Journal of Current Research in Engineering and Science
JCRES
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 7-Issue 1, January 2024
Paper: 62

Applying Optimized LSTM Deep Learning for the Prediction of Retail Store Sales

Dr. N.Valliammal Assistant Professor (SG),
Rathna A, PG Student, Master of Computer Applications,
Department of Computer Science
Avinashilingam Institute for Home Science and Higher Education for Women
Coimbatore, India.
vallinrayancbe@gmail.com
rathu2001dhanam@gmail.com

***ABSTRACT*— Store sales are essential to the retail industry. Administrators thoroughly evaluate the most dependable models to aid in predicting future sales. Predicting store sales can assist in projecting the future variations or increases in sales to make decision based on past and present data. In recent days, Machine Learning (ML) models like Autoregressive Integrated Moving Average (ARIMA), Random Forest (RF) and Long Short-Term Memory (LSTM) were utilized for predicting retail store sales from training on previously transitioned data. But, ARIMA model experienced poorer performance for long term forecasts. RF model is only suitable for limited datasets. This paper develops an optimised LSTM based on the Enhanced Mine Blast Algorithm (EMBA) to choose the best hyper-parameters of LSTM like number of neurons and hidden layer, number of units in a dense layer, number of dropout level, weight setup, decay percentage, learning rates, activating function, momentum, number of the epoch and batch size in order to improve the structure of the model for the accurate prediction of retail store sales. Mean Squared Error (MSE) is created as the measure of fitness in the EMBA-LSTM model, and EMBA is utilised to globally optimise the weights between neuron nodes of the model. An important principle of EMBA is derived from the mine bombs explosion in real-time applications. The initial population of shrapnel fragments represents the initial hyper-parameter and the computation of their subsequent locations represents the search for the best hyper-parameter. For the model evaluation, P- Values, (RMSE) and (MAPE) are employed to compare the performance of these models like ARIMA, LSTM and EMBA-LSTM. Lower prediction values for RMSE and MAPE and higher values of P-values indicate the efficient accuracy performance in selecting the retail store sales. Finally, the test results show that the EMBA-LSTM on collected sales forecasting dataset achieves a p-value of 0.051%, RMSE value of 0.092% and MAPE value of 0.085 compared to the other existing models.**

Keywords: Enhanced Mine Blast Algorithm, Long Short-Term Memory, Consumer experience, Mean Squared Error, Shrapnel fragments.

I. INTRODUCTION

Every successful retail firm in today's competitive marketplace depends on precise sales forecasting. By avoiding overproduction and reducing overstock, it can help in inventory management. In order to prevent running out of sale products during a specific season, every shop wants to know what their consumers desire in advance [1, 2]. A sales forecasting approach is used to do this, estimating future sales based on previous sales data. For retail chain organisations, being able to predict item sales is essential. Sales projections are also essential inputs for many administrative choices, including pricing, shop allocation, listing/delisting, purchasing, and inventory management.

These standards are able to be established by analysing the overall retail shop sales pattern or the sales of a particular item. It is essential to remember that each product has a different level of predicting complexity. Since demand for some items, like milk, is consistent year-round, it is simple to predict their sales [5, 6]. Forecasting is made more difficult by the seasonality and trends in the sales patterns of some products, such as clothing and furniture. in earlier studies. Previous studies on the prediction of sales have only ever used one prediction model. Consequently, it is impossible for a single model prediction to be the most effective for all product categories [7]. For offline companies in particular, sales forecasting is essential.

With ML model, retailers may well know themselves and their intrants, alter their sales strategies, and continue unbeatable if they can uncover the secret of enormous data sets. [9, 10]. Some ML-predictive models like ARIMA, RF and LSTM are utilized to predict future trends and sales demand. Using time series data, the statistical analysis model ARIMA can forecast future patterns or help one comprehend the current data set. Statistical models called autoregressive forecast future values using data from the past [11]. The datasets that can be transformed to time series that are stationary are the ones that the ARIMA algorithm performs best. Time invariant statistics of stationary

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 7-Issue 1, January 2024
Paper: 62

time series include autocorrelations and other statistical features. Signal and noise are frequently present in datasets with time that is stationary series. There might be a seasonal component or a sinusoidal oscillation pattern in the signal. As a filter, ARIMA separates the signal from the background noise before extrapolating the signal into the future to create forecasts. But for long-term projections, this model performed worse.

The decision tree structure is created by creating a branch for each decision and a leaf for the result, and then repeating this process to create a tree [12]. The attribute selection measure is used by the algorithm to select each attribute. Because of its high accuracy in forecasting via decision trees and its ability to handle large data sets, the Random Forest process is frequently used in prediction and aggression. LSTM networks are categorized as Recurrent Neural Networks (RNN) and are utilized for time series analysis prediction due to its ability to retain and preserve information [13]. A LSTM network functions similarly to a computer's memory in that it stores and accumulates information using cells. LSTM can be used to predict the sales of a single product with limited prior data by analyzing the sales of like products. LSTM models are frequently employed in retail sales forecasting due to their high accuracy. RF and LSTM both are functional to the forecast of time series which is a mainly interesting topic payable to the presence of long-term trends, seasonal and cyclical oscillations, and random noise [14]. While LSTM is a relatively new model, there are no guideline for configuration requirements for a time series dataset.

On considering this, an optimizing model called EMBA-LSTM is proposed in this paper to solve the above-mentioned issue in LSTM. This model optimizes the LSTM structure using EMBA by selecting the best hyper-parameters of LSTM such as the number of nodes and hidden layer, number of units in a dense layer, number of dropout layer, weight initialization, decay rate, learning rate, activation function, momentum, number of epochs, and batch size. The MSE is employed in the EMBA-LSTM model as the fitness function, and the EMBA is used to globally optimize the weights between the LSTM model's neuron nodes to increase generalizability. A principle of EMBA is derived from the burst of mine bombs in real-time applications. The initial population of shrapnel fragments is the initial hyper-parameter, and the computation of their successive placements is the search for the optimal hyper-parameter. In order evaluate these models like ARIMA, LSTM, and EMBA-LSTM, P-Values, RMSE and MAPE are utilized. The efficient accuracy performance is signified by lower prediction values for RMSE and MAPE and higher values of P-values in predicting the retail shop sales.

The outstanding article is prepared as follows: Section II presents the studies related to retail sales prediction using ML algorithms. Section III discusses the proposed algorithm, and Section IV shows its performance compared to the existing algorithms. Section V concludes this study and suggests future enhancements.

II. LITERATURE SURVEY

Pavlyshenko [15] constructed a Machine learning model are to predicting sales time series. The result of ML generalisation in this model was to identify patterns throughout the whole dataset. When there were few previous data points available for a certain sales time series, as may be the case if a new product or shop was launched, this effect might be applied to create sales predictions. Multiple model predictions were made on the validation set using the stacking method, which was then used as an input regressor for models at higher levels. In order to uncover trends in the time series for sales prediction, the Lasso regression method was applied. However, the result was a slower rate of convergence.

Ji et al. [16] presented a XGBoost (CXGBoost) and ARIMA (AXGBoost) models based on clustering are used to anticipate sales for online international companies. The forecasts made by this CXGBoost model are going to take into account the characteristics of goods sales. For various clusters, the relevant C-XGBoost models were created using the XGBoost. Then, using the ARIMA model for the linear component and the XGBoost model for the nonlinear half, the A-XGBoost model was utilised to anticipate the tendency. By giving weight to the C-XGBoost and A-XGBoost outcome predictions, the final results are determined. However, this model performs poorly with sparse and unstructured data.

Verstraete et al. [17] created a methodology based on leading macroeconomic variables to produce tactical sales projections automatically. A framework for predicting sales was developed to automatically choose the necessary factors. The seasonal aspect was then forecast using the seasonal naïve technique and the long-term pattern was predicted utilising LASSO regression with macroeconomic data, while attempting to minimise the size of the indication set. The fact that this model was evaluated solely using monthly macroeconomic data, the same level of aggregation as the dependant variable, somewhat decreases its accuracy of long-term predictions.

Huber &Stuckenschmidt [19] created a daily calculate for retail demand using machine learning, focusing on calendar special days. The situation for

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 7-Issue 1, January 2024
Paper: 62

large-scale demand forecasting in the retail industry that is presented by this model calls for daily projections at the shop level. Data was first acquired using aggregated information based on organisational (store level vs. business level) and chronological (on a weekly data vs. everyday data) hierarchy. Finally, the daily retail data was forecast by ML models, with a focus on special days. However, because of coarser data and a changing data mixture (new goods), the error in forecasting was larger.

Dong et al. [24] created a hybrid human-machine sales forecasting method for e-commerce businesses. Using an advanced ML model and a feature selection model, this model develops a Short-term load forecasting (STLF) system for the prediction of sales a few hours to a week in advance. Multiple activation functions, including sigmoid, rectifier linear unit (ReLU) variations, hyperbolic tangent (tanh), and others, were used to create an intelligent forecasting model. The ability of the smart load forecasting algorithms was further assessed utilising a variety of parametric and variable information. The operating expenses associated with this type are higher, too.

III. PROPOSED METHODOLOGY

In this section, the proposed EMBA-LSTM model is illustrated briefly. The figure 1 depict the block diagram of proposed work.

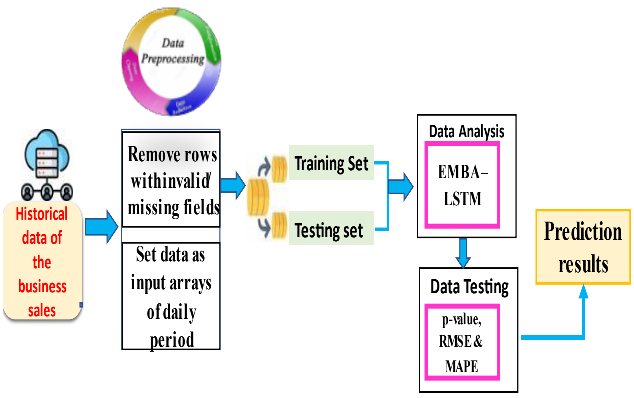

Figure 1 Block diagram of proposed work

A. Data Collection

For sales forecasting, historical sales data for 45 retailers located in different places are provided. Each store possesses a number of categories. During the year, the business additionally conducts many promotional discount events. These sales usually take place before big holidays, the most important of which being the Super Bowl, Workers Day, the day of Thanksgiving, and Christmas. Weeks featuring these major holidays feature five times more highly in the assessment than non-holiday weeks. The retail sales data utilized for prediction will be acquired from the Kaggle database [26] and modified to suit the characteristics.

B. Data Pre-processing

In data pre-processing, cleaning the data is the initial process which removes noisy data, missing value, outlier and inconsistent data such as duplicate data like seller name, product added date, tags. Following that, normalization is proceeded to minimize the modifications in the data. The standardization process is important for time-series data like retail sale data, to reduce the model's dependency on very significant data variations. After that, the normalized data will be processed for feature engineering. Feature engineering has a crucial task while preparing time series data for the machine learning process. The final stage of pre-processing is to divide the dataset into two parts as training and testing, so that the training set does not underfit or overfit the model. In this study, 80% of the data set as the training set and the remaining 20% as the testing set.

C. LSTM model

The LSTM Model has its basis on the Recurrent Neural Network (RNN). RNN is a machine learning technique that employs repeating interconnections in every concealed layer units that link together at various points. LSTM was created to overcome the vanishing a gradient issue in RNN. Unlike RNN, LSTM may store long-term temporal dependency as well as the map across both input and output data. The LSTM architecture consists of a gate that forgets, an input gate, a cell state, and an output gate. The remember gate is the sigmoid layer which decides whether data about the cell state should be stored or replaced. The forget gate is determined by Eq. (1).

$$f_t = \sigma(w_f * [h_{t-1}, x_t] + b_f) \qquad (1)$$

In Eq. (1) f_t is the sigmoid layer that is considered from the product of mass matrixes forget gatew_f and h_{t-1}which is the output value at a time point h_{t-1} and input value $x_t$which is then added to bias matrixes forget gate b_f. For each number in the cell state, the forget gate produces a number between 0 and 1.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \qquad (2)$$

$$\tilde{C}_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \qquad (3)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \qquad (4)$$

In Eq. (2), i_t is the input gate i_t which chooses which new information should be saved in the cell state.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 7-Issue 1, January 2024
Paper: 62

The candidate cell state that will be stored at time $t$is represented by Eq. (3). Eq. (4) is a formula to calculate the cell state at a given point in time t, by considering whether it is necessary to forget the previous cell state $C_{t-1}$and what to consider in the cell state at the current time $\tilde{C}_t$. The value selected by the sigmoid gate as illustrated in Eq. (5) is the output value.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * \tanh C_t \quad (6)$$

In Eq. (6), $tanh$ is used to scale values in the range -1 to 1, Wi is the weight of the input gate, $b_i$is the bias value of the input gate, $\tilde{C}_t$is the new value added to the cell state, C_t is the cell state, $b_o$is the bias value in the cell state and o_t is the output gate. The output generated from the model is the output that is selected based on the state of the cell state model.

D. Hyperparameter Optimization of LSTM using EMBA Process involved in EMBA

One of the most recent metaheuristic algorithms, MBA, has been enhanced to become EMBA. In comparison to other metaheuristic optimisation techniques, this algorithm has generated good results. The landmine detonation mechanism served as the model for this MBA method. According to the MBA, whenever a mine explodes, the fragments of metal that is created spreads and collides with other landmines to destroy them. To find the most effective method, it is crucial to identify the initial blasting point when enough impacts are generated to thoroughly sweep the region of all mines. As a result of this, when an explosion occurs for the first time at an initial location (X_0), shrapnel fragments are generated and spread across the (search) area.

The possibility of the initial shot points observes the key space to be explored and this initial shot point will be selected arbitrarily. This EMBA utilizes the lower bound values of a assumed problem, as if the initial round point's position is not relevant, and to perform a new first shot. EMBA utilizes a adjusting subjectively accomplished value using Eq. (7),

$$X_0^{current} = \ell b + r * (ub - \ell b) \quad (7)$$

Suppose, $X_0^{current}$,ℓb, ub and r are the point of the initial explosion point, lower bound, upper bound search space of LSTM hyper-parameters and a uniform random number between 0 and 1. This means them minimum and maximum range of values for hyper parameter set listed in table 1. The present explosion location$X$ of a mine bomb with n^{th}iteration is given as

$$X_n = \{X_T\}, T = 1, 2, 3, \ldots, n_d \quad (8)$$

Where,n_d is the number of parameters is to be optimized in this study. It is considered that n_s shrapnel pieces are constructed causing another mine explosion at location X_{n+1}

$$X_{n+1} = X_{z(n+1)} + \exp\left(-\sqrt{\frac{T_{n+1}}{d_{n+1}}}\right) * X_n \quad (9)$$

$n = 0,1,2, \ldots$d

In above Eq. (10), $X_{z(n+1)}$ is the explosion mine bomb place, T_{n+1} and d_{n+1} are the direction and remoteness of the accomplished thrown shrapnel pieces in each iteration. The exploding landmine location is determined in Eq. (10) as flows,

$$X_{z(n+1)} = d_{n+1} \times r \times \cos(\theta) \quad (10)$$

Where, r is the random integer between [0, 1] andθ is the shrapnel angle. This angle is equal to $360/_{n_s}$.$n_s$is the number of shrapnel pieces.

The distance and the directions of shrapnel pieces are computed by Eqs. (11) & (12), the location of the detonated landmine is computed by utilizing the Eq. (10).

$$d_{n+1} = \sqrt{(X_{n+1} - X_n)^2 + (F_{n+1} - F_n)^2} \quad (11)$$

$$T_{n+1} = \frac{F_{n+1} - F_n}{X_{n+1} - X_n} \quad (12)$$

$F$is the value of the fitness purpose at the site of X. In the exploitation phase, the EMBA uses a fitness value of the current solution which represents the classification error for the parameters selected at location X. There are two approaches are used to improve the exploration and exploitation of EMBA.

Exploration process

The user specifies an initial the distance, that is utilised to determine the best answer between the range that is calculated using the distance's result and a generated at random number. Additionally, investigation is included in order to take on design exploration of space at closer and farther distances. This factor, which the algorithm utilises in its early iterations, gets compared to a repetition number., and if it is higher than the iteration number the exploration of the solution space is given as:

$$d_{n+1} = X_n * |r^2| \quad (13)$$

$$X_{z(n+1)} = d_{n+1} * \cos(\theta) \quad (14)$$

Exploitation process

This process depends upon the exploration factor (γ), if it is smaller than the iteration number (b). In this instance, the distance explored progressed by the remains in the explosion is minimalized by decreasing the value of the constant ε which is the reduction factor. This constant is strongminded by the candidates. This elimination in distance is calculated by Eq. (15) as follows:

$$X_n = \frac{\gamma_{n-1}}{\exp\left(\frac{b}{\varepsilon}\right)} \quad (15)$$

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 7-Issue 1, January 2024
Paper: 62

Convergence criteria

In most metaheuristic algorithms, the best outcome is determined when the finish condition might be expected as the determined number of iterations, search distance or slight value ε defined as an acceptable tolerance between the last two results. This is the main concept and assume best solution so far X_{best}. The enhancement is considered is as follows for differentiating EMBA from MBA. The updating of explosion positions is proceeded until the below convergence criteria is satisfied. The alteration realized by having the distance between present exploded point $X_{z(n+1)}$

$$X_{n+1} = X_{z(n+1)} + exp(-\sqrt{\tfrac{1}{ED}}) \times r \otimes \{X_{best} - X_{z(n+1)}\} j = 1,2, \ldots n \qquad (16)$$

$$ED = \left[\sum_{i=1}^{T} (X_{best} - X_{z(n+1)})^2\right]^{1/2} \qquad (17)$$

Where ED represents Euclidean distances between position of optimal solution X_{best} and present point of explosion $X_{z(n+1)}$in T dimensions. The presented method did not use data of previous best location; therefore, it advances the speed the convergence of the algorithm.

The Figure 2 depicts the hyper-parameter optimization of LSTM by EMBA.

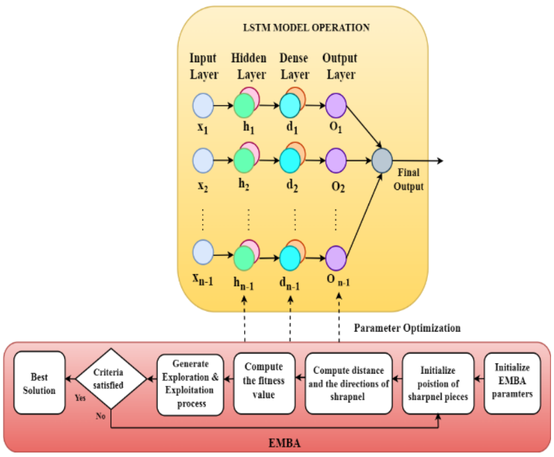

Figure 2 Hyperparameter optimization of LSTM by EMBA

The EMBA-LSTM model utilizes the MSE value as a suitable value function which is computed in Eq. (18), obtained by the initial shrapnel fragments as an optimal solution during iteration process.

$$MSE = \frac{1}{t_N} \sum_{j=1}^{t_N} (k_{ij} - d_{ij}) \qquad (18)$$

In Eq. (18), N is the number of populations, t_N is the number of test set, $k_{ij}$is the predictionresult of LSTM model corresponding to mine bomb's explosion i, $a_{ij}$is the actual value.

E. Classification by EMBA-LSTM

To enhance the model's capacity for generalisation and forecasting, the EMBA is utilised in this instance to globally adjust the weights between the neuron nodes of the LSTM model. This EMBA-LSTM model utilises the EMBA method to transfer the weight values between the LSTM model nodes, translating each weight value to a particular direction of shrapnel pieces and using the shrapnel pieces as a solution set of candidate weights for the LSTM neural network. The EMBA is then used to create an optimisation search area and repeatedly iterate to improve the weights between the neurons, hence improving the LSTM model's effectiveness and precision in making predictions.

IV. PERFORMANCE EVALUATION

This section presents the efficiency of the EMBA-LSTM model compared with the other classical models like ARIMA [11], RF [12], LSTM [13], LGBM [18], STLF [24] and TLSTM [25] on the considered historical sales data (discussed in Section 3.1). The performance metrics used to evaluate the proposed and existing algorithms are described below:

A. P-Values

A measurement of statistics known as a p-value can be used to verify a hypothesis in light of the facts actually observed. When the null hypothesis is considered to be true, it evaluates the likelihood of getting the observed findings. The statistically significant nature of the observable variation is inversely proportional to the p-value. The formula for the calculation for P-value is depicted in Eq. (19)

$$p = \frac{x - \hat{x}}{\sqrt{\frac{x(1-x)}{N}}} \qquad (19)$$

Where, x is the assumed population proportion in the null hypothesis, $\hat{x}$ is the sample proportion, and N represents the sample size.

B. RMSE

It is resolute by separating the total variation between the actual data y and the predicted data $\hat{y}$ by the number of data N. A lesser RMSE number is usually preferable to a better value, with RMSE values ranging from 0 to +∞. The RMSE is formulated in Eq. (20)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y - \hat{y})^2} \qquad (20)$$

C. MAPE

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 7-Issue 1, January 2024
Paper: 62

MAPE is a metrics is calculated by the average percentage nonconformity error between predicted and actual data. MAPE is determining by dividing the difference between real data y and expected data $\hat{y}$ by actual data y and also by the number of data N. A lower MAPE value is preferred to a higher value.Eq. 21 depicts the MAPE formula.

$$MAPE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left|\frac{y-\hat{y}}{y}\right|} \qquad (21)$$

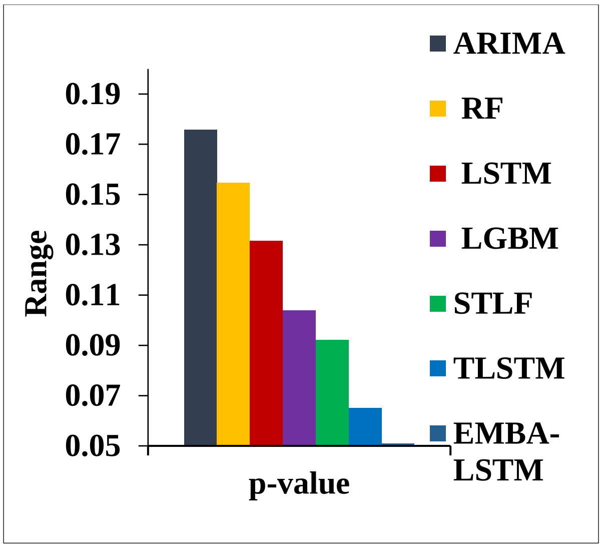

Figure 3. Comparison of p-value for given historical sales dataset

Figure 3 portrays the p-values obtained by the different models for retail store sales prediction. It is shown that the proposed EMBA-LSTM model results in decreased p-value by 70.89%, 67.01%, 61.22%, 50.91%, 44.56% and 21.54% respectively compared to the ARIMA, RF, LSTM, LGBM, STLF and TLSTM models.

Thus, it is concluded that the EMBA-LSTM model enhances the prediction performance by learning all kinds of features about historical sales data in contrast with the existing ARIMA, RF, LSTM, LGBM, STLF and TLSTM model.

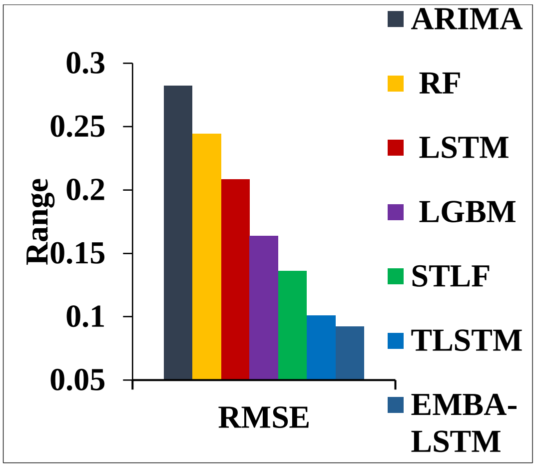

Figure 4. Comparison of RMSE for given historical sales dataset

Figure 4 portrays the RMSE values obtained by the different models for retail store sales prediction. When compared to the ARIMA, RF, LSTM, LGBM, STLF, and TLSTM models, the EMBA-LSTM model can reduce the RMSE by 67.28%, 62.21%, 55.68%, 43.62%, 32%, and 8.3%, respectively. As a result, it can be said that the EMBA-LSTM model improves prediction performance by learning all kinds of information about past sales data in contrast to the existing ARIMA, RF, LSTM, LGBM, STLF, and TLSTM models.

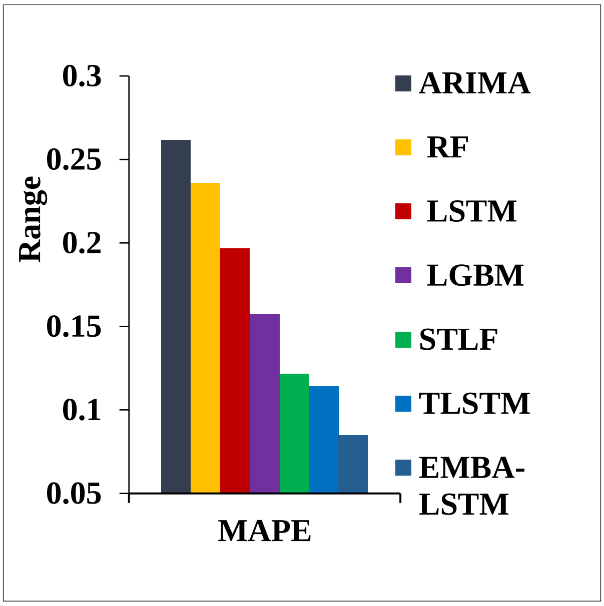

Figure 5. Comparison of MAPE for given historical sales dataset

The MAPE values generated by the various models for forecasting retail shop sales are shown in Figure 5. In comparison to the ARIMA, RF, LSTM, LGBM, STLF, and TLSTM models, it is demonstrated that the EMBA-LSTM model can reduce the MAPE by 68.51%,

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)

Volume 7-Issue 1, January 2024
Paper: 62

65.07%, 58.11%, 47.61%, 32.23%, and 27.91%, respectively. In contrast to the current ARIMA, RF, LSTM, LGBM, STLF, and TLSTM models, the EMBA-LSTM model improves prediction performance regarding historical sales data.

V. CONCLUSION

Forecasting retail sales is crucial for commercial operations across all sectors. Sales forecasting will assist in obtaining the information required to predict both the revenue and the profits with the use of retail sales projections. In order to estimate sales and increase income, a variety of ML models have been employed to analyze sale data and identify the crucial variables that affect sales. In this study, EMBA-LSTM is created to choose the best LSTM hyper-parameters and enhance the model's framework for accurate retail sales in stores prediction. In order to improve generalization, the EMBA is used to maximize the weights between the neuron nodes in the LSTM model. P-values, RMSE, which and the MAPE are utilized for evaluation. At last, the experiment results demonstrate that the EMBA-LSTM achieves a p-value of 0.051%, RMSE value of 0.092%, and MAPE value of 0.085 on the collected sales forecasting dataset when compared to the other models.

REFERENCES

1) Gupta, S., & Ramachandran, D. (2021). Emerging market retail: transitioning from a product-centric to a customer-centric approach. Journal of Retailing, 97(4), 597-620.
2) Minbo, L., Haipeng, W., Songkui, C., & Chang, L. (2017). Industrial big data analysis technology and tire sales data forecasting. ComputEngAppl, 53(11), 100-109.
3) Sethuraman, R., Gázquez-Abad, J. C., &Martínez-López, F. J. (2022). The effect of retail assortment size on perceptions, choice, and sales: Review and research directions. Journal of Retailing.
4) Bonfrer, A., Chintagunta, P., &Dhar, S. (2022). Retail store formats, competition and shopper behavior: A Systematic review. *Journal of Retailing*.
5) Ensafi, Y., Amin, S. H., Zhang, G., & Shah, B. (2022). Time-series forecasting of seasonal items sales using machine learning–A comparative analysis. International Journal of Information Management Data Insights, 2(1), 100058.
6) Padilla, W. R., García, J., & Molina, J. M. (2021). Improving time series forecasting using information fusion in local agricultural markets. Neurocomputing, 452, 355-373.
7) Bajaj, P., Ray, R., Shedge, S., Vidhate, S., &Shardoor, N. (2020). Sales Prediction Using Machine Learning Algorithms. International Research Journal of Engineering and Technology (IRJET), 7(06).
8) Tony, A., Kumar, P., &Rohith Jefferson, S. (2021). A study of demand and sales forecasting model using machine learning algorithm. Psychology and Education Journal, 58, 10182-10194.
9) Theresa, I., Medikonda, V. R., & Reddy, K. N. (2020). Prediction of Big Mart Sales Using Exploratory Machine Learning Techniques. International Journal of Advanced Science and Technology, 29(06), 2906-2911.
10) Raizada, S., & Saini, J. R. (2021). Comparative Analysis of Supervised Machine Learning Techniques for Sales Forecasting. International Journal of Advanced Computer Science and Applications, 12(11).
11) Almasarweh, M., &Alwadi, S. (2018). ARIMA model in predicting banking stock market data. Modern Applied Science, 12(11), 309.
12) Raditya, D., &Hanafiah, N. (2021). Predicting sneaker resale prices using machine learning. Procedia Computer Science, 179, 533-540.
13) Elmasdotter, A., &Nyströmer, C. (2018). A comparative study between LSTM and ARIMA for sales forecasting in retail.
14) Sarkar, M., & De Bruyn, A. (2021). LSTM response models for direct marketing analytics: Replacing feature engineering with deep learning. Journal of Interactive Marketing, 53(1), 80-95.
15) Pavlyshenko, B. M. (2019). Machine-learning models for sales time series forecasting. *Data*, *4*(1), 15.
16) Ji, S., Wang, X., Zhao, W., &Guo, D. (2019). An application of a three-stage XGBoost-based model to sales forecasting of a cross-border e-commerce enterprise. Mathematical Problems in Engineering, 2019.
17) Verstraete, G., Aghezzaf, E. H., &Desmet, B. (2020). A leading macroeconomic indicators' based framework to automatically generate tactical sales forecasts. Computers & Industrial Engineering, 139, 106169.
18) Wang, J. (2020, December). A hybrid machine learning model for sales prediction. In 2020 International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI) (pp. 363-366). IEEE.
19) Huber, J., &Stuckenschmidt, H. (2020). Daily retail demand forecasting using machine learning with emphasis on calendric special days. International Journal of Forecasting, 36(4), 1420-1438.
20) Ma, S., &Fildes, R. (2021). Retail sales forecasting with meta-learning. European Journal of Operational Research, 288(1), 111-128.
21) Saha, P., Gudheniya, N., Mitra, R., Das, D., Narayana, S., & Tiwari, M. K. (2022). Demand Forecasting of a Multinational Retail Company using Deep Learning Frameworks. IFAC-PapersOnLine, 55(10), 395-399.
22) Rohaan, D., Topan, E., &Groothuis-Oudshoorn, C. G. (2022). Using supervised machine learning for B2B sales forecasting: A case study of spare parts sales forecasting at an after-sales service provider. Expert systems with applications, 188, 115925.
23) Elalem, Y. K., Maier, S., & Seifert, R. W. (2022). A machine learning-based framework for forecasting sales of new products with short life cycles using deep neural networks. International Journal of Forecasting.
24) Dong, L., Zheng, H., Li, L., &Hao, L. (2022). Human-machine hybrid prediction market: A promising sales forecasting solution for E-commerce enterprises. Electronic Commerce Research and Applications, 56, 101216.
25) Li, D., Li, X., Lin, K., Liao, J., Du, R., Lu, W., & Madden, A. (2023). A Multiple Long short-term model for Product Sales Forecasting based on Stage Future Vision with Prior Knowledge. *Information Sciences*.
26) https://www.kaggle.com/datasets/manjeetsingh/retaildataset