

# Classifiers Optimization Evaluation

Jien Wei, Lim (100276935)

November 21, 2022

## 1 Introduction

This report summarizes the optimization and evaluation of different classifiers with different parameters using 5-fold cross validation. The classifiers used are Decision Trees (DT), Random Forest (RF), and K Nearest Neighbor (KNN) Classifiers.

## 2 Process

Before any evaluation can be done, arrays are defined to hold the parameters to be used for the classifiers:-

```
# Array for DT classifier parameters
dt_classifier_params = [2, 3, 5, 7, 10]
# Array for RF classifier parameters
rf_classifier_params = [100, 200, 500]
# Array for KNN classifier parameters
knn_classifier_params = [[1, 11, 21, 31, 51],
                        ["euclidean", "manhattan"]]
```

To evaluate and optimize the classifiers, their respective array is iterated to set the parameters of each classifier and 5-fold cross validation is performed on them to obtain a list of accuracy scores. The median of the accuracy scores is then obtained as the overall score for the classifier which are then stored in another array to be plotted. The model with the best score will be trained to predict the test data and obtain the accuracy.

Table 1: Classifier Parameters and 5-Fold Cross Validation Accuracy Scores

Decision Tree		Random Forest		KNN	
Max Depth	Accuracy	Estimators	Accuracy	Estimators+Metric	Accuracy
2	0.9367	100	0.9620	1+Euclidean	0.8875
3	0.9250	200	0.9500	1+Manhattan	0.9125
5	0.9494	500	0.9375	11+Euclidean	0.9125
7	0.9494			11+Manhattan	0.925
10	0.9494			21+Euclidean	0.925
				21+Manhattan	0.9375
				31+Euclidean	0.9000
				31+Manhattan	0.9250
				51+Euclidean	0.9125
				51+Manhattan	0.9125

Table 2: Best-parameter Classifiers with Test Data Accuracy Score

Decision Tree		Random Forest		KNN	
Max Depth	Accuracy	Estimators	Accuracy	Estimators+Metric	Accuracy
5	0.9006	100	0.9474	21+Manhattan	0.9240

