****

**“Metodología MLOps para la entrega continúa de un modelo de Machine Learning para el reconocimiento y control de las plagas Stenoma catenifer y heilipus lauri en el cultivo de aguacate Hass”**

**Juan Felipe Rodríguez**

**Anteproyecto presentado como requisito parcial para optar al título de:**

**Magister en Ingeniería de Software**

**Director:**

**Ph. David Arango**

**Pontificia Universidad Javeriana**

**Facultad de Ingeniería**

**Departamento de Electrónica y Ciencias de la Computación**

**Cali-Colombia**

**26 mayo de 2023**

FICHA RESUMEN

ANTEPROYECTO DE TRABAJO DE GRADO DE MAESTRÍA

TITULO: “\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Metodología MLOps para la entrega continua de un  
modelo de Machine Learning para el reconocimiento y control de las plagas Stenoma catenifer y heilipus lauri en el cultivo de aguacate Hass.\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_”

1. ÉNFASIS: Ingeniería \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Ingeniería de Software\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2. ÁREA DE INVESTIGACIÓN: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_área\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

3. ESTUDIANTE: \_\_\_\_\_\_\_\_\_\_\_nombre estudiante\_\_\_\_\_\_\_\_\_\_\_\_\_

4. CORREO ELECTRÓNICO: \_\_\_\_\_\_\_\_correo electrónico\_\_\_\_\_\_\_\_\_

5. DIRECTOR: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_nombre director\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

6. CO-DIRECTOR(ES): \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_nombre co-director\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

7. GRUPO QUE LO AVALA: \_\_\_\_\_\_\_\_nombre grupo de investigación\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

8. OTROS GRUPOS: \_\_\_\_\_\_\_\_nombre grupo de investigación\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

9. PALABRAS CLAVE: Big Data, Machine Learning, MLOps, Cultivo de aguacate, Plagas Stenoma catenifer y heilipus lauri

10. FECHA DE INICIO: 23 de 04 de 2023 DURACIÓN ESTIMADA: 6 Meses

11. RESUMEN (máximo una página).

La creciente aplicación de avances tecnológicos en diversos ámbitos de la vida ha llevado a la adopción de tecnologías innovadoras en la agricultura para mejorar la productividad y la eficiencia. Dentro de estas tecnologías, la metodología MLOps se destaca por su capacidad para mantener la operación del modelo y el despliegue mientras se mejora, se reentrena los modelos de aprendizaje automático, donde se optimiza la toma de decisiones y el aumento de la precisión de los resultados.

La investigación se centra en el cultivo del aguacate Hass, un componente crucial para la economía y el desarrollo socioeconómico en países como México y Colombia. Sin embargo, este cultivo enfrenta desafíos significativos, como las plagas Stenoma catenifer y heilipus lauri. Para combatir este problema, se plantea una serie de objetivos que giran en torno a la implementación de la metodología MLOps y un modelo de Machine Learning. La metodología MLOps se presenta como una solución prometedora para combatir este problema, proporcionando un marco de trabajo apropiado para los científicos de datos, lo que les permite integrar, automatizar y monitorear los modelos de Machine Learning.

En concreto, se busca implementar una metodología MLOps que permita la integración, automatización y monitoreo de un modelo de Machine Learning, específicamente diseñado para el reconocimiento y control de dichas plagas en el cultivo del aguacate Hass. Esta metodología seria validada mediante su despliegue en un entorno controlado, lo que permitiría monitorear y mejorar continuamente el rendimiento del modelo.

Además, se planea desarrollar y entrenar este modelo de Machine Learning utilizando técnicas apropiadas de preprocesamiento y selección de características para garantizar una detección de las plagas en el cultivo de aguacate Hass. Con esto se tendría una herramienta digital accesible para los científicos de datos, que facilite la predicción y prevención de la aparición de plagas, lo que constituirá un recurso valioso para ellos.

Finalmente, se espera que los resultados de este proyecto incluyan un informe detallado sobre el diseño, ejecución y evaluación de la metodología MLOps, así como la creación de una metodología MLOps que permita el monitoreo y la mejora continua del rendimiento del modelo de Machine Learning a desarrollar. Este enfoque MLOps permitiría un seguimiento más preciso y eficiente del cultivo de aguacate Hass, contribuyendo así a la sostenibilidad y productividad del sector agrícola.

**Tabla de contenido**

[**1.** **INTRODUCCIÓN** 5](#_Toc137575298)

[**2.1** **Planteamiento del problema** 6](#_Toc137575299)

[**2.2** **Formulación del problema** 8](#_Toc137575300)

[**2.3** **Alcance del trabajo de grado** 8](#_Toc137575301)

[**3.** **OBJETIVOS DEL PROYECTO** 9](#_Toc137575302)

[**3.1** **Objetivo general** 9](#_Toc137575303)

[**3.2** **Objetivos específicos** 9](#_Toc137575304)

[**3.3** **Resultados esperados** 9](#_Toc137575305)

[**4.** **Justificación** 10](#_Toc137575306)

[**5.** **MARCO DE REFERENCIA** 12](#_Toc137575307)

[**5.1** **Estado del arte** 12](#_Toc137575308)

[**5.2** **Marco teórico-conceptual** 16](#_Toc137575309)

[***5.2.1*** ***Metodología MLOps*** 16](#_Toc137575310)

[***5.2.2*** ***Modelo de Machine Learning*** 18](#_Toc137575311)

[***5.2.3*** ***Cultivo de aguacate Hass*** 21](#_Toc137575312)

[***5.2.4*** ***Plagas Stenoma catenifer y heilipus lauri*** 22](#_Toc137575313)

[**6.** **METODOLOGÍA DE LA INVESTIGACIÓN** 23](#_Toc137575314)

[**7.** **RECURSOS A EMPLEAR** 24](#_Toc137575315)

[**8.** **REFERENCIAS BIBLIOGRÁFICAS** 25](#_Toc137575316)

[**9.** **CRONOGRAMA DE TRABAJO** 28](#_Toc137575317)

[**10.** **GLOSARIO DE TERMINOS ESPECIALES** 29](#_Toc137575318)

# **INTRODUCCIÓN**

En los últimos años alrededor del mundo se viene implementando la metodología MLOps que permiten la implementación y el despliegue eficiente y escalable de modelos de aprendizaje automático (Machine Learning) para los científicos de datos en diferentes entornos productivos como los agrícolas. Este tipo de intervención tecnológica se traduce en la capacidad de desarrollar modelos de predicción y análisis de datos agrícolas, como pronósticos climáticos, análisis de suelos, monitoreo de cultivos y detección temprana de enfermedades o plagas (Organización de las Naciones Unidas para la Alimentación y la Agricultura, FAO, 2021).

El uso de los modelos de Machine Learning en la agricultura ofrece varias ventajas significativas, como por ejemplo permite aprovechar los datos recopilados de sensores y dispositivos IoT para tomar decisiones basadas en datos en tiempo real, asimismo optimizar el riego en el proceso de cosecha, la aplicación de fertilizantes y pesticidas y la planificación de la cosecha, que son aspectos clave para la producción agrícola necesarios para los científicos de datos en este ámbito económico.

Con la ayuda de la metodología MLOps, estos modelos propician la automatización de tareas repetitivas y complejas, como el procesamiento de grandes volúmenes de datos, la generación de informes y la gestión de la logística (Arley y Llano, 2016; Monsalve, 2021), acciones que ahorran tiempo y recursos, permitiendo a los científicos de datos se enfoquen en actividades estratégicas y en la toma de decisiones de manera eficaz.

En este sentido, el MLOps para el aprendizaje automático, es un enfoque que combina prácticas y herramientas de desarrollo de software con técnicas de aprendizaje automático, brindando a los científicos de datos una serie de beneficios significativos como la automatización de tareas repetitivas, el entrenamiento y despliegue de modelos, además, facilita la colaboración entre los equipos de ciencia de datos y operaciones, promoviendo la comunicación fluida y el intercambio de conocimientos.

El MLOps proporciona a los científicos de datos una infraestructura sólida y procesos eficientes para desarrollar, implementar y mantener modelos de aprendizaje automático, en los distintos escenarios económicos y productivos, donde estas estrategias favorecen a las prácticas de control de versiones y monitoreo continuo, garantizando la trazabilidad y el control de calidad de los modelos.

La adopción de la metodología MLOps en la agricultura obedece a la capacidad para mejorar la calidad y la precisión de los resultados, ya que los modelos de aprendizaje automático pueden analizar patrones complejos en los datos y generar predicciones más precisas sobre el rendimiento de los cultivos, la salud de los suelos y otros aspectos agrícolas.

1. **DEFINICIÓN DEL PROBLEMA DE INVESTIGACIÓN**

## **Planteamiento del problema**

Los aportes económicos, laborales, alimenticios y sociales de los procesos agrícolas son fundamentales para el desarrollo de un país, situación que pone de manifiesto la mirada de los entes gubernamentales y no gubernamentales internacionales y nacionales en el renglón de la economía agrícola y de la comunidad que se encuentra articulada en este ámbito económico.

La producción agrícola del aguacate Hass para el caso de México en el año 2019 correspondió a un 2.4 millones de toneladas, aportando el 45% en las exportaciones de este país y aumentando la cantidad de exportaciones en un 22% en el año 2020 (Cruz, Caamal, Pat y Reza, 2022). Estos aportes a nivel nacional se reflejan en PIB, contribuyendo al avance socioeconómico de las regiones agrícolas.

El aguacate Hass corresponde cerca del 82% de todos los aguacates el más consumido a nivel mundial. De acuerdo con los datos de la Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAOSTAT, 2021) el primer país productor de aguacate Hass es México con unas 2.393.849 toneladas al año, seguido de Colombia con unas 876.754 toneladas al año y de República Dominicana con 676.373 toneladas al año.

Los avances en la agroindustria en Colombia contribuyen al ámbito económico y laboral, siendo en la actualidad la producción agrícola del cultivo de aguacate Hass un producto de alta demanda a nivel nacional e internacional. Para el Departamento Administrativo Nacional de Estadística (DANE, 2016), Las problemáticas a tener en cuenta en el cultivo del aguacate Hass corresponden a: factores atmosféricos relacionados con la temperatura, las precipitaciones, el viento, la altitud, los factores de las condiciones del terreno y los factores relacionados con la siembra donde se encuentra la fertilización, los abonos y el tratamiento de las plagas y enfermedades.

Dentro de las plagas más importantes en el cultivo del aguacate Hass se encuentran la Stenoma catenifer y el heilipus lauri, insectos y larvas que introducen sus huevos provocando el daño en las semillas de los frutos en crecimiento. Además, el Stenoma catenifer impacta en el fruto al perforar el brote terminal y los laterales del aguacate, formando túneles de hasta 25cm, corta los pedúnculos y la base de los frutos pequeños, como resultado los frutos verdes y pequeños caen.

Para prevenir y controlar estas plagas, se recomienda implementar prácticas agrícolas adecuadas, como el manejo integrado de plagas, la selección de variedades resistentes y el control de la humedad en el suelo. Además, se deben realizar monitoreos constantes para detectar y tratar a tiempo cualquier plaga o enfermedad que pueda aparecer en el cultivo del aguacate Hass.

Para la detección de este tipo de plaga en la producción agrícola del cultivo del Hass existe el método Manejo Integrado de Plagas (MIP) en el cual se señala el monitoreo de manera manual y de observación constante partiendo de tres elementos, el primero corresponde a la prevención como cuidados, restricciones y limpieza del personal y sus utensilios de trabajo, el segundo al control donde se utiliza evaluaciones y registros manuales, instalación de trampas y el tercero es el manejo de la enfermedad en el cual se genera una protección y cuidado de las plantas dependiendo de los patógenos dañinos (Instituto Agropecuario Colombiano, 2012)

En este sentido el Machine Learning permite a los científicos de datos, a través de imágenes el reconocimiento de patrones de concentración y expansión de las plagas de manera óptima en todo el cultivo generando una reducción económica y mejorando la calidad del producto agrícola.

El cultivo del aguacate Hass en Colombia ha tenido una gran demanda a nivel nacional e internacional, generando un crecimiento del 34% del total de área sembrada de aguacate, además al ser un producto que presenta una cosecha constante por las condiciones del relieve y climáticas del país viene en un crecimiento de área sembrada de un 65% (Colombia Mide, 2021). Dadas estas circunstancias, para los científicos de datos la implementación de la metodología MLOps se presenta como una oportunidad para mejorar la calidad, confiabilidad y eficiencia de los modelos de Machine Learning utilizados en la detección de plagas, evaluación del nivel de daño y reconocimiento de deformaciones y coloraciones específicas en las áreas afectadas. Al someter los modelos a rigurosos procesos de control de calidad, los científicos de datos con esta metodología garantizan la trazabilidad y transparencia a lo largo de todo el ciclo de vida del modelo, brindando así resultados más precisos y confiables.

La utilización de MLOps se ha convertido en una práctica cada vez más extendida en el campo de la ciencia de datos, y se ha demostrado que mejora significativamente la eficiencia y la seguridad en la implementación de modelos de Machine Learning (Géron, 2019). Su aplicación en el contexto de la agricultura y la predicción de plagas y enfermedades puede ser un paso importante para mejorar la productividad y sostenibilidad del cultivo de aguacate Hass y otros cultivos agrícolas.

Es importante destacar la relevancia de utilizar metodologías de MLOps para garantizar el correcto desarrollo, implementación y mantenimiento continuo de un modelo de control y cuidado de plagas permitiendo mejorar los procesos productivos agrícolas en Colombia. El modelo de Machine Learning al ser un programa de automatización y actualización constante de sus tareas avanza en el mejoramiento y la eficiencia de su procesamiento de información de manera continua a través de la metodología MLOps aportando a los procesos de análisis a los científicos de datos.

## **Formulación del problema**

En este contexto la investigación busca desarrollar una herramienta digital que ayude a pronosticar y prevenir la posible presencia o no de las plagas como el Stenoma catenifer y el heilipus lauri en el cultivo de aguacate Hass, entendiendo que es crítico la detección temprana del brote en un cultivo, se propone la creación de un software accesible para los científicos de datos que les permita abordar esta problemática. Con base en esto, surge la siguiente pregunta de investigación: ¿Cómo el uso de la metodología MLOps en el desarrollo de un modelo de Machine Learning facilita la integración, la actualización y el despliegue continuo del reconocimiento de las plagas Stenoma catenifer y heilipus lauri en el cultivo de aguacate Hass, contribuyendo a mejorar los modelos agrícolas de forma automática y brindando beneficios económicos y sociales a la comunidad de científicos de datos? asimismo ¿Cómo mantener el programa de Machine Learning de forma automatizada y con supervisión continua, de modo que no se vea comprometido su rendimiento?

## **Alcance del trabajo de grado**

Los alcances que se buscan desarrollar en esta investigación corresponden a una implementación de una metodología de MLOps para contribuir a un modelo de Machine Learning permitiendo la integración, la actualización y el despliegue continuo del reconocimiento de plagas Stenoma catenifer y heilipus lauri en el cultivo de aguacate Hass, partiendo de las siguientes fases:

* ***Experimentación / Desarrollo / Pruebas***.

Feature Store: validación y preparación de los datos

Código de fuente

Model Registry

* ***Pre-Producción y Producción***.

El modelo se despliega y se monitoriza

Pipeline

* ***Fase de despliegue del modelo***

El desarrollo de una metodología MLOps que permita la implementación de modelos con mayor rapidez con procesos automatizados, contribuyendo a acelerar el tiempo de creación de valor al entregar información de manera ágil.

Fuera del alcance se encuentra la optimización del modelo de manera constante y la reutilización del modelo a medida que los datos evolucionan con el tiempo generando de manera efectiva mejoraras al rendimiento y la adaptabilidad del modelo a partir de técnicas de transferencia de aprendizaje, donde se ajustan los pesos y las capas del modelo para adaptarse a los nuevos datos.

# **OBJETIVOS DEL PROYECTO**

## **Objetivo general**

Implementar la metodología MLOps para la entrega continua de un modelo de Machine Learning para el reconocimiento y control de las plagas Stenoma catenifer y heilipus lauri en el cultivo de aguacate Hass.

## **Objetivos específicos**

* Desarrollar una metodología MLOps que permita la integración, automatización y monitoreo del modelo de Machine Learning diseñado para el reconocimiento y control de las plagas Stenoma catenifer y heilipus lauri en el cultivo de aguacate Hass.
* Validar el uso de MLOps mediante despliegue en un ambiente controlado, con la capacidad de monitorear y mejorar continuamente el rendimiento del modelo.
* Desarrollar y entrenar un modelo de Machine Learning utilizando técnicas apropiadas de preprocesamiento y selección de características, así como algoritmos de aprendizaje supervisado o no supervisado, para lograr una detección de las plagas Stenoma catenifer y heilipus lauri en el cultivo de aguacate Hass.
* Implementar técnicas de procesamiento de imágenes para extraer características relevantes y mejorar la capacidad del modelo de Machine Learning en el reconocimiento y detección de las plagas Stenoma catenifer y heilipus lauri en el cultivo de aguacate Hass a partir de imágenes capturadas en campo.

## **Resultados esperados**

En esta propuesta de investigación para la maestría en desarrollo en ingeniería de software se plantearon los siguientes resultados esperados:

1. Creación de un modelo de Machine Learning para detectar las plagas Stenoma catenifer y heilipus lauri en el cultivo de aguacate hass.
2. La implementación de la estrategia MLOps que permitirá la integración, automatización y monitoreo continuo del modelo de Machine Learning.
3. Generación de una vista de despliegue que proporcione los componentes de la metodología MLOps.
4. La demostración del modelo de Machine Learning utilizando la metodología MLOps, con la capacidad de monitorear y mejorar continuamente el rendimiento del modelo.

# **Justificación**

Los avances técnicos y tecnológicos a nivel mundial han permitido procesos de interacción en todos los ámbitos de la vida de las personas como la educación, la economía, la política, la cultura y el medio ambiente. Estas transformaciones sociales a nivel mundial generan acciones de aprendizajes y de mercados globales, donde las tecnologías y los aportes de nuevos o mejorados software “generan reducciones significativas de costos por la experiencia y utilización de patentes o porque aportan beneficios por la capacidad de vender variedades similares de productos en diversos mercados” (Vela, 2012).

En la actualidad, la agricultura es un renglón económico que viene identificando diseños de políticas públicas para la intervención de aplicaciones tecnológicas y generar el aprovechamiento y mejoras de la productividad agraria a nivel mundial. Estas iniciativas resultan de las dificultades económicas y de productividad agraria, las cuales deben de ser transformadas para el desarrollo del campo, como explica la FAO (2021):

La incorporación de estas tecnologías supone la generación de información (basada en la recopilación y procesamiento de datos) e indicaciones que permiten el monitoreo, el análisis, la planificación y el control inteligente de procesos de producción, transformación, distribución y comercialización de productos agrícolas (p.88).

El uso de tecnología y datos se ha vuelto cada vez más importante para mejorar la eficiencia y la productividad en la agricultura. En este contexto, el empleo de la metodología MLOps (Machine Learning Operations) se justifica como una herramienta poderosa para potenciar la agricultura de precisión y lograr una gestión más eficiente de los recursos. Estas mejoras en América Latina proporcionan una sostenibilidad en la producción agrícola y contribuye a procesos más saludable para el medio ambiente (Ver imagen 1).

Grafica 1 Acceso y aprovechamiento de tecnología digital en la agricultura de América del Sur

Fuente: Información tomada de FAO (2021)

El uso de la metodología MLOps en la agricultura ofrece mejoras debido a su capacidad para optimizar la toma de decisiones, automatizar tareas y mejorar la precisión de los resultados. Esta tecnología puede marcar una gran diferencia en la eficiencia y la sostenibilidad de la agricultura, ayudando a enfrentar los desafíos actuales y futuros del sector.

De allí que, la metodología MLOps tiene un gran potencial para transformar la agricultura y brindar beneficios significativos a los científicos de datos involucrados en esta industria, debidoa pueden aprovechar varias ventajas, como por ejemplo el MLOps permite una mejor gestión de los modelos de aprendizaje automático, es posible utilizar prácticas de control de versiones para rastrear y gestionar los cambios en los modelos, lo que facilita la colaboración y la reproducibilidad de los resultados en las áreas de producción agrícola. Además, el MLOps garantiza un monitoreo continuo de los modelos en producción, lo que permite identificar y solucionar problemas rápidamente.

Al utilizar técnicas de aprendizaje automático, los científicos de datos pueden analizar grandes volúmenes de datos agrícolas para identificar patrones y tomar decisiones informadas, ayudando a implementar estos modelos en sistemas integrados, lo que permite la automatización de tareas agrícolas como el riego, la fertilización y la detección de enfermedades. Esto conduce a un uso más eficiente de los recursos, reduciendo costos y minimizando el impacto ambiental.

En este sentido la tecnología constituye una herramienta relevante para el mundo actual, de allí que este tipo de investigaciones permitan abordar métodos aplicables en la ingeniería de software tanto para la generación de conocimientos, como métodos de aplicación en escenarios reales, como la agricultura y, en particular, el cultivo del aguacate Hass. Asimismo, esta clase de investigaciones abre la vía para la articulación entre los profesionales, las universidades y la industria agrícola para fomentar prácticas económicas y tecnológicas, mejorar y fortalecer actividades de investigación de laboratorios de computación y programas de alfabetización digital, con el objetivo de contribuir al desarrollo económico, científico y académico, asegurando que la ingeniería de software se emplee de forma concreta y provechosa en la solución de problemas prácticos en diversas áreas, como la agricultura.

# **MARCO DE REFERENCIA**

El marco de referencia está organizado en un primer momento con el análisis del estado del arte para la comprensión del objeto de estudios sobre MLOps para el ámbito agrícola y un segundo momento con los elementos teórico-conceptuales que permiten la argumentación teórica sobre Machine Learning, la metodología MLOps y el cultivo de aguacate Hass con las plagas detalladas para esta investigación.

## **Estado del arte**

Las investigaciones a nivel internacional se describen a continuación

Tabla 1 Investigaciones internacionales

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Autor** | **Titulo** | **Objetivo** | **Metodología** | **Resultado** |
| Aimacaña C y Columba G (2021) | Análisis comparativo de algoritmos de Machine Learning para la detección de plagas en los cultivos representativos de la sierra ecuatoriana | Realizar un análisis comparativo entre distintos algoritmos de Machine y Deep Learning, que permita determinar cuál de ellos ofrece mejores resultados en la detección de plagas en cultivos endémicos de la serranía ecuatoriana. | Metodología CRISPDM la cual sus siglas significan Proceso Estándar Entre Industrias para la Minería de Datos, | El estudio realizado se aplicó el mejor modelo para la construcción de un prototipo el cual fue la red neuronal convolucional InceptionV3; el prototipo realizado, permite al usuario enviar imágenes para ser procesadas por el modelo y este a su vez reciba una predicción he información de los síntomas, además de una sugerencia de un posible tratamiento |
| Castañeda V., Guerrero M., Renteros P y Villanueva M  (2021) | Detección de nutrientes del suelo y planta, y pestes en campos de cultivo de banano orgánico con Machine Learning | Diseño una plataforma que sirva de apoyo al agricultor para poder estimar que macronutrientes tiene en deficiencia su planta de banano, con la finalidad de obtener un mejor producto en la cosecha del banano orgánico, así como una posible interfaz para la detección de plagas en el banano | Se utilizó un conjunto de fotografías público, el cual fue encontrado en un repositorio y se escogieron aquellas plagas que afectaban al banano. Posteriormente, se realizó un aumento a este conjunto de fotografías mediante transformaciones lineales y las imágenes resultantes fueron pre procesadas en diferentes espacios de color para ser utilizadas como entradas a la red neuronal. | El modelo permitió una alta precisión a través de diferentes métricas. Continuamente se desarrolló un prototipo de plataforma web para que los agricultores en un futuro pudieran acceder al sistema. |
| Valenzuela, C (2022) | Detección y Clasificación de Enfermedades en el Tomate Mediante Deep Learning y Computer Visión | Aplicar el aprendizaje profundo a problemas de visión de computadora tales como detección y clasificación específicamente en las enfermedades del tomate mediante el procesamiento de imágenes digitales. | Neuronales pre-entrenadas (transfer learning), para hacer la detección de la hoja de tomate (primera: Faster Mask R-CNN) y luego a partir de la detección, realizar la clasificación de la enfermedad (segunda: red neuronal convolucional), tomando como entrada el área de la imagen donde se encuentra la hoja, luego de realizar la clasificación el sistema desarrollado proporciona información de los síntomas asociados a la enfermedad, como también como proceder con la prevención y el tratamiento a seguir en tiempo real | Con el advenimiento de las redes neuronales, ha permitido mejorar drásticamente la precisión, tanto de la detección de objetos como de la clasificación de imágenes. |
| García F y Ulloa L (2022) | Implementación de modelo Machine Learning aplicado al estudio de enfermedades de café en el centro de investigación Sacha Wiwa, perteneciente a la parroquia Gasaganga, Cantón la Maná, provincia de Cotopaxi | Implementar modelo de clasificación de imágenes empleando técnicas de Machine Learning, para la detección de enfermedades del Cafeto (Coffea arábica) en huertas del Centro de Investigación Sacha Wiwa. | Cualitativa experimental permite la selección, análisis y presentación de los datos documentados de una manera ordenada y siguiendo los objetivos del proyecto | Mediante el modelo de Machine Learning desarrollado en el entorno virtual “Jupyter lab” se logró crear y entrenar las redes neuronales convencionales para la clasificación de las enfermedades del café expuestas en el caso de estudio en conjunto con la clasificación sana del cafeto |

Fuente: Elaboración del investigador

En las investigaciones a nivel internacional se pude observar que el desarrollo de Machine Learning, en particular el uso de redes neuronales, ha demostrado ser de gran importancia en el control de plagas y enfermedades en diversos campos, como la agricultura, asimismo este utiliza un modelo de red neuronal convolucional para la detección de enfermedades como por ejemplo el cultivo de café, tomate y banano es un ejemplo claro de cómo el Machine Learning puede contribuir al control de plagas.

Uno de los principales beneficios del Machine Learning en el control de plagas es su capacidad para analizar grandes volúmenes de datos y extraer patrones y características relevantes, en las investigaciones se observaron que utilizaban imágenes de plantas enfermas como sanas, lo que le permitió aprender a reconocer los síntomas y distinguir entre diferentes enfermedades. Esto es especialmente útil en el control de plagas, ya que puede facilitar la detección temprana de enfermedades y ayudar a los agricultores a tomar medidas preventivas o correctivas de manera oportuna.

Para el caso de las investigaciones nacionales se detallan en la siguiente tabla:

Tabla 2 Investigaciones a nivel nacional

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Autor** | **Titulo** | **Objetivo** | **Metodología** | **Resultado** |
| Tovar, M(2019) | Diseño e implementación de un aplicativo móvil para realizar la detección temprana de la enfermedad de la Sigatoka Negra en los cultivos de plátanos | Diseñar e implementar un aplicativo móvil, para realizar la detección temprana de la enfermedad de la Sigatoka Negra en los cultivos de plátanos. | La metodología utilizada para la detección de esta enfermedad, consistió en construir una base de datos que contiene imágenes de los posibles estados de la misma, teniendo en cuenta la escala estándar de Fouré utilizada para analizar la forma de evolución de la Sigatoka Negra. Luego se inició con la creación del algoritmo, en donde se realizó la segmentación de cada una de las imágenes para generar un realce en la enfermedad, para posteriormente obtener las características respectivas y finalmente a partir de las técnicas de Machine Learning generar una clasificación que identifique el estadío en el que se encuentra una hoja al tomar una imagen que no se ubique dentro de la base de datos. | El aplicativo móvil, demostró un porcentaje de acierto del 80% en la detección de la Sigatoka Negra por lo tanto, la aplicación puede llegar a ayudar a los productores de plátano, evitando así perdidas económicas. |
| Avila, P  (2022) | Entrenamiento de redes neuronales para la identificación de plagas en cultivos de café. | Implementar un aplicativo basado en visión artificial para la identificación de Cercospora, Phoma, Roya, minador de hojas y Arañita roja, en el café a partir del análisis de las hojas de la planta | Esta base será introducida a modelos de redes neuronales: densa, neuronal convolucional y neuronal convolucional con overfittting, usando el lenguaje Python y las bibliotecas: Keras y TensorFlow. Una vez entrenado el modelo, se procede al análisis de las imágenes que contienen la morfología de la planta y que servirán de entrada a una red neuronal para poder comparar los resultados y determinar cuál de los modelos presenta mejores resultados | El modelo convolucional con overfitting, presenta el mejor resultado de predicción, pero aun así, es un modelo limitado que podría ser remplazado por otros modelos de mayor alcance, como: ResNet50, CIFAR-10, VGG19.  Al tener una cantidad de plagas tan limitada, es evidente que no se puede hacer un reconocimiento más detallado de las plagas, por lo que sería necesaria ampliar la base de dato |

Fuente: Elaboración del investigador

A nivel nacional los diferentes modelos operativos en el contexto de la detección de plaga hace referencia un modelo convolucional con overfitting que presenta los mejores resultados de predicción, pero se plantea la posibilidad de reemplazarlo por otros modelos de mayor alcance, como ResNet50, CIFAR-10 o VGG19.

La elección del modelo operativo es crucial en cualquier aplicación de Machine Learning, ya que tiene un impacto directo en la precisión y el rendimiento del sistema. En el caso mencionado, el modelo convolucional con overfitting logró un alto nivel de acierto en la detección de plaga

## **Marco teórico-conceptual**

### ***Metodología MLOps***

Machine Learning Operations (en sus siglas en ingles MLOps) es un conjunto de prácticas que combinan Machine Learning, DevOps e Ingeniería de datos, cuyo objetivo es implementar y mantener sistemas ML en producción de manera confiable y eficiente.

En esa dirección Rivero (2022) explica que, MLOps se refiere a las prácticas y metodologías que permiten gestionar y optimizar el ciclo de vida de los sistemas de aprendizaje automático (Machine Learning) de forma que se puede desarrollar confiablemente. Se tiene en cuenta que, la metodología MLOps combina técnicas y herramientas de desarrollo de software con conocimientos de ciencia de datos y operaciones de infraestructura, para permitir la implementación, el monitoreo y el mantenimiento escalable de los modelos de aprendizaje automático en producción.

En el contexto del desarrollo de modelos de Machine Learning, MLOps aborda los desafíos asociados con la gestión de datos, la experimentación, el entrenamiento, la implementación, el monitoreo y la reevaluación continua del rendimiento del modelo, que lo hacen diferente a DeOps y la Ingenieria de Datos como se describe en la Tabla1. Esto implica el uso de técnicas como la integración y entrega continua (CI/CD), la orquestación de flujos de trabajo, la automatización de pruebas y la gestión de versiones.

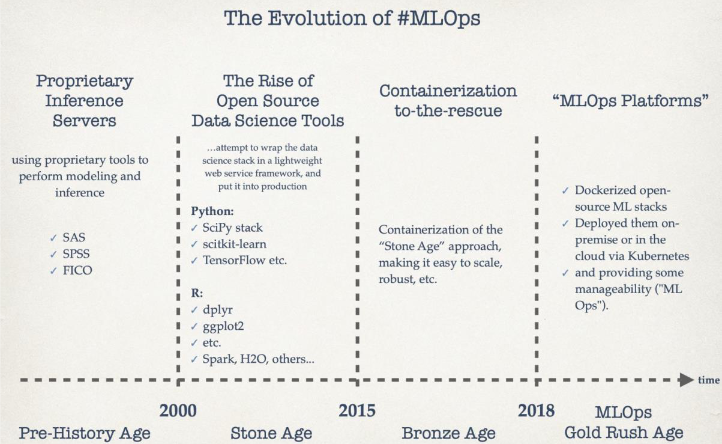
*Tabla 1 Comparativo entre DevOps, Ingeniería de Datos y MLOps*

|  |  |  |  |
| --- | --- | --- | --- |
| **Práctica** | **DevOps** | **Ingeniería de Datos** | **MLOps** |
| Control de versiones | Versionamiento de código | Versionamiento de código linaje de datos | Versionamiento de código + datos + modelos (conectados) |
| Pipeline | n/a | Flujo de datos/ETL | Flujo de ML de entrenamiento, flujo de ML para predicción |
| Validación de comportamiento | Pruebas unitarias | Pruebas unitarias | Validación de modelo |
| CI/CD | Despliegue de código en producción | Despliega código en flujo de daros | Despliegue de código a producción + Flujo de datos de ML |
| Validación de datos | n/a | Validación de negocio y formato | Validación estatística |
| Monitoreo | Basados en SLOs | Basados en SLOs | SLOs + monitoreo diferencial y estatístico |

Fuente: Rivero (2022)

El objetivo fundamental de MLOps es garantizar que los modelos de aprendizaje automático sean confiables, reproducibles y escalables en entornos de producción. Al adoptar prácticas de MLOps, las organizaciones pueden acelerar el tiempo de comercialización de sus modelos, mejorar la colaboración entre equipos multidisciplinarios, minimizar los riesgos asociados con los modelos en producción y facilitar la implementación de mejoras y actualizaciones. Estos avances en la metodología MLOps se vienen desarrollando desde el año 2000 para resolver los distintos problemas para los contextos ML (ver imagen 1).

Imagen 1 Evolución temporal de MLOps



Fuente: Visengeriyeva, et al., (2020)

La metodología MLOps es una disciplina que fusiona las mejores prácticas de desarrollo de software con los desafíos específicos del Machine Learning, con el objetivo de facilitar la implementación y gestión de modelos de aprendizaje automático. Con el auge del software de código abierto y la disponibilidad de datos, más profesionales de software comenzaron a usar bibliotecas Python o R para entrenar modelos ML.

A medida que iba surgiendo la tecnología de contenerización, se resolvió el despliegue del modelo de forma escalable mediante el uso de contenedores Docker y Kubernetes. Recientemente, vemos la evolución de esas soluciones hacia plataformas de implementación de ML que cubren toda la iteración de experimentación, capacitación, implementación y monitoreo de modelos (Visengeriyeva, et al., 2020).

***Proceso de la metodología MLOps***

* **Flujos de ML:** Los flujos de datos, también llamados pipelines de datos o procesos ETL (extraer, transformar y cargar), son procesos en que la data es obtenida desde una fuente, para ser transformada y luego cargada en la misma fuente o en otro sistema. Las transformaciones de los datos se refieren a todos los cambios necesarios sobre los mismos, para que quede con el formato que es requerido en el destino final (el modelo de ML).
* **Versionado de modelos y de datos**: En aprendizaje automático es igual de importante mantener registro de las diferentes versiones del código, pero se le suma también la necesidad de registrar las versiones del modelo, así como de los datos que fueron utilizados para entrenarlo, y de otros datos como hiper-parámetros del modelo.
* **Validación de modelos**: Se debe tener en cuenta:

***Precisión del modelo***: de los casos que el modelo clasificó como TP

***Recall del modelo***: la precisión abierta por categorías.

***Medición de sesgo en los datos***: es importante realizar mediciones que ayuden a entender si el modelo está sesgado

***Casos críticos***: un modelo puede tener buena precisión, recall y no estar sesgado, pero no agarrar los casos más importantes en la historia reciente.

***Medición de un modelo de regresión***:

- Métricas globales del modelo como RMSE, MAE Y R2.

- Las mismas métricas pero aperturadas por diferentes cortes, para entender si hay algún segmento en el cual el modelo tenga oportunidades de mejora.

- Medición del sesgo en los datos, igual que en los problemas de clasificación.

- Casos críticos, igual que en los problemas de clasificación.

* **Validación de datos**: se genera en dos elementos, el primero corresponde al más básico es la calidad de los mismos, lo cual incluye verificar cantidad de nulos, el SLA (service level agreement) de actualización de los datos, que todos los datos sean del tipo esperado, entre otros. El segundo es el más complejo consiste en monitorear la distribución de los datos. Un cambio en la distribución puede ser originada por una falla en alguno de los pipelines o por cambios en los datos (Rivero, 2022).

### ***Modelo de Machine Learning***

De acuerdo con Murphy (2012), el modelo Machine Learning se refiere a la localización automatizada de datos que se encuentran estandarizados, además es una herramienta constante en el proceso de extracción de grandes volúmenes de información para su organización y análisis, definido también como Big Data. En la actualidad el Machine Learning se encuentra en los motores de búsqueda con los cuales se desarrolla resultados instantáneos como por ejemplo los anuncios, software antispam, las transacciones con la tarjeta de crédito, etc.

Estas nuevas tecnologías, a través del internet, permiten la recopilación de bases de datos de un tamaño considerable, pueden ser históricos o datos en tiempo real, siendo posible el análisis de datos debido a las mejoras constantes en la tecnológica y el desarrollo de software. Para Shalev & Ben (2014), este desarrollo ha permitido que las bases de datos se encuentren en distintos formatos, lo que es un desafío para orientar estrategias que permitan resolver distintos problemas de acuerdo a las necesidades de cada formato y de esta manera avanzar en el desarrollo de métodos para el Machine Learning.

El Machine Learning al ser un proceso automatizado que retoma patrones en grandes cantidades de datos, de allí que se construyan los modelos utilizados en aplicaciones de análisis de datos predictivos teniendo en cuenta:

1. Técnicas de aprendizaje automáticos siendo un modelo de la relación entre un conjunto de características descriptivas y una característica objetivo basado en un conjunto de ejemplos históricos o instancias comunes.
2. Posteriormente, es posible con este modelo establecer nuevas predicciones de manera constante (Shalev & Ben, 2014).

Por esta razón con el Machine Learning es posible establecer tres tipos de modelos (ver tabla 3) y tres tipos de aprendizajes (ver tabla 4), ya que se busca el aprendizaje continuo, conocer el progreso y la veracidad de la información que se desarrolla, porque:

Su aplicación está determinada en diferentes ámbitos dependiendo de las necesidades del problema, permitiendo a los científicos y especialistas tomar decisiones referentes a temas tan puntales como si un tipo de material puede ser separado por características de resistencia a la fragmentación o capacidad de absorción de energía pudiendo determinar con esto si es recomendable su uso en una u otra tarea específica (Salamanca y Castro, 2021, p.42).

Tabla 3 Modelos para el desarrollo de Machine Learning

|  |  |
| --- | --- |
| **Modelo** | **Características** |
| Geométrico | El espacio de instancias es el conjunto de todas las instancias posibles o descriptibles, estén presentes en nuestro conjunto de datos o no. Por lo general, este conjunto tiene alguna estructura geométrica. Por ejemplo, si todas las características son numéricas, entonces se puede usar cada característica como una coordenada en un sistema de coordenadas cartesiano. |
| Probabilístico | Muchos de estos modelos se basan en la siguiente idea. Sea X las variables que se conoce por ejemplo, los valores de las características de nuestra instancia; y deje que, Y denote las variables de destino que nos interesan, por ejemplo, la clase de la instancia. La pregunta clave en el aprendizaje automático es cómo modelar la relación entre X e Y. |
| Lógico | Se inspira en la informática y la ingeniería. A este tipo lo llamo “lógico” porque los modelos de este tipo pueden traducirse fácilmente en reglas comprensibles para los humanos, como si   * if Viagra=1thenClass=Y=spam· |

Fuente: Salamanca y Castro (2021)

*Tabla 2 Los tipos de aprendizaje en el ML*

|  |  |
| --- | --- |
| **Tipos** | **Fundamento** |
| Aprendizaje supervisado | Consiste en encontrar el mejor clasificador posible f: X→ Y para un problema determinado. El algoritmo responsable de encontrar este mapeo se denomina algoritmo de clasificación, que infiere un modelo de cada ejemplo de entrada x є X y su respectiva clase y є Y. Este modelo es una aproximación para la distribución de probabilidad conjunta (DPC) de las variables X e Y. |
| Aprendizaje no supervisado | Son aquellos en los que no estamos interesados en ajustar pares (entrada, salida), sino en aumentar el conocimiento estructural de los datos disponibles (y posibles datos futuros que provengan del mismo fenómeno), por ejemplo, dando una agrupación de los datos según su similaridad (clustering), simplificando las estructura de los mismos manteniendo sus características fundamentales (como en los procesos de reducción de la dimensionalidad), o extrayendo la estructura interna con la que se distribuyen los datos en su espacio original (aprendizaje topológico). |
| Aprendizaje semi supervisado | El objetivo es la clasificación, pero la entrada contiene datos etiquetados y sin etiquetar. El aprendizaje semi-supervisado, debido a su estructura, se puede dividir principalmente en dos tipos, según el objetivo del análisis que se busque hacer a la base de datos:  1. *Aprendizaje transductivo*, que separa el conjunto dado en conjunto de entrenamiento (donde la variable objetivo es conocida), y conjunto test donde ésta no es dada.  2. *Aprendizaje inductivo* trata de obtener una función  de predicción usando todo el conjunto {(X, Y )}, es decir, usando tanto los datos donde la variable objetivo es conocida como aquellos en los que no |

Fuente: Salamanca y Castro (2021)

Autores como Contreras et., al. (2022), Salamanca y Castro (2021) y Jara, Giral y Martínez (2016) explican que, en el campo del Machine Learning los métodos de predicción presentan un papel fundamental al permitir que las máquinas realicen predicciones precisas sobre datos no vistos. Existen varios métodos utilizados en el Machine Learning para generar predicciones, entre ellos se destacan los siguientes:

1. Regresión: Es utilizado para predecir un valor numérico continuo basado en variables independientes. Algunos ejemplos comunes son la regresión lineal y la regresión logística, que se utilizan para predecir la relación entre variables (Salamanca y Castro, 2021).
2. Clasificación: Se utiliza para predecir la pertenencia a una clase o categoría específica. Los algoritmos de clasificación, como el árbol de decisiones, las máquinas de vectores de soporte (SVM) y los clasificadores de Naive Bayes, se emplean para clasificar datos en categorías predeterminadas (Salamanca y Castro, 2021).
3. Agrupamiento: Es utilizado para agrupar conjuntos de datos similares en clusters o grupos. Algunos métodos comunes de agrupamiento son el algoritmo k-means y el agrupamiento jerárquico, que ayudan a encontrar patrones y estructuras ocultas en los datos (Contreras et al., 2022).
4. Redes neuronales: Estas técnicas se basan en modelos inspirados en el funcionamiento del cerebro humano. Las redes neuronales profundas, como las redes neuronales convolucionales (CNN) y las redes neuronales recurrentes (RNN), han logrado grandes avances en áreas como la visión por computadora y el procesamiento del lenguaje natural (Salamanca y Castro, 2021).
5. Máquinas de soporte vectorial: Se utilizan para encontrar un hiperplano que separe los datos en diferentes categorías. Este método es útil tanto para clasificación como para regresión, y se basa en encontrar la mejor separación posible entre los datos (Jara, Giral y Martínez, 2016).

Estos métodos de predicción en el Machine Learning han demostrado ser eficientes en una amplia gama de aplicaciones, desde la detección de fraudes hasta el diagnóstico médico y la recomendación de productos. La elección del método adecuado depende del tipo de datos, el problema a resolver y los objetivos específicos del proyecto de Machine Learning.

### ***Cultivo de aguacate Hass***

Cultivo del aguacate HANS es una práctica cada vez más popular en la agricultura debido a sus características y beneficios, del aguacate HANS es una variedad de aguacate que se distingue por su adaptabilidad a diferentes condiciones climáticas y su alta productividad. Según estudios realizados por el Departamento Administrativo de Nacional de Estadística (DANE, 2016), esta variedad se ha destacado por su resistencia a enfermedades y plagas comunes en el cultivo del aguacate.

Una de las ventajas de cultivar el aguacate Hass es su capacidad para adaptarse a diferentes climas, según los autores Reyes, Castro, López et., al. (2022), esta variedad es apta para regiones con temperaturas que oscilan entre los 10 y 35 grados Celsius, además, puede tolerar tanto condiciones secas como húmedas, lo que la convierte en una opción viable para diferentes áreas geográficas.

Cabe destacar, que otra de las características del aguacate Hass es su alta productividad, esta variedad puede alcanzar rendimientos superiores a los 20 kilogramos de fruta por árbol al año, esto la convierte en una opción rentable para los agricultores, ya que pueden obtener una mayor producción por unidad de superficie cultivada (Reyes et., al, 2022).

Además de su adaptabilidad y productividad, la aguacate Hass presenta resistencia a enfermedades comunes en el cultivo del aguacate, lo revelado por los autores Agapito, et., al. (2022), esta variedad ha mostrado resistencia al hongo Phytophthora cinnamomi, causante de la pudrición de raíz, lo que reduce la necesidad de aplicar productos químicos para su control.

### ***Plagas Stenoma catenifer y heilipus lauri***

El cultivo del aguacate puede verse amenazado por diversas plagas que pueden afectar tanto las hojas, los frutos como las raíces de la planta, Según el Dane (2016) las plagas más comunes es el ácaro del aguacate llamado Oligonychus yothersi, este ácaro se alimenta de las hojas del aguacate, lo que provoca un daño en la apariencia de la planta y puede reducir su capacidad de fotosíntesis, para controlar esta plaga, se pueden emplear insecticidas específicos o realizar una poda adecuada de las hojas afectadas.

Otra plaga importante en el cultivo del aguacate es la polilla del aguacate (Stenoma catenifer), según Carabalí, Caicedo y Holguín (2022) las larvas de esta polilla se alimentan de los frutos del aguacate, perforando galerías y causando daños considerables en la calidad de la cosecha, para prevenir y controlar esta plaga, otra plaga del aguate es el barrenador del hueso del aguacate Heilipus spp, que puede causar daños graves al cultivo, las larvas de este insecto se alimentan del interior del hueso del aguacate, lo que afecta el desarrollo adecuado del fruto (Carabalí, Caicedo y Holguín , 2022).

# **METODOLOGÍA DE LA INVESTIGACIÓN**

La metodología exploratoria aplicada es un enfoque de investigación que se utiliza para explorar y comprender fenómenos o problemas que son poco conocidos o no han sido investigados previamente de manera exhaustiva (Zafra, 2006). Se emplea cuando existe una falta de información o teorías bien establecidas sobre el tema de estudio, como lo son la identificación y monitoreo de plagas en el cultivo de aguacata Hass a traves de metodologías MLOps.

De acuerdo con Zafra (2006), la metodología exploratoria aplicada es flexible y adaptable, permitiendo a los investigadores ajustar su enfoque a medida que se obtiene nueva información. Al principio del proceso, el investigador puede tener una comprensión limitada del fenómeno, pero a medida que avanza la investigación, se van generando nuevas preguntas e ideas que guían el proceso de recolección y análisis de datos.

La metodología exploratoria aplicada es un enfoque flexible y creativo que permite a los investigadores en arquitectura de software explorar y comprender fenómenos como el Big Data y las metodolgías MLOps que son relativamente nuevas para generar conocimiento y a establecer las bases para investigaciones futuras más rigurosas y proporciona una base sólida para futuras investigaciones y puede contribuir a generar explicaciones y enfoques más avanzados en relación con la producción agrícola.

Para lograrlo, se requerirá la capacitación y evolución del modelo de pronóstico y su respectivo encapsulamiento y despliegue. Para el resguardo de todos los datos indispensables para el correcto funcionamiento del sistema, será imprescindible también implementar una base de datos que se adapte a los requerimientos del proyecto.

Para el despliegue de procesos se llevará a cabo el aprovisionamiento de una base de datos relacionada con el cultivo de aguacate y las condiciones para la aparición o no de plagas como lo son Stenoma Catenifer y Heilipus Lauri. Se requerirá, además, la implementación de los pipelines necesarios para la automatización de las tareas de encapsulamiento y despliegue. Por último, se pondrán en funcionamiento medios para la monitorización del sistema.

Todo este proceso se realizará aplicando los principios fundamentales de MLOps, siguiendo las medidas de seguridad informática requeridas para salvaguardar los datos personales, y además, cumpliendo con el conjunto de mejores prácticas en el desarrollo de software en general.

# **RECURSOS A EMPLEAR**

A continuación, se resumen los recursos necesarios para llevar a cabo la propuesta del presente proyecto.

* + 1. **Humanos**
    2. **Director**

Ph. David Arango

* + 1. **Físicos**

Sala de sistemas Pontificia Universidad Javeriana Cali.

Software libre y especializado para el análisis de datos.

Computador

* + 1. **Bibliográficos**

Bases de datos, tales como:

* Science Direct
* Springer
* Proquest
* ASTM
* Dialnet
* Scielo
* Redalyc

# **REFERENCIAS BIBLIOGRÁFICAS**

Agapito, A., Cibrián, L., Gutiérrez, R., Ruiz, J., López ,C y Rueda, P. (2022). Phytophthora cinnamomi Rands en aguacate. Revista Mexicana de Ciencias Agricolas, 1(28), 331-341. <https://dialnet.unirioja.es/descarga/articulo/8624636.pdf>

Aimacaña, C. y Columba ,G. (2021). Análisis comparativo de algoritmos de Machine Learning para la detección de plagas en los cultivos representativos de la sierra ecuatoriana. (Tesis de pregrado). Universidad Central del Ecuador. Ecuador <http://www.dspace.uce.edu.ec/bitstream/25000/24228/1/UCE-FING-ISI-AIMACA%c3%91A%20JEFFERSON-COLUMBA%20ALEXANDER.pdf>

Arley, O. y Llano, R. (2016). Sistemas de información enfocados en tecnologías de agricultura de precisión y aplicables a la caña de azúcar, una revisión. Revista Ingenierías Universidad de Medellín, 15(28), 103-124.

Ávila, P. (2022). Entrenamiento de redes neuronales para la identificación de plagas en cultivos de café. ( Tesis de Grado). Universidad Pedagogica. Colombia.

<http://repository.pedagogica.edu.co/bitstream/handle/20.500.12209/18306/entrenamiento%20redes%20neuronales.pdf?sequence=4&isAl>

Carabalí, M., Caicedo, V. y Holguín, M. (2022). Guía para el reconocimiento y manejo de las principales plagas de aguacate CV. Hass en Colombia. Instituto Colombiana Agropecuario. Colombia. <https://editorial.agrosavia.co/index.php/publicaciones/catalog/download/226/208/1422-1?inline=1>

Castañeda, V., Guerrero, M., Renteros, P y Villanueva, M. (2021). Detección de nutrientes del suelo y planta, y pestes en campos de cultivo de banano orgánico con Machine Learning(Tesis de Grado).Universidad Piura. Peru. <https://pirhua.udep.edu.pe/bitstream/handle/11042/5204/T_IME_2104.pdf?sequence=1&isAllowed=y>

Contreras, C., Medina, D., Acevedo, J. y Guevara, Ibarra. (2022). Metodología de desarrollo de técnicas de agrupamiento de datos usando aprendizaje automático. Revista Tecnura, 26(72), 42-58. <http://www.scielo.org.co/pdf/tecn/v26n72/0123-921X-tecn-26-72-42.pdf>

Cruz, L., Caamal, C., Pat., F. y Reza, S. (2022). Competitividad de las exportaciones de aguacate Hass de México en el mercado mundial. *Revista Mexicana de Ciencias Agrícolas*, 13(2), 355-362. <http://cienciasagricolas.inifap.gob.mx/index.php/agricolas/article/view/2885/4723>

Departamento Administrativo Nacional de Estadística. (DANE, 2016). Cultivo del aguacate Hass (Persea americana Mill; Persea nubigena var. Guatemalensis x Persea americana var. drymifolia), plagas y enfermedades durante la temporada de lluvias. *Boletín Mensual Insumos y Factores asociados a la Producción Agropecuaria*, (50), 1-102. <https://www.dane.gov.co/files/investigaciones/agropecuario/sipsa/Bol_Insumos_ago_2016.pdf>

García, F. y Ulloa, L. (2022). Implementación de modelo Machine Learning aplicado al estudio de enfermedades de café en el centro de investigación Sacha Wiwa, perteneciente a la parroquia Gasaganga, Cantón la Maná, provincia de Cotopaxi. (Tesis). Ecuador. Universidad Técnica de Cotopaxi. <http://repositorio.utc.edu.ec/bitstream/27000/8981/1/UTC-PIM-000530.pdf>

Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems (2nd ed.). O'Reilly Media, Inc.

Instituto Colombiano Agropecuario. (2012). Manejo Fitosanitario del cultivo del aguacate Hass. Mediadas para la temporada invernal. Colombia. Ministerio de Agricultura y Desarrollo Rural.

Jara, E., Giral, D. y Martínez, S. (2016). Implementación de algoritmos basados en máquinas de soporte vectorial (SVM) para sistemas eléctricos: revisión de tema. Revista Tecnura, 20(48), 149-170. <https://www.redalyc.org/pdf/2570/257046835012.pdf>

Monsalve, S. (2021). Agricultura de precisión en la predicción de la merma de defectos en cultivos de banano. (Tesis Especialización). Antioquia. Universidad de Antioquia. <https://bibliotecadigital.udea.edu.co/bitstream/10495/20593/10/SebastianMonsalve_2021_AgriculturaPrecisi%C3%B3n.pdf>

Murphy, K. (2012). Machine Learning A Probabilistic Perspective. 3rd ed. New York. The Massachusetts Institute of Technolog-MIT Press. <http://noiselab.ucsd.edu/ECE228/Murphy_Machine_Learning.pdf>

Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAOSTAT, 2021). Hacia una agricultura sostenible y resiliente en América Latina y el Caribe - Análisis de siete trayectorias de transformación exitosas. Chile. Organización de las Naciones Unidas para la Alimentación y la Agricultura –FAO. <https://www.fao.org/3/cb4415es/cb4415es.pdf>

Organización de las Naciones Unidas para la Alimentación y la Agricultura. (FAO, 2021). Perspectivas de la Agricultura y del Desarrollo Rural en las Américas: una mirada hacia América Latina y el Caribe 2021-2022. Costa Rica. CEPAL, FAO, IICA. <http://repositorio.cepal.org/bitstream/handle/11362/47208/CEPAL-FAO21-22_es.pdf?sequence=1&isAllowed=y>

Proyecto Colombia Mide. (2021). informe ejecutivo: Estudio sobre las necesidades y brechas de la calidad en la cadena productiva de aguacate Hass y plan de acción. Antioquía y su zona de influencia. Bogotá. Instituto Nacional de Metrología INM, Programa de la Unión Europea.

<https://colombiamide.inm.gov.co/wp-content/uploads/2021/05/Informe-Aguacate-Hass_VFinal_20210506.pdf>

Reyes, A., Castro A., López, F., Mendoza, B y Secundino J. (2022). Cambio climático y producción de aguacate. Revista de la Universidad de Chapingo. <https://www.researchgate.net/publication/357662730_Cambio_climatico_y_produccion_de_aguacate>

Rivero, D. (2022). Investigación de mercado y aplicación práctica de MLOps: machine learning operations. (Tesis Maestría). Uruguay. Universidad ORT Uruguay. <https://dspace.ort.edu.uy/bitstream/handle/20.500.11968/4836/Material%20completo.pdf?sequence=1&isAllowed=y>

Salamanca, R. y Castro, E. (2021), Técnicas de aprendizaje automático aplicadas en los sistemas de predicción. Revista Tecnología Investigación y Academia TIA, 8(1), 39-53.

Shalev, S. & Ben, D. (2014). Understanding Machine Learning: From Theory to Algorithms. Cambridge. Cambridge University Press. <https://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/understanding-machine-learning-theory-algorithms.pdf>

Tovar, M. (2019). Diseño e implementación de un aplicativo móvil para realizar la detección temprana de la enfermedad de la Sigatoka Negra en los cultivos de plátanos. (Tesis de Maestría). Colombia. Universidad Tecnológica de Pereira.

<https://repositorio.utp.edu.co/server/api/core/bitstreams/5de73e8c-547d-4d43-bcf9-bff99d8ea88c/content>

Vela, C. (2012). La industria del software: Una experiencia de empresas, gobiernos y universidades en Uruguay y Ecuador. Ecuador. FLACSO. <https://biblio.flacsoandes.edu.ec/libros/digital/52735.pdf>

Valenzuela, C. (2022). Detección y Clasificación de Enfermedades en el Tomate Mediante Deep Learning y Computer Vision. (Tesis de Maestria). Argentina. Universidad Nacional de la Plata. <http://sedici.unlp.edu.ar/bitstream/handle/10915/139770/Documento_completo.pdf-PDFA.pdf?sequence=1&isAllowed=y>

Visengeriyeva, L. Kammer, A., Bär, I., Kniesz, A & Michael Plöd, M. (2020). Why you Might Want to use Machine Learning. <https://ml-ops.org/content/motivation>

Zafra, G. (2006). Tipos de Investigación. Revista Científica General José María Córdova, 4(4), 13-14. <https://www.redalyc.org/pdf/4762/476259067004.pdf>

# **CRONOGRAMA DE TRABAJO**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ACTIVIDADES** | **SEMESTRE 1** | | | | | | | | | | | | | | | | | | | **SEMESTRE 2** | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Mar | | | | Abr | | | | | May | | | | | Jun | | | | | Ago | | | | | | | Sep | | | | | | Oct | | | | | Nov | | | | | Dic | | | | |
| 1 | 2 | 3 | 4 | | 1 | 2 | 3 | 4 | | 1 | 2 | 3 | 4 | | 1 | 2 | 3 | 4 | | 1 | 2 | | 3 | | 4 | | | **1** | **2** | **3** | **4** | | **1** | **2** | **3** | **4** | | **1** | **2** | **3** | **4** | | **1** | **2** | **3** | **4** |
| Problema |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  | |  | | |  | |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Objetivos |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  | |  | | |  | |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Metodología |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  | |  | | |  | |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Proyecto |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  | |  | | |  | |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Desarrollar una metodología MLOps |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  | |  | | |  | |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Validar el uso de MLOps |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  | |  | | |  | |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Desarrollar y entrenar un modelo |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  | |  | | |  | |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |
| Implementar técnicas de procesamiento |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  | |  | | |  | |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |  |  |

# **GLOSARIO DE TERMINOS ESPECIALES**

**Big Data**: Se refiere a los datos que son tan grandes, rápidos o complejos que es difícil o imposible procesarlos con los métodos tradicionales.

**Machine Learning**: Sub-área de la inteligencia artificial que consiste en desarrollar algoritmos con la capacidad de aprender y mejorar a través de experiencias, sin necesidad de programarlas explícitamente.

**MLOp**s: es una extensión de la metodología DevOps que busca incluir los procesos de aprendizaje automático y ciencia de datos en la cadena de desarrollo y operaciones para hacer que el desarrollo del ML sea más confiable y productivo.

**Feature Store**: validación y preparación de los datos, donde una capa de gestión de datos para el aprendizaje automático que permite compartir y descubrir funciones y crear canalizaciones de aprendizaje automático más eficaces.

**Código de fuente**: colección de líneas de texto, escritas en un lenguaje de programación, que guían el proceso de ejecución de un programa. Estas instrucciones, que son comprensibles por humanos, están redactadas por un programador.

**Model Registry**: Modelo requerido para los atributos se les asigna un clave para distinguirlos de los demás registros. Relación (el vínculo que define la dependencia entre varias entidades)

**Pipeline**: Tuberías virtuales se crean para segmentar los datos y, de este modo, incrementar el rendimiento de un sistema digital.

**Plagas**: Cualquier ser vivo que resulta perjudicial para otro ser vivo de interés para el ser humano.