# RFE- 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:

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

| | Logistic | SVMl | SVMnl | KNN | Navie | Decision | Random |
|---|---|---|---|---|---|---|---|
| Logistic | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 |
| SVC | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| Random | 0.97 | 0.97 | 0.98 | 0.97 | 0.91 | 0.96 | 0.98 |
| DecisionTree | 0.95 | 0.98 | 0.93 | 0.94 | 0.85 | 0.97 | 0.98 |

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

| | Logistic | SVMl | SVMnl | KNN | Navie | Decision | Random |
|---|---|---|---|---|---|---|---|
| Logistic | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.99 | 0.98 |
| SVC | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 |
| Random | 0.98 | 0.98 | 0.99 | 0.96 | 0.92 | 0.95 | 0.98 |
| DecisionTree | 0.96 | 0.96 | 0.97 | 0.97 | 0.85 | 0.97 | 0.96 |

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

| | Logistic | SVMl | SVMnl | KNN | Navie | Decision | Random |
|---|---|---|---|---|---|---|---|
| Logistic | 0.98 | 0.98 | 0.98 | 0.96 | 0.98 | 0.99 | 0.98 |
| SVC | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 1.0 | 0.99 |
| Random | 0.98 | 0.98 | 0.98 | 0.98 | 0.91 | 0.96 | 0.96 |
| DecisionTree | 0.95 | 0.96 | 0.96 | 0.94 | 0.88 | 0.97 | 0.95 |

Here , The accuracy of all the models decreases when we calculate RFE values with Decision Tree and random forest models while using 7 features so we can go with 6 features and we

could see 99% of accuracy with RFE score using SVC models on 6 features (['sc', 'hrmo', 'pcv', 'sg_c', 'rbc_normal', 'dm_yes'])


**Regression result:**

```
# result of  k = 3
result
```

|  | Linear | SVMI | Decision | Random |
|---|---|---|---|---|
| **Linear** | 0.441961 | 0.262153 | 0.441961 | 0.441816 |
| **SVC** | 0.441961 | 0.262153 | 0.441961 | 0.441816 |
| **Random** | 0.664893 | 0.609652 | 0.965961 | 0.916304 |
| **DecisionTree** | 0.676174 | 0.670691 | 0.933504 | 0.887256 |

```
# result of  k = 4
result
```

|  | Linear | SVMI | Decision | Random |
|---|---|---|---|---|
| **Linear** | 0.60401 | 0.457046 | 0.776711 | 0.776492 |
| **SVC** | 0.60401 | 0.457046 | 0.776711 | 0.776492 |
| **Random** | 0.671727 | 0.628963 | 0.835247 | 0.8403 |
| **DecisionTree** | 0.681563 | 0.614992 | 0.96711 | 0.923559 |

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

|  | Linear | SVMI | Decision | Random |
|---|---|---|---|---|
| **Linear** | 0.620124 | 0.457136 | 0.77924 | 0.780135 |
| **SVC** | 0.604508 | 0.456871 | 0.776474 | 0.776745 |
| **Random** | 0.674403 | 0.628206 | 0.696181 | 0.815538 |
| **DecisionTree** | 0.686361 | 0.643365 | 0.836806 | 0.845303 |

Here , the accuracy of all the algorithms increases but few of the algorithms value decreases so we can't choose 3 and 5 features

Here , the accuracy of all the algorithms increases and the accuracy of decision tree value remains same when we choose 4 features (['al', 'hrmo', 'sg_c', 'sg_d'])

In conclusion, the accuracy of all the algorithms got increases when we choose 6 features (['sc', 'hrmo', 'pcv', 'sg_c', 'rbc_normal', 'dm_yes']) in classification and 4 features (['al', 'hrmo', 'sg_c', 'sg_d']) in regression models

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

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