**ROC (Receiver Operating Characteristic) CURVES**

**ROC Curve for SVM**

Chart

Description automatically generated

**ROC CURVE FOR NAÏVE BAYES**

Chart, line chart

Description automatically generated

**ROC FOR KNN**

**Chart, line chart

Description automatically generated**

**AREAS UNDER ROC CURVE**

|  |  |  |
| --- | --- | --- |
| SVM | Naïve BAYES | KNN |
| 0.7915117570373232 | 0.690454347931051 | 0.7513735595857441 |

**CROSS VALIDATION FOR PARAMETER SELECTION**

* **SVM (SUPPORT VECTOR MACHINES)**
* LINEAR KERNEL:

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 error percentage of: **27.430555555555557.**

* RBF (RADIAL BASIS FUNCTION) KERNEL:

Fitting 10 folds for each of 20 candidates, totalling 200 fits.

SVM implementation in Sklearn for RBF kernel with cross validation and C value 10, gamma/spread value of, 0.1 gives a model that yields an error percentage of: **25.694444444444443**

* **NAÏVE BAYES**

**Total number of FOLDS= 10**

|  |  |
| --- | --- |
| FOLD NUMBER | SCORE |
| 1 | 34.48275862068966 |
| 2 | 42.5287356321839 |
| 3 | 40.229885057471265 |
| 4 | 34.883720930232556 |
| 5 | 36.04651162790697 |
| 6 | 32.55813953488372 |
| 7 | 40.69767441860465 |
| 8 | 33.72093023255814 |
| 9 | 40.69767441860465 |
| 10 | 34.883720930232556 |

Mean of errors over model training is: **37.07297514033681**

Standard Deviation over model training is: **3.574240479489708**

Final error that the cross validated model of Naive Bayes classifier yields is: **36.11111111111111**

* **KNN (K-NEAREST NEIGHBOURS)**
* KNN Using Cross-Validation:

The Accuracy Rates of the different folds are:

|  |  |
| --- | --- |
| FOLD NUMBER | ACCURACY RATE |
| 1 | 63.00578035 |
| 2 | 58.95953757 |
| 3 | 59.53757225 |
| 4 | 65.69767442 |
| 5 | 68.02325581 |

The Average Accuracy is: 63.044764081193705

* KNN Using Grid-Search:

Fitting 10 folds for each of 50 candidates, totalling 500 fits

The Training Time is: 12.2240s

The optimal number of neighbours to use for best accuracy is: KNeighborsClassifier**(n\_neighbors=33**)

Accuracy for the Training Set with the most optimal number of neighbours is: **66.39802191927294**

**Project Members**

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