Original Article

Improving Demand Forecasting by Implementing Machine Learning in Poultry Production Company

Joaquin Garcia-Arismendiz¹, Sandra Huertas-Zúñiga¹, Carlos Augusto Lizárraga-Portugal¹, Juan Carlos Quiroz-Flores^{1*} and Yvan Jesus Garcia-Lopez¹

¹Facultad de Ingeniería y Arquitectura, Universidad de Lima, Perú.

*Corresponding Author: Jcquiroz@ulima.edu.pe

Received: 14 November 2022

Revised: 14 January 2023

Accepted: 23 January 2023

Published: 25 February 2023

Abstract - The use of manual methods to forecast demand in perishable food companies is generally subject to the variability of internal and external factors in the company, causing excess inventories and significant monetary losses, so it is relevant to carry out this research with the objective of to demonstrate that by implementing Machine Learning it is possible to improve the accuracy of the demand forecast. A case study in a company in the poultry sector in Peru, forecasting the last quarter of 2022, based on a real sales database and applying the time series method, comparing the results of the Machine Learning model, and obtaining as a result in a model with high Forecast Accuracy (FA) of 97.56% and a high Forecast Bias (FB) of 2.44%. The research is an important contribution to knowledge, demonstrating that Machine Learning is an ideal tool to project the demand for perishable food products, ideal for its application in various fields, such as loss reduction control, preventive maintenance of machines and control of supplies such as water and energy, among others.

Keywords - Machine learning, Demand forecasting, Poultry company.

1. Introduction

Demand forecasting in the food industry is an important issue to ensure commercial, economic and sustainable development success. [1]. Moreno et al. mentioned that the fresh and refrigerated food industry must take special care to meet high-quality standards in their production, storage and distribution processes, which is why they must apply very accurate and reliable demand forecasting techniques in order to develop a production plan that minimizes sales losses due to lack of products and reduce returns due to expiration date, better; besides controlling their waste environmental effects, and also improving availability to customers, thus reflecting better results in the company and even reducing environmental [2]. It explains that food production has complex processes subject to uncertainty in demand forecasting and variability in the yield of materials and their supply, among other aspects that generate important differences between planned and actual production [3]. The food industry applies engineering solutions through to improve its performance, industrial automation supported by artificial intelligence (AI) or machine learning (ML) or deep learning (DL) algorithms to manage a sales plan efficiently while improving operational competence. [4].

Given the demands of a highly competitive market, it is of vital importance that companies incorporate market intelligence tools that enable them to make proactive decisions to ensure the achievement of their strategic objectives [5]. Moreover, making the right strategic decisions depends on the accuracy of demand forecasting, avoiding costs due to overestimation or underestimation of demand [6].

The main reasons for not being able to obtain an accurate demand forecast are variability and outliers, external factors, and not considering the necessary events and variables. Also, suppliers are highly dependent on delivering products to the final consumer. In this sense, the main objective of this research is to apply a Machine Learning model for forecasting the demand for a Peruvian company's poultry product to improve its accuracy, reducing the economic loss caused by the waste of lost sales and excess stock of products.

2. Literature Review

This research followed an in-depth literature review to recognize relevant works related to machine learning and demand forecasting using models based on neural networks, time series and sales forecasting, showing the advantages and methodology of their use.

A study to evaluate various Machine Learning models in forecasting the demand for new credits in agriculture; demonstrated that not all Machine Learning models are ideal for forecasting, although most of them have greater predictive power than standard models, improving their performance when the analysis variables are well selected [7]. Another article presents the benefits of implementing Machine Learning in sales forecasting in the fresh food and highly perishable products sector; obtaining great support for decision-making, improving supply chain management and reducing the lack of stock at the point of sale, increasing the profitability and competitiveness of the company [14, 25].

A study that sought to solve the problem of the low level of the bicycle rental service for urban transport, determining the inaccuracy of the demand as the main cause, for which a neural network model with Machine Learning was proposed, proving to have better results in comparison with other learning algorithms [8,9].

A study proposes a new short-term approach to water demand forecasting by applying a two-stage smart learning process combining time series clustering with gene expression programming (GEP), resulting in a multi-scale model starting from a time series to optimize the proposed solution (water demand) [10].

This research work proposes a predictive model to monitor the status of industrial machines. Using a Machine Learning model, it was possible to help the early detection of failures thanks to the historical data extracted from the machines. With this detection, it was possible to change in advance the necessary steps so that the machine does not present unnecessary failures in the future. Thus the production lines are not affected [11]. The problem presented by this study is that the electricity supply is insufficient for modern systems. Therefore, the objective was to use a Machine Learning model to improve the generation scheduling in electricity markets. The development of the solution framework and its implementation is simple and flexible and will greatly help GENCOs with quick decision-making in the current competitive scenario [12].

3. Materials and Methods

This research follows the Machine Learning methodology, where a poultry production company performs an objective comparison of traditional forecasting methods against Machine Learning models. Python code was used to create an efficient Machine Learning model. The block diagram methodology is shown in Figure 1.

3.1. Business Understanding & Performance Metrics Definition

As a starting point, the company's facilities were visited, interviewing the general manager, compiling reports on the company's situation, and documents with sales plan data. A diagnosis was made by applying the problem tree technique to identify the main problem and its root causes to establish alternative solutions. As indicators to generate the demand prediction model for each product, the minimum value of the mean absolute error (MAE), the mean absolute percentage error (MAPE), the mean square error (RMSE) and the coefficient of determination were determined. (R-squared) that measures the percentage of variation of the response variable explained by the explanatory variable [1]. These indicators were defined as follows:



Fig. 1 Phases for the construction of the Machine Learning model [13]

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |\hat{y} - y_i|$$
 (1)

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\hat{y} - y_i}{y_i} \right| * 100\%$$
 (2)

Predicted data minus actual data divided by actual data by 100%.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y} - y_i)^2}$$
(3)

Actual data minus predicted data squared.

$$R^{2} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\hat{Y} - \underline{y}\right)^{2}}}{\sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(y_{i} - \underline{y}\right)^{2}}}$$
(4)

Where n is the out-of-sample number; y_i and \hat{y}_i are the observed (actual) values and adjusted values of the dependent values of the Y variable for the i-th case, respectively; furthermore \hat{Y} is the arithmetic mean of Y [1]

In addition, a performance indicator used by the company will be used, which is called as "compliance indicator", which is shown below:

$$Compliance \ indicator = \frac{Actual \ sales}{Planned \ sales} * 100\%$$

It is interpreted as:

- If the percentage is greater than 100%, more was produced than was sold.
- If the percentage is less than 100%, less was produced than was sold.

Before generating the statistical projections, the historical product demand data was analyzed using statistical tests such as The Ljung-Box (LB) test to determine whether or not there are autocorrelations in model residuals of adjusted time series [26] and; Shapiro test to determine if a data sample comes from a population with a normal distribution [15]; Kwiatkowski Phillips Schmidt Shin (KPSS) test for stationarity of a time series [16]; and the Augmented Dickey-Fuller (ADF) for the stationarity of a time series [17].

For the sales forecast, the results will be subjected to the following indicators:

Forecast accuracy (FA): measures how accurate the forecast is [27].

$$FA = 1 - \frac{|SALES - FORECAST|}{SALES}$$

Forecast bias (FB): measures the accuracy between the margin between forecast and actual sales [19].

$$FB = \frac{FORECAST}{ACTUAL} - 1$$

To achieve this, the formulas must be included in the Python code.

3.2. Data Preparation Phase

The following phases were carried out

- Data insertion for demand forecasting: the data must be analyzed constantly and frequently, ensuring that the raw data flows through a solid and reliable data management process, available when required.
- Data exploration and understanding: To achieve a reliable forecast, quality data must be used, specifying its characteristics of size, precision, initial patterns, missing values, measurement precision, measurement time, synchronization, and latency.
- Data and output variable preprocessing: Establishing proper data preparation and cleansing is critical to having features that will make machine learning algorithms perform better and reproduce more accurate results.

3.3. Model Construction and Selection

In the modeling phase, the data is converted according to the forecast method, using advanced algorithms that scan the historical data (training data), patterns are extracted, completing the model that then predicts different data from those used to build the model.

- Training data set: Uses an amount of data to fit machine learning models, training a machine learning algorithm that "learns" from historical data to be able to predict future data points.
- Validation data set: For this point, a "validation data" set is used to impartially evaluate the fit of a model on the "training data" set while fitting machine learning model parameters. Validation data sets should be leveraged to fit a machine learning model's parameters by observing its performance to improve the models.
- Test data set: when the model is fully trained and validated through the "training and validation data" set using the "test data" tested on different machine learning models, evaluating their fit below or above normal.

3.4. Model Implementation

The implementation of the model must be executed using a code editor and a programming language, in this case, Python, adjusting the data under study and then executing the Machine Learning algorithm to predict the desired period finally.

3.5. Results Validation

This is the last stage of the methodology, where the error metrics obtained by the proposed model will be compared with the metrics provided by the company. To then perform a hypothesis test to determine if there is a significant difference between the two samples. In addition, predictions will be made for certain periods, which have no data, and then compared with actual sales data provided by the company in order to validate whether the model is working well or not.

4. Results and Discussion

Following the best performance of the models executed in Python, the following was selected:





Fig. 4 Dataset import

[]	<pre>from pycaret.time_series import * setup(data,target = 'UNIDADES', fold_strategy='sliding',fh = 8,</pre>
	Fig. 5 Setup time series

Fig. 6 Comparison of prediction models

To achieve the study, the following forecast models were used: a) Random Forest, the result of combining prediction trees [20]; b) Gradient Boost Algorithm, which shows high precision in forecasting and future sales, using Intelligent Decision Analytical System methods [21,24]; exponential smoothing method (ETS) computes a weighted average over all observations in the input time series dataset as its prediction [22].

4.1. Forecast Performance

Table 2 shows the statistical parameters of sales between the periods 2016-01 to 2022-07. The mean, median, standard deviation and variance are presented.

Table 2. Dataset statistics				
Property	Value			
Length	80.00			
Mean	1723452.56			
Median	1718817.00			
Standard Deviation	101424.90			
Variance	10287009391.87			

Table 2. Dataset statistics

The sales data analysis was subjected to statistical tests where the p-value can be observed. The results are presented in Table 3.

Table 3. Statistical test					
Test Name	Property	erty Setting			
Tinna	p-value	{'alpha': 0.05, 'K': 24}	1.5x10-18		
Ljung-	Correlation		24		
вох	period		months		
ADE	p-value	{'alpha': 0.05, 'K': 24}	0.97		
ADF	Stationarity	{'alpha': 0.05}	FALSO		
	p-value	{'alpha': 0.05}	0.041		
KPSS	Trend Stationarity	{'alpha': 0.05}	FALSO		
Shapiro	p-value	{'alpha': 0.05}	0.91		
	Normality	{'alpha': 0.05}	TRUE		

Table 3 shows the training and test distribution applied to the historical sales data. As seen in Figure 7, the period 2016 to 2021 was used as training data and January to August 2022 as test data.



Fig. 7 Train test split



Fig. 8 Forecast Random Forest Regressor **T 11 4 F**

..

Model	MAE	RMSE	MAPE	SMAPE	MASE	RMSSE
Random Forest w/ Cond. Deseasonalize & Detrending	71575.43	82170.81	0.0392	0.0401	1.0658	0.8946
Gradient Boosting w/ Cond. Deseasonalize & Detrending	72919.38	87076.74	0.0399	0.0406	1.0754	0.945
ETS	72790.05	83432.45	0.04	0.0409	1.0834	0.907

Table 5. FB and FA						
DATE	Current		Current Model			
	Bias	Accuracy	Bias	Accuracy		
2022-02	0.11%	99.89%	1.35%	98.65%		
2022-03	11.61%	88.39%	0.05%	99.95%		
2022-04	7.19%	92.81%	4.08%	95.92%		
2022-05	15.62%	84.38%	3.88%	96.12%		
2022-06	3.09%	96.91%	0.59%	99.41%		
2022-07	9.63%	90.37%	3.89%	96.11%		
2022-08	13.71%	86.29%	3.26%	96.74%		

DATE	Compliance indicator			
	Current	Model		
2022-02	100.11%	98.65%		
2022-03	88.39%	99.95%		
2022-04	92.81%	95.92%		
2022-05	115.62%	96.12%		
2022-06	96.91%	100.59%		
2022-07	90.37%	96.11%		
2022-08	113.71%	96.74%		

Table 4 shows the performance of the Auto Arima, Arima and AdaBoost w/ Cond. Deseasonalize & Detrending, where it can be observed, according to their indicators, that Auto Arima is the best fit using the test data. The table shows the MAE, RMSE, MAPE, SMAPE, MASE and RMSSE of these models, and the graph of the best model can also be seen in Figure 8.

Table 5 compares forecast bias (FB) and forecast accuracy (FA) between the current and ML models. It can be observed that in February and June, the FB and FA of the ML model did not improve against the company's current model. However, good results were obtained in the FA and the FB in the remaining months with the ML model, with a better average performance.

Table 6 shows the compliance indicator throughout the year 2022, where it can be seen that, on average, the LC model achieves better results with an indicator closer to 100%.

4.1. Validation

At this stage, the student's t-hypothesis test was used and applied to the FB indicator in the table to validate if there is a significant sample between the samples, where the following hypotheses were established:

- H0: There is no significant difference between the error rates of Machine Learning and those obtained from the company.
- H1: If there is a significant difference between the error rates of Machine Learning and those obtained from the company.



Date	Prediction	Real	Bias	Accuracy
2022-09	1796243.6	1831770	1.94%	98.06%
2022-10	1854039.4	1994368	7.04%	92.96%

Table 7. Machine Learning model prediction

Table 8. Company prediction

Date	Prediction	Real	Bias	Accuracy		
2022-09	1728122	1831770	5.66%	94.34%		
2022-10	1738652	1994368	12.82%	87.18%		

Figure 9 shows the T-test calculation that yields the value of -2.93, which concludes that the null hypothesis must be rejected, so H0 is accepted. There is a significant difference between the error percentages of Machine Learning and those obtained from the company, validating that the error obtained by the Machine Learning model obtains an improvement with respect to that obtained by the company's model.

To complete the validation, we made the prediction for two periods, September and October, and then asked for sales data for those same periods and made the comparison, which can be seen in Table 7, where it can be seen that the model is predicting correctly by having accurate results.

As seen in Tables 7 and 8, the prediction of the Machine Learning model is more accurate in both periods than the prediction of the poultry company. This gives us one more validation tool, which indicates that the Machine Learning model is working correctly.

On the other hand, this research allowed us to carry out an exhaustive search of how Machine Learning can be implemented not only in issues of environmental aspects or essential services, as shown in other scientific articles but also allowed Machine Learning to be implemented in the projection of the demand of a poultry production company, a product which is perishable, which leaves the forecast susceptible to changes and with a short period of time for handling it. In addition, it was possible to determine the importance of the Machine Learning tool and how it can be applied in various fields, such as loss reduction control, preventive maintenance of machines and control of supplies such as water, electricity, etc.

5. Conclusion

Based on the results of this study, it has been possible to demonstrate that with the implementation of the Machine Learning model, it was possible to reduce the error metrics by 9.7% compared to the model currently used by the company. In addition, the performance of the Machine Learning model was successfully measured through the compliance KPI and the MAPE, which helped to know that this model does show a significant improvement in the sales plan forecast of the poultry production company.

By comparing the results of the Machine Learning model with the company's actual sales, it is shown that the model has a high Forecast Accuracy (FA) of 97.56% and a high Forecast bias (FB) of 2.44%, achieving a reduction of economic losses by 64.68% and a reduction of inventories by 49.77%.

Similar evidence from other sectors shows that reducing the error metrics in the demand forecast is possible to be more accurate, which in effect reduces economic losses and can lead to better inventory management in relation to overstocking or shortages. This study fulfills the objective of providing new knowledge and tools to minimize the error metrics in a predictive estimation of a sales plan in the poultry sector. In addition, how it positively affects other areas, such as production, which will allow a reduction in inventory management costs.

Funding Statement

The authors declare that they have financed their article's research and publication with their personal financial resources.

References

- Krzysztof Lutoslawski et al., "Food Demand Prediction Using the Nonlinear Autoregressive Exogenous Neural Network," *IEEE Access*, vol. 9, pp. 146123-146136, 2021. *Crossref*, https://doi.org/10.1109/ACCESS.2021.3123255
- [2] Carlos A. Moreno-Camacho et al., "Sustainability Metrics for Real Case Applications of the Supply Chain Network Design Problem: A Systematic Literature Review," *Journal of Cleaner Production*, vol. 231, pp. 600-618, 2019. *Crossref*, https://doi.org/10.1016/j.jclepro.2019.05.278
- [3] Alberto Garre, Mari Carmen Ruiz, and Eloy Hontoria, "Application of Machine Learning to Support Production Planning of a Food Industry in the Context of Waste Generation Under Uncertainty," *Operations Research Perspectives*, vol. 7, pp. 100147, 2020. *Crossref*, https://doi.org/10.1016/j.orp.2020.100147
- [4] Indrajeet Kumar et al., "Opportunities of Artificial Intelligence and Machine Learning in the Food Industry," *Journal of Food Quality*, pp. 1-10, 2021. *Crossref*, https://doi.org/10.1155/2021/4535567
- [5] Rafael Mota Viana, Tiago Bittencourt Nazaré, and Mário César Fialho, "Market Intelligence Tools for Improving Decision-Making in a Textile Company," *International Journal of Engineering Trends and Technology (IJETT)*, vol. 67, no. 10, pp. 141-146, 2019. *Crossref*, https://doi.org/10.14445/22315381/IJETT-V67I10P224

- [6] P. Kumar Sahu, and R. Kumar, "The Evaluation of Forecasting Methods for Sales of Sterilized Flavoured Milk in Chhattisgarh," *International Journal of Engineering Trends and Technology (IJETT)*, vol. 8, no. 2, pp. 98-104, 2014. *Crossref*, https://doi.org/10.14445/22315381/IJETT-V8P219
- [7] Jennifer Ifft, Ryan Kuhnsb, and Kevin Patrick, "Can Machine Learning Improve Prediction An Application with Farm Survey Data," International Food and Agribusiness Management Review, vol. 21, no. 8, pp. 1083-1098, 2018. Crossef, https://doi.org/10.22434/IFAMR2017.0098
- [8] Celsus murenjekha shilehwa, Sibilike khamala makhanu, and Alex Wabwoba khaemba, "Application of Artificial Neural Network Model in Forecasting Water Demand: Case of Kimilili Water Supply Scheme, Kenya," *SSRG International Journal of Civil Engineering*, vol. 6, no. 9, pp. 6-11, 2019. *Crossref*, https://doi.org/10.14445/23488352/IJCE-V6I9P102
- [9] Fan Wu et al., "Neural Networks with Improved Extreme Learning Machine for Demand Prediction of Bike-Sharing," *Nature Mobile Networks and Applications*, vol. 26, no. 5, pp. 2035-2045, 2021. *Crossref*, https://doi.org/10.1007/s11036-021-01737-1
- [10] Sina Shabani et al., "Gene Expression Programming Coupled with Unsupervised Learning: A Two-Stage Learning Process in Multi-Scale, Short-Term Water Demand Forecasts," *Water*, vol. 10, no. 2, pp. 1-13, 2018. *Crossref*, https://doi.org/10.3390/w10020142
- [11] R Srivatsan et al., "Early Maintenance and Diagnosis of Connected Machines Using Machine Learning," International Journal of Electrical Engineering and Technology (IJEET), vol. 11, no. 4, pp. 412-421, 2020. Crossref, https://doi.org/10.34218/IJEET.11.4.2020.046
- [12] Dr. P G Latha, "A Machine Learning Approach for Generation Scheduling in Electricity Markets," *International Journal of Electrical Engineering and Technology (IJEET)*, vol. 9, no. 3, pp. 69-79, 2018.
- [13] Francesca Lazzeri, "Machine Learning for Time Series Forecasting with Python," Indianapolis, Indiana, pp. 46-68, 2021.
- [14] Han-Yu XUE, Shang XIANG, and Jun LI, "An Innovative Design of the Internet of Things for Supply Chain Management of Fresh Agricultural Products," SSRG International Journal of Computer Science and Engineering, vol. 7, no. 12, pp. 1-4, 2020. Crossref, https://doi.org/10.14445/23488387/IJCSE-V7I12P101
- [15] Berna Yazici, and Senay Yolacan, "A Comparison of Various Tests of Normality," *Journal of Statistical Computation and Simulation*, vol. 77, no. 2, pp. 175-183, 2007. *Crossref*, https://doi.org/10.1080/10629360600678310
- [16] Piotr Kokoszka, and Gabriel Young, "KPSS Test for Functional Time Series," Statistics, vol. 50, no. 5, pp. 957-973, 2016. Crossref, https://doi.org/10.1080/02331888.2015.1128937
- [17] Rizwan Mushtaq, "Testing Time Series Data for Stationarity Augmented Dickey Fuller Test," SRN Electronic Journal, 2011 Crossref, https://doi.org/10.2139/Ssrn.1911068
- [18] Archana M. Nayak, Akhilesh Ladha, and Nirbhay Kumar Chaubey, "A Comprehensive Comparative Analysis of Passenger Demand Prediction for Improving the Urban Bus Transportation System (UBTS)," *International Journal of Engineering Trends and Technology*, vol. 70, no. 9, pp. 269-279, 2022. *Crossref*, https://doi.org/10.14445/22315381/IJETT-V70I9P227
- [19] Machine Learning Mastery website, 2022. [Online]. Available: https://machinelearningmastery.com/random-forest-for-time-series-forecasting/
- [20] L. Breiman, "Random Forests," Machine Learning, vol. 45, pp. 5-32, 2001.
- [21] Sunitha Cheriyan et al., "Intelligent Sales Prediction Using Machine Learning Techniques," *International Conference on Computing, Electronics & Communications Engineering (ICCECE)*, pp. 53-58, 2018. *Crossref*, https://doi.org/10.1109/Iccecome.2018.865911
- [22] Awswebsite, 2022. [Online]. Available: https://docs.aws.amazon.com/forecast/latest/dg/aws-forecast-recipe-ets.html.
- [23] Physical Layer (PHY) Specification, IEEE Std, vol.12, no. 11, pp. 260-280, 1997.
- [24] Ayesha Saleem et al., "Challenges and Strategies for Sales Prediction in Apparel Industry," SSRG International Journal of Computer Science and Engineering, vol. 6, no. 8, pp. 7-11, 2019. Crossref, https://doi.org/10.14445/23488387/IJCSE-V6I8P102
- [25] Elcio Tarallo et al., "Machine Learning in Predicting Demand for Fast-Moving Consumer Goods: An Exploratory Research," International Federation of Automatic Control IFAC-Papers online, vol. 52, no. 13, pp. 737-742, 2019. Crossref, https://doi.org/10.1016/j.ifacol.2019.11.203
- [26] Taewook Lee, "Wild Bootstrap Ljung–Box Test for Residuals of ARMA Models Robust to Variance Change," Journal of the Korean Statistical Society, vol. 51, pp. 1005–1020, 2022. Crossref, https://doi.org/10.1007/s42952-022-00172-6
- [27] The Inventorowebsite, 2022. [Online]. Available: Https://Inventoro.Com/Biased-Forecasts-Can-Cause-Higher-Inventory-and-Logistics-Costs/