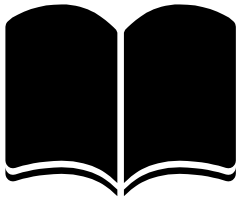
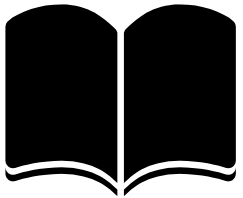
HOMEWORK 2
DECISION TREE

MY PROBLEM



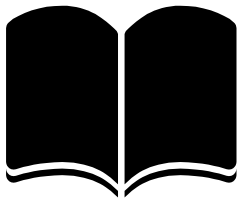
CLASSIFY WHETHER OR NOT A GIVEN PERSON HAS A HEART DISEASE

ATTRIBUTES



1. Age
2. Sex
3. Weight
4. Activity level (exercise)
5. Smoking
6. Family heart disease history

MY RULES



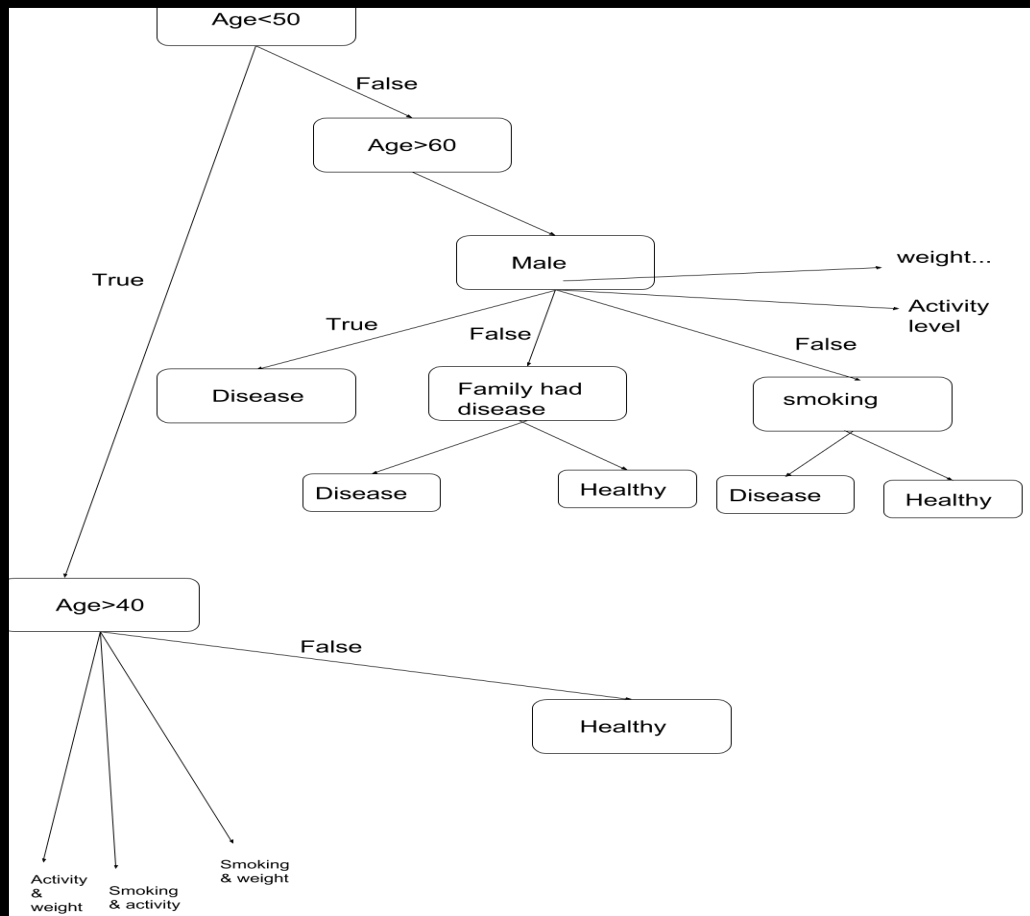
Disease:

1. If age > 60 AND Male
2. If 50 < age < 60 *If Family heart disease history
    - If Smoking
    - If Weight > 100
    - If Activity level = 0
3. If 40 < age < 50
    - If Activity level = 0 AND Weight > 100
    - If Smoking AND Activity level = 0
    - If Smoking AND Weight > 100
4. If Family heart disease history AND Male

Healthy:

1. All other possibilities

MY TREE

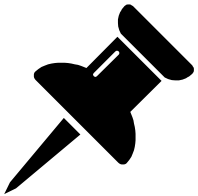

Age<50
False
Age>60
Male
weight...
Activity level
True
True
False
False
Disease
Family had disease
smoking
Disease
Healthy
Disease
Healthy
True
Age>40
False
Healthy
Activity & weight
Smoking & activity
Smoking & weight

PRODUCED TREE

COMPARISON

Of the hand made, and generated trees

PRIORITIES



Hand-generated tree priorities

Age > Sex > Weight/Activity level/Smoking/Family history

Code-generated tree priorities

Sex > Age/Family history > Age > Weight/Age > Smoking/Age

- I tried to put emphasis on the attributes such as sex, age, and family heart disease history, and it was reflected in the code-generated tree, by being on top levels of the tree, and having large impact on the result.
- The secondary attributes - smoking/weight/activity level, were used for creating rules for more specific classification. By creating these statistics, I hoped to reduce underfitting, and I believe I largely succeeded.
- The code-generated tree also put the same amount of priority to these secondary statistics as I did, since they appear at 4th+ level of the tree.

CONCLUSION

94.4%

It seems that the rules that I was thinking about during the creation of 'right' data, were very similar to the rules which code-generated tree used.

code-generated decision tree was successful at identifying the trends in data, and creating the rules that dictate how each person should be classified

Thanks!