



UNIVERSIDAD AGRARIA DE LA HABANA
“Fructuoso Rodríguez Pérez”

FACULTAD DE CIENCIAS TÉCNICAS

TESIS EN OPCIÓN AL TÍTULO DE MÁSTER EN BIOMATEMÁTICA

**Sistema para el apoyo a la toma de decisiones
agrícolas en la Unidad Docente Productiva “El
Guayabal”**

Autora: Ing. Claudia Aguilar Rajme

MAYABEQUE, 2023



UNIVERSIDAD AGRARIA DE LA HABANA
“Fructuoso Rodríguez Pérez”

FACULTAD DE CIENCIAS TÉCNICAS

TESIS EN OPCIÓN AL TÍTULO DE MÁSTER EN BIOMATEMÁTICA

Sistema para el apoyo a la toma de decisiones agrícolas en la Unidad Docente Productiva “El Guayabal”

Autora: Ing. Claudia Aguilar Rajme

Tutoras: Dra.C Neili Machado García

MAYABEQUE, 2022

Resumen

La agricultura es una actividad fundamental para la supervivencia de la humanidad, ya que proporciona alimentos, materias primas y recursos naturales para la vida cotidiana, además de su importancia económica, la agricultura también tiene un gran impacto en el medio ambiente y en la sociedad. Es una actividad que debe ser gestionada de manera responsable y sostenible para garantizar el bienestar de las generaciones presentes y futuras. Los sistemas de soporte a la toma de decisiones son herramientas valiosas para este sector, que pueden ayudar a mejorar la toma de decisiones y aumentar la eficiencia y la productividad del sector agrícola. En esta investigación se define el procedimiento a seguir para apoyar a la toma de decisiones sobre la producción agrícola en la Unidad Docente Productiva “El Guayabal”. En este procedimiento se tienen en cuenta los indicadores medioambientales de dicha área geográfica y los resultados productivos históricos para ayudar a los expertos y productores a decidir dónde y cuándo sembrar qué cultivos para obtener los mejores resultados. Para ello se hace necesario almacenar de forma estructurada la información de los cultivos del centro y vincularla con los datos de la zona que ya se encuentran disponibles en un Sistema de Información Geográfica para luego analizarlo en conjunto y puedan llegar a conclusiones basadas en los resultados arrojados por dicho análisis.

Palabras clave: toma de decisiones, agricultura, datos espaciales, sistemas de información geográfica

Abstract

Agriculture is a fundamental activity for the survival of humanity, as it provides food, raw materials and natural resources for daily life. In addition to its economic importance, agriculture also has a great impact on the environment and society. It is an activity that must be managed in a responsible and sustainable manner to ensure the well-being of present and future generations. Decision support systems are valuable tools for this sector, which can help improve decision making and increase the efficiency and productivity of the agricultural sector. This research defines the procedure to be followed to support decision making on agricultural production in the Productive Teaching Unit "El Guayabal". This procedure takes into account the environmental indicators of this geographical area and the historical production results to help experts and producers decide where and when to plant which crops to obtain the best results. For this purpose, it is necessary to store in a structured way the information of the crops of the center and link it with the data of the area that are already available in a Geographic Information System in order to analyze it together and reach conclusions based on the results of such analysis.

Keywords: decision support, agriculture, spatial data, geographic information systems.

Índice

INTRODUCCIÓN	8
CAPÍTULO I: FUNDAMENTACIÓN TEÓRICA	13
1.1. Sistemas de Soporte a la Toma de Decisiones	13
1.2. La inteligencia artificial en la agricultura	17
1.3. Algoritmos de Inteligencia Artificial	20
1.4. Weka para el análisis de datos	25
CAPÍTULO II:	MATERIALES Y MÉTODOS
28	
2.1. Caracterización del área de estudio.	28
2.2. Datos utilizados en la investigación	28
2.3. Procedimiento a seguir para el apoyo a la toma de decisiones	31
2.3.1. Identificar los indicadores que forman parte del proceso de toma de decisiones	32
2.3.2. Diseñar la base de datos que almacenará la información	33
2.3.3. Establecer relaciones entre los datos de las diferentes fuentes	34
2.3.4. Procesar los datos	37
2.3.5. Analizar los resultados	38
2.4. Herramientas informáticas empleadas	38
CAPÍTULO III: RESULTADOS Y DISCUSIÓN	39
3.1. Resultados asociados a la estructuración de los datos	39
3.2. Resultados asociados al análisis de los datos	40
CONCLUSIONES	47
RECOMENDACIONES	48
BIBLIOGRAFÍA	49

Índice de tablas

TABLA 1 LISTADO DE LOS INDICADORES PRESENTES EN EL ANUARIO ESTADÍSTICO DEL GUAYABAL	29
--	----

Índice de ilustraciones

ILUSTRACIÓN 1: BASE DE DATOS CULTIVOS_GUAYABAL	34
ILUSTRACIÓN 2: FRAGMENTO DE CÓDIGO CON LA DCLARACIÓN DE LA CLASE ARFFFILE	35
ILUSTRACIÓN 3: IMPLEMENTACIÓN DE LA CLASE CONEXIÓN.JAVA	36
ILUSTRACIÓN 4: SECCIÓN INICIAL DEL ARCHIVO DATA-2022-8.ARFF	40
ILUSTRACIÓN 5: INTERFAZ DE VISUALIZACIÓN DE LOS DATOS RECIÉN CARGADOS EN WEKA	41
ILUSTRACIÓN 6: INTERFAZ CON LOS HISTOGRAMAS DE CADA UNO DE LOS ATRIBUTOS DEL ARCHIVO DATA-2022-8.ARFF	42
ILUSTRACIÓN 7: VISUALIZACIÓN DE TODOS LOS DATOS	43
ILUSTRACIÓN 7: ANÁLISIS DE LOS RESULTADOS DEL ALGORITMO DE CLASIFICACIÓN J48	44
ILUSTRACIÓN 8: ÁRBOL DE DECISIÓN OBTENIDO CON EL ALGORITMO J48	44
ILUSTRACIÓN 9: ANÁLISIS DE LOS RESULTADOS DEL ALGORITMO DE AGRUPAMIENTO SIMPLEKMEANS.....	45
ILUSTRACIÓN 10: VISUALIZACIÓN DE LA ASIGNACIÓN DE CLÚSTERS CON EL ALGORITMO SIMPLEKMEANS	46

Introducción

Según informe Programa Mundial de Alimentos de las Naciones Unidas (WFP) , el número de personas que padecen hambre en el mundo alcanzaba los 828 millones en 2021, mientras que el sobre peso y la obesidad continúan aumentando en todas las regiones del mundo (World Food Programme, 2023). Esto evidencia un marcado desbalance en las condiciones de vida y alimentación. Según los expertos de la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO, por sus siglas en inglés), el crecimiento y la transformación continua del sector agropecuario ofrecen oportunidades sustanciales para el desarrollo agrícola, la reducción de la pobreza, los avances en materia de seguridad alimentaria y la mejora de la nutrición humana (World Health Organization, 2020).

En Cuba debido a que existen dificultades para la adquisición de recursos por las restricciones políticas y económicas aplicadas a la isla, el gobierno cubano ha trazado el Plan Nacional de Soberanía Alimentaria y Educación Nutricional (Consejo de Ministros de Cuba, 2020), aprobado el 22 de julio de 2020 siendo el primero de su tipo en la historia del país, el cual se sustenta en la capacidad de la nación para producir alimentos de forma sostenible y dar acceso a la población a una alimentación balanceada, nutritiva e inocua, reduciendo la dependencia de medios e insumos externos, con respeto a la diversidad cultural y responsabilidad ambiental. Así mismo se aprobaron en abril de 2021 sesenta y tres medidas (Castro *et al.*, 2021) de las cuales treinta se consideran prioritarias y algunas de implementación inmediata, todas orientadas al sector agropecuario y se propone impulsar el desarrollo local entre productores y cooperativas agropecuarias para la gestión conjunta de actividades productivas que incrementen el valor agregado de las producciones primarias.

Si a esto se le suma la política de informatización de la sociedad (Consejo de Ministros de Cuba, 2017) que se aplica actualmente en el país se hace necesario dar un enfoque científico de la agricultura. La inteligencia artificial (IA) tiene el potencial de mejorar significativamente la eficiencia y la productividad de este

sector, y proporcionar una mayor seguridad alimentaria para la población mundial. Puede ayudar a los agricultores a tomar decisiones informadas sobre el uso de recursos como el agua y los nutrientes, y a identificar y controlar plagas y enfermedades. Por otra parte, los avances tecnológicos han llevado al desarrollo de drones y sensores que generan grandes cantidades de datos y la IA permite procesar todo ese volumen de información más fácilmente de forma tal que se obtengan datos útil para los agricultores (Ramón Fernández, 2020).

Existen varias formas de procesar y analizar estos datos, una de ellas es utilizar algoritmos de aprendizaje automático (ML, por sus siglas en inglés) para detectar patrones y tendencias en los datos. Por ejemplo, los algoritmos de ML pueden utilizarse para analizar imágenes de cultivos, detectar plagas o enfermedades, analizar datos meteorológicos y predecir las condiciones climáticas futuras. Estas herramientas permiten a los agricultores visualizar y analizar fácilmente los datos, lo que les ayuda a tomar decisiones informadas sobre el uso de recursos y el manejo de cultivos. Además, algunas de estas herramientas y algoritmos están disponibles en plataformas en la nube, lo que permite a los agricultores acceder a los datos y analizarlos desde cualquier lugar y en cualquier momento (Ponte *et al.*, 2021).

Al momento de tomar decisiones en el sector agrícola es importante poder analizar los diferentes factores que inciden sobre una cosecha como pueden ser tipo de suelo, los minerales presentes en él, la época del año, características climatológicas, tipo de cultivo, fertilizantes utilizados, entre otros. En muchas ocasiones estos datos provienen de diferentes fuentes y la forma que existe para poder conocer cuáles corresponden al área de interés es necesario poderlos ubicar geográficamente. La georreferencia es la ubicación concreta de un punto en un sistema de coordenadas geográficas (ASALE, 2022) por lo que al tener la información georreferenciada, se pueden visualizar en un mapa y relacionar con otros datos siempre y cuando estos también estén georreferenciados. La georreferenciación de los datos permite a los agricultores ver cómo los diferentes factores ambientales, como el clima, el suelo y el relieve, afectan a los cultivos en

diferentes áreas. También ayuda a planificar y administrar mejor los recursos, como el agua y los fertilizantes, al poder identificar las áreas que requieren una mayor atención y los recursos necesarios para su rendimiento sea mayor (Angulo Orobio, 2023). Los datos georreferenciados pueden ser utilizados como una fuente valiosa de información para mejorar los procesos de toma de decisiones en la agricultura.

Algunas de las herramientas orientadas a ayudar a las personas a tomar mejores decisiones se conocen como Sistemas de Soporte de Decisión (DSS, por sus siglas en inglés). El apoyo a una decisión significa ayudar a las personas a reunir conocimiento, generar alternativas y sugerir decisiones basados fundamentalmente en IA y ML. A nivel internacional se pueden encontrar múltiples ejemplos de estos sistemas, algunos de ellos en las investigaciones de (Jacobs, 2019), (Charania y Li, 2020), (Trichkova-Kashamova y Paunova-Hubenova, 2021), (Malvar *et al.*, 2022) y (Climate FieldView, 2023).

En el contexto de Cuba los procesos de digitalización de la sociedad se enfocan hacia la llamada transformación digital; sin embargo, en el sector de la agricultura esas transformaciones son mucho más lentas, aunque se aprecian avances. Ejemplos de estos avances se pueden encontrar en los trabajos de (Machado-García *et al.*, 2014), (Molina *et al.*, 2015), (Machado-García *et al.*, 2016), (Franco *et al.*, 2022), , y (Martínez *et al.*, 2022).

La Universidad Agraria de La Habana “Fructuoso Rodríguez Pérez” (UNAH) cuenta con la Unidad Docente Productiva (UDP) “El Guayabal”, la cual es una granja agrícola ubicada en el municipio de San José de Las Lajas que posee áreas de producción agropecuaria centradas fundamentalmente en mostrar la aplicación de la ciencia y la tecnología en el campo agrario. En estas áreas, se llevan a cabo investigaciones que contribuyen a la formación teórica y práctica de estudiantes y al mejoramiento de la producción de alimentos. En este centro se han realizado recientemente tres investigaciones enfocadas a la creación de una infraestructura que permita informatizar los procesos que tienen lugar en la UDP teniendo como premisa la georreferencia como puerta de enlace entre los datos que de ellos se deriven. Como resultado de estos trabajos se obtuvo un Sistema de Información

Geográfica (SIG) que integra datos de la zona geográfica El Guayabal desde diferentes formatos y fuentes de información (Pérez Guerra *et al.*, 2022), un método para realizar consultas mediante lenguaje natural a dicho SIG que permita su uso a personas que no estén familiarizados con el lenguaje estructurado de consultas (SQL) (Sosa Franco, 2022) y una plataforma geoespacial basada en la cartografía colaborativa para la caracterización y el seguimiento de la flora arvense, propiciando la colaboración entre los observadores y los expertos en el tema (Espinosa Pérez, 2022).

Estas investigaciones tienen especial relevancia ya que han resultado en herramientas informáticas que contienen información correspondiente al área de la granja y que anteriormente solo estaba disponible en formato físico. Entre ellas destaca por su gran relevancia para la presente investigación la base de datos del SIG con todas las capas de información georreferenciada que posee de la zona. El clima y los recursos naturales son factores que afectan la producción agrícola, su influencia en un cultivo determinado depende de las características de la localidad geográfica y de las condiciones específicas de producción por lo que tener conocimiento de estas puede hacer la diferencia a la hora de obtener resultados en la producción. Por otra parte, a pesar de que la UDP cuenta con un registro histórico de las producciones obtenidas desde 2001 hasta la actualidad estos valores no están asociados a un área geográfica (no están georreferenciados) por lo que actualmente no se puede establecer relación alguna entre esos resultados productivos y las características del entorno que pueden o no incidir sobre ellos y por tanto solo se puede llevar el registro de lo que ya se produjo sin posibilidad de valorar la incidencia de otros factores para decisiones futuras.

Lo anteriormente descrito deriva en el siguiente **problema**: ¿Cómo ayudar a los productores a tomar mejores decisiones para la producción agrícola en la UDP “El Guayabal”?

La investigación tiene como **objeto de estudio**: los sistemas espaciales de soporte a la toma de decisiones en el sector agropecuario.

Para la realización de esta investigación se propone como **hipótesis** que el empleo de un DSS para la gestión de recursos permitirá mejorar la toma de decisiones en las producciones agropecuarias en la UDP “El Guayabal”.

La presente investigación se sustenta en la **línea de investigación** de Agrofísica y Geoinformática en la Agricultura.

Para resolver el problema se traza el siguiente **objetivo general**: Definir un procedimiento para el apoyo a la toma de decisiones en la producción agrícola de la UDP “El Guayabal”.

Asimismo, se denotan los siguientes **objetivos específicos**:

- Identificar los indicadores que intervienen en los procesos de toma de decisiones en la producción agropecuaria.
- Desarrollar una arquitectura orientada a servicios para la obtención de los datos necesarios para la toma de decisiones de las diferentes fuentes que existen en la UDP.
- Aplicar los algoritmos de minería de datos para clasificar y agrupar los datos de las producciones de la UDP como parte del proceso de apoyo a la toma de decisiones basadas en el conocimiento.

Capítulo I: Fundamentación Teórica

La toma de decisiones es un factor clave en los resultados productivos de cualquier sector, más aún en el agrícola donde es tan necesario cada día obtener mejores resultados para satisfacer la demanda de alimentos. Usar la tecnología en función de apoyar esa toma de decisiones permite aumentar la eficiencia y productividad del sector. Los sistemas para el apoyo a la toma de decisiones se basan en la utilización de datos y algoritmos de inteligencia artificial para analizar patrones y tendencias en los datos. En este capítulo se abordan elementos relacionados con la aplicación de estas técnicas en la agricultura y la importancia de la toma de decisiones sobre la base de la información y el conocimiento de los diferentes factores que influyen en los resultados productivos.

1.1. Sistemas de Soporte a la Toma de Decisiones

Los sistemas de soporte a la toma de decisiones (DSS, por sus siglas en inglés) son sistemas informáticos diseñados para ayudar a los especialistas en el proceso de toma de decisiones. Estos sistemas pueden incluir modelos matemáticos, algoritmos y tecnologías de inteligencia artificial, así como también bases de datos y herramientas de visualización de datos. Los DSS ayudan a los usuarios a analizar y comprender mejor los datos relevantes para un problema particular y a generar soluciones y estrategias de acción (Brenes Carranza *et al.*, 2020). En la agricultura, los sistemas de soporte a la toma de decisiones pueden ayudar a los agricultores a tomar decisiones informadas sobre cómo optimizar su uso de los recursos y a mejorar la eficiencia y productividad de sus operaciones agrícolas.

Algunos ejemplos incluyen:

- **Sistemas de riego inteligentes:** Estos sistemas utilizan sensores y algoritmos de aprendizaje automático para analizar los datos meteorológicos y de humedad del suelo, y proporcionar información sobre el riego óptimo para los cultivos. Algunos ejemplos de estos se pueden encontrar en las investigaciones de (Pedroza Salazar y Fajardo Tascon, 2019), (Bautista Andocilla, 2022), (González-Teruel *et al.*, 2021), (Lerma Sánchez, 2020),

(Paredes *et al.*, 2021), (Hernández Rodríguez, 2019) y (Chugá Rosero, 2022).

- **Análisis de rendimiento de cultivos:** Los sistemas de análisis de rendimiento de cultivos utilizan imágenes satelitales y algoritmos de IA para analizar los datos georreferenciados y proporcionar información sobre el rendimiento de los cultivos en diferentes áreas. Las publicaciones de (Bonilla Bolaños y Singaña Tapia, 2019), (Cadena Zumárraga, 2021), (Márquez-Vasallo *et al.*, 2020), (Mederos-Ramírez y Ortiz-Pérez, 2021) y (Munévar *et al.*, 2020) abordan este tema desde diferentes perspectivas y razonamientos.
- **Predicción de plagas y enfermedades:** Los sistemas de predicción de plagas y enfermedades utilizan algoritmos de IA para analizar los datos meteorológicos y de cultivos, y proporcionar información sobre las condiciones favorables para el desarrollo de plagas y enfermedades. Los estudios de (Aimacaña Chuquimarca y Columba Guanoluisa, 2021), (Carranza Arévalo, 2020), (Guillen-Climent *et al.*, 2020), (Mendoza Zambrano, 2022), (Solano Robles y Caballero Cruz, 2022) y (Valerieva Ivanova, 2022) son muestras de ello.
- **Optimización de fertilización:** Los sistemas de optimización de fertilización utilizan algoritmos de IA para analizar los datos georreferenciados de los cultivos y proporcionar información sobre el uso óptimo de los fertilizantes. Ejemplos de estos los podemos encontrar en las publicaciones de (Brenes Carranza *et al.*, 2020), (Plazas Aquite *et al.*, 2022), (Ramón Fernández, 2020) y (Velasco Mata *et al.*, 2022).
- **Agricultura de precisión:** La agricultura de precisión utiliza tecnologías como drones, sensores y sistemas de posicionamiento global (GPS) para recopilar datos georreferenciados precisos sobre los cultivos y utilizar algoritmos de IA para analizar estos datos y proporcionar información útil para los agricultores. Algunos estudios que abordan este tema son (Berenguel y Gil, 2021), (Brenes Carranza *et al.*, 2020), (Burbano y Peñaranda, 2020), (Camila Andrea y Riveros Rey, 2023), (Otero Barrera, 2019) y (Rodríguez, 2019).

Estos sistemas permiten a los agricultores analizar grandes cantidades de datos agrícolas, como información sobre el clima, el suelo, los cultivos y los insumos, y utilizar esta información para tomar decisiones informadas sobre cómo optimizar su uso de los recursos y mejorar la eficiencia y productividad de sus operaciones agrícolas. Además, los sistemas de soporte a la toma de decisiones pueden ayudar a identificar patrones y tendencias en los datos agrícolas, lo que puede resultar en la identificación de oportunidades de mejora y la implementación de soluciones más efectivas. Esto puede ayudar a aumentar la productividad, reducir los costos y mejorar la rentabilidad de la tierra.

Para que un sistema de soporte a la toma de decisiones (DSS) sea eficiente, se requieren ciertos componentes clave. Algunos de ellos incluyen:

- **Datos relevantes y actualizados:** El DSS debe tener acceso a una amplia gama de datos agrícolas relevantes, como información sobre el clima, el suelo, los cultivos y los insumos. Además, estos datos deben ser actualizados con regularidad para garantizar que el DSS brinde información precisa y relevante.
- **Algoritmos de análisis:** El DSS debe contar con algoritmos de análisis sofisticados que permitan analizar grandes cantidades de datos agrícolas y generar información útil y valiosa para el agricultor.
- **Interfaz de usuario intuitiva:** El DSS debe tener una interfaz de usuario intuitiva que permita acceder a la información generada por el sistema de una manera fácil y eficiente.
- **Integración con otras herramientas y tecnologías:** Es importante que el DSS sea compatible con otras tecnologías y herramientas utilizadas en la agricultura, como sensores, drones y sistemas de seguimiento de cultivos, para proporcionar una imagen completa y precisa de las operaciones agrícolas.
- **Capacidad de personalización:** Es importante que el DSS sea flexible y personalizable para que pueda adaptarse a las necesidades específicas de cada agricultor.

Una parte indispensable del DSS son los datos con los que opera y para un sistema de este tipo, que opere en el área agrícola, es de gran valor que esos datos estén georreferenciados. La georreferencia es la ubicación concreta de un punto en un sistema de coordenadas geográficas (ASALE, 2022), esta permite asociar una ubicación geográfica específica a cada dato, lo que facilita la visualización y el análisis espacial de los datos agrícolas. Esto es especialmente importante en la agricultura, ya que muchos factores, como el clima, la topografía y las condiciones del suelo, varían a nivel local. Al georreferenciar los datos, los DSS pueden generar mapas de calor y otras visualizaciones que muestren la distribución de diferentes variables agrícolas en una zona determinada, lo que permite a los agricultores tomar decisiones informadas sobre dónde cultivar, cómo cultivar y qué insumos utilizar (Angulo Orobio, 2023). Por lo tanto, la georreferencia es un componente importante para lograr una eficiencia óptima en los sistemas de soporte a la toma de decisiones en agricultura.

Los sistemas espaciales de soporte a la toma de decisiones (SDSS, por sus siglas en inglés) son una subcategoría de los DSS que utilizan información georreferenciada. Estos sistemas combinan información espacial con modelos matemáticos y algoritmos para generar recomendaciones y soluciones a problemas específicos. En la agricultura, los SDSS son especialmente útiles porque permiten a los agricultores visualizar y analizar información georreferenciada relacionada con el cultivo, el suelo, el clima y otros factores importantes. Por ejemplo, un SDSS puede utilizar imágenes de satélite para identificar las áreas de un campo que requieren más atención, o para recomendar el momento óptimo para realizar ciertos trabajos agrícolas en función de la ubicación específica. Por tanto, un SDSS es un DSS que utiliza información georreferenciada para generar recomendaciones y soluciones a problemas específicos, lo que los hace especialmente útiles en aplicaciones agrícolas que requieren una comprensión precisa y detallada de la relación entre los factores ambientales y la producción agrícola (MacDonald y Faber, 2019).

Implementar un SDSS requiere una serie de pasos:

1. Identificación de problemas y objetivos: La primera etapa es identificar claramente qué problemas o desafíos se quieren abordar con el DSS.
2. Recopilación de datos: Es importante recopilar la mayor cantidad posible de datos relevantes y precisos. Esto incluye información sobre los recursos, las condiciones climáticas, la producción, los costos, etc.
3. Modelado y análisis de datos: Los datos recopilados deben ser modelados y analizados para generar información útil y relevante para la toma de decisiones.
4. Desarrollo del sistema: Una vez analizados los datos, se puede desarrollar el DSS, utilizando herramientas como modelos matemáticos, algoritmos y software especializado.
5. Pruebas y validación: Antes de su implementación, es importante realizar pruebas exhaustivas y validar que el sistema funciona adecuadamente y proporciona resultados precisos.
6. Implementación: Finalmente, se implementa el sistema y se brinda capacitación a los usuarios para que puedan utilizarlo de manera efectiva.
7. Seguimiento y mejora continua: Una vez implementado el DSS, es importante seguir monitoreando su desempeño y realizar mejoras continuas para mantener su relevancia y efectividad.

Es importante destacar que implementar un SDSS eficiente requiere un equipo interdisciplinario con conocimientos en agricultura, informática, matemáticas y estadísticas. Como ya se ha dicho estos sistemas manejan grandes cantidades de datos que son difíciles de procesar de forma manual por lo que emplean herramientas de procesamiento de datos, principalmente aquellas que pertenecen al área de la inteligencia artificial.

1.2. La inteligencia artificial en la agricultura

La inteligencia artificial (IA) es una rama de la informática que se centra en el desarrollo de algoritmos y sistemas capaces de realizar tareas que requieren inteligencia humana, como el aprendizaje, la toma de decisiones, la resolución de

problemas y el reconocimiento de patrones. Se basa en la idea de que una máquina puede ser programada para simular la inteligencia humana y realizar tareas que normalmente requieren un ser humano para hacerlas (Ricardo *et al.*, 2021). Los sistemas de inteligencia artificial se construyen a partir de una combinación de algoritmos matemáticos, estadísticos y de aprendizaje automático, y pueden funcionar de forma autónoma o en colaboración con los seres humanos. La inteligencia artificial se utiliza en una variedad de aplicaciones, incluyendo la robótica, la automatización industrial, la medicina, las finanzas y, por supuesto, la agricultura (Martínez Devia, 2019).

El uso de la inteligencia artificial en la agricultura se remonta a las décadas de 1950 y 1960, cuando los investigadores comenzaron a explorar la posibilidad de utilizarla para mejorar la eficiencia y la productividad en la agricultura. En ese momento, los primeros sistemas de IA utilizados en la agricultura se centraron en el monitoreo y el control del cultivo, y en la identificación y diagnóstico de enfermedades y plagas. A medida que la tecnología evolucionó, también lo hizo su aplicación en la agricultura. Entre 1980 y 1990, los sistemas se volvieron más sofisticados y comenzaron a utilizar técnicas avanzadas de análisis de datos, incluyendo el aprendizaje automático y la inteligencia artificial profunda. En la década de 2000, con el auge de la tecnología de sensores y dispositivos móviles, evolucionó aún más, permitiendo a los agricultores recopilar y analizar datos en tiempo real sobre su cultivo y su ganado (Valdivieso Duarte, 2021).

En la actualidad, se ha convertido en una herramienta esencial para la agricultura moderna, y los avances tecnológicos continúan impulsando nuevas aplicaciones y usos en el sector agrícola. Algunas formas en que la inteligencia artificial se utiliza en la agricultura incluyen:

- Monitoreo y control del cultivo: para recopilar y analizar datos relacionados con el crecimiento y la salud de los cultivos.
- Diagnóstico de enfermedades y plagas: para identificar y diagnosticar enfermedades y plagas en los cultivos y tomar medidas preventivas para reducir los costos relacionados a ellas.

- Optimización de la fertilización: para determinar la cantidad correcta de fertilizantes y nutrientes que se deben aplicar a los cultivos.

En concordancia con lo anterior (Astudillo Troya, 2022) plantea que el uso de la inteligencia artificial en la agricultura ofrece una serie de ventajas y desventajas que es importante tener en cuenta.

Ventajas:

- Mejora de la eficiencia: La inteligencia artificial permite a los agricultores tomar decisiones más informadas sobre cómo manejar sus cultivos y recursos, lo que resulta en una mayor eficiencia y productividad.
- Monitoreo en tiempo real: Los sistemas de inteligencia artificial pueden utilizarse para monitorear de manera constante el estado de los cultivos y la tierra, lo que permite a los agricultores reaccionar rápidamente ante cualquier problema que pueda surgir.
- Predicción climática: La inteligencia artificial puede utilizarse para analizar grandes cantidades de datos climáticos y predecir las condiciones climáticas futuras, lo que ayuda a los agricultores a tomar decisiones informadas sobre cuándo sembrar, cosechar y proteger sus cultivos.
- Mejora de la calidad de los cultivos: La inteligencia artificial puede utilizarse para optimizar las condiciones de cultivo, lo que resulta en una mejora de la calidad y producción de los cultivos.
- Reducción de costos: La automatización de tareas repetitivas y la toma de decisiones informadas mediante la inteligencia artificial pueden reducir los costos y mejorar la rentabilidad de la agricultura.

Desventajas:

- Costo de implementación: La implementación de tecnologías de inteligencia artificial en la agricultura puede ser costosa y requerir una inversión significativa en hardware y software.
- Falta de conocimiento técnico: Muchos agricultores pueden tener dificultades para entender y utilizar efectivamente la tecnología de inteligencia artificial,

lo que puede resultar en una implementación ineficiente y una disminución de los resultados esperados.

- Dependencia de la tecnología: La agricultura que depende de la inteligencia artificial puede ser más vulnerable a fallos en la tecnología y a interrupciones en la comunicación, lo que puede resultar en pérdidas significativas.
- Riesgo de pérdida de empleo: La automatización de tareas a través de la inteligencia artificial puede resultar en la pérdida de empleos para trabajadores agrícolas.
- Falta de regulación: Actualmente, existe una falta de regulación clara en cuanto al uso de la inteligencia artificial en la agricultura, lo que puede resultar en un uso inadecuado de las técnicas y herramientas si no se cuentan con los conocimientos necesarios para su desarrollo.

En resumen, la inteligencia artificial es una herramienta valiosa que permite a los agricultores tomar decisiones informadas y mejorar la eficiencia y la sostenibilidad de su producción agrícola, pero se hace necesario emplearla sobre la base del conocimiento y las condiciones tecnológicas adecuadas.

1.3. Algoritmos de Inteligencia Artificial

Un algoritmo de inteligencia artificial es un conjunto de reglas o procedimientos que permiten a una computadora o máquina imitar la inteligencia humana o resolver problemas complejos de manera automática. Los algoritmos de inteligencia artificial se basan en conceptos matemáticos y estadísticos, y su funcionamiento se apoya en la programación de computadoras (Ricardo *et al.*, 2021). Los algoritmos de inteligencia artificial se utilizan en una amplia variedad de aplicaciones, incluyendo la robótica, la visión por computadora, el reconocimiento de patrones, el aprendizaje automático y la inteligencia artificial en la nube. Algunos ejemplos de algoritmos de inteligencia artificial incluyen:

- Algoritmos de aprendizaje automático: se utilizan para analizar grandes cantidades de datos y aprender de ellos para hacer predicciones precisas o

clasificaciones (Vite Cevallos *et al.*, 2020). Ejemplos incluyen regresión lineal, árboles de decisión y redes neuronales.

- Algoritmos de reconocimiento de patrones: se utilizan para identificar patrones o tendencias en los datos y para detectar anomalías (Solano *et al.*, 2019). Ejemplos incluyen el análisis de clustering y la detección de outliers.
- Algoritmos de visión por computadora: se utilizan para analizar imágenes y videos y para realizar tareas como el seguimiento de objetos, la detección de objetos y la segmentación de imágenes (DEL ROCIO, 2022).
- Algoritmos de robótica: se utilizan para controlar robots y para permitir que los robots realicen tareas complejas en un entorno físico.
- Algoritmos de NLP (Procesamiento del Lenguaje Natural, por sus siglas en inglés): se utilizan para analizar y comprender el lenguaje humano y para realizar tareas como la traducción automática, la generación de texto y la respuesta a preguntas (Sosa Franco, 2022).

Para un sistema de soporte a la toma de decisiones agrícola, se recomienda usar algunos de los siguientes algoritmos de inteligencia artificial:

- Redes Neuronales Artificiales (RNA): son un tipo de algoritmo de aprendizaje automático que se pueden utilizar para predecir la productividad de los cultivos o la presencia de enfermedades, basadas en la estructura y funcionamiento del cerebro humano. Están compuestas por nodos interconectados, llamados neuronas artificiales, que procesan la información y realizan predicciones. Las RNA se entrenan con datos de entrenamiento para aprender a realizar tareas específicas, como la identificación de patrones o la clasificación de objetos. A través de múltiples iteraciones, la red ajusta sus pesos y ajustes para mejorar su precisión en la tarea. Existen diferentes tipos de RNA, incluyendo Perceptrón, Redes Neuronales Multicapa (MLP), Redes Convolucionales (CNN) y Redes Recurrentes (RNN). Cada tipo de red es más adecuado para tareas específicas y requiere una cantidad diferente de datos de entrenamiento y tiempo de entrenamiento (Harrington, 2012). En la agricultura, las RNA se utilizan en tareas como la

clasificación de cultivos, la detección de plagas y enfermedades, y la estimación de la producción. También se utilizan para la optimización de las prácticas agrícolas, como la elección de los mejores momentos para sembrar y cosechar, y la identificación de los mejores métodos de riego y fertilización. Son una herramienta valiosa en la agricultura debido a su capacidad para procesar grandes cantidades de datos y realizar predicciones precisas. Sin embargo, es importante tener en cuenta que el éxito de su uso depende en gran medida de la calidad y cantidad de datos de entrenamiento disponibles y de la correcta configuración y entrenamiento de la red.

- **Regresión Lineal:** es un algoritmo que se puede utilizar para predecir la producción agrícola basándose en una serie de variables climáticas y de cultivo. La regresión lineal es un algoritmo de aprendizaje automático que se utiliza para modelar la relación entre una variable dependiente (variable objetivo) y una o más variables independientes (predictoras). Este modelo se basa en la idea de que la relación entre las variables es lineal, trata de encontrar la mejor línea recta que se ajusta a los datos y se utiliza para predecir el valor de la variable dependiente a partir de los valores de las variables independientes. Es uno de los algoritmos más utilizados y bien comprendidos en el aprendizaje automático y es ampliamente utilizado en diferentes aplicaciones, incluyendo la agricultura. En la regresión lineal, se asume que la relación entre la variable dependiente y las variables independientes es lineal. Esto significa que la variable dependiente puede ser modelada como una combinación lineal de las variables independientes. La tarea en la regresión lineal es encontrar los coeficientes que minimizan el error entre los valores predichos y los valores reales de la variable dependiente. Una vez estimados los coeficientes, se pueden utilizar para hacer predicciones de la variable dependiente a partir de nuevos valores de las variables independientes (Harrington, 2012). Este algoritmo tiene varias aplicaciones en la agricultura, como la predicción del rendimiento de las cosechas en función de factores climáticos o la estimación del precio de los productos agrícolas en función de la oferta y la demanda. Además, se puede

utilizar para modelar la relación entre la calidad de los suelos y los rendimientos de las cosechas o para predecir la incidencia de enfermedades en función de factores climáticos. Sin embargo, es importante tener en cuenta que puede ser insuficiente cuando la relación entre las variables no es lineal o cuando existe una alta interacción entre las variables independientes. En estos casos, pueden ser necesarios otros modelos de regresión, como la regresión no lineal o la regresión polinómica.

- Clasificación: se pueden utilizar algoritmos de clasificación, como el Árbol de Decisión o la Regresión Logística, para clasificar los cultivos en categorías, por ejemplo, cultivos afectados por enfermedades o no afectados. Los algoritmos de clasificación son un tipo de algoritmos de aprendizaje automático que se utilizan para predecir una variable categórica a partir de un conjunto de datos de entrada. Se trata de un proceso supervisado, lo que significa que el algoritmo se entrena con un conjunto de datos etiquetados y se evalúa posteriormente en un conjunto de datos de prueba para determinar su eficacia en la clasificación. Hay varios tipos de algoritmos de clasificación, incluyendo:
 - Regresión logística: es una técnica estadística que se utiliza para predecir una probabilidad de una variable categórica binaria a partir de un conjunto de datos de entrada.
 - Árboles de decisión: es un tipo de algoritmo de clasificación que se basa en una serie de reglas de decisión para predecir la clase de un objeto a partir de un conjunto de datos de entrada.
 - Naive Bayes: es un algoritmo probabilístico que utiliza una hipótesis simplista sobre la independencia de las variables para predecir una variable categórica a partir de un conjunto de datos de entrada.
 - Máquinas de vectores de soporte (SVM, por sus siglas en inglés): es un tipo de algoritmo de clasificación que se basa en una técnica de optimización para encontrar una frontera de decisión que separe las diferentes clases de un conjunto de datos de entrada.

- **Algoritmos Genéticos:** son un tipo de algoritmo que imitan el proceso de evolución natural para optimizar la producción agrícola. Los algoritmos genéticos son una técnica de optimización inspirada en el proceso de evolución de la vida. Funcionan creando una población de soluciones posibles a un problema y utilizando técnicas basadas en la selección natural y la mutación para mejorar la población a lo largo del tiempo. La selección natural implica evaluar cada solución en la población y seleccionar aquellas con mejores resultados para ser utilizadas como base para la siguiente generación. La mutación consiste en hacer pequeñas alteraciones aleatorias a las soluciones seleccionadas para añadir variabilidad a la población. Este proceso se repite varias veces hasta que se encuentre una solución óptima o hasta que se alcance un criterio de parada (Harrington, 2012). Los algoritmos genéticos se utilizan en una variedad de aplicaciones, incluida la agricultura. Por ejemplo, se pueden utilizar para optimizar los planes de cultivo y para encontrar las mejores combinaciones de fertilizantes, pesticidas y otros insumos. También pueden utilizarse para mejorar la disposición de los sensores de monitoreo en una granja para maximizar la recolección de datos. Sin embargo, también hay desventajas en el uso de algoritmos genéticos, incluyendo tiempos de procesamiento largos y la necesidad de una cantidad significativa de datos para entrenar el algoritmo.
- **Algoritmos de agrupamiento:** son un tipo de algoritmo de aprendizaje no supervisado que se utiliza para agrupar datos similares en clústers. Estos algoritmos funcionan analizando las características de los datos y asignando cada registro a un cluster en función de su similitud con los demás. Hay varios tipos de algoritmos de clustering, incluyendo k-means, agrupamiento jerárquico, agrupamiento basado en densidad y modelos basados en gaussianos. Cada uno funciona de manera diferente y puede ser más adecuado para diferentes tipos de datos y aplicaciones (Harrington, 2012). En la agricultura, los algoritmos de clustering pueden utilizarse para agrupar parcelas de tierra según sus características, como la topografía, la calidad del suelo y el uso anterior. Estos grupos pueden utilizarse para guiar la toma

de decisiones sobre cómo mejorar la producción agrícola y aumentar la eficiencia en el uso de recursos como fertilizantes y agua. Además, estos algoritmos pueden utilizarse para identificar patrones en los datos meteorológicos y de crecimiento de las plantas, lo que puede ayudar a predecir las condiciones ideales para el cultivo de diferentes cultivos y a optimizar la planificación de la cosecha.

Estos son solo algunos ejemplos de algoritmos que se pueden utilizar en un sistema de soporte a la toma de decisiones agrícola. La elección final depende de los datos disponibles y de las necesidades específicas del sistema. Actualmente existen diferentes herramientas que ya traen estos algoritmos incorporados para permitir su uso a usuarios que no sean expertos en el área de la inteligencia artificial y que solo necesiten los resultados para analizarlos, una de ellas es el software Weka.

1.4. Weka para el análisis de datos

Weka es una colección de algoritmos de inteligencia artificial para tareas de análisis de datos. Contiene herramientas para la preparación, clasificación, regresión, agrupación, minería de reglas de asociación y visualización de datos. Es un software de código abierto emitido bajo la Licencia Pública General GNU (Universidad de Waikato, 2023). Todas las técnicas de Weka se fundamentan en la asunción de que los datos están disponibles en un fichero plano o una relación, en la que cada registro de datos está descrito por un número fijo de atributos (normalmente numéricos o nominales, aunque también se soportan otros tipos). El formato soportado por Weka es el formato de archivo de atributos-relación (ARFF, por sus siglas en inglés).

Los ARFF son archivos de texto ASCII¹ que describen una lista de instancias con atributos comunes. Se componen de dos secciones distintas. La información de cabecera está presente en la primera sección, que contiene detalles sobre el nombre de las relaciones, las listas de atributos y sus tipos. El nombre de la relación se indica en la primera línea del archivo ARFF. Las declaraciones de atributos

¹ ASCII: código de caracteres basado en el alfabeto latino

siguen al nombre de la relación mediante una secuencia ordenada de declaraciones @atributo. Cada atributo perteneciente al conjunto de datos se denota mediante una declaración @atributo independiente que lleva el nombre del atributo y el tipo de datos. La segunda sección del archivo ARFF es la sección Datos. Esta sección consta de la línea de declaración de datos y de las líneas de instancia propiamente dichas. La línea de declaración de datos es una sola línea que denota el comienzo de la sección de datos del archivo. Cada instancia que sigue a esta línea se denota utilizando una sola línea, con salto de línea para indicar que la instancia ha terminado.

Entre los algoritmos más utilizados en Weka dependiendo del tipo de datos y propósito de la evaluación se encuentran el árbol de decisión J48 y el de agrupamientos SimpleKMeans (Mina, 2018). Los árboles de decisión proveen de una herramienta de clasificación muy potente. Su uso en el manejo de datos la hace ganar en popularidad dadas las posibilidades que brinda y la facilidad con que son comprendidos sus resultados por cualquier usuario. El árbol en sí mismo, al ser obtenido, determina una regla de decisión (Bouza y Santiago, 2012). Un árbol de decisión como J48 aplicado a un conjunto de datos con una lista de predictores o variables independientes y una lista de objetivos o variables dependientes, permitirá predecir la variable de destino de un nuevo registro de conjunto de datos (Mina, 2018).

Los algoritmos de agrupación son una herramienta que permite la experimentación de análisis de datos mediante la aplicación, análisis y evaluación de las técnicas provenientes del aprendizaje automático, sobre cualquier conjunto de datos del usuario. Estos algoritmos buscan grupos de instancias con características similares, según un criterio de comparación entre valores de atributos de las instancias definidos en los algoritmos (Chakkaravarthy *et al.*, 2020). SimpleKMeans es un algoritmo de clasificación no supervisada (clusterización) que agrupa objetos en k grupos basándose en sus características. El agrupamiento se realiza minimizando la suma de distancias entre cada objeto y el centroide de su grupo o cluster, se suele usar la distancia cuadrática. El algoritmo consta de tres pasos:

1. Inicialización: una vez escogido el número de grupos, k , se establecen k centroides en el espacio de los datos, por ejemplo, escogiéndolos aleatoriamente.
2. Asignación objetos a los centroides: cada objeto de los datos es asignado a su centroide más cercano.
3. Actualización centroides: se actualiza la posición del centroide de cada grupo tomando como nuevo centroide la posición del promedio de los objetos pertenecientes a dicho grupo.

Se repiten los pasos 2 y 3 hasta que los centroides no se mueven, o se mueven por debajo de una distancia umbral en cada paso. Este algoritmo resuelve un problema de optimización, siendo la función a optimizar (minimizar) la suma de las distancias cuadráticas de cada objeto al centroide de su clúster (Patil *et al.*, 2015).

Por todas las opciones que brinda, el amplio conjunto de algoritmos que contiene y su condición de software libre Weka es una de las herramientas más utilizadas para el procesamiento y análisis de datos. Es un software de fácil acceso para los usuarios con una interfaz de usuario sencilla por lo que goza de gran aceptación por parte de quienes lo emplean.

Capítulo II: Materiales y métodos

En este capítulo se definen los materiales empleados para el desarrollo de la investigación tales como: la información del área de estudio, el procedimiento a seguir para el apoyo a la toma de decisiones y las herramientas de software que permiten llevarlo a cabo.

2.1. Caracterización del área de estudio.

La presente investigación se realizó en la Unidad Docente Productiva (UDP) “El Guayabal”, la cual se encuentra ubicada en las coordenadas 22°59'57.49"N 82°10'09.42"W a las afueras del poblado de San José de las Lajas. Pertenece a la provincia Mayabeque y más específicamente a la Universidad Agraria de La Habana “Fructuoso Rodríguez Pérez” (UNAH) de la cual se encuentra a solo 5,3 km. Esta unidad docente combina el trabajo de investigación de estudiantes y profesores del Complejo Científico Docente² con la producción agropecuaria y permite la validación de dichas investigaciones en un escenario controlado. Las producciones resultantes de esta entidad se destinan fundamentalmente a contribuir la alimentación de los estudiantes, profesores y trabajadores del complejo (Pérez Guerra *et al.*, 2022).

Posee una extensión de 665,7 ha, equivalente a 49,6 caballerías. Su área agrícola es de 72,2 ha y sus suelos son predominantemente Ferralíticos Rojos. En este centro se desarrollan procesos de ingeniería, agrícolas, de salud y reproducción animal. Así mismo se trabaja en la ceba de toros, recría, apicultura, acuicultura y en la comercialización de estos. Su área productiva cuenta con cría de ganado mayor y menor como: vacas, cerdos, aves, conejos y peces. Dicho centro se caracteriza por la labor investigativa y científica, así como por la producción agropecuaria que se pretende sea cada vez mayor (Sosa Franco, 2022).

2.2. Datos utilizados en la investigación

En la Facultad de Ciencias Económicas y Empresariales (FCEE) de la UNAH, como parte del proyecto “Gestión sustentable de la producción agraria en la Granja

² El Complejo Científico Docente está formado por la UNAH, el Centro Nacional de Salud Animal (CENSA), el Instituto Nacional de Ciencias Agrícolas (INCA) y el Instituto de Ciencia Animal (ICA).

Universitaria Guayabal” se confecciona el Anuario Estadístico del Guayabal con la información brindada por los directivos del centro. Este anuario en su primera versión comprende datos entre 2002 y 2011 de los 50 indicadores (Tabla 1) asociados a la producción en este período conteniendo así un total de 2380 registros.

Tabla 1 Listado de los indicadores presentes en el Anuario Estadístico del Guayabal

No.	Indicador	Unidad de Medida
1	Producción de huevo	unidad
2	Cantidad de gallinas ponedoras al inicio del mes	cabezas
3	Cantidad de gallinas ponedoras al final del mes	cabezas
4	Cantidad de gallinas ponedoras promedio en el mes	cabezas
5	Rendimiento de gallinas ponedoras	unidad/cabezas
6	Carne de Gallina	kg
7	Gallinas sacrificadas	unidad
8	Rendimiento de gallinas sacrificadas	cabezas
9	Peso promedio total gallinas sacrificadas	kg
10	Cantidad de pollos de ceba	cabezas
11	Pollos de ceba sacrificados	cabezas
12	Rendimiento pollos de ceba sacrificados	cabezas
13	Cantidad carne de conejo	kg
14	Cantidad conejos sacrificados	cabezas
15	Rendimiento de conejos sacrificados	cabezas
16	Peso promedio de conejos sacrificados	kg
17	Cantidad carne de cerdo	kg
18	Cantidad de cerdos sacrificados	cabezas
19	Rendimiento de cerdos sacrificados	cabezas
20	Peso total de cerdos sacrificados	cabezas
21	Cantidad de carne vacuna	kg
22	Cantidad de sacrificios vacunos	cabezas
23	Rendimiento de sacrificios vacunos	kg/cabezas

24	Peso total de vacunos sacrificados	kg
25	Cantidad de carne de búfalos	kg
26	Cantidad de sacrificios de búfalos	cabezas
27	Rendimiento de sacrificios de búfalos	kg/cabezas
28	Peso total de sacrificio de búfalos	kg
29	Plan de producción de leche	litros
30	Producción real de leche	litros
31	Rendimiento de vacas en ordeño	litros/cabeza
32	Nacimientos	cabezas
33	Existencias de vacas	cabezas
34	Muertes de vacas	cabezas
35	Cantidad de carne ovinos-caprinos	kg
36	Cantidad de ovinos-caprinos sacrificados	cabezas
37	Rendimiento de ovinos-caprinos sacrificados	kg/cabezas
38	Peso total de ovinos-caprinos sacrificados	kg
39	Cantidad de carne equina	kg
40	Cantidad de equinos sacrificados	cabezas
41	Rendimiento de equinos sacrificados	kg/cabezas
42	Peso total de equinos sacrificados	kg
43	Cantidad de viandas en huertos y organopónicos	kg
44	Cantidad de hortalizas en huertos y organopónicos	kg
45	Cantidad de granos en huertos y organopónicos	kg
46	Cantidad de frutales en huertos y organopónicos	kg
47	Cantidad de viandas en áreas de hasta 100ha	kg
48	Cantidad de hortalizas en áreas de hasta 100ha	kg
49	Cantidad de granos en áreas de hasta 100ha	kg
50	Cantidad de frutales en áreas de hasta 100ha	kg

Estos datos se encuentran distribuidos en tres libros Excel y para su presentación como anuario lo estructuraron en un documento Word con imágenes de las tablas de los libros.

Por otra parte, como ya se dijo, existe un Sistema de Información Geográfica que cuenta con una base de datos de información referente al área que ocupa la UDP. En ella se pueden encontrar datos sobre el espacio que abarcan las máquinas de riego, los límites geográficos de las áreas agrícolas que actualmente son 13 (un cantero, dos plantas proteicas, tres huertos y siete grandes áreas de cultivos), de cada una de estas áreas se conoce su extensión en hectáreas, concentración de fósforo, magnesio, calcio, potasio, pH y tipo de suelo. Otro dato que se almacena es para una fecha dada el cultivo que se encuentra sembrado en cada una de ellas. Sobre el clima se tiene registro diario de las temperaturas máxima y mínima, porcentaje de humedad, velocidad de los vientos y el registro de lluvia. Como ya se dijo anteriormente el clima y los recursos naturales son factores que afectan la producción agrícola por lo tanto la información con la que actualmente se cuenta en dicho SIG es importante y debe ser tomada en cuenta a la hora de decidir cuándo y dónde comenzar un nuevo ciclo de cultivos.

2.3. Procedimiento a seguir para el apoyo a la toma de decisiones

Para la presente investigación fue necesario analizar los diferentes aspectos que componen el proceso de apoyo a la toma de decisiones, para lo cual se tomaron como referentes los trabajos de (Gálvez y GÁLVEZ, 1994), (Pérez, 1998) y (Lamelas, 2009), y finalmente se siguió el procedimiento que se muestra a continuación:

1. Identificar los indicadores que forman parte del proceso de toma de decisiones.
2. Diseñar la base de datos que almacenará la información.
3. Establecer relaciones entre los datos de las diferentes fuentes.
4. Procesar los datos
5. Analizar los resultados

2.3.1. Identificar los indicadores que forman parte del proceso de toma de decisiones

Teniendo en cuenta los datos que ya se recogían para el anuario y tras realizar varios análisis de conjunto con los directivos y especialistas de la granja fue necesario añadir otros valores que se manejan en el centro pero que no se llevaban al reporte final de producción, como son el desglose por tipo de cultivo para que además de la producción total de viandas, granos, frutas y hortalizas sea posible determinar, por ejemplo, cuánto corresponde a boniato y cuánto a malanga, siendo estos cultivos los que se relacionan directamente con el área donde están sembrados en el SIG y no el tipo de cultivo en sí. También es necesario almacenar qué cultivo hay sembrado en las áreas y la fecha de siembra y de cosecha para poder enmarcar el período de tiempo en el cual estuvo sembrado y poder asociarlo a los registros del clima que se obtienen de forma diaria para la zona que comprende la granja. Se debe registrar, además, el plan de producción de dicha siembra para luego cuando se coseche y se calcule el total de producción real determinar si este se cumplió o no, y tomarlo como variable de referencia para la toma de decisiones. De esta forma quedaron identificados un total de 21 indicadores que intervienen en este proceso de toma de decisiones, los cuales se relacionan a continuación:

1. Área geográfica
2. Nivel de fósforo del suelo
3. Nivel de magnesio del suelo
4. Nivel de materia orgánica del suelo
5. Nivel de calcio del suelo
6. Tamaño del área en hectáreas
7. Nivel de ph del suelo
8. Nivel de potasio (k₂o) del suelo
9. Nivel de ácido fosfórico (p₂o₅) del suelo
10. Tipo de suelo
11. Año
12. Mes
13. Día

14. Temperatura máxima
15. Temperatura mínima
16. Por ciento de humedad
17. Velocidad del viento
18. Cantidad de lluvias
19. Tipo de cultivo
20. Tipo de área
21. Cumplimiento del plan de producción

Una vez identificados es necesario poseer una estructura que permita tener toda esta información disponible para poder procesarla posteriormente.

2.3.2. Diseñar la base de datos que almacenará la información

A partir de los datos que se recopilan para el anuario se diseñó una base de datos (cultivos_guayabal) que permita gestionarlos y añadir los que identificaron que faltaban para un correcto análisis de la información. Para lograr esto la base de datos resultante (ver **¡Error! No se encuentra el origen de la referencia.**) cuenta con 4 entidades: producción, área, cultivo y área_cultivo. En producción se almacenan los tipos de producción como son viandas, frutas, hortalizas y granos. Los registros de área y área_cultivo debe corresponderse con las áreas geográficas en las cuales se encuentra dividida la UDP que están geográficamente delimitadas en el SIG y los cultivos que hay sembrados en cada una de ellas según la fecha. Los cultivos responderán a un tipo de producción y podrán estar sembrados en diferentes áreas, así como en un área se podrán tener varios cultivos en diferente fecha. Para un cultivo sembrado en un área en una fecha específica se debe almacenar la fecha de recogida, el plan de producción y una vez cosechado se debe registrar la producción real, la cual se divide en producción de cultivos permanentes o temporales según los registros que actualmente se tienen, y la suma de estos será la producción real que habrá que valorar con respecto al plan de producción para determinar si se cumplió o no.

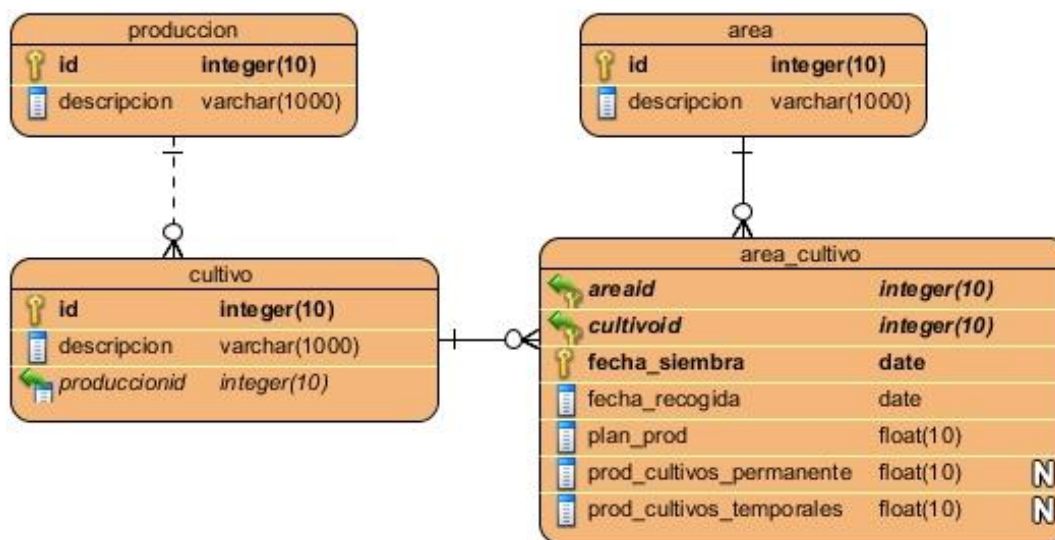


Ilustración 1: Base de Datos cultivos_guayabal

Una vez se tuvo la información almacenada en esta base de datos entonces se podrá unir a la que ya posee el SIG pasar al análisis, para que a partir de lo que ya se conoce los especialistas puedan tomar decisiones.

2.3.3. Establecer relaciones entre los datos de las diferentes fuentes

Cuando se realiza análisis de datos estos deben estar presentados según las necesidades de la herramienta o algoritmo que se utilizará para procesarlos. Para la presente investigación se emplea el software Weka en su versión de escritorio 3.8.6, al cual se le deben introducir los datos mediante un archivo ARFF. Como se tienen los datos almacenados en dos bases de datos diferentes (cultivos_guayabal y geo_guayabal que es la base de datos del SIG) fue necesario implementar una arquitectura orientada a servicios que se comunicara con ambas, extrajera la información necesaria y creara el archivo ARFF. Esta arquitectura se desarrolló con el framework Spring Boot y está compuesta por los siguientes elementos:

DssApplication.java: Clase principal por donde se inicia la ejecución del programa y que verifica la disponibilidad del servicio de postgresql para establecer contacto con las bases de datos.

ArffController.java: Controlador donde se define el endpoint `"/createFileByDate/{anno}/{mes}"` el cual se usa filtrar los datos correspondientes a un mes y un año específicos.

ArffFile.java: Clase donde se define la estructura del fichero ARFF que contiene el nombre de la relación, la lista con los nombres de los atributos en el orden que serán declarados y luego la lista de registros donde los valores deben corresponderse con el orden en que fueron declarados los atributos (ver Ilustración 2). Es importante destacar, ya que el archivo se crea con el objetivo de utilizarlo como fuente de datos en Weka, que el último atributo que se declare es el que se asume como variable dependiente a no ser que se indique lo contrario durante la ejecución de los algoritmos en el software.

```
1 package cu.unah.guayabal.dss.data;
2
3 import java.util.ArrayList;
4 import java.util.List;
5
6 public class ArffFile {
7     String relation;
8     List<String> attributes;
9     List<List<String>> data;
10 }
```

Ilustración 2: Fragmento de código con la declaración de la clase ArffFile

Conexión.java: clase donde se configura el acceso al sistema gestor de bases de datos, en este caso se parametrizaron los valores correspondientes al servidor de bases de datos, puerto, usuario y contraseña para acceder a él, así como el nombre de la base de datos (ver Ilustración 3).

```

10 import java.sql.*;
11 import java.util.logging.Level;
12 import java.util.logging.Logger;
13
14 public class Conexion {
15
16     private static Connection con = null;
17
18     public static Connection Conexion(String host, String puerto, String database, String username, String password) {
19         try {
20             Class.forName("org.postgresql.Driver");
21             con = DriverManager.getConnection("jdbc:postgresql://" + host + ":" + puerto + "/" + database, username, password);
22         } catch (ClassNotFoundException ex) {
23             Logger.getLogger(Conexion.class.getName()).log(Level.SEVERE, null, ex);
24         } catch (SQLException ex) {
25             Logger.getLogger(Conexion.class.getName()).log(Level.SEVERE, null, ex);
26         }
27         return con;
28     }
29
30     public void CerrarConexion() {
31         try {
32             if (con != null)
33                 con.close();
34         } catch (SQLException ex) {
35             Logger.getLogger(Conexion.class.getName()).log(Level.SEVERE, null, ex);
36         }
37     }
38 }

```

Ilustración 3: Implementación de la Clase Conexión.java

DBService.java: Es la clase mediante la cual se crean instancias de las conexiones a las bases de datos, se cierran estas conexiones y, además, se implementó un método que devuelve una lista con los nombres de las tablas de la base de datos. Este método es necesario para poder filtrar las tablas del SIG de las cuales se extraerá información para crear el archivo ARFF ya que cada vez que se añaden capas de información al SIG se hace en forma de una nueva tabla y si se trabaja solo con las que se tengan predefinidas se podría perder información relevante y actualizada.

ArfffService.java: es donde se implementa la lógica para la creación del archivo, se crea la conexión a cada una de las bases de datos, se llaman a los métodos que ejecutan las consultas para extraer la información y se estructura dicha información teniendo como punto de enlace el componente geográfico, es decir, un registro se conforma teniendo en cuenta los resultados de ambas bases de datos. De la base de datos del SIG se extrae para un área geográfica el nivel de fósforo, magnesio, materia orgánica, calcio, ph, potasio y ácido fosfórico del suelo, tamaño en hectáreas, tipo de área, tipo de suelo, y por cada uno de los días la temperatura máxima, temperatura mínima, porcentaje de humedad, velocidad del viento, cantidad

de lluvia y tipo de cultivo, todo esto correspondiente al mes y año que se pasa en la llamada del endpoint de la clase `ArffController.java`. De la otra base de datos se extrae para cada área el cumplimiento o no del plan de producción para el mismo mes y año que en la otra consulta. Todo esto se persiste en un fichero que se crea en la carpeta raíz del proyecto cuyo nombre es *data-<año>-<mes>.arff*.

2.3.4. Procesar los datos

Para analizar toda esta información se utilizan técnicas de minería de datos, haciendo uso del software Weka se pueden aplicar varias herramientas para extraer conocimiento de la información que se tiene, en este caso se hace énfasis en tres de ellas: visualización, clasificación y agrupación.

Visualización

Una vez se tiene el archivo con los datos se le pasa al software Weka donde se puede visualizar la información que contiene como el nombre de la relación, la cantidad de atributos e instancias. Si se selecciona uno de los atributos se ven en la parte derecha los valores que toma y un histograma que los representa.

Clasificación

El próximo paso es la clasificación en la cual se debe escoger el algoritmo de clasificación que se aplicará, en este caso dentro del grupo de árboles se utiliza el *J48*, indicando que la información será utilizada como conjunto de entrenamiento y se selecciona la variable dependiente. Los resultados se ven en la parte derecha de la interfaz y para una mejor comprensión se puede visualizar el árbol que se obtiene en la sección izquierda inferior Lista de Resultados.

Agrupamiento

Para la agrupación se selecciona, al igual que en la clasificación, que los datos serán usados como conjunto de entrenamiento y se selecciona el algoritmo *SimpleKMeans*. Esta función lo que hace es fijar de forma aleatoria cierta cantidad de puntos de referencia y agrupar las instancias en estas dependiendo de las semejanzas entre sus atributos. Los resultados del algoritmo se muestran en la

sección derecha y se puede visualizar la asignación de instancias a los clusters mediante las opciones de la Lista de Resultados.

2.3.5. Analizar los resultados

Los resultados del procesamiento de los datos sirven para que los encargados de tomar decisiones en la granja puedan hacerlo sobre la base del conocimiento ya que se puede saber, por ejemplo, si la cantidad de lluvia influye en que se pueda cumplir o no el plan de producción de los cultivos del mes de agosto. Esto es posible gracias a las estadísticas y relaciones que se observan en los resultados arrojados por Weka.

2.4. Herramientas informáticas empleadas

Para llevar a cabo el procedimiento anteriormente descrito se emplearon los softwares informáticos que se describen a continuación:

Visual Paradigm for UML (v8.0) para el modelado de la base de datos de los resultados de producción ya que es una herramienta CASE³ muy valorada por su capacidad de integración con los sistemas gestores de bases de datos y lenguajes de programación

PostgreSQL (v10.4) como sistema gestor de bases de datos relacionales de software libre, con el cliente de escritorio PgAdmin4 (v6.14) para facilitar la gestión de las bases de datos. Se incorporó además la extensión Postgis (v3.1.4) para el trabajo con componentes espaciales. Esta combinación es la misma que se utiliza en para el almacenamiento de datos del SIG de la granja.

Java (v1.8.0 update 202) es el lenguaje de programación utilizado, con Maven (v3.6.0) como gestor de dependencias, Spring Boot (v3.0) como framework para el desarrollo de servicios web e IntelliJ Idea (v2020.1) como entorno de desarrollo integrado.

³ CASE: Ingeniería de Software Asistida por Computación

Capítulo III: Resultados y discusión

En este capítulo se describen los resultados obtenidos durante la investigación. Estos se dividieron en dos partes para su valoración, por un lado todo lo referente a la obtención y estructuración de los datos necesarios para la toma de decisiones y por el otro los resultados del análisis propiamente dicho.

3.1. Resultados asociados a la estructuración de los datos

Luego del análisis de los datos que conforman el anuario y de las opiniones de los especialistas se logró definir el listado de indicadores que intervienen en el proceso de toma de decisiones del área agrícola de la granja. Se definieron 21 entre los que se encuentran características climatológicas como la velocidad del viento, porcentaje de humedad, temperaturas máximas y mínimas, especificaciones del terrero como el tipo de suelo, los niveles de fósforo, materia orgánica y calcio, así como el tipo de cultivo sembrado y el cumplimiento o no del plan de producción previsto. Todos estos datos están relacionados mediante el área geográfica a la que están asociados, lo que hace posible establecer relaciones entre estas informaciones.

La base de datos resultante con la información agrícola cuenta con un total de 864 registros correspondientes a los cultivos del período 2006-2010 y del 2022. Los que se usan para el análisis con Weka fueron los de 2022 ya que este es el único período de tiempo del cual se tienen datos en la base de datos del SIG. Esta base de datos tiene gran valor ya que no solo sirve para el apoyo a la toma de decisiones, sino que puede ser fuente de datos para otras investigaciones y también como estructura de almacenamiento para un sistema de gestión para la informatización del anuario de la granja. Otro elemento a destacar es la relación que se puede establecer entre los datos que en ella se almacenan y los del SIG teniendo como punto de articulación el componente espacial del área a la cual correspondan los valores.

La infraestructura de servicios web permite extraer información de las dos fuentes de datos

Teniendo la arquitectura descrita en el capítulo anterior funcionando correctamente en el servidor se hace una petición de tipo GET a la url http://<nombre_servidor>:<puerto>/arff/createFileByDate/2022/8 la cual devolverá *true* si se creó el archivo correctamente o *false* en caso contrario. Este servicio selecciona los datos correspondientes al mes de agosto del 2022 tanto de la base de datos del SIG como de la base de datos cultivos_guayabal de la cual se extraen los valores de plan de producción y producción real para saber si se cumplió el plan o no. Con esta información se genera el archivo *data-2022-8.arff* (ver fragmento en Ilustración 4).

Ilustración 4: Sección inicial del archivo data-2022-8.arff

40

valor desconocido y el porcentaje que representan del total que en este caso es cero, la cantidad de valores distintos que toma el atributo en el conjunto de datos que es 12 y el porcentaje de valores que están en solo una de las instancias, o sea que son únicos, que en este caso no hay ninguno (ver Ilustración 5). Por cada uno de los atributos se puede conocer los valores que toman y observar los histogramas de ellos (ver Ilustración 6).

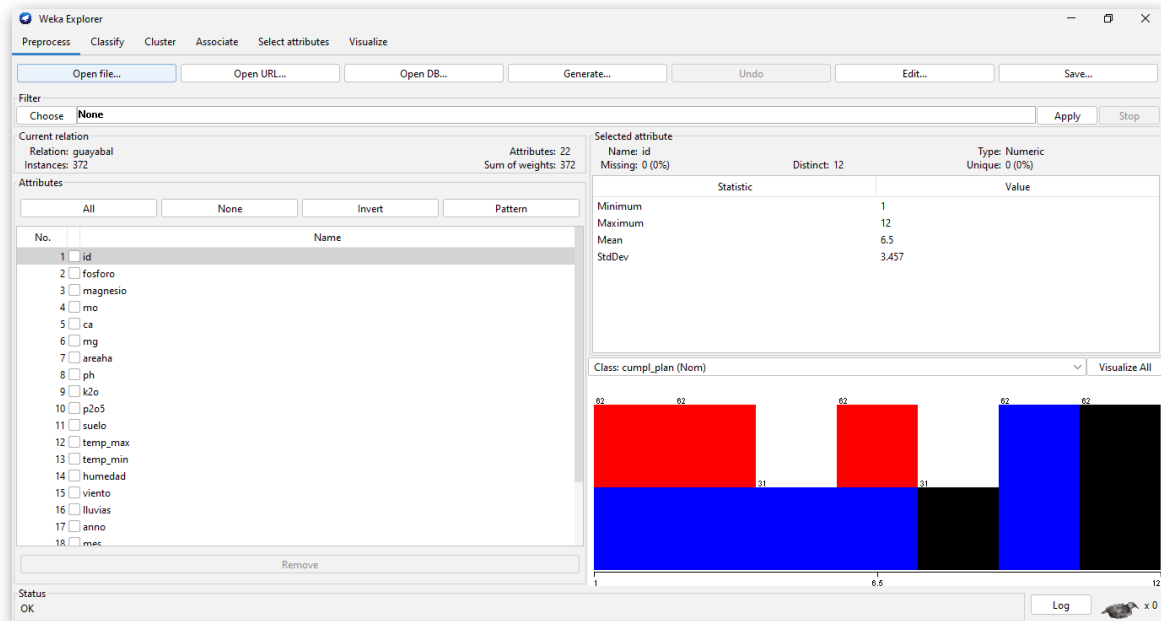


Ilustración 5: Interfaz de visualización de los datos recién cargados en Weka

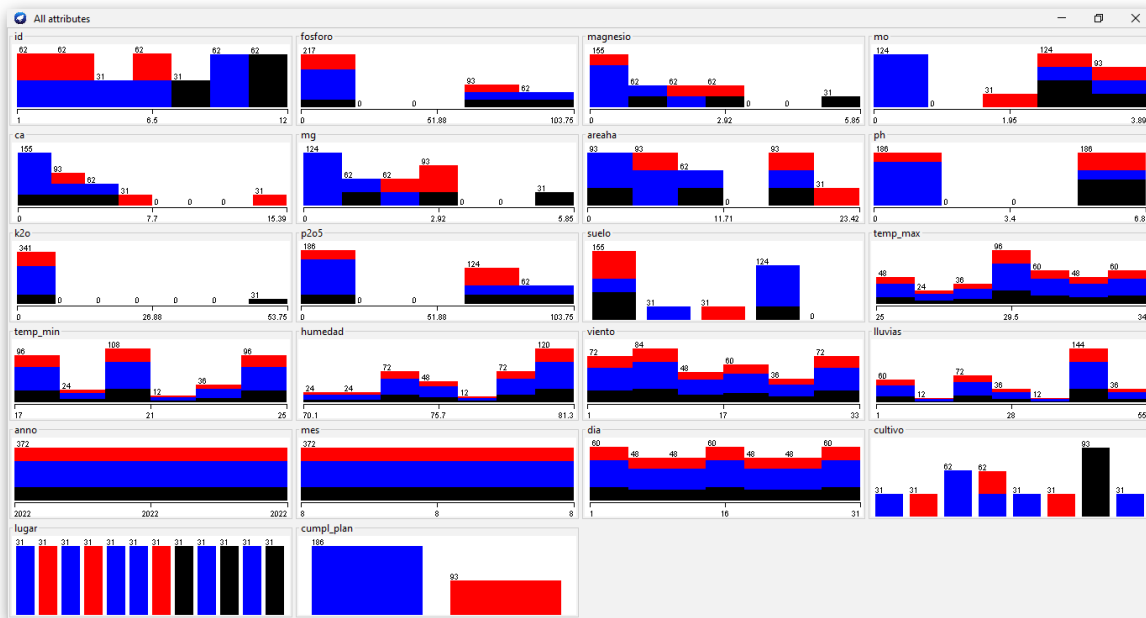


Ilustración 6: Interfaz con los histogramas de cada uno de los atributos del archivo data-2022-8.arff

Además, con solo cargar los datos se puede visualizar una matriz donde se ve la relación que existe entre todos los atributos. Aquí también es posible aumentar el tamaño de los puntos y la cantidad de instancias para ver cómo sería la interacción y si sufrirían grandes cambios (ver Ilustración 7).

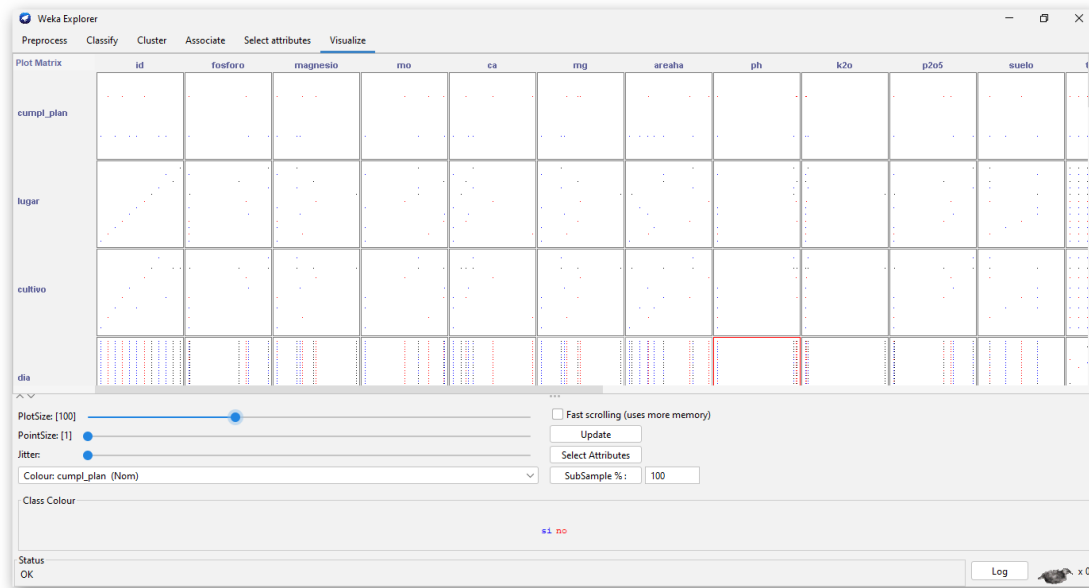


Ilustración 7: Visualización de todos los datos

Luego se realiza la clasificación para la cual se realiza un árbol de decisión, específicamente con el algoritmo J48. En los datos de salida del algoritmo se puede observar que el grado de confiabilidad del árbol es del 100% ya que de un total de 372 instancias fueron ignoradas 93 (tenían asignado el valor ? que identifica la clase como desconocida debido a que en esas áreas no se sembró y por tanto no había un plan de producción que cumplir) y las 279 restantes fueron clasificadas correctamente, teniendo 186 en el grupo de las que sí cumplieron el plan y 93 en el de las que no cumplieron (ver Ilustración 8). El árbol resultante de este proceso es el que se observa en la Ilustración 9.

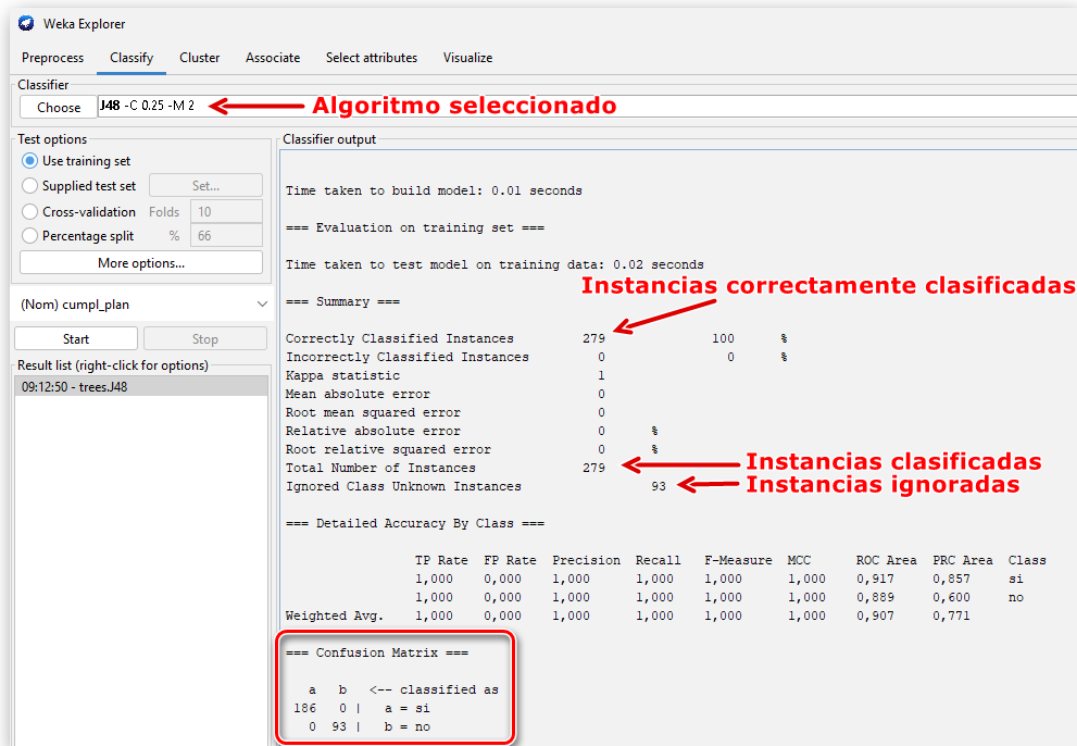


Ilustración 8: Análisis de los resultados del algoritmo de clasificación J48

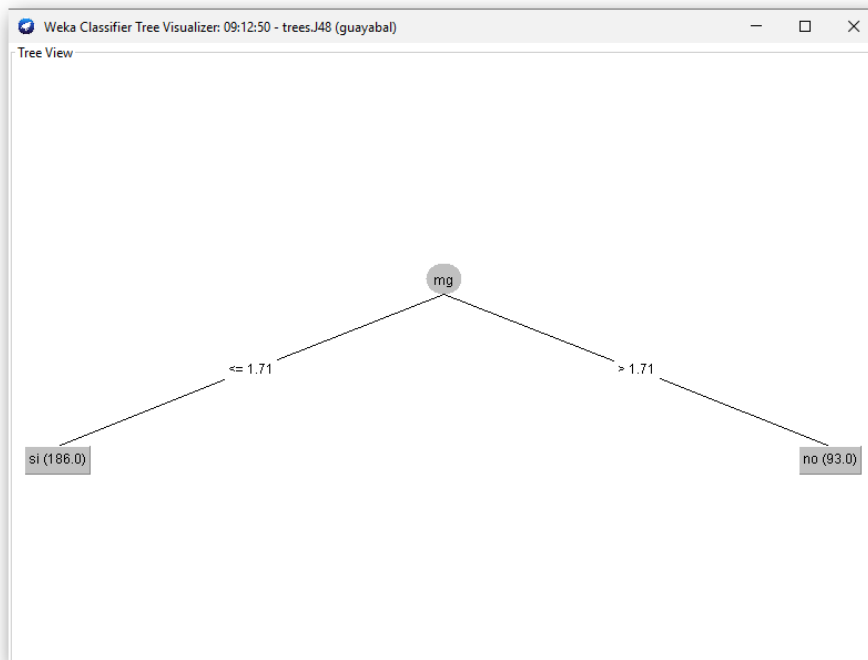


Ilustración 9: Árbol de Decisión obtenido con el algoritmo J48

Con este análisis se identificó que la variable que realmente influye en que se cumpla o no el plan de producción era la concentración de magnesio en el suelo ya que cuando esta era menor o igual a 1.71 sí se cumplió, pero cuando era mayor no.

Posteriormente se realizó el agrupamiento o clusters, donde al aplicar el algoritmo SimpleKMeans se obtuvo un total de dos clústers luego de tres iteraciones encontrando 186 instancias en cada uno de ellos lo que significa que en cada uno se puede ubicar el 50% del total de instancias (ver Ilustración 10).

Algoritmo seleccionado

Cantidad de iteraciones

Puntos de referencia iniciales (seleccionados de forma aleatoria)

Cantidad de instancias en el cluster

Attribute	Full Data (372.0)	Cluster# 0 (186.0)	1 (186.0)
id	6.5	4.5	8.5
fosforo	35.5408	12.9167	58.165
magnesio	1.4975	0.6	2.395
mo	2.0375	1.165	2.91
ca	3.3633	3.06	3.6667
mg	1.7225	0.6	2.845
areaaha	10.0492	5.1867	14.9117
ph	3.3167	0	6.6333
k2o	4.7708	0	9.5417
p2o5	40.9575	12.9167	68.9983
suelo	F_Rojo_Tipico F_Rojo_Hidratado	F_Rojo_Tipico	F_Rojo_Tipico
temp_max	29.8387	29.8387	29.8387

Ilustración 10: Análisis de los resultados del algoritmo de agrupamiento SimpleKMeans

También se puede visualizar esta información mediante gráficos para su mejor comprensión, permitiendo seleccionar qué variables se desea obtener en cada uno de los ejes de coordenadas, así como la posibilidad de aumentar el número de instancias (las genera con valores similares a los de los datos reales y ubicándolos en el clúster que le corresponde) para ver cómo sería el comportamiento de los clústers (ver Ilustración 11).

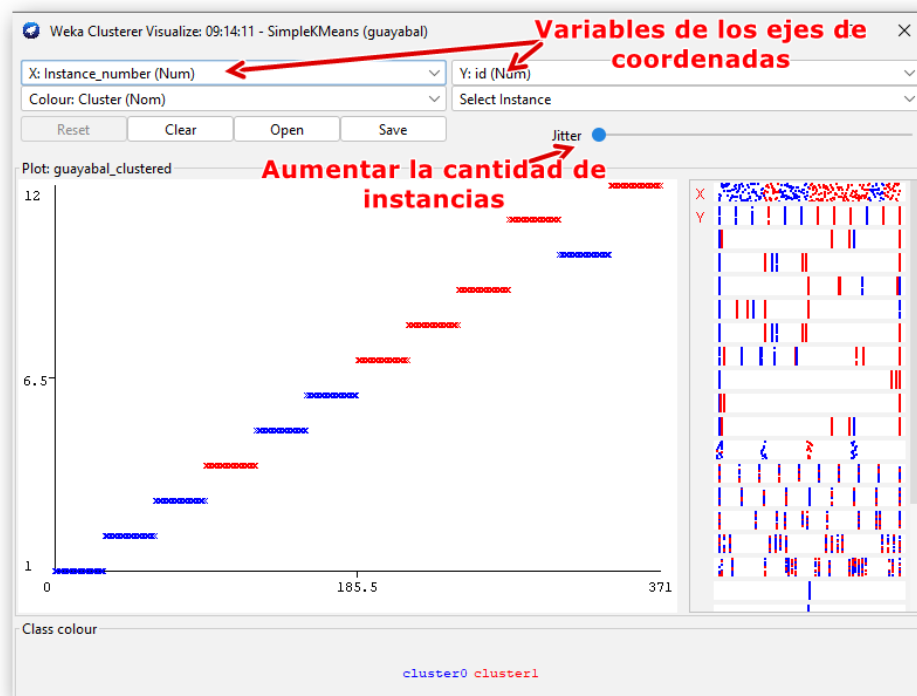


Ilustración 11: Visualización de la asignación de clústers con el algoritmo SimpleKMeans

Esta es solo una parte de la información que se pueden obtener de un conjunto de datos gracias a Weka y que puede servir como punto de partida para tomar decisiones sobre la base del conocimiento.

Conclusiones

- Se estableció un procedimiento a seguir para el apoyo a la toma de decisiones en el área agrícola de la UDP “El Guayabal” que consta de cinco pasos: identificar indicadores que determinan el cumplimiento o no del plan de producción, almacenar la información de producción asociándola al área geográfica donde está sembrada, relacionar esa información con el resto de indicadores mediante el componente espacial, procesar el conjunto de datos resultante y tener en cuenta los resultados durante la toma de decisiones.
- Luego del análisis de los referentes bibliográficos y del intercambio con directivos y productores de la granja se identificaron un total de 21 indicadores entre los que destacan aspectos climáticos, características de los suelos y valores de producción.
- Se obtuvo una arquitectura orientada a servicios que extrae información de dos fuentes diferentes, establece relaciones entre los datos por medio del componente espacial que poseen y los estructura en un archivo ARFF.
- Se analizaron las diferentes técnicas y algoritmos que posee Weka para el procesamiento de datos resultando elegidos el árbol de decisión J48 para la técnica de clasificación y el SimpleKMeans para el agrupamiento y sus resultados constituyen elementos a tener en cuenta para la toma de decisiones en el área agrícola de la UDP.

Recomendaciones

1. Realizar el análisis de los datos desde la misma infraestructura de servicios web utilizando la librería Weka en lugar del software de escritorio.
1. Incorporar al SIG funcionalidades que permitan consumir los servicios web y mostrar los resultados de los análisis desde su misma interfaz.

Bibliografía

Aimacaña Chuquimarca, J. P. y Columba Guanoluisa, A. R. (2021) *Análisis comparativo de algoritmos de Machine Learning para la detección de plagas en los cultivos representativos de la sierra ecuatoriana*. B.S. thesis. Quito: UCE.

Angulo Orobio, J. R. (2023) "El papel de la georreferenciación y las tecnologías en el campo de la zootecnia".

ASALE, R.- (2022) "georreferencia", «*Diccionario de la lengua española*» - Edición del Tricentenario. Disponible en: <https://dle.rae.es/georreferencia> (Accedido: 17 de febrero de 2023).

Astudillo Troya, W. G. (2022) *Agricultura 4.0 tecnologías empleadas en labores agrotécnicas*. B.S. thesis. BABAHOYO: UTB, 2022.

Bautista Andocilla, C. S. (2022) *Desarrollo de un sistema de riego inteligente para un terreno multicultivo de la parroquia "11 de noviembre" en Cotopaxi*. B.S. thesis.

Berenguel, M. M. y Gil, T. (2021) "Aplicación del IoT en la agricultura intensiva protegida", *II Jornadas de Doctorado en Informática de la UAL*. Universidad Almería, 115, p. 178.

Bonilla Bolaños, A. G. y Singaña Tapia, D. A. (2019) "La productividad agrícola más allá del rendimiento por hectárea: análisis de los cultivos de arroz y maíz duro en Ecuador", *LA GRANJA. Revista de Ciencias de la Vida*. Universidad Politécnica Salesiana, 29(1), pp. 70-83.

Bouza, C. y Santiago, A. (2012) "La minería de datos: árboles de decisión y su aplicación en estudios médicos", *Modelación Matemática de Fenómenos del Medio Ambiente y la Salud*, 2, pp. 64-78.

Brenes Carranza, J. A., Martínez Porras, A., Quesada López, C. U. y Jenkins Coronas, M. (2020) "Sistemas de apoyo a la toma de decisiones que usan inteligencia artificial en la agricultura de precisión".

Burbano, C. y Peñaranda, G. (2020) "Breve visión de los aspectos evolutivos de la agricultura de precisión como herramienta de gestión", *Revista Lumen Gentium*, 4(1), pp. 57-67.

Cadena Zumárraga, F. M. (2021) *Análisis del impacto económico del cambio climático en cultivos de quinua (Chenopodium quinua), en la provincia de Chimborazo*. Master's Thesis. Quito, EC: Universidad Andina Simón Bolívar, Sede Ecuador.

Camila Andrea, R. B. y Riveros Rey, C. E. (2023) "Estado del arte: aplicaciones de la agricultura 4.0".

Carranza Arévalo, G. E. (2020) *Predicción de los efectos del cambio climático sobre el daño potencial de una especie de cogollero en el cultivo de maíz (Zea mays) en la provincia de Tungurahua*. Master's Thesis.

Castro, Y., Ramos, G. L., del Sol González, Y. y Veloz Placencia, G. (2021) *Sesenta y tres medidas para incrementar la producción de alimentos en el país*, Granma.cu. Disponible en: <https://www.granma.cu/cuba/2021-04-15/sesenta-y-tres-medidas-para-incrementar-la-produccion-de-alimentos-en-el-pais> (Accedido: 14 de diciembre de 2022).

Chakkaravarthy, A. P., Pugalenth, R., Ramya, J. y Dhanalakshmi, J. (2020) "An Automatic Classification of COVID with J48 and Simple K-Means using Weka", *Int. J. Future Gener. Commun. Netw.*, 13, pp. 490-500.

Charania, I. y Li, X. (2020) "Smart farming: Agriculture's shift from a labor intensive to technology native industry", *Internet of Things*. Elsevier, 9, p. 100142.

Chugá Rosero, M. A. (2022) "Diseño e implementación de un sistema de riego inteligente y monitoreo de variables edafológicas y meteorológicas activado por IOT". Universidad Internacional SEK.

Climate FieldView (2023) *Digital Farming decisions and insights to maximize every acre*. Disponible en: <https://climate.com/> (Accedido: 31 de enero de 2023).

Consejo de Ministros de Cuba (2017) "Política Integral para el Perfeccionamiento de la Informatización de la Sociedad en Cuba 2017".

Consejo de Ministros de Cuba (2020) "Plan Nacional de Soberanía Alimentaria y Educación Nutricional", *La Habana: Gaceta Oficial de la República*.

DEL ROCIO, M. L. K. (2022) *IMPLEMENTACIÓN DE UNA APLICACIÓN WEB Y MÓVIL BASADO EN VISIÓN POR COMPUTADOR PARA EL CONTROL DE LA ENFERMEDAD SIGATOKA NEGRA EN EL CULTIVO DEL BANANO EN LA HACIENDA LOS ÁNGELES*. PhD Thesis. UNIVERSIDAD AGRARIA DEL ECUADOR.

Espinosa Pérez, R. R. (2022) *Plataforma de Cartografía Colaborativa para el manejo de las arvenses en el agroecosistema de la Granja Universitaria "El Guayabal"*. TESIS EN OPCIÓN AL TÍTULO DE MÁSTER EN BIOMATEMÁTICA. Universidad Agraria de La Habana «Fructuoso Rodríguez Pérez».

Franco, I. S., Guerra, G. A. P., García, D. N. M. y Pérez, D. M. E. R. (2022) "Procesamiento de consultas en un Sistema de Información Geográfica", *Anuario Ciencia en la UNAH*, 20(3).

Gálvez, J. J. T. y GÁLVEZ, J. J. T. (1994) "Metodología para el desarrollo de DSS". Instituto Tecnológico y de Estudios Superiores de Monterrey.

González-Teruel, J. D., Blanco, V., Blaya-Ros, P. J., Domingo, R., Soto-Valles, F. y Torres-Sánchez, R. (2021) "Estimación del nivel de estrés hídrico en frutales mediante técnicas machine learning para aplicación en sistemas de riego inteligentes", en *XLII Jornadas de Automática*. Universidade da Coruña, Servizo de Publicacións, pp. 477-484.

Guillen-Climent, M. L., Mas, H., Fernández-Landa, A., Algeet-Abarquero, N. y Tomé, J. L. (2020) "Uso de imágenes hiperespectrales para la predicción del marchitamiento de *Pinus halepensis* (Mill.) en el bosque mediterráneo", *Revista de Teledetección*, (55), pp. 59-69.

Harrington, P. (2012) *Machine learning in action*. Simon and Schuster.

Hernández Rodríguez, C. (2019) "Sistema de riego inteligente de bajo coste".

Jacobs, F. (2019) "Plan de negocios: empleo de drones en la agricultura de precisión". Universidad de San Andrés. Escuela de Negocios.

Lamelas, M. T. (2009) "Esquema metodológico para la toma de decisiones sobre el uso sostenible del suelo: Aplicación a la localización de suelo industrial", *GeoFocus. International Review of Geographical Information Science and Technology*, (9).

Lerma Sánchez, S. (2020) "Sistema de riego inteligente para jardines verticales". Universitat Politècnica de València.

MacDonald, M. L. y Faber, B. G. (2019) "Exploring the potential of multi-criteria spatial decision support systems: A system for sustainable land-use planning and design", en *Spatial Multicriteria Decision Making and Analysis*. Routledge, pp. 353-377.

Machado-García, N., González-Ruiz, L. y Balmaseda-Espinosa, C. (2014) "Recuperación de objetos geoespaciales utilizando medidas de similitud semántica", *Revista Cubana de Ciencias Informáticas*. Universidad de las Ciencias Informáticas, 8(2), pp. 132-144.

Machado-García, N., Machado-Molina, M., Rodríguez-Lohuiz, I. y Balmaseda-Espinosa, C. (2016) "Técnicas semánticas en la localización geoespacial de los suelos en Cuba utilizando medidas de similitud espacial", *Revista Ciencias Técnicas Agropecuarias*. 1986, Universidad Agraria de La Habana, 25(4), pp. 47-54.

Malvar, S., Badam, A. y Chandra, R. (2022) "FarmBeats: Digital Water for Agriculture", *Resource Magazine*. American Society of Agricultural and Biological Engineers, 29(4), pp. 40-42.

Márquez-Vasallo, Y., Salomón-Díaz, J. L. y Acosta-Roca, R. (2020) "Análisis de la interacción genotipo ambiente en el cultivo de la papa (*Solanum tuberosum* L.)", *Cultivos Tropicales*. Ediciones INCA, 41(1).

Martínez Devia, A. (2019) "La Inteligencia Artificial, el Big Data y la Era Digital: Una Amenaza para los Datos Personales", *Rev. Prop. Inmaterial*. HeinOnline, 27, p. 5.

Martínez, M. Y., Molina, M. M., García, N. M. y López, E. V. (2022) "Técnicas de aprendizaje supervisado para la detección y clasificación de enfermedades y defectos en imágenes de frutas: revisión", *Magazine de las Ciencias: Revista de Investigación e Innovación*, 7(1), pp. 1-16.

Mederos-Ramírez, A. y Ortiz-Pérez, R. (2021) "Análisis de la interacción genotipo ambiente en el cultivo de la soya (*Glycine max* (L) Merrill)", *Cultivos Tropicales*. Ediciones INCA, 42(1).

Mendoza Zambrano, A. M. (2022) *Aplicación del aprendizaje automático para la predicción de los tiempos de retención de pesticidas medidos en cromatografía líquida de fluidos supercríticos en muestras de espinacas*. B.S. thesis. Universidad del Azuay.

Mina, M. A. E. (2018) "Weka, áreas de aplicación y sus algoritmos: una revisión sistemática de literatura", *REVISTA CIENTÍFICA ECOCIENCIA*, 5, pp. 1-26. doi:10.21855/ecociencia.50.153.

Molina, M. M., Pereira, A. G. y García, N. M. (2015) "Sistema automatizado para la determinación del estado de maduración en fruta bomba", *Revista Ciencias Técnicas Agropecuarias*. Universidad Agraria de La Habana Fructuoso Rodríguez Pérez, 24, pp. 56-61.

Munévar, D. E., Ruiz, E., Díaz, W. D., Báez, D. E., Hernández, J. S., Salamanca, Ó. y Montoya, M. M. (2020) "Cosecha en cultivos de palma de aceite mediante el uso del grabber: caso de estudio en una plantación de Colombia", *Palmas*, 41(2), pp. 13-26.

Otero Barrera, G. (2019) *Sistemas de administración para la agricultura de precisión*. PhD Thesis. Universidad Central "Marta Abreu" de Las Villas, Facultad de Ingeniería

Paredes, M., Zúñiga, W., Caiza, A. F. M. y Mendoza, M. (2021) "Agricultura de precisión mediante WSN con nodos inteligentes aplicada a un sistema de riego en cultivo de mora", *Revista Perspectivas*, 3(2), pp. 26-30.

Patil, R., Deshmukh, S. y Rajeswari, K. (2015) "Analysis of SimpleKMeans with Multiple Dimensions using WEKA", *International Journal of Computer Applications*. Citeseer, 110(1).

Pedroza Salazar, J. y Fajardo Tascon, J. C. (2019) "Diseño de un prototipo de sistema de riego inteligente para el mejoramiento del uso del agua en invernaderos.". Unidad Central del Valle del Cauca (UCEVA).

Pérez, F. G. (1998) *Metodología para el desarrollo de sistemas DSS en una plataforma empresarial para la toma de decisiones*. PhD Thesis. Universidad Politécnica de Madrid.

Pérez Guerra, G. A. P., Sosa Franco, I., Ruiz Pérez, M. E. y Machado García, N. (2022) "Integración de distintas fuentes de datos en un Sistema de Información Geográfica", en. *XVI Conferencia Científica Ingeniería Agrícola AGRING 2022*, San José, Mayabeque.

Plazas Aquite, L. O., Candro Cuineme, D. F. y Díaz Rojas, M. J. (2022) *Plan de negocios para la creación de una empresa de servicios de captura y analítica de datos, basados en tecnología UAV (drones) apoyado en inteligencia artificial (IA) para la planeación, seguimiento y control de proyectos agrícolas*. Master's Thesis. Maestría en Gerencia de Proyectos.

Ponte, D., Espinosa, A., Gibeaux de González, S. y González, C. (2021) "Estado actual del aprendizaje automatizado aplicado al Internet de las cosas para automatizar procesos agrícolas". Universidad Autónoma de Chiriquí.

Ramón Fernández, F. (2020) "Inteligencia Artificial y Agricultura: nuevos retos en el sector agrario", *Campo Jurídico (Online)*. Faculdade São Francisco de Barreiras, 8(2), pp. 123-139.

Ricardo, J. E., Vázquez, M. Y. L., Palacios, A. J. P. y Ojeda, Y. E. A. (2021) "Inteligencia artificial y propiedad intelectual", *Universidad y Sociedad*, 13(S3), pp. 362-368.

Rodríguez, M. M. (2019) "Aplicación del IoT en la agricultura intensiva protegida", // *JORNADAS*, p. 178.

Solano, J. J. B., Salazar, E. P. y Solano, S. B. (2019) "Reconocimiento automático de patrones y características de las imágenes de los cultivos como alternativa para el desarrollo agrícola", *Pensamiento Actual*, 19(33), pp. 90-105.

Solano Robles, C. A. y Caballero Cruz, I. del P. (2022) "Modelo de predicción de plagas en el cultivo de palto utilizando metodología de aprendizaje automático supervisado, empresa Virú SA, 2019-2021". Universidad Privada Antenor Orrego.

Sosa Franco, I. (2022) *Método para la realización de consultas en un Sistema de Información Geográfica. Caso de estudio UDP "El Guayabal"*. TESIS EN OPCIÓN AL TÍTULO DE MÁSTER EN BIOMATEMÁTICA. Universidad Agraria de La Habana «Fructuoso Rodríguez Pérez». Disponible en: https://www.researchgate.net/profile/lvett-Sosa-2/publication/367165680_Metodo_para_la_realizacion_de_consultas_en_un_Sistema_de_Informacion_Geografica_Caso_de_estudio_UDP_El_Guayabal/links/63c4b6cad9fb5967c2de09f0/Metodo-para-la-realizacion-de-consultas-en-un-Sistema-

de-Informacion-Geografica-Caso-de-estudio-UDP-El-Guayabal.pdf (Accedido: 2 de febrero de 2023).

Trichkova-Kashamova, E. y Paunova-Hubenova, E. (2021) "Integrated software solutions in animal husbandry", en *2021 International Conference Automatics and Informatics (ICAI)*. IEEE, pp. 248-251.

Universidad de Waikato (2023) *Weka 3 - Data Mining with Open Source Machine Learning Software in Java*. Disponible en: <https://www.cs.waikato.ac.nz/~ml/weka/> (Accedido: 12 de febrero de 2023).

Valdivieso Duarte, J. A. (2021) *Estudio comparativo de las técnicas de inteligencia artificial para el diagnóstico de enfermedades en la agricultura*. B.S. thesis.

Valerieva Ivanova, D. (2022) "Evaluación de modelos para la predicción de la plaga Roya".

Velasco Mata, A., Vallez, N., Ruiz-Santaquiteria Alegre, J., Pedraza, A., Bueno, G. y Deniz, Ó. (2022) "Métodos de inteligencia artificial para la predicción de componentes químicos a partir de imágenes hiperespectrales", en *XLIII Jornadas de Automática*. Universidade da Coruña. Servizo de Publicacións, pp. 1056-1062.

Vite Cevallos, H., Carvajal Romero, H. y Barrezueta Unda, S. (2020) "Aplicación de algoritmos de aprendizaje automático para clasificar la fertilidad de un suelo bananero", *Conrado*. Universidad de Cienfuegos., 16(72), pp. 15-19.

World Food Programme (2023) *Una crisis alimentaria mundial*. Disponible en: <https://es.wfp.org/crisis-global-hambre> (Accedido: 17 de febrero de 2023).

World Health Organization (2020) *The state of food security and nutrition in the world 2020: transforming food systems for affordable healthy diets*. Food & Agriculture Org.