

## RFE SUMMARY 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 Machine Learning, CLASSIFICATION mainly performs by **CONFUSION MATRIX TO GET THE ACCURACY SCORE.**

ML Feature Selection -> Only supports for Machine Learning.

#### Feature Selection Methods:-

### 1. RECURSIVE FEATURE ELIMINATION

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

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

```
[6]: result  
#4
```

[6]:	Logistic	SVMl	SVMnl	KNN	Navie	Decision	Random
Logistic	0.95	0.95	0.95	0.95	0.95	0.95	0.95
SVC	0.96	0.96	0.96	0.96	0.96	0.96	0.96
Random	0.97	0.97	0.97	0.98	0.87	0.95	0.97
DecisionTree	0.98	0.98	0.92	0.98	0.81	0.98	0.98

---

- Decision Tree and SVM (linear) have the highest accuracy at 0.98 across most comparisons
- Random Forest performs well, with an accuracy of 0.97 for SVM (non-linear) and Logistic Regression
- KNN achieves the highest accuracy of 0.98 for Random Forest
- Naive Bayes has the lowest accuracy scores, with 0.81 being the lowest accuracy achieved

The **Decision Tree** and **SVM** are the best performing algorithms for this dataset, while **Naive Bayes** has the poorest performance.

```
[8]: result
#5
```

	Logistic	SVMl	SVMnl	KNN	Navie	Decision	Random
Logistic	0.98	0.98	0.98	0.98	0.98	0.98	0.98
SVC	0.99	0.99	0.99	0.99	0.99	0.99	0.99
Random	0.97	0.97	0.98	0.97	0.91	0.96	0.98
DecisionTree	0.95	0.98	0.93	0.94	0.85	0.97	0.98

- **SVM (linear and non-linear) and Logistic Regression** have the **highest** accuracy scores across all algorithms, consistently achieving 0.98 or 0.99 accuracy.
- Random Forest also performs well, with an accuracy of 0.98 for SVM (non-linear) and 0.97 for Logistic Regression.
- Decision Tree shows varying performance, with accuracies ranging from 0.85 to 0.98 across different comparisons.
- Naive Bayes has the lowest accuracy scores, with 0.91 being the highest accuracy achieved.

```
[10]: result  
#3
```

	Logistic	SVMl	SVMnl	KNN	Navie	Decision	Random
Logistic	0.94	0.94	0.94	0.94	0.94	0.94	0.94
SVC	0.87	0.87	0.87	0.87	0.87	0.87	0.87
Random	0.94	0.94	0.94	0.94	0.9	0.91	0.92
DecisionTree	0.98	0.98	0.98	0.98	0.79	0.97	0.97

- **Decision Tree** has the **highest** accuracy at 0.98 across all comparisons.
- **Logistic Regression and SVM** (linear and non-linear) have the same accuracy of 0.94.
- **Random Forest** performs better than Naive Bayes and Decision Tree, with an accuracy of 0.92.
- **Naive Bayes** has the **lowest** accuracy, with 0.79 being the lowest score.

The results suggest that Decision Tree is the best performing algorithm for this dataset, while Naive Bayes has the poorest performance. Logistic Regression and SVM (linear and non-linear) have similar accuracy scores, followed by Random Forest.

## 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 (linear and non-linear) and Logistic Regression** have the **highest** accuracy scores across all algorithms, consistently achieving 0.98 or 0.99 accuracy.