Applications of supervised algorithms for sales prediction in small business - Santo Domingo, Ecuador

Mikel Ugando Peñate [1[0000-0002-3021-0717]] , Ángel Ramón Sabando García [1[0000-0001-5438-9590]], Reinaldo Armas Herrera [2[0000-0002-3477-5838]], Angel Alexander Higuerey-Gómez [2[0000-0003-0031-8898]], Elvia Rosalía Inga Llanez [2[0000-0002-9853-4363]], Pierina D'Elia-Di Michele [2[0000-0001-7763-7577]], Antonio Villalón Peñate [1[0000-0002-5746-1145]], Kent Bryan Gualapuro Burga [1[0009-0003-4402-9067]], Cristian Mauricio Tinoco Diaz [1[0000-0001-5392-9743]]

[1] Pontificia Universidad Católica del Ecuador- Sede Santo Domingo (PUCESD) Av. Chone Km 2, Santo Domingo 230203, Ecuador
[2] Universidad Técnica Particular de Loja (UTPL), San Cayetano Alto, Loja, 110150, Ecuador
mugandop@pucesd.edu.ec, arsabando@pucesd.edu.ec, ahreinaldo@utpl.edu.ec, aahiguerey@utpl.edu.ec, eringa1@utpl.edu.ec, pdelia@utpl.edu.ec, avillalonp@pucesd.edu.ec, kbgualapurob@pucesd.edu.ec , cmtinocod@pucesd.edu.ec

Abstract. Small and medium businesses in Ecuador contribute greatly to the generation of employment and important contributions to the economy in general of the country. The objective of the following investigation is the application of supervised algorithms for the prediction of sales in a small business minimarket located in the province of Santo Domingo de Los Tsáchilas.
The research methodology follows a predictive quantitative approach, through the use of ARIMA time series, with a non-experimental and cross-sectional design, having as available information 1,825 sales dynamics that left the business during the years 2018 to 2022, obtaining a financial sample of 60 observations. The partial results allow to visualize a non-stationary time series with seasonality and became stationary with a difference of moving means. The final predictive model was an ARIMA (0,1,0) (1,0,1).
As a general conclusion of this investigation, it was obtained that the sales forecasts in the small business are simultaneous in the timeline, evidencing seasonality and a sales growth trend of more than 3%.

Keywords: Data analysis, scientific statistics, company, mathematical model, economic forecast

1 Introduction

Nowadays, Artificial Intelligence methods used in forecasting accuracy are higher except for irregular series. Traditional supervised ARIMA methods were found to be more

accurate in forecasting financial variables, unlike CNN (Convolutional Neural Networks), LSTM (Long Short-Term Memory), and MLP (Multilayer Perceptron) machine learning models [1]. It has been observed that the supervised model that performs best with different data sets is the Random Forest [2].

The autoregressive integrated moving average model (ARIMA) has a stochastic characteristic that differs from an autoregressive conditional heteroskedasticity (ARCH) model. For the first model case, it is required that the analyzed series be stationary (that complies with constant mean and variance and the existence of covariance between periods, but that the latter approaches) [3].

The theoretical usefulness of ARIMA models in modeling historical sales series is novel and essential in the business field; anticipating the future is the desire of any businessman that, combined with empirical knowledge, strengthens economic and financial science [4].

The scientific literature on supervised and unsupervised algorithms has been used extensively in recent years. The prediction of patterns as accurately as possible proposed by [5], able to predict future behaviors, semi-supervised as a middle ground between supervised and unsupervised learning [6], approaching patterns of historical data events that impact forecasts of financial and non-financial variables, [7].

Supervised or semi-supervised models are more efficient to have the best accuracy in categorizing documents set [8], thus in the prediction of financial variables [9]. In this sense, semi-supervised models would become more valuable than unsupervised models for theory-based text analysis [10]. Machine learning is a powerful branch of AI that has been successfully used in several industries, such as manufacturing [2].

Even though it is evident that the use of algorithmic methods can improve the quality of systems and processes, an adequate monitoring process is still needed to validate the implementation of these technologies [11]. With the aim of having parameters and assumptions for a reliable supervised econometric model for predictions, [12] have analyzed the traditional and non-traditional fruit exports over time for Ecuador.

The framework of the prediction, the supervised models with the ARIMA time series algorithm, generates an optimistic trend to forecast financial variables in the future [13]. The Dickey-Fuller unit root test is used to determine the stationarity for the residuals of the ARIMA model due that the series of the null hypothesis has a unit root is rejected and concluded that it is a stationary series because it shows a probability of less than 5% [14]. In the health sector, the best outcome of predicted Ebola victims is a Random Tree Classifier, with a mean absolute error (MAE) score of 7.85%, a root mean squared error (RMSE) value of 61.14% and a directional accuracy of 85.99% [15].

The conjecture underlying most semi-supervised learning algorithms is closely related to each other and how they relate to the well-known assumption of semi-supervised clustering [16]. RStudio software can work with supervised and unsupervised algorithms, and its capabilities are equal to or better than those provided in many end-user software applications, [17] and [18]. The supervised or semi-supervised algorithm uses statistical techniques in a new mode, leading to an exhaustive applicable algorithm [19].

The semi-supervised method estimates forecasts more effectively, showing significant improvement in prediction accuracy compared to traditional methods [20].

According to the technique of supervised methods for a time series model, the outcomes presented by [21] show that the implemented model exceeds the existing models in terms of qualitative parameters such as mean squared error (MSE), causing efficiency in the predictions of Ecuadorian incomes.

Several studies show that individual models are still appropriate for forecasting. However, the best performance comes from composite models like Conv-LSTM-MLP, [22], Sarima model ARIMA (2, 1, 0) ×ARIMA (1, 1, 0), created through a supervised technique, where the upward trend in the fit is highlighted, with a relative mistake of around 2%, which is close to optimal [14]. [23] realize essential contributions with the application of three supervised learning algorithms in the prediction of sales of Ecuadorian shrimp, [24] highlighted the usefulness of temporary models for APPLE share price prediction, being the more significant market capitalization and where textual information also improved the level of share price prediction.

In brief, from the review analysis of several articles, the Decision Tree, Support Vector Machine, and Bayes Algorithms are the most cited, discussed, and applied supervised techniques. On the other hand, K-Medias, Hierarchical Clustering, and Principal Component Analysis also emerged as the most often used unsupervised techniques [25]. Extreme Learning Machines (ELM) are efficient and effective learning mechanisms for the classification and regression of patterns. However, ELM is mainly applied to supervised learning problems [26].

Various articles have studied Ecuadorian SMEs and how to apply the prediction of relevant variables through ARIMA models. Thus, [27] predicted the demand for flowers in a macro way and for a particular company, reaching the conclusion that the Bayesian network is the model with the least error. [28] carried out a case study on an Ecuadorian textile company, comparing different Machine Learning techniques, determining that neural networks have the best performance. [29] applied neural networks to a water bottling company, determining that they are feasible for forecasting the demand for bottled water.

The other sections of this article are organized as follows: in the second section are the objetives; in the third, the methodology; the results in the fourth section; in the fifth, the conclusions and the last section are the references.

2 Objective

According to [30], minimarkets are self-service stores dedicated to selling mass consumer products, offering as advantages their convenience, variety of products and proximity to customers. In Ecuador, a high percentage of microenterprises are predominant, a hugely important part of them as a business. Our research aims to apply supervised algorithms to predict sales for 2023 and 2024 of a small local business, a minimarket located in Santo Domingo de Los Tsáchilas province in Ecuador.

The use of ARIMA models in small businesses is quite important because it allows predicting variables of interest such as sales, so that the entrepreneur can plan his financial management more efficiently and avoid incurring redundant costs. [12] and [13]

have used this methodology to generate forecasts, obtaining good results in terms of precision.
Data from a single micromarket was used because it was the data available. In general, companies and even more so SMEs do not share sales data because they are considered a sensitive variable within management. Some articles have used the analysis of a company to predict the relevant variables, [27] and [28].
The importance of this article lies in the fact that it is one of the first articles in Ecuador that tries to predict the sales of a commercial company, since to our knowledge, other articles have dealt with the agricultural sector [27] or the manufacturing sector, [28] and [29]. The commercial sector has its particularities from the point of view of management and does not behave in the same way as other economic sectors.

3 Methodology

3.1 Methodology and data subsection

This research used an exploratory quantitative, confirmatory and predictive approach. ARIMA time series has allowed the prediction of sales levels until 2024. The data was collected from the beginning of 2018 until the end of 2022 for a total of 1825 sales dynamics that were then grouped every month, with a sample of 60 observations.
The hypothesis of the ARIMA models is that the series can be adjusted to an autoregressive, integrated and moving average scheme, so that the time series becomes stationary, and is explained by the lags of the variable of interest and by the error. in regression. To determine which is the best model, information loss criteria such as Akaike or Bayesian are used.
The retail industry in Ecuador is very fragmented, where large commercial chains coexist with micromarkets, which are often self-starters. By 2023, sales are expected to grow compared to the previous year after having left the Covid-19 pandemic behind. It is the sector of the economy with the largest number of employees [31].

3.2 Statistical analysis

The statistical and predictive process was realized through RStudio software; the first step was to conduct an exploratory analysis of quantitative sales data, with the purpose of determining the average and the standard deviation from 2018 to 2020, for the parametric hypothesis, [32]; and the time series ARIMA in financial or econometric process, [33].

Through this process, the normality test was executed using the Shapiro Wilk test because there are 12 sales data or each year, and by default, a normality test should be applied for small samples, in accordance with the criteria of [34].

Consequently, the sales variable was transformed to a time series in months, in which it was noticed that it was a non-stationary event and applying the Dickey-Fuller test to a difference, it was made stationary, i.e., achieved that the mean and variance are constant, [4]. In the framework of the time series for sales, a forecast of the data was made

and contrasted with the actual data, having a similar behavior of sales as a function of time.

In RStudio's algorithm for the ARIMA model, the Auto Arima function was used. By default, the model was selected in a series of combinations adjusted to the Akaike (AIC) and Bayes criteria (BIC) and with the maximum likelihood technique because the past data help forecast the future [35]. Having obtained the best model, sales forecasts were made. In addition, the Ljung Box test must guarantee the independence of the residuals of the ARIMA predictive model [36].

4 Results

Table 1 shows the sales achieved by the minimarket, registered monthly for 2018 until 2022. Beginning with an exploratory data analysis to determine the average, standard deviation, minimum, maximum, summation, and sequentially the normal distribution of the data, with the idea of applying the best time series technique.

For 2018, the minimarket presented an average of $ 29497.79 with a standard deviation of $ 6977.34, demonstrating a low data variability. In the year 2019 showed an average of 22696.49 with a variation of $ 3065.77; where there is a minimum standard deviation, it means that for this year, sales are normalized.

The sales for the year 2020 registered an average of $ 16,012.48 and a dispersion of $ 6,032.08, and this high variance is due to the effect of the pandemic on the consumer that sometimes had to buy most of their food for a month to avoid infected by COVID-19. Throughout 2021 sales at the minimarket followed a decrease, representing an average of $ 14560.62 and a standard deviation of $ 1469.97; with low variability, the data is normal distributed.

In 2022, sales increased with an average of $ 17917.28 and a standard deviation of $ 4,175.99. This increase is probably because the people of Santo Domingo de Los Tsáchilas are recovering from the economic crisis generated by COVID-19.

The normal distribution for sales is confirmed through the Shapiro-Wilk test of normality. Table 1 shows the normality for each study period, exceeding the 5% significance of the probabilistic value.

The nature of this normality test is that the number of cases is small for each year. Although in the year 2020 sales presented irregular movements, i.e., with high and low sales, this phenomenon is attributed to COVID-19, which caused atypical conditions for consumers and suppliers fear of getting sick when making a buying or selling transaction.

Table 1. Financial dynamics for sales of Minimarket.

Months	Sales/2018	Sales/2019	Sales/2020	Sales/2021	Sales/2022
1	20 646,57	21 451,40	23 328,26	12 307,10	15 516,02
2	16 651,20	16 179,18	21 934,28	12 632,34	14 873,69
3	25 657,77	18 187,22	22 206,05	14 051,76	14 559,89
4	250 94,63	24 984,24	6 835,23	13 348,18	17 901,47

5	32 197,14	22 382,10	7 270,00	14 414,36	15 585,72
6	32 910,42	27 322,02	6 698,33	14 343,94	11 419,37
7	39 105,13	23 852,12	18 460,54	15 128,27	18 342,48
8	36 885,48	24 282,91	19 377,17	15 020,16	17 908,62
9	38 836,86	24 804,77	18 164,42	15 495,45	17 885,34
10	29 477,23	23 396,77	16 500,77	17 128,15	20 585,53
11	27 284,36	21 433,46	14 365,23	16 783,09	23 789,08
12	29 226,65	24 081,67	17 009,50	14 074,60	26 640,10
Total	**353 973,44**	**272 357,88**	**192 149,77**	**174 727,41**	**215 007,31**
Average	29 497,79	22 696,49	16 012,48	14 560,62	17 917,28
Standard deviation	6 977,34	3 065,77	6 032,08	1 469,97	4 175,99
Shapiro-Wilk	0,839	0,317	0,067	0,823	0,458

As highlighted by the results presented above, the normality for sales, the ARIMA model is the right one to forecast the financial future of small businesses in Santo Domingo.

Figure 1 shows the behavior of the observed values of sales for the period from 2018 to 2022, showing that during 2018 the highest sales were reported and causing a slight decrease for 2019, while the worst scenario was visualized for 2020. The COVID-19 pandemic most likely caused this high decline in sales. After that, sales have had minimal growth until the end of 2022.

Afterward, a figure is made between the predicted values (see Fig. 1), clarifying that these events have a similar behavior from 2018 to 2022, as they have simultaneous behaviors with the observed values. This growth is evident from the latest data and the data from the recent past that impact the coefficients of the ARIMA model. This behavior of the data allows to see the creation of an ARIMA model or, on the other hand, the generation of a seasonal autoregressive integrated moving average model (SARIMA) by witnessing a behavior of events that repeat at a specific date, known as seasonality.

As for the financial variable sales (see Tab. 2), it is considered that this time series for a financial prediction calculation presents properties of stationarity, whose behavior defines the time variable with constant mean and variance, and in turn, this process is achieved with the Dickey-Fuller test.

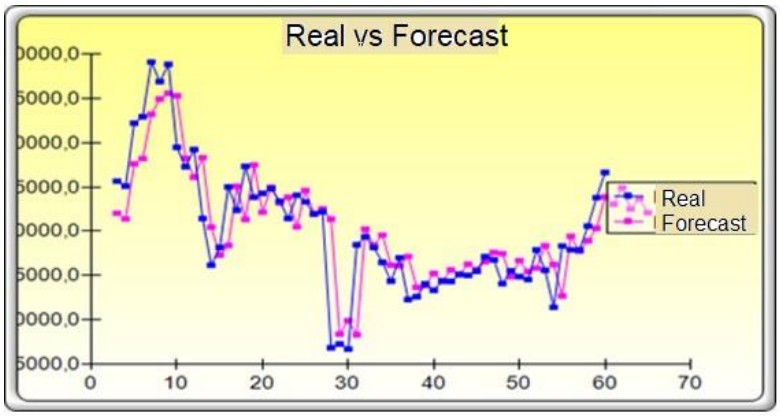


Fig. 1. Dynamics of observed and expected sales.

Table 2. Dickey-Fuller test for sales.

Test de Dickey-Fuller			
Variable	DF	Delay	p-value
Sales (times series)	-2,2439	3	0,4762
Sales (1 difference)	-4,2512	3	0,0100

In this way, in the beginning, temporary sales present a non-stationary event ($p<0.05$); with a mean and variance that are not constant, and, by default, this series is not advisable for forecasting. Therefore, applying a unit root of the Dickey-Fuller test was necessary to generate stationary sales through a difference in the moving average ($p<0.05$). In this same setting, when presenting this condition, the sales present a trend and, in the same way, integrate the seasonality, producing a model of integration of order 1 (I=1).

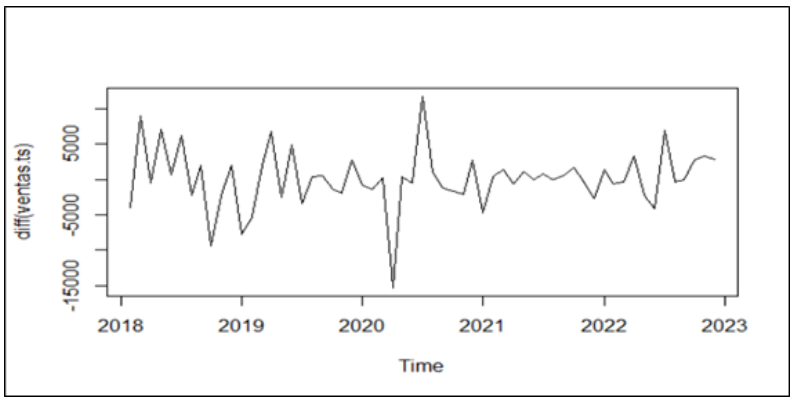


Fig. 2. Stationary series of sales, by a difference of moving average.

In Figure 2, it is evident that the sales follow a stationary process, and this was achieved through a moving average difference (1 difference = I), with the Dickey-Fuller test shown in Table 2. Thus, predictions can be made with higher reliability by having a constant average and variance. Moreover, in the same way, data from the recent past contributes more to the future. In the same way, parametric prediction model is used, as in this case is the Box-Jenkins model applying the ARIMA model.

Later, a figure is made between the observed and forecasting values, making it clear that these events have a similar behavior from 2018 to 2022. By having simultaneous behaviors, growth is evidenced from the year 2020, and the recent past data impacts the coefficients of the ARIMA model, generating the seasonality, AR order (1), and MA order (1). Therefore, this also leads to forming a SARIMA model (0,1,0) (1,0,1).

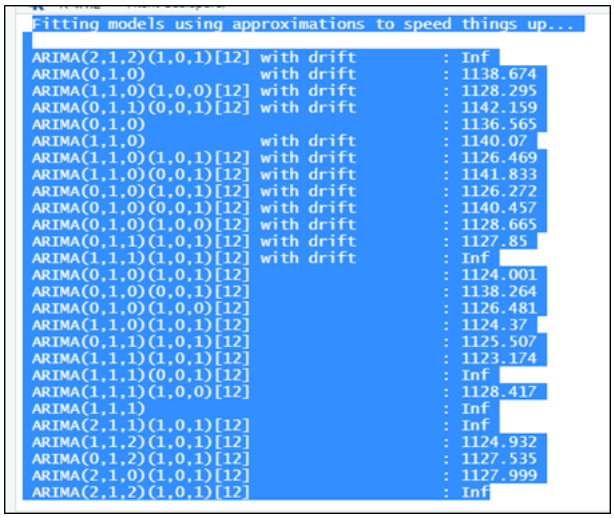


Fig. 3. Sales Iterations for ARIMA Simulation.

Regarding generating ARIMA models (see Fig. 3), the RStudio programming package gives countless financial simulations with the self-modeling function. In this way, it generated 25 iterations for sales through the autoregressive moving average model (ARIMA), with and without derivatives. In this stage, the simulations that generate higher reliability are those that have derived processes and, by default, show the lowest information criterion (Inf=1126,272), which represents the model ARIMA (0,1,0) (1,0,1).

It can be deduced that small businesses have seasonal behavior and light growth that can be said for all financial variables to realize forecasts. In this way, the SARIMA simulation is similar to the same properties of an ARIMA due to its autocorrelation, arithmetic average, seasonality component, and sometimes trend component when performing the stationary series [3]. In the coefficients frame, (see Fig. 4), autoregressive stationarity (SAR=-0.8828) was formed, with an intense change for the average sales (SMA=0.7710) and with the presence of stationarity.

```
 ARIMA(1,1,1)(1,0,1)[12]                    : Inf
 ARIMA(0,1,0)(1,0,1)[12]                    : 1155.786

 Best model: ARIMA(0,1,0)(1,0,1)[12]

Series: ventas.ts
ARIMA(0,1,0)(1,0,1)[12]

Coefficients:
         sar1    sma1
      -0.8828  0.7710
s.e.   0.5414  0.6944

sigma^2 = 17063819:  log likelihood = -574.67
AIC=1155.35   AICc=1155.79   BIC=1161.58
```

Fig. 4. Sales coefficients for the ARIMA simulation (0,1,0) (1,0,1).

In the coefficients frame, (see Fig. 4), autoregressive stationarity (SAR=-0.8828) was formed, with an intense change for the average sales (SMA=0.7710) and with the presence of stationarity.

In the same line, the predictive model SARIMA (0,1,0) (1,0,1), through maximum likelihood test, showed an AIC=1155.35; AICc=1155.79 and BIC=1161.58). These coefficients and model reliability tests confirm that the parametric Box-Jenkins models make trust forecasts for small businesses while showing that these businesses show seasonality, i.e., sales cause optimistic or pessimistic scenarios in a specific period.

Following the reliability of the ARIMA model (0,1,0) (1,0,1), figure 5 shows the graphical representation of the residuals. The graph of the residuals shows few dispersions; In the same way, the simple autoregressive graph (PCA), whose bars are below the upper and lower bands, deducts the normality of the data. On the other hand, the histogram graph shows the highest concentration of data around zero, although some outliers can be seen.

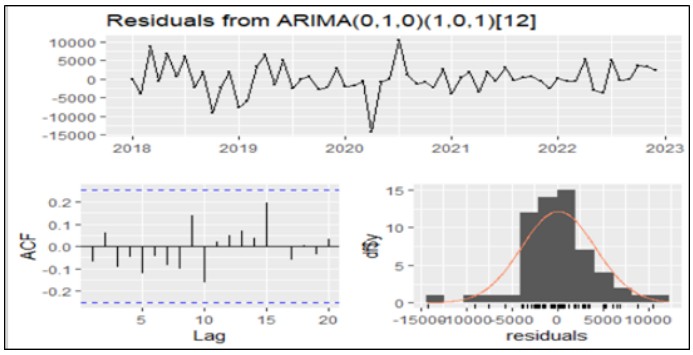


Fig. 5. Residuals test for model errors.

Posteriorly, the Kolmogorov-Smirnov normality test was applied for the errors of the ARIMA model (0,1,0) (1,0,1), confirming the hypothesis the null hypothesis is not rejected, that the residuals follow a normal distribution.

The Ljung Box test represents a stage of no autocorrelation of the residuals ($p>0.05$); they are random walk behavior for the time series model for sales (see Tab. 3). In other words, the residuals are independent for each sale, with zero mean and constant variance.

Table 3. Ljung Box test for forecasts of the ARIMA model (0,1,0) (1,0,1).

Ljung-Box test			
Model residuals	Q	df	p-value
ARIMA (0,1,0) (1,0,1)	7,1544	10	0,7108

The prediction generated monthly by the SARIMA model (0,1,0) (1,0,1) from 2023 to the end of 2024 for the sales of a minimarket. Figure 6 shows a positive variability of the average sales of the minimarket. For the prediction case, at 80% reliability for the financial variable sales, at the beginning of January 2023, we have an average of $ 25561.23 with a lower limit of $ 20250.687 for a pessimistic scenario and an upper limit of $ 30871.77 for an optimistic scenario.

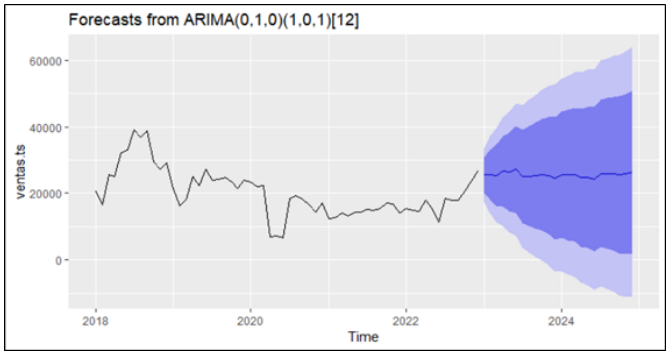


Fig. 6. Sales forecasts for the ARIMA model (0,1,0) (1,0,1).

In the same way, (see Fig. 7), about the monthly forecasts for the year 2024, shows an average of $ 26402.63 with a lower confidence interval of $ 1874.838 and an upper confidence interval of $ 50930.43. Besides, if we evaluate the impact of the variation in sales between January 2023 and December 2024, we have a growth rate of 3.19%.

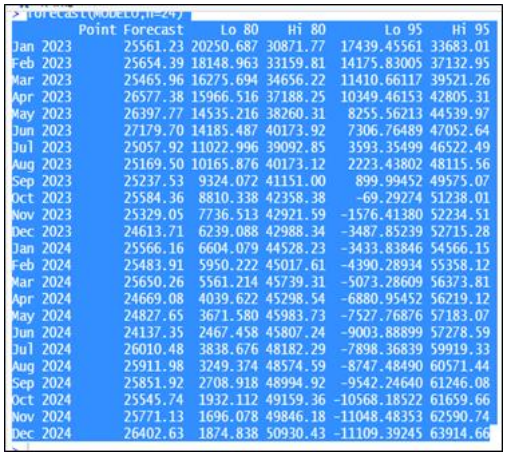
The results obtained allow us to predict an average sales growth above 3 percent, although they are below the latest estimates [31]. This is because micromarkets do not have the same ability to sell as large stores that can give discounts, give credit and other commercial strategies that encourage sales. The methodology used has been tested by [27] and [28], comparing the ARIMA method with other Machine Learning techniques,

but in other sectors. The results of the ARIMA model do not differ excessively from those obtained with other prediction techniques.

This is one of the first articles used to predict sales of small businesses and the ARIMA methodology is justified by its temporal development, making it a highly contrasted methodology, and the most administrative-economic of the managers of commercial companies, compared to the high knowledge of statistics and computer science that other Machine Learning methods imply. The end of this article is to demonstrate that through relatively simple techniques, managers of any business can predict the sales of their businesses.

Given the current predictions, it will help managers make decisions in favor of other financial variables such as increasing inventory, increasing financial debt in banks, adjusting accounts receivable, or incorporating or improving market strategies to overcome the estimated situation.

If the company correctly predicts sales, it can carry out consistent financial planning and develop business models that are more adjusted to fluctuating demand, that is, it can optimize its business model, not run out of inventory and increase sales based on new sales policies. credit adjusted to the customer's credit score.



	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2023	25561.23	20250.687	30871.77	17439.45561	33683.01
Feb 2023	25654.39	18148.963	33159.81	14175.83005	37132.95
Mar 2023	25465.96	16275.694	34656.22	11410.66117	39521.26
Apr 2023	26577.38	15966.516	37188.25	10349.46153	42805.31
May 2023	26397.77	14535.216	38260.31	8255.56213	44539.97
Jun 2023	27179.70	14185.487	40173.92	7306.76489	47052.64
Jul 2023	25057.92	11022.996	39092.85	3593.35499	46522.49
Aug 2023	25169.50	10165.876	40173.12	2223.43802	48115.56
Sep 2023	25237.53	9324.072	41151.00	899.99452	49575.07
Oct 2023	25584.36	8810.338	42358.38	-69.29274	51238.01
Nov 2023	25329.05	7736.513	42921.59	-1576.41380	52234.51
Dec 2023	24613.71	6239.088	42988.34	-3487.85239	52715.28
Jan 2024	25566.16	6604.079	44528.23	-3433.83846	54566.15
Feb 2024	25483.91	5950.222	45017.61	-4390.28934	55358.12
Mar 2024	25650.26	5561.214	45739.31	-5073.28609	56373.81
Apr 2024	24669.08	4039.622	45298.54	-6880.95452	56219.12
May 2024	24827.65	3671.580	45983.73	-7527.76876	57183.07
Jun 2024	24137.35	2467.458	45807.24	-9003.88899	57278.59
Jul 2024	26010.48	3838.676	48182.29	-7898.36839	59919.33
Aug 2024	25911.98	3249.374	48574.59	-8747.48490	60571.44
Sep 2024	25851.92	2708.918	48994.92	-9542.24640	61246.08
Oct 2024	25545.74	1932.112	49159.36	-10568.18522	61659.66
Nov 2024	25771.13	1696.078	49846.18	-11048.48353	62590.74
Dec 2024	26402.63	1874.838	50930.43	-11109.39245	63914.66

Fig. 7. Sales forecasts for the minimarket store.

5 Conclusions

In this paper, we have sought to predict sales for the years 2023 and 2024 in small local businesses through time series in Ecuador, choosing like the study case 1,825 sales generated by a minimarket in the Santo Domingo de Los Tsáchilas province of Ecuador during the years 2018 to 2022.

It has been obtained as a result that the predictive sales of small businesses are simultaneous in the timeline, evidencing seasonality and a slight growth trend in sales, throwing a predictive ARIMA model (0,1,0) (1,0,1). This model has been chosen using

goodness-of-fit coefficients which has generated acceptable reliability for sales predictions.

The results of the present study may be applied as a reference for those personal businesses engaged in commerce, considering strategies to improve income, and the model applied is an innovative contribution to small Ecuadorian businesses, which are predominant in the country's economy.

For future research, we expect to extend the model with more data and compare it with similar companies to improve the prediction of sales in this type of business. On the other hand, it is interesting to add, besides sales, the movement of other financial accounts, which would allow modeling a complete system in terms of predicting the financial situation of small Ecuadorian businesses.

References

1. Mukherjee, S., Chittipaka, V., Mohan Baral, M.: A structural equation modeling approach to Big Data Analytics adoption by SMEs in India. In: Gupta, D., Goswami, RS, Banerjee, S., Tanveer, M., Pachori, RB (eds) Pattern recognition and data analysis with applications. Lecture Notes in Electrical Engineering, vol 888. Springer, Singapur. (2022).
2. Herrera Roldan, G., Castillo Brito, Y. Review of supervised machine learning applications in the manufacturing industry. Qualitas Journal 21(21), 044 - 056 (2021).
3. León Anaya, L. M. Application of empirical mode decomposition to stock market forecasting with ARIMA-ARCH models and evolutionary artificial neural networks. Universidad Autónoma del Estado de México. (2017).
4. Altamirano Pérez H. R., Morales A., Tovar Pinzón M. E., Yance Gómez L.E.: Application of the ARCH model to sales forecasting, a business approach. Journal of the School of Economics 28(1), 149 - 170 (2022).
5. Salamanca Rativa, I. N.: Machine learning techniques applied to forecasting systems. Technology Research & Academia 8(1), 37–53 (2021).
6. Yan, X., Bai, Y., Fang, S. C., Luo, J. A kernel-free quadratic surface support vector machine for semi-supervised learning. Journal of the Operational Research Society 67(7), 1001-1011 (2016).
7. Padmanabha, YCA., Viswanath, P., Eswara B. Semi-supervised learning: a brief review. International Journal of Engineering &Technology 7 (1.8) 81-85, Núm Especial 8. (2018).
8. Cevallos-Culqui, A., Pons, C., Rodriguez, G.: Semi-supervised learning models for document classification: A systematic review and meta-analysis. Inteligencia Artificial 26(72), 81–111 (2023).
9. Sabando García, A.R., Ugando Peñate, M., Armas Herrera, R., Higuerey Gómez, Ángel, A., Espín Estrella, G. M., Villalón Peñate, A.: Econometric and stochastic modeling of ginger sales forecasts in Ecuador. Research & Development Engineering 22(1), 25–43 (2022).
10. Watanabe, K., Zhou, Y.: Theory-Driven Analysis of Large Corpora: Semisupervised Topic Classification of the UN Speeches. Social Science Computer Review 40(2), 346–366 (2022).
11. Álvarez Vega, M., Quirós Mora, L. M., Cortés Badilla, M. V. (2020). Artificial intelligence and machine learning in medicine. Synergy Medical Journal 5(8), e557 (2020).
12. Sabando García, Á. R., Ugando Peñate, M., Cueva Torres, E. Y., Villalón Peñate, A. Mendoza Esmeralda, G. E., Arias Minda J.: Production modeling and sales forecasts for pitahaya cultivation in Ecuador. Synapsis 12(1), 106-121, (2020).

13. Ugando Peñate, M., Sabando García, A. R., Armas Herrera, R., Higuerey Gómez, A. A., Villalón Peñate, A.: Applied econometric modeling and forecasting of exportable levels for the barraganete plantain in Santo Domingo de los Tsáchilas, Ecuador. Zulia University Journal 14(39), 139-161. (2022).
14. Ruiz Hernández, J. A., Barrios Puente, G., Gómez Gómez, A. A.: Apple price analysis through a SARIMA model. Mexican Journal of Agricultural Sciences 10(2), 225–237. (2019).
15. Pandey, M.K., Subbiah, K.: Performance analysis of time series prediction using machine learning algorithms for Ebola casualty prediction. In: Deka, G., Kaiwartya, O., Vashisth, P., Rathee, P. (eds) Applications of Computing and Communication Technologies. ICACCT 2018. Communications in Computer and Information Sciences, vol. 899. Springer, Singapur (2018).
16. Van Engelen, J.E., Hoos, H.H. A.: Survey on semi-supervised learning. Mach Learn 109, 373–440 (2020).
17. Baturo A., Dasandi N., Mikhaylov S. J.: Understanding state preferences with text as data: Introducing the UN general debate corpus. *Research & Politics* 4(2), 2053168017712821 (2017).
18. Benoit K., Watanabe K., Wang H., Nulty P., Obeng A., Müller S., Matsuo A.: Quanteda: An R package for the quantitative analysis of textual data. Journal of Open-Source Software 3(30), 774 (2018).
19. King G., Lam P., Roberts M. E.: Computer-assisted keyword and document set discovery from unstructured text. *American Journal of Political Science* 61(4), 971–988 (2017).
20. Yao, L., Ge, Z. (2017). Deep learning of semisupervised process data with hierarchical extreme learning machine and soft sensor application. IEEE Transactions on Industrial Electronics 65(2), 1490-1498 (2017).
21. Solís, E., Noboa, S., Cuenca, E. Financial time series forecasting by applying deep learning algorithms. In: Conference on Information and Communication Tech of Ecuador, pp. 46–60. Springer (2021).
22. Noboa, S., Solís, E., Cuenca, E.: Price forecast for Ecuador agricultural products: a comparative study of deep learning models. In: Herrera-Tapia, J., Rodriguez-Morales, G., Fonseca C., E.R., Berrezueta-Guzman, S. (eds) Information and Communication Technologies. TICEC 2022. Communications in Computer and Information Sciences, vol 1648. Springer, Cham (2022).
23. Cordero-Torres, B. P.: Supervised Learning Algorithms for Sales Projection of Ecuadorian Shrimp with Python Programming Language. Business & Economics 13(2), 30–51 (2022).
24. Evaristo Broncano, R.: Application of machine learning techniques for the prediction of APPLE share prices. Research Journal of Systems and Informatics 15(1), 13–22 (2022).
25. Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., Aljaaf, A.J.: A systematic review on supervised and unsupervised machine learning algorithms for data science. In: Berry, M., Mohamed, A., Yap, B. (eds) Supervised and unsupervised learning for data science. Unsupervised and semisupervised learning. Springer, Cham. (2020).
26. Huang, G., Song, S., Gupta, J. N., Wu, C.: Semi-supervised and unsupervised extreme learning machines. IEEE transactions on cybernetics 44(12), 2405-2417 (2014).
27. Herrera-Granda, I. D., Lorente-Leyva, L. L., Peluffo-Ordóñez, D. H., Alemany, M. M. E. A forecasting model to predict the demand of roses in an Ecuadorian small business under uncertain scenarios. In International Conference on Machine Learning, Optimization, and Data Science (pp. 245-258). Cham: Springer International Publishing. (2020, July).
28. Lorente-Leyva, L. L., Alemany, M. M. E., Peluffo-Ordóñez, D. H., Herrera-Granda, I. D. A Comparison of machine learning and classical demand forecasting methods: a case study of

Ecuadorian textile industry. In International Conference on Machine Learning, Optimization, and Data Science (pp. 131-142). Cham: Springer International Publishing. (2020, July).
29. Herrera-Granda, I. D., Chicaiza-Ipiales, J. A., Herrera-Granda, E. P., Lorente-Leyva, L. L., Caraguay-Procel, J. A., García-Santillán, I. D., Peluffo-Ordóñez, D. H. Artificial neural networks for bottled water demand forecasting: a small business case study. In Advances in Computational Intelligence: 15th International Work-Conference on Artificial Neural Networks, IWANN 2019, Gran Canaria, Spain, June 12-14, 2019, Proceedings, Part II 15 (pp. 362-373). Springer International Publishing. (2019).
30. Rojas Adames, L. A., Medina Rojas, F., Sánchez Medina, I. I., Malqui Cabrera, J.: Social Inclusion Engineer: management software for minimarkets. In UTP Congress Reports (51-54) (2016).
31. Primicias (3 October 2023). The productive sector projects a sales increase of 6% in 2023. https://www.primicias.ec/noticias/economia/ventas-incremento-empresas-sectores/
32. Flores Muñoz, P. J.: Compare the efficiency of hypothesis testing and confidence intervals in the inference process. Average study. Science Journal 22(2), 65-85 (2018).
33. Bazán Ramirez, W.: Modeling of the monthly average of quota values by AFP and type 2 fund with the Box and Jenkins or ARIMA methodology. Industrial Data 24(1), 243-276 (2021).
34. Gandica de Roa, E. M.: Potency and Robustness in Normality Tests with MonteCarlo Simulation. Scientific Journal 5(18), 108–119 (2020).
35. Ruiz-Benito, P., Andivia, E., Archambeaou, J., Astigarraga, J., Barrientos, R., Cruz-Alonso, V., Florencio, M., Gómez, D., Martínez-Baroja, L., Quiles, P., Rohrer, Z., Santos, A. M., Velado, E., Villén-Pérez, S., Morales-Castilla, I.: Advantages of Bayesian statistics vs. frequentist statistics: why do we resist to use it?. Ecosystem 27(2), 136-139 (2018).
36. Gutiérrez, R.: Predicting Optimal Cross-Hedging Ratios in the Mexican Oil Market. Mexican Journal of Economics and Finance REMEF 13(1) (2017).