K- Algorithm Assignments

Dataset:

```
# print the updated data
df2
```

	age	bp	al	su	bgr	bu	sc	sod	pot	hrmo	...	pc_normal	pcc_present	ba_present	htn_yes	d
0	2.000000	76.459948	3.0	0.0	148.112676	57.482105	3.077356	137.528754	4.627244	12.518156	...	False	False	False	False	
1	3.000000	76.459948	2.0	0.0	148.112676	22.000000	0.700000	137.528754	4.627244	10.700000	...	True	False	False	False	
2	4.000000	76.459948	1.0	0.0	99.000000	23.000000	0.600000	138.000000	4.400000	12.000000	...	True	False	False	False	
3	5.000000	76.459948	1.0	0.0	148.112676	16.000000	0.700000	138.000000	3.200000	8.100000	...	True	False	False	False	
4	5.000000	50.000000	0.0	0.0	148.112676	25.000000	0.600000	137.528754	4.627244	11.800000	...	True	False	False	False	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
394	51.492308	70.000000	0.0	0.0	219.000000	36.000000	1.300000	139.000000	3.700000	12.500000	...	True	False	False	False	
395	51.492308	70.000000	0.0	2.0	220.000000	68.000000	2.800000	137.528754	4.627244	8.700000	...	True	False	False	True	
396	51.492308	70.000000	3.0	0.0	110.000000	115.000000	6.000000	134.000000	2.700000	9.100000	...	True	False	False	True	
397	51.492308	90.000000	0.0	0.0	207.000000	80.000000	6.800000	142.000000	5.500000	8.500000	...	True	False	False	True	
398	51.492308	80.000000	0.0	0.0	100.000000	49.000000	1.000000	140.000000	5.000000	16.300000	...	True	False	False	False	

399 rows × 28 columns

Classification result:

```
# get all the accuracy of each algorithm and print it in a dataframe
result=selectk_Classification(acclog,accsvml,accsvmnl,accknn,accnav,accdes,accrf)
```

```
# result of  k = 5
result
```

	Logistic	SVMl	SVMnl	KNN	Navie	Decision	Random
ChiSquare	0.94	0.94	0.95	0.89	0.83	0.96	0.95

```
# result of  k = 6
result
```

	Logistic	SVMl	SVMnl	KNN	Navie	Decision	Random
ChiSquare	0.95	0.96	0.96	0.93	0.89	0.97	0.97

```
# result of  k = 7
result
```

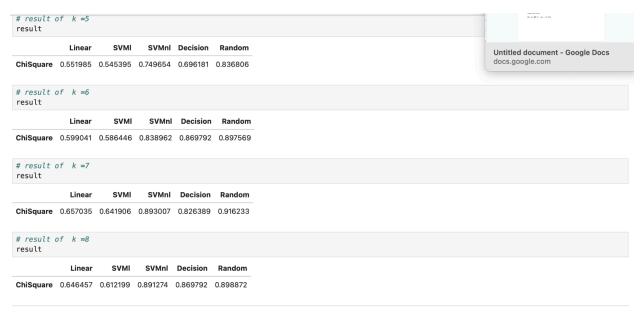|	Logistic	SVMl	SVMnl	KNN	Navie	Decision	Random
ChiSquare	0.97	0.97	0.97	0.97	0.91	0.98	0.97

Here , the accuracy of all the algorithms increases and the accuracy of random forest value remains same when we choose 7 features ['al', 'bgr', 'bu', 'sc', 'hrmo', 'pcv', 'wc']

Regression result:

```
# result of  k =5
result
```

	Linear	SVMl	SVMnl	Decision	Random
ChiSquare	0.551985	0.545395	0.749654	0.696181	0.836806

```
# result of  k =6
result
```

	Linear	SVMl	SVMnl	Decision	Random
ChiSquare	0.599041	0.586446	0.838962	0.869792	0.897569

```
# result of  k =7
result
```

	Linear	SVMl	SVMnl	Decision	Random
ChiSquare	0.657035	0.641906	0.893007	0.826389	0.916233

```
# result of  k =8
result
```

	Linear	SVMl	SVMnl	Decision	Random
ChiSquare	0.646457	0.612199	0.891274	0.869792	0.898872

Here , the accuracy of all the algorithms increases but few of the algorithms value decreases so we can't choose 8 features

Here , the accuracy of all the algorithms increases and the accuracy of decision tree value remains same when we choose 7 features (['al', 'bgr', 'bu', 'sc', 'hrmo', 'pcv', 'wc'])

In conclusion, the accuracy of all the algorithms got increases when we chooses the below 7 features ['al', 'bgr', 'bu', 'sc', 'hrmo', 'pcv', 'wc'] in both regression and classification models

In regression, Random Forest gives best accuracy than the other regression models

In classification, Decision Tree gives best accuracy than the other classification models