REPORT :

As seen from the output summary , **k-nearest neighbor** had the best accuracy of 0.99 .

**Best parameters :**

algorithm='auto’,

leaf\_size=30,

metric='minkowski',

metric\_params=None,

n\_jobs=1,

n\_neighbors=1,

p=2,

weights='uniform'

The above parameters gave the best results after going through a 5 – fold cross validation and auto tuning using grid search .

REASON :

* Works well with multiclass data set .
* Non-parametric setting and simple classification based on just nearest neighbor.

IMPROVEMENT :

* Changing the distance metric also improves the performance .
* The accuracy maybe improved if the data is rescaled and normalized .
* Inverted indexing which includes all neighbors in the training set ,maybe performed to improve the model

APPROACH USED :

* The data was initially split into train and test data with train\_test\_split to prevent data leak into test data.
* Then the training data was fit into each classifier , along with a set of parameters and further grid search using to find the best set of parameters for each classifier .
* Then the data went through cross-validation with 5 folds after training the model with best parameters .
* Each time the accuracy was calculated on the test data set and finally the mean value was calculated .

EVALUATION METRICS PERFORMED :

* Confusion matrix
* Classification Report
* Accuracy measure