International Journal of Latest Engineering Research and Applications (IJLERA) ISSN: 2455-7137

Volume – 10, Issue – 12, December 2025, PP – 01-10

Predictive Model Based on Machine Learning Techniques for Estimating Blackberry Production in a Region of Mexico

JAL Méndez[1], JRG Pulido[2], JRH Morales[3]

[1]*The Davinci University, México*

[2,3]*The University of Colima, México*

I. Introduction

Food production, particularly of berries, has experienced a notable increase in recent years, driven by its growing market demand. Mexico has established itself as the world's leading producer of blackberries, with the states of Michoacán and Jalisco standing out for their high production. Within Michoacán, the municipalities of Los Reyes, Peribán, Tacámbaro, Tocumbo, and Ario concentrate the highest production, with Los Reyes being the national leader.

The main objective of the project is to create an advanced machine learning model to accurately predict blackberry production in selected regions of Mexico, based on the country's highest-producing regions. This will provide producers with reference information to help them decide whether to introduce the crop to a new region.

This project implemented the PRISMA framework for the systematic review and applied the CRISP-DM methodology for development. Data related to temperature, humidity, and wind were collected, along with information on production and cultivated land area. This data was used to create a dataset that served as input for the Linear Regression, Support Vector Machine, and Decision Tree algorithms. Various evaluation metrics were configured, with the squared correlation (r^2) chosen as the primary criterion to weigh and select the most suitable algorithm.

In the first application of the CRISP methodology, using data from the three highest-producing regions, satisfactory results were not achieved with any of the algorithms, as the squared correlation for all three was below 0.9. In a second iteration, a more exhaustive analysis revealed a correlation between production and cultivated land area. Therefore, the municipalities of Tocumbo and Tacámbaro, which have similar characteristics, were selected to create a new dataset. A correlation matrix was applied to analyze the relationship between the variables, and to homogenize the data, the values were normalized on a scale from 0 to 1.

This adjusted dataset was used with the models generated in the first iteration. Upon evaluating the results, a squared correlation value of 0.95 was obtained for Linear Regression, 0.90 for Decision Trees, and 0.83 for Support Vector Machine, leading to the selection of Linear Regression for implementation.

For validation, the model obtained was used with the climate and production values from the municipality of Ario to generate a production prediction. This predictive result was compared with the actual production values, thereby validating the model and fulfilling the established objective. For deployment, a prediction tool was created using the resulting model. This tool is easy for producers to use, allowing them to input the base data, after which it displays the prediction.

II. Related Work

Fashoto et al. (2021) implemented machine learning techniques to estimate maize yields for a single season in Eswatini, located in Southern Africa. A machine learning model was developed and evaluated using both publicly available and local data. This process was carried out using three distinct datasets comprising 48 observations, each with 7 features. The adjusted r^2 values were 0.784, 0.849, and 0.878; after normalization and backward elimination, the values were 0.846, 0.886, and 0.885. In a second attempt, the process was repeated using the combined predictor data of 68 data points with 7 attributes each, with the same data splits and methods. The adjusted r^2 values were 0.966, 0.972, and 0.978; after normalization and backward elimination, the values were 0.967, 0.973, and 0.978.

Lontsiet al. (2020) present a prediction system based on machine learning techniques, designed to estimate the yield of crops such as rice, maize, cassava, cotton seed, and plantains at a national level in West African countries. They integrate climatic, meteorological, agricultural yield, and chemical data to provide decision-makers and farmers with tools to anticipate annual crop yields in their respective nations. For the system's construction, methods such as Decision Trees, Multivariate Logistic Regression, and k-Nearest Neighbors models were employed. They implemented a hyperparameter tuning technique during the cross-validation process, with the goal of developing a model that minimizes the risk of overfitting. They analyzed the

correlation between the generated predictions and the actual results. The findings indicated that the predictions obtained from the Decision Tree model and the k-Nearest Neighbors model showed a significant correlation with the expected data.

Kamir (2020) presents a study with the objective of providing accurate estimates of actual wheat yields in the Australian wheat belt, using machine learning regression methods, climate data, and time series of satellite images. The benchmark approaches based on the Normalized Difference Vegetation Index (NDVI) and a harvest index were significantly outperformed by the machine learning regression models. Climatic variables, such as maximum temperatures and accumulated rainfall, provided complementary information to the 16-day NDVI time series, resulting in a notable improvement in yield predictions. Variables observed up to and around the flowering period showed high predictive power compared to information obtained during the grain-filling phase.

III. Methodology

This project is based on the PRISMA methodology for the documentary review and the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology, as it is an industry-independent model for Data Mining. Its phases in the project development are presented below.

Phase 1 Problem Understanding: A case study is conducted on the highest-producing municipalities in Mexico: Los Reyes, Peribán, Tacámbaro, Tocumbo, and Ario in the state of Michoacán de Ocampo, with the purpose of analyzing their production and the factors involved. Interviews are conducted with producers from the Los Reyes region to cross-reference the datasets with the results of these surveys. Interviews are also held with agronomists and professors from Educational Institutions in the region linked to the field to understand the factors that affect production.

Phase 2 Data Understanding: Another purpose of conducting this case study is to locate the sources of data for the historical climate data of recent years and the historical production data. This involves visiting regional institutions to obtain statistical climatological data for the region and downloading government data on blackberry production. The tasks include identifying the type, format, and meaning of the data; and assessing whether the data is complete, correct, how frequently it contains errors, and if there are missing values.

Phase 3 Data Preparation: Systematize the collected data into datasets to ensure the Machine Learning model is appropriate. This involves extracting data from the source bases, homogenizing it, and creating a base repository by applying techniques such as data normalization, handling null values, treating duplicates, and data imputation if necessary.

Phase 4 Modeling: The Machine Learning process is applied to the dataset obtained in the previous step, generating models for evaluation. The dataset corresponds to climate records (Temperature, Humidity, and Wind) and production (production and yield). Based on an analysis of these repositories, the appropriate techniques will be decided.

Phase 5 Evaluation: Exhaustive testing of the algorithms and error correction are performed. Once the best algorithm is obtained, it is implemented with the input data to obtain baseline results. The degree to which the model meets the stated objectives is measured, and if the model is deficient, the reason is detected. The data mining work and the steps followed are reviewed to determine if any important factors were omitted and to analyze aspects of model quality assurance. Based on the conclusions from the evaluation of the results and the process review, a decision is made to move to the deployment phase to put the model into operation or to conduct new iterations of the previous phases.

Phase 6 Deployment: Generate a solution where data can be input and, using the obtained model, the production of a given area can be predicted. Document the obtained results.

IV. Development

This project was carried out based on the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology, developing its six iterative phases. The Problem Understanding, Data Understanding, Modeling, and Evaluation phases of the first iteration were completed. During the evaluation, it was found that the solution was not optimal; therefore, in compliance with the CRISP-DM methodology, a second iteration of the model was carried out.

In this second iteration, no changes were made to the problem understanding phase, but the data understanding was adjusted. Based on this adjustment, we proceeded to data preparation, modeling, and evaluation. In this last phase, it was determined that an optimal solution had been achieved, so we continued with the deployment phase, thus concluding the process. Each of the phases is detailed below.

Iteration 1 Phase 1 Problem Understanding: The regions of Los Reyes, Peribán, Tacámbaro, Tocumbo, and Ario in the state of Michoacán cultivate berries, primarily blackberries. Together, these zones are currently the largest producers of this berry in the state, with Los Reyes being the national leader in production.

Iteration 1 Phase 2 Data Understanding: To begin the field work, a first survey was designed, focused on obtaining baseline data from producers, blackberry cultivation, and problems affecting it. Producers showed reluctance and little confidence to answer, which led to a redesign. A total of 60 adjusted surveys were applied. In parallel to the surveys, personal interviews were conducted with producers and professors from the Sustainable Agricultural Innovation Engineering program at the Instituto Tecnológico Superior de Los Reyes. Analyzing the survey results, it was proposed to use climatological data and add production data to implement Data Mining and generate a model that would allow us to predict blackberry production in a given area.

For the collection of meteorological information, an inquiry was made into the database of conventional climatological stations that make up the National Network of the National Water Commission (CONAGUA, 2024). It was found that this information corresponded to old periods from stations that are currently suspended. Educational institutions such as the Instituto Tecnológico Superior de Los Reyes, CBTA 49, and CONALEP Campus Los Reyes were contacted; the first two have a meteorological station, but only from 2022 and 2023 onwards. The search for this meteorological information continued, leading to the NASA POWER Project portal (NASAPOWER, 2024), which is described as "parameters related to meteorology and solar energy formulated for evaluating renewable energy systems." The climatological information for the studied municipalities, weighted by their monthly average, was obtained.

To obtain information on berry production, interviews were held with producers and packers in the regions, who were reluctant to provide this information, citing security reasons. Although it was emphasized that it was an academic study, they did not agree to provide the information. Therefore, the official portal of SAGARPA (SIAP, 2024) was used, which contains the monthly production averages for localities across the country.

The study is conducted with the variables: month, cultivated land area, production, production coefficient, temperature, humidity, and wind, obtained from the mentioned repositories. The municipalities of Los Reyes, Peribán, and Tocumbo were selected for being the highest producers to generate the dataset to be used in the models.

Climatic Data: The climatology information was downloaded in a .csv file for each of the study regions. These files were opened in a commercial spreadsheet, and data analysis and cleaning were performed, obtaining the variables Temperature, Humidity, and Wind in monthly aggregates. With these aggregates, the corresponding tables and graphs were created. The figures show examples of the monthly temperature aggregate (Table 1) and the monthly temperature graph (Figure 1).

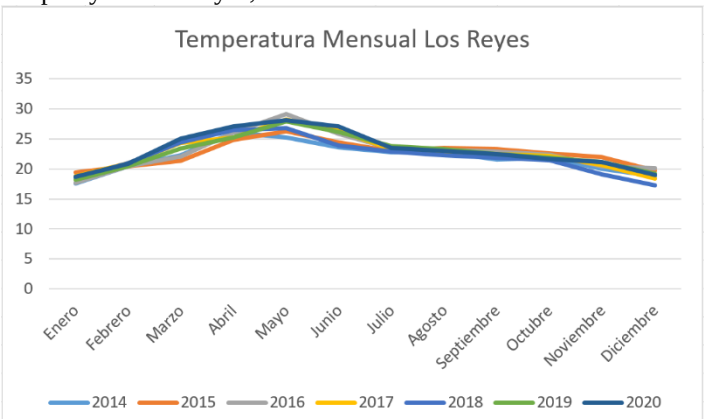
Table 1: Monthly temperature summary.

	2014	2015	2016	2017	2018	2019	2020
Enero	17.55	19.42	17.68	18.79	18.44	18.23	18.74
Febrero	20.46	20.43	20.44	20.91	20.99	20.44	20.88
Marzo	22.3	21.31	22.05	24.31	24.38	23.4	25.03
Abril	26.08	24.82	26.11	25.26	26.43	25.2	27.05
Mayo	25.27	26.3	29.13	28.18	26.83	27.9	28.12
Junio	23.63	24.45	25.87	26.39	23.94	26.19	27.09
Julio	22.83	22.95	23.4	23.3	22.86	23.81	23.5
Agosto	22.75	23.55	23.23	22.77	22.25	23.28	22.99
Septiembre	21.55	23.33	22.85	22.44	21.89	22.45	22.49
Octubre	21.85	22.62	22.33	22.14	21.44	21.87	21.65

Noviembre	20.05	21.95	20.44	20.68	19.1	21.18	21.12
Diciembre	18.76	19.74	20.12	18.42	17.3	19.23	19.02

Monthly temperature from 2014 to 2020 in the municipality of Los Reyes, own elaboration based on NASAPOWER.

Figure 1: Los Reyes Temperature. Comparison of the monthly temperature from 2014 to 2020 in the municipality of Los Reyes, own elaboration based on NASAPOWER.

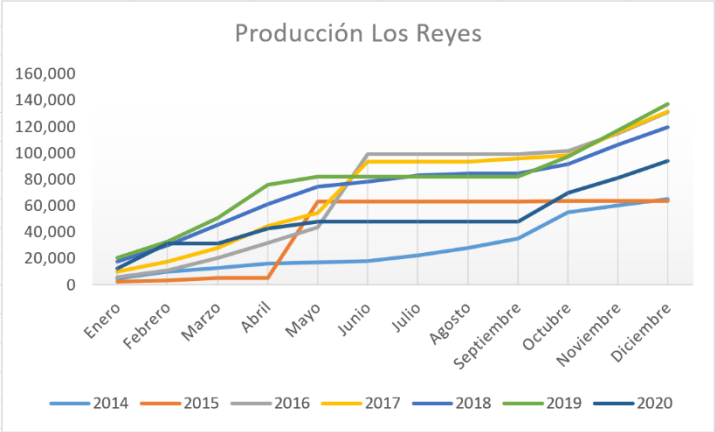


Production Data In the case of agricultural production, the portal did not have a download button; it only displayed the information on the screen for each month by selecting the state, locality, and crop characteristics. Therefore, web scraping was performed, manually downloading the information, a process that required considerable time as the query and information retrieval operation had to be done for each record. Subsequently, data analysis and cleaning were performed, obtaining the variables Harvested Area, Production, and Yield, creating the corresponding tables and graphs. An example is shown in the figures: monthly blackberry production in Los Reyes (Table 2), graph of monthly blackberry production in Los Reyes (Fig.2), and Historical production of blackberries in Los Reyes (Fig.3):

Table 2: Monthly blackberry production from 2014 to 2020 in the municipality of Los Reyes, own elaboration based on SIAP.

	2014	2015	2016	2017	2018	2019	2020
Enero	5,020	2,660	6,050	9,948	17,632	20,735	12,570
Febrero	10,040	3,302	11,151	17,623	30,057	33,040	31,570
Marzo	13,100	5,500	20,723	27,931	45,807	50,785	31,598
Abril	16,142	5,500	32,017	44,851	61,550	76,000	42,636
Mayo	17,142	63,069	43,966	54,666	74,472	82,050	48,123
Junio	18,260	63,069	99,327	93,523	78,541	82,050	48,123
Julio	22,260	63,069	99,327	93,523	83,041	82,050	48,123
Agosto	28,260	63,069	99,327	93,523	84,541	82,050	48,123
Septiembre	35,260	63,069	99,327	95,723	84,541	82,050	48,123
Octubre	55,340	63,630	101,613	98,094	91,735	97,400	69,663
Noviembre	60,240	63,630	114,827	116,017	106,485	117,250	81,450
Diciembre	65,260	63,630	130,941	131,465	119,944	137,246	94,374

Graph 2: Monthly blackberry production comparison from 2014 to 2020 in the municipality of Los Reyes, own elaboration based on SIAP.



Graph 3: Historical Blackberry Production Los Reyes from 2014 to 2020 in the municipality of Los Reyes, own elaboration based on SIAP.

Iteration 1 Phase 3 Data Preparation: A repository was created with the climatological data: temperature, humidity, and wind; and the production data: cultivated area, production, production coefficient, and production month, corresponding to the municipalities of Los Reyes, Peribán, and Tacámbaro in the state of Michoacán. Missing data were found in the municipality of Tacámbaro; these were omitted. No erroneous data were found. Information from the years 2014, 2015, 2016, 2017, 2018, 2019, and 2020 was consolidated.

Iteration 1 Phase 4 Modeling: For the analysis, the Rapid Miner Studio platform was used. This tool stands out for allowing the creation, delivery, and analysis of analytical predictions with high value. Linear Regression, Decision Trees, and Support Vector Machine were selected as they are the most suitable for the desired prediction model. The metrics used were: mean squared error, normalized absolute error, correlation, squared correlation, and prediction average, prioritizing the squared correlation for model selection.

Iteration 1 Phase 5 Evaluation: Once the models were defined, they were implemented, a process detailed below:

Linear Regression: The measures obtained were: mean squared error: 10504.64, normalized absolute error: 0.34, correlation: 0.94, squared correlation: 0.88, and prediction average: 31430.90. The analysis shows a squared correlation of 0.887, a value with a low range for model acceptance, demonstrating that the data prediction is not adequate.

Decision Trees: The measures obtained were: mean squared error: 14642.13, normalized absolute error: 0.46, correlation: 0.90, squared correlation: 0.82, and prediction average: 31430.90. The analysis shows a squared correlation of 0.825, a value with a lower range than Linear Regression.

Support Vector Machine: The measures obtained were: mean squared error: 31951.74, normalized absolute error: 1.01, correlation: 0.86, squared correlation: 0.75, and prediction average: 31429.55. The analysis shows a squared correlation of 0.755, a value with a lower range than Linear Regression and Decision Trees.

According to the CRISP methodology, since the desired result was not obtained, we returned to the initial phase of the model.

Iteration 2 Phase 1 Problem Understanding: The first phase was reviewed without modification.

Iteration 2 Phase 2 Data Understanding: Information on Climatic data and production data for the municipalities of Tocumbo and Ario was downloaded, as they were the next in production volume. For these municipalities, information from the years 2015, 2016, 2017, 2018, 2019, and 2020 was used.

Iteration 2 Phase 3 Data Preparation: A repository was created with the climatological data: temperature, humidity, and wind; and the production data: month, cultivated area, production, and production coefficient, corresponding to the municipalities of Tacámbaro and Tocumbo in the state of Michoacán. To determine how the different selected variables relate to each other, a correlation matrix with its respective graph was created. Upon analyzing the correlation matrix and its graph, it was observed that there are no correlations greater than 0.9, so the study variables were maintained as initially projected.

Iteration 2 Phase 4 Modeling: The models from Iteration 1, Phase 4, created in RapidMiner, were modified. In all three models, Linear Scaling normalization was used to change the values of the dataset columns to a common scale, reducing the variance and skewness of the data. This was done because values with large differences in scale and measurement were observed, and this difference could affect the modeling. The input dataset used was the one adjusted for the municipalities of Tacámbaro and Tocumbo.

Iteration 2 Phase 5 Evaluation: The evaluation was then carried out with the adjustments made.

Linear Regression: The measures obtained were: mean squared error: 1033.75, normalized absolute error: 0.26, correlation: 0.97, squared correlation: 0.95, and prediction average: 5740.02. The analysis shows a squared correlation of 0.955, a value with a high range for model acceptance, demonstrating that the data prediction is adequate.

Decision Trees: The measures obtained were: mean squared error: 1188.53, normalized absolute error: 0.26, correlation: 0.95, squared correlation: 0.90, and prediction average: 5740.02. The analysis shows a squared correlation of 0.90, a value that does not reach the range for model acceptance, demonstrating that the data prediction is not adequate.

Support Vector Machine: The measures obtained were: mean squared error: 3895.70, normalized absolute error: 1.03, correlation: 0.84, squared correlation: 0.832, and prediction average: 5736.36. A squared correlation of 0.83 is observed, a low value that does not reach the range for model acceptance, demonstrating that the data prediction is not adequate. A comparative table of the modeling results was created to facilitate a better analysis (Table 3).

Table 3: Model Comparison of the metrics obtained in the implemented models.

	Error cuadrático Medio:	Error absoluto normalizado	Correlación	Correlación Cuadrada	Promedio Predicción
Regresión Lineal	1033.752 +/- 312.672	0.261 +/- 0.053	0.977 +/- 0.014	0.955 +/- 0.028	5740.024 +/- 441.799
Árboles de Decisión	1188.530 +/- 386.576	0.266 +/- 0.098	0.952 +/- 0.041	0.908 +/- 0.075	5740.024 +/- 441.799
SVM	3895.708 +/- 1547.109	1.032 +/- 0.078	0.845 +/- 0.074	0.832 +/- 0.083	5736.363 +/- 1319.465

Analyzing the results, it is observed that Linear Regression has a high squared correlation, while the Decision Trees and Support Vector Machine models have a low value in this metric. The Decision Trees and

Support Vector Machine models were discarded, obtaining a suitable Linear Regression model for evaluation. The generated model was applied to the data obtained from the municipality of Ario, and the model's output was compared with the real data (Table 4).

Table 4: Model Comparison: Real vs. Estimated Production obtained from the Linear Regression model.

Mes	Cosechada	Temp	Hum	Viento	Coeficiente	Produccion	Estimada
2	432	24.4	9.52	90.92	6.05	2,613	3,411
3	432	27.93	10.25	90.93	6.05	2,613	3,380
4	432	29.73	10.93	90.83	6.05	2,613	3,558
5	432	30.43	12.27	90.78	6.05	2,613	3,889
6	432	29.1	14.47	90.81	8.106	3,501	5,298
7	432	26.69	14.59	90.9	8.106	3,501	5,457
8	432	25.39	15.08	90.84	8.106	3,501	5,761
9	432	24.45	15.08	90.82	9.17	3,960	6,350
10	504	23.91	14.65	90.81	10.31	5,196	7,332
11	504	24.05	14.77	90.91	12.837	6,470	8,367
12	504	22.51	12.21	90.95	15.649	7,887	9,163

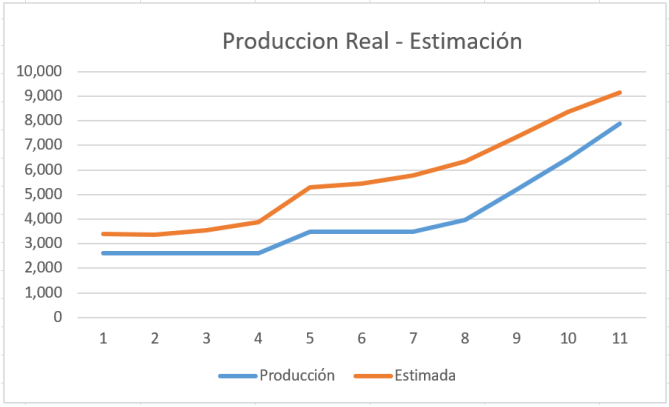
For this implementation, the real values were normalized using the Linear Scaling formula (Table 5).

Table 5: Normalized values, taking the real values and applying the Linear Scaling formula

Valores Normalizados						
Mes	cosechada	Produccion	Temp	Hum	Viento	Coeficiente
0.091	0.188	2,613	0.542	0.183	0.681	0.183
0.182	0.188	2,613	0.811	0.213	0.684	0.183
0.273	0.188	2,613	0.949	0.242	0.649	0.183
0.364	0.188	2,613	1.002	0.298	0.631	0.183
0.455	0.188	3,501	0.901	0.389	0.642	0.260
0.545	0.188	3,501	0.717	0.394	0.674	0.260
0.636	0.188	3,501	0.618	0.415	0.652	0.260
0.727	0.188	3,960	0.546	0.415	0.645	0.299
0.818	0.220	5,196	0.505	0.397	0.642	0.341
0.909	0.220	6,470	0.515	0.402	0.677	0.435
1.000	0.220	7,887	0.398	0.295	0.691	0.539

The estimated value and the real value were graphed (Fig.4).

Figure 4: Comparative Selected Model: Real Data vs. Estimation of the data obtained with the Linear Regression model and the real data



Iteration 2 Phase 6 Deployment: A normalized Linear Regression model was obtained with the formula:
Prediction = (-723.052 * Month) + (14630.250 * HarvestedArea)
+ (-853.748 * Temperature) + (-4695.418 * humidity)
+ (-2722.052 * wind) + (11228.187 * coefficient)

An electronic prediction tool was created from this model, where data for the region to be studied can be entered, and the values to be predicted are calculated. On the main sheet, the real data is captured, and the system shows the prediction and its corresponding graph.

Iteration Comparison: Two iterations were performed. In the first, during the evaluation stage, adequate results were not obtained. In the second, a result that met the objective was achieved. A comparison of the iterations in the evaluation phase is presented below (Table 6).

Table 6: Iteration Comparison at the Evaluation Phase.

	Error cuadrático Medio	Error absoluto normalizado	Correlación	Correlación cuadrada	Promedio Predicción
Regresión Lineal Iteración 1	10504.641 +/- 1452.368	0.342 +/- 0.069	0.942 +/- 0.019	0.887 +/- 0.035	31430.907 +/- 7032.908
Regresión Lineal Iteración 2	1033.752 +/- 312.672	0.261 +/- 0.053	0.977 +/- 0.014	0.955 +/- 0.028	5740.024 +/- 441.799
Árboles de Decisión Iteración 1	14642.131 +/- 2715.530	10993.236 +/- 1849.192	0.908 +/- 0.026	0.825 +/- 0.047	31430.907 +/- 7032.908
Árboles de Decisión Iteración 2	1188.530 +/- 386.576	0.266 +/- 0.098	0.952 +/- 0.041	0.908 +/- 0.075	5740.024 +/- 441.799
SVM Iteración 1	31951.742 +/- 6225.445	1.012 +/- 0.052	0.872 +/- 0.041	0.761 +/- 0.071	31391.847 +/- 5769.435
SVM Iteración 2	3895.708 +/- 1547.109	1.032 +/- 0.078	0.845 +/- 0.074	0.832 +/- 0.083	5736.363 +/- 1319.465

V. Discussion

During the development of this research, the Linear Regression, Decision Trees, and Support Vector Machine models were used, normalized with Linear Scaling on the Rapid Miner platform. The metrics implemented were: root mean squared error, normalized absolute error, correlation, squared correlation, and prediction average, with squared correlation being weighted as the base metric. The results obtained for the metrics per model are presented below (Table 7).

Table 7: Comparison of the Model Metrics obtained, own elaboration.

	Error cuadrático Medio:	Error absoluto normalizado	Correlación	Correlación Cuadrada	Promedio Predicción
Regresión Lineal	1033.752 +/- 312.672	0.261 +/- 0.053	0.977 +/- 0.014	0.955 +/- 0.028	5740.024 +/- 441.799
Árboles de Decisión	1188.530 +/- 386.576	0.266 +/- 0.098	0.952 +/- 0.041	0.908 +/- 0.075	5740.024 +/- 441.799
SVM	3895.708 +/- 1547.109	1.032 +/- 0.078	0.845 +/- 0.074	0.832 +/- 0.083	5736.363 +/- 1319.465

Performing a comparative analysis of the metrics, Linear Regression is the suitable metric for the prediction model. Decision Trees and Support Vector Machine show low values in the metrics, specifically in squared correlation.

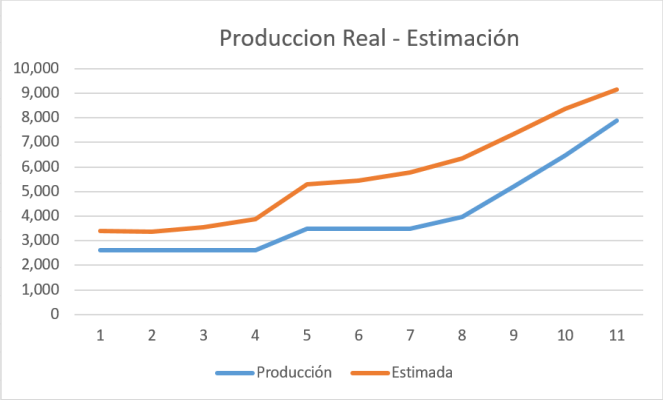
Selected Model: The Linear Regression model was selected as it met the highest values in the prediction metrics (Table 8).

Table 8: Comparison of the selected Model Metricsobtained, own elaboration.

	Raíz del error cuadrático medio	Error normalizado absoluto	correlación	Correlación al cuadrado	Promedio de predicción
Regresión Lineal	1033.752 +/- 312.672	0.261 +/- 0.053	0.977 +/- 0.014	0.955 +/- 0.028	5740.024 +/- 441.799

Model Implementation: Once the model was selected, we proceeded with its implementation by taking a municipality for which its production and input values are known. The model was applied to this municipality, comparing the real results with the modeled ones, as shown below (Graph 5).

Graph 5: Model Production vs. Real Production comparison, own elaboration.



An electronic prediction tool was created where the climatology and production values for the target area of prediction can be input. A **spreadsheet was chosen** to provide an easily accessible and user-friendly tool, avoiding the tedious process of entering data point by point or managing data loads.

VI. Future Work

The following research lines are proposed for future work:

Integration of soil factors into the predictive model: The incorporation of soil characteristic variables into the predictive model is recommended to improve its accuracy and applicability. Specifically, it is suggested to: analyze the nutritional composition of the soil (considering soil properties and nutrients such as: pH, organic matter content, macronutrient levels: nitrogen, phosphorus, and potassium, micronutrient levels: iron, zinc, boron, manganese, copper, and molybdenum) and its correlation with production levels.

Optimization of geographic segmentation in predictive modeling: To improve the model's accuracy without resorting to segmentation based solely on production areas, it is proposed to: Investigate advanced segmentation methods that incorporate multiple variables (climatic, edaphic, topographic) to define micro-production zones. Develop a hierarchical modeling approach that allows for accurate predictions at different geographical scales; and Explore machine learning techniques to identify complex spatial patterns in blackberry production.

VII. Referencias Bibliográficas

[1]. Beller EM, Glasziou PP, Altman DG, et al. (2013). PRISMA for Abstracts Group. PRISMA for Abstracts: reporting systematic reviews in journal and conference abstracts. PLoS Med 16–32

[2]. CONAGUA. (2024). Normales Climatológicas por estado. Recuperado de https://smn.conagua.gob.mx/es/informacion-climatologica-por-estado?estado=mich.

[3]. Data Mexico. (2024). Explora, visualiza, compara, y descarga datos mexicanos. Recuperado de: https://www.economia.gob.mx/datamexico/

[4]. Fashoto, Stephen & Mbunge, Elliot & Opeyemi, Ogunleye & Burg, Johan. (2021). Implementation of machine learning for predicting maize crop yields using multiple linear regression and backward elimination. Malaysian journal of computing. 6. 679-697.

[5]. Gobierno de México. (2017). Agroindustria en México, recuperado de; https://www.gob.mx/firco/articulos/agroindustria-en-mexico?idiom=e

[6]. Gobierno de México. (2024). Zarzamora, la frutilla número uno de México. Recuperado de: https://www.gob.mx/agricultura/articulos/zarzamora-la-frutilla-nume-ro-uno-de-mexico

[7]. IBM. (2021). Guía de CRISP-DM de IBM SPSS Modeler, recuperado de https://www.ibm.com/docs/es/SS3RA7_18.4.0/pdf/ModelerCRISPDM.pdf

[8]. Kamir, Elisa & Waldner, Francois & Hochman, Zvi. (2020). Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. ISPRS Journal of Photogrammetry and Remote Sensing. 160. 124-135

[9]. Lontsi Saadio, Cedric & Adoni, Hamilton & Aworka, Rubby & Zoueu, Jeremie & Kalala Mutombo, Franck &Krichen, Moez & Mberi Kimpolo, Charles. (2022). Crops Yield Prediction Based on Machine Learning Models: Case of West African Countries. Smart Agricultural Technology

[10]. NASAPOWER. (2024). NASA Prediction of Worldwide Energy Resources.

https://power.larc.nasa.gov/data-access-viewer/

[11]. Page MJ, Moher D, Bossuyt PM, et al. (2020). PRISMA explanation and elaboration: updated guidance and exemplars for reporting systematic reviews.

[12]. Ponce Pedro. (2010). Inteligencia artificial con aplicaciones a la ingeniería Primera Edición, Alfaomega Grupo Editor, S.A. de C.V., México, ISBN: 978-607-7854-83-8

[13]. SAGARPA. (2017). Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Agricultura. Planeación Agrícola Nacional 2017-2030. Frutas del bosque. Recuperado de https://www.gob.mx/cms/uploads/attachment/file/257076/potencial-frutas-del-bosque.pdf.

[14]. SIAP. (2024). Avance de Siembras y Cosechas Resumen por estado. Recuperado de http://infosiap.siap.gob.mx:8080/agricola_siap_gobmx/ResumenProducto.do

