

Grado en Ciencia e Ingeniería de Datos

Aprendizaje Automático I

Práctica 2

Objetivo

La práctica consiste en la realización y presentación de un informe de un proyecto de Ciencia de Datos, aplicando técnicas de Aprendizaje Automático. Durante el desarrollo del proyecto, se utilizarán las técnicas vistas en el curso y se aplicarán a un conjunto de datos seleccionado.

- El informe se entregará en formato **R Markdown y HTML** y debe ser claro, autocontenido, bien organizado y justificado. Además, se debe especificar claramente las partes de la práctica realizadas por cada integrante del grupo.
- Algunas de las técnicas a aplicar deberán ser programadas por el alumnado. Esto se realizará en otro archivo de R y se entregará también en formato **R Markdown y HTML**.
- Adicionalmente, se debe entregar una **presentación en PDF** con un máximo de **15 diapositivas**, que será defendida el día de la entrega. Cada equipo tendrá un máximo de **15 minutos** para la exposición.

Todos los archivos se entregarán en un único **archivo comprimido**.

Instrucciones de entrega

- Cada grupo deberá entregar el archivo comprimido a través del aula virtual, indicando en el mismo el nombre de los integrantes y el número del grupo asignado por los profesores.
- La entrega debe ser realizada por uno de los miembros del grupo.
- **Fecha límite de entrega:** 8 de mayo de 2025, a las 14:00 horas.

Rúbrica

1. (1 puntos) **Planteamiento del problema.** Objetivos y enfoque de la práctica.
2. (6 puntos) **Entrenamiento de modelos y programación.** El entrenamiento implica realizar búsqueda de hiperparámetros.
 - Se valorará la inclusión de los siguientes modelos (siempre y cuando sea procedente aplicarlos): k-NN, Análisis Discriminante Lineal, Árboles de Decisión, Bagging, Boosting, Random Forest y Naïve Bayes.
 - Implementación por parte del alumnado de los siguientes modelos: k-NN, Análisis Discriminante Lineal, Bagging, Boosting (versión Adaboost) y Naïve Bayes.
 - Para los modelos que requieran selección de hiperparámetros, se debe programar también la búsqueda de hiperparámetros mediante **grid search** con validación cruzada y distintas medidas de rendimiento.
 - Comparación de los modelos programados con las funciones propias de R.

- Documentación detallada del código en un archivo independiente (**.rmd**) con explicaciones claras y comentadas línea a línea. En dicho archivo debe estar explicado previamente cómo se plantea el código.
- Para integrar las funciones implementadas en el archivo principal de estudio de los datos, se usará el comando `source()` de R.

3. (1 punto) **Medidas de rendimiento y comparación de modelos.**

- Justificación de las medidas de rendimiento utilizadas.
- Evaluación comparativa de los modelos entrenados con base en estas medidas.

4. (1 puntos) **Conclusiones y trabajo a futuro.**

- Elección del modelo final
- Análisis en el conjunto de validación del modelo seleccionado.
- Conclusiones y propuestas de mejoras y trabajo futuro.

5. (1 punto) **Exposición de resultados.**

- Claridad y organización de la presentación.
- Calidad visual y estructural de las diapositivas.
- Capacidad de explicar y defender el trabajo realizado.
- Correcta gestión del tiempo (no exceder los 15 minutos).