



**UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE QUITO**

CARRERA DE COMPUTACIÓN

SISTEMA DE PREDICCIÓN EN LA PRODUCCIÓN DE ARÁNDANOS

Trabajo de titulación previo a la obtención del
Título de Ingeniero en Ciencias de la Computación

AUTOR: SANCHEZ APRÁEZ ANDY KEVIN

TUTOR: GUSTAVO ERNESTO NAVAS RUILOVA

Quito - Ecuador
2024

**CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE
TITULACIÓN**

Yo, Andy Kevin Sánchez Apráez con documento de identificación N° 1725636847, manifiesto que:

Soy el autor y responsable del presente trabajo; y, autorizo a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total parcial el presente trabajo de titulación.

Quito, 15 de agosto de 2024

Atentamente,



Andy Kevin Sánchez Apráez
1725636847

**CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE
TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA**

Yo, Andy Kevin Sánchez Apráez con documento de identificación N° 1725636847, expreso mi voluntad y por medio del presente documento cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que soy autor del Proyecto Técnico: “Sistema de predicción en la producción de arándanos”, el cual ha sido desarrollado para optar por el título de: Ingeniero en Ciencias de la Computación, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribo este documento cuando entregaré el trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Quito, 15 de agosto de 2024

Atentamente,



Andy Kevin Sánchez Apráez
1725636847

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, Gustavo Ernesto Navas Ruilova con documento de identificación N° 1705675625, docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: SISTEMA DE PREDICCIÓN EN LA PRODUCCIÓN DE ARÁNDANOS, realizado por Andy Kevin Sánchez Apráez con documento de identificación N° 1725636847, obteniendo como resultado final el trabajo de titulación bajo la opción de Proyecto Técnico que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Quito, 15 de agosto de 2024

Atentamente,



Ing. Gustavo Ernesto Navas, MSc.
1705675625

DEDICATORIA

A mis padres y abuelo por haberme dado una buena crianza y haberme enseñado los valores que han hecho de mí no solo una persona profesional en las tareas que realizo, sino también una persona humanista que siente la necesidad de ayudar a los demás y que esto sea el principio de dejar un legado que aporte esperanza a la sociedad para que más estudiantes piensen en un título, no por necesidad, sino como una esperanza de que algún día podremos crear ideas a través de la ciencia y tecnología que puedan aportar al crecimiento colectivo.

Sánchez Apráez Andy Kevin.

CARTA DE AUTORIZACIÓN

Por medio del presente documento, el abajo firmante:

Por medio de la presente, queremos expresar nuestra conformidad con el trabajo realizado en relación con su tesis titulada "SISTEMA DE PREDICCIÓN EN LA PRODUCCIÓN DE ARÁNDANOS", además, autorizamos formalmente la publicación de su tesis.

Agradecemos su dedicación y esfuerzo en este proyecto y esperamos que continúe teniendo éxito en sus futuras investigaciones.

Para constancia de lo anterior, se firma en QUITO día 08/08/2024

Atentamente, Ing. Andrés Jorge Orozco Sánchez

Representante de Empresa Privada Mirandanos

Firma:



AGRADECIMIENTOS

A mis padres, que, con sus consejos y apoyo, me han motivado durante todo el trayecto de mi carrera.

A la Carrera en Ciencias de la Computación de la Universidad Politécnica Salesiana, donde me formé como profesional y persona.

A la Productora Ilaru por darme la oportunidad de poder realizar el presente trabajo de titulación.

Al Ing. Manuel Jaya, quien, con su amistad, apoyo y profesionalismo, me alentaron para culminar con éxito mi carrera.

Un agradecimiento muy especial al Ing. Gustavo Navas, que, con su guía, paciencia y sabios consejos me han encaminado la culminación del presente proyecto de grado.

A todos los profesores de la Carrera que compartieron sus conocimientos y sabiduría.

Sánchez Apráez Andy Kevin

ÍNDICE DE CONTENIDO

CAPÍTULO I.....	1
ANTECEDENTES Y GENERALIDADES	1
1.1 INTRODUCCIÓN	1
1.2 PROBLEMA DE ESTUDIO	3
1.2.1 <i>Antecedentes</i>	4
1.2.2 <i>Importancia</i>	6
1.2.3 <i>Delimitación</i>	8
1.3 JUSTIFICACIÓN	8
1.4 OBJETIVOS	10
1.4.1 <i>Objetivo General</i>	10
1.4.2 <i>Objetivos Específicos</i>	10
1.5 ALCANCE.....	12
CAPÍTULO II.....	14
MARCO TEÓRICO.....	14
1.6 ESTRUCTURA ORGANIZATIVA DE LA PRODUCTORA	14
1.6.1 <i>Área de Producción</i>	15
1.6.2 <i>Área de Logística</i>	16
1.7 VARIEDADES DE ARANDANOS.....	18
1.7.1 <i>Ciclo de cultivo</i>	19
1.7.2 <i>Fases del ciclo de cultivo</i>	20
1.7.3 <i>Variedades de arándanos y sus especificidades</i>	23
1.7.4 <i>Técnicas de cultivo</i>	27
1.8 FACTORES QUE AFECTAN EN LA PRODUCCIÓN.....	28
1.8.1 <i>Factores ambientales</i>	28
1.8.2 <i>Clima</i>	29
1.8.3 <i>Suelo</i>	29
1.8.4 <i>Riego</i>	30

1.8.5	<i>Factores biológicos</i>	31
1.8.6	<i>Polinización</i>	31
1.8.7	<i>Plagas y Enfermedades</i>	31
1.8.8	<i>Factores Agronómicos</i>	32
1.8.9	<i>Poda</i>	32
1.8.10	<i>Manejo del Suelo</i>	32
1.8.11	<i>Factores de Manejo Postcosecha</i>	32
1.8.12	<i>Transporte</i>	32
1.9	TECNOLOGÍA Y AUTOMATIZACIÓN	33
1.9.1	<i>Sistemas de Riego Automatizados</i>	33
1.9.2	<i>Sensores y Tecnología</i>	33
1.10	SISTEMAS DE PREDICCIÓN	33
1.10.1	<i>Aprendizaje automático</i>	33
1.10.2	<i>Modelos de predicción</i>	35
1.10.3	<i>Máquinas de Soporte Vectorial (SVM)</i>	35
1.10.4	<i>Redes Neuronales Artificiales (ANN)</i>	35
1.10.5	<i>Bosques Aleatorios (RF)</i>	35
1.11	RECOPIACIÓN DE DATOS	35
1.11.1	<i>Datos Locales</i>	38
1.11.2	<i>Datos Públicos</i>	40
1.11.3	<i>Monitoreo de Datos</i>	41
CAPÍTULO III		46
METODOLOGÍA		46
1.12	DISEÑO DE LA INVESTIGACIÓN	46
1.12.1	<i>Investigación preliminar</i>	46
1.12.2	<i>Recolección de datos</i>	47
1.12.3	<i>Selección de Datos</i>	48
1.12.4	<i>Enfoque Metodológico</i>	50

1.13	MODELADO.....	51
1.13.1	<i>Métricas de validación</i>	53
1.13.2	<i>Mean Squared Error (MSE)</i>	53
1.13.3	<i>R-squared (R²)</i>	54
1.13.4	<i>Mean Absolute Error (MAE)</i>	54
1.13.5	<i>Mean Absolute Percentage Error (MAPE)</i>	55
1.13.6	<i>Root Mean Squared Error (RMSE)</i>	55
1.13.7	<i>Explained Variance Score (EVS)</i>	55
1.14	DESARROLLO DEL SISTEMA	55
1.14.1	<i>Arquitectura del Sistema</i>	55
CAPÍTULO IV		57
RESULTADOS		57
1.15	ANÁLISIS DE DATOS.....	57
1.15.1	<i>Matriz de Correlación</i>	58
1.15.2	<i>Evaluación de Métricas</i>	62
1.16	RESULTADOS DE LA PREDICCIÓN	66
1.17	IMPACTO DEL SISTEMA DE PREDICCIÓN	70
CONCLUSIÓN		71
RECOMENDACIONES		72
BIBLIOGRAFÍA		73
GLOSARIO		75

ÍNDICE DE TABLAS

Tabla 1 Numero de Plantas por Variedad	20
Tabla 2 Dataset y Características	36
Tabla 3 Tecnología y Características	39
Tabla 4 Variables y Características.....	43
Tabla 5 Datos Usados	49
Tabla 6 Modelos Utilizados	52
Tabla 7 Influencia e Importancia de las Variables.....	59
Tabla 8 Interpretación y Comparación de Resultados	63

ÍNDICE DE FIGURAS

Figura 1 Historiografía de Producción vs Variedad.	6
Figura 2 Estructura Organizacional en las áreas de la Productora	15
Figura 3 Riego y Cuidado	16
Figura 4 Transporte del Producto	17
Figura 5 Control de Calidad.....	18
Figura 6 Planta Biloxi	23
Figura 7 Planta Emerald	24
Figura 8 Planta Legacy	25
Figura 9 Planta Star.....	26
Figura 10 Bloque 1 de Producción	29
Figura 11 Suelo Sustrato de Coco.....	30
Figura 12 Tanques de Mezcla	31
Figura 13 Sistema de Mezcla.....	40
Figura 14 Sistema de Agroclimatología	42
Figura 15 Arquitectura del Sistema	56
Figura 16 Matriz de Correlación.....	58
Figura 17 Predicciones vs Valores Reales (Regresión Polinómica).....	64
Figura 18 Predicciones vs Valores Reales (Árbol de Decisiones).....	65
Figura 19 Predicciones vs Valores Reales (Random Forest).....	66
Figura 20 Importancia de las Variables Basadas en la Ganancia	67
Figura 21 Importancia de las Variables Basadas en la Precisión	68

RESUMEN

La producción de arándanos en Ecuador enfrenta desafíos significativos debido a la variabilidad climática y la necesidad de optimizar el uso de recursos agrícolas. Este trabajo tiene como objetivo desarrollar un sistema de predicción para la producción de arándanos en la empresa Ilaru, utilizando técnicas de aprendizaje automático como Máquinas de Soporte Vectorial (SVM) y Bosques Aleatorios (RF). Se recopilaron y analizaron datos de clima, suelo y manejo agrícola para identificar las variables más influyentes en la producción de arándanos. La metodología incluyó la creación de un conjunto de datos que integrara información de diferentes fuentes y el entrenamiento de modelos predictivos. Los resultados obtenidos muestran una mejora significativa en la precisión de las predicciones, lo que permite a la empresa tomar decisiones más informadas y eficientes sobre la gestión de recursos y la planificación de la producción. En conclusión, el sistema propuesto no solo aumenta la productividad de los arándanos, sino que también promueve prácticas agrícolas más sostenibles y responsables.

Palabras Clave: Predicción, Arándanos, Aprendizaje Automático, Agricultura de Precisión, Optimización de Recursos.

ABSTRACT

Blueberry production in Ecuador faces significant challenges due to climate variability and the need to optimize the use of agricultural resources. This study aims to develop a prediction system for blueberry production at Ilaru, using machine learning techniques such as Support Vector Machines (SVM) and Random Forests (RF). Climate, soil, and agricultural management data were collected and analyzed to identify the most influential variables in blueberry production. The methodology included creating a dataset integrating information from various sources and training predictive models. The results obtained show a significant improvement in prediction accuracy, allowing the company to make more informed and efficient decisions regarding resource management and production planning. In conclusion, the proposed system not only increases blueberry productivity but also promotes more sustainable and responsible agricultural practices.

Keywords: Prediction, Blueberries, Machine Learning, Precision Agriculture, Resource Optimization.

CAPÍTULO I

ANTECEDENTES Y GENERALIDADES

1.1 INTRODUCCIÓN

El consumo de arándanos está en aumento a nivel mundial, y Ecuador no es una excepción. A pesar de ser un mercado emergente, los productores ecuatorianos están implementando diversas estrategias para impulsar la exportación, preocupados por los problemas del contrabando, este cultivo, prácticamente desconocido hace menos de una década se ha visto un crecimiento significativo gracias a pioneros como Diego Garzón, fundador de Bluebite en 2015. Actualmente, su empresa produce 10 toneladas de arándanos al mes para el mercado local.

La primera exportación de arándanos ecuatorianos por Hortifruit a Europa y Estados Unidos en 2022 marcó un hito importante, demostrando la calidad del producto nacional. Sin embargo, los productores están principalmente enfocados en el mercado local, vendiendo a supermercados, empresas y directamente a consumidores (Revista Líderes, 2023).

Aunque el mercado interno está en crecimiento, se anticipa una saturación en uno o dos años, lo que impulsa a la Federación a buscar mercados internacionales para los pequeños y medianos productores por lo que, para fortalecer la industria, Garzón cofundó la Federación Ecuatoriana de Productores y Exportadores de Arándano (Fepexa), promoviendo el cultivo y trabajando en la reducción de aranceles sobre insumos agrícolas (Revista Líderes, 2023).

Este es el caso de la productora Ilaru donde se ha enfocado el presente trabajo de estudio propiedad de Juan Francisco Canelos, que se enfrenta a estos desafíos significativos en la

planificación de la producción para salir a mercados internacionales. La empresa busca optimizar la producción de sus variedades de arándanos: Emerald, Biloxi, Legacy y Star, cada una con requerimientos específicos de riego y condiciones de suelo. Actualmente, Ilaru utiliza

plataforma Aazuli para gestionar y recopilar datos de producción, que incluyen información detallada sobre fecha, variedad, bloque, empleado, gramaje/calidad, cantidad y código.

Uno de los mayores desafíos que enfrentan los productores es el contrabando desde Perú, donde frutas de menor calidad se venden a precios significativamente más bajos, afectando la competitividad del producto ecuatoriano.

A pesar de las ventajas comparativas de Ecuador, como la capacidad de producir arándanos durante todo el año gracias a su clima y la alta calidad del agua y suelo, la variabilidad en las condiciones climáticas y las características específicas de cada variedad complican la planificación de la producción y el uso de recursos.

Sin un sistema de predicción preciso, la empresa enfrenta dificultades para anticipar los arándanos que cosechará en determinadas temporadas, lo que provoca un uso ineficiente de insumos agrícolas y afecta negativamente la toma de decisiones estratégicas, al implementar algoritmos para manejar estos datos y utilizar computadoras capaces de soportar el procesamiento de esta información es fundamental, esto puede ayudar a solventar necesidades de la productora, como la prevención de enfermedades, la decisión de cuándo regar, el control del pH, la humedad y la temperatura, entre otras variables módulos estadísticos y modelos de aprendizaje automático.

La implementación de tecnologías avanzadas no solo mejora la precisión en la predicción de rendimientos, sino que también optimiza el uso de recursos. (Herrera-Díaz, 2016). Exploró la integración de análisis estadísticos y predictivos con el uso de Big Data y Machine Learning en

la agricultura, enfatizando la automatización de procesos para mejorar la eficiencia en el cultivo, al utilizar dispositivos como sensores y actuadores ayuda a regular y administrar los recursos necesarios.

El aprendizaje automático y el análisis predictivo son componentes clave de esta práctica, permitiendo analizar grandes volúmenes de datos y extraer patrones útiles para la toma de decisiones, mientras más factores se aborden con relación a las afectaciones de los arándanos ya sean positivas o negativas, más variables se encontrarán para desarrollar un sistema más preciso y confiable en el entrenamiento de datos y, posteriormente, en la precisión de las predicciones.

Estos avances tecnológicos permiten a los agricultores responder más eficientemente a las necesidades de los cultivos, ajustando las variables que se puedan controlar en el proceso de crecimiento de la planta esto de acuerdo con las predicciones de producción como consecuencia se pretende tener ingresos en base a la reducción los costos operativos y mejora la sostenibilidad de las operaciones agrícolas y claro está en el aumento de la producción.

1.2 PROBLEMA DE ESTUDIO

Ilaru. enfrenta desafíos significativos en el rendimiento y calidad de su producción de arándanos debido a la falta de comprensión sobre las variables que afectan este rendimiento, la empresa necesita identificar las variables que influyen en el gramaje y la calidad de los arándanos de las cuatro variedades que se cultivan, esto es crucial para poder anticipar con precisión la cantidad de arándanos que se podrá cosechar en cada temporada y reducir la producción de frutos de calidad inferior, como los de tipo deforme, pequeña y muy pequeña. Al mejorar la proporción de frutos de calidad mediana y grande, la empresa puede incrementar su producción total y optimizar sus prácticas de cultivo.

El principal problema radica en la incapacidad de prever con exactitud la cantidad de insumos agrícolas necesarios para la producción de arándanos en un período determinado. Esta

incertidumbre dificulta la optimización de recursos y eleva los costos operativos.

Este desafío es especialmente crítico para las cuatro variedades de arándanos cultivadas por Ilaru. (Emerald, Biloxi, Legacy y Star), cada una con requerimientos específicos de agua y condiciones de suelo. Implementar un sistema de predicción preciso permitirá a Ilaru. mejorar la eficiencia de sus operaciones y asegurar un mayor rendimiento y calidad en su producción de arándanos.

Desarrollar mejores prácticas agrícolas basadas en las características específicas de cada variedad de arándano implicara el ajuste del uso de recursos agrícolas, como agua, fertilizantes y pesticidas, así como modificar los tiempos de cosecha según las predicciones derivadas de análisis precisos con estos se reducirán significativamente los costos operativos y mejorar la eficiencia en el uso de insumos teniendo un impacto positivo en la toma de decisiones estratégicas, abarcando la asignación de recursos, la planificación logística y la gestión de la comercialización de los productos.

1.2.1 Antecedentes

Juan Francisco Canelos, dueño de la productora Ilaru donde se ha enfocado el presente estudio, identificó esta oportunidad y en 2022 cambió el cultivo de flores por arándanos en su finca en Tabacundo. Su empresa, junto con seis socios, cultiva arándanos en 18 hectáreas. El 80% de su producción se vende en supermercados, el 5% en panaderías, pastelerías y fruterías, y el resto a domicilio. Aunque actualmente se enfocan en el mercado local, planean iniciar exportaciones este año (Canelos, 2022).

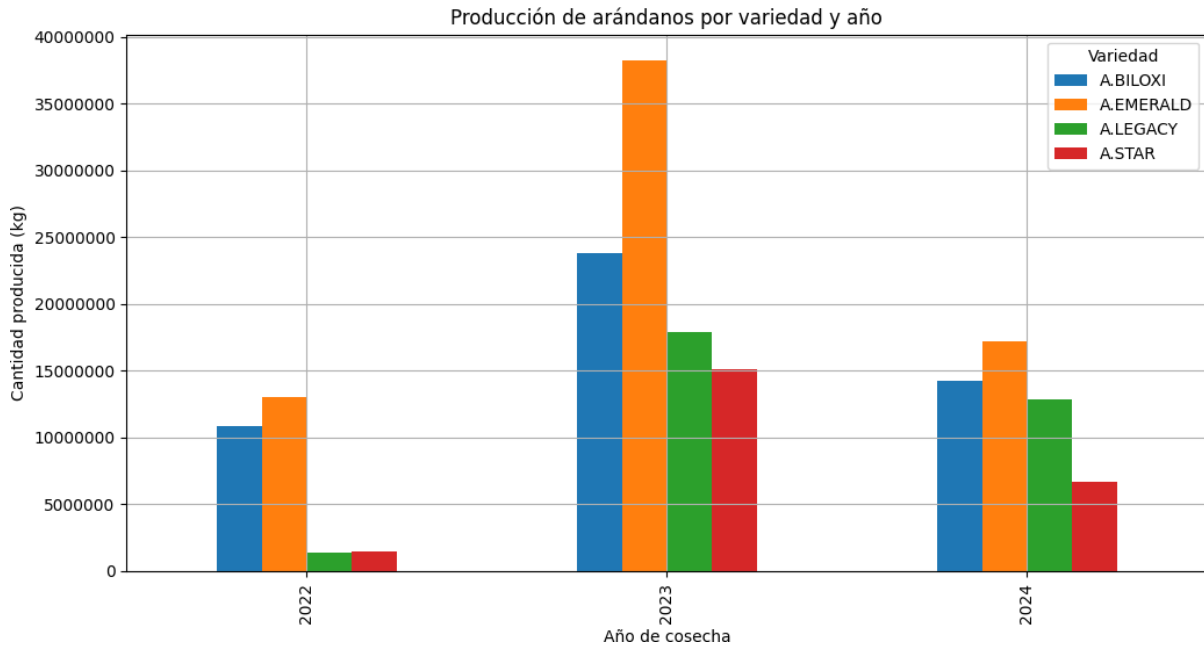
No obstante, el mercado de arándanos en Ecuador enfrenta desafíos, como el contrabando desde Perú, donde los arándanos se venden a precios muy bajos (USD 4 por kilo) en comparación con el producto ecuatoriano (USD 18-24 por kilo en supermercados). Este contrabando afecta a

los productores locales, quienes necesitan una inversión inicial alta de aproximadamente USD 120,000 por hectárea para cubrir costos de planta, sistema de riego e invernaderos (Revista Líderes, 2023). A pesar de estos retos, el mercado local sigue siendo competitivo debido a la alta calidad del producto ecuatoriano.

Para abordar estos desafíos, es esencial desarrollar un sistema de predicción que mejore la capacidad de planificación y gestión de la productora la implementación de un sistema de predicción basado en aprendizaje automático y análisis de datos permitirá a Ilaru. anticipar con mayor precisión la producción de arándanos, optimizando el uso de recursos y mejorando la planificación de la producción, con un sistema que permita el ingreso de una fecha de siembra y de cosecha y que dará como resultado una predicción en la producción de los arándanos de acuerdo a cada variedad que maneja la productora ayudando así a prever que recurso utilizar optimizando así estos de acuerdo a un previo análisis y de la madurez del código desarrollado.

Figura 1

Historiografía de Producción vs Variedad.



Nota: La grafica muestra la cantidad de arándanos producida desde el año 2022 al presente acorde a la variedad. Elaborado por: El autor.

1.2.2 Importancia

La falta de una predicción precisa complica la gestión logística y la comercialización, ya que no se puede garantizar un suministro constante y adecuado para satisfacer la demanda del mercado local.

Sin un sistema de predicción que identifique las variables que afectan el rendimiento de la producción, es difícil producir frutos de alta calidad con el gramaje adecuado para la comercialización.

La disparidad en el tiempo de cosecha entre las cuatro variedades de arándanos (Emerald, Biloxi, Legacy y Star) incrementa los costos operativos y dificulta la planificación. Al armonizar

los tiempos de cosecha, se podría optimizar la producción y mejorar la calidad del producto, asegurando frutos uniformes en tamaño y calidad, y reduciendo el número de arándanos descartados por ser demasiado pequeños. Esto aumentaría el número de frutos de calidad mediana, grande y muy grande, y reduciría el desperdicio, incrementando así el número de cajas producidas y la satisfacción del cliente.

La importancia de un buen rendimiento agrícola es crítica, dado el vínculo entre economía y agricultura. Predecir los niveles de producción en Sudamérica es particularmente desafiante debido a la sensibilidad de la agricultura a los cambios climáticos, como ondas de frío y calor, que afectan la producción y seguridad alimentaria de la región (Marengo et al., 2014). Estos cambios climáticos no solo tienen un impacto a largo plazo, sino también a corto plazo, incrementando los costos y afectando especialmente a las regiones con menores ingresos económicos.

Realizar predicciones en cultivos como la caña de azúcar, se requiere la evaluación de múltiples factores, a menudo impredecibles, como el clima. Esto obliga a los agricultores a tomar decisiones difíciles sobre la siembra, lo que puede resultar en decisiones subóptimas y grandes pérdidas económicas (Laurentin, 2020).

El desarrollo de un sistema de predicción contribuirá significativamente al avance de la agricultura de precisión en Ecuador. Este sistema proporcionará un modelo replicable en otras operaciones agrícolas, permitiendo a los agricultores integrar datos públicos y privados en un modelo predictivo robusto para tomar decisiones más informadas. Promoverá prácticas agrícolas más sostenibles y rentables, enfrentando de manera efectiva los desafíos de la variabilidad climática y la gestión eficiente de los recursos.

La implementación de este sistema no solo beneficiará a Ilaru., sino que también sentará las bases para futuras investigaciones y aplicaciones en el sector agrícola, mejorando la eficiencia y

sostenibilidad de la producción de arándanos y otros cultivos.

1.2.3 Delimitación

Problemas operativos no relacionados con la predicción: El sistema no abordará problemas operativos que no estén directamente relacionados con la predicción de la producción de arándanos.

Cambios en la infraestructura física: No se contemplarán cambios en la infraestructura física de la productora.

Factores externos impredecibles: El sistema no considerará factores externos impredecibles que puedan afectar la producción de arándanos, como desastres naturales o eventos climáticos extremos.

Problemas de gestión no relacionados con la predicción: No se abordarán problemas de gestión que no estén directamente relacionados con la predicción de la producción de arándanos.

El sistema se ajustará al mejor modelo y se realizarán pruebas hasta optimizar su predicción al mayor porcentaje posible mas no se espera tener una predicción perfecta.

1.3 JUSTIFICACIÓN

La motivación para desarrollar un sistema de predicción en la producción de arándanos surge de la necesidad de enfrentar diversos desafíos y oportunidades en la industria agrícola. Los productores de arándanos se enfrentan a variabilidad climática, fluctuaciones en la calidad del suelo, y la necesidad de optimizar el uso de recursos como agua, fertilizantes y pesticidas. Un sistema de predicción preciso permitirá a los agricultores anticipar las necesidades de las plantas, mejorar la eficiencia operativa y aumentar la rentabilidad, realizar tareas de predicción es complejo debido a la necesidad de evaluar múltiples factores, como el clima, que muchas veces

son impredecibles.

Esta complejidad dificulta la toma de decisiones agrícolas, llevando a pérdidas monetarias significativas (Laurentino, 2020). Además, la implementación de tecnologías avanzadas de aprendizaje automático y análisis de datos no solo mejorará la precisión en la predicción de rendimientos, sino que también promoverá prácticas agrícolas más sostenibles.

Este proyecto contribuirá al avance de la agricultura de precisión, proporcionando herramientas innovadoras para los agricultores y fomentando el desarrollo de soluciones tecnológicas en el sector agrícola.

En última instancia, este sistema beneficiará a los agricultores, consumidores e investigadores, mejorando la producción y calidad de los arándanos y fortaleciendo la industria agrícola en general.

Este sistema permitirá:

- Optimización de recursos: Regular la cantidad de insumos agrícolas necesarios, como agua, fertilizantes y pesticidas, para optimizar la producción y minimizar los costos.
- Mejora en la toma de decisiones: Proporcionar información precisa y oportuna para la toma de decisiones estratégicas.
- Aumento de la eficiencia y rentabilidad: Fomentar prácticas agrícolas más inteligentes y sostenibles mediante el uso de técnicas de aprendizaje automático y análisis predictivo.

- Avance en la agricultura de precisión: Contribuir al desarrollo de tecnologías avanzadas en el sector agrícola, beneficiando no solo a Ilaru., sino también sentando las bases para futuras investigaciones y aplicaciones en el país.

La implementación de tecnologías avanzadas, como el aprendizaje automático, en la agricultura no solo mejora la eficiencia y la productividad, sino que también fomenta la adopción de prácticas agrícolas más sostenibles y responsables, permitiendo a los agricultores planificar sus operaciones de manera más eficiente, regulando la cantidad de insumos agrícolas necesarios y optimizando el uso de recursos como el agua, los fertilizantes y los pesticidas. Además, el sistema contribuirá a la sostenibilidad y la rentabilidad de las prácticas agrícolas, promoviendo el uso inteligente de los recursos y reduciendo el impacto ambiental.

1.4 OBJETIVOS

1.4.1 Objetivo General

El objetivo general de esta tesis es desarrollar un sistema de predicción para la producción de arándanos.

Esto se logrará mediante la creación de un dataset que integre datos públicos y privados relevantes para la producción agrícola. El propósito de este sistema es aumentar la producción de arándanos, optimizando así la planificación y gestión de los recursos agrícolas.

1.4.2 Objetivos Específicos

- Identificación de variables clave:

Este objetivo se centra en identificar las variables con mayor correlación con la producción de arándanos. Para lograrlo, se utilizarán técnicas de análisis estadístico y minería de datos con el fin de obtener las variables más relevantes que servirán para entrenar el modelo de predicción,

mejorando así su precisión.

- Evaluación de modelos predictivos:

Se evaluaron diferentes modelos de predicción utilizando la métrica del error cuadrático medio (MSE), este análisis permitirá determinar cuál de los modelos ofrece un mayor nivel de precisión en las predicciones con el objetivo de seleccionar el modelo más efectivo para la predicción.

- Desarrollo e implementación del sistema:

Este objetivo implica desarrollar e implementar el sistema de predicción, integrando las variables relevantes y los modelos seleccionados es decir que el sistema se diseñó para realizar predicciones basadas en las variables más correlacionadas con la producción de arándanos, facilitando la toma de decisiones en la gestión agrícola.

- Validación y calibración del sistema:

Se validará y calibrará el sistema de predicción utilizando técnicas de evaluación de modelos asegurando que el sistema sea preciso y confiable, garantizando que las predicciones sean útiles para optimizar la producción de arándanos y mejorar la eficiencia en el uso de recursos agrícolas.

Estos objetivos buscan mejorar la planificación y gestión en la producción de arándanos, promoviendo prácticas agrícolas más sostenibles y rentables mediante el uso de tecnologías avanzadas y análisis de datos.

1.5 ALCANCE

El objetivo principal de este proyecto es desarrollar un sistema de predicción para medir el nivel de producción de arándanos en la empresa Ilaru. Este sistema será construido mediante la integración de datos históricos de producción, variables climáticas, condiciones geográficas y variables de cultivo.

A continuación, se detallan los componentes específicos del alcance:

Recolección y Análisis de Datos Históricos:

Integración de datos públicos y privados de producción de arándanos, incluyendo información sobre fechas de siembra y cosecha, variedades cultivadas y bloques de cultivo.

Recopilación de datos climáticos detallados, tales como temperatura, humedad, precipitación y radiación solar, utilizando fuentes como la NASA y sistemas de monitoreo locales.

Evaluación de las condiciones del suelo y su calidad, incluyendo pH, composición y estructura.

Selección y Entrenamiento de Modelos Predictivos:

Evaluación de distintos algoritmos de aprendizaje automático (Support Vector Machines, Redes Neuronales Artificiales, Random Forest) para determinar cuál se adapta mejor a la complejidad de los datos recolectados.

Uso de técnicas de análisis estadístico y minería de datos para identificar las variables más relevantes que influyen en la producción de arándanos.

Desarrollo e Implementación del Sistema:

Integración del modelo predictivo seleccionado en una plataforma informática que permita ingresar y procesar datos relevantes.

Diseño de una interfaz de usuario para facilitar la entrada de datos y la visualización de predicciones.

Implementación de mecanismos para actualizar el sistema con nuevos datos y recalibrar el modelo según sea necesario.

Validación y Calibración del Sistema:

Uso de técnicas de validación cruzada y métricas como el Error Cuadrático Medio (MSE) y el Coeficiente de Determinación (R^2) para evaluar la precisión del modelo.

Realización de pruebas piloto en campos de arándanos para comparar las predicciones del modelo con los resultados reales y ajustar los parámetros del sistema.

Optimización del Uso de Recursos Agrícolas:

Identificación de prácticas agrícolas más eficientes basadas en las predicciones del sistema.

Ajuste de los regímenes de riego, fertilización y control de plagas para optimizar la producción y minimizar el desperdicio de recursos.

Propuestas de mejoras en la planificación logística y la gestión de la comercialización de los productos.

Impacto y Portabilidad:

Al ser un proyecto de valor para las empresas agrónomas busca no solo mejorar la eficiencia y la rentabilidad de la producción de arándanos en Ilaru, sino también proporcionar un modelo replicable que pueda beneficiar a otros agricultores y contribuir al desarrollo de la agricultura de precisión en Ecuador para futuras investigaciones y aplicaciones del sistema en otros cultivos o en diferentes condiciones agrícolas.

CAPÍTULO II

MARCO TEÓRICO

1.6 ESTRUCTURA ORGANIZATIVA DE LA PRODUCTORA

La productora Ilaru sigue una estructura organizativa vertical, en la cual la gestión se organiza por procesos que permiten mantener la información de cada área bien estructurada. Este enfoque facilita una mayor flexibilidad y capacidad de adaptación a los cambios.

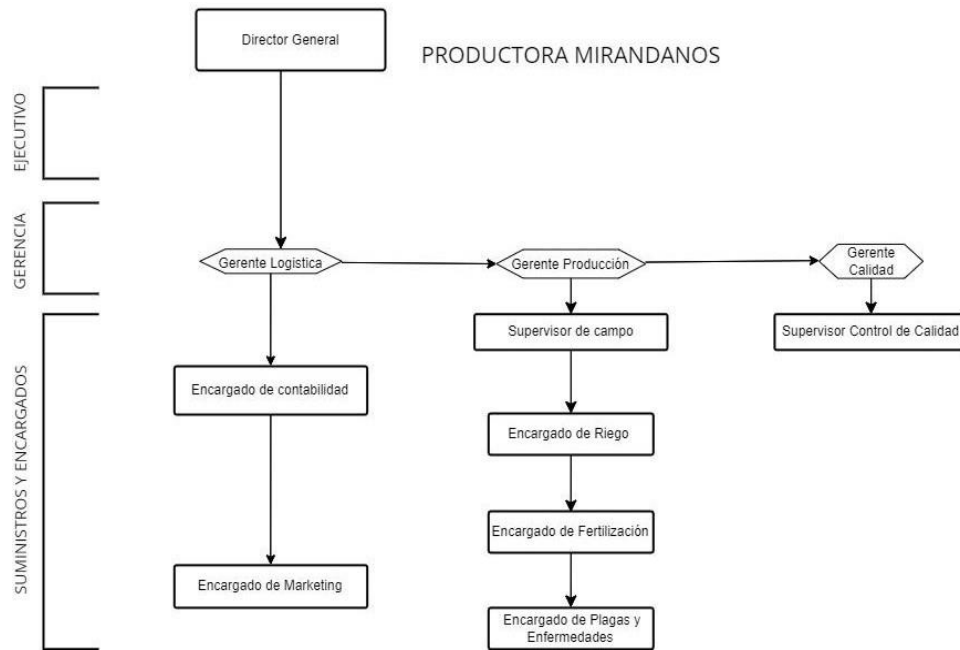
La organización vertical de la productora se estructura de abajo hacia arriba, lo que implica que los datos son recopilados y gestionados por responsables que administran diferentes procesos en tres áreas clave: logística, producción y calidad. Estas áreas son esenciales para el estudio, ya que proporcionan datos históricos precisos sobre el desarrollo de sus procesos, basados en las condiciones específicas del entorno en el momento de la recolección.

Tener la información dividida en diferentes áreas permitirá analizar las relaciones entre ellas y crear un conjunto de datos consolidado y robusto, al disponer de un mayor número de variables para entrenar que el modelo sea más robusto al momento de entrenar el modelo y así tener más datos a comparar.

A continuación, se muestra el modelo de gestión de procesos aplicado en la productora:

Figura 2

Estructura Organizacional en las áreas de la Productora



Nota: Se muestra el modelo de gestión de procesos aplicado en la productora. Elaborado por: El autor.

1.6.1 Área de Producción

El Área de Producción es fundamental para asegurar la calidad y cantidad de la cosecha de arándanos. Esta área se encarga de todas las actividades relacionadas con el cultivo y mantenimiento de las plantas, desde la siembra hasta la cosecha, así como el manejo de plagas y enfermedades. Entre las funciones principales del Área de Producción están llevar a cabo todos los procesos que manejan las diferentes subáreas que la conforman.

Estas tareas incluyen la supervisión de la siembra y el manejo continuo del cultivo, la

administración adecuada de riego y fertilizantes, la planificación y ejecución de la cosecha en el momento óptimo, y el monitoreo y tratamiento para prevenir y controlar plagas y enfermedades. Además, se destaca la precisión en el registro de datos, el monitoreo continuo del cultivo, la gestión de calidad y cantidad de la producción, y el registro de variables climáticas.

Figura 3

Riego y Cuidado



Nota: La Imagen muestra al supervisor de campo haciendo una revisión general del estado de las plantas y del cultivo en general. Elaborado por: El autor.

1.6.2 Área de Logística

El Área de Logística se encarga de gestionar todo el proceso postcosecha, asegurando que los arándanos se almacenen y transporten de manera eficiente y en condiciones óptimas para mantener su frescura y calidad hasta que lleguen al consumidor final.

Sus funciones principales incluyen la planificación de rutas y coordinación del transporte, la gestión de inventarios y el control de condiciones de almacenamiento, y el registro detallado de

gastos e ingresos, así como la elaboración de presupuestos.

Las características clave del Área de Logística son la gestión integral de la cadena de suministro, la eficiencia en el transporte y distribución, y el control estricto de calidad durante el almacenamiento.

Figura 4

Transporte del Producto



Nota: La Imagen muestra el transporte del producto final para su posterior distribución.
Elaborado por: El autor.

Área de Calidad

Esta área realiza inspecciones y controles de calidad en todas las etapas de producción y distribución para garantizar que el producto final sea de la más alta calidad. Las funciones principales del Área de Calidad incluyen la verificación de la calidad mediante controles de calidad, la evaluación del producto final a través de inspecciones y registros de la calidad de los productos.

Los datos manejados por esta área incluyen los resultados de inspecciones, las normas de

calidad aplicadas de acuerdo con las normas de calidad que maneja la productora sobre el producto. Sus características incluyen un control estricto de calidad, inspección detallada en cada etapa, registro de cumplimiento de normas de calidad, gestión de datos para mejora continua con un registro detallado del registro de los frutos que hayan pasado por este proceso.

Figura 5

Control de Calidad.



Nota: En la empresa Ilaru, el 75% de sus trabajadores en campo son mujeres porque son más delicadas y eficientes con el manejo de la fruta. Fuente: Revista Líderes (2023)

1.7 VARIEDADES DE ARANDANOS

Biloxi, Emerald, Legacy y Star son las cuatro variedades de arándanos que maneja la productora las cuales presentan características distintivas que influyen en su adaptabilidad a diferentes condiciones de cultivo y necesidades de los productores.

Biloxi se destaca por su resistencia y bajo consumo de agua, siendo ideal para suelos esponjosos y sustrato de coco.

Emerald, conocida por sus frutos grandes y ácidos, requiere un manejo cuidadoso del riego

para evitar el estrés hídrico.

Legacy produce frutos grandes y menos ácidos, adecuados para la exportación, y demanda riegos prolongados debido a su alto nivel de hidratación.

Finalmente, Star, con el tiempo de maduración más largo entre las variedades, ofrece frutos dulces y necesita menos agua, siendo adecuada para mercados especializados que valoran sus características de sabor y calidad.

Cada una de estas variedades ofrece ventajas y desafíos específicos, permitiendo a los productores seleccionar la más adecuada según sus objetivos y condiciones de cultivo.

1.7.1 Ciclo de cultivo

El análisis de las variables a seleccionar durante el proceso de ciclo de cultivo van depender mucho no solo de condiciones o estándares establecidos para un producto determinado, sino que esto dependerá mucho de la variedad de la planta de arándano ya que cada variedad de las cuatro que se manejan en la productora tiene ciclos de siembra diferentes con tiempos que difieren de acuerdo a su mes de siembra y cosecha por lo que ciertos métodos

serán variables dependiendo en qué estado se encuentre la planta ya que como se mencionó anteriormente van a necesitar un tratado diferente desde el tiempo de riego el tipo de suelo y la poda estos campos principalmente son los que influyen de manera directa a la maduración de la planta por lo que fue importante haber realizado la exploración de campo previa donde se pudo obtener datos certeros de la mano del encargado de producción donde supo explicar las pequeñas variantes que manejan de manera general para cada una de las hileras dependiendo del ciclo en el cual se encuentre la planta, entre las principales etapas del ciclo incluyen la plantación, crecimiento vegetativo, floración, fructificación y cosecha.

Tabla1

Numero de Plantas por Variedad

VARIEDAD	NUMERO DE PLANTAS
BILOXI	11.090
EMERALD	63.458
LEGACY	14.424
START	27.046

Nota: Presentación de los data set hallados en la productora con sus características. Elaborado por: El autor.

1.7.2 Fases del ciclo de cultivo

A continuación, se describen las fases que se dan durante el ciclo de cultivo, es importante trataresto ya que de aquí surgen los datos de las variables recolectadas a lo largo de las diferentes fases del ciclo de cultivo que son esenciales para el desarrollo de un sistema en la exploración de campo que fue de la guía de Jaime Maldonado encargado de controlar las condiciones

fitosanitarias, se obtuvo conocimiento acerca de las prácticas agrícolas que se dan para cada fase.

1.7.2.1 Plantación

La plantación es la etapa inicial y puede realizarse en diferentes épocas del año debido a las buenas condiciones, dependiendo del clima y la variedad del arándano.

Es crucial seleccionar el sitio adecuado con suelos bien drenados por lo que se ha adoptado una forma de cultivo de hileras es decir se cultivan en masetas siguiendo un orden de filas.

1.7.2.2 Establecimiento y Crecimiento Vegetativo

Durante los primeros años después de la plantación, las plantas de arándano se centran en establecer un sistema de raíces robusto y un crecimiento vegetativo saludable. En esta fase, es importante:

Riego regular: Mantener el suelo consistentemente húmedo, especialmente durante los primeros dos años y con un pH entre 4.5 y 5.5, ideal para el crecimiento de los arándanos. La preparación del suelo incluye la adición de materia orgánica y la corrección de su pH.

Control de malezas: Para evitar la competencia por nutrientes y agua.

Fertilización Adecuada: Proveen nutrientes esenciales en pequeñas dosis frecuentes.

Floración: La floración generalmente ocurre en primavera, cuando las temperaturas empiezan a subir. Las flores de arándano son pequeñas y blancas o rosadas, y su polinización es fundamental para una buena cosecha.

1.7.2.3 Maduración

Los arándanos maduran en verano, dependiendo de la variedad y las condiciones climáticas, la maduración se caracteriza por el cambio de color de los frutos, generalmente de verde a azul oscuro diferentes variedades tienen diferentes tiempos de maduración por lo que se sumó un total de días para el dataset usado donde se saca el intervalo de días que hay entre la fecha de siembra y fecha de cosecha ya que algunas variedades pueden madurar más rápidamente, mientras que otras requieren más tiempo para alcanzar su máximo potencial. En esta etapa, es crucial.

1.7.2.4 Cosecha

La cosecha se realiza generalmente a mano para asegurar la calidad de los arándanos, aunque también se pueden usar máquinas en plantaciones grandes. La cosecha se lleva a cabo en varias rondas, ya que los frutos no maduran todos al mismo tiempo. Es vital:

Recolección Cuidadosa: Para minimizar el daño a los frutos.

Manejo Postcosecha: Mantener los frutos en condiciones óptimas de almacenamiento para prolongar su vida útil.

1.7.2.5 Poda y Mantenimiento Postcosecha

Después de la cosecha, las plantas de arándano requieren poda para eliminar ramas viejas, enfermas o dañadas. La poda ayuda a mejorar la aireación y la penetración de la luz, lo que favorece el crecimiento de nuevos brotes y la productividad en la próxima temporada.

1.7.3 Variedades de arándanos y sus especificidades

1.7.3.1 Biloxi

Tiene un tiempo de maduración relativamente largo con promedio de 1153 días, esta variedad puede ser adecuada para productores que buscan una variedad resistente y duradera, sin embargo, la espera más prolongada podría ser una desventaja para aquellos que buscan una producción más rápida, son plantas jóvenes y vegetativas que requieren riegos más cortos debido a su tamaño y desarrollo. Ideal para suelos esponjosos y sustrato de coco por su baja consumo.

Figura 6

Planta Biloxi



Nota: El número de plantas Biloxi borda los 11.090 siendo la variedad con menor número de plantas sembradas. Elaborado por: El autor.

1.7.3.2 *Emerald*

Presenta un tiempo de maduración similar al de BILOXI, pero ligeramente más corto de 1139.12 días, al tener un tiempo de maduración más corto, EMERALD podría ser preferible para productores que buscan reducir el tiempo hasta la cosecha sin sacrificar mucho en términos de durabilidad y resistencia por lo que requiere más agua y produce frutos grandes y ácidos. Necesita un manejo cuidadoso del riego para evitar el estrés hídrico su regado es de 960cm³ por día.

Figura 7

Planta Emerald



Nota: El número de plantas Emerald borda los 63.458 siendo la variedad con mayor número de plantas sembradas. Elaborado por: El autor.

1.7.3.3 Legacy

Esta variedad tiene un tiempo de maduración comparable a BILOXI y EMERALD, pero un poco más largo de 1163.32 días, puede ser ideal para productores interesados en una variedad tradicionalmente confiable y con características bien establecidas, aunque el tiempo de maduración más largo produce frutos grandes y menos ácidos, adecuados para exportación.

Requiere riegos más prolongados al igual que Emerald de 960cm^3 por día.

Figura8

Planta Legacy



Nota: El número de plantas sembradas de Legacy es de ligeramente mayor a Biloxi con un total 14.424, su alto nivel de hidratación debido a su tallo leñoso. Elaborado por: El autor.

1.7.3.4 *Star*

Tiene el tiempo de maduración más largo entre las variedades analizadas de 1272.69 días, STAR es la variedad con el tiempo de maduración más extenso, lo que puede hacerla menos atractiva para aquellos que necesitan una producción más rápida.

Sin embargo, este tiempo adicional podría traducirse en características de sabor y calidad específicas que podrían ser apreciadas en mercados especializados, necesita menos agua su riego es de 720cm³ por día y produce frutos dulces. Requiere un manejo específico del riego y el suelo para asegurar su calidad es la más nueva en plantarse son plantas muy jóvenes.

Figura 9

Planta Star



Nota: Star es la segunda mayor en número de plantas sembradas con 27. 046. Elaborado por: El autor.

1.7.4 Técnicas de cultivo

Para poder adquirir conocimiento acerca de las técnicas que implementan para las 4 variedades de plantas que maneja la productora no solo bastaba con recopilar la información dada por la base de datos otorgada acerca de los planos de riego sino que también era necesario adentrarse de primera mano junto al supervisor de campo para así corroborar que los datos dados coincidan con las técnicas de cultivo de arándanos que incluyen prácticas como la poda, la polinización asistida, el tipo de tierra y abono usado, el uso de coberturas protectoras. A continuación, detallare las técnicas de cultivo que maneja cada variedad de planta:

Emerald: Esta variedad de arándanos son plantas de copa alta por lo que necesitan abundante riego y manejo cuidadoso para asegurar frutos grandes y ácidos ya que esta variedad posee tallos más gruesos también son más frondosos el riego de estas plantas es de 2 minutos debido al tipo de desarrollo que maneja esta variedad y la producción final son frutos grandes y de sabor un poco ácido.

Biloxi: Plantas jóvenes requieren menos riego, pero un manejo más delicado debido a su tamaño no tan alto más vale medianas y de desarrollo vegetativo esta producción es medianamente nueva por lo que la mayoría de plantas son pequeñas es decir vegetativas a estarse le da una técnica de 1 rama de pulmón es decir que esto de ramas se la deja en producción el suelo que tiene es de sustrato de coco ese tipo de suelo se caracteriza por ser esponjoso es decir que absorbe bastante agua estas manejan 1 minuto de riego al ser más pequeñas.

Legacy: Plantas altas producen arándanos en la copa maneja un riego de 2 minutos es de tallo leñoso fruto grande son menos ácidas que las Emerald y no son tan dulces son resistentes y adecuado para exportación, requiere un manejo específico para mantener la calidad del fruto.

Star: Plantas pequeñas, medias, altas, frondosas con frutos que poseen una textura suave y son bien dulces manejan un riego de 1 minuto y medio las más pequeñas necesitan 25% menos de agua, pero una gestión precisa para asegurar la dulzura y tamaño de los frutos el suelo que usan también es de tipo sustrato de coco.

1.8 FACTORES QUE AFECTAN EN LA PRODUCCIÓN

1.8.1 Factores ambientales

En Ecuador, una de las ventajas es la capacidad de producir arándanos durante todo el año debido a la estabilidad climática, que evita variaciones severas. La excelente calidad del agua, una de las mejores de Latinoamérica, y la radiación solar constante, gracias a la ubicación en la mitad del mundo, contribuyen a que el sabor sea más dulce y que las plantas crezcan más rápidamente (Revista Líderes, 2023).

Los arándanos obtenidos son de buen tamaño y presentan características deseadas como firmeza y dulzura, lo que los distingue de los producidos en otros países, la calidad del suelo en Ecuador ha permitido la producción exitosa de hasta 12 variedades de arándanos, algunas ya cultivadas y otras que se cosecharán este año (Revista Líderes, 2023).

Figura 10

Bloque 1 de Producción



Nota: Se aprecia el bloque 1 que tiene una carpa que protege a las plantas de cualquier factor o amenaza externa que interfiera en su desarrollo. Elaborado por: El autor.

1.8.2 Clima

La temperatura, la humedad, la precipitación y la luz solar son cruciales para el desarrollo de los arándanos. Las temperaturas extremas pueden afectar la floración y la fructificación por lo que en cada bloque se han levantado carpas que protegen de posibles cambios drásticos, los arándanos requieren climas templados con inviernos fríos para romper la dormancia y promover un buen crecimiento en primavera.

1.8.3 Suelo

Los arándanos prefieren suelos ácidos (pH 4.5-5.5), bien drenados y ricos en materia orgánica. La textura del suelo también es importante; suelos arenosos o con buen drenaje como son ideales para evitar problemas de encharcamiento es decir que no se estanque el agua en la maceta.

Figura 11

Suelo Sustrato de Coco



Nota: Fibra de coco es un tipo de suelo usado para todos los cultivos de la productora.
Elaborado por: El autor.

1.8.4 Riego

La cantidad y la calidad del agua de riego afectan directamente el crecimiento y la producción de los arándanos en especial el tipo de regado que se da en esta productora se hacen base a la mezcla de soluciones para el control de plagas y pesticidas esto lleva una medida de mezcla exacta por litro de agua para cada bloque esto se parametriza en el sistema de regado donde se mezcla en los 4 tanques de mezcla estas medidas van a variar de acuerdo a cada variedad por lo que él envió de estas soluciones se lo asigna en el sistema para las 63 válvulas que conectan los 17 bloques donde se está dando la producción actual estas medidas de riego sino se las sabe tratar pueden hacer que se reduzca significativamente el rendimiento, mientras que el exceso de agua puede provocar enfermedades radiculares

Figura 12

Tanques de Mezcla



Nota: Se muestra los 4 tanques de mezcla los cuales combinan las soluciones químicas de fertilizantes y control de plagas. Elaborado por: El autor.

1.8.5 Factores biológicos

Cada variedad tiene características específicas en cuanto a requerimientos de agua, nutrientes y resistencia a enfermedades. Por ejemplo, la variedad Emerald requiere más agua y produce frutos más grandes y ácidos, mientras que la variedad Star necesita menos agua y produce frutos más dulces.

1.8.6 Polinización

La polinización adecuada es esencial para una buena fructificación. Los arándanos dependen de polinizadores como las abejas para transferir el polen entre las flores.

1.8.7 Plagas y Enfermedades

Las plagas como los pulgones y las enfermedades fúngicas pueden reducir el rendimiento.

Un manejo integrado de plagas (MIP) y el uso de variedades resistentes son estrategias clave para minimizar estas amenazas.

1.8.8 Factores Agronómicos

El suministro adecuado de nutrientes es vital para el desarrollo de la planta y la producción de frutos. El nitrógeno, el fósforo y el potasio son esenciales, pero también son importantes los micronutrientes como el hierro y el magnesio.

1.8.9 Poda

La poda regular ayuda a mantener la planta en una forma óptima para la producción de frutos. La eliminación de ramas viejas, enfermas o dañadas mejora la aireación y la penetración de la luz, lo que favorece un mejor crecimiento y rendimiento.

1.8.10 Manejo del Suelo

La estructura y la fertilidad del suelo deben mantenerse mediante prácticas como la rotación de cultivos, la incorporación de materia orgánica y el control de malezas.

1.8.11 Factores de Manejo Postcosecha

La técnica y el momento de la cosecha afectan la calidad y el rendimiento final. La cosecha manual es preferible para evitar daños en los frutos, aunque es más laboriosa y costosa.

La conservación de los arándanos en condiciones adecuadas de temperatura y humedades crucial para mantener su calidad y prolongar su vida útil. Los arándanos son sensibles a la deshidratación y a la pudrición, por lo que un manejo adecuado durante el almacenamiento es esencial.

1.8.12 Transporte

La logística de transporte debe garantizar que los frutos lleguen en buenas condiciones al

mercado. Los arándanos deben ser transportados en contenedores adecuados que minimicen los daños físicos y mantengan la frescura.

1.9 TECNOLOGÍA Y AUTOMATIZACIÓN

1.9.1 Sistemas de Riego Automatizados

La implementación de sistemas de riego por goteo controlados automáticamente puede optimizar el uso del agua y mejorar el rendimiento.

1.9.2 Sensores y Tecnología

El uso de sensores para monitorear las condiciones del suelo y del clima en tiempo real permite ajustes inmediatos en las prácticas de manejo, mejorando la eficiencia y el rendimiento.

1.10 SISTEMAS DE PREDICCIÓN

La utilización de modelos predictivos basados en inteligencia artificial puede ayudar a anticipar las necesidades de las plantas y ajustar las prácticas agrícolas en consecuencia.

El rendimiento de los arándanos depende de una combinación de factores ambientales, biológicos, agronómicos y de manejo postcosecha. La integración de tecnologías avanzadas y prácticas agrícolas sostenibles puede optimizar estos factores y mejorar significativamente la producción. Un sistema de predicción robusto, que incorpore datos de clima, suelo y manejo agrícola, puede ser una herramienta invaluable para maximizar el rendimiento y la rentabilidad de los cultivos de arándanos.

1.10.1 Aprendizaje automático

El aprendizaje automático puede ser categorizado principalmente en dos métodos:

supervisado y no supervisado. (Vidhate y Kulkarni, 2012). El aprendizaje supervisado opera con conjuntos de datos que incluyen entradas y salidas predefinidas, facilitando la modelación de relaciones entre variables. Por el contrario, el aprendizaje no supervisado se centra en agrupar conjuntos de datos basándose en las similitudes entre las muestras sin etiquetas previas, identificando estructuras ocultas dentro de los datos. Estos métodos son paralelos a las técnicas utilizadas en minería de datos, diferenciándose principalmente en la terminología empleada; las técnicas predictivas y descriptivas corresponden al aprendizaje supervisado y no supervisado, respectivamente.

El propósito del aprendizaje automático es descubrir y adaptarse continuamente a cambios en el entorno, mejorando la capacidad de una máquina para analizar datos y resolver problemas. (Calvo Valverde, 2016). Este campo utiliza métodos que detectan patrones automáticamente en los datos, que luego se usan para hacer predicciones o tomar decisiones bajo incertidumbre. Esta capacidad de automejora y adaptación es fundamental para el progreso en áreas como la toma de decisiones autónoma y el análisis predictivo, donde la precisión y la capacidad de adaptación son cruciales.

Estudios anteriores, como el de García-Arteaga et al. (2020), han demostrado que el aprendizaje automático puede predecir con precisión el rendimiento de los cultivos. Estas investigaciones han utilizado diversas técnicas, incluyendo redes neuronales, máquinas de soporte vectorial y bosques aleatorios, para modelar y predecir variables agrícolas críticas. La investigación de Crane-Droesch (2018) destaca la importancia de considerar el impacto del cambio climático en la predicción del rendimiento, subrayando la necesidad de modelos adaptables a condiciones variables.

1.10.2 Modelos de predicción

Los modelos predictivos son herramientas esenciales para anticipar el rendimiento de los cultivos y optimizar la gestión agrícola. Estos modelos utilizan datos históricos y variables actuales para prever resultados futuros. Los principales modelos predictivos en la agricultura incluyen.

1.10.3 Máquinas de Soporte Vectorial (SVM)

Eficientes para manejar grandes conjuntos de datos y capturar patrones complejos.

1.10.4 Redes Neuronales Artificiales (ANN)

Capaces de aprender y adaptarse a diferentes condiciones y factores que afectan el cultivo.

1.10.5 Bosques Aleatorios (RF)

Útiles para manejar datos con alta dimensionalidad y variabilidad.

1.11 RECOPIACIÓN DE DATOS

Para la recopilación de datos del sistema de predicción fue necesario hacer un análisis previo de las tres áreas que maneja la productora por lo que se realizó una exploración de campo donde se pudo presenciar que las zonas de cultivo llamadas que son un total son 16 manejan características diferentes unos a otros es decir no manejan una sola especie por bloque por lo que pueden ser varias especies de los 4 tipos contenidas en un solo bloque o contrario a continuación se muestra un croquis de los bloques de las áreas de cultivo.

Visitando una por una se ha corroborado la información obtenida del dataset de plano de siembra que se utilizó en el conjunto de datos final implementado en el sistema el cual detalla las principales propiedades de la planta al inicio de su siembra como son fecha de siembra , variedad , bloque y numero de plantas estas variables son inversamente proporcionales a la cantidad a

predecir por lo que fue de vital importancia recolectarlas ahí salte de igual manera su posterior plan de cosecha el cual se concentraba en un software propio de la empresa llamado Aazuli el cual tiene como principal función registrar el total de producción que se ha tenido por bloque, especie y gramaje o calidad este dataset es muy importante para el diseño del sistema ya que al tener la variable dependiente producción podremos entrenar al modelo con valores reales que ya hayan sucedido y más que todo ligados a la realidad de las características del entorno donde se desarrollan los datos de las variables.

A continuación, se muestra una tabla donde se describe los datos climáticos, de suelo y de manejo agrícola, obtenidos de fuentes como la NASA:

Tabla 2

Dataset y Características

Data Set	Fuente	Características	Variables
	Aazurisoftware	Detalles sobre fechas y bloques de cosecha, variedades, y el calibre/gramaje de los arándanos cosechados.	Fecha de cosecha, Variedad, Bloque, Gramaje/Calidad, Cantidad
Producción	Aazurisoftware	Información detallada sobre la plantación de arándanos, incluyendo las áreas, las naves, las hileras, y las fechas de siembra.	Bloque, Área, Naves, Hileras, Plantas, Variedad, Proveedor, Fecha de siembra

Control Plagas y Enfermedades	Supervisor de Campo	Datos no proporcionados	Datos no proporcionados
Económico	Encargado de Contabilidad	Datos mensuales de ingresos durante el año 2022, incluyendo categorías detalladas de ingresos.	Código, Nombre, Meses (Enero, Febrero, Marzo, etc.), Total 2022, Porc. 2022
Climáticos	NASA Prediction Of Worldwide Energy Resources (POWER)	Contiene registros diarios de diferentes parámetros climáticos, como temperaturas, humedad, precipitación y velocidad del viento.	Fecha, Temp 2m, Temp Dew 2m, Temp Humeda 2m, Temp Superficie, Humedad Específica 2m, Humedad Relativa 2m, Precipitación Corregida, Presión.

Nota: Presentación de los data set hallados en la productora con sus características. Elaborado por: El autor.

2.6.1 Recolección de datos

Para la recopilación de datos primero se tuvo que hacer un análisis previo de las áreas que maneja la productora por lo que se realizó una exploración de campo donde se presenciaron las áreas de cultivo que en la productora se las conoce como bloques que son un total de 16 cada uno con diferentes dimensiones de área, estos bloques manejan características diferentes unos a otros es decir no manejan una sola especie por bloque por lo que pueden ser varias especies de los 4 tipos

contenidas en un solo bloque o contrario a continuación se muestra un croquis de los bloques de las áreas de cultivo.

Visitando una por una se ha corroborado la información obtenida del dataset de plano de siembra que se utilizó para el conjunto de datos final el cual detalla las principales propiedades de la planta al inicio de su siembra como son fecha de siembra, variedad, bloque y número de plantas. Estas variables son inversamente proporcionales a la cantidad a predecir por lo que fue de vital importancia de igual manera su posterior plan de cosecha el cual se concentraba en un

software propio de la empresa llamado Aazuli el cual tiene como principal función registrar el total de producción que se ha tenido por bloque, especie y gramaje o calidad. Estos datos se obtienen al final de cada cosecha.

Este dataset es muy importante para el diseño del sistema ya que al tener la variable dependiente a predecir que es la producción se pudo entrenar al sistema con valores reales que han sucedido a lo largo de línea de tiempo y más que todo ligados a la realidad de los datos que manejan las distintas variables que son condicionadas por cada uno de los bloques y los tipos de cuidado que manejan.

1.11.1 Datos Locales

La tecnología que se maneja en la productora a base de sensores tienen funcionalidades avanzadas que se rigen por softwares los cuales manejan su propia fuente de información. Estas fueron las principales fuentes que nos otorgaron los datos que van a ser clave al momento del levantamiento de los datos históricos e incluso ver que datos se están manejando en tiempo real. De esta forma se puede corroborar que los datos históricos sean de acuerdo a las medidas usadas en el diario cuidado de las plantas.

En la siguiente tabla se muestra los datos de acuerdo con la tecnología usada en cada área de la productora:

Tabla3

Tecnología y Características

Área	Tecnología	Datos
Producción	Aaazuli Software	Recoge todos los datos de las características del producto final acorde a su fecha de cosecha
Control	Sistema de Mezcla	En esta área se obtiene las variables con los datos de los niveles de mezcla que se dan por bloques.
Economía	Servidores	Manejan dataset donde detallan la fecha de siembra y también los gastos dados durante el proceso de inicio y postcosecha

Nota: Muestra la tecnología usada en cada área y sus características. Elaborado por: El autor.

Monitoreo de los tanques de mezcla: a monitorear el nivel de químicos que se van a ingresar en la mezcla de los 4 tanques de preparación de fertilizantes químicos y los químicos usados para el control de plagas.

La agricultura de precisión es una práctica que utiliza tecnología avanzada para gestionar los cultivos de manera más eficiente. Esto incluye el uso de sensores, sistemas de información geográfica (SIG) y herramientas de análisis de datos para monitorear y optimizar las prácticas agrícolas. (Pierce y Nowak, 1999).

Figura 13

Sistema de Mezcla



Nota: Este Software de mezcla puede asignar el porcentaje de químicos a usar por mezcla y también por que tubería enviar dicha solución. Elaborado por: El autor.

1.11.2 Datos Públicos

2.6.1.1.1 Sistema de información geográfica

Los SIG son herramientas poderosas para gestionar y analizar datos espaciales en la agricultura. Permiten:

La herramienta Informes que usamos para obtener datos de las temperaturas proporciona informes basados en la ubicación de un único punto para un período de tiempo especificado por el usuario desde la interfaz de programación de aplicaciones (API) dePOWER. Los informes disponibles incluyen el informe de indicadores climáticos y la captación de otras variables relacionadas con la agro climatología.

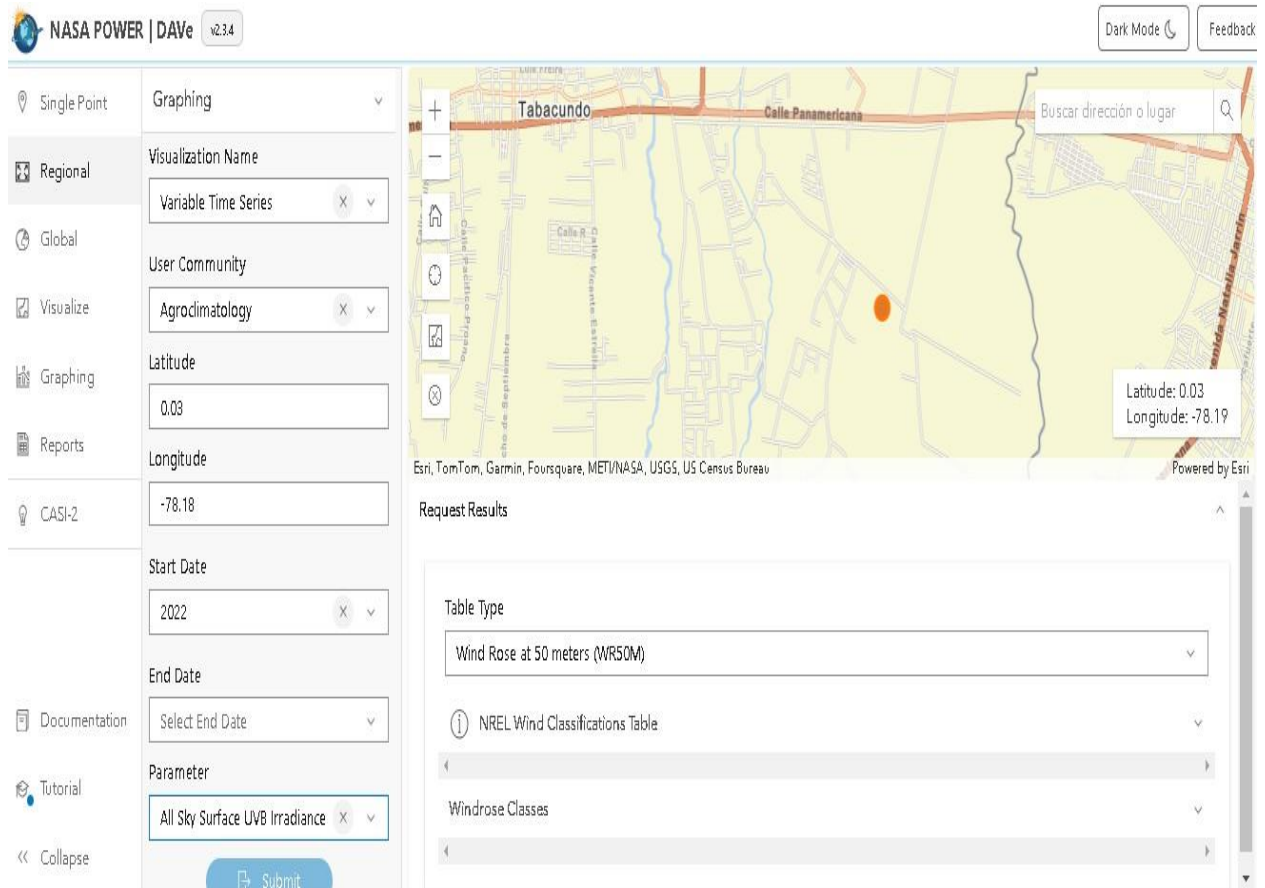
1.11.3 Monitoreó de Datos

Atraves de la coordenadas otorgadas por imágenes satelitales de Google Maps ubicamos los distintos bloques dentro del campo de producción donde se obtuvo las coordenadas 0.0341921866982889, -78.1829210596979 que nos van a ser esenciales al momento del obtener los datos históricos que se han monitoreados a través de los satélites de la herramienta que utilizaremos de ayuda como es el sistema de NASA POWER (Prediction Of Worldwide Energy Resources) este sistema utiliza una combinación de datos satelitales y modelos climáticos para proporcionar información climática precisa y detallada.

Este sistema es versátil al momento de elegir que permito obtener datos agro climatológico es decir variables que afectan que afectan directamente al crecimiento y proliferación de las plantas como son los llamados rayos UVB estos son los que hacen que la fotosíntesis en la planta.

Figura14

Sistema de Agroclimatología



Nota: Sistema información Agroclimática Elaborado por: El autor.

NASA POWER utiliza datos de satélites y modelos numéricos para recopilar información sobre variables climáticas y variables acerca de la agroclimatología esto quiere decir que este sistema no solo da una vista general de los fenómenos climáticos, sino que nos otorga las variables específicas que influyen en el desarrollo agrónomo a través de mapeo y análisis de la variabilidad del suelo en diferentes áreas del campo.

Estos datos, disponibles desde 2017 a través del Data Access Viewer de la NASA, son

fundamentales para desarrollar modelos predictivos precisos que permitan a Ilaru. optimizar su producción de arándanos y mejorar la eficiencia en el uso de recursos agrícolas.

A continuación, se presenta una tabla con la descripción y la importancia de cada variable en el conjunto de datos agroclimáticos de acuerdo con las obtenidas en el monitoreo de datos:

Tabla4

Variables y Características

Variable	Descripción	Importancia
Temp 2m	Temperatura a 2 metros de altura	Afecta la fotosíntesis, respiración y otros procesos fisiológicos de las plantas
Temp Dew 2m	Temperatura de rocío a 2 metros de altura	Influencia en la humedad del aire y el riesgo de enfermedades fúngicas
Temp Humeda 2m	Temperatura húmeda a 2 metros de altura	Refleja la cantidad de humedad en el aire, afectando la transpiración de las plantas
Temp Superficie	Temperatura de la superficie	Influencia en la germinación de semillas y desarrollo de raíces
Temp Mínima 2m	Temperatura mínima a 2 metros de altura	Influencia en el riesgo de heladas que pueden dañar las plantas
Temp Máxima 2m	Temperatura máxima a 2 metros de altura	Posibilidad de causar estrés térmico en las plantas
Rango Temp 2m	Rango de temperatura a 2 metros de altura	Información sobre la variabilidad térmica diaria

Radiación Solar Superficie	Radiación solar en la superficie	Fundamental para la fotosíntesis, afectando el crecimiento y productividad de las plantas
Evapotranspiración Superficie	Evapotranspiración en la superficie	Mide la pérdida de agua del suelo y las plantas, crucial para la gestión del riego
Evapotranspiración 2m	Evapotranspiración a 2 metros de altura	Similar a la superficie, pero considerando la altura, relevante para el microclima
Humedad Relativa 2m	Humedad relativa a 2 metros de altura	Afecta la transpiración de las plantas y el riesgo de enfermedades
Precipitación Corregida	Precipitación corregida	Proporciona agua necesaria para las plantas, importante para el suministro de agua disponible
Presión Superficie	Presión atmosférica en la superficie	Influencia en las condiciones climáticas que afectan el crecimiento de las plantas
Velocidad Viento 2m	Velocidad del viento a 2 metros de altura	Afecta la polinización, la dispersión de enfermedades y la transpiración de las plantas
Velocidad Viento Máxima 2m	Velocidad máxima del viento a 2 metros	Las ráfagas de viento pueden causar daños físicos a las plantas y afectar la polinización
Velocidad Viento Mínima 2m	Velocidad mínima del viento a 2 metros	Las bajas velocidades del viento pueden afectar la ventilación natural
Rango Velocidad Viento 2m	Rango de velocidad del viento a 2 metros	Proporciona una imagen más completa de las condiciones del viento que afectan a las plantas

Dirección Viento 2m	Dirección del viento a 2 metros de altura	Influencia en la dispersión de enfermedades y distribución de humedad y temperatura
Humedad Suelo Superficie	Humedad del suelo en la superficie	Crucial para el establecimiento de las plantas y la disponibilidad de agua para la absorción de nutrientes
Humedad Suelo Raíces	Humedad del suelo en la zona de raíces	Fundamental para el desarrollo radicular y la absorción de agua y nutrientes.

Nota: Muestra la tecnología usada en cada área y sus características. Elaborado por: El autor.

CAPÍTULO III

METODOLOGÍA

1.12 DISEÑO DE LA INVESTIGACIÓN

1.12.1 Investigación preliminar

La investigación es de tipo descriptiva, explicativa y experimental. Se enfoca en identificar y analizar las variables que afectan la producción de arándanos. La investigación descriptiva permitirá documentar las prácticas actuales de cultivo y manejo de arándanos. La investigación explicativa se centrará en entender cómo estas variables influyen en el rendimiento de los cultivos. Finalmente, la investigación experimental se implementará para probar y validar el sistema de predicción desarrollado.

Para adentrarse más a esta obtención y corroboración del estudio se tuvo varias entrevistas con diferentes miembros de las áreas que se manejan en la productora principalmente era necesario el tipo de afectaciones que pueden tener la producción dentro del área donde se cultiva al ser una producción que se maneja en hileras es decir manejan filas de masetas dentro de bloque que son protegidas con carpas gigantes librando de afectaciones graves externas es decir fuertes vientos o plagas que no

Para investigar las variables que afectan la producción de arándanos, se llevará a cabo una investigación de campo exhaustiva. Se recopilará información detallada sobre el manejo y tratamiento de las plantas de arándano desde su fase de siembra, maduración y cosecha. Esta información se obtendrá a través de herramientas como cuestionarios, entrevistas y observaciones directas. El objetivo es identificar las principales variables que pueden influir en

el rendimiento y aumentar la producción de arándanos.

1.12.2 Recolección de datos

Para la recolección de datos se obtuvo los datos de todas las áreas que maneja procesos de recopilación de datos que tienen que ver con los procesos y tratamientos que manejan en la siembra, en el intervalo de crecimiento de la planta y en la cosecha final estos procesos que lleva la planta son los que arrojan los datos esenciales para el conjunto de datos recolectados para su posterior evaluación al tener un gran conjunto de datos que no solo afectan de manera directa con la producción sino que hay datos que nos daremos cuenta que aunque pareciera que no tienen un grado de afectación va a tener una influencia indirecta con otras variables que si lo son por lo que es necesario agregarlas al momento de la recolección de datos.

Para recolectar los datos necesarios se extrajeron de bases de datos propias de sistemas usados en la productora:

- Sensores y dispositivos de tecnología: Utilizados para monitorear en tiempo real las condiciones climáticas y del suelo.
- Activos de información físicos: Incluyen registros y documentos mantenidos por los productores.
- Activos de información digitales: Bases de datos y sistemas de gestión de información agrícola.
- Bases de datos y SIG: Sistemas de Información Geográfica (SIG) para analizar la variabilidad espacial de los datos recolectados.

En total se seleccionaron un total de 22 variables del conjunto de datos de las cuales se trató a la variable cantidad como la variable dependiente ya que esta nos da la información de la producción de la cantidad de arándanos en kg dada su previa cosecha y las otras 21 como variables independientes la variable dependiente se usa para la predicción del rendimiento de cultivos.

1.12.3 Selección de Datos

Para la recopilación y obtención de datos que nos ayuden a crear un sistema predicción eficaz, robusto, es decir un big data donde se hallen todos los conjuntos de datos que se analizaron de acuerdo al grado de correlación que exista entre estas variables.

En base a estudios anteriores se ha dado retroalimentación a lo dicho en la recolección de datos que algunas variables no influyen de manera directa en la producción así mismo se demostró que los factores genéticos no afectan tanto como los climáticos, llegando a la conclusión de que dos de los factores que más afectan la producción de cultivos son la temperatura y la precipitación que se ha considerado esto en el presente estudio (Khaki,2019).

En este caso los factores genéticos se muestran en las variedades es interesante este enfoque aplicado en el presente sistema, ya que así se tendrá una visión diferente de esta variable para este caso ya que al manejar 4 variedades va a ser necesario ver que variables también se correlacionan por bloque y este a su vez ligado a su variedad mostrara esas pequeñas variantes que afectan a las plantas que no producen por bloque independiente de la especie.

A continuación, se muestra una tabla de los datos implementados en el entrenamiento del sistema:

Tabla5

Datos Usados en el entrenamiento del sistema:

Variable	Tipo de dato	Descripción
Fecha de siembra	Catagórico	Fecha en que se sembró el cultivo
Fecha de cosecha	Catagórico	Fecha en que se cosechó el cultivo
Variedad	Catagórico	Variedad del cultivo
Bloque	Catagórico	Bloque o área del cultivo
Gramaje/Calidad	Catagórico	Calidad o tamaño del grano
Cantidad	Continuo	Cantidad producida
Temp 2m	Continuo	Temperatura 2 metros de altura
Temp Dew 2m	Continuo	Temperatura de rocío a 2 metros de altura
Temp Humeda 2m	Continuo	Temperatura humedad a 2 metros de altura
Temp Superficie	Continuo	Temperatura de la superficie
Temp Mínima 2m	Continuo	Temperatura minima a 2 metros de altura
Temp Máxima 2m	Continuo	Temperatura máxima a 2 metros de altura
Rango Temp 2m	Continuo	Rango de temperatura a 2 metros de altura
Radiación Solar Superficie	Continuo	Radiación solar en la superficie
Evapotranspiración Superficie	Continuo	Evapotranspiración en la superficie
Evapotranspiración 2m	Continuo	Evapotranspiración a 2 metros de altura
Humedad Relativa 2m	Continuo	Humedad relativa a 2 metros de altura

Precipitación Corregida	Continuo	Precipitación corregida
Presión Superficie	Continuo	Presión atmosférica en la superficie
Velocidad Viento 2m	Continuo	Velocidad del viento a 2 metros de altura
Velocidad Viento Máxima 2m	Continuo	Velocidad máxima del viento 2 metros de altura
Velocidad Viento Mínima 2m	Continuo	Velocidad mínima del viento a 2 metros de altura
Rango Velocidad Viento 2m	Continuo	Rango de velocidad del viento a 2 metros de altura
Dirección Viento 2m	Continuo	Dirección del viento a 2 metros de altura
Humedad Suelo Superficie	Continuo	Humedad del suelo en la superficie
Humedad Suelo Raíces	Continuo	Humedad del suelo en la zona de raíces

Nota: Variables s usados en el entrenamiento del modelo por su nivel de importancia. Elaborado por: El autor.

1.12.4 Enfoque Metodológico

La investigación adopto un enfoque metodológico mixto, combinando tanto métodos cuantitativos como cualitativos. Los métodos cuantitativos se utilizarán para analizar datos numéricos y obtener conclusiones estadísticas sobre las variables que afectan la producción de arándanos.

Los métodos cualitativos se emplearán para obtener una visión más clara de las prácticas agrícolas y las percepciones de los productores sobre los factores que influyen en el rendimiento de los cultivos.

Los datos cualitativos son los que darán una descripción más detallada del resultado de acuerdo con los grupos que manejen estas variables categóricas.

1.13 MODELADO

Para abordar el diseño del modelado de nuestro sistema y tener una idea más clara de cómo se va a abordar el modelado de nuestro sistema de predicción enfocado a la producción agrícola es importante analizar estudios anteriores que se asemejen a nuestro estudio como el de (Carrillo & Parraga-Alava, 2018; Parraga-Alava, et al., 2020), donde se han aplicado algoritmos de aprendizaje automático para tareas de clasificación y predicción, en el caso de predicción de cultivos, (Crane & Droesch, 2018), analizó el rendimiento del maíz en el medio oeste de Estados Unidos, aplicando un modelado que utiliza una variable semi paramétrica de una red neuronal

profunda, para de esta manera ser capaces de capturar la no linealidad de los datos. El enfoque de este estudio, al igual que (Khaki & Wang, 2019), se centró en ser capaz de tratar con el problema de la alta dimensionalidad de los datos, y ser capaces de capturar gran parte de la no linealidad de estos.

Para el desarrollo del sistema de predicción se emplearán diversos algoritmos de aprendizaje automático:

Algoritmos de aprendizaje automático: Incluyen regresión lineal, máquinas de soporte vectorial (SVM), redes neuronales artificiales (ANN) y bosques aleatorios (RF).

Validación de modelos: Los modelos predictivos se validarán mediante técnicas de validación cruzada y se evaluarán utilizando métricas como el error cuadrático medio (MSE).

Tabla6

Modelos Utilizados y Parámetros

Modelo	Parámetro	Valor a Evaluar
Random Forest	Bootstrap	Activo
Random Forest	Bootstrap	Inactivo
Random Forest	Estimadores	100
Random Forest	Estimadores	200
Random Forest	Estimadores	500
Random Forest	Estimadores	1000
Random Forest	Criterio	MSE
Random Forest	Criterio	MAE
Árboles de decisión	Profundidad	10
Árboles de decisión	Profundidad	15
Árboles de decisión	Profundidad	20
Árboles de decisión	Splitter	Random
Árboles de decisión	Splitter	Best
Árboles de decisión	Criterio	MSE
Árboles de decisión	Criterio	FMSE
Árboles de decisión	Criterio	MAE
SVM	C	0.01
SVM	C	0.1

SVM	C	1
SVM	Gamma	10
SVM	Gamma	0.1
SVM	Gamma	1
DNN	Kernel	10
DNN	Kernel	RBF
DNN	Capas ocultas	Linear
DNN	Capas ocultas	-100
DNN	Capas ocultas	(100, 100)
DNN	Funciones de activación	(100, 100, 100)

Nota: Esas medidas fueron utilizadas de acuerdo a los parámetros considerados para evaluar en las pruebas de cada modelo de aprendizaje automático. Elaborado por: El autor.

1.13.1 Métricas de validación

Para asegurar la funcionalidad y precisión del sistema, se llevarán a cabo las siguientes métricas que son utilizadas para evaluar la precisión y exactitud de tus modelos de predicción en la cantidad de arándanos. Cada una de estas métricas proporciona una perspectiva diferente sobre el rendimiento del modelo, permitiendo una evaluación completa y detallada.

1.13.2 Mean Squared Error (MSE)

Esta métrica mide el promedio de los cuadrados de los errores, es decir, la diferencia promedio cuadrada, donde y representa la cantidad real de arándanos que representa la cantidad predicha por el modelo.

Penaliza los errores grandes más que los pequeños debido al cuadrado de las diferencias.

Esto es útil para identificar modelos que tienen grandes desviaciones en las predicciones.

En la predicción de la cantidad de arándanos, si una predicción está muy alejada del valor real, el MSE reflejará fuertemente este error, ayudando a ajustar el modelo para mejorar la precisión.

1.13.3 R-squared (R^2)

Donde (\bar{y}) es la media de los valores reales. Esta métrica representa la proporción de la varianza total de la cantidad de arándanos que es explicada por el modelo.

Proporciona una medida de cuánta de la variación total en la variable dependiente (cantidad de arándanos) es explicada por las variables independientes en el modelo. Un valor cercano a 1 indica que el modelo explica bien la variabilidad.

Ayuda a entender la eficacia general del modelo para explicar la producción de arándanos y puede ser comparado entre diferentes modelos para seleccionar el mejor.

1.13.4 Mean Absolute Error (MAE)

Mide el promedio de los errores absolutos entre las cantidades reales y las cantidades predichos.

Es una métrica simple que mide el promedio de los errores absolutos. Es menos sensible a valores atípicos que el MSE.

Proporciona una interpretación directa del error promedio en las predicciones de cantidad de arándanos, siendo fácil de entender y comunicar.

1.13.5 Mean Absolute Percentage Error (MAPE)

Esta métrica mide el error en términos porcentuales entre las cantidades reales y las predichas.

Mide el error en términos porcentuales, lo que facilita la interpretación del rendimiento del modelo en un contexto relativo.

En la predicción de la producción de arándanos, MAPE es útil para entender el error en relación con el tamaño de la cantidad predicha, permitiendo comparaciones relativas entre diferentes escenarios de producción.

1.13.6 Root Mean Squared Error (RMSE)

Es la raíz cuadrada del error cuadrático medio y proporciona una medida de la magnitud del error.

1.13.7 Explained Variance Score (EVS)

Mide la proporción de la varianza de la cantidad de arándanos que es explicada por las predicciones.

1.14 DESARROLLO DEL SISTEMA

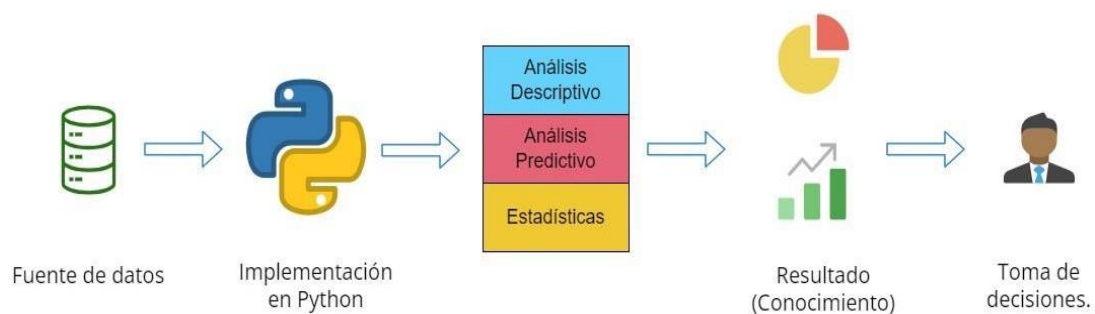
1.14.1 Arquitectura del Sistema

Para el diseño del sistema se utilizó una estructura organizativa de acuerdo a un orden jerárquico. Es decir, el procesamiento se da a través de un dataset que ha recopilado información de varias bases de datos de las cuales se han escogido las variables de acuerdo a su nivel de correlación. Luego, estas serán obtenidas por el lenguaje de programación Python donde usamos las librerías necesarias para evaluar a través de un análisis descriptivo el mejor modelo de predicción.

a usar en el entrenamiento de datos una vez escogido se utilizara el modelo que mejor se ajuste para la predicción para de esta manera dar paso a un análisis predictivo el cual lo daremos a través de diagramas donde muestren los valores reales con los valores predichos para ver qué tan exacto es el modelo en una serie de tiempo dada estos resultados se tomaran como el nuevo conocimiento al saber que variables son más correlacionales es decir las principales variables que afectan al rendimiento de la producción también con el mejor modelose podrá ingresar por consola las variables de prueba que serían las más relevantes encarecterísticas del nivel de correlación y con estas dar paso a la ejecución del modelo con todas las variables de entrenamiento y así tener otro resultado de conocimiento con un sistema de predicción que se ira desarrollando de acuerdo a una estructura organizacional y que utiliza metodologías agiles para ver el progreso constante de acuerdo a metas semanales que se manejaran de acuerdo a la arquitectura del sistema:

Figura15

Arquitectura del Sistema



Nota: Este diseño muestra no solo la arquitectura sino también el alcance y los pasos de implementación para el mismo. Elaborado por: El autor.

CAPÍTULO IV

RESULTADOS

1.15 ANÁLISIS DE DATOS

En esta sección se presentarán los resultados del análisis de los datos recolectados. Se describirán las principales tendencias y patrones observados, así como las relaciones identificadas entre las variables estudiadas y la producción de arándanos para lo cual se han evaluado las variables que tienen más relación con métricas que miden su grado de importancia tanto como para ver las variables que mejoraran el modelo para la predicción como las más importantes al momento de predecir esto se ha logrado evaluándolas, también habla de la importancia de seleccionar adecuadamente las variables predictoras y el modelo de predicción más adecuado para optimizar la producción de arándanos.

Las conclusiones de los análisis de las variables con respecto al rendimiento del fruto serán acordes a los modelos evaluados, el Random Forest ofrece la mejor combinación de precisión y capacidad explicativa, lo que lo convierte en la herramienta más eficaz para predecir la cantidad de producción de arándanos bajo las condiciones y explica la importancia de las variables analizadas tanto en la ganancia de la información como en cuanto ayuda a mejorar la precisión del modelo.

A continuación, se presenta una tabla con los resultados obtenidos de acuerdo con el grado de correlación que existe entre las variables independientes o predictoras y la variable dependiente a predecir cantidad donde de acuerdo con la importancia va a depender su influencia y de igual manera se describe su importancia en la aplicación:

Tabla7

Influencia e Importancia de las Variables

Variable	Grado de Correlación	Importancia	Influyente en la Predicción	Razón
Gramaje/Calidad	0.54	Alta	Sí	Importante para la calidad del producto final
regado(cm3)	0.08	Media	Sí	Crucial para el crecimiento de las plantas
Cantidad	1.0	Alta	Sí	Directamente relacionada con la cantidad de producción
Temp 2m	0.45	Alta	Sí	Impacta en el desarrollo de la planta
Temp Dew 2m	0.59	Alta	Sí	Afecta el punto de rocío y la humedad del ambiente
Temp Húmeda 2m	0.99	Alta	Sí	Indicador de la humedad en el aire
Temp Superficie	0.96	Alta	Sí	Crucial para el metabolismo de las plantas
Rango Temp 2m	0.48	Media	Sí	Rango de temperatura afecta la salud de las plantas

Temp Máxima 2m	0.99	Alta	Sí	Temperatura máxima afecta la fotosíntesis
Temp Mínima 2m	0.39	Media	Sí	Temperatura mínima puede impactar negativamente
Humedad Específica 2m	0.75	Alta	Sí	Humedad específica es crucial para la transpiración
Humedad Relativa 2m	0.4	Media	Sí	Relación directa con la disponibilidad de agua
Precipitación Corregida	0.03	Baja	No	Precipitación afecta el riego natural
Presión Superficie	0.66	Media	Sí	Impacta en la presión de vapor y salud del suelo
Velocidad Viento 2m	0.04	Baja	No	Velocidad del viento puede impactar la polinización
Velocidad Viento Máxima 2m	0.2	Baja	No	No es una variable principal
Velocidad Viento Mínima 2m	0.05	Baja	No	No es una variable principal

Rango Velocidad Viento 2m	0.07	Baja	No	No es una variable principal
Dirección Viento 2m	0.13	Baja	No	No es una variable principal
Humedad Suelo Superficie	0.73	Alta	Sí	Humedad del suelo es crucial para el riego adecuado
Humedad Suelo Raíces	0.7	Alta	Sí	Humedad de las raíces afecta la absorción de nutrientes
Mes de siembra	0.07	Baja	No	No es una variable principal
Mes de cosecha	0.43	Media	Sí	Puede influir en el ciclo de crecimiento
Año de siembra	0.04	Baja	No	No es una variable principal
Año de cosecha	0.08	Baja	No	No es una variable principal
Variedad	0.03	Baja	No	No es una variable principal

Nota: Resultados obtenidos de los promedios obtenidos de acuerdo con el grado de correlación, importancia y de influencia en la predicción. Elaborado por: El autor.

Las variables climáticas y de manejo altamente correlacionadas deben ser monitoreadas y gestionadas cuidadosamente, ya que tienen un impacto significativo en la producción de arándanos como la relación entre la humedad del suelo, la temperatura y la presión superficial sugiere que la gestión del riego y la presión pueden optimizarse para mejorar las condiciones de cultivo y, por ende, la producción.

Algunas variables, como Bloque, Variedad, Mes de siembra" y "Mes de cosecha, no muestran correlaciones significativas con las variables de producción, lo que sugiere que estos factores temporales pueden no ser tan influyentes en la variabilidad de la producción de arándanos en el contexto analizado.

1.15.2 Evaluación de Métricas

A continuación, se muestra los resultados de la implementación de los modelos de predicción para las variables para lo cual se evaluaron los siguientes modelos Árbol de Decisiones, Random Forest, SVM, Redes Neuronales (ANN) y Regresión Polinómica, para la comparación entre modelos para elegir el modelo más adecuado utilizando las siguientes métricas: MSE, R^2 , MAE, MAPE, RMSE y EVS.

El R^2 es el coeficiente que indica qué tan bien el modelo explica la variabilidad en la producción de arándanos:

Tabla8

Interpretación y Comparación de Resultados

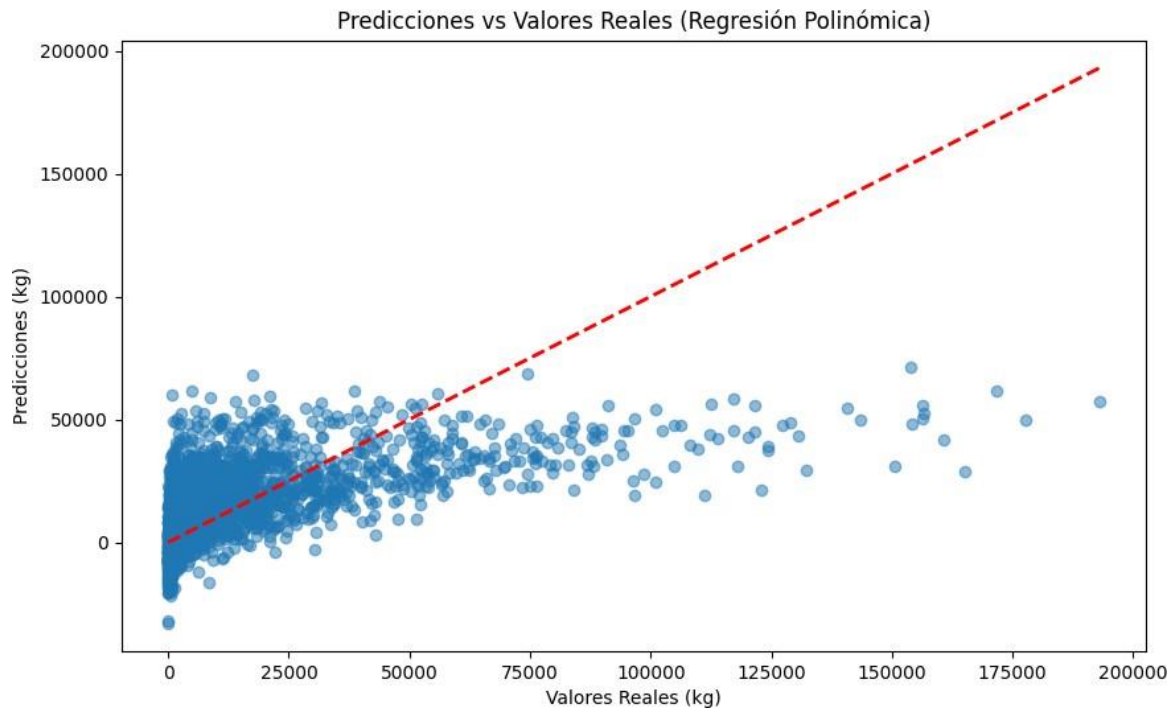
Modelo	MSE	R²	MAE	MAPE	RMSE	EVS
Árbol deDecisiones	219.83	0.4	10.47	0.15	14.82	0.4
RandomForest	161.62	0.53	9.23	0.13	12.71	0.53
SVM	233.77	0.36	11.47	0.18	15.29	0.36
ANN	151.67	0.56	8.92	0.12	12.32	0.56
Regresión Polinómica	207.32	0.44	10.87	0.16	14.4	0.44

Nota: Resultados obtenidos de las métricas de los modelos de predicción evaluados. Elaborado por: El autor.

El MSE de 278195070.72 es el más bajo entre todos los modelos, lo que significa que tiene los errores promedio más pequeños en sus predicciones, regresión Polinómica y ANN tienen valores moderados de R² (0.33 y 0.3 respectivamente), árbol de Decisiones y SVM tienen valores negativos de R² (-0.07 y -0.14 respectivamente), lo que significa que estos modelos no son útiles y su desempeño es peor que simplemente usar la media de los datos.

Figura17

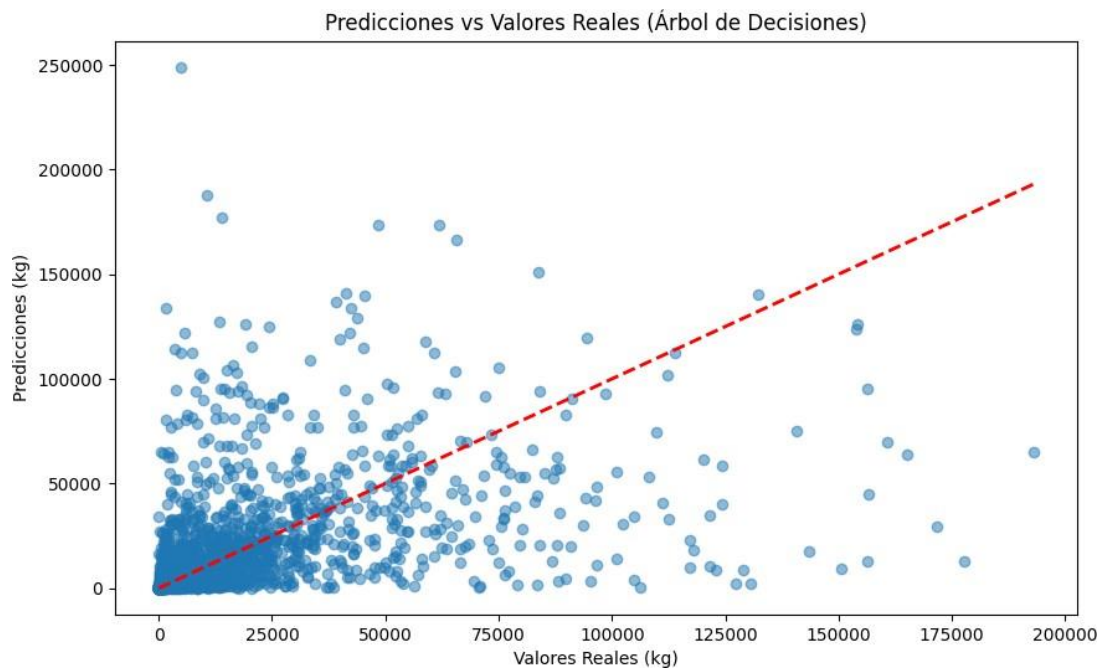
Predicciones vs Valores Reales (Regresión Polinómica)



Nota: Regresión Polinómica y ANN tienen MSE más altos que Random Forest, pero significativamente más bajos que Árbol de Decisiones y SVM. Elaborado por: El autor.

Figura18

Predicciones vs Valores Reales (Árbol de Decisiones)



Nota: Árbol de Decisiones y SVM tienen los valores de MSE más altos, indicando grandes errores en sus predicciones. Elaborado por: El autor.

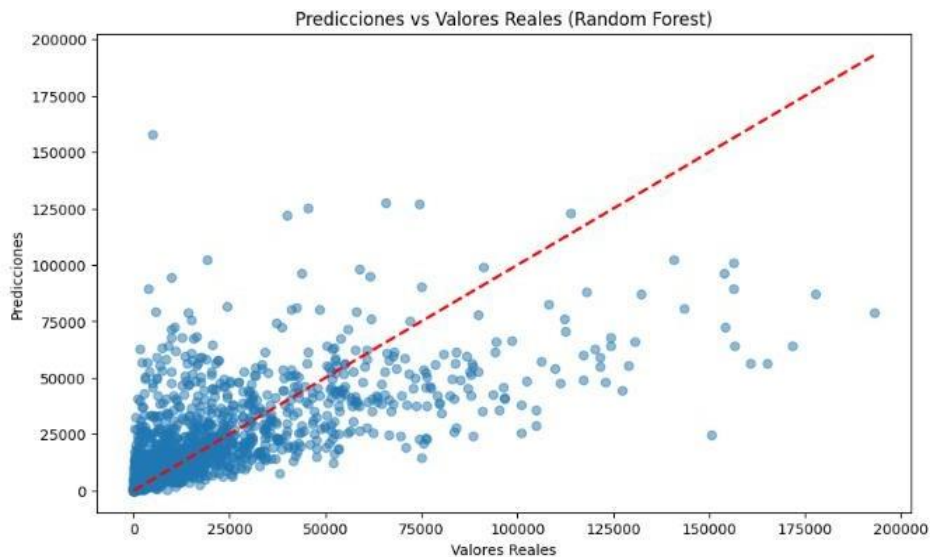
Random Forest tiene el MSE más bajo (278195070.72), lo que indica que sus predicciones están más cercanas a los valores reales en promedio.

R^2 de 0.5: Indica que explica el 50% de la variabilidad en la producción de arándanos, lo cual es considerablemente mejor que los otros modelos.

Basado en el análisis de R^2 y MSE, el modelo de Random Forest es claramente el más adecuado para predecir la producción de arándanos, ya que ofrece un equilibrio óptimo entre explicación de la variabilidad y los otros modelos, aunque algunos tienen valores moderados de R^2 , no alcanzan la precisión y capacidad explicativa del modelo Random Forest.

Figura19

Predicciones vs Valores Reales (Random Forest)



Nota: Las predicciones están más cercanas a los valores reales en promedio. Elaborado por: El autor.

1.16 RESULTADOS DE LA PREDICCIÓN

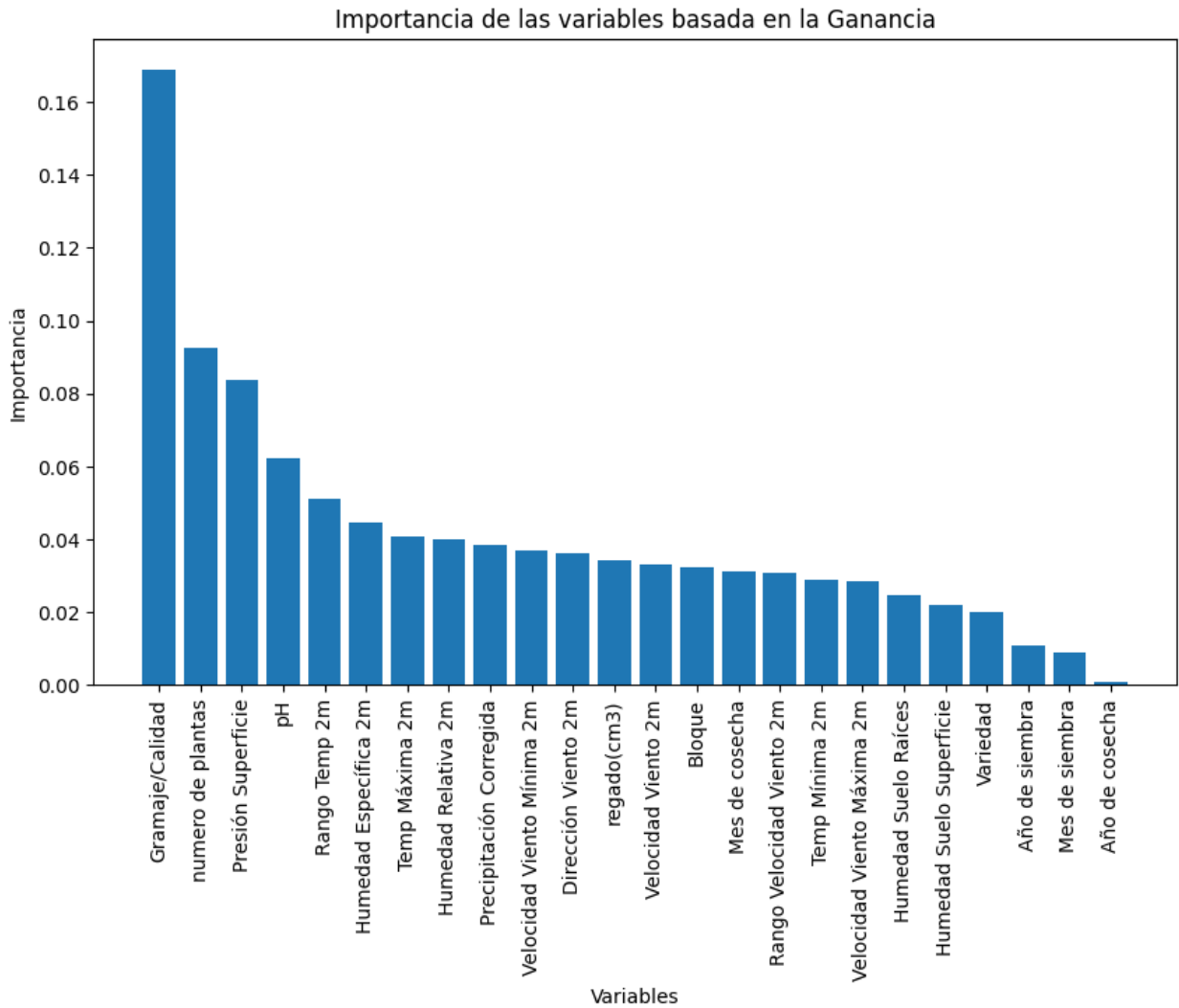
Esta gráfica muestra la importancia de las variables en función de la reducción de la impureza que cada variable proporciona al árbol de decisión. En otras palabras, se mide cuánto mejora la predicción cada variable cuando se usa para dividir los datos en los nodos del árbol. La métrica utilizada aquí es la ganancia de información o disminución de la ganancia.

La importancia de permutación evalúa cuánto disminuye la precisión del modelo cuando se permutan (mezclan) los valores de una característica. Si la permutación de una característica reduce significativamente la precisión del modelo, esa característica se considera importante.

La métrica utilizada aquí es la reducción en la precisión del modelo (por ejemplo, reducción en el R^2 o aumento en el error) cuando los valores de la característica se permutan.

Figura 20

Importancia de las Variables Basadas en la Ganancia

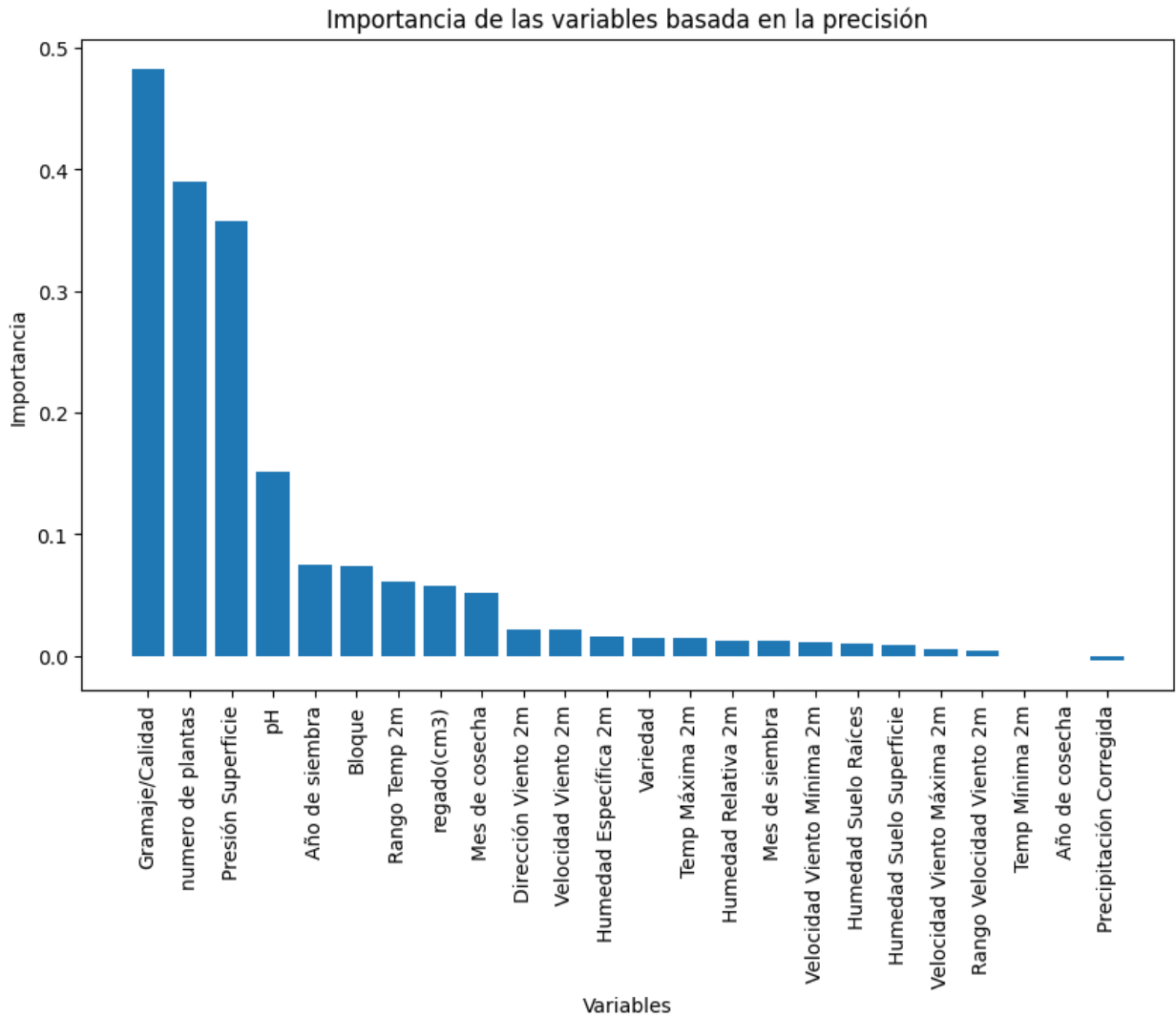


Nota: Esta gráfica utiliza la importancia de permutación para medir la importancia de las variables. Elaborado por: El autor.

En esta gráfica, Gramaje/Calidad sigue siendo la variable más importante, seguida de numero de plantas, regado(cm3) y "Presión Superficie.

Figura 21

Importancia de las Variables Basadas en la Precisión



Nota: Esta gráfica utiliza la importancia de permutación para medir la importancia de las variables. Elaborado por: El autor.

Comparando las cantidades reales con las predichas para distintas categorías de gramaje, se observa que el modelo ofrece una buena aproximación en términos generales, aunque existen discrepancias notables en algunas categorías como "MUY PEQ." y "GRANDE."

Tabla 9*Cantidad de Producción Real por Gramaje*

Gramaje	Cantidad Real
MUY PEQ.	2809,7
PEQUEÑA	82745,2
MEDIANA	109302,11
GRANDE	3682,9
DEFORME	18206,3

Nota: Muestra real de la cantidad de arándanos obtenidos de acuerdo con un periodo de tiempo.
Elaborado por: El autor.

Tabla 10*Predicciones de la Cantidad de Producción*

Gramaje	Cantidad Predicha
MUY PEQ.	4664.12
PEQUEÑA	19160.02
MEDIANA	94109.68
GRANDE	62552.34
DEFORME	2493.67

Nota: Muestra predicha de la cantidad de arándanos dados el mismo periodo anterior utilizando Random Forest. Elaborado por: Autor.

1.17 IMPACTO DEL SISTEMA DE PREDICCIÓN

El sistema de predicción basado en Random Forest tiene el potencial de optimizar la producción de arándanos de varias maneras:

Aumento del Rendimiento: Al identificar las variables más influyentes en la producción, los agricultores pueden centrarse en optimizar estas variables para mejorar el rendimiento de los cultivos.

Eficiencia en el Uso de Recursos: El sistema ayuda a utilizar los recursos de manera más eficiente, por ejemplo, optimizando el riego según las necesidades específicas del cultivo, lo que puede reducir el desperdicio de agua.

Reducción de Riesgos: Al predecir de manera precisa la producción, los agricultores pueden tomar decisiones más informadas para mitigar los riesgos asociados con factores climáticos y de manejo.

Fomento de Prácticas Agrícolas Sostenibles: El sistema no solo aumenta la productividad, sino que también promueve la adopción de prácticas agrícolas más sostenibles y responsables. Al optimizar el uso de insumos y minimizar el impacto ambiental, contribuye a la sostenibilidad a largo plazo de la producción de arándanos y otros cultivos.

Beneficios para la Industria Agrícola: Además de mejorar la producción de arándanos en Ilaru, el sistema desarrollado tiene un potencial de replicabilidad en otras operaciones agrícolas. Esto puede beneficiar a otros agricultores y contribuir al desarrollo de la agricultura de precisión en Ecuador, sentando las bases para futuras investigaciones y aplicaciones tecnológicas en el sector agrícola.

CONCLUSIÓN

La fortaleza del Sistema de predicción en la producción de arándanos utilizando el modelo Random Forest se destaca por su alto R^2 y su baja MSE, lo que indica una alta precisión y capacidad explicativa. Estos indicadores reflejan que el modelo es capaz de explicar la variabilidad en los datos de producción de manera efectiva, proporcionando predicciones confiables que pueden ser utilizadas para la toma de decisiones en la gestión del cultivo.

Una de las fortalezas más significativas del modelo Random Forest es su capacidad para identificar las variables más influyentes en la producción de arándanos, variables como "Gramaje/Calidad", "regado(cm3)", "número de plantas" y "Bloque" han sido destacadas como determinantes principales. Esta identificación permite una gestión más focalizada y efectiva, permitiendo a los agricultores optimizar estas variables clave para mejorar la cantidad de producción.

Aunque el modelo ha mostrado un buen desempeño general, se han observado discrepancias en las predicciones para ciertas categorías de gramaje. Por ejemplo, las predicciones para las categorías "MUY PEQ." y "GRANDE" difieren significativamente de las cantidades reales, estas discrepancias indican que hay margen para mejorar la precisión del modelo en estos casos específicos, posiblemente ajustando los parámetros del modelo o incluyendo variables adicionales.

RECOMENDACIONES

La precisión y efectividad del modelo dependen en gran medida de la calidad y cantidad de los datos disponibles como datos insuficientes, desactualizados o de mala calidad pueden afectar negativamente las predicciones. Es fundamental contar con una base de datos robusta y actualizada para asegurar la fiabilidad del modelo.

Incrementar la cantidad y la calidad de los datos recolectados es esencial para mejorar las predicciones del modelo, se recomienda incluir más variables que puedan influir en la producción de arándanos, así como asegurar la precisión y consistencia en la recolección de datos. Esto permitirá al modelo captar mejor las complejidades del sistema de producción y realizar predicciones más precisas.

Aunque el modelo Random Forest ha demostrado ser efectivo, explorar y comparar otros modelos de predicción puede proporcionar una perspectiva más amplia y potencialmente mejores resultados, modelos como la regresión polinómica, redes neuronales y máquinas de soporte vectorial pueden ser investigados para complementar o mejorar los resultados obtenidos con Random Forest.

Fue importante probar el modelo en diferentes escenarios prácticos y ajustar los parámetros según las necesidades específicas de cada contexto de producción, la implementación en campo permitirá identificar áreas de mejora y ajustar el modelo para que sea más adaptable y preciso en diversas condiciones de cultivo.

BIBLIOGRAFÍA

- Carrillo, J. M., & Parraga-Alava, J. (2018). How Predicting the Academic Success of Students of the ESPAM MFL? A Preliminary Decision Trees Based Study. *2018 IEEE Third Ecuador Technical Chapters Meeting (ETCM)*, 1-6. <https://doi.org/10.1109/ETCM.2018.8580296>
- Crane-Droesch, A. (2018). Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environmental Research Letters*, *13*(11), 1-12. <https://doi.org/10.1088/1748-9326/aae159>
- Ferrero, R., Lima, M., & Gonzalez-Andujar, J. (2018). Crop production structure and stability under climate change in South America. *Annals of Applied Biology*, *172*, 65-73. <https://doi.org/10.1111/aab.12402>
- Food and Agriculture Organization (FAO). (2002). Agricultura mundial: hacia los años 2015/2030 [World agriculture: towards the years 2015/2030]. Technical Report 1. Recuperado de <https://n9.cl/n29i7>
- Garzón, D. (2023, 10 de abril). Arándano, un mercado naciente con grandes oportunidades. *Revista Líderes*. Recuperado de <https://www.revistalideres.ec/lideres/arandano-mercado-ecuador-oportunidades.html>
- Herrera-Díaz, C. A. (2016). Implementación de un módulo de análisis estadístico y predictivo para agricultura utilizando bigdata y machine learning, integrado al sistema Iotmach [Implementation of a statistical and predictive analysis module for agriculture using bigdata and machine learning, integrated to the Iotmach system]. *Tesis de pregrado, Universidad Técnica de Machala*. Recuperado de <https://n9.cl/abp1>
- Khaki, S., & Wang, L. (2019). Crop Yield Prediction Using Deep Neural Networks. *Frontiers in Plant Science*, *10*:621. <https://doi.org/10.3389/fpls.2019.00621>
- Laurentin, H. (2020). Importancia de la predicción del rendimiento en caña de azúcar en un contexto de transformación digital. *SofOS*. Recuperado de <https://n9.cl/fl2vh>
- Marengo, J. A., Chou, S. C., Torres, R. R., Giarolla, A., Alves, L. M., & Lyra, A. (2014). Climate change in Central and South America: Recent trends, future projections, and impacts on regional agriculture. *CCAFS Working Paper no. 73*. Copenhagen, Denmark: CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS). Recuperado de <https://hdl.handle.net/10568/41912>
- Parraga-Alava, J., Alcivar-Cevallos, R., Riascos, J. A., & Becerra, M. A. (2020). Aphids Detection on Lemons Leaf Image Using Convolutional Neural Networks. In Botto-Tobar, M., Zamora, W., Larrea Plúa, J., Bazurto Roldan, J., & Santamaría Philco, A. (Eds.), *Systems and Information Sciences. ICCIS 2020. Advances in Intelligent Systems and Computing*, 1273. Springer, Cham. https://doi.org/10.1007/978-3-030-59194-6_2

- Pierce, F. J., & Nowak, P. (1999). Aspects of precision agriculture. *Advances in Agronomy*, 67, 1-85. [https://doi.org/10.1016/S0065-2113\(08\)60513-1](https://doi.org/10.1016/S0065-2113(08)60513-1)
- Seo, S. N., & Mendelsohn, R. (2008). An analysis of crop choice: Adapting to climate change in South American farms. *Ecological Economics*, 67(1), 109-116. <https://doi.org/10.1016/j.ecolecon.2007.12.007>
- Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177. <https://doi.org/10.1016/j.compag.2020.105709>

GLOSARIO

R-squared (R^2): Métrica que representa la proporción de la varianza total de la cantidad de arándanos que es explicada por el modelo. Proporciona una medida de cuánta de la variación total en la variable dependiente es explicada por las variables independientes en el modelo.

Mean Absolute Error (MAE): Mide el promedio de los errores absolutos entre las cantidades reales y las cantidades predichas. Es una métrica simple y menos sensible a valores atípicos.

Mean Absolute Percentage Error (MAPE): Mide el error en términos porcentuales entre las cantidades reales y las predichas, facilitando la interpretación del rendimiento del modelo en un contexto relativo.

Root Mean Squared Error (RMSE): Es la raíz cuadrada del error cuadrático medio y proporciona una medida de la magnitud del error.

Explained Variance Score (EVS): Mide la proporción de la varianza de la cantidad de arándanos que es explicada por las predicciones.

Sistema de mezcla: Software que puede asignar el porcentaje de químicos a usar por mezcla y determinar por qué tubería enviar dicha solución.

NASA POWER (Prediction Of Worldwide Energy Resources): Sistema que utiliza una combinación de datos satelitales y modelos climáticos para proporcionar información climática precisa y detallada.

Ciclo de cultivo: Fases que incluyen la plantación, crecimiento vegetativo, floración, fructificación y cosecha, cruciales para el desarrollo del sistema de producción de arándanos.

Agricultura de precisión: Práctica que utiliza tecnología avanzada para gestionar los cultivos de manera más eficiente, incluyendo el uso de sensores y sistemas de información

geográfica (SIG).

Aaazuli Software: Plataforma utilizada para gestionar y recopilar datos de producción de arándanos, incluyendo información detallada sobre fecha, variedad, bloque, empleado, gramaje/calidad, cantidad y código.

Factores ambientales: Variables como temperatura, humedad, precipitación y radiación solar que afectan directamente el crecimiento y la proliferación de las plantas.

Sensores y tecnología: Dispositivos utilizados para monitorear en tiempo real las condiciones climáticas y del suelo, esenciales para la agricultura de precisión.

Modelos de predicción: Algoritmos y técnicas de aprendizaje automático utilizados para predecir la producción de arándanos basados en variables relevantes.

Variables climáticas: Datos sobre temperatura, humedad, precipitación y otros parámetros climáticos que influyen en la producción agrícola.

Sistema de información geográfica (SIG): Herramienta para gestionar y analizar datos espaciales en la agricultura, facilitando la optimización de prácticas agrícolas.

Pruebas piloto: Evaluaciones realizadas en condiciones reales de campo para validar el rendimiento del sistema de predicción.

Control de calidad: Inspecciones y verificaciones en todas las etapas de producción y distribución para asegurar la calidad del producto final.