

Actividad 2. Software, Personal y Procesos

Minería de datos

Ingeniería en Desarrollo de Software

Tutor: Félix Acosta Hernández

Alumno: José Luis Pacheco González

Fecha: 20 de junio 2024

Índice

Introducción	3
Descripción	4
Justificación	5
Software de Data Mining	6
Perfiles y roles.....	8
Proceso del proyecto	14
Conclusión	15
Referencias.....	16

Introducción

En la era de la información, el data mining se ha convertido en una herramienta esencial para las organizaciones que buscan transformar datos crudos en conocimientos valiosos. Esta actividad se centra en la selección del software adecuado para realizar data mining, así como en la identificación de los roles y perfiles necesarios y la planificación del proceso de desarrollo del proyecto. Elegir el software correcto es crucial, ya que debe alinearse con los objetivos del proyecto, la naturaleza de los datos y las capacidades técnicas del equipo. Además, es fundamental definir los roles y perfiles clave, como el del científico de datos, el analista de datos, el ingeniero de datos y el administrador de proyectos, cada uno con habilidades específicas que contribuyen al éxito del proyecto. A lo largo del proceso de desarrollo, se seguirán etapas estructuradas que incluyen la identificación del problema, la recolección y preparación de datos, la selección de modelos, la evaluación y validación de resultados, y la implementación de las soluciones. Esta actividad no solo proporcionará una comprensión profunda de las herramientas y técnicas de data mining, sino que también destacará la importancia de la colaboración interdisciplinaria y la gestión efectiva del proyecto para alcanzar los objetivos establecidos.

Descripción

Juan ha sido designado como el nuevo director del área de Inteligencia de Negocios y Minería de Datos Nacional, con la misión de diseñar e implementar un proyecto innovador para su departamento. Su primera tarea es presentar una propuesta de proyecto que incluya la selección de tres programas de data mining, siguiendo las recomendaciones del curso. Estos programas deben ser evaluados según su capacidad para satisfacer las necesidades del proyecto y su compatibilidad con la infraestructura tecnológica actual. Además, Juan debe identificar los roles y perfiles esenciales para el desarrollo del proyecto, incluyendo científicos de datos, analistas de datos, ingenieros de datos y gestores de proyectos, cada uno con responsabilidades específicas clave para el éxito del proyecto. Esta planificación detallada garantizará que el proyecto esté bien fundamentado desde el principio, facilitando una implementación eficiente y alineada con los objetivos estratégicos del área de Inteligencia de Negocios y Minería de Datos Nacional. La estructura propuesta asegurará una integración efectiva de las herramientas y técnicas de data mining, promoviendo la colaboración interdisciplinaria y una gestión de proyecto eficaz.

Justificación

La selección del software adecuado para realizar data mining es crucial para maximizar el valor de los datos y alcanzar los objetivos del proyecto. Optar por herramientas que ofrezcan robustez, flexibilidad y compatibilidad con la infraestructura existente asegura un análisis eficiente y preciso. Además, es fundamental seleccionar programas con una buena relación costo-beneficio y soporte técnico confiable. En cuanto a los roles y perfiles, contar con científicos de datos, analistas de datos, ingenieros de datos y gestores de proyectos es esencial. Los científicos de datos diseñan modelos predictivos y algoritmos avanzados; los analistas de datos interpretan los resultados para tomar decisiones estratégicas; los ingenieros de datos gestionan y optimizan las infraestructuras de datos; y los gestores de proyectos coordinan el trabajo del equipo y aseguran el cumplimiento de plazos y objetivos. El proceso de desarrollo del proyecto debe seguir una metodología estructurada que incluya la identificación de objetivos, recolección y preparación de datos, selección y evaluación de modelos, y finalmente la implementación y monitoreo de las soluciones. Este enfoque sistemático no solo facilita la colaboración interdisciplinaria, sino que también garantiza una gestión eficiente del proyecto, minimizando riesgos y optimizando los recursos para obtener resultados efectivos y alineados con las metas de la organización.

Software de Data Mining

Tecnología de Minería de datos	¿Por qué propondrías ese software?	¿Qué procesos de minería de datos puede realizar el software?
SQL server	por su escalabilidad, rendimiento, seguridad, integración con herramientas de Microsoft, funcionalidades avanzadas, y sólida comunidad de soporte, lo que facilita la gestión de bases de datos y desarrollo de aplicaciones empresariales.	puede realizar clasificación, clustering, regresión, asociación, análisis de series temporales, detección de anomalías y análisis predictivo, ayudando a las organizaciones a identificar patrones y tomar decisiones informadas basadas en datos.
Tableau	por su facilidad de uso, integración con múltiples fuentes de datos, visualizaciones interactivas, análisis en tiempo real y sólidas opciones de colaboración y soporte.	puede realizar exploración y preparación de datos, análisis descriptivo, análisis predictivo (con R y Python), segmentación, análisis de series temporales, y creación de dashboards interactivos.
Python	es ideal para análisis y minería de datos por su popularidad, sintaxis clara, y amplia gama de	permite realizar desde la limpieza y visualización de datos hasta la aplicación de algoritmos avanzados como clustering, clasificación, regresión, y detección de anomalías, utilizando

	bibliotecas especializadas como pandas y scikit-learn, que facilitan la integración y el manejo eficiente de grandes volúmenes de datos.	bibliotecas como numpy y matplotlib, y herramientas de automatización como PyCaret para optimizar el proceso de análisis de datos. 3.5
--	--	---

Perfiles y roles

Gerente de Proyecto (Project Manager)

Funciones: Liderar el proyecto, coordinar al equipo, gestionar los recursos y tiempos, y asegurar la alineación con los objetivos estratégicos de la organización.

Analista de Negocios (Business Analyst)

Funciones: Recopilar y analizar los requisitos del negocio, traducir las necesidades empresariales en especificaciones técnicas, y asegurar que las soluciones desarrolladas cumplan con los objetivos comerciales.

Ingeniero de Datos (Data Engineer)

Funciones: Diseñar, construir y mantener la infraestructura de datos, desarrollar pipelines de datos, y asegurar la calidad y accesibilidad de los datos.

Científico de Datos (Data Scientist)

Funciones: Analizar y modelar datos, desarrollar algoritmos y modelos predictivos, y generar insights accionables para el negocio.

Analista de Datos (Data Analyst)

Funciones: Realizar análisis descriptivo de los datos, generar reportes y dashboards, y apoyar en la toma de decisiones basada en datos.

Especialista en Visualización de Datos (Data Visualization Specialist)

Funciones: Crear visualizaciones efectivas y atractivas, diseñar dashboards interactivos, y comunicar los insights de manera clara y comprensible.

Desarrollador de BI (BI Developer)

Funciones: Desarrollar y mantener soluciones de inteligencia de negocios, incluyendo ETL, reportes, y dashboards.

Administrador de Bases de Datos (Database Administrator)

Funciones: Gestionar las bases de datos, asegurar su rendimiento, seguridad y disponibilidad, y realizar tareas de mantenimiento y optimización.

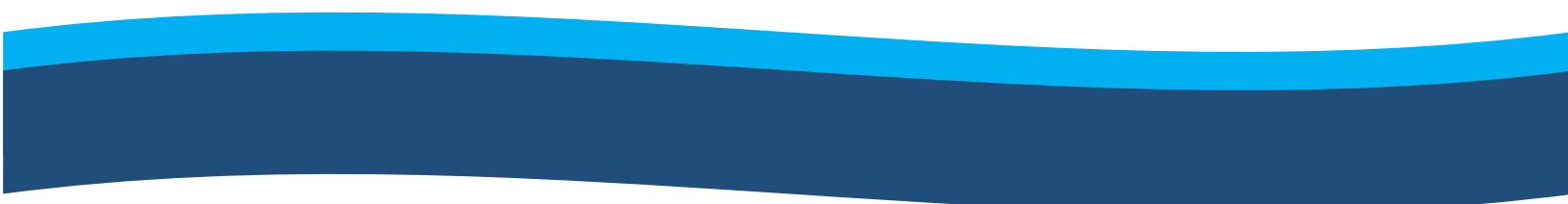


Arquitecto de Soluciones (Solutions Architect)

Funciones: Diseñar la arquitectura general del sistema, asegurar la integración de las distintas tecnologías, y definir las mejores prácticas y estándares.

Especialista en Seguridad de Datos (Data Security Specialist)

Funciones: Asegurar la protección de los datos, implementar políticas y procedimientos de seguridad, y cumplir con las regulaciones y normativas de privacidad.



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

- Gerente de Proyecto
- Analista de Negocios
- Ingeniero de Datos
- Científico de Datos
- Analista de Datos
- Especialista en Visualización de Datos
- Desarrollador de BI
- Administrador de Bases de Datos
- Arquitecto de Soluciones
- Especialista en Seguridad de Datos

• ¿Por qué son fundamentales?

Por qué proporcionan una cobertura completa de todas las áreas críticas para lograr el éxito del proyecto de inteligencia de negocios y minería de datos, asegurando que todos los aspectos técnicos, de análisis, de seguridad y de gestión estén adecuadamente atendidos y cubiertos.

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

Sumando los perfiles y el número de personas necesarias para cada rol, se tiene un total de 20 personas.

1 Gerente de Proyecto.

2 Analistas de Negocios.

3 Ingenieros de Datos.

3 Científicos de Datos.

3 Analistas de Datos.

2 Especialistas en Visualización de Datos.

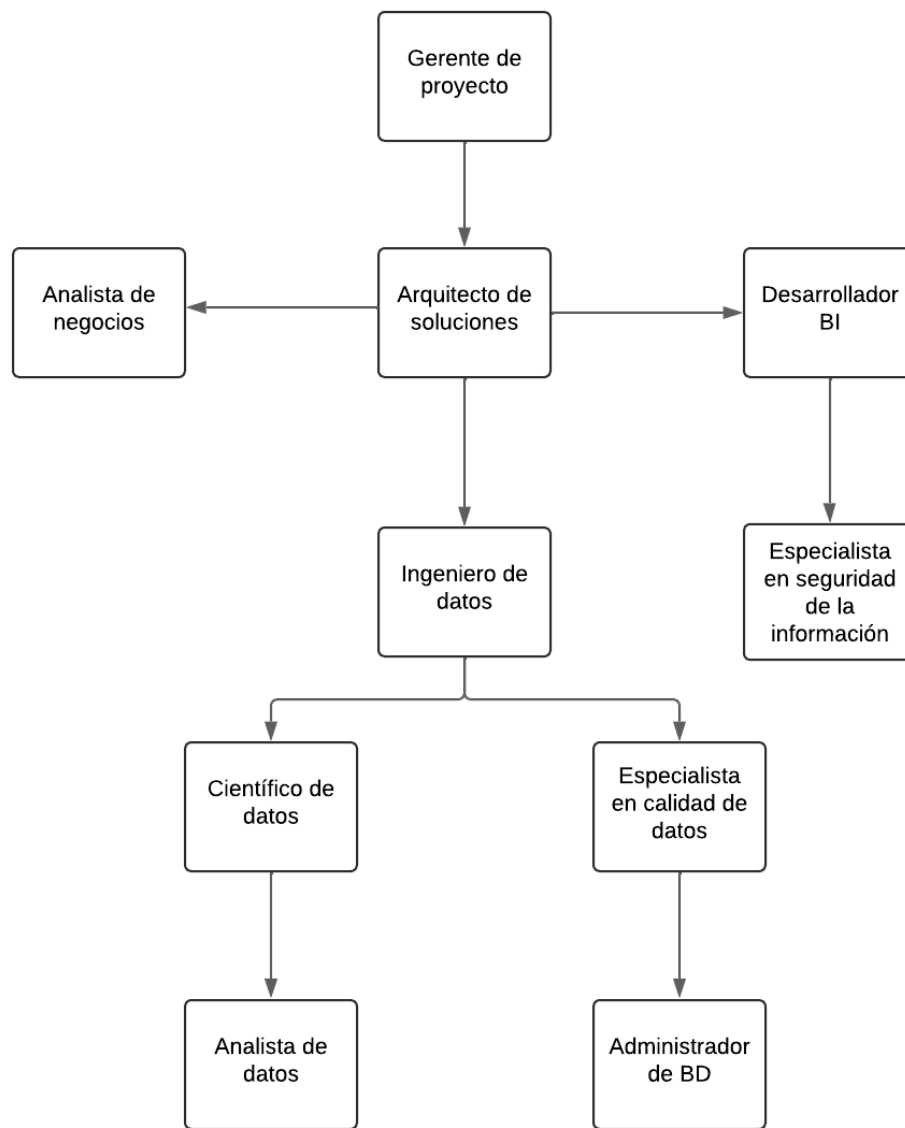
2 Desarrolladores de BI.

2 Administradores de Bases de Datos.

1 Arquitecto de Soluciones.

1 Especialista en Seguridad de Datos.

A continuación, se muestra el organigrama del proyecto.



Proceso del proyecto

Aquí se muestra el cronograma del proceso del proyecto, donde se detallan las actividades a realizar desde el inicio hasta el final del proyecto.

Aa Nombre	Asignar	Estado	Fecha
Definir objetivos del proyecto y asignación de roles.		Listo	@1 de junio de 2024 → 8 de junio de 2024
Establecer alcance, cronograma y presupuesto detallado		Listo	@9 de junio de 2024 → 16 de junio de 2024
Elaborar el plan de proyecto con análisis de riesgos		En curso	@17 de junio de 2024 → 24 de junio de 2024
Revisión y aprobación del plan por parte de las partes interesadas.		Sin empezar	@25 de junio de 2024 → 2 de julio de 2024
Implementación de ETL y configuración de almacenamiento		Sin empezar	@3 de julio de 2024 → 10 de julio de 2024
Desarrollo y configuración de pipelines de datos		Sin empezar	@11 de julio de 2024 → 18 de julio de 2024
Configuración inicial de la plataforma de BI y herramientas de análisis.		Sin empezar	@19 de julio de 2024 → 26 de julio de 2024
Inicio de la captura y procesamiento de datos		Sin empezar	@27 de julio de 2024 → 3 de agosto de 2024
Desarrollo de modelos y algoritmos de análisis		Sin empezar	@4 de agosto de 2024 → 11 de agosto de 2024
Creación de dashboards y visualizaciones preliminares		Sin empezar	@12 de agosto de 2024 → 19 de agosto de 2024
Pruebas y validación de sistemas y modelos implementados		Sin empezar	@20 de agosto de 2024 → 27 de agosto de 2024
Implementación piloto de soluciones de BI y minería de datos		Sin empezar	@28 de agosto de 2024 → 4 de septiembre de 2024
Capacitación del personal en nuevas herramientas y plataformas		Sin empezar	@5 de septiembre de 2024 → 12 de septiembre de 2024
Ajuste de procesos y procedimientos según resultados piloto		Sin empezar	@13 de septiembre de 2024 → 20 de septiembre de 2024
Evaluación de los resultados obtenidos frente a los objetivos		Sin empezar	@21 de septiembre de 2024 → 28 de septiembre de 2024
Documentación final del proyecto y preparación de informes		Sin empezar	@28 de septiembre de 2024 → 5 de octubre de 2024
Identificación de lecciones aprendidas y recomendaciones		Sin empezar	@6 de octubre de 2024 → 13 de octubre de 2024
Presentación de resultados y cierre formal del proyecto		Sin empezar	@14 de octubre de 2024 → 18 de octubre de 2024

Conclusión

En conclusión, la selección cuidadosa del software de data mining y la identificación precisa de los roles y perfiles necesarios son fundamentales para el éxito de un proyecto en el área de Inteligencia de Negocios y Minería de Datos. Elegir herramientas robustas y compatibles asegura que los análisis sean eficientes y precisos, mientras que la selección de personal con habilidades específicas en ciencia de datos, análisis de datos, ingeniería de datos y gestión de proyectos garantiza una ejecución fluida y efectiva del proyecto. Un proceso de desarrollo bien estructurado, que incluya etapas claras desde la identificación de objetivos hasta la implementación y monitoreo de soluciones, facilita la colaboración interdisciplinaria y la gestión eficiente de recursos. Este enfoque no solo minimiza los riesgos asociados con la implementación del proyecto, sino que también optimiza el uso de los recursos disponibles, asegurando que los resultados obtenidos estén alineados con las metas estratégicas de la organización. En última instancia, una planificación meticulosa y una ejecución bien coordinada permiten transformar los datos en conocimientos valiosos, potenciando la capacidad de la organización para tomar decisiones informadas y estratégicas, lo que resulta en una ventaja competitiva significativa en el mercado.

Referencias

Herramientas de minería de datos. (n.d.). Microsoft.com. Retrieved June 21, 2024, from

<https://learn.microsoft.com/es-es/analysis-services/data-mining/data-mining-tools?view=asallproducts-allversions>

Roles principales en proyectos de integración de datos. (2022, October 19). Datapeaker;

[datapeaker.com. https://datapeaker.com/data-warehouse/roles-principales-en-proyectos-de-integracion-de-datos/](https://datapeaker.com/data-warehouse/roles-principales-en-proyectos-de-integracion-de-datos/)

(N.d.). Sap.com. Retrieved June 21, 2024, from

<https://www.sap.com/latinamerica/products/technology-platform/hana/what-is-data-mining.html>

Enlace Github