

# **Sistema de recomendación para la aplicación de fertilizantes nitrogenados en el cultivo de café colombiano**



Proyecto de Trabajo de Grado

**Luis Felipe Bastidas Torres**

**Edinson David León Chilito**

Director: PhD. Cristhian Nicolás Figueroa Martínez

Codirector: PhD. Juan Carlos Corrales

Asesor: Mg. Juan Fernando Casanova

Departamento de Telemática

Facultad de Ingeniería Electrónica y Telecomunicaciones

Universidad del Cauca

Popayán, Cauca

Marzo de 2022

# **Sistema de recomendación para la aplicación de fertilizantes nitrogenados en el cultivo de café colombiano**

**Luis Felipe Bastidas Torres**

**Edinson David León Chilito**

Trabajo de grado presentado a la Facultad de Ingeniería Electrónica y Telecomunicaciones de la Universidad del Cauca para optar por el título de: Ingeniero en Electrónica y Telecomunicaciones

Director: PhD. Cristhian Nicolás Figueroa Martínez

Codirector: PhD. Juan Carlos Corrales

Asesor: Mg. Juan Fernando Casanova

Departamento de Telemática

Facultad de Ingeniería Electrónica y Telecomunicaciones

Universidad del Cauca

Popayán, Cauca

Marzo de 2022

# Agradecimientos

Queremos agradecer a Dios por nuestras vidas, a nuestras madres, padres, hermanos y amigos por todo su apoyo a lo largo de la carrera. También a nuestro director, el Ing. Cristhian Nicolás Figueroa Martínez, nuestros asesores, el Ing. Juan Fernando Casanova, la Bióloga Cristina Ordóñez y la empresa Ecotecma por su colaboración en este proyecto. La cual permitió tener un acceso a datos de cultivos de café en los cuales ellos trabajan, siendo un insumo importante para el desarrollo de nuestra investigación.

# Resumen

El crecimiento demográfico ha impactado fuertemente en la demanda de alimentos a nivel mundial y ha obligado a los productores agrícolas a mejorar sus niveles de producción para solventar las necesidades alimentarias y mantener el crecimiento económico. En consecuencia, muchos productores han incrementado la aplicación de fertilizantes minerales en sus cultivos de manera excesiva e indiscriminada debido a la falta de capacitación en innovaciones continuas. Lo anterior tiene profundos efectos en el medio ambiente, ya que la aplicación de fertilizantes es una de las principales causas de contaminación del agua por nitratos, fosfatos y plaguicidas, y de emisión de gases de efecto invernadero como metano y óxido nitroso. En especial, los fertilizantes nitrogenados representan una de las fuentes de contaminación que más impactan sobre la atmósfera, el suelo y el agua.

En Colombia uno de los cultivos más importantes en la economía es el café, el cual ha sido objeto de amplia investigación con el fin de mejorar su producción reduciendo su impacto en el medio ambiente. Por esta razón, es fundamental comprender variables relacionadas con este cultivo como el estado del suelo, la siembra, la cobertura de sombrero, el control de plagas, las enfermedades, la fertilización, la afectación por factores climáticos, entre otros aspectos; sirve para sacarle el mejor provecho posible a la producción del café. Para esto, es necesaria la implementación de tecnologías de la información que faciliten al agricultor el acceso a estas variables y tomar decisiones sobre su cultivo.

El presente trabajo de grado, propuso un sistema de recomendación de la cantidad de fertilizantes a aplicar a un cultivo de café considerando aspectos agroclimáticos y una base de conocimiento desarrollada a partir de entrevistas con expertos del dominio y estudios científicos. Dicha base de conocimiento fue consolidada en un sistema de razonamiento por casos (CBR) con el cual se pudo generar las recomendaciones de fertilizante de N, P y K a aplicar en el cultivo. Adicionalmente, el sistema realizó un análisis de temporada climática para recomendar la frecuencia de aplicación de estos fertilizantes más adecuada de acuerdo a las condiciones climáticas.

Finalmente, el trabajo realizado hace una contribución importante en el desarrollo agrícola, ya que brindar información de clima, suelo y cultivo ayuda a los agricultores a tomar mejores decisiones para beneficio del manejo y producción de sus cultivos.

# Abstract

Population growth has strongly impacted global food demand and has forced farmers to improve their production levels to meet food needs and maintain economic growth. Consequently, many producers have increased the application of mineral fertilizers on their crops in an excessive and indiscriminate manner due to the lack of training in continuous innovation. This has profound effects on the environment, since fertilizer application is one of the main causes of water pollution by nitrates, phosphates and pesticides, and the emission of greenhouse gases such as methane and nitrous oxide. In particular, nitrogen fertilizers are one of the sources of pollution that have the greatest impact on the atmosphere, soil and water.

In Colombia, one of the most important crops in the economy is coffee, which has been the object of extensive research with the aim of improving its production and reducing its impact on the environment. For this reason, it is fundamental to understand variables related to this crop such as the state of the soil, planting, shade cover, pest control, diseases, fertilization, and the effects of climatic factors, among other aspects, in order to take the best possible advantage of coffee production. For this, it is necessary to implement information technologies that facilitate the farmer's access to these variables in order to make decisions about his crop.

The present work proposed a system for recommending the amount of fertilizers to apply to a coffee crop considering agro-climatic aspects and a knowledge base developed from interviews with experts in the field and scientific studies. This knowledge base was consolidated in a case-based reasoning (CBR) system that was used to generate the recommendations for N, P and K fertilizer to be applied to the crop. Additionally, the system performed a climatic season analysis to recommend the most appropriate application frequency of these fertilizers according to climatic conditions.

Finally, the work performed makes an important contribution to agricultural development, since providing climate, soil and crop information helps farmers make better decisions to benefit the management and production of their crops.

# Tabla de contenido

<b>1. INTRODUCCIÓN</b>	11
1.1. Planteamiento del problema	11
1.2. Antecedentes	12
1.3. Aportes	13
1.4. Objetivos	13
1.4.1. Objetivo general	13
1.4.2. Objetivos específicos	13
1.5. Estructura del documento	13
<b>2. CONTEXTO TEÓRICO Y ESTADO DEL ARTE</b>	15
2.1. Contexto teórico	15
2.1.1. Sistema de recomendación	15
2.1.2. Sensores en la agricultura	18
2.1.3. Tecnologías de redes inalámbricas existentes	18
2.2. Estado del arte	22
2.2.1. Revisión sistemática de la literatura	22
2.2.2. Extracción de datos	24
2.2.3. Trabajos relacionados	25
2.2.4. Brechas encontradas	29
<b>3. INTERPRETACIÓN DEL SISTEMA Y DATOS NECESARIOS</b>	31
3.1. Definición del sistema	31
3.2. Metodología	32
3.3. Comprensión del negocio	33
3.3.1. Fertilización en el cultivo de café	33
3.3.2. Períodos de fertilización	35
3.4. Comprensión de los datos	36
3.4.1. Análisis de fuentes de datos	36
3.4.2. Recopilación documental	36
3.4.3. Recolección de datos	36
3.4.3.1. Determinación de variables	37
3.4.3.2. Captura de datos	41
3.4.4. Descripción de los datos	45
<b>4. PREPARACIÓN DE LOS DATOS</b>	47
4.1. Datos de modelado del sistema	47
4.1.1. Estructura de un caso	47
4.1.2. Base de casos	49
4.2. Datos de entrada del sistema	51
4.2.1. Clasificación de temporada	52

4.2.2. Construcción de un nuevo problema	56
4.2.3. Normalización del problema	57
<b>5. MODELADO</b>	<b>58</b>
5.1. Recuperación	59
5.2. Reutilización	62
5.3. Recomendaciones	64
<b>6. EVALUACIÓN</b>	<b>67</b>
6.1. Determinación del factor k	67
6.2. MAE y RMSE	70
6.3. Prueba del sistema	72
<b>7. CONCLUSIONES Y TRABAJOS FUTUROS</b>	<b>77</b>
7.1. Conclusiones	77
7.2. Limitaciones	78
7.3. Trabajos futuros	78
<b>REFERENCIAS</b>	<b>80</b>

# Lista de figuras

Figura 1	Proceso de RSL. Adapdado de la metodología Kitchenham	23
Figura 2.	Metodología CRISP-DM. Adaptada de [60].	32
Figura 3.	Diagrama de un año cafetero en el Cauca. Adaptado de [71].	35
Figura 4.	Diagrama del proceso de determinación de variables del sistema. Fuente propia.	37
Figura 5.	Variables de entrada más comunes encontradas en la RSL. Fuente propia.	38
Figura 6.	Proceso de captura de datos del sistema. Fuente propia.	41
Figura 7.	Arquitectura de red de sensores Ecotecma. Adaptado de [83].	44
Figura 8.	Proceso de construcción de un nuevo problema. Fuente propia.	51
Figura 9.	Registros históricos de la NWS. Fuente [96].	52
Figura 10.	Proceso de clasificación de temporada. Fuente propia.	54
Figura 11.	Diagrama de modelado del sistema CBR. Fuente propia.	58
Figura 12.	Proceso de recuperación del sistema CBR. Fuente propia.	60
Figura 13.	Proceso de reutilización del sistema CBR. Fuente propia.	62
Figura 14.	División de la base de casos. Fuente propia.	67
Figura 15.	Precisión del sistema ante diferentes valores de k.	68
Figura 16.	Precisión del sistema ante una variación de la base de casos.	69
Figura 17.	Casos más similares encontrados respecto al problema del caso real. Fuente propia.	73



# Lista de tablas

Tabla 1. Artículos resultantes según los criterios definidos.....	24
Tabla 2. Brechas encontradas. Fuente propia. ....	30
Tabla 3. Variables determinadas para el desarrollo del sistema. Fuente propia. ....	39
Tabla 4. Sensores de captura de datos de suelo. Fuente [81]. ....	42
Tabla 5. Sensores que cumplen los criterios de selección. Adaptado de [81]. ....	43
Tabla 6. Rangos de clasificación de valores en las variables determinadas. Fuente [77]. ....	47
Tabla 7. Base de casos con 5 ejemplos. Fuente propia.....	50
Tabla 8. Clasificación de temporada para los años 2006-2021. Fuente propia.....	53
Tabla 9. Acumulados de PP en los años históricos catalogados con temporada normal. Fuente propia.....	55
Tabla 10. Nuevo problema que ingresa al sistema. Fuente propia. ....	56
Tabla 11. Relevancia asignada a cada variable del problema. Fuente propia. ....	61
Tabla 12. Peso asignado según el valor de distancia resultante. Fuente propia.....	64
Tabla 13. Cantidades de fertilizante a recomendar según la condición de suelo y densidad de siembra. Fuente propia.....	65
Tabla 14. MAE de cada tasa de fertilizante recomendado. Fuente propia. ....	71
Tabla 15. RMSE de cada tasa de fertilizante recomendado. Fuente propia.....	72
Tabla 16. Trabajos relacionados que evaluaron con MAE y RMSE. Fuente propia. ....	72
Tabla 17. Datos de cultivo que brindó el caficultor. Fuente propia.....	73
Tabla 18. Análisis de datos de PP dada la fecha de última floración. Fuente propia. ....	73
Tabla 19. Nuevo problema construido. Fuente propia. ....	74
Tabla 20. Nueva solución calculada por el sistema para el problema entrante. Fuente propia .	75
Tabla 21. Datos de tasa convertidos a cantidad de fertilizante en mg/Kg por año. Fuente propia .....	75

# Anexos

ANEXO A. Análisis de resultados de la revisión sistemática en el software de Parsifal.

ANEXO B. Hoja de cálculo con la extracción de datos realizada en la RSL.

ANEXO C. Gráficas con los resultados de la extracción de datos.

ANEXO D. Base de conocimiento construida.

ANEXO E. Comparación de las tecnologías inalámbricas existentes.

ANEXO F. Proceso de obtención de datos realizada.

ANEXO G. Base de casos.

ANEXO H. Registros meteorológicos históricos de la estación climática seleccionada.

ANEXO I. Script para la clasificación de la temporada climática.

ANEXO J. Código del desarrollo del sistema CBR.

ANEXO K. Código de la evaluación del sistema CBR.

ANEXO L. Artículo científico con resultados de la investigación.



# CAPÍTULO 1

## INTRODUCCIÓN

### 1.1. Planteamiento del problema

La producción agrícola en Colombia es uno de los principales pilares para su desarrollo económico, según el Banco Mundial para el año 2020 este sector aportó el 7.6 % del PIB nacional [1]. Además, genera el 15% de los empleos en el país [2]. Sus exportaciones alcanzaron los US 9.418,6 millones, cifra récord en el año 2021. Entre dichas exportaciones está el café, uno de los productos agrícolas principales en la economía del país y uno de los cuatro productos de mayor exportación junto con las flores, el banano y el azúcar [3].

Cabe resaltar el alto impacto social y económico que tiene el café en el sector de la agricultura colombiana, tan solo en el año 2019 generó aproximadamente 785,000 empleos directos, lo que equivale al 26 % del empleo agrícola colombiano, una cifra mayor a los empleos generados por los cultivos de arroz, papa y maíz juntos [4].

No obstante, existen múltiples retos que enfrenta la agricultura colombiana y para ello han sido realizados varios estudios científicos con el propósito de mejorar la productividad agrícola para cubrir la demanda creciente, abordar el cambio climático y la intensificación de los desastres naturales, prevenir las plagas y enfermedades, entre otros [5]. En efecto, la producción agrícola debe garantizar una base sostenible de recursos naturales debido a que gran parte de estos muestran preocupantes signos de degradación y para esto uno de los retos más importantes es el de mejorar las técnicas en la aplicación de insumos a los cultivos [6].

La aplicación de insumos es una de las técnicas usadas por los agricultores que tiene un gran impacto sobre la producción agrícola, constituye un factor primordial en el desempeño de los cultivos. En las producciones agrícolas a pequeña escala, cada agricultor ejecuta diferentes técnicas de fertilización y riego, según la experiencia obtenida en producciones anteriores o la disponibilidad de fertilizantes con la que cuentan al momento de realizar la aplicación [7]. Dichas técnicas de fertilización suelen ser manuales y las cantidades calculadas son de acuerdo a la experiencia previa que hayan tenido, por lo cual puede traducirse en una aplicación ineficiente de fertilizantes minerales. Y en algunos casos no están relacionadas con los requerimientos de nutrientes del cultivo y disponibilidad del suelo. En consecuencia, se puede producir una pérdida en el rendimiento de los cultivos o un aumento en los costos de producción [8].

Adicionalmente, se identificó que los fertilizantes nitrogenados representan una de las fuentes de contaminación que más impactan sobre la atmósfera, el suelo y el agua. Esto se debe a que los cultivos usualmente aprovechan entre el 50-60 % del nitrógeno aplicado, dejando una porción residual que después de la cosecha es liberado al medio ambiente a través de lixiviación, desnitrificación, escorrentía superficial, erosión del suelo y emisiones de N<sub>2</sub>O [9] [10].

La mayoría de las soluciones existentes están ubicadas en otras regiones diferentes al territorio colombiano, en las cuales predominan los cultivos de maíz o caña de azúcar. Además, prevalecen tecnologías de alto costo como por ejemplo, los sensores industriales que superan valores por encima de 5 millones de pesos y que sólo se encuentran al alcance de grandes empresas agrícolas. La propuesta abordada en este trabajo de grado está ubicada en este contexto tecnológico que consiste en brindar recomendaciones a pequeños agricultores colombianos a través de un sistema de recomendación apoyado en sensores de bajo costo, que les informe de manera oportuna la aplicación de la cantidad de fertilizante nitrogenado requerido que deben utilizar en sus cultivos.

Por lo tanto, el desarrollo del presente trabajo de grado busca dar respuesta a la siguiente pregunta de investigación:

*¿Cómo sugerir la cantidad necesaria de aplicación de fertilizantes nitrogenados para pequeños cultivos de café colombianos?*

## **1.2. Antecedentes**

El presente trabajo se enmarca en dos trabajos de investigación desarrollados bajo la modalidad de post doctorado, realizados en la empresa Ecotecma y financiados por recursos del Ministerio de Ciencias. Cabe mencionar que los sensores que fueron utilizados para capturar algunas de las variables de entrada del sistema de recomendación realizado en este trabajo de grado hacen parte de los proyectos en mención, como también, el servicio de captura automática utilizado para los datos meteorológicos.

### **1.3. Aportes**

- Reporte de la revisión sistemática de la literatura realizada.
- Base de conocimiento experta acerca de la fertilización, estado del suelo y condiciones agroclimáticas para los cultivos de café en Colombia.
- Repositorio GIT con los códigos implementados en el desarrollo del sistema.
- Construcción de un sistema de recomendación basado en casos para fertilizantes nitrogenados en cultivos de café, que puede ser de utilidad a pequeños caficultores de la región.
- Artículo que contiene los resultados de la investigación (Anexo L).

### **1.4. Objetivos**

#### **1.4.1. Objetivo general**

Determinar la cantidad de aplicación de fertilizantes nitrogenados en cultivos de café utilizando un sistema de recomendación que tenga en cuenta variables agroclimáticas obtenidas a partir de una red inalámbrica de sensores de bajo costo.

#### **1.4.2. Objetivos específicos**

1. Determinar las variables agroclimáticas que van a usarse para el desarrollo del sistema de recomendación.
2. Adaptar una red inalámbrica con sensores de bajo costo para la recolección de datos agroclimáticos.
3. Proponer un sistema de recomendación que sugiera la cantidad requerida de fertilizante nitrogenado para los cultivos de café.
4. Evaluar el sistema de recomendación para determinar su grado de error de predicción utilizando medidas como el MAE o RMSE.

### **1.5. Estructura del documento**

Este documento ha sido dividido en los capítulos descritos a continuación.

- **CAPÍTULO 1: INTRODUCCIÓN**, expone las bases del presente trabajo de grado, presenta el planteamiento detallado del problema, define los objetivos y aportes investigativos del trabajo.
- **CAPÍTULO 2: MARCO TEÓRICO Y ESTADO DEL ARTE**, describe una base conceptual para entrar en contexto en torno a los conceptos básicos del estudio y

expone los trabajos relacionados a sistemas de recomendación de fertilizantes en la agricultura.

- **CAPÍTULO 3: INTERPRETACIÓN DEL SISTEMA Y DATOS NECESARIOS**, presenta la metodología desarrollada, la contextualización de la situación, los datos que se necesitan para cumplir los objetivos y cómo es su recolección.
- **CAPÍTULO 4: PREPARACIÓN DE LOS DATOS**, presenta el estructurado de los datos, el análisis realizado con los datos históricos y la construcción de la base de casos.
- **CAPÍTULO 5: MODELADO**, explica el desarrollo del modelo del sistema, el algoritmo de aprendizaje automático utilizado y el proceso de predicción de un nuevo resultado.
- **CAPÍTULO 6: EVALUACIÓN**, presenta la evaluación del sistema a partir de métricas como la precisión, el MAE y RMSE. También muestra un ejemplo con un caso real.
- **CAPÍTULO 7: CONCLUSIONES Y TRABAJOS FUTUROS**, presentan las conclusiones, limitaciones y trabajos futuros de la investigación.
- **REFERENCIAS**: Lista todas las referencias bibliográficas que se tomaron para la realización del trabajo.
- **ANEXOS**: Lista los anexos que muestran el trabajo realizado.

# CAPÍTULO 2

## CONTEXTO TEÓRICO Y ESTADO DEL ARTE

### 2.1. Contexto teórico

#### 2.1.1. Sistema de recomendación

Un sistema de recomendación (*Recommender System* o *RS*) es una herramienta y técnica de software que ofrece sugerencias sobre los elementos que pueden ser utilizados por un usuario. Las sugerencias proporcionadas tienen el objetivo de apoyar a los usuarios en diversos procesos de toma de decisiones, como la compra de artículos, preferencias musicales o noticias convenientes [11]. Por lo tanto, un RS analiza todos los datos posibles para hallar la opción más pertinente en el momento adecuado haciendo este resultado cada vez más personalizado para el usuario. Los RS se clasifican según el tipo de contenido que analizan para generar las recomendaciones al usuario, estas clasificaciones se les conoce como modelos básicos de RS [12], que se describen a continuación:

**Filtrado colaborativo:** Se considera la técnica más popular y ampliamente implementada en los RS, la cual basa sus predicciones y recomendaciones en las calificaciones o el comportamiento de otros usuarios en el sistema [13]. La suposición fundamental detrás de este método es que las opiniones de otros usuarios pueden seleccionarse y agregarse de tal manera que proporcionen una predicción razonable de la preferencia del usuario activo. Intuitivamente, suponen que, si los usuarios están de acuerdo con la calidad o relevancia de algunos artículos, entonces probablemente estarán de acuerdo con otros artículos [14]. Cabe resaltar que existen dos tipos de métodos que son usualmente utilizados para el filtrado colaborativo y se les conoce como métodos basados en memoria y métodos basados en modelos. El método basado en memoria, también conocido como filtrado colaborativo en vecindarios, hace sus predicciones basándose en la comparación de las calificaciones de un usuario con las calificaciones de otros usuarios vecinos (similares al primer usuario). Por otra parte, los basados en modelos utilizan técnicas de aprendizaje automático y de minería de datos para aplicarlos en modelos predictivos, entre los más implementados se encuentran: árboles de decisión, modelos basados en reglas, métodos bayesianos y modelos de factores latentes [15].

**Basados en contenido:** Los sistemas que implementan un enfoque de recomendación basado en el contenido analizan un conjunto de documentos o descripciones de



elementos previamente calificados por un usuario y crean un perfil moderado de los intereses del usuario en función de las características de los objetos calificados por ese usuario, en otras palabras, analizan el gusto y/o preferencias que haya tenido ese usuario en el pasado [11]. El proceso de recomendación consiste básicamente en hacer coincidir o encontrar una fuerte relación entre los atributos del perfil de usuario con los atributos del ítem de contenido. El resultado representa el posible nivel de interés del usuario en ese ítem. Diferenciándose así de los sistemas de filtrado colaborativo, que no requieren acceso al contenido del ítem o elemento para generar una recomendación [16]. Si un perfil refleja con precisión las preferencias del usuario, es una gran ventaja para la efectividad de un proceso de acceso a la información [17].

**Basados en conocimiento:** Las recomendaciones en estos sistemas se calculan sin depender de calificaciones de usuarios o enfoques colaborativos, pero sí de una fuente de conocimiento por lo que se diferencian en este aspecto de los sistemas basados en contenido y filtrado colaborativo [18]. Esta fuente de conocimiento puede estar formada por reglas, ontologías, casos históricos o contenido más detallado de artículos recomendables, lo que hace que se pueda tener un nuevo tipo de sistema, que puede ser basado en casos, basado en reglas, basado en restricciones, basado en ontologías, entre otros. Es por esto que estos sistemas son usados ampliamente en situaciones donde hay escasez de información, como por ejemplo, un sistema que no tiene suficientes calificaciones de un artículo o un perfil de usuario con un historial casi vacío que no basta para brindarle una recomendación [15]. A continuación, se introduce una base contextual de los tipos de RS basados en conocimiento más comunes encontrados en la revisión sistemática que se explica más adelante:

- **RS basado en reglas:** El uso de reglas corresponde a la forma natural de los humanos en modelar el conocimiento para resolver problemas [19]. Las reglas que construyen una base de conocimiento se caracterizan por usar condicionales de Si-Entonces (*IF-THEN*), que es el método más común y fácil de comprender. Un sistema basado en reglas trabaja bajo un estado actual de memoria de trabajo, que consiste en un proceso de análisis de reglas y coincidencia de patrones entre reglas para elegir cuál es la regla más adecuada para ejecutarse en el sistema. Cuando ocurren casos donde diferentes reglas pueden coincidir, generando un conflicto de qué regla elegir, el sistema debe contar con un mecanismo que debe estudiar las diferencias entre esas reglas y elegir una, haciendo que la memoria de trabajo del sistema sea actualizada, para evitar futuros conflictos [20]. Esta funcionalidad hace que el sistema se vuelva más adaptable y preciso, a medida que reciba nueva información.

- **RS basado en casos:** Utiliza el razonamiento basado en casos (*Case Based Reasoning* o *CBR*), que según Richter y Weber [21] es una metodología para solucionar problemas que van desde lo cotidiano hasta los estudios científicos más avanzados que existan. Su funcionamiento se basa en reunir o coleccionar un conjunto de casos o problemas resueltos, donde dado un nuevo problema por resolver, se procede a analizar y sacar conclusiones utilizando los casos base, a este enfoque se le denomina razonamiento. Así, el CBR generará una nueva solución para el nuevo problema, basándose en los casos anteriormente registrados. Sin embargo, el CBR hace que estas soluciones no se tomen como conclusiones verdaderas, es decir, si el nuevo caso es muy similar a uno de los casos base, entonces la solución también será similar. Aunque esta reutilización no es la solución correcta, sí es la más cercana. Los casos generalmente son representados como una ontología de conocimiento y experiencia, definiendo las entidades y estructuras de la información (por ejemplo, clases, relaciones, atributos, tipos de datos) la cual es analizada y procesada durante el ciclo CBR.

Este ciclo CBR consta de cuatro pasos secuenciales organizados en torno al conocimiento almacenado en los casos [22]. Primero, la fase de recuperación selecciona uno o varios casos similares de la base de casos. Luego la fase de reutilización adapta las soluciones contenidas en esos casos según la consulta. En la fase de revisión, la solución determinada hasta el momento es verificada en el mundo real y posiblemente corregida o mejorada, esto lo puede hacer un experto en el dominio. Finalmente, la fase de retención toma la retroalimentación de la fase de revisión y actualiza la base de casos con el nuevo problema y su solución encontrada.

- **RS basado en ontologías:** Son sistemas de recomendación que se apoyan en una base de conocimiento para enriquecer sus conocimientos implementando ontologías, quienes son representadas a través de lenguajes de marcado RDF (Resource Description Framework) y OWL (Ontology Web Language) [23].

**Basados en contexto:** Los RS basados en contexto están dedicados a modelar, predecir y recomendar los gustos del usuario mediante tres paradigmas algorítmicos diferentes (pre-filtrado contextual, post-filtrado y modelado) [13], los cuales tienen la función de realizar la incorporación de información contextual tales como: El tiempo, la ubicación, información social, el propósito de una compra, etc. [24].

**Híbridos:** Los RS híbridos son combinaciones de varios tipos de RS con el objetivo de utilizar las diferentes características y fortalezas de cada uno de ellos. Según el contexto o requerimientos que sean necesarios, los RS híbridos pueden utilizar técnicas que

pueden funcionar de manera más robusta en una amplia variedad de configuraciones [25].

### 2.1.2. Sensores en la agricultura

Son dispositivos electrónicos que tienen la capacidad de medir cambios físicos en su entorno y convertirlos en señales eléctricas apropiadas para su manipulación [26]. En la literatura existen varias clasificaciones de sensores dependiendo de la tecnología que utilizan (ópticos, eléctricos, radiométricos, basados en fuerza, electroquímicos), los parámetros que miden (sensores de parámetros del suelo, salud de las plantas, parámetros meteorológicos), la ubicación de los sensores (terrestres, bajo tierra, bajo el agua, móviles, satelitales, proximales, aerotransportados).

- **Redes inalámbricas de sensores:** La agricultura moderna está en gran medida automatizada, controlada y constantemente monitoreada por sensores y varios dispositivos de detección. Por consiguiente, estos dispositivos generan un extenso volumen de datos que crece continuamente y que posteriormente es gestionado a través de sistemas de información especializados que contienen mecanismos para recopilar, almacenar, analizar y simular datos que son el insumo que permite a los agricultores comprender y predecir el crecimiento y el comportamiento de los cultivos en diversas circunstancias [27].

### 2.1.3. Tecnologías de redes inalámbricas existentes

En la agricultura, las aplicaciones normalmente requieren que las redes de sensores sean inalámbricas (WSN por sus siglas en inglés) por las facilidades que éstas ofrecen para la comunicación remota y sin cables, lo cual además de sus características de fácil implementación, bajo costo y bajo consumo de energía las hace ideales para ser utilizadas en lugares de difícil acceso para la implementación de redes cableadas [28]. Algunas tecnologías que existen son:

- **Zigbee:** Debido a algunas desventajas que estaba presentando el estándar para redes inalámbricas IEEE 802.11, como el costo, la confiabilidad y las pérdidas en la red, nació la necesidad de crear una nueva tecnología o estándar que proporcionará bajo costo, bajo consumo de energía, baja velocidad de datos y soluciones económicas en cuanto a implementación. Por consiguiente, la IEEE desarrolló el protocolo 802.15.4 enfocado en una área personal inalámbrica de baja velocidad (LR-WPAN) [29]. En base a este protocolo se define la tecnología ZigBee quien aparte de tener las características ya mencionadas presenta una topología de red tipo malla. Es de mencionar que las aplicaciones de esta tecnología se pueden encontrar en diversos ámbitos como lo son: automatización

del hogar, control de equipos industriales, recopilación de datos médicos, sistemas de monitoreo agrícolas y otras necesidades de bajo ancho de banda. En cuanto a la velocidad de datos, esta maneja 250 kb/s siendo baja en cuanto a otras tecnologías inalámbricas, su rango de transmisión varía entre 10 a 20 mts dependiendo de la potencia de salida y las características medioambientales. Con respecto a la frecuencia donde funciona ZigBee dependerá del país donde se implemente, en la mayoría de los países opera sobre la banda (ISM) 2.4 GHz.

ZigBee presenta muchas ventajas, sin embargo es importante nombrar sus desventajas al igual que cualquier otra tecnología inalámbrica, como por ejemplo que esta tecnología comparte la banda de radiofrecuencia ISM con otros dispositivos que hacen uso de WiFi y Bluetooth, por lo tanto sufren de interferencias involuntarias con los otros dispositivos coexistentes lo cual puede irrumpir la comunicación, además debido a la apertura del medio inalámbrico las comunicaciones pueden sufrir ataques de inferencia de radio, es decir si un equipo emite una señal de inferencia de alta potencia todos los dispositivos ZigBee en red no podrán comunicarse [30].

- **Wifi (Wireless Fidelity):** Es una red de área local inalámbrica y se define como una de las principales tecnologías inalámbricas utilizadas en la actualidad bajo el estándar de la IEEE 802.11. Proporciona conexiones para millones de dispositivos por medio de frecuencias de radio y es una de las principales opciones de los usuarios para acceder a internet reemplazando escalonadamente el acceso por cable. Funciona sobre la banda de frecuencias 2.4 GHz y 5 GHz, su cobertura puede ir desde los 10 metros hasta los 3 Kilómetros y dependerá del número de obstáculos que encuentre la señal en su camino y de los dispositivos que se utilizan en la conexión. Sus velocidades dependen de la versión del protocolo 802.11 y pueden ir desde la primera versión que alcanza los 2 Mbps hasta la más actual que llega a los 2,4 Gbps teóricos [31].

Entre sus ventajas encontramos que al ser una red inalámbrica puede ser estéticamente mejor que una red cableada, además que sus costos se reducen al no utilizar cables. Una de sus principales características es su compatibilidad ya que los dispositivos con la marca Wi-Fi existen a nivel global, con lo que no se tiene ningún problema al usarse en cualquier parte del mundo. Por otra parte, sus desventajas están en el nivel de velocidad que manejan, ya que comparándola con la red cableada las velocidades son menores debido a que es propensa a las interferencias o pérdidas de señal del propio entorno, otra desventaja y no menos importante es su seguridad puesto que existen software especializados capaces de capturar los paquetes de información [32].

- **LoraWan:** LoraWan es una tecnología inalámbrica bajo el estándar LPWAN que es empleado para comunicar y gestionar dispositivos LoRa, fue desarrollada por la compañía LoRA Alliance. Las frecuencias ISM bajo las que funciona, van a depender de la zona del planeta donde se utilice, por ejemplo para europa trabaja sobre los 868 MHz, para américa en los 915 MHz y para Asia en los 433 MHz y se caracteriza por tener eficiencia energética, comunicación de larga distancia, gestión de canales, velocidad de datos adaptativa, geolocalización sin GPS y seguridad integrada. Sus aplicaciones se pueden encontrar en diversos campos como las Smart Cities (Ciudades inteligentes), Smart Farming (Agricultura inteligente), Health Care (Cuidado de la salud), Smart Grid (Redes inteligentes), Environmental Monitoring (Supervisión ambiental) [33].

Entre sus ventajas se puede encontrar que tiene conexiones bidireccionales seguras mediante encriptación de extremo a extremo, un bajo consumo de energía en sus dispositivos alcanzando una vida útil en sus baterías de hasta 10 años, alta inmunidad a interferencias, su alcance está entre los 5 km en urbano a 20 km, en zona rural de distancia, se pueden lograr conexiones de hasta 1 millón de nodos en la red, la transferencia de datos es baja alcanzando un máximo de hasta los 255 bytes, baja frecuencia de transmisión, movilidad y servicios de localización e interoperabilidad entre las diferentes redes LoraWan de todo el mundo [34]. También cabe mencionar sus desventajas o limitaciones que posee esta tecnología como por ejemplo: No toma los datos en tiempo real ya que solo envía información una vez cada tantos minutos y no permite envío de información multimedia.

- **NB-IoT (NarrowBand Internet of Things):** NB-IoT es una tecnología inalámbrica de grado móvil que nace de la iniciativa del 3GPP, también se le llama LTE Cat-NB1 (NB2) y hace parte de las tecnologías de área amplia de baja potencia (LPWA). Se puede implementar LTE para reducir los costos de implementación y lograr actualizaciones sin problemas, puede operar de tres modos diferentes: dentro de la banda de protección de un operador existente, independiente como un operador dedicado y dentro del ancho de banda de un operador LTE de banda ancha. Sus velocidades pueden alcanzar hasta los 20 Kbps en transmisión de datos para un enlace de datos ascendente y 200 Kbps para un enlace descendente [35].

Se caracteriza por un bajo consumo de energía en sus terminales, una cobertura mejorada debido a que se fundamenta en la red móvil, sensibilidad y baja latencia. Sus aplicaciones se encuentran en diferentes industrias que incluyen, servicios

públicos, atención médica, ciudades inteligentes, consumo, agricultura inteligente, logística y almacenamiento, edificios inteligentes y fabricación [36]. Cabe mencionar que sus ventajas se encuentran en la amplia cobertura en sectores urbanos, además sus tiempos de respuesta son más rápidos que otras tecnologías similares como LoRa y puede garantizar una mejor calidad del servicio, sin embargo no todo es positivo, NB-IoT tiene ciertas desventajas con respecto a sus competidores las cuales son las siguientes: La implementación de su firmware es compleja para la transferencia de archivos, los costos de implementación son altos, en zonas rurales el acceso es limitado al depender de la red móvil, y su inmunidad a interferencias es bajo [34].

- **SigFox:** La tecnología SigFox va muy de la mano con el concepto del Internet de las Cosas (IoT) que se encuentra día tras día en un continuo desarrollo y enfrentándose cada vez más a desafíos encontrados en los trabajos investigativos, tales como, elementos de bajo costo, su vida útil, su rango de comunicación, los espectros sin licencia, entre otros. Pero SigFox se diferencia de IoT por la razón de que esta tecnología tiene un enfoque a convertirse en un modelo de negocio para operadores de red. Las aplicaciones que usan esta tecnología solicitan pequeñas transferencias de datos representadas mediante la tecnología de banda ultra estrecha (UNB). El ancho de banda de la comunicación es de 100 bps aproximadamente (ancho de banda de 100 Hz), pero para el Instituto Europeo de Normas de Telecomunicaciones (ETSI) y la Comisión Federal de Comunicaciones (FCC), el ancho de banda puede llegar hasta los 600 bps (ancho de banda de 600 Hz) [37].

Algunas ventajas de la tecnología SigFox es que se caracteriza por ser de bajo costo, baja potencia y se complementa muy bien con otras tecnologías, tales como Bluetooth, Wifi y GPS 2G/3G/4G. Cada elemento o sensor conectado en una red SigFox tiene tres canales de comunicación para enviar de forma pseudoaleatoria los datos mediante paquetes, por lo que de esta manera se evitan interferencias y colisiones en la transmisión de datos, lo que hace que sea una gran ventaja al momento de elegir esta tecnología. Así mismo, otro gran beneficio de SigFox es la seguridad en cada transferencia de datos, ya que cada paquete contiene una clave con un único id del elemento de la red, lo que hace que los datos no lleguen de forma desordenada o repetida [38].

## 2.2. Estado del arte

### 2.2.1. Revisión sistemática de la literatura

El presente trabajo de grado realizó una investigación literaria a través de una metodología denominada revisión sistemática de literatura para ingeniería del software propuesta por Kitchenham [39]. Los pasos llevados a cabo en esta metodología se muestran en la figura no. 1.



Figura 1 Proceso de RSL. Adaptado de la metodología Kitchenham.

El primer paso es el planteamiento de la pregunta de investigación, la cual se encuentra en la sección 1.1. Luego se procede a definir las palabras clave y una cadena de búsqueda que estén relacionadas con el problema definido. Las palabras clave seleccionadas fueron escritas en inglés y son presentadas a continuación: sistemas de recomendación, sistemas para la toma de decisiones, sistemas de predicción y sistemas expertos, cabe resaltar que en la búsqueda se hace un enfoque en trabajos que se centran en agricultura, nutrientes y fertilizantes o fertilización.

Las fuentes de información de trabajos investigativos utilizadas en esta revisión fueron Web Of Science (WoS), Scopus y la base de datos de la IEEE. A continuación se presenta la cadena de búsqueda utilizada: **("recommender system" OR "recommendation system" OR "prediction system" OR "expert system" OR**

"decision support system") AND (Farm\* OR Agri\* OR Nutri\* OR fertili\*)) la cual entregó un total de 1536 documentos.

Posteriormente, la fase de selección de artículos utilizó los siguientes filtros:

**Primer filtro:** En este paso se definieron los siguientes criterios de exclusión obteniendo como resultado 220 artículos.

- Artículos de fuentes secundarias o terciarias.
- Artículos que no están escritos en el idioma inglés.
- Artículos que sólo son resúmenes de libros o revisiones de otros trabajos.
- Artículos presentados en conferencias.

**Segundo filtro:** Los 220 artículos resultantes del primer filtro se ingresan al programa Parsifal, que es una herramienta de apoyo a las revisiones sistemáticas. El proceso realizado se describe más a detalle en el anexo A. En este programa, se realizó una evaluación de la calidad donde se calificó cada artículo con unos criterios de evaluación o también denominados preguntas de calidad, que se muestran a continuación:

- ¿Realiza una revisión comprensiva del estado actual del conocimiento o trabajos relacionados existentes?
- ¿Muestra y describe los algoritmos o pseudocódigos utilizados o por lo menos indica claramente las técnicas utilizadas?
- ¿Analiza los principales hallazgos y resultados empíricos obtenidos?
- ¿Describe claramente la evaluación realizada de los resultados obtenidos?

Cada pregunta contiene tres posibles valores de respuesta, los cuales corresponden a una puntuación que se la asigna al artículo revisado, estas respuestas son:

- “*Sí*”: Indica una puntuación de 5, se refiere a cuando las preguntas se responden de forma completa, sin dejar lugar a dudas.
- “*Parcialmente*”: indica una puntuación de 3, quiere decir que el trabajo muestra de forma muy resumida o general, la característica que se pretende responder en la pregunta.
- “*No*” con una puntuación de 0, ocurre cuando no se encuentra en el trabajo, la característica que se está buscando.

Entonces, la máxima calificación que puede recibir un artículo es 20 y la mínima es 0. Se asume un valor aceptable de calificación de 10, donde los artículos que obtengan una calificación de 10 o más, son aceptados para ingresarlos al proceso de extracción de



datos. Finalmente, de los 220 artículos sólo 105 pasaron la calificación establecida, iniciando así, el proceso que se describe a continuación.

### 2.2.2. Extracción de datos

El proceso de extracción de datos de los artículos inicia identificando todas las características que se quieren saber de cada trabajo, con el fin de hacer una identificación de brechas. Las características definidas para identificar en cada artículo son:

- El tipo de sistema desarrollado en el trabajo.
- El país de estudio del trabajo.
- La técnica o el algoritmo utilizado en el trabajo.
- El o los cultivos que fueron estudiados en el trabajo.
- Las fuentes de donde recolectaron los datos trabajados.
- Los datos o variables de entrada del sistema.
- Los datos o variables de salida del sistema.
- Al sistema se le realizó evaluación o no.

En el anexo B, se observan los resultados de esta extracción de datos para los 102 artículos resultantes de la evaluación de calidad que se realizó en el programa de Parsifal. Así mismo, el anexo C muestra los resultados en gráficas para identificar las características más comunes resultantes para cada parámetro al que se le realizó una identificación dentro de cada trabajo. Pero, para los objetivos del presente trabajo, se realiza un tercer filtro donde se seleccionan sólo los RS que recomiendan fertilizantes y que trabajan con datos históricos, redes de sensores o información experta, dando un resultado de 19 trabajos relacionados, descritos de forma resumida en la siguiente sección. En la tabla 1 se puede observar el número de artículos resultantes por cada filtro realizado.

*Tabla 1. Artículos resultantes según los criterios definidos.*

<b>Filtros</b>	<b>Artículos resultantes</b>
1. Criterios de exclusión.	220
2. Criterios de evaluación.	102
3. RS de fertilizantes.	19

### 2.2.3. Trabajos relacionados

De los más de 100 artículos que aprobaron la calificación planteada en el programa de Parsifal, se procede a seleccionar sólo aquellos que se enfocaron en sistemas de recomendación de fertilizantes. De los cuales, sólo se encontraron 19, estos trabajos son los que se describen a continuación, con el fin de encontrar unas brechas más específicas en cuanto a RS de fertilizantes.

El trabajo de Chougule et. al. [40] desarrollaron un RS basado en ontologías para cualquier tipo de cultivo, en el estado de Maharashtra, India. El RS fue alimentado con datos históricos e información experta de bases de datos del gobierno, la cual fue convertida y almacenada en ontologías, posteriormente los datos fueron analizados y estudiados con dos algoritmos de aprendizaje automáticos: que fueron el agrupamiento de k-medias (k-means clustering) y bosque aleatorio (random forest). Los datos que componen la base del conocimiento fueron un historial de los últimos tres años del contenido de NPK en el suelo, los tipos de cultivo que crecieron en ese suelo, condiciones climáticas y cómo fue la producción de esos cultivos. Finalmente, con este historial, el RS brinda una recomendación a los agricultores basada en la región, el contenido de NPK en el suelo y el tipo de cultivo, afirmando que el desempeño del sistema tiene una alta precisión y esperan que logre el objetivo de mejorar la producción agrícola en esa zona del país.

Los fertilizantes en la agricultura pueden estar compuestos de los principales nutrientes que necesitan los cultivos, pero también pueden existir fertilizantes concentrados en un sólo nutriente, los tres más importantes son nitrógeno, fósforo y potasio. Vieira et al. [41] Implementaron un RS enfocado sólo en el nutriente del nitrógeno, diseñado en cultivos de trigo y cebada en la región de Paraná, Brasil. El RS propuesto está basado en contenido con datos de 70 experimentos de campo realizados entre 2007 y 2012. Estos datos corresponden a valores de materia orgánica (OM), pH, P, K, tasas de N aplicado y el rendimiento de los diferentes cultivos que se sembraron en ese periodo de tiempo. Así, el sistema mediante análisis de correlación de datos trata de obtener la máxima eficiencia económica de tasas de aplicación de N en los cultivos.

El trabajo de Kumar et. al. [42] consiste en un RS híbrido en la agricultura que se alimenta de variables del suelo capturadas por sensores, con el fin de brindar recomendaciones de fertilización a pequeños agricultores en India en cultivos como el trigo, la cebada, el maíz, la caña de azúcar entre otros. El sistema toma como datos de entrada el nivel de NPK del suelo, el pH, la humedad y temperatura del suelo, quienes a través de un proceso de reacciones químicas implementando la técnica de colorimetría y con la ayuda de un sensor de colores detectan los niveles de fertilizantes.

El RS basado en contenido de Wickramasinghe et. al [43], presenta un sistema con sensores integrados que miden la fertilidad del suelo y desarrollan una aplicación móvil multiplataforma para sugerir los mejores cultivos según la fertilidad del suelo, pero también, brinda una recomendación de un plan de fertilización para optimizar el uso de fertilizantes. El sistema es desarrollado en Sri Lanka y probado en cultivos que no tienen grandes extensiones, donde podemos encontrar el maíz, maní, frijol, banano, tomate, caña de azúcar, entre otros. El funcionamiento del sistema consiste en registrar el nivel de los nutrientes en el suelo, tales como, el pH, humedad, temperatura, conductividad eléctrica (CE) y el NPK, todo esto mediante el uso de sensores. Luego, estos registros son enviados a un servidor en la nube y con los niveles de fertilizante adecuados definidos por los autores, se entrena el modelo utilizando el algoritmo de máquinas de vectores de soporte (SVM) para identificar cuáles son los valores más óptimos de NPK en el suelo para los diferentes tipos de cultivo estudiados.

Suchithra y Maya Pai [44] también desarrollaron un RS basado en contenido que brinda recomendaciones de tipo y cantidad de fertilizantes en el cultivo de arroz. El área de estudio es todo el estado de Thrissur en India. El trabajo se basó en un registro histórico de variables entre los años 2015 a 2017, que incluyen el pH del suelo, CE, nutrientes en el suelo (OC, P, K, S, Zn, B, Cu, Mn y Fe) y fertilizantes como la Urea, MOP, Fosfato y Fertilizante orgánico. Estos datos fueron procesados para identificar parámetros y analizar su eficacia, formando un dataset mejorado, el cual se divide en tres porcentajes: El 70% de los datos son para entrenamiento, 15% para validación y el otro 15% para pruebas. El modelo de predicción desarrollado se basa en el algoritmo SVM y su fin es recomendar la tasa de aplicación de fertilizantes óptima para los cultivos de arroz. Los autores plantearon que en el futuro, el RS se pueda aplicar en otras tierras y cultivos.

Otro RS de fertilizantes basado en contenido es el estudio de Sujithra et. al. [45] el cual fue desarrollado con el objetivo de brindar apoyo para la práctica agrícola eficaz en cultivos de coco en India. Sus parámetros de entrada consisten en variables de suelo recolectadas por sensores inalámbricos, donde se destacan, el NPK, pH, temperatura y humedad. El sistema utiliza un modelo de clasificación, que experimenta con los algoritmos de árboles de decisión J48, SVM y k-means para seleccionar el clasificador más adecuado. Los resultados obtenidos indicaron que el algoritmo J48 hizo una mejor clasificación de la disponibilidad de NPK en el suelo que los demás, por lo que fue elegido para hacer una clasificación más precisa. Posteriormente, los datos que recopilaron en el campo fueron tomados como datos de prueba y comparados con los datos entrenados ya introducidos en el sistema, con el fin de sugerir fertilizantes para los casos donde haya deficiencia de macronutrientes en el suelo.

Zhang y Ren [46] desarrollaron un RS en línea de fertilización de cultivos de maíz, trigo y maní. Este RS está basado en un modelo de balance de nutrientes y un sistema de toma de decisiones soportado en un servidor ArcGIS a escala de lote para el condado de Huaen China. La base de conocimiento de este trabajo consiste en datos de estaciones meteorológicas del gobierno, de donde recolectaron variables climáticas como la temperatura máxima, media y mínima diaria, la precipitación media diaria y las horas de exposición solar diaria, también datos de suelo como tipo de suelo, textura y diferentes nutrientes. Además, trabajaron con datos de cultivo, como especies de cultivo, variedades de cultivo, región de plantación, área de plantación, rendimiento por hectárea, producción anual y el índice de cultivos múltiples, con el fin de brindar recomendaciones sobre el tipo de cultivo y la gestión en la aplicación de fertilizantes a los agricultores.

Ren y Lu [47] propusieron un RS para cultivos de arroz en la región de Wenling(China). Para el desarrollo del RS usaron un sistema de toma de decisiones soportado en las siguientes bases de conocimiento históricas que incluyeron datos de: Tipo de suelo, nutrientes disponibles, capacidad de suministro, idoneidad de entre los fertilizantes, el crecimiento de los cultivos, la consistencia del tiempo entre la fertilización y el cultivo que requiere, la cantidad de fertilizante utilizada en los últimos años, la variedad y la producción de cultivos y las características del fertilizante necesario, con el objetivo de recomendar los fertilizantes más adecuados para dicho cultivo.

Hossain y Siddique [48] plantearon RS basado en conocimiento de fertilizantes y cultivos genéricos en Bangladesh. El RS está basado en un sistema experto que aprovecha una base de conocimiento con datos históricos del suelo obtenidos de la base de datos nacional de suelos desarrollada por el gobierno de Bangladesh. Además, considera datos de muestras del suelo como: Nitrógeno, Potasio, fósforo, además del tipo de suelo, que a su vez sirvieron como entrada para el RS. Este fue desplegado en la web y puede ayudar a los campesinos brindándoles recomendaciones sobre el tipo de fertilizante que deben usar para sus cultivos y la tasa de aplicación.

Cholissodin, Dewi, y Eunike Endariahna Surbakti [49] desarrollaron un RS basado en conocimiento con datos experimentales de fertilizantes para cultivos de maíz usando los siguientes dos algoritmos: ANN para analizar la influencia de varias de las entradas del sistema como: las combinaciones de tasas de aplicación de fertilizantes experimentales, y optimización de enjambre de partículas mejorado (*BIPSO* por sus sigla en inglés), con el fin de determinar la dosis de fertilizantes necesaria.

Islam y Tuhser [50] definieron un RS basado en contenido para cultivos de arroz en Bangladesh. Para esto, se basaron en una tabla de colores de hojas (LCC) propuesta por el Instituto Internacional de Investigación de Arroz (IRRI), la cual permite a los

agricultores determinar la demanda de nitrógeno de las plantas en cultivos de arroz [51]. Islam y Tuhser obtuvieron 6000 imágenes de muestras de hojas de arroz, que posteriormente fueron analizadas, segmentadas semánticamente y clasificadas mediante dos algoritmos, una red neuronal convolucional (CNN) y un árbol de decisiones (DT), con el fin de determinar la cantidad necesaria de nitrógeno que deben aplicar los agricultores. Abd Razak et al., [52] desarrollaron un RS de fertilizantes nitrogenados para el cultivo de arroz en Malasia, pero a diferencia del trabajo previo, utilizaron algoritmos de lógica difusa para analizar la tabla de LCC e identificar la cantidad óptima de aplicación de fertilizante nitrogenado.

Sumaryanti y Lamalewa [53] propusieron el desarrollo de un RS basado en conocimiento haciendo uso del algoritmo de lógica difusa para recomendar la dosis de fertilizante NPK para cultivos de arroz en Indonesia. El estudio utilizó como base de conocimiento algunas variables de entrada de una tabla de colores de plantas que mide el nivel de clorofila y el rendimiento del cultivo. A continuación, a partir de un modelo difuso generaron un sistema de toma de decisiones de criterios múltiples. Y luego, por medio de la técnica para el orden de preferencia por similitud con la solución ideal (TOPSIS) realizaron la recomendación de dosis de fertilizante nitrogenado.

Garg y Aggarwal [54] plantearon un RS de fertilizantes para cultivos de arroz y trigo en India. Para esto utilizaron conjuntos de datos de laboratorio acerca de los nutrientes del suelo con los cuales entrenaron varios modelos. Estos modelos fueron desarrollados a partir de cinco algoritmos de ML, a saber, descenso del gradiente automático (SGD), ANN, RF, KNN, árboles de regresión (RT) y SVM. Los resultados fueron evaluados con las medidas MAE y RMSE, con lo cual pudieron concluir que el SGD fue el que mejor se comportó con un MAE de 0.41 y un RMSE de 0.45.

Qin y Myers [55] propusieron un RS basado en contenido para la predicción de la tasa óptima de nitrógeno para cultivos de maíz en USA. Para esto, capturaron datos de estaciones meteorológicas, sensores y muestras de perfil de suelo, luego probaron algunos algoritmos de ML como la regresión lineal (LR), regresión de crestas (RR), contracción mínima absoluta y operador de selección (LASSO) y árboles de regresión de aumento de gradiente (GBRT). Para evaluar sus resultados utilizaron R<sup>2</sup>, MAE y RMSE. En la evaluación el algoritmo de regresión de crestas presentó el mejor comportamiento con un 70 % de acierto.

Jiang y Xing [56] desarrollaron un RS basado en contenido para cultivos de maíz en China. Este RS fue construido a partir del modelo cuadrático utilizando datos obtenidos entre los que se encuentran: Tasas de nitrógeno aplicado, rendimiento del cultivo, ubicación, precipitación de lluvias, temperatura, pH, suelo orgánico de carbono, fósforo y potasio, de 31 parcelas experimentales de 60 m<sup>2</sup> durante 5 años ubicadas en la principal

región productora de maíz del noroeste de China, con el fin de tener la tasa óptima de fertilizantes de NPK.

Puntel y Pagani [57] desarrollaron un RS basado en contenido para cultivos de maíz en Argentina. Para el desarrollo del RS implementan el algoritmo de ramificación y modelos de regresión de datos de suelo recolectados de 54 variables estáticas (tales como, pH, materia orgánica, elevación, profundidad, rendimiento de cultivo anterior, entre otras) y variables dinámicas (entre las que se encuentran la humedad del suelo, el nitrato del suelo, la precipitación, la temperatura del aire, entre otras) que se obtienen a partir de sensores y estaciones meteorológicas, con el objetivo de recomendar a los agricultores la tasa óptima de nitrógeno a aplicar.

McFadden, Brorsen y Raun [58] desarrollaron un mecanismo de recomendación de la cantidad de nitrógeno a aplicar sobre un cultivo genérico basado en un método Bayesiano que combina una estimación previa de la función de producción realizada por el agricultor, calculada a partir de la ganancia esperada (función del precio de salida y el costo del nitrógeno), y datos obtenidos de sensores de reflectancia óptica (ORM).

Ramson y Kitchen [59] desarrollaron un RS basado en contenido para cultivos de maíz en Estados Unidos. El estudio utilizó los algoritmos de ML como regresión contraída, paso a paso, LASSO, regresión de red elástica, componente principal de regresión (PCR), regresión de mínimos cuadrados parciales (PLSR), Árbol de decisión y bosque aleatorio. Los datos de entrada al sistema fueron datos históricos entre los que se encuentran: precipitación de lluvia, tipo de suelo, Nitrógeno en el suelo, Nitrógeno de carbono y pH, con el fin de recomendar las cantidades necesarias que necesita el cultivo.

La mayoría de los trabajos estudiados son en realidad RS híbridos pues generalmente utilizan datos contextuales obtenidos a partir de sensores, que son posteriormente complementados con datos específicos de cada cultivo (contenido) y bases de conocimiento experto. Además, estos trabajos fueron realizados para cultivos extensivos y por lo tanto, existe aún la necesidad de adaptarlos a las necesidades y contexto de pequeños agricultores en países en vía de desarrollo para que su impacto sea mayor en cuanto al beneficio económico y la protección del medio ambiente.

#### **2.2.4. Brechas encontradas**

La siguiente tabla muestra las brechas detectadas en la revisión sistemática general, que es la que incluye los sistemas de predicción, sistemas de soporte de decisiones, sistemas expertos, entre otros. Y las brechas encontradas en estudios de sistemas de recomendación que sólo recomiendan fertilizantes.

Tabla 2. Brechas encontradas. Fuente propia.

Sección-Trabajos	Brechas
Sistemas de recomendación de fertilizantes	<ul style="list-style-type: none"> <li>- No están desarrollados o adaptados para Colombia.</li> <li>- No están desarrollados para pequeños cultivos y no usan sensores de bajo costo.</li> <li>- No están enfocados para cultivos de café.</li> <li>- Los datos trabajados son recolectados de cultivos estacionales o registros históricos. No trabajan con datos recientes.</li> </ul>
Sistemas expertos, de predicción y de apoyo para la toma de decisiones en la agricultura	<ul style="list-style-type: none"> <li>- No presentan evaluación del sistema desarrollado.</li> </ul>

Cabe mencionar que el presente trabajo aborda las brechas encontradas en la Tabla 2. ya que el desarrollo del sistema propuesto se realizó en el Cauca, se enfocó en pequeños cultivos de café, es decir en extensiones de tierra menores a 1 hectárea y además implementó sensores de bajo costo para la recolección de datos recientes, también cabe nombrar que el sistema muestra una evaluación con sus respectivos resultados.

# CAPÍTULO 3

## INTERPRETACIÓN DEL SISTEMA Y DATOS NECESARIOS

### 3.1. Definición del sistema

Para comenzar con el desarrollo del RS, es importante tener claro la estructura del sistema, cómo se va a componer y cómo va a funcionar. Es por esto que para el desarrollo del sistema propuesto, el primer paso fue definir el tipo de sistema que se iba a desarrollar. En el marco teórico se observaron que existen diferentes tipos de RS que tienen diferentes funcionalidades dependiendo del contexto en el que se vayan a utilizar. Entonces se identificaron las herramientas y datos que va a utilizar el sistema con el fin de definir qué tipo de sistema se construyó.

En la propuesta se tiene que el sistema trabaja con variables agroclimáticas, es decir, datos que se recolectan del cultivo, suelo y clima para su posterior procesamiento y análisis. Esto hace que sea un sistema basado en contenido, en el que se tienen unos datos iniciales que son pre procesados y analizados con los datos de background del sistema (datos almacenados en el sistema y con los que el sistema trabaja), y que haciendo uso de un algoritmo o técnica de recomendación, el sistema brindará las respectivas recomendaciones para esos datos iniciales ingresados.

Por otro lado, el presente trabajo contó con la ayuda de dos expertos de Ecotecma, para la verificación y validación del desarrollo del sistema. Lo que fue útil para crear una base del conocimiento y generar las recomendaciones a partir de la información experta recolectada. Esto hace que el sistema también sea basado en conocimiento. Es decir, el sistema desarrollado es un sistema híbrido que analiza contenido ingresado al mismo, a partir de la recolección de las variables agroclimáticas, y se apoya de una base de conocimiento para brindar las recomendaciones.

Finalmente, se definió la técnica de recomendación utilizada en el desarrollo del RS, donde se analizaron los diferentes tipos de sistemas basados en conocimiento que existen, tales como los sistemas basados en reglas, los sistemas basados en ontologías y los sistemas basados en casos. Se hizo un breve estudio de los tres, descartando los primeros dos métodos, por razones de que no se adaptan de la mejor forma al sistema propuesto y que no disponían de facilidad al momento de realizar correcciones. Las reglas, como se vió en el marco teórico, trabajan bajo una estructura de IF-THEN, haciendo que ante múltiples variables de entrada (datos iniciales), se tenga una ramificación de reglas que aumenta ante el aumento de las variables de entrada. Por lo que un sistema basado en reglas no es viable para situaciones donde se tienen muchas



variables de entrada. Así mismo, las ontologías no se adaptaron bien para el desarrollo del presente sistema, debido a que estas usan relaciones entre entidades (que se componen de parámetros, clases, entre otras cosas). Haciendo que este tipo de sistemas sea más viable cuando se trabajan con múltiples cultivos y múltiples suelos. Por lo tanto, el sistema basado en casos fue la técnica seleccionada para el desarrollo del sistema. Esta técnica se basó en el razonamiento basado en casos (CBR).

### 3.2. Metodología

Este capítulo presenta la metodología trabajada llevada a cabo para el desarrollo del presente trabajo. La cual se basa en el marco de trabajo de CRISP-DM (Cross Industry Standard Process for Data Mining). La mencionada metodología proporciona una estructura y orden en el desarrollo de un proyecto de minería de datos. Este contiene las fases del proyecto, así como sus respectivas tareas y resultados.

El ciclo de vida de la metodología se desglosa en seis fases interconectadas entre sí por flechas que expresan las relaciones más relevantes y frecuentes: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue. La figura 2 muestra una adaptación de las fases propuestas que fueron implementadas en la presente investigación.

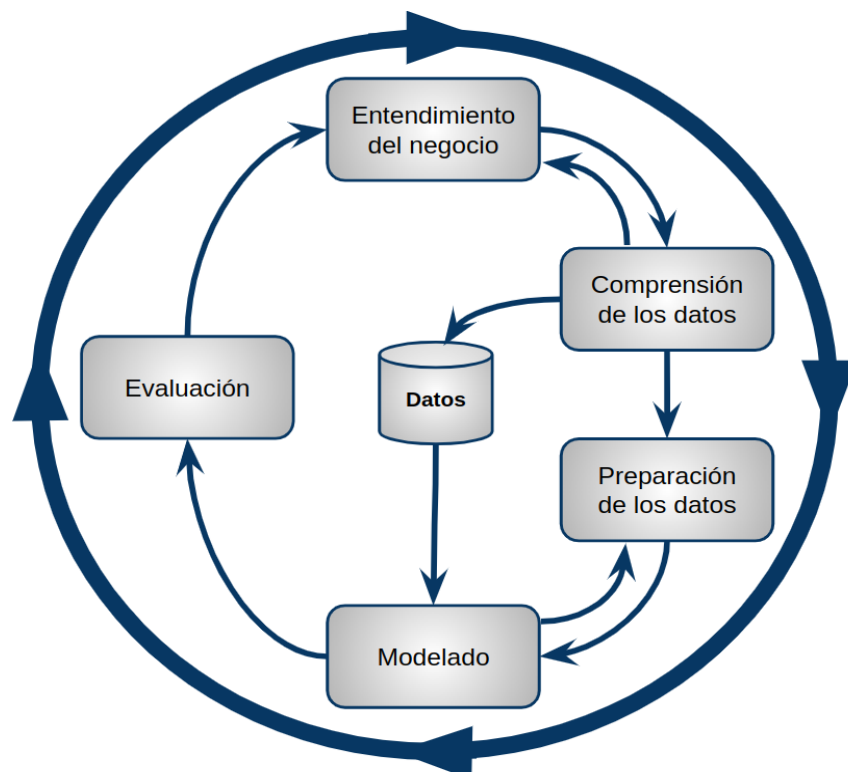


Figura 2. Metodología CRISP-DM. Adaptada de [60].

Las dos primeras fases son la base para la consecución del conjunto de datos meteorológicos y de suelo para el presente estudio. En este orden de ideas, el capítulo se encuentra dividido en los siguientes apartados:

- **Comprensión del negocio:** se expone el objetivo que se busca con la construcción de los casos del sistema CBR. Como parte de esto se estudia el proceso de fertilización del cultivo de café, así como las variables que influyen en este.
- **Comprensión de los datos:** se presenta el proceso de determinación y captura de los datos iniciales, así como la descripción de estos.
- **Preparación de los datos:** muestra la clasificación de los datos, la estructuración de un caso y la construcción de la base de casos del CBR. Además, explica el proceso de armado de un nuevo caso y cómo se preparan los datos para que ingresen al modelado.
- **Modelado:** describe todos los pasos desarrollados en el sistema CBR, así como la definición de las recomendaciones finales que brinda dicho sistema. También se muestra un ejemplo de cómo funciona el sistema con un caso real.
- **Evaluación:** explica el proceso de la evaluación del sistema, cómo se determinó el factor  $k$ , cómo se calculó la precisión del sistema y cuáles fueron el MAE y RMSE hallados.

### 3.3. Comprensión del negocio

Se busca construir un sistema que recomiende al agricultor la tasa de aplicación de fertilizantes adecuada en cultivos de café caturra en su etapa de producción, la cual incluye el proceso de floración, maduración del fruto y producción [61]. En la presente investigación, es necesario entender el proceso de fertilización del cultivo ya mencionado, así como las variables que pueden incidir en él.

#### 3.3.1. Fertilización en el cultivo de café

El cultivo de café como cualquier otra especie vegetal requiere de elementos primordiales para su desarrollo y crecimiento, tres de ellos llamados constituyentes orgánicos, se encuentran libremente en el ambiente, que son el carbono (C), el hidrógeno (H) y oxígeno (O), estos se obtienen del agua y la atmósfera, representando así el 95% del peso de la

planta. Mientras que el 5% restante se encuentran en el suelo y son clasificados como minerales, los cuales se clasifican en macronutrientes, que son Nitrógeno (N), Fósforo (P), Potasio (K), Calcio (Ca), Magnesio (Mg) y Azufre (S), y Micronutrientes como el Hierro (Fe), Manganeso (Mn), Cobre (Cu), Boro (B), Cloro (Cl), Molibdeno (Mo) y Níquel (Ni) [62].

Cabe mencionar que los macronutrientes son necesarios en mayor cantidad, además se clasifican en primarios (N, P y K) y secundarios (S, Ca y Mg), los primarios se necesitan en un porcentaje más alto que los secundarios, por el contrario los micronutrientes son necesarios en cantidades muy pequeñas (partes por millón) [63]. Adicionalmente, debido a cambios y fenómenos climáticos que ocurren con el paso del tiempo, se pueden presentar diferentes procesos naturales nombrados a continuación, los cuales hacen que se produzca una pérdida de estos nutrientes en el suelo [64].

- **Lixiviación:** entendida como el desplazamiento de nutrientes y sustancias por debajo de la zona de las raíces del cultivo hacia los cuerpos de agua debido al exceso de humedad en el suelo [65].
- **Escorrentía superficial:** se define como la precipitación que sobre la superficie del suelo corre por acción de la gravedad sin infiltrarse en el suelo [66].
- **Erosión:** desgaste que sufre la superficie de la tierra debido a diferentes eventos naturales como la lluvia, el sol y desastres naturales, además de las causas generadas por el hombre cuando hace uso inadecuado de los recursos del suelo [67].
- **Desnitrificación:** debido la gran cantidad de microorganismos que utilizan el nitrito y nitrato en lugar del oxígeno, se genera la producción de formas gaseosas de nitrógeno entre los que se encuentra el óxido nitroso quien es uno de los principales contaminantes del aire [68].

Para reducir el déficit de los nutrientes en el suelo, se realiza un proceso de fertilización el cual aporta los complementos necesarios para que la planta sea productiva y no ocurran pérdidas [62]. Uno de los factores a tener en cuenta en este proceso es la etapa del cultivo que ya es mencionada en la sección 4.1. Teniendo la etapa de enfoque elegida, se considera que el N es el nutriente más limitante de la producción, pues cuando se excluye de los planes de fertilización, el rendimiento se puede reducir hasta en un 80% [69]. La materia orgánica del suelo al mineralizarse (descomponerse) proporciona cantidades considerables de N, a pesar de esto y por lo general, no alcanza a cubrir por completo la demanda del cultivo, motivo por el cual es necesario proporcionar N al suelo de manera frecuente a lo largo del año [70].

### 3.3.2. Períodos de fertilización

Entendiendo la problemática de la necesidad de fertilizar los cultivos, se tiene que identificar los momentos de aplicación de la fertilización en los cultivos. Para el café, se debe identificar y comprender el tiempo transcurrido desde la floración hasta la cosecha. La figura 3 muestra un período cafetero o también denominado “*año cafetero*”, donde se observa que en un año cafetero, existen dos momentos de floración del café y por ende dos cosechas.

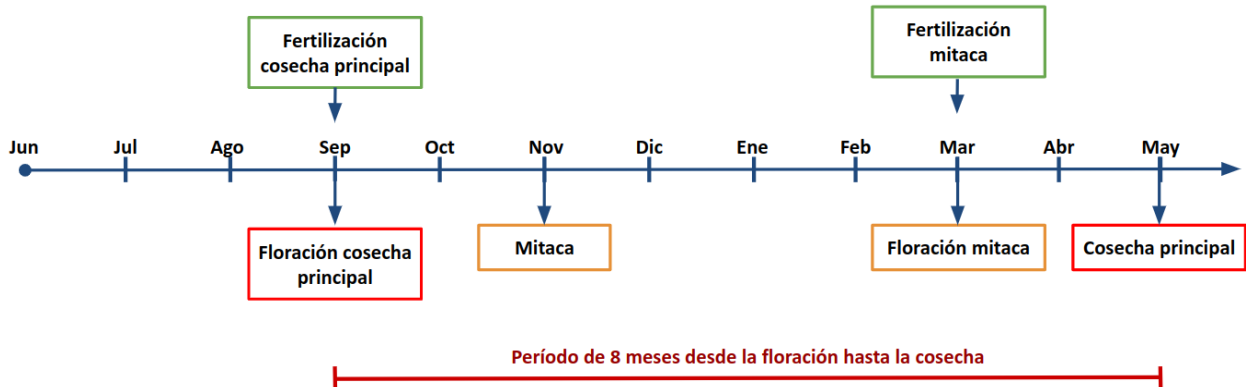


Figura 3. Diagrama de un año cafetero en el Cauca. Adaptado de [71].

Las dos cosechas se conocen como cosecha primaria y secundaria. La cosecha primaria también se denomina cosecha principal y contiene entre el 50-75% de la cosecha total en el año cafetero. La cosecha secundaria también conocida como “*mitaca*” corresponde al 25-50% de la producción total en el año [72]. Con esto, se identifica que existen dos períodos de fertilización en un año cafetero.

La región cafetera colombiana, es afectada históricamente por fenómenos climáticos que suelen ocurrir de manera natural conocidos como El Niño y La Niña [73]. Un año o temporada afectada por el fenómeno de El Niño indica una temporada con una ausencia considerable de lluvias por lo que también se le conoce como “*Temporada seca*”. Una temporada afectada por el fenómeno de La Niña se refiere a “*Temporada lluviosa*” o temporada con altas precipitaciones pluviales [74]. Adicionalmente, los años o temporadas que no se ven afectados por ninguno de los dos fenómenos, se les conoce como años normales o neutrales. Con esta clasificación de la temporada, según el estudio de Cenicafé [75], es posible definir el número de aplicaciones de fertilizantes para cada período de fertilización en el año cafetero.

En resumen, al identificar la importancia de la fertilización, así como los períodos y la frecuencia de este proceso, se comprende el problema o negocio que abarca este

trabajo. Entonces, el grado de éxito es encontrar las cantidades más adecuadas de fertilizantes a aplicar en el cultivo de café, teniendo en cuenta las diferentes condiciones posibles que pueden haber en el clima y suelo. Por lo tanto, se inicia la siguiente fase de la metodología CRISP-DM, con el fin de identificar las variables que se van a trabajar, así como su recolección.

### **3.4. Comprensión de los datos**

La segunda fase de CRISP-DM es la comprensión de los datos. En esta fase se llevó a cabo la recolección de los datos base para la construcción de los casos del modelo, además de realizar una exploración y descripción de estos.

#### **3.4.1. Análisis de fuentes de datos**

En la primera fase de CRISP-DM, se explicó la importancia de los datos meteorológicos y de suelo en las zonas donde hayan cultivos sembrados. Con esta información, se realizó un análisis de posibles fuentes de datos históricos en cultivos de café, así como datos meteorológicos de la zona de estudio (la región del Cauca). Como resultado de la búsqueda se obtuvieron las siguientes opciones: Federación nacional de cafeteros, el Centro Nacional de Investigaciones de Café (Cenicafé), Agronet del Ministerio de Agricultura y Desarrollo Rural, el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), entre otros. Con estas fuentes y la colaboración de expertos de Ecotecma, se definió utilizar los datos meteorológicos registrados en los anuarios de Cenicafé y una red inalámbrica implementada por la misma empresa.

#### **3.4.2. Recopilación documental**

En esta parte se reúne toda la información investigada sobre el manejo de cultivos de café. Con esta información, se forma una base de conocimiento que se compone de todo el proceso de fertilización en un cultivo, las etapas del cultivo, la composición de nutrientes en el suelo, la absorción de esos nutrientes en una planta de café, las cantidades de fertilizante que suelen aplicar los caficultores al año, cómo ocurren los cambios en los factores climáticos en la región, entre otras variables. Esta información se detalla en el anexo D.

#### **3.4.3. Recolección de datos**

La recolección de los datos se llevó a cabo en dos etapas. La primera etapa fue la obtención de información para determinar las variables a trabajar en el sistema propuesto con el fin de crear los casos para el modelado del sistema. La segunda etapa fue la captura de datos a partir de una red inalámbrica de sensores y el consumo de un servicio para la extracción de datos meteorológicos históricos.

### 3.4.3.1. Determinación de variables

El proceso de la determinación de las variables consta de 4 pasos que van desde la identificación de las variables de entrada más comunes en los trabajos relacionados, hasta la priorización de las variables que se consideran necesarias e importantes para el desarrollo de un RS de fertilizantes en un cultivo de café. Este proceso se muestra en la siguiente figura:

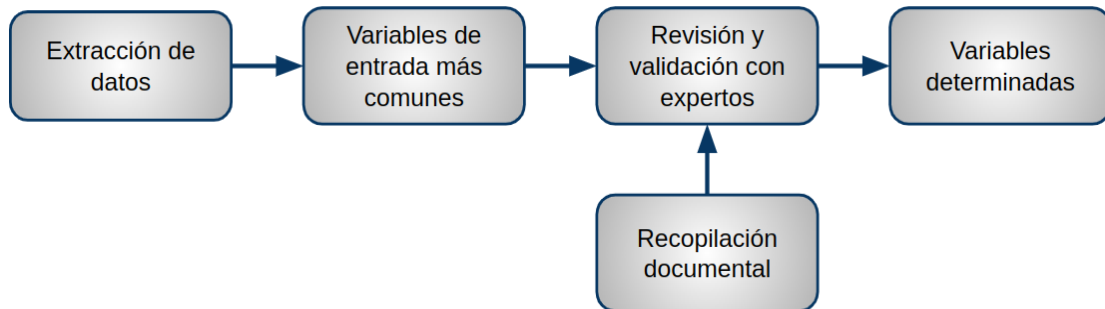


Figura 4. Diagrama del proceso de determinación de variables del sistema. Fuente propia.

- **Extracción de datos:** Este proceso ya explicado en el estado del arte, consistió en revisar cada uno de los trabajos relacionados e identificar cuáles variables o qué datos se trabajaron como entrada al sistema desarrollado, con el fin de hacer una recomendación de fertilizantes. Así, los resultados obtenidos sirven de base para determinar las variables más comunes trabajadas en este tipo de estudios.
- **Variables más comunes:** La figura 5, muestra los resultados obtenidos en la extracción de datos de las variables de entrada estudiadas en los trabajos relacionados.

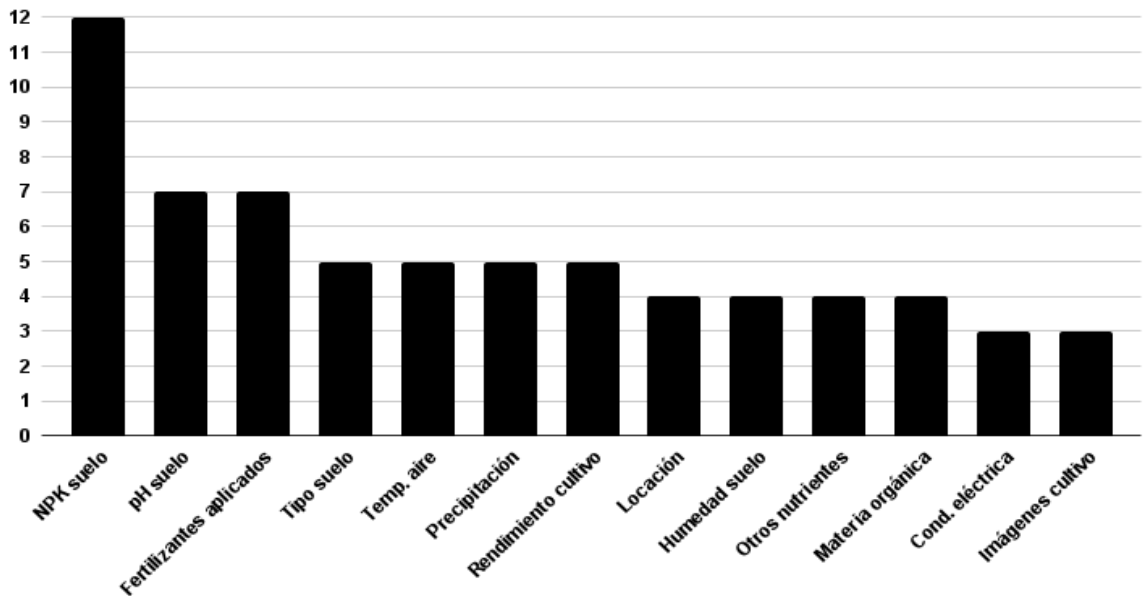


Figura 5. Variables de entrada más comunes encontradas en la RSL. Fuente propia.

En la figura anterior se observa que las variables más comunes son el nivel en el suelo de los tres macronutrientes primarios. Luego está el pH del suelo, una variable que está directamente relacionada con la presencia o ausencia de nutrientes en el suelo [76]. Los fertilizantes aplicados son la tercera variable más utilizada, ya que muchos de los trabajos desarrollan sistemas que no sólo recomiendan la cantidad de fertilizante sino también el tipo de fertilizante a aplicar, teniendo en cuenta que hay muchos tipos de fertilizantes con diferentes composiciones de nutrientes y para diferentes cultivos. Otras variables encontradas incluyen datos climáticos como la temperatura y la precipitación; datos de suelo como el tipo de suelo, humedad, materia orgánica, entre otras; y también datos de cultivo como el rendimiento anterior del cultivo.

- **Base de conocimiento:** Esta parte tiene en cuenta la base de conocimiento construida y explicada en la sección 3.4.2. Con esta información se plantean unas bases de cuáles variables pueden considerarse prioritarias en el trabajo, donde posteriormente se validan con los expertos.
- **Revisión y validación con expertos:** el presente trabajo contó con la colaboración de la empresa Ecotecma, como se explica en los antecedentes. Entre esta colaboración también se contó con reuniones con dos expertos en la agricultura y el estudio del suelo, como lo son el ingeniero ambiental Juan Fernando Casanova Olaya y la bióloga María Cristina Ordoñez Díaz. En estas reuniones se revisaron todas las variables encontradas, donde a través de un

proceso de aceptación, la variable se considera prioritaria. Caso contrario, si la variable se considera no prioritaria, entonces se descarta. Esto se hace con el fin de utilizar las variables que más prioritarias se consideren, para brindar una recomendación de la cantidad adecuada de fertilizante inorgánico en la etapa de producción del café.

- **Variables determinadas:** en esta sección, los expertos determinaron las variables a trabajar en el desarrollo del sistema, las cuales se muestran en la tabla 3. Posteriormente, se describen las razones por las cuales algunas variables de la figura 5 fueron descartadas.

*Tabla 3. Variables determinadas para el desarrollo del sistema. Fuente propia.*

Tipo de variable	Variable	Unidad de medida
Cultivo	<b>Densidad de siembra</b>	No. de plantas por hectárea (plant/ha)
Cultivo	<b>Cobertura de sombrío</b>	Porcentaje (%)
Cultivo	<b>Fecha de floración</b>	Fecha
Clima	Precipitación Pluvial (PP)	Milímetros de lluvia (mm)
Suelo	Nivel de N	mg/Kg
Suelo	Nivel de P	mg/Kg
Suelo	Nivel de K	mg/Kg
Suelo	Humedad (H)	Porcentaje (%)
Suelo	pH	Nivel de pH

De las 9 variables mostradas en la tabla, se tienen 6 que se encontraron en la extracción de datos y se aprobaron con los expertos. Así mismo hay 3 variables resaltadas en la tabla, las cuales no se encontraron en la extracción de datos. Sin embargo, con la revisión de los expertos, estas 3 variables se consideraron importantes en el estudio de cultivos de café, las cuales son la densidad de siembra, cobertura de sombrío y fecha de floración [77]. Sin embargo, estas variables no fueron encontradas en los trabajos relacionados, esto se debe a que ninguno de estos trabajos realizó su estudio en el café.



A continuación, se describen las razones de descarte de las variables de la figura 5 que no están en la tabla 3, es decir, variables que no fueron consideradas para el presente trabajo.

- Fertilizantes aplicados: Es un factor que se utiliza en trabajos que recomiendan el tipo de fertilizante a aplicar, donde se considera importante el historial de los tipos de fertilizantes aplicados en el cultivo, ya que dependiendo de la composición que éstos tenían, el suelo se veía afectado para la siguiente producción [44]. Sin embargo, los objetivos del presente trabajo se acotaron hasta la cantidad de fertilizante a aplicar. Para realizar una recomendación de tipo y composición de fertilizante, se debe hacer una nueva investigación acerca de los fertilizantes existentes en el mercado y según sus costos, cuáles pueden ser más beneficiosos para el caficultor. Lo anterior, se considera como una extensión al RS en un trabajo futuro.
- Rendimiento del cultivo: Este dato no se considera debido al alcance del presente trabajo. Es un factor que puede ser muy útil en un trabajo futuro para evaluar la utilidad y la calidad de las recomendaciones.
- Tipo de suelo: Es una variable mayormente trabajada para sistemas que no se enfocan en un sólo cultivo. En el presente trabajo sólo se tiene en cuenta el cultivo de café, el cual suele sembrarse en suelos arcillosos y derivados de cenizas volcánicas [78] [70].
- Temperatura (T): Esta variable climática es importante en los estudios de cultivos, pero se descarta en el presente trabajo por sugerencia de los expertos, quienes sugieren darle más relevancia a la precipitación, la cual tiene mayor relevancia en los periodos de floración del café.
- Locación: Es un factor utilizado en sistemas generalizados, es decir, sistemas que pueden adaptarse para usarse en diferentes regiones. En el presente trabajo, la locación se define como la región del Cauca. Un estudio futuro, puede considerar la ampliación del sistema hacia otras regiones cafeteras de Colombia.
- Materia orgánica: Variable de suelo que se descarta por la etapa del cultivo de café en la que se enfoca el presente trabajo (etapa de producción). Esto se debe a que la materia orgánica, es más importante en la etapa vegetativa, donde se relaciona directamente con los fertilizantes orgánicos [79].

- Conductividad eléctrica (CE): Este parámetro del suelo que al igual que la salinidad, mide el contenido de sales que puede tener el suelo. Generalmente este contenido ocurre cuando hay una mala calidad del agua de riego o se ha fertilizado el cultivo en exceso con nutrientes que se componen de altos contenidos de sales, afectando directamente las raíces de la planta y la absorción de nutrientes [80]. Es un parámetro que se utiliza principalmente en estudios de sistemas de riego y aplicación de tipos o composiciones de fertilizantes, lo cual dista de los objetivos del presente trabajo enfocado principalmente en la aplicación de la cantidad de fertilizantes NPK.

### 3.4.3.2. Captura de datos

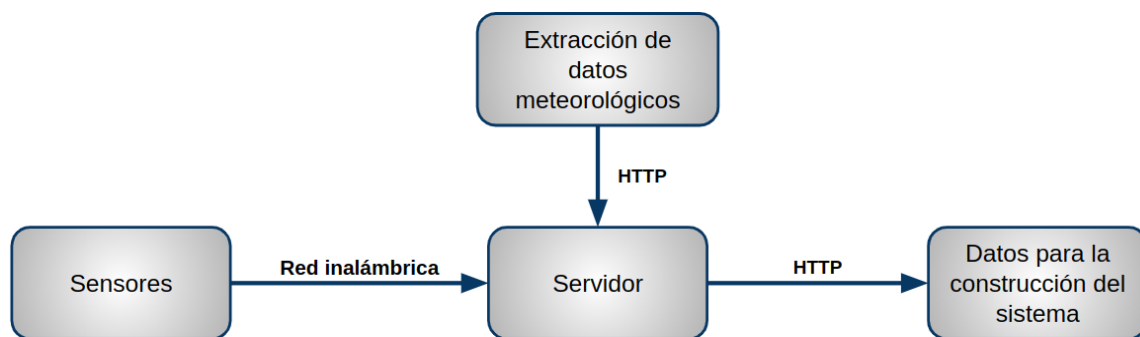


Figura 6. Proceso de captura de datos del sistema. Fuente propia.

La figura 6 muestra el proceso de captura de datos relacionados con las variables de suelo definidas en la sección anterior. Por lo tanto, para elegir los dispositivos que mejor se adaptaron a la solución propuesta, se realizó un estudio de sensores agrícolas, tomando como base una investigación ya realizada por la empresa Ecotecma y que se visualiza en la tabla 4. Con los resultados de dicha investigación se inició una fase de filtrado de características teniendo en cuenta los criterios para el cumplimiento del requerimiento, tales como: sensores que puedan medir las variables definidas en la sección anterior y que el precio sea de bajo costo. Cabe mencionar que la definición para el costo de los sensores se asumió de la siguiente manera:

- Costo bajo: < 400 USD.
- Costo medio: 400-1000 USD.
- Costo alto: > 1000 USD.

Tabla 4. Sensores de captura de datos de suelo. Fuente [81].

Sensor	Parámetros de medida	Costo	Precio (USD)
Watermark 200ss	H	BAJO	50.24
FieldScout TDR 350	H, CE, T	ALTO	1.524
ECH2O EC-5	H	BAJO	162.56
ECH2O HS-10	H	BAJO	193.98
5TE/5TM	H, CE, T, S	BAJO	153.45
Teros 10	H, CE, T, S	BAJO	153.45
Teros 11/12	H, CE, T, S	BAJO	241.8
LaquaTwin CE	CE	BAJO	196
Hydra Probe II	H, T, S	MEDIO	425
WaterScout SMEC 300	H, T, S	BAJO	239
WaterScout SM 100	H	BAJO	95
WaterMark soil moisture sensor	H	BAJO	42
THERM 200	T	BAJO	54
ZD-06	H, pH	BAJO	58.90
LaquaTwin pH	pH	BAJO	196
FieldScout soilstik Ph	pH	BAJO	169
LightScout silicon pyranometer	Radiación Solar	BAJO	269
LightScout quantum light sensor	Radiación Fotosintéticamente Activa (PAR)	BAJO	239
Chlorophyll meter	Clorofila (CH)	ALTO	2900
LaquaTwin Na	Na	MEDIO	435
LaquaTwin K	K	MEDIO	435
LaquaTwin NO3	NO3	MEDIO	435
LaquaTwin Ca	Ca	MEDIO	495
GS1 DH	H	BAJO	88.64
EC-5	H	BAJO	124.39
RT-1	T	BAJO	121.48
TERALYTIC	N, P, K	BAJO	120
HONDETEC	CE, T, H, Salinidad, N,P, K	BAJO	400

Además del estudio de los sensores, también se realizó un estudio de las tecnologías inalámbricas existentes, para conocer las características, ventajas y limitaciones de cada una. Esta investigación se detalla en el anexo E, donde se evidenció que la tecnología Lora es la más común en el campo agrícola debido a sus ventajas de bajo costo, bajo consumo de energía y una cobertura no tan pequeña. Teniendo en cuenta que fue esta tecnología la que ya estaba implementada por Ecotecma, haciendo que la investigación realizada tenga concordancia con la tecnología escogida por la empresa para realizar sus investigaciones. Cabe mencionar que la principal desventaja de Lora es que no maneja datos en tiempo real, aspecto que no es relevante en el desarrollo del presente trabajo.

Luego, como resultado del filtrado y de acuerdo con los criterios de sensores de bajo costo y que midan las variables definidas anteriormente, se obtuvieron los sensores mostrados en la tabla 5. De lo anterior se pudo determinar que el sensor que cumplió con las características necesarias para el desarrollo del proyecto fue el HONDETEC, que además opera bajo la tecnología inalámbrica Lora [82].

*Tabla 5. Sensores que cumplen los criterios de selección. Adaptado de [81].*

Sensor	Parámetros de medida	Manuales de operación	Costo	Precio (USD)
Watermark 200ss	H	Si	BAJO	50.24
ECH2O EC-5	H	SI	BAJO	162.56
ECH2O HS-10	H	SI	BAJO	193.98
5TE/5TM	H, EC, T, S	SI	BAJO	153.45
Teros 10	H, EC, T, S	SI	BAJO	153.45
Teros 11/12	H, EC, T, S	SI	BAJO	241.8
THERM 200	T	SI	BAJO	54
LightScout silicon pyranometer	RS	SI	BAJO	269
LightScout quantum light sensor	PAR	SI	BAJO	239
EC-5	H	SI	BAJO	124.39
RT-1	T	SI	BAJO	121.48
HONDETEC	EC, T, H, Salinidad, N, P, K	SI	BAJO	400

Una vez definido el sensor a utilizar se procedió a la instalación de éste para la transmisión de datos por la red inalámbrica ya implementada por la empresa Ecotecma. La figura 7 muestra la arquitectura de la red inalámbrica de sensores, la cual se implementó en una pequeña finca en el municipio de Piendamó y se le agregó un servicio de captura de datos meteorológicos de las estaciones climáticas de Cenicafé.

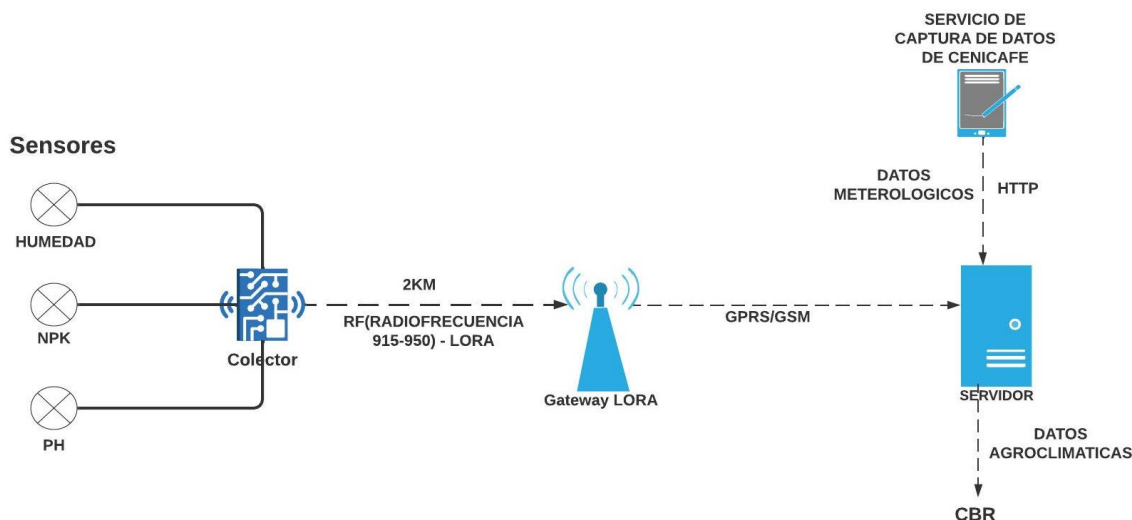


Figura 7. Arquitectura de red de sensores Ecotecma. Adaptado de [83].

Posteriormente de tener los sensores instalados se procedió a capturar y enviar datos de las variables de suelo mencionadas anteriormente en la tabla 3 a un colector que cumplió con la función de transmitir por radiofrecuencia (915Mhz) la información recolectada a una gateway Lora. Esta a su vez, vía GSM envía esos datos a un servidor de la empresa quien es el encargado de almacenarlos para su siguiente uso.

En cuanto a los datos meteorológicos históricos, fueron extraídos de los anuarios de Cenicafé que corresponden a datos provenientes de la red de estaciones meteorológicas que posee la Federación Nacional de Cafeteros en el país [84]. El rango temporal de estos datos se ubica a partir del año 2006 hasta la actualidad. Para acceder a estos datos, en el presente trabajo de grado, en colaboración con una de las pasantías de Ecotecma, se realizó una petición HTTP al servidor de almacenamiento de los datos de Cenicafé para su posterior análisis y posterior preparación junto con las variables de entrada del sistema. El proceso de la obtención de los datos de la red de sensores y las estaciones climáticas, se especifica de forma más detallada en el anexo F.

### 3.4.4. Descripción de los datos

Habiendo determinado las variables a trabajar en la sección 4.2.3.1, se realizó una investigación para determinar los rangos de valores posibles de cada una de estas variables. Así como los rangos de valores bajos, altos y normales (que serían los valores más adecuados). También se indican los valores máximos y mínimos de cada variable, que pueden darse en un cultivo de café. Estos rangos se calculan con base a los estudios de [77] y [85], los cuales realizan análisis de suelos y determinan los valores más adecuados para las variables de N, P, K y pH del suelo. A continuación, se describen estos valores:

- Nitrógeno del suelo: sus valores se pueden encontrar entre 0 y 225 mg/kg para cultivos de café. El valor adecuado para cultivos de café se encuentra entre 51 mg/kg y 87 mg/kg. Cuando el valor del nitrógeno está por debajo de 51 mg/kg se puede decir que hay una deficiencia de este nutriente y por lo tanto la planta va a presentar efectos adversos en la clorofila (importante para la fotosíntesis) y la planta dejará de crecer [64]. Si el nitrógeno excede el valor de 87 mg/kg, significa que habrán residuos o desperdicios del nutriente, lo que afectaría económicamente al agricultor y contaminaría el ambiente.
- Fósforo del suelo: el rango posible de valores en el suelo se encuentra entre 0 mg/kg y 80 mg/kg, para cultivos de café los valores adecuados se encuentran en el rango entre 10 mg/kg y 20 mg/kg. Si están por debajo de los 10 mg/kg la planta puede producir un amarillamiento desuniforme en las hojas más viejas, acompañado de manchas rojizas y en casos severos defoliación [64]. Si el valor excede los 20 mg/kg se le denomina un valor alto de P en el suelo y puede ocasionar el bloqueo en la absorción del boro en las plantas [86].
- Potasio del suelo: sus valores se pueden encontrar entre 0 mg/kg y 546 mg/kg. Los valores adecuados para cultivos de café están entre 78 mg/kg y 156 mg/kg. Si el valor del potasio está por debajo de los 78 mg/kg el tamaño del fruto de la planta se reducirá y producirá defoliación de las hojas [64]. Si el valor del potasio es superior a 156 mg/kg lo que ocasionaría un bloqueo en la absorción de micronutrientes en las plantas [87].
- pH: se mide en una escala de niveles que va desde 0 a 14. El valor adecuado para cultivos de café debe estar entre 4.5 y 5.5, si el pH del suelo se encuentra por debajo de 4.5 se dice que el suelo es más ácido y afecta el crecimiento de las raíces de la planta lo que conlleva a que la planta no pueda absorber de manera adecuada los nutrientes; cabe mencionar que los suelos ácidos también bloquean la absorción de K y N, así como favorecen la absorción de Mn a niveles que

pueden causar toxicidad a los cultivos. Si está por encima de 5.5 se dice que el suelo es alcalino y va generar un bloqueo en la absorción de P, Fe, Zn y Cu, es decir que habrá menos cantidad de estos nutrientes [64], [88].

- Densidad de siembra: para una densidad de siembra baja se toman valores menores a 5000 árboles, para una media valores entre 5000 y 6000 y para una alta valores mayores a 6000 y se define como el número de plantas por unidad de área del terreno. Su efecto es importante sobre la producción del cultivo y se considera como un insumo. Se argumenta que está relacionada con los efectos que produce en la planta, la competencia por el consumo de nutrientes con otras plantas de la misma especie o de otra especie [89].
- Cobertura de sombrero: se mide en porcentajes y sus valores se pueden encontrar bajos cuando están por debajo del 30%, medios entre el 30% y el 60% y altos para valores mayores a 60%. Cabe resaltar que para un buen desarrollo en los cultivos de café se debe controlar tanto la disponibilidad de agua como la incidencia de luz solar, ya que un buen manejo del sombrero en el cultivo contribuye a mantener la fertilidad del suelo, reciclar nutrimentos y reducir la erosión (importante en temporadas secas) [90].
- Temporada: Es una variable que indica el comportamiento del clima en un periodo determinado y se clasifica en normal, lluviosa o seca [70].

# CAPÍTULO 4

## PREPARACIÓN DE LOS DATOS

En el presente capítulo se desarrolló la fase de correspondiente a la preparación de los datos para la construcción del modelado del sistema y para la entrada del sistema.

### 4.1. Datos de modelado del sistema

En esta sección se explica cómo se estructura un caso y se construye la base de casos partiendo de la recolección realizada en la sección anterior.

#### 4.1.1. Estructura de un caso

Un caso en el CBR se compone de un problema o situación determinada y la solución encontrada a ese problema. Así mismo, el problema se compone de las variables que se determinaron en la sección anterior, con excepción de las variables de fecha de floración y precipitación pluvial (PP), las cuales se usan en un análisis por aparte para determinar la temporada climática. La tabla 6 muestra los rangos de clasificación de las variables del problema, así como los valores mínimos y máximos posibles, que se identificaron en la descripción de los datos.

*Tabla 6. Rangos de clasificación de valores en las variables determinadas. Fuente [77].*

Variable	Bajo	Medio	Alto	Unidad	Mín - Máx
Densidad de siembra	> 5000	5000 - 6000	> 6000	No. de plantas por ha	2500 - 7500
Cobertura sombrío	< 0.3	0.3 - 0.4	> 0.4	%	0 - 100
Humedad del suelo	< 0.3	0.3 - 0.6	> 0.6	%	0 - 100
Nitrógeno	< 51	51 - 87	> 87	mg / Kg	0 - 225
Fósforo	< 10	10 - 20	> 20	mg / Kg	0 - 80
Potasio	< 78	78 - 156	> 156	mg / Kg	0 - 546
pH	< 4.5	4.5 - 5.5	> 5.5	nivel	3 - 7

Adicional a estas variables, se tiene la temporada que se representa de la siguiente manera en la base de casos:

- Temporada seca: Se representa con el valor de 0.
- Temporada normal: Se representa con el valor de 0.5.



- Temporada lluviosa: Se representa con el valor de 1.

La solución del problema de un caso del CBR se determina mediante la información experta que se recolecta en la recopilación documental (sección 4.2.2). Con esta base de conocimiento se determinaron las condiciones de los nutrientes en el suelo ante diferentes situaciones en el problema. Esta solución se compone de 5 variables descritas a continuación:

- Tasa de fertilizante N: se refiere a la cantidad de fertilizante nitrogenado que se recomienda aplicar al cultivo.
- Tasa de fertilizante P: se refiere a la cantidad de fertilizante fosfórico que se recomienda aplicar al cultivo.
- Tasa de fertilizante K: se refiere a la cantidad de fertilizante potásico que se recomienda aplicar al cultivo.
- Recomendación adicional A: se refiere a prácticas agrícolas que brinda el sistema de forma opcional, para mejorar las condiciones de fertilidad en el suelo.
- Recomendación adicional B: se refiere a prácticas agrícolas que brinda el sistema de forma opcional, para mejorar las condiciones de pH en el suelo.

Las primeras 3 variables se refieren a 3 tasas base de fertilizantes ante ciertas condiciones en el suelo, cada una corresponde a los 3 macronutrientes principales en los cultivos (N, P, K) y los posibles valores que pueden tener se describen a continuación:

- **Tasa baja:** Se determina cuando hay una condición favorable en el suelo, es decir, un buen nivel de fertilidad en el suelo (valores altos del nutriente), no hay riesgos de pérdidas de nutrientes por factores climáticos (es decir, temporada y humedad del suelo normales) y el pH del suelo está en el rango adecuado. Representada con un 1 en la base de casos, para facilidad en el manejo de datos.
- **Tasa normal:** Asignada cuando la fertilidad del suelo está normal (valores medios) y la temporada como la humedad del suelo están normales. A esto se le denomina condición normal en el suelo. Representada con un 2 en la base de casos.
- **Tasa alta:** Ocurre cuando la fertilidad en el suelo no tiene los niveles adecuados o cuando la temporada, humedad y ph en el suelo no es lo adecuado. A esto se le denomina una condición no favorable en el suelo. Por ejemplo, un nivel de N en el suelo normal pero con temporada seca, en este caso la humedad baja y un suelo más ácido de lo normal hacen que sea una condición de suelo no favorable para el nutriente de N. Representada con un 3 en la base de casos.
- **Tasa muy alta:** Aquí se asignan los peores casos posibles para los nutrientes en el suelo. Por ejemplo, cuando hay escasez de nutrientes en el suelo, altos niveles de humedad en temporadas lluviosas (ocasionando pérdidas de nutrientes por

escorrentía y lixiviación) o bajos niveles de humedad en temporadas secas (ocasionando pérdidas por erosión), así como bajos o altos niveles de pH. A estas condiciones se les denomina condiciones críticas en el suelo. Representada con un 4 en la base de casos.

Luego, las otras 2 variables corresponden a prácticas agrícolas recomendadas por la información experta bajo ciertas condiciones. Debido a que en situaciones de sequía y déficit de humedad en el suelo, las plantas tendrán dificultades de absorción de nutrientes, sin importar la aplicación de cantidades altas de fertilizante [91]. Estas variables se consideran opcionales o extras, la recomendación A se denominó como “*Plus A*” y la recomendación B se denominó como “*Plus B*” en la base de casos. Estas recomendaciones se describen a continuación:

- **Plus A:** Si los datos del problema indican humedad baja en el suelo y la temporada es seca, se recomiendan prácticas agrícolas como tratar de mejorar el riego en el cultivo o aplicar el fertilizante de forma húmeda o líquida [91]. Esta recomendación se representa con un 0 en la base de casos. En el caso contrario, para una humedad alta en una temporada lluviosa, se recomienda mejorar el sistema de drenaje en el cultivo para evitar acumulaciones de agua en las plantas, ya que un exceso de humedad puede afectar el oxígeno en la planta [92]. En la base de casos, esto se representa con un 2.
- **Plus B:** Si el pH indica un suelo muy ácido se recomienda hacer uso de cal, yeso o algún otro material alcalinizante para corregir la acidez [77]. En la base de casos esto se representa con un 0. Si el pH indica un nivel alto en el suelo, se recomienda acidificar el suelo aplicando azufre, sulfato de hierro o materia orgánica [93], esta recomendación se representa en la base de casos con un 2.

Cabe mencionar, que el valor de 1 en ambas variables indicaría que no es necesario realizar estas recomendaciones. Finalmente, con todas las variables explicadas en el problema y la solución de un caso, se procede a construir la base de casos.

#### **4.1.2. Base de casos**

La base de casos es el núcleo de todo sistema CBR, porque dependen directamente de la existencia de una base de casos para poder realizar todos los pasos en un ciclo CBR [94]. Un caso puede ser construido por expertos humanos o experiencias pasadas que haya tenido el sistema. Los casos se pueden representar de muchas maneras, como reglas, fórmulas lógicas, marcos y registros de bases de datos [95]. En el presente trabajo, la base de casos fue construida teniendo en cuenta la base de conocimiento que se explica en el proceso de recolección de datos y la determinación de las variables a usar en el sistema. Posteriormente, esta base de casos se validó con los expertos.

A continuación, se tomaron diferentes combinaciones posibles de datos del problema y se determinaron las soluciones más adecuadas de acuerdo con la base de conocimiento experto. La tabla 7 muestra un ejemplo con 5 casos construidos con el proceso anteriormente descrito y se explican posteriormente. Las celdas que se muestran en color rojo, indican que el valor de la variable correspondiente se clasifica como bajo, si es color verde indica valor medio o normal, y el color amarillo indica un valor alto.

Tabla 7. Base de casos con 5 ejemplos. Fuente propia.

Problema								Solución				
Densidad	Sombrío	Temporada	Humedad	N	P	K	pH	Tasa N	Tasa P	Tasa K	Plus A	Plus B
4500	0.3	0	0.3	60	6	100	4.5	2	3	2	1	1
5500	0.3	1	0.4	70	27	100	4	2	1	2	1	0
6500	0.2	1	0.4	50	21	180	6	3	1	1	1	2
5700	0.38	1	0.75	45	8	125	5	4	4	3	2	1
6500	0.18	0	0.1	105	11	145	4.2	2	3	3	0	0

El primer caso mostrado en la tabla anterior, muestra una densidad de siembra baja con una cobertura de sombrero normal, la temporada es seca pero con una humedad normal, el N y K tienen valores normales pero el valor de P está bajo. Por esto se asigna una condición de suelo normal para el N y K pero no favorable para el P y no se requieren recomendaciones adicionales.

El segundo caso muestra todos los valores normales excepto el P que está alto y el pH que está bajo. Además de una temporada lluviosa. Lo que indica nuevamente una condición normal para N y K pero una condición favorable para el P. También se recomienda tratar la acidez del pH.

El tercer caso tiene una densidad alta, sombrero bajo, temporada lluviosa, humedad normal, N bajo, P, K y pH altos, lo que indica una condición no favorable para el N y condiciones favorables para el P y K, junto con el extra de tratar el valor alto de pH.

El cuarto caso consta de una densidad y sombrero normales, humedad alta en temporada lluviosa, lo cual no es un factor positivo, además, el N y P están bajos. Por lo tanto, las condiciones climáticas pueden incidir en la pérdida de nutrientes en el suelo por factores como la escorrentía superficial o la lixiviación, lo que hace que se tenga una condición crítica para el N y P, y condición no favorable para el K. Además de recomendar que se mejore el sistema de drenaje.

Por último, el quinto caso de ejemplo mostrado en la tabla, tiene una densidad alta de siembra, la temporada es seca, el sombrero es bajo y la humedad es baja. Al tener un

sombrío bajo, se ven afectados negativamente los nutrientes en el suelo, ya que pueden ocurrir pérdidas por erosión. Lo que significa que, a pesar de tener niveles normales de P y K, es mejor aplicar tasas altas en estos fertilizantes, con el fin de suplir las condiciones climáticas no favorables que hay en este caso. Así mismo, se recomienda mejorar el riego en el cultivo y tratar el nivel bajo de pH.

De manera similar, el presente trabajo de grado desarrolló una base de casos con un total de 300 problemas con diversas combinaciones de datos para representar varias situaciones y sus respectivas soluciones expertas, consolidando así una base de 300 casos. Posteriormente la base fue validada con los expertos y se almacenó en formato CSV que se puede detallar en el anexo G.

## 4.2. Datos de entrada del sistema

Los datos de entrada del sistema CBR propuesto en este trabajo de grado se pueden considerar un nuevo problema que requiere de una solución que puede no estar considerada en la base de casos. Estos datos son las variables que se determinaron en la comprensión del negocio, las cuales en esta fase se van a preparar y transformar para la construcción del nuevo problema. En la figura 8 se muestra el proceso de preparación y transformación de las variables determinadas para la construcción del problema.

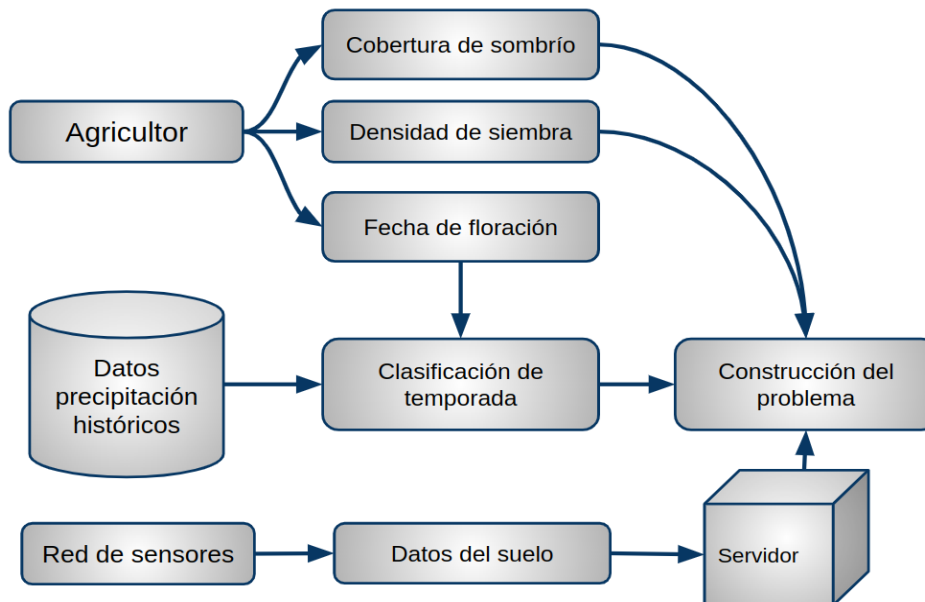


Figura 8. Proceso de construcción de un nuevo problema. Fuente propia.

El problema se empieza a construir desde los datos que brinda el agricultor, los cuales son la densidad de siembra, el sombrío y la última fecha de donde ocurrió la floración del café. Así mismo, con los datos de PP históricos, se clasifica la temporada climática para los dos meses siguientes a esa fecha de floración. Además, con los datos de la red de

sensores almacenados en un servidor, se determina el estado del suelo para esa misma fecha. Finalmente, con todos estos datos, se arma el problema de un nuevo caso que se ingresará al sistema CBR. Pero antes, se explica el proceso de cómo se clasifica una temporada climática a partir de una fecha de floración.

#### 4.2.1. Clasificación de temporada

Para realizar una clasificación de la temporada climática con base a la fecha de la última floración ocurrida en el cultivo, se hace necesario hacer una clasificación de cómo han sido las temporadas climáticas desde el año 2006 hasta la fecha establecida. Estas temporadas fueron identificadas con base en los registros históricos de periodos secos o lluviosos a lo largo de cada año, que se toman del *Servicio Meteorológico Nacional de EE.UU.* (NWS, por sus siglas en inglés). Este servicio contiene registros desde la década de los 50 hasta la actualidad [96]. Donde, dichos registros se refieren a los períodos secos, normales o lluviosos ocurridos en el océano pacífico. Ya que una clasificación de temporada climática depende de la ocurrencia de los fenómenos del El Niño y La Niña, y dicha ocurrencia depende de las temperaturas registradas en el océano pacífico. La figura 9 muestra los registros mencionados:

2006	-0.9	-0.8	-0.6	-0.4	-0.1	0.0	0.1	0.3	0.5	0.8	0.9	0.9
2007	0.7	0.2	-0.1	-0.3	-0.4	-0.5	-0.6	-0.8	-1.1	-1.3	-1.5	-1.6
2008	-1.6	-1.5	-1.3	-1.0	-0.8	-0.6	-0.4	-0.2	-0.2	-0.4	-0.6	-0.7
2009	-0.8	-0.8	-0.6	-0.3	0.0	0.3	0.5	0.6	0.7	1.0	1.4	1.6
Year	DJF	JFM	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDJ
2010	1.5	1.2	0.8	0.4	-0.2	-0.7	-1.0	-1.3	-1.6	-1.6	-1.6	-1.6
2011	-1.4	-1.2	-0.9	-0.7	-0.6	-0.4	-0.5	-0.6	-0.8	-1.0	-1.1	-1.0
2012	-0.9	-0.7	-0.6	-0.5	-0.3	0.0	0.2	0.4	0.4	0.3	0.1	-0.2
2013	-0.4	-0.4	-0.3	-0.3	-0.4	-0.4	-0.4	-0.3	-0.3	-0.2	-0.2	-0.3
2014	-0.4	-0.5	-0.3	0.0	0.2	0.2	0.0	0.1	0.2	0.5	0.6	0.7
2015	0.5	0.5	0.5	0.7	0.9	1.2	1.5	1.9	2.2	2.4	2.6	2.6
2016	2.5	2.1	1.6	0.9	0.4	-0.1	-0.4	-0.5	-0.6	-0.7	-0.7	-0.6
2017	-0.3	-0.2	0.1	0.2	0.3	0.3	0.1	-0.1	-0.4	-0.7	-0.8	-1.0
2018	-0.9	-0.9	-0.7	-0.5	-0.2	0.0	0.1	0.2	0.5	0.8	0.9	0.8
2019	0.7	0.7	0.7	0.7	0.5	0.5	0.3	0.1	0.2	0.3	0.5	0.5
Year	DJF	JFM	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDJ
2020	0.5	0.5	0.4	0.2	-0.1	-0.3	-0.4	-0.6	-0.9	-1.2	-1.3	-1.2
2021	-1.0	-0.9	-0.8	-0.7	-0.5	-0.4	-0.4	-0.5	-0.7	-0.8	-1.0	-1.0
2022	-1.0											

Figura 9. Registros históricos de la NWS. Fuente [96].

La figura muestra los períodos secos (en rojo) y lluviosos (en azul) que han ocurrido desde el año 2006 hasta la actualidad. Tomando como base la Información sugerida por

expertos, para cada año se identificaron secuencias de 5 períodos secos o lluviosos, para clasificar esos años como secos o lluviosos respectivamente. Si en algún año, no ocurrieron estos 5 períodos consecutivos, entonces el año se catalogó como normal. A continuación, se muestra la tabla con la clasificación histórica de los años 2006 hasta el 2021.

*Tabla 8. Clasificación de temporada para los años 2006-2021. Fuente propia.*

<b>Año</b>	<b>Temporada</b>	<b>Año</b>	<b>Temporada</b>
2006	Normal	2014	Normal
2007	Lluvioso	2015	Seco
2008	Lluvioso	2016	Normal
2009	Seco	2017	Normal
2010	Lluvioso	2018	Normal
2011	Lluvioso	2019	Seco
2012	Normal	2020	Normal
2013	Normal	2021	Lluvioso

Con estas clasificaciones históricas, se realizó el análisis de datos históricos para definir una nueva clasificación de la temporada climática del año donde se haya dado la última fecha de floración en el cultivo. Estos datos son los que se tomaron de los anuarios de Cenicafé y se pueden observar en detalle en el anexo H. La clasificación mencionada fue realizada con base en la información sugerida por expertos. La figura 10 describe este proceso.

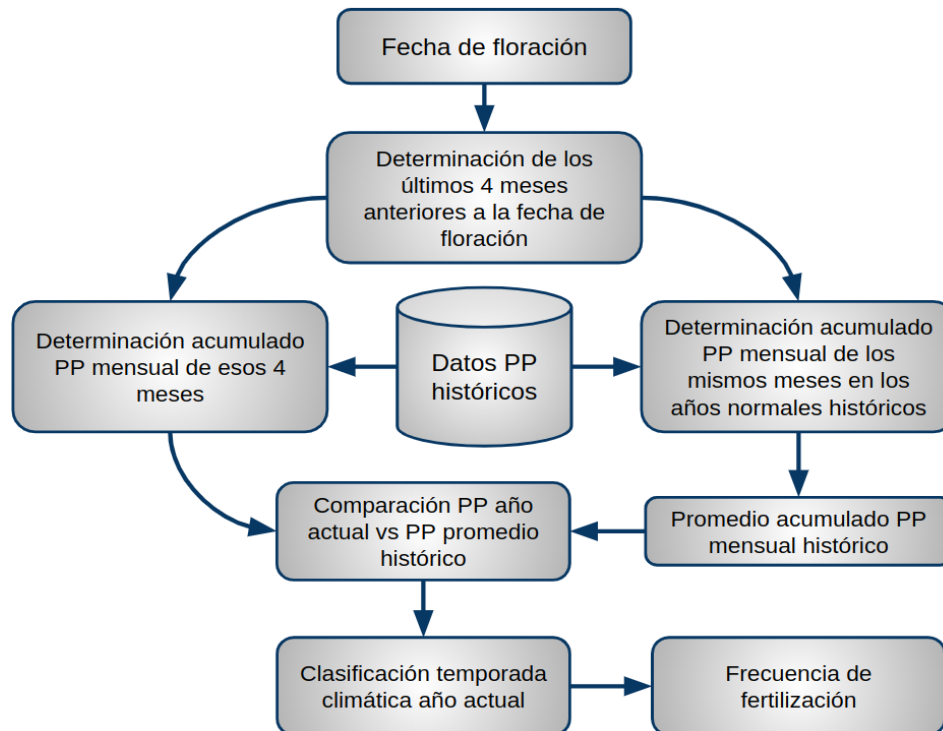


Figura 10. Proceso de clasificación de temporada. Fuente propia.

La figura anterior muestra paso a paso la realización de una clasificación de temporada dada una fecha de floración. A continuación, se describe cada paso mostrado en la figura:

- **Determinación de los últimos 4 meses anteriores a la fecha de floración:** recibiendo como entrada la fecha de la última floración, el sistema toma las 120 precipitaciones registradas en los 120 días anteriores al día de la floración (que serían los 4 meses anteriores a la fecha dada).
- **Datos PP históricos:** se refiere a los datos de PP almacenados en los anuarios de Cenicafé. En la sección 4.2.3.2 se menciona que estos anuarios recolectan datos de más de 30 estaciones meteorológicas en el país. De las cuales sólo 2 están ubicadas en el Cauca. Se escogió la estación climática ubicada en el municipio de Piendamó por razones de cercanía y accesibilidad al transporte.
- **Determinación del acumulado de PP mensual:** Estos 120 datos fueron divididos en 4 conjuntos de 30 datos cada uno y para cada conjunto se realizó una suma de los valores, de facilidad en el transporte a esa zona. Los datos de PP históricos de esta estación obteniendo así un acumulado de PP para los últimos 30 días con respecto al día de la floración (denominado acumulado PP mes 1). Así mismo, un acumulado entre los 31 y 60 días anteriores (acumulado PP mes 2), un acumulado

entre los 61 y 90 días anteriores (acumulado PP mes 3) y un acumulado entre los 91 y 120 días anteriores (acumulado PP mes 4).

- **Determinación del acumulado de PP mensual en los años históricos:** el sistema replica el mismo proceso para los años entre el 2006 y el año anterior al año de la fecha de floración. Se toman los años que se catalogaron normales, y se obtienen 4 acumulados en cada año para esos mismos meses de los acumulados calculados en el año de la fecha de floración.
- **Promedio acumulado de PP mensual histórica:** en esta parte, se calculó el promedio de PP acumulada histórica para el respectivo mes. Obteniendo 4 promedios de PP acumulada mensual histórica. La tabla 9 muestra que para el cálculo de cada promedio, se toman 8 acumulados históricos.

*Tabla 9. Acumulados de PP en los años históricos catalogados con temporada normal. Fuente propia.*

2006	2012	2013	2014	2016	2017	2018	2020	Promedio
Acumulado PP mes 1	Acumulado PP mes 1	Acumulado PP mes 1	Acumulado PP mes 1	Acumulado PP mes 1	Acumulado PP mes 1	Acumulado PP mes 1	Acumulado PP mes 1	Promedio PP mes 1
Acumulado PP mes 2	Acumulado PP mes 2	Acumulado PP mes 2	Acumulado PP mes 2	Acumulado PP mes 2	Acumulado PP mes 2	Acumulado PP mes 2	Acumulado PP mes 2	Promedio PP mes 2
Acumulado PP mes 3	Acumulado PP mes 3	Acumulado PP mes 3	Acumulado PP mes 3	Acumulado PP mes 3	Acumulado PP mes 3	Acumulado PP mes 3	Acumulado PP mes 3	Promedio PP mes 3
Acumulado PP mes 4	Acumulado PP mes 4	Acumulado PP mes 4	Acumulado PP mes 4	Acumulado PP mes 4	Acumulado PP mes 4	Acumulado PP mes 4	Acumulado PP mes 4	Promedio PP mes 4

- **Comparación PP y clasificación de temporada:** con la PP acumulada mensualmente en cada uno de los 4 meses analizados y el promedio de PP acumulada durante esos mismos meses en los años históricos, el sistema procedió a realizar una comparación entre estos datos. Si el resultado de esa comparación muestra que la PP acumulada en el año de la fecha de floración supera por más de un 35% al promedio histórico de PP calculada, entonces la temporada climática al momento de esa fecha de floración se cataloga como lluviosa. Por el contrario, si la comparación muestra que la PP acumulada es menor en un 25% al promedio histórico, entonces la temporada se cataloga como



seca. Si estas dos condiciones no se cumplen, la temporada se cataloga como normal.

- **Frecuencia de fertilización:** como se explica en la sección 4.1.2, la temporada indica cuántas veces se deben aplicar los fertilizantes en el cultivo durante el período de fertilización. Entonces la frecuencia de aplicación para el presente trabajo se define así:
- **Para una temporada lluviosa:** Aplicar tres veces. A los 2, 4 y 6 meses después de la floración del cultivo de café.
- **Para una temporada seca:** Aplicar dos veces. A los 3 y 6 meses después de la floración del cultivo de café.

Se tiene en cuenta que, si la temporada se clasifica como normal, se prioriza la frecuencia de 3 aplicaciones por período de fertilización. Cabe mencionar que el script que realiza este análisis de datos, se puede detallar en el anexo I.

#### 4.2.2. Construcción de un nuevo problema

Habiendo clasificado la temporada correspondiente al momento de la floración del cultivo, se procede a reunir todas las variables que se consideran como entradas al sistema CBR, recordando que una entrada al CBR se denomina “*Nuevo problema*”. El problema se compone de 8 variables descritas en la sección 4.2.4 y se muestran de forma resumida en la tabla 10 junto con la fuente de donde se captura cada variable.

Tabla 10. Nuevo problema que ingresa al sistema. Fuente propia.

Tipo variable	Variable	Fuente de recolección
Cultivo	Densidad de siembra	Agricultor
Cultivo	Cobertura de sombrío	Agricultor
Clima	Temporada	Clasificación previamente realizada
Suelo	Humedad	Red de sensores
Suelo	Nivel de N	Red de sensores
Suelo	Nivel de P	Red de sensores
Suelo	Nivel de K	Red de sensores
Suelo	pH	Red de sensores

Finalmente, habiendo identificado cada una de las variables de entrada del sistema de CBR, se procede a hacer una normalización de datos.

### 4.2.3. Normalización del problema

Los datos del problema representan de forma independiente una característica del ambiente donde se encuentra el cultivo, por ende, trabajan en unidades diferentes. La normalización se refiere a cambiar los valores de cada variable a un rango entre 0 y 1, así se evita que el análisis de los datos en el modelado no sea sensible a la escala de las características del ambiente. Las escalas no contribuyen por igual a la función de ajuste y aprendizaje del modelo, ocasionando un sesgo; por lo que la normalización permite obtener una mejor precisión del sistema [97], [98]. Entonces, el primer paso es determinar los valores mínimos y máximos que pueda tomar cada variable en la realidad, por lo que en la recolección documental también se revisaron trabajos, estudios, datos recolectados e históricos para la determinación de estos valores. La ecuación (1) describe el proceso de normalización para cada dato en el problema entrante.

$$x_{Normalizado} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

El proceso se repite para todos los datos del problema, teniendo así un problema normalizado, que es el que ingresa al modelado del CBR, que se explica en el siguiente capítulo.

# CAPÍTULO 5

## MODELADO

La fase 4 de la metodología CRISP-DM consta del modelado del sistema, esta fase obtiene y describe el modelo del sistema CBR desarrollado. En general el funcionamiento de un sistema CBR se compone de los siguientes pasos: recuperación, reutilización, revisión y recepción [99]. Estos pasos, también son conocidos como las 4 R, y describen todo el proceso de funcionamiento de un CBR. Teniendo en cuenta lo anterior, el presente trabajo de grado abarca la construcción de los dos primeros pasos del CBR, que son la obtención de los casos más similares y la determinación de una nueva solución.

Como el sistema se enfoca en la recomendación de fertilizantes en el cultivo de café, una validación de las recomendaciones sólo se puede hacer cuando se tenga el resultado de la producción o cosecha de ese cultivo, que es el momento donde se conocerá si el sistema brindó una buena recomendación o no. Por esta razón, el sistema no alcanza a abarcar el tercer paso del ciclo CBR, el cual depende de una validación o calificación de la solución determinada en el segundo paso, y el cuarto paso depende del tercer paso para agregar ese nuevo caso a la base.

El código del sistema fue desarrollado en el lenguaje de programación Python, con ayuda de las librerías de análisis y estructuras de datos Pandas y Numpy. La plataforma de trabajo fue seleccionada Google Colab, una plataforma colaborativa en la nube, para realizar tareas de aprendizaje automático, análisis de datos y educación. Las ventajas de trabajar en esta plataforma son las siguientes:

- No requiere descarga ni instalación en equipos locales, todo lo hace en línea.
- Ya está configurado para trabajar con las librerías de pandas y python, lo que evita las descargas locales de dichas librerías.
- Todo el procesamiento del código realizado se ejecuta en un servidor en línea.
- La facilidad de trabajar colaborativamente.

El código desarrollado se encuentra en el anexo J. La figura 11 muestra el diagrama del modelado del sistema.

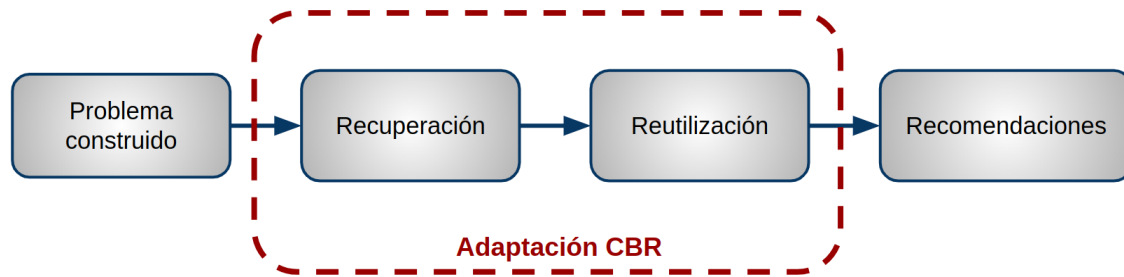


Figura 11. Diagrama de modelado del sistema CBR. Fuente propia.

Con el problema que se construyó en la sección anterior, que contiene todos los datos de entrada al sistema CBR, se procedió a la realización de los pasos de recuperación y reutilización.

### 5.1. Recuperación

Dado un nuevo problema, el CBR utiliza un algoritmo para recuperar los casos más similares al nuevo problema. Este algoritmo de recuperación se basa en los índices y la organización de la base de casos del sistema para dirigir la búsqueda a casos potencialmente útiles. Los métodos más conocidos para la recuperación de casos son: el vecino más cercano, la inducción, la inducción guiada por conocimiento y la recuperación de plantillas [100]. A continuación, se describen las razones del método seleccionado y el descarte de los demás.

- **Vecino más cercano:** se enfoca en evaluar la similitud entre los casos almacenados y el nuevo caso entrante. La principal ventaja es que permite ponderar cada característica de un caso, haciendo que la similitud pueda variar de acuerdo con los pesos asignados. Por esta razón, se escoge este método para el desarrollo del RS basado en CBR.
- **Inducción:** este método genera una estructura de tipo árbol de decisiones para organizar los casos almacenados de acuerdo de acuerdo con una selección previa de características. Resulta útil cuando se requiere una sola característica de caso como solución. Lo que hace que este método no se pueda adaptar a los objetivos del presente trabajo.
- **Inducción guiada por conocimiento:** este método realiza el mismo proceso del método anterior, pero con la diferencia de que la identificación de características se hace de forma manual (por medio de un experto o con base en información experta). Haciendo que sea enfocado a sistemas no automatizados, haciendo que no se adapte al sistema planteado en este trabajo.

- **Recuperación de plantillas:** este método devuelve todos los casos que se ajustan a ciertos parámetros. Lo que hace que no sea eficaz para casos donde se ingresan todas las características que componen el problema de un caso.

Habiendo identificado el método a trabajar en el paso de recuperación, se determina que el algoritmo a utilizar es el KNN (*K-Nearest Neighbors*), el cual puede ser utilizado para clasificación o regresión [101]. Este algoritmo utiliza un factor  $k$ , que indica el número de vecinos más cercanos (similares) que se deben tener en cuenta para posteriormente hacer una predicción. Luego, se inició el proceso de encontrar los casos más similares, el cual se describe en la siguiente figura:

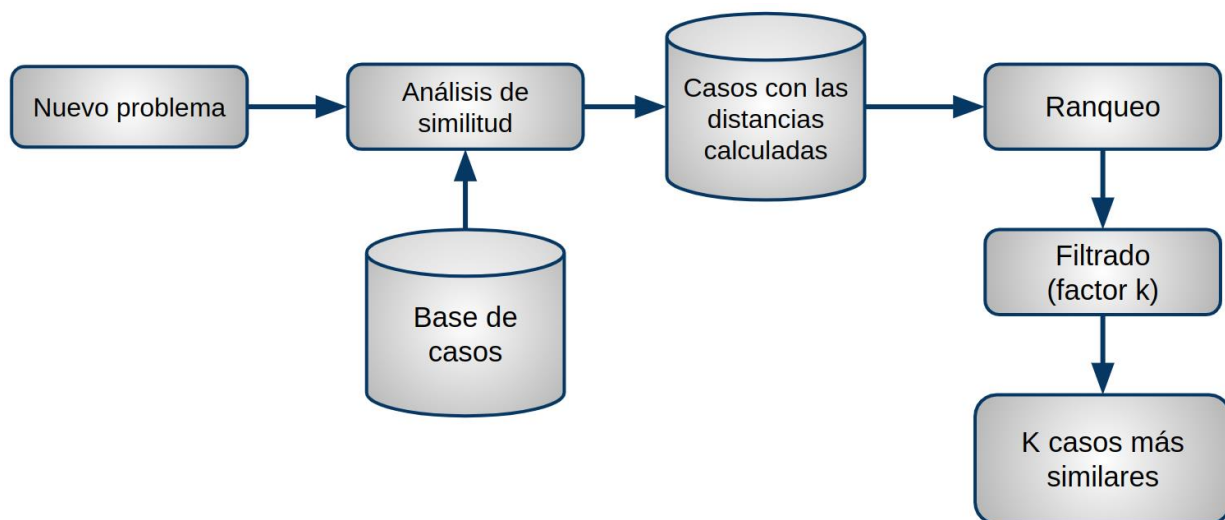


Figura 12. Proceso de recuperación del sistema CBR. Fuente propia.

**Análisis de similitud:** para encontrar los casos más similares (vecinos más cercanos) dada una nueva entrada, se realizó una comparación entre cada uno de los datos de cada caso, haciendo uso de una medida de similitud. Según Gabel [102], las medidas de similitud en un sistema CBR dependen de la composición de las variables del problema, si los valores son discretos, se suele usar la *Distancia Hamming* o el *Coefficiente de Coincidencia Simple (Simple Matching Coefficient o SMC)*; si los datos son simbólicos o cadenas de caracteres, utilizamos la *Medida de similitud de cosenos*; así mismo, si son valores numéricos reales se utilizan las distancias *Manhattan*, *Euclidiana* y *Mahalanobis*. En el presente trabajo se escogió la distancia euclidiana debido a la estructura de los datos del problema (datos numéricos) y la ventaja de ponderación de características que posee esta medida de similitud.

Luego, al calcular una distancia entre cada dato de dos casos, se obtienen similitudes entre cada una de las variables del problema, lo que significa que dos problemas pueden ser similares desde el punto de vista de una variable en particular, pero completamente diferentes si se visualizan desde otra variable. Por esto, se consideró asignar un peso a cada variable para fortalecer la similitud en algunas y en otras no. Esto se hace, partiendo del nivel de incidencia que tienen las variables en el cultivo, no todas afectan en la misma proporción. Por ej. Dos problemas que sólo se diferencian en el sombrío del cultivo no pueden tener la misma similitud a dos problemas que sólo se diferencian en la temporada climática; es decir la temporada incide más para considerar esos dos problemas diferentes. La asignación de pesos se hizo de la siguiente manera:

- Peso 1: Para diferencias consideradas nada relevantes.
- Peso 2: Para diferencias consideradas poco relevantes.
- Peso 3: Para diferencias consideradas relevantes.
- Peso 4: Para diferencias consideradas muy relevantes.
- Peso 5: Para diferencias consideradas fundamentales.

Cabe resaltar que los expertos validaron la importancia de ponderar algunas variables sobre otras. La tabla 11 muestra el proceso de asignación de pesos para todas las variables del problema.

*Tabla 11. Relevancia asignada a cada variable del problema. Fuente propia.*

<b>Diferencia entre variables de:</b>	<b>Peso asignado</b>
Densidad de siembra	2
Cobertura de sombrío	2
Temporada	5
Humedad suelo	5
N suelo	2
P suelo	2
K suelo	2
pH	3

Comprendiendo la razón de la importancia de la ponderación en cada variable del problema y la elección de la medida de similitud, se procede a describir esta medida. La distancia euclidiana es una medida geométrica que mide la distancia entre dos puntos en un plano bidimensional [103, p. 1]. Esta medida tiene la característica de asignar una

ponderación a las variables, denominándose la distancia euclidiana ponderada [104]. La ecuación (2) muestra la fórmula matemática de esta distancia.

$$d_{ep} = \sqrt{\sum_{i=1}^n w_i (x_i - y_i)^2} \quad (2)$$

Donde,  $n$  es el total de datos o muestras a comparar,  $x$  representa los datos del nuevo problema que ingresa,  $y$  representa los datos de cada caso de la base y  $w$  representa el peso de cada dato del problema.

**Casos con las distancias calculadas:** Al finalizar el análisis de similitud entre el nuevo problema y los casos almacenados en la base, se obtuvieron las distancias euclidianas ponderadas para cada caso. Con estas distancias se procede a hacer un ranqueo de casos.

**Ranqueo:** en esta parte se organizaron los casos de la base de acuerdo a los valores de distancia calculados. Donde el primer caso es el que obtuvo la menor distancia calculada y el último caso es el que obtuvo la mayor distancia.

**Filtrado (factor k):** luego de organizar los casos, se procedió a escoger los  $k$  casos más similares. Es decir, se escogieron los  $k$  casos que tuvieron la menor distancia calculada.

**K casos más similares:** finalmente, el resultado del filtrado arrojó los  $k$  casos más similares obtenidos ante el nuevo problema.

## 5.2. Reutilización

Con la identificación de los  $k$  casos más similares, se procedió a revisar las soluciones de esos casos con el fin de determinar o adaptar una nueva solución para el nuevo problema. En el CBR, existen diferentes métodos de encontrar una nueva solución que pueden basarse en reglas, condiciones, fórmulas, indicaciones expertas, restricciones, entre otros [102]. Además, uno de los métodos para determinar una nueva solución es la técnica de regresión, la cual se usó para predecir las tasas de fertilizante de los tres macronutrientes estudiados. Y la técnica de clasificación fue utilizada en las predicciones de las dos recomendaciones extra que realiza el sistema. Lo anterior se explica en la siguiente figura:

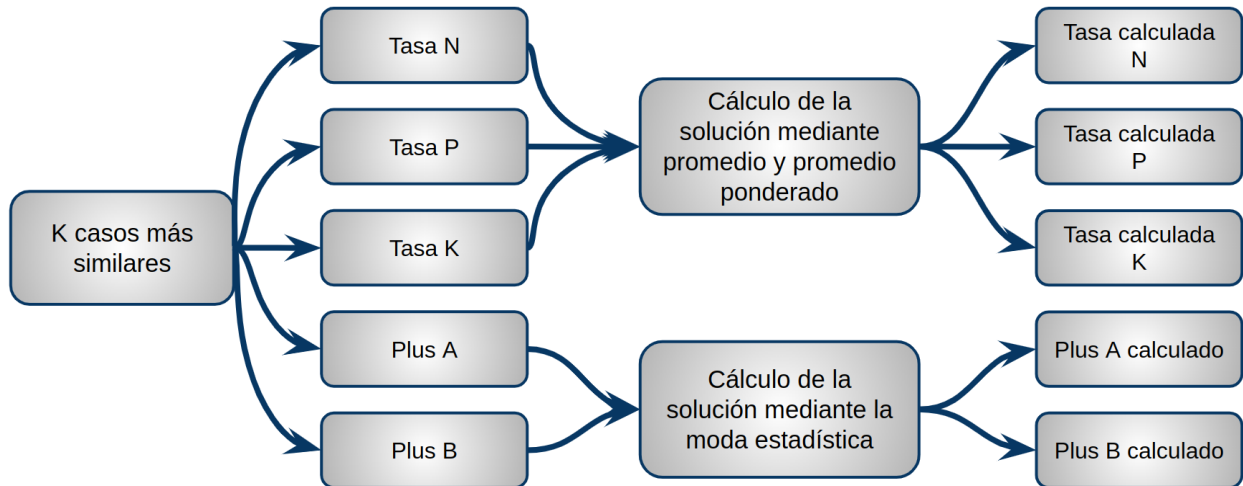


Figura 13. Proceso de reutilización del sistema CBR. Fuente propia.

En la figura se observa que se utilizaron dos métodos de regresión y uno de clasificación para determinar los datos de la nueva solución. El método de clasificación se utilizó para los 2 datos de las recomendaciones adicionales en la solución del caso, que consiste en hallar la moda estadística de estos datos. Para las otras 3 variables, se utilizaron dos métodos de regresión que se describen a continuación:

- **Promedio normal o media**

Este método es el que utiliza el algoritmo KNN comúnmente. Consistió en definir la media de los datos obtenidos en los casos más similares hallados en la sección anterior. Esta media calculada se determina como el dato predicho por el algoritmo.

$$Prom = \frac{\sum_{i=1}^n x_i}{n} \quad (3)$$

Donde  $x_i$  son los datos similares encontrados de cada variable en la solución de un caso y  $n$  es el total de datos similares encontrados. Por lo tanto, se tuvieron 3 predicciones de acuerdo con las 3 tasas de la cantidad de fertilizante que compone la solución de un caso.

- **Promedio ponderado**

El segundo método determina una nueva solución haciendo un promedio ponderado de las  $k$  soluciones encontradas. Como se explicó en la sección 5.1.1, una solución de un caso se compone de 5 variables, 3 indican las tasas de la cantidad de aplicación de fertilizante y las otras 2 son recomendaciones adicionales de prácticas



agrícolas. La ecuación (3), muestra la fórmula del promedio ponderado que se utiliza en cada tasa de fertilizante de las soluciones similares.

$$P_{ponderado} = \frac{\sum_{i=1}^n x_i * w_i}{\sum_{i=1}^n w_i} \quad (4)$$

Donde  $x_i$  son los datos similares encontrados de cada variable en la solución de un caso,  $w_i$  son los pesos asignados en cada dato y  $n$  es el total de datos similares encontrados. Estos pesos son asignados con base en los valores de las distancias calculadas. La tabla 12, muestra estas ponderaciones que se asignan de acuerdo al valor de distancia resultante.

Tabla 12. Peso asignado según el valor de distancia resultante. Fuente propia.

Distancia resultante	Peso asignado
$d < 0.5$	10
$0.5 \leq d < 1$	8
$1 \leq d < 1.5$	6
$1.5 \leq d < 2$	4
$2 \leq d < 2.5$	2
$2.5 \leq d < 3$	1
$d \geq 3$	0

Luego, se calculó el promedio ponderado para cada variable de tasa de fertilizante de la solución de un caso. El paso final es convertir esas variables a las recomendaciones finales que brinda el sistema de CBR.

### 5.3. Recomendaciones

Como se explicó en la sección 4.1.1, las primeras 3 variables de la solución de un caso, indican las tasas (baja, media, alta o muy alta) de la cantidad de fertilizante NPK a aplicar según la condición general del suelo. Las cuales, después de obtener las predicciones se convierten en cantidad de fertilizante expresada en kilogramo por hectárea (Kg/ha) durante un año. Las tasas de fertilizante recomendadas fueron obtenidas a partir de las

revisiones con los expertos y estudios encontrados de Cenicafé [77] [105]. Estos estudios determinaron las cantidades máximas de fertilizante de N, P y K requeridas para un cultivo de café en Colombia, basándose en un caso de densidad muy alta y sombrero bajo. Con esto, en el presente trabajo, se determinaron las cantidades de fertilizante para las condiciones variadas de suelo que pueden darse en un caso.

Así mismo, estas cantidades de fertilizante se escalan de acuerdo a la densidad de siembra del cultivo. Ya que no es lo mismo recomendar una tasa media, por ejemplo, en una densidad de 3000 plantas por ha a hacer la misma recomendación en una densidad de 7000 plantas por ha. Por lo tanto, se tiene en cuenta definir 3 tasas altas para una densidad de siembra baja, media y alta. Así mismo, 3 tasas medias para densidad baja, media y alta Y así, sucesivamente. La tabla 13 muestra las cantidades determinadas para cada nutriente estudiado, las cuales se validaron con los expertos en el dominio.

*Tabla 13. Cantidades de fertilizante a recomendar según la condición de suelo y densidad de siembra. Fuente propia.*

Nutriente	Condición suelo	Tasa de fertilizante	Cantidad de fertilizante (Kg/ha por año)		
			Densidad baja	Densidad media	Densidad alta
Nitrógeno	Crítica	Muy alta	150	180	210
	No favorable	Alta	180	210	240
	Normal	Media	210	240	270
	Favorable	Baja	240	270	300
Fósforo	Crítica	Muy alta	30	36	42
	No favorable	Alta	36	42	48
	Normal	Media	42	48	54
	Favorable	Baja	48	54	60
Potasio	Crítica	Muy alta	135	165	195

	No favorable	Alta	165	195	225
	Normal	Media	195	225	255
	Favorable	Baja	225	255	295

Se debe tener en cuenta que estas cantidades de fertilizante son para un año cafetero entero, lo que significa que para un período de fertilización, sería la mitad de este valor, y dependiendo de la temporada climática, en cada aplicación sólo se usaría la cuarta o sexta parte de estas cantidades de fertilizante.

# CAPÍTULO 6

## EVALUACIÓN

Este capítulo corresponde a la quinta fase de la metodología CRISP-DM, que consta de la evaluación del sistema. Ésta depende de las métricas de evaluación que se usen como la precisión de los resultados, la tasa de error o la correlación entre los datos. También se muestra un ejemplo con un caso real, donde se probó el sistema desarrollado.

En este trabajo de grado, la evaluación del sistema se realizó con las métricas del error medio absoluto (*Mean Absolute Error* o *MAE*) y la raíz del error cuadrático medio (*Root Mean Squared Error* o *RMSE*). Se debe tener en cuenta que estas métricas tienen un resultado diferente para distintos valores del factor  $k$ , por lo que se debe definir cuál es el valor  $k$  más apropiado. El anexo K muestra el código desarrollado en este proceso de evaluación del sistema.

### 6.1. Determinación del factor $k$

Para realizar la evaluación del sistema, se asignaron unos casos como entrada del sistema o también denominados casos de validación (o conjunto de datos de validación) y el resto de los casos quedaron como base, que se denominan conjunto de datos de entrenamiento. Por lo tanto, de la base de casos completa que se construyó en la fase 3, se asignaron el 80% de esos casos para el conjunto de entrenamiento y el 20% para el conjunto de validación, de acuerdo a lo establecido en el estudio de Kumar [106]. La figura 14 muestra la división de la base de casos:

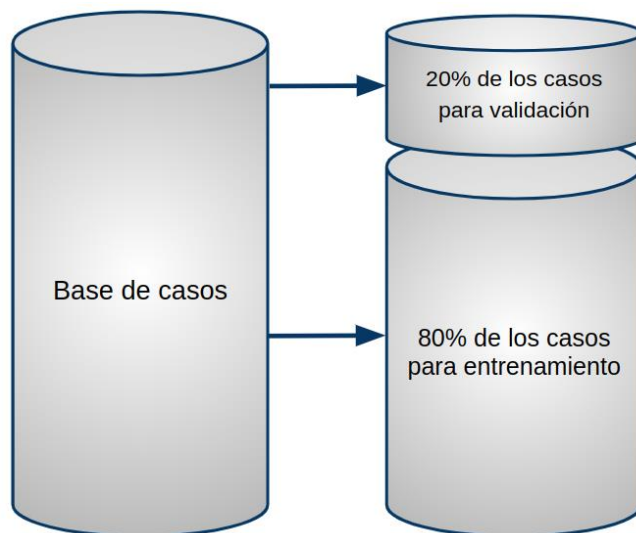


Figura 14. División de la base de casos. Fuente propia.

La figura anterior muestra la proporción de la división que le hizo a la base de casos para realizar la evaluación del sistema. Se tienen 240 casos de entrenamiento y 60 casos para la validación. En el aprendizaje automático, por lo general las variables de entradas se conocen como características (*features*) y existe sólo una variable de salida, la cual se le denomina variable objetivo, que puede ser numérica o categórica dependiendo del modelo de aprendizaje automático que se va a construir (regresión o clasificación) [107]. En el presente trabajo, la salida del modelo está compuesta por 5 variables, en lugar de una sola, por lo que el modelo debe hacer 5 predicciones para encontrar una nueva solución a cada caso entrante.

Para la precisión de las tasas NPK que predice el modelo, se calcularon basándose en el MAE de cada una de esas tasas de fertilizante. Primero se hizo una identificación de los valores máximos y mínimos posibles que puede tomar el MAE. En un caso ideal, el sistema da como resultado un MAE igual a 0, lo cual quiere decir que el modelo ha predicho de forma exacta la tasa base que se indicó al momento de la construcción de la base de casos (tasa base experta), indicando una precisión del 100%. En caso contrario, si el MAE es cercano a 0, es porque el modelo ha predicho un valor cercano a la tasa base experta, por lo que para calcular la precisión (qué tan cercano está el valor predicho del valor experto) se hace necesario conocer el valor máximo posible que pueda tomar el MAE.

Este valor máximo depende de la diferencia máxima posible que puede haber entre el valor predicho y el valor experto, la cual es 3. Es decir, si el valor experto indica una tasa baja (que se representa con un 0) y el modelo predice una tasa muy alta (representada con un 3), se tiene un 3 como la diferencia máxima posible en una predicción (el peor de los casos). Por lo tanto, un MAE de 3 (en términos de tasas) implica una precisión del 0%.

Por otra parte, el cálculo de las precisiones en las dos variables restantes de la solución, se realiza teniendo en cuenta el número de aciertos que tuvo el modelo, dividido sobre el total de datos de validación. Este proceso es el estándar que se realiza en los modelos de clasificación [108]. Como el valor estadístico usado fué la moda, se recomienda usar valores de  $k$  impares para evitar conflictos entre dos valores que se repiten en igual cantidad.

Finalmente, todos los procesos de cálculo de precisiones en los 5 datos de una solución dada, fueron ejecutados para diferentes valores de  $k$  (desde  $k=1$  hasta  $k=29$ ), con el fin de encontrar la mejor precisión del sistema. La figura 15 muestra los resultados obtenidos, teniendo en cuenta los dos métodos de regresión para determinar las tasas de fertilizante a partir del promedio y el promedio ponderado:

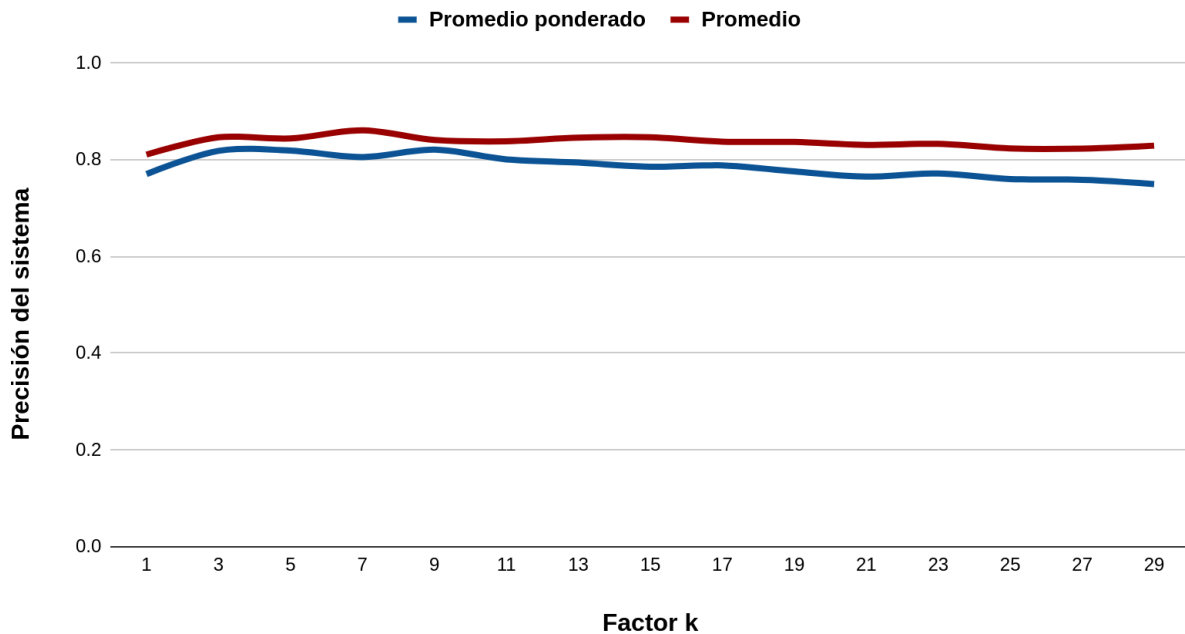


Figura 15. Precisión del sistema ante diferentes valores de k.

La figura muestra que el método de la determinación de la solución usando el promedio normal tuvo una mayor precisión del sistema que usando el promedio ponderado. Lo que definió ese primer método como el más adecuado a usar en la fase de reutilización del CBR. Así mismo, se observa que la precisión del sistema es más alta para los valores de  $k=7$ , en el método que usa el promedio normal. Estableciendo así, que el sistema de recomendación de fertilizantes para cultivos de café construido, tiene una precisión aproximada del 86% para una base de 300 casos y bajo la condición de usar un valor de  $k=7$ .

Además, cabe mencionar que esta precisión depende del tamaño de la base de casos, es decir, el número de casos base iniciales afecta la eficiencia que pueda tener un sistema basado en el CBR [109]. Para entender cómo afecta el tamaño de dicha base, se realizaron nuevamente las pruebas con las que se determinó el factor k, pero esta vez se varió el número de la base de casos y se fijó el factor con  $k=7$ . La figura 16 muestra el proceso descrito.

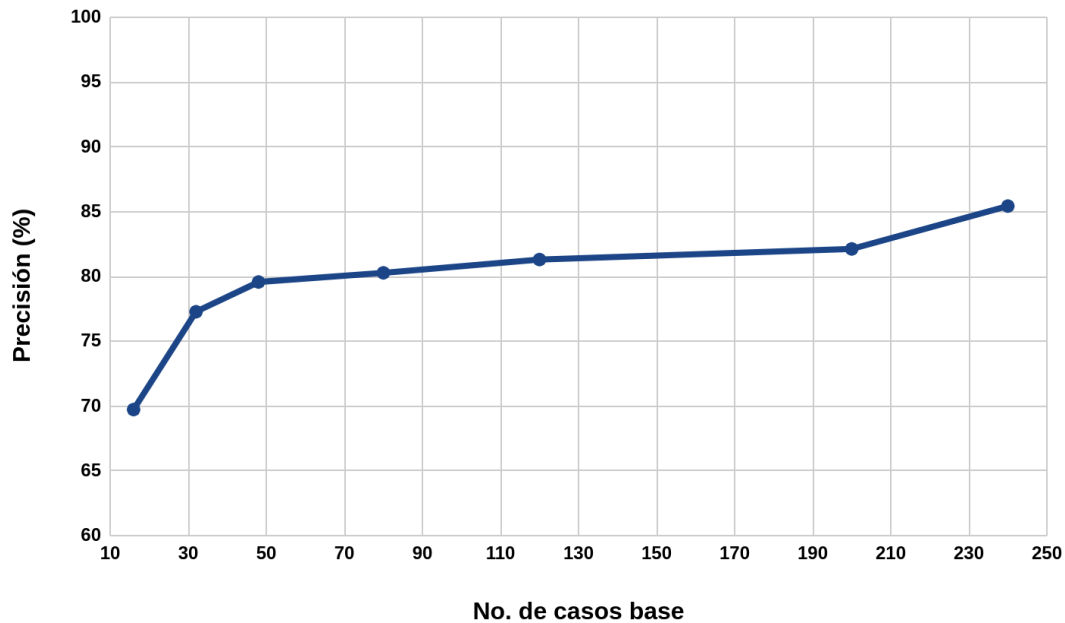


Figura 16. Precisión del sistema ante una variación de la base de casos.

Para cada tamaño de base de casos probado, se tuvo en cuenta el 20% para los casos de validación. Así, la gráfica anterior muestra un comportamiento ascendente de la precisión del sistema a medida que la base de casos aumenta. Esto quiere decir que el CBR funciona mejor cuando tiene mayor disponibilidad de casos para analizar.

## 6.2. MAE y RMSE

Habiendo calculado el factor  $k$ , se procedió a hallar las métricas de MAE y RMSE del sistema para  $k=7$ . Obteniendo los resultados descritos a continuación.

- **Error medio absoluto (MAE por sus siglas en inglés):** Es una métrica estadística que mide el promedio de los errores presentes en un conjunto de valores conocidos y estimados, sin tener en cuenta la dirección de cada uno de ellos [110]. En la ecuación (4) se presenta la fórmula para calcular dicha métrica.

$$MAE = \frac{1}{n} \sum_{i=1}^n |c_i - e_i| \quad (5)$$

Donde  $c_i$  indica el valor conocido en la posición  $i$ ,  $e_i$  el valor estimado y  $n$  como el número de datos estudiados, en este caso, el número de muestras.

Antes de empezar con el cálculo del MAE, se procedió a convertir los datos predichos y los expertos (datos conocidos) que están en términos de tasas de fertilizante (baja, media, alta y muy alta) a los valores de cantidad de fertilizante (en Kg/ha al año) de N, P y K respectivamente. Con esto, el MAE resulta del valor absoluto de la diferencia promedio entre los datos expertos y los datos predichos por el sistema. Finalmente, el MAE respectivo para cada variable de la solución, se muestra en la siguiente tabla:

Tabla 14. MAE de cada tasa de fertilizante recomendado. Fuente propia.

Variable	MAE
Fertilizante de N	16.67 Kg/ha al año
Fertilizante de P	3.14 Kg/ha al año
Fertilizante de K	19.16 Kg/ha al año

- **Error cuadrático medio (RMSE por sus siglas en inglés):** Es una métrica estadística que también se le conoce como raíz de la desviación cuadrática media, la cual permite estimar la magnitud promedio del error entre un valor conocido y un valor estimado y se usa comúnmente en aplicaciones que requieren comparaciones entre dos conjuntos de datos [111]. A continuación, se presenta la fórmula para calcular esta métrica.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (c_i - e_i)^2} \quad (6)$$

Donde  $c_i$  indica el valor conocido en la posición  $i$ ,  $e_i$  el valor estimado y  $n$  como el número de datos estudiados.

Siguiendo el mismo proceso del MAE, se procedió a calcular el RMSE para cada recomendación brindada por el sistema desarrollado. Los resultados se muestran en la tabla 15.



Tabla 15. RMSE de cada tasa de fertilizante recomendado. Fuente propia.

Variable	RMSE
Fertilizante de N	20.54 Kg/ha al año
Fertilizante de P	3.87 Kg/ha al año
Fertilizante de K	23.22 Kg/ha al año

Los resultados de MAE y RMSE obtenidos pasaron por una revisión con los expertos, los cuales determinaron que eran valores aceptables en la tasa de fertilizante de N y P. Sin embargo, recomendaron mejorar la precisión de la tasa de K. Esto puede hacerse creando más casos en la base, para diferentes valores de K, con base en la información experta.

Cabe mencionar que no se tiene otro sistema idéntico para comparar los resultados obtenidos. Por eso se tomaron los 19 trabajos más relacionados al presente trabajo y que fueron descritos en el estado del arte, con el fin de observar si los resultados del presente sistema tienen concordancia con los resultados de los trabajos relacionados. Se identificó que, de los 19 trabajos, sólo 9 realizaron una evaluación del sistema desarrollado, pero sólo 3 utilizaron las métricas de MAE y RMSE. La tabla 16 muestra estos trabajos y los valores obtenidos.

Tabla 16. Trabajos relacionados que evaluaron con MAE y RMSE. Fuente propia.

Trabajo	MAE	RMSE
Puntel, 2019	59 Kg/ha al año	73 Kg/ha al año
Qin, 2018	33.6 Kg/ha al año	-
Ransom, 2019	-	41 Kg/ha al año

Observando que los resultados de nuestro sistema no se desvían considerablemente con los resultados de los trabajos relacionados, es decir, son acordes. Con esto, el desarrollo del sistema y su evaluación se da finalizado, por lo que el siguiente paso es la validación del sistema en un cultivo real.

### 6.3. Prueba del sistema

Finalmente, para probar el sistema se buscó un lugar cercano a la estación climática escogida de los anuarios de Cenicafé (estación de Piendamó). La finca a la que se pudo

tener acceso fue la del caficultor Diego López y se ubica en la vereda de Alto Piendamó. En este lugar se instaló el sensor que se detalla en la sección 4.2.3.2 para la transmisión de datos del suelo. Posteriormente, el caficultor brindó los datos de fecha de floración, densidad de siembra y cobertura de sombrío, los cuales se detallan en la siguiente tabla.

*Tabla 17. Datos de cultivo que brindó el caficultor. Fuente propia.*

<b>Datos de cultivo</b>	<b>Valores dados</b>
Densidad de siembra	5500 plantas/ha aprox.
Cobertura de sombrío	30% aprox.
Última fecha de floración	15 octubre 2021

La tabla anterior, indica que el agricultor sembró cada planta con las medidas de 1m por 1.81 m, lo que da un valor de 1.81 metros cuadrados por planta, sembrando así un número aproximado de 5500 plantas de café por hectárea. Para el sombrío, el agricultor utiliza árboles de guamo sembrados en puntos específicos, cubriendo alrededor del 30% del suelo en cada hectárea de cultivo. Así mismo, el agricultor brinda la fecha de la última floración que se dió en el cultivo, la cual ocurrió el 15 de octubre del 2021.

Posteriormente, se procedió a ingresar esa fecha de floración al sistema, para realizar una clasificación de la temporada climática en esa fecha de floración dada. Los resultados fueron los siguientes:

*Tabla 18. Análisis de datos de PP dada la fecha de última floración. Fuente propia.*

<b>Acumulado PP año 2021</b>	<b>Acumulado PP histórico (2006-2020)</b>
(14 oct. - 15 sept.) = 165.49	(14 oct. - 15 sept.) = 161.1
(14 sept. - 15 ago.) = 79.1	(14 sept. - 15 ago.) = 74.4
(14 ago. - 15 jul.) = 87.99	(14 ago. - 15 jul.) = 65.88
(14 jul. - 15 jun.) = 179.7	(14 jul. - 15 jun.) = 51.09

En la tabla anterior se observa que las precipitaciones para el año 2021 (año donde se dió la última floración del cultivo estudiado) son mayores a los acumulados históricos, superando en un 35% estos valores (promediando los 4 acumulados). Por lo tanto, la temporada se clasificó como lluviosa.

Luego, se construyó un nuevo problema para que ingrese al sistema CBR, el cual se compuso con los datos que se alcanzaron a recolectar del sensor (estuvo en funcionamiento por 15 días), los datos del cultivo y la temporada que se clasificó previamente. La tabla 18 muestra el problema construido.

*Tabla 19. Nuevo problema construido. Fuente propia.*

<b>Variable</b>	<b>Valor</b>
Densidad	5500
Sombrío	30 %
Temporada	Lluviosa
Humedad	50 %
Nitrógeno	40
Fósforo	12
Potasio	196
pH	5.2

La tabla anterior, indica el problema construido a partir de la información recolectada del cultivo, suelo y clima. Las variables de cultivo fueron las indicadas por el agricultor, las variables de clima fue la temporada climática clasificada a partir de los datos meteorológicos de la estación climática de Piendamó y los datos de suelo fueron los valores promedio obtenidos de la lectura de los sensores realizada a lo largo de 7 días.

Posteriormente, los datos del problema fueron normalizados e ingresados al sistema. Donde se inició el proceso de recuperación, encontrando los 7 casos más similares, los cuales se describen en la siguiente figura.

Densidad	Sombrío	Temporada	Humedad	Nitrógeno	Fósforo	Potasio	pH	Distancia
5500.0	0.40	1.0	0.60	60.0	6.0	180.0	5.0	0.318868
6000.0	0.30	1.0	0.40	70.0	12.0	160.0	5.0	0.358940
5400.0	0.35	1.0	0.35	45.0	11.0	185.0	4.8	0.388622
5700.0	0.38	1.0	0.55	45.0	13.0	145.0	4.4	0.415557
5000.0	0.40	1.0	0.50	40.0	21.0	80.0	5.0	0.449862
5400.0	0.35	1.0	0.55	45.0	6.0	175.0	6.2	0.465393
5600.0	0.15	1.0	0.35	65.0	7.0	125.0	5.2	0.469657

Figura 17. Casos más similares encontrados respecto al problema del caso real. Fuente propia.

En la figura se observan los problemas más similares encontrados, junto con las distancias calculadas en cada caso. Entonces, se inició el paso de reutilización para determinar una nueva solución, calculando el promedio normal ante las tasas de fertilizante NPK de los 7 casos de la figura 17, así como la moda de las recomendaciones adicionales. Donde se obtuvieron los siguientes resultados:

Tabla 20. Nueva solución calculada por el sistema para el problema entrante. Fuente propia

Tasa N calculada	Tasa P calculada	Tasa K calculada	Plus A calculado	Plus B calculado
2.71	2.43	1.42	1	1

Indicando una tasa de 2.71 para N, lo que indica un valor entre tasa normal y tasa alta pero más aproximado a tasa alta. La tasa de P también está entre normal y alta pero más aproximada a la tasa normal. Para la tasa de K, es un valor que se encuentra entre tasa baja y tasa normal, siendo más cercano a una tasa baja de K. Para el Plus A y B, que fue 1, indica que no hay recomendaciones de Plus A y B a tener en cuenta.

A los resultados de la tabla anterior se les hizo una conversión posteriormente a las cantidades a aplicar de cada fertilizante, tomando como referencia la tabla 13 y la densidad de siembra (que es media). Los resultados de esta conversión se muestran a continuación:

Tabla 21. Datos de tasa convertidos a cantidad de fertilizante en mg/Kg por año. Fuente propia

Tasa N Kg/ha por año	Tasa P Kg/ha por año	Tasa K Kg/ha por año	Plus A calculado	Plus B calculado
231.3	44.58	177.6	1	1

Se debe tener en cuenta que estos resultados se refieren a la cantidad de fertilizante durante un año cafetero, por lo que toca tomar la mitad de este valor para referirse a la cantidad de fertilizante para un sólo período de fertilización. Luego, con la temporada, se indica que se debe aplicar a los 2 meses después de la fecha de floración (15 de diciembre del 2021). A los 4 meses después de la fecha de floración (15 de febrero del 2022) y a los 6 meses después de la fecha de floración (15 de abril del 2022). En cada una de estas fechas la recomendación fue la siguiente: *Se recomienda aplicar 38.55 Kg/ha de fertilizante de N, 7.43 Kg/ha de fertilizante de P y 29.6 Kg/ha de fertilizante de K. No hay recomendaciones adicionales.*

# CAPÍTULO 7

## CONCLUSIONES Y TRABAJOS FUTUROS

### 7.1. Conclusiones

Los sistemas de recomendación han aportado datos importantes a los agricultores en diversos cultivos a nivel mundial, que les han permitido mejorar la producción, reducir riesgos como enfermedades y plagas e incluso reducir el impacto ambiental relacionado. No obstante, su implementación en entornos reales se ha visto impedido por limitaciones tecnológicas para capturar los datos necesarios para que este tipo de sistemas funcionen adecuadamente. En este aspecto, el presente trabajo de grado propuso la generación de un sistema de recomendación que a diferencia de los trabajos existentes está basado en información experta obtenida a partir de entrevistas a expertos del dominio y de investigaciones científicas de instituciones colombianas relacionadas con el cultivo del café. Lo anterior, permitió que el sistema de recomendación desarrollado no requiera de información histórica referente al suelo ni del historial de producciones del cultivo, sino que solamente tenga en cuenta el estado actual del cultivo e información de la temporada climática en la que se encuentre el café en el momento de la recomendación. De esta manera, el sistema de recomendación realizado logró solventar el problema de escasez de datos necesarios para generar las recomendaciones y demostró en su evaluación que los resultados obtenidos fueron lo suficientemente cercanos a las predicciones realizadas por expertos.

En cuanto a la implementación del sistema de recomendación se determinó que el uso de conocimiento experto en el dominio de la agricultura es fundamental para aquellos cultivos en los cuales se cuenta con un número limitado de datos, en los cuales sea difícil acceder a datos históricos de parámetros importantes para su desarrollo como datos climáticos, de manejo del cultivo, del suelo entre otros. El presente trabajo propuso la implementación del sistema de recomendación a partir de un CBR, los cuales, han demostrado su efectividad en otros dominios de aplicación. La adaptación del CBR al dominio del problema propuesto en este trabajo requirió de la participación y colaboración de dos expertos de la empresa Ecotecma, quienes aportaron a la construcción de la base de casos, la cual fue la piedra angular del sistema desarrollado.

Por último y no menos importante cabe mencionar que en cuanto a la validación del sistema, este solo se puede realizar evaluando el rendimiento del café, pero para lograrlo se requieren muchos años debido a que se necesita la retroalimentación de los

caficultores para cada cosecha indicando el valor del nitrógeno recomendado y el rendimiento obtenido.

## **7.2. Limitaciones**

En el presente trabajo se presentan diversas limitaciones debido a los problemas de la pandemia (en el 2020), ya que por esta circunstancia la llegada de los sensores se retrasó, debido a la crisis logística de contenedores. Por ello la captura de datos fue escasa y se planteó una solución a ello con la propuesta de un RS basado en conocimiento (usando el CBR).

## **7.3. Trabajos futuros**

Teniendo en cuenta las oportunidades de investigación que se abren con el desarrollo de este proyecto de investigación, se proponen los siguientes trabajos futuros.

1. El sistema de recomendación actualmente sólo recomienda la cantidad y la frecuencia de aplicación de los fertilizantes. Un trabajo futuro podría incluir el tipo de fertilizante en las recomendaciones. Ya que en el mercado existen muchos tipos de fertilizantes, que tienen diferentes costos y diferentes composiciones de nutrientes.
2. Para ofrecer a caficultores y expertos del dominio, información más detallada, que vaya más allá de la recomendación de la cantidad de fertilizantes a aplicar, se propone como trabajo futuro el desarrollo de una interfaz gráfica tipo dashboard. Esta interfaz serviría para indicar alertas tempranas relativas a la fertilización del café, el análisis de las temporadas y su relación con la absorción y pérdida de fertilizantes, además de los estados del suelo.
3. A pesar de que el trabajo propuesto demostró que sus recomendaciones eran cercanas a las entregadas por los expertos, aún hace falta realizar más investigación al respecto, ya que para obtener una evaluación completa del sistema, es necesario analizar cómo las recomendaciones de fertilización entregadas afectan la producción del café, para lo cual sería necesario una evaluación durante varios años, en los cuales un caficultor debería registrar, las fechas y tasas de fertilización y el resultado de producción relacionado. Por esta razón, se propone a futuro complementar el sistema desarrollado con módulos que capturen automáticamente estos datos y realimenten el CBR para generar una base más robusta de casos que estén más directamente relacionados con la realidad de un cultivo en específico.

4. Por último y no menos importante vale la pena mencionar que el sistema de recomendación desarrollado en este trabajo de grado, no queda limitado al cultivo del café, sino que se puede aplicar a otros cultivos. Para esto, es necesario realizar cambios en la base de casos, considerando las características específicas de cada cultivo y adaptarlas en la generación de las recomendaciones.



## REFERENCIAS

- [1] «Agricultura, valor agregado (% del PIB) - Colombia», *Datos Grupo Banco Mundial*, 2022. <https://datos.bancomundial.org/indicador/NV.AGR.TOTL.ZS?locations=CO> (accedido 26 de marzo de 2022).
- [2] «Empleos en agricultura (% del total de empleos) - Colombia», 2019. <https://datos.bancomundial.org/indicador/SL.AGR.EMPL.ZS?locations=CO> (accedido 26 de marzo de 2022).
- [3] «Exportaciones agropecuarias de Colombia cerraron con cifras récord en 2021 con ventas por US\$9.418 millones y un crecimiento de 19,9%», *Ministerio de Agricultura*, 2022. [https://www.minagricultura.gov.co/noticias/Paginas/Exportaciones-agropecuarias-de-Colombia-cerraron-con-cifras-r%C3%A9cord-en-2021-con-ventas-por-US\\$9-418-millones-y-un-crecimient.aspx](https://www.minagricultura.gov.co/noticias/Paginas/Exportaciones-agropecuarias-de-Colombia-cerraron-con-cifras-r%C3%A9cord-en-2021-con-ventas-por-US$9-418-millones-y-un-crecimient.aspx) (accedido 26 de marzo de 2022).
- [4] C. Velásquez Agudelo y M. Trávez Velásquez, «Café especial, una alternativa para el sector cafetero en Colombia», 2019.
- [5] (FAO) Organización de las Naciones Unidas para la Agricultura y la Alimentación, «Tendencias y Desafíos», *El Futuro Agric. Aliment.*, p. 44, 2017, doi: 10.1515/nleng-2015-0013.
- [6] (FAO), «La agricultura mundial en la perspectiva del año 2050», *Foro Expert. Alto Niv.*, p. 4, 2009.
- [7] (IFA) Asociación Internacional de la Industria de los Fertilizantes, «Los fertilizantes y su uso», *Clin. Exp. Pharmacol. Physiol.*, vol. 20, n.º 12, pp. 801-808, 1993, doi: 10.1111/j.1440-1681.1993.tb03018.x.
- [8] G. M. Martín Alonso, A. Pérez Díaz, R. R. Espinosa, C. B. González, R. Viñals Núñez, y M. V. Nualles, «Comparison of two models of response to nitrogen doses in corn and coffee», *Cultiv. Trop.*, vol. 37, n.º 2, pp. 155-164, 2016, doi: 10.13140/RG.2.1.2631.8322.
- [9] S. Lenka, V. Coumar, y M. L. Dotaniya, «Impact of Fertilizers use on Environmental Quality», Kalyani, 2016. [En línea]. Disponible en: <https://www.researchgate.net/publication/298388032>
- [10] U. M. Sainju, «Determination of nitrogen balance in agroecosystems», *MethodsX*, vol. 4, pp. 199-208, 2017, doi: 10.1016/j.mex.2017.06.001.
- [11] M. J. Mokarrama y M. S. Arefin, «RSF: A recommendation system for farmers», en *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, Dhaka, dic. 2017, pp. 843-850. doi: 10.1109/R10-HTC.2017.8289086.
- [12] C. Aggarwal, *Recommender Systems - The Textbook*. 2016. doi: 10.1145/245108.245121.
- [13] F. Ricci, L. Rokach, B. Shapira, y P. B. Kantor, Eds., *Recommender Systems Handbook*. Boston, MA: Springer US, 2011. doi: 10.1007/978-0-387-85820-3.

- [14]M. D. Ekstrand, J. T. Riedl, y J. A. Konstan, «Collaborative filtering recommender systems», *Found. Trends Hum.-Comput. Interact.*, vol. 4, n.º 2, pp. 81-173, 2010, doi: 10.1561/1100000009.
- [15]C. C. Aggarwal, *Recommender Systems*. Cham: Springer International Publishing, 2016. doi: 10.1007/978-3-319-29659-3.
- [16]J. Konaté, A. G. Diarra, S. O. Diarra, y A. Diallo, «Syragri: A recommender system for agriculture in mali», *Inf. Switz.*, vol. 11, n.º 12, pp. 1-24, 2020, doi: 10.3390/info11120561.
- [17]P. Lops, M. D. Gemmis, y G. Semeraro, *Content-based Recommender Systems: State of the Art and Trends*. 2011. doi: 10.1007/978-0-387-85820-3.
- [18]D. Jannach, M. Zanker, A. Felfernig, y G. Friedrich, *Recommender Systems: An Introduction*. Cambridge: Cambridge University Press, 2010. doi: 10.1017/CBO9780511763113.
- [19]L. Sumaryanti, T. Istanto, y S. Pare, «Rule Based Method in Expert System for Detection Pests and Diseases of Corn», *J. Phys. Conf. Ser.*, vol. 1569, n.º 2, p. 022023, jul. 2020, doi: 10.1088/1742-6596/1569/2/022023.
- [20]V. J. Kovarik, «Chapter 12 - Cognitive Research: Knowledge Representation and Learning», en *Cognitive Radio Technology (Second Edition)*, B. A. Fette, Ed. Oxford: Academic Press, 2009, pp. 367-399. doi: 10.1016/B978-0-12-374535-4.00012-6.
- [21]M. Richter y R. Weber, *Case-based reasoning: a textbook*. 2013.
- [22]R. Bergmann, K.-D. Althoff, M. Minor, M. Reichle, y K. Bach, «Case-Based Reasoning - Introduction and Recent Developments», *Künstl. Intell.*, vol. 1/2009, pp. 5-11, ene. 2009.
- [23]D. G. Huamán Acuña, «Sistema de recomendación de libros basado en ontologías asociadas a tesauros: el caso de la Biblioteca de la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional Mayor de San Marcos». 2019.
- [24]J. Vera, K. Villalba-Condori, y S. Castro Cuba-Sayco, «Model of a Recommendation System Based on the Context from the Analysis of Static Code for the Development of Computational Thinking: A Web Programming Case», *Educ. Knowl. Soc.*, vol. 19, n.º 2, pp. 103-126, 2018, doi: 10.14201/eks2018192103126.
- [25]M. A. Molina Crespo, «Sistema de recomendación híbrido para la predicción de calificaciones en Yelp.com», PhD Thesis, Universidad Politécnica de Madrid, 2016.
- [26]B. Basnet y J. Bang, «The State-of-the-Art of Knowledge-Intensive Agriculture: A Review on Applied Sensing Systems and Data Analytics», *J. Sens.*, vol. 2018, pp. 1-13, sep. 2018, doi: 10.1155/2018/3528296.
- [27]T. Ojha, S. Misra, y N. S. Raghuwanshi, «Wireless sensor networks for agriculture: The state-of-the-art in practice and future challenges», *Comput. Electron. Agric.*, vol. 118, pp. 66-84, oct. 2015, doi: 10.1016/j.compag.2015.08.011.
- [28]J. Granda-Cantuna, C. Molina-Colcha, S.-E. Hidalgo-Lupera, y C.-D. Valarezo-Varela,

«Design and Implementation of a Wireless Sensor Network for Precision Agriculture Operating in API Mode», en *2018 International Conference on eDemocracy & eGovernment (ICEDEG)*, Ambato, abr. 2018, pp. 144-149. doi: 10.1109/ICEDEG.2018.8372346.

- [29] P. P. Saraswala, S. B. Patel, y J. K. Bhalani, «Performance metric analysis of transmission range in the ZigBee network using various soft computing techniques and the hardware implementation of ZigBee network on ARM-based controller», *Wirel. Netw.*, vol. 27, n.º 3, pp. 2251-2270, abr. 2021, doi: 10.1007/s11276-021-02568-4.
- [30] H. Pirayesh, P. Kheirkhah Sangdeh, y H. Zeng, «Securing ZigBee Communications Against Constant Jamming Attack Using Neural Network», *IEEE Internet Things J.*, vol. 8, n.º 6, pp. 4957-4968, mar. 2021, doi: 10.1109/JIOT.2020.3034128.
- [31] L. B. Del Mundo, R. L. D. Ansay, C. A. M. Festin, y R. M. Ocampo, «A comparison of Wireless Fidelity (Wi-Fi) fingerprinting techniques», en *ICTC 2011*, Seoul, Korea (South), sep. 2011, pp. 20-25. doi: 10.1109/ICTC.2011.6082543.
- [32] «Advantages of WiFi | Disadvantages of WiFi». <https://www.rfwireless-world.com/Terminology/Advantages-and-Disadvantages-of-WiFi.html> (accedido 22 de marzo de 2022).
- [33] M. A. Ertürk, M. A. Aydın, M. T. Büyükakkaşlar, y H. Evirgen, «A Survey on LoRaWAN Architecture, Protocol and Technologies», *Future Internet*, vol. 11, n.º 10, p. 216, oct. 2019, doi: 10.3390/fi11100216.
- [34] H. Caballero, «Estudio en detalle de LoRaWAN. Comparación con otras tecnologías LPWAN considerando diferentes patrones de tráfico», p. 89.
- [35] M. Ballerini, T. Polonelli, D. Brunelli, M. Magno, y L. Benini, «NB-IoT Versus LoRaWAN: An Experimental Evaluation for Industrial Applications», *IEEE Trans. Ind. Inform.*, vol. 16, n.º 12, pp. 7802-7811, dic. 2020, doi: 10.1109/TII.2020.2987423.
- [36] M. Chen, Y. Miao, Y. Hao, y K. Hwang, «Narrow Band Internet of Things», *IEEE Access*, vol. 5, pp. 20557-20577, 2017, doi: 10.1109/ACCESS.2017.2751586.
- [37] A. Lavric, A. I. Petrariu, y V. Popa, «SigFox Communication Protocol: The New Era of IoT?», en *2019 International Conference on Sensing and Instrumentation in IoT Era (ISSI)*, Lisbon, Portugal, ago. 2019, pp. 1-4. doi: 10.1109/ISSI47111.2019.9043727.
- [38] «Technology | Sigfox». [https://www.sigfox.com/en/what-sigfox/technology#id\\_security](https://www.sigfox.com/en/what-sigfox/technology#id_security) (accedido 11 de enero de 2022).
- [39] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, y S. Linkman, «Systematic literature reviews in software engineering – A systematic literature review», *Inf. Softw. Technol.*, vol. 51, n.º 1, pp. 7-15, ene. 2009, doi: 10.1016/j.infsof.2008.09.009.
- [40] A. Chougule, V. K. Jha, y D. Mukhopadhyay, «Crop Suitability and Fertilizers Recommendation Using Data Mining Techniques», 2019. doi: 10.1007/978-981-13-0224-4\_19.
- [41] S. M. Vieira Fontoura, F. Viero, R. P. de Moraes, y C. Bayer, «Nitrogen Fertilization of No-

Tillage Winter Cereals in the South-Central Region of Parana, Brazil», *Rev. Bras. Cienc. SOLO*, vol. 41, 2017, doi: 10.1590/18069657rbcscs20170009.

- [42] A. Kumar, A. Kumar, A. De, S. Shekhar, y R. K. Singh, «IoT based farming recommendation system using soil nutrient and environmental condition detection», *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, n.º 11, pp. 3055-3060, 2019, doi: 10.35940/ijitee.K2335.0981119.
- [43] C. P. Wickramasinghe, P. L. N. Lakshitha, H. P. H. S. Hemapriya, A. Jayakody, y P. G. N. S. Ranasinghe, «Smart Crop and Fertilizer Prediction System», en *2019 International Conference on Advancements in Computing (ICAC)*, Malabe, Sri Lanka, dic. 2019, pp. 487-492. doi: 10.1109/ICAC49085.2019.9103422.
- [44] M. S. Suchithra y M. L. Pai, «Improving the Performance of Sigmoid Kernels in Multiclass SVM Using Optimization Techniques for Agricultural Fertilizer Recommendation System», en *2nd Int. Conf. on Soft Computing Systems, ICSCS 2018*, vol. 837, 2018, pp. 857-868. doi: 10.1007/978-981-13-1936-5\_87.
- [45] T. Sujithra, M. Thanjaivadivel, y S. Durai, «Fertilizer recommendation system for coconut cultivation», *Int. J. Civ. Eng. Technol.*, vol. 8, n.º 9, pp. 813-819, 2017.
- [46] H. Zhang *et al.*, «Design and implementation of crop recommendation fertilization decision system based on WEBGIS at village scale», 2011. doi: 10.1007/978-3-642-18336-2\_44.
- [47] Z. Ren y X. Lu, «Design of fertilization recommendation knowledge base and application», *2012 1st Int Conf Agro-Geoinformatics Agro-Geoinformatics 2012*, pp. 203-207, 2012, doi: 10.1109/Agro-Geoinformatics.2012.6311640.
- [48] M. A. Hossain y M. N. A. Siddique, «Online Fertilizer Recommendation System (OFRS): A Step Towards Precision Agriculture And Optimized Fertilizer Usage By Smallholder Farmers In Bangladesh», *Eur. J. Environ. Earth Sci.*, vol. 1, n.º 4, ago. 2020, doi: 10.24018/ejgeo.2020.1.4.47.
- [49] I. Cholissodin, C. Dewi, y E. E. Surbakti, «Integrated ANN and Bidirectional Improved PSO for optimization of fertilizer dose on Palawija plants», en *2016 2nd Int. Conf. on Science in Information Technology (ICSITech)*, oct. 2016, pp. 193-197. doi: 10.1109/ICSITech.2016.7852632.
- [50] T. Islam, R. U. B. Rizan, Y. A. Tusher, M. Shafiuzzaman, Md. A. Hossain, y S. Galib, «Nitrogen Fertilizer Recommendation for Paddies through Automating the Leaf Color Chart (LCC)», *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, n.º 8, 2020, doi: 10.14569/IJACSA.2020.0110891.
- [51] C. Witt, J. M. C. A. Pasuquim, R. Mutters, y R. J. Buresh, «New leaf color chart for effective nitrogen management in rice», *Better Crops*, 2005.
- [52] M. S. Abd Razak, S. Abdul-Rahman, S. Mutalib, y Z. Abd Aziz, «Nitrogen Fertilizer Recommender for Paddy Fields», en *3rd Int. Conf. on Soft Computing in Data Science, SCDS 2017*, vol. 788, 2017, pp. 230-240. doi: 10.1007/978-981-10-7242-0\_20.
- [53] L. Sumaryanti, L. Lamalewa, y T. Istanto, «Implementation of Fuzzy Multiple Criteria Decision Making for Recommendation Paddy Fertilizer», *Int. J. Mech. Eng. Technol. IJMET*,

2019.

- [54] R. Garg, H. Aggarwal, P. Centobelli, y R. Cerchione, «Extracting Knowledge from Big Data for Sustainability: A Comparison of Machine Learning Techniques», *Sustainability*, vol. 11, n.º 23, p. 6669, nov. 2019, doi: 10.3390/su11236669.
- [55] Z. Qin *et al.*, «Application of Machine Learning Methodologies for Predicting Corn Economic Optimal Nitrogen Rate», *Agron. J.*, vol. 110, n.º 6, pp. 2596-2607, nov. 2018, doi: 10.2134/agronj2018.03.0222.
- [56] W. Jiang, Y. Xing, X. Wang, X. Liu, y Z. Cui, «Developing a Sustainable Management Strategy for Quantitative Estimation of Optimum Nitrogen Fertilizer Recommendation Rates for Maize in Northeast China», *Sustainability*, vol. 12, n.º 7, p. 2607, mar. 2020, doi: 10.3390/su12072607.
- [57] L. A. Puntel, A. Pagani, y S. V. Archontoulis, «Development of a nitrogen recommendation tool for corn considering static and dynamic variables», *Eur. J. Agron.*, vol. 105, pp. 189-199, abr. 2019, doi: 10.1016/j.eja.2019.01.003.
- [58] B. R. McFadden, B. W. Brorsen, y W. R. Raun, «Nitrogen fertilizer recommendations based on plant sensing and Bayesian updating», *Precis. Agric.*, vol. 19, n.º 1, pp. 79-92, 2018, doi: 10.1007/s11119-017-9499-4.
- [59] C. J. Ransom *et al.*, «Statistical and machine learning methods evaluated for incorporating soil and weather into corn nitrogen recommendations», *Comput. Electron. Agric.*, vol. 164, p. 104872, sep. 2019, doi: 10.1016/j.compag.2019.104872.
- [60] P. Chapman *et al.*, «Step-by-step data mining guide», p. 76, 2000.
- [61] J. Arcila P., F. Farfán V., A. Moreno B., L. F. Salazar G., y E. Hincapié G., *Sistemas de producción de café en Colombia*, 1.ª ed. Editorial Blanecolor Ltda., 2007.
- [62] J. S. Pozas, «Fertilidad del suelo y nutrición del café en Colombia - Guía práctica», vol. N° 32, p. 45, 2008.
- [63] «Micronutrientes (nutrición vegetal) | Crop Care». <https://www.crodacropcare.com/es-mx/market-areas/micronutrients> (accedido 27 de marzo de 2022).
- [64] S. Sadeghian Khalajabadi, «Síntomas visuales de deficiencias nutricionales en café», *Cenicafé*, vol. 878, p. 12, 2017.
- [65] S. S. Khalajabadi, H. G. Osorio, y E. A. Suárez, «Prácticas que ayudan a reducirla», p. 36.
- [66] A. Valdivielso, «¿Qué es la escorrentía?», *iAgua*, 2 de septiembre de 2020. <https://www.iagua.es/respuestas/que-es-escorrentia> (accedido 27 de marzo de 2022).
- [67] É. H. Gómez y F. A. R. Ortiz, «RIESGO A LA EROSIÓN EN SUELOS DE LADERA DE LA ZONA CAFETERA», p. 8.
- [68] Daniel Montero Blanco, «Acciones en el sector agropecuario», Costa Rica, p. 16, 2017.

- [69] «Nutrición de cafetales y manejo de la acidez del suelo | [www.cenicafe.org](http://www.cenicafe.org)». [https://www.cenicafe.org/es/index.php/cultivemos\\_cafe/nutricion\\_de\\_cafetales](https://www.cenicafe.org/es/index.php/cultivemos_cafe/nutricion_de_cafetales) (accedido 27 de marzo de 2022).
- [70] F. Gast H. *et al.*, *Manual del cafetero colombiano: Investigación y tecnología para la sostenibilidad de la caficultura*, vol. 1, 3 vols. LEGIS, 2013.
- [71] V. H. Ramirez Builes, «La fenología del café, una herramienta para apoyar la toma de decisiones», *Cenicafé*, Manizales, Caldas, Informe técnico 441, 2014.
- [72] Federación nacional de cafeteros de Colombia, «Recolección de café», en *Cartilla cafetera*, *Cenicafé*, 2004, p. 22.
- [73] N. O. and A. A. US Department of Commerce, «What are El Nino and La Nina?», *National Ocean Service*, 2021. <https://oceanservice.noaa.gov/facts/ninonina.html> (accedido 24 de marzo de 2022).
- [74] Federación nacional de cafeteros de Colombia, «Evento de la niña en Colombia: Recomendaciones para la caficultura», *Cenicafé*, Informe técnico 467, 2016.
- [75] Federación nacional de cafeteros de Colombia, «Épocas recomendables para la fertilización de cafetales», *Cenicafé*, Informe técnico 442, 2014.
- [76] G. Catalán Salas, «El pH del suelo en la agricultura», *Agropal*, 8 de noviembre de 2016. <http://www.agropal.com/es/el-ph-del-suelo/> (accedido 25 de marzo de 2022).
- [77] Federación nacional de cafeteros de Colombia, *Manual del cafetero colombiano: Investigación y tecnología para la sostenibilidad de la caficultura*, vol. 2, 3 vols. LEGIS, 2013.
- [78] G. Valencia A., I. F. Carrillo P., y L. I. Estrada H., «Fertilización del cafetal según el análisis de suelos», *Cenicafé*, Informe técnico, 1990.
- [79] S. Sadeghian, *La materia orgánica: componente esencial en la sostenibilidad de los agroecosistemas cafeteros*. *Cenicafé*, 2010. doi: 10.38141/cenbook-0018.
- [80] «Conductividad Eléctrica del Suelo y su Influencia en la Agricultura», *Maher Electrónica*, 22 de julio de 2020. <https://www.maherelectronica.com/conductividad-electrica-agricultura/> (accedido 26 de marzo de 2022).
- [81] J. F. Casanova O., «Sensores inalámbricos para la captura de datos de suelo en la agricultura», *Ecotecma SAS*, Informe técnico 1, 2020.
- [82] LoRaWan, «LoRaWAN: Descubre la mayor revolución para el Internet de las Cosas (IoT)». [En línea]. Disponible en: <https://lorawan.es/>
- [83] Campo Juan Camilo-Ecotecma, «Sistema de monitorización de parámetros de suelo para un sistema productivo de café en el Departamento del Cauca».
- [84] CENICAFE, *Anuario meteorológico cafetero*. 2019. Accedido: 26 de marzo de 2022. [En línea]. Disponible en: <https://biblioteca.cenicafe.org/handle/10778/660>

- [85] S. Sadeghian-Khalajabadí, «CALIBRACIÓN DE ANÁLISIS DE SUELO PARA N, P, K Y Mg EN CAFETALES AL SOL Y BAJO SEMISOMBRA», *Cenicafé*, p. 18, 2009.
- [86] Á. G. J.-T. L. Fertibox, «El fósforo y su importancia en el crecimiento vegetal», *fertibox*, 10 de diciembre de 2019. <https://www.fertibox.net/single-post/fosforo-agricultura> (accedido 28 de marzo de 2022).
- [87] «POTASIO», *DF Grupo*. <https://www.dfgrupo.com/convencionales/potasio/> (accedido 28 de marzo de 2022).
- [88] Siavosh Sadeghian Khalajabadi-Centro Nacional de Investigaciones de café(CenicaféCeni), «La acidez del suelo, una limitante común para la producción de café.», *Abril 2016*, vol. 466, p. 12.
- [89] J. Arcila P., «Densidad de siembra y productividad de los cafetales», en *Sistemas de producción de café en Colombia*.
- [90] F. Farfán-Valencia y A. Mestre-Mestre, «MANEJO DEL SOMBRÍO Y FERTILIZACIÓN DEL CAFÉ EN LA ZONA CENTRAL COLOMBIANA», p. 8.
- [91] V. H. Ramírez Builes *et al.*, «Recomendaciones para la reducción del riesgo en la caficultura de Colombia ante un evento climático de El Niño», Caldas, Colombia, Informe técnico 445, 2014.
- [92] J. Arcila P., «Factores que determinan la productividad del cafetal», en *Sistemas de producción de café en Colombia*, 1.<sup>a</sup> ed., Editorial Blanecolor Ltda., 2007, p. 26.
- [93] J. A. Mula, «Cómo cambiar el pH del suelo», *Agromática*, 2 de febrero de 2013. <https://www.agromatica.es/como-cambiar-el-ph-del-suelo/> (accedido 26 de marzo de 2022).
- [94] M. Sánchez-Marrè, «PRINCIPLES OF CASE-BASED REASONING», p. 13, 2001.
- [95] Y. Shang, «5 - Expert Systems», en *The Electrical Engineering Handbook*, W.-K. Chen, Ed. Burlington: Academic Press, 2005, pp. 367-377. doi: 10.1016/B978-012170960-0/50031-1.
- [96] N. C. P. Center, «NOAA's Climate Prediction Center». [https://origin.cpc.ncep.noaa.gov/products/analysis\\_monitoring/ensostuff/ONI\\_v5.php](https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php) (accedido 26 de marzo de 2022).
- [97] S. Loukas, «Everything you need to know about Min-Max normalization in Python», *Medium*, 8 de octubre de 2021. <https://towardsdatascience.com/everything-you-need-to-know-about-min-max-normalization-in-python-b79592732b79> (accedido 21 de marzo de 2022).
- [98] S. Pandey, «Normalization of Data in Python». <https://linuxhint.com/normalization-of-data-in-python/> (accedido 21 de marzo de 2022).
- [99] J. L. Kolodner, «An introduction to case-based reasoning», p. 32, 1992.
- [100] I. Watson y F. Marir, «Case-Based Reasoning: A Review», presentado en AI-CBR, Dept.

of Computer Science, University of Aukland, 2000.

- [101] M. Miller, «The Basics: KNN for classification and regression», *Medium*, 18 de octubre de 2019. <https://towardsdatascience.com/the-basics-knn-for-classification-and-regression-c1e8a6c955> (accedido 24 de marzo de 2022).
- [102] T. Gabel, «Problem Solving by Case-Based Reasoning», p. 10, 2010.
- [103] B. O'Neill, «Chapter 1 - Calculus on Euclidean Space», en *Elementary Differential Geometry (Second Edition)*, B. O'Neill, Ed. Boston: Academic Press, 2006, pp. 3-42. doi: 10.1016/B978-0-12-088735-4.50005-5.
- [104] I. H. Witten, E. Frank, M. A. Hall, y C. J. Pal, «Chapter 7 - Extending instance-based and linear models», en *Data Mining (Fourth Edition)*, I. H. Witten, E. Frank, M. A. Hall, y C. J. Pal, Eds. Morgan Kaufmann, 2017, pp. 243-284. doi: 10.1016/B978-0-12-804291-5.00007-6.
- [105] S. Sadeghian Khalajabadi, «Fertilidad del suelo y nutrición del café en Colombia», *Programa Investig. Científica - Cenicafé*, n.º 32, p. 45, 2008.
- [106] A. Kumar, «Machine Learning - Training, Validation & Test Data Set», *Data Analytics*, 13 de junio de 2021. <https://vitalflux.com/machine-learning-training-validation-test-data-set/> (accedido 24 de marzo de 2022).
- [107] J. Rocca, «A simple introduction to Machine Learning», *Medium*, 19 de marzo de 2020. <https://towardsdatascience.com/introduction-to-machine-learning-f41aabc55264> (accedido 25 de marzo de 2022).
- [108] J. Martínez Heras, «Precision, Recall, F1, Accuracy en clasificación - IArtificial.net», 17 de noviembre de 2019. <https://www.iartificial.net/precision-recall-f1-accuracy-en-clasificacion/> (accedido 25 de marzo de 2022).
- [109] S. El-Sappagh y M. Elmogy, «Case Based Reasoning: Case Representation Methodologies», *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, pp. 192-208, nov. 2015, doi: 10.14569/IJACSA.2015.061126.
- [110] L. Kouadio, R. C. Deo, V. Byrareddy, J. F. Adamowski, S. Mushtaq, y V. Phuong Nguyen, «Artificial intelligence approach for the prediction of Robusta coffee yield using soil fertility properties», *Comput. Electron. Agric.*, vol. 155, pp. 324-338, dic. 2018, doi: 10.1016/j.compag.2018.10.014.
- [111] E. G. L. Sambony, «SISTEMA EXPERTO BASADO EN EMPAREJAMIENTO DE PATRONES», p. 93.



# Anexo A

Este anexo presenta el proceso realizado en el software de Parsifal, una herramienta de apoyo a revisiones sistemáticas. Las figuras 1 y 2 del presente anexo, muestra la sección de planeación de la RSL, donde se especifican los objetivos de la investigación, la fuentes de búsqueda, la cadena de búsqueda, los criterios de inclusión y exclusión, y las preguntas de calidad.

Parsifal Blog About Help leoc8697

Cristhian Figueroa / SLR Recommender Systems in Agriculture

Review **Planning** Conducting Reporting

Protocol **Quality Assessment Checklist** Data Extraction Form

Protocol

Objectives

PICOC

Research Questions

Keywords and Synonyms

Search String

Sources

Selection Criteria

Objectives

To study main recomender systems in agriculture and find technological gaps that limit their implementation in Colombia

Save

PICOC

Separate the terms used in the PICOC using commas. This will make possible to save them separately as keywords so we can help you design your search string.  
If any of the sections of PICOC doesn't apply to your research, please leave it blank.

Figura 1. Sección de planeación en el Parsifal.

Protocol

Objectives

PICOC

Research Questions

Keywords and Synonyms

Search String

Sources

Selection Criteria

Search String

Use uppercase for boolean operators (AND, OR), double quotes for composite words and parentheses to logically separate the keywords and synonyms.

((("recommender system" OR "recommendation system" OR "prediction system" OR "expert system" OR "decision support system") AND (Farm\* OR Agri\* OR Nutri\* OR fertili\*))

Save Suggested Search String

Sources

Name	URL		
Mendeley		edit	remove
Scopus	<a href="http://www.scopus.com">http://www.scopus.com</a>	edit	remove
WoS	<a href="http://webofknowledge.com.acceso.unicauca.edu.co">http://webofknowledge.com.acceso.unicauca.edu.co</a>	edit	remove

+ Add Source Add a Digital Library

Figura 3. Cadena de búsqueda y fuentes seleccionadas.

Como se explica en la descripción de la RSL en el capítulo 2, el primer filtrado se realizó con base en los criterios de inclusión y exclusión. Este proceso consistió de forma manual, clasificar cada trabajo encontrado en un criterio, este fue el primer filtrado realizado. Luego, para el segundo filtrado se califica cada paper de acuerdo a las preguntas de calidad, este proceso se realizó en la sección de “*Quality Assessment Checklist*” en Parsifal. La figura 3 muestra dicha sección, donde se agregan todas las preguntas de calidad con las que se va a calificar cada trabajo de la búsqueda. La figura 4 muestra los puntos de calificación que cada trabajo recibirá en cada pregunta, aceptando los trabajos que obtengan una calificación acumulada de al menos 10 puntos.

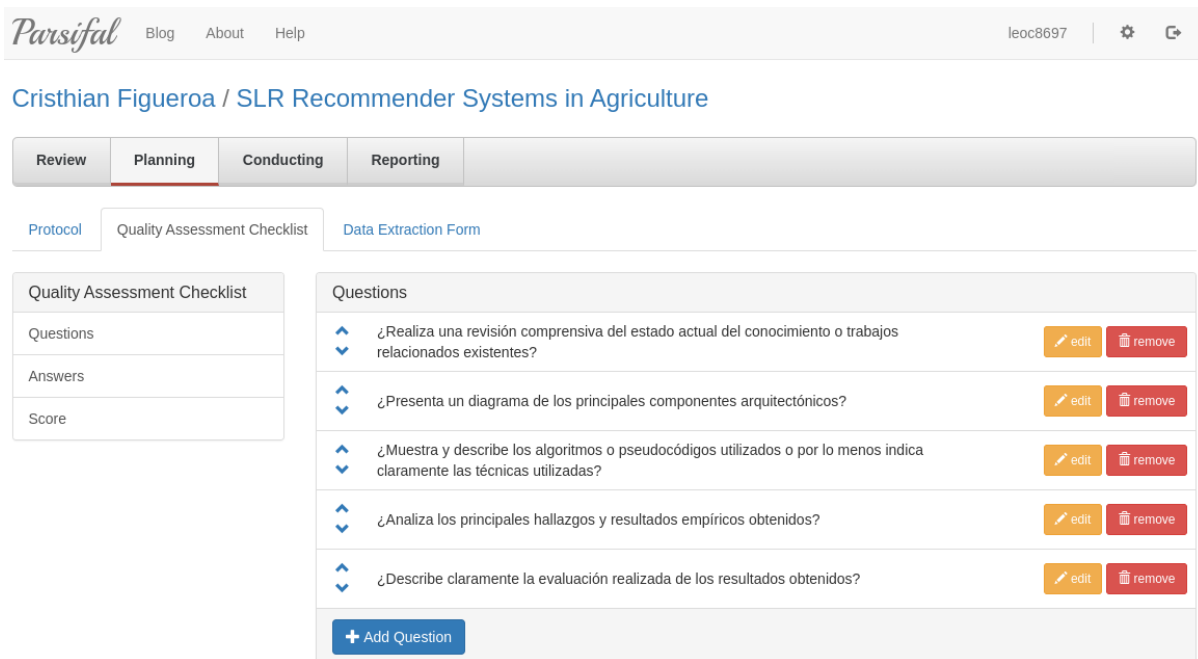


Figura 3. Preguntas de calidad definidas en el Parsifal.

Answers		
Description	Weight	
Sí	5.0	edit  remove
Parcialmente	3.0	edit  remove
No	0.0	edit  remove
<a href="#">+ Add Answer</a>		

Quality Assessment Scores		
<b>Max Score</b>	<input type="text" value="25.0"/>	Calculated based on the number of questions and on the answer of greater weight
<b>Cutoff Score</b>	<input type="text" value="9.0"/>	save

Figura 4. Calificación establecida en el Parsifal.

Luego, en la sección de Conducción de Parsifal, se realizó en importe de los trabajos y se clasificaron en estado “Aceptado”, “Duplicado” o “Rechazado” según los criterios de inclusión y exclusión. La figura 5 muestra la vista de esta sección.

Parsifal [Blog](#) [About](#) [Help](#) leoc8697

[Cristhian Figueroa / SLR Recommender Systems in Agriculture](#)

**Review** **Planning** **Conducting** **Reporting**

1. Search 2. Import Studies 3. Study Selection 4. Quality Assessment 5. Data Extraction 6. Data Analysis

### Study Selection

All Sources [Mendeley](#) [Scopus](#) [WoS](#)

Find Duplicates Export Articles

Action:   0 of 207 selected

Show:  All  Accepted  Rejected  Unclassified  Duplicated

<input type="checkbox"/>	Bibtex Key	Title	Author	Journal	Year	Added by	Added at	Status
<input type="checkbox"/>	ISI:000315607200036	A new Expert System for greenness identification in agricultural images	Romeo, J and Pajares, G and Montalvo, M and Guerrero, J M and Guijarro, M and de la Cruz, J M	EXPERT SYSTEMS WITH APPLICATIONS	2013	leoc8697	14 May 2021 20:15:55	
<input type="checkbox"/>	ISI:000337111700006	A Machine Vision-Based Maturity Prediction System for Sorting of Harvested Mangoes	Nandi, Chandra Sekhar and Tudu, Bipan and Koley, Chiranjib	IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT	2014	leoc8697	14 May 2021 20:16:09	
<input type="checkbox"/>	Yanuar201445	Decision support system based	Yanuar, H and Basro, W and	International Journal of	2014	leoc8697	14 May 2021	

Figura 5. Selección de trabajos según los criterios establecidos.

Finalmente, la figura 6 muestra la sección donde se califica cada trabajo de acuerdo a las preguntas de calidad.

1. Search   2. Import Studies   3. Study Selection   4. Quality Assessment   5. Data Extraction   6. Data Analysis

### Quality Assessment

**Detailed**   Summary

Show:  All    Done    Pending    Score higher than 9.0    Score lower or equal to 9.0   Order by: Title (a - z) ▼

To answer the form you may click on the desired answer on the following tables.

A Fuzzy Logic-Based Crop Recommendation System (2021) <span style="float: right;">16.0</span>			
¿Realiza una revisión comprensiva del estado actual del conocimiento o trabajos relacionados existentes?	Sí	Parcialmente	No
¿Presenta un diagrama de los principales componentes arquitectónicos?	Sí	Parcialmente	No
¿Muestra y describe los algoritmos o pseudocódigos utilizados o por lo menos indica claramente las técnicas utilizadas?	Sí	Parcialmente	No
¿Analiza los principales hallazgos y resultados empíricos obtenidos?	Sí	Parcialmente	No
¿Describe claramente la evaluación realizada de los resultados obtenidos?	Sí	Parcialmente	No

A Machine Vision-Based Maturity Prediction System for Sorting of Harvested Mangoes (2014) <span style="float: right;">25.0</span>			
¿Realiza una revisión comprensiva del estado actual del conocimiento o trabajos relacionados existentes?	Sí	Parcialmente	No
¿Presenta un diagrama de los principales componentes arquitectónicos?	Sí	Parcialmente	No
¿Muestra y describe los algoritmos o pseudocódigos utilizados o por lo menos indica claramente las técnicas utilizadas?	Sí	Parcialmente	No
¿Analiza los principales hallazgos y resultados empíricos obtenidos?	Sí	Parcialmente	No
¿Describe claramente la evaluación realizada de los resultados obtenidos?	Sí	Parcialmente	No

Figura 6. Calificación de los trabajos resultantes del primer filtrado.

Como se mencionó en el capítulo 2, resultaron 105 trabajos que tuvieron una calificación de 10 puntos o más. A los cuales, se les realizó el siguiente paso, que consistía en la extracción de datos.

## Anexo B

Este anexo presenta una hoja de cálculo con todos los trabajos relacionados encontrados en la RSL y que tuvieron una calificación mínima de 10 puntos en el Parsifal. En total fueron 102 artículos a los que se les extrajeron datos como, el año de publicación, el país de estudio, la técnica o algoritmo utilizado, el tipo de sistema desarrollado, el cultivo estudiado, las fuentes de recolección de datos, las variables de entrada del sistema, las variables de salida del sistema, entre otras cosas.

El anexo está disponible en el siguiente enlace:

<https://docs.google.com/spreadsheets/d/1g7v8ix77MPI97pX1qaxZvpeOjnpAfQ6UIRUhYKawxxA/edit?usp=sharing>

## Anexo C

Este anexo presenta los resultados obtenidos de la extracción de datos realizada. La Figura. 1 muestra los años de publicación de los trabajos.

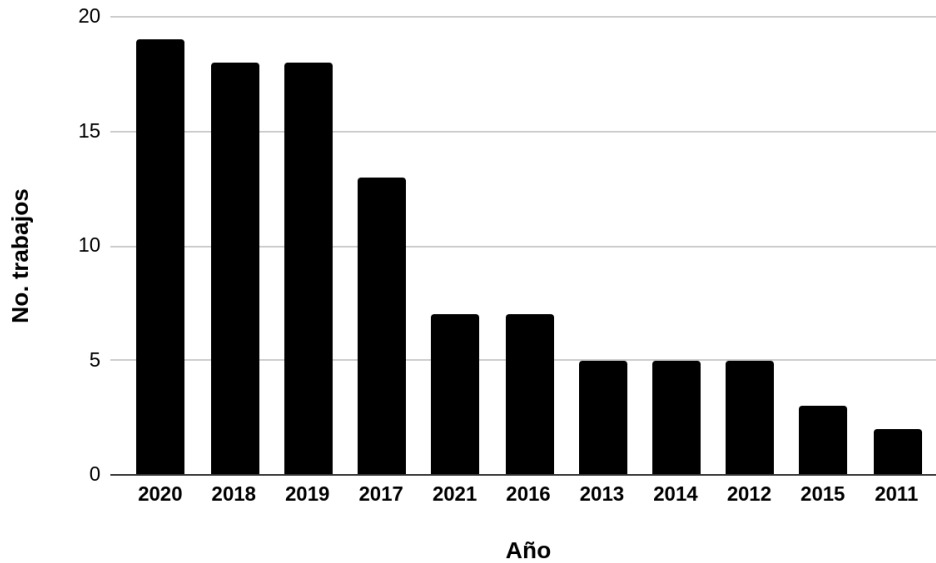


Figura 1. Extracción de los años de los trabajos relacionados.

La figura muestra que la mayoría de los trabajos de sistemas de apoyo a las decisiones que se encontraron en la RSL, han sido elaborados en los últimos 5 años. Lo que indica un aumento en la investigación agrícola, con el fin de mejorar las prácticas de los agricultores. La figura 2, muestra los resultados de los países de estudio.

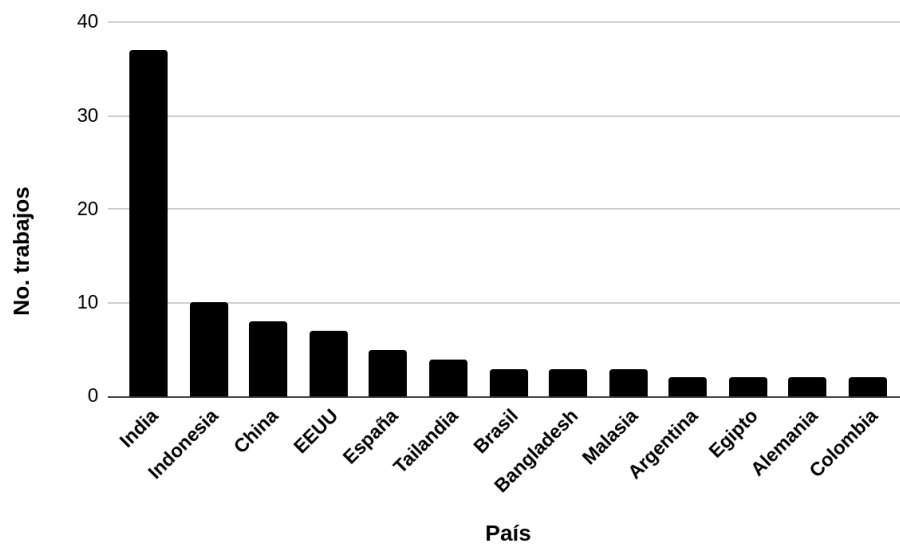


Figura 2. Extracción de los países de estudio de los trabajos relacionados.

Esta figura muestra que los estudios de este tipo de sistemas (expertos, recomendación, predicción, etc) en la agricultura son frecuentes en las potencias agrícolas que existen, como India, China y USA. También se observa que son muy pocos los estudios en la región suramericana, lo que indica que es un campo poco explorado en la agricultura de la región. La figura 3 muestra los algoritmos más comunes utilizados.

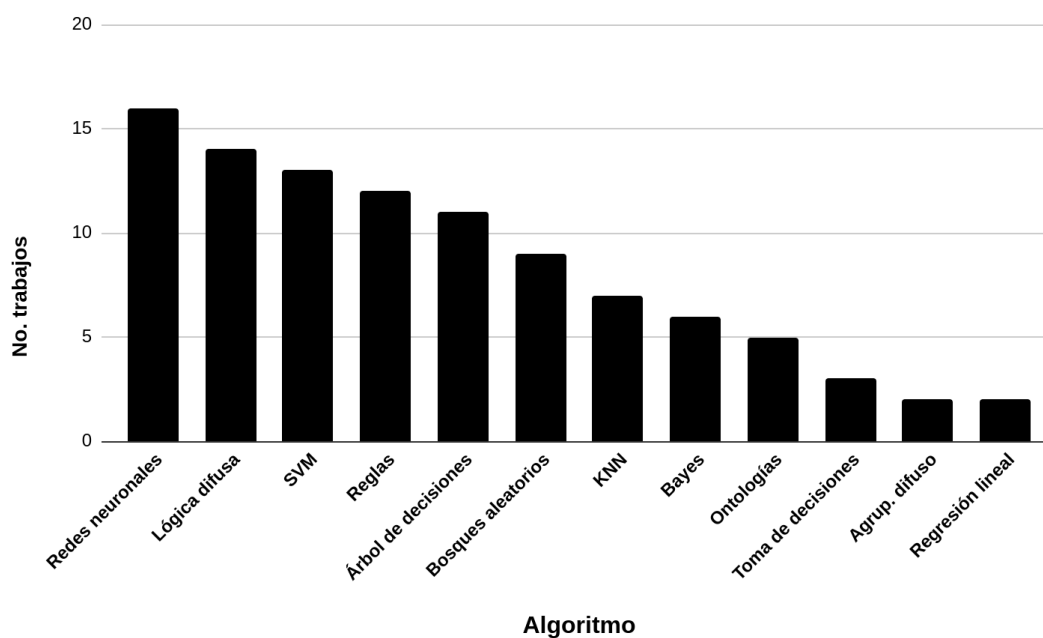


Figura 3. Extracción de los algoritmos utilizados en los trabajos relacionados.

La figura 3 muestra que los principales algoritmos utilizados son los algoritmos de aprendizaje automático supervisado más comunes que existen. Incluyendo el algoritmo de los k-vecinos más cercanos, que fue el utilizado en el presente trabajo. Así mismo, se observa que más de 20 trabajos utilizan técnicas relacionadas con expertos para el desarrollo de su sistema. Validando la importancia de contar con información experta en estudios agrícolas. La figura 4 muestra los tipos de sistemas encontrados en la RSL.

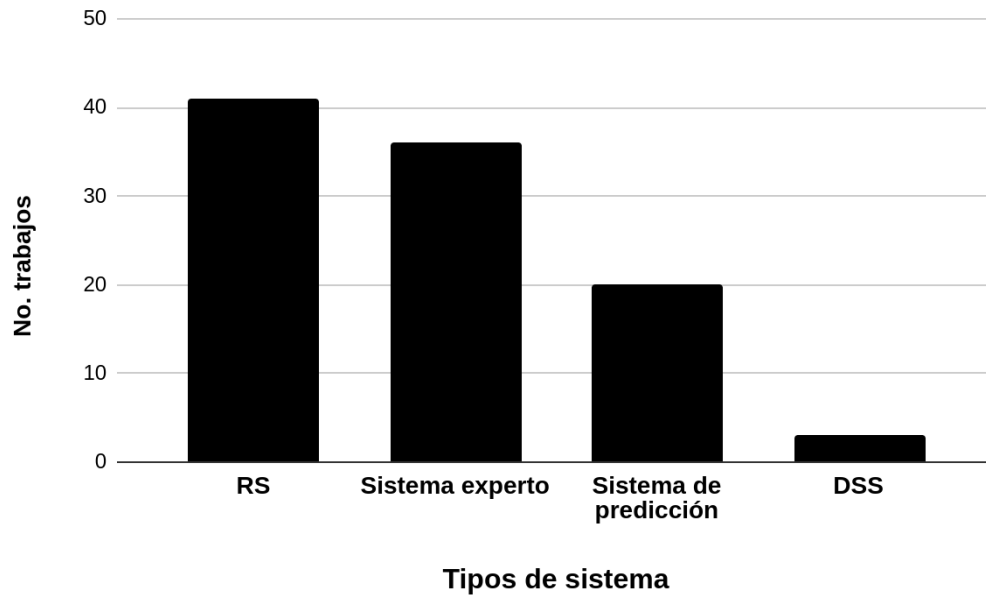


Figura 4. Tipos de sistema encontrados en los trabajos relacionados.

Se observa que los RS y los sistemas expertos son los más trabajados en este campo, ya que brindar recomendaciones a los agricultores y teniendo en cuenta información experta es lo que más ayuda a mejorar el cuidado y manejo de los cultivos.

La figura 5 muestra las variables más comunes encontradas en los 102 trabajos.

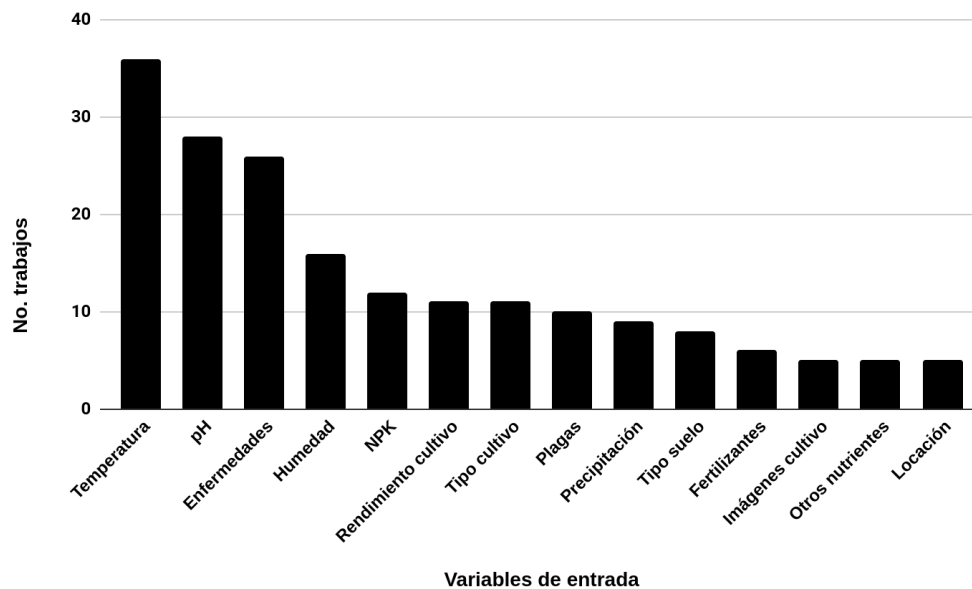


Figura 5. Extracción de las variables de entrada más comunes utilizadas en los trabajos relacionados.



## Anexo D

En este anexo se presenta la base de conocimiento construida para la realización de nuestro sistema. El cual, esta base se compone de una biblioteca virtual con artículos, documentos, informes, libros, etc. que tratan de información de los cultivos de café. En síntesis, los documentos hablan las etapas de cultivo de café, la siembra, el manejo de plagas y enfermedades, el proceso de fertilización, el cuidado del suelo, el sombrero, la descripción del año cafetero, los fertilizantes más utilizados, entre otras cosas. Por lo que, es a partir de esta información experta, que se determinan cuáles factores son más importantes al momento de realizar una recomendación de fertilizantes. Además de construir la base de casos del sistema, identificando cómo puede ser la tasa de fertilizante ante diversas condiciones de clima, suelo y cultivo.

[https://drive.google.com/drive/folders/1sNHXbDdA8w\\_RNwAFVaL\\_F0VEBoHKrRFb?usp=sharing](https://drive.google.com/drive/folders/1sNHXbDdA8w_RNwAFVaL_F0VEBoHKrRFb?usp=sharing)

## Anexo E

Este anexo presenta los resultados obtenidos de la comparación de tecnologías de redes inalámbricas para sensores más comunes. En la revisión bibliográfica realizada en el estado del arte, se evidenció que las tecnologías para redes de sensores inalámbricos más frecuentes o que sugieren los autores son: ZigBee, Wifi, NB-IoT, SigFox y LoraWan. En la siguiente tabla presenta un breve resumen comparativo con las principales características de cada una.

<b>Tecnologías</b>	<b>ZigBee</b>	<b>Wifi</b>	<b>Nb-IoT</b>	<b>SigFox</b>	<b>Lora</b>
Costos	BAJO	BAJO	ALTO	ALTO	BAJO
Consumo de energía	BAJO	ALTO	BAJO	BAJO	BAJO
Transmisión de datos	BAJO	ALTO	ALTO	ALTO	BAJO
Cobertura (Rural)	BAJA	ALTA	ALTA	MEDIA	MEDIA
Compatibilidad con otras tecnologías	SÍ	SÍ	SÍ	NO	SÍ

**Tabla comparativa de tecnologías para redes de sensores inalámbricos**

Los ítems a tener en cuenta para definir la tecnología que más utilidad y ventajas podía aportar al presente trabajo fueron:

- **Costos:** Es apropiado que la tecnología de red inalámbrica usada no sea costosa o no involucre pagar licenciamiento.
- **Consumo de energía:** Es apropiado un bajo consumo de energía debido a que un factor primordial es reducir costos, ya que el público objetivo no cuenta con altos ingresos.
- **Transmisión de datos:** Se requiere una transmisión de datos periódica ya que no es necesario tomar datos constantemente.
- **Cobertura:** Es necesario una cobertura pequeña o media debido a que el trabajo está enfocado en terrenos no muy extensos.

Con lo anterior, se definió la tecnología Lora como la más factible a utilizar en el presente trabajo.

# Anexo F

Este anexo presenta el proceso de la obtención de los datos que conforman la entrada del sistema, teniendo en cuenta que se recolectaron tres clases de datos (suelo, clima y cultivo). Para las variables de cultivo, se obtuvieron directamente de información brindada por el caficultor, dueño de la finca donde se realizó el estudio. Para las variables de clima, se utilizó un servicio de captura de datos de las estaciones climáticas, en colaboración con una de las pasantías de la empresa Ecotecma. La figura 1 del presente anexo muestra el boletín diario de la plataforma Agroclima, que recolecta diferentes datos climáticos, entre ellos la precipitación o lluvia. En el que se utilizó la técnica de Web Scraping para extraer los datos del boletín diario y almacenarlos en un archivo CSV.

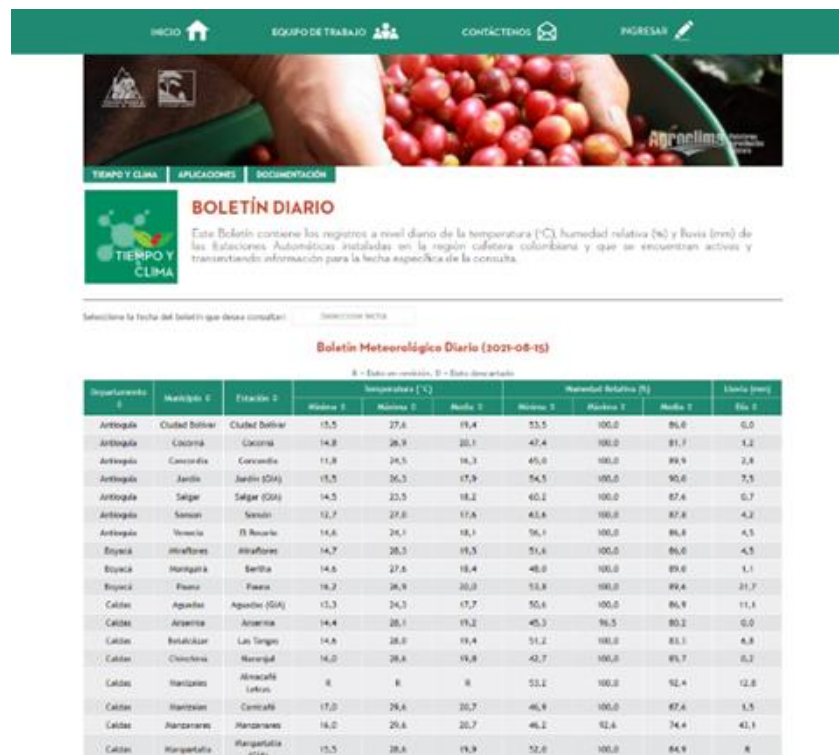


Figura 1. Boletín diario Agroclima.

El script desarrollado por una de las pasantías de Ecotecma, automatiza el envío de los datos del boletín a un servidor de Ecotecma. La figura 2 muestra este proceso.



Figura 2. Proceso de captura automática de los datos del boletín.

Así, con este servicio se obtienen los datos climáticos históricos de las diferentes estaciones climáticas ubicadas a lo largo del país. Donde para el presente trabajo, sólo se utilizaron los datos de la estación La Trinidad del municipio de Piendamó. Datos que se muestran en el anexo H.

Por otro lado, los datos del suelo que son los que recolectan los sensores de la red que se adaptó al presente sistema, son obtenidos a través del sitio web oficial del sensor HONDETEC. Teniendo en cuenta que el sensor con el que trabaja la empresa Ecotecma es un sensor chino que registra los datos en un servidor chino. Las siguientes figuras muestran el sensor utilizado, el cultivo estudiado y el servidor donde se almacenan los datos del sensor.



Figura 3 y 4. Finca vereda Alto Piendamó.

Las figuras anteriores muestran el cultivo de café que fue estudiado para brindar las respectivas recomendaciones de fertilizantes. En el que se accedió a este lugar por un período de aproximadamente 7 días, en el que se recolectaron algunos datos de suelo y de cultivo. El sensor HONDETEC trabajado se muestra en la siguiente imagen:



Figuras 5 y 6. Sensor HONDETEC utilizado.

Cabe mencionar que el HONDETEC es un kit de múltiples sensores de suelo, entre los que se destacan los sensores NPK, humedad y pH del suelo, que fueron los datos utilizados por el sistema desarrollado. La obtención de estos datos se hace a través del sitio web oficial de la marca del sensor, mostrada en la siguiente figura.

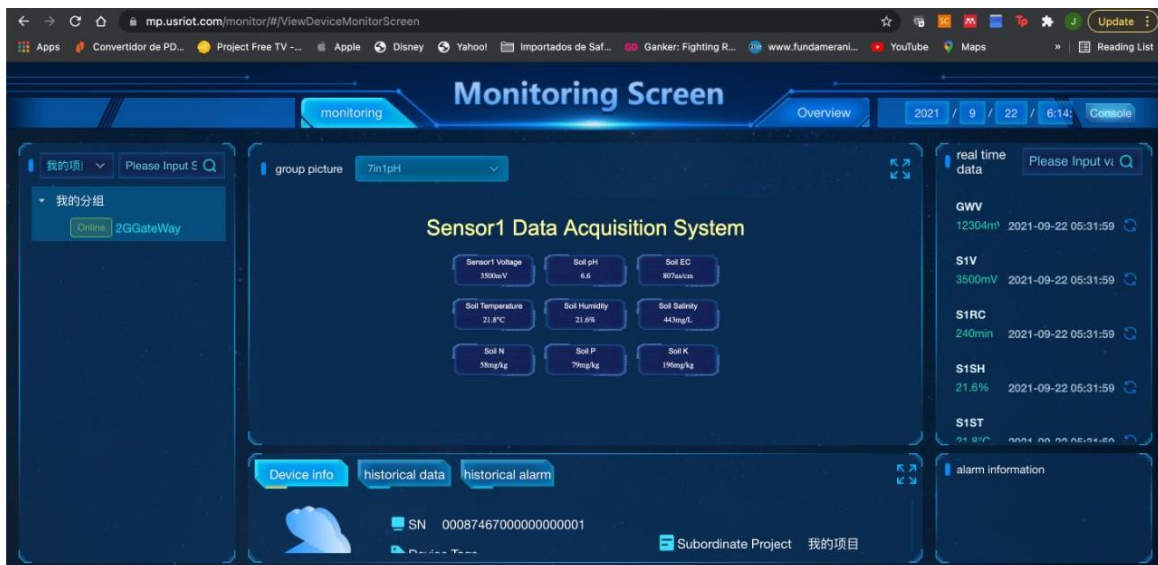


Figura 7. Sitio web de visualización de los datos de los sensores.

No se contó con el acceso al sitio web para mostrar de forma más detallada los datos recolectados por los sensores, debido a que es la empresa Ecotecma la que tiene este acceso. Sin embargo, se logró visualizar los datos de suelo recolectados para la posterior construcción del problema y entrada al sistema desarrollado.

## Anexo G

En este anexo se presenta la base de casos construida para el sistema de recomendación que utiliza el razonamiento basado en casos (CBR). Se construyeron un total de 300 casos que se pueden detallar en un archivo CSV compartido en el siguiente enlace:

<https://drive.google.com/drive/folders/1ptbrYIIT31R4uClk2ilsleft1zq5yNmY?usp=sharing>

## Anexo H

En este anexo se agregan los registros de precipitación pluvial históricos que recolecta la estación climática seleccionada en el presente trabajo. La cual se ubica en el municipio de Piendamó. El CSV con los datos están disponibles en el siguiente enlace:

<https://drive.google.com/drive/folders/1V2KJcagklbdddOpysnrRs4r6tzRuwL64?usp=sharing>



# Anexo I

En este anexo se encuentra el script o Notebook del código realizado para clasificar una temporada climática a partir de una fecha de floración dada. Se debe tener en cuenta, que este script se realizó en la plataforma de Google Colab por lo que se encuentra en el siguiente enlace compartido:

<https://colab.research.google.com/drive/1hR7c4s5voFSJUaUWBYsXieEfYciizdCQ?usp=sharing>

## Anexo J

En este anexo se presenta el código del sistema CBR que encuentra una nueva solución ante un problema que ingresa. En este código se detalla el proceso de normalización de datos, organización de la base de casos, la fase de recuperación y reutilización. Al igual que el anexo anterior, el Notebook también se comparte a través de la plataforma de Google Colab mediante este enlace:

<https://colab.research.google.com/drive/18F8AioKk4hNIJ8cAXUVSSJ3zbr4fYa-B?usp=sharing>

## Anexo K

En este anexo se presenta el script realizado para determinar el factor  $k$  y calcular la precisión del sistema desarrollado. También se detalla el proceso del cálculo del MAE y RMSE.

<https://colab.research.google.com/drive/1TMawL7SM8mqZEPL-aptVPpEzveMG37-Q?usp=sharing>

## Anexo L

Artículo titulado 'Recommender Systems for Nitrogen Fertilizers: a literature review' pendiente para mandar a revisión de revista.

Se puede encontrar en el siguiente enlace:

<https://drive.google.com/drive/u/0/folders/1LTYy5pktpI2cYnbC7gOaoy8Ybpkjdy6D>