**Ingeniería de Sistemas**

**Computación Científica y Analítica**

**Taller 3**

Estudiante: Oscar Fabian Espinosa Pinilla

1. **Del script de forest fires:**
   1. ¿Se desea resolver el problema utilizando aprendizaje supervisado o no supervisado? ¿Es un problema de clasificación o de regresión?

RTA/ Es un aprendizaje supervisado y es un problema de regresión

* 1. ¿Qué interpretación le puede dar a los resultados obtenidos?

RTA/ El coeficient de determinación esta fuera del rango, normalmente debería estar entre 0 y 1, pero para este caso está por debajo de 0, por lo tanto, no podemos confiar en el modelo para nada.

Patrón de fondo

Descripción generada automáticamente

1. **Del script de recursos humanos (rrhh):**
   1. ¿Cuál es la clase para la que el modelo más se equivoca? ¿Por qué?

RTA/ La clase que más se equivoca, creo que es Satifaction Level, porque hay una alta varianza en los datos.

* 1. ¿Cuál cree que es el propósito del parámetro *max\_depth* usado al momento de instanciar el modelo de árbol de decisión?

RTA/ Para definir la profundidad del árbol de decisión.

* 1. Para este caso particular, ¿por qué cree que es difícil obtener un buen clasificador?

RTA/ Es un modelo que, aunque no tenga muchos parámetros, se usan la mayoría, si o es que todos, pero los datos considero tienen poca correlación entre sí, lo cual provoca que cada parámetro altera la clasificación de distinta manera.

1. **Identificación de géneros musicales:** Tenga en cuenta que hay dos scripts: *music.ipynb* y *music-multiclass.ipynb*. En el primero se intenta crear un modelo clasificador solo para dos clases (caso binario) y en el segundo se entrena uno para todas las clases (géneros musicales) del dataset.
   1. Para el caso binario (*jazz and blues* vs. *soul and reggae*) ¿Es posible obtener mejores métricas entrenando un modelo basado en Random Forest?
   2. Escoja otro par de géneros, entrene un conjunto de modelos y documente los resultados del mejor que se haya obtenido.
   3. Para el caso multi-clase, ¿cuál es la clase para la que el modelo más se equivoca? ¿Por qué?
   4. Para el caso multi-clase, el modelo basado en red neuronal parece estar mayoritariamente sesgado hacia un género particular ¿Cuál género cree que es?
2. **Segmentación de cajas de compensación familiar (subsidio):**
   1. ¿Qué cajas de compensación parecen ser mayoritariamente diferentes a las demás?
   2. ¿A partir de qué características utilizadas para el entrenamiento del modelo se podría explicar la razón por la que las cajas anteriores fueron agrupadas en clusters tan pequeños?
   3. ¿Se pueden obtener resultados más homogéneos utilizando cantidades diferentes de clusters para el entrenamiento? Entienda homogeneidad como clusters con cantidades similares de instancias de datos.