**FORMATO PARA EL DESARROLLO DE COMPONENTE FORMATIVO**

| PROGRAMA DE FORMACIÓN | Servicios Comerciales y Financieros |
| --- | --- |

| COMPETENCIA | 220501114 - Sistematizar datos masivos de acuerdo con métodos de analítica y herramientas  tecnológicas | RESULTADOS DE APRENDIZAJE | 220501114-2 Construir recursos visuales de acuerdo a información requerida. |
| --- | --- | --- | --- |

| NÚMERO DEL COMPONENTE FORMATIVO | CF007 |
| --- | --- |
| NOMBRE DEL COMPONENTE FORMATIVO | Modelamiento y relacionamiento de los datos |
| BREVE DESCRIPCIÓN | En la presente unidad temática el aprendiz descubrirá la importancia de la gestión de los datos para los servicios comerciales y financieros con el propósito de mitigar los diferentes riesgos asociados a su gestión como la pérdida de clientes, el lavado de activos, transacciones ilícitas, fraudes y fallas operacionales en la entidad por medio del análisis y el uso de técnicas y herramientas para ello. |
| PALABRAS CLAVE | Datos, análisis, almacenamiento de datos, Big Data, información, toma de decisiones, gestión, software. |

| ÁREA OCUPACIONAL | 1 - Finanzas Y Administración |
| --- | --- |
| IDIOMA | Español |

1. **TABLA DE CONTENIDOS:**

**Introducción**

**1. Ciencia de los datos**

1.1 Ciclo de vida.

1.2 Técnicas de análisis de datos.

1.3 Modelos y metodologías de analítica.

1.4 Minería de datos.

1.5 Métodos de recopilación e integración de datos.

1.6 Preparación de datos estructurados.

1.7 Ecosistema de datos abiertos.

**2. Bases de datos**

2.1 Clasificación

2.2 Características

2.3 Modelos

2.4 Sistemas de Gestión de bases de datos

**3. Big Data**

3.1 Características

3.2 Tipos de datos

3.3 Principios del análisis masivo de datos

3.4 Tecnologías para analizar grandes volúmenes de datos.

**4. Técnicas de gestión de datos masivos**

4.1 Metodologías

4.2 Almacenamiento

4.3 Lenguajes de consulta y manipulación de datos

1. **Introducción**

Es común hablar de estudios exitosos que han aportado a la innovación y creación de diferentes productos o servicios de acuerdo a las necesidades de los clientes o inclusive hablar de la mejora de los procesos dentro de las organizaciones gracias a los mismos.

Pero de lo que no es común hablar es los métodos que utilizan para llegar a estos resultados y cómo generaron esa información útil a la humanidad o empresas ya que de los datos que se recopilaron en cada estudio o investigación, se generó un análisis que dieron fórmulas o estrategias de solución a diferentes fenómenos presentados.

Conocer las bases de datos en una organización es importante ya que de esa información que recopilan las empresas se pueden dar soluciones a la eficacia y eficiencia de las actividades productivas, talento humano, atención al cliente, mercadeo y ventas entre otras.

En el caso del área de servicios comerciales y financieros la recolección de datos y su adecuado análisis puede mitigar fraudes, lavado de activos, financiación del terrorismo y cualquier tipo de actividad ilícita o de riesgo para sus operaciones, cabe también resaltar que ayuda a conocer nuevos clientes y disminuir la probabilidad de perder los actuales.

Teniendo en cuenta lo anterior, en la presente unidad temática se conocerá sobre la ciencia de datos, técnicas, gestión masiva de bases y Big data, con el fin de mejorar la capacidad de gestión en la clasificación y análisis de la información que le ayuden en su proceso comercial y financiero dentro de su entidad.

¡Bienvenidos!

A continuación se les invita a ver el siguiente Video:



1. **DESARROLLO DE CONTENIDOS:**

**1. Ciencia de los datos**

Para iniciar este tema es importante contextualizar sobre el significado de estas dos palabras:



Una vez teniendo claros estos dos conceptos, es importante hacerse este cuestionamiento ¿Qué es ciencia de datos?



* 1. **Ciclo de vida**

Los datos tienen una importancia en la gestión diaria de las entidades, ya que de esa información que se recoge se genera un análisis para planes de acción o estrategias según la razón por la cual se recolecta.

Como todas las actividades de las organizaciones, la Ciencia de datos también tiene unas etapas o ciclos, que se presentan a continuación:





**Nota:** cabe resaltar que la información obtenida queda en una base de datos creada por la entidad por si en algún momento se requiere consultar.

* 1. **Técnicas de análisis de datos**

Hace unas décadas el hablar de obtener información, requería de grandes actividades para conseguirla y su forma de adquisición era escasa, ahora gracias a las diferentes herramientas digitales y canales, la obtención de la misma es abundante. En la actualidad el reto es administrar bien los datos e interpretar de forma correcta su análisis.

Teniendo en cuenta lo anterior, las entidades aplican una serie de métodos de recolección de datos desde las más básicas hasta las más modernas, los medios utilizados son la entrevista, la observación, encuesta, análisis documental y demás instrumentos que son manejados para su análisis por medio de las siguientes técnicas:



Las técnicas anteriores se hacen de forma manual y por tal motivo pese a su gran aporte pueden tener ciertas desviaciones ya que interviene el factor humano, por otro lado, se presentan otras más avanzadas apoyadas por la inteligencia artificial, como se verán a continuación:



* 1. **Modelos y metodologías de analítica**

Para analizar la información obtenida por la entidad, las organizaciones emplean un conjunto de modelos y metodologías cuantitativas y cualitativas, como se observó en los puntos anteriores y para ampliar el conocimiento de modelos y metodologías de analítica, se invita a ver el siguiente video:



* 1. **Minería de datos**

Este proceso consiste en tomar o extraer información requerida por las entidades de diferentes bases de datos o fuentes, su nombre está asociado a la minería ya que la extracción es vital en ambas actividades; por otro lado, su propósito es extraer datos útiles que ayuden a la resolución de las diferentes situaciones que se originan en las entidades y su gestión se puede hacer a través de herramientas estadísticas tanto básicas como avanzadas que proporciona la inteligencia artificial.



La minería de datos como disciplina estudia las herramientas y algoritmos que le permitan acceder y recolectar datos a las organizaciones con el fin que esa información sea predictiva para que se facilite su análisis para la toma de decisiones ante situaciones presentadas y mejorar los procesos de las entidades.

****

**Ejemplo** de minería de datos en los servicios comerciales y financieros:

Una entidad financiera implementa una herramienta de minería de datos para evaluar a sus clientes a la hora de hacer los préstamos extrayendo información de los hábitos de pago, historial crediticio, movimientos bancarios entre otros datos que le servirán para su análisis y clasificar a cada uno según su capacidad de endeudamiento.

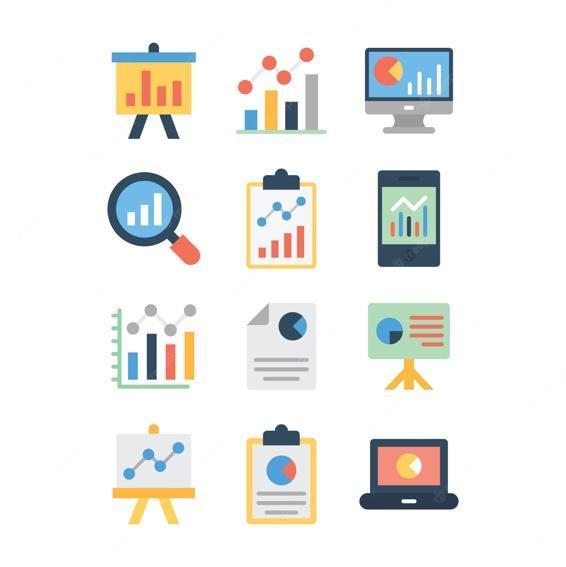
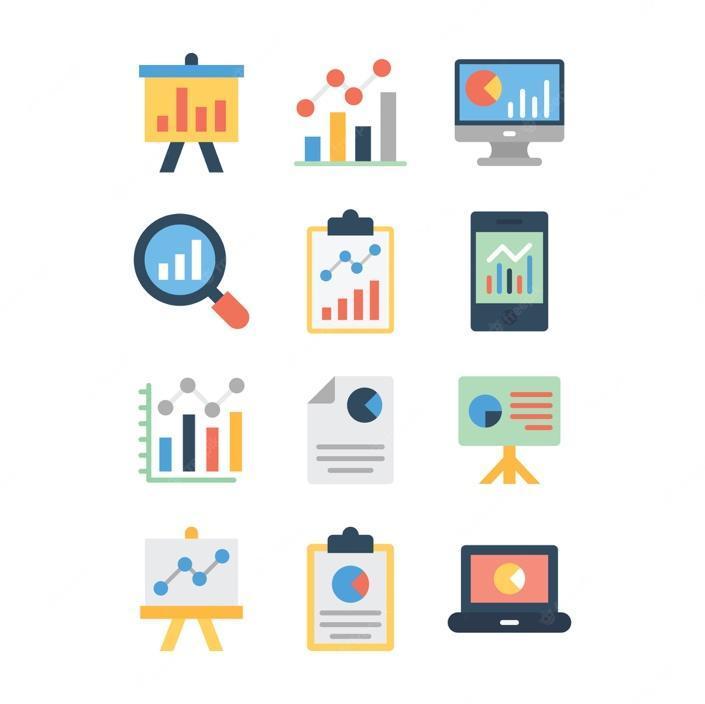
Teniendo en cuenta el ejemplo presentado, la gestión que se debe hacer con la minería de datos es:



La gestión anteriormente presentada, se debe realizar a través de:



**Proceso de datos**, es decir revisar la información y eliminar lo que no sirve.

* **Elección de características**, es decir tomar las variables influyentes a solucionar el problema.
* **Uso de algoritmos**, es decir utilizar el modelo de conocimiento más acorde según las técnicas.
* **Análisis y evaluación**, es decir comprobar que las validaciones sean reales.

En el siguiente recurso se expondrán las técnicas que utiliza esta herramienta de recolección de datos:



* 1. **Métodos de recopilación e integración de datos.**

Debe entenderse que los procesos de recolección de la información constan una gran multiplicidad de técnicas o herramientas para la recopilación de la misma, las técnicas más usadas son:



* 1. **Preparación de datos estructurados.**

Esta actividad se enfoca en limpiar y recopilar los datos que van a ser utilizados, consultados, investigados, extraídos o buscados, con el propósito de que la información sea precisa y válida para los diferentes análisis que se requieren hacer en el proceso, garantizando la seguridad de los datos obtenidos.

Las organizaciones en la preparación de datos tienen en cuenta:

* Claridad de la información y seguridad en el acceso de un lugar seguro.



* Confiabilidad de los datos obteniendo información fidedigna veraz, visible y auditable por si se debe modificar.



* Comprensible y con capacidad de repetición asegurando su entendimiento para generar estrategias.



Ahora para preparar los datos estructurados se siguen las siguientes fases:

* Identificar los datos que se necesitan, adquirirlos, compilarlos y generar un canal o línea de acceso permanente.
* Revisar la calidad de los datos compilados según la relación con la actividad o situación que se quiere solucionar.
* Limpieza de la información que consiste en corregir, eliminar duplicidad, suprimir datos incompletos y salvaguardar información confidencial.



**Nota:** estimados aprendices esta actividad se realiza por medio de las herramientas digitales, informáticas o de inteligencia artificial según lo estipule cada organización.

* 1. **Ecosistema de datos abiertos**

Como su nombre lo indica es un conjunto de información que está disponible al público en general bajo unos lineamientos de reproducción y reutilización para que los usuarios hagan buen provecho de esos datos y les sirvan según la necesidad o requerimiento.

En Colombia la ley 1712 de 2014 indica la obligación de las organizaciones públicas de emitir datos públicos con excepciones si es información clasificada.



También se comparten datos como resultados de elecciones, contrataciones estatales, pruebas saber entre otros.

El Ecosistema de datos abiertos presenta unas características y principios particulares, que conocerá a continuación:



**2. Bases de datos**

Es una herramienta encargada de guardar los datos y entrelazarlos de forma lógica según su contenido o temática.

Es muy útil en las entidades que buscan gestionar gran cantidad de información por medio de herramientas digitales, buscando soluciones al gran volumen de datos que administran.

**2.1 Clasificación**

Las bases de datos en las entidades usualmente se clasifican según la necesidad de las organizaciones de la siguiente manera:



**2.2 características**

A continuación, se enumeran las características de las bases de datos para un mayor entendimiento:



**2.3 Modelos**

Los modelos de administración de datos se basan en ubicar donde se almacena la información, así como la extracción o consulta de la misma.

Los modelos son:



****

**Nota:** se aclara que los modelos van asociados a la tecnología, características y diferentes herramientas que se integran como la recolección de los datos, según cada entidad, presupuesto y necesidades.

**2.4 Sistemas de Gestión de bases de datos**

La administración y gestión que se hace de la información se realiza a través de un software diseñado para extraer, modificar y programar los datos recolectados. El uso de esta herramienta es elegido según diferentes características y modelos existentes, configurándola según las necesidades de cada entidad.

Por otro lado, estos softwares garantizan la confiabilidad de la fuente de datos, su seguridad, extracción, recuperación y gestión, así como la facilidad de su administración.

En las organizaciones el uso de las bases de datos se gestiona por medio de herramientas digitales que le permiten su administración, esto por un software que responde a las necesidades según la magnitud de las bases que maneje cada entidad, a continuación, se presentan las más utilizadas.



**3. Big Data**

Este término hace referencia al conjunto de grandes volúmenes de datos estructurados y no estructurados, que permiten a las organizaciones tomar mejores decisiones tras el análisis de datos complejos de una forma más comprensible y que necesitan aplicaciones informáticas apropiadas para el tratamiento adecuado de los datos, cada día se procesan más de 2,5 quintillones de bytes de datos en el mundo, que pueden ser desde consultas en el navegador de internet hasta complejos informes sobre el proceso de fabricación de un producto. Este término hace referencia a conjuntos de datos tan grandes y complejos que necesitan aplicaciones informáticas apropiadas para el tratamiento adecuado de los datos (Barranca, 2012).

El ejemplo más claro es la pandemia por el Coronavirus, en la que, mediante herramientas de Big Data e Industria 4.0 para combatir el COVID, se pudieron generar sistemas de geolocalización para conocer la propagación del virus en tiempo real. Dentro de la industria, las aplicaciones son más puntuales como, por ejemplo, en donde la planta está recibiendo la información de los pedidos ajustados, basados en las preferencias de color o modelo, de los consumidores que visualizan los productos en buscadores de internet. Otra aplicación, es el seguimiento de la eficiencia de los procesos en cada una de las fases en tiempo real.

Los beneficios más destacados que aporta el Big Data a la industria es la velocidad de entrega de la información en grandes volúmenes y variedad de formatos dispuestos en distintas fuentes como redes sociales, internet de las cosas, bases de datos internas y externas, Web Service entre otras. Además de estos beneficios, hay otros como el seguimiento de equipos en tiempo real, para identificación de eventos que tengan probabilidad de ocurrencia de una parada de producción por daños mecánicos o bajas de energía e identificación de cuellos de botella en el proceso productivo.

Cuando se hace referencia al *Big data*, también se asocia a las “V” que lo caracterizan, las cuales se presentan en el siguiente recurso de aprendizaje:



**3.1 Características**

El contenido de información que se transmite funciona bajo la siguiente metodología conocida como las 3 V:



**3.2 Tipos de datos**

Los datos que se presentan en la gestión del Big Data son:



**3.3 Principios del análisis masivo de datos**

Para la administración y análisis de datos se presentan los siguientes principios:



**3.4 Tecnologías para analizar grandes volúmenes de datos**

En la siguiente tabla se presentan las herramientas para el proceso de análisis y administración del Big Data:



Imágenes de Diana | Vectores, fotos de stock y PSD gratuitos

**Nota**: el uso de cualquiera de estas herramientas para analizar datos por medio del Big Data depende de la complejidad, volumen y tipo de información que maneje cada entidad.

**4. Técnicas de gestión de datos masivos**

El proceso para realizar la gestión de datos está inmerso en cuatro técnicas generales que son:

* **Asociación**: hace la actividad de relacionar los datos según sus variables.
* **Minera de datos**: extrae datos que contienen relación con las variables relacionadas.
* **Agrupación**: toma los grupos grandes y los divide en pequeños para facilitar el análisis y encontrar similitudes con el propósito de conocer la estructura de la información.
* **Análisis de texto**: realiza la modelación de la información extraída por contenido textual.

Cabe resaltar que estas técnicas van acompañadas por las diferentes herramientas digitales y métodos que escoge cada empresa.

**4.1 Metodologías**

En la actualidad existen varias metodologías que se utilizan para el análisis de datos masivos de las cuales el aprendiz conocerá el método, su característica y ejemplo para que pueda relacionar en su entorno laboral mediante la siguiente tabla.



Imágenes de Diana | Vectores, fotos de stock y PSD gratuitos**Nota:** cada entidad utiliza el método según requerimientos en su gestión.

**4.2 Almacenamiento**

El almacenamiento es la actividad de guardar los datos por medio de herramientas que se utilizan para conservar la información. Los tipos de herramientas pueden ser magnéticas, ópticas, sólidas, informáticas y digitales con las cuales la entidad puede almacenar los datos.

A continuación en el siguiente video, se presentan las principales herramientas de almacenamiento de datos:



**4.3 Lenguajes de consulta y manipulación de datos**

Es el lenguaje que utilizan los sistemas de gestión de la información, para que los navegantes o usuarios puedan ingresar a las bases de datos revisarlas y editarlas si es necesario. Es decir, por medio de este idioma se manipulan las bases de datos.

A continuación se presenta el lenguaje de administración de datos:



1. **SÍNTESIS**

Para concluir con la unidad temática se retroalimenta sobre la importancia de la información en las organizaciones y que el debido análisis de los datos aporta en la sostenibilidad de las entidades y se invita a observar el siguiente mapa conceptual.



Esperamos que el contenido temático de esta formación sea de gran apoyo cada día en la gestión de su aprendizaje y lo invitamos a resolver la actividad didáctica que le servirá como refuerzo de lo aprendido.

¡Muchos éxitos!

1. **ACTIVIDADES DIDÁCTICAS (OPCIONALES SI SON SUGERIDAS)**

| DESCRIPCIÓN DE ACTIVIDAD DIDÁCTICA | |
| --- | --- |
| Nombre de la actividad | Completa los espacios y refuerza el conocimiento. |
| Objetivo de la actividad | Retroalimentar mientras se evalúa por medio de búsqueda de palabras que hacen falta en los espacios las cuales están en el contenido temático. |
| Tipo de actividad sugerida | Completar los espacios |
| Archivo de la actividad  (Anexo donde se describe la actividad propuesta) | Anexo6\_ActividadDidactica1\_CF07 |

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 |
| --- | --- | --- | --- |
| 1.2 Técnicas de análisis de datos. | Ecosistema de Recursos Educativos Digitales SENA. [Ecosistema de Recursos Educativos Digitales]. (2022, 4 de marzo). Analítica de datos. [video]. Youtube | Video | <https://www.youtube.com/watch?v=eK0QG1SUk9A> |
| 2. Bases de datos | Ecosistema de Recursos Educativos Digitales SENA. [Ecosistema de Recursos Educativos Digitales SENA]. (2021, 24 de mayo). Sistemas de gestión de bases de datos: Ventajas MySQL. [video].Youtube | Video | <https://www.youtube.com/watch?v=DubJRQFmLLk> |
| 3. Big Data | Ecosistema de Recursos Educativos Digitales SENA. [Ecosistema de Recursos Educativos Digitales SENA] . (2020, 20 de octubre). Analítica de Datos [video]. Youtube | Video | <https://www.youtube.com/watch?v=LFupF0du7b4> |
| 4. Técnicas de gestión de datos masivos | Huarcaya Almeyda, R. [Remigio Huarcaya Almeyda]. (2020, 18 de octubre). Relación muchos a muchos | Modelamiento de datos [video]. Youtube | Video - | <https://www.youtube.com/watch?v=HL7EHau2tbA> |

1. **GLOSARIO**

| TÉRMINO | SIGNIFICADO |
| --- | --- |
| Almacenamiento de datos | Guardar datos en un software u otra herramienta, generalmente de forma ordenada, para poder disponer cuando se necesite o convenga. |
| Análisis | Examen detallado de una cosa para conocer sus características o cualidades, o su estado, y extraer conclusiones, que se realiza separando o considerando por separado las partes que la constituyen. |
| Big Data | Datos que contienen una mayor variedad y que se presentan en volúmenes crecientes y a mayor velocidad. |
| Datos | Información concreta sobre hechos, elementos, etc., que permite estudiarlos, analizarlos o conocerlos. |
| Gestión | Acción o trámite que se lleva a cabo para conseguir o resolver una situación. |
| Herramienta digital | Paquetes informáticos que están en las computadoras, o en dispositivos electrónicos como celulares y tabletas, entre otros. |
| Información | Conjunto de datos, ya procesados y ordenados para su comprensión, que aportan nuevos conocimientos a un individuo o sistema sobre un asunto, materia, fenómeno o ente determinado. |
| Software | Conjunto de programas y rutinas que permiten a la computadora realizar determinadas tareas. |
| Toma de decisiones | Proceso en el cual un individuo o grupo de individuos deben elegir entre varias opciones. Este proceso se activa cuando hay que resolver un conflicto o situación que requiere de una acción concreta |

1. **REFERENCIAS BIBLIOGRÁFICAS**

Big Data. Técnicas, herramientas y aplicaciones María Pérez Marqués Copyright © 2015 RC Libros

Elmasri, R. & Shamkant, B. (2001). Introducción a los sistemas de bases de datos. Prentice Hall 7ª edición.

Connolly. T. & Begg, C. (2007). Sistemas de bases de datos. Addison-Wesley, 5ª edición.

**G. CONTROL DEL DOCUMENTO**

| Autor (es) | Nombre | Cargo | Dependencia  *(Para el SENA indicar Regional y Centro de Formación)* | Fecha |
| --- | --- | --- | --- | --- |
| Diego Villalobos Barreto | Experto Temático | Sena Distrito Capital Grupo De Apoyo Administrativo - Cgi | Junio de 2022 |
| Caterine Bedoya Mejía | Diseñadora Instruccional | Regional Distrito Capital – Centro de Gestión Industrial | Junio de 2022 |
| Carolina Coca Salazar | Asesor Metodológico | Regional Distrito Capital- Centro de Diseño y Metrología | Junio de 2022 |
| Rafael Neftalí Lizcano Reyes | Responsable equipo de Desarrollo Curricular Ecosistema de Recursos Educativos Digitales | Regional Santander - Centro Industrial del Diseño y la Manufactura | Junio de 2022 |
| Sandra Patricia Hoyos Sepúlveda | Corrección de estilo | Regional Distrito Capital - Centro de Diseño y Metrología | junio de 2022 |

1. **CONTROL DE CAMBIOS**

**(Diligenciar únicamente si realiza ajustes a la Unidad Temática)**

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