

## 1 Descripción del problema.

## 2 Datos.

## 3 Herramientas.

Las herramientas que se pueden usar para abordar la resolución del problema es amplia. Es por ello, que resulta difícil elegir la más adecuada. Tras un estudio de las bibliotecas disponibles para los distintos lenguajes de programación que domino, me selecciono weka(java),sklearn(python) y R. Para determinar cual de las 3 es la más conveniente en este caso, realizaré pruebas de ejecución de los diferentes algoritmos y compararemos el tiempo de ejecución y el uso de memoria. Además, se debe tener en cuenta la facilidad para tratar los datos tanto para su lectura, como para la generación de los archivos con las predicciones.

Los resultados obtenidos se muestran en las siguientes gráficas:

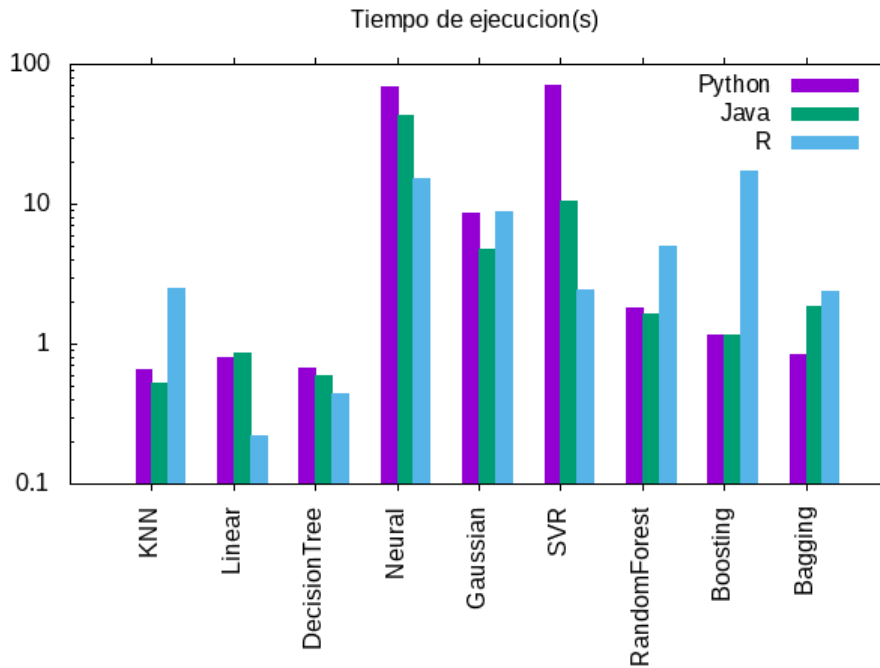


Figure 1: Test algoritmos:Tiempo de ejecucion

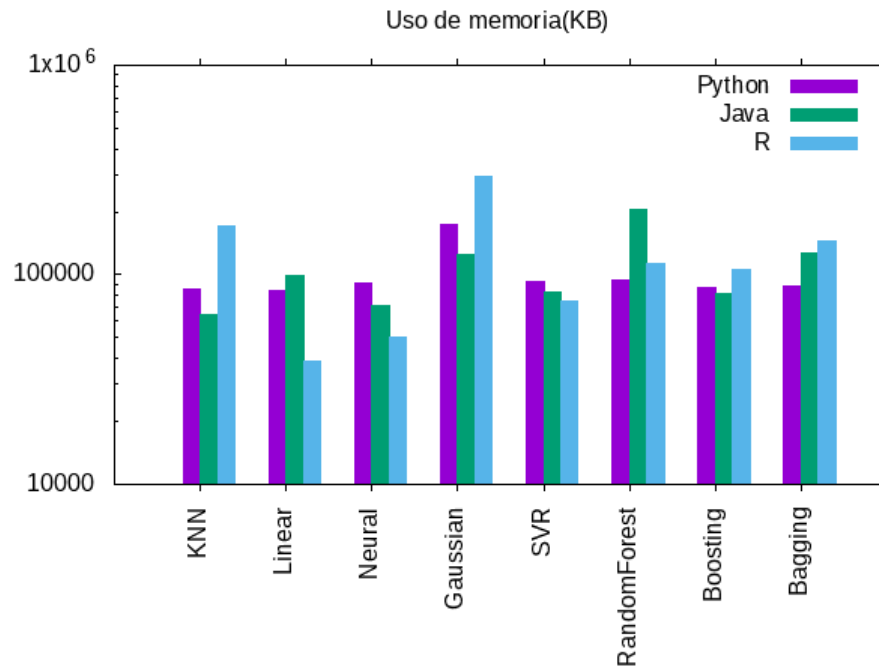


Figure 2: Test algoritmos:Memoria usada

Observando los resultados podemos concluir que sklearn y weka ofrecen (en media) un rendimiento similar, mientras que R ofrece un mejor rendimiento para los algoritmos básicos pero es inferior en la ejecución de multclasificadores. Por otro lado, R y sklearn permiten tratar los datos de forma cómoda. Otra cuestión a tener en cuenta, es la buena documentación con la que cuenta sklearn.

En conclusión, el lenguaje a usar será Python ya que permite un manejo de los datos flexible y un código legible propio de este lenguaje. Esta decisión se apoya en que en términos de rendimiento no hay una alternativa "mucho mejor".

En particular, se usaran las siguientes bibliotecas:

- NumPy.
- Pandas.
- Scikit Learn.
- XGBoost.

## 4 Algoritmos.

**Nota:** Para la ejecución de los algoritmos los datos nominales se han pasado a numerico.

A priori desconocemos que algoritmo se adapta mejor a nuestro problema, es por ello que realizaremos un estudio comparando varios algoritmos. Los algoritmos elegidos son los siguientes:

- algoritmos básicos como KNN, Linear Regression y Árboles de regresión
- algoritmos "potentes" como Neural Network, Gaussian y SVM
- multclasificadores como Random Forest, Boosting y Bagging

En la siguiente gráfica se muestran los resultados obtenidos.

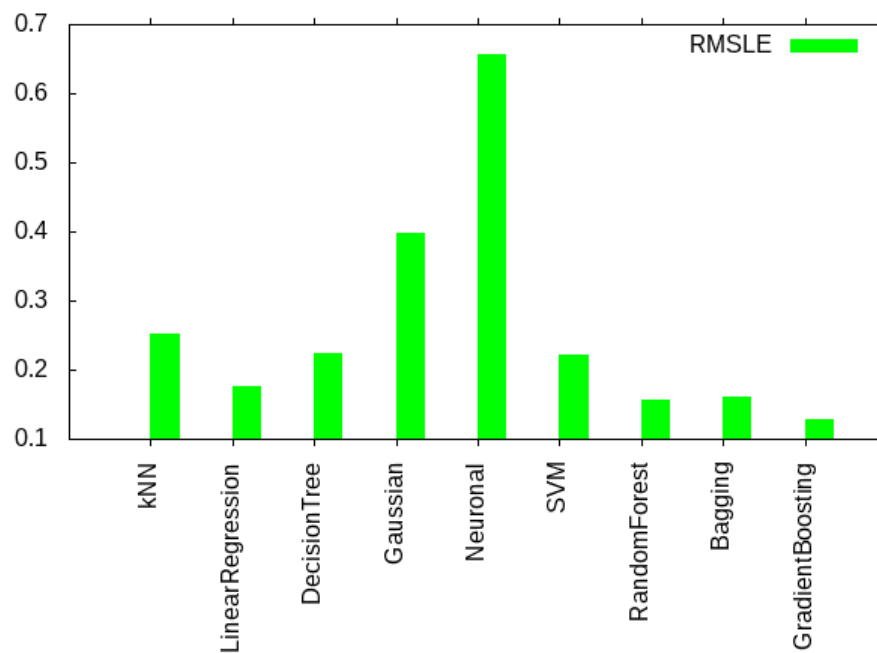


Figure 3: Test algoritmos:Error.

Se puede observar que los algoritmos que mayor precisión proporcionan en este caso son los multclasificadores. Por tanto, al igual que los competidores, nos decantamos por esta opción.