Forecasting Retail Sales in the Indian Market: Integrating Time-Series and Machine Learning Approaches

Amir Ahmad Dar[1*] , Shouvik Sanyal[2] , Khaliquzzaman Khan[3] , Mohammad Shahfaraz Khan[4] ,
Aseel Smerat[5] , Vanshita Arora[1]

[1] School of Chemical Engineering and Physical Sciences, Lovely Professional University, Punjab 144411, India
[2] Department of Marketing and Entrepreneurship, Dhofar University, Salalah 211, Sultanate of Oman
[3] School of Business, American University in Dubai, Dubai 28282, UAE
[4] College of Economics and Business Administration, University of Technology and Applied Sciences-Salalah, Salalah 211, Oman
[5] Faculty of Educational Sciences, Al-Ahliyya Amman University, Amman 19328, Jordans

Corresponding Author Email: sagaramir200@gmail.com

ABSTRACT

Accurate sales forecasting is essential for large organizations because it supports effective planning and strategic decision-making. Forecasting techniques are effective tools that could be utilized to extract the underlying information in the mass data to ensure that the validity of the forecast can be maximized. It is a study that uses various predictive algorithms to be able to breakthrough in future sales. In addition to a more conservative approach, such as ARIMAX, machine learning techniques were used to forecast sales data and evaluate predictive performance in terms of Random Forest, XGBoost, as well as Support Vector Regression (SVR). The paper proceeds to give a discussion on various prediction methods and assessment criteria after giving a short introduction regarding sales data and forecasting concepts. The ARIMAX model achieved the best forecasting performance among all models. It produced the lowest error values, with MAE = 14.20, RMSE = 16.60, and MAPE = 11.56%. The SVR model showed comparable performance, with a MAPE of 11.66%. Ensemble models such as Random Forest and XGBoost exhibit comparatively higher prediction errors. Statistical validation using the Diebold–Mariano test confirms that ARIMAX provides reliable and competitive forecasts. These findings highlight the continued relevance of interpretable time-series models with exogenous variables for sales forecasting in data-driven retail environments.

1. INTRODUCTION

Accurate sales forecasting plays a crucial role in good supply chain management. Errors in forecasting can lead to overestimating and underestimating the demand, which in turn affects inventory levels, cash flow, company reputation, and overall profitability. Therefore, it has attracted significant attention from both researchers and business practitioners. Most importantly, accurate sales forecasting can rescue any business from going under, a business of any size, and they can become even more essential for small and medium-sized businesses by giving a clear-cut estimation of the increase in revenue and thereby making short-term demand and supply planning very simple [1]. Such projections may prove to be a lighthouse in the financial management voyages which will be instrumental for the business objectives at large. Apart from that, they can also expedite the strategic decision-making process by pinpointing the immediate investment areas through delivering the insights based on the expected sales trends. To better understand the dataset, we conducted exploratory data analysis (EDA) on the given dataset. A time series is a sequence of observations recorded at regular time intervals. The objectives of time series analyses are several such as recognizing the reasons for the observed figures, forecasting the future values based on the historical data, or simply showing the main patterns and features in the series [2]. On the other hand, machine learning (ML) is an AI subfield that focuses on the development of algorithms and models that can identify data patterns and improve their performance without the need for explicit programming. ML can reveal hidden patterns in complex datasets. It can also be used to predict future outcomes and to develop accurate models for tasks such as classification and clustering. By using statistical methods and computer power, ML makes it possible for systems to be self-correcting and even draw data-driven conclusions in various domains [3]. A lot of recent research has shown that the combination of XGBoost and lagged features has excellent results. Correspondingly, great performances of Autoregressive Integrated Moving Average (ARIMA) models and LSTM-based networks in similar tasks of forecasting have been reported [4, 5]. Based on these findings, we selected four forecasting algorithms:

Autoregressive Integrated Moving Average with exogenous variables (ARIMAX), Random Forest, XGBoost, and SVR. We will compare their performances by means of their error metrics. Sales prediction is essentially a time series forecasting problem. It is commonly addressed using autoregressive (AR) and autoregressive moving average (ARMA) models [6-8]. These models perform well for stationary time series. However, most real-world data are non-stationary. As a result, this has caused the emergence of both linear and nonlinear models such as autoregressive conditional heteroscedasticity (ARCH), generalized autoregressive conditional heteroskedasticity (GARCH), seasonal ARIMA, seasonally decomposed ARIMA (STL-ARIMA), and autoregressive integrated moving average (ARIMA) [9-11]. Moreover, the employment of autoregressive models that manage cointegration, for instance, autoregressive distributed lag (ARDL) models, or estimate covariance functions, e.g., stochastic ARMA processes [12, 13], has also been recommended. The above-listed regression methods are general and thus can be applied in any domain. They have already been used with success in various domains such as wind speed prediction [14] and sales forecasting [15] to mention only a few. For example, the open-source additive model developed by Taylor and Letham [16] is a means that analysts can use to cope with different time series forecasting tasks. Even models and codes from other research works have been made accessible to everyone. Recent ML research increasingly focuses on time-series prediction in fields such as finance, healthcare, and transportation. Sometimes this entails that the raw data be transformed into features that are manually created. SVR was used by Yu et al. [17] in their work to forecast the sales of newspapers and magazines; meanwhile, Kazem et al. [18] applied the same method to predict stock prices. Similarly, Gumus and Kiran [19] selected the XGBoost algorithm to both accurately and efficiently forecast the crude oil prices. For instance, when it came to time-series predictions, Babai et al. [20] used autoregressive models while Khandelwal et al. [21] opted to merge ARIMA with artificial neural networks (ANN). Tsai et al. [22] and Bandara et al. [4] both employed long short-term memory (LSTM) networks for their respective predictions PM2.5 concentrations and sales demand. Chen et al. [23] developed the trend alignment dual-attention multi-task recurrent neural network (TADA) that raises the accuracy of sales forecasting by using two attention mechanisms, one for aligning the future trend with the matching past pattern and the other for considering the unknown future factors.

2. LITERATURE REVIEW

Accurate sales forecasting is a central concern in marketing analytics, as it supports demand planning, inventory control, and budget allocation. Traditional forecasting approaches rely heavily on statistical time-series models, particularly the Autoregressive Integrated Moving Average (ARIMA) framework. ARIMA models capture temporal dependence, trend, and autocorrelation in historical sales data and have been widely applied for short- to medium-term forecasting across sectors. Hyndman and Athanasopoulos [24] emphasize that ARIMA performs well when sales series exhibit stable patterns and linear dynamics. Empirical studies further support its effectiveness. Fattah et al. [25] demonstrated strong forecasting accuracy of ARIMA in food demand prediction,

while Dar et al. [26], using the Box–Jenkins methodology, showed reliable performance of ARIMA in volatile financial markets, highlighting its robustness across domains.

Despite its strengths, ARIMA has limitations, particularly when sales data exhibit nonlinear relationships, complex seasonality, or structural changes driven by evolving consumer behavior and marketing activities. To address these limitations, machine learning (ML) methods—especially deep learning models—have gained prominence in sales forecasting research. Long Short-Term Memory (LSTM) networks are designed to capture long-range dependencies and nonlinear temporal patterns. Siami-Namini and Namin [27] reported that LSTM models significantly outperformed ARIMA in financial time-series forecasting by reducing prediction errors. Similarly, recent retail and e-commerce studies indicate that ML-based models better adapt to fluctuating demand patterns influenced by digital marketing and rapidly changing consumer preferences [28].

Comparative forecasting studies increasingly emphasize that no single model is universally superior; rather, performance depends on data characteristics, forecast horizon, and market dynamics. While ARIMA remains computationally efficient and interpretable, ML models offer higher flexibility and predictive power in complex environments. As marketing channels expand and sales dynamics become more data-intensive, researchers advocate benchmarking classical time-series models against ML approaches to identify optimal forecasting strategies. Accordingly, this study focuses on a systematic comparison of ARIMA and machine learning–based forecasting models to evaluate their relative predictive accuracy and practical usefulness in sales forecasting, thereby aligning methodological analysis with the core objective of forecasting performance evaluation rather than causal advertising impact.

3. METHODOLOGY

This study aimed to predict sales across several media channels based on current and anticipated performance by applying a data-driven technique. The systematic approach of the methodology was focused on forecasting, evaluation, model choice and data preprocessing. The machine learning and statistics libraries of the Python programming environment were used to implement the entire pipeline.

The three advertising expenditures (TV, Radio, and Online) were treated as independent variables (features), while sales were modeled as the dependent variable. A Date variable was used to chronologically order the observations and to resample the data at a weekly frequency, ensuring consistency with the time-series forecasting framework. In order to determine the performance of the model in an objective manner, the data was divided into training and testing using a random 80/20 split. The first 80% of observations were used for model training, while the remaining 20% were reserved for out-of-sample testing. For ML models, hyperparameter selection was conducted using time-series cross-validation with an expanding window approach, ensuring that model evaluation respected the time-ordering of the data.

Comparative modeling method was adopted to discover optimal algorithm to predict sales. The four models that were investigated were ARIMAX, Rand Forest, XGBoost, and SVR. These models have been chosen to balance between the traditional time-series algorithms and the most recent machine

learning algorithms.

3.1 ARIMAX

ARIMAX extends the traditional ARIMA framework by incorporating external regressors. In this study, TV, radio, and online advertising expenditures were included as exogenous variables to capture their influence on sales.

The ARIMA order was determined using two diagnostics. First, we examined the autocorrelation (ACF) and partial autocorrelation (PACF) plots of the sales series. Second, we compared Akaike Information Criterion (AIC) values across competing models. Together, these diagnostics guided the final model selection. These diagnostics indicated a dominant first-order autoregressive component, leading to the selection of an ARIMAX(1,0,0) specification. Seasonal components were excluded due to the absence of pronounced seasonal patterns in the weekly data. The model was estimated using the training dataset, and forecasts were generated for the test period using observed advertising expenditures as exogenous inputs.

3.2 Random Forest Regressor

Random Forest is a ML method that uses bootstrap aggregation to combine multiple decision trees. In this study, the Random Forest Regressor was trained using 100 estimators. It provided a strong non-linear model capable of capturing complex relationships between advertising expenditure and sales. It was a good choice for comparison because of its capacity to deal with multi-dimensional predictors without subjecting to rigid assumptions. Model hyperparameters, including the number of trees and maximum tree depth, were selected using grid search with an expanding-window time-series cross-validation strategy to ensure a fair comparison while respecting the temporal structure of the data.

3.3 XGBoost

The gradient-boosting algorithm XGBoost has been tuned for speed and efficiency. Instead of using bagging, as Random Forest does, XGBoost builds trees one after the other, trying to fix the mistakes of its predecessors. Highly accurate models are frequently produced using this method, especially when dealing with tabular data. The method used the squared-error objective function and was set up with 100 boosting iterations to maximize the prediction of continuous sales data. By adding XGBoost, an enhanced ensemble technique was made available that has proven to perform better in a variety of predicting applications. XGBoost hyperparameters were tuned using grid search. The key parameters included learning rate, maximum tree depth, and the number of boosting iterations. An expanding-window time-series cross-validation strategy was applied to ensure that the tuning process respected temporal ordering and improved generalization performance.

3.4 SVR

SVR uses Support Vector Machines (SVM) concepts to address regression problems. In order to find a function that approximates the target variable within a given margin of tolerance, SVR maps features into a high-dimensional space. Because it can model non-linear relationships, the radial basis function (RBF) kernel was used in this study. Both the target

variable and the input features were normalized before training, as SVR is sensitive to variable scaling. For evaluation, predictions were rescaled to the original sales units once the model was fitted to scaled training data. The regularization parameter and kernel width were selected via grid search with expanding-window time-series cross-validation, ensuring robust model performance.

The team was using three well-known error measures to measure and compare the behavior of the different models, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). This set of indicators was used on the test set predictions to develop a comparative evaluation scheme. The table overview of the data makes it possible to identify the most appropriate model in a short period of time. The model with the smallest RMSE was considered the best-performing forecasting model, subject to further statistical validation using the Diebold–Mariano (DM) test.

To make models comparative to each other, the relative values of each model were presented as bar plots to depict their relative MAE values, RMSE values, and MAPE values. These visualizations allowed one to get a clear picture of the pros and cons of the model in relation to different error metrics. Apart from that, monthly aggregated plots were created to compare actual and forecasted sales in the test period. These kinds of visualizations indicated the degree to which each model captured the underlying sales trends, thereby giving the error figures a context.

Once the performance of the best model was established, sales forecasting was incorporated into the workflow. The ARIMAX model was the one chosen for this task as it was the most interpretable and the best fit for time-series forecasting with exogenous variables. The average historical TV, radio, and internet spending were taken as exogenous inputs to predict future advertising expenditures. The final prediction together with the historical sales were plotted not only to show the predicted path for the next year but also the continuation of the previous trends. This visualization supported decision-making by providing stakeholders with insight into expected future sales performance.

It was a systematic approach that allowed the study to achieve a balance between statistical rigor, machine learning accuracy, and practical interpretability. Besides finding the best model to explain sales volatility, the method exemplified how one could predict future results under given scenarios for advertising expenditures.

4. RESULT AND ANALYSIS

The findings of the research are divided into two parts. Firstly, the forecasting performance of four models, i.e., ARIMAX, Random Forest, XGBoost, and SVR, was judged on the test dataset through three error measures (MAE, RMSE, and MAPE). Secondly, the model with the best performance was employed to create a sales projection, and the output was visually examined by comparing the past and predicted patterns.

4.1 Descriptive statistics

The dataset consists of 200 weekly observations from December 2021 to September 2025, covering sales and advertising expenditures across three media channels (TV,

Radio, Online). Table 1 presents basic descriptive statistics. Weekly sales averaged 148.57 units (SD = 53.45), with a minimum of 14.82 and a maximum of 272.07 units. TV spend averaged 150.36, Radio 103.35, and Online 217.85, indicating variability in advertising allocation.

4.2 ARIMAX model justification

By examining the sales series' ACF and PACF, the proper ARIMA order was determined. The plots are shown in Figure 1.

Table 1. Descriptive statistics

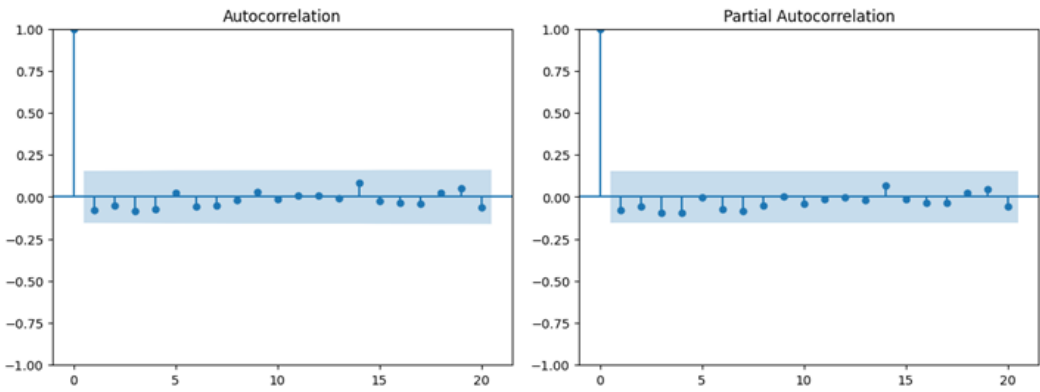| Variables	Total Count	Average	Standard Deviation	Minimum	Quartile 1	Quartile 2	Quartile 4	Maximum
TV Spend	200	150.36	85.52	11.60	76.29	153.40	229.48	296.19
Radio Spend	200	103.35	57.14	5.99	55.98	110.62	149.73	198.15
Online Spend	200	217.86	116.78	24.12	117.10	219.65	328.22	399.89
Sales	200	148.57	53.45	14.82	110.10	147.67	187.05	272.07

Figure 1. The ACF and PACF plots

ACF and PACF plots of the training set were examined (lags = 20), revealing a strong first-order autocorrelation, supporting the selection of AR(1) in the ARIMAX(1,0,0) model. The AIC for this model was 1409.50, indicating an adequate fit relative to alternative specifications. The three advertising spend variables were included as exogenous regressors to capture their influence on sales.

4.3 Comparative model performance

Table 2 illustrates the comparative results of the models on the test dataset. The ARIMAX model recorded the lowest scores across all three error indices with an MAE of 14.20, RMSE of 16.60, and MAPE of 11.56%. It means that, on average, ARIMAX predictions were off from the actual values by about 14 sales units and accounted for the smallest proportionate error when compared to other methods. The SVR model ranked second and thus, slightly underperformed ARIMAX but managed to achieve a competitive performance with a MAPE of 11.65%. The performance of Random Forest and XGBoost as powerful ensemble algorithms led to higher errors; thus, XGBoost had the highest RMSE at 27.73, implying lower generalization to unknown data.

These findings are further illustrated in Figure 2, which displays the bar chart comparison of model errors across the three metrics.

Figure 2 and Table 2 illustrate the continual superior performance of ARIMAX over the other models, especially in RMSE, and demonstrates its capability to reduce large deviations. Although SVR had a marginally higher RMSE than the Random Forest and XGBoost models that showed higher error levels, it was a close second to ARIMAX in terms

of MAE and MAPE.

Table 2. Model comparison on test dataset using MAE, RMSE, and MAPE

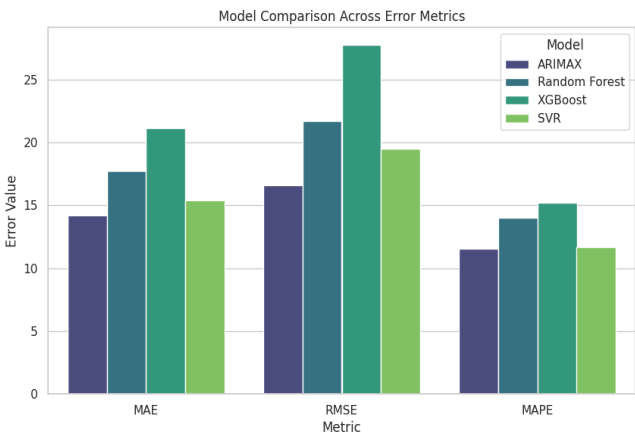| Model	MAE	RMSE	MAPE
ARIMAX	14.20	16.60	11.56%
Random Forest	17.74	21.70	13.99%
XGBoost	21.14	27.73	15.21%
SVR	15.41	19.47	11.66%

Figure 2. Chart comparison of model errors

Besides the numeric measures of error, model forecasts and actual sales for the test period were compared visually as well. Figure 3 shows the monthly sums of the real and predicted values for each of the four models.
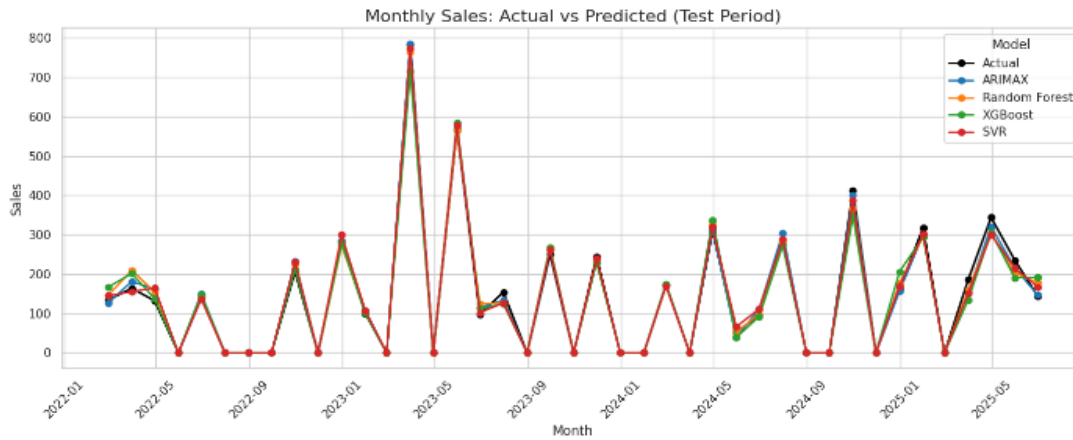
Figure 3. Plots the monthly aggregation of actual and predicted values across all four models

Figure 4. Forecasted monthly sales for next year using ARIMAX

The comparisons reveal that ARIMAX and SVR depict the reality more accurately as they can show the sales peaks and troughs properly. Random Forest and XGBoost were more like trend followers, however, their tendency to overestimate during some spikes and underestimate during dips caused them to have higher error figures.

4.4 Statistical validation of model superiority

To test whether the observed performance differences between ARIMAX and SVR were statistically significant, the Diebold-Mariano (DM) test was applied to the forecast errors. The DM statistic was -0.890 with a p-value of 0.3734, indicating that the difference between ARIMAX and SVR forecasts is not statistically significant at conventional levels. This reinforces the reliability of ARIMAX while acknowledging that SVR remains competitive.

4.5 Forecasting

The ARIMAX model was chosen as the final forecasting model, which is a first-ever decision based on its better performance in both statistical metrics and visual inspection. The model was estimating revenues for the next fifty-two weeks with the average amount spent on advertising as an exogenous input. Monthly totals were obtained by summing these sales. Both the historical sales series and the 12-month forecast are shown in Figure 4.

ARIMAX turned out to be the best and clearest model in terms of accuracy because it could maintain good predictive

performance while also considering advertising expenditures as exogenous inputs. The comparative study, however, emphasizes the significance of matching model selection with the data structure, a fact that is still consistent with their predictions being a dependable basis for operational planning.

5. DISCUSSION

The results of the study make it clear that the correct model selection is one of the major considerations in accurately estimating the sales response to the expenditures on advertising. The comparison of ARIMAX, Random Forest, XGBoost, and SVR demonstrates substantial differences in how traditional time-series models and ML algorithms represent temporal structure and exogenous influences.If sales data have temporal dependencies and are affected by external factors, time-series methods are still viable and interpretable tools as shown by the dominant performance of ARIMIX.

In the context of the Indian retail market, sales data often display unique characteristics such as weekly and festival-driven fluctuations, promotional campaigns aligned with major holidays, and varying consumer demand across regions. For instance, spikes in consumer purchases during Diwali, Holi, or end-of-season sales can create short-term peaks in sales, which time-series models like ARIMAX can capture effectively when exogenous advertising spend is included. Although ML models are effective in capturing complex non-linear relationships, they do not explicitly model temporal dependence and may smooth short-term fluctuations, which

can reduce interpretability.

An excellent illustration of how ARIMIX might be applied in practice to support managerial decision-making is the forecasting problem. If the advertising scenario stays the same, the 12-month estimates give businesses a starting point picture of their future sales. Although the forecast showed smoother and easier-to-understand trends, it was unable to adequately reflect the actual volatility of the sales. Given that firms typically need precise estimates to guide their budget allocations and policy decisions, this capability could be useful in strategic planning.

The study also emphasizes the necessity of taking into account local factors in India, such as the influence of government-driven promotional programs on consumer behavior, diverse media reach, and regional festival calendars. Future forecasting efforts could improve predicted accuracy and provide Indian retailers with more actionable findings by incorporating these aspects.

However, the use of average advertising expenditure as an exogenous input represents a simplification; future work should incorporate dynamic advertising strategies to better reflect real-world decision environments. The reasons raised above indicate that ARIMIX is a suitable and effective tool for generating sales data scenarios in which advertising acts as an exogenous influence. By demonstrating that sophisticated algorithms should not be only viewed as a substitute for conventional statistical models without taking into account data aspects and the decision-making environment, the findings promote a balanced perspective on predictive modeling. Future research can expand on this and develop even more precise and adaptable forecasting frameworks by assessing hybrid models that combine the benefits of ML with time-series analysis.

Although the study provides useful information regarding the anticipated relationship between advertising budget and sales, it is crucial to acknowledge its limitations. First, the dataset's scope was rather constrained because just three advertising channels were included as predictors. The omission of other important variables, such as pricing strategies, competition activity, seasonal events, and economic situations, may have limited the models' capacity to explain. Additionally, the forecasting stage's supposition of a fixed average advertising expenditure is simplistic and ignores how dynamic real marketing strategies are. The selection of parameters, particularly for ML models, is another limitation; while suitable defaults were used, careful tuning may have improved their performance. Finally, the reliance on a single train–test split, rather than more rigorous time-series cross-validation procedures, may limit the robustness of the performance evaluation.

Future study could incorporate a broader range of predictors, such as macroeconomic data, promotional efforts, and rival activity, in order to overcome these constraints and successfully capture the complex elements driving sales performance. Using more complex or hybrid models, like combining ensemble approaches with ARIMAX or experimenting with deep learning techniques like LSTM or Prophet, could further increase projected accuracy while preserving interpretability. If alternate scenarios for advertising expenditure were examined rather than assuming constant averages, managers would also have more valuable insights into the potential outcomes of other budget allocation approaches. To further investigate model stability over time, rolling or extending window validation techniques may be useful in future research. By expanding the range of predictors and the analytical toolkit, future research can build on the results of this study and produce more accurate and useful sales forecasting solutions.

6. CONCLUSIONS

The outcome of this research can become a transformational game-changer among the management and law makers with regard to decision making. When appropriately combined with the analysis of advertising spending, correct sales forecasts can be converted to powerful budgeting, inventory, and resource allocation orders. The adoption of these instruments can therefore serve as an engine in the execution of evidence-based marketing strategies and supply chain efficiency in the business that relies on the Indian market to produce consumer demand by advertisement. The findings are particularly relevant for the Indian retail sector, characterized by high sales variability around festivals, regional promotional campaigns, and diverse consumer behaviors. Accurate sales projections in this context can help Indian retailers optimize inventory, allocate marketing resources effectively, and plan promotional campaigns aligned with local demand patterns.

In practice, these insights can help Indian politicians and shop managers better allocate resources, increase the effectiveness of the supply chain, and apply evidence-based marketing techniques. Additionally, sector-specific laws that support sustainable growth in the Indian retail market and promote efficient media spending can be informed by predictive modeling. This study connects theoretical forecasting models with real-world applications by specifically taking into account the Indian market's dynamics, such as seasonal demand, cultural events, and heterogeneous advertising effects. This allows Indian businesses to operate in competitive and quickly changing retail environments with practical advice.

The need behind this is significant since it will be combining the time-series approaches with the sophisticated machine learning approaches to understand the effect of advertisement on sale prediction in the Indian retail market, which is a field that has not been studied extensively. With a combination of explanatory and predictive views, the study does not only contribute to the theoretical perspectives of the forecasting literature but also provides useful solutions in strategic decisions within the retail industry.

The research exercise shifts forecasting into the Indian retail business, the Indian retail market that is a developing economy, and hence offers research-based information that the Indian business can use to improve sales forecasting, advertising techniques, and competitive edge in a dynamic marketplace.

TRANSPARENCY STATEMENT

The authors affirm that this manuscript provides an accurate, transparent, and honest account of the study being reported.

DATA AVAILABILITY STATEMENT

The data supporting this study are available upon reasonable request from the corresponding author.

AUTHORS' CONTRIBUTIONS

Amir Ahmad Dar: Conceptualized the study design, coordinated the research, and supervised the overall project. Shouvik Sanyal: Performed the ARIMA and LSTM modeling, prepared figures and tables. Khaliquzzaman Khan and Fayaz Ahamed: Drafted the introduction and literature review, integrated references. Mohammad Shahfaraz Khan and Aseel Smerat: Contributed to the discussion, conclusion, and manuscript revisions. Vanshita Arora: Revision. All authors: Contributed to writing, editing, and approved the final manuscript.

REFERENCES

[1] Cheriyan, S., Ibrahim, S., Mohanan, S., Treesa, S. (2018). Intelligent sales prediction using machine learning techniques. In 2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE), Southend, UK, pp. 53-58. https://doi.org/10.1109/iCCECOME.2018.8659115

[2] Box, G. (2013). Box and Jenkins: Time series analysis, forecasting and control. In A Very British Affair. Palgrave Advanced Texts in Econometrics, pp. 161-215. https://doi.org/10.1057/9781137291264_6

[3] Tyagi, A.K., Chahal, P. (2022). Artificial intelligence and machine learning algorithms. In Research Anthology on Machine Learning Techniques, Methods, and Applications. https://doi.org/10.4018/978-1-6684-6291-1.ch024

[4] Bandara, K., Shi, P., Bergmeir, C., Hewamalage, H., Tran, Q., Seaman, B. (2019). Sales demand forecast in e-commerce using a long short-term memory neural network methodology. In Neural Information Processing: 26th International Conference, ICONIP 2019, Sydney, NSW, Australia, pp. 462-474. https://doi.org/10.1007/978-3-030-36718-3_39

[5] Kihoro, J., Otieno, R., Wafula, C. (2004). Seasonal time series forecasting: A comparative study of ARIMA and ANN models. African Journal of Science and Technology, 5(2): 41-49. https://doi.org/10.4314/AJST.V5I2.15330

[6] Friedlander, B. (1982). A recursive maximum likelihood algorithm for ARMA spectral estimation. IEEE Transactions on Information Theory, 28(4): 639-646. https://doi.org/10.1109/TIT.1982.1056531

[7] Cadzow, J. (1980). High performance spectral estimation--A new ARMA method. IEEE Transactions on Acoustics, Speech, and Signal Processing, 28(5): 524-529. https://doi.org/10.1109/TASSP.1980.1163440

[8] Yule, G.U. (1927). VII. On a method of investigating periodicities disturbed series, with special reference to Wolfer's sunspot numbers. Philosophical Transactions of the Royal Society A. Mathematical, Physical and Engineering Sciences, 226(636-646): 267-298. https://doi.org/10.1098/rsta.1927.0007

[9] Box, G.E.P., Pierce, D.A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. Journal of the American Statistical Association, 65(332): 1509-1526. https://doi.org/10.2307/2284333

[10] Buhl, J., Liedtke, C., Schuster, S., Bienge, K. (2020). Predicting the material footprint in Germany between 2015 and 2020 via seasonally decomposed autoregressive and exponential smoothing algorithms. Resources, 9(11): 125. https://doi.org/10.3390/resources9110125

[11] Shah, I., Iftikhar, H., Ali, S. (2020). Modeling and forecasting medium-term electricity consumption using component estimation technique. Forecasting, 2(2): 163-179. https://doi.org/10.3390/forecast2020009

[12] Schubert, T., Korte, J., Brockmann, J.M., Schuh, W.D. (2020). A generic approach to covariance function estimation using ARMA-models. Mathematics, 8(4): 591. https://doi.org/10.3390/math8040591

[13] Busu, M. (2020). Analyzing the impact of the renewable energy sources on economic growth at the EU level using an ARDL model. Mathematics, 8(8): 1367. https://doi.org/10.3390/math8081367

[14] Erdem, E., Shi, J. (2011). ARMA based approaches for forecasting the tuple of wind speed and direction. Applied Energy, 88(4): 1405-1414. https://doi.org/10.1016/j.apenergy.2010.10.031

[15] Kechyn, G., Yu, L., Zang, Y.G., Kechyn, S. (2018). Sales forecasting using WaveNet within the framework of the Kaggle competition. arXiv preprint arXiv:1803.04037. https://doi.org/10.48550/arXiv.1803.04037

[16] Taylor, S.J., Letham, B. (2017). Forecasting at scale. PeerJ Preprints, 5: e3190v2. https://doi.org/10.7287/peerj.preprints.3190v2

[17] Yu, X.D., Qi, Z.Q., Zhao, Y.M. (2013). Support vector regression for newspaper/magazine sales forecasting. Procedia Computer Science, 17: 1055-1062. https://doi.org/10.1016/j.procs.2013.05.134

[18] Kazem, A., Sharifi, E., Hussain, F.K., Saberi, M., Hussain, O.K. (2013). Support vector regression with chaos-based firefly algorithm for stock market price forecasting. Applied Soft Computing, 13(2): 947-958. https://doi.org/10.1016/j.asoc.2012.09.024

[19] Gumus, M., Kiran, M.S. (2017). Crude oil price forecasting using XGBoost. In 2017 International Conference on Computer Science and Engineering (UBMK), Antalya, Turkey, pp. 1100-1103. https://doi.org/10.1109/UBMK.2017.8093500

[20] Babai, M.Z., Ali, M.M., Boylan, J.E., Syntetos, A.A. (2013). Forecasting and inventory performance in a two-stage supply chain with ARIMA(0,1,1) demand: Theory and empirical analysis. International Journal of Production Economics, 143(2): 463-471. https://doi.org/10.1016/j.ijpe.2011.09.004

[21] Khandelwal, I., Adhikari, R., Verma, G. (2015). Time series forecasting using hybrid ARIMA and ANN models based on DWT decomposition. Procedia Computer Science, 48: 173-179. https://doi.org/10.1016/j.procs.2015.04.167

[22] Tsai, Y.T., Zeng, Y.R., Chang, Y.S. (2018). Air pollution forecasting using RNN with LSTM. In 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), Athens, Greece, pp. 1074-1079. https://doi.org/10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00178

[23] Chen, T., Yin, H.Z., Chen, H.X., Wu, L., Wang, H., Zhou, X.F. (2018). TADA: Trend alignment with dual-

attention multi-task recurrent neural networks for sales prediction. In 2018 IEEE International Conference on Data Mining (ICDM), Singapore, pp. 49-58. https://doi.org/10.1109/ICDM.2018.00020

[24] Hyndman, R.J., Athanasopoulos, G. (2021). Forecasting: Principles and Practice. OTexts. Melbourne, Australia. https://otexts.com/fpp3/.

[25] Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., Lachhab, A. (2018). Forecasting of demand using ARIMA model. International Journal of Engineering Business Management, 10. https://doi.org/10.1177/1847979018808673

[26] Dar, A.A., Jain, A., Malhotra, M., Farooqi, A.R., Albalawi, O., Khan, M.S., Hiba. (2024). Time Series analysis with ARIMA for historical stock data and future projections. Soft Computing, 28: 12531-12542. https://doi.org/10.1007/s00500-024-10309-w

[27] Siami-Namini, S., Namin, A.S. (2018). Forecasting economics and financial time series: ARIMA vs. LSTM. arXiv preprint arXiv:1803.06386. https://doi.org/10.48550/arXiv.1803.06386

[28] Nithya, V., Rajesh, S., Nithyasri, S. (2025). Sales data forecasting using Arima model. In Proceedings of 5th International Conference on Artificial Intelligence and Smart Energy. ICAIS 2025. Information Systems Engineering and Management, pp. 128-138. https://doi.org/10.1007/978-3-031-90482-0_10