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:

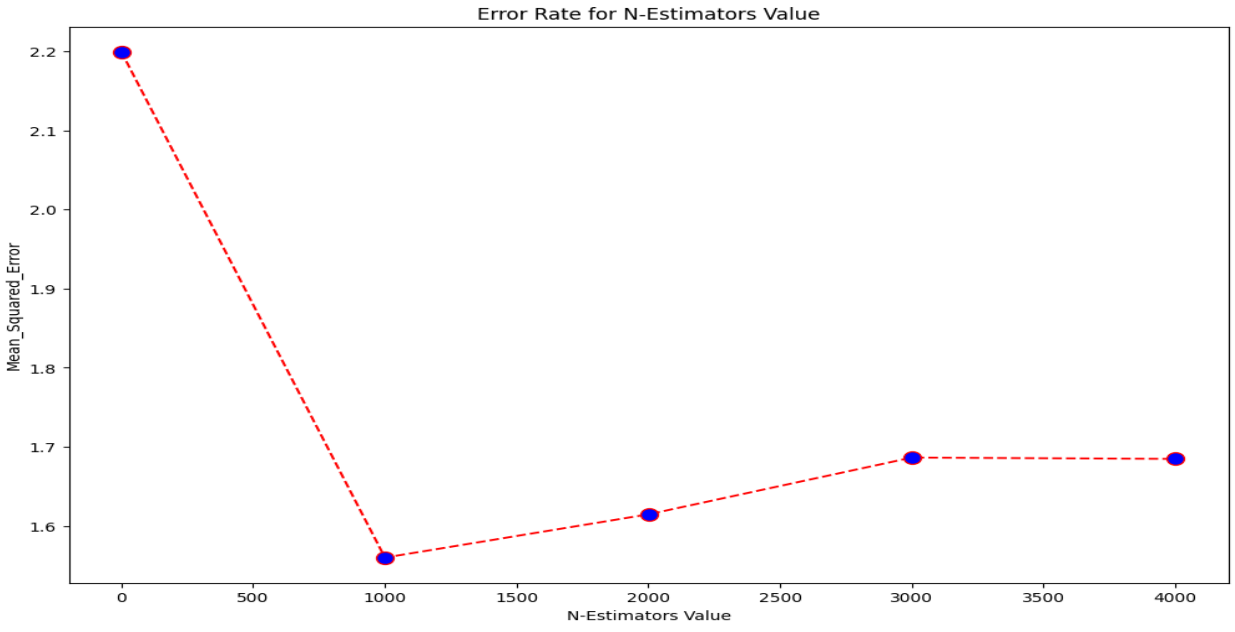
	Predicted h	Predicted g
Actual h	1582	424
Actual g	400	1606

```
model_accuracy =  0.7946161515453639
model_precision =  0.7911330049261084
model_recall =  0.8005982053838484
model_f1 =  0.7958374628344895
```

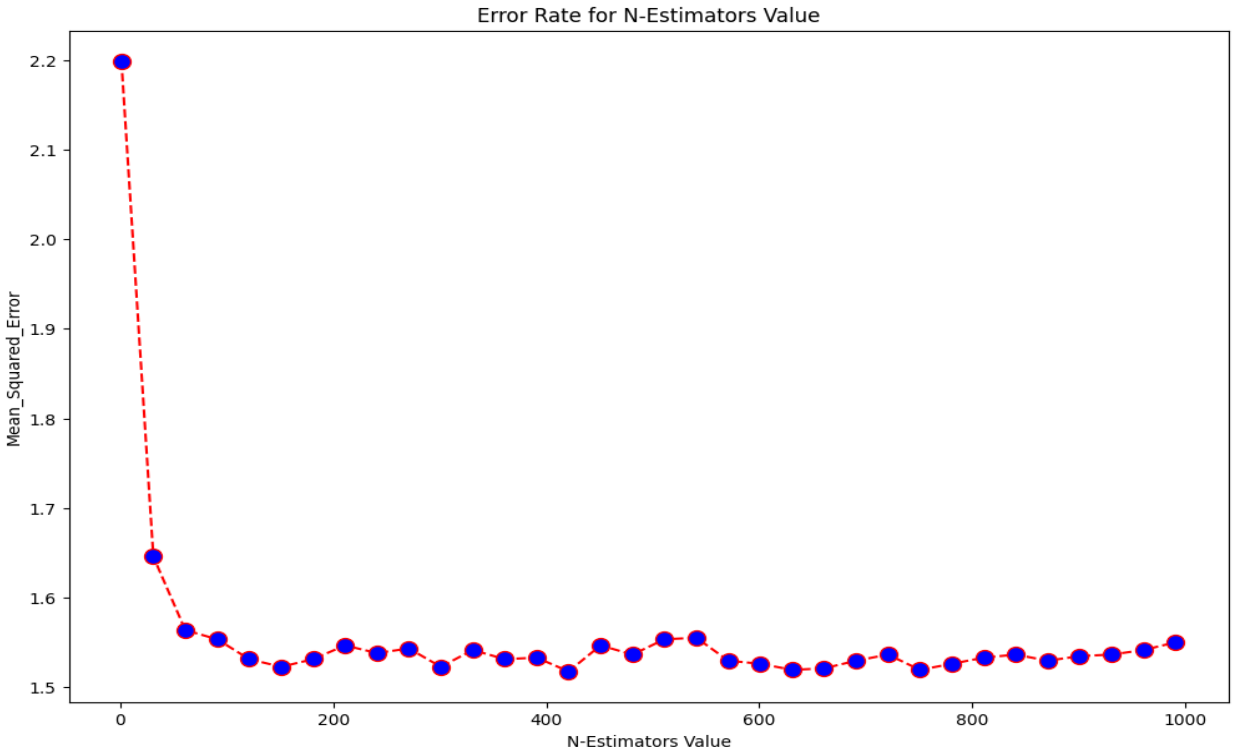
AdaBoost:

Parameter tuning:

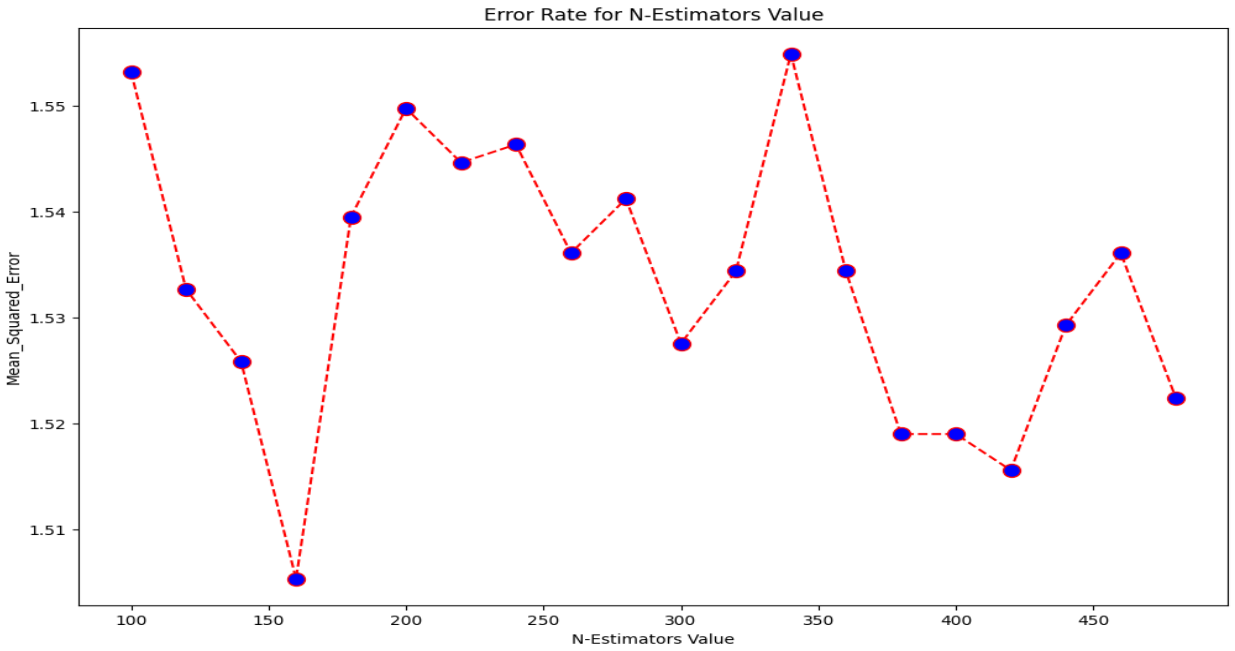
1- Cross validation with range(1, 50000) with step 1000



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



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



Therefore, n_estimators = 160 for this run.

Confusion Matrix:

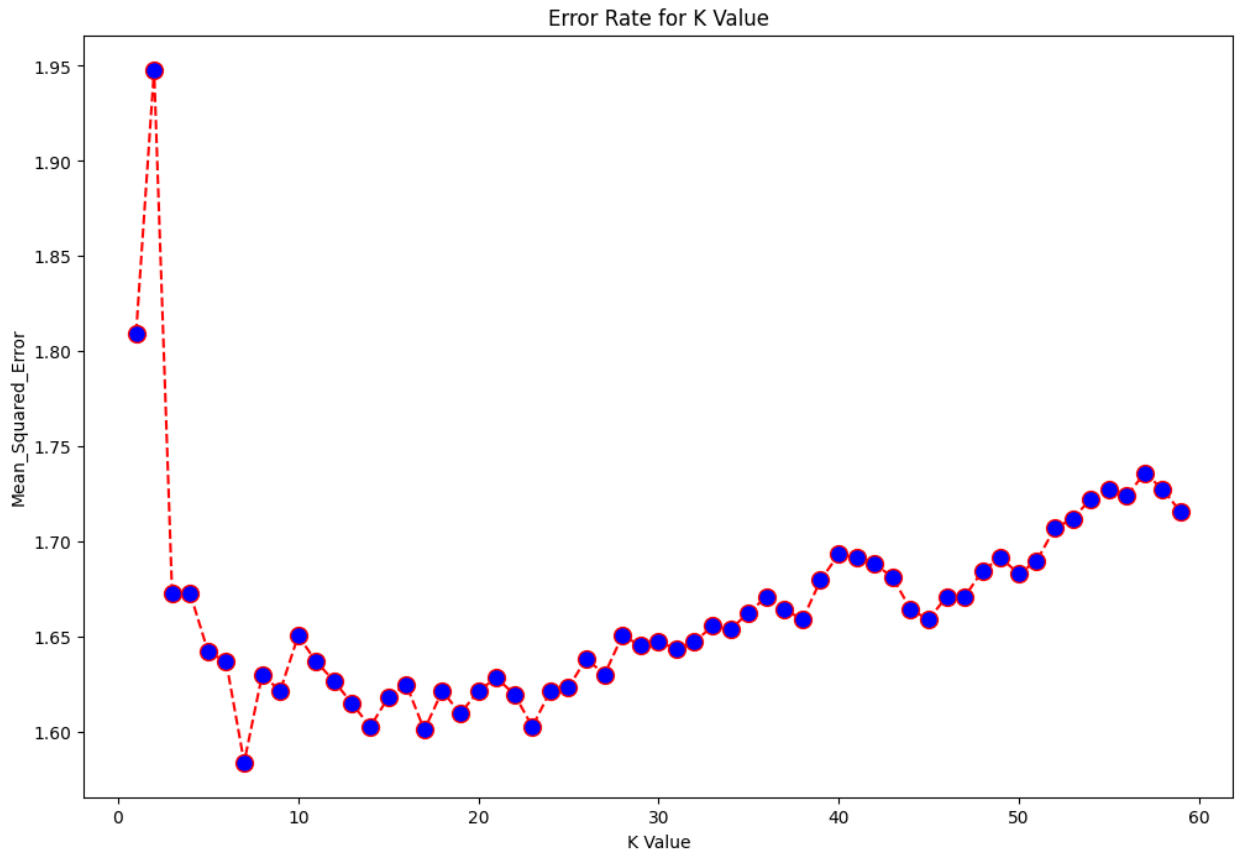
	Predicted h	Predicted g
Actual h	1656	350
Actual g	389	1617

```
model_accuracy =  0.81580259222333
model_precision =  0.8220640569395018
model_recall =  0.8060817547357926
model_f1 =  0.8139944626227031
```

K-NN:

Parameter tuning:

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



Therefore, k = 7 for this run.

Confusion Matrix:

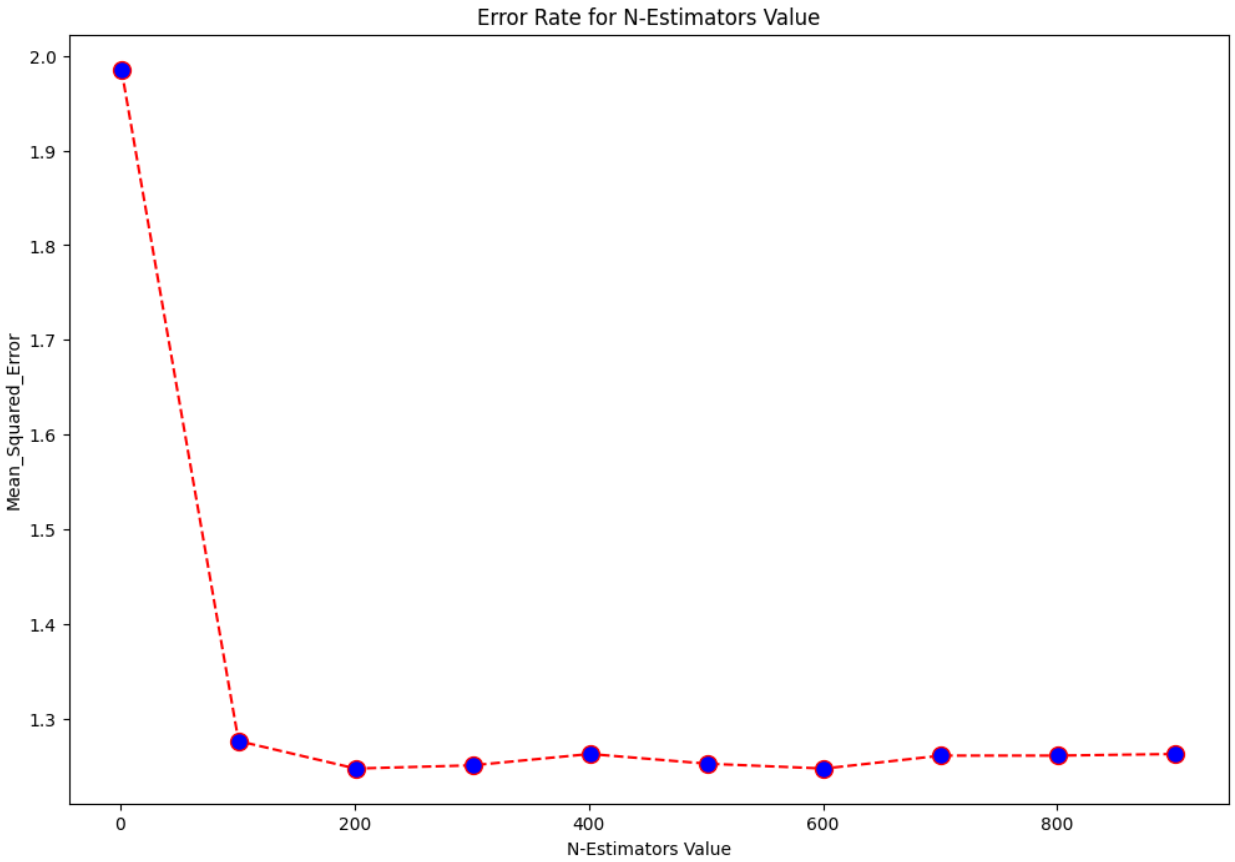
	Predicted h	Predicted g
Actual h	1507	499
Actual g	282	1724

```
model_accuracy =  0.805333998005982
model_precision =  0.7755285650022492
model_recall =  0.8594217347956131
model_f1 =  0.8153227713407424
```

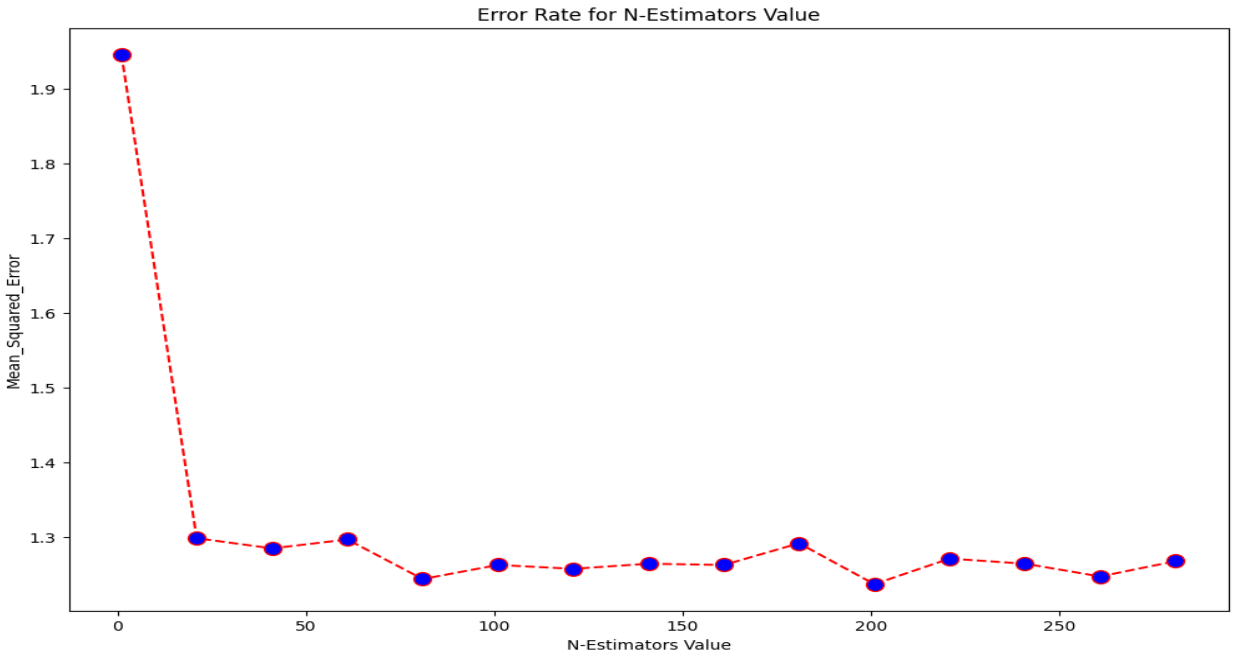
Random Forests:

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 = 201 for this run.

Confusion Matrix:

	Predicted h	Predicted g
Actual h	1669	337
Actual g	265	1741

```
model_accuracy =  0.849950149551346
model_precision =  0.8378248315688162
model_recall =  0.8678963110667997
model_f1 =  0.8525954946131244
```

Naive Bayes:

Confusion Matrix:

	Predicted h	Predicted g
Actual h	805	1201
Actual g	224	1782

```
model_accuracy =  0.6448155533399801
model_precision =  0.5973851827019778
model_recall =  0.8883349950149552
model_f1 =  0.714371617558629
```

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 = 0.7911330049261084

2- ada_boost → model_precision = 0.8220640569395018

3- knn → model_precision = 0.7755285650022492

4- random_forests → model_precision = 0.8378248315688162

5- naïve_bayes → model_precision = 0.5973851827019778

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