**Chi square classification**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **ChiSquare** | **Logistic** | **SVMl** | **SVMnl** | **KNN** | **Navie** | **Decision** | **Random** |
| 6 | 0.95 | 0.96 | 0.96 | 0.93 | 0.89 | 0.97 | 0.97 |
| 5 | 0.94 | 0.94 | 0.95 | 0.89 | 0.83 | 0.96 | 0.95 |
| 4 | 0.85 | 0.82 | 0.83 | 0.86 | 0.79 | 0.89 | 0.89 |

**Based on the Chi square classification**

**Selecting 6 features** – All algorithms performance is well and with the highest number of performances is done in **Decision Tree and Random forest** and the percentage is **97%**.

**Selecting 5 features** – All algorithms performance is well and with the highest number of performances is done in **Decision Tree** and the percentage is **96%**.

**Selecting 4 features** – All algorithms performance is well and with the highest number of performances is done in **Decision Tree and Random forest** and the percentage is **89%**.

Considered with this table we can take **6 or 5** features with the algorithm **Decision Tree or Random** **forest** to have the best performance of **97%.**

**Chi Square Regression**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ChiSquare** | **Linear** | **SVMl** | **SVMnl** | **Decision** | **Random** |
| 7 | 0.66 | 0.64 | 0.89 | 0.83 | 0.92 |
| 6 | 0.60 | 0.59 | 0.84 | 0.87 | 0.89 |
| 5 | 0.55 | 0.55 | 0.75 | 0.70 | 0.84 |
| 4 | 0.30 | 0.25 | 0.43 | 0.48 | 0.60 |

**Based on the Chi square Regression**

**Selecting 7 features** – SVMnl, Decision Tree and Random Forest algorithms performance is well and with the highest number of performances is done in **Random forest** and the percentage is **92%**.

**Selecting 6 features** – SVMnl, Decision Tree and Random Forest algorithms performance is well and with the highest number of performances is done in **Random forest** and the percentage is **89%**.

**Selecting 5 features** – Random forest algorithms performance is well and with the highest number of performances is done in **Decision Tree and Random forest** and the percentage is **89%**.

Considered with this table we can take **7** features with the algorithm **Random** **forest** to have the best performance of **92%.**

**RFE Classification**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **RFE 5** | **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 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **RFE 4** | **Logistic** | **SVMl** | **SVMnl** | **KNN** | **Navie** | **Decision** | **Random** |
| Logistic | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 | 0.95 |
| SVC | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 |
| Random | 0.97 | 0.97 | 0.97 | 0.98 | 0.87 | 0.95 | 0.97 |
| DecisionTree | 0.98 | 0.98 | 0.92 | 0.98 | 0.81 | 0.98 | 0.98 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **RFE 3** | **Logistic** | **SVMl** | **SVMnl** | **KNN** | **Navie** | **Decision** | **Random** |
| Logistic | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 |
| SVC | 0.87 | 0.87 | 0.87 | 0.87 | 0.87 | 0.87 | 0.87 |
| Random | 0.94 | 0.94 | 0.94 | 0.94 | 0.9 | 0.91 | 0.92 |
| DecisionTree | 0.98 | 0.98 | 0.98 | 0.98 | 0.79 | 0.97 | 0.97 |

**Based on RFE classification**

**Selecting 5 features** - All algorithms performance is well and with the percentage of **90**-**99%**.

**Selecting 4 features** - All algorithms performance is well and with the percentage of **90**-**98%**.

**Selecting 3 features** - All algorithms performance is well and with the percentage of **90**-**98%**.

Considered based on above table we can select 5 features to have a high performance since all the algorithms performance are close to the same range.

**RFE Regression**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **RFE 6** | **Linear** | **SVMl** | **Decision** | **Random** |
| Linear | 0.62 | 0.46 | 0.82 | 0.81 |
| SVC | 0.61 | 0.53 | 0.80 | 0.81 |
| Random | 0.70 | 0.67 | 0.78 | 0.83 |
| DecisionTree | 0.71 | 0.67 | 0.84 | 0.88 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **RFE 5** | **Linear** | **SVMl** | **Decision** | **Random** |
| Linear | 0.62 | 0.46 | 0.78 | 0.78 |
| SVC | 0.60 | 0.46 | 0.78 | 0.78 |
| Random | 0.67 | 0.63 | 0.70 | 0.82 |
| DecisionTree | 0.69 | 0.64 | 0.84 | 0.85 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **RFE 4** | **Linear** | **SVMl** | **Decision** | **Random** |
| Linear | 0.60 | 0.46 | 0.78 | 0.78 |
| SVC | 0.60 | 0.46 | 0.78 | 0.78 |
| Random | 0.67 | 0.63 | 0.84 | 0.84 |
| DecisionTree | 0.68 | 0.61 | 0.98 | 0.92 |

**Based on RFE Regression**

**Selecting 6 Features** – Linear and SVMl show average performance and Decision tree and Random forest have higher performance with the **80-88%**.

**Selecting 5 Features** - Linear and SVMl show average performance and Decision tree and Random forest have higher performance with the **70-85%**.

**Selecting 4 Features** -Linear and SVMl show average performance and Decision tree and Random forest have higher performance with the **78-98%**.

Considering with above table we can take 4 features with the algorithm **Decision Tree** to have the best performance of **98%.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Chi Square** | **RFE** | **Features selected** | **Percentage** |
| **Classification** | Decision Tree or Random Forest |  | 6 or 5 | 97% |
| **Regression** | Random Forest |  | 7 | 92% |
| **Classification** |  | All Algorithm | 5 | 90-99% |
| **Regression** |  | Decision Tree | 4 | 98% |