**FORMATO PARA EL DESARROLLO DE COMPONENTE FORMATIVO**

|  |  |
| --- | --- |
| PROGRAMA DE FORMACIÓN | Desarrollo e implementación de soluciones para la transformación digital |

|  |  |  |  |
| --- | --- | --- | --- |
| COMPETENCIA | 220501114- Sistematizar datos masivos de acuerdo con métodos de analítica y herramientas tecnológicas. | RESULTADOS DE APRENDIZAJE | 220501114-03 Procesar datos de acuerdo con procedimiento técnico y metodología estadística.  220501114-04 Presentar los datos de acuerdo herramientas y técnicas de visualización. |

|  |  |
| --- | --- |
| NÚMERO DEL COMPONENTE FORMATIVO | 15 |
| NOMBRE DEL COMPONENTE FORMATIVO | Analítica de datos y visualización |
| BREVE DESCRIPCIÓN | El presente componente formativo aborda aspectos generales y claves sobre elementos, métodos y herramientas empleados para el desarrollo de reportes y tableros, a partir de los datos. Con su estudio responsable, el aprendiz se afianzará en fuentes de datos, transformación, *machine learning* y desarrollo de gráficos, usando datos nativos y cálculos con lenguajes de consulta. |
| PALABRAS CLAVE | Analítica, *dashboard*, *machine learning*, inteligencia de negocio, visualización de datos. |

|  |  |
| --- | --- |
| ÁREA OCUPACIONAL | VENTAS Y SERVICIOS |
| IDIOMA | Español |

1. **TABLA DE CONTENIDOS:**

**Introducción**

**1. Modelos y metodologías de analítica**

**2. Aprendizaje automático (*machine learning*)**

2.1 Aprendizaje supervisado

2.2 Aprendizaje no supervisado

2.3 Algoritmos y modelos matemáticos

2.4 Predictores

**3. Herramientas de analítica de datos y visualización**

**4. Gestión de datos masivos**

4.1 Lenguajes de consulta

4.2 Manipulación de datos

**5. Tableros estadísticos, informes y resultados de visualización**

1. **INTRODUCCIÓN**

Se da la bienvenida al estudio del componente formativo “**Analítica de datos y visualización**”. Para comenzar satisfactoriamente esta experiencia de aprendizaje, se le invita a explorar el video que se muestra a continuación. ¡**Adelante**!

|  |
| --- |
| **DI\_CF15\_0\_Video\_Introduccion** |

1. **DESARROLLO DE CONTENIDOS:**

## AI ethics or AI Law concept. Developing AI codes of ethics. Compliance, regulation, standard , business policy and responsibility for guarding against unintended bias in machine learning algorithms.1. Modelos y metodologías de analítica

A partir del contexto de la inteligencia de negocio, se debe comenzar por entender bien los requerimientos específicos de las soluciones analíticas; si bien toma muchos elementos del desarrollo del *software* e incluso, se toma como una rama del desarrollo de aplicaciones informáticas, hay ciertas especificidades en los requisitos para la toma de decisiones que se deben asumir.

Los desarrollos de *software* se limitan a cumplir los requisitos de un área o un proceso determinado, como una isla en la empresa, donde no se integra información, ni formatos. Hoy en día, aún las empresas usan sistemas para cada área, por ejemplo, contabilidad, recursos humanos, área financiera, ventas, ofertas, *marketing*, producción, etc.

En los modelos y metodologías de analítica, se debe tener presente:

|  |
| --- |
| **DI\_CF15\_1\_Slide\_ModelosYMetodologiasDeAnalitica** |

Una de las condiciones iniciales más importantes son las preguntas, es decir, qué se necesita saber del negocio, qué decisiones se planean tomar y qué insumos se requieren para la toma de estas decisiones.

En sí, se deben seleccionar varios elementos para el desarrollo de soluciones analíticas, tales como:

|  |
| --- |
| **DI\_CF15\_1\_Tajertas\_ElementosParaDesarrolloDeSoluciones** |

**2. Aprendizaje automático (*machine learning*)**

En el diario vivir, en un hogar promedio actual, podría fácilmente suceder que el chico de la escuela, quien hace las tareas, pregunta en voz alta, ¿cuántos departamentos tiene Colombia?, a lo que el padre, quien tradicionalmente debería saberlo, tiene en su memoria que son 33 departamentos, pero mientras está en la mitad de su proceso cognitivo, asegurándose para responder, se oye primero el dispositivo electrónico Alexa, dando la respuesta de 32 departamentos.

Adicionalmente, amplía esta información con algunos datos complementarios. En este instante, la inteligencia artificial, es tomada como fuente de información rápida y confiable. Es así como los padres y profesores, que han sido tradicionalmente las fuentes de conocimiento, van siendo desplazados en este sentido (para adquirir conocimientos), y empiezan otros retos y roles igual de importantes que deben asumir.

En relación con el aprendizaje automático, es importante tener en cuenta aspectos como:

|  |
| --- |
| **DI\_CF15\_2\_Slide\_AprendizajeAutomatico** |

### 2.1 Aprendizaje supervisado

*Machine Learning* (ML) es el área de las ciencias computacionales que hace parte de la IA, su enfoque es que las computadoras, en vez de ser programadas paso a paso, aprendan a partir de los datos. Cada solución de ML es específica para cada necesidad, tal y como se tiene el enfoque con la programación convencional. Los profesionales de ML están dedicados al desarrollo de algoritmos genéricos que pueden extraer patrones de diferentes tipos de datos.



El ML enfocado a la ciencia de datos apunta a desarrollar procesos específicos como la identificación de la fuente de datos, desechar información inválida o no útil, limpiar, normalizar, relacionar, datos sesgados, etc.

**Nota importante**. Todas estas tareas podrían encontrar solución eficiente en la selección de soluciones de *machine learning,* cuya aplicación resulte apropiada, la elección del algoritmo más adecuado, el ajuste de los parámetros del método elegido, el análisis de los resultados, la identificación de comportamientos incorrectos, la vuelta a procesos anteriores con el fin de cambiar lo que resulte necesario para mejorar los resultados. (Bobadilla, 2020)

Si bien la aplicación de *Business Intelligence* (BI) es una manera interesante para el conocimiento de las organizaciones y otras bondades, la implementación de ML es ir más allá. Preste atención a los aspectos clave que se muestran a continuación:

|  |
| --- |
| **DI\_CF15\_2-1\_Acordeon\_BusinessIntelligenceYMachineLearning** |

### ChatGPT Chat concept AI, Artificial Intelligence. businessman using technology smart robot AI, enter command prompt, contact for business information analysis, Futuristic technology transformation.2.2 Aprendizaje no supervisado

Este paradigma de aprendizaje toma como base, únicamente, los datos de entrada, sin explicarle al sistema qué resultado es el que se espera obtener. Podría ser un poco difícil de concebir, porque si no hay una referencia previa, ¿de qué manera podrían los sistemas aprender?

Este concepto es menos usado, pues sostiene mayores retos a la ciencia y a los algoritmos, donde a partir de un parámetro, el sistema deberá tratar de descubrir qué resultado o resultados posibles daría ese dato de entrada.

Estas son algunas generalidades que se deben tener en cuenta, respecto del aprendizaje no supervisado:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tiene una ventaja, porque el entrenamiento en aprendizaje supervisado implicaría miles de horas, humanos enseñando y altos costos, debido a que para que un sistema esté bien entrenado requiere al menos 100 mil ejemplos, esto es una tarea larga y costosa. | Por su parte, el aprendizaje no supervisado solo requiere de los datos de entrada, dar unos pocos parámetros de lo que se quiere y dejar todo a la máquina. | La desventaja es que requiere mucho procesamiento, puntos generales que pueda asociar a lo que se le parezca, y avance tecnológico con el fin que la máquina vislumbre y descubra el dato de salida. | Si bien, las técnicas ML más usadas están basadas en referencias de salidas, el aprendizaje no supervisado será el futuro, porque de cierta manera los sistemas usarán estas referencias y gran capacidad computacional para empezar a asociar. | Lograr que una máquina tenga sentido común es un objetivo muy difícil, sin embargo, los algoritmos actuales se van acercando un poco, por ejemplo, hay palabras que si bien es la misma para dos cosas o significados diferentes (palabras polisémicas), para los humanos es fácil según el contexto, pero para una máquina es difícil definir estas cosas. |

### 2.3 Algoritmos y modelos matemáticos



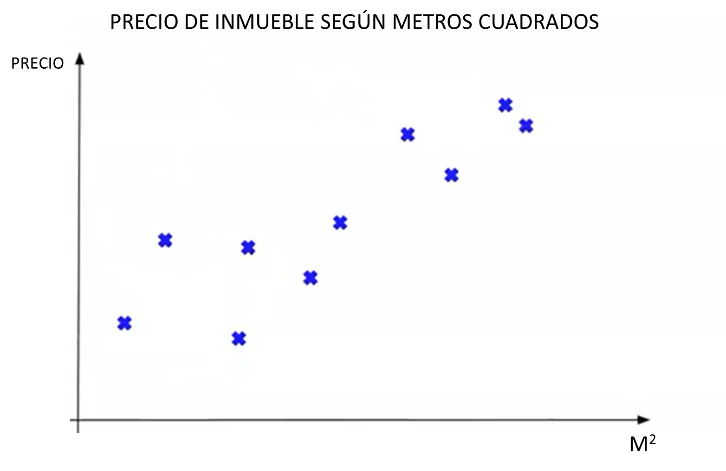
Los métodos matemáticos y algorítmicos empleados en la Inteligencia Artificial, pueden variar según los desarrollares. Aunque existen otros, los más comunes o típicos para **aprendizaje supervisado** son las Regresiones lineales y logísticas, Máquinas de vectores de soporte, árbol de decisiones y K-Media.

**Regresiones lineales y logísticas**

Estos algoritmos tienen un comportamiento basado en el historial de los datos, por ejemplo, se requiere predecir o asignar correctamente el valor de un inmueble; teniendo como base datos históricos, se tendría una gráfica como esta:

**Figura 1**

Datos base para los algoritmos regresiones lineales

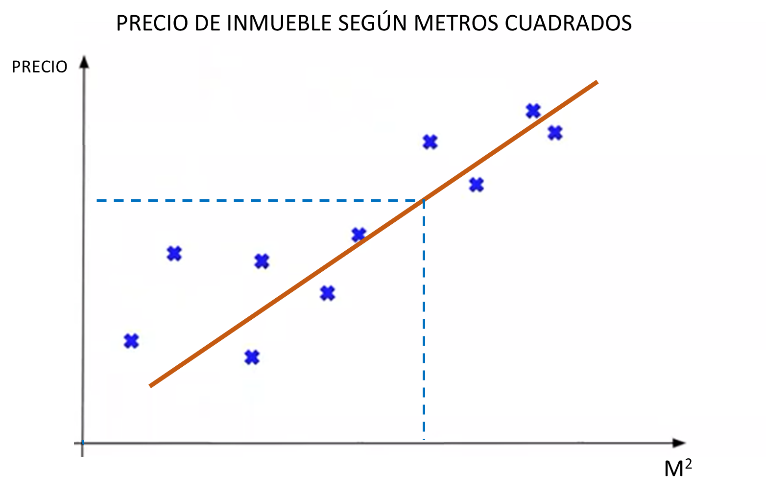


Luego, a partir de una la regresión lineal, se calcula la línea que abarca los valores y se establece con mejor exactitud el valor del inmueble.

De esa manera, cuando una propiedad tiene ciertos metros cuadrados específicos, la salida que ofrece la máquina será un precio que coincida con la regresión lineal, como se expresa en la siguiente gráfica:

**Figura 2**

Línea a partir de valores conocidos



Con estas técnicas se pueden establecer grupos, por ejemplo, es posible definir si un inmueble es costoso o no lo es, con base en los datos y sus grupos.

**K-MEANS cluestering o K-media de agrupamiento**

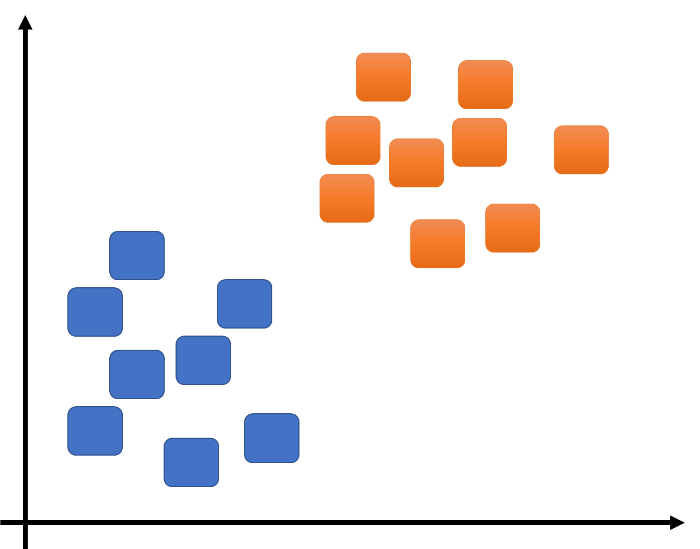
En el caso anterior, se tenían valores históricos, sin embargo, ¿qué tal si se presenta el caso de que los datos no están categorizados y no hay un historial?

Se deberá usar uno de los algoritmos o métodos **no supervisados** para que la máquina aprenda a identificar patrones y arroje respuestas.

El algoritmo por agrupación usa como base **centroides** o **puntos de datos base,** que procura detectar patrones similares y de esa manera identifica grupos, como se presenta a continuación.

**Figura 3**

Agrupando (clustering)

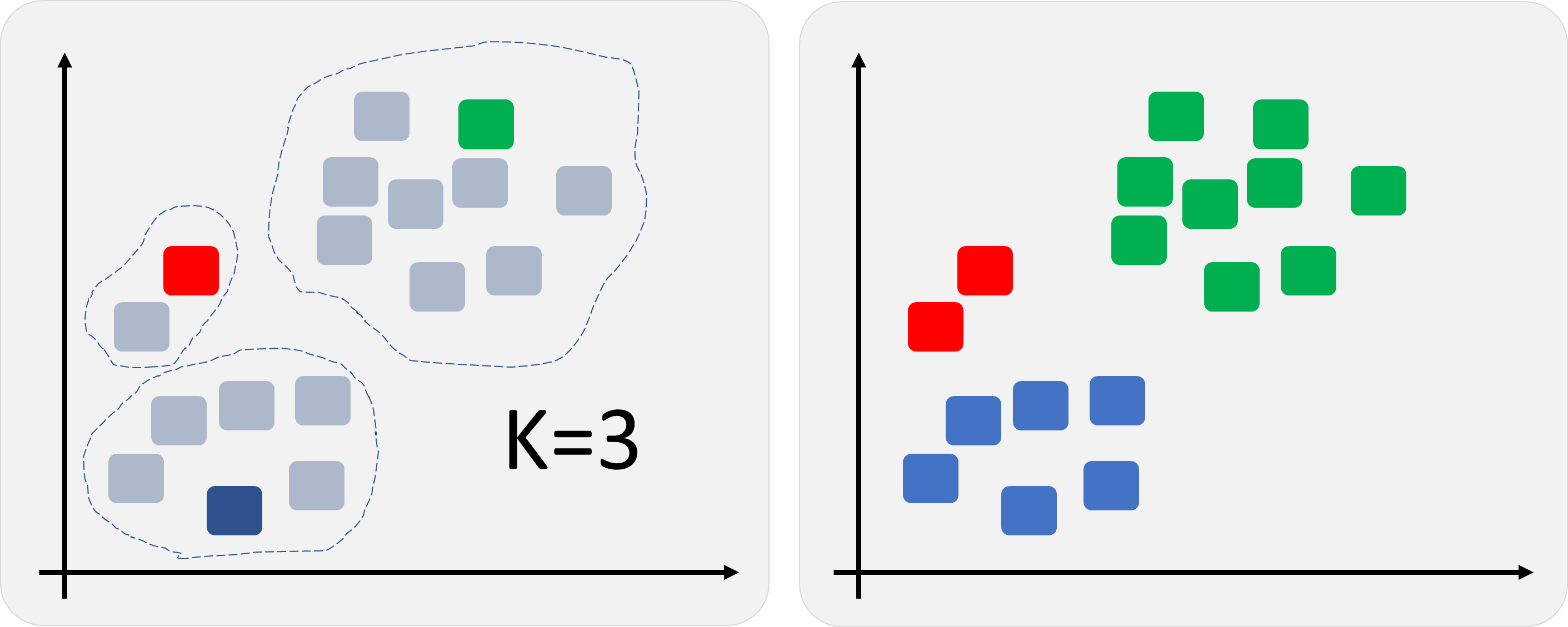


En un primer momento, los datos no están clasificados, sin embargo, el algoritmo determina similitud según la distancia entre los mismos (en estos datos se nota que hay dos grupos por su cercanía).

Para empezar a usar este tipo de algoritmos, se debe elegir el número de *clusters*, representado por K, Luego, aleatoriamente, se asignan centroides y se calcula, uno a uno, la cercanía de los datos al centroide, como se presenta en la siguiente figura.

**Figura 4**

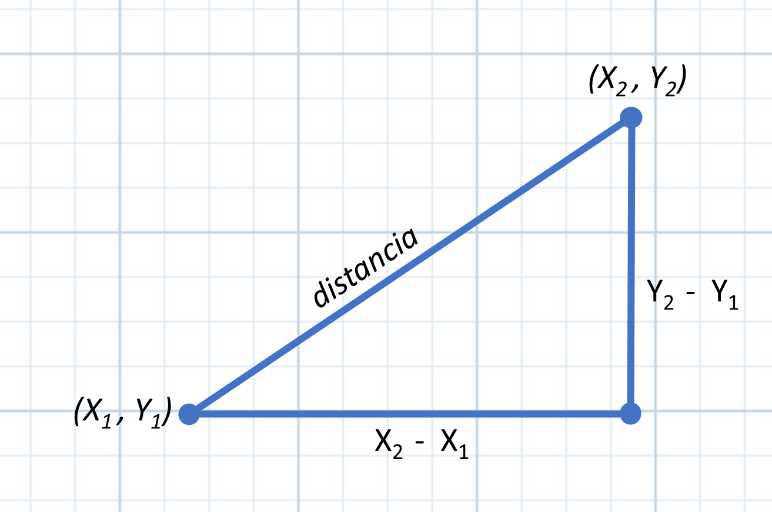
Clusters y centroides



El algoritmo calcula estas distancias usando la **Distancia Euclidiana**. Se trata de una variación del teorema de Pitágoras y es una de las maneras más fáciles para calcular distancia, a partir de las posiciones de dos objetos en un plano cartesiano, como se presenta a continuación.

**Figura 5**

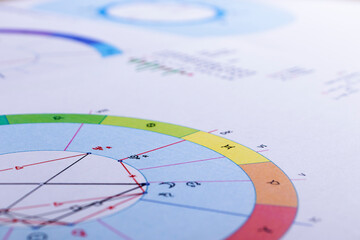
Distancia euclidiana



Esta función se repite las veces que sea necesario o se hayan parametrizado, para que el algoritmo autónomamente determine qué dato estaría en cuál clúster o qué clasificación.

Esto es solo un ejemplo de métodos o algoritmos que emplea el *machine learning*, pero hay muchos más tanto para aprendizaje supervisado como no supervisado.

### 2.4. Predictores



Adivinar el futuro ha sido uno de los factores más fascinantes que, a lo largo de toda la historia humana, ha ocupado a las civilizaciones y líderes, hasta el punto de desarrollar ansiedades por tal incertidumbre. Es por ello que es tan frecuente que el futuro sea un insumo para historias de cine, y atrapen y fascinen a todo el público, aparecen objetos como oráculos, bolas de cristal y profecías alrededor de múltiples *films* e historias.

Si bien la predicción, culturalmente, se contempla como saber el futuro, no necesariamente es así, la predicción enfatiza la capacidad de ver información oculta, ya sea en el pasado, el presente o el futuro, tal como se concibe la bola de cristal de algunas historias mágicas, donde este objeto no solo permite ver el futuro, sino también ver qué sucede en el presente en otros lugares; esta utilidad ya dejó de ser fascinante porque se convirtió en algo normal después del avance de los medios de comunicación, pero igual se sigue llamando predicción.

**Predictores como funciones de herramientas**



Un ejemplo de predictores usados por todos, ha sido la función autocompletar de Google en sus búsquedas y los teclados de celular donde, al escribir una palabra o parte de ella, el sistema procura autocompletar lo que se desea escribir, ahorrando tiempo y disminuyendo errores de digitación de manera ostensible. Esto se hace a partir de unas métricas y del aprendizaje que la máquina detectó por las tendencias propias del usuario o tendencias de búsqueda masiva localizada o mundial.

En las aplicaciones, también la predicción avanza según esta va aprendiendo; es el caso del traductor de Google: cuando esta herramienta era nueva, su traducción no tenía buena calidad, pero, actualmente, con el aporte de los mismos usuarios, las traducciones tienen en cuenta incluso ciertos contextos que la máquina ha aprendido a diferenciar.

En relación con los predictores, tenga en cuenta algunas generalidades como las que enuncian a continuación:

|  |
| --- |
| **DI\_CF15\_2-4\_Pasos\_Predictores** |

## 3. Herramientas de analítica de datos y visualización



Por practicidad y funcionalidad, las organizaciones deciden usar un ecosistema integrado que abarque todas las funcionalidades y extras de la inteligencia de negocio, desde el contexto de gestión de datos hasta el modelamiento y visualización de reportes. En este caso, la orientación estará enfocada a las herramientas propias para la visualización de datos.

Existen muchas opciones para las soluciones BI, las más populares son Tableau de la marca Salesforce, Power BI de Microsoft y otras como Qlik Sense de la empresa Qlik.

Sobre las herramientas de analítica de datos y visualización, tenga en cuenta los siguientes ítems:

|  |
| --- |
| **DI\_CF15\_3\_Slide\_HerramientasAnaliticaDatosYVisualizacion** |

## 4. Gestión de datos masivos



La gestión del flujo de datos de extremo a extremo (*end to end*), es un término usado por algunas marcas y consiste en establecer la tubería por donde fluyen los datos; de esta manera, existe una toma de datos (extracción de datos origen), sigue su flujo haciendo filtraciones, depuraciones, limpieza y mejora en la calidad de los datos, para que lleguen a un gran estanque (bodega de datos), y luego puedan ser consumidos según los requerimientos del negocio.

### Herramientas de desarrollo

Las soluciones *Open Source* es una sugerencia, al menos, para el proceso de aprendizaje, porque muchas organizaciones se alinean bajo ecosistemas de pago tales como Microsoft con **SQL Server Integration Services** (SSIS), **Qlik analityc** u otras herramientas comerciales.

Las herramientas *Open Source* cumplen con el principio de ser abiertas, es decir, que pertenecen o se matriculan a una comunidad que tiene acceso libre para su uso y participación del código fuente. Si bien son escasos los soportes técnicos, existen foros donde la comunidad misma ofrece su ayuda en temas o situaciones específicas.

¡**Importante**!

Muchas empresas usan este tipo de tendencias en las herramientas informáticas no por el ahorro del costo, sino también por su funcionalidad e impacto en el funcionamiento y utilidad.

**Pentaho**

Nacida en el año 2006, es una plataforma BI y se trata de una multiplataforma. Al ser un proyecto *Open Source*, hay una gran comunidad que brinda apoyo y ayuda con los errores que se puedan detectar en los proyectos desarrollados.

Es una plataforma integral con una suite de herramientas para completar las tareas y flujo de datos por toda la tubería de datos, transformación y carga.

Los principales componentes de Pentaho son:

|  |  |  |  |
| --- | --- | --- | --- |
| **Pentaho server**  Núcleo de la plataforma. | **Pentaho report designer**  Para desarrollar reportes a través de consultas de diversos orígenes de datos. | **Pentaho schema workbench**  Para administrar y crear cubos OLAP. Usa motor de datos llamado Mondrian, también de código abierto. | **Pentaho data integration kettle**  Es una de las herramientas más usadas de Pentaho, permite crear procesos ETL diseñando y alimentando los *Data warehouse*. |

Pentaho puede usarse como suite completa, es decir, usando todo el ecosistema de extremo a extremo o como componente individual de toda la solución BI, integrándose con otras herramientas.

**Spoon (Pentaho Data Integration - Kettle)**

Para poder realizar los procesos ETL, es necesario un entorno gráfico, eso es Spoon, una interfaz para realizar todos los procesos y tareas de extracción, transformación y carga de datos hacia las bodegas de datos.

### Application of Smartphone with business graph and analytics data on isometric mobile phone. Analysis trends and software development coding process concept. Programming, testing cross platform code4.1 Lenguajes de consulta

Los lenguajes de consulta son lenguajes de programación orientados al manejo y gestión de datos; cuando se desarrollan aplicaciones bajo lenguajes de programación como C++, Java, y otros, los lenguajes de consulta son incluidos en algunas funciones de las aplicaciones, sin embargo, dejar todo el trabajo de reportes y consultas en los desarrollos de *software,* podría resultar limitado y con problemas de desempeño, cuando se procesan grandes cantidades de datos.

A continuación, se describen los tres principales lenguajes de consulta que todo profesional de datos deberá tener en cuenta para desarrollar proyectos analíticos y transformación digital:

|  |
| --- |
| **DI\_CF15\_4-1\_LineaDeTiempo\_LenguajesDeConsulta** |

La mayoría de estos lenguajes a diferencia de los lenguajes para el desarrollo de aplicaciones, es que son **lenguajes interpretados,** es decir, ejecutan las instrucciones directamente sin una previa compilación de las líneas de código.

### 4.2. Manipulación de datos



En sí, la manipulación de datos es un elemento que permite varios elementos. La gestión de extremo a extremo comienza con los datos de origen de las áreas o procesos del negocio. El proceso realiza ingesta de datos de las fuentes, esto da como salida datos integrados en una herramienta específica para, luego, a través de la administración de los datos, se ordenen y se sincronicen.

En esta parte se preparan y optimizan los datos para aplicarles técnicas de analítica, visualizando datos y generando modelos de inteligencia artificial para predecir comportamientos y, en casos avanzados, sugerir acciones para que, en la siguiente etapa, ya con ***insigth*** y datos claros, se generen alertas programadas y en algunos casos, que el sistema tome algunas decisiones automatizadas (como subir o bajar precios a productos, publicar o no productos, entre otras acciones que pueden automatizarse). Al final, se toman acciones ya sean humanas o por los sistemas de información, de manera inteligente.

## 5. Tableros estadísticos, informes y resultados de visualización



Los tableros estadísticos, también llamados cuadros de mandos o *dashboard*, permiten mostrar información consolidad a alto nivel. Los *dashboard* son las maneras más populares en BI de presentar datos, debido a que permite la compresión de manera sencilla los hechos del negocio y las situaciones presentadas.

Los tableros estadísticos se centran en diferentes aspectos, como:

|  |  |  |  |
| --- | --- | --- | --- |
| **Emplear recursos gráficos**  Emplear recursos gráficos como elemento principal para representar datos y cifras. | **Presentar aspectos del negocio**  Presentar aspectos del negocio de manera general, pero que con la interactividad pueda llegar al detalle hasta que los datos lo permitan. | **Crear diseños amigables**  Por lo general todas las herramientas permiten crear diseños muy atractivos visualmente y amigables para la navegación e interacción con los datos. | **Generar actualizaciones**  Por lo generalestos recursos permiten actualizaciones de manera fácil incluyendo o quitando datos o elementos visuales según las necesidades y sugerencias de los clientes o líderes. |

**Tableros estadísticos como herramienta para monitorizar el negocio**

Un *dashboard*, permite visualizar de manera ágil y actualizada los procesos de negocio, pues muestra información clave de fácil entendimiento, los *dashboard* modernos tienen un “*refresh* de datos” casi en tiempo real, lo que les permite a las personas del nivel de decisión o coordinación tener el negocio y sus procesos a la mano con el detalle que requiera, permitiendo tomar con más velocidad las decisiones diarias.

Los tableros de mando contienen los siguientes elementos:

* Permiten empelar variedad de elementos (gráficos, tablas, alertas, mapas, etiquetas, etc.).
* La interacción se aplica de manera integral a todos los elementos del tablero.
* La información que se presenta se basa en indicadores claves del negocio, dando idea clara de cómo van las áreas o aspectos del negocio o proceso.
* Exhibe las tendencias del negocio para tomar decisiones inteligentes, con base en datos y cifras.
* Con la analítica predictiva, los tableros mostrarán hechos futuros con cierta precisión que ayudará también a la toma de decisiones más acertadas.

**Usuarios de los tableros de mando**

El perfil de los usuarios que usan estos *dashboard* son, por lo general:

* Alta dirección, con el objetivo de comprender lo que sucede en el negocio.
* Gerentes, que deben monitorizar procesos de negocio.
* Usuarios de negocio: estos necesitan realizar análisis exploratorio de datos.

En este sentido, los tableros o cuadros de mando aportan valor al nivel estratégico, táctico y operativo.

## Informes y resultados de visualización



Los *dashboard* no son la única manera de ver resultados del negocio o visualizar los datos. Cada área del negocio podría generar sus propios reportes a partir del desarrollo de BI, incluso desde la propia *data warehouse* se puede suministrar a usuarios, datos filtrados según sus interesas particulares del negocio, para que ellos mismos puedan elaborar sus reportes con los datos del despliegue analítico, garantizando que las cifras, independiente del área coincidan entre ellas.

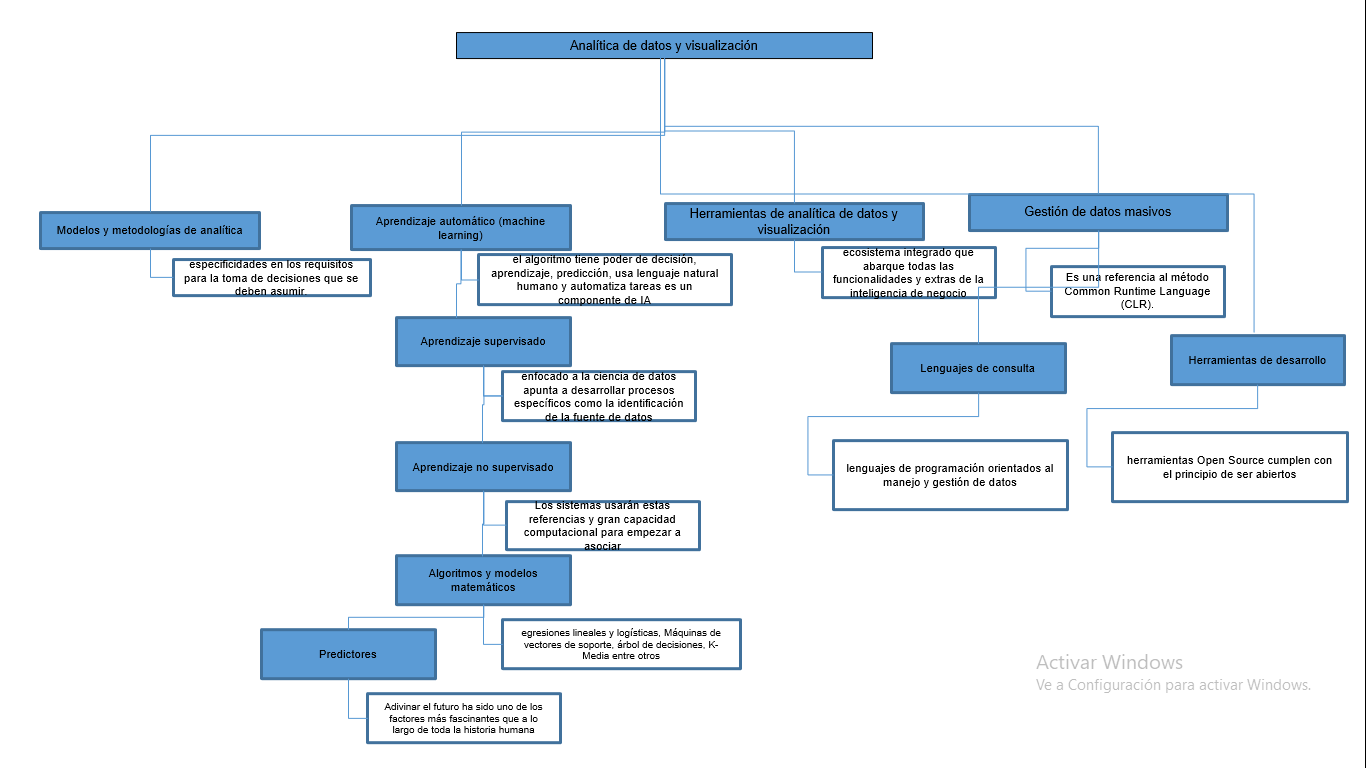
Lo anterior, dando respuesta a que las organizaciones y los usuarios, cada vez más, tienen más habilidades digitales y dominio de herramientas analíticas así no sean usuarios informáticos propiamente dicho.

En relación con los informes y resultados de visualización, tenga presente:

|  |
| --- |
| **DI\_CF15\_5\_Pestañas\_InformesYResultadosVisualizacion** |

1. **SÍNTESIS**

Aquí finaliza el estudio de los temas de este componente formativo. En este punto, explore con atención el esquema que se muestra enseguida y realice su propia síntesis de los conceptos estudiados. ¡**Adelante**!



1. **ACTIVIDADES DIDÁCTICAS (Se debe incorporar mínimo 1, máximo 2)**

|  |  |
| --- | --- |
| DESCRIPCIÓN DE ACTIVIDAD DIDÁCTICA | |
| Nombre de la Actividad | Analizando y visualizando datos |
| Objetivo de la actividad | Fortalecer los conceptos de la analítica y visualización de datos, y el desarrollo de soluciones visuales de inteligencia de negocio. |
| Tipo de actividad sugerida | COMPLETAR ESPACIOS |
| Archivo de la actividad  (Anexo donde se describe la actividad propuesta) | Carpeta Anexos:  Actividad\_Didactica\_1 |

1. **MATERIAL COMPLEMENTARIO:**

|  |  |  |  |
| --- | --- | --- | --- |
| Tema | Referencia APA del Material | Tipo de material  (Video, capítulo de libro, artículo, otro) | Enlace del Recurso o  Archivo del documento o material |
| Algoritmos y modelos matemáticos | Data Silvestre. (2021). *Distancia Euclidiana en Python | Métricas y Distancias #1* [video]. YouTube. <https://youtu.be/vklKtK5oCfg> | Video | <https://youtu.be/vklKtK5oCfg> |
| Gestión de datos masivos | The QA Testing Channel. (2017). *Comandos SQL Básicos en Base de Datos - En 1 minuto* [video]. YouTube. <https://www.youtube.com/watch?v=YuqqGaPLrjY&ab_channel=TheQATestingChannel> | Video | <https://www.youtube.com/watch?v=YuqqGaPLrjY&ab_channel=TheQATestingChannel> |
| Gestión de datos masivos | Learning BI. (2017). *Introducción Pentaho Data Integration* [video]. YouTube. <https://youtu.be/o7If1a-gkyI> | Video | <https://youtu.be/o7If1a-gkyI> |
| Gestión de datos masivos | Contreras, P. (2018). *Business intelligence con Excel y Power BI.* [Conjunto de datos]. Sistema de bibliotecas SENA. <https://login.bdigital.sena.edu.co/login?qurl=https://www.ebooks7-24.com%2f%3fil%3d10579> | Libro | <https://www-ebooks7-24-com.bdigital.sena.edu.co/?il=10579> |

1. **GLOSARIO:**

|  |  |
| --- | --- |
| **TÉRMINO** | **SIGNIFICADO** |
| **Algoritmo:** | pasos programados para que las máquinas realicen una función o tarea. Los algoritmos pueden programarse o en caso de la IA se auto ajustan o calibran sin necesidad de intervención humana. |
| **Aprendizaje Profundo:** | *Deep Learning,* se refiere a los programas que emplean redes neuronales programadas para tener procesos de machine learning más avanzados y complejos. |
| **AWS:** | Amazon Web Service, *p*lataforma de computación o servicios en la nube, cuenta con múltiples servicios entre los cuales muchas herramientas tienen grandes componentes de inteligencia artificial y gestión de datos. |
| ***Chatbots:*** | chats operados por robots o chats inteligentes que interactúan con personas o clientes sin necesidad de intervención humana. |
| ***Datasets:*** | conjunto de datos guardados en un sistema, ya sea en una o varias bases de datos. Por lo general son datos estructurados y están disponibles para gestión y uso que se desee dar según los objetivos del negocio. |
| ***Insigth:*** | en términos de informática y marketing se refiere a las verdades (a veces relativa según tiempo y condiciones), que generan los datos o los comportamientos de consumo. |
| **Minería de datos:** | *Data mining,* es el uso de grandes volúmenes de datos para la obtención de situación, circunstancias o verdades a partir de patrones y características de los datos. Existen varias técnicas para la aplicación de minería de datos. |
| **Palabras polisémicas:** | palabras que tienen varios significados, el significado lo da el contexto en el que se da la comunicación. |

1. **REFERENCIAS BIBLIOGRÁFICAS:**

Agrawal, A., Joshua, G., & Avi, G. (2019). *Máquinas predictivas: la sencilla economía de la inteligencia artificial* [Conjunto de datos]. Sistema de bibliotecas SENA. <https://elibro-net.bdigital.sena.edu.co/es/lc/senavirtual/titulos/122310>

Bobadilla, J. (2020). *Machine Learning y Deep Learning.* [Conjunto de datos]. Sistema de bibliotecas SENA. <https://www-ebooks7-24-com.bdigital.sena.edu.co/?il=10277>

Curto Díaz, J. (2016). *Introducción al business intelligence.* [Conjunto de datos]. Sistema de bibliotecas SENA. <https://elibro-net.bdigital.sena.edu.co/es/lc/senavirtual/titulos/101030>

Dot CSV. (2019). *¿Qué es el Aprendizaje Supervisado y No Supervisado?* [video]. YouTube. <https://youtu.be/oT3arRRB2Cw>

1. **CONTROL DEL DOCUMENTO**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Nombre | Cargo | Dependencia | Fecha |
| Autor (es) | Jaime Hernán Tejada | Experto Temático | Centro de la Industria, la empresa y los servicios - CIES | Noviembre de 2022 |
| Javier Ricardo Luna Pineda | Diseñador Instruccional | Centro De La Industria, La Empresa Y Los Servicios | Noviembre de 2022 |
| Fabián Leonardo Correa Díaz | Diseñador Instruccional | Regional Santander – Centro Industrial del Diseño y la Manufactura | Junio de 2023 |
| Ana Catalina Córdoba Sus | Metodólogo para Formación Virtual | Regional Santander – Centro Industrial del Diseño y la Manufactura | Junio de 2023 |
| Rafael Neftalí Lizcano Reyes | Responsable Equipo de Desarrollo Curricular | Regional Santander – Centro Industrial del Diseño y la Manufactura | Junio de 2023 |

1. **CONTROL DE CAMBIOS**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Nombre | Cargo | Dependencia | Fecha | Razón del Cambio |
| Autor (es) |  |  |  |  |  |