Proceedings of the Fourth International Conference on Sentiment Analysis and Deep Learning (ICSADL-2025)
IEEE Xplore Part Number: CFP25UU5-ART; ISBN: 979-8-3315-2392-3

Stacking based Enhanced Sales Forecasting for E-commerce using Ensemble Learning approaches with Multi-Scaling Data

S.Balaji
Department of Computer Science Engineering
Saveetha School of Engineering
Saveetha Institute of Medical and Technical Sciences
Chennai, Tamilnadu, India.
kpm.balaji@gmail.com

D. Manikavelan*
Department of Computer Science Engineering
Saveetha School of Engineering
Saveetha Institute of Medical and Technical Sciences
Chennai, Tamilnadu, India
manikaveland.sse@saveetha.com

Abstract*—E-commerce and demand forecasting for large retail chains have taken the center stage in many business spheres. Sales prediction is one of the most important tools in inventory management and optimization of operations in retailing. This paper rates the applicability of some of the machine learning algorithms including Linear Regression (LR), Import Vector Machine (IVM), and Multi-Scaled Long Short-Term Memory (MLSTM) based on the ability to predict sales of big marts. Though the basic idea of LR was satisfactory, it was deficient in terms of non-linearity and categorical features, which resulted in suboptimal performance on the problem set with a high value of RMSE as 1142.004 and R^2 value as very low at 0.567 on the test data. Traditional models like Linear Regression struggle with non-linear relationships in sales data, leading to suboptimal performance. The superior handling of complex relationships through SVM turned out to be better than others with RMSE at 1133.55 and an R^2 value of 0.574. SVM is Effective for handling high-dimensional data but limited in capturing temporal dependencies. The MLSTM model designed to capture the temporal dependencies has had an improved accuracy with RMSE = 1086.291 and R^2 = 0.6. LSTM networks are good at capturing temporal patterns but may struggle with multi-scale time dependencies. An ensemble model of LR, SVM, and MLSTM using stacking was developed. In the proposed method, each base model has been orchestrated to exploit its strength, achieving better performance of RMSE = 1050.20, MAE = 720.35 and with the R^2 score of the test data as 0.65. The developed ensemble model that combines Linear Regression, SVM, and Multi-Scaled LSTM to overcome these challenges and improve sales forecasting accuracy. The scatter plot of the predicted values against the observed ones, for the ensemble model, closely followed the ideal line, therefore, showing strong predictability and stability. The method easily overcomes the weaknesses in standalone individual models and provides a more reliable and accurate tool for the forecasting of e-commerce sales and demand planning.

Keywords*— *Sales Forecasting, E-commerce Analytics, Machine Learning Algorithms, Ensemble Learning, Linear Regression and IVM

I. Introduction

Due to advancements in information technology and statistical analysis, business insights are now a crucial component of company support technologies[1]. To that respect, demand and sales forecasting is an integral element in the solution for business analytics whereby a company requires an accurate prediction of sales for them to use their S & OP, or operations and sales processes. In the wake of the growth over recent years in the industries of e-commerce & logistics, supply chains have taken a hit both in size and speed[2]. To manufacturers and retailers, a comprehensive forecast of a product's possible sale will help to make better judgments with regards to marketing and sales involvement, manufacturing, and even procurement plans.

Earlier, companies manufactured products without giving any consideration to sales volume or consumers' demand. Each manufacturing company needs market information on the demand for the product to decide whether to increase or decrease the production of millions of units [3]. Sales predictions by the business managers are usually arbitrary. However, capable managers are becoming less common and less trustworthy. In these regards, computer programmers can aid the process of sales forecasting either by acting in lieu of gifted managers in their absence or by providing those managers with the information they need to make the best decision by supplying possible sales forecasts. One way the above idea may be brought about would be to attempt to develop a computer program that imitates the abilities of expert managers [4]. Businesses operating in today's market face the risk of failure if these ideas are not carefully considered. When analyzing demand and sales, various metrics are used for each company. Due to the intense economic competition and ever-changing customer environment, businesses involved in manufacturing, distribution, or retail can greatly benefit from accurate and timely revenue forecasts, commonly known as revenue or sales forecasting [5].

Sales forecasting refers to the process of estimating sales in subsequent periods to assist decision-makers with better planning, supplying, producing, or marketing of the operations. Businesses use a number of strategies towards maintaining their levels of sales throughout the fiscal year. One such strategy is organizing sales promotions, whereby a bouquet of products is sold to retailers at discounted prices in return for the retailers displaying more goods for a certain period of time. In this study, we will carry out predictive analytics on the sales of Rossman using a large dataset. Sales forecasting, in general, can be viewed as a time-series problem when models and statistical analysis are applied to build forecasts. Regression analysis can also be used when using machine learning methods to uncover hidden patterns in historical period of data and then leverage those patterns and trends to predict future transactions, be it in the near or distant future.

Sales forecasting advantages vary depending on supply chain activities. In manufacturing, it helps optimize planning and decision-making across areas such as marketing, sales, production scheduling, and inventory management. However, inaccurate forecasts often result from the use of insufficient demand prediction software, despite the critical role of forecasting in business operations [7].

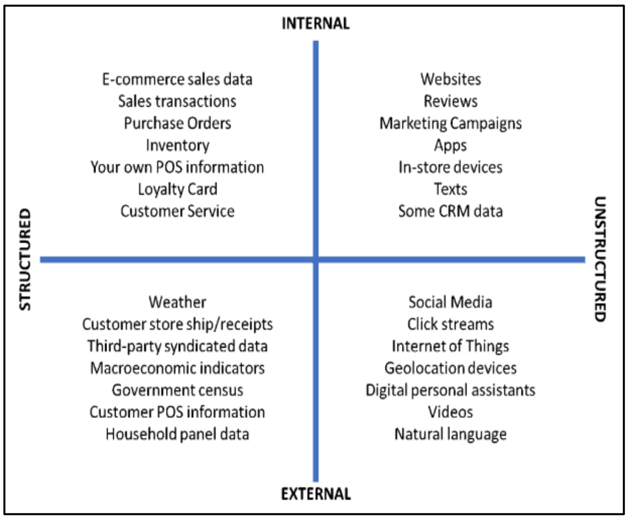

Fig. 1. Basic structure of demand forecasting

Sales forecasting is a major concern for every retailer, distributor, and manufacturer in the retail industry. For many companies in the supply chain, it also happens to be a major responsibility. Figure 1 demand and sales forecast provides a host of benefits with regard to the activity of supply and demand chains. The manufacturing companies can discover the fact that sales forecasting helps organizations in managing supply, planning production, marketing, and sales campaigns as well as other planning and decisionmaking procedures [8].

A time series is a set of data points that were collected at regular intervals among linked points in time. Time series forecasting makes forecasts based solely on past data. Enough reliable historical data must be accessible [6]. Traditional time series forecasting methods are used in this thesis to create a standard by which machine learning methods can be evaluated. A time series is a general problem that has huge practical importance in many domains. It enables one to make relatively rational predictions, based on previous values, about what the future values of the sequence might be, although not with absolute precision [9].

The previous system's design was unable to provide accurate results data that would have allowed the sales to be classified correctly [10]. Several machine learning Regressions have demonstrated predictive capabilities through dataset testing. We must create a system which can forecast sales more accurately, and we can do so by utilizing machine learning approaches to boost the system's effectiveness and accuracy.

We have identified three primary contributions that our work has made:

- The ensemble model comprising LR, SVM, and MLSTM models improves the accuracy of forecasting by a great degree. Using an ensemble model leverages the strengths of each model to make much better and more accurate predictions of sales while working with complex patterns in e-commerce data.
- The performance of the ensembled model is robust on diversified data features. To be specific, Linear Regression provides a baseline on linear relationships, while SVM manages the feature space in high-dimensional data effectively and MLSTM captures temporal dependencies in sequential data. This would mean a variety of ways by which the model would make sure it was effective at handling various facets of sales data; this would range from, but not be limited to, trends and seasonality.
- The integration of numerous machine learning algorithms into one ensemble model reduces the risk of overfitting inherent in one individual model. It thus creates a better generalization to new unseen data, increasing the reliability of sales forecasts.

In this dynamic e-commerce environment, ensembling balances strengths and weaknesses of each model for more stable and reliable predictions. A thorough introduction to machine learning, sales forecasting, its advantages, and its goals may be found in Section I. The literature review is discussed in Section II. This section looked at a few researches that used machine learning methods to predict sales. In Section III, the applied methodologies were covered. Section IV presents the recommended algorithms as well as the experiment results. The Conclusion was delivered in Section V.

II. Literature Review

In the recent work [11], the importance of efficient pricing methodology has been elaborated on, with especial consideration being given to the dynamic environment which in general was exacerbated by the rising online sales throughout the COVID-19 pandemic. In addition to reviewing the existing literature, this article examines some recent research on pricing techniques and market evaluation. The study also aims to provide repeatable pricing methodologies that are scalable price solutions and support the quick decision-making by merchants. To enhance artificial intelligence achievement, it entails extensive information preparation, feature engineering, data analytic investigation, and characteristic generation. models. While LightGBM, CatBoost, and XGBoost make their way into the world of ensemble learning approaches, the basis is struck for the new algorithm: X-NGBoost, a hybrid approach combining XGBoost with natural gradient boosting; this, in general, shows extraordinary speed and accuracy. Based on a number of given criteria, the current algorithms will be deeply investigated, and here, the study concludes by stating that X-NGBoost outperforms its opponents, which can be reflected by the reduction of RMSE.

The conclusion highlights the importance of the suggested methodology for small-scale businesses and proposes further research with the extension to forecast various e-commerce platforms. [12] started a large-scale effort to create a framework and efficient procedures, leveraging powerful machine learning algorithms to enhance the way customers choose well-priced products on e-commerce sites. The study proposes adaptability, which is applicable in online marketplaces without stocks and mainly targeted for inventory-centric e-commerce companies. Adapt or dynamic pricing techniques are proposed in the study to predict buying decisions through the use of machine learning and statistical models. Such architecture is developed on multiple sources of data, including but not limited to visit features, visitor features, purchase history, online data, and context awareness. It therefore combines machine learning with big data technologies and web mining into one solution landscape that addresses the challenges of e-commerce pricing optimization.

III. THE PROPOSED SYSTEM FOR FORECASTING

The processes of the dataset for Big Mart sale undergoes to build up the suggested model in order to produce accurate findings are depicted in the process flow diagram in Figure 2. There are seven steps in all, and each one is essential to developing the suggested model, which is a two-level statistics model. Preparing the data for development of both the single model and the stack model constitutes the first of five processes. After the model is built, its accuracy is evaluated by running it on fictitious data. A model that is superior is indicated by a lesser average absolute difference value.

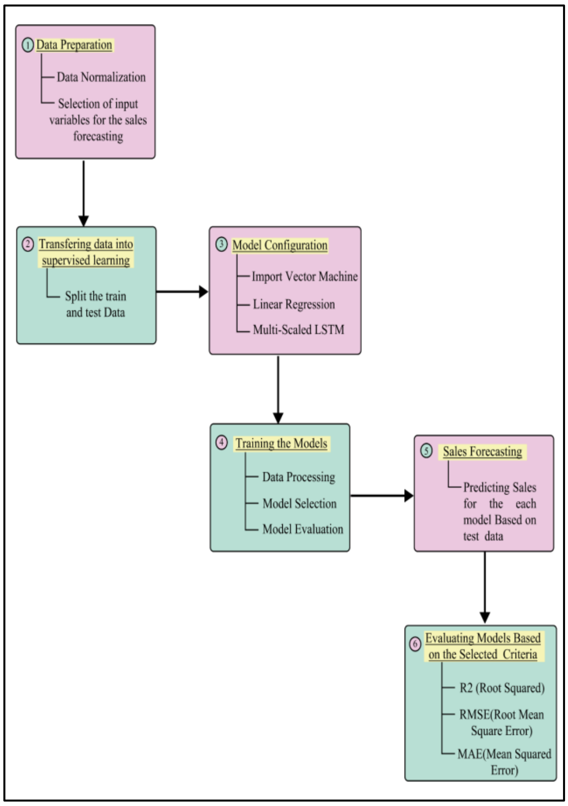

Fig. 2. Flow Chart for the Proposed System

A. Bigmart DataBase

Global Retail Corporation - Big Mart. Here, Big Mart Sales data from 2013 is utilized. Item identification, item fat, piece accessibility, piece kind, outlet category, piece MRP, item identification, item pounds, outlet size, outlets position kind, outlets organization year, and item outlet earnings are the 12 parameters that make up Big Mart revenue information.Except for the remainder, all of the qualities are predictor variables, while item outlet sales serve as the response variable. The following dataset contains 8523 observations. In the 80:20 split, the dataset was utilized for this purpose.

B. Hypothesis Generation

In any kind of evaluation, this is the most crucial stage. Knowing the issue description is the initial and most significant phase. Finding the features of something and the merchant (outlet) that affect a certain product's sales is the aim. Though the data is incomplete, many more characteristics would be helpful to understand the problem better

C. Data Exploration

Based on the study of this dataset, the following information of the data was looked upon: The item mass and outlet width require additional data. Since an object requires area to exist, it is theoretically inconceivable for the minimal value of item accessibility to be 0. Item accessibility can never be 0 as a result. The time period that these outlets were established, from 1985 to 2009, are listed. The values may not be suitable in this form. Therefore, we changed them to express the age of that particular outlet. The dataset has 10 unique outlets and 1559 unique goods. There are two types of item fat content, although some of them have the words "low fat" or "LF" typed incorrectly, and "regular" instead of "Regular."

D. Data Cleaning

We found missing Outlet Dimensions and Item Weight data throughout our study. Component weights are averaged to fill in missing values. Blank data for that specific type of outlet capacity are entered into the outlet size.

E. Feature Engineering

Throughout the phase of data research, some nuances were detected within the collected data. Therefore, in this stage, all the subtleties have been worked out, and hence, the data was fit for modeling. Further, item visibility was recorded to be zero. Item content of fat is corrected by replacing all of the incorrectly coded items with the correct ones. It is noted that sometimes the feature for fat content was on non-consumables, which just isn't possible. Therefore, "none" became the third category for item fat content. Each unique ID in the item identification property was found to start with either FD, DR, or NC. Thus, Foods, Drinks, and Non-consumables were added in the new column "Item Type New."A column "Year" is added that tells how many years have passed from the particular source. Import Vector Machines in Sales Predictions and Demand Forecasting: Import Vector Machines was the supervised learning technique developed by Boser, Guyon, and Vapnik in 1992 to build complex models using data. SVMs can be used as a model for simplifying complex relationships of several features of a variable with target variables in the sales predictions for e-commerce businesses and also for large retail chains' demand forecasting.

- Kernel Function: SVM uses kernel functions to transform low-dimensional data into high dimensional space so that the model detects complex patterns related to sales and demand datasets. Transformation helps identify non-linear relationships between features like seasonality, promotions, and consumer behavior.
- Hyperplane: In sales forecasting application domains, a hyperplane is the decision boundary that splits into two very different categories of sales or indicates a continuous value for the quantity of sales predicted based on the chosen input features. In terms of demand forecast, it deals with the identification of suitable hyperplanes for finding future demand levels by capturing the interaction between past sales and multiple influencing factors.
- Boundary Lines: They are used to form a boundary around the hyperplane. In regression, the lines forming the boundary outline the error margin (ε) of predictions. All points within this margin are support vectors important in calculating the appropriate model that best reduces the error of prediction.
- Support Vectors: These are the points nearest to the boundary lines that can define the decision boundary. In the case of sales prediction and demand forecasting, the support vectors represent the most important cases which can influence the forecast accuracy and therefore, this model is fine-tuned to handle the critical data.

F. Model Building

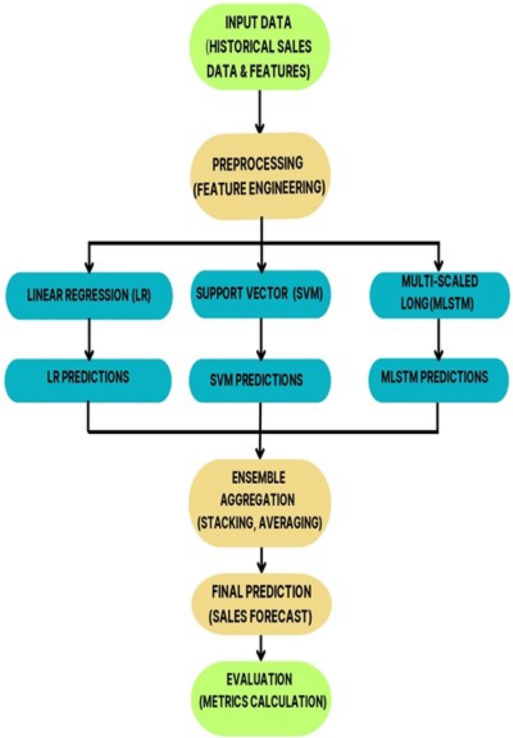

Fig. 3. A Two-tier Statistical Framework

Using SVM can enhance the accuracy calculation of sales forecasting for any business as well as help in stock management. This model can perform complex and high-dimensional data to analyze all related consumer behavior patterns, market trends, and seasonal fluctuations. SVM implementation in the above areas will optimize the stock levels, promotional activities, and overall business strategies. It is quite strong against local minima, uses kernels for the processing of non-linearity, and cares for support vectors for precision, which makes it a highly valuable algorithm for the prediction of sales and forecasting demand in the e-commerce and retail sectors. Once all the stages of data cleansing, exploration, and feature engineering are complete, the dataset is now ready for predictive model building. Model building basically means constructing a model that provides the most explanation of the correlation between the predictor and response variables. In this paper, the model-building process consists of two phases. First, only one model was developed for all general prediction methods such as support vector regression [17], k-nearest neighbor [18], cubist [14], regression tree [13], linear regression, and cubist [15, 16], etc. The second step involved the development of a two-level statistical model. The two-level statistical model embedded cubist, support vector regression, and linear regression as machine learning methods. Essentially, it is an assemblage of many models that are treated as a single model. Stacking can include more than one layer, which makes the model complex, but it may help with the making of an accurate prediction.

The operation of the two-level statistical framework is shown in Figure 3. The information set's initial characteristics feed bottom layer models cube, SVR, and regression analysis. Cubist then performs the role of a top layer approach, generating the last forecast utilizing as its input the predictions of the bottom layer architecture.

Linear regression: A time series is a sequence of measurements of the same variable taken over time, typically at regular intervals such as monthly or yearly. For instance, if we measure global temperature annually, we would use yt to represent the temperature at time t. In an autoregressive model, the current value of the time series is predicted based on its previous values. For example, in a first-order autoregressive model, the value yt is regressed on the value from the preceding period yt-1:

$$\mathrm{yt} = \beta 0 + \beta 1\mathrm{yt} - 1 + \epsilon\mathrm{t} \quad (1)$$

In this model, yt-1 serves as the predictor for yt, and ϵt represents the error term. The autoregressive order indicates the number of previous time periods used for making the current prediction; thus, an LR model uses only the immediately preceding value.

Multi-Scaled Long Short-Term Memory networks: An extension of traditional LSTM models to capture and model time patterns across multiple scales of time. They are greatly used in processing time series data where patterns occur at different time scales, like financial market trends, seasonal variation in sales, or varying frequencies in sensor measurements. Traditional LSTMs are inherently good at learning long-term dependencies within sequential data in overcoming issues of vanishing and exploding gradients inherent to standard RNNs. However, it may not be well suited to handle data that involves more than one time resolution. Multi-scaled LSTMs handle this problem by developing the ability to handle different scales over time simultaneously.

- Hierarchical Time Scales: Multiple LSTM layers or units can be used to improve the architecture of a simple LSTM. Different time scales can be achieved using many LSTM layers or units. For instance, one layer captures the short-term dependencies, for example, daily fluctuations, while another layer captures only long-term trends, such as seasonal patterns.
- Time-Scale Adaptive Units: This proposed model includes time-scale adaptive units. Such units can change their learning dynamics based on the appropriate timescale under which the input data is input. Moreover, the internal states and weights of the units are also modified to make better captures on patterns relevant to each of their scales.
- Multi-Scale Integration: The outputs of the various units at different time scales are integrated using a fusion mechanism such as concatenation or weighted averaging to produce an integrated representation of the time series. This integration enables the model to use information from multiple time scales to make more accurate predictions or classifications.
- Attention Mechanisms: Other advanced multi-scaled LSTM architectures introduce the use of attention mechanisms that would dynamically focus on different time scales according to contexts. That is, the model would selectively focus its attention on the importance of temporal features across various scales.

Finally, stacking, or stacked generalization, several base models are all trained on the same dataset in order to make a prediction. Those predictions then become features for a meta-model. The meta-model learns how to optimally combine the outputs from the base models to make a final prediction. It works through strengths of various models combined, aiming to widen the predictive performance by catching those aspects of the data which may have been overlooked by individual models. Aggregating the predictions of diverse models, stacking often produces better results than any one model in isolation. Thereby, forecasting performance has been improvised.

IV. Empirical Analysis

A. Models and Mertrics

The Linear Regression model, SVM &, MLSTM with proposed model was the three deciding models that fitted the data. Other methods, such as stochastic gradient descent were also tested; however, the results were not possible to take into consideration, most likely due to the small size of the training set and hence are not included here. All models are implemented in Python 3.9 and derived from the scikit-learn module R^2.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (2)$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}| \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (4)$$

Operating on the assumption that the relationship among the dependent variable y with the independent variable x is linear, this will yield a set of coefficients β_i that minimizes the remaining total of squares among the observed values in the dataset with the targets estimated by the model. In this case, the inclusion of the term intercept was for better results. The equation is given by:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots (5)$$

SVM in proposed method that works by constructing a number of decision trees during training, then averaging the results to reduce overfitting and improve prediction accuracy. In the case of regression tasks, it returns the average of projected values for each tree. In this experiment, the function to quantify the quality of splits was chosen as the mean squared error, and estimator was the parameter which defined the total amount of trees in the forest with the value 200. The technique of gradient boosting is a method of producing composite models from multiple simple models, usually decision trees. This follows a progressive stage-wise approach because the basic models are added one at a time, with every additional model moving the system closer to lowering the prediction error. The final overall full model would then be stronger by including more basic models. The word "regression" refers to a method that this algorithm uses to reduce losses: gradient descent. The learning rate & the total amount of boosting steps was set to 0.1 & 100, respectively, during the model's training

B. Procedure for Assessment

Selecting relevant features and training the models came after data processing & model selection. Based on how well a feature correlated with the target variable, it was chosen. Cramér's V, or Cramér's φ, is a statistic that quantified the relationship between nominal variables [19]. It was utilized as the measurement criterion for calculating correlation coefficients. Additionally, Python's dython module was used for the computation. Moreover, throughout the learning process, several hyperparameters were tried; the chosen one was displayed in the preceding section. Following the instruction, their outcomes were contrasted and examined

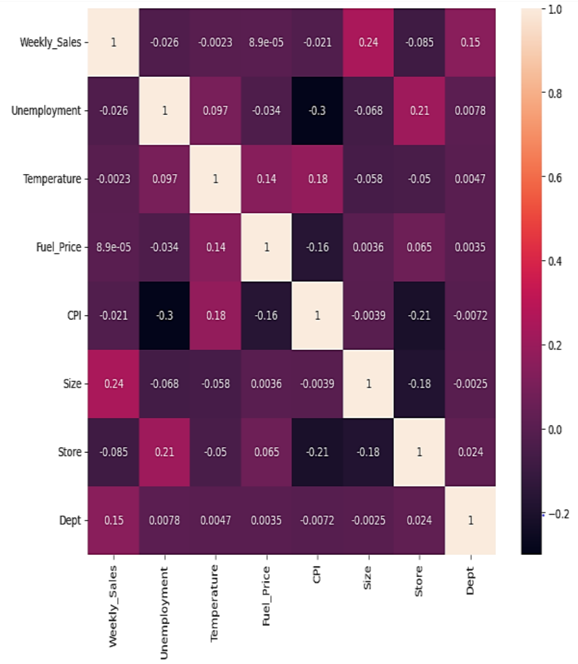

Fig. 4. Correlation coefficient heatmap

C. Correlation and Scatter plot analysis

This heat map that shows the association between each variable was created using the nominal association function in the Python library. Of course, the two features used for identification-which is the product ID and the outlet ID-were excluded in the analyses since they did not pertain to the objectives of the study. Figure 4 shows the association. Most of the parameters had a low relationship with the target variable, outlet sales, with the exception of MRP and outlet type. This may partly be explained because these two factors have strong relations to the type of outlet, themselves contributing to the moderate relationship of 0.21 among outlet size and sales. Among them, weight was 0.01, fat content was 0.03, and the year of establishment was -0.05-less correlated parameters under test during model training based on the coefficients of correlation obtained to put into being. According to the test results, the model accuracy was positively influenced by the two test parameters, namely decade of establishment and amount of fat content. Hence, the training of the model was done using the remaining 8 features by excluding weight.

In comparing and contrasting the performances of different models, scatter plots of modeled versus observed values are very good tools in determining what the accuracy is. In the case of the Linear Regression model, the scatter plot presented a very poor fit wherein predictions were quite far removed from observed values. This plot revealed a curve that had diminishing positive slope and an upper bound that seemed to be around 6000, which gave to incorrect predictions in those cases of low values observed. This easily resulted in situations such as the prediction of negative sales values-this is impossible in practical scenarios. These problems arise from the fact that the model mainly functioned with categorical features, which are obviously not appropriate for a linear regression model.

Figure 5 showed the SVM model is a better alignment in the scatter plot compared to Linear Regression, showing a much better fit with both training and testing datasets. It showed great robustness with an R^2 score of 0.937 while training but its predictive accuracy on test data showed limitations as it reported values around 1133.55 RMSE. Here, we have seen the SVM on data sets involving complex relationships; still, its performance against outliers was demonstrated while sometimes it was struggling with some of the data. Figure 6 displaying the MLSTM model scatter plot was much better than the one in the regression and SVM models; it had a closer fit to the test data, revealing an R^2 value of 0.6, which indicated its ability to cope with complex patterns in the data and temporal dependencies. The MLSTM model had been underperforming in some instances, with its RMSE standing at 1086.291, which should be refined for predicting in other areas.In this case, the best overall performance was clearly demonstrated by the proposed ensemble model of Linear Regression, SVM, and MLSTM combined through stacking. The ensemble model systematically benefited from the strengths of its base models and could reduce the RMSE and MAE for test data while also raising the R^2 score.

Table 1. RMSE measures the average magnitude of the error. MAE provides a clear interpretation of the average error. R^2 indicates the proportion of variance explained by the model. The proposed ensemble model achieved the lowest RMSE (1050.20) and MAE (720.35) compared to individual models (Linear Regression, SVM, and MLSTM). The R^2 score of the ensemble model (0.65) was higher than that of the individual models, indicating better predictive accuracy. The proposed ensemble model outperformed all these techniques, achieving an RMSE of 1050.20, MAE of 720.35, and R^2 of 0.65. This hybrid method improved forecast precision and reliability on outlier values while accounting for some weaknesses that could be attributed to the standalone models.

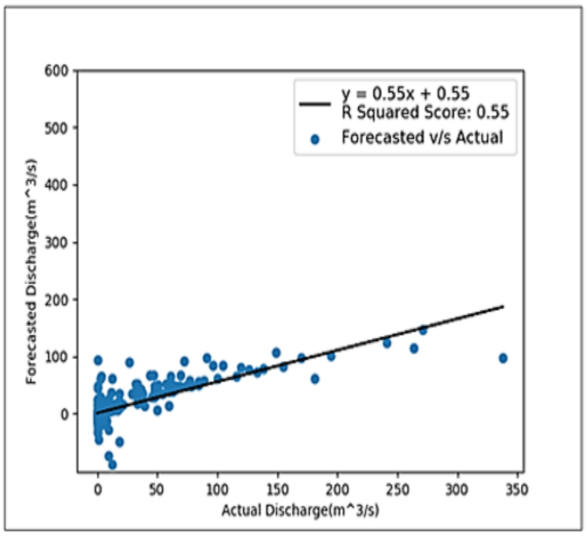

Fig. 5. Proposed ensemble findings displayed as scatter

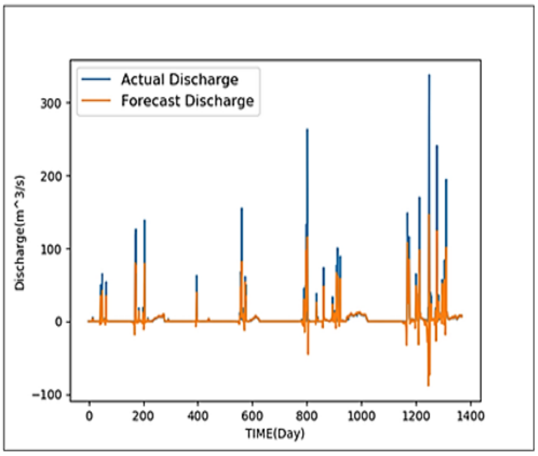

Fig. 6. Proposed ensemble findings displayed as plots

TABLE I. PERFORAMNCE ASSESSMENT USING ERROR RATE

Metric	Linear Regression	SVM	MLSTM	Proposed Ensemble Model
RMSE (train)	1124.926	425.269	1029.046	**396.215**
MAE (train)	834.714	293.638	727.225	**281.485**
R2 (train)	0.562	0.937	0.633	**0.949**
RMSE (test)	1142.004	1133.55	1086.291	**1050.20**
MAE (test)	840.73	790.548	753.112	**720.35**
R2 (test)	0.567	0.574	0.67	**0.651**

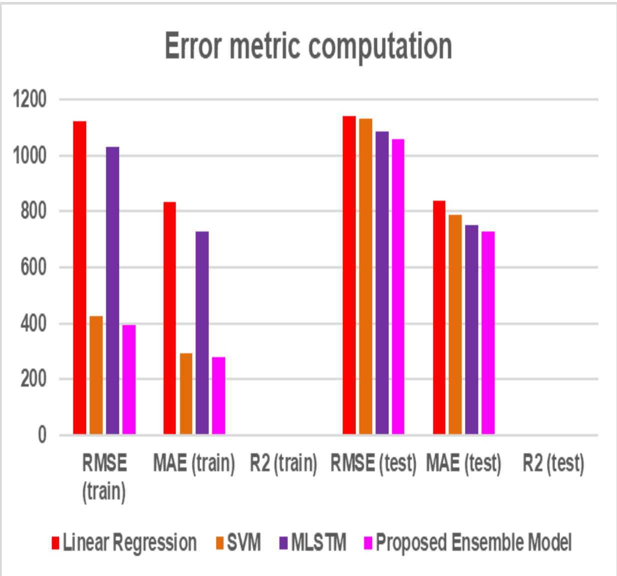

Fig. 7. Error computation

Among the three as per figure 7, the best overall performance was that of Gradient Boosting; it had the least mistakes on the test and the highest R^2 score. Indeed, the results proved that trees which are gradient boosted are superior on average to random forests, and random forests did somewhat worse than gradient boosting in general [20]. The overall lowest accuracy corresponds to a linear regression model according to RMSE and MAE. While making the comparison between these three, its RMSE was similar to those of the remaining two models because the number of outliers in their results was higher as a result of failing to correctly estimate data with large values.

V. CONCLUSTION

The suggested model of the ensemble encompasses LR, SVM, and MLSTM-a really big enhancement in the ability of e-commerce business sales forecasting and demand prediction for large retail chains. The potential of an ensemble approach to combine the advantages of each of the individual models in this task increases the level of accuracy and robustness of the predictions compared with stand-alone methods. While the integration of LR provides a foundation for linear trends, SVM is superior in dealing with high-dimensional complex data, whereas MLSTM captures intrinsic temporal dependencies in data. Each of the diverse algorithms is combined within an ensemble model to realize improved generalization and reduced overfitting for more reliable and accurate forecasts. In fact, this holistic approach further improves the predictive performance and enables the business to have a strong tool for better inventory management and strategic planning. This makes the ensemble model stand out as an important contribution within the realm of machine learning applied to sales predictions, helping in driving more informed decisions and resource allocation in the e-commerce sector.

REFERENCES

[1] G. Behera, A. Bhoi, and A. K. Bhoi, “A Comparative Analysis of Weekly Sales Forecasting Using Regression Techniques,” Lect. Notes Networks Syst., vol. 431, pp. 31–43, 2022, doi: 10.1007/978-981-19-0901-6_4/COVER

[2] T. Weng, W. Liu, and J. Xiao, “Supply chain sales forecasting based on lightGBM and LSTM combination model,” Ind. Manag. Data Syst., vol. 120, no. 2, pp. 265–279, 2020, doi: 10.1108/IMDS-03-2019-0170.

[3] G. Behera and N. Nain, “A comparative study of big mart sales prediction,” Commun. Comput. Inf. Sci., vol. 1147 CCIS, no. October, pp. 421–432, 2020, doi: 10.1007/978-981-15-4015-8_37.

[4] Asokan, R., & Preethi, P. (2021). Deep learning with conceptual view in meta data for content categorization. In Deep Learning Applications and Intelligent Decision Making in Engineering (pp. 176-191). IGI global.

[5] Y. jiang Li, Y. Yang, K. Zhu, and J. Zhang, “Clothing Sale Forecasting by a Composite GRU-Prophet Model With an Afention Mechanism,” IEEE Trans. Ind. Informatics, vol. 3203, no. c, pp. 1–9, 2021, doi: 10.1109/TII.2021.3057922.

[6] B. Sri, S. Ramya, and K. Vedavathi, “An Advanced Sales Forecasting Using Machine Learning Algorithm,” Int. J. Innov. Sci. Res. Technol., vol. 5, no. 5, pp. 342–345, 2020.

[7] A. Lasek, N. Cercone, and J. Saunders, “Sales and customer demand forecasting: Literature survey and categorization of methods,” Lect. Notes Inst. Comput. Sci. Soc. Telecommun. Eng. LNICST, vol. 166, pp. 479–491, 2016, doi: 10.1007/978-3-319-33681-7_40.

[8] J. T. Menker and M. A. Moon, “Sales forecasting management : a demand management approach,” p. 347, 2005.

[9] A. Tealab, “Time series forecasting using artificial neural networks methodologies: A systematic review,” Futur. Comput. Informatics J.,

vol. 3, no. 2, pp. 334–340, Dec. 2018, doi: 10.1016/J.FCIJ.2018.10.003.

[10] S. Kohli, G. T. Godwin, and S. Urolagin, “Sales Prediction Using Linear and KNN Regression,” pp. 321–329, 2021, doi: 10.1007/978-981-15-5243-4_29

[11] Namburu A, Selvaraj Varsha M. (2022). Product pricing solutions using hybrid machine learning algorithm. Innov Syst Softw Eng. 2022 Jul 25:1-12. Doi: 10.1007/s11334-022-00465-3. Epub ahead of print. PMID: 35910813; PMCID: PMC9309595

[12] Gupta, R., & Pathak, C. (2014). A machine learning framework for predicting purchase by online customers based on dynamic pricing. Procedia Computer Science, 36, 599-605

[13] Wei-Yin Loh, “Classification and Regression Trees”, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Vol. 1(1), pp. 14-23, January, 2011.

[14] J.R. Quinlan, “Learning with Continuous Classes”, Proceedings of the 5th Australian Joint Conference on Artificial Intelligence, pp. 343- 348, Tasmania, November, 1992.

[15] J.R. Quinlan, “Combining Instance Based and Model Based Learning”, Proceedings of the Tenth International Conference on Machine Learning, pp. 236-243, San Mateo, June, 1993.

[16] Yong Wang and Ian H. Witten, “Inducing Model Trees for Continuous Classes”, Proceedings of the Ninth European Conference on Machine Learning, pp. 128-137, Czech Republic, April, 1997.

[17] Preethi, P., & Asokan, R. (2020, December). Neural network oriented roni prediction for embedding process with hex code encryption in dicom images. In Proceedings of the 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida, India (pp. 18-19).

[18] Padraig Cunningham and Sarah Jane Delany, “k-Nearest Neighbour Classifiers”, Multiple Classifier Systems, Vol. 34, pp. 1-17, March, 2007.

[19] Cramér, H.: Mathematical Methods of Statistics (PMS-9), Princeton: Princeton University Press, 2016.

[20] Caruana, R., Alexandru, N. M.: An Empirical Comparison of Supervised Learning Algorithms. Proceedings of the 23rd International Conference on Machine Learning – ICML, 06, (2006).