**CONFUSION MATRICES**

**BEFORE CROSS VALIDATION**

**SVM NAÏVE BAYES**

**ACTUAL**

**ACTUAL**

**PREDICTION**

|  |  |  |
| --- | --- | --- |
|  | 0 | 1 |
| 0 | 109 | 55 |
| 1 | 30 | 94 |

|  |  |  |
| --- | --- | --- |
|  | 0 | 1 |
| 0 | 111 | 77 |
| 1 | 28 | 72 |

**PREDICTION**

True Positives:94 False Positive:30 True Positives: 72 False Positives:28

True Negatives:109 False Negatives:55 True Negatives:111 False Negatives:77

|  |  |  |
| --- | --- | --- |
|  | 0 | 1 |
| 0 | 94 | 66 |
| 1 | 37 | 91 |

**KNN (K-NEAREST NEIGHBOURS**)

**ACTUAL**

**PREDICTION**

True Positives:91 False Positive:37

True Negatives:94 False Negatives:66

**AFTER CROSS VALIDATION**

**SVM NAÏVE BAYES**

**ACTUAL**

**ACTUAL**

**PREDICTION**

|  |  |  |
| --- | --- | --- |
|  | 0 | 1 |
| 0 | 104 | 39 |
| 1 | 35 | 110 |

|  |  |  |
| --- | --- | --- |
|  | 0 | 1 |
| 0 | 114 | 79 |
| 1 | 25 | 70 |

**PREDICTION**

True Positives:110 False Positive:35 True Positives: 70 False Positives:25

True Negatives:104 False Negatives:39 True Negatives:114 False Negatives:79

|  |  |  |
| --- | --- | --- |
|  | 0 | 1 |
| 0 | 114 | 74 |
| 1 | 17 | 83 |

**KNN (K-NEAREST NEIGHBOURS**)

**ACTUAL**

**PREDICTION**

True Positives:83 False Positive:17

True Negatives:114 False Negatives:74

**ACCURACIES**

**SVM ACCURACY RESULTS:**

**BEFORE CROSS VALIDATION**

* SVM implementation in sklearn for **linear kernel** without cross validation gives a model that yields an accuracy of **70.48611111111111 %**
* SVM implementation in sklearn for **RBF (Radial Basis Function) kernel** without cross validation gives a model that yields an accuracy of **69.09722222222222 %**

**AFTER CROSS VALIDATION**

* Cross validation for Linear Kernel by fitting 10 folds for each of 5 candidates, totalling 50 fits, SVM implementation in sklearn for Linear kernel with cross validation and C value 1000 gives a model that yields an accuracy of: **72.56944444444445 %**
* Cross validation for RBF Kernel by fitting 10 folds for each of 20 candidates, totalling 200 fits SVM implementation in sklearn for linear kernel with cross validation and C value 10, gamma/spread value of, 0.1 gives a model that yields an accuracy of **74.30555555555556%**

**NAÏVE BAYES ACCURACY RESULTS:**

**BEFORE CROSS VALIDATION**

* Naive Bayes implementation in sklearn without cross validation gives a model that yields an accuracy of 63.5416666666667 **%**

**AFTER CROSS VALIDATION**

* Final accuracy that the cross validated model of Naive Bayes classifier yields is : **63.88888888888889%**

**KNN ACCURACY RESULTS**

Before Optimizing Parameters:

|  |  |
| --- | --- |
| Accuracy on Testing Set Without Optimizing Parameters | 64.23% |
| Accuracy on Training Set Using 5-Fold CV (K=20): | [63.005, 58.959, 59.537, 65.697, 68.023] |
| Average Accuracy on Training Set Using 5-Fold CV (K=20): | 63.04% |

While optimizing Parameters:

No. of Neighbours vs Accuracy:

|  |  |  |  |
| --- | --- | --- | --- |
| **No. Of Neighbours Considered** | **Accuracy** | **No. Of Neighbours Considered** | **Accuracy** |
| 1 | 0.606028 | 51 | 0.651203 |
| 3 | 0.592208 | 53 | 0.650027 |
| 5 | 0.589936 | 55 | 0.644253 |
| 7 | 0.601497 | 57 | 0.644253 |
| 9 | 0.613045 | 59 | 0.636140 |
| 11 | 0.631569 | 61 | 0.632652 |
| 13 | 0.625795 | 63 | 0.633855 |
| 15 | 0.629217 | 65 | 0.629177 |
| 17 | 0.629217 | 67 | 0.640765 |
| 19 | 0.638532 | 69 | 0.637289 |
| 21 | 0.646659 | 71 | 0.641954 |
| 23 | 0.646645 | 73 | 0.633814 |
| 25 | 0.637330 | 75 | 0.631502 |
| 27 | 0.647821 | 77 | 0.626851 |
| 29 | 0.648931 | 79 | 0.626838 |
| 31 | 0.660505 | 81 | 0.625702 |
| 33 | 0.663980 | 83 | 0.625702 |
| 35 | 0.651243 | 85 | 0.632678 |
| 37 | 0.652406 | 87 | 0.625728 |
| 39 | 0.646632 | 89 | 0.632652 |
| 41 | 0.646632 | 91 | 0.626891 |
| 43 | 0.643130 | 93 | 0.626878 |
| 45 | 0.648877 | 95 | 0.622267 |
| 47 | 0.648864 | 97 | 0.623456 |
| 49 | 0.654665 | 99 | 0.622280 |

After optimizing Parameters:

|  |  |
| --- | --- |
| Best Parameter: | 33 Neighbours |
| Accuracy on Training Set After Optimizing Parameters | 66.39% |
| Accuracy on Testing Set After Optimizing Parameters: | 68.40% |

INFERENCE:

1. It is visible that as we start with a small number of neighbours, we get low accuracy. Accuracy increases as we increase the number of neighbours.
2. The Accuracy reaches a peak point after which it stays stagnant for a few increases in neighbours after which the accuracy is inconsistent with the increase of neighbours.

**Project Members**

|  |  |  |
| --- | --- | --- |
| Akashdeep Singh | singa190@mcmaster.ca | 400353314 |
| Sulaiman Ghouse Mohiuddin | mohius3@mcmaster.ca | 400418055 |
| Vikram Chawla | chawlv3@mcmaster.ca | 400354764 |