UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

Machine Learning vs Traditional Statistical Method: A Comparison in Demand Forecasting for an Agricultural Company.

Stephano Vallejo Freire

Ingeniería Industrial

Trabajo de fin de carrera presentado como requisito
para la obtención del título de

INGENIERO INDUSTRIAL

Quito, 20 de diciembre de 2022

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

HOJA DE CALIFICACIÓN DE TRABAJO DE FIN DE CARRERA

Machine Learning vs Traditional Statistical Method: A Comparison in Demand Forecasting for an Agricultural Company.

Stephano Vallejo Freire

Nombre del profesor, Título académico Gabriela Baldeón Calisto, Ph.D.

Quito, 20 de diciembre de 2022

© DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica deEducación Superior del Ecuador.

Nombres y apellidos: Bryan Stephano Vallejo Freire

Código: 00203767

Cédula de identidad: 1725837841

Lugar y fecha: Quito, 20 de diciembre de 2022

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en http://bit.ly/COPETheses.

UNPUBLISHED DOCUMENT

Note: The following capstone project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on http://bit.ly/COPETheses.

RESUMEN

El siguiente estudio tiene como finalidad poner a prueba el desempeño de dos métodos para pronosticar la demanda como es Machine Learning y un Software Estadístico Tradicional en una planta de productos agrícolas de Quito-Ecuador. Este trabajo investigativo es dirigido a las personas que estan interesadas en temas como la visualización de datos, la programación y la estadística. Dentro del estudio se aplicó la herramienta de redes neuronales que pertenece a la subcategoría de modelos supervisados para la predicción, llamado LSTM y por otro lado, el modelo Holt-Winters como una herramienta tradicional que se ha destacado a lo largo de los años. El resultado del estudio fue que el modelo LSTM logró destacar bajo las mismas condiciones que el Holt- Winters. Adicionalmente, los resultados demuestran que el futuro de las predicciones apunta al uso de los modelos de aprendizaje automatizado para la industria agrícola. Asimismo, se logra mejorar la planeación agregada, los niveles de producción y la calidad de los procesos productivos de la agroindustria.

Palabras clave: Machine Learning, Redes neuronales, Método estadístico tradicional, Holt-Winters, industria agrícola.

ABSTRACT

Forecasting methods are essential for controlling production levels, inventory overstocks, and inaccurate replenishment. This work picks the best application reviewed for the Machine Learning (ML) method that is most used in the last decade to forecast demand such as Long-Short term memory (LSTM) and compares it with the best traditional forecasting method known as Holt-Winters that fits the features of the given data. According to the applications of forecasting demand, the performance between ML and traditional statistical method is still not fully explored in the Ecuadorian Agricultural Industry. The purpose of this work is the comparison of the LSTM model and Holt-Winters method for improving the efficiency, productivity, and aggregate plan for manufacturing companies of this sector. The data is obtained from historical sales of an Agricultural Company located in Quito, Ecuador. The results show that forecasting with the LSTM model (ML) offers better performance than the traditional statistical method Holt-Winters.

Keywords: Machine Learning, Neural networks, Traditional statistical method, Holt-Winters, Agro-industry.

TABLA DE CONTENIDO

Introduction

The Ecuadorian agro-industrial sector has been an important engine of the nation's productive system, representing 8.6% of the total Ecuadorian production and the 7.7% of the gross national product (FLACSO, 2010). Nevertheless, the agro-industrial sector has not seen important progress in the last decades. Its growth has been affected by environmental degradation, droughts, the El Niño phenomenon, little financing, and the absence of technological resources (Juca, 2021). The lack of technology is essential to understanding the problems this sector faces, as it impacts the efficiency, quality, and competitiveness of the products produced by the industry (Flórez Martínez, 2013).

The agro-industry presents an uncertain demand for its productive sectors caused greatly by the fact that the national food products market has changed for many reasons, such as advances in technology, increases in prices, increase in quality standards, and uncontrolled external factors. In Ecuador, some productive sectors, such as meat, dairy, corn, rice, and sugar cane have a more constant and established demand. However, other productive sectors such as potato, wheat and milling, flour and starch, tomato, legumes, and bamboo have more variability and less competitiveness, and there is insufficient information about their demand (Baquero, 2010). Agricultural companies manage most of these regional sectors, ranging from pre-harvest and harvest to post-harvest.

Ecuador has promoted investment in the agricultural sector to improve national production to medium and small-scale producers, but the evolution of these markets shows that the concentration of the agricultural sector remains constancy and a transformation in the food production cycle is needed for the food production cycle (Muthusinghe, Palliyaguru, Weerakkody, Saranga, & Rankothge, 2018). There are a total of 670.402 agricultural specialized families (small producers) in the 4 regions of Ecuador, representing 58.8% of agricultural small producers (Valle, 2013). The implementation of technological methods by small producers could level up the productivity of small-scale agriculture businesses.

Agriculture companies in Ecuador started using traditional forecasting methods such as linear regression, moving averages, and exponential smoothing almost a decade ago. Since then, with recent technology, these methods have been enhanced through the years, but the industry has not been able to implement all the current tools in their operational and strategic organization (Chávez & Damián, 2021).

Demand forecasting is one of the most essential components of supply chain management. Using the predicted demand, companies elaborate their budget and operation plan, which directly influences their overall performance and competitiveness (Yue, Wangwei, Jianguo, Junjun, & Jiazhou, 2016). Demand forecasting aims to determine the demand pattern of products based on a company´s historical data. Forecasting methods can be divided into traditional statistical methods and machine learning models. Some popular traditional statistical methods include simple moving averages, exponential smoothing, double exponential smoothing, and triple exponential smoothing or Holt-Winters. The model to be applied is selected based on the historical data patterns and the best prediction error (Makridakis S, 2018). The selection of each method will depend on different types of demand patterns, which are classified into linear, stationary, or seasonal trends. Nowadays, there are different methods to forecast each one, but their performance could be better in some cases than others. The traditional methods were the first methods applied in the last decade. However, they reach some conditions regarding the variability and uncertainty of the demand when there is no determined pattern to follow. It turns difficult to forecast an accurate model, these statistical methods work on limited scenarios when there is a certain historical pattern because they need a piece of background information to predict patterns, and the more information the better forecast.

On the other hand, ML models are more precise than traditional methodologies and allow the analysis of bigger databases (Priyadarshi, Panigrahi, Routroy, & Garg, 2019). Supervised machine learning is a subcategory of artificial intelligence and is characterized by using labeled datasets to train a model to predict the response variable. Specifically, the models are trained by optimizing their parameters to minimize an objective function that measures the prediction error. Once a model has been trained, it can be used to forecast unseen observations (Cruz Negrete, 2018). The most common ML algorithm used for demand prediction is the LSTM model which belongs to the Artificial Neural Networks (ANN) models. The LSTM model is well known for processing and making predictions, including important events regardless of the time series trend. In addition, the model unscrambles the variability in the patterns between the predictions and actual values and keeps background information with a feedback connection.

Considering the important economic impact the agricultural sector has on the Ecuadorian industry, it is essential to develop and implement new methods for controlling production levels, inventory overstocks, and inaccurate replenishment. These indicators contribute to overall productivity because controlling them can increase the global efficiency of the agro-industrial production units (Lemos, 2019).

In this work, we compare the ML and traditional statistical models for demand prediction in an Ecuadorian agricultural company. The latter has seen an increase in demand in the last years and therefore needs an accurate model to estimate the sales and adopt a better operational strategy. The applications to be compared are the Long-Short Term Memory (LSTM) model against the Holt-Winters method using the Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE) as evaluation metrics and following the CRISP-DM methodology. The dataset is from a small Ecuadorian company offering a broad range of natural mineral products for agriculture, where 2 lines represent 80% of the demand. The top 5 most demanded products were specially selected from January 2019 to August 2022 for forecasting. The results from the experiments demonstrated that LSTM neural network had a better performance than traditional agricultural product applications.

1.1 Objectives

Develop an accurate model to forecast the demand in an Ecuadorian agricultural company.

Compare the performance of traditional statistical methods against machine Learning algorithms in demand forecasting.

Determine which method is better at forecasting demand.

2. Literature Review

In this section we present the most used method for forecasting demand, dividing them into traditional statistical methods and machine learning methods.

2.1 Machine Learning Methods

ML models have been widely used in the prediction of numerical response variables. In the case of demand forecasting, the last years have seen an increase in the application of various ML methods in different industrial sectors as these models can learn from past examples and capture functional relationships among the data, even if the underlying relationships are hard to describe (Zhang, 1998). A study conducted by Gunawan implemented ANN to predict demand for plastic packaging products in Dynaplastics manufacturing company from 2017 to December 2019, the ANN model gave the smallest error value. It results show an MSE of 25871.97, 28155,91, 78374,01, and 94923,49 for each product and a MAPE of 3.5%, 3%, 4.1%, and 2%. A recent study was conducted in agriculture, where Extreme Learning Machine ELM and Support Vector Regression SVR models were implemented to forecast. The results show that ELM can learn distinct prediction patterns and forecast with a lower error, the value of MSE = 0.0102 and RMSE = 0.01010 in the testing stage of ELM is much less than the SVR with an MSE = 503.4801 and RMSE = 22.4384. (Mostafaeipour, et al., 2020). The LSTM and SVR performed well in terms of accuracy and the degree of association between the predicted and real demand, the LTSM develop better for its ability to transfer the forecast patterns and minimize the error in the test dataset (Kantasa-Ard, Nouiri, Bekrar, Ait el Cadi, & Sallez, 2021). The work proposed by Priyadarshi, Panigrahi, Routroy, & Gatg presents the models of LSTM, SVR, Gradient Boosting Regression GBR, and Xtreme Gradient Boosting Regression XGBR are compared in the prediction of vegetable demand. The LSTM and SVR models had the best results (Priyadarshi, Panigrahi, Routroy, & Garg, 2019). In (Muthusinghe, Palliyaguru, Weerakkody, Saranga, & Rankothge, 2018) compared prediction modules using two machine learning algorithms, namely Recurrent Neural Network RNN and LSTM. The LSTM model had a 0.17 training MSE value, with a 0.37 test MSE value.

The LSTM model is a kind of NN capable of learning long-term dependencies. Remembering information for long periods of time is practically their default behavior. Four layers are interacting instead of one as typical NN are structured. Feature vectors are defined where the benchmark LSTM used hyperparameters settings include the number of timesteps, the number of LSTM layers, the number of neurons, and the limit of maximum length of the feature vector. Therefore, selecting a good set of hyperparameters must be done carefully (Kim & Choi, 2021). Exploring optimal hyperparameters determine the internal structure and affect the performance of LSTM. In general, hyperparameters have a strong correlation, which means that the effect of one feature can be explained through another feature.

2.2 Traditional Statistical Methods

Traditional statistical Methods create predictions based on historical periods (Fattah, Ezzine, Aman, El Moussami, & Lachhab, 2018). The appropriate statistical method to implement depends on the characteristics of the historical demand, such as increasing or decreasing linear trends, stationarity, such as Linear regression, double exponential smoothing, and seasonality, such as the triple exponential smoothing method called Holt-Winters. The study conducted by Mgale, Yan, & Timothy predicted the price of rice with Holt-Winters shows a MAPE 5.33% of prediction error, the data used are from wholesale rice prices in Tanzania for the period January 2004 to September 2019 (2021). Also, the study conducted by Mor, Jaiswal, Singh, & Bhardwaj (2019) applied the forecasting models such as moving average, regression, multiple regression, and Holt-Winters for a group of perishable dairy products in the milk processing industry, the results for winter were an MSE 73561.36 and, MAPE 3.55%. The applications mentioned before are the most used technical method to achieve immediate and accessible forecasting. At this point, it is important to understand the historical patterns of the products to shorten the implementation of traditional methods that are not fitted by their trends. Its results are still limited by the difficulty of the patterns such as seasonality and trend. The intermittent demand makes it difficult to forecast in some cases. Researchers assume some causes that present the results of the models. For example, a study from Waugh on demand forecasting in agriculture mentions that the projections are estimates of the economic situation (1964). The results could be biased on these empirical estimates which is not enough information to process information under uncertain demand to forecast.

The Holt-Winters method captures seasonality and comprises the forecast equation with three smoothing equations: one for the level with the corresponding parameter (α), one for the trend with the corresponding parameter (β), and one for the seasonal component with the corresponding parameter (ɤ). There are two variations, the multiplicative method is preferred when the seasonal variations change proportionally to the level and in the additive method, the seasonal component is expressed in absolute terms in the scale of the observed series (Kotsialos, Papageorgiou, & Poulimenos, 2005). Setting the parameters in the Holt-Winters method must be between 0 and 1. For seasonal and level constants, a high value gives the recent period more weight and low values give more weight to past data. The trend constant must be fitted last for practical purposes.

3. Methodology

The Cross Industry Standard Process for Data Mining (CRISP-DM) methodology is used in this study (Wirth & Hipp, 2000). CRISP-DM was specially designed for Machine Learning projects and is comprised of six steps as shown in Figure 1.

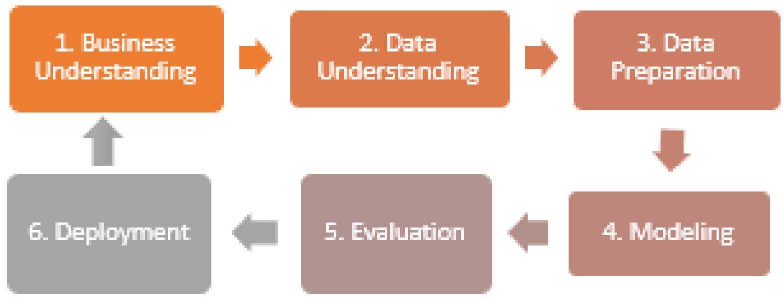


Figure 1. CRISP-DM methodology

The description of each step is in Table 1, by the user guide of CRISP-DM (Schröer, Kruse, & Gómez, 2021).

Table 1. CRISP-DM process model description.

Steps	Description
Business Understanding	This step should assess the business situation to get an overview of the available and required resources.
Data Understanding	In this step, data should be collected from data sources. These data should be explored, described, and their quality assessed. Collecting data from data sources, exploring, and describing it, and checking the data quality are essential tasks in this phase. To make it more concrete. We should describe the data description task by using statistical analysis and determining attributes and their collations.
Data preparation	In this step, the dataset is cleaned. It should be conducted by defining inclusion and exclusion criteria. The bad data quality can be handled by cleaning data.
Modeling	The data modeling step involves selecting the modeling technique and splitting the dataset into train, test, and validation data. For model building, the corresponding hyperparameters must be selected. Moreover, evaluation criteria must be determined to select the best models.
Evaluation	In this step, the results are checked against the defined business objectives and the evaluation criteria. Moreover, the results must be interpreted.
Deployment	The deployment phase consists of planning the deployment of the model, monitoring the results, and maintenance.

4. Data Collection

4.1 Business Understanding

The agricultural company is considered a medium-scale producer of different lines of products such as fertilizers, and agrochemicals for wheat, potato, tomatoes, rice, fruits, citric, and legumes harvests. The company oversees several activities throughout its supply chain, including logistics, distribution, production, sales, and purchase. The company covers the Coast and Sierra regions of Ecuador, and it is growing constantly in both regions offering the most important products and distributing new brand products.

4.2 Data Understanding

The dataset provided by the Ecuadorian agricultural company was from January 2019 to August 2022. And it had production and sales information. The dataset contained 14 variables and 25,847 records. The 14 columns corresponded to the Year, Month, Date, Clients, Cod RTC, Vendor, Number of Receipts, Product Name, Product Line, Type, Quantity, Price, Discounts, and Total. The quantity sold (column 10) is the response variable used for prediction. The description of variables is explained in Table 2.

Table 2. Dataset description

Number of columns	**Column name**	**Description**	**Type of information**
1	Year	The variable Year shows the year of each record.	Date information
2	Month	The variable Month shows the month of each record.	Date information
3	Date	The variable Date shows the date of each record.	Date information
4	Client	The variable Client shows the full name of each client.	Client information
5	Cod. RTC	The variable Cod. RTC shows the zoning code of the records.	Vendor information
6	Vendor	The variable Vendor shows the name of the zone of the records.	Vendor information
7	Receipt	The variable Receipt shows the number of receipts of the records.	Client information
8	Product Name	The variable Product Name shows the product name.	Product information
9	Line	The variable Line shows the product line of the products.	Product information
10	Type	The variable Type shows the type of product either inventory or production. for each record.	Product information
11	Quantity	The variable Quantity shows the quantity of products sold.	Response variable
12	Price	The variable Price shows the price given to the product quantity.	Product information
13	%Discount	The variable %Discount shows the discount on each sale.	Product information
14	Total	The variable Total shows the total amount of sales after discounts and tax.	Product information

4.3 Data preparation

Firstly, the columns with the year, month, clients, cod. RTC, vendor, number of receipts, product name, product line, type, price, and discounts were deleted as they do not present valuable information for the models. The atypical values, defined as those records with inconsistencies such as a date with future years, the atypical values were replaced with the recent year 2022 to take them into account in the model and null values were analyzed before and deleted from the main dataset as well. The null values were deleted because they represented only 2% of the total records in the dataset so there is no significant number of nulls to interfere with the model results.

Once the dataset was cleaned, Pareto charts were created, as shown in figure 2, where a column with the accumulative percentage was calculated as quantity and grouped by the product´s lines 1, 2, 3, 4,5, and 6. This was done by focusing on the best seller’s line of products and ensuring the production planning for the most important line of products.

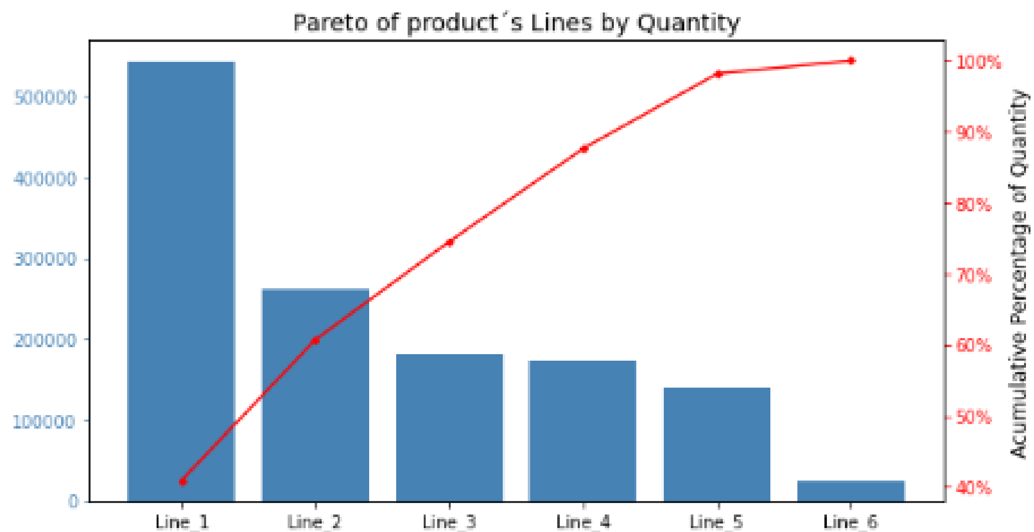


Figure 2. Pareto chart of demand by the line of products.

Secondly, after finding that line 1 and line 2 of products present most of 80% of demand, the top five products were selected for each line to focus on the forecast methods representing 50%.of demand. After preprocessing the dataset, the subsets with the top 5 product information were created. In this step, the date column was summarized in monthly and quarterly periods to observe how the demand is behaving over time as shown in figure 3 and figure 4 below.

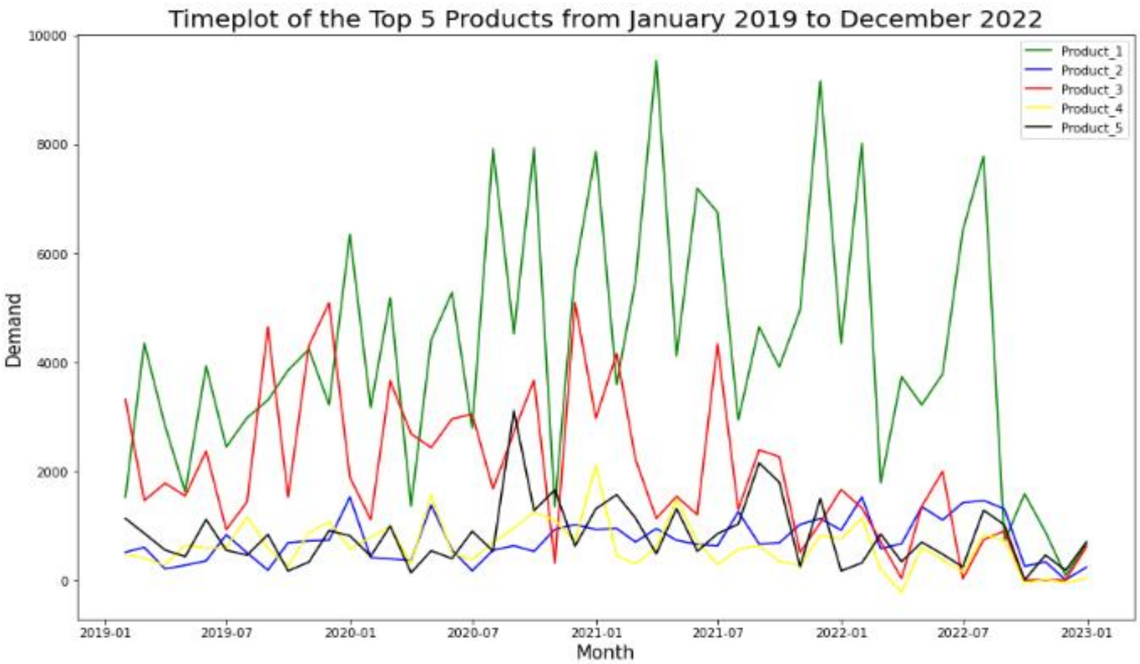


Figure 3. Monthly time plot for the top 5 products.

The monthly time plot shows that there is a growing trend plus a stationary trend with a seasonal period of 3 over the 4 years of historical data. The trend for product number 1,2, and 3 are more stationary than for product 4 and 5. So, the Holt-Winters method fits with the characterized trends. Product 1 is the most important product for the company, it showed large peaks since 2020 as the demand is increasing. Product 3 has a similar pattern to product 1, both belong to the same line of products. On the other hand, for products 2, 4, and 5 there is no growing trend like products 1 and 2 but they still present a stationary trend over time.

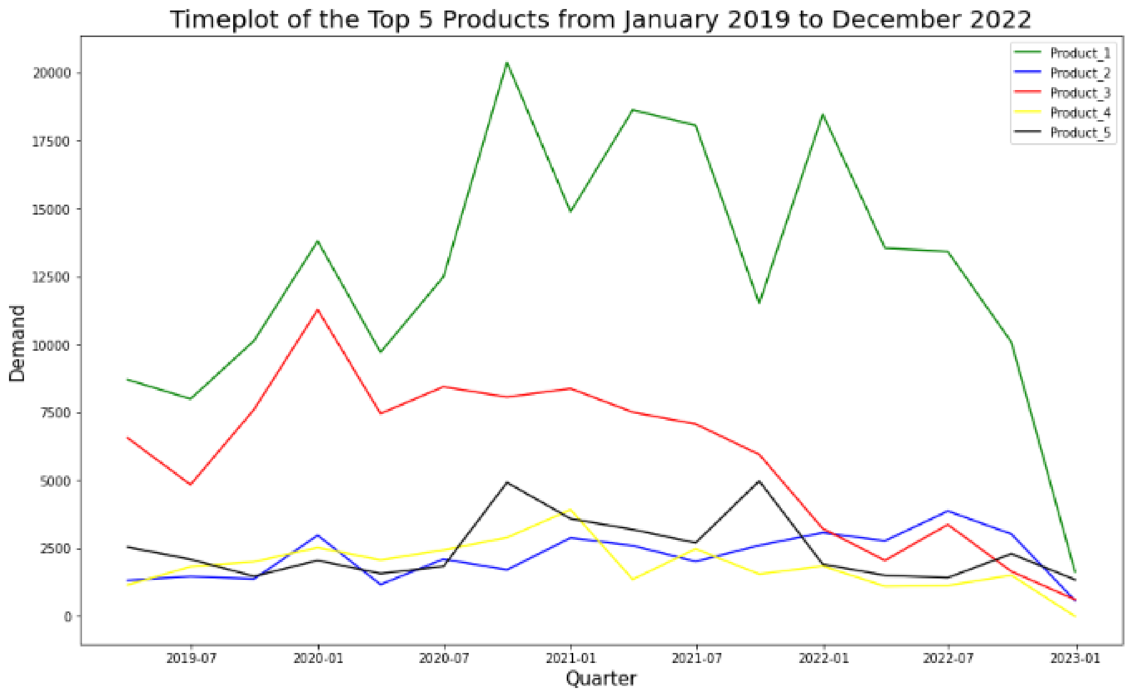


Figure 4. Quarterly time plot for the top 5 products.

The quarterly time plot shows that there is no trend to follow and there is not enough data to forecast either. So, it helps to understand the behavior of the 5 products every 3 months and a constant trend over time is observed but it cannot be modeled for Holt-Winters and LSTM because there are not enough records to process and there is a risk of overfitting the results. The monthly historical data was selected for the Holt-Winters and LSTM models. Before, verifying the trends observed in figure 3, the decomposition of the time series for each product was created for the monthly time plot. This will help to understand how the trend grows and decreases over time and the sequence of the stational trend in the following figures 5,6,7,8, and 9.

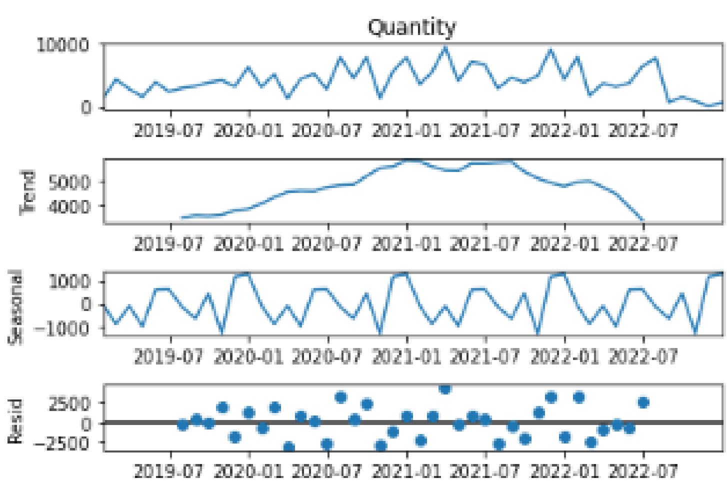


Figure 5. Decomposition of time series of product 1.

In figure 5, product 1 presents a growing trend as its peaks from January 2021 to July 2021, and then a decreasing trend shows until July 2022 and the seasonal trend has 3 high peaks and 3 low peaks over time as well.

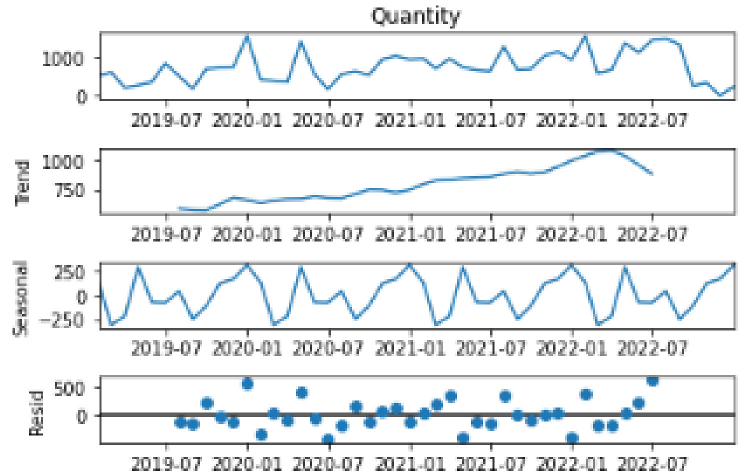


Figure 6. Decomposition of time series of product 2.

In figure 6, product 2 shows a growing trend until July 2022 and the seasonal trend has the same peaks as product 1.

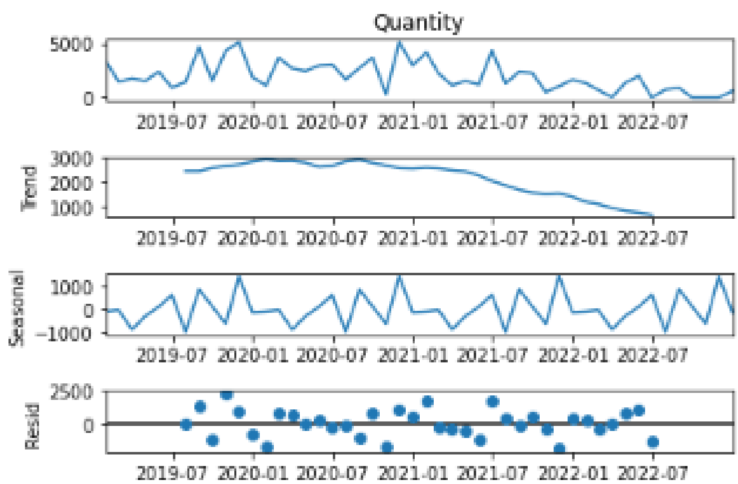


Figure 7. Decomposition of time series of product 3.

In figure 7, product 3 presents a decreasing trend from January 2019 to July 2022 and the seasonal trend has between 3 and 4 peaks each year.

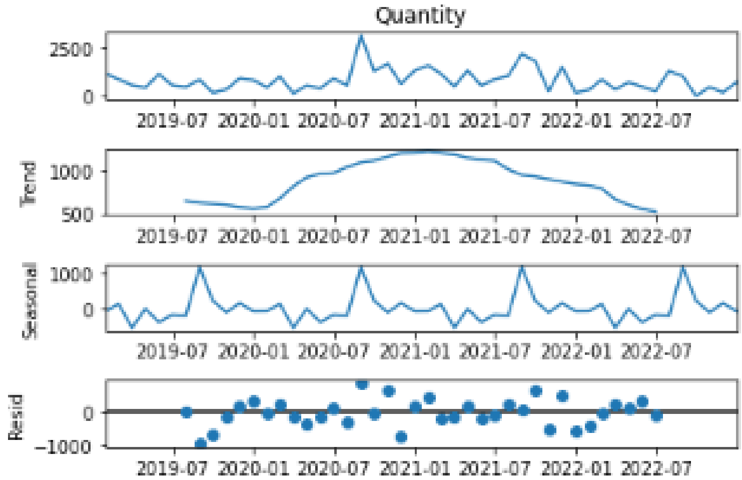


Figure 8. Decomposition of time series of product 4.

In figure 8, product 4 presents a growing trend until March 2021 and after that a decreasing trend until July 2022, its seasonal trend shows 2 peaks over time.

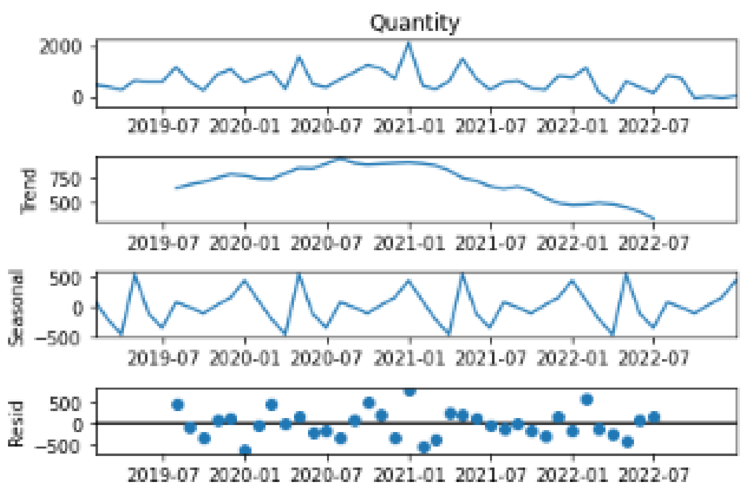


Figure 9. Decomposition of time series of product 5.

In figure 9, product 5 presents a constant trend until February 2021 after that a decreasing trend is showing until June 2022, its seasonal trend shows 3 peaks as products 1 and 2.

4.4 Modeling

For the Machine Learning models, the test set gets the last 12 months of 2021, representing 25% of the dataset, the validation set gets the other 25%, and the remainder 50% belongs to the training set (Torres, 2020).

On the other hand, the dataset was reviewed monthly and quarterly for statistical applications to better understand their patterns and establish initial parameters.

4.4.1 LSTM model

The best fit for hyperparameters and parameters for the models were found to optimize the prediction error. The hyperparameters selected for the LSTM model were a Sequential layer, an LSTM between 90 and 100 neurons, an activation function tanh, a 3-1 blocks for the monthly period, and a Dense layer. The model required also compiling the optimizer Adam and a loss function of MSE to evaluate the model. Additionally, the MAPE metric is used as well (Yu, Si, Hu, & Zhang, 2019).

4.4.3 Holt-Winters

This model uses three exponential smoothing constants, (ɤ) for the trend, (β) for the seasonality, and (α) for the level, the number of periods with a multiplicative method that check the seasonality and trend is also required and it depends on the season period that follows each product (Koehler, Snyder, & Ord, 2001). These constants will change depending on the product's historical data.

4.5 Evaluation metrics

The Evaluation metrics used in this work consist of the Mean Square Error (MSE) and Mean Average Percentage Error (MAPE). "Surveys show that the mean absolute percentage error (MAPE) is the most widely used measure of prediction accuracy in businesses and organizations" (Sanyal, Bosch, & Paquette, 2020).. However, it is important to calculate also the MSE because represents the variation between the sample means. The equations for each metric will be shown below with a brief explanation.

$$\text{MSE}= \frac{1}{N}\sum_{j=1}^{n}\left(y_j - \acute{y}_j\right)^2 \qquad (1)$$

Where:

y_j = real value

$\acute{y}_j$= prediction value

N = number of periods

$$\text{MAPE} = \frac{\sum_{t=1}^{n} \frac{|A_t - F_t|}{|A_t|}}{N} \quad (2)$$

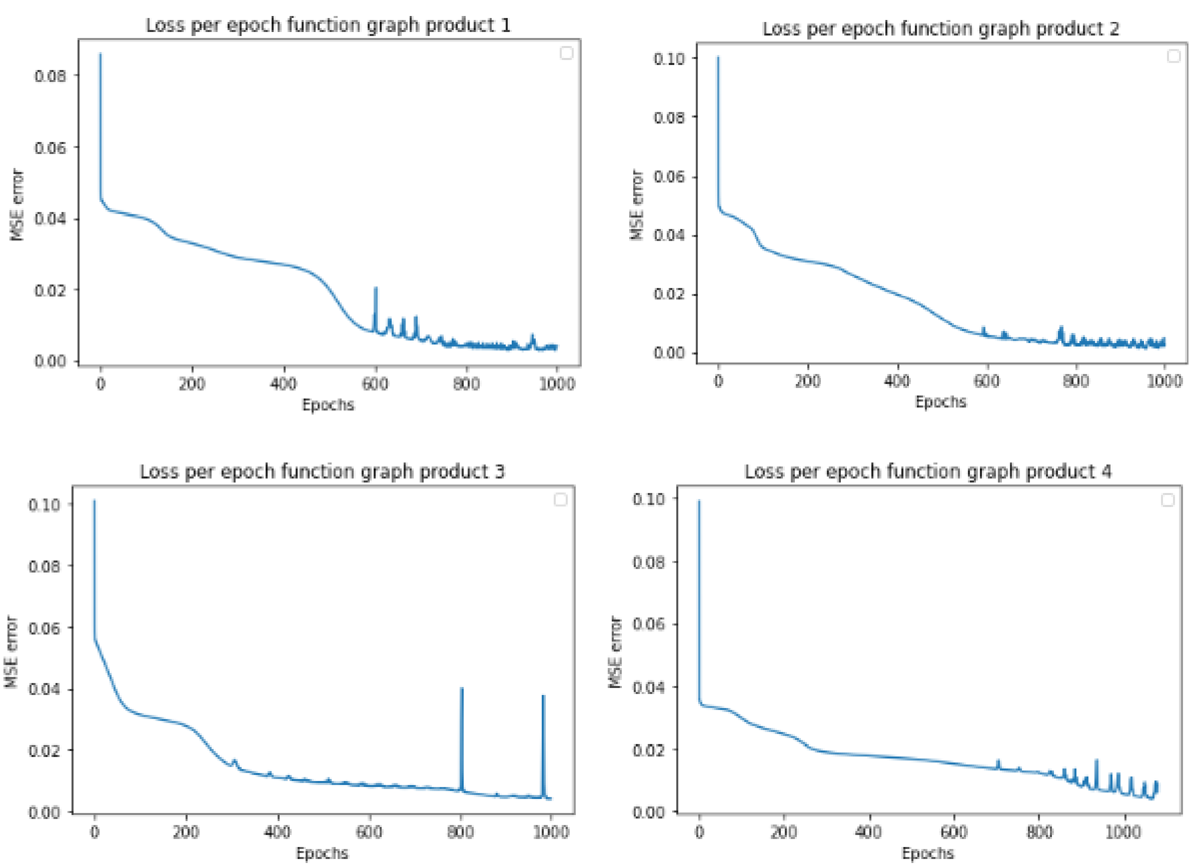
Where:

At = real value

Ft = prediction value

N= number of periods

Additionally, the LSTM model is evaluated by the loss per epoch function, it refers to the loss value over the training data after each epoch. The optimization process is trying to minimize the training set. The lower, the better. For this case, 1000 epochs were needed for each product to get closer to 0.001 MSE. In figure 10, the objective function by MSE iterations for each epoch is minimized over the 5 products as shown below.



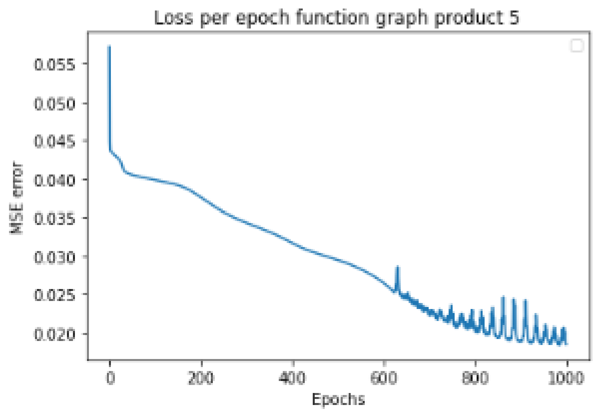


Figure 10. Iterations for epoch for the top 5 products.

4.6 Validation

To validate the results in both methods LSTM model and Holt-Winters, the literature from Industrial and business forecasting methods (Lewis, 1982) was used with a range of MAPE errors is interpreted in the following Table 3. This table has been useful to validate a minimum range of errors for studies and to compare the approach of the results obtained.

Table 3. The acceptable range of MAPE errors.

MAPE	**Interpretation**
<10	Highly accurate forecasting
10-20	Good forecasting
20-50	Reasonable forecasting
>50	Inaccurate forecasting

5. Results

The following figures 11,12,13,14 and 15 show the predictions vs actual values to compare the accuracy of the LSTM and Holt-Winters model for each product.

5.1 Graphical Results

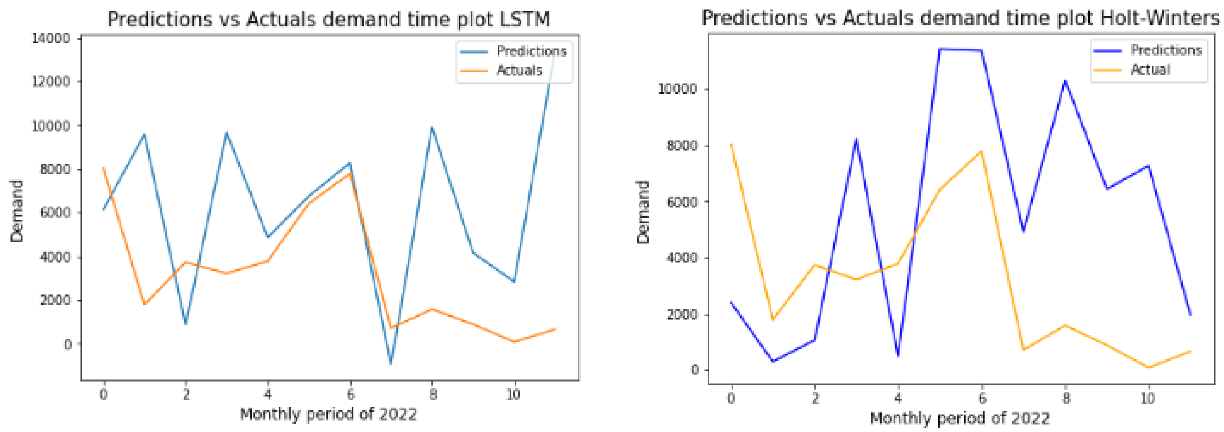


Figure 11. Predictions vs Actual demand time plot for the monthly period of 2022 using LSTM and Holt-Winters model for product 1.

The comparison of prediction and actual errors for product 1 using the LSTM model is highly accurate, the period of March, April, May, and June indicates that LSTM can predict as closely as possible. On the other hand, the comparison of prediction and actual error using the Holt-Winters method presents a little more difference between prediction values. Therefore, the variability of errors increases, and the prediction trend is affected as well.

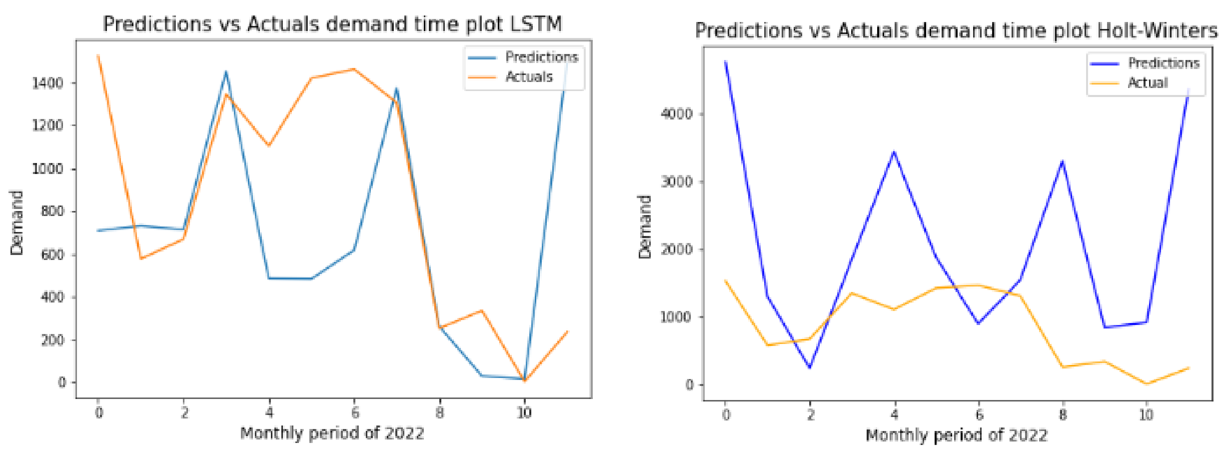


Figure 12. Predictions vs Actual demand time plot for the monthly period of 2022 using LSTM and Holt-Winters model for product 2.

The comparison of prediction and actual errors for product 2 using the LSTM model is highly accurate, the period of March, April, July, August, September, and October indicates that LSTM can predict as closely as possible. On the other hand, the comparison of prediction and actual error using the Holt-Winters method presents a big difference between prediction values. Therefore, the variability of errors increases, and the prediction trend is affected as well as product 1.

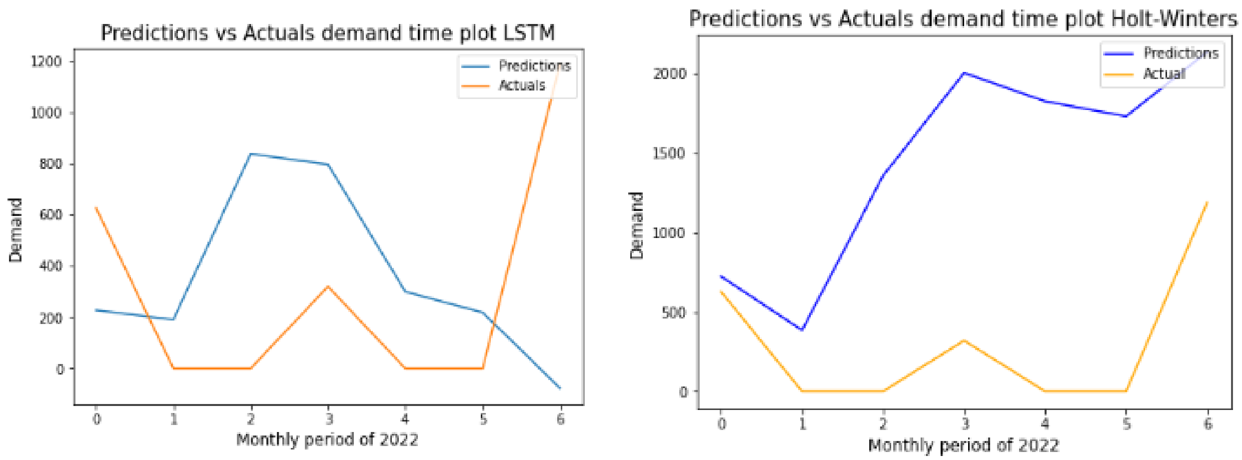


Figure 13. Predictions vs Actual demand time plot for the monthly period of 2022 using LSTM and Forecast results for each product are summarized for the Holt-Winters model in Table 5.

The comparison of prediction and actual errors for product 3 using the LSTM model is good and still accurate, all the periods show that are close. However, the month of June seems to be further, but the LSTM can predict as closely as possible. On the other hand, the comparison of prediction and actual error using the Holt-Winters method presents an underfitting from March to May. Therefore, the variability of errors increases, and the prediction trend is affected as well as products 1 and 2.

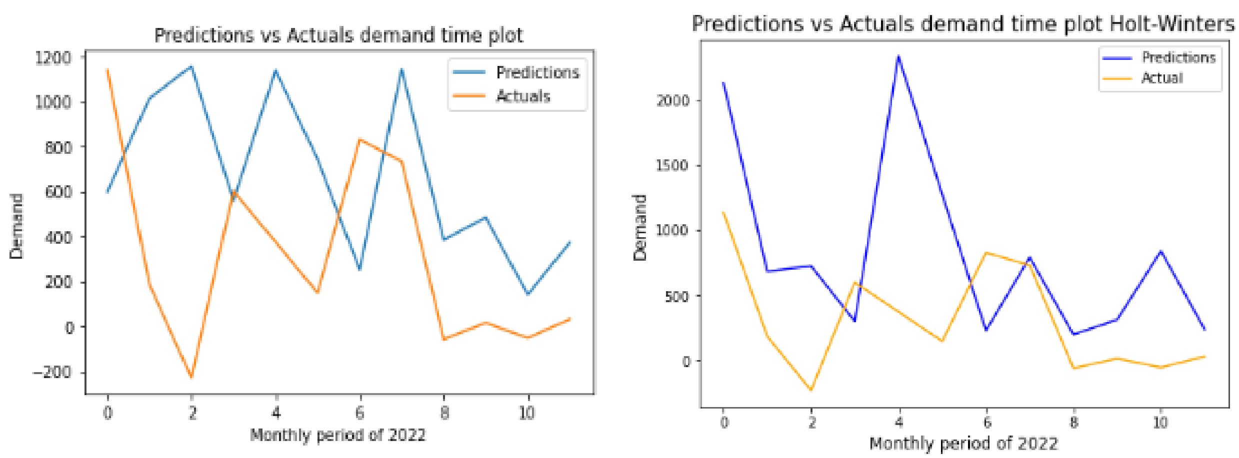


Figure 14. Predictions vs Actual demand time plot for the monthly period of 2022 using LSTM and Holt-Winters model for product 4.

The comparison of prediction and actual errors for product 4 using the LSTM model is good and still accurate, all the periods show that are close. Even though there is a trend with many peaks up and down, the LSTM results follow the trend patterns of the current values. On the other hand, the comparison of prediction and actual error using the Holt-Winters method presents an underfitting in the month of January and May.

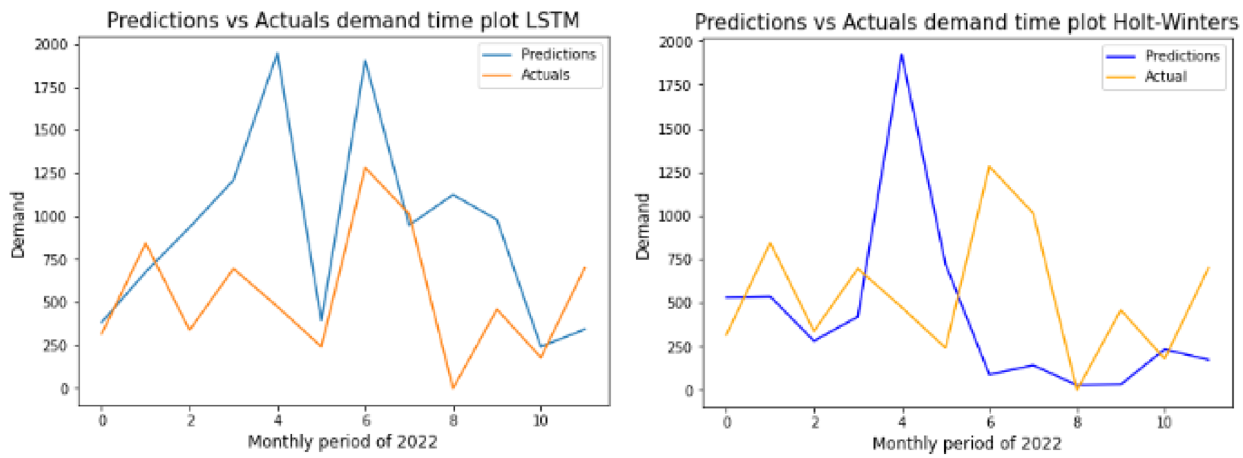


Figure 15. Predictions vs Actual demand time plot for the monthly period of 2022 using LSTM and Holt-Winters model for product 5.

The comparison of prediction and actual errors for product 5 using the LSTM model is good and still accurate, all the periods show that is close unless March, April, May, and August. Even though there is a trend with many peaks up and down, the LSTM results still follow the trend patterns of the current values as well as product 4. On the other hand, the comparison of prediction and actual error using the Holt-Winters method presents an underfitting in May and June, the predictions are closer although always below the actual demand.

5.2 Numerical Results

Forecast results for each product are summarized for the LSTM model in Table 4.

Table 4. Long Short-Term Memory for each product error results.

Product	**Line**	**MSE**	**MAPE**
Product 1	Adjuvants	5544.04	5.89
Product 2	Adjuvants	601.36	0.96
Product 3	Adjuvants	639.05	9.95
Product 4	Biostimulant	639.75	4.22
Product 5	Biostimulant	639.69	5.97

Forecast results for each product are summarized for the Holt-Winters model in Table 5.

Table 5. Holt/Winters for each product error results.

Product	**Line**	**MSE**	**MAPE**
Product 1	Adjuvants	6652.507	8.587
Product 2	Adjuvants	1311.566	10.019
Product 3	Adjuvants	1539.24	11.64
Product 4	Biostimulant	1842.53	12.34
Product 5	Biostimulant	902.93	9.35

The results obtained in the study show that, in all products, the prediction error is explained well within the highly accurate forecasting range of less than 10% compared with the acceptable range of MAPE errors explained in the validation section. The LSTM model performed better than Holt-Winters in all cases. In the case of Product 2 was the best accurate model with the smallest error with 0.96% MAPE and 601.36 MSE. The next product with the smaller error was product 4 with 4.22% and 639.75 MSE. Product 1 scored third place, with 5.89% MAPE and 5544.04 MSE but the same product will show better results using the Holt-Winters method. Product 4 with 5.97% MAPE and 639.69 MSE and the last product 3 with 9.95% MAPE and 639.05 MSE. In addition, the MSE indicates how the variation within the predictions and the actual values are explained, the Product 1 showed a big number of MSE which could be for large differences in samples over time.

In a comparison of the results such as the case of Dynaplastics products (Gunawan, 2021), the MAPE was a little smaller than our 5 products, but there was no significant difference between both studies because they ranged highly accurate forecasting in every case. LSTM theory shows that hidden neurons play an important role in neural algorithms, this role may not necessarily involve iterative tuning of neuron parameters (Mostafaeipour, et al., 2020). However, searching the hyperparameters for the LSTM model was hard to find, to solve this problem, the trial-and-error method was applied because there was no direct method to find an acceptable prediction error. Besides, LSTM hyperparameters from the hyperparameters reviewed in the literature from sugar consumption did not work for this case but it gave a baseline to start from 500 neurons (Kantasa-Ard, Nouiri, Bekrar, Ait el Cadi, & Sallez, 2021). After running the code many times, Analysts could see the coding process as a difficult step to build. Therefore, it presents more effort than the Holt-Winters method eventually does, and this can be considered the main disadvantage of the LSTM model.

On the other hand, the Holt-Winters method results show that the prediction error is explained quite well for products 1 and 5 are within the range of highly accurate forecasting with 8.58%, 9.35% MAPE, and 6652.5, 902.93 MSE. Products 2,3,4 are within the range of good forecasting. The MSE for each product has a sharp increase than the LSTM model, the reason for the MSE increase is given by the loss of demand information over time. These results are still considered good enough for industries. The Holt-Winters method is not harder compared to the LSTM model. Nevertheless, the trial-and-error method was also applied to settle the initial parameters. One possible way of forecasting improvement represents a relaxation of space parameter restrictions, and the recommended values of space are dependent on times series characteristics such as error, trend, and seasonality (Tratar & Strmčnik, 2016), it is important to mention that in each case, the sequence for setting the exponential smoothing constants must go first, the season's constant because it stabilizes most of the prediction error. Secondly, the trend's constant explains the rest of the prediction error, and last, the level constant ensures the optimal error. The main reason to set the parameters in order is to reduce the number of runs for each case; this will also help as a baseline for future works.

The results of both methods were discussed, and the Holt-Winters method still gave accurate forecasts with less effort than the LSTM model. However, when the data set is bigger, the LSTM model has the advantage of its capacity to preprocess and analyze a big data set, the LSTM model can be used to forecast other agricultural products, and the results showed that the proposed method can improve forecast accuracy and can be generalized, it could be used for other crops in which different data would be required. (Mostafaeipour, et al., 2020). The code helped delete those records that are not significant in the model and the data visualization related to the business information also found the most important products and decisions to be taken. Besides, the code can be optimized to save costs related to time and labor. Another advantage of forecasting demand using the ML tool is the variety of analyses that we can create based on the company's goals giving a lot of valuable decisions.

5.4 Proposed Improvements

For proposed improvements, I suggest adding information to the dataset of the company and including more variables that can be considered relevant information for future studies. For example, a new variable could be the harvest destination that is being used in a specific harvest product or the sale season for each product such as winter, or summer. So, the dataset will get more relevant information. On the other hand, I also propose a pre-validation process for the collection of company data. The records should have enough quality to be used in any forecasting method, this step will speed up the preprocessing and visualization steps before modeling the hyperparameters.

6. Conclusion

The agricultural industry in Ecuador has an important engine of the nation's productive system of total Ecuadorian production and a significant value of the Gross National Product. Improving the agroindustry in Ecuador will level up the efficiency, competitiveness, and quality of the national markets. The forecasting method is one of the most important components that industries must apply in their production plans. An appropriate forecasting model offers many opportunities for process optimization and strategic decisions. Monthly and Quarterly values were calculated by the historical demand from the company using the LSTM model and Holt-Winters method. MSE and MAPE have been used as statistics metrics for forecasting accuracy. This work has been emphasized by the comparison of forecasting performance between the most cited methods in the last decade such as the Machine Learning LSTM model vs Holt-Winters a traditional statistic method. The results from both methods showed good forecasting also, it is the most appropriate method applied in each product, whereas the LSTM model is suitable for long-term forecasting as well as short-term forecasting. Besides, the Holt-Winters still shows good forecasting for each product. The previous process shows that the Holt-Winters method has less difficulty than the LSTM model. The presented study was conducted using real data from an Ecuadorian company, and it represents additional value in this paper, the proposed improvements are useful immediately in the real business sector.

References

Baquero, M. L.-P. (2010). La Agroindustria ecuatoriana: un sector importante que requiere de una ley que promueva su desarrollo. *La Granja*.

Chávez, N., & Damián, S. (2021). Deep learning for agricultural products price forecasting: the case of Ecuador. *Bachelor's thesis, Universidad de Investigación de Tecnología Experimental Yachay*.

Cruz Negrete, J. B. (2018). Concentración económica de la agroindustria en Ecuador 2006-2013: un estudio empírico. *Estado & comunes, revista de políticas y problemas públicos*.

Fattah, J., Ezzine, L., Aman, Z., El Moussami, H., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. *International Journal of Engineering Business Management,*.

FLACSO. (2010). *Boletín Mensual de Análisis Sectorial de MIPYMES.*

Flórez Martínez, D. H. (2013). Agenda prospectiva de investigación de la cadena productiva de la panela y su agroindustria. *Tecnura*.

Gunawan, P. A. (2021). Production planning and capacity control with demand forecasting using artificial neural network (Case study PT. Dynaplast) for industry 4.0. *Proceedings of the 11th Annual International Conference on Industrial Engineering and Operations Management*.

Juca, L. A. (2021). Ecuador: análisis económico del desarrollo del sector agropecuario e industrial en el periodo 2000-2018. *Revista Científica y Tecnológica UPSE*.

Kantasa-Ard, Nouiri, M., Bekrar, A., Ait el Cadi, A., & Sallez, Y. (2021). . Machine learning for demand forecasting in the physical internet: a case study of agricultural products in Thailand. *International Journal of Production Research*.

Kim, D. G., & Choi, J. Y. (2021). Optimization of Design Parameters in LSTM Model for Predictive Maintenance. *Applied Sciences*.

Koehler, A., Snyder, R., & Ord, J. K. (2001). Forecasting models and prediction intervals for the multiplicative Holt–Winters method. *International Journal of Forecasting,*.

Kotsialos, A., Papageorgiou, M., & Poulimenos, A. (2005). Long-term sales forecasting using holt–winters and neural network methods. *Journal of Forecasting,*.

Lemos, S. V. (2019). Agroindustrial best practices that contribute to technical efficiency in Brazilian sugar and ethanol production mills. *Energy*.

Lewis, C. (1982). *Industrial and business forecasting methods.* London: Butterworth.

Makridakis S, S. E. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS ONE*.

Mgale, Y. J., Yan, Y., & Timothy, S. (2021). A Comparative Study of ARIMA and Holt-Winters Exponential Smoothing Models for Rice Price Forecasting in Tanzania. *Open Access Library Journal*.

Mor, R. S., Jaiswal, S. K., Singh, S., & Bhardwaj, A. (2019). Demand forecasting of the short-lifecycle dairy products. *In Understanding the role of business analytics*.

Mostafaeipour, A., Fakhrzad, M., Gharaat, S., Jahangiri, M., Dhanraj, J., Band, S., & Mosavi, A. (2020). Machine learning for prediction of energy in wheat production. *Agriculture*.

Muthusinghe, M., Palliyaguru, S., Weerakkody, W., Saranga, A., & Rankothge, W. (2018). Towards smart farming: accurate prediction of paddy harvest and rice demand. *Sri Lanka Institute of Information Technology*.

Priyadarshi, R., Panigrahi, A., Routroy, S., & Garg, G. K. (2019). Demand forecasting at retail stage for selected vegetables: a performance analysis. *Journal of Modelling in Management*.

Sanyal, D., Bosch, N., & Paquette, L. (2020). Feature Selection Metrics: Similarities, Differences, and Characteristics of the Selected Models. *International Educational Data Mining Society.*

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, 527.

Torres, J. (2020). Python deep learning. *Marcombo*.

Tratar, L. F., & Strmčnik, E. (2016). *The comparison of Holt–Winters method and Multiple regression method:*. A case study Energy.

Valle, L. (2013). La Agricultura Familiar en El Ecuador. *Centro Latinoamericano para el Desarrollo Rural*.

Waugh, F. V. (1964). Demand and price analysis: some examples from agriculture. En F. V. Waugh, *Demand and price analysis: some examples from agriculture.* Illinois: Economic and statistical analysis division US Department of agriculture.

Wirth, R., & Hipp, J. (2000). CRISP-DM methodology. *CRISP-DM: Towards a standard process model for data mining*.

Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. *Neural computation.*

Yue, L., Wangwei, J., Jianguo, Z., Junjun, G., & Jiazhou, Z. (2016). Product life cycle based demand forecasting by using artificial bee colony algorithm optimized two-stage polynomial fitting. *Journal of Intelligent & Fuzzy Systems*.

Zhang, G. P. (1998). Forecasting with artificial neural networks: The state of the art. *International journal of forecasting*.