

Actividad | 2 |

Software, Personal y Procesos

Minería y Análisis de Datos

Ingeniería en Desarrollo de Software



TUTOR: Iván Paredes

ALUMNO: Hernán Everardo Velázquez Zavala

FECHA: 25/10/2025

Índice

Desarrollo	2
Software en Data Mining	2
Perfiles y roles	4
Proceso del proyecto	6
Conclusión	8
Referencias	9
Portafolio de GitHub.....	9

Desarrollo

Software en Data Mining

El proyecto tiene como principal objetivo definir e implementar un sistema de minería de datos que transforme la información en conocimiento de gran utilidad para mejorar la toma de decisiones, es por esto que el área de Inteligencia de Negocios y Minería de Datos Nacional en conjunto con su director, son los responsables de esta investigación, análisis e implementación.

Para lograr esto, se investigó y se filtraron las propuestas de software hasta llegar a tres herramientas de acuerdo a sus prestaciones, costo-beneficio e integración con las necesidades; también se revisaron gestores de base de datos para definir el más adecuado para este proyecto.

A continuación se presenta la tabla con el resultado de la investigación.

Tecnología de Minería de Datos	Justificación de la propuesta del software	Procesos de minería de datos que puede realizar el software
KNIME Analytics Platform	Es un programa gratuito y open source (código abierto); este permite diseñar los flujos de trabajo de manera visual facilitando su revisión y análisis; también, tiene la opción de la limpieza y la transformación de datos sin tener que programar.	Análisis predictivo, clasificación, clustering, minería de texto, regresión, análisis de social networks, entre otros.
RapidMiner Studio	Este programa ofrece una interfaz relativamente intuitiva y el fácil acceso a distintos algoritmos alojados en su propia biblioteca. Es muy útil en la creación de modelos predictivos y de una manera rápida incluso sin tener amplia experiencia en programación.	Clasificación, regresión de minería de texto, segmentación, análisis de series temporales, entre otros.

Microsoft Azure Machine Learning Studio	Este programa en nube permite la escalabilidad de procesos y la creación de modelos de IA, además, cuenta con soporte a nivel empresarial.	Análisis predictivo, alimentación de su aprendizaje automático, minería de texto, procesa volúmenes de datos muy grandes, entre otros.
---	--	--

¿Cuál es el mejor gestor de base de datos para este proyecto?
<p>Para este proyecto se define que PostgreSQL es la mejor alternativa, principalmente debido a que es un software gratuito y muy estable, además, funciona correctamente con diferentes herramientas para la minería de datos, asegurando el buen funcionamiento con cualquiera de las tres alternativas previamente presentadas.</p> <p>Este software permite manejar grandes cantidades de información y resultar consultas tanto simples como complejas con un buen rendimiento.</p> <p>En conclusión, se elige este software por su flexibilidad y funcionamiento sin el requerimiento de costosas licencias.</p>

Perfiles y roles

Una vez realizada la investigación, se pueden responder las siguientes preguntas, las cuales nos brindan un resumen de la información recabada.

¿Qué roles o perfiles se eligen para el desarrollo del proyecto?

Para este proyecto decidí que se requieren cuatro diferentes perfiles: Data Scientist, Data Engineer, Business Analyst y Project Manager.

A continuación se presentan en una tabla una breve descripción de cada cargo.

Rol	Descripción
Data Scientist (Científico de datos)	Responsable del análisis de datos, el desarrollo de modelos predictivos y filtrado y evaluación de resultados.
Data Engineer (Ingeniero de Datos)	Responsable del diseño y el mantenimiento de la infraestructura de los datos y desarrollo de procesos referentes a los datos como lo son la extracción, transformación y carga de información.
Business Analyst (Analista de Negocios)	Responsable de analizar e interpretar los resultados de los análisis técnicos y diseñar las presentaciones de información de manera comprensible y ágil a directivos y clientes.
Project Manager (Administrador del proyecto ó Líder de Proyecto)	Responsable de coordinar y supervisar el proyecto de inicio a fin, programando los tiempos de trabajo y administrando los recursos.

¿Por qué son fundamentales?

Considero que estos roles son fundamentales porque cubren todas las etapas del proyecto, desde la gestión de los datos hasta la toma de decisiones. Sin un *Data Engineer* no habría datos limpios ni organizados; sin el *Data Scientist* no se podrían generar modelos ni análisis útiles; el *Business Analyst* traduce los resultados técnicos en información práctica para la dirección; y el *Project Manager* mantiene todo el proceso bajo control. En conjunto, garantizan que el sistema de minería de datos funcione correctamente y aporte valor real a la organización.

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

Para iniciar el proyecto propongo contratar una persona por cada rol, es decir, un total de cuatro integrantes. Este equipo es suficiente para desarrollar la primera versión del sistema de minería de datos y evaluar sus resultados. En el futuro, si el proyecto crece, se podrían añadir más analistas o ingenieros según las necesidades del área.

Proceso del proyecto

El proyecto se enfocará en transformar la información de la organización en conocimiento estratégico, mediante la aplicación de técnicas de minería de datos para detectar patrones, predecir tendencias y sustentar la toma de decisiones basadas en datos y conocimiento útil.

Después de la investigación, análisis y comparación de las tres propuestas de herramientas de minería de datos (KNIME, RapidMiner y Azure Machine Learning Studio), propongo utilizar KNIME Analytics Platform como la opción más conveniente para el proyecto.

Esta decisión se basa principalmente en que es un software gratuito y open source (de código abierto), además de su alta compatibilidad con diferentes fuentes de datos, otro punto importante es su entorno visual ya que facilita el diseño de flujos de trabajo sin requerir programación avanzada, permitiendo que diferentes perfiles de usuarios puedan utilizarlo o aprender de manera relativamente sencilla.

Además, KNIME permite la integración con Python, R y bases de datos como PostgreSQL, lo que le otorga flexibilidad y capacidad de expansión a largo plazo. Su comunidad es muy activa en internet y la documentación accesible también facilita el aprendizaje y soporte técnico, reduciendo los costos iniciales de capacitación y mantenimiento.

El gestor de base de datos seleccionado es PostgreSQL, por su estabilidad, seguridad y excelente desempeño en el manejo de grandes volúmenes de datos. Al ser también de código abierto, evita gastos en licencias y permite personalización según las necesidades del proyecto.

El desarrollo del proyecto está planeado en seis fases; el objetivo es implementar el sistema y los procesos un paso a la vez para asegurar que todo funcione correctamente hasta concluir con la implementación en el área correspondiente.

Fase	Descripción	Responsable
1.- Análisis de necesidades y planteamiento de objetivos	Se realiza un diagnóstico identificando las necesidades del sistema, los tipos de datos disponibles, los objetivos estratégicos del proyecto y los indicadores que se desean optimizar. Se define también el alcance y los entregables así como la programación de fechas para el plan de trabajo.	Project Manager y Business Analyst
2.- Arquitectura de datos	Se diseña la estructura de almacenamiento de datos utilizando PostgreSQL como gestor principal. Se planifican los procesos de recolección, limpieza y organización de la información. Además, se establece la infraestructura técnica (servidores y estaciones de trabajo) y se preparan los entornos de desarrollo y pruebas.	Data Engineer y Project Manager
3.- Integración y preparación de los datos	Se implementan los procesos de extracción, transformación y carga (ETL) para consolidar la información de distintas fuentes. KNIME se utiliza para automatizar tareas de limpieza, normalización y validación de datos, garantizando su calidad y consistencia antes del análisis.	Data Engineer
4.- Modelado y análisis de datos	Se desarrollan modelos de minería de datos utilizando KNIME. Se aplican técnicas de clasificación, regresión, segmentación y minería de texto. Los resultados se validan mediante métricas de rendimiento y se seleccionan los modelos más precisos y eficientes.	Data Scientist
5.- Interpretación y comunicación de resultados	Se convierten los resultados técnicos en información útil para la toma de decisiones. Se elaboran reportes visuales, dashboards y presentaciones ejecutivas que muestran hallazgos clave, tendencias y predicciones relevantes para la dirección.	Business Analyst y Data Scientist
6.- Implementación, capacitación y seguimiento	Se implementa la solución en el entorno productivo. Se capacita al personal del área en el uso de la herramienta y se establecen procedimientos de mantenimiento y mejora continua. Se realizan evaluaciones periódicas para medir el impacto del sistema en la eficiencia y toma de decisiones.	Equipo

Conclusión

La propuesta presentada busca que el área de Inteligencia de Negocios y Minería de Datos Nacional cuente con una solución moderna, funcional y escalable. El objetivo principal del proyecto es transformar la información en conocimiento útil para la toma de decisiones estratégicas, aprovechando herramientas de minería de datos confiables y un equipo especializado.

Con esta base sólida, la organización podrá identificar patrones, optimizar procesos y anticipar tendencias de manera eficiente.

Referencias

Video conferencing, web conferencing, webinars, screen sharing. (s. f.-p). Zoom. Consultado el 18 noviembre de 2024.

https://academiaglobal-mx.zoom.us/rec/share/DNZnWUUIKvUr9u4LeOT-pWHITaQ3hoo2SKePbGpCTkhaS7WYtdbJYFg399G2FSge.cU9iVh8gliBAM_Pl

KNIME AG. (s. f.). *Home* | *KNIME*. KNIME. Consultado el 25 de octubre de 2025.

[://www.knime.com/](https://www.knime.com/)

Data Analytics and AI Platform | *Altair RapidMiner*. (s. f.). Consultado el 25 de octubre de 2025.

<https://rapidminer.com/>

Sdgilley. (s. f.). *Azure Machine Learning documentation*. Microsoft Learn. Consultado el 25 de octubre de 2025. <https://learn.microsoft.com/en-us/azure/machine-learning/>

PostgreSQL: about. (s. f.). The PostgreSQL Global Development Group. Consultado el 25 de octubre de 2025. <https://www.postgresql.org/about/>

Portafolio de GitHub

Se comparte la actividad para revisión y consulta a través de Git Hub

<https://github.com/IDS-H/MineriaYAnalisisdeDatos.git>