Recent Trends in Intensive Computing
M. Rajesh et al. (Eds.)

doi:10.3233/APC210210

Hybrid Customer-Centric Sales Forecasting Model Using AI ML Approaches

Sanket Londhe [a,1], Sushila Palwe [a]

[a]Department of Computer Science and Engineering, MIT World Peace University, Pune, MH

Abstract. Business Intelligence is a process of preparing, analyzing, presenting, and maintaining the data to gain insights for the decision-makers to make informed decisions. While there are many approaches to predict the growth based on the sales figures a very few consider the influence of customer data on the forecasting and the relevance of the same while making the predictions. So, in this study, we will look at some of the existing techniques used so far to make predictions and studies used to understand the customer data. With the analysis, we shall try to devise a hybrid approach to the traditional sales prediction, which would include a customer-centric data analysis. We shall look at some of the techniques which are traditionally used in Market Basket analysis and at the same time look at the techniques like classification, segmentation, regression, etc. to get a perception of the impact of customer data on sales forecasting. We shall highlight all the pros and cons of the algorithms and try to come up with an intelligent approach that would give accurate results

Keywords. Business Intelligence, Sales Forecasting, Customer Analysis, Classifiers, Regression, Segmentation, ARIMA, LSTM, RFM

1. Introduction

Business intelligence (BI) is an analytics-driven process that combines analytics and data processing techniques along with data visualization tools and best practices to help organizations to make data-driven decisions. The refined processes are supported by data to make a case. In practice, data-driven operations eliminate inefficiencies, and quickly adapt to market or supply changes. [1,6] These processes include: Data mining techniques and preparation; Knowledge discovery and transfer by reports; Benchmarking to Gain actionable insights; Querying and Statistical analysis prior to Visual analysis

Analytics experts use BI concepts and techniques to make informed decisions using historical data and its impact on the present modes of operations. Using various data visualization techniques analysts gain insights and observe patterns to ensure the growth of the organizations by taking relevant actions quickly. Today BI is used not only to increase sales but also to ensure the security of the data, understand consumer behavior, devise marketing strategies, and much more. [1]

AI and ML approaches contribute to BI strategy by integrating business insights into the strategy.

[1]Sanket Londhe, Department of Computer Science and Engineering, MIT World Peace University

Email: sanketlondhe99@gmail.com.

As companies use more and more data for their business operations, efforts would be made for a shared data strategy and collaborative data projects. A Business Intelligence Framework becomes a critical part of the BI Strategy across the organization.[2]

In this paper, we will look at some of the work done in the field of Customer Analytics and Sales forecasting and see various techniques used in their respective fields. Based on the studies we will come up with an approach that will be a combination of the said concept and compliant with the Business Intelligence concept.

2. Related Work

Customer analysis is a process having huge importance at different stages of a business plan. It is very important to know your product and the target audience who may want or benefit from your product [3,5]. An article from Forbes states that 81% of businesses are dependent on customer analytics insights to improve their understanding of the customers [4].

The approach used in [9] uses various models like Logistic Regression, Linear SVM, Decision Trees, Random Forest, Quadratic Discriminant Analysis, and Neural Networks to estimate sales using customer classification. For sales forecasting the authors in [9] used various regression trees and Ridge Regression. These classifiers can be implemented using the python Scikit-Learn library [13].

Another study [10] uses the RFM (Recency, Frequency & Monetary Value) model for segmentation using the k-means algorithm to understand customer purchase patterns and behavior.

Time-series forecasting is one of the fundamental concepts when it comes to making predictions and analyzing trends based on historic values [7]. Since the intent of this survey is to forecast sales, the data we are considering is traditionally non-stationary in nature. ARIMA (Autoregressive Integrated Moving Average) is one such commonly used technique used to analyze non-stationary data. The study in [11] emphasizes the benefits of ANNs in various fields like finance, engineering, and others where time series forecasting holds critical importance.

We intend to define our model to work on data of a smaller size hence the work done in [11] makes sense and they have proposed their model based on RNN, ARIMA, and part sentiment analysis. They claim their model works well in the short term without compromising the results. The focus of the study in [11] is to assess the consumer perception about the quality of the product and whether they would be interested to purchase the same or not.

Authors in [11] used MAPE (Mean Absolute Percentage Error) to measure the accuracy of the forecast based on the ARIMA-RNN model. MAPE is used as it compares different models and signifies if one is better than the other based on the deviation of predicted values from the observed values. The smaller is the MAPE, the more accurate is the model.

$$MAPE = \left(\frac{|\,(\boldsymbol{Actual\,Value} - \boldsymbol{Predicted\,Value})\,|}{\boldsymbol{Actual\,Value}} \right) * 100$$

Another study [12] uses a combination of ARIMA (Autoregressive Integrated Moving Average), LSTM (Long short-term memory) models, and concepts of wavelet de-noising on Hydrological time series data. Hydrological time series data is data related to water flow and mainly used in water management studies. This data is not stationary in nature and also contains lots of noise making it very difficult to forecast. The proposed model in [12] first uses the wavelet de-noising to eliminate the noise from the data. Then this de-noised data is fit to the model using ARIMA and subsequently forecasted. The result if ARIMA produces some residuals which are used to train the LSTM network. Results of the LSTM network are then used to fine-tune the forecasts by the ARIMA model.

3. Mathematical Model for K-Means Clustering

K-means clustering is a method of forming a specific number of clusters (k) represented by their centroids. Based on the k (no. of clusters required) centroids, each point in the data is segregated into a specified cluster. Clustering is done based on the distance between the data point and the centroid. We do this until all the data points are part of a cluster.

Input: k (number of clusters), D (set of lift ratios)
Output: a set of k subgroups (clusters)
Method: Randomly choose k as initial centroids;
Repeat: Assign each object to the cluster centroid based on the similarity features i.e., the mean value of objects in the cluster and update the cluster means
Until no change;

*3.1. **Data in Euclidean Space:** The quality of a cluster is measured using Sum of Squared Error (SSE). It is nothing but the Summation of the Euclidean distances between the centroid of the cluster and the data points. It is also known as scatter.*

*3.2. **Cohesion**: The main aim is to increase the similarity between the centroid and the data points in the cluster. For the objective function the centroid is the mean. This ensures that the algorithm is not limited to the Euclidean space. This is called Cohesion.*

*3.3. **Time and Space Complexity** - The space complexity of the K-means algorithm is BigO((d+K)n), where d are the data points and n are the attributes. Time taken to run the algorithm is linear in terms of data points in the cluster. The Time Complexity is BigO(i*K*d*n), where i are the no. of iterations.*

*3.4. **Evaluation Criteria** - For clusters K=3 and K=5 in [10] we compare the Recency values with Frequency and Monetary values of the sales. Comparing the clusters, we find which set of customers have the highest RFM value.*

4. Mathematical Model for LSTM Model

A typical LSTM network contains memory blocks which are called cells. There are two states which are transferred to the cell next to it. These two states are the hidden state and the cell state. The cell state holds the main flow of data and it allows an unchanged flow of data forward. However, some minor linear transformations may take place. A sigmoid gate is used to add or remove data from the state in the chain. A gate contains different individual weights. LSTM models avoid the long-term dependency issues. As discussed, before it uses sigmoid gates to deal with the memorizing process. [15]

An LSTM network is initialized by getting rid of the unwanted data in the first step. This Sigmoid Function decides which data to remove. It takes the result of the previous unit (h_{t-1}) at time $t - 1$ and the input (X_t) at time t. The forget gate (f_t) of the sigmoid function determines which of the previous outputs needs to be omitted relative to the number in C_{t-1} state. [15]

$$f_t = \sigma\ (W_f [h_{t-1}, X_t] + b_f)$$

where σ is the sigmoid function, W_f is the weight matrices and b_f are the bias.

The next step is to select and store data in the next input (X_t) and then update the cell state. One of the two layers decides whether or not to update (0 or 1) the new information. The other one is the tanh function which assigns weight (-1 to 1) to the value which passed by. The product of two is updated in the new cell. This result is added to the previous memory (C_{t-1}) which updates to C_t.

$$i_t = \sigma\ (W_i [h_{t-1}, X_t] + b_i),$$
$$N_t - tanh\ (W_n [h_{t-1}, X_t] + b_n),$$
$$C_t = C_{t-1} f_t + N_t i_t$$

where, C_{t-1}, C_t are the cell states, $t - 1$, t is time at cell states C_{t-1} and C_t respectively, W is the weight matrices and b is the bias of the cell state.

The last step shows the filtered version on the output (h_t) based on the state (O_t). Here initially the sigmoid function finalized the part of cell state which subsequently becomes output and then the sigmoid gate (O_t) and the result if tanh function from cell state (C_t) is multiplied. The resulting value is between -1 and 1.

$$O_t = \sigma\ (W_o [h_{t-1}, X_t] + b_o),$$
$$h_t = O_t \tanh(C_t)$$

where, W_o is the weight matrices and b_o is bias of the output gate.

4.1. ***Evaluation Criteria*** *– The performance of the model is evaluated by statistical methods. NSE and RMSE are popularly used to evaluate predicted and observed values.* ***The Nash-Sutcliffe Efficiency (NSE)*** *measures the magnitude of predicted data variance compared to measured values. NSE values range from −∞ to 1.* ***The Root Mean Square Error (RMSE)*** *is the error-index that shows how closely the predicted and observed values match, based on the range of data. Lower is the RMSE better if the fit of the model and zero implies a perfect fit. An LSTM model gives highly reliable results with NSE values close to 1 and a minimal value of RMSE.*

$$NSE = \left(1 - \frac{\sum_{i=1}^{n} (O_i - P_i)^2}{\sum_{i=1}^{n} \left(O_i - \underline{O_i}\right)^2}\right) * 100 \qquad RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (O_i - P_i)^2}$$

where, O_i is the observed discharges, P_i is the simulated discharges at time t, n is the total number of observations.

5. Results and Discussion

The study in [9] shows that algorithms like Linear SVM, Logistic Regression, Neural Network, AdaBoost, and Naïve Bayes showed classifier accuracy of more than 80% as shown in Table 1.

TABLE 1. CLASSIFIERS ACCURACY [9]

Classifier	Success Ratio	Classifier	Success Ratio
Linear SVM	84%	K-Nearest Neighbour	78%
Logistic Regression	84%	RBF SVM	75%
Neural Network	84%	Decision Tree	74%
AdaBoost	84%	Random Forest	74%
Naive Bayes	84%	Quad. Discriminant Analysis	43%

Similarly, work in [10] shows the RFM log on two different values of K (using k-means clustering) and presents an optimal way of getting the result as shown in Table 2. The authors in [10] took reference of the Silhouette Score (It is a metric used to calculate the goodness of a cluster. Its value ranges from -1 to 1.) to present the most optimal results.

TABLE 2. SILHOUETTE SCORE [10]

K = 3				K = 5			
#Clusters	Recency log	Frequency log	Monetary log	#Clusters	Recency log	Frequency log	Monetary log
0	161.1914	16.7619	291.8525	0	210.331	10.2652	170.5903
1	11.3732	209.3714	5316.8004	1	20.3099	95.2432	1913.8276
2	20.323	48.8775	894.3214	2	10.7796	280.0861	7456.3908
				3	81.2908	32.5154	619.0144
				4	3.8476	41.6503	708.6397
Silhouette Score for K= 3 is 0.3621				Silhouette Score for K= 5 is 0.3491			

The experiment carried out by authors in [11] used the Mac Revenue dataset consisting of 6757 entries. They divided the data into various batches to train and test the models. The experiments showed that on certain dates in the data ARIMA model performed better and on others RNN models. However, the proposed hybrid model outperformed them all with a minimum Average MAPE value of 3.7354 compared to ARIMA (6.3762) and RNN (9.2642) models as shown in Table 3.

TABLE 3. MODEL ACCURACY TESTING [11]

Model	Mean MAPE Value
Proposed Hybrid RNN Model	3.7354
RNN Model	9.2642
ARIMA Model	6.3762

As the study in [11], here also authors have used MAPE along with MSE (Mean Square Error) to measure the accuracy of the forecast. The experiments in [12] show that the predicted values by the proposed model based on de-noising-ARIMA-LSTM are significantly close to the observed values. Table 4. shows the results of MAPE and MSE and we can see that the values of the proposed model are fairly smaller compared to standalone ARIMA and LSTM models.

TABLE 4. RESULT COMPARISON [12]

Model	MSE	MAPE
de-noising-ARIMA-LSTM	0.0044	0.51%
ARIMA	0.0178	0.72%
LSTM	0.0049	0.56%

There is a need to understand customer behavior and apply the same not only in the marketing segment but also to accurately predict the sales and gain real insight. The research related to sales prediction is very old and is going on for years now. The traditional approaches do not fit right in the cross functional analysis. Considering the growing trend of social media, the customers are more informed of the offerings provided and they have a lot of options to choose from for any given product. Hence there is a need for the companies to incorporate customer analysis within their sales forecast which is where the traditional approaches fail.

6. Proposed Approach

We propose a hybrid approach to traditional sales forecasting, the whole process of prediction is going to be customer-centric. Figure 1 following are the steps to implement the proposed plan:

Step 1: Customer Segmentation

Pre-process and aggregate the data at the customer level. Build RFM Features for each customer. Apply K-means clustering to identify different groups (clusters) and evaluate the nature of each group. Calculate RM score and sort customers. Visualize the results and explore some key numbers.

Step 2: Segment Selection

The main aim is to identify the segments which impact the sales and not include values for the sales of average churn customers. Based on the RMF score and we would define the segments which have relevance to the sales prediction. Since Low RMF customers are likely to churn their sales value has no impact on the prediction hence we must discard them from subsequent predictions.

Step 3: Sales Prediction

Once we have removed the unwanted customer segments, we shall do the time series analysis on the remaining customers using appropriate ARIMA or LSTM model. In the end, all the independent values are consolidated to get the final predicted values.

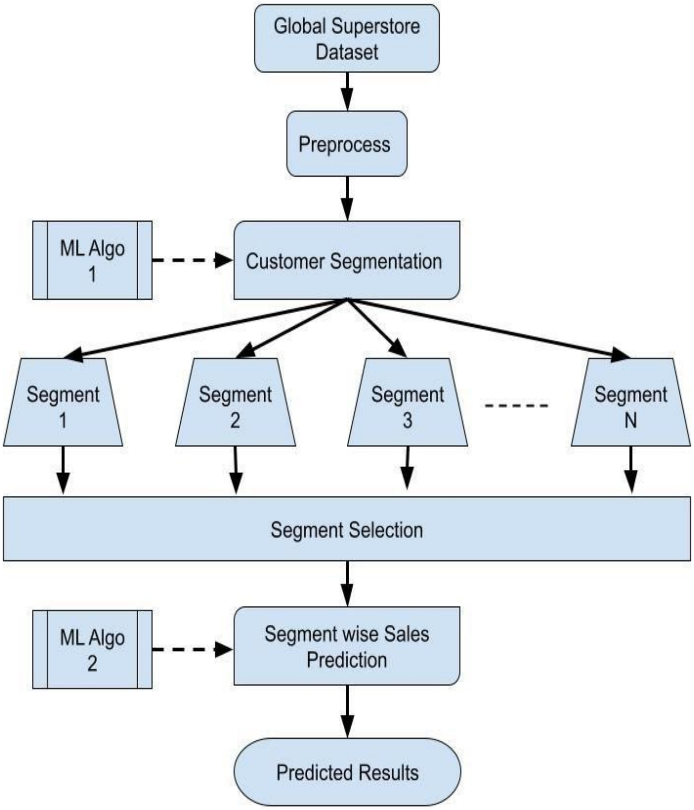

Figure 1. Proposed Model

7. Conclusion

We saw there are many fundamental studies that apply the concepts of Business Intelligence using various AI and machine learning techniques. We saw many algorithms to get concrete results. Many studies are available which presents different ways of segmenting customer behavior, but none to blend the idea for predicting sales. There is a need to understand customer behavior and apply the same not only in the marketing segment but also to accurately predict the sales and gain real insight. The survey work intends to present the analysis in a simplistic way to fuel the growth of any organization by applying the predictive approach. It will be a combination of programming, data analysis, and machine learning concepts; from the customer standpoint.

References

[1] J. Su and Y. Tang, "Business Intelligence Revisited," 2017 Fifth International Conference on Advanced Cloud and Big Data (CBD), 2017, pp. 1-6, doi: 10.1109/CBD.2017.9.

[2] S. Atul Khedkar and S. K. Shinde, "Customer Review Analytics for Business Intelligence," 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2018, pp. 1-5, doi: 10.1109/ICCIC.2018.8782305.
[3] Customer Analysis | Alameda SBDC (2020). Available at: https://www.acsbdc.org/resources/small-business-topics/marketing/customer-analysis (Accessed: 12 December 2020).
[4] Customer Analysis | CMG Consulting (2020). Available at: https://www.cmgconsulting.com/post/what-is-customer-analysis (Accessed: 12 December 2020).
[5] McColl-Kennedy JR, Zaki M, Lemon KN, Urmetzer F, Neely A. Gaining Customer Experience Insights That Matter. Journal of Service Research. 2019;22(1):8-26. doi:10.1177/1094670518812182
[6] S. Khan, M. R. Qader, T. K and S. Abimannan, "Analysis of Business Intelligence Impact on Organizational Performance," 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), 2020, pp. 1-4, doi: 10.1109/ICDABI51230.2020.9325610.
[7] Mentzer, J.T. and Cox, J.E., JR. (1984), Familiarity, application, and performance of sales forecasting techniques. J. Forecast., 3: 27-36. https://doi.org/10.1002/for.3980030104
[8] Insights, S. (2020) Predictive Analytics: What it is and why it matters, Sas.com. Available at: https://www.sas.com/en_in/insights/analytics/predictive-analytics.html (Accessed: 14 December 2020).
[9] R. Valero-Fernandez, D. J. Collins, K. P. Lam, C. Rigby and J. Bailey, "Towards Accurate Predictions of Customer Purchasing Patterns," 2017 IEEE International Conference on Computer and Information Technology (CIT), Helsinki, 2017, pp. 157-161, doi: 10.1109/CIT.2017.58.
[10] P. Anitha, Malini M. Patil, "RFM model for customer purchase behavior using K-Means algorithm", Journal of King Saud University - Computer and Information Sciences, 2019, ISSN 1319-1578, https://doi.org/10.1016/j.jksuci.2019.12.011.
[11] R. G. Hiranya Pemathilake, S. P. Karunathilake, J. L. Achira Jeewaka Shamal and G. U. Ganegoda, "Sales Forecasting Based on AutoRegressive Integrated Moving Average and Recurrent Neural Network Hybrid Model," 2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Huangshan, China, 2018, pp. 27-33, doi: 10.1109/FSKD.2018.8686936.
[12] Z. Wang and Y. Lou, "Hydrological time series forecast model based on wavelet de-noising and ARIMA-LSTM," 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Chengdu, China, 2019, pp. 1697-1701, doi: 10.1109/ITNEC.2019.8729441
[13] Verhoef, P. C., & Donkers, B. (2001). Predicting customer potential value an application in the insurance industry. Decision support systems, 32(2), 189-199.
[14] K. T. Ratnayaka, K. N. Seneviratne, W. Jianguo and H. I. Arumawadu, "A Hybrid Statistical Approach for Stock Market Forecasting Based on Artificial Neural Network and ARIMA Time Series Models".
[15] Le, Xuan-Hien; Ho, Hung V.; Lee, Giha; Jung, Sungho. 2019. "Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting." Water 11, no. 7: 1387.
[16] Statistical tree-based feature vector for content-based image retrieval ,Sushila Aghav-Palwe; Dhirendra Mishra, IJCSE 2020,21(4)