ADROID Conference Series: Engineering Sciences and Technology (ACS: EST)

International Conference on Artificial Intelligence for Sustainable Development, Jaypee University Anoopshahr, India 2025

Volume 01, October 2025, pp. 34-43

DOI: https://doi.org/10.63503/c.acset.2025.4

COMPARATIVE ANALYSIS OF LSTM, ARIMA, AND LINEAR REGRESSION FOR E-COMMERCE SALES PRICE PREDICTION

Hare Ram Sah [1] and Apoorv Vyas [2]

[1, 2] Institute of Advance Computing, SAGE University, INDORE, INDIA

apoorvvyas9@gmail.com

Abstract. In the rapidly evolving landscape of e-commerce, accurate price prediction and sales forecasting have become vital for operational efficiency, inventory management, and strategic decision-making. This study presents a comprehensive methodology employing preprocessing techniques, normalization strategies, data splitting, and advanced modeling—specifically Linear Regression, Long Short-Term Memory (LSTM), and AutoRegressive Integrated Moving Average (ARIMA) to predict product sales prices using an E-Commerce Sales dataset. The raw data undergoes missing-value handling and column pruning to enhance relevancy. Subsequently, normalization is applied through both MinMax and Standard Scaler techniques to ensure scale uniformity. The preprocessed dataset is then partitioned into training and test sets, facilitating both model evaluation and predictive capability assessment. Performance is quantified using key metrics—Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) accompanied by comparative visualizations. Our findings demonstrate the strengths and trade-offs of each algorithm in forecasting accuracy, showcasing their potential for real-world e-commerce applications. By illustrating the comparative performance of traditional statistical techniques and deep learning models, the study provides a roadmap for stakeholders to choose appropriate forecasting tools aligned with their business needs.

Keywords: E-Commerce, Sales Price Prediction, Machine Learning, Time Series Forecasting, LSTM (Long Short-Term Memory), ARIMA

1. INTRODUCTION

E-commerce has transformed retail by enabling continuous, dynamic pricing strategies across global platforms such as Amazon, eBay, and Walmart. These platforms harness vast amounts of historical price data, coupled with external signals like search trends and customer sentiment, to influence pricing decisions and competitiveness. Rising consumer expectations around discounts, fast delivery, and price transparency have intensified the need for accurate automated forecasting tools that help brands and marketplaces adapt in real time.

Strategic price prediction improves inventory management, demand planning, and revenue optimization, which together support profitability in highly volatile retail environments. The emergence of price trackers like CamelCamelCamel—which monitor and alert consumers about pricing dips—reflects the growing value of price-forecasting analytics for both consumers and sellers.

Nowadays, the development of e-commerce, there is a wide variety of data available online. Typically, descriptive information like photos is still present, even when new goods don't have prior interaction data. On the flip side, you'll see a lot more user evaluation information on more established products. Being both visual and textual, e-commerce relies heavily on these two types of data. Traditional methods of price prediction, on the other hand, run into Citing this article Hua H. in the year 2024. Utilizing deep neural networks and variational mode decomposition for online product price prediction. Difficulty in providing sufficient backing for all-encompassing price prediction systems due to difficulties in deriving useful insights from text and visuals. Companies rely heavily on accurate product pricing predictions when making strategic decisions in today's complicated and unpredictable market economy.Market complexity, nonlinearity, and the lack of stationary time series features present numerous hurdles to traditional price forecasting methodologies (Cortez et al., 2018; Wang et al., 2020). Traditional linear models have a hard

time reflecting the different behavioural patterns shown by product pricing over time, which is a fundamental drawback. Predicting commodity prices is a complex and important task in the world of online shopping. Price fluctuations are introduced by the dynamic life cycle of commodities (Agnello et al., 2020). While established commodities typically have a wealth of price history, newer ones may struggle with data scarcity when trying to forecast their future value owing to a lack of past interaction data. These differences may be too much for conventional price prediction systems to handle.

The majority of existing price forecasting algorithms stick to a single computational technique, ignoring the impact of the commodity life cycle. Nevertheless, the intricacies of price prediction for both new and established commodities may be too much for a single algorithm to handle, which could lead to subpar forecast results (Lago et al., 2021). As a result, a complete framework for price forecasting that incorporates many advanced forecasting methods is recommended. In order to improve the precision of price forecasts, this framework should be able to determine the stage of the commodity's life cycle using its attributes and past price data. Currently, media text sentiment, keyword, or event feature extraction is the main method for integrating unstructured data into price prediction. This requires forecasting futures prices at the same time. According to Pan and Zhou (2020), Ramkumar et al. (2023), and Sun et al. (2022), the core of relevant study is the skillful transformation of unstructured data into e-commerce product prediction. It is worth investigating the following points: firstly, there is a lot of noise in the extracted effective price features from things like analytical reports and social comments, which could affect the model's ability to spot price fluctuations. Secondly, current methods for extracting event features from text data require a lot of human annotations on certain corpora, which is prone to subjective judgment and could cause other relevant information to be overlooked. Lastly, there is still a lot of debate about how to combine structured price data with features from unstructured textual information. A number of research fail to take into consideration the fact that different studies use different financial indicator datasets, sentiment traits, and trading data when feeding into their prediction models. The decomposition integration process is said to be a great way to improve the accuracy of predictions in complex time series predictive modelling. Its core idea is to simplify modelling by dividing complicated time series into smaller, more manageable subsequences using signal decomposition algorithms (Da Silva RG et al., 2020). Doi: 10.7717/peerj-cs.2353, The Variational Wu (2024), PeerJ Computer Science. 2/22 In order to gain a more nuanced understanding of the implicit modes in the data, a signal processing approach called Mode Decomposition (VMD) skillfully breaks down complex signals into many eigenmodes. At the same time, the complex structure of the deep neural network (DNN) allows it to handle unstructured data, identify complex relationships in price fluctuations, and improve the accuracy of predictions by varying the input data (Güven¸c, Çetin & Ko¸cak, 2021). Deep neural networks (DNNs) are great at extracting features from images and text codes; this allows them to better adapt to dynamic changes in e-commerce product prices and discover nonlinear correlations. When it comes to predicting the prices of products sold on online marketplaces, the integration of VMD and DNN presents a novel method.

2. RELETED WORK

Wang, C et.al. (2025) the traditional e-commerce business chain is being reconstructed around the content of short videos and live streams, and the interest e-commerce is thriving as a new trend in the e-commerce industry. Diversified content promotes the rapid development of interest e-commerce. For consumers, their preferences for different content reflect their consumption level to a certain extent. The purpose of this study is to accurately predict the purchasing power level with the consumer content preference, and provide new ideas for interest e-commerce business. In this paper, the new swarm intelligence algorithm is used to find the optimal misclassification cost, and three cost-sensitive models are established. On this basis, the content preference of interest e-commerce consumers is used to predict the level of purchasing power. The results show that the content preference of interest e-commerce consumers, such as “fashion”, “photography” and “interpretation”, have a significant effect on the prediction of purchasing power at the 95% confidence level. The accuracies of the optimized cost-sensitive support vector machine in predicting consumer purchasing power are all above 0.9, and the highest is 0.9792. This study effectively alleviates the problem that the classification results tend to be biased towards negative samples, especially when the imbalanced rate of the sample is high. It not only provides researchers with an efficient parameter optimization method, but also reflects the relationship between consumer content preference and purchasing power, providing data support for interest e-commerce operations.

Nowak, M et.al. (2024) This study deeply integrates multimodal data analysis and big data technology, proposing a multimodal learning framework that consolidates various information sources, such as user geographic location, behavior data, and product attributes, to achieve a more comprehensive understanding and prediction of consumer behavior. By comparing the performance of unimodal and multimodal

approaches in handling complex cross-border e-commerce data, it was found that multimodal learning models using the Adam optimizer significantly outperformed traditional unimodal learning models in terms of prediction accuracy and loss rate. The improvements were particularly notable in training loss and testing accuracy. This demonstrates the efficiency and superiority of multimodal methods in capturing and analyzing heterogeneous data. Furthermore, the study explores and validates the potential of big data and multimodal learning methods to enhance customer satisfaction in the cross-border e-commerce environment. Based on the core findings, specific applications of big data technology in cross-border e-commerce operations were further explored. A series of innovative strategies aimed at improving operational efficiency, enhancing consumer satisfaction, and increasing global market competitiveness were proposed.
Gkikas, D et.al. (2024) Effective sales prediction for e-commerce would assist retailers in developing accurate production and inventory control plans, which would further help them to reduce inventory costs and overdue losses. This paper develops a systematic method for e-commerce sales prediction, with a particular focus on predicting the sales of products with short shelf lives. The short-shelf-life product sales prediction problem is poorly addressed in the existing literature. Unlike products with long shelf lives, short-shelf-life products such as fresh milk exhibit significant fluctuations in sales volume and incur high inventory costs. Therefore, accurate prediction is crucial for short-shelf-life products. To solve these issues, a stacking method for prediction is developed based on the integration of GRU and LightGBM. The proposed method not only inherits the ability of the GRU model to capture timing features accurately but also acquires the ability of LightGBM to solve multivariable problems. A case study is applied to examine the accuracy and efficiency of the GRU-LightGBM model. Comparisons among other sales prediction methods such as ARIMA and SVR are also presented. The comparative results show that the GRU-LightGBM model is able to predict the sales of short-shelf-life products with higher accuracy and efficiency. The selected features of the GRU-LightGBM model are also useful due to their interpretability while developing sales strategies.

3. PROPOSED METHODOLOGY

The architecture diagram outlines the workflow for ecommerce price prediction and management using machine learning techniques. It begins with Data Selection, where a diabetes dataset in CSV format is chosen. Data Preprocessing follows, addressing missing values and applying label encoding to prepare the data for analysis. The data is then split into Training and Testing sets to evaluate model performance. Classification models, includingLinear Regression (LR) and LSTM, are employed to analyze the data. The Result Generation step involves calculating error metrics to assess model effectiveness. Finally, Prediction determines whether an individual prediction is based on the trained models.

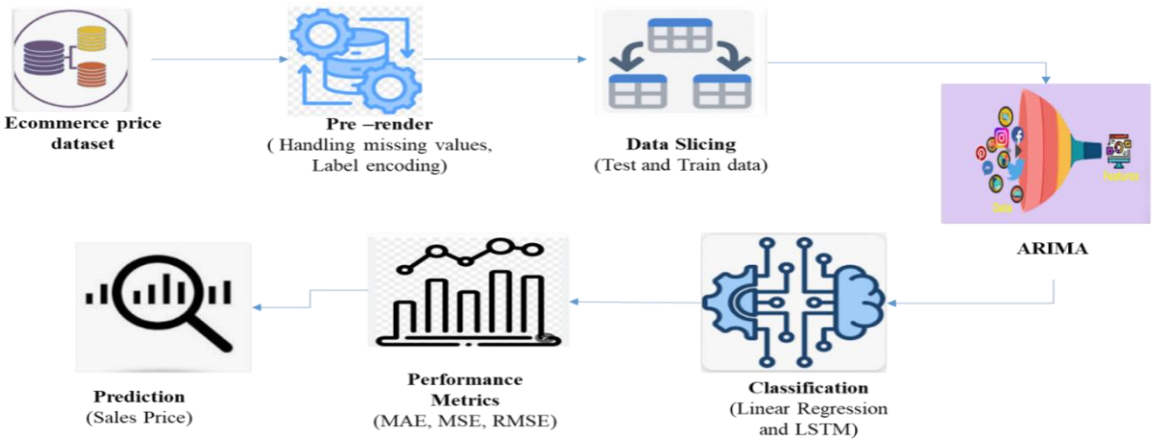


Figure. 1 Proposed Architecture

The proposed system offers a multi-stage framework for e-commerce price prediction. Initially, input data (CSV/XLSX) undergoes preprocessing: missing entries are imputed (e.g., mean, median), and irrelevant columns (like ID or textual descriptions) are dropped. Next, the cleaned data is normalized using both MinMax and Standard Scaler to compare scaling effects on model accuracy. The normalized dataset is split (e.g., 80/20) into training and test sets. Three predictive models are then developed: Linear Regression for baseline trend capture; ARIMA for short-term univariate forecasting and seasonality detection; and LSTM for modeling complex, time-dependent patterns. Each model is trained on historical data and evaluated using

MSE, MAE, and RMSE on the test set. Visualizations—such as prediction vs actual plots, comparison graphs, and residual analyses—illustrate model strengths and weaknesses. Finally, a comparative framework highlights key trade-offs in accuracy, computational cost, interpretability, and scalability. This enables practitioners to select the most appropriate model based on product specificity, data availability, and forecast horizon.

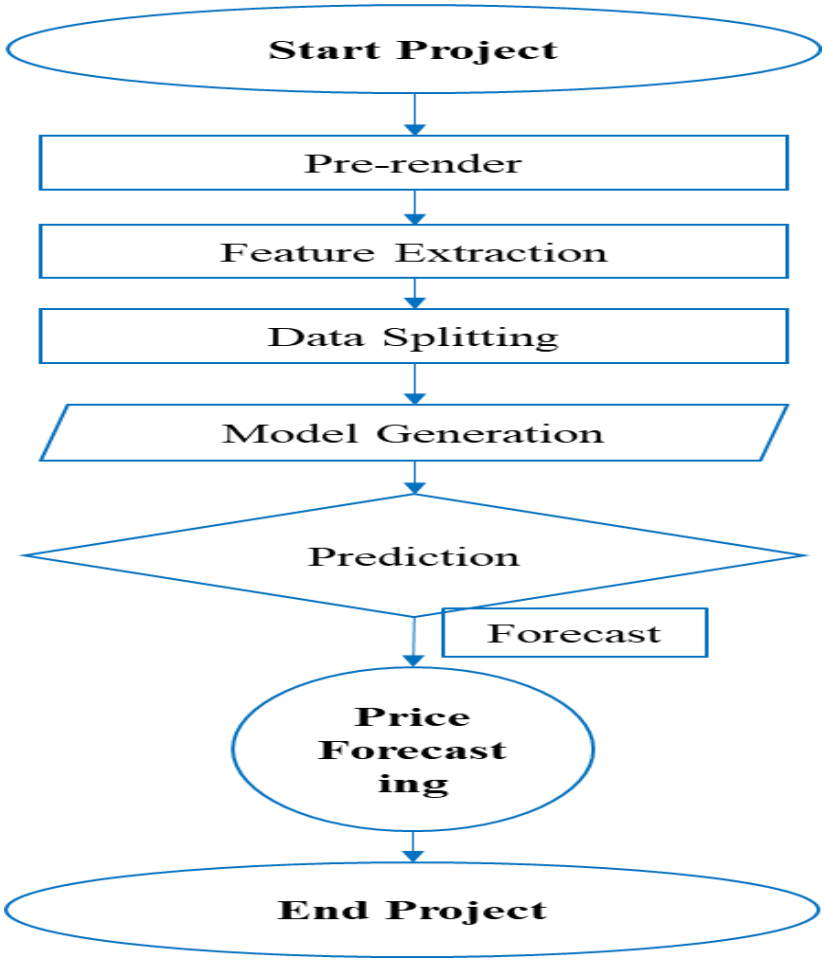


Figure 2. Block diagram

The Block diagram outlines the workflow for ecommerce price prediction and management using machine learning techniques. It begins with Data Selection, where a diabetes dataset in CSV format is chosen. Data Preprocessing follows, addressing missing values and applying label encoding to prepare the data for analysis. The data is then split into Training and Testing sets to evaluate model performance. Classification models, including Linear Regression (LR) and LSTM, are employed to analyze the data. The Result Generation step involves calculating error metrics to assess model effectiveness. Finally, Prediction determines whether an individual prediction is based on the trained models.

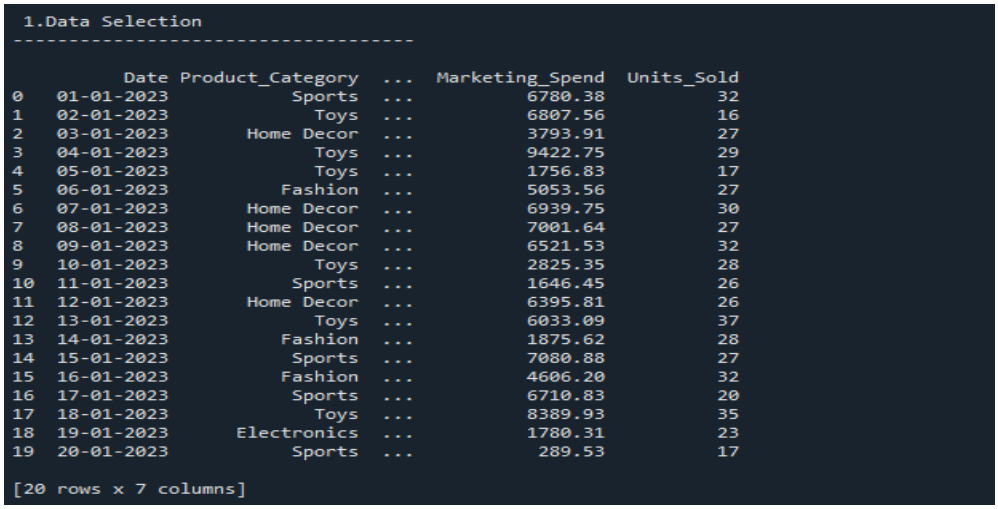
4. SIMULATION RESULT

This module evaluates the forecasted price values against actual prices using quantitative metrics.

- Key metrics include Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). These metrics reveal how far off the predictions are from real prices, helping to assess model accuracy and reliability.
- Lower error scores indicate better performance. Additionally, visualizations such as line plots, prediction vs. actual graphs, and residual plots provide an intuitive understanding of model behavior.
- Performance across the three models is compared through bar charts or tables to identify the most effective method.
- This module also helps diagnose problems like underfitting or overfitting. Time-series residuals are plotted to inspect cyclic errors or drift. The visualization helps stakeholders (e.g., business analysts)

interpret outcomes even without deep technical knowledge. It bridges the gap between raw data science and actionable insights.

The Data Flow Diagram (DFD) Level 2 provides a detailed view of the ecommerce price prediction system. It starts with Data Selection of the CSV dataset, followed by Data Pre-processing to handle missing values and apply label encoding. Data Splitting divides the data into training and testing sets. Classification involves training models like Linear Regression (LR) and LSTM. Result Generation assesses model accuracy, and Prediction determines if an individual prediction is based on the model outputs.

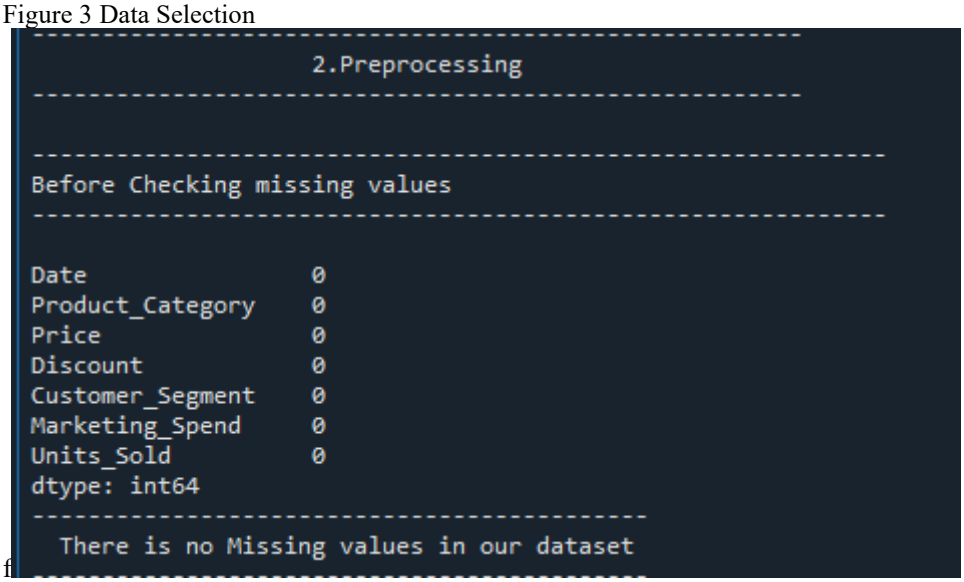


```
 1.Data Selection
----------------------------------

          Date Product_Category  ...  Marketing_Spend  Units_Sold
0   01-01-2023           Sports  ...          6780.38          32
1   02-01-2023             Toys  ...          6807.56          16
2   03-01-2023       Home Decor  ...          3793.91          27
3   04-01-2023             Toys  ...          9422.75          29
4   05-01-2023             Toys  ...          1756.83          17
5   06-01-2023          Fashion  ...          5053.56          27
6   07-01-2023       Home Decor  ...          6939.75          30
7   08-01-2023       Home Decor  ...          7001.64          27
8   09-01-2023       Home Decor  ...          6521.53          32
9   10-01-2023             Toys  ...          2825.35          28
10  11-01-2023           Sports  ...          1646.45          26
11  12-01-2023       Home Decor  ...          6395.81          26
12  13-01-2023             Toys  ...          6033.09          37
13  14-01-2023          Fashion  ...          1875.62          28
14  15-01-2023           Sports  ...          7080.88          27
15  16-01-2023          Fashion  ...          4606.20          32
16  17-01-2023           Sports  ...          6710.83          20
17  18-01-2023             Toys  ...          8389.93          35
18  19-01-2023      Electronics  ...          1780.31          23
19  20-01-2023           Sports  ...           289.53          17

[20 rows x 7 columns]
```

Figure 3 Data Selection

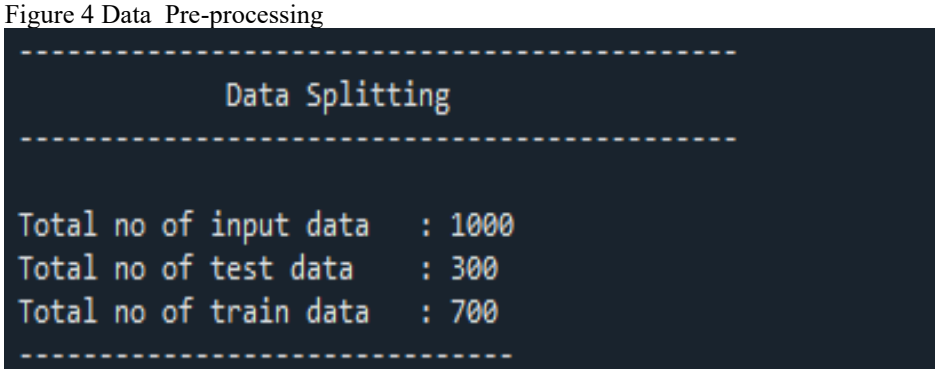


```
                    2.Preprocessing
-----------------------------------------------------


------------------------------------------------------------
Before Checking missing values
------------------------------------------------------------

Date                0
Product_Category    0
Price               0
Discount            0
Customer_Segment    0
Marketing_Spend     0
Units_Sold          0
dtype: int64
-------------------------------------------
  There is no Missing values in our dataset
-------------------------------------------
```

Figure 4 Data Pre-processing



```
-----------------------------------------------
            Data Splitting
-----------------------------------------------


Total no of input data   : 1000
Total no of test data    : 300
Total no of train data   : 700
--------------------------------
```

Figure 5 Data Splitting Process

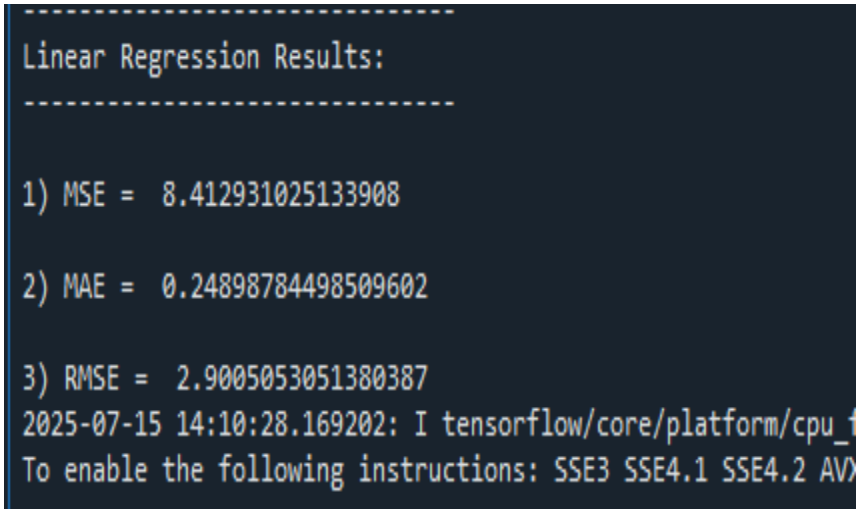


Figure 6 Result of Linear Legression

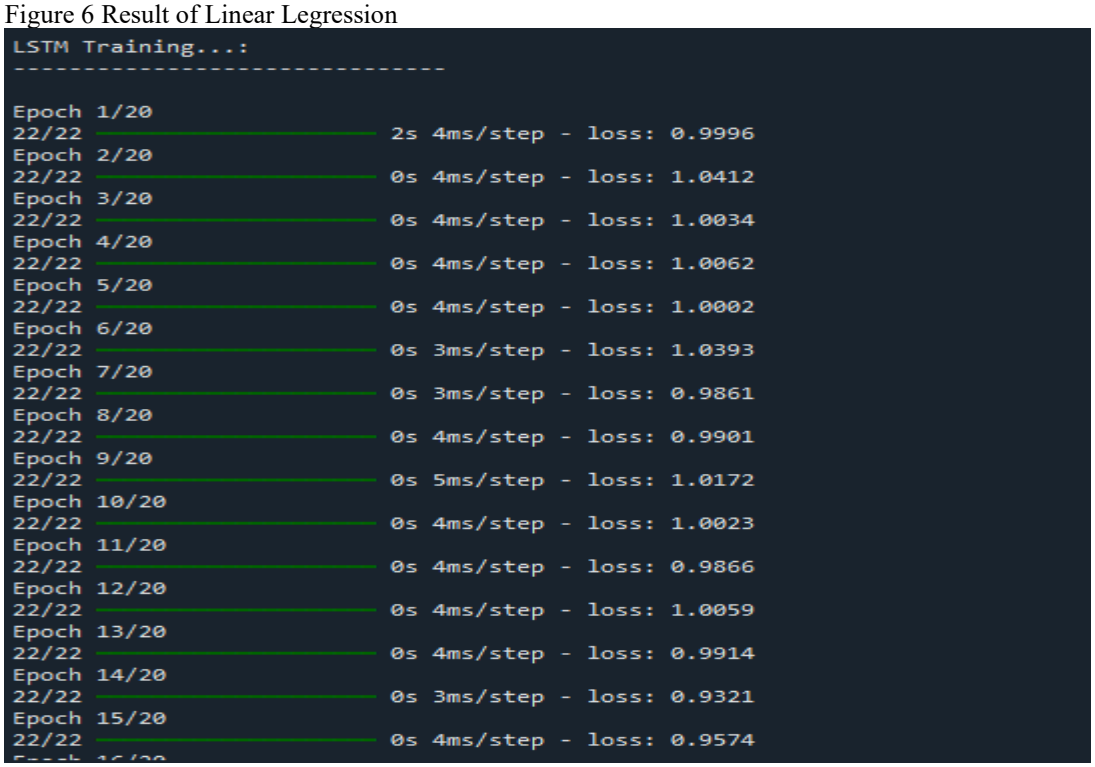


Figure 7 LSTM Training Process

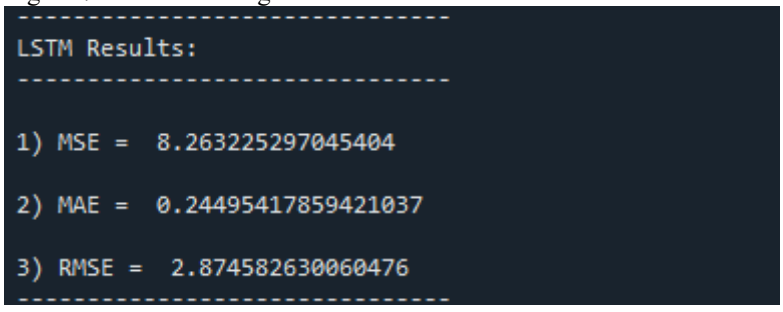


Figure 8 Result of LSTM algorithm

```
ARIMA Results:
------------------------------

1) MSE =  11.084014022525274

2) MAE =  0.27174571965188316

3) RMSE =  3.3292662889179163
```

Figure 9 ARIMA Result

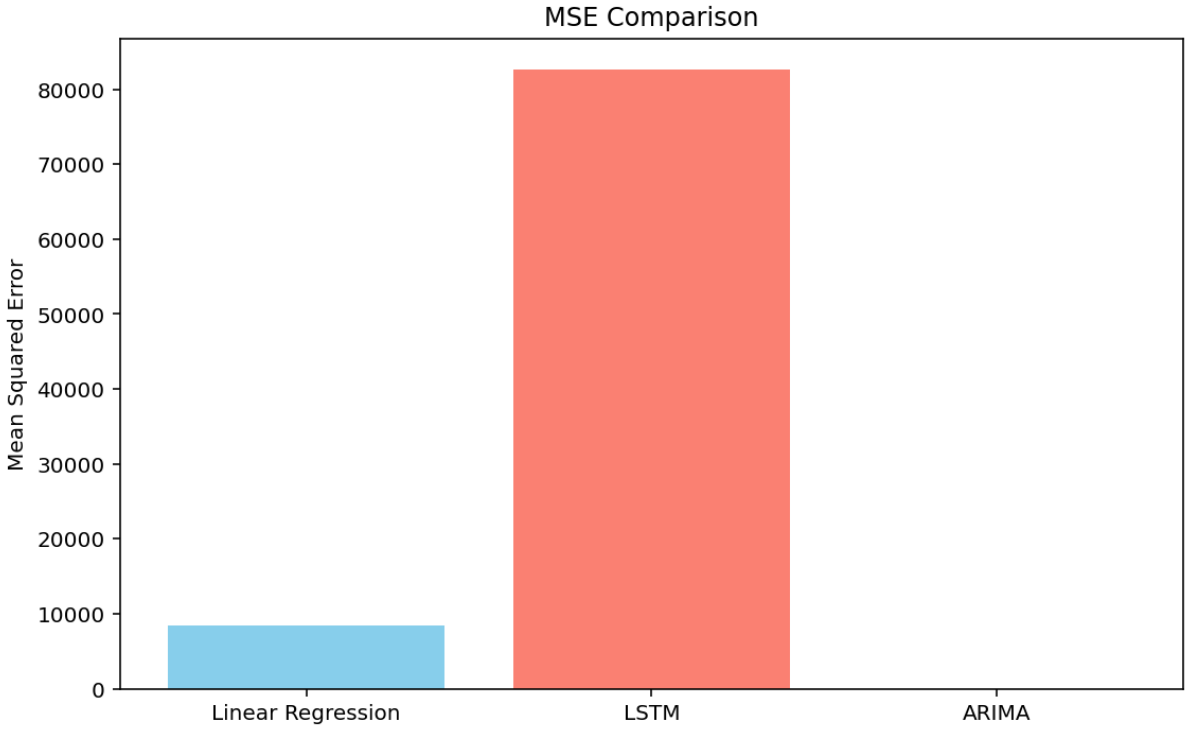


Figure 10 MSE Comparison

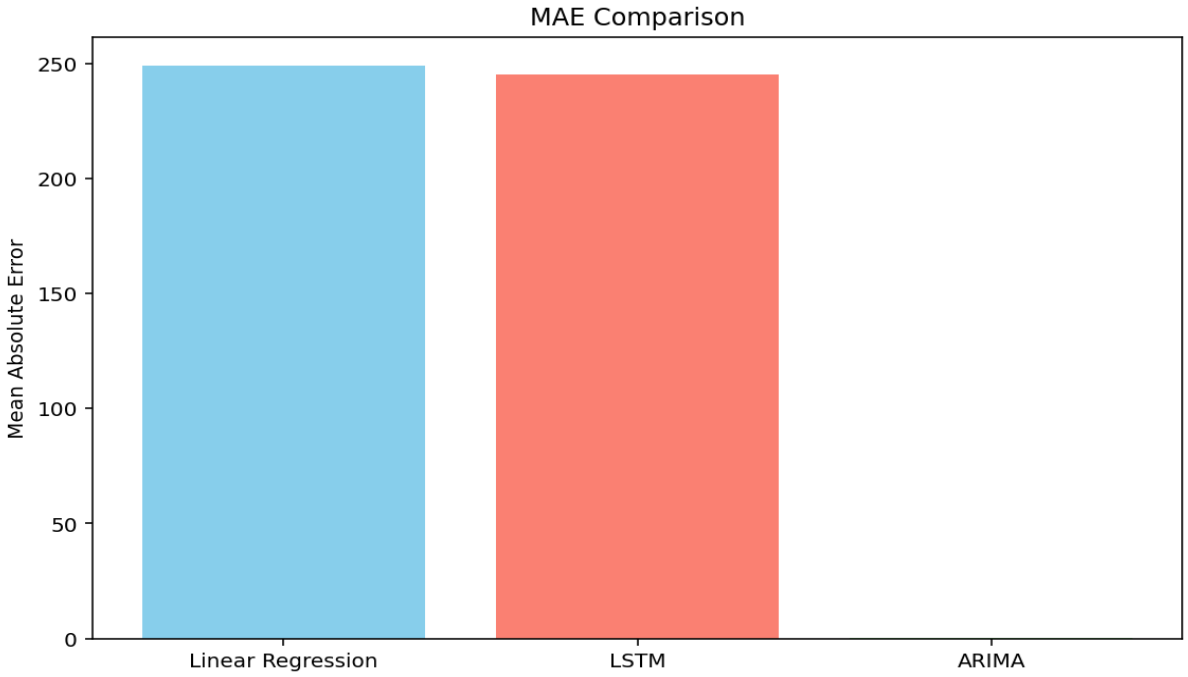


Figure 11 MAE Comparison

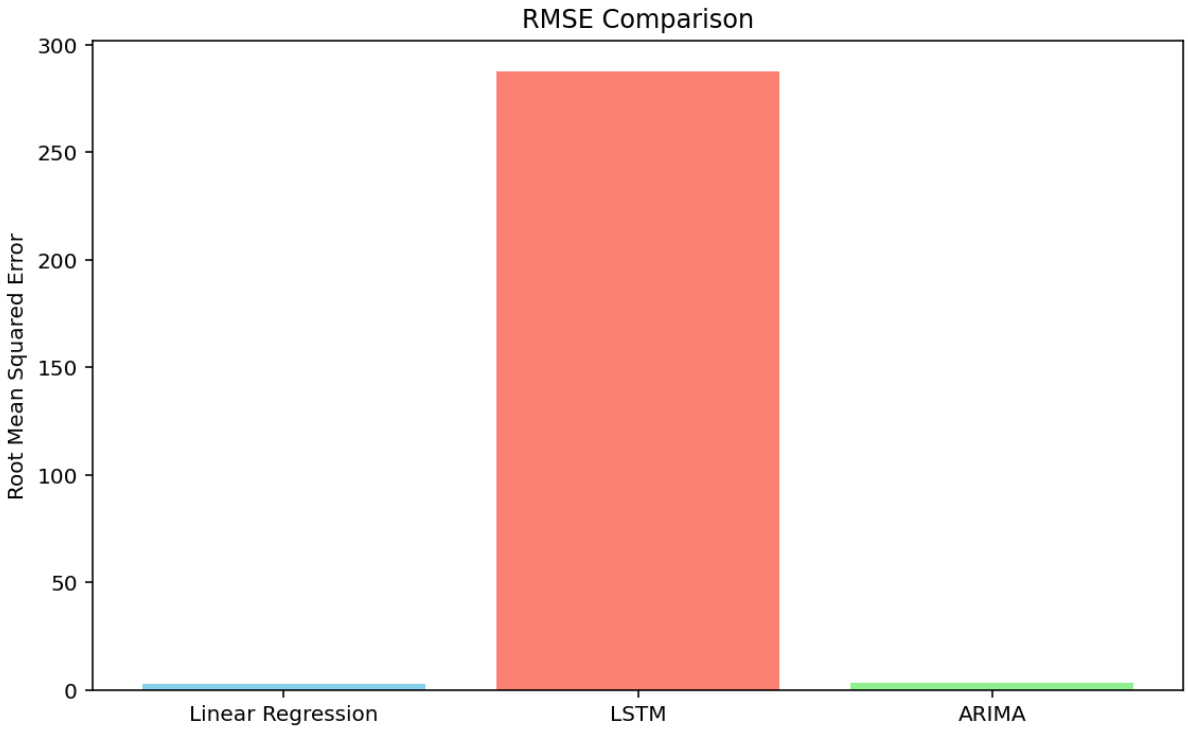


Figure 12 RMSE Comparison

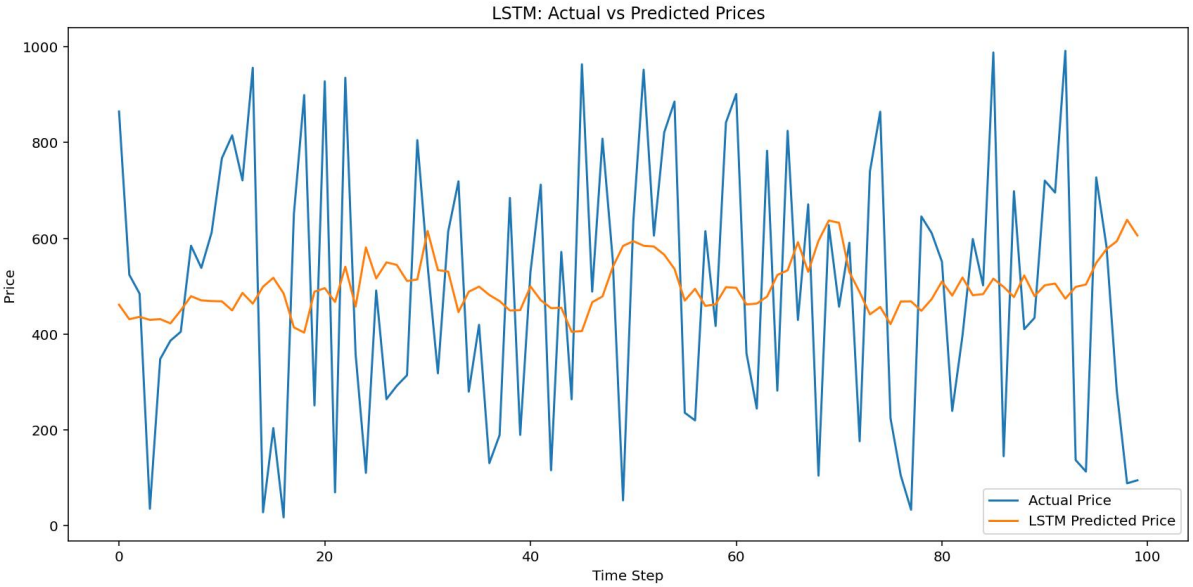


Figure 13 LSTM Product Price Prediction

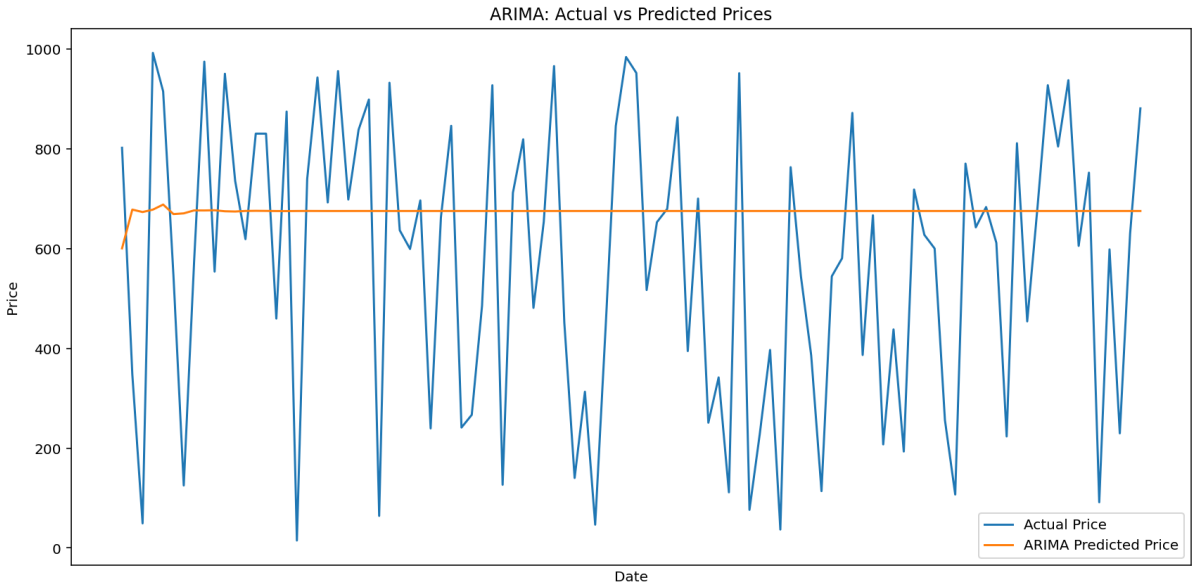


Figure 14 ARIMA Product Price Predictions

5. CONCLUSION

The e-commerce price prediction system developed in this project provides a comprehensive approach to forecasting product prices using advanced machine learning and statistical models. By systematically applying data preprocessing, normalization, and model training techniques, the system ensures high data quality and model performance. The comparative analysis of Linear Regression, ARIMA, and LSTM models offers valuable insights into their respective strengths and limitations. Linear Regression serves as a simple and interpretable baseline for identifying basic trends, while ARIMA excels in capturing linear and seasonal time-based patterns. On the other hand, LSTM demonstrates superior performance in modeling complex, nonlinear dependencies in dynamic sales environments. The use of multiple evaluation metrics, including MAE, MSE, and RMSE, ensures a robust performance assessment. Visualizations such as prediction graphs and error comparison charts enhance the interpretability of the results and support better decision-making. The system is adaptable, scalable, and applicable across various e-commerce platforms, making it a valuable tool for businesses to optimize pricing strategies, inventory management, and promotional planning. Despite challenges like model complexity and data dependency, the proposed framework offers a balanced blend of statistical rigor and deep learning innovation. Future enhancements could include hybrid models, real-time data integration, and incorporation of external features like marketing campaigns or competitor pricing. Ultimately, this project demonstrates that intelligent forecasting systems can significantly enhance business intelligence and operational efficiency in modern digital commerce.

6. FUTURE SCOPE

Future enhancements to the e-commerce price prediction system can focus on increasing model accuracy, adaptability, and real-world applicability. One promising direction is the development of hybrid models that combine the strengths of statistical techniques like ARIMA with deep learning models like LSTM to better capture both linear trends and complex nonlinear patterns. Integrating external data sources—such as Google Trends, competitor pricing, customer reviews, weather data, and promotional calendars—can significantly improve forecasting accuracy by adding richer context to the prediction process. Another area of improvement is implementing real-time data streaming and prediction, enabling the system to adapt dynamically to sudden market changes, flash sales, or seasonal spikes. Enhancing the model interpretability using tools like SHAP or LIME can help non-technical users understand why certain predictions were made. Additionally, applying AutoML and hyperparameter optimization frameworks like Optuna or GridSearchCV can automate the process of finding the best-performing model configurations. Expanding the system to support multivariate time-series forecasting would allow simultaneous analysis of prices, sales volume, customer behavior, and return rates. Improving scalability to handle millions of SKUs across categories using cloud-based platforms like AWS or Azure would enable deployment at an enterprise level. Incorporating forecast uncertainty estimation using Bayesian models or confidence intervals can provide

risk-aware insights for inventory and pricing decisions. Furthermore, embedding the prediction engine within a user-friendly dashboard using tools like Power BI, Streamlit, or Dash can allow business users to interact with the models and visualizations easily. Lastly, including continuous learning mechanisms that retrain models automatically as new data becomes available will keep the system adaptive to changing trends.

CONFLICT OF INTEREST

The authors declare no conflicts of interest regarding the current research.

REFERENCES

1. Banerjee, P., Prasad, V., Thakur, K., Mitra, D., Gaurav, K., & Kanrar, S. Predictive modeling and dynamic analysis of price trends in e-commerce using ML technique. In *IEEE International Conference on Networks, Knowledge, and Convergence (NKCon)*. 2024. 21–22 September 2024. DOI:10.1109/NKCon62728.2024.10774870
2. Hu, Y., Zhou, Z., & Wang, T. Stock price prediction of e-commerce platforms under COVID-19's influence based on machine learning. In *IEEE International Conference on Machine Learning for Intelligent Systems Engineering (MLISE)*. 2022. 05–07 August 2022. DOI:10.1109/MLISE57402.2022.00092
3. Telkar, K., Kandarkar, G., Tola, A., Bagade, J., & Bhimanpallewar, R. Prediction of laptop prices with recommendations based on user specifications. In *IEEE International Conference on Computing, Communication, Control and Automation (ICCUBEA)*. 2023. DOI:10.1109/ICCUBEA58933.2023.10392000
4. Goyal, H., Sukhija, A., Bhatia, K., Guleria, K., & Sharma, S. Demand prediction of consumer intention to buy edible items using machine learning techniques. In *IEEE International Conference on Smart Generation Computing, Communication and Networking (SMARTGENCON)*. 2023. 29–31 December 2023. DOI:10.1109/SMARTGENCON60755.2023.10442766
5. Alrumiah, S.S., & Hadwan, M. Implementing big data analytics in e-commerce: Vendor and customer view. *IEEE Access*, 2021. 9: p.3063615. DOI:10.1109/ACCESS.2021.3063615
6. Almuqren, L., Alruwais, N., Alhashmi, A.A., Alzahrani, I.R., Salih, N., & Assiri, M. WSN-assisted consumer purchasing power prediction via Barracuda Swarm Optimization-driven deep learning for e-commerce systems. *IEEE Transactions on Consumer Electronics*, 2024. 70(2): p.1694–1701. DOI:10.1109/TCE.2024.3371249
7. Zhang, C., Wang, X., Zhao, C., Ren, Y., Zhang, T., & Peng, Z. PromotionLens: Inspecting promotion strategies of online e-commerce via visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 2022. 29(2): p.767–777. DOI:10.1109/TVCG.2022.3209440
8. Dritsas, E., & Trigka, M. Machine learning in e-commerce: Trends, applications, and future challenges. *IEEE Access*, 2025. 13: p.99048–99067. DOI:10.1109/ACCESS.2025.3572865
9. Trần Quang, P.L., Ngo, U., & Vo Nhi, A.D. E-commerce price suggestion algorithm: A machine learning application. *Innovation in Industry Journal*, 2022. DOI:10.46254/IN02.20220609
10. Akram, A., & Arun, M. E-commerce product sales prediction using machine learning. *International Journal of Innovative Research in Computer Science & Technology*, 2025. 6(2): p.2149–2151.
11. Chen, Y., Xie, X., Pei, Z., Yi, W., Wang, C., Zhang, W., & Ji, Z. Development of a time series e-commerce sales prediction method for short-shelf-life products using GRU-LightGBM. *Applied Sciences*, 2024. 14(2):866. https://doi.org/10.3390/app14020866
12. Wang, C., & Wang, J. Research on e-commerce inventory sales forecasting model based on ARIMA and LSTM algorithm. *Mathematics*, 2025. 13(11):1838. https://doi.org/10.3390/math13111838
13. Nowak, M., & Pawłowska-Nowak, M. Dynamic pricing method in the e-commerce industry using machine learning. *Applied Sciences*, 2024. 14(24):11668. https://doi.org/10.3390/app142411668
14. Gkikas, D.C., & Theodoridis, P.K. Predicting online shopping behavior: Using machine learning and Google Analytics to classify user engagement. *Applied Sciences*, 2024. 14(23):11403. https://doi.org/10.3390/app142311403