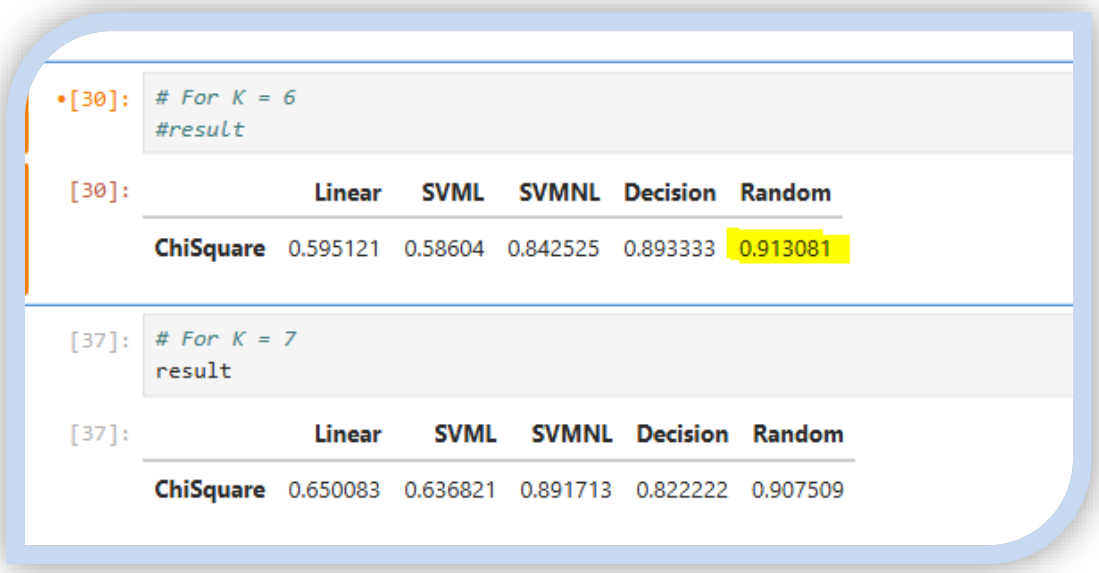
Feature Selection

SELECT K

1. Select K Regression:



```
[30]: # For K = 6
      #result
```

[30]:

	Linear	SVML	SVMNL	Decision	Random
ChiSquare	0.595121	0.58604	0.842525	0.893333	0.913081

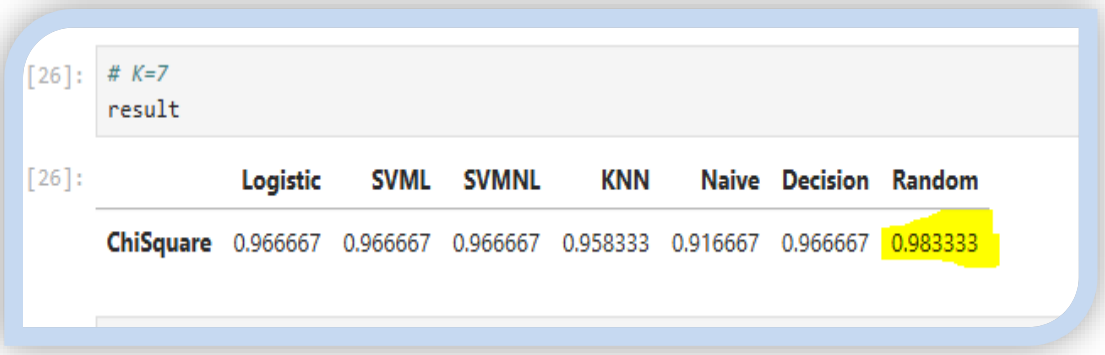
```
[37]: # For K = 7
      result
```

[37]:

	Linear	SVML	SVMNL	Decision	Random
ChiSquare	0.650083	0.636821	0.891713	0.822222	0.907509

In Select K Regression, columns = 6 gives 0.91as highest chi square value in Random Forest algorithm therby we found the best model..

2.a) Select K Classification(Before Grid_Search):



```
[26]: # K=7
      result
```

[26]:

	Logistic	SVML	SVMNL	KNN	Naive	Decision	Random
ChiSquare	0.966667	0.966667	0.966667	0.958333	0.916667	0.966667	0.983333

In Select k Classification, if we take feature selection for 7 columns, we get the highest chi square value of 0.98 for Random Forest Algorithm before grid search therby we found the best model.

2.b) Select K Classification(After Grid_Search):

```
# K=7
result
```

	Logistic	SVML	SVMNL	KNN	Naive	Decision	Random
ChiSquare	0.966667	0.966667	0.966667	0.958333	0.916667	0.925	0.991667

In Select k Classification, if we take feature selection for 7 columns, we get the highest chi square of 0.99 after the grid search for Random Forest Algorithm therby we found the best model.

RECURSIVE FEATURE ELIMINATION

1a). RFE Regression(Before GridSearch):

```
# For K=7
result
```

	Linear	SVML	SVMNL	Decision	Random
Linear	0.638025	0.572635	0.786516	0.816771	0.815285
SVM	0.630993	0.572802	0.783481	0.815304	0.813284
Decision	0.713978	0.693422	0.932137	0.928889	0.959996
Random	0.711696	0.688855	0.92742	0.786667	0.869703

In RFE regression, if we take feature selection for 7 columns, we get the highest accuracy_score value of 0.95 for Decision Tree featured with Random Forest Algorithm before Grid Search therby we found the best model.

1b). RFE Regression(After GridSearch):

```
# For K = 7
result
```

	Linear	SVML	SVMNL	Decision	Random
Linear	0.638025	0.572635	0.786516	0.782681	0.808108
SVM	0.630993	0.572802	0.783481	0.74988	0.809859
Decision	0.713978	0.693422	0.932137	0.786667	0.972267
Random	0.711696	0.688855	0.92742	0.857778	0.929931

In RFE regression, if we take feature selection for 7 columns, we get the highest accuracy_score value of 0.97 for Decision Tree featured with Random Forest Algorithm therby we found the best model..

2. RFE(Classification):

```
# For K=6
#result
```

	Logistic	SVML	SVMNL	KNN	Naive	Decision	Random
Logistic	0.975	0.958333	0.975	0.975	0.975	0.975	0.966667
SVM	0.983333	0.983333	0.983333	0.958333	0.983333	0.983333	0.983333
Decision	0.966667	0.966667	0.975	0.858333	0.966667	0.975	0.966667
Random	0.975	0.983333	0.991667	0.925	0.966667	0.941667	0.975

In
RFE classification, if we take feature selection for 6 columns, we get the highest accuracy_score value of 0.99 for Random Forest featured with SVM non- linear Algorithm therby we found the best model..