

## Tarea Parte 2 de Machine Learning

### Ensamblajes de algoritmos

El proyecto del curso consiste en el desarrollo de sistemas basados en el aprendizaje de máquinas, particularmente en la segunda parte de curso. El proyecto conforma parte de la calificación del estudiante por lo que tiene múltiples hitos con el objeto de lograr desarrollar el sistema de forma satisfactoria. En especial, como parte final de Tarea se espera el uso de ensamblajes.

Como primera opción de proyecto se elegirá:

A). La predicción de deserción de un estudiante dadas variables educativas y socioeconómicas. En particular, este dataset tiene 3 clases: deserción (dropout), en curso (enrolled) y graduado (graduated). La data debe ser descargada de: <https://archive.ics.uci.edu/dataset/697/predict+students+dropout+and+academic+success>. Considerar el uso de cualquiera de las técnicas vistas en el curso

B). Además puedes optar por continuar proyecto previo con proyecto de profesor Schwarzenberg basado en energía eléctrica. **En este caso, deben plantear técnicas distintas a las dadas en última entrega de este proyecto.**

Esta Tarea constará de 2 entregas:

- a. Avance inicial de proyecto (Viernes 7/6 a las 11:59 pm, **peso: 35%**)
  - i. Definición de proyecto. Importancia, contexto. En el caso de seguir proyecto de Parte 1 de curso, debe quedar clara la diferencia técnica.
  - ii. Descripción de datos. Fuente de data, cantidad de registros, número de variables y número de clases.
  - iii. Diseño de experimentos. Experimentos a realizar. Se sugiere plantear uso de enfoque validación cruzada con 10 folds. Métricas para validar propuesta (al menos 2).
  - iv. Implementación de un algoritmo de clasificación en tarea. Pueden reusar código existente.
  - v. Reporte de resultados obtenidos.
- b. Implementación de algoritmo básico (Viernes 21/6 a las 11:59 pm, **peso: 50%**)
  - i. Revisión de métodos relacionados con tarea, al menos 3 nuevos trabajos después del 2018.
  - ii. Descripción e implementación de algoritmo considerando **ensamblajes**. Los algoritmos y tarea pueden discutirse con profesor de curso.

- iii. Reporte de resultados obtenidos y comparación con resultados previos.
- iv. Conclusiones de resultados

Nótese que se ha indicado la fecha y ponderación de cada fase. Las entregas consisten en i. la documentación (**peso 10%**) y ii. código de proyecto (**peso 5%**).

- 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 0.1 puntos sobre nota de **hito a**.
- En caso de seguir proyecto de Parte 1 tendrán bonificación de 0.1 puntos sobre **nota final** de proyecto.
- Para código se sugiere reportar usando notebook de Python. Para ello pueden usar Anaconda como Google Colab.

En cada entrega deben entregar dentro de archivo ZIP con formato **rutcompleto1\_rutcompleto2\_rutcompleto3.zip**; dentro de buzón asignado a esta Tarea.

1. La documentación, la cual contendrá al menos 5 páginas (Excepto primera que no tiene límite inferior), preferentemente con impresiones de pantallas, y un máximo de 25 páginas por entrega. Se recomienda el uso de Latex, pero no es obligatorio. Una página debe ser para caratula. El informe se debe entregar en formato fuente (Latex o Word) así como PDF.
2. Código fuente de avances de proyectos. **Nótese que es obligatorio; en otro caso tendrá nota mínima dado que es evidencia de trabajo realizado.**
3. El proyecto se deberá subir al Sistema Canvas en buzón habilitado para ello.

El mínimo/máximo de alumnos por grupo es de 1/4 integrantes; a excepción que sigan mismo grupo de entrega anterior, donde pueden repetir los mismos grupos. La copia es penalizada siguiendo indicatrices de UNAB. Cualquier consulta práctica se hará con el profesor de curso ya sea vía mail o en oficina.

Saludos y suerte en este reto,

Billy