

Assignment 1

Import packages

```
In [1]: import pandas as pd
import numpy as np
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.ensemble import AdaBoostClassifier
from sklearn.ensemble import BaggingClassifier
from sklearn.ensemble import ExtraTreesClassifier
from sklearn.model_selection import train_test_split
from autorank import autorank, create_report, plot_stats
```

Read data

```
In [2]: ar = pd.read_csv('arrhythmia.csv')
ca = pd.read_csv('caesarian.csv')
wb = pd.read_csv('website-phishing.csv')
```

Arrhythmia

```
In [3]: ar.drop('J', axis=1, inplace=True)
ar.replace('?', np.nan, inplace=True)
ar.dropna(inplace=True)
y = ar['class']
X = ar.drop('class', axis=1)

RANDOM_STATE = 1234
np.random.seed(RANDOM_STATE)
df = pd.DataFrame(columns=['DecisionTree', 'RandomForest', 'AdaBoost', 'Bagging'])
for x in np.random.rand(10):
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=x, random_state=RANDOM_STATE)

    dt = DecisionTreeClassifier(random_state=RANDOM_STATE)
    dt.fit(X_train, y_train)

    rf = RandomForestClassifier(max_depth=10, random_state=RANDOM_STATE)
    rf.fit(X_train, y_train)

    ab = AdaBoostClassifier(random_state=RANDOM_STATE)
    ab.fit(X_train, y_train)

    bagging = BaggingClassifier(random_state=RANDOM_STATE)
    bagging.fit(X_train, y_train)

    et = ExtraTreesClassifier(random_state=RANDOM_STATE)
    et.fit(X_train, y_train)

    df = df.append({'DecisionTree':dt.score(X_test, y_test), 'RandomForest':rf.score(X_test, y_test), 'AdaBoost':ab.score(X_test, y_test), 'Bagging':bagging.score(X_test, y_test)})
```

```
In [4]: result = autorank(df)
print(result)
print(df)
```

```
RankResult(rankdf=
              meanrank    median    mad    ci_lower    ci_upper    effect_size \
```

Bagging	1.50	0.702358	0.087555	0.550868	0.803419	0.0
ExtraTrees	2.00	0.682712	0.064756	0.565757	0.752137	0.255126
RandomForest	2.65	0.680737	0.074962	0.563275	0.735043	0.265281
AdaBoost	4.15	0.624769	0.04542	0.563275	0.666667	1.112455
DecisionTree	4.70	0.604606	0.039609	0.183623	0.679012	1.438564

magnitude
 Bagging negligible
 ExtraTrees small
 RandomForest small
 AdaBoost large
 DecisionTree large

pvalue=3.885067488345842e-06

cd=1.9288111473713958

omnibus=friedman

posthoc=nemenyi

all_normal=False

pvals_shapiro=[0.0004982678801752627, 0.07376586645841599, 0.1636943072080612, 0.5802350640296936, 0.57322096824646]

homoscedastic=True

pval_homogeneity=0.6385993426914864

homogeneity_test=levene

alpha=0.05

alpha_normality=0.01

num_samples=10

posterior_matrix=

None

decision_matrix=

None

rope=None

rope_mode=None

effect_size=akinshin_gamma)

	DecisionTree	RandomForest	AdaBoost	Bagging	ExtraTrees
0	0.679012	0.728395	0.654321	0.777778	0.703704
1	0.625954	0.717557	0.656489	0.725191	0.709924
2	0.641304	0.722826	0.663043	0.766304	0.733696
3	0.618182	0.627273	0.606061	0.663636	0.648485
4	0.573171	0.621951	0.600610	0.652439	0.646341
5	0.582609	0.721739	0.643478	0.756522	0.747826
6	0.606838	0.735043	0.666667	0.803419	0.752137
7	0.602374	0.643917	0.599407	0.679525	0.661721
8	0.183623	0.563275	0.563275	0.550868	0.565757
9	0.491848	0.595109	0.573370	0.586957	0.614130