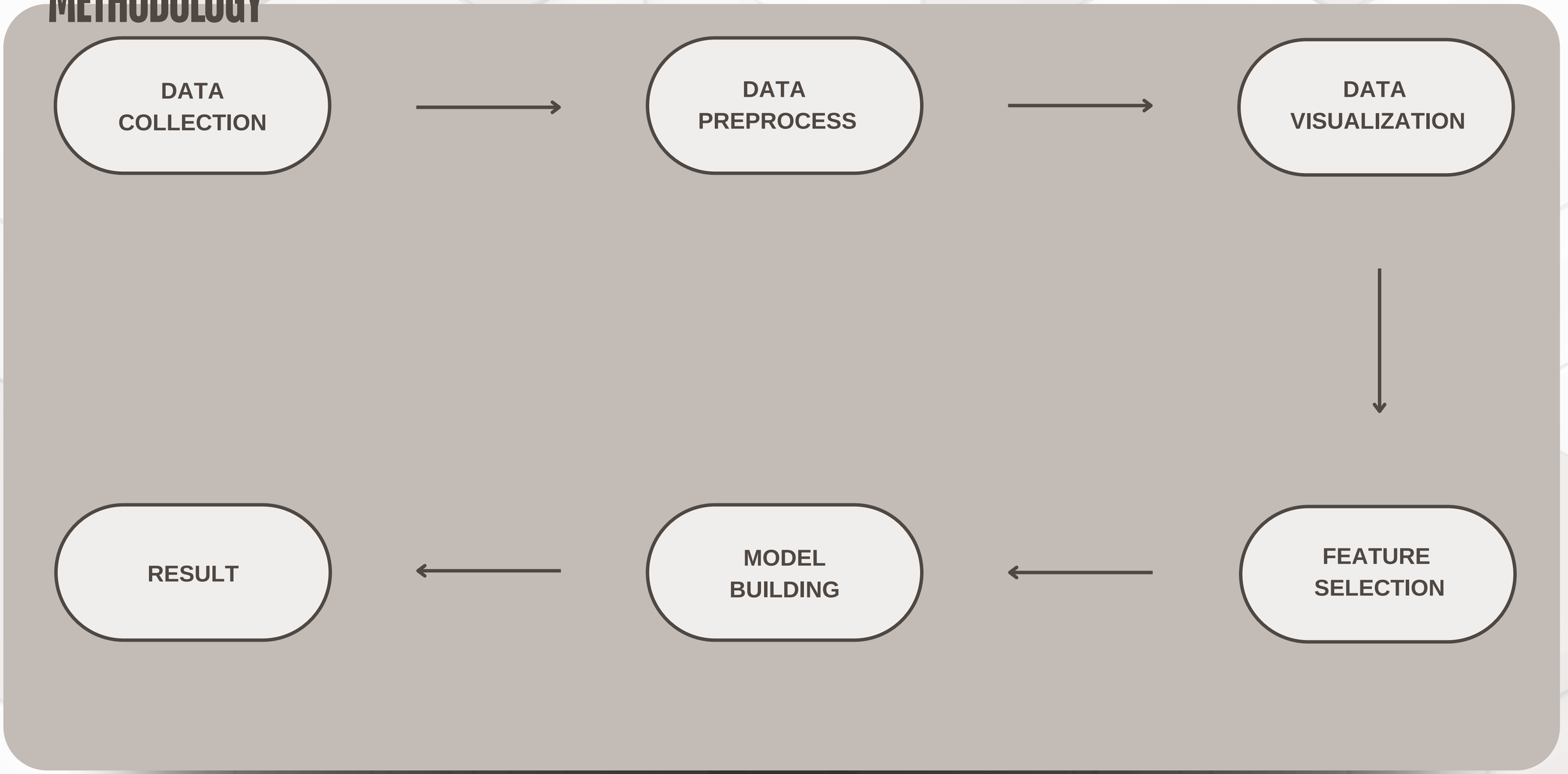
Models the linear relationship between independent variables and the target variable.
Strength: Simple, efficient, and interpretable for linear data.
Limitation: Fails to capture non-linear patterns and interactions.

SUPPORT VECTOR REGRESSION (SVR)

Applies kernel functions to map data to a higher dimensional space, allowing a hyperplane to be fitted for separation.
Strength: Good for non-linear data and outliers.
Limitation: Requires careful tuning; slow for large datasets.

METHODOLOGY

METHODOLOGY



RESULTS

RESULTS

SMAPE (Symmetric Mean Absolute Percentage Error):

- **Best Performer:** Linear Regression with the lowest average SMAPE (3.09%).
- **SVR** should be avoided due to its high error variance and poor performance for several stocks.
- For models like **Decision Tree and LSTM**, while they perform decently, their effectiveness varies based on the stock.

R^2 (Coefficient of Determination):

- **Highest Accuracy:** Linear Regression with the best average R^2 (0.9759).
- **LSTM:** Consistently good but less robust in certain cases.
- **Negative R^2:** Notable for SVR and Decision Tree on specific stocks (e.g., ICICIBANK, LT).

RMSE (Root Mean Square Error):

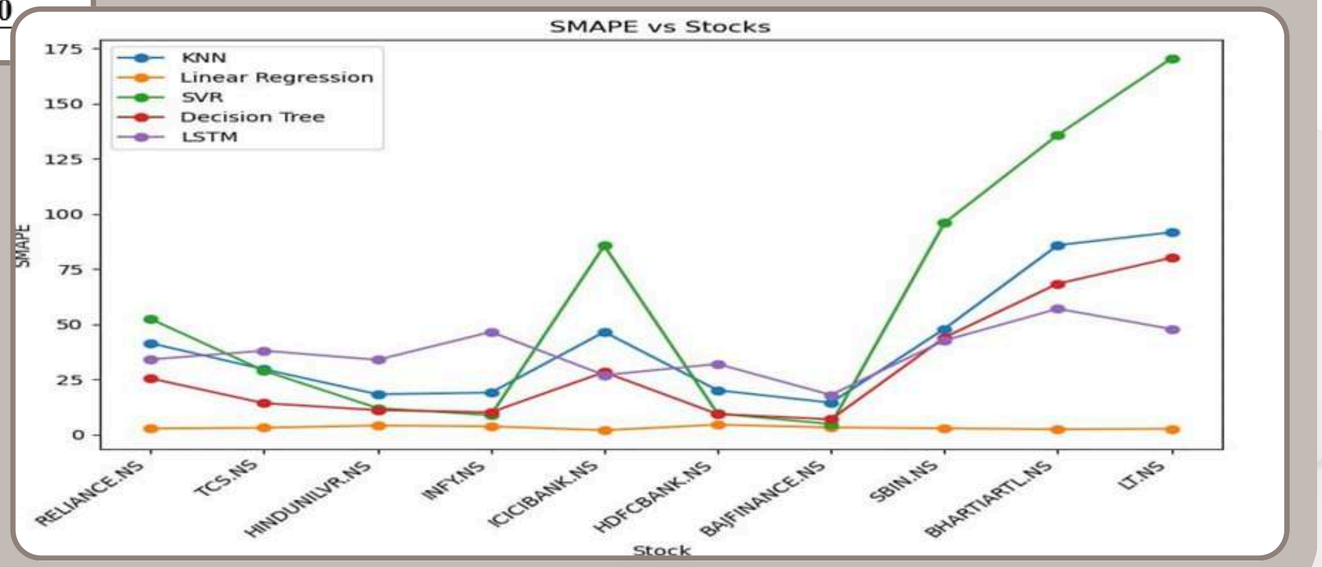
- **Lowest Error:** Linear Regression achieved the best average RMSE.
- **LSTM:** Also performed strongly but marginally less effective than LSTM.
- **High Errors:** SVR and Decision Tree Regression had significantly higher RMSE in volatile stocks (e.g., BHARTIARTL, LT).

RESULTS

SMAPE

Parameter	Symmetric Mean Absolute Percentage Error (SMAPE)				
Algorithm	**KNN**	**Linear Regression**	**SVR**	**DTR**	**LSTM**
RELIANCE	41.3695	2.7071	52.3628	25.4178	34.0639
TCS	29.5534	3.0585	28.8791	14.1810	37.9727
HINDUNILVR	18.2559	4.0809	11.9810	11.0723	33.8652
INFY	18.9797	3.6932	8.7636	10.0778	46.4977
ICICIBANK	46.5055	1.9456	85.6289	28.2761	26.9509
HDFCBANK	20.0494	4.4390	9.5102	9.2743	31.9707
BAJFINANCE	14.4219	3.2114	4.7401	6.8845	17.9349
SBIN	47.8709	2.7910	95.8431	43.9610	42.6176
BHARTIARTL	85.8475	2.3957	135.8424	68.3537	57.0099
LT.	91.6570	2.5856	170.4141	80.1099	47.8466
Average	**41.4510**	**3.0908**	**60.3965**	**29.7608**	**37.6730**

Table 1: SMAPE Scores for Stock Prediction Models

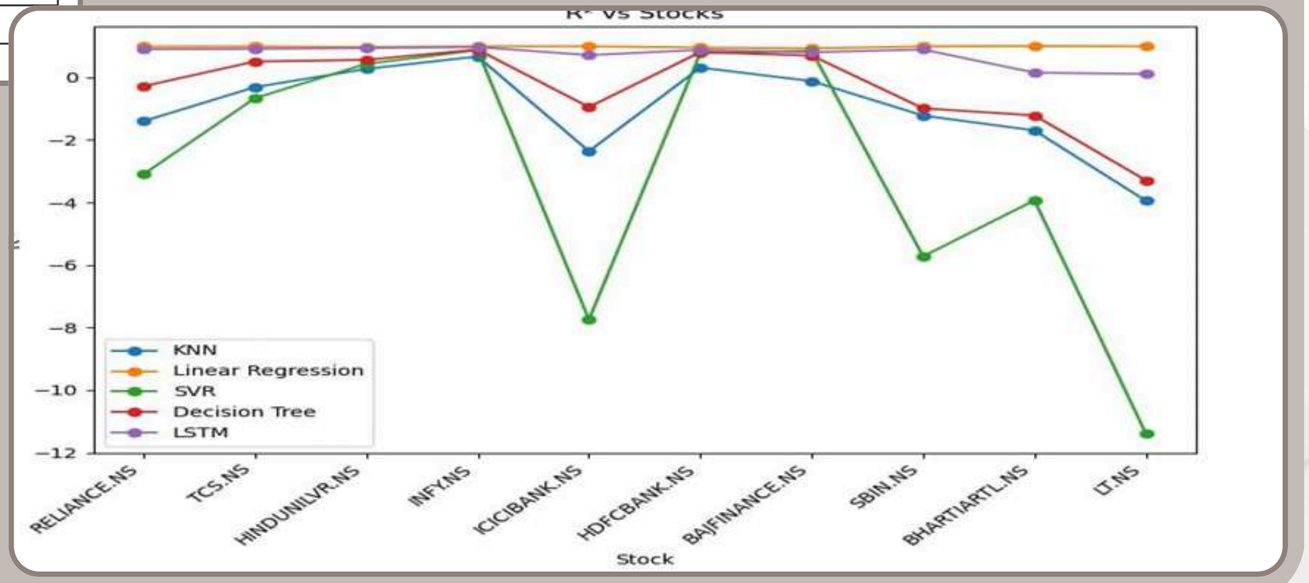


RESULTS

R2

Parameter	R Squared (R^2)				
Algorithm	**KNN**	**Linear Regression**	**SVR**	**DTR**	**LSTM**
RELIANCE	-1.4042	0.9834	-3.1007	-0.2860	0.8984
TCS	-0.3160	0.9853	-0.6694	0.5042	0.9091
HINDUNILVR	0.2669	0.9683	0.4279	0.5573	0.9347
INFY	0.6646	0.9879	0.8682	0.8852	0.9669
ICICIBANK	-2.3640	0.9889	-7.7445	-0.9518	0.7092
HDFCBANK	0.3092	0.9520	0.7726	0.8038	0.8758
BAJFINANCE	-0.1220	0.9212	0.8436	0.6943	0.7994
SBIN	-1.2247	0.9879	-5.7241	-0.9925	0.8884
BHARTIARTL	-1.6987	0.9965	-3.9343	-1.2185	0.1483
LT	-3.9465	0.9877	-11.3863	-3.2930	0.1056
Average	**-0.9835**	**0.9759**	**-2.9647**	**-0.3297**	**0.7235**

Table 2: R^2 Scores for Stock Prediction Models

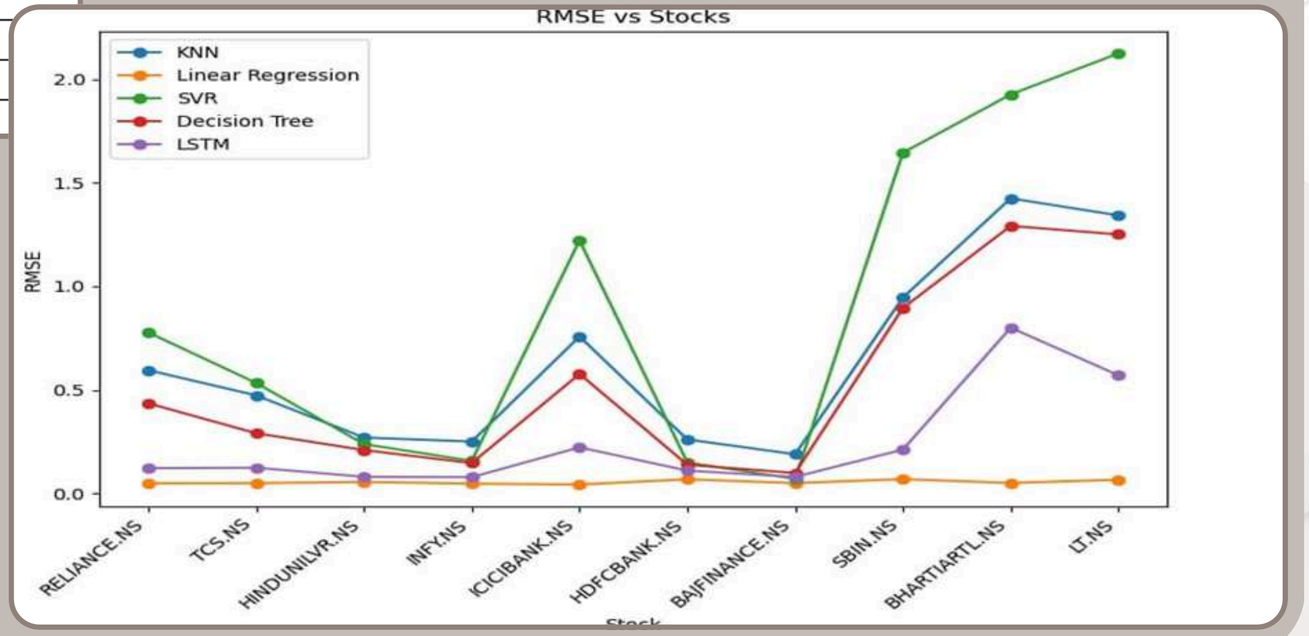


RESULTS

RMSE

Parameter	RMSE (Root Mean Square Error)				
Algorithm	**KNN**	**Linear Regression**	**SVR**	**DTR**	**LSTM**
RELIANCE	0.5951	0.0494	0.7772	0.4352	0.1222
TCS	0.4737	0.0500	0.5336	0.2907	0.1244
HINDUNILVR	0.2700	0.0561	0.2385	0.2098	0.0805
INFY	0.2507	0.0476	0.1572	0.1467	0.0787
ICICIBANK	0.7578	0.0433	1.2118	0.5772	0.2228
HDFCBANK	0.2607	0.0686	0.1495	0.1389	0.1105
BAJFINANCE	0.1894	0.0501	0.0707	0.0988	0.0801
SBIN	0.9459	0.0696	1.6444	0.8951	0.2117
BHARTIARTL	1.4244	0.0506	1.9261	1.2915	0.8002
LT	1.3423	0.0668	2.1241	1.2505	0.5707
Average	**0.6510**	**0.0552**	**0.8833**	**0.5334**	**0.2401**

Table 3: RMSE Scores for Stock Prediction Models



KEY FINDINGS

KEY FINDINGS

Top-Performing Models:

- LSTM: Demonstrated strong accuracy in handling complex and volatile stock trends.
- Linear Regression: Reliable and efficient for linear and less volatile datasets.

Algorithm Limitations:

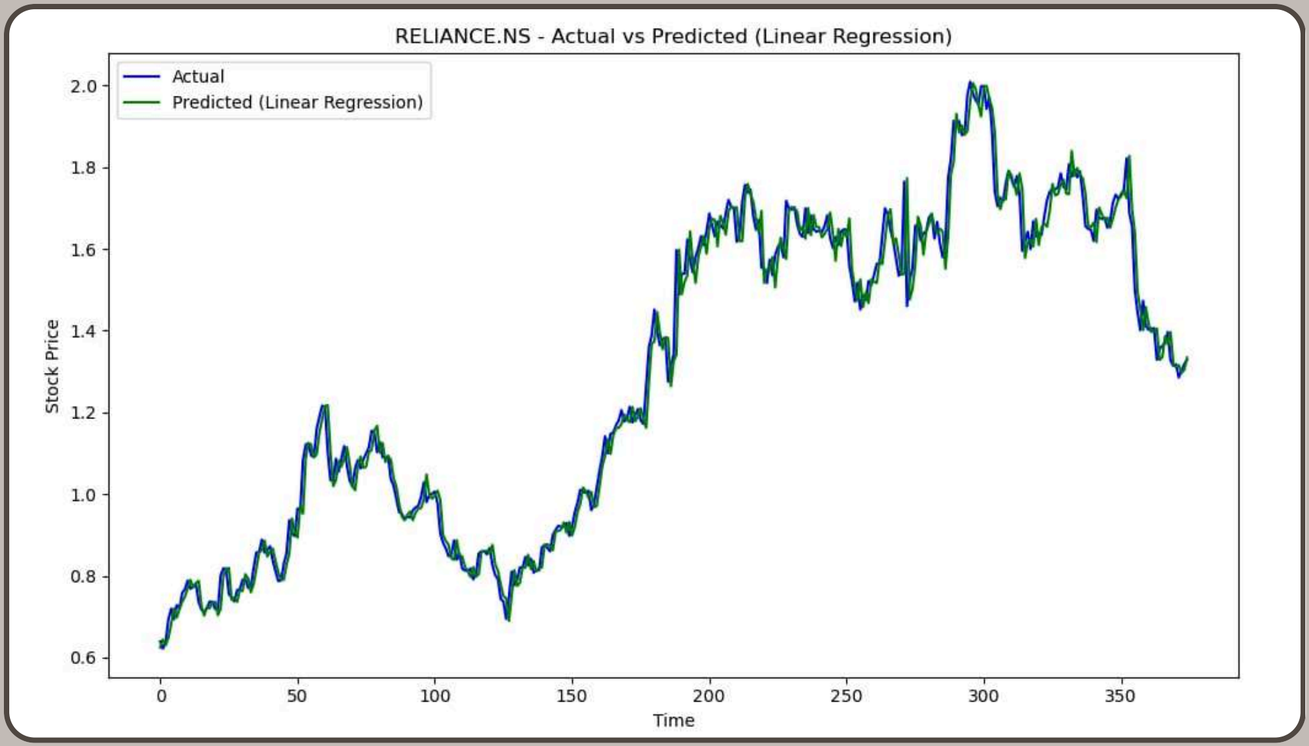
- SVR and Decision Tree Regression: Struggled with volatile stocks, showing inconsistent and poor performance.
- KNN: Moderate accuracy but faced scalability issues with high-dimensional data.

Metric Summary:

- Linear Regression had the lowest SMAPE and highest R2.
- LSTM and Linear Regression achieved the lowest RMSE, making it highly precise in dynamic markets.

KEY FINDINGS

ACTUAL VS PREDICTED PRICE



Model: Linear Regression effectively predicts stock prices with close alignment between actual and predicted values.

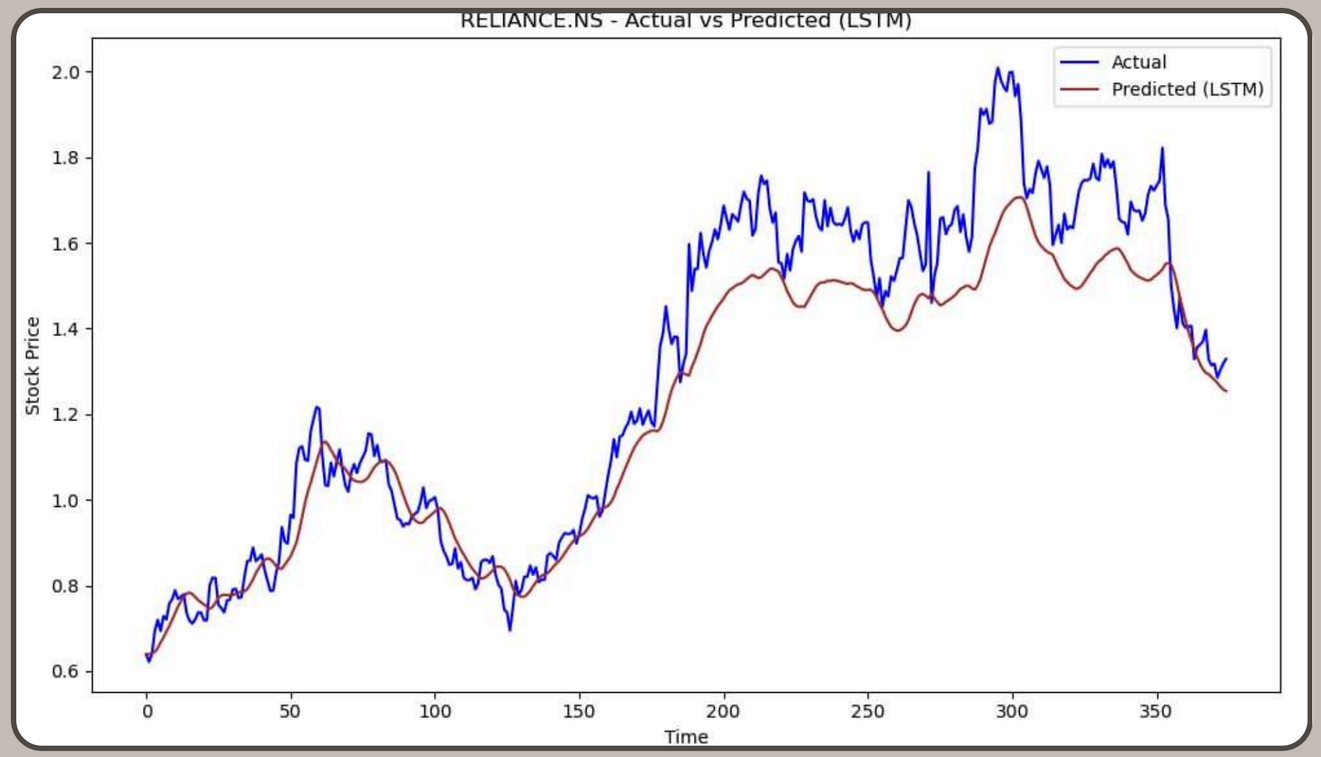
Accuracy: Minimal deviations indicate strong performance for this dataset.

Trend Capture: Captures overall trends and fluctuations in stock prices accurately.

Applicability: Well-suited for stable and linear market conditions.

KEY FINDINGS

ACTUAL VS PREDICTED PRICE



Model: LSTM effectively predicts stock prices, showcasing strong alignment with actual values.

Accuracy: Captures overall trends well, though slight deviations appear during rapid market fluctuations.

Trend Capture: Smooth prediction curve highlights its ability to handle dynamic and volatile conditions.

Applicability: Well-suited for non-linear and complex market patterns.

FUTURE DIRECTIONS

FUTURE DIRECTIONS

Hybrid Models:

- Combine ML and DL techniques to leverage their strengths (e.g., Linear Regression for trends, LSTM for sequential data).

Integration of Additional Data:

- Incorporate sentiment analysis from news, social media, and economic indicators to enhance predictive accuracy.

Real-Time Predictions:

- Implement models for real-time forecasting to assist in day-to-day trading decisions.

CONCLUSION

CONCLUSION

- **Linear Regression Reliability:** Strong performance in linear scenarios with highest R2 and lowest SMAPE.
- **LSTM Superiority**: Best accuracy for volatile, dynamic markets due to its ability to capture sequential and temporal patterns.
- **SVR and Decision Tree Limitations:** Inconsistent results and poor handling of extreme volatility.
- **KNN Challenges:** Scaling issues with large, high-dimensional datasets.
- **Key Insight:** Model selection should align with dataset characteristics and prediction goals.
- **Future Research:** Focus on hybrid models and incorporating sentiment data and economic indicators.
- **Balance ML and DL:** Linear Regression works for simple cases; LSTM excels in complex financial data.

REFERENCES

REFERENCES

- **Bansal, M., Goyal, A., & Choudhary, A. (2022).** Stock market prediction with high accuracy using machine learning techniques. Procedia Computer Science, 215, 247–265. [doi: 10.1016/j.procs.2022.12.028]
- **Adebayo, S. A., Olayemi, Y. A., & Adeyemi, O. O. (2021).** Stock market prediction using machine learning algorithms: A comparative study of KNN, SVM, and Decision Tree models. International Journal of Financial Studies, 9(3), 45-59.
- **Choudhury, S., & Garg, A. (2020).** Stock price prediction using deep learning techniques: A review and future prospects. International Journal of Data Science and Analytics, 10(2), 123-136. [d2oi: 10.1007/s41060-020-00209-6]
- **Singh, A., & Patel, A. (2021). Predicting stock market trends with machine learning:** A study on support vector machines and linear regression. Procedia Computer Science, 185, 214-219. [doi: 10.1016/j.procs.2021.01.031]
- **Zhao, Y., & Wang, J. (2019).** Comparison of deep learning and traditional machine learning algorithms for stock price prediction. Journal of Finance and Data Science, 5(1), 17-31. [doi: 10.1016/j.jfds.2018.10.003]

THANK YOU