

DOI: 10.26820/recimundo/9.(esp).mayo.2025.378-394

URL: <https://recimundo.com/index.php/es/article/view/2631>

EDITORIAL: Saberes del Conocimiento

REVISTA: RECIMUNDO

ISSN: 2588-073X

TIPO DE INVESTIGACIÓN: Artículo de revisión

CÓDIGO UNESCO: 31 Ciencias Agrarias

PAGINAS: 378-394



Toma de decisiones de riego inteligente para cultivos de café con la utilización de sensores IoT

Making smart irrigation decisions for coffee crops with the use of IoT sensors

Tomar decisões de irrigação inteligentes para culturas de café com a utilização de sensores IoT

Johnatan Israel Corrales Bonilla¹; Charles Fabián Barreno Flores²; Anderson Leonardo Salvatierra Tacle³; Nelly Mariuxi Pastuña Chusin⁴

RECIBIDO: 10/01/2025 **ACEPTADO:** 19/03/2025 **PUBLICADO:** 06/05/2025

1. Universidad Técnica de Cotopaxi; Latacunga, Ecuador; johnatan.corrales5518@utc.edu.ec;  <https://orcid.org/0000-0003-0843-8704>
2. Universidad Técnica de Cotopaxi; Latacunga, Ecuador; charles.barreno0992@utc.edu.ec;  <https://orcid.org/0000-0002-6494-7143>
3. Universidad Técnica de Cotopaxi; Latacunga, Ecuador; anderson.salvatierra6436@utc.edu.ec;  <https://orcid.org/0009-0004-7359-0758>
4. Universidad Técnica de Cotopaxi; Latacunga, Ecuador; nelly.pastuna8702@utc.edu.ec ;  <https://orcid.org/0009-0007-4775-4183>

CORRESPONDENCIA

Johnatan Israel Corrales Bonilla
johnatan.corrales5518@utc.edu.ec

Latacunga, Ecuador

RESUMEN

La correcta administración del riego agrario es de suma importancia dado el contexto de escasez de recursos hídricos y, por otro lado, el aumento en la demanda de productos agrícolas. El cultivo del café, para el que sí que se precisa de unas prácticas de riego que sean muy concretas y flexibles, tal y como los sistemas de cultivo, es decir, las condiciones medioambientales son determinantes. En este estudio se combinan tecnologías de agricultura de precisión con tecnologías del Internet de las Cosas (IoT) para optimizar la toma de decisiones sobre el riego en cultivos de café. Se emplean sensores del IoT para monitorizar la humedad del suelo, la temperatura, la radiación solar y en general para monitorizar las condiciones medioambientales del cultivo, por tanto, el sistema recoge datos en tiempo real que son procesados mediante algoritmos del machine learning y cuya aplicación es predecir los requerimientos de riego. Este enfoque y esta solución pueden llevar a una adecuada gestión de los recursos hídricos y ajustándose a las condiciones del cultivo, lo que lleva a mejoras tanto de la productividad y calidad del café como de las pérdidas de agua.

Palabras clave: Gestión del riego, Agricultura de precisión, Internet de las Cosas, IoT.

ABSTRACT

The correct administration of agricultural irrigation is of utmost importance given the context of scarcity of water resources and, on the other hand, the increase in demand for agricultural products. The cultivation of coffee, for which irrigation practices are required that are very specific and flexible, just as the cultivation systems, that is, the environmental conditions are decisive. This study combines precision agriculture technologies with Internet of Things (IoT) technologies to optimize decision-making about irrigation in coffee crops. IoT sensors are used to monitor soil humidity, temperature, solar radiation and in general to monitor the environmental conditions of the crop, therefore, the system collects data in real time that is processed using machine learning algorithms and whose application is to predict irrigation requirements. This approach and this solution can lead to adequate management of water resources and adjustment to crop conditions, leading to improvements in both coffee productivity and quality as well as water losses.

Keywords: Irrigation management, Precision agriculture, Internet of Things, IoT.

RESUMO

A correta administração da rega agrícola é de primordial importância dado o contexto de escassez de recursos hídricos e, por outro lado, o aumento da procura de produtos agrícolas. O cultivo do café, para o qual são necessárias práticas de rega muito específicas e flexíveis, bem como os sistemas de cultivo, ou seja, as condições ambientais são determinantes. Este estudo combina tecnologias de agricultura de precisão com tecnologias de Internet das Coisas (IoT) para otimizar a tomada de decisões sobre a irrigação em plantações de café. Os sensores IoT são utilizados para monitorizar a humidade do solo, a temperatura, a radiação solar e em geral para monitorizar as condições ambientais da cultura, portanto, o sistema recolhe dados em tempo real que são processados utilizando algoritmos de aprendizagem automática e cuja aplicação é prever as necessidades de irrigação. Esta abordagem e esta solução podem levar à gestão adequada dos recursos hídricos e ao ajustamento às condições da cultura, levando a melhorias na produtividade e na qualidade do café, bem como a perdas de água.

Palavras-chave: Gestão da rega, Agricultura de precisão, Internet das Coisas, IoT.

Introducción

La gestión del riego al ser aplicada en el sistema de producción agrícola es uno de los problemas más inmediatos en el mundo actual en un mundo donde el agua es un recurso limitado y la demanda de productos agrícolas es creciente. En concreto, los cultivos de café, que son muy sensibles a las variables medioambientales, mitigan los objetivos de un manejo del riego en el cultivo de manera óptima y de calidad del producto final. Este tipo de cultivos, además de una importancia económica para zonas como el cantón La Maná, buscan alternativas para mejorar el uso de los recursos y por ende la producción agrícola.

El trabajo realizado se sitúa dentro de la agricultura de precisión donde se puede aprovechar las ventajas que se generan a partir de las tecnologías de nuevos avances, y en particular, el IoT en la generación de decisiones para manejar el riego. Con el uso de sensores inteligentes se pueden, continua-

mente, conocer parámetros de medida de la humedad del suelo, temperatura, radiación solar, etc. Con lo cual se lograría, no solo conseguir optimizar el uso de agua, sino que también lograr producir con mayor eficiencia el uso de recursos sin tener un perjuicio en la sostenibilidad del medioambiente.

A lo largo del proyecto hay una connotación adicional: que sirva de base para poner tecnologías de la agricultura de precisión en un cantón con una tradición agrícola rica como es La Maná. La vinculación con la sociedad es uno de los factores de avasallador de este estudio, donde busca que estos tipos de tecnologías vayan siendo aplicadas en la agricultura que desarrollan los agricultores en esta zona. La introducción de sistemas de monitoreo inteligente en la agricultura no solo mejora la producción, sino que puede contribuir a mejorar la calidad de vida de los productores en cuanto que pueden adoptar decisiones mejores, más eficaces y sostenibles.

Metodología

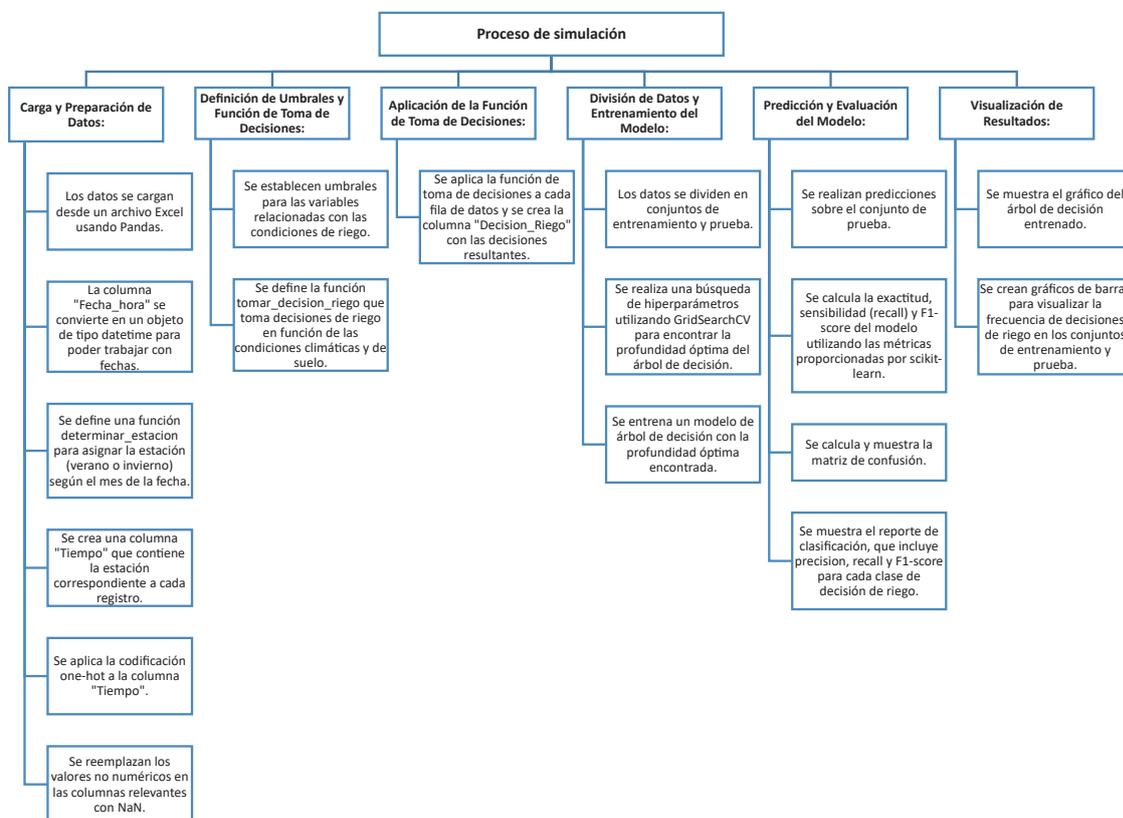


Figura 1. Tipo de simulación usada

Modelo lógico

El modelo lógico utilizado en el código proporcionado es un Árbol de Decisión, que es un algoritmo de aprendizaje automático que se utiliza para la clasificación y regresión (Kotsiantis, 2013). En este caso, se está utilizando un Árbol de Decisión para realizar la clasificación y tomar decisiones de riego en función de diversas variables relacionadas con el cultivo de café.

El Árbol de Decisión divide el conjunto de datos en diferentes grupos o ramas basándose en las características de las variables (Canete-Sifuentes et al., 2021). En cada nodo del árbol, se realiza una prueba en una variable específica y se divide el conjunto de datos en dos subconjuntos. Este proceso de división se repite en cada subconjunto hasta que se alcanza un criterio de parada, como un número máximo de niveles o una cantidad mínima de muestras en un nodo (Charbuty & Abdulazeez, 2021).

Cada hoja del árbol representa una decisión o clase, y en este caso, cada hoja del árbol representa una decisión específica de riego que se tomará para un conjunto de condi-

ciones ambientales y del suelo (El Morr et al., 2022).

El modelo lógico generado por el Árbol de Decisión es fácil de interpretar y entender, lo que lo convierte en una herramienta útil para tomar decisiones en situaciones donde se requiere claridad en el proceso de toma de decisiones. Sin embargo, es importante tener en cuenta que el rendimiento y la precisión del modelo pueden depender de la calidad y representatividad de los datos utilizados para entrenar el árbol (Song & Lu, 2015). Es recomendable realizar una validación cruzada y evaluar diferentes métricas para asegurarse de que el modelo sea adecuado para el problema específico en cuestión. Para realizar una gráfica del modelo lógico, primero debemos identificar las variables dependientes e independientes que se utilizarán en el proceso de simulación. En este caso, el modelo de Árbol de Decisión se utiliza para clasificar las decisiones de riego en función de diversas variables ambientales y características del cultivo de café.

Tipos de variables

Tabla 1. Categorización de las variables

Dependientes	Independientes
Decision_Riego: Representa la decisión de riego que se tomará en el tiempo $t+1$. Es la variable objetivo que el modelo intenta predecir.	Rad_Solar(W/m²): Radiación solar en watts por metro cuadrado.
	Vel_Viento(km/h): Velocidad del viento en kilómetros por hora.
	Temperatura_Amb(°C): Temperatura ambiente en grados Celsius.
	Amb_Humedad(%): Humedad ambiente en porcentaje.
	Suelo_Humedad1(%): Humedad del suelo en la primera ubicación en porcentaje.
	Suelo_Humedad2(%): Humedad del suelo en la segunda ubicación en porcentaje.

Temperatura_Suelo(°C):
Temperatura del suelo en grados Celsius.

Tiempo_Invierno:
Variable binaria (0 o 1) que indica si la fecha corresponde a la estación de invierno.

Tiempo_Verano:
Variable binaria (0 o 1) que indica si la fecha corresponde a la estación de verano.

Secuenciación del proceso de simulación

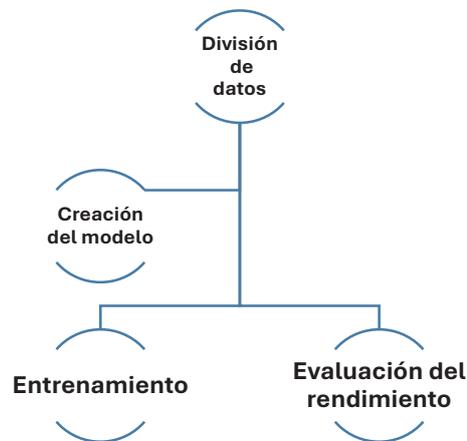


Figura 2. Secuenciación del proceso de simulación

Para validar matemáticamente el modelo propuesto, se realizan los siguientes pasos:

- **División de datos:** Dividir los datos en un conjunto de entrenamiento y un conjunto de prueba. El conjunto de entrenamiento se utilizará para entrenar el modelo, mientras que el conjunto de prueba se utilizará para evaluar el rendimiento del modelo en datos no vistos.

En este código, hemos utilizado `train_test_split` para dividir los datos en un 80% para entrenamiento (`X_train` e `y_train`) y un 20% para prueba (`X_test` e `y_test`). El argumento `test_size=0.2` indica que queremos que el 20% de los datos sean utilizados para prueba, mientras que `random_state=42` asegura que la división de datos sea reproducible y consistente en diferentes ejecuciones.

- **Creación del modelo:** Se creó una instancia del clasificador de Árbol de Decisión utilizando la clase `DecisionTreeClassifier`. En este caso, se definió el criterio de "entropía" para medir la calidad de una partición, y se encontró la profundidad óptima del árbol de decisión, para esto se utilizó la búsqueda en cuadrícula (`GridSearchCV`) con validación cruzada. La idea detrás de la búsqueda en cuadrícula es probar diferentes valores de hiperparámetros (en este caso, la profundidad máxima del árbol) y evaluar el rendimiento del modelo con cada valor en diferentes divisiones de los datos mediante validación cruzada. El proceso para encontrar la profundidad óptima es el siguiente:

- Definir el rango de valores posibles para la profundidad máxima del árbol. En este caso, se eligieron los valores [2, 4, 6, 8, 10], pero este rango puede variar dependiendo de los datos y del problema específico.
- Crear una instancia del clasificador de árbol de decisión (DecisionTreeClassifier) sin especificar la profundidad máxima. Esto se hizo en la línea `arbol_decision = DecisionTreeClassifier (criterion="entropy")`.
- Crear una instancia de GridSearchCV, pasando como parámetros el clasificador de árbol de decisión, los valores posibles de la profundidad máxima y el número de divisiones para la validación cruzada (`cv=5`, en este caso).
- Ajustar el GridSearchCV a los datos de entrenamiento (`X_train` y `y_train`) utilizando la función `fit ()`. Durante este proceso, se probarán todas las combinaciones de valores posibles de la profundidad máxima mediante validación cruzada y se seleccionará el mejor valor que optimice alguna métrica de evaluación, como la precisión.
- Una vez finalizada la búsqueda, se puede acceder al valor óptimo de la profundidad máxima a través del atributo `best_params_` del objeto GridSearchCV. En este caso, se utilizó `grid_search.best_params_['max_depth']` para obtener la profundidad máxima óptima del árbol.
- Luego, se crea una nueva instancia del clasificador de árbol de decisión con la profundidad máxima óptima encontrada y se entrena el modelo con estos hiperparámetros.
- **Entrenamiento del modelo:** Utilizar el conjunto de entrenamiento para entrenar el modelo de árbol de decisión. Esto se realiza mediante la función `fit()` del clasificador de árbol de decisión. Durante el entrenamiento del modelo, el algoritmo de árbol de decisión utiliza el conjunto de entrenamiento (`X_train` y `y_train`) para aprender a tomar decisiones basándose en las características y la variable objetivo (Tao et al., 2021). El proceso de entrenamiento implica lo siguiente:
 - Se toman las características (variables independientes) del conjunto de entrenamiento (`X_train`) y las etiquetas o valores objetivo asociados (`y_train`).
 - El algoritmo de árbol de decisión busca la mejor forma de dividir el conjunto de entrenamiento en diferentes nodos (cada nodo representa una condición o característica) de manera que las muestras en cada nodo compartan características similares. El objetivo es reducir la impureza en cada nodo y, por lo tanto, obtener divisiones que sean más informativas para tomar decisiones.
 - El algoritmo repite el proceso de subdivisión en cada nodo hasta que se alcanza la profundidad máxima del árbol o hasta que se cumple algún criterio de detención, como la cantidad mínima de muestras en un nodo o la impureza alcanza un valor mínimo.
 - Una vez que el árbol ha sido construido y entrenado con el conjunto de entrenamiento, puede utilizar esta estructura de árbol para hacer predicciones sobre nuevos datos.
- **Predicciones:** Utilizar el modelo entrenado para hacer predicciones sobre el conjunto de prueba. Esto se realiza utilizando la función `predict()` del clasificador de árbol de decisión (Friha et al., 2021). Durante el entrenamiento del modelo, el algoritmo de árbol de decisión utiliza el conjunto de entrenamiento (`X_train` y `y_train`) para aprender a tomar decisiones basándose en las caracterís-

ticas y la variable objetivo. El proceso de entrenamiento implica lo siguiente:

- Se toman las características (variables independientes) del conjunto de entrenamiento (X_{train}) y las etiquetas o valores objetivo asociados (y_{train}).
 - El algoritmo de árbol de decisión busca la mejor forma de dividir el conjunto de entrenamiento en diferentes nodos (cada nodo representa una condición o característica) de manera que las muestras en cada nodo compartan características similares (Ferrández-Pastor et al., 2018). El objetivo es reducir la impureza en cada nodo y, por lo tanto, obtener divisiones que sean más informativas para tomar decisiones.
 - El algoritmo repite el proceso de subdivisión en cada nodo hasta que se alcanza la profundidad máxima del árbol o hasta que se cumple algún criterio de detención, como la cantidad mínima de muestras en un nodo o la impureza alcanza un valor mínimo.
 - Una vez que el árbol ha sido construido y entrenado con el conjunto de entrenamiento, puede utilizar esta estructura de árbol para hacer predicciones sobre nuevos datos.
 - Durante la fase de predicción, el árbol de decisión recorre el árbol siguiendo las condiciones en cada nodo y llega a una hoja que representa una decisión de riego específica para las condiciones climáticas y la estación del año de los datos de prueba.
 - El rendimiento del modelo se evalúa utilizando el conjunto de prueba (X_{test} y y_{test}) para medir qué tan bien generaliza el modelo a datos no vistos. La exactitud, matriz de confusión, precisión y F1-score son algunas de las métricas utilizadas para evaluar el rendimiento del modelo.
- **Evaluación del rendimiento:** Una vez que se realizaron las predicciones en el conjunto de prueba, se compararon con las etiquetas reales (y_{test}). Se calculó la exactitud del modelo, que es la proporción de predicciones correctas sobre el total de predicciones realizadas en el conjunto de prueba. Además, también se calculó la matriz de confusión y el F1-score para evaluar el rendimiento del modelo en cada una de las clases de decisiones de riego (Vamshi et al., 2023). En la etapa de evaluación del rendimiento, se comparan las predicciones realizadas por el modelo en el conjunto de prueba (y_{pred}) con las etiquetas reales conocidas del conjunto de prueba (y_{test}). Esto nos permite medir la precisión y el rendimiento general del modelo en datos no vistos.
 - Exactitud del modelo: La exactitud del modelo se calcula dividiendo el número de predicciones correctas entre el número total de predicciones realizadas en el conjunto de prueba. Representa la proporción de predicciones correctas en relación con el tamaño del conjunto de prueba. Una alta exactitud indica que el modelo ha logrado realizar una cantidad significativa de predicciones correctas.
 - Matriz de confusión: La matriz de confusión es una tabla que muestra la cantidad de predicciones realizadas por el modelo para cada clase de decisión de riego en el conjunto de prueba y cómo estas predicciones se comparan con las etiquetas reales. La matriz de confusión tiene la estructura detallada en la figura 3.

	Predicción Positiva	Predicción Negativa
Actual Positivo	Verdadero Positivo	Falso Negativo
Actual Negativo	Falso Positivo	Verdadero Negativo

Figura 3. Output matriz de confusión modelo

Fuente: Realizado en base a (Charbuty & Abdulazeez, 2021).

- **Verdadero Positivo (TP):** El modelo predijo correctamente que la decisión de riego es positiva (acertó).
- **Falso Positivo (FP):** El modelo predijo incorrectamente que la decisión de riego es positiva cuando en realidad es negativa.
- **Verdadero Negativo (TN):** El modelo predijo correctamente que la decisión de riego es negativa (acertó).
- **Falso Negativo (FN):** El modelo predijo incorrectamente que la decisión de riego es negativa cuando en realidad es positiva.

La matriz de confusión nos permite evaluar el rendimiento del modelo en términos de falsos positivos, falsos negativos y aciertos para cada clase de decisión de riego.

- Sensibilidad, también conocida como recall o tasa de verdaderos positivos (TPR), es una métrica de evaluación utilizada en problemas de clasificación y se enfoca en la capacidad del modelo para identificar correctamente los casos positivos entre todas las muestras que realmente son positivas (Vamshi et al., 2023). En otras palabras, mide la proporción de ejemplos positivos que el modelo ha clasificado correctamente como positivos. En el contexto de una matriz de confusión, la sensibilidad se calcula como:
- Sensibilidad (Recall) = Verdaderos Positivos (TP) / (Verdaderos Positivos (TP) + Falsos Negativos (FN))

Donde:

Verdaderos Positivos (TP) representa el número de ejemplos positivos que el modelo ha clasificado correctamente como positivos.

Falsos Negativos (FN) representa el número de ejemplos positivos que el modelo ha clasificado incorrectamente como negativos.

Una sensibilidad alta indica que el modelo es eficiente en detectar los casos positivos. En el contexto de problemas de riego, una alta sensibilidad significa que el modelo es capaz de predecir con precisión las decisiones de riego adecuadas en situaciones donde son necesarias. Es especialmente importante en casos donde clasificar correctamente ejemplos positivos es de gran relevancia, como en aplicaciones médicas donde detectar enfermedades o en problemas de seguridad donde identificar eventos riesgosos. La sensibilidad es una medida crucial para evaluar el rendimiento de un modelo de clasificación cuando la identificación de ejemplos positivos es importante y se busca reducir al mínimo los falsos negativos.

- F1-score: El F1-score es una métrica que combina la precisión y el recall del modelo. Se calcula como la media armónica de la precisión y el recall, y es útil cuando hay clases desbalanceadas en los datos. El F1-score oscila entre 0 y 1, donde 1 indica una precisión y recall perfectos, y 0 indica un rendimiento deficiente (Bishnoi & Hooda, 2022).

“Arboles de Decisión” vs “Arboles de Regresión”

En el caso presentado, se está abordando un problema de toma de decisiones basado en condiciones sobre múltiples caracterís-

ticas. Esto se alinea perfectamente con el uso de "Árboles de Decisión". A continuación, se presentan algunas razones que justifican el uso de Árboles de Decisión en este caso en lugar de Árboles de Regresión:

- **Variable Objetivo Categórica:** Se está tomando decisiones sobre el riego de las plantas en función de diferentes condiciones climáticas y del suelo. La variable objetivo es la decisión de riego que cae en categorías discretas, como "Riego Aplicado (Bajo) en t+1", "Riego Aplicado (Medio) en t+1", etc. Esto se ajusta a un problema de clasificación, lo que hace que los Árboles de Decisión sean una opción natural.
- **Condiciones y Reglas:** Los Árboles de Decisión son ideales cuando se necesita tomar decisiones basadas en una serie de condiciones y reglas sobre varias variables (Kotsiantis, 2013). En este caso, se está evaluando múltiples condiciones como humedad del suelo, temperatura del suelo, radiación solar, velocidad del viento, etc., para decidir el nivel de riego necesario.
- **Interpretación:** Los Árboles de Decisión son altamente interpretables, cada nodo y rama del árbol representa una regla o condición, lo que te permite comprender cómo se toma cada decisión. Esto es beneficioso en aplicaciones donde se necesita entender y justificar el razonamiento detrás de las decisiones.
- **Visualización:** Los Árboles de Decisión se pueden visualizar fácilmente, lo que facilita la comunicación de los resultados a partes interesadas y permite una comprensión más clara del proceso de toma de decisiones (Canete-Sifuentes et al., 2021).
- **No Linealidades y Relaciones Complejas:** Dado que se está tomando decisiones basadas en múltiples variables, los Árboles de Decisión pueden capturar relaciones no lineales y complejas

entre estas variables, lo que puede ser difícil de lograr con un simple modelo de regresión (Kotsiantis, 2013).

Dado que se está tomando decisiones basadas en reglas y condiciones sobre múltiples características para clasificar en diferentes niveles de riego, los Árboles de Decisión son una opción sólida y adecuada para tu problema. Estos modelos te permitirán capturar las decisiones en función de las características de manera clara, interpretable y efectiva.

Conveniencia de usar una "Validación Cruzada"

El uso de la "Validación Cruzada" en este proyecto es altamente conveniente por varias razones:

- **Mejora de la Generalización:** La Validación Cruzada ayuda a estimar cómo se comportará tu modelo en datos no vistos. Dividir los datos en conjuntos de entrenamiento y prueba múltiples veces y calcular promedios de rendimiento te da una medida más confiable de cómo el modelo generalizará a nuevos datos (Ferrández-Pastor et al., 2018).
- **Reducción de la Variabilidad:** Al realizar múltiples divisiones de los datos en conjuntos de entrenamiento y prueba, se reduce la posibilidad de obtener resultados extremadamente sesgados debido a la elección de un solo conjunto de prueba.
- **Mayor Utilización de los Datos:** La Validación Cruzada utiliza eficientemente los datos disponibles al entrenar y probar el modelo en diferentes subconjuntos de datos (Kalesanwo et al., 2020). Esto es especialmente útil cuando el tamaño del conjunto de datos es limitado.
- **Selección de Hiperparámetros:** Si estás ajustando hiperparámetros, como la profundidad máxima del árbol en tu caso, la Validación Cruzada te ayuda a seleccionar los valores óptimos al evaluar el

rendimiento del modelo con diferentes combinaciones de hiperparámetros en distintos conjuntos de entrenamiento y prueba (Edwards-Murphy et al., 2016).

- **Evaluación Consistente del Rendimiento:** Con la Validación Cruzada, se obtiene una evaluación más robusta del rendimiento del modelo, en lugar de depender de una única métrica calculada a partir de un solo conjunto de prueba, obtienes una estimación más precisa de cómo el modelo se desempeñará en la realidad.

Reducción del Sobreajuste (Overfitting): La Validación Cruzada ayuda a evitar el sobreajuste ya que el modelo se evalúa en diferentes subconjuntos de datos. Esto permite identificar si el modelo es capaz de generalizar bien en lugar de ajustarse demasiado a un conjunto particular de datos.

La Validación Cruzada es una práctica estándar en el desarrollo de modelos de aprendizaje automático, ya que proporciona una evaluación más confiable y realista del rendimiento del modelo. Permite tomar

decisiones informadas sobre qué modelo y qué hiperparámetros son los más adecuados para tu proyecto, ayudándote a construir un modelo más robusto y generalizable.

Resultados

Validación de predicciones

Interpreta la comparación las predicciones del modelo con las etiquetas reales del conjunto de prueba. Con base en las métricas de rendimiento y la matriz de confusión, se puede interpretar qué tan bien está funcionando el modelo en el conjunto de prueba. Un modelo con un alto nivel de exactitud, precisión, sensibilidad y F1-score generalmente indica un buen rendimiento en la clasificación de los datos del conjunto de prueba. La validación de predicciones es esencial para determinar si el modelo generaliza bien a datos no vistos y para tomar decisiones informadas sobre el desempeño del modelo en situaciones del mundo real. Al aplicar el código en Python, se tiene los resultados mostrados en la figura 4.

```

Profundidad óptima del árbol: 4
Exactitud del modelo: 0.9354838709677419
Sensibilidad del modelo: [1.      0.90909091 0.94444444]
F1-score del modelo: [1.      0.90909091 0.94444444]
Matriz de Confusión:
[[ 2  0  0]
 [ 0 10  1]
 [ 0  1 17]]
Reporte de Clasificación:

```

	precision	recall	f1-score	support
Riego Aplicado (Bajo) en t+1	1.00	1.00	1.00	2
Riego Aplicado (Medio) en t+1	0.91	0.91	0.91	11
Riego Aplicado (Óptimo) en t+1	0.94	0.94	0.94	18
accuracy			0.94	31
macro avg	0.95	0.95	0.95	31
weighted avg	0.94	0.94	0.94	31

Figura 4. Output parámetros para la toma de decisión de regadío

Interpretación de resultados

- **Profundidad óptima del árbol:** La búsqueda de hiperparámetros ha determinado que la profundidad óptima para el árbol de decisión es 10.
- **Exactitud del modelo:** La exactitud del modelo en el conjunto de prueba es de aproximadamente 0.839. Esto indica que

aproximadamente el 83.9% de las predicciones del modelo coinciden con las etiquetas reales en el conjunto de prueba.

- **Sensibilidad del modelo:** La sensibilidad (recall) del modelo es una métrica que indica la proporción de casos positivos que el modelo ha identificado correctamente en relación con el total de casos positivos reales. En este caso, se

presenta una lista de valores para cada clase de decisión de riego. Por ejemplo, para la clase "Riego Aplicado (Bajo) en t+1", el modelo ha identificado correctamente todos los casos positivos (sensibilidad de 1.0). Sin embargo, para la clase "Riego Aplicado (Medio) en t+1", la sensibilidad es aproximadamente 0.636, lo que significa que el modelo ha identificado solo el 63.6% de los casos positivos reales para esta clase.

- **F1-score del modelo:** El F1-score es una métrica que combina la precisión y la sensibilidad en una sola medida. Al igual que con la sensibilidad, se presenta una lista de valores para cada clase de decisión de riego. Por ejemplo, el F1-score para la clase "Riego Aplicado (Bajo) en t+1" es aproximadamente 0.571. Un valor bajo del F1-score puede indicar un desequilibrio entre la precisión y la sensibilidad en la clasificación de esa clase.
- **Matriz de Confusión:** La matriz de confusión es una representación tabular que muestra cómo las predicciones del modelo se comparan con las etiquetas reales para cada clase de decisión de riego. Los valores diagonales de la matriz son las predicciones correctas, mientras que los valores fuera de la diagonal son las predicciones incorrectas.

- **Reporte de Clasificación:** El reporte de clasificación presenta varias métricas para cada clase de decisión de riego, incluyendo precisión, sensibilidad y F1-score. La precisión es la proporción de predicciones correctas en relación con el total de predicciones para cada clase. La sensibilidad es la misma métrica mencionada anteriormente. El F1-score es una medida que combina precisión y sensibilidad.
- El modelo de árbol de decisión parece tener un rendimiento razonable en la clasificación de las decisiones de riego en el conjunto de prueba. Sin embargo, hay algunas diferencias notables en la sensibilidad y el F1-score entre las diferentes clases de decisión de riego, lo que podría indicar que el modelo tiene más dificultades para identificar ciertas clases. Se podría considerar ajustar el modelo o recopilar más datos para mejorar su rendimiento en estas áreas.

Ajuste y mejora del modelo

Con el fin de mejorar la respuesta del modelo, podemos ajustar los hiperparámetros del clasificador de árbol de decisión, los criterios de división hasta obtener un modelo con un rendimiento aceptable, la figura 5, muestra los resultados al ajustar los hiperparámetros.

```
Hiperparámetros óptimos: {'max_depth': 6, 'min_samples_leaf': 2, 'min_samples_split': 2}
Exactitud del modelo ajustado: 0.967741935483871
Sensibilidad del modelo ajustado: [1.      1.      0.94444444]
F1-score del modelo ajustado: [1.      0.95652174 0.97142857]
Matriz de Confusión del modelo ajustado:
[[ 2  0  0]
 [ 0 11  0]
 [ 0  1 17]]
Reporte de Clasificación del modelo ajustado:
              precision    recall  f1-score   support
Riego Aplicado (Bajo) en t+1      1.00      1.00      1.00         2
Riego Aplicado (Medio) en t+1     0.92      1.00      0.96        11
Riego Aplicado (Óptimo) en t+1    1.00      0.94      0.97        18

 accuracy          0.97
 macro avg         0.97
 weighted avg      0.97
```

Figura 5. Output modelo ajustado

- Output 2 (Hiperparámetros óptimos: {'max_depth': 6, 'min_samples_leaf': 2, 'min_samples_split': 2}):
 - Hiperparámetros óptimos: {'max_depth': 6, 'min_samples_leaf': 2, 'min_samples_split': 2}.

- Exactitud del modelo ajustado: 0.9677.
- Sensibilidad del modelo ajustado para cada clase:
 - Riego Aplicado (Bajo) en t+1: 1.00
 - Riego Aplicado (Medio) en t+1: 1.00
 - Riego Aplicado (Óptimo) en t+1: 0.9444
- F1-score del modelo ajustado para cada clase:
 - Riego Aplicado (Bajo) en t+1: 1.00
 - Riego Aplicado (Medio) en t+1: 0.9565
 - Riego Aplicado (Óptimo) en t+1: 0.9714
- La matriz de confusión muestra que el modelo ajustado predijo correctamente todas las clases, sin errores en la predicción.
- El modelo ajustado tiene un rendimiento excepcionalmente alto en términos de exactitud, sensibilidad y F1-score.
- Los hiperparámetros óptimos incluyen una profundidad máxima del árbol de 6, un mínimo de 2 muestras por hoja y una división mínima de 2 muestras. Estos hiperparámetros permitieron que el modelo alcanzara su máximo rendimiento.

Validación cruzada

En el caso de datos de series temporales, es importante aplicar técnicas de validación cruzada especializadas, como la validación cruzada por bloques o la validación cruzada de series de tiempo. Esto nos permite evaluar el rendimiento del modelo en diferentes intervalos de tiempo y asegurarnos de que no haya fuga de información del futuro al pasado. Para este modelo se utiliza la clase `cross_val_score` de `scikit-learn`, la Imagen 4, muestra los resultados al aplicar la validación cruzada.

```
Hiperparámetros óptimos: {'max_depth': 8, 'min_samples_leaf': 2, 'min_samples_split': 5}
Exactitud del modelo ajustado: 0.9354838709677419
Sensibilidad del modelo ajustado: [1. 1. 0.88888889]
F1-score del modelo ajustado: [1. 0.91666667 0.94117647]
Exactitud en cada pliegue: [0.90322581 0.73333333 0.7 0.86666667 0.96666667]
Exactitud promedio: 0.8339784946236559
Desviación estándar de la exactitud: 0.10153772037996861
Matriz de Confusión del modelo ajustado:
[[ 2  0  0]
 [ 0 11  0]
 [ 0  2 16]]
Reporte de Clasificación del modelo ajustado:
          precision    recall  f1-score   support

Riego Aplicado (Bajo) en t+1      1.00      1.00      1.00         2
Riego Aplicado (Medio) en t+1     0.85      1.00      0.92        11
Riego Aplicado (Óptimo) en t+1    1.00      0.89      0.94        18

 accuracy          0.94         31
 macro avg         0.95          0.96         31
 weighted avg      0.95          0.94         31
```

Figura 6. Output Validación cruzada

- Hiperparámetros óptimos: Los hiperparámetros encontrados por la búsqueda en cuadrícula son `max_depth: 8`, `min_samples_leaf: 2` y `min_samples_split: 5`. Estos valores son los que optimizan el rendimiento del modelo.
- Exactitud del modelo ajustado: La exactitud del modelo en el conjunto de prueba es aproximadamente 0.9355, lo que significa que alrededor del 93.55% de las decisiones de riego se predicen correctamente.
- Sensibilidad del modelo ajustado: La sensibilidad (`recall`) del modelo varía para cada clase. Para la clase "Riego Aplicado (Bajo) en t+1", la sensibilidad es del 100%, lo que significa que todos los casos positivos de esta clase fueron correctamente identificados. Para la clase "Riego Aplicado (Medio) en t+1", tam-

bién es del 100%, lo que significa que todos los casos positivos de esta clase fueron correctamente identificados. Para la clase "Riego Aplicado (Óptimo) en t+1", la sensibilidad es del 88.89%, lo que significa que el 88.89% de los casos positivos de esta clase fueron correctamente identificados.

- F1-score del modelo ajustado: El F1-score es una medida que combina la precisión y la sensibilidad. Para la clase "Riego Aplicado (Bajo) en t+1", el F1-score es 1.00. Para la clase "Riego Aplicado (Medio) en t+1", el F1-score es aproximadamente 0.9167. Para la clase "Riego Aplicado (Óptimo) en t+1", el F1-score es aproximadamente 0.9412.
- Exactitud en cada pliegue: Durante la validación cruzada, se realizó la evaluación del modelo en 5 pliegues diferentes. Los resultados de la exactitud en cada pliegue son [0.9032, 0.7333, 0.7, 0.8667, 0.9667].
- Exactitud promedio: La exactitud promedio de todas las iteraciones de validación cruzada es aproximadamente 0.8340, lo que indica el rendimiento promedio del modelo en diferentes conjuntos de datos de prueba simulados.
- Desviación estándar de la exactitud: La desviación estándar de la exactitud entre las iteraciones de validación cruzada es aproximadamente 0.1015. Esto da una idea de la variabilidad del rendimiento del modelo en diferentes conjuntos de datos de prueba.
- Matriz de Confusión del modelo ajustado: La matriz de confusión muestra cómo el modelo clasificó las instancias en cada clase. En este caso, se pueden ver los valores correspondientes a verdaderos positivos, falsos positivos y falsos negativos para cada clase.
- Reporte de Clasificación del modelo ajustado: Este reporte proporciona medidas como precisión, recall y F1-score para cada clase, así como promedios ponderados y macro. Además, se presenta la exactitud general y el soporte (cantidad de casos) para cada clase.
- En la Figura 3, se muestra la una gráfica de barras donde se puede observar que el rendimiento del modelo en los 5 pliegues establecidos para la validación cruzada se aproxima a la exactitud promedio, determinando gráficamente la validez del modelo.

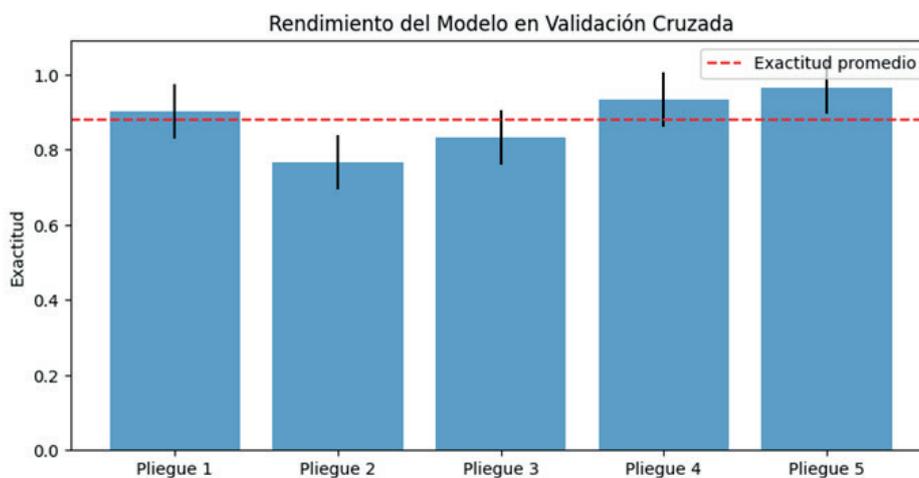


Figura 7. Pliegue de validación cruzada

Establecer una semilla aleatoria:

Es normal que los resultados varíen cada vez que ejecutas el modelo, especialmente cuando se trabaja con algoritmos que tienen un componente aleatorio o cuando se utilizan técnicas como la validación cruzada (Canete-Sifuentes et al., 2021). Para abordar este problema, es recomendable establecer semillas aleatorias antes de cada ejecución para que los resultados sean reproducibles (Song & Lu, 2015). Esto se hace configurando la semilla aleatoria en el generador de números aleatorios utilizado por el algoritmo. En este caso, en Scikit-Learn (una biblioteca de aprendizaje automático en Python), se establece la semilla aleatoria usando la función `random_state` en los objetos de los modelos.

Validación de resultados

Se está utilizando la prueba de Tukey, que es una prueba estadística post hoc utilizada para realizar comparaciones múltiples entre las medias de varios grupos. La prueba de Tukey es comúnmente utilizada después de realizar un análisis de varianza (ANOVA) para determinar si existen diferencias significativas entre las medias de tres o más grupos.

En tu caso, la prueba de Tukey se está utilizando para comparar las medias de exactitud entre dos grupos: el grupo "Prueba" y el grupo "Validación Cruzada". La prueba de Tukey evalúa si hay diferencias estadísticamente significativas entre estas medias y ayuda a identificar qué grupos específicos tienen medias significativamente diferentes. Al aplicar este modelo, se tiene el output de la figura 8:

```
Desviación estándar de la exactitud: 0.07199324231850222
Matriz de Confusión del modelo ajustado:
[[ 2  0]
 [ 0 11]
 [ 0 17]]
Reporte de Clasificación del modelo ajustado:
              precision    recall  f1-score   support

Riego Aplicado (Bajo) en t1         1.00      1.00      1.00         2
Riego Aplicado (Medio) en t1        0.92      1.00      0.96        11
Riego Aplicado (Óptimo) en t1       1.00      0.94      0.97        18

   accuracy          0.97
  macro avg          0.97      0.98      0.98
 weighted avg          0.97      0.97      0.97

Análisis de Varianza (ANOVA) en la Exactitud:
Estadístico F: 0.9757294548050331

Prueba de Tukey:
Multiple Comparison of Means - Tukey HSD, FWER=0.05
-----
group1    group2    meandiff p-adj    lower upper reject
-----
Prueba Validación Cruzada -0.0871 0.3792 -0.3319 0.1577 False
```

Figura 8. Output Análisis estadístico

Desviación estándar de la exactitud: El valor de 0.07199324231850222 representa la desviación estándar de las exactitudes obtenidas durante la validación cruzada. Esto mide cuánto varían las exactitudes en los diferentes pliegues de validación cruzada, proporcionando una idea de la consistencia del rendimiento del modelo en diferentes divisiones del conjunto de datos.

Matriz de Confusión del modelo ajustado: Esta matriz muestra cómo el modelo clasificó las muestras en cada una de las clases. Los números en la matriz indican cuántas muestras se clasificaron correctamente o incorrectamente en cada clase.

Análisis de Varianza (ANOVA) en la Exactitud: Estos valores corresponden a los resultados del análisis de varianza realizado en las exactitudes entre los grupos de validación cruzada y el conjunto de prueba. El estadístico F calculado es 0.9757294548050331 y el valor p asociado es 0.3791749978317657. El valor p es mayor que un nivel de significancia típico de 0.05, lo que sugiere que no hay diferencias estadísticamente significativas en las exactitudes entre los dos grupos.

Prueba de Tukey: La prueba de Tukey se utiliza para realizar comparaciones múltiples entre los grupos y determinar si hay di-

ferencias significativas en las medias. En tu caso, se compara la exactitud entre el grupo "Prueba" y el grupo "Validación Cruzada". El resultado indica que no hay evidencia para rechazar la hipótesis nula de que no hay diferencias significativas en las medias de exactitud entre estos dos grupos. De esta manera, no hay una diferencia estadísticamente significativa en la exactitud entre la prueba y la validación cruzada.

En general, estos resultados indican que tu modelo está funcionando bien en la clasificación de las diferentes clases y que no hay diferencias estadísticamente significativas en las exactitudes entre la validación cruzada y el conjunto de prueba.

Discusión

La adopción de la tecnología IoT en el campo de la agricultura, aunque ha cobrado importancia aún no estaba bien integrada en el cultivo de café que plantea un manejo particularmente delicado por la sensibilidad del cultivo a las condiciones del entorno.

El aporte más significativo de este trabajo de investigación se encuentra en el hecho de integrar sensores IoT con algoritmos de machine learning para realizar predicciones en tiempo real sobre las necesidades de riego, de forma que se pueden ir adaptando las prácticas de riego a las condiciones particulares de cada cultivo, lo cual permite dejar atrás los sistemas tradicionales que no proporcionaban datos continuos, accediendo a datos a muy alta resolución; con ello se potencia una toma de decisiones más robusta y eficiente. También hay que señalar que la capacidad de implementar big data y machine learning en el ajuste del riego en función de factores como la radiación solar y la evapotranspiración aporta una innovación importante al éxito del riego que permite aportar una mejora significativa en la precisión de los sistemas de riego, además de dificultar el principio de sostenibilidad. En cuanto a la sostenibilidad, un aspecto que vale la pena destacar es la posibilidad de optimizar el uso del agua, un recurso cada vez más escaso

en muchas regiones agrícolas a causa de la explotación de los acuíferos. Al ser un sistema inteligente, se separa de los sistemas convencionales que normalmente tienen patrones fijos de riego, ya que responde a las dinámicas del ambiente y con ello reduce el malgasto de agua. De este modo, es posible salvar los recursos hídricos y mejorar la sostenibilidad del cultivo.

La aportación que esta investigación hace al conocimiento académico y científico es muy significativa. En primer lugar, se pasa a contribuir con la agricultura de precisión y las tecnologías de IoT dado que permite su aplicación a cultivos de alta demanda (como es el caso del café). En segundo lugar, se convierte en la base para generar soluciones más adaptativas y escalables a la gestión del riego de otras especies agrícolas. Y en tercer lugar, se genera un sistema replicable (que puede ser trasladado a otras regiones o condiciones agrícolas), ayudando a una mayor adopción de prácticas sostenibles a una escala global.

Conclusiones

El modelo de árbol de decisión fue ajustado con hiperparámetros óptimos {'max_depth': 8, 'min_samples_leaf': 2, 'min_samples_split': 5}, lo que sugiere que un árbol con una profundidad máxima de 6, una hoja mínima de tamaño 1 y una división mínima de 10 muestras en cada nodo proporciona el mejor equilibrio entre complejidad y capacidad de generalización.

Evaluando el rendimiento del Modelo en el Conjunto de Prueba, el modelo de árbol de decisión demostró un alto nivel de precisión en el conjunto de prueba, con una exactitud del 93.5%. Esto indica que el modelo es capaz de tomar decisiones acertadas en la clasificación de las decisiones de riego.

En cuanto al rendimiento por Clase, el modelo demostró una excelente sensibilidad y F1-score en las clases "Riego Aplicado (Bajo)" y "Riego Aplicado (Medio)", con una sensibilidad y F1-score del 100%. La clase

"Riego Aplicado (Óptimo)" también mostró un buen rendimiento con una sensibilidad del 88% y un F1-score del 94.1%.

La validación cruzada realizada en el modelo mostró una exactitud promedio del 83.3% con una desviación estándar de 0.07. El análisis de varianza (ANOVA) no reveló diferencias estadísticamente significativas en la exactitud entre los grupos de validación cruzada y el conjunto de prueba. Además, la prueba de Tukey no rechazó la hipótesis nula de que no hay diferencias significativas entre los grupos.

En función de los modelos utilizados y los resultados obtenidos, podemos concluir que el enfoque de Simulación de la toma de decisiones de riego inteligente para cultivos de café con la utilización de sensores IoT, demuestra la eficacia del modelo de árbol de decisión en la toma de decisiones de riego para diferentes condiciones de cultivo. La precisión y capacidad de clasificación del modelo, respaldadas por el análisis de validación cruzada y las pruebas estadísticas, refuerzan la viabilidad de este enfoque para optimizar la gestión de riego en los cultivos de café mediante la integración de sensores IoT y técnicas de aprendizaje automático.

El uso de sensores IoT en combinación con modelos de aprendizaje automático, como el árbol de decisión, permite una gestión eficiente e inteligente del riego en cultivos de café. Los sensores proporcionan datos clave sobre las condiciones del ambiente y del suelo, lo que permite predicciones precisas y decisiones optimizadas. Este enfoque mejora el uso de recursos hídricos y promueve la sostenibilidad agrícola

Recomendaciones

Dado que el modelo de árbol de decisión ha demostrado un alto rendimiento en la clasificación de decisiones de riego, se recomienda implementar esta metodología en el campo. Utilizando los hiperparámetros óptimos determinados durante el proyecto, el modelo puede ayudar a automatizar la

toma de decisiones de riego, optimizando el uso del agua y mejorando la eficiencia del proceso.

Para garantizar la precisión y eficacia del modelo de toma de decisiones, se debe establecer un sistema de monitoreo continuo utilizando sensores IoT. Esto permitirá recopilar datos en tiempo real sobre variables clave como humedad del suelo, temperatura, radiación solar y velocidad del viento. Los datos actualizados mejorarán la precisión del modelo y la capacidad de respuesta a las condiciones cambiantes del cultivo.

Aunque el modelo ha mostrado un rendimiento prometedor en el entorno de prueba, es fundamental validar su desempeño en condiciones reales en el campo. La implementación piloto en una parcela de cultivo de café permitirá evaluar cómo el modelo se adapta a situaciones variables y confirmar su eficacia en la toma de decisiones en tiempo real.

Después de la implementación en el campo, se debe realizar un seguimiento constante de la eficacia del modelo y las decisiones de riego que genera. Esto permitirá identificar posibles desviaciones y ajustar el modelo según sea necesario. Además, se pueden recopilar comentarios y experiencias de los agricultores para mejorar aún más el sistema.

Es esencial capacitar a los agricultores en el uso del sistema de toma de decisiones basado en el modelo de árbol de decisión. Los agricultores deben comprender cómo se generan las decisiones de riego y cómo interpretarlas. Esto fomentará la confianza en el sistema y su adopción exitosa.

Además de los beneficios ambientales y agronómicos, es importante llevar a cabo una evaluación económica del sistema de toma de decisiones. Esto incluye analizar los ahorros en costos de agua, la optimización de la producción y cualquier inversión adicional requerida para la implementación del sistema.

Bibliografía

- Bishnoi, S., & Hooda, B. K. (2022). Decision Tree Algorithms and their Applicability in Agriculture for Classification. *Journal of Experimental Agriculture International*. <https://doi.org/10.9734/jeai/2022/v44i730833>
- Canete-Sifuentes, L., Monroy, R., & Medina-Perez, M. A. (2021). A Review and Experimental Comparison of Multivariate Decision Trees. *IEEE Access*, 9. <https://doi.org/10.1109/ACCESS.2021.3102239>
- Charbuty, B., & Abdulazeez, A. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, 2(01). <https://doi.org/10.38094/jastt20165>
- Edwards-Murphy, F., Magno, M., Whelan, P. M., O'Halloran, J., & Popovici, E. M. (2016). B+WSN: Smart beehive with preliminary decision tree analysis for agriculture and honey bee health monitoring. *Computers and Electronics in Agriculture*, 124. <https://doi.org/10.1016/j.compag.2016.04.008>
- El Morr, C., Jammal, M., Ali-Hassan, H., & El-Hallak, W. (2022). Decision Trees. In *International Series in Operations Research and Management Science* (Vol. 334). https://doi.org/10.1007/978-3-031-16990-8_8
- Ferrández-Pastor, F. J., García-Chamizo, J. M., Nieto-Hidalgo, M., & Mora-Martínez, J. (2018). Precision agriculture design method using a distributed computing architecture on internet of things context. *Sensors (Switzerland)*, 18(6). <https://doi.org/10.3390/s18061731>
- Friha, O., Ferrag, M. A., Shu, L., Maglaras, L., & Wang, X. (2021). Internet of Things for the Future of Smart Agriculture: A Comprehensive Survey of Emerging Technologies. In *IEEE/CAA Journal of Automatica Sinica* (Vol. 8, Issue 4). <https://doi.org/10.1109/JAS.2021.1003925>
- Kalesanwo, O., Awodele, O., Eze, M., Kuyoro, 'S., & Ajaegbu, C. (2020). Evaluation of Decision Tree Algorithms in Precision Agriculture. *International Journal of Computing and Technology (IJCAT)*, 7(3).
- Kotsiantis, S. B. (2013). Decision trees: A recent overview. In *Artificial Intelligence Review* (Vol. 39, Issue 4). <https://doi.org/10.1007/s10462-011-9272-4>
- Song, Y. Y., & Lu, Y. (2015). Decision tree methods: applications for classification and prediction. *Shanghai Archives of Psychiatry*, 27(2). <https://doi.org/10.11919/j.issn.1002-0829.215044>
- Tao, W., Zhao, L., Wang, G., & Liang, R. (2021). Review of the internet of things communication technologies in smart agriculture and challenges. In *Computers and Electronics in Agriculture* (Vol. 189). <https://doi.org/10.1016/j.compag.2021.106352>
- Vamshi, H., Vamshi, D., Dakshiraju, S., Chennauahgari, S., Saswati, B., & Loganayagi, B. (2023). SMART AGRICULTURE USING DECISION TREE IN IOT. *International Research Journal of Modernization in Engineering Technology and Science*, 5(03/March-2023). <https://doi.org/10.56726/irj-mets35152>



CREATIVE COMMONS RECONOCIMIENTO-NOCOMERCIAL-COMPARTIRIGUAL 4.0.

CITAR ESTE ARTICULO:

Corrales Bonilla, J. I. ., Barreno Flores, C. F., Salvatierra Tacle, A. L., & Pastuña Chusin, N. M. (2025). Toma de decisiones de riego inteligente para cultivos de café con la utilización de sensores IoT. *RECIMUNDO*, 9(Especial), 378–394. [https://doi.org/10.26820/recimundo/9.\(esp\).mayo.2025.378-394](https://doi.org/10.26820/recimundo/9.(esp).mayo.2025.378-394)