



Actividad | 2 | C Software, Personal y Procesos

Minería y Análisis de Datos

Ingeniería en Desarrollo de
Software

TUTOR: Ing. Iván Paredes

ALUMNO: Adriana Esteban López

FECHA: 26 de octubre de 2025

DESARROLLO

1. Software de Data Mining

La Minería de Datos es una técnica para encontrar patrones, mientras que la analítica de datos utiliza estos patrones para su interpretación y toma de decisiones; y para llegar al objetivo la Minería de Datos se apoya en el uso de diferente software, en el desarrollo de esta actividad estaremos mencionando algunos de ellos:

RapidMiner

Tecnología de Minería de Datos	¿Porqué propondrías ese Software?	¿Qué procesos de Minería de Datos puede realizar el Software?
RapidMiner	Interfaz fácil de usar. Fácil conexión a Bases de Datos. Ideal para proyectos de diferentes magnitudes. El lenguaje base en el que esta desarrollado en Java. Buena predicción de escenarios futuros en base a los datos recopilados.	1. Preparación de los datos 2. Modelado y Análisis 3. Evaluación e implementación
Weka	Software de código abierto. Interfaz intuitiva. Integración con Java. Herramientas de procesamiento y visualización. Facilidad de uso	1. Clasificación 2. Agrupamiento 3. Regresión 4. Selección de características 5. Preprocesamiento de datos
Orange	Permite una programación visual y versátil para un análisis de datos. Interfaz intuitiva Software libre Multiplataforma	1. Carga y preparación de datos. 2. Visualización de datos. 3. Modelado. 4. Evaluación y prueba de modelos. 5. Análisis de texto. 6.

¿Cuál es el mejor gestor de base de datos para este proyecto?

El elegir un gestor de Base de Datos va a depender del proyecto en cuanto al volumen de datos que se vayan a utilizar y/o analizar, los objetivos del proyecto y sobre todo el presupuesto, sin embargo, la opción más atinada sería **My SQL Server** ya que cuenta con una gran variedad de herramientas.

2. Perfiles y roles

➤ *Líder del proyecto*

Funciones	Perfil
1. Planificación. 2. Coordinación del equipo. 3. Gestión de Riesgos. 4. Gestión del Presupuesto. 5. Supervisión. 6. Entrega final del proyecto	Escolaridad: Maestría Liderazgo efectivo Uso avanzado de herramientas digitales Gestión de equipos remotos Disposición al cambio Habilidad para la negociación Trabajo bajo presión para la resolución de problemas

➤ *Científico de Datos*

Funciones	Perfil
Desarrollar hipótesis Modelado de datos Análisis de información	Escolaridad: Ingeniería Dominio de lenguajes de programación Dominio de Estadísticas y Matemáticas Conocimiento en algoritmos de aprendizaje

➤ *Ingeniero de Datos*

Funciones	Perfil
Creación y mantenimiento de Bases de Datos Seguridad e integración de los datos Resolución de problemas técnicos	Escolaridad: Ingeniería Dominio de herramientas tecnológicas relacionadas con Bases de Datos Disponibilidad a la comunicación y colaboración

➤ *Analista de Negocios*

Funciones	Perfil
<p>Recopilación de datos</p> <p>Identificación de patrones y tendencias.</p> <p>Presentación de resultados del análisis de la información</p>	<p>Formación académica en áreas como Estadística, Matemáticas, Ingeniería o Economía</p> <p>Dominio de herramientas y técnicas de estadísticas</p>

➤ *Administrador de Base de Datos*

Funciones	Perfil
<p>Evaluación del software de base de datos</p> <p>Identificar las necesidades y permisos de los usuarios</p> <p>Realizar y probar modificaciones en la estructura de la base de datos</p> <p>Mantener la integridad y el rendimiento de la base de datos</p> <p>Fusionar bases de datos antiguas con bases de datos nuevas</p> <p>Realizar copias de seguridad y restaurar datos para evitar la pérdida de datos</p> <p>Probar y evaluar la base de datos periódicamente para garantizar la privacidad y seguridad de los datos</p> <p>Proporcionar soporte y capacitación en gestión de datos a los usuarios</p>	<p>Dominio de bases de datos, sistemas de gestión de bases de datos (DBMS) y lenguajes de consulta.</p> <p>Conocimiento sobre protección de datos, derechos de acceso, firewalls y software antivirus, con actualización constante sobre la legislación en esta área.</p> <p>Resolución de problemas:</p> <p>Habilidades específicas para gestionar bases de datos en entornos de nube (AWS, Azure, Google Cloud) y optimizar costos.</p>

¿Qué roles o perfiles escogiste para el desarrollo del proyecto?

Líder del proyecto, Científico de Datos, Ingeniero de Datos, Analista de Negocios, Administrador de Base de Datos.

¿Por qué son fundamentales?

Los perfiles propuestos son cruciales para el éxito del proyecto ya que cada uno va aportar sus conocimientos dentro de su área de expertis para garantizar la calidad dentro del desarrollo del proyecto a través del uso de técnicas establecidas.

Con base en los roles seleccionados, ¿cuánto personal se va a contratar?

Se considera oportuno y prudente la contratación de personas.

3. Procesos del proyecto

- **Definición del problema:** Establecer la problemática a solucionar a través de la comprensión de los objetivos y exigencias del negocio.
- **Preparación de los Datos:** Recolección de datos inicial y la cual continua conforme se van desarrollando las actividades.
- **Exploración de los datos:** Realización de todas las actividades necesarias para construir el conjunto de datos final se incluyen las tareas de preparación de selección de los datos, la limpieza de éstos, la construcción de nuevas variables, la integración de los datos y el formateo de los mismos.
- **Generación de modelos:** Aplicación del software *RapidMiner* para la aplicación de las técnicas de minería de datos a nuestros datos.
- **Validación de modelos:** Se mide la viabilidad del modelo generado en la etapa anterior el cual debe tener una alta calidad desde una perspectiva de análisis de datos.
- **Implementación de modelos:** Aplicación del modelo generado y validado dentro de la aplicación directa con el usuario final.

Propuesta del personal que se requiere:

Rol a desempeñar	Función principal	Perfil
Líder del proyecto	Definición del alcance y los objetivos, gestión y coordinación al equipo, controlar los riesgos, y mantener una comunicación constante con todas las partes interesadas.	Escolaridad: Maestría Liderazgo efectivo Uso avanzado de herramientas digitales Gestión de equipos remotos Disposición al cambio Habilidad para la negociación Trabajo bajo presión para la resolución de problemas
Científico de Datos	Recopilación y preparación de datos para su análisis	Escolaridad: Ingeniería Dominio de lenguajes de programación Dominio de Estadísticas y Matemáticas Conocimiento en algoritmos de aprendizaje
Ingeniero de Datos	Seguridad e integración de los datos	Escolaridad: Ingeniería Dominio de herramientas tecnológicas relacionadas con Bases de Datos Disponibilidad a la comunicación y colaboración
Analista de Negocios	Presentación de resultados del análisis de la información	Formación académica en áreas como Estadística, Matemáticas, Ingeniería o Economía Dominio de herramientas y técnicas de estadísticas
Administrador de Base de Datos	Garantizar que los datos sean accesibles, seguros y estén disponibles de forma eficiente	Dominio de bases de datos, sistemas de gestión de bases de datos (DBMS) y lenguajes de consulta. Conocimiento sobre protección de datos, derechos de acceso, firewalls y software antivirus, con actualización constante sobre la legislación en esta área. Resolución de problemas: Habilidades específicas para gestionar bases de datos en entornos de nube (AWS, Azure, Google Cloud) y optimizar costos.

RapidMiner Studio

Requisitos mínimos:

- **Procesador:** Dual core, 2 GHz.
- **RAM:** 4 GB.
- **Espacio en disco:** Más de 1 GB de espacio libre.
- **Resolución de pantalla:** 1280x1024.
- **Sistema operativo:**
 - Windows 7 o posterior (se recomienda 64 bits).
 - Linux (solo 64 bits).
 - macOS X 10.10 o posterior.
- **Plataforma Java:** Se recomienda altamente una plataforma Java de 64 bits. Se puede usar Oracle Java 8 u OpenJDK 8.

Requisitos recomendados:

- **Procesador:** Quad core, 3 GHz o más rápido.
- **RAM:** 16 GB o más.
- **Espacio en disco:** Más de 100 GB de espacio libre.

Conectividad:

- Conexión a bases de datos relacionales con controlador JDBC.
- Conexión a bases de datos NoSQL, servicios en la nube y tipos de archivos como CSV, Excel, XML, entre otros.

¿Cómo se realizaría la ejecución del proyecto para que sea exitoso al implementarlo en el área en el que Juan está a cargo?

Juan debe de asumir el rol como líder y tener claro cuál es el objetivo del proyecto y por ende su participación atinada en cada una de la fases del desarrollo del proyecto, es importante la delegación de las funciones para cada uno de los perfiles involucrados, pero no debe de dejar de lado la supervisión de los mismo.

Se agrega dicha actividad a la plataforma de GitHub a través del siguiente link:

<https://github.com/22HADRIA/Miner-a-y-An-lisis-de-Datos>

REFERENCIAS

Las 8 mejores herramientas de minería de datos en 2025: características, ventajas, desventajas y casos de uso

<https://www.astera.com/es/type/blog/data-mining-tools/>

Proceso de minería de datos

<https://www.ibm.com/docs/es/db2/11.1.0?topic=miner-data-mining-process>

Software de data mining: realiza análisis de datos más efectivos

<https://www.ionos.mx/ofertas-especiales?itc=2E5H5WQ0-TH53Y8-TS1AJ68>

Ventajas y desventajas de RapidMiner: descripción completa

<https://www.theiotacademy.co/blog/rapidminer-advantages-and-disadvantages/>

RapidMiner para ciencia de datos y minería de datos

<https://medium.com/image-processing-with-python/rapidminer-for-data-science-and-data-mining-1547bbc3b475>

Los múltiples beneficios de Weka

<https://www.opensourceforu.com/2024/04/the-many-benefits-of-weka/>

¿Qué es Weka?

<https://www.appliedaicourse.com/blog/what-is-weka/#:~:text=It%20supports%20graphical%20user%20interfaces,techniques%20without%20extensive%20programming%20knowledge.>

Orange Data Mining

<https://prezi.com/bsxip-n8gph8/orange-data-mining/>

<https://www.linkedin.com/advice/0/heres-how-you-can-define-roles-responsibilities-team-pevgc?lang=es&originalSubdomain=es#cient%C3%ADfico-de-datos>

¿Cuáles son las funciones básicas de un Project Leader?

<https://www.obsbusiness.school/blog/cuales-son-las-funciones-basicas-de-un-project-leader>

Habilidades clave que debe tener un Project Manager

<https://www.ui1.es/blog-ui1/habilidades-clave-que-debe-tener-un-project-manager-en-2025>

¿Qué es un Data Scientist?

<https://hireline.io/mx/enciclopedia-de-perfiles-de-tecnologia/data-scientist>

Data Engineer: Qué es y qué funciones tiene

<https://www.michaelpage.es/advice/profesi%C3%B3n/tecnolog%C3%ADa/perfil-de-data-engineer>

Analista de datos: Qué hace, sueldo y formación

<https://www.michaelpage.es/advice/profesi%C3%B3n/tecnolog%C3%ADa/perfil-de-analista-de-datos>

Analista de datos: Qué hace, sueldo y formación

<https://www.michaelpage.es/advice/profesi%C3%B3n/tecnolog%C3%ADa/perfil-de-analista-de-datos>

Administrador de base de datos

<https://www.cambridgehealth.edu/data-management/database-administrator/>