5th African International Conference on Industrial Engineering and Operations Management, Johannesburg/Pretoria, South Africa, April 23 - 25, 2024
Publisher: IEOM Society International, USA
Published: April 23, 2024 DOI: 10.46254/AF05.20240169

Implementation of Machine Learning in the Forecasting of International Demand for Peruvian Organic Coffee

Fernando Walter Scaler Alcocer
Carrera de Ingeniería Industrial
Universidad de Lima
Lima, Peru
20142273@aloe.ulima.edu.pe

Abstract

This research aims to ensure efficiency in the Peruvian organic coffee supply chain. Advanced Machine Learning techniques will be applied to improve the accuracy of the international demand forecast. For the forecast, data has been collected from the annual export record in tons from 2003 to 2022. The application of Machine Learning in this context provides important benefits, such as the ability to adapt to changing data patterns, identify non-linear relationships between variables and the ability to continuously improve as new data is incorporated, providing more intelligent decisions that directly contribute to the competitiveness and sustainability of the Peruvian organic coffee export industry.

Keywords
Organic coffee, Supply chain, Forecasting, Machine Learning, Software Colab.

1. Introduction
Currently, according to the National Coffee Board, Peru is positioned as the third country in South America with the highest participation in coffee exports, behind Brazil and Colombia. In the case of organic coffee, Peru is the world's leading producer, thanks to the combination of sustainable agricultural practices, favorable altitudes and ideal climatic conditions that have allowed the coffee to be recognized for its unique qualities. This recognition is reflected not only in the growth of international demand, but also clearly in the expectations of consumers who value quality and sustainability. Among the main producing departments we have: San Martin, Junin, Cajamarca, Amazonas, Cuzco, Ucayali, Huanuco, Pucallpa. This Peruvian agricultural product is considered a flagship product and one of the main productive activities, in which approximately 230 thousand families are dedicated to its cultivation. However, the coffee situation was affected at the beginning of 2020, since the World Health Organization declared that the COVID-19 virus was a worldwide epidemic, which caused some concern in all countries, forcing them to make the decision to close their borders. This caused a rupture in the supply chain, directly affecting the producer, importer and exporter, since they had a paralyzed coffee stock and high price volatility. However, according to the Peruvian Foreign Trade Society, in the first half of 2022, exports rose 328% compared to the same period of 2021, showing a recovery after the negative trend experienced due to the pandemic.

Muhuri et al. (2019) comments that machine learning is one of the technologies of industry 4.0. that has successfully contributed in strategic decision making in various companies. Since it has a high potential in process optimization. That is why the objective of this work is the implementation of Machine Learning in the agribusiness sector for the forecast of future international demand for coffee, presenting the lowest possible prediction error, in order to increase the level of service and be more competitive worldwide.

1.1 Objectives
To develop and evaluate the feasibility of an international demand forecasting system for Peruvian organic coffee with the help of Machine Learning, in order to increase the level of service and increase efficiency in the management of the supply chain.

2. Literature Review
There is an extensive literature that raises important positive perspectives on the implementation of machine learning in demand forecasting. These studies coincide with the in-depth study of the importance of production, marketing and supply chain optimization.

Aguilar et al. (2020) and Dairu et al. (2021) address the importance of demand forecasting, through machine learning, in the retail sector with multiple operations. Specifically, they focused on the implementation of the XGBoost algorithm, obtaining results that were validated under real market data, which were efficient and accurate. Likewise, Satya et al. (2023) emphasize the critical connection between revenue estimation and sales forecasting. The analysis examines seasonal patterns and differences between time series studies. They concluded that the implementation of machine learning models, specifically XGBoost, demonstrated improved resource allocation, business planning, and decision making.

On the other hand, Syed et al. (2023) aim at optimization in sales forecasting using machine learning algorithms such as XGBoost, linear regression, ridge regression, decision tree regression, and random forest regression. After analyzing Walmart's Kaggle sales data sets, they concluded that XGBoost and modified outliers were the best algorithm to provide accurate sales forecasts without breaking most of the unused ones. Ultimately, this approach provides an effective tool to improve stock and avoid losses in their retail business. Finally, Madrie and Sempertiga (2019) highlight the importance between the relationships that organic coffee producers and sellers have towards end consumers, this is of utmost importance when observing how organic coffee is currently doing in Peru and what characteristics or types of relationship would improve this index. An important factor for the consumer when selecting an organic product is to know if it is certified by a recognized brand, as this will provide a guarantee of the quality of the product. This aspect is important when developing a national strategy for the promotion of products certified as environmentally friendly.

3. Methods
According to Sanchez et al. (2018) a research design is considered as the research technique to be used to execute the control of variables, relate them and observe them in order to fulfill the research objective.
In the present work, through machine learning, the accuracy of the demand projection will be increased, optimizing the supply chain, increasing customer satisfaction, productivity, yield, and minimizing the logistic cost in the production and commercialization process of Peruvian organic coffee. This mathematical model will be developed through the use of Colab software.

Under this concept, a non-experimental design of longitudinal type was applied, since we collect historical export data at different moments in time and we will analyze how the variables evolve and relate to each other. Therefore, the research work has a correlational scope, since the objective is to evaluate the relationship between two or more variables, achieving to propose, through an optimization model, the accuracy in the demand projection. In conclusion, our research approach is quantitative, since we will use analysis techniques or tools over time.

3.1 Artificial Intelligence
Dwivedi et al. (2019) conclude that in the last decades artificial intelligence has taken vital importance, due to the variety of applications in companies. In addition, it has a wide field of knowledge, which is comprised by subareas such as: Machine Learning and Deep Learning. Machine Learning (ML), also known as machine learning, has become a trend in digital development, making processes more reliable and efficient. A basic outline of the relationship between the areas of knowledge within artificial intelligence is shown in Figure 1.

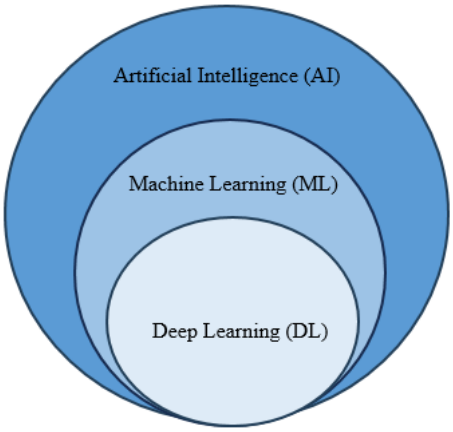

Figure 1. Sub-areas of artificial intelligence knowledge

Regondi et al. (2021) explain that machine learning algorithms for machine learning fall into three categories: reinforced learning, supervised learning and unsupervised learning (Figure 2).

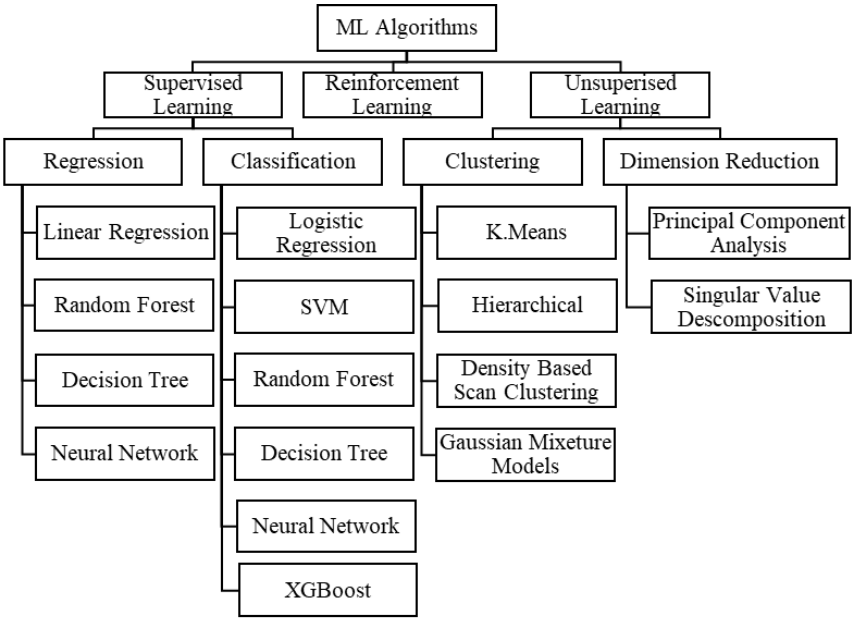

Figure 2. Classification of machine learning algorithms

3.2 Population and sampling

Vara (2012) mentions the importance of obtaining informants or direct sources of information in order to meet the objectives of the research work. This source of direct information is known as population, and it is the grouping of all the people, documents, events, situations, etc. to be investigated. This population has common properties, is found in the same territory and varies over time. It also mentions the sample as the set of cases extracted from the population and which have been selected through a rational method. Taking into account the information provided in Vara (2012), we decided that, for this research work, the population to be studied, under a quantitative approach, is the production and export of Peruvian organic coffee, this product has the tariff heading 0901119000.

3.4 Variables, Dimensions and Indicators

The main variables that we consider for the development of our problem are the following: national market conditions, international market conditions and the characteristics of organic coffee. Within the national market, the most important dimensions, we consider the productive situation, in this we will report on the production yield and the main producing departments. The next variable, international market conditions, we will evaluate the main exporting countries, these values indicate the annual quantity exported in tons (Table 1).

Table 1. Production by departments (tn)

Variables	Dimensions	Indicators
Domestic market conditions	Production situation of Peruvian organic coffee.	The hectares planted for its production and its yield.
	Main regions where coffee production takes place.	Tons produced by each region.
International market conditions	Export situation of Peruvian organic coffee in recent years.	Tons exported annually.
	Main countries where organic coffee is exported.	Tons exported annually to each country.
Characteristics of organic coffee	Quality of Peruvian organic coffee.	Necessary characteristics for the optimal development of organic coffee.

3.4 Supply Chain Links

The supply chain is a vital point in the commercialization of organic coffee, since having an optimal supply chain will reduce costs, improve quality, service time, competitiveness in the market, and improve the level of service. Among the links in the supply chain are: the supplier, farmer, intermediary, collection center, processing company, retailer, wholesaler, exporting companies and the final consumer. Figure 3 below shows the links in the supply chain.

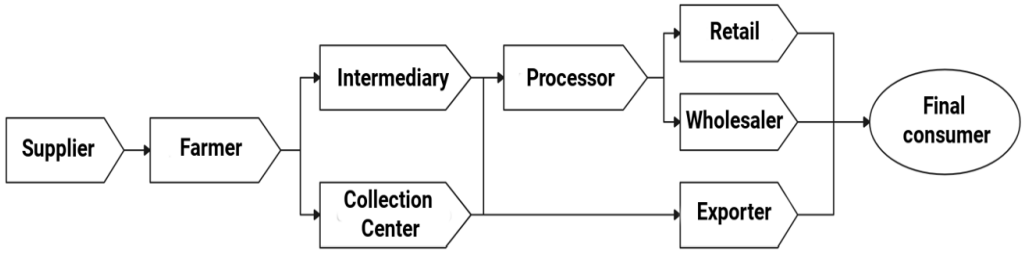

Figure 3. Ranking of coffee bean producing departments

4. Data Collection

This chapter presents the main data for the research process, such as the production of Peruvian coffee beans by department and the yield in tons per hectare of harvest in the period 2015 to 2021. Subsequently, the evolution of exports from 2003 to 2022 will be presented. Likewise, information will be presented regarding the presence by continents and the main consuming countries. Likewise, the fundamental and physicochemical characteristics for the optimal development of production in our region will be presented. Finally, the modeling process for demand forecasting will be presented (Table 2).

4.1 Production by department in tons

Table 2. Production by departments (tn)

Departments	2021	2020	2019	2018	2017	2016	2015
San Martín	77,994	82,809	85,439	91,423	91,197	91,197	82,164
Junín	68,463	72,335	80,430	89,837	75,100	75,100	39,275
Cajamarca	76,758	71,793	71,794	63,893	62,863	62,863	46,083
Amazon	51,087	44,991	42,843	43,946	38,893	38,893	35,101
Cusco	26,265	27,627	28,264	30,754	26,615	26,615	18,413
Pasco	20,430	13,193	11,484	13,610	11,669	11,669	6,898
Huanuco	12,409	11,921	11,699	10,782	9,427	9,427	5,109
Ucayali	12,659	10,968	13,622	8,325	4,004	4,004	5,442
Puno	8,314	8,105	8,122	7,784	7,754	7,754	6,504
Piura	4,854	4,987	4,731	3,660	4,050	4,050	2,677
Ayacucho	3,353	2,212	2,450	3,430	3,781	3,781	3,051
Lambayeque	2,222	1,838	2,009	1,748	1,553	1,553	863
La Libertad	203	220	226	225	227	227	188
Loreto	192	189	183	178	171	171	150
Madre de Dios	19	16	12	13	14	14	13
Huancavelica	0	1	12	12	12	12	7

Figure 4 shows the distribution of production in the most important producing departments of Peru between 2015 and 2020.

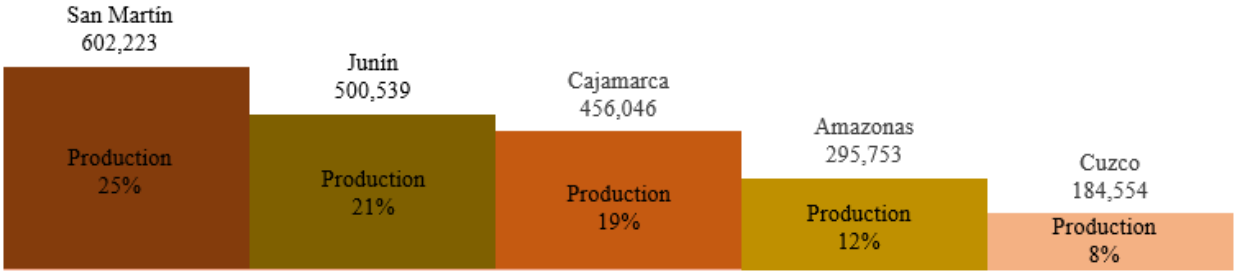

Figure 4. Ranking of coffee bean-producing departments

4.2 Performance

Table 3 presents the coffee yield, which consists of the ratio of harvested area data with respect to the production of Peruvian coffee beans between 2015 and 2021. Yield is an important factor to evaluate, since it will be possible to analyze how climate, pests, temperature, etc. affect the cultivation of coffee beans.

Table 3. Production by departments (tn)

Year	Production (Tn)	Surface area (Ha)	Yield (Tn/Ha)
2015	251,938	379,282	66%
2016	227,760	383,973	59%
2017	337,330	424,129	80%
2018	369,622	447,426	83%
2019	363,320	438,177	83%
2020	353,206	430,820	82%

2021	365,221	427,433	85%

4.3 Export evaluation

Table 4 below shows the evolution, between 2003 and 2022, of Peruvian coffee exports in tons. According to the Ministry of Agrarian Development and Risk, the commercialization of coffee beans did not show signs of recovery. For this reason, this scientific article provides a strategy in the supply chain to generate a significant increase; this means an increase in jobs, in the sales of farmers, recognition in the world market and the country's economy. It is worth noting that, in 2011, farmers faced a plague, which was one of the factors in the fall of exports until 2015 reaching 176,177 tons. After this relapse, there was an intensification of exports, reaching 213,385 tons of organic coffee by the end of 2020.

Table 4. Exports and Evolution (tn)

Year	Exports (Tn)	Variation
2003	150,544	-
2004	191,140	27%
2005	142,166	-26%
2006	238,084	67%
2007	173,624	-27%
2008	225,090	30%
2009	197,664	-12%
2010	230,052	16%
2011	296,416	29%
2012	266,394	-10%
2013	238,690	-10%
2014	185,418	-22%
2015	176,177	-5%
2016	239,631	36%
2017	246,019	3%
2018	256,361	4%
2019	226,922	-11%
2020	213,385	-6%
2021	191,965	-10%
2022	237,235	24%

Figure 5 shows the presence of Peruvian coffee by continent in the last 20 years, with Europe as the main destination with 51.49%, followed by South America and North America with 20.33% and 20.05% respectively. This is followed by Asia with 5.16% and Oceania with 2.57%. Finally, Central America and Africa have smaller shares with 0.35% and 0.06% respectively.

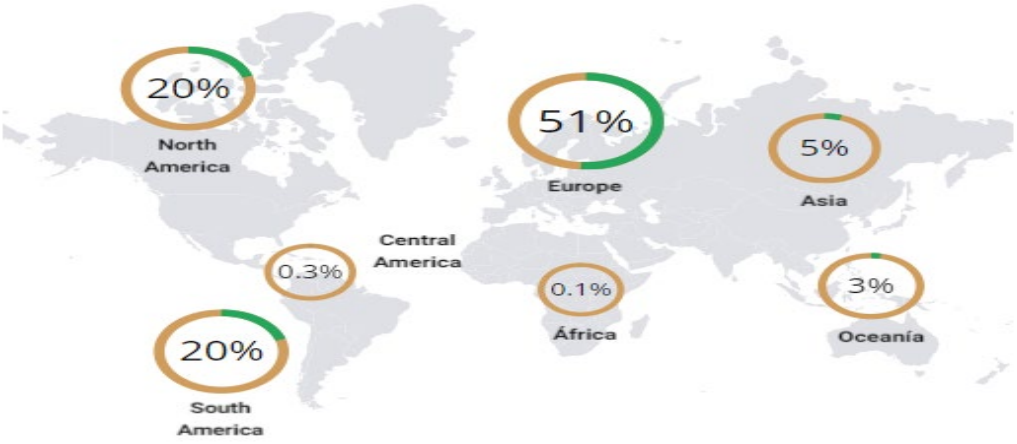

Figure 5. Presence of Peruvian coffee by continent

4.4 Main consuming countries

Figure 6 below shows the main export destinations of Peruvian coffee in the last 20 years.

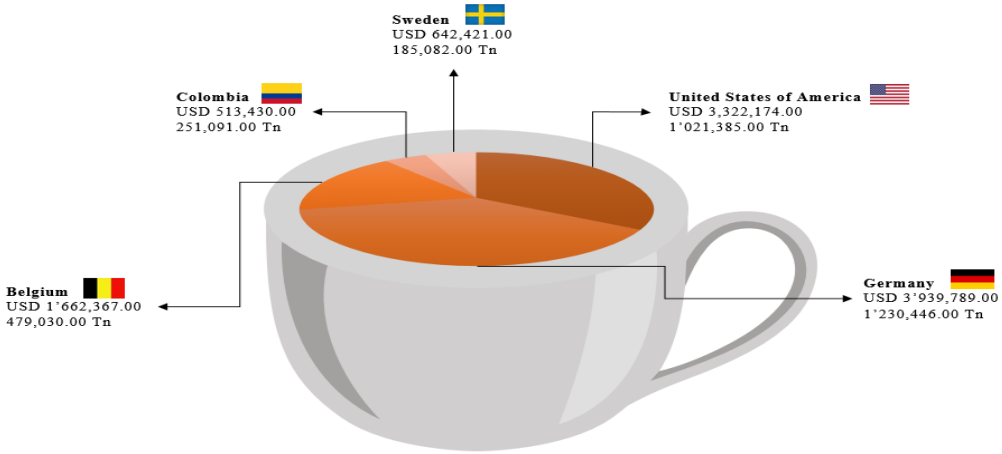

Figure 6. Main organic coffee consuming countries

5. Results and Discussion

5.1 Numerical Results

In the evaluation of the polynomial regression and XGBoost models to predict the annual coffee export volume for 2022, significant differences in performance were appreciated. The polynomial regression shows obvious limitations with a low R^2 of 0.0958 and a high MAE of 30,317.68, indicating difficulty in accounting for data variability and lack of prediction accuracy. On the other hand, the XGBoost model performs satisfactorily, showing an R^2 of 0.9066 and MAE of 4,061.33. These results demonstrate the ability of XGBoost to correctly adapt to the dynamics, providing accurate predictions with little deviation from the real values. Finally, the fitted XGBoost model has had mixed results, as the MAE is better with a value of 25,003.26, but the R^2 is 0.3835. Therefore, under the annual representation scheme, the superiority of the unadjusted XGBoost model for forecasting annual Peruvian coffee exports is reinforced. The results of the model are detailed and illustrated below (Table 5).

Table 5. XGBoost model forecast

Year	Real	Polynomial Regression	XGBoost	XGBoost Adjusted
2003	150,545	84,282	142,170	202,336
2004	191,140	126,023	142,170	202,336
2005	142,170	159,989	142,170	202,336
2006	238,081	186,879	238,081	223,598
2007	173,626	207,395	173,626	210,102
2008	225,093	222,238	225,093	217,352
2009	197,667	232,108	197,667	217,352
2010	230,052	237,706	230,052	219,906
2011	296,415	239,734	296,415	234,717
2012	266,397	238,892	266,397	234,717
2013	238,691	235,881	238,691	232,860
2014	185,420	231,402	185,420	208,740
2015	176,181	226,156	176,181	208,740
2016	239,632	220,844	239,632	222,510
2017	246,020	216,167	246,020	222,510
2018	256,362	212,825	246,020	222,510
2019	226,927	211,521	226,927	222,510
2020	213,388	212,954	226,927	222,510
2021	191,963	217,825	191,963	222,510
2022	237,235	226,836	237,235	222,510

5.2 Graphical Results

The main objective was the implementation of machine learning to improve demand accuracy, and with the help of the XGBoost model it was possible to propose and achieve the expected results. In both studies, advanced programming language was applied and it was possible to solve the problem of demand forecasting, which is a very important factor in the efficiency of the supply chain, since a bad forecast could generate a high operating cost, poor service level, high number of lost customers and poor decision making (Figure 7 and 8).

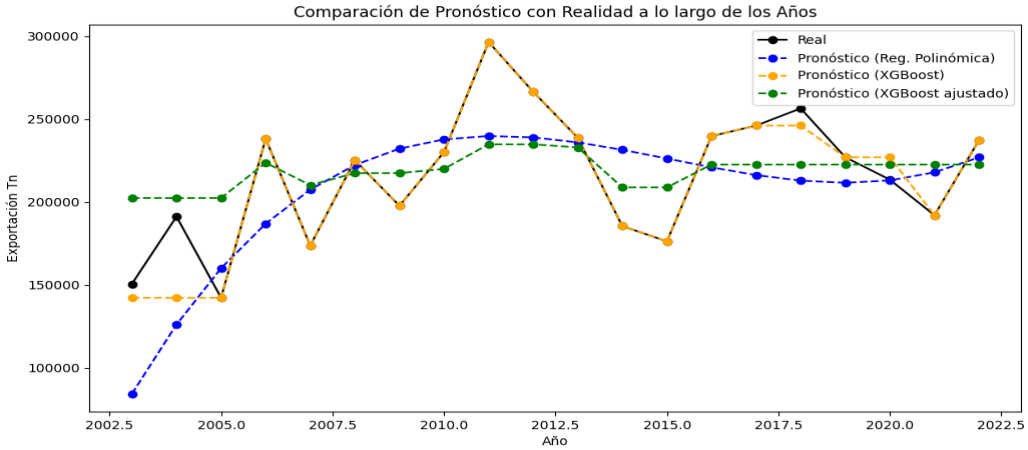

Figure 7. Predicted vs. actual values over the years.

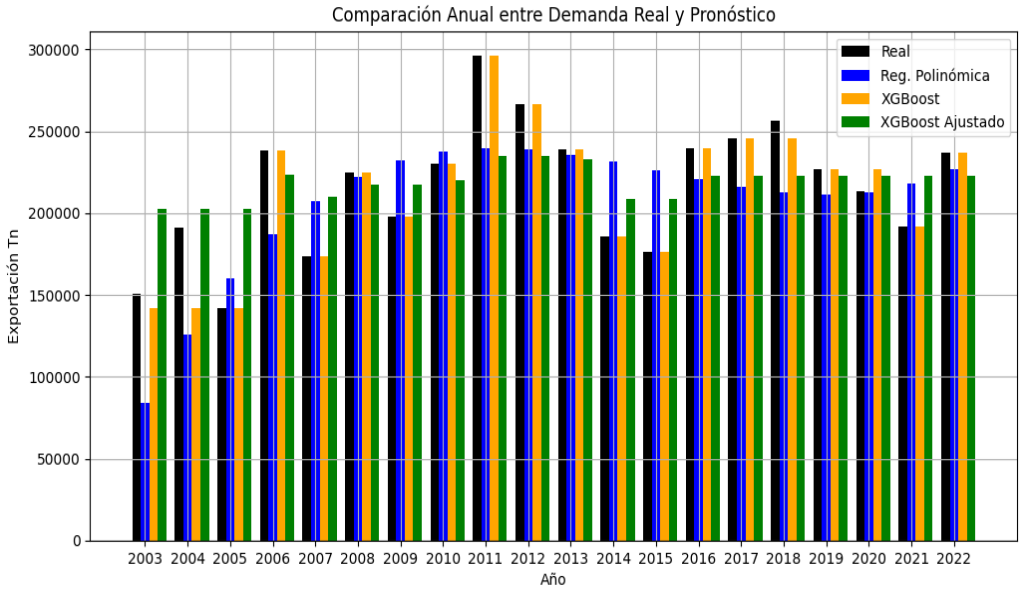

Figure 8. Yearly comparison between actual and forecasted demand

6. Conclusion

The study developed and evaluated the importance of sample selection for forecasting international demand for Peruvian organic coffee. The results show that the XGBoost model demonstrated a high coefficient of determination (R^2) and low absolute error (MAE), making it the most effective option compared to polynomial regression. The successful implementation of the XGBoost model will generate multiple benefits for Peruvian organic coffee exports, such as guaranteeing a reliable and timely supply, strengthening trade relations, and enhancing Peru's reputation as a supplier. Likewise, this positive impact will contribute to the sustainable economic growth of the producing families and the country in general, which is why the importance of optimizing the supply chain is a fundamental element for the competitiveness of companies.

Finally, companies can use this tool to reduce costs, increase customer satisfaction and reduce negative impacts on the supply chain.

References

Aguilar-Palacios, C., S. Munoz-Romero, and J. L. Rojo-Alvarez, "Cold-Start Promotional Sales Forecasting through Gradient Boosted-Based Contrastive Explanations," IEEE Access, vol. 8, pp. 137574-137586, 2020.

Arjomandi, A. A. Ahmadi, S. Taheri, M. Fotuhi, M. Moeini, and M. Lehtonen, "Pandemic-Aware Day-Ahead Demand Forecasting Using Ensemble Learning," IEEE Access, vol. 10, pp. 7098-7106, 2022.

Dairu X.and Z. Shilong, "Machine learning model for sales forecasting by using XGBoost," IEEE Access, pp. 480-483, Jan. 2021.

Dwivedi, Y., L. Hughes, and Elvira. Ismagilova, "Artificial intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy," Int J Inf Manage, 2019.

Espinosa, J., "Application of random forest and XGBoost algorithms based on a credit card applications database," Engineering Research and Technology, vol. 21, no. 3, pp. 1-16, 2020.

Madrie, J. and I. Sempertiga, "Analysis of factors involved in the purchase intention of organic coffee," Universidad San Ignacio de Loyola, Peru, 2019.

Muhuri, P., A. Shukla, and A. Abraham, "Industry 4.0: A bibliometric analysis and detailed overview," Eng Appl Artif Intell, vol. 78, pp. 218-235, 2019.

Regondi, S., R. Pugliese, and R. Marini, "Machine learning-based approach: Global trends, research directions, and regulatory standpoints," Data Science and Management, vol. 4, pp. 19-29, 2021.

Sanchez, H., C. Reyes, and K. Mejia, "Handbook of terms in scientific, technological, and humanistic research." 2018.

Satya, K., S. Durga, M. Hemanth, M. Vinay, and B. Sai, "Sales Forecasting Using Xgboost," International Journal of Creative Research Thoughts, vol. 11, no. 4, p. 726, 2023.

Syed, H., K. Alice, and S. Anoop, "Sales Forecasting using XGBoost," Institute of Science and Technology, 2023.

Vara, A., "7 steps to a successful thesis," Universidad de San Martin de Porres, Peru, 2012.

Villafuerte, F., "Comparative analysis of ARIMA and XGBoost forecasting models applied to monthly sales series in a certifying company," Universidad Nacional Mayor de San Marcos, Peru, 2021.

Biographies

Fernando Scaler Alcocer holds a Bachelor's Degree in Industrial Engineering from Universidad de Lima. Enriching his education, he has successfully completed the specialization in Supply Chain Management and the Diploma in Integrated Management Systems at the Universidad Nacional Agraria la Molina. His research interests focus on the implementation of machine learning and supply chain improvement. He has worked in important companies in Peru and has been able to develop and improve his knowledge in the industrial, construction and automotive sectors. He stands out for his contribution to the development of effective solutions for logistics challenges, reflecting a commitment to excellence.