Determination of Amazon Stock Price Using Novel LASSO Algorithm Comparing with Accuracy of Linear Regression Algorithm

P. H. Bhargav* and F. Mary Harin Fernandez†
Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, India

Keywords: Amazon Stock Price, Novel LASSO, Linear Regression, Machine Learning, Price Prediction, Stock Market.

Abstract: This research sought to compare the efficacy of the Novel LASSO algorithm (Group 1) and the Linear Regression algorithm (Group 2) in predicting Amazon's stock price, aiming to identify which technique offers superior accuracy. The study was bifurcated into two groups, each comprising 10 samples, wherein each group applied either the Novel LASSO technique or the Linear Regression algorithm to an extensive dataset of Amazon's all-time stock prices. Sample sizes were calculated using ClinCalc software, setting α at 0.05 and a pretest power at 0.8. The results demonstrated that the Novel LASSO technique achieved a higher mean accuracy of 85.31% compared to Linear Regression's 77.44%, a difference found to be statistically significant (p=0.046). In summary, the Novel LASSO method outperformed the Linear Regression algorithm in predicting stock prices.

1 INTRODUCTION

Stock price prediction involves using historical stock price data to construct a model capable of forecasting future prices (Ji, Wang, and Yan 2021). It requires the analysis of various financial and economic indicators, such as earnings per share, the price-to-earnings ratio, and market trends, to discern patterns that might indicate future stock prices. Such predictions are valuable for multiple reasons (Schöneburg 1990). For investors, predicting future stock prices accurately can guide investment decisions. Companies also value these predictions as indicators of their financial health and performance. A myriad of factors, ranging from financial performance and market conditions to investor sentiment, can sway a company's stock price (Hu, Zhao, and Khushi 2021). Numerous studies have explored stock price prediction at the finance level, with some harnessing ML and artificial intelligence for forecasting, whilst others lean on traditional statistical finance methods. Both investors, to guide their investments, and companies, to gauge factors affecting their stock value, use stock price predictions (Obthong et al. 2020).

Research on predicting Amazon's stock price currently sees an average of 50 articles published annually on IEEE Xplore and 800 on Science Direct (Bayu Distiawan Trisedya.2015). One study utilised neural networks' learning functions to process financial data from the internet and then conduct relevant stock data research (Yu and Yan 2020) (Palanivelu et al. 2022). Another employed the machine learning Linear Regression approach for Amazon stock price prediction (J. A. Cook. 2018) (Ramkumar G. et al. 2021). The significance of machine learning in forecasting stock prices was underscored by another study (Obthong et al. 2020). A holistic method, inclusive of stock market dataset pre-processing, diverse feature engineering techniques, and a specialised deep learning algorithm for stock market trend prediction, was proposed in another study (Cakra, Yahya Eru 2015) (Vickram AS. et al. 2021). The author's comprehensive evaluations revealed their suggested solution's superiority, attributed to their meticulous feature engineering (Soni 2020).

Predicting stock closing prices presents challenges and limitations (Matloob Khushi 2021). Its objective is to leverage historical data and other variables to anticipate a stock's future closing value, which serves various purposes, from investment decisions to strategy formulation. Given the stock

* Research Scholar
† Research Guide, Corresponding Author

Bhargav, P. and Fernandez, F.
Determination of Amazon Stock Price Using Novel LASSO Algorithm Comparing with Accuracy of Linear Regression Algorithm.
DOI: 10.5220/0012518000003739

In *Proceedings of the 1st International Conference on Artificial Intelligence for Internet of Things: Accelerating Innovation in Industry and Consumer Electronics (AI4IoT 2023)*, pages 467-473
ISBN: 978-989-758-661-3

market's complexity and dynamism, influenced by numerous factors, predicting future prices accurately is arduous. Its inherent volatility further complicates short-term predictions. This research aims to predict Amazon's stock price using Novel LASSO and Linear Regression techniques, comparing the data to ascertain the most accurate method.

2 MATERIALS AND METHOD

The experiment was conducted at the Artificial Intelligence Laboratory within the Saveetha School of Engineering, Department of CSE, SIMATS. The dataset titled "Amazon Stock Price (All Time) Updated Microsoft Stock Price" was sourced from [Kaggle](https://www.kaggle.com/datasets/kannan131/amazon-stock-price-all-time/code) and stored as a `.csv` file comprising 10,486 data tuples, with the intent of using the Novel LASSO technique for value predictions.

In the proposed methodology, the data was both trained and tested employing a Jupyter notebook. The SPSS software was utilised to visualise forecasting graphs, while G Power was employed to compute the pretest for the algorithms, determining their likely performance (Cleophas & Zwinderman, 2011).

The algorithm ran on a machine equipped with a 50 GB hard drive and 8 GB RAM, operating on a 64-bit Windows OS.

2.1 Novel LASSO

The Novel LASSO is an advancement of the regularisation technique applied in Linear Regression (LR). This method acts to curtail the model's complexity by penalising it for excessive feature usage (Ranstam and Cook 2018). Within the Novel LASSO, the LR model's objective function is augmented to incorporate a penalty term, which is directly proportional to the absolute values of the model coefficients. This can cause certain coefficients to be nullified, effectively excluding the associated features from the model.

This yields a streamlined and more comprehensible model with fewer features. For stock price prediction, LASSO can be integrated into a linear regression framework (Bhattacharjee and Bhattacharja 2019). In such an application, it's imperative to collate and preprocess data prior to model training. This data encompasses financial metrics like stock prices and volumes, supplemented by pertinent indicators like economic trends and corporate news.

Using the Novel LASSO, the most salient features from this data pool can be identified, forming the foundation for the linear regression model. Once trained, the model is then poised to forecast future stock prices.

2.2 Linear Regression

Linear regression statistically models the relationship between a dependent variable and one or more independent variables (Montgomery, Peck, and Geoffrey Vining 2021). This method is termed 'linear' based on the underlying assumption that the variables share a linear correlation. Essentially, this means a change in the dependent variable is associated with a proportional change in the independent variables. Linear regression serves several purposes, from forecasting future values, delineating relationships between variables, to pinpointing key variables influencing specific outcomes. It's applicable for stock price prediction by constructing a linear regression model that projects stock prices based on historical data (Cakra and Distiawan Trisedya 2015). The initial step involves data collection and preprocessing. Post-preprocessing, the relevant predictor variables are selected. Once the model is set up, it becomes instrumental in anticipating future stock prices.

Procedure

Step 1: Take note of both dependent and independent factors when collecting data.
Step 2: Clean, normalize, and divide data into training and testing sets as part of the preprocessing step.
Step 3: Select a model: Based on the variables, decide between basic and multiple linear regression.
Step 4: Use Mean Squared Error (MSE) as the loss function to define loss.
Step 5: Apply Gradient Descent to reduce MSE in order to optimize coefficients.
Step 6: Train the model by utilizing training data to iteratively change the coefficients.
Step 7: Evaluation: Model performance is evaluated using testing data (MSE, R-squared, etc.).
Step 8: Interpret Coefficients: Examine coefficients for correlations between variables.
Step 9: Predict: Use a trained model to provide predictions.

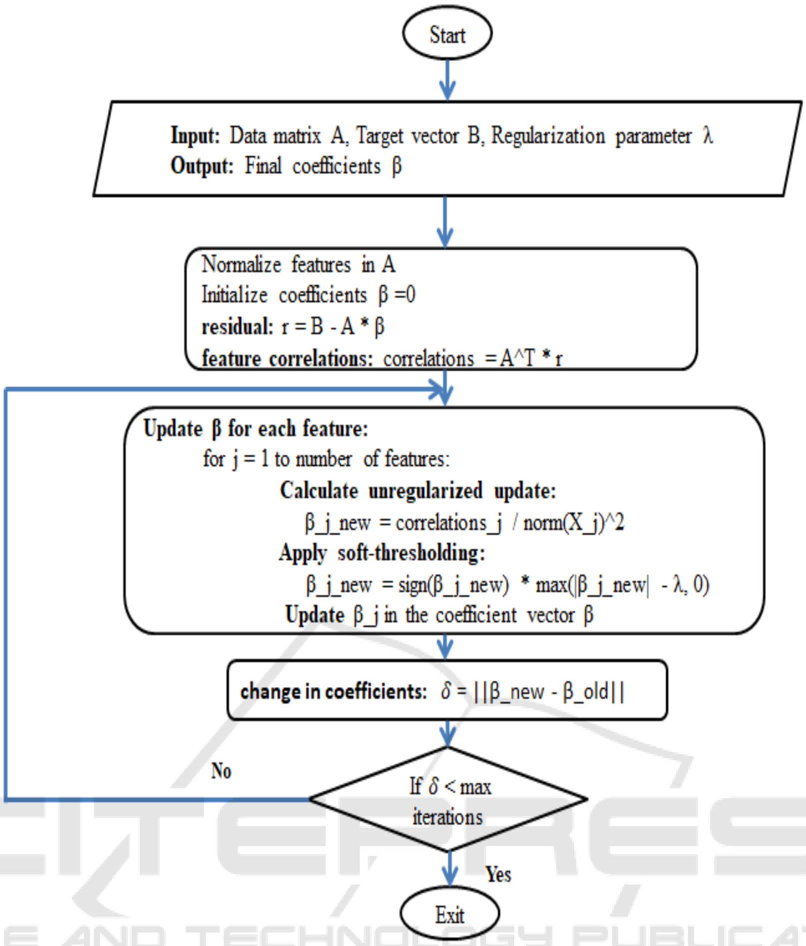

Figure 1: Flowchart.

3 STATISTICAL ANALYSIS

IBM SPSS Version 21 was utilised for the data analysis (Zhang 2022). In this analysis, accuracy served as the dependent variable, while the independent variables comprised price, objects, low, modulation, medium, and high. Iterations, capped at 10 samples, were conducted for both the proposed and existing algorithms. For every iteration, the predicted accuracy was recorded to analyse performance. The data gathered from these iterations was then subjected to the Independent Sample T-test

4 RESULT

Table 1 details the accuracy analysis of features extracted statistically from the data for training both the Novel LASSO and Linear Regression algorithms. The statistics extracted encompass the mean, standard deviation, minimum, 25% quantile, 50% quantile, 75% quantile, and maximum.

As illustrated in Table 2, the LASSO technique achieved an accuracy with a mean value of 85.31, a standard deviation of 0.74162, and a mean standard error of 0.23452. In contrast, the Linear Regression method presented a mean of 77.44, standard deviation of 0.73270, and a mean standard error of 0.23170. Notably, the independent sample t-test from Table 3 produced a significance value below 0.046 ($p<0.05$), affirming our hypothesis's validity.

Figure 2 displays the average accuracy in forecasting Amazon's stock price via both the LASSO and linear regression methods. The LASSO model's mean accuracy stood at 85.31%, while the Linear Regression's was 77.44%. Evidently, the LASSO technique surpassed the Linear Regression in performance.

Table 1: Accuracy Analysis of Novel LASSO and Linear Regression Algorithm.

Iterations	Novel LASSO Accuracy (%)	Linear Regression Accuracy (%)
1	84.31	76.34
2	84.54	76.45
3	84.79	76.67
4	84.98	76.98
5	85.14	77.25
6	85.43	77.58
7	85.69	77.89
8	85.93	78.17
9	86.24	78.28
10	86.57	78.56

Table 2: Group Statistics Results for Novel LASSO and Linear Regression algorithms.

	Group	N	Mean	Std. Deviation	Std. Error Mean
Accuracy	LASSO	10	85.3100	0.74162	0.23452
	Linear Regression	10	77.4430	0.73270	0.23170

Table 3: Independent sample test for the determination of the standard error. Independent samples t-tests were used to obtain the P-value, which was found to be 0.046, less than the 0.05 threshold for statistical significance. The 95% confidence intervals were also calculated.

<table>
<tr><th colspan="2" rowspan="2"></th><th colspan="2">Levene's test for equality of variances</th><th colspan="7">T-test for equality means with 95% confidence interval</th></tr>
<tr><th>f</th><th>Sig.</th><th>t</th><th>df</th><th>Sig. (2-tailed)</th><th>Mean difference</th><th>Std.Error difference</th><th>Lower</th><th>Upper</th></tr>
<tr><td rowspan="2">Accuracy</td><td>Equal variances Assumed</td><td rowspan="2">0.011</td><td rowspan="2">0.919</td><td>0.098</td><td>18</td><td>0.0461</td><td>0.13500</td><td>1.38137</td><td>-2.76714</td><td>3.03714</td></tr>
<tr><td>Equal Variances not Assumed</td><td>23.863</td><td>17.997</td><td>0.001</td><td>7.86700</td><td>0.32967</td><td>7.17437</td><td>8.55963</td></tr>
</table>

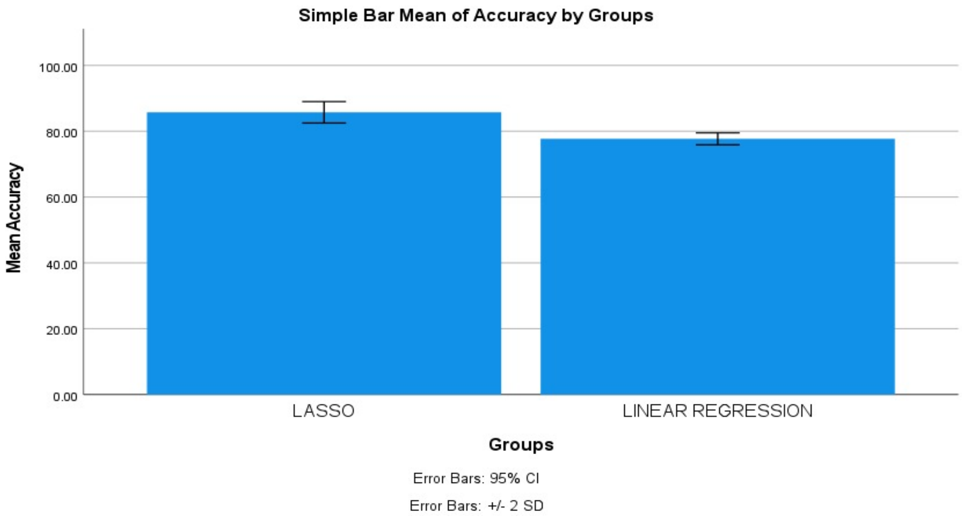

Figure 2: The mean accuracy of the LASSO method and linear regression are compared in a bar graph. A bar graph showing the mean accuracy gain comparison between the LASSO method and Linear Regression. LASSO approach has a higher mean precision than linear regression. Mean detection accuracy is +/- 2SD.

5 DISCUSSION

In the study under discussion, the LASSO algorithm predicted Amazon's stock price with 85.31% accuracy, whereas the linear regression method achieved 77.44%. Using independent samples t-tests for statistical analysis, a significant difference in the accuracy of the two algorithms was identified, with a value of 0.046 ($p<0.05$).

Predicting stock prices is a complex endeavour, given the myriad factors like market conditions, economic shifts, corporate performance, and investor sentiment, all potentially influencing a stock's value (Obthong et al. 2020). Despite the plethora of data, many scholars agree that no single prediction model can consistently forecast stock prices with complete accuracy, given the multifarious variables at play (Sen and Chaudhuri 2018). With myriad variables affecting stock prices, any decision based on predictions must be made judiciously, considering a spectrum of possible outcomes. While linear regression is beneficial for discerning relationships between variables, it might be constrained in capturing intricate relationships (Vijh et al. 2020). Thus, there may be a need to adopt advanced approaches, such as machine learning algorithms, for heightened predictive accuracy. Like other companies, Amazon's stock predictions are not immune to market vagaries, which can impinge on the precision and dependability of stock price forecasts (Bhimani 2019). However, Amazon's continued growth, forays into new domains like cloud computing and digital advertising, offer a promising future for stock price predictions (Siahaan and Sianipar 2022). If Amazon sustains its innovation and market expansion, it might well uphold or augment its stock value (Ta 2020).

In conclusion, while Amazon's stock price predictions carry inherent uncertainties, the company's potential for growth remains intact. It's essential to be cognisant of these uncertainties when considering investments and to treat stock price predictions as one among many decision-making tools. Furthermore, seeking advice from financial experts prior to investment decisions is always prudent.

6 CONCLUSION

Predicting stock prices has always been a challenge for analysts and investors alike. The complexity arises from the numerous variables that can influence a stock's price, ranging from a company's financial performance to broader economic trends and even investor sentiment. This study attempted to decipher some of this complexity, specifically targeting Amazon's stock price using two prediction models: LASSO and Linear Regression. A deeper dive into the results and surrounding contexts reveals several points of note.

1. Model Complexity and Interpretability: LASSO's regularization property aids in simplifying the model by selecting only the most crucial features,

making it more interpretable. This contrasts with Linear Regression, which might consider all variables, potentially overcomplicating the model.
2. Feature Selection: LASSO inherently performs feature selection. By introducing a penalty to the absolute values of model coefficients, it reduces some of them to zero, effectively removing less important features, which might be beneficial in stock price predictions where numerous variables can influence the outcome.
3. Robustness Against Overfitting: LASSO tends to be more resilient against overfitting compared to traditional Linear Regression, especially in scenarios where there's a risk of fitting the model too closely to the training data. This is crucial in stock price predictions, given the volatile nature of stock markets.
4. Statistical Significance: The p-value of 0.0461 from the independent sample t-test is crucial as it indicates a statistically significant difference between the results of the two algorithms. This adds weight to the conclusion that one method might be superior to the other.
5. Future Applications: The results suggest that LASSO might be a more suitable technique for other stock price predictions or in scenarios with vast amounts of data and numerous features. The adaptability of LASSO to other complex predictive scenarios warrants further exploration.
6. External Influences: It's crucial to understand that while the LASSO algorithm shows higher accuracy in this study, stock price prediction is influenced by numerous external factors. Market conditions, economic shifts, and company performance, among other variables, will always introduce an element of unpredictability.

In conclusion, while the mean accuracy of the LASSO algorithm stands at 85.31% in predicting Amazon's stock price, the Linear Regression approach lags slightly with 77.44%. It becomes evident that, in this context, LASSO seems to offer a more accurate prediction model. The independent sample t-test, backed by a p-value of 0.0461, further solidifies the significance of this difference. Given the insights, it's compelling to lean towards models like LASSO when faced with multifaceted prediction tasks such as stock price forecasting.

REFERENCES

Bhattacharjee, Indronil, and Pryonti Bhattacharja. 2019. "Stock Price Prediction: A Comparative Study between Traditional Statistical Approach and Machine Learning Approach." 4th International Conference on Electrical Information and Communication Technology 1–6.

Bhimani, Shabbir. 2019. Right Stock at Right Price for Right Time. Independently Published.

Cakra, Yahya Eru, and Bayu Distiawan Trisedya. 2015. "Stock Price Prediction Using Linear Regression Based on Sentiment Analysis." In 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS), 147–54.

Cleophas, Ton J., and Aeilko H. Zwinderman. 2011. Statistical Analysis of Clinical Data on a Pocket Calculator: Statistics on a Pocket Calculator. Springer Science & Business Media.

Hu, Zexin, Yiqi Zhao, and Matloob Khushi. 2021. "A Survey of Forex and Stock Price Prediction Using Deep Learning." Applied System Innovation 4 (1): 9.

Ji, Xuan, Jiachen Wang, and Zhijun Yan. 2021. "A Stock Price Prediction Method Based on Deep Learning Technology." International Journal of Crowd Science 5 (1): 55–72.

Kishore Kumar, M. Aeri, A. Grover, J. Agarwal, P. Kumar, and T. Raghu, "Secured supply chain management system for fisheries through IoT," Meas. Sensors, vol. 25, no. August 2022, p. 100632, 2023, doi: 10.1016/j.measen.2022.100632.

Montgomery, Douglas C., Elizabeth A. Peck, and G. Geoffrey Vining. 2021. Introduction to Linear Regression Analysis. John Wiley & Sons.

Obthong, Mehtabhorn, Nongnuch Tantisantiwong, Watthanasak Jeamwatthanachai, and Gary Wills. 2020. "A Survey on Machine Learning for Stock Price Prediction: Algorithms and Techniques." In , 9.

Palanivelu, J., Thanigaivel, S., Vickram, S., Dey, N., Mihaylova, D., & Desseva, I. (2022). Probiotics in functional foods: survival assessment and approaches for improved viability. Applied Sciences, 12(1), 455.

Ramkumar, G. et al. (2021). "An Unconventional Approach for Analyzing the Mechanical Properties of Natural Fiber Composite Using Convolutional Neural Network" Advances in Materials Science and Engineering vol. 2021, Article ID 5450935, 15 pages, 2021. https://doi.org/10.1155/2021/5450935

Ranstam, J., and J. A. Cook. 2018. "LASSO Regression." The British Journal of Surgery 105 (10): 1348–1348.

Schöneburg, E. 1990. "Stock Price Prediction Using Neural Networks: A Project Report." Neurocomputing 2 (1): 17–27.

Sen, J., and T. D. Chaudhuri. 2018. "Stock Price Prediction Using Machine Learning and Deep Learning Frameworks." On Business Analytics and Intelligence. https://www.academia.edu/download/57743652/ICBAI_2018_Final_Paper.pdf.

Siahaan, Vivian, And Rismon Hasiholan Sianipar. 2022. Amazon Stock Price: Visualization, Forecasting, And Prediction Using Machine Learning with Python Gui. Balige Publishing.

Soni, Vishal Dineshkumar. 2020. Prediction of Stock Market Values Using Artificial Intelligence.

Ta, Ton Viet. 2020. Price-Forecasting Models for Amazon Inc AMZN Stock.

Vickram, A. S., Srikumar, P. S., Srinivasan, S., Jeyanthi, P., Anbarasu, K., Thanigaivel, S., ... & Rohini, K. (2021). Seminal exosomes–an important biological marker for various disorders and syndrome in human reproduction. Saudi journal of biological sciences, 28(6), 3607-3615.
Vijh, Mehar, Deeksha Chandola, Vinay Anand Tikkiwal, and Arun Kumar. 2020. "Stock Closing Price Prediction Using Machine Learning Techniques." Procedia Computer Science 167 (January): 599–606.
V. P. Parandhaman, "An Automated Efficient and Robust Scheme in Payment Protocol Using the Internet of Things," 2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), Chennai, India, 2023, pp. 1-5, doi: 10.1109/ICONSTEM56934.2023.10142797.
Yu, Pengfei, and Xuesong Yan. 2020. "Stock Price Prediction Based on Deep Neural Networks." Neural Computing & Applications 32 (6): 1609–28.
Zhang, Haorui. 2022. "Stock Price Prediction Using Linear Regression and LSTM Neural Network." 2022 International Conference on Machine Learning and Intelligent Systems Engineering (MLISE). https://doi.org/10.1109/mlise57402.2022.00043.