

Log File

This file shows the various tuning observations applied in the dataset for each classifier.

Finding the best set of parameters

1. SVM:

Kernel	Number of splits	Training accuracy(%)	Test accuracy (%)
Sigmoid	10	72.41	75.0
	5	72.41	73.68
	8	72.72	77.22
Poly	10	74.32	71.42
	5	78.01	71.92
	8	71.41	73.68
RBF	10	67.20	72.3
	5	86.63	70.5
	8	68.0	70.8
Linear	10	75.78	69.28
	5	77.64	71.63
	8	72.22	71.05

The kernel chosen is sigmoid,
number of splits =10,
Gamma=0.1,
decision_function_shape='ovo'

This tuning of parameters will lead to **training accuracy of 72.41% and testing accuracy of 75.0%.**

2. Naïve Bayes:

We used Gaussian Naïve Bayes since we have a classification problem and we obtained the maximum accuracy of for **training dataset which is 75.86% and testing accuracy of 72.22%**

The best split for K fold cross validation is 2.

3. Decision Trees:

Criterion	Splitter	Max_depth	Training accuracy (%)	Testing accuracy (%)
Gini	Best	7	71.42	86.89
Gini	Best	5	82.36	77.63

The best parameter consists of the set including
Criterion='gini',
splitter='best' and

max_depth=7 where the number of splits = 2 which yielded best results for **testing accuracy i.e., 86.89%**

4.Perceptron:

Finding the best number of split for cross validation

Number of Splits	Training accuracy (%)	Test accuracy (%)
10	72.96%	75.0%
5	70.89	66.66
6	65	61.71

On tuning parameters it lead to final set of best parameters where

penalty='elasticnet',

alpha=0.001

n_iter=10

which gives the accuracy of **75.0% on test dataset.**

5.Neural Network:

Number of splits (K fold)	Training accuracy (%)	Test accuracy (%)
10	72.41	75.0
5	72.41	73.68
8	72.72	72.22
Number of Hidden layers	Training accuracy (%)	Test accuracy (%)
(5,2)	71.24	74.23
(5,5)	72.41	75.0
Activation Function	Training accuracy (%)	Test accuracy (%)
Relu	72.41	75.0
Solver		
Lbfgs	72.41	75.0

The best set of parameters obtained includes:

N_splits=10,

N_hiddenlayers: (5,5)

Activation_function='relu',

Solver='lbfgs',

Which gives the best accuracy in **test dataset of 75.0 %**

6.Logistic regression:

N_splits (cross validation)	Training accuracy (%)	Test accuracy (%)
5	74.56	73.68
10	74.32	71.42
8	73.91	72.22
Solver	Training accuracy (%)	Test accuracy (%)
liblinear	74.56	73.68
Lbfgs	74.41	71.14

Tuning the hyperparameters lead us the best results where

C=1.0, class_weight=None, dual=False, fit_intercept=True,
intercept_scaling=1, max_iter=100, multi_class='ovr', n_jobs=2,
penalty='l2', random_state=None, solver='lbfgs', tol=0.0001,
verbose=0, warm_start=False

And the **test accuracy is 73.68%**.

7. KNN:

N_splits(K fold)	Training accuracy (%)	Test accuracy (%)
10	81.99	71.42
5	82.75	66.66
P	Training accuracy (%)	Test accuracy (%)
10	81.99	71.42
2	82.37	67.85

N tuning the parameters we get the best **test accuracy of 71.42%**, considering

8. Bagging:

N_splits(K fold)	Training accuracy(%)	Test accuracy(%)
10	95.83	60.71
2	91.72	66.66
8	93.91	58.33
n_estimators= 5	91.91	72.16

Keeping the n_estimators=5 we get the best accuracy on **test dataset=72.16%**.

9. Random Forest:

n_splits(K fold)	Training accuracy(%)	Test accuracy(%)
8	73.0	69.80
5	73.33	71.8
N_estimators	Training accuracy	Test accuracy
8	71.89	69.76
10	73.33	71.8
Max_depth	Training accuracy	Test accuracy
5	71.71	73.77

2	73.33	71.8
Max_features	Training accuracy	Test accuracy
Auto	73.27	75.43
Sqrt	72.02	72.45

After tuning we found the best set of parameters which gives the **test accuracy of 75.43%** where the best parameters are

N_splits=5	n_estimators=10	max_depth=2	max_features="auto"	min_samples_split=2
-------------------	------------------------	--------------------	----------------------------	----------------------------

10. Deep Learning:

n_splits(K fold)	Training accuracy (%)	Test accuracy (%)
5	90.21	71.22
10	72.41	81.22
7	72.41	79.43
Activation	Training accuracy (%)	Test accuracy (%)
Relu	72.41	77.63
Sigmoid	72.41	71.54
Number of layers	Training accuracy (%)	Test accuracy (%)
2	70.85	77.63
3	72.41	69.34

Tuning the parameters gives us the best **testing accuracy of 81.22 %**, where the best set of parameters are:

n_splits=10	Number of layers=3	Kernel_initializer='uniform'	Activation='relu'and sigmoid	Optimizer='adam'	Loss='binary_crossentropy'
--------------------	---------------------------	-------------------------------------	-------------------------------------	-------------------------	-----------------------------------

11. Adaboosting:

n_split(K fold)	Training accuracy(%)	Test accuracy(%)
	0	
10	78.54	78.57
5	78.54	56.81
8	79.01	74.22
N_estimators	Training accuracy (%)	Test accuracy (%)
40	76.20	78.57
50	79.69	75.50

Alpha	Training accuracy (%)	Test accuracy (%)
0.9	79.69	78.57
0.7	79.69	75

After tuning the hyperparameters we got the best accuracy on **testing dataset of 78.57%** where the best set of parameters are:

base_estimator=None, n_estimators=40, learning_rate=0.9, algorithm='SAMME.R',
random_state=None

12. Gradient Boosting

n_splits(K fold)	Training accuracy(%)	Test accuracy(%)
10	88.71	71.42
5	89.65	65.81
8	89.22	68.26
Loss	Training accuracy	Test accuracy
Deviance	80.32	79.16
Exponential	81.22	82.14
N_estimators	Training accuracy	Test accuracy
130	79.47	79.16
100	81.22	82.14

Tuning the parameters gives the best **testing accuracy of 82.14%**, where the best set of parameters are:

loss='exponential', learning_rate=0.1, n_estimators=100, criterion='friedman_mse',
min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_depth=2,
min_impurity_split=None, init=None, random_state=None, max_features=None, verbose=0,
max_leaf_nodes=None, warm_start=False, presort='auto')