****

**“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: \_\_\_\_\_\_\_\_palabras clave\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

10. FECHA DE INICIO: Día de Mes de Año DURACIÓN ESTIMADA: ## 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 3](#_Toc136436819)

[2.1 Planteamiento del problema 4](#_Toc136436820)

[2.2 Formulación del problema 6](#_Toc136436821)

[2.3 Alcance del trabajo de grado 6](#_Toc136436822)

[3. OBJETIVOS DEL PROYECTO 7](#_Toc136436823)

[3.1 Objetivo general 7](#_Toc136436824)

[3.2 Objetivos específicos 7](#_Toc136436825)

[3.3 Resultados esperados 8](#_Toc136436826)

[4. Justificación 8](#_Toc136436827)

# **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) 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.

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 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, 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, esta metodología garantiza 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.

## **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.

**REFERENCIAS BIBLIOGRÁFICAS**

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.

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>

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.

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>

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>

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>

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.

ttps://colombiamide.inm.gov.co/wp-content/uploads/2021/05/Informe-Aguacate-Hass\_VFinal\_20210506.pdf