

# Ciencia de datos

## Proyecto grupales

El proyecto del curso consiste en la aplicación de los conceptos vistos durante el curso Ciencia de Datos en diversos tópicos, los cuales deberán involucrar algoritmos vistos en clases. Los tópicos esperados involucran marketing y ventas online. La tarea de cada proyecto depende del grupo y depende del tema elegido.

Se espera que cumplan múltiples hitos. En particular tendrán 3 entregas y 1 exposición:

- a. Análisis y preprocesamiento de datos (8/5 a las 11:59 pm, **peso: 20%**)
  - i. Definición de proyecto
  - ii. Descripción de datos
  - iii. Análisis 1D y 2D de datos (EDA).
  - iv. Preprocesamiento de datos.
  - v. Diseño de experimentos. Es decir que experimentos plantean realizar.
- b. Implementación de algoritmos básicos (29/5 a las 11:59 pm, **peso: 30%**)
  - i. Revisión de métodos relacionados con tarea, al menos 4 trabajos después del 2015.
  - ii. Descripción de algoritmos básicos para tarea. Los algoritmos y tarea dependerán de discusión con profesor de curso.
  - iii. Implementación de algoritmos básico en tarea.
  - iv. Reporte de resultados obtenidos.
- c. Resultados de innovación de algoritmos (17/6 a las 11:59 pm, **peso: 30%**)
  - i. Revisión de métodos relacionados con tarea, al menos 3 trabajos después del 2015.
  - ii. Descripción e implementación de algoritmo con innovación. Los algoritmos y tarea dependerán de discusión con profesor de curso.
  - iii. Reporte de resultados obtenidos.
  - iv. Conclusiones de resultados
- d. *Exposición final de proyectos (fecha a definir a partir del 18/6, **peso: 20%**)*
  - i. *Presentación de grupal de proyectos.*
    - *Fecha a sortear en últimas semanas de curso.*
    - *Duración máxima 20 minutos.*
    - *Tiempo para preguntas: 3 minutos*
    - *Notas tanto individual como grupal.*

Nótese que se ha indicado la fecha y ponderación de cada fase. La nota de este proyecto equivale a nota de proyecto de investigación (ver syllabus).

Las entregas consisten en i. la documentación y ii. código de proyecto.

- Para documentación se sugiere usar Latex dado que es lenguaje de comunicación académico. Especialmente pueden usar web overleaf. En caso de seguir recomendación tendrán bonificación de 0.5 puntos sobre nota de hito.
- Para código se sugiere reportar usando notebook de Python. Para ello pueden usar Anaconda como Google colab.

Los resultados serán entregados vía online usando

En cada entrega deben entregar dentro de archivo ZIP:

1. La documentación, la cual contendrá al menos 7 páginas, preferentemente con impresiones de pantallas, y un máximo de 25 páginas. Se recomienda el uso de Latex, pero no es obligatorio. Una página debe ser para caratula. El informe se debe entregar en formato Latex así como PDF.
2. Código fuente de avances de proyectos.

Se debe enviar archivo ZIP a correo de profesor: [billy.peralta@unab.cl](mailto:billy.peralta@unab.cl). Opcionalmente pueden usar servicios en la nube si los archivos fueran muy voluminosos. El mínimo/máximo de alumnos por grupo es de 2/4 integrantes. La copia es Cualquier consulta práctica se hará con el profesor de curso ya sea vía mail o en oficina.

Saludos y suerte en este reto,

Billy