REVISTA DIGITAL

RIISDS | Revista Interdisciplinaria De Ingeniería Sustentable Y Desarrollo Social

ISSN 2448-8003

Análisis y Evaluación Económica de la Producción Agrícola en el Estado de Sinaloa para el Año 2022 utilizando Ciencia de Datos

Economic Analysis and Evaluation of Agricultural Production in the State of Sinaloa for the Year 2022 using Data Science

Gilberto Bojorquez-Delgado[1], Jesús Bojorquez-Delgado[1], Manuel-Alfredo Flores-Rosales[1], Marco-Antonio-de-Jesús Rojas-Orozco[1], José-Antonio Sandoval-Acosta[1], Christian-Guillermo Reyes-Zúñiga[1]

[1] Tecnológico Nacional de México – ITS de Guasave, Sinaloa, México.

Recibido: 25-10-2023
Aceptado: 05-12-2023

Autor corresponsal: gilberto.bd@guasave.tecnm.mx

Resumen

La agricultura en Sinaloa desempeña un papel crucial, contribuyendo significativamente a la economía agrícola de México. En este estudio, basado en datos del SIAP para 2022, se emplearon técnicas de ciencia de datos para analizar y evaluar la producción agrícola y su impacto económico en Sinaloa. A través de un análisis exploratorio, se descubrió que cultivos como el maíz (Zea mays), sorgo (Sorghum bicolor) y frijol (Phaseolus vulgaris) dominan en términos de área sembrada. Por otro lado, en cuanto a volumen de producción, el maíz (Zea mays) y la alfalfa (Medicago sativa) resaltan como los más producidos. Es importante mencionar que, aunque el tomate (Solanum lycopersicum) no lidera en extensión o volumen, demostró ser económicamente valioso, subrayando su importancia en el mercado agrícola. Un análisis de correlación reveló una fuerte relación entre el área sembrada y el volumen de producción, indicando que las grandes áreas generalmente producen volúmenes mayores. Adicionalmente, usando la técnica de clustering, se segmentaron los cultivos en tres grupos distintos, que evidencian la diversidad de la agricultura en Sinaloa. Estos grupos abarcan desde cultivos en pequeñas parcelas hasta cultivos comerciales a gran escala. Este estudio desentraña las complejidades de la agricultura sinaloense, subraya la importancia de ciertos cultivos y brinda información clave para futuros análisis y políticas agrarias.

Palabras clave: Producción Agrícola en Sinaloa, Ciencia de Datos, Clustering.

Abstract

Agriculture in Sinaloa plays a crucial role, significantly contributing to Mexico's agricultural economy. In this study, based on 2022 data from SIAP, data science techniques were used to analyze and evaluate agricultural production and its economic impact in Sinaloa. Through exploratory analysis, it was discovered that crops such as maize (Zea mays), sorghum (Sorghum bicolor), and beans (Phaseolus vulgaris) dominate in terms of planted area. On the other hand, in terms of production volume, maize (Zea mays) and alfalfa (Medicago sativa) stand out as the most produced. It is important to mention that, although tomato (Solanum lycopersicum) does not lead in extent or volume, it proved to be economically valuable, underlining its importance in the agricultural market. A correlation analysis revealed a strong relationship between planted area and production volume, indicating that larger areas generally produce

higher volumes. Additionally, using the clustering technique, crops were segmented into three distinct groups, demonstrating the diversity of agriculture in Sinaloa. These groups range from small-scale plots to large-scale commercial crops. This study unravels the complexities of Sinaloan agriculture, highlights the importance of certain crops, and provides key insights for future analyses and agricultural policies.
Keywords: Agricultural Production in Sinaloa, Data Science, Clustering.

Introduction

Sinaloa, commonly recognized as the "breadbasket of Mexico" (Bojorquez and Fiscal, 2016), boasts a rich agricultural tradition that has played a pivotal role in food security both locally and nationally (Araujo, Ramírez, & Bautista, 2022). The blend of its diversified climate, favorable geography, and fertile alluvial soils has established it as a stronghold in agricultural production, meeting the demand of both domestic and export markets (Cota-Montes, Valenzuela-Losoya & García-López, 2023).

However, the agricultural tapestry of Sinaloa is as varied as it is complex. Each crop, with its unique characteristics in terms of market demand, production costs, yields, and climate adaptability, outlines a distinct set of challenges and opportunities for farmers and, consequently, for the state economy (Cuadras-Berrelleza, 2021).

In an era where sustainability, operational efficiency, and the maximization of returns are imperatives in the agricultural domain (Araujo & Ponce, 2023), it becomes essential to unravel the current dynamics governing agricultural production in Sinaloa. Rigorous analysis becomes crucial since decisions based on these analyses can have significant repercussions on the local economy and food security (Cuadras-Berrelleza & Portillo-Molina, 2021). This study aims to explore Sinaloa's agricultural production during 2022, employing sophisticated data science techniques like Exploratory Data Analysis (Monsalve Solis, 2021), Descriptive Analysis (Ibarra, 2022), Correlation Analysis (Hinojosa, De la Cruz & Espinoza, 2020), and Segmentation or Clustering techniques (Giménez, 2020) to unveil patterns, trends, and relevant segmentations in the data provided by the Service of Agri-food and Fisheries Information (Hernández & Alcaraz, 2021).

With the support of advanced software and analytical tools like Python (Python Software Foundation, 2023), and specific libraries such as Pandas (McKinney, 2023), Seaborn (Waskom, 2023), Matplotlib (Hunter, Año), and Scikit-learn (Shahrin, 2020), this study aspires to provide a deeper understanding of the agricultural dynamics in Sinaloa. By unraveling the relationships between key variables like the sown

area, production volume, and production value, it seeks to offer valuable insights that can guide informed agricultural decisions, sound government policies (Serra, 2021), and robust investment strategies, thus ensuring that Sinaloa perpetuates its legacy as an agricultural pillar in Mexico. Moreover, the importance of data preprocessing to manage inherent uncertainty and limitations in the analysis is acknowledged (Represa, 2020), thus ensuring the accuracy and relevance of the findings (Lara, 2022).

Materials and Methods

A detailed understanding of any phenomenon or trend requires not only careful observation but also a rigorous methodology. In the realm of agricultural research, this is even more essential, as decisions based on these analyses can have significant repercussions on the economy, food security, and the well-being of communities. This section outlines the tools, techniques, and procedures used to analyze and evaluate agricultural production in Sinaloa during 2022.

Materials:

- **Dataset:** We worked with a dataset named "Cierre_agr_mun_2022.csv", which comes from the Service of Agri-food and Fisheries Information (SIAP, 2022). This file provides a detailed view of agricultural production in Mexico for the year 2022, comprising a total of 37,555 entries distributed across 24 variables. From this dataset, specifically 846 entries corresponding to the state of Sinaloa were extracted.
- **Data Preprocessing:** Before analysis, data cleaning was carried out to handle missing and outlier values. Records with missing data were imputed using the mean method, while outliers were identified and managed using the interquartile range (IQR) method.
- **Software and Tools:** Python was used for the analysis. Specific libraries include Pandas for data manipulation, Seaborn and Matplotlib for visualization, and Scikit-learn for machine learning techniques and statistical analysis.

Methods:

- **Exploratory Data Analysis (EDA):** An EDA was performed to understand the distributions and relationships between variables. Outliers and missing values were identified and managed. Distributions of key variables were visualized using histograms and density plots.

- **Descriptive Analysis:** Measures of central tendency and dispersion were calculated for the main variables. This included the mean, median and standard deviation for variables such as area planted, production volume and production value.
- **Correlation Analysis:** A correlation matrix was generated to identify relationships between variables. Strong correlations informed subsequent steps of the analysis and helped to identify possible factors influencing production.
- **Segmentation or Clustering:** The KMeans algorithm was used to segment the crops. The optimal number of clusters was determined using the elbow method. The quality of segmentation was validated using the silhouette metric.
- **Limitations:** Despite the efforts made to ensure the accuracy of the analysis, there are intrinsic limitations associated with the data set and techniques used. Missing data imputations, although necessary, may introduce certain levels of uncertainty. In addition, the clustering algorithm, although robust, is sensitive to initialization and may vary in different runs.

Results and discussion

Through the detailed analysis of Sinaloa's agricultural data for 2022, several significant findings were obtained that provide an in-depth perspective of the agricultural situation in the state.

1. Main Crops in Sinaloa.

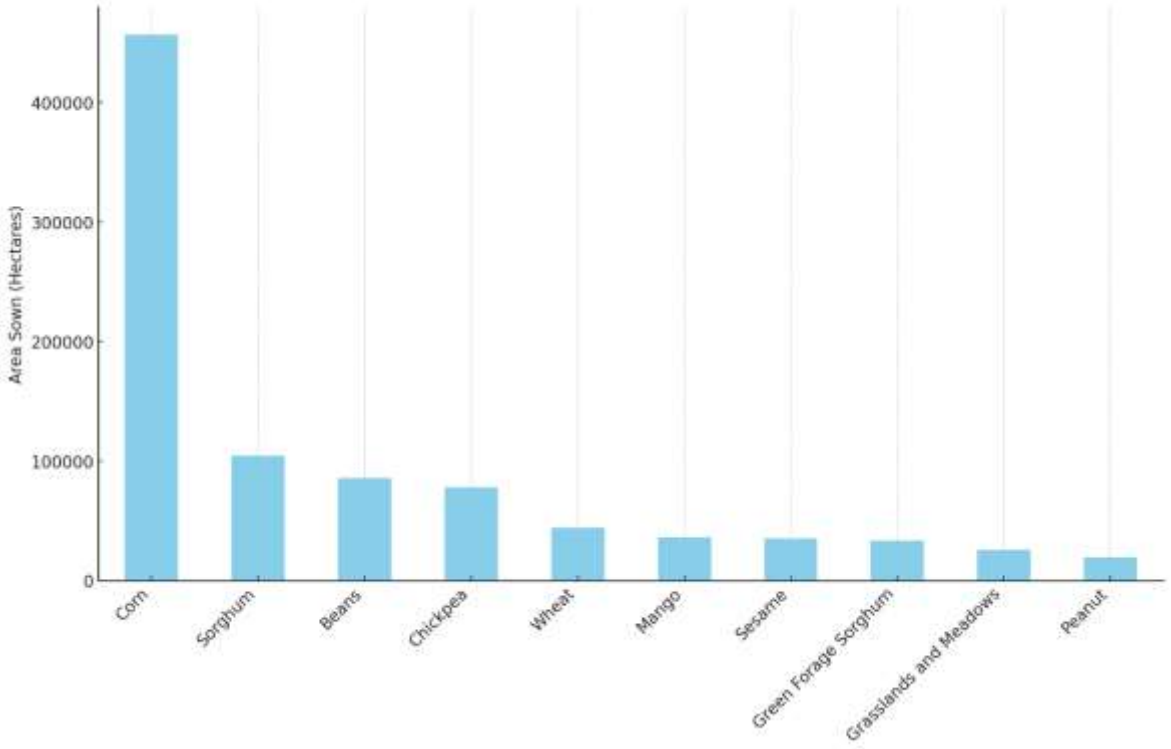

Figure 1. Main crops in Sinaloa by planted area

Figure 1 reveals that the dominant crops in terms of area planted in Sinaloa are corn, sorghum and beans. These crops have traditionally been predominant in the state, reflecting their importance for both the local economy and food security.

2. Performance Analysis

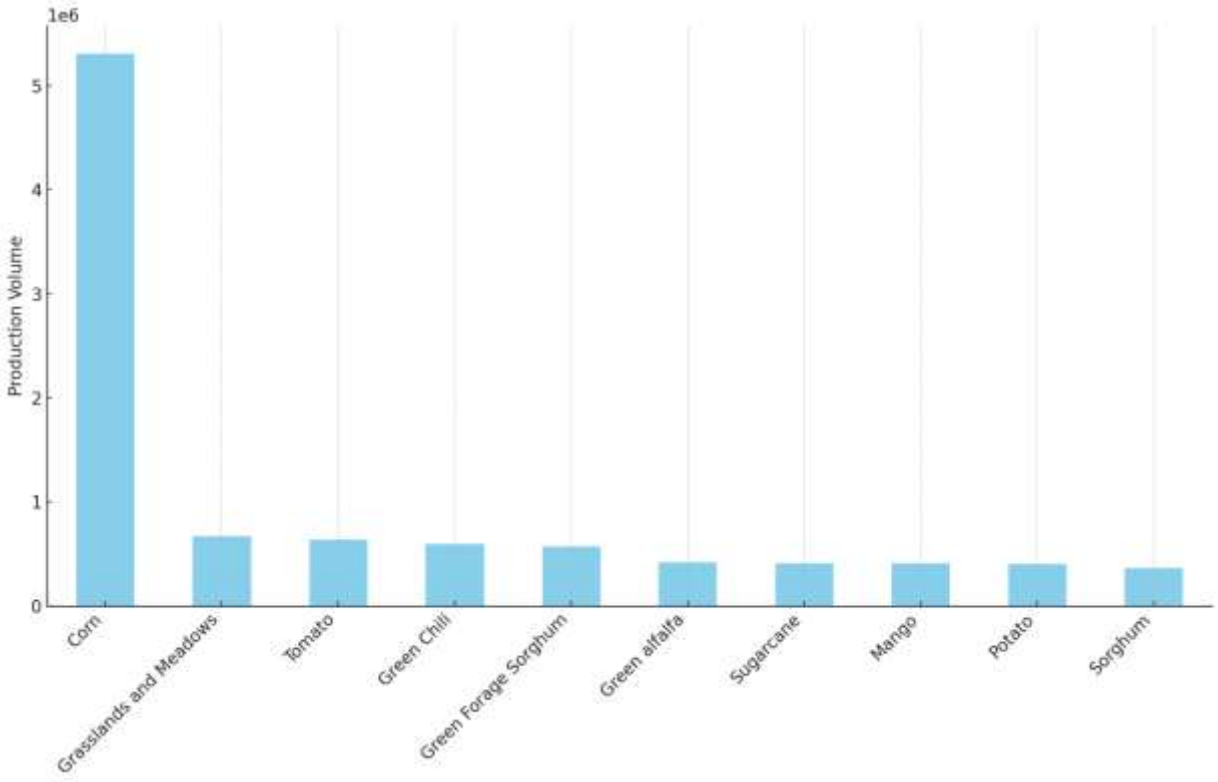

Figure 2. Main crops in Sinaloa by volume of production.

As shown in Figure 2, yields vary considerably among crops. While some crops, such as wheat, exhibit consistently high yields, others show greater variability. These yield differences can be attributed to a combination of factors, including farming practices, climatic conditions, and pest and disease resistance.

3. Economic Evaluation

The economic analysis, as shown in Figure 3, sheds light on the crops that have the highest economic value in terms of production. Among these crops, maize stands out prominently, occupying the highest position in terms of production value. This is not surprising, considering its extensive planting throughout the region and its strong demand in both local and national markets. While corn dominates the production landscape, it is critical to recognize the intricate web of factors that contribute to its economic importance. Factors such as planting area, yield per hectare, market prices and production costs interact to determine its economic impact.

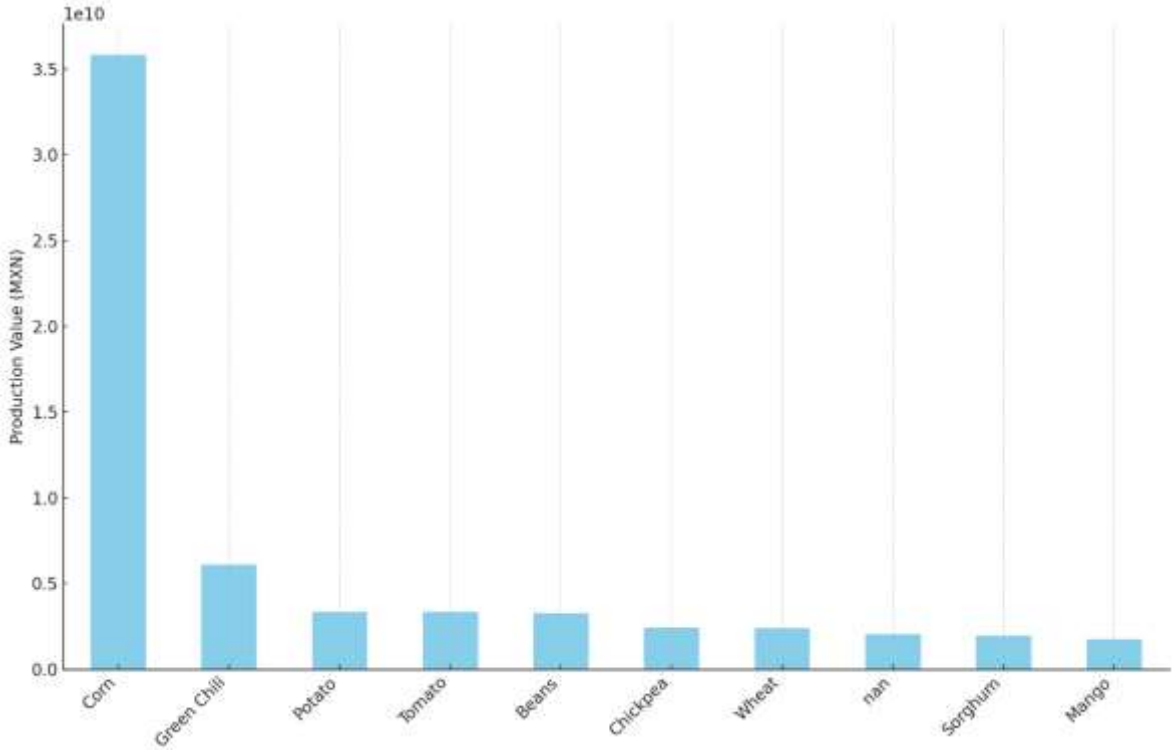

Figure 3. Main crops in Sinaloa by value of production.

4. Correlation analysis:

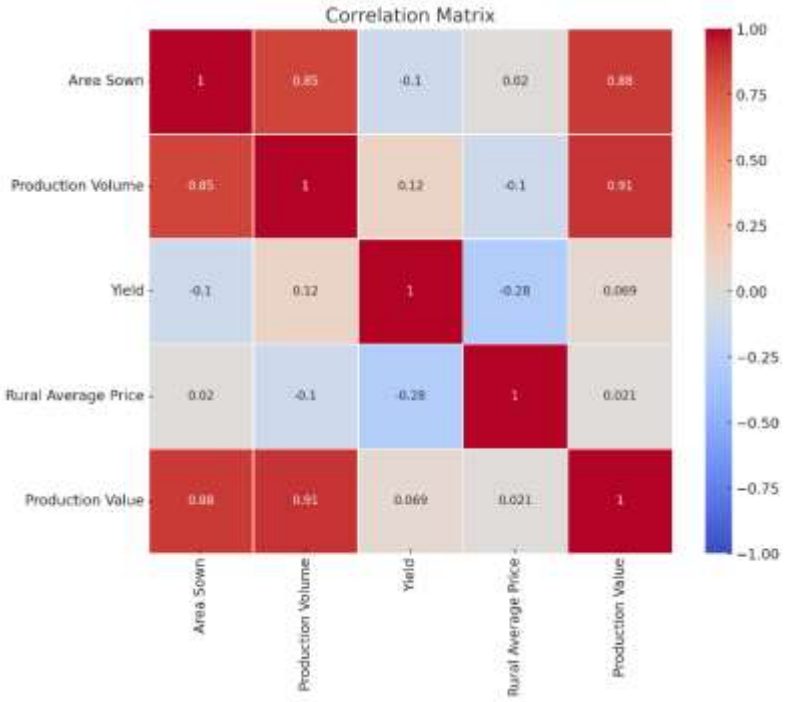

Figure 4. Correlation matrix

The correlation matrix shown in Figure 4 reveals interesting relationships. A strong positive correlation between area planted and yield volume indicates that, in general, the larger the area planted, the higher the yield volume. However, a direct correlation between area planted and yield is not always observed, suggesting that factors other than crop area size influence yield.

5. Segmentation or Clustering

The segmentation analysis, using clustering techniques, provided us with a detailed view of how crops are grouped in the state of Sinaloa during the year 2022. This approach allowed us to identify underlying patterns in the agricultural data and reveal valuable information for agricultural and economic decision making. The clusters identified revealed interesting patterns in the distribution of crops in Sinaloa. Although each cluster grouped a variety of crops, we observed notable trends in terms of area planted, yield, production value and average rural price.

- Cluster 1: This cluster includes the cultivation of maize (Zea mays) and alfalfa (Medicago sativa), characterized by high-yielding crops in relation to the area planted. Crops in this cluster stood out for their production efficiency.
- Cluster 2: Comprising tomato (Solanum lycopersicum) cultivation, Cluster 2 groups crops with high production value despite a smaller production volume, suggesting a focus on crops with high economic value.
- Cluster 3: Formed by the cultivation of sorghum (Sorghum bicolor) and beans (Phaseolus vulgaris), this cluster showed crops with a large planted area but moderate yields, indicating that these crops may require specific growing conditions.Cluster 1: Includes the cultivation of maize (Zea mays) and alfalfa (Medicago sativa), notable for their high production efficiency relative to the planted area.

Agricultural and Economic Implications

The segmentation results have important implications. Farmers can benefit by adjusting their farming practices according to the cluster to which they belong. In addition, agricultural authorities can use this information to plan and prioritize agricultural development strategies. Agriculture-related businesses can also make informed investment and marketing decisions.

Conclusions

The analysis of agricultural production and economic evaluation in Sinaloa during 2022 has revealed several crucial findings:

1. Corn dominance: Corn stands out as the main crop in Sinaloa, not only in terms of area planted and volume of production, but also in terms of economic value. Its prominence reinforces its central role in the state's agriculture.
2. Economic Importance of Tomato: Despite not being the crop with the largest planted area or production volume, tomato proved to be one of the most economically valuable crops, highlighting its significant contribution to Sinaloa's agricultural value added.
3. Area-Volume-Value Relationship: There is a strong positive relationship between area planted, production volume and production value. This indicates that, in general, crops that occupy larger areas not only produce more, but also generate greater economic value.
4. Leading Municipalities: Guasave, Ahome and Culiacán emerge as the leading municipalities in terms of agricultural production and economic value. These regions are essential to Sinaloa's agricultural dynamics and economy.
5. Crop Segmentation: Through clustering, three distinct crop segments were identified, each with unique characteristics. These segments provide a detailed perspective of the diversity and complexity of agricultural production in the state.
6. Variability in Yields and Prices: Although some crops, such as tomatoes, showed significant yields, they did not necessarily translate into the highest prices in the rural market. This suggests the presence of other factors, such as market demand and supply dynamics, which influence prices.

These findings reflect Sinaloa's rich agricultural tapestry and its essential role in Mexico's agricultural production. Analysis based on data science provides a deep understanding of these patterns, offering a solid basis for decision making, planning and agricultural policy.

Bibliographic references

Araujo, L. M. V., & Ponce, E. L. (2023). Estrategias de seguridad alimentaria: Contribuciones al desarrollo sostenible de jornaleras agrícolas de Navolato y El Fuerte, Sinaloa, México. Contextualizaciones Latinoamericanas, 2(29)

Araujo, L. M. V., Ramírez, L. T., & Bautista, J. C. C. (2022). Participación de mujeres de campos agrícolas en la seguridad alimentaria de sus familias y comunidades: estudio de una localidad en Navolato, Sinaloa, México. Margen: revista de trabajo social y ciencias sociales, (105), 9.

Bojorquez Camacho, O., y Bruno Fiscal, C. (2016). La paradoja de la economía Sinaloense en la globalización: Inseguridad alimentaria en el granero de México, 1994-2014. CONACYT.

Cota-Montes, D., Valenzuela-Losoya, F., & Adaid García-López, P. (2023). Identificación y caracterización de indicadores para evaluar la sustentabilidad ambiental en sistemas de producción agrícola en la zona del valle del municipio de Sinaloa. Ra Ximhai, 19.

Cuadras-Berrelleza, A. A., Peinado-Guevara, V. M., Peinado-Guevara, H. J., López-López, J. D. J., & Herrera-Barrientos, J. (2021). Agricultura intensiva y calidad de suelos: retos para el desarrollo sustentable en Sinaloa. Revista mexicana de ciencias agrícolas, 12(8), 1401-1414.

Cuadras-Berrelleza, A. A., Peinado-Guevara, V. M., Portillo-Molina, R., Ahumada-Cervantes, R., Belisario Leyva-Morales, J., José Peinado-Guevara, H., & Cervantes-Rosas, M. D. L. Á. (2021). Prácticas agrícolas y sustentabilidad en cultivos de maíz en Guasave, Sinaloa. Ra Ximhai, 17.

Giménez-Gallego, J., González-Teruel, J. D., Honorio Navarro-Hellín, A. B., Toledo-Moreo, F. S. V., & Torres-Sánchez, R. (2020). Segmentación de imagen de múltiples hojas de árboles con fondos naturales usando Deep Learning para aplicaciones en agricultura basadas en imagen. En II Simpósio Ibérico de Engenharia Hortícola 2020 (p. 64).

Hernández Santoyo, A., & Alcaraz Vera, J. V. (2021). Tendencia de cambio climático para la región Cuitzeo de Michoacán y su asociación con la producción de maíz de temporal.

Hinojosa Benavides, R. A., De la Cruz Marcos, R. N., & Espinoza Quispe, C. E. (2020). Revistas de ciencias agrícolas de Suramérica indexadas en SciELO: relación con indicadores de desarrollo de la agricultura. Información, Cultura y Sociedad, (43), 53-68

Hunter, J. D. (2023). Matplotlib: A 2D Graphics Environment, versión 3.8. Disponible en https://matplotlib.org/

McKinney, W. (2023). Pandas: Python Data Analysis Library, versión 2.1.4. Disponible en https://pandas.pydata.org/

Shahrin, F., Zahin, L., Rahman, R., Hossain, A. J., Kaf, A. H., & Azad, A. A. M. (2020, December). Agricultural analysis and crop yield prediction of habiganj using multispectral bands of satellite imagery with machine learning. In 2020 11th International Conference on Electrical and Computer Engineering (ICECE) (pp. 21-24). IEEE.

Ibarra, G. E. R. (2022). Agricultura de Precisión: La integración de las TIC en la producción Agrícola. Computer and Electronic Sciences: Theory and Applications, 3(1), 34-38.

Monsalve Solis, S. (2021). Agricultura de precisión en la predicción de la merma de defectos en cultivos de banano.

Python Software Foundation. (2023). Python Language Reference, versión 3.10. Disponible en https://www.python.org

Represa, N. S. (2020). Elaboración e implementación de una propuesta metodológica para la evaluación y gestión de la calidad del aire mediante el enfoque de la ciencia de datos (Doctoral dissertation, Universitat Politècnica de València).

Serra, S. (2021). Data Science for Agriculture. Revista del Grupo de Investigacion-ITDATA, 1(1), 19-20.

Servicio de Información Agroalimentaria y Pesquera (SIAP). (2022). Cierre agrícola por municipio 2022 (Versión 1.0) [Conjunto de datos]. Gobierno de México. http://infosiap.siap.gob.mx/gobmx/datosAbiertos.php

Waskom, M. (2023). Seaborn: Statistical Data Visualization, versión 0.13. Disponible en https://seaborn.pydata.org/