FORWARD SELECTION

Forward Selection is a feature selection method which is used to identify the important features from a huge dataset that improves the model's performance.

Scikit Learn library doesn't provide Forward Selection function but this functionality can be achieved using "**SequencialFeatureSelector**" from "**mlxtend**" library.

Forward Selection works by evaluating all the features present in the dataset and adds a single feature to a temporary feature set. Model is trained using this temporary feature set and evaluates the accuracy score. Next another feature is added, trains the model and evaluates the accuracy score. This process continues until it reaches the predefined number or further addition of features leads to no improvement in the model.

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

	Logistic	SVML	SVMNL	KNN	NB	DecisionTree	RandomForest
RandomForest	0.98	0.96	0.96	0.99	0.91	0.99	0.99
DecisionTree	0.97	0.97	0.97	0.99	0.82	0.99	0.99

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

	Logistic	SVML	SVMNL	KNN	NB	DecisionTree	RandomForest
RandomForest	0.95	0.96	0.96	0.99	0.93	0.99	0.99
DecisionTree	0.97	0.97	0.97	0.99	0.83	0.99	0.99

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

	Logistic	SVML	SVMNL	KNN	NB	DecisionTree	RandomForest
RandomForest	0.96	0.96	0.96	0.99	0.96	0.99	0.99
DecisionTree	0.99	0.99	0.99	0.99	0.88	0.99	0.99

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

	Logistic	SVML	SVMNL	KNN	NB	DecisionTree	RandomForest
RandomForest	0.97	0.96	0.96	0.98	0.96	0.99	0.99
DecisionTree	0.98	0.96	0.96	0.99	0.91	0.99	0.97

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

	Logistic	SVML	SVMNL	KNN	NB	DecisionTree	RandomForest
RandomForest	0.97	0.96	0.96	0.97	0.96	0.99	0.99
DecisionTree	0.98	0.96	0.96	0.99	0.91	0.99	0.99

Conclusion:

From the above displayed tables, the accuracy score results are varying as the number of features increased. For n = 5, the results are stable. Hence, we can finalize the number of features required for the prediction of CKD using Forward Selection is n=5.