Proceedings of the 11th World Congress on Mechanical, Chemical, and Material Engineering (MCM'25)
Barcelona, Spain - Paris, France - August, 2025
Paper No. ICMIE 128
DOI: 10.11159/icmie25.128

Implementation Of Machine Learning and BPM To Increase Profitability in A Travel Agency

Guillermo Bernardo Huapaya Villegas[1], Aiko Maria Makino Aguilar[2], Edilberto Avalos Ortecho[3]
[1,2,3]Carrera de Ingeniería Industrial, Universidad de Lima, Perú
Av. Javier Prado Este 4600 – Santiago de Surco, Lima, Perú
20194779@aloe.ulima.edu.pe; 20193026@aloe.ulima.edu.pe; eavalos@ulima.edu.pe

Abstract – This research addresses the low accuracy in monthly tourist projections for a travel agency, with a Mean Absolute Percentage Error (MAPE) of 29.49%, resulting in an opportunity cost of $3.9 million in the travel package purchasing process. A solution was implemented using Machine Learning (ML) techniques, Support Vector Regression - Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors (SVR-SARIMAX), and Business Process Management (BPM) to reduce the MAPE and improve the purchasing process by redesigning and adding a negotiation activity. The ML model was trained on historical data of international tourist arrivals from 2013 to 2023, and 2023 purchase records. Using Python and Arena software, the MAPE was reduced to 18.18%, with key performance indicators improving by 16%, reaching 19,907 customers, $20.7 million in revenue, and $1.8 million in opportunity cost. Furthermore, a conversion rate over 90% was achieved after implementing process improvements, demonstrating the effectiveness of ML and BPM in resource allocation and process enhancement.

***Keywords*:** Machine Learning (ML), Business Process Management (BPM), Support Vector Regression - Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors (SVR-SARIMAX), tourism.

1. Introduction

1.1. Overview and Objectives

The present study focuses on the application of Machine Learning (ML) and Business Process Management (BPM) tools to enhance profitability in a travel agency. Literature reviews demonstrate the use of ML for accurate tourist projections by leveraging seasonal patterns, exogenous variables, and analyzing factors that influence sector behavior [1-4]. Additionally, the literature supports the effectiveness of integrating ML with BPM simultaneously [5][6].

The aim of the research is to design a solution that integrates ML techniques (SVR-SARIMAX) for tourist projections and BPM to improve the travel package purchasing process. Eleven years of historical data are used to feed the Python model, along with process analysis to identify improvements through BPM. The study includes an introduction and justification of the problem, followed by the methodology to be used. Next, the results are presented, followed by a discussion to compare them with the literature and address encountered limitations. Finally, the conclusions aim to answer the question: Can ML and BPM tools improve profitability in a travel agency?

1.2. Problem and Justification

The tourism sector faces challenges in projecting tourist numbers, as behavior is influenced by factors such as seasonality, exogenous variables, and economic fluctuations. It is crucial for companies in the sector to make projections, as this allows for better resource allocation, service improvement, and increased profitability. Authors [7][8] consider a projection with a Mean Absolute Percentage Error (MAPE) lower than 10% to be optimal; however, this figure is difficult to achieve for companies in the sector.

In 2022, tourism represented 2.3% of Peru's GDP and over one million jobs, with strong links to hospitality and local production [9]. Digital platforms like Airbnb and TripAdvisor have increased demand for tourism services, reflecting a 25.7% increase in international tourist arrivals for 2023, equivalent to $2.6 million [10].

Within the same company, there is a projection with a MAPE of 29.49% for 2023 and an average of 44.14% over the past five years, reflecting a problem that results in opportunity costs of up to $3.9 million. The goal is to reduce the MAPE value to below 10% to decrease opportunity costs while simultaneously increasing the number of customers and total revenue in the process [17].

Table 1: MAPE value per year

Year	MAPE %	Year	MAPE %
2019	19.13	2022	34.57
2020	78.71	2023	29.49
2021	58.8	**MEAN**	**44.14**

1.3. Literature Review

Authors such as [1] use the SVR-SARIMA model to project the number of tourists in the Philippines, evaluating magnitude and seasonality. Additionally, the study by [2] employs SARIMAX to incorporate exogenous variables such as the calendar and holiday dates within SARIMA-FTSMC to capture seasonal patterns in train passengers in Indonesia.

Literature reviews such as [6] provide information on the use of ML within BPM applications throughout the process lifecycle, highlighting its potential to automate processes, reduce costs, and better allocate resources through anomaly detection and decision-making support.

Figure 2 uses the problem tree to define root causes and identify the tool to be used as a solution.

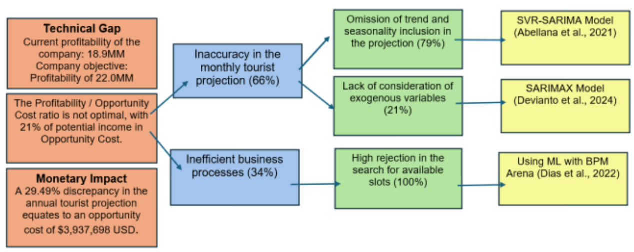

Fig. 1: Problem Tree

2. Methodology

2.1. Design and Scope of the Investigation

The scope of the research is correlational and explanatory, employing a pure experimental design with a quantitative approach. The data sources include 11 years (2013-2023) of historical data on international tourists coming to Peru, collected by the agency, along with purchase records for all packages in 2023. Unit of evaluation chosen is tourist numbers over packages for their consistency in analysis.

Figure 2 presents the design of the solution used. It includes inputs that reflect the reality of the travel agency, followed by the steps to be implemented and the metrics to be achieved once completed.

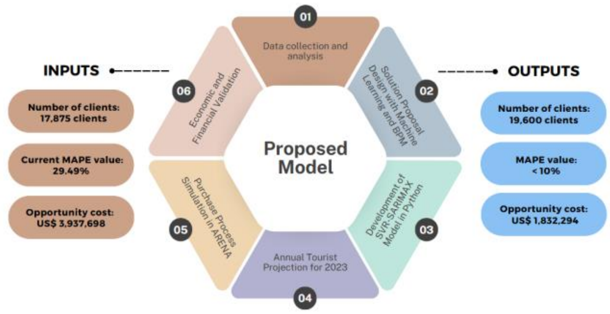

Fig. 2: Solution proposal design

2.2. Solution Evaluation

Studies support the use of ML for projections in the tourism sector. The authors [1] combine Support Vector Regression (SVR) with Seasonal Autoregressive Integrated Moving Average (SARIMA) to create SVR-SARIMA, leveraging the linear and seasonal modeling capabilities of SARIMA with the nonlinear modeling capabilities of SVR, achieving a MAPE value of 5.45%.

Another study by [11] combines SVR with ARIMA to capture nonlinear relationships with SVR and linear trends with with ARIMA, achieving a MAPE of 10.22%. The authors [2][3] demonstrate in their studies how incorporating exogenous exogenous variables into SARIMA is beneficial for capturing the impact of external variables. Specifically, [2] uses SARIMAX to project train passengers using holidays as an exogenous variable, achieving a MAPE of 5%.

Business Process Management (BPM) is used to improve the travel package purchasing process, aiming to increase the conversion rate. The authors [6] demonstrate that the integration of ML into BPM reduces operational errors by 15% and improves resource allocation by 25%, making the process more efficient and enhancing customer satisfaction.

2.3. Data Collection

The research uses two data sets provided by the travel agency. The first set is a collection of historical data on monthly tourists over a period of 11 years (2013-2023), making a total of 132 observations. The second set consists of purchase records for the entire year of 2023, with 567 observations, used to understand customer behavior during the process. Both data sets are complete and serve as the foundation for building the SVR-SARIMAX model and implementing BPM, allowing for a rigorous analysis of tourist trends and purchasing patterns.

2.4. Solution Development

The development of SVR-SARIMAX begins in Python, following the architecture built on the foundation used by [1] in their research. The process starts with uploading input data, including historical data on international inbound tourists and the data used as an exogenous variable, which in this research is the accumulated precipitation for the city of Cusco, as it is the destination for more than 80% of tourists visiting Peru. The authors [13] confirm that extreme precipitation negatively impacts tourism. The next step is the separation of the data into training and testing sets, following a 10:1 ratio, with 2023 as the test set. The number of years used to feed the model is considered appropriate according to [12].

Next, the SVR model is adjusted by removing the trend to generate residuals, which are then used to feed SARIMAX. Both SVR and SARIMAX make projections for 2023 of tourists and residuals, respectively, which are then combined to obtain a projection based on the SVR-SARIMAX hybrid model. Finally, the model is evaluated with the test set, and its accuracy is determined by calculating the MAPE value.

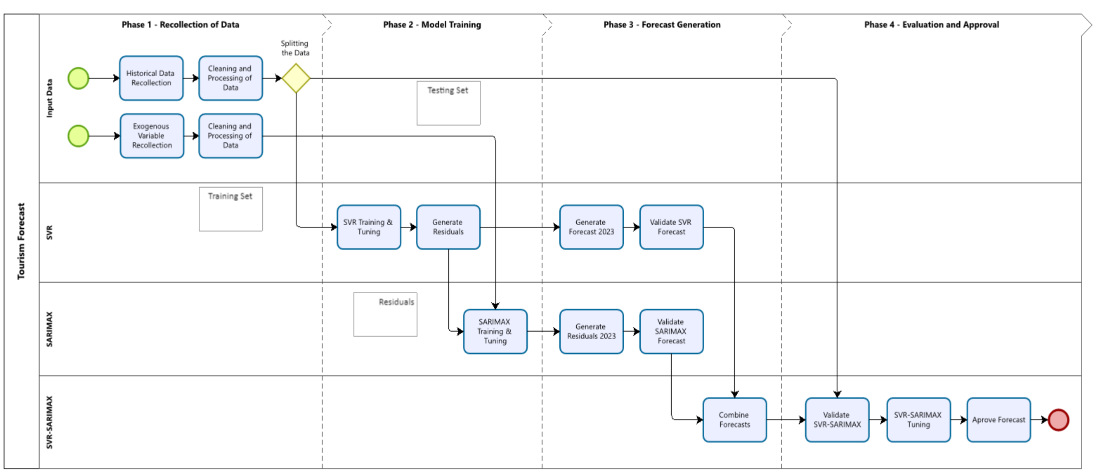

Fig. 3: SVR-SARIMAX Methodology design

Next is the application of BPM in the travel package purchasing process, starting with defining the entities to model, attributes, and variables. Arena software is used to simulate the process; the authors [14] recommend using Arena for its ability to build realistic simulations of processes, making analysis and improvement easier. The KPIs to be modeled are: total revenue from the process, number of customers, opportunity cost, and conversion rate.

Once the simulation is built, it is used to thoroughly analyze the process and identify areas for improvement [6]. on this analysis, the changes to be implemented in the process are defined, such as the inclusion of the SVR-SARIMAX results as variables for available slots per month and the addition of negotiation activities to increase the customer rate. Once implemented, the simulation is run again, and the KPIs are measured once more.

Finally, an evaluation of the results is conducted to confirm whether the changes and additions were beneficial to the process. Output Analyzer is used for the comparison of means and to analyze the results of both simulations. The authors [18] support the use of Output Analyzer for conducting analyses, as seen in their research on applying Lean principles to hospitals. Based on the evaluation, the effectiveness of the SVR-SARIMAX and BPM contribution to the process is validated.

3. Results

3.1. Forecast Results

The results of SVR-SARIMAX are the projection that the model has for 2023, which is compared with the actual data from 2023 to determine the model's accuracy with its MAPE value.

Figure 4 presents a graph that illustrates the behavior of SVR-SARIMAX and how it aligns with the sector's behavior to provide the 2023 projection. The forecast line successfully follows seasonal trends, especially during peak

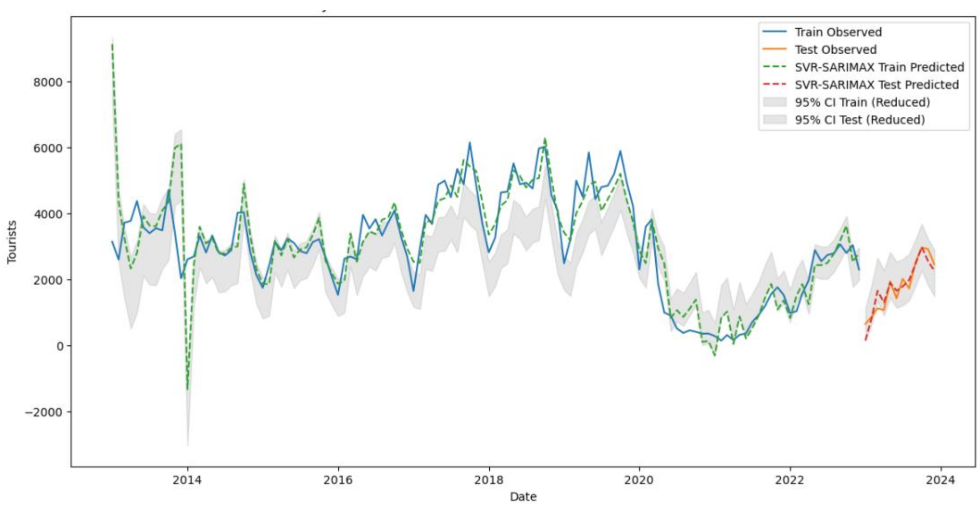

Fig. 4: SVR-SARIMAX Forecast

Table 2 presents the results obtained through the SVR-SARIMAX projection for 2023 alongside the actual numbers. The MAPE value of the projection is also provided at the end of the table.

Table 2: Monthly forecast results

Month	Real	Forecast	Month	Real	Forecast
January	642	141	July	2016	1782
February	871	841	August	1722	1977
March	1120	1651	September	2510	2461
April	1067	1295	October	2964	2977
May	1940	1897	November	2918	2552
June	1420	1653	December	2453	2259
				MAPE	**18.18%**

The projection made by SVR-SARIMAX shows an 11% improvement over the agency's current projection, reducing from 29.49% to 18.18%. The highest deviations were in January and July, possibly due to post-holiday behavior and rainfall variability in Cusco, respectively.

3.2. BPM Results

The implementation of BPM involves building a simulation of the agency's current travel package purchasing process to identify areas for improvement. The process design is created in Bizagi and simulated using Arena software.

Within the process, the variables of group size in the package, purchase month, and slot availability per month are determined. 100 replicas are run in the simulation, following the recommendation of [15], which states that the half-width should not exceed 5% of the average result.

Table 3 presents the simulation results and compares them with the agency's actual data to observe the variation.

Table 3: As Is model simulation results

KPI	Real	Simulation	Variation
Number of Clients	17,815.00	17,095.72	**4.04%**
Total Income	18,916,609.80	17,777,013.00	**6.02%**
Opportunity Costs	4,065,279.76	4,744,320.50	**16.70%**
Conversion Rate	82.31	78.90	**4.14%**

Based on the analysis, it is decided to incorporate the negotiation activity into the process during the search for reservations, including steps to ask the customer about their availability to travel in another month.

Figure 5 shows the updated process, with the customer negotiation activity included in the diagram.

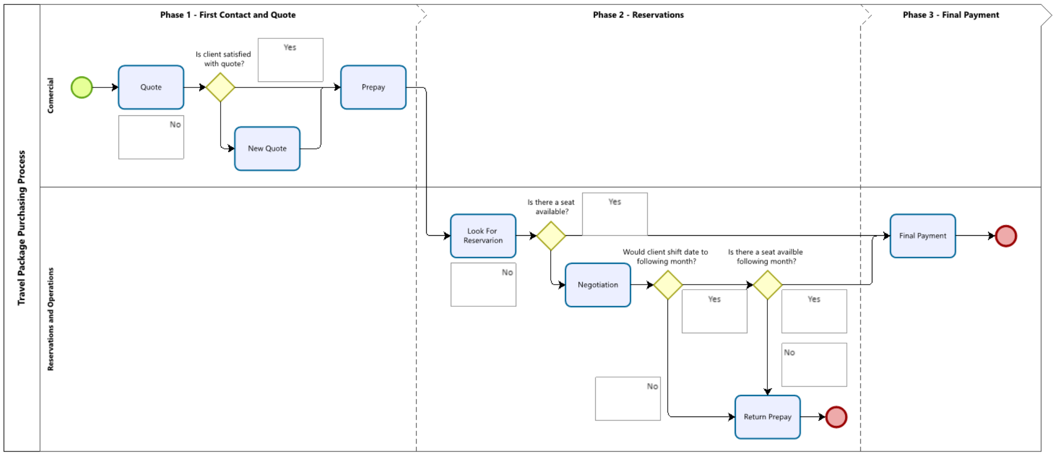

Fig. 5: Travel agency To Be purchase process

Table 4 presents the simulation results of the process with both improvements: the SVR-SARIMAX projection in the available slot variables and BPM with the incorporated negotiation activity. These changes led to a consistent rise in KPIs, especially in conversion rate, which surpassed 90%.

Table 4: To Be model simulation results

KPI	As Is	To Be	Improvement
Number of Clients	17,095.72	19,907.49	**16.45%**
Total Income	17,777,013.00	20,703,908.50	**16.46%**
Opportunity Costs	4,744,320.50	1,814,480.50	**61.75%**
Conversion Rate	78.90	91.90	**16.48%**

The simulation confirmed improvements. The forecast model optimized slot allocation, and the negotiation activity raised flexibility, leading to more customers and lower opportunity costs. Figure 6 shows the comparison of means using Output Analyzer.

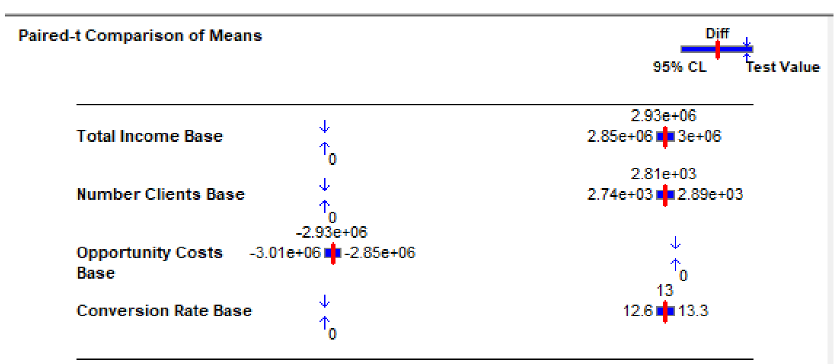

Fig. 6: Output analyzer means comparison

4. Discussion

A MAPE value of 18.18% reflects an 11% improvement over the company's current value of 29.49% for 2023. According to the authors [8][16], this projection would be classified as "good." Based on this, planning and resource allocation can be improved [17]. However, the target value of less than 10% as seen in the studies by [1][2] has not been achieved, indicating that there is still room for improvement.

The integration of BPM with ML achieves a 16% improvement in the indicators, with over 19,000 customers and total total process revenue exceeding $20.7 million. Additionally, a significant 61.75% reduction in opportunity costs is observed, with only $1.8 million. The addition of the negotiation activity contributes to the process by providing a conversion rate of 91.9%, offering a 16% improvement over the current process.

The comparison of means in Output Analyzer shows that, according to the findings of [19][20], the indicators of total revenue, number of customers, and conversion rate will always be higher in the improved process. This is because their intervals are entirely positioned to the right, clearly avoiding zero, which indicates better performance compared to the current process. In contrast, the opportunity cost will always be lower in the improved process, as its interval is fully positioned to the left, demonstrating a significant reduction in costs. These results highlight the effectiveness of the improvements made in the process.

The limitations of the research are mainly centered around the projection with SVR-SARIMAX, as it is not possible to obtain more than 132 observations due to the platform change made by the agency in 2013. Although 10 years of data are sufficient to feed the model, other studies [1][2] use between 20 to 30 years, which helps better capture long-term trends [12]. Another limitation is in the exogenous variable data, where reliable and complete data on accumulated precipitation is required, something that can be complex in the current operational context [13], as well as the inclusion of more variables to obtain even more accurate projections.

Integrating SVR-SARIMAX with BPM improved process performance and resource management [6]. Additionally, it positions the company as a pioneer in technological innovation within the tourism sector, offering a replicable approach that could generate sustainable competitive advantages in the market.

5. Conclusion

The research aimed to increase the profitability of a tourism agency through the implementation of SVR-SARIMAX and BPM. The results show a significant improvement in all indicators, with a 16% increase in total revenue, number of customers, and conversion rate, as well as up to a 61% reduction in opportunity costs. Although the target of achieving a MAPE lower than 10% was not reached, with only 18.18% achieved due to the limitation of using just 10 years of historical data, the positive impact of the tools used is still evident.

Both ML and BPM tools successfully improved the allocation of available slots each month, optimizing resource distribution. They also reinforced operational flexibility, enhancing the company's ability to quickly adapt to demand fluctuations. This, in turn, boosted the company's capacity to respond to market changes, ultimately improving its competitiveness and positioning in the industry.

To strengthen the obtained results, it is recommended to expand the number of observations used to feed the ML model to between 20 and 30 years, in order to better capture trends and approach the MAPE values achieved by other authors [1][2] in the literature, which are below 10%. Additionally, it is recommended to include more exogenous variables in the model, not limiting it to just one, and incorporating other factors such as economic stability and political situation. These initiatives would consolidate the progress made and create new opportunities for improvement in tourism sector management.

References

[1] D. P. M. Abellana, D. M. C. Rivero, Ma. E. Aparente, and A. Rivero, "Hybrid SVR-SARIMA model for tourism forecasting using PROMETHEE II as a selection methodology: a Philippine scenario," *Journal of Tourism Futures*, vol. ahead-of-print, no. ahead-of-print, May 2020, doi: https://doi.org/10.1108/jtf-07-2019-0070.

[2] D. Devianto *et al.*, "An innovative model for capturing seasonal patterns of train passenger movement using exogenous variables and fuzzy time series hybridization," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 10, no. 1, p. 100232, Mar. 2024, doi: https://doi.org/10.1016/j.joitmc.2024.100232.

[3] W. H. K. Tsui, C. K. W. Chow, Y.-H. Lin, and P.-L. Chen, “Econometric analysis of factors influencing Chinese tourist visits to New Zealand,” *Tourism Management Perspectives*, vol. 39, p. 100861, Jul. 2021, doi: https://doi.org/10.1016/j.tmp.2021.100861.
[4] Y. Feng, J. Hao, X. Sun, and J. Li, “Forecasting short-term tourism demand with a decomposition-ensemble strategy,” *Procedia Computer Science*, vol. 199, pp. 879–884, 2022, doi: https://doi.org/10.1016/j.procs.2022.01.110.
[5] P. Gallo, J. Dobrovič, V. Čabinová, P. Pártlová, J. Straková, and B. Mihalčová, “Increasing the Efficiency of Enterprises in Tourism Sector Using Innovative Management Methods and Tools,” *Social Sciences*, vol. 10, no. 4, p. 132, Apr. 2021, doi: https://doi.org/10.3390/socsci10040132.
[6] S. Weinzierl, S. Zilker, S. Dunzer, and M. Matzner, “Machine learning in business process management: A systematic literature review,” *Expert Systems with Applications*, vol. 253, pp. 124181–124181, Nov. 2024, doi: https://doi.org/10.1016/j.eswa.2024.124181.
[7] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, “Time Series Analysis,” *Wiley Series in Probability and Statistics*, Jun. 2008, doi: https://doi.org/10.1002/9781118619193.
[8] M. Mamula, “Modelling and Forecasting International Tourism Demand – Evaluation of Forecasting Performance,” *International Journal of Business Administration*, vol. 6, no. 3, Apr. 2015, doi: https://doi.org/10.5430/ijba.v6n3p102.
[9] MINCETUR, “Balance de Gestión 2022 del Pliego 035: Ministerio de Comercio Exterior y Turismo,” Mar. 2023. Available: https://transparencia.mincetur.gob.pe/documentos/newweb/Portals/0/transparencia/MemoriasMincetur/Balance_Gestion_2022_Mincetur.pdf
[10] ComexPerú, “Reporte de Desempeño Turístico en el Perú al tercer trimestre de 2022,” 2022. Available: https://www.comexperu.org.pe/upload/articles/reportes/reporte-turismo-009.pdf
[11] L. Yao, R. Ma, and H. Wang, “Baidu index-based forecast of daily tourist arrivals through rescaled range analysis, support vector regression, and autoregressive integrated moving average,” *Alexandria Engineering Journal*, vol. 60, no. 1, pp. 365–372, Feb. 2021, doi: https://doi.org/10.1016/j.aej.2020.08.037.
[12] H. Hewamalage, K. Ackermann, and C. Bergmeir, “Forecast evaluation for data scientists: common pitfalls and best practices,” *Data Mining and Knowledge Discovery*, Dec. 2022, doi: https://doi.org/10.1007/s10618-022-00894-5.
[13] O. Meseguer-Ruiz, J. A. Lopez-Bustins, L. Arbiol-Roca, J. Martin-Vide, J. Miró, and M. J. Estrela, “Temporal changes in extreme precipitation and exposure of tourism in Eastern and South-Eastern Spain,” *Theoretical and Applied Climatology*, vol. 144, no. 1–2, pp. 379–390, Feb. 2021, doi: https://doi.org/10.1007/s00704-021-03548-6.
[14] A. S. M. E. Dias *et al.*, “Utilization of the Arena simulation software and Lean improvements in the management of metal surface treatment processes,” *Procedia Computer Science*, vol. 204, pp. 140–147, Jan. 2022, doi: https://doi.org/10.1016/j.procs.2022.08.017.
[15] Byoung Kyu Choi and D. Kang, *Modeling and Simulation of Discrete Event Systems*. John Wiley & Sons, 2013.
[16] Baggio R, Klobas J. *Quantitative Methods in Tourism: A Handbook*. Bristol, Blue Ridge Summit: Channel View Publications; 2017. https://doi.org/10.21832/9781845416201
[17] S. C. Thushara, J.-J. Su, J. S. Bandara, and W. C. Poon, “Forecasting international tourist arrivals in formulating tourism strategies and planning: The case of Sri Lanka,” *Cogent Economics & Finance*, p. 1699884, Dec. 2019, doi: https://doi.org/10.1080/23322039.2019.1699884.
[18] H. Camgoz Akdag, C. O. Kaya, G. Savuran, and N. Z. Canturk, “Application of Lean Principles in Hospitals: A Process Design in an Emergency Department,” *Lecture Notes in Management and Industrial Engineering*, pp. 265–278, Dec. 2017, doi: https://doi.org/10.1007/978-3-319-71225-3_23.
[19] W. David Kelton, R. P. Sadowski, and D. T. Sturrock, *Simulation with Arena*. McGraw-Hill Science, Engineering & Mathematics, 2004.
[20] Torres Vega PJ, *Simulación de sistemas con el software Arena*. Default journal, Jan. 2016.