BEST MODELSUMMARY REPORT

RFE CLASSIFICATION

Feature selection:-

In Company user project, example medical diagnosis 27 columns of input no need.we have to reduce the columns(Features).output is (label). To overcome this we going for Advanced ML Feature selection.

Important columns(Features) selected and Model Performances.

In MachineLearning, CLASSIFICATION mainly performs by **CONFUSION MATRIX TO GET THE ACCURACY SCORE.**

ML Feature Selection -→ Only supports for Machine Learning.

Feature Selection Methods:-

1. ### SELECT K ALGORITHM

   The Important columns(features) are selected by **Chi-square model.** Using this model we applying for different Algoithms, Getting the Accuracy score.

   Finally **DecisionTree** got the Highest Accuracy

GETFEATURENAME OUT

In KBEST Classification Using **DecisionTree**, **n=5** has highest Accuracy.

```
[9]: result
     #5
```

	Decision
ChiSquare	0.96

```
21]: from sklearn.feature_selection import SelectKBest
     from sklearn.feature_selection import chi2

     skbest = SelectKBest(score_func=chi2, k=5)

     skbest.fit(indep_X, dep_Y)
     # Get the selected feature names
     selected_features = indep_X.columns[skbest.get_support()]
     print(selected_features)

     Index(['bgr', 'bu', 'sc', 'pcv', 'wc'], dtype='object')
```

Now Display the Feature Names (Column Names Taken for n=5).

Summary:

Using **k=5**, The Algorithm Performance is good. Decision Tree Getting **0.96** Accuracy compare to other Algorithms.

2. RECURSIVE FEATURE ELIMINATION

The Important columns(features) are selected by **Algorithm model itself.** Using this Algorithm model we **cross validation** for different Algoithms, Getting the Accuracy score.

If one algorithm selecting the feature(column), **cross validation** for another respective algorithms.

Finally **Logistc Regression-classification** got the highest accurcy

GETFEATURENAME OUT

In RFE Classification Using **LogisticRegression**, **n=5** has highest Accuracy.

```
[43]:  result
       #4
```

```
[43]:            Logistic
       _____
       Logistic    0.95
```

```
[49]:  result
       #3
```

```
[49]:            Logistic
       _____
       Logistic    0.94
```

```
[55]:  result
       #5
```

```
[55]:            Logistic
       _____
       Logistic    0.98
```

Now Display the Feature Names (Column Names Taken for n=5)

```
[15]:  print(selected_features)
       #5

       Index(['al', 'sg_c', 'sg_d', 'htn_yes', 'dm_yes'], dtype='object')
```

Summary

Compare to all Accuracy, **Using n=5**, **SVM Logistic Regression** have the **highest** accuracy scores across all algorithms, consistently achieving 0.98 accuracy.