 [Traducido del inglés al español - www.onlinedoctranslator.com](https://www.onlinedoctranslator.com/es/?utm_source=onlinedoctranslator&utm_medium=docx&utm_campaign=attribution)

Minería de datos - Curso de pregrado

B.Sc. en Ingeniería de Sistemas - Universidad del Norte

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |

*"La minería de datos es como buscar oro en vastas montañas de información. Descubre información valiosa y patrones que pueden transformar las empresas e impulsar la innovación". - Anónimo*

| Instructor | Horarios de Atención |
| --- | --- |
| Elías D. Niño Ruiz, Ph.D.  [enino@uninorte.edu.co](mailto:enino@uninorte.edu.co)  <https://enino84.github.io/> |  |

Directrices del proyecto

En este curso, tendrá la emocionante oportunidad de colaborar como equipo y participar en la extracción de datos en un conjunto de datos de su preferencia. Tiene la libertad de elegir entre una variedad de conjuntos de datos diversos, incluidos, entre otros, los siguientes:

1. https://www.datos.gov.co/

2. https://www.policia.gov.co/grupo-informacion-criminalidad/estadistica-delictiva

3. https://www.kaggle.com/datasets

Si bien los dos primeros conjuntos de datos son particularmente interesantes debido a su relevancia para nuestra vida social, siéntase libre de explorar cualquier otra fuente de datos que capte su curiosidad. Lo animo a seleccionar conjuntos de datos que ofrezcan una cantidad sustancial de muestras y columnas (características) para garantizar un análisis significativo y perspicaz.

A lo largo de este curso, estará equipado con las herramientas y la orientación necesarias para descubrir patrones, tendencias y conocimientos valiosos a partir de los datos. Trabajando juntos, obtendrá una valiosa experiencia en el campo de la minería de datos y desarrollará sus habilidades para extraer información valiosa.

Presentaciones introductorias - presentación de 15 minutos

1. Introducción (1-2 minutos):

* Presentar al equipo y sus respectivos roles.
* Proporcione una breve descripción general del curso de minería de datos y su importancia.
* Establecer el contexto para la presentación.

2. Selección del tema del proyecto y conjunto de datos (1-2 minutos):

* Indique claramente el tema del proyecto elegido relacionado con la minería de datos.
* Introduzca el conjunto de datos seleccionado, ya sea desde las opciones proporcionadas o desde otro lugar.
* Explique por qué el conjunto de datos es relevante para el proyecto y los objetivos de la minería de datos.

3. Objetivos (1-2 minutos):

* Presentar los principales objetivos del proyecto de minería de datos.
* Describa lo que el equipo espera lograr a través del análisis y la exploración de datos.

4. Exploración y preparación de datos (1-2 minutos):

* Discuta el proceso de exploración de datos, incluida una descripción general de la estructura del conjunto de datos y, potencialmente, un panel de control.
* Mencione cualquier paso de preprocesamiento o limpieza de datos tomado para garantizar la calidad de los datos.

5. Plan del proyecto y cronograma (1-2 minutos):

* Comparta el plan del proyecto, incluidas las tareas que deben completarse y la división de responsabilidades entre los miembros del equipo.
* Proporcione un cronograma estimado para cada fase del proyecto.

6. Resultados esperados e impacto (1-2 minutos):

* Discuta los posibles resultados y conocimientos que el equipo espera obtener.
* Explique cómo estos hallazgos podrían ser valiosos o impactantes en el contexto dado.

7. Conclusión y preguntas y respuestas (1-2 minutos):

* Resume los puntos clave de la presentación.
* Abra la palabra para las preguntas de la audiencia (si corresponde) y responda a ellas.

Recuerde, la clave para realizar una presentación exitosa de 15 minutos es ser claro, conciso y centrado en los aspectos más importantes del proyecto. Asegúrese de que cada miembro del equipo conozca su función y contribuya de manera efectiva al contenido de la presentación. Practicar la presentación de antemano también puede ayudar a garantizar que se ajuste al límite de tiempo y que esté bien estructurada. ¡Buena suerte con las presentaciones de proyectos!

Presentaciones intermedias - presentación de 15 minutos

1. Introducción (1-2 minutos):

* Describe brevemente el proyecto.

2. Ingeniería de características (1-2 minutos):

* Explicar las técnicas de ingeniería de características aplicadas para extraer características relevantes del conjunto de datos.
* Resalte las características que fueron más influyentes en el modelo de aprendizaje automático.

3. Selección del modelo (2-3 minutos):

* Presentar los diferentes modelos de aprendizaje automático considerados para la tarea.
* Discuta la razón detrás de la selección de un modelo o conjunto de modelos en particular.

4. Capacitación y evaluación del modelo (2-3 minutos):

* Describa el proceso de entrenamiento de los modelos elegidos en el conjunto de datos.
* Discutir las métricas de evaluación utilizadas para medir el desempeño de los modelos.

5. Resultados y perspectivas (2 minutos):

* Comparta los resultados del análisis de aprendizaje automático, incluidos los hallazgos o patrones interesantes descubiertos.
* Discuta cómo los resultados se alinean con la declaración inicial del problema y si se lograron los objetivos.

8. Desafíos y lecciones aprendidas (1-2 minutos):

* Aborde cualquier desafío que haya enfrentado durante el proyecto y cómo el equipo los superó.
* Comparta lecciones clave aprendidas a través del proceso de uso de BigQuery Machine Learning.

9. Sesión de preguntas y respuestas (1 minuto):

* Abra la palabra para preguntas de la audiencia y responda consultas relacionadas con el proyecto y la presentación.

Al seguir esta estructura, cada equipo puede presentar una descripción general completa de su trabajo sobre el uso de BigQuery Machine Learning para el análisis de datos dentro del marco de tiempo asignado de 15 minutos. Recuerde concentrarse en los aspectos más significativos del proyecto y use elementos visuales para mejorar la claridad de la presentación. Practique la entrega para garantizar una presentación fluida y atractiva. ¡Mucha suerte con la presentación!

Presentaciones finales - presentación de 15 minutos

1. Introducción (1-2 minutos):

* Comience con una breve introducción de su proyecto.

2. Descripción general de los componentes (1 minuto):

* Proporcione una descripción general de alto nivel de cada componente: dbt, BigQuery ML, SQLite y GraphQL.
* Explique sus respectivos roles en la arquitectura de datos y cómo se complementan entre sí.

3. Flujo de datos e integración (2-3 minutos):

* Presente el flujo de datos general dentro de la arquitectura, destacando cómo se mueven los datos a través de los distintos componentes.
* Ilustre cómo se usa dbt para la transformación y preparación de datos, BigQuery ML para el entrenamiento de modelos de aprendizaje automático, SQLite para necesidades de bases de datos locales y GraphQL para consultas de API.

4. Beneficios y ventajas (1 minuto):

* Explique cómo este enfoque integrado mejora la eficiencia, la escalabilidad y la facilidad de mantenimiento.

5. Estrategia de producción (2 minutos):

* Describa la estrategia para poner en producción toda la arquitectura de datos.
* Aborde las consideraciones para la implementación, el monitoreo y el escalado de los componentes en un entorno de producción.

7. Desafíos y Mitigaciones (1-2 minutos):

* Aborde cualquier desafío que haya encontrado el equipo mientras desarrollaba la arquitectura.
* Comparte estrategias o soluciones implementadas para superar estos desafíos.

8. Seguridad y Gobierno de Datos (2 minutos):

* Resalte la importancia de la seguridad y el gobierno de los datos en la arquitectura de datos en producción.
* Discuta brevemente las medidas tomadas para garantizar la privacidad y el cumplimiento de los datos.

9. Conclusión (1 minuto):

* Resuma los puntos principales de la presentación, enfatizando los beneficios de la arquitectura de datos integrados.
* Resuma el análisis de datos: hallazgos de su conjunto de datos.
* Resuma el rendimiento de su modelo ML.

10. Sesión de preguntas y respuestas (1 minuto):

* Abra la palabra para preguntas de la audiencia y responda consultas relacionadas con la arquitectura y la presentación.

Al seguir esta estructura, el equipo puede presentar una descripción general integral de su trabajo en el diseño de una arquitectura de datos lista para la producción mediante dbt, BigQuery ML, SQLite y GraphQL en un período de 15 minutos. Use ayudas visuales y explicaciones concisas para transmitir la información clave de manera efectiva. Practique la entrega para asegurar una presentación clara y atractiva. ¡Mucha suerte con la presentación final!

Observaciones adicionales:

Para garantizar una colaboración y un control de versiones efectivos, se debe realizar un seguimiento de cada paso del proceso mediante GitHub. Todo el código desarrollado para el proyecto debe almacenarse en el repositorio designado, y los archivos Léame explicativos pueden acompañar a los archivos fuente para proporcionar una documentación clara.

Además, debe enviar un informe conciso de dos páginas para cada presentación a través de Blackboard. Siéntase libre de usar cualquier plantilla que considere adecuada para presentar sus ideas de manera efectiva. El informe debe abarcar los principales hallazgos, metodologías y puntos de vista derivados del proyecto.

La participación en el equipo es crucial para completar con éxito el curso. A cualquier estudiante que no participe activamente con el equipo se le asignará una calificación de 0 para el componente del curso respectivo. Es fundamental informarme si algún miembro del equipo no puede contribuir al proyecto.

Por último, todas las tareas deben cumplir con el Código de Honor para estudiantes de UniNorte. El plagio y las prácticas deshonestas no se tolerarán y pueden tener consecuencias académicas graves.

Colaboremos y aseguremos una experiencia de aprendizaje fructífera y ética a lo largo del proyecto. Si tiene alguna pregunta o inquietud, no dude en comunicarse. ¡Buena suerte con el proyecto y el envío de informes!