11/03/2021

Danielle sherman

CTO, Blackwell Electronics

Overview:

As per your request, I have further analyzed the Demographic Data provided. The objectives are as follows: 1. Are there differences in the age of customers between regions? 2. Is there a correlation between age of a customer and if the transaction was online or in-store?

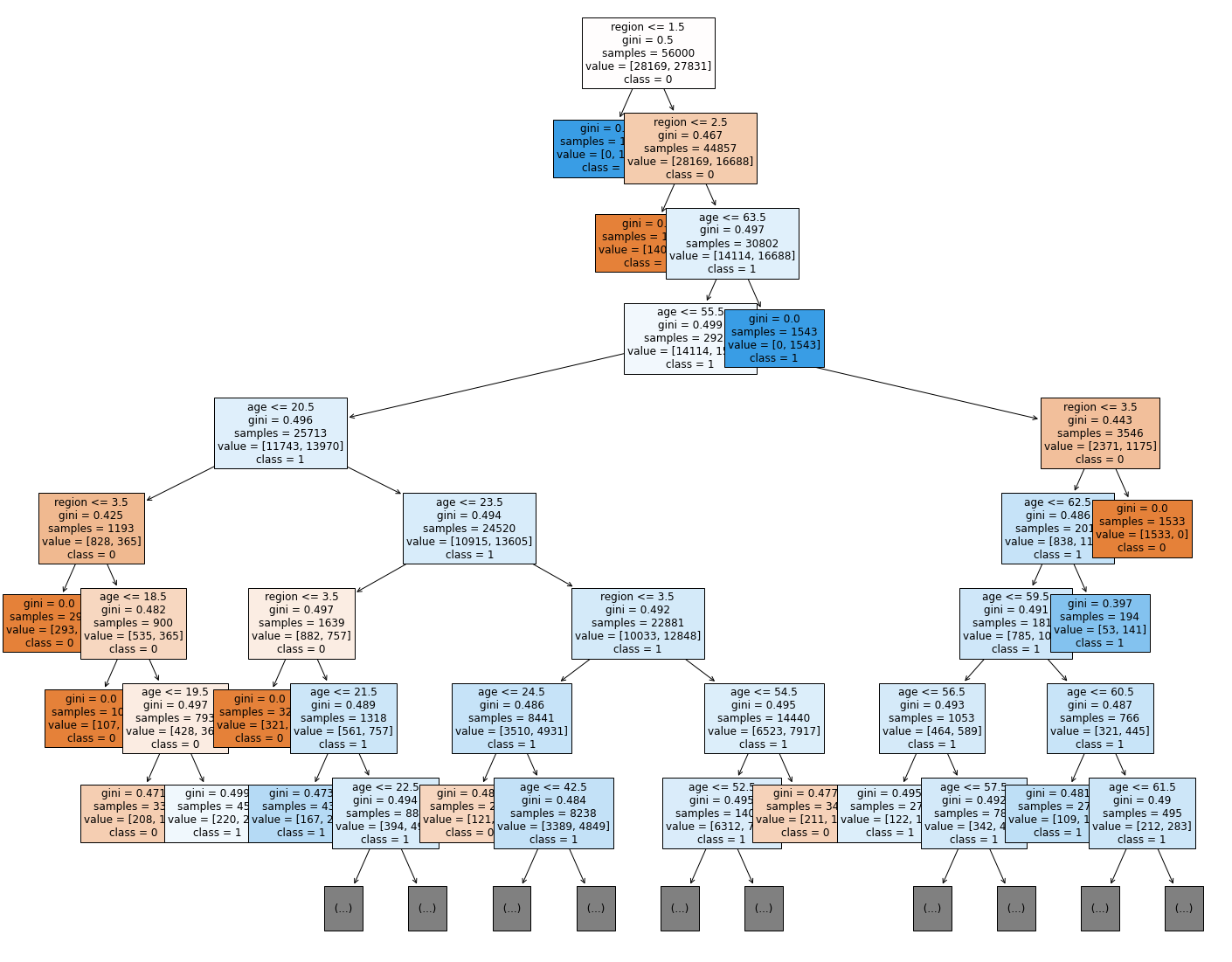
DATA:

As this is an extension of the previous task, I am still utilizing the transaction data you provided. The data contains 80,000 transactions and records the following features: in-store, age, items, amount, and region. In-store is coded 0 for an online purchase and 1 for in stores. Region is also coded as 1,2,3,4 signifying North, South, East, West, respectively.

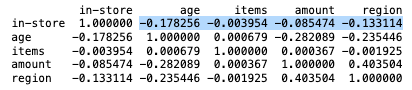
Analysis:

I utilized Random Forest Classifier and Decision Tree Classifier models in an attempt to predict outcomes pertaining to the objectives. Using Age and Region as features and In-store as the dependent variable, I build a reliable model (Decision Tree Classifier 0.7923500202478625).

The modeling aided in determining the strength of the result.



The table below indicates the correlation between features



Conclusion:

Based on my findings, there are no standout differences in the age of customers between regions. The prediction models return values that are less than a 50/50 guess.

There is a very weak correlation between the age of a customer and where the item(s) were purchased. The correlation is not reliable.

In addition, Mark’s hypothesis regarding the age of customers and the amount spent is a stronger correlation than age/purchase location, but still too low to give a definitive correlation. Region is the strongest indicator of whether a sale has been made online or in-store. However, more sales were in-store versus online.

GRANT JOHNSON