

Modelo de agricultura inteligente utilizando minería de datos para identificar las condiciones que afectan un cultivo de café

Ing. Jocelyne Prieto Montiel¹, Dr. Eduardo Roldán Reyes², Dr. Guillermo Cortés Robles³, Dra. Edna Araceli Romero Flores⁴,
M.A.E. Maricela Gallardo Córdova⁵

Resumen –

En la actualidad, la caficultura se enfrenta a algunos problemas: el cambio medioambiental, la escasez de agua, las plagas, entre otros. Ante estas problemáticas es crucial incorporar procesos innovadores que permitan resolver y proponer soluciones que aumenten el valor agregado de los cultivos.

En este artículo se aborda a la agricultura inteligente mediante la aplicación de un enfoque basado en la minería de datos para explotar la información proveniente del monitoreo en campo de variables de cultivo. Para ello se utilizaron los datos recopilados por una estación de monitoreo y utilizando el algoritmo de agrupamiento K-Means y de clasificación K-NN, se identifican los factores que afectan al crecimiento del café. Se recopila la base de datos sobre el sensor donde se realizó la investigación. Se describe el desarrollo del sistema de análisis, así como los resultados obtenidos.

Palabras claves – Agricultura inteligente, Minería de datos, Agrupamiento, Clasificación, Cultivo de café.

Abstract –

Currently, coffee growing is facing some problems: environmental change, water scarcity, pests, among others. In the face of these problems, it is crucial to incorporate innovative processes to solve and propose solutions that increase the added value of crops.

This article addresses smart agriculture by applying an approach based on data mining to exploit information from field monitoring of crop variables. For this purpose, data collected by a monitoring station were used and using the K-Means clustering and K-NN classification algorithm, the factors affecting coffee growth are identified. The database on the sensor where the research was conducted is compiled. The development of the analysis system is described, as well as the results obtained.

Keywords – Intelligent agriculture, Data mining, Clustering, Classification, Coffee cultivation.

I. INTRODUCCIÓN

Hoy en día, conocer, monitorizar y registrar la información en campo permite identificar factores que pueden ser aprovechados para mejorar las condiciones de cultivo.

Se desarrolló un modelo de agricultura inteligente el cual integra la minería de datos con un prototipo que facilita la obtención de los datos de las variables monitoreadas. Además, la monitorización de los datos obtenidos a través de sensores situados en los campos permite ayudar a los agricultores a adquirir mejores resultados y beneficios.

Generalmente, los agricultores han dependido de su experiencia e instinto para completar las diferentes diligencias en el campo, siguiendo un calendario fijo año con año. Gracias a la utilización de las nuevas tecnologías en materia de macrodatos, se pueden tomar decisiones de inmediato mediante la detección, el almacenamiento y el análisis de

¹ F. A. Author is with Tecnológico Nacional de México - Campus Orizaba, Orizaba, Ver 94500 MX (corresponding author to provide phone: 271-211-4382; e-mail: jprietom@orizaba.tecnm.mx).

² S.B Author is with Tecnológico Nacional de México - Campus Orizaba, Orizaba, Ver 94500 MX (corresponding author to provide e-mail: jprietom@orizaba.tecnm.mx).

³ T. C. Author is with Tecnológico Nacional de México - Campus Orizaba, Orizaba, Ver 94500 MX (corresponding author to provide e-mail: gc_robles@hotmail.com).

⁴ F. D. Author is with Tecnológico Nacional de México - Campus Orizaba, Orizaba, Ver 94500 MX (corresponding author to provide e-mail: eromerof@orizaba.tecnm.mx).

⁵ F. E. Author is with Tecnológico Nacional de México - Campus Orizaba, Orizaba, Ver 94500 MX (corresponding author to provide e-mail: maricelagal@gmail.com).

datos en tiempo real procedentes de los sensores que se distribuyeron por todo el terreno.

Las ventajas que se obtienen de la agricultura inteligente se pueden resumir en la optimización de recursos, el aumento de la producción, la disminución de los costes relacionados con ésta, y la mejora de la calidad del producto final.

El modelo de agricultura inteligente, en conjunto con el prototipo, facilitará la obtención de datos de las variables y por consiguiente al analizarlas detectarán si es necesario intervenir en el cultivo.

Una de las técnicas de minería de datos, consiste en la división de los datos en grupos de objetos similares. Cuando se representan la información obtenida a través de clusters se pierden algunos detalles de los datos, pero a la vez se simplifica dicha información.

Se dispone de un conjunto de datos de entrenamiento y cada dato está asociado a una etiqueta. Se construye un modelo en la fase de entrenamiento utilizando dichas etiquetas, que dicen si está clasificada correcta o incorrectamente por el modelo. Una vez construido el modelo se puede utilizar para clasificar nuevos datos que, en esta fase, ya no necesitan etiqueta para su clasificación, aunque sí la necesitan para evaluar el porcentaje de objetos bien clasificados.

El resultado del modelo nos arroja dos clases, las cuales son cluster 0 y cluster 1, que significan en el caso del cluster 0 que no requiere intervención, por otra parte, el cluster 1 necesita intervención del usuario.

II. TRANSFONDO TEÓRICO

A. Agricultura inteligente

La agricultura inteligente es un enfoque reciente que busca incorporar tecnologías de información para hacer más eficientes las actividades en el contexto de la agricultura.

Además, es complementaria de la agricultura de precisión, la cual ha tenido amplia aplicación para mejorar el rendimiento de los cultivos, utilizando por ejemplo tecnología de vehículos aéreos no tripulados (VANT). [1] Por ejemplo, (Morales, 2016) proponen la utilización de estos vehículos, conocidos como drones, para capturar imágenes de los campos de cultivo. Las imágenes de los vuelos se procesan para generar un orto mosaico en formato Geo TIFF.

Una aplicación de la agricultura inteligente es con respecto al clima, en donde se busca mejorar la gestión de los paisajes (las tierras de cultivo, la ganadería, los bosques y la pesca) que representan dos desafíos interrelacionados: la seguridad alimentaria y el cambio climático.[2] Por otro lado, [3] describe un dispositivo construido en una minicomputadora Raspberry, capaz de controlar sensores y recopilar mediciones de las variables observadas. Este trabajo previo sirve de base para la propuesta que se describe en este artículo, el cual es una continuación para incorporar la minería de datos como un enfoque para aplicar la agricultura inteligente.

B. Minería de datos

Por su parte, la Minería de Datos intenta descubrir a partir de bases de datos, conocimiento que puede ser potencialmente útil para la toma de decisiones. Surge como una nueva herramienta que propone una solución en el tratamiento de la información, particularmente apta cuando las herramientas estadísticas convencionales no son suficientes.

La Minería de Datos (del inglés *Data Mining*) es una disciplina que ha tenido un desarrollo vertiginoso gracias a los avances en diversas áreas tales como la informática, la estadística y la inteligencia artificial. Muchas de las metodologías utilizadas en la Minería de Datos provienen principalmente de dos ramas de investigación, la primera es la desarrollada dentro de la comunidad del aprendizaje automático (del inglés “*machine learning*”) y la otra de la comunidad estadística, particularmente en estadística multivariada y computacional.

El aprendizaje automático, o también llamado aprendizaje de máquinas, está íntimamente relacionado con la ciencia computacional y la inteligencia artificial. Consiste en desarrollar algoritmos que permitan encontrar relaciones y regularidades en los datos y que pueden transformarse en verdades generales. [4]

Una definición [5] plantea a la Minería de Datos como un proceso para detectar conocimiento en grandes cantidades de información que se encuentran almacenadas en bases de datos y bajo una estructura formal, que se vuelve útil para la toma de decisiones inteligentes.

Algunos de los conceptos que se manejan en la Minería de Datos son:

- *Algoritmos*: mecanismos o series de instrucciones que crean modelos de Minería de Datos. Para crear un modelo, un algoritmo analiza primero un conjunto de datos, buscando patrones y tendencias específicos, después, el algoritmo utiliza los resultados de este análisis para definir los parámetros del modelo de Minería de Datos.
- *Base de datos*: conjunto de datos que pertenecen al mismo contexto, almacenados sistemáticamente para su posterior uso.
- *Patrones*: la forma esperada del comportamiento de un conjunto de datos analizados previamente.
- *Instancia*: se refiere a un elemento particular (i.e. registro o ítem) dentro de un conjunto o base de datos.
- *Atributo*: es una característica particular que posee un determinado elemento o instancia.
- *Clase*: es el atributo o variable que se desea conocer mediante la búsqueda de un patrón en una base de datos.

C. El proceso de minería de datos.

Para llevar a cabo una aplicación exitosa de la Minería de Datos, es necesaria la implementación de un proceso sistemático bien organizado. Diferentes actores del medio industrial y académico han propuesto diversos procesos para maximizar el éxito de los proyectos. Estos esfuerzos han dado lugar a varios procesos estandarizados. Un ejemplo es la metodología “Cross-Industry Standard Process for Data Mining - CRISP-DM”. [6]

Esta metodología está compuesta por seis fases:

Entendimiento de negocios. En esta fase inicial se realiza una comprensión de los objetivos y requerimientos de los proyectos de Minería de Datos desde el punto de vista de la empresa y sus necesidades de conocimiento para establecer una definición del problema y un plan preliminar.

Comprensión de Datos. El principal objetivo de esta fase es familiarizarse con los datos con el fin de detectar problemas de calidad, u otros grupos de datos interesantes que puedan sumar al proyecto, así como la formulación de hipótesis.

Preparación de datos. Aquí se realiza una previa preparación donde se construye un conjunto de datos a partir de los iniciales (datos brutos), esto se realiza con diversos modelos que se encargan de realizar una depuración de datos.

Modelado. Fase en la cual se hace una selección de las diversas técnicas de modelado. Existen técnicas que muy a menudo resuelven los problemas de Minería de Datos, algunas de estas solicitan los datos con una cierta estructura o requerimientos.

Evaluación. Es importante realizar una evaluación del modelo que se utilizó y analizar los pasos que se llevaron a cabo para asegurar que se alcanzaron los objetivos establecidos en la primera fase, así como determinar que no existe algún conflicto o problema importante con la empresa, que no se haya considerado.

Despliegue. La información deberá ser presentada de tal forma que la empresa pueda utilizar estos datos. El despliegue puede ser tan rápido y fácil o complejo dependiendo de los requerimientos. En la mayoría de los casos es la empresa la encargada de efectuar la implementación.

D. Las aplicaciones para la minería de datos

La Minería de Datos se ha convertido en una herramienta muy conocida para resolver problemas en las empresas. Ha demostrado ser útil en diferentes áreas como se muestra a continuación: [7]

Administración basada en la relación del cliente. CRM siglas del término en inglés *Customer Relationship Management*. Su objetivo principal es crear un trato uno a uno con el cliente, estableciendo una relación íntima de sus necesidades y deseos. A través de la construcción de una relación en el tiempo identificando una variedad de transacciones como: solicitudes de servicios, servicios de garantía y todos aquellos datos que van en aumento. Estos documentos traen consigo importante información, que a simple vista es difícil detectar, pero con estos datos es posible identificar a compradores potenciales de nuevos productos y/o servicios, entender cuáles son los motivos por los que los clientes están dejando de comprar, así como identificar a clientes rentables y sus necesidades, dando la oportunidad de fortalecer relaciones e incrementar las ventas.

Finanzas. La Minería de Datos puede ayudar a los bancos a procesar las solicitudes de préstamos e identificar a las personas morosas con mayor exactitud y rapidez, así mismo ayuda a la detección de operaciones fraudulentas, también identifica que productos el cliente pueda llegar a adquirir con base a los servicios y/o productos con los que cuenta. Puede además predecir cuándo y cómo la lista de precios y cotizaciones cambiará. Analizar el comportamiento de los diferentes mercados bursátiles, entre otros.

Ventas y logística: Se utiliza para predecir los volúmenes de venta con el fin de establecer niveles de inventarios, además de permitir la identificación de las relaciones entre las ventas de productos con el fin de determinar la mejor distribución de estos. Por otro lado, puede servir para la previsión de niveles de consumo de diferentes tipos de productos (basados en condiciones estacionales o ambientales) o para mejorar aspectos logísticos y de marketing.

Manufactura y producción: Es un excelente medio para la predicción de fallas en las máquinas utilizando la información arrojada por sensores. Sirve además para identificar las anomalías y similitudes en los sistemas de producción con el fin de optimizar la capacidad de fabricación. Otra aplicación puede ser la de descubrir nuevos patrones para identificar y mejorar la calidad en los productos.

Agricultura:

En [8] se describe un estudio que presenta algunas de las técnicas de minería de datos más utilizadas en el campo de la agricultura. Se discuten algunas de estas técnicas, como la k-means, la k nearest neighbor, las redes neuronales artificiales y las máquinas de vectores de apoyo, y se presenta una aplicación en agricultura para cada una de estas técnicas. La minería de datos en la agricultura es un campo de investigación relativamente novedoso. En nuestra opinión, es posible desarrollar y adaptar técnicas eficaces para resolver problemas agrícolas complejos mediante la minería de datos. Al final de este estudio se ofrecen recomendaciones para futuras direcciones de investigación en campos relacionados con la agricultura.

La agricultura de precisión [9] es la aplicación de la tecnología GPS más avanzada en relación con el tratamiento del cultivo basado en sensores y específico para cada lugar. También puede describirse como un enfoque de la agricultura basado en datos, que está fuertemente relacionado con una serie de problemas de minería de datos. Uno de ellos es también una tarea inherentemente importante en la agricultura: la predicción del rendimiento.

Este artículo ofrece una breve visión general de los datos disponibles, señala en detalle el principal problema de los enfoques de aprendizaje clásicos y presenta una novedosa técnica de validación cruzada espacial para superar los problemas del enfoque clásico en la mencionada tarea de predicción del rendimiento.

En [10] el sistema WEKA proporciona un conjunto completo de facilidades para aplicar técnicas de minería de datos a grandes conjuntos de datos. Este documento habla de un modelo de interacción para la investigación de la información y describe la ayuda que WEKA acomoda a este modelo. El modelo de dominio "aprendido" por el algoritmo de minería de datos puede ser efectivamente consolidado a una aplicación de software. Este proceso de análisis y construcción de aplicaciones basado en WEKA.

E. Herramientas para la aplicación de la minería de datos.

Actualmente, existen muchas herramientas de software para el desarrollo de modelos de Minería de Datos tanto libres como comerciales, por ejemplo:

Weka. Es una de las herramientas más conocida en el medio. Es de distribución libre y de código abierto. Contiene varios algoritmos para Minería de Datos, así como herramientas para el preprocesamiento, clasificación, regresión, agrupamiento, reglas de asociación y visualización. Su principal ventaja es que es un software libre de código abierto. [11]

RapidMiner. En los últimos años, esta herramienta está teniendo una gran difusión en el medio académico. En ella se incluye además de los algoritmos de Minería de Datos, extensiones que permiten realizar análisis de Minería de Textos y Minería de la Web (Web Mining), además de poder interpretar los algoritmos de Weka. Su entorno es más robusto, y a la vez, fácil de manejar por los usuarios. Es de distribución libre bajo licencia GNU.

InfiniteInsight. Herramienta de análisis predictivo que reduce el esfuerzo de realizar actividades de forma manual y repetitiva para tomar decisiones. Algunos de sus enfoques son la optimización del ciclo de vida del cliente, la reducción del riesgo y prevención del fraude, la recomendación de productos, la optimización de palabras claves, el marketing, la recomendación de productos y el mejoramiento del diseño de un producto utilizando la voz del cliente. [12] Este software

tiene como objetivo solucionar los problemas reales de la empresa. Su principal desventaja es que solo se enfoca al área del cliente, olvidando las demás áreas de la empresa.

III. METODOLOGÍA

La metodología que se utilizó en este trabajo se inspira de la metodología CRISP y está basada en 4 etapas. El modelo se desarrolló en el software RapidMiner.

1. Adquisición de datos: Los datos se obtuvieron de 3 fuentes:
 - a. Las mediciones realizadas con el prototipo desarrollado RaspBerry [3]. Las variables que se tomaron en cuenta para las mediciones son: Humedad, Temperatura ambiental, Suelo (ph) y Precipitación, las cuales se monitorizan cada 2 horas.
 - b. Los expertos agricultores del café.
 - c. La generación de datos mediante la distribución triangular. Debido a las condiciones derivadas de la pandemia, fue difícil continuar con el monitoreo en campo, por lo que se generaron datos haciendo uso del procedimiento de la distribución triangular.
2. Preprocesamiento: En esta etapa la calidad de los datos es mejorada mediante la aplicación de subprocesos de filtrado, eliminación de outliers, normalización.
3. Procesamiento: en esta etapa es donde se aplican los algoritmos de minería de datos para la identificación de las condiciones que requieren intervención en campo. Se aplicó el algoritmo k-Means para la identificación de grupos. Se generaron dos grupos, 1) para el conjunto de observaciones cuyas condiciones no requieren intervención (Clúster_0) y 2) un grupo para aquel conjunto en los que se requiere intervención (Clúster_1). El segundo algoritmo que se aplica es el K vecinos más cercanos (K-NN), este algoritmo se aplica solo en el segundo grupo, es decir aquél en el que se requiere la intervención. Para ello, se compara con otra base en la cual previamente han sido reconocidas las condiciones en las que se requiere intervención y en qué consistió dicha intervención. Entonces el algoritmo clasifica las nuevas instancias y determina que tipo de intervención en campo se requiere tomando como parámetro la similitud entre la base de observaciones de referencia. En este paso se realiza la capitalización de experiencias previas, al aprovechar las experiencias pasadas.
4. Resultados y evaluación: Los resultados obtenidos son desplegados y validados por el experto. Estos resultados son a su vez integrados en la base de datos, y servirán como referencias para futuras intervenciones.

IV. CASO DE ESTUDIO Y RESULTADOS

El experimento se desarrolló en un cultivo de café ubicado en la zona centro del estado de Veracruz, México. Las variables que se tomaron en cuenta para las mediciones son: Humedad, Temperatura ambiental, Suelo(ph) y Precipitación.

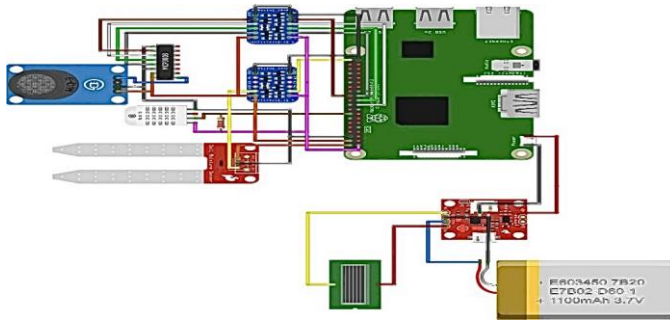


Figura 1. Prototipo

Cabe mencionar que, por causas ligadas a la pandemia, fue difícil la intervención en campo, por lo que una vez generadas ciertas variables se recurrieron a la generación de datos mediante la distribución triangular.

Temperatura	Humedad	Suelo	Lluvia
18	90	35	0.9
19	90	36	0.9
19	92	38	0.3
19	91	30	0.3
19	85	25	0.3
20	84	25	0
20	81	27	0
20	79	27	0
23	82	27	0
24	83	22	0
26	76	15	0
27	65	12	1.2
27	62	5	1.2
26	61	5	1.2
26	60	3	0.6
25	60	3	0.6
22	61	2	0.6

Tabla 1 Extracto de las variables monitoreadas y sus valores.

En la siguiente figura se puede observar el modelo de minería de datos que, en conjunto con el prototipo, permitieron la obtención de las variables, lo que nos permitirá analizarlas y por consiguiente detectar si es necesario intervenir en la producción del café.

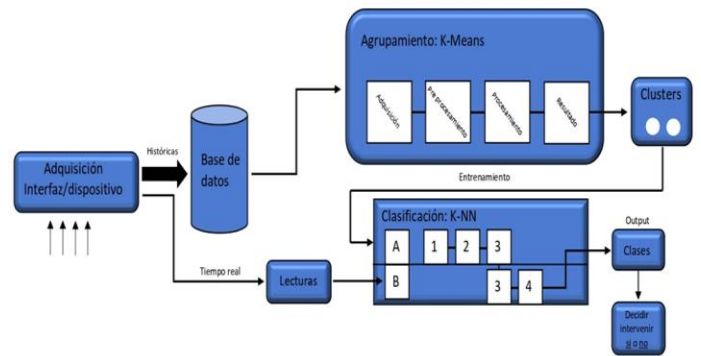


Figura 2. Modelo de minería de datos

Agrupamiento

Para el agrupamiento de los datos observados se aplicó el algoritmo K-Means. El valor de k se estima en 2 ya que se busca identificar los dos grupos, el primero como aquellos que requieren intervención y el segundo aquellos que no lo requieren. Para la generación de grupos se calculan las distancias de las instancias a cada uno de los centroides mediante la distancia euclidiana (Ecuación 1). Las instancias se asignan a aquellos grupos cuya distancia es mínima con respecto a todos los centroides.

$$(A, B) = \sqrt{\sum_i (x_A - x_B)^2 + (y_A - y_B)^2}$$

Ecuación 1 Distancia euclidiana

Luego se actualizan los centroides con el valor medio de todas las instancias asignadas a ese grupo. Se repitió el paso 2 y 3 hasta que se satisface el criterio de convergencia. Las razones de utilizar el algoritmo k-means son que es computacionalmente rápido para generar los grupos, y trabaja bien con datos faltantes.

Clasificación

Para la clasificación se utilizan el algoritmo k-NN, el cual clasifica un nuevo punto en función de las clases conocidas de los puntos más próximos a él. Para ello se siguen los siguientes pasos:

1. Se escoge un número (k) de vecinos próximos. En el experimento se escogieron los valores de 1, 3, y 5.
2. Se elige la medida de calcular la similitud, es decir, una función para calcular la distancia entre dos instancias. Se utiliza la distancia euclidiana (Ecuación 1)
3. Para la instancia a clasificar:
 - Se calcula la distancia con el resto de las instancias.
 - Se seleccionan los k vecinos más cercanos.

Se determina la clase de la nueva instancia como la más representada entre los k vecinos. Resolución de empates. Si coincide el número de vecinos de dos o más clases, se consideran más valores para k y se evalúan las clasificaciones.

Resultados del modelo

El resultado del modelo arroja dos clases, las cuales son cluster 0 y cluster 1, que significan en el caso del cluster 0 que no requiere intervención, por otra parte, el cluster 1 necesita intervención del usuario. En la tabla 2 se presenta un extracto del agrupamiento de las observaciones en campo.

Temperatura	Humedad	Suelo	Precipitación	label
22.0	68.0	15.0	.5	cluster_0
20.0	72.0	14.0	.3	cluster_0
24.0	82.0	4.0	.4	cluster_0
22.0	86.0	18.0	.7	cluster_0
22.0	83.0	12.0	.9	cluster_0
21.0	81.0	14.0	1.1	cluster_0
20.0	75.0	13.0	.8	cluster_0
18.0	79.0	31.0	1.0	cluster_1
19.0	73.0	28.0	.4	cluster_1
23.0	81.0	26.0	.3	cluster_0

Tabla 2 Extracto del agrupamiento de las observaciones en campo.

Posteriormente, las clasificaciones realizadas a las instancias del clúster_1 se despliegan en la tabla 3.

Temperatura	Humedad	Suelo	Precipitación	Instancias	label
24	82	4	0.37	3	vigilar ph
22	86	18	0.66	4	verificar humedad
22	83	12	0.89	5	vigilar ph
21	81	14	1.05	6	Verificar humedad
19	75	13	0.81	7	vigilar temperatura
18	79	31	1.02	8	vigilar temperatura
19	73	28	0.4	9	vigilar temperatura

Tabla 3 Clasificaciones realizadas a las instancias del clúster_1

V. CONCLUSIONES

Anteriormente, era necesario recurrir a prácticas manuales, las cuales además de ser laboriosas y lentas, se prestaban a diversos errores debidos a la falta de registros y analistas, además de que, si algún dato llegaba a perderse, podía afectar toda la investigación.

La minería de datos trajo la solución a estos problemas, proponiendo un análisis veloz, sencillo y automatizado. También cabe mencionar que, entre más complejos son los datos recopilados, es más sencillo realizar análisis con resultados relevantes para la investigación realizada. Esta técnica se puede aplicar a cualquier investigación que requiera el descubrimiento de un patrón de comportamiento.

Del mismo modo, con esta herramienta es posible encontrar la relación entre un evento y los datos registrados, o bien, la relación que existe entre las mismas variables obtenidas, permitiendo a los investigadores y estudiosos manipular la información para conseguir una solución al problema que se les presenta.

Al momento en que estos mismos se ven capaces de gestionar y manejar el riesgo, pueden generar un cambio, para

que el resultado de la minería de datos, culmine en la erradicación del problema o el nacimiento de una nueva propuesta para la corrección y mejoramiento de las variables obtenidas.

Sin duda, con este análisis, el estudio del comportamiento del suelo en cafetales permitirá a las personas encargadas del cuidado de los cultivos, dar mayor atención a ciertos aspectos, permitiendo así una mejora en el producto que se estudia.

TRABAJOS FUTUROS

Como trabajo futuro se pretende mejorar el prototipo de monitoreo, incorporar un mecanismo más sofisticado para la propuesta de soluciones. Se plantea además utilizar el enfoque de los sistemas de recomendación los cuales permiten realizar recomendaciones ante situaciones futuras utilizando algoritmos específicos.

VI. REFERENCIAS

- [1] A. Morales, «Ejemplos de uso de drones en GIS», *MappingGIS*, sep. 16, 2014. <https://mappinggis.com/2014/09/ejemplos-de-uso-de-drones-en-gis/> (accedido abr. 15, 2021).
- [2] «Agricultura inteligente con respecto al clima», *World Bank*. <https://www.bancomundial.org/es/topic/climate-smart-agriculture> (accedido abr. 15, 2021).
- [3] B. Suazo Debernardi, «Aplicación de agricultura de precisión para medir el rendimiento del cultivo del café». 2020.
- [4] P. Giudici, «Applied Data Mining: Statistical Methods for Business and Industry | Wiley», *Wiley.com*. <https://www.wiley.com/en-us/Applied+Data+Mining%3A+Statistical+Methods+for+Business+and+Industry-p-9780470871393> (accedido abr. 15, 2021).
- [5] M. J. A. Berry y G. Linoff, *Data mining techniques: for marketing, sales, and customer relationship management*, 2nd ed. Indianapolis, Ind: Wiley Pub, 2004.
- [6] C. Catley, K. Smith, C. McGregor, y M. Tracy, «Extending CRISP-DM to incorporate temporal data mining of multidimensional medical data streams: A neonatal intensive care unit case study», en *2009 22nd IEEE International Symposium on Computer-Based Medical Systems*, ago. 2009, pp. 1-5, doi: 10.1109/CBMS.2009.5255394.
- [7] E. Turban, R. Sharda, y D. Delen, *Decision Support and Business Intelligence Systems*, 9th ed. USA: Prentice Hall Press, 2010.
- [8] A. Mucherino, P. Papajorgji, y P. M. Pardalos, «A survey of data mining techniques applied to agriculture», *Oper. Res.*, vol. 9, n.º 2, pp. 121-140, ago. 2009, doi: 10.1007/s12351-009-0054-6.
- [9] G. Ruß y A. Brenning, «Data Mining in Precision Agriculture: Management of Spatial Information», en

Computational Intelligence for Knowledge-Based Systems Design, Berlin, Heidelberg, 2010, pp. 350-359, doi: 10.1007/978-3-642-14049-5_36.

- [10] S. Cunningham y G. Holmes, «Developing innovative applications in agriculture using data mining», 1999. /paper/Developing-innovative-applications-in-agriculture-Cunningham-Holmes/85ad11fb3c4c29cb1c775d9ca47d2adccfe42408 (accedido abr. 15, 2021).
- [11] «Weka 3 - Data Mining with Open Source Machine Learning Software in Java». <https://www.cs.waikato.ac.nz/ml/weka/> (accedido abr. 15, 2021).
- [12] «SAP Software Solutions | Business Applications and Technology», SAP. <https://www.sap.com/index.html> (accedido abr. 15, 2021).