



**UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO**  
**FACULTAD DE CIENCIAS DE LA INGENIERÍA**  
**CARRERA DE TELEMÁTICA**

Trabajo de Integración Curricular  
previa la obtención del Grado  
Académico de Ingeniera en  
Telemática.

**Proyecto de Investigación:**

**“ESTIMACIÓN DEL PESO DE RACIMOS DE BANANO POR MEDIO DE  
ALGORITMOS DE APRENDIZAJE AUTOMÁTICO”**

**Autor:**

JAZMÍN JAMILETH MORA SÁNCHEZ

**Director de Proyecto de Investigación:**

ING. NÉSTOR RAFAEL SALINAS BUESTAN, M. Sc

**Quevedo – Los Ríos – Ecuador.**

**2024**





## DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

Yo, **JAZMÍN JAMILETH MORA SÁNCHEZ**, declaro que la investigación aquí descrita es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

La Universidad Técnica Estatal de Quevedo, puede hacer uso de los derechos correspondientes a este documento, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

---

**JAZMÍN JAMILETH MORA SÁNCHEZ**

**C.I: 0929418275**



## CERTIFICACIÓN DE CULMINACIÓN DEL PROYECTO DE INVESTIGACIÓN

El suscrito, **Ing. Néstor Rafael Salinas Buestan M. Sc**, Docente de la Universidad Técnica Estatal de Quevedo, certifica que la estudiante **Jazmín Jamileth Mora Sánchez**, realizó el Proyecto de Investigación de grado titulado **“ESTIMACIÓN DEL PESO DE RACIMOS DE BANANO POR MEDIO DE ALGORITMOS DE APRENDIZAJE AUTOMÁTICO”**, previo a la obtención del título de **Ingeniero en Telemática**, bajo mi dirección, habiendo cumplido con las disposiciones reglamentarias establecidas para el efecto.

---

**Ing. Néstor Rafael Salinas Buestan. M. Sc**

**DIRECTOR DEL PROYECTO DE INVESTIGACIÓN**



## CERTIFICADO DEL REPORTE DE LA HERRAMIENTA DE PREVENCIÓN DE COINCIDENCIA Y/O PLAGIO ACADÉMICO

El suscrito, **Ing. Néstor Rafael Salinas Buestán. M. Sc**, mediante el presente cumpto en presentar a usted, el informe de proyecto de Investigación titulado “ESTIMACIÓN DEL PESO DE RACIMOS DE BANANO POR MEDIO DE ALGORITMOS DE APRENDIZAJE AUTOMÁTICO” Presentado por la estudiante **Jazmín Jamileth Mora Sánchez**, egresada de la Carrera de Telemática, que fue revisado bajo mi dirección según resolución del Consejo Directivo de la Facultad de Ciencias de la Ingeniería, que se ha desarrollado de acuerdo al Reglamento de la Unidad de Integración Curricular de la Universidad Técnica Estatal de Quevedo y cumple con el requerimiento de análisis de COMPILATIO el cual avala los niveles de originalidad en un 96% y similitud 4%, del trabajo investigativo. Valido este documento para que el estudiante siga con los trámites pertinentes, de acuerdo como lo estable el Reglamento.

 CERTIFICADO DE ANÁLISIS  
magister

**ESTIMACIÓN DEL PESO DE RACIMOS DE BANANO POR MEDIO DE ALGORITMOS DE APRENDIZAJE AUTOMÁTICO**

**4%**  
Textos sospechosos

**3%** Similitudes  
1% similitudes entre comillas  
< 1% entre las fuentes mencionadas  
**1%** Idiomas no reconocidos

Nombre del documento: Modelo predictivo para estimar el peso del racimo de banano.pdf  
ID del documento: 11ba080ead608d667fb365c94a21261413775139  
Tamaño del documento original: 1,56 MB  
Autor: Jazmin Mora

Depositante: Jazmin Mora  
Fecha de depósito: 10/5/2024  
Tipo de carga: url\_submission  
fecha de fin de análisis: 10/5/2024

Número de palabras: 10.470  
Número de caracteres: 70.237

Ing. Néstor Rafael Salinas Buestan. M. Sc  
DIRECTOR DEL PROYECTO DE INVESTIGACIÓN



**UNIVERSIDAD TÉCNICA ESTATAL DE QUEVEDO  
FACULTAD DE CIENCIAS DE LA INGENIERÍA  
CARRERA DE TELEMÁTICA**

**PROYECTO DE INVESTIGACIÓN**

**Título:**

**“ESTIMACIÓN DEL PESO DE RACIMOS DE BANANO POR MEDIO DE  
ALGORITMOS DE APRENDIZAJE AUTOMÁTICO”**

Presentado al Consejo Directivo de Facultad como requisito previo a la obtención del título de Ingeniero en Telemática.

Aprobado por:

**PRESIDENTE DEL TRIBUNAL**

Ing. Byron Oviedo Bayas. PhD

**MIEMBRO DEL TRIBUNAL**

Ing. Eduardo Samaniego Mena. M. Sc

**MIEMBRO DEL TRIBUNAL**

Ing. Ángel Torres Quijije. M. Sc

QUEVEDO – LOS RÍOS – ECUADOR  
2024

## AGRADECIMIENTO

Agradezco a Dios por darme la fortaleza, sabiduría y guía necesarias para alcanzar este significativo logro académico.

A mis padres, hermanos (Robinson y Jandry) y amigos, por su amor incondicional, apoyo constante y sacrificios que han realizado para permitirme alcanzar mis metas educativas. Su aliento y orientación han sido pilares fundamentales en mi trayecto hacia la culminación de esta tesis.

A la Universidad Técnica Estatal de Quevedo, mi alma máter, le expreso mi gratitud por brindarme la oportunidad de adquirir conocimientos, desarrollar habilidades y crecer tanto académica como personalmente durante mi tiempo en esta institución.

No puedo dejar de reconocer y agradecer el invaluable respaldo de mi director de tesis, el Ing. Néstor Rafael Salinas Buestan, M.Sc. La orientación experta y constante apoyo fueron esenciales en cada etapa de este proceso de investigación. Su conocimiento y paciencia fueron elementos clave para el éxito de este proyecto.

También quiero expresar mi gratitud a mi asesor el Ing. Juan Carlos Elizalde. Msc y a mi tribunal el Ing. Byron Oviedo. PhD, Ing. Eduardo Samaniego, M.Sc y el Ing. Ángel Torres, M.Sc, por su valiosa retroalimentación y su dedicación a mi trabajo. Su conocimiento y experiencia han enriquecido mi investigación y me han ayudado a desarrollar una comprensión más profunda de mi tema.

*“En este momento de culminación, concluyo que el viaje académico que me ha llevado hasta aquí y reconozco que cada experiencia, cada desafío y cada persona que ha cruzado mi camino ha contribuido a mi crecimiento y aprendizaje. Agradezco a todos los que han formado parte de este camino y me han brindado su apoyo incondicional. Que este logro sea un recordatorio que, con determinación, perseverancia, el apoyo de seres queridos y mentores, podemos alcanzar nuestras metas”.*

***Jazmín Jamileth Mora Sánchez***

## DEDICATORIA

Queridos padres. Con cada paso que doy en esta travesía llamada vida, me doy cuenta de lo afortunada que soy de tenerlos como guías, apoyo incondicional y fuente inagotable de amor. Su dedicación y sacrificio son la brújula que me orienta en los momentos difíciles y la inspiración que me impulsa hacia mis metas. Gracias por estar siempre ahí, por creer en mí incluso cuando yo dudaba de mis propias capacidades.

A mis queridos hermanos. Su presencia en mi vida es un regalo que valoro más de lo que las palabras pueden expresar. Juntos hemos compartido risas, lágrimas, alegrías y desafíos. Su cercanía y apoyo incondicional son un recordatorio constante de que nunca estoy sola.

*“En la travesía de la vida, nuestros padres y hermanos son como estrellas que iluminan nuestro camino en las noches oscuras y nos acompañan en los días soleados. Su amor y apoyo son los pilares que sostienen nuestra existencia, recordándonos que, aunque el viaje pueda ser difícil en ocasiones, nunca estamos solos mientras tengamos a quienes amamos cerca”.*

Con mucho amor y cariño.

***Jazmin Jamileth Mora Sánchez.***

## **RESUMEN EJECUTIVO Y PALABRAS CLAVES**

Esta investigación desarrolla un modelo predictivo mediante técnicas de aprendizaje automático para estimar el peso del racimo de banano, dada la importancia como el alimento clave y liderazgo mundial de Ecuador en exportaciones. Conducida en Mata de Cacao, provincia de Los Ríos, empleando métodos de juicio, exploratorios y descriptivos para identificar variables significativas en el peso del racimo, como calibración superior, manos, semana y temperatura mínima. Se aplican cinco algoritmos, entre ellos Regresión de Gradiente Potenciado, Regresión de Ridge, Regresión de Vecinos Más Cercanos, Regresión Lasso y Gradient Boosting Machine con Implementación de Árboles Ligeros. Destacando la eficacia del algoritmo Gradient Boosting Machine con Implementación de Árboles Ligeros, este enfoque contribuye a mejorar la gestión agrícola y la planificación de cosechas.

### **Palabras Clave.**

Predicción, Banano, Algoritmos, Aprendizaje Automático, Agrícola.

## **ABSTRACT AND KEYWORDS**

This research develops a predictive model using machine learning techniques to estimate the weight of the banana bunch, given its importance as the key food and Ecuador's world leadership in exports. Conducted in Mata de Cacao, Los Ríos province, using exploratory, judgment and descriptive methods to identify significant variables in bunch weight, such as upper and lower calibration, hands, lot, age, and maximum and minimum temperature. Five algorithms are applied, including Enhanced Gradient Regression, Ridge Regression, Nearest Neighbor Regression, Lasso Regression and Gradient Boosting Machine with Light Tree Implementation. Highlighting the effectiveness of the Gradient Boosting Machine with Light Tree Implementation algorithm, this approach contributes to improved agricultural management and crop planning.

### **Keywords**

Prediction, Banana, Algorithms, Machine Learning, Agricultural.

## TABLA DE CONTENIDO

DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS... ¡Error! Marcador no definido.	
CERTIFICACIÓN DE CULMINACIÓN DEL PROYECTO DE INVESTIGACIÓN ..... ¡Error! Marcador no definido.	
CERTIFICADO DEL REPORTE DE LA HERRAMIENTA DE PREVENCIÓN DE COINCIDENCIA Y/O PLAGIO ACADÉMICO..... ¡Error! Marcador no definido.	
CERTIFICADO DE APROBACIÓN POR TRIBUNAL DE SUSTENTACIÓN . ¡Error! Marcador no definido.	
AGRADECIMIENTO .....	ii
DEDICATORIA .....	iii
RESUMEN EJECUTIVO Y PALABRAS CLAVES.....	iv
ABSTRACT AND KEYWORDS .....	v
TABLA DE CONTENIDO.....	vi
ÍNDICE DE FIGURA.....	ix
ÍNDICE DE ANEXO.....	x
ÍNDICE DE TABLA .....	xi
CÓDIGO DUBLÍN.....	xii
INTRODUCCIÓN.....	1
CAPÍTULO I.....	2
MARCO CONTEXTUAL DE LA INVESTIGACIÓN .....	2
1.1. Problema de la investigación.....	3
1.1.1. Planteamiento del problema.....	3
Pronóstico .....	3
Diagnóstico .....	4
1.1.2. Formulación del problema .....	4
1.1.3. Sistematización del problema.....	4
1.2. Objetivos.....	4

1.2.1.	<i>Objetivo general</i> .....	4
1.2.2.	<i>Objetivos específicos</i> .....	4
1.3.	Justificación .....	5
CAPÍTULO II .....		3
FUNDAMENTACIÓN TEÓRICA DE LA INVESTIGACIÓN .....		3
2.1.	Marco Referencial .....	7
2.2.	Marco Conceptual .....	9
2.2.1.	<i>El banano</i> .....	9
2.2.3.	<i>Python</i> .....	10
2.2.4.	<i>¿Qué es Flask?</i> .....	11
2.2.5.	<i>¿Qué es un modelo predictivo?</i> .....	11
2.2.6.	<i>¿Qué es Machine Learning (LM)?</i> .....	11
2.2.6.1.	<i>Tipos de LM</i> .....	11
2.2.7.	<i>Modelos de regresión</i> .....	13
2.2.8.	<i>Ley orgánica de protección de datos personales</i> .....	15
2.2.9.	<i>Ley orgánica de agrobiodiversidad, semillas y fomento de agricultura</i> .....	15
CAPÍTULO III.....		16
METODOLOGÍA DE LA INVESTIGACIÓN .....		16
3.1.	<i>Localización</i> .....	17
3.2.	Tipo de investigación .....	17
3.2.1.	<i>Investigación exploratoria</i> .....	17
3.2.2.	<i>Investigación descriptiva</i> .....	17
3.3.	Método de investigación.....	18
3.3.1.	<i>Método de juicio</i> .....	18
3.3.2.	<i>Método analítico</i> .....	18
3.3.3.	<i>Método inductivo</i> .....	18
3.4.	Fuente de recopilación de información .....	18

3.5.	Diseño de la investigación.....	19
3.6.	Recursos humanos y materiales.....	20
3.6.1.	<i>Recursos humanos</i> .....	20
3.6.2.	<i>Materiales</i> .....	20
CAPÍTULO IV.....		17
RESULTADOS Y DISCUSIÓN .....		17
4.1.	Resultados.....	23
4.1.1.	<i>Variables influyentes en la estimación del peso del racimo de banano mediante un análisis de correlación múltiple</i> .....	23
4.1.1.1.	<i>Procesamiento de datos.</i> .....	24
4.1.1.2.	<i>Exploración de datos.</i> .....	25
4.1.1.3.	<i>Matriz de correlación.</i> .....	26
4.1.2.	<i>Algoritmo adecuado y entrenamiento del modelo predictivo</i> .....	27
4.1.2.1.	<i>Algoritmos de aprendizaje supervisado para estimar el peso del racimo de banano.</i> .....	28
4.1.3.	<i>Desempeño del modelo de predicción entrenado a través de pruebas de rendimiento para validar la estimación calculada</i> .....	40
4.1.3.1.	<i>Desarrollo de la página web del modelo predictivo.</i> .....	41
4.2.	Discusión.....	43
CAPÍTULO V .....		23
CONCLUSIONES Y RECOMENDACIONES .....		23
5.1.	Conclusiones.....	45
5.2.	Recomendaciones .....	46
CAPÍTULO VI.....		47
BIBLIOGRAFÍA .....		47
6.1.	Bibliografía.....	48
CAPÍTULO VII .....		36
ANEXOS .....		36

## ÍNDICE DE FIGURA

Figura 1 Manos del racimo de banano .....	9
Figura 2 Calibración del banano.....	10
Figura 3 Ubicación de la zona mata de cacao.....	17
Figura 4 Fase del diseño de la investigación .....	19
Figura 5:Procesamiento de datos. ....	24
Figura 6:Matriz de correlación de las variables preliminar. ....	26
Figura 7:Diagrama de flujo de la realización del modelo predictivo. ....	28
Figura 8:Regresión Ridge, el peso en función de manos y calibración superior.....	29
Figura 9:Regresión Lasso, el peso en función de calibración superior y manos.....	31
Figura 10:KNN, peso en función de calibración superior y manos.....	33
Figura 11:Regresión de Gradiente Potenciado, en peso en función de semana y manos...	34
Figura 12:LightGBM, en peso en función de manos y calibración superior.....	36
Figura 13: Comparación de la precisión de los modelos entrenado. ....	38
Figura 14: Comparación del tiempo de entrenamiento de los modelos de ML.....	39
Figura 15:Diagrama de comprobación del desempeño del modelo de predicción entrenado. .....	40
Figura 16: DataSet de validación.....	41
Figura 17: Interfaz del ingreso de los valores de las variables predictoras. ....	42
Figura 18:Resultado de la predicción del peso del racimo de banano.....	42

## ÍNDICE DE ANEXO

Anexo 1 Interfaz de la aplicación web .....	52
Anexo 2 Codificación de la página web .....	53
Anexo 3 Índex de la página web.....	53
Anexo 4 Codificación del algoritmo entrenado.....	54

## ÍNDICE DE TABLA

Tabla 1 Hardware .....	20
Tabla 2 Software.....	20
Tabla 3 Variables del DataSet de producción de banano .....	23
Tabla 4 Variables preliminar seleccionadas del DataSet.....	24
Tabla 5 Descripción de las variables preliminar utilizadas para el análisis de correlación múltiple.....	25
Tabla 6 Variables identificadas en el análisis de correlación, para el modelo predictivo ...	27
Tabla 7 comparación de las métricas de los modelos.....	37

## CÓDIGO DUBLÍN

<b>Título:</b>	Estimación del peso de racimos de banano por medio de algoritmos de aprendizaje automático		
<b>Autor:</b>	Jazmín Jamileth Mora Sánchez		
<b>Palabras claves:</b>	Predicción Agrícola	Banano Algoritmos	Aprendizaje Automático
<b>Fecha de publicación:</b>	Mayo 2024		
<b>Editorial:</b>	Quevedo- UTEQ “La María”, 2023		
<b>Resumen:</b>	<p>Esta investigación desarrolla un modelo predictivo mediante técnicas de aprendizaje automático para estimar el peso del racimo de banano, dada la importancia como el alimento clave y liderazgo mundial de Ecuador en exportaciones. Conducida en Mata de Cacao, provincia de Los Ríos, empleando métodos de juicio, exploratorios y descriptivos para identificar variables significativas en el peso del racimo, como calibración superior, manos, semana y temperatura mínima. Se aplican cinco algoritmos, entre ellos Regresión de Gradiente Potenciado, Regresión de Ridge, Regresión de Vecinos Más Cercanos, Regresión Lasso y Gradient Boosting Machine con Implementación de Árboles Ligeros. Destacando la eficacia del algoritmo Gradient Boosting Machine con Implementación de Árboles Ligeros, este enfoque contribuye a mejorar la gestión agrícola y la planificación de cosechas.</p>		
<b>Keywords:</b>	<p>This research develops a predictive model using machine learning techniques to estimate the weight of the banana bunch, given its importance as the key food and Ecuador's world leadership in exports. Conducted in Mata de Cacao, Los Ríos province, using exploratory and descriptive methods to identify significant variables in bunch weight, such as upper and lower calibration, hands, lot, age, and maximum and minimum temperature. Five algorithms are applied, including Enhanced Gradient Regression, Ridge Regression, Nearest Neighbor Regression, Lasso Regression and Gradient Boosting Machine with Light Tree Implementation. Highlighting the effectiveness of the Gradient Boosting Machine with Light Tree Implementation algorithm, this approach contributes to improved agricultural management and crop planning.</p>		
<b>Descripción:</b>	74 hojas: dimensiones, 29 x 21 cm + CD-ROM 6162		
<b>URI:</b>			

## INTRODUCCIÓN

El banano (*Musa Paradisiaca*), el cuarto alimento más importante y siendo la fruta más consumida a nivel mundial [1]. El comercio internacional en muchos países está enfocándose en eliminar barreras comerciales, viéndolo como un motor clave para el desarrollo económico sostenido [2]. Ecuador es destacado como el principal exportador de banano en el mundo cubriendo un 29% el mercado internacional, actualmente abarca aproximadamente 200,000 hectáreas de terreno que se riegan y fumigan de manera continua para garantizar la cosecha semanal [1],[3].

Andrade y Fernando indican que entre las estrategias que se pueden implementar en la agricultura, consideran principalmente las tecnologías de procesos [4]. Entre esa tecnología destacan la de procesamiento de datos. Juan Carlos Elizalde experto en la agricultura indica que un modelo predictivo para estimar el peso del racimo de banano es una herramienta valiosa en la agricultura moderna, que contribuye en optimizar la gestión agrícola y mejorar la precisión en la planificación de la cosecha.

El trabajo de integración curricular llevó a cabo el desarrollo de un modelo predictivo con el objetivo de estimar el peso del racimo de banano. Este modelo fundamenta técnicas sofisticadas de aprendizaje automático. Para identificar el algoritmo óptimo que se alineó eficientemente con los datos, procediendo a realizar una selección y evaluación rigurosa de cinco algoritmos distintos: Regresión de Gradiente Potenciado, Regresión de Ridge, Regresión de Vecinos Más Cercanos, Regresión Lasso y Regresión gradiente potenciado con Implementación de Árboles Ligeros. Tras un análisis exhaustivo, determinando que LightGBM sobresalía por su precisión superior. Este documento proporciona una descripción detallada y metódica del proceso implementado para este desarrollo.

## **CAPÍTULO I**

### **MARCO CONTEXTUAL DE LA INVESTIGACIÓN**

## **1.1. Problema de la investigación**

### ***1.1.1. Planteamiento del problema***

Ecuador destacado como el principal exportador de banano en el mundo, cubre el 29% del mercado internacional. Actualmente, abarca aproximadamente 200,000 hectáreas de terreno que se riegan y fumigan de manera continua para garantizar la cosecha semanal del banano. [1],[3]. A pesar de su importancia, también se enfrentan a grandes desafíos, por ejemplo, las plantas de banano son vulnerables al viento y fácilmente derribadas por el peso propio del racimo, el manejo incorrecto resulta una pérdida significativa para los agricultores [5]. También en cultivos de banano la sigatoka negra<sup>1</sup>, la reducción del follaje<sup>2</sup> causa reducción entre el 20% y 40 % en el peso de la fruta [6].

Los agricultores enfrentan desafíos en determinar con precisión el peso del racimo del banano, debido a la falta de herramientas o conocimientos adecuados. Esta situación lleva a una subestimación o sobreestimación del peso del racimo del banano, afectando el precio de venta, planificación de cosecha y transporte. Este problema se ve exagerado por factores como el daño causado por plagas de insectos, afectando calidad y peso de los racimos, haciendo que no sea apto para exportación.

### **Pronóstico**

El Modelo Predictivo para Estimar el Peso del Racimo de Banano representa un avance significativo en la gestión agrícola. Su potencial reside en optimizar la gestión agrícola y mejorar la precisión en la planificación de la cosecha. Así mismo, se anticipa que este modelo introducirá innovación tecnológica en el sector bananero, elevando su competitividad a nivel mundial.

---

<sup>1</sup> La Sigatoka negra, ocasionada por el hongo *Mycosphaerella fijiensis* (también conocido como *Pseudocercospora fijiensis* en su fase anamórfica), es la principal dificultad fitopatológica que afecta al cultivo del banano, siendo una enfermedad foliar.

<sup>2</sup> El follaje del banano se refiere a las hojas de la planta de banano.

## **Diagnóstico**

La falta de herramientas precisas para estimar el peso de los racimos de banano genera incertidumbre en la gestión agrícola, limitando eficiencia y toma de decisiones informadas. La complejidad de varios factores constituye desafíos para optimizar el cultivo de banano.

### ***1.1.2. Formulación del problema***

¿Cómo estimar el peso del racimo de banano, integrando múltiples variables, con el propósito de optimizar la gestión agrícola y mejorar la precisión en la planificación de la cosecha?

### ***1.1.3. Sistematización del problema***

- ¿Cuáles son las variables que muestran una correlación significativa con el peso del racimo de banano?
- ¿Cuál es la metodología correcta para adoptar un modelo predictivo empleando las variables identificadas?
- ¿Cómo se comprueba la eficiencia del modelo de predicción seleccionado?

## **1.2. Objetivos**

### ***1.2.1. Objetivo general***

Desarrollar un modelo de predicción para la estimación del peso de racimos de banano a través de técnicas de aprendizaje automático.

### ***1.2.2. Objetivos específicos***

- Identificar las variables más influyentes para la estimación del peso del racimo de banano mediante un análisis de correlación múltiple.
- Aplicar el algoritmo más efectivo mediante técnicas avanzadas de aprendizaje automático para estimar el peso del racimo del banano.
- Comprobar el desempeño del modelo de predicción entrenado a través de pruebas de rendimiento para validar la estimación calculada.

### **1.3.Justificación**

Según el reporte de la Organización para la Alimentación y Agricultura de las Naciones Unidas (FAO), la aplicación de tecnologías innovadoras en agricultura tiene el potencial de incidir positivamente en reducción de hambruna y extrema pobreza. Las avanzadas tecnologías, como el aprendizaje automático con sus algoritmos predictivos, están ganando relevancia y abriendo camino hacia la implementación de una agricultura inteligente y tecnológicamente avanzada [7].

El modelo predictivo para estimar el peso el racimo del banano es de gran importancia en la agricultura, permite a los agricultores hacer elecciones fundamentadas sobre cómo y cuándo manejar y cosechar bananos, gracias a las predicciones del peso del racimo del banano. También ayuda a identificar que práctica agrícola puede aumentar en el peso de los racimos, mejorando la productividad general [8] .

El modelo predictivo desarrollado para estimar el peso del racimo del banano es una herramienta esencial para optimizar la gestión agrícola y mejorar la precisión en la planificación de la cosecha. Tiene, la capacidad de analizar grandes volúmenes de datos permitiendo a los agricultores tomar decisiones informadas sobre el manejo de sus cultivos.

El modelo desarrollado es fundamental, aporta beneficios significativos en varios aspectos. Al estimar con precisión el peso de los racimos de banano, facilita optimizar los recursos necesarios para la cosecha, incluyendo la mano de obra, materiales y transporte. Esta optimización conduce a una asignación de recursos más eficiente, lo que a su vez reduce los costos y aumenta la rentabilidad. Adicionalmente, el modelo permite una planificación del momento óptimo para la cosecha, garantizando así una mejora significativa en la calidad del producto final.

**CAPÍTULO II**  
**FUNDAMENTACIÓN TEÓRICA DE LA INVESTIGACIÓN**

## 2.1.Marco Referencial

Este artículo realizado por Soares, JDR, Pasqual, M., Lacerda, WS, Silva, SO y Donato, SLR. "Utilización de redes neuronales artificiales en la predicción del peso de los racimos en plantas de banano" propone una metodología para predecir el peso de los racimos de plátano utilizando redes neuronales artificiales. El estudio evalúa el impacto de diferentes características agronómicas en el peso de los racimos, y demuestra que las redes neuronales artificiales tienen potencial de predecir con precisión el peso de los racimos de banano. Este documento aporta una herramienta prometedora para la estimación del rendimiento de los cultivos de plátano a través del uso de redes neuronales artificiales [9].

El artículo aporta una metodología basada en redes neuronales artificiales para predecir el peso de los racimos de banano, sirve de gran utilidad para el estudio propuesto, brindando una herramienta novedosa y precisa para estimar el rendimiento de los cultivos de banano.

Este artículo realizado por Jayasinghe, SL, Ranawana, C., Liyanage, I. y Kaliyadasa, P. (2022). Trata de la revisión sistemática que aborda la estimación del crecimiento y rendimiento del plátano a través de modelos matemáticos. Mencionan diversos enfoques, incluyendo modelos no lineales que tienen en cuenta las complicadas relaciones no lineales entre las variables predictoras y el rendimiento del plátano. También discuten modelos de regresión, análisis de series temporales, modelos híbridos lineales, no lineales y técnicas como regresión bayesiana, entre otros [10].

El documento aporta una revisión sistemática de diversos enfoques y modelos matemáticos utilizados para estimar el crecimiento y rendimiento del plátano. Teniendo relación con el tema propuesto desarrollado.

Este artículo realizado por Ogunsua, J., Saengrayap, R., Ullah, H., & Chaiwong, S. (2019). Analiza la predicción del peso de las bananas Cavendish utilizando un modelo de red neuronal con arquitectura (10-10-1), estudia la relación entre la calidad de la fruta, temperatura en las temporadas de verano y lluvia. También se encontró que el modelo de red neuronal seleccionado tiene un MPE (Mean Predictive Error) bajo y un R2 (coeficiente de determinación) alto, indica una mejor precisión en la predicción en las bananas Cavendish [11].

Este artículo proporciona información específica sobre cómo se puede aplicar este modelo para estimar el peso del banano. También analiza la relación entre la calidad de la fruta y temperatura en temporadas de verano y lluvia.

En esta investigación realizada por Muñoz Torres, Pedro Santiago (2022). Aborda el uso de algoritmos de aprendizaje automático y análisis de datos en el sector bananero. El documento plantea el problema de predicción del peso de los racimos de plátano y propone el uso de modelos predictivos basados en datos recopilados a través de sensores en las plantaciones de plátano. Se describe la metodología utilizada para recopilar y procesar los datos, así como los algoritmos de aprendizaje automático utilizados para entrenar y validar los modelos. Además, se presentan recomendaciones para futuras investigaciones, como la incorporación de un mayor conjunto de variables y el uso de sensores para mejorar la precisión de los modelos predictivos [8].

Este documento presenta una correlación bien directa con el estudio realizado, utilizan algoritmos de aprendizaje automático y análisis de datos del sector bananero.

## **2.2.Marco Conceptual**

### ***2.2.1. El banano***

El banano, cultivado en regiones tropicales globalmente y en muchos países en desarrollo, desempeña un papel crucial en el crecimiento económico y desarrollo social local. A nivel mundial, es un alimento esencial y se destaca como uno de los cultivos alimentarios más importantes, siendo también un producto destacado en las exportaciones [12].

### ***2.2.2. Variables.***

A continuación, se detallan las principales variables que influyen en el peso del racimo del banano en este estudio.

#### ***2.2.2.1.Temperatura.***

La temperatura del suelo es un factor crucial en la agricultura, ya que el éxito de la germinación y el desarrollo de la planta está directamente vinculado al calor que el suelo pueda proporcionar[13] . La magnitud de los cambios de temperatura varía de acuerdo con las características climáticas locales, la proximidad al mar y la topografía, entre otros factores [14].

#### ***2.2.2.2.Manos del racimo del banano.***

Las manos en el contexto del racimo de banano se refieren a las agrupaciones de frutos individuales (también llamados dedos) que cuelgan de un racimo. Cada racimo de banano está formado por varias filas de frutas, y cada fila se conoce como una mano. Cada mano está compuesta por múltiples plátanos o bananos unidos entre sí [15].

**Figura 1** *Manos del racimo de banano*



**Nota:** El gráfico representa las filas que está conformado el racimo de banano. Tomado de: <https://n9.cl/xxswb>

### 2.2.2.3. Calibración del banano.

En términos de procesos de cosecha y poscosecha, se toman en cuenta varios aspectos como la edad fisiológica del racimo, el tiempo adecuado para la cosecha, y el grado de crecimiento de la fruta. La calibración es una parte importante del proceso, definida como la medición del grado de llenado de la fruta. Esta se realiza a partir de la semana 11 del desarrollo de la fruta y ayuda a determinar el momento óptimo para la cosecha. Durante la cosecha, se evalúa el grado de la fruta en la última y segunda mano del racimo, y se rechaza la fruta que no cumpla con las exigencias del embarque [16].

**Figura 2** Calibración del banano



**Nota:** El gráfico representa la medida del grosor banano. Tomado de: <https://n9.cl/rxe9s>

### 2.2.3. Python

Python es un lenguaje de programación con propósito general que fue creado en los años 90 por Guido Van Rosum, quien ha trabajado tanto en Google como en la actualidad en Dropbox. Este lenguaje se distingue por su sintaxis limpia y legible. Además, presenta tipado dinámico, lo que significa que una variable puede contener datos de diversos tipos. Su naturaleza interpretada lo convierte en un lenguaje idóneo para aquellos que están dando sus primeros pasos en la programación. Al ser un lenguaje interpretado, no es necesario compilar el código fuente antes de ejecutarlo, lo que proporciona diversas ventajas [17].

#### ***2.2.4. ¿Qué es Flask?***

Flask, clasificado como un micro framework, facilita la creación de aplicaciones web con Python de manera sencilla. La designación "micro" no implica que sea un proyecto pequeño o destinado exclusivamente a páginas web de escala reducida. Al implementar Flask, se accede a las herramientas esenciales para desarrollar una aplicación web completamente funcional. En caso de requerir nuevas funcionalidades que no estén disponibles de inmediato con la instalación estándar, existe un extenso conjunto de extensiones (plugins) compatible con Flask. Estas extensiones pueden ser integradas de manera fácil para ampliar las capacidades de la aplicación según las necesidades específicas [18], [19].

#### ***2.2.5. ¿Qué es un modelo predictivo?***

Un modelo predictivo es un modelo estadístico o matemático que utiliza datos existentes para predecir valores futuros o para clasificar datos en categorías, se pueden aplicar en diversos campos, como la salud, la educación y la investigación científica [20],[21]. También utiliza técnicas de estadísticas, aprendizaje computacionales o minerías de datos para desarrollar los modelos que predicen eventos futuros o conductas, permitiendo aprovechar los patrones de comportamiento.

#### ***2.2.6. ¿Qué es Machine Learning (LM)?***

LM es una forma de programación que permite a los ordenadores aprender de los datos. En lugar de escribir código específico para cada tarea, los algoritmos de aprendizaje automático pueden aprender a realizar tareas a partir de un conjunto de datos [22].

LM es un campo multidisciplinar que combina métodos de inteligencia artificial, estadística y sistemas de bases de datos para descubrir patrones en grandes conjuntos de datos [23]. También permite a las computadoras resolver problemas por sí misma a medida que ML tiene acceso a más datos, sus resultados mejoran. Para realizar el análisis, utiliza algoritmos que generan nuevos datos según sea necesario [24].

##### ***2.2.6.1. Tipos de LM.***

Para realizar cualquier tarea específica, se debe tener conocimiento de conceptos importantes de LM, así como las diferentes opciones existentes, las medidas de calidad más utilizadas, entre otros.

- **Aprendizaje supervisado**

Tipo de aprendizaje automático en el que el modelo recibe datos etiquetados, es decir, datos que incluyen una entrada y una salida. El modelo aprende a asociar una entrada con una salida determinada. Este tipo de aprendizaje ofrece los mejores resultados porque el modelo tiene toda la información necesaria para aprender [25].

En el aprendizaje supervisado pueden utilizar técnicas por regresión y clasificación. La primera técnica se utiliza para predecir valores numéricos y la segunda consiste en asignar con precisión datos de prueba en categorías específicas.

- **Aprendizaje no supervisado**

El aprendizaje no supervisado se utiliza cuando el problema requiere una cantidad masiva de datos sin etiquetar. Este aprendizaje es aplicable en situaciones que necesitan procesar grandes volúmenes de datos sin etiquetar. La comprensión del significado detrás de estos datos se necesita de algoritmos que clasifican los datos con base en los patrones o clústeres que encuentra. También lleva a cabo un proceso repetido, analizando los datos sin mediación humana [26].

### **Clustering (agrupamiento)**

En situaciones de este tipo, la meta consiste en organizar elementos en conjuntos o clusters según sus similitudes. Por ejemplo, clasificar al cliente en función de sus costumbres de compras [27].

### **Reducción de dimensiones**

La meta consiste en disminuir la cantidad de variables o atributos en los datos, conservando la máxima cantidad de información pertinente. Este enfoque resulta beneficioso para la visualización de datos en dimensiones reducidas y para simplificar la complejidad de los modelos predictivos [27].

- **Aprendizaje semisupervisado**

El aprendizaje semisupervisado es una rama del aprendizaje automatizado que resulta de combinar el aprendizaje supervisado y el no supervisado [28].

- **Aprendizaje adaptativo**

Tipo de aprendizaje en el que se parte de un modelo previo cuyos parámetros se modifican o adaptan usando los nuevos datos de entrenamiento [22].

- **Aprendizaje por refuerzo**

El aprendizaje por refuerzo es un tipo de aprendizaje automático en el que el modelo no recibe datos etiquetados, es decir, datos que incluyen una entrada y una salida. El modelo aprende a tomar decisiones para maximizar una recompensa, sin saber cuál es la salida correcta [29].

### ***2.2.7. Modelos de regresión***

Los modelos de regresión lineal ejecutan predicciones numéricas acerca de un objeto, utilizando información proveniente de un conjunto de datos que está categorizado según una o más características [22]. Así como ocurre con la clasificación, la regresión también constituye una técnica predictiva, pero se distingue por tener como característica principal que la variable objetivo es de naturaleza continua. Sus usos abarcan diversos sectores y se centran en la realización de tareas como la identificación de correlaciones, la cuantificación de la causalidad y la detección de tendencias [30].

#### **Regresión de vecinos más cercanos (KNN)**

KNN representa un algoritmo versátil que se puede aplicar en diversos contextos de aprendizaje automático, para realizar tareas de regresión o clasificación. En el caso de la regresión, su funcionamiento es directo: Se generan predicciones tomando como referencia las muestras más cercanas a la instancia que intenta predecir [22].

#### **Regresión de contracción y selección de operadores menos absolutos (LASSO)**

El enfoque LASSO, conocido como Least Absolute Shrinkage and Selection Operator en inglés, fue presentado por (Tibshirani, 1996) como una opción al método Ridge con la meta de aumentar la precisión y facilitar la interpretación del modelo de regresión lineal [31].

## **Regresión de Ridge**

Esta técnica fue inicialmente presentada por Hoerl y Kennard como una estrategia para evitar las consecuencias negativas asociadas al problema de colinealidad en un modelo lineal estimado mediante el uso de mínimos cuadrados [32].

## **Regresión de gradiente potenciado (GBR)**

El GBR, también conocido como Gradient Tree Boosting o Gradient Boosted Regression Trees (GBRT), es especialmente efectivo en escenarios de datos tabulares y se destaca por su capacidad de combinar múltiples modelos predictivos débiles para crear un modelo más fuerte y preciso. Los modelos débiles suelen ser árboles de decisión poco profundos y cada nuevo árbol se crea para corregir los errores del anterior. Un aspecto clave de esta técnica es la tasa de aprendizaje, que controla cuánto mejora cada nuevo árbol respecto al anterior [33].

## **Gradient Boosting Machine con Implementación de Árboles Ligeros (LightGBM)**

LightGBM, en inglés significa Light Gradient Boosting Machine, es una técnica de aprendizaje automático que utiliza el algoritmo de Gradient Boosting para construir un modelo predictivo a partir de árboles de decisión más simples y menos complejos, resultando en un modelo eficiente en tiempo y con un buen rendimiento predictivo [34].

A continuación, se presenta el marco normativo que respalda y sitúa en contexto el estudio actual. Las leyes descritas a continuación son: Ley orgánica de protección de datos personales (con el art 7 y 8) y ley orgánica de agrobiodiversidad, semillas y fomento de agricultura (con el Art 13 y 14).

#### ***2.2.8. Ley orgánica de protección de datos personales***

**Art. 7.** Establece las condiciones para el tratamiento legítimo de datos personales, definiendo las circunstancias en las cuales dicho tratamiento es considerado legal y ético [35].

**Art. 8.** El tratamiento y comunicación de datos personales requiere el consentimiento del titular, el cual debe ser libre, específico, informado e inequívoco. El titular tiene el derecho de revocar el consentimiento en cualquier momento sin justificación, y el responsable debe establecer procedimientos eficientes para ello. El tratamiento previo a la revocación es válido, ya que el consentimiento no tiene efectos retroactivos [35].

#### ***2.2.9. Ley orgánica de agrobiodiversidad, semillas y fomento de agricultura***

**Art. 13.** Establece el derecho de las personas y colectividades al acceso seguro y permanente de alimentos sanos, suficientes y nutritivos; preferentemente producidos a nivel local y en correspondencia con sus diversas identidades y tradiciones culturales [36].

**Art. 14.** Establece el derecho a vivir en un medio ambiente sano y ecológicamente equilibrado, también declara de interés público la preservación del ambiente, la conservación de los ecosistemas, la biodiversidad y la integridad del patrimonio genético del país, así como la prevención del daño ambiental y la recuperación de los espacios naturales degradados [36].

## **CAPÍTULO III**

### **METODOLOGÍA DE LA INVESTIGACIÓN**

### 3.1. Localización

El proyecto de integración curricular se llevó a cabo en Mata de Cacao, ubicado en Babahoyo, provincia de Los Ríos. Las coordenadas geográficas precisas de la zona son 1.9865393141312058, -79.35248547008646, lugar donde se recolectó el conjunto de datos de los racimos de banano cortado.

**Figura 3** Ubicación de la zona mata de cacao



**Nota:** El gráfico representa el lugar donde se realizó el estudio. Tomado de: <https://n9.cl/qy0z2t>

### 3.2. Tipo de investigación

#### 3.2.1. Investigación exploratoria

La investigación exploratoria centrada primordialmente en la identificación preliminar de factores que inciden en el peso del racimo de banano. Dicho enfoque facilitó la realización de una revisión profunda de la literatura especializada en el ámbito agrícola, abarcando también aspectos relevantes del Aprendizaje Automático.

#### 3.2.2. Investigación descriptiva

Esta investigación aportó en un análisis detallado de correlación y relevancia entre las variables identificadas. Se emplearon técnicas estadísticas para cuantificar la fuerza y dirección de las relaciones entre estas variables y el peso del racimo de banano. Este análisis descriptivo permitió identificar las variables más influyentes en la estimación del peso de banano.

### **3.3.Método de investigación**

#### **3.3.1. Método de juicio**

Este método, también denominado juicio de expertos, es empleado para consultar a expertos agrónomos, cuyas opiniones y conocimientos se consideran esenciales a considerar en el modelo.

#### **3.3.2. Método analítico**

El método analítico se utilizó para examinar la correlación entre las variables identificadas, con el peso del racimo de banano. Este enfoque permitió comprender la estructura y naturaleza de las influencias que afectan la estimación del peso del racimo de banano mediante técnicas de análisis estadístico.

#### **3.3.3. Método inductivo**

Este método permitió observar los resultados de las diferentes métricas de los algoritmos escogidos. A partir de esas métricas, se verifica la eficacia de los algoritmos del modelado predictivo en la estimación del peso del racimo de banano.

### **3.4.Fuente de recopilación de información**

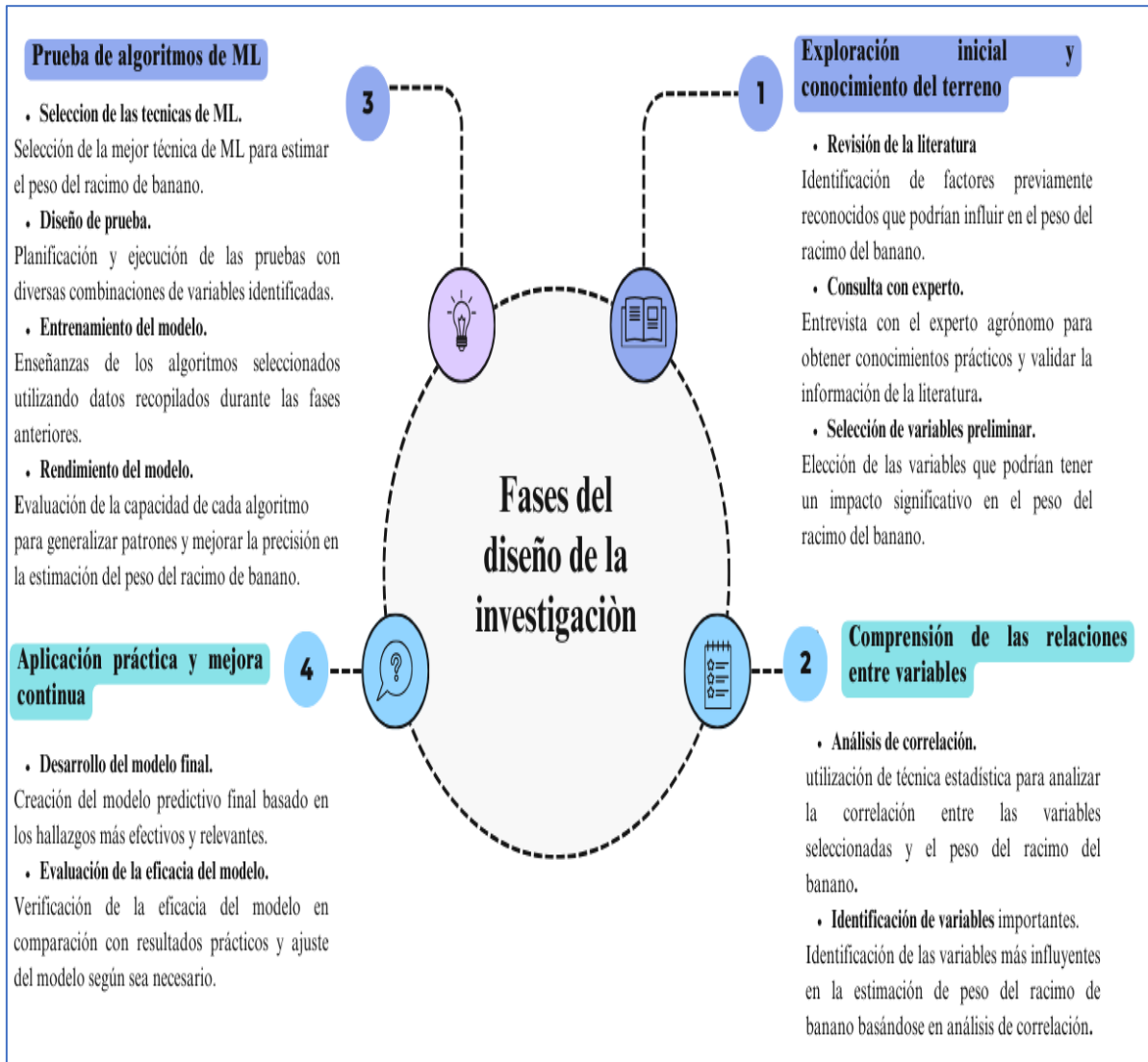
Se llevó a cabo una investigación centrada en temas de máxima relevancia, empleando una variedad de fuentes especializadas. Como artículos científicos, tesis, libros y sitios web de relevancia. Además, se destacan las bases de datos reconocidos por su rigurosidad académica, tales como:

- Google académico.
- Scopus.
- SpringerLink.
- SciELO.

### 3.5. Diseño de la investigación

A continuación, se describe el proceso que se llevó a cabo para realizar el modelo predictivo para estimar el peso del racimo del banano.

**Figura 4** Fase del diseño de la investigación



**Nota:** El gráfico representa el diseño que se llevó a cabo en el estudio. Elaborado por: Jazmín Mora

### 3.6. Recursos humanos y materiales

#### 3.6.1. Recursos humanos

**Autor:** Mora Sánchez Jazmín Jamileth.

**Director del proyecto:** Ing. Néstor Rafael Salinas Buestan. M.Sc.

**Experto en la agricultura:** Ing. Juan Carlos Elizalde. M.Sc.

#### 3.6.2. Materiales

**Tabla 1** Hardware

Cantidad	Equipo	Características	Costo
1	Laptop	Core i7-8565U 12GB RAM 1TB	\$ 1000

**Fuente:** Investigación

**Elaborado:** Jazmín Mora

Para desarrollar un modelo predictivo destinado a estimar el peso del racimo de banano, se considera una computadora Dell procesador Core i7-8565U, 12GB de RAM y 1TB de almacenamiento, con acceso a Internet de alta velocidad. La elección de este hardware justifica varios aspectos relacionadas con la tarea específica de construir un modelo predictivo para la estimación del peso del racimo del banano.

#### 3.6.3. Softwares

**Tabla 2** Software

Tipo	Descripción	Costo
MS Office	Herramientas esenciales para productividad y colaboración.	\$ 25.00
Visual Studio Code	Desarrollo de software eficiente y escalable.	Cd abierto
Tableau Prep	Preparación avanzada de datos para análisis preciso	Cd abierto
Python	Automatización y análisis de datos.	Cd abierto
Flask	Desarrollo ágil de aplicaciones web y APIs	Cd abierto

**Fuente:** Investigación

**Elaborado:** Jazmín Mora

La tabla 2 muestra todas las herramientas utilizadas para el desarrollo del modelo predictivo para estimar el peso del racimo de banano. MS Office (Excel), se utilizó para organizar y visualizar los datos. Visual Studio Code es un editor de código fuente que soporta múltiples lenguajes de programación, empleado para escribir y editar los scripts de Python.

Tableau Prep herramienta de preparación de datos, utilizada para limpiar el DataSet. Python lenguaje de programación muy utilizado en análisis de datos y ML, ofrece una amplia gama de bibliotecas. El Flask micro framework empleado para el desarrollo de la aplicación web en Python.

La suma total de inversión para el desarrollo del proyecto asciende a \$1025, incluyendo el equipo como el software necesario. Se destaca por su notable eficiencia económica y accesibilidad, gracias a la implementación estratégica de software de código abierto.

**CAPÍTULO IV**  
**RESULTADOS Y DISCUSIÓN**

## 4.1. Resultados

### 4.1.1. Variables influyentes en la estimación del peso del racimo de banano mediante un análisis de correlación múltiple

Para identificar las variables a utilizar en el modelo predictivo, se obtuvo un DataSet de producción de banano periodo 2022, abarcando un total de once variables con 146,662 registros. A continuación, la tabla 3 describe las variables.

**Tabla 3** Variables del DataSet de producción de banano

N°	Campo	Descripción de campo
1	Semana	Semana del año
2	Peso	Peso del racimo
3	Mano	Número de frutas en el racimo
4	Calibración superior	Medida superior del grosor del fruto
5	Calibración Inferior	Medida inferior del grosor del fruto
6	Año	Periodo lectivo
7	Fecha	Fecha del año
8	Edad	Edad para cortar el racimo
9	Lote	Ubicación de la hacienda del racimo cortado
10	Temperatura Mínima	Temperatura máxima del aire
11	Temperatura máxima	Temperatura mínima del aire

**Fuente:** Investigación

**Elaborado:** Jazmín Mora

La selección de variables preliminar se llevó a cabo mediante el método de juicio, basándose en el criterio del Ing. Juan Carlos Elizalde, experto en agricultura. Él indica, para estimar el peso del racimo de banano las variables independientes que podrían influir en el peso son: semana, mano, calibración superior e inferior, edad, lote, temperatura mínima y máxima. Con la finalidad de aplicar un análisis de correlación múltiple para determinar cuáles tienen mayor influencia en el peso del racimo de banano.

**Tabla 4** Variables preliminar seleccionadas del DataSet.

N°	Campo	Descripción de campo	Tipo
1	Semana	Semana del año	Independiente
2	<b>Peso</b>	<b>Peso del racimo</b>	<b>Dependiente</b>
3	Mano	Número de frutas en el racimo	Independiente
4	Calibración superior	Medida superior del grosor del fruto	Independiente
5	calibración Inferior	Medida inferior del grosor del fruto	Independiente
6	Edad	Edad para cortar el racimo	Independiente
7	Lote	Ubicación de la hacienda del racimo cortado	Independiente
8	Temperatura Mínima	Temperatura máxima del aire	Independiente
9	Temperatura máxima	Temperatura mínima del aire	Independiente

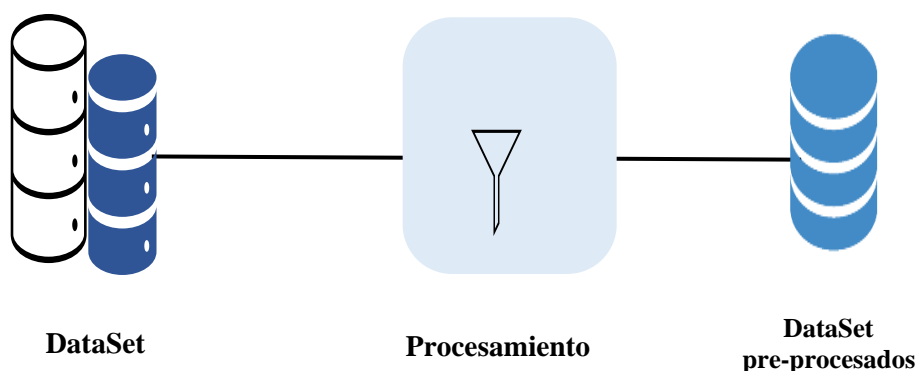
**Fuente:** Investigación

**Elaborado:** Jazmín Mora

#### 4.1.1.1. *Procesamiento de datos.*

En esta fase se realiza la depuración del DataSet de las variables preliminar seleccionadas para cargarlos en la herramienta informática Tableau Prep donde se ejecuta rutinas como problemas de escritura, espacios en blanco, fecha con diferentes formatos, valores nulos, repetidos entre otros, que permiten la limpieza y transformación de los datos mostrando como resultado un DataSet con los datos pre-procesados. Tal como se visualiza en la figura siguiente

**Figura 5** *Procesamiento de datos*



**Nota:** El gráfico representa el proceso que se llevó a cabo para pre-procesar el DataSet.

Elaborado por: Jazmín Mora

#### 4.1.1.2. Exploración de datos.

Se muestra un resumen estadístico del DataSet, la tabla 5 proporciona una visión general del estado y la calidad de los datos antes de realizar el análisis de correlación múltiple, para ser aplicados al modelo predictivo. Permite evaluar si los datos son coherentes, identificar posibles anomalías y comprender mejor las características fundamentales de las variables preliminar. A continuación, se explican cada una de las medidas.

**Count (cantidad):** Indica el número total de observaciones para cada variable. Como se visualiza en la tabla 5, muestra que todas las variables tienen 146,662 datos. Es un buen indicador de consistencia y asegura que no hay datos faltantes en las variables.

**Mean (media):** Promedio de todas las observaciones para una variable específica. La media es útil para entender el valor central típico de los datos.

**Std (Desviación estándar):** Mide la cantidad de dispersión de un conjunto de valores. La desviación estándar baja indica que los valores tienden a estar cerca de la media, mientras que un alta indica que los valores están más dispersos.

**Cuartiles :** Los cuartiles dividen datos ordenados en cuatro partes. El primer cuartil (25%) es el valor por debajo del cual se encuentra un cuarto de los datos. El segundo cuartil (mediana, 50%) es el punto medio, y el tercer cuartil (75%) es el valor por debajo del cual se encuentra el 75% de las observaciones. Estos valores indican la distribución y el rango de los datos.

**Min (Mínimo) y Max (máximo):** Valor que muestra el máximo y mínimo de los valores.

**Tabla 5** Descripción de las variables preliminar utilizadas para el análisis de correlación múltiple.

Detalle	Peso	Semana	Manos	Calibración sup	Calibración inf	Edad	Lote	t. <sup>a</sup> Max	t. <sup>a</sup> Min
count	146.662	146.662	146.662	146.662	146.662	146.662	146.662	146.662	146.662
mean	84	16.2	9	44.6	40.6	12.3	5.8	28.08	21.6
std	16.25	12.5	0.960	1.817	3.137	0.719	3.422	2.13	1.24
min	11.37	1	4	0	0	8	1	23.60	18.50
25%	73.78	7	8	44	40	12	3	26.40	20.70
50%	84.45	16	9	45	41	12	6	28	21.70
75%	94.46	30	9	46	42	13	8	29.80	22.80
max	247.58	39	15	58	51	16	12	32.20	23.60

**Fuente:** Investigación

**Elaborado:** Jazmín Mora

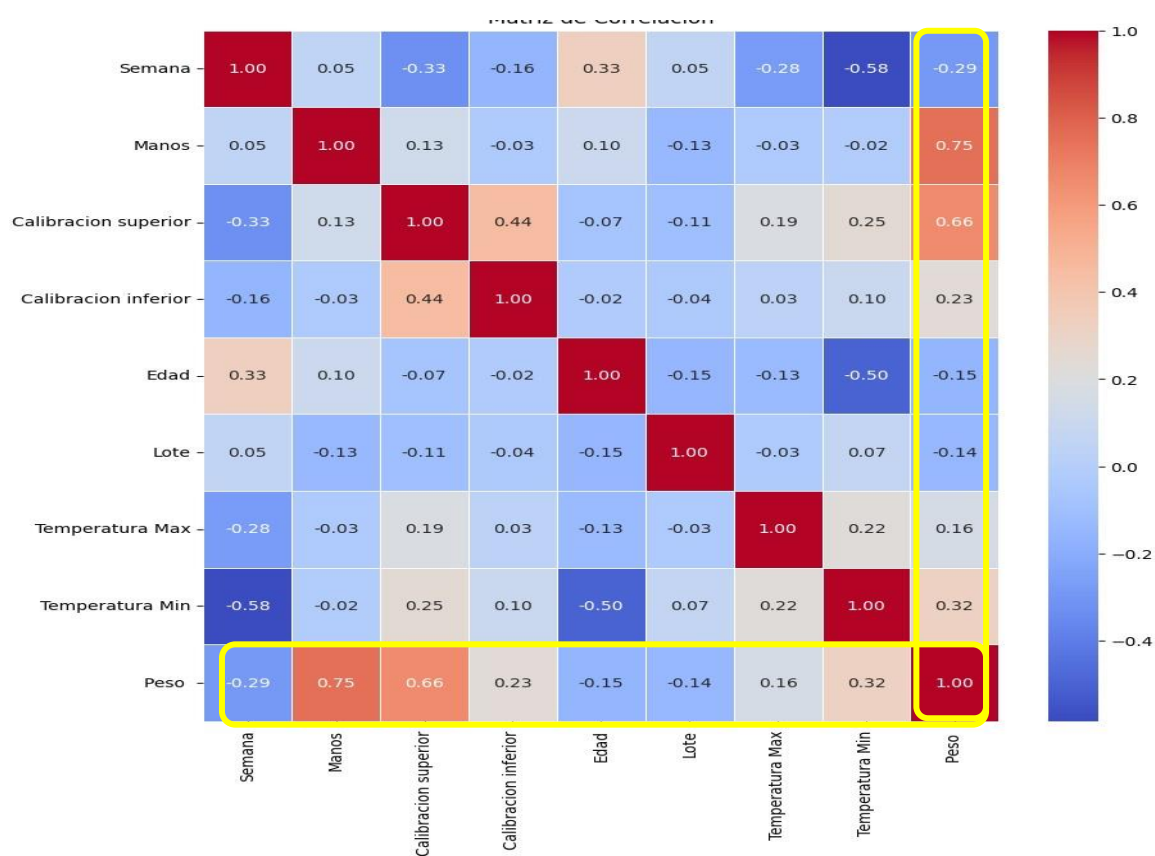
#### 4.1.1.3. Matriz de correlación.

Las variables preliminares mencionadas, se le aplicó un análisis de correlación múltiple para identificar las de mayor influencia en el peso del racimo de banano. Los resultados se muestran en la figura 6.

Los valores en la matriz de correlación varían entre -1 y +1.

- El valor de 1 indica una correlación positiva perfecta (cuando una variable aumenta, la otra también lo hace).
- El valor de -1 indica una correlación negativa perfecta (cuando una variable aumenta, la otra disminuye).
- Un valor de 0 indica que no hay correlación entre las variables.

**Figura 6** Matriz de correlación de las variables preliminar



**Nota:** El gráfico representa el análisis de correlación entre las variables. Elaborado por: Jazmín Mora

La fuerza de correlación es representada mediante una escala de colores, que va desde el rojo (correlación positiva fuerte), azul (correlación negativa fuerte) y blanco (correlación nula). Destacando las correlaciones significativas se encuentran en las siguientes variables: manos, calibración superior y temperatura mínima. También, la mayor correlación negativa en semana. El análisis de correlación múltiple permitió identificar las variables de mayor influencia en el peso del racimo del banano, se detallan en la tabla 6.

**Tabla 6** *Variables identificadas en el análisis de correlación, para el modelo predictivo*

<b>Variabes más influyentes en el peso del racimo de banano</b>	<b>Valor de correlación</b>
Manos	0.75
Calibración superior	0.66
Temperatura Mínima	0.32
Semana	-0.29

**Fuente:** Investigación

**Elaborado:** Jazmín Mora

#### **4.1.2. Algoritmo adecuado y entrenamiento del modelo predictivo**

Se aplicaron cinco algoritmos de ML con el fin de determinar cuál se adapta mejor entrenamiento del modelo predictivo, destinado a estimar el peso del racimo de banano. La revisión literaria de ML reveló que el aprendizaje supervisado es el enfoque que mejor se adecua al conjunto de datos en este estudio. En esta técnica, los algoritmos entrenados utilizando datos etiquetados, permite aprender la relación entre las entradas y salidas. Dentro de las modalidades del aprendizaje supervisado, se pudo aplicar clasificación y regresión, optando por regresión debido a la naturaleza continua de la variable de salida (el peso del racimo de banano) y las variables predictoras que son etiquetadas.

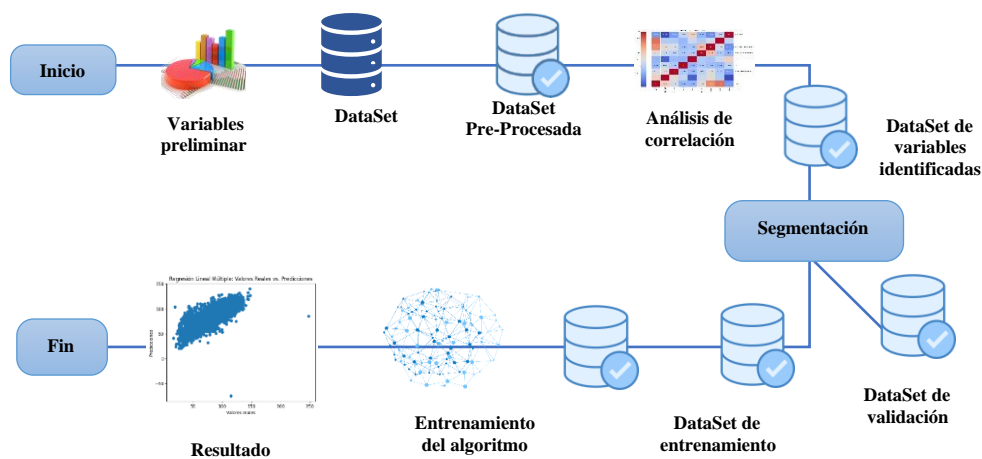
#### 4.1.2.1. Algoritmos de aprendizaje supervisado para estimar el peso del racimo de banano.

Los algoritmos aplicados en este estudio para abordar la tarea de estimar el peso del racimo de banano son: regresión de gradiente potenciado, regresión de Ridge, regresión de vecinos más cercanos, regresión Lasso y máquina de potenciación de gradiente con implementación de árboles ligeros. Cada uno de estos algoritmos desempeña un papel crucial, ofrecen diferentes enfoques y, técnicas para modelar y predecir el peso del racimo de banano.

#### Pruebas de los algoritmos seleccionados

Para evaluar los distintos algoritmos, se utilizó el entorno de desarrollo integrado Visual Studio Code, facilitando la implementación y prueba de cada uno de ellos. Este proceso permitió identificar el algoritmo óptimo para la construcción del modelo predictivo. Cada paso llevado a cabo en el análisis de los algoritmos se encuentra detallado y esquematizado en la figura 7, proporcionando una visión clara y estructurada del procedimiento seguido.

Figura 7 Diagrama de flujo de la realización del modelo predictivo



**Nota:** El gráfico representa el proceso que se llevó a cabo para entrenar el modelo. Elaborado por: Jazmín Mora

**VARIABLES preliminar:** Mediante el método del juicio se procedió a identificar las variables que podrían ser influyentes en peso del racimo del banano, con el experto en la agricultura Ing. Juan Carlos Elizalde. M.Sc.

**DataSet:** Variables preliminar de producción del banano periodo 2022.

**DataSet pre-procesado:** variables preliminar con los datos pre-procesados, para realizar el análisis de correlación múltiples.

**Análisis de correlación múltiple:** Identificación de las variables más influyentes en el peso del racimo del banano.

**DataSet de variables identificadas:** variables a utilizar en el modelo.

**Segmentación:** El flujo se divide en dos ramas, un DataSet de validación (20%) y el otro de entrenamiento (80%).

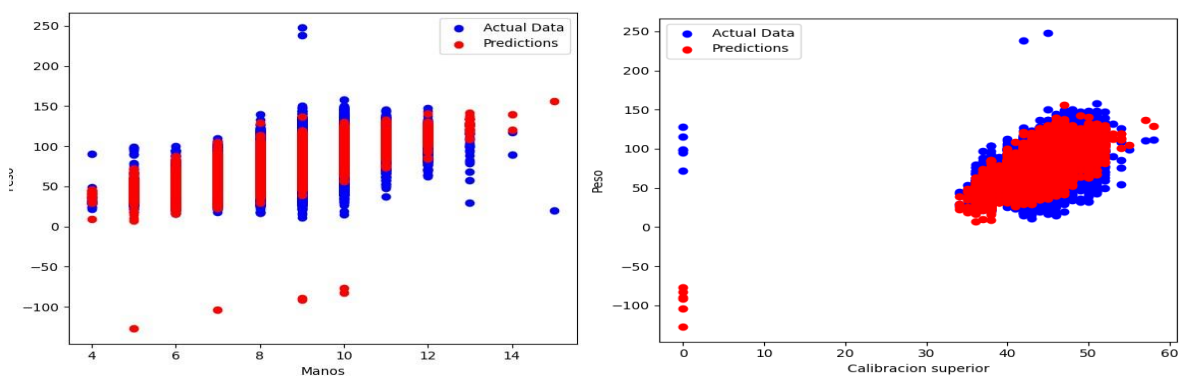
**Entrenamiento del algoritmo:** Se realiza el proceso del entrenamiento del modelo.

**Resultado:** Se muestran los resultados de una gráfica de dispersión que muestra valores reales vs predicciones y las métricas de error cuadrático medio (MSE), raíz del error cuadrático medio (RMSE), error absoluto medio (MAE) y coeficiente de determinación ( $R^2$ ). Ayudando a determinar cuál de los algoritmos es el más adecuado para este estudio.

- **Regresión de Ridge**

Técnica de regularización ampliamente empleada en estadísticas y, aprendizaje automático con el propósito de mejorar la estabilidad y la capacidad de generalización de los modelos de regresión lineal. En este contexto específico, se aplica para estimar el peso de racimos de banano. Esta metodología posibilita la comparación entre los datos reales y las predicciones generadas por un modelo de regresión.

**Figura 8** Regresión Ridge, el peso en función de manos y calibración superior



**Nota:** La figura representa la comparación de los valores reales vs las predicciones realizadas por el modelo. Elaborado por: Jazmín Mora

Las cuatro variables de mayor correlación identificadas en este estudio, se escogieron aquellas que mostraron la correlación más significativa, positiva (manos y calibración superior) para su representación gráfica. La Figura 8 presenta dos gráficos de dispersión que contrastan los valores observados (puntos azules) con las predicciones del modelo (puntos rojos). En el eje  $x$  se ubican las variables predictoras mano y calibración superior, mientras que en el eje  $y$  se representa la variable objetivo peso.

#### **El error cuadrático (MSE) o la Raíz del Error Cuadrático Medio (RMSE).**

En el análisis del algoritmo de regresión de Ridge, se observa que MSE alcanza un valor de 82.951, RMSE corresponde a 9.107. Estas métricas revelan la discrepancia entre los valores reales y las predicciones. Es evidente que el modelo no logra cumplir con la meta de minimizar la magnitud de los errores. Por ende, la elección del modelo ideal implica la selección del MSE o RMSE más bajo, este algoritmo no satisface los criterios establecidos.

#### **Error absoluto medio (MAE).**

El algoritmo de regresión de Ridge, se obtiene un MAE de 6.89. Esta métrica es empleada para minimizar la sensibilidad a errores significativos y reducir la influencia de valores atípicos. La selección del modelo ideal se basa en minimizar el MAE. En este caso, el algoritmo de regresión de Ridge no satisface dicho criterio.

#### **Coefficiente de Determinación ( $R^2$ ).**

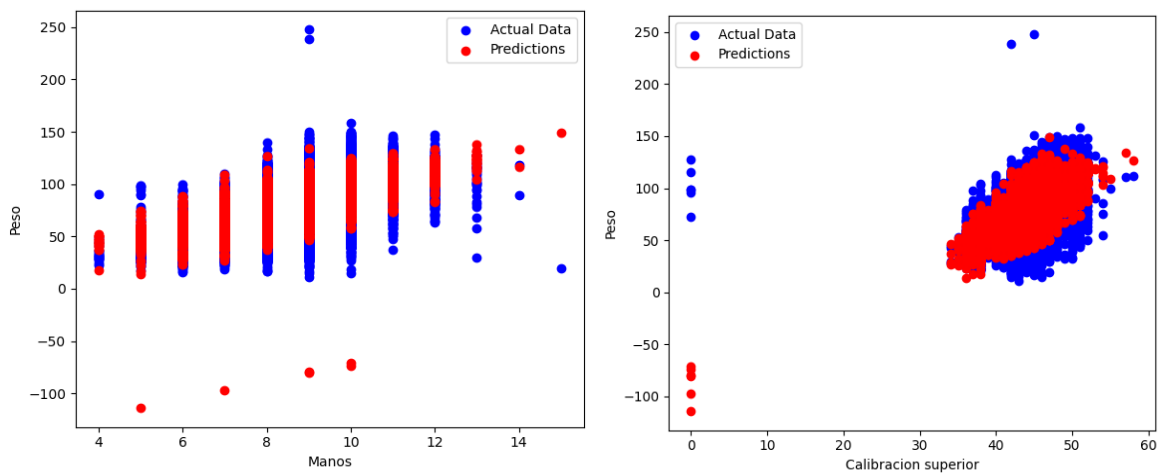
El  $R^2$  de 0.689, esta métrica evalúa la capacidad del modelo para explicar la variabilidad en los datos reales. Un valor de 0.689 indica que el 69% de la variabilidad en los datos reales puede ser explicada por el algoritmo. Se busca seleccionar el modelo con el  $R^2$  más alto y este no cumple con lo establecido.

La figura 8 muestra gráficos de dispersión, se aprecia discrepancia entre los pesos predichos (puntos rojos) y los valores reales (puntos azules), indicando errores en las predicciones. Esto se evidencia en las métricas de evaluación como MSE, RMSE, MAE y  $R^2$ . Es importante destacar que este algoritmo no resulta ser la opción más idónea para la estimación del peso de racimos de banano.

- **Regresión Lasso**

Método que encuentra un equilibrio entre ajustar los datos de entrenamiento de manera precisa y mantener un modelo simple que generalice bien a nuevos datos. En este contexto específico, se aplica para estimar el peso de racimos de banano. Esta metodología posibilita la comparación entre los datos reales y las predicciones generadas por el modelo.

**Figura 9** Regresión Lasso, el peso en función de calibración superior y manos



**Nota:** La imagen representa la comparación de los valores reales vs las predicciones realizadas por el modelo. Elaborado por: Jazmín Mora

Las cuatro variables de mayor correlación identificadas en este estudio, se escogieron aquellas que mostraron la correlación más significativa, positiva (manos y calibración superior) para su representación gráfica. La Figura 9 presenta dos gráficos de dispersión que contrastan los valores observados (puntos azules) con las predicciones del modelo (puntos rojos). En el eje  $x$  se ubican las variables predictoras mano y calibración superior, mientras que en el eje  $y$  se representa la variable objetivo peso.

**El error cuadrático (MSE) o la Raíz del Error Cuadrático Medio (RMSE).**

El rendimiento del algoritmo de regresión Lasso se evalúa mediante el MSE, que alcanza un valor de 87.125, y el RMSE, que se sitúa en 9.334. Estas métricas reflejan la disparidad entre los valores reales y las predicciones. Se observa que no logra cumplir con la meta de minimizar la magnitud de los errores. En consecuencia, este algoritmo no satisface los criterios deseados para la selección del modelo, que busca minimizar el MSE o RMSE.

### **Error absoluto medio (MAE).**

El valor obtenido del MAE mediante el algoritmo de regresión Lasso es de 7.094. Esta métrica se elige para minimizar la sensibilidad a errores significativos y reducir la influencia de valores atípicos. La meta es seleccionar el modelo con el MAE más bajo. En este contexto el algoritmo de regresión Lasso no satisface este objetivo.

### **Coefficiente de Determinación ( $R^2$ ).**

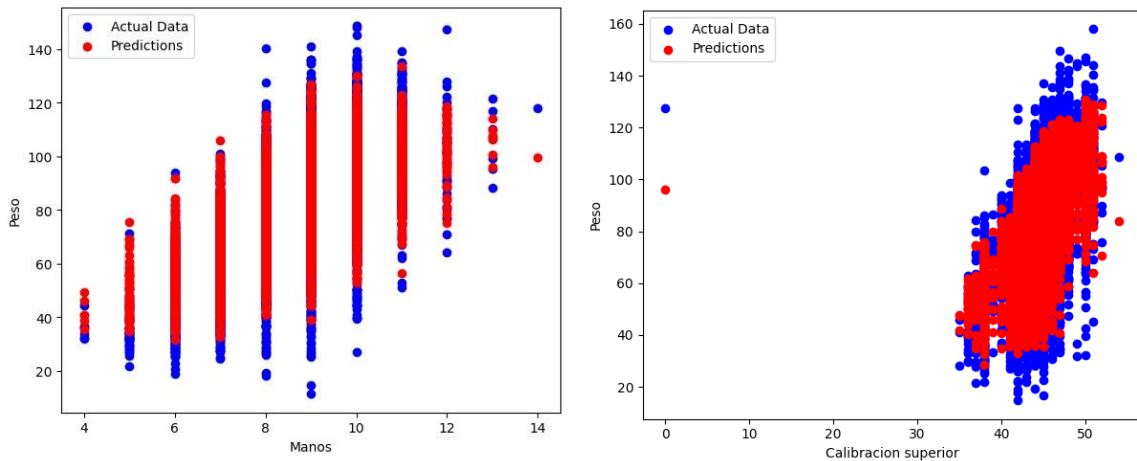
El coeficiente de determinación  $R^2$  obtenido mediante el algoritmo de regresión Lasso es de 0.669. Esta métrica evalúa la capacidad del modelo para explicar la variabilidad en los datos reales. Un valor de 0.669 indica que el 67% de la variabilidad en los datos reales puede ser explicada por el algoritmo de regresión Lasso. Dado que se busca seleccionar el modelo con el  $R^2$  más alto, este algoritmo no satisface completamente dicho criterio.

La figura 9 muestra gráficas de dispersión, se aprecia discrepancia entre los pesos predichos (puntos rojos) y los valores reales (puntos azules), indicando errores en las predicciones. Esto se evidencia en las métricas de evaluación como MSE, RMSE, MAE y  $R^2$ . Es importante destacar que este algoritmo no resulta ser la opción más idónea para la estimación del peso de racimos de banano.

- **Regresión de Vecinos Más Cercanos (K-NN)**

Algoritmo de aprendizaje supervisado utilizado tanto para clasificación y regresión. Es simple pero efectivo, se basa en la idea de que objetos similares están cerca unos de otros. En este contexto específico, utilizado para estimar el peso de racimos de banano. Esta metodología posibilita la comparación entre los datos reales y las predicciones generadas por el modelo.

**Figura 10** KNN, peso en función de calibración superior y manos



**Nota:** La imagen representa la comparación de los valores reales vs las predicciones realizadas por el modelo. Elaborado por: Jazmín Mora

Las cuatro variables de mayor correlación identificadas en este estudio, se escogieron aquellas que mostraron la correlación más significativa, positiva (manos y calibración superior) para su representación gráfica. La Figura 10 presenta dos gráficos de dispersión que contrastan los valores observados (puntos azules) con las predicciones del modelo (puntos rojos). En el eje  $x$  se ubican las variables predictoras mano y calibración superior, mientras que en el eje  $y$  se representa la variable objetivo peso.

#### **El error cuadrático (MSE) o la Raíz del Error Cuadrático Medio (RMSE).**

El algoritmo KNN tiene un MSE de 85.672 y un RMSE de 9.255. Estas métricas revelan la discrepancia entre los valores reales y las predicciones, indicando que el modelo no logra minimizar de manera efectiva la magnitud de los errores. Por lo tanto, el algoritmo no satisface los criterios deseados para la selección del modelo, busca minimizar el MSE o el RMSE.

#### **Error absoluto medio (MAE).**

El MAE obtenido del algoritmo KNN de 7.025, esta métrica busca ser poco sensible a errores significativos y minimizar valores atípicos. La selección del modelo se basa en la minimización del MAE, por lo tanto, el algoritmo no cumple con esta métrica.

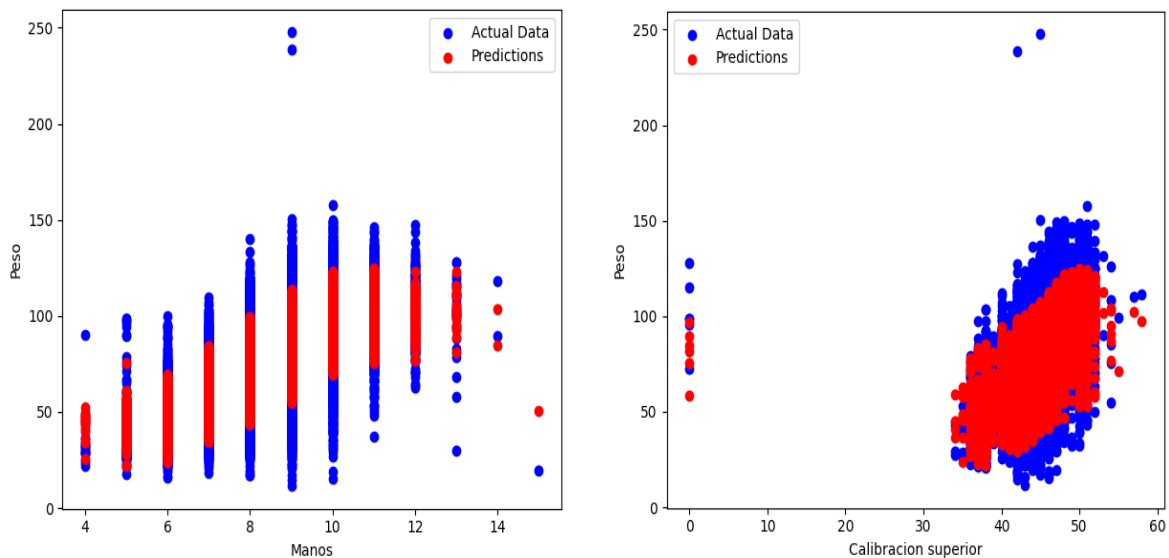
### **Coefficiente de Determinación ( $R^2$ ).**

El algoritmo de regresión KNN obtuvo un  $R^2$  de 0.674, significa que puede explicar el 68% de la variabilidad en los datos reales. Sin embargo, se busca un modelo con un  $R^2$  más alto, por lo que el desempeño del algoritmo KNN no cumple con el criterio deseado. La figura 10 muestra gráficas de dispersión, apreciando diferencia entre los pesos predichos (puntos rojos) y los valores reales (puntos azules), indican errores en las predicciones. Esto se evidencia en las métricas de evaluación como MSE, RMSE, MAE y  $R^2$ . Es importante destacar que este algoritmo no resulta ser la opción más idónea para la estimación del peso de racimos de banano.

- **Regresión de Gradiente Potenciado.**

Método avanzado de aprendizaje automático utilizado para tareas de regresión y clasificación. Combina las predicciones de varios modelos más simples para generar una predicción final más precisa y robusta. En este contexto específico, se aplica para estimar el peso de racimos de banano. Esta metodología posibilita la comparación entre los datos reales y las predicciones generadas por el modelo.

**Figura 11** *Regresión de Gradiente Potenciado, en peso en función de semana y manos*



**Nota:** La imagen representa la comparación de los valores reales vs las predicciones realizadas por el modelo. Elaborado por: Jazmín Mora

Las cuatros variables de mayor correlación identificadas en estes estudio, se escogieron aquellas que mostraron la correlación más significativa, positiva (manos y calibración superior) para su representación gráfica. La Figura 11 presenta dos gráficos de dispersión que contrastan los valores observados (puntos azules) con las predicciones del modelo (puntos rojos). En el eje  $x$  se ubican las variables predictoras mano y calibración superior, mientras que en el eje  $y$  se representa la variable objetivo peso.

#### **El error cuadrático (MSE) o la Raíz del Error Cuadrático Medio (RMSE).**

El algoritmo de regresión gradiente potenciado tiene un MSE de 71.915 y un RMSE de 8.480. Estas métricas revelan la discrepancia entre los valores reales y las predicciones, indicando que el modelo logra minimizar de manera no tan efectiva la magnitud de los errores. Por lo tanto, este algoritmo no satisface los criterios deseados para la selección del modelo, que busca minimizar el MSE o el RMSE.

#### **Error absoluto medio (MAE).**

El MAE de 6.388, esta métrica se diseñó para ser poco susceptible a errores y reducir valores atípicos. La elección del modelo busca el menor MAE, por lo tanto, el algoritmo de Regresión Gradiente Potenciado no satisface el criterio métrico.

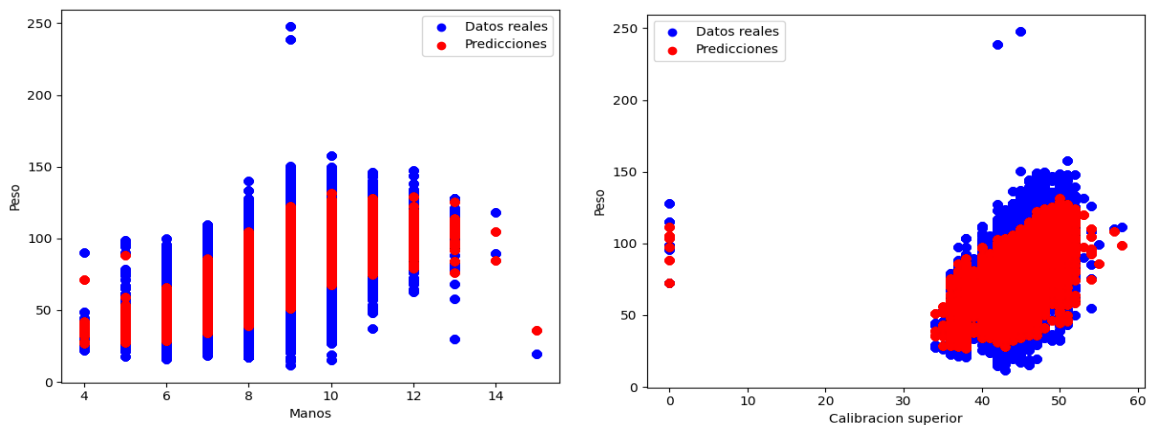
#### **Coefficiente de Determinación ( $R^2$ ).**

El algoritmo de regresión gradiente potenciado obtuvo un  $R^2$  de 0.727, significa que puede explicar el 73% de la variabilidad en los datos reales. Sin embargo, se busca un modelo con un  $R^2$  más alto, por lo que el desempeño del algoritmo no cumple con el criterio deseado. La figura 11 muestra gráficas de dispersión, apreciando diferencia entre los pesos predichos (puntos rojos) y los valores reales (puntos azules), indican errores en las predicciones. Esto se evidencia en las métricas de evaluación como MSE, RMSE, MAE y  $R^2$ . Es importante destacar que este algoritmo no resulta ser la opción más idónea para la estimación del peso de racimos de banano.

- **Gradient Boosting Machine con Implementación de Árboles Ligeros (LightGBM).**

Este algoritmo es un método avanzado de aprendizaje automático se utiliza tanto para tareas de regresión y clasificación. Combina las predicciones de varios modelos simples para generar una predicción final más precisa y robusta. En este contexto específico, es aplicado para estimar el peso de racimos de banano. Esta metodología posibilita la comparación entre los datos reales y las predicciones generadas por el modelo.

**Figura 12** *LightGBM, en peso en función de manos y calibración superior*



**Nota:** La imagen representa la comparación de los valores reales vs las predicciones realizadas por el modelo. Elaborado por: Jazmín Mora

Las cuatro variables de mayor correlación identificadas en este estudio, se escogieron aquellas que mostraron la correlación más significativa, positiva (manos y calibración superior) para su representación gráfica. La Figura 12 presenta dos gráficos de dispersión que contrastan los valores observados (puntos azules) con las predicciones del modelo (puntos rojos). En el eje  $x$  se ubican las variables predictoras mano y calibración superior, mientras que en el eje  $y$  se representa la variable objetivo peso.

**El error cuadrático (MSE) o la Raíz del Error Cuadrático Medio (RMSE).**

El algoritmo de LightGBM tiene un MSE de 50.43 y un RMSE de 7.10. Estas métricas revelan la disparidad entre los valores reales y las predicciones, indicando que el modelo logra minimizar de manera efectiva la magnitud de los errores. En consecuencia, este algoritmo satisface los criterios deseados para la selección del modelo, que busca minimizar el MSE o el RMSE.

### **Error absoluto medio (MAE).**

El MAE de 4.21, esta métrica se diseñó para ser poco susceptible a errores significativos y para reducir al mínimo la influencia de valores atípicos. La elección del modelo se fundamenta en la minimización del MAE, por lo tanto, el algoritmo LightGBM satisface eficazmente este criterio métrico.

### **Coefficiente de Determinación (R<sup>2</sup>).**

El R<sup>2</sup> obtenido mediante el algoritmo LightGBM asciende a 0.85 significa que puede explicar el 85% de la variabilidad en los datos reales. Al buscar la elección del modelo con el R<sup>2</sup> más elevado, LightGBM cumple con este criterio métrico.

La figura 12 de dispersión, se aprecia una notoria convergencia entre los datos reales y las predicciones, sugiriendo que el modelo posee la capacidad de realizar predicciones con un grado de precisión significativo. A pesar de ello, se identifican algunas predicciones (puntos rojos) que se encuentran considerablemente distantes de los datos reales (puntos azules). Estas desviaciones, la gran mayoría de las predicciones se sitúan en cercanía a los valores reales. En comparación con otros modelos evaluados, LightGBM exhibe el RMSE bajo y el R<sup>2</sup> más alto. Estos indicadores sugieren, que entre todos los algoritmos LightGBM es más preciso y se ajusta de manera óptima a los datos.

En la tabla 7, se observan las métricas de los algoritmos que obtuvieron los mejores resultados.

**Tabla 7** comparación de las métricas de los modelos.

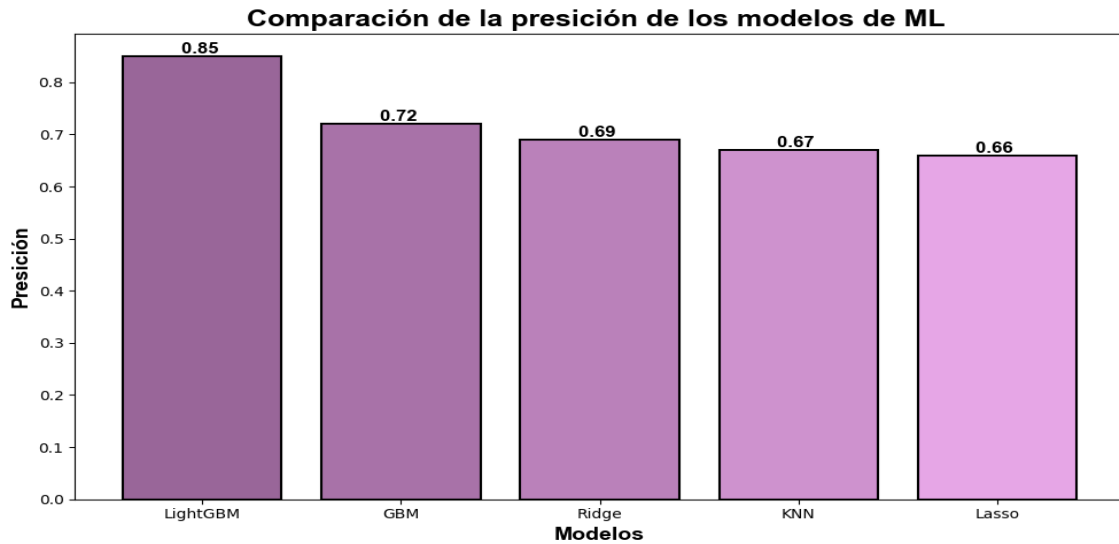
<b>MODELO</b>	<b>MSE</b>	<b>RMSE</b>	<b>MAE</b>	<b>R<sup>2</sup></b>
<b>LightGBM</b>	<b>50.43</b>	<b>7.10</b>	<b>4.21</b>	<b>0.85</b>
<b>GBR</b>	71.915	8.480	6.388	0.727
<b>Ridge</b>	82.951	9.107	6.896	0.685
<b>KNN</b>	85.672	9.255	7.025	0.674
<b>LASSO</b>	87.125	9.334	7.094	0.669

**Fuente:** Investigación

**Elaborado:** Jazmín Mora

En la tabla 7, LightGBM tiene el MAE, RMSE, MSE bajo y el  $R^2$  alto frente a sus competidores. Estas métricas permitieron definir que LightGBM resulto ser el mejor en este estudio.

**Figura 13** Comparación de la precisión de los modelos entrenado.



**Nota:** El gráfico representa la presión de los modelos con cada algoritmo empleado. Elaborado por: Jazmín Mora

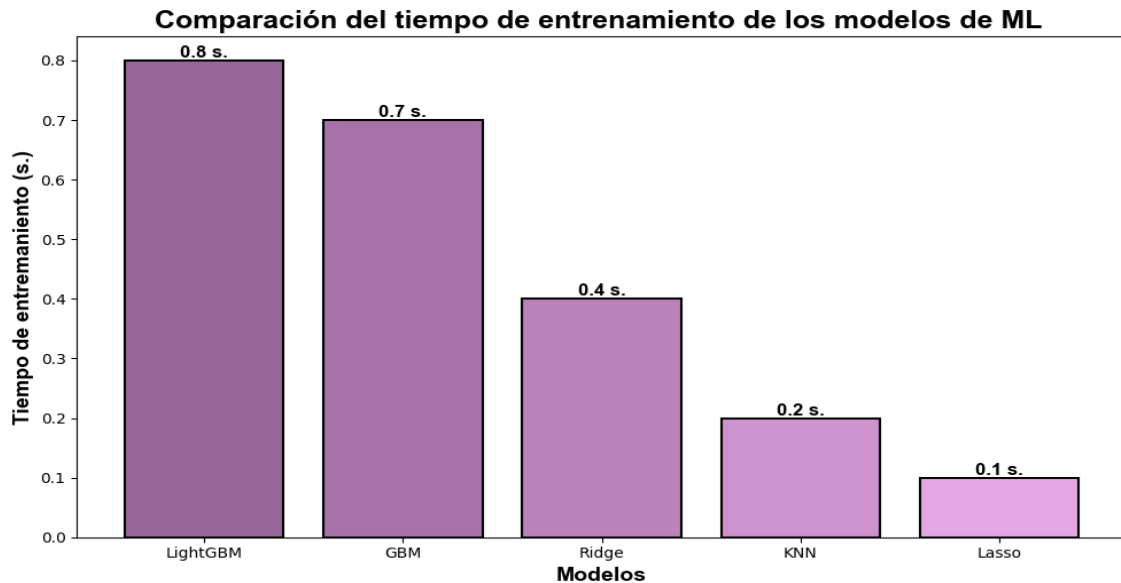
La figura 13 de barra muestra la precisión de cada uno de los diferentes modelos de ML. En el eje  $x$  están etiquetados los modelos y en  $y$  la precisión. Permite interpretar rápidamente que el LightGBM tiene la mejor precisión entre los modelos comparados.

- LightGBM: Este modelo tiene la precisión más alta, aproximadamente 0.85.
- GBM: Con una precisión ligeramente superior, alrededor de 0.72.
- RR: Con una precisión aproximada de 0.69.
- KNN: Con una precisión cerca de 0.67.
- RLS: Con una precisión cercana a 0.66.

En la Figura 14, ofrece una comparación visual del tiempo requerido para entrenar distintos modelos de aprendizaje automático, cada uno utilizando un algoritmo específico. El eje vertical, se aprecia el tiempo de entrenamiento expresado en segundos, funcionando como un indicador del periodo de preparación. Por otro lado, el eje horizontal muestra los diferentes modelos evaluados.

- LightGBM: fue el que tomó más tiempo de entrenamiento de 0.8 segundos.
- GBM: se tomó un tiempo de entrenamiento alrededor de 0.7 segundos.
- Ridge: El tiempo de entrenamiento es de aproximadamente 0.4 segundos.
- KNN: Con un tiempo de entrenamiento alrededor de 0.2 segundos.
- Lasso: Este es el modelo que menos tiempo tomó entrenar, con solo 0.1 segundos.

**Figura 14** Comparación del tiempo de entrenamiento de los modelos de ML.



**Nota:** El gráfico representa el tiempo de entrenamiento de los modelos con cada algoritmo seleccionado. Elaborado por: Jazmín Mora

El que tomó más tiempo de entrenamiento fue LightGBM, se debe a su naturaleza de árbol de decisión y complejidad de su proceso de entrenamiento. Los algoritmos basados en árboles, como LightGBM y GBM, requieren más tiempo de entrenamiento en comparación con modelos lineales como Ridge o métodos más simples como KNN y Lasso.

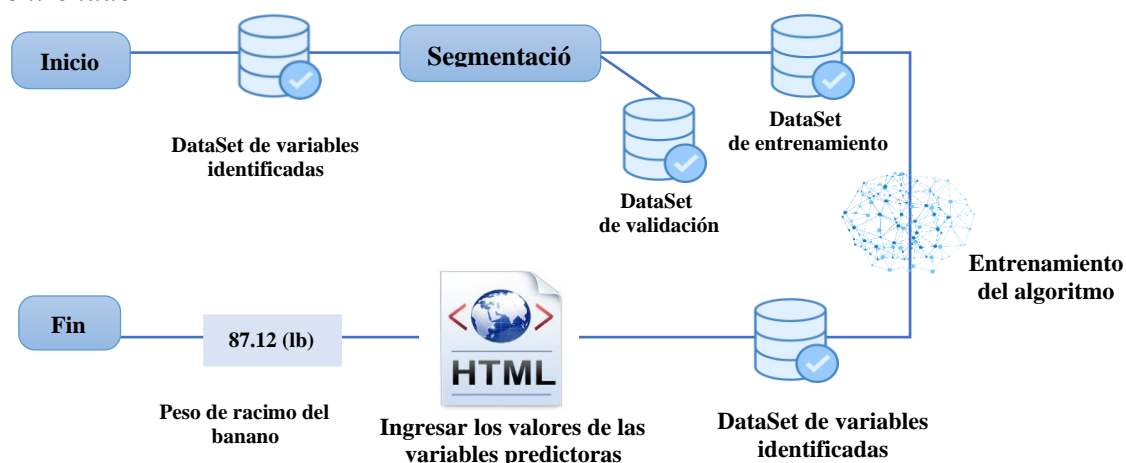
Por estas razones, LightGBM fue considerado el más preciso y efectivo para las predicciones. La elección del algoritmo es un paso fundamental para poder desarrollar el modelo de aprendizaje automático exitoso y útil. La elección de LightGBM tiene las mejores métricas a diferencias de sus competidores en este estudio. Contribuye significativamente en la robustez y la calidad de las predicciones, respaldando así la capacidad del modelo para proporcionar información valiosa en el ámbito de estimar el peso de los racimos de banano.

#### 4.1.3. Desempeño del modelo de predicción entrenado a través de pruebas de rendimiento para validar la estimación calculada

Este proceso garantiza que el modelo de predicción no solo se adapte de manera adecuada a los datos empleados durante su entrenamiento, sino que también mantenga la habilidad de efectuar predicciones exactas al enfrentarse a nuevos conjuntos de datos. Dicha habilidad resulta fundamental para asegurar la aplicabilidad y relevancia del modelo en escenarios del mundo real.

La Figura 15 presenta el esquema del diagrama de flujo diseñado para verificar el rendimiento del modelo de predicción entrenado a través de pruebas de rendimiento, con el propósito de validar la estimación calculada del peso del racimo de banano.

**Figura 15** Diagrama de comprobación del desempeño del modelo de predicción entrenado



**Nota:** El gráfico representa el proceso que se realizó para comprobar el desempeño del modelo. Elaborado por: Jazmín Mora

El DataSet de variable identificadas para estimar el peso del racimo de banano, se procedió a realizar la segmentación de los datos: un 80% se destinó al entrenamiento, permitiendo que el algoritmo LightGBM descubra y aprenda las interacciones entre las variables predictoras y la variable dependiente. El 20% restante se reservó para la validación, es esencial para verificar la habilidad del modelo en generar predicciones precisas con datos que no participaron en el entrenamiento. Este proceso asegura que el modelo no solo aprenda eficazmente, sino que también demuestre su capacidad de generalización en situaciones nuevas.

Tras el entrenamiento exitoso del modelo, se procedió a evaluar su rendimiento utilizando el DataSet de las variables identificadas. A continuación, se ingresaron los datos de las variables predictoras y finalmente obteniendo el peso del racimo de banano.

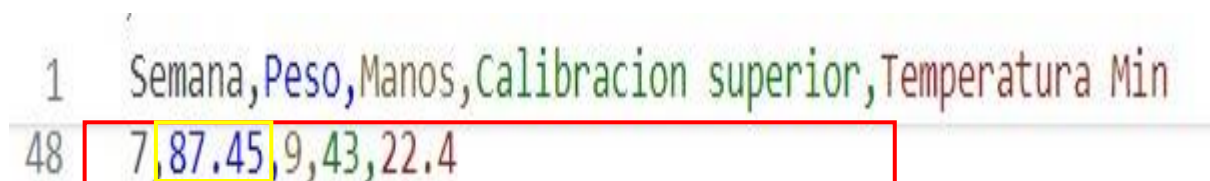
#### 4.1.3.1. Desarrollo de la página web del modelo predictivo.

La página web desarrollada en Python tiene como objetivo principal facilitar la prueba y validación del modelo predictivo diseñado para estimar el peso de los racimos de banano. Este modelo se ha creado con el propósito de optimizar y mejorar la precisión en la estimación del peso de los racimos, contribuyendo así a la eficiencia en la producción bananera.

La aplicación cuenta con una interfaz de usuario fácil de usar, permitiendo a los usuarios interactuar de manera sencilla con las funciones de prueba del modelo. También les permite a los usuarios cargar los datos de las variables predictoras para realizar las pruebas. Además, facilita la prueba del modelo predictivo utilizando los datos cargados.

Una fila del DataSet variables predictoras es seleccionada para realizar prueba de rendimiento del modelo, tal como se muestra en la Figura 16.

**Figura 16** DataSet de validación.



1	Semana	Peso	Manos	Calibracion superior	Temperatura Min
48	7	87.45	9	43	22.4

**Nota:** El gráfico representa la fila del DataSet que se utilizó, para realizar las pruebas del desempeño del modelo. Elaborado por: Jazmín Mora

La selección de la fila 48 se llevó a cabo con el propósito de ejecutar una prueba de rendimiento del modelo predictivo. Esta acción implicó la introducción manual de los valores correspondientes a cada variable predictora en la interfaz proporcionada por la página web. El objetivo es calcular estimación para el peso del racimo de banano. El dato registrado en la fila 48 indicaba un peso real de 87.45 libras. Tras la aplicación del modelo, se obtuvo una estimación que sugiere un peso para el racimo de banano de 87.12 libras, denota una proximidad notable al peso real y subraya la precisión del modelo entrenado.

**Figura 17** Interfaz del ingreso de los valores de las variables predictoras



**Nota:** El gráfico representa la predicción realizada en la página web. Elaborado por: Jazmín Mora

**Figura 18** Resultado de la predicción del peso del racimo de banano.



**Nota:** El gráfico representa el resultado de la predicción realizada en la página web. Elaborado por: Jazmín Mora

## 4.2. Discusión.

El proyecto magistral realizado por Ortiz Mata Jhonny Darwin, titulado “*Diseño de un modelo predictivo basado en algoritmos de Machine Learning para la estimación del peso de racimos de banano, caso de estudio hacienda bananera en ecuador*”. Comparando los resultados con el tema “Estimación del peso del racimo de banano por medio de algoritmos de aprendizaje automático”, se hacen evidentes varias diferencias clave y ventajas potenciales.

- Jhonny Darwin Ortiz Mata emplea seis variables, seleccionadas directamente y concluye al terminar su proyecto que solo tres de estas tienen una influencia significativa en el peso del racimo de banano. A diferencia, en el desarrollo de este estudio se optó por seleccionar nueve variables mediante el criterio del experto en agricultura el Ing. Juan Carlos Elizalde. Posteriormente, se llevó a cabo un análisis de correlación múltiple, para determinar que cuatro variables destacan como las más influyentes en el peso del racimo del banano y utilizarlas para construir el modelo predictivo.
- En el trabajo de Ortiz Mata Jhonny Darwin, empleo dos algoritmos: Random Forest y eXtreme Gradient Boosting, con el fin de identificar el más eficaz. Los resultados indicaron que eXtreme Gradient Boosting fue superior. En el presente estudio, utilizo cinco algoritmos, destacando el rendimiento notable de LightGBM. Este algoritmo es preferido por su capacidad de manejar extensos conjuntos de datos, logrando modelos precisos y rápido en comparación con Random Forest y eXtreme Gradient Boosting. Ambas investigaciones utilizaron las mismas métricas para evaluar el rendimiento de los modelos. El estudio actual ha logrado los resultados más sobresalientes en estas métricas de evaluación.
- Jhonny Darwin Ortiz Mata también lleva a cabo la segmentación del DataSet en validación y entrenamiento. A diferencia de este estudio, se empleó el DataSet preprocesado utilizando las variables identificadas en el análisis de correlación múltiple. Además, se integró una aplicación web de fácil acceso y uso, permitiendo a los usuarios finales utilizar el modelo predictivo de manera sencilla.

## **CAPÍTULO V**

### **CONCLUSIONES Y RECOMENDACIONES**

## 5.1. Conclusiones

- El análisis de correlación múltiple permitió identificar las variables que influyen en el peso del racimo de banano, como el número de manos, calibración superior, temperatura mínima y semana del año. Estos factores han resultado fundamentales en este estudio. La correlación entre estas variables y el peso del racimo ha proporcionado una base sólida para el modelado predictivo, permitiendo comprender los elementos clave que impactan en el peso del racimo en la producción bananera.
- La selección del algoritmo máquina de potenciación de gradiente con implementación de árboles ligeros resultó ser la más acertada para el entrenamiento del modelo predictivo en este estudio. Tras una exhaustiva comparación con otros algoritmos, tales como Regresión de Gradiente Potenciado, Regresión de Ridge, Regresión de Vecinos Más Cercanos y Regresión Lasso. Mediante un análisis riguroso de su desempeño, se ha demostrado que el algoritmo de LightGBM sobresale frente a sus competidores en términos de precisión y eficiencia. Su bajo MSE, RMSE, MAE bajo y el  $R^2$  más alto destaca su capacidad superior para manejar la complejidad del conjunto de datos.
- La comprobación del modelo con el DataSet en la aplicación ha demostrado eficacia y fiabilidad, asegurando que el modelo no solo es teóricamente sólido sino también prácticamente viable y útil para la comunidad agrícola. La capacidad de la aplicación para procesar datos de entrada y ofrecer estimaciones de peso del banano en tiempo real representa un avance significativo en la tecnología agrícola.

## 5.2. Recomendaciones

Para futuros desarrollo se recomienda lo siguiente:

- Aunque se identificaron las variables importantes para predecir el peso del racimo de banano, futuros trabajos podrían ampliar el DataSet que incluyan datos de diferentes zonas de producción y condiciones climáticas. Esto permitirá al modelo aprender y adaptarse a una gama más amplia de situaciones, mejorando su capacidad de generalización y precisión en diferentes contextos.
- Aunque máquina de potenciación de gradiente con implementación de árboles ligeros (LightGBM) demostró ser el algoritmo más eficaz en este estudio, podrían explorar algoritmos adicionales.
- Realizar pruebas de campo, donde el modelo se aplique bajo condiciones reales de producción bananera, resulta esencial para evaluar su desempeño y realizar los ajustes pertinentes según sea necesario. También es crucial desarrollar una aplicación móvil que incorpore el modelo predictivo, con el fin de estimar el peso del racimo directamente en el campo.

## **CAPÍTULO VI**

### **BIBLIOGRAFÍA**

## 6.1. Bibliografía

- [1] L. Galarza Suárez, “Tierra, trabajo y tóxicos: sobre la producción de un territorio bananero en la costa sur del Ecuador,” *Estudios atacameños*, no. 63, pp. 341–364, 2019.
- [2] W. Arias Arana, N. Palma, and D. Riccio, “La apertura comercial y su incidencia en el crecimiento económico del Ecuador,” *Empresarial*, vol. 13, no. 1, pp. 29–34, Aug. 2019, doi: 10.23878/empr.v13i1.148.
- [3] J. P. León Ajila, M. A. Espinosa Aguilar, H. R. Carvajal Romero, and J. Quezada Campoverde, “Análisis de la producción y comercialización de banano en la provincia de El Oro en el periodo 2018-2022,” *Ciencia Latina Revista Científica Multidisciplinar*, vol. 7, no. 1, pp. 7494–7507, Mar. 2023, doi: 10.37811/cl\_rcm.v7i1.4981.
- [4] F. H. Andrade, “Los desafíos de la agricultura,” International Plant Nutrition Institute, 2016.
- [5] “I.-Las enfermedades del Banano.”
- [6] Alvaro Segura Monges, “Procesos Fisiológicos Involucrados en el Daño Producido por *Pseudocercospora fijiensis* en las Plantas de Banano (*Musa* spp): Efecto sobre el Follaje y Disminución del Peso del Racimo (parte,” jun. 2021.
- [7] A. Chlingaryan, S. Sukkariéh, and B. Whelan, “Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review,” *Comput Electron Agric*, vol. 151, pp. 61–69, 2018, doi: <https://doi.org/10.1016/j.compag.2018.05.012>.
- [8] M. En, T. De, and L. A. Información, “REPÚBLICA DEL ECUADOR VICERRECTORADO DE INVESTIGACIÓN Y POSGRADO PROYECTO DE INVESTIGACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO.”
- [9] J. D. R. Soares, M. Pasqual, W. S. Lacerda, S. O. Silva, and S. L. R. Donato, “Utilization of artificial neural networks in the prediction of the bunches’ weight in banana plants,” *Sci Horti*, vol. 155, pp. 24–29, May 2013, doi: 10.1016/j.scienta.2013.01.026.
- [10] S. L. Jayasinghe, C. J. K. Ranawana, I. C. Liyanage, and P. E. Kaliyadasa, “Growth and yield estimation of banana through mathematical modelling: a systematic

- review,” *Journal of Agricultural Science*, vol. 160, no. 3–4. Cambridge University Press, pp. 152–167, Jun. 23, 2022. doi: 10.1017/S0021859622000259.
- [11] T. Ogusua, “Application of artificial neural networks in the prediction of fruit damages and hand weight in Cavendish banana,” *Program in Postharvest Technology, School of Agro-Industry*, vol. 5, pp. 140–145, 2019.
- [12] L. Galarza Suárez, “Tierra, trabajo y tóxicos: Sobre la producción de un territorio bananero en la costa sur del Ecuador,” *Estudios Atacameños*, Nov. 2019, doi: 10.22199/issn.0718-1043-2019-0034.
- [13] Vasyl Cherlinka, “Temperatura Del Suelo Para La Siembra Y El Cultivo.”
- [14] A. Elbehri *et al.*, “Cambio climático y sostenibilidad del banano en el Ecuador,” *Roma: Organización de las Naciones Unidas para la Alimentación y la Agricultura*, 2015.
- [15] M. B. Anne Vézina, “Morfología de la planta del banano.”
- [16] Medardo Palacios, “LABORES DE COSECHA EN CULTIVO DE BANANO.”
- [17] J. S. Nolasco Valenzuela, *Python: aplicaciones prácticas*. RA-MA Editorial, 2018. [Online]. Available: <https://elibro.net/es/lc/uteq/titulos/106523>
- [18] EPITCH, “Qué es Flask (Python) y cuáles son sus principales ventajas.”
- [19] José Domínguez Muñoz, “Qué es Flask.”
- [20] J. M. Pineda, “Modelos predictivos en salud basados en aprendizaje de maquina (machine learning),” *Revista Médica Clínica Las Condes*, vol. 33, no. 6, pp. 583–590, Nov. 2022, doi: 10.1016/j.rmcl.2022.11.002.
- [21] F. Chen *et al.*, “Design of an Intelligent Variable-Flow Recirculating Aquaculture System Based on Machine Learning Methods,” *Applied Sciences*, vol. 11, no. 14, p. 6546, Jul. 2021, doi: 10.3390/app11146546.
- [22] J. Bobadilla, *Machine Learning y Deep Learning: Usando Python, Scikit y Keras*. RA-MA Editorial, 2020. [Online]. Available: <https://elibro.net/es/lc/uteq/titulos/222698>
- [23] A. M. Vázquez, “Introducción a machine learning.” Dialnet, 2018.
- [24] E. M. Rojas, “Machine Learning: análisis de lenguajes de programación y herramientas para desarrollo,” *Revista Ibérica de Sistemas e Tecnologías de Informação*, no. E28, pp. 586–599, 2020.
- [25] J. M. Ortega Candel, *Big data, machine learning y data science en Python*. RA-MA Editorial, 2022. [Online]. Available: <https://elibro.net/es/lc/uteq/titulos/230290>

- [26] H. R. Rincón, *Lo que callan los amantes del mañana: Descubre lo que hace la Inteligencia Artificial y el Metaverso en las personas y en las empresas*. Hipertexto, 2022. [Online]. Available: <https://books.google.com.ec/books?id=j6eSEAAAQBAJ>
- [27] M. Souza, *ESSENTIAL PYTHON: DE LA CIENCIA DE DATOS A LA AUTOMATIZACIÓN*. GAVEA LAB. [Online]. Available: [https://books.google.com.ec/books?id=Tx\\_QEAAAQBAJ](https://books.google.com.ec/books?id=Tx_QEAAAQBAJ)
- [28] Y. González Valle, D. Galpert, R. Molina-Ruiz, and G. Agüero-Chapin, “Integración de rasgos y aprendizaje semi-supervisado para la clasificación funcional de enzimas utilizando K-medias de Spark,” *Revista Cubana de Ciencias Informáticas*, vol. 14, no. 4, pp. 134–161, 2020.
- [29] aprende machinelearning, “Aprendizaje por Refuerzo,” Dec. 2020.
- [30] V. Lopez Fandino, *Sistemas de Big Data*. RA-MA Editorial, 2023. [Online]. Available: <https://elibro.net/es/lc/uteq/titulos/235054>
- [31] M. E. Mamani Ccalla, “Métodos de regularización LASSO y LARS para el modelo de regresión lineal general,” 2022.
- [32] R. W. Hoerl, “Ridge Regression: A Historical Context,” *Technometrics*, vol. 62, no. 4, pp. 420–425, Oct. 2020, doi: 10.1080/00401706.2020.1742207.
- [33] interactivechaos, “Gradient Boosting”.
- [34] Juan Ignacio Barrios Arce, “Light GBM vs XGBoost . ¿Cuál es mejor el algoritmo?”
- [35] Ley, “LEY ORGÁNICA DE PROTECCIÓN DE DATOS PERSONALES,” 2021. [Online]. Available: [www.lexis.com.ec](http://www.lexis.com.ec)
- [36] Ley, “LEY ORGANICA DE AGROBIODIVERSIDAD, SEMILLAS Y FOMENTO DE AGRICULTURA,” 2017. [Online]. Available: [www.lexis.com.ec](http://www.lexis.com.ec)

**CAPÍTULO VII**  
**ANEXOS**

## Anexo 1 Interfaz de la aplicación web



### Banano

El banano (*Musa Paradisiaca*), el cuarto alimento más importante y siendo la fruta más consumida a nivel mundial. El comercio internacional en muchos países está enfocándose en eliminar barreras comerciales, viéndolo como un motor clave para el desarrollo económico sostenido. Ecuador es destacado como el principal exportador en el mundo cubriendo un 29% el mercado internacional, actualmente abarca aproximadamente 200,000 hectáreas de terreno que se riegan y fumigan de manera continua para garantizar la cosecha semanal.

### Predicción



Semana del año.

7



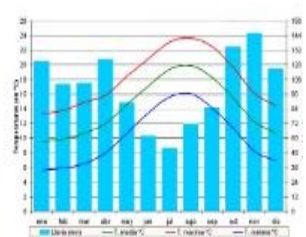
Número de frutas en el racimo.

9



Medida superior del grosor del fruto.

43



Temperatura mínima del aire.

22.4

Realizar Predicción

¡Predicción exitosa!

El peso estimado es: 87.12381544969885 libras

**Elaborado:** Jazmín Mora

## Anexo 2 Codificación de la página web

```

JAZMIN
├── static\img
│   ├── Banano_mano.png
│   ├── Calendario.png
│   ├── calibracion.png
│   ├── fci.png
│   ├── logo-uteq.png
│   ├── portada.jpg
│   └── Temperatura.png
├── templates
│   ├── index.html
│   └── app.py
└── modelo.joblib

app.py > ...
1 from flask import Flask, render_template, request
2 import joblib
3 import numpy as np
4
5 app = Flask(__name__)
6
7 # Cargar el modelo entrenado
8 modelo = joblib.load('modelo.joblib')
9
10 @app.route('/', methods=['GET', 'POST'])
11 def index():
12     if request.method == 'POST':
13         try:
14             # Obtener los datos ingresados desde el formulario
15             semana = float(request.form['semana'])
16             manos = float(request.form['manos'])
17             calibracion_superior = float(request.form['calibracion_superior'])
18             temperatura_min = float(request.form['temperatura_min'])
19
20             # Realizar la predicción utilizando el modelo cargado
21             datos = np.array([[semana, manos, calibracion_superior, temperatura_min]])
22             prediccion = modelo.predict(datos)
23
24             # Renderizar la plantilla 'index.html' con el resultado de la predicción
25             return render_template('index.html', prediccion=prediccion[0])
26
27         except ValueError:
28             # Si hay un error al convertir a float, mostrar un mensaje de error en la misma página
29             mensaje_error = "Por favor, ingrese valores numéricos en todos los campos."
30             return render_template('index.html', error=mensaje_error)
31
32     else:
33         # Si es una solicitud GET (primera visita a la página), simplemente renderiza el formulario
34         return render_template('index.html')
35
36 if __name__ == '__main__':
37     app.run(debug=True, port=5006)
```

Elaborado: Jazmín Mora

## Anexo 3 Index de la página web

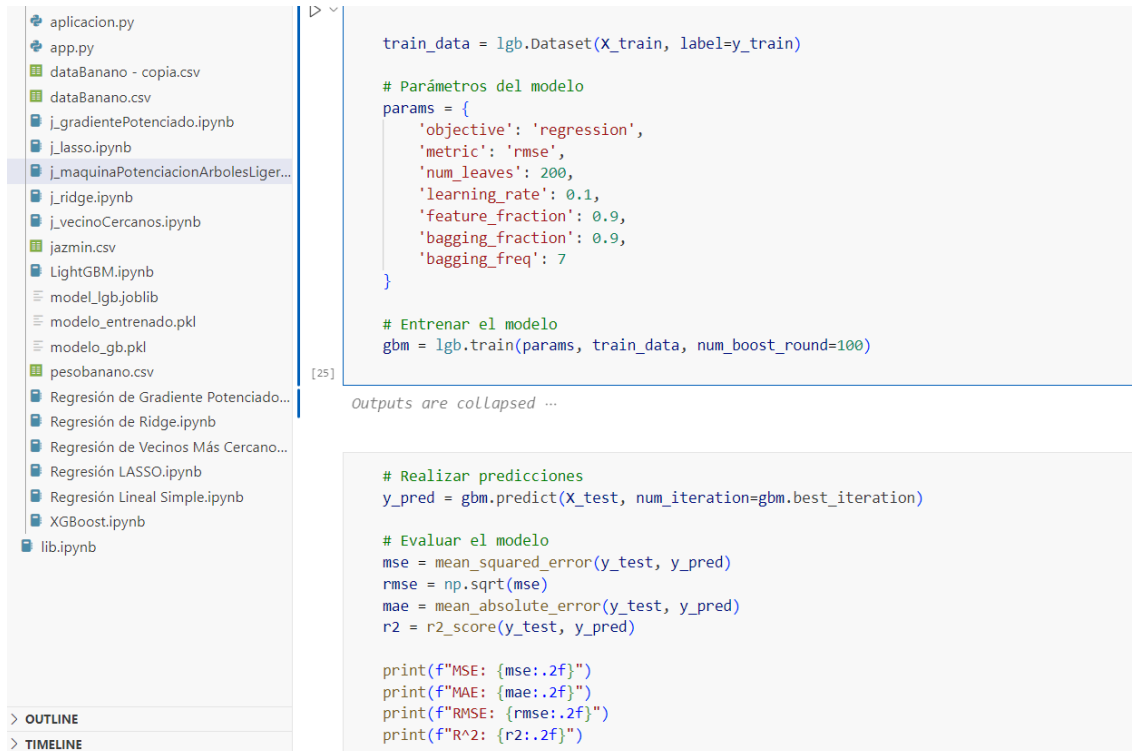
```

static\img
├── Banano_mano.png
├── Calendario.png
├── calibracion.png
├── fci.png
├── logo-uteq.png
├── portada.jpg
├── Temperatura.png
├── templates
│   ├── index.html
│   └── app.py
└── modelo.joblib

index.html
1 <html lang="es">
2 <body>
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47 <div class="container mt-4">
48 <hr>
49 <center>
50 <h3 style="color: #808080; font-family: Arial, sans-serif;"> <b>Predicción</b></h3>
51 </center>
52 <hr>
53
54 <form id="formulario" method="post">
55 <div class="row">
56 <!-- Descripción e imagen 1 -->
57 <div class="col-md-3">
58 <figure class="figure">
59 
61 <center>
62 <figcaption class="figure-caption">Semana del año.</figcaption>
63 </center>
64 </figure>
65 <input type="text" name="semana" class="form-control" placeholder="Semana">
66 </div>
67 <!-- Descripción e imagen 2 -->
68 <div class="col-md-3">
69 <figure class="figure">
70 
72 <center>
73 <figcaption class="figure-caption">Número de frutas en el racimo.</figcaption>
74 </center>
75 </figure>
76 <input type="text" name="manos" class="form-control" placeholder="Manos">
77 </div>
78 </div>
79
80 </form>
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
```

Elaborado: Jazmín Mora

## Anexo 4 Codificación del algoritmo entrenado



```
aplicacion.py
app.py
dataBanano - copia.csv
dataBanano.csv
j_gradientePotenciado.ipynb
j_lasso.ipynb
j_maquinaPotenciacionArbolesLiger...
j_ridge.ipynb
j_vecinoCercanos.ipynb
jazmin.csv
LightGBM.ipynb
model_lgb.joblib
modelo_entrenado.pkl
modelo_gb.pkl
pesobanano.csv
Regresión de Gradiente Potenciado...
Regresión de Ridge.ipynb
Regresión de Vecinos Más Cercano...
Regresión LASSO.ipynb
Regresión Lineal Simple.ipynb
XGBoost.ipynb
lib.ipynb
> OUTLINE
> TIMELINE

train_data = lgb.Dataset(X_train, label=y_train)

# Parámetros del modelo
params = {
    'objective': 'regression',
    'metric': 'rmse',
    'num_leaves': 200,
    'learning_rate': 0.1,
    'feature_fraction': 0.9,
    'bagging_fraction': 0.9,
    'bagging_freq': 7
}

# Entrenar el modelo
gbm = lgb.train(params, train_data, num_boost_round=100)

Outputs are collapsed ...

# Realizar predicciones
y_pred = gbm.predict(X_test, num_iteration=gbm.best_iteration)

# Evaluar el modelo
mse = mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)
mae = mean_absolute_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"MSE: {mse:.2f}")
print(f"MAE: {mae:.2f}")
print(f"RMSE: {rmse:.2f}")
print(f"R^2: {r2:.2f}")
```

Elaborado: Jazmín Mora