

Actividad | 2 | Software, personal y procesos.

Minería de datos.

Ingeniería en Desarrollo de
Software.



academi**ag**lobal

TUTOR: Félix Acosta Hernández

ALUMNO: Carlos Ariel Nicolini

FECHA: 15/10/2024

Índice

Introducción	3
Descripción	4
Justificación	5
Desarrollo.....	6
• Software de Data Mining.....	6
• Perfiles y roles.....	8
• Proceso del proyecto	12
Conclusión.....	14
Referencias.....	15

Introducción

La minería de datos o data mining es un proceso técnico, automático o semiautomático, que analiza grandes cantidades de información dispersa para darle sentido y convertirla en conocimiento. Busca anomalías, patrones o correlaciones entre millones de registros para predecir resultados, como indica el SAS institute, uno de los referentes mundiales en analítica de negocios.

Mientras tanto la información crece sin parar. Un estudio de 2017 sobre big data revela que el 90% de los datos del mundo posteriores a 2014 y su volumen se duplica cada 1.2 años. En este contexto, el data mining es una práctica estratégica considerada importante por casi el 80% de las organizaciones que aplican inteligencia empresarial (business intelligence), según Forbes.

Gracias a la acción conjunta de analítica y minería de datos, que combina estadística, Inteligencia Artificial y aprendizaje automático, las empresas pueden crear modelos para descubrir conexiones entre millones de registros. El data mining posibilita, entre otros aspectos:

- Limpiar datos de ruido y repeticiones.
- Extraer la información relevante y utilizarla para evaluar posibles resultados.
- Tomar mejores decisiones de negocio con mayor rapidez.

El data mining está presente en la industria actual, estos son algunos ejemplos:

- Marketing
- Comercio minorista
- Banca
- Medicina
- Televisión y radio

Descripción

Contextualización:

Juan ha sido nombrado como el nuevo director del área de inteligencia de negocios y minería de datos nacional. Por lo tanto, se le solicita un proyecto para diseñar e implementar en el área, como primer punto se pretende identificar los requisitos básicos como son el software necesario y el requisito del personal, posterior a esto en la actividad 3, Juan debe estimar los costos para la implementación de la propuesta del proyecto.

Actividad:

- Presentar una propuesta del proyecto
- Seleccionar 3 programas de Data Mining (considerar los que se mencionaron en el curso)
- Identificar los roles y/o perfiles requeridos para el desarrollo del proyecto.

Justificación

En esta actividad realizaremos un proyecto para diseñar e implementar Data Mining, en el cual tendremos que seleccionar una herramienta para tal fin, los requisitos básicos y el personal que será necesario para dicha implementación, además de como desarrollaremos dicho proyecto.

Esta clase me resulto muy productiva pero no por eso un poco más difícil, ya que estoy acostumbrado más a lo técnico, me siento más seguro realizando por decirlo de una manera trabajo más manual (como configuraciones en programas), pero noto el gran potencial que representa dicho proyecto, ya que me permite aprender mucho sobre cómo llevar adelante un proyecto, que cosas deben tenerse en cuenta, desde la herramienta a utilizar, el motor de base de datos, cuanto personal se necesita etc, es diseñar todo antes de ponerlo a ambientar.

En el ámbito laboral es muy importante contar con estas habilidades para poder seguir escalando en puestos y poder tener más oportunidades laborales con mejores beneficios.

Este trabajo fue subido al siguiente enlace de GitHub

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

Desarrollo

Software de Data Mining

Para área de inteligencia de Negocios y Minería de datos Nacional necesitamos tener 3 propuestas de herramientas de Data Mining para el área de inteligencia de Negocios y Minería de datos Nacional, para lo cual se tiene las siguientes opciones, donde se mencionan sus puntos fuertes y que procesos de minería podríamos realizar con dichas herramientas, tras lo cual debemos elegir la que mejor nos parezca. En el siguiente grafico presentamos dicha información:

Tecnología de Minería de datos	¿Por qué propondrías ese software?	¿Qué procesos de minería de datos puede realizar el software?
Microsoft Azure Machine Learning	<ul style="list-style-type: none"> * Implementación rápida * Seguridad y cumplimiento * Integración de servicios * Automatización * Escalabilidad * Herramientas visuales * Monitoreo y mantenimiento 	<ul style="list-style-type: none"> * Exploración de datos * Limpieza y Preprocesamiento de datos * Análisis predictivo, de agrupamiento, de redes neuronales y de series temporales. * Optimización y visualización de modelos * Implementación y despliegue
Oracle Data Mining (ODM)	<ul style="list-style-type: none"> * Integración directa en la base de datos * Amplia gama de algoritmos * Automatización de procesos * Manejo de grandes volúmenes * Trazabilidad * Soporte para procesos de negocio 	<ul style="list-style-type: none"> * Exploración de datos * Limpieza y preparación de datos * Clasificación * Análisis de asociación. * Optimización y visualización de resultados * Detección de anomalías * Validación, evaluación e implementación de modelos
SAP Data Intelligence	<ul style="list-style-type: none"> * Integración de datos * Orquestación de flujo de datos * Interfaz visual * Automatización de procesos * Modelo de datos unificado * Funciones de Machine Learning y AI * Escalabilidad * Cumplimiento y seguridad 	<ul style="list-style-type: none"> * Preparación de datos * Exploración de datos * Modelo predictivo * Evaluación, validación e implementación de modelos * Monitoreo y Mantenimiento * Automatización de flujos de trabajo

Para nuestro proyecto se seleccionó Microsoft Azure Machine Learning.

Gestor de base de datos:

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

El gestor de base de datos elegido es el Microsoft SQL server, el cual tiene las siguientes ventajas:

- Proporciona capacidades robustas para minería de datos y análisis predictivo, con variedad de algoritmos.
- Interfaz amigable y de fácil uso.
- Integración con Power Bi, el cual facilita la creación de dashboards interactivos y visualizaciones.
- Costo de licenciamiento más accesible en comparación con otros productos.
- Cuenta con herramientas de desarrollo y gestión que simplifican la implementación y mantenimiento.

Perfiles y roles

A continuación mencionaremos los tipos de perfiles y roles necesario para nuestro proyecto de Data Mining, los cuales mencionaremos y explicaremos a continuación:

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

1- Director del departamento:

- Responsabilidades: Definición y alineación de objetivos, planificación del proyecto, gestión de equipos y roles, supervisión de procesos técnicos, control del presupuesto, cumplimiento de normativas y seguridad de datos.
- Habilidades: Comunicación y liderazgo, experiencia en gestión de proyectos, conocimiento técnico en Data Mining y Machine Learning, capacidad de análisis de negocio, conocimiento de seguridad y privacidad de datos

2- Project Manager:

- Responsabilidades: Coordina y gestiona el proyecto, establece plazos, asegura la comunicación entre los equipos y administra los recursos.
- Habilidades: Conocimiento en gestión de proyectos, capacidad de liderazgo, comprensión de minería de datos y análisis de datos.

3- Científico de datos:

- Responsabilidades: Diseña, implementa y evalúa modelos de minería de datos y machine learning. Es responsable de la selección de algoritmos, la optimización de modelos y la interpretación de resultados.
- Habilidades: Dominio del lenguaje de programación como Python o R, habilidades estadísticas, experiencia en machine learning y conocimientos sobre técnicas de minería de datos.

4- Ingeniero de datos:

- 1- Responsabilidades: Crea y gestiona la infraestructura de datos, asegura la disponibilidad y calidad de los datos, optimiza el flujo de datos desde diversas fuentes hasta su procesamiento.
- 2- Habilidades: Conocimiento en base de datos (SQL y NoSQL), ETL (extracción, transformación y carga de datos), experiencia en Big Data y conocimiento de almacenamiento en la nube.

5- Analista de datos:

- Responsabilidades: Realiza el análisis de los datos, visualiza patrones y tendencias, colabora en la preparación de los datos.
- Habilidades: Conocimiento en herramientas de análisis (Excel, PowerBI), dominio de SQL, habilidades en estadística y en lenguajes como de scripting como Python o R para análisis básicos y limpieza de datos.

6- Arquitecto de datos:

- Responsabilidades: Diseña la arquitectura de datos del proyecto, asegurando que la estructura soporte tanto el flujo de datos como las necesidades analíticas y de escalabilidad.
- Habilidades: Dominio en diseño de base de datos y estructuras de almacenamiento de datos, conocimiento en sistemas de gestión de base de datos y familiaridad con tecnologías de Big Data y arquitectura en la nube.

7- Especialista en seguridad de datos:

- Responsabilidades: Asegura que se cumplan los estándares, regulaciones de seguridad y privacidad en el manejo de los datos, implementando medidas para proteger los datos en todo el proceso.
- Habilidades: Conocimiento en normas de seguridad de datos, experiencia en herramientas de encriptación y anonimización de datos, habilidades en seguridad informática.

8- Administrador de base de datos:

- Responsabilidades: Gestiona las bases de datos asegurando su rendimiento, disponibilidad y seguridad.
- Habilidades: Experiencia en administración de base de datos (SQL, Oracle, MongoDB), habilidades en optimización de rendimiento y conocimiento en seguridad de base de datos.

¿Por qué son fundamentales?

Contar con el personal con el perfil necesario es fundamental en este y cualquier proyecto de Data Mining para asegurar su efectividad y éxito. Este proyecto de Data Mining involucra distintas fases que requieren habilidades especializadas para manejar grandes volúmenes de datos, transformarlos en información útil y después traducirla en resultados para la toma de decisiones.

¿Cuánto personal se va a contratar?

En base a los roles seleccionados, se contratará

- Director del departamento: 1
- Project Manager: 1
- Científico de datos: 1
- Ingeniero de datos: 1
- Analista de datos: 3
- Arquitecto de datos: 1
- Especialista en seguridad de datos: 1
- Administrador de base de datos: 2

Desarrollo del proyecto

Para la correcta implementación del proyecto, es necesario una planificación estructurada, una ejecución organizada y una evaluación continua. Las etapas a seguir son:

- 1- **Definir los objetivos del proyecto:** Determinar los alcances del proyecto, especificar los indicadores de rendimiento y definir qué éxito se quiere lograr en términos de métricas.
- 2- **Selección y contratación del equipo:** Realizar las entrevistas, revisar que los candidatos sean los idóneos para los perfiles solicitados, contratación del personal y proporcionarles los medios para que puedan desarrollar sus actividades.
- 3- **Configurar el entorno de Azure Machine Learning:** Crear en el portal de azure un nuevo workspace de AML que servirá como el entorno de desarrollo para el proyecto. Configurar un clúster de computación para entrenar modelos en la nube.
- 4- **Preparar y explorar datos:** Conectar AML a fuentes de datos (como Azure Database) para cargar los datos del entorno. Utilizar las herramientas de procesamiento de datos para limpiar, transformar y estructurar los datos, eliminando valores nulos, realizando transformaciones de características y gestionando variables categóricas. Utilizar notebooks de Python o R para realizar un EDA y visualizar los datos, obteniendo un mejor rendimiento de las variables claves y su relación con el objetivo.
- 5- **Selección de variables y creación de características:** Usar técnicas de características para identificar las variables que aporten más valor, crear nuevas variables o una combinación para optimizar el modelo.
- 6- **Entrenar el modelo:** Elegir algoritmos según el tipo de proyecto. Usar el SDK de Azure ML o la interfaz gráfica para entrenar al modelo. Configurar un experimento para optimizar los parámetros y mejorar la precisión del modelo.

- 7- **Evaluacion del modelo:** Separar los datos en conjunto de entrenamiento, validación y prueba. Evaluar el modelo con métricas según el objetivo. Ejecutar experimentos para comprar modelos y elegir el más adecuado.
- 8- **Implementación del modelo:** Implementar el modelo en un entorno de producción usando contenedores Docker o un servicio de inferencia en tiempo real. Creación de endpoint de API que faciliten la integración con aplicaciones y sistemas. Realizar pruebas para verificar que el modelo funciona correctamente en producción y ajustar parámetros de ser necesario.
- 9- **Monitoreo y mantenimiento:** Configurar alertas para monitorear el rendimiento y la presicion del modelo. Utilizar AML para detectar cambios en la distribución de datos. Implementar versionado.
- 10- **Visualización de resultados e informes:** Integrar con Power Bi para crear dashboards interactivos y visualización de resultados. Usar informes detallados de métricas de modelo y el proceso de minería de datos para documentar los resultados y recomendar acciones basadas en los resultados obtenidos.
- 11- **Evaluación y lecciones aprendidas:** Documentar cada fase del proyecto, desde el ingreso de datos hasta el despliegue y monitoreo. Evaluar el proyecto y realizar una sesión de revisión con el equipo para identificar mejoras y lecciones aprendidas.

Conclusión

La correcta organización y ejecución de un proyecto es muy importante en todos los aspectos de nuestra vida, no solo en el laboral.

Por eso es de vital importancia a la hora de implementar un proyecto de Data Mining tener una correcta organización del proyecto, desde las herramientas que se van a utilizar, el presupuesto con el que se cuenta o se necesita, el personal necesario y que habilidades deben presentar entre otras cosas, para que dicho proyecto llegue a ser un éxito.

Un proyecto de Data Mining ayuda a las empresas a comprender las necesidades de sus clientes y anticiparse a las tendencias del mercado, se pueden detectar patrones de conductas de los clientes, optimizar ventas y logística, diseñar tiendas, predecir ventas entre muchas otras ventajas, por tales motivos es de vital importancia que este aplicado de una manera correcta y que sea totalmente funcional lo cual se va a ver reflejado en mejores ganancias y menos costos.

Muchas gracias profesor por las clases y sus explicaciones.

Este trabajo fue subido al siguiente enlace de GitHub

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

Referencias

Data mining: 6 etapas para generar insights valiosos. (2021, February 11). Zendesk.

<https://www.zendesk.com.mx/blog/data-mining-que-es/>

Khan, F. (2024, May 7). Las mejores herramientas de minería de datos en 2024. *Astera*.

<https://www.astera.com/es/type/blog/data-mining-tools/>

Oracle México. (n.d.). Oracle.com. Retrieved October 27, 2024, from

<https://www.oracle.com/mx/>

Shankar, K. V. (2019). *Azure Machine Learning*. [https://azure.microsoft.com/es-](https://azure.microsoft.com/es-mx/products/machine-learning/)

[mx/products/machine-learning/](https://azure.microsoft.com/es-mx/products/machine-learning/)

(N.d.-a). Iberdrola.com. Retrieved October 27, 2024, from

<https://www.iberdrola.com/innovacion/data-mining-definicion-ejemplos-y-aplicaciones#:~:text=%C2%BFQu%C3%A9%20es%20el%20'data%20mining,sentido%20y%20convertirla%20en%20conocimiento.>

(N.d.-b). Sap.com. Retrieved October 27, 2024, from

<https://www.sap.com/latinamerica/index.html>