**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**!

|  |
| --- |
| **Agregar video** |

1. **DESARROLLO DE CONTENIDOS:**

****

2. Aprendizaje automático (machine learning)

En la vida cotidiana de un hogar promedio actual, podría ocurrir fácilmente que un niño de escuela, mientras hace sus tareas, pregunte en voz alta: "¿Cuántos departamentos tiene Colombia?". El padre, quien tradicionalmente debería saber la respuesta, recuerda que son 33 departamentos. Sin embargo, mientras se asegura de su respuesta, el dispositivo electrónico Alexa interrumpe primero, proporcionando la respuesta correcta de 32 departamentos.

**Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente**



**Gráfico, Gráfico de dispersión

Descripción generada automáticamente**

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:

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

**Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

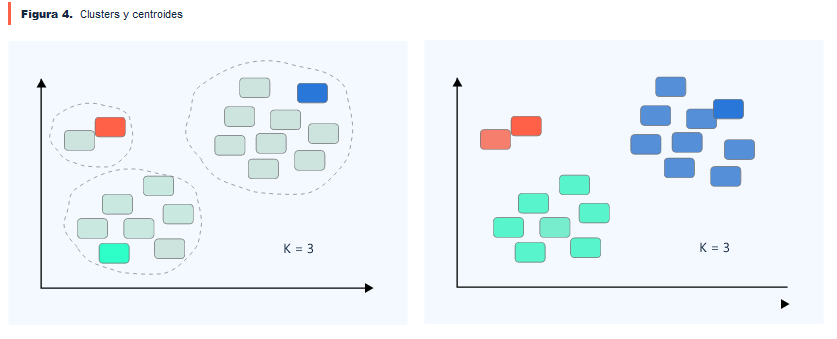
Descripción generada automáticamente**

**Gráfico

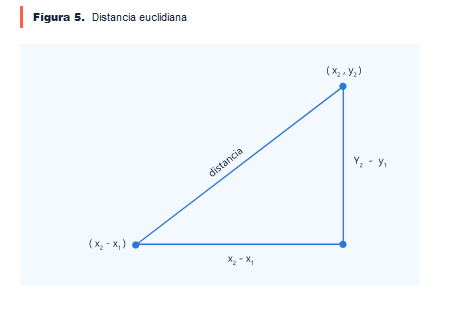
Descripción generada automáticamente**

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.

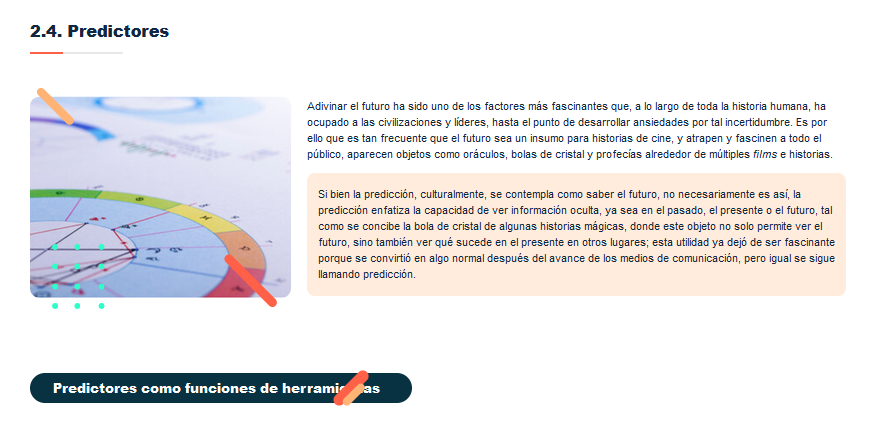
****

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.

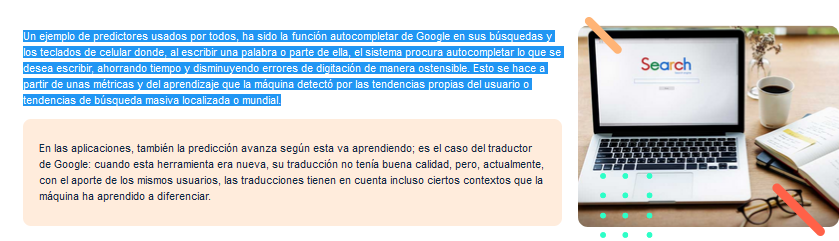
****

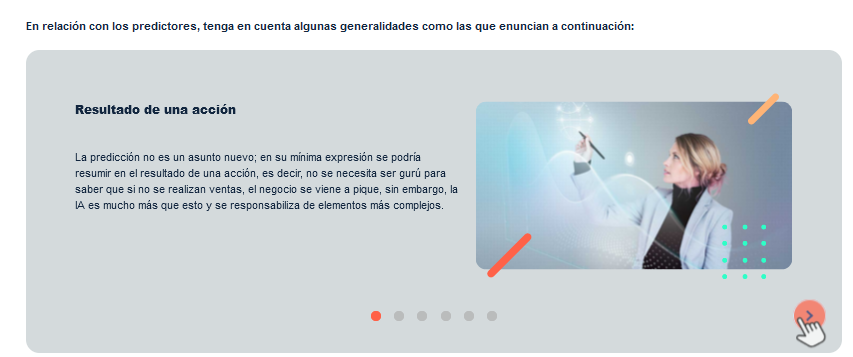
Esta función se repite las veces que sea necesario o se hayan parametrizado, para que el algoritmo autónomamente determine qué dato estaría en cuál cluster o qué clasificación.

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



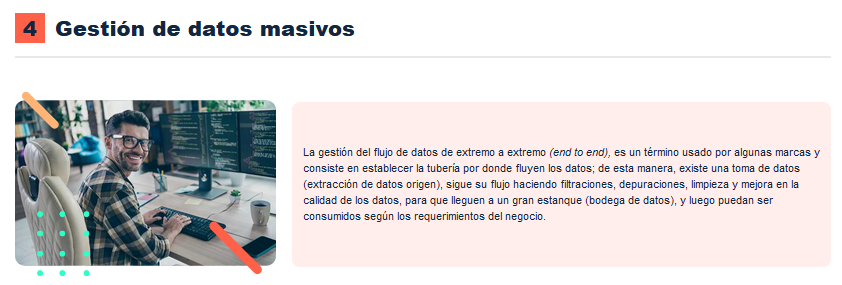
Un ejemplo de predictores usados por todos, ha sido la función autocompletar de Google en sus búsquedas y los teclados de celular donde, al escribir una palabra o parte de ella, el sistema procura autocompletar lo que se desea escribir, ahorrando tiempo y disminuyendo errores de digitación de manera 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.











Interfaz de usuario gráfica

Descripción generada automáticamente con confianza media

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.



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.

Texto

Descripción generada automáticamente con confianza media

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.



Tabla

Descripción generada automáticamente con confianza media

A diferencia de los lenguajes utilizados para el desarrollo de aplicaciones, la mayoría de estos lenguajes son interpretados. Esto significa que ejecutan las instrucciones directamente, sin necesidad de compilar previamente las líneas de código.





Interfaz de usuario gráfica, Texto

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

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:**



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**!

Diagrama

Descripción generada automáticamente

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]. | Video | <https://youtu.be/vklKtK5oCfg> |
| Gestión de datos masivos | The QA Testing Channel. (2017). *Comandos SQL Básicos en Base de Datos -* | Video | <https://youtu.be/S93ln0XFPuY> |
| Gestión de datos masivos | Learning BI. (2017). *Introducción Pentaho Data Integration* [video]. | Video | <https://youtu.be/o7If1a-gkyI> |
| Gestión de datos masivos | *Be Intellligence.*  *Business intelligence* | Video | <https://youtu.be/Pmdps2kK_5M> |

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:**

**Browner. M. (2020).** *Máquinas predictivas: la sencilla economía de la inteligencia artificial.* [*https://dokumen.pub/maquinas-predictivas-prediction-machines-spanish-edition-la-sencilla-economia-de-la-inteligencia-artificial-1nbsped-8494949381-9788494949388.html*](https://dokumen.pub/maquinas-predictivas-prediction-machines-spanish-edition-la-sencilla-economia-de-la-inteligencia-artificial-1nbsped-8494949381-9788494949388.html)

**Zambelli. R** (2024). ¿Qué es el Machine Learning y cómo usarlo en la gestión industrial?*.* <https://blog-es.checklistfacil.com/machine-learning/?utm_term=&utm_campaign=LATAM-ES-PAID-CF-GOOGLE-SEARCH-LM_NOVOS_LEADS-DSA-BLOG&utm_source=google&utm_medium=cpc&hsa_acc=6707140990&hsa_cam=21096577828&hsa_grp=160250856136&hsa_ad=693261475572&hsa_src=g&hsa_tgt=dsa-2284541207217&hsa_kw=&hsa_mt=&hsa_net=adwords&hsa_ver=3&gad_source=1&gclid=CjwKCAjwqMO0BhA8EiwAFTLgIAqb8LevzY8hmeCkm9H9GcC_TwoE49Fnr6dqoYkrvwRMnj4720jy7BoChYAQAvD_BwE>

Curto Díaz, J. (2016). *Introducción al business intelligence.* <https://cursos.yura.website/wp-content/uploads/2020/03/Introduccion_al_Business_Intelligence.pdf>

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) | Sandra Paola Morales Paez | Evaluador Instruccional | Regional  Santander  Centro Agroturístico | 5 de junio de 2024 | Adecuaciones a 2024 |
| Claudia Johanna Gómez Pérez | Responsable Línea Santander |