

Actividad | #2

Software, Personal y Procesos

Desarrollo De Sistemas Web I

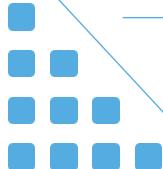
Ingeniería en Desarrollo de Software



TUTOR: Jessica Hernández Romero

ALUMNO: Alex Aldahir García Hernández

FECHA: 22/02/2026



Contenido

Desarrollo.....	2
1. Software de Data Mining	2
2. Perfiles y roles	3
3. Proceso del proyecto	5
4. Repositorio GitHub.....	8
Referencias.....	9

Desarrollo

1. Software de Data Mining

Para la tabla consideré únicamente servicios en la nube para evitar temas de DevOps e infraestructura local lo que permitirá al equipo enfocarse en el análisis de datos.

<i>Tecnología de minería de datos</i>	<i>¿Por qué propondrías ese software?</i>	<i>¿Qué procesos de minería de datos puede realizar el software?</i>
<i>Amazon SageMaker</i>	Por su gran escalabilidad técnica en la nube. Además de que cuenta con herramientas que permiten a los equipos construir, entrenar y desplegar modelos rápidamente.	Cubre todo el ciclo de vida de la minería y análisis de datos. Ingesta de datos a gran escala, almacenamiento de características, entrenamiento distribuido, optimización y monitoreo de calidad del modelo en producción.
<i>IBM watsonx.ai</i>	Tiene un enfoque distinto, una IA confiable, ética y transparente. Principalmente utilizado en sectores regulados como en los gobiernos que requieren mayor trazabilidad y control de sus procesos y sistemas de inteligencia.	Permite el entrenamiento, validación y despliegue responsable de modelos fundacionales, gobernanza de IA, automatización de pipelines de recuperación aumentada (RAG) y detección de sesgos.
<i>Azure Machine Learning</i>	Cuenta con integración nativa y fluida con el ecosistema empresarial de Microsoft, ofreciendo niveles de seguridad y cumplimiento de grado gubernamental. Es ideal para organizaciones que buscan una gestión del ciclo de vida de ML (MLOps) robusta.	Realiza manipulación de grandes volúmenes de datos en la nube, creación y entrenamiento de diversos modelos de aprendizaje automático, implementación escalable y gestión de seguridad integrada en todo el proceso analítico.

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

- Puesto que elegí Amazon SageMaker se podrían utilizar varias bases de datos para cubrir todos los casos de uso utilizando una arquitectura ‘Lakehouse’, unificando Amazon Redshift para el análisis de grandes volúmenes de datos históricos y Amazon Aurora para la ingesta de datos operativos, todos ellos orquestados bajo el nuevo SageMaker Unified Studio.

2. Perfiles y roles

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

Director de BI y Minería de Datos: Se encargará de la planificación, cumplimiento normativo y alineación estratégica acorde al negocio.

Científico de Datos: Se requiere de un perfil que cuente con experiencia en SageMaker Studio, se encargará del diseño de modelos predictivos y algoritmos de aprendizaje automático.

Ingeniero de Datos: Responsable de construir los pipelines que conceten fuentes como Redshift o Aurora con SageMaker, asegurando que la información fluya de forma limpia y segura.

Analista BI: Se enfocarán en traducir los hallazgos en visualización dentro de dashboards o tableros de control.

Ingeniero de MLOps / DevOps Cloud: Especializado en automatizar el ciclo de vida de los modelos y monitorear su rendimiento en producción para evitar derivas de datos.

2. ¿Por qué estos roles son fundamentales?

Director de BI y Minería de Datos: Es importante para controlar los modelos y evitar "Shadow AI" (IA sin control), a demás de asegurar que el proyecto cumpla con la gobernanza.

Científico de Datos: Capas de realizar el análisis y utilizar las capacidades de las herramientas para descubrir patrones ocultos que un análisis tradicional no detectaría.

Ingeniero de Datos: Realizan el tratamiento de los datos y los preparan para que los científicos los puedan consumir, si ellos los científicos perderían mucho tiempo en el tratamiento en lugar de modelar.

Analista BI: Garantiza que la inversión tecnológica se traduzca en decisiones. Es el puente que explica a los directivos nacionales el significado de las predicciones generadas.

Ingeniero de MLOps / DevOps Cloud: Controla los costos del servicio para no generar sorpresas en la factura, se encarga de la optimización de recursos y garantiza que el sistema sea escalable.

3. ¿cuánto personal se va a contratar?

Este sería el esquema inicial y se podrá ir aumentando con forme la carga de trabajo.

Rol	Cantidad	Justificación
Director / Gerente de Área	1	Liderazgo único para la toma de decisiones estratégicas.
Científicos de Datos	2	Uno enfocado en investigación y otro en la implementación de modelos predictivos.
Ingenieros de Datos	2	Requeridos para gestionar la ingesta masiva de múltiples fuentes nacionales.
Analistas de BI	2	Encargados de la creación de reportes para diferentes ministerios o departamentos.
Ingeniero de MLOps / DevOps	1	Especialista transversal para toda la infraestructura de SageMaker.

3. Proceso del proyecto

1. Propuesta de Software y Motivo de Selección

Para un proyecto de escala nacional que requiere alta disponibilidad y procesamiento de volúmenes masivos de información, la tecnología seleccionada es Amazon SageMaker.

Software Propuesto: Amazon SageMaker AI (integrado en el nuevo SageMaker Unified

Studio).

Motivo de Selección:

- Ecosistema Unificado: SageMaker Unified Studio permite que ingenieros, analistas y científicos colaboren en una sola interfaz, eliminando el "cambio de contexto" entre herramientas de SQL, Python y Spark.

- Tecnología Zero-ETL: Permite conectar fuentes de datos nacionales como Amazon Redshift (para analítica masiva) y Amazon Aurora (para datos operativos) sin necesidad de construir tuberías manuales complejas, reduciendo el tiempo de preparación de semanas a horas.
- Escalabilidad y MLOps: Ofrece capacidades de gobernanza centralizada para evitar la "IA en la sombra" (Shadow AI) y herramientas de MLOps para automatizar el reentrenamiento de modelos en producción.

2. Propuesta de Personal y Requisitos

La estructura operativa se basa en un modelo de Centro de Excelencia (COE), diseñado para garantizar la calidad y la ética en el manejo de datos ciudadanos.

Rol	Cantidad	Skills
Director / Gerente de Área	1	Planificación estratégica de datos, gobernanza y ética en IA.
Científicos de Datos	2	Machine Learning, Deep Learning (TensorFlow, PyTorch), Python y R.
Ingenieros de Datos	2	Arquitecturas distribuidas (Spark), SQL, NoSQL y herramientas de integración (AWS Glue).
Analistas de BI	2	Herramientas de visualización (QuickSight, Power BI), SQL avanzado y análisis de KPIs.
Ingeniero de MLOps / DevOps	1	Automatización de pipelines (SageMaker Pipelines), CI/CD y monitoreo de infraestructura.

Total, de personal inicial: **8 personas**.

Fase 1: Comprensión del Negocio (Definición de Impacto)

- Identificar los objetivos estratégicos nacionales (ej. detección de fraudes, optimización de recursos públicos).
- Definir los indicadores clave de rendimiento (KPIs) que el proyecto debe mejorar.

Fase 2: Comprensión de los Datos (Inventario y Calidad)

- Recopilar datos de diversas fuentes institucionales (Redshift, Aurora, S3).
- Evaluar la calidad de la información mediante SageMaker Data Wrangler para detectar anomalías iniciales.

Fase 3: Preparación de Datos (Integración Zero-ETL)

- Utilizar la integración Zero-ETL para sincronizar datos transaccionales en tiempo real hacia el Lakehouse analítico.
- Limpiar, transformar y crear el "Feature Store" (repositorio de características) para asegurar que los modelos usen datos consistentes.

Fase 4: Modelado (Experimentación Inteligente)

- Entrenar modelos utilizando SageMaker Autopilot para generar automáticamente los mejores algoritmos predictivos.
- Utilizar cuadernos Jupyter unificados en SageMaker Studio para ajustes finos y experimentación personalizada.

Fase 5: Evaluación (Validación de Resultados)

- Medir la precisión del modelo contra los objetivos originales.
- Utilizar SageMaker Clarify para detectar sesgos algorítmicos y asegurar una IA ética y transparente.

Fase 6: Implementación y Monitoreo (Puesta en Marcha)

- Desplegar los modelos en "endpoints" escalables para obtener predicciones en tiempo real.
- Configurar SageMaker Model Monitor para recibir alertas automáticas si el rendimiento del modelo decae (concept drift) debido a cambios en el entorno nacional.

4. Repositorio GitHub.

https://github.com/AlxAldahir/mineria_y_analisis_de_datos.git

Referencias

Guan, S. (2025, agosto 19). *Los 15 Mejores Software de Minería de Datos para Empresas en 2025*. Thunderbit.com; Thunderbit. <https://thunderbit.com/es/blog/data-mining-software-tools>

Habilidades Esenciales para un Científico de Datos en 2025. (2025, febrero 28).

Máster en Data Science. <https://www.master-data-scientist.com/habilidades-esenciales--cientifico-datos-2025/>

Perfil y Salario de un Perfil de Business Intelligence Analyst en México en 2026. (s/f).

Hireline. Recuperado el 23 de febrero de 2026, de <https://hireline.io/mx/enciclopedia-de-perfiles-de-tecnologia/perfil-de-business-intelligence-analyst>

Ries, R. (2023, junio 9). Amazon Aurora vs. Redshift: What you need to know.

Missioncloud.com. <https://www.missioncloud.com/blog/amazon-aurora-vs-redshift-what-you-need-to-know>

Top 20+ data mining tools in 2026: Types, features & how to choose. (s/f).

Netcomlearning.com. Recuperado el 23 de febrero de 2026, de <https://www.netcomlearning.com/blog/data-mining-tools>

(S/f-a). Amazon.com. Recuperado el 23 de febrero de 2026, de <https://aws.amazon.com/es/sagemaker/unified-studio/>

(S/f-b). Amazon.com. Recuperado el 23 de febrero de 2026, de <https://aws.amazon.com/es/blogs/aws/simplify-analytics-and-aiml-with-new-amazon-sagemaker-lakehouse/>