



Actividad | 2 |Software, Personal y

Procesos

Minería de Datos

Ingeniería en Desarrollo de Software

TUTOR: Félix Acosta Hernández

ALUMNO: Hazaél Atlai Hervert Martínez

FECHA: 21/10/2024



Índice

Portada-----	1
Índice -----	2
introducción -----	3
Descripción -----	3
Justificación -----	3
Desarrollo	
Software de data mining,perfiles y roles, proceso del proyecto-----	4-9
Conclusión -----	10
Referencias -----	10

Introducción

En el presente proyecto se abordará el diseño e implementación de un plan para la creación del área de Inteligencia de Negocios y Minería de Datos, a cargo de Juan, recientemente nombrado director de esta unidad. El objetivo de este trabajo es identificar los requisitos esenciales para el desarrollo del área, centrándose en tres aspectos fundamentales: la selección del software de minería de datos, los roles y perfiles necesarios, y los procesos a seguir para lograr una implementación exitosa. Además, se detallará cómo estos componentes contribuyen a la optimización de la toma de decisiones estratégicas mediante el análisis de grandes volúmenes de información, esenciales para el crecimiento organizacional en un entorno competitivo.

Descripción

El proyecto tiene como propósito establecer los pilares fundamentales para la implementación de un sistema de minería de datos dentro de la organización. Para ello, se procederá a la selección de tres herramientas de software especializadas en minería de datos, justificando su elección en función de las necesidades del proyecto. Asimismo, se identificarán los perfiles clave que deben formar parte del equipo, como científicos de datos, ingenieros de datos y analistas de negocio, quienes serán responsables de llevar a cabo las actividades relacionadas con el procesamiento y análisis de datos. Finalmente, se describirán los pasos necesarios para la ejecución del proyecto, detallando el proceso de implementación y las metodologías que asegurarán su éxito.

Justificación

El uso de minería de datos se ha convertido en una herramienta indispensable para las empresas que buscan obtener una ventaja competitiva mediante la explotación de grandes volúmenes de datos. Este proyecto está justificado en la necesidad de transformar la información en conocimiento valioso para la organización. Al implementar una solución de minería de datos, se podrán identificar patrones ocultos, predecir tendencias y tomar decisiones más informadas, lo que optimizará los procesos internos y mejorará la eficiencia operativa. Además, contar con un equipo especializado y herramientas adecuadas garantiza que el análisis de datos se realice de manera precisa y efectiva, contribuyendo al éxito a largo plazo de la empresa.

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?
Oracle SQL	<p>Como jefe del proyecto yo usaría Oracle SQL por estas razones:</p> <p>Integración directa: Acceso y manipulación de grandes volúmenes de datos sin moverlos a otras plataformas.</p> <p>Funciones avanzadas: Herramientas para análisis complejos como agregaciones y análisis de series temporales.</p> <p>Escalabilidad: Capacidad para manejar millones o miles de millones de registros.</p> <p>Modelos de Data Mining: Algoritmos de clasificación, regresión, clustering, entre otros, integrados en Oracle Database.</p> <p>Optimización del rendimiento: Uso de Oracle Optimizer para ejecutar consultas y análisis de manera eficiente.</p> <p>Seguridad: Controles de acceso robustos y características de seguridad para proteger los datos.</p> <p>Visualización y reporting: Integración con herramientas para representar los resultados gráficamente.</p>	<p>Clasificación: Asigna categorías basadas en patrones.</p> <p>Regresión: Predice valores continuos como ingresos o demanda.</p> <p>Clustering: Agrupa datos según similitudes.</p> <p>Asociación: Identifica relaciones entre elementos, como productos comprados juntos.</p> <p>Análisis de series temporales: Estudia patrones a lo largo del tiempo.</p> <p>Detección de anomalías: Encuentra datos fuera de lo esperado, útil para detectar fraudes.</p> <p>Reducción de dimensionalidad: Simplifica conjuntos de datos eliminando atributos irrelevantes.</p>
Amazon web services	<p>Como jefe de proyecto, usaría Amazon Web Services (AWS) para minería de datos por las siguientes razones:</p> <p>Escalabilidad infinita: Permite ajustar recursos bajo demanda según las necesidades del proyecto.</p> <p>Amplia gama de servicios: Ofrece herramientas como Amazon SageMaker, Amazon Redshift y Amazon EMR para análisis de datos.</p> <p>Costo eficiente: Modelo de pago por uso que optimiza los costos según el consumo.</p> <p>Seguridad y cumplimiento: Garantiza altos niveles de seguridad y cumplimiento de normativas internacionales.</p> <p>Disponibilidad global: Infraestructura en múltiples regiones que asegura alta disponibilidad y tiempos de respuesta rápidos.</p> <p>Integración con IA: Servicios de inteligencia artificial que facilitan el análisis y la detección de patrones.</p> <p>Facilidad de uso: Ecosistema amigable y bien documentado que acelera el desarrollo del proyecto.</p>	<p>Clasificación: Asignar categorías a datos.</p> <p>Regresión: Predecir valores continuos.</p> <p>Clustering: Agrupar datos similares.</p> <p>Asociación: Identificar relaciones entre variables.</p> <p>Ánálisis de series temporales: Estudiar datos a lo largo del tiempo para detectar tendencias.</p> <p>Detección de anomalías: Encontrar datos que se desvian de patrones esperados.</p> <p>Reducción de dimensionalidad: Simplificar conjuntos de datos eliminando atributos irrelevantes.</p> <p>Optimización: Encontrar soluciones óptimas en problemas complejos.</p>
Azure	<p>Como jefe del proyecto yo usaría Azure por estas razones:</p> <p>Integración en la nube: Facilita el acceso a datos y servicios de análisis.</p> <p>Herramientas de machine learning: Proporciona algoritmos para construir y desplegar modelos fácilmente.</p> <p>Escalabilidad: Permite ajustar recursos según las necesidades del proyecto.</p> <p>Ánalisis en tiempo real: Permite decisiones rápidas con Azure Stream Analytics.</p> <p>Seguridad y cumplimiento: Garantiza la protección de datos sensibles y el cumplimiento de regulaciones.</p> <p>Integración con otras herramientas: Se conecta con servicios de Microsoft y terceros para un flujo de trabajo ágil.</p> <p>Facilidad de uso: Interfaz intuitiva y plantillas que simplifican la creación de modelos.</p> <p>Capacidades de visualización: Herramientas como Power BI para presentar resultados de forma clara.</p>	<p>Clasificación: Asignar etiquetas a datos.</p> <p>Regresión: Predecir valores continuos.</p> <p>Clustering: Agrupar datos similares.</p> <p>Ánálisis de asociación: Identificar relaciones entre variables.</p> <p>Ánálisis de series temporales: Estudiar datos a lo largo del tiempo para detectar tendencias.</p> <p>Detección de anomalías: Encontrar datos que se desvian de patrones esperados.</p> <p>Reducción de dimensionalidad: Simplificar conjuntos de datos eliminando atributos irrelevantes.</p> <p>Optimización: Encontrar soluciones óptimas en problemas complejos.</p>

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

Yo como jefe usaría Amazon Web Services (AWS) para el proyecto de minería de datos debido a su amplia gama de servicios diseñados para análisis y machine learning, como Amazon SageMaker y Redshift. La escalabilidad de AWS permite ajustar recursos fácilmente según las necesidades del proyecto, lo que optimiza costos con su modelo de pago por uso. Además, ofrece alta disponibilidad y confiabilidad, garantizando el acceso continuo a nuestros datos. Su robusta seguridad protege información sensible y cumple con regulaciones. La capacidad de realizar análisis en tiempo real con herramientas como Kinesis facilita la toma de decisiones rápida y fundamentada, haciendo de AWS una opción ideal para nuestras necesidades. Aunque claro ay algunos puntos a tomar a cuenta y que debemos de fortalecer para poder obtener buenos resultados.

DESVENTAJAS

Complejidad de la plataforma: La variedad de servicios puede ser abrumadora, aumentando la curva de aprendizaje para los usuarios sin experiencia.

Costos acumulativos: Aunque es flexible, el modelo de pago por uso puede resultar en costos inesperados si no se gestionan adecuadamente.

Dependencia de la conectividad a Internet: Requiere una conexión constante, lo que puede ser problemático en áreas con acceso limitado.

Gestión de seguridad y cumplimiento: La responsabilidad compartida en seguridad implica que la organización debe mantenerse vigilante para cumplir con normativas.

Curva de aprendizaje: La capacitación del personal puede llevar tiempo y recursos, retrasando la implementación efectiva del proyecto.

Perfiles y Roles

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

- **Director del Departamento**
- **Project Manager (IT, Negocios)**
- **Product Manager**
- **Arquitecto de Soluciones (Ingenieros en Software)**
- **Ingeniero de Redes (DevOps)**
- **Científico de Datos (Matemáticos, Físicos)**
- **Analistas de Datos (x2)**
- **Desarrollador Frontend**
- **Desarrollador Backend**
- **Full Stack Develop**

¿Por qué son fundamentales?

- **Director del Departamento:** Es fundamental para asegurar que el proyecto esté alineado con los objetivos generales de la empresa y para tomar decisiones estratégicas.
- **Project Manager:** Coordina todos los esfuerzos del equipo y gestiona los recursos para que el proyecto sea ejecutado en tiempo y forma.
- **Product Manager:** Asegura que el producto resultante de la minería de datos sea útil y cumpla con las expectativas del cliente o usuario final.
- **Arquitecto de Soluciones:** Garantiza que la estructura técnica sea adecuada y escalable para manejar grandes volúmenes de datos.
- **Ingeniero de Redes:** Es esencial para garantizar que la infraestructura necesaria para el análisis de datos funcione de manera óptima.
- **Científico de Datos:** Responsable de extraer valor a partir de los datos a través de técnicas matemáticas y estadísticas avanzadas.
- **Analistas de Datos:** Son cruciales para transformar datos sin procesar en conjuntos que puedan ser usados por el científico de datos.
- **Desarrolladores (Frontend, Backend, Full Stack):** Son indispensables para crear la infraestructura digital que permitirá interactuar con los resultados de la minería de datos.

¿Cuánto personal se va a contratar?

- **Director del Departamento:** 1
- **Project Manager:** 1
- **Product Manager:** 1
- **Arquitecto de Soluciones:** 1
- **Ingeniero de Redes (DevOps):** 1
- **Científico de Datos:** 1
- **Analistas de Datos:** 2
- **Desarrollador Frontend:** 1
- **Desarrollador Backend:** 1
- **Full Stack Developer:** 1

Propuesta de Software: Amazon Web Services (AWS)

Motivo de selección:

AWS es una de las plataformas de servicios en la nube más robustas y ampliamente utilizadas en el mercado. Su selección para este proyecto de minería de datos se basa en los siguientes puntos:

1. **Escalabilidad:** AWS ofrece infraestructura elástica que permite escalar los recursos según el crecimiento del proyecto.
2. **Seguridad:** Proporciona un amplio conjunto de herramientas de seguridad y cifrado de datos, cumpliendo con estándares internacionales.
3. **Variedad de Servicios:** Incluye una amplia gama de servicios relacionados con almacenamiento de datos, análisis y machine learning, tales como Amazon S3, Redshift, AWS Glue y SageMaker.
4. **Costo Eficiente:** Su modelo de pago por uso permite optimizar el presupuesto del proyecto, ya que solo se paga por lo que se consume.
5. **Compatibilidad con Herramientas de Minería de Datos:** AWS se integra fácilmente con herramientas de minería de datos como Apache Spark, Hadoop, Jupyter Notebooks y más.
6. **Disponibilidad Global:** Gracias a su infraestructura distribuida, AWS garantiza alta disponibilidad y baja latencia para la ejecución del proyecto.

Propuesta del Personal Necesario:

El organigrama presentado establece los siguientes roles y personal para el proyecto:

1. **Director del Departamento:** Juan supervisará y garantizará que el proyecto esté alineado con

- los objetivos estratégicos.
2. **Project Manager (IT, Negocios)**: Se encargará de la gestión integral del proyecto, supervisando el cumplimiento de plazos y objetivos.
 3. **Product Manager**: Definirá las características del producto y priorizará las funcionalidades.
 4. **Arquitecto de Soluciones (Ingeniero en Software)**: Diseñará la arquitectura técnica del sistema, asegurando la integración de AWS con las soluciones internas.
 5. **Ingeniero de Redes (DevOps)**: Se encargará de configurar y mantener la infraestructura necesaria en AWS.
 6. **Científico de Datos**: Desarrollará los algoritmos y modelos de minería de datos utilizando los servicios de AWS.
 7. **Analistas de Datos (x2)**: Ayudarán en la limpieza y preparación de los datos.
 8. **Desarrolladores (Frontend, Backend, Full Stack)**: Crearán las interfaces necesarias y configurarán las bases de datos para interactuar con los resultados.

Requisitos del personal:

- **Conocimiento en AWS**: El personal técnico debe tener experiencia con la plataforma de AWS y certificaciones como AWS Certified Solutions Architect o AWS Certified Big Data.
- **Experiencia en Minería de Datos y Ciencia de Datos**: Deben estar familiarizados con herramientas como Python, R, Spark y otros frameworks de análisis de datos.
- **Habilidades en Desarrollo Web y API**: Para garantizar la correcta interacción entre la plataforma de análisis y las interfaces de usuario.

Proceso del proyecto

Ejecución del proyecto

Para garantizar el éxito del proyecto, se seguirán los siguientes pasos:

1. **Fase de Planificación:**
 - a. **Definición de Objetivos y KPIs**: Juan y su equipo definirán los objetivos de negocio específicos que se deben alcanzar con la minería de datos.

- b. **Análisis de Requisitos:** El Project Manager y el Arquitecto de Soluciones identificarán los requisitos técnicos y del negocio para la infraestructura de AWS.
- c. **Estimación de Costos:** Se desarrollará un presupuesto estimado basado en el uso de AWS y el costo de personal necesario.

2. Fase de Diseño:

- a. **Diseño de Arquitectura Técnica:** El Arquitecto de Soluciones diseñará la arquitectura en la nube, especificando los servicios de AWS que se utilizarán, como S3 para almacenamiento, Redshift para el análisis de datos y SageMaker para machine learning.
- b. **Plan de Desarrollo:** Se definirá el cronograma de desarrollo, dividiendo las tareas entre los equipos de backend, frontend y ciencia de datos.

3. Fase de Implementación:

- a. **Desarrollo e Implementación Técnica:** Los ingenieros y desarrolladores comenzarán a implementar la infraestructura en AWS, configurando bases de datos, integrando herramientas de análisis y desarrollando las interfaces necesarias.
- b. **Despliegue de Modelos de Minería de Datos:** El Científico de Datos implementará modelos predictivos utilizando SageMaker y herramientas de machine learning.
- c. **Pruebas Unitarias y de Integración:** Cada componente desarrollado será sometido a pruebas para asegurar su correcto funcionamiento.

4. Fase de Evaluación:

- a. **Pruebas Piloto:** Se realizarán pruebas piloto con datos reales para verificar el rendimiento del sistema.
- b. **Evaluación de Resultados:** El equipo analizará los resultados obtenidos para validar la efectividad de los modelos de minería de datos.
- c. **Optimización:** Se harán ajustes y optimizaciones en los modelos y la infraestructura según sea necesario.

5. Fase de Despliegue Final:

- a. **Lanzamiento a Producción:** Una vez probada la solución, se procederá al despliegue final, haciendo que el sistema esté disponible para su uso en producción.
- b. **Capacitación del Personal:** Se capacitará al personal operativo sobre cómo usar las herramientas y sistemas desarrollados.
- c. **Monitoreo y Soporte:** El Ingeniero de Redes y DevOps se encargará de monitorear el rendimiento del sistema y de realizar tareas de mantenimiento cuando sea necesario.

6. Fase de Mantenimiento y Optimización Continua:

- a. **Monitoreo de KPIs:** El Project Manager y el Product Manager revisarán periódicamente el desempeño del sistema basado en los KPIs definidos al inicio del proyecto.
- b. **Mejoras Incrementales:** Dependiendo del crecimiento de los datos y los resultados obtenidos, el equipo de desarrollo y los científicos de datos realizarán mejoras continuas.

Conclusión

En conclusión, realizar esta actividad sobre minería de datos me ha ayudado a comprender mejor cómo funcionan las herramientas tecnológicas y los roles necesarios para aprovechar los datos en las empresas. Aunque solo fue una actividad, aprendí que la minería de datos es algo que realmente puede ayudar a tomar decisiones más inteligentes y eficientes. Investigar sobre software como AWS, Oracle SQL y Azure me mostró lo importante que es escoger las herramientas adecuadas y tener un equipo especializado. Esta experiencia me hizo ver cómo estas tecnologías pueden transformar la manera en que las empresas trabajan y toman decisiones, lo que será cada vez más importante en el futuro.

Referencias

El material de apoyo.

ChatGpt.

Amazon Web Services, Inc. (2023). *What is Cloud Computing?* Retrieved from <https://aws.amazon.com/what-is-cloud-computing/>

Amazon Web Services, Inc. (2022). *Amazon Web Services: Overview of Security Processes* (Whitepaper). Retrieved from https://d1.awsstatic.com/whitepapers/Security/AWS_Security_Whitepaper.pdf

Oracle Corporation. (2023). *Oracle Database Documentation*. Retrieved from <https://docs.oracle.com/en/database/oracle/oracle-database/>

Price, J. (2021). *Oracle Database 19c SQL: A Beginner's Guide*. McGraw-Hill Education.

Microsoft Corporation. (2023). *What is Azure?* Retrieved from <https://azure.microsoft.com/en-us/overview/what-is-azure/>

Microsoft Corporation. (2022). *Azure Data Fundamentals*. Microsoft Press.

