**Simulador de datos climáticos para el desarrollo de proyectos IoT**

Camilo Andrés Diaz Gómez

Juan Esteban Contreras Diaz

Jhonatan Mauricio Villarreal Corredor

Trabajo de Grado presentado para optar al título de Ingeniero de Sistemas

Asesor: Nikolay Lenin Reyes Jalizev, Magíster (MSc) en Ingeniería Industrial

****

Universidad de San Buenaventura

Facultad de Ingeniería (Bogotá)

Ingeniería de Sistemas

Bogotá D.C., Colombia

2023

|  |  |
| --- | --- |
| Citar/How to cite | (Diaz et al., 2023) |
| Referencia/Reference  Estilo/Style:  APA 7ma ed. (2020) | Diaz, C.A; Contreras, J.E y Villarreal, J. M. (2023). *Simulador de datos climáticos para el desarrollo de proyectos IoT*. [Trabajo de grado profesional]. Universidad de San Buenaventura Seleccione sede / seccional y/o extensión. |

|  |  |
| --- | --- |
|  | Biblioteca Digital (Repositorio)  www.bibliotecadigital.usb.edu.co |

**Bibliotecas Universidad de San Buenaventura**

Biblioteca Fray Alberto Montealegre O.F.M. - Bogotá.

Biblioteca Fray Arturo Calle Restrepo O.F.M. **-** Medellín, Bello, Armenia, Ibagué.

Departamento de Biblioteca - Cali.

Biblioteca Central Fray Antonio de Marchena – Cartagena.

**Universidad de San Buenaventura Colombia** - www.usb.edu.co

Bogotá **-** www.usbbog.edu.co

Medellín **-** www.usbmed.edu.co

Cali -www.usbcali.edu.co

Cartagena - www.usbctg.edu.co

Editorial Bonaventuriana - www.editorialbonaventuriana.usb.edu.co

Revistas científicas – www.revistas.usb.edu.co

**Dedicatoria**

Texto de dedicatoria centrado.

**Agradecimientos**

Texto de agradecimientos centrado.

Tabla de contenido

[Resumen 10](#_Toc129721736)

[Abstract 12](#_Toc129721737)

[Introducción 14](#_Toc129721738)

[Planteamiento del Problema 16](#_Toc129721739)

[Antecedentes 17](#_Toc129721740)

[Justificación 20](#_Toc129721741)

[Objetivos 22](#_Toc129721742)

[Objetivo General 22](#_Toc129721743)

[Objetivos Específicos 22](#_Toc129721744)

[Marco Teórico 23](#_Toc129721745)

[Internet of Things (IoT) 23](#_Toc129721746)

[Analítica de Datos 31](#_Toc129721747)

[Simulación 37](#_Toc129721748)

[Machine Learning 45](#_Toc129721749)

[Metodología 49](#_Toc129721750)

[Comprensión del negocio 49](#_Toc129721751)

[Comprensión de datos 49](#_Toc129721752)

[Preparación de datos 49](#_Toc129721753)

[Modelado 50](#_Toc129721754)

[Evaluación 50](#_Toc129721755)

[Despliegue. 50](#_Toc129721756)

[Resultados 57](#_Toc129721757)

[Conclusiones 62](#_Toc129721758)

[Recomendaciones 63](#_Toc129721759)

[Referencias 64](#_Toc129721760)

[Anexos 71](#_Toc129721761)

**Lista de Tablas**

[Tabla 1 32](#_Toc129628014)

[Tabla 2 52](#_Toc129628015)

[Tabla 3 58](#_Toc129628016)

**Lista de Figuras**

[Figura 1](#_Toc129721771) *[Árbol de Problema](#_Toc129721771)* [21](#_Toc129721771)

[Figura 2](#_Toc129721772) *[Modelo general de la arquitectura de una red IoT](#_Toc129721772)* [26](#_Toc129721772)

[Figura 3](#_Toc129721773) *[Funcionamiento de actuadores IoT](#_Toc129721773)* [30](#_Toc129721773)

[Figura 4](#_Toc129721774) *[Cadena de valor](#_Toc129721774)* [34](#_Toc129721774)

[Figura 5](#_Toc129721775) *[Estructura de la metodología CRISP-DM](#_Toc129721775)* [50](#_Toc129721775)

[Figura 6](#_Toc129721776) *[Temperaturas obtenidas por modelo](#_Toc129721776)* [59](#_Toc129721776)

**Lista de Anexos**

[Anexo 1 71](#_Toc129628046)

## Resumen

Las redes IoT1, han permitido mejorar diferentes aspectos para la humanidad, con el uso de ecosistemas tecnológicos que de manera eficiente, recopilen e intercambien información o datos que permitan aumentar mejorar la calidad de vida, aumentar la productividad en las grandes empresas, tener un control más preciso de los procesos, entre muchas funciones. Sin embargo, la implementación de este tipo de redes puede llegar a ser muy costoso, por lo que es preferible hacer un diseño previamente que permita validar y probar el correcto funcionamiento antes de su construcción e implementación.

Ya que este tipo de redes están compuestos por dispositivos que se encargan de recopilar información del exterior o el ambiente que los rodea, llamados sensores, la compra de estos dispositivos conlleva a un gasto monetario elevado, sin embargo, es posible generar los datos ambientales con los que la red podría trabajar, así que al simular los datos que harán funcionar la red, también es posible simular toda una red IoT basada en datos ambientales. Para que los datos sean consistentes entre si es necesario desarrollar un modelo de simulación que genere los datos apropiados teniendo en cuenta los otros aspectos del ambiente que los rodean.

Para lograr una simulación precisa de los datos ambientales, es necesario crear un modelo que tenga en cuenta las diferentes variables del ambiente. En este sentido, el Aprendizaje Automático (Machine Learning), una disciplina de la inteligencia artificial, es clave para entrenar el modelo usando datos recopilados de diversas estaciones hidrometeorológicas de la ciudad de Bogotá. Al obtener todos los datos acumulados por dichas estaciones, se podrán relacionar entre sí, considerando otras variables obtenidas de cada estación, como la fecha y hora en la que se extrajeron los datos. De esta manera, se creará un conjunto de datos organizado con el que el modelo será entrenado. Finalmente, el modelo será capaz de predecir los datos ambientales en base a la información del entorno que rodea la simulación.

*Palabras clave*: Redes IoT, ecosistemas tecnológicos, sensores, datos ambientales, modelo de simulación, Machine Learning, inteligencia artificial, consistencia de datos, predicción de datos.

## Abstract

IoT networks have allowed for improving different aspects for humanity, through the use of technological ecosystems that efficiently collect and exchange information or data to increase the quality of life, productivity in large companies, more precise control of processes, among many functions. However, the implementation of this type of network can be very expensive, so it is preferable to design it beforehand to validate and test proper functioning before construction and implementation.

Since these types of networks are composed of devices that are responsible for collecting information from the outside or the environment that surrounds them, called sensors, the purchase of these devices entails a high monetary expense. However, it is possible to generate environmental data with which the network could work, so by simulating the data that will make the network work, it is also possible to simulate an entire IoT network based on environmental data. In order for the data to be consistent with each other, it is necessary to develop a simulation model that generates the appropriate data taking into account the other aspects of the surrounding environment.

To achieve a correct simulation of environmental data, it is necessary to create a model that takes into account the different variables of the environment. For this reason, the use of Machine Learning, one of the branches of artificial intelligence, is the tool that will allow training the model using data collected from different stations in the city of Bogotá. By obtaining all the data accumulated by these stations, they can be related to each other, thanks to other variables obtained from each station, for example, the date and time in which the data were extracted by the station. This is done in order to create an organized data set with which the model will be trained. Finally, the model will be able to predict the data based on information from the environment surrounding the simulation.

*Keywords***:** IoT networks, technological ecosystems, sensors, environmental data, simulation model, Machine Learning, artificial intelligence, data consistency, data prediction.

## Introducción

Con la aparición del internet, el crecimiento tecnológico de las comunicaciones ha avanzado considerablemente, permitiendo así que las personas puedan mantenerse comunicadas fácilmente, además, estos grandes avances han permitido llegar a áreas que, hasta hace algunos años, era innecesaria una comunicación, pero al ver el valor que podían tener los datos, surgió la necesidad de que los objetos como electrodomésticos y dispositivos de uso industrial puedan recopilar e intercambiar entre si información permitiendo expandir aún más los usos que se pueden hacer con el internet y las redes de comunicaciones que lo componen, a esto se le conoce como Internet of Things (IoT) o también como internet de las cosas.

IoT permite que los dispositivos estén integrados en una red, intercambiando datos recopilados entre sí para aportar más beneficios a las personas e incluso las industrias actuales, pero al requerir una correcta integración de dispositivos que cumplan objetivos que aporten a la red los datos necesarios para su correcto funcionamiento y lograr aprovechar al máximo los beneficios que este tipo de redes son capaces de aportar, es importante que estén diseñadas de la mejor manera posible, y así mismo probar que su funcionamiento sea correcto. Sin embargo, los costos de este tipo de redes y los dispositivos que la componen pueden llegar a ser exageradamente caros.

Por esta razón, el proyecto “Simulador de datos climáticos para el desarrollo de proyectos IoT”; se obtendrá un modelo para la generación de datos climáticos, inicialmente temperatura, con ello podrán ser usados para posteriores proyectos donde sea necesario una validación de funcionamiento de dispositivos IoT. Para esto se desarrollarán algoritmos, modelos y pruebas para generar un óptimo funcionamiento del modelo, con ello, llegar al objetivo general y suplir la necesidad de los desarrolladores, investigadores y técnicos de redes IoT que buscan probar sus dispositivos IoT previo a la instalación de las redes en la ciudad de Bogotá para lugares cercanos al Jardín Botánico, Universidad Nacional y a la sede central del IDEAM.

## Planteamiento del Problema

En la actualidad, un gran número de dispositivos electrónicos de uso diario, como celulares, computadores, televisores, neveras, aires acondicionados, sensores, entre otros, están conectados a internet con diversos propósitos. A este concepto se le conoce como IoT “Internet of Things”, el cual, “se refiere a la interconexión en red de objetos cotidianos, que a menudo están equipados con inteligencia ubicua” (Xia, Yang, Wang & Vinel, 2012).

Gracias a esto, la conexiones IoT se han enlazado a la vida cotidiana por el avance exponencial y transversal de las tecnologías en la sociedad, que han evolucionado a una velocidad en la que muchas personas, empresas, negocios, entre otros, están adquiriendo estas nuevas tecnologías con el fin de mejorar el rendimiento en diferentes aspectos.

Los beneficios son amplios, adelante en el documento se podrán encontrar, y esto es gracias a que los dispositivos IoT están encargados de la obtención de datos y su envió a la nube (cloud), permitiendo la conexión e intercambio de información entre estos objetos.

Según el centro de investigación SAP, estos objetos interconectados están perfectamente integrados a la red de información, lo que hace que se pueda interactuar con los mismos a través de internet, pudiendo consultar o editar su estado a tiempo real (Sandy &Abasolo, 2013)

Actualmente, hay demasiadas personas que se desempeñan en el diseño, instalación y mantenimiento de estas conexiones como desarrolladores, investigadores, técnicos, entre otros en las redes IoT, estos a su vez buscan alguna ayuda para poder desarrollar sus pruebas sin la necesidad de gastar muchos recursos; por esta razón, una de las dificultades recurrentes para la realización de estas conexiones y su posterior análisis es su alto costo económico la cual limitan la optimización, eficiencia y el tiempo en la construcción de estas conexiones que es una de las razones por la que estos proyectos de redes IoT pueden costar más dependiendo del propósito y presupuestos.

### **Antecedentes**

En la actualidad existe una gran variedad de plataformas y proyectos que están dirigidos a la contextualización, implementación y desarrollo de las redes IoT gracias al gran crecimiento en la actualidad junto al desarrollo de la tecnología. En el mercado se puede encontrar con software como:

#### **Cisco Packet Tracer**

Es un software dirigido especialmente a la enseñanza y aprendizaje del comportamiento de las redes. En esta herramienta se puede desarrollar y simular redes gráficamente; uno de sus grandes servicios o usabilidad es la posibilidad de simular redes IoT, en donde se pueden encontrar distintos dispositivos tanto domésticos como industriales.

#### **IoT Device Simulator (AWS)**

Gracias a la plataforma AWS (Amazon Web Services) la cual ofrece servicios en la web como lo dice su propio nombre, se puede encontrar el servicio IoT Device Simulator, en el cual, como lo menciona “ayuda a los clientes a probar la integración de dispositivos y a mejorar el rendimiento de sus servicios Backend de IoT, a través de una interfaz gráfica basada en la web. La solución permite a los clientes crear y simular cientos de dispositivos conectados, sin tener que configurar y administrar dispositivos físicos, o desarrollar scripts que consumen mucho tiempo.” (AWS, s. f.).

#### **Bluemix (IBM)**

**Es una servicio completamente almacenado y gestionado en la nube** que facilita la derivación de valor de los dispositivos de Internet de las cosas (IoT)**, para el uso de este Bluemix es necesarios tener los dispositivos en físico, al tenerlos se conectan con la plataforma y esta empezara enviar datos de forma segura por medio de MQQT,** Puede configurar y gestionar los dispositivos mediante el panel de control en línea(IBM , s.f.)**.**

#### **Iotify**

**Simulador IoT que te permite desarrollar rápidamente soluciones IoT en la nube. Es una herramienta de simulación IoT de gran crecimiento. Es un software libre, todo esto con la creación de proyectos de simulación, esto llevo a la creación de proyectos gratis pero después cambiaron a tener una membrecía para poder utilizarla** (Hardwarelibre, s.f.)**.**

#### **Matlab**

**Una herramienta que tiene módulos de simulación de IoT para poder desarrollar y modelar dispositivos inteligentes,** también puedes crear prototipos de tus dispositivos inteligentes, también permite la manipulación de archivos y su visualización.

De igual manera, muchos otros proyectos se encuentran dirigidos al desarrollo o implementación de redes IoT, como lo pueden ser:

#### **Control y simulación de una planta piloto de laboratorio docente con integración de plataformas IoT para subida de datos a la nube**

Este trabajo de grado es una continuación del anterior TFG, que lleva por nombre Monitorización y Seguimiento de Simuladores de Procesos Industriales con Fines Educativos creado por John Paúl Mayorga Jines. En TFG, el control y la monitorización se realizan en SIMATIC Manager y WinCC Flexible 2008.TFG se divide en dos partes: simular el modelo en la simulación SIMIT y subir los datos a la nube por medio de la plataforma IBM Cloud, tanto en la simulación como en el modelo real. Los datos del modelo real se cargan a través de la puerta de enlace IoT2040.

#### **Simulación realista de comunicaciones IoT en entornos urbanos**

Pensando en la implementación de ciudades inteligentes, este proyecto utiliza redes IoT para que los dispositivos que integraran la red que abarca la ciudad estén interconectados, ya que al tener un bajo consumo energético y un rango de conexión de larga distancia permite ahorrar costes en instalación y mantenimiento.

Este es un trabajo de grado donde se presenta una solución que incluye un simulador de red para medir el rendimiento en las comunicaciones LPWAN mediante un entorno regulado, un motor 3D para la construcción de este entorno y un motor 3D. Incluyendo el trazado de rayos que ayuda a mejorar los patrones de propagación y ofrece los resultados de rendimiento esperados cuando se utiliza la tecnología de red de área amplia (LoRaWAN).

##### Análisis de la simulación de dispositivos, circuitos y sistemas electrónicos paraInternet de las cosas (IoT)

**Los dispositivos que componen una red IoT son conocidos como nodos, al funcionar con energía eléctrica, por lo que puede llegar a ser importante cuanta energía eléctrica puede llegar a consumir una red IoT, por esto, este proyecto de grado pretende simular el consumo energético de los nodos, tomando como referencia la plataforma hardware Cookies, desarrollada en el Centro de Electrónica industrial de la universidad Politécnica de Madrid.**

**Este proyecto ha sido realizado mediante modelos de consumo parametrizables permitiendo así al usuario adaptar dichos modelos a las especificaciones de una manera correcta en una red IoT.**

Con lo anterior, se generó la siguiente pregunta:

¿Cómo simular datos ambientales mediante el uso de Analítica de Datos para ayudar en la generación de proyectos para modelos IoT?

## Justificación

Las redes IoT están compuestas por dispositivos como: sensores, actuadores, cloud, entre otros. Estos dispositivos se encargan de recopilar e intercambiar información entre ellos y de esta manera cumplir con los propósitos planteados por los usuarios. Por esta razón es muy importante seleccionar los dispositivos que más se adapten y cumplan los propósitos con los que la red IoT está construida; con esto es necesario previamente diseñar la red para posteriormente realizar las pruebas y verificaciones del correcto funcionamiento y que cumpla con las necesidades planteadas en su diseño, a su vez se deben probar cada uno de los dispositivos que integran la red para verificar que cumplan con los propósitos para los que fue creada. Pero cuando las pruebas no cumplen con las necesidades planteadas, es necesario reestructurar y rediseñar la red de nuevo e iniciar el proceso desde el principio, generando así perdidas de recursos como tiempo y dinero entre otros, es por esto que se tiene una estructura de un árbol de problema que permite explicar mejor la justificación del proyecto (ver Figura 1).

**Figura 1** *Árbol de Problema*

Diagrama

Descripción generada automáticamente

*Fuente: Elaboración propia.*

## Objetivos

### **Objetivo General**

Realizar un modelo para simular datos IoT mediante un lenguaje de programación para la posterior validación de estos y así facilitar pruebas de proyectos IoT para una región de Colombia.

### **Objetivos Específicos**

* + Diseñar los algoritmos de simulación para los diferentes sensores y actuadores.
  + Desarrollar los modelos de simulación de sensores IoT sobre un lenguaje de programación que permitirá a los usuarios hacer uso de estos.
  + Generar datos e información mediante los algoritmos desarrollados.

Realizar pruebas funcionales para comprobar la validez de los datos simula

## Marco Teórico

### **Internet of Things (IoT)**

Internet of Things o también conocido como el Internet de las cosas es un nuevo paradigma del mundo moderno el cual es la conexión de varios nodos IoT que contienen integrados sensores los cuales son los dispositivos, por ejemplo, de recopilación de información, electrodomésticos, celulares, computadores, entre otros que tienen la capacidad de conectarse al internet, que recopilan información de su entorno, por ejemplo, sensores de temperatura, gas, humo, humedad, velocidad del viento, etc. Por otro lado, están los actuadores que son los que reciben como dato de entrada la información proporcionada por los sensores, estos pueden ser enseres domésticos, motores, ventiladores y demás. Esta información transmitida desde el sensor al actuador se hace por medio de las puertas de enlace (gateways) y de esta manera llegar a las plataformas (software) para su respectivo procesamiento y así suplir el objetivo final de la red. En la actualidad, este concepto es muy ubicuo en el día a día, aunque no sea muy percibido, ya que la mayoría de las cosas de uso diario como lo son los electrodomésticos, celulares, entre otras cosas; lo integran (Xia, Yang, Wang, & Vinel, 2012) (CambioDigital, 2018).

Actualmente, IoT es muy versátil y se puede aplicar a varios usos, por lo que la mayoría de los dispositivos lo integran. Como resultado, la humanidad anticipa un cambio tecnológico donde IoT se volverá cada vez más importante e indispensable para la vida diaria, similar a la situación actual.

#### **Áreas de Aplicación**

Las redes IoT tienen una amplia gama de áreas de aplicación en diferentes industrias y dominios, algunas de las cuales incluyen

##### ***Hogares inteligentes***

Las redes IoT se pueden usar para automatizar y controlar electrodomésticos y dispositivos, como termostatos, sistemas de iluminación, cámaras de seguridad y sistemas de entretenimiento.

##### ***Automatización industrial***

Las redes IoT se pueden usar para monitorear y controlar máquinas y equipos en fábricas, almacenes y otros entornos industriales, mejorando la eficiencia y reduciendo el tiempo de inactividad.

##### ***Atención médica***

las redes IoT se pueden usar para monitorear pacientes de forma remota, recopilar datos de salud de dispositivos portátiles y rastrear equipos médicos en hospitales.

##### ***Agricultura***

Las redes IoT se pueden usar para monitorear los niveles de humedad del suelo, la temperatura y otros factores ambientales para optimizar el rendimiento de los cultivos y reducir el uso de agua.

##### ***Transporte***

Las redes IoT se pueden usar para rastrear la ubicación y el estado de los vehículos, monitorear el flujo de tráfico y optimizar las operaciones logísticas.

##### ***Ciudades inteligentes***

Las redes IoT se pueden usar para monitorear y controlar el alumbrado público, el flujo de tráfico, la gestión de desechos y otros servicios en áreas urbanas.

##### ***Gestión de energía***

Las redes IoT se pueden utilizar para monitorear y controlar el consumo de energía en hogares y negocios, lo que ayuda a reducir costos y mejorar la sostenibilidad.

Estos son solo algunos ejemplos de las muchas áreas de aplicación de las redes IoT. A medida que la tecnología continúa evolucionando y mejorando, podemos esperar ver casos de uso aún más innovadores en el futuro.

#### **Beneficios**

La aplicación de IoT es variada y funcional, ya que es versátil para cumplir su objetivo en función de los datos que genera el entorno donde se despliega. Se analiza la red en función de la funcionalidad prevista para cubrir la necesidad deseada (Universidad de Alcalá, 2019).

Por ejemplo, se ha demostrado que la aplicación industrial de IoT minimiza los esfuerzos y los riesgos al tiempo que aumenta la productividad, lo que hace que los procesos sean más efectivos y eficientes a través de la gestión y el control remotos de la maquinaria. En el estudio Industry 4.0 de Deloitte, se muestra que la tecnología IoT brinda información sobre el funcionamiento general del trabajo para analizarlo y realizar mejoras, incluido el monitoreo del estado de las máquinas, el control de calidad, el inventario y otros análisis.

El sector agrícola también se ha beneficiado de la tecnología IoT en forma de Smart Agriculture, que combina herramientas tecnológicas para digitalizar y mejorar la productividad agrícola. Los agricultores pueden obtener información actualizada sobre el estado de sus cultivos las 24 horas del día, los 7 días de la semana, y se pueden operar diferentes equipos agrícolas de forma remota para mejorar la cantidad y calidad de los cultivos. Otros sectores como gobierno, transporte, educación, entre otros, también pueden beneficiarse de las redes IoT.

#### **Modelos de Referencia**

En el Modelo general de la arquitectura de una red IoT ( Ver Figura 2) se puede evidenciar un modelo general estas redes, asimismo, se puede ver dividido en 3 grandes campos. Hay que resaltar que los nombres de cada parte del modelo pueden variar, pero su definición y función serán igual. En primer lugar, se puede ver la percepción, el cual contiene todos los sensores que recolectan la información para ser enviada. Gracias a la percepción esta información se envía a la red o nube, la cual, es la segunda capa, por medio de los gateways que son los encargados de transmitir, con toda la seguridad necesaria, los datos obtenidos entre las dos capas. Finalmente en la tercera capa se encuentran los actuadores o aplicación en donde se recibe toda la información recogida en la capa de percepción a través de la red y gracias a esto actuar dependiendo el requerimiento o el objetivo al que se quiera llegar; como se mencionó anteriormente toda la información es enviada a la red o nube, la cual almacena toda esta información para ser procesada y con su posterior análisis de los datos, estar lista para la aplicación en las “cosas” (Crespo Moreno, 2018).

**Figura 2** *Modelo general de la arquitectura de una red IoT*

Escala de tiempo

Descripción generada automáticamente

*Fuente: Elaboración propia.*

#### **Protocolos**

Los protocolos son una guía y/o pasos para saber cómo realizar una acción, de esta manera, los protocolos en IoT son los métodos en la que dos o más componentes se comunican por medio de la red, regulando las condiciones en que se transporta, el direccionamiento, enrutamiento y controles de los datos para que de esta manera se garantice la consistencia y transferencia de la información Machine2Machine (M2M).

##### **Tipos de Protocolos IoT**

Con el crecimiento, desarrollo e implementación de dispositivos IoT, se han establecido diferentes protocolos IoT para la gestión de comunicaciones. Para determinar el tipo de protocolo que se debe implementar en una red IoT, se deben considerar los dispositivos interconectados, su función u objetivo y la distancia que recorrerán para la transición de datos. Sin embargo, se utilizan comúnmente dos tipos principales de protocolos:

###### ***Protocolos de acceso de red***

Se utilizan en la capa inferior para permitir la conexión de dispositivos, que se puede realizar a través de Wi-Fi, Ethernet, 3G, 4G, 5G, etc

###### ***Protocolos de transmisión***

Utilizados en la transmisión de datos, que codifica la información enviada a través de las redes.

Hay que recalcar que la elección del protocolo depende de los requisitos específicos de la red a diseñar, los tipos de datos que se transmiten, la cantidad y tipos de dispositivos en la red y la distancia en la que los dispositivos deben transmitir los datos. Algunos de los protocolos comúnmente utilizados en las redes IoT incluyen.

###### ***MQTT (Transporte de telemetría de Message Queue Server)***

Este protocolo se usa ampliamente en aplicaciones IoT por su baja sobrecarga y distribución eficiente de mensajes. Está diseñado para aplicaciones con redes de bajo ancho de banda y alta latencia.

###### ***CoAP (Protocolo de aplicación restringida)***

Este protocolo está diseñado para dispositivos y redes con recursos limitados, y se utiliza para transferir datos entre dispositivos.

###### ***HTTP (Protocolo de transferencia de hipertexto)***

Aunque originalmente se diseñó para la web, HTTP se usa en aplicaciones de IoT por su amplio soporte, compatibilidad y facilidad de uso.

###### ***DDS (servicio de distribución de datos)***

DDS es un protocolo de mensajería de publicación/suscripción en tiempo real utilizado para sistemas centrados en datos como IoT.

###### ***AMQP (Protocolo de cola de mensajes avanzado)***

Este protocolo está diseñado para la entrega confiable de mensajes y admite múltiples patrones de mensajería.

###### ***ZigBee***

Este protocolo se utiliza en redes inalámbricas de baja potencia y está diseñado para aplicaciones con velocidades de datos bajas y un número limitado de dispositivos.

###### ***LoRaWAN (red de área amplia de largo alcance)***

Este protocolo se utiliza en aplicaciones IoT que requieren comunicación de largo alcance, como las aplicaciones de ciudades inteligentes.

#### **Dispositivos**

Como se definió anteriormente, IoT no puede funcionar sin los nodos IoT lo cuales son los diferentes dispositivos, en su mayoría físicos, los cuales garantizan el tratamiento de datos rápido, seguros, y eficientes.

##### ***Sensores***

Los sensores son los dispositivos que recogen la información del ambiente en donde se encuentra funcionando o responden a una salida de un sistema, detectando así los cambios que ocurren, de esta manera, por medio de los Gateway es enviada toda la información para luego ser procesada y analizada en la red.

##### ***Actuadores***

Los actuadores son los dispositivos que reciben o responden a la información producida por los sensores y analizada en la red, todo esto por medio de los gateways, para posteriormente por medio de la aplicación de control, funcionar dependiendo de la situación que en el que esté funcionando.

Por ejemplo, como se podrá observar en la Figura 3 Funcionamiento de actuadores IoT, un actuador muy común son los aires acondicionados que tiene la capacidad por medio de redes IoT recibir información de la temperatura por medio de un sensor que recoge estos datos, de esta manera, el actuador va a funcionar dependiendo la temperatura y los gustos de la persona que haga uso del aire acondicionado.

Figura 3 *Funcionamiento de actuadores IoT*

Diagrama

Descripción generada automáticamente

*Fuente: Elaboración propia.*

##### ***Gateway***

Los gateways son dispositivos intermediarios entre los sensores, actuadores y la red. Generalmente son físicos o software los cuales son los que reciben la información gracias a los sensores para posteriormente enviarla a la red (Crespo Moreno, 2018).

##### ***Teléfonos inteligentes y tablets***

Estos dispositivos se pueden usar para controlar y monitorear dispositivos IoT de forma remota.

##### ***Wearables***

Estos dispositivos, como relojes inteligentes y rastreadores de actividad física, recopilan y transmiten datos sobre la actividad física y la salud de un usuario.

##### ***Electrodomésticos inteligentes***

Estos dispositivos, como termostatos y refrigeradores inteligentes, se pueden controlar y monitorear de forma remota a través de un teléfono inteligente o una tableta.

##### ***Sistemas integrados***

Estos son sistemas informáticos especializados que están integrados en otros dispositivos, como automóviles o sistemas de seguridad para el hogar, para permitir la funcionalidad de IoT.

#### **Plataformas IoT**

Las plataformas IoT son el software en el cual los dispositivos IoT se pueden comunicar o conectar y crear información de valor por medio del entorno digital de la misma plataforma, gracias a esto, el desarrollo y el funcionamiento de la red se produce de la mejor manera. Teniendo en cuenta lo anterior, hoy en día en cuanto a plataformas para el monitoreo de estas conexiones existe una gran variedad. De cierto modo, IoT se está convirtiendo en un pilar para la sociedad y por esta razón han desarrollado plataformas nuevas y cada día mejores para el buen uso, desarrollo, gestión y mantenimiento de estas (Quiñones Cuenca, González Jaramillo, Torres, & Miguel , 2017).

### **Analítica de Datos**

La analítica de datos es el uso de la información que se puede obtener de manera digital, con el propósito de extraer la mejor información para poder tomar las mejores decisiones (Gibbs, 2012); varios autores vinculan la analítica con el manejo de variables con el uso de algoritmos.

La analítica de datos puede clasificarse en tres grandes categorías: analítica descriptiva, analítica predictiva y analítica prescriptiva (Pusala, Amini, Katukuri, 2016).

***La analítica de datos descriptiva***

Como estado inicial en el que los diferentes tomadores de decisiones profundizan en los respectivos datos históricos con el fin de detectar patrones de comportamiento en las variables para realizar análisis de correlación (ibertech, 2006).

***La analítica de datos predictiva***

Donde las empresas con datos registrados anteriormente generan modelos de pronósticos sobre las tendencias y así poder realizar cambios para mejorar (ibertech, 2006).

***La analítica de datos prescriptiva***

Donde las compañías utilizan modelos de simulación de escenarios, para la optimización de diferentes fuentes de interés (ibertech, 2006).

#### **Áreas de Aplicación de Analítica.**

Las áreas donde se puede aplicar la analítica son extensas, ya que muchas actividades o procedimientos que realizan es necesario hacer una investigación anteriormente, para poder obtener unos resultados apropiados con el fin de analizar y así poder utilizarlos en un propósito de sacar conclusiones sobre la información tratada; todo esto para tomar las mejores decisiones, verificar teorías y modelos existentes, la clasificación de conjuntos de datos para obtener una relación y utilizarlos en el mejoramiento de campañas (Joyanes Aguilar, 29 de mayo del 2019); muchas de las áreas pueden ser de economía, probabilidad, administración, web, inteligencia artificial etc (Gomez-Aguilar, Garcia-Peñalvo, & Theron, 2014).

#### **Cadena de Valor de Datos**

Como su propio nombre indica, son una gran cantidad de datos que representan cierta información (Quintero, 2006); en el cual ciertas empresas ya están destinadas a prestar estos servicios; la cadena tiene varias etapas para su realización las cuales son:

1. Priorizar: Determina por su posición en la lista, el primer conjunto de datos tiene la prioridad más alta.
2. Recolectar: Proceso de recopilación y medición de información sobre variables establecidas de una manera sistemática.
3. Integrar: Combinación de procesos técnicos y de negocio que se utilizan para combinar información de diferentes fuentes.
4. Procesar y analizar: Manipulación de elementos de datos para producir información significativa.
5. Visualizar: Presentación de datos en formato ilustrado o gráfico, el cual permite a los tomadores de decisiones poder optar la mejor opción.
6. Impactar: es el uso que se le pueden dar a los datos frente al impacto que pude tener frente a la empresa u organización.

La cadena de valor proporciona cierto modelo de aplicación el cual permite representar todas las actividades de cualquiera empresa y también proporciona un procedimiento para el desarrollo de ventajas como la debilidad o fortaleza de los datos a tratar, para poder extraer la mayor cantidad posible de información, la identificación de datos relevantes y el planteamiento de estrategias o contingencias para el manejo de los respectivos datos pulidos (ver Figura 4) asimismo en la Tabla 1. Se muestran las etapas para realizar la cadena de valor (Hergert & Morris, 1989)

Figura 4 *Cadena de valor*

Diagrama

Descripción generada automáticamente

***Fuente*** *(Muñoz, 2019)****.***

Tabla 1 *Etapas para realizar cadena de valor*

|  |  |  |  |
| --- | --- | --- | --- |
| Recopilación | Publicación | Consumo | Impacto |
| Identificar | Analizar | Conectar | Usar |
| Manipular | Liberación | Incentivar | Cambio |
| Proceso | Diseminar | Influencia | Reutilizar |

*Fuente: Elaboración propia.*

#### **Datos**

Representación simbólica de alguna información o procedimiento, en la cual, puede ser almacenado y analizado para poder realizar ciertas operaciones y así poder generar información adecuada para la toma de decisiones frente al mejoramiento de procesos en los diferentes campos para el obtener mejor desempeño de empresas o establecimientos.

##### ***Tipos de datos***

Los tipos de datos son una forma de clasificar diferentes valores que se pueden almacenar y manipular en un programa de computadora. La mayoría de los lenguajes de programación proporcionan varios tipos de datos predefinidos, como números enteros, decimales, caracteres y valores booleanos. Cada tipo de datos tiene un conjunto específico de valores que puede aceptar y representar de una manera específica en la memoria de la computadora. Elegir el tipo de datos correcto es muy importante para garantizar que las operaciones realizadas en el programa sean precisas y eficientes.

###### **Estructurados**

Son la mayoría de los datos que se pueden encontrar almacenados en una base de datos; se muestran en fila y columnas, tienen definido su longitud en el formato que se encuentran los respectivos datos (Hernandez & Rodriguez, 2008).

###### **No Estructurados**

Son datos binarios que no están organizados que no tienen algún valor al utilizarlos hasta que son organizados y almacenados, el cual su manejo es mucho más complejo que en los demás, estos datos no se pueden usar en una base tradicional como son las tablas ya que es imposible poder organizarlos o ajustarlos en filas y columnas estandarizadas, pero se encuentran muchos tipos de datos no estructurados de uso común como archivos PDF, imágenes o archivos de texto (Hassan, Domingo-Ferrer, & Soria-comas, 2018).

###### **Semi – estructurados**

Son datos que no son organizados en un repositorio, pero tiene información importante como metadatos (datos que están cerca de los datos) la cual hace que se pueda procesar más fácilmente los datos (Raposo, 2007).

#### **Tipos de analítica**

Con el aumento de la cantidad de datos que generan actualmente las organizaciones, los negocios pueden extraer grandes cantidades de información las cuales se pueden utilizar para mejorar a las empresas como predecir qué mes se puede obtener más ganancias.

“*un conjunto de métodos de análisis matemático y estadístico que sirve para identificar patrones de comportamiento, pronósticos, escenarios “que pasaría si”, entre otros*” (Davenport y Harrys, 2017).

##### ***Descriptiva***

Se pueden diferenciar de los otros tipos por medio de la pregunta inicial sea ¿qué sucedió? o ¿Qué está pasando? un resumen del desempeño del total de las actividades empresariales, el cual permite ver la composición principal dentro de un negocio como por ejemplo observar las ganancias o pérdidas en el mes (cuatro tipos de analítica de retail que todo comercio necesita en 2018, 2018). Consiste en almacenar y realizar información con datos históricos para obtener un análisis del estado actual y anterior del negocio, permite detectar, visualizar, observar e identificar el camino a tomar para la mejor desciño posible.

##### ***Diagnostica***

Se pueden diferenciar de los otros tipos por medio de la pregunta inicial sea ¿Por qué está pasando? Tiene en cuenta los antecedentes de lo que se quiere analizar para dar un informe más acertado con sus respectivas herramientas y así poder eliminar el problema y así tener los elementos necesarios para que el respectivo análisis en los datos se pueda obtener la causa de los problemas (cuatro tipos de analítica de retail que todo comercio necesita en 2018, 2018).

##### ***Predictiva***

Se pueden diferenciar de los otros tipos por medio de la pregunta inicial sea ¿qué podría pasar? o ¿Qué es lo más probable que pueda pasar? tienen como objetivo identificar la probabilidad que ocurra algo en el futuro y que no perjudique el análisis realizado; estos modelos se suelen utilizar los datos o variables los cuales se puedan realizar predicción y así tomar mejores decisiones por esto es uno de los más importantes (cuatro tipos de analítica de retail que todo comercio necesita en 2018, 2018).

##### ***Prescriptiva***

Se pueden diferenciar de los otros tipos por medio de la pregunta inicial sea ¿qué deberíamos hacer? o ¿Qué necesito hacer?, entendimiento de lo que ha sucedido, por qué ha sucedido y un procedimiento en el cual podría suceder con el paso del tiempo, ayudar al usuario a determinar el mejor curso de acción a tomar (cuatro tipos de analítica de retail que todo comercio necesita en 2018, 2018).

#### **Técnicas de Analítica**

Muchas veces la base o la raíz que tiene la analítica es entender lo que ha pasado o lo que está pasando en el momento; todo esto para poder adquirir el conocimiento para mejorar las decisiones hacia el futuro, muchas veces se utilizan técnicas de estadística y matemáticas para poder lograr el objetivo (Garcia, 2006).

Análisis exploratorio de datos (EDA): Este es un método para analizar y resumir datos utilizando técnicas gráficas y estadísticas, ayuda a descubrir patrones, relaciones y tendencias en los datos e identificar valores atípicos o datos faltantes (EOI, 2007).

##### Análisis de regresión: Esta técnica se utiliza para determinar la relación entre una variable dependiente y una o más variables independientes, ayuda a determinar la fuerza y ​​la dirección de las relaciones entre las variables (EOI, 2007).

##### Análisis de series temporales: Este método se utiliza para analizar datos que cambian con el tiempo y predecir tendencias futuras, ayuda a identificar tendencias y tendencias a lo largo del tiempo y predecir la demanda futura (EOI, 2007).

##### Análisis de clustering: Este método se utiliza para organizar datos similares en grupos o grupos. Ayuda a identificar tendencias y piezas de datos que se pueden utilizar para tomar decisiones comerciales (EOI, 2007).

##### Análisis de texto: Este método se utiliza para analizar y extraer información de grandes cantidades de datos no estructurados, como correos electrónicos, tweets y comentarios en redes sociales, ayudar a identificar temas, ideas y puntos en común (EOI, 2007).

### **Simulación**

La simulación es una representación exacta o casi exacta del comportamiento de un evento o fenómeno, con el fin de obtener el mismo resultado, características, información entre otros, aspectos, consumiendo menos recursos de los que consumiría ejecutar el modelo real, sin necesidad de realizar dicho evento para obtener un análisis o estudio del resultado con una menor inversión (Castellanos Hernández & Chacon Osorio, 2006).

#### **Tipos de Simulación**

La simulación puede ser dividida y clasificada según la naturaleza del sistema que será representado en el modelo de simulación, por ejemplo:

##### **Simulación de situaciones**

Permite simular una situación física o real y observar su comportamiento, con el fin de extraer información de hechos existentes (Castellanos Hernández & Chacon Osorio, 2006).

(EOI, 2007)

##### **Simulación de Realizar Alguna Situación**

Son aquellos que permiten experimentar una situación como si el usuario u sujeto estuviera en ella, un simulador de vuelo es un ejemplo de esto, permite al usuario pilotear un avión sin estar en uno realmente (Castellanos Hernández & Chacon Osorio, 2006).

#### **Fases de Estudio de Simulación**

Para hacer un modelo de simulación efectivo, es necesario tener en cuenta distintas fases que permitirán la construcción de un modelo eficiente.

##### ***Definición de objetivos***

Se deben establecer los objetivos que se pretenden conseguir con la simulación, los efectos que causara y las respuestas a responder con este estudio (García Sánchez & Ortega Mier, 2006).

##### ***Definición del sistema***

Definir los elementos que harán parte del sistema teniendo en cuenta el sistema a emular (García Sánchez & Ortega Mier, 2006).

##### ***Elaboración del modelo conceptual***

A partir de los objetivos planteados anteriormente se crea un modelo conceptual, el cual debe ser sencillo (solo enfocarse en lo necesario para simular) y específicamente diseñado para cumplir dichos objetivos. El modelo conceptual debe representar sencillez y a su vez representar el realismo del sistema a emular (García Sánchez & Ortega Mier, 2006).

Este modelo conceptual debe ser evaluado y comprobar que refleje fielmente el sistema que se desea emular teniendo en cuenta los objetivos que debe cumplir (Coss Bu, 2003).

##### ***Elaboración del sistema comunicativo***

Los diseñadores del modelo conceptual son distintos muchas veces a los programadores del simulador. Para su comunicación entre si debe ser eficaz, por esta razón los diagramas de flujo son una opción útil para representar los eventos en el simulador como lo son los datos, el proceso, una decisión un avance en la simulación etc (García Sánchez & Ortega Mier, 2006).

##### ***Construcción y verificación de modelo informático***

Una vez verificado el modelo conceptual se escoge un lenguaje apto para para la programación del simulador, este lenguaje debe permitir la correcta emulación como fue planeada, se debe escoger el más conveniente (García Sánchez & Ortega Mier, 2006).

##### ***Validación final***

Una vez construido el modelo de simulación creado anteriormente, es necesario hacer pruebas para verificar su correcto funcionamiento, en las cuales los resultados deberán ser similares a los esperados y si es posible se comparará con los resultados del sistema real al cual se está simulando (Coss Bu, 2003).

#### **Modelos de simulación**

Hay diversos modelos de simulación los cuales serán mencionados a continuación, haciendo mayor énfasis en el modelo Estocástico:

##### **Estático**

La simulación no depende del tiempo, por ejemplo, un sistema que se encuentra en un estado de equilibrio o reposo (García Sánchez & Ortega Mier, 2006).

##### **Dinámico**

En contraparte a la simulación estática, el modelo dinámico depende del tiempo, sus procesos pueden variar respecto al tiempo que va transcurriendo, por ejemplo, un modelo que represente el incremento poblacional al cabo de n años (García Sánchez & Ortega Mier, 2006).

##### **Determinístico**

Su ejecución será siempre igual, el valor de su resultado será el un resultado ya esperado debido a las condiciones iniciales, por lo tanto, no tiene ninguna variable o proceso al azar, por ejemplo, el uso de un dispositivo programado para realizar una tarea determinada al momento de que el usuario oprima un botón (García Sánchez & Ortega Mier, 2006).

##### **Estocástico**

A diferencia del modelo determinístico, los procesos estocásticos contienen variables o procesos al azar que cual puede variar el resultado de la simulación, de esta manera funciona todo en la vida real, por ejemplo, un modelo que permita predecir el tiempo que puede tardar en realizarse una transacción bancaria, en la mayoría de veces es posible calcular un tiempo promedio pero su vez hay factores que pueden afectar la velocidad en la que se realiza la transacción, uno de estos factores es el tráfico en la red afectado por la cantidad de usuarios que está realizando una transferencia al mismo tiempo (García Sánchez & Ortega Mier, 2006).

##### **Discreto**

Varía dependiendo de sucesos que ocurran en la simulación del modelo y en un tiempo determinado, por ejemplo, en una sala de urgencias los pacientes son remitidos a diferentes especialistas y tratamientos según la incidencia que hayan sufrido, por lo que el tiempo y otros factores son alterados con respecto a el tipo de atención que dicho paciente requiera (García Sánchez & Ortega Mier, 2006).

##### **Continuo**

Tiene un rango de tiempo el cual es previamente establecido. Sin importar los sucesos que ocurran seguirá ejecutándose y sus variables estarán cambiando en cualquier momento, por ejemplo, al comprar productos en internet, la compra será realizada según la disponibilidad del producto, ya que este puede ser vendido por completo a otro cliente en cualquier momento.

##### **Físicos**

Se basan en eventos físicos o fenómenos que ocurren y no son posibles de controlar y estudiar, por ejemplo, el cambio climático en una región de Colombia (García Sánchez & Ortega Mier, 2006).

#### **Procesos estocásticos**

Los procesos estocásticos es una colección de variables aleatorias infinitas que se basan en el cambio o evolución de una variable con respecto al tiempo o en función de otra variable como lo puede ser la temperatura, el cambio climático entre otras (Myriam Muñoz de ózak, 1991).

Las variables aleatorias que dependen del tiempo son aquellos fenómenos que evoluciona al azar a lo largo del tiempo, el tiempo tomará diferentes valores en dicho conjunto donde la colección de variables aleatorias se verá afectadas por esto tomando diferentes valores según T (Myriam Muñoz de ózak, 1991).

El conjunto de variables que dependen de otra variable son aquellos que su evolución no dependen del tiempo sino de otro fenómeno (Myriam Muñoz de ózak, 1991), estas variables pueden tomar valores directamente proporcionales a este fenómeno no perteneciente al conjunto dicho anteriormente es decir que los valores de las variables aumentaran si el valor del fenómeno aumenta y disminuirá si este disminuye, por otro lado las variables aleatorias que son indirectamente proporcionales al fenómeno harán todo lo contrario si el valor del fenómeno aumenta, el valor de las variables disminuirá y si este disminuye las variables aumentaran.

#### **Variables aleatorias**

Las variables aleatorias son parte fundamental de una simulación, ya que los sistemas requieren diferentes tipos de datos no siempre serán los mismos para ejecutar un evento simulado, por esta razón es de vital importancia crear variables aleatorias ya que necesitamos que la simulación sea lo más apegado posible a la realidad (Eduardo García Dunna, 2013).

El modelo que se quiere construir debe estar compuesto de variables aleatorias que interactúen entre sí, para asemejarlo a la realidad (Eduardo García Dunna, 2013). Una variable aleatoria es una representación de un suceso o un numero de una parte del evento que se está intentando emular, es decir que un suceso en el sistema puede variar en ese mismo proceso u otro proceso independiente, por ejemplo, si se quiere emular un sensor de temperatura, una de las variables aleatorias será la temperatura ya que esta puede cambiar con el paso del tiempo o ser diferente en otra prueba del simulador (Eduardo García Dunna, 2013).

Existen dos tipos de variables aleatorias, la primera es la variable aleatoria discreta, la cual se caracteriza por ser de solo números enteros es decir no puede tomar valores como 10,97 u otro tipo de número que no es un entero (Eduardo García Dunna, 2013), por otra parte la variable aleatoria continua es más caracterizada por su uso para las mediciones en este caso si puede contener valores decimales, retomando el ejemplo del medidor de temperatura podemos decir que esta simulación consta de variable aleatoria continua ya que la temperatura es una medida y puede tener parte decimal (Eduardo García Dunna, 2013).

#### **Modelos probabilísticos**

Cuando se habla de un modelo probabilístico se hace referencia a un conjunto de datos obtenidos por diversas repeticiones de un evento aleatorio usados para poder predecir el comportamiento de este evento con los mismos o diferentes datos para las futuras repeticiones de dicho evento (Leónardo Darío Bello Parias, 2000), esta serie de repeticiones permiten asemejar el modelo que se está construyendo con datos aleatorios a un conjunto de datos de una población mayor, con esto se hace referencia a la simulación más acercada posible de un evento real mediante la prueba y repetición del modelo que se está simulando.

Existen varios modelos probabilísticos para variables aleatorias:

* Distribución Uniforme
* Distribución Gamma
* Distribución Exponencial
* Distribución Ji-dos
* Distribución Normal
* Distribución t Student
* Distribución F de Sendecos
* Distribución normal bivariante

Los modelos probabilísticos son basados en hipótesis y se compone por ecuaciones las cuales relacionan las diversas variables aleatorias (Carlos Gamero Burón, 2015), estos modelos son la representación más viable de una hipótesis para un evento que este compuesto de variables aleatorias por lo cual debe ser rectificado correctamente y probado una y otra vez.

#### **Números pseudoaleatorios**

Una simulación, muchas veces se compone de variables aleatorias es decir números al azar, para conseguir esto los números pseudoaleatorios son parte fundamental en este proceso de simulación, su nombre está compuesto de dos palabras, “Pseudo” lo cual significa falso y “aleatorio”, se le denomina falso debido a que es imposible generar números completamente aleatorios, al no ser posible generar números completamente aleatorios los números pseudoaleatorios son creados a partir de algoritmia determinística con parámetros de arranque, esto permitirá generar números que se comportaran similarmente a números totalmente aleatorios es decir números sin correlación entre ellos mismos permitiéndonos simular el comportamiento aleatorio de las variables en el evento que queremos simular (Eduardo García Dunna, 2013).

##### ***Generación de números de pseudoaleatorios***

Para hacer la generación de los números Pseudoaleatorios se debe tomar un espacio o rango lo suficientemente grande para ello, es decir cuente con demasiados números en secuencia para una vida útil prolongada (Eduardo García Dunna, 2013). Es necesario este conjunto tan grande porque al hacer una simulación pequeña se necesitarán un conjunto de números mínimo, pero si se quiere hacer aun mayor este número incrementara, pero al hacer la simulación no puede basarse en solo un resultado para ello es necesaria la simulación una y otra vez con números distintos es por esto por lo que es necesario dicho conjunto los suficientemente grande para satisfacer esta necesidad (Eduardo García Dunna, 2013).

Para aprobar el uso de estos números el conjunto de números pseudoaleatorios se debe someter a ciertas pruebas que nos permitan comprobar la independencia entre ellos y que estos sean uniformes, para ellos se mencionaran unas pruebas estadísticas para la aprobación de este conjunto se debe asegurar que los números de un conjunto deben ser uniformemente distribuidos lo cual significa que en los subintervalos haya la misma cantidad de números del conjunto, deben ser continuos, la media del conjunto debe ser equivalente a ½ y la varianza también debe ser ½ (Eduardo García Dunna, 2013).

#### **Ventajas de la simulación**

La simulación permite ahorrar recursos para obtener los posibles resultados del comportamiento de un evento (García Sánchez & Ortega Mier, 2006).

A partir de la simulación es posible trabajar mejor los experimentos debido a su mejor manejo en las condiciones de dicho experimento (García Sánchez & Ortega Mier, 2006).

Es posible a partir de la simulación comparar y escoger el sistema más viable dependiendo de una necesidad (García Sánchez & Ortega Mier, 2006).

La simulación puede permitir una mejor comprensión del evento que está simulando (García Sánchez & Ortega Mier, 2006).

Con una simulación es posible hacer diferentes experimentos y su reacción a estos, los cuales no son posibles con el modelo físico el cual se pretende obtener esta nueva información (García Sánchez & Ortega Mier, 2006).

#### **Desventajas de la simulación**

Una vez creada la simulación es posible ahorrar tiempo en la obtención de los datos del modelo simulado, pero para crear la simulación lleva tiempo y estudios los cuales no son mayores los recursos que requerirá usar el modelo real (García Sánchez & Ortega Mier, 2006).

La simulación debe ser exacta al modelo real pero aun así se puede generar datos no correctos o no exactos algunas veces a los reales (García Sánchez & Ortega Mier, 2006).

### **Machine Learning**

La creación de las computadoras tiene como objetivo realizar las tareas que las personas no son capaces de hacer o no pueden ejercer de una manera rápida y eficiente, a pesar de que este objetivo ha sido cumplido por las maquinas, nunca han podido tener en ellas una inteligencia al nivel o superior de la mente humana que permita a los dispositivos pensar por sí mismos y tomar decisiones en los diferentes procesos y tareas que ejecutan. Sin embargo, el hombre ha logrado hacer que las maquinas imiten la inteligencia humana, a esto se le conoce como inteligencia artificial (IA) (Oracle, 2022), una característica fundamental del razonamiento humano es el aprendizaje de hechos o experiencias. Esto permite al humano tomar decisiones lógicas sobre las acciones que realiza, permitiendo así obtener mejores resultados de las mismas.

Para lograr que un dispositivo de cómputo sea capaz de aprender, son necesarias grandes cantidades de datos ya que estos son equivalentes al conocimiento necesario para que a partir de algoritmos que tengan la capacidad tomar decisiones y realizar acciones que permitan cambiar y mejorar según el modelo de datos utilizado(Carlemany, 2022), es decir que los algoritmos integrados en un sistema de aprendizaje automático deben tener la capacidad de “alimentarse” de datos con el fin de ser entrenados y de esta manera lograr aprender de ellos encontrando patrones que permitan predecir un comportamiento normal o común entre dichos datos y así tomar decisiones coherentes o acertadas e incluso predecir eventos futuros de manera eficiente. Para lograr el entrenamiento de los algoritmos de Machine Learning, se pueden utilizar tres métodos de aprendizaje: supervisado, no supervisado y por esfuerzo.

#### **Supervisados**

El aprendizaje supervisado se lleva a cabo mediante conjuntos de datos que contienen tanto parámetros característicos que toman el rol de preguntas, como parámetros de etiquetas que toman el rol de respuestas (Sandoval, 2018). De esta manera, el modelo aprende de todas las características que componen al conjunto de datos para interpretar y predecir las respuestas o características faltantes de manera automática y precisa en el futuro. El aprendizaje supervisado se divide en dos tipos de algoritmos: el algoritmo de clasificación y el algoritmo de regresión.

Este tipo de aprendizaje está dividido en dos tipos de algoritmos que son:

##### ***Algoritmo de Clasificación***

El algoritmo de clasificación se utiliza para clasificar conjuntos de datos gracias a etiquetas de clase categórica. En otras palabras, este algoritmo determina si el dato a predecir pertenece o no a un grupo específico de datos en función de las características que lo componen (Sandoval, 2018)

##### ***Algoritmo de Regresión***

El algoritmo de regresión, por otro lado, se utiliza para predecir una característica de un elemento, generalmente un número, en función de las demás características que lo componen (Sandoval, 2018).

#### **No Supervisados**

Por otro lado, el aprendizaje no supervisado no se basa en la experiencia proporcionada por datos etiquetados, sino que se lleva a cabo a partir de elementos no etiquetados. En este tipo de aprendizaje, se intenta agrupar los datos no estructurados en función de las características que los componen (Dra. Marianella Álvarez Vega, 2020). El aprendizaje no supervisado se divide en dos métodos: agrupamiento y reducción de la dimensionalidad.

##### ***Agrupamiento***

El agrupamiento se utiliza para interpretar los datos y agruparlos por similitudes en función de las características en las que están compuestos. Esto permite estructurar y organizar los datos que no tienen una etiqueta que los defina en un grupo específico, como ocurre en los algoritmos de clasificación en el aprendizaje supervisado (González, 2015).

##### ***Reducción de la dimensionalidad***

La reducción de la dimensionalidad se utiliza para seleccionar las características más relevantes o representativas para la clasificación objetivo de los diferentes conjuntos de datos. Al mismo tiempo, se descartan las características menos importantes, lo que permite una mayor eficiencia en el modelo de aprendizaje (Softtek, 2021).

#### **Por esfuerzo**

El aprendizaje por esfuerzo (también conocido como aprendizaje activo o aprendizaje semisupervisado) es una técnica de aprendizaje automático en la que el modelo se entrena a partir de un conjunto inicial de datos etiquetados y luego solicita al usuario que proporcione etiquetas adicionales para una selección estratégica de los datos no etiquetados para mejorar el modelo. En lugar de requerir grandes conjuntos de datos etiquetados, el aprendizaje por esfuerzo permite entrenar modelos precisos con una cantidad relativamente pequeña de datos.

## Metodología

Esta sección del documento describe los pasos involucrados en el desarrollo del modelo, detallando las actividades desde la documentación y el desarrollo hasta las pruebas. La metodología utilizada es el enfoque CRISP-DM (Cross Industry Standard Process for Data Mining) para la minería de datos.

Las técnicas de ciencia de datos y análisis de datos se originaron en la década de 1990 con el concepto de Knowledge Discovery in Databases (KDD), que implicaba extraer valor de los datos. Hacia finales de los 90, en un esfuerzo por estandarizar los proyectos de ciencia de datos, se desarrollaron dos metodologías: CRISP-DM (Proceso Estándar de la Industria Cruzada para la Minería de Datos) y SEMMA (Muestree, explore, modifique, modele y evalúe).

CRISP-DM (Cross-Industry Standard Process for Data Mining) es una metodología ampliamente utilizada en proyectos de ciencia de datos y minería de datos que proporciona un enfoque estructurado para guiar el desarrollo de un modelo predictivo. La metodología consta de seis fases

### **Comprensión del negocio**

Comprender los objetivos y requisitos del proyecto, identificar el problema a resolver y definir los objetivos de la minería de datos.

### **Comprensión de datos**

Recopilación y exploración de datos, identificación de problemas de calidad de datos y verificación de la idoneidad de los datos para el modelado.

### **Preparación de datos**

Limpieza, transformación y selección de los datos más apropiados para ser utilizados en el modelado.

### **Modelado**

Seleccionar las técnicas de modelado a utilizar, construir y evaluar el modelo, y seleccionar el mejor.

### **Evaluación**

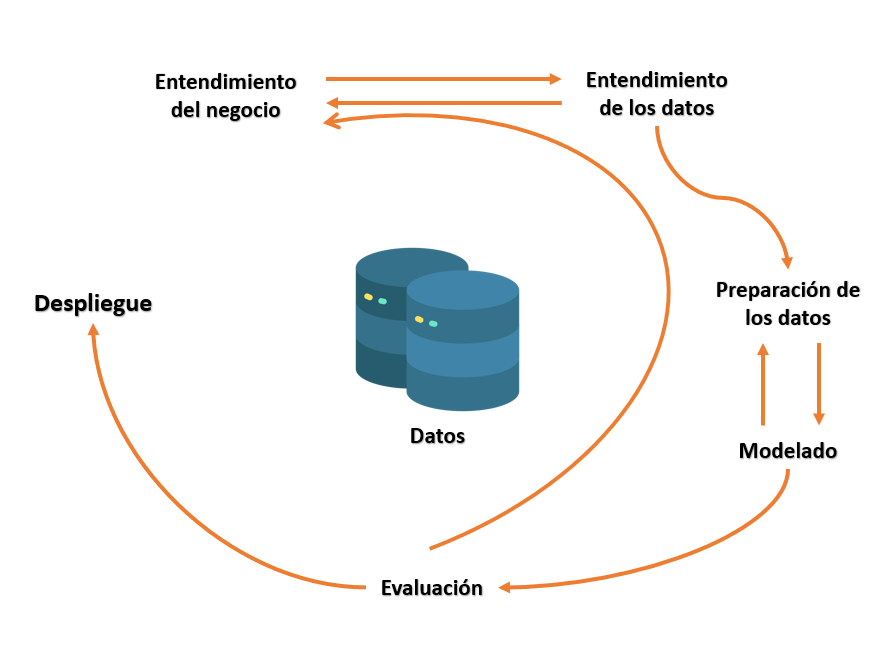
Evaluar la efectividad del modelo, probar su generalización y asegurarse de que cumpla con los objetivos comerciales.

### **Despliegue.**

Integración del modelo en el proceso de negocio y seguimiento de su rendimiento.

Para el presente proyecto de simulación de datos ambientales, la metodología CRISP-DM será utilizado para guiar el desarrollo del modelo de simulación. En la fase de Business Understanding, se identificarían los objetivos y requisitos del proyecto, como simular el comportamiento de un ecosistema en particular bajo diferentes escenarios de cambio climático. En la fase de comprensión de datos, se recopilarían y explorarían datos relacionados con el ecosistema y el clima. En la fase de preparación de datos, los datos se limpiarían, transformarían y seleccionarían para crear el modelo de simulación. En la fase de Modelado se construiría y evaluaría el modelo de simulación, seleccionando el más adecuado. En la fase de Evaluación, se evaluaría y probaría la efectividad del modelo de simulación para garantizar que cumpla con los objetivos del proyecto. Finalmente, en la fase de Despliegue, el modelo sería integrado al proyecto y monitoreado para asegurar su desempeño.

**Figura 5** *Estructura de la metodología CRISP-DM*



*Fuente: Elaboración propia*

**Compresion del negocio**

En esta fase se entenderá el propósito del modelo, contexto del ámbito ambiental y climatológico, terminología, objetivos y búsqueda de la información.

Para esta fase se tuvo previsto una X semanas de investigación, entendimiento de los diferentes temas y búsqueda de los datos.

***Tareas***

**Actividad 1.**

Análisis de objetivos.

**Actividad 2.**

Diseño de justificación y pregunta problema.

**Actividad 3.**

Generación de alcances y limitaciones.

**Actividad 4.**

Marco conceptual.

**Actividad 5.**

Búsqueda de antecedentes.

**Actividad 6.**

Socialización y aprobación del documento.

**Actividad 7.**

Correcciones del documento.

**Actividad 8.**

Base de conocimiento.

**Entendimiento de los datos**

En esta fase se entenderán los datos obtenidos después de haberlos obtenido en las fuentes correspondientes, en el caso del proyecto y la fuente más confiable es Datos Abiertos del modelo, contexto del ámbito ambiental y climatológico, terminología, objetivos y búsqueda de la información.

Para esta fase se tuvo previsto unas 4 semanas de investigación, entendimiento de los diferentes temas y búsqueda de los datos.

***Tareas***

**Actividad 9.**

Web Scrapping de los datos.

**Actividad 10.**

Analítica descriptiva de los datos.

**Preparación de los datos**

Además de comprender los datos, la preparación de datos es uno de los procesos más críticos en la preparación de modelos. Por lo tanto, se asignó suficiente tiempo para preparar y estandarizar los datos para pasar a la siguiente fase. Esta fase tardó 5 semanas en completarse.

***Tareas***

**Actividad 11.**

Estandarización de datos.

**Actividad 12.**

Obtención de datos extras con respecto a la fecha (año, mes, día y hora).

**Actividad 13.**

Carga de datos al sistema de gestión de bases de datos relacional llamado MySQL.

**Actividad 14.**

Obtención de promedios horarios por variable.

**Actividad 15.**

Unión de variables para la construcción del conjunto de datos final.

**Actividad 16.**

Verificación y carga de datos a Google Collab.

**Modelado**

Vale la pena señalar que los pasos antes mencionados debían llevarse a cabo con mucho cuidado, atención a los detalles y una comprensión profunda de cada etapa para garantizar un modelado preciso basado en datos de entrada de alta calidad. Por lo tanto, el equipo del proyecto dedicó 5 semanas a la recopilación de información, comprensión y selección del modelo y lenguaje de programación más adecuados que se alinean con los objetivos del proyecto.

***Tareas***

**Actividad 17.**

Entendimiento del conjunto de datos por medio del lenguaje de programación Python.

**Actividad 18.**

Realización de analítica descriptiva a los datos por medio del lenguaje de programación Python.

**Actividad 19.**

Realización de los diferentes modelos de Machine Learning.

**Evaluación**

Si los pasos anteriores se llevaron a cabo correctamente, la evaluación debería arrojar resultados satisfactorios. Dado que esta es una tarea meticulosa, se necesitaron 3 semanas para realizar las pruebas funcionales y las pruebas en tiempo real.

***Tareas***

**Actividad 20:** Pruebas funcionales al código.

**Actividad 21:** Comparación de datos reales con los datos obtenidos mediante el modelo realizado.

**Despliegue**

Finalmente, como se menciona en el alcance, la información obtenida del modelo se cargará en una base de datos para su consumo y manipulación. Por lo tanto, fueron necesarias 2 semanas para la construcción de la estructura de la base de datos y la carga de datos.

***Tareas***

**Actividad 22.** Desarrollo de la estructura de la base de datos, también conocido como DDL, la cual, contendrá los datos previamente obtenidos con el modelo.

Teniendo en cuenta lo anteriormente mencionado, se definieron las siguientes tareas y cronograma realizadas durante todo el proceso.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Ago. | | | | | Sept | | | | Oct | | | | Nov | | |
| Actividades | S. 1 | S. 2 | S. 3 | S. 4 | S. 5 | S. 6 | S. 7 | S. 8 | S. 9 | S. 10 | S. 11 | S. 12 | S. 13 | S. 14 | S. 15 | S. 16 |
| Act. #1 | X | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Act. #2 | X | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Act. #3 |  | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Act. #4 |  | X | X | X |  |  |  |  |  |  |  |  |  |  |  |  |
| Act. #5 |  |  | X | X |  |  |  |  |  |  |  |  |  |  |  |  |
| Act. #6 |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |  |
| Act. #7 |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |  |
| Act. #8 |  |  |  | X | X |  |  |  |  |  |  |  |  |  |  |  |
| Act. #9 |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |
| Act. #10 |  |  |  |  | X | X |  |  |  |  |  |  |  |  |  |  |
| Act. #11 |  |  |  |  | X | X |  |  |  |  |  |  |  |  |  |  |
| Act. #12 |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |
| Act. #13 |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |
| Act. #14 |  |  |  |  |  |  | X | X |  |  |  |  |  |  |  |  |
| Act. #15 |  |  |  |  |  |  |  | X | X |  |  |  |  |  |  |  |
| Act. #16 |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |
| Act. #17 |  |  |  |  |  |  |  |  | X | X |  |  |  |  |  |  |
| Act. #18 |  |  |  |  |  |  |  |  |  | X | X |  |  |  |  |  |
| Act. #19 |  |  |  |  |  |  |  |  |  |  | X | X | X |  |  |  |
| Act. #20 |  |  |  |  |  |  |  |  |  |  |  |  | X |  |  |  |
| Act. #21 |  |  |  |  |  |  |  |  |  |  |  |  | X | X | X |  |
| Act. #22 |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X | X |
|  | Ago. | | | | | Sept | | | | Oct | | | | Nov | | |
| Actividades | S. 1 | S. 2 | S. 3 | S. 4 | S. 5 | S. 6 | S. 7 | S. 8 | S. 9 | S. 10 | S. 11 | S. 12 | S. 13 | S. 14 | S. 15 | S. 16 |
| Act. #1 | X | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Act. #2 | X | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Act. #3 |  | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Act. #4 |  | X | X | X |  |  |  |  |  |  |  |  |  |  |  |  |
| Act. #5 |  |  | X | X |  |  |  |  |  |  |  |  |  |  |  |  |
| Act. #6 |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |  |
| Act. #7 |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |  |
| Act. #8 |  |  |  | X | X |  |  |  |  |  |  |  |  |  |  |  |
| Act. #9 |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |  |
| Act. #10 |  |  |  |  | X | X |  |  |  |  |  |  |  |  |  |  |
| Act. #11 |  |  |  |  | X | X |  |  |  |  |  |  |  |  |  |  |
| Act. #12 |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |  |
| Act. #13 |  |  |  |  |  |  | X |  |  |  |  |  |  |  |  |  |
| Act. #14 |  |  |  |  |  |  | X | X |  |  |  |  |  |  |  |  |
| Act. #15 |  |  |  |  |  |  |  | X | X |  |  |  |  |  |  |  |
| Act. #16 |  |  |  |  |  |  |  |  | X |  |  |  |  |  |  |  |
| Act. #17 |  |  |  |  |  |  |  |  | X | X |  |  |  |  |  |  |
| Act. #18 |  |  |  |  |  |  |  |  |  | X | X |  |  |  |  |  |
| Act. #19 |  |  |  |  |  |  |  |  |  |  | X | X | X |  |  |  |
| Act. #20 |  |  |  |  |  |  |  |  |  |  |  |  | X |  |  |  |
| Act. #21 |  |  |  |  |  |  |  |  |  |  |  |  | X | X | X |  |
| Act. #22 |  |  |  |  |  |  |  |  |  |  |  |  |  |  | X | X |



## Resultados

Después de haber diseñado, programado y probado los modelos de aprendizaje automático, se obtuvieron resultados satisfactorios gracias a la gran cantidad de datos que se insertaron. Estos datos corresponden desde el año 2015 hasta agosto del 2022. Es importante destacar que los modelos de aprendizaje automático dan mejores resultados en función de la cantidad de datos con los que se entrenan.

Inicialmente, los modelos estaban siendo entrenados con datos horarios de las variables con las que tenían más correlación: temperatura, humedad y hora. Esto tiene sentido gracias a un informe proporcionado por el IDEAM denominado "Características Climatológicas de Ciudades Principales y Municipios Turísticos" desde el año 2018 hasta julio del 2022.

Durante este lapso, se obtuvieron resultados satisfactorios, pero no como se esperaba, ya que había una diferencia promedio de 1.06°C centígrados. Por esta razón, se decidió procesar una mayor cantidad de datos, tal como se mencionó anteriormente, y gracias a esta decisión se lograron mejores resultados en las predicciones de temperaturas para el mes de septiembre de 2022 (el mes en el que se realizaron las pruebas). El modelo que dio los mejores resultados tuvo un promedio de 0.5239, lo cual posiblemente se debió a que se le pasaron las variables que tenían una mejor correlación con la variable a predecir.

Se llevaron a cabo pruebas utilizando tanto datos en tiempo real como datos previamente procesados para entrenar modelos de Machine Learning. Se presentarán a continuación los resultados obtenidos mediante la comparación de estos datos y el análisis en tiempo real.

Para iniciar, se generaron promedios diarios correspondientes al mes de septiembre por cada modelo, utilizando la columna de temperatura real. Como resultado de este proceso, se obtuvo la siguiente tabla:

Tabla 3

*Promedio de datos diarios del mes de septiembre*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Dia** | **Regresión** | **KNN** | **Red Neuronal** | **Temperatura Real** |
| 1 | 11,58500565 | 14,8468652 | 11,93305542 | 12,44827586 |
| 2 | 13,10572763 | 15,66800964 | 12,97155187 | 14,40606061 |
| 3 | 14,00038708 | 14,5059219 | 13,94292552 | 13,72535211 |
| 4 | 13,8839865 | 15,30038412 | 13,83148508 | 12,65352113 |
| 5 | 14,45140117 | 14,75294494 | 14,41887829 | 13,44084507 |
| 6 | 14,94901863 | 14,57443182 | 15,0269089 | 14,22142857 |
| 7 | 15,44891863 | 14,76535354 | 15,78945256 | 14,90444444 |
| 8 | 16,22580546 | 14,75277778 | 16,27810271 | 14,77222222 |
| 9 | 15,09425634 | 14,76265783 | 15,09664726 | 13,27083333 |
| 10 | 13,16260775 | 17,46681818 | 13,87530648 | 12,06 |
| 11 | 16,44574154 | 15,08876263 | 16,50773894 | 14,56666667 |
| 12 | 12,59657752 | 14,53181818 | 13,03182398 | 9,119230769 |
| 13 | 14,39138087 | 14,8208042 | 14,6356572 | 14,62051282 |
| 14 | 14,55638747 | 14,61255088 | 14,51098877 | 14,98059701 |
| 15 | 14,84836203 | 15,24057239 | 15,05490849 | 15,57962963 |
| 16 | 12,02611449 | 14,17376033 | 12,36879407 | 11,92272727 |
| 17 | 15,14094452 | 15,44753467 | 15,28326219 | 14,36271186 |
| 18 | 13,88204393 | 15,24247995 | 13,82148561 | 13,23823529 |
| 19 | 13,62902781 | 15,30674931 | 13,50815916 | 14,15 |
| 20 | 11,76027029 | 14,84061265 | 12,14688717 | 10,87826087 |
| 21 | 13,35087066 | 14,86994573 | 13,28653228 | 13,8119403 |
| 22 | 12,21693965 | 14,55942761 | 12,52110213 | 12,07407407 |
| 23 | 14,71228764 | 15,14236771 | 14,65237174 | 14,44925373 |
| 24 | 15,42424684 | 14,70319499 | 15,42093813 | 14,37536232 |
| 25 | 16,45876036 | 15,2681479 | 16,47694129 | 14,2641791 |
| 26 | 15,39763814 | 15,50956439 | 15,41340955 | 14,72777778 |
| 27 | 14,20176254 | 15,11073232 | 14,37979109 | 13,79444444 |
| 28 | 13,65478998 | 15,10619501 | 13,70350618 | 14,83709677 |
| 29 | 11,43839434 | 15,12045455 | 12,10371437 | 11,8 |
| 30 | 11,58500565 | 14,8468652 | 11,93305542 | 12,44827586 |

*Fuente: Elaboración propia*

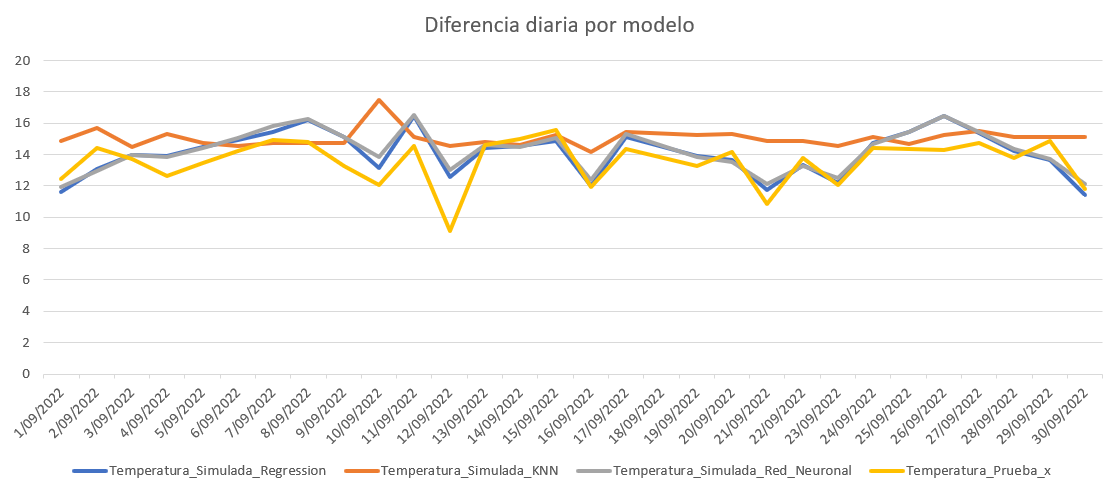
La tabla presenta los resultados de diferentes técnicas de modelado aplicadas a la predicción de la temperatura real en un período de 30 días. Las técnicas de modelado incluyen Regresión, K-Nearest Neighbors (KNN) y Red Neuronal.

Cada fila representa un día en particular y las columnas corresponden a las diferentes técnicas de modelado y la temperatura real registrada para ese día. Los valores en las celdas de cada columna son las predicciones realizadas por la técnica correspondiente para la temperatura real en ese día.

En general, se puede observar que la regresión y la red neuronal logran predecir la temperatura real con mayor precisión que KNN. Además, la regresión y la red neuronal tienen un rendimiento similar en términos de precisión de la predicción.

De igual manera, se representó la anterior tabla en una gráfica en donde representa mejor las temperaturas obtenidas.

**Figura 6** *Temperaturas obtenidas por modelo*



*Fuente: Elaboración propia*

Como fue mencionado anteiormente, la grafica representa los resultados de cada dia de los 30 dias del mes de septiembre, de esta manera, se puede observar que los modelos de Regresión, KNN y Red Neuronal presentan resultados variados en la predicción de la temperatura real. Algunos modelos se desempeñan mejor en ciertos días que en otros, y algunos incluso presentan valores extremos por ejemplo, la temperatura obtenida por medio de KNN el 10 de septiembre, estos resultados pueden ser considerados como anomalías.

Es necesario realizar un análisis más detallado de los resultados para determinar qué modelo es el más adecuado para la predicción de la temperatura. Se deben considerar aspectos como la precisión de la predicción, la velocidad de procesamiento y el costo del modelo.

Se realizaron cálculos de diferencia entre los datos horarios obtenidos y los datos del conjunto de datos para cada uno de los modelos de machine learning (regresión, KNN y red neuronal). Se calcularon los promedios de estas diferencias, obteniendo como resultado un valor de 0.523 para el modelo de regresión, 1.052 para el modelo KNN y 0.576 para el modelo de red neuronal.

Estos resultados sugieren que el modelo de regresión tiene una menor discrepancia con el conjunto de datos original en comparación con los modelos KNN y red neuronal. Sin embargo, es importante tener en cuenta que una discrepancia mayor no necesariamente indica una menor precisión del modelo en la predicción de la temperatura real.

Es necesario realizar una evaluación más detallada de los resultados y considerar otros factores, como la precisión de la predicción, el tiempo de procesamiento y la eficiencia del modelo en diferentes condiciones climáticas.

Y comparando con las temperaturas en tiempo real se obtuvieron resultados gratificantes (ver Anexo 1)

## Conclusiones

En resumen, el modelo de Machine Learning desarrollado para predecir variables climáticas y utilizarlos en el desarrollo de proyectos IoT ha demostrado ser una herramienta eficiente y precisa para analizar y pronosticar el comportamiento del clima. Los resultados obtenidos a través de la evaluación de tres modelos diferentes, Regresión Lineal, KNN y Red Neuronal, han sido evaluados con diferentes métricas de error, como el error absoluto, el error cuadrático medio, el pseudo huber y el promedio de diferencia con la temperatura real, y han arrojado resultados cercanos por muy poca diferencia a los datos reales. En general, el modelo por Regresion Lineal ha mostrado la mayor precisión en las predicciones, con el menor promedio de diferencia con la temperatura real, seguido por el modelo de Red Neuronal y KNN, pero la elección del modelo dependerá de los datos específicos y las necesidades del usuario. Aunque los modelos de Red Neuronal y KNN también han proporcionado resultados alentadores, los resultados de Regresión lineal sugieren que este modelo tiene un mayor potencial para su uso en pruebas de proyectos IoT pero la elección del modelo dependerá de los datos específicos y las necesidades del usuario.

Los resultados indican que el modelo de machine learning puede ser utilizado para proporcionar pronósticos precisos del clima y de esta manera realizar un entorno simulado de datos climaticos, lo que puede ser especialmente importante en el contexto de proyectos IoT, donde la precisión en la toma de decisiones es fundamental. En conclusión, el modelo de machine learning para predecir variables climáticas tiene un gran potencial para mejorar la eficiencia y precisión en proyectos IoT y podría ser una herramienta valiosa en la planificación del desarrollo sostenible y la gestión del clima en general.

## Recomendaciones

Para entrenar modelos de Machine Learning de manera efectiva, es fundamental tener en cuenta algunas recomendaciones clave. En primer lugar, es importante tener en cuenta que los modelos de Machine Learning funcionan mejor cuando se trabaja con una gran cantidad de datos de alta calidad. Por lo tanto, es fundamental obtener una gran cantidad de datos para entrenar el modelo y hacer que sea más preciso.

Sin embargo, la cantidad de datos por sí sola no es suficiente. También es necesario prestar atención a la calidad de los datos y asegurarse de que estén limpios y procesados de manera adecuada. Esto es especialmente importante en el caso de la obtención de datos climáticos, donde los sensores pueden fallar y dar valores atípicos. Por lo tanto, una buena calidad, procesamiento y elección de datos es la segunda recomendación clave para entrenar modelos de Machine Learning de manera efectiva.

A pesar de la importancia de los dos pasos anteriores, es importante tener en cuenta que no serían posibles en gran medida si la máquina en la que se trabajará para procesar los datos y entrenar el modelo es limitada en recursos. Por esta razón, se recomienda hacer todo el proceso en una máquina virtual proporcionada por una plataforma de servicios de computación en la nube como GCP, Azure, AWS, entre otros. Esto permitirá que el proceso sea más rápido y eficiente, lo que a su vez aumentará la precisión del modelo.

## Referencias

Ahrenholz, J., Danilov, C., Herderson, T., & Kim, J. (2008). A real-time network emulator.In Military Communications Conference. *MILCON*, 1-7.

Amine Khelif, M., Lorandel, J., Romain, O., Regnery, M., & Baheux, D. (2019). A Versatile Emulator of MitM for the identification of vulnerabilities of IoT devices, a case of study: smartphones. In A. f. Machinery (Ed.), *3rd International Conference on Future Networks and Distributed Systems*, (pp. 1-6). Paris. Retrieved Abril 2, 2020, from https://dl.acm.org/doi/pdf/10.1145/3341325.3342019

AprendiendoArduino. (2018, Noviembre 17). *Protocolos IoT Capa Aplicación*. Retrieved Abril 14, 2020, from Aprendiendo Arduino: https://aprendiendoarduino.wordpress.com/2018/11/17/protocolos-iot-capa-aplicacion/

AWS. (n.d.). *IoT Device Simulator*. Retrieved from IoT Device Simulator: https://aws.amazon.com/es/solutions/implementations/iot-device-simulator/

Barbara IoT. (2021, Abril 29). *barbara*. Retrieved from Protocolos de comunicación en IoT que deberías conocer: https://barbaraiot.com/blog/protocolos-iot-que-deberias-conocer/

*CambioDigital*. (2018, diciembre 12). Retrieved marzo 13, 2020, from IoT: Qué necesitan saber los profesionales de la red: https://cambiodigital-ol.com/2018/12/iot-que-necesitan-saber-los-profesionales-de-la-red/

Carlemany, U. (2022, 7 19). *universitatcarlemany*. Retrieved from https://www.universitatcarlemany.com/actualidad/cual-es-la-diferencia-entre-inteligencia-artificial-y-machine-learning#:~:text=La%20IA%2C%20se%20puede%20decir,all%C3%A1%20de%20la%20inteligencia%20humana.

Carlos Gamero Burón, J. L. (2015). *Modelos probabilísticos para Variables aleatorias continuas.* Malaga, España.

Castellanos Hernández, W. E., & Chacon Osorio, M. E. (2006, Abril 17). Utilización de herramientas software para el modelado y la simulación de redes de comunicaciones. *GTI, V*(11), 74-75. Retrieved Marzo 26, 2020, from https://revistas.uis.edu.co/index.php/revistagti/article/view/1624/2014

Corso, C., & Lorena, C. (2009). *Aplicacion de algoritmos de clasificacion sepervisan y no supervisada usando Weka.* cordoba: Universidad Tecnologi Nacional.

Coss Bu, R. (2003). *Simulación un enfoque practico.* Monterrey, Mexico: Limusa. Retrieved Marzo 25, 2020, from https://books.google.es/books?hl=es&lr=&id=iY6dI3E0FNUC&oi=fnd&pg=PA11&dq=simulacion&ots=uKV85h0Scu&sig=fMdImFTXdSYn3HghHIZ7HbFXQhg#v=onepage&q=simulacion&f=false

Crespo Moreno, J. E. (2018, Noviembre 11). *Aprendiendo Arduino*. Retrieved Abril 2, 2020, from Arquitecturas IoT: https://aprendiendoarduino.wordpress.com/2018/11/11/arquitecturas-iot/

Cuatro tipos de analítica de retail que todo comercio necesita en 2018. (2018, Agosto 7). *Analitica de retail*. Retrieved from Analitica de retail: http://analiticaderetail.com/tipos-de-analitica-de-retail/

Deloitte. (n.d.). *IoT - Internet Of Things*. Retrieved from Deloitte: https://www2.deloitte.com/es/es/pages/technology/articles/IoT-internet-of-things.html

Diaz, T., Escriba, F., & Murgui, M. (2002). La base de datos BD. *MORES. Revista de Economia Aplicada*, 165-184.

Dra. Marianella Álvarez Vega, D. L. (2020, 8 8). *Revista Medica Sinergia.* Retrieved from https://www.revistamedicasinergia.com/index.php/rms/article/view/557/923

Dunkles, A., Schmidt, O., Finne, N., Erikson, J., Osterlind, F., Tsiftes, N., & Durvy, M. (2011). *The contiki os: The operating system for the internet of things.* Retrieved from Online: http://www. contikios. org

editorial, E. (2018, Noviembre 1). *REPORTEGIGITAL*. Retrieved from REPORTEGIGITAL: https://reportedigital.com/cloud/analitica-de-datos/

Eduardo García Dunna, H. G. (2013). *Simulación y análisis de sistemas con ProModel.* Naucalpan d e Juárez, Estado de México , México : PEARSON. Retrieved from https://jrvargas.files.wordpress.com/2015/04/libro-simulacic3b3n-y-anc3a1lisis-de-sistemas-2da-edicic3b3n.pdf

Gan, S. (2017). *An IoT simulator in NS3 and a key-based authentication architecture for IoT devices using blockchain.* Tesis, Instituto Indio de Tecnología Kanpur, Ciencias informáticas e ingeniería, Kanpur. Retrieved Abril 2, 2020, from https://security.cse.iitk.ac.in/sites/default/files/12807624\_0.pdf

García Sánchez, Á., & Ortega Mier, M. (2006). *Introducción a la simulación de sistemas discretos.* Retrieved Marzo 25, 2020, from http://www.iol.etsii.upm.es/arch/intro\_simulacion.pdf

Garcia, P. (2006). *TECNICAS DE ANALITICA.*

Gibbs, G. (2012). *Analisis de datos en investigaciones cualitativas .* Ediciones Morata.

Gomez-Aguilar, D., Garcia-Peñalvo, F., & Theron, R. (2014). Analitica visual en learning. *El profesional de la informatica* , 23(3).

*Hardwarelibre*. (n.d.). Retrieved from Hardwarelibre: https://www.hwlibre.com/iotify-servicio-web-desarrolladores-hardware-libre/#:~:text=As%C3%AD%20hace%20poco%20hemos%20conocido,compatible%20con%20cualquier%20hardware%20libre.

Hassan, F., Domingo-Ferrer, J., & Soria-comas, J. (2018). Anominacion de datos no estrucutrados a traves del reconocimiento de entidades nominadas. *Actas de la XV Reunni Espaola sobre Criptologa y Seguridad de la informcin-RECSI*, 102-106.

Hergert, M., & Morris, D. (1989). Datos contables para el analisis de la cadena de valor . *Diario de gestion estrategica*, 10(29,175-188.

Hernandez Perez, A. (2013). *Datos abiertos y repositorios de datos .* nuevo reto para los bibliotecarios.

Hernandez, C., & Rodriguez, J. (2008). Preprocesamiento de datos estructurados. *Vinculos*, 27-48.

Huang, Y., Wang, L., Hou, Y., Zhang, W., & Zhang, Y. (2018). *A prototype IOT based wireless sensor network for traffic information monitoring.* International Journal of Pavement Research & Technology.

*ibertech*. (2006, 06 45). Retrieved from ibertech: https://www.ibertech.org/analitica-descriptiva-predictiva-y-prescriptiva/

*ibertech*. (2006, 06 18). Retrieved from ibertech: https://www.ibertech.org/analitica-descriptiva-predictiva-y-prescriptiva/

*IBM* . (n.d.). Retrieved from IBM : https://www.ibm.com/es-es/cloud?

Isaac Lera, C. G. (2019). *YAFS.* Palma. Retrieved from https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8758823

*Itop*. (2018, agosto 20). Retrieved marzo 13, 2020, from IoT: origen, importancia en el presente y perspectiva de futuro: https://www.itop.es/blog/item/iot-origen-importancia-en-el-presente-y-perspectiva-de-futuro.html

Joyanes Aguilar, L. (29 de mayo del 2019). *Inteligencia de negocios y anlitica de datos.* Bogota: Alfaomega.

Leónardo Darío Bello Parias, L. C. (2000). *Libro de estadística descriptiva.* Medellin, Antioquia, Colombia: Editorial Amistad ISBN. Retrieved from http://aprendeenlinea.udea.edu.co/lms/moodle/pluginfile.php/128508/mod\_resource/content/0/Tema\_4/Modelos\_probabilisticos\_Caucasia.pdf

Levis, P., Lee, N., Welsh, M., & Culler, D. (2003). *TOSSIM: Accurate and scalable simulation of entire Tinyos applcations.* In Proceedings of the 1st international conference on Embedded networked sensor systems.ACM.

Loukides, M. (2011). *¿Que es la ciencia de datos?* O'Reilly Media, Inc.

Mäkinen, A. (2016). *Emulation of IoT Devices.* Espoo. Retrieved from https://aaltodoc.aalto.fi/bitstream/handle/123456789/23951/master\_M%c3%a4kinen\_Alli\_2016.pdf?sequence=1&isAllowed=y

Martin Iglesia, M. (2019, Febrero). *Archivo Digital UPM.* Retrieved from Análisis de la simulación de dispositivos, circuitos y sistemas electrónicos para Internet de las cosas (IoT): https://oa.upm.es/54136/

*MathWorks*. (n.d.). Retrieved from MathWorks: https://www.mathworks.com/solutions/internet-of-things.html

Mehmood, T. (n.d.). *COOJA Network Simulator.* Islamabad. Retrieved from https://arxiv.org/ftp/arxiv/papers/1712/1712.08303.pdf

Micrsoft. (n.d.). *Micrsoft*. Retrieved from https://azure.microsoft.com/es-es/overview/machine-learning-algorithms/#overview

Myriam Muñoz de ózak, S. F. (1991). PROCESOS ESTOCÁSTICOS CON DOS PARÁMETROS I. 72-74. Retrieved from https://revistas.unal.edu.co/index.php/estad/article/viewFile/9955/10486

NEO.LCC. (n.d.). *Protocolos de transporte*. Retrieved Abril 14, 2020, from Herramientas WEB para la enseñanza de protocolos de comunicación: http://neo.lcc.uma.es/evirtual/cdd/tutorial/transporte/protrans.html

Nigro, O., Corti, D., & Terren, D. (2004). *Knowledge Discovery in Databases.* Un proceso centrado ene el usuario. In VI Woorkshop de investigadores en Ciencias de lamComputacion.

NSNAM. (2011). *NS-3 Network Simulator*. Retrieved Abril 2, 2020, from NS-3: https://www.nsnam.org/

Oracle. (2022). *www.oracle.com*. Obtenido de https://www.oracle.com/co/artificial-intelligence/what-is-ai/

Peña, S. (2017). *Análisis de Datos.* Bogota D.C: Areandino.

Prado, J. (n.d.). *VALTX*. Retrieved from VALTX: https://www.valtx.pe/blog/que-es-la-analitica-de-datos-y-como-puede-impactar-positivamente-en-tu-negocio

Quintero, J. (2006). La cadena de valor : Una herramienta de pensamiento estraegico. *Telos*, 14.

Quiñones Cuenca, M., González Jaramillo, V., Torres, R., & Miguel , J. (2017). *Sistema de Monitoreo de variables medioambientales usando una red de sensores inalámbricos y plataformas de Internet de las Cosas.* Universidad Técnica Particular de Loja, Departamento de Ingeniería, Loja. Retrieved Abril 14, 2020, from http://scielo.senescyt.gob.ec/scielo.php?script=sci\_arttext&pid=S1390-65422017000100329

Raposo, J. (2007). *Tecnicas de mantenimeiento automatico de programas envoltorio para fuentes de datos web semiestrucuturadas.* Coruña: Doctoral dissertation.

Riverder Tecnologies. (2017). *opnet simulator*. Retrieved from https://www.riverbed.com/in/products/steelcentral/opnet.html

Robles Solano, D. (2021, Marzo). *Repositorio Digital.* Retrieved from Control y simulación de una planta piloto de laboratorio docente con integración de plataformas IoT para subida de datos a la nube: https://repositorio.upct.es/handle/10317/9259

Rodríguez Moreno, E. S., & López Ordoñez, V. F. (2017). *Diseño e implmentación de un sistema inteligente para un edificio.* Tesis de grado, Universidad Francisco Jose de Caldas, Facultad de Ingeniería, Bogotá. Retrieved Marzo 25, 2020

Roldán Carrasco, Á. (2007). *Emulador de Gameboy para dispositivos móviles.* Tesis, Escuela Superior de Ingeniería Informática, Departamento de Informática, Ciudad Real. Retrieved Marzo 25, 2020, from https://www.esi.uclm.es/www/cglez/downloads/pfc/pfcaroldan.pdf

Ruz Nieto, A. (2021). *Repositorio Digital.* Retrieved from Simulación realista de comunicaciones IoT en entornos urbanos: https://repositorio.upct.es/handle/10317/9646

Sánchez Martín, A. A., Barreto Santamaría, L. E., Ochoa Ortiz, J. J., & Villanueva Navarro, S. E. (2019). *Emulador para desarrollo de proyectos IoT y analiticas.* PREGUNTAR, Universidad Distrital Francisco José de Caldas, Bogotá. Retrieved marzo 13, 2020

Sandoval, L. J. (2018, 4 16). ALGORITMOS DE APRENDIZAJE AUTOMÁTICO PARA ANÁLISIS Y. *ITCA FEPADE*, pp. 36-40.

Sandy E. Abasolo, M. A. (2013). *Evaluación del modelo de referencia de “Internet of Things” (IoT), mediante la implantación de arquitecturas basadas en plataformas comerciales, open hardware y conectividad IPv6.*

Semle, A. (2016, Septiembre). Protocolos IIoT para considerar. *AADECA Revista*, 32-35. Retrieved Marzo 25, 2020, from https://editores-srl.com.ar/sites/default/files/aa2\_semle\_protocolos\_ilot.pdf

Snoke, J., & HalanLarochelle. (2012). Practica y optimizacion de algoritmos de aprendizaje automatico. *Avances en sistemas de procesamiento de informacion neuronal*, 2951-2959.

Tetcos. (2017). *Netsim emulator*. Retrieved from http://tetcos.com/

Timarán-Pereira. (2016). *The Process of Knowledge Discovery on Databases .* Bogota : Ediciones .

Torres Bataller, J. (2016). *Desarrollo de una solucion para la simulacion de entornos IoT.*

Universidad de Alcalá. (2019). *¿Por qué actualmente es tan importante el IoT?* Retrieved marzo 13, 2020, from Máster en industria 4.0: https://www.masterindustria40.com/importancia-iot-master/

Varga, A. (2016). *In Modeling and tools for network simulation.* Berlin,Heidelberg: Springer.

Xia, F., Yang, L., Wang, L., & Vinel, A. (2012). *Internet of Things.* International journal of communication systems. doi:10.1002/dac.2417

Yacchirema Vargas, D. C., & Palau Salvador, C. E. (n.d.). *Smart IoT Gateway For Heterogeneous Devices Interoperability* (Octava ed., Vol. 14). IEEE Latin America Transactions. doi:10.1109/TLA.2016.7786378

## Anexos

Anexo 1

*Resultados en tiempo real*