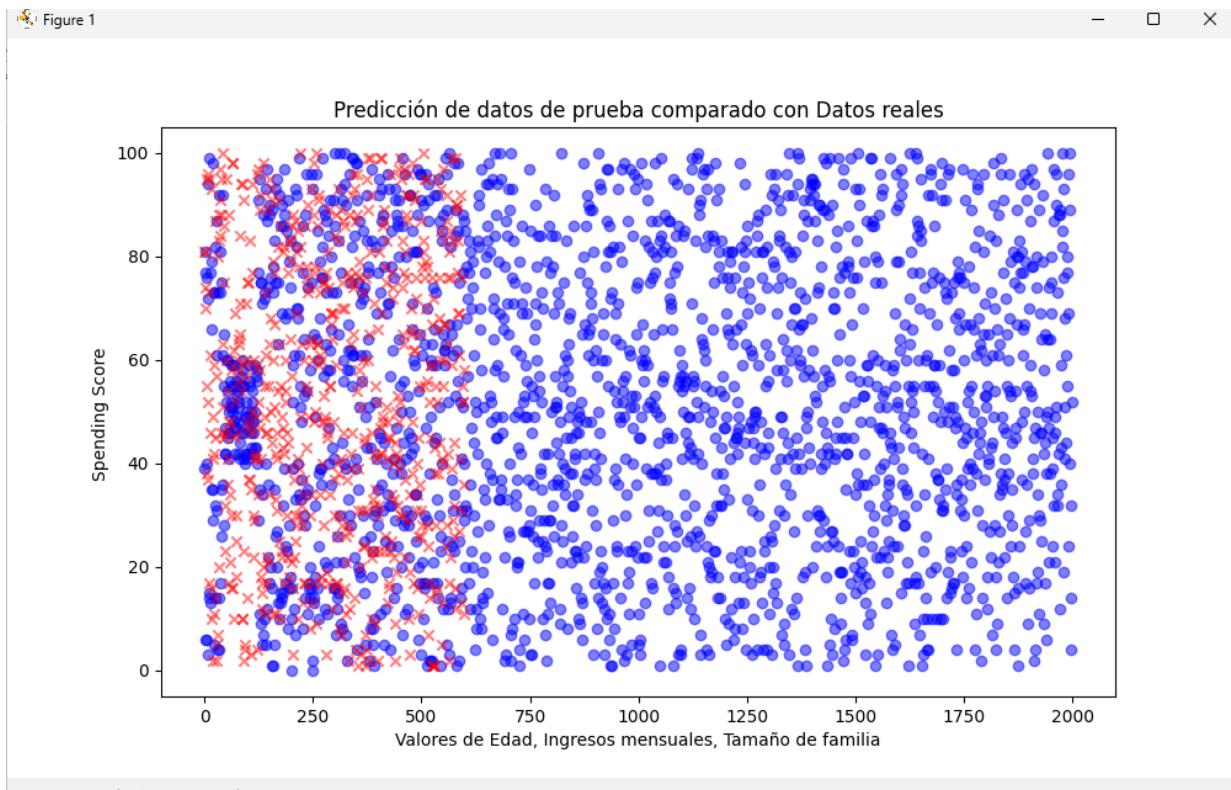


Esta es la gráfica de árbol. Esta se tuvo que mostrar limitando la cantidad de datos que recibe el modelo, ya que al ser alrededor de 2 mil datos, el diagrama puede incluso no mostrarse.

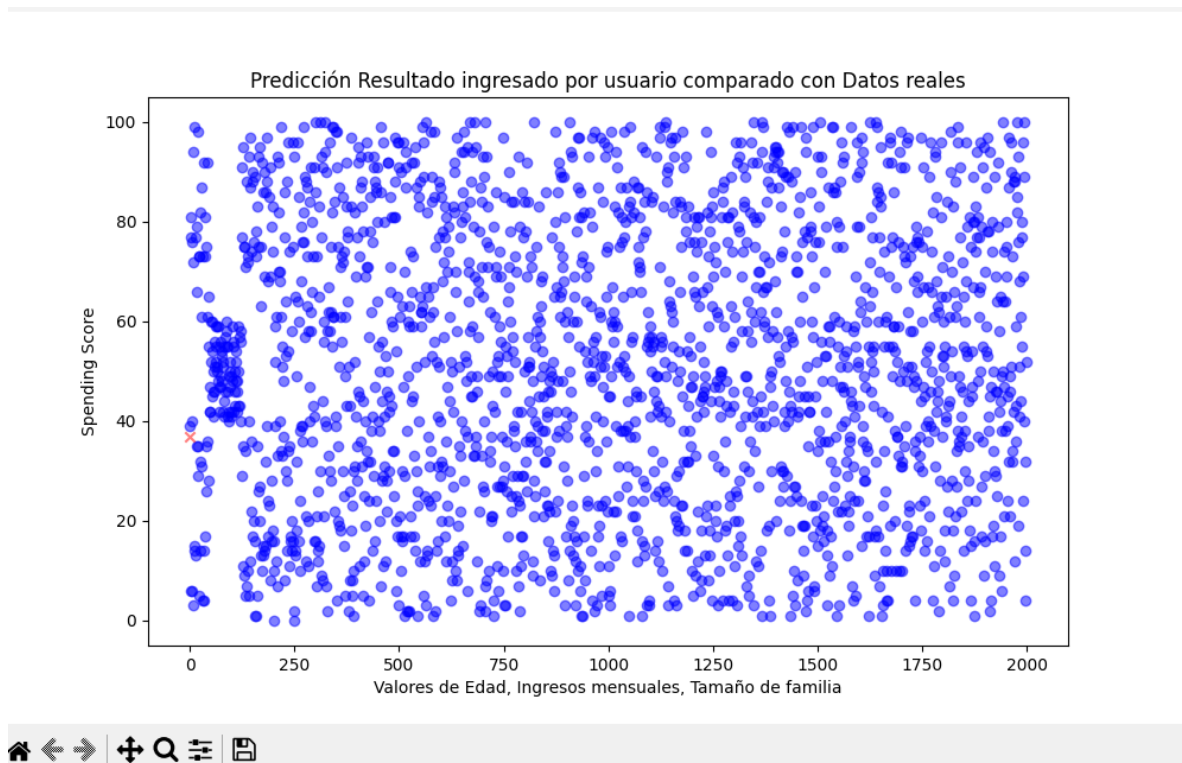
Vemos que dan las posibles caminos que puede tomar los valores. Vemos que hay un dato que aparece como “Gini” el cual es un indicador que nos permite saber qué tan “puros” son los datos que hay. Un Gini cercano o igual a 1 significa que hay mucha impureza en los datos, cercano o igual a cero es que los datos son puros o se distribuyen de forma uniforme entre las clases. Por ejemplo, en la raíz vemos que tenemos un Gini alto, 0.72, el cual ya nos puede decir que hay datos que no están bien distribuidos. Mientras que en las ramificaciones vemos que el Gini es menor.

Por lo tanto, se podría concluir que el modelo necesita revisar si realmente está separando de forma correcta los datos.



Acá comparamos los datos de entrenamiento con los que el modelo predice cuando es entrenado. Las X rojas son los valores que se predijeron, los círculos azules son los datos reales.

En el eje X hay valores que serían los valores de Edad, Ingresos Mensuales, Tamaño de familia (variables explicativas) y en el eje Y vemos al Spending Score, lo que queremos predecir (variable objetivo)



Y en esta última gráfica vemos el resultado predicho por el modelo usando las variables con datos ingresados por el usuario comparado con los valores reales. Las X rojas son los valores que se predijeron, los círculos azules son los datos reales.

En el eje X hay valores que serían los valores de Edad, Ingresos Mensuales, Tamaño de familia (variables explicativas) y en el eje Y vemos al Spending Score, lo que queremos predecir (variable objetivo)

Podemos ver en ambas gráficas que pueden haber una considerable imprecisión de predicción,, no totalmente mala pero sí bastante mejorable ya que hay muchos valores predichos que no concuerdan con los reales.