Laboratorio 6

**Encoding:**

No hubo una necesidad de codificar las variables ya que las que elegimos ya venían codificadas o eran cuantitativas.

**Balanceo del dataset:**

En el dataset de LOL, utilizamos valuecount() de pandas para balancear las variables en valores de 0 y 1. Se llegó a la conclusión de que si está balanceado por lo que así se quedó. Por otro lado, en el dataset de FIFA la variable objetivo era cuantitativa por lo que no se hizo nada.

**Selección de las variables:**

* Lol:
  + blueWins: variable objetivo
  + blueKills: ventaja de habilidad de un equipo en relación a otro, al mismo tiempo son las muertes de los rojos
  + bueDeaths: representa la habilidad de los rojos sobre los azules y son las kills de los rojos
  + blueTowersDestroyed: vulnerabilidad de los azules
  + redTowersDestroyed: vulnerabilidad de los azules
  + blueTotalGold: representa la capacidad de comprar items
  + redTotalGold: representa la capacidad de comprar ítems
* Fifa:
  + Age
  + Overall
  + Value
  + Wage
  + Special
  + Acceleration
  + Dribbling
  + Finishing
  + Agility
  + Ball control
  + Omposure
  + Potential

**Métrica de desempeño:**

* Lol:

Ya que la predicción se hará con valores booleanos, la métrica se calculará con un accuracy de cuantas veces acertó el valor deseado.

* Fifa:

Como se utilizará la media del sector que seleccione el árbol, la mejor métrica es la media R2.

**Comparación:**

¿Qué implementación fue mejor? ¿Por qué?

* Lol:

Ambas implementaciones tuvieron un accuracy relativamente igual, por lo que con tal de realizar bien el DTC, las probabilidades de acertar aumentan.

* Fifa:

Lo más posible es que la manera de predicción de sklearn es un poco distinta a nuestra implementación, ya que el r2 de sklearn fue mucho menor.