

**Prototipo de monitoreo de cultivos de plantas aromáticas usando redes  
de sensores inalámbricos y minería de datos**

**Rafael Mauricio Navarro Casadiegos  
Pedro Javier León Uribe**

**PROYECTO DE GRADO**

**Universidad Pontificia Bolivariana  
Escuela De Ingeniería  
Bucaramanga  
2021**

**Prototipo de monitoreo de cultivos de plantas aromáticas usando redes  
de sensores inalámbricos y minería de datos**

**Rafael Mauricio Navarro Casadiegos  
Pedro Javier León Uribe**

**PROYECTO DE GRADO  
Presentado ante la Universidad Pontificia Bolivariana  
Como requisito para optar al título de  
INGENIERO ELECTRÓNICO**

**Director del proyecto:  
Jhon Jairo Padilla Aguilar. PhD.**

**Universidad Pontificia Bolivariana  
Escuela De Ingeniería  
Bucaramanga  
2021**

Página reservada para el acta de evaluación

## **DEDICATORIA**

A nuestras familias, profesores y todos aquellos que de una u otra forma hicieron este paso posible y creyeron en nosotros.

## **AGRADECIMIENTOS**

Al doctor Jhon Jairo Padilla Aguilar, que siendo nuestro director de proyecto nos orientó y motivó usando su invaluable conocimiento y experiencia. A la doctora Sandra Cuervo, cuya asesoría fue clave durante las primeras etapas de desarrollo de este trabajo. A todos quienes fueron nuestros profesores, que nos brindaron su conocimiento como un escalón para lograr esta meta. A nuestras familias y a todos los que creyeron en nosotros, porque completar este camino solo no es imposible, pero gracias a su compañía y apoyo se hizo más agradable.

## TABLA DE CONTENIDO

TABLA DE CONTENIDO.....	6
PLANTEAMIENTO DEL PROBLEMA .....	17
1.1 Definición del problema.....	17
1.2 Objetivo General .....	17
1.3 Objetivos Específicos .....	17
1.4 Justificación.....	18
ANTECEDENTES .....	19
MARCO TEÓRICO.....	20
3.1 Plantas aromáticas.....	20
3.1.1 Hierbabuena .....	20
3.2 Variables .....	21
3.3 Smart farming (Agricultura inteligente).....	21
3.4 Protocolos de comunicación .....	22
3.4.1 Protocolo Modbus .....	22
3.4.2 Arquitectura de protocolos TCP/IP.....	22
3.4.3 ZigBee .....	23
3.5 Controladores.....	23
3.5.1 MPC-134.....	23
3.5.2 UCM- 316 .....	24
3.5.3 IAQM-THCO2 .....	25
3.5.4 DM-124 .....	25
3.5.5 Higrómetro Arduino FC-28 (YL 36 y YL 69) .....	26
3.5.6 Sensor de luminosidad -fotorresistencia LDR GL55 .....	26
3.6 Software.....	26
3.6.1 Power BI .....	26
3.6.2 Python.....	27
3.6.2.1 Pandas .....	27
3.6.2.1 Numpy .....	27
3.6.2.2 Scikit-Learn.....	27
3.6.2.3 Pickle.....	27

3.7 Minería de datos (Data mining) .....	28
3.8 Machine Learning.....	28
3.8.1 Aprendizaje supervisado.....	29
3.8.2 Modelos de Entrenamiento .....	29
3.8.2.1 Árboles de decisión .....	29
3.8.2.2 Bosques aleatorios (random forest).....	32
3.9 Métricas de rendimiento .....	32
3.9.1 Matriz de confusión.....	32
3.9.2 Precisión .....	33
3.9.3 Recall.....	33
3.9.4 f1-score.....	33
3.10 Normalización de datos.....	34
3.11 Test de rasgos de importancia.....	34
3.12 Crossvalidation.....	34
METODOLOGÍA.....	35
4.1. Experimento .....	35
4.1.1. Fase exploratoria .....	35
4.1.2. Diseño e implementación de la estructura física.....	35
4.1.2.1 Configuración de dispositivos .....	36
4.1.3. Adquisición de datos y puesta en funcionamiento .....	58
4.1.4. Creación de algoritmos, limpieza y manipulación con técnicas de machine learning .....	59
4.1.4.1 Data Wrangling.....	60
4.1.4.2 Modelado.....	62
4.1.4.3 Evaluación .....	65
4.1.4.4 Análisis de resultados y selección del modelo.....	68
4.1.5 Diseño del dashboard .....	69
4.1.5.1 Instalación y configuración de Power BI .....	69
RESULTADOS .....	82
CONCLUSIONES Y RECOMENDACIONES .....	85
6.1 Conclusiones.....	85
6.2 Recomendaciones .....	85
BIBLIOGRAFÍA .....	86



## INDICE DE FIGURAS

Figura 1. Flor y hojas de la hierbabuena o Mentha Sativa L.....	20
Figura 2. Controlador MPC-134 + cable USB A/B.....	23
Figura 3. Controlador UCM- 316 + conexión por RJ-45.....	24
Figura 4. Sensor digital IAQM-THCO2.....	25
Figura 5. Puente de interconexión alámbrica e inalámbrica DM-124 .....	25
Figura 6. Estructura de un árbol de decisión .....	30
Figura 7. Estructura de una matriz de confusión.....	33
Figura 8. Proceso de aplicación de k-fold crossvalidation.....	34
Figura 9. FT_Prog .....	37
Figura 10. Pasos par configuración UD 100 mediante consola.....	38
Figura 11. Visualización de la consola luego de terminar todo el proceso .....	38
Figura 12. Ventana principal del WM Config Tool.....	39
Figura 13. Programa de configuración WM Config Tool, utilizando UD1000 como intercomunicador. ....	40
Figura 14. Configuración y parámetros del DM 124 nodo de conexiones. ....	41
Figura 15. Parámetros de configuración para nodos esclavos .....	41
Figura 16. Esquema del puerto UART .....	42
Figura 17. Configuración 2 MPC 134 ubicación y tiempo de lectura para los datos .....	42
Figura 18. Configuración 3 MPC 134 asignación de trabajo en cada uno de los puestos UART`s.....	43
Figura 19. Configuración de comunicación de puertos UART RS485 para el MPC 134 .....	44
Figura 20. Configuración de sensores análogos en MPC 134.....	44
Figura 21. Registro de entradas analógicas MPC 134.....	45
Figura 22. Configuración de lectura de datos transmitidos vía UART RS485 para UCM 316.....	45
Figura 23. Registro de entradas analógicas UCM 316.....	46
Figura 24. Visualización de los valores recogidos por el UCM 316 .....	46
Figura 25. Configuración de comunicación del UCM 316.....	47
Figura 26. Configuración de puerto de entrada análogo de sensor LDR GL55. ....	48
Figura 27. Configuración 3 UCM 316 organización de parámetros de lectura, almacenamiento y carga .....	49
Figura 28. Configuración 3 UCM 316 tiempo y cantidad de lecturas.....	49
Figura 29. Configuración de envío de información vía FTP para UCM 316 .....	50
Figura 30. Configuración inicial para servidor FTP en Filezilla Server .....	51
Figura 31. Pantalla principal Filezilla Server .....	52
Figura 32. Menú General settings de Filezilla Server.....	52
Figura 33. Menú Passive mode settings de Filezilla Server.....	53
Figura 34. Menú FTP over TLS settings de Filezilla Server .....	54
Figura 35. Menú Autoban de Filezilla Server .....	55
Figura 36. Menú General de usuarios de Filezilla Server .....	55
Figura 37. Configuración de nuevo usuario en Filezilla Server.....	56
Figura 38. Menú Shared folders de Filezilla Server.....	56
Figura 39. Búsqueda de carpeta a compartir con el usuario.....	57
Figura 40. Arquitectura física del experimento .....	57
Figura 41. Montaje realizado para las pruebas .....	58

Figura 42. Datos obtenidos a partir de los archivos.....	59
Figura 43. Gráfica de datos de temperatura obtenidos.....	60
Figura 44. Gráfica de datos de CO2 obtenidos.....	60
Figura 45. Gráfica de datos de humedad de aire obtenidos.....	61
Figura 46. Gráfica de datos de humedad de suelo obtenidos.....	61
<i>Figura 47. Gráfica de datos de luminosidad obtenidos.....</i>	<i>61</i>
Figura 48. Gráfica de datos de Estados planta 1 obtenidos.....	61
Figura 49. Gráfica de datos de Estados planta 2 obtenidos.....	61
Figura 50. Gráfica de datos de Numero de hojas planta 1 obtenidos (fuente: Autor) ....	61
Figura 51. Gráfica de datos de Numero de hojas planta 2 obtenidos.....	61
Figura 52. Tabla de datos usada para la creación del modelo.....	62
Figura 53. Etapas de desarrollo del algoritmo.....	62
Figura 54. Proceso de desarrollo de árbol de decisión sin podar.....	63
Figura 55. Proceso de desarrollo de árbol de decisión con poda.....	63
Figura 56. Proceso de desarrollo de random forest.....	64
Figura 57. Proceso de desarrollo de random forest con datos balanceados.....	64
Figura 58. Proceso de desarrollo de random forest con datos balanceados y normalizados.....	65
Figura 59. Resultados predichos contra resultados reales.....	69
<i>Figura 60. Descarga Powe Bi paso 1.....</i>	<i>70</i>
Figura 61. Descarga Powe Bi paso 2. Zona de descargas Power Bi de manera gratuita.....	70
Figura 62. Descarga Powe Bi paso 3.....	71
Figura 63. Descarga Powe Bi paso 4.....	71
Figura 64. Icono de instalación.....	71
Figura 65. Pantalla de inicio Power Bi versión de escritorio.....	72
Figura 66. Obtención y origen de los datos a utilizar.....	72
Figura 67. Recuadro de conexión.....	73
Figura 68. Ventana de ajuste de datos – operaciones básicas Excel.....	73
Figura 69. Ventana de conexión con el punto de origen.....	74
Figura 70. Ventana de conexión puesta a punto con el origen del archivo.....	74
Figura 71. Plataforma de trabajo.....	75
Figura 72. Ejemplo de CO2 vs Hora.....	75
Figura 73. Área de trabajo y herramientas de la plataforma.....	76
Figura 74. Accesos a las configuraciones y programación de Power Bi web.....	76
Figura 75. Ruta de majeo y configuraciones de los datos adjuntos al panel/Dashboard.....	77
Figura 76. Parámetros y herramientas para actualización remota automática.....	77
Figura 77. Credencial de acceso.....	78
Figura 78. Puerta de enlace lista para usarse.....	78
Figura 79. Configuración de la puerta de enlace y frecuencia de actualización automática.....	79
Figura 80 Información general de cada una de las variables visualizada por fecha.....	80
Figura 81. Información correspondiente al sujeto de pruebas 1.....	80
Figura 82. Información correspondiente al sujeto de pruebas 2.....	81
Figura 83. Diagramas de cajas correspondiente a los modelos desarrollados en la fase 2 del desarrollo del modelo.....	82

Figura 84. Diagramas de cajas correspondiente a los modelos desarrollados en la fase 3 del desarrollo del modelo.....	83
Figura 85. Diagramas de cajas correspondiente a los modelos desarrollados en la fase 4 del desarrollo del modelo.....	84

## INDICE DE TABLAS

Tabla 1. Resultados del test de rasgos de importancia.....	66
Tabla 2. Métricas para modelos de la fase.....	66
Tabla 3. Métricas para modelos con datos normalizados de la fase 2 .....	66
Tabla 4. Métricas para modelos de la fase 3.....	67
Tabla 5. Métricas para modelos con datos normalizados de la fase 3 .....	67
Tabla 6. Métricas para modelos de la fase 4.....	67
Tabla 7. Métricas para modelos con datos normalizados de la fase 4 .....	67
Tabla 8. Comparación de los mejores resultados obtenidos en cada fase .....	68
Tabla 9. Resultados del modelo entrenado para el valor 1 con datos de test ..	68
Tabla 10. Resultados del modelo entrenado para el valor 2 con datos de test.	68

## LISTA DE ACRÓNIMOS Y ABREVIACIONES

HMI: Human Machine Interface (Interfaz humano máquina)  
PLC: Programmable Logic Controller (Control lógico programmable)  
TIC: Tecnologías de la información y comunicación  
TCP: Transmission Control Protocol (Protocolo de control de transmisión)  
UDP: User Datagram Protocol (Protocolo de datagramas de usuario)  
FTP: File Transfer Protocol (Protocolo de transferencia de archivos)  
HTTP: Hypertext Transfer Protocol (protocolo de transferencia de hipertextos)  
SMTP: Simple Mail Transfer Protocol (protocolo simple de transferencia de correo)  
DNS: Domain Name System (sistema de nombres de dominio)  
NFS: Network File System (sistema de archivos de red)  
USB: Universal Serial Bus (Bus Universal en Serie)  
RS232: Recommended Standard 232 (Estándar Recomendado 232)  
RS485: Recommended Standard 485 (Estándar Recomendado 485)  
GSM: Global System for Mobile communications (sistema global para las comunicaciones móviles)  
GPRS: General Packet Radio Service (servicio general de paquetes vía radio)  
MQTT: Message Queing Telemetry Transport  
IEEE: Instituto de Ingenieros Eléctricos y Electrónicos  
NAT: Network Address Translation

## RESUMEN GENERAL DE TRABAJO DE GRADO

**TITULO:** PROTOTIPO DE MONITOREO DE CULTIVOS DE PLANTAS AROMÁTICAS USANDO REDES DE SENSORES INALÁMBRICOS Y MINERÍA DE DATOS

**AUTOR(ES):** Rafael Mauricio Navarro Casadiegos  
Pedro Javier León Uribe

**PROGRAMA:** Facultad de Ingeniería Electrónica

**DIRECTOR(A):** Jhon Jairo Padilla Aguilar

### RESUMEN

Este proyecto plantea la creación de un prototipo de monitoreo de cultivos de hierbabuena usando una Red de sensores de los cuales se obtienen datos destinados al entrenamiento de un modelo de Machine Learning, que es el encargado del monitoreo y la predicción del estado de la planta, pudiendo estar saludable, enferma o muerta. Inicialmente se presentan trabajos previos que sirven como antecedentes. Luego se presenta la información teórica acerca de la yerbabuena, de la infraestructura física que se implementa, del software y las librerías claves para el desarrollo del modelo de Machine Learning y del proyecto, así como de las distintas métricas que se utilizan para monitorear el rendimiento de los posibles modelos, de donde se obtiene que el modelo elegido es un random forest con datos balanceados y normalizados, que obtuvo un valor de 0.997 de precisión, 0.741 de recall y 0.850 de f1-score a la hora de predecir el valor 1 del estado de la planta, que corresponde a las plantas en buen estado; y un 0.989 de precisión, 0.419 de recall y 0.589 de f1-score a la hora de predecir el valor 2 del estado de la planta, que corresponde a las plantas enfermas o muertas. Posteriormente, los datos medidos por la red de sensores, variables que son agregadas de forma manual y el estado de salud de la planta son accesibles mediante la HMI desarrollada en Power BI.

### PALABRAS CLAVE:

Machine Learning, random forest, árboles de decision, precision, recall, f1-score, monitoreo



Vº Bº DIRECTOR DE TRABAJO DE GRADO

## GENERAL SUMMARY OF WORK OF GRADE

**TITLE:** AROMATIC PLANTS CROP MONITORING PROTOTYPE USING WIRELESS SENSOR NETWORK AND DATA MINING

**AUTHOR(S):** Rafael Mauricio Navarro Casadiegos  
Pedro Javier León Uribe

**FACULTY:** Facultad de Ingeniería Electrónica

**DIRECTOR:** Jhon Jairo Padilla Aguilar

### ABSTRACT

This project proposes the creation of a prototype for monitoring mint crops using a sensor network from which data is obtained for training a Machine Learning model, which is responsible for monitoring and predicting the state of the plant, which can be healthy, sick or dead. Initially, previous works that serve as background are presented. Then, the theoretical information about the spearmint, the physical infrastructure to be implemented, the software and key libraries for the development of the Machine Learning model and the project are presented, as well as the different metrics used to monitor the performance of the possible models, from which it is obtained that the model chosen is a random forest with balanced and normalized data, which obtained a value of 0.997 of precision, 0.741 of recall and 0.850 of f1-score when predicting the value 1 of plant condition, which corresponds to plants in good condition; and 0.989 of precision, 0.419 of recall and 0.589 of f1-score when predicting the value 2 of plant condition, which corresponds to diseased or dead plants. Subsequently, the data measured by the sensor network, variables that are manually aggregated and the plant health status are accessible through the HMI developed in Power BI.

### KEYWORDS:

Machine Learning, random forest, decision trees, precision, recall, f1-score, monitoring.



Vº Bº DIRECTOR OF GRADUATE WORK

## INTRODUCCIÓN

Con el descubrimiento de la agricultura el ser humano cambió drásticamente la forma en la que conseguía sus alimentos, pasando de ser cazadores-recolectores a lo que actualmente se conoce como sociedades agrarias. A partir de este punto, la agricultura fue evolucionando a la par de la humanidad, como sucedió en la antigüedad en sitios como Mesopotamia y Egipto, donde se utilizaban sistemas de irrigación para proveer de agua a sus cultivos, o a inicios del siglo XX con el uso de fertilizantes sintéticos, y pesticidas, o la implementación del tractor de propulsión a gasolina, cambiando los animales de tiro que hasta el momento se usaban para el arado[1][2][3][4].

Como resultado de la manipulación humana, diversos alimentos han sufrido numerosos cambios. Cereales como el trigo y el maíz han visto alteradas drásticamente su forma y tamaño, ya que a través de selección se ha optado por utilizar los ejemplares con granos más grandes, más resistentes a las plagas y que se adapten a las condiciones climáticas del sitio donde se planta[5], pero esto no las hace completamente inmunes y en un mundo donde la producción a gran escala.

# CAPITULO 1

## PLANTEAMIENTO DEL PROBLEMA

### 1.1 Definición del problema

Los países de Sudamérica poseen grandes riquezas en el sector agropecuario. Estas riquezas se han manejado mediante técnicas que dependen, de forma significativa, de la intervención y experticia humana para ejecutarse, esto incluye la detección y reacción ante fenómenos como las plagas, cambios en el clima, enfermedades; teniendo consecuencias a nivel de productividad y competitividad. En la actualidad la tecnología ha llegado a desarrollarse a niveles que años antes hubieran sido inimaginables, encontrándose actualmente inmersa en casi todos los ámbitos de la vida, y el sector agropecuario no es ajeno a dicho fenómeno. La necesidad en la que se enfoca este proyecto es buscar que, con ayuda de la tecnología basada en minería de datos y redes de sensores inalámbricos, se pueda prevenir y alertar de prácticas poco eficientes, que puedan afectar las plantas (crecimiento, salud, esperanza de vida). Para efectos de este estudio se optó por tomar a la hierbabuena como objeto de los distintos análisis y experimentos a realizar, ya que además de ser una planta que no ocupa mucho espacio y no consume muchos recursos, se cuenta con la asesoría de un docente con amplios conocimientos sobre esta.

### 1.2 Objetivo General

Diseñar e implementar un prototipo de monitoreo de cultivos de hierbabuena usando una Red de sensores inalámbricos, y minería de datos.

### 1.3 Objetivos Específicos

- Diseñar, instalar y realizar la puesta a punto de los equipos necesarios para el monitoreo de las variables CO<sub>2</sub>, temperatura y humedad relativa del aire y la humedad del suelo en el cultivo de hierbabuena
- Diseñar e implementar el software de minería de datos y su respectiva HMI para el análisis de variables agrícolas.
- Analizar con técnicas de minería de datos la información adquirida, con el fin de mostrar alertas preventivas del estado del cultivo.

## 1.4 Justificación

Este proyecto se realiza no solo con el fin de promover dinámicamente la tecnología en el sector agropecuario, sino con el fin de prevenir y tomar decisiones acertadas y oportunas, y poner a prueba los diversos protocolos de comunicación en sí, con el fin de recopilar información vital para su posterior análisis de los cultivos. Cabe destacar que en Colombia se producen infinidad de productos, pero debido a la falta de implementación tecnológica sobre todo en el sector agro, no se puede explotar al máximo el potencial del país. Por otra parte, este proyecto pretende recolectar información de las variables temperatura, nivel de CO<sub>2</sub>, Humedad y luminosidad, para realizar posteriores análisis y poder deducir patrones y modelos a través de minería de datos que permitan predecir las eventualidades que puedan ocurrir a las plantaciones en primera instancia. Este proceso se realiza a través de la minería de datos, la cual está basada en algoritmos de aprendizaje automático de máquinas (machine learning). Por otro lado, considerado como el pivote de una venta, el productor, es el encargado de lograr la mayor calidad en sus artículos. Los gastos y la eficiencia de la producción se pueden mejorar gracias a los avances tecnológicos que se pueden encontrar en la actualidad. Por otra parte, la versatilidad para realizar comunicaciones entre sensores, actuadores y controladores de manera eficiente y articulada entre los diferentes protocolos permite mejorar las condiciones de las plantas. Sin embargo, hoy en día existen tecnologías de menor costo y muy buena calidad que permiten el uso de redes de sensores inalámbricos (WSN) junto con tecnologías de redes industriales como PLCs y comunicaciones MODBUS para generar una automatización eficiente en cultivos agrícolas.

## CAPÍTULO 2

### ANTECEDENTES

El desarrollo que ha tenido la tecnología y la cada vez más notable utilidad en los distintos campos de la vida humana, donde la agricultura no ha sido ajena, ha provocado que se desarrollen estudios como realizado por Sally Jo Cunningham et al, en la universidad de Waikato, en Hamilton Nueva Zelanda , titulado “Developing innovative applications in agriculture using data mining” [6].donde se buscó, utilizando minería de datos y mediante la implementación de un software llamado WEKA inducir un sistema de clasificación capaz de catalogar champiñones en grados de calidad y lograr una precisión similar a la obtenida por inspectores humanos, obteniendo que los resultados arrojados por los algoritmos de minería de datos permitieron clasificar los champiñones en tres grados distintos de calidad.

En el ámbito nacional, se tiene como referencia un estudio en cultivos de arroz de regadío y arroz de secano (que no tiene riego y solo de beneficia del agua de lluvia) en los departamentos de Meta y Tolima; realizado por Sylvain Delerce et al, titulado “Assessing Weather-Yield Relationships in Rice at Local Scale Using Data Mining Approaches” [7] donde se buscó analizar, mediante el uso de datos de observación y de minería de datos, el papel de la variabilidad climática como un limitante en la producción en grandes cultivos de arroz en dichas regiones y qué factores de esta variabilidad fueron los que más impactaron, además de validar los datos de observación como una fuente confiable de información; obteniendo como resultado destacable que la variabilidad climática podía afectar entre el 32%, 39% e incluso llegando al 60% dependiendo de las combinaciones de cultivos , y la validación de los datos de observación como una fuente fiable de la cual extraer información valiosa para un manejo optimizado del arroz.

## CAPITULO 3

### MARCO TEÓRICO

Para comprender de mejor manera el proyecto que se pretende realizar, es importante conocer todos los elementos involucrados en este, mediante el presente marco teórico.

#### 3.1 Plantas aromáticas

Como paso inicial partamos de conocer que las plantas aromáticas (que son los sujetos de estudio) son plantas medicinales, cuyos principios activos están constituidos total o parcialmente por esencias[8]. Se escogió como sujeto de estudio a la Hierbabuena, por lo cual se procede a dar una explicación más detallada de esta.

##### 3.1.1 Hierbabuena



*Figura 1. Flor y hojas de la hierbabuena o Mentha Sativa L.*

(fuente: [https://laenergiadelasflores.files.wordpress.com/2018/04/peppermint\\_20180304\\_231052.png?w=900&h=900](https://laenergiadelasflores.files.wordpress.com/2018/04/peppermint_20180304_231052.png?w=900&h=900) [Último acceso 25 noviembre 2020])

De acuerdo con la RAE (Real Academia Española de la lengua), la hierbabuena es una planta herbácea (de crecimiento rápido, que no tiene tallo leñoso y cuyas hojas son de color verde[9]) con tallos erguidos, poco ramosos, hojas vellosas, elípticas, agudas, nerviosas y aserradas, flores rojizas, fruto seco con cuatro semillas, frecuentemente cultivada en huertas, que desprende un olor agradable y tiene un uso extendido como condimento[10].

La hierbabuena es una planta poco tolerante a las sequías, que se desarrolla mejor en climas húmedos y templados, siendo plantada mediante la extracción de rizomas (tallos horizontales y subterráneos[11]) provistos de raíces, de plantas madres, colocados a una distancia de entre 20 y 25 cm, a una profundidad de 5 cm. Su primera cosecha se da alrededor de 5 o 6 meses después de germinada la planta[12].

### 3.2 Variables

Según la RAE, una variable se puede definir como un elemento o causa que afecta algún proceso[13]. Para nuestro caso de estudio, una variable es un elemento o un factor que puede alterar para bien o para mal, el estado de la hierbabuena. Dentro de las variables que se contemplan analizar se encuentran:

- Temperatura: Es la energía cinética interna debido al movimiento de las moléculas del cuerpo[14].
- Nivel de CO<sub>2</sub>: El dióxido de carbono es considerado la forma más estable del carbono, aparte de ello es uno de los elementos más utilizado en la creación y cambio energético sobre las plantas y animales, puesto que con ayuda de agua y un proceso fotosintético es transformado en carbohidratos y liberando como residuo partículas de oxígeno. Las grandes concentraciones el CO<sub>2</sub> puede resultar peligroso para la vida animal[15].
- Humedad: La humedad es el porcentaje o nivel de agua que se encuentra en el aire por unidad volumétrica. Cuando el nivel alcanza el 100% decimos que el aire está saturado, este exceso lleva a la condensación del vapor convirtiéndolo en gotas de agua. Existen 3 tipos de humedad: humedad absoluta, humedad relativa y humedad específica. La primera consiste en la densidad de vapor de agua suspendida en el aire a temperatura y presión determinada, la segunda es la porción de aire que contiene vapor de agua real comparada con la cantidad de vapor de agua necesaria para la saturación a una temperatura correspondiente y es medida en porcentaje, y finalmente esta la humedad específica que consiste en una muestra de aire húmedo y representa la fracción de masa de vapor con respecto a la masa de aire húmedo[16].

### 3.3 Smart farming (Agricultura inteligente)

Este concepto de última tendencia tiene como propósito la implementación de las TIC's (tecnologías de la información y comunicación) para el control y manipulación de cultivos agrícolas, por otro lado, la utilización de GPS, automatización, robótica y comunicaciones electrónicas buscan incrementar la sostenibilidad y producción agrícola, basada en el monitoreo de las variables claves de la buena producción[17].

Por otro lado, la rentabilidad, respecto al consumo de fertilizantes y venenos es reducido pues el control y monitoreo se realiza en tiempo real, realizando los respectivos correctivos a tiempo. Un aspecto importante a resaltar es la salubridad del mismo producto, pues las condiciones de cultivo son controladas en su mayoría, entretanto las plantas son monitoreadas, adicionalmente la producción aumenta debido a esas mejoras de condiciones, no obstante, dicha tecnología es realmente articulada, gracias a la diversidad de tecnologías y sensores que se van desarrollando a través del tiempo[18].

### **3.4 Protocolos de comunicación**

Los protocolos de comunicación son un conjunto de reglas para la transmisión o intercambio de datos dentro de una red [19]. Los protocolos que van a usarse durante el desarrollo del proyecto son:

#### **3.4.1 Protocolo Modbus**

Es un protocolo diseñado para aplicaciones industriales, en la cual su comunicación se realiza de manera serial, y basados en una arquitectura maestro esclavo, brindando así una buena velocidad a bajo costo, por otro lado, este dispositivo permite conectar hasta 247 esclavos a un solo maestro, esto permite la división en varias capas permitiendo controlar PLC's o incluso drivers de motores. ModBus posee 2 variantes para su conexión, RTU una representación binaria compacta de datos y ASCII una representación más legible pero poco eficiente[20].

Una de las grandes ventajas de la comunicación ModBus es que al no presentar cambios o codificaciones de información de tal manera que el flujo de información binaria sea plana y rápida. Un carácter de silencio tiene la duración de un byte de datos enviado por el medio, pero no transporta datos, y su duración (T) depende de la velocidad (Vt) y del número bits que se usen para su codificación (N) según  $T=N/Vt$ . Según el estándar de MODBUS, para velocidades de hasta 19.200bps, el tiempo entre tramas debe ser como mínimo 3,5 veces la duración de un carácter, y para velocidades superiores se recomienda un tiempo fijo de 1,75ms [21]. Su medio de conexión consta de un cable semidúplex (RS-422, BC 0-20mA o fibra óptica), manejando velocidades desde los 75 a los 19200 baudios, su comunicación se realiza de manera asíncrona y con un rango de distancia de 1.2 kilómetros de distancia sin uso de repetidora [22]

#### **3.4.2 Arquitectura de protocolos TCP/IP**

En 1968 la agencia de investigación avanzada DARPA desarrolla un método de transmisión de información entre redes de distintos tipos y se implementó una red para interconectar redes de diferentes tipos, llamada ARPANET. Generalmente Se encuentran entre 3 y 5 niveles funcionales en la construcción del protocolo TCP/IP:

1. Nivel físico
2. Nivel de vínculo de datos
3. Nivel internet
4. Nivel de transporte
5. Nivel de aplicación

El nivel de acceso a la red (nivel físico y nivel de vínculo de datos) es el nivel más bajo dentro de la jerarquía TCP/IP, en él se encapsula la información, en tramas distribuidas a través de la red a su destino final. En la siguiente capa se encuentra el nivel de internet, en él los paquetes son orientados a través de las rutas más adecuadas para llegar a su destinatario; en esta capa es muy

importante la dirección IP para hacer un enrutamiento del origen al destino. A continuación, se encuentra el nivel de transporte el cual facilita que los paquetes lleguen en secuencia y sin errores; sus principales ítems son la TCP, UDP. Finalmente, en el nivel de aplicación se encuentran todos los procesos que hacen uso del nivel de transporte; protocolos tales como el FTP, HTTP, SMTP, DNS, NFS, Telnet, son los existentes en este nivel [23].

### 3.4.3 ZigBee

Es una de las nuevas tecnologías inalámbricas de baja latencia y consumo, originaria de la antigua alianza HomeRF. Es utilizada en el hogar para seguridad y automatización. Por otro lado, también se suele encontrar esta tecnología en juguetes interactivos reconocimiento remoto y medicina entre otros. Dicha tecnología está diseñada para trabajar en tasas que van desde los 20 a los 250 Kbps en un rango máximo de 13 a 75 metros con un consumo de potencia 1mW el cual cubre un rango de 13 m de radio y 154 m con 1 W de consumo. Es importante resaltar que este método tiene buena relación señal-ruido puesto que es casi inmune al ruido.

En cuanto seguridad ZigBee utiliza encriptación ASE de 128 bits la cual permite encriptaciones en las comunicaciones. Además de ello utiliza claves de seguridad en los enlaces de red cuando se ejecuta una comunicación[24].

En lo que respecta a la aplicación en la agricultura, Zigbee es muy útil, ya que mediante su uso se pueden conformar redes de sensores inalámbricos con cobertura de varios kilómetros, sin necesidad recurrir al uso de cables[25]

## 3.5 Controladores

Los controladores son dispositivos mecánicos, electromecánicos o electrónicos que usan una señal de entrada para censar y/o cambiar condiciones o valores en procesos [26]. A continuación, se muestran una serie de controladores que se van a usar durante el desarrollo del proyecto.

### 3.5.1 MPC-134



*Figura 2. Controlador MPC-134 + cable USB A/B.  
(Fuente: Autor)*

Es un controlador fabricado por Advanticsys que cuenta con un banco de registros, elaborado con el propósito de medir y almacenar información mediante la recepción de señales ya sean, analógicas o digitales, mediante protocolos de conexión. Dicho dispositivo se compone de interfaces ModBus RS232, RS485 y GSM / GPRS. Además, cuenta con 4 entradas analógicas (de voltaje, corriente, resistencias o sensores de temperatura PT100), 16 entradas digitales, 8 salidas digitales, horario programable y un sistema de alarma reportado por el sistema de mensajería[27].

### 3.5.2 UCM- 316



*Figura 3. Controlador UCM- 316 + conexión por RJ-45  
(Fuente: Autor)*

El UCM-316 funciona como banco de almacenamiento de datos y controlador remoto elaborado con el fin de registrar, medir y controlar datos medidos a través de los protocolos ModBus que contiene. Posee entradas analógicas (como corriente, voltaje, resistencia o sensores de temperatura PT100) y entradas discretas (como pulsos, alarmas o eventos), conserva un control de salidas discretas. Una de sus grandes ventajas es el suministro de información de manera remota utilizando el protocolo abierto Modbus TCP/IP o utilizando GPRS utilizando el formato csv (Excel).

Por otro lado, cuenta con un reloj real que se puede sincronizar con el tiempo a través del protocolo NTP, para realizar previas programaciones en la recolección de datos. Sus protocolos de conexión son Modbus RS232 y RS 485, USB, Internet y 3G (UMTS/HSPA+).

Su función como maestro le permite conectar en simultaneo hasta 20 dispositivos por M-bus, compone de 6 entradas analógicas (corriente, voltaje, sonda de temperatura), 4 entradas digitales (contacto, contador de impulsos y alarmas), también posee 2 salidas de relé de hasta 230 voltios alternos con 3 amperios, posee un atenuador de 0 a 10 voltios, contiene una memoria de 8gb en uSD finalmente admite los protocolos ya mencionados además el TCP/IP, FTP y MQTT [28].

### 3.5.3 IAQM-THCO2



*Figura 4. Sensor digital IAQM-THCO2  
(Fuente: Advanticsys)*

Este dispositivo es un sensor inalámbrico con protocolo IEEE 802.15.4 compatible con redes DM-124 Modbus. Mide específicamente la calidad del aire como la temperatura, humedad relativa, CO2. Posee un sistema auto-configurado. Sus principales características son:

- Mediciones de temperatura, humedad y CO2
- Comunicación inalámbrica basada en IEEE 802.15.4 (2.4 GHz)
- Auto configurable
- Transmisión inalámbrica Modbus RTU
- Incluye adaptador de corriente AC / DC
- Pantalla LCD que muestra los niveles de sensor en tiempo real [29].

### 3.5.4 DM-124



*Figura 5. Puente de interconexión alámbrica e inalámbrica DM-124  
(Fuente: Advanticsys)*

Su función es interconectar los puestos de trabajo de manera inalámbrica utilizando IEEE802.15.4 y traduciéndola del protocolo analógico RTU cableado como RS485 (es un puente entre Wireless Modbus RS485 - IEEE 802.15.4 2.4 GHz), permitiendo acciones con sensores inalámbricos. Este dispositivo maneja una autoconfiguración con respecto al tráfico de sus paquetes pues es quien decide la mejor ruta para el envío. Su rango de alimentación esta entre los 110 y 230 Vca.

Al ser instalado, el DM-124 es transparente para los equipos en la red RS485, no solo manteniendo los requisitos eléctricos del bus sino también la rectitud de los paquetes de RTU Modbus transmitidos.

Características:

- Comunicación inalámbrica basada en IEEE 802.15.4 (2.4 GHz).
- Auto configurable.
- Transmisión inalámbrica Modbus RTU.
- Entrada de alimentación de 110-230 VCA.
- Dispositivo de riel DIN.
- Conexión fácil (terminales Modbus A y B +).
- Contador de impulsos configurable, con 3 modos de funcionamiento: contacto seco, colector abierto, tipo S0 (contadores de gas y agua).
- Nodo coordinador - Puente Modbus dentro de las redes de sensores inalámbricos ADVANTICSYS[30].

### **3.5.5 Higrómetro Arduino FC-28 (YL 36 y YL 69)**

Consiste en un dispositivo capaz de arrojar valores numéricos de tipo digital o análogo correspondientes a la humedad del suelo, dicho dispositivo trabaja a través de un divisor de voltaje, el cual varía con la resistividad que genera la humedad o sequía del suelo, cabe resaltar que de manera digital se debe sintonizar con el potenciómetro, el umbral que se desea poner y es recomendable utilizar un segundo dispositivo para su calibración, su protocolo de comunicación es 0-5 voltios para ambos casos (digital-análogo)[31], [32].

### **3.5.6 Sensor de luminosidad -fotorresistencia LDR GL55**

Formado por sulfuro de cadmio este dispositivo permite hacer mediciones de luz según su intensidad, sus siglas LDR traducen resistor dependiente de luz, de tal forma que este equipo al recibir la luz solar, algunos de sus fotones se absorben generando un flujo de electrones en la banda de conducción, la cual genera una disminución resistiva de tipo inversamente proporcional a la luz recibida[33].

## **3.6 Software**

De acuerdo con la definición de la RAE, software hace referencia a un conjunto de programas, instrucciones y reglas informáticas para ejecutar tareas en un computador[34]. A continuación, se describe el software que se va a usar para el desarrollo del proyecto.

### **3.6.1 Power BI**

Power BI es un software de Microsoft, permite crear informes y paneles, producto de modelizar y analizar datos. Power BI permite conectar varias fuentes de datos (Excel, SAP, SQL Server, entre otros), además de conectar y acceder a todos sus sistemas de forma intuitiva [35].

### 3.6.2 Python

Python es un lenguaje de programación que hace hincapié en combinar un poder notable con una sintaxis clara, soporta entre otras cosas programación orientada a objetos. Es un lenguaje multiplataforma (se puede trabajar en más de un sistema operativo) y posee una licencia de código abierto. [36].

Python posee un gran abanico de librerías para manejo de datos, data mining y machine learning, dentro de las cuales se encuentran:

#### 3.6.2.1 Pandas

Considerado un paquete de Python, es el encargado del manejo estructural y análisis especializado de datos, depende de la librería Numpy, el cual permite generar arreglos matriciales complejos. Dispone de formatos básicos de trabajo como CSV, Excel, bases de datos SQL entre otros. Contiene 3 formas o tipos de trabajo; en primer lugar están las *Series*, que básicamente se comporta como una lista con etiquetas de filas de una única dimensión; en segundo lugar se encuentran los *DataFrames*, siendo más elaborado con 2 dimensiones se compone de tablas matriciales utilizando filas y columnas y por último se encuentran los *Paneles*, que son considerados una sucesión de tablas superpuestas entre ellas tomando forma de cubo o tercera dimensión[37].

#### 3.6.2.1 Numpy

Numpy es un paquete de Python que permite crear matrices y vectores en arreglos multidimensionales, así como realizar operaciones de matemáticas avanzadas (eje: Fourier, algebra lineal, simulación aleatoria, entre otras). Es altamente utilizado en la computación científica por su velocidad, facilidad a la hora de implementar avanzadas operaciones matemáticas en presencia de grandes cantidades de datos[38].

#### 3.6.2.2 Scikit-Learn

Herramienta básica utilizada para la predicción de resultados de manera eficiente, funciona de manera articulada para otros contextos, contiene la mayoría de módulos de algoritmos de aprendizaje supervisado; modelos lineales generalizados, clustering, reducción dimensional, regresión, vectores de soporte, árbol de decisiones entre otros, incluso su versatilidad le permite trabajar con otras aplicaciones como Pandas o PyBrain[39].

#### 3.6.2.3 Pickle

Es un módulo implementado para serializar y deserializar una estructura de objetos Python. Pickle realiza una operación para el almacenamiento de los datos manejados, donde convierte la información de objetos creados en una secuencia binaria, permitiendo así transferencia a través de la red, a este proceso se le conoce como serialización o pickling de objetos, este luego debe ser revertido para obtener nuevamente la información original, a dicho proceso se le conoce como deserialización o unpickling[40].

### **3.7 Minería de datos (Data mining)**

Básicamente es un concepto innovador que estudia los métodos y algoritmos, que permiten la extracción de la información de manera automática y sintetizada, que permite denotar y relacionar características escondidas en dicha base de datos. Esto se realiza con el fin de obtener predicciones futuras mediante el análisis de los mismos. La minería de datos usa como herramienta básica las diversas técnicas de estadísticas, para deducir patrones que conlleven a hipótesis valederas. Es también importante el histórico que pueda generar la información debido que sirve para posteriores comparaciones o simplemente deducir un comportamiento con anticipación[41].

### **3.8 Machine Learning**

Machine Learning o aprendizaje de máquina es el campo de estudio donde los computadores pueden aprender de los datos que les proveen, con el fin de ejecutar determinada instrucción o set de instrucciones, sin que estas estén explícitamente codificadas.

A menudo cuando se tienen grandes cantidades de datos, como pueden ser la información de numerosos clientes de un supermercado y los artículos que adquirieron, a simple vista pueden parecer aleatorios, pero con el uso de herramientas de Machine Learning se pueden llegar a detectar patrones o hábitos que se repiten bajo ciertas condiciones – como puede ser artículos que tienen alta demanda en cierta temporada – .Estos patrones permiten, ya sea entender el proceso o hacer predicciones (asumiendo que los patrones vistos en el pasado, se repetirán en el futuro).

Existen tres tipos de división de algoritmos de Machine Learning cruciales para un entendimiento básico del proyecto, y del campo del Machine Learning en sí mismo, estos son: el aprendizaje no supervisado, el aprendizaje semisupervisado y el aprendizaje supervisado. Estas divisiones están basadas en el tipo de datos de los cuales se va a alimentar el modelo.

El aprendizaje no supervisado o unsupervised learning comprende una serie de algoritmos que son alimentados únicamente con datos de entrada (también llamados variables de entrada o predictores) pero sin información acerca de la variable o variables de salida, también llamadas target[42]. Un ejemplo de estos algoritmos se puede observar en un sistema para fraccionar los pacientes atendidos en un hospital en grupos homogéneos, pero sin saber qué grupos se desean obtener. Por otra parte, en el aprendizaje semisupervisado, se encuentran algoritmos donde se mezclan los datos que alimentan a los no supervisados y supervisados, teniendo variables de entrada etiquetadas y no etiquetadas con la variable de salida, dentro de los datos[43], [44].

### **3.8.1 Aprendizaje supervisado**

El aprendizaje supervisado o Supervised Learning comprende una serie de algoritmos donde se tiene un set de datos en el que se conocen los predictores y la variable de salida. Las primeras van etiquetadas con las segundas, y a partir de esto el modelo de Machine Learning establece una relación de qué características o parámetros de entrada pueden pertenecer a una etiqueta o parámetro de salida[45]. Un ejemplo puede ser un modelo alimentado por una base de datos donde se guarde información del tamaño, peso y color de limones, naranjas, limas y manzanas, donde en dichos datos se especifique a qué fruta pertenece cada conjunto de características

Los algoritmos de aprendizaje supervisado tienen varias clasificaciones, donde las dos más conocidas y de las que se va a hacer mención, son los modelos de clasificación y los modelos de regresión

Los modelos de regresión son algoritmos de aprendizaje supervisado, que son implementados cuando se desea predecir un valor cuantitativo, un ejemplo es la predicción del precio de un automóvil a partir de sus características (kilometraje, modelo, marca, etc.

Los modelos de clasificación son algoritmos que van destinados a la predicción de datos cualitativos, un ejemplo puede ser un sistema para detectar si un correo electrónico es spam o no[46].

### **3.8.2 Modelos de Entrenamiento**

Dentro de los modelos de clasificación se encuentran los árboles de decisión y los bosques aleatorios (también pueden ser usados como modelos de regresión). Estos algoritmos son los que se proceden a implementar en el desarrollo del proyecto.

#### **3.8.2.1 Árboles de decisión**

El árbol de decisión es un algoritmo que se usa para predecir las etiquetas de los datos que son ingresados. Su funcionamiento se basa en la clasificación de características que se pueden extraer de la información suministrada. Dicha clasificación inicia en un nodo, llamado nodo raíz, cuya salida o respuesta apunta a otro nodo, llamado nodo secundario o nodo hijo (puede existir un nodo hijo por cada posible clasificación que exista), este puede tener a su vez otros nodos secundarios, y estos a otros, hasta llegar a uno sin nodos hijos (a este se le conoce con el nombre de nodo hoja) estableciendo así una jerarquía que toma la forma de árbol. Cada nodo hoja tiene asignado un ítem o etiqueta de la variable de salida, o la posibilidad que el elemento clasificado por este nodo pertenezca a uno[47].

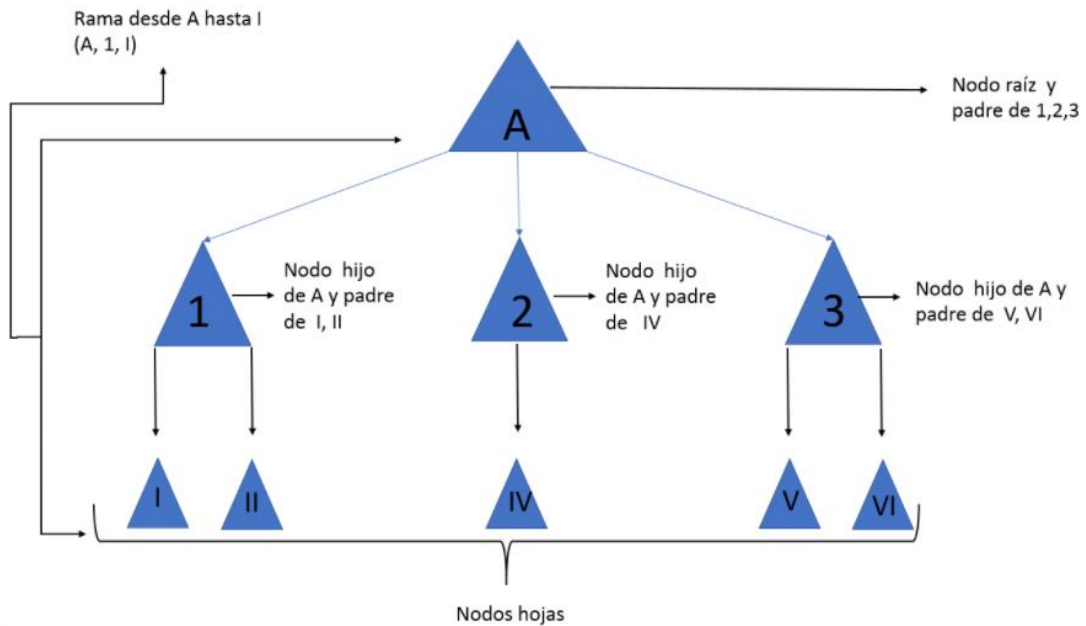


Figura 6. Estructura de un árbol de decisión (Fuente:Autor).

Los árboles de decisión intentan identificar una variable y establecer una clasificación. La métrica generalmente utilizada para dicha categorización, es la homogeneidad, que tiene la finalidad de dar una distribución de la variable objetivo (variable que se desea predecir) tan pareja como sea posible, es decir, la homogeneidad busca que los valores similares que se hacen presentes en esta variable sean agrupados.

Con el fin de hallar la mejor forma de realizar la clasificación y lograr el mayor rendimiento posible del árbol de decisión, existen distintas técnicas o algoritmos. A continuación, se proceden a explicar los más representativos o importantes.

### Algoritmos para crear árboles de clasificación

- Algoritmo ID3

Se basa en el concepto de la ganancia de información basado en la entropía, que está dada por la fórmula:

$$H(S) = \sum_{i=1}^n -P_i * \log_2(P_i)$$

Donde n son las categorías para una variable a predecir, y  $P_i$  es la probabilidad de que sea dicha variable i[46]

El primer paso del algoritmo ID3 es, basados en la variable objetivo, calcular la entropía inicial del sistema, luego se calcula la ganancia de información para cada variable candidata para un nodo y se selecciona como nodo de decisión la que provee más ganancia de información. Se realiza el cálculo de la ganancia de información para cada rama o valor de cada nodo o variable candidata, para obtener un nodo hoja. Se comprueba si el nodo hoja clasifica de forma adecuada los datos. De no

ser así, se vuelve a calcular la ganancia de información para continuar ramificando[48].

- Algoritmo CART

El algoritmo CART funciona tanto para variables de entradas categóricas como continuas. En variables continuas trabaja hallando el promedio de dos valores adyacentes pertenecientes a la variable de entrada, enviando los menores de este promedio a un nodo y los mayores a otro.

En variables categóricas con varios niveles o categorías, funciona enviando un subconjunto de niveles a un nodo, y dejando uno en otro nodo.

Una vez hallados los nodos, el algoritmo CART se encarga de calcular las probabilidades de que determinado elemento pertenezca o no a cada nodo. Según sea el caso, puede utilizar el criterio de entropía o el Criterio de Gini[49][46].

- *Criterio de Gini (Gini Impurity)*

Es una medida que se utiliza para determinar cómo las características de los datos deben ser divididas, con el fin de formar nodos para constituir un árbol de decisión. Concretamente esta medida es un valor entre 0 y 0.5, que señala la probabilidad de que nuevos datos sean clasificados de forma errónea, si se les asigna una etiqueta de clase aleatoria acorde con la distribución de los datos. La etiqueta con menor valor determinará la mejor característica para dividir el nodo actual. Es comúnmente utilizado si la variable objetivo es binaria.

Matemáticamente, el índice de Gini para un dataset con k clases está representado por la fórmula:

$$Gini(D) = 1 - \sum_{i=1}^k P_i^2$$

Donde  $P_i$  es la probabilidad de una determinada clasificación  $i$ [46][43].

- Detector Automático de la Interacción con Chi Cuadrado (CHAID)

Utiliza Chi Cuadrado para encontrar divisiones óptimas. Esto se hace utilizando tablas de tabulación cruzada entre las variables de entrada y los resultados, luego se hace la comprobación de independencia de chi cuadrado con el objetivo de corroborar su significación. CHAID se encarga de seleccionar la variable de entrada con mayor relevancia.[50].

Dos problemas comunes que se suelen encontrar a la hora de trabajar con árboles de decisión y otros algoritmos de machine learning son el overfitting, y el underfitting. Se denomina overfitting a la situación donde el modelo se ajusta demasiado bien a los datos proporcionados para el entrenamiento, pero presenta un rendimiento pobre con los datos de test, o datos que el modelo no ha visto

antes; este problema surge porque el modelo está captando muy bien los datos de entrenamiento, pero no es capaz de extrapolar las relaciones existentes a otros datos. Por otro lado, el underfitting es la situación en la que el rendimiento del modelo con los datos que se le han proporcionado para su entrenamiento es pobre, esto se debe a que el modelo es incapaz de detectar las relaciones que pueden existir entre los datos de entrada y los datos de salida. Mientras que la solución para el underfitting suele ser tomar buenas métricas de rendimiento y tener una buena cantidad de datos de entrenamiento, la solución para el overfitting suele ser algo más específica según el modelo de machine learning que se implemente. En el caso de los árboles de decisión, la solución suele ser la poda del árbol[43], [51].

El proceso de poda de un árbol consiste en la simplificación de este, bien sea estableciendo un límite mínimo de datos de entrada a un nodo que, de no cumplirse, el nodo no se dividirá más (a esto se le conoce como prepruning); o bien dejando crecer el árbol sin ningún tipo de restricciones, para luego reemplazar las ramas por nodos hoja, donde dicho reemplazo no afecte las métricas de rendimiento del árbol y cada nodo hoja esté etiquetado con las variables que la rama abarcaba (a esto se le conoce como postpruning)[43].

### **3.8.2.2 Bosques aleatorios (*random forest*)**

Se denominan bosques aleatorios a la unión de varios árboles de decisión, estos votan por una clase resultando elegida la que cuenta con más votos. Cada árbol de decisión perteneciente al bosque se caracteriza por tener baja correlación con los demás, lo que permite que este modelo sea más resistente al ruido producido por los datos (al cual los árboles de decisión son bastante más sensibles) y al overfitting[52].

La baja correlación existente entre los árboles se produce debido a que en los bosques aleatorios se implementa una técnica llamada bagging, en la que se toma el conjunto de datos de entrenamiento, se divide en subconjuntos y se deja crecer un árbol con cada subconjunto. Estos subconjuntos usados pueden tener datos repetidos entre sí, es decir, datos del conjunto de entrenamiento usados en más de un subconjunto[53].

## **3.9 Métricas de rendimiento**

Para medir el comportamiento de los modelos de Machine Learning, se deciden emplear las siguientes métricas.

### **3.9.1 Matriz de confusión**

La matriz de confusión es una herramienta ampliamente utilizada para medir la cantidad de aciertos y errores que tiene un modelo de aprendizaje supervisado. En su forma básica está compuesta de la siguiente manera

		Predicción	
		Positivos	Negativos
Observación	Positivos	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	Negativos	Falsos Positivos (FP)	Verdaderos Negativos (VN)

Figura 7. Estructura de una matriz de confusión

(Fuente: [RPubs - Matriz de Confusión - Evaluación de modelos de predicción](#))

Donde cada columna de la matriz representa las predicciones hechas para cada clase y cada fila representa a la clase real; la sección de Verdaderos Positivos indica la cantidad de valores correspondientes a la clase, que fueron correctamente clasificados; la sección de Verdaderos Negativos indica la cantidad de valores no correspondientes a la clase, que fueron correctamente clasificados; la sección de Falsos Negativos indica la cantidad de valores correspondientes a la clase que fueron clasificados como no pertenecientes a esta; y la sección de Falsos Positivos indica la cantidad de valores no pertenecientes a la clase, que fueron clasificados como si pertenecieran a esta. [46]

### 3.9.2 Precisión

En machine Learning hace referencia a la cantidad de valores pertenecientes a la clase clasificada correctamente, sus valores van, desde 0 a 1, siendo 0 el indicador de que el modelo no es nada preciso, y 1 el indicador de que el modelo siempre que escoja bien una clase, esta va a ser confiable[46]. Se halla mediante la fórmula:

$$\text{Precisión: } \frac{\text{Verdaderos positivos}}{\text{Verdaderos positivos} + \text{Falsos Positivos}}$$

### 3.9.3 Recall

El recall o sensibilidad, en Machine Learning indica la proporción de casos identificados correctamente como pertenecientes a la clase, que el modelo fue capaz de realizar, sus valores van, desde 0 a 1, siendo 1 el indicador de que el modelo tiene el recall más alto[46]. Se halla mediante la fórmula:

$$\text{Recall: } \frac{\text{Verdaderos positivos}}{\text{Verdaderos positivos} + \text{Falsos Negativos}}$$

### 3.9.4 f1-score

El f1-score es una medida que abarca tanto la precisión como el recall, siendo la media armónica entre estas dos medidas, sus valores van, desde 0 a 1, siendo 1 el indicador de mayor f1-score[46]. Está dado por la fórmula:

$$f1 - score: 2x \frac{Recall \times Precision}{Recall + Precision}$$

### 3.10 Normalización de datos

En el campo de la estadística, la normalización de datos es una técnica que se encarga de cambiar de escala los datos pertenecientes a una variable, con el fin de poder ser comparados con otros. En Machine Learning la normalización de datos es muchas veces requerida, ya que el comportamiento del modelo puede cambiar de forma significativa si la distribución de una o más de sus variables de entrada no tienen una distribución Gaussiana de media cero y varianza unitaria[54]. La estandarización se lleva a cabo mediante la fórmula:

$$Z = \frac{(x - u)}{s}$$

Donde x es la muestra, u es la media de las muestras (cero, si se requiere una media igual a cero) y s es la desviación estándar de las muestras[55].

### 3.11 Test de rasgos de importancia

También conocido como importancia de Gini, se encarga de calcular la importancia de cada característica (variable de entrada), y se suele tomar como la suma de la disminución de la impureza del nodo (o disminución de la entropía), promediada en todos los árboles del árbol de decisión[56].

### 3.12 Crossvalidation

Crossvalidation (también conocido como k-fold crossvalidation) es un proceso de resampléo que se caracteriza principalmente por ser menos propenso a los sesgos que puedan presentar los datos, y se trata de dividir un dataset en un número (k) determinado de grupos, una vez se tienen estos grupos, el modelo se procede a entrenar con una parte de dicho subgrupo y la parte restante se utiliza para realizar pruebas; este proceso se repite para cada uno de los grupos y al final se saca una media aritmética de las métricas de rendimiento. Este proceso se puede ver en la figura 8[57].

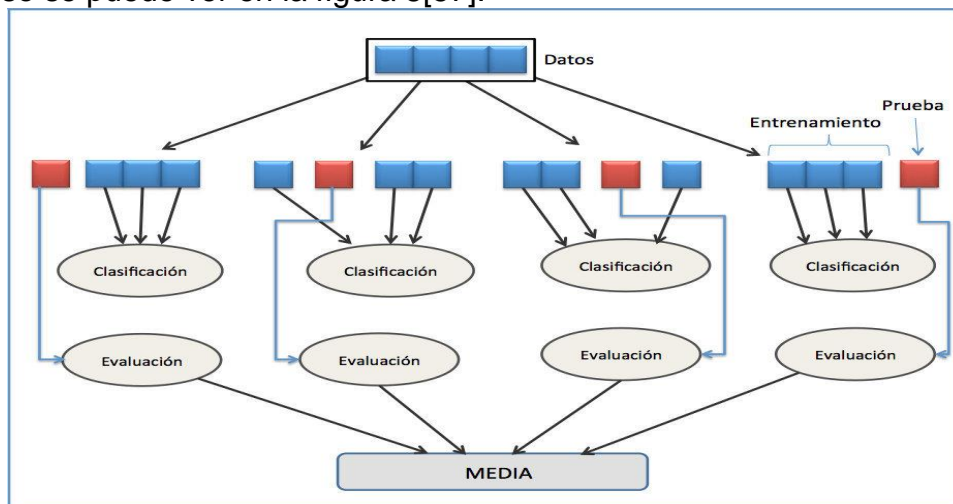


Figura 8. Proceso de aplicación de k-fold crossvalidation (fuente: [Esquema castellà - Validación cruzada - Wikipedia, la enciclopedia libre](#))

## **CAPITULO 4**

### **METODOLOGÍA**

A lo largo de este capítulo se proceden a explicar a detalle los pasos, inconvenientes y las decisiones que llevaron a desarrollar el experimento de la forma en la que se presenta, conexiones de hardware, uso de software entre otros. Para este capítulo fue crucial la asesoría y las recomendaciones dadas por la profesora Sandra Cuervo, docente de la Universidad Pontificia Bolivariana, quien tiene experiencia previa con el manejo de la yerbabuena.

#### **4.1. Experimento**

El experimento se aborda desde cinco puntos principales para su elaboración: fase exploratoria, el diseño y la implementación de la estructura física, la adquisición de los datos y puesta en funcionamiento, creación de algoritmos, limpieza y manipulación con técnicas de machine learning y finalmente presentación de resultados. Es importante mencionar que cada uno de estos puntos tuvo varias iteraciones antes de llegar a las decisiones finales, ya que se presentaron inconvenientes que obligaron a realizar cambios sobre la marcha, tales como el cambio de ubicación de los sujetos de prueba, cambio en tecnologías y hardware, entre otros, que se explican a medida que se desarrolla este capítulo del libro.

##### **4.1.1. Fase exploratoria**

En la fase exploratoria se estudian los antecedentes mencionados en el capítulo dos de este libro, se estudia la bibliografía que se considera pertinente acerca de la yerbabuena, se combinan esta información con las recomendaciones dadas por la profesora Sandra Cuervo, las características y limitaciones de los dispositivos que se decidió utilizar, y datos recopilados mediante pruebas empíricas, para obtener las primeras condiciones y características de diseño. Teniendo en cuenta lo anteriormente dicho, durante este punto se encuentra la existencia de un solo microclima y se opta por el uso de dos sujetos de experimento, cada sujeto estando compuesto por tres plantas y dos masetas de 78 centímetros de largo, 18 centímetros de alto y 20 centímetros de profundidad cada una. También se encontró que, para las condiciones del microclima, una posible cantidad de agua a suministrar, con el fin de mantener las plantas saludables, era de 800 mililitros por sujeto de experimento, en intervalos de tiempo de 2 días.

##### **4.1.2. Diseño e implementación de la estructura física**

En la instalación de hardware se proceden a realizar las primeras conexiones y a verificar si las primeras ideas obtenidas en la fase exploratoria eran acertadas, qué ideas se debían modificar y cuales se podían mantener.

Durante la instalación de hardware inicialmente se pensó utilizar terrenos de la universidad para sacar provecho a la existencia de plantas de yerbabuena en el campus, que son debidamente mantenidas, pero debido a la facilidad y

comodidad en la realización del experimento, y que se requería dejar los elementos a la intemperie, se optó por cambiar el sitio donde se realiza el experimento al sitio de residencia de uno de los estudiantes que desarrollan el proyecto. En este punto y debido a la limitación de número de sensores que se podían conectar al UCM 316 y MPC 134 se optó por el uso de dos sensores de humedad de suelo, un sensor de humedad y un sensor inalámbrico IAQM-THCO2.

Antes de ubicar los sujetos de prueba, se realizaron experimentos previos, obteniendo como resultado que la luz solar podía incidir directamente en la planta pero durante periodos no muy prolongados de tiempo; también se detectó que los sujetos de pruebas presentan un crecimiento radicular veloz y abundante, esto generaba la autodestrucción de la planta ya que, a pesar que se agregaba abono a las plantas estas terminaban sucumbiendo al no tener como expandir sus raíces, Este inconveniente se solucionó parcialmente cambiando a masetas con mayor profundidad.

La versión definitiva de la arquitectura física se puede observar en el esquema mostrado en la figura 9, en el cual se observa el UCM 316 actuando como nodo maestro, conectado al sensor de luminosidad de forma análoga (0V a 5V), al computador donde se va a trabajar con los datos a través de cable ethernet, y al MPC 134 a través de protocolo modbus. El MPC 134 funciona como nodo esclavo y tiene conexión con los dos sensores de humedad de tierra de forma análoga (0V a 5V), y con el DM 124 a través de protocolo modbus. El DM 134 está conectado al sensor IAQM-THCO2 mediante protocolo Zigbee.

Tanto el UCM 316, el MPC 134, como la conexión entre el DM 124 y el sensor IAQM-THCO2 requieren ser configurados vía software.

#### **4.1.2.1 Configuración de dispositivos**

Con el fin de configurar la red de sensores, el primer paso que se debe realizar es la configuración del DM-124 y el IAQM-THCO2. Esta configuración se realiza por medio del UD 1000. Como prerrequisito se deben instalar o tener instalados los programas WM Config Tool y Microsoft.NET Framework 3.5 o más reciente. También se deben instalar los drivers del UD 1000, luego se procede a su programación.

Para realizar la programación del UD 1000 se debe instalar el programa FT\_Prog 2.4.2. Una vez instalado se conecta el UD 1000, se abre el FT\_Prog y se da click en el ícono de la lupa, como se observa en la siguiente figura, luego se procede a dar click derecho en Device > apply Template > From file, se selecciona el archivo UD 1000, se presiona CTRL + P o en su defecto el ícono al lado de la lupa y se da click en Aceptar para completar la instalación. Si la programación del dispositivo se realizó de forma correcta, aparecerá el nombre AS UD 1000 en la parte superior derecha

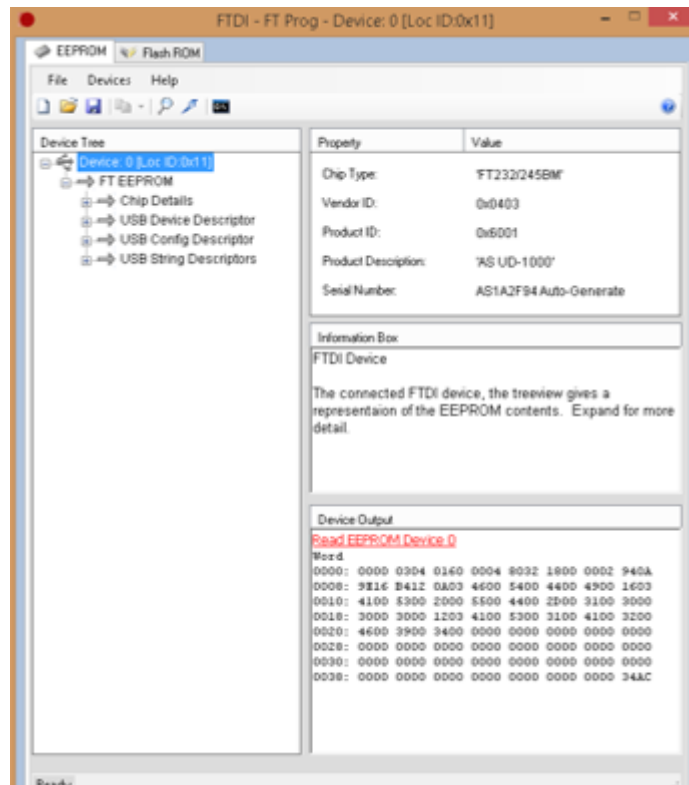


Figura 9. FT\_Prog  
(fuente: Autor)

La siguiente parte de la configuración del UD 1000 requiere el uso del sistema operativo TinyOS, que es de código abierto y está basado en componentes para redes de sensores inalámbricos. Se puede realizar la instalación de este sistema, siguiendo las instrucciones encontradas en [Use TinyOS in a virtual machine - TinyOS Wiki \(stanford.edu\)](http://TinyOS Wiki (stanford.edu)) (se recomienda hacerlo con Ubuntu).

Una vez instalado TinyOS, se debe configurar el puerto COM donde irá el UD 1000, para esto, y si se hizo desde Linux, se debe copiar el archivo llamado RadioConfigurator.ihex, abrir el terminal de la consola, se accede a carpeta BaseStation usando la introduciendo la línea "cd /home/advanticsys/Desktop/TinyOS/tinyos-2.1.2/apps/BaseStation", se crea una carpeta bajo el nombre telosb usando la línea "make telosb", se usa la línea "Motelist" para observar los dispositivos.

```

Player
Applications Places
advanticsys@ubuntu: ~/Desktop/TinyOS/tinyos-2.1.2/apps/BaseStation
File Edit View Search Terminal Help
TinyOS version: TinyOS-2.1.2
advanticsys@ubuntu:~/Desktop/TinyOS/tinyos-2.1.2/apps/BaseStation$ cd /home/advanticsys/Desktop/TinyOS/tinyos-2.1.2/apps/BaseStation
advanticsys@ubuntu:~/Desktop/TinyOS/tinyos-2.1.2/apps/BaseStation$ make telosb
mkdir -p build/telosb
compiling BaseStationC to a telosb binary
ncc -o build/telosb/main.exe -Os -fnesc-separator=__ -Wall -Wshadow -Wnesc-all -target=telosb -fnesc-cfile=build/telosb/app.c -board= -DDEFINED_TOS_AM_GROUP=0x22 -DCC2420_NO_ACKNOWLEDGEMENTS -DCC2420_NO_ADDRESS_RECOGNITION -DTASKLET_IS_TASK -DIDENT_APPNAME="BaseStationC" -DIDENT_USERNAME="advanticsys" -DIDENT_HOSTNAME="ubuntu" -DIDENT_USERHASH=0x5d565464L -DIDENT_TIMESTAMP=0x58a36546L -DIDENT_UIDHASH=0xf3ef83eeL BaseStationC.nc -ln /home/advanticsys/TinyOS/tinyos-2.1.2/tos/chips/cc2420/lpl/DummyLplC.nc:39:2: warning: #warning "*** LOW POWER COMMUNICATIONS DISABLED ***"
compiled BaseStationC to build/telosb/main.exe
14438 bytes in ROM
1818 bytes in RAM
msp430-objcopy --output-target=ihex build/telosb/main.exe build/telosb/main.ihex
writing TOS image
advanticsys@ubuntu:~/Desktop/TinyOS/tinyos-2.1.2/apps/BaseStation$ motelist
Reference Device Description
-----
AS18IAXZ /dev/ttyUSB0 FTDI AS UD-1000
advanticsys@ubuntu:~/Desktop/TinyOS/tinyos-2.1.2/apps/BaseStation$ make telosb i
ninstall bsl,/dev/ttyUSB0
mkdir -p build/telosb
compiling BaseStationC to a telosb binary
ncc -o build/telosb/main.exe -Os -fnesc-separator=__ -Wall -Wshadow -Wnesc-all -target=telosb -fnesc-cfile=build/telosb/app.c -board= -DDEFINED_TOS_AM_GROUP=0x22 -DCC2420_NO_ACKNOWLEDGEMENTS -DCC2420_NO_ADDRESS_RECOGNITION -DTASKLET_IS_TAS
K -DIDENT_APPNAME="BaseStationC"
advanticsys@ubuntu:~/Desktop/TinyOS/tinyos-2.1.2/apps/BaseStation

```

Figura 10. Pasos par configuración UD 1000 mediante consola (fuente: Autor)

Luego se procede a instalar el puerto COM utilizando la línea “make telosb install bsl,/dev/ttyUSB0”. Cabe aclarar que el fragmento de la línea /dev/ttyUSB0 cambia dependiendo del COM al cual esté conectado el UD 1000, por ejemplo, si este se encuentra conectado al COM2, el fragmento del comando sería “/dev/ttyUSB1”. Luego de instalar el puerto COM se copia el contenido del archivo RadioConfigurator.ihex y reemplazarlo en main.ihex en la carpeta build/telosb. Como paso final se procede a reinstalar el puerto COM usando la línea “make telosb reinstall bsl,/dev/ttyUSB0”

```

advanticsys@ubuntu:~/Desktop/TinyOS/tinyos-2.1.2/apps/BaseStation$ make telosb reinstall bsl,/dev/ttyUSB0
cp build/telosb/main.ihex build/telosb/main.ihex.out
installing telosb binary using bsl
tos-bsl --telosb -c /dev/ttyUSB0 -r -e -I -p build/telosb/main.ihex.out
MSP430 Bootstrap Loader Version: 1.39-goodfet-8
Mass Erase...
Transmit default password ...
Invoking BSL...
Transmit default password ...
Current bootstrap loader version: 1.61 (Device ID: f16c)
Changing baudrate to 38400 ...
Program ...
14582 bytes programmed.
Reset device ...
rm -f build/telosb/main.exe.out build/telosb/main.ihex.out
advanticsys@ubuntu:~/Desktop/TinyOS/tinyos-2.1.2/apps/BaseStation$

```

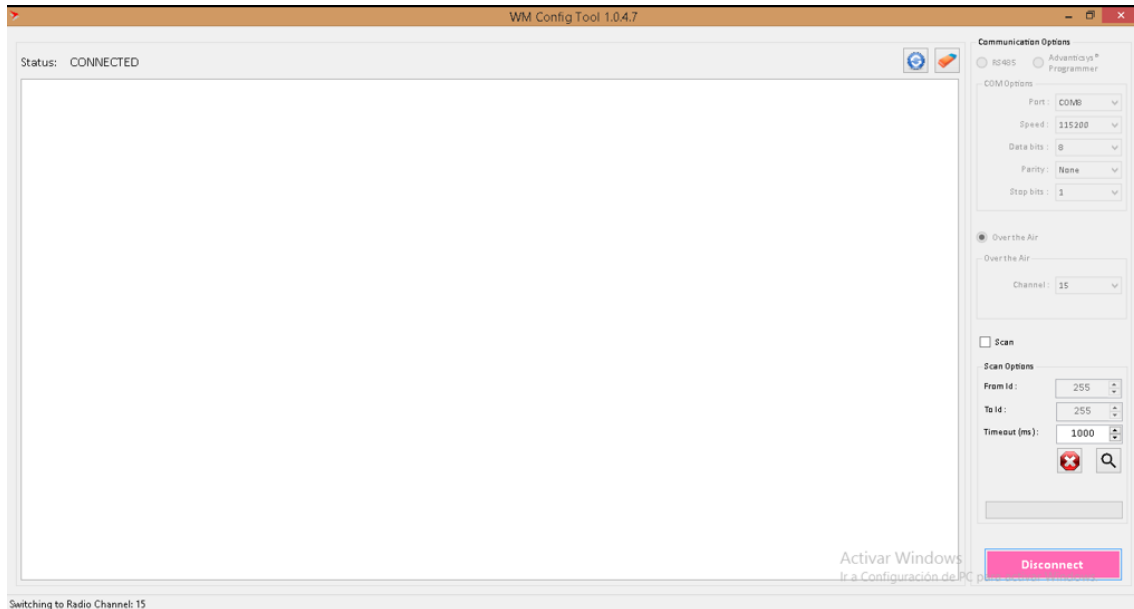
Figura 11. Visualización de la consola luego de terminar todo el proceso (fuente: Autor)

Una vez configurado el UD 1000 se proceden a configurar el IAQM-THCO2 y el DM 124.

### Configuración IAQM-THCO2

Como primer paso se debe detectar el dispositivo con el software WM config tool, para ello se conecta el UD 1000 al computador, y luego se abre el programa, una vez desplegado, WM Config Tool desplegará una ventana como la mostrada en

la siguiente figura. En esta ventana se muestran una serie de opciones en el costado derecho, donde se proceden a seleccionar el COM en el cual quedó configurado el UD 1000, luego la opción Over the Air ya que el dispositivo se enlaza vía inalámbrica, y para finalizar se selecciona Channel 15 ya que este es el número de canal que viene configurado por defecto para este sensor. Luego se da click al botón Connect.



*Figura 12. Ventana principal del WM Config Tool  
(fuente: Autor)*

Para finalmente poder detectar el sensor, se debe cerciorar que el Status mostrado en la parte superior izquierda de la ventana diga CONNECTED, en cuyo caso se procede a escanear dando click en el botón Scan, botón con una lupa como imagen.

Una vez detectado el IAQM-THCO2, se despliega un menú como el mostrado en la siguiente figura; en esta ventana se configuran los parámetros Modbus id, Radio id, Group id y Radio Channel. El Modbus id que fue asignado al sensor fue el 22 (este debe ser único en la red de sensores conectados), el Radio id asignado es el 22 (esta número debe ser único para todos los dispositivos conectados ya que es la identificación del dispositivo en la red de sensores), como Group id se asigna el número 33 y como Radio Channel se asigna el 15 (estos dos últimos parámetros sirven para identificar la red a la cual se va a conectar el dispositivo, por lo que deben ser los mismos valores para todos los dispositivos de la red de sensores).

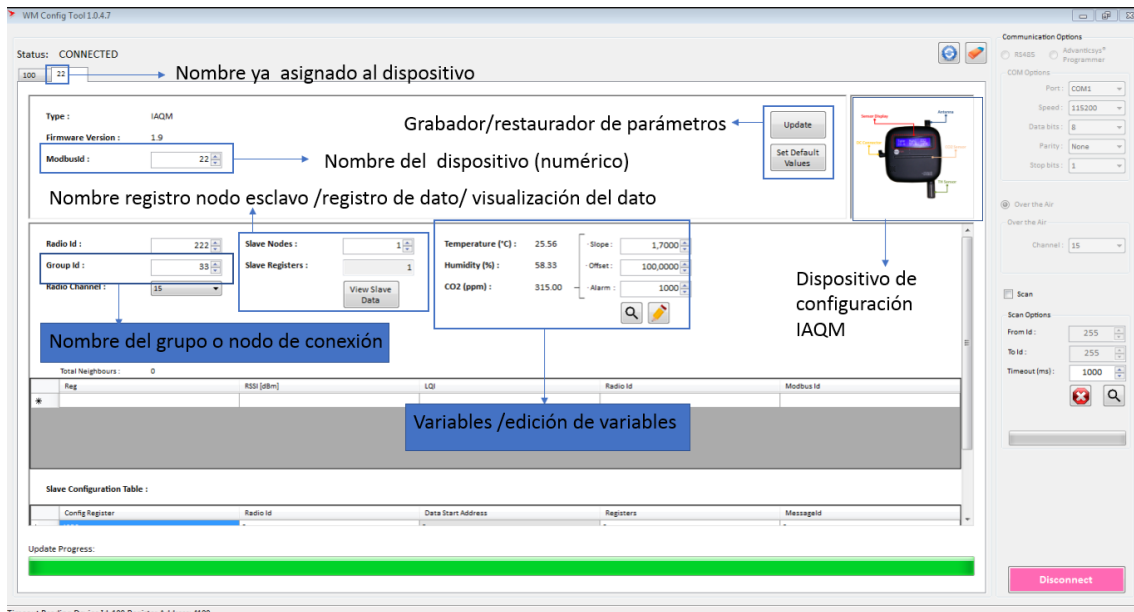


Figura 13. Programa de configuración WM Config Tool, utilizando UD1000 como intercomunicador.

(fuente:Autor)

Una opción sencilla para comprobar que el sensor IAQM-THCO2 que se está enlazando a la red está siendo reconocido correctamente, es mediante el chequeo de los campos Temperature, Humidity y CO<sub>2</sub> que corresponden a las variables que este sensor mide, y que al dar click en el botón scan que se encuentra en la parte inferior de los parámetros, estos se actualizan.

## Configuración DM 124

Al igual que con el IAQM-THCO2, para detectar el DM 124 se debe escanear con el WM Config Tool, bajo los mismos parámetros. Una vez encontrado el dispositivo, se despliega una ventana como la mostrada en la siguiente figura.

Se asigna un Modbus id de 100, Radio id de 101, y el mismo Group id y Radio Channel que a los demás dispositivos conectados a la red de sensores. Algo a tener en cuenta es que, si se desea que el DM 124 haga la función de coordinador de red, el Modbus id asignado debe ser un número múltiplo de 10.

Al asignarse el DM 124 como coordinador de red, se configuran también los parámetros Slave Nodes y Slave Registers. Slave nodes sirve para establecer la comunicación con los demás dispositivos inalámbricos de la red (El número asignado en esta opción debe corresponder con el número de dispositivos inalámbricos conectados a la red, aparte de este DM 124), por su parte Slave Registers permite elegir cuántos registros guardar en memoria.

Para finalizar, en los parámetros RS485, se configuran el Baud rate, Parity, Data bits y Stop bits; es importante tener presentes los valores configurados aquí, ya que deben ser los mismos para los demás elementos de la red.

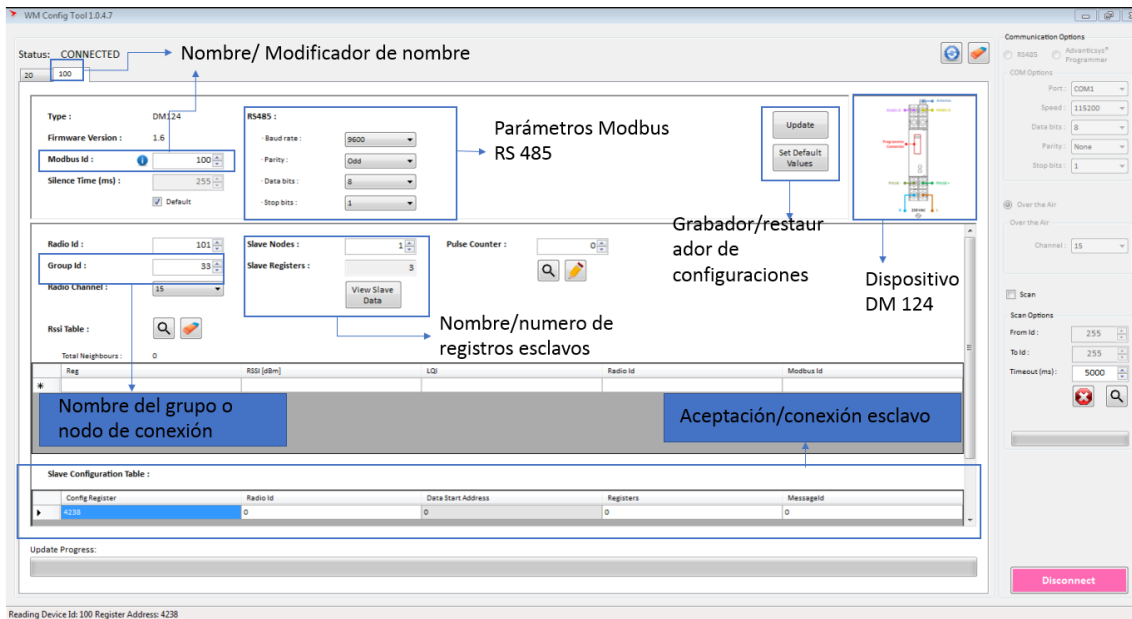


Figura 14. Configuración y parámetros del DM 124 nodo de conexiones.  
(fuente: Autor)

Al configurar las opciones de esclavos en el DM 124 se activan una serie de parámetros para los nodos esclavos habilitados, dichos parámetros se pueden ver en la siguiente figura.

Slave Configuration Table:

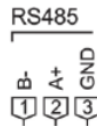
Config Register	Radio Id	Data Start Address	Registers	MessageId
4238	2222	4100	6	0

Figura 15. Parámetros de configuración para nodos esclavos  
(fuente: Autor)

Los parámetros a tener en cuenta son la opción Radio id, donde se coloca el Radio id del dispositivo que va a ir como esclavo y Registers donde se coloca la cantidad de registros que el DM 124 debe almacenar.

## Enlace y configuración entre DM 124, UCM 316, IAQM-THCO2 y MPC 134

Una característica física con la que cuentan tanto el UCM 316 como el MPC 134 es que no cuentan con antena de señal IEEE 802.15.4, por lo que se hace necesario el uso del DM 124 como conexión inalámbrica del MPC 134 a la red. Para llevarlo a cabo, se deben conectar ambos dispositivos vía UART RS485. El MPC 134 posee otro puerto UART RS485, mediante el cual se conecta al UCM 316.



Puerto RS485

Figura 16. Esquema del puerto UART  
(fuente: Advanticsys)

Para realizar las configuraciones tanto del UCM 316 como del MPC 134, estos deben conectarse vía USB a un computador. En el caso del MPC 134, una vez conectado, se abre su software y se da click en Communication > Modbus Devices > Configuring Modbus Devices > Modbus RTU Client, como se muestra en la siguiente figura. Se procede a habilitar la opción Enabled en las opciones del Modbus RTU Client1, se asigna un periodo de lectura de 5 y un número de peticiones de 1. En la opción Address se procede a poner el Modbus id asignado al sensor IAQM-THCO2, se le asigna como valor de función 3 ya que es el valor para realizar lecturas analógicas, en Amount of registers se coloca un valor de 3 ya que este es el número de registros necesarios para tomar los valores de las tres variables medidas por el sensor, y en MPC reg. Index se coloca la posición de memoria en la cual se van a guardar los datos; una vez asignados los valores se da click en el botón Set.

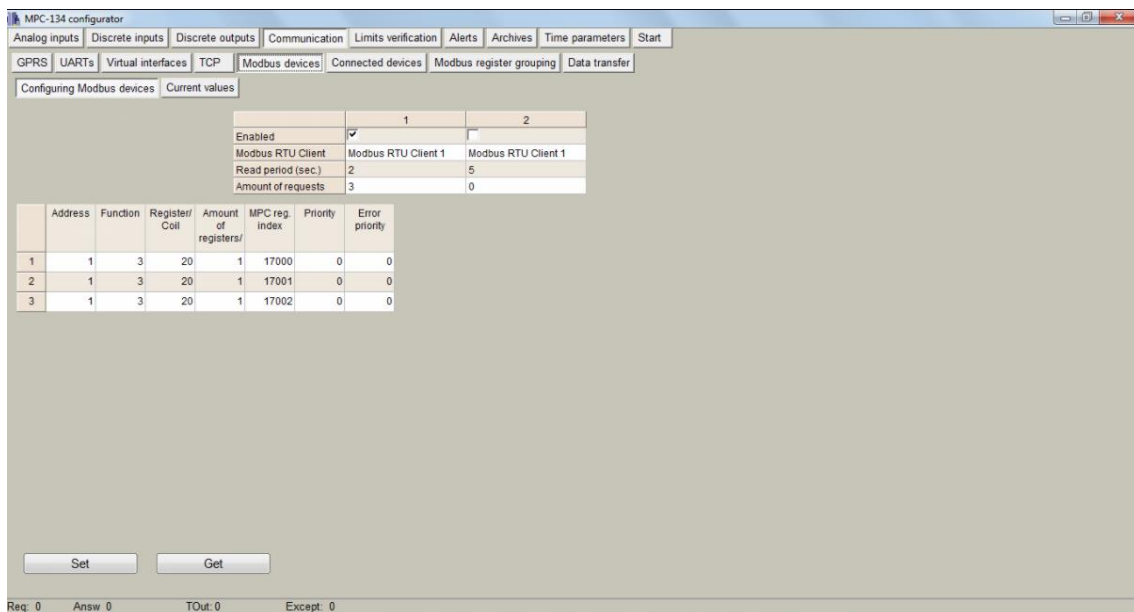


Figura 17. Configuración 2 MPC 134 ubicación y tiempo de lectura para los datos  
(fuente: Autor)

Luego se proceden a configurar las conexiones con los puertos UART RS485, el primero se encuentra conectado al DM 124 y el segundo al UCM 316, para ellos primero se accede a Communication > Virtual interfaces, donde se configura como Modbus RTU Client al puerto donde se conecta el DM 124, se da click en la opción Enabled y se da click en el botón Set, luego se configura como Modbus

RTU Server el puerto que se conecta con el UCM 316, se selecciona la opción Enabled y se da click en el botón Set

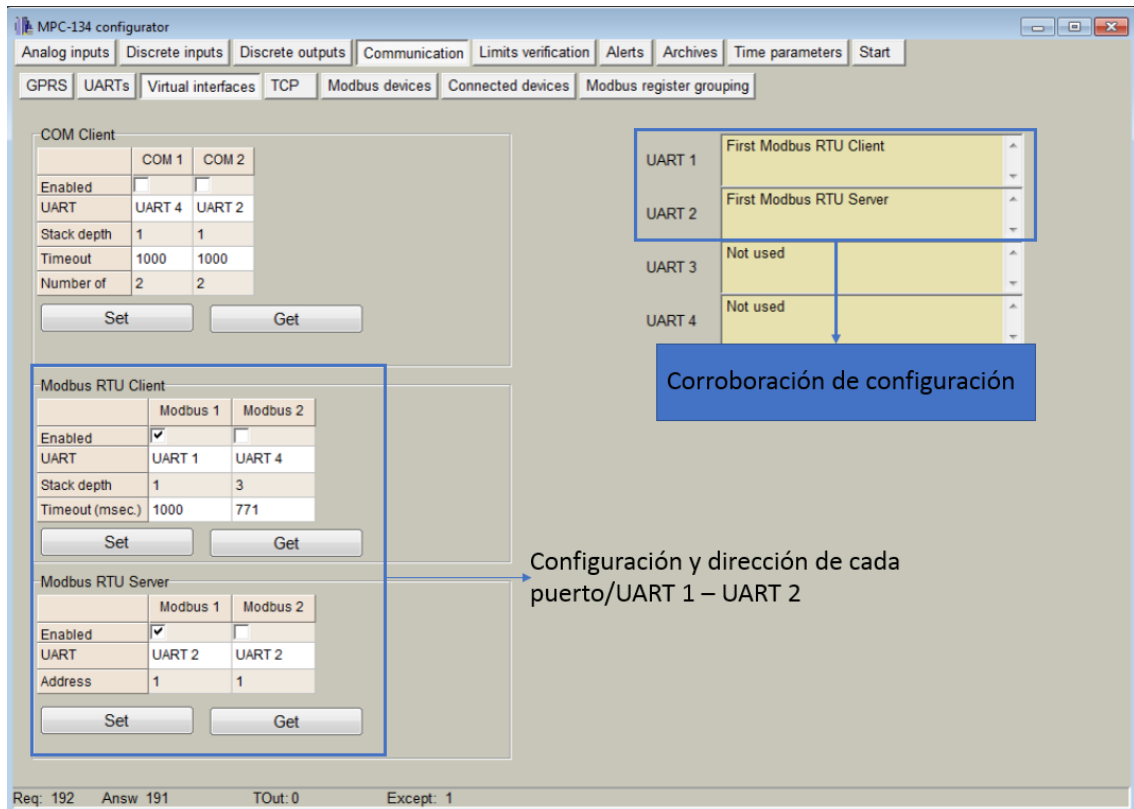


Figura 18. Configuración 3 MPC 134 asignación de trabajo en cada uno de los puestos UART's  
(fuente: Autor)

Luego se accede a Communication > UARTs y en las opciones que se muestran, se configuran las mismas con las que se configuró el DM 124 previamente, en seguida se procede a dar click en el botón Set.

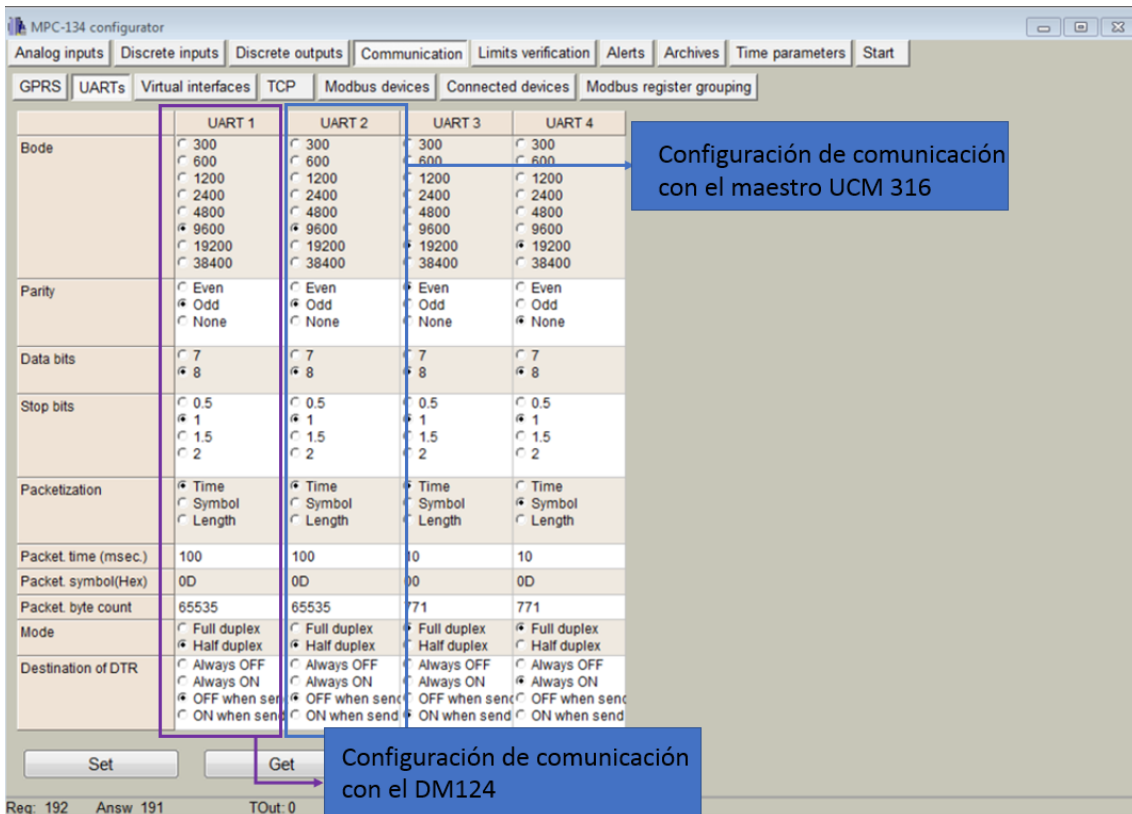


Figura 19. Configuración de comunicación de puertos UART RS485 para el MPC 134

Al MPC 134 van también enlazados los sensores de humedad FC-28; estos se conectan a las entradas de voltaje análogas. Una vez conectados, en el software del MPC 134 se da click en Analog inputs > Configuration; en las opciones que se muestran en la ventana se habilita la opción Enabled de los puertos a los que fueron conectados los sensores y se procede a dar click en el botón Set.

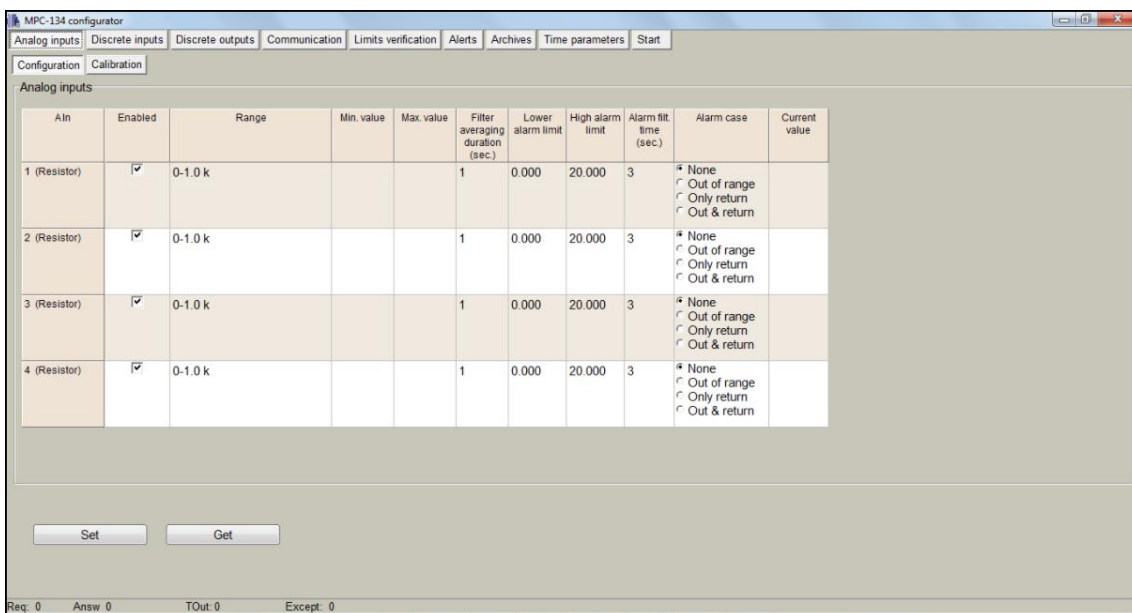


Figura 20. Configuración de sensores análogos en MPC 134

Description	Modbus Register list					
	AI_1	AI_2	AI_3	AI_4	Data format	Type
Analog input use (0- disabled, 1 – enabled)	4524	4525	4526	4527	Char	R/W
Measuring type (0-voltage, 1-current, 2-resistance)	4528	4529	4530	4531	Char	R/W
Min recalculated value Float[4]	4600-4601	4602-4603	4604-4605	4606-4607	Float	R/W
Max recalculated value Float[4]	4608-4609	4610-4611	4612-4613	4614-4615	Float	R/W
Physical value (Float[4])	4504-4505	4506-4507	4508-4509	4510-4511	Float	R
Recalculated value (Float[4])	4512-4513	4514-4515	4516-4517	4518-4519	Float	R

Figura 21. Registro de entradas analógicas MPC 134  
(Fuente: Advanticsys)

Para el UCM 316 se abre su software y se da click a Communication > Modbus Devices > Configuring Modbus Devices > Modbus RTU Client; en esta ventana se selecciona Modbus RTU Client 1, se habilita la opción Enabled, se asigna un período de lectura (en segundos) y en Amount of request se asigna un valor de 2. En el parámetro Adress se pone el número de identificación del dispositivo esclavo, en Funtion de acuerdo al manual se pone el valor 3 ya que es el valor para realizar lecturas analógicas, en Register/Coil se coloca la dirección de donde se va a extraer la información, en el caso de los datos del IAQM-THCO2 se coloca el número 17000 ya que es la posición de memoria donde fue guardado el primer registro, para los sensores FC-28, se colocan los números 4516 y 4518 ya que estas son las posiciones de memoria asignadas a las entradas análogas 3 y 4 del MPC 134; y en Amount of registers/Coils se configuran la cantidad de variables que se van a leer de cada registro y para finalizar en MPC reg index se selecciona el registro de memoria del controlador se va a guardar la información censada y se procede a dar click en el botón Set.

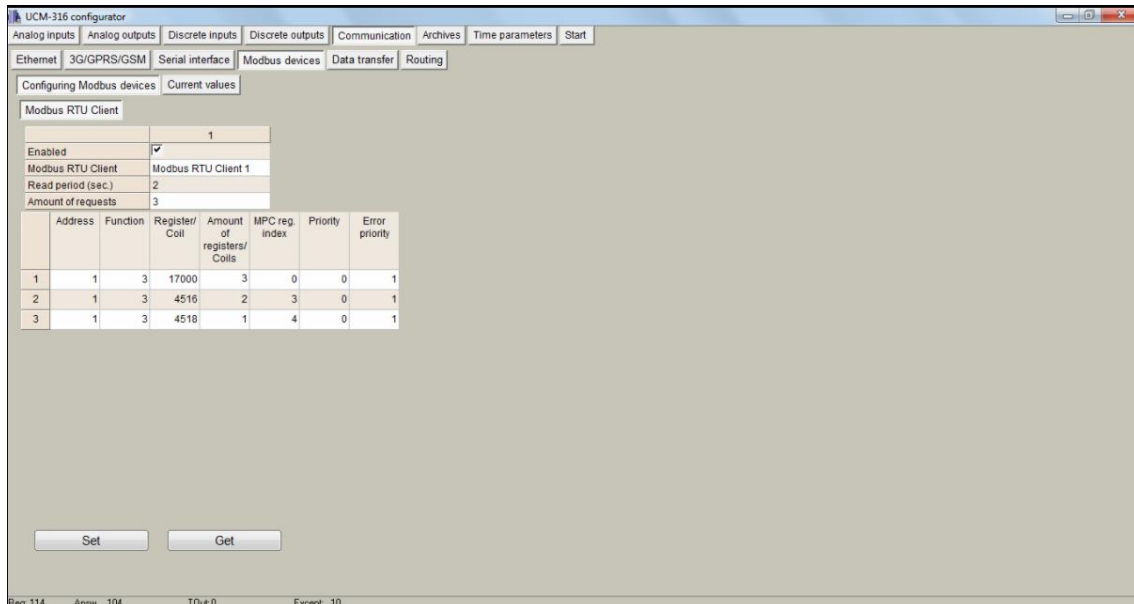
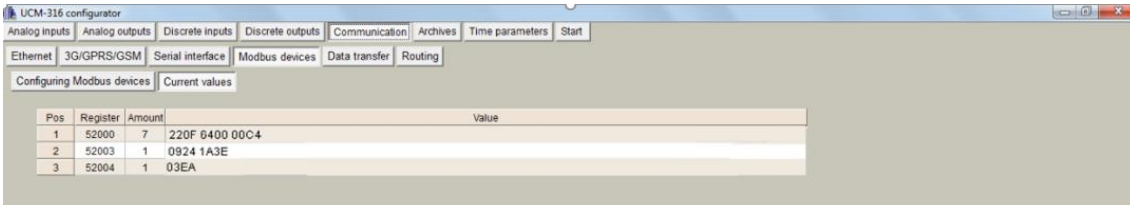


Figura 22. Configuración de lectura de datos transmitidos vía UART RS485 para UCM 316.  
(fuente: Autor)

Description	Modbus Register list						Data format	Type
	AI_1	AI_2	AI_3	AI_4	AI_5	AI_6		
Analog input use (0- disabled, 1 – enabled)	16200	16201	16202	16203	16204	16205	Char	R/W
Measuring type (0-voltage, 1-current, 2-resistance)	16206	16207	16208	16209	16210	16211	Char	R/W
Physical value (Float[4])	4418-4419	4420-4421	4422-4423	4424-4425	4426-4427	4428-4429	Float	R
Recalculated value (Float[4])	4430-4431	4432-4433	4434-4435	4436-4437	4438-4439	4440-4441	Float	R

Figura 23. Registro de entradas analógicas UCM 316  
(Fuente: Advanticsys)

Una vez hechas las anteriores configuraciones, los valores de los dispositivos se pueden observar dando click en Communication > Modbus devices > Current values; cada medida está representada por cuatro números en escala hexadecimal, tal y como se ve en la siguiente figura, donde ambos registros corresponden a los valores medidos por los sensores conectados al MPC 134 (el primer registro corresponde a los valores medidos por el IAQM-THCO2 y el segundo corresponde a los valores de los sensores de humedad FC-28).



Pos	Register	Amount	Value
1	52000	7	220F 0400 00C4
2	52003	1	0924 1A3E
3	52004	1	03EA

Figura 24. Visualización de los valores recogidos por el UCM 316  
(fuente: Autor)

Para poder lograr la comunicación entre el UCM 316 y el MPC 134, es importante configurar los parámetros comunicación, para ello se accede a Communication > Serial Interface; En las opciones de Bode, Parity, Data bits, y Stop bits se colocan los mismos valores configurados en el DM 124, en Packetization (tipo de paquetizado) se selecciona la opción Time (se espera un tiempo de paquete y se considera que la transmisión es exitosa), en Packet time (duración de la transmisión de paquetes en mili segundos) se colocó la opción en 8 y en Mode se selecciona la opción Half Duplex, las opciones Packet byte count y Packet symbol mantienen los valores que vienen por defecto; una vez configuradas las opciones se procede a dar click en el botón Set (el que se encuentra inmediatamente debajo de las opciones mencionadas).

Para el UCM 316, en las opciones de Purpose of UART se selecciona Modbus RTU Master, ya que este dispositivo va a funcionar como maestro en la comunicación entre él y el MPC 134; en la opción de Timeout (que corresponde a un retardo de lectura en milisegundos) se coloca el valor de 1000, es decir un retardo de 1 segundo (este número no es muy relevante para la configuración ya que la lectura de datos que se necesita realizar tiene un intervalo de varios minutos).

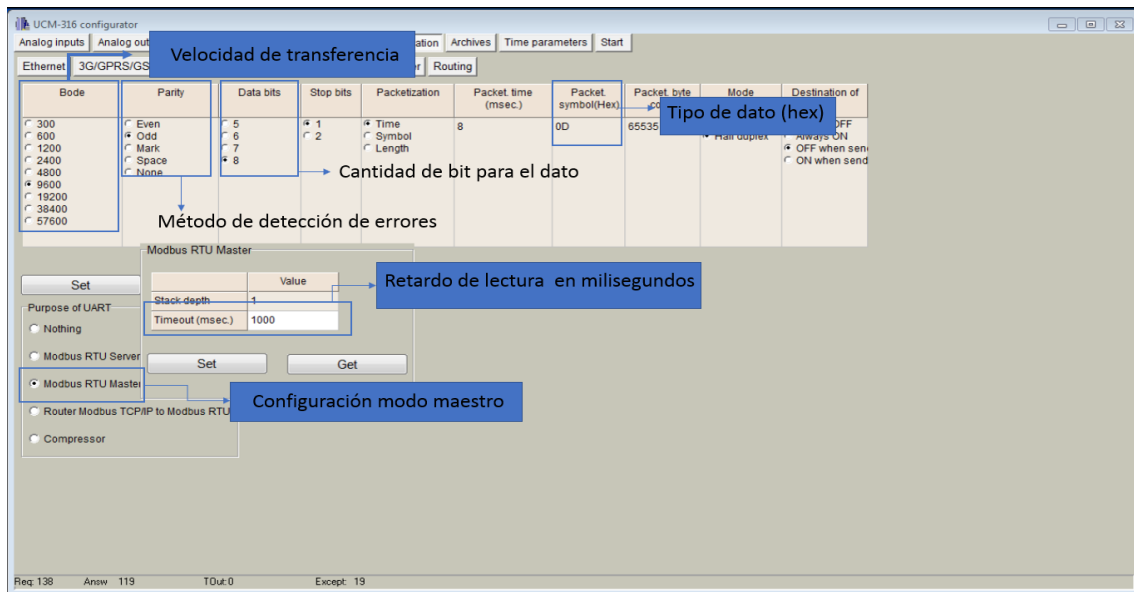


Figura 25. Configuración de comunicación del UCM 316 (fuente: Autor)

Una vez realizada esta configuración en el UCM 316 se procede a realizar la configuración para la lectura de datos del sensor lumínico LDR GL55; este sensor se conecta en uno de los puertos de entrada de voltaje análogos que posee el UCM 316, para ello el sensor debe estar conectado y en el software del UCM 316 se da click en Analog inputs > Configuration; en las opciones que aparecen se selecciona el puerto en el cual se conectó el sensor y se da click en la opción Enabled; también se puede configurar una alarma en la opción Alarme case pero para este caso dicha opción se mantuvo en None (sin alarma). Una vez realizada la configuración, se da click en el botón Set.

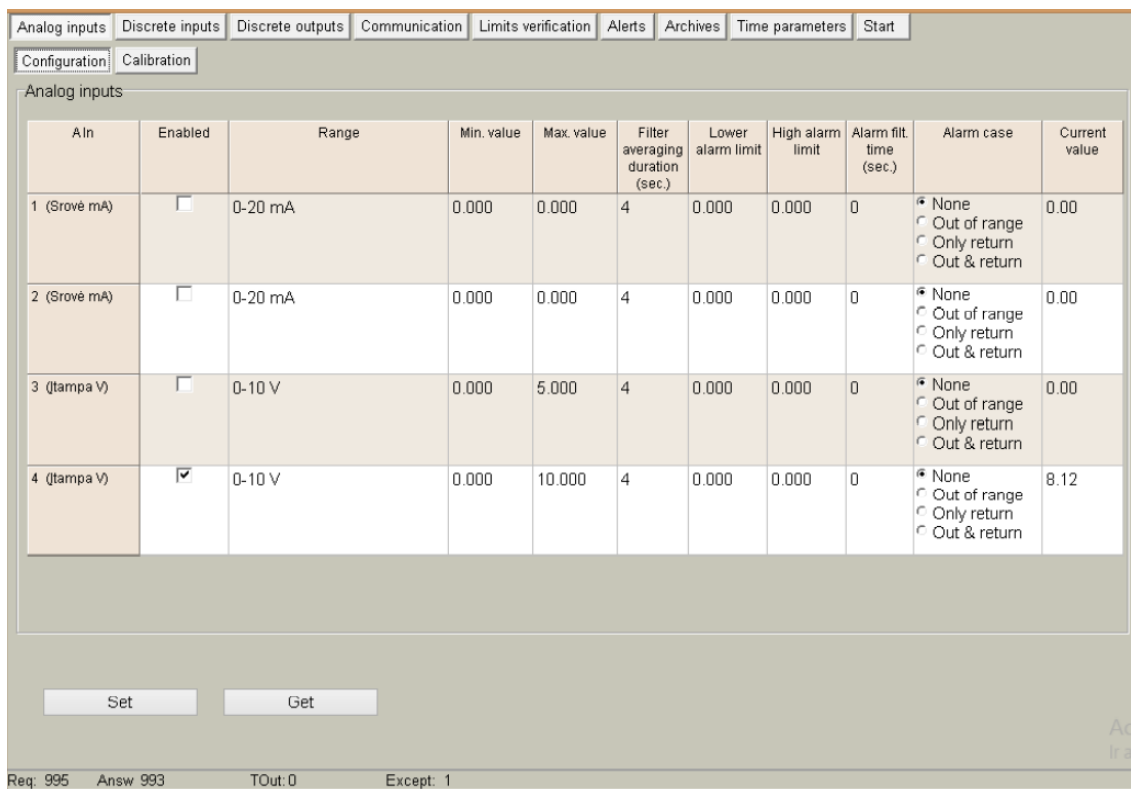


Figura 26. Configuración de puerto de entrada análogo de sensor LDR GL55. (fuente: Autor)

El siguiente paso a realizar en la configuración del UCM 316 es la organización de los datos que se van a almacenar y enviar; para ello se da click en Communication > Data Transfer > Common parameters. En primer lugar, se colocan los Register Coil donde la información está siendo almacenada dentro del dispositivo, para el caso de los datos que entran vía Modbus la información queda almacenada en los registros 52000, 52001, 52002, 52003, 52004, y para el caso del sensor lumínico LDR GL55 se coloca la dirección de registro de la entrada análoga en la cual fue conectado (en nuestro caso, como fue conectado en la entrada análoga uno y de acuerdo con la figura de registro de entradas analógicas UCM 316, se coloca el register coil 4418); luego se configura la opción Header, que es el encabezado que va a llevar cada columna donde se va a guardar cada dato; en la opción Dimension se pueden datos extra de cada columna (como pueden ser las unidades) . Finalmente, se da click en el botón Set

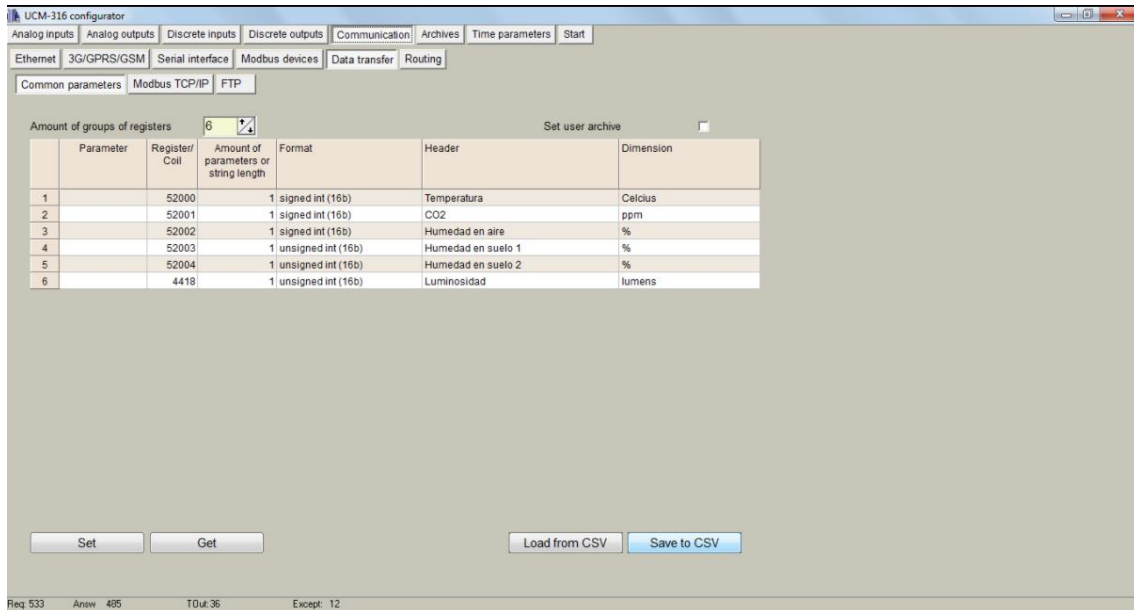


Figura 27. Configuración 3 UCM 316 organización de parámetros de lectura, almacenamiento y carga (fuente: Autor)

Al dar click en Archives > Configuration, en la opción Period se puede configurar el periodo (en minutos) en el cual se va a realizar la toma de información, así como agregar un tiempo de Delay si es necesario, luego de estas configuraciones se da click en el botón Set.

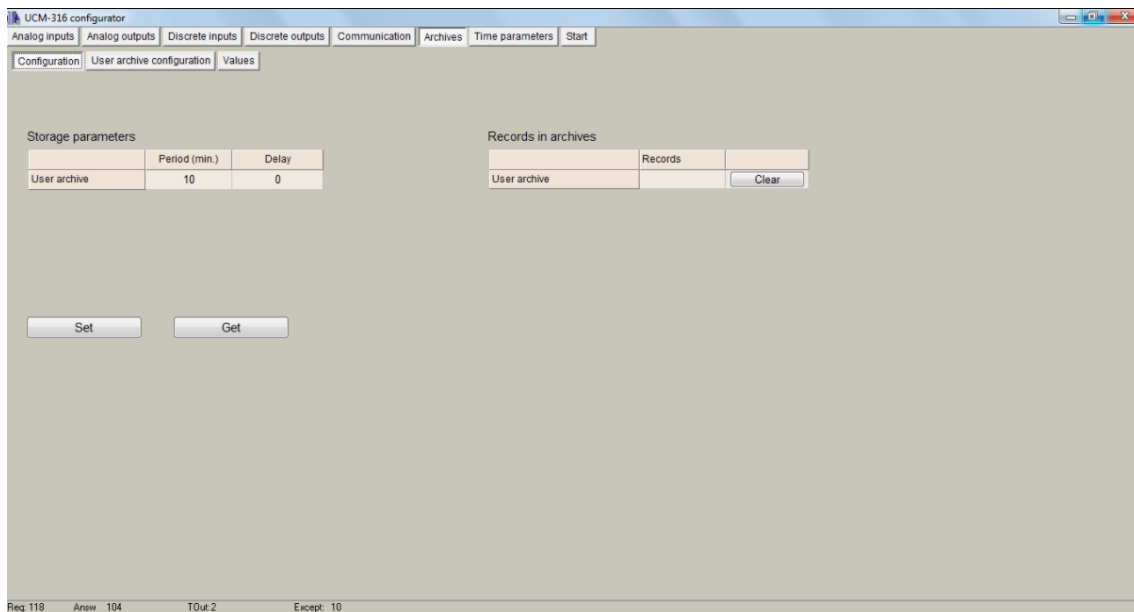


Figura 28. Configuración 3 UCM 316 tiempo y cantidad de lecturas (fuente: Autor)

## Transmisión de datos entre la red y el computador

Para realizar la transmisión de datos entre la red de sensores y el computador para su uso, se escogió el uso de un servidor FTP. Los datos se transmiten desde

el UCM 316 realizando la siguiente configuración. Se da click en Communication > Data transfer> FTP, luego se selecciona la opción Enabled de la configuración del servidor que se va a utilizar, en la opción FTP port se configura el número del puerto a través del cual se va a acceder al servidor (en este caso se escogió el puerto 21), en Transfer Period se elige el periodo (en minutos) en el que se desea realizar la transferencia (se eligió 10 minutos, que es el mismo período de toma de datos de los sensores), en Repeat transfer if failure se puede configurar, el tiempo en el cual quiere que se repita un intento de transmisión de datos si esta falla, en Max number of last records to be sent se define el número máximo de datos anteriores almacenados y no enviados debido a error de comunicación, en Transmission channel se configura el tipo de conexión mediante el cual se va a hacer la transmisión de datos (se decidió escoger la opción Ethernet). Al documento en el que se van a enviar los datos, se le puede asignar un nombre en la casilla Name template of the file to be sent (este tiene formato .csv).

En la opción URL and directory se coloca la dirección de acceso al servidor FTP, que en este caso se trata de la dirección IP pública del computador, ya que este además de procesar los datos, es usado como servidor FTP; en las opciones User name y Password se colocan el nombre de usuario y la clave de acceso al servidor FTP. Una vez configuradas las opciones mencionadas se procede a dar click en el botón Set.

En la parte inferior derecha de la ventana se puede una ventana donde se muestran los siguientes datos, en la opción Status se puede ver el estatus de la conexión al servidor, y en la opción Successful/attempt/last record transfer times se pueden ver las fechas de la última transferencia exitosa, el último intento y el último registro

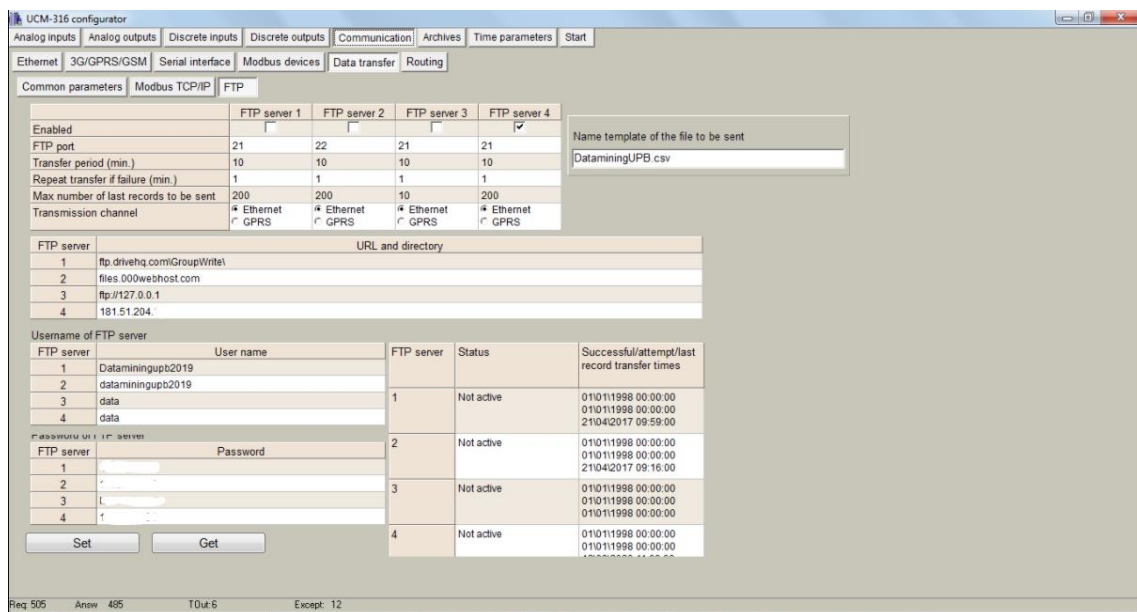
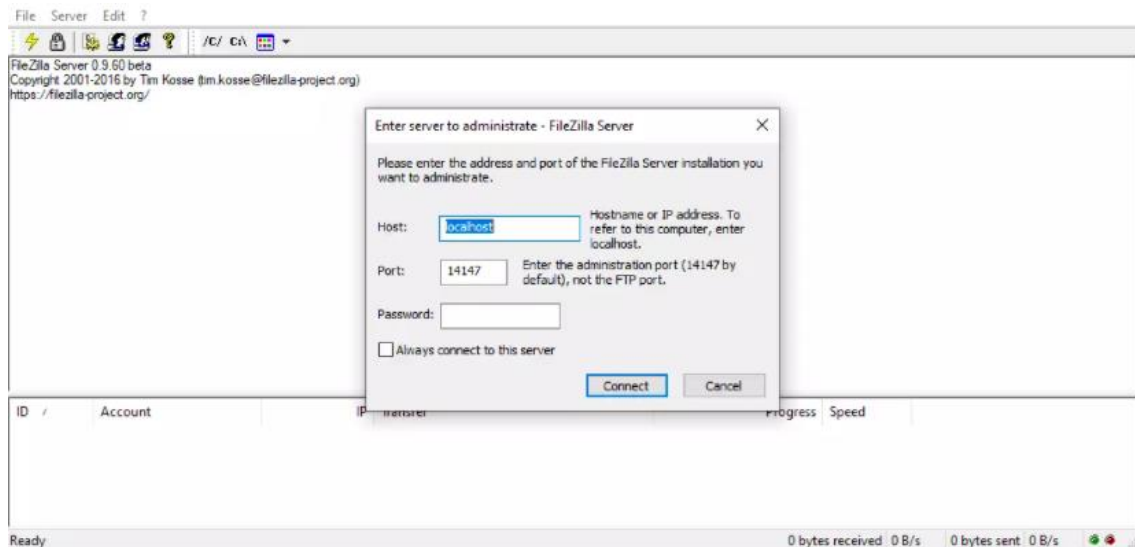


Figura 29. Configuración de envío de información vía FTP para UCM 316 (fuente: Autor)

El siguiente paso es crear el servidor FTP en el computador. Esto se hace utilizando el software Filezilla Server que, una vez instalado se deben realizar los siguientes pasos.

Al iniciar Filezilla Server se despliega una ventana donde se configuran las opciones de acceso al servidor, en Host se introduce el nombre o dirección IP donde está instalado el servidor, en este caso como el servidor es el mismo equipo, se coloca la palabra localhost; luego se configuran el puerto que utiliza el administrador del servidor (en este caso se deja el que viene por defecto que es el 14147) y la clave de acceso al servidor.



*Figura 30. Configuración inicial para servidor FTP en Filezilla Server (fuente: Autor)*

Una vez hecho esto el servidor quedará perfectamente operativo. Aparecerá un aviso en rojo donde se indica que se debe configurar el FTP en modo pasivo y hacer un reenvío de puertos hacia el servidor FTP, y una advertencia que dice que el FTP over TLS no está habilitado, por lo que los usuarios no pueden iniciar sesión de forma segura, ya que el envío de credenciales se hace mediante texto plano.

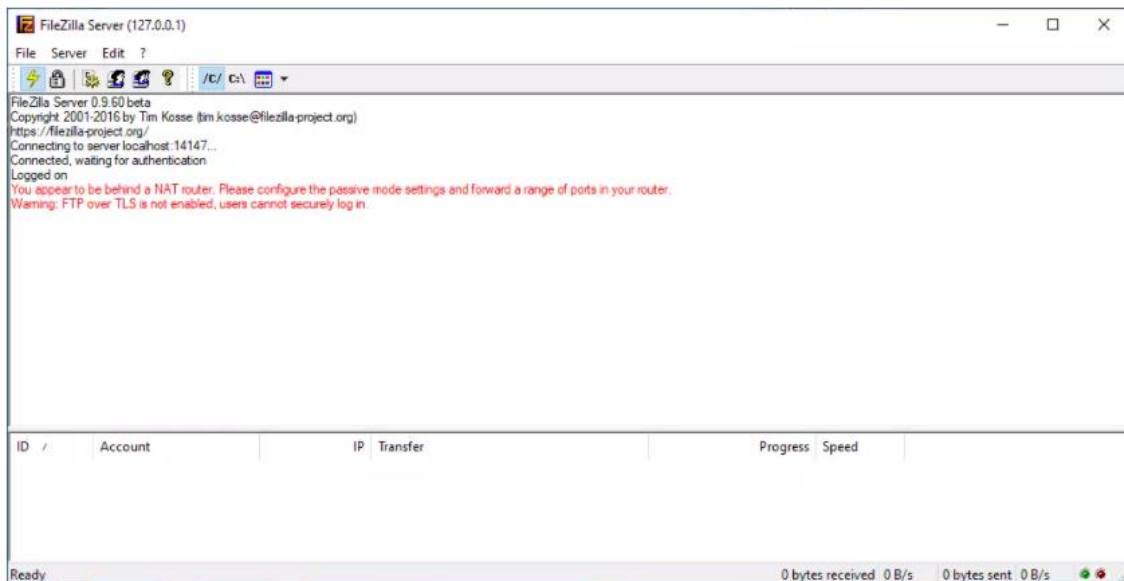


Figura 31. Pantalla principal Filezilla Server  
(fuente: Autor)

Una vez hecha la configuración precisa se da click en Edit > Settings donde se despliega el menú mostrado en la siguiente figura. En General Settings se procede a configurar el puerto mediante el cual se van a conectar al servidor FTP (es el mismo puerto configurado en el UCM 316, que en este caso es el 21). En este menú también se pueden configurar el Máximo número de usuarios (se optó por 3). Las Performance settings, Timeout settings y las opciones de los submenús Welcome message, IP biddings e IP Filter se dejaron por defecto.

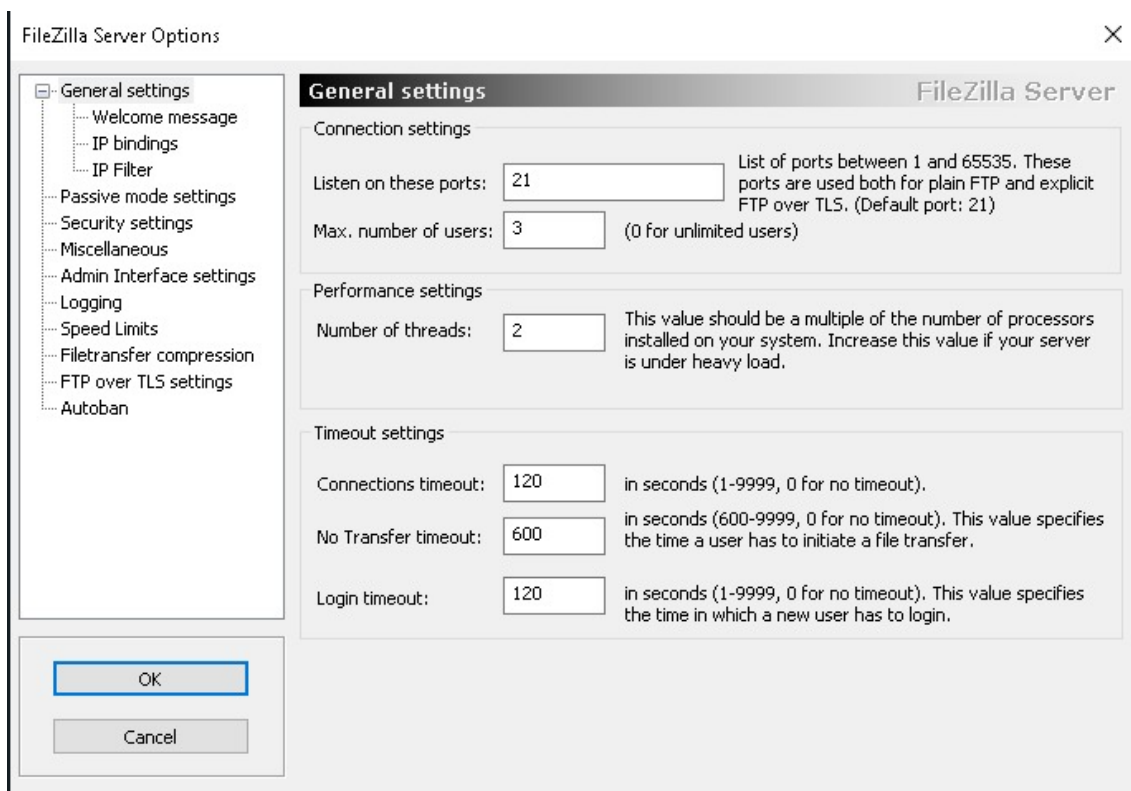


Figura 32. Menú General settings de Filezilla Server

(fuente: Autor)

En el submenú Passive mode settings se da click a la opción Use custom port range, que es un rango de puertos que el FTP va a utilizar (en este caso se eligió la el rango entre 1200 y 1500 y se elige la opción Retrieve external IP address from, ya que el FTP está operando detrás de una NAT (IP pública desde la cual se conectan a internet dispositivos con IPs privadas).

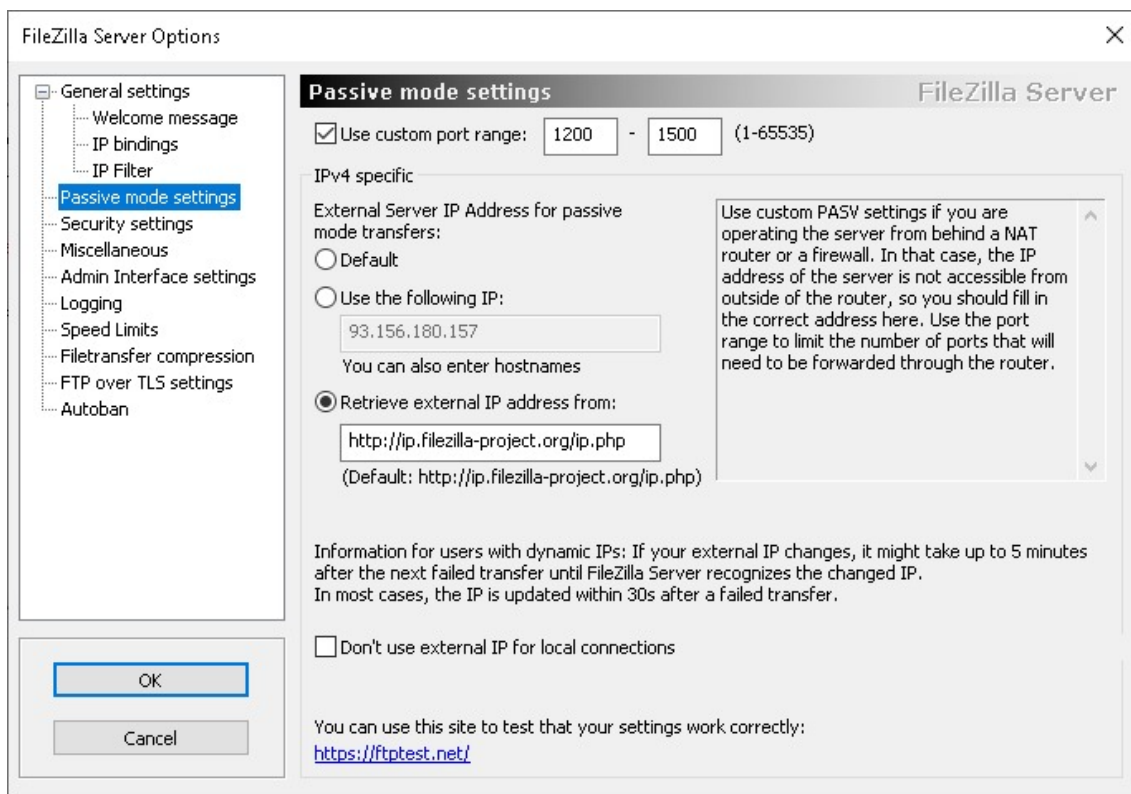


Figura 33. Menú Passive mode settings de Filezilla Server (fuente: Autor)

Luego en el submenú FTP over TLS settings, se habilita la opción Enable FTP over TLS support y se configura el puerto que va a utilizar el servidor para escuchar las conexiones implícitas FTP over TLS, se escoge el puerto 990.

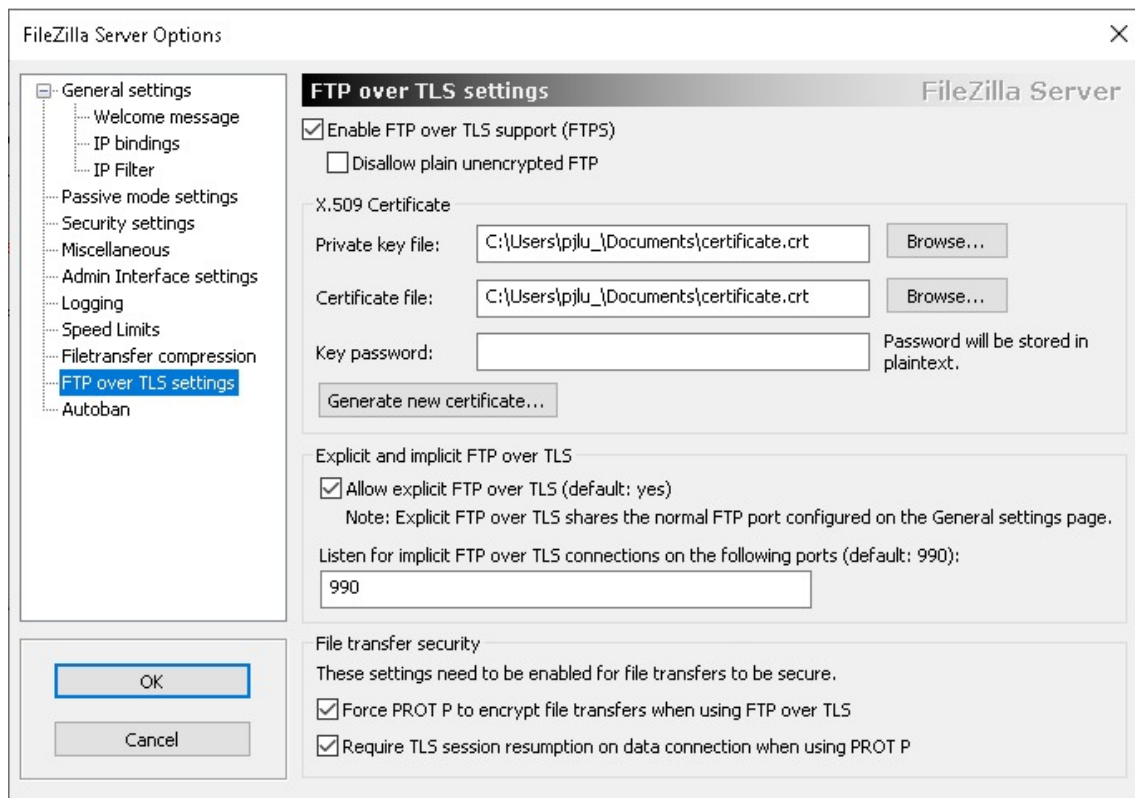


Figura 34. Menú FTP over TLS settings de Filezilla Server (Fuente:Autor)

En la opción Autoban se configura un límite de intentos de conexión fallidos en una hora, después de los cuales se banea la dirección IP y se establece el tiempo de baneo. Luego se da click en el botón OK.

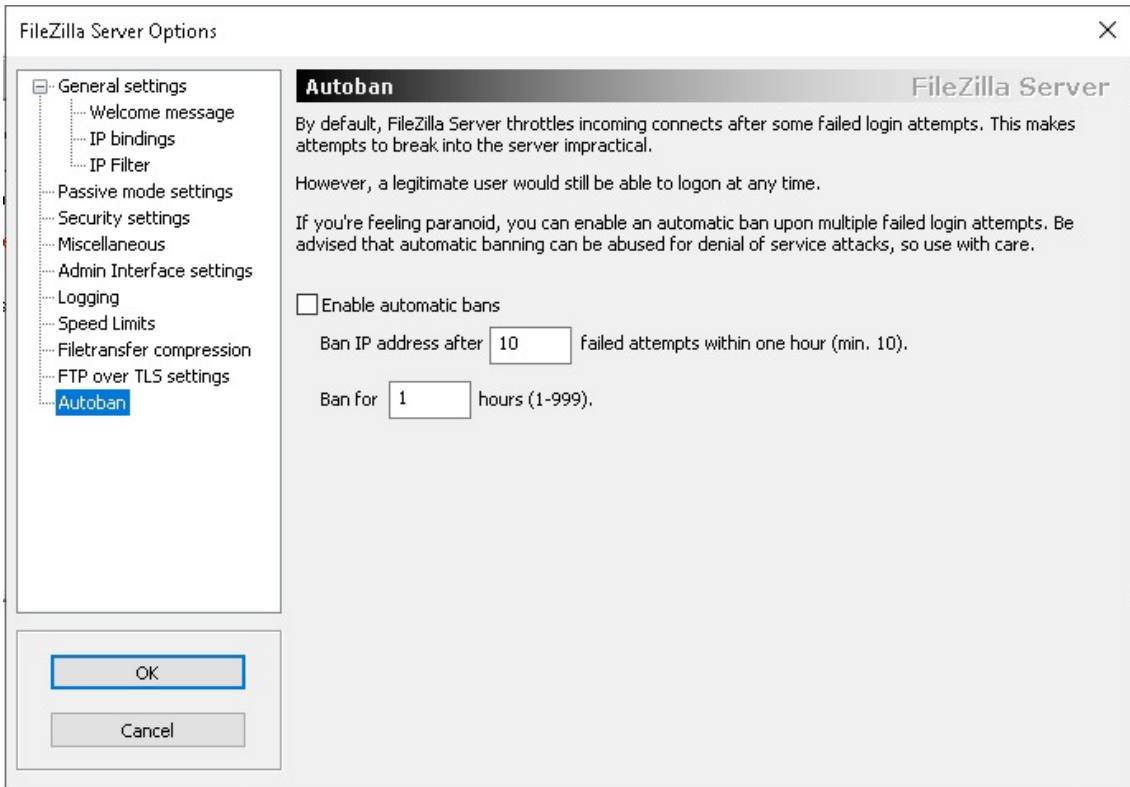


Figura 35. Menú Autoban de Filezilla Server  
(fuente: Autor)

**Nota:** Se debe cerciorar que los puertos usados en esta configuración están abiertos, para ello comuníquese con su proveedor de internet.

Como siguiente paso se procede a configurar usuarios para acceso al FTP, para lo cual se da click en Edit > Users, donde se despliega el menú mostrado en la siguiente figura.

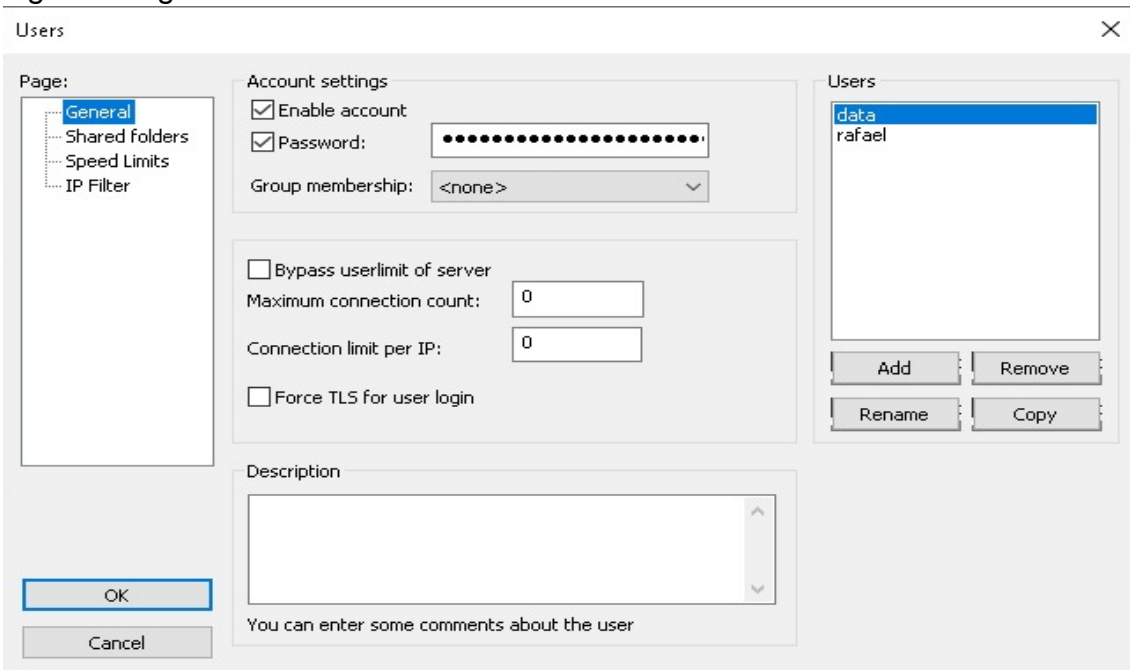


Figura 36. Menú General de usuarios de Filezilla Server

(fuente: Autor)

En el submenú General se procede a dar click en el botón Add, donde se procede a agregar un nuevo usuario.



Figura 37. Configuración de nuevo usuario en Filezilla Server  
(fuente: Autor)

Luego de agregado el nuevo usuario, se seleccionan las opciones Enable account y Password y se introduce la misma clave que fue puesta en la configuración del UCM 316. Si se desea se pueden agregar más usuarios, pero no se debe olvidar que solo se pueden conectar al tiempo, el máximo de usuarios configurados previamente en Edit > Settings.

Luego se configura la carpeta del computador a la que se le va a permitir el acceso a cada usuario. Para esto se da click en el submenú Shared folders donde se despliegan las opciones mostradas en la siguiente figura.

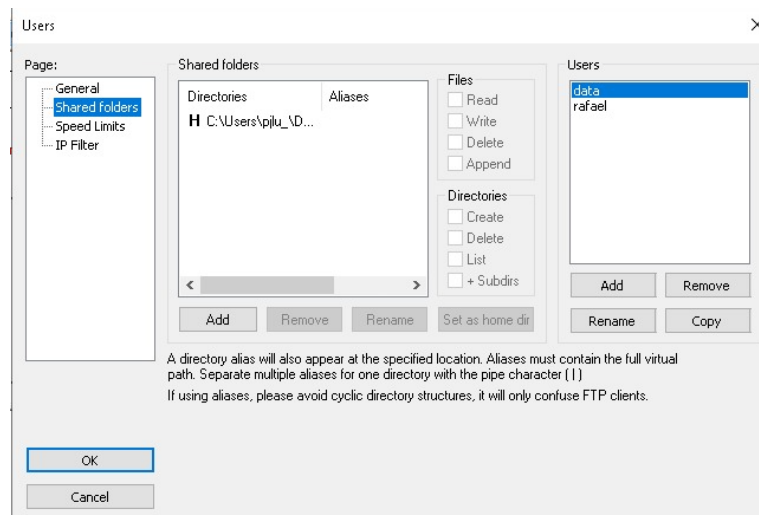


Figura 38. Menú Shared folders de Filezilla Server  
(Fuente: Autor)

Se da click sobre el usuario deseado, luego se da click al botón Add que se encuentra ubicado en la parte central de la ventana, lo que desplegará una

ventana donde se busca la carpeta, donde una vez seleccionada se da click en Aceptar

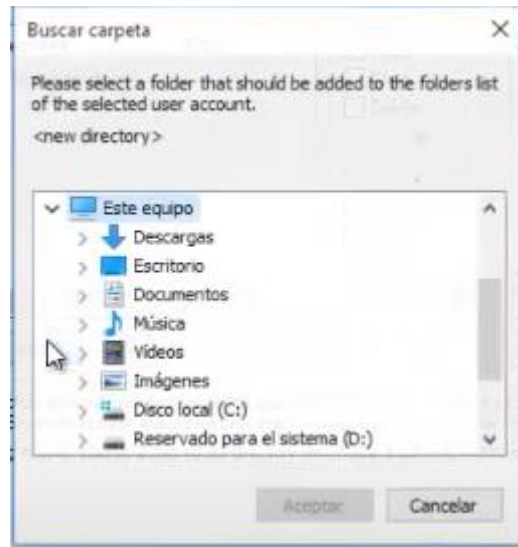


Figura 39. Búsqueda de carpeta a compartir con el usuario (fuente: Autor)

La opción Files permite seleccionar el tipo de permisos que se le desea dar al usuario, en este caso se le conceden todo, ya que este usuario es el que permite al UCM 316 conectar al servidor FTP. La opción Directories no es utilizada ya que para este caso no se requiere ningún tipo de manipulación de carpetas o subdirectorios. Se procede a dar click en Ok.

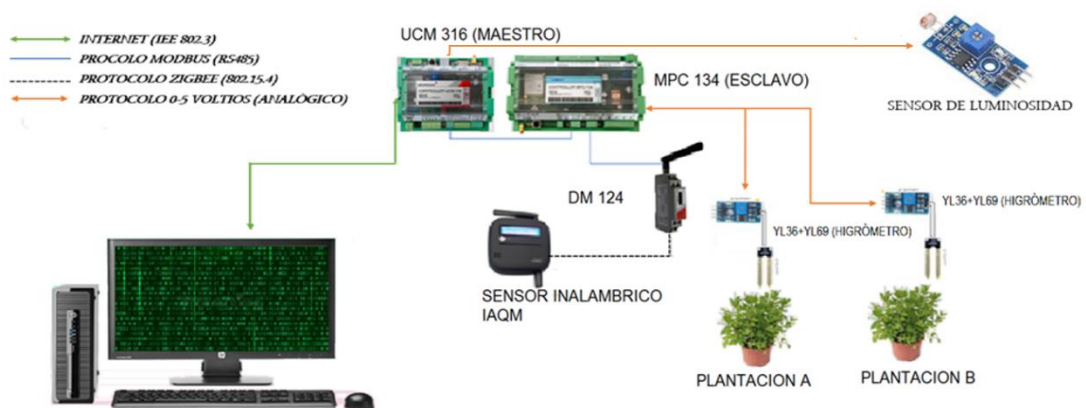


Figura 40. Arquitectura física del experimento (fuente: Autor)



*Figura 41. Montaje realizado para las pruebas  
(fuente: Autor)*

#### **4.1.3. Adquisición de datos y puesta en funcionamiento**

En la adquisición de datos se pone en funcionamiento la versión configuración definitiva de hardware, mostrada en la figura 40, y se dispone a la recolección de los datos producto de las condiciones a las que se exponen los sujetos de prueba. Estas condiciones establecen que uno de los dos sujetos es expuesto a en sequía, mientras el otro se mantiene en condiciones favorables, también se establece que por practicidad del experimento solo se van a manipular la humedad del suelo y que, para que los datos recopilados sean válidos, un sujeto no puede verse afectado por alguna enfermedad o plaga y solo puede ser estresado o morir producto de la manipulación del suministro de agua. Se considera que un sujeto está en sequía cuando ha pasado un tiempo mayor a dos días sin recibir agua.

La adquisición de los datos y puesta en funcionamiento como inicialmente se pensó, contemplaba el uso del servidor ubicado en el laboratorio de comunicaciones (K301). La adquisición de ellos iría directamente a través del protocolo Zigbee, interconectados a través de nodos DM 124 hasta el servidor. Posteriormente, debido al cambio de la ubicación de los sujetos de pruebas, se pensó en enviar la información desde casa hasta el servidor para la manipulación de la información; sin embargo, se encontró que, debido a motivos de seguridad, para tener acceso remoto al servido se debía gestionar ante la universidad la obtención de una IP estática, tomando en cuenta esta dificultad se opta por usar un ordenador con el fin de procesar la información. También se pensó en el uso de más sensores de humedad, aprovechando las conexiones análogas con protocolo 4 a 20 mA que poseen tanto el UCM 316 como el MPC 134, pero fue imposible conseguirlos.

El UCM 316 es el encargado de la recolección final de los datos, y su posterior envío a la computadora mediante el uso de un servidor FTP en un archivo .CSV

en intervalos de 10 minutos. El UCM 316 recolecta la información suministrada por el sensor de luminosidad y por el MPC 134, que a su vez recolecta la información del IAQM-THCO2 y de los sensores de humedad de suelo.

Si bien en los datos tomados aparecen unidades como los lumens y el % en el caso de la humedad del suelo, no fue posible establecer unidades o formato alguno en los datos que los sensores (Higrómetro Arduino FC-28 (YL 36 y YL 69) y Sensor de luminosidad -fotorresistencia LDR GL55) enviaban al MPC 134 y al UCM 316, debido a que a pesar de que estos funcionan de forma análoga bajo el estándar de 0 a 5V, son sensores pensados para trabajo en la plataforma Arduino. Sin embargo, se optó por su uso ya que, en los datos recopilados se reflejan los cambios en las variables medidas por estos, además de la no existencia de sensores de humedad y luminosidad diseñados para el MPC 134 y el UCM 316 en el inventario del laboratorio de electrónica de la universidad y la imposibilidad de adquirirlos. Las unidades inicialmente mencionadas no se borraron, ya que para hacerlo desde donde se establecieron, habría requerido reiniciar la toma de datos, y su presencia no afecta el desarrollo de las fases siguientes. Por su parte, el sensor IAQM-THCO2 sí trabaja con unidades establecidas, siendo grados Celsius (°C) para la temperatura, partes por millón (ppm) para la concentración de CO<sub>2</sub> en el aire, y como porcentaje para la humedad en el aire (%).

La estructura de los datos obtenidos durante la adquisición se puede observar en la siguiente figura.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Time	Temperatura	CO2			Humedad en aire	Humedad en suelo 1	Humedad en suelo 2	Luminosidad	EstadosPlanta_1	EstadosPlanta_2	N_hojas_1	N_hojas_2				
2	2020.01.01 00:	2599 Celcius	6315 ppm			704 %	50709 %	50710 %	16389 lumens	1	1	72	97				
3	2020.01.01 00:	2598 Celcius	6388 ppm			700 %	50709 %	50710 %	16390 lumens	1	1	72	97				
4	2020.01.01 00:	2598 Celcius	6517 ppm			698 %	50710 %	50710 %	16390 lumens	1	1	72	97				
5	2020.01.01 00:	2591 Celcius	6493 ppm			694 %	50710 %	50710 %	16397 lumens	1	1	72	97				
6	2020.01.01 00:	2590 Celcius	6315 ppm			693 %	50709 %	50710 %	16398 lumens	1	1	72	97				
7	2020.01.01 00:	2589 Celcius	6388 ppm			689 %	50709 %	50710 %	16399 lumens	1	1	72	97				
8	2020.01.01 00:	2589 Celcius	6517 ppm			682 %	50710 %	50710 %	16399 lumens	1	1	72	97				

Figura 42. Datos obtenidos a partir de los archivos (fuente: Autor)

Se puede observar la fecha de toma de datos en la columna Time; columnas con las variables registradas (temperatura, CO<sub>2</sub>, humedad en el aire, humedad en el suelo del sujeto 1 y del sujeto 2, luminosidad, Número de hojas del sujeto 1 y del sujeto 2), así como columnas con sus respectivas unidades.

#### 4.1.4. Creación de algoritmos, limpieza y manipulación con técnicas de machine learning

Durante la limpieza y manipulación con técnicas de machine learning, se toman los datos ya recopilados de los sujetos de prueba, se crean los algoritmos y son aplicados a esta información, con el fin de sacar datos que sean relevantes y concluyentes para el cumplimiento de los objetivos planteados. Estos datos por lo general son las relaciones existentes entre las variables que se deciden tomar y el estado. Para la elaboración de este trabajo se usó Jupyter Notebook 6.1.4 con las librerías Pandas, Matplotlib, Scipy y Numpy, corriendo en un computador Asus X405U bajo sistema operativo Windows 10, equipado con 4 GB de RAM, disco magnético de 1 TB de capacidad, un procesador Intel Core i7 – 7500U. Las

librerías y el software anteriormente mencionados son de distribución gratuita. El trabajo en torno a los modelos de Machine Learning se estructuró en los siguientes pasos.

#### 4.1.4.1 Data Wrangling

El primer paso para el diseño de los modelos fue el manejo y la limpieza de los datos. Basados en la tabla de la Figura 8, se procede a determinar qué columnas van a ser las variables de entrada, cuál va a ser la variable de salida, qué modificaciones se deben realizar y qué columnas o filas se pueden eliminar.

Como variables de entrada, inicialmente se toman las columnas de temperatura, CO<sub>2</sub> que corresponde al nivel de CO<sub>2</sub> presente en el aire, Humedad de aire, Humedad en suelo 1 que corresponde a la humedad medida en el suelo donde se encuentra plantado el sujeto 1, Humedad en suelo 2 que corresponde a la humedad medida en el suelo donde se encuentra plantado el sujeto 2, Luminosidad, N\_hojas\_1 que corresponde al número de hojas del sujeto 1 al momento de tomar la medida y N\_hojas\_2 que corresponde al número de hojas del sujeto 2 al momento de tomar la medida; como variable de salida se toman las columnas EstadosPlanta\_1 y EstadosPlanta\_2 que corresponden a los estados de los sujetos, donde 1 significa que el sujeto está en buen estado, 2 que el sujeto está enfermado (presenta colores más opacos, se está marchitando). Se optó por eliminar la columna Time, así como aquellas donde por defecto vienen las unidades de los datos medidos, teniendo en cuenta la cantidad de datos registrados se optó por eliminar aquellas filas que presentaran un dato vacío (pasando de tener 21312 registros en la tabla a 16133); también se modificaron los encabezados de las columnas Humedad en aire, Humedad en suelo 1, Humedad en suelo 2 por hum\_aire, hum\_gnd\_1 y hum\_gnd\_2 respectivamente, debido a que estos presentan espacios y en algunos casos, al requerir estas columnas pueden presentar errores.

Posteriormente se realiza una fase exploratoria de los datos, graficando las variables medidas, como se puede ver a continuación

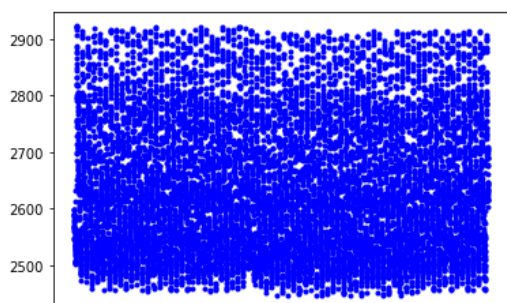


Figura 43. Gráfica de datos de temperatura obtenidos (fuente: Autor)

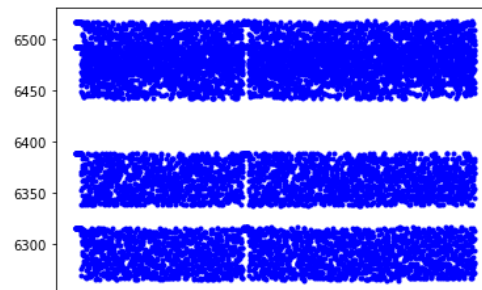


Figura 44. Gráfica de datos de CO<sub>2</sub> obtenidos (fuente: Autor)

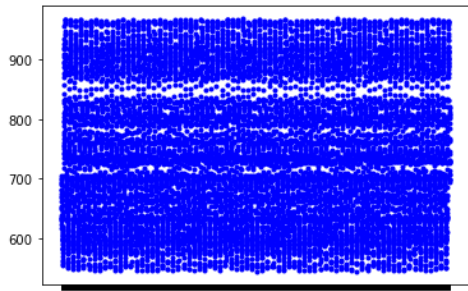


Figura 45. Gráfica de datos de humedad de aire obtenidos  
(fuente: Autor)



Figura 46. Gráfica de datos de humedad de suelo obtenidos  
(fuente: Autor)



Figura 47. Gráfica de datos de luminosidad obtenidos  
(fuente: Autor)

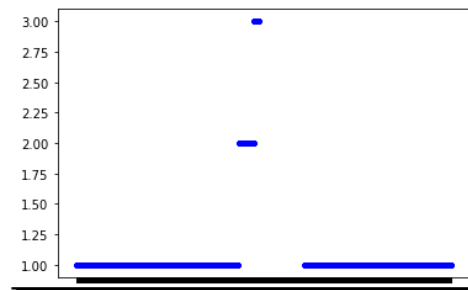


Figura 48. Gráfica de datos de Estados planta 1 obtenidos  
(fuente: Autor)

(fuente: Autor)

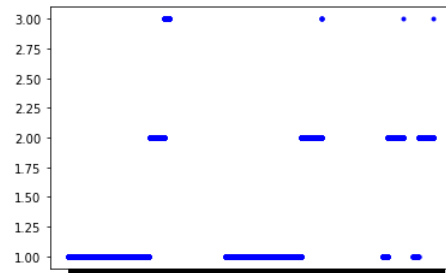


Figura 49. Gráfica de datos de Estados planta 2 obtenidos  
(fuente: Autor)

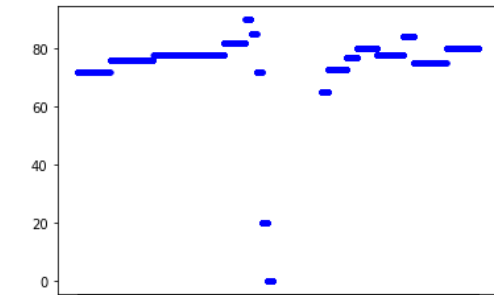


Figura 50. Gráfica de datos de Numero de hojas planta 1 obtenidos  
(fuente: Autor)

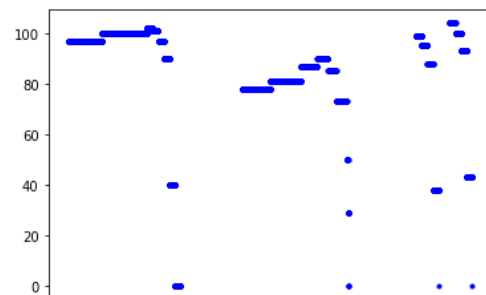


Figura 51. Gráfica de datos de Numero de hojas planta 2 obtenidos  
(fuente: Autor)



En la primera etapa se lleva a cabo primer acercamiento a los modelos de machine learning, se construye el árbol de decisión y se deja crecer sin ningún límite, su desarrollo se puede observar en la siguiente figura.

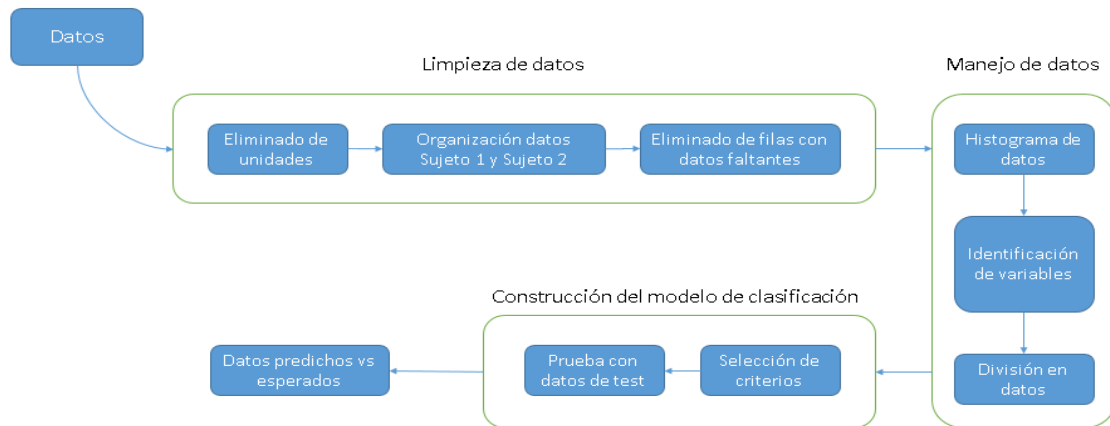


Figura 54. Proceso de desarrollo de árbol de decisión sin podar (fuente: Autor).

El árbol sin podar tuvo una profundidad de veintidós niveles. Posteriormente se procede a realizar la poda de acuerdo con la medida de precisión, de la cual se concluye que, para este caso se iba a podar en su cuarto nivel de profundidad; teniendo un valor de 95,49% de precisión y evitando la presencia de overfitting al no tener tantos niveles de profundidad. El proceso de construcción del árbol junto a la poda se puede ver en la siguiente figura.

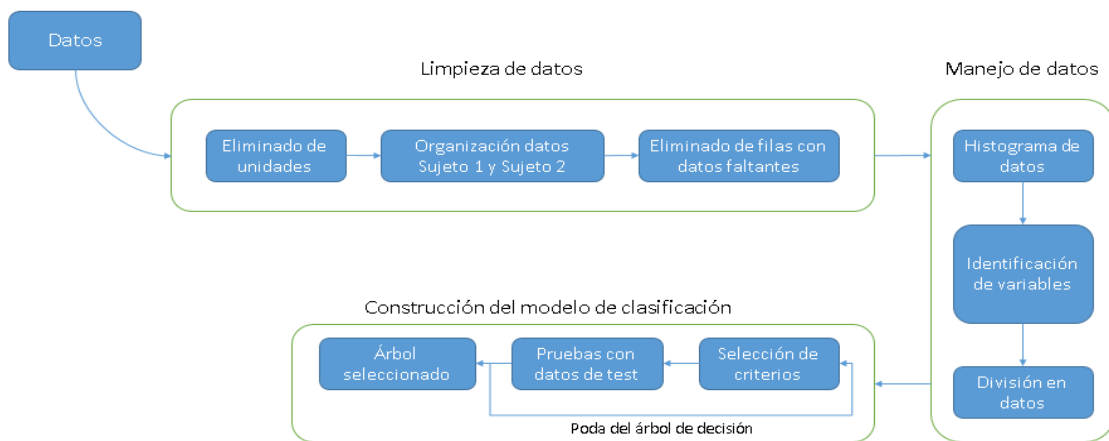


Figura 55. Proceso de desarrollo de árbol de decisión con poda (fuente: Autor).

Con el fin de comparar el desempeño del árbol de decisión, en la segunda etapa de desarrollo del algoritmo, se implementa random forest. Su implementación se puede ver representado en la siguiente figura.

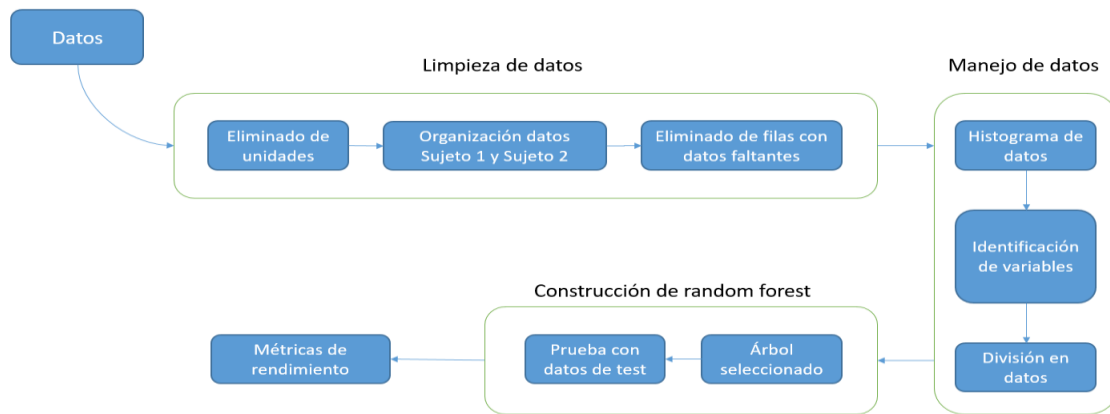


Figura 56. Proceso de desarrollo de random forest (fuente: Autor).

A partir de este punto de desarrollo, se decide agregar más métricas (recall y f1-score) con el fin de obtener mejores datos a la hora de comparar los resultados obtenidos en ambos modelos.

Una de las características que tiene el dataset manejado, es que posee datos no balanceados (existen más muestras de la clase 1 que de la 2 y la 3), por lo que se procede a aprovechar una opción que viene por defecto en el código fuente del random forest, la cual permite ajustar los pesos de forma inversamente proporcional a la frecuencia de cada clase en los datos[58], con el fin de ver si el uso de un dataset más balanceado, mejoraba las métricas del modelo. Su implementación se puede ver representado en la siguiente figura.

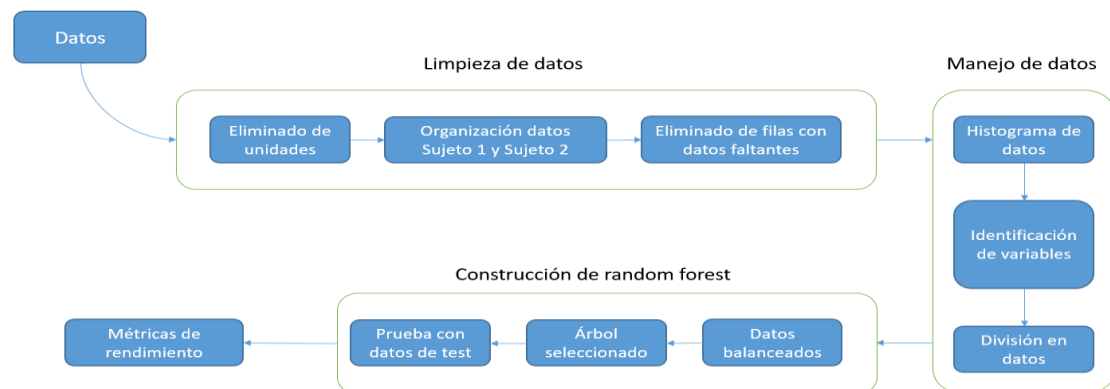


Figura 57. Proceso de desarrollo de random forest con datos balanceados (fuente: Autor).

Otra característica de este dataset es el manejo de distintas escalas de valores en los datos de las variables, por lo cual se procede a realizar la normalización de los datos y medir que impacto tiene esto en las métricas de los modelos. La implementación del algoritmo random forest con datos balanceados y normalizados se puede ver representado en la siguiente figura.

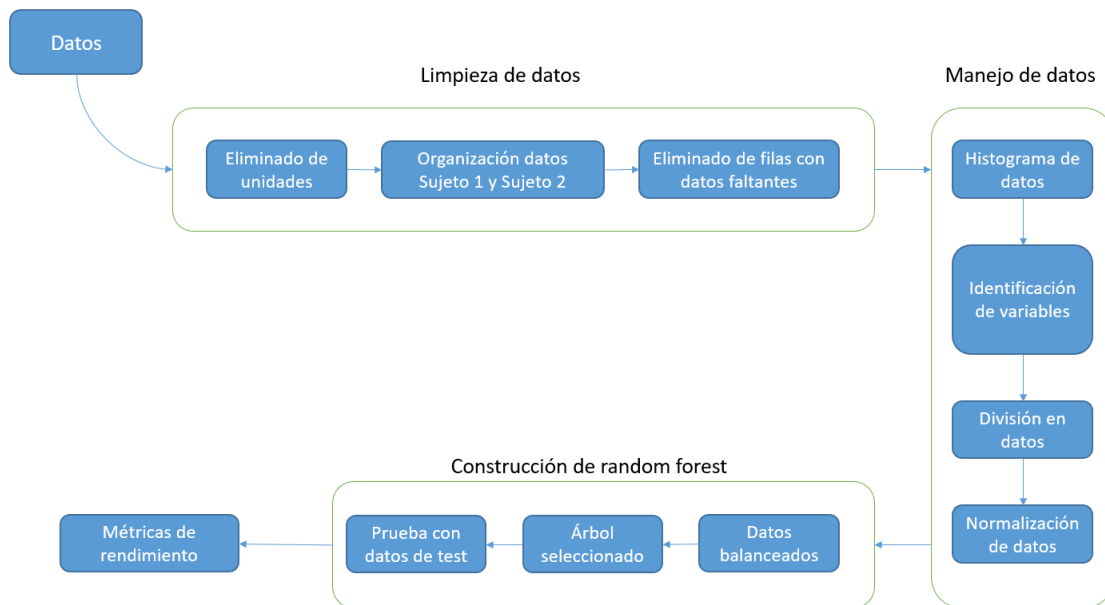


Figura 58. Proceso de desarrollo de random forest con datos balanceados y normalizados (fuente: Autor).

Se observa mediante la aplicación de un test de rasgos de importancia, que la variable más relevante para el modelo es el Número de hojas, sin embargo, variables como Temperatura, CO<sub>2</sub>, Humedad del aire o Luminosidad poseen una importancia casi nula, por lo que, en la fase 3 de desarrollo se implementan los mismos procedimientos realizados en la fase dos, pero sin tener en cuenta la variable de entrada Número de Hojas, esto con el fin de medir el comportamiento de los modelos respecto a las demás variables de entrada. Los resultados obtenidos se pueden ver en la siguiente tabla.

En la fase cuatro se fusionaron los valores 2 y 3 de la variable de salida, ya que estas presentan baja cantidad de muestras, si se comparan con la variable 1, lo que puede provocar problemas al modelo. Los modelos implementados son los mismos que en las fases dos y tres.

#### 4.1.4.3 Evaluación

Para el proceso de evaluación de los modelos, se tomaron en cuenta los valores de la precisión, recall, y el f-1 score, así como los resultados del test de rasgos de importancias.

Modelo	Fase 2					Fase 3					Fase 4				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Temperatura	0	0.01	0.01	0	0	0.08	0.08	0.08	0.08	0.08	0	0.02	0.01	0	0.01
CO <sub>2</sub>	0	0	0.01	0	0	0.09	0.08	0.09	0.09	0.09	0	0.01	0.01	0	0.01

hum_aire	0	0.01	0	0	0	0.08	0.08	0.08	0.08	0.08	0.01	0.02	0.01	0.01	0.01
hum_gnd	0.17	0.29	0.26	0.18	0.18	0.35	0.38	0.40	0.35	0.35	0.48	0.52	0.49	0.48	0.49
Luminosidad	0	0.06	0.05	0	0	0.17	0.17	0.19	0.17	0.17	0.15	0.12	0.14	0.15	0.14
N_hojas	0.82	0.49	0.58	0.82	0.82										
sujeto	0	0.15	0.09	0	0	0.23	0.22	0.17	0.23	0.23	0.36	0.32	0.34	0.36	0.34

Tabla 1. Resultados del test de rasgos de importancia.

En la anterior tabla, en la fila llamada Modelo, cada número corresponde a un modelo evaluado, donde el número 1 corresponde al Árbol de decisión, el número 2 al Random Forest, el número 3 al Random Forest con datos balanceados, el 4 al Árbol de decisión con datos normalizados, y el 5 al Random Forest con datos balanceados y normalizados.

## Rendimiento de modelos en la fase 2

	Árbol de decisión			Random Forest			Random forest data balanceada		
Valor de salida	Precisión	Recall	f1-score	Precisión	Recall	f1-score	Precisión	Recall	f1-score
1	0.95	0.86	0.90	0.89	0.96	0.92	0.98	0.74	0.84
2	0.46	0.72	0.56	0.50	0.28	0.36	0.38	0.92	0.54
3	1	1	1	0.84	0.99	0.91	1	1	1

Tabla 2. Métricas para modelos de la fase 2

	Árbol de decisión datos normalizados			Random Forest datos normalizados y balanceados		
Valor de salida	Precisión	Recall	f1-score	Precisión	Recall	f1-score
1	0.95	0.86	0.90	0.98	0.74	0.84
2	0.46	0.72	0.56	0.37	0.91	0.53
3	1	1	1	1	1	1

Tabla 3. Métricas para modelos con datos normalizados de la fase 2

Durante la fase 2, se observa que los modelos con mejor desempeño son el árbol de decisión y el árbol de decisión con datos normalizado, siendo estos los que tienen mejores valores de precisión, recall y f1-score para todos los valores de la variable de salida menos para la uno (siendo sus mejores resultados los obtenidos en el random forest con datos balanceados y en el random forest con datos balanceados y normalizados) , pero observando el test de rasgos de importancia también se observa que estos modelos se basan únicamente en dos variables de entrada: humedad de suelo(hum\_gnd) y número de hojas (N\_hojas), siendo esta última la que más importancia tiene para el modelo.

## Rendimiento de modelos en la fase 3

	Árbol de decisión	Random Forest	Random forest data balanceada
--	-------------------	---------------	-------------------------------

Valor de salida	Precisión	Recall	f1-score	Precisión	Recall	f1-score	Precisión	Recall	f1-score
1	0.90	0.89	0.90	0.89	0.91	0.90	0.97	0.84	0.90
2	0.36	0.40	0.40	0.33	0.32	0.33	0.37	0.60	0.46
3	0.02	0.01	0.01	0	0	0	0.02	0.06	0.03

*Tabla 4. Métricas para modelos de la fase 3*

Valor de salida	Árbol de decisión datos normalizados			Random Forest datos normalizados y balanceados		
	Precisión	Recall	f1-score	Precisión	Recall	f1-score
1	0.91	0.89	0.90	0.99	0.75	0.85
2	0.36	0.40	0.38	0.29	0.60	0.40
3	0.02	0.01	0.02	0.09	0.38	0.15

*Tabla 5. Métricas para modelos con datos normalizados de la fase 3*

Durante la fase 3 se observa que los modelos con mejores resultados en todos los valores de la variable de salida es el random forest con datos balanceados y en el random forest con datos balanceados y normalizados. Si bien sus métricas con la variable de salida son similares, para las variables de salida dos y tres empeoran considerablemente respecto a la fase 2. Se observa en los resultados del test de rasgos de importancia, que los modelos en esta fase se basan en mayor o menor medida en todas las variables, siendo la humedad del suelo (hum\_gnd) la que más importancia tiene.

#### Rendimiento de modelos en la fase 4

Variable de salida	Árbol de decisión			Random Forest			Random forest data balanceada		
	Precisión	Recall	f1-score	Precisión	Recall	f1-score	Precisión	Recall	f1-score
1	0.89	0.93	0.91	0.86	0.93	0.89	0.99	0.75	0.86
2	0.51	0.40	0.45	0.33	0.18	0.23	0.42	0.96	0.59

*Tabla 6. Métricas para modelos de la fase 4*

Valor de salida	Árbol de decisión datos normalizados			Random Forest datos normalizados y balanceados		
	Precisión	Recall	f1-score	Precisión	Recall	f1-score
1	0.89	0.93	0.91	0.99	0.75	0.86
2	0.51	0.40	0.45	0.42	0.96	0.59

*Tabla 7. Métricas para modelos con datos normalizados de la fase 4*

Durante la fase 4 se observa que los modelos con mejores resultados son el random forest con datos balanceados y el random forest con datos balanceados y normalizados, teniendo mejores métricas para todas las variables que los modelos desarrollados en las anteriores fases. También se observa en el test de rasgos de importancia, que al igual que en la fase 3, los modelos toman en cuenta más variables de entrada que en la fase 2, siendo la humedad del suelo (hum\_gnd) la que más importancia tiene.

#### 4.1.4.4 Análisis de resultados y selección del modelo

Valor de salida	Modelo con mejor rendimiento fase 2			Modelo con mejor rendimiento fase 3			Modelo con mejor rendimiento fase 4		
	Precisión	Recall	f1-score	Precisión	Recall	f1-score	Precisión	Recall	f1-score
1	0.95	0.86	0.90	0.91	0.89	0.90	0.99	0.75	0.86
2	0.46	0.72	0.56	0.36	0.40	0.38	0.42	0.96	0.59
3	1	1	1	0.02	0.01	0.02	-	-	-

Tabla 8. Comparación de los mejores resultados obtenidos en cada fase

Al analizar los modelos que obtienen mejores resultados en cada fase, se descarta la implementación de los modelos de la fase 3 debido a que las métricas obtenidas para los valores 2 y 3 de la variable de salida son peores que en las demás fases, no siendo valores muy buenos (incluso presentando menores a 0.1 para el valor 3 de la variable de salida) mostrando que el desempeño de estos modelos a la hora de predecir los valores 2 y 3 va a ser muy pobre.

Los modelos obtenidos tanto en la fase 2 como en la fase 4 presentan métricas de rendimiento bastante similares, siendo, dentro de los modelos analizados, buenas opciones a implementar. Sin embargo, se optó por el uso del modelo de random forest con datos normalizados y balanceados debido a que presenta unas métricas de rendimiento un poco mejores para todos los valores de la variable de salida, también, porque en sí el algoritmo de random forest es más resistente a posibles sesgos presentes en los datos, que un algoritmo de árbol de decisión (algoritmo que presenta mejores resultados en la fase 2) y al tener el tratamiento de normalización de datos, disminuye el riesgo de que en algún momento las distintas escalas de la información de entrada, pueda afectar el desempeño del modelo.

Una vez entrenado el modelo, los resultados para el valor 1 de la variable de salida, de evaluar este con los datos de test fueron:

Precisión	0.9972868217
F1-score	0.8500165180
Recall	0.7406447899

Tabla 9. Resultados del modelo entrenado para el valor 1 con datos de test

Para el valor dos de la variable de salida, los datos de test fueron:

Precisión	0.9893617021
F1-score	0.5891402715
Recall	0.4194587629

Tabla 10. Resultados del modelo entrenado para el valor 2 con datos de test.

Adicionalmente el modelo predijo correctamente 2573 de 3370 posibles correspondientes a al valor 1 de la variable de salida, y 651 de 667 posibles del valor 2 de la variable de salida, graficados en la siguiente figura.

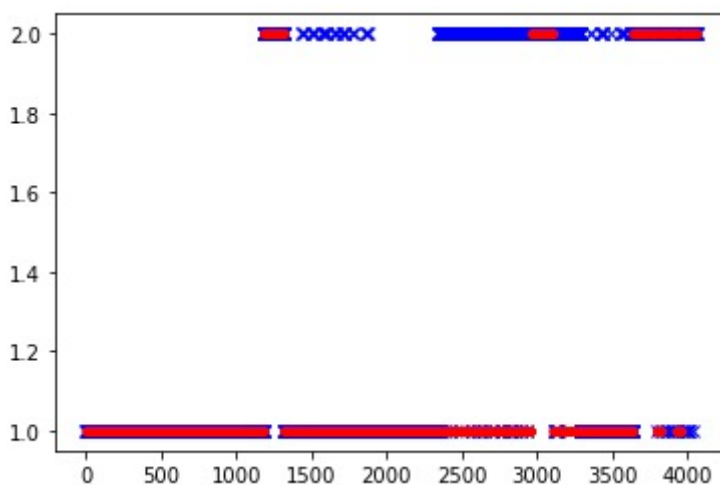


Figura 59. Resultados predichos contra resultados reales  
(fuente: Autor)

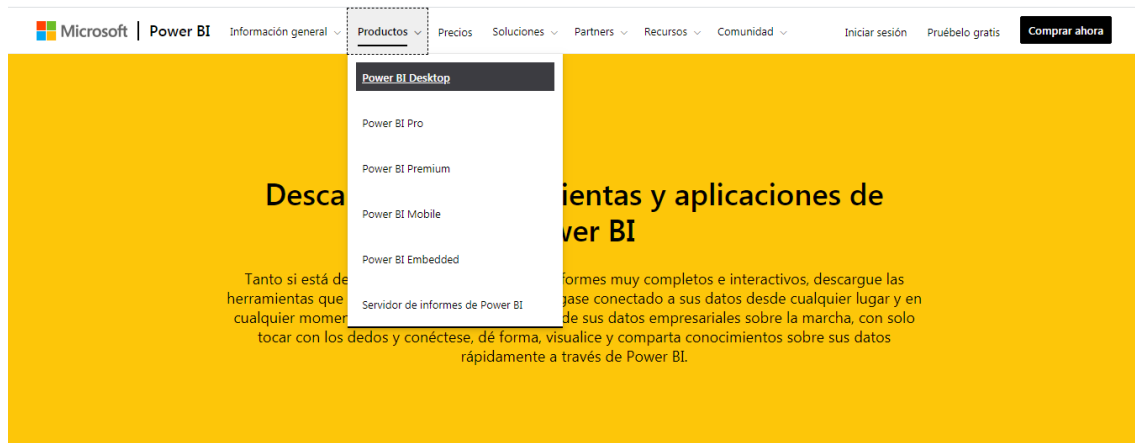
En la anterior figura, los resultados traslapados son donde los resultados predichos por el modelo (gráfica color azul) predijo correctamente los resultados reales (color rojo).

#### 4.1.5 Diseño del dashboard

En la visualización de resultados se realiza la presentación de los datos procesados por el algoritmo de machine learning. En primera instancia se pensó utilizar la plataforma Pentaho debido a la existencia de literatura y proyectos previamente desarrollados con esta herramienta, pero al notar que la plataforma Pentaho no posee motores de búsqueda de información de manera automatizada, se llega a la conclusión de optar por el uso de otra plataforma, para ello se seleccionó la Power Bi de Microsoft. Esta plataforma contiene 2 modos de operación uno desde la plataforma en la nube y otro desde el escritorio, siendo este último la herramienta base, pues en él es donde se elaboran algunos filtrados, ajustes y de la información, para luego ser archivados en la nube como informes, estos pueden ser visibles de manera remota sin dificultades. Un aspecto clave por el cual se decide el uso de esta plataforma es debido a que ofrece memberships gratuitas de prueba a correos de tipo empresarial o institucional, dando así un margen de trabajo hasta de 60 días con todas las herramientas posibles, además de esto da la oportunidad de gestionar de manera automática la actualización de los datos de trabajo, permitiendo acceder a información reciente de lo que ocurre, también se tiene la posibilidad de generar alertas que pueden ser enviadas en forma de correo electrónico.

##### 4.1.5.1 Instalación y configuración de Power BI

Se descarga del programa para ordenador directamente desde la página de Power Bi –“<https://powerbi.microsoft.com/es-es/desktop/>”



<https://powerbi.microsoft.com/es-es/desktop/>

Figura 60. Descarga Power Bi paso 1.  
(fuente: Autor)

En la parte superior derecha se encuentra el apartado de productos, en ello se despliega las opciones en la que seleccionaremos Power Bi desktop.

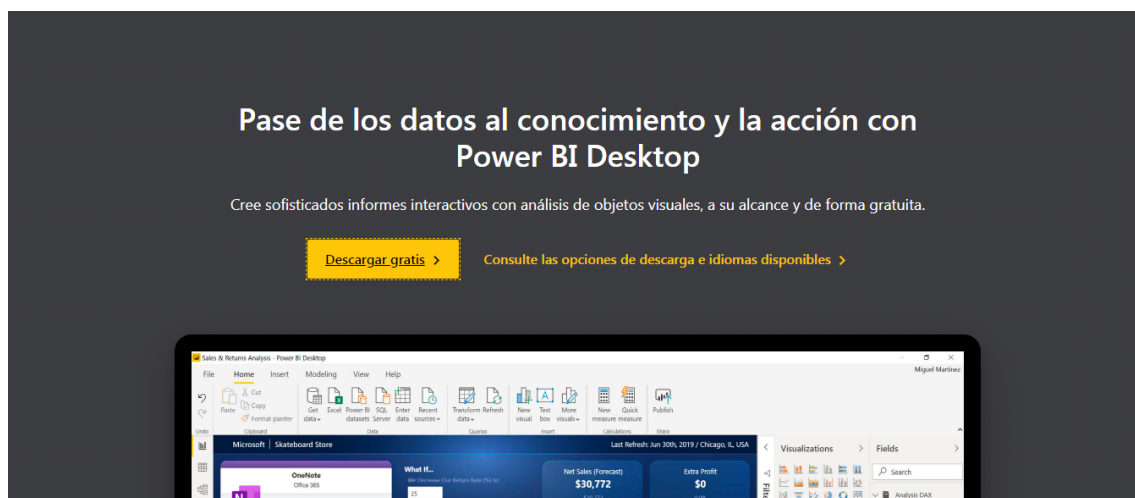
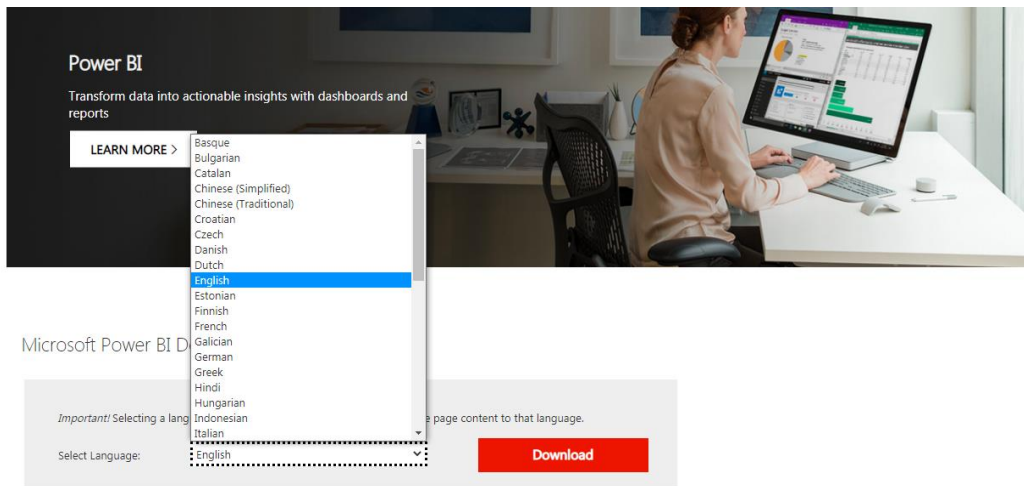
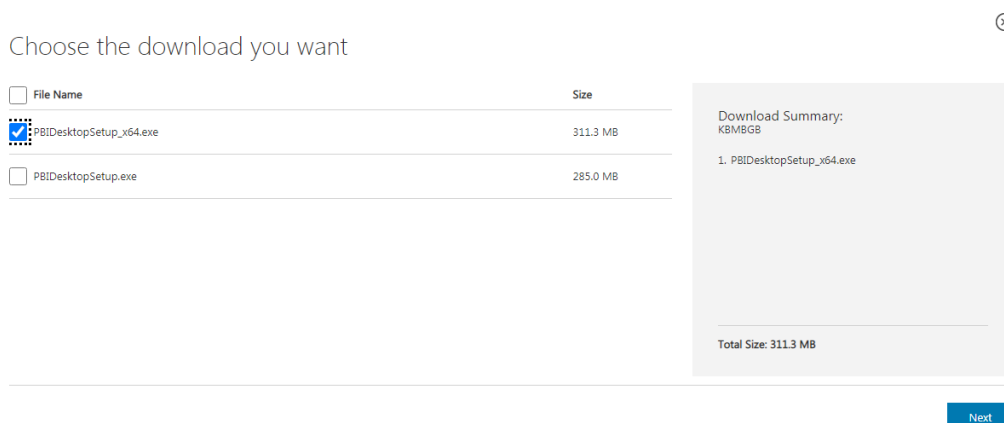


Figura 61. Descarga Power Bi paso 2. Zona de descargas Power Bi de manera gratuita  
(fuente: Autor)



**Figura 62. Descarga Powe Bi paso 3**  
(fuente: Autor)

En este apartado se elige idioma y se accede a la plataforma de descarga a través de Download



**Figura 63. Descarga Powe Bi paso 4**  
(fuente: Autor)

Se elige la versión de ordenador 64 o 32. Después de descargar el archivo PBIDesktopSetup\_x64seprocede a ejecutar como administrador



**Figura 64. Icono de instalación**  
(Fuente: Autor)

Después de instalar el servicio Power Bi Desktop, se puede abrir o se abrirá de manera automática el programa de escritorio. A continuación, una breve descripción de la entrada al programa

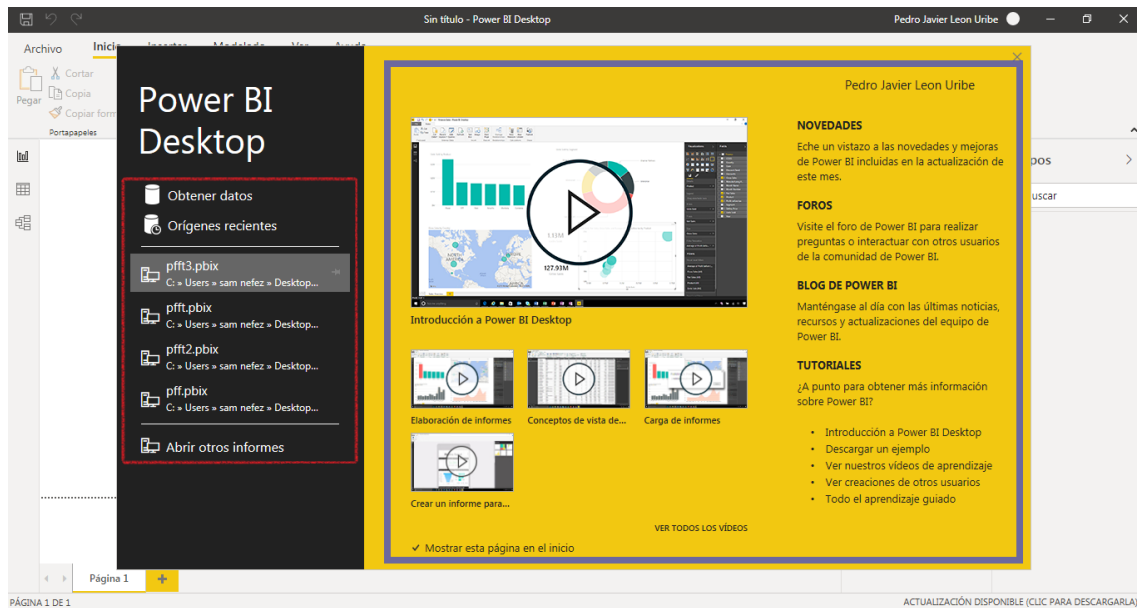


Figura 65. Pantalla de inicio Power Bi versión de escritorio (fuente: Autor)

La pantalla principal de inicio muestra varios aspectos, en el cuadro demarcado de rojo se encuentra: obtener datos este apartado es utilizado para tomar la ubicación madre del archivo a leer bien sea plataformas, servicios en línea o simplemente archivos como CSV, XML entre otros. En la siguiente figura se puede apreciar algunos formatos y plataformas que esta misma trae las siguientes opciones muestra informes realizados y debajo de ellos mismos la ubicación exacta del archivo madre y finalmente la opción de abrir informes que no aparecen en el listado reciente.

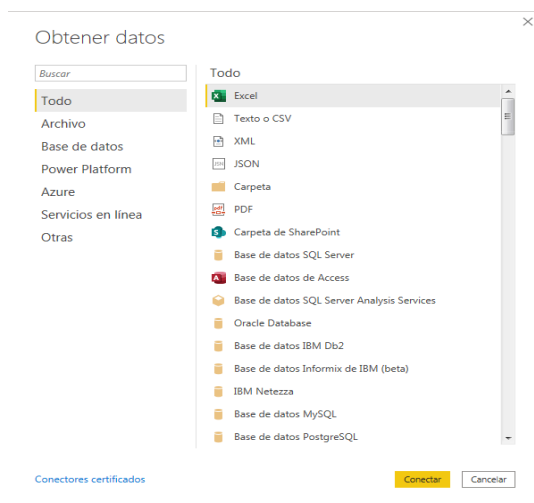


Figura 66. Obtención y origen de los datos a utilizar

(fuente: Autor)

Después de ubicar el archivo madre con los datos aparece un recuadro indicando la conexión con la ruta donde se encuentra el archivo deseado

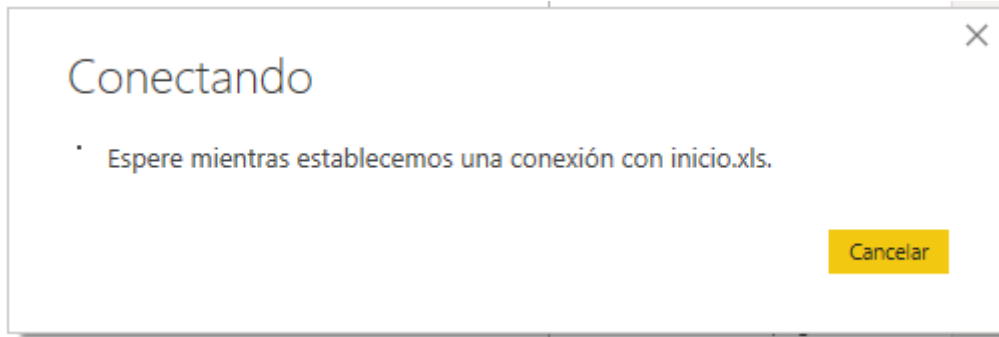


Figura 67. Recuadro de conexión  
(fuente: Autor)

Posterior a ello aparece un recuadro en el cual se pueden ajustar los datos de manera rápida y fácil, como pueden ser las normalizaciones de los mismos, filtraciones, e incluso ajustes de escritura.

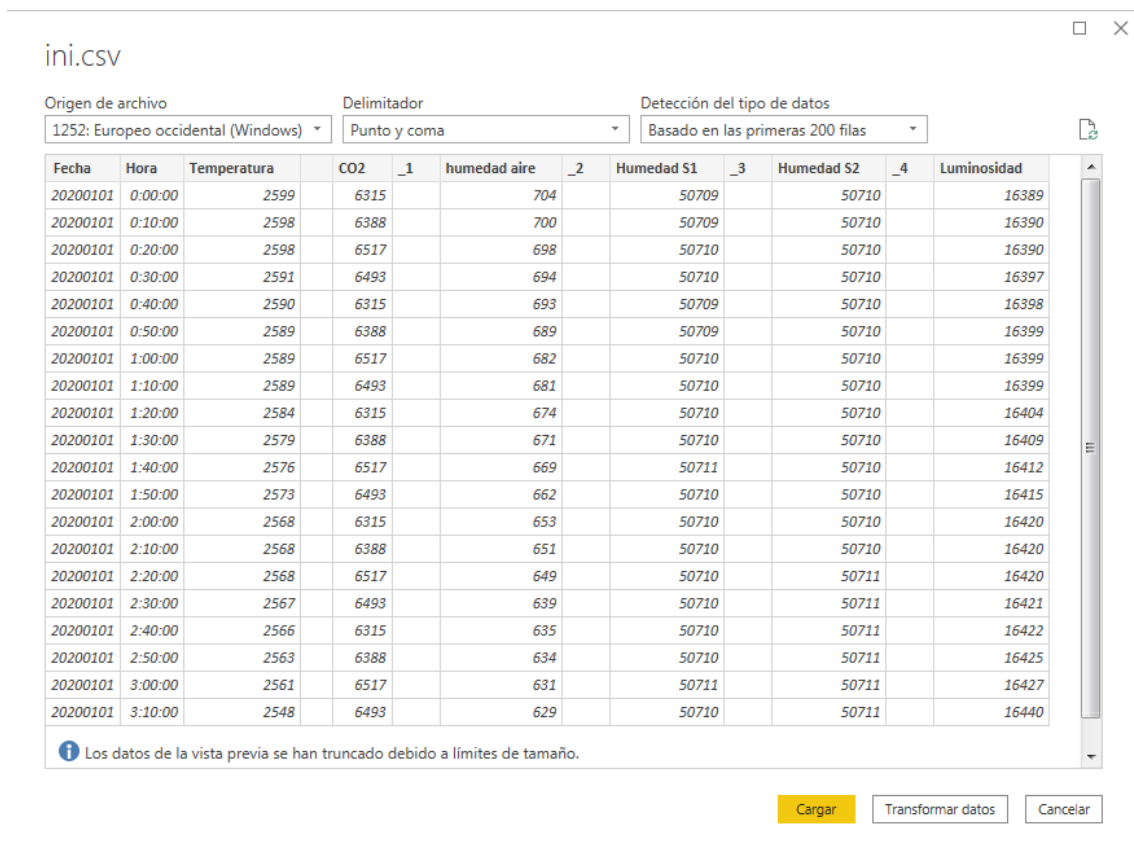
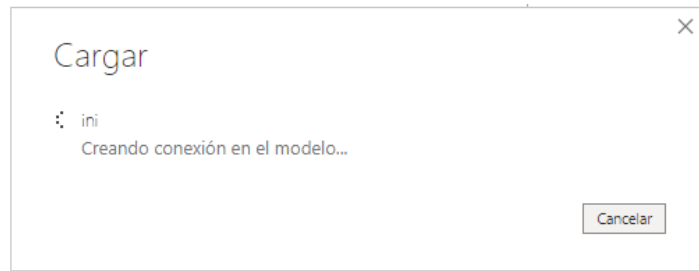


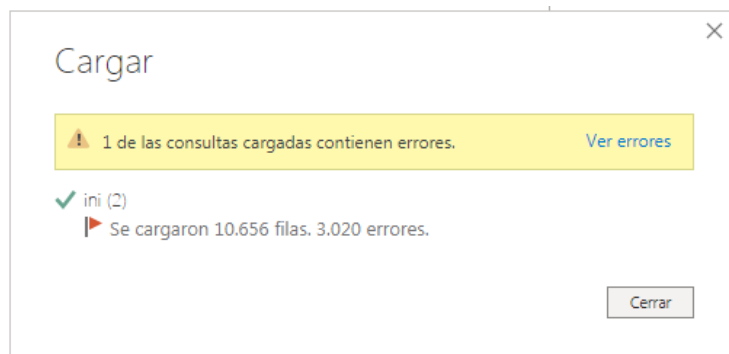
Figura 68. Ventana de ajuste de datos – operaciones básicas Excel  
(fuente: Autor)

Tan pronto se modifiquen los datos y estos sean cargados aparecerá la ventana de conexión directa con la ubicación de la información de trabajo



*Figura 69. Ventana de conexión con el punto de origen.  
(fuente: Autor)*

Finalmente aparece la información donde anuncia la conexión correcta con el punto de origen y anuncia algunos errores con respecto a escrituras.



*Figura 70. Ventana de conexión puesta a punto con el origen del archivo.  
(fuente: Autor)*

En la pantalla principal de trabajo se encuentran variedad de herramientas y elementos a disposición para ser utilizados de manera articulada

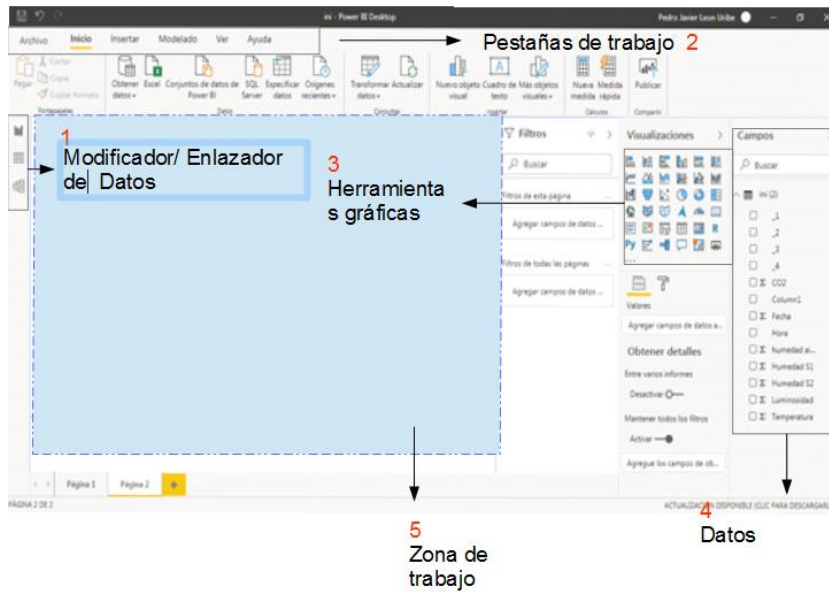


Figura 71. Plataforma de trabajo (fuente: Autor)

En la sección mostrada en el número 1 de la anterior figura, se encuentran los sub módulos de modificación de datos destinado para arreglos después de cargados, cruces de datos y enlaces entre los mismos, en la sección 2 las pestañas de trabajo como insertar y modelado, que permiten jerarquización del esquema, ediciones de texto y modificaciones rápida de datos, en el 3 se muestra el panel de herramientas graficas se encuentran 36 formas de representar y cruzar los datos, en la 4 se encuentra de manera organizada y por columnas los datos ya transformados y listos para ser utilizados, en la zona 5 se diseña el estilo del informe tan solo arrastrando las piezas del panel gráfico.

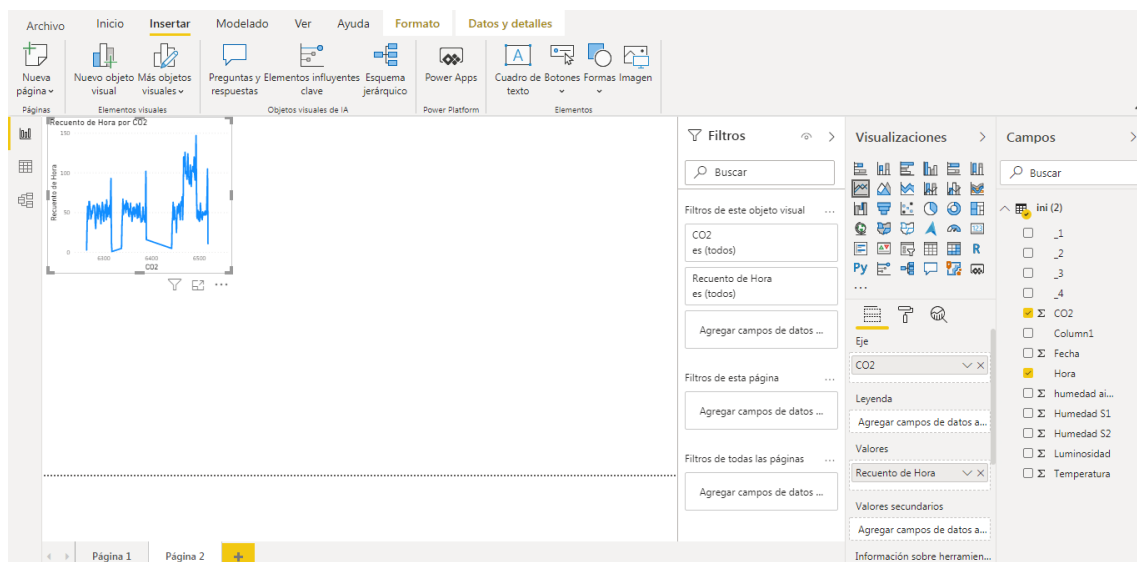


Figura 72. Ejemplo de CO2 vs Hora (fuente: Autor)

## Enlace, automatización y actualización de datos Power Bi

En la plataforma virtual Power Bi se encuentran varias herramientas que no están en la versión de escritorio, una de ellas es poder enlazar la base de datos y actualizarle periódicamente (lapsos de 10 minutos), de tal forma que la visualización remota este prácticamente en tiempo real.

A continuación, los pasos para generar el Gateway directamente a la base de datos:

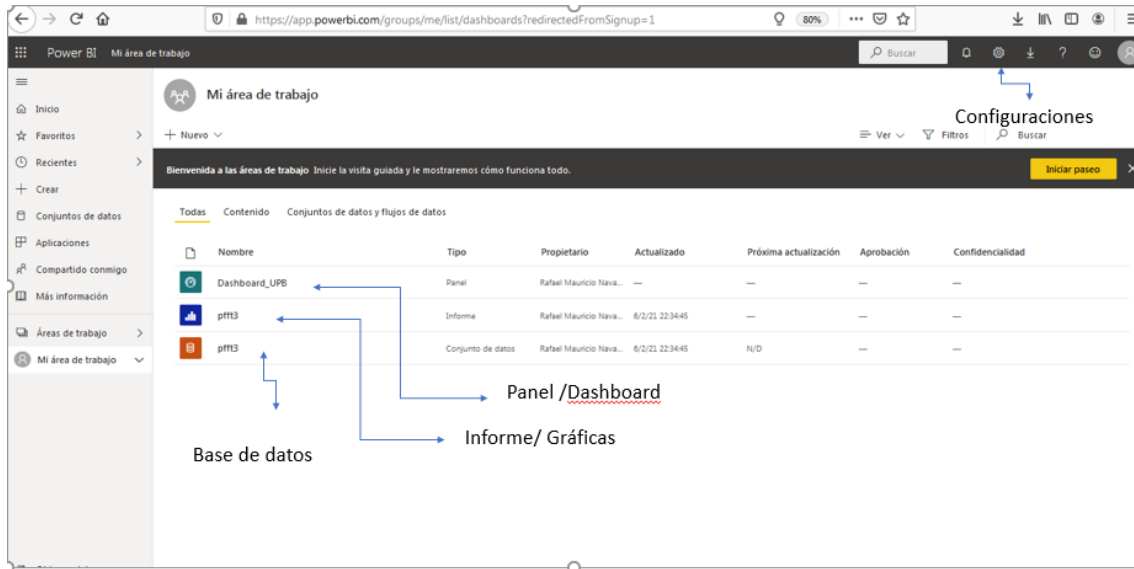



Figura 73. Área de trabajo y herramientas de la plataforma.  
(fuente: Autor)

Se da click en el ícono  > Configuración

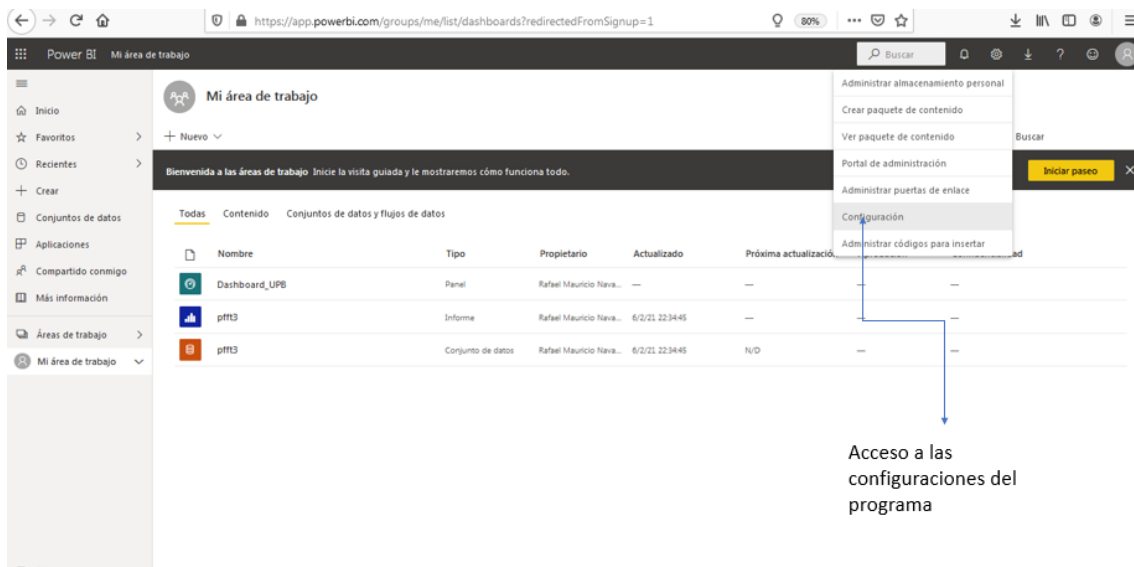


Figura 74. Accesos a las configuraciones y programación de Power Bi web

(fuente: Autor)

Se da click en Conjunto de datos

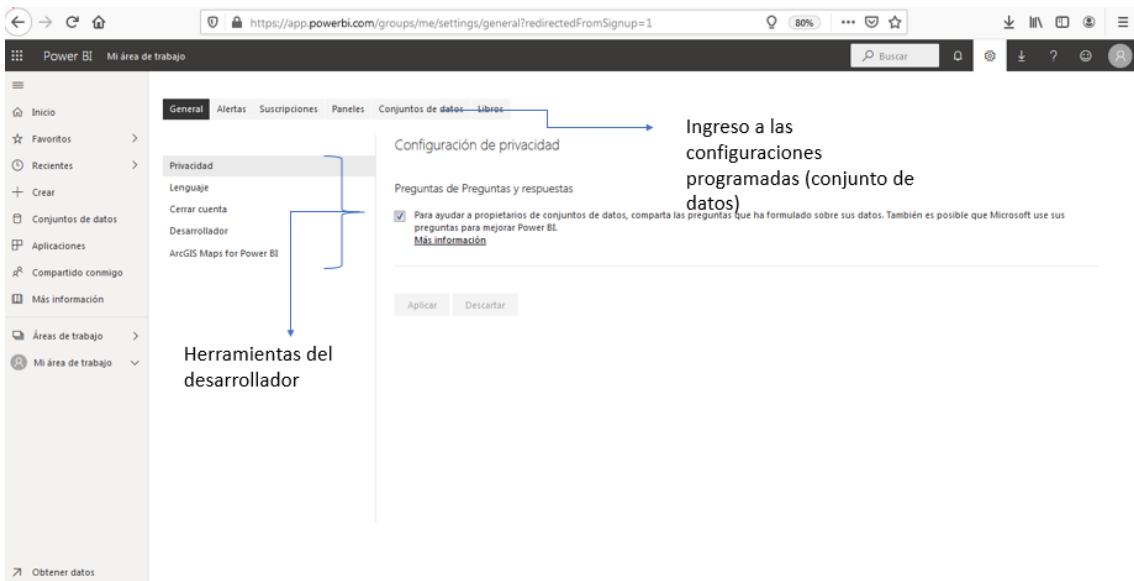


Figura 75. Ruta de mapeo y configuraciones de los datos adjuntos al panel/Dashboard (fuente: Autor)

Se selecciona el informe de gráfico, en la sección de Conexión de puerta de enlace se habilita la opción Usar una puerta de enlace de datos, y se instala el Gateway recomendado

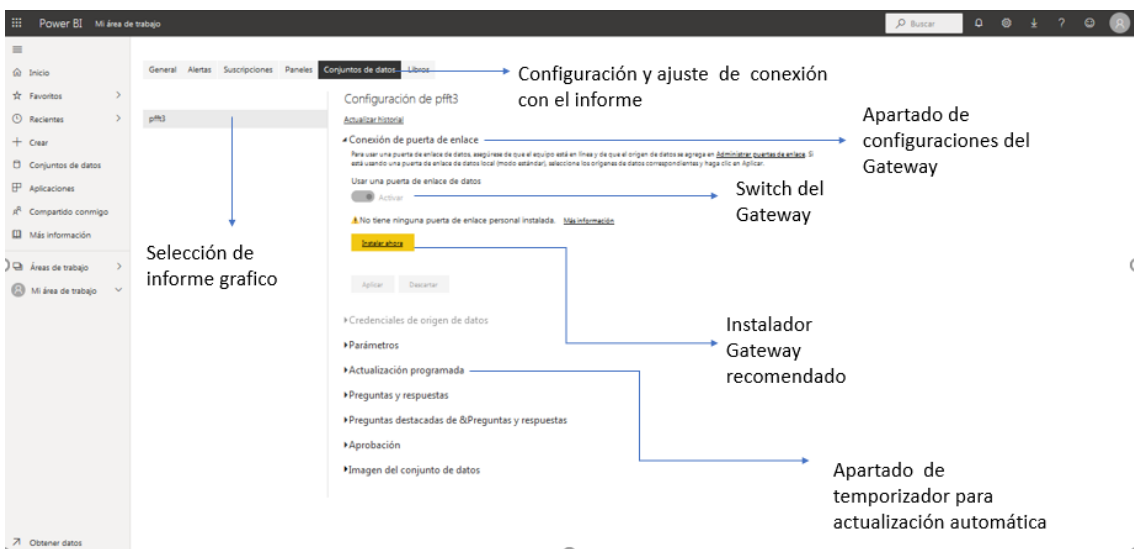
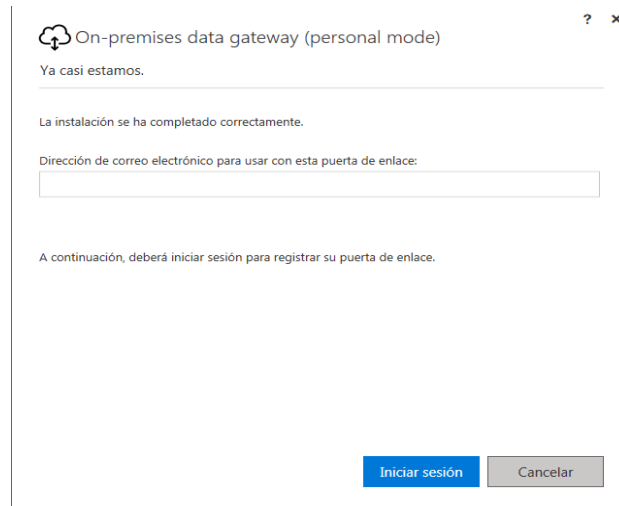


Figura 76. Parámetros y herramientas para actualización remota automática. (fuente: Autor)

**Nota :** se recomienda utilizar el Gateway que otorga Microsoft debido a su fácil configuración puesto que solo deberemos incluir la cuenta de trabajo asociada.

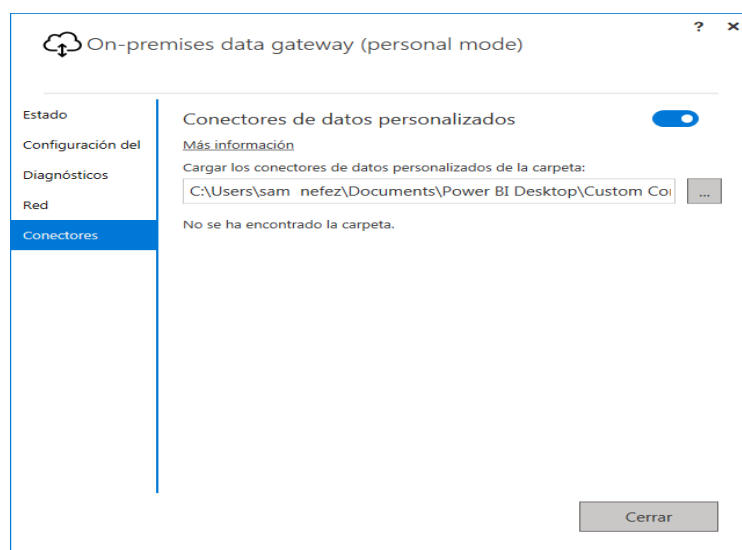
Una vez se empieza a instalar el Gateway, se debe introducir una dirección de correo electrónico para realizar la conexión, en este caso se usa el correo institucional de la universidad, ya que estos proporcionan un tiempo de prueba de la herramienta de 60 días



*Figura 77. Credencial de acceso.  
(fuente: Autor)*

Una vez introducida la dirección de correo, la configuración se hace de forma automática.

El siguiente paso es configurar la carpeta desde la cual se carga el archivo con los datos, como se muestra en la siguiente figura.



*Figura 78. Puerta de enlace lista para usarse.*

(fuente: Autor)

Una vez configurado, la ventana principal debe verse como en la siguiente figura.

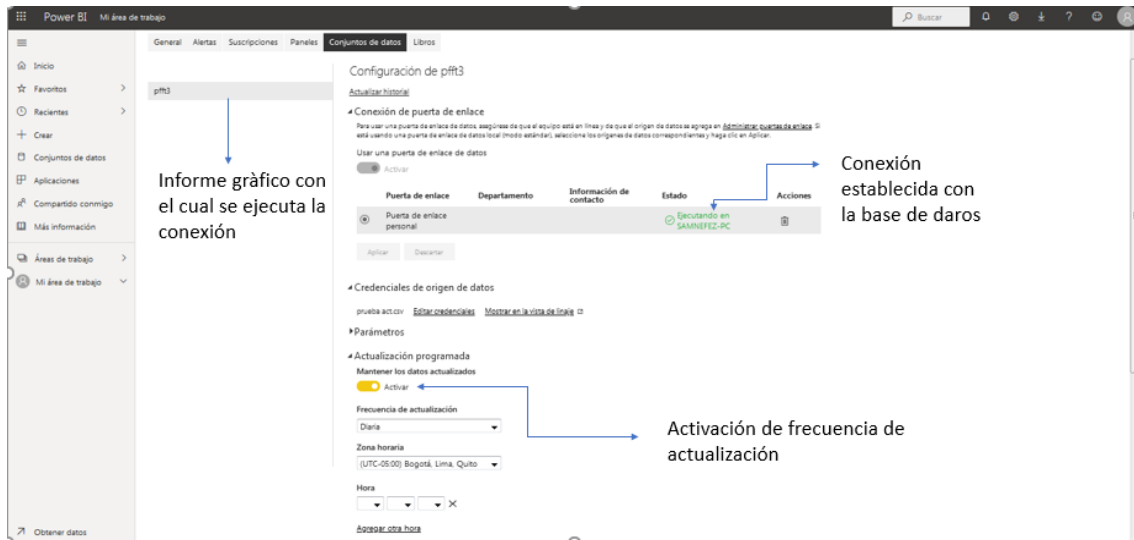


Figura 79. Configuración de la puerta de enlace y frecuencia de actualización automática.

(fuente: Autor)

**Nota:** Se recomienda utilizar ordenadores o servidores que no se apaguen pues al finalizar el equipo se pierde el enlace y deberá volverse a reconfigurar incluso su periodicidad de trabajo.

## Visualización de datos

La visualización de los datos se hace mediante tres pestañas en la herramienta de escritorio de Power BI, la primera muestra información general del sistema, como la gráfica de la Temperatura por Hora, el Promedio de la temperatura por día, el promedio de humedad por día. Esta pestaña se muestra en la figura 80.

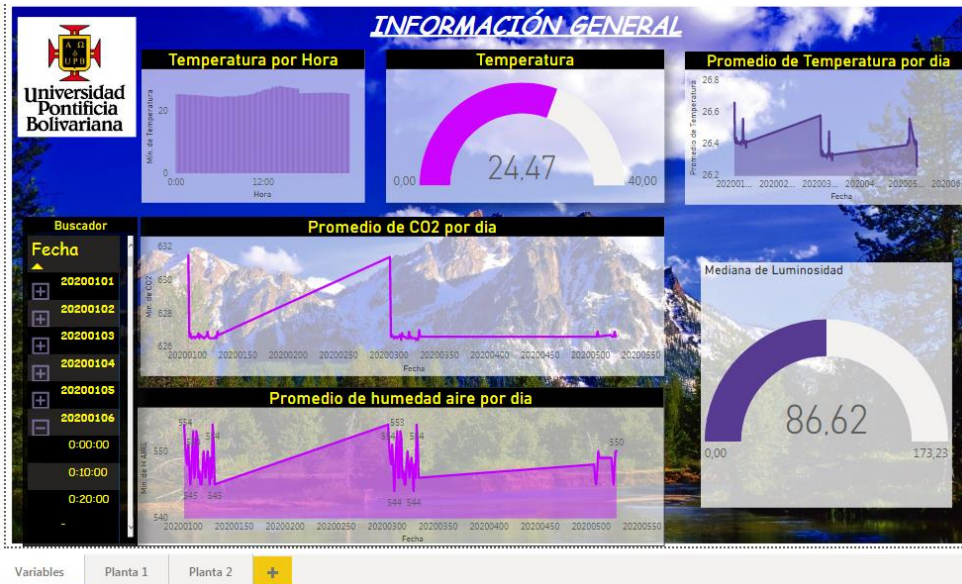


Figura 80 Información general de cada una de las variables visualizada por fecha (fuente: Autor).

La segunda pestaña se encarga de mostrar la información correspondiente al sujeto de pruebas 1 (temperatura, CO<sub>2</sub>, Humedad en el suelo) así como si se encuentra viva o enferma/muerta. La pestaña se observa en la figura 81.

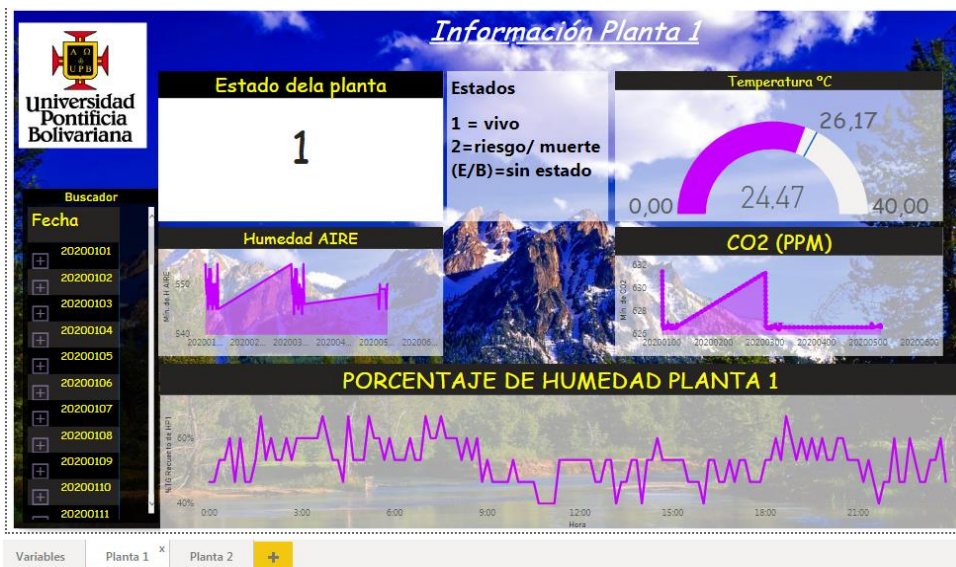


Figura 81. Información correspondiente al sujeto de pruebas 1 (fuente: Autor)

La tercera pestaña se encarga de mostrar la información correspondiente al sujeto de pruebas 2 (temperatura, CO<sub>2</sub>, Humedad en el suelo) así como si se encuentra viva o enferma/muerta. La pestaña se observa en la figura 82.

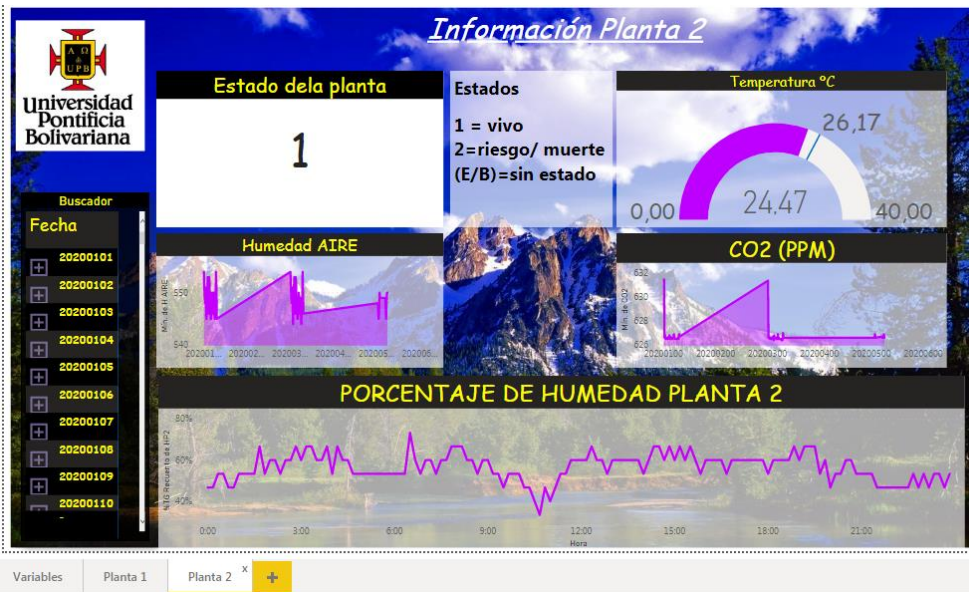


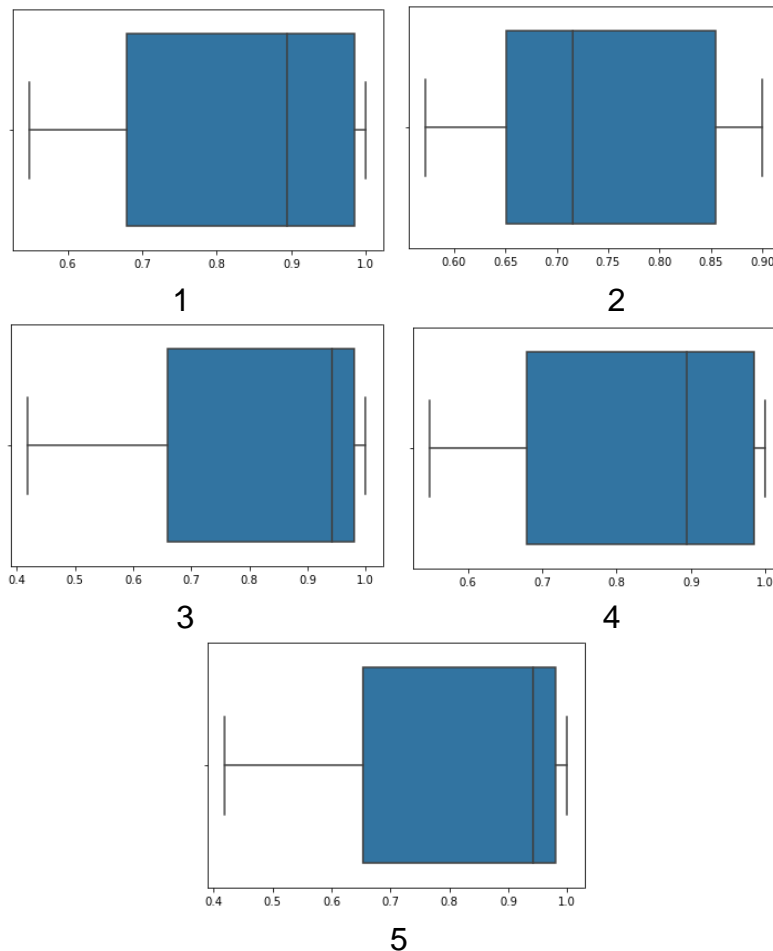
Figura 82. Información correspondiente al sujeto de pruebas 2 (fuente: Autor)

## CAPITULO 5

### RESULTADOS

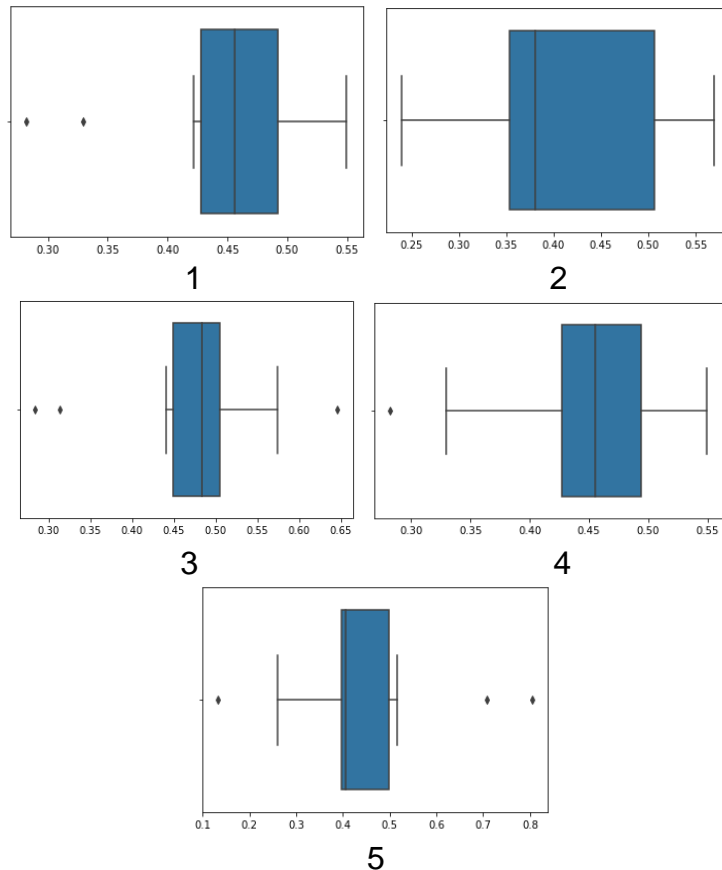
Los resultados obtenidos en la sección 4.1.4.4. Evaluación, se ratifican mediante el uso de los diagramas de cajas mostrados a continuación. En los cuales se muestra el comportamiento de la métrica f1-score con distintos datasets pequeños. Todo esto es posible gracias a la aplicación de crossvalidation con el dataset de entrenamiento.

En las figuras 31, 32 y 33, el primer diagrama de cajas corresponde al modelo de árbol de decisión, el segundo corresponde al random forest, el tercero corresponde al random forest con datos balanceados, el cuarto al árbol de decisión con datos normalizados y el quinto al random forest con datos normalizados y balanceados.



*Figura 83. Diagramas de cajas correspondiente a los modelos desarrollados en la fase 2 del desarrollo del modelo (fuente: Autor)*

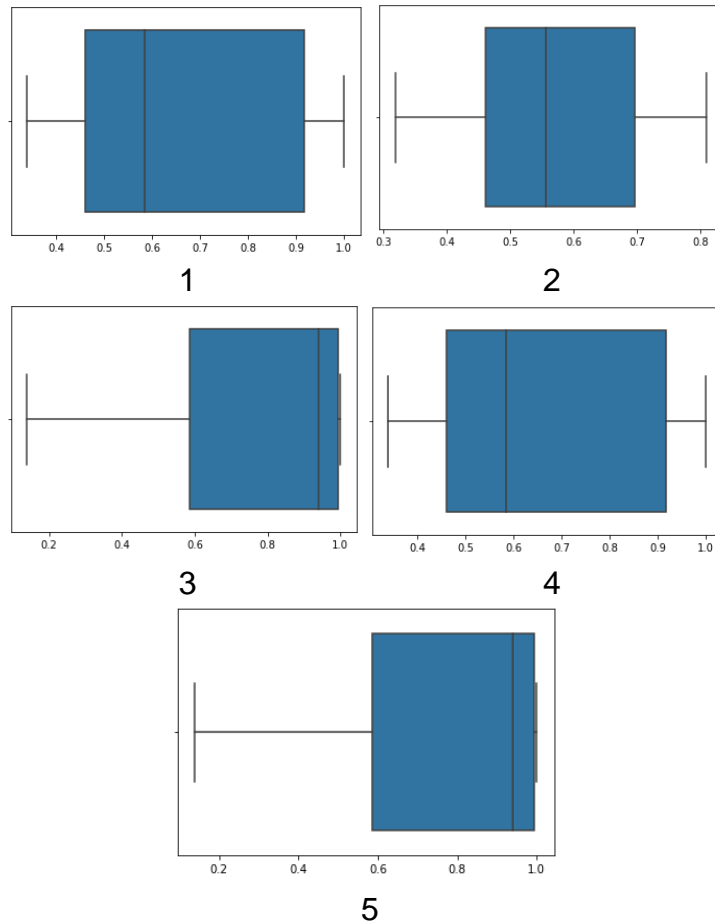
El comportamiento de los modelos, de acuerdo al f1-score, en la fase 2 es bueno, ya que en el modelo que presenta peor rendimiento, el 50% central de los valores se encuentran entre 0.65 y 0.85, con una mediana ligeramente superior a 0.7; y en el modelo con mejor rendimiento, estos se encuentran en valores muy cercanos a 0.7 y valores muy cercanos a 1, con una mediana ligeramente inferior a 0.95.



*Figura 84. Diagramas de cajas correspondiente a los modelos desarrollados en la fase 3 del desarrollo del modelo (fuente: Autor)*

En los diagramas de cajas de los modelos de la fase 3 se ratifica que son los que presentan peor comportamiento, debido a que ninguno tiene el 50% de datos centrales en un rango que sea superior a 0.5, y con una mediana más cercana de 0.4 que de 0.5. Además, cuatro de los cinco modelos presentan valores extremos en los diagramas (valores que son o muy altos o muy bajos respecto a los demás).

El efecto negativo que tuvo el eliminar la variable con mayor peso para la predicción de los modelos es visible, sobre todo si se comparan con los diagramas de cajas de los modelos de las fases 2 y 4.



*Figura 85. Diagramas de cajas correspondiente a los modelos desarrollados en la fase 4 del desarrollo del modelo (fuente: Autor)*

En los diagramas de los modelos correspondientes a la fase 4 se observa que estos tienen al 50% de sus datos centrales en un rango más amplio que el mismo modelo, pero implementado en las fases anteriores, pero sus valores y su media están más cerca de 1, lo que hace que sea más probable la obtención de un valor de f1-score más alto si se utiliza el mismo modelo, pero con las condiciones de esta fase. Teniendo en cuenta lo previamente mencionado, y analizando únicamente los modelos de esta fase, se observa que los modelos con mejor desempeño son los modelos correspondientes a los diagramas 3 y 5 correspondientes a los random forest con datos balanceados y random forest con datos balanceados y normalizados, respectivamente, por lo que se puede decir que la elección del random forest con datos balanceados y normalizados como modelo a implementar fue acertada.

## CAPITULO 6

### CONCLUSIONES Y RECOMENDACIONES

#### 6.1 Conclusiones

Se pudo llevar a cabo de forma satisfactoria el diseño e implementación de un prototipo de monitoreo de cultivos de hierbabuena usando una red de sensores, y minería de datos.

Se logró la puesta en funcionamiento de la infraestructura física destinada para el proyecto, logrando hacer transferencia de datos desde los PLCs hasta una computadora, vía servidor FTP

Se realizó el monitoreo de las variables de propuestas en los objetivos específicos (CO<sub>2</sub>, temperatura y humedad relativa del aire y la humedad del suelo en el cultivo de hierbabuena) y se adicionaron otras como el número de hojas de cada sujeto y el sujeto al que pertenece cada variable.

Se diseñó e implementó un modelo de Machine Learning cuyo rendimiento se consideró como satisfactorio de acuerdo a las métricas de rendimiento empleadas, por parte del equipo que lo desarrolló.

Se comprobó la viabilidad en la aplicación modelos de Machine Learning para el monitoreo del estado de cultivos.

Se logró la implementación y el enlace de forma satisfactoria, de una HMI para la visualización de los datos previamente obtenidos.

#### 6.2 Recomendaciones

Se recomienda para futuros proyectos, el uso de especies de plantas más resistentes, o realizar el experimento en un ambiente mucho más controlado, como puede ser un invernadero.

Se recomienda el uso de datasets balanceados o con un menor desbalance.

Se recomienda experimentar con otros algoritmos de Machine Learning.

## BIBLIOGRAFÍA

- [1] H. Kees, *Ancient Egypt: A Cultural Topography*. University of Chicago Press., 1961.
- [2] J. Janick, "Ancient Egyptian Agriculture and the Origins of Horticulture," West Lafayette, Indiana. [Online]. Available: [https://www.hort.purdue.edu/newcrop/Hort\\_306/text/lec06.pdf](https://www.hort.purdue.edu/newcrop/Hort_306/text/lec06.pdf).
- [3] K. D. M. Snell, *Annals of the Labouring Poor, Social Change and Agrarian England 1660–1900*. Cambridge University Press, 1985.
- [4] R. Tannahill, *The fine art of food*. 1970.
- [5] A. Bhargava and S. Srivastava, "Human Civilization and Agriculture," 2019, pp. 1–27.
- [6] S. Cunningham and G. Holmes, "Developing innovative applications in agriculture using data mining," Hamilton, New Zealand. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.41.8457&rep=rep1&type=pdf>.
- [7] S. Delerce *et al.*, "Assessing Weather-Yield Relationships in Rice at Local Scale Using Data Mining Approaches," 2016. [Online]. Available: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0161620#sec016>.
- [8] F. Muñoz, *Plantas medicinales y aromáticas: estudio, cultivo y procesado*. Mundi-Prensa Libros, 2002.
- [9] M. Sánchez, "¿Qué son las plantas herbáceas?" <https://www.jardineriaon.com/que-son-las-plantas-herbaceas.html> (accessed Nov. 23, 2020).
- [10] RAE, "Hierbabuena." <http://dle.rae.es/?id=KMaNjyn> (accessed Nov. 23, 2020).
- [11] RAE, "Rizoma." <http://dle.rae.es/?id=WX0RW2W> (accessed Nov. 23, 2020).
- [12] J. Japón Quintero, *Cultivo del perejil y de la hierbabuena*. Ministerio de Agricultura, Pesca y Alimentación, 1985.
- [13] RAE, "Variable." <https://dle.rae.es/variable> (accessed Nov. 23, 2020).
- [14] Universidad de Nebrija, "Termodinámica Temperatura y calor." [https://www.nebrija.es/~cmalagon/Fisica\\_Aplicada/transparencias/02-Termodinamica/07\\_-\\_Termodinamica.pdf](https://www.nebrija.es/~cmalagon/Fisica_Aplicada/transparencias/02-Termodinamica/07_-_Termodinamica.pdf) (accessed Nov. 23, 2020).
- [15] W. Jansen, "Dióxido de carbono: Información básica." [https://www.scienceinschool.org/sites/default/files/teaserMaterial/issue20\\_CO2\\_background\\_spanish.pdf](https://www.scienceinschool.org/sites/default/files/teaserMaterial/issue20_CO2_background_spanish.pdf) (accessed Nov. 23, 2020).
- [16] C. Meruane and R. Garreaud, "Determinación de Humedad en la Atmósfera," 2006. [Online]. Available: <http://mct.dgf.uchile.cl/AREAS/modHR.pdf>.
- [17] "¿Que es Smart Farming?" .
- [18] A. WALTER, R. FINGER, and N. HUBER, Robert. BUCHMANN, "Smart farming is key to developing sustainable agriculture." [www.pnas.org/content/114/24/6148.full](http://www.pnas.org/content/114/24/6148.full) (accessed Nov. 23, 2020).
- [19] Universidad Nacional Autónoma de México., "Capítulo 3: Protocolos de

- comunicación.”  
<http://www.ptolomeo.unam.mx:8080/xmlui/bitstream/handle/132.248.52.100/734/A6.pdf?sequence=6> (accessed Nov. 23, 2020).
- [20] Radio Observatorio de Jicamarca, “Protocolos de comunicación utilizados en sistemas de transmisores ROJ.”  
[http://jro.igp.gob.pe/teps/pdf/2009/ryaya\\_ProtocolosTX.pdf](http://jro.igp.gob.pe/teps/pdf/2009/ryaya_ProtocolosTX.pdf) (accessed Nov. 23, 2020).
- [21] A. Ruiz, A. Barandica, and F. Guerrero, “Implementación de una Red MODBUS/TCP,” *Revista Ingeniería y Competitividad*, pp. 35–34, 2011.
- [22] Universidad de Sevilla, “Anexo 1: Protocolo MODBUS.”  
<http://bibing.us.es/proyectos/abreproy/11676/fichero/Volumen+1%252FAnexo.pdf> (accessed Nov. 23, 2020).
- [23] “Curso de administrador de servicios de Internet/ Extranet/ Intranet.”  
[http://www.juntadeandalucia.es/empleo/recursos/material\\_didactico/especialidades/materialdidactico\\_administrador\\_servidores/Content/2-redes\\_tcp/3-LosProtocolosTCP-IP.pdf](http://www.juntadeandalucia.es/empleo/recursos/material_didactico/especialidades/materialdidactico_administrador_servidores/Content/2-redes_tcp/3-LosProtocolosTCP-IP.pdf) (accessed Nov. 23, 2020).
- [24] J. Martin and D. Ruiz, “Informe técnico: Protocolo Zigbee IEEE 802.15.4.”  
[https://rua.ua.es/dspace/bitstream/10045/1109/7/Informe\\_ZigBee.pdf](https://rua.ua.es/dspace/bitstream/10045/1109/7/Informe_ZigBee.pdf) (accessed Nov. 23, 2020).
- [25] W. Pinto, “Monitoreo de cultivos con redes de sensores xbee, arduino, y dispositivos de medicion de suelos.”  
<http://repositorio.utp.edu.co/dspace/handle/11059/5712> (accessed Nov. 23, 2020).
- [26] “Types of Controls and Controllers - A ThomasNet Buying Guide.”  
<https://www.thomasnet.com/articles/instruments-controls/types-of-controllers/> (accessed Nov. 23, 2020).
- [27] Advanticsys, “Controlador programable MPC-134.”  
<https://www.advanticsys.com/shop/mpc134-p-35.html>.
- [28] Advanticsys, “Controlador programable UCM-316.”  
<https://www.advanticsys.com/shop/ucm316-p-50.html>.
- [29] Advanticsys, “Sensor emisor de datos IAQM-THCO2.”  
<https://www.advanticsys.com/shop/iaqmthco2-p-48.html>.
- [30] Advanticsys, “Puente inalámbrico DM-124.”  
<https://www.advanticsys.com/shop/dm124-p-47.html>.
- [31] “MEDIR LA HUMEDAD DEL SUELO CON ARDUINO E HIGRÓMETRO FC-28.” <https://www.luisllamas.es/arduino-humedad-suelo-fc-28/> (accessed Aug. 07, 2021).
- [32] “SENSOR DE HUMEDAD DE SUELO FC28 - HIGRÓMETRO ARDUINO.” <https://www.arcaelectronica.com/blogs/tutoriales/sensor-de-humedad-de-suelo-fc28-higrometro-arduino> (accessed Aug. 07, 2021).
- [33] “MEDIR NIVEL DE LUZ CON ARDUINO Y FOTORESISTENCIA LDR (GL55).” <https://www.luisllamas.es/medir-nivel-luz-con-arduino-y-fotoresistencia-ldr/> (accessed Aug. 07, 2021).
- [34] RAE, “Software.” <https://dle.rae.es/software> (accessed Nov. 23, 2020).
- [35] “Power BI, ¿qué es Power BI?” <https://www.makesoft.es/powr-bi-que-es-power-bi/> (accessed Nov. 23, 2020).
- [36] Python Software Foundation, “What is Python?”  
<https://docs.python.org/3/faq/general.html#what-is-python>.

- [37] Universitat Politècnica de València, “Bioinformatics at COMAV.” <https://bioinf.comav.upv.es/courses/linux/python/pandas.html> (accessed May 14, 2021).
- [38] The SciPy community, “What is Numpy?” <https://numpy.org/doc/stable/user/whatisnumpy.html> (accessed May 14, 2021).
- [39] Universidad de Alcalá, “Scikit-learn, herramienta básica para el data science en Python.” <https://www.master-data-scientist.com/scikit-learn-data-science/> (accessed May 14, 2021).
- [40] P. S. Foundation, “Pickle — Serialización de objetos Python.” <https://docs.python.org/es/dev/library/pickle.html#:~:text=> (accessed May 14, 2021).
- [41] B. Beltrán, “Minería de datos.”
- [42] O. Bousque, U. Von Luxburg, and G. Rätsch, Eds., *Advanced Lectures on Machine Learning*. .
- [43] E. Alpaydin, *Introduction to Machine Learning Adaptive computation and machine learning*. 2004.
- [44] T. Mitchell, *Machine Learning*. .
- [45] P. Norvig and S. Russell, *Artificial Intelligence: A Modern Approach*, 3rd ed. .
- [46] A. Gerón, *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow*, 2nd ed. .
- [47] C. Kingsford and S. Salzberg, “What are decision trees?,” *Nature*, [Online]. Available: <https://www.nature.com/articles/nbt0908-1011>.
- [48] J. R. Quinlan, “Induction of decision trees,” *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986, doi: 10.1007/BF00116251.
- [49] “Método de división de nodos: Cart Clasificación.” <https://support.minitab.com/es-mx/minitab/19/help-and-how-to/statistical-modeling/predictive-analytics/how-to/cart-classification/methods-and-formulas/node-splitting-methods/> (accessed Apr. 26, 2021).
- [50] IBM, “Nodo CHAID.” <https://www.ibm.com/docs/es/spss-modeler/SaaS?topic=nodes-chaid-node> (accessed Apr. 25, 2021).
- [51] Amazon, “Amazon Machine Learning Developer Guide.” <https://docs.aws.amazon.com/machine-learning/latest/dg/machinelearning-dg.pdf#model-fit-underfitting-vs-overfitting> (accessed May 14, 2021).
- [52] T. K. Ho and Laboratories AT&T Bell, “Random Decision Forests,” 1995. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/598994>.
- [53] L. Breiman, “Bagging predictors,” *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996, doi: 10.1007/BF00058655.
- [54] “Preprocessing data.” <https://scikit-learn.org/stable/modules/preprocessing.html> (accessed Aug. 07, 2021).
- [55] “sklearn.preprocessing.StandardScaler.” <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html> (accessed Aug. 07, 2021).
- [56] M. Sandri and P. Zuccolotto, “A Bias Correction Algorithm for the Gini Variable Importance Measure in Classification Trees.”
- [57] D. Berrar, “Cross-validation,” [Online]. Available:

- [https://www.researchgate.net/profile/Daniel-Berrar/publication/324701535\\_Cross-Validation/links/5cb4209c92851c8d22ec4349/Cross-Validation.pdf](https://www.researchgate.net/profile/Daniel-Berrar/publication/324701535_Cross-Validation/links/5cb4209c92851c8d22ec4349/Cross-Validation.pdf).
- [58] “sklearn.ensemble.RandomForestClassifier.” [https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html?highlight=feature\\_importances\\_#sklearn.ensemble.RandomForestClassifier.feature\\_importances\\_](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html?highlight=feature_importances_#sklearn.ensemble.RandomForestClassifier.feature_importances_) (accessed Aug. 07, 2021).

Floridablanca, 25 de Agosto de 2021

Señores  
COMITÉ DE TRABAJOS DE GRADO  
Facultad de Ingeniería Eléctrica y Electrónica

Cordial Saludo,

Por medio de la presente informo que los estudiantes **Rafael Mauricio Navarro Casadiegos y Pedro Javier León Uribe**, han realizado el informe final del proyecto titulado: **“Prototipo de monitoreo de cultivos de plantas aromáticas usado redes de sensores inalámbricos y minería de datos”**, y doy mi aprobación para que lo entreguen, pues cumple con los requisitos de un informe final de un proyecto de grado. Esta entrega se hace con el fin de que se le asignen los evaluadores y su posterior sustentación una vez haya sido aprobado.

Agradezco su valiosa colaboración.

Atentamente,



JHON JAIRO PADILLA AGUILAR  
Director Proyecto