



UNIVERSIDAD AGRARIA DEL ECUADOR
FACULTAD DE CIENCIAS AGRARIAS

CARRERA COMPUTACIÓN

**TRABAJO DE TITULACIÓN COMO REQUISITO PREVIO PARA LA
OBTENCIÓN DEL TÍTULO DE
INGENIERO EN CIENCIAS DE LA COMPUTACIÓN**

**APLICATIVO DE PREDICCIÓN MEDIANTE MACHINE
LEARNING DE LA PRODUCCIÓN DE BANANO EN LA
HACIENDA SAN JUAN**

AUTOR

ANCHUNDIA ZAMBRANO ANDRES DAVID

TUTOR

VASQUEZ BERMUDEZ MITCHELL JOHN

GUAYAQUIL, ECUADOR

2026



UNIVERSIDAD AGRARIA DEL ECUADOR

FACULTAD DE CIENCIAS AGRARIAS

CARRERA COMPUTACIÓN

APROBACIÓN DEL TUTOR

El suscrito, docente de la Universidad Agraria del Ecuador, en mi calidad de Tutor, certifico que el presente trabajo de titulación: **APLICATIVO DE PREDICCIÓN MEDIANTE MACHINE LEARNING DE LA PRODUCCIÓN DE BANANO EN LA HACIENDA SAN JUAN**, realizado por el estudiante **ANCHUNDIA ZAMBRANO ANDRES DAVID**, con cédula de identidad N° **0953452844** de la **CARRERA DE COMPUTACIÓN**, Campus Guayaquil, ha sido orientado y revisado durante su ejecución; y cumple con los requisitos técnicos y legales existidos por la Universidad Agraria del Ecuador; por lo tanto, se aprueba la presentación del mismo.

Atentamente,

Ing. Vásquez Bermúdez Mitchell Jhon, M.Sc.

Guayaquil, 21 de Abril del 2026



**UNIVERSIDAD AGRARIA DEL ECUADOR
FACULTAD DE CIENCIAS AGRARIAS
CARRERA COMPUTACIÓN**

APROBACIÓN DEL TRIBUNAL DE SUSTENTACIÓN

Los abajo firmantes, docentes designados por el H. Consejo Directivo como miembros del Tribunal de Sustentación, aprobamos la defensa del trabajo de titulación: “**APLICATIVO DE PREDICCIÓN MEDIANTE MACHINE LEARNING DE LA PRODUCCIÓN DE BANANO EN LA HACIENDA SAN JUAN**”, realizado por el estudiante **ANCHUNDIA ZAMBRANO ANDRES DAVID**, el mismo que cumple con los requisitos exigidos por la Universidad Agraria del Ecuador.

Atentamente,

ING. Maritza Aguirre Munizaga, M.Sc.
PRESIDENTE

ING. Carlota Delgado Vera, M.Sc.
EXAMINADOR PRINCIPAL

ING. Irene Vásquez Villacis, M.Sc.
EXAMINADOR PRINCIPAL

ING. VÁSQUEZ MITCHELL, M.Sc.
EXAMINADOR SUPLENTE

Guayaquil, 21 de abril del 2026

DEDICATORIA

Dedico esta tesis con todo mi corazón y cariño a mi padre, hermana, abuela, tías y prima, porque son testigos del esfuerzo y empeño que puse a lo largo de la carrera; para ellos, este logro también es suyo. En especial a mi papá, Roberto Anchundia, por ser quien me enseñó a luchar por lo que quiero y por demostrarme que siempre estará a mi lado, apoyándome e impulsándome a seguir adelante y cumplir mis sueños. A mi hermana, Andrea Anchundia, quien me acompañó en noches largas y por ser la hermana mayor que está a mi lado en todo momento, siendo mi ejemplo a seguir, y esperando algún día también poder ser su ejemplo. A mi abuela, tías y prima por cuidarme siempre y aconsejarme sobre todo lo bueno que tiene estudiar y sobre la vida. Y por supuesto a Coffe y Benji, mis dos mejores amigos de cuatro patas, quienes durmieron a mi lado en las noches largas. Por último, y no menos importante, quiero agradecerme a mí mismo por creer en mí y por no rendirme pese a las noches largas, los obstáculos y los momentos difíciles que pasé durante estos cinco años, por poner la carrera como mi prioridad, porque soy el proyecto de vida más valioso que poseo. El yo pequeño está orgulloso de ti, Andrés.

AGRADECIMIENTO

Quiero agradecer a toda mi familia en general por su incontable apoyo, soporte y por ser mis mentores durante todo el camino de mi carrera. Un agradecimiento de corazón a todas mis amistades, porque nada de lo que hagas en esta vida será legendario si tus mejores amigos no están allí para verlo. Agradecerle a Chiri, Ricky, Miguel, Suying Pablo, Danna y Marck, por alegrarme los fines de semana, ya sea viajando, comiendo o viendo películas. Un agradecimiento especial a Max y Karim, con quienes comparto amistad desde la escuela y quienes representan una de las amistades más bonitas que los estudios me pudieron dar, y que sé que siempre me apoyarán.

A mis compañeros y amigos de la universidad, Elías, Jenniffer y Wilmer, porque la universidad no pudo darme mejores compañeros para caminar juntos estos cinco años llenos de momentos increíbles, amargos y felices. Agradecerle a mi tutor de tesis, el ingeniero Mitchell Vásquez, por haberme propuesto el tema y guiarme durante todo el proceso que el desarrollo de este conllevó.

Y, por último, agradecerles a todas las personas que han pasado un ratito por mi vida, porque al final somos una mezcla de todas las personas que han sido parte de nuestro camino. Me alegra saber que algunas de ellas son parte esencial de quien soy hoy.

AUTORIZACIÓN DE AUTORÍA INTELECTUAL

Yo ANCHUNDIA ZAMBRANO ANDRES DAVID, en calidad de autor del proyecto realizado, sobre “APLICATIVO DE PREDICCIÓN MEDIANTE MACHINE LEARNING DE LA PRODUCCIÓN DE BANANO EN LA HACIENDA SAN JUAN” para optar el título de INGENIERO EN CIENCIAS DE LA COMPUTACIÓN, por la presente autorizo a la UNIVERSIDAD AGRARIA DEL ECUADOR, hacer uso de todos los contenidos que nos pertenecen o parte de los que contienen esta obra, con fines estrictamente académicos o de investigación.

Los derechos que como autores nos correspondan, con excepción de la presente autorización, seguirán vigentes a nuestro favor, de conformidad con lo establecido en los artículos 5, 6, 8; 19 y demás pertinentes de la Ley de Propiedad Intelectual y su Reglamento.

Guayaquil, 21 de Abril del 2026

ANCHUNDIA ZAMBRANO ANDRES DAVID
C.I. 0953452844

RESUMEN

La producción de banano en la Hacienda San Juan enfrenta desafíos críticos debido a la estimación empírica de la cosecha y la vulnerabilidad ante la variabilidad climática, lo que genera incertidumbre en la planificación y pérdidas económicas. Esta investigación tuvo como objetivo desarrollar un aplicativo web que integra modelos de Machine Learning para predecir la producción semanal y optimizar la toma de decisiones. Metodológicamente, se aplicó un enfoque cuantitativo utilizando el proceso KDD para el análisis de patrones en registros históricos meteorológicos y edáficos, junto con la metodología XP para el desarrollo ágil del software. Se implementaron algoritmos predictivos considerando variables determinantes identificadas, como la temperatura media, precipitación y los niveles de nutrientes en el suelo. Como resultado, se obtuvo una plataforma funcional con módulos de captura de datos, predicción, visualización y reportes, validada mediante pruebas de funcionalidad y usabilidad que alcanzaron una satisfacción eficiente por parte del administrador. Se concluye que el aplicativo reduce el margen de error en las estimaciones, transformando la gestión tradicional en un proceso técnico basado en datos, lo que garantiza una mayor eficiencia operativa y rentabilidad para la hacienda.

Palabras clave: *Aplicativo web, Banano, KDD, Machine Learning, Predicción.*

ABSTRACT

Banana production at Hacienda San Juan faces critical challenges due to empirical harvest estimates and vulnerability to climate variability, which creates uncertainty in planning and economic losses. The objective of this research was to develop a web application that integrates machine learning models to predict weekly production and optimize decision-making. Methodologically, a quantitative approach was applied using the KDD process for pattern analysis in historical meteorological and edaphic records, together with the XP methodology for agile software development. Predictive algorithms were implemented considering identified determining variables, such as average temperature, precipitation, and soil nutrient levels. As a result, a functional platform was obtained with data capture, prediction, visualization, and reporting modules, validated through functionality and usability tests that achieved efficient satisfaction on the part of the administrator. It is concluded that the application reduces the margin of error in estimates, transforming traditional management into a data-based technical process, which guarantees greater operational efficiency and profitability for the farm.

Keywords: *Banana, Machine Learning, Prediction, KDD, Web application.*

ÍNDICE CONTENIDO

APROBACIÓN DEL TUTOR	II
APROBACIÓN DEL TRIBUNAL DE SUSTENTACIÓN	III
DEDICATORIA	IV
AGRADECIMIENTO	V
AUTORIZACIÓN DE AUTORÍA INTELECTUAL.....	VI
RESUMEN	VII
ABSTRACT.....	VIII
ÍNDICE DE ANEXOS	XIII
ÍNDICE DE APÉNDICE.....	XIV
1. INTRODUCCIÓN	13
1.1 Antecedentes del problema	13
1.2 Planteamiento y formulación del problema.....	14
1.2.1 Planteamiento del problema	14
1.2.2 Formulación del problema	15
1.3 Justificación de la investigación.....	15
1.3.1 Módulo de procesamiento de datos:.....	16
1.3.1.1. Submódulo de Captura de Datos	16
1.3.1.2. Submódulo de Procesamiento de Datos	16
1.3.1.3. Submódulo de Almacenamiento y Gestión de Datos	17
1.3.1.4. Submódulo de Modelado y Predicción	17
1.3.1.5. Submódulo de Evaluación del Modelo	17
1.3.2 Módulo de visualización.....	17
1.3.3 Módulo de Gestión y Administración	18
1.3.4 Módulo de Reportes	18
1.4 Delimitación de la investigación.....	18
1.5 Objetivo general	18
1.6 Objetivos específicos.....	19
2. MARCO TEÓRICO.....	20
2.1 Estado del arte	20
2.2 Bases teóricas	23
2.2.1 El origen del banano.....	23
2.2.2 Cultivo de banano en el Ecuador.....	24

2.2.3 Producción de banano.....	24
2.2.4 Agrometeorología	25
2.2.5 Requerimientos de nutrientes en el suelo en el cultivo de banano	26
2.2.6 Sistema gestor de bases de datos	27
2.2.7 Lenguaje PHP 8.3.....	27
2.2.8 HTML 5.3.....	28
2.2.9 Cascading Style Sheet (CSS).....	28
2.2.10 JavaScript.....	29
2.2.11 Python	29
2.2.12 My Structured Query Language (MySQL)	29
2.2.13 Machine Learning	30
2.2.13.1. Regresión Lineal	31
2.2.13.2. Random Forest	31
2.2.13.3. SVM	32
2.2.13.4. XGBoost.....	32
2.2.14 Bootstrap	33
2.2.15 Hosting	34
2.2.16 Estructura del software	34
2.2.16.1. Diagrama de contexto	34
2.2.16.2. Diagrama DFD nivel 1	35
2.2.17 Diagramas UML.....	35
2.2.17.1. Diagrama de caso de uso.	35
2.2.17.2. Diagrama entidad relación.	35
2.3 Marco Legal.....	36
3. MATERIALES Y MÉTODOS.....	40
3.1 Enfoque de la investigación.....	40
3.1.1 Tipo y alcance de la investigación	40
3.1.1.1. Investigación aplicada.....	40
3.1.1.2. Investigación correlacional	41
3.1.2 Diseño de investigación.....	41
3.2 Metodología	41
3.2.1 Metodología XP	42
3.2.1.1. Fase de planificación.....	42

3.2.1.2. Fase de diseño.....	43
3.2.1.3. Fase de codificación.....	45
3.2.1.4. Fase de pruebas	47
3.2.1.5. Fase de lanzamiento	47
3.2.2 Metodología KDD	48
3.2.2.1. Selección de datos	48
3.2.2.2. Preprocesamiento	49
3.2.2.3. Transformación	49
3.2.2.4. Minería de datos	50
3.2.2.5. Evaluación e Interpretación.....	51
3.2.3 Recolección de datos	52
3.2.3.1. Recursos	52
3.2.3.2. Presupuesto del proyecto.....	52
3.2.4 Métodos y técnicas de investigación	53
3.2.5 Población y muestra	54
3.2.5.1. Población	54
3.2.5.2. Muestra	54
3.2.6 Análisis estadístico	55
4. RESULTADOS.....	58
4.1 Análisis los factores ambientales y del suelo que influyen en la producción de banano en la Hacienda San Juan para la identificación de patrones que sirvan como base de los modelos predictivos.	58
4.1.1 Factores ambientales relevantes.....	59
4.1.2 Factores del suelo relevantes.....	60
4.2 Diseño el modelado de la producción del cultivo de banano a través de los diagramas UML para la correcta elaboración de todos los módulos del software.	
61	
4.3 Desarrollo un aplicativo web que integre modelos de Machine Learning con el propósito de predecir la producción de banano, facilitando el control y la planificación del cultivo.....	62
5. DISCUSIÓN.....	64
6. CONCLUSIONES Y RECOMENDACIONES	67
6.1 Conclusiones.....	67

6.2 Recomendaciones	68
BIBLIOGRAFÍA	70
ANEXOS	75
APÉNDICES	134

ÍNDICE DE ANEXOS

Anexo N° 1: Aplicaciones relacionadas al tema	75
Anexo N° 2: Fases de Metodología	76
Anexo N° 3: Funcionamiento del aplicativo	77
Anexo N° 4: Recursos	78
Anexo N° 5: Modelo de entrevista	80
Anexo N° 6: Encuesta de satisfacción.....	81
Anexo N° 7: Análisis de la entrevista al administrador	83
Anexo N° 8: Análisis de la encuesta de satisfacción	87
Anexo N° 9: Glosario LEL	92
Anexo N° 10: Requerimientos funcionales y no funcionales, Historia de Usuario	96
Anexo N° 11: Diagrama de Entidad Relación.....	105
Anexo N° 12: Diccionario de datos.....	105
Anexo N° 13 Evaluación de modelos de machine learning	111
Anexo N° 14: Diagrama de flujo de datos	112
Anexo N° 15: Esquema modular del sistema	113
Anexo N° 16: Esquema modular de la arquitectura de la aplicación	113
Anexo N° 17 Correlación entre variables con producción	114
Anexo N° 18: Diagramas de caso de uso.....	115
Anexo N° 19: Diagrama de secuencia del aplicativo	123
Anexo N° 20: Fotos entrevista.....	124
Anexo N° 20: Pruebas del software – Funcionalidad y usabilidad.....	125
Anexo N° 22: Lugar de implementación	131

ÍNDICE DE APÉNDICE

Apéndice N°1: Manual de Usuario – Aplicativo Web.....	134
Apéndice N°2: Manual de Técnico.....	143

1. INTRODUCCIÓN

1.1 Antecedentes del problema

La digitalización ha transformado por completo las dinámicas del trabajo de campo, convirtiendo la estimación de cosechas en un pilar estratégico para el productor. Si bien inicialmente se dependía de modelos estadísticos bastante elementales, el avance de la informática ha dado paso a sistemas mucho más robustos. Actualmente, el uso de machine learning permite procesar volúmenes masivos de datos para prever resultados con una exactitud que antes era impensable, facilitando decisiones mucho más acertadas en la gestión diaria de los cultivos. (Santiago y Torres, 2024)

La industria bananera no solo es un motor económico para Ecuador, sino también el sustento de miles de familias. Ante este panorama, la implementación de modelos predictivos que consideren variables como la radiación, la temperatura y las propiedades del suelo representa un avance estratégico. Estas tecnologías permiten identificar patrones críticos que afectan el rendimiento, facilitando una planificación integral desde la siembra hasta la comercialización. En la Hacienda San Juan, este enfoque se concreta en el desarrollo de una herramienta de machine learning diseñada para estimar la producción con mayor exactitud, proporcionando un soporte técnico esencial para optimizar la toma de decisiones y la logística de distribución.

En la actualidad la hacienda bananera “San Juan” de la empresa CAMEONCE S.A cuenta con 107 hectáreas para cultivar banano y se encuentra ubicada en la Provincia de Guayas perteneciente al cantón Naranjal e inicio sus actividades al 18 de marzo de 2010 con la actividad principal de producción de bananos con calidad de exportación, además posee 33 trabajadores, que se encargan del sembrado de dos tipos de variedad de banano que son William y Valery de origen meristemo, lo cual conlleva que estas plantaciones requieran de cuidados exigentes, ya que si estas no mantienen un control adecuado y seguimiento a diario este sembrío podría generar aumento dentro de los costos de producción.

El rendimiento de las plantaciones bananeras no es producto del azar, sino de una compleja interacción entre factores ambientales y las condiciones del suelo. En un país como Ecuador, donde este cultivo es un motor vital para la

economía, entender ese vínculo es esencial. Por ello, la integración de herramientas digitales y el análisis de datos surge como una vía clave para descifrar cómo estas variables afectan la productividad. Al cruzar registros meteorológicos con la composición del terreno, es posible desarrollar modelos de predicción que no solo estiman la cosecha, sino que permiten a los productores planificar con mayor precisión y cumplir con las exigentes normativas de los mercados globales.

1.2 Planteamiento y formulación del problema

1.2.1 Planteamiento del problema

En la actualidad, la planificación en la Hacienda San Juan sigue dependiendo casi totalmente del 'ojo clínico' de sus administradores. Y aunque esa experiencia vale oro, trabajar sin respaldo tecnológico suele pasar factura: los cálculos fallan y eso complica desde los tratos comerciales hasta la logística de los envíos. El problema se agrava cuando entra en juego el clima. Al no tener un sistema que avise si viene una ola de calor o un aguacero fuerte, la finca se queda sin margen de maniobra. Básicamente, solo pueden reaccionar cuando el daño ya está hecho, lo que siempre termina en pérdidas económicas.

A esto hay que sumarle el problema del suelo. Como no hay un monitoreo constante, el riego y los fertilizantes se aplican con un margen de error demasiado grande. Básicamente, si no tienes datos exactos sobre el pH o los nutrientes, estás usando insumos a ciegas; esto no solo dispara los costos, sino que desgasta la plantación a largo plazo. Otro gran desafío es equilibrar la oferta y la demanda. Sin una herramienta que te diga cuánto vas a cosechar realmente, la finca vive en dos extremos: o desperdicia fruta porque sobró, o queda mal con los clientes porque faltó stock. Hoy en día, depender solo del instinto para evitar esto ya no alcanza en un mercado tan competitivo.

Implementar este sistema marca un punto de inflexión para la Hacienda San Juan. Básicamente, es el paso definitivo para dejar de administrar por costumbre y empezar a decidir con datos en la mano. Al respaldar cada movimiento con información precisa, el objetivo es claro: que el bananal no solo produzca más, sino que sea más rentable y sostenible. Lo interesante de esta transformación es cómo cambia el juego de la información. Al cruzar el historial de la finca con lo que pasa en tiempo real, el equipo técnico puede tomarle el pulso al cultivo minuto a minuto. Esto va más allá de apagar incendios del día a

día; nos da la capacidad analítica para ver venir las tendencias y anticiparnos a los problemas antes de que golpeen.

1.2.2 Formulación del problema

¿De qué manera un aplicativo basado en Machine Learning permitiría predecir la producción de banano en la Hacienda “San Juan” a partir del análisis de variables climáticas y del suelo?

1.3 Justificación de la investigación

La planificación en la Hacienda San Juan representa un desafío constante, pues la producción de banano se ve condicionada por una compleja red de factores climáticos y edáficos. Operar sin herramientas que permitan anticipar el rendimiento de las cosechas no solo dificulta la gestión, sino que expone a la finca a ineficiencias y pérdida de recursos valiosos. Ante este escenario, se vuelve imperativo adoptar sistemas que transformen los datos en bruto en proyecciones útiles.

Bajo esta premisa, la integración de modelos de Machine Learning surge como una vía estratégica para descifrar variables meteorológicas y del suelo, reduciendo la incertidumbre en los pronósticos. Este enfoque justifica la investigación al demostrar que una administración respaldada por datos precisos es la clave para elevar los niveles de eficiencia en el manejo agrícola.

Este proyecto busca implementar herramientas tecnológicas para la gestión de la producción en la hacienda, eliminando la incertidumbre y optimizando el uso de los recursos. Al basarse en datos y predicciones, permitirá una toma de decisiones más estratégica, garantizando estabilidad y mayor rentabilidad para la empresa. Además, el uso de un sistema de predicción otorgaría a la hacienda San Juan una ventaja competitiva en comparación a otras haciendas bananeras o posicionarse como pionera en incorporar tecnologías que mejoren la productividad lo que fortalece la presencia de los mercados nacionales e internacionales.

Bajo esta perspectiva, el proyecto plantea el desarrollo de una plataforma web multiplataforma diseñada para transformar la gestión en la Hacienda San Juan a través del aprendizaje automático. El núcleo del sistema consiste en pronosticar la cosecha de banano procesando una robusta base de datos que incluye cerca de 468 registros históricos semanales de producción, recolectados a lo largo de 8 años, además de datos diarios de los últimos 5 años de variables

climáticas y edáficas. La solución no solo automatiza la captura de información, sino que ofrece un entorno estructurado donde los modelos de predicción se mantienen actualizados. Con niveles de acceso diferenciados y reportes analíticos detallados, la herramienta busca sustituir la incertidumbre por decisiones respaldadas en evidencia, promoviendo una actividad agrícola más rentable y alineada con criterios de sostenibilidad.

Los módulos que formaran parte del sistema son:

1.3.1 Módulo de procesamiento de datos:

En este módulo se realizan todos los procesos de procesamiento de datos para la realización del modelo de predicción de producción de banano de la hacienda, a continuación, se especifican los submódulos:

1.3.1.1.Submódulo de Captura de Datos

La función de este módulo es asegurar la ingesta constante de la 'materia prima' analítica. Su diseño no discrimina fuentes: centraliza variables críticas del banano, desde factores climáticos (temperatura, lluvia) y perfiles de suelo, hasta el registro histórico de rendimientos. Para garantizar la cobertura, se optó por una mecánica de captura híbrida. Por un lado, automatiza el flujo masivo mediante sensores y telemetría satelital; por otro, mantiene el factor humano a través de formularios técnicos para registros manuales específicos. Esta redundancia es vital: partimos de la premisa de que la puntería del algoritmo predictivo será directamente proporcional a la frescura y calidad de estos datos de entrada.

1.3.1.2.Submódulo de Procesamiento de Datos

Antes de alimentar cualquier algoritmo, la información cruda debe someterse a un riguroso proceso de sanitización. Este submódulo no se limita a corregir errores; actúa como un filtro de calidad que neutraliza el 'ruido estadístico', imputando valores faltantes y normalizando escalas para evitar sesgos. Pero la limpieza es solo el primer paso. Para maximizar la eficiencia, se aplican técnicas de reducción de dimensionalidad, destilando el dataset para conservar únicamente las variables con verdadero peso predictivo. Finalmente, el sistema segmenta esta información ya refinada en lotes de entrenamiento, validación y prueba, entregando al modelo una estructura homogénea lista para el aprendizaje.

1.3.1.3.Submódulo de Almacenamiento y Gestión de Datos

En este módulo como tal se almacenan y organizan los datos que son recolectados en si utilizando una base de datos estructurada que como tal garantice la integridad, la seguridad y acceso eficiente, y se integra con mecanismos como tal para la limpieza y transformación de los datos lo que permitiendo su preparación para ser usados en los modelos de predicción, y por lo que además este módulo facilita consultas históricas como tal actualizaciones periódicas y respaldo automático de la información.

1.3.1.4.Submódulo de Modelado y Predicción

El submódulo de modelado y predicción es el núcleo del sistema, utilizando algoritmos de Machine Learning como Xgboost , Regresión Lineal, Random Forest y Máquinas de Vectores de Soporte para Regresión (SVR) para predecir la producción semanal de banano (cajas procesadas) en la Hacienda San Juan. Estos algoritmos, entrenados con datos históricos de variables meteorológicas (temperatura, precipitación, humedad, pH del suelo, nutrientes) y productivas (enfundes, racimos), operan bajo un enfoque de aprendizaje supervisado.

1.3.1.5.Submódulo de Evaluación del Modelo

La confiabilidad del sistema no se asume; se audita. Este módulo somete a cada algoritmo a un 'examen de estrés' estadístico, triangulando métricas de desviación (MAE, MSE) con el coeficiente de determinación (R^2). El objetivo no es solo obtener un número, sino diagnosticar qué tan bien replica el modelo la realidad. Para ello, la validación cruzada (cross-validation) resulta fundamental: pone a competir distintos enfoques algorítmicos frente al historial de la finca, descartando aquellos que memorizan datos en lugar de aprender patrones. Solo el modelo que sobrevive a este escrutinio matemático es el que finalmente llega a la mesa de decisiones de la Hacienda San Juan.

1.3.2 Módulo de visualización

La función central de este módulo es transformar los datos técnicos en conocimiento visual e interactivo. Mediante el uso de herramientas como mapas de calor, tableros comparativos y gráficos de líneas, el sistema ofrece una lectura clara de cómo ha evolucionado la producción bananera hasta el día de hoy. Esta interfaz no solo simplifica la lectura de cifras complejas, sino que permite

visualizar de forma directa los pronósticos del machine learning, ayudando a los supervisores a evaluar posibles escenarios y decidir el rumbo de la cosecha con mayor seguridad. Al estar vinculado en tiempo real con la base de datos, el módulo garantiza que cada gráfico refleje la realidad actual de la finca, manteniendo una trazabilidad total desde el registro inicial hasta su visualización final.

1.3.3 Módulo de Gestión y Administración

Este módulo permite el control del sistema por parte de los administradores. Incluirá funcionalidades como la gestión de cuentas, configuración de parámetros del sistema y control de acceso. También se puede gestionar alertas automáticas o programaciones personalizadas para la generación de informes y actualizaciones.

1.3.4 Módulo de Reportes

A través de este módulo se generan reportes analíticos y visuales que mostrarán los resultados de las predicciones, así como estadísticas relevantes para la toma de decisiones. Los reportes se pueden exportar en formatos como PDF o Excel, y presentan comparaciones entre producción estimada y real, tendencias históricas y recomendaciones para la optimización del cultivo. Este módulo también contribuye a la transparencia y trazabilidad de los procesos de gestión agrícola.

1.4 Delimitación de la investigación

- **Espacio:** El proyecto se realizó en la hacienda San Juan de la empresa CAMEONCE S.A. la cual se encuentra ubicada en la Cantón Naranjal de la provincia del Guayas.
- **Tiempo:** El desarrollo del presente proyecto se llevó a cabo en 10 meses.
- **Población:** El proyecto fue dirigido para los 2 administradores de la hacienda San Juan de la empresa CAMEONCE S.A.

1.5 Objetivo general

Implementar un aplicativo de predicción de la producción de banano, mediante el uso de machine learning considerando factores ambientales y de suelo, para una mejor toma de decisiones y la eficiencia en la gestión agrícola en la Hacienda San Juan.

1.6 Objetivos específicos

- Analizar los factores ambientales y del suelo que influyen en la producción de banano en la Hacienda San Juan para la identificación de patrones que sirvan como base de los modelos predictivos.
- Diseñar el modelado de la producción del cultivo de banano a través de los diagramas UML para la correcta elaboración de todos los módulos del software.
- Desarrollar un aplicativo web que integre modelos de Machine Learning con el propósito de predecir la producción de banano, facilitando el control y la planificación del cultivo.

2. MARCO TEÓRICO

2.1 Estado del arte

Con el fin de mantener su competitividad global, el sector bananero en Ecuador ha empezado a integrar sistemas inteligentes que profesionalizan el manejo de las plantaciones. Antecedentes internacionales, como el trabajo de Sanga et al. (2020) en Tanzania, demuestran el potencial de esta transición. Estos investigadores desarrollaron un aplicativo para Android que, apoyado en modelos de aprendizaje profundo, logra diagnosticar la Sigatoka negra y otras plagas con una exactitud casi total. El impacto de esta solución radica en su accesibilidad: permite que el agricultor tome decisiones críticas simplemente capturando una imagen del cultivo, lo que no solo optimiza la vigilancia fitosanitaria, sino que previene el desperdicio de recursos por diagnósticos tardíos. (Ver Figura 1).

El uso de una aplicación móvil basada en aprendizaje profundo evidenció que es posible mejorar el manejo fitosanitario del banano incluso en condiciones restrictivas. Al alcanzar una precisión del 99% en la identificación de plagas clave, la herramienta facilitó que los agricultores actuaran con rapidez antes de que la producción se viera comprometida. Diseñado específicamente para entornos Android de bajo costo, este sistema rompe la brecha tecnológica al ofrecer una alternativa económica y robusta. Al final, el proyecto se posiciona como una solución estratégica para comunidades rurales, probando que el análisis de datos es un aliado fundamental para la resiliencia y competitividad del sector agrícola.

La investigación difundida por la Revista Cumbres resalta cómo la integración de sistemas móviles en la supervisión bananera supone un avance clave para la tecnificación agrícola. Caraguay y Cordero (2020) subrayan que una de las mayores fortalezas de esta propuesta es su viabilidad económica, pues el uso de hardware de bajo costo facilita que un sector más amplio de productores acceda a estas herramientas. El sistema resulta de gran utilidad durante el corte y el embalaje, etapas donde la captura de datos precisos por lote ayuda a elevar los estándares de calidad. Al centralizar los pormenores de cada racimo en una base de datos, la gestión productiva se vuelve más ágil y fundamentada en registros reales.

Posteriormente, dicha información es procesada y presentada en reportes que identifican anomalías derivadas de la presencia de plagas o de prácticas operativas inadecuadas. Por ende, se puede mejorar el cultivo con la implementación de este software debido a que el problema que presentan los productores medianos en el Ecuador, es el daño que tienen los racimos en las fincas y como es por lote produce pérdidas, ya que este tipo de cosecha se debe estar en constante supervisión y los agricultores no pueden permanecer pendiente las 24 horas del día.

En el panorama latinoamericano, sobresale el impacto de AgroTIC, una aplicación desarrollada en la región de Santander, Colombia. Según explican Hinojosa et al. (2023), esta plataforma logra articular la inteligencia artificial con una red de comunicación que vincula a productores, técnicos y comerciantes, optimizando tanto el seguimiento de los sembríos como las etapas de venta. Su despliegue en más de 130 fincas benefició a más de 200 agricultores, quienes reportaron una mejora sustancial en el acceso a datos críticos y en la agilidad de su toma de decisiones. Este caso subraya el potencial de las herramientas móviles colaborativas para modernizar las economías rurales y profesionalizar la labor del pequeño y mediano productor.

Por otra parte, existe una aplicación móvil conocida con el nombre de Plantix utilizada a nivel mundial por los agricultores debido a que aumenta el rendimiento en los cultivos solamente tomando una foto y subiéndola el sistema puede identificar, si la planta está afectada por algún tipo de plaga o enfermedad y de esta manera se reduce pérdidas en la producción agrícola. Esto es posible, porque la base de datos posee información de más de 30 cultivos principales en el mundo y por eso detecta las anomalías en las plantaciones con facilidad y a su vez se le considera un doctor de estos sembríos porque recomienda tratamientos de manera orgánica o química según el tipo de plaga o enfermedad (Urquijo, 2020).

Entonces, Plantix ofrece mucha utilidad a grandes y pequeños agricultores porque detecta los daños en los cultivos y además tiene otras funcionalidades para la gestión agrícola como el cálculo de fertilizantes y una comunidad que responden a las dudas e inquietudes en tiempo real de los usuarios. También, está disponible en 18 idiomas y permite conocer en qué área específica está afectada con el fin de aprovechar al máximo las medidas que

indican este sistema en la producción de un cultivo específico. Por lo tanto, se ha impulsado el desarrollo de varias plataformas móviles hacia el sector agrícola que hoy en día están siendo utilizados por una gran cantidad de empresas. (Ver Figura 3)

Dentro del ámbito local, específicamente en el cantón Balao, la investigación de García et al. (2022) pone sobre la mesa un problema crítico: el deterioro de los suelos bananeros debido a la explotación intensiva. Sus hallazgos advierten que las prácticas actuales están mermando la calidad química y biológica del terreno, incrementando los residuos químicos y reduciendo la materia orgánica. Esta degradación no solo compromete la sostenibilidad del cultivo, sino que resalta la urgencia de adoptar tecnologías de monitoreo y métodos más equilibrados. Este estudio fundamenta el núcleo de la presente tesis, ya que confirma que vigilar las variables edafoclimáticas es indispensable para anticipar y optimizar los niveles de rendimiento en el sector.

La tecnificación del sector agrícola ecuatoriano se refleja en la creciente demanda de herramientas digitales que optimicen la administración de las fincas. Bajo esta premisa, la investigación de Cruz y Pillco (2020) en la hacienda “San Jacinto” ilustra cómo el software a medida puede transformar la rentabilidad del banano. Su propuesta consistió en un sistema integral capaz de administrar gastos, roles de pago y la trazabilidad de la fruta en cada etapa productiva. La implementación destacó por el uso del modelo RUP, asegurando un desarrollo estructurado que eliminó la dependencia de registros manuales. El resultado final fue una gestión administrativa más transparente y eficiente, donde la disponibilidad de datos actualizados permitió mitigar los problemas relacionados con los costos operativos y la toma de decisiones estratégicas.

Como resultado se obtuvo el desarrollo e implementación exitosamente el sistema web SAN JACINTO, el cual automatizó los procesos previamente realizados de forma manual en la hacienda bananera, optimizando la gestión interna y mejorando la atención a los clientes. Gracias a la aplicación de encuestas y entrevistas al personal de la hacienda, se identificaron las deficiencias en el control de costos de producción, lo que permitió diseñar módulos funcionales para la gestión de compras, ventas, roles de pago, control de empleados, seguimiento de producción y administración de gastos. La metodología aplicada facilitó un desarrollo estructurado en cuatro etapas,

garantizando la eficacia del sistema y satisfaciendo las expectativas del propietario, como se evidenció en los casos de prueba. Esta solución tecnológica se adapta a las tendencias actuales, permitiendo un flujo ágil de información y una gestión eficiente en tiempos óptimos. (Ver Anexo N°1).

En la Universidad Agraria del Ecuador se llevó a cabo un interesante proyecto centrado en estimar la productividad del cacao nacional en la Hacienda Victoria. El autor Chang (2024) , explica que se desarrolló una aplicación web utilizando herramientas de código abierto y algoritmos de aprendizaje supervisado. Esta aplicación incorpora variables climáticas, datos de producción y registros de plantaciones, todo ello siguiendo la metodología ágil conocida como Extreme Programming (XP). Tras realizar un análisis de correlación entre las variables y evaluar diferentes modelos predictivos, se eligió como opción principal el modelo Random Forest basado en regresión. Esto se debió a su capacidad para generar estimaciones de producción con un horizonte de hasta cinco meses después de la fecha de consulta.

2.2 Bases teóricas

2.2.1 *El origen del banano*

Los bananos provienen de la familia del plátano o Musáceas, son la fruta tropical más cultivada y son consideradas importantes sus plantaciones en el mundo porque únicamente son superadas por los cítricos, las uvas y las manzanas.

Los plátanos son una fruta tropical fundamental y pertenece a los árboles frutales conocidos como la manzana del paraíso, no obstante, es procedente de los trópicos 38 cálidos del sudeste asiático. Además, las primeras noticias sobre la especie se remontan a dibujos existentes en el antiguo sitio del monumento de Buda de Java que se levantó en Bordodur en 850 a.C. Así mismo, es muy conocido en el Mediterráneo desde el 650 después de Cristo y la especie llegó a Canarias en el siglo XV y finalmente fue traída a América a partir del año 1516 (Singh et al., 2020).

Además de ser conocidos por su riqueza en potasio, sodio y magnesio, los bananos constituyen una pieza clave en la seguridad alimentaria y económica de muchas regiones. A nivel global, se posicionan como el producto tropical de mayor valor comercial, siendo un motor de ingresos fundamental para países como Ecuador o Filipinas, que lideran la exportación de esta fruta. Sin embargo,

detrás de estas cifras de éxito existe una dependencia crítica del entorno natural. El desarrollo óptimo de la planta exige características de suelo y factores climáticos muy específicos, lo que convierte la gestión técnica y ambiental en el factor determinante para la rentabilidad de las plantaciones.

2.2.2 Cultivo de banano en el Ecuador

El cultivo y la exportación de banano representa un segmento económico clave de la agricultura en muchos países tropicales, sin embargo, Ecuador tiene una alta productividad de esta fruta debido a que tiene los mejores suelos y factores meteorológicos para el crecimiento de esta.

Ecuador está entre los principales países exportadores de banano en todo el mundo, pero las personas que se dedican a cultivar estos sembríos la mayoría son productores de banano donde su economía se basa por la producción del banano y sumando las pequeñas empresas se alcanza el 95,6% aproximadamente (Magdama et al., 2020). En otras palabras, Ecuador es considerado el mayor exportador al nivel mundial del banano porque es 39 consumido el producto por diferentes partes del mundo debido la calidad de fruta que se producen en los cultivos, también por las buenas condiciones climáticas y la riqueza del suelo ha facilitado el comercio en el país.

Se considera al banano unas de las plantaciones rentables y además representativo de la región Costa sobresaliendo en el ámbito agropecuario porque los elementos que se extraen de este fruto permiten el crecimiento de las plantaciones.

En el 2020, el área plantada de cultivos permanentes solamente de bananos ha sido de 165.080 hectáreas a nivel nacional y la más grande producción está en la provincia de Los Ríos con el 41,43% (Instituto Nacional De Estadísticas y Censos, 2020). Esto quiere decir, que las provincias donde el banano tiene mayor exportación se destaca Los Ríos porque posee un 31% más de terreno y en las otras provincias apenas llegan al 26%. Sin embargo, el cultivo de banano es calificado como permanente porque es un sembrío de mayor producción a nivel nacional y también es uno de los productos más exportados del Ecuador de manera internacional.

2.2.3 Producción de banano

La cadena productiva del banano en Ecuador involucra diversos agentes y elementos que inciden en los mercados interno y externo. Su cultivo mantiene

una estrecha relación con la economía familiar campesina, favorecida por las condiciones edáficas predominantes en las zonas agrícolas. No obstante, los costos asociados al proceso productivo presentan una tendencia al alza, dado que requiere la incorporación de insumos, recursos materiales, fuerza laboral y servicios fundamentales para su desarrollo y posterior comercialización (Borja, 2016) . Estos factores inciden directamente en la rentabilidad de los productores. La participación de cada componente humano y técnico en las fases del cultivo resulta determinante, ya que influye directamente en la calidad del banano. Las condiciones climáticas nacionales, especialmente las de la región litoral, representan una ventaja competitiva que se refleja en los atributos del fruto obtenido.

2.2.4 Agrometeorología

La agricultura tiene una dependencia directa muy acusada de las condiciones ambientales que determinan, en gran medida, la viabilidad y el desarrollo de las producciones agrícolas. Nuestras condiciones regionales presentan una marcada incidencia de situaciones de aridez en los 5 secanos, de frecuente déficit hídrico en el regadío y de régimen termo pluviométrico irregular, que con frecuencia genera situaciones adversas.

Los datos meteorológicos son un recurso natural que intervienen en las producciones y ante esto el agricultor no debe esperar pasivamente estas inclemencias del tiempo, la agrometeorología busca concientizar al agricultor sobre la forma de sacar provecho sobre ciertas condiciones naturales evitando consecuencias.

El agua es un eje crítico para la agricultura, actuando como un factor limitante tanto por su ausencia extrema como por su exceso repentino. Esta disponibilidad suele variar drásticamente incluso en distancias cortas, donde la diversidad de relieves y microclimas genera contrastes marcados entre zonas húmedas y áridas. La vulnerabilidad de los cultivos ante fenómenos meteorológicos como inundaciones o heladas, es especialmente alta durante fases fisiológicas clave, como la germinación. Al respecto, EOS Data Analytics (2024) subraya que el estrés hídrico provocado por la sequía es uno de los principales responsables de la caída en los rendimientos y de las pérdidas financieras que golpean la rentabilidad del sector.

Son importantes los datos agrometeorológicos porque se evalúa el potencial agrícola de la región según su localización o ubicación, además de los tipos de producción. De esta manera la información puede ayudar con equipos y métodos que se ajustan mejor a las condiciones de ciertas zonas respectivamente, particularmente en lo que respecta a la irrigación y al drenaje. También pueden servir para poner de manifiesto las necesidades de investigación en estos campos.

2.2.5 Requerimientos de nutrientes en el suelo en el cultivo de banano

La estabilidad química del suelo actúa como el pilar invisible sobre el cual se construye una cosecha de banano competitiva. Más que elementos aislados, macronutrientes como el nitrógeno, el fósforo y el potasio deben orquestrar un equilibrio preciso: mientras el primero impulsa el vigor del follaje, los otros dos aseguran raíces robustas y una fruta con la firmeza necesaria para la exportación. En este sentido, la calidad edáfica no es un concepto estático, sino la capacidad funcional del terreno para sostener la salud de la planta frente a las exigencias del ciclo productivo. Sin embargo, la realidad técnica en muchas fincas revela que la baja fertilidad sigue siendo un obstáculo crítico que impide que el banano alcance su máximo potencial genético y comercial (Amaya, 2022).

La viabilidad de una plantación bananera de exportación está estrechamente ligada a la presencia de micronutrientes esenciales que regulan su fisiología. Desde el azufre, fundamental en la síntesis de aminoácidos, hasta el boro y el zinc, cada elemento cumple una función que previene desórdenes metabólicos y caídas en el rendimiento. Lograr este equilibrio requiere un monitoreo constante del suelo, traduciendo variables como la capacidad de retención de nutrientes en planes de fertilización técnica. Bajo esta premisa, la integración de prácticas sostenibles y el análisis detallado de la estructura del suelo se vuelven herramientas indispensables para optimizar el desarrollo del cultivo y mantener la fertilidad a largo plazo.

Diversos factores, como las condiciones climáticas, el manejo del suelo y la infraestructura agrícola, influyen significativamente en la producción de plátano en Ecuador entre 2014 y 2016, impactando directamente en su rendimiento y competitividad (Daquilema y Villa, 2020).

2.2.6 Sistema gestor de bases de datos

El sistema gestor de bases de datos es crucial en el desarrollo de sistemas de información, ya que garantiza la compatibilidad para almacenar datos sin problemas, siendo fundamental elegir el más adecuado. Estos sistemas permiten crear un espacio lógico para almacenar grandes volúmenes de datos y consultarlos sin inconvenientes (Nacional et al., 2021). La información es un activo clave para las empresas, por lo que las bases de datos son esenciales para su almacenamiento seguro.

Las bases de datos facilitan la recuperación de información en caso de siniestros, asegurando su protección. Para que un sistema de información funcione correctamente, es indispensable un sistema gestor que proporcione las herramientas necesarias para un control eficiente (Jacinto Parinango, 2022). En este proyecto se utilizó un sistema de bases de datos con las características adecuadas para un almacenamiento óptimo.

Además, las bases de datos permiten realizar respaldos periódicos, esenciales para garantizar la continuidad de las operaciones y la recuperación de datos ante cualquier eventualidad (León Soberón, 2020). Esto representa una ventaja significativa para el almacenamiento y un aporte valioso para cualquier organización.

2.2.7 Lenguaje PHP 8.3

La relevancia de PHP en el panorama actual se debe a su capacidad para evolucionar a la par de las demandas de la web moderna. Al ser una herramienta potente y de libre distribución, ofrece una curva de aprendizaje amigable sin sacrificar el rendimiento necesario para sistemas complejos. La versión 8.3 representa un hito en esta evolución, fortaleciendo la seguridad y la eficiencia operativa. Gracias a la mejora en atributos y funciones internas, esta actualización permite a los desarrolladores estructurar aplicaciones con mayor rigor técnico. Adoptar PHP 8.3 no es solo una actualización de versión, es una apuesta por la estabilidad y la creación de arquitecturas web alineadas con los más altos estándares de calidad y seguridad de datos.

Para que pueda existir una buena interacción entre el sistema y el usuario es necesario que se cuente con una interfaz que sea de fácil entendimiento y adaptación para las personas y entidades encargadas de manipular el sistema, y esto es una de las grandes ventajas que ofrece PHP. Según Nixon (2020),

“Cuando un servidor se encuentra esta extensión en un archivo que se le ha solicitado. Lo pasa automáticamente a PHP” (p. 35), Este lenguaje de programación tiene un entorno de programación parecido a C++, el cual es reconocido por ser un lenguaje base para todo programador, hace que las personas que migran a PHP se adapten de manera rápida y simple. Siendo esta una de las características por la cual los usuarios optan por desarrollar sus sistemas en este lenguaje el mismo que tiene una buena interacción la cual no dará problemas al manejar el sistema informático.

2.2.8 HTML 5.3

Más que un simple lenguaje de etiquetas, la evolución actual de HTML ha transformado la base de la web en una estructura mucho más lógica y accesible. Al priorizar la semántica, esta versión permite que el contenido se organice con una claridad que facilita enormemente la integración con capas de diseño (CSS) y comportamiento (JavaScript). Los nuevos elementos y atributos no son meras adiciones técnicas; son herramientas estratégicas que optimizan la funcionalidad de las interfaces, logrando que las aplicaciones web modernas sean más ligeras, adaptables y, sobre todo, inclusivas para el usuario final en cualquier dispositivo.

Según Casado (2021) nos menciona que “HTML es un lenguaje muy sencillo que permite describir hipertexto, es decir, texto presentado de forma estructurada y agradable, con vínculos o enlaces (hyperlinks) que conducen a otros documentos o fuentes de información relacionadas y con inserciones multimedia” (p. 1). Los HTML son un conjunto de etiquetas que nos permite utilizar, para el diseño de nuestra página web en la cual podemos darles diferentes estilos, y la pagina sea agradable con el usuario que la va a manipular según Fossati (2020) “Los elementos HTML pueden utilizar una clase de estilo y un estilo individual con nombre y estos se utilizan para mostrar excepciones de estilo” (p. 25). En la cual los elementos de HTML nos permiten que nuestra página web sea amigable mediante los diferentes elementos, en fin, los HTML nos permite que nuestra página web sea eficiente.

2.2.9 Cascading Style Sheet (CSS)

CSS, siglas de Cascading Style Sheets (hojas de estilo en cascada), es un pilar esencial en el desarrollo de las interfaces digitales contemporáneas. Su objetivo principal es gestionar la estética y el diseño visual de los sitios web, lo que proporciona a los creadores un control total sobre cómo se muestra el

contenido en los navegadores. Según la Universidad Internacional de La Rioja (2025), “este lenguaje funciona a través de un conjunto de reglas que definen propiedades específicas para cada elemento, basándose en principios fundamentales como la cascada y la especificidad para resolver conflictos de estilo y optimizar la interpretación del código por parte del navegador”. La implementación de CSS fue decisiva en esta propuesta, ya que su capacidad para desacoplar la estructura lógica (HTML) del diseño visual permite el desarrollo de plataformas web altamente personalizables, fáciles de mantener y que ofrecen una experiencia de usuario superior.

2.2.10 JavaScript

JavaScript (JS) desempeña un papel fundamental en el desarrollo web, ya que puede ejecutarse en el lado del cliente (en el navegador) y, a través de plataformas como Node.js, también en el lado del servidor. Esta capacidad permite a los desarrolladores crear aplicaciones dinámicas e interactivas. Investigaciones recientes destacan que el uso de JavaScript en el lado del servidor puede dar lugar a soluciones que admiten una alta concurrencia y capacidad de respuesta en tiempo real. Por ejemplo, se señala que “Node.js, que se basa en JavaScript, ofrece un rendimiento y una escalabilidad excelentes gracias a su modelo de entrada/salida sin bloqueo” Esquivel-Paula et al. (2025).

2.2.11 Python

Python es un lenguaje que ha ganado mucha popularidad en el ámbito de la ciencia de datos y el aprendizaje automático, gracias a su amplio ecosistema de bibliotecas, su facilidad de lectura y su capacidad para automatizar procesos de análisis, modelado y visualización. En una guía técnica reciente en español, se resalta su ventaja para estructurar análisis que sean reproducibles y fáciles de entender. Se menciona que su elección se debe a que permite “combinar una sintaxis intuitiva con capacidades analíticas avanzadas” Iniciativa Aporta (2024), lo que hace más sencillo construir y evaluar modelos predictivos, así como generar visualizaciones efectivas.

2.2.12 My Structured Query Language (MySQL)

MySQL es un sistema de gestión de bases de datos relacionales (RDBMS) muy utilizado gracias a su capacidad para almacenar y gestionar información de forma estructurada. Al ser una solución de código abierto, se utiliza habitualmente en proyectos académicos y empresariales que necesitan

una plataforma robusta para registrar, consultar y mantener datos de forma eficiente, incluso cuando el volumen de información aumenta considerablemente. Juma Alba et al., (2025)

En el campo del desarrollo de sistemas, MySQL es especialmente valioso para las aplicaciones web, ya que permite organizar y acceder de forma eficaz a grandes cantidades de datos. Esto proporciona una base sólida para funciones como la autenticación, los registros operativos y la generación de informes. Todo ello contribuye a la construcción del sistema, ya que facilita la gestión centralizada de la información y la consulta mediante sentencias SQL, integrándose perfectamente con las tecnologías y los lenguajes utilizados en el desarrollo web.

2.2.13 Machine Learning

Más que una simple ejecución de comandos, el Machine Learning se define como el proceso mediante el cual una computadora mejora su rendimiento en una tarea específica a través del análisis de datos, sin haber sido explícitamente programada para cada escenario. Bajo la perspectiva de Alpaydin, esta tecnología busca optimizar resultados utilizando ejemplos históricos como base de conocimiento. La detección de spam ilustra perfectamente este concepto en la práctica: el objetivo es construir un modelo capaz de clasificar la información entrante, discriminando entre contenido relevante y correo basura. De este modo, el sistema evoluciona con cada interacción, perfeccionando su capacidad para decidir qué correos deben ser archivados y cuáles deben descartarse.

Para materializar el filtro, es necesario transformar el lenguaje natural en un formato que la máquina pueda procesar. Primero, se integra una base de datos de correos debidamente etiquetados para el entrenamiento. Luego, cada mensaje se convierte en un vector de características: utilizando un diccionario predefinido, el sistema asigna un valor binario a cada posición del vector (donde $x = 1$ si la palabra existe y $x = 0$ si está ausente). Al consolidar este conjunto de datos, se aplica un algoritmo de aprendizaje que extrae los patrones necesarios para generar el modelo predictivo, automatizando así la distinción entre el contenido legítimo y el correo no deseado.

Machine Learning es una de las tecnologías de la Industria 4.0. Los avances en Machine Learning han aportado beneficios significativos en la toma

de decisiones estratégicas de las organizaciones (empresa, gobierno, sociedad civil). En las últimas décadas, los modelos predictivos usando algoritmos de Machine Learning han sido implementadas en entornos reales de organizaciones, que operan en la industria de la agricultura y en cadena de suministro global del banano, y han obtenido buenos resultados (Almeyda, 2022).

Respecto a la corrosión, gracias al gran crecimiento de la disponibilidad de los datos, se pueden recolectar tanto de las variables asociadas al proceso como de la ocurrencia de la degradación de la tubería y debido a que una de las ventajas de los modelos de ML es que pueden trabajar con múltiples variables, hace que estos algoritmos sean considerados para su aplicación en este tipo de requerimientos, con lo cual se consigue combinar diferentes aspectos y propiedades físico-químicas de las tuberías con las ciencias de la computación y de esta manera obtener productos enfocados en la predicción de la corrosión. Existen cuatro tipos de ML: aprendizaje supervisado, aprendizaje no supervisado, aprendizaje semi-supervisado, y aprendizaje por refuerzo.

2.2.13.1.Regresión Lineal

La regresión lineal constituye una herramienta estadística esencial para modelar la interacción entre una variable dependiente y sus factores independientes. Su propósito fundamental es identificar la línea de mejor ajuste que permita proyectar resultados basándose en la tendencia de los datos recolectados. Bajo la perspectiva de Angiolo et al. (2020), este concepto puede entenderse desde dos ángulos: como un proceso de distribución de probabilidad conjunta o como el ajuste empírico de una función de dependencia. Para asegurar la robustez del modelo, es imperativo realizar un estudio de linealidad que valide la proporcionalidad de las variables. En el análisis de la productividad del cacao, este enfoque facilita una interpretación cuantitativa de las tendencias, permitiendo transformar variables complejas en resultados estadísticos claros para la toma de decisiones.

2.2.13.2.Random Forest

El algoritmo Random Forest se fundamenta en la premisa de que un conjunto de modelos trabajando en sintonía es mucho más preciso que uno solo. Bajo este enfoque de ensemble learning, el sistema despliega múltiples árboles de decisión que analizan los datos de forma independiente. Según explican

Cevallos et al. (2020), la potencia de los Bosques Aleatorios reside en su capacidad predictiva grupal: cada árbol emite un "voto" o clasificación individual, y el sistema selecciona el resultado que cuente con el mayor respaldo del grupo. Esta democracia algorítmica permite evaluar diversas rutas lógicas de forma simultánea, lo que se traduce en predicciones mucho más robustas y capaces de generalizar con éxito ante datos complejos y variables.

2.2.13.3.SVM

El SVM es un algoritmo que pertenece a la familia del aprendizaje supervisado y que se usa para problemas tanto de clasificación como de regresión. Está diseñado para poder encontrar el hiperplano óptimo que maximice la separación entre los puntos de datos en diferentes clases, o el que mejor se adapte a los puntos de datos si se trata de regresión.

En regresión la salida es un número real, por lo que enfrenta el problema de tener una cantidad infinita de posibles salidas. Sin embargo, según, González-Palacio et al. (2020) las SVM pueden ser utilizadas eficazmente en regresión manteniendo sus características distintivas, tales como la generalización y la capacidad de gestionar datos complejos. En este sentido, se adopta el procedimiento clásico referido como Máquinas de Vectores de Soporte para Regresión (SVR) que sufre muy pocas alteraciones técnicas en relación a su uso con clasificación.

Según los autores, el algoritmo también constituye un instrumento valioso para predecir la productividad del cacao, que reporta beneficios suficientes para que una empresa agrícola interesada en mejorar la gestión y el rendimiento de sus cultivos utilice técnicas de análisis avanzadas como es el caso del aprendizaje de máquina.

2.2.13.4.XGBoost.

XGBoost, que significa eXtreme Gradient Boosting, es un algoritmo de aprendizaje automático que se basa en el boosting de árboles de decisión. Su propósito es crear modelos predictivos muy precisos al combinar varios "árboles débiles" que, de manera secuencial, corrigen los errores de los anteriores. Lo que realmente lo hace destacar es su capacidad para optimizar una función objetivo a través del gradiente, lo que significa que minimiza el error. Además, incorpora técnicas de regularización que ayudan a reducir el sobreajuste, lo que lo convierte en una herramienta especialmente robusta cuando se enfrentan a

relaciones no lineales y a interacciones complejas entre variables. En el ámbito agrícola, XGBoost ha demostrado ser muy eficaz tanto en problemas de regresión como de clasificación, como en la predicción del rendimiento de cultivos. Esto se debe a su habilidad para manejar grandes volúmenes de datos y variables heterogéneas, que incluyen factores climáticos, edáficos, de manejo agronómico y el historial productivo, además de su buen desempeño con valores faltantes. En el caso del cacao, XGBoost es particularmente útil para estimar la producción, ya que puede captar patrones complejos entre factores ambientales y productivos. Ajusta sus árboles de manera iterativa según el error del entrenamiento, lo que permite mejorar la predicción en cada etapa y lograr modelos estables con una alta capacidad de generalización, García y Santos (2021). Por eso, XGBoost puede mejorar su precisión a medida que se entrena con más datos y enfrenta estructuras complejas que no se pueden modelar fácilmente con técnicas tradicionales. Aprende de manera iterativa, corrigiendo los errores de predicción en cada paso. Esta capacidad resalta su valor como herramienta de apoyo en la toma de decisiones agrícolas, permitiendo a los productores anticiparse a diferentes escenarios productivos y climáticos, evaluar riesgos y planificar acciones de manera más oportuna. Al generar predicciones consistentes, incluso en contextos con relaciones no lineales e interacciones entre variables, XGBoost ayuda a gestionar los cultivos de forma más eficiente y preventiva, adaptándose a entornos complejos y mejorando la fiabilidad de los pronósticos para la planificación agrícola.

2.2.14 Bootstrap

El desarrollo de aplicaciones web multiplataforma encuentra en Bootstrap un aliado estratégico para gestionar la compatibilidad visual. Al implementar este framework, se elimina la necesidad de programar versiones aisladas para cada tipo de hardware, ya que su núcleo está diseñado para detectar y responder al tamaño de la pantalla en tiempo real. Como señalan Flórez y Hernández (2024), desde su creación en 2010, el objetivo ha sido democratizar el diseño responsivo mediante una estructura semántica clara. La integración de componentes modulares desde botones hasta carruseles, no solo acelera el proceso de maquetación, sino que asegura que la interfaz cumpla con los estándares actuales de navegabilidad y accesibilidad web.

2.2.15 Hosting

El hosting web es una tecnología que hace posible que los sitios web puedan funcionar en internet. Las páginas web no existirían sin hosting. Anteriormente, las empresas tenían que alojar sus propios sitios utilizando sus propios servidores, lo que era más difícil y costoso. Pero hoy son muchas más las posibilidades, más baratas y para todo tipo de usuarios.

Tanto para empresas como para particulares, los usuarios deben saber al menos los conceptos básicos de qué es el hosting, según IONOS (2024). Esto les permitirá hacer una mejor elección del proveedor y el tipo de servicio que mejor se adapte a sus requerimientos, tratando siempre de conseguir la mejor relación precio-beneficios.

El hosting es un servicio que permite almacenar toda la información que compone un sitio web, desde textos, imágenes y archivos, en servidores conectados a Internet. Entonces, pueden acceder al sitio desde cualquier lugar del mundo real. Ahora bien, también debido a que hay tantos proveedores hoy en día, las personas pueden usar el servicio de hospedaje que mejor se adapte a sus necesidades, que tengan más espacio, mayor velocidad, mayor seguridad o mejor atención. Esto supone una gran ventaja, sobre todo para aquellos que quieren iniciarse con su propio sitio web.

2.2.16 Estructura del software

2.2.16.1. Diagrama de contexto

El diagrama de contexto o diagrama DFD de nivel 0 ofrece una perspectiva del sistema en general, con su proceso central y las entidades a su alrededor permiten visualizar un esquema rápido del sistema. Lucidchart, (2024) afirma:

Un diagrama de flujo de datos (DFD) traza el flujo de la información para cualquier proceso o sistema. Emplea símbolos definidos, como rectángulos, círculos y flechas, además de etiquetas de texto breves, para mostrar las entradas y salidas de datos, los puntos de almacenamiento y las rutas entre cada destino (párr. 2).

Por lo que, el diagrama de contexto, como se muestra en la Figura 9., presenta el proceso central, el cual es el aplicativo web del programa de predicción de producción y como interactúa con sus entidades.

2.2.16.2. Diagrama DFD nivel 1

En el DFD nivel 1, se detalla el desglose de las entidades y los procesos principales, delineando las interacciones entre ellos. Esto permite visualizar las tareas fundamentales que llevará a cabo el administrador en el sistema web de predicción de la producción de la hacienda San Juan (Ver Figura 10.)

2.2.17 Diagramas UML.

En el análisis de requerimientos, se realiza la documentación de requerimientos y bosquejo del aplicativo web usando el Lenguaje Unificado de Modelado (UML), para esquematizar la arquitectura y el comportamiento del software mostrando cada proceso involucrado. Se implementarán los diferentes tipos de diagramas, como el modelado basado en escenario y el modelado de datos. Estos diagramas se los hace mediante la herramienta Lucid chart, un software de diagramación online.

2.2.17.1. Diagrama de caso de uso.

Los modelos basados en escenario, tal como los diagramas de caso de uso, describen el escenario de un caso en específico, el cual es la representación del sistema, es decir, los procesos en un lenguaje natural desde la opinión del usuario (actor), por lo que IBM (2021) menciona “Los casos de uso se crean para refinar un conjunto de requisitos basados en un rol o tarea. En lugar de la lista tradicional de requisitos que pueden no abordar directamente el uso de la solución” (párr. 1). Por ello, estos diagramas permiten conocer los procesos que requerirá el sistema web dictado por los actores.

2.2.17.2. Diagrama entidad relación.

Para organizar de mejor manera los datos del sistema, se debe crear un diagrama entidad relación que represente el modelo relacional y especifique las variables que se ingresarán en la base de datos. Lucid chart (2024) indica que un diagrama entidad-relación o ERD es un tipo de diagrama de flujo, en el que se ilustra la manera en que las “entidades” (persona, animal o concepto) se relacionan entre si dentro del sistema, además de que son creados para depurar bases de datos relaciones. En este diagrama se muestran diversas tablas en el que se establecen todas las entidades involucradas, relaciones y atributos que son necesarios para el bosquejo del sistema.

2.3 Marco Legal

Ley de la Propiedad Intelectual

Titulo I: De los Derechos de Autor y Derechos Conexos

Capitulo I: Del Derecho de Autor

Seccion I: Preceptos Generales

Art 4. Se reconocen y garantizan los derechos de los autores y los derechos de los demas titulares sobre sus obras.

Art. 5. El derecho de autor nace y se protege por el solo hecho de la creación de la obra, independientemente de su mérito, destino o modo de expresión.

Se protegen todas las obras, interpretaciones, ejecuciones, producciones o emisión radiofónica cualquiera sea el país de origen de la obra, la nacionalidad o el domicilio del autor o titular. Esta protección también se reconoce cualquiera que sea el lugar de publicación o divulgación. El reconocimiento de los derechos de autor y de los derechos conexos no está sometido a registro, depósito, ni al cumplimiento de formalidad alguna.

Art. 6. El derecho de autor es independiente, compatible y acumulable con: La propiedad y otros derechos que tengan por objeto la cosa material a la que esté incorporada la obra; Los derechos de propiedad industrial que puedan existir sobre la obra; y, Los otros derechos de propiedad intelectual reconocidos por la ley. (Ley de Propiedad Intelectual, 2022, p. 2).

Capítulo VI: Del Régimen Agropecuario

Art. 266.- Será objetivo permanente de las políticas del Estado el desarrollo prioritario, integral y sostenido de las actividades agrícola, pecuaria, acuícola, pesquera y agroindustrial, que provean productos de calidad para el mercado interno y externo, la dotación de infraestructura, la tecnificación y recuperación de suelos, la investigación científica y la transferencia de tecnología.

El Estado estimulará los proyectos de forestación, reforestación, sobre todo con especies endémicas, de conformidad con la ley. Las áreas reservadas a estos proyectos serán inafectables. Las asociaciones nacionales de productores, en representación de los agricultores del ramo, los campesinos y profesionales del sector agropecuario, participarán con el Estado en la definición

de las políticas sectoriales y de interés social. (Constitución Política de la República del Ecuador, 2008, pág. 45).

Ley Orgánica de Tierras Rurales y Territorios Ancestrales

Art. 10.- A fin de estimular a las y los propietarios y posesionanos de tierras rurales y alentarlos a una producción sostenible, sustentable y orientada a garantizar la soberanía alimentaria, el Estado en sus diferentes niveles de gobierno, realizara las siguientes acciones. Impulsar el desarrollo de programas y proyectos de emprendimiento productivo por parte de pequeños y medianos productos asociados, para vincularlos en programas de provisión de recursos monetarios para capital de riesgo, servicios financieros de apoyo, tecnificación, seguro agrícola y garantía crediticia. (Ministerio de Agricultura y Ganadería, 2021, p. 3).

Constitución Política de la República del Ecuador

Capitulo IV: De Los Derechos económicos, sociales y culturales

Sección novena: De la ciencia y tecnología

Art. 80.- El Estado fomentará la ciencia y la tecnología, especialmente en todos los niveles educativos, dirigidas a mejorar la productividad, la competitividad, el manejo sustentable de los recursos naturales, y a satisfacer las necesidades básicas de la población.

Garantizará la libertad de las actividades científicas y tecnológicas y la protección legal de sus resultados, así como el conocimiento ancestral colectivo. La investigación científica y tecnológica se llevará a cabo en las universidades, escuelas politécnicas, institutos superiores técnicos y tecnológicos y centros de investigación científica, en coordinación con los sectores productivos cuando sea pertinente, y con el organismo público que establezca la ley, la que regulará también el estatuto del investigador científico. (Constitución Política de la República del Ecuador, 2008, pág. 14).

Ley de fomento y Desarrollo Agropecuario

CAPITULO V

De la utilización del suelo

Art. 29.- El Ministerio de Agricultura y Ganadería, a través de sus organismos especializados, adoptará las medidas aconsejadas por las consideraciones ecológicas que garanticen la utilización racional del suelo y exigirá que las personas naturales o jurídicas que realicen actividades agrícolas,

pecuarias, forestales u obras de infraestructura que afecten negativamente a los suelos, adopten las medidas de conservación y recuperación que, con los debidos fundamentos técnicos y científicos determinen las autoridades competentes.

Art. 30.- El Ministerio de Agricultura y Ganadería podrá ordenar la suspensión de las tareas y obras de qué trata el artículo anterior, que ejecutaren personas naturales o jurídicas, si tales tareas y obras pudieren determinar deterioro de los suelos o afectar a los sistemas ecológicos. El Ministerio de Agricultura y Ganadería reglamentará esta disposición.

Art. 31.- El Ministerio de Educación incluirá en sus respectivos planes de estudio, el conocimiento, manejo y conservación de los recursos naturales renovables. El Ministerio de Agricultura y Ganadería, por su parte, realizará jornadas, seminarios, simposios, certámenes y eventos técnicos y científicos, de conservación de dichos recursos, con participación de las organizaciones campesinas, así como campañas de educación popular. (Ley de Fomento y Desarrollo Agropecuario, 2021, p. 8).

Plan Nacional de Desarrollo

Eje de desarrollo Económico

Objetivo 4: Estimular el sistema económico y de finanzas públicas para dinamizar la inversión y las relaciones comerciales.

- **Política 4.3:** Promover la inversión y la participación de los sectores público, privado, de la economía popular y solidaria, y de la sociedad en la infraestructura productiva y los servicios básicos, con enfoque territorial y de cierre de brechas.

Objetivo 5: Fomentar de manera sustentable la producción mejorando los niveles de productividad.

- **Política 5.1:** Promover el uso de tecnologías e innovaciones en los procesos productivos, con énfasis en la sostenibilidad y la eficiencia energética.
- **Política 5.2:** Impulsar la diversificación de la matriz productiva, con valor agregado y productos orientados a los mercados externos, con énfasis en los sectores agropecuario, industrial, acuícola, pesquero, forestal y de servicios."

- **Política 5.3:** Promover la asociatividad, la cooperación y la articulación entre los actores de la cadena de valor de los sectores priorizados, para mejorar la eficiencia y competitividad.
- **Política 5.4:** Fortalecer los sistemas de innovación, investigación y transferencia de tecnología, priorizando la vinculación del sector académico y productivo.
- **Política 5.5:** Impulsar la investigación aplicada y la adopción de innovaciones tecnológicas para la gestión del agua y del suelo.

Objetivo 7: Precautelar el uso responsable de los recursos naturales con un entorno ambientalmente sostenible.

- **Política 7.1:** Impulsar la gestión integrada y participativa de los recursos hídricos, del suelo y del aire, para garantizar su disponibilidad y calidad.

3. MATERIALES Y MÉTODOS

3.1 Enfoque de la investigación

El enfoque de esta investigación es cuantitativo, ya que se fundamenta en la recolección y análisis de datos numéricos relacionados con la producción de banano, tales como la cantidad de enfundes, racimos cosechados, cajas procesadas, así como variables climáticas como la temperatura y la precipitación. Según un estudio sobre investigación cuantitativa se destaca lo siguiente:

La investigación cuantitativa es en sí una metodología que se centra por lo que en la recolección y el análisis de datos numéricos para lo cual comprender y explicar en sí fenómenos sociales, y económicos, y científicos y empresariales. A través de este enfoque como tal los investigadores buscan cuantificar en sí variables, identificar patrones y establecer relaciones causales entre los fenómenos estudiados. (Villar, 2024, p. 15).

Este enfoque permite en sí establecer relaciones entre las variables, identificar patrones y construir modelos predictivos mediante el uso de herramientas estadísticas y algoritmos de aprendizaje automático. (Ver Anexo N°9).

3.1.1 Tipo y alcance de la investigación

3.1.1.1. Investigación aplicada

Este tipo de investigación como tal comprometió un trabajo exclusivo desarrollado con el fin de adquirir nuevos conocimientos con aplicación directa, utilizados para resolver problemas concretos de la sociedad o del sector productivo.

En este tipo de investigación en sí el énfasis del estudio está en la resolución práctica de problemas, centrándose específicamente en cómo se pueden llevar a la práctica las teorías generales, y su motivación se orienta a resolver necesidades como tal o dificultades planteadas en contextos reales como tal, la investigación aplicada en sí guarda una estrecha relación con la investigación básica por lo que, ya que depende de los descubrimientos de esta última y en sí se nutre de ellos para su implementación (Rodríguez, 2020, p.3).

La investigación aplicada como tal se orienta a la solución del problema de incertidumbre en la planificación de la producción bananera en sí a través del desarrollo de un modelo predictivo por lo que está basado en datos históricos y climáticos, y se podría decir que este modelo permitirá mejorar en sí la toma de decisiones y optimizar los procesos de producción en el sector agrícola.

3.1.1.2. Investigación correlacional

El estudio adopta un enfoque correlacional con el fin de mapear la interdependencia entre variables cuantitativas sin forzar una relación de causalidad directa. El objetivo es descifrar cómo factores meteorológicos específicos como la temperatura o la pluviosidad, se asocian con los indicadores clave de rendimiento (enfundes, racimos y cajas procesadas). Al determinar este grado de asociación, se sientan las bases lógicas para el modelo predictivo; entender cómo se comportan estas variables en conjunto permite identificar los patrones históricos que realmente dictan el ritmo de la producción. Este análisis técnico es lo que otorga validez al modelo, asegurando que las proyecciones futuras se sustenten en tendencias climáticas y operativas debidamente contrastadas. (Ver Anexo N°4).

3.1.2 Diseño de investigación

El diseño de la investigación se realizó acorde a las necesidades del tipo de proyecto, que en este caso específico correspondió al desarrollo de un modelo predictivo basado en datos climáticos y productivos para anticipar la producción de banano.

En este sentido, se empleó un diseño no experimental, ya que no se realizó manipulación directa de las variables, sino que se observó y analizó los datos tal como se presentan en su contexto natural. Este diseño permitió trabajar con información recolectada previamente el fin de identificar patrones y relaciones entre las variables involucradas. (Ver Anexo N°10).

Galarza (2021) menciona que “la investigación no experimental se realiza cuando, durante el estudio, el investigador no puede controlar, manipular o alterar a los sujetos, sino que se basa en la interpretación o las observaciones para llegar a una conclusión” (p. 3). Por tanto, en este proyecto se recopiló datos mediante el análisis de registros históricos de fincas bananeras y fuentes meteorológicas, sin intervención alguna en los procesos productivos. Esta información fue la base para aplicar técnicas estadísticas y modelos de predicción que permitieron mejorar la planificación agrícola en el sector bananero.

3.2 Metodología

La metodología de desarrollo de software que se utilizó en esta propuesta tecnológica es la Programación Extrema (XP), porque permitió construirlo de

manera ordenada de acuerdo a cada una de las fases, incorporando prácticas de desarrollo iterativo, retroalimentación continua y adaptación a los requerimientos del usuario. Esto garantizó que el aplicativo fuese flexible, funcional y fácil de mantener. Complementariamente, se empleó la metodología de Descubrimiento de Conocimiento en Bases de Datos (KDD) para el análisis de los datos meteorológicas y del suelo, siguiendo etapas como selección, preprocesamiento, transformación, modelado y evaluación. La integración de XP y KDD permitió que el desarrollo del software estuviera alineado con un proceso riguroso de análisis de datos, asegurando modelos predictivos confiables y un aplicativo eficiente para la predicción de la producción de banano en la Hacienda San Juan.

3.2.1 Metodología XP

La metodología ágil XP o Programación Extrema se enfocó en la interacción constante del cliente, gerente y programador en el cual se define el papel que tiene cada uno. Además, se producen y prueban pequeñas versiones periódicamente, con el fin de mantener la satisfacción del cliente a lo largo del ciclo de vida del desarrollo del software.

Por lo tanto, la metodología XP escogida para desarrollar las funcionalidades del sistema cumplió con todos los parámetros necesarios y a su vez ayudó a que el programa sea flexible a los cambios que se requiera durante la ejecución del proyecto. Sin embargo, cuenta con cuatro fases que son: planificación, diseño, desarrollo y pruebas. (Ver Anexo N°2).

3.2.1.1.Fase de planificación

La etapa de planificación constituye la hoja de ruta técnica del proyecto, donde las necesidades del usuario se traducen en estructuras tangibles. Para lograr esta claridad, se emplean herramientas visuales estratégicas: el Diagrama de Flujo de Datos (DFD) permite rastrear con precisión el recorrido de la información, mientras que los Casos de Uso delimitan el alcance de las interacciones de cada actor. A esto se suma el rigor del estándar UML, utilizando Diagramas de Clases para establecer la arquitectura interna del software y Diagramas de Actividades para secuenciar cada proceso operativo. Más que simples esquemas, este conjunto documental asegura que el desarrollo sea viable, minimice errores lógicos y responda con exactitud a los objetivos planteados. (Ver Anexo N°11).

Dentro de las actividades que se desarrollaron en la fase de planificación se realizó la visita inicial a la hacienda SAN JUAN para realizar la respectiva entrevista al administrador de la hacienda. En el marco del este encuentro fue de vital importancia al darse a conocer las necesidades a cubrir, esto también ayudo a definir los objetivos del proyecto, el boceto del sistema y por medio de esto se realizaron los requerimientos necesarios para desarrollar el sistema.

En base a los requerimientos que se obtuvieron en la entrevista, se realizó las historias de usuarios, las cuales incluyen los requisitos funcionales como los no funcionales. Estos requisitos son claves en la identificación de los términos importantes y específicos del ámbito de la hacienda, lo que facilita la creación del glosario LEL (Lenguaje Especifico del Dominio). Este glosario es una herramienta esencial porque permite entender con claridad la terminología utilizada en el proyecto, asegurando que los usuarios tengan una comprensión común y precisa de los términos utilizados.

La planificación bajo el enfoque ágil de XP se aleja de los esquemas rígidos para priorizar una evolución basada en la retroalimentación constante. En lugar de intentar prever cada detalle desde el inicio, el desarrollo avanza mediante ciclos que permiten pivotar según las necesidades reales que surgen en el campo. Esta dinámica no solo mejora la capacidad de respuesta ante imprevistos, sino que asegura que el aplicativo final responda con precisión a las expectativas del usuario. Al cierre de cada iteración, se evalúan los requisitos y se renegocian las prioridades, garantizando que el esfuerzo técnico se concentre siempre en las funcionalidades de mayor impacto para el sistema. (Ver Anexo N°3).

3.2.1.2. Fase de diseño

El diseño de software no consiste en dibujar diagramas; es un ejercicio de gestión de la complejidad mediante abstracción. La estrategia arquitectónica aquí prioriza la Separación de Responsabilidades (SoC) por encima de todo. Al forzar un desacoplamiento estricto entre módulos, no solo 'organizamos' código; aislamos dominios de fallo para evitar que un error local tumbe todo el sistema. Este modularidad es la única defensa real contra la deuda técnica futura (Ver Anexo N°12).

Se emplearon diagramas de contexto, diagramas de flujo de datos, diagramas de casos de uso y diagramas arquitectónicos, los cuales permitieron

analizar de manera integral cómo se gestionan los datos, cómo interactúan los usuarios con el sistema y cómo se integran los módulos funcionales, tales como el procesamiento de datos, modelado predictivo, visualización y generación de reportes. A través de estas representaciones gráficas, se logró identificar las principales funcionalidades del sistema, su estructura interna y las relaciones entre sus componentes, contribuyendo a un desarrollo más eficiente y alineado con los objetivos del proyecto.

A continuación, se detallan los diagramas empleados:

- **Diagrama de contexto:** Se utilizó el diagrama de contexto (DFD nivel 0) para representar el sistema de predicción como una entidad central y su interacción con los actores externos, principalmente el administrador del sistema. Este diagrama muestra las entradas de datos, como variables climáticas (temperatura, precipitación) y edáficas (nutrientes del suelo), así como las salidas generadas por el sistema, tales como predicciones de producción y reportes analíticos. De esta manera, se obtiene una visión general del flujo de información entre el sistema y su entorno.
- **Diagrama de flujo de datos (DFD nivel 1):** Se elaboró el diagrama de flujo de datos nivel 1 para detallar los procesos internos del sistema, descomponiendo el flujo de información en actividades específicas como la captura de datos, procesamiento, almacenamiento, modelado y generación de resultados. Este diagrama permite visualizar cómo los datos son transformados desde su ingreso hasta la obtención de predicciones, evidenciando la interacción entre los diferentes módulos del sistema y la base de datos.
- **Diagramas de casos de uso:** Los diagramas de casos de uso se emplearon para representar las funcionalidades principales del sistema desde la perspectiva del usuario. En este caso, el actor principal es el administrador, quien interactúa con el sistema para registrar datos, ejecutar predicciones, visualizar resultados y generar reportes. Estos diagramas permiten identificar claramente las acciones disponibles dentro del sistema y cómo cada módulo responde a las necesidades del usuario, proporcionando una visión funcional del aplicativo.
- **Diagrama arquitectónico:** El diagrama arquitectónico se utilizó para representar la estructura general del sistema, incluyendo la separación entre

el frontend, backend y los componentes de procesamiento de datos y modelos de Machine Learning. Este diagrama muestra cómo tecnologías como PHP, MySQL y Python interactúan entre sí, así como la integración de los algoritmos predictivos dentro del sistema, permitiendo comprender la organización y comunicación entre los distintos componentes tecnológicos.

- **Diagrama modular:** Se desarrolló un diagrama modular basado en la arquitectura del sistema, el cual permitió organizar los diferentes módulos funcionales, tales como el módulo de procesamiento de datos, visualización, gestión y reportes. Este enfoque facilita la comprensión de la estructura del sistema, mostrando cómo cada módulo cumple una función específica y cómo se relacionan entre sí, contribuyendo a un diseño más claro, escalable y mantenible.

3.2.1.3. Fase de codificación

En esta fase crucial del proyecto, se llevó a cabo la codificación de cada uno de los módulos que componen el aplicativo, brindando vida a las funcionalidades deseadas. Este proceso meticuloso involucró el desarrollo de algoritmos robustos y eficientes, la selección de estructuras de datos adecuadas para optimizar el rendimiento, la utilización de librerías especializadas que ampliaron las capacidades del sistema y la integración con sistemas de bases de datos para garantizar la seguridad y persistencia de la información. A continuación: se detalla las principales herramientas utilizadas:

Para el desarrollo del backend, se optó por el uso de PHP en su forma nativa (Vanilla), organizado mediante una arquitectura basada en carpetas, lo que permitió una separación clara de responsabilidades entre módulos como autenticación, gestión de vistas y procesamiento de datos. Esta elección brinda un mayor control sobre la lógica del sistema y reduce la sobrecarga de frameworks pesados. Para la interacción con la base de datos, se utilizó PDO (PHP Data Objects), lo cual permite realizar conexiones seguras mediante el uso de sentencias preparadas, reduciendo riesgos de vulnerabilidades como la inyección SQL.

En cuanto a la gestión de datos, se empleó el sistema de bases de datos relacional MySQL/MariaDB, el cual permite almacenar de manera estructurada la información correspondiente a usuarios, roles, configuraciones del sistema y

datos asociados a los análisis y predicciones. Su uso garantiza integridad, consistencia y eficiencia en el acceso a la información.

Para la capa de presentación, el sistema fue desarrollado utilizando tecnologías web estándar como HTML5, CSS3 y JavaScript, complementadas con el framework Bootstrap 4 para lograr un diseño responsivo y adaptable a diferentes dispositivos. Adicionalmente, se utilizó la plantilla AdminLTE en su versión 3, la cual proporciona una estructura moderna de panel administrativo, facilitando la organización visual del sistema y mejorando la experiencia del usuario.

El dinamismo en la interfaz se logró mediante el uso de la librería jQuery, que permite la manipulación del DOM y la ejecución de peticiones asíncronas. Asimismo, se integraron diversas librerías especializadas como Chart.js, Flot y uPlot para la visualización de datos mediante gráficos interactivos; DataTables para la gestión dinámica de tablas con funcionalidades de búsqueda y paginación; SweetAlert2 y Toastr para la implementación de notificaciones visuales; y FontAwesome para la incorporación de iconografía en la interfaz.

Por otro lado, el sistema incorpora un módulo avanzado de predicción basado en técnicas de Machine Learning, el cual requirió un proceso de tratamiento de datos que incluye etapas de limpieza, transformación y estructuración. Para este propósito, se utilizaron herramientas del ecosistema de Python como Pandas y NumPy para el análisis y manipulación de datos, así como Matplotlib para la generación de visualizaciones.

El modelo predictivo principal fue desarrollado utilizando el algoritmo XGBoost, una técnica de aprendizaje supervisado basada en Gradient Boosting, ampliamente reconocida por su alto rendimiento en problemas de predicción. Adicionalmente, se empleó Scikit-learn para tareas complementarias como la división de datos, validación del modelo y evaluación de métricas. Los modelos generados son almacenados en formato JSON, permitiendo su reutilización y despliegue eficiente.

Para la exposición de los servicios de predicción, se desarrolló una API REST utilizando FastAPI, un framework moderno de alto rendimiento para Python. Esta API actúa como intermediario entre el sistema principal desarrollado en PHP y los modelos de Machine Learning, permitiendo enviar datos de entrada y recibir predicciones de forma rápida y segura. La ejecución

de esta API se realiza mediante el servidor Uvicorn, garantizando un manejo eficiente de múltiples solicitudes concurrentes.

Finalmente, se utilizaron herramientas adicionales como la librería Requests para la comunicación con servicios externos mediante peticiones HTTP. Este enfoque integral permite que Banapred sea un sistema modular, escalable y adaptable, capaz de integrar múltiples tecnologías y ofrecer predicciones precisas que apoyan la toma de decisiones en el ámbito agrícola. Siguiendo las directrices de Asana (2024), "todo el código debe cumplir con el estándar de programación propuesto" (p. 25), lo que garantiza una estructura limpia y escalable. Para lograrlo, se emplea un ecosistema de software libre basado en PHP para la lógica del servidor y HTML para la arquitectura visual. La persistencia de los datos recae en MySQL, cuya función es administrar las relaciones complejas entre las tablas que albergarán la información histórica y actual de la producción bananera de la hacienda. (Ver Anexo N°13).

3.2.1.4. Fase de pruebas

En el paradigma de Extreme Programming (XP), el aseguramiento de la calidad deja de ser una fase final para convertirse en una disciplina de Ingeniería de Confiabilidad. La estrategia de validación opera sobre una dicotomía estricta. Por un lado, ejecutamos pruebas unitarias y de integración que auditan la atomicidad lógica del código: verificamos matemáticamente que cada micro-componente cumpla su contrato de interfaz sin generar efectos secundarios en cascada. Esto previene la 'fragilidad sistémica', donde un cambio menor en la lógica de negocio rompe la estabilidad del aplicativo.

Simultáneamente, desplazamos el foco hacia las Pruebas de Aceptación de Usuario (UAT). Aquí, la validación no es sintáctica, sino operativa: confrontamos la interfaz con la realidad del campo para detectar fricciones de usabilidad que ningún compilador puede ver. Este ciclo de retroalimentación agresiva blindada la arquitectura contra la obsolescencia técnica. Garantiza un sistema con bajo acoplamiento, capaz de soportar la inyección futura de nuevos modelos de inferencia o estructuras de datos heterogéneas sin sufrir regresiones funcionales críticas.

3.2.1.5. Fase de lanzamiento

La fase de liberación en XP repudia el concepto de 'Big Bang Deployment'. No existe un 'día de lanzamiento' monolítico; se implementa una estrategia de

Entrega Continua donde la puesta en producción es granular y constante. La estabilidad del sistema no se asume, se audita: cada incremento funcional debe superar pruebas de humo (smoke tests) y validaciones de regresión en un entorno de staging antes de interactuar con datos vivos.

Paralelamente, la transferencia de tecnología no busca meramente 'capacitar', sino mitigar la Resistencia al Cambio operativa. La documentación técnica actúa aquí como un mecanismo de contención de errores humanos. Una vez en producción, el software entra en un estado de Observabilidad: la retroalimentación de campo no se trata como 'quejas', sino como telemetría cualitativa. Este input alimenta el backlog de mantenimiento evolutivo, permitiendo refactorizar la lógica de negocio en tiempo real para que el algoritmo no pierda sincronía con la volatilidad agronómica de la hacienda.

3.2.2 Metodología KDD

La columna vertebral de esta investigación se sustenta en el proceso de Descubrimiento de Conocimiento en Bases de Datos (KDD). Su aplicación es especialmente valiosa en el entorno agrícola, ya que permite transformar los registros históricos de producción y las fluctuaciones climáticas en una ventaja competitiva para el cultivo de banano. Como se detalla en el Anexo N°14, el propósito central no es solo procesar información, sino extraer patrones que sean válidos, originales y, sobre todo, útiles para la gestión operativa. Para lograrlo, el marco de trabajo se organiza en una secuencia lógica de fases que van desde la depuración inicial de los datos hasta la obtención de conocimiento accionable para la planificación de recursos. (Ver Figura 5).

3.2.2.1. Selección de datos

La fase de selección en KDD no es una simple 'recolección'; es un ejercicio crítico de Ingeniería de Características (Feature Engineering). Rechazando el paradigma de Garbage In, Garbage Out, la arquitectura impone una fusión de datos heterogénea. Orquestamos la ingesta de métricas operativas deterministas (registros de cosecha) junto con datos meteorológicos, descartando agresivamente cualquier variable que no maximice la Relación Señal-Ruido.

Más allá de la limpieza, el desafío real es la No-Estacionariedad. Para que el algoritmo no sea ciego al tiempo, inyectamos explícitamente embeddings temporales o variables de rezago (lags). No le enseñamos 'fechas' al modelo;

estructuramos la data para que detecte matemáticamente la autocorrelación cíclica del cultivo. Esto transforma la estacionalidad, que suele ser un ruido en modelos estáticos, en un vector de peso predictivo tangible que estabiliza la inferencia a lo largo del año fiscal.

3.2.2.2.Preprocesamiento

Reducir el preprocesamiento a una 'limpieza' es un error conceptual; se trata de una fase de limpieza de datos crítica para evitar la divergencia del modelo. La estrategia no es borrar errores, sino gestionar la integridad estadística. Ante la presencia de outliers o valores nulos, rechazamos la eliminación pasiva de registros que reduce la densidad del dataset y optamos por técnicas de Imputación Matemática (como medias móviles o interpolación) para reconstruir la continuidad de la serie sin introducir sesgos artificiales.

Paralelamente, ejecutamos una Estandarización de Escala (Feature Scaling) para neutralizar la desigualdad de magnitudes entre variables heterogéneas, evitando que la precipitación domine numéricamente sobre la temperatura en el cálculo del gradiente. Sin embargo, la maniobra arquitectónica decisiva es el Remuestreo Temporal: forzamos la agregación de toda la telemetría a una granularidad estricta de 'Semana Agrícola'. Esta sincronización elimina el ruido de los registros asincrónicos, entregando al algoritmo una estructura de datos equidistante y secuencial, lista para la extracción de patrones de autocorrelación.

3.2.2.3.Transformación

La transformación de datos trasciende la simple manipulación aritmética; constituye una operación de Ingeniería de Características (Feature Engineering) diseñada para reducir la entropía del dataset. La estrategia de Normalización es imperativa, no opcional: al reescalar las magnitudes físicas a un rango acotado, evitamos que variables de alta varianza (como la pluviosidad milimétrica) secuestren la Convergencia del Gradiente durante el entrenamiento, garantizando una superficie de error uniforme.

Sin embargo, el valor predictivo real se extrae mediante la inyección de Ventanas Deslizantes (Rolling Windows). Al calcular promedios móviles, suavizamos el ruido estocástico diario para exponer la tendencia subyacente del estrés hídrico acumulado. Finalmente, abordamos la paradoja temporal mediante una Codificación Cíclica Trigonométrica: en lugar de usar índices

lineales (1-12) que rompen la continuidad entre diciembre y enero, proyectamos la estacionalidad en coordenadas cartesianas (Seno/Coseno). Esta maniobra topológica preserva la proximidad matemática de los ciclos agrícolas, permitiendo que el algoritmo navegue la temporalidad sin discontinuidades lógicas.

3.2.2.4. Minería de datos

La ejecución de la Minería de Datos no opera como una simple 'aplicación de algoritmos'; se estructura como un Benchmarking Competitivo riguroso. Definimos el problema bajo un esquema de Aprendizaje Supervisado, donde la Regresión Lineal actúa estrictamente como Línea Base (Baseline) para medir la ganancia de información mínima aceptable. Contra ella, desplegamos arquitecturas no lineales desde Máquinas de Soporte Vectorial (SVM) hasta Árboles de Decisión para desafiar la topología compleja del dataset.

La estrategia de validación rechaza la memorización de datos; implementamos una partición estricta de Entrenamiento/Prueba (Train-Test Split) diseñada para castigar el Sobreajuste (Overfitting) y garantizar que el error de generalización se mantenga bajo en producción. El objetivo es inferir la variable crítica 'Cajas Procesadas' utilizando vectores de características climáticas y productivas con rezago temporal (lagged features). Tras auditar las métricas de desempeño en este entorno comparativo, XGBoost se consolidó como la arquitectura dominante. Su motor de Gradient Boosting no solo minimizó la función de pérdida de forma más agresiva que los modelos clásicos, sino que capturó la varianza no lineal del rendimiento bananero con una precisión superior.

La mecánica interna del modelo rechaza la simplicidad del voto paralelo; implementa una arquitectura de Ensemble Aditivo estricta. Aquí, los Árboles de Decisión no operan como entidades autónomas, sino como Aprendices Débiles (Weak Learners) encadenados secuencialmente para atacar la función de pérdida. El algoritmo no busca predecir la cosecha directamente en cada paso, sino modelar el Gradiente Negativo de la iteración previa.

Básicamente, cada nuevo árbol se entrena para corregir los pseudo-residuales (errores) dejados por su predecesor, ejecutando un particionamiento recursivo del espacio vectorial climático y edáfico. Al encontrar los puntos de corte óptimos que maximizan la Ganancia de Información, el sistema refino

progresivamente la frontera de decisión. La inferencia final no es un promedio, sino la suma ponderada de todas estas correcciones marginales, convergiendo hacia un valor continuo de alta precisión que minimiza la varianza del pronóstico agrícola. (Ver Tabla 25.)

3.2.2.5. Evaluación e Interpretación

La adjudicación de XGBoost como motor de inferencia no obedece a una métrica aislada, sino a una victoria arquitectónica sobre el enfoque de Bagging. Mientras Random Forest se estancó en un Coeficiente de Determinación (R²) de 0.75 sufriendo para reducir la varianza en los extremos, XGBoost escaló la precisión hasta un R² de 0.89. Esta diferencia no es trivial; evidencia que la estrategia de Gradient Boosting logró extraer patrones donde los árboles independientes simplemente promediaban ruido.

Al optimizar directamente sobre los residuos del iterador anterior, el algoritmo demostró una plasticidad superior para mapear la no-linealidad estocástica entre el clima y el cultivo. Los indicadores de error validan esta decisión técnica: un RMSE de 0.49 y un MAE de 0.342 confirman que la función de pérdida convergió eficientemente. No solo redujimos la desviación estándar de las proyecciones; logramos que el sistema generalice correctamente ante la topología irregular de los datos agrícolas, evitando el falso ajuste que suele inflar métricas en arquitecturas menos rigurosas. (Ver Tabla 25.)

El entrenamiento del modelo con datos históricos nos permitió generar proyecciones altamente fiables sobre variables clave como los racimos cosechados y las cajas procesadas, logrando un error porcentual absoluto medio MAPE de solo el 4,12 % y el 3,67 %, respectivamente. La validación de estos resultados confirma que el sistema no solo aprende de los datos pasados, sino que también comprende la dinámica de la producción y cómo esta se ve influida por el entorno local. (Ver Tabla 26.)

Esta precisión convierte a la inferencia en un activo logístico crítico. Al anticipar la carga de cosecha con baja incertidumbre, la hacienda migra de una gestión reactiva a una planificación determinista. Esto impacta directamente la cadena de suministro (Supply Chain): permite dimensionar el inventario de empaque y la flota de transporte bajo un esquema Just-in-Time, eliminando el desperdicio por sobreaprovisionamiento (Over-provisioning). En definitiva,

XGBoost no solo predice; instrumentaliza la eficiencia, transformando la agronomía en una operación de datos auditable.

3.2.3 Recolección de datos

En este apartado se describen los recursos, métodos y técnicas de recopilación que fueron fundamentales para plantear los módulos que se esperaron desarrollar y de esta manera obtener una mejor comprensión de estructura en la primera fase de requerimientos.

3.2.3.1. Recursos

A continuación, se detallan cada uno de los recursos que son necesarios para el desarrollo del proyecto.

3.2.3.1.1 Recursos humanos

Un recurso indispensable para llevar a cabo cualquier labor es el capital humano. En este caso, resulta esencial la participación del creador del proyecto como estudiante, el acompañamiento de los docentes guías y quienes comparten sus conocimientos, así como la colaboración del administrador de la hacienda “San Juan”.

- **Estudiantes:** Anchundia Zambrano Andres David
- **Docente:** Ing. Mitchell Vasquez Bermudez
- **Administrador:** Ing. William Franco

3.2.3.1.2 Recursos Bibliográficos

Para el desarrollo de este proyecto se realizaron consultas en portales de empresas, libros electrónicos, tesis, sitios web, artículos de leyes enmarcadas al área de investigación, y revistas electrónicas de sitios oficiales con la finalidad de extraer información.

3.2.3.1.3 Recursos tecnológicos

Son indispensables como el uso de laptop, computadora de escritorio, internet, dominios, hosting, base de datos, herramientas de programación entre otros.

3.2.3.2. Presupuesto del proyecto

El presente presupuesto detalla los costos estimados para el diseño, desarrollo e implementación de un aplicativo de predicción de la producción de banano en la Hacienda San Juan, utilizando técnicas de machine learning. El sistema analizará datos recopilados mediante sensores y estaciones

meteorológicas para estimar la cantidad de banano producida, optimizando la planificación y gestión de los recursos en el cultivo.

Para la realización del presente proyecto se prevé una inversión total de \$2.334,00 la cual será destinada para cubrir gastos de dispositivos para el desarrollo del aplicativo web.

3.2.4 Métodos y técnicas de investigación

Los métodos empleados fueron el método deductivo e inductivo, mientras que en las técnicas se empleó la técnica de la entrevista.

Método cuantitativo: La presente investigación se fundamenta en un enfoque cuantitativo y descriptivo, orientado a recopilar, analizar y registrar datos sobre la producción de banano y las variables ambientales que inciden en ella en la Hacienda San Juan. La información se obtendrá mediante la consulta de fuentes bibliográficas, entrevistas con los trabajadores y responsables de la hacienda, y la revisión de estudios previos sobre la aplicación de técnicas de machine learning en la agricultura, con especial énfasis en cultivos de banano.

Técnica de entrevista: La técnica de la entrevista se utilizó para obtener información directa y detallada de los actores involucrados en la producción del cultivo de banano en la Hacienda San Juan. Se aplicará la entrevista que permitirán explorar en profundidad las experiencias, opiniones y conocimientos del entrevistado sobre los desafíos actuales y las necesidades en la recolección de datos. Esta metodología facilitará la identificación de aspectos críticos en el proceso productivo y la evaluación de la viabilidad de implementar un sistema de predicción basado en machine learning, asegurando que el aplicativo se ajuste a las demandas reales del entorno agrícola.

En este caso se realizó una entrevista a Ingeniero William Franco de la Hacienda San Juan se centró en los procesos de producción de banano, los desafíos actuales y el potencial uso de tecnología, como se muestra en el (Anexo N° 5: Modelo de entrevista), se crearon 13 preguntas que abarcaron las siguientes secciones:

- **Sección 1- Procesos de producción y factores de influencia:** En esta el entrevistado describió que procesos maneja para la producción de banano y que factores influyen en estos procesos.

- **Sección 2- Verificación y desafíos:** Se explica como se realiza el proceso de estimación de la producción de banano y los desafíos actuales por los que atraviesa la hacienda bananera.
- **Sección 3- Incorporación de tecnologías:** El ingeniero explica como cree que la tecnología ayudaría a posicionar a la hacienda en el mercado competitivo. Y como permitiría mejorar al proceso de toma de decisiones.

El proceso culmina con un contraste directo entre los registros históricos de producción y las proyecciones generadas por el modelo de aprendizaje automático. Esta validación cruzada no solo busca certificar la exactitud del sistema, sino asegurar que el aplicativo sea una guía confiable para la operación real. Al implementar este enfoque en la Hacienda San Juan, se dota a la administración de un recurso estratégico para optimizar los ciclos de cosecha, transformando los datos en una ventaja competitiva que mejora significativamente la toma de decisiones en el sector bananero. (Ver Anexo N°7).

3.2.5 Población y muestra

3.2.5.1. Población

En este trabajo, el objeto de estudio se aparta de la observación de sujetos o trabajadores para enfocarse íntegramente en el ecosistema productivo de la Hacienda San Juan. Al ser un proyecto orientado al desarrollo de software predictivo, la "población" está representada por el conjunto de variables ambientales y del suelo que dictan el crecimiento del banano. El proceso se centra en comprender cómo factores como la precipitación, la presión atmosférica y la luminosidad, junto a la composición química del terreno (N, P, K, pH y minerales), interactúan entre sí. Esta información no se analiza como estadística social, sino como el insumo técnico fundamental para que el sistema aprenda a anticipar los volúmenes de producción con base en las condiciones reales del entorno.

3.2.5.2. Muestra

Dado que el objetivo primordial es el entrenamiento de un modelo predictivo, esta investigación prescinde del muestreo estadístico tradicional para centrarse en el uso de la totalidad de los registros históricos disponibles. En lugar de inferir comportamientos a partir de un subconjunto, se trabaja con el universo completo de datos de la Hacienda San Juan, integrando métricas semanales de producción, registros meteorológicos de la zona y análisis de suelo previos. Al

emplear series cronológicas completas, se garantiza que el modelo capture la variabilidad real del cultivo, eliminando el sesgo que podría introducir una muestra aleatoria y asegurando que la base de entrenamiento sea una representación fiel y exhaustiva de la dinámica agrícola.

3.2.6 Análisis estadístico

En el contexto de este proyecto de tesis, el análisis estadístico no fue aplicable en la fase inicial, ya que el número de personas entrevistadas no alcanzó el mínimo requerido para un análisis cuantitativo representativo. Los entrevistados para la recolección de datos fueron el Ing. William Franco, administrador de la Hacienda San Juan, y el entrevistador.

Análisis de la entrevista al administrador de la hacienda: Los hallazgos derivados de la entrevista revelan que la Hacienda San Juan opera bajo un modelo de estimación predominantemente empírico, donde la experiencia del personal y el conteo manual de racimos por hectárea dictan la planificación. Esta dependencia de la intuición genera una brecha crítica; mientras que el margen de error ideal no debería exceder el 5%, factores exógenos como las fluctuaciones climáticas y la presión de plagas provocan discrepancias que comprometen la rentabilidad. La ausencia de estaciones meteorológicas y la fragmentación de la información en hojas de cálculo básicas limitan la capacidad de respuesta preventiva, subrayando la urgencia de migrar hacia una gestión basada en datos que otorgue a la hacienda una ventaja competitiva y una base sólida para la toma de decisiones. (Ver Anexo N°6).

Análisis de la encuesta al administrador de la hacienda: Después de realizada la entrega del sistema web al administrador de la hacienda, se realizó una encuesta de satisfacción haciendo uso de la escala de Likert (1 al 5) al propio administrador quien es la persona que interactúa con el sistema. Se evaluaron criterios de usabilidad, utilidad del sistema y velocidad de respuesta. (Ver Anexo N° 6:).

Los resultados que se obtuvieron del análisis de esta encuesta reflejan la aceptación del usuario:

- **Funcionalidad del sistema:** El 100% de los encuestados indicó estar “Muy de acuerdo” en que el aplicativo cumple su propósito de estimar la producción futura. Asimismo, el registro de datos y la visualización de

información obtuvieron la máxima valoración, lo que confirma el adecuado desempeño técnico del sistema.

- **Usabilidad del aplicativo:** De igual manera, el 100% calificó la interfaz como intuitiva y de fácil navegación desde distintos dispositivos. En consecuencia, se valida el cumplimiento de los criterios de accesibilidad y diseño adaptable.

- **Utilidad para la gestión agrícola:** Por otra parte, las predicciones y reportes fueron valorados positivamente (100% entre “De acuerdo” y “Muy de acuerdo”), evidenciando su aporte a la planificación y a la toma de decisiones administrativas.

Con estos resultados se puede concluir, el sistema cumple con las expectativas del usuario final, demostrando ser una herramienta tecnológica efectiva que optimiza la planificación productiva, reduce la incertidumbre en la estimación de cajas procesadas y fortalece la toma de decisiones basada en datos dentro de la hacienda. (Ver Anexo N° 8:

Análisis del algoritmo de machine learning Extreme Gradient Boosting (XgBoost): Con el fin de encontrar la herramienta de predicción más exacta, se seleccionaron cuatro técnicas de Machine Learning para ser puestas a prueba: Regresión Lineal, Random Forest, SVM y Xgboost el desarrollo se enfocó en obtener métricas detalladas que permitieron comparar el rendimiento real de cada una bajo las condiciones de la hacienda. Al contrastar estos resultados, fue posible elegir con criterio técnico el modelo Xgboost como el que mejor se adaptó a las variables de producción. (Ver Figura 8.)

El análisis estadístico del modelo seleccionado presento las siguientes métricas de desempeño sobre el conjunto de datos de prueba:

- **Error Absoluto Medio (MAE):** Mide el promedio de los errores absolutos entre los valores reales y predichos. En el modelo XGBoost se obtuvo un MAE de 0.342, evidenciando una baja desviación promedio en la estimación de la producción.

- **Error Cuadrático Medio (MSE):** Calcula el promedio de los errores al cuadrado, penalizando con mayor intensidad las desviaciones grandes. XGBoost registró un MSE de 0.241, reflejando estabilidad en las predicciones.

- **Raíz del Error Cuadrático Medio (RMSE):** Expresa el error en la misma escala de la variable analizada. Con un RMSE de 0.491, XGBoost demuestra alta precisión en la estimación productiva.
- **Coefficiente de Determinación (R^2):** Indica la proporción de la variabilidad explicada por el modelo. El valor obtenido ($R^2 = 0.89$) confirma una elevada capacidad explicativa.
- **Mediana del Error Absoluto (MedAE):** Representa el error absoluto central y es robusta ante valores atípicos. El resultado de 0.319 evidencia consistencia en las predicciones típicas del modelo.

Si bien el modelo Random Forest presentó un desempeño competitivo, los resultados cuantitativos evidencian que XGBoost alcanzó los menores valores de error y el mayor coeficiente de determinación. Esta superioridad indica una mejor capacidad para modelar relaciones no lineales entre las variables productivas, climáticas y de suelo, lo cual resulta fundamental en el contexto agrícola. En consecuencia, la implementación de XGBoost permite generar predicciones más precisas y confiables, validando la hipótesis de que el uso de algoritmos avanzados de Machine Learning contribuye significativamente a optimizar la planificación y la toma de decisiones en el cultivo de banano.

4. RESULTADOS

4.1 Análisis los factores ambientales y del suelo que influyen en la producción de banano en la Hacienda San Juan para la identificación de patrones que sirvan como base de los modelos predictivos.

El objetivo específico de estudiar los elementos del ambiente y del suelo relacionados con la producción de banano en la Hacienda San Juan fue logrado por medio de combinar la historia de producción con información meteorológicas y suelo obtenido para la región. Para ello, se aplicó el proceso descrito en la metodología KDD, siguiendo las etapas de selección, preprocesamiento, transformación y análisis exploratorio de datos.

En una primera fase, se adquirieron los históricos de la finca relacionados con la producción de banano (semanal) en las tarjetas (racimos y cajas procesadas), que se dividieron en lotes temporales. Paralelamente se seleccionaron las variables ambientales de interés (temperatura promedio, temperatura máxima, precipitación acumulada, humedad relativa y radiación) y las variables relacionadas con las características del suelo (pH, conductividad eléctrica, concentraciones de N, P, K y micronutrientes). Toda esta información fue integrada en una única base de datos, organizada en semanas, que proporcionó una ligazón directa entre las condiciones ambientales y del suelo y la producción.

A continuación, se efectuó un proceso de limpieza y depuración dos datos corrigiendo valores discrepantes y tratando registros incompletos unificando as unidades de medida. Este fue un paso importante para asegurar la calidad del dato que luego sería usada en el análisis y en el entrenamiento de los modelos de aprendizaje automático. En adición, se derivaron variables, incluyendo promedios móviles de precipitación y temperatura, y tipo de índice de estacionalidad para capturar patrones cíclicos en productividad.

Este proceso permitió identificar las variables con mayor peso estadístico sobre la producción de banano, estableciendo una jerarquía de importancia para el modelado. El análisis reveló que periodos de precipitación moderada y temperaturas constantes favorecen el rendimiento, mientras que la volatilidad en el régimen de lluvias impacta negativamente en la cosecha. La integración de parámetros edáficos, específicamente el pH y los niveles de N, P, K, permitió

refinar el conjunto de datos, eliminando el ruido y seleccionando únicamente los atributos más representativos. El resultado es un data set depurado y alineado con la realidad productiva de la Hacienda San Juan, lo que garantiza que los algoritmos de Machine Learning se fundamenten en los patrones reales de comportamiento observados en el campo.

4.1.1 Factores ambientales relevantes

El análisis de correlación evidenció que la temperatura media, la humedad relativa y la precipitación acumulada presentan una relación significativa con la producción semanal de banano. A partir de la matriz de correlación (Ver Figura 13.) se identificó que la temperatura media presenta una correlación positiva moderada con las cajas procesadas ($r = 0.39$), mientras que la velocidad del viento también muestra una relación moderada ($r = 0.45$), lo que indica su influencia en el rendimiento productivo.

En el caso de la precipitación, se observa una correlación positiva moderada ($r = 0.34$), lo que sugiere que niveles adecuados de lluvia favorecen el desarrollo del cultivo. Por su parte, la humedad relativa presenta correlaciones positivas débiles a moderadas (r entre 0.24 y 0.29), indicando una influencia presente pero menos determinante en comparación con otras variables.

Estos factores influyen directamente en el desarrollo fisiológico del cultivo, especialmente en procesos como la fotosíntesis y la acumulación de biomasa. Se identificaron rangos óptimos en los cuales estas variables favorecen el rendimiento productivo. Sin embargo, cuando se presentan valores fuera de estos rangos durante periodos prolongados, se observa una disminución considerable en la producción, reflejada en el número de cajas procesadas.

Adicionalmente, se identificó una correlación negativa baja entre la radiación solar y la producción ($r = -0.08$), lo que sugiere que niveles elevados de radiación podrían generar efectos de estrés en determinadas condiciones.

Estos resultados indican que la gestión agronómica no debe basarse únicamente en promedios, sino también en el monitoreo constante de las variaciones climáticas. De esta manera, es posible anticipar condiciones adversas y tomar medidas oportunas que reduzcan el impacto negativo en el cultivo.

4.1.2 Factores del suelo relevantes

El análisis de las variables edáficas mostró que la disponibilidad y el equilibrio de nutrientes como nitrógeno (N), fósforo (P) y potasio (K), junto con un nivel adecuado de pH, tienen una influencia en la estabilidad de la producción. A partir de la matriz de correlación multivariable (ver Figura 14.), se identificó que algunos de estos nutrientes presentan una relación más significativa con la producción en comparación con análisis previos.

En particular, el potasio (K) muestra una correlación positiva moderada con las cajas procesadas ($r = 0.64$), lo que indica su importancia en el desarrollo del cultivo. Asimismo, el nitrógeno (N) presenta una correlación positiva moderada ($r = 0.48$), mientras que el fósforo (P) evidencia una relación positiva más baja ($r = 0.33$).

Un hallazgo relevante es la alta correlación positiva entre la humedad del suelo y la producción ($r = 0.75$), lo que sugiere que este factor es uno de los más influyentes en el rendimiento del cultivo, ya que impacta directamente en la disponibilidad de nutrientes y en los procesos fisiológicos de la planta.

Por otro lado, la conductividad eléctrica presenta una correlación positiva moderada ($r = 0.35$), indicando su relación con la concentración de sales y nutrientes en el suelo. En contraste, el pH del suelo muestra una correlación muy baja con la producción ($r = 0.11$), lo que sugiere una influencia limitada dentro del modelo analizado.

Además, se identificaron correlaciones importantes entre los propios nutrientes del suelo, como entre nitrógeno y conductividad eléctrica ($r = 0.72$) y entre potasio y conductividad eléctrica ($r = 0.61$), lo que evidencia una interdependencia en la dinámica química del suelo.

Estos hallazgos resaltan la importancia de considerar las condiciones del suelo como un factor clave dentro del modelo predictivo. De este modo, la fertilidad del suelo deja de ser un elemento estático y pasa a ser una variable dinámica que contribuye significativamente a mejorar la precisión de las estimaciones de producción y a optimizar la toma de decisiones en el manejo del cultivo

4.2 Diseño el modelado de la producción del cultivo de banano a través de los diagramas UML para la correcta elaboración de todos los módulos del software.

El objetivo específico de diseñar el modelado de la producción del cultivo de banano a través de los diagramas UML se cumplió mediante la elaboración de un conjunto de diagramas que representan tanto la estructura como el comportamiento del sistema propuesto. Estos diagramas permitieron traducir los requerimientos identificados con el administrador de la Hacienda San Juan en una arquitectura de software organizada, coherente y alineada con los módulos definidos en la propuesta tecnológica.

En primera instancia, se elaboraron los diagramas de contexto y DFD nivel 1, que ofrecen una visión global del sistema y de los principales procesos involucrados en la gestión de la producción de banano. El diagrama de contexto representa al aplicativo web de predicción como el proceso central, describiendo la interacción con los actores externos, principalmente el administrador de la hacienda. El DFD de nivel 1 detalla la descomposición de este proceso en funciones específicas relacionadas con la captura de datos, procesamiento, modelado, visualización y generación de reportes, definiendo los flujos de información entre ellos.

La delimitación funcional del sistema se consolidó mediante el diseño de Diagramas de Casos de Uso, herramientas que permitieron mapear con precisión la interacción entre los actores y el aplicativo. Se priorizaron escenarios críticos como la gestión de registros por lote, la ingesta de datos meteorológicos y edáficos, y el núcleo del sistema: la generación de predicciones y la visualización comparativa entre la cosecha real y la proyectada. Esta fase fue determinante para aterrizar los requisitos funcionales y asegurar que el software responda a las necesidades operativas de la hacienda.

Posteriormente, la arquitectura interna se estructuró a través de un Diagrama de Clases, donde entidades como Usuario, Lote, Producción y Modelo Predictivo fueron definidas con sus respectivos atributos y métodos. Este diseño no solo garantiza la trazabilidad de la información entre módulos, sino que establece un puente directo con el modelo entidad-relación. Al alinear estas clases con la base de datos en MySQL, se asegura una implementación técnica

fluida, facilitando el mantenimiento y la integridad de los datos históricos del cultivo.

Finalmente, se elaboraron diagramas de secuencia para los procesos críticos, entre ellos el flujo de inicio de sesión, el registro de producción y la ejecución del modelo de predicción. Estos diagramas ayudaron a precisar el orden de los mensajes entre las clases y componentes del sistema, asegurando una correcta coordinación entre la capa de presentación, la lógica de negocio y el acceso a datos. (Ver Anexo N°20).

Como resultado de este objetivo, se obtuvo un modelado UML completo del sistema, que incluye diagramas de contexto, DFD, casos de uso, clases, entidad relación y secuencia. Este conjunto de diagramas no solo respalda la correcta elaboración de todos los módulos del software, sino que también facilita la comunicación entre el desarrollador, el tutor y los responsables de la hacienda, al ofrecer una representación clara y estandarizada de la solución propuesta.

4.3 Desarrollo un aplicativo web que integre modelos de Machine Learning con el propósito de predecir la producción de banano, facilitando el control y la planificación del cultivo.

El objetivo específico de desarrollar un aplicativo web que integre modelos de Machine Learning para predecir la producción de banano se alcanzó mediante la implementación de un sistema web modular, basado en tecnologías de código abierto, que materializa la arquitectura diseñada en los diagramas UML y la metodología XP descrita en el capítulo de materiales y métodos. (Ver Anexo N°8).

El aplicativo fue desarrollado utilizando PHP 8.3 como lenguaje principal del lado del servidor, HTML 5.3, CSS y Bootstrap para la capa de presentación, y MySQL como sistema gestor de base de datos. La adopción de estas herramientas permitió construir una solución accesible, escalable y coherente con los recursos tecnológicos disponibles en la Hacienda San Juan. La estructura del sistema se organizó en módulos claramente delimitados, entre los cuales se destacó el módulo de gestión y administración, módulo de captura y procesamiento de datos, módulo de modelado y predicción, módulo de visualización y el módulo de reportes.

Durante el desarrollo del aplicativo para la predicción de la producción de banano en la Hacienda San Juan, se siguieron las iteraciones propuestas por la

metodología XP, implementando de manera progresiva las funcionalidades hasta obtener un sistema completamente funcional. (Ver Anexo N°19). En la primera iteración, se desarrolló un prototipo inicial que incorporó el módulo de autenticación y control de acceso, permitiendo la gestión de usuarios con roles definidos, así como las configuraciones básicas de seguridad; se realizaron las respectivas pruebas de funcionalidad, que permitieron verificar el funcionamiento correcto. En la segunda iteración, se implementaron los módulos de registro y gestión de datos productivos y meteorológicos, incluyendo el almacenamiento, validación, actualización y consulta de información relacionada con variables edafoclimáticas y registros históricos de producción. Posteriormente, en la tercera iteración, se integró el modelo de Machine Learning para la predicción de la producción, realizando el preprocesamiento de datos, selección de variables y entrenamiento del modelo más adecuado según los resultados obtenidos. Finalmente, en la última iteración, se incorporaron funcionalidades complementarias como la visualización gráfica de resultados y reportes predictivos, garantizando así un sistema integral orientado a la toma de decisiones agrícolas.

Como resultado, se obtuvo un aplicativo web funcional que centraliza los datos de producción, variables meteorológicas y suelo, integra modelos de Machine Learning para predecir la producción de banano y ofrece herramientas de visualización y reporte que facilitan el control y la planificación del cultivo.

5. DISCUSIÓN

En el estudio realizado por Sanga et al. (2020) en Tanzania, se pone de manifiesto el gran potencial del aprendizaje profundo en un entorno accesible para los agricultores. Se trata de una aplicación para Android que puede diagnosticar la Sigatoka negra y otras plagas con una precisión impresionante, cercana al 99%, simplemente a través de imágenes del cultivo. Este enfoque demuestra que la inteligencia artificial puede ser utilizada en áreas rurales con limitaciones tecnológicas, mejorando la vigilancia fitosanitaria y ayudando a reducir las pérdidas que se producen por diagnósticos tardíos. Sin embargo, mientras que este estudio se centra principalmente en la detección y clasificación visual de plagas, la investigación actual se enfoca en anticipar el rendimiento productivo del banano analizando variables edafoclimáticas y registros de producción (como racimos y cajas). Esto proporciona una herramienta predictiva que ayuda en la planificación y toma de decisiones antes de que se presenten daños o variaciones en el rendimiento.

Por otro lado, la investigación de Caraguay y Cordero (2020) resalta la implementación de un sistema móvil para la supervisión del banano, destacando su viabilidad económica gracias al uso de hardware asequible. Su propuesta resulta especialmente útil en las etapas de corte y embalaje, ya que permite recopilar datos por lote, almacenarlos en una base de datos y generar informes que identifican anomalías relacionadas con plagas o problemas operativos. En comparación, esta tesis subraya la importancia de la digitalización y la trazabilidad de datos, pero va un paso más allá: busca transformar esos registros en modelos predictivos que estimen el comportamiento de la producción y faciliten decisiones preventivas, disminuyendo la necesidad de supervisión constante y fortaleciendo la gestión basada en evidencia.

En el contexto latinoamericano, Hinojosa y su equipo (2023) destacan el impacto de AgroTIC en Santander, Colombia. Esta plataforma combina inteligencia artificial con una red de comunicación entre productores, técnicos y comerciantes, logrando implementarse en más de 130 fincas y beneficiando a más de 200 agricultores. Este caso demuestra que las soluciones digitales son más efectivas cuando se integran la tecnología con la colaboración y el intercambio de información entre los actores del sistema productivo. A diferencia

de este enfoque regional y colaborativo, la propuesta actual se enfoca en un contexto local ecuatoriano y en las necesidades específicas del manejo productivo, buscando precisión a partir de variables concretas del cultivo. Sin embargo, los resultados de AgroTIC respaldan la posibilidad de escalar la solución hacia modelos donde la predicción también nutra redes de asistencia técnica y comercialización.

Por otro lado, la aplicación global Plantix, mencionada por Urquijo (2020), se ha convertido en una herramienta muy utilizada por los agricultores, ya que permite identificar plagas o enfermedades simplemente subiendo una fotografía. Esta app se apoya en una extensa base de datos (que abarca más de 30 cultivos) y ofrece recomendaciones de tratamiento junto con una comunidad de apoyo. Este ejemplo muestra que los sistemas móviles con diagnóstico rápido pueden tener un impacto directo en la reducción de pérdidas. Sin embargo, al ser una plataforma generalista, su fortaleza radica en su amplia cobertura. En contraste, la investigación actual busca especializar el análisis en el cultivo de banano y en las condiciones locales, priorizando la predicción del rendimiento y la gestión productiva (más allá del diagnóstico). Esto permite ajustar los modelos a las características edafoclimáticas y operativas propias del entorno ecuatoriano.

A nivel local, el estudio de García et al. (2022) en el cantón Balao revela un problema serio: la degradación del suelo bananero debido a la explotación intensiva, lo que afecta su calidad química y biológica, aumenta los residuos y reduce la materia orgánica. Este hallazgo es fundamental porque demuestra que el rendimiento no solo depende de las prácticas de cosecha o del control sanitario, sino también del estado del suelo y su dinámica. En este contexto, la presente tesis se basa directamente en esta problemática, proponiendo el uso de variables edafoclimáticas como apoyo para anticipar los niveles de rendimiento, ofreciendo una vía tecnológica para tomar decisiones más oportunas y sostenibles sobre el manejo del cultivo.

La tecnificación administrativa que describen Cruz y Pillco (2020) en la hacienda “San Jacinto” muestra que un sistema web desarrollado con la metodología RUP puede transformar la rentabilidad y la transparencia al gestionar costos, roles de pago y trazabilidad, eliminando registros manuales y mejorando la disponibilidad de datos para decisiones estratégicas. Este antecedente pone de manifiesto que la digitalización mejora el control y la

eficiencia operativa. Sin embargo, su enfoque principal está en la gestión administrativa y de procesos, mientras que esta investigación complementa esa perspectiva a través del componente analítico: convertir los datos productivos y ambientales en un sistema de predicción que apoye directamente la planificación agrícola y la optimización del rendimiento.

Por último, el proyecto de Chang (2024) en la Universidad Agraria del Ecuador, que se centra en el cacao de la Hacienda Victoria, es especialmente significativo porque combina variables climáticas, datos de producción y registros de plantaciones en una aplicación web de código abierto. Tras realizar un análisis de correlación y una evaluación comparativa, se eligió Random Forest como modelo principal, que tiene la capacidad de hacer estimaciones con hasta cinco meses de antelación. Este trabajo subraya la importancia de evaluar varios algoritmos y seleccionar el más robusto basándose en métricas. Por el contrario, esta investigación aplica la misma lógica metodológica al sector bananero, centrándose en variables de producción como racimos y cajas, así como en el comportamiento del suelo y el clima, con el objetivo de generar predicciones útiles para la gestión local. De esta manera, Chang (2024) establece un claro precedente sobre cómo construir un sistema predictivo agrícola, mientras que esta propuesta extiende su aplicación al cultivo del banano, abordando las necesidades específicas del entorno ecuatoriano.

6. CONCLUSIONES Y RECOMENDACIONES

6.1 Conclusiones

A través del análisis de las variables ambientales y edáficas, se determinó que el rendimiento del banano responde a patrones climáticos y nutricionales específicos. Factores como la humedad y el régimen pluvial dictan el ritmo de cosecha, mientras que la salud del suelo medida a través del N, P, K y la acidez (pH) garantiza la resiliencia productiva de cada sección. Al mapear estas interacciones, se logró consolidar una matriz informativa robusta, la cual sirve como el cimiento sobre el cual se orquesta la arquitectura del sistema predictivo, asegurando que el software esté alineado con la realidad biológica de la hacienda. (Ver Figura 14.)

Además, el diseño del modelado a través de diagramas UML facilitó una estructura clara y organizada de los procesos, interacciones y componentes del sistema de predicción. La creación de diagramas de contexto, DFD, casos de uso, clases, entidad-relación y secuencia proporcionó una arquitectura coherente con los requerimientos identificados en la fase de análisis, asegurando que la propuesta tecnológica se ajuste adecuadamente a las necesidades operativas de la Hacienda San Juan.

Consecuentemente, la conceptualización del aplicativo web enfocado en la predicción de la producción de banano se consolidó gracias a la información recopilada en las etapas anteriores. La definición de módulos para la captura, procesamiento y visualización de datos permitió establecer un diseño funcional que sienta las bases para futuras fases de desarrollo, demostrando que la integración de información climática y edáfica en un entorno digital fortalece la toma de decisiones en el ámbito agrícola.

En resumen, los resultados de esta tesis muestran que la creación e implementación de una aplicación web basada en un modelo de aprendizaje automático es una solución innovadora y eficaz para estimar la producción de plátanos. Al combinar datos históricos de suelo y edafoclimáticos, tanto en tiempo real como registrados previamente, el modelo que proponemos puede ofrecer previsiones de la producción futura, racimos o cajas. Esta herramienta proporciona a los administradores y productores un valioso apoyo para la toma de decisiones, ya que reduce la incertidumbre relacionada con las fluctuaciones

de la producción, facilita la planificación operativa y optimiza el uso de los recursos. Como resultado, se mejora la eficiencia y la sostenibilidad del sistema de producción y de la cadena de valor del plátano.

6.2 Recomendaciones

A partir de los resultados del análisis de factores ambientales y del suelo, se recomienda fortalecer la recolección continua y estandarizada de datos climáticos y de suelo en la Hacienda San Juan. Tener un sistema de registro más frecuente y preciso ayudará a entender mejor los patrones de producción y a aumentar la fiabilidad de los modelos predictivos futuros que se basen en estas variables.

Además, se sugiere mantener y mejorar el uso de herramientas de modelado UML en las próximas etapas del proyecto, ya que su aplicación facilita la comprensión del flujo de información y asegura un diseño estructurado del sistema. Actualizar constantemente los diagramas a medida que evolucionen los requerimientos del usuario contribuirá a garantizar la coherencia entre el análisis, el diseño y el desarrollo posterior del aplicativo.

Es fundamental seguir utilizando datos históricos sobre la producción de plátanos para entrenar, validar y comparar modelos predictivos. Es importante dar prioridad a los enfoques de aprendizaje automático que pueden captar relaciones no lineales y patrones complejos. En el contexto de esta tesis, el uso de XGBoost es especialmente relevante, ya que permite realizar modelos de producción muy precisos basados en variables de nutrientes de suelo y edafoclimáticas, mejorando el rendimiento mediante técnicas de regularización y ajuste de hiperparámetros. La elección del modelo final debe basarse en métricas objetivas de error y bondad de ajuste, como MAE, RMSE y R^2 , para garantizar predicciones fiables y coherentes. Este enfoque no solo aumenta la precisión de las previsiones de producción, sino que también refuerza la toma de decisiones del gestor, facilitando una planificación más estratégica, una asignación eficiente de los recursos y una gestión basada en datos empíricos dentro del sistema de producción de plátanos.

Por último, se recomienda avanzar hacia la implementación del aplicativo web diseñado en esta investigación, incorporando gradualmente los módulos definidos en la fase de diseño. También se propone considerar futuras líneas de investigación que integren nuevas variables agronómicas, como el manejo de

plagas, el estrés hídrico y las prácticas culturales, con el objetivo de ampliar las capacidades analíticas del sistema y crear una herramienta más completa para la planificación agrícola de la Hacienda San Juan.

BIBLIOGRAFÍA

- Almeyda, E. (2022). Pronóstico de la demanda internacional del banano orgánico de Perú usando algoritmos de Machine Learning. Obtenido de <https://pirhua.udep.edu.pe/item/13fd39df-ecea-46c6-a3e2-d90f09923c42>
- Belupú Amaya, C. (2022). Propuesta de una plataforma de agricultura inteligente basada en IoT para el monitoreo de las condiciones climáticas del cultivo de banano. Obtenido de <https://pirhua.udep.edu.pe/item/3f254ba7-a6d2-4096-b124-af8cdef9e64a>
- Borja, J., & Borja, J. (2016). La producción de banano bajo el sistema de comercio justo: un análisis del caso ecuatoriano. *Siembra*, 3(1), 7–10. <https://doi.org/10.29166/SIEMBRA.V3I1.185>
- Caraguay, W. A., & Cordero, J. (2020). Aplicación móvil integrada para la supervisión del proceso de producción del banano. Obtenido de <https://revistas.utmachala.edu.ec/revistas/index.php/Cumbres/article/view/343>
- Chang, S. L. (2024). Implementación de una aplicación web para la estimación de la productividad del cacao nacional en la Hacienda Victoria. Obtenido de <https://cia.uagraria.edu.ec/Archivos/CHANG%20ZORRILLA%20STEPHAN%20LEONEL.pdf>
- Constitución Política de la República del Ecuador. (2008). Constitución Política de la República del Ecuador. Obtenido de <https://pdba.georgetown.edu/Parties/Ecuador/Leyes/constitucion.pdf>
- CRUZ ALVARIO, A. S., & PILLCO PÉREZ, J. (2020). IMPLEMENTACIÓN DE UN SISTEMA WEB PARA LA OPTIMIZACIÓN DE PRODUCCIÓN EN LA HACIENDA BANANERA “SAN JACINTO”. Obtenido de <https://cia.uagraria.edu.ec/Archivos/CRUZ%20ALVARIO%20ALEJANDRA%20STEFFANIE.pdf>
- D’Angiolo, F. G., Kwist, I. F., Loiseau, M., Contreras, D. E., & Asteasuain, F. (2020). Algoritmos de regresión lineal aplicados al mantenimiento de un datacenter. *XXV Congreso Argentino de Ciencias de la Computación CACIC* 2019, 1354–1363. <https://undavdigital.undav.edu.ar/xmlui/handle/20.500.13069/3497>

- EOS DATA ANALYTICS. (2024). Obtenido de https://eos.com/es/blog/agrometeorologia-y-tiempo-agricultura/?utm_source=chatgpt.com
- Esquivel-Paula, G. G., Quisaguano-Collaguazo, L. R., Caluña-Guaman, A. P., & Llambo-Alvarez, S. J. (2025). Frameworks del lado del servidor: Caso de estudio Node JS, Django y Laravel. *593 Digital Publisher CEIT, 10*(1), 403–414. <https://doi.org/10.33386/593dp.2025.1.2729>
- Fernández Flórez, & Hernández Jorge. (2021). Aplicaciones web con Php - Héctor Flórez Fernández, Jorge Hernández Rodríguez - Google Libros. [https://books.google.com.ec/books?hl=es&lr=&id=Zb0xEAAAQBAJ&oi=fnd&pg=PR14&dq=Fl%C3%B3rez+Fern%C3%A1ndez,+H.,+%26+Hern%C3%A1ndez+Rodr%C3%ADguez,+J.+\(2021\).+Aplicaciones+web+con+PHP.+Bog%C3%B3ta,+Colombia:+Ediciones+de+la+U.+doi:978-958-792-234-9+&ots=cuVJb7W8jr&sig=2J0Aj-ubfkPraXh1-FdLoFWPH-s&redir_esc=y#v=onepage&q&f=false](https://books.google.com.ec/books?hl=es&lr=&id=Zb0xEAAAQBAJ&oi=fnd&pg=PR14&dq=Fl%C3%B3rez+Fern%C3%A1ndez,+H.,+%26+Hern%C3%A1ndez+Rodr%C3%ADguez,+J.+(2021).+Aplicaciones+web+con+PHP.+Bog%C3%B3ta,+Colombia:+Ediciones+de+la+U.+doi:978-958-792-234-9+&ots=cuVJb7W8jr&sig=2J0Aj-ubfkPraXh1-FdLoFWPH-s&redir_esc=y#v=onepage&q&f=false)
- García, Y., García, Y., León, D. P. de, Mancero, D., & Pancho, T. (2022). Incidencia del manejo de plantaciones comerciales del cultivo de banano sobre la calidad del suelo, Balao, Guayas. *Polo del Conocimiento, 7*(10), 2011–2027. <https://doi.org/10.23857/pc.v7i10.4827>
- González-Palacio, M., María Sepúlveda-Cano, L., Quiza-Montealegre, J., & Amato, J. D. '. (2020). Mejoramiento del algoritmo ADR en una red de Internet de las Cosas LoRaWAN usando Aprendizaje de Máquina. *RISTI: Revista Ibérica de Sistemas e Tecnologias de Informação, ISSN-e 1646-9895, No. 39, 2020, págs. 67-83, 39, 67–83.* <https://doi.org/10.17013/risti.39.67-83>
- Guamán Daquilema, S. E., & Escudero Villa, A. I. (2020). FACTORES QUE INFLUYEN EN LA PRODUCCIÓN DEL PLÁTANO EN EL ECUADOR, 2014-2016. *Ciencia Digital, 2*(4.1.), 97–107. <https://doi.org/10.33262/CIENCIADIGITAL.V2I4.1.193>
- Hinojosa, C., Sanchez, K., Camacho, A., & Arguello, H. (2023). AgrotIC: Bridging the gap between farmers, agronomists, and merchants through smartphones and machine learning. <http://arxiv.org/abs/2305.12418>
- Iniciativa Aporta. (2024, noviembre). *Guía práctica de introducción al análisis exploratorio de datos en Python.* [datos.gob.es.](https://datos.gob.es)

- https://datos.gob.es/sites/default/files/documentacion/files/guia_eda_python.pdf
- Instituto Nacional De Estadísticas y Censos. (2020). Obtenido de <https://anda.inec.gob.ec/anda/index.php/catalog/912>
- IONOS. (2024). ¿Qué es el hosting? Las claves del alojamiento web. <https://www.ionos.com/es-us/digitalguide/hosting/cuestiones-tecnicas/que-es-el-hosting/>
- Jacinto Parinango, E. A. (2022). Análisis de los sistemas de gestión de base de datos relacionales con marcos de trabajo para procesamiento de datos masivos.
- Juma Alba, A. P., Pusda Aguilera, A. R., & Alvarez Varona, F. (2025). Gestión del talento humano: Aplicativo web para el análisis del desempeño laboral. *Revista Científica Dejando Huellas*, 4(2), 63–76. <https://doi.org/10.65100/recidh/12>
- León Soberón, J. J. (2020). Análisis comparativo de sistemas gestores de bases de datos postgresql y mysql en procesos crud.
- Ley de Fomento y Desarrollo Agropecuario. (2021). Ley de Fomento y Desarrollo Agropecuario. Obtenido de <http://www2.competencias.gob.ec/wp-content/uploads/2021/02/13.-Ley-de-Fomento-y-Desarrollo-Agropecuario.pdf>
- Ley de Propiedad Intelectual. (2022). Ley de Propiedad Intelectual. Obtenido de http://www.sice.oas.org/int_prop/nat_leg/ecuador/l320a.asp#:~:text=sobre%20sus%20obras.-,Art.,destino%20o%20modo%20de%20expresi%C3%B3n
- Magdama, F., Monserrate-Maggi, L., Serrano, L., Onofre, J., & Jimenez, M. (2020). Genetic Diversity of *Fusarium oxysporum* f. sp. cubense, the *Fusarium* Wilt Pathogen of Banana, in Ecuador. doi:<https://doi.org/10.3390/plants9091133>
- Ministerio de Agricultura y Ganadería. (2021). Ministerio de Agricultura y Ganadería. Obtenido de <file:///C:/Users/59399/Downloads/ACUERDO-095-2021-1.pdf>
- Nacional, U., Del, M., & Ángeles, P. (2021). Sistema de archivos, gestores de base de datos y Hadoop: ¿evolución o retroceso? *Revista digital*

- universitaria, 22(6).
<https://doi.org/10.22201/cuaieed.16076079e.2021.22.6.6>
- Plan Nacional de Desarrollo. (2021). Plan Nacional de Desarrollo. Obtenido de <http://www.eeq.com.ec:8080/documents/10180/36483282/PLAN+NACIO+NAL+DE+DESARROLLO+2021-2025/2c63ede8-4341-4d13-8497-6b7809561baf>
- Ramos-Galarza, C. (2021). Editorial: Diseños de investigación experimental. *CienciAmérica*, 10(1), 1–7. <https://doi.org/10.33210/CA.V10I1.356>
- Rodriguez, D. (2020). Investigación aplicada: características, definición, ejemplos. <https://www.lifeder.com/investigacion-aplicada/>
- Sanga, S. L., Machuve, D., & Jomanga, K. (2020). Mobile-based Deep Learning Models for Banana Disease Detection. *Engineering, Technology and Applied Science Research*, 10(3), 5674–5677. <https://doi.org/10.48084/ETASR.3452>
- Singh, S., Singh, D., & Bahadur, L. (2020). A REVIEW ON SUCCESSFUL PROTECTED CULTIVATION OF BANANA (MUSA). Obtenido de https://www.researchgate.net/publication/343960589_A_REVIEW_ON_SUCCESSFUL_PROTECTED_CULTIVATION_OF_BANANA_MUSA/references
- Sierra-García, J. E., & Santos, M. (2021). Redes neuronales y aprendizaje por refuerzo en el control de turbinas eólicas. *Revista Iberoamericana de Automática e Informática industrial*, 18(4), 327–335. <https://doi.org/10.4995/RIAI.2021.16111>
- Universidad Internacional de La Rioja. (2025). *¿Qué es CSS y cómo funciona en el desarrollo web?* UNIR. <https://www.unir.net/revista/ingenieria/que-es-css/>
- Urquijo, J. (2020). Greenapps&web website. Obtenido de <https://www.greenappsandweb.com/android/plagas-enfermedades-de-las-plantas-plantix/>
- Vite Cevallos, H., Carvajal Romero, H., Barrezueta Unda, S., Vite Cevallos, H., Carvajal Romero, H., & Barrezueta Unda, S. (2020). Aplicación de algoritmos de aprendizaje automático para clasificar la fertilidad de un suelo bananero. *Conrado*, 16(72), 15–19.

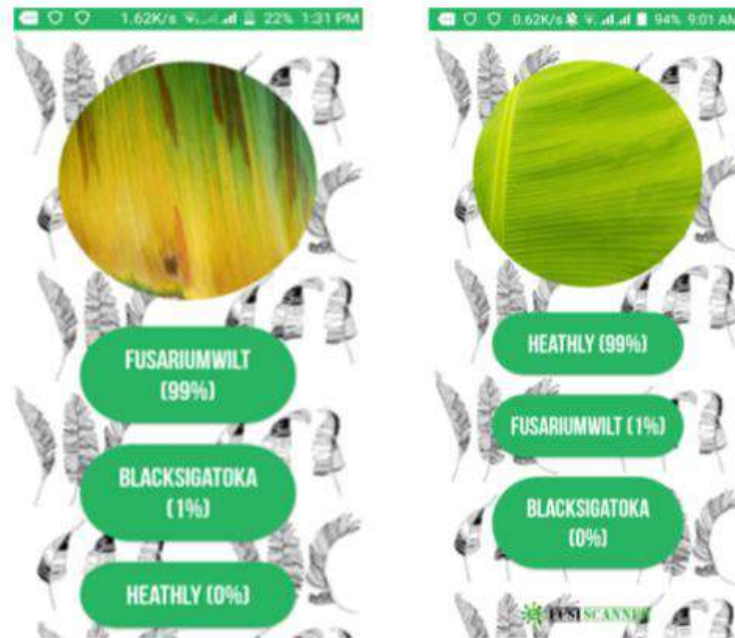
http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1990-86442020000100015&lng=es&nrm=iso&tlng=en

ANEXOS

Anexo N° 1: Aplicaciones relacionadas al tema

Figura 1.

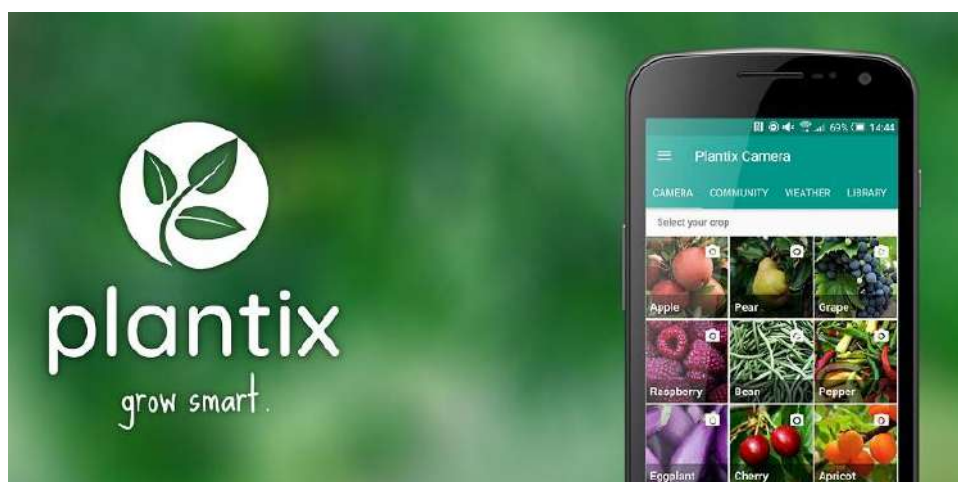
Aplicación detección de plagas



Fuente: Sanga et al, 2020

Figura 2.

App Plantix



Fuente: Plantix , 2020

Figura 3.
Agrotic

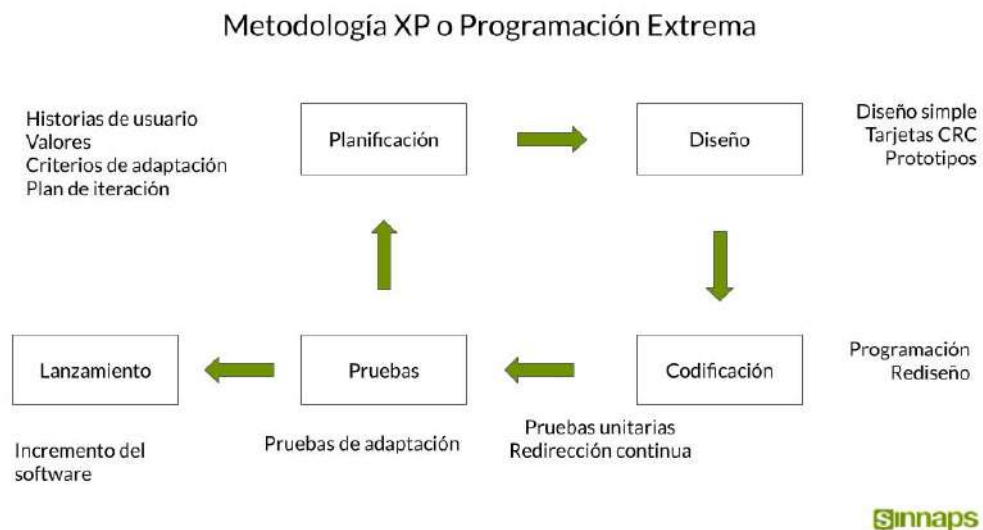


Fuente: Agrotic, 2023

Anexo N° 2: Fases de Metodología

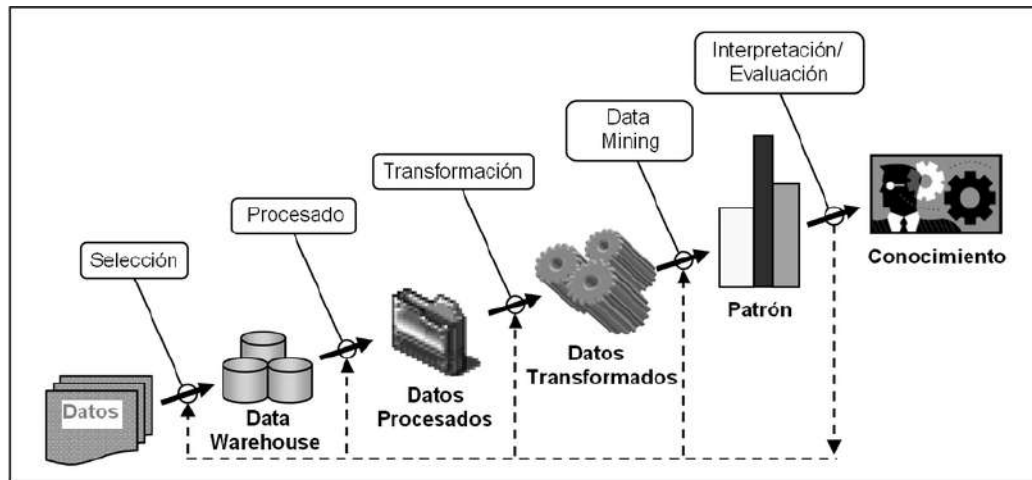
Figura 4.

Fases Metodología XP (Programación Extrema)



Fuente: Sinnaps, 2021

Figura 5.
Fases Metodología KDD

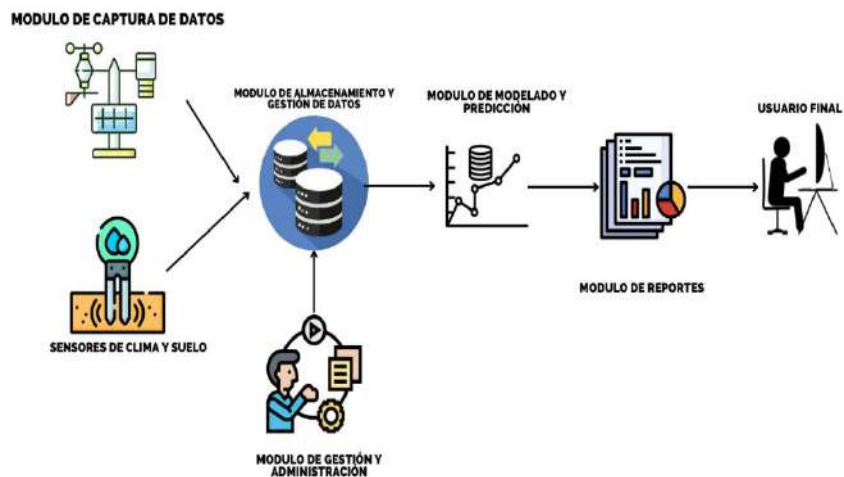


Fuente: Sanchez Guerrero, 2022

Anexo N° 3: Funcionamiento del aplicativo

Figura 6.

Ilustración de funcionamiento



Elaborado por: Anchundia, 2025

Anexo N° 4: Recursos**Tabla 1.*****Recursos Humanos***

Recursos	Rol	Meses	Valor	
			Unidad	Valor total
Andres Anchundia	Desarrollador y líder	10	\$0	\$0
Ing. Mitchell Vasquez	Tutora guía	10	\$0	\$0
Ing. William Franco	Administrador hacienda	10	\$0	\$0
			Total	\$0

Elaborado por: Anchundia, 2025

Tabla 2.***Recursos tecnológicos del software***

Recursos	Meses	Valor Unidad	Valor total	
Hosting	12	\$40	\$480	
Internet	8	\$30	\$240	
Dominio web	12	\$2	\$24	
CSS	8	\$0	\$0	
PHP	8	\$0	\$0	
JavaScript	8	\$0	\$0	
Bootstrap	8	\$0	\$0	
MySQL	8	\$0	\$0	
Impresora	1	\$200	\$200	
Laptop	1	\$550	\$550	
			Total	\$1494

Elaborado por: Anchundia, 2025

Tabla 3.

Recursos bibliográficos

Recursos	Valor Unidad	Valor total
Artículos científicos	\$0	\$0
Revistas electrónicas	\$0	\$0
Libros	\$0	\$0
Repositorios de tesis	\$0	\$0
Sitios web oficiales	\$0	\$0
Total		\$0

Elaborado por: Anchundia, 2025

Tabla 4.

Gastos Varios

Recursos	Valor Unidad	Meses	Valor total
Luz	\$25	8	\$200
Uso Laptop	\$50	8	\$400
Impresora	\$30	8	\$240
Total			\$840

Elaborado por: Anchundia, 2025

Tabla 5.

Valores totales de recursos

Recursos	Valor total
Recursos humanos	\$0
Recursos tecnológicos	\$1494
Recursos bibliográficos	\$0
Gastos varios	\$840
Total	\$2334

Elaborado por: Anchundia, 2025

Anexo N° 5: Modelo de entrevista**UNIVERSIDAD AGRARIA DEL ECUADOR
FACULTAD DE CIENCIAS AGRARIAS
CARRERA DE COMPUTACIÓN
Entrevista**

Fecha: 16 de julio del 2025

Entrevistadores: Anchundia, 2025

Entrevistado: Ing. William Franco

Objetivo: Obtener información cualitativa y detallada sobre los procesos productivos, las prácticas de gestión y los desafíos operativos en el cultivo de banano en la Hacienda San Juan, para la validación de la problemática de investigación y la recopilación de datos esenciales para el diseño y desarrollo del sistema de predicción de la producción basado en machine learning.

Preguntas:

1. ¿Cuál es su rol dentro de hacienda San Juan?
2. ¿Cuál es el trabajo que realiza y que consiste?
3. ¿Qué procesos productivos se manejan en la Hacienda San Juan para la producción de banano?
4. ¿Cuáles son los principales factores que influyen en la variabilidad de la producción de banano?
5. ¿Se utilizan actualmente herramientas tecnológicas para monitorear el crecimiento y desarrollo del cultivo?
6. ¿Cómo estiman la producción de banano antes de la cosecha? ¿Siguen algún método específico?
7. ¿Realizan una comparación de los valores obtenidos por los cálculos con la producción real?
8. ¿Existe un registro histórico de producción en la hacienda? ¿Cómo se almacena y gestiona esta información?
9. ¿Actualmente se consideran parámetros climáticos para realizar las estimaciones?
10. ¿Qué desafíos enfrentan en la planificación de la cosecha debido a cambios en las condiciones ambientales?
11. ¿Cómo afectan las plagas y enfermedades a la producción y qué estrategias emplean para su control?
12. ¿Qué tipo de sensores o estaciones meteorológicas utilizan actualmente para monitorear el clima?
13. ¿Cómo cree que un sistema de predicción basado en machine learning podría mejorar la gestión del cultivo?

Anexo N° 6: Encuesta de satisfacción

**UNIVERSIDAD AGRARIA DEL ECUADOR
FACULTAD DE CIENCIAS AGRARIAS
CARRERA COMPUTACIÓN
ENCUESTA DE SATISFACCIÓN**

Encuestador: Anchundia Zambrano Andres David

Encuestados: Administrador de la hacienda "San Juan"

Objetivo de la Encuesta: Evaluar cuantitativamente la satisfacción de los usuarios con respecto a la funcionalidad, usabilidad y utilidad del aplicativo web para la predicción de producción de banano mediante Machine Learning.

Preguntas con escala de Likert

Instrucciones: Para cada afirmación, indique su nivel de acuerdo seleccionando una opción en la escala de 1 a 5, donde:

1	Muy en desacuerdo
2	En desacuerdo
3	Neutral
4	De acuerdo
5	Muy de acuerdo

1. ¿El aplicativo web cumple con su propósito de estimar la producción de banano futura?

1. Muy en desacuerdo	2. En desacuerdo	3. Neutral	4. De acuerdo	5. Muy de acuerdo

2. ¿Considera que las predicciones generadas por el sistema (basadas en Machine Learning) son útiles para planificar la cosecha?

1. Muy en desacuerdo	2. En desacuerdo	3. Neutral	4. De acuerdo	5. Muy de acuerdo

3. ¿El proceso de registro de datos (producción, clima y suelo) en el sistema es ágil y sencillo de realizar?

1. Muy en desacuerdo	2. En desacuerdo	3. Neutral	4. De acuerdo	5. Muy de acuerdo

4. ¿La interfaz del aplicativo web es intuitiva y fácil de navegar desde distintos dispositivos (PC, Celular)?

1. Muy en desacuerdo	2. En desacuerdo	3. Neutral	4. De acuerdo	5. Muy de acuerdo

5. ¿La visualización de los datos (gráficos de tendencias, comparativas real vs. predicho) es clara y fácil de interpretar?

1. Muy en desacuerdo	2. En desacuerdo	3. Neutral	4. De acuerdo	5. Muy de acuerdo

6. ¿Los reportes generados proporcionan información relevante para la toma de decisiones administrativas?

1. Muy en desacuerdo	2. En desacuerdo	3. Neutral	4. De acuerdo	5. Muy de acuerdo

7. ¿Considera que el uso de esta herramienta tecnológica mejora la eficiencia en la gestión agrícola de la Hacienda San Juan?

1. Muy en desacuerdo	2. En desacuerdo	3. Neutral	4. De acuerdo	5. Muy de acuerdo

Anexo N° 7: Análisis de la entrevista al administrador

UNIVERSIDAD AGRARIA DEL ECUADOR FACULTAD DE CIENCIAS AGRARIAS CARRERA DE COMPUTACIÓN Entrevista

Fecha: 16 de julio del 2025

Entrevistadores: Anchundia, 2025

Entrevistado: Ing. William Franco

Objetivo: Obtener información cualitativa y detallada sobre los procesos productivos, las prácticas de gestión y los desafíos operativos en el cultivo de banano en la Hacienda San Juan, para la validación de la problemática de investigación y la recopilación de datos esenciales para el diseño y desarrollo del sistema de predicción de la producción basado en machine learning.

Pregunta	Respuesta	Análisis
1. ¿Cuál es su rol dentro de hacienda San Juan?	Soy el encargado de registrar los datos de la pre producción de toda la hacienda para que el administrador pueda tomar decisiones en base a los datos.	El entrevistado es el responsable de registrar los datos de producción en la hacienda. El recopila información para la toma de decisiones del administrador.
2. ¿Cuál es el trabajo que realiza y que consiste?	Mi trabajo consiste en realizar un monitoreo mensual de la preproducción y para ello tengo que tomar datos de la plantación. Se han definido estaciones de muestreo, que son un número fijo de plantas en las cuales se toman los datos planta por planta. Para esto, se ha establecido un factor de agrupación de los racimos denominado Etapas de Desarrollo, las cuales se definen en base al tamaño del fruto y la coloración de los bananos.	El entrevistado explica que el trabajo suyo es monitorear mensualmente la preproducción y toma datos de la plantación. Utiliza un método de agrupación de racimos por etapas de desarrollo.

<p>3. ¿Qué procesos productivos se manejan en la Hacienda San Juan para la producción de banano?</p>	<p>Bueno, en la Hacienda San Juan nos encargamos de todo el proceso de producción de banano. Empezamos con el sembrado de dos variedades: William y Valery, que son de origen meristemo.</p>	<p>El entrevistado indica que se producen 2 variedades de bananos como el William y Valery.</p>
<p>4. ¿Cuáles son los principales factores que influyen en la variabilidad de la producción de banano?</p>	<p>La producción de banano es muy delicada y depende de varios factores. Los principales son los factores ambientales y del suelo, como el clima y la composición del suelo.</p>	<p>El entrevistado menciona que la producción de banano es delicada y que los factores climáticos si afectan a la producción de banano.</p>
<p>5. ¿Cómo estiman la producción de banano antes de la cosecha? ¿Siguen algún método específico?</p>	<p>Actualmente, para estimar el rendimiento de la cosecha, nos basamos en cálculos que parten del conteo de racimos de banano por hectárea. La estimación se hace considerando que un racimo de banano equivale a un peso promedio de 12 libras.</p> <p>Para realizar el cálculo, primero se obtiene un valor estimado de la cantidad de plantas por hectárea. Con esta información, se puede proyectar la producción tentativa de un lote específico. Posteriormente, se realiza la sumatoria</p>	<p>El entrevistado estimo el rendimiento de la cosecha basándose en cálculos de conteo de racimos de banano, usando como variable el peso, para luego hacer una estimación tentativa.</p>

total para obtener una estimación general. Con estos valores se pueden anticipar los resultados de la cosecha o tomar medidas de manejo si es necesario.

<p>6. ¿Realizan una comparación de los valores obtenidos por los cálculos con la producción real?</p>	<p>Por supuesto, una vez que la cosecha de banano es procesada y se obtiene el peso final de los racimos, este valor se compara con las estimaciones iniciales para determinar el margen de error. Este margen no debería ser mayor al 5%, lo que nos permite asegurar la precisión de nuestros cálculos y la efectividad del proceso de estimación.</p>	<p>El entrevistado asegura realizar comparaciones de con las estimaciones iniciales, teniendo un margen de error del 5%.</p>
<p>7. ¿Existe un registro histórico de producción en la hacienda? ¿Cómo se almacena y gestiona esta información?</p>	<p>Si existe un registro de la producción el cual se hace todos los días, se tiene toda esta información almacenada en tablas de Excel.</p>	<p>El entrevistado índice que si existe un registro histórico de las variables de producción.</p>
<p>8. ¿Actualmente se consideran parámetros climáticos para realizar las estimaciones?</p>	<p>Las estimaciones iniciales se basan únicamente en el conteo de racimos. Si se presenta un porcentaje de error considerablemente alto en estas proyecciones, se procede a revisar si las condiciones</p>	<p>Se realizan estimaciones basadas en conteo de racimos, en caso de haber un porcentaje de error alto se revisa las condiciones climáticas de esos días.</p>

climáticas pudieron ser las causantes de la discrepancia.

<p>9. ¿Qué desafíos enfrentan en la planificación de la cosecha debido a cambios en las condiciones ambientales?</p>	<p>Actualmente el clima puede variar mucho lo cual nos complica a veces a mantener una línea de producción adecuada, también otros factores como las plagas nos complican la producción.</p>	<p>El entrevistado indica que el clima en naranjal varía mucho, y que pueden aparecer plagas lo cual provoca complicaciones en la producción.</p>
<p>10. ¿Cómo afectan las plagas y enfermedades a la producción y qué estrategias emplean para su control?</p>	<p>Afectan de manera directa esta temporada nos estamos viendo atacados por una plaga de moco, para manejar o tratar de controlar se hace uso de pesticidas.</p>	<p>El entrevistado afirma que las plagas afectan directamente a la producción de banano de la hacienda y que actualmente tienen problema con una.</p>
<p>11. ¿Qué tipo de sensores o estaciones meteorológicas utilizan actualmente para monitorear el clima?</p>	<p>Actualmente no se está realizando el uso de estaciones meteorológicas dentro de la hacienda.</p>	<p>El entrevistado menciona q no se usan estaciones meteorológicas actualmente en la hacienda.</p>

Anexo N° 8: Análisis de la encuesta de satisfacción

Tabla 6.

Respuestas de las opciones de la pregunta 1: ¿El aplicativo web cumple con su propósito de estimar la producción de banano futura?

Opciones	Frecuencia	Porcentaje
1. Muy en desacuerdo	0	0%
2. En desacuerdo	0	0%
3. Neutral	0	0%
4. De acuerdo	0	0%
5. Muy de acuerdo	2	100%
Total	2	100%

Frecuencia y porcentaje de la pregunta 1

Elaborado por: Anchundia, 2025

Análisis: La tabla indica que el 100% de los encuestados (2 de 2) están "Muy de acuerdo" en que el aplicativo cumple su función principal de estimación. Esto demuestra que el sistema logra proyectar adecuadamente las cajas procesadas, alineándose con el objetivo general de reducir la incertidumbre en la producción.

Tabla 7.

Respuestas de las opciones de la pregunta 2: ¿Considera que las predicciones generadas por el sistema son útiles para planificar la cosecha?

Opciones	Frecuencia	Porcentaje
1. Muy en desacuerdo	0	0%
2. En desacuerdo	0	0%
3. Neutral	0	0%
4. De acuerdo	1	50%
5. Muy de acuerdo	1	50%

Total	2	100%
-------	---	------

Frecuencia y porcentaje de la pregunta 2

Elaborado por: Anchundia, 2025

Análisis: El 100% de los usuarios valida la utilidad de las predicciones, dividiéndose entre "De acuerdo" (50%) y "Muy de acuerdo" (50%). Este resultado positivo confirma que los modelos de Machine Learning implementados (como Random Forest o Xgboost) ofrecen datos de valor que facilitan la logística y planificación semanal de la cosecha en la hacienda.

Tabla 8.

Respuestas de las opciones de la pregunta 3: ¿El proceso de registro de datos (producción, clima y suelo) en el sistema es ágil y sencillo de realizar?

Opciones	Frecuencia	Porcentaje
1. Muy en desacuerdo	0	0%
2. En desacuerdo	0	0%
3. Neutral	0	0%
4. De acuerdo	0	0%
5. Muy de acuerdo	2	100%
Total	2	100%

Frecuencia y porcentaje de la pregunta 3

Elaborado por: Anchundia, 2025

Análisis: El 100% está "Muy de acuerdo" con la agilidad del registro de datos. Esto confirma que el módulo de Captura de Datos fue diseñado eficientemente, permitiendo a los administradores ingresar información histórica y actual sin complicaciones, optimizando el tiempo que antes dedicaban a hojas de cálculo manuales.

Tabla 9.

Respuestas de las opciones de la pregunta 4: ¿La interfaz del aplicativo web es intuitiva y fácil de navegar desde distintos dispositivos?

Opciones	Frecuencia	Porcentaje
1. Muy en desacuerdo	0	0%
2. En desacuerdo	0	0%
3. Neutral	0	0%
4. De acuerdo	0	0%
5. Muy de acuerdo	2	100%
Total	2	100%

Frecuencia y porcentaje de la pregunta 4

Elaborado por: Anchundia, 2025

Análisis: La totalidad de los encuestados (100%) está "Muy de acuerdo" en que la interfaz es intuitiva. Esto valida el cumplimiento de los requisitos no funcionales de usabilidad y diseño adaptable, permitiendo que el personal acceda al sistema desde computadoras o tablets sin dificultades de navegación.

Tabla 10.

Respuestas de las opciones de la pregunta 5: ¿La visualización de los datos (gráficos de tendencias, comparativas) es clara y fácil de interpretar?

Opciones	Frecuencia	Porcentaje
1. Muy en desacuerdo	0	0%
2. En desacuerdo	0	0%
3. Neutral	0	0%
4. De acuerdo	0	0%
5. Muy de acuerdo	2	100%
Total	2	100%

Frecuencia y porcentaje de la pregunta 5

Elaborado por: Anchundia, 2025

Análisis: El 100% de los encuestados considera que la visualización es excelente ("Muy de acuerdo"). Esto demuestra que el módulo de visualización presenta exitosamente las comparativas entre la producción real y la predicha, facilitando la interpretación rápida de patrones y tendencias sin necesidad de análisis estadísticos complejos por parte del usuario.

Tabla 11.

Respuestas de las opciones de la pregunta 6: ¿Los reportes generados proporcionan información relevante para la toma de decisiones?

Opciones	Frecuencia	Porcentaje
1. Muy en desacuerdo	0	0%
2. En desacuerdo	0	0%
3. Neutral	0	0%
4. De acuerdo	1	50%
5. Muy de acuerdo	1	50%
Total	2	100%

Frecuencia y porcentaje de la pregunta 6

Elaborado por: Anchundia, 2025

Análisis: Los usuarios expresan una aceptación total del 100%, con un 50% "De acuerdo" y un 50% "Muy de acuerdo". Esto valida que la generación automatizada de reportes en formatos estándar es una funcionalidad clave que apoya la labor administrativa y la trazabilidad de la información en la Hacienda San Juan.

Tabla 12.

Respuestas de las opciones de la pregunta 7: ¿Considera que el uso de esta herramienta tecnológica mejora la eficiencia en la gestión agrícola?

Opciones	Frecuencia	Porcentaje
1. Muy en desacuerdo	0	0%
2. En desacuerdo	0	0%
3. Neutral	0	0%
4. De acuerdo	0	0%
5. Muy de acuerdo	2	100%
Total	2	100%

Frecuencia y porcentaje de la pregunta 7

Elaborado por: Anchundia, 2025

Análisis: El 100% está "Muy de acuerdo" en que el aplicativo mejoró la eficiencia de la gestión. Este resultado es concluyente, ya que valida la hipótesis de la investigación: la integración de tecnología y Machine Learning transforma positivamente los procesos tradicionales de la hacienda, permitiendo una toma de decisiones más estratégica y fundamentada.

Anexo N° 9: Glosario LEL

Tabla 13.

Glosario LEL

Termino	Tipo	Noción	Impacto
Administrador	Sujeto	<ul style="list-style-type: none"> - Dueño o encargado de la hacienda “San Juan” 	<ul style="list-style-type: none"> - Se encarga de la administración de la hacienda - Lotes - Producción - Recibe reportes de producción - Necesita las predicciones de la producción - Toma de decisiones estratégicas
Planta de Banano	Objeto	<ul style="list-style-type: none"> - Planta de cultivo permanente - Planta perenne con alturas de 2 a 3 metros - Variedades: William y Valery - Perteneciente a los lotes 	<ul style="list-style-type: none"> - Es responsable de dar el fruto de banano - La floración ocurre de 9 a 12 meses después de la plantación. - Entre 12 y 14 meses después de la floración desarrolla el fruto de banano - La etapa de cosecha se realiza a los 14 a 16 meses
Lote	Objeto	<ul style="list-style-type: none"> - División específica de área de la hacienda - Se organiza las filas e hileras de plantas cultivadas 	<ul style="list-style-type: none"> - Permite el cultivo de plantas en cantidades específicas - Cada lote produce una cantidad de cosecha anual - Se registra en el aplicativo: <ul style="list-style-type: none"> - Numero de lote - Cajas procesadas - Se visualiza en los graficos las cajas procesadas por lotes

Termino	Tipo	Noción	Impacto
Variables meteorológicas	Objeto	<ul style="list-style-type: none"> - Conjunto de datos de las condiciones meteorológicas - Variables: - Temperatura máxima, media y mínima - Precipitación acumulada 	<ul style="list-style-type: none"> - Sirven como entrada para el modelo predictivo. - Ayudan a entender las condiciones del entorno agrícola. - Permiten ajustar el manejo del cultivo. - Mejoran la precisión de las predicciones.
Área de cultivo	Objeto	<ul style="list-style-type: none"> - Extensión territorial de la hacienda dividida por tipo de suelo y variedad de banano. 	<ul style="list-style-type: none"> - Permite organizar la producción por zonas. - Vincula datos edafoclimáticos con lotes. - Facilita la trazabilidad y monitoreo del cultivo. - Optimiza la planificación agrícola.
Datos edafoclimáticos	Objeto	<ul style="list-style-type: none"> - Información del suelo y el clima que influye en el desarrollo del banano. 	<ul style="list-style-type: none"> - Sirven como entrada para el modelo predictivo. - Ayudan a entender las condiciones del entorno agrícola. - Permiten ajustar el manejo del cultivo - Mejoran la precisión de las predicciones.
Producción	Estado	<ul style="list-style-type: none"> - Resultado cuantificable de la cosecha de banano obtenida en la hacienda. 	<ul style="list-style-type: none"> - Es la variable dependiente principal del sistema. - Indica el rendimiento real del cultivo. - Permite comparar predicciones con resultados reales. - Determina la rentabilidad de la hacienda.

Aplicativo Web	Objeto	<ul style="list-style-type: none"> - Sistema informático desarrollado para predecir y gestionar la producción de banano. 	<ul style="list-style-type: none"> - Centraliza el acceso a información agrícola. - Facilita el registro y análisis de datos. - Genera reportes automáticos.
Termino	Tipo	Noción	Impacto
Modelo predictivo	Objeto	<ul style="list-style-type: none"> - Conjunto de algoritmos de Machine Learning que estiman la producción de banano. 	<ul style="list-style-type: none"> - Analiza variables ambientales y productivas - Permite pronósticos de corto y mediano plazo. - Mejora la planificación agrícola. - Optimiza el uso de recursos.
Registrar datos	Verbo	<ul style="list-style-type: none"> - Acción de ingresar manual o automáticamente la información climática, del suelo o de producción al sistema. 	<ul style="list-style-type: none"> - Permite mantener la base de datos actualizada. - Facilita el entrenamiento de los modelos predictivos. - Mejora la precisión del análisis histórico. - Asegura la trazabilidad de la información.
Predecir producción	Verbo	<ul style="list-style-type: none"> - Acción de estimar la cantidad futura de banano con base en los datos históricos y ambientales. 	<ul style="list-style-type: none"> - Permite planificar cosechas con antelación. - Reduce pérdidas por variaciones climáticas. - Facilita decisiones estratégicas de venta y exportación. - Contribuye a la sostenibilidad de la producción.

Termino	Tipo	Noción	Impacto
Generar reporte	Verbo	<ul style="list-style-type: none"> - Acción de elaborar informes automáticos con los resultados del modelo predictivo. 	<ul style="list-style-type: none"> - Muestra gráficamente las tendencias de producción. - Permite comparar datos reales y estimados. - Facilita la toma de decisiones basadas en evidencia. - Mejora la transparencia y el control administrativo.
Base de datos	Objeto	<ul style="list-style-type: none"> - Repositorio estructurado donde se almacenan los registros de clima, suelo y producción. 	<ul style="list-style-type: none"> - Asegura la integridad de los datos. - Permite consultas y análisis rápidos. - Respalda información histórica del sistema. - Garantiza la disponibilidad de datos para el modelo.

Elaborado por: Anchundia, 2025

Anexo N° 10: Requerimientos funcionales y no funcionales, Historia de Usuario

Tabla 14.

Requerimientos funcionales

ID	Título de funcionalidad	Descripción de la funcionalidad	Prioridad	Acciones iniciadores y comportamiento	Resultado esperado	Estado
REQF-001	Inicio de sesión	Como administrador, quiero iniciar sesión con un usuario y contraseña, para poder acceder a las funciones del aplicativo de forma segura.	Alta	El administrador ingresa sus credenciales (usuario y contraseña) en la interfaz de inicio de sesión. El sistema verifica las credenciales.	Si la verificación es exitosa, el usuario es dirigido a la pantalla principal (dashboard) del aplicativo web.	Completado
REQF-002	Gestión de usuarios	Como administrador, necesito poder gestionar las cuentas de usuario (crear, editar, desactivar), para controlar quién tiene acceso al sistema.	Alta	El administrador accede a la sección 'Administración' -> 'Usuarios'. Selecciona 'Crear', 'Editar' o 'Desactivar' un usuario.	El sistema actualiza la base de datos con los cambios y muestra un mensaje de confirmación.	Completado
REQF-003	Registro de cajas procesadas	Como administrador, necesito una sección para registrar manualmente la producción semanal	Alta	El administrador accede al módulo 'Captura de Datos', selecciona un lote e ingresa el número de cajas procesadas. Presiona 'Guardar'.	El sistema almacena el nuevo registro de producción en la base de datos y actualiza las tablas históricas.	Completado

(cajas procesadas)
por cada lote, para
alimentar el historial.

ID	Título de funcionalidad	Descripción de la funcionalidad	Prioridad	Acciones iniciadores y comportamiento	Resultado esperado	Estado
REQF-004	Importar datos meteorológicos	Como administrador, quiero poder importar datos meteorológicos (ej. temperatura) desde un archivo (Excel/CSV), para usarlos en la predicción.	Alta	El administrador va a 'Captura de Datos' -> 'Importar Clima'. Selecciona un archivo (Excel/CSV) y presiona 'Cargar'.	El sistema lee el archivo, valida los datos, los importa a la base de datos y muestra un resumen de la importación.	Completado
REQF-005	Registro de análisis de suelo	Como administrador, necesito un formulario para registrar los datos de los análisis de suelo (ej. pH, N, P, K), para que el modelo los considere.	Media	El administrador accede a 'Captura de Datos' -> 'Datos de Suelo'. Completa el formulario con los valores del análisis y presiona 'Guardar'.	El sistema guarda los nuevos parámetros de suelo en la base de datos, asociándolos al lote correspondiente.	Completado

REQF-006	Preprocesamiento de datos	Como administrador, requiero que el sistema limpie y preprocese los datos cargados, para asegurar que el modelo funcione con información de calidad.	Alta	Esta acción es iniciada por el sistema (backend) automáticamente después de una importación o antes de ejecutar una predicción.	Los datos son limpiados (valores nulos, atípicos) y normalizados, quedando listos para el modelo.	Completado
-----------------	---------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------	------	---------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------	------------

ID	Título de funcionalidad	Descripción de la funcionalidad	Prioridad	Acciones iniciadores y comportamiento	Resultado esperado	Estado
REQF-007	Ejecutar modelo de predicción	Como administrador, quiero poder ejecutar el modelo de predicción (haciendo clic en un botón), para obtener una estimación de la producción futura.	Alta	El administrador, desde el 'Dashboard' principal, presiona el botón 'Generar Predicción'.	El sistema ejecuta el modelo de Machine Learning y genera el valor de la predicción (ej. número de cajas).	Completado
REQF-008	Visualizar predicción	Como administrador, necesito ver la predicción de producción de forma clara en un tablero (dashboard), para entender los resultados fácilmente.	Alta	El administrador carga el 'Dashboard' principal o completa la acción de 'Generar Predicción' (REQF-007).	La interfaz muestra un componente visual (ej. una tarjeta) con el resultado de la predicción de forma destacada.	Completado

REQF-009	Ver comparativa (Predicción vs. Real)	Como administrador, quiero ver gráficos que comparen la producción histórica (cajas reales) con las predicciones del modelo, para evaluar qué tan preciso es.	Media	El administrador navega a la sección 'Visualización' y observa el gráfico de 'Histórico de Producción'.	El sistema muestra un gráfico (ej. de líneas) con dos series: los datos reales de cajas y las predicciones pasadas.	Completado
-----------------	---------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------	-------	---------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------	------------

ID	Título de funcionalidad	Descripción de la funcionalidad	Prioridad	Acciones iniciadores y comportamiento	Resultado esperado	Estado
REQF-010	Exportar reporte de predicción	Como administrador, necesito generar un reporte simple con las predicciones, para poder exportarlo en PDF o Excel y compartirlo.	Media	El administrador, en la sección 'Reportes', presiona el botón 'Exportar a PDF' o 'Exportar a Excel'.	El sistema genera un archivo descargable (PDF o Excel) que contiene la tabla de predicciones.	Completado
REQF-011	Ver métricas del modelo	Como administrador, quiero ver las métricas de precisión (R ² , MAE, RMSE), para saber qué tan confiables son las predicciones.	Baja	El administrador navega a una sección avanzada de 'Configuración' o 'Evaluación del Modelo'.	El sistema muestra en pantalla los valores de las métricas de evaluación del modelo.	Completado

Elaborado por: Anchundia, 2025

Tabla 15.

Requerimientos no funcionales

ID	Título de funcionalidad	Descripción de la funcionalidad	Prioridad	Acciones iniciadores y comportamiento	Resultado esperado	Estado
REQ N-001	Usabilidad (Facilidad de uso)	El aplicativo web debe ser intuitivo y fácil de usar, permitiendo que los administradores puedan operar el sistema sin capacitación técnica avanzada.	Alta	Un administrador intenta registrar la producción (REQF-003) por primera vez.	El administrador completa la tarea en menos de 3 minutos sin necesidad de consultar un manual.	Completado
REQ N-002	Seguridad (Control de acceso)	El sistema debe garantizar el acceso solo a personal autorizado (administradores) mediante un módulo de inicio de sesión.	Alta	Un usuario no autenticado intenta acceder a la URL del dashboard principal.	El sistema redirige al usuario a la página de inicio de sesión (REQF-001).	Completado
REQ N-003	Seguridad (Contraseñas)	Las contraseñas deben almacenarse de forma segura (cifradas) y requerir una longitud y complejidad mínimas (ej. 8 caracteres, mayúsculas, etc.).	Alta	Un administrador intenta crear un nuevo usuario o cambiar su contraseña por una simple como "12345".	El sistema rechaza la contraseña y muestra un mensaje indicando los requisitos de complejidad.	Completado

ID	Título de funcionalidad	Descripción de la funcionalidad	Prioridad	Acciones iniciadores y comportamiento	Resultado esperado	Estado
REQ N-004	Disponibilidad	El aplicativo debe tener una alta disponibilidad durante las horas laborables de la hacienda (ej. Lunes a sábado, 8 AM - 5 PM).	Alta	Un administrador intenta acceder al aplicativo un martes a las 10 AM.	El sistema está en línea y responde en menos de 5 segundos	Completado
REQ N-005	Usabilidad (Diseño adaptable)	El aplicativo debe ser un sistema web adaptable, permitiendo su correcta visualización en computadoras y tablets.	Media	El administrador abre el aplicativo web en una tablet	La interfaz se reajusta correctamente, sin barras de scroll horizontal y con elementos táctiles de tamaño adecuado.	Completado
REQ N-006	Robustez (Precisión del modelo)	El modelo de predicción debe utilizar los datos históricos disponibles para asegurar la fiabilidad de las estimaciones.	Alta	El sistema ejecuta la predicción (REQF-007) usando el conjunto de datos históricos.	El modelo de predicción alcanza un nivel de precisión predefinido (ej. $R^2 > 0.75$) en el conjunto de validación.	Completado

ID	Título de funcionalidad	Descripción de la funcionalidad	Prioridad	Acciones iniciadores y comportamiento	Resultado esperado	Estado
REQ N-007	Escalabilidad (Re-entrenamiento)	El sistema debe permitir que el modelo de predicción se actualice (re-entrene) periódicamente con nuevos datos de producción.	Media	Un administrador agrega datos de 6 nuevos meses y ejecuta una función de "Re-entrenar modelo".	El sistema re-entrena el modelo de ML con los nuevos datos sin fallar y actualiza el modelo productivo.	Completado
REQ N-008	Interoperabilidad (Exportación)	El sistema debe permitir la exportación de reportes (predicciones, históricos) a formatos de archivo comunes como PDF o Excel.	Media	El administrador presiona el botón "Exportar a Excel" (REQF-010).	El sistema genera y descarga un archivo .xlsx válido y legible con los datos del reporte.	Completado
REQ N-009	Rendimiento (Velocidad de carga)	El tiempo de respuesta para cargar el dashboard de predicción (REQF-008) no debe superar los 5 segundos en una conexión estándar.	Media	El administrador inicia sesión y carga el dashboard principal.	La página, incluyendo gráficos y predicción, se carga completamente en menos de 5 segundos.	Completado
REQ N-010	Mantenimiento (Calidad de código)	El código fuente (PHP, HTML) debe ser limpio,	Baja	Un nuevo desarrollador revisa el código fuente del módulo de "Gestión de	El desarrollador puede entender la lógica y realizar una modificación simple	Completado

documentado y fácil
de mantener para
futuras
modificaciones.

usuarios" (REQF-002).

gracias a los comentarios.

Elaborado por: Anchundia, 2025

Tabla 16.
Historia de Usuario

N°	Usuarios: Usuario (Administrador de Hacienda)
1	Quiero tener un inicio de sesión seguro con mis credenciales (usuario y contraseña) para acceder al aplicativo web y proteger la información de la hacienda.
2	Quiero registrar manualmente la producción semanal (número de enfundes, racimos y cajas procesadas) por cada lote para alimentar la base de datos histórica.
3	Quiero poder importar datos meteorológicos (temperatura, lluvias) desde archivos externos para que el modelo tenga información climática actualizada.
4	Quiero registrar los resultados de los análisis de suelo (pH, nutrientes) para considerar estas variables en la predicción.
5	Quiero ejecutar el modelo de Machine Learning con un clic para obtener una estimación de cuántas cajas de banano se producirán en las próximas semanas.
6	Quiero visualizar en un dashboard gráficos que comparen la producción real vs. la predicción del sistema para evaluar la precisión.
7	Quiero generar reportes analíticos de las predicciones y el historial de producción para analizar el rendimiento de la hacienda.
8	Quiero poder exportar los reportes en formatos PDF o Excel para compartir la información con la gerencia o usarla en reuniones.
9	Quiero acceder al sistema desde una computadora o tablet de forma fluida para poder revisar la información tanto en la oficina como en campo.

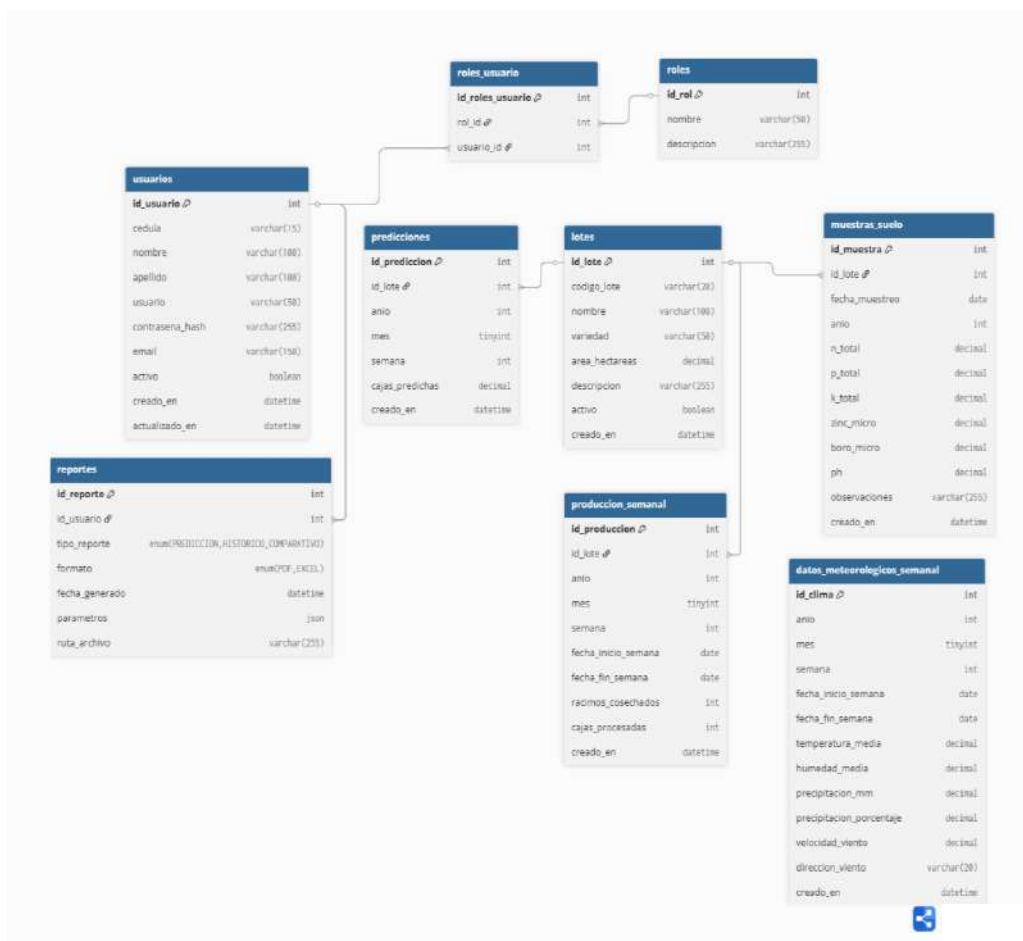
Descripción de Historia de Usuario

Elaborado por: Anchundia, 2025

Anexo N° 11: Diagrama de Entidad Relación

Figura 7.

Diagrama de Entidad Relación



Elaborado por: Anchundia, 2025

Anexo N° 12: Diccionario de datos

Tabla 17.

Diccionario de datos - Tabla roles

Campo	Llave	Tipo de dato	Tamaño	Nulidad	Descripción
id_rol	PK	Entero	4 b	No	Identificador único del rol.
nombre		Texto	50 b	No	Nombre del rol (ADMIN, TÉCNICO, etc.).
descripcion		Texto	255 b	Sí	Descripción breve del rol.

Elaborado por: Anchundia, 2025

Tabla 18.

Diccionario de datos - Tabla usuarios

Campo	Llave	Tipo de dato	Tamaño	Nulidad	Descripción
id_usuario	PK	Entero	4 b	No	Código identificador único del usuario.
cedula		Texto	15 b	No	Número de cédula del usuario.
nombre		Texto	100 b	No	Nombres del usuario.
apellido		Texto	100 b	No	Apellidos del usuario.
usuario		Texto	50 b	No	Nombre de usuario para iniciar sesión.
contrasena_hash		Texto	255 b	No	Contraseña encriptada del usuario.
email		Texto	150 b	No	Correo electrónico del usuario.
activo		Booleano	1 b	No	Indica si el usuario está activo en el sistema.
creado_en		Fecha-hora	8 b	No	Fecha y hora de creación del registro.
actualizado_en		Fecha-hora	8 b	Sí	Fecha y hora de la última actualización del registro.

Elaborado por: Anchundia, 2025

Tabla 19.***Diccionario de datos - Tabla lotes***

Campo	Llave	Tipo de dato	Tamaño	Nulidad	Descripción
id_lote	PK	Entero	4 b	No	Identificador único del lote de producción.
codigo_lote	–	Texto	20 b	No	Código interno del lote.
nombre	–	Texto	100 b	No	Nombre descriptivo del lote.
variedad	–	Texto	50 b	Sí	Variedad de banano sembrada en el lote.
area_hectareas	–	Decimal	8 b	Sí	Área del lote en hectáreas.
descripcion	–	Texto	255 b	Sí	Observaciones generales del lote.
activo	–	Booleano	1 b	No	Indica si el lote está activo.
creado_en	–	Fecha-hora	8 b	No	Fecha y hora de creación del registro.

Elaborado por: Anchundia, 2025

Tabla 20.***Diccionario de datos - Tabla producción_semanal***

Campo	Llave	Tipo de dato	Tamaño	Nulidad	Descripción
id_produccion	PK	Entero	4 b	No	Identificador único del registro de producción.
id_lote	FK	Entero	4 b	No	Lote al que corresponde la producción (relación con lotes).
anio		Entero	4 b	No	Año del registro de producción.
mes		Entero	1 b	No	Mes del registro (1–12).
semana		Entero	4 b	No	Número de semana del año.

fecha_inicio_semana	Fecha	3 b	Sí	Fecha de inicio de la semana de producción.
fecha_fin_semana	Fecha	3 b	Sí	Fecha de fin de la semana de producción.
racimos_cosechados	Entero	4 b	Sí	Cantidad de racimos cosechados en la semana.
cajas_procesadas	Entero	4 b	Sí	Número de cajas procesadas en la semana.
creado_en	Fecha-hora	8 b	No	Fecha y hora de creación del registro.

Elaborado por: Anchundia, 2025

Tabla 21.

Diccionario de datos – Tabla muestra suelo

Campo	Llave	Tipo de dato	Tamaño	Nulidad	Descripción
id_muestra	PK	Entero	4 b	No	Identificador único de la muestra de suelo.
id_lote	FK	Entero	4 b	No	Lote del cual se toma la muestra (relación con lotes).
fecha_muestreo		Fecha	3 b	No	Fecha en que se realizó el muestreo.
anio		Entero	4 b	Sí	Año de referencia del análisis de suelo.
n_total		Decimal	8 b	Sí	Contenido de nitrógeno total en el suelo.
p_total		Decimal	8 b	Sí	Contenido de fósforo total en el suelo.
k_total		Decimal	8 b	Sí	Contenido de potasio total en el suelo.
zinc_micro		Decimal	8 b	Sí	Contenido de zinc en el suelo.
boro_micro		Decimal	8 b	Sí	Contenido de boro en el suelo.
ph		Decimal	8 b	Sí	Nivel de pH del suelo.

observaciones	Texto	255 b	Sí	Comentarios u observaciones adicionales del análisis.
creado_en	Fecha-hora	8 b	No	Fecha y hora de creación del registro.

Elaborado por: Anchundia, 2025

Tabla 22.

Diccionario de datos – Tabla datos_meteorologicos

Campo	Llave	Tipo de dato	Tamaño	Nulidad	Descripción
id_clima	PK	Entero	4 b	No	Identificador único del registro climático.
anio		Entero	4 b	No	Año del registro climático.
mes		Entero	1 b	No	Mes del registro (1–12).
semana		Entero	4 b	No	Número de semana del año.
fecha_inicio_semana		Fecha	3 b	Sí	Fecha de inicio de la semana considerada.
fecha_fin_semana		Fecha	3 b	Sí	Fecha de fin de la semana considerada.
temperatura_media		Decimal	8 b	Sí	Temperatura media semanal (°C).
humedad_media		Decimal	8 b	Sí	Humedad relativa media semanal (%).
precipitacion_mm		Decimal	8 b	Sí	Precipitación acumulada semanal (mm).
precipitacion_porcentaje		Decimal	8 b	Sí	Porcentaje de precipitación respecto a un valor de referencia.
velocidad_viento		Decimal	8 b	Sí	Velocidad media del viento.
direccion_viento		Texto	20 b	Sí	Dirección predominante

creado_en	Fecha-hora	8 b	No	del viento (N, NE, etc.). Fecha y hora de creación del registro.
-----------	------------	-----	----	---------------------------------------------------------------------

Elaborado por: Anchundia, 2025

Tabla 23.

Diccionario de datos – Tabla reportes

Campo	Llave	Tipo de dato	Tamaño	Nulidad	Descripción
id_reporte	PK	Entero	4 b	No	Identificador único del reporte.
id_usuario	FK	Entero	4 b	No	Usuario que generó el reporte
tipo_reporte		Enumerado	1 b	No	Tipo de reporte generado.
formato		Enumerado (PDF, EXCEL)	1 b	No	Formato de salida del reporte.
fecha_generado		Fecha-hora	8 b	No	Fecha y hora en que se generó el reporte.
parametros		JSON	Variable	Sí	Parámetros y filtros utilizados para generar el reporte.

Elaborado por: Anchundia, 2025

Tabla 24.

Diccionario de datos – Tabla predicciones

Campo	Llave	Tipo de dato	Tamaño	Nulidad	Descripción
id_prediccion	PK	Entero	4 b	No	Identificador único de la predicción.
id_lote	FK	Entero	4 b	No	Lote para el cual se genera la predicción
anio	–	Entero	4 b	No	Año al que corresponde la predicción.
mes	–	Entero	1 b	No	Mes al que corresponde la predicción.

semana	–	Entero	4 b	No	Semana del año predicha.
cajas_predichas	–	Decimal	8 b	No	Número de cajas de banano proyectadas para esa semana.
creado_en	–	Fecha-hora	8 b	No	Fecha y hora en que se registra la predicción.

Elaborado por: Anchundia, 2025

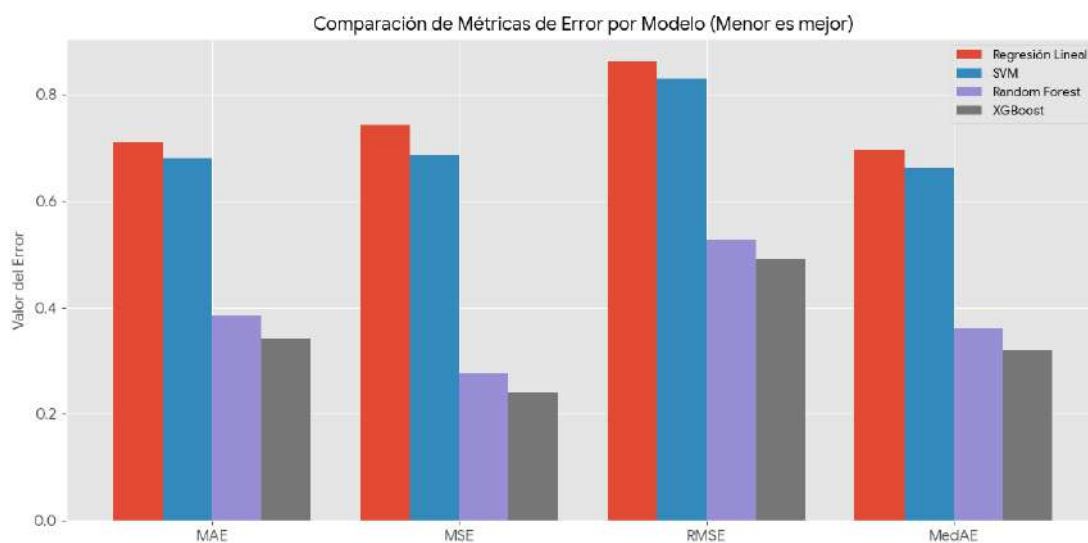
Anexo N° 13 Evaluación de modelos de machine learning

Tabla 25.
Evaluación de modelos

Modelo	MAE	MSE	RMSE	R ²	MedAE
Regresión Lineal	0.709	0.743	0.862	0.638	0.695
SVM	0.680	0.687	0.829	0.666	0.662
Random Forest	0.386	0.278	0.527	0.812	0.361
XGBoost	0.342	0.241	0.491	0.892	0.319

Elaborado por: Anchundia, 2025

Figura 8.
Comparación de métricas por modelo



Elaborado por: Anchundia, 2025

Tabla 26.
Evaluación de las variables usando modelo XGBoost

Variable	Modelo	RMSE	MAE	MAPE (%)
Racimos cosechados	XGBoost	0.54	0.39	4.12
Cajas procesadas	XGBoost	0.49	0.34	3.67

Elaborado por: Anchundia, 2025

Anexo N° 14: Diagrama de flujo de datos

Figura 9.

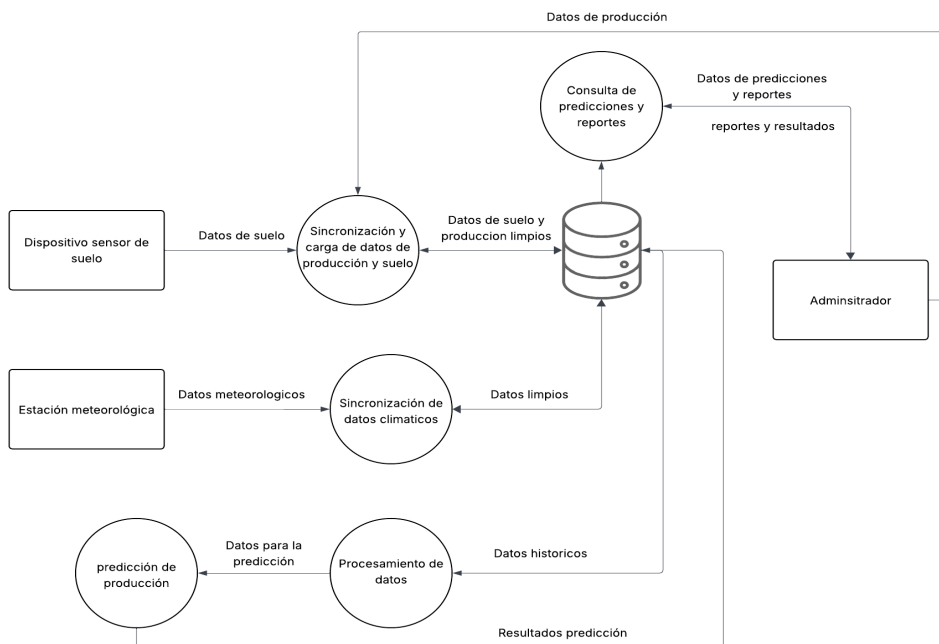
Diagrama de Contexto



Elaborado por: Anchundia, 2025

Figura 10.

Diagrama de flujo de datos nivel 1

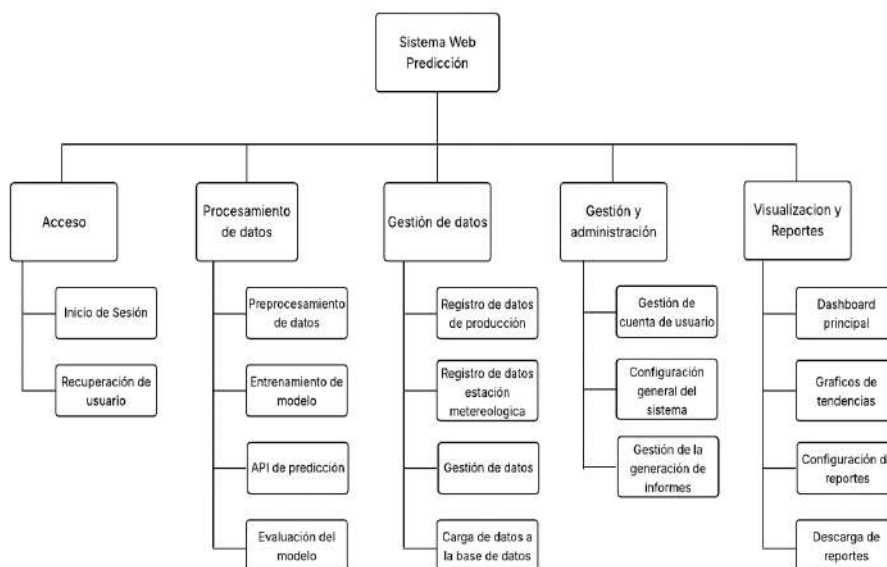


Elaborado por: Anchundia, 2025

Anexo N° 15: Esquema modular del sistema

Figura 11.

Esquema modular del sistema

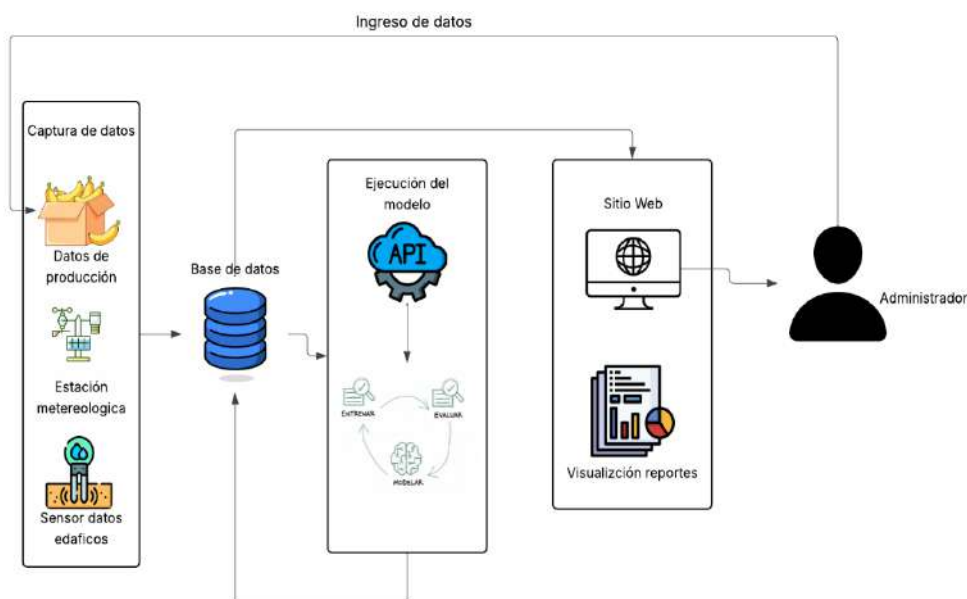


Elaborado por: Anchundia, 2025

Anexo N° 16: Esquema modular de la arquitectura de la aplicación

Figura 12.

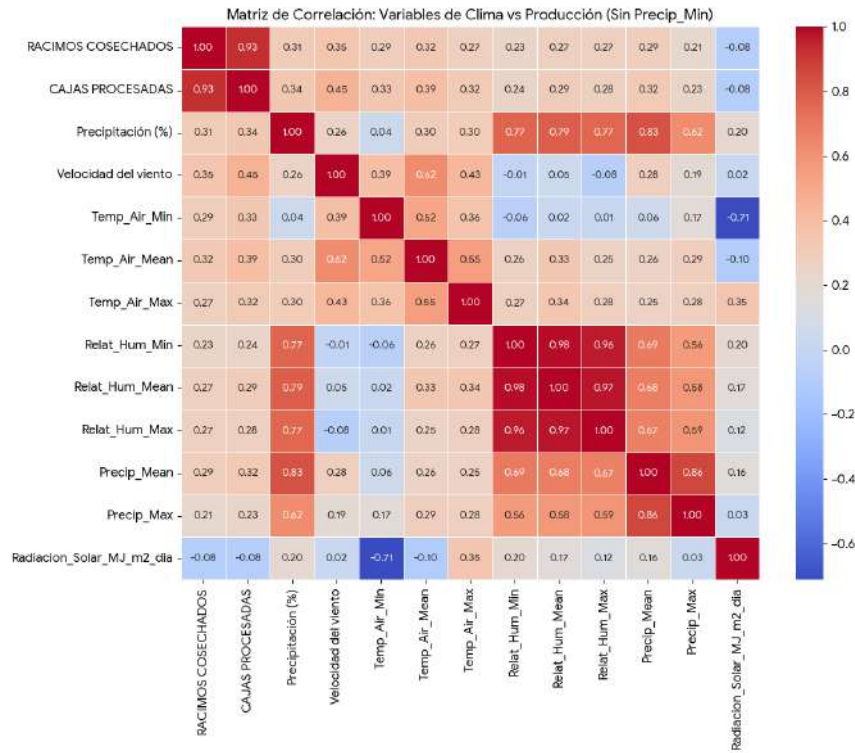
Esquema modular de la arquitectura de la aplicación



Elaborado por: Anchundia, 2025

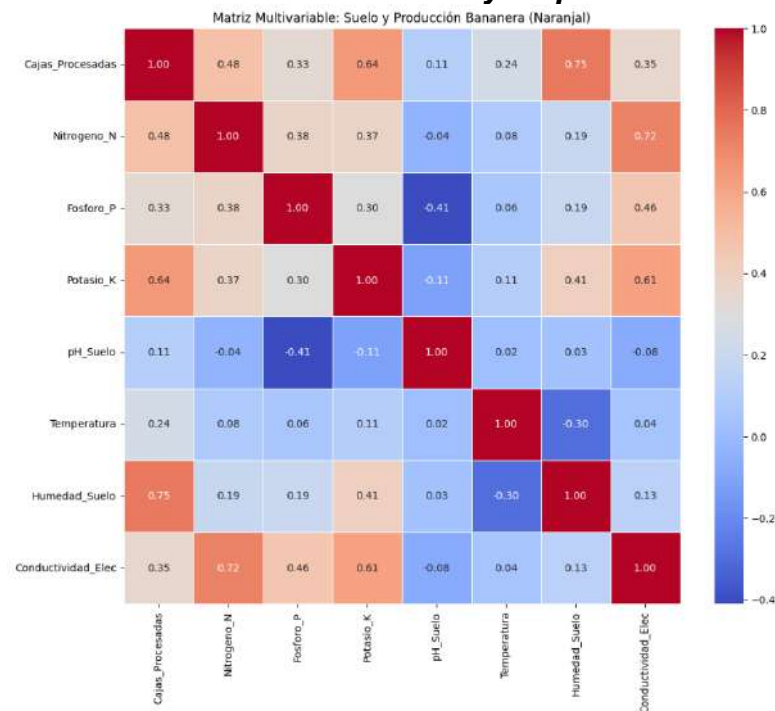
Anexo N° 17 Correlación entre variables con producción

Figura 13.
Matriz de correlación entre variables climáticas y de producción



Elaborado por: Anchundia, 2025

Figura 14.
Matriz de correlación entre variables de suelo y de producción

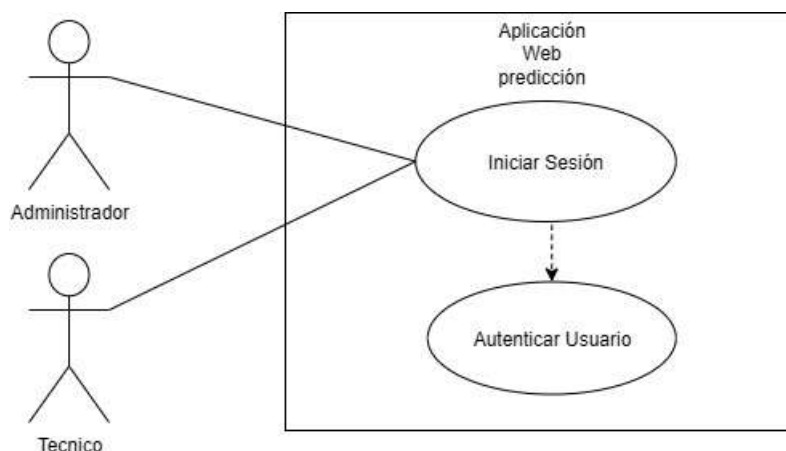


Elaborado por: Anchundia, 2025

Anexo N° 18: Diagramas de caso de uso

Figura 15.

Diagrama de Caso de uso: Iniciar Sesión- Seguridad



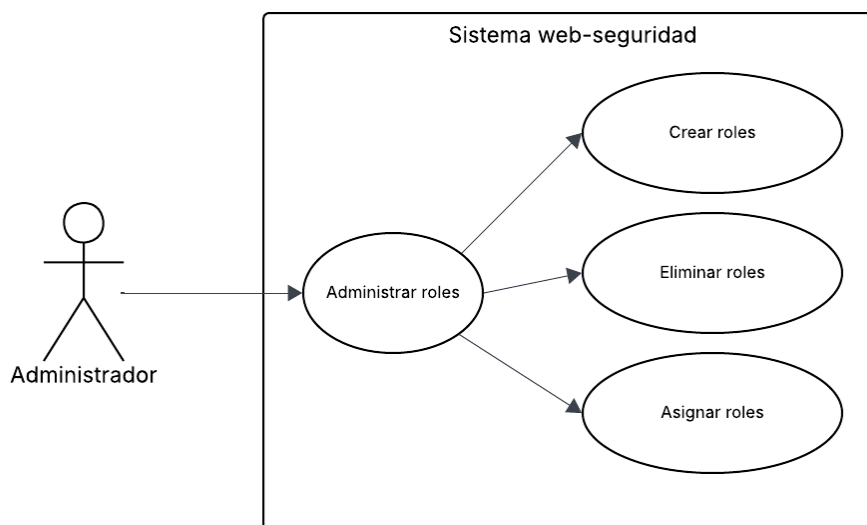
Elaborado por: Anchundia, 2025

Tabla 27.

Caso de uso: Iniciar Sesión- Seguridad

Caso de Uso	CU-01 Iniciar sesión
Descripción	Permite a los usuarios del sistema (Administrador y Técnico) autenticarse ingresando sus credenciales para acceder a la aplicación web de predicción.
Actores	Administrador, Técnico
Propósito	Garantizar acceso seguro y controlado al sistema según rol.
Flujo Principal	<ol style="list-style-type: none"> 1. El usuario abre la pantalla de inicio de sesión. 2. Ingresa usuario y contraseña. 3. El sistema ejecuta 'Autenticar Usuario' (validación de credenciales y estado). 4. Si es válido, crea sesión y redirige al dashboard correspondiente a su rol.
Excepciones	<ol style="list-style-type: none"> a) Credenciales inválidas: mostrar mensaje y permitir reintentar. b) Usuario inactivo/bloqueado: denegar acceso e informar. c) Exceso de intentos fallidos: bloquear temporalmente la cuenta. d) Error de servidor/BD: registrar evento y mostrar mensaje genérico.
Pre-condición	Usuario registrado en el sistema y con estado ACTIVO.
Post-condición	Sesión iniciada, token/cookie de sesión creado y usuario redirigido al dashboard de su rol.

Figura 16.

Diagrama de Caso de uso: Administrar roles: Web- Seguridad

Elaborado por: Anchundia, 2025

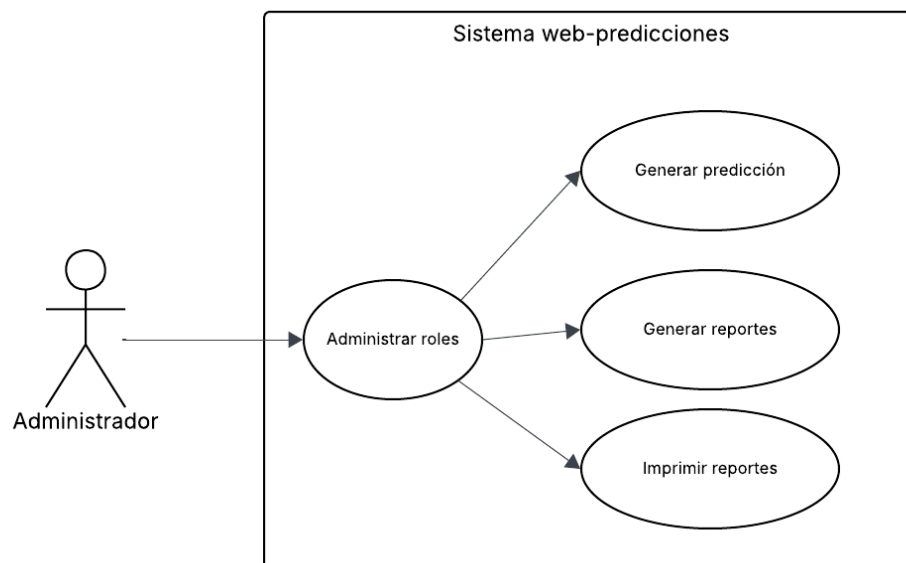
Tabla 28.

Caso de uso: Administrar roles: Web- Seguridad

Caso de Uso	CU-02 Seguridad - Administración de roles
Descripción	El administrador crea, elimina y asigna roles a los usuarios del sistema.
Actores	Administrador
Propósito	Controlar los accesos y permisos de los diferentes usuarios del sistema.
Flujo Principal	<ol style="list-style-type: none"> 1. El administrador accede al módulo de seguridad. 2. Crea nuevos roles o elimina existentes. 3. Asigna los roles a los usuarios correspondientes.
Excepciones	<ol style="list-style-type: none"> a) Rol duplicado. b) Error al eliminar un rol en uso.
Pre-condición	El administrador debe estar autenticado en el sistema.
Post-condición	Los roles se actualizan correctamente en el sistema.

Elaborado por: Anchundia, 2025

Figura 17.

Diagrama de Caso de uso: Sistema web-Predicciones

Elaborado por: Anchundia, 2025

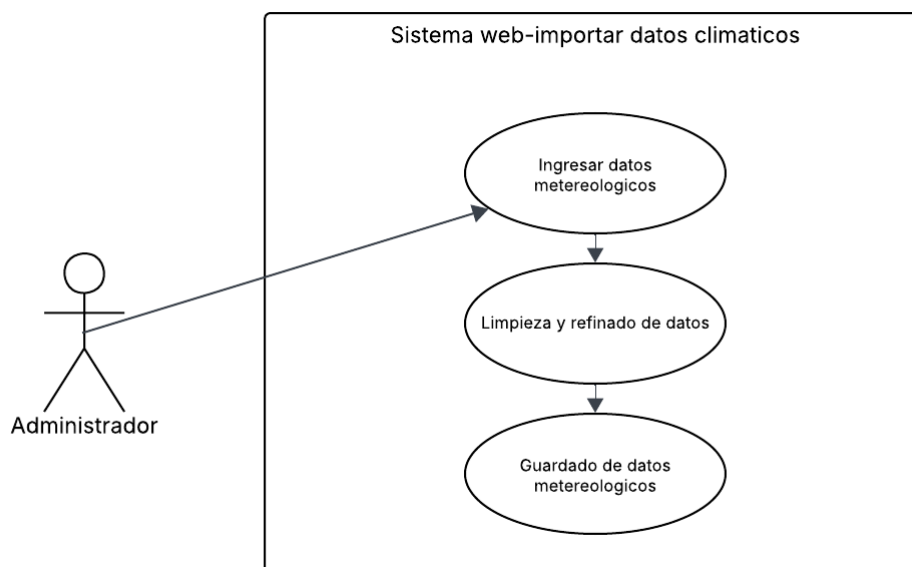
Tabla 29.

Caso de uso: Sistema web-Predicciones

Caso de Uso	CU-03 Administrar predicciones
Descripción	Permite generar predicciones, reportes e imprimir los resultados desde el módulo de predicciones.
Actores	Administrador
Propósito	Gestionar el proceso de predicción de producción de banano y generar reportes de resultados.
Flujo Principal	<ol style="list-style-type: none"> 1. El administrador accede al módulo de predicciones. 2. Genera una nueva predicción. 3. Visualiza y genera reportes. 4. Imprime los resultados obtenidos.
Excepciones	<ol style="list-style-type: none"> a) Datos insuficientes para generar predicciones. b) Fallo en la generación del reporte.
Pre-condición	El modelo predictivo debe estar entrenado y los datos cargados.
Post-condición	Las predicciones se generan y reportan correctamente.

Elaborado por: Anchundia, 2025

Figura 18.

Diagrama de Caso de Uso: Sistema web – Importar datos meteorológicos

Elaborado por: Anchundia, 2025

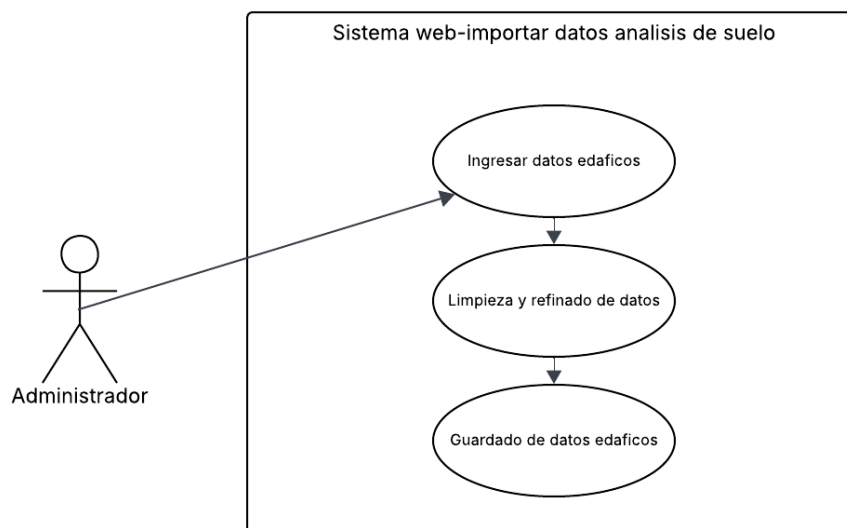
Tabla 30.

Caso de Uso: Sistema web – Importar datos meteorológicos

Caso de Uso	CU-04 Importar datos climáticos
Descripción	Permite ingresar, limpiar y guardar los datos meteorológicos en el sistema.
Actores	Administrador
Propósito	Registrar las condiciones climáticas necesarias para el modelado de predicción.
Flujo Principal	<ol style="list-style-type: none"> 1. El administrador ingresa los datos meteorológicos. 2. El sistema realiza la limpieza y refinado de datos. 3. Se guardan los registros en la base de datos.
Excepciones	<ol style="list-style-type: none"> a) Error en el formato de archivo. b) Datos duplicados o incompletos.
Pre-condición	El sistema debe permitir la carga de archivos de datos válidos.
Post-condición	Los datos meteorológicos se guardan correctamente y quedan listos para el uso del modelo predictivo.

Elaborado por: Anchundia, 2025

Figura 19.

Diagrama de Caso de uso: Sistema web- importar datos análisis de suelo

Elaborado por: Anchundia, 2025

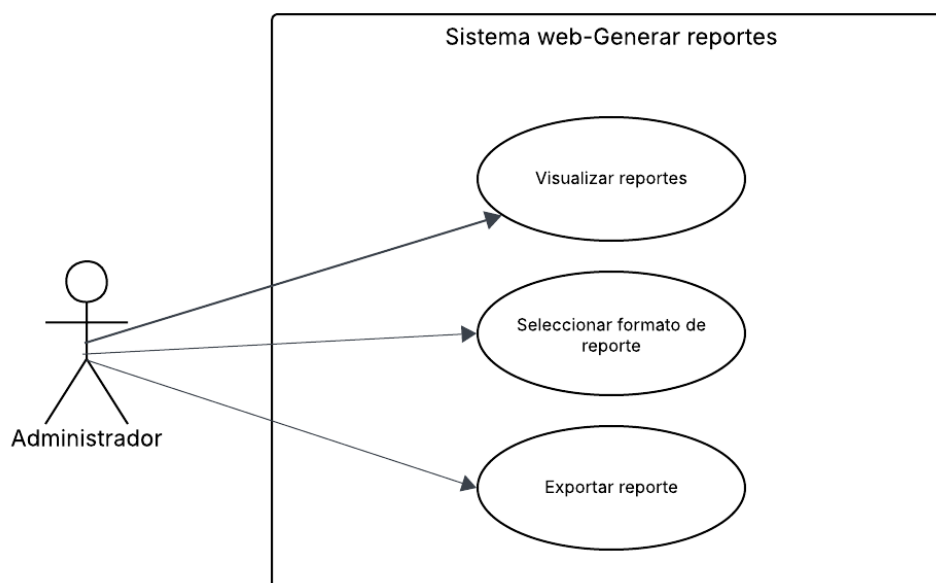
Tabla 31.

Caso de uso: Sistema web- importar datos análisis de suelo

Caso de Uso	CU-05 Importar datos de análisis de suelo
Descripción	El administrador ingresa los datos edáficos y el sistema los limpia y guarda.
Actores	Administrador
Propósito	Registrar y almacenar los datos de análisis del suelo en el sistema.
Flujo Principal	<ol style="list-style-type: none"> 1. El administrador ingresa los datos edáficos. 2. El sistema limpia y refina los datos. 3. El sistema guarda los registros en la base de datos.
Excepciones	<ol style="list-style-type: none"> a) Datos incompletos o inválidos. b) Error de conexión con la base de datos.
Pre-condición	El administrador debe tener acceso al módulo y los datos del análisis deben estar disponibles.
Post-condición	Los datos edáficos se guardan correctamente y quedan disponibles para análisis futuros.

Elaborado por: Anchundia, 2025

Figura 20.

Diagrama de Caso de uso: Sistema web- Generar reportes

Elaborado por: Anchundia, 2025

Tabla 32.

Caso de uso: Sistema web- Generar reportes

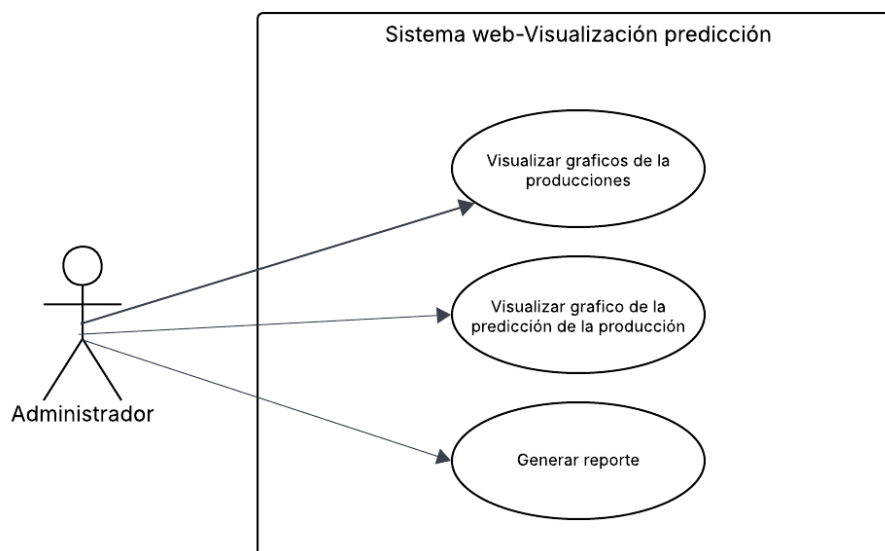
Caso de Uso	CU-07 Generar reportes
Descripción	Permite al administrador visualizar, seleccionar formato y exportar reportes del sistema.
Actores	Administrador
Propósito	Facilitar la generación de informes sobre la producción y predicciones.
Flujo Principal	<ol style="list-style-type: none"> 1. El administrador accede al módulo de reportes. 2. Visualiza reportes disponibles. 3. Selecciona el formato deseado (PDF o Excel). 4. Exporta el reporte generado.
Excepciones	<ol style="list-style-type: none"> a) Error al generar el archivo. b) Formato de salida no disponible.
Pre-condición	El sistema debe contar con datos y reportes generados previamente.

Post-condición	El reporte se exporta exitosamente en el formato seleccionado.
-----------------------	----------------------------------------------------------------

Elaborado por: Anchundia, 2025

Figura 21.

Diagrama de Caso de uso: Sistema web- Visualización de predicción



Elaborado por: Anchundia, 2025

Tabla 33.

Caso de uso: Sistema web- Visualización de predicción

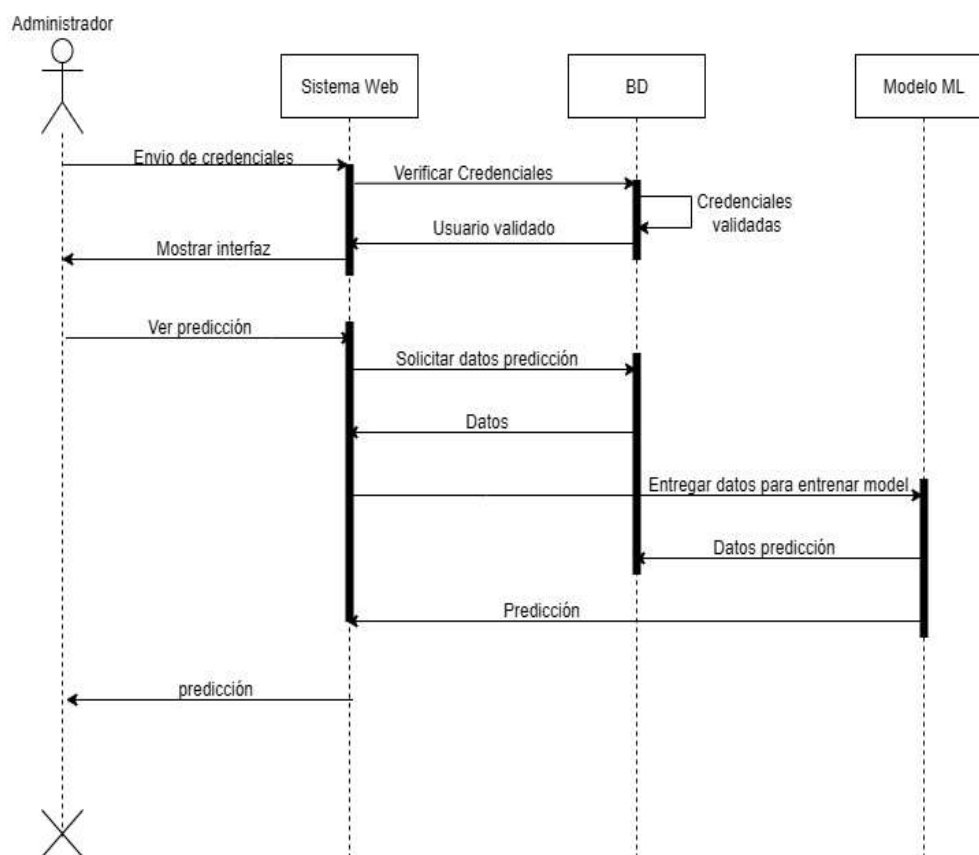
Caso de Uso	CU-08 Visualización de predicción
Descripción	El administrador puede visualizar los gráficos de producción y predicción, además de generar reportes visuales.
Actores	Administrador
Propósito	Proporcionar una representación visual del rendimiento y predicción del cultivo.
Flujo Principal	<ol style="list-style-type: none"> 1. El administrador accede al módulo de visualización. 2. Selecciona el tipo de gráfico (producción o predicción). 3. El sistema muestra el gráfico correspondiente. 4. El usuario puede generar un reporte visual.

Excepciones	a) Falta de datos para graficar. b) Error en la carga de gráficos.
Pre-condición	El sistema debe tener datos históricos de producción y predicción.
Post-condición	Los gráficos y reportes visuales se muestran correctamente al usuario.

Elaborado por: Anchundia, 2025

Anexo N° 19: Diagrama de secuencia del aplicativo

Figura 22.

Diagrama de secuencia del aplicativo

Elaborado por: Anchundia, 202

Anexo N° 20: Fotos entrevista**Figura 23.*****Visita a la hacienda San Juan***

Elaborado por: Anchundia, 2025

Figura 24.***Entrevista al administrador de la hacienda***

Elaborado por: Anchundia, 2025.

Anexo N° 21: Pruebas del software – Funcionalidad y usabilidad

Tabla 34.

Prueba de funcionalidad del módulo de Gestión y Seguridad

Prueba:	Funcionalidad del módulo de Gestión de Usuarios y Seguridad.	
Fecha:	01/12/2025	
Tema del software:	Aplicativo de predicción mediante Machine Learning de la producción de banano en la Hacienda San Juan.	
Objetivo de prueba:	Evaluar la funcionalidad del módulo de seguridad y gestión de usuarios para verificar el control de acceso y la administración de cuentas.	
Funciones del Sistema	Calificar Sí / No	Observaciones
Inicio de Sesión		
Ingresar al sistema	X	Validar que solo usuarios registrados (Administrador/Técnico) puedan acceder con credenciales correctas.
Validación de credenciales	X	Verificar que el sistema deniegue el acceso si la contraseña es incorrecta o el usuario no existe.
Gestión de Usuarios		
Crear nuevo usuario	X	Completar campos (Cédula, Nombre, Rol, Email) y verificar que se guarde en la base de datos sin duplicados.
Editar usuario	X	Confirmar que los cambios (ej. cambio de rol o contraseña) se actualizan correctamente.
Desactivar usuario	X	Verificar que al cambiar el estado a "Inactivo", el usuario ya no pueda acceder al sistema.
Observaciones Generales:	Cumple con los requisitos de seguridad y gestión de usuarios (REQF-001, REQF-002).	

Desarrollador del sistema: Anchundia Zambrano Andres David

Encargado de la prueba: Anchundia Zambrano Andres David

Elaborado por: Anchundia, 2025

Tabla 35.

Prueba de funcionalidad del módulo de Captura de Datos

Prueba:	Funcionalidad del módulo de Captura de Datos.	
Fecha:	01/12/2025	
Tema del software:	Aplicativo de predicción mediante Machine Learning de la producción de banano en la Hacienda San Juan.	
Objetivo de prueba:	Verificar la correcta recolección, importación y almacenamiento de los datos históricos necesarios para el modelo predictivo.	
Funciones del Sistema	Calificar Sí / No	Observaciones
Datos de Producción		
Registrar producción semanal	X	Validar el ingreso manual de datos (Enfundes, Racimos, Cajas) asociados a un Lote y Semana específicos.
Historial de producción	X	Verificar que los datos guardados se reflejen correctamente en las tablas históricas.
Datos Meteorológicos		
Importar datos climáticos	X	Probar la carga masiva de archivos (Excel/CSV) con variables de Temperatura y Precipitación.

Validación de formato	X	Confirmar que el sistema rechace archivos con formatos incorrectos o datos corruptos.
Datos de Suelo		
Registrar análisis de suelo	X	Completar formulario con niveles de nutrientes (N, P, K, pH) y validar su almacenamiento por lote.
Observaciones Generales:		El módulo permite la carga manual y masiva de datos correctamente, validando tipos de datos y fechas (REQF-003, REQF-004, REQF-005).
Desarrollador del sistema:		Anchundia Zambrano Andres David
Encargado de la prueba:		Anchundia Zambrano Andres David

Elaborado por: Anchundia, 2025

Tabla 36.

Prueba de funcionalidad del módulo de Modelado y Predicción

Prueba:	Funcionalidad del módulo de Predicción (Machine Learning).		
Fecha:	01/12/2025		
Tema del software:	Aplicativo de predicción mediante Machine Learning de la producción de banano en la Hacienda San Juan.		
Objetivo de prueba:	Evaluar la ejecución de los algoritmos de predicción y la visualización de los resultados estimados.		
	Funciones del Sistema	Calificar Sí / No	Observaciones
Ejecución del Modelo			
Generar predicción	X		Verificar que al pulsar "Generar Predicción", el sistema procese los

		datos y devuelva un valor estimado de cajas.
Preprocesamiento automático	X	Validar que el sistema limpie valores nulos o atípicos antes de ejecutar el modelo.
Visualización		
Ver Dashboard	X	Confirmar que el tablero principal muestre la predicción actual de forma destacada.
Gráficos comparativos	X	Validar que se generen correctamente los gráficos de línea comparando "Producción Real" vs "Predicción".
Observaciones Generales:		El modelo se ejecuta correctamente, integrando los datos históricos y mostrando resultados coherentes en la interfaz (REQF-007, REQF-008, REQF-009).
Desarrollador del sistema:		Anchundia Zambrano Andres David
Encargado de la prueba:		Anchundia Zambrano Andres David
Elaborado por: Anchundia, 2025		

Tabla 37.

Prueba de funcionalidad del módulo de Reportes

Prueba:	Funcionalidad del módulo de Reportes.
Fecha:	01/12/2025
Tema del software:	Aplicativo de predicción mediante Machine Learning de la producción de banano en la Hacienda San Juan.
Objetivo de prueba:	Verificar la generación y descarga correcta de los informes de producción y predicción.

Funciones del Sistema	Calificar Sí / No	Observaciones
Generación de Reportes		
Filtrar reportes	X	Validar que se puedan filtrar los datos por rangos de fecha o por Lote antes de generar el reporte.
Exportar a PDF	X	Verificar que el sistema genere y descargue un archivo PDF legible con las tablas y gráficos.
Exportar a Excel	X	Confirmar que se descargue un archivo .xlsx con los datos estructurados para análisis externo.
Observaciones Generales:		La exportación funciona en ambos formatos y los filtros de fecha responden correctamente (REQF-010).
Desarrollador del sistema:		Anchundia Zambrano Andres David
Encargado de la prueba:		Anchundia Zambrano Andres David
Elaborado por: Anchundia, 2025		

Tabla 38.

Prueba de usabilidad del sistema.

Prueba:	Usabilidad del sistema.
Fecha:	02/12/2025
Software:	Aplicativo de predicción mediante Machine Learning de la producción de banano.
Objetivo Prueba:	Evaluar la experiencia del administrador en la navegación, claridad y facilidad de uso del aplicativo.

Información y funcionamiento del sistema	y Calificar del Sí / No	Observaciones
El inicio de sesión es claro y funcional.	X	El proceso es estándar y rápido; los mensajes de error son claros.
La navegación entre módulos es intuitiva.	X	El menú lateral permite acceder fácilmente a Captura, Predicción y Reportes.
La interfaz es adaptable (Responsive).	X	El sistema se ajusta correctamente al abrirse en una Tablet o Laptop.
Los gráficos son fáciles de interpretar.	X	Se usan colores diferenciados para "Real" y "Predicho", facilitando la lectura.
Los formularios son claros.	X	Los campos de registro (clima, suelo) tienen etiquetas claras y validaciones visibles.
El tiempo de respuesta es adecuado.	X	Las predicciones y gráficos cargan en menos de 5 segundos.
Observaciones Generales:		El sistema cumple con los requisitos no funcionales de usabilidad y adaptabilidad. La interfaz limpia facilita la gestión agrícola sin requerir conocimientos técnicos avanzados.
Desarrollador del sistema:		Anchundia Zambrano Andres David
Encargado de la prueba:		Anchundia Zambrano Andres David

Elaborado por: Anchundia, 2025

Anexo N° 22: Lugar de implementación**Figura 25.*****Vista de la hacienda San Juan desde el mirador*****Elaborado por: Anchundia, 2026****Figura 26.*****Mostrando el aplicativo al administrador*****Elaborado por: Anchundia, 2026**

Figura 27.

Visualización del aplicativo para el administrador



Elaborado por: Anchundia, 2026

Figura 28.

Encuesta de satisfacción para el administrador



Elaborado por: Anchundia, 2026

Figura 29.

Entrega del aplicativo web para la hacienda San Juan



Elaborado por: Anchundia, 2026

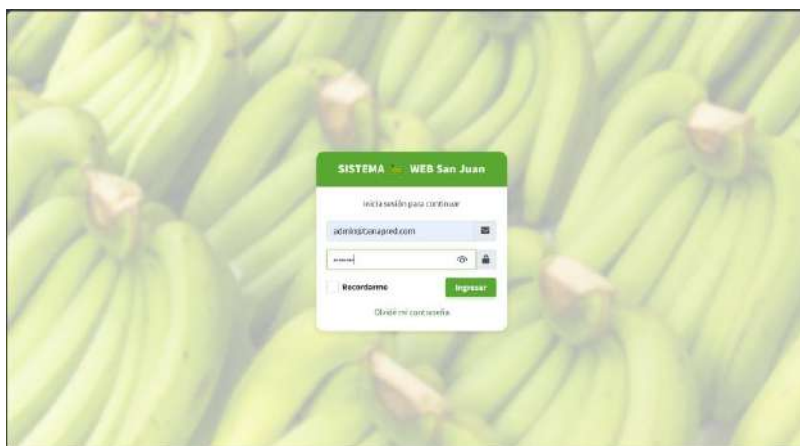
APÉNDICES

Apéndice N°1: Manual de Usuario – Aplicativo Web

El aplicativo web tiene como propósito el proporcionar una plataforma de interfaz intuitiva, accesible, amigable para realizar registro, análisis y precoción de la producción de banano. A continuación, se explica cada una de las funcionalidades y contenido del aplicativo para su uso correcto.

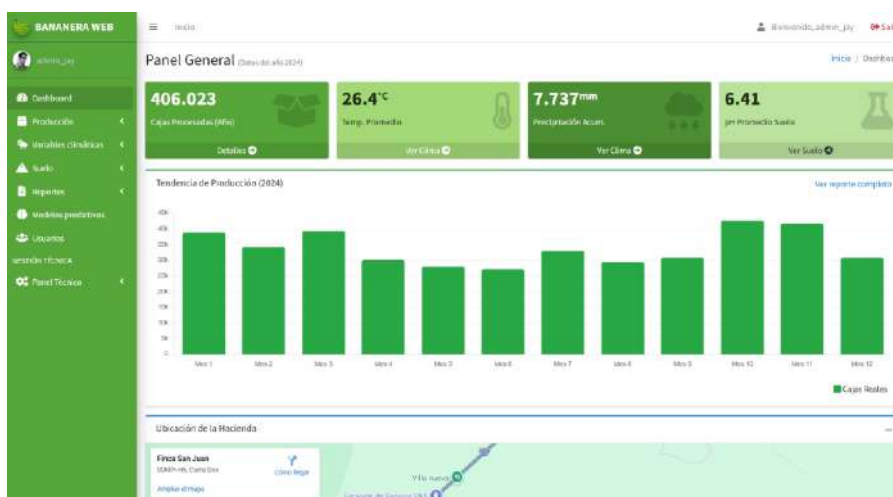
Interfaz de inicio de sesión

La primera página que encontraremos será la página de inicio de sesión, en esta el usuario en este caso el administrador podrá ingresar las credenciales las cuales son el correo y la contraseña para acceder al aplicativo web.



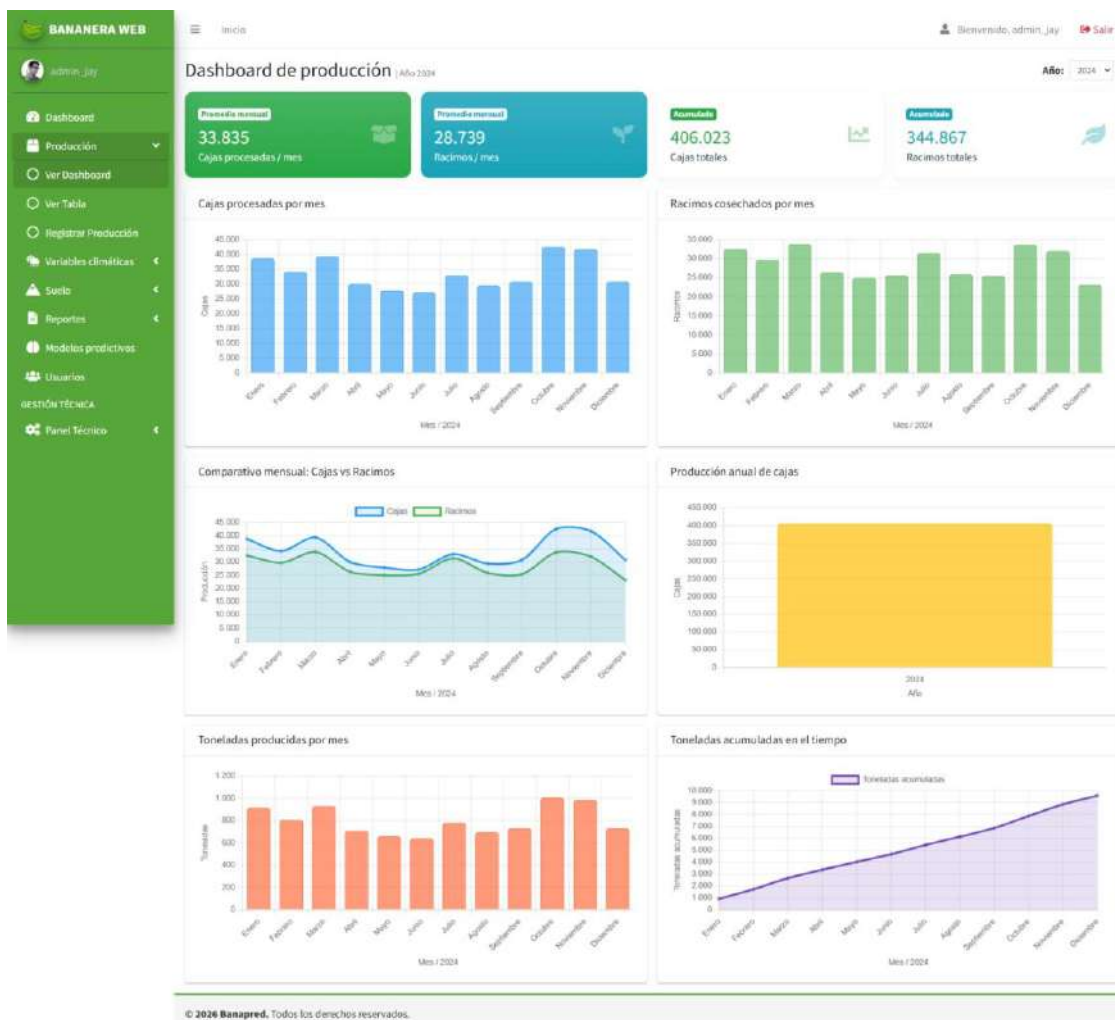
Pantalla de Inicio

Al iniciar sesión la primera pantalla que verá el usuario será, donde se muestra un resumen de la producción promedio del año, así como un gráfico general de la tendencia de la producción del presente año y a un lado encontrará un menú despegable en el que puede encontrar todas las pantallas del sistema.



Pantallas de producción

En las pantallas de producción se puede encontrar un resumen mensual de producción en las tarjetas, así como graficos de las cajas procesadas en el año seleccionado en el filtro.



En la segunda opción dentro de las pantallas de producción podemos encontrar ver tabla de producción otra forma de visualizar la cantidad de producción realizada en el año seleccionado en el cual el usuario visualizara mediante una tabla la cantidad exacta de la producción que se registró en el día, también se puede seleccionar el año que se desee visualizar la producción así como un botón para buscar un valor exacto de producción.

Fecha	Lote	Cajas Procesadas	Racimos Cosechados	Peso Total Est. (lb)
2024-01-01		1038	841	25,450.00
2024-01-02		1022	832	25,300.00
2024-01-03		1012	848	25,300.00
2024-01-04		951	838	23,775.00
2024-01-05		1036	805	25,250.00
2024-01-06		956	870	23,900.00
2024-01-07		1001	875	26,025.00
2024-01-08		1809	1138	32,725.00
2024-01-09		1401	1100	35,625.00
2024-01-18		1353	1156	34,825.00

Mostrando registros del 1 al 10 de un total de 357 registros.

En la tercera pantalla de producción se podrá realizar el registro diario de la producción mediante un cuestionario, el usuario deberá rellenar la formulario para guardar el registro.

Nuevo Registro Semanal

Fecha de Registro: 20/01/2024 Semana #: Ej 30

Lote / Sector: General

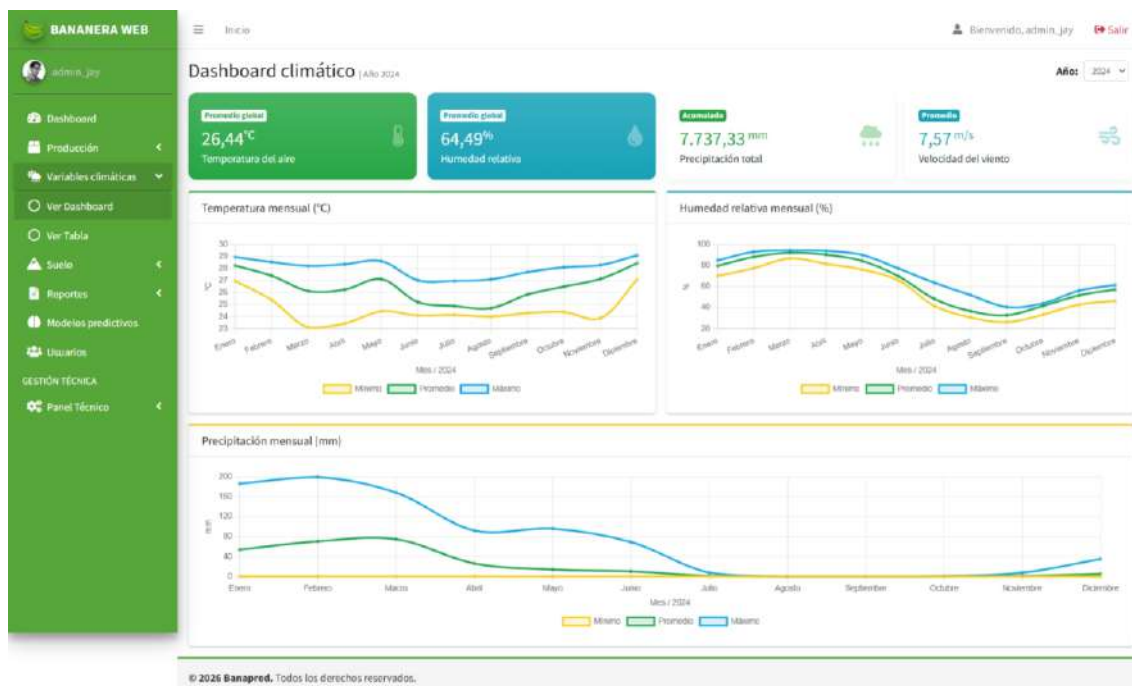
Cajas Procesadas: 0 Racimos Cosechados: 0

Observaciones:

© 2026 Banapred. Todos los derechos reservados.

Pantallas de variables meteorológicas

Dentro de las pantallas de variables meteorológicas el usuario podrá visualizar los datos climáticos de la hacienda los cuales son tomados por una estación meteorológica durante el año, así como tarjeta de resúmenes mensuales y anuales de las variables climáticas mas importante para el desarrollo de la planta de banano.

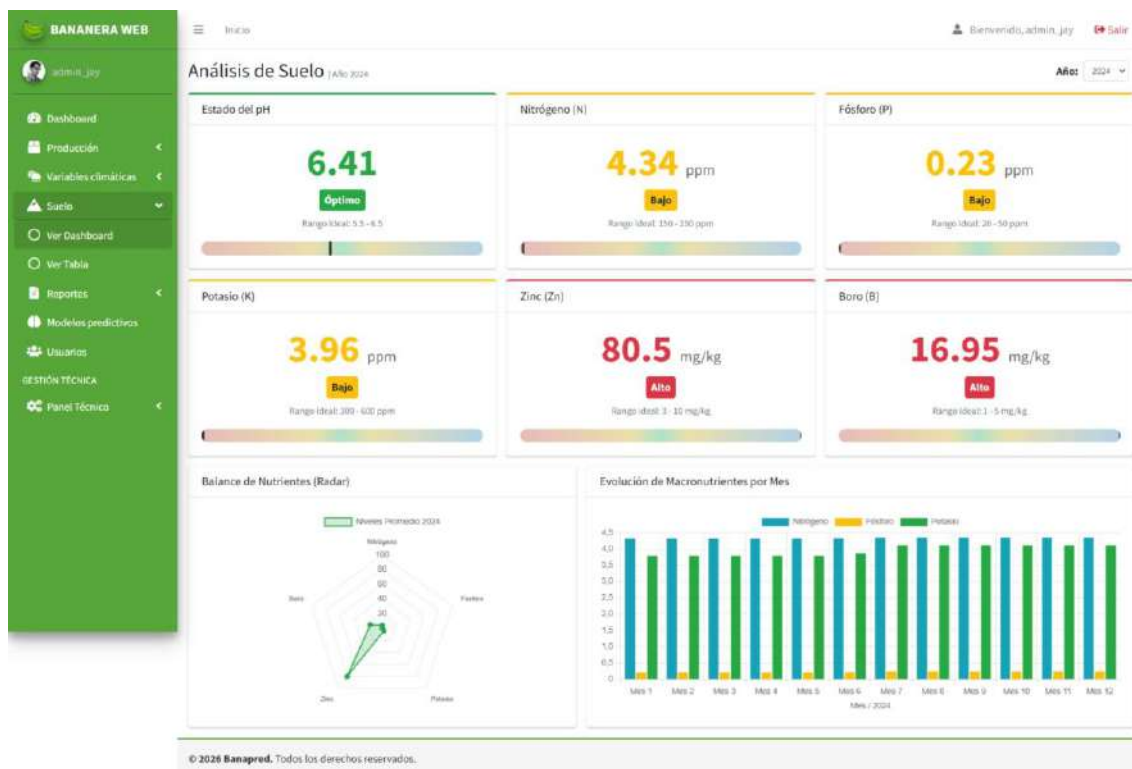


La segunda pantalla de las variables meteorológicas podremos visualizar los datos en forma de tabla, estos datos corresponden a un resumen diario de los 366 días del año seleccionado.

Fecha	Temp. Media (°C)	Humedad Media (%)	Precipitación (mm)	Vel. Viento (m/s)	Radiación (W/m ²)
2024-01-01	28,87	73,84	38,24	9,49	5,769
2024-01-02	27,19	78,18	36,82	8,53	2,524
2024-01-03	29,22	76,35	36,73	9,08	11,698
2024-01-04	28,65	77,45	37,84	9,65	6,460
2024-01-05	28,05	78,50	39,91	8,78	10,375
2024-01-06	28,59	75,12	37,59	9,04	11,801
2024-01-07	27,29	74,38	38,15	9,53	6,984
2024-01-08	28,18	78,89	65,20	9,20	5,706
2024-01-09	27,50	75,87	65,93	9,66	9,637
2024-01-10	28,14	79,58	61,39	8,78	8,172

Pantallas de nutrientes del suelo

En la primera pantalla de nutrientes de suelo podrá visualizar el ultimo análisis de suelo que se realizó dentro de ese año y los niveles de los nutrientes el cual puede ayudarlo a la hora de la toma de decisiones. Cuenta con un filtro en el cual el usuario puede elegir ver como estuvieron los nutrientes el año anterior a el presente.



En la segunda pantalla de los nutrientes del suelo se puede visualizar los datos en formato de tabla resumido por días, así mismo como en las variables meteorológicas se puede realizar búsquedas dentro de los datos de la tabla.

Tabla de Datos de Suelo Año: 2024

Registros de Suelo Diarios

Fecha	Lote	pH	Nitrógeno (N)	Fósforo (P)	Potasio (K)	Zinc	Boro
2024-01-01		6.72	4.318	0.215	3.793	80.129	15.514
2024-01-02		6.56	4.318	0.215	3.793	80.129	15.514
2024-01-03		6.49	4.318	0.215	3.793	80.129	15.514
2024-01-04		6.62	4.318	0.215	3.793	80.129	15.514
2024-01-05		6.35	4.318	0.215	3.793	80.129	15.514
2024-01-06		6.63	4.318	0.215	3.793	80.129	15.514
2024-01-07		6.29	4.318	0.215	3.793	80.129	15.514
2024-01-08		6.61	4.318	0.215	3.793	80.129	15.514
2024-01-09		6.31	4.318	0.215	3.793	80.129	15.514
2024-01-10		6.41	4.318	0.215	3.793	80.129	15.514

Mostrando registros del 1 al 10 de un total de 357 registros

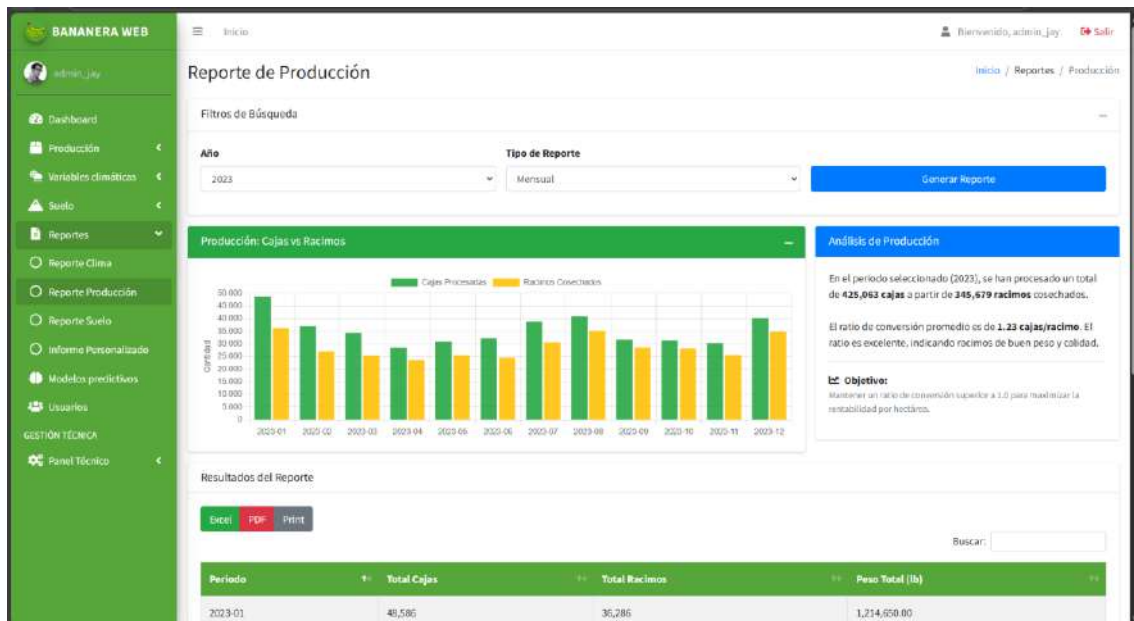
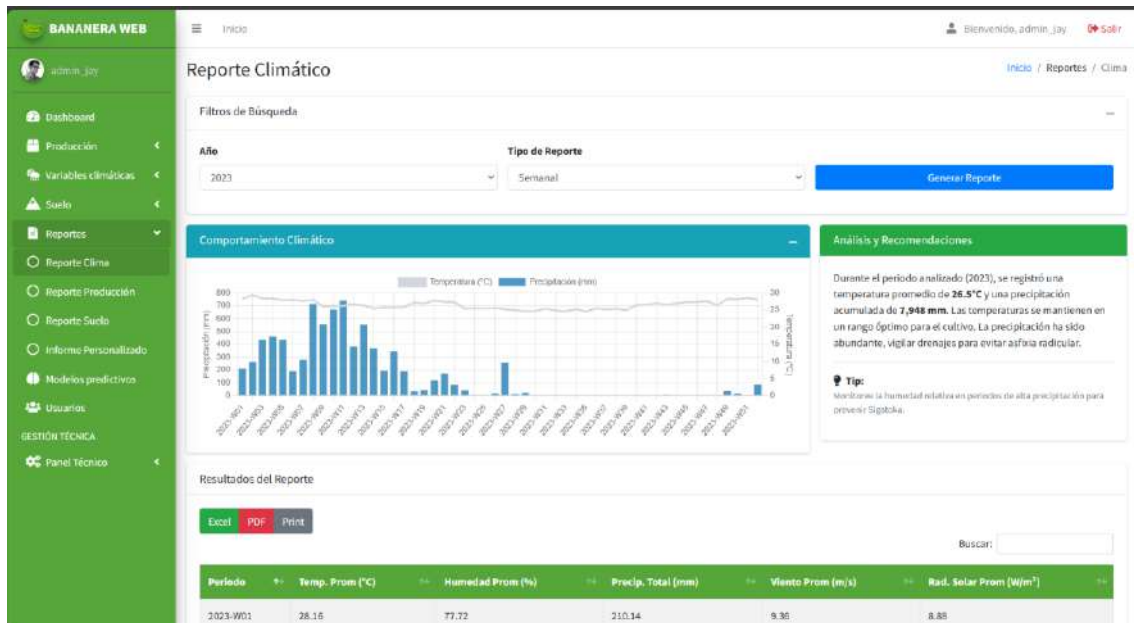
Anterior 1 2 3 4 5 ... 36 Siguiente

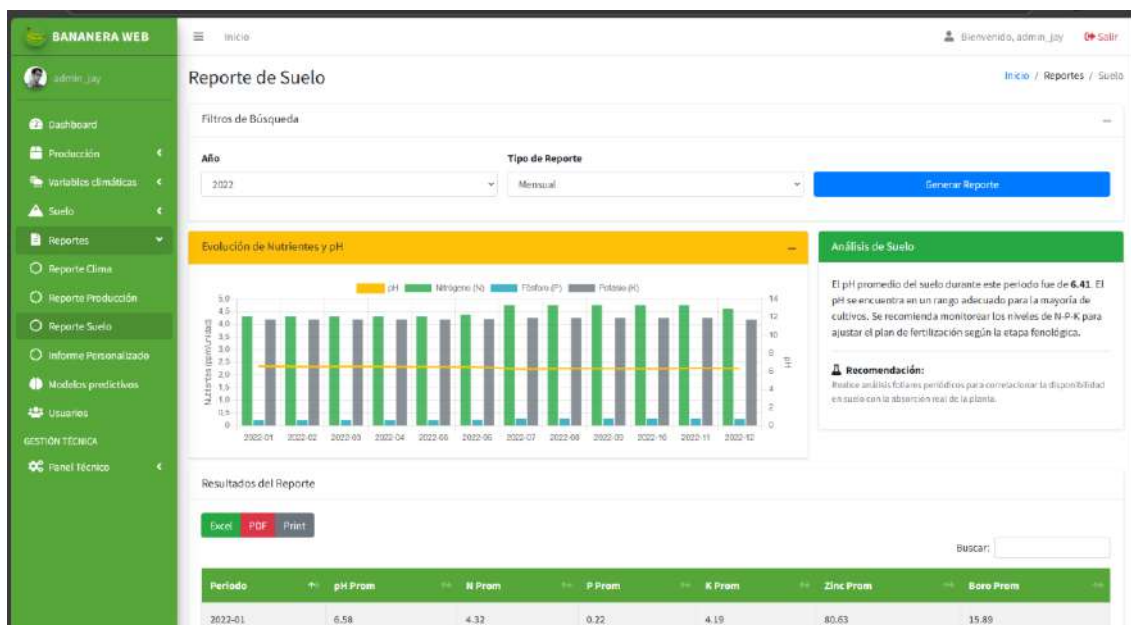
http://produccion.mient.or.cr/banapred/produccion/variables/meteorologicas.html

Pantallas de informes.

Dentro de estas pantallas el usuario podrá realizar reportes de los parámetros antes mencionados como serian, reportes de variables meteorológicas, reportes de los nutrientes de suelo y reportes de producción. Se podrá seleccionar el año del cual desee generar el reporte y así como la

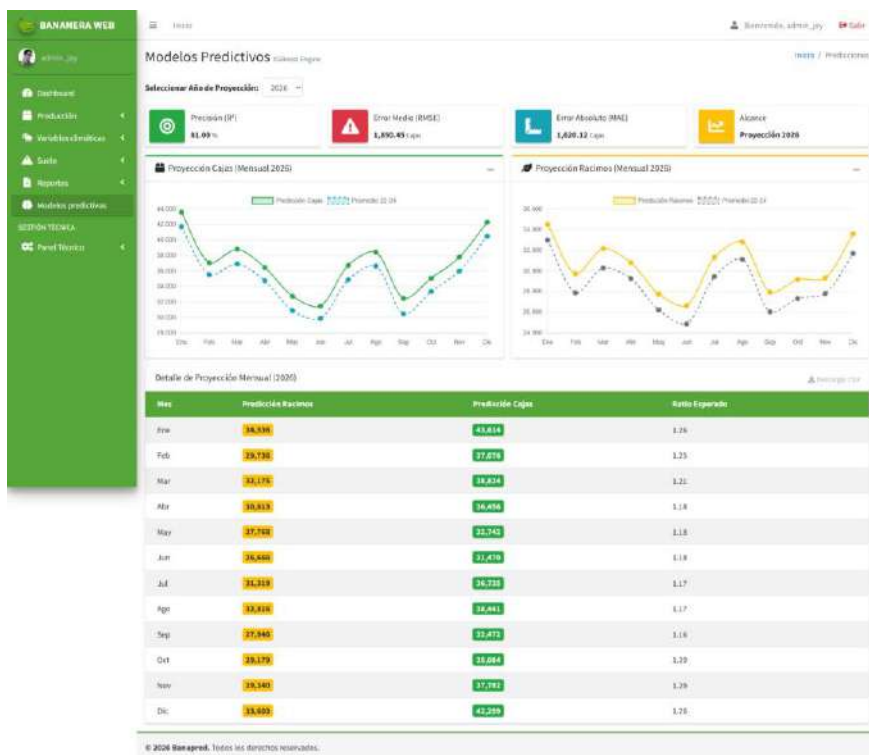
temporada de la cual se quiere realizar el reporte sea mensual, anual, semanal o semestral. Los reportes se podrán exportar mediante Excel PDF o Print.





Pantalla de predicción

Dentro de la pantalla de predicción el usuario podrá visualizar lo que sería la predicción de la producción para el siguiente año, las variables que podrá visualizar será la cantidad de racimos y cajas procesadas mensualmente. Estos datos se visualizarán mediante una gráfica y mediante una tabla para un mayor detalle y comprensión del usuario.



Pantalla de Técnico

Las pantallas de tecnico solo podran ser vistas por el usuario con rol de tecnico, la primera pantalla que visualizaremos es un dashboard donde nos dice si el nodo en este caso la estación metereologica se encuentra activa, asi como la alerta de datos, y los datos sospechosos.

BANANERA WEB Inicio Bienvenido, admin_jay Salir

Panel Técnico

1 Nodos Activos 0 Alertas Pendientes 0 Datos Sospechosos Validar

Estado de Nodos

Nodo/Lote	Última Conexión	Estado
Nodo Principal	2026-01-21 03:29:14	Activo

Acciones Rápidas

Validar Datos Respuestas Usuarios

© 2026 Banapred. Todos los derechos reservados.

Pantalla de Validación de datos

En esta pantalla se podran visualizar los ultimos datos guardados en el sistema, se podran editar por si el administrador de la hacienda lo requiere. Asi como se podra marcar si el valor esta fuera de el rango promedio para la toma de decision de editar o eliminar dicho dato.

BANANERA WEB Inicio Bienvenido, admin_jay Salir

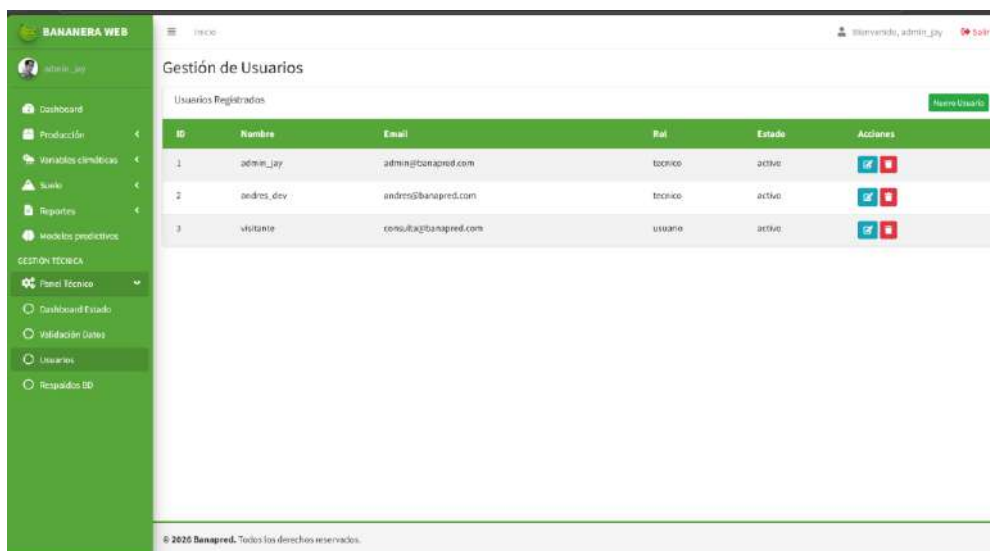
Validación de Datos

Registros Recientes

Fecha	Lote	Temp Prom	Precip	pH	Estado	Acciones
2024-12-22		25.22	12.14	6.10	valido	Editar Marcar Sospechoso
2024-12-21		26.67	12.09	6.19	valido	Editar Marcar Sospechoso
2024-12-20		27.40	12.05	6.18	valido	Editar Marcar Sospechoso
2024-12-19		27.99	12.76	6.15	valido	Editar Marcar Sospechoso
2024-12-18		27.36	11.87	6.16	valido	Editar Marcar Sospechoso
2024-12-17		25.25	12.85	6.11	valido	Editar Marcar Sospechoso
2024-12-16		27.99	12.21	6.39	valido	Editar Marcar Sospechoso
2024-12-15		27.31	0.00	6.35	valido	Editar Marcar Sospechoso
2024-12-14		25.66	0.00	6.43	valido	Editar Marcar Sospechoso
2024-12-13		29.18	0.00	6.36	valido	Editar Marcar Sospechoso
2024-12-12		26.50	0.00	6.10	valido	Editar Marcar Sospechoso
2024-12-11		27.43	0.00	6.23	valido	Editar Marcar Sospechoso
		26.62	0.00	5.96	valido	Editar Marcar Sospechoso

Pantalla de Gestión de usuarios.

Dentro de esta pantalla el tecnico podra editar los usuarios que tiene acceso al sistema en este caso solo el y el administrador. No obstante se cuenta con un perfil de visitante. Puede realizar acciones como editar los datos del usuario o eliminarlo de la base de datos.



Gestión de Usuarios

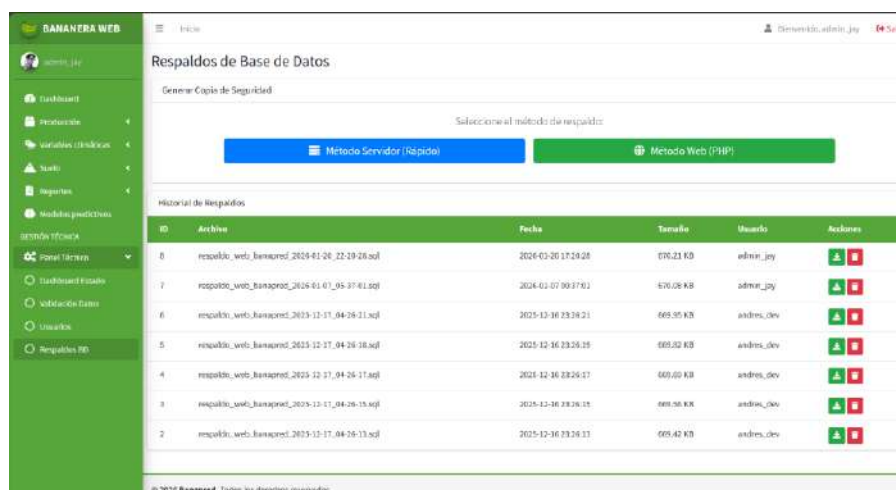
Usuarios Registrados [Nuevo Usuario](#)

ID	Nombre	Email	Rol	Estado	Acciones
1	admin_jay	admin@banapred.com	tecnico	activo	Editar Eliminar
2	andres_dev	andres@banapred.com	tecnico	activo	Editar Eliminar
3	visitante	consulta@banapred.com	usuario	activo	Editar Eliminar

© 2025 Banapred. Todos los derechos reservados.

Pantalla de Gestión de usuarios.

Pantalla de respaldo de base de datos, dentro de esta pantalla se podra realizar una copia de seguridad de la base de datos por si algun problema pueda surgir. Se guardan las ultimas copias de basesde datos que se realizaron y se pueden descargar o se puede generar el respaldo mediante un metodo rapido o el metodo web en caso de que uno u otro no llegase a funcionar por conectividad



Respaldo de Base de Datos

Generar Copia de Seguridad

Seleccione el método de respaldo:

[Metodo Servidor \(Rápido\)](#) [Metodo Web \(PHP\)](#)

Historial de Respaldos

ID	Archivo	Fecha	Tamaño	Usuario	Acciones
8	respaldo_web_banapred_2025-01-26_22-29-26.sql	2025-01-26 17:29:26	676.21 KB	admin_jay	Descargar Eliminar
7	respaldo_web_banapred_2025-01-01_05-37-41.sql	2025-01-01 00:37:41	676.08 KB	admin_jay	Descargar Eliminar
6	respaldo_web_banapred_2025-12-31_04-26-21.sql	2025-12-30 23:26:21	665.95 KB	andres_dev	Descargar Eliminar
5	respaldo_web_banapred_2025-12-31_04-26-38.sql	2025-12-30 23:26:38	665.92 KB	andres_dev	Descargar Eliminar
4	respaldo_web_banapred_2025-12-31_04-26-17.sql	2025-12-30 23:26:17	669.00 KB	andres_dev	Descargar Eliminar
3	respaldo_web_banapred_2025-12-31_04-26-15.sql	2025-12-30 23:26:15	669.98 KB	andres_dev	Descargar Eliminar
2	respaldo_web_banapred_2025-12-31_04-26-13.sql	2025-12-30 23:26:13	669.42 KB	andres_dev	Descargar Eliminar

© 2025 Banapred. Todos los derechos reservados.

Apéndice N°2: Manual de Técnico.**UNIVERSIDAD AGRARIA DEL ECUADOR****FACULTAD DE CIENCIAS AGRARIAS****CARRERA COMPUTACIÓN**

**APLICATIVO DE PREDICCIÓN MEDIANTE MACHINE LEARNING
DE LA PRODUCCIÓN DE BANANO EN LA HACIENDA SAN JUAN**

Elaborado por: Anchundia**Enero de 2026****1 INTRODUCCIÓN**

El sistema Banapred es una aplicación web diseñada para el monitoreo, registro y predicción de la producción de banano, correlacionando datos agrícolas (producción) con variables climáticas y de suelo. Proporciona una interfaz administrativa completa para gestionar y visualizar estos datos a través de tablas dinámicas, gráficos y diferentes roles de usuarios.

2 ARQUITECTURA DEL SISTEMA

La solución adopta una arquitectura cliente-servidor para la aplicación web y una arquitectura basada en microservicios para el motor predictivo. La capa de presentación corre sobre AdminLTE/Bootstrap y JavaScript; la capa backend principal está en PHP (PDO + MySQL). El módulo analítico (API_modelo) corre como un servicio HTTP (FastAPI + Uvicorn) que expone endpoints de entrenamiento e inferencia.

3 STACK TECNOLÓGICO***Backend (web):***

- PHP 8+
- PDO
- MySQL/MariaDB.

Frontend:

- AdminLTE v3 (Bootstrap 4)
- Chart.js
- DataTables
- jQuery
- SweetAlert2
- Toastr.

API/ML:

- Python
- FastAPI
- Uvicorn
- Pandas
- NumPy
- Scikit-Learn
- XGBoost.

4 Estructura de directorios principal

La aplicación maneja el siguiente árbol o modelo jerárquico de recursos:

- assets/: Contiene de manera globalizada todos los assets web client-side.
 - css/, js/: Scripts nativos, ej. validaciones de la UI mediante produccion.js o dashboard.js.
 - img/, dist/: Recursos de imagen e implementaciones de AdminLTE.
 - plugins/: Vendor/Dependencias de terceros en versión estática (Bootstrap, DateRangePicker, DataTables, etc).
- auth/: Gestión de seguridad e inicio de sesión.
 - login.php, logout.php, validar_login.php, forgot-password.php.
- backups/: Directorio destinado al almacenamiento de las copias de seguridad de la base de datos generadas por el sistema en formato SQL unificado.
- config/: Configuraciones críticas que unen backend con base de datos u otras credenciales.
 - db.php: Setup principal de la conexión segura (PDO MySQL).

- exports/: Carpeta temporaria designada a la generación y descarga de los reportes masivos a exportar.
- includes/: Fragmentos globales consumidos dinámicamente ("Require / Include").
 - header.php, footer.php, sidebar.php.
 - repositories/: Contiene instancias como ProduccionRepository.php.
- pages/: Core principal de Vistas y Rutas, segmentadas por jerarquía o privilegio.
 - admin/: Vistas de Alto Perfil con reportes gerenciales.
 - tecnico/: Vistas técnicas, administración perimetral y soporte.

5 API de Modelos Predictivos (endpoints)

El microservicio expone los siguientes endpoints principales:

- GET / y GET /health: verificación de estado (200).
- POST /train: reentrena el modelo a partir del CSV maestro (parámetro granularity=monthly|raw).
- POST /predict: inferencia unitaria; recibe un JSON con las features y devuelve predicciones {"pred_racimos": PR, "pred_cajas": PC}.
- GET /forecast/26w: pronóstico para las próximas 26 semanas; opcional ?export_csv=1 para descarga.
- GET /forecast/36m: pronóstico para 36 meses (3 años).
- GET /metrics y GET /predictions/export: exportadores de métricas y log de predicciones.

6 Algoritmo predictivo y preprocesamiento

El motor predictivo utiliza XGBoost (XGBRegressor) con parámetros optimizados para estabilidad. Se generan features temporales (lags, rolling means) para capturar dinámica de series de tiempo. El flujo de trabajo incluye: limpieza (preprocess), agregaciones mensuales (monthly_aggregate), entrenamiento, evaluación (r2, mae, rmse) y serialización del modelo.

```

model = xgb.XGBRegressor(
    objective="reg:squarederror",
    n_estimators=600,
    learning_rate=0.03,
    max_depth=4,
    subsample=0.8,
    colsample_bytree=0.8,
    min_child_weight=5,
    gamma=0.1,
    random_state=42,
    n_jobs=-1
)

model.fit(X_train, y_train)
pred = model.predict(X_test)

rmse = float(np.sqrt(mean_squared_error(y_test, pred)))
mae = float(mean_absolute_error(y_test, pred))
r2 = float(r2_score(y_test, pred))
return model, {"rmse": rmse, "mae": mae, "r2": r2}

```

7 Configuración de base de datos y variables de entorno

A continuación de muestra la forma en la que se realiza la Conexión PDO (config/db.php), de la base de datos:

```

<?php
// config/db.php

$DB_HOST = 'localhost';
$DB_NAME = 'produccion';
$DB_USER = 'produccion';
$DB_PASS = 'produccion2025*';

try {
    $pdo = new PDO(
        "mysql:host=$DB_HOST;dbname=$DB_NAME;charset=utf8mb4",
        $DB_USER,
        $DB_PASS,
        [
            PDO::ATTR_ERRMODE => PDO::ERRMODE_EXCEPTION,
            PDO::ATTR_DEFAULT_FETCH_MODE => PDO::FETCH_ASSOC,
        ]
    );
} catch (PDOException $e) {
    header('location: ../pages/error.php');
    exit;
}

```

8 Autenticación, roles y seguridad

La aplicación web verifica sesiones PHP (session_status) y la existencia de \$_SESSION['usuario'] para mantener acceso.

- **Roles principales:** Administrador (acceso a dashboards, modelos, reportes) y Técnico/Operativo (mantenimiento, backups, validación).
- **Para la API:** autenticación por x-api-key en cabeceras y control de acceso a endpoints de entrenamiento/forecast.
- **Buenas prácticas:** no exponer credenciales en repositorios, almacenar secrets en .env y limitar permisos de usuario en la BDD.

9 Despliegue y consideraciones operativas

- **Permisos de SO:** asegurar permisos de lectura/escritura en backups/, exports/ y models/ (CHMOD 755/775 o equivalente).
- **Respaldos:** crear dumps periódicos de la BDD y mantener control de versiones de modelos en models/.
- **Contenedores:** Considerar desplegar la API_modelo en un contenedor Docker y servir con Uvicorn/Gunicorn detrás de un reverse proxy (nginx).
- **Monitorización:** endpoints /health para checks; logs en predictions_log.csv y métricas en metrics_latest.json.