FEATURE IMPORTANCE

Feature Importance is an attribute, present in the tree based models like, **Decision Tree, Random Forest, and Gradient Boosting Classifier**. Since the feature importance from these models are impurity based importance, it can be biased towards highly unique value features. Hence we use **PERMUTATION FEATURE IMPORTANCE.**

Permutation Feature Importance is provided by scikit-learn library that uses agnostic technique and fits any estimator. This PFI works by randomly shuffling the values present in a single feature and checks the score. The larger the score decreases, higher the importance. This takes place for every single features present in the dataset.

The accuracy result for n number of important features fetched from Permutation Feature Importance is given below.

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

	Logistic	SVML	SVMNL	KNN	NB	DecisionTree	RandomForest
DecisionTree	0.95	0.96	0.96	0.96	0.85	0.94	0.96
RandomForest	0.94	0.94	0.94	0.95	0.84	0.95	0.96

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

	Logistic	SVML	SVMNL	KNN	NB	DecisionTree	RandomForest
DecisionTree	0.95	0.96	0.96	0.95	0.88	0.96	0.94
RandomForest	0.95	0.96	0.96	0.95	0.88	0.98	0.96

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

	Logistic	SVML	SVMNL	KNN	NB	DecisionTree	RandomForest
DecisionTree	0.95	0.95	0.95	0.96	0.87	0.95	0.97
RandomForest	0.96	0.96	0.96	0.98	0.88	0.94	0.97

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

	Logistic	SVML	SVMNL	KNN	NB	DecisionTree	RandomForest
DecisionTree	0.95	0.96	0.96	0.93	0.87	0.96	0.97
RandomForest	0.96	0.97	0.97	0.96	0.88	0.98	0.98

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

	Logistic	SVML	SVMNL	KNN	NB	DecisionTree	RandomForest
DecisionTree	0.95	0.97	0.97	0.95	0.81	0.95	0.96
RandomForest	0.95	0.97	0.97	0.96	0.84	0.94	0.97

Conclusion:

From the above displayed tables, the difference in accuracy score is not positively huge as number of features is increased from n = 3 to 7. So we can finalize the number of features required for CKD prediction obtained from PFI is n=4