Title: Hotel select prediction

Attributes:

Hotel rating(1-5 stars),

Price,

the closest train station to the hotel,

pay first

Rules:

hotel rating >= 3stars,

Price <= 2000,

Closest station <= 200,

No need to pay first

Stars >= 3

yes

no

Not accept

accept

Not accept

Need to pay first

Not accept

Closest station <= 200

Not accept

Price <= 2000

yes

no

yes

no

yes

no

Data:

I use numpy to generate numbers randomly for the attributes, And I get the the result of accept by using if-condition. Build them as a train data csv file. Then I do the same thing to build a test data csv file.

A decision tree generated by sklearn:

tree.pdf

Accuracy:

In this model, I did not limit the max depth of the decision tree. The max depth of this tree is 4 and the accuracy is 0.994. I discovered that the accuracy decreases when the max depth of the decision tree gets shorter. When depth is 3, the accuracy decreases to 0.976. When depth is 2, the accuracy decreases to 0.954. However, when depth is 1 only, the accuracy remains at 0.954. probably it is because my attributes and rules are few, so that the accuracy is so high and the different depth of the decision tree does not have many effects on it.