



Pontificia Universidad Católica De Chile  
Escuela De Ingeniería  
Departamento de Ciencia de la Computación  
IIC2433 - Minería de Datos  
Primer semestre 2022

# Proyecto Final

## Fechas restantes

**Fecha 3:** presentación final [8 de julio] 40% de la nota del proyecto

Cada grupo deberá presentar su trabajo terminado (máximo 7 minutos). Se espera una presentación con material de apoyo visual (*Power Point*, *LaTeX* o alguna herramienta en formato diapositiva). La asistencia es obligatoria para todos los integrantes del grupo durante todo el módulo.

Las presentaciones son el **jueves 8 de julio de 8:30 a 11:30**. Se harán vía el servidor de Discord del curso. En cuanto a los detalles de cómo se realizarán las presentaciones en discord, se informará próximamente.

Como parte de la evaluación, cada grupo deberá reunirse con su ayudante a cargo como **mínimo en una ocasión** antes de la fecha de las presentaciones, con el fin de presentar su progreso y recibir retroalimentación que deberán incluir en la presentación y fecha final.

**Fecha 4:** entrega final [9 de julio] 40% de la nota del proyecto

Cada grupo deberá subir a su repositorio de Github un documento en formato **PDF** con un reporte detallado de su proyecto. Se aceptarán *commits* hasta las 23:59 horas del **9 de julio**.

**Además**, se pedirá subir también todo el código utilizado y como **mínimo**, documentar el código que permita utilizar el modelo o proceso que hayan desarrollado en el proyecto.

## Formato presentaciones

Duración máxima de 7 minutos, más 3 minutos para preguntas. Las presentaciones serán evaluadas de acuerdo a los siguientes puntos (**¡no olvidar ninguno!**):

1. Motivación: ¿Cuál es el contexto general del tema/problema/datos de estudio? ¿Por qué es de interés o cuál es su relevancia?
2. Hipótesis/Problemáticas iniciales: ¿Qué es lo que se desea analizar? o ¿Qué problema queremos resolver?
3. Descripción de los datos que se van a utilizar: Sus características más relevantes e interesantes (estadísticas, gráficos, etc).
4. Descripción breve del proceso: ¿Qué hicimos para resolver el problema?
5. Principales resultados y conclusiones del proyecto.
6. Calidad de la presentación: Preparación del grupo y claridad en la exposición.

**Tip:** Se aconseja practicar la presentación antes para cumplir con el tiempo indicado.

## Formato entrega final

1. Todo código utilizado en el proyecto.
2. Modelos entrenados<sup>1</sup>.
3. Código documentado para hacer uso del modelo final entrenado. Este código debe ser capaz de leer los datos que entregue el usuario e imprime en consola la clasificación/cluster/recomendación según el problema abordado. La entrada de datos que entrega el usuario y el formato queda a su decisión, puede ser el *path* de algún archivo, llenar alguna variable del código, etc.
4. Un informe en formato PDF, de máximo 12 planas sin considerar portada, ni referencias, anexo o apéndice, que incluye:
  - a. Introducción al tema. ¿Cuál es el contexto general del tema/problema/datos de estudio? ¿Por qué es de interés? Debe quedar clara la motivación de su proyecto.

---

<sup>1</sup> Se recomienda investigar la librería Pickle o como guardar modelos con sklearn.

- b. Temática o problemática que están abordando. ¿Qué es lo que se desea analizar? o ¿Qué problema queremos resolver?
- c. Descripción de los *datasets* ocupados y de sus columnas utilizadas.<sup>2</sup>
- d. Exploración de los datos. Sus características más relevantes e interesantes (estadísticas, gráficos, etc).
- e. Descripción del proceso para llegar a los resultados. Esta descripción incluye la obtención de los *datasets*, el preprocesamiento realizado, modelos utilizados y métricas para comparar.
- f. Evaluación de resultados. Mediante métricas explicar cómo su propuesta enfrentó el problema.
- g. Conclusiones a partir de los resultados. ¿Se logró solucionar el problema abordado?
- h. Trabajo futuro. ¿Cómo se podría continuar este proyecto? ¿Qué cosas se podrían hacer que por tiempo y/o dificultad se dejaron de lado?
- i. Descripción de principales dificultades (al menos 2). ¿Qué problemas tuvieron durante el desarrollo del proyecto? ¿El dataset fue difícil de conseguir? ¿El modelo propuesto inicialmente funcionó como esperaban?. A partir de estos problemas, indique 2 consejos para abordar de mejor forma su proyecto. Justifique sus consejos en función de la experiencia que tuvieron en el proyecto.
- j. Referencias y links a las páginas que recurrieron. El formato para las referencias queda a su decisión.
- k. Anexo/Apéndice. Sección para poner gráficos y tablas imprescindibles que serán referenciados durante el informe.

---

<sup>2</sup> Si hay alguna columna que no ocupan, no la mencionen.