

# Proyecto Final

## Minería de Datos

### IIC2433 - 2020-2

## Descripción General

El proyecto final del curso se realizará en grupos de cuatro alumnos. Si alguien quiere trabajar solo o en pares, también puede hacerlo. No se admitirán grupos de 5 o más personas. Se deberá aplicar alguno de los contenidos vistos a lo largo del curso, en un tema de interés de los miembros del grupo. En primera instancia, los alumnos deberán presentar un set de datos, un contenido relevante del curso para aplicar y un problema de investigación, de forma que haga sentido aplicar este contenido particular en los datos elegidos. Se sugiere fuertemente trabajar con datos relevantes a los intereses laborales o académicos del estudiante. Si no tiene acceso a una base de datos, puede revisar los siguientes enlaces:

- <https://www.kaggle.com/datasets>
- <https://archive.ics.uci.edu/ml/datasets.php>
- <https://www.reddit.com/r/datasets/>

Luego, deberán ejecutar su propuesta usando *Python* (el uso de otro lenguaje de programación debe ser conversado con el profesor) y cualquier librería necesaria, con el fin de encontrar conclusiones relevantes que respondan al problema planteado, poniendo en práctica lo aprendido durante el semestre. Finalmente, deberán presentar su trabajo en un informe final y una presentación oral.

A continuación se muestran dos ejemplos de propuestas válidas:

1. Buscar reglas de asociación para recomendar libros según categorías en base de datos de biblioteca municipal.  
**Set de datos:** Base de datos de biblioteca municipal.  
**Contenido relevante:** Reglas de asociación.  
**Problema de investigación:** Asociaciones entre préstamos de libros según categoría.
2. Evaluación de clasificadores para predecir si los clientes pagan o no sus deudas.  
**Set de datos:** Base de datos de pagos en tienda X.  
**Contenido relevante:** Clasificación y evaluación de clasificadores  
**Problema de investigación:** Predecir qué clientes pagan sus deudas y encontrar cuál de los clasificadores vistos en clases predice esto de mejor manera.

## Contenidos adicionales a considerar

Dada la cantidad de ayudantes y sus diversas especialidades, es posible aplicar el proyecto a contenidos que no veremos o profundizaremos en el curso. A continuación se presenta un listado de contenidos adicionales que podrían utilizar en su proyecto ya que una o más ayudantes podrá guiarlos.

- *Text mining*
- Visualización
- Sistemas recomendadores
- *Deep Learning*
- Astronomía: clasificación automática, manejo de imágenes, curvas de luz, caracterización de features.
- Procesamiento de imágenes
- Reconocimiento de patrones
- *Process mining*.

En caso de haber otro tema de interés que no esté en esta lista, puede hacer una issue preguntando al respecto.

## Fechas

**Fecha 0:** propuesta de grupos [9 de octubre]

Los alumnos deberán responder un [formulario de Google](#), indicando los integrantes del grupo. Esta propuesta debe ser de 4 integrantes. En otro caso, el cuerpo docente se encargará de formar los demás equipos. Se abrirá una [issue en el Syllabus](#) para que puedan comunicarse en la búsqueda de grupo. Durante el fin de semana se formarán los equipos faltantes y se notificará el lunes a primera hora. Luego se les enviará una invitación a github classroom para que creen su repositorio de grupo. Se aceptarán respuestas del formulario hasta las 23:59 horas del 9 de octubre.

**Fecha 1:** propuestas preliminares [16 de octubre]

Cada grupo deberá subir a su repositorio de Github un documento con dos párrafos de máximo 100 palabras cada uno con una propuesta preliminar. Cada propuesta debe ser de un contenido diferente y debe indicar el conjunto de datos a utilizar, el contenido relevante del curso a aplicar, problema de investigación y un orden de prioridad si lo desea pertinente. Durante la semana del 19 al 23 de octubre se evaluarán las propuestas y el equipo docente podrá sugerir modificaciones o un cambio de proyecto, en caso de que las propuestas no se adecue a lo esperado en el curso. También se le notificará a cada grupo cuál propuesta deberán realizar y el ayudante que estará a cargo de su proyecto. Se aceptarán *commits* hasta las 23:59 horas del 16 de octubre.

**Fecha 2:** presentación de propuestas y avance [20 de noviembre]

Esta entrega pondera un 20% de la nota del proyecto.

Cada grupo deberá subir a su repositorio de Github un documento en formato PDF y todo código (.ipynb o .py) realizado hasta la fecha. En términos de código, se espera mínimo que el *dataset* ya esté pre-procesado y analizado, es decir, características básicas del dataset como cantidad de filas, columnas, distribución de *features* más importantes, etc. En términos del PDF, este documento debe poseer un máximo de 5 páginas y su contenido debe ser:

1. Introducción a la problemática abordada
2. Descripción de los datos y exploración inicial.
3. Temática o problemática central y describir cómo se abordará inicialmente.
4. Trabajo pendiente para finalizar el proyecto

Se aceptarán *commits* hasta las 23:59 horas del 20 de noviembre.

**Fecha 3:** presentación final [semana del 7 de diciembre]

Esta entrega pondera un 40% de la nota del proyecto.

Cada grupo deberá inscribirse en un horario en donde deberán presentar su trabajo terminado (máximo 7 minutos). Se espera una presentación con material de apoyo visual (*Power Point*). La asistencia es obligatoria para todos los integrantes del grupo. El formato del informe final se avisará posteriormente.

**Fecha 4:** entrega final [13 de diciembre]

Esta entrega pondera un 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 13 de diciembre. El formato de la presentación se avisará posteriormente.