



# Customer-Centric Sales Forecasting Model: RFM-ARIMA Approach

**Sanket Tanaji Londhe, Sushila Palwe**

*School of Computer Engineering and Technology, MIT-World Peace University, Pune, Maharashtra, India*

## Abstract

**Background:** Decision makers use the process of determining the best course of action by processing, analysing & interpreting the data to gain insights, known as Business Intelligence. Some decision support systems use sales figures to predict future expansion, but few consider the effect of customer data. **Objectives:** The main objective of this study is to build a model that will give a forecast based on fine-tuned sales numbers using some customer-centric features. **Methods/Approach:** We first use the RFM model to segment the customers into distinct segments based on customer buying characteristics and then discard the segments that are irrelevant to the business. Then we use the ARIMA model to do the sales forecasting for the remainder of the data. **Results:** Using this model, we were able to achieve a better fitment of the data for the prediction model and achieved a better accuracy when used after RFM analysis. **Conclusions:** We tried to merge two different concepts to do a cross-functional analysis for better decision-making. We were able to present the RFM-ARIMA model as a better metric or approach to fine-tune the sales analysis.

**Keywords:** Business Intelligence; Customer Analysis; Sales Forecasting; Exploratory Analysis; Segmentation; Decision Support System; Recency, Frequency & Monetary Value (RFM); Auto-Regressive Integrated Moving Averages (ARIMA); Long short-term memory (LSTM)

**JEL classification:** C53, E30, E37

**Paper type:** Research paper

**Received:** 29 Dec 2021

**Accepted:** 08 May 2022

**Citation:** Londhe, S.T., Palwe, S. (2022), "Customer-Centric Sales Forecasting Model: RFM-ARIMA Approach", Business Systems Research, Vol. 13 No. 1, pp.35-45.

**DOI:** 10.2478/bsrj-2022-0003

## Introduction

In today's data-driven world, it is almost impossible for businesses to operate without an intelligent strategy. With so many data sources accessible, organisations can now leverage data to make informed business decisions, predict future outcomes, and take actionable steps toward reaching their strategic goals. Operating a successful business requires an organisation to have constant visibility into its business performance and the ability to measure key metrics that indicate how well they are doing. Business intelligence (BI) is an analytics-driven process that enables companies to monitor performance indicators. It identifies opportunities for growth by collecting, storing, analysing, and reporting data from various sources (Tomičić Furjan et al., 2020). The digital age has transformed the way businesses operate and compete, as well as how they measure success. In this new era of constant information, organisations must adapt their operations to capture data and drive measurable results. The effective management of operational processes optimises performance based on efficiency and effectiveness. The BI domain has grown to include various activities and processes that increase system performance (Akbar et al., 2020). Data visualisation techniques are critical for analysts to see and make sense of data. These techniques help them extract insights from raw data, construct charts and diagrams, create interactive visualisations, and share the insights with others. Recently, Business Intelligence has been used to bump up sales revenue and improve data security, understand customer purchase intent, build various reach-out strategies, and more (Vinod, 2008). The most common process which helps organisations make informed decisions includes ideas and processes to boost profits, understanding consumer behaviour, competitor analysis, performance analysis, data and process optimisation, success prediction, identifying trends and activities, identifying gaps and taking corrective measures. As companies use more and more data to run their business, they are working on shared data strategies and joint data projects. A business intelligence framework becomes an important part of an enterprise-wide business intelligence strategy (Calzon, 2020).

## *Problem Statement & Hypothesis*

Analysts use the power of BI to improve results and establish competitive benchmarks that make companies more efficient. Analysts use BI to make informed decisions using historical data and its impact on current operations. Analysts leverage the power of BI to improve results and set competitive benchmarks that make companies run smoother. Companies rely only on a small subset of data acquired through customer purchases/transactions (Mirza et al., 2022). This data is primarily related to the products sold and the sales value over time. Any machine learning technique applied to this subset of data may provide some useful insights into the future, but it is not enough. The insights from such analysis are very generalised if a company is catering to a diverse audience. To deal with this issue, it is important to identify the nature of your customers; if the same sales forecasting can be done based on the nature of the diverse customer base, the companies can gain a lot.

We believe that any business has different types of customers with different requirements. They purchase goods at a certain frequency, at different periods, and spend varying amounts based on the offerings and requirements. Over time we can identify various segments of the customers based on this Recency, Frequency and Monetary values and identify these customers. We also may identify some customers who no longer are buying and have churned; these customers act as noise to the whole data and needs to be removed before doing the sales analysis on the rest of the segments.

## Customer Analytics

Customer Analytics is the process of mining consumer data for profiling and target market analysis (Pejic Bach et al., 2021). Forbes magazine reported that 81% of companies rely on customer analytics to improve their customer knowledge. These insights help organisations create loyalty programs and expand their business (R. Valero et al., 2017). This analysis, in a way, ensures that customers' specific needs are met or even exceed those needs. These strategies not only help organisations get to know their customer but also help acquire new ones. Companies study why people buy certain products, how they buy them and in what various ways they make the purchase, how frequently they buy their products, and what influences them to buy them (Khedkar et al., 2018). The following reasons stand out as relevant for customer analytics:

- Social media makes consumers more informed, demanding, and tuned to everybody's opinions. Companies need to be up to date with what their customers want and should be in line with the social trends at a given time (Istrefi-Jahja et al., 2020).
- Increase response rates by contacting the right customers. Reduce campaign costs by targeting a suitable audience and sending appropriate marketing messages by segregating the consumers efficiently. Such strategies provide a better insight into the target audience (CMG Consulting, 2020).

The study article by (Anitha et al., 2022) understands customer buying patterns and behaviours using RFM (recency, frequency, and monetary Value) models for segmentation with k-means algorithms. This work presents the RFM protocol for two different values of K (using k-means clustering) and presents the best way to obtain the results using Silhouette Score ("It is a metric used to calculate the goodness of a cluster. Its value ranges from -1 to 1"). In their study, the Silhouette Score for K= 3 was 0.3621, while for K=5, it was 0.3491.

## Sales Forecasting

Sales forecasting is the process of predicting sales growth in the future using relevant data. Sales forecasting usually uses historic sales data and economic trends of sales. The sales forecasting model allows organisations to predict short-term and long-term performance (Vijayaraghavan, 2019). Time series forecasting is one of the core concepts in making forecasts and analysing trends based on past values. There are different techniques for making predictions based on the nature of the data (Le, 2019). Sales forecasting enables organisations to manage resources like inventory, billings, and workforce. It also helps companies allocate the right resources to the right tasks without creating an overhead on the budget (Skyword Staff, 2020). Sales forecasting allows companies to predict sales growth and help set an achievable benchmark; effectively allocate company resources; Build an action plan for the coming future. The following reasons stand out the importance of predictive analytics:

- Predictive Analytics offers businesses a unique opportunity to identify various insights into the data and use these patterns to build a plan to act upon them. The surge of data demands more AI and machine learning approaches, which would offer organisations valuable insights (Sas, 2020).
- Some predictive techniques identify customer responses to a service and their purchase behaviour. Such studies help build offers around the products and services and promote cross-sell opportunities (Sas, 2020).

A paper by (Hu et al., 2020) used a combination of the LSTM (long short-term memory) model, ARIMA (autoregressive integrated moving average) model, and wavelet denoising on hydrological time series data. The Hydrological data contains values related to water flow, and this data is nonstationary and contains a lot of noise. This noise makes it very difficult to process and subsequently predict. The authors (Hu et al., 2020) first denoise the data using wavelet denoising, then fitted it to a model using

his ARIMA to make predictions. An LSTM network is trained using the residual values from the ARIMA results. They then used this as feedback to improve the predictions of the ARIMA model. The results in the paper (Hu et al., 2020) show that the predicted values of the proposed model based on Denoising-ARIMA-LSTM are fairly close to the observed values.

### The goal of the paper

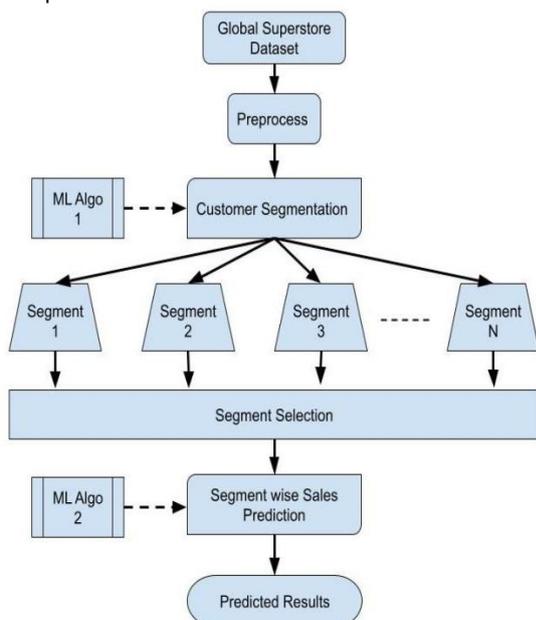
In this paper, we perform a set of experiments driven by business intelligence concepts discussed above which will help organisations get a realistic idea regarding their business growth. This two-step process includes segmentation of customers who are relevant to the business growth using the RFM model and then forecasting the growth based on the sales numbers of these segmented customers using the ARIMA model. This will not only help organisations identify the customers who are going to churn and technically act as noise, but they will also get a realistic view of the growth of their sales figures.

## Methodology

### Hybrid Customer-centric Approach

We implemented a hybrid multi-step model based on traditional sales forecasting models, aiming to keep the whole prediction model customer-centric. We divided the work into two key parts 1. Customer Analysis using RFM Model and 2. Sales Forecasting using ARIMA Model. For this study, we use Tableau's Global Superstore data set. These extensive datasets provide a basis for exploring data processing details and examples for performing data transformations. The dataset contains data on the sales of multiple products a business sells along with additional features like customer segments, categories of products, purchase geographies, revenue generated, profit made, etc. The data is also customer-centric, which is suitable given the context of the problem we intend to solve. The data contains details of the orders by each customer between the years 2011 through 2015. Fig 1. Shows the implementation approach:

Figure 1  
Proposed Model



Source: Author's illustration

### Customer Analysis using RFM Model

RFM is mainly used for campaign targeting to reach out to relevant customers. These customers are segregated from the data based on recency, frequency and monetary values (RFM). RFM analysis helps understand the type of customer and helps organisations answer questions like who are my top paying customers? Who is about to churn? Who must the business retain?

Using the transactional data to devise a metric that will divide the customer into certain groups based on their current purchase behaviour and potential Value to the business.

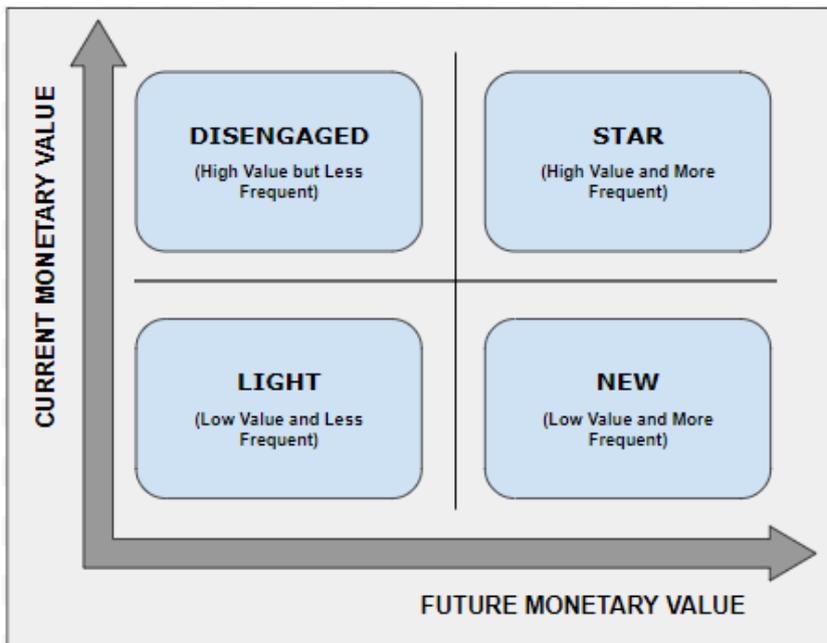
The RFM scores are calculated using the quantile regression model. The quantile regression equation is given as:

$$u_i = v_i \beta_q + c_i \tag{1}$$

where  $\beta_q$  is vector associated with  $q^{th}$  quantile ( $0 < q < 1$ ),  $c_i$  is the prediction error.

Based on the RFM scores, the customers are segmented into four categories per their current and potential values (Figure 2).

Figure 2  
Customer Segmentation Matrix



Source: Adapted from Smartkarrot.com (2020)

We pre-processed the data and aggregated the same at the customer level. We then built each customer's RFM (Recency, Frequency & Monetary-value) features. We used the 80% quantile regression model for Recency and Monetary-value to automate the segmentation. We then calculated the RM score and sorted the customers. Finally, we visualised the results to explore some key numbers.

Figure 3 shows the Value matrix by (a) Avg. Monetary Value, (b) Number of Customers and (c) Recency.

Figure 4 shows a Segment interpretation of the same.

Figure 3  
Customer Segmentation Matrix

		R				R				R	
		1	2			1	2			1	2
M	2	2189	2378	M	2	423	3061	M	2	1136	301
	1	279	330		1	3059	10872		1	1147	360
<b>(a) Avg. Monetary Value</b>				<b>(b) Number of Customers</b>				<b>(c) Recency</b>			

Source: Author's illustration

Figure 4  
Value Matrix Interpretation

Disengaged		Star	
\$ Value	2189	\$ Value	2378
# Cust	423	# Cust	3061
# Days	1136	# Days	301
Light		New	
\$ Value	279	\$ Value	330
# Cust	3059	# Cust	10872
# Days	1147	# Days	360

Source: Author's illustration

Based on the combination of RM Scores calculated, we have divided customers into 4 groups.

- **Disengaged** – There are 423 customers with an average dollar value of \$2189 and have last purchased approximately 1136 days ago. These customers are highly valued but have purchased less frequently.
- **Star** – There are 3061 customers with an average dollar value of \$2378 who purchased approximately 301 days ago. These customers are highly valued and frequent.
- **New** – There are 10872 customers with an average dollar value of \$330 who purchased approximately 360 days ago. These customers are relatively new and have started to make purchases. Low dollar value but frequent purchasers.
- **Light** – There are 3059 customers with an average dollar value of \$279 who purchased approximately 1147 days ago. These customers are low-valued and less frequent.

We observed that a subset of Light customers with low RM value (R=1 & M=1) is likely to churn. We will discard these values before we move on to forecasting, as these customers made their last purchase long ago. These values are likely to skew our analysis and must be processed from the forecast.

### Sales Prediction using ARIMA model

Once we had removed the Light/churn-able customers, we moved on to the next step and analysed the remaining customers independently. We used ARIMA to do the time series forecasting. ARIMA is a time-series forecasting model that explicitly works well with nonstationary time series. Auto-Regressive Integrated Moving Averages (ARIMA) has three key terms: **AR** – Lags of series which is made stationary; **I** – Used to make the series stationary using the order of differencing; **MA** – Lag of the error of forecast.

Consider X, which refers to a time series. The expression for ARIMA can be given as:

$$X_t = c + e_t + \sum_{i=1}^n \varphi_i X_{t-i} + \sum_{i=1}^m \theta_i e_{t-i} \tag{2}$$

where  $X_t$  is the original time series,  $X_{t-i}$  is the lag  $i$  in the time series and  $e_{t-i}$  is the lag  $i$  forecast error.

## Results

Since we wanted to present the impact of customer data on sales prediction accuracy, we implemented the ARIMA Model on the data without pre-processing the data with the RFM model. Secondly, we implemented the ARIMA model after pre-processing it with the RFM model.

We compared the results in terms of data's feasibility and usefulness of features for predicting by looking at the Adjusted R-squared Value. Secondly, we compared the actual vs predicted sales values on visual plots.

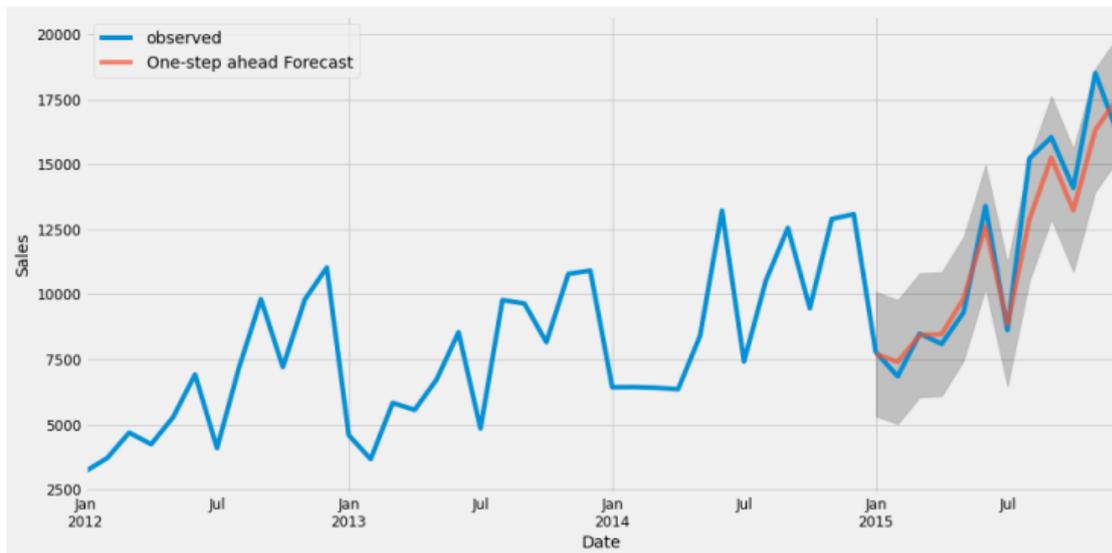
Table 1 shows the Adjusted R-squared values for two ARIMA models derived from OLS (Ordinary Least Square) values. The RFM-ARIMA model shows a better fit for the prediction compared to the standalone ARIMA.

Table 1  
Data Feasibility Comparison

Model	Adjusted RS
RFM-ARIMA model	0.71
ARIMA	0.67

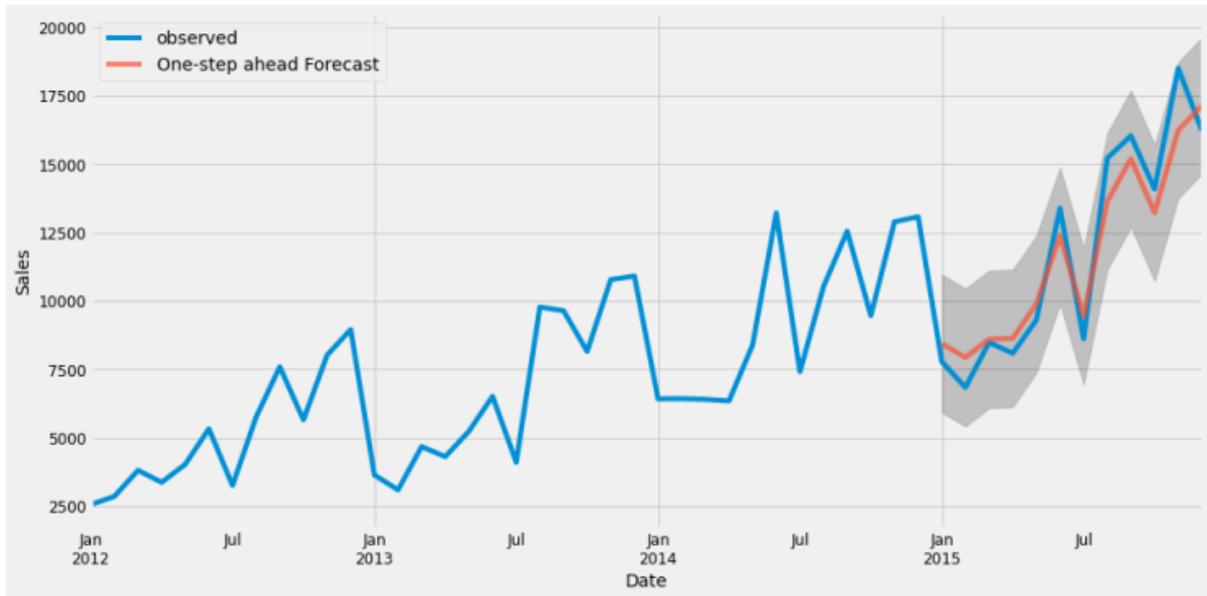
The following figures show the prediction output on visual plots. Figure 5a shows the sales prediction using the standalone ARIMA model. The model initially showed a good fit but then deviated a lot. However, in the RFM-ARIMA model, the model fit improved due to reduced noise by eliminating churned customers, as shown in Figure 5b.

Figure 5a  
Sales Prediction using standalone ARIMA Model



Source: Author's illustration

Figure 5b  
Sales Prediction using RFM-ARIMA Model



Source: Author's illustration

We evaluated the model's accuracy using statistical methods. **Root mean squared error (RMSE)** is an error metric that indicates how well the predicted and observed values match based on the data range. The lower the RMSE, the better the model fit, and RMSE is popularly used to evaluate predicted and observed values. **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.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2} \quad (3)$$

where  $O_i$  = observed discharges,  $P_i$  = simulated discharges at time  $t$ , and  $n$  = total number of observations. Table 2. shows the MSE and RMSE Values for the two models.

Table 2  
Model Accuracy Comparison

Model	MSE	RMSE
RFM-ARIMA model	1762925.07	1327.75
ARIMA	2219903.64	1498.93

Table 2 contains the MSE and RMSE estimates for forecast models. As can be observed, the RFM-ARIMA model has a lower prediction error rate than the standalone His ARIMA model, confirming the higher efficiency of the RFM-ARIMA model than the generalised model for revenue forecasting. Another observation from the Figures 5(a) and 5(b) is that, for the RFM-ARIMA model, the forecast values are fairly close to the actual values. And towards the end, the predicted model is inclined toward the actual Value for the RFM-ARIMA model while it is deviating in the case of the standalone ARIMA model, which is a positive sign.

## Conclusion

### *Summary of results*

This research merges two concepts to do a cross-functional analysis for better decision-making. Our Customer-centric Sales Forecasting model using a combination of RFM and ARIMA model was able to provide results with better accuracy, as shown by the RMSE Value of 1327.75 compared to 1498.93 for the standalone ARIMA model.

The analysis of actual and predicted data shows that implementing the RFM model on customer data before Sales Forecasting using ARIMA model is a better approach to identify business gaps and create an actionable plan. We were able to present the RFM model as a better metric or approach to fine-tune the sales analysis. The main advantage of using the RFM model is that it not only gives an insight into the customer behaviour towards organisation offerings but also identifies the churnable customers. They should be discarded from the data as they act as noise. This, in turn, reduces the data's size and subsequently improves the model's performance. While the scope of this research is limited to a small amount of data with only marginal improvements in the results, future studies can expand this scope. This research highlights the importance of understanding consumer purchase intent and applying the same not only to market to the target audience but also to predict sales revenue accurately and gain important insights.

Results show that the Hybrid Customer-centric approach for Sales Forecasting is more refined and targeted than standalone Sales Forecasting. Even the change is marginal but is significant in terms of the value-added insights from the RFM analysis. Results also indicate that the RFM-ARIMA model is more accurate. The overall observation indicates that this Hybrid approach has some promising scope in BI applications.

### *Implications*

In this research, we used two models, RFM and ARIMA models, and applied them in series to get the results. The RFM model can give companies insight into their customer behaviour and how they react to various offerings, and this model indirectly provides customer purchase efficiency. This approach allows businesses to understand how diverse their customers are and can even target each of them through targeted marketing campaigns and making their products and services more lucrative. Conversely, the ARIMA model can help companies see growth opportunities. Models like ARIMA help businesses understand what is working for them and what isn't. These two models together form a business intelligence framework for customer-centric sales analysis to get actionable insights.

### *Limitations & Scope for Future Work*

The specificity of the model itself is one of its limitations. There are many techniques which may prove to give a better analysis. This research also lacks concepts of Artificial Intelligence. While we can work with Machine Learning models on Time Series data, AI can bring a lot on how the data is structured and provide some hidden insights that could be a game changer. It can also help automate the process of predicting how businesses grow in different areas. Hence, there is a lot to be explored in this area.

## References

1. Akbar, R., Silvana, M., Hersyah, M. H., Jannah, M. (2020), "Implementation of Business Intelligence for Sales Data Management Using Interactive Dashboard Visualization in XYZ Stores", in 2020 International Conference on Information Technology Systems and Innovation (ICITSI), pp. 242-249.
2. Anitha, P., Patil, M. (2022), "RFM model for Customer Purchase Behavior using K-Means Algorithm", Journal of King Saud University - Computer and Information Sciences, Vol. 34, No. 5, pp. 1785-1792.
3. Calzon, B. (2021), "See Top 10 Analytics & Business Intelligence Trends For 2021", available at <https://www.datapine.com/blog/business-intelligence-trends/> (12 December 2020)
4. CMG Consulting (2020), "What is Customer Analysis? Explaining one of the critical components of any business plan", available at <https://www.cmgconsulting.com/post/what-is-customer-analysis> (12 December 2020)
5. Hu, X., Shi, Z., Yang, Y., Chen, L. (2020), "Classification Method of Internet Catering Customer Based on Improved RFM Model and Cluster Analysis", in 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), pp. 28-31.
6. Istrefi-Jahja, A., Zeqiri, J. (2021), "The Impact of Digital Marketing and Digital Transformation on Brand Promotion and Brand Positioning in Kosovo's Enterprises", in IRENET - Society for Advancing Innovation and Research in Economy, Zagreb, pp. 244-255,
7. Khedkar, S. A., Shinde, S. K. (2018), "Customer Review Analytics for Business Intelligence", in 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICIC), 2018, pp. 1-5.
8. Le, X. H., Hung, V. H., Giha, L., Sungho, J. (2019), "Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting", Water, Vol. 11 No. 7, pp. 1-19.
9. Mirza, S., Mahmood, A., Waqar, H. (2022). "The interplay of open innovation and strategic innovation: Unpacking the role of organisational learning ability and absorptive capacity", International Journal of Engineering Business Management, Vol. 14, pp. 1-15.
10. Pejic Bach, M., Pivar, J., Jaković, B. (2021), "Churn Management in Telecommunications: Hybrid Approach Using Cluster Analysis and Decision Trees", Journal of Risk and Financial Management, Vol. 14 No. 11, pp. 1-25.
11. Sas (2020), "Predictive Analytics: What it is and why it matters", available at [https://www.sas.com/en\\_in/insights/analytics/predictive-analytics.html](https://www.sas.com/en_in/insights/analytics/predictive-analytics.html) (14 December 2020)
12. Skyword Staff (2020), "Sales Forecasting", available at <https://www.skyword.com/marketing-dictionary/sales-forecasting/> (14 December 2020)
13. Smarkarrot.com (2020), "Win the new game of Intelligent Customer Success", available at <https://www.smartkarrot.com/> (14 December 2020).
14. Tomičić Furjan, M., Tomičić-Pupek, K., Pihir, I. (2020), "Understanding digital transformation initiatives: case studies analysis", Business Systems Research, Vol. 11 No. 1, pp.125-141.
15. Valero, R., Collins, D., Rigby, C., Bailey, J. (2017), "Towards Accurate Predictions of Customer Purchasing Patterns", in IEEE International Conference on Computer and Information Technology (CIT), Helsinki, 2017, pp. 157-161.
16. Vijayaraghavan (2019), "How Sales Forecasting can help your online Business Growth?", available at <https://productimize.com/blog/how-sales-forecasting-can-help-your-online-business-growth/#:~:text=Sales%20forecasting%20is%20the%20process,wide%20comparisons%2C%20and%20economic%20trends> (14 December 2020)
17. Vinod, B. (2008), "The continuing evolution: Customer-centric revenue management", Journal of Revenue and Pricing Management, Vol. 7, pp. 27-39.

## About the authors

Sanket Londhe is an MIT-World Peace University student and an analytics enthusiast with around 4 years of experience working in a private analytics company. He is passionate about driving insights from both measurable and non-measurable data. He contributes to the efficiency of the Business Operations by delivering visual insights and analysing the team's engagement with the customers. The author can be contacted at [sanketlondhe99@gmail.com](mailto:sanketlondhe99@gmail.com)

Sushila Palwe is a PhD in Computer Engineering with more than 18 years of teaching experience. She is a faculty member at the Department of Computer Science Engineering at MIT- World Peace University. She specialises in teaching Machine Learning, Databases, Data Science, Data Mining, and Business Intelligence. On the research front, she received funding for a research project and international workshop from the Board of College and University Development, Pune. She supervised many undergraduate and postgraduate students and published research papers and book chapters in various reputed conferences, journals and books. The author can be contacted at [Sushila.Palwe@mitwpu.edu.in](mailto:Sushila.Palwe@mitwpu.edu.in)