

UNIVERSIDAD CATÓLICA SANTO TORIBIO DE MOGROVEJO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN



Sistema inteligente basado en minería de datos para predecir la producción de fertilizante en la empresa Nutrition Vegetable Corporation Figal SAC

**TESIS PARA OPTAR EL TÍTULO DE
INGENIERO DE SISTEMAS Y COMPUTACIÓN**

AUTOR

Carlos Alexander Seclen Custodio

ASESOR

Jessie Leila Bravo Jaico

<https://orcid.org/0000-0001-6841-2536>

Chiclayo, 2025

**Sistema inteligente basado en minería de datos para predecir la
producción de fertilizante en la empresa Nutrition Vegetable
Corporation Figal SAC**

PRESENTADA POR
Carlos Alexander Seclen Custodio

A la Facultad de Ingeniería de la
Universidad Católica Santo Toribio de Mogrovejo
para optar el título de

INGENIERO DE SISTEMAS Y COMPUTACIÓN

APROBADA POR

Gregorio Manuel Leon Tenorio
PRESIDENTE

Ricardo David Iman Espinoza
SECRETARIO

Jessie Leila Bravo Jaico
VOCAL

Dedicatoria

Dedico este trabajo, ante todo, a Dios, por haberme dado la fuerza, la sabiduría y la perseverancia para alcanzar este momento tan importante en mi vida profesional. A mi madre, por su amor incondicional, por creer siempre en mí y por ser mi mayor inspiración en los días difíciles; sin su presencia y entrega, este sueño no habría sido posible.

Agradecimientos

Expreso mi más profundo agradecimiento a mi asesora, cuyos consejos y experiencia guiaron cada paso de este proyecto. A mi familia y amigos, que siempre me alentaron a seguir adelante, gracias por su paciencia y amor. Finalmente, a la institución que me brindó los recursos y el ambiente propicio para desarrollar este trabajo, les debo una parte fundamental de este logro.

SISTEMA INTELIGENTE BASADO EN MINERÍA DE DATOS PARA PREDECIR LA PRODUCCIÓN DE FERTILIZANTE EN LA EMPRESA NUTRITION VEGETABLE CORPORATION FIGAL SAC

INFORME DE ORIGINALIDAD

21%

INDICE DE SIMILITUD

20%

FUENTES DE INTERNET

4%

PUBLICACIONES

7%

TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	hdl.handle.net Fuente de Internet	6%
2	tesis.usat.edu.pe Fuente de Internet	1%
3	laccei.org Fuente de Internet	1%
4	repository.uamerica.edu.co Fuente de Internet	1%
5	repositorio.puce.edu.ec Fuente de Internet	<1%
6	www.coursehero.com Fuente de Internet	<1%
7	repositorio.ucv.edu.pe Fuente de Internet	<1%
8	Submitted to Universidad Nacional del Centro del Peru Trabajo del estudiante	<1%
9	revistas.udec.cl Fuente de Internet	

Índice

Resumen	6
Abstract	7
Introducción.....	8
Revisión de literatura.....	10
Materiales y métodos.....	11
Resultados y discusión.....	13
FASE #1: Comprensión del negocio	13
FASE #2: Comprensión de los datos	15
FASE #3: Preparación de los datos	20
FASE #4: Modelado	23
FASE #5: Evaluación.....	28
FASE #6: Despliegue.....	30
Conclusiones.....	35
Recomendaciones.....	36
Referencias.....	37
Anexos.....	40

Resumen

En este estudio se desarrolló un sistema predictivo basado en minería de datos para la empresa Nutrition Vegetable Corporation Figal SAC, enfocado en optimizar su producción de fertilizantes mediante el análisis avanzado de datos históricos. Siguiendo la metodología CRISP-DM, se analizaron y evaluaron diversos modelos predictivos, aplicando técnicas de validación cruzada y métricas de rendimiento como precisión, exactitud y recall, donde la red neuronal alcanzó una capacidad predictiva del 53%. El sistema fue implementado a través de una interfaz web interactiva, validado bajo la norma ISO 25010, asegurando su usabilidad y eficiencia operativa en condiciones reales. Los resultados demostraron una mejora significativa en la planificación de la producción, evidenciada por la reducción de pérdidas operativas y una gestión más eficiente de recursos. La plataforma desarrollada permitió visualizar datos y generar predicciones en tiempo real, facilitando la toma de decisiones estratégicas y contribuyendo a prácticas más sostenibles en la cadena productiva, lo que posibilitó a la empresa anticipar y ajustar sus niveles de producción de manera más precisa, minimizando el desperdicio de recursos y mejorando su competitividad en el mercado. Las pruebas realizadas fueron llevadas a cabo mediante SonarCloud.

Palabras clave: Sistema inteligente, Minería de datos, Producción de fertilizantes, Redes neuronales, ISO 25010.

Abstract

In this study, a predictive system based on data mining was developed for the company Nutrition Vegetable Corporation Figal SAC, focused on optimizing its fertilizer production through advanced analysis of historical data. Following the CRISP-DM methodology, various predictive models were analyzed and evaluated, applying cross-validation techniques and performance metrics such as precision, accuracy and recall, where the neural network reached a predictive capacity of 53%. The system was implemented through an interactive web interface, validated under the ISO 25010 standard, ensuring its usability and operational efficiency in real conditions. The results demonstrated a significant improvement in production planning, evidenced by the reduction of operational losses and more efficient resource management. The developed platform made it possible to visualize data and generate predictions in real time, facilitating strategic decision making and contributing to more sustainable practices in the production chain, which enabled the company to anticipate and adjust its production levels more precisely, minimizing the waste of resources and improving its competitiveness in the market. The tests carried out were carried out through SonarCloud.

Keywords: Intelligent system, Data mining, Fertilizer production, Neural networks, ISO 25010.

Introducción

En la actualidad, los fertilizantes foliares se encuentran dentro del rango de elementos fundamentales en la productividad agrícola [1]. En este caso, el Ministerio de Desarrollo Agrario y Riego – MIDAGRI había recalcado la importancia de los fertilizantes foliares, destacando su capacidad para proporcionar nutrientes y mejorar la salud de las plantas mediante la absorción de nutrientes procedentes del suelo [2]. Desde este punto, la falta de las tecnologías de la Industria 4.0 junto con la ejecución inadecuada de los procesos de producción podría influenciar en los daños y pérdidas para la empresa [3].

A nivel mundial, la producción de fertilizantes foliares representaba el 11.3% del valor total del mercado de fertilizantes en el año 2021 [4]. En 2024, el mercado alcanzó los 20,18 mil millones de dólares, con predicciones de crecimiento hasta los 28,59 mil millones de dólares para el año 2030 [5]. En el Perú, la Encuesta Nacional Agraria del 2022, mencionó que 2,243,811 millones de personas reportaron alguna actividad agrícola siendo solo el 69,6% son productores agrarios y 1.7 millones de hectáreas dependían de la fertilización, siendo los cultivos como arroz, maíz y papa los de alta dependencia. [6] En este contexto, la empresa Nutrition Vegetable Corporation Figal SAC, en la Región Lambayeque, producía aproximadamente 300 mil litros anuales de ácido fólico, contribuyendo a la oferta de fertilizantes foliares en el mercado.

Los factores que influían en la producción de fertilizantes foliares, comenzando por, la variación estacional de la demanda del producto, la cual generó desequilibrio en los niveles de producción requerida en diferentes épocas del año [7]. La calidad del producto jugaba un papel fundamental en la percepción y satisfacción del cliente, convirtiéndose en un elemento clave para la aceptación y éxito comercial. [7]. Además, la competencia en el mercado añadía presión sobre los precios, provocando fluctuaciones que también impactaban la producción y la capacidad de mantener costos estables. [8]. Los costos de producción, como mano de obra, energía, mantenimiento de máquinas, capacitación del personal y gastos de transporte, eran factores críticos que influían directamente en la rentabilidad y viabilidad de la producción [9], [10]. La insatisfacción de los clientes por la calidad del producto y el incumplimiento de los pedidos son las consecuencias que generaría pérdida para la empresa [11]. Los problemas presentados, afectaban la rentabilidad y la capacidad de producir volúmenes deseados; mismo

modo, los retrasos en la entrega de materias primas o la calidad deficiente de esta, influían en la producción [12].

Diversas investigaciones habían empleado sistemas inteligentes con un enfoque específico en la mejora de la producción, entre estos estudios, se destacaba el diseño de un modelo predictivo utilizando técnicas de Machine Learning para predecir los pesos de racimos en la agricultura bananera en Ecuador, lo que permitía a los productores tomar decisiones sobre riego, fumigación y fertilización, mejorando así la productividad de las plantaciones [13]. Otro estudio se centraba en la creación de un modelo predictivo automatizado que permitía mejorar la planificación de la producción de arroz en la región San Martín, reduciendo la imprecisión en las estimaciones de producción, lo que beneficiaba a las instituciones y organizaciones que utilizaban estos indicadores [14]. Además, se desarrolló un aplicativo web que utilizaba algoritmos de Machine Learning para predecir la producción lechera en una ganadería en la Sierra Norte del Ecuador, utilizando inteligencia artificial para mejorar el proceso de selección genética y facilitar la toma de decisiones por parte de los administradores de la ganadería [15].

Frente a los desafíos operativos en la producción de fertilizante foliar, que llevaban a una variabilidad en la calidad del producto y a pérdidas tanto para la empresa como para los agricultores, se había formulado la siguiente cuestión de investigación: ¿Cómo puede un sistema inteligente basado en minería de datos mejorar la predicción de la producción de fertilizantes foliares en la empresa Nutrition Vegetable Corporation Figal SAC, optimizando la eficiencia y reduciendo pérdidas?

Desde el punto de vista científico, la investigación representaba un aporte significativo al abordar la predicción de la producción de fertilizante foliar mediante el desarrollo de un sistema predictivo respaldado por inteligencia artificial. Este trabajo sentaba un valioso precedente para futuros investigadores interesados en explorar sistemas predictivos y aplicaciones de inteligencia artificial en la industria. El propósito específico de la predicción de la producción era mantener un control preciso, lo que prometía generar conocimiento innovador y beneficios potenciales para las empresas del sector.

La investigación proponía que la precisión en la predicción de la producción de fertilizantes mediante un sistema predictivo tenía un impacto financiero significativo. Al reducir los costos de producción, incluyendo mano de obra, mantenimiento, materias primas y capacitación, se

abría la posibilidad de optimizar recursos y maximizar ganancias, lo que representaba un beneficio considerable para las empresas del sector de fertilizantes.

Asimismo, el sistema brindaba beneficios a los agricultores, ya que les permitía mantener la calidad óptima de los fertilizantes disponibles para la venta. Además, al evitar la sobreproducción, se contribuía a reducir el impacto ambiental, disminuyendo las emisiones nocivas y promoviendo prácticas más sostenibles en la agricultura, con un impacto positivo en la comunidad.

Desde una perspectiva tecnológica, esta investigación se justificaba al utilizar tecnologías de vanguardia, particularmente la inteligencia artificial aplicada a un sistema predictivo para la producción de fertilizante foliar. La adopción de tecnologías actuales reflejaba el compromiso de mantenerse a la par del progreso tecnológico, aprovechando su potencial para mejorar la eficiencia y precisión en la producción de fertilizantes, un factor clave para el éxito en la industria.

En el planteamiento de la investigación como objetivo general de la investigación fue la implementación de un sistema inteligente basado en minería de datos para predecir la producción de fertilizantes en la empresa Nutrition Vegetable Corporation Figal SAC. Para el desarrollo del estudio, se identificó las técnicas más adecuadas para predecir la producción de fertilizante foliar basada en la caracterización de las variables clave que influían en su producción. Esto permitió el desarrollo de un modelo de predicción idóneo utilizando la técnica de inteligencia artificial seleccionada. Posteriormente, se diseñó una interfaz web para facilitar la interacción del usuario. Finalmente, se aseguró la calidad del sistema a través de la opinión de especialistas, siguiendo la normativa ISO 25010.

Revisión de literatura

Muñoz en su investigación [13] abordó la problemática de la gestión agrícola, la cual carecía de precisión y eficiencia, resultando en menor productividad y un uso ineficiente de recursos como riego, fertilización y fumigación. Como alternativa de solución, se planteó la aplicación de técnicas de Machine Learning para predecir los pesos de racimos en la agricultura bananera de Ecuador, con el objetivo de ayudar a los productores a tomar decisiones informadas. Los

resultados demostraron que el modelo Xgboost fue el más efectivo, con buenos indicadores de precisión, como el Error Absoluto Medio (MAE) y el Error Porcentual Absoluto Medio (MAPE).

Ordoñez en su estudio [15] se centró en mejorar el rendimiento genético y la gestión de la producción lechera en el Instituto Agropecuario Superior Andino (IASA), donde existía una ineficiencia en la gestión de producción debido a la falta de predicción precisa. Se propuso como solución la implementación de la metodología CRISP-DM y el uso de algoritmos de regresión lineal múltiple para predecir la producción lechera. El modelo logró un coeficiente de determinación ajustado (R^2) de 0,80, mostrando efectividad en la predicción.

Jiménez, et al. [16], enfrentaron la problemática de la planificación ineficiente en la producción y venta de alimentos procesados en la panadería "La Cascada", lo que generaba excesos y escasez en productos, afectando las ganancias. Se desarrolló un modelo de pronóstico de demanda basado en Machine Learning y un modelo de programación multiobjetivo para mejorar la planificación de producción. Los resultados indicaron un aumento del 17,23% en las ganancias y una reducción del 6,36% en los costos operativos.

Porras y Ovalle [17] investigaron la necesidad de prevenir la sobreproducción de cacao en Perú, que podría llevar a una crisis de precios y perjudicar a los agricultores. Como solución, desarrollaron un modelo de predicción de la producción de cacao usando técnicas de Deep Learning y LSTM. El modelo resultante logró una precisión del 99,6% en la estimación de la producción, ayudando a prevenir la sobreoferta y estabilizando el mercado.

Materiales y métodos

Se planeó llevar a cabo esta investigación de tipo de aplicada, según lo establecido en el Manual de Frascati [18]. Se logrará examinar trabajos de investigación originales para adquirir nuevos conocimientos que se enfocan principalmente en un objeto o propósito específico [18]. Se busca desarrollar un sistema inteligente que pueda predecir la producción del fertilizante foliar evitando pérdidas para la empresa.

Métodos de investigación

Mencionar los métodos de investigación a emplear:

TABLA I
MÉTODOS DE INVESTIGACIÓN

Método	Sustento por el cual será empleado en la investigación
Deductivo	Se empleará este método debido a que nos permitirá identificar la problemática
Análisis bibliográfico	Se utilizará este método ya que nos permitirá investigar y recopilar diferentes antecedentes y teorías que serán esenciales para fundamentar los argumentos en nuestra investigación.
Inductivo	La razón por la que aplicaremos este método es para poder proponer una solución al problema identificado.
Implementación	Se aplicará este método porque permite dar nuevas soluciones, tales como algoritmos, con el objetivo de abordar y resolver el problema identificado.

TABLA II
TÉCNICAS DE INVESTIGACIÓN

Técnicas	Instrumentos
Análisis de documentos	Analizaremos las fuentes bibliográficas para utilizarlos y llegar al conocimiento del tema a investigar.
Revisión Sistemática de la Literatura RSU	Se trata de evaluar e interpretar exhaustiva de toda la investigación disponible y relevante en relación con una pregunta específica de investigación.
Juicio de experto	Lista de cotejos

Metodología de desarrollo

La metodología elegida para esta investigación fue el CRISP-DM (Cross-Industry Standard Process for Data Mining), una de las más empleadas en el ámbito de la minería de datos [19]. Esta metodología se distinguió por su enfoque estructurado y sistemático, lo que facilitó la gestión integral de proyectos de minería de datos, abarcando desde la identificación del problema hasta la implementación de soluciones. Las fases del proceso fueron las siguientes:

1. Entendimiento del negocio
2. Entendimiento de los datos
3. Preparación de datos
4. Modelado
5. Evaluación
6. Despliegue

Resultados y discusión

En base a la metodología utilizada, La fases de la metodología son:

FASE #1: Comprensión del negocio

El objetivo principal es comprender a fondo el contexto y los objetivos del negocio para los cuales se está realizando el proyecto. Esto implica interactuar con los principales interesados, como los gerentes de la empresa Nutrition Vegetable Corporation Figal SAC, para entender sus necesidades, problemas y objetivos específicos. Algunas actividades clave en esta fase incluyen:

Determinar objetivos del negocio

A través de un análisis exhaustivo, se logró comprender detalladamente el proceso de producción de fertilizantes foliares en Nutrition Vegetable Corporation Figal SAC, identificando puntos críticos donde un sistema de predicción impactaría positivamente en la toma de decisiones operativas y estratégicas. Este entendimiento profundo del ciclo productivo reveló la necesidad de implementar una solución tecnológica basada en modelos predictivos,

que no solo optimizaría la eficiencia en la producción, sino que también proporcionaría información valiosa para la planificación de recursos y la gestión del inventario. La clara identificación de estas necesidades estableció los cimientos para desarrollar un sistema que se alinee perfectamente con los objetivos de mejora continua y expansión de la empresa en el mercado de fertilizantes.

Evaluar la situación actual del negocio

La empresa con la que se trabajó estaba enfocada en el rubro de la industria química y la agricultura, lo que buscaba era la efectividad de fertilizantes para así lograr la satisfacción de sus clientes al usarlos. Por ello, necesitaba analizar las necesidades de sus clientes, lo cual le permitiría cubrir la demanda que se tenía por temporadas.

La demanda de los productos variaba por temporadas, lo que ocasionaba que para la empresa fuera un problema, porque en ocasiones no podía cubrir la demanda y en otros casos excedía su producción, lo que le ocasionaba pérdidas. Del mismo modo, el exceso de producción producía daños físicos para la empresa, así como también contaminación del aire, provocada por la emisión de gases del producto.

De esta forma, se buscó predecir la cantidad de producción que requería la empresa por las temporadas del año, llegando a solucionar problemas con la demanda y logrando reducir las pérdidas para la empresa, así como también la contaminación del aire que se podía ocasionar.

Determinar objetivos de minería de datos

- Se manejó la información que fue proporcionada por la empresa para evaluar tendencias, así como también qué variables eran importantes.
- Se encontraron patrones que relacionaban las variables con las tendencias que se tenían en los datos históricos.
- Se elaboró un modelo de predicción utilizando los datos disponibles de la empresa.
- Se elaboraron reportes en base a los datos obtenidos.

Realizar plan del proyecto

Se planeó un plan de proyecto para así poder manejar tiempos y realizar el proyecto de manera eficiente, se detalla en la siguiente tabla III.

TABLA III
PLAN DE PROYECTO

FASE	TIEMPO PROMEDIO	RIESGOS
Comprensión del negocio	2 semanas	Dificultades para situar los objetivos comerciales en el contexto adecuado. Restricciones en cuanto a la disponibilidad de tiempo de los participantes.
Comprensión de los datos	2 semanas	Complicaciones relacionadas con la comprensión de los datos.
Preparación de los datos	2 semanas	Desafíos vinculados a la depuración y fusión de los datos.
Modelado	5 semanas	Falta de capacidad para identificar un modelo apropiado.
Evaluación	3 semanas	Dificultades para llevar a cabo la implementación de los resultados.
Despliegue	2 semanas	Inconvenientes al poner en marcha el modelo o al elaborar el informe final.

FASE #2: Comprensión de los datos

En esta fase, el objetivo principal fue comprender los datos de la empresa. Algunas actividades clave en esta fase incluyeron:

Recolectar los datos iniciales

Como parte de la metodología utilizada, la fuente inicial provino de los datos históricos que la empresa poseía. Todo eso reflejaba la producción de la empresa y el nivel de servicio que brindaba. De esa forma, con la entrevista ya realizada (Ver Anexo 1), se buscó comprender a profundidad los datos y las variables.

Descripción de los datos

La empresa nos proporcionó un conjunto de datos en formato xlsx, un tipo de documento asociado a Microsoft Excel. Cada registro en el archivo xlsx se organizaba en columnas, facilitando así el procesamiento e intercambio de datos. El conjunto de datos en cuestión constaba de 10 columnas, las cuales representaban variables cruciales para el área comercial del negocio, abarcando un total de 533 registros hasta el momento.

Con el fin de obtener una visualización más comprensible y manejable de los datos, se recurrió a la herramienta Google Colab, una plataforma de computación en la nube basada en Jupyter Notebook, como se mostró en la figura 1. Google Colab no solo fue la herramienta para visualizar los datos, sino que también se utilizó posteriormente como entorno de trabajo para el procesamiento y análisis de estos.

FECHA	FERTILIZANTE	ENTRADA/SALIDA	CANTIDAD	INVENTARIO	NOTA	TIEMPO_PROD	COND_CLIM	TIPO_CULTIVO	PH
2022-12-13	FIGAL KILLIFER	Entrada	268	268.0	NaN	29.0	27.0	General (multiuso)	NaN
2022-12-13	SULFATO DE COBRE	Entrada	40	40.0	NaN	29.0	27.0	Frutales, hortalizas	3.7
2022-12-13	FIGAL HIERRO	Entrada	30	30.0	NaN	30.0	27.0	Hortalizas, plantas verdes	3.2
2022-12-13	FOSFITO DE COBRE	Entrada	40	40.0	NaN	29.0	27.0	Frutales, hortalizas	3.5
2022-12-13	FOSFITO DE POTASIO	Entrada	140	140.0	NaN	29.0	27.0	Frutales, hortalizas	4.2
2022-12-13	RESISTMAS	Entrada	51	51.0	NaN	29.0	27.0	General (multiuso)	NaN

Fig. 01. Visualización de los datos en Google Colab

El conjunto de datos proporcionado por la empresa comprendía un grupo de variables importantes y significativas dentro del área comercial. Con el objetivo de lograr una comprensión exhaustiva de cada variable en particular y mitigar posibles complicaciones durante la fase de modelado, se procedió a definir minuciosamente cada una de ellas, evaluando sus respectivos tipos de datos, como se detalla en la tabla IV.

La definición clara de las variables y la identificación de sus tipos de datos representaron un paso fundamental en el procesamiento y análisis de los conjuntos de datos. Esto sentó las bases necesarias para el desarrollo de modelos predictivos y la extracción de información relevante. Esta etapa preliminar fue esencial para interpretar adecuadamente los datos, lo que facilitó la toma de decisiones informadas en el contexto comercial de la empresa.

TABLA IV
DICCIONARIO DE DATOS

Variable	Tipo	Escala de Medida	Descripción
FECHA	Cuantitativa Discreta	Razón	Fecha de registro de la entrada y salida de fertilizante.
FERTILIZANTE	Cualitativa Nominal	Nominal	Nombre del fertilizante.
ENTRADA/SALIDA	Cualitativa Nominal	Nominal	Indica si el registro corresponde a una entrada o salida de fertilizante.
CANTIDAD	Cuantitativa Continua	Razón	Cantidad de fertilizante registrada.
INVENTARIO	Cuantitativa Continua	Razón	Cantidad de fertilizante disponible en inventario.
NOTA	Cualitativa Nominal	Nominal	Notas adicionales sobre el registro.
TIEMPO_PROD	Cuantitativa Continua	Razón	Tiempo de producción relacionado con el fertilizante registrado.
COND_CLIM	Cualitativa Nominal	Nominal	Condiciones climáticas promedio que se dio para la producción del fertilizante.
TIPO_CULTIVO	Cualitativa Nominal	Nominal	Tipo de cultivo al que se aplica el fertilizante.
PH	Cuantitativa Continua	Razón	Nivel de pH del fertilizante.

Exploración de los datos

Se utilizó la herramienta de visualización de Google, Colab, para explorar el conjunto de datos desde diferentes perspectivas, lo cual permitió profundizar en el conocimiento de este. Cabe resaltar que se verificaron los nombres de las columnas para asegurar que no existieran complicaciones futuras al trabajar con ellas.

Se destacó que la variable FECHA al ser de tipo “datetime64”, se separó por Año, Mes y Día para poder trabajarla adecuadamente durante la exploración de los datos. Asimismo, se validó que la variable NOTA no era una variable de gran importancia ni aportaba al desarrollo del modelo, por lo que se decidió eliminar la variable en la fase de limpieza de datos.

Después de estos ajustes, no se encontraron mayores inconvenientes, y se verificó que todos los campos estaban correctamente nombrados. Este proceso tuvo como fin evidenciar los datos que se trabajarían en el proyecto.

Posteriormente, para un mejor análisis, se agruparon las variables en dos subgrupos: el primero como columnNumeric (8) y el segundo como columnString (4). Esta división se realizó para asegurar el procesamiento correcto de los datos, como se detalla en el anexo 2.

Después de haber separado los tipos de datos, se inició el análisis de cada subgrupo creado. En el subgrupo columnNumeric, se determinó que los datos presentaban problemas debido a la presencia de datos atípicos o vacíos. Para un análisis más detallado, se utilizó la función describe() de la librería pandas, la cual describe estadísticamente los datos de cada variable. Este análisis permitió obtener un conteo de los datos, la cantidad de valores únicos, la moda, entre otros aspectos relevantes. Este proceso fue fundamental para la limpieza de los datos, lo que facilitará la posterior construcción del modelo, como se detalla en la figura 2.

```
datos[columnNumeric].describe()
```

	CANTIDAD	INVENTARIO	TIEMPO_PROD	COND_CLIM	PH	YEAR	MONTH	DAY
count	533.000000	530.000000	249.000000	532.000000	447.000000	533.000000	533.000000	533.000000
mean	284.709193	581.018868	28.923695	27.037594	7.990336	2023.268293	5.988743	15.540338
std	638.794818	1145.427878	0.446989	1.468597	2.518848	0.532115	3.592477	8.635952
min	1.000000	0.000000	28.000000	23.000000	1.000000	2022.000000	1.000000	1.000000
25%	19.000000	25.000000	29.000000	26.000000	7.000000	2023.000000	3.000000	8.000000
50%	40.000000	80.000000	29.000000	27.000000	9.500000	2023.000000	6.000000	15.000000
75%	200.000000	480.000000	29.000000	28.000000	9.710000	2024.000000	9.000000	23.000000
max	4800.000000	6200.000000	30.000000	33.000000	12.400000	2024.000000	12.000000	31.000000

Fig. 2. Descripción de las variables numéricas en Google Colab

En los resultados de las variables numéricas, se observó que algunas variables presentaban datos faltantes o valores nulos. Esto se evidenció en el conteo (count) de los valores para cada variable, donde se mostró un número menor al total de registros del conjunto de datos. La presencia de datos faltantes representaba un desafío en el procesamiento y modelado, por lo que fue necesario abordar esta situación mediante técnicas adecuadas de imputación o eliminación de registros.

Además, los resultados mostraron la media (mean) y la desviación estándar (std) para cada variable numérica. La media representaba el valor promedio de la distribución de los datos, mientras que la desviación estándar medía la dispersión de los valores individuales respecto al promedio.

También, en el subgrupo columnString, se determinó que los datos presentaban problemas debido a la presencia de datos atípicos o vacíos. Para un análisis más detallado, se utilizó la función describe()

```
datos[columnString].describe(include='o')
```

	FERTILIZANTE	TIPO_CULTIVO	ENTRADA/SALIDA
count	533	533	533
unique	29	7	2
top	ACIDO HUMICO	General (multiuso)	Salida
freq	68	215	348

Fig. 3. Descripción de las variables categóricas en Google Colab

De manera similar al análisis anterior, en los resultados obtenidos de columnString se pudo identificar la presencia de datos faltantes o valores nulos. Esta situación se evidenció en el conteo (count) de los valores para cada variable categórica, donde se mostró un número menor al total de registros del conjunto de datos.

Además de identificar los datos faltantes, el análisis descriptivo de las variables categóricas proporcionó información valiosa, como el conteo de la cantidad de datos y el valor único de cada variable. Asimismo, se mostró el valor más frecuente o moda para cada variable categórica, lo que indicaba cuál era el dato con mayor frecuencia.

Verificar la calidad de los datos

En esta etapa, se evidenció la presencia de datos faltantes en las variables numéricas y categóricas. Se determinó que los datos presentados se encontraban en un 70% limpios, mientras que el porcentaje restante presentaba datos faltantes. Se identificó que no existían datos faltantes en la variable numérica analizada (CANTIDAD).

Se concluyó que la calidad de los datos era la adecuada. De esta manera, se permitió trabajar con los datos originales sin ninguna modificación que pudiera influir en los resultados finales de esta fase. Es importante destacar que este paso preliminar de análisis de calidad de datos es crucial para el proyecto de análisis de datos y desarrollo de modelos predictivos.

FASE #3: Preparación de los datos

Seleccionar los datos

Tras reflexionar sobre los hallazgos obtenidos durante la fase anterior y evaluar el comportamiento de las variables, se destacó que todo el análisis y trabajo de comprensión de los datos se realizó gracias a la colaboración de la empresa Nutrition Vegetable Corporation FINGAL SAC. Se determinó trabajar con todas las variables en columnNumeric y columnString, con excepción de la variable NOTA. Esta decisión se fundamentó en el análisis previo, donde se concluyó que dicha variable no representaba una contribución sustancial en el contexto del problema abordado.

Limpieza de datos

En esta sección, se comenzó a trabajar con los registros faltantes (missing data) en ambos subgrupos. Se inició el proceso de imputación de los registros vacíos, utilizando la mediana como factor de reemplazo para los datos numéricos y, en el caso de los datos categóricos, se empleó el valor más frecuente. Esta elección se fundamentó en la forma de distribución de los valores en las variables, evitando que los datos atípicos pudieran representar un problema potencial. Al utilizar la mediana como medida de tendencia central, se mitigó el impacto que estos valores atípicos podrían tener, ya que la mediana se basa principalmente en los valores centrales de la distribución. Por otro lado, el valor más frecuente se utilizó para dar valor al dato

con mayor frecuencia en la distribución. El proceso de imputación de los registros faltantes se puede verificar en la anexo 3.

En el anexo 3, se visualizó la forma en que se trabajó la imputación de los dos subgrupos, utilizando la librería SimpleImputer de Python. Esta librería permitió generar un objeto imputador de datos, el cual recibió como parámetros la expresión de valor faltante (NaN) en la variable, así como la selección de la forma de resolverlo. Se utilizaron la mediana y el valor más frecuente como estrategias de imputación debido a su robustez frente a valores atípicos.

Posteriormente, se aplicó el imputador generado sobre el subconjunto de datos que contenía las variables numéricas y categóricas (almacenadas en los dataframes correspondientes `data_imp_num` y `data_imp_cat`). Finalmente, se comprobó que la imputación se realizó correctamente, verificando que los registros faltantes ya no aparecían como se observa en la figura 4.

<p>(a) <code>imputed_data_univ_num.isnull().sum()</code></p> <pre>CANTIDAD 0 INVENTARIO 0 TIEMPO_PROD 0 COND_CLIM 0 PH 0 YEAR 0 MONTH 0 DAY 0 dtype: int64</pre>	<p>(b) <code>imputed_data_univ_cat.isnull().sum()</code></p> <pre>FERTILIZANTE 0 TIPO_CULTIVO 0 ENTRADA/SALIDA 0 dtype: int64</pre>
---	--

Fig. 4. Resultados de la verificación de datos después de la imputación de datos (a) Resultados de las variables numéricas, (b) Resultados de las variables categóricas.

Estructurar los datos

En esta sección, se inició el proceso de recodificación de los datos categóricos a valores numéricos que podrían ser utilizados por el modelo predictivo en desarrollo. Tras un análisis previo de los datos, se identificaron las variables categóricas que necesitaban ser recodificadas: `TIPO_CULTIVO`, `FERTILIZANTE` y `ENTRADA/SALIDA`. Una vez determinadas estas variables, se utilizó la librería LabelEncoder de Python para codificar cada registro, asignando un valor numérico a cada uno de los valores únicos en dichas variables. Los resultados de esta codificación se almacenaron en la variable `desarrll_imp`, como se mostró en la figura 5.

```

import pandas as pd
from sklearn.preprocessing import LabelEncoder

# Columnas categoricas que vamos a cambiar a numericas
columnTC = 'TIPO_CULTIVO'
columnF = 'FERTILIZANTE'

# Convertir todos los valores de la columna en cadenas
imputed_data_univ_cat[columnTC] = imputed_data_univ_cat[columnTC].astype(str)

# Ahora puedes intentar codificar la columna problemática
label_encoderTC = LabelEncoder()
imputed_data_univ_cat[columnTC] = label_encoderTC.fit_transform(imputed_data_univ_cat[columnTC])

# Ahora puedes intentar codificar la columna problemática
label_encoderF = LabelEncoder()
imputed_data_univ_cat[columnF] = label_encoderF.fit_transform(imputed_data_univ_cat[columnF])

# Convertir 'Entrada/Salida' a variable binaria (0 para Salida, 1 para Entrada)
imputed_data_univ_cat['ENTRADA/SALIDA'] = (imputed_data_univ_cat['ENTRADA/SALIDA'] != 'SALIDAD').astype(int)

# Verificar el resultado
imputed_data_univ_cat.head()

```

Fig. 5. Recodificación de las variables categóricas en Google Colab

Integración de datos

En esta etapa, se unificó el conjunto de datos inicial, ya limpio y sin datos atípicos o vacíos. El objetivo de este paso fue consolidar un único conjunto de datos integrado y preparado para la construcción del modelo predictivo en la siguiente etapa. Dado que el orden de los registros no se alteró en los procesos previos, la integración de los subconjuntos de datos se realizó sin complicaciones, ver figura 6.

```

# Consolidamos los subset!
desarrll_imp = pd.concat([imputed_data_univ_num,imputed_data_univ_cat],axis=1)

desarrll_imp.head()

```

	CANTIDAD	INVENTARIO	TIEMPO_PROD	COND_CLIM	PH	YEAR	MONTH	DAY	FERTILIZANTE	TIPO_CULTIVO	ENTRADA/SALIDA
0	268.0	268.0	29.0	27.0	9.71	2022.0	12.0	13.0	11	3	1
1	40.0	40.0	29.0	27.0	3.70	2022.0	12.0	13.0	25	2	1
2	30.0	30.0	30.0	27.0	3.20	2022.0	12.0	13.0	10	6	1
3	40.0	40.0	29.0	27.0	3.50	2022.0	12.0	13.0	19	2	1
4	140.0	140.0	29.0	27.0	4.20	2022.0	12.0	13.0	20	2	1

Fig. 6. Resultado de la recodificación de las variables categóricas en Google Colab

Dar formato a los datos

En esta etapa, se dividió el conjunto de datos en dos segmentos: uno para entrenamiento y otro para prueba. Los datos de entrenamiento se utilizaron para ajustar el modelo predictivo, mientras que los datos de prueba se reservaron para evaluar la calidad del modelo generado a partir del entrenamiento. En este caso, se decidió asignar el 70% de los datos al conjunto de entrenamiento y el 30% restante al conjunto de prueba, como se detalló en el figura 7.

```
# Creación de la data de train y la data de test
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X,
                                                    Y,
                                                    test_size=0.30,
                                                    random_state=100)

X_train.shape, X_test.shape, y_train.shape, y_test.shape
((182, 10), (78, 10), (182,), (78,))
```

Fig. 7. Determinar data de entrenamiento y prueba en Google Colab

Por último, es importante destacar que los algoritmos seleccionados para este proyecto no exigieron un formato particular de los datos, ya que fueron capaces de manejar la información sin restricciones de formato. Teniendo en cuenta este aspecto, se procedió a la fase de modelado.

FASE #4: Modelado

Seleccionar técnica de modelamiento

Dado que en la fase 2 de la metodología se realizó la caracterización detallada de las variables, se procedió a una investigación exhaustiva sobre las técnicas de modelado disponibles. Esto permitió identificar aquellas con mayor potencial y pertinencia para este proyecto, seleccionando las que cumplen con los criterios más adecuados para abordar el problema planteado.

TABLA V
CARACTERÍSTICAS COMPARATIVAS DE LOS ALGORITMOS EMPLEADOS

Técnica	Descripción	Ventaja	Desventaja	Flexibilidad y adaptabilidad
Regresión Lineal	Modelo predictivo que encuentra una relación lineal entre variables dependientes e independientes.	Fácil de interpretar e implementar para relaciones lineales.	Ineficiente para relaciones no lineales o complejas.	Limitada a relaciones lineales, poca flexibilidad en datos no lineales.
Árboles de Decisión	Modelo jerárquico que realiza predicciones a través de decisiones secuenciales.	Interpretables, flexibles, y manejan variables continuas y categóricas.	Alta susceptibilidad al sobreajuste sin poda adecuada.	Muy flexible para relaciones complejas, pero requiere poda y ajustes.
Random Forest	Conjunto de árboles de decisión que mejora la precisión combinando predicciones de varios árboles.	Reduce sobreajuste y mejora precisión respecto a un solo árbol.	Alta demanda computacional y difícil interpretación.	Bastante flexible, pero requiere más recursos computacionales y procesamiento.
(ANN) Redes Neuronales Artificiales	Modelo inspirado en el cerebro que captura relaciones complejas a través de capas de procesamiento.	Captura relaciones no lineales y complejas en grandes volúmenes de datos.	Alta demanda computacional y difícil interpretación.	Muy flexible y adaptable a datos complejos, pero requiere muchos recursos.

SVM (Máquinas de Vectores de Soporte)	Método que encuentra el hiperplano óptimo para clasificar o predecir valores continuos.	Excelente para datos con alta dimensionalidad y relaciones complejas.	Difícil de interpretar y ajustar con alta complejidad.	Flexible para datos no lineales, pero complicado de ajustar correctamente.
Análisis de Componentes Principales (PCA)	Técnica de reducción de dimensionalidad que transforma variables en un conjunto de variables no correlacionadas.	Reduce dimensionalidad manteniendo la mayor parte de la varianza.	Puede perder interpretabilidad al transformar las variables originales.	Adaptable para la reducción de dimensionalidad, pero puede perder información importante si no se ajusta correctamente.

Luego del análisis de las técnicas, se mencionó que la Regresión Lineal era adecuada para capturar relaciones lineales entre las variables, siendo eficiente y fácilmente interpretable en escenarios donde la correlación entre los datos era clara. Por otro lado, se destacó que las Redes Neuronales Artificiales (ANN) y las Máquinas de Vectores de Soporte (SVM) eran especialmente valiosas para manejar relaciones no lineales y complejas, lo cual resultaba fundamental cuando se trabajaba con grandes volúmenes de datos o interacciones intrincadas, como en la producción industrial. En cambio, se observó que otras técnicas, como los Árboles de Decisión y Random Forest, aunque útiles para clasificación y otros contextos, tendían a ser propensas al sobreajuste o requerían un procesamiento excesivo sin ofrecer mejoras significativas en precisión para este tipo de problema predictivo. Además, se concluyó que el Análisis de Componentes Principales (PCA) no sería óptimo en este caso, ya que podría sacrificar la interpretabilidad y no era una técnica predictiva por sí misma.

Llegando a definir, que la Regresión Lineal, las Redes Neuronales Artificiales (ANN) y las SVM por su capacidad para manejar tanto relaciones lineales como no lineales, siendo las más adecuadas para predecir la producción de fertilizantes. Otras técnicas, como los Árboles de

Decisión y PCA, se descartaron por sus limitaciones en precisión e interpretabilidad para este caso.

Generar plan de prueba

Detallamos cómo se midió el rendimiento de los modelos implementados, utilizando las siguientes métricas de evaluación:

- **TP (Positivos Verdaderos):** Cantidad de casos positivos correctamente predichos por el modelo.
- **TN (Negativos Verdaderos):** Cantidad de casos negativos correctamente predichos.
- **FP (Positivos Falsos):** Cantidad de casos negativos que fueron incorrectamente predichos como positivos.
- **FN (Negativos Falsos):** Cantidad de casos positivos que fueron incorrectamente predichos como negativos.

Precisión: Mide la proporción de verdaderos positivos frente al total de valores predichos como positivos.

$$precisión = \frac{TP}{TP + FP}$$

Accuracy: La exactitud mide el porcentaje de casos que el modelo ha acertado.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Recall: La métrica de exhaustividad nos informa sobre la cantidad que el modelo es capaz de identificar.

$$recall = \frac{TP}{TP + FN}$$

Construir el modelo

Al comparar los algoritmos, se eligió el modelo más eficiente, considerando en primera instancia los resultados obtenidos en el análisis de datos. Se recalibró el modelo para resaltar la efectividad, ver figura 8.

```

# *****
# **  IMPORTAMOS LIBRERIAS  **
# *****
import tensorflow as tf # Libreria de IA
import matplotlib.pyplot as plt # Para hacer graficas
import numpy as np
from sklearn.metrics import mean_squared_error, r2_score, accuracy_score, precision_score, recall_score, f1_score, confusion_matrix, roc_auc_score

# *****
# **  Declararemos arreglo de numeros  **
# *****
# X = Variables independientes
# Y = Variables dependientes
X = desarrll_imp.drop('CANTIDAD', axis=1)
Y = desarrll_imp['CANTIDAD']

# Convertir Y a una representación binaria para clasificación (si es necesario)
# Aquí asumo que Y ya está en formato binario (0 o 1). Si no es así, deberás convertirlo.
# *****
# **  Usaremos Keras para las redes neuronales  **
# *****
modelo = tf.keras.Sequential([
    tf.keras.layers.Dense(units=128, input_shape=[10], activation='relu'), # Más unidades en la primera capa oculta
    tf.keras.layers.Dropout(0.3), # Dropout del 30%
    tf.keras.layers.Dense(units=64, activation='relu'), # Más unidades en la segunda capa oculta
    tf.keras.layers.Dropout(0.3), # Dropout del 30%
    tf.keras.layers.Dense(units=32, activation='relu'), # Más unidades en una tercera capa oculta, si es necesario
    tf.keras.layers.Dropout(0.3), # Dropout del 30%
    tf.keras.layers.Dense(units=1) # Capa de salida
])
# *****
# **  Compilaremos el modelo  **
# *****
modelo.compile(
    optimizer=tf.keras.optimizers.Adam(learning_rate=0.0005), # Reducir la tasa de aprendizaje
    loss='binary_crossentropy',
    metrics=['accuracy']
)

print("*****")
print("Comenzando entrenamiento...")
print("*****")
# fit = valores de entrada y salida dando vueltas (1500) + vueltas mejor.
historial = modelo.fit(X, Y, epochs=1500, verbose=False, validation_split=0.2) # Se agrega validación para monitorear la precisión durante el entrenamiento
print("Modelo entrenado!")
print("*****")
print(" ")

```

Fig. 8. Creación del modelo de predicción

Evaluar el modelo

Describimos a continuación los resultados obtenidos del modelo seleccionado, que fueron:

TABLA VI
RESULTADO DE LAS METRICAS DE EVALUACION DEL ALGORITMO

Modelo	Accuracy	Precision	Recall
Red Neuronal	0.53	0.53	0.53

FASE #5: Evaluación

Evaluar los resultados

Con base en los resultados obtenidos tras la evaluación del modelo implementado, se concluyó que el modelo de red neuronal fue el que mejor se ajustó a la información y a la realidad presentada en este proyecto. Este modelo alcanzó una precisión y una exactitud (accuracy) del 53%, superando a los demás modelos que se desarrollaron.

Revisar el proceso

Para validar la efectividad del modelo y su precisión, se crearon escenarios aleatorios utilizando los valores de cada variable de los datos proporcionados por la organización. Estos escenarios sirvieron para evaluar la capacidad del modelo de hacer predicciones precisas en diferentes condiciones. Los resultados y el procedimiento de verificación se presentan en la figura 9, evidenciando así la robustez y confiabilidad del modelo implementado.

```

✓ datos = pedir_datos()
↕ Ingrese el valor de FERTILIZANTE: 4
  Ingrese el valor de TIPO_CULTIVO: 1
  Ingrese el valor de ENTRADA/SALIDA: 1
  Ingrese el valor de INVENTARIO: 300
  Ingrese el valor de TIEMPO_PROD: 29
  Ingrese el valor de COND_CLIM: 18
  Ingrese el valor de PH: 6
  Ingrese el valor de YEAR: 2024
  Ingrese el valor de MES: 7
  Ingrese el valor de DIA: 9

✓ [6] print(datos)
↕ {'FERTILIZANTE': '4', 'TIPO_CULTIVO': '1', 'ENTRADA_SALIDA': '1', 'INVENTARIO': '300', 'TIEMPO_PROD': '29', 'COND_CLIM': '18', 'PH': '6', 'YEAR': '2024', 'MES': '7', 'DIA': '9'}

✓ [9] new = new.astype(float)
  resultado = modelo.predict(new)

  print("*****")
  print("El resultado es " + str(resultado[0][0]) + " CANTIDAD!")
  print("*****")

↕ 1/1 [=====] - 1s 757ms/step
  *****
  El resultado es 1405.6681 CANTIDAD!
  *****

```

Fig. 9. Evaluación del modelo

Determinar próximos pasos

Teniendo ya el modelo ajustó con los objetivos del proyecto, se procedió a desarrollar la arquitectura necesaria para su implementación y funcionamiento en la web. Este paso resultó fundamental para integrar los resultados del modelo final en el proceso comercial de la organización, facilitando la generación de informes accesibles para el usuario final.

El despliegue del modelo se organizó en varias fases clave: primero, se creó el archivo ejecutable del modelo entrenado; luego, se desarrolló y subió la API a la web; y, por último, se diseñó la interfaz final para el usuario. La arquitectura utilizada y su implementación detallada se ilustraron en la anexo 3. Este proceso permitió consolidar los conocimientos adquiridos y aplicar de manera práctica las habilidades desarrolladas en el manejo de modelos predictivos y su integración en entornos comerciales.

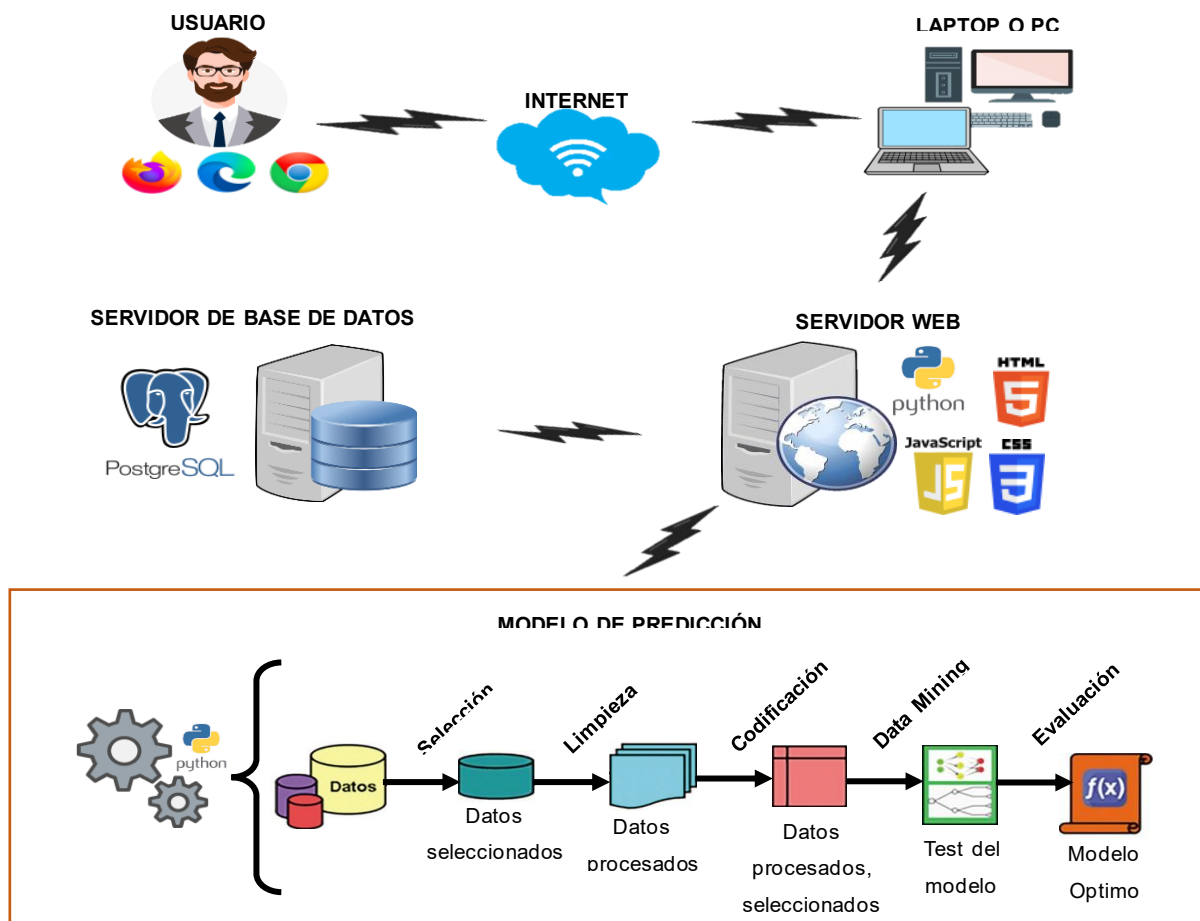


Fig. 10. Arquitectura del producto final

El despliegue del modelo predictivo se estructuró en tres fases principales. La primera fase involucró la elaboración, entrenamiento y evaluación del modelo, culminando en su exportación como un archivo ejecutable, que sirvió como base para todo el sistema. En la segunda fase, se construyó y cargó al servidor web la API, que actuó como intermediaria entre el modelo y la interfaz final del usuario, gestionando las solicitudes y generando las predicciones. La tercera y última fase consistió en desarrollar la interfaz de usuario y conectarla directamente a la API, permitiendo que el usuario realizara solicitudes y recibiera las predicciones y reportes de manera eficiente.

FASE #6: Despliegue

Desplegar del plan

El despliegue del modelo predictivo se organizó en tres fases principales. La primera fase incluyó la creación, entrenamiento y evaluación del modelo, finalizando con su exportación como un archivo ejecutable, que constituyó la base de todo el sistema. En la segunda fase, se desarrolló y subió la API al servidor web, actuando como intermediaria entre el modelo y la interfaz de usuario, gestionando las solicitudes y generando las predicciones. La tercera y última fase consistió en diseñar la interfaz del usuario y conectarla directamente a la API, lo que permitió que el usuario realizara solicitudes y recibiera las predicciones y reportes de manera eficiente.

Esta versión facilitó la conexión de la API con la interfaz de usuario final. El método GET se vinculó directamente a la interfaz, lo que permitió una comunicación fluida entre el cliente y el sistema. La construcción de la API con ambos métodos fundamentales se detalló en el anexo 4.

Se desarrollaron funciones específicas para listar y gestionar los datos, además de otras funciones auxiliares para optimizar la interacción con el modelo. Estas funciones se integraron en la API, la cual se conectó al archivo .h5 del modelo entrenado. Esta conexión permitió que el modelo se activara y utilizara dentro de diferentes endpoints de la API, facilitando su integración y funcionalidad en el sistema.

Monitoreo y mantenimiento

Se realizó la verificación con Postman para asegurarse de que los métodos GET y POST de la API se habían subido al servidor correctamente. Esta herramienta permitió probar las solicitudes y confirmar que la API respondía adecuadamente a las peticiones. (ver anexo 5)

Desarrollar el reporte final

Para el desarrollo de la interfaz final, se utilizó el framework Flask y el lenguaje de programación Python, complementados con HTML para la estructura de las páginas. Se empleó CSS para el diseño visual y la biblioteca Bootstrap 4 para asegurar un diseño responsive y moderno. Además, se incorporó JavaScript para añadir funcionalidades dinámicas a la interfaz.

Se prestó especial atención a que la interfaz fuera simple, visualmente agradable e intuitiva para el usuario. Se automatizó el proceso de entrada de datos tanto como fue posible, reduciendo la necesidad de digitación manual y mejorando la experiencia del usuario final. La integración fluida de estos elementos facilitó una interacción eficiente y amigable con el sistema. Se puede consultar el Anexo 2 para ver las imágenes del sistema desplegado.

Revisión del proyecto

El proyecto culminó con éxito tras comparar distintos modelos predictivos y seleccionar el que mejor se adaptó a la realidad de la compañía. Se implementó y puso en producción este modelo, integrándolo con una interfaz web para facilitar su uso por parte del usuario final. Esta interfaz permite cargar datos actuales y predecir la posibilidad de deserción de clientes con mayor eficacia y eficiencia en comparación con los análisis manuales y estadísticos realizados por el área de Inteligencia Comercial de la compañía.

La interfaz web también proporciona la capacidad de analizar a los clientes tanto de manera específica como global, destacando aquellos con una alta probabilidad de deserción según los criterios de la organización. Además, informa sobre la importancia relativa de las variables en la predicción, ayudando al usuario final a identificar los factores más influyentes en la deserción

de clientes. Esta funcionalidad permite una toma de decisiones más informada y estratégica, centrada en retener a los clientes más propensos a abandonar.

En base a los objetivos del proyecto

Objetivo 1: Identificar las técnicas más adecuada para predecir la producción del fertilizante foliar basado en la caracterización de las variables clave que influyen en su producción

En la fase 4 (Modelado), específicamente durante la actividad de selección de técnicas de modelamiento, se evaluaron las posibles metodologías con base en el análisis previo de las variables clave, que se realizó en la fase 2 de la metodología. Estas variables fueron determinantes para identificar y comparar las técnicas de inteligencia artificial (IA) más adecuadas, considerando aspectos como su descripción, ventajas, limitaciones y otros factores críticos. De esta forma, se seleccionaron los métodos más prometedores para el desarrollo del modelo final.

Durante este proceso de selección, se compararon los tres modelos predictivos, utilizando métricas clave de rendimiento para determinar su idoneidad en la predicción de la producción del fertilizante foliar. A continuación, se presentó un resumen de los resultados obtenidos:

**TABLA VII
RESULTADO DE LAS METRICAS DE EVALUACION DEL ALGORITMO**

Modelo	Accuracy	Precision	Recall
Lineal	0.023	0.023	0.023
Red Neuronal	0.523	0.523	0.523
SVM	0.628	1.000	0.341

Tras analizar los resultados obtenidos al evaluar los diferentes modelos implementados, se determinó que la red neuronal fue la técnica más adecuada para este problema específico. Aunque inicialmente mostró un desempeño moderado en términos de precisión, su capacidad para capturar patrones complejos y no lineales en los datos resultó ser superior en comparación

con los otros modelos evaluados. La red neuronal alcanzó un Accuracy de 0.523, una precisión de 0.523 y un Recall de 0.523. Estos valores reflejan una capacidad robusta para predecir y clasificar correctamente la producción del fertilizante foliar.

La elección de la red neuronal se fundamentó en su capacidad para adaptarse y aprender de la estructura subyacente de los datos, lo que la hace especialmente adecuada para problemas donde la relación entre las variables predictoras y la variable objetivo puede ser compleja o no lineal. Además, su flexibilidad y la posibilidad de ajustar diferentes parámetros permitieron optimizar su rendimiento durante el proceso de modelado. En conclusión, la red neuronal se destacó como la mejor opción debido a su capacidad para ofrecer predicciones precisas y confiables en la producción del fertilizante foliar, cumpliendo con los objetivos de este estudio.

Objetivo 2: Desarrollar un modelo de predicción idóneo para la producción del fertilizante foliar utilizando la técnica de inteligencia artificial seleccionada.

En la fase 4 (Modelado), durante la actividad de construcción del modelo, se procedió a desarrollar un modelo de predicción de la producción del fertilizante foliar utilizando la técnica seleccionada: una red neuronal. Esta elección se basó en los resultados previos que indicaron la superioridad de la red neuronal respecto a otros modelos evaluados. La implementación del modelo se realizó utilizando TensorFlow y Keras, frameworks especializados en el desarrollo de redes neuronales en Python. Esta descripción detallada se observa en la figura 9, donde se presentan las especificaciones técnicas y el proceso de implementación de la red neuronal.

En la fase de evaluación del modelo para predecir la producción del fertilizante foliar, se observó que el modelo de red neuronal alcanzó una precisión del 52.3%. Este valor destacó la capacidad del modelo para realizar predicciones acertadas sobre la cantidad de producción de fertilizante foliar, proporcionando así una herramienta confiable para optimizar los procesos de planificación y producción en el sector agrícola.

Objetivo 3: Diseñar una interfaz web para facilitar la interacción del usuario en la empresa de producción de fertilizante.

En la fase 6 (Despliegue), durante la actividad de monitoreo y mantenimiento, se diseñó una interfaz web intuitiva y funcional para facilitar la interacción del usuario final. En el anexo 5, se presentaron las interfaces que fueron desarrolladas cuidadosamente para cumplir con los requisitos y expectativas del usuario. Esto incluyó la presentación clara de los resultados de predicción de producción de fertilizante foliar, permitiendo al usuario cargar datos y obtener fácilmente los pronósticos necesarios. La interfaz se diseñó con un enfoque en la usabilidad y la accesibilidad, asegurando que los usuarios pudieran interpretar y actuar sobre la información proporcionada de manera efectiva y eficiente.

Objetivo 4: Asegurar la calidad del sistema a través de la opinión de los especialistas, siguiendo la normativa ISO 25010.

Este objetivo se evaluó en la fase 6 (Despliegue), durante la actividad de monitoreo y mantenimiento del sistema. Dado que este proceso se implementó una vez que el sistema fue desplegado, se desarrolló una ficha de evaluación que se utilizó para asegurar la calidad del sistema de predicción, la cual se basó en las directrices de la normativa ISO 25010.

La evaluación fue realizada por especialistas seleccionados en el área, quienes aportaron su experiencia y conocimientos para identificar áreas de mejora y garantizar que el sistema cumpliera con los estándares de calidad establecidos. La tabla que detalló los criterios de evaluación y los indicadores específicos considerados por los especialistas se encuentra en el Anexo 7.

Discusión

En el estudio de Muñoz, se abordó la gestión agrícola mediante el modelo XGBoost, prediciendo con precisión el peso de racimos de banano. Sin embargo, mi investigación amplía el enfoque al integrar redes neuronales para optimizar la producción de fertilizantes en un entorno industrial, abordando la prevención de la sobreproducción y mejorando la eficiencia operativa, aspectos no considerados en el trabajo de Muñoz.

Ordoñez empleó regresión lineal múltiple y CRISP-DM para predecir la producción lechera. Aunque ambos coincidimos en la metodología, mi investigación utiliza redes neuronales, que

capturan patrones más complejos en la producción de fertilizantes. Además de mejorar la predicción, mi trabajo se enfoca en la reducción de pérdidas y variabilidad en la producción, lo que le da un alcance más completo.

El estudio de Jiménez utilizó Machine Learning para mejorar la planificación de alimentos procesados, optimizando producción y costos. Mientras mi tesis se centra en buscar la eficiencia, de un proceso químico más complejo, donde la precisión en la predicción es clave para reducir pérdidas y evitar impactos ambientales. Mientras Jiménez se centra en ganancias, mi investigación prioriza la estabilidad en la producción y la satisfacción de la demanda, con una mirada hacia la sostenibilidad.

Porras y Ovalle emplearon Deep Learning para predecir la producción de cacao, alcanzando una precisión del 99.6%. Aunque mi modelo de red neuronal no alcanza ese nivel, es más flexible y adaptable al contexto de los fertilizantes foliares. Además de evitar la sobreproducción, mi enfoque integra factores como la estacionalidad de la demanda y la optimización de recursos, con un enfoque holístico que busca mejorar la calidad del producto y minimizar el impacto ambiental.

Conclusiones

En la identificación de las variables críticas que influyen en la producción de fertilizante foliar, se consideraron aspectos como la variación estacional de la demanda, la calidad del producto y las condiciones del mercado. Basándonos en la bibliografía sobre los principales algoritmos de clasificación, se seleccionaron varias técnicas, incluyendo regresión lineal, árboles de decisión, Random Forest, redes neuronales artificiales (RNA) y máquinas de soporte vectorial (SVM). Entre estos, la regresión lineal, las RNA y las SVM se destacaron como las más adecuadas para el desarrollo del modelo predictivo. Estas metodologías no solo permiten una selección precisa de los datos más relevantes y significativos, sino que también ofrecen una flexibilidad considerable para capturar patrones complejos en la producción, lo que resulta crucial para optimizar la toma de decisiones en este ámbito.

Se logró construir exitosamente los modelos utilizando las técnicas seleccionadas,

destacando que las redes neuronales artificiales representan la metodología más efectiva, alcanzando una capacidad de predicción del 53%. Este modelo, resultado de un exhaustivo proceso de entrenamiento y evaluación, se constituye como una herramienta valiosa para la empresa, capaz de ofrecer predicciones precisas que optimizan significativamente la planificación de la producción y la toma de decisiones operativas. La implementación de este modelo no solo mejora la capacidad de respuesta ante la variabilidad del mercado de fertilizantes, sino que también permite un análisis más profundo de la demanda de los clientes, alineándose con los objetivos estratégicos de la organización.

Se diseñó e implementó una interfaz web intuitiva y funcional que optimiza la interacción de los usuarios con el sistema predictivo de producción de fertilizante. Esta interfaz, desarrollada utilizando HTML, JavaScript y Python. Con esta herramienta, el usuario puede realizar predicciones de producción, como también acceder a gráficas visualizables y descargables que representan diferentes escenarios de producción, lo que mejora la toma de decisiones estratégicas y la planificación operativa en el contexto diario de la empresa.

Se realizó una rigurosa validación del sistema predictivo, en colaboración con un QA especializado, conforme a la norma ISO 25010. Este proceso confirmó su alta calidad en términos de fiabilidad, usabilidad y eficacia. La validación no solo asegura el cumplimiento de estándares internacionales, sino que también otorga a la empresa la confianza necesaria para implementar y utilizar el sistema en sus operaciones diarias, con la certeza de que contribuirá de manera efectiva a la optimización de la producción de fertilizantes. Así, el sistema se establece como una herramienta fundamental para la mejora continua, garantizando resultados sostenibles y de calidad en la producción.

Recomendaciones

Se recomienda establecer un proceso continuo de mejora y actualización del sistema predictivo implementado, incluyendo el seguimiento regular del rendimiento del modelo y la incorporación de nuevos datos y variables relevantes. Además, es esencial proporcionar capacitación continua al personal para maximizar el uso del sistema. Se sugiere explorar la adaptación de este sistema en otras áreas de la empresa o en empresas similares del sector, generando nuevas oportunidades de investigación y desarrollo. También se aconseja integrar el

sistema con tecnologías del Internet de las Cosas (IoT) para recopilar datos en tiempo real, mejorando la precisión de las predicciones y la capacidad de respuesta ante fluctuaciones en la demanda y condiciones de producción. Por último, es fundamental mantenerse actualizado con los avances en inteligencia artificial y minería de datos, considerando la incorporación de nuevas técnicas y algoritmos para mejorar aún más la precisión y eficacia del sistema predictivo.

Referencias

- [1] V. Fernández, T. Sotiropoulos, y P. Brown, “Fertilización foliar. Principios científicos y prácticas de campo en Asociación Internacional de la Industria de Fertilizantes (IFA),” Paris, Francia, pp. 49-82, 2015.
- [2] P. González, “Consecuencias ambientales de la aplicación de fertilizantes,” Asesoría Técnica Parlamentaria, vol. 1, no. 1, pp. 1-5, 2019.
- [3] A. G. Caballero Torres y M. Á. Suarez Villamizar, "Potencial para la implementación de la tecnología 4.0 en el sector del cacao en Santander," 2023.
- [4] MordorIntelligence “Análisis de tamaño y participación del mercado global de fertilizantes foliares: tendencias y pronósticos de crecimiento hasta 2028” [En línea]. Disponible en: <https://www.mordorintelligence.com/market-analysis/foliar-fertilizers>
- [5] Custom Market Insights, “Mercado mundial de fertilizantes foliares 2023-2032”, 2023. [En línea]. Disponible en: <https://www.custommarketinsights.com/report/foliar-fertilizer-market>
- [6] Centro Peruano de Estudios Sociales (CEPES), “Crisis de fertilizantes: los problemas en Perú para satisfacer su demanda”, 2022. [En línea]. Disponible en: <https://cepes.org.pe/2022/09/06/crisis-de-fertilizantes-los-problemas-en-peru-para-satisfacer-su-demanda>
- [7] M. Espinoza Campos, “Efecto de la fertilización foliar orgánica en la producción de papa (*Solanum tuberosum* L.) variedad Canchan INIA en condiciones edafoclimáticas de Huaripampa-Marañon-Huánuco 2020,” Tesis de Licenciatura, Ecuador, 2021.

- [8] J. Vásquez Sánchez, “Efecto de cuatro fertilizantes foliares en cultivo establecido de alfalfa (*Medicago sativa* L.) en el valle de Cajamarca-2019,” Tesis de Licenciatura, Perú, 2021.
- [9] Bemíagro, “La importancia de los fertilizantes foliar para las plantas”, 2018. [En línea]. Disponible en: <https://bemíagro.com/la-importancia-del-fertilizante-foliar-las-plantas/>
- [10] T. Santos y D. Aguilar Manjarrez, "Fertilización foliar, un respaldo importante en el rendimiento de los cultivos," *Terra Latinoamericana*, vol. 17, no. 3, pp. 247-255, jul.-sep. 1999. [En línea]. Disponible: <http://www.redalyc.org/articulo.oa?id=57317309>.
- [11] P. González, “Consecuencias ambientales de la aplicación de fertilizantes,” *Asesoría Técnica Parlamentaria*, vol. 1, no. 1, pp. 1-5, 2019.
- [12] C. Cevallos Carvajal, et al., “Efectividad de momentos y fuentes de aplicaciones foliares de calcio, boro y zinc en el rendimiento y rentabilidad del cacao nacional,” *Chilean Journal of Agricultural & Animal Sciences*, vol. 38, no. 3, pp. 304-317, 2022. <https://doi.org/10.29393/CHJAA38-29PVAR10029>
- [13] P. S. Muñoz Torres, “Diseño de un Modelo Predictivo basado en Algoritmos de Machine Learning para la estimación del Peso de Racimos de Banano, caso de Estudio Hacienda Bananera en Ecuador,” Tesis de Maestría, Ecuador, 2022.
- [14] C. Rodríguez Grandez, y F. Velásquez Carranza, “Modelo predictivo automatizado para la planificación de la producción de arroz en la Región San Martín,” Tesis de Licenciatura, Perú, 2018.
- [15] J. I. Ordoñez Merino, “Modelo predictivo aplicando algoritmos de Machine Learning para la producción lechera en la hacienda el prado, del Instituto Agropecuario Superior Andino (IASA), ” Tesis de Maestría, Ecuador, 2023
- [16] M. Afanador, et al. Diseño de un modelo de pronóstico de demanda basado en machine learning y un modelo multi-objetivo para planeación de la producción en una industria panificadora. [en línea]. Disponible en: <http://hdl.handle.net/10554/59123>.
- [17] R. Porras, C. Ovalle, “Prediction model for Peruvian cocoa production using Deep Learning and Long Short-Term Memory techniques,” *LACCEI*, pp. 1-9, 2023. <https://dx.doi.org/10.18687/LACCEI2023.1.1.314>
- [18] Organisation for Economic Co-operation and Development (OECD). “Manual de Frascati 2015: Guía para la recopilación y presentación de información sobre la investigación y el desarrollo experimental” OECD Publishing, Paris/FEYCT, Madrid, <https://doi.org/10.1787/9789264310681-es>



- [19] L. P. Vieira, et al. “Introducción a la Minería de Datos” 6ta ed., Brasil: E-papers Servicios Editoriales. 2009. [En línea]. Disponible: https://www.google.com.pe/books/edition/Introducci%C3%B3n_a_la_Miner%C3%ADa_de_Datos/jJEhHyESFsC?hl=es&gbpv=1
- [20] Norma ISO 25000 “ISO 25010”. [En línea]. Disponible en: <https://iso25000.com/index.php/normas-iso-25000/iso-25010>

Anexos

ANEXO N° 01. AUDIOS DE LA ENTREVISTA REALIZADA AL GERENTE GENERAL DE LA EMPRESA NUTRITION VEGETABLE CORPORATION FIGAL SAC

Mi unidad > TESIS > AUDIOS DE LA PRIMERA... ▾

1 seleccionados

Nombre ▾	Propietario	Última modificación ▾	Tamaño del i	
 18 oct, 4.12 p. m...aac	 yo	20:42 yo	29 MB	⋮
 18 oct, 3.58 p. m...aac	 yo	20:42 yo	15 MB	⋮
 18 oct, 3.51 p. m...aac	 yo	20:42 yo	10.8 MB	⋮

<https://drive.google.com/drive/folders/1zPQabz1mh0pjQRulNzhns5bFyVpK29s?usp=sharing>

ANEXO N° 02. SUBGRUPOS: VARIABLES NUMÉRICAS Y VARIABLES CATEGÓRICAS EN GOOGLE COLAB

```
columnNumeric=['CANTIDAD', 'INVENTARIO', 'TIEMPO_PROD', 'COND_CLIM', 'PH', 'YEAR', 'MONTH', 'DAY']
columnString=['FERTILIZANTE', 'TIPO_CULTIVO', 'ENTRADA/SALIDA']
```

ANEXO N° 03. IMPUTACIÓN DE VALORES NULOS EN LAS VARIABLES CATEGÓRICAS Y NUMÉRICAS EN GOOGLE COLAB

```

# Usamos los metodos de imputacion aprendidos!
from sklearn.impute import SimpleImputer
# Generamos el imputador iterativo - Imputacion Univariada Numerica
imp_univ_num = SimpleImputer(missing_values=np.nan, strategy='median')

# Generamos el imputador iterativo - Imputacion Univariada Categorica
imp_univ_cat = SimpleImputer(missing_values=np.nan, strategy='most_frequent')

# Generamos los subset de variables categoricas y numericas
data_impt_num = datos[columnNumeric]
data_impt_cat = datos[columnString]

# Realizamos la imputación univariada en una nueva base de datos - Variables Numericas
imp_univ_num.fit(data_impt_num)
imputed_data_univ_num = pd.DataFrame(data=imp_univ_num.transform(data_impt_num),
                                     columns=data_impt_num.columns, dtype='float')

# Realizamos la imputación univariada en una nueva base de datos - Variables Categoricas
imp_univ_cat.fit(data_impt_cat)
imputed_data_univ_cat = pd.DataFrame(data=imp_univ_cat.transform(data_impt_cat),
                                     columns=data_impt_cat.columns, dtype='object')

```

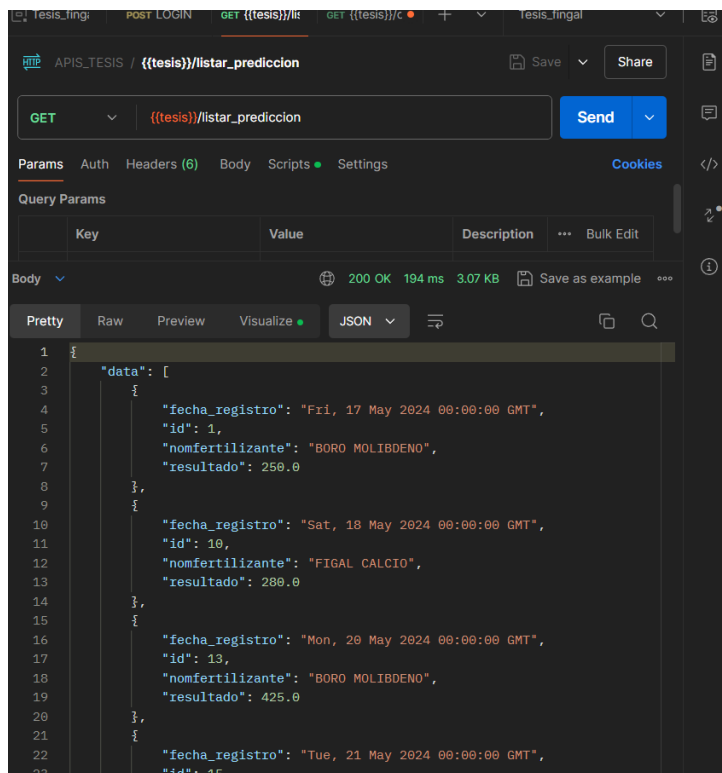
ANEXO N° 04. CÓDIGO PARA LA INICIACIÓN DEL SISTEMA

```

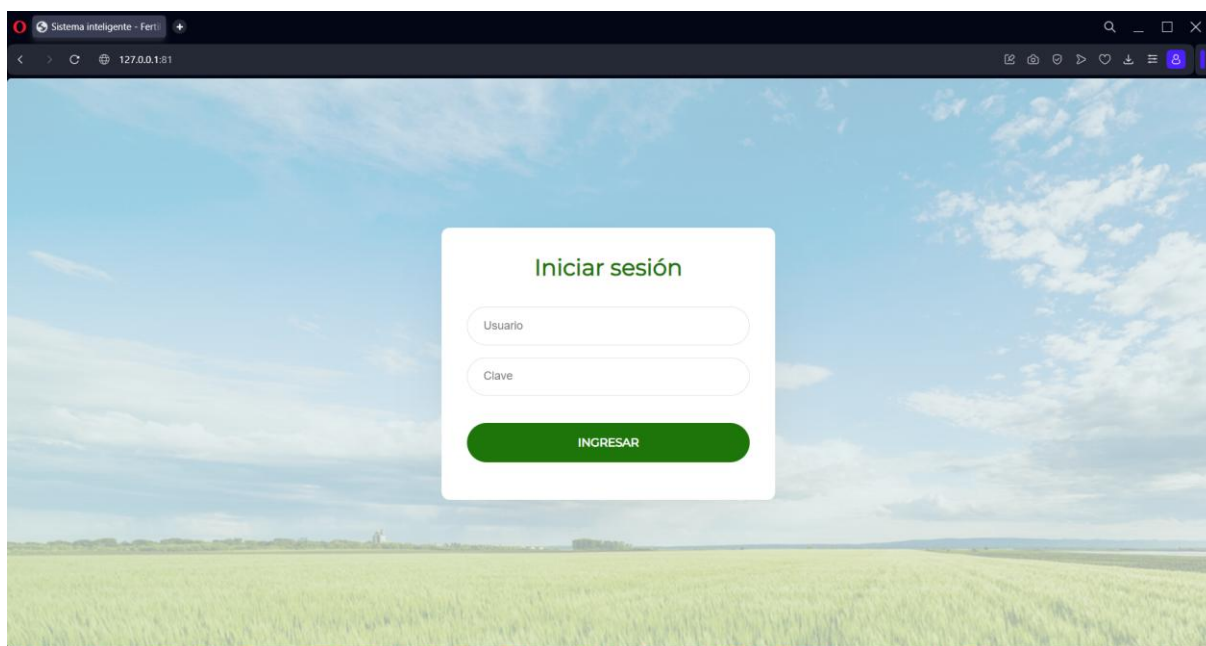
1 @ws_predecir.route("/pre_prod")
2 def mostrar_pre_prod():
3     return render_template('pre_prod.html')
4
5 @ws_predecir.route("/preprod")
6 def mostrar_preprod():
7     return render_template('preprod.html')
8
9 @ws_predecir.route("/listar_prediccion", methods=['GET'])
10 def listar_prediccion():
11     try:
12         resultado = Prediccion.listar_preprod()
13
14         if not resultado['status']:
15             print(f"Error al obtener la lista de predicciones: {resultado['message']}")
16             return jsonify({'status': False, 'message': resultado['message'], 'data': []}), 500
17
18         #print("Lista de predicciones:", resultado['data'])
19         return jsonify({'status': True, 'data': resultado['data']}), 200
20
21     except Exception as e:
22         print(f"Error al obtener la lista de predicciones: {e}")
23         return jsonify({'status': False, 'message': str(e), 'data': []}), 500
24
25
26 # Ruta para cargar los datos de fertilizantes en el combo
27 @ws_predecir.route("/cmb_fertilizante", methods=['GET'])
28 def cmb_fertilizante():
29     fertilizantes = Prediccion.obtener_fertilizantes()
30     return jsonify(fertilizantes)
31
32
33 # Asegurate de que la ruta al modelo es correcta
34 model_path = './static/modelo/modeloRN.h5'
35 if not os.path.exists(model_path):
36     raise FileNotFoundError(f"El archivo del modelo no se encuentra en la ruta: {model_path}")
37
38 try:
39     modelo = load_model(model_path)
40     print(f"Modelo cargado exitosamente desde {model_path}")
41 except Exception as e:
42     raise RuntimeError(f"Error cargando el modelo: {e}")
43

```

ANEXO N° 05. PRUEBA DEL POSTMAN PARA EVALUAR LA LISTA DE PREDICCIÓN



ANEXO N° 06. CAPTURAS DEL SISTEMA INTELIGENTE DESPLEGADO



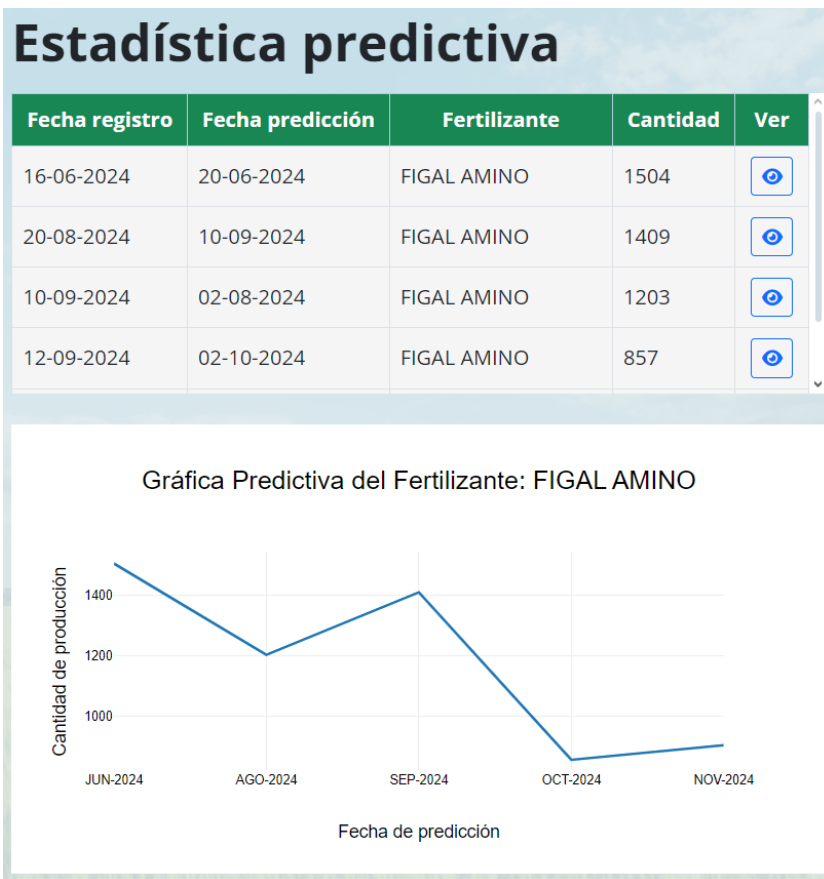
Iniciar sesión del sistema



Pantalla de inicio del sistema

Código	Fertilizante	Predicción	Generar gráfica
FER001	ACIDO HUMICO	?	
FER002	BORO MOLIBDENO	?	
FER003	CALVIVA	?	
FER004	CITOBLOOM	?	
FER005	CLOROFILA	?	
FER006	FIGAL AMINO	?	
FER007	FIGAL CALBOZINC	?	
FER008	FIGAL CALCIO	?	
FER009	FIGAL CALCIO-200LT	?	
FER010	FIGAL CALCIO-20LT	?	

Sistema de predice listando los fertilizante que tiene actualmente.

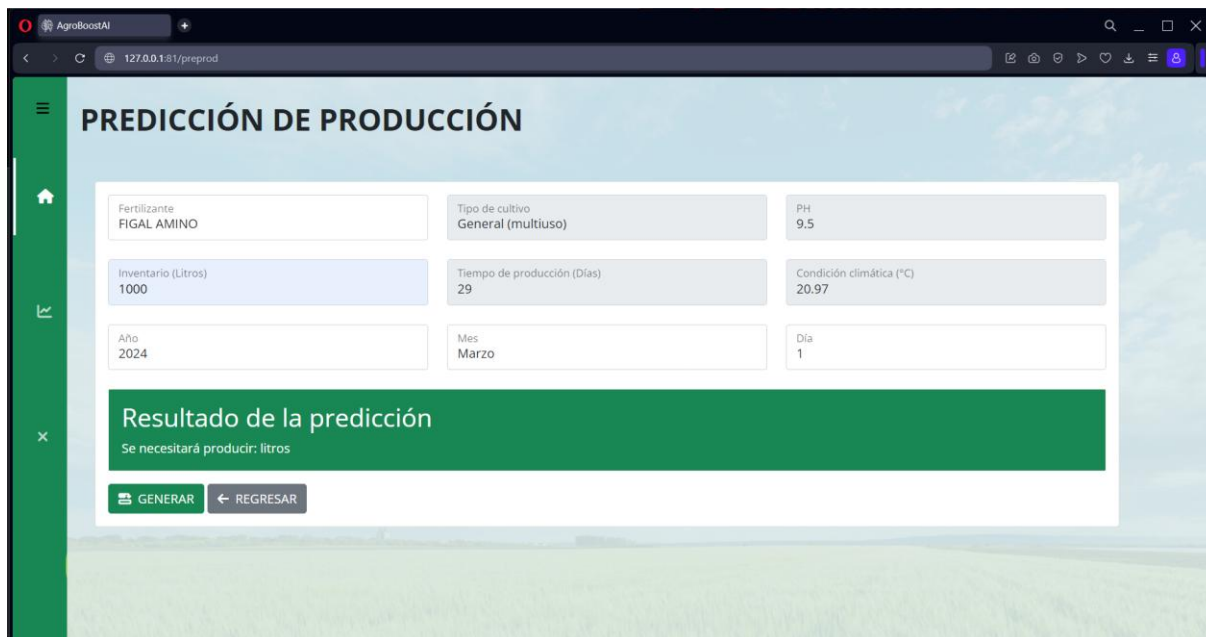


Estadística donde se ve las predicciones realizadas por fertilizante con su gráfica

DETALLE DE PREDICCIÓN

Fecha de Registro	20-08-2024
Fecha de Predicción	10-09-2024
Nombre del Fertilizante	FIGAL AMINO
Resultado	1409 litros
PH	9.5
Tiempo de Producción	29 días
Tipo de Cultivo	General (multiuso)

Detalle de las predicciones que se tiene de cada fertilizante



The screenshot displays the 'AgriBoostAI' web application interface. The browser address bar shows '127.0.0.1:81/preprod'. The page title is 'PREDICCIÓN DE PRODUCCIÓN'. The interface includes a sidebar with navigation icons (home, list, close) and a main content area with a form for inputting fertilizer details. The form fields are:

Fertilizante FIGAL AMINO	Tipo de cultivo General (multiuso)	PH 9.5
Inventario (Litros) 1000	Tiempo de producción (Días) 29	Condición climática (°C) 20.97
Año 2024	Mes Marzo	Día 1

Below the form, a green banner displays the prediction result: 'Resultado de la predicción' and 'Se necesitará producir: litros'. At the bottom of the form, there are two buttons: 'GENERAR' and 'REGRESAR'.

Sistema de predicción del fertilizante.

**ANEXO N° 7. RUBRICA DE LA NORMA ISO 25010 DESARROLLADA POR LOS
ESPECIALISTAS SELECCIONADOS**

Lista de Chequeo ISO 25010 [20] Evaluar la Calidad del Producto Software					
Software por evaluar:	Ing. Marlon Eugenio Vilchez Rivas				
Fecha:	05 – 07 – 24				
FUNCIONABILIDAD					
Definición	1	2	3	4	5
Conectividad de usuario correctamente.				X	
Posee un apartado de inteligencia de negocios.				X	
Despliegue de los datos correctamente.				X	
Creación y diseño la gráfica con su respectiva escala de calificación.				X	
USABILIDAD					
Definición	1	2	3	4	5
Despliegue y correcto funcionamiento en dispositivos web.				X	
Los usuarios pueden conectarse a las aplicaciones disponibles sin importar la plataforma que utilicen.				X	
Despliegue y correcto funcionamiento en diferentes navegadores.			X		
Presenta interfaces de usuario intuitivas y amigables.				X	
Presenta mensajes de errores claros y personalizados				X	
EFICIENCIA					
Definición	1	2	3	4	5
El tiempo de respuesta apropiada por el usuario				x	
Descarga de la gráfica correctamente.				x	
Tarda el sistema en generar el mensaje de confirmación.			x		

MANTEBILIDAD O MODIFICABILIDAD					
Definición	1	2	3	4	5
Facilidad para detectar fallos.			x		
Flexibilidad de la herramienta a los cambios de los recursos para realizar cambios como actualizaciones.			x		
FIABILIDAD					
Definición	1	2	3	4	5
Tolerancia a fallas.			x		
SEGURIDAD					
Definición	1	2	3	4	5
Es confiable y seguro el cargue de información utilizando.				x	
El sistema se encuentra disponible y puede ser utilizado con alto grado de fiabilidad por los usuarios.			x		
Capacidad de protección contra el acceso de datos e información no autorizados.				x	

ANEXO N° 8. PRUEBA DE SOFTWARE POR LA PLATAFORMA SONARCLOUD

The screenshot displays the SonarCloud interface for a project named PGWEB_TESIS. The main section is titled 'Main Branch Summary' and shows a 'Quality Gate' status of 'Passed'. The interface includes a sidebar with navigation options like Overview, Main Branch, Pull Requests, and Branches. The main content area shows various metrics: Security (0 Open Issues, A), Reliability (4 Open Issues, E), Maintainability (24 Open Issues, A), Accepted Issues (0), Coverage (0.0%), and Duplications (0.0%). A warning message at the top right indicates 'The last analysis has a warning. See details'. The interface is dark-themed and includes a search bar and user profile icons at the top.

Metric	Value	Grade
Security	0 Open Issues	A
Reliability	4 Open Issues	E
Maintainability	24 Open Issues	A
Accepted Issues	0	
Coverage	0.0%	
Duplications	0.0%	