



UTMACH

UNIDAD ACADÉMICA DE INGENIERÍA CIVIL

CARRERA DE INGENIERÍA DE SISTEMAS

IMPLEMENTACIÓN DE UN MÓDULO DE ANÁLISIS ESTADÍSTICO Y
PREDICTIVO PARA AGRICULTURA UTILIZANDO BIGDATA Y
MACHINE LEARNING, INTEGRADO AL SISTEMA IOTMACH

HERRERA DÍAZ CARLOS ALBERTO

MACHALA
2016



UTMACH

UNIDAD ACADÉMICA DE INGENIERÍA CIVIL

CARRERA DE INGENIERÍA DE SISTEMAS

IMPLEMENTACIÓN DE UN MÓDULO DE ANÁLISIS
ESTADÍSTICO Y PREDICTIVO PARA AGRICULTURA
UTILIZANDO BIGDATA Y MACHINE LEARNING, INTEGRADO
AL SISTEMA IOTMACH

HERRERA DÍAZ CARLOS ALBERTO

MACHALA
2016



UTMACH

UNIDAD ACADÉMICA DE INGENIERÍA CIVIL

CARRERA DE INGENIERÍA DE SISTEMAS

TRABAJO DE TITULACIÓN
PROPUESTAS TECNOLÓGICAS

IMPLEMENTACIÓN DE UN MÓDULO DE ANÁLISIS ESTADÍSTICO Y PREDICTIVO
PARA AGRICULTURA UTILIZANDO BIGDATA Y MACHINE LEARNING,
INTEGRADO AL SISTEMA IOTMACH

HERRERA DÍAZ CARLOS ALBERTO
INGENIERO DE SISTEMAS

MAZÓN OLIVO BERTHA EUGENIA

Machala, 18 de octubre de 2016

MACHALA
2016

Nota de aceptación:

Quienes suscriben MAZÓN OLIVO BERTHA EUGENIA, ZEA ORDOÑEZ MARIUXI PAOLA, REDROVAN CASTILLO FAUSTO FABIAN y JUMBO CASTILLO FREDDY ANIBAL, en nuestra condición de evaluadores del trabajo de titulación denominado IMPLEMENTACIÓN DE UN MÓDULO DE ANÁLISIS ESTADÍSTICO Y PREDICTIVO PARA AGRICULTURA UTILIZANDO BIGDATA Y MACHINE LEARNING, INTEGRADO AL SISTEMA IOTMACH, hacemos constar que luego de haber revisado el manuscrito del precitado trabajo, consideramos que reúne las condiciones académicas para continuar con la fase de evaluación correspondiente.




MAZÓN OLIVO BERTHA EUGENIA
0603100512
TUTOR



ZEA ORDOÑEZ MARIUXI PAOLA
0702801598
ESPECIALISTA 1



REDROVAN CASTILLO FAUSTO FABIAN
0702739228
ESPECIALISTA 2



JUMBO CASTILLO FREDDY ANIBAL
0704167949
ESPECIALISTA 3

Machala, 18 de octubre de 2016

Urkund Analysis Result

Analysed Document: HERRERA DIAZ CARLOS ALBERTO.docx (D21637207)
Submitted: 2016-09-07 05:12:00
Submitted By: caherrerad_est@utmachala.edu.ec
Significance: 2 %

Sources included in the report:

HONORES CUN JENNIFER LISSETH - SANCHEZ CHUICO DANILO ANTONIO.docx (D21540099)
Informe Caso de Estudio Jaime Michilena.pdf (D19319297)
Informe Caso de Estudio Jaime Michilena.pdf (D19346683)
1472836549_Versión-final-Informe-Proyecto-Integrador-FERNANDEZ-Y-RAMIREZ.docx (D21590250)
1447822511_907__Resumen_WSN.pdf (D16256057)
1447822416_353__grupo%252BB.pdf (D16256054)
TESIS_OLALE_FINAL 5.docx (D14920196)
TESIS_OLALE_FINAL 9.docx (D14916953)
<http://www.scielo.org.co/pdf/itec/v12n2/v12n2a07.pdf>
<https://www.computer.org/csdl/proceedings/pci/2008/3323/00/3323a100-reference.bib>
<https://www.computer.org/csdl/proceedings/pci/2008/3323/00/3323zTOC.pdf>
http://www.dataversity.net/not-your-type-big-data-matchmaker-on-five-data-types-you-need-to-explore-today/?cm_mc_uid=07202058171914656643101&cm_mc_sid_50200000=1466444811
<http://www.analyticsvidhya.com/blog/2014/03/sas-vs-vs-python-tool-learn/>

Instances where selected sources appear:

CLÁUSULA DE CESIÓN DE DERECHO DE PUBLICACIÓN EN EL REPOSITORIO DIGITAL INSTITUCIONAL

El que suscribe, HERRERA DÍAZ CARLOS ALBERTO, en calidad de autor del siguiente trabajo escrito titulado IMPLEMENTACIÓN DE UN MÓDULO DE ANÁLISIS ESTADÍSTICO Y PREDICTIVO PARA AGRICULTURA UTILIZANDO BIGDATA Y MACHINE LEARNING, INTEGRADO AL SISTEMA IOTMACH, otorga a la Universidad Técnica de Machala, de forma gratuita y no exclusiva, los derechos de reproducción, distribución y comunicación pública de la obra, que constituye un trabajo de autoría propia, sobre la cual tiene potestad para otorgar los derechos contenidos en esta licencia.

El autor declara que el contenido que se publicará es de carácter académico y se enmarca en las disposiciones definidas por la Universidad Técnica de Machala.

Se autoriza a transformar la obra, únicamente cuando sea necesario, y a realizar las adaptaciones pertinentes para permitir su preservación, distribución y publicación en el Repositorio Digital Institucional de la Universidad Técnica de Machala.

El autor como garante de la autoría de la obra y en relación a la misma, declara que la universidad se encuentra libre de todo tipo de responsabilidad sobre el contenido de la obra y que él asume la responsabilidad frente a cualquier reclamo o demanda por parte de terceros de manera exclusiva.

Aceptando esta licencia, se cede a la Universidad Técnica de Machala el derecho exclusivo de archivar, reproducir, convertir, comunicar y/o distribuir la obra mundialmente en formato electrónico y digital a través de su Repositorio Digital Institucional, siempre y cuando no se lo haga para obtener beneficio económico.

Machala, 18 de octubre de 2016



HERRERA DÍAZ CARLOS ALBERTO
0706582756

DEDICATORIA

Primeramente a Dios por brindarme salud y darme los conocimientos suficientes para lograr cumplir las metas que me he propuesto a lo largo de mi carrera.

A mis padres por haberme dado la vida y por estar conmigo siempre, en especial en esos malos momentos en los que todo ser humano puede pasar; resaltar su apoyo incondicional, los ánimos que día a día me supieron dar y llenándome de consejos para superarme con el pasar del tiempo.

A mis hermanos por todo su apoyo, su comprensión, por motivarme en cada una de las cosas que he realizado.

A mi tutora, Ing. Bertha Mazón Olivo, que con sus orientaciones, amplios conocimientos y la confianza que puso en mí, supe llevar a cabo este proceso que es uno de los más importantes en esta etapa de mi vida.

A la Universidad Técnica de Machala, que por medio de la Unidad Académica de Ingeniería Civil y la carrera de Ingeniería de Sistemas, pude adquirir los conocimientos que me permitirán desenvolverme en el campo laboral.

Carlos Herrera Díaz

AGRADECIMIENTO

Agradezco a Dios por iluminarme en el transcurso de mi vida y guiarme en todas las cosas que hago y las decisiones que he tomado en mi convivir diario.

A mis padres, las personas más importantes que tengo, porque supieron inculcarme el buen ejemplo, respeto, humildad y demás valores que me han servido en cada etapa de mi vida. Por apoyarme en cada decisión que tomo y por estar siempre a mi lado; gracias por eso y por mucho más.

A la Universidad Técnica de Machala por abrirme las puertas, a mi tutora Ing. Bertha Mazón Olivo, que gracias a sus conocimientos, experiencias supo guiarme en cada momento, por tenerme paciencia y estar pendiente en cada paso de esta etapa académica.

Carlos Herrera Díaz

RESUMEN

El contexto que abarca la minería de datos y aprendizaje de máquina (Data Mining y Machine Learning en inglés respectivamente) tienen como principal objetivo hacer que una máquina de procesamiento con buenas capacidades que demanda esta era tecnológica, ejecute acciones para resolver problemas, empleando técnicas que a posterior podrían ser utilizadas para la toma de decisiones relacionadas a las actividades de una organización. Tanto la minería de datos y el aprendizaje de máquina necesitan información en gran cantidad, aquí es donde entra otro vocablo: Big data, el cual hace referencia a la abundancia de datos obtenidos de distintas fuentes que posee una empresa sea cual sea su sector de comercio; aplicando dichas técnicas, se pueden descubrir patrones, tendencias, similitudes entre datos que existen pero no están visibles de manera clara. Estos descubrimientos van a permitir realizar predicciones, clasificaciones, segmentación o agrupación en el marco de un conjunto de datos, que satisfagan necesidades o problemas que a diario surgen dentro de un negocio.

En el campo de la agricultura, se realizan muchos procesos que hoy en día son automatizados, en los que permiten entre varias cosas, regular, controlar y administrar de una manera aceptable los recursos que en un cultivo se requiere para una buena producción, mediante dispositivos como sensores, actuadores, que, comunicados entre sí deben operar de cierta manera para optimizar entre varios aspectos: tiempo, dinero y mano de obra. Esta manera de comunicación entre los objetos se denomina Internet de las cosas (IOT-Internet of Things en inglés), el cual ha tenido auge últimamente en la agricultura. Aprovechando la automatización y las tecnologías modernas que actualmente existen, se puede realizar análisis a los datos de un sistema de información o de diversas fuentes de datos, mediante aplicaciones que tengan como finalidad facilitar la manipulación de información dentro del negocio, aplicando técnicas descriptivas y predictivas.

La información que puede generarse en todo el tiempo de vida de un cultivo, es abundante, por lo que debe ser analizada y sometida a un análisis empleando algoritmos para solventar necesidades, tales como prevenir enfermedades, decidir cuánto regar, controlar la humedad y temperatura, clasificar plantas, agrupar especies de cultivos, entre otras. La implementación de algoritmos o técnicas se pueden plasmar en distintos software dedicados exclusivamente al aprendizaje de máquina, en la actualidad el más

popular y el que destaca es el lenguaje R, el mismo que se utilizó para el desarrollo de este proyecto.

Este trabajo tuvo como finalidad , diseñar e implementar una aplicación que permita el análisis de la información que se genera dentro del sistema IOTMACH y de datasets externos, de tal manera que se pueda visualizar, manipular, interpretar y aprender de los datos, teniendo en cuenta que la extracción del conocimiento que guardan los datos, pueden ser útiles para que el agricultor o cualquier persona encargada de la administración de los cultivos tome decisiones acertadas respecto a las situaciones que se presenten en un futuro. Con la herramienta desarrollada, se puede constituir un apoyo en el área de agricultura como también en distintas áreas o sectores económicos.

PALABRAS CALVE: agricultura, aprendizaje de máquina, big data, internet de las cosas, minería de datos.

ABSTRACT

The context includes data mining and machine learning are aimed to make a processing machine with good skills demanded by this technological age, run actions to solve problems, using techniques that to later could be used for decision making related to the activities of an organization. Both data mining and machine learning need information in large amount, this is where another word comes: Big data, which refers to the abundance of data obtained from different sources that owns a company whatever its sector of trade; applying these techniques, you can discover patterns, trends, similarities between data that exist but are not clearly visible. These findings will allow make predictions, classifications, segmentation or clustering within a data set that meet needs or problems that arise within daily business.

In the field of agriculture, many processes today are automated, which allow, among many things, regulate, control and manage in a manner acceptable resources in plot is required for good production through devices are made as sensors, actuators, which communicate with each other must operate in a way to optimize between several aspects: time, money and labor. This way of communication between objects is called Internet of Things, which has been booming lately in agriculture. Taking advantage of automation and modern technologies that exist today, you can perform analysis data of an information system or various data sources, through applications which are designed to facilitate handling of information within the business, using descriptive and predictive techniques.

The information that can be generated throughout the lifetime of a crop is abundant, so must be analyzed and subjected to analysis using algorithms to address needs such as disease prevention, decide how much to water, control moisture and temperature, classify plants, grouped plants species, among others. The implementation of algorithms or techniques can translate into different software exclusively dedicated to machine learning, currently the most popular and highlighting is the R language, the same used for the development of this project.

This paper aims to design and implement an application that allows the analysis of the information generated within the IOTMACH system and external datasets, so you can visualize, manipulate, interpret and learn from the data, taking into that the extraction of

knowledge stored data may be useful for the farmer or anyone responsible for administering crop make wise decisions about situations that arise in the future. With the tool developed, it can be a support in the area of agriculture as well as in different areas or economic sectors.

KEYWORDS: agriculture, bigdata, data mining, internet of things, machine learning.

CONTENIDO

	Pág.
DEDICATORIA	2
AGRADECIMIENTO	3
RESUMEN	4
ABSTRACT	6
CONTENIDO	8
LISTA DE CUADROS	10
LISTA DE FIGURAS	11
LISTA DE ANEXOS	13
GLOSARIO	14
INTRODUCCIÓN	19
1. DIAGNÓSTICO DE NECESIDADES Y REQUERIMIENTOS	22
1.1 Ámbito de Aplicación: descripción del contexto y hechos de interés	22
1.2 Establecimiento de requerimientos	24
1.3 Justificación del requerimiento a satisfacer	25
2. DESARROLLO DEL PROTOTIPO	27
2.1 Definición del prototipo tecnológico	27
2.2 Fundamentación teórica del prototipo	28
2.2.1 Machine Learning	28
2.2.2 Big Data	30
2.2.3 Data Mining	32
2.2.4 Técnicas predictivas	36
2.2.5 Técnicas descriptivas	38
2.2.6 Internet de las cosas	39
2.2.7 Agricultura de precisión	42
2.2.8 Red de sensores inalámbricos	43
2.2.9 Proyecto IOTMACH	44
2.3 Objetivos del prototipo	44
2.3.1 Objetivo general	44
2.3.2 Objetivos específicos	45
2.4 Requisitos, análisis y diseño	45
2.4.1 Requisitos funcionales	45
2.4.2 Requisitos no funcionales	48
2.4.3 Submódulos implementados en la interfaz web	49
2.4.4 Suposiciones y restricciones	50
2.4.5 Análisis y diseño	51
2.5 Desarrollo e implementación del prototipo	60

2.5.1 Lenguaje de programación R	60
2.5.2 Herramientas y librerías	62
2.6 Ejecución del prototipo	75
2.6.1 Interfaces de la aplicación	75
3.EVALUACIÓN DEL PROTOTIPO	81
3.1. Plan de evaluación	81
3.1.1 Pruebas unitarias	81
3.1.2 Pruebas de integración	82
3.1.3 Pruebas de usabilidad	83
3.1.4 Pruebas de rendimiento	84
3.1.5 Pruebas de regresión	84
3.2. Resultados de la evaluación	85
CONCLUSIONES	91
RECOMENDACIONES	93
BIBLIOGRAFÍA	94
ÍNDICE	949
ANEXOS	100

LISTA DE CUADROS

	Pág.
Cuadro 1. Áreas de aplicación en el análisis predictivo	23
Cuadro 2. Desafíos de Machine Learning	29
Cuadro 3. Técnicas de Minería de datos	34
Cuadro 4. Comparación de las metodologías	35
Cuadro 5. Diferencia entre aprendizaje supervisado y no supervisado	38
Cuadro 6. Requisito funcional REQ_001	45
Cuadro 7. Requisito funcional REQ_002	46
Cuadro 8. Requisito funcional REQ_003	46
Cuadro 9. Requisito funcional REQ_004	47
Cuadro 10. Requisito funcional REQ_005	47
Cuadro 11. Requisito funcional REQ_006	48
Cuadro 12. Caso de uso CU_001	53
Cuadro 13. Caso de uso CU_002	54
Cuadro 14. Caso de uso CU_003	54
Cuadro 15. Caso de uso CU_004	55
Cuadro 16. Caso de uso CU_005	56
Cuadro 17. Comparación entre R y Python	62
Cuadro 18. Elementos utilizados en la aplicación Shiny	67
Cuadro 19. Descripción de pruebas unitarias	81
Cuadro 20. Descripción de pruebas de integración	82
Cuadro 21. Descripción de pruebas de usabilidad	83
Cuadro 22. Descripción de pruebas de rendimiento	84
Cuadro 23. Descripción de pruebas de regresión	85

LISTA DE FIGURAS

	Pág.
Figura 1. IOTMach Analysis	27
Figura 2. Tipos de datos de Big Data	30
Figura 3. Las 3 magnitudes claves de Big Data	32
Figura 4. Líneas de desarrollo Data Minig	33
Figura 5. Descubrimiento de conocimiento en base de datos	34
Figura 6. Extracción del conocimiento de una base de datos	36
Figura 7. Aplicaciones IOT	41
Figura 8. Red de sensores inalámbricos	43
Figura 9. Diagrama de caso de uso general	51
Figura 10. Caso de uso gestionar conexiones	51
Figura 11. Caso de uso gestionar repositorio local	52
Figura 12. Caso de uso realizar análisis descriptivo	52
Figura 13. Caso de uso realizar análisis predictivo	52
Figura 14. Caso de uso gestión de estadísticas	53
Figura 15. Interfaz administración de conexiones	57
Figura 16. Interfaz repositorio local	57
Figura 17. Interfaz análisis descriptivo	58
Figura 18. Interfaz análisis predictivo	58
Figura 19. Interfaz de estadísticas	59
Figura 20. Interfaz de estadísticas	59
Figura 21. Logo R Project	60
Figura 22. Hola mundo en R	61
Figura 23. Logo de RStudio	63
Figura 24. Interfaz de RStudio	64
Figura 25. Estructura de una aplicación Shiny	66
Figura 26. Creación de un proyecto Shiny	66
Figura 27. Estructura de datos para el algoritmo Apriori	69
Figura 28. Estructura de datos para el algoritmo Kmeans	70
Figura 29. Estructura de datos para el algoritmo Prcomp	71
Figura 30. Estructura de datos para el algoritmo Predict	72
Figura 31. Estructura de datos para el algoritmo Random forest	74
Figura 32. Estructura de datos para el algoritmo Naive bayes	75

Figura 33. Menú inicio	76
Figura 34. Menú administrar conexiones	77
Figura 35. Submenú clusterización dentro del análisis descriptivo	77
Figura 36. Submenú clusterización pestaña variables	78
Figura 37. Submenú clusterización pestaña grupos	78
Figura 38. Submenú regresión lineal dentro del análisis predictivo	79
Figura 39. Submenú regresión lineal pestaña variables	79
Figura 40. Submenú regresión lineal pestaña resultados	80
Figura 41. Subida de archivos repositorio local	80
Figura 42. Respuesta sobre el funcionamiento de la aplicación	87
Figura 43. Respuesta sobre los gráficos de la aplicación	87
Figura 44. Respuesta sobre ortografía y textos	88
Figura 45. Respuesta sobre la responsividad de la aplicación	88
Figura 46. Tiempo de respuesta carga de archivos a la aplicación	89
Figura 47. Resultados prueba de concurrencia	90
Figura 48. Estimación de los tiempos de carga	90

LISTA DE ANEXOS

	Pág.
ANEXO A. Prueba unitaria: Administración de conexiones	98
ANEXO B. Prueba unitaria: Análisis descriptivo	99
ANEXO C. Prueba unitaria: Análisis predictivo	100
ANEXO D. Prueba unitaria: Repositorio local	101
ANEXO E. Prueba de usabilidad: Interfaces de usuario	102
ANEXO F. Prueba de integración: Análisis descriptivo	103
ANEXO G. Prueba de integración: Análisis predictivo	104
ANEXO H. Prueba de integración: Análisis con datasets de ejemplo	105
ANEXO I. Prueba de rendimiento: Carga	106
ANEXO J. Prueba de rendimiento: Concurrencia	107
ANEXO K. Prueba de regresión: Estimación de carga	108

GLOSARIO

AGRICULTURA DE PRECISIÓN: se lo define como el conjunto de técnicas de cultivo que utiliza tecnologías de la información para solucionar problemas en el campo de la agricultura de una manera eficiente, optimizando la utilización de recursos que se emplearán para el proceso de vida de un determinado cultivo.

ALGORITMO: es una secuencia de instrucciones que se deben seguir para la resolución de un problema. Comúnmente este término es muy conocido en la programación en la que permite crear funciones o métodos.

APLICACIÓN WEB: son aquellas aplicaciones en las que es necesario el uso del internet por parte de los usuarios, accediendo a lo que se denomina Servidor web; una aplicación web es codificada en lenguajes soportados por navegadores web.

BIG DATA: se denomina a toda la información que tiene características de volumen, velocidad y variedad que no pueden ser analizadas o procesadas con herramientas tradicionales que existen actualmente.

CLASIFICACIÓN: es un algoritmo que resuelve problemas de clasificación de la información proporcionada, en donde se debe conocer de antemano el número de campos o variables y se debe tener datos de entrenamiento para la solución en una muestra nueva de datos.

CLUSTERIZACIÓN: se conoce así a la técnica de agrupamiento en la minería de datos, en la que consiste en la división de los datos en grupos con características similares entre ellos.

CRISP-DM: es una metodología utilizada en proyectos para la resolución de problemas relacionados a la minería de datos, constituida por un conjunto de tareas que se deben cumplir para llevar a cabo un objetivo.

CSV: comma-separated values (valores separados por coma), es un tipo de archivo que almacena datos que pueden tener cualquier clase de valor (numérico, fecha, carácter, booleano) en texto plano, cuyos datos están separados por coma generalmente. Cada

línea en el archivo representa un registro, del que se componen uno o más campos que pueden estar delimitados también por tabuladores o punto y coma.

DATA FRAME: es un término en inglés que significa marco de datos, se la utiliza en la construcción de aplicaciones en el lenguaje R. Es usada para almacenar la estructura de las tablas de datos, compuesta por una lista de arreglos (vectores) que deben tener la misma longitud de campos.

DATASET: en español quiere decir conjunto de datos y representa el contenido que puede ser de una tabla de base de datos, archivos de texto plano, delimitados por coma, tabuladores o punto y coma; en donde cada columna representa una variable o campo y cada fila representa el registro de dicha estructura.

DENODO: es una plataforma de software que se utiliza para la integración de manera virtualizada de los datos de diversa variedad o naturaleza que proceden de diferentes fuentes de datos, sean éstas tablas de bases de datos, vistas, hojas de cálculo, web o archivos delimitados.

DESCRIPTIVO: son técnicas que hacen uso de los datos para realizar observaciones de las características que contienen. Se utiliza para realizar análisis de la información en el área de minería de datos.

ESTADÍSTICA: se trata de una ciencia que permite estudiar los datos de una manera resumida, con el objetivo de realizar comparaciones y llegar a un resultado para la toma de decisiones en una determinada área.

FRAMEWORK: en español: marco de trabajo, es un entorno que permite el desarrollo de aplicaciones en un determinado lenguaje de programación, permitiendo la fácil realización de proyectos, en donde facilitan aspectos como soporte del programa, plantillas, bibliotecas, entre otros.

IOT: abreviatura de Internet of Things (Internet de las cosas en español), es un sistema de máquinas que se compone de tecnología para la recopilación de datos (protocolos de comunicación), de manera que los objetos se puedan comunicar entre sí, generalmente sensores, actuadores y dispositivos electrónicos similares.

IOTMACH: es un proyecto integrador de carácter investigativo enfocado principalmente en la solución de problemas relacionados con la agricultura de precisión. Es un trabajo que nace desde las aulas, con estudiantes comprometidos día a día con todo lo que hacen.

KDD: es la abreviatura de Knowledge Discovery in Databases en español: proceso de extracción de conocimiento en bases de datos, lo cual se refiere al proceso de descubrir información útil dentro de los datos que se pueden encontrar en cualquier fuente o repositorio de datos.

KMEANS: es un método de agrupamiento (clusterización) utilizado en la minería de datos, que tiene como finalidad la distribución de N observaciones en K grupos en donde cada observación o pertenece al grupo o clúster más cercano a la media.

LENGUAJE R: es un lenguaje y entorno de programación orientado al análisis estadístico y la representación gráfica de los resultados que se obtengan aplicando cualquier tipo de algoritmo. Es un proyecto GNU por lo tanto es open source y su estructura de programación se representa por comandos, dando como resultado una serie de líneas denominadas scripts.

MACHINE LEARNING: en español aprendiza de máquina o aprendizaje automático, es una disciplina científica que se encarga de que los sistemas aprenden de una manera instintiva, por medio de algoritmos predice comportamientos futuros.

MAIN PANEL: es un tipo de interfaz de usuario utilizada para la administración de las funciones que se pueden ejecutar en las aplicaciones, ya sean escritorio, web o móvil.

MINERÍA DE DATOS: es el proceso que mediante el conjunto de técnicas y tecnologías permite explorar o detectar la información que se detecta en grandes volúmenes de datos, con el objetivo de encontrar patrones o tendencias que puedan explicar el comportamiento de todos los datos en un determinado argumento. Normalmente dichos patrones no se pueden ver mediante el análisis tradicional debido a que existen relaciones complejas o por el simple hecho de haber abundancia de datos.

MONGODB: es un sistema de base de datos No SQL, orientada a documentos, desarrollado con código abierto. Permite guardar los datos con una estructura parecida al JSON, no utiliza tablas como en caso de bases de datos relacionales.

MYSQL: es un sistema de gestión de base de datos relacional, de código abierto y utiliza el lenguaje de consultas SQL. Permite obtener la información guardada en las bases de datos mediante sentencias de una manera rápida y eficiente.

NAIVE BAYES: es un algoritmo utilizado en la minería de datos, en la cual permite clasificar y predecir la probabilidad de hipotéticos resultados. Hace uso de datos históricos para encontrar tanto asociaciones como relaciones y hacer predicciones, basándose en modelos de probabilidad.

ORACLE: es una potente herramienta que se basa en el modelo cliente/servidor para la gestión de bases de datos relacionales. Por la potencia que lo caracteriza y su precio elevado en el mercado, por lo que su uso se da generalmente en grandes empresas y multinacionales.

PLOTLY: es una librería (API) que se encuentra disponible para ser aplicada en distintos lenguajes de programación. Básicamente se trata de una herramienta para crear y visualizar gráficos para diversas plataformas, más comúnmente para sitios o aplicaciones web.

POSTGRESQL: es un sistema de gestión de bases de datos relacionales, con código fuente disponible libremente. Es uno de los más potentes dentro del mercado, el cual es muy utilizado y popular dentro del almacenamiento de información.

PREDICTIVO: engloba un conjunto de técnicas utilizadas dentro de la minería de datos. Se encarga del análisis de datos históricos y nuevos, para realizar predicciones sobre futuros eventos que puedan ocurrir. Permite extraer conclusiones una vez realizado un tratamiento de los datos, mediante la aplicación de métodos estadísticos, matemáticos y el de reconocimiento de tendencias o patrones.

REGLA DE ASOCIACIÓN: es un término utilizado en la minería de datos y Machine Learning, se emplean para revelar hechos que ocurren en un determinado conjunto de datos. Se ubica en el orden de las técnicas descriptivas.

REGRESIÓN: es una técnica que se utiliza para simular la relación que existe entre variables (dependientes e independientes), por lo que permite predecir el comportamiento de una variable mediante la construcción de modelos.

RSTUDIO: es un entorno de desarrollo integrado (IDE), además es software libre, utilizado para construir aplicaciones en el lenguaje R. Permite acceder de manera sencilla a todas las funcionalidades que posee R, ya que incorpora un sinnúmero de funciones para dicho propósito.

SEMMA: es una metodología utilizada en proyectos para la resolución de problemas relacionados a la minería de datos, comparada con la metodología CRISP-DM es menos extensa y se enfoca menos en los objetivos empresariales.

SHINY: es un framework para el desarrollo de aplicaciones web para el lenguaje de programación R. Permite crear app interactivas, permitiéndoles a los usuarios manipular los datos sin necesidad de trabajar con el código. Una aplicación Shiny consta básicamente en dos archivos: ui.R y server. R, los cuales permiten ordenar el código que será presentado para el cliente y el último permite crear toda la lógica del negocio, resolver peticiones de los usuarios que puedan realizar cuando se encuentren interactuando en dicha aplicación.

SQLITE: es un sistema de gestión de bases de datos relacionales, es un software de código libre. Soporta múltiples tablas, procedimientos almacenados, triggers y vistas, s una pequeña biblioteca escrita en C.

VARIABLE DEPENDIENTE: denominadas a las variables que resultan de un estudio o experimento, es decir que los valores que se obtienen dependerán de otras variables denominadas independientes.

VARIABLE INDEPENDIENTE: denominadas a las variables que se conocen desde el inicio de un estudio o experimento, es decir, su valor no va a depender de ninguna otra variable.

WSN: son las siglas de Wireless Sensor Networks, traducido al español como red de sensores inalámbricos. Es un conjunto de dispositivos que se encuentran ubicados estratégicamente en lugares autónomos, utilizando sensores, actuadores o cualquier tipo de aparato que permita el monitoreo de condiciones físicas y ambientales dependiendo del campo de aplicación de la red.

INTRODUCCIÓN

En la actualidad, las empresas dedicadas a cualquier tipo de actividad comercial, dependiendo del tamaño de las mismas, generan información diaria, podemos hablar que puede llegar a ser cientos y miles de datos. Las empresas privadas tanto como las públicas deben tener la capacidad de ser adaptativas, aprender cómo resolver problemas y generar conocimiento, para establecer nuevos métodos en pro de la resolución de los mismos [1]. La minería de datos se trata de un área interdisciplinar fuertemente relacionada con el aprendizaje automático y la estadística [2], que como resultado permite que los datos sean usados para realizar predicciones y poder sacar conclusiones para futuras tomas de decisiones.

Los términos de aprendizaje de máquina, estadística y minería de datos pueden ser aprovechados en cualquier área o sector económico, porque permiten la extracción del conocimiento de los datos y en cada una se tiene la posibilidad de aplicar innumerables técnicas, su objetivo es el mismo y varía únicamente en la forma como los expertos hacen referencia a la hora de utilizarlas [3].

En toda empresa, el aprovechamiento de los datos debe ser el pilar fundamental para el éxito de la misma, para lo cual se debe contar con herramientas dedicadas al análisis y entendimiento de dichos datos, una mejor vía es hacerlo con los recursos computacionales que solventen esta necesidad, más precisamente el aprendizaje de máquina. El aprendizaje de máquina o también llamado Machine Learning, proporciona a las computadoras la capacidad de aprender, es decir, se puede desarrollar programas informáticos que puedan “auto enseñarse” para adaptarse a las situaciones a medida que vayan ingresando datos nuevos al sistema. Cabe destacar que el aprendizaje de máquina emplea algoritmos que permiten resolver un problema a partir de información histórica que hayan sido aprendidos en el pasado o extraer sus propias conclusiones en base a los que exista en un conjunto de datos. Machine Learning tiene aplicaciones muy interesantes dentro del campo de las tecnologías, un ejemplo claro es cuando un usuario está decidido a comprar un libro en algún sitio de internet, al momento de hacerlo, el sitio le arroja unas “sugerencias” acerca de otros libros de similar contenido que otros usuarios han comprado junto con el suyo; esto quiere decir que si este usuario decide comprar no solo su libro sino otro que le pareció interesante, la técnica de Machine

Learning que se ha aplicado tuvo los resultados deseados y por ende las ganancias por la venta de dos libros ayudará al negocio a seguir produciendo.

El aprendizaje de máquina manipula diversas fuentes de información que pueden encontrarse en sistemas automatizados, hojas de cálculos, archivos planos, datos de sensores inalámbricos, Big data, entre otros. Big data involucra características de volumen, veracidad y velocidad de datos; actualmente todo esto se encuentra englobado en el Internet de las cosas (IoT)

IOT percibe un mundo donde los dispositivos que lo conforman pueden ser identificados en el Internet y está creciendo a un ritmo acelerado con nuevos dispositivos que se van conectando [4]. Hablar de IOT, hace referencia a una arquitectura de comunicación que puede abarcar muchas tecnologías, compuesta principalmente por una red de sensores inalámbricos, redes móviles, actuadores. Cada uno de estos componentes debe tener la capacidad de comunicarse entre sí operando en base a reglas que se le hayan suministrado para que realicen o se abstengan de ejecutar alguna acción y enviando la información obtenida cuando se quiera acceder a ellos desde cualquier parte, a toda hora.

La automatización de los procesos en la agricultura está hoy en día en gran auge para revolucionar el mundo en este sector importante de la economía. Un concepto más preciso de este término es la gestión de parcelas agrícolas con el fin de aumentar la productividad desde los pilares de la sostenibilidad ambiental [5]. El Internet de las cosas ya está envolviendo el área de la agricultura, el cual permite un control óptimo sobre cada parcela de cultivos, realizando acciones automatizadas como: gestión del riego, medir temperatura, humedad, PH, alertar en caso de algún evento que esté ocurriendo a su alrededor, entre otros. Toda esa información que es generada mediante la comunicación entre los sensores, actuadores y demás dispositivos, debe ser analizada para descubrir la existencia de patrones, obtener respuestas y sacar conclusiones de lo qué está pasando con los cultivos.

Una vez más se hace referencia a la minería de datos, aprendizaje automático y estadísticas, para llevar a otro nivel la obtención de dicha información, procesarla y como resultado crear conocimientos que ayudarán a la toma de decisiones de algún hecho en particular.

Por lo tanto existe la necesidad de manipular los datos que el sistema IOTMACH genere dentro de las actividades concernientes a la agricultura, con ello dicha información no quede simplemente almacenada y olvidada con el transcurso del tiempo.

Con todo lo descrito anteriormente, se ha decidido presentar la solución a este problema con el tema: “IMPLEMENTACIÓN DE UN MÓDULO DE ANÁLISIS ESTADÍSTICO Y PREDICTIVO PARA AGRICULTURA UTILIZANDO BIGDATA Y MACHINE LEARNING, INTEGRADO AL SISTEMA IOTMACH”.

Para cubrir aspectos de análisis estadístico y predictivo, el lenguaje R [6] es la herramienta indicada para cumplir esa misión, es la que se usó en el trabajo práctico, debido a que permite, mediante sus algoritmos y gráficos estadísticos, visualizar de manera clara, organizada todo lo que se ha descubierto con la información que se le ha proporcionado.

El objetivo que tiene este proyecto es ayudar a la toma de decisiones a las personas que están encargadas de administrar cualquier actividad del sector productivo, teniendo como referencia toda la información que se genera a diario, y por ende ayudar en el normal desarrollo de la misma.

El presente trabajo de titulación se ha organizado de la siguiente manera:

Capítulo 1: Diagnóstico de necesidades y requerimientos. En donde se detalla las necesidades que se intentan satisfacer.

Capítulo 2: Desarrollo del prototipo. Se detallan los objetivos, la construcción del módulo y la ejecución de la aplicación que se va implementar.

Capítulo 3: Evaluación del prototipo. En esta parte se realizan pruebas a la aplicación y así comprobar que es un producto con calidad.

1. DIAGNÓSTICO DE NECESIDADES Y REQUERIMIENTOS

1.1 Ámbito de Aplicación: descripción del contexto y hechos de interés

El aprendizaje automático o aprendizaje de máquina (Machine Learning) es una tecnología muy interesante que se aplica hoy en día en diferentes sectores de la industria y de la ciencia. Se lo puede catalogar como un tipo de inteligencia artificial, debido a que pretende crear máquinas capaces de comprender cosas que los seres humanos saben por naturaleza o por adquisición de conocimiento que lo han hecho durante el trayecto de su vida. Para lograr este objetivo, la única manera es hacerlo mediante algoritmos de aprendizaje que traten de imitar cómo el cerebro humano trabaja, empleando tecnologías de punta que permitan la solución de un determinado problema y muchas de las veces van a tratar de suplir las actividades que realiza una persona, porque una máquina de aprendizaje haría todo eso por sí misma.

Las razones por las que el aprendizaje automático se ha difundido en gran proporción tienen que ver con el crecimiento de la web y de la automatización de los procesos en distintas organizaciones; es decir que ahora los datos que se generan están siempre aumentando en volumen y variedad. Por ejemplo una empresa puede recolectar datos acerca de las preferencias que tienen los usuarios en un sitio web e intentan aplicar algoritmos de aprendizaje para explorar esos datos y así poder entender mejor a los usuarios, brindándoles mejores servicios de calidad.

En los sectores de la medicina, comunicaciones, servicio al cliente, biología y en todas las áreas de ingeniería se tienen conjunto de datos cada vez más y más grandes que a menudo se los trata de comprender empleando técnicas que existen actualmente para cada tipo de necesidad.

La información generada de distintas fuentes de datos puede llegar a tener un tamaño extremo, las características de volumen, variedad y velocidad están ligadas al concepto de Big Data [7]. Todo esto exige diferentes maneras nuevas de procesamiento de la información, que permita un mejor enfoque, toma de decisiones y automatización de procesos; cuando se habla de la magnitud volumen, se está refiriendo a la cantidad de datos que puede existir, la variedad hace hincapié al número de tipos de datos y por último la velocidad apunta a la rapidez en la que los datos son procesados. La forma con que estas tres características se adapten a los nuevos cambios dentro de éste

contexto, será una clave del éxito dentro los sistemas de inteligencia de negocios en un futuro muy cercano.

El éxito radica en la buena utilización de herramientas o técnicas que permitan aprovechar los aspectos que se esconden en toda la información generada. Para ello, entre las principales técnicas que son empleadas en el proceso de Machine Learning se encuentran la descriptiva y predictiva. Un modelo predictivo es un método estadístico para estimación (predecir) el resultado de una nueva observación (un dato registro) basada en el conocimiento obtenido a partir de unos observaciones anteriores [8]. Las predicciones dan una visión de cómo se verán afectadas ciertas características en un futuro, anticipándose al hecho para que se tenga tiempo y así poder elegir la mejor manera una acción a tomar.

El análisis predictivo tiene el potencial de ofrecer un valor en una diversidad de áreas [9]. Entre los cuales figuran los siguientes:

Cuadro 1. Áreas de aplicación en el análisis predictivo

Área	Casos en los que se utiliza
Energía	Patrones y gestión de consumo de energía
Servicios financieros	Identificación de fraudes, morosidad, entre otros
Comida y bebidas	Predicción de la demanda de suministros para empaque de productos para su posterior envío
Transporte	Optimización de rutas de entrega
Marketing	Predicción del comportamiento, consumo o propensión a pagar.
Manufacturación	Calidad y optimización de fallos de máquinas, tiempo de inactividad, entre más.

Fuente: Elaboración propia

Otra técnica empleada es la descriptiva, la que está orientada a la caracterización de un conjunto de datos, a diferencia de las técnicas predictivas las cuales están orientadas a estimar valores. Los modelos descriptivos buscan particularidades entre todas las variables (campos) que compone una estructura de datos, no disponen de la necesidad de tener información histórica para extraer el conocimiento para la toma de decisiones en el futuro.

El proyecto ha sido desarrollado utilizando algoritmos de minería de datos y Machine Learning con técnicas descriptivas y predictivas, supervisadas por grupos de investigación que día a día trabajan continuamente dentro del ámbito de aprendizaje automático. Como resultado, el módulo desarrollado está integrado al sistema IOTMACH, el cual es un proyecto que engloba una variedad de tecnologías que son aplicadas exclusivamente al sector de la agricultura, pudiendo realizar procesos que ayuden a aprovechar los recursos para mantener la producción de cada cultivo.

El aporte del presente trabajo de investigación es la construcción y posterior implementación de un software para el análisis estadístico, descriptivo, predictivo de datos utilizando algoritmos de aprendizaje de máquina, descubrimientos de patrones mediante la minería de datos y Big Data enfocado a la agricultura e integrado al sistema IOTMACH, que será útil para ayudar en la toma de decisiones de acuerdo a la necesidad que se pretenda satisfacer.

1.2 Establecimiento de requerimientos

Los datos que se generan dentro de un sistema de algunas empresas o instituciones generalmente son almacenados y olvidados. Hoy en día la información va creciendo cada vez más, por lo cual se debe usar herramientas que permitan aprovechar la información para descubrir conocimiento útil y tomar decisiones.

La necesidad de aplicar análisis a los datos dentro una organización es muy importante debido a que permite conocer en qué situación se encuentran los procesos dentro de cada área de un sistema, es decir la relevancia que guardan los datos harán que se llegue a tomar decisiones acertadas una vez realizado el tratamiento de los datos. Al contrario de organizaciones o negocios que no cuentan con las herramientas de análisis tanto descriptivo como predictivo, viven situaciones de incertidumbre sobre qué les deparará el futuro con relación a su permanencia dentro del mercado, la competitividad frente a otras empresas, y demás factores que una buena organización debería conocer para ser competente en el mercado.

El campo de la agricultura de la cual se enfoca el sistema IOTMACH, no posee un módulo que se encargue del análisis que se describió anteriormente, para un sistema con estas características es indispensable el estudio de los datos de una manera diferente en comparación con otros sectores productivos. Entre los datos más importantes que se deben analizar de un cultivo, se encuentran aspectos relacionados

con la temperatura del ambiente [10], humedad [11], salinidad, condiciones físicas del suelo, mediciones de Ph, longitud de las plantas y todo lo relacionado a cada parte de la misma. Todos estos datos provienen de distintas fuentes, una de ellas es IOT, más precisamente de una red de sensores inalámbricos que se encargan de recolectar toda la información posible de lo que está ocurriendo en un determinado cultivo.

Esa abundante información no tendría significancia si sólo se queda en simples archivos, bases de datos o en cualquier otro modo de almacenamiento que permitan únicamente su visualización de manera estática. La alternativa para aprovechar el valor de los datos es tener a mano una herramienta con la capacidad de generar conocimiento a partir del descubrimiento de los patrones o semejanzas que guardan los conjuntos de datos, conocimiento que va a ayudar a controlar y aprovechar los recursos que deben ser aplicados a una determinada área, logrando la eficiencia en el desenvolvimiento de cada uno de los procesos que a diario se realizan. A parte del óptimo uso de recursos que se logra con dicha herramienta, se puede predecir cualquier tipo de circunstancias beneficiosas o adversas que vayan a suceder dentro de un determinado tiempo, clasificar o agrupar características entre los datos, y mucho más; de esta manera la toma de decisiones dentro de este sector no se echará a la suerte, conociendo de antemano que es lo que se deberá hacer y qué gestiones se deben abstener de hacerse.

Es por esa razón que se planteó la implementación del módulo que se encargue del análisis estadístico y predictivo integrado al sistema IOTMACH, aplicando técnicas de minería de datos conjunto con algoritmos de Machine Learning.

1.3 Justificación del requerimiento a satisfacer

Actualmente el análisis de datos es indispensable en todo sistema automatizado que diariamente derive flujos de información, sean estas transacciones o cualquier tipo de operación en la que se manipule un sistema de gestión de base de datos, sistemas estructurados, no SQL, textos planos, registros en archivos con formato CSV, en fin, todo aquello que intervenga en la generación de datos.

Nos encontramos en una era en la que las herramientas informáticas ayudan al desarrollo sostenible en una organización, las mismas que pueden realizar cualquier función tan eficazmente que logran cumplir aspectos indispensables para con el negocio, así mismo la necesidad de aprovechar cada recurso obliga a buscar técnicas que resuelvan y satisfagan los problemas que se presentan en el transcurso del tiempo;

el éxito de cómo se lo haga radica en el conocimiento valioso que los datos poseen, los cuales permanecen recónditos, no se los puede apreciar a simple vista, teniendo en cuenta varios factores que son entendibles por las personas.

La minería de datos se encarga precisamente de extraer lo más relevante de un conjunto de datos que posteriormente será útil para realizar análisis y tratamientos que podrán mejorar gran parte de las decisiones tomadas para solventar algo que se vive con incertidumbre. Junto al aprendizaje automático aplican técnicas capaces de comprender el problema que se está analizado y brindar una respuesta acertada en el momento indicado que a priori, será en beneficio del bien común dentro de una empresa. Dentro del campo de la agricultura se pueden optimizar muchos recursos si se cuenta con técnicas de aprendizaje que supervisen cada área de un cultivo; se puede hablar de predecir si un cultivo enfermará dependiendo de los factores que directamente estén ligados a éste, uso eficiente de los suelos, clasificar plantas con ciertas características, utilización correcta del agua en los riegos, controlar el uso de químicos, verificar la productividad que se genera cuando un cultivo está en fase de cosecha, entre un sinnúmero de actividades que a la larga, prometerán un desenvolvimiento acertado en el sector de la agricultura.

La implementación del módulo contribuye a la exploración, gestión y mantenimiento de la información que se genera desde diferentes fuentes de datos en el sistema IOTMACH. Se empleó el lenguaje de programación R, creado exclusivamente para tratar los datos y analizarlos de una manera estadística, haciendo uso de los algoritmos de clusterización, creación de reglas de asociación, realización de predicciones, clasificación de los datos y búsqueda de patrones propios de las técnicas de Data Mining y Machine Learning.

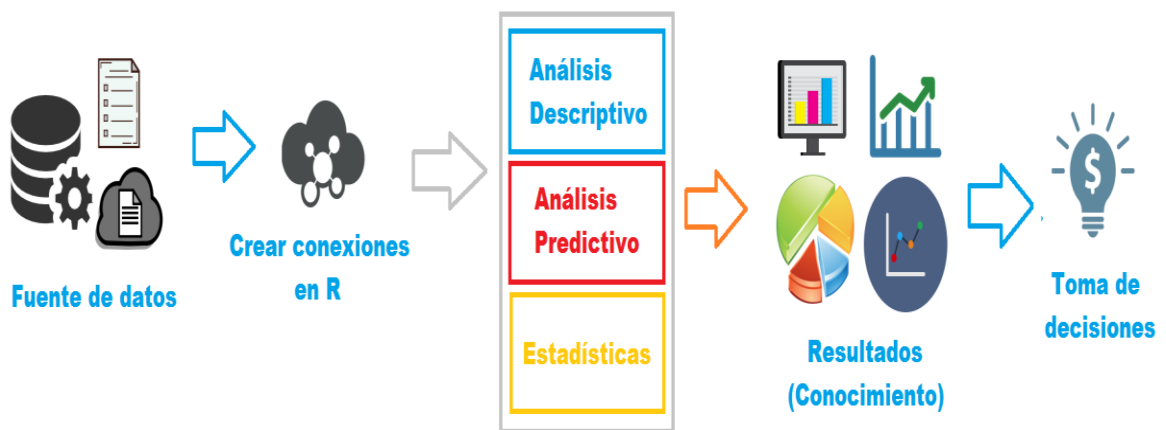
La herramienta desarrollada está orientada a la web, construida bajo el framework Shiny, una librería de R, cuya aplicación se compone de una interfaz con sutiles menús de fácil manejo para los usuarios, en donde se podrá aplicar algoritmos predictivos y descriptivos para cierto conjunto de datos en donde cada proceso irá acompañado de ayudas para el entendimiento total y aprovechamiento de la aplicación. El resultado que arroje cada algoritmo ayudará reforzar las decisiones que se vayan a tomar, reduciendo en cierta forma recursos como tiempo, dinero, esfuerzos; el sistema IOTMACH necesitaba de la implementación de este módulo porque su excelente funcionamiento reside en la riqueza que se pueden encontrar en los datos obtenidos desde la WSN (red de sensores inalámbricos) [12] y la gestión de cultivos agrícolas.

2. DESARROLLO DEL PROTOTIPO

2.1 Definición del prototipo tecnológico

El módulo implementado en el sistema IOTMACH se lo ha desarrollado ambientado a una aplicación web interactiva, utilizando el framework Shiny para R, permitiendo aprovechar todas las librerías de análisis de este potente lenguaje de programación. El nombre de la aplicación se ha denominado IOTMach Analysis.

Figura 1. IOTMach Analysis



Fuente: Elaboración propia

Lo conforma un panel de administración con varios menús, siendo los más significativos los que se detallan a continuación:

Administración de conexiones, encargado de crear los enlaces con distintas fuentes de datos, siendo éstas a través de controladores de bases de datos como PostgreSQL, MySQL, Oracle; también por medio de drivers que permiten la administración de datos virtualizados como es el caso de Denodo [13] y mediante archivos con gran volumen de información, sean éstos en formatos de texto plano, csv, archivos alojados en la nube (DropBox o Google Drive) y permitiendo la subida de archivos locales en un repositorio dentro de la aplicación. Las conexiones que se vayan creando serán almacenadas en un archivo que a posterior podrá ser usado para conectarse a una fuente de datos para su análisis en cualquier algoritmo que se encuentre trabajando.

El análisis de datos, cuando se especifique una conexión a una fuente de datos o se seleccione un ejemplo predefinido en la aplicación, entre las técnicas que componen este menú se encuentran: predictivas y descriptivas.

Las técnicas descriptivas constan de algoritmos de segmentación o clusterización (kmeans), algoritmos para crear reglas de asociación de datos transaccionales (a priori) y análisis de componentes principales (ACP).

Las técnicas predictivas se estructuran de algoritmos de regresión (modelo de regresión lineal), predicción (random forest) que permite analizar los datos con árboles de decisión y clasificación mediante algoritmos bayesianos (naive bayes).

El menú Estadísticas, en donde se podrá apreciar de manera gráfica las situaciones que están pasando en un cultivo, riego utilizado, predominio de tipos de suelo, productiva de un cultivo con el transcurso del tiempo, temperatura, humedad, enfermedades y distinta técnicas propias de la estadística que el usuario puede hacer uso para cualquiera que sea su necesidad.

Repositorio local, se encarga de almacenar los archivos subidos localmente por los usuarios, permitiendo tener acceso a los mismos en cualquier momento.

Dependiendo de la opción que se elija, el algoritmo arrojará información sobre los campos tratados, detalles de ejecución, gráficos estadísticos que muestran el resultado que al final permitirá conocer las relaciones o semejanzas que se encontraron en los conjuntos de datos, impulsando al usuario a tomar decisiones con el conocimiento extraído de todo el proceso de análisis.

2.2 Fundamentación teórica del prototipo

2.2.1 *Machine Learning*. El Aprendizaje Automático (Machine Learning), es más precisamente descrita como la unión de estadística y la inteligencia artificial [14]. Se refiere a la capacidad que tienen los programas para que aprendan con base en los datos que estudian, los algoritmos analizan los datos e identifican la existencia de relaciones entre atributos y entidades para construir los modelos que permiten a los expertos entender ciertas relaciones que guardan los datos. Dichos algoritmos mejoran su rendimiento en base a la experiencia que adquieran conforme a ejemplos o datos

históricos verídicos que se guardan a lo largo del tiempo como resultado de las actividades de una empresa [15], los cuales se denominan conjuntos de entrenamiento.

Cuadro 2. Desafíos de Machine Learning

	Características
Datos	<ul style="list-style-type: none"> - Gran cantidad de datos. - Muchas variables o campos (en aprendizaje supervisado existen variables predictoras). - Datos demasiados complejos.
Aprendizaje	<ul style="list-style-type: none"> - Experiencia técnica significativa es requerida. - No existe una solución para todo, requiere una perspectiva diferente. - Probar múltiples algoritmos y ver cuál trabaja mejor (mayor tiempo invertido).

Fuente: Elaboración propia

Una máquina puede instruirse si se aplican dos conocidos modos de aprendizaje: supervisado y no supervisado [16]; este concepto está estrechamente ligado a la minería de datos, debido a que utiliza técnicas similares, por no decir las misma pero con nomenclatura diferente (aprendizaje supervisado = técnicas predictivas y aprendizaje no supervisado = técnicas descriptivas, de Machine Learning y minería de datos respectivamente). El aprendizaje supervisado requiere de datos de entrenamiento con entradas y salidas definidas, representa la relación que poseen los parámetros descritos anteriormente. El aprendizaje no supervisado se encarga de clasificar el conjunto de muestras proporcionadas en diversos grupos, mediante la indagación de la similitud que existen entre ellos [17].

El objetivo de aprendizaje automático es descubrir, aprender y luego adaptarse a las circunstancias que puedan cambiar con el tiempo, por lo tanto mejorar el rendimiento de una máquina. Para concluir con el análisis de los datos en buscar a una solución de los problemas con aspectos que hayan sido obtenido en dicho análisis.

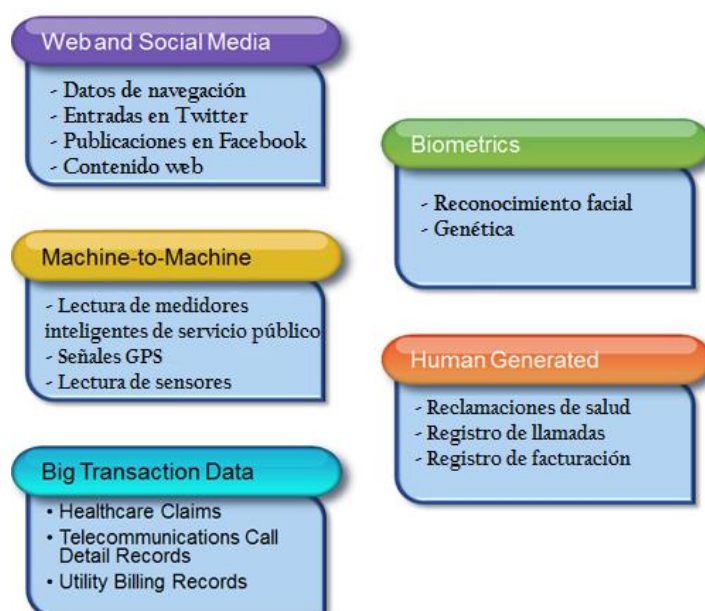
Es un conjunto de métodos que automáticamente pueden detectar patrones en los datos y usar los patrones descubiertos para predecir datos futuros o para ejecutar otra clase de toma de decisión bajo incertidumbre [18]. En la actualidad las grandes empresas han

aplicado el aprendizaje de máquina para desarrollar diferentes proyectos, como por ejemplo podemos encontrar vehículos que pueden ser conducidos sin una persona dentro, programas de reconocimiento y extracción de voz, búsquedas muy efectivas en la web, preferencias de usuarios cuando se realiza alguna transacción en internet. Según los científicos es el avance que van a tener las máquinas en el aspecto de la inteligencia artificial, ya que actuará y tomará decisiones como si fuera una persona real.

2.2.2 *Big Data*. Esta terminología hace referencia a la gran cantidad de información que se genera a través de algún sistema automatizado y hoy en día la fuente principal de esos datos es el internet de las cosas. Para tener un claro entendimiento de lo que es Big Data, podemos decir que es el manejo de gran cantidad de datos que, a diferencia de la información tradicional, no puede ser analizada fácilmente, sino que se deberán utilizar herramientas y técnicas para tratarlas.

Los datos más relevantes que son obtenidos mediante procesos propios de análisis de Big Data, ayudan a la toma de decisiones. Por eso las empresas desde ya deberán aplicarla sobre sus transacciones diarias, sus datos históricos, en fin sobre cualquier tipo de información provenientes de distintas fuentes. Hablar de fuentes de datos, es mencionar la información que se genera diariamente en distintos sistemas de información, redes sociales, el mismo internet de las cosas, sitios web. La representación se puede ilustrar en la siguiente figura:

Figura 2. Tipos de datos de Big Data



Fuente: Referencia [19]

Las grandes cantidades de datos se pueden encontrar en:

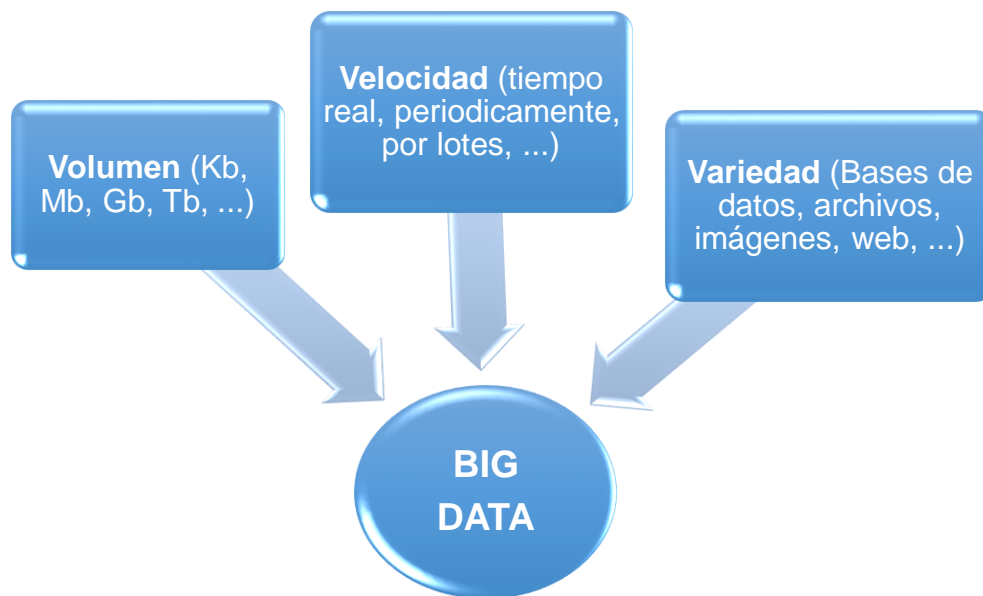
- Web and Social Media. La web y medios sociales, es el contenido que se generan en las distintas páginas web y redes sociales como Facebook, Twitter, blogs, entre otros.
- Machine-to-Machine. Máquina a máquina, son las tecnologías que permiten la comunicación entre sensores, actuadores, medidores, cualquier dispositivo dentro de una red inalámbrica, que pueden capturar eventos que pasan a su alrededores, estos eventos pueden ser la temperatura, humedad, PH, salinidad, presión; toda esta información viaja a distintas aplicaciones que transforman los eventos en información muy significativa.
- Big Transaction Data. Gran cantidad de transacciones de datos, puede incluir registros de facturación, registro de llamadas (en telecomunicaciones), entre otros.
- Biometrics. Biometría, los datos que se generan en las huellas digitales, reconocimiento de retina, reconocimiento facial.
- Human Generated. Generada por personas, lo que incluye mensajes, notas de voz, llamadas telefónicas, registro de atenciones en un Call Center, correos electrónicos, documentos electrónicos.

Para apoyar a la minería de Big data, se requiere plataformas de computación de alto rendimiento que imponen los diseños sistemáticos para liberar todo el poder del Big Data [20]. Este grupo de fuentes de datos puede ir creciendo con el avance de las tecnologías. Las magnitudes que definen “los grandes datos” se encuentran: volumen, velocidad y variedad (llamadas también las 3 Vs de Big data) [21]

Actualmente se puede ver el crecimiento de la cantidad de datos que pueden generarse en distintas fuentes de información, con tamaños que oscilan desde kilobytes hasta petabytes de un sistema de almacenamiento con que cuenta una empresa, esa magnitud se denomina volumen; la velocidad se refiere a la frecuencia con las que se pueden generar nuevos datos, se refiere también a un análisis pertinente de la información y con respuestas que sean en tiempo real. Está muy ligado a la explosión de los medios sociales, debido a que los teléfonos se volvieron más inteligentes, y por esta razón las personas tienen la necesidad de estar siempre conectados, de una u otra manera generan constantemente mensajes de texto, publicaciones en redes sociales, llamada, entre otras. Por último la tercera V de Big data es su variedad, y es que los datos pueden almacenarse en formatos diferentes, con una naturaleza diversa, el volumen de los datos que se manejan comúnmente es digital, por ejemplo las bases de

datos, archivos de texto, hojas de cálculo. Otros formatos que son poco tradicionales se encuentran en vídeos, mensajes de texto, archivos PDF, imágenes, datos de sensores, GPS y un sinnúmero de fuentes de información generadas día a día.

Figura 3. Las 3 magnitudes claves de Big Data



Fuente: Elaboración propia

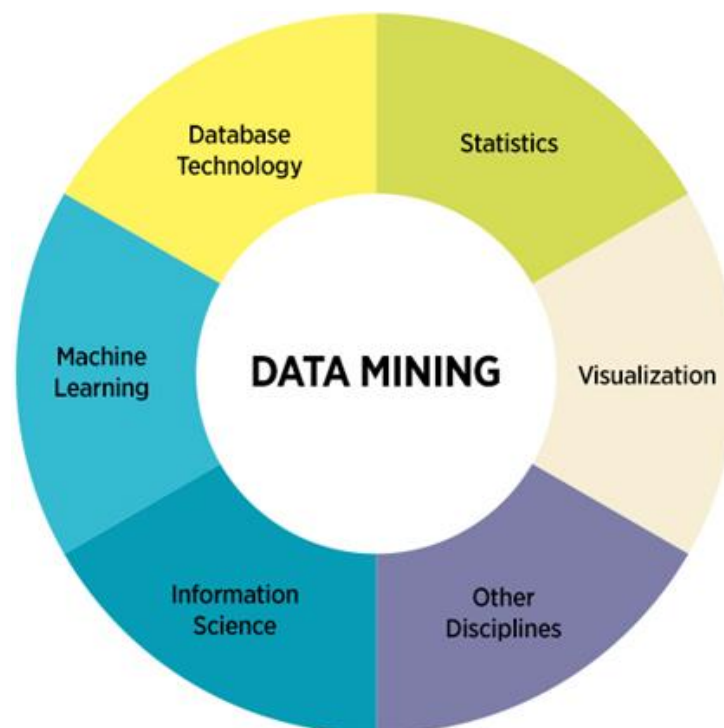
2.2.3 Data Mining. La disponibilidad de grandes volúmenes de información y el uso de herramientas informáticas ha transformado el análisis de datos orientándolos hacia determinadas técnicas especializadas, las cuales se enmarcan en el término de Minería de Datos o Data Mining [22]

La minería de datos es el proceso de descubrir el conocimiento interesante de grandes cantidades de datos [23]. Consistente en la comprensión y extracción de la información más relevante para el desarrollo del negocio en una organización, cuya manera de obtenerla, no se podría hacer de otra manera, debido a que permanece oculta, sin oportunidad de percibirla a simple vista. Abarca muchos campos de estudio, entre los cuales destacan la estadística, el aprendizaje de máquina, recuperación de información y el reconocimiento de patrones entre los datos. La minería de datos está presente en áreas de impacto económico y social, en donde se puede mencionar: el comercio,

comunicaciones, agricultura, economía, organizaciones bancarias, redes sociales y sitios o aplicaciones web, entre otros.

Las líneas de desarrollo que la minería de datos fundamenta, se originan en los siguientes conceptos: la estadística, la inteligencia artificial y el aprendizaje automático [24], que permiten tomar mejores decisiones sobre la información extraída y aplicarla en lo que se crea más conveniente. Las técnicas de minería de datos persiguen un objetivo, el cual es descubrir de manera automática el conocimiento que está “oculto” en toda la información almacenada, transacciones y datos históricos de una organización. Lo que se pretende es obtener patrones que existen entre la relación de unos datos con otros, perfiles, segmentación en grupos que tienen cierta similitud; todo esto permitirá realizar predicciones para conocer cuáles serán los valores futuros en los nuevos datos que vayan ingresando a las bases de datos empresariales, recuperar datos que por alguna razón se encontrasen incompletos y poder rellenarlos con lo que se obtiene de las distintas técnicas que se vayan a aplicar. Las técnicas que se utilizan en la minería de datos se componen de herramientas que facilitan el descubrimiento de la información.

Figura 4. Líneas de desarrollo Data Minig



Fuente: Referencia [25]

Las principales técnicas de Minería de Datos se suelen clasificar según su tarea de descubrimiento [26]. Con esta acotación, podemos mencionar que las técnicas de Minería de Datos se clasifican en dos grandes categorías: técnicas descriptivas y técnicas predictivas [22].

Cuadro 3. Técnicas de Minería de datos

TÉCNICAS DESCRIPTIVAS	Asociación
	Segmentación
	Dependencias
TÉCNICAS PREDICTIVAS	Regresión
	Clasificación
	Predicciones

Fuente: Elaboración propia

Todo proceso para el desarrollo de aplicaciones de minería de datos, necesita aspectos metodológicos que permitan tener una guía a la hora de trabajar con técnicas y herramientas. La minería de datos se rige por los conceptos del ciclo de vida del software, debido a que es netamente área de la ingeniería. Existen varias metodologías para el desarrollo de aplicaciones de minería de datos, entre ellas destacan KDD (Descubrimiento de conocimiento en base de datos), SEMMA y CRISP-DM [27]; las mismas que permiten la estructuración de los proyectos en fases, que tienen un conjunto de tareas.

El descubrimiento del conocimiento en base de datos podrá ser un recurso para la toma de decisiones. La interpretación de ese conocimiento va a proporcionar una comprensión acerca de un universo [28], y por ende, aumentar nuestro conocimiento sobre ese universo. Se lo puede resumir de la siguiente manera:

Figura 5. Descubrimiento de conocimiento en base de datos



Fuente: Elaboración propia

Los datos no suelen estar expresados en una forma simple o entendible, puede estar en forma plana, valores separados por comas, espacios, tabulaciones, entre otros.

El proceso de minería de datos pasará por las siguientes fases:

- Filtrado de datos.
- Selección de variables.
- Extracción de conocimiento.
- Interpretación y evaluación. [29]

Cuadro 4. Comparación de las metodologías

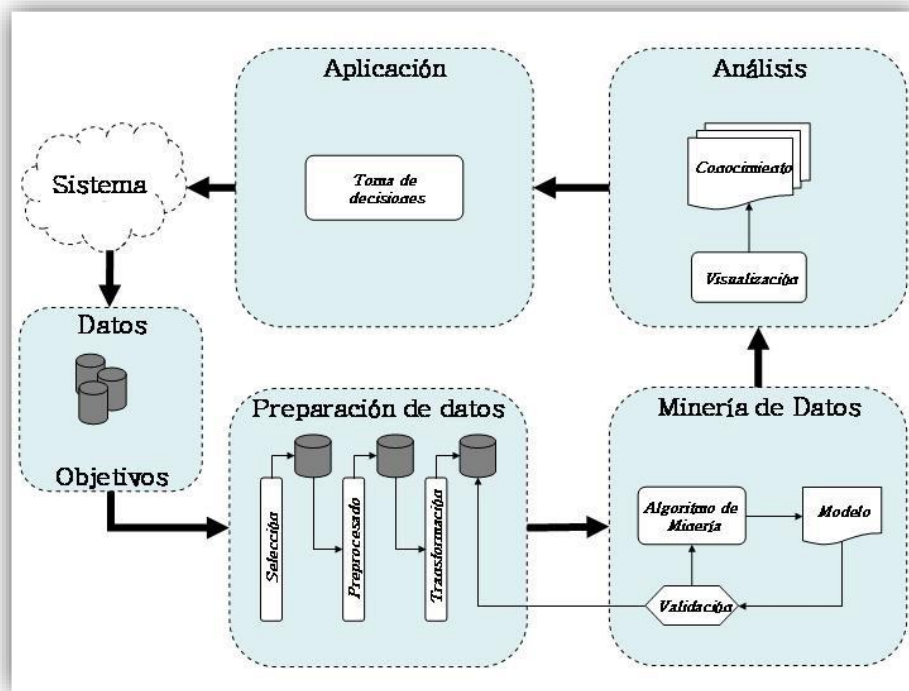
KDD	SEMMA	CRISP-DM
Pre KDD	-----	Conocimiento del negocio
Selección	Muestreo	Comprensión de datos
Pre procesamiento	Exploración	
Transformación	Modificación	Preparación de los datos
Minería de datos	Modelo	Modelado
Interpretación / Evaluación	Valoración	Evaluación
Post KDD	-----	Despliegue

Fuente: Referencia [30]

Todas las técnicas de Data Mining tienen el fin de asumir la capacidad de extracción de patrones, predicción de comportamientos, descubrimiento de tendencias, aprovechando toda la información con la que cuentan las organizaciones; generalmente los datos se encuentra en grandes volúmenes y por ello se dice que la minería de datos está allí para la extracción de conocimiento en bases de datos.

La información de la base de datos que se recogen se consideran conjunto de hechos y los patrones que se descubren son expresiones que describen un grupo de datos. El descubrimiento del conocimiento en las bases de datos involucra un proceso iterativo e interactivo de búsqueda de modelos o patrones [31]. Los patrones que se vayan descubriendo tienen que ser válidos, interesantes y útiles.

Figura 6. Extracción del conocimiento de una base de datos



Fuente: Referencia [31]

El objetivo final de todo esto es incorporar el conocimiento obtenido en algún sistema real, tomar decisiones a partir de los resultados alcanzados o, simplemente, registrar la información conseguida y suministrársela a quien esté interesado.

En conclusión podemos decir que la técnica de minería de datos con todos los procesos que se pueden aplicar, aporta beneficios para elevar los niveles de competencia de los negocios sobre un mercado determinado, atraer y mantener clientes, mejorar la calidad de sus servicios o productos, detectar nuevas oportunidades para el mercado, determinar una ubicación que sea conveniente para abrir una sucursal, entre otros, y su base está en la rapidez para la identificación, procesamiento y extracción de la información que realmente tiene relevancia, por ello se generará el conocimiento a partir de patrones que estén ocultos en distintas fuentes de datos.

2.2.4 *Técnicas predictivas.* Las técnicas predictivas o también llamados métodos de aprendizaje supervisado predicen el valor de un atributo (etiqueta) de un dataset, conocidos otros atributos (atributos descriptivos). A partir de datos cuya etiqueta se conoce se induce una relación entre dicha etiqueta y otra serie de atributos [32]. Están orientadas a estimar valores de salida.

El aprendizaje supervisado necesita de un conocimiento previo, funciona en dos fases, la primera se denomina entrenamiento, la cual consiste en diseñar un modelo en el que utilizemos una parte de los datos con variables conocidas y reales que se obtengan de alguna actividad o transacción de una base de datos. La segunda fase se la denomina testeo, y no es más que la prueba a la cual son sometidos los datos en sus totalidad; todo este proceso arrojará resultados, los mismos que deberán ser interpretados para tomar decisiones que permitirán hacer alguna acción en beneficio de nuestros intereses. Las relaciones entre los datos aplicando técnicas predictivas, ayudan a conocer cuál será la siguiente variable que hasta ese entonces es desconocida.

Los modelos que se construyen con estas técnicas siempre buscan encontrar las respuestas a muchas interrogantes en cuanto a las actividades del negocio en el futuro, como por ejemplo ¿Cuáles serán las ventas para el año 2017?, ¿Esta transacción es válida?, ¿Qué tipo de producto consumirá el cliente X?, ¿Cuál será el costo de mi casa?, entre otros. Podemos citar algunos casos en donde estas técnicas de aprendizaje supervisado permiten tener el panorama claro para ciertas situaciones inciertas que pueden ocurrir a priori; respondiendo a las interrogantes que se plantearon como ejemplos anteriormente, podemos predecir cuánto se venderá el otro año si tenemos de antemano todos los registros de las ventas que se hicieron día a día, mes a mes, que son los datos que ayudarán a resolver esa suposición. También podemos saber cuál será el valor de una casa con X m² si ya tenemos datos de otras casas vendidas, teniendo en cuenta su tamaño con el precio en el que ha sido vendida, lo que cambia el enfoque de análisis histórico mirando hacia atrás para pronosticar el futuro y ofrecer una gama de posibles cursos de acción [33].

Conocer a priori como irá evolucionando una variable en el futuro, será una información muy valiosa e indiscutiblemente llegará a ser una ventaja competitiva para las otras organizaciones. Disponer de dichas predicciones va a permitir adecuar de manera óptima las respuestas en un tiempo determinado, esto puede resultar muy útil en los campos más diversos:

- Detección de oportunidades.
- Prevención de problemas.
- Gestión óptima del personal.
- Optimización de stocks. [29]

2.2.5 *Técnicas descriptivas.* Las técnicas descriptivas o también llamados métodos de aprendizaje no supervisado descubren patrones y tendencias en los datos actuales (no utilizan datos históricos). El descubrimiento de esa información sirve para llevar a cabo acciones y obtener un beneficio (científico o de negocio) de ellas [32]. Están orientadas a describir un conjunto de datos.

El aprendizaje no supervisado no necesita de datos de entrenamiento con las respuestas conocidas de antemano, ya que analiza una gran cantidad de variables y por sí solo determina características comunes que pueden ser segmentadas en grupos pudiendo formar conceptos que previamente no eran conocidos. Se dice que estas técnicas observan los patrones que hay dentro de un volumen de información y llegan al objetivo de descubrimiento.

No se supone que existan variables dependientes ni independientes y tampoco se supone que haya un modelo para tratar los datos previamente. Los modelos se crean automáticamente por medio del reconocimiento de patrones, proporcionan información sobre las relaciones que existen entre los datos con sus respectivas características.

Cuadro 5. Diferencia entre aprendizaje supervisado y no supervisado

Aprendizaje supervisado	Aprendizaje no supervisado
Predice (comportamientos y preferencias).	Captura y analiza.
Es predictivo.	Es descriptivo.
Trata sobre el futuro.	Trata sobre el pasado.
Resuelve interrogantes como: - ¿Qué clientes están por comprar? - ¿Qué clientes pagarán?	Resuelve interrogantes como: - ¿Cuántos productos se vendieron? - ¿Cuál es el porcentaje de morosidad?

Fuente: Referencia [34]

Las interrogantes que pueden responder las técnicas descriptivas ante las tareas de un negocio, pueden ser, por ejemplo ¿Qué clientes que comprar pan suelen comprar mantequilla?, ¿El tabaco y las drogas son los factores más relevantes para que ocasione X enfermedad?, ¿Qué grupos de clientes componen las ventas en la sucursal 1?, ¿Cómo puede categorizar a mis empleados?, entre otras. El aprendizaje no

supervisado, aplicándolo a una interrogante en particular, la categorización de los empleados, supongamos que se tiene un registro de todos los empleados con sus respectivos datos personales, domiciliarios, legales, se requiere que la técnica a utilizar pueda segmentar todos esos datos y darse cuenta por sí solo cuales son las relaciones entre cada empleado, creando grupos que se ajusten a la información que se encuentre en la base de datos. Una vez que finalice el proceso de segmentación, se podrá tener el número de grupos que fueron descubiertos, por ejemplo, supongamos que fueron tres, un grupo puede constar de empleados que no tengan hijos, vivan en casa de alquiler y no tengan vehículo, otro grupo puede manifestar los empleados que tienen hijos, casa y carro propio, por último los empleados que no tengan hijos, tengan casa propia y no cuenten con un vehículo.

Son varios algoritmos que existen para las técnicas descriptivas entre ellos el asociativo, que permite establecer posibles relaciones entre datos o sucesos que a simple vista se ven independientes; como también puede reconocer como la acción de algún suceso tenga como consecuencia la aparición de otro. Otro algoritmo muy ligado es el agrupamiento o clusterización, es una tarea descriptiva común donde se busca identificar un conjunto finito de categorías o grupos para describir los datos [35], permite la identificación de grupos, en donde los elementos tienen una gran similitud entre sí y se diferencian mucho de los demás.

En conclusión podemos decir que, a diferencia de las técnicas predictivas descritas anteriormente, el aprendizaje supervisado permite reconocer patrones entre los datos sin previamente tener un modelo de entrenamiento, para que se puedan relacionar los elementos entre sí, agruparse, buscar si existen dependencia entre cada uno, y muchas cosas más. Con el conocimiento obtenido se pueden tomar acciones que beneficien las actividades diarias del negocio.

2.2.6 *Internet de las cosas.* ha venido a tener más repercusiones en los últimos años, se considera como una tendencia debido a que ha revolucionado el mundo de las comunicaciones y la electrónica digital [36]. Es necesario que las organizaciones empiecen desde ya a incursionar en las tecnologías que puede abarcar el internet de las cosas, porque va a quedarse y consolidarse como una alternativa para mejorar de una u otra manera la producción de las actividades en el mercado.

En estos momentos, con la demanda de la telefonía móvil, tenemos más facilidades de realizar tareas que a diario facilitan nuestras vidas, como por ejemplo comprar algún

producto en internet, bloquear un vehículo desde el trabajo, encender el aire acondicionado, entre otros. El internet de las cosas permite realizar todas esas actividades, mantenernos enterados de lo que sucede a nuestro alrededor, comunicar cualquier evento de gran relevancia, emergencias, y así. Pero el internet va mucho más allá de las simples cosas cotidianas, abarca muchas áreas en las que es muy utilizado. La podemos denominar como una red global de información y comunicación, en donde los dispositivos que tenemos en nuestro alrededor se encuentran conectados permanentemente al internet [37]; la información que se genera mediante sensores y actuadores permite la automatización y procesamiento de las actividades, permite analizar qué datos serán más útiles que otros, que servirán para más adelante darles un buen uso.

Una de las características de la aplicación del internet de las cosas es la información, con características de ser precisa, automatizada y en tiempo real, por lo cual la interpretación de los datos recabados dará lugar a la implementación de servicios nuevos o mejorar los que ya se tengan en el negocio, la conclusión de todo esto, es el impacto en la parte económica y social.

El internet de las cosas hoy en día abarca un sinnúmero de áreas en donde se puede aplicar todo lo que se ha explicado. En las más populares podemos encontrar casas inteligentes, también llamadas Smart Home, ciudades inteligentes, aplicaciones enfocadas a la medicina, química, ingeniería, agricultura, entre muchas.

Un casa inteligente o llamada Smart Home, se refiere a la facilidad con que una persona realiza sus actividades cotidianas dentro de cualquier parte de su casa, podemos mencionar ejemplos como el control de la temperatura, que ésta se ajuste dependiendo del ambiente en el que se encuentre, podemos tener una refrigeradora inteligente, la cual de aviso cuando algún alimento se esté agotando y permita su compra en cualquier tienda por medio del internet. Estas y otras acciones son automatizadas para gusto del usuario y en todo momento se busca la comodidad de las personas con ayuda de la tecnología.

Las ciudades inteligentes (Smart City) [38] tienen el objetivo de facilitar la vida de los ciudadanos, ya sea contando con un parqueadero inteligente en donde una red de sensores indiquen si hay algún lugar para un vehículo, prevenir a los conductores sobre el tráfico en tiempo real y brindar rutas alternativas para que evite un congestionamiento vehicular. Las ciudades inteligentes en un futuro ayudarán a impulsar el crecimiento

económico sostenible, brindar prosperidad para así ayudar al buen vivir de los habitantes.

Figura 7. Aplicaciones IOT



Fuente: Referencia [39]

A todo este ambiente de tecnología dentro de las tareas cotidianas de las personas, se puede catalogar como computación ubicua, la cual consiste en la integración de la informática en el entorno de la persona, de forma que los ordenadores no se perciban como objetos diferenciados [40]

En el campo de la agricultura, el internet de las cosas tiene poca incursión, ya que existen pocas empresas que se dedican a trabajar en esta área. Una "Agricultura inteligente" permite tener un control exacto de cultivos sea cual sea la especie que se siembre, permite el monitoreo constante, manejo automático del riego, identificación y tratamiento de los tipos de suelo, entre otras, por medio de plataformas que permitan la toma de decisiones. El futuro de la agricultura inteligente será muy importante en las regiones en donde la agricultura es el principal sector para ingresos monetarios.

2.2.7 *Agricultura de precisión.* La urgente necesidad de aumentar la producción de la agricultura, sobre todo en una pieza cada vez menor de tierra, así como la reducción del consumo de recursos [41] requiere del uso de nuevas tecnologías y métodos de primerísima prioridad. Durante los últimos años se han producido notables avances tecnológicos para el uso de técnicas de control y gestión en este campo, una de ellas es la agricultura de precisión.

Si bien es cierto el costo para mantener un cultivo sano, estable, además de la adquisición de equipos para las fuentes de riego, fertilización del suelo suele ser elevado pero necesario para llevar a cabo esta actividad; el objetivo de la agricultura inteligente o denominada comúnmente agricultura de precisión es optimizar los recursos aplicados a cada parcela, con ello se reducirá el uso en cantidad de cada elemento y los beneficios que se pueden ver como resultado es la calidad de los productos, ahorro de dinero y reducir el impacto hacia el medio ambiente.

El concepto de agricultura de precisión se lo define como el conjunto de técnicas de cultivo que utilizan tecnologías de la información para ajustar el uso de semillas y de agroquímicos considerando la diversidad del medio físico y del medio biológico [42]. En resumen es el uso de tecnologías que ayuden a controlar un cultivo, mediante el uso de sensores, bombas, actuadores pueden trabajar de manera coordinada y comunicarse entre sí para solventar una necesidad que conlleva un exacto manejo de las explotaciones agrícolas.

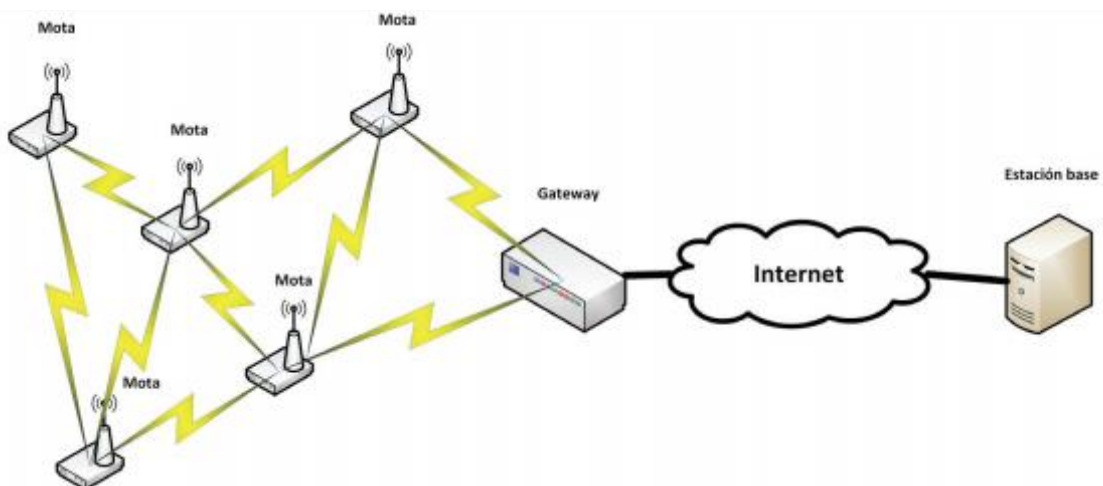
Los sensores remotos han tenido un rol importante en la clasificación de cultivos, sanidad vegetal y evaluación de rendimientos [43]. Dependiendo del área geográfica, tipos de suelo, condiciones climáticas o de la diversidad de los cultivos, se aplica un tipo de sensor programados de tal forma que cumplan con óptimo rendimiento, por ejemplo, se puede enfocar una red de sensores exclusivamente en algún sector de los ya nombrados anteriormente para que cumpla con la disponibilidad del agua para su posterior riego, lectura de la salinidad del suelo, informar sobre la cantidad de nutrientes que pueden absorber los cultivos, comparación de datos acerca de un cultivo con otro para conocer si existe alguna enfermedad [44] o alguna plaga que esté consumiendo parte de la parcela, entre otros. El sector de la agricultura llegará a ser más productiva si se aplican todos los conceptos de IOT, como ya se dijo desde el principio, busca optimizar los recursos que se administran en cada parcela para de esta manera permite la eficiente explotación de las áreas agrícolas y generar más productividad para la organización.

2.2.8 *Red de sensores inalámbricos.* Los sensores han venido evolucionando de manera considerable, teniendo características que cubren las necesidades de la comunicación de una forma flexible, bajo costo, con una vital útil mayor. Una red de sensores inalámbricos se compone de dispositivos que se encuentran ubicados estratégicamente de manera distribuida, en donde se utiliza sensores para monitorear aspectos ambientales o físicos de un determinado lugar [45].

Las tecnologías de comunicación que una red de sensores puede implementar son varias, pero las más usadas en el mundo son las siguientes:

- Wi-Fi, cuyo estándar de comunicación es muy utilizado por la mayoría de computadoras y teléfonos inteligentes.
- Bluetooth, permite realizar transmisión de voz y datos entre teléfonos móviles.
- Zigbee, netamente desarrollado para una WSN y es muy utilizado debido a su costo y menor consumo de energía.
- Tecnologías inalámbricas sub-Ghz, al igual que la tecnología Zigbee es usada exclusivamente en red de sensores, aunque más potente. [46]

Figura 8. Red de sensores inalámbricos



Fuente: Referencia [47]

Las aplicaciones que se le dan a las redes de sensores en muchas áreas como son salud, servicios básicos, agricultura, trabajos civiles. Independientemente del estándar de comunicación que tenga una red de sensores, su funcionamiento es el mismo, el cual permite realizar lecturas, inspecciones a ciertas variables que se encuentren en un determinada lugar (temperatura, movimientos, humedad, luz) y con los datos que se

retoman, los sensores por sí pueden enviarlos hacia distintas fuentes de información, aplicaciones, sistemas; ya con todo eso, existen reglas que permiten realizar cierta acción a otros dispositivos llamados actuadores, los que ejecutan las instrucciones que le fueron enviadas para realizar, por ejemplo, dejar pasar agua para el riego de un cultivo, apagar una bomba dependiendo de la humedad que el sensor haya obtenido del suelo, entre muchas cosas.

Todo el funcionamiento en una arquitectura IOT permite controlar cualquier actividad que ocurra en un negocio, por medio de una computadora, celular o cualquier otra interfaz. Al final se podrá realizar un análisis de los datos de interés, pudiendo así llegar a tomar las mejores decisiones para preservar la producción.

2.2.9 Proyecto IOTMACH. Es un proyecto integrador de carácter investigativo enfocado principalmente en la solución de problemas relacionados con el internet de las cosas, red de sensores inalámbricos y la agricultura de precisión. Es un trabajo que nació desde las aulas, con estudiantes comprometidos día a día con todo lo que hacen; el objetivo es la implementación de un sistema robusto, fácil de manejar, amigable y con muchas herramientas que permiten la gestión de la información, brindando seguridad, confiabilidad, calidad.

Las tecnologías que se están usando son estándares a nivel mundial que se utilizan en este campo, es lo que la comunidad y desarrolladores recomiendan a los demás usuarios, para hacer cada día más sencillo comunicarse desde cualquier parte que nos encontremos.

2.3 Objetivos del prototipo

2.3.1 Objetivo general

Implementar un módulo de análisis estadístico y predictivo para agricultura utilizando Big Data y Machine Learning integrado al sistema IOTMACH, utilizando R como lenguaje de programación, con el propósito de contar con una herramienta útil para la toma de decisiones.

2.3.2 *Objetivos específicos*

- Analizar las diferentes técnicas de minería de datos para resolver problemas más comunes que se pueden presentar en la agricultura.
- Desarrollar modelos de aprendizaje supervisado y no supervisado en Machine Learning para el análisis predictivo y descriptivo.
- Diseñar una interfaz web que permita la fácil utilización de la herramienta de análisis.
- Realizar pruebas unitarias conforme a las etapas que el desarrollo del software lo requiera.
- Evaluar la calidad de la aplicación con pruebas de seguridad, rendimiento, integración y pruebas totales del sistema.

2.4 **Requisitos, análisis y diseño**

2.4.1 *Requisitos funcionales.* Se declararán los servicios que la aplicación web realizará y en donde se describe el comportamiento de las salidas como producto de entradas recibidas.

Cuadro 6. Requisito funcional REQ_001

Cód.	REQ_001
Nombre	Gestión de conexiones
Descripción	El usuario deberá ingresar los datos en los campos que la aplicación solicite.
Requisitos asociados	NINGUNO
Tipo	Funcional
Condición	NINGUNA
Observaciones	El sistema registrará la conexión únicamente cuando los datos ingresados fueron correctos, es decir, se probará la conexión y se mostrará un mensaje del resultado de dicha operación.

Fuente: Elaboración propia

Cuadro 7. Requisito funcional REQ_002

Cód.	REQ_002
Nombre	Gestión de repositorio local
Descripción	El usuario podrá subir archivos locales desde su ordenador cuando lo crea necesario, para ello deberá llenar los campos requeridos.
Requisitos asociados	NINGUNO
Tipo	Funcional
Condición	NINGUNA
Observaciones	Se podrá subir archivos con distintos formatos, no hay restricción en cuanto al tamaño y los archivos se guardarán en un directorio el cual podrá ser accedido por el usuario cuando lo crea conveniente.

Fuente: Elaboración propia

Cuadro 8. Requisito funcional REQ_003

Cód.	REQ_003
Nombre	Realizar análisis descriptivo
Descripción	El usuario deberá cargar un archivo o referencia a una conexión en donde se encuentre la información que se va a someter a un análisis descriptivo y en donde podrá encontrar técnicas para hacerlo, las cuales son: clusterización, asociación y análisis de componentes principales.
Requisitos asociados	REQ_001 y REQ_002
Tipo	Funcional
Condición	Para acceder a las funcionalidades de este requisito, se deberá primeramente haber creado una conexión o gestionado un archivo en el repositorio local.
Observaciones	Para visualizar el resultado, detalles y ejecución del análisis, el usuario deberá completar los pasos que se encuentran en cada pestaña de la interfaz web. Se presentarán alertas para notificar mensajes en el caso de no cumplir con lo requerido por el sistema.

Fuente: Elaboración propia

Cuadro 9. Requisito funcional REQ_004

Cód.	REQ_004
Nombre	Realizar análisis predictivo
Descripción	El usuario deberá cargar un archivo o referencia a una conexión en donde se encuentre la información que se va a someter a un análisis predictivo y en donde podrá encontrar técnicas para hacerlo, las cuales son: regresión lineal, clasificación y predicciones.
Requisitos asociados	REQ_001 y REQ_002
Tipo	Funcional
Condición	Para acceder a las funcionalidades de este requisito, se deberá primeramente haber creado una conexión o gestionado un archivo en el repositorio local.
Observaciones	Para visualizar el resultado, detalles y ejecución del análisis, el usuario deberá completar los pasos que se encuentran en cada pestaña de la interfaz web. Se presentarán alertas para notificar mensajes en el caso de no cumplir con lo requerido por el sistema.

Fuente: Elaboración propia

Cuadro 10. Requisito funcional REQ_005

Cód.	REQ_005
Nombre	Manipulación de gráficos estadísticos
Descripción	El usuario deberá cargar un archivo o referencia a una conexión en donde se encuentre la información, el sistema permitirá la selección de los campos a utilizar, así también como los tipos de gráficos que se deseen mostrar en el análisis estadístico.
Requisitos asociados	REQ_001 y REQ_002
Tipo	Funcional
Condición	Para acceder a las funcionalidades de este requisito, se deberá primeramente haber creado una conexión o gestionado un archivo en el repositorio local.
Observaciones	Se deben utilizar archivos válidos y con los formatos soportados para el conjunto de datos con el fin de poder observar los resultados.

Fuente: Elaboración propia

Cuadro 11. Requisito funcional REQ_006

Cód.	REQ_006
Nombre	Refrescar/Actualizar
Descripción	El usuario deberá elegir la opción en del menú de la interfaz web.
Requisitos asociados	NINGUNO
Tipo	Funcional
Condición	NINGUNA
Observaciones	En caso de que hubiera información con la que se estuviera trabajando, al realizar esta opción, dicha información desaparecerá y se cargarán los valores que se encuentran por defecto al iniciarse la interfaz.

Fuente: Elaboración propia

2.4.2 *Requisitos no funcionales.* Se detallan las características y ámbito que posee la aplicación web, entre las cuales se detallan las siguiente:

2.4.2.1 *Mantenibilidad.* La manera en la que se encuentra estructurada la aplicación web va a permitir adaptarse a los cambios que en un futuro puedan suceder en los módulos desarrollados. De igual forma, el código con el que se programó la aplicación, está documentado para el fácil entendimiento por parte de los programadores que quieran realizar alguna nueva versión del prototipo.

Se garantiza también que los cambios que se efectúen en un determinado módulo o componente, éste no afectará el correcto funcionamiento de los demás submenús de la aplicación web.

2.4.2.2 *Usabilidad.* Este requisito está ambientado a la fácil utilización de la aplicación por parte de los usuarios, que, con o sin mucha experiencia en informática van a poder manejar la herramienta de manera sencilla.

Los menús y paneles se encuentran distribuidos de tal manera que se pueda acceder rápidamente a una determinada tarea. En cada submenú se encuentran textos de ayuda que permitirán al usuario realizar sin ningún inconveniente análisis con los datos que

haya seleccionado anteriormente. La interfaz web posee una estructura y colores llamativos con funcionamiento interactivo en cada acción que se vaya a ejecutar.

2.4.2.3 *Confiabledad.* El desempeño y tiempos de respuesta que se generan en cada ejecución serán muy buenos, debido a que si se trabaja con gran cantidad de información, el sistema responderá todas las peticiones de manera rápida, obteniendo así los resultados esperados. Los datos que se almacenen dentro de la aplicación se mantendrán disponibles para cuando el usuario decida acceder a ello.

2.4.2.4 *Escalabilidad.* La aplicación podrá conectarse con distintos sistemas de bases de datos o archivos que se encuentren alojados en la nube. Permite también trabajar con distintos formatos de archivos y, a medida que los datos crecen, no se notará cambios en cuanto a los tiempos de ejecución o respuesta.

2.4.3 *Submódulos implementados en la interfaz web.* Los submódulos que se han implementado en la interfaz web son los siguientes:

- **Submódulo administración de conexiones:** se encargará de gestionar las conexiones, es decir los archivos o tablas de una base de datos. Para ello se requerirá que el usuario ingrese los datos de conexión tales como: nombre de conexión, nombre de base de datos, nombre o dirección del host, puerto, nombre de usuario, y contraseña (en caso de seleccionar una conexión a base de datos). Los datos que se deben ingresar para la conexión a archivos son: nombre de conexión, nombre del archivo, nombre del directorio en donde se decida alojar dichos archivos.

Es importante recalcar que para crear una conexión se debe verificar que los datos ingresados sean correctos, se puede realizar esta acción gracias a una opción que permite probar la conexión antes de crearla.

- **Submódulo análisis descriptivo:** permitirá realizar análisis de los datos con técnicas de clusterización, reglas de asociación y análisis de componentes principales. Cabe recalcar que los resultados obtenidos de este análisis se podrán visualizar en las pestañas de: resultados, detalles de ejecución, estadísticas y detalles de los modelos que se han generado.

- Submódulo análisis predictivo permitirá realizar análisis de los datos con técnicas de regresión lineal, predicción y clasificación. De igual manera que el submódulo anterior, los resultados obtenidos de este análisis se podrán visualizar en las pestañas de: resultados, detalles de ejecución, estadísticas y detalles de los modelos que se han generado.

En los dos submódulos anteriores, se localiza la pestaña principal orígenes de datos y es en dónde se encuentran organizadas todas las conexiones o archivos que han sido creados por el usuario.

- Submódulo estadísticas, se encarga de mostrar los datos en forma de gráficos estadísticos, lo mismos que van a permitir al usuario analizar la información generada para que pueda realizar acciones para la toma de decisiones o resolver alguna necesidad que haya sido encontrada.

2.4.4 *Suposiciones y restricciones.* Para la realización del prototipo se han tenido que tomar algunas consideraciones, como son las siguientes:

2.4.4.1 *Suposiciones.* Para el prototipo:

- Se asume que el equipo en donde se va a almacenar y ejecutar la aplicación cumple con los requisitos necesarios para su correcto funcionamiento.
- La aplicación será utilizada por personas con conocimientos básicos para este tipo de software.
- Los usuarios utilizarán navegadores actuales que permitan acceder a páginas o aplicaciones web, tales como Google Chrome o Mozilla Firefox.
- La aplicación será utilizada para los propósitos con la que fue desarrollada.

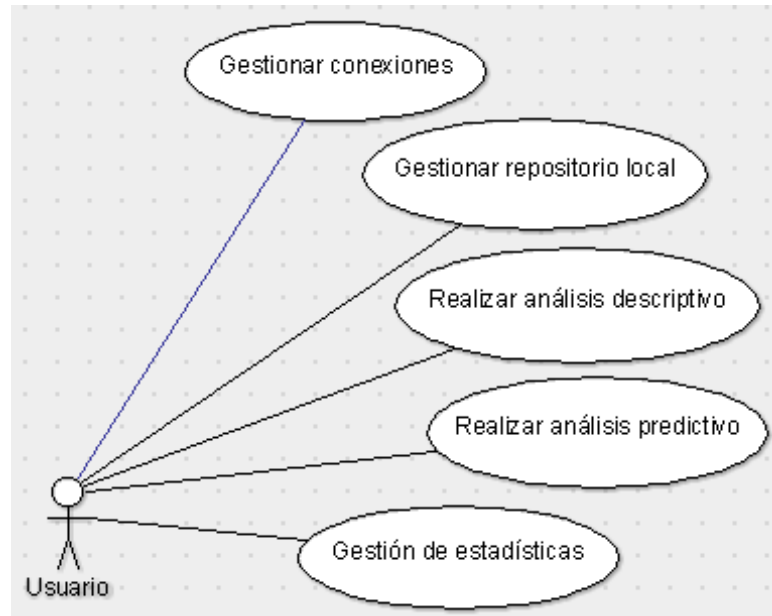
2.4.4.2 *Restricciones.* En cuanto al diseño y posterior integración:

- El framework de desarrollo es Shiny 0.13.2.
- La plataforma de alojamiento de aplicaciones de este tipo es limitada, debido a que es una versión Open Source.

2.4.5 *Análisis y diseño*. Se detallarán las funciones del sistema.

2.4.5.1 *Diagramas de casos de uso*. Representan las funciones de la aplicación.

Figura 9. Diagrama de caso de uso general

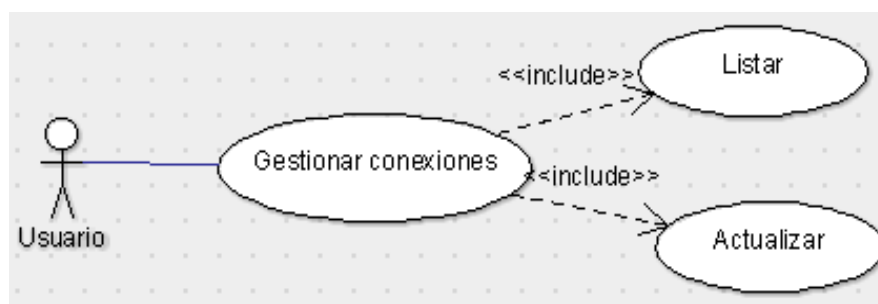


Fuente: Elaboración propia

A continuación se detallarán cada uno de los casos de uso en los que interviene el usuario dentro de la aplicación:

- Gestionar conexiones. Permite conectarse a las distintas fuentes de datos.

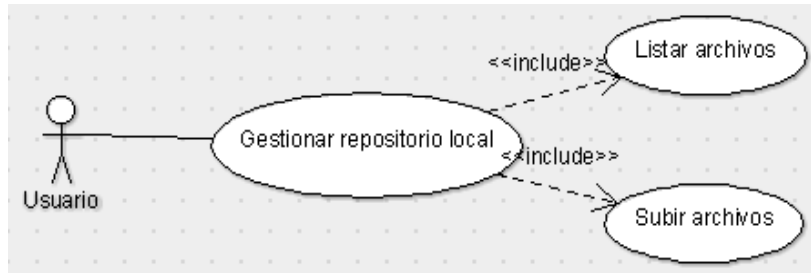
Figura 10. Caso de uso gestionar conexiones



Fuente: Elaboración propia

- Gestionar repositorio local. Permite almacenar los archivos que el usuario suba.

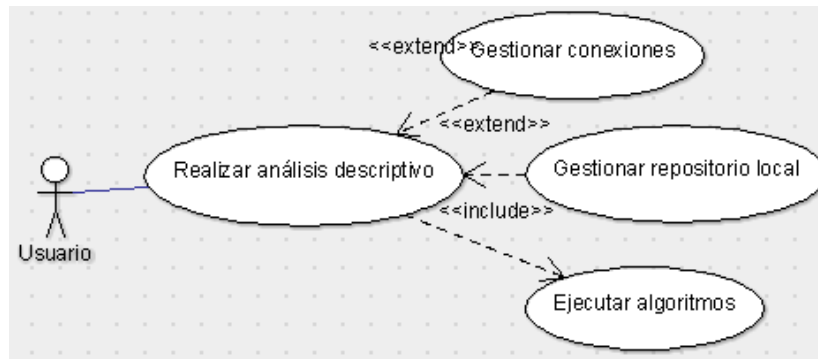
Figura 11. Caso de uso gestionar repositorio local



Fuente: Elaboración propia

- Realizar análisis descriptivo. Permite realizar análisis descriptivo de los datos.

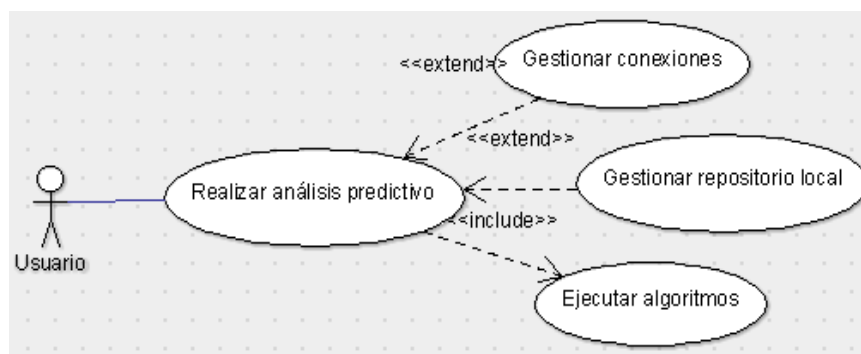
Figura 12. Caso de uso realizar análisis descriptivo



Fuente: Elaboración propia

- Realizar análisis predictivo. Permite realizar análisis predictivo de los datos.

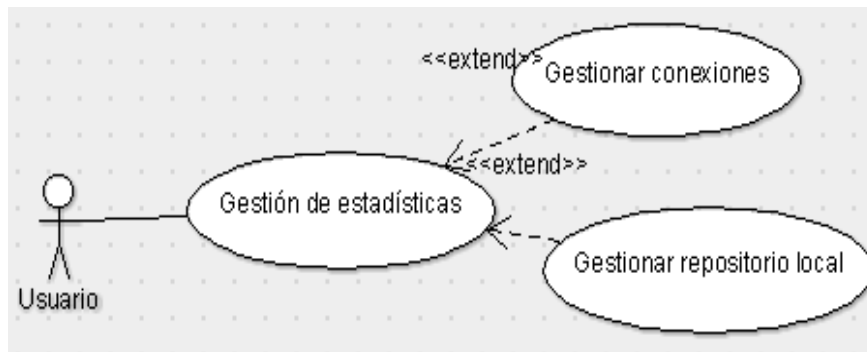
Figura 13. Caso de uso realizar análisis predictivo



Fuente: Elaboración propia

- Gestión de estadísticas. Permite realizar estadísticas y gráficas de los datos.

Figura 14. Caso de uso gestión de estadísticas



Fuente: Elaboración propia

2.4.5.2 Descripción de los casos de uso. Se detalla cada caso de uso.

Cuadro 12. Caso de uso CU_001

Cód.: CU_001	
Caso de uso	Nombre: Gestionar conexiones
Descripción	Conecta los diferentes orígenes de datos establecidos con la aplicación para su posterior análisis.
Actor primario	Usuario
Actor secundario	
Supuestos	Se cuentan con los datos necesarios para realizar la conexión a una determinada fuente de datos.
Recursos	
Pasos	<ol style="list-style-type: none"> 1. El usuario ingresará los datos solicitados 2. Se probará si la información es correcta mediante un test de conexión 3. Si la información es válida, se procederá a guardar dichos datos 4. En caso de no ser válidos los datos, se mostrará un mensaje notificándolo 5. Se enlistarán todas las conexiones creadas por el usuario en la respectiva tabla de visualización.
Requisitos no funcionales	Tiempos de respuesta cortos para realizar la conexión a distintas fuentes de datos.

Fuente: Elaboración propia

Cuadro 13. Caso de uso CU_002

Cód.: CU_002	
Caso de uso	Nombre: Gestionar repositorio local
Descripción	Aloja diferentes tipos de archivo en un directorio dentro de la aplicación para su posterior análisis.
Actor primario	Usuario
Actor secundario	
Supuestos	Se cuentan con los datos necesarios para subir el archivo al directorio de la aplicación.
Recursos	
Pasos	<ol style="list-style-type: none"> 1. El usuario ingresará los datos solicitados 2. Se validará la información 3. Si la información es válida, se procederá a almacenar el archivo en el repositorio de la aplicación 4. En caso de no ser válidos los datos, se mostrará un mensaje notificándolo 5. Se creará una vista previa de la fuente de datos subida
Requisitos no funcionales	Tiempos de respuesta cortos para realizar la conexión a distintas fuentes de datos.

Fuente: Elaboración propia

Cuadro 14. Caso de uso CU_003

Cód.: CU_003	
Caso de uso	Nombre: Realizar análisis descriptivo
Descripción	Análisis de los datos de un dataset que permita identificar grupos, comportamientos o asociaciones entre los diferentes campos que componen dicho dataset.
Actor primario	Usuario
Actor secundario	
Supuestos	Se cuentan con los datos necesarios para aplicar los algoritmos descriptivos y realizar la extracción del conocimiento.
Recursos	Datos de un determinado dataset

Cuadro 14. (Continuación)

Cód.: CU_003	
Caso de uso	Nombre: Realizar análisis descriptivo
Pasos	<ol style="list-style-type: none"> 1. Se tendrá que especificar el algoritmo que se empleará 2. El usuario deberá seleccionar el dataset que utilizará para realizar el análisis 3. El usuario elegirá los campo con los cuales trabajará 4. El sistema analizará la estructura del dataset y aplicará el algoritmo que se ha seleccionado 5. En caso de que la información sea correcta, se visualizarán los resultados del análisis 6. En caso de existir inconsistencias en el dataset, se informará de lo ocurrido por medio de un mensaje 7. El usuario puede acceder a la ayuda que se encuentra ubicada en la interfaz de cada algoritmo
Requisitos no funcionales	Requisitos de confiabilidad, debido a qué al aplicar un algoritmo se debe presentar los resultados de la forma más óptima posible.

Fuente: Elaboración propia

Cuadro 15. Caso de uso CU_004

Cód.: CU_004	
Caso de uso	Nombre: Realizar análisis descriptivo
Descripción	Análisis de los datos de un dataset que permita realizar regresiones, aplicar clasificaciones y predecir valores para datos nuevos.
Actor primario	Usuario
Actor secundario	
Supuestos	Se cuentan con los datos necesarios para aplicar los algoritmos y realizar las predicciones de forma precisa.
Recursos	Datos de un determinado dataset
Pasos	<ol style="list-style-type: none"> 1. Se tendrá que especificar el algoritmo que se empleará 2. El usuario deberá seleccionar el dataset que utilizará para realizar el análisis 3. El usuario elegirá los campo con los cuales trabajará 4. El sistema analizará la estructura del dataset y aplicará el algoritmo que se ha seleccionado

Cuadro 15. (Continuación)

Cód.: CU_004	
Caso de uso	Nombre: Realizar análisis descriptivo
Pasos	<ol style="list-style-type: none"> 5. En caso de que la información sea correcta, se visualizarán los resultados del análisis 6. En caso de existir inconsistencias en el dataset, se informará de lo ocurrido por medio de un mensaje 7. El usuario puede acceder a la ayuda que se encuentra ubicada en la interfaz de cada algoritmo
Requisitos no funcionales	Requisitos de confiabilidad, debido a que al aplicar un algoritmo se debe presentar los resultados de la forma más óptima posible.

Fuente: Elaboración propia

Cuadro 16. Caso de uso CU_005

Cód.: CU_005	
Caso de uso	Nombre: Gestión de estadísticas
Descripción	Solicita la visualización de los resultados que se han generado a partir de un dataset.
Actor primario	Usuario
Actor secundario	
Supuestos	Se cuentan con los datos necesarios para realizar los gráficos estadísticos.
Recursos	
Pasos	<ol style="list-style-type: none"> 1. El usuario elegirá el dataset para realizar el análisis 2. Seleccionará el tipo de gráfico que desea visualizar 3. Se validará la estructura del dataset 4. Se procederá a presentar los resultados para el respectivo análisis por parte del usuario
Requisitos no funcionales	Adaptabilidad, debido a que los resultados se deben mostrar de la manera más comprensible.

Fuente: Elaboración propia

2.4.5.3 *Interfaces de usuario*. Permite la interacción entre el usuario y las funciones de la aplicación.

- Administración de conexiones. Bosquejo de los componentes del formulario.

Figura 15. Interfaz administración de conexiones

IOTMach Analysis

Adm. Conexiones

- Análisis descriptivo
- Análisis predictivo
- Estadísticas
- Actualizar
- Acerca de

[Nueva conexión](#)

Tipo de conexión

Nombre conexión Combo Box

Nombre host Puerto

Nombre base de datos

Nombre usuario Contraseña

Probar

Guardar

[Todas las conexiones](#)

Colum...	Column 2
Content 1	Content 2
Content 3	Content 4

Fuente: Elaboración propia

- Repositorio local. Bosquejo del panel para subir archivos localmente.

Figura 16. Interfaz repositorio local

IOTMach Analysis

Rep. Local

- Adm. Conexiones
- Análisis descriptivo
- Análisis predictivo
- Estadísticas
- Actualizar
- Acerca de

Origenes de datos Ayuda

[Fuente de datos](#)

Subir archivo

Examinar ...

Nueva carpeta

Nombre carpeta

Nombre archivo Extensión

Guardar

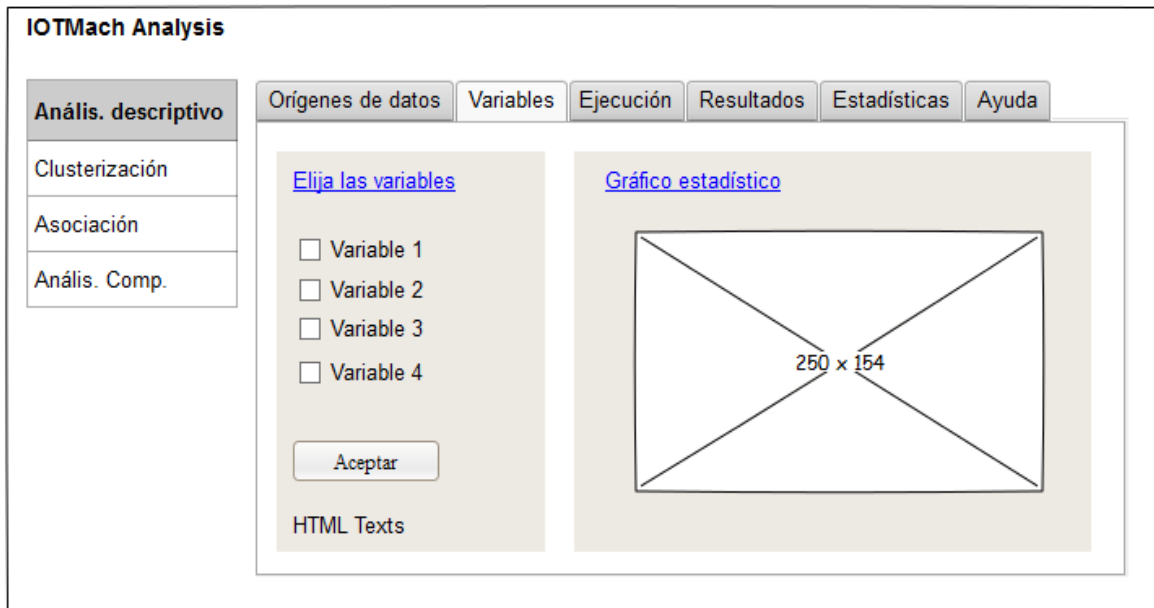
[Vista previa](#)

<input type="checkbox"/>	Column 2
<input checked="" type="checkbox"/>	Cell Content 1
<input type="checkbox"/>	Cell content 2
<input type="checkbox"/>	Cell content 3
<input type="checkbox"/>	Cell content 4

Fuente: Elaboración propia

- Análisis descriptivo. Bosquejo de las opciones para realizar análisis descriptivo.

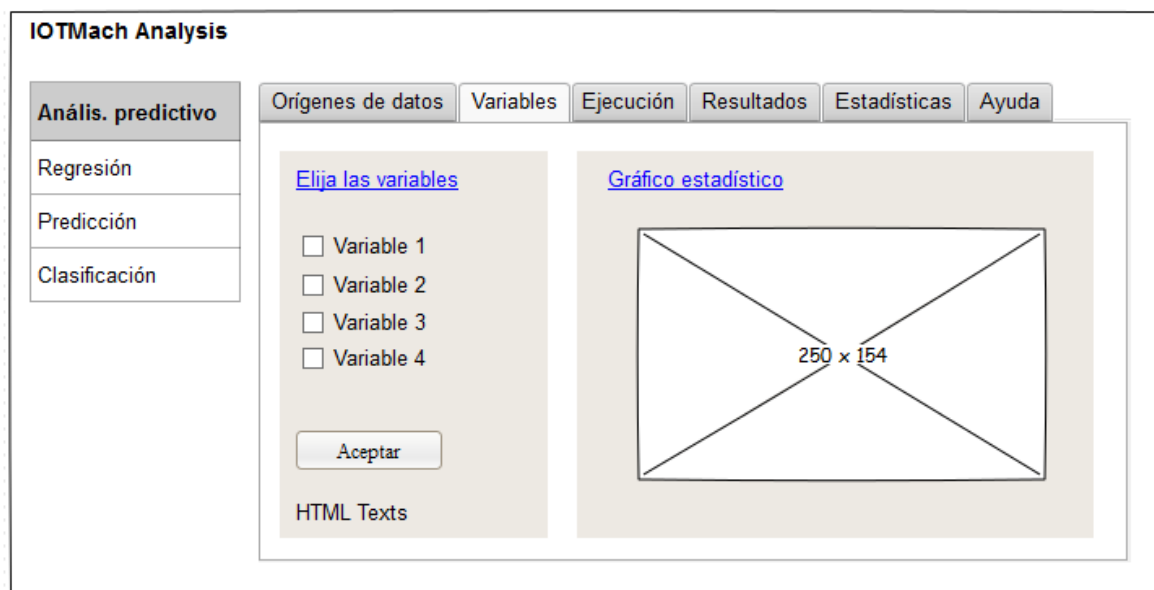
Figura 17. Interfaz análisis descriptivo



Fuente: Elaboración propia

- Análisis predictivo. Bosquejo de las opciones para realizar análisis predictivo.

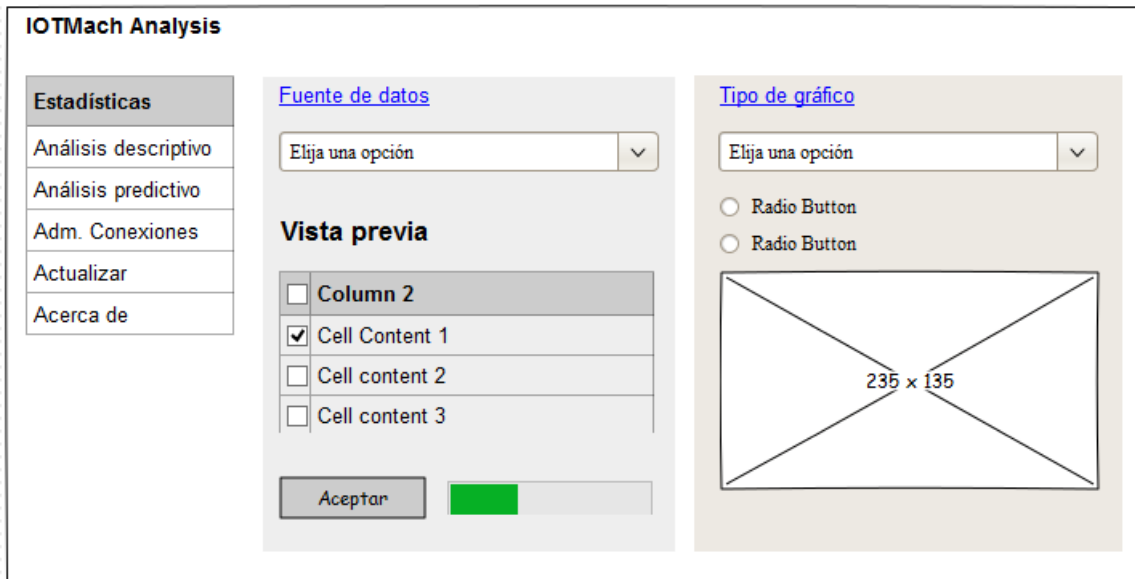
Figura 18. Interfaz análisis predictivo



Fuente: Elaboración propia

- Estadísticas. Bosquejo de los componentes para realizar estadísticas a los datos.

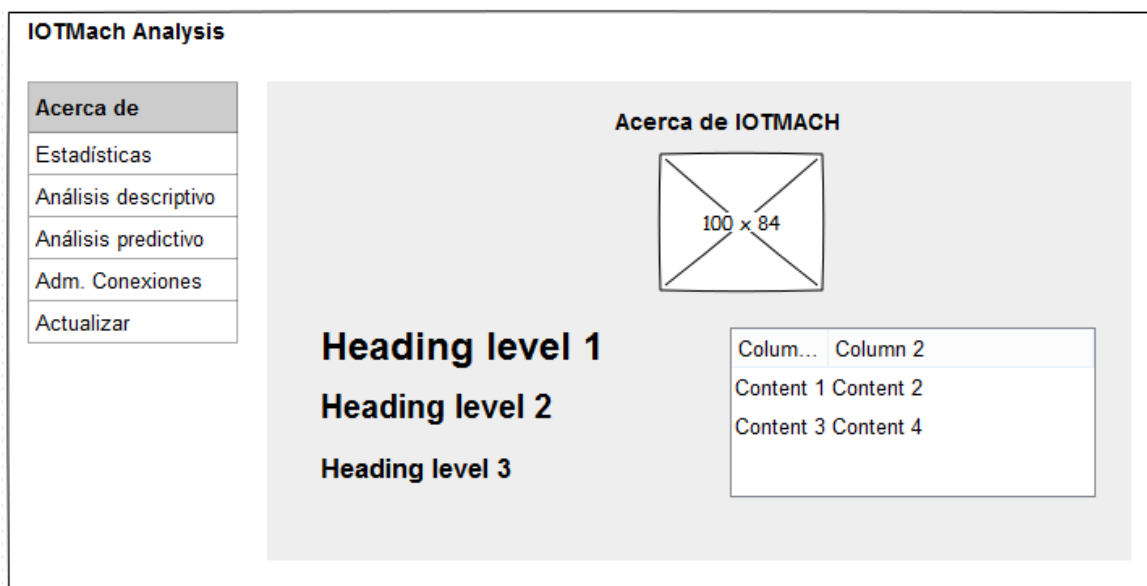
Figura 19. Interfaz de estadísticas



Fuente: Elaboración propia

- Acerca de. Bosquejo de la estructura para mostrar información de la aplicación.

Figura 20. Interfaz acerca de



Fuente: Elaboración propia

2.5 Desarrollo e implementación del prototipo

Para el diseño y desarrollo de la aplicación IOTMach Analysis se necesitó de una serie de recursos para obtener los resultados esperados, los cuales se detallarán a continuación:

2.5.1 *Lenguaje de programación R.* Es un lenguaje de programación que se utiliza para el análisis estadístico, los cálculos y en el aspecto gráfico [48]. Fue desarrollado por Robert Gentleman y Ross Ihaka Universidad de Auckland, en el año de 1993; el nombre con que oficialmente se lo conoce es R Project [6]

Es tanto un lenguaje de programación como un entorno de para desarrollo de software de código abierto para problemas en el área de computación y resultados en el plano estadístico. Este lenguaje provee innumerables técnicas para la simulación, construcción de modelos lineales y no lineales, realizar pruebas estadísticas tradicionales, análisis de series temporales, clasificación, agrupación en clústeres, predicciones, entre otros.

Figura 21. Logo R Project



Fuente: Referencia [6]

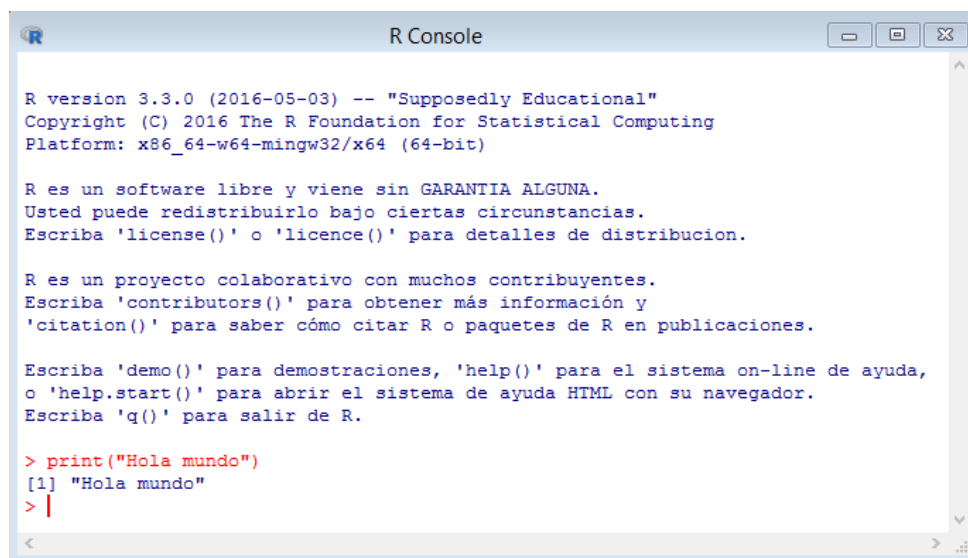
La programación se la realiza en líneas de comando, las cuales pueden ser ejecutadas por medio de una interfaz gráfica llamada Rcommander, aunque también existe un IDE llamado RStudio, el cual proporciona una potente interfaz de usuario para diseño de aplicaciones en R.

Es muy utilizado en el área de la minería de datos, debido a que posee múltiples métodos y técnicas para llevar ese propósito. La versión más reciente de este software es la 3.3.1 (actualizado el 21 de junio del 2016). La organización sin fines de lucro que se encuentra

detrás de esta potente herramienta es “The R Foundation”, dicha organización está encargada de brindar apoyo a R, así como también innovaciones en el cálculo estadístico; en la actualidad R se ha convertido en una valiosa herramienta para investigadores, empresas o universidades que la utilizan para el desarrollo de futuras invenciones en software para el área de la estadística computacional [49].

Una de las grandes fortalezas de este lenguaje radica en sus librerías que usuarios de la comunidad desarrollan día tras día, permitiendo reducir tiempo y código para la construcción de aplicaciones de minería de datos, estadísticas o cualquier problema relacionado a este campo. De igual manera existen interfaces que permiten la comunicación entre otros lenguajes, controladores de bases de datos, archivos en la nube, librerías para la graficación, en donde se pueden generar gráficos de alta calidad; aparte de estas fabulosas características, R también posee su propia documentación y permite de igual forma construir proyectos (llamados paquetes) para que puedan ser usados por otros usuarios.

Figura 22. Hola mundo en R



```
R Console

R version 3.3.0 (2016-05-03) -- "Supposedly Educational"
Copyright (C) 2016 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R es un software libre y viene sin GARANTIA ALGUNA.
Usted puede redistribuirlo bajo ciertas circunstancias.
Escriba 'license()' o 'licence()' para detalles de distribucion.

R es un proyecto colaborativo con muchos contribuyentes.
Escriba 'contributors()' para obtener más información y
'citation()' para saber cómo citar R o paquetes de R en publicaciones.

Escriba 'demo()' para demostraciones, 'help()' para el sistema on-line de ayuda,
o 'help.start()' para abrir el sistema de ayuda HTML con su navegador.
Escriba 'q()' para salir de R.

> print("Hola mundo")
[1] "Hola mundo"
> |
```

Fuente: Elaboración propia

La eficiencia que tiene la programación en R se compara con otros software con similares características tales como GNU Octave, MatLab o Python [50]. Existe un repositorio de los paquetes disponibles para ser usados, el contenido se aloja en su página oficial denominada The Comprehensive R Archive Network (CRAN) [51], dicho de otra manera es una red de servidores FTP y Web de todo el mundo en donde se

almacenan, hasta a la fecha, las versiones de código y la documentación de proyectos realizados en R.

Cuadro 17. Comparación entre R y Python

Parámetro	R	Python
Disponibilidad/Costo	5 de 5	5 de 5
Curva de aprendizaje	2.5 de 5	3.5 de 5
Capacidad de manejo de datos	4 de 5	4 de 5
Capacidades gráficas	4.5 de 5	4 de 5
Avances en herramientas	4.5 de 5	4 de 5
Comunidad	3.5 de 5	3 de 5

Fuente: Elaboración propia

En el análisis se puede observar varios aspectos que tienen los dos lenguajes de programación en donde se puede concluir que R supera a Python en cuanto a sus capacidades gráficas y herramientas funcionales, con una comunidad más extensa pro con una curva de aprendizaje muy baja. El costo de realizar un proyecto en las dos herramientas es muy factible, debido a su disponibilidad la facilidad para el manejo de los datos.

2.5.2 Herramientas y librerías. Para la interacción con el código R se utilizaron paquetes o librerías conjunto con un IDE que agilitó el proceso de construcción de la aplicación, los mismos que se describen en el siguiente apartado.

2.5.2.1 IDE RStudio. Es un entorno de desarrollo integrado (IDE) para R. Incluye una consola, editor de resaltado de sintaxis que soporta la ejecución de código directa, así como herramientas para el trazado, la historia, la depuración y la gestión del espacio de trabajo [52]. Permite acceder de manera fácil a toda la potencia que caracteriza a R, por lo que se debe haber instalado R previamente para el correcto funcionamiento de la herramienta.

Figura 23. Logo de RStudio



Fuente: Referencia [52]

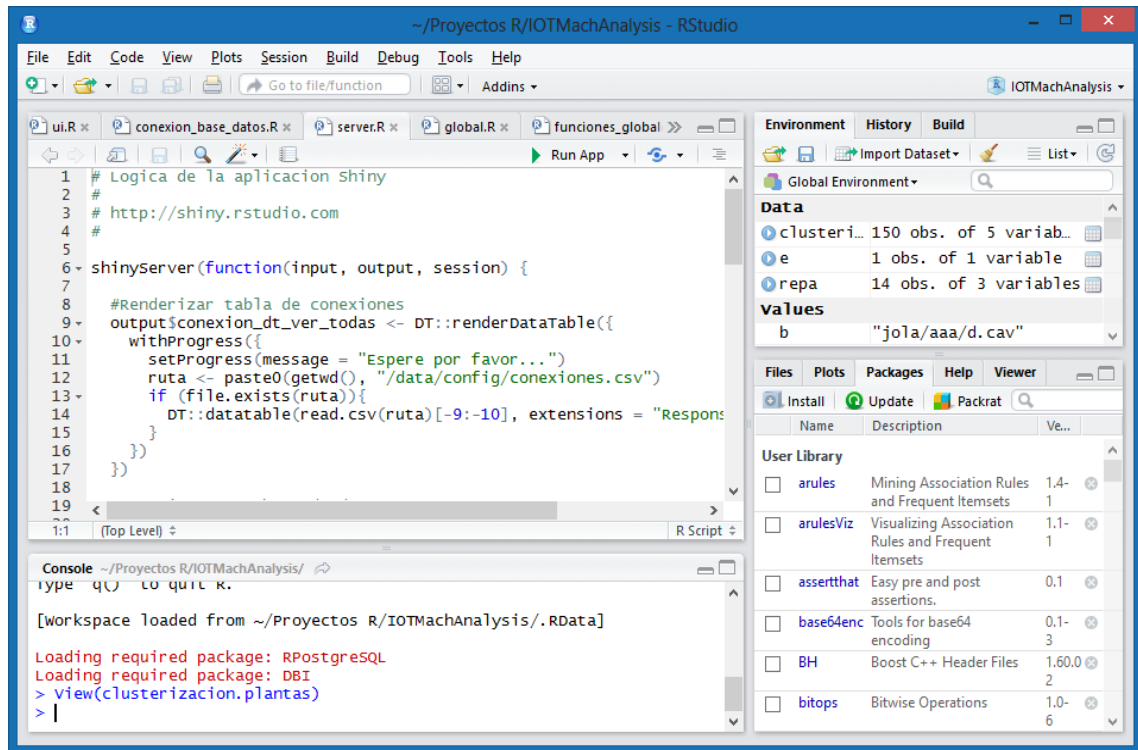
El objetivo de los creadores del IDE RStudio fue desarrollar una herramienta potente que soporte la alta gama de procedimientos y técnicas requeridas para realizar un análisis de calidad con una alta confiabilidad. Del mismo modo, pretenden que RStudio sea lo más sencillo posible, de forma muy intuitiva para proporcionar un entorno de trabajo amigable, tanto para los usuarios catalogados como experimentados como para los nuevos usuarios que pretenden incursionar en el análisis de los datos con R.

Se puede utilizar RStudio porque es una herramienta potente y también porque la comunidad la recomienda, debido a que:

- Respeto la filosofía que tradicionalmente posee la consola de R (RCommander).
- Visualiza todos los objetos dentro del marco de trabajo.
- Muestra un historial de los comandos que han sido utilizados.
- Cuenta con una vista previa de los gráficos que se hayan programado.
- Se compone con un visor de paquetes los cuales pueden ser instalados o cargados simplemente con escribir su nombre.
- Ofrece la facilidad de ejecutar fragmentos de código únicamente con marcarlo en la ventana de scripts o simplemente ir probando cada línea de programación utilizando la combinación de teclas Ctrl + Enter.
- Dispone de autocompletado de código.
- Permite las presentaciones web mediante HTML5.
- Tiene la ventaja de trabajar con varios proyectos al mismo tiempo.
- Incluye ventanas de ayuda para cada paquete que se ha instalado, únicamente con escribir el comando `help()` y enviarle como parámetro el nombre del paquete o el nombre de la función con que estemos trabajando.

- Permite crear conexiones a datasets o archivos delimitados de forma ágil y generando las líneas de código correspondientes.
- Dispone de versiones para distintos sistemas operativos, entre los cuales se figuran Windows, MAC OS y Linux.

Figura 24. Interfaz de RStudio



Fuente: Elaboración propia

Esta pantalla se encuentra dividida en cuatro partes:

1. La ventana superior izquierda, se pueden visualizar los archivos R, los cuales se los llama scripts. Es en donde el usuario se dedica a la codificación de su programa, teniendo como principal ayuda, el autocompletado de código, localización de errores e indentación del bloque de código.
2. La ventana inferior izquierda, donde se encuentra situado el prompt “>”, se llama Consola, aquí se puede ejecutar los comandos pero una línea a la vez y se logra apreciar los resultados que se pueden generar.
3. La ventana superior derecha, se divide en tres:

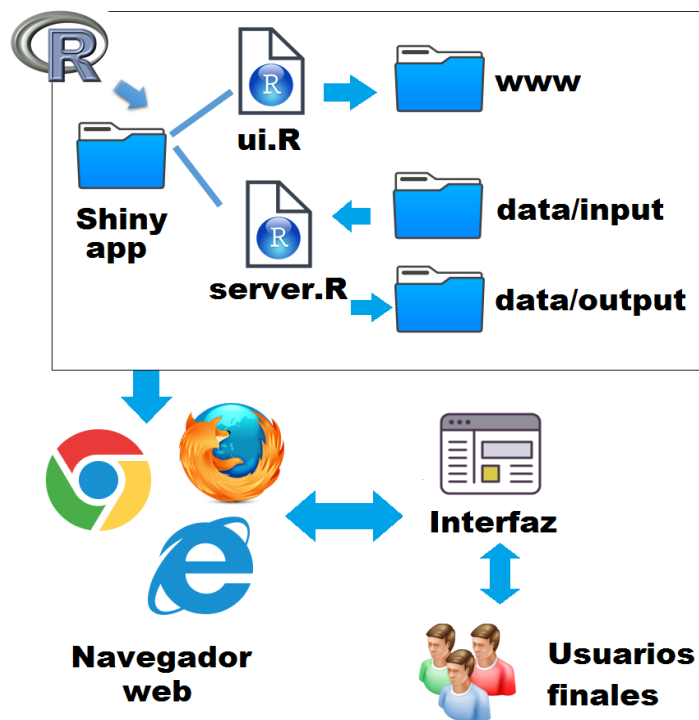
1. En la primera pestaña se pueden observar los detalles de todas las variables que se están ejecutando dentro de un script, pueden estar categorizados en Data Frames, valores (numéricos, caracteres, booleanos) y funciones creadas.
2. En la segunda pestaña se encuentra el historial de los objetos que se encuentran almacenados en memoria.
3. En la tercera pestaña se observan detalles de la ejecución de un script al momento de dar clic en el botón compilar.
4. La ventana inferior derecha, RStudio presenta el directorio de trabajo, los gráficos que se van generando, librerías disponibles para ser cargadas y poder instalarlas directamente, pestaña de ayuda y un visor web para verificar la documentación que trae una librería en formato HTML.

Por todas las características que se han revisado, esta herramienta fue la ideal para trabajar con aplicaciones en R, también permitió entender y profundizar más sobre el paradigma que sigue este tipo de lenguaje de programación.

2.5.2.2 Framework Shiny. Es una librería para el desarrollo de aplicaciones web en R [53]. Básicamente se trata de una librería desarrollada por usuarios de la comunidad que permite la creación de aplicaciones de análisis orientadas a la web de manera interactiva, como resultado se obtienen app brillantes, responsivas y de un gran alcance, todo esto realizado con el menor esfuerzo, debido a que los conocimientos de HTML, JavaScript o CSS son poco necesarios.

La construcción de proyectos Shiny se lo realiza mediante RStudio, el cual da la facilidad de programación, uso de librerías gráficas, útiles para una buena experiencia de aplicación web. A este framework se lo puede catalogar como un preprocesador HTML, ya que permite escribir pseudocódigo que más adelante será convertido en etiquetas del lenguaje de marcas de hipertexto, estilos CSS y JavaScript; el objetivo de este preprocesador es de mantener ordenado y fácil de editar. El funcionamiento de las aplicaciones creadas con Shiny obedece la arquitectura cliente-servidor por medio de dos archivos, el primero permite receptor las peticiones del usuario, es decir el fronted de la app web (ui.R) y el otro archivo (server.R) se encarga de responder a las peticiones de los usuarios, es decir el backend, en donde se encuentra la parte lógica que mantiene el correcto funcionamiento de una Shiny App.

Figura 25. Estructura de una aplicación Shiny

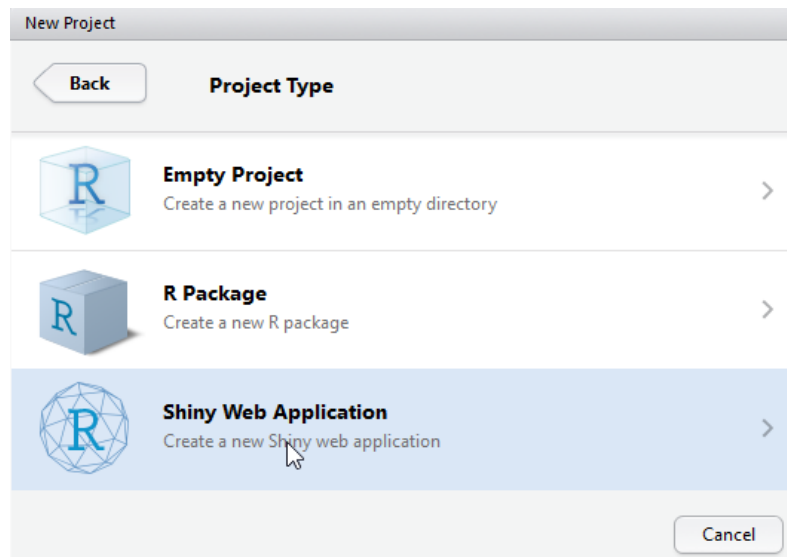


Fuente: Elaboración propia

Siguiendo las buenas prácticas de la comunidad de desarrolladores en R, se debe estructurar un proyecto de tal manera como se pudo observar en la figura anterior, con el fin de llevar de una manera ordenada el diseño de aplicaciones. Los archivos que obligatoriamente deben existir en una carpeta de proyecto son: ui.R y server.R, debido a que al momento de compilar, el servidor que trae incluido RStudio, detectará la presencia de dichos archivos para continuar con su normal ejecución. Aunque pueden existir más archivos o scripts, éstos se los debe llamar en el archivo que se lo requiera, de igual manera ocurre con las demás librerías sean gráficas, matemáticas o de conexiones a fuentes de datos.

Se pueden incluir archivos multimedia como imágenes, para lo cual se debe crear un directorio al mismo nivel que los archivos ui.R y server.R, con el nombre www; dependiendo de las demás librerías utilizadas en la lógica de programación, también se pueden incluir directorios para el ingreso y/o salida de datos.

Figura 26. Creación de un proyecto Shiny



Fuente: Elaboración propia

Para crear una interfaz de usuario se necesitan de elementos de entrada y de salida, denominados inputs y outputs [54]. Los inputs permitirán realizar el ingreso de los datos que serán utilizados para ejecutar las líneas de comando en el servidor, con ello presentar los respectivos resultados en los outputs.

Cuadro 18. Elementos utilizados en la aplicación Shiny

Elemento/Función	Tipo	Ubicación	Descripción
textInput	input	ui.R	Campo para ingresar texto
selectInput			Crea una lista despegable
radioButtons			Permite seleccionar una opción
checkboxInput			Permite seleccionar varias opciones
numericInput			Ingreso de valores numéricos
sliderInput			Fija un rango entre números
dateInput			Ingreso de fechas
actionButton	output	ui.R	Ejecuta acciones
plotOutput			Crea un elemento gráfico de salida
tableOutput			Crea una tabla como elemento de salida
uiOutput			Crear etiquetas HTML de salida

Cuadro 18. (Continuación)

Elemento/Función	Tipo	Ubicación	Descripción
textOutput	output	ui.R	Crea una cadena de texto de salida
htmlOutput			Crea un elemento de salida HTML
renderPlot		server.R	Crea un gráfico reactivo
renderDataTable			Crea una tabla reactiva
renderText			Mostrar un vector de caracteres
renderUI			Genera código HTML dinámico
renderPrint			Captura la salida en una cadena de valores

Fuente: Elaboración propia

2.5.2.3 *Técnicas utilizadas.* Las técnicas que se encuentran disponibles para el análisis descriptivo y predictivo en la aplicación son los siguientes:

1. Asociación

Esta técnica se la emplea principalmente para generar posibles relaciones entre sucesos y acciones dentro de un conjunto de datos; se puede observar como la ocurrencia de una acción puede desencadenar la aparición o alteración de otras. Por ejemplo, aplicando esta técnica podemos identificar si la compra de unos productos está asociada con la compra de otros: Si un cliente X compra pan, también comprará mantequilla ($\{Pan\} \rightarrow \{Mantequilla\}$) o si otro cliente compra pan y mantequilla, también comprará leche ($\{Pan, Mantequilla\} \rightarrow \{Leche\}$).

- Algoritmo:

El algoritmo que se empleará en esta técnica es: APRIORI. Extrae de un conjunto de datos transaccionales los elementos frecuentes, el algoritmo Apriori emplea a nivel de búsqueda en cuanto a los conjuntos de elementos frecuentes.

El código R:

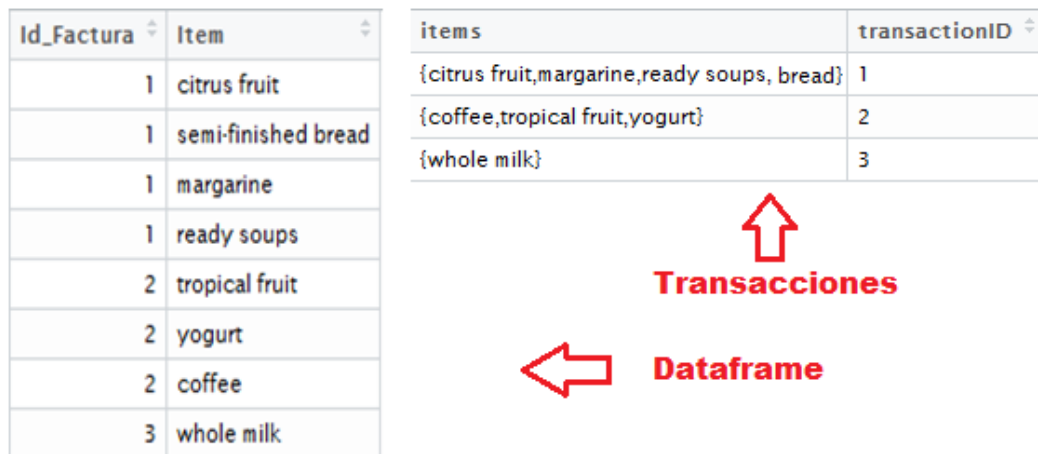
```
Trans <- Compras #Data Frame
Trans <- split(Transacciones$Item, Transacciones$Id_Factura)
Trans <- as(Trans, "transactions")
library(arulesViz);
library(arules)
```

```
reglas<-apriori(Transacciones, parameter=list(support=0.01,
confidence = 0.01))
```

- Estructura de los datos:

En R el conjunto de datos se los llama Data Frame, y es la estructura que va a permitir trabajar con gran cantidad de datos. El Data Frame debe ser tipo transaccional, es decir, debe ser transformado de acuerdo a la siguiente figura:

Figura 27. Estructura de datos para el algoritmo Apriori



Fuente: Elaboración propia

2. Clusterización

Esta técnica permite identificar segmentos o grupos donde los datos guardan gran similitud entre sí y muchas diferencias con los de otros clúster (grupos). Por ejemplo, se pueden agrupar ciertos empleados teniendo en cuenta características que sean semejantes (Edad, Estado civil, Ocupación, entre otros) o agrupar clientes dependiendo de las compras que realicen.

- Algoritmo:

El algoritmo que se empleará en esta técnica es: KMEANS, el cual es uno de los más usados en cuanto a clusterización. Usa un método de agrupamiento por vecindad, es decir que para su funcionamiento necesita saber el número de grupos y de un conjunto de datos a agrupar sin estar etiquetados.

Para conocer cuántos grupos necesita el método para obtener un resultado óptimo, se aplica los siguientes comandos:

```
set.seed(123)
Error <- NULL
K_Max <- 9 #Máximos clusters permitidos
# Guarda el error de cada ejecución en el vector "Errores"
for (i in 1:K_Max){
  Error[i] <- sum(kmeans(dataframe, centers=i)$withinss)
}
```

Para aplicar el algoritmo KMEANS en un Data Frame, podemos hacer uso de todas las columnas siempre y cuando sean campos con valores numéricos, para saber qué aspecto en común tienen dichos campos.

```
#Aplicar el algoritmo de k medias
Clusters <- 3 #Se usarán 3 cluster
kmeans(dataframe, Clusters)
```

El algoritmo automáticamente agrupará los datos que le hemos proporcionado.

- Estructura de los datos:

Para la aplicación de esta técnica, la estructura tiene que ser un Data Frame. Para ello no va a importar el número de campos, y el tipo de dato debe ser diferente a caracteres, valores booleanos; sólo se admitirá valores numéricos, en este caso, el ejemplo se lo llevará a cabo con un Data Frame que consta de los tamaños y longitudes de pétalos y sépalos de tres tipos de flores.

Figura 28. Estructura de datos para el algoritmo Kmeans

Longitud_sepalo ↕	Ancho_sepalo ↕	Longitud_petalo ↕	Ancho_petalo ↕
5.1	3.5	1.4	0.2
4.9	3.0	1.4	0.2
4.7	3.2	1.3	0.2
4.6	3.1	1.5	0.2
5.0	3.6	1.4	0.2
5.4	3.9	1.7	0.4
4.6	3.4	1.4	0.3
5.0	3.4	1.5	0.2

Fuente: Elaboración propia

3. Análisis de componentes principales (ACP)

Esta técnica permite sintetizar la información o también se la conoce como técnica de reducción de dimensiones (los campos de un dataframe); es decir, si se tiene un conjunto de datos con muchas variables, el objetivo de ACP es de reducir las variables a un menor número y con ello mantener la significancia de la información [55].

- Algoritmo:

El algoritmo que se utilizó es: PRCOMP, define los factores de manera que el primero extraiga la mayor significancia de las variables originales, el segundo la más alta entre las demás y así progresivamente. De todos los factores que se han creado, se eligen los que recojan un porcentaje suficiente de variabilidad, denominándolos así componentes principales (CP). Por tanto, el dataframe que se generará tendrá un número de columnas como CP existan y filas como el número de variables.

El código R:

```
analisis_PCA<-prcomp(dataframe)
DatosCP<-data.frame(Componente1 = analisis_PCA$x[,1],
                    Componente2 = analisis_PCA$x[,2] )
```

- Estructura de los datos:

Para la aplicación de esta técnica, deben existir un mayor número de variables y de registros; de igual manera los datos pueden ser de cualquier tipo.

Figura 29. Estructura de datos para el algoritmo Prcomp

Area	Perimetro	Densidad	Longitud_Grano	Ancho_Grano	Coef_Asimetria	Ranura_Grano
15.26	14.84	0.8710	5.763	3.312	2.2210	5.220
14.88	14.57	0.8811	5.554	3.333	1.0180	4.956
14.29	14.09	0.9050	5.291	3.337	2.6990	4.825
13.84	13.94	0.8955	5.324	3.379	2.2590	4.805
16.14	14.99	0.9034	5.658	3.562	1.3550	5.175
14.38	14.21	0.8951	5.386	3.312	2.4620	4.956
14.69	14.49	0.8799	5.563	3.259	3.5860	5.219
14.11	14.10	0.8911	5.420	3.302	2.7000	5.000
16.63	15.46	0.8747	6.053	3.465	2.0400	5.877
16.44	15.25	0.8880	5.884	3.505	1.9690	5.533

Fuente: Elaboración propia

4. Regresión lineal

Esta técnica permite predecir los valores de una variable continua (dependiente) a partir de la evolución sobre otra variable continua (independiente) [56], muchas veces el tiempo, ganancias, precio, unidades de producto, entre otros. Por ejemplo, se intenta predecir cuánto demorará (minutos) un técnico en reparar cierta cantidad de aparatos (unidades), teniendo registros de los resultados de las reparaciones que se han hecho anteriormente.

- Algoritmo:

El algoritmo que se empleará en esta técnica es: PREDICT, el cual emplea modelos para la resolución de una predicción basándose en los datos que se le administren para llevar a cabo su proceso.

```
minutos = reparaciones$minutos
unidades = reparaciones$unidades
#Crear el modelo lineal de regresion
modelo = lm(minutos~unidades)
#Ingresamos 3 unidades para predecir su tiempo
unidades_prediccion <- 3
#Predicción para saber cuánto se tardará en reparar 3 unidades
Prediccion <- predict(modelo, newdata =
data.frame(unidades=unidades_prediccion))
```

- Estructura de los datos:

El Data Frame debe tener valores continuos, preferible tiempo, costos, precio, unidades de productos entre otros. En este caso los datos que utilizaremos serán históricos sobre un centro de reparaciones de aparatos electrónicos.

Figura 30. Estructura de datos para el algoritmo Predict

minutos	unidades
23	1
29	2
49	3
64	4
74	4
87	5

Fuente: Elaboración propia

5. Clasificación

Esta técnica permite predecir datos futuros a partir de un modelo en el cual se utilizan datos que ya han sido realizados o que sean conocidos. Consiste, como su nombre lo indica, clasificar individuos (registros) en una determinada clase de acuerdo a las características que posean. Por ejemplo predecir si un cliente comprará otra vez en una tienda, detectar enfermedades de una persona de acuerdo a sus síntomas, entre otros.

- Algoritmo:

El algoritmo que se utilizará en esta técnica es: RANDOMFOREST, fue desarrollado por Breiman [57]. Crea árboles de decisiones para datos aleatorios dentro de un Data Frame, dando como resultado el voto mayoritario que serán representados en la columna de resultado booleano. Puede ser usado tanto para la clasificación como la regresión. El algoritmo irá seleccionando los campos de mayor importancia para predecir los valores de Si/No en el Data Frame. Se crea un modelo con los datos:

```
library(randomForest) library(C50)
Modelo<-randomForest(churn ~ .,
                      data=Datos_Entrenamiento, # datos para entrenar
                      ntree=30,                 # cantidad de arboles
                      mtry=6,                   # cantidad de variables
                      replace=T)                 # muestras con reemplazo
```

Después de crear el modelo, se procede a predecir los datos de Sí o No. Luego se compara mediante una matriz de confusión, si los datos predichos se acercan a la realidad.

```
Clasificacion <- predict (Modelo , Datos_Test);
Matriz_Confusion <- table(Datos_Test[, "churn"],Clasificacion)
```

- Estructura de los datos:

El Data Frame con el que trabajará este algoritmo puede contener tantas columnas de diferentes tipos, como pueden existir registros que estén en blanco. El campo a predecir debe estar expresado en valores de Si/No o de 1/0. En este caso los datos que se usarán corresponden a un centro de atención al cliente, en donde tienen registros de los clientes que han abandonado la empresa y los que aún continúan en ella.

En el Data Frame está estructurado con 20 campos o columnas.

Figura 31. Estructura de datos para el algoritmo Random forest

total_intl_calls	total_intl_charge	number_customer_service_calls	churn
3	2.70	1	no
3	3.70	1	no
5	3.29	0	no
7	1.78	2	no
3	2.73	3	no

Fuente: Elaboración propia

6. Predicción

Esta técnica intenta determinar los valores de una o más variables, partiendo de un conjunto de datos y pueden calcularse por varios algoritmos predictivos. Por ejemplo se pretende predecir si un cliente comprará en un local, el tiempo de inactividad de un servidor, saber que cliente atenderá un correo de publicidad. Hace uso de datos históricos con resultados que ya se han realizado.

- Algoritmo:

El algoritmo que se utilizará en esta técnica es: NAIVE BAYES, el cual calcula las probabilidades condicionales a-posterior de una variable categórica dada una variables predictoras (generalmente Si o No) independientes utilizando la regla de Bayes [58].

```
library(e1071) # librerias de naiveBayes
# crea tabla Probabilidades excluyendo la columna ID
Probabilidades <- naiveBayes(Datos_Historicos$Compra ~.,
data=Datos_Historicos[-1])
# crea prediccion a datos nuevos, excluyendo el campo compra
Prediccion <- predict (Probabilidades , Nuevos_Clientes[,-6])
# Escribe en la comlumna compra la prediccion obtenida
Nuevos_Clientes$Compra <- NULL
Compra_Prediccion <- Nuevos_Clientes
Compra_Prediccion$Prediccion_Compra <- Prediccion
```

- Estructura de los datos:

El Data Frame puede tener pocos o muchos campos, se debe procurar que dichos campos no estén relacionados con otros y muy importante, debe tener un campo de destino, el cual debe ser un dato discreto (Si o No, por ejemplo).

En este caso se utilizará un Data Frame que contiene registros de clientes que han aceptado o han rechazado una compra que cierta empresa ha comunicado mediante llamadas y correos electrónicos.

Figura 32. Estructura de datos para el algoritmo Naive bayes

IdCliente	EstadoCivil	Profesion	Universitario	TieneVehiculo	Compra
1	Casado	Empresario	Si	No	No
2	Casado	Empresario	Si	Si	No
3	Soltero	Empresario	Si	No	Si
4	Viudo	Desempleado	Si	No	Si
5	Viudo	Empleado	No	No	Si
6	Viudo	Empleado	No	Si	No
7	Soltero	Empleado	No	Si	Si
8	Casado	Desempleado	Si	No	No

Fuente: Elaboración propia

2.5.2.4 *Otras librerías.* Para el funcionamiento de la aplicación, se recurrió a la utilización de librerías para representar los datos en interactivos gráficos estadísticos, pastel, histogramas, puntos de dispersión y muchas más; se las puede encontrar en la librería Plotly [59]. De la misma manera la aplicación consta de librerías para la comunicación entre R y los demás sistemas de bases de datos, archivos externos alojados en repositorios web, interfaces entre los drivers JDBC, PostgreSQL, MySQL, Oracle, entre otros.

El resto de librerías se las ha ocupado para dar una brillante apariencia a la aplicación web, estilos llamativos, eventos y animaciones (shinydashboard, shinyBS, shinyjs respectivamente). Finalmente para la estructuración de los datos en las distintas tablas presentadas en la interfaz web se ha utilizado el paquete DT y datatable, los cuales permiten hacer tablas responsivas acopladas perfectamente en los respectivos paneles de organización.

2.6 Ejecución del prototipo

2.6.1 *Interfaces de la aplicación.* A continuación se detallan los menús más significativos de la aplicación IOTMach Analysis.

2.6.1.1 *Inicio.* El menú inicio muestra la información sobre la estructura y las necesidades que pretende satisfacer el software, de la misma manera es el punto de partida para navegar por toda la interfaz de la aplicación.

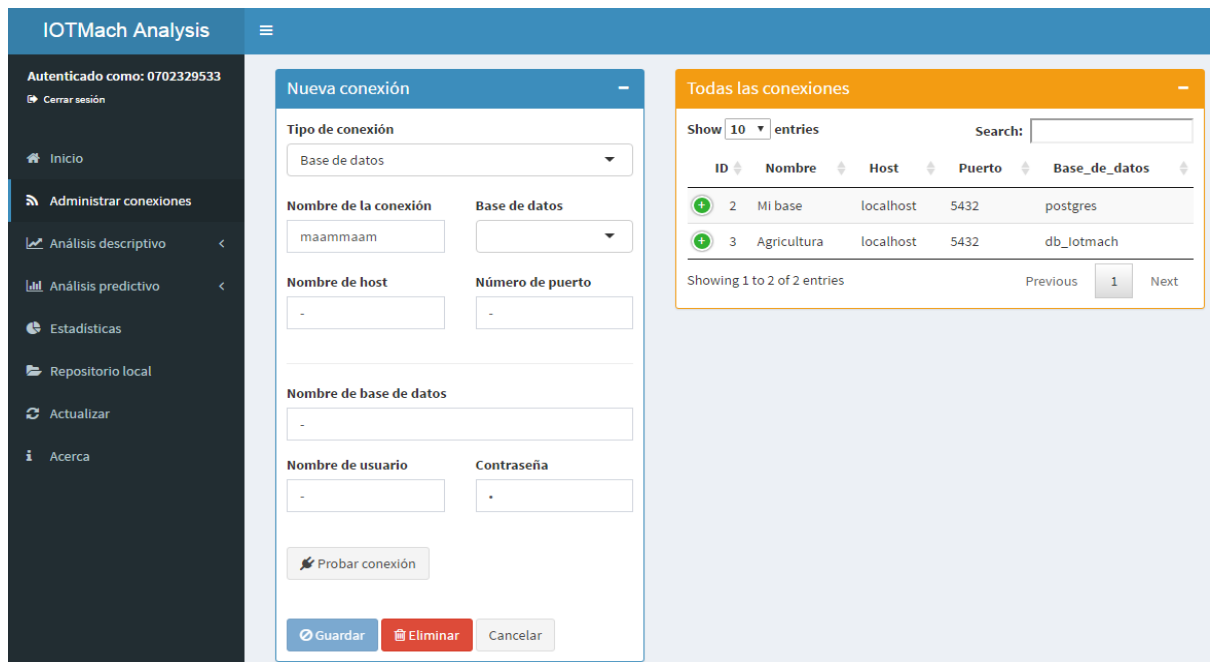
Figura 33. Menú inicio

The screenshot shows the IOTMach Analysis web application interface. On the left is a dark sidebar with the following menu items: 'Inicio', 'Administrar conexiones', 'Análisis descriptivo', 'Análisis predictivo', 'Estadísticas', 'Repositorio local', 'Actualizar', and 'Acerca'. The main content area features the IOTMach logo (Research Group 2015) and text describing the application as a web-based interactive system using R and Shiny. It highlights its use of Machine Learning and data mining for statistical and predictive analysis. Below the text is the IOTMach 2016 Research Group logo. On the right, there are two panels: 'Análisis de la información' (Analysis of information) and 'Estadísticas' (Statistics). The 'Análisis de la información' panel lists six analysis types: Clusterización (Descriptivo), Reglas de asociación (Descriptivo), Componentes principales (Descriptivo), Regresión (Predictivo), Predicción (Predictivo), and Clasificación (Predictivo). The 'Estadísticas' panel lists two options: Resumen de los datos (Data Summary) and Gráficos (Plots).

Fuente: Elaboración propia

2.6.1.2 *Administrar conexiones.* Se crearán y listarán las conexiones a las bases de datos y/o los archivos externos, para lo cual se debe llenar los campos que son requeridos, después se prueba la conexión dando clic en el botón probar. Por último se validarán los datos ingresados y se procederá a guardar dicha conexión en el archivo de configuraciones que posee la aplicación.

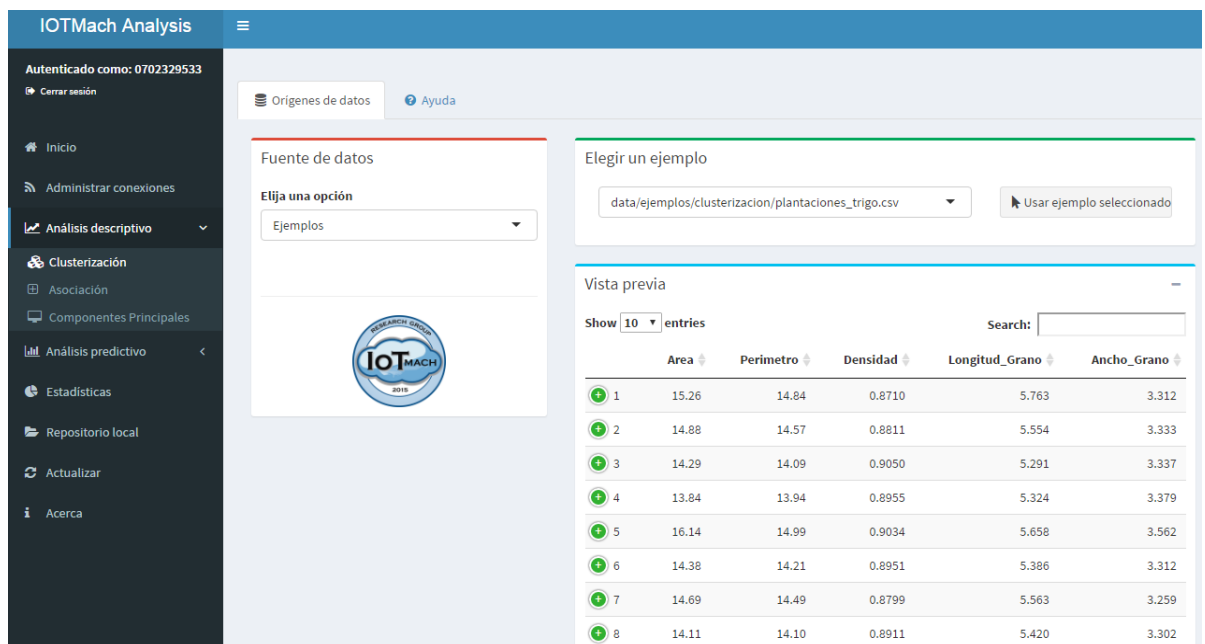
Figura 34. Menú administrar conexiones



Fuente: Elaboración propia

2.6.1.3 *Análisis descriptivo*. Este menú consta de 3 submenús: clusterización, asociación y componentes principales. Al hacer clic en una de estas opciones, se mostrará un panel en donde se tendrá que seleccionar la fuente de datos

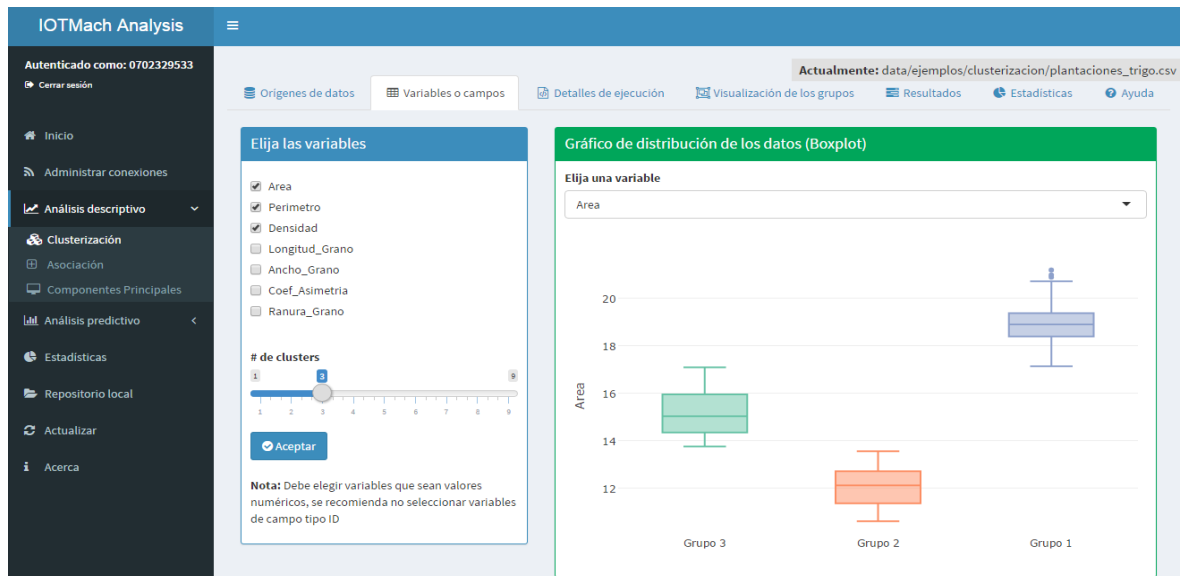
Figura 35. Submenú clusterización dentro del análisis descriptivo



Fuente: Elaboración propia

De inmediato se activará una pestaña en donde debemos elegir los campos o variables del dataframe que se van a someter a análisis, de igual forma seleccionamos el número de clusters que se quieran encontrar y se da clic en Aceptar (Ver Figura 36)

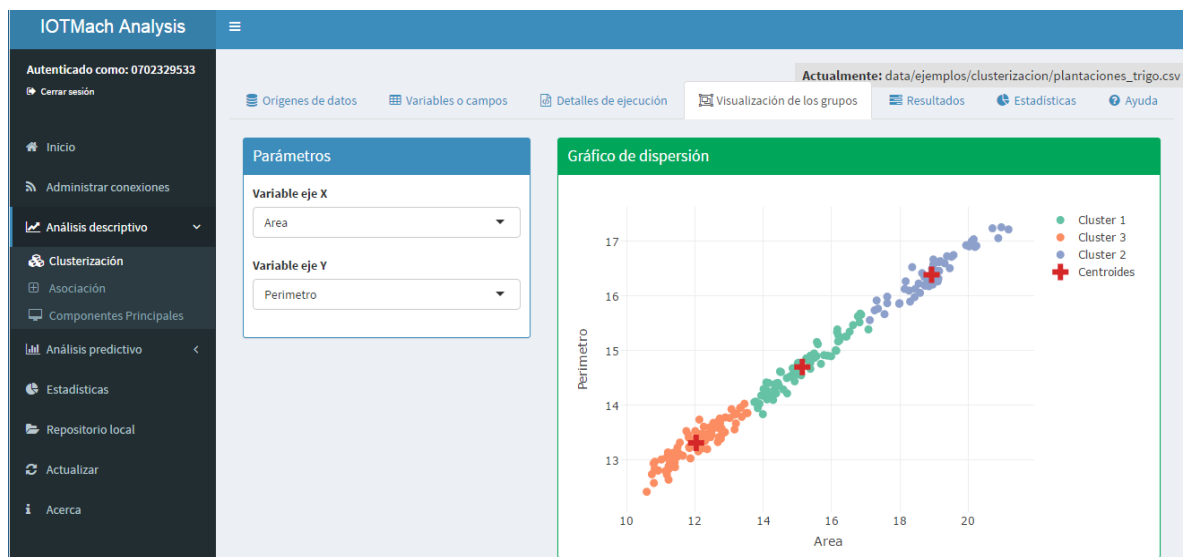
Figura 36. Submenú clusterización pestaña variables



Fuente: Elaboración propia

Las demás pestañas se habilitarán y se podrá visualizar el detalle de ejecución del algoritmo, los clusters creados, los resultados del dataframe con sus respectivos grupos, estadísticas y la respectiva ayuda que brinda este submenú.

Figura 37. Submenú clusterización pestaña grupos



Fuente: Elaboración propia

2.6.1.4 *Análisis predictivo*. Este menú consta de 3 submenús: regresión lineal, predicción bayesiana y clasificación. Al hacer clic en una de estas opciones, se mostrará un panel en donde se tendrá que seleccionar la fuente de datos

Figura 38. Submenú regresión lineal dentro del análisis predictivo

The screenshot shows the IOTMach Analysis web application. The left sidebar contains navigation options: Inicio, Administrar conexiones, Análisis descriptivo, Análisis predictivo (expanded), Regresión Lineal (selected), Predicción Bayesiana, Clasificación, Estadísticas, Repositorio local, Actualizar, and Acerca. The main content area has a header with 'Origenes de datos' and 'Ayuda'. Below this, there's a 'Fuente de datos' section with a dropdown menu set to 'Ejemplos'. To the right, an 'Elegir un ejemplo' section shows a dropdown menu with 'data/ejemplos/regresion/area_abono.csv' and a 'Usar ejemplo seleccionado' button. Below that, a 'Vista previa' section shows a table with 7 entries. The table has columns for ID, AREA_M2, and ABONO_KG.

ID	AREA_M2	ABONO_KG
1	500	42
2	600	52
3	720	61
4	900	74
5	1000	8
6	1800	95
7	2500	111

Fuente: Elaboración propia

De inmediato se activará una pestaña en donde debemos elegir las variables del dataframe tanto dependiente como independiente, se da clic en Aceptar. (Ver Figura 39)

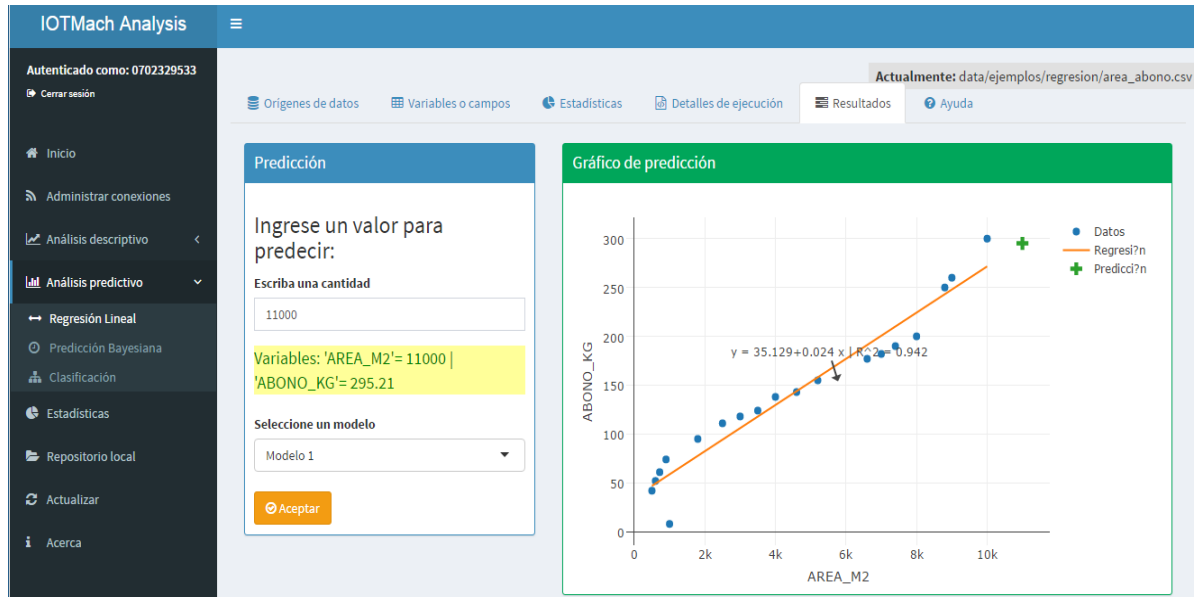
Figura 39. Submenú regresión lineal pestaña variables

The screenshot shows the IOTMach Analysis web application with the 'Variables o campos' tab active. The 'Elija las variables' panel has 'ABONO_KG' selected under 'Predictoria (Dependiente)'. Under 'En base a (Independiente)', 'AREA_M2' is selected. There is an 'Aceptar' button and a note: 'Nota: Debe elegir variables que sean valores numéricos, se recomienda no seleccionar variables de campo tipo ID o la misma variable dos veces'. To the right, a 'Gráfico de dispersión' shows a scatter plot of ABONO_KG (y-axis, 0 to 300) versus AREA_M2 (x-axis, 0 to 10k). The data points show a positive correlation.

Fuente: Elaboración propia

Las demás pestañas se habilitarán y se podrá visualizar el detalle de ejecución del algoritmo, los resultados y la respectiva ayuda que brinda este submenú.

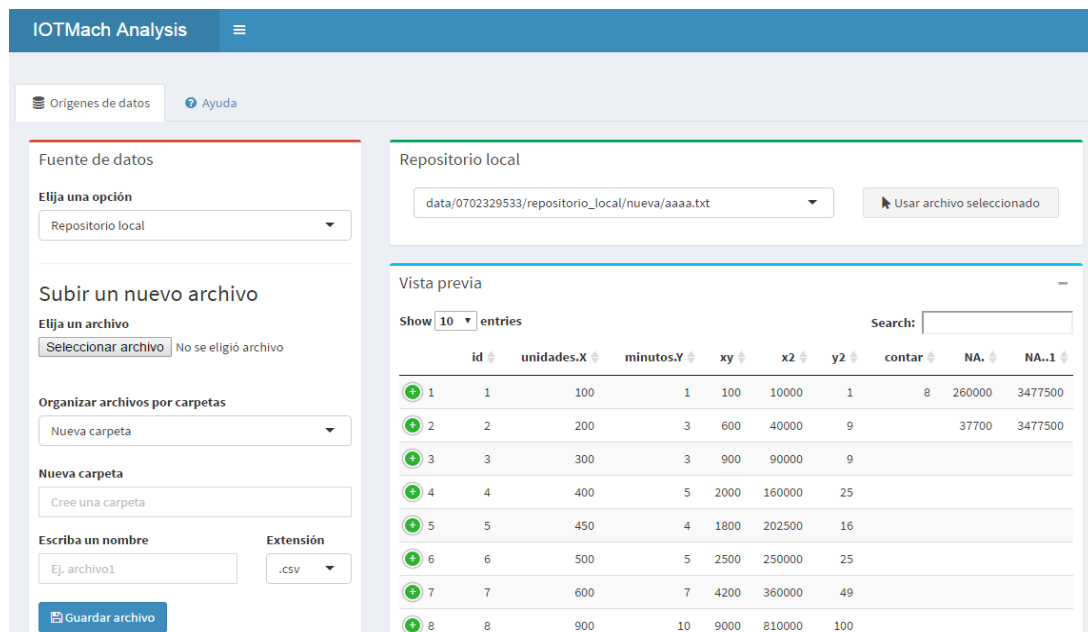
Figura 40. Submenú regresión lineal pestaña resultados



Fuente: Elaboración propia

En cada submenú, en la pestaña orígenes de datos tiene la opción de subir archivos locales para ser analizados, la interfaz para su uso es muy sencilla. (Ver Figura 41)

Figura 41. Subida de archivos repositorio local



Fuente: Elaboración propia

3. EVALUACIÓN DEL PROTOTIPO

3.1. Plan de evaluación

Una vez terminada la construcción de la aplicación web, fue necesario someterla a una serie de pruebas, con el objetivo de comprobar su calidad antes de que pase por la fase de producción. Por medio de las pruebas, fue posible medir la robustez frente a distintas acciones que los usuarios finales realizarán desde el lugar en donde se encuentren.

Fue necesario probar el desenvolvimiento de la aplicación, funcionalidad, rendimiento, tiempos de respuesta, aspectos de amigabilidad de las interfaces con los usuarios, entre otras pruebas. Los planes de prueba tienen por objeto detectar errores y solucionarlos de forma organizada y técnica.

3.1.1 *Pruebas unitarias.* Cuando se concluya con el desarrollo de un software, la primera etapa de pruebas que se deben tener en cuenta son precisamente las unitarias. Estas pruebas se encargan de comprobar si un módulo de la aplicación se encuentra preparado para que pueda ser anexado a los demás módulos. Estas pruebas logran cubrir la totalidad de código programado en la aplicación.

Cuando se ejecuten las pruebas se deben hacer independientemente de lo que suceda con toda la aplicación, en sí cada prueba no debe afectar la ejecución de otra, se recomienda realizar las pruebas unitarias las veces que sean necesarias. Se han aplicado las pruebas a los módulos de Administración de conexiones, análisis descriptivo y análisis predictivo; la descripción de la prueba es la siguiente:

Cuadro 19. Descripción de pruebas unitarias

Aplicación IOTMach Analysis	
Objetivo	Verificar que el código funcione conforme ha sido programado y de acuerdo a las especificaciones.
Descripción	<ul style="list-style-type: none">• Probar los módulos separadamente para facilitar el proceso de testeo• Entre los aspectos a evaluar se encuentran: validaciones, valores requeridos y opcionales, valores con rango, mensajes de alerta, manejo correcto de archivos y sus formatos

Cuadro 19. (Continuación)

Aplicación IOTMach Analysis	
	<ul style="list-style-type: none"> • Abarcar todo el proceso de ejecución posible en el código para la unidad que se esté evaluando
Criterios	Comprobar los resultados obtenidos con los que se esperan de la aplicación. En caso de existir errores se los debe reportar para tenerlos en cuenta para su posterior retroalimentación.

Fuente: Elaboración propia

El formato para llevar a cabo el control de las pruebas unitarias se lo puede revisar desde el Anexo A hasta el Anexo D.

3.1.2 *Pruebas de integración.* Se las debe realizar cuando se hayan completado las pruebas unitarias de cada módulo o submódulo. Estas pruebas ayudan a prevenir aparición de fallas que tienen que ver con la mala arquitectura de las interfaces que componen la aplicación.

Una vez que se han finalizado la detección de errores en los submódulos, se procedió a integrarlos junto con los demás módulo. Las actividades que hacen referencia a este tipo de pruebas, consisten en las revisiones completas de todas las funciones que la aplicación realiza, que van desde la validación de datos de entradas por medio de formularios, validación de archivos y extensiones hasta la conexión con distintas bases de datos.

Cuadro 20. Descripción de pruebas de integración

Aplicación IOTMach Analysis	
Objetivo	Identificar los posibles errores que fueron producidos por la unión de los módulos probados individualmente.
Descripción	<ul style="list-style-type: none"> • Verificar el funcionamiento entre los componentes que conforman la aplicación • Probar las conexiones creadas, utilizando cada dataset para introducirlo en cada uno de los análisis • Decidir las respectivas acciones a tomar cuando se presenten problemas

Cuadro 20. (Continuación)

Aplicación IOTMach Analysis	
Criterios	Comprobar los resultados obtenidos con los que se esperan de la aplicación. Empezar con los módulos de nivel superior para así verificar que las llamadas a los otros módulos sean correctas, con los parámetros indicados.

Fuente: Elaboración propia

El formato para llevar a cabo el control de las pruebas de integración se lo puede revisar desde el Anexo F hasta el Anexo H.

3.1.3 *Pruebas de usabilidad.* Estas pruebas se encargan de analizar la facilidad con la que los usuarios finales pueden manejar la aplicación sin ningún problema. En el caso de los aspectos que se tuvieron en cuenta para emplear dicha prueba en la aplicación fueron: determinar si el usuario puede operar en el sistema, navegando entre los menús, comprendiendo todas las funciones que se deben realizar, determinar si la interfaz de usuario es amigable tanto para las personas con poca o nula experiencia en una aplicación web y con todos estos aspectos, determinar si el prototipo creador puede requerir cambios para cumplir los objetivos.

Cuadro 21. Descripción de pruebas de usabilidad

Aplicación IOTMach Analysis	
Objetivo	Determinar la aceptación, la manera de usar y entender la aplicación por parte de los usuarios.
Descripción	<ul style="list-style-type: none"> • Medir el grado de satisfacción de los usuarios cuando utilizan por primera vez la aplicación • Evidenciar que las características de apariencia, texto, mensajes, calidad de los gráficos y la confiabilidad de cada análisis sean los deseados • Asegurar que la ayuda proporcionada por la aplicación permita orientar a los usuarios en caso de requerirla
Criterios	<ul style="list-style-type: none"> • Comprobar los resultados obtenidos con los que se esperan de la aplicación.

Fuente: Elaboración propia

El formato para llevar a cabo el control de las pruebas de usabilidad se lo puede revisar en el Anexo E.

3.1.4 *Pruebas de rendimiento.* Estas pruebas se encargan medir el performance de las aplicaciones, es decir se su función es dar a conocer los aspectos relacionados con la carga al sistema, los tiempos de respuesta, conexión a través del internet, entre otros. Con las pruebas de funcionalidad no se miden estas características, debido a que no incluyen ámbitos de concurrencia y capacidad máxima con la que la aplicación debe funcionar adecuadamente.

Cuadro 22. Descripción de pruebas de rendimiento

Aplicación IOTMach Analysis	
Objetivo	Medir el performance de la aplicación mediante pruebas de carga y concurrencia
Descripción	<ul style="list-style-type: none"> • Realizar las debidas pruebas para comprobar la capacidad de la aplicación para el correcto funcionamiento bajo distintas condiciones de carga • Realizar las pruebas de concurrencia bajo distintas peticiones esperadas para evitar colapsos en el funcionamiento de la aplicación
Criterios	<ul style="list-style-type: none"> • Comprobar los resultados obtenidos con los que se esperan de la aplicación.

Fuente: Elaboración propia

El formato para llevar a cabo el control de las pruebas de rendimiento, tanto de carga como de concurrencia, se lo puede revisar en: Anexo I y Anexo J, respectivamente.

3.1.5 *Pruebas de regresión.* Estas pruebas se las aplica después de haber realizado las pruebas de funcionamiento, en donde buscan verificar si las funciones de un sistema siguen intactas una vez que se haya modificado algún aspecto relacionado con la carga, concurrencia u otra característica de performance que se haya realizado anteriormente. Para llevar a cabo este tipo de pruebas, se requiere el uso de técnicas de regresión lineal que, mediante la información recolectada en las pruebas de rendimiento, permitan conocer el comportamiento de la aplicación cuando su carga va incrementando.

Cuadro 23. Descripción de pruebas de regresión

Aplicación IOTMach Analysis	
Objetivo	Comprobar si se mantienen los niveles medidos en las pruebas de rendimiento a partir de nuevos cambios
Descripción	<ul style="list-style-type: none"> • Medir los tiempos de respuesta de la aplicación relacionados al almacenamiento y a la visualización de la información de un dataset • Aplicar técnicas de regresión lineal para conocer el tiempo que se obtendrá con el ingreso de nuevos datasets con mayor capacidad
Criterios	<ul style="list-style-type: none"> • Comprobar los resultados obtenidos con los que se esperan de la aplicación.

Fuente: Elaboración propia

El formato para llevar a cabo el control de las pruebas de regresión se lo puede revisar en el Anexo K.

3.2. Resultados de la evaluación

Cada una de las pruebas que se han realizado de manera independiente y de manera conjunta ayudó a detectar ciertas vulnerabilidades, muchas de ellas estaban ligadas más con las validaciones de datos dentro de los formularios. Hubo pequeñas fallas en cuanto a la presentación de los datos, debido a que ciertas líneas de código en la parte de diseño del fronted, estaban mal referenciadas, lo que ocasionaba conflictos a la hora de visualizar los resultados de tipo tablas, gráficos estadísticos y renderización de textos como detalles de ejecución para los algoritmos.

Las pruebas se las realizó a un grupo de 20 estudiantes que cursan el séptimo semestre de la carrera de Ingeniería de Sistemas de la UTMACH, en donde se aprovechó los conocimientos que tenían sobre programación y Machine Learning que adquirieron en el transcurso del módulo de Base de datos III impartido por la Ing. Bertha Mazón.

Pruebas unitarias. Al momento de realizar las pruebas se encontró con formularios que no se encontraban validados, campos que no especificaban si era obligatorio o no; para ello se hace referencia a lo siguiente:

- Al momento de subir un archivo en el menú repositorio local, los campos no se vacían para permitir la subida de más archivos.

- La validación de los campos numéricos permitían el ingreso de números negativos en donde era requerido cantidades mayores a 0, por ende los resultados gráficos no se mostraban de una buena manera
- En el submenú regresión lineal, al momento de ingresar un valor erróneo en un campo numérico, no presentaba ninguna advertencia de error.
- Cuando se creaba una nueva conexión, la aplicación no tenía en cuenta las duplicaciones de nombres asignados para cada registro y por ende al momento de listar todas las conexiones, no se apreciaba la conexión que fue introducida.

Pruebas de integración. En donde se observó que en cuanto a los módulos de análisis tanto descriptivo como predictivo, las conexiones creadas recientemente no eran mostradas en el panel de selección de datos. Por lo cual se procedió a examinar dicha vulnerabilidad para su posterior corrección.

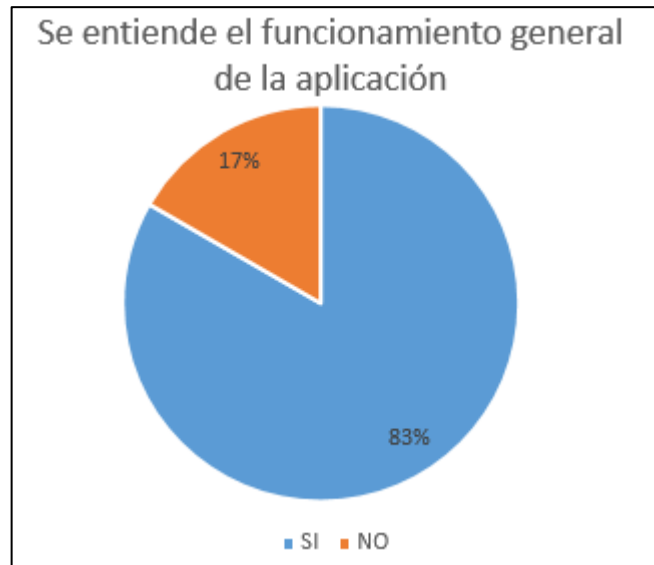
- En el módulo administrar conexiones, al momento de ejecutar la acción eliminar conexión no se recargaba la lista con los datos que fueron afectados; para continuar con la operación era necesario actualizar la página nuevamente para visualizar los cambios que se han realizado
- En el módulo análisis descriptivo se detectó que al momento de suceder algún error con las fuentes de datos seleccionadas, no aparecía un mensaje que confirma la causa de dicha anomalía y la aplicación se detenía.
- En el módulo análisis predictivo se encontró un error al momento de querer extraer los campos de un dataframe, los cuales eran de tipo fecha. La aplicación se caía y no permitía continuar con el correcto funcionamiento de los demás menús.
- Cuando se consultaba las tablas de una base de datos, no se cerraban las conexiones y la aplicación tendía a colgarse.

Pruebas de usabilidad. Con esta prueba se pudo verificar que el uso las interfaces son intuitivos y amigables, debido a que el grupo de estudiantes pudieron navegar dentro de la aplicación sin ningún problema. Los resultados obtenidos en los aspectos más importantes planteados para evaluar la usabilidad de la aplicación fueron:

En cuanto a la pregunta sobre el funcionamiento general de la aplicación, los estudiantes indicaron en un 83% que si entendían, en cambio el 17% supo manifestar que se les

dificultó al inicio, pero se les aclaró el panorama con las explicaciones que se les supo dar posteriormente. (Ver Figura 42).

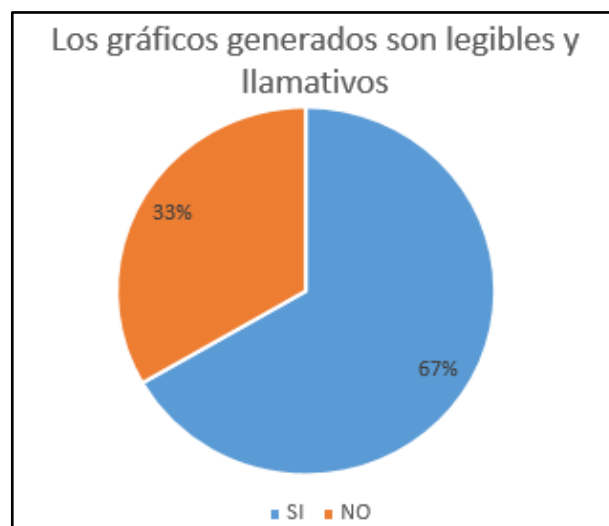
Figura 42. Respuesta sobre el funcionamiento de la aplicación



Fuente: Elaboración propia

Con respecto a la pregunta sobre la calidad de los gráficos, los estudiantes indicaron en un 67% que son llamativos y legibles, en cambio el 33% no estaban conformes con los resultados estadísticos que se generaron. (Ver Figura 43).

Figura 43. Respuesta sobre los gráficos de la aplicación



Fuente: Elaboración propia

En cuanto a la pregunta sobre el contenido textual y/o faltas de ortografía dentro de la aplicación, el 80% del grupo de estudiantes notó que no existían faltas o algún tipo de ilegibilidad, en cambio el 20% observó algunas palabras con inconsistencia y supieron dar sus observaciones. (Ver Figura 44).

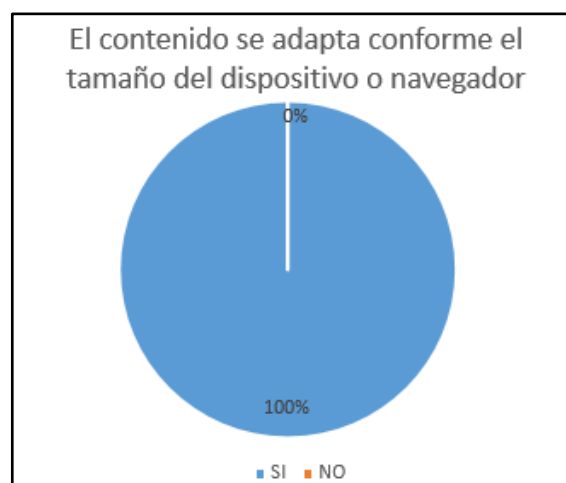
Figura 44. Respuesta sobre ortografía y textos



Fuente: Elaboración propia

Con respecto a la pregunta acerca de la adaptación del contenido de la aplicación en el navegador, todos los estudiantes se dieron cuenta de la responsividad de la interfaz, esto se corroboró cuando se mostró la aplicación en una computadora de escritorio, una portátil y en un Smartphone; en los tres escenarios el contenido se mantuvo visible y claro, ajustándose de acuerdo al tamaño del dispositivo. (Ver Figura 45).

Figura 45. Respuesta sobre la responsividad de la aplicación

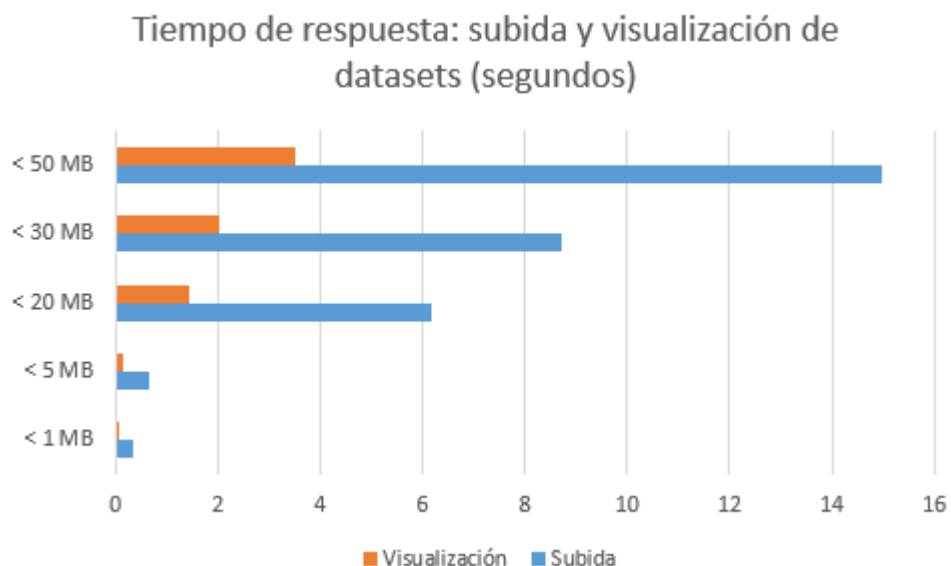


Fuente: Elaboración propia

Pruebas de rendimiento. Mediante estas pruebas se pudo medir el tiempo que tarda la aplicación en crear y cargar un dataset desde archivos que cada estudiante subió en el repositorio local o en una conexión. También se midió el comportamiento de la aplicación cuando se encuentra funcionando con varios usuarios a la vez, cuando se la utilizaba con una conexión inalámbrica o cableada, los aspectos de la respuesta oportuna y de inactividad en caso de que existiera.

- Pruebas de carga: al finalizar la subida de archivos con distintos tamaños expresados en megabytes, se pudo deducir que los archivos menores a 50 MB tardan en guardarse en el repositorio de la aplicación en tiempos menores a los 20 segundos; en cambio para archivos más pesados el tiempo varía, llegando hasta un minuto. La visualización del dataset es más rápido, en menos de un segundo para archivos livianos y menor a 35 segundos para archivos más pesados (Ver Figura 46)

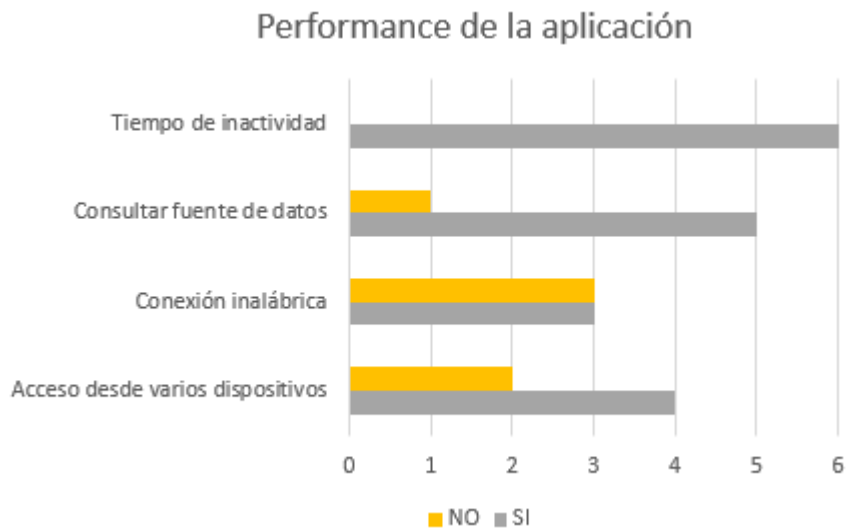
Figura 46. Tiempo de respuesta carga de archivos a la aplicación



Fuente: Elaboración propia

- Prueba de concurrencia: con los resultados obtenidos al aplicar la prueba, se pudo constatar que al ingresar desde varios dispositivos a la aplicación, ésta tiende a demorar un poco más de lo esperado, debido a que si se trabaja con conexión inalámbrica afecta el funcionamiento y carga, según los estudiantes pudieron manifestar. En cambio para acceder a una fuente de datos almacenada en la aplicación los tiempos son cortos, y por último, la aplicación no presentó síntomas de inactividad cuando los estudiantes se encontraban ejecutando sus funciones. (Ver Figura 47)

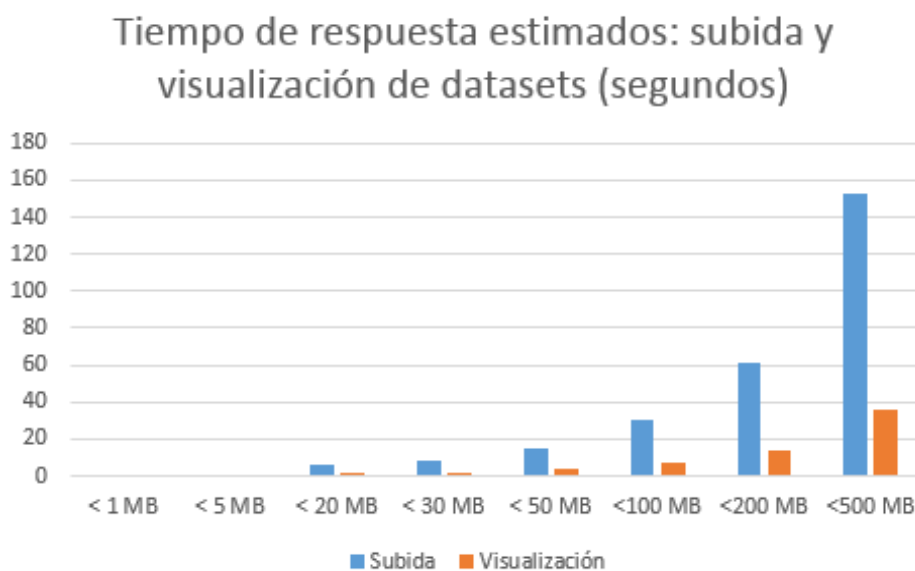
Figura 47. Resultados prueba de concurrencia



Fuente: Elaboración propia

Pruebas de regresión. Mediante la técnica de regresión lineal y los datos obtenidos de la prueba de carga realizada con anterioridad, se pudo observar cuáles serán los tiempos de carga y visualización que demorará la aplicación en realizar para distintos archivos con capacidades mayores a los 50 megabytes. En cuanto a los resultados, se estima que los archivos muy pesados pueden procesarse entre uno y hasta dos minutos. (Ver Figura 48)

Figura 48. Estimación de los tiempos de carga



Fuente: Elaboración propia

CONCLUSIONES

Una vez finalizado el desarrollo e implementación del módulo de análisis para agricultura, aplicación de técnicas de minería de datos y algoritmos de Machine Learning, se puede concluir que:

- La minería de datos dentro del funcionamiento normal de un negocio, cualquiera que éste sea, permite convertir la información extraída en información valiosa para que al final se llegue a tomar una posición en el mercado diferenciándose de los demás competidores en el campo que se encuentre incursionado.
- Los datos que se generan en los procesos de la agricultura de precisión, las diferentes mediciones en una red de sensores dentro de un esquema IOT son abundantes con características de volumen, velocidad y variedad, los mismos que sometidos a la herramienta desarrollada permiten realizar análisis de confiabilidad con bajos márgenes de error para total entendimiento de los usuarios finales.
- La aplicación de cada algoritmo y su posterior análisis, dependerán del conjunto de datos que se le proporcionen a la herramienta, debido a que cada técnicas trabaja con una estructura diferente, algunos con campos de respuesta, otros con información transaccional y los resultados que se arrojen estarán acopladas a la necesidad que tenga el usuario para solventar cualquier aspecto dentro de un cultivo y qué podrá hacer para ejecutarlo de la mejor manera.
- La programación en R permitió llevar a cabo el proceso de minería de datos en una forma sólida, aplicando algoritmos de Machine Learning que, gracias a sus numerosas librerías que implementan funciones de estadística, métodos de aprendizaje supervisado y no supervisado, construcción de gráfica interactivas, acoplamiento de los datos cuando ocurra un evento que afecte al modelo armado, archivos de ayuda para cada librería que fue empleada y muchas características más que hacen grande la construcción de programas en este lenguaje. Se pudo diseñar varios análisis para cada técnica tanto descriptiva como predictiva.
- Trabajar con aplicaciones web en R es tan fácil, que solo se necesita instalar una librería muy potente para este propósito, Shiny. Se requiere poco o nada de una

línea de conocimiento en desarrollo web, para una persona que no está muy familiarizada con HTML, JavaScript o CSS, programar aquí, será una tarea sencilla, únicamente se debe tener el objetivo de lo que se va a analizar y los resultados que se obtendrán son tan buenos como la manera en el que se presentan.

- Es imprescindible trabajar con un lenguaje que tenga este nivel, debido a que se pueden realizar muchas tareas, trabajar en ambientes web y escritorio, realizar análisis más rápidamente, incluir librerías que ahorran el tiempo y líneas de código, amplia comunidad que está siempre en constante desarrollo para mejorar día a día la potencia de R. Todo esto, sumado a su legible forma de escribir código, a diferencia de otros programas que también tienen este propósito; lo más rescatable del trabajo con R, es que realiza el proceso de extracción de conocimiento de datos en una sola herramienta.
- Con la ayuda del plan de pruebas se pudo corregir algunas fallas tanto en la parte de interfaz de usuario como también en aspectos relacionados con el código fuente.

RECOMENDACIONES

- Un aspecto importante es tener claro el panorama que se pretende descubrir en un proyecto de Minería de datos, porque en base al objetivo planteado se podrá elegir el tipo de datos para su análisis, así mismo determinar las técnicas que involucran la resolución de problemas en la agricultura.
- Al utilizar la herramienta con cualquier análisis se debe acatar las sugerencias que puedan existir dentro de la aplicación; todo esto con el fin de una mejor experiencia en la manipulación de los datos y resultados sin ningún tipo de inconvenientes.
- Los datos que usará la herramienta deben ser confiables, deben presentarse libre de inconsistencias esto permitirá un correcto funcionamiento del algoritmo y por ende un óptimo resultado de análisis, asemejándose a la realidad que se pretende descubrir. También es importante tener conocimientos en estadísticas para que ayuden al entendimiento total del módulo implementado.
- Realizar un estudio entre diferentes herramientas que permitan hacer usos de técnicas de minería de datos, definir comparaciones entre sus pros y contras, de hecho existen más herramientas aparte de R que puedan realizar este propósito. Mediante dicho estudio se evaluarán aspectos que serán muy interesantes para seguir trabajando en esta área de la ingeniería tan importante
- La herramienta es muy general, por lo cual si se continua con el proyecto de IOTMACH se recomienda enfocarse más en un solo algoritmo, lo cual involucra muchas más cosas que en este prototipo no incluyen, como pueden ser pruebas de hipótesis, técnicas estadísticas más avanzada, construcción de gráficos en cada fase de la técnica seleccionada. Debido a que un determinado análisis permite conocer la situación de actual de los conjuntos de datos, se tiene que plantear un objetivo en el cual abarque un área en específico.
- Para las personas que quieran incursionar en la programación de aplicaciones estadística en R, se recomienda revisar una librería llamada swirl, la cual incorpora una serie de ejemplos, estructura de codificación y un asistente que evaluará los resultados de cada línea que se vaya escribiendo.

BIBLIOGRAFÍA

- [1] R. M. A. Yelitzza Josefina and Talavera Pereira, "Minería de Datos como soporte a la toma de decisiones empresariales," *Opción*, vol. 23, no. 52, pp. 104–118, 2007.
- [2] R. Roberto, G. Karina, and R. J. C., "Presentación: Minería de Datos," *Rev. Iberoam. Intel. Artif.*, no. 29, pp. 7–9, 2006.
- [3] L. S. Sight, "R-bloggers," 2016. [Online]. Available: <http://www.r-bloggers.com/whats-the-difference-between-machine-learning-statistics-and-data-mining/>.
- [4] A. Cama Pinto, E. De la Hoz Franco, and D. Cama Pinto, "Las redes de sensores inalámbricos y el internet de las cosas," *Inge Cuc*, vol. 8, no. 1, pp. 163–172, 2012.
- [5] S. R. Cabrero and C. Heliodoro, "Agricultura de precisión: ¿verdad o humo?," *Agric. Rev. Agropecu.*, no. 940, pp. 350–354, 2011.
- [6] R. Project, "The R Project for Statistical Computing," 2016. [Online]. Available: <https://www.r-project.org/>.
- [7] D. Becker, T. D. King, and B. McMullen, "Big data, big data quality problem," in *2015 IEEE International Conference on Big Data (Big Data)*, 2015, pp. 2644–2653.
- [8] D. E. Peacock and G. Hu, "Analyzing Grammy, Emmy, and Academy Awards Data Using Regression and Maximum Information Coefficient," in *2013 Second IIAI International Conference on Advanced Applied Informatics*, 2013, pp. 74–79.
- [9] W. I. Matters, *Predictive Analytics 101: Next-Generation Big Data Intelligence*. 2013.
- [10] D. Ramesh and B. V. Vardhan, "Data mining techniques and applications to agricultural yield data," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 2, no. 9, pp. 3477–3480, Jun. 2013.
- [11] S. A. Imam, A. Choudhary, and V. K. Sachan, "Design issues for wireless sensor networks and smart humidity sensors for precision agriculture: A review," in *2015 International Conference on Soft Computing Techniques and Implementations (ICSCIT)*, 2015, no. Table 1, pp. 181–187.
- [12] S. Ivanov, K. Bhargava, and W. Donnelly, "Precision Farming: Sensor Analytics," *IEEE Intell. Syst.*, vol. 30, no. 4, pp. 76–80, Jul. 2015.
- [13] D. Granados, "Denodo: la integración de datos financieros," 2012. [Online]. Available: <http://www.eoi.es/blogs/20dato/2012/05/31/denodo-la-integracion-de-datos-financieros/>.
- [14] B. Vargas Govea, *Introducción a R con fundamentos de minería de datos Contenido*. 2014.
- [15] P. Dixit and G. I. Prajapati, "Machine Learning in Bioinformatics: A Novel Approach for DNA Sequencing," in *2015 Fifth International Conference on Advanced Computing & Communication Technologies*, 2015, vol. 2015-April, pp. 41–47.
- [16] D. Vidhate and P. Kulkarni, "Cooperative Machine Learning with Information Fusion for Dynamic Decision Making in Diagnostic Applications," in *2012 International Conference on Advances in Mobile Network, Communication and Its Applications*, 2012, pp. 70–74.
- [17] M. A. Alsheikh, S. W. Lin, D. Niyato, and H. Tan, "Machine Learning in Wireless Sensor Networks: Algorithms, Strategies, and Applications," *IEEE Commun. Surv. Tutorials*, vol. 16, no. 4, pp. 1996–2018, 2014.
- [18] L. A. Calvo Valverde, "Estrategia basada en el aprendizaje de máquina para tratar con conjuntos de datos no etiquetados usando conjuntos aproximados y / o ganancia de información," *Rev. Tecnol. en Marcha*, vol. 29, no. 5, pp. 4–15, 2016.

- [19] S. Soares, "DATAVERSITY Data Education for Business and IT Professionals," 2012. [Online]. Available: http://www.dataversity.net/not-your-type-big-data-matchmaker-on-five-data-types-you-need-to-explore-today/?cm_mc_uid=07202058171914656643101&cm_mc_sid_50200000=1466444811. [Accessed: 03-Jun-2012].
- [20] Xindong Wu, Xingquan Zhu, Gong-Qing Wu, and Wei Ding, "Data mining with big data," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 1, pp. 97–107, Jan. 2014.
- [21] S. Madden, "From Databases to Big Data," *IEEE Internet Comput.*, vol. 16, no. 3, pp. 4–6, May 2012.
- [22] C. PEREZ LOPEZ and D. SANTIN GONZALEZ, *Minería de datos: técnicas y herramientas*. Paraninfo, 2007.
- [23] S. D. E. Manjula, "Analysis of Data Mining Techniques for Agriculture Data," *Int. J. Comput. Sci. Eng. Commun.*, vol. 4, no. 2, pp. 1311–1313, 2016.
- [24] C. Logreira, "Minería de datos y su incidencia en la toma de decisiones empresariales en el contexto de CRM," *Ing. Solidar.*, vol. 7, no. 12–13, pp. 68–71, 2011.
- [25] M. Ramichetty, "Linked IN," 2016. [Online]. Available: <https://www.linkedin.com/pulse/data-mining-vs-statistics-how-different-mahesh-ramichetty>. [Accessed: 06-Jun-2016].
- [26] F. Valenga, E. Fernández, H. Merlino, D. Rodríguez, C. Procopio, P. Britos, and R. García-martínez, "Minería de Datos Aplicada a la Detección de Patrones Delictivos en Argentina," in *VII Jornadas Iberoamericanas de Ingeniería del Software e Ingeniería del Conocimiento*, 2008, pp. 31–39.
- [27] J. A. Gallardo Arancibia, *Metodología para la definición de requisitos en proyectos de data mining (ER-DM)*. 2009.
- [28] S. I. Chowdhury, R. Gustavsson, and O. Wigertz, "Approaches for knowledge discovery in databases," in *Proceedings of ANZIIS '94 - Australian New Zealand Intelligent Information Systems Conference*, 1994, pp. 457–461.
- [29] DAEDALUS, *Data, decisions and Language*. Madrid, 2002.
- [30] A. Azevedo and M. F. Santos, "KDD, SEMMA and CRISP-DM: a parallel overview," *IADIS Eur. Conf. Data Min.*, no. January, pp. 182–185, 2008.
- [31] J. A. Freitez, "Desarrollo de un Modelo Predictivo del Brote de la Sigatoka Negra para las Plantaciones de Plátanos al Sur del Lago de Maracaibo," vol. 9, no. 1, pp. 191–198, 2009.
- [32] M. J. Moreno García, María; Miguel Quintales, Luis; García, Francisco J.; Polo Martín, "Obtención y Validación de Modelos de Estimación de Software Mediante Técnicas de Minería de Datos," *Rev. Colomb. Comput.*, vol. 3, no. 1, pp. 53–71, 2002.
- [33] S. Earley, "Big Data and Predictive Analytics: What's New?," *IT Prof.*, vol. 16, no. 1, pp. 13–15, Jan. 2014.
- [34] Anónimo, "Evaluando Software.com," 2013. [Online]. Available: <http://www.evaluandosoftware.com/que-es-el-analisis-predictivo-2/>.
- [35] J. M. D. Bruxella, S. Sadhana, and S. Geetha, "Categorization of Data Mining Tools Based on Their Types," *Int. J. Comput. Sci. Mob. Comput.*, vol. 3, no. 3, pp. 445–452, 2014.
- [36] O. Bello and S. Zeadally, "Intelligent Device-to-Device Communication in the Internet of Things," *IEEE Syst. J.*, vol. 10, no. 3, pp. 1172–1182, Sep. 2016.
- [37] C. Perera, R. Ranjan, L. Wang, S. U. Khan, and A. Y. Zomaya, "Big Data Privacy in the Internet of Things Era," *IT Prof.*, vol. 17, no. 3, pp. 32–39, May 2015.
- [38] CISCO, "Internet of Everything," 2014. [Online]. Available: http://www.cisco.com/web/ES/tomorrow-starts-here/ioe/tomorrow_starts_here_infographic.html.
- [39] IBM, "IBM Smart Cities," 2016. [Online]. Available: http://www.ibm.com/smarterplanet/es/es/smarter_cities/overview/.
- [40] M. Aragon, "Computación Ubicua y Domotica," in *Prezi.com*, 2012.

- [41] S. Dimitriadis and C. Goumopoulos, "Applying Machine Learning to Extract New Knowledge in Precision Agriculture Applications," in *2008 Panhellenic Conference on Informatics*, 2008, pp. 100–104.
- [42] S. T. Sonka, *Precision agriculture in the 21st century*. Washington D.C: National Academy Press, 1997.
- [43] M. Jiménez-López, Andrés Fernando; Jiménez-López, Fabián Rolando; Jiménez-López, "Multispectral analysis of vegetation for remote sensing applications Análisis multiespectral de la vegetación para aplicaciones de sensado remoto," *ITECKNE*, vol. 12, no. 2, pp. 156–167, 2015.
- [44] A. Jiménez, D. Ravelo, and J. Gómez, "Sistema de adquisición, almacenamiento y análisis de información fenológica para el manejo de plagas y enfermedades de un duraznero mediante tecnologías de agricultura de precisión System acquisition and analysis of information storage and handling phenol," *Tecnura*, vol. 14, no. 27, pp. 41–51, 2010.
- [45] N. INSTRUMENTS, "NATIONAL INSTRUMENTS," 2016. [Online]. Available: <http://www.ni.com/white-paper/7142/en/>. [Accessed: 05-May-2016].
- [46] J. Capella Hernández, *Redes inalámbricas de sensores: una nueva arquitectura eficiente y robusta basada en jerarquía dinámica de grupos*. Valencia: UNIVERSIDAD POLITECNICA DE VALENCIA, 2010.
- [47] J.-A. Sánchez-Alcón, L. López-Santidrián, and J.-F. Martínez, "Solución Para Garantizar La Privacidad En Internet De Las Cosas.," *Solut. to ensure Priv. internet things.*, vol. 24, no. 1, pp. 62–70, 2015.
- [48] I. Keka and M. Hamiti, "Load Profile Analyses Using R Language," in *Proceedings of the ITI 2013, 35th International Conference on Information Technology Interfaces*, 2013, pp. 245 – 250.
- [49] C. Wagner, S. Miller, and J. M. Garibaldi, "A fuzzy toolbox for the R programming language," in *2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011)*, 2011, pp. 1185–1192.
- [50] J. Kunal, "SAS vs. R (vs. Python) – which tool should I learn?," 2014. [Online]. Available: <http://www.analyticsvidhya.com/blog/2014/03/sas-vs-vs-python-tool-learn/>.
- [51] R. Project, "The Comprehensive R Archive Network," 2016. [Online]. Available: <https://cran.r-project.org/>.
- [52] R. Project, "RStudio," *info@rstudio.com*, 2016. [Online]. Available: <https://www.rstudio.com/products/RStudio/>.
- [53] Pro, "Shiny framework," 2016. [Online]. Available: <http://shiny.rstudio.com/>.
- [54] K. K. Nisa, H. A. Andrianto, and R. Mardhiyyah, "Hotspot clustering using DBSCAN algorithm and shiny web framework," in *2014 International Conference on Advanced Computer Science and Information System*, 2014, pp. 129–132.
- [55] T. Wang, F. Liu, W. Zhou, X. Zhu, and S.-U. Guan, "Neural incremental attribute learning based on principal component analysis," in *2016 IEEE International Conference on Big Data Analysis (ICBDA)*, 2016, pp. 1–5.
- [56] H. Fetanat, L. Mortazavifarr, and N. Zarshenas, "The analysis of agricultural data with regression data mining technique," *Cienc. e Nat.*, vol. 37, no. 2, pp. 102–107, 2015.
- [57] J. Cabezas, M. Galleguillos, and J. F. Perez-Quezada, "Predicting Vascular Plant Richness in a Heterogeneous Wetland Using Spectral and Textural Features and a Random Forest Algorithm," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 5, pp. 646–650, May 2016.
- [58] H. R. Seth and H. Banka, "Hardware implementation of Naïve Bayes classifier: A cost effective technique," in *2016 3rd International Conference on Recent Advances in Information Technology (RAIT)*, 2016, pp. 264–267.
- [59] R. Project, "Plotly," 2016. [Online]. Available: <https://plot.ly/r/>.

ÍNDICE

Agricultura, 6, 7, 16, 18, 22, 23, 26, 28, 35, 42, 43, 44, 45, 46, 47, 93, 95, 98
Big data, 7, 96, 97
Data mining, 8, 9, 97, 98
Dataset, 17
Descriptivo, 13, 14, 15, 26, 40, 47, 48, 51, 54, 56, 57, 58, 60, 70, 79, 83, 88, 101, 103,
105, 107
IOTMACH, 3, 7, 8, 10, 18, 23, 26, 27, 28, 29, 46, 95
Lenguaje R, 18
Machine learning, 8, 9
Predictivo, 12, 13, 14, 15, 23, 25, 26, 27, 40, 46, 47, 49, 52, 54, 60, 70, 81, 83, 88, 97,
102, 103, 106, 107
Shiny, 12, 13, 20, 28, 29, 52, 67, 68, 69, 93, 98

ANEXO A. Prueba unitaria: Administración de conexiones

Prueba Unitaria Ciclo o proceso: Módulo administración de conexiones Objetivo: Ingresar los datos que se van a utilizar para realizar las conexiones a las diferentes fuentes de datos para su posterior análisis. Tipo de prueba: Funcional [X] Técnica [] Módulo o programa: IOTMach Analysis Tiempo estimado: 1 hora				
Responsables		Firma		
Desarrollador				
Usuario final:.....				
Descripción de la prueba				
A través de la prueba se requiere validar las funciones que cumple el menú administrar conexiones, con los respectivos parámetros.				
Plan de pruebas				
N.	Descripción	Observación	Cumple	
			SI	NO
1	Ingresar información para conectarse al servidor de base de datos	Tipo de conexión base de datos		
2	Probar conexión antes de guardar los datos			
3	Guardar el registro de conexión	Probar conexión para activar esta opción		
4	Ingresar información para crear conexión a archivos externos	Tipo de conexión fuentes externas		
5	Probar conexión antes de guardar los datos			
6	Guardar el registro de conexión	Probar conexión para activar esta opción		
Observaciones al plan de pruebas				
Criterios de prueba				

Fuente: Elaboración propia

ANEXO B. Prueba unitaria: Análisis descriptivo

Prueba Unitaria Ciclo o proceso: Módulo análisis descriptivo Objetivo: Realizar análisis descriptivo utilizando los algoritmos disponibles con cualquier fuente de datos. Tipo de prueba: Funcional [X] Técnica [] Módulo o programa: IOTMach Analysis Tiempo estimado: 1 hora				
Responsables		Firma		
Desarrollador				
Usuario final:.....				
Descripción de la prueba				
A través de la prueba se requiere validar las funciones que cumple el menú análisis descriptivo, con los respectivos parámetros.				
Plan de pruebas				
N.	Descripción	Observación	Cumple	
			SI	NO
1	Seleccionar una fuente de datos	Puede ser conexión, local o ejemplo		
2	Dar clic en usar datasets seleccionado	Habilitará pestaña variables		
3	Elegir los campos a utilizar			
4	Dar clic en aceptar	Habilitará las demás pestañas y resultados		
5	Volver y cargar otra fuente de datos			
Observaciones al plan de pruebas				
Criterios de prueba				

Fuente: Elaboración propia

ANEXO C. Prueba unitaria: Análisis predictivo

Prueba Unitaria

Ciclo o proceso: Módulo análisis predictivo

Objetivo: Realizar análisis predictivo utilizando los algoritmos disponibles con cualquier fuente de datos.

Tipo de prueba: Funcional [X] Técnica []

Módulo o programa: IOTMach Analysis

Tiempo estimado: 1 hora

Responsables	Firma
Desarrollador	
Usuario final:.....	

Descripción de la prueba

A través de la prueba se requiere validar las funciones que cumple el menú análisis predictivo, con los respectivos parámetros.

Plan de pruebas

N.	Descripción	Observación	Cumple	
			SI	NO
1	Seleccionar una fuente de datos	Puede ser conexión, local o ejemplo		
2	Dar clic en usar datasets seleccionado	Habilitará pestaña variables		
3	Elegir los campos a utilizar			
4	Dar clic en aceptar	Habilitará las demás pestañas y resultados		
5	Volver y cargar otra fuente de datos			

Observaciones al plan de pruebas

Criterios de prueba

ANEXO D. Prueba unitaria: Repositorio local

Prueba Unitaria Ciclo o proceso: Repositorio local Objetivo: Comprobar que los archivos se almacenen en el repositorio de la aplicación para su posterior análisis Tipo de prueba: Funcional [X] Técnica [] Módulo o programa: IOTMach Analysis Tiempo estimado: 1 hora				
Responsables		Firma		
Desarrollador				
Usuario final:.....				
Descripción de la prueba				
A través de la prueba se requiere validar las funciones que cumple para subir un archivo, con los respectivos parámetros.				
Plan de pruebas				
N.	Descripción	Observación	Cumple	
			SI	NO
1	Seleccionar un análisis	Puede ser descriptivo o predictivo		
2	En la pestaña orígenes de datos elegir repositorio local			
3	Ingresar los parámetros requeridos	Se puede almacenar en un directorio existente o crear uno nuevo		
4	Dar clic en guardar			
5	Verificar su existencia en el panel Vista previa			
Observaciones al plan de pruebas				
Criterios de prueba				

Fuente: Elaboración propia

ANEXO E. Prueba de usabilidad: Interfaces de usuario

Prueba de Usabilidad Ciclo o proceso: Interfaces de usuario Objetivo: verificar si la aplicación es fácil de utilizar y qué tan amigable puede ser. Tipo de prueba: Funcional [<input type="checkbox"/>] Técnica [<input checked="" type="checkbox"/>] Módulo o programa: IOTMach Analysis Tiempo estimado: 1 hora					
Responsables		Firma			
Desarrollador					
Usuario final:					
Descripción de la prueba					
A través de la prueba se busca conocer la aceptación que tiene la aplicación, con los aspectos de amigabilidad y facilidad de uso					
Plan de pruebas					
N.	Descripción	Observación	Cumple		Resultados
			SI	NO	
1	La ayuda proporcionada en la aplicación es clara				
2	Existen falta de ortografía o inconsistencia en los textos				
3	Los mensajes generados son entendibles				
4	Es fácil la navegabilidad entre los menús				
5	Los gráficos generados son legibles y llamativos				
6	El contenido se adapta conforme el tamaño del dispositivo o navegador				
7	Se entiende el funcionamiento general de la aplicación				
Observaciones al plan de pruebas					
Criterios de prueba					

Fuente: Elaboración propia

ANEXO F. Prueba de integración: Análisis descriptivo

Prueba de Integración Ciclo o proceso: Módulo análisis descriptivo Objetivo: Utilizar un dataset creado en el menú Administración de conexiones para realizar el respectivo análisis con los algoritmos descriptivos Tipo de prueba: Funcional [X] Técnica [] Módulo o programa: IOTMach Analysis Tiempo estimado: 1 hora				
Responsables		Firma		
Desarrollador				
Usuario final:.....				
Descripción de la prueba				
A través de la prueba se requiere validar las funciones que cumple el menú administrar conexiones, con los respectivos parámetros.				
Plan de pruebas				
Descripción	Resultado esperado	Cumple		
		S	P	N
1. Acceder al menú administración de conexiones 2. Ingresar los campos requeridos 3. Crear nueva conexión 4. Acceder al menú descriptivo y elegir un algoritmo 5. En la pestaña orígenes de datos seleccionar conexión creada 6. Utilizar el datasets seleccionado para continuar con el análisis de los datos	- Crear conexión exitosamente - Listar todas las conexiones creadas para ser utilizadas en los demás menús	SATISFACTORIO	POCO SATISFACTORIO	NO SATISFACTORIO
Resultado obtenido				
Observaciones/Recomendaciones				

Fuente: Elaboración propia

ANEXO G. Prueba de integración: Análisis predictivo

Prueba de Integración Ciclo o proceso: Módulo análisis predictivo Objetivo: Utilizar un dataset creado en repositorio local para realizar el respectivo análisis con los algoritmos descriptivos Tipo de prueba: Funcional [X] Técnica [] Módulo o programa: IOTMach Analysis Tiempo estimado: 1 hora				
Responsables	Firma			
Desarrollador				
Usuario final:.....				
Descripción de la prueba A través de la prueba se requiere comprobar que los archivos subidos en el repositorio local sean accedidos por los usuarios y sin tener pérdida de datos.				
Plan de pruebas				
Descripción	Resultado esperado	Cumple		
		S	P	N
1. Acceder a menú análisis predictivo 2. Elegir un algoritmo cualquiera 3. En la pestaña orígenes de datos seleccionar repositorio local en la lista de opciones 4. Ingresar los parámetros requeridos 5. Proceder a subir el archivo 6. Comprobar si el archivo se encuentra en la lista de todos los archivos subidos en la aplicación	<ul style="list-style-type: none"> - El archivo se subió sin ningún problema - Listar los archivos y mostrar los datos mediante una vista previa 	SATISFACTORIO	POCO SATISFACTORIO	NO SATISFACTORIO
Resultado obtenido <div style="border: 1px solid black; height: 40px; width: 100%;"></div>				
Observaciones/Recomendaciones <div style="border: 1px solid black; height: 40px; width: 100%;"></div>				

Fuente: Elaboración propia

ANEXO H. Prueba de integración: Análisis con datasets de ejemplo

Prueba de Integración Ciclo o proceso: Módulo análisis predictivo/descriptivo Objetivo: Utilizar un dataset de ejemplo para realizar el respectivo análisis con los algoritmos descriptivos Tipo de prueba: Funcional [X] Técnica [] Módulo o programa: IOTMach Analysis Tiempo estimado: 1 hora																					
Responsables		Firma																			
Desarrollador																					
Usuario final:.....																					
Descripción de la prueba A través de la prueba se requiere comprobar que los archivos de ejemplo sean accedidos por los usuarios y sin tener pérdida de datos.																					
Plan de pruebas <table border="1" style="width: 100%; border-collapse: collapse; margin-top: 5px;"> <thead> <tr> <th rowspan="2" style="width: 45%;">Descripción</th> <th rowspan="2" style="width: 30%;">Resultado esperado</th> <th colspan="3" style="text-align: center;">Cumple</th> </tr> <tr> <th style="text-align: center;">S</th> <th style="text-align: center;">P</th> <th style="text-align: center;">N</th> </tr> </thead> <tbody> <tr> <td style="padding: 5px;"> 1. Acceder a menú análisis predictivo 2. Elegir un algoritmo cualquiera 3. En la pestaña orígenes de datos seleccionar ejemplo en la lista de opciones 4. Comprobar la existencia del archivo de ejemplo mediante la vista previa del mismo 5. Utilizar el dataset para realizar el análisis </td> <td style="padding: 5px; vertical-align: middle;"> - Cargar el ejemplo para que funcione con el algoritmo seleccionado </td> <td style="text-align: center; vertical-align: middle;">S</td> <td style="text-align: center; vertical-align: middle;">P</td> <td style="text-align: center; vertical-align: middle;">N</td> </tr> <tr> <td></td> <td></td> <td style="text-align: center;">SATISFACTORIO</td> <td style="text-align: center;">POCO SATISFACTORIO</td> <td style="text-align: center;">NO SATISFACTORIO</td> </tr> </tbody> </table>				Descripción	Resultado esperado	Cumple			S	P	N	1. Acceder a menú análisis predictivo 2. Elegir un algoritmo cualquiera 3. En la pestaña orígenes de datos seleccionar ejemplo en la lista de opciones 4. Comprobar la existencia del archivo de ejemplo mediante la vista previa del mismo 5. Utilizar el dataset para realizar el análisis	- Cargar el ejemplo para que funcione con el algoritmo seleccionado	S	P	N			SATISFACTORIO	POCO SATISFACTORIO	NO SATISFACTORIO
Descripción	Resultado esperado	Cumple																			
		S	P	N																	
1. Acceder a menú análisis predictivo 2. Elegir un algoritmo cualquiera 3. En la pestaña orígenes de datos seleccionar ejemplo en la lista de opciones 4. Comprobar la existencia del archivo de ejemplo mediante la vista previa del mismo 5. Utilizar el dataset para realizar el análisis	- Cargar el ejemplo para que funcione con el algoritmo seleccionado	S	P	N																	
		SATISFACTORIO	POCO SATISFACTORIO	NO SATISFACTORIO																	
Resultado obtenido <div style="border: 1px solid black; height: 40px; width: 100%;"></div>																					
Observaciones/Recomendaciones <div style="border: 1px solid black; height: 40px; width: 100%;"></div>																					

Fuente: Elaboración propia

ANEXO I. Prueba de rendimiento: Carga

Prueba de rendimiento (Carga)

Ciclo o proceso: Repositorio local/Conexiones

Objetivo: Comprobar la capacidad de la aplicación para su correcto funcionamiento bajo distintas condiciones de carga

Tipo de prueba: Funcional [X] Técnica []

Módulo o programa: IOTMach Analysis

Tiempo estimado: 2 horas

Responsables	Firma
Desarrollador	
Usuario final:.....	

Descripción de la prueba

A través de la prueba se medirá el tiempo que la aplicación tarda en cargar y visualizar la información de un determinado dataset o un archivo subido en el repositorio local.

Plan de pruebas

N.	Descripción	Tiempo de respuesta (*)	Observación
1a	Crear una conexión con un dataset menor a 1 Mb		
1b	Visualizar el dataset		
2a	Crear una conexión con un dataset menor a 10 Mb		
2b	Visualizar el dataset		
3a	Crear una conexión con un dataset menor a 30 Mb		
3b	Visualizar el dataset		

* Medido en segundos (s)

Observaciones al plan de pruebas

Criterios de prueba

Fuente: Elaboración propia

ANEXO J. Prueba de rendimiento: Concurrencia

Prueba de rendimiento (Concurrencia) Ciclo o proceso: Repositorio local/Conexiones/Análisis Objetivo: Comprobar la capacidad de la aplicación bajo cierta cantidad de peticiones esperadas, de esta manera detectar posibles cuellos de botellas. Tipo de prueba: Funcional [X] Técnica [] Módulo o programa: IOTMach Analysis Tiempo estimado: 2 horas				
Responsables		Firma		
Desarrollador				
Usuario final:.....				
Descripción de la prueba A través de la prueba se verificará si en la aplicación existen limitaciones o algún tipo de inconveniente para su correcto funcionamiento.				
Plan de pruebas				
N.	Descripción	Observación	Cumple	
			SI	NO
1	Al ingresar desde varios dispositivos, se observa que la aplicación demora más de lo normal en realizar su operación			
2	La conexión inalámbrica afecta o cuelga las funciones de la aplicación			
3	Los tiempos de espera para realizar consultas o crear conexiones a una fuente de datos son cortos			
4	Existe tiempo de inactividad de la aplicación cuando se la deja de utilizar			
Observaciones al plan de pruebas <div style="border: 1px solid black; height: 40px; width: 100%;"></div>				
Criterios de prueba <div style="border: 1px solid black; height: 40px; width: 100%;"></div>				

Fuente: Elaboración propia

ANEXO K. Prueba de regresión: Estimación de carga

Prueba de Regresión

Ciclo o proceso: Repositorio local/Conexiones

Objetivo: Comprobar si se mantienen los niveles medidos en las pruebas de carga tras un posible cambio en la aplicación (tiempos de respuesta)

Tipo de prueba: Funcional [X] Técnica []

Módulo o programa: IOTMach Analysis

Tiempo estimado: 2 horas

Responsables	Firma
Desarrollador	
Usuario final:.....	

Descripción de la prueba

A través de la prueba se observará si existen cambios adversos en el sistema si la capacidad de un dataset aumenta

Plan de pruebas

N.	Descripción	Tiempo de respuesta (*)	Observación
1a	Lectura dataset 1.16 Mb		
1b	Visualizar el dataset		
2a	Lectura dataset 2.45 Mb		
2b	Visualizar el dataset		
3a	Lectura dataset 19.5Mb		
3b	Visualizar el dataset		
4a	Lectura dataset 28.7 Mb		
4b	Visualizar el dataset		
5a	¿Lectura dataset > 40 Mb?		Realizar la predicción del tiempo de respuesta un dataset mayor a 40 Mb
5b	¿Visualizar el dataset?		

* Medido en segundos (s)

Observaciones al plan de pruebas

Criterios de prueba

Fuente: Elaboración propia