|  |
| --- |
| McMaster University |
| 4SL3 Assignment 3 |
|  |

|  |
| --- |
| Uday Sharma - 400139248  11-13-2021 |

## Precision/Recall Plots.

Chart, histogram

Description automatically generated

**

Chart

Description automatically generated**

Cross Validation errors for KNN Classifiers:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| K | 1 | 2 | 3 | 4 | 5 |
| SKlearn | 0.093122185995 | 0.094938674118 | 0.0790715727 | 0.08087253532 | 0.073792889302 |
| Manual | 0.08782797702 | 0.2176680639 | 0.0965378046 | 0.16848315478 | 0.101800962583 |

* For the SKLearn Implementation: K = 5 was chosen
* For the Manual Implementation: K = 1 was chosen

This disparity is due to the implementation specifics for the manual method, a score was assigned across multiple features, and a majority classifier was used.

Ties in the number of K neighbors in each class were dealt with in a method to favour false-positives as opposed to false-negatives and ties in the distances were dealt with to favour positive classifications as opposed to negative.

Misclassification Rates/Test Errors:

|  |  |
| --- | --- |
| **Logistic Regression SKLearn** | 0.02631578947368418 |
| **Logistic Regression Manual** | 0.06140350877192982 |
| **KNN SKLearn** | 0.07379288930290327 |
| **KNN Manual** | 0.0878279770222015 |

As we can see above, the SKLearn implementations are generally better than the ones implemented manually. Of the four models, the Logistic Regression SKLearn implementation has the smallest misclassification rate of the 4 models built in this assignment.

F1 Scores:

|  |  |
| --- | --- |
| **Logistic Regression SKLearn** | 0.9787234042553192 |
| **Logistic Regression Manual** | 0.9503546099290779 |
| **KNN SKLearn** | 0.943661971830986 |
| **KNN Manual** | 0.732919254658385 |

As per the f-scores, the winner remains the same between the 4 models. With the Logistic Regression SKLearn is still the best of the 4 models.