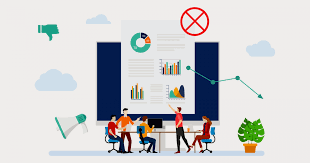
**Evaluación y comparación de**

**Modelos básicos**

****

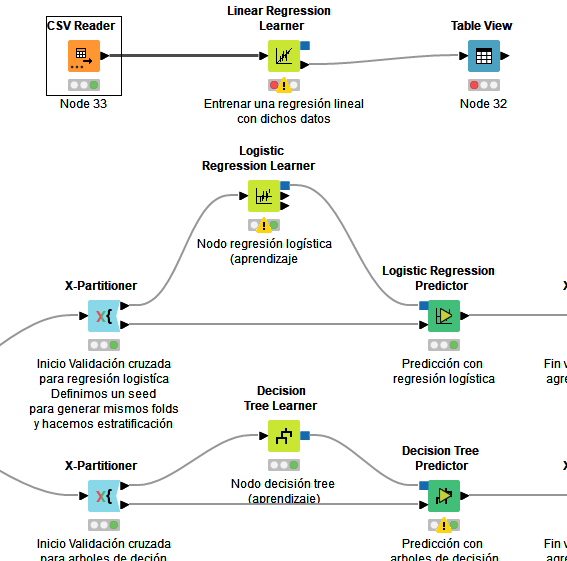
Profesor:

Francesc Busquet

**Introducción**

Durante este trabajo se estudió y analizo una base de datos llamada **adult.csv**, la cual se ordeno, limpio y analizo. En esta base de datos podemos ver que trata de información de personas alrededor del mundo, mostrándonos su nivel de estudios, cuantos ingresos genera, sus estatus personales, género, raza y que tipo de trabajo ejerce.

Luego de estudiar y comprender la base de datos, eh comprendido que debemos tratar los “**Missing Values**” para esto procesaremos los datos y le mostraremos a KNIME que valores tomar en cuenta para así poder llevarlos a los modelos que deseamos investigar.

* **Naïve Bayes**
* **KNN**
* **Regresión logística**

Para este estudio, tomare en cuenta principalmente las variables de **income, education, country, working class y ocupation.**

Después de trabajar en KNIME y probar los distintos modelos y como van funcionando con los nodos y leyendo la información estadística que obtenemos de estos. Podemos comprender por qué hay modelos más adecuados que otros dependiendo la información que estemos procesando. Para este caso nos damos cuenta que el modelo de una **regresión lineal** es poco completo, no hay variables binarias y esto le da poca precisión a esta base de datos. La **regresión logarítmica** funciona bastante bien ya que podemos elegir varias categorías o variables para estudiar y aumentar así la potencialidad de aprendizaje de la maquina dando mayor asertividad debido a que maneja más información. Por ultimo con el modelo de KNN, este busca los k puntos más cercanos a un punto en concreto para poder inferir su valor más comúnmente en problemas de clasificación o de regresión.

**Conclusión**

En Conclusión durante esta actividad hemos desarrollado las habilidades para procesar datos y limpiarlos, esto para que el computador o la plataforma KNIME entienda lo que deseamos procesar y poder llevarla así a modelos de aprendizaje, de los cuales además eh aprendido que depende del problema que tengamos cual deseamos usar, durante esta actividad me parece que el modelo más correcto para elegir es una regresión logarítmica, tiene buena asertividad en manejar múltiples variables que fue a los que nos vimos enfrentados en adult.csv.

Ha sido un trabajo complejo y cometí muchos errores, pero esto me llevo a profundizar y estudiar más el contenido para comprenderlo de una mejor manera, un desafío el cual queda un largo camino por enfrentar.