DETERMINING AND PREDICTING THE MOST EFFICIENT RETAIL SALES STRATEGY

Routhu Srinivas Chaitanya, Jagarapu Neelima, Medisetti Pavan Sai, Mrs. G. Indira Devi (Assistant Professor) Department of Electronics and Computer Engineering, Vignan's Institute of Information Technology,Visakhapatnam-India.

Abstract

The sales forecast is used to track the quantity of limited commodities' so that the dearth or else redundant commodities can be reduced, the demand of any retailercan be fulfilled while prepare their demand on schedule, and following proper arrangements with suppliers can be done with co-operation. The main approach of this paper is to consider case studies of deals soothsaying using artificial literacy. This paper evaluates and compares colorful Machine literacy models like Random Forest, Linear regression, ARIMA and XG Boost to read retail store deals. Training data set contains once deals where Colorful data cleaning and exploratory data analysis algorithms were enforced over raw datasets before used for modeling. Mean square errors, R square error and Root mean square errors were estimated for individual prognostications from time series and machine literacy models. In this the best model was selected.

Keywords:online Linear retrogression, XG Boost, ARIMA, Random Forest

Introduction

Ultramodern business intelligence includes deals soothsaying as a crucial element.It can be delicate content to break, especially when there is a deficit of matter, missing data, or outliers. A series of time deals can be considered. At this moment, colorful time series models have been developed. It discusses colorful time series ways. The authors investigate the predictability of time series and compare the results of several time series soothsaying algorithms.. Any visit to a company that uses deals soothsaying in its operations or marketing indicates that the most significant element of soothsaying for people in charge of organizational soothsaying is a commodity much more abecedarian than the approach chosen. Vaticinating deals bring value to a company's nethermost line. Soothsaying is used by finance to induce budgets for capacity plans and employment, while deals soothsaying is used by products to plan their cycles. Deals operations with home and share planning, force chain with material procurement and the product capacity, and deals strategy with channel and mate strategy all benefit from soothsaying. Different conditions for effective cast combining were addressed in this study. They lagged variable selection, hyperactive parameter optimization, and a comparison of classical and machine literacy-grounded time series ways. The author demonstrated that both classical and machine literacy-grounded ways may be employed on temperature time-series dataset.

Time-series approaches to deals soothsaying have some downsides-

- Need to consider a large number of exogenous rudiments that impact deals.
- To capture seasonality, have to literal data over a lengthy period. Still, need to constantly warrant literal data for a target variable, similar to when a new product is introduced. At the same time, they have a deals time series for an analogous product, and they may anticipate an analogous deals pattern for their new product.
- The important ways of Deals Soothsaying- Check of Buyers Intention, Opinion bean of deals force, Expert opinion, Request test system, Protuberance of once deals, Products in use analysis, Time series analysis, Statistical demand analysis.

For the retail shop data sets, a comparative analysis was performed. The sales of a retail store were obtained and subjected to several analysis in order to determine the optimum approach for retail store sales analysis. X.G. Boost, Random forest, ARIMA, linear regression model, where X.G.Boost is quite accurate.

Data Used

This collection contains the retail store's sales for the previous three years. A fragment of the dataset that was used is shown in Figure 1

	date	store	item	sales
0	2013-01-01	1	1	13
1	2013-01-02	1	1	11
2	2013-01-03	1	1	14
3	2013-01-04	1	1	13
4	2013-01-05	1	1	10

Figure 1. The figure is comprised of slaes of items on a specific day.

Review of Literature

Making the appropriate business decisions is incredibly crucial in today's environment, when competition is severe. It assists in determining product quantities by taking into account the vacuity of installations such as the outfit, capital, force, space, and so on. Being able to directly predict the sales of a specific product in a specific nation may be quite beneficial to the business owner. Many individuals have worked on developing a model to predict agreements, and they've created dozens of models using various algorithms and approaches.

In [1] the author used the Walmart sales information from the kaggel website to demonstrate that the NN model performs better than the linear regression model.

In [3,] the author proposed a novel approach for predicting sales by combining the X.G. Boost algorithm with feature engineering processing on a sales dataset. The newly proposed model outperformed existing models.

In [5] author analysed all of the methodologies and determined that the ARIMA model outperforms the others. Hybrid models were also included in the investigation, and hybrid models provided the best forecasting results.

Block Diagram

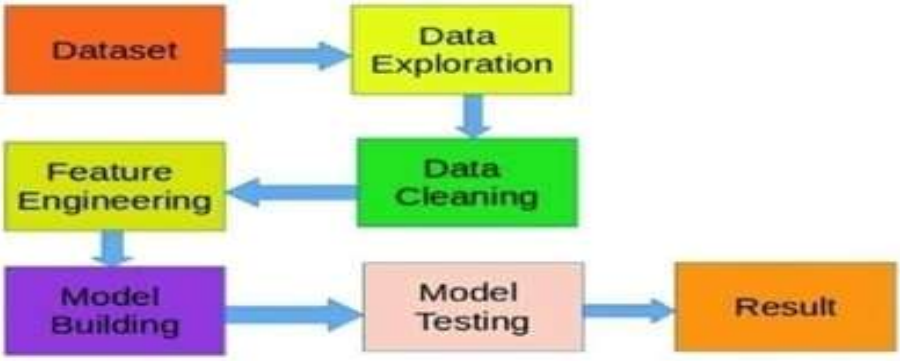

Figure 2. Overall Experiment Workflow

Research Methodology

A. Importing Data Set

The data set is utilised to train a vaticination engine on how to use the information. A recordset is a collection of mutual records that deal with a set of time-varying parameters. Use the data you've gathered to train predictions. A recordset can only hold three records at a time. The target time series is one, followed by the related time series and the metadata of the record type item.

B. Linear Regression

The utilization of a direct regression procedure to illustrate the relationship between the dependent and independent variables, which is why it is called Linear Regression. These variables are represented by x and y. It depicts the direct link between the value's dependent variable and the independent variable.

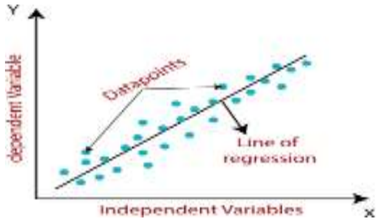

Figure 3. Linear regression model

Applying Statistical regression to a given dataset gives the plot

Figure 4. Linear regression analysis

C. X.G.Boost

XGBOOST was introduced by TIANQI CHEN, a great machine learning library. It extends Grade Boosting Machines, which has been conceded as one of the most effective supervised literacy algorithms. XGBoost stands for “Extreme Gradient Boosting," according to the abstract. Grade Boosting Machines are one of the stylish performing algorithms in supervised literacy, and XGBoost is an extension of them. XGBoost stands for “Extreme Gradient Boosting," according to the abstract. The model and parameters are the abecedarian aspects of XGBoost, as it's a supervised literacy strategy. The model is a fine model that's used to prognosticate issues grounded on input values, and the parameters must be learned from the data set. We are trying to produce an objective function to quantify the performance of a particular model with specific parameters since weneed to determine the optimal parameters for a specific training data set.obj (Θ) = L (θ) Ω (Θ). The loss of the training and the regularization element are two crucial factors of the objective function. The training loss is a metric that measures how well the model fits the training data. The model's complexity is measured via regularization.

Applying XG Boost to a given dataset gives the plot

Figure 5.X G Boost analysis

D. Random Forest

Random Forest is known for its artificial literacy. It uses the algorithm of a supervised literacy system. In artificial literacy, it can be done for bracket as well asretrogression issues. It is grounded in to use of multiple learning algorithms, it is a system of several classifiers to break a difficult problem and to increase the model'sperformance. If the timber has a bigger number of trees, then the problem of overfitting is avoided and it is more efficient and accurate.

Figure 6. random forest analysis

E. ARIMA

The RIMA (r, s, t) soothsaying equation is as follows: (if necessary). If the statistical properties of a time series' arbitrary variable stay harmonious throughout time, it is considered to be stationary. A desk-bound collection has no trend, has consistent breadth variations around its mean, and jiggles in a pleasing manner, i.e., its arbitrary temporal styles seem statistically identical across short time periods. The ultimate criteria implies that its serial correlations are consistent over time, or that its power diapason is balanced over time. The sign of being a sample of quick or sluggish indicating regression, sinusoidal oscillation, or rapid-hearthplace signal alternation, with a seasonal aspect, can be noticed in an arbitrary variable. A pattern of quick or slow mean regression, sinusoidal oscillation, or rapid-fire sign alternation with a seasonal element is the signal (if one exists). The ARIMA soothsaying equation comprises a direct, retrogression-type equation for a constant time series. That is the predicted value of Y = a, which is a weighted sum equal to one or more recent values. If the predictors are merely lagged values of Y, the model is a pure autoregressive (" tone- regressed") model, which may be filled with standard retrogression tools.

Applying the ARIMA model to a given dataset gives the plot

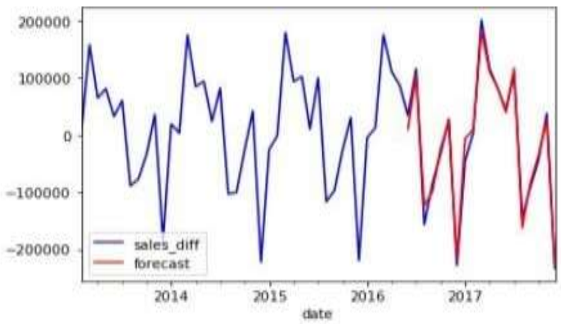

Figure 7. ARIMA model analysis.

Results and Discussion

MSE, RMSE And R Square are calculated for every algorithm. MSE can be calculated as MSE=(M+1)/M*sum (k to M) (sqr (Yk-(AkXk+A0) Where, M is the total number of observations, Yi = Actual value, (a1xi+a0) = Predicted value, MAE = sum (YK – xk)/m; m = number of instances of each observation set. R- squared is a statistical approach for determining virtuousness of fit between two sets of data.

S No.	Model Name	Mean Square Error(MSE)

1	Linear regression	433.0000000
2	Random Forest	15832.75
3	X.G.Boost	11649.66667
4	ARIMA	11265.33749

Table 1. MSE values

S No.	Model Name	R square error
1	Linear regression	0.990716
2	Random Forest	0.987794
3	X.G.Boost	0.993498
4	ARIMA	0.993498

Table 2. R square error values

S No.	Model Name	Root Mean Square Error (RMSE)
1	Linear regression	16221.040791
2	Random Forest	18599.232966
3	X.G.Boost	135474.79262
4	ARIMA	14959.893467

Table 3. RMSE values

Model Representation of MSE&RMSE.

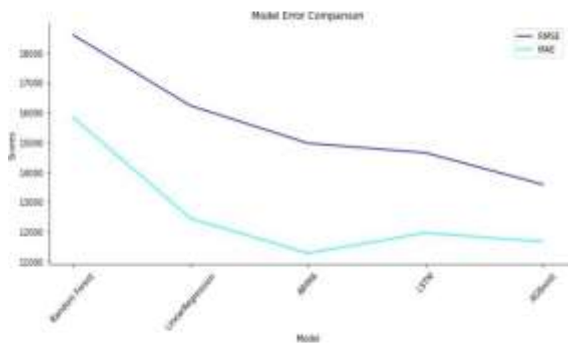

Figure 8. Resulting output of all models

Conclusion

Sales forecasting is a significant issue of the strategy implementation process because it helps a firm to foresee its future performance. The following is a summary of the models' performance:

S No.	Model	MSE	R square	RMSE
1	Linear regression	433.0000000	0.990716	16221.040791
2	Random Forest	15832.75	0.987794	18599.232966
3	X.G.Boost	11649.67	0.993498	135474.79262
4	ARIMA	11265.33749	0.993498	14959.893467

Table 4. MSE,Rsquare,RMSE values comparision

All the regression models are performing with an excellent R-squared and stable RMSEvalue. The most accurate is the one which has least RMSE value, R-square value neat to 1 and average MSE. From table[4] among Linear regression, ARIMA, Random Forest, and XGBoost Algorithm, X.G.Boost has least RMSE,average MSE and R-square value near to 1, So XGBoost is the accurate algorithm for retail stores.

References

[1] Khan, Muhammad Adnan, Shazia Saqib, Tahir Alyas, Anees Ur Rehman, Yousaf Saeed, Asim Zeb, Mahdi Zareei, and Ehab Mahmoud Mohamed. "Effective demand forecasting model using business intelligence empowered with machine learning." *IEEE Access* 8 (2020): 116013-116023

[2] Chen, Junhang, Wang Koju, Shurui Xu, and Zhe Liu. "Sales Forecasting Using Deep Neural Network And SHAP techniques." In *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, pp. 135-138. IEEE, 2021.

[3] Y. Niu, "Walmart Sales Forecasting using XGBoost algorithm and Feature engineering," *2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE)*, 2020, pp. 458-461, doi: 10.1109/ICBASE51474.2020.00103

[4] Pavlyshenko, Bohdan M. "Machine-learning models for sales time series forecasting." *Data* 4, no. 1 (2019): 15

[5] Gurnani, Mohit, Yogesh Korke, Prachi Shah, Sandeep Udmale, Vijay Sambhe, and Sunil Bhirud. "Forecasting of sales by using fusion of machine learning techniques." In *2017 International Conference on Data Management, Analytics and Innovation (ICDMAI)*, pp. 93-101. IEEE, 2017

[6] Krishna, Akshay, V. Akhilesh, Animikh Aich, and Chetana Hegde. "Sales-forecasting of retail stores using machine learning techniques." In *2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, pp. 160-166. IEEE, 2018.

[7] Kiran, Manju, Amit Kumar Sharma, and Divakar Venkatesh. "Automation of Best-Fit Model Selection using a Bag of Machine Learning Libraries for Sales Forecasting." (2021).

[8] Gustriansyah, Rendra, Dian Palupi Rini Ermatita, and Reza Firsandaya Malik. "Integration of decision-making method and data-mining method as a preliminary study of novel sales forecasting method." *International Journal* 9, no. 4 (2020).

[9] Dalrymple, Douglas J. "Sales forecasting practices: Results from a United States survey." *International journal of Forecasting* 3, no. 3-4 (1987): 379-391.

[10] Peng, Lifang, Qinyu Liao, Xiaorong Wang, and Xuanfang He. "Factors affecting female user information adoption: an empirical investigation on fashion shopping guide websites." *Electronic Commerce Research* 16, no. 2 (2016): 145-169.

[11] Panzhi, Ni, Binhui Peng, Yitong Zhang, and Yuping Yan. "An Automatic System for Sale

Forcasting with Memory Reduction Technique." In *2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, pp. 367-370. IEEE, 2021

[12] Tugay, Resul, and Sule Gunduz Oguducu. "Demand prediction using machine learning methods and stacked generalization." *arXiv preprint arXiv:2009.09756 (2020).*

[13] Kharfan, Majd, and Vicky Wing Kei Chan. "Forecasting Seasonal Footwear Demand Using Machine Learning." (2018).

[14] Wisesa, Oryza, Andi Adriansyah, and Osamah Ibrahim Khalaf. "Prediction analysis sales for corporate services telecommunications company using gradient boost algorithm." In *2020 2nd International Conference on Broadband Communications, Wireless Sensors and Powering (BCWSP)*, pp. 101-106. IEEE, 2020

[15] Paul, Justin, and Mark Rosenbaum. "Retailing and consumer services at a tipping point: New conceptual frameworks and theoretical models." *Journal of Retailing and Consumer Services* 54 (2020): 101977

[16] Di Pillo, Gianni, Vittorio Latorre, Stefano Lucidi, and Enrico Procacci. "An application of support vector machines to sales forecasting under promotions." *4OR* 14, no. 3 (2016): 309-325.

[17] Sarhani, Malek, and Abdellatif El Afia. "Electric Load Forecasting Using Hybrid Machine Learning Approach Incorporating Feature Selection." In *BDCA*, pp. 1-7. 2015

[18] Lu, Chi-Jie, and Ling-Jing Kao. "A clustering-based sales forecasting scheme by using extreme learning machine and ensembling linkage methods with applications to computer server." *Engineering Applications of Artificial Intelligence* 55 (2016): 231-238

[19] Mortensen, Stephen, Michael Christison, BoChao Li, AiLun Zhu, and Rajkumar Venkatesan. "Predicting and Defining B2B Sales Success with Machine Learning." In *2019 Systems and Information Engineering Design Symposium (SIEDS)*, pp. 1-5. IEEE, 2019.

[20] Megahed, Aly, Peifeng Yin, and Hamid Reza Motahari Nezhad. "An optimization approach to services sales forecasting in a multi-staged sales pipeline." In *2016 IEEE International Conference on Services Computing (SCC)*, pp. 713-719. IEEE, 2016.

[21] Helmini, Suleka, Nadheesh Jihan, Malith Jayasinghe, and Srinath Perera. "Sales forecasting using multivariate long short term memory network models." *PeerJ PrePrints 7 (2019): e27712v1.*

[22] Catal, Cagatay, E. C. E. Kaan, Begum Arslan, and Akhan Akbulut. "Benchmarking of regression algorithms and time series analysis techniques for sales forecasting." *Balkan Journal of Electrical and Computer Engineering* 7, no. 1 (2019): 20-26.

[23] Spiliotis, Evangelos, Spyros Makridakis, Artemios-Anargyros Semenoglou, and Vassilios Assimakopoulos. "Comparison of statistical and machine learning methods for daily SKU demand forecasting." *Operational Research (2020): 1-25*

[24] Kaneko, Yuta, and Katsutoshi Yada. "A deep learning approach for the prediction of retail store sales." *In 2016 IEEE 16th International conference on data mining workshops (ICDMW), pp. 531-537. IEEE, 2016.*

[25] Ma, Shaohui, and Robert Fildes. "Retail sales forecasting with meta-learning." *European Journal of Operational Research* 288, no. 1 (2021): 111-128.

