**Problem Identification:**

Stage 1:

Domain Selection – Machine Learning

Reason: Requirement output is Numeric only

Stage 2:

Learning Selection – Supervised Learning

Reason: Clear requirement and data set input and output well defined.

Stage 3:

Supervised Learning – Classification

Reason: Requirement expected output yes or no, not a numeric value.

**Data Set Details:**

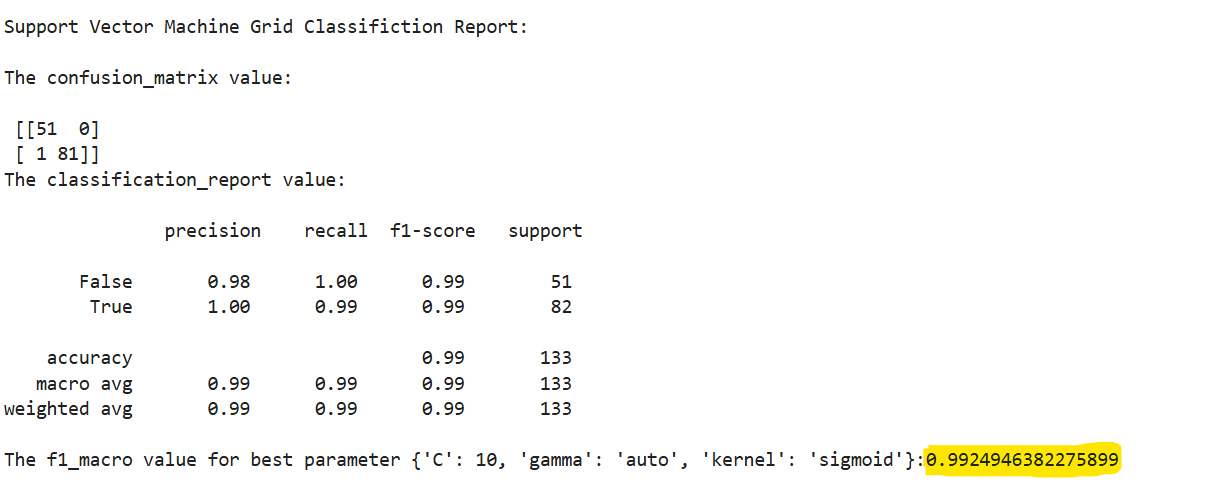
Use the panda module **‘shape’** attribute find the rows and columns details.

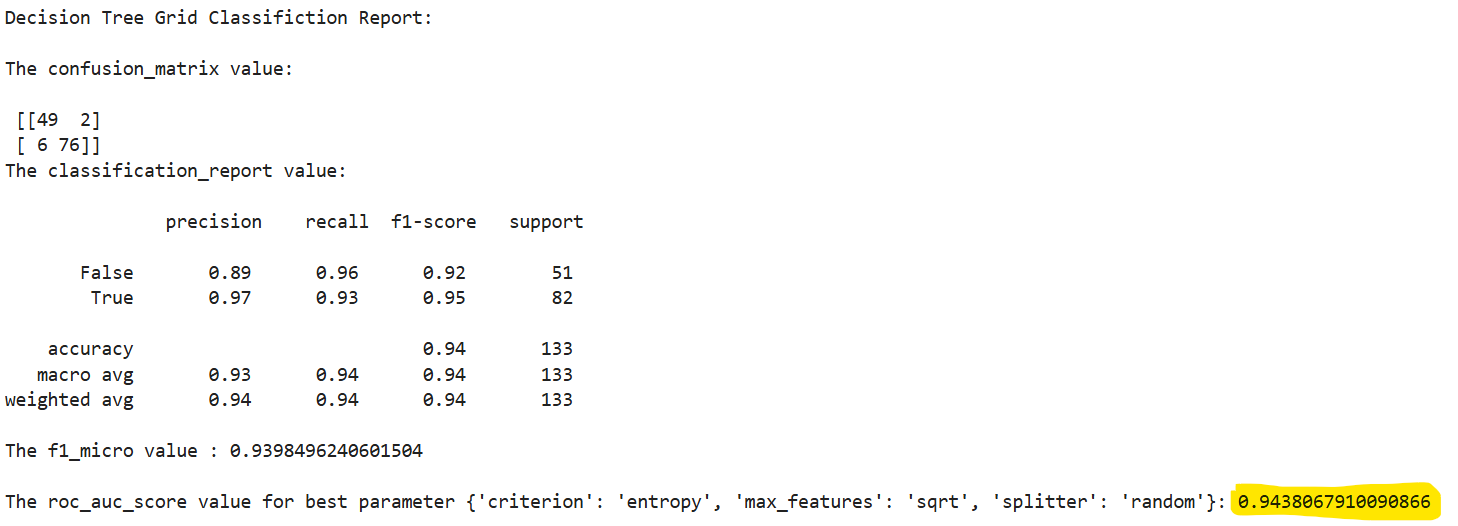
Number of Row : 339

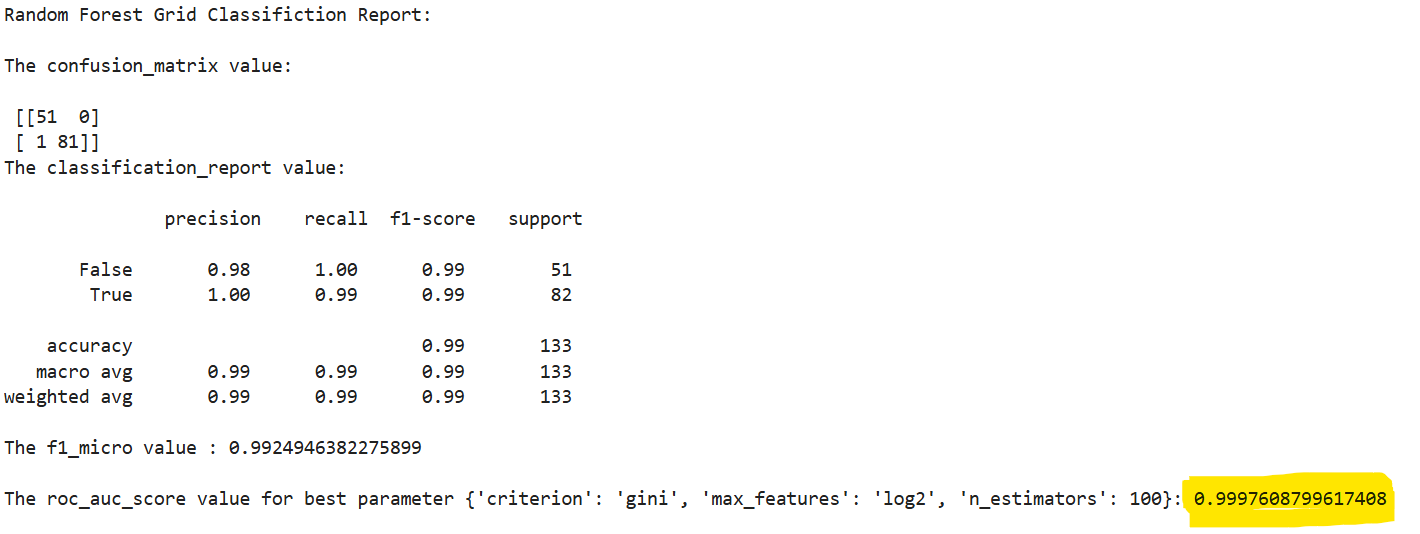
Number of Column : 25

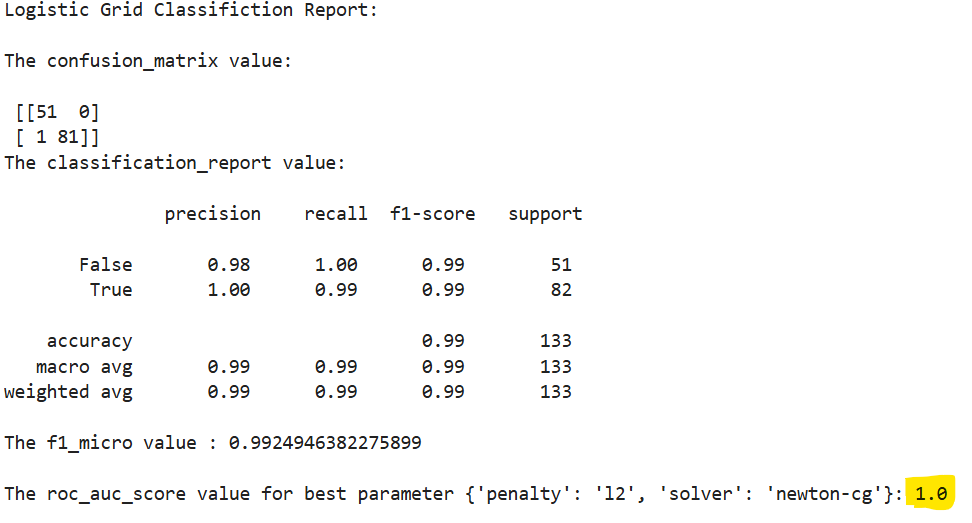
**Pre-Processing the Data Set:**

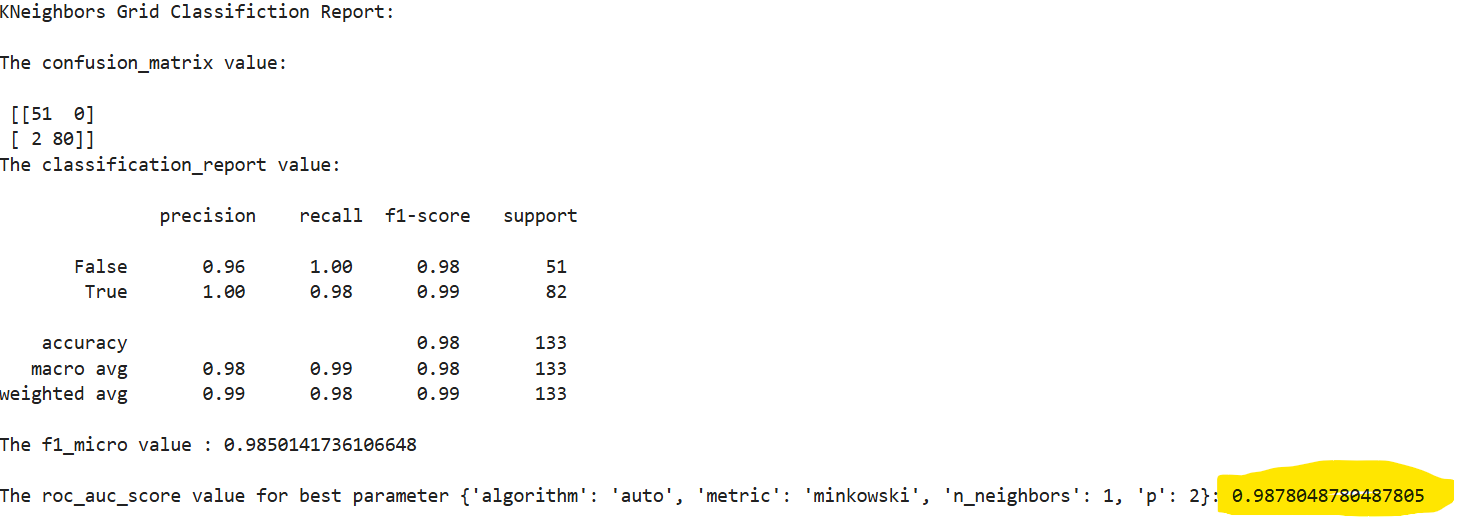
In the input dataset, some of the columns contain categorical data. Therefore, we need to convert them to numeric data using the nominal 'One Hot Encoding' algorithm.

****

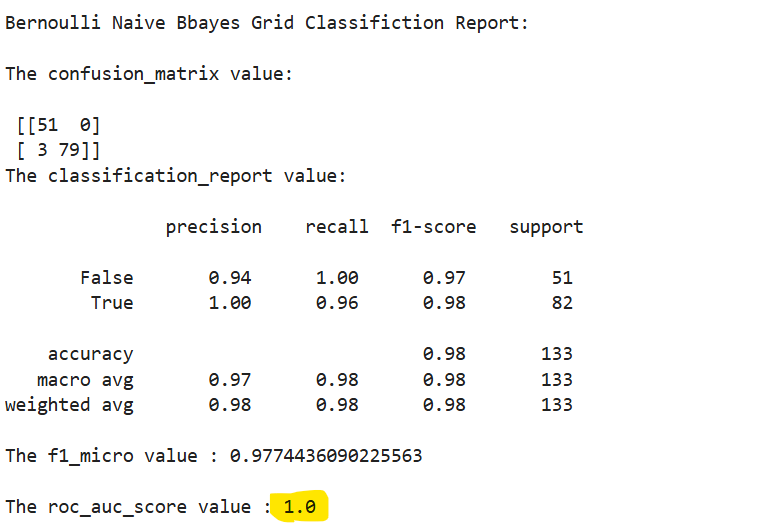
****







**Naive\_bayes Alogrithm:**



**Final Model:**

We have predicted the final model using a Logistic algorithm with the f1 score is 0.99 and ROC and AUC score is 1

|  |  |  |
| --- | --- | --- |
| **Classification Algorithms** | **f1\_score** | **roc\_auc\_score** |
| SVM | 0.992494638 |  |
| Decision Tree | 0.939849624 | 0.943806791 |
| Random Forest | 0.992494638 | 0.99976088 |
| Logistic | 0.992494638 | 1 |
| KNN | 0.985014174 | 0.987804878 |
| Bernoulli Naive Bayes | 0.977443609 | 1 |