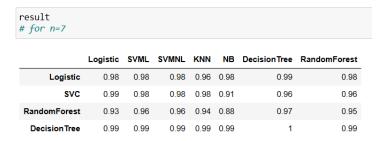# RECURSIVE FEATURE ELIMINATION (RFE)

Recursive Feature Elimination is a method that identifies the most important feature by recursively training the model and eliminating the least important feature.

We've got the dataset for the prediction of Chronic Kidney Disease. The dataset contains 25 columns in total, out of which 24 are independent variables and 1 dependent variable. RFE is used to fetch the significant features out of these 24 available features in the dataset.

In this RFE, we've brought four models into the RFE "Logistic Regression", "Random Forest Classifier", "Decision Tree Classifier", "SVC" to fetch the important features from each models. All these important features are then tested with each models to study how the selected features performs when implemented to the individual models, and the results are given below. Here, n is the number of features.

```
result
# for n=3
```

	Logistic	SVML	SVMNL	KNN	NB	Decision Tree	RandomForest
Logistic	0.94	0.94	0.94	0.94	0.94	0.94	0.94
SVC	0.94	0.94	0.94	0.94	0.9	0.91	0.92
RandomForest	0.97	0.98	0.98	0.98	0.79	0.97	0.97
Decision Tree	0.87	0.87	0.87	0.87	0.87	0.87	0.87

```
result
# for n=4
```

	Logistic	SVML	SVMNL	KNN	NB	Decision Tree	RandomForest
Logistic	0.95	0.95	0.95	0.95	0.95	0.95	0.95
SVC	0.97	0.97	0.97	0.97	0.87	0.95	0.97
RandomForest	0.91	0.92	0.92	0.98	0.81	0.98	0.98
Decision Tree	0.96	0.96	0.96	0.96	0.96	0.96	0.96

```
result
# for n=5
```

	Logistic	SVML	SVMNL	KNN	NB	Decision Tree	RandomForest
Logistic	0.98	0.98	0.98	0.98	0.98	0.98	0.98
SVC	0.97	0.98	0.98	0.97	0.91	0.96	0.98
RandomForest	0.92	0.93	0.93	0.94	0.85	0.97	0.98
Decision Tree	0.99	0.99	0.99	0.99	0.99	0.99	0.99

```
result
# for n=6
```

	Logistic	SVML	SVMNL	KNN	NB	Decision Tree	RandomForest
Logistic	0.98	0.98	0.98	0.98	0.98	0.99	0.98
SVC	0.97	0.99	0.99	0.96	0.92	0.95	0.98
RandomForest	0.96	0.97	0.97	0.94	0.85	0.97	0.96
Decision Tree	0.99	0.99	0.99	0.99	0.99	0.99	0.99

```
result
# for n=7
```

	Logistic	SVML	SVMNL	KNN	NB	Decision Tree	RandomForest
Logistic	0.98	0.98	0.98	0.96	0.98	0.99	0.98
SVC	0.99	0.98	0.98	0.98	0.91	0.96	0.96
RandomForest	0.93	0.96	0.96	0.94	0.88	0.97	0.95
Decision Tree	0.99	0.99	0.99	0.99	0.99	1	0.99

Conclusion:

From the above table, the accuracy score increases as the number of feature value increase. But after the value n=6, the accuracy score almost remains unchanged. So, we can finalize that the number of features required for the prediction of Chronic Kidney Disease is n = 6.