**Simulador para el desarrollo de proyectos IoT y analítica de datos**

**Simulador de 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**

El título debe responder algunos de los siguientes interrogantes:

* ¿Qué? (proceso)
* ¿Sobre qué o quién? (sujeto u objeto)
* ¿Dónde? (localización geográfica)
* ¿Cómo? (técnica o método)
* ¿Cuándo? (épocas, fechas….)

**Camilo Andres Diaz Gomez**

**Jhonatan Mauricio Villarreal Corredor**

**Juan Esteban Contreras Diaz**

**Universidad de San Buenaventura, Sede Bogotá.**

**Facultad de Ingeniería.**

**Programa de Ingeniería de Sistemas**

**Bogotá D.C, Colombia**

**2022**

**Simulador para el desarrollo de proyectos IoT y analítica de datos**

**Camilo Andres Diaz Gomez**

**Jhonatan Mauricio Villarreal Corredor**

**Juan Esteban Contreras Diaz**

**Nikolay Lenin Reyes Jalizev**

**Universidad de San Buenaventura, Sede Bogotá.**

**Facultad de Ingeniería.**

**Programa de Ingeniería de Sistemas**

**Bogotá, Colombia**

**2022**

**DEDICATORIA**

ESTA PÁGINA ES OPCIONAL.

**AGRADECIMIENTOS**

ESTA PÁGINA ES OPCIONAL

**Tabla De Contenido**

[Introducción 3](#_Toc101123023)

[Capítulo 1 1](#_Toc101123024)

[Antecedentes 1](#_Toc101123025)

[Planteamiento del problema 2](#_Toc101123026)

[Justificación y pregunta de Investigación 2](#_Toc101123027)

[Árbol problema 3](#_Toc101123028)

[Objetivo General 4](#_Toc101123029)

[Objetivos Específicos 4](#_Toc101123030)

[Alcances y Limitaciones 4](#_Toc101123031)

[Marco Conceptual 5](#_Toc101123032)

[Internet of Things (IoT) 5](#_Toc101123033)

[Analítica de datos 10](#_Toc101123034)

[Simulación 17](#_Toc101123035)

[Desarrollo de Ingeniería 27](#_Toc101123036)

[Figuras y tablas 27](#_Toc101123037)

[Ecuaciones 27](#_Toc101123038)

[Capítulo 3 29](#_Toc101123039)

[Análisis de Resultados 29](#_Toc101123040)

[CONCLUSIONES 30](#_Toc101123041)

[RECOMENDACIONES 31](#_Toc101123042)

[REFERENCIAS 32](#_Toc101123043)

[Anexo I 33](#_Toc101123044)

**Lista de tablas**

[Tabla 1. Título de Tabla 3](#_Toc536104043)

**Lista de Figuras**

[Figura 1. Formas y descripción de las formas. 3](#_Toc536103760)

**Lista de Ilustraciones**

[Ilustración 1Arbol Problema 3](#_Toc101119164)

[Ilustración 2 Modelo general de la arquitectura de una red IoT 7](#_Toc101119165)

[Ilustración 3 Funcionamiento de actuadores IoT 10](#_Toc101119166)

[Ilustración 5 Cadena de valor 13](#_Toc101119167)

**Nomenclatura**

# Introducción

En la introducción, el autor presenta y señala la importancia, el origen (los antecedentes teóricos y prácticos), los objetivos, los alcances, las limitaciones, la metodología empleada, y las principales conclusiones del trabajo. No debe confundirse con el resumen y se recomienda que la introducción tenga una extensión de máximo de 2 páginas Proporciona al lector una idea completa del trabajo a realizar y su naturaleza. Debe ser claro, preciso y concreto, evitando los títulos demasiado extensos.

En la introducción, el autor presenta y señala la importancia, el origen (los antecedentes teóricos y prácticos), los objetivos, los alcances, las limitaciones, la metodología empleada, y las principales conclusiones del trabajo. No debe confundirse con el resumen y se recomienda que la introducción tenga una extensión de máximo de 2 páginas.

La presente plantilla tiene en cuenta aspectos importantes de la Norma Técnica Colombiana - NTC 1486 y el Manual de publicaciones de la APA, con el fin que sean usadas para la presentación final del proyecto de grado.

# Capítulo 1

## 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 que ha tenido este término en la actualidad junto al desarrollo de la tecnología. En el mercado nos podemos encontrar con software como:

<https://aws.amazon.com/es/solutions/implementations/iot-device-simulator/>

* **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**: 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 de usuario (GUI) intuitiva 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 de IBM: Es un entorno de plataforma como servicio desarrollado por IBM. Admite múltiples lenguajes de programación y servicios1 y metodologías de desarrollo DevOps de manera integrada para crear, ejecutar, implementar y administrar aplicaciones en la nube. Bluemix se basa en la tecnología abierta de Cloud Foundry y se ejecuta en la infraestructura SoftLayer.** (IBM , s.f.)

**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 es con dispositivos en físico y al tenerlos solo se conectan y este empezara enviar datos de forma segura por medio de MQQT,** Puede configurar y gestionar los dispositivos mediante el panel de control en línea o nuestras API seguras.

* **Iotify: Simulador IoT que te permite desarrollar rápidamente soluciones IoT en la nube. Esta herramienta le permite simular instalaciones de IoT a gran escala en su propio laboratorio de IoT virtual.**

**Simular una flota de vehículos conectados o una red de cámaras de vigilancia nunca ha sido tan fácil. Puede desarrollar sus modelos de IO utilizando plantillas de Javascript y generar tráfico en tiempo real a cualquier proveedor de plataforma en la nube a través de MQTT, HTTP o CoAP.** (Hardwarelibre, s.f.)

**Es un hardware libre al no tenerlo físicamente, todo esto con la creación de proyectos elegir porque medios se realiza la simulación, ayuda a las personas con pocos recursos con esto obtuvo un gran impacto, pero con el consumo de esto es necesario tener ganancias con esto colocan que pueden realizar un solo proyecto por la prueba gratis y si desean más en necesario comprar una suscripción por mes para poder seguir utilizándola.**

* **Matlab: Esta herramienta que tiene módulos de simulación de IO para poder desarrollar y modelar dispositivos inteligentes para así poder recopilar y analizar los datos en la nube.**

Más específicamente, puede desarrollar algoritmos en Simulink y luego implementarlos en su hardware embebido. También puedes crear prototipos de tus dispositivos inteligentes usando Arduino y Frambuesa Pi. (MathWorks, s.f.)

Esta herramienta permite acceder a archivos y mediante una interfaz poder almacenarlos en la nube, bases de datos, y permite los protocolos de MQTT o REST, el análisis de algoritmos de IoT personalizados para probar miles de funciones prediseñadas y tener una limpieza de daos para un mejor control y optimización; con esto admite datos no estructurados y con marca de tiempo de muchas fuentes, incluidos los servicios de almacenamiento en la nube.

De igual manera, se encuentran proyectos que están dirigidos al desarrollo o implementación de redes IoT, como lo son:

https://repositorio.upct.es/handle/10317/9259

https://repositorio.upct.es/handle/10317/9646

https://oa.upm.es/54136/

PICO…..

## Planteamiento del problema

### Justificación y pregunta de Investigación

En la actualidad, una gran parte de los objetos de uso diario por las personas están conectados a internet por diferentes propósitos, este concepto es conocido 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 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 técnicas que les ayudan a mejorar el rendimiento en diferentes aspectos por que los dispositivos IoT están encargados de la obtención de datos y su envió a la nube, 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 (Abasolo, Carrera, Gordillo, & Romero, 2013). Ver de donde salió este.

Actualmente, hay demasiadas personas que se desempeñan en el diseño, instalación y mantenimiento de estas conexiones como desarrolladores, investigadores, técnicos y demás en redes IoT, estos a su vez buscan alguna ayuda para poder desarrollar sus pruebas sin la necesidad de gastar muchos recursos; por esta razón, las dificultades más 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.

### Árbol problema

Diagrama

Descripción generada automáticamente

Ilustración Arbol Problema

En la ilustración 1 Arbol Problema, se tiene una estructura de un árbol para poder explicar mejor la justificación del proyecto, con esto se entiende que 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 el o los usuarios. Por esta razón es muy importante seleccionar los dispositivos que más se adapten a las necesidades de la red IoT; 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.

Con lo anterior, se generó la siguiente pregunta:

¿Cómo simular datos ambientales en un modelo IoT para validar herramientas de analítica de datos?

### 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 simulados.

### Alcances y Limitaciones

Alcances:

* Toda la información generada será almacenada en una base de datos.
* El emulador se basará en un modelo realístico para la generación de datos IoT.

Limitaciones:

* El emulador no tendrá todos los sensores y actuadores que hay disponibles en la actualidad.
* El emulador no será totalmente preciso, aun así, se buscará la cercanía a la realidad.
* El emulador no realizara analítica de los datos obtenidos.
* El emulador no tendrá manual de uso.

## Marco Conceptual

### Internet of Things (IoT)

El Internet de las cosas es un nuevo paradigma del mundo moderno el cual es la conexión de varios nodos IoT (sensores de recopilación de información, electrodomésticos, celulares, computadores, entre otros que tienen la capacidad de conectarse al internet) que contienen integrado sensores los cuales son los dispositivos que recopilan información de su entorno por ejemplo, sensores de temperatura, gas, humo, humedad, velocidad del viento, etc, por otro lado, los actuadores 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).

Hoy en día, como fue mencionado anteriormente, la mayoría de los dispositivos integran este nuevo concepto y esto por su gran versatilidad a los diferentes usos que se le puedan aplicar a cada uno y no es extraño que la humanidad se está anticipando a un cambio tecnológico en donde IoT será tan importante e indispensable para la vida cotidiana de los humanos, algo no tan diferente a lo que estamos viviendo ahora.

#### Áreas de aplicación

La humanidad está pasando por una transformación en el cual en un futuro dependerá de la tecnología directamente, por esta razón, IoT va a ser un gran pilar para este cambio, gracias a esto, está siendo implementado en diferentes sectores como lo son agricultura, salud, transporte y logística, seguridad, entre otros.

#### Beneficios

La aplicación de este concepto es variada y funcional, ya que, es versátil para el objetivo que se quiera cumplir con estas conexiones por que al ser generados datos por el ambiente donde se encuentre montado y por último ser analizados dependiendo el objetivo cuya ejecución haya sido planeada (Universidad de Alcalá, 2019).

Por ejemplo, industrialmente se ha evidenciado la importancia de hacer la transformación a entornos IoT (PREGUNTAR ESTUDIO DELOITTE) ya que estos minimizan esfuerzos y riesgos para aumentar productividad, de esta manera, ser más efectivos y eficientes en los procesos, esto gracias a la gestión y monitorización remota de la maquinaria, también conocimiento del funcionamiento general del trabajo (gracias a los datos obtenidos) para poder realizar análisis de este y aplicar mejoras, y demás como conocimiento del estado de las maquinas, control de calidad, inventario y demás.

También en la agricultura se ha visto beneficiada de esta tecnología ya que este sector puede aplicarlo en la digitalización y mejora de la productividad agrícola, a esto se le conoce como Agricultura Inteligente a la combinación de herramientas tecnológicas que permite a los productores obtener información actualizada las 24 horas del día, los 7 días de la semana sobre el estado de sus cultivos, gracias a esto se puede operar de forma remota diferentes equipos agrícolas con el objetivo de mejorar la cantidad y calidad de los cultivos. Así como estos sectores obtienen beneficios de las redes IoT se pueden encontrar más en los sectores como el gubernamental, transporte, educación y demás.

#### Modelos de referencia



Ilustración Modelo general de la arquitectura de una red IoT

En la anterior Ilustración 1 Modelo general de la arquitectura de una red IoT se puede evidenciar un modelo general de IoT, donde 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, podemos ver percepción, el cual contiene todos los sensores que recolectan la información para ser enviada a la red o nube, la cual, es la segunda capa, por medio de los gateways y finalmente en la tercera capa se encuentran los actuadores que son los que reciben 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).

#### 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**

Gracias al crecimiento, desarrollo e implementación de dispositivos IoT se han establecido diferentes protocolos IoT para la gestión de la comunicación.

Para establecer el tipo de protocolo que se debe implementar en una red IoT se tiene que tener en cuenta los dispositivos que se interconectaran, función o objetivo a cumplir, la distancia que viajaran para la transición de los datos. Aun así, existen dos tipos de protocolos:

* **Protocolos de acceso de red:** es usada en la capa inferior que permite la conexión entre los dispositivos, la cual, se puede hacer por medio de Wifi, Ethernet, 3G, 4G, 5G, etc.
* **Protocolos de transmisión:** se usado en la transmisión de los datos, este codifica la información que se envía a través de las redes.

#### 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 gateways es enviada todos los datos 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 la situación que en el que esté funcionando.

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



Ilustración Funcionamiento de actuadores IoT

**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 y después de ser procesada recibirla y enviarlo a los actuadores (Crespo Moreno, 2018).

#### Plataformas IoT

En la actualidad, 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 la utilización de información que se puede tener u 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 en la gran búsqueda de variables.

La analítica de datos puede clasificarse en tres grandes categorías: analítica descriptiva, analítica predictiva y analítica prescriptiva (Pusala, Amini, Katukuri, Xie y Raghavan (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 respectivas variables con el de realizar analisis de correlación.
* La **analítica de datos predictiva** donde las empresas, compañías con datos anteriores registrados generan modelos de pronósticos sobre las tendencias y así poder realizar cambios para mejorar.
* La **analítica de datos prescriptiva** donde las compañías utilizan modelos de simulación de escenarios, de optimización sobre diferentes fuentes de interés.

#### Áreas de aplicación de analítica

Las áreas donde se puede aplicar la analítica son un poco extensas ya que muchas actividades o procedimientos que realizamos es necesario hacer una investigación anteriormente para poder obtener unos resultados apropiados para poder analizar y así poder utilizarlos en un propósito de sacar conclusiones sobre la información tratada; el uso en industrias para tomar mejores decisiones, el usos verificar teorías y modelos existentes, la clasificación de conjuntos de datos para obtener un relación y utilizarlos en el mejoramiento de campañas, empresas. (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 los datos

La 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 origen de datos de la lista 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.
6. Impactar: es el uso que se le pueden dar a los datos frente al impacto que pude tener frente a la empresa o organización



Ilustración Cadena de valor

Esto también lleva a una extensa investigación mediante el cual se utilizan ciertas etapas para poder realizar (Hergert & Morris, 1989).

Tabla Etapas para poder realizar una cadena de valor

|  |  |  |  |
| --- | --- | --- | --- |
| Collection | Publication | Uptake | impact |
| Indentify | Analyze | Connect | Use |
| Collet | Reléase | Inccentivize | Change |
| Proces | Disseminate | Influence | reuse |

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 la plantación de estrategias o contingencias para el manejo de los respectivos datos pulidos (Hergert & Morris, 1989).

#### Datos

Representación simbólica de alguna información o procedimiento en la cual puede ser almacenado y analizado para poder realizar ciertas operaciones para poder generar información adecuada para la implementación en campañas para poder tomar decisiones frente a el mejoramiento de procesos en los diferentes campos o predicciones al futuro para el desempeño de empresas o establecimientos. Toda Información debe estar disponible en todo momento y estos tiene un valor justo en el cual los datos deben tener privilegios para poder distribuirlos, reutilizarlos. (Loukides, 2011)

**Datos abiertos**

Los datos abiertos son datos o información que se pueden utilizar, publicar y reutilizar tantas veces quiera sin ninguna repercusión, pero los datos no pueden ser modificados (Hernandez Perez, 2013).

**Fuentes de datos**

La fuente de datos son conjuntos de información con sus respectivos datos recolectados para su respectivo análisis en la cual son fuentes de información para nivel informático y analítico (Diaz, Escriba, & Murgui, 2002).

**Estructurados**

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

**No estructurados**

Son esencialmente datos binarios que no tiene una estructura u organización que no tiene algún valor al utilizarlos hasta que son organizados y almacenados, el cual su manejo es mucho más dificultoso que en los demás, estos datos no se pueden usar en una base tradicional 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).

#### Ciclo de la analítica KDD – Metodologías

El proceso de extraer conocimiento a partir de grandes volúmenes de datos la cual es un proceso automático en el que se combinan descubrimiento y análisis. La cual tiene unos pasos a seguir (Nigro, Corti, & Terren, 2004).

* Selección: se crea un conjunto de datos objetivo seleccionando todo el conjunto de datos o una muestra representativa de este.
* Preprocesamiento/limpieza: Se analiza la calidad de los datos, se aplican operaciones básicas como la remoción de datos ruidosos, se seleccionan estrategias para el manejo de datos desconocidos.
* Transformación/reducción: Se buscan características útiles para representar los datos dependiendo de la meta del proceso.
* Minería de datos: Es la búsqueda y descubrimiento de patrones insospechados y de interés, aplicando tareas de descubrimiento como clasificación.
* Interpretación/evaluación: Se interpretan los patrones descubiertos y posiblemente se retorna a las anteriores etapas para posteriores iteraciones.

#### Tipos de analítica

Es necesario saberlos ya que con el aumento de la cantidad de datos que generan actualmente empresas, negocios se pueden extraer grandes cantidades de información las cuales se pueden utilizar para mejoramiento en empresas como predecir que mes se puede obtener más ganancias o el que menos se obtiene.

“*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**

¿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).

**Diagnostica**

¿Por qué está pasando? o ¿Por qué está pasando? Tiene en cuenta los antecedentes de lo que se quiere analiza para dar un informe más acertado con sus respectivas herramientas para poder eliminar el problema, con esto tener las herramientas necesarias para que el respectivo análisis en los datos se puedan obtener la causa de los problemas (cuatro tipos de analítica de retail que todo comercio necesita en 2018, 2018).

**Predictiva**

¿qué podría pasar? o ¿Qué es lo más probable que pueda pasar? Tienen como objetivo identificar la probabilidad que ocurra algo en el futuro que no perjudique o perjudique a análisis realizado; estos modelos se suelen utilizar con datos o variables que se puedan hacer predicción el cual se puede 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**

¿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

pronosticar resultados potenciales sobre la base de posibles variaciones en las variables, 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).

**Machine Learning**

La técnica de machine learning es un apoyo para el conocimiento de ciertas generaciones, tiene una organización en el cual es el auto aprendizaje que muestra estadísticas con una gran velocidad de respuesta. Esto es una disciplina la cual trata de crear o construir modelos complejos, también algoritmos que buscan llegar o alcanzar una predicción, el cual funciona sin dar alguna instrucción u orden la cual busca que se actualice automáticamente con los datos que adquirió para poder adaptarse a la situación que se les presente (Snoke & HalanLarochelle, 2012).

#### Tipos de algoritmos

Los algoritmos de aprendizaje automático son fragmentos de código que ayudan a los usuarios a explorar y analizar conjuntos de datos complejos y encontrar significado en ellos. Cada algoritmo es un conjunto limitado de instrucciones explícitas paso a paso que una máquina puede seguir para lograr un objetivo determinado. (Micrsoft, s.f.)

**Supervisados**

*“los algoritmos trabajan con datos “etiquetados” (*labeled data)*, intentado encontrar una función que, dadas las variables de entrada (*input data*)”* (Corso & Lorena, 2009).

**No supervisados**

*“no se dispone de datos “etiquetados” para el entrenamiento. Sólo conocemos los datos de entrada, pero no existen datos de salida que correspondan a un determinado*input” (Corso & Lorena, 2009).

### Simulación

La simulación es una representación exacta del comportamiento interno de un evento o fenómeno, diseñado para obtener el mismo resultado, características, información entre otros, 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

* Simulación de situaciones: Permite simular una situación física o real y observar su comportamiento (Castellanos Hernández & Chacon Osorio, 2006).
* 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

* 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, pero nos centraremos más en el modelo Estocástico:

* Estático: La simulación no depende del tiempo (García Sánchez & Ortega Mier, 2006).
* Dinámico: La simulación depende del tiempo, sus procesos pueden variar respecto al tiempo que va transcurriendo (García Sánchez & Ortega Mier, 2006).
* Determinístico: Su ejecución será siempre igual, el valor de su resultado será ya esperado, no tiene ninguna variable o proceso al azar (García Sánchez & Ortega Mier, 2006).
* Estocástico: Contiene variables o procesos al azar el cual puede variar el resultado de la simulación (García Sánchez & Ortega Mier, 2006).
* Discreto: Varía dependiendo de sucesos que ocurran en la simulación del modelo (García Sánchez & Ortega Mier, 2006).
* Continuo: Tiene un rango de tiempo el cual es previamente establecido. Sin importar los sucesos que ocurran seguirá ejecutándose.
* Físicos: Se basan en eventos físicos o fenómenos que ocurren y no son posibles de controlar y estudiar (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 queremos 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 estamos intentando emular, es decir que un suceso en el sistema puede variar en ese mismo proceso u otro proceso independiente, por ejemplo, si queremos 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 hablamos de un modelo probabilístico nos referimos 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 una conjunto de datos de una población mayor, con esto nos referimos 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 nos 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 debemos 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 queremos hacer una 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).

Es la base o fundamentación sistémica, organizada y objetiva para afrontar el problema con la suficiencia requerida, ya que allí se relacionan el conocimiento básico y el especializado que existe en torno al tema específico del mismo. Como características esenciales de este marco cabe mencionar: precisión y concisión. No se trata de transcribir grandes textos sobre la materia, sino de sintetizar los principales conceptos y teorías alrededor del problema de investigación.

Establece relaciones entre los estudios consultados y el planteamiento del problema. Hace explícito el aporte de los estudios (antecedentes) consultados, al objeto de estudio de la investigación; siendo estos actuales y pertinentes con el tema.

#### Metodología

Depende del enfoque de investigación que se ha estimado conveniente para la naturaleza del problema de investigación a resolver. La metodología se expresa en una estrategia de trabajo organizada y precisa, junto con las diferentes actividades, técnicas, métodos, procesos, procedimientos, ensayos, pruebas, diseño experimental, modelo estadístico, etc. necesarios para alcanzar cada uno de los objetivos propuestos y dar respuesta al problema planteado incluyendo la recolección de la información, su organización, sistematización, análisis y presentación de resultados.

En cuanto al enfoque a emplear en la investigación y de acuerdo a las políticas para la investigación establecidas por la Universidad, éste puede ser:

* **Empírico-analítico:** cuyo interés es el técnico, orientado a la interpretación y transformación del mundo material. Proporciona una estructura particular a la metodología de investigación en tanto que orienta el trabajo a la contrastación permanente de las aseveraciones teóricas con la verificación experimental, de manera que los cálculos generados a través de modelos matemáticos y simulaciones computacionales se deben retroalimentar con la experimentación, en la búsqueda de información cada vez más confiable y práctica para la solución del problema. Esta simbiótica debe llevar consigo una relación teórica al menos presumible entre variables, de manera que se puedan establecer relaciones funcionales entre ellas; igualmente y de acuerdo con los medios experimentales, también se deben establecer los parámetros experimentales convenientes.
* **Histórico-hermenéutico:** Cuya estrategia de trabajo se centra en el estudio y análisis de textos, así como la confrontación de ideas, opiniones y razonamientos de diversos autores en marcos históricos diversos que permiten la construcción de nuevas ideas y opiniones que significan o reportan para la solución del problema de investigación un aporte al marco del conocimiento en el área de debate. En el área de ingeniería este enfoque es pertinente a la construcción de documentos de ingeniería que buscan constituirse como manuales de trabajo, manuales de procedimientos de certificación, normas de calidad, de diseño o experimentación, etc. surgidos del estudio de documentación y otros estudios pre-existentes sin involucrar experimentación o simulación computacional.
* **Crítico- social**: encaminado a la crítica y a la transformación del mundo social8; en particular, se entiende para ingeniería, el enfoque de investigación que orienta proyectos hacia el análisis de las consecuencias o impacto social, logrado a partir de la implantación y uso de nuevas tecnologías (software, dispositivos electromecánicos, entre otros).

Capítulo 2

## Desarrollo de Ingeniería

Descripción clara, detallada y completa de los procedimientos realizados para cumplir los objetivos. Se debe describir de forma clara, concisa, lógica y organizada cada uno de los experimentos/pruebas que se diseñaron para validar los diferentes elementos/bloques funcionales/fases/componentes del sistema/software/hardware/metodología desarrollados

### Figuras y tablas

*Tabla 1. Título de Tabla*

|  |  |
| --- | --- |
| ***Columna X*** | ***Columna X*** |
| Dato | Dato |
| Dato | Dato |
| Dato | Dato |

#### 

Figura . Descripción de la Figura.

### Ecuaciones

Las ecuaciones deben estar centradas y numeradas a la derecha. Se debe insertar una línea vacía antes y después de la ecuación:

 (1)

Asegúrese de que los símbolos de su ecuación se definen antes o inmediatamente después de la ecuación o agregue una nomenclatura.

# Capítulo 3

## Análisis de Resultados

Mostrar de forma concisa, organizada y siguiendo un orden lógico la información (figuras, tablas, etc.) relacionada con cada una de las pruebas definidas. Se sugiere la siguiente estructura:

(i) Analiza los resultados obtenidos a la luz de lo que predice la teoría y de lo que se esperaba del experimento y describiendo explícitamente si el resultado obtenido permite o no verificar el correcto funcionamiento del sistema y el cumplimiento de los objetivos específicos asociados.

(ii) Compara con otros trabajos realizados previamente (si aplica) a nivel local, nacional o internacional; esto resulta de gran utilidad pues permite evidenciar el aspecto innovador del trabajo (y permitiría además identificar su potencial para ser publicado posteriormente).

# CONCLUSIONES

Las conclusiones deben ser la respuesta a los objetivos o propósitos planteados. Deben contemplar las perspectivas de la investigación, las cuales son sugerencias, proyecciones o alternativas que se presentan para modificar, cambiar o incidir sobre una situación específica o una problemática encontrada. Pueden presentarse como un texto con características argumentativas, resultado de una reflexión acerca del trabajo de investigación.

# RECOMENDACIONES

ESTA SECCIÓN ES OPCIONAL

Se presentan como una serie de aspectos que se podrían realizar en un futuro para emprender investigaciones similares o fortalecer la investigación realizada.

# REFERENCIAS

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

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

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

Blankenberg, D., Kuster, G. V., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., ... & Taylor, J. (2010). Galaxy: a web‐based genome analysis tool for experimentalists. Current protocols in molecular biology, 19-10.

Bolger, A., & Giorgi, F. Trimmomatic: A Flexible Read Trimming Tool for Illumina NGS Data. URL http://www. usadellab. org/cms/index. php.

https://barbaraiot.com/blog/protocolos-iot-que-deberias-conocer/

Verificar que las referencias se ajusten a la normatividad APA y sean utilizadas dentro del documento

# Anexo I

Utilice el anexo para incluir datos, instrumentos de investigación y material adicional que aporte a la consecución de los objetivos y alcances del proyecto de grado. Se debe incluir el cronograma de actividades y presupuesto (si aplica).