

Trabajo Práctico de Machine Learning Supervisado y No Supervisado

Contexto

En este ejercicio deberás poner en práctica tus conocimientos sobre creación de modelos de machine learning con el lenguaje de programación R. De manera particular, crearas un modelo para realizar una clasificación binaria y otro para agrupar datos.

Para este ejercicio, deberás emplear los datos de alumnos y calificaciones proporcionados para esta asignatura en [este directorio de Google Drive](#).

Los datos de partida para este ejercicio, serán el conjunto de datos que previamente preparaste para el trabajo práctico de regresión lineal que tiene como nombre de columnas: Legajo,Codigo_Materia, Cond_Regular, Cursada_Final.

Instrucciones

1. Agrega una quinta columna llamada "Situacion_Final" en la cuál deberás añadir un 1 si la calificación en la columna "Cursada_Final" es igual o mayor a 4, y un 0, si la calificación es menor a 4. El 1 representa que el alumno ha aprobado, mientras que el 0 indica que el alumno ha reprobado.
2. Considerando que las variables "Legajo", "Codigo_Materia" y "Cond_Regular" son independientes y que "Situación_Final" es la variable dependiente, diseña un modelo de regresión logística para predecir si un alumno aprobará o no una asignatura.
3. Empleando las variables "Legajo", "Codigo_Materia", "Cond_Regular" y "Situacion_Final" utiliza el modelo K-Means para identificar grupos dentro de los alumnos de ingeniería de sistemas. Emplea distintos valores de "K" y grafica los grupos (clusters) identificados para decidir que valor de "K" puede categorizar mejor a los alumnos. Como valores de "K" prueba con 2, 3 y 4.

Entrega

Deberás compartir en la carpeta compartida de Google Drive de la asignatura, un archivo en formato ZIP con los programas en el lenguaje R que hayas creado así como el archivo de datos utilizado.