

## INTELIGENCIA ARTIFICIAL GRADO EN INGENIERÍA INFORMÁTICA

2021-2022

### **Práctica 3 (OBLIGATORIA):**

#### **1. Objetivo**

Esta práctica consiste en aplicar diversos conceptos vistos relativos al tratamiento y pre-procesado de datos, la selección de métricas de comportamiento y la construcción de modelos de Aprendizaje Automático en un problema de clasificación, de forma que el alumno desarrolle su capacidad crítica a la hora de emplear y comparar estos modelos.

#### **2. Planteamiento**

El alumno empleará la base de datos elegida en el último apartado de la práctica entregable anterior, referida a un problema de diagnóstico médico o un problema similar. En caso de que el alumno desee cambiar de base de datos, siempre en un contexto similar, podrá hacerlo previa motivación clara de porqué es preciso dicho cambio. Una justificación insuficiente será penalizada. Adicionalmente, el problema debe ser necesariamente uno de clasificación supervisada (es decir, nunca será posible tratar un problema de regresión o no supervisado).

#### **3. Tareas a realizar y preguntas**

1. Describir brevemente en qué consiste el problema (1 punto, 10 líneas de texto máximo)
2. Sobre el conjunto de datos original proponer algún tratamiento justificado de los datos, alternativas posibles son el escalado de los datos, la compleción de la base de datos para evitar desbalanceo entre las clases, la eliminación de observaciones influyentes u otras más complejas como la generación mediante Montecarlo de observaciones subrogadas (2 puntos, 30 líneas de texto)
3. Dividir la nueva base de datos generada considerando algún criterio razonable en tres conjuntos: entrenamiento, test y validación y motivar dicha división. En el resto de la práctica, en lo relativo a la selección de modelos óptimos nunca se emplearán los conjuntos de test y validación. (1 punto, 10 líneas de texto máximo)
4. Seleccionar, empleando exclusivamente el conjunto de entrenamiento los modelos óptimos para cada uno de los modelos siguientes, los hiper parámetros a considerar aparecen entre paréntesis: k-nn(número de vecinos), Bayes naïve

(parámetro de suavizado), árboles de clasificación (profundidad del árbol) y redes alimentadas hacia delante de (número de capas -máximo 3- y número de neuronas por capa). Para dicha selección se aplicará necesariamente algún procedimiento de re-muestreo como validación cruzada (2 puntos, 20 líneas máximo).

5. Comparar, con criterios basado en la matriz de confusión los resultados obtenidos con los distintos modelos sobre el conjunto de test. Se deberá seleccionar el mejor modelo (knn, Bayes, etc.) indicando claramente los motivos por los cuales se hace dicha elección (2 puntos, 20 líneas máximo)
6. Empleando el conjunto de validación indicar si la elección realizada anteriormente es consistente explicando, en caso positivo o negativo los resultados obtenidos y las posibles mejoras al procedimiento. (1 punto, 10 líneas máximo)
7. Emplear un modelo lineal considerando las características del problema dado (no continuidad de la variable objetivo) y empleando algún criterio de selección de parámetros óptimos necesariamente explicado en clase. Comparar los resultados con el modelo óptimo seleccionado considerando las métricas propuestas anteriormente (1 punto, 10 líneas máximo).

#### 4. Formato

La práctica deberá entregarse **obligatoriamente** en formato pdf, sin comprimir, sin incluir portada e indicando los nombres y apellidos de los miembros del grupo en el encabezado de la primera página. Es posible incluir algunas gráficas y tablas sencillas pero en ningún caso código. Este documento debe ser producido a partir de los resultados obtenidos con la ejecución del código pero nunca puede referirse a resultados de la ejecución del mismo, es decir, debe ser auto-contenido indicando explícitamente los resultados obtenidos.

Adicionalmente es preciso entregar el archivo de código empleado (formato .ipynb, sin comprimir). La carga y manejo de los datos empleados **obligatoriamente**, deberán estar en formato .csv, sin comprimir.

#### 5. Fecha

La práctica se realizará a través de un buzón de entrega del código en *Blackboard* que estará abierto hasta las **22.00 horas del domingo 16 de enero de 2022**.

**Importante:** Cualquier intento de copia de contenido o código entre grupos o de recursos externos de cualquier tipo (códigos de *github*, *kaggle*, etc.) será penalizado, pudiendo conllevar desde la anulación completa de esta práctica al suspenso en la asignatura.