**Support Vector Machine**

From the results obtained in SVM for a n fold cross validation, based on the accuracy, precision, recall and F-Measure values, the ***radial*** kernel is giving the best results for the SVM model that we created. The parameters that we have taken into consideration are ***cost, kernel, gamma*** and ***tolerance***.

For a linear kernel, the measure of accuracy, precision, recall or F-measure are almost similar irrespective of the value change in cost and tolerance. We cannot place a gamma value for a linear kernel, because the gamma value comes into picture only if the kernel is not a linear kernel.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Expt #** | **Fold** | **Parameter1 Cost** | **Parameter2 kernel** | **Parameter3 gamma** | **Parameter4 tolerance** | **Average Accuracy(in %)** | **Average Precision** | **Average Recall** | **F-measure** |
| 1 | 10 | 10 | linear | - | 0.01 | 78.69 | 0.767 | 0.776 | 0.771 |
| 2 | 10 | 12 | linear | - | 0.05 | 78.74 | 0.767 | 0.776 | 0.772 |
| 3 | 10 | 18 | linear | - | 0.05 | 78.55 | 0.769 | 0.776 | 0.772 |
| **4** | **10** | **20** | linear | - | **0.01** | **78.73** | **0.769** | **0.780** | **0.775** |
| 5 | 10 | 15 | linear | - | 0.01 | 78.59 | 0.776 | 0.779 | 0.773 |

In polynomial kernel, if we are either increasing the cost or the value of gamma, the time required for processing the algorithm is huge comparatively. We are getting the best results in polynomial for cost 10, gamma 0.25 and tolerance 0.01. If we closely observe the values for polynomial kernel, the measures like accuracy, F-measure, precision and recall are better when the gamma values are less comparatively.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Expt #** | **Fold** | **Parameter1 Cost** | **Parameter2 kernel** | **Parameter3 gamma** | **Parameter4 tolerance** | **Average Accuracy(in %)** | **Average Precision** | **Average Recall** | **F-measure** |
| 1 | 10 | 10 | polynomial | 0.25 | 0.01 | 81.36 | 0.788 | 0.813 | 0.800 |
| 2 | 10 | 12 | polynomial | 0.33 | 0.05 | 80.96 | 0.783 | 0.807 | 0.795 |
| 3 | 10 | 18 | polynomial | 0.5 | 0.05 | 79.14 | 0.762 | 0.792 | 0.776 |
| **4** | **10** | **15** | polynomial | 0.67 | **0.01** | **80.23** | **0.779** | **0.797** | **0.788** |
| 5 | 10 | 10 | polynomial | 0.75 | 0.01 | 79.40 | 0.768 | 0.790 | 0.779 |
| 6 | 10 | 20 | polynomial | 0.33 | 0.05 | 81.01 | 0.784 | 0.805 | 0.794 |
| 7 | 10 | 12 | polynomial | 0.67 | 0.01 | 78.64 | 0.763 | 0.784 | 0.773 |
| 8 | 10 | 15 | polynomial | 0.25 | 0.05 | 79.66 | 0.769 | 0.788 | 0.778 |
| 9 | 10 | 18 | polynomial | 0.5 | 0.01 | 79.99 | 0.774 | 0.796 | 0.784 |
| 10 | 10 | 20 | polynomial | 0.5 | 0.01 | 78.40 | 0.760 | 0.774 | 0.767 |

Radial kernel is producing consistent and better measures of precision, recall, accuracy and F-measure. With the decrease in gamma value and increase in cost value, the measures are increasing. If we closely observe, the tolerance value effects the measures of the experiment. Lower the value of tolerance, better the values of accuracy, precision, recall and F-measures.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Expt #** | **Fold** | **Parameter1 Cost** | **Parameter2 kernel** | **Parameter3 gamma** | **Parameter4 tolerance** | **Average Accuracy(in %)** | **Average Precision** | **Average Recall** | **F-measure** |
| 1 | 10 | 10 | radial | 1 | 0.001 | 81.25 | 0.783 | 0.815 | 0.798 |
| 2 | 10 | 15 | radial | 0.5 | 0.01 | 81.31 | 0.782 | 0.816 | 0.799 |
| 3 | 10 | 18 | radial | 0.25 | 0.05 | 80.59 | 0.775 | 0.805 | 0.790 |
| **4** | **10** | **12** | **radial** | **0.33** | **0.01** | **80.31** | **0.772** | **0.805** | **0.788** |
| 5 | 10 | 20 | radial | 0.45 | 0.001 | 80.52 | 0.774 | 0.805 | 0.789 |
| 6 | 10 | 12 | radial | 0.67 | 0.01 | 81.19 | 0.780 | 0.817 | 0.798 |
| 7 | 10 | 15 | radial | 0.25 | 0.05 | 81.10 | 0.788 | 0.814 | 0.801 |
| 8 | 10 | 10 | radial | 0.33 | 0.01 | 80.91 | 0.778 | 0.813 | 0.795 |
| 9 | 10 | 18 | radial | 0.09 | 0.01 | 81.22 | 0.782 | 0.813 | 0.795 |
| 10 | 10 | 20 | radial | 0.25 | 0.05 | 80.18 | 0.769 | 0.806 | 0.787 |

Out of all the observations that we have observed, we can closely relate the data as:

Lower the values of gamma, tolerance and higher the value of cost we are obtaining the best results in the respective kernels.

Test Results:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **svm.pred** | **Pclass** | **Sex** | **Age** | **SibSp** | **Parch** | **Embarked** |
| 1 | 3 | 2 | 1 | 0 | 0 | 4 |
| 1 | 2 | 2 | 3 | 0 | 0 | 4 |
| 0 | 2 | 1 | 2 | 0 | 0 | 4 |
| 0 | 2 | 1 | 2 | 0 | 0 | 4 |
| 0 | 1 | 1 | 2 | 0 | 0 | 4 |
| 1 | 3 | 2 | 2 | 0 | 0 | 3 |
| 1 | 1 | 2 | 2 | 1 | 0 | 2 |
| 1 | 1 | 1 | 2 | 1 | 0 | 2 |
| 0 | 3 | 2 | 2 | 1 | 0 | 4 |
| 0 | 3 | 1 | 2 | 1 | 1 | 2 |
| 0 | 3 | 1 | 2 | 1 | 0 | 2 |
| 1 | 2 | 2 | 1 | 0 | 0 | 4 |
| 0 | 3 | 2 | 2 | 0 | 0 | 4 |
| 0 | 3 | 1 | 2 | 0 | 0 | 4 |
| 0 | 3 | 2 | 2 | 1 | 0 | 4 |
| 0 | 1 | 1 | 2 | 0 | 1 | 2 |
| 0 | 2 | 1 | 2 | 1 | 0 | 2 |
| 0 | 3 | 1 | 2 | 0 | 0 | 4 |
| 1 | 1 | 2 | 2 | 0 | 2 | 4 |
| 0 | 3 | 2 | 2 | 0 | 0 | 4 |
| 0 | 2 | 1 | 1 | 0 | 0 | 4 |
| 0 | 3 | 1 | 2 | 0 | 0 | 4 |
| 0 | 3 | 1 | 3 | 0 | 2 | 4 |
| 1 | 3 | 1 | 1 | 0 | 2 | 4 |
| 0 | 2 | 1 | 2 | 0 | 0 | 4 |
| 0 | 3 | 1 | 2 | 0 | 0 | 4 |
| 0 | 3 | 1 | 2 | 0 | 0 | 2 |
| 0 | 2 | 1 | 2 | 0 | 0 | 4 |
| 0 | 3 | 1 | 2 | 0 | 0 | 4 |
| 1 | 1 | 1 | 3 | 2 | 0 | 3 |
| 0 | 1 | 1 | 2 | 1 | 1 | 4 |
| 0 | 3 | 1 | 2 | 0 | 0 | 3 |
| 0 | 1 | 1 | 3 | 1 | 1 | 4 |
| 0 | 1 | 1 | 2 | 0 | 0 | 4 |
| 1 | 1 | 2 | 2 | 0 | 0 | 4 |
| 0 | 3 | 1 | 2 | 0 | 0 | 4 |
| 1 | 1 | 2 | 2 | 1 | 0 | 2 |
| 1 | 2 | 2 | 2 | 1 | 1 | 4 |
| 1 | 2 | 2 | 2 | 1 | 0 | 4 |
| 0 | 3 | 1 | 2 | 0 | 0 | 4 |
| 0 | 3 | 1 | 2 | 0 | 0 | 4 |
| 1 | 1 | 2 | 2 | 0 | 0 | 2 |
| 0 | 3 | 1 | 3 | 0 | 0 | 4 |
| 0 | 3 | 1 | 3 | 0 | 0 | 4 |
| 1 | 3 | 2 | 2 | 0 | 0 | 3 |
| 0 | 3 | 1 | 1 | 5 | 2 | 4 |
| 0 | 3 | 2 | 2 | 0 | 0 | 4 |
| 0 | 3 | 1 | 2 | 0 | 0 | 2 |
| 0 | 1 | 1 | 2 | 0 | 0 | 4 |
| 0 | 3 | 2 | 2 | 1 | 0 | 4 |
| 0 | 3 | 1 | 2 | 0 | 0 | 2 |
| 1 | 2 | 2 | 3 | 0 | 0 | 4 |
| 1 | 2 | 2 | 2 | 1 | 2 | 4 |
| 1 | 2 | 2 | 2 | 0 | 0 | 2 |
| 0 | 1 | 1 | 2 | 0 | 0 | 4 |
| 0 | 3 | 1 | 2 | 0 | 0 | 4 |
| 0 | 2 | 1 | 2 | 0 | 0 | 4 |
| 0 | 3 | 1 | 2 | 0 | 0 | 4 |
| 1 | 3 | 2 | 2 | 0 | 0 | 3 |
| 0 | 1 | 1 | 2 | 0 | 0 | 4 |
| 1 | 3 | 1 | 1 | 1 | 1 | 2 |
| 0 | 3 | 1 | 2 | 0 | 0 | 2 |
| 0 | 3 | 1 | 2 | 0 | 0 | 4 |
| 0 | 2 | 1 | 3 | 0 | 0 | 4 |
| 0 | 2 | 1 | 2 | 1 | 0 | 4 |
| 0 | 3 | 1 | 3 | 0 | 0 | 4 |
| 1 | 1 | 1 | 3 | 1 | 0 | 2 |
| 0 | 1 | 1 | 2 | 0 | 0 | 4 |
| 0 | 1 | 1 | 2 | 0 | 0 | 4 |
| 0 | 3 | 1 | 2 | 0 | 0 | 4 |
| 0 | 3 | 1 | 2 | 0 | 0 | 4 |
| 0 | 2 | 1 | 2 | 1 | 1 | 4 |
| 0 | 3 | 1 | 2 | 1 | 0 | 4 |
| 1 | 3 | 2 | 1 | 0 | 0 | 3 |
| 0 | 1 | 1 | 3 | 0 | 0 | 4 |
| 0 | 1 | 1 | 2 | 0 | 0 | 2 |
| 0 | 3 | 1 | 3 | 0 | 0 | 4 |
| 1 | 2 | 2 | 3 | 0 | 0 | 4 |
| 0 | 1 | 1 | 3 | 1 | 0 | 4 |
| 1 | 1 | 2 | 2 | 0 | 0 | 4 |
| 0 | 1 | 1 | 2 | 0 | 0 | 2 |
| 0 | 3 | 1 | 2 | 0 | 0 | 3 |
| 0 | 3 | 1 | 2 | 0 | 0 | 4 |
| 0 | 3 | 1 | 2 | 0 | 0 | 2 |
| 0 | 3 | 1 | 1 | 0 | 0 | 4 |
| 0 | 1 | 1 | 2 | 0 | 0 | 4 |
| 0 | 3 | 1 | 2 | 0 | 0 | 2 |
| 1 | 2 | 1 | 1 | 0 | 2 | 2 |
| 0 | 3 | 1 | 2 | 0 | 0 | 3 |
| 0 | 3 | 1 | 2 | 0 | 0 | 4 |
| 0 | 3 | 1 | 3 | 0 | 0 | 4 |
| 1 | 1 | 2 | 2 | 1 | 0 | 2 |
| 1 | 1 | 2 | 3 | 0 | 1 | 2 |
| 0 | 3 | 1 | 2 | 0 | 0 | 3 |

* From the results of testing data, svm model has predicted 27 survivals out of 94 instances.
* Ratio of number of males survived to number of females survived is 6:21
* Age 0-18:19-40:41+ ratio of survival is 6:15:6

The above 2 attributes show significant differences that can be seen from the results. The statistics show that the ***females*** and the ***middle-aged*** people had more chances of survival, compared to the other categories. Even, the training data set shows similar significant results. The training data has 303 entries who survived. 214 ***females*** survived and 188 ***middle-aged*** people survived in the training set.