

MINERÍA DE DATOS

Maximiliano Ojeda

muojeda@uc.cl



IIC-2433

Proyecto

- El curso contempla trabajo grupal en torno a un proyecto (40%).
- El proyecto final es en grupos de hasta tres personas.
- El proyecto consta de tres etapas: Propuesta (10%), avance de proyecto (30%) y entrega final (60%).

Proyecto

Propuesta

Haga una presentación introduciendo su proyecto. Además, produzca un ‘graphical abstract’ que explique con claridad en qué consiste su proyecto. Finalmente, deberá definir un plan de actividades declarando cuál será su avance para la entrega “Avance de proyecto”. Luego, grabe un video de 3 minutos donde presente su proyecto y súbalo a canvas.

Fecha de entrega: 17 de octubre (canvas)

Avance de proyecto

Consiste en una presentación de avance del proyecto que debe hacer en clases. Dispondrá de 5 minutos para presentar un video + un tiempo para preguntas. El avance del proyecto debe estar comprometido en la propuesta. Se espera que al menos tenga la etapa de recolección de datos y el setting experimental (códigos/herramientas) listo para esta entrega.

Fechas de presentaciones: 28 y 30 de octubre.

Entrega final

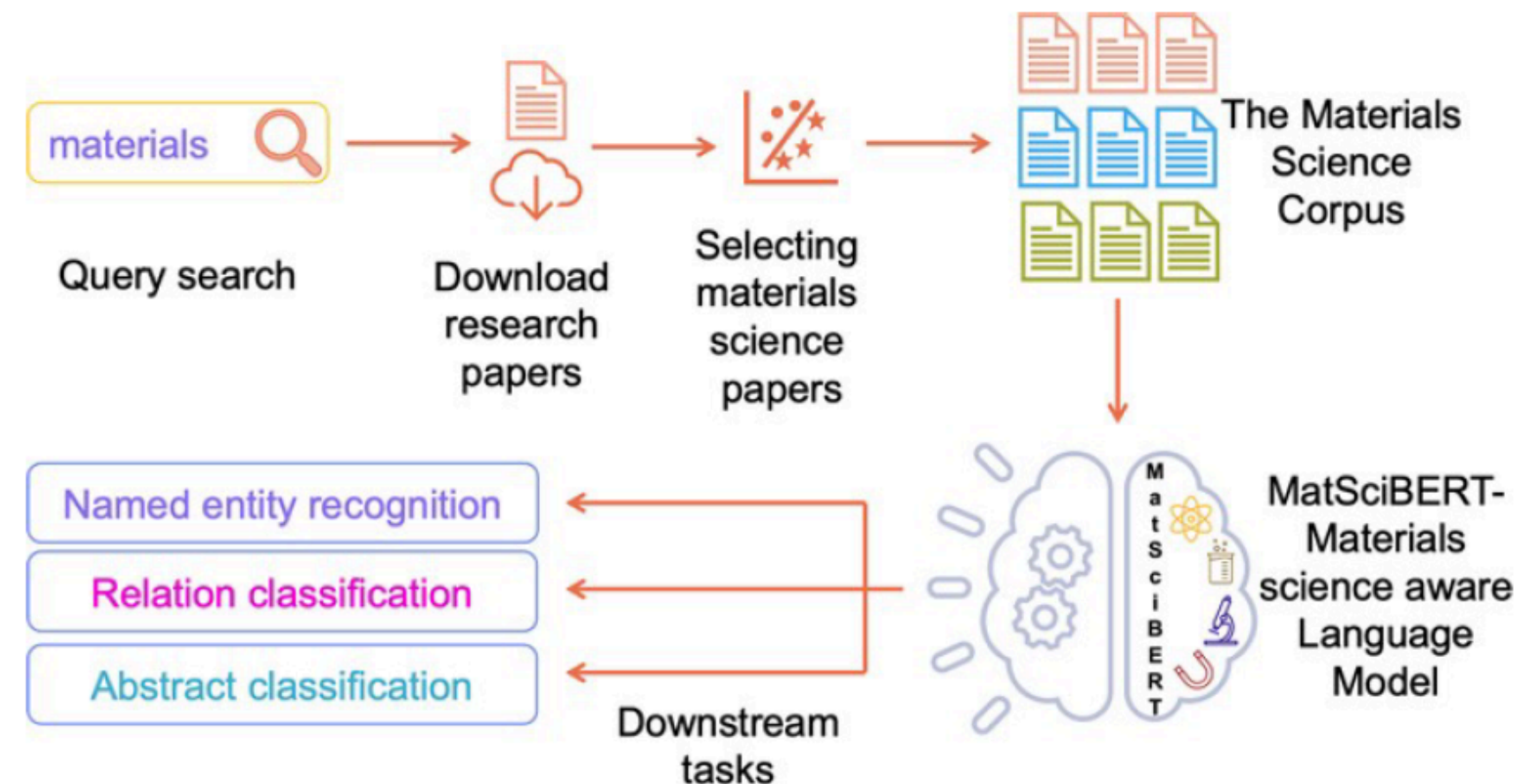
Consiste en la presentación final del proyecto. Debe presentar el proyecto en clases. Dispondrá de 7 minutos para presentar un video + un tiempo para preguntas. Debe hacer un live demo, mostrando el tipo de resultados que obtiene. Incluya una etapa de evaluación del modelo, con métricas que permitan comparar sus resultados con un baseline. Incluye elementos novedosos en relación al avance, considerando las últimas clases (por ejemplo, causalidad).

Fechas de presentaciones: 25 y 27 de noviembre

Proyecto

¿Qué es un graphical abstract?

Es una representación visual de las ideas de su proyecto. A menudo se utiliza en publicaciones académicas para resumir de manera concisa y visual el contenido del artículo y capturar la atención del lector. El Graphical Abstract suele ser una imagen que incluye gráficos, diagramas, ilustraciones y/o fotografías, acompañado de texto breve que resume los puntos principales del proyecto. El objetivo principal del Graphical Abstract es comunicar de manera visualmente atractiva y accesible.



¿Cuáles son los elementos de un proyecto exitoso en Minería de Datos?

Definición clara de objetivos: Es importante tener una comprensión clara de los objetivos y metas del proyecto de MD. De esta manera, se puede seleccionar la mejor técnica y herramienta para cumplir con esos objetivos.

Recopilación de datos adecuada: La calidad y cantidad de los datos son cruciales para el éxito del proyecto. Es importante seleccionar el conjunto de datos adecuado y asegurarse de que sea lo suficientemente grande y diverso para entrenar y validar el modelo. Es importante tener acceso a datos de alta calidad y saber cómo limpiarlos, procesarlos y prepararlos para el análisis. Esto puede incluir el manejo de datos faltantes, la detección de anomalías y la transformación de datos.

Selección de algoritmos y técnicas de minería de datos: Existen diversas técnicas de MD, como clustering y clasificación, entre otras. Es importante seleccionar la técnica adecuada que se ajuste a los objetivos del proyecto.

Entrenamiento y validación del modelo: Después de seleccionar la técnica adecuada, es necesario entrenar el modelo con los datos recopilados.

Evaluación del modelo: Una vez que se ha construido un modelo, es necesario evaluar su rendimiento y precisión. Esto puede implicar la validación cruzada, la evaluación de la matriz de confusión, la precisión, la curva ROC, entre otros.



¿Cuáles son los elementos diferenciadores de su proyecto?

Dominio: algo novedoso que puede hacer es trabajar en un dominio no convencional. La mayoría de los proyectos MD disponibles en git abordan el dominio de las finanzas, noticias o datos clínicos, entre otros en los cuales el MD se aplica de forma evidente. Es menos evidente aplicarlo en deportes, entretenimiento, gobierno, educación, minería, etc. Abordar un problema en un dominio no convencional es una contribución.

Integración: algo novedoso que puede hacer es extender un modelo. En clases hemos visto los modelos convencionales más usados en MD. Extender una técnica, integrándola con otra, es muy novedoso. Abordar un problema que integre algoritmos es una contribución.

Resolver el problema de alguien concreto: algo relevante es abordar un problema real cuyos efectos puedan generar beneficios a la sociedad. Resolver un problema concreto generando un beneficio para la sociedad es una contribución.



¿Qué debe tener la primera presentación?

- Nombre del proyecto
- Integrantes
- ¿Cuál es el problema que aborda en su proyecto?
- ¿Cuál(es) técnica(s) va a usar?
- ¿Cuáles son los datos con los cuales va a trabajar? Declare si son datos abiertos. Indique volumen y tipo de datos con los que trabajará.
- ¿Cuáles son los elementos diferenciadores (contribuciones) de su proyecto?
- Plan de actividades.
- Entregables al avance y a la entrega final.

MINERÍA DE DATOS

Maximiliano Ojeda

muojeda@uc.cl



IIC-2433