

Actividad | 2 | Software, Personal y Procesos

Minería de datos

Ingeniería en Desarrollo de Software



TUTOR: Iván Paredes

ALUMNO: Juan Carlos Legorreta Ramos




FECHA: 22-10-2025

Índice

Desarrollo.....	1
Software de Data Mining.....	1
Perfiles y roles.....	2
Proceso del proyecto.....	4

Desarrollo

Software de Data Mining

Tecnología de Minería de datos	¿Por qué propondrías ese software?	¿Qué procesos de minería de datos puede realizar el software?
<p>KNIME Analytics Platform</p> 	<p>Es un software gratuito y de código abierto permitiendo Crear flujos de trabajo y éste se puede integrar fácilmente con Python, R y PostgreSQL siendo el más adecuado para limpieza y preparación de datos.</p>	<p>Este software nos permite tanto unida como preparar los datos al igual que explorar información y creando modelos predictivos, así como agrupar datos y mostrarlos en gráficos.</p>
<p>RapidMiner Studio</p> 	<p>Este software nos ofrece una interfaz muy intuitiva que consiste en arrastrar y soltar permitiendo así crear modelos predictivos ya que no sea necesario programar con un soporte de aprendizaje automático y una minería avanzada.</p>	<p>Este nos permitirá clasificar y predecir los datos como validar modelos, reduciendo información y que sea compleja a solo lo esencial haciendo análisis predictivo y generando los reportes.</p>
<p>SAS Enterprise Miner</p> 	<p>Este es un software profesional utilizado para grandes cantidades de datos ofreciendo así soporte técnico, Minería de datos y un modelado avanzado.</p>	<p>Con este software podremos explorar datos, segmentar clientes, así como analizar textos, hacer predicciones y así crear reportes ejecutivos.</p>

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

Para este proyecto el mejor gestor de bases de datos es PostgreSQL pues varias razones:

- Es gratuito y de código abierto ya que no se necesita pagar por licencias y este puede ser modificado según sus necesidades.
- Es seguro ya que tiene mecanismos de control de acceso y protección de datos de modo avanzado.
- Escalable ya que se adapta a proyectos pequeños, así como grandes en datos.
- Tiene una gran compatibilidad con varios lenguajes y como lo puede ser Python, java, R entre otros facilitando su integración con programas por herramientas de análisis.
- Se adapta fácilmente las herramientas de minería de datos siendo ya las mencionadas.

Perfiles y Roles

Referencias de las paginas investigadas para perfiles y roles:

Herranz, F. (2023b, agosto 28). Los 12 perfiles de Big Data e Inteligencia Artificial que más demanda la industria 4.0. Industry TALKS. Diario Digital de la Industria Española.

<https://industrytalks.es/los-12-perfiles-big-data-e-ia-que-mas-demanda-la-industria-4-0/>

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

Es parte del desarrollo de este proyecto de minería de datos los perfiles a seleccionar basándose en las funciones más comunes en proyectos de inteligencia de negocios y ciencia de datos:

Project manager encargado de coordinar las actividades, así como la definición de objetivos y controlar el cumplimiento en los plazos del proyecto.

Analista de negocios encargado de Expresar las necesidades del negocio en cuestión de objetivos técnicos al igual que define los indicadores clave del desempeño.

Ingeniero de datos el encargado de preparar, así como limpia e integra los datos de diversas fuentes.

Administrador de base de datos es el encargado de gestionar y mantener la base de datos, así como asegurar su integridad y seguridad.

Desarrollador ETL creado y automatiza los flujos de extracción la transformación y carga de datos.

Científico de datos responsable de analizar la información, así como diseñar modelos predictivos y así generar conocimiento útil para la toma de decisiones.

- **¿Por qué son fundamentales?**

En cuestión cada uno cumplirá con una función en específico.

El **Project manager** mantiene el control y dirección del proyecto.

El **analista de negocios** se da el que conectará los objetivos estratégicos con las soluciones tecnológicas.

El **ingeniero de datos** así como el desarrollador ETL serán los encargados de garantizar la calidad y disponibilidad de los datos.

El **administrador de base de datos** encargado de asegurar el correcto funcionamiento de la base de datos.

Los **científicos de datos** Encargados de transformar los datos en conocimiento y dándole valor espada la organización.

- **¿Cuánto personal se va a contratar?**

El proyecto estará contando con 7 integrantes en donde en cada rol habrá una persona a excepción de científico de datos siendo 2 elementos para este rol.

Proceso del proyecto

La propuesta para que el área de inteligencia de negocios y minería de datos funcione de una forma correcta se propone el uso de 3 herramientas principales las cuales se complementan entre ellas siendo:

KNIME Analytics Platform el cual se usará para limpiar, como preparar y unir los datos y y se empleará **Rapidminer Studio** para analizar la información y crear modelos que ayuden Apple decir los resultados y con la ayuda de **SAS Enterprise Miner** podemos generar reportes más completos y realiza análisis avanzados de manera profesional.

Con el uso de esas herramientas el área podrá trabajar de una forma ordenada y confiable esto desde la recolección de los datos hasta la creación de reportes. Al igual se recomienda utilizar

PostgreSQL como el gestor de base de datos ya que este es gratuito, seguro y permite guardar grandes cantidades de información sin algún problema.

Para la propuesta del personal y los requisitos este equipo se está formando por 7 integrantes donde cada uno tendrá una tarea en específico para que el proyecto avance de una forma correcta siendo:

1. El **project manager** el cual coordinará el trabajo de todos.
2. El **analista de negocios** el que definirá los objetivos del proyecto.
3. El **ingeniero de datos** y el **desarrollador ETL** preparada y organizada los datos.
4. El **administrador de base de datos** cuidada la información y los accesos a esta
5. Los **científicos de datos** encargados de analizar la información y generar resultados útiles.

Los requisitos para el desarrollo del proyecto:

1. **Infraestructura** contando con computadoras o servidores con una buena capacidad en almacenamiento y conexión a internet.
2. **Software** contar con instalación de las herramientas de KENIME, RapidMiner y SAS enterprise Miner.
3. **Base de datos** con la configuración de PostgreSQL para guardar y consultar los datos.
4. **Capacitación** con un breve curso para que el personal aprenda a usar las herramientas.
5. **Seguridad** al cuidar la información con contraseñas y respaldos frecuentemente.
6. **Fuente de datos** conectando los sistemas internos de la empresa y las bases de datos externas que sean útiles para el análisis.

Para que esta propuesta sea exitosa se debe tener la combinación de estos 3 programas eficientes y tener un equipo profesional completo lo que permitirá a la organización analizar de dónde es

cantidades de información de forma rápida y confiable esto de forma sencilla, con todo esto se podrían tomar mejores decisiones y así detectar oportunidades de mejora fomentando así una cultura basada en el uso de datos.

Ejecución del Proyecto

Para poder tener éxito en el proyecto en el área de inteligencia de negocios y minería de datos nacional se propone el seguir el proceso ordenado y por etapas con esto se busca que las herramientas tanto el personal y los datos que se utilicen sean de una manera correcta y eficiente.

Pasos para el desarrollo del proyecto:

- Planeación y definición de objetivos donde se realizará una reunión con el equipo esto para definir y qué es lo que se quiere lograr al igual que qué información se analizará y cuáles son los indicadores de éxito.
- Diseño de la arquitectura tecnológica para elegir las herramientas que se usarán al igual definir cómo se conectarán entre sí.
- La recolección y organización de los datos en donde el ingeniero de datos y el desarrollador ETL entender la función de obtener la información de distintas fuentes limpiarla y dejarla lista para su análisis.
- Análisis y modelo de los datos donde los científicos de datos usada en las herramientas de minería espada descubrí así patrones y queda modelos predictivos para la generación de reportes.
- Validación de resultados donde se revisará que los modelos sean precisos y que los resultados sean útiles para la toma de decisiones dentro del área.
- Implementación de las soluciones en donde los modelos validados se están aplicando en los sistemas internos para que la haya de inteligencia de negocios puedan utilizarlos en sus análisis diarios.

- Capacitación del personal se estará entrenando a los trabajadores del área para que así puedan comprender los reportes y aprendan a usar las herramientas de minería de datos.
- Monitoreo y mejora continua dando seguimiento constante al desempeño de los modelos y a la calidad de los datos en donde se detectan errores o nuevas oportunidades se realicen los ajustes pertinentes para mejorar los resultados.

Con todos estos pasos se está buscando que el proyecto tenga la función correcta desde el inicio y en donde el equipo sea capaz de mantenerlo en operación a largo plazo y para que el proyecto tenga éxito deberá tener una buena planificación así como la comunicación entre los integrantes seguido de la capacitación continua y revisión constante de los resultados.

Enlace a GitHub

<https://github.com/juanitolamaravilla/Miner-a-de-datos.git>