

UNIVERSIDAD POLITÉCNICA ESTATAL DEL CARCHI



FACULTAD DE INDUSTRIAS AGROPECUARIAS Y CIENCIAS AMBIENTALES

CARRERA DE INGENIERÍA EN INFORMÁTICA

Tema: “Minería de datos para series temporales y su aplicación en las precipitaciones pluviales en la Finca Experimental San Francisco de Huaca”

Trabajo de titulación previa la obtención del
título de Ingeniero en Informática

AUTOR: Montenegro Argoti Carlos Fernando

TUTOR: Yandún Velasteguí Marco Antonio, MSc.

Tulcán, 2021

CERTIFICADO JURADO EXAMINADOR

Certificamos que el estudiante Montenegro Argoti Carlos Fernando con el número de cédula 0401754965 ha elaborado el trabajo de titulación: “Minería de datos para series temporales y su aplicación en las precipitaciones pluviales en la Finca Experimental San Francisco de Huaca”

Este trabajo se sujeta a las normas y metodología dispuesta en el Reglamento de Titulación, Sustentación e Incorporación de la UPEC, por lo tanto, autorizamos la presentación de la sustentación para la calificación respectiva.

f.....

Yandún Velasteguí Marco Antonio, MSc

TUTOR

Tulcán, septiembre de 2021

AUTORÍA DE TRABAJO

El presente trabajo de titulación constituye requisito previo para la obtención del título de **Ingeniero** en la Carrera de ingeniería en informática de la Facultad de Industrias Agropecuarias y Ciencias Ambientales

Yo, Montenegro Argoti Carlos Fernando con cédula de identidad número 0401754965 declaro: que la investigación es absolutamente original, auténtica, personal y los resultados y conclusiones a los que he llegado son de mi absoluta responsabilidad.



f.....

Montenegro Argoti Carlos Fernando

AUTOR

Tulcán, septiembre de 2021

ACTA DE CESIÓN DE DERECHOS DEL TRABAJO DE TITULACIÓN

Yo, Montenegro Argoti Carlos Fernando declaro ser autor/a de los criterios emitidos en el trabajo de investigación: “Minería de datos para series temporales y su aplicación en las precipitaciones pluviales en la Finca Experimental San Francisco de Huaca” y eximo expresamente a la Universidad Politécnica Estatal del Carchi y a sus representantes legales de posibles reclamos o acciones legales.



f.....

Montenegro Argoti Carlos Fernando

AUTOR

Tulcán, septiembre de 2021

AGRADECIMIENTO

Agradecimiento a la Universidad Politécnica Estatal del Carchi por abrirme sus puertas y permitirme culminar mis estudios de la mejor manera, conociendo a buenos docentes compañeros y amigos, a mi madre, mi hermana y toda mi familia quienes han estado pendiente de mi durante todo este proceso académico, al MSc. Marco Yandún quien con su conocimiento y paciencia fue de gran ayuda para realizar el presente proyecto investigativo, a mis docentes quienes demostraron su profesionalismo en cada clase impartida además de inculcar valores éticos y morales, al MSc. Diego Caicedo por su gran aporte al presente trabajo investigativo.

DEDICATORIA

Dedicado a la memoria de mi padre Carlos Fernando y mi abuelita Clarita Elena quienes espiritualmente siempre me han acompañado a todo lado, a ellos se los debo todo.

A mi madre, hermana, sobrinito, tía y abuelito quienes con su amor y esfuerzo me han sido de gran ayuda para poder seguir adelante, al apoyo incondicional de mis mejores amigos Cristian, Gabi, Geovanny y mi compañera de vida Dani quien me ha tenido mucha paciencia y me ha brindado su amor en todo momento, a mi buen amigo Mateo quien ha demostrado que la distancia no es un impedimento para tener una buena amistad.

A toda mi familia, mis amigos y mis compañeros que depositaron su confianza en mí.

ÍNDICE

I. PROBLEMA	16
1.1. PLANTEAMIENTO DEL PROBLEMA	16
1.2. FORMULACIÓN DEL PROBLEMA	17
1.3. JUSTIFICACIÓN	18
1.4. OBJETIVOS Y PREGUNTAS DE INVESTIGACIÓN	19
1.4.1. Objetivo General.....	19
1.4.2. Objetivos Específicos	19
1.4.3. Preguntas de Investigación	19
II. FUNDAMENTACIÓN TEÓRICA	21
2.1. ANTECEDENTES INVESTIGATIVOS	21
2.2. MARCO TEÓRICO	22
2.2.1. Agricultura.....	22
2.2.2. Eventos que afectan la agricultura.....	22
2.2.3. Precipitaciones pluviales	23
2.2.4. Estación meteorológica	23
2.2.5. Variables meteorológicas	24
2.2.6. Big Data.....	25
2.2.7. Minería de datos	26
2.2.8. Series temporales	31
III. METODOLOGÍA.....	34
3.1. ENFOQUE METODOLÓGICO	34
3.1.1. Enfoque.....	34
3.1.2. Tipo de Investigación	34
3.2. IDEA A DEFENDER.....	35
3.2.1. Variables.....	35

3.3. DEFINICIÓN Y OPERACIONALIZACIÓN DE VARIABLES	36
3.4. MÉTODOS UTILIZADOS.....	37
3.4.1. Método	37
3.5. TECNICAS PARA LA RECOLECCIÓN DE DATOS	37
3.5.1. La entrevista.....	37
3.5.2. La encuesta.....	38
3.5.3. Análisis estadístico.....	38
3.6. RECURSOS	40
IV. RESULTADOS Y DISCUSIÓN	42
4.1. RESULTADOS	42
4.1.1. Análisis de las encuestas	42
4.1.2. Propuesta.....	53
4.1.3. Metodología y herramientas.....	54
4.1.4. Fases Metodología CRISP-DM	54
4.1.5. Elaboración de un Dashboard	68
4.2. DISCUSIÓN.....	71
V. CONCLUSIONES Y RECOMENDACIONES.....	74
5.1. CONCLUSIONES.....	74
5.2. RECOMENDACIONES	75
IV. REFERENCIAS BIBLIOGRÁFICAS	76
V. ANEXOS.....	79

ÍNDICE DE FIGURAS

Figura 1. Número de días con precipitación y máxima de 24 horas	16
Figura 2. Flujo de procesos metodología KDD.....	27
Figura 3. Flujo de procesos metodología CRISP-DM.....	28
Figura 4. La actividad agrícola constituye su principal fuente de economía	42
Figura 5. Pérdida de cultivos por eventos meteorológicos	43
Figura 6. Tipos de precipitaciones pluviales se da en su zona	44
Figura 7. Precipitaciones pluviales que afectan la calidad de los cultivos	45
Figura 8. Precipitaciones pluviales afectan el precio de productos agrícolas	46
Figura 9. Métodos para la predicción del cambio climático.....	47
Figura 10. Datos que pueden medir una estación meteorológica	48
Figura 11. Cuenta con una estación meteorológica en su sector agrícola	49
Figura 12. De acuerdo con los datos que se generaron en años anteriores haciendo uso de predicciones, cree que se puede generar los mismos valores en los siguientes años.	50
Figura 13. Beneficiaria a usted un análisis predictivo sobre las futuras precipitaciones pluviales	51
Figura 14. Existe disponibilidad de internet en su zona.....	52
Figura 15. Cuenta con un dispositivo electrónico para acceder a los datos analizados de la estación meteorológica	53
Figura 16. Base de datos estación meteorológica original	56
Figura 17. Base de datos estación meteorológica transformada.....	57
Figura 18. Serie de tiempo.....	57
Figura 19. Serie de tiempo con corte de información.....	58
Figura 20. Serie de tiempo 2019 – 2021.....	58
Figura 21. Estacionalidad en serie de tiempo	59
Figura 22. Descomposición aditiva de series de tiempo	59
Figura 23. Predicción con promedios móviles SMA.....	60
Figura 24. Dashboard Shinyapps.io server.....	62
Figura 25. Página principal Prediction	63
Figura 26. Serie de tiempo Exploración	63
Figura 27. Tablas de Exploración.....	64
Figura 28. Archivo descargado en formato xlsx.....	65
Figura 29. Histograma SMA Pronostico	66

Figura 30. Tablas de datos valores ponderados y pronostico	66
Figura 31. Página principal vista desde móvil	67
Figura 32. Menú pronostico vista desde móvil	68
Figura 33. Dashboard forecast	69
Figura 34. Indicadores de colores	69
Figura 35. Promedios de precipitaciones orden ascendente	70
Figura 36. Valores Justados 2019-2020	70
Figura 37. Tabla pronóstico 2021	71
Figura 38. Selección de datos	89
Figura 39. Guarda como CSV	90
Figura 40. Extensión delimitado por comas .CSV	90
Figura 41. Interfaz principal.....	91
Figura 42. Carga de datos	91
Figura 43. Total de instancias	92
Figura 44. Función botón de búsqueda	92
Figura 45. Rango de fechas.....	92
Figura 46. Meses de pronóstico	93
Figura 47. Menú exploración	93
Figura 48. Histograma precipitación.....	93
Figura 49. Tablas de exploración.....	94
Figura 50. Tendencia precipitación.....	94
Figura 51. Pronóstico SMA	95
Figura 52. Página principal del dashboard.....	96
Figura 53. Indicadores de colores	96
Figura 54. Promedio de precipitaciones.....	97
Figura 55. Filtro en tabla de promedios	97
Figura 56. Filtro de orden en tabla de precipitaciones	97
Figura 57. Tabla de pronósticos	98
Figura 58. Botones de descarga, compartir y copiar	99
Figura 59. Formatos de descarga	99

ÍNDICE DE TABLAS

Tabla 1. Diferencia de software más utilizado para minería de datos	29
Tabla 2. Operacionalización de Variables	36
Tabla 3. Nivel de Confianza (NC).....	39
Tabla 4. Parámetros para formula finita	40
Tabla 5. Recursos para proyecto investigativo	41
Tabla 6. Resultados encuesta pregunta 1	42
Tabla 7. Resultados encuesta pregunta 2.....	43
Tabla 8. Nivel de conocimiento tipos de precipitación en su zona	44
Tabla 9. Nivel de conocimiento precipitaciones pluviales que afectan la calidad de los cultivos	45
Tabla 10. Nivel de conocimiento precipitaciones pluviales que afectan al precio de productos agrícolas.....	46
Tabla 11. Nivel de conocimiento métodos que utiliza para la predicción del cambio climático	47
Tabla 12. Nivel de conocimiento datos que pueden medir una estación meteorológica.....	48
Tabla 13. Nivel de conocimiento cuenta con una estación meteorológica.....	49
Tabla 14. Nivel de conocimiento usted cree que se pueda generar los mismos valores en los siguientes años.....	50
Tabla 15. Nivel de conocimiento beneficiaria a usted un análisis predictivo sobre las futuras precipitaciones que puedan ocurrir	51
Tabla 16. Nivel de conocimiento disponibilidad de internet en su zona.....	52
Tabla 17. Nivel de conocimiento cuenta con un dispositivo electrónico para acceder a los datos analizados de la estación meteorológica.....	53
Tabla 18. Variables estación meteorológica.....	55
Tabla 19. Fitted años 2019 – 2020	60
Tabla 20. Forecast año 2021	61

ÍNDICE DE ANEXOS

Anexo 1: Acta de predefensa	79
Anexo 2: Informe de Turnitin	80
Anexo 3: Certificado validación del Abstract.....	81
Anexo 4: Certificado de conformidad.....	83
Anexo 5: Guion entrevista	84
Anexo 6: Guion Encuesta	85
Anexo 7: Manual de usuario	87

RESUMEN

En el cantón San Pedro de Huaca existe la Finca Experimental San Francisco, cuenta con una estación meteorológica que recolecta datos relacionados con el clima como: la precipitación pluvial, donde se hace un análisis mínimo a los datos históricos que almacena, no sé a evidenciado el uso de herramientas de minería de datos aplicado en las series temporales, sirviendo de ayuda para los diferentes agricultores evitando que el riesgo sea menor por los fenómenos climáticos relacionados. Haciendo uso del enfoque cualitativo se pudo conocer mediante una entrevista a personal de la finca el tratamiento que se hace a los datos almacenados por la estación meteorológica como también las variables y sus unidades de medida, mientras que el enfoque cuantitativo permitió conocer la factibilidad de saber el pronóstico de precipitaciones pluviales a través de una encuesta, para ello la investigación estará basada en la aplicación de técnicas de minería de datos predictiva con el uso de la metodología CRISP-DM, dando cumplimiento a todas sus fases incluyendo el despliegue. Para la etapa de modelado se aplica los pronósticos móviles, Este modelo se lo aplica para el pronóstico de series de tiempo, donde hace predicciones basadas en comportamientos pasados, para llevar a cabo todo este proceso se hizo uso del lenguaje de programación R ya que presenta grandes ventajas en el análisis de datos. Finalmente, los resultados obtenidos serán mostrados en un aplicativo web shinyapp.io que permitirá interactuar y descargar el análisis de los datos, quedando estos a disposición del público.

Palabras claves: Minería de datos, análisis de datos, CRISP-DM, series de tiempo, promedios móviles

ABSTRACT

There is San Francisco Experimental Farm in San Pedro de Huaca canton, it has a meteorological station that collects weather-related data such as: rainfall, where a minimal analysis is made to the historical data that it stores, the use of data mining tools applied in the time series has not been evidenced, serving as help for the different farmers avoiding that the risk is less due to the related climatic phenomena. It was possible to know, using the qualitative approach through an interview with farm staff the treatment that is done to the data stored by the meteorological station as well the variables and their units of measurement, while the quantitative approach allowed to know the feasibility of knowing the rainfall forecast through a survey, for this, the research will be based on the application of predictive data mining techniques with the use of the CRISP-DM methodology complying with all its phases including deployment. For the modeling stage, mobile forecasts are applied. This model is applied for forecasting time series, it makes predictions based on past behaviors, the programming language R was used to carry out all this process since it has great advantages in data analysis. Finally, the results obtained will be displayed in a web application shinnyapp.io that will allow you to interact and download the data analysis, keeping it available to the public.

Key words: Data mining, data analysis, CRISP-DM, time series, moving averages.

INTRODUCCIÓN

La finca Experimental San Francisco de la Universidad Politécnica Estatal del Carchi, objeto de estudio, es un espacio brindado por la casona para hacer diferentes estudios relacionados con la agricultura y ganadería, se encuentra ubicada en el cantón San Pedro de Huaca y cuenta con una estación meteorológica sistematizada cuyo propósito es recolectar datos de las diferentes variables meteorológicas que existen.

El presente proyecto tuvo como finalidad la creación de un aplicativo informático de análisis de datos históricos de la Finca Experimental San Francisco de Huaca, con el cual se pretende predecir futuras precipitaciones pluviales haciendo uso de la minería de datos para series temporales aplicando modelos como los promedios móviles simples, además de poner estos resultados a la vista del público en general mediante un aplicativo web interactivo.

El proyecto de investigación consta de 5 capítulos que se describen a continuación:

El capítulo uno expone la problemática, justificación de la investigación, objetivos determinados y preguntas de investigación. El capítulo dos indica la fundamentación teórica, donde constan los antecedentes de investigación que hacen referencia al estudio y el marco teórico donde se define términos técnicos, tales como: minería de datos, series de tiempo. El tercer capítulo describe el enfoque metodológico utilizado, los tipos de investigación en las cuales se fundamentó, idea a defender, Operacionalización de variables de estudio, de igual forma los métodos necesarios para la recolección de información. El capítulo cuatro describe los resultados de la investigación, para ello se da a conocer la metodología para el desarrollo del proyecto, llegando finalmente a la discusión de estos. Por último, capítulo quinto expone las conclusiones y recomendaciones que se ha obtenido después de finalizar el proyecto.

I. PROBLEMA

1.1. PLANTEAMIENTO DEL PROBLEMA

Ecuador un país único en fauna y flora donde la actividad agrícola constituye una de las principales fuentes de economía que sostienen a la nación, además de ser una gran fuente de empleo que representa un gran porcentaje de la población. La agricultura depende directamente del factor clima, existiendo un alto riesgo de pérdida en las cosechas por diferentes eventos hidro – meteorológicos como lo indica (Vazquez, 2018) “las precipitaciones pluviales producen eventos como sequías, heladas, granizadas, inundaciones, etc.” Que ponen en vulnerabilidad los productos agrícolas de esta forma afectando la calidad y los precios en el mercado, así como también llevando a pérdidas significantes de los diferentes sectores.

Existen datos históricos del comportamiento de las precipitaciones que son un fenómeno físico natural de la caída del agua al suelo que se encuentra en las nubes.(Ilbay, 2019) señala que:

La zona centro de los Andes del Ecuador presentan tendencias al incremento de precipitaciones anuales y disminución en la zona centro sur, debido a que la precipitación se concentra en un número reducido de meses y por el contrario una muy baja cantidad de precipitación puede estar cayendo en un mayor número de meses, causando periodos de sequía e inundaciones respectivamente.

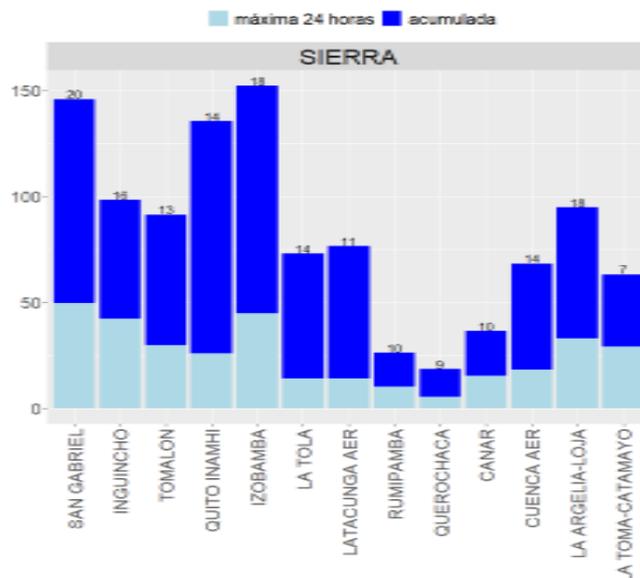


Figura 1. Número de días con precipitación y máxima de 24 horas

Fuente: Reporte mensual febrero 2021 (INAMHI)

Al norte del país se encuentra la provincia del Carchi, una de las principales productoras agrícolas por su situación geográfica, donde la mayoría de su población se dedican a los cultivos alimenticios en base a conocimientos empíricos para la predicción climatológica los cuales fueron transmitidos a través de generaciones y no en base a estudios o investigaciones.

Este fenómeno produce que se acumule gran cantidad de agua con material pétreo, material de rellenos sanitarios, provocando que los canales de riego se taponen y desborden, afectando a los cultivos en su área de influencia principalmente en la provincia del Carchi, esto también origina afectación en la salubridad del sector ya que se encuentran los desechos que caen del relleno sanitario.

Al existir cambios en las precipitaciones en las diferentes zonas del país es muy importante conocer esta variación para cumplir con un punto de equilibrio entre la demanda y la oferta para que exista una utilidad y no una pérdida económica además de obtener un producto de buena calidad.

En la provincia del Carchi, Cantón San Pedro de Huaca existe la Finca Experimental San Francisco de Huaca de propiedad de la Universidad Politécnica Estatal del Carchi, donde existe una estación meteorológica que recolecta datos relacionados con el clima como: la humedad, la temperatura, velocidad del viento entre otros, donde se hace un análisis mínimo a los datos históricos que genera, no sé a evidenciado el uso de herramientas de minería de datos aplicado en las precipitaciones pluviales, que sirvan de ayuda para los diferentes moradores que se dedica a los cultivos para evitar que el riesgo sea menor por la volatilidad del clima y otros factores de la naturaleza evitando pérdidas económicas.

1.2. FORMULACIÓN DEL PROBLEMA

El bajo análisis e interpretación de los datos históricos de las precipitaciones pluviales en la estación meteorológica de la Finca Experimental San Francisco de Huaca afecta la planificación y la toma de decisiones en el ámbito agrícola.

1.3. JUSTIFICACIÓN

Los datos históricos que son almacenados por estaciones meteorológicas son información importante para el análisis de datos, uno de estos análisis que se puede realizar son las predicciones de precipitaciones pluviales en series de tiempo, las mismas que sirven para saber cuándo es buen tiempo para realizar actividades agrícolas teniendo un producto de gran calidad y sin perjudicar los cultivos.

La minería de datos consiste en una gran herramienta de estrategia la cual su funcionalidad es analizar los datos históricos desde los diferentes puntos estratégicos, para así poder transformar esta información, ordenando, clasificando y filtrando. (Rivera, Lema, Freire, Rojas, y Villa, 2018), estos datos serán facilitados y tomados de la estación meteorológica, de la finca experimental San Francisco de Huaca de la Universidad Politécnica Estatal del Carchi, ya que se cuenta con una autorización respectiva.

Según el autor (Ortiz, 2017) afirma que: Dentro del área agrícola se puede aplicar la minería de datos con la finalidad de proponer herramientas que permitan al usuario tener acceso a la información precisa donde se realicen predicciones sobre la evolución futura de actividades que se desarrollan, obteniendo resultados a corto plazo, que permitirá asegurar la confiabilidad de estos, sirviendo de apoyo para las decisiones futuras que se puedan tomar sobre las mismas.

Los beneficiarios directos serán los agricultores del cantón San Pedro de Huaca, al igual que otros agricultores de diferentes zonas que cuenten con datos históricos sobre las precipitaciones, ya que la propuesta del aplicativo informático permitirá visualizar los resultados de predicciones obtenidos con sus propias variables almacenadas mediante el uso de un modelo predictivo realizado con la técnica de promedios móviles, permitiendo tomar mejores decisiones para obtener productos de mejor calidad, resultando con mayor demanda para el mercado logrando un mayor desarrollo económico.

Esta investigación ayudaría a crear un nuevo instrumento para recolectar y analizar datos, ya que el proceso de extracción del conocimiento se la realiza a través de varias fases (preparación de datos, exploración, auditoria, minería de datos, evaluación, difusión y utilización de modelos) y la incorporación de diversas técnicas (árboles de decisión, regresión lineal, redes neuronales artificiales, máquinas de soporte vertical) además permite una presentación variada

del problema mediante su clasificación, categorización, estimación, regresión, agrupamiento etc. (Pérez y Santín, 2017).

Además de que poder realizar diversas técnicas para obtener los patrones pudiendo así hacer una comparativa de cual técnica es la más viable al momento de obtener datos, el número de técnica de minería de datos es muy grande además que cabe indicar no existe una técnica específica para resolver un problema determinado, como también la aplicación de diferentes metodologías de minería de datos las cuales todas llevan a un mismo propósito.

1.4. OBJETIVOS Y PREGUNTAS DE INVESTIGACIÓN

1.4.1. Objetivo General

- Desarrollar un aplicativo informático de análisis de datos históricos de la Finca Experimental San Francisco de Huaca.

1.4.2. Objetivos Específicos

- Fundamentar bibliográficamente como las precipitaciones pluviales afectan el rendimiento y la calidad de productos agrícolas además de los tipos de cultivos.
- Elaborar un marco metodológico para la investigación del proyecto Minería de datos para series temporales y su aplicación en las precipitaciones pluviales en la Finca Experimental San Francisco de Huaca
- Interpretar el análisis de los datos históricos de la estación meteorológica en la finca experimental San Francisco de Huaca para la predicción de futuras precipitaciones pluviales.
- Diseñar el paquete para el trabajo con series temporales con aplicativo de analítica de datos.

1.4.3. Preguntas de Investigación

- ¿Cómo la fundamentación bibliográfica ayudara a entender como las precipitaciones pluviales afectan el rendimiento y la calidad de productos agrícolas y conocer los tipos de cultivos?
- ¿Cómo un marco metodológico favorece a la investigación del proyecto Minería de datos para series temporales y su aplicación en las precipitaciones pluviales en la Finca Experimental San Francisco de Huaca?

- ¿Cómo se interpreta un análisis de datos históricos de una estación meteorológica para predecir futuras precipitaciones pluviales?
- ¿Qué es un aplicativo de analítica de datos?

II. FUNDAMENTACIÓN TEÓRICA

2.1. ANTECEDENTES INVESTIGATIVOS

Como primer antecedente investigativo tomamos en cuenta el trabajo de investigación del Instituto Mexicano de Tecnología del Agua que pertenece al autor Miguel Ángel Vázquez Zavaleta publicado en el año 2018 con el título “Minería de datos para generación de reglas de tendencia de precipitación pluvial en el estado de Morelos cuyo objetivo principal es Aplicar técnicas de minería de datos a la información de las redes de estaciones climatológicas que permita obtener conocimiento de gran utilidad para la toma de decisiones.”

Como segundo documento de ayuda para la realización de este plan de titulación, se encuentra el trabajo investigativo de la Universitat Oberta de Catalunya perteneciente al autor Carlos Espino Timón publicado en el año 2017 con el título “Análisis predictivo: técnicas y modelos utilizados y aplicaciones del mismo – herramientas Open Source que permiten su uso.” Cual objetivo general es Conocer que es el análisis predictivo, así como los modelos y técnicas aplicables.

El tercer trabajo investigativo de gran relevancia para la ejecución de este plan de titulación, se encuentra en el trabajo de la Universidad Central “Marta Abreu” de las Villas realizado por César Soto Valero publicado en el año 2013, con el título “Minería de datos para series temporales en Weka y su aplicación en el pronóstico de precipitaciones” cuya finalidad es Diseñar e implementar un paquete para Weka que facilite la tarea de clasificación de series temporales, y aplicar tanto los nuevos métodos implementados como los tradicionales al problema del pronóstico de precipitaciones en la provincia de Villa Clara.

Como cuarto antecedente investigativo tomamos en cuenta el trabajo de investigación del Instituto Superior Politécnico “José Antonio Echeverría” que pertenece a los autores Gil Cruz Lemus, Alejandro Rosete-Suárez, Leonor Turtós Carbonell publicado en el año 2010 con el título “Estimación de parámetros meteorológicos secundarios aplicando minería de datos” cuyo objetivo principal es Realizar un estudio sobre patrones y tendencias de los parámetros meteorológicos secundarios.

Como quinto documento de ayuda para la realización de este plan de titulación, se encuentra el trabajo investigativo de la Escuela Superior Politécnica de Chimborazo perteneciente a los autores Silvia Haro, Lourdes Zúñiga, Luis Vera, Antonio Meneses, Amalia Escudero publicado

en el año 2018 con el título “Métodos de clasificación en minería de datos meteorológicos” Cual objetivo general es clasificar una variable dentro de una de las categorías de una clase, considerando variables meteorológicas para determinar el modelo adecuado al conjunto de datos, aplicando los modelos de clasificación.

2.2. MARCO TEÓRICO

2.2.1. Agricultura

La agricultura en la provincia del Carchi corresponde a una actividad activa económicamente y de gran relevancia para la población de esta zona. Desde el descubrimiento de la agricultura se ha ido mejorando cada día sus técnicas a la hora de producir bienes alimenticios, haciendo uso de productos químicos y maquinas que ayudan a realizar el tratamiento de la tierra y el cultivo.

La actividad agrícola constituye uno de los sectores más importantes dentro de la economía global ya que la finalidad de este desarrollo es acabar con la pobreza extrema y llevar alimentos a millones de habitantes (Lapo, 2019).

En el Ecuador se aplica la agricultura familiar campesina que consiste en no hacer uso de tecnología, sin embargo tiene gran relevancia ya que atribuye alimentos de la canasta básica y es el tipo de agricultura que más se utiliza dentro del área rural, es considerada de gran relevancia para el ámbito económico moderno, evitando la pobreza dentro de los habitantes de la zona rural (Lapo, 2019).

2.2.2. Eventos que afectan la agricultura

Uno de los principales eventos que afecta la agricultura en la Latinoamérica y por lo tanto en Ecuador es el cambio climático, esto debido a su situación geográfica. La producción agrícola depende de estos y otros factores climáticos, como la humedad del suelo, temperatura, precipitaciones pluviales, luz del sol y fertilidad de la tierra encontrando efectos diferentes para cada región (Guamán, 2020).

Los cambios del clima principalmente afectan a los nutrientes del suelo y su superficie, los eventos afectan a la producción y distribución de alimentos ocasionando efectos negativos no solo en los cultivos si no también en agricultores, en el ámbito económico y social (Guamán ,2020).

2.2.3. Precipitaciones pluviales

Las precipitaciones pluviales se encuentran en grandes cantidades dentro de la provincia del Carchi esto debido a su situación geográfica y al cambio climático que se encuentra en la región sierra, provocando efectos perjudiciales en el clima y en toda actividad que dependa de esta variable.

“La precipitación pluvial se considera como la variable principal en los estudios hidrogeológicos, ya que es la fuente fundamental para el cálculo de balances hídricos y la generación de alertas tempranas por riesgo de sequía en la región.” (Herrera, Campos, y Carrillo, 2017)

Las precipitaciones se las puede encontrar en sus diferentes formas y estados como la lluvia, granizo, nieve dependiendo de la posición geográfica y la región en donde se encuentre.

2.2.4. Estación meteorológica

Según el autor (Anastacio y Barzola 2019) define a las estaciones meteorológicas como lugares donde se realizan diversos estudios sobre el comportamiento del medio ambiente y la atmosfera mediante los instrumentos meteorológicos.

2.2.4.1. Tipos de estaciones meteorológicas

Existen dos tipos de estaciones meteorológicas que son:

- **Convencionales.** “Estación que está constituida por instrumentos convencionales, donde la variable meteorológica es convertida en un movimiento mecánico que hace desplazar una plumilla sobre una banda de papel que se avanza continuamente por un sistema de relojería, y genera una gráfica” (Bolaños, 2018).

La estación meteorológica convencional necesita de un observador meteorológico quien debe estar capacitado para emitir el registro y lectura de la información meteorológica. La información es generada por los diferentes sensores con los que la estación cuenta: termómetros, pluviómetros, veletas, entre otros sensores que cumplen una determinada función.

- **Automáticas.** Las estaciones meteorológicas automáticas se caracterizan por hacer uso de herramientas electrónicas las cuales son encargadas de registrar las lecturas diarias de los datos en almacenamientos sólidos. Este tipo de estaciones tiene poca intervención de personal ya que los sensores son los encargados de generar y almacenar los datos mediante un servidor (Bolaños, 2018).

Actualmente en la Finca Experimental San Francisco se encuentra en funcionamiento una estación meteorológica sistematizada automática, cuenta con doce canales en donde almacena diferentes variables meteorológicas realizando la toma de mediciones cada hora. Estos datos se almacenan en un dispositivo USB (*Bus Universal en Serie*) teniendo acceso a ellos cuando se necesite.

2.2.5. Variables meteorológicas

La Agencia Estatal de Meteorología indica que las variables meteorológicas son la: “Cantidad o magnitud meteorológica que puede adquirir distintos valores numéricos dentro de un conjunto de números especificado. Algunas de las principales variables meteorológicas son la presión atmosférica, la temperatura, la humedad, la velocidad del viento, la nubosidad y la precipitación” (Agencia Estatal de Meteorología, 2018).

Las variables meteorológicas como su nombre lo indica es todo aquello que pueda ser medido por medio de los instrumentos meteorológicos, como la temperatura que es medida en grados Fahrenheit o centígrados.

2.2.5.1. Temperatura

La temperatura es la variable meteorológica que mide la cantidad de calor que se encuentra en la atmósfera, la unidad de medidas para la temperatura es los grados (Fahrenheit o centígrados) (Anastacio y Barzola, 2019). La variable temperatura corresponde al canal 9 de la estación meteorológica de la Finca Experimental San Francisco.

2.2.5.2. Humedad relativa

La humedad relativa corresponde al promedio de todas las mediciones ejecutadas en un determinado tiempo de 10 minutos, esta variable meteorológica indica la cantidad de agua que está presente en el aire (Anastacio y Barzola, 2019). La variable humedad corresponde al canal 10 de la estación meteorológica de la Finca Experimental San Francisco.

2.2.5.3. Precipitación

“La precipitación puede ser entendida como cualquier fenómeno físico que cae desde la atmósfera (lluvia, granizo, nieve) hacia la corteza terrestre. Esta variable se expresa en milímetros (mm)” (Anastacio y Barzola, 2019). La precipitación corresponde al canal 11 de la estación meteorológica de la Finca Experimental San Francisco.

2.2.6. Big Data

El autor (Paucar, 2019) define al Big Data como: “una nueva generación de tecnologías y arquitecturas diseñadas para extraer económicamente el valor de volúmenes muy grandes de una amplia variedad de datos permitiendo la captura de alta velocidad, el descubrimiento y/o análisis”.

El Big Data tiene como finalidad brindar beneficios a las organizaciones por medio de la recopilación de grandes cantidades de datos que pueden ser estructurados, semi estructurados y no estructurados que son generados por las mismas, para ser descubiertos y optimizados para una mejor toma de decisiones.

2.2.6.1. Business Intelligence

El Business Intelligence o la inteligencia de negocios consiste en un proceso de análisis de datos utilizando herramientas tecnológicas que brinda soluciones tanto a empresas como organizaciones a la toma de decisiones correctas para una mejor gestión de la información.

Según (Tixe, 2021) afirma que: BI es un proceso participativo que permite explorar y analizar información de un área determinada que es almacenada en un data warehouse, con el propósito de revelar patrones o tendencias, para posteriormente sacar ideas y conclusiones.

Para la realización de este proceso es necesario contar con un buen manejo de la cantidad de datos que tengan almacenados y que estos sean de calidad, otras de las características que deben de cumplir para aplicar BI según el autor (Tixe , 2021) son: “Eficiencia en minimizar recursos para el tratado de la información, proporcionar información relevante, adecuada, concreta y clara en la cual su acceso no debe ser complejo, orientado a la gestión para la toma de decisiones.”

2.2.6.2. Machine Learning

El machine learning o también conocido como aprendizaje de máquinas se define como diversos métodos computacionales que hacen uso de la experiencia con la finalidad de mejorar el desempeño de las predicciones con la finalidad de que estas sean más precisas, la experiencia consiste en la información de bases de datos históricas que es utilizada para los procesos de entrenamiento (Bonilla Ambrossi, 2019).

2.2.7. Minería de datos

Según el autor (Guayasamín , 2018) indica que: la minería de datos corresponde a un proceso sistemático de extracción de información útil y oculta, que es utilizada estratégicamente para la toma de decisiones y así poder entender de mejor manera los fenómenos que ocurren. La minería de datos consta de las siguientes fases que son: definición de problema, preparación, selección de datos, procesamiento de los datos, generación y validación del modelo.

La minería de datos corresponde a una potente herramienta del descubrimiento de nuevos datos a partir de información histórica determinado diferentes patrones para finalmente realizar predicciones, este proceso se lo puede utilizar en diferentes campos y ámbitos siendo la agricultura uno de ellos.

2.2.7.1. Metodologías para minería de datos

Existen tres metodologías principales para llevar a cabo el proceso de minería de datos

- **KDD (*Knowledge Discovery in Databases*)**. La metodología KDD fue propuesta por Fayyad en el año de 1996, según el autor (Guzmán, 2017) consta de varias fases que son:
 - **Selección**. Obtener conocimientos previos desde las necesidades del cliente para así poder identificar el propósito de la extracción de conocimientos KDD. Crear un conjunto de datos objetivo, esto quiere decir seleccionar los datos (variables) necesarias para llevar al descubrimiento de nuevos datos.
 - **Preprocesamiento y limpieza**. Para la limpieza y el procesamiento de datos se hace uso de diferentes operaciones como la eliminación de ruido en los campos donde se encuentran datos vacíos. Otro paso importante es hacer una reducción de la dimensionalidad para interpretar datos en función del objetivo del trabajo.

- **Transformación.** Llevar los datos preprocesados a un método de minería de datos, seleccionando un algoritmo para la búsqueda de patrones de datos.
- **Interpretación, evaluación.** La interpretación de los datos corresponde a visualizar tanto los modelos y patrones extraídos como los datos nuevos que fueron extraídos, hacer uso de estos datos para realizar informes, documentos o para adoptar nuevas medidas con los datos tratados.

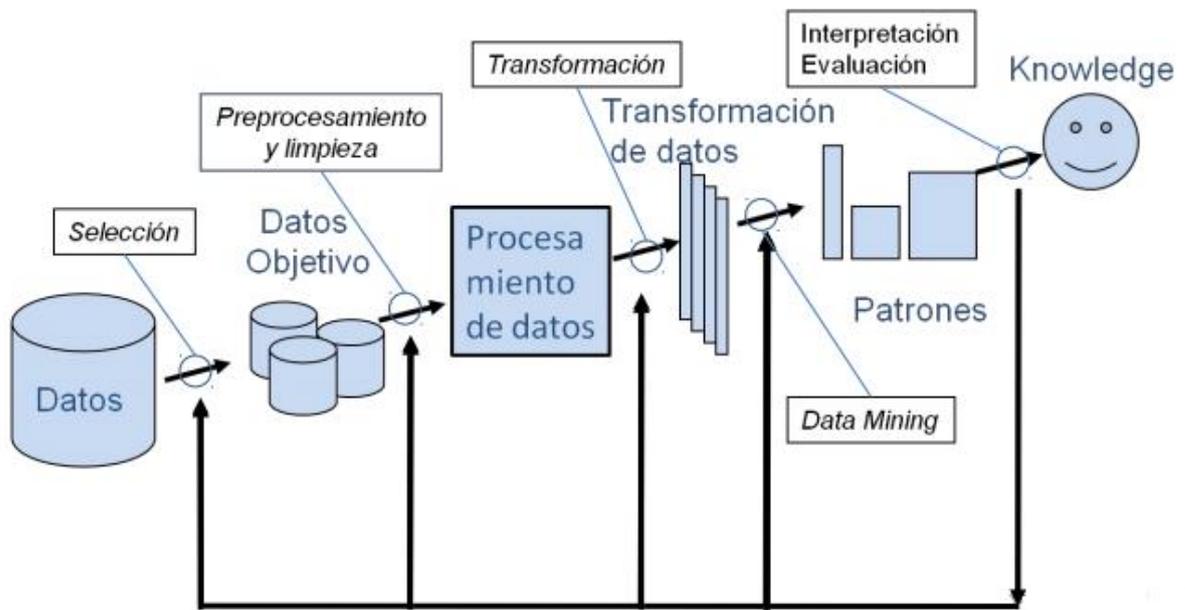


Figura 2. Flujo de procesos metodología KDD

Fuente: Módulo minería de datos (2016)

- **CRISP-DM.** Cross Industry Standard Process for Data Mining (Proceso estándar de la industria para la minería de datos) Es el modelo analítico más utilizado por los expertos en el área de minería de datos, el objetivo de esta metodología es permitir la compatibilidad de las herramientas a través de todo el proceso de minería de datos (Guzmán, 2017).

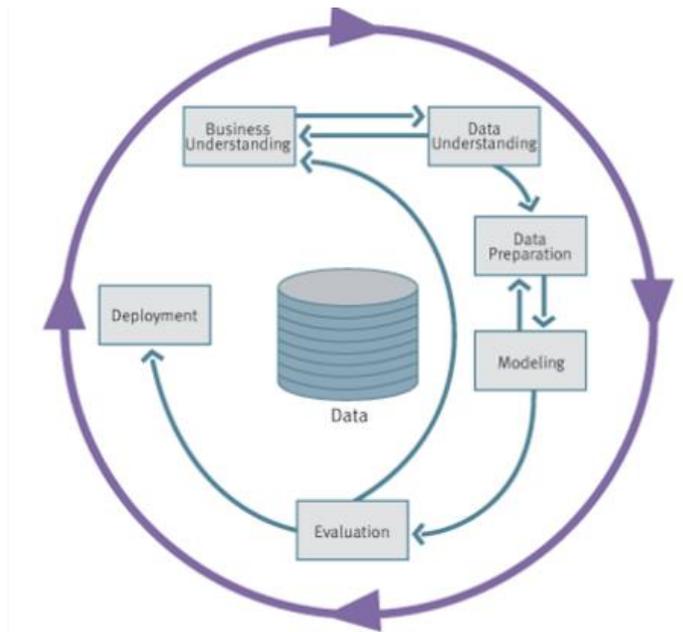


Figura 3. Flujo de procesos metodológica CRISP-DM

Fuente: Módulo de minería de datos (2016)

La metodología CRISP-DM consta de seis fases que son:

- **Análisis del problema.** Definir el problema, entender los objetivos y requisitos para poder llevar a cabo el proyecto.
- **Comprensión de los datos.** Obtener datos iniciales y realizar una exploración, identificar características de los datos y sus especificaciones.
- **Preparación de los datos.** Selección y limpieza de los datos.
- **Modelamiento.** Implementar en alguna de las herramientas disponibles para minería de datos.
- **Evaluación.** Identificar si los resultados concuerdan y cumplen con el objetivo del proyecto.
- **Despliegue.** Hacer uso del resultado de los modelos en nuevas prácticas y/o adaptaciones.

La metodología CRISP-DM es la más comúnmente usada en proyectos de minería de datos, ya que es completa y entendible en cada una de sus fases, además de que cuenta con una fase adicional de despliegue donde explota más a fondo el modelo realizado por esta metodología llevándolo a lo práctico e interactivo.

- **SEMMA.** SEMMA son las siglas de las cinco fases que componen a esta metodología (Sample, Explore, Modify, Model, Assess), el autor (Guzmán, 2017) define a esta metodología como “un proceso de selección exploración y modelamiento de enormes cantidades de datos para obtener y descubrir nuevos patrones que fueron desconocidos”
 - **SAMPLE.** Entradas de datos, partición de datos
 - **EXPLORE.** Exploración de datos, intuición, asociación, selección de variables
 - **MODIFY.** Transformación de variables, agrupamiento, ruido
 - **MODEL.** Regresión lineal, arboles de decisión, redes neuronales
 - **ASSESS.** Evaluación, medidas, reportes

2.2.7.2. Herramientas tecnológicas para minería de datos

Las herramientas de minería de datos sirven para gestionar los datos identificando las tendencias y patrones que son relevantes. Cada día existen nuevos y mejores softwares de minería de datos que también son conocidos como data mining tools, el autor corporativo (IONOS, 2018) presenta en una vista general algunos de los software más usados.

Tabla 1. Diferencia de software más utilizado para minería de datos

Nombre	Características	Lenguaje de programación	Sistema operativo	Precio/Licencia
RapidMiner	Apto para todos los procesos. Destaca en el análisis predictivo	Java	Windows, macOS, Linux	Freeware, diferentes versiones de pago
WEKA	Muchos métodos de clasificación	Java	Windows, macOS, Linux	Software libre (GPL)
Orange	Crea una visualización de datos atractiva sin que se requieran muchos conocimientos previos para ello	Núcleo del software: C++, ampliación y lenguaje de entrada: Python	Windows, macOS, Linux	Software libre (GPL)
KNIME	Software de data mining de código abierto que ha democratizado el acceso a los análisis predictivos	Java	Windows, macOS, Linux	Software libre (GPL) (a partir de la versión 2.1)

SAS	Caro, pero potente para grandes empresas	Lenguaje SAS	Windows, macOS, Linux	Freeware limitado a instituciones públicas, el precio se establece tras solicitud, diferentes modelos disponibles
-----	--	--------------	-----------------------	---

En la tabla 1 encontramos los principales softwares de minería de datos con sus características más esenciales.

2.2.7.3. Enfoque de minería de datos con R

R es un lenguaje de programación interpretado (con licencia GNU GPL). Esto indica que, el comando se ejecuta directamente sin recompilar el programa en instrucciones en lenguaje de máquina. Este entorno se utiliza comúnmente para cálculos estadísticos y gráficos debido a sus múltiples métodos estadísticos (modelado lineal y no lineal, pruebas estadísticas clásicas, análisis de series de tiempo, análisis, etc.) (Unir, 2019).

Estas son algunas de las características principales de R:

- Almacenamiento y gestión de datos eficiente.
- Un conjunto de operadores para realizar cálculos matriciales.
- Una gran colección de herramientas para el análisis de datos.
- Utilidad gráfica para visualización de datos.

Se utiliza en todas las fases del análisis de datos.

- Recopila datos de fuentes disponibles como bases de datos y archivos de texto.
- Preparación de datos: Elimina duplicados, datos incorrectos, extremos, etc.
- Análisis de datos: Construir modelos predictivos, clasificar, agrupar
- Comunicar resultados: Crear informes que presenten resultados y conclusiones.

Las funciones y aplicaciones versátiles del R lo convierten en una herramienta básica para los analistas de datos.

Las diferentes librerías con las que cuenta el lenguaje de programación R son de gran eficacia al tratarse de análisis de datos permitiendo obtener resultados con mayor precisión y creando un entorno de trabajo ágil y versátil, permitiendo también realizar todas las etapas que corresponden a la minería de datos.

2.2.7.4. Tableau Public

Para la elaboración del dashboard final y entregable de los datos resultantes de la aplicación se utiliza la herramienta tableau public, que consiste en una plataforma virtual gratuita que permite explorar visualizar y compartir datos con el público general. Es una plataforma en línea sin embargo cuenta con un aplicativo de escritorio para poder trabajar de mejor manera.

Tableau public permite generar buenos gráficos estadísticos como mapas de calor, barras, tablas de manera ordenada y entendible para que cualquier usuario pueda interpretar los datos de predicciones de las precipitaciones pluviales para el presente año

2.2.8. Series temporales

El autor (Mancero, 2017) indica que: “las series temporales son sucesiones de variables aleatorias que se desarrollan a través de otras variables”, el estudio del comportamiento de estas variables es de gran importancia al momento de querer predecir sucesos a futuro, aunque muchas de las veces no sea con mayor exactitud, permite tomar acciones que decidan continuar o modificar la evolución actual.

Las series temporales son de gran importancia en el estudio de precipitaciones pluviales ya que mediante estas sucesiones podemos observar el comportamiento que han tomado desde la aparición de la variable y como ha ido evolucionando mediante el tiempo y los diferentes factores que pueden interferir en ella.

2.2.8.1. Modelos de series temporales

El termino modelo hace referencia a la forma de representación de un proceso físico mediante la abstracción, donde permite a un fenómeno ser repetido en varias ocasiones, este proceso se lo conoce también como simulación.

Los modelos de series temporales lineales se clasifican en dos categorías como explica (Mancero, 2017):

- **Modelos univariantes.** Emplea una única variable que se va a predecir en la serie temporal.
- **Modelos multivariantes.** Utiliza una o más series temporales que pueden ser influyentes para la predicción de la nueva variable.

2.2.8.2. Descomposición aditiva de series de tiempo

El análisis clásico de series de tiempo se basa en el supuesto de que los valores asumidos por las variables observadas son el resultado de los tres componentes y que sus acciones comunes se traducen en medidas (Escutia, 2019).

Los componentes de la serie temporal son:

- Tendencia T_t
- Estacional S_t
- Componente aleatorio E_t

Los dos primeros componentes son deterministas y el tercero es aleatorio. En el modelo de suma, la serie de tiempo se puede representar de la siguiente manera: $X_t = T_t S_t E_t$

El autor (Escutia, 2019) define a los componentes de la serie de tiempo como:

- **Tendencia.** Esta función se usa para trazar e interpretar tendencias en los datos ts (series de tiempo), mientras se adquieren vectores de tendencia. La serie temporal no tiene tendencia.
- **Estacionalidad.** Esta función se utiliza para obtener el vector de estacionalidad, para obtener una serie de tiempo sin él, o para representar gráficamente la estacionalidad de datos de tipo ts (serie de tiempo).
- **Aleatorio.** Esta función elimina el componente aleatorio de un objeto de tipo ts que es una serie de tiempo. Este componente estocástico se puede trazar y calcular a partir de la serie de tiempo sin preprocesamiento o introduciendo la tendencia y los componentes estáticos de la serie original.
- **Descomposición.** Esta función elimina objetos de tipo ts , específicamente tendencia de series de tiempo, componentes fijos y aleatorios, y los devuelve en un objeto de lista. Cabe señalar que los devuelven como objetos de series de tiempo. También devuelve el valor del tipo de decaimiento utilizado y le permite agregar un histograma de sus componentes.

2.2.8.3. Promedios Móviles

Para la aplicación de la técnica de promedios móviles la serie de tiempo debe de cumplir ciertos requisitos como que esta tenga estabilidad, esto quiere decir que los datos que la componen se generan sin diferencia significativa entre un dato y otro (error aleatorio) lo que quiere decir que

el comportamiento de los datos incluso si muestra una ganancia o una pérdida, lo hará con una tendencia constante (Zamora, 2018).

“Cuando se utiliza el método de la media móvil, se supone que todas las observaciones de la serie temporal tienen la misma importancia para la estimación de los parámetros de predicción. Por lo tanto, el promedio de valores pertenecientes a los datos de la última serie de tiempo se utiliza como pronóstico para el próximo período de tiempo” (Zamora, 2018).

Se utiliza la siguiente expresión matemática para los promedios móviles que es:

$$\text{Promedio Móvil SMA} = \frac{\sum(n \text{ valores más recientes de la serie de tiempo})}{n}$$

III. METODOLOGÍA

3.1. ENFOQUE METODOLÓGICO

3.1.1. Enfoque

La presente investigación tendrá un enfoque cuantitativo, ya que según los autores (Hernández Sampieri, Fernández Collado, y Baptista Lucio 2018) el enfoque cuantitativo usa la recolección de datos para probar hipótesis, con base en la medición numérica y el análisis estadístico, para establecer patrones de comportamiento y probar teorías.

Este tipo de enfoque está presente en las diferentes encuestas que se va a realizar para determinar la problemática del proyecto, permitiendo conocer si los agricultores de la zona se ven perjudicados por las precipitaciones pluviales que llegan a suceder.

A su vez, utiliza un enfoque cualitativo ya que “utiliza la recolección de datos sin medición numérica para descubrir o afinar preguntas de investigación en el proceso de interpretación” (Hernández Sampieri et al., 2018).

Este enfoque se ve plasmado en las diferentes entrevistas que se realizara al personal de la Finca Experimental San Francisco, para conocer más sobre el mecanismo en el manejo de datos de la estación meteorológica.

3.1.2. Tipo de Investigación

3.1.2.1. Investigación Descriptiva

La presente investigación hace uso de la investigación descriptiva ya que establece una descripción lo más completa posible de un fenómeno, situación, midiendo las características y procesos que componen a dicho suceso, el uso de la investigación descriptiva se encuentra en el acceso a los datos dentro de la estación meteorológica.

3.1.2.2. Investigación Experimental

Se emplea la investigación experimental basándose en la manipulación de variables en condiciones controladas, observando el grado en la o las variables implicadas y manipuladas producen un efecto determinado. Se obtiene los datos de manera aleatoria, de esta forma permitiendo establecer diferentes hipótesis, la parte experimental hace énfasis en que se

realizara diferentes análisis con los datos obtenidos de la estación con la finalidad de encontrar predicciones pluviales.

3.1.2.3. Investigación Documental

La investigación documental o teórica tiene como principal objetivo la obtención de conocimientos de diferentes ámbitos, sin tomar en cuenta la aplicabilidad de los conocimientos obtenidos, este tipo de investigación será de ayuda para la realización de del marco teórico y conceptual del presente plan de titulación.

3.2. IDEA A DEFENDER

Los datos históricos de las precipitaciones pluviales en la estación meteorológica de la Finca Experimental San Francisco de Huaca permitirán generar reglas de tendencia para la predicción de precipitaciones futuras y evitar pérdidas económicas.

3.2.1. Variables

3.2.1.1. Variable Dependiente. Series temporales, precipitaciones pluviales

3.2.1.2. Variable Independiente. Minería de datos

3.3. DEFINICIÓN Y OPERACIONALIZACIÓN DE VARIABLES

Tabla 2. Operacionalización de Variables

Variables	Dimensión	Indicadores	Técnica	Instrumento o reactivos
Independiente: Minería de datos	Comprensión	Selección	Metodología CRISP-DM	Lenguaje de programación R
	Preparación	Tipo de dato		
	Modelado	Modelo matemático		
	Evaluación	Resultados		
Dependiente:				
Series temporales	Patrón estacional	Histograma	Descomposición aditiva de serie de tiempo	Tendencia Estacional Componente aleatorio
Precipitaciones pluviales	Variable meteorológica	Unidad de medida	Milímetros por metro cuadrado de terreno	Estación meteorológica

En la tabla 2 constituye la operacionalización de variables dependiente e independiente, además de indicar su definición conceptual, dimensión e indicador

3.4. MÉTODOS UTILIZADOS

3.4.1. Método

3.4.1.1. Método Deductivo

“El método deductivo consiste en extraer una conclusión con base en una premisa o a una serie de proposiciones que se asumen como verdaderas.” (Westreicher, 2020)

El método deductivo hace presencia en esta investigación ya que haciendo uso de razonamiento lógico se logró plantear la idea a defender, basándose en una teoría comprobada y validada permitiendo de igual manera tomar decisiones más acertadas para brindar mejores resultados u conclusiones en el proceso de minería de datos para series temporales aplicado en las precipitaciones pluviales.

3.4.1.2. Método Descriptivo

El método descriptivo es el encargado de recopilar datos medibles para ser analizados en una población y muestra específica, este método tuvo gran importancia en el desarrollo de esta investigación definiendo y midiendo con gran precisión los problemas de precipitaciones pluviales que los agricultores del cantón San Pedro de Huaca tenían en sus cultivos posteriormente teniendo perdidas de calidad en sus productos, de esta forma conociendo la realidad del problema planteado en la investigación, así como también brindando una posible solución haciendo uso de la técnica de minería de datos para predicción de series temporales.

3.5. TECNICAS PARA LA RECOLECCIÓN DE DATOS

3.5.1. La entrevista

El autor (Caballero, 2017) afirma que la entrevista: “consiste en la comunicación verbal entre el entrevistador y entrevistado con el fin de obtener datos. Debe ser previamente diseñada en función al tema de estudio, a la vez que planteada por el entrevistador”.

La entrevista fue aplicada al MSc. Diego Caicedo responsable de los datos meteorológicos de la finca experimental San Francisco de la Universidad Politécnica Estatal del Carchi con la finalidad de conocer y tener más información sobre el tratamiento de los datos proporcionados por la estación meteorológica de este establecimiento para de esta manera analizar la

factibilidad de realizar minería de datos para obtener nueva información para la toma de decisiones.

3.5.2. La encuesta

La entrevista es un instrumento de recolección de datos que funciona mediante una consulta o un cuestionario de preguntas que hace referencia a cualquier aspecto en común (Caballero, 2017). Esta técnica de recolección de información fue aplicada a agricultores de la zona con el propósito de conocer que tipos de precipitaciones pluviales se presentan y como estas afectan en sus cultivos, así como también si generan perdidas y si están de acuerdo en la ayuda del modelo de minería de datos para la predicción de futuras precipitaciones que puedan ocurrir.

3.5.3. Análisis estadístico

3.5.3.1. Población

La población o universo es el conjunto total de individuos que comparten características y especificaciones que son comunes (Hernández Sampieri et al., 2018).

Como población se ha tomado al cantón San Pedro de Huaca ubicado en la provincia del Carchi, cuyos datos fueron obtenidos del Instituto Nacional de Estadísticas y Censos INEC realizado en el 2010 en donde se pudo observar que cuenta con 6868 habitantes de los cuales 2331 corresponden a población activa económicamente desde hace cinco años o más, eligiendo a 324 habitantes que se dedican a la agricultura como actividad ocupacional.

3.5.3.2. Muestra

La muestra corresponde a un subconjunto de la población. Para realizar el muestreo se utilizó la siguiente fórmula para una población finita.

$$n = \frac{N * z_a^2 * p * q}{e^2 * (N - 1) + z_a^2 * p * q}$$

Donde los parámetros corresponden a los siguientes valores:

n = Tamaño de muestra

N = Tamaño de población o universo

Z = Parámetro estadístico que depende el Nivel de Confianza (NC)

e = Error de estimación aceptado

p = Probabilidad de que ocurra el evento estudiado (éxito)

$q = (1-p)$ Probabilidad de que no ocurra el evento estudiado

Para calcular el parámetro estadístico (Z) hacemos uso de la tabla del Nivel de Confianza:

Tabla 3. Nivel de Confianza (NC)

N.Confianz	Z alfa
99,70%	3
99%	2,58
98%	2,33
96%	2,05
95%	1,96
90%	1,645
80%	1,28
50%	0,674

En la tabla 3 se encuentra el nivel de confianza junto al valor de Z que va ser utilizado como parámetro para la fórmula de población finita y poder determinar la muestra.

Asignamos los valores para cada parámetro con un margen de error de estimación máximo aceptado del 5% y un nivel de confianza del 95%, para p y q asignamos valores del 50% a cada uno siendo esta la probabilidad de que pueda ocurrir o no.

Partiendo de la información previamente vista se conoce que 324 personas corresponden a la población que se dedica a la agricultura como actividad principal de ingreso económico en el cantón San Pedro de Huaca en la provincia del Carchi.

Por lo tanto, el valor de N (Tamaño de la población) toma el valor de 324

Tabla 4. Parámetros para formula finita

Parámetro	Valor
N	324
Z	1,96
p	50
q	50
e	5

En la Tabla 4 encontramos los valores para cada parámetro para posteriormente ser reemplazados en la fórmula de muestra finita.

$$n = \frac{324 * 3.8416 * 50 * 50}{25 * 323 + 3.8416 * 50 * 50}$$

$$n = \frac{3111696}{17679}$$

$$n = 176.01086$$

Una vez aplicada la formula con los valores para cada parámetro, el resultado fue de 176 habitantes agricultores del cantón San Pedro de Huaca, muestra a la cual se destinó una encuesta estructurada realizada por la plataforma de Google Forms.

3.6. RECURSOS

Para la realización del presente proyecto de minería de datos para series temporales y su aplicación en las precipitaciones pluviales se utilizarán herramientas digitales que permitirán desarrollar el aplicativo informático para el tema a tratar.

Tabla 5. Recursos para proyecto investigativo

Tipo	Nombre	Propiedades	Costo
Institucionales	Finca	Infraestructura, Estación meteorológica	-
	Experimental San Francisco (UPEC)		
Materiales	Computadora	Intel Core i5 8 th Generation 8GB RAM, ASUS	800\$
	Teléfono Móvil	Smartphone con acceso a internet	200\$
Tecnológicos	lenguaje de Programación R		-
	Librerías data mining		-
	Host de shinyapp.io		-
Total			1000\$

La tabla 5 hace referencia a los recursos que van a ser utilizados durante la investigación divididos en categorías como institucionales, materiales y tecnológicos mostrando sus propiedades y costos respectivos.

IV. RESULTADOS Y DISCUSIÓN

4.1. RESULTADOS

4.1.1. Análisis de las encuestas

1.- ¿La actividad agrícola constituye su principal fuente de economía?

Tabla 6. Resultados encuesta pregunta 1

		Frecuencia	Porcentaje
Válido	SI	139	79,0%
	NO	37	21,0%
Total		176	100,0%

La tabla 6 muestra la frecuencia y el porcentaje de los resultados obtenidos de la pregunta 1

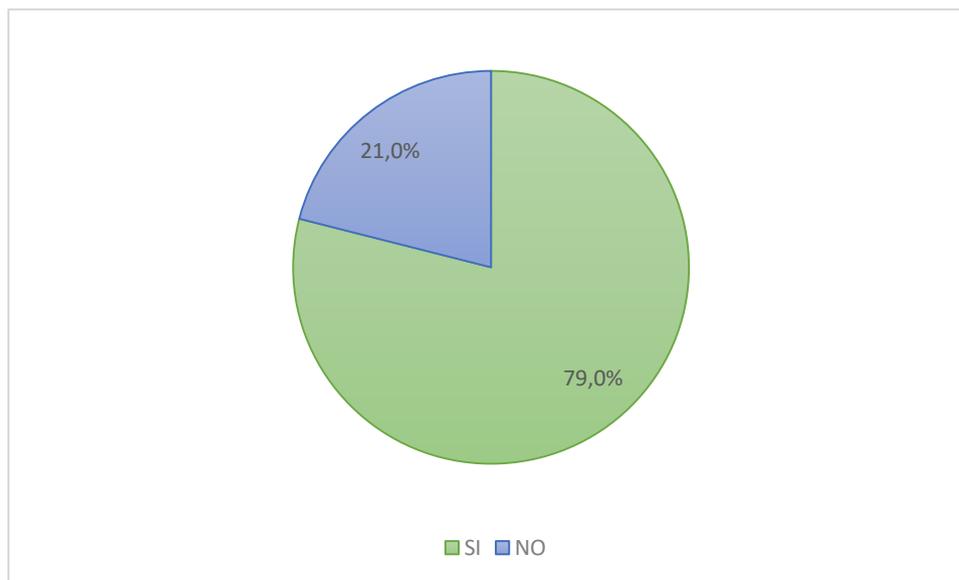


Figura 4. La actividad agrícola constituye su principal fuente de economía

La primera pregunta de la encuesta corresponde a conocer si las actividades agrícolas son la principal fuente de economía para numerosas familias, los resultados favorecen la mayor parte de encuestados quienes indicaron que la agricultura es su fuente de ingresos, mientras que en un menor porcentaje indicaron que la agricultura no corresponde a una actividad principal de generar dinero, esto quiere decir que el mayor porcentaje de personas encuestadas tiene una dependencia del campo agrícola.

2.- ¿A tenido pérdidas en sus cultivos por eventos meteorológicos?

Tabla 7. Resultados encuesta pregunta 2

		Frecuencia	Porcentaje
Válido	SI	131	74,4%
	NO	45	25,6%
	Total	176	100,0%

La tabla 7 muestra la frecuencia y el porcentaje de los resultados obtenidos de la pregunta 2

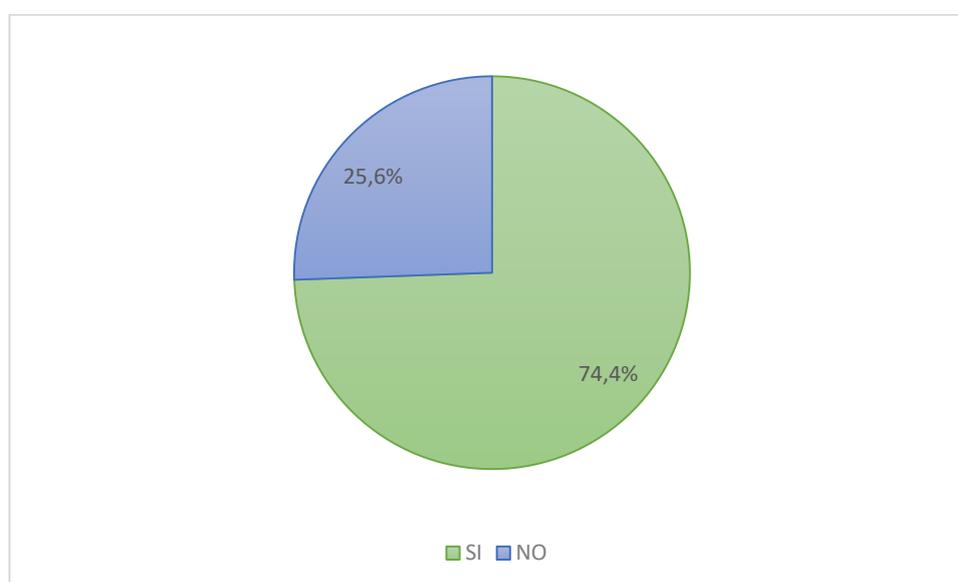


Figura 5. Pérdida de cultivos por eventos meteorológicos

La segunda pregunta va encaminada a las pérdidas económicas ocasionadas por cultivos, los porcentajes son alarmantes ya que determinan que la mayor población agrícola ha sufrido afectaciones a sus cultivos por culpa de eventos meteorológicos, mientras que una menor parte de la muestra no ha tenido ningún inconveniente.

3.- Cuáles de estos tipos de precipitaciones pluviales se da en su zona

Tabla 8. Nivel de conocimiento tipos de precipitación en su zona

		Frecuencia	Porcentaje
Válido	Sequias	124	70,5%
	Heladas	122	69,3%
	Granizadas	137	77,8%
	Inundaciones	142	80,7%
	Total	176	

La tabla 8 muestra la frecuencia y el porcentaje de los resultados obtenidos de la pregunta 3

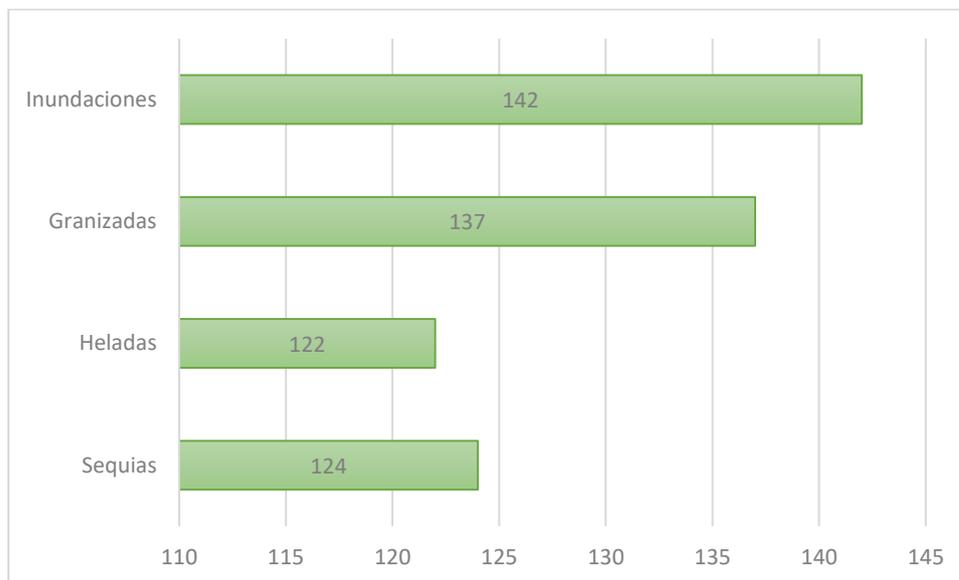


Figura 6. Tipos de precipitaciones pluviales se da en su zona

En la pregunta tres se pudo identificar cual es el tipo de precipitación pluvial que más se da en la zona donde fue aplicada la encuesta encontrando en primer lugar a las inundaciones con un con el mayor porcentaje, seguida por las granizadas con, sequias y finalmente las heladas siendo la precipitación menos influyente con el porcentaje mínimo del total.

4.- ¿Las precipitaciones pluviales afectan a la calidad de los cultivos?

Tabla 9. Nivel de conocimiento precipitaciones pluviales que afectan la calidad de los cultivos

		Frecuencia	Porcentaje
Válido	SI	134	76,1%
	NO	42	23,9%
	Total	176	100,0%

La tabla 9 muestra la frecuencia y el porcentaje de los resultados obtenidos de la pregunta 4

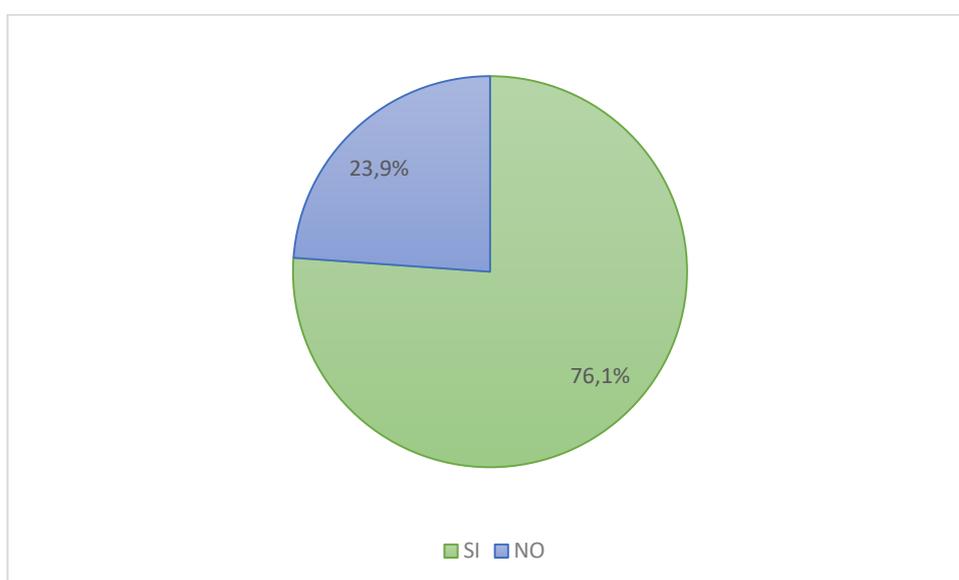


Figura 7. Precipitaciones pluviales que afectan la calidad de los cultivos

La pregunta cuatro de la encuesta hace énfasis en que si las precipitaciones afectan a la calidad de los cultivos en donde se puede notar que un mayor porcentaje de la población encuestada ha tenido deterioros en sus cultivos ocasionando grandes pérdidas en la calidad de los mismos, mientras que el menor porcentaje restante indica que las precipitaciones no afectan para nada en la calidad de los productos.

5.- ¿Las precipitaciones pluviales afectan al precio de los productos agrícolas?

Tabla 10. Nivel de conocimiento precipitaciones pluviales que afectan al precio de productos agrícolas.

		Frecuencia	Porcentaje
Válido	SI	130	73,9%
	NO	46	26,1%
	Total	176	100,0%

La tabla 10 muestra la frecuencia y el porcentaje de los resultados obtenidos de la pregunta 5

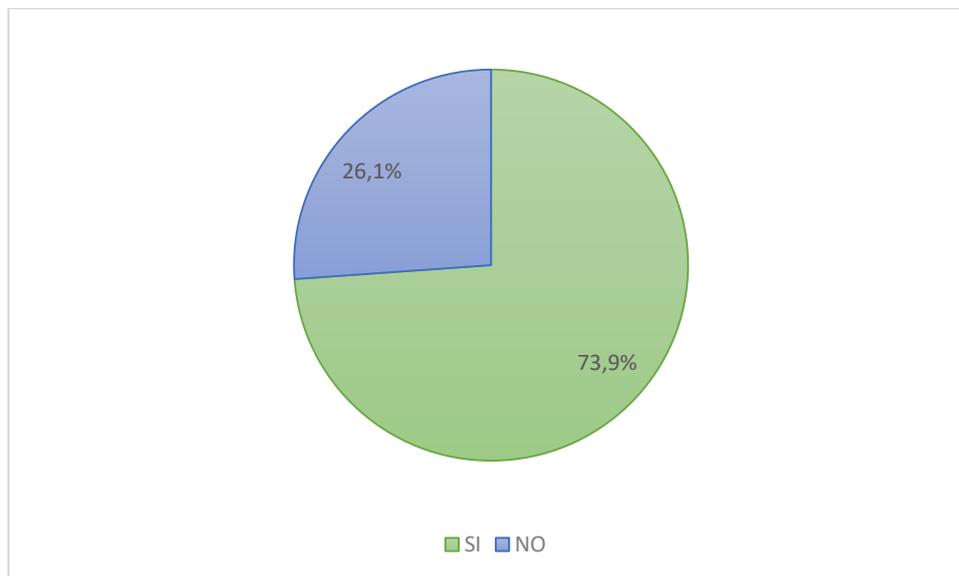


Figura 8. Precipitaciones pluviales afectan el precio de productos agrícolas

La quinta pregunta tiene mucha relación con la pregunta anterior, ya que permite identificar si las precipitaciones pluviales afectan al precio de estos cultivos el porcentaje es alto en las personas que están a favor quienes alguna vez han tenido consecuencias negativas por este motivo, mientras que la menor cantidad de personas restante no ha presentado pérdidas en los precios de productos agrícolas.

6.- ¿Qué métodos utiliza para la predicción del cambio climático?

Tabla 11. Nivel de conocimiento métodos que utiliza para la predicción del cambio climático

		Frecuencia	Porcentaje
Válido	Internet	151	85,8%
	Almanaques	125	71,0%
	Ancestrales	116	65,9%
	Noticieros	132	75,0%
	Total	176	

La tabla 11 muestra la frecuencia y el porcentaje de los resultados obtenidos de la pregunta 6

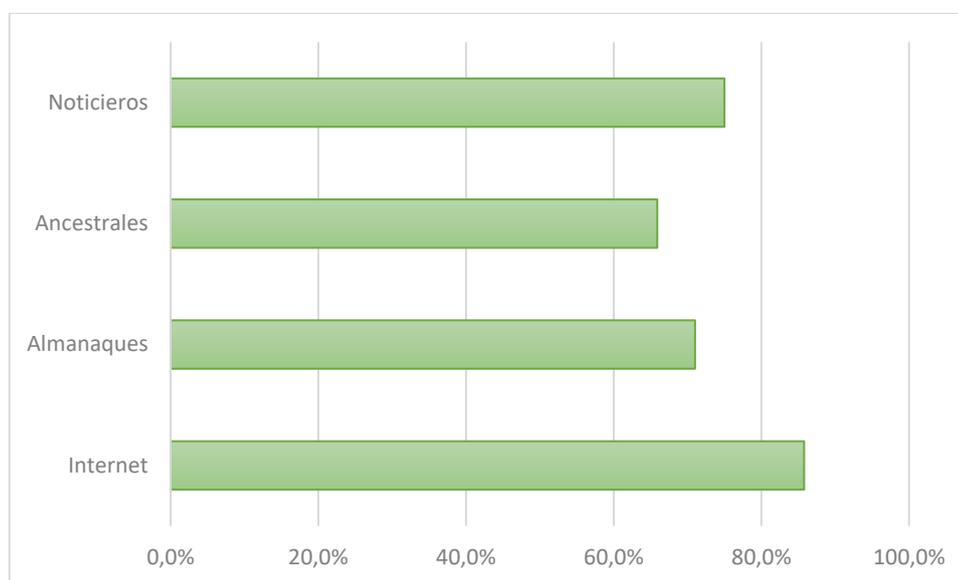


Figura 9. Métodos para la predicción del cambio climático

La sexta pregunta hace referencia a los métodos que más comunes utilizados para predecir los cambios climáticos, siendo internet el método más frecuentado por los usuarios, seguido de los noticieros, almanaques y finalmente los métodos ancestrales siendo los menos populares en la encuesta con el menor porcentaje del total de encuestados.

7.- ¿Qué datos puede medir una estación meteorológica?

Tabla 12. Nivel de conocimiento datos que pueden medir una estación meteorológica

		Frecuencia	Porcentaje
Válido	Viento	150	85,2%
	Temperatura	123	69,9%
	Humedad	135	76,7%
	Presión atmosférica	108	61,4%
	Atmosfera - Sol	101	57,4%
	Total	176	

La tabla 12 muestra la frecuencia y el porcentaje de los resultados obtenidos de la pregunta 7

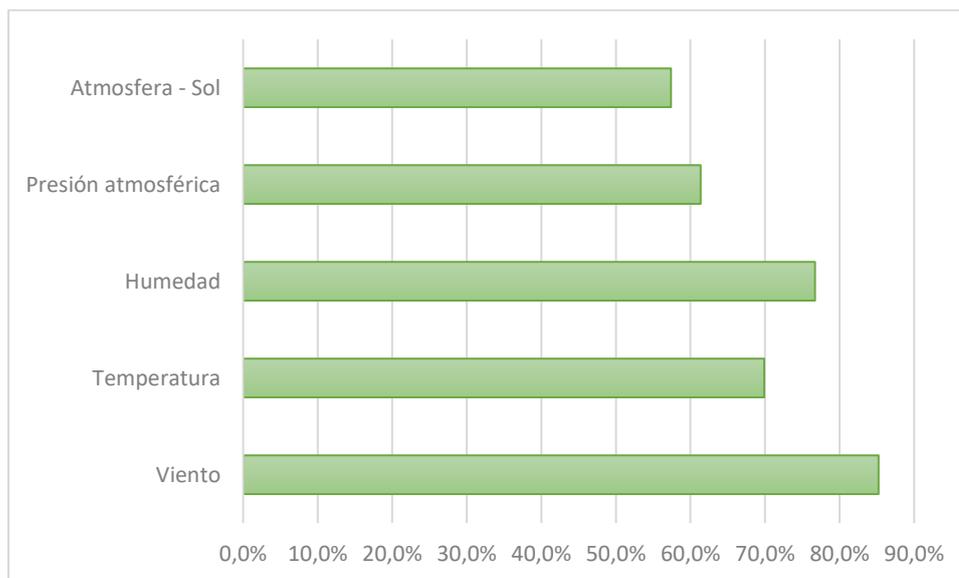


Figura 10. Datos que pueden medir una estación meteorológica

La pregunta número siete hace referencia a las variables medibles que se puede obtener con una estación meteorológica, entre todas las opciones las variables que más tienen conocimiento son el viento, seguido por la humedad, temperatura, presión – atmosférica; finalmente encontramos la variable de atmosfera – sol siendo la menos conocida por los usuarios encuestados con un leve porcentaje.

8.- ¿Cuenta con una estación meteorológica en su sector agrícola?

Tabla 13. Nivel de conocimiento cuenta con una estación meteorológica

		Frecuencia	Porcentaje
Válido	SI	108	61,4%
	NO	68	38,6%
	Total	176	100,0%

La tabla 13 muestra la frecuencia y el porcentaje de los resultados obtenidos de la pregunta 8

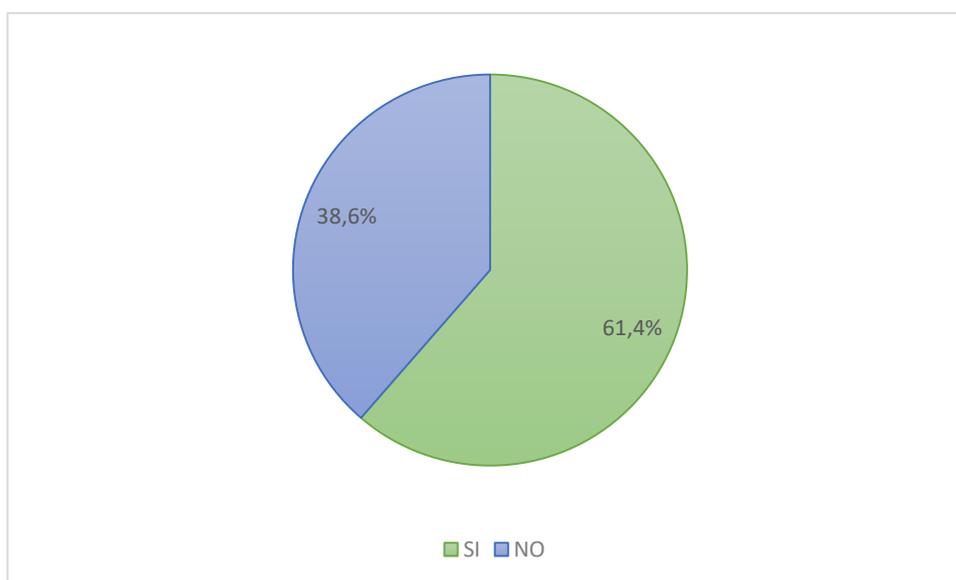


Figura 11. Cuenta con una estación meteorológica en su sector agrícola

La pregunta número ocho es un indicador de que los agrícolas en su lugar de trabajo cuentan con una estación meteorológica, el porcentaje favorable supera con un poco más de la mitad del total de encuestas que tienen acceso a una estación meteorológica de cualquier tipo, mientras que el un porcentaje menor no tiene acceso a una estación meteorológica.

9.- ¿En qué le beneficiaría a usted tener los datos meteorológicos?

La novena pregunta sirve para tener claro la inquietud de qué manera se beneficiarían los agricultores al tener datos meteorológicos, al ser una pregunta abierta las respuestas que más se repitieron fueron que les ayudaría a descubrir cual mes es el mejor para cultivar, así como también les permite mejorar los cultivos en tiempo y calidad, llevando un manejo mas cuidadoso de los mismos.

10.- De acuerdo con los datos que se generaron en años anteriores haciendo uso de predicciones, usted cree que se puede generar los mismos valores en los siguientes años.

Tabla 14. Nivel de conocimiento usted cree que se pueda generar los mismos valores en los siguientes años

		Frecuencia	Porcentaje
Válido	SI	135	76,7%
	NO	41	23,3%
	Total	176	100,0%

La tabla 14 muestra la frecuencia y el porcentaje de los resultados obtenidos de la pregunta 10

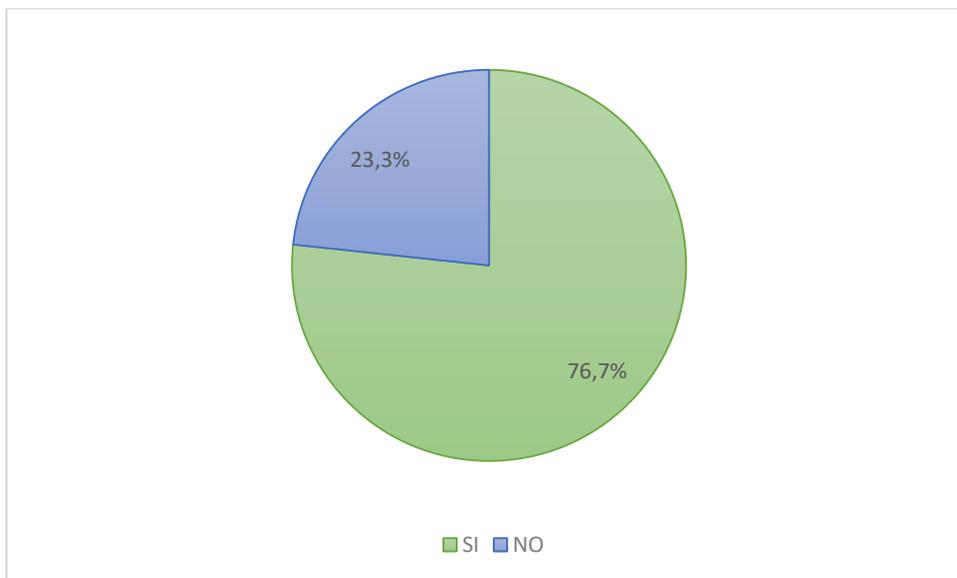


Figura 12. De acuerdo con los datos que se generaron en años anteriores haciendo uso de predicciones, cree que se puede generar los mismos valores en los siguientes años.

La décima pregunta hace hincapié en la credibilidad de que se pueda generar valores a partir de valores que fueron tomados de años anteriores, un alto porcentaje de encuestados está de acuerdo en que si se puede generar conocimiento a partir de datos históricos mientras que un bajo porcentaje se encuentra en desacuerdo.

11.- ¿Beneficiaria a usted un análisis predictivo sobre las futuras precipitaciones pluviales que puedan ocurrir?

Tabla 15. Nivel de conocimiento beneficiaria a usted un análisis predictivo sobre las futuras precipitaciones que puedan ocurrir

		Frecuencia	Porcentaje
Válido	SI	133	75,6%
	NO	43	24,4%
Total		176	100,0%

La tabla 15 muestra la frecuencia y el porcentaje de los resultados obtenidos de la pregunta 11

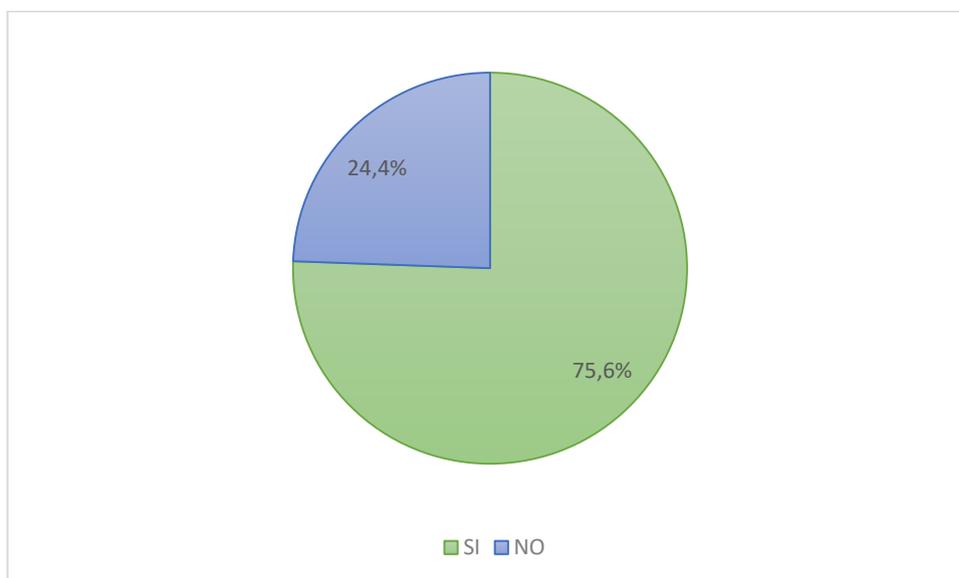


Figura 13. Beneficiaria a usted un análisis predictivo sobre las futuras precipitaciones pluviales

La pregunta once se la realizó con la finalidad de conocer la factibilidad que tendría un desarrollo de un análisis predictivo sobre futuras precipitaciones pluviales, un alto porcentaje de encuestados confía en que los datos que se obtendrán mediante el análisis predictivo, mientras que un mínimo porcentaje se encuentra en total desacuerdo.

12.- ¿Existe disponibilidad de internet en su zona?

Tabla 16. Nivel de conocimiento disponibilidad de internet en su zona

		Frecuencia	Porcentaje
Válido	SI	99	56,3%
	NO	77	43,8%
Total		176	100,0%

La tabla 16 muestra la frecuencia y el porcentaje de los resultados obtenidos de la pregunta 12

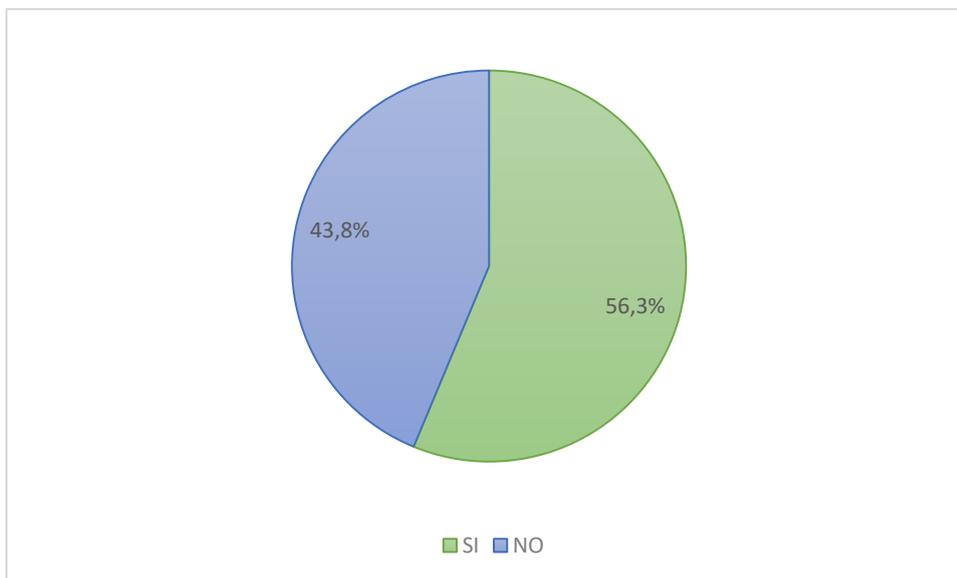


Figura 14. Existe disponibilidad de internet en su zona

La penúltima pregunta indica si se cuenta con conexión a internet dentro de su zona, ya que es muy importante presentar conectividad para revisar los análisis predictivos, las personas que cuentan con conectividad son un poco más de la mitad del total, y casi la otra mitad no cuenta con la conectividad en su zona de trabajo.

13.- ¿Cuenta con un dispositivo electrónico (computadora, celular, Tablet) para acceder a los datos analizados de la estación meteorológica?

Tabla 17. Nivel de conocimiento cuenta con un dispositivo electrónico para acceder a los datos analizados de la estación meteorológica

		Frecuencia	Porcentaje
Válido	SI	125	71,0%
	NO	51	29,0%
Total		176	100,0%

La tabla 17 muestra la frecuencia y el porcentaje de los resultados obtenidos de la pregunta 13

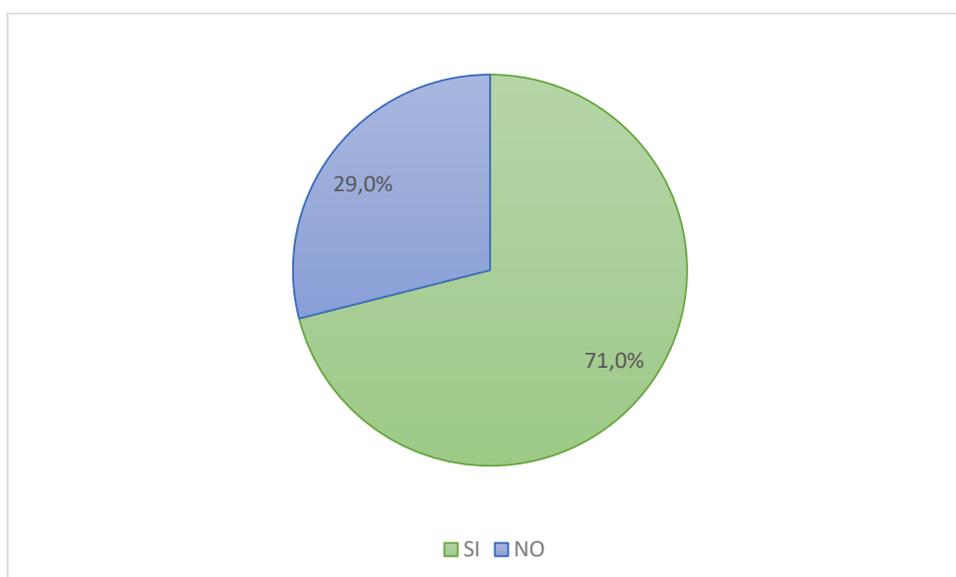


Figura 15. Cuenta con un dispositivo electrónico para acceder a los datos analizados de la estación meteorológica

La décimo tercera y última pregunta tiene la finalidad de conocer cuántos de los encuestados cuenta con un dispositivo electrónico con acceso a internet como una computadora, celular o Tablet, el mayor porcentaje de los encuestados si cuenta con alguna de estas herramientas tecnológicas, y un mínimo de encuestados no cuenta con aparatos electrónicos.

4.1.2. Propuesta

En concordancia con las variables de investigación y los análisis estadísticos realizados la propuesta consiste en crear un modelo y un aplicativo informático de analítica de datos para series temporales aplicado en las precipitaciones pluviales, haciendo uso de los datos históricos

de la finca experimental San Francisco con la finalidad de saber, conocer y predecir de mejor manera y más exacta las futuras precipitaciones pluviales que puede darse dentro de su zona, permitiendo solucionar problemas como la pérdida de cultivos, así como también sus ganancias ya que muchas personas dependen de esta actividad como fuente de ingreso primaria.

4.1.3. Metodología y herramientas

La metodología de programación utilizada es la CRISP-DM, cumpliendo con todas las etapas, haciendo énfasis en las etapas de modelado, evaluación y despliegue ya que es ahí donde se concentra la problemática de este proyecto.

La herramienta utilizada para la elaboración del modelado y aplicativo informático de analítica de datos es el lenguaje de programación R ya que en el ámbito de datos permite manipular, procesar y visualizar gráficos del análisis de datos de alta calidad, además de ser un lenguaje que puede ser utilizado en todas las fases de análisis de datos. También se hace uso de R Studio que no es más que un entorno de desarrollo.

4.1.4. Fases Metodología CRISP-DM

4.1.4.1. Análisis del Problema

En esta etapa se determinar exactamente cuál es la problemática que existe y se debe solucionar, para la presente investigación se pretende crear e interpretar un modelo de minería de datos en series temporales para generar conocimiento en cuanto al pronóstico de precipitaciones pluviales que pueden ocurrir a futuro, para evadir pérdidas en los cultivos.

4.1.4.2. Comprensión de los datos

Posterior a entender el tema se recolectan todos los datos que sean necesarios para la investigación y generación del conocimiento referentes a la problemática de estudio.

Los datos con los que se cuenta corresponden al histórico de datos de la estación meteorológica de la Finca Experimental San Francisco de la Universidad Politécnica Estatal del Carchi, cuenta con datos desde junio 2018 hasta agosto 2021, con un registro de variables a cada hora, es decir que se obtiene 24 registros por día, los parámetros de variables se almacenan en 12 canales que son:

Tabla 18. Variables estación meteorológica

	Variable	Indicador	Unidad de Medida
Canal 1	WS (ave)	Promedio Velocidad Viento	m/s
Canal 2	WD (ave)	Dirección	grados
Canal 3	WS (max)	Viento máximo	m/s
Canal 4	WD (most)	Puntos cardinales	Norte, Sur, Este, Oeste
Canal 5	WS (inst_m)	Viento sostenido	m/s
Canal 6	WD (inst_m)	dirección viento sostenido	grados
Canal 7	Max_time	Hora viento máximo	Formato hora
Canal 8	Solar_rad	Radiación solar	MJ MegaJouls
Canal 9	TEMP	Temperatura	Grados Centigrados
Canal 10	Humidity	Porcentaje de humedad	Porcentaje
Canal 11	Rainfall	Precipitación	mm
Canal 12	Bar_press	Presión atmosférica	hPa Pascal

La tabla 18 indica los canales donde se almacenan las diferentes variables meteorológicas de la estación, su interpretador y la unidad de medida

La predicción se realizará para las precipitaciones pluviales (rainfall) el cual es el principal problema de pérdidas de cultivos en el cantón san Pedro de Huaca ya que existen meses en donde las precipitaciones a través de sus diferentes manifestaciones son de alto grado.

La base de datos fue facilitada por el MSc. Diego Caicedo quien es el responsable de las diferentes estaciones meteorológicas que son parte de la universidad, se encuentra en Excel con la extensión .xlsx y luce de la siguiente forma:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	No	Date	Time	Unit,0 ColdJunc0 BC	PowerVolt V	PowerKind	CH001 WS(ave) m/s	CH002 WD(ave) β	CH003 WS(max) m/s	CH004 WD(most)	CH005 WS(inst_m) m/s	CH006 WD(inst_m) β	CH007 Max_time	CH008 Solar_rad MJ	CH009 TEMP BC	CH010 Humidity %	CH011 Rainfall mm	CH012 Bar_press, hPa
2	454	3/6/2018	20:40:00	11,9	13,1	2	0	74	0	NE	0,7	162	20:33:58	0	11,5	85,43	0	719,9
3	455	3/6/2018	20:50:00	11,8	13,1	2	0	219	0	W	0,7	200	20:44:16	0	11	89,43	0	719,9
4	456	3/6/2018	21:00:00	11,7	13	2	0	271	0	W	0	265	20:50:01	0	10,6	92,53	0	720,1
5	457	3/6/2018	21:10:00	11,5	13,1	2	0,4	301	0,4	WNW	1,9	304	21:07:14	0	10,4	94,73	0	720,1
6	458	3/6/2018	21:20:00	11,4	13	2	0	84	0,4	E	0,7	76	21:13:02	0	10,4	94,33	0	720,1
7	459	3/6/2018	21:30:00	11,2	13	2	0,1	294	0,1	WNW	1,3	274	21:21:57	0	10,3	95,33	0	720,2
8	460	3/6/2018	21:40:00	11	13	2	0	311	0,1	NW	0,7	312	21:32:13	0	10,2	95,43	0	719,9
9	461	3/6/2018	21:50:00	10,8	13	2	0	308	0	WNW	0	295	21:40:01	0	10,1	96,12	0	720,1
10	462	3/6/2018	22:00:00	10,7	13	2	0	287	0	WNW	0	292	21:50:01	0	10,1	96,33	0	720,2
11	463	3/6/2018	22:10:00	10,6	13	2	0	289	0	WNW	0	282	22:00:01	0	10,1	96,43	0	720,4
12	464	3/6/2018	22:20:00	10,6	13	2	0	295	0	WNW	0	295	22:10:01	0	10,1	96,53	0	720,4
13	465	3/6/2018	22:30:00	10,6	13	2	0	290	0	WNW	0	295	22:20:01	0	9,9	97,13	0	720,4
14	466	3/6/2018	22:40:00	10,5	13	2	0	261	0	W	0	266	22:30:01	0	9,8	97,53	0	720,4
15	467	3/6/2018	22:50:00	10,4	13	2	0	269	0	W	0,7	54	22:48:30	0	9,9	97,52	0	720,4
16	468	3/6/2018	23:00:00	10,4	13	2	0,1	41	0,1	NE	1,3	49	22:53:59	0	10	97,53	0	720,5
17	469	3/6/2018	23:10:00	10,4	13	2	0	5	0,1	NE	0,7	294	23:09:12	0	9,9	97,62	0	720,4
18	470	3/6/2018	23:20:00	10,4	13	2	0	247	0	SW	1,3	252	23:14:36	0	9,8	97,73	0	720,4
19	471	3/6/2018	23:30:00	10,4	13	2	0	78	0	NE	0	217	23:20:01	0	9,9	97,83	0	720,4
20	472	3/6/2018	23:40:00	10,4	13	2	0	332	0	WNW	0	103	23:30:01	0	9,9	97,82	0	720,4
21	473	3/6/2018	23:50:00	10,4	13	2	0	284	0	WNW	0	296	23:40:01	0	9,9	97,73	0	720,2
22	474	4/6/2018	00:00:00	10,3	13	2	0	277	0	W	0	271	23:50:01	0	9,8	97,73	0	720,1

Figura 16. Base de datos estación meteorológica original

4.1.4.3. Preparación de los datos

Ya obtenido los datos históricos, hay que proceder con la transformación y la selección de los datos para un mejor manejo y que se encuentren óptimos para realizar el análisis de estos.

Como primer paso se realiza la transformación de los datos, quitando el encabezado y llevando del archivo de origen en Excel (.xlsx) a un archivo de valores separados por comas (.csv), ya que este es formato adecuado para proceder con el modelado de datos.

	A	B	C	D	E	F	G	H	I					
1	No	Date	Time	ColdJunc0	PowerVolt	PowerKind	WS(ave)	WD(ave)	WS(max)	WD(most)	WS(inst_m)	WD(inst_n		
2	454	03/06/2018	20:40:00	"11,9"	"13,1"	2,0	74,0	NE	"0,7"	162	20:33:58	0,"11,5"	"85,43"	0,"719,9"
3	455	03/06/2018	20:50:00	"11,8"	"13,1"	2,0	219,0	W	"0,7"	200	20:44:16	0,11,"89,43"	0,"719,9"	
4	456	03/06/2018	21:00:00	"11,7"	"13,2"	0,271	0,W	0,265	20:50:01	0,"10,6"	"92,53"	0,"720,1"		
5	457	03/06/2018	21:10:00	"11,5"	"13,1"	2,"0,4"	301,"0,4"	WNW	"1,9"	304	21:07:14	0,"10,4"	"94,73"	0,"720,1"
6	458	03/06/2018	21:20:00	"11,4"	"13,2"	0,84,"0,4"	E,"0,7"	76	21:13:02	0,"10,4"	"94,33"	0,"720,1"		
7	459	03/06/2018	21:30:00	"11,2"	"13,2"	"0,1"	294,"0,1"	WNW	"1,3"	274	21:21:57	0,"10,3"	"95,33"	0,"720,2"
8	460	03/06/2018	21:40:00	11,13	2,0	311,"0,1"	NW,"0,7"	312	21:32:13	0,"10,2"	"95,43"	0,"719,9"		
9	461	03/06/2018	21:50:00	"10,8"	"13,2"	0,308	0,WNW	0,295	21:40:01	0,"10,1"	"96,12"	0,"720,1"		
10	462	03/06/2018	22:00:00	"10,7"	"13,2"	0,287	0,WNW	0,292	21:50:01	0,"10,1"	"96,33"	0,"720,2"		
11	463	03/06/2018	22:10:00	"10,6"	"13,2"	0,289	0,WNW	0,282	22:00:01	0,"10,1"	"96,43"	0,"720,4"		
12	464	03/06/2018	22:20:00	"10,6"	"13,2"	0,295	0,WNW	0,295	22:10:01	0,"10,1"	"96,53"	0,"720,4"		
13	465	03/06/2018	22:30:00	"10,6"	"13,2"	0,290	0,WNW	0,295	22:20:01	0,"9,9"	"97,13"	0,"720,4"		
14	466	03/06/2018	22:40:00	"10,5"	"13,2"	0,261	0,W	0,266	22:30:01	0,"9,8"	"97,53"	0,"720,4"		
15	467	03/06/2018	22:50:00	"10,4"	"13,2"	0,269	0,W	"0,7"	54	22:48:30	0,"9,9"	"97,52"	0,"720,4"	
16	468	03/06/2018	23:00:00	"10,4"	"13,2"	"0,1"	41,"0,1"	NE	"1,3"	49	22:53:59	0,10,"97,53"	0,"720,5"	
17	469	03/06/2018	23:10:00	"10,4"	"13,2"	0,5,"0,1"	NE,"0,7"	294	23:09:12	0,"9,9"	"97,62"	0,"720,4"		
18	470	03/06/2018	23:20:00	"10,4"	"13,2"	0,247	0,SW	"1,3"	252	23:14:36	0,"9,8"	"97,73"	0,"720,4"	
19	471	03/06/2018	23:30:00	"10,4"	"13,2"	0,78	0,NE	0,217	23:20:01	0,"9,9"	"97,83"	0,"720,4"		
20	472	03/06/2018	23:40:00	"10,4"	"13,2"	0,332	0,WNW	0,103	23:30:01	0,"9,9"	"97,82"	0,"720,4"		
21	473	03/06/2018	23:50:00	"10,4"	"13,2"	0,284	0,WNW	0,296	23:40:01	0,"9,9"	"97,73"	0,"720,2"		

Figura 17. Base de datos estación meteorológica transformada

Se analizó los mm de precipitación pluvial a través de una base de datos que contiene información diaria desde 2018-06-03 hasta 2021-08-02.

Se procedió a realizar una limpieza de la información al determinar la cantidad de datos faltantes o fuera de los rangos estándares.

Para ello se creó una secuencia de fechas artificial que contiene los datos completos desde 2018-06-03 hasta 2021-08-02, a la cual se añadieron los datos completos de precipitación, seguido por una agregación de datos por día y no por hora como los datos crudos de origen.

La serie de tiempo se ve de la siguiente forma:

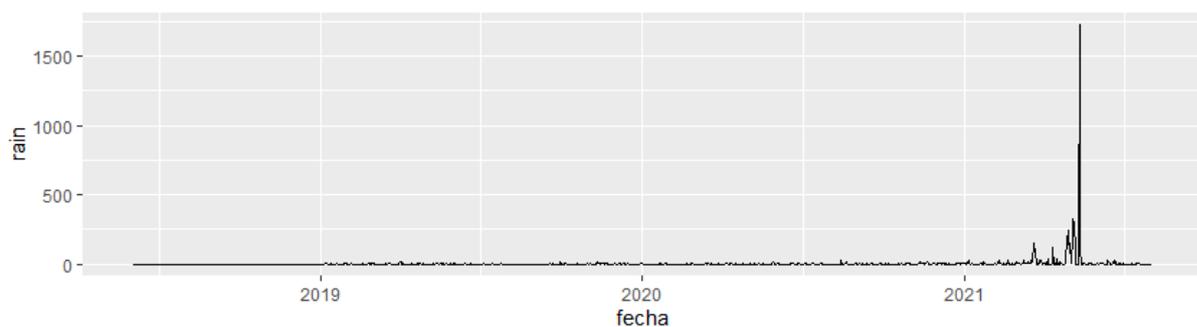


Figura 18. Serie de tiempo

Como se puede observar los datos a partir del año 2021 contienen valores muy elevados, llegando hasta un valor máximo de 1725mm, lo cual es irreal en comparación con los datos históricos e incompatible con los rangos de precipitación documentados.

Se procedió entonces a realizar un corte de la información tomando solo los valores anteriores a 2021-01-01, quedando una serie de tiempo de agregados diarios, como sigue:

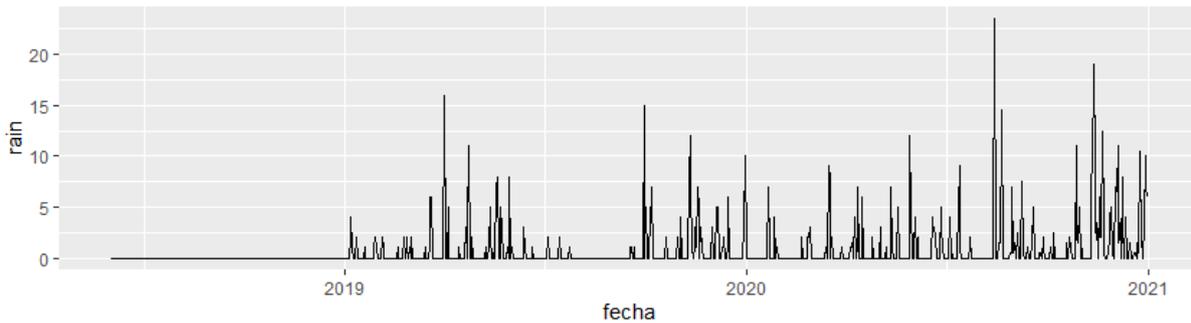


Figura 19. Serie de tiempo con corte de información

Como se puede distinguir, los datos del 2018 no contienen valores diferentes de 0, lo cual podría ser real o bien error de registro de datos, por lo cual se procedió a eliminar los datos anteriores al 2019, quedando como sigue:

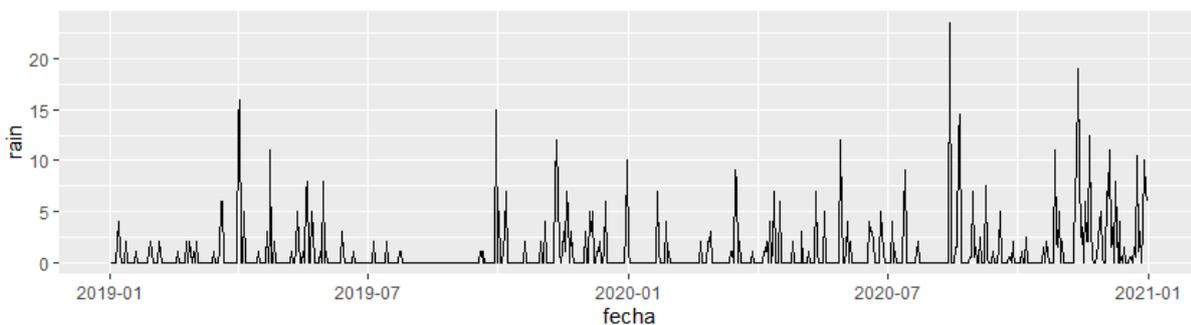


Figura 20. Serie de tiempo 2019 – 2021

Con esta delimitación de información podemos observar una mejor continuidad de la información, con lo cual haremos agregados mensuales para verificar si podemos encontrar alguna estacionalidad en la serie de tiempo:

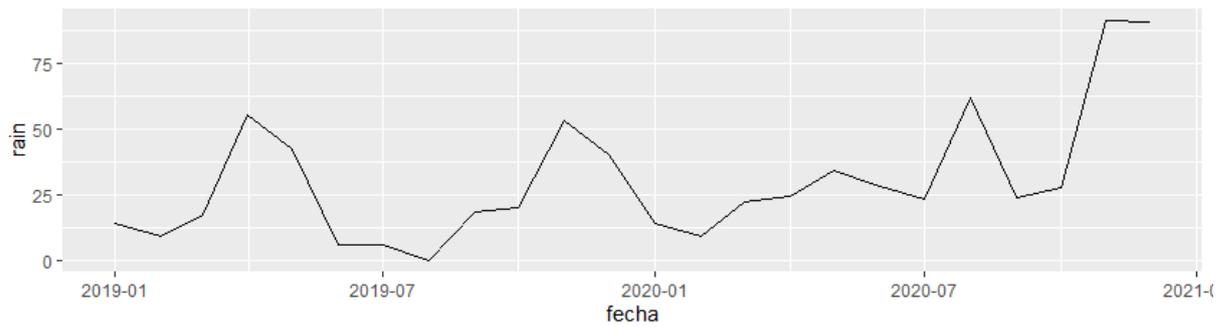


Figura 21. Estacionalidad en serie de tiempo

4.1.4.4. Modelamiento

Dado que se tienen dos años completos de datos, se procedió a realizar una descomposición aditiva de la serie de tiempo de agregados mensuales para determinar el comportamiento general de los datos, quedando como sigue:

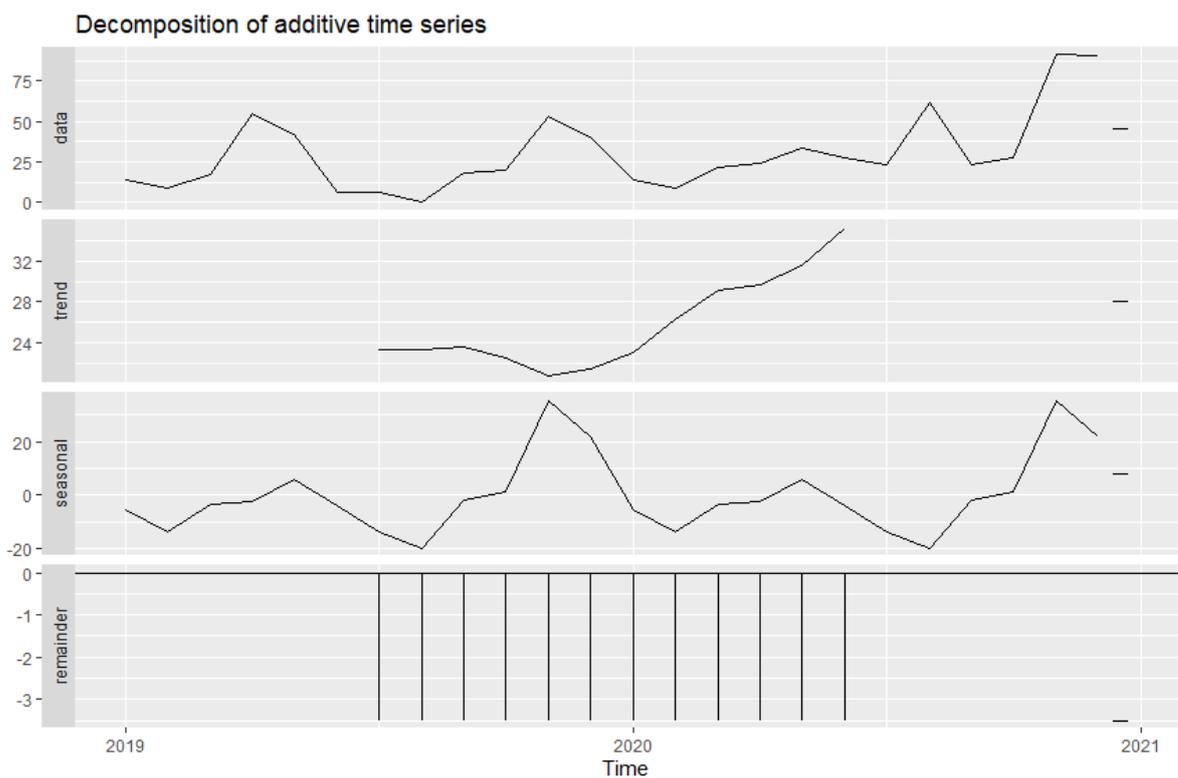


Figura 22. Descomposición aditiva de series de tiempo

Con esta gráfica podemos observar como la parte tendencial (trend) tiene un crecimiento notable, lo que significaría que la precipitación pluvial tiende a aumentar con el paso de los años. Por otro lado, vemos como la parte estacional (seasonal) refleja que aproximadamente durante los segundos y los últimos trimestres de cada año se tienen las mayores cantidades de precipitación.

Con los datos agregados mensuales procedimos a realizar una proyección de datos a 6 meses usando un pronóstico a partir de promedios móviles (SMA).

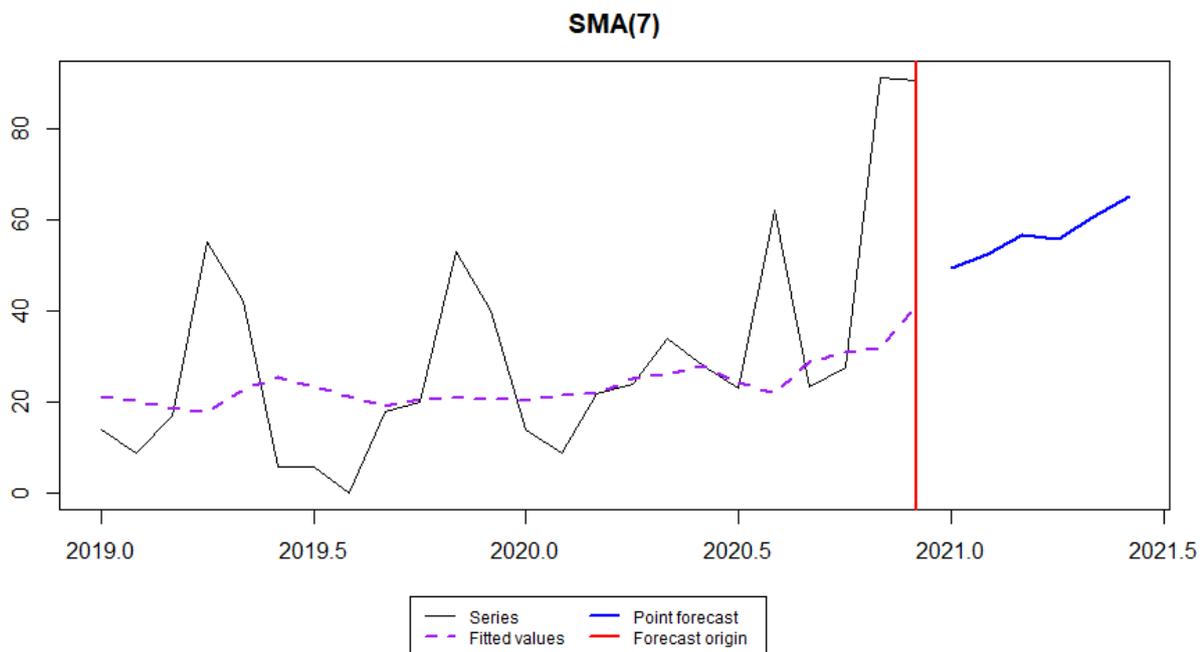


Figura 23. Predicción con promedios móviles SMA

4.1.4.5. Evaluación

Para la evaluación se realiza un promedio de precipitación mm por cada mes en los dos años que son con los cuales se a trabajado para este modelo predictivo, esto se lo hace para identificar la tendencia de aumento de precipitaciones en los diferentes meses y años.

Tabla 19. Fitted años 2019 – 2020

Fitted	Jan	Feb	Mar	Apr	May	Jun
2019	21,28571	20,2449	18,4898	17,87755	22,69388	25,65306
2020	20,42857	21,57143	22	25,14286	26	28
	Jul	Aug	Sep	Oct	Nov	Dec
2019	23,46939	21,28571	19,28571	20,57143	21	20,71429
2020	24,42857	22	28,85714	30,92857	31,71429	41,28571

En la Tabla 19 observamos los valores promedio para cada mes en cuanto a precipitaciones pluviales, cabe recalcar que la unidad de medida es (mm) milímetros cuadrados.

El siguiente paso es interpretar los datos del forecast (predicción) que se encontraron realizando el análisis predictivo de serie de tiempo, encontrando un enorme ascenso de los valores para las fechas correspondientes a los doce meses próximos del año 2021 tal como se ve en la figura de SMA.

Tabla 20. Forecast año 2021

Forecast	Jan	Feb	Mar	Apr	May	Jun
2021	49,35714	52,40816	56,60933	55,83923	60,45912	65,16757
	Jul	Aug	Sep	Oct	Nov	Dec
2021	61,47722	57,33111	58,47025	59,33626	59,72583	60,28105

En la tabla 20 observamos los valores encontrados mediante la predicción basada en sma, para los próximos meses del año 2021.

4.1.4.6. Despliegue

En la etapa de despliegue el objetivo es llevar estos modelos a la parte práctica e interactiva, a hacer uso de esto modelo de forma repetida y continúa siendo útil no solamente para los datos con los que se trabajó si no para todo aquel que tenga una base de datos histórica similar a la utilizada para el desarrollo que tenga los mismos parámetros.

Es por eso que para cumplir con este punto de la metodología CRISP-DM se ha realizado un aplicativo informático en forma de página web con shiny apps en el lenguaje de R, permitiendo realizar y descargar dashborad dinámicos para el usuario, generando predicciones con los datos que el usuario quiera descargar.

Una Shiny App es una página web (UI) en el lenguaje de programación R, la cual está conectada a un servidor de R, los usuarios pueden manipular el UI, y la función del servidor es actualizar la UI a medida que se vaya usando.

Como primer paso se registra de forma gratuita en la página de shinyapps.io server para tener acceso al servidor en donde nuestra shinyapp.io par a ser alojada, una vez creada la cuenta

encontramos un dashboard con toda la información referente a nuestro aplicativo web para tener un manejo más cauteloso.

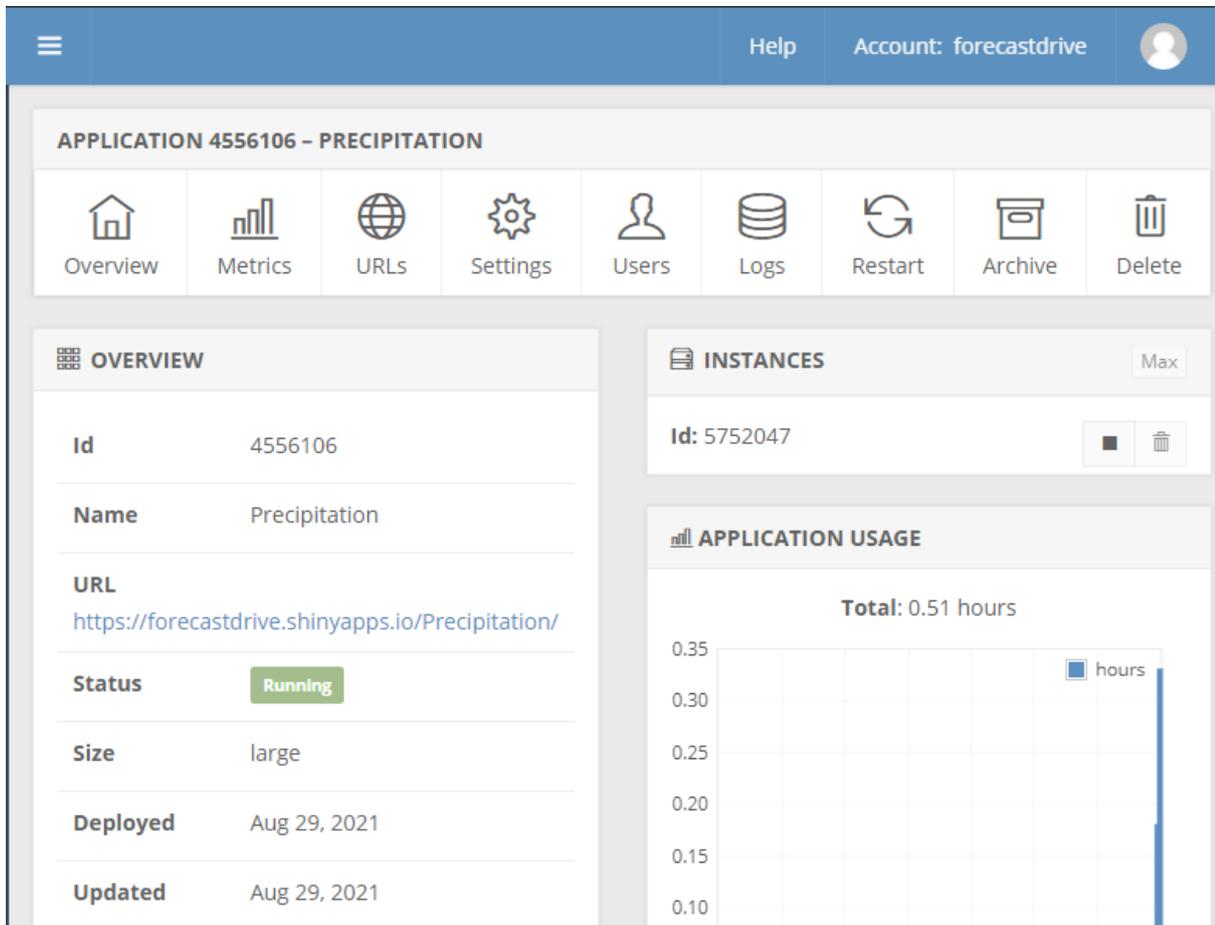


Figura 24. Dashboard Shinyapps.io server

La página principal de la aplicación cuenta un informativo de las funciones que presenta, en la parte inferior se encuentra un botón (Browse) el cual cumple la función de buscar y subir el archivo separado con comas para realizar el análisis predictivo, mientras que en la parte izquierda encontramos dos menús adicionales exploración y pronóstico, un input del rango de fechas indicando las que van a ser tomadas en cuenta, y un sidebar que indicará los meses de pronóstico que son requeridos.



Figura 25. Página principal Prediction

En el menú de pronóstico lo primero que observamos es la serie de tiempo de la relación fecha y precipitación de nuestra tabla de datos base, a medida que modificamos los campos del rango de tiempo estos cambiarán mostrando al usuario la serie de tiempo deseada en el momento deseado.

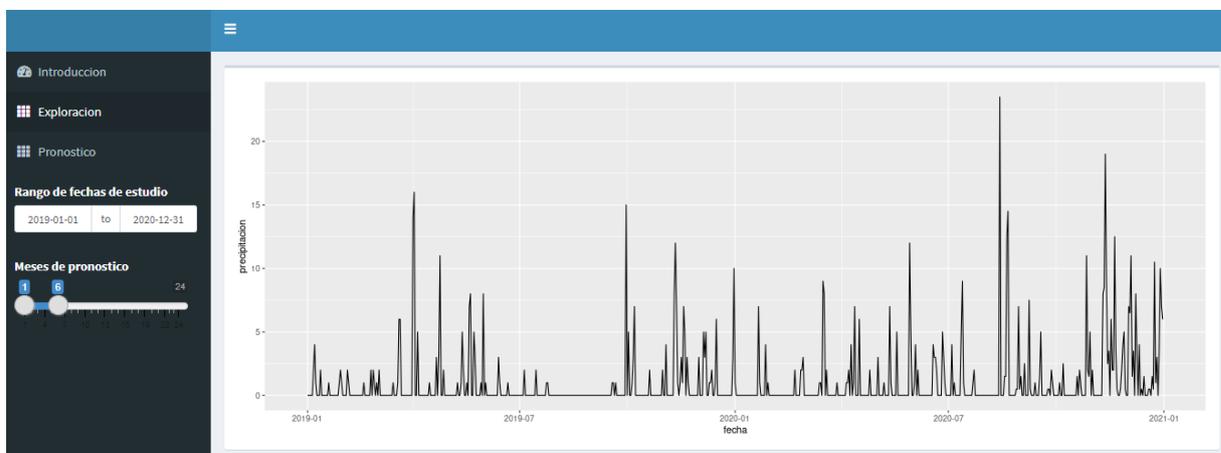


Figura 26. Serie de tiempo Exploración

En la parte superior del menú de exploración dos tablas que muestran 10 entradas este campo pueden ser modificados, en esta instancia se muestra la fecha y el valor de precipitación por día, y la precipitación acumulada, en esta parte de la exploración ya comienza a trabajar el modelo anteriormente realizado.

Además, este apartado cuenta con tres botones de copia, CSV, Excel permitiendo descargar en estos dos tipos de formatos los datos que sean requeridos.

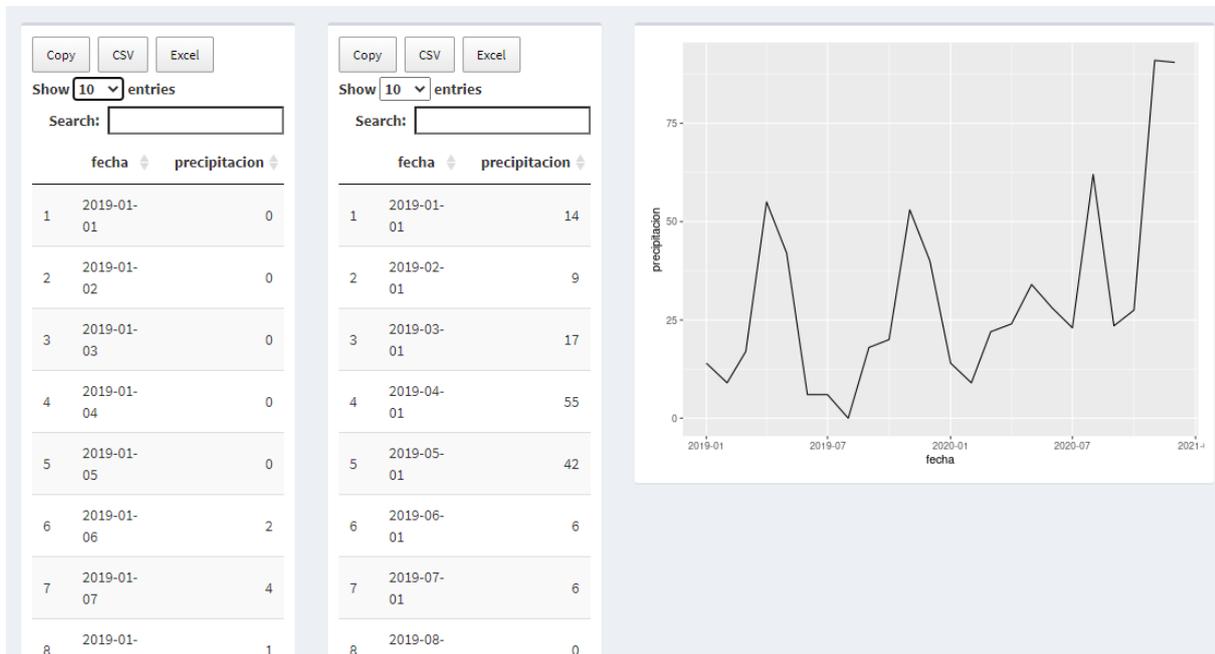


Figura 27. Tablas de Exploración

	A	B	C	D
1	Exported data			
2		fecha	precipitacion	
3	1	2019-01-01	14	
4	2	2019-02-01	9	
5	3	2019-03-01	17	
6	4	2019-04-01	55	
7	5	2019-05-01	42	
8	6	2019-06-01	6	
9	7	2019-07-01	6	
10	8	2019-08-01	0	
11	9	2019-09-01	18	
12	10	2019-10-01	20	
13				
14				

Figura 28. Archivo descargado en formato.xlsx

El ultimo menú muestra el histograma con el pronóstico de precipitaciones pluviales aplicado con el modelo de SMA, además de igual forma muestra las tablas con los mismos botones de descarga, pero con la diferencia de que aquí obtenemos los datos de los valores ajustados y de los pronósticos para cada mes respectivamente.

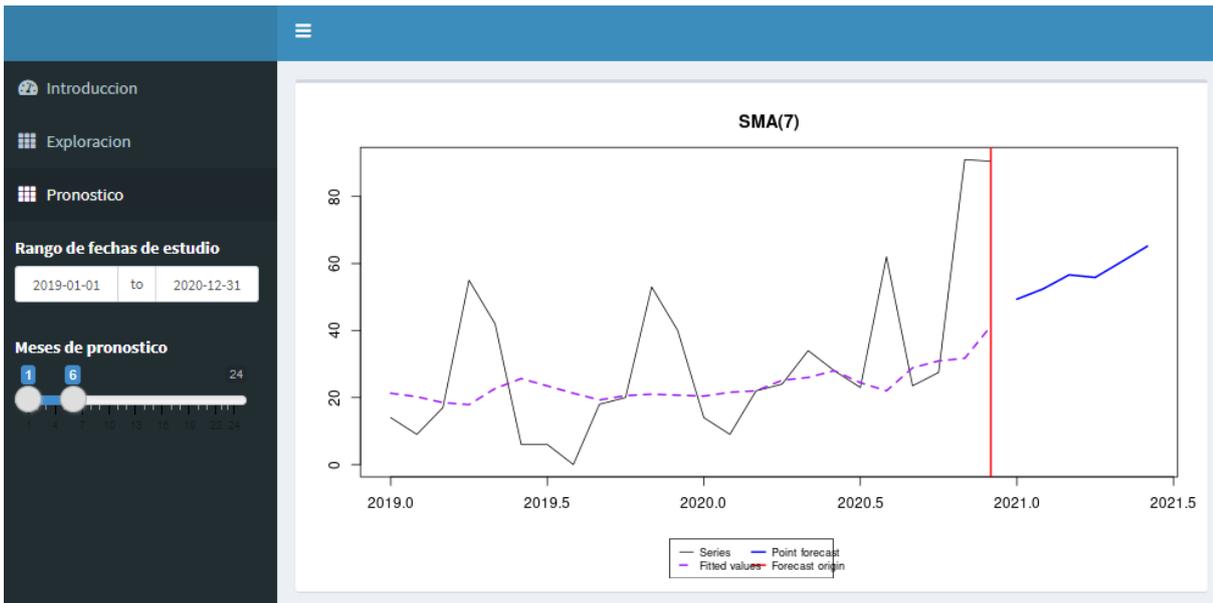


Figura 29. Histograma SMA Pronostico

Copy CSV Excel Show 10 entries

Search:

	fecha	fited_values
1	2019-01-01	22.1578947368421
2	2019-02-01	21.7285318559557
3	2019-03-01	21.0360110803324
4	2019-04-01	20.7645429362881
5	2019-05-01	22.4930747922438
6	2019-06-01	23.5373961218837
7	2019-07-01	22.6869806094183
8	2019-08-01	21.8365650969529
9	2019-09-01	20.6703601108033
10	2019-10-01	20.4515235457064

Showing 1 to 10 of 19 entries

Previous 1 2 Next

Copy CSV Excel Show 10 entries

Search:

	fecha	pronostico
1	ago 2020	22.1578947368421
2	sep 2020	22.5872576177285
3	oct 2020	23.3023764397142
4	nov 2020	23.6340804628571
5	dic 2020	21.9832425924812
6	ene 2021	20.929729044717

Showing 1 to 6 of 6 entries Previous 1 Next

Figura 30. Tablas de datos valores ponderados y pronostico

Cabe destacar que el aplicativo informático es responsive esto quiere decir que la aplicación puede ser abierta desde el navegador web de cualquier dispositivo móvil con acceso a internet sin perder ninguna de las funcionalidades que tiene, adaptándose al tamaño de pantalla disponible.

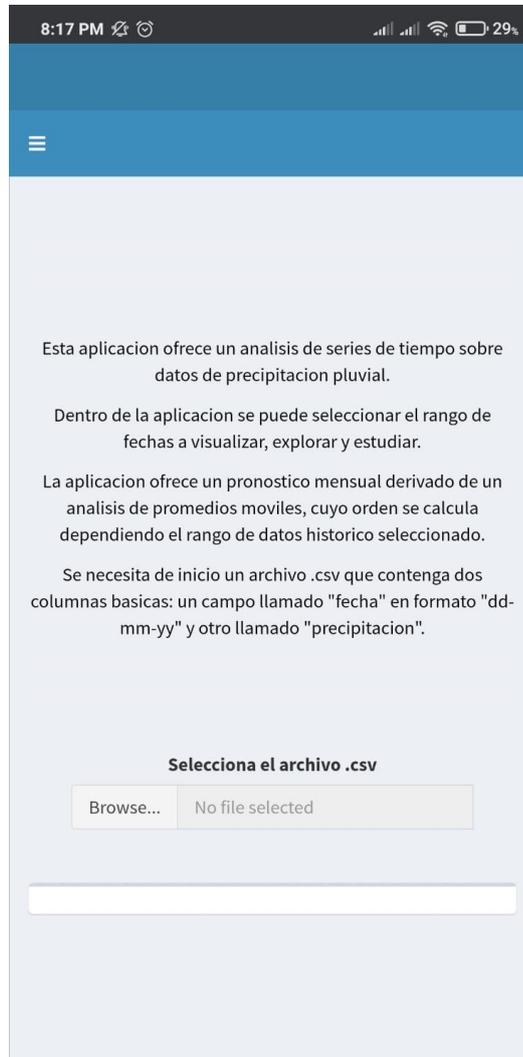


Figura 31. Página principal vista desde móvil

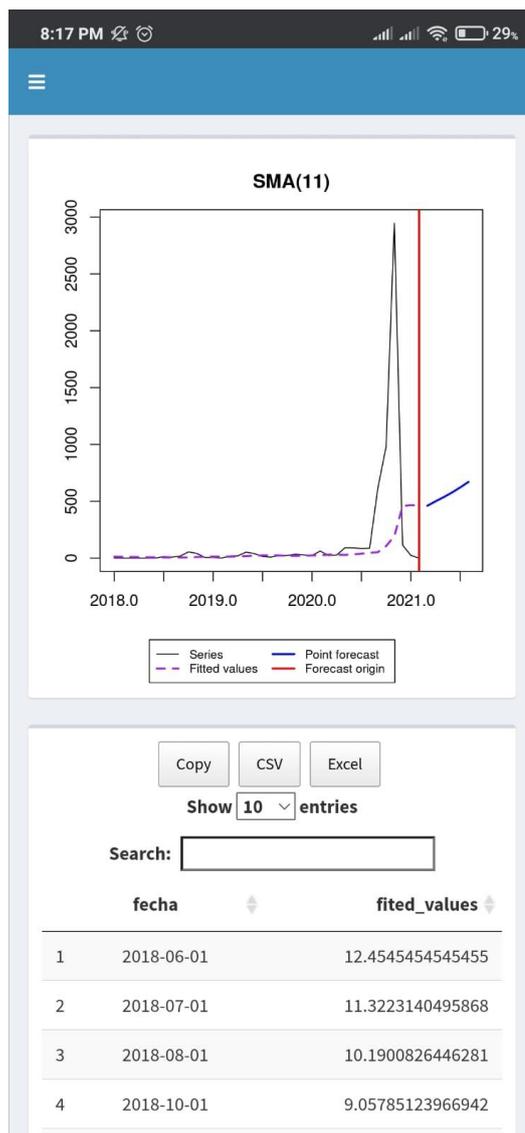


Figura 32. Menú pronostico vista desde móvil

Fuente: Elaboración propia

4.1.5. Elaboración de un Dashboard

Para la visualización, interpretación e interacción de los datos con los usuarios, se realizó un dashboard con la herramienta tableau public, la cual nos permite subir las bases de datos generadas con el aplicativo para hacer un tratamiento de los datos en gráficos estadísticos que sean más digeribles y fáciles de entender para el público en general permitiendo conocer los máximos y bajos de las precipitaciones pluviales en cada mes, como también los pronósticos del presente año y los valores ajustados.



Figura 33. Dashboard forecast

Una vista general del dashboard realizado para las precipitaciones pluviales, cuenta con diferentes tablas y gráficos estadísticos que contienen información sobre sobre promedios, valores ajustados de los años que fueron tomados en cuenta para el análisis estadístico para obtener la tabla de pronósticos del año 2021.



Figura 34. Indicadores de colores

En la parte superior encontramos los indicadores de color para cada variable que se encuentra en el dashboard estos indicadores van de un color claro que indica el menor valor hasta uno más oscuro que indica el mayor valor para esa variable.

Al ser una herramienta interactiva podemos organizar mediante filtros la información presentada en el dashboard según sea la necesidad o el orden.

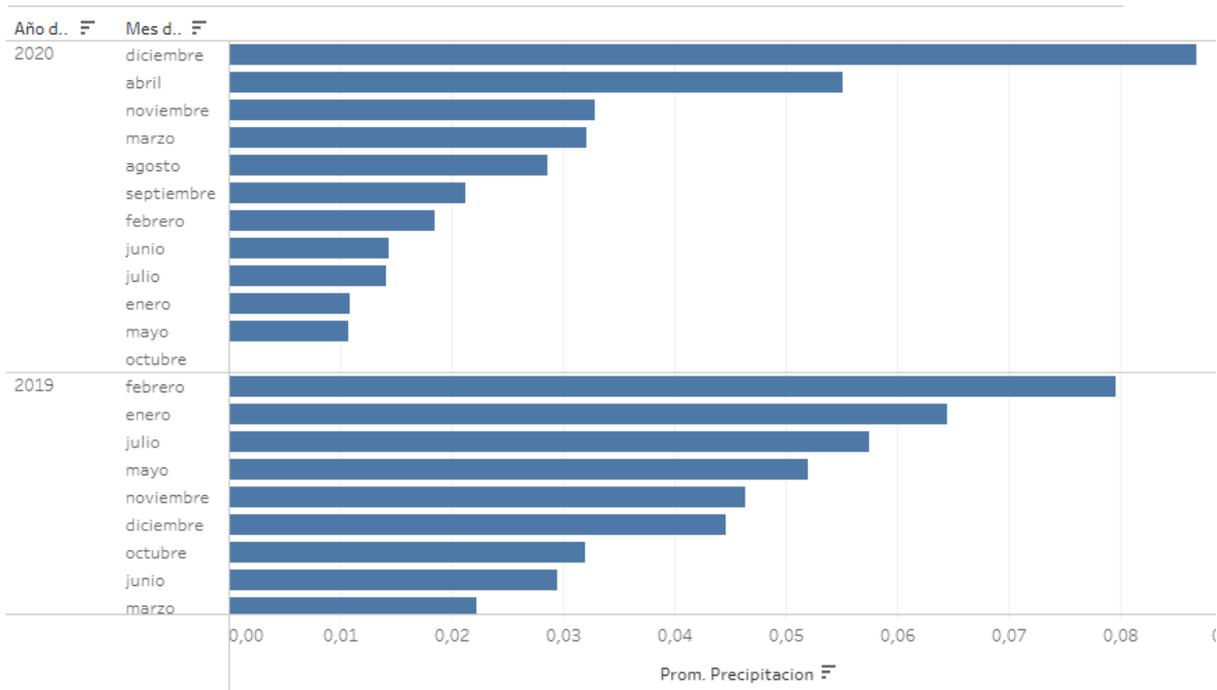


Figura 35. Promedios de precipitaciones orden ascendente

La tabla azul indica los valores promedio que fueron obtenidos en los años anteriores organizados por meses y años, esta figura cuenta con un filtro el cual permite ordenar los datos de forma ascendente, descendiente o en el orden cronológico de los meses, además cuenta con filtros para cambiar a trimestres por cada año.



Figura 36. Valores Justados 2019-2020

La tabla de color naranja indica los valores ajustados, estos datos hacen correspondencia a la media estimaciones de punto de la respuesta media para los valores dados de los predictores.

Pronóstico 2021)mm= por metro cuadrado

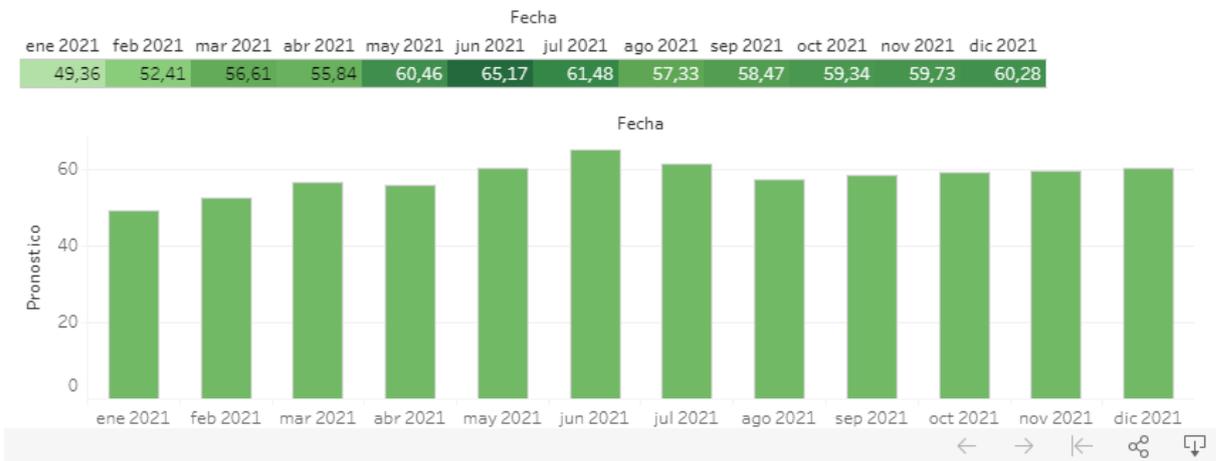


Figura 37. Tabla pronóstico 2021

La tabla de color verde indica los pronósticos que fueron encontrados haciendo uso de modelo predictivo para los meses del año 2021, indicado un valor acumulado por mes cuya medida es los milímetros por metro cuadrado de superficie.

4.2. DISCUSIÓN

El objetivo principal de la presente investigación fue desarrollar un aplicativo informático de análisis de datos históricos de la Finca Experimental San Francisco de Huaca, utilizando minería de datos para series temporales.

Se realizó un sondeo de la situación actual de la localidad, conociendo cual es el tratamiento posterior que se les da a los datos históricos que almacena la estación meteorológica de la Finca Experimental. Coincidiendo con el autor Vázquez (2017), donde menciona que es posible aplicar técnicas de minería de datos para crear reglas de tendencia haciendo uso de variables meteorológicas como la temperatura, unidad de calor y las precipitaciones pluviales de una estación meteorológica.

Con la aplicación de la herramienta de recolección de datos encuesta, fue aplicada a 176 personas que se dedican a la actividad agrícola dentro del cantón San Pedro de Huaca, se obtuvo como resultado la factibilidad y la ayuda que brindaría un aplicativo informático que facilite la interpretación y visualización de los datos analizados para la predicción de precipitaciones pluviales, como menciona Soto (2013) quien indica que diseñar e implementar un paquete de minería de datos en WEKA facilita las tareas de clasificación de series temporales, permitiendo

aplicar métodos nuevos como los tradicionales al problema de las precipitaciones, en concordancia con lo mencionado también se hizo uso de los métodos tradicionales del pronóstico para la evaluación del mismo.

Por otra parte en desacuerdo a lo mencionado por Soto (2013) no se utilizó la herramienta para minería de datos Open Source WEKA, ya que es muy limitado en su uso, por esto se utilizó el lenguaje de programación R y el intérprete R estudio haciendo uso de las diferentes librerías de análisis de datos que están disponibles para realizar un modelo de la mejor manera posible, permitiendo tener un pronóstico más exacto e interactivo permitiendo visualizar de mejor manera histogramas y tablas de los datos obtenidos.

Coincidiendo con la autora Zamora (2018), quien indica que en su proyecto hizo uso de la metodología CRISP-DM ya que ayuda a tener mejores resultados y una mejor evaluación, llevando a cabo el proyecto de una manera más ordenada y comprendiendo cada fase que lo conforma prestando atención en cada etapa, en especial en el modelado y evaluación, al igual que lo mencionado por la autora se hizo uso de la metodología CRISP-DM para la elaboración de nuestro análisis, como primera parte identificando el problema la cual se plantea dar una solución mediante la extracción del conocimiento, analizar y conocer los datos con los que se van a trabajar para posteriormente entrar en la fase de transformación donde se realiza una limpieza de los datos al determinar que existen datos faltantes o que se encuentran fuera de rango, para el modelamiento se utilizó una descomposición aditiva de la serie de tiempo de agregados mensuales determinado en general el comportamiento de los datos por separado, para luego utilizar promedios móviles para obtener una predicción acertada, finalmente las fases de evaluación donde se compara los resultados obtenidos por el histograma y una tabla de pronósticos, la fase de despliegue donde se indica que el modelo puede ser usado no solo a una determinada base de datos si no a cualquiera que cumpla con los parámetros requeridos, haciendo uso del aplicativo informático en tiempo real desde cualquier dispositivo con acceso a internet.

El aplicativo informático y el modelo de análisis de datos está desarrollado en base a las diferentes herramientas y tecnologías como: Lenguaje R, R studio metodología CRISP-DM y el uso de la técnica de promedios móviles, además de shinyapps.io server.

El aplicativo informático contiene tres módulos para que los usuarios agricultores puedan interactuar con el modelo y los datos históricos que almacenan en sus propias estaciones meteorológicas con la finalidad de cumplir con la meta propuesta de predecir las futuras precipitaciones.

V. CONCLUSIONES Y RECOMENDACIONES

5.1. CONCLUSIONES

- Se pudo concluir que mediante la fundamentación teórica permitió conocer como las precipitaciones afectan la calidad, oferta y demanda de los cultivos al ser una actividad que depende mucho de los fenómenos climáticos.
- La utilización de diferentes métodos y metodologías fueron de gran ayuda para la presente investigación, al igual que las herramientas de recolección de datos que permitieron generar conocimiento a partir de la información obtenida.
- El presente modelo de pronóstico basado en promedios móviles es una buena aproximación hacia el pronóstico acumulado mensual de precipitación mensual.
- Su implementación en aplicaciones interactivas permite una exploración de datos muy práctica y flexible, permitiendo interpretar el análisis de datos históricos de la estación meteorológica de la finca experimental San Francisco de Huaca de una manera más digerible y entendible para predecir futuras precipitaciones.
- Finalmente, el análisis de series de tiempo resultan ser una poderosa herramienta para la evaluación proyectos de investigación.

5.2. RECOMENDACIONES

- Conseguir una base de datos con tanto histórico como sea posible, ya que mientras con más datos se tenga mayor será la precisión de la predicción permitiendo de esta manera tener resultados óptimos y tomar mejores decisiones.
- Es posible generar mejores modelos a partir de más puntos de colección de datos, por lo que agregar información de otras estaciones meteorológicas sería recomendable.
- Hacer pruebas de calidad sobre la obtención de datos, ya que en el presente trabajo se observa como los valores a partir del año 2021 se vuelven fuera del comportamiento normal, y esto puede ser debido a fallas técnicas de la estación.
- Generar modelos de pronóstico diario es posible si se agrega información con filtros de calidad adecuados, así como también sistematizar las estaciones meteorológicas para tener un mejor manejo de los datos.
- Generar modelos de pronóstico a partir de otras técnicas y metodologías de minería de datos para obtener diferentes resultados y hacer un estudio de que técnica es más aceptable en la aplicación de predicciones en precipitaciones pluviales.

VI. REFERENCIAS BIBLIOGRÁFICAS

- Agencia Estatal de Metereología. (2018). meteoglosario.aemet. Recuperado de https://meteoglosario.aemet.es/es/termino/417_variable-meteorologica
- Anastacio, L., & Barzola, M. (2019). *IMPLEMENTACIÓN DE UNA ESTACIÓN METEOROLÓGICA PARA LA FACULTAD DE CIENCIAS NATURALES DE LA UNIVERSIDAD DE GUAYAQUIL UTILIZANDO HERRAMIENTAS OPEN SOURCE PROYECTO*. Universidad De Guayaquil.
- Bolaños Logroño, P. F. (2018). *Predicción de la velocidad del viento mediante redes neuronales, para la Estación Meteorológica Chimborazo*. (ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO). ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO. Recuperado de <http://dspace.esPOCH.edu.ec/handle/123456789/8308>
- Bonilla Ambrossi, A. F. (2019). *CLASIFICACIÓN SONORA DE INSTRUMENTOS ANDINOS A TRAVÉS DE MACHINE LEARNING*.
- Caballero, L. (2017). El camino del éxito de las encuestas y entrevistas. *Documentos de Docencia* N° 30, 32. Recuperado de <https://revistas.ucc.edu.co/index.php/dodo/article/view/2282/2326>
- Escutia, I. (2019). Descomposición de series de tiempo. Recuperado de https://rstudio-pubs-static.s3.amazonaws.com/546278_da6272365edd4444a4e7ccf6b17978fe.html
- Guamán Pozo, J. M. (2020). *Índice de cambio climático y su afectación a la agricultura, caso de estudio cantón Ambato*. Universidad Politécnica Salesiana Sede Quito.
- Guayasamín Guanca, C. A. (2018). *DATA MINING Y ANÁLISIS DE DATOS APLICADO AL MINISTERIO DEL CONOCIMIENTO Y TALENTO HUMANO EN ECUADOR DEL 1993 A 2015* (Universidad de las Fuerzas Armadas). Universidad de las Fuerzas Armadas. Recuperado de <http://repositorio.espe.edu.ec/bitstream/21000/16067/1/T-ESPE-038590.pdf>
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, M. del P. (2018).

METODOLOGÍA de la Investigación (Quinta Edi). México: Mc Graw Hill.

Herrera Oliva, C. S., Campos Gaytán, J. R., & Carrillo González, F. M. (2017). Estimación de datos faltantes de precipitación por el método de regresión lineal: Caso de estudio Cuenca Guadalupe, Baja California, México. *Investigación y Ciencia de la Universidad Autónoma de Aguascalientes*, (71), 34–44. <https://doi.org/10.33064/iycuaa201771598>

Ilbay Yupa, M. L. (2019). “*TENDENCIA ESPACIO-TEMPORAL DE LA PRECIPITACIÓN, SU AGRESIVIDAD Y CONCENTRACIÓN EN LA REGIÓN INTERANDINA DEL ECUADOR*”. UNIVERSIDAD NACIONAL AGRARIA LA MOLINA.

IONOS. (2018). Software de data mining: realiza análisis de datos más efectivos. Recuperado de <https://www.ionos.es/digitalguide/online-marketing/analisis-web/software-de-data-mining-las-mejores-herramientas/>

Lapo Paredes, L. E. (2019). *Dinámica de la Agricultura Familiar Campesina en el Cantón Catamayo* (Universidad Nacional de Loja). Universidad Nacional de Loja. Recuperado de [https://dspace.unl.edu.ec/jspui/bitstream/123456789/22705/1/GABRIELA MAYTTE ORDOÑEZ ORDOÑEZ.pdf](https://dspace.unl.edu.ec/jspui/bitstream/123456789/22705/1/GABRIELA_MAYTTE_ORDOÑEZ_ORDOÑEZ.pdf)

León Guzmán, E. (2017). Minería de Datos Módulo Diplomado. *Módulo Universidad Nacional de Colombia*, 1–19. Recuperado de http://disi.unal.edu.co/~eleonguz/cursos/md/presentaciones/Sesion5_Metodologias.pdf

Mancero, H. (2017). *ESTIMACIÓN DEL COEFICIENTE DE HURST DE LAS SERIES TEMPORALES DE TRÁFICO VEHICULAR EN ZONAS URBANAS*. Universidad de las Fuerzas Armadas.

Ortiz Farro, P. E. (2017). “*MINERÍA DE DATOS CON SERIES DE TIEMPO EN EL DESARROLLO E IMPLEMENTACIÓN DEL SISTEMA INTELIGENTE QUE PREDICE LA PRODUCCIÓN DE ARROZ EN EL ÁMBITO DE LA GERENCIA REGIONAL DE AGRICULTURA – LAMBAYEQUE*.” AUTOR(A): Universidad Señor De Sipán.

Paucar Carrión, K. Y. (2019). *Big Data y media*. Universidad Técnica Particular de Loja.

Rivera, S. H., Lema, L. Z., Freire, A. M., Rojas, L. V., & Villa, A. E. (2018). MÉTODOS DE CLASIFICACIÓN EN MINERÍA DE DATOS METEOROLÓGICOS. *Revistas ESPOCH*, 107–113.

Tixe Bustamante, M. S. (2021). *MODELAMIENTO DIMENSIONAL DE LA INFORMACIÓN DE PRODUCCIÓN DE CRUDO DE DOS CAMPOS PETROLEROS DEL GRUPO PETROGAS S.A CON QLIK PLATAFORMA DE ANALÍTICA DE DATOS PARA BUSINESS INTELLIGENCE*. (Universidad Politécnica Salesiana Sede Quito). Universidad Politécnica Salesiana Sede Quito. Recuperado de <http://dspace.ups.edu.ec/bitstream/123456789/5081/1/UPS-CYT00109.pdf>

Unir. (2019). Lenguaje R, ¿qué es y por qué es tan usado en Big Data? Recuperado de <https://www.unir.net/ingenieria/revista/lenguaje-r-big-data/>

Vazquez, M. (2018). Minería de datos para generación de reglas de tendencia de precipitación pluvial en el estado de Morelos. *Ingeniería-Revista Académica de la facultad de Ingeniería*, 22, 9–24.

Westreicher, G. (2020). Método deductivo. Recuperado de <https://economipedia.com/definiciones/metodo-deductivo.html>

Zamora Villalobos, T. F. (2018). Aplicación De Tecnicas De Minería De Datos Para Pronósticos Del Sector Agrícola. *Pontificia Universidad Católica De Valparaíso*.

Anexo 2: Informe de Turnitin

TESIS

INFORME DE ORIGINALIDAD

4%	4%	1%	3%
INDICE DE SIMILITUD	FUENTES DE INTERNET	PUBLICACIONES	TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	repositorio.upec.edu.ec Fuente de Internet	2%
2	www.unir.net Fuente de Internet	1%
3	www.ionos.es Fuente de Internet	1%

Excluir citas Activo Excluir coincidencias < 1%
Excluir bibliografía Activo



Firmado
Electronicamente por:
MSC. MARCO
ANTONIO YANDÚN
VELASTEGUÍ
2021-09-03
19:59-05:00

Anexo 3: Certificado validación del Abstract



UNIVERSIDAD POLITÉCNICA ESTATAL DEL CARCHI
FOREIGN AND NATIVE LANGUAGE CENTER

ABSTRACT- EVALUATION SHEET				
NAME: Carlos Fernando Montenegro Argoti				
DATE: 15 de septiembre de 2021				
TOPIC: "Minería de datos para series temporales y su aplicación en las precipitaciones pluviales en la Finca Experimental San Francisco de Huaca"				
MARKS AWARDED		QUANTITATIVE AND QUALITATIVE		
VOCABULARY AND WORD USE	Use new learnt vocabulary and precise words related to the topic	Use a little new vocabulary and some appropriate words related to the topic	Use basic vocabulary and simplistic words related to the topic	Limited vocabulary and inadequate words related to the topic
	EXCELLENT: 2 <input checked="" type="checkbox"/>	GOOD: 1,5 <input type="checkbox"/>	AVERAGE: 1 <input type="checkbox"/>	LIMITED: 0,5 <input type="checkbox"/>
WRITING COHESION	Clear and logical progression of ideas and supporting paragraphs. <input checked="" type="checkbox"/>	Adequate progression of ideas and supporting paragraphs. <input type="checkbox"/>	Some progression of ideas and supporting paragraphs. <input type="checkbox"/>	Inadequate ideas and supporting paragraphs. <input type="checkbox"/>
	EXCELLENT: 2 <input checked="" type="checkbox"/>	GOOD: 1,5 <input type="checkbox"/>	AVERAGE: 1 <input type="checkbox"/>	LIMITED: 0,5 <input type="checkbox"/>
ARGUMENT	The message has been communicated very well and identify the type of text <input checked="" type="checkbox"/>	The message has been communicated appropriately and identify the type of text <input type="checkbox"/>	Some of the message has been communicated and the type of text is little confusing <input type="checkbox"/>	The message hasn't been communicated and the type of text is inadequate <input type="checkbox"/>
	EXCELLENT: 2 <input checked="" type="checkbox"/>	GOOD: 1,5 <input type="checkbox"/>	AVERAGE: 1 <input type="checkbox"/>	LIMITED: 0,5 <input type="checkbox"/>
CREATIVITY	Outstanding flow of ideas and events <input type="checkbox"/>	Good flow of ideas and events <input checked="" type="checkbox"/>	Average flow of ideas and events <input type="checkbox"/>	Poor flow of ideas and events <input type="checkbox"/>
	EXCELLENT: 2 <input type="checkbox"/>	GOOD: 1,5 <input checked="" type="checkbox"/>	AVERAGE: 1 <input type="checkbox"/>	LIMITED: 0,5 <input type="checkbox"/>
SCIENTIFIC SUSTAINABILITY	Reasonable, specific and supportable opinion or thesis statement <input type="checkbox"/>	Minor errors when supporting the thesis statement <input checked="" type="checkbox"/>	Some errors when supporting the thesis statement <input type="checkbox"/>	Lots of errors when supporting the thesis statement <input type="checkbox"/>
	EXCELLENT: 2 <input type="checkbox"/>	GOOD: 1,5 <input checked="" type="checkbox"/>	AVERAGE: 1 <input type="checkbox"/>	LIMITED: 0,5 <input type="checkbox"/>
TOTAL/AVERAGE	9 - 10: EXCELLENT 7 - 8,9: GOOD 5 - 6,9: AVERAGE 0 - 4,9: LIMITED		TOTAL 9	



**UNIVERSIDAD POLITÉCNICA ESTATAL DEL
CARCHI FOREIGN AND NATIVE LANGUAGE
CENTER**

Informe sobre el Abstract de Artículo Científico o Investigación.

Autor: Carlos Fernando Montenegro Argoti

Fecha de recepción del abstract: 15 de septiembre de 2021

Fecha de entrega del informe: 15 de septiembre de 2021

El presente informe validará la traducción del idioma español al inglés si alcanza un porcentaje de: 9 – 10 Excelente.

Si la traducción no está dentro de los parámetros de 9 – 10, el autor deberá realizar las observaciones presentadas en el ABSTRACT, para su posterior presentación y aprobación.

Observaciones:

Después de realizar la revisión del presente abstract, éste presenta una apropiada traducción sobre el tema planteado en el idioma Inglés. Según los rubrics de evaluación de la traducción en Inglés, ésta alcanza un valor de 9, por lo cual se valida dicho trabajo.

Atentamente



Firmado electrónicamente por:
EDISON BOANERGES
PENAFIEL ARCOS

Ing. Edison Peñafiel Arcos MSc
Coordinador del CIDEN

Anexo 4: Certificado de conformidad



**Universidad Politécnica
Estatal del Carchi**
Educamos para transformar el mundo



Msc. Diego Caicedo
Responsable de la estación meteorológica
UNIVERSIDAD POLITÉCNICA ESTATAL DEL CARCHI

ASUNTO: INFORME DE CONFORMIDAD

De mi consideración,

*Saludos cordiales, por medio del presente, me permito informar, y **CERTIFICAR** que se realizó un modelo para pronóstico de precipitaciones pluviales enfocado en los datos históricos obtenidos por la estación meteorológica, del proyecto de culminación de carrera cuyo tema es: “**Minería de datos para series temporales y su aplicación en las precipitaciones pluviales en la Finca Experimental San Francisco de Huaca**”.*

Particular que me permito poner en su conocimiento, para los fines académicos correspondientes.

Atentamente,

DIEGO MARCELO CAICEDO ROSERO  Firmado digitalmente por
DIEGO MARCELO CAICEDO
ROSERO

Ing. Diego Caicedo. MSc.
DOCENTE UPEC



UNIVERSIDAD POLITÉCNICA ESTATAL DEL CARCHI
FACULTAD DE INDUSTRIAS AGROPECUARIAS
Y CIENCIAS AMBIENTALES
CARRERA DE INGENIERIA EN INFORMÁTICA



Entrevista MSc. Diego Caicedo responsable de la estación meteorológica en la finca experimental San Francisco del cantón San Pedro de Huaca

Fecha:

1. ¿En qué forma aporta la estación meteorológica al sector agrícola?
2. ¿Existe otra estación meteorológica en la zona por parte de otras instituciones?
3. ¿Cómo se obtenían las variables meteorológicas antes de la existencia de la estación meteorológica?
4. ¿Cuál es el propósito de implementar una estación meteorológica?
5. ¿De qué tipo es la estación meteorológica y cuál es su función?
6. ¿Qué variables meteorológicas almacena la estación meteorológica y cuál es su valor de medición?
7. ¿Cuál es el historial de datos de la estación meteorológica y que procesos se realiza con la información o datos que almacenan?
8. ¿Cuáles son las ventajas de contar con una estación meteorológica en la finca experimental San Francisco de Huaca?



UNIVERSIDAD POLITÉCNICA ESTATAL DEL CARCHI
FACULTAD DE INDUSTRIAS AGROPECUARIAS Y
CIENCIAS AMBIENTALES
CARRERA DE INGENIERIA EN INFORMÁTICA



ENCUESTA AGRICULTORES

Marque una o las respuestas que usted crea que son convenientes

1. ¿La actividad agrícola constituye su principal fuente de economía?

SI NO

2. ¿Ha tenido pérdidas en sus cultivos por eventos meteorológicos?

SI NO

3. Cuáles de estos tipos de precipitaciones pluviales se dan en su zona

Sequias

Heladas

Granizadas

Inundaciones

Otros: _____

4. ¿Las precipitaciones pluviales afectan a la calidad de los cultivos?

SI NO

5. ¿Las precipitaciones pluviales afectan al precio de los productos agrícolas?

SI NO

6. ¿Qué métodos utiliza para la predicción del cambio climatológico?

Internet

Almanaques

Ancestrales

Noticieros

Otros: _____

7. ¿Qué datos puede medir una estación meteorológica?

Viento

Temperatura

Humedad

Presión atmosférica

Atmosfera – sol

Otros: _____

8. ¿Cuenta con una estación meteorológica en su sector agrícola?

SI NO

9. ¿En qué le beneficiaría a usted tener los datos meteorológicos?

MANUAL DE USUARIO

FORECAST DRIVE

APLICATIVO INFORMÁTICO PARA LA
PREDICCIÓN DE PRECIPITACIONES
PLUVIALES

TULCÁN – ECUADOR

2021

Contenido

Introducción.....	90
Selección de datos.....	90
Preparación de los datos.....	90
Acceso al sitio web.....	91
Interfaz principal de ForecastDrive.....	91
Rango de fecha.....	93
Meses de pronóstico.....	93
Exploración.....	93
Pronostico.....	95
Dashboard Tableau Public.....	96
Indicadores de colores.....	96
Promedios Precipitaciones	97
Valores Ajustados.....	98
Pronostico.....	98
Descarga.....	99

Introducción

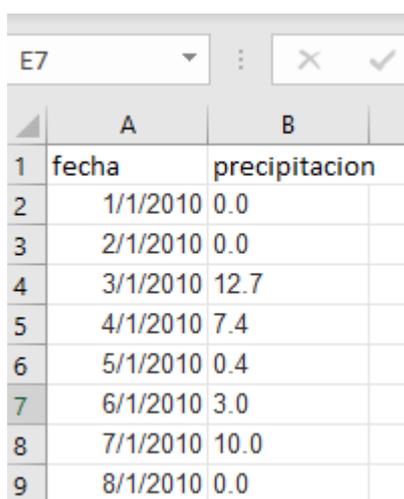
El presente manual del usuario tiene como finalidad explicar de manera detallada, verbal y grafica el uso del aplicativo informático, como también la preparación de los datos. Para que la experiencia de uso sea de mayor provecho y satisfactoria.

Selección de datos

La aplicación ForecastDrive hace uso de datos históricos de la variable precipitación (rainfall) con la unidad de medida milímetros por metro cuadrado (mm), esta variable debe ir acompañada de una fecha donde fue hecho el registro de ese dato en el formato dd/mm/yyyy.

01/01/2010 0.7

Para que la aplicación haga el reconocimiento y la lectura correcta de estos datos se debe colocar las variables en el orden *fecha, precipitacion*. En una hoja de Excel almacenamos cada variable en una columna en el orden y con el nombre especificado.



	A	B
1	fecha	precipitacion
2	1/1/2010	0.0
3	2/1/2010	0.0
4	3/1/2010	12.7
5	4/1/2010	7.4
6	5/1/2010	0.4
7	6/1/2010	3.0
8	7/1/2010	10.0
9	8/1/2010	0.0

Figura 38. Selección de datos

En el caso de que la medida de precipitación haga uso de decimales se debe utilizar el punto (.)

NOTA: Mientras con más datos históricos se cuente mayor será la efectividad de la predicción

Preparación de los datos

Una vez tengamos los datos seleccionados, el siguiente paso es transformar los datos de la hoja Excel a un archivo CSV (Archivo separado por comas).

Para ello en el documento de Excel nos dirigimos al botón de ARCHIVO ubicado en la parte superior izquierda, luego damos clic en GUARDAR COMO

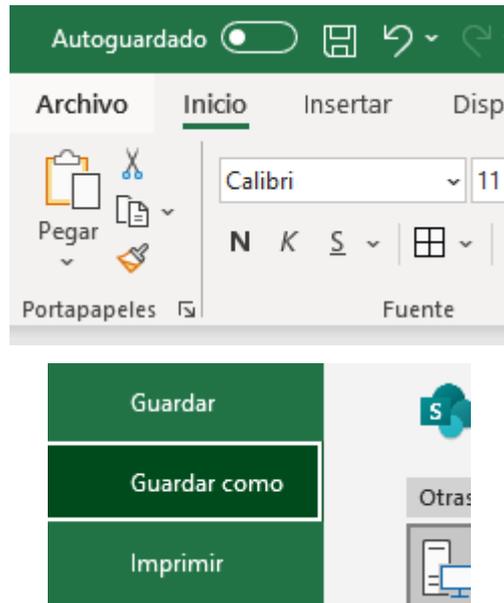


Figura 39. Guarda como CSV

En este menu indicamos el nombre con que el vamos a guardar nuestro archivo, en el menu desplegable que se encuentra seguido de este apartado buscamos el formato “Archivo separado por comas” (CSV), y guardamos nuestro nuevo archivo.



Figura 40. Extensión delimitado por comas .CSV

De esta manera ya tenemos listos nuestros datos para subir a la aplicación

Acceso al sitio web

Para acceder al aplicativo web ingresamos a la url:

<https://forecastdrive.shinyapps.io/Precipitation/>

desde cualquier navegador de nuestra preferencia. ForecastDrive es un aplicativo responsive por lo tanto es compatible con navegadores de dispositivos móviles.

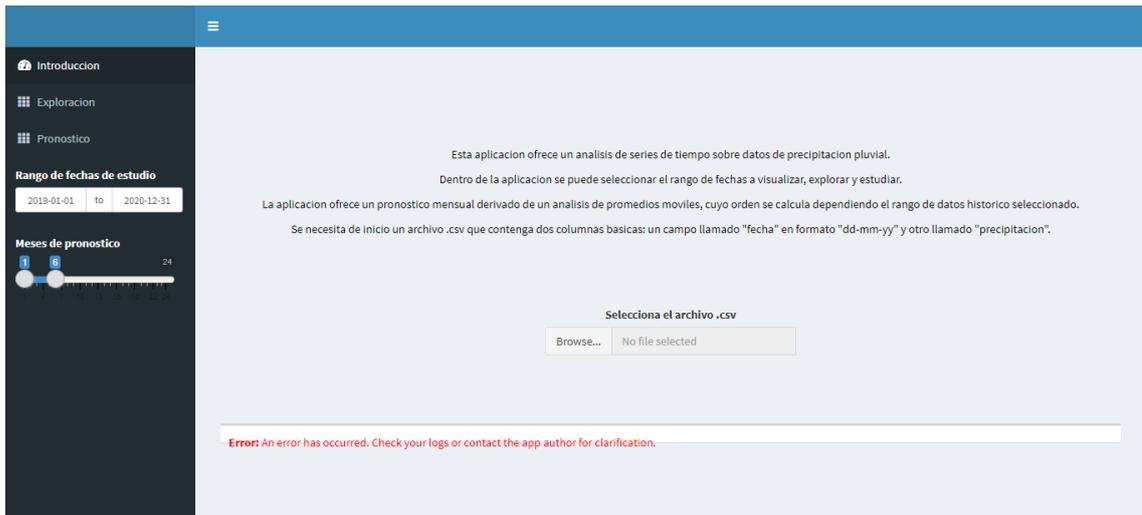


Figura 41. Interfaz principal

Interfaz principal de ForecastDrive

Una vez hayamos accedido a la página principal del aplicativo informático damos clic en el botón BROWSE y seleccionamos el archivo CSV anteriormente preparado.

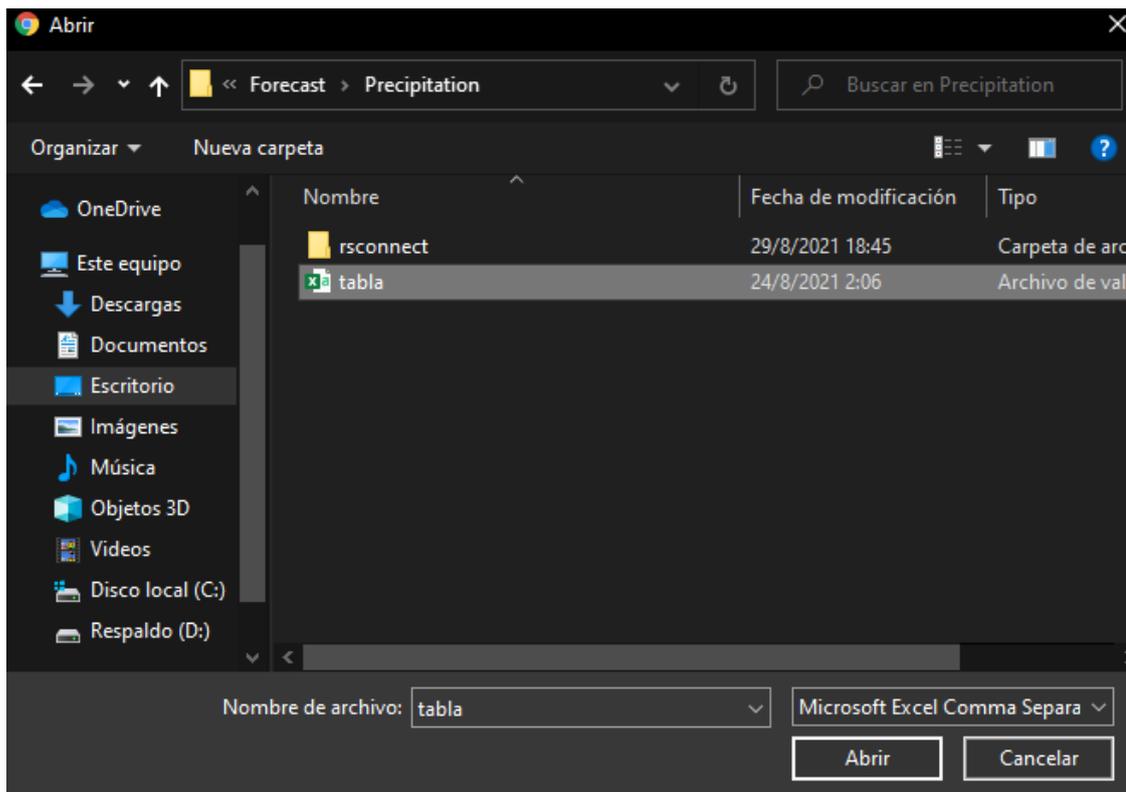


Figura 42. Carga de datos

Cargamos el archivo y observamos como los datos fueron cargados satisfactoriamente mediante una tabla, además en este apartado podemos ver en la parte inferior el número de instancias que fueron seleccionadas.

Show entries Search:

	fecha	precipitacion
1	2018-06-03	0
2	2018-06-03	0
3	2018-06-03	0
4	2018-06-03	0
5	2018-06-03	0
6	2018-06-03	0
7	2018-06-03	0
8	2018-06-03	0
9	2018-06-03	0
10	2018-06-03	0

Showing 1 to 10 of 27,698 entries Previous 2 3 4 5 ... 2770 Next

Figura 43. Total de instancias

En la tabla de datos podemos seleccionar los datos que podemos observar por pantalla dando clic en la ventana desplegable, esta opción nos permite ver datos desde 10 hasta 100 datos por página. También cuenta con un cajo SEARCH donde podemos filtrar datos por fecha.

Show entries Search:

	fecha	precip
5747	2019-01-31	
5748	2019-01-31	
5749	2019-01-31	

Figura 44. Función botón de búsqueda

Rango de fecha de estudio y meses de pronostico

Rango de fecha

Para comenzar con la exploración de los datos es importante determinar el rango de fecha de los datos con los que se van a trabajar.

En el apartado izquierdo encontramos un menú donde existe la opción de “RANGO DE FECHA DE ESTUDIO”, donde debemos escribir la fecha de inicio y la fecha de fin de los datos que fueron ingresados.

Rango de fechas de estudio

2019-01-01

to

2020-12-31

Figura 45. Rango de fechas

Meses de pronóstico

Seguido del apartado de rango de fecha, encontramos una barra deslizante la cual indica los meses de pronóstico que queremos tener. Esta barra va desde los 0 meses hasta los 24.



Figura 46. Meses de pronóstico

Exploración

Una vez ingresado los rangos de fecha y los meses de pronóstico, nos dirigimos al menú de EXPLORACIÓN en donde encontramos el grafico de precipitaciones de la primera serie de tiempo.

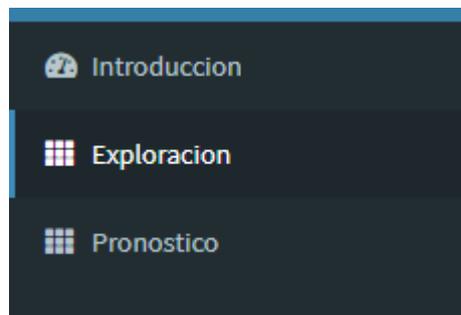


Figura 47. Menú exploración

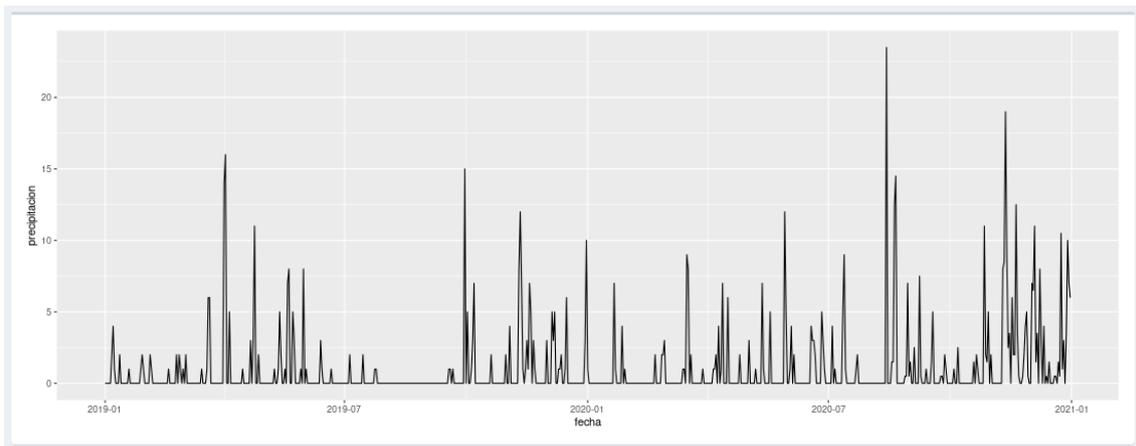


Figura 48. Histograma precipitación

En la parte inferior de este apartado encontramos dos tablas las cuales corresponden a los datos tomados por día (primera tabla) y al promedio de datos por cada día (segunda tabla)

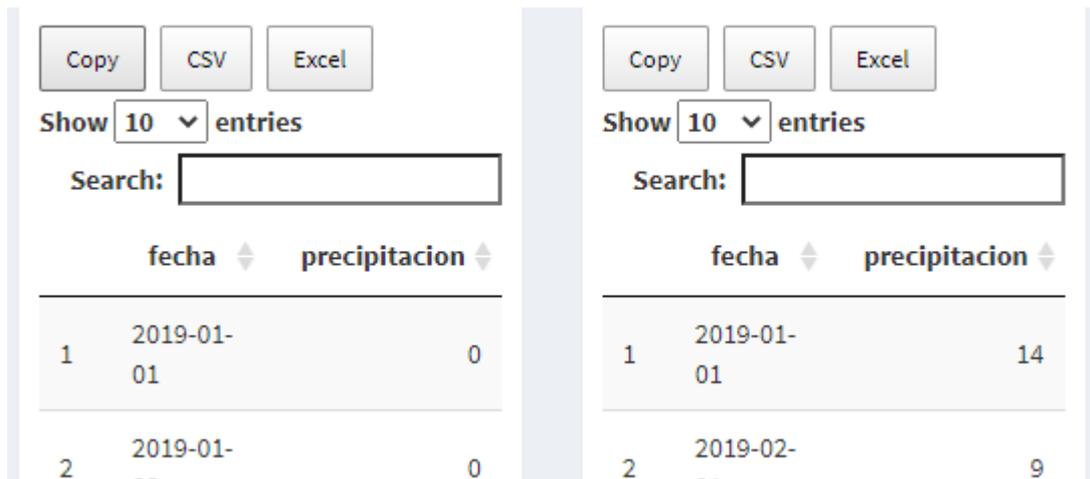


Figura 49. Tablas de exploración

Estas tablas tienen las mismas opciones de búsqueda y vistas como la primera tabla, adicional esta tabla cuenta con tres botones que son: “COPY” el cual copia los datos que se encuentran en la tabla, “CSV” descarga la tabla en un archivo separado por comas y el botón “EXCEL” descarga la tabla en una hoja de cálculo de Excel.

En la parte izquierda observamos un gráfico de tendencia de las precipitaciones pluviales.

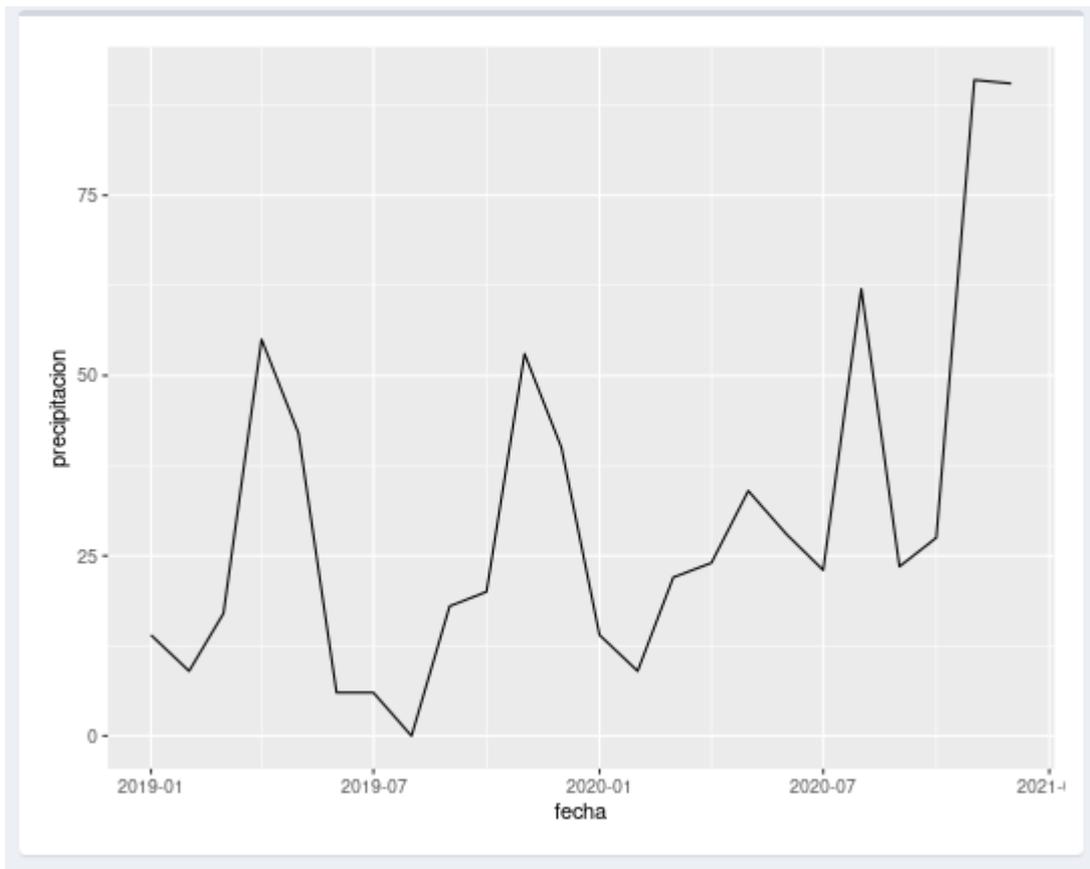


Figura 50. Tendencia precipitación

Pronósticos

Finalmente tenemos el menú de pronóstico, en este apartado se encuentra el gráfico de donde se determina la predicción (color azul) y los valores ajustados (color morado).

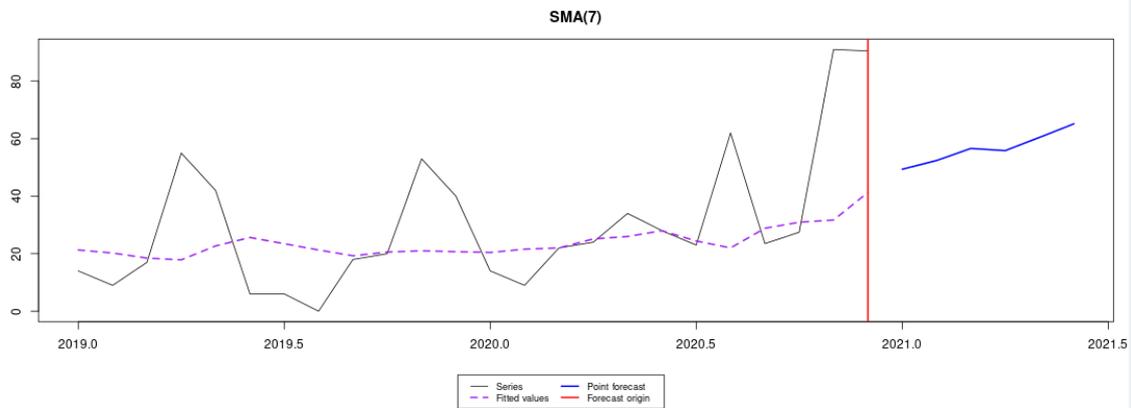


Figura 51. Pronóstico SMA

En la parte superior encontramos dos tablas: Valores ajustados (primera tabla) y la tabla final con los resultados de los pronósticos por los meses seleccionados (segunda tabla)

Estas tablas cuentan con los mismos botones y funcionalidades descritas en los anteriores pasos.

Las tablas muestran los valores ajustados y los pronósticos para los meses de enero a junio de 2021. Ambas tablas tienen botones para Copiar, CSV, Excel y un menú desplegable para mostrar 10 entradas. Hay un campo de búsqueda en la parte superior de cada tabla.

	fecha	fited_values
1	2019-01-01	21.2857142857143
2	2019-02-01	20.2448979591837
3	2019-03-01	18.4897959183673
4	2019-04-01	17.8775510204082
5	2019-05-01	22.6938775510204
6	2019-06-01	25.6530612244898

	fecha	pronostico
1	ene 2021	49.3571428571429
2	feb 2021	52.4081632653061
3	mar 2021	56.6093294460641
4	abr 2021	55.8392336526447
5	may 2021	60.4591241744511
6	jun 2021	65.167570485087

Dashboard Tableau Public

Para acceder al dashboard con los resultados obtenidos accedemos al siguiente link:

<https://public.tableau.com/app/profile/carlos6372/viz/ForecastDrive/Dashboard1>

Vista general del dashboard desde la página de public tableau

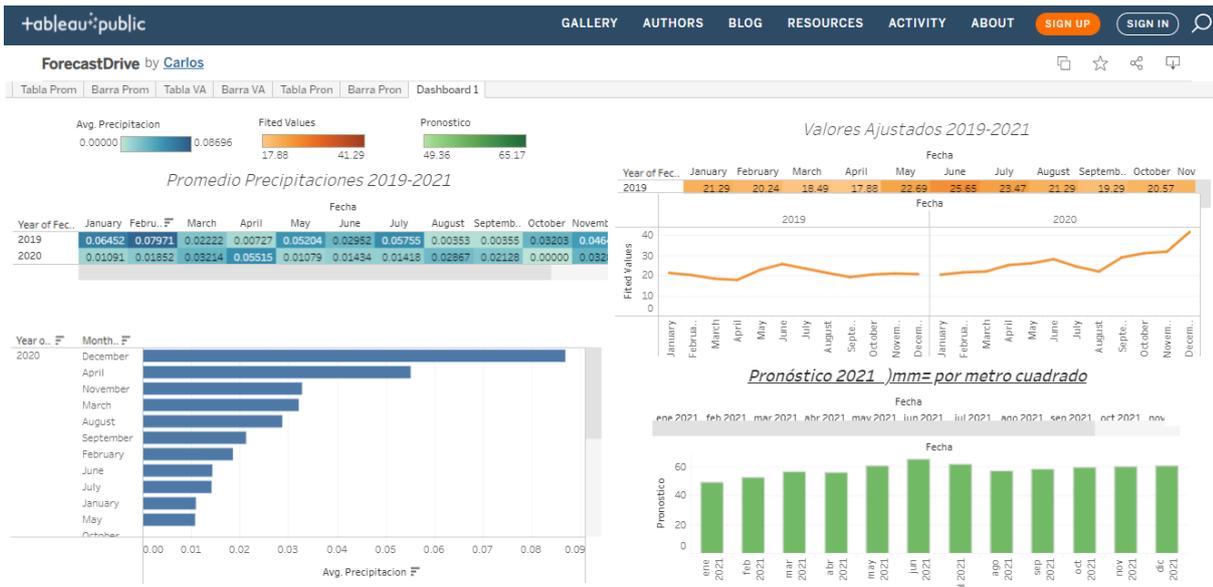


Figura 52. Página principal del dashboard

Indicadores de color



Figura 53. Indicadores de colores

En la parte superior del dashboard encontramos los indicadores de colores. El color más claro indica el menor valor mientras que el color más oscuro indica el valor más alto.

El dashboard cuenta con tres tablas de análisis que son:

- Color Azul – Promedio de Precipitaciones 2019-2020
- Color Naranja – Valores ajustados 2019-2020
- Color Verde – Pronostico Predicciones 2021

Promedios Precipitaciones 2019-2020

Promedio Precipitaciones 2019-2021

Year	January	February	March	April	May	June	July	August	September	October	November
2019	0.06452	0.07971	0.02222	0.00727	0.05204	0.02952	0.05755	0.00353	0.00355	0.03203	0.046
2020	0.01091	0.01852	0.03214	0.05515	0.01079	0.01434	0.01418	0.02867	0.02128	0.00000	0.032

Figura 54. Promedio de precipitaciones

En la tabla de promedio de precipitaciones encontramos los años con los que se trabajó en filas y los meses en columnas obteniendo el promedio de precipitación para el mes correspondiente.

Promedio Precipitaciones 2019-2021

Year of Fec..	1	2	3	4	5	6	7	8	9	10	11
2020	0.0435	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0833
2019	0.0000	0.0000	0.0417	0.6667	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Figura 55. Filtro en tabla de promedios

Cuenta con filtros que permiten visualizar la tabla en trimestres, meses, o hasta días dependiendo la necesidad del usuario.

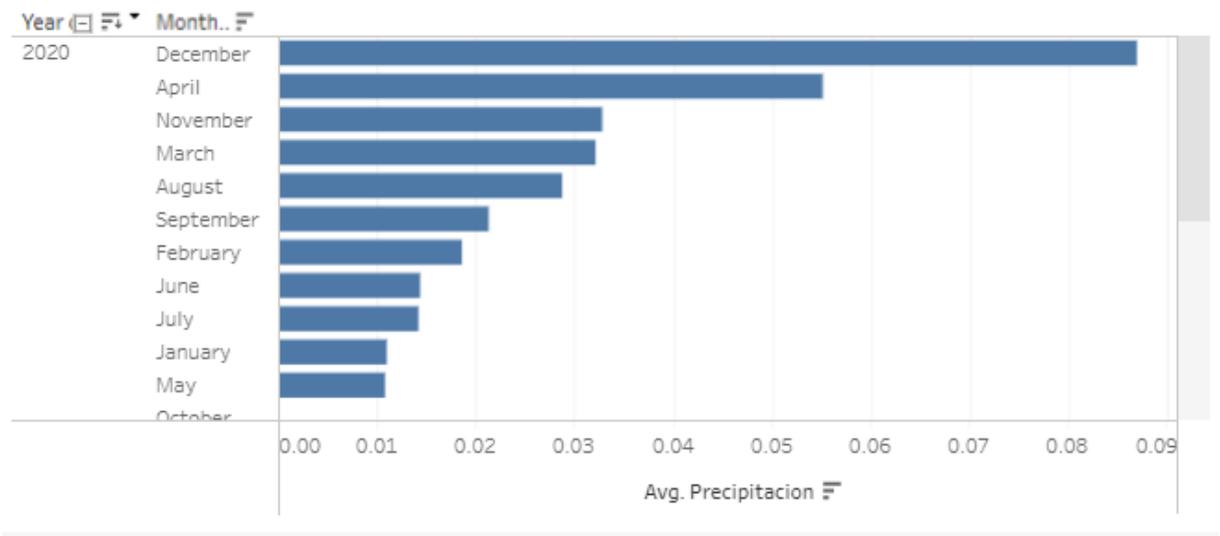
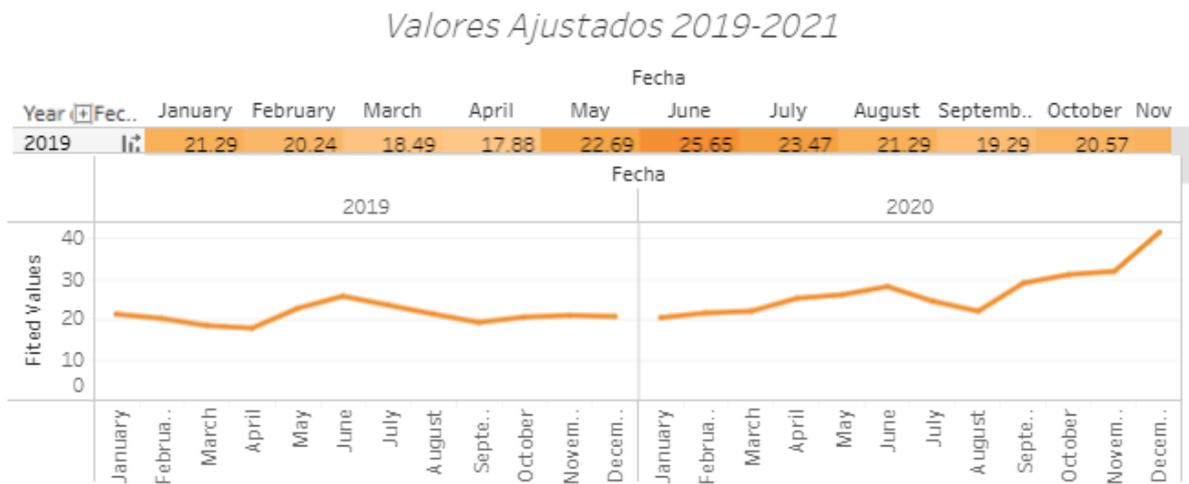


Figura 56. Filtro de orden en tabla de precipitaciones

Los promedios de precipitaciones también se pueden ver mediante un gráfico de barras horizontales, el cual separa los años con sus correspondientes meses, dando clic en el icono  permitirá ordenar de forma ascendente, descendente y ordenada cronológicamente según el usuario necesite.

Valores ajustados



La tabla y el grafico naranja indican los valores ajustados que sirven para estimar de punto de la respuesta media para los valores dados de las predicciones.

Pronostico 2021



Figura 57. Tabla de pronósticos

La tabla y el gráfico verde indican los valores de los pronósticos para el año 2021 en meses. La unidad de medida es milímetro por metro cuadrado de superficie en terreno, se encuentra un promedio mensual debido a los anteriores datos.

NOTA: Cabe mencionar que las funciones de filtros se encuentran disponibles para los tres gráficos con los que cuenta el dashboard.

Descarga



Figura 58. Botones de descarga, compartir y copiar

Tableau public permite descargar en diferentes formatos el dashboard creado, dando clic en el último icono de descarga, además de permitir copiar, guardar como favoritos y compartir.

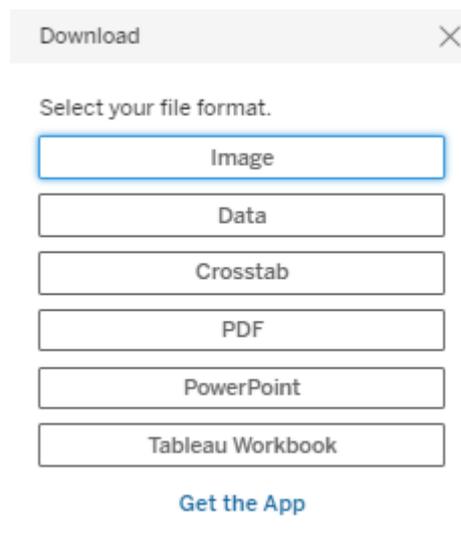


Figura 59. Formatos de descarga

Estos son algunos de los formatos en que tableau public nos permite descargar el dashboard.