**Análisis y Reporte sobre el Desempeño del Modelo**

Russel Rosique Rodríguez

A01283727

En esta entrega intermedia, se utilizó el dataset de kaggle llamado “Students performance in exams”. Este dataset consiste de 8 columnas: “gender”, “race/ethnicity group”, “parental level of education”, “lunch”, “test preparation course”, “math score”, “writing score”, “reading score”. Se decidió implementar un modelo de aprendizaje para este dataset, de tal forma que el modelo pudiera predecir el promedio de los tres puntajes de un alumno, en base a el resto de las columnas. Se consideró a este conjunto de datos como apto para la implementación ya que consiste de un total de 1000 datos únicos, y registrados de forma completa.

Separación y Evaluación del Modelo

Como modelo, se eligió utilizar la herramienta de random forests, ya que es un gran modelo de aprendizaje, además que cuenta con diversas ventajas. Para evaluar este modelo con el conjunto de datos, se realizó un split de los datos en un conjunto de entrenamiento y uno de evaluación. Se tomaron el 80% de los datos de forma aleatoria para el conjunto de entrenamiento, y el 20% restante para el conjunto de evaluación. Una vez hecho esto, se implementó el modelo con el algoritmo de random forest con los siguientes parámetros:

Posterior a esto se realizó la evaluación del modelo con ambos conjuntos de datos (entrenamiento y prueba). Con el conjunto de entrenamiento, se obtuvieron las siguientes métricas:

MSE= 121.92537893770688

MAE= 8.909239624927874

R2= -0.7067617240634501

Por otro lado, con el conjunto de prueba se obtuvieron las siguientes métricas:

MSE= 215.06806740794644

MAE= 11.36903290962244

R2= -1.9521491364554615