Proceedings of the 2nd International Conference on Financial Technology and Business Analysis
DOI: 10.54254/2754-1169/60/20231182

Stock Price Prediction of GM Company: Comparison Based on KNN, Linear Regression and LSTM

Congyan Yin[1,a,*]

[1]*College of Arts and Sciences, The Ohio State University, Columbus, USA*
a. Yin.750@osu.edu
**corresponding author*

Abstract: Stock price prediction holds significant importance in the financial sector. It not only aids investors in making informed buy and sell decisions to achieve potential profits but also supports enterprises and financial institutions in risk assessment and management. This study utilized the stock prices of GM stocks spanning from 2013 to 2023 as dataset, with the goal of predicting their closing prices. This paper carefully identifies the parameters in the three model and one indicator of RMSE outcomes were computed. Visualizations of the results of the three model predictions are also provided. After comparing the indicator of the three models, it is found that the RMSE for LSTM is the smallest, indicating that the LSTM outperforms the KNN and linear regression from the perspective of forecast accuracy. The research highlights the application of the machine learning algorithm of LSTM in the prediction of the prices of financial assets and its practical value in financial market analysis.

Keywords: LSTM, KNN, linear regression

1. Introduction

The automotive sector is undeniably a cornerstone of the global economy. As potentially the most extensive manufacturing sector worldwide, its management strategies and structural approaches are meticulously crafted. The products of this industry touch the lives of millions daily, offering mobility and, in turn, presenting multifarious challenges [1]. Serving as a primary engine of the global economy, the automotive industry not only generates a vast number of jobs but also stands at the forefront of technological innovation and infrastructure development. Consequently, delving into the intricacies of the automotive industry is of paramount importance. Furthermore, predicting the stock trends of the automotive sector is not only equally crucial but also multifaceted in its significance. This not only helps investors more accurately assess the future performance of the industry but also offers deep insights into market dynamics and potential investment opportunities. Through meticulous analysis of this sector, investors can better understand how various factors such as consumer demand, technological advancements, and policy changes impact stock prices. Therefore, accurate prediction of stock trends in the automotive sector is not only key to asset allocation but also an effective way to maximize investment returns.

Historically, the prediction of stock prices has predominantly relied on individual machine learning models, such as nonlinear time series models [2], artificial neural networks [3-5], decision trees [6,7], genetic algorithms [8,9], and support vector machines [10]. While each method has its merits, there has been a noticeable lack of lateral comparisons in existing research. To address this

DOI: 10.54254/2754-1169/60/20231182

gap, this paper adopts three models: KNN, LSTM, and Linear Regression. In this study, a decade's worth of historical stock price data for General Motors Company (GM) was utilized to employ three different models—KNN, Linear Regression, and LSTM—for forecasting its stock prices. The RMSE values of these models were then compared, and the LSTM model, having the lowest RMSE, was selected for its superior accuracy in predicting this stock data. Finally, an optimized LSTM model was used to backtrack the past 100 days of data, generating forecasts for GM's stock prices, complete with data visualization.

2. Methodology

2.1. KNN

KNN is a non-parametric technique primarily used for classification and regression tasks in pattern recognition [11]. This algorithm classifies objects based on the closest training samples in the feature space. It adopts an instance-based learning strategy, also known as lazy learning, where all computations are deferred until classification [11].

For classification, KNN examines the k-nearest training samples of an input vector and returns the most common class label. Specifically, when K=1, the classification of a sample is like its neighboring samples [11]. In regression, the output is the average of the values of k neighbors (see Figure 1).

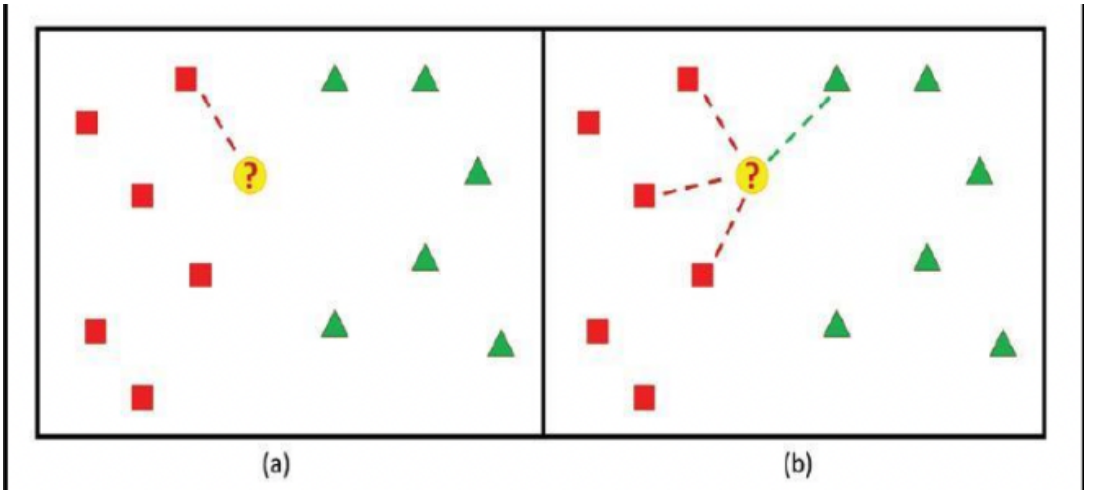

Figure 1: KNN model [11].

The first step in using KNN is to choose a distance metric, such as Euclidean, Manhattan, and Minkowski distances. The choice of K is crucial for model performance, and its performance largely depends on the value of K and the chosen distance metric [11].

When determining the optimal K value, cross-validation is typically used. In short, multiple k-values can be tested, their performance evaluated, and the best value selected [11].

2.2. Linear Regression

Linear Regression is a cornerstone in statistical modeling, pivotal for predicting stock prices based on various indicators. This method aims to decipher relationships between stock prices and predictors via an optimally fitted regression line.

Before utilizing linear regression for stock predictions, one must be mindful of key assumptions, including linearity, observation independence, and error normality. Breaching these can skew model accuracy.

DOI: 10.54254/2754-1169/60/20231182

For evaluating model performance in stock prediction, the R-squared metric is instrumental, indicating how much stock price variance is explained by the model. Meanwhile, adjusted R-squared provides a refined view, factoring in the number of predictors, ensuring a robust assessment of model effectiveness.

2.3. LSTM

Long Short-Term Memory (LSTM) networks, a sophisticated variant of Recurrent Neural Networks (RNN), arose as a groundbreaking solution in 1997. Their inception was primarily to address the vanishing gradient challenge, a notorious issue that beset traditional RNNs [12]. The crux of LSTMs lies in their meticulously designed architecture. Central to this design is the memory cell, a unique feature that empowers LSTMs with the ability to remember and utilize long-term dependencies in sequential data. This memory cell operates in harmony with three pivotal gates:

Input Gate: Modulates the entry of new information into the memory cell.

Forget Gate: Deciphers which facets of the stored data ought to be retained or jettisoned, ensuring relevance in the memory cell's content.

Output Gate: Influences how the memory cell's content impacts the LSTM unit's current output (see Figure 2).

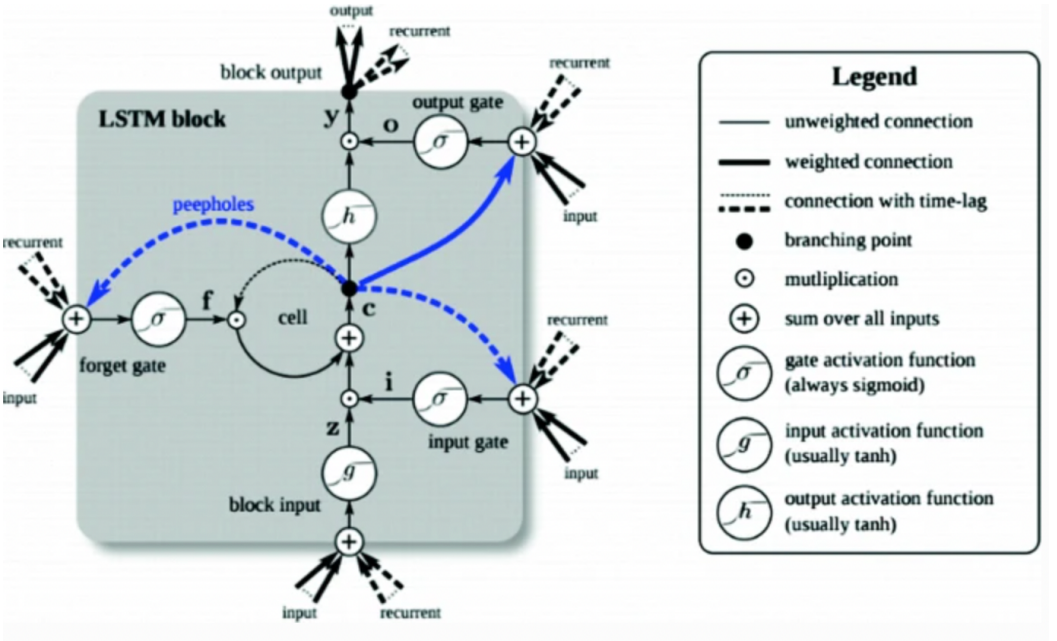

Figure 2: LSTM model [13].

These gates, namely input, forget, and output, meticulously regulate the information flow, thereby ensuring that the LSTM not only learns efficiently from historical data but also sidesteps redundancy. Such attributes have catapulted LSTMs to the forefront, especially when it comes to deciphering patterns in sequential datasets like stock prices [14].

3. Results

3.1. KNN

In this paper, the objective was to predict the closing price of the 'GM' stock using the KNN model, with data spanning from July 25, 2013, to July 25, 2023. To ensure a rigorous approach, the Time Series Cross-Validation method was employed, segmenting the dataset into three

DOI: 10.54254/2754-1169/60/20231182

distinct portions. All feature data, namely 'High', 'Low', 'Open', and 'Volume', were standardized to maintain consistency in scale. A significant aspect of the methodology was determining the optimal K value for the KNN model. Through cross-validation, a range of K values from 1 to 150 was explored, and the RMSE for each was calculated. Based on the analysis, this paper identified the optimal K value as 111 (see Figure 3).

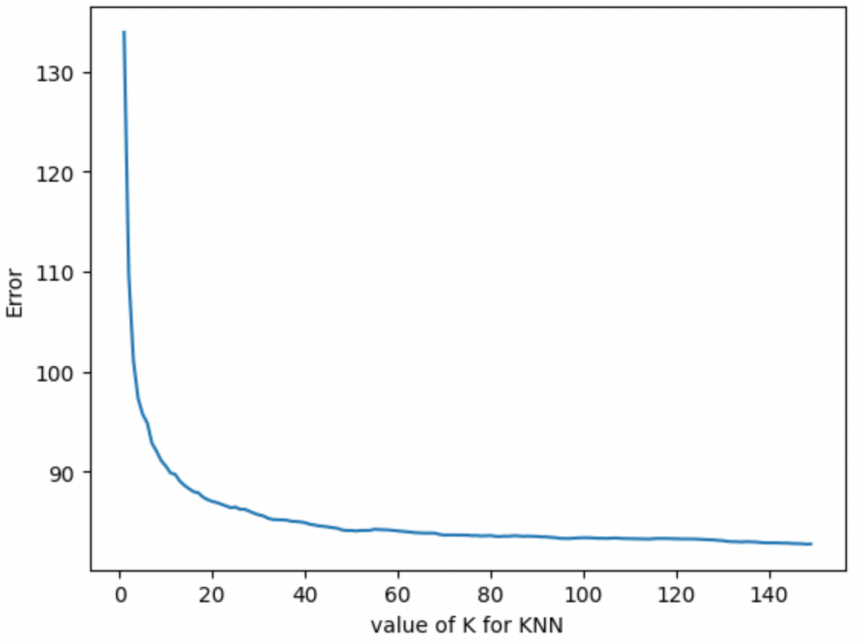

Figure 3: Error Value VS. K Value.

Using this value, predictions were made, and when compared with the actual closing prices, the model yielded an RMSE of 8.3389, highlighting the model's performance on this dataset (see Figure 4).

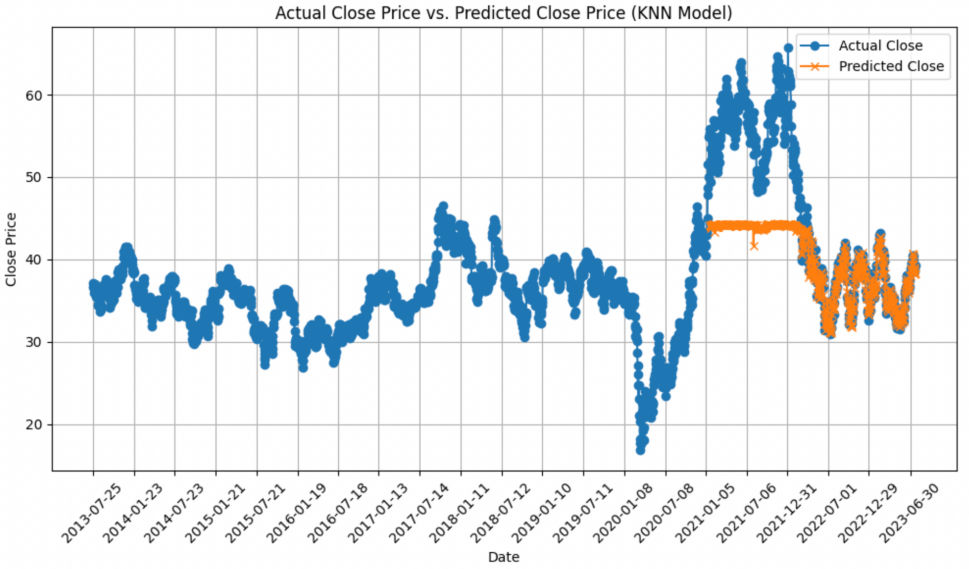

Figure 4: KNN Predict Result.

DOI: 10.54254/2754-1169/60/20231182

3.2. Linear Regression

In this study, the primary objective was to harness the capabilities of a Linear Regression model for predicting future stock prices. The dataset, encompassing features such as 'High', 'Low', 'Open', and 'Volume', was meticulously processed. Before modeling, these features underwent a standardization process using the StandardScaler, ensuring they operated on a consistent scale, which is pivotal for the accuracy of linear models. After training the model on this standardized data, predictions were made on a test set. The model's performance was then visually represented through a graph, juxtaposing the actual and predicted closing prices. This figure 5, titled "Actual Close Price vs. Predicted Close Price (Linear Regression Model)", provides a clear visual representation of the model's predictive prowess and its alignment with actual values. The RMSE is calculated to be 5.4635.

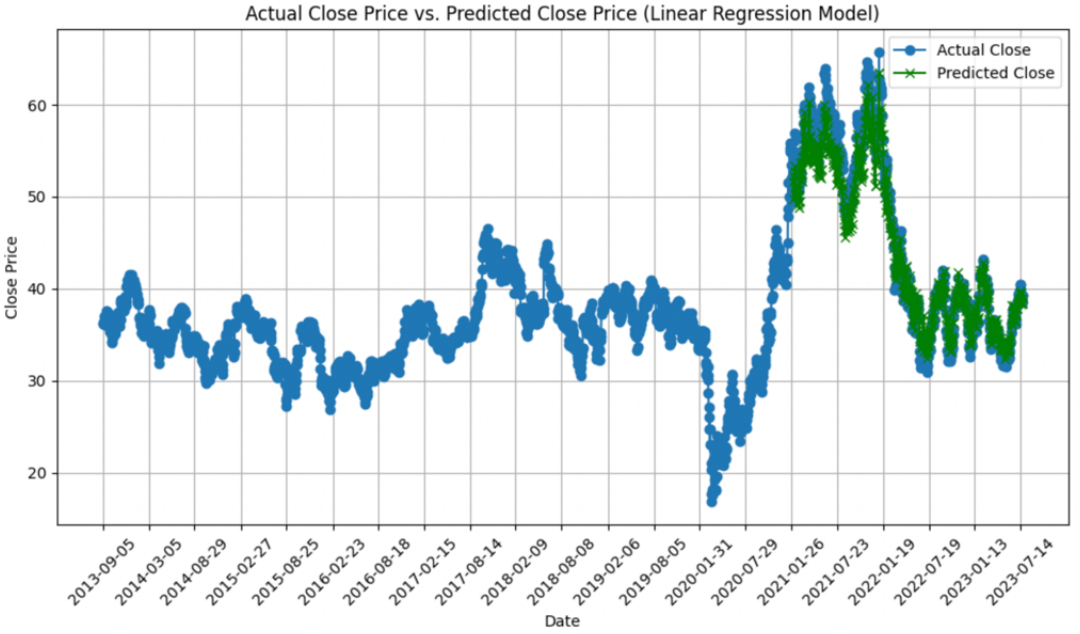

Figure 5: Linear Regression Predict Result.

3.3. LSTM

In the presented study, the primary objective was to predict the closing prices of automotive industry stocks using a dataset spanning from 2013 to 2023. The dataset, primarily composed of the 'Close' prices of the stocks, underwent preprocessing where the closing prices were scaled between 0 and 1 using the MinMaxScaler. Unlike traditional random partitioning methods, the dataset was divided using TimeSeriesSplit with 3 splits to ensure the temporal sequence of the data. This resulted in multiple training and testing sets, each with a different time range. The training and testing sets were then transformed into a format suitable for LSTM models, with each sequence containing 100 time steps.

For model input, the data was transformed into sequences with a time step of 100 days, implying that the stock prices of the preceding 100 days were utilized to forecast the subsequent day's closing price.

The LSTM model architecture was designed with three layers, each containing 50 units. To mitigate overfitting, a dropout rate of 0.2 was incorporated after each layer. The model was compiled with the mean squared error as the loss function and employed the Adam optimizer. During its training phase, the model was subjected to 10 epochs with a batch size of 32.

DOI: 10.54254/2754-1169/60/20231182

Additionally, a grid search approach was adopted to optimize the model's parameters. This exploration considered LSTM units of 50 and 100, while other parameters such as dropout rate, batch size, number of layers, activation function, and optimizer remained constant (see Figure 6).

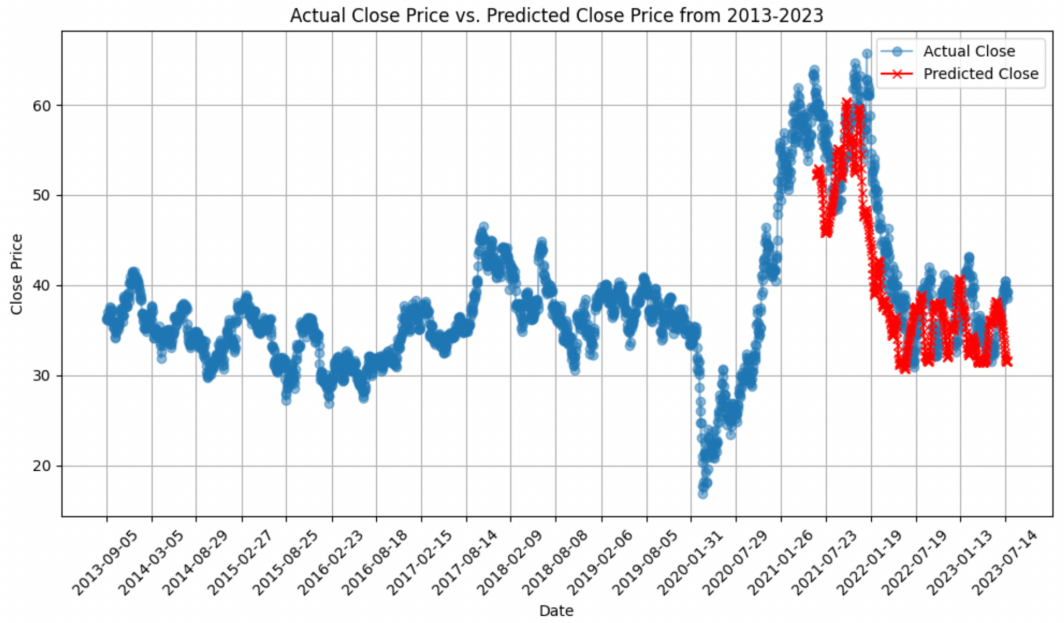

Figure 6: LSTM Predict Result.

Upon the culmination of the training phase, the model was deployed to predict the closing prices on the test set. The predicted values underwent inverse transformation to derive the actual stock price predictions. A juxtaposition of the predicted and actual prices yielded an RMSE of (2.8882), signifying the model's efficacy in forecasting stock prices (see Table 1).

Table 1: The RMSE Comparison Results.

Model	RMSE Value
KNN	8.3389
Linear Regression	5.4635
LSTM	2.8882

4. Conclusion

This study embarked on the pivotal task of stock price prediction, emphasizing its profound relevance in guiding investment decisions and bolstering risk management in the financial realm. Utilizing a decade-long dataset of GM automotive Company stock, three predictive models - KNN, linear regression, and LSTM - were evaluated. Based on the RMSE metrics, the performance of the three models varies significantly. The KNN model has an RMSE value of 8.3389, the Linear Regression model stands at 5.4635, while the LSTM model boasts the lowest RMSE of 2.8882. This data clearly indicates that the LSTM model outperforms its counterparts in terms of prediction accuracy. The LSTM model was chosen to forecast for stock prices, yielding insightful visual representations. However, while the LSTM model demonstrated superiority in this context, it's essential to acknowledge the study's confinement to a specific

DOI: 10.54254/2754-1169/60/20231182

dataset and sector. Exploring diverse datasets and integrating other advanced models might offer even more nuanced insights in future endeavors.

References

[1] *Orsato, R. J., & Wells, P. (2007). The automobile industry & sustainability. Journal of cleaner production, 15(11-12), 989-993.*

[2] *Peng, Z. Y. (2004). Non-linear characteristics and prediction models of Shanghai and Shenzhen stock markets (Doctoral dissertation, Wuhan University of Technology).*

[3] *Hu, P. (2014). Stock prediction model based on ANN method (Doctoral dissertation, Chongqing University).*

[4] *Li, L. (2015). A comparative study on stock price prediction ability based on GARCH and BP-ANN (Doctoral dissertation, Southwest Jiaotong University).*

[5] *Zhou, B. (2009). A study on stock price index prediction based on GA's ANN (Doctoral dissertation, Wuhan University of Technology).*

[6] *Tao, Y. Y. (Year). Application of decision tree and neural network algorithms in stock classification prediction (Doctoral dissertation, Hangzhou Dianzi University).*

[7] *Wang, Y., Chen, D., & Tang, Y. (2019). Stock prediction based on CART decision tree and boosting methods. Journal of Harbin University of Science and Technology, 24(6), 6.*

[8] *Meng, X., & Liu, X. (1997). Stock market modeling and prediction based on fuzzy neural networks and genetic algorithms. Information and Control, 26(5), 5.*

[9] *Lu, Q., Ye, D., & Nan, M. (2010). Stock price prediction based on genetic algorithms and neural networks. Computer Development and Applications, 23(2), 2.*

[10] *Chen, Y. (Year). Research on stock prediction models (Doctoral dissertation, Harbin Engineering University).*

[11] *Imandoust, S. B., & Bolandraftar, M. (2013). Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background. International journal of engineering research and applications, 3(5), 605-610.*

[12] *Svozil, D., Kvasnicka, V., & Pospichal, J. (1997). Introduction to multi-layer feed-forward neural networks. Chemometrics and intelligent laboratory systems, 39(1), 43-62.*

[13] *Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A search space odyssey. IEEE transactions on neural networks and learning systems, 28(10), 2222-2232.*

[14] *Lipton, Z. C., Kale, D. C., Elkan, C., & Wetzel, R. (2015). Learning to diagnose with LSTM recurrent neural networks. arXiv preprint arXiv:1511.03677.*