LAB3 AI REPORT

MADE BY:

Ahmad Said Nouh (7086)

Ahmed Alaa Tawfik (6842)

Hussein Mourad Kassem (6729)

Hossam Eldin Ahmed Eissa (6747)

Contents

	Predicted (h)	Predicted (g)
Actual (h)	TN	FP
Actual (g)	FN	TP

Algorithms:

Decision Tree:

Confusion Matrix:

	Predicted h	Predicted g
Actual h	1594	412
Actual g	422	1584

```
model_accuracy =  0.792123629112662
model_precision =  0.7935871743486974
model_recall =  0.7896311066799602
model_specificity =  0.7946161515453639
model_f1 =  0.7916041979010495
```

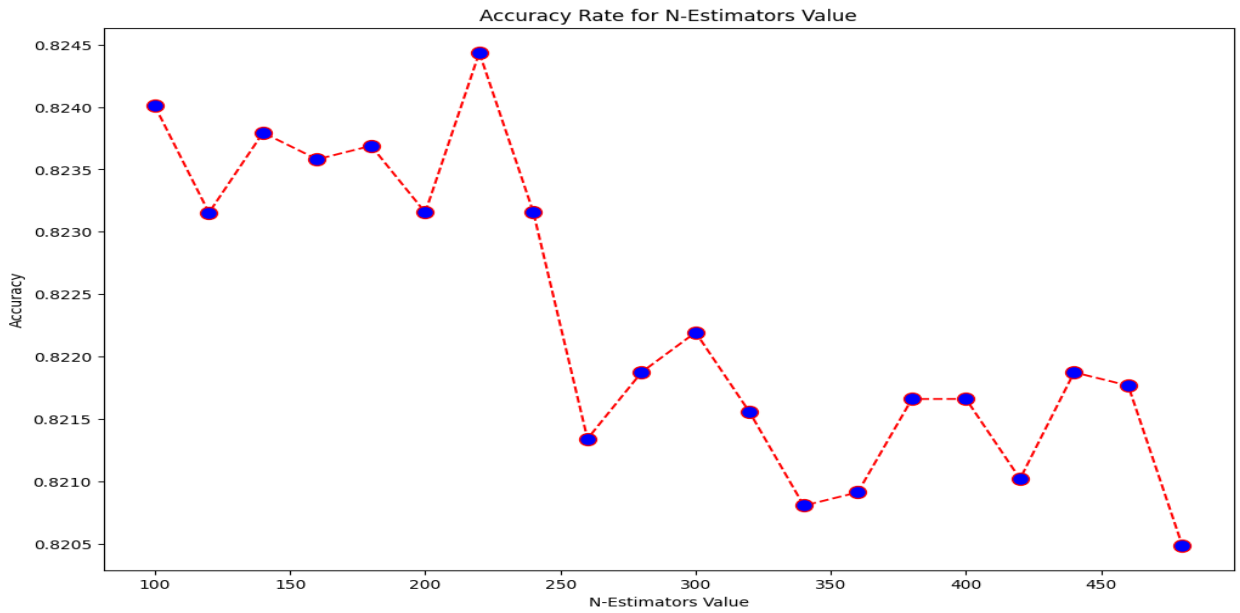
AdaBoost:

Parameter tuning:

1- Cross validation with range(1, 50000) with step 1000

2- Cross validation with range(1, 1000) with step 30

3- Cross validation with range(100, 500) with step 20



Therefore, n_estimators = 220 for this run.

Confusion Matrix:

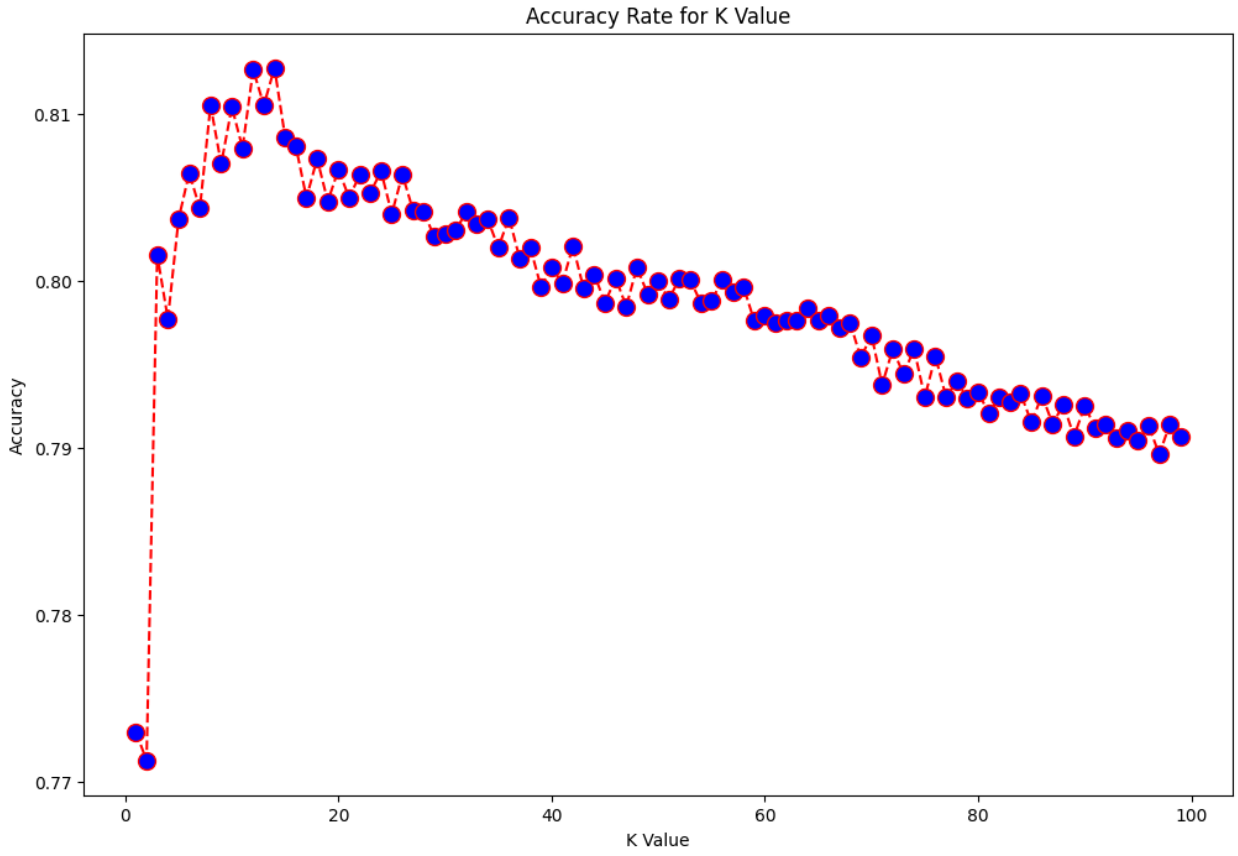
	Predicted h	Predicted g
Actual h	1668	338
Actual g	360	1646

```
model_accuracy =  0.8260219341974078
model_precision =  0.8296370967741935
model_recall =  0.8205383848454636
model_specificity =  0.8315054835493519
model_f1 =  0.825062656641604
```

K-NN:

Parameter tuning:

1- Cross validation with range(1, 60) with step 1



Therefore, k = 14 for this run.

Confusion Matrix:

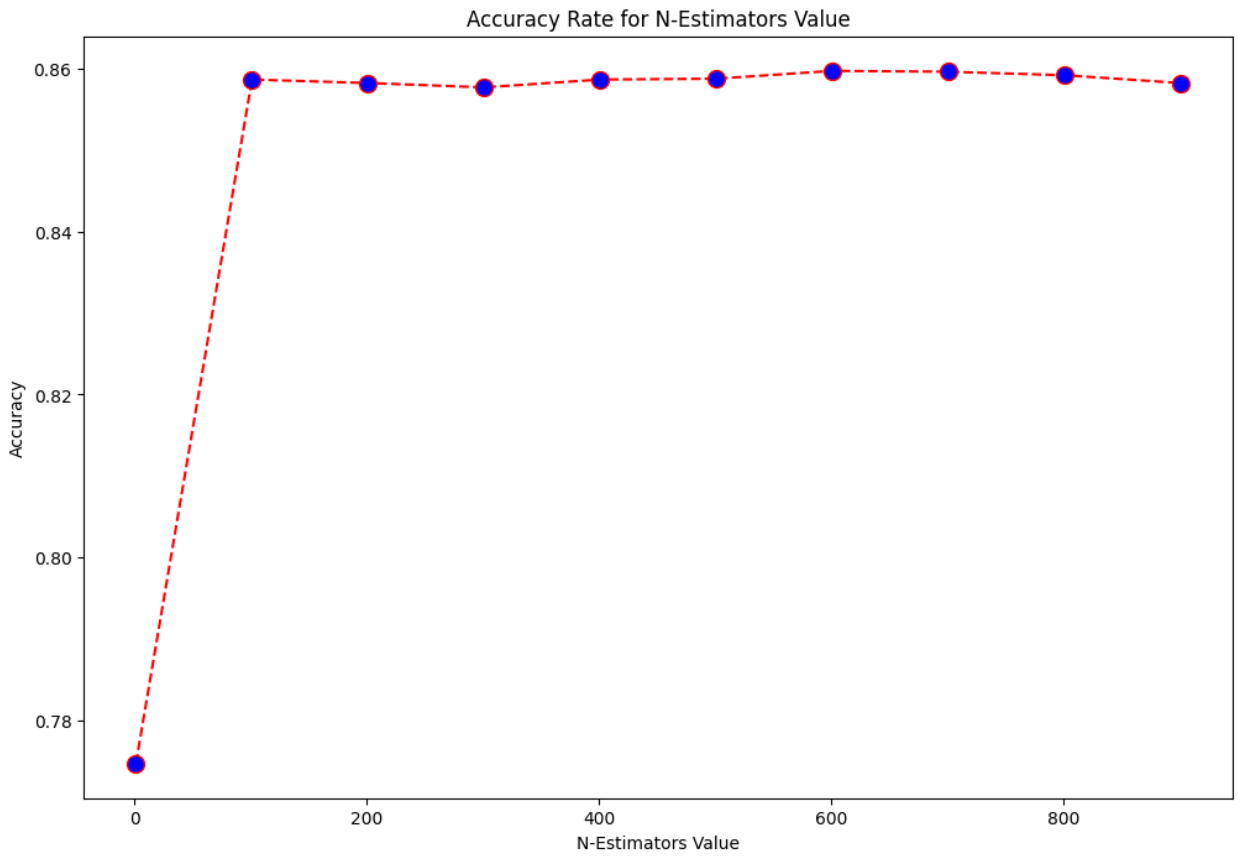
	Predicted h	Predicted g
Actual h	1560	446
Actual g	299	1707

```
model_accuracy =  0.8143070787637089
model_precision =  0.7928471899674873
model_recall =  0.8509471585244267
model_specificity =  0.7776669990029911
model_f1 =  0.8208704015388315
```

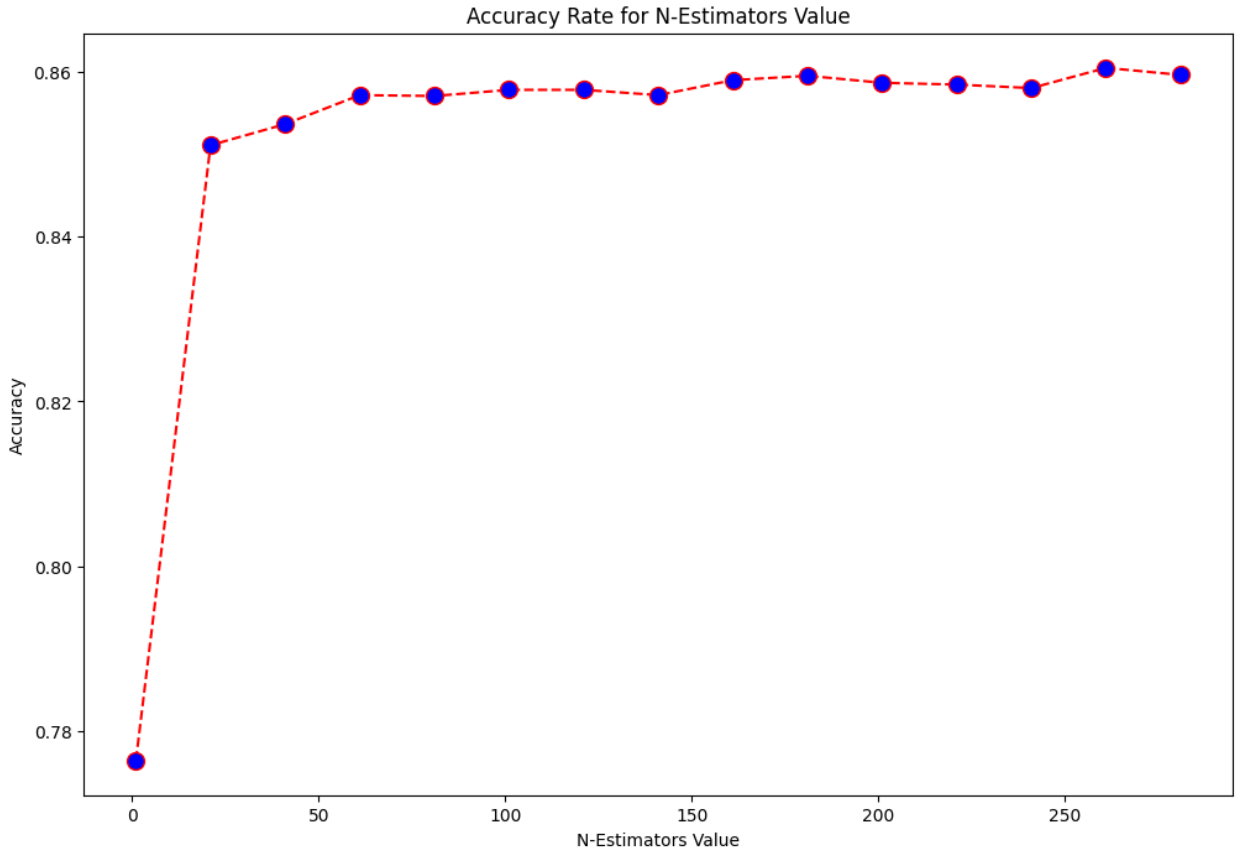
Random Forests:

Parameter tuning:

1- Cross validation with range(1, 1000) with step 100



2- Cross validation with range(1, 300) with step 20



Therefore, n_estimators = 261 for this run.

Confusion Matrix:

	Predicted h	Predicted g
Actual h	1688	318
Actual g	241	1765

```
model_accuracy =  0.8606679960119641
model_precision =  0.8473355736917907
model_recall =  0.8798604187437687
model_specificity =  0.8414755732801595
model_f1 =  0.8632917583761311
```

Naive Bayes:

Confusion Matrix:

	Predicted h	Predicted g
Actual h	787	1219
Actual g	184	1822

```
model_accuracy =  0.6502991026919243
model_precision =  0.5991450180861558
model_recall =  0.9082751744765702
model_specificity =  0.39232303090727816
model_f1 =  0.7220130770754905
```

Comment:

The objective of modeling this dataset is to classify a new event as either a gamma (signal) or a hadron (background). It is clearly stated in the objective of this modelling that simple classification accuracy is not meaningful for this data, since classifying a background event as signal is worse than classifying a signal event as background. Since precision is the percentage of positive predictions that are correct, it is exactly the indicator that we need. Therefore, we will judge the classifier model on its precision and select the classifier with the highest precision, if tie occurred then on the highest accuracy.

From the above information, we will find the precisions of each model is as follows:

1- decision_tree → model_precision = 0.7935871743486974

2- ada_boost → model_precision = 0.8296370967741935

3- knn → model_precision = 0.7928471899674873

4- random_forests → model_precision = 0.8473355736917907

5- naïve_bayes → model_precision = 0.5991450180861558

Therefore, **random_forests** classifier is the best model for this dataset.