

Evaluation

Arnold Jiadong Yu

March 8, 2019

Result:

```
10-Fold on removeMissing_cat
Accuracy : [0.8080875041431886, 0.802784222737819, 0.809413324494531, 0.8017898574743122,
0.8100762346702022, 0.8097447795823666, 0.8090818694066954, 0.8213457076566125, 0.8133907855485581,
0.8024527676499834]
Average Accuracy : 0.808816705336427
10-Fold on replaceMissing_cat
Accuracy : [0.8105618667485416, 0.8047282775560332, 0.8096407737181456, 0.803193122505373,
0.8077985876573534, 0.8056493705864293, 0.8084126496776174, 0.8185446730119742, 0.815781393920786,
0.8041142155357691]
Average Accuracy : 0.8088424930918021
10-Fold on removeMissing
Accuracy : [0.8034471329134902, 0.802784222737819, 0.8100762346702022, 0.8044414981769971,
0.8080875041431886, 0.8186940669539278, 0.811070599933709, 0.8167053364269141, 0.8239973483592973,
0.8104076897580378]
Average Accuracy : 0.8109711634073582
10-Fold on replaceMissing
Accuracy : [0.8065704636168253, 0.802886091495241, 0.8130181148295977, 0.8068774946269573,
0.8120970217992017, 0.8142462388701259, 0.8136321768498619, 0.8176235799815781, 0.8243782622044826,
0.816088424930918]
Average Accuracy : 0.8127417869204792
```

Apply the training data files to weka data mining suite, and gain an accuracy of 0.818. Since our result uses train and test data together, it is not bad at all. The results are acceptable.

In order to improve the result, we can apply laplacian to all categorical attributes, some categories showed in the training set may not show up in the testing set. This will cause inaccuracy.

Instruction on compiling and running program:

IDE: Anaconda-Spyder with python 3.6

1. Open k-foldCross.py
2. Set to current directory (where the k-foldCross.py is).
3. Execute the program.