**Student’s Name: Vikas Dangi**

**Roll Number: B20238**

**Mobile No: 9406661661**

**Branch:EE**

# a.

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 81 | 27 |
| 27 | 201 |

Figure 1 KNN Confusion Matrix for K = 1

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 83 | 25 |
| 12 | 216 |

Figure 2 KNN Confusion Matrix for K = 3

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 82 | 26 |
| 9 | 219 |

Figure 3 KNN Confusion Matrix for K = 5

**b.**

Table 1 KNN Classification Accuracy for K = 1, 3 and 5

|  |  |
| --- | --- |
| **K** | **Classification**  **Accuracy (in %)** |
| 1 | **0.839** |
| 3 | **0.889** |
| 4 | **0.896** |

# Inferences:

1. The highest classification accuracy is obtained with K =.5
2. Increasing the value of K increases the prediction accuracy.
3. Increasing the value of K increases the prediction accuracy because as we increase the value of K, it takes in account more values which means it checks the distance from greater part of the distribution giving us more accuracy. Also, it removes the possibility of prediction getting affected by deviated value.
4. As the classification accuracy increases with the increase in value of K does the magnitude of diagonal elements increase.
5. The increase in diagonal elements due to a greater number of elements getting predicted accurately.
6. As the classification accuracy increases with the increase in value of K infer the number of off-diagonal elements decrease because the total number of elements is constant.
7. The reason for decrease in off-diagonal elements is because the total number of elements is constant and the accuracy increases due to the increase in k hence more values lie in the diagonal side reducing the off-diagonal values.

# a.

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 104 | 4 |
| 9 | 219 |

Figure 4 KNN Confusion Matrix for K = 1 post data normalization

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 105 | 3 |
| 7 | 221 |

Figure 5 KNN Confusion Matrix for K = 3 post data normalization

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 104 | 4 |
| 7 | 221 |

Figure 6 KNN Confusion Matrix for K = 5 post data normalization

**b.**

Table 2 KNN Classification Accuracy for K = 1, 3 and 5 post data normalization

|  |  |
| --- | --- |
| **K** | **Classification**  **Accuracy (in %)** |
| 1 | **0.961** |
| 3 | **0.970** |
| 5 | **0.967** |

# Inferences:

1. The data normalization increases classification accuracy.
2. The reason for increase in classification accuracy after data normalization, is that the normalization helps attributes prevent overweighting attributes with smaller ranges
3. The highest classification accuracy is obtained with K =3
4. Increasing the value of K increases the prediction accuracy.
5. Increasing the value of K increases the prediction accuracy because more the value of K more it represents the greater part of the distribution also it prevents the exceptional values to create a large impact on the outcome.
6. As the classification accuracy increases with the increase in value of K infer the number of diagonal elements increase.
7. The number in the diagonal elements represent the elements which are correctly predicted hence more the accuracy more the value of diagonal elements.
8. As the classification accuracy increases with the increase in value of K infer does the off-diagonal elements decrease.
9. The reason for decrease in off-diagonal elements is because the total number of elements is constant and the accuracy increases due to the increase in k hence more values lie in the diagonal side reducing the off-diagonal values.

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 105 | 13 |
| 5 | 214 |

Figure 7 Confusion Matrix obtained from Bayes Classifier

The classification accuracy obtained from Bayes Classifier is 94.6 %.

Table 3 Mean for class 0 and class 1

|  |  |  |  |
| --- | --- | --- | --- |
| **S. No.** | **Attribute Name** | **Mean** | |
| **Class 0** | **Class 1** |
|  | X\_Maximum | 273.418 | 723.656 |
|  | Y\_Maximum | 1583169.659 | 1431588.69 |
|  | Pixels\_Areas | 7779.663 | 585.967 |
|  | X\_Perimeter | 393.835 | 54.491 |
|  | Y\_Perimeter | 273.183 | 45.658 |
|  | Sum\_of\_Luminosity | 843350.275 | 62191.126 |
|  | Minimum\_of\_Luminosity | 53.326 | 96.236 |
|  | Maximum\_of\_Luminosity | 135.762 | 130.452 |
|  | Length\_of\_Conveyer | 1382.762 | 1480.018 |
|  | Steel\_Plate\_Thickness | 40.073 | 104.214 |
|  | Edges\_Index | 0.123 | 0.385 |
|  | Empty\_Index | 0.459 | 0.427 |
|  | Square\_Index | 0.592 | 0.513 |
|  | Outside\_X\_Index | 0.108 | 0.02 |
|  | Edges\_X\_Index | 0.55 | 0.608 |
|  | Edges\_Y\_Index | 0.523 | 0.831 |
|  | Outside\_Global\_Index | 0.288 | 0.608 |
|  | LogOfAreas | 3.623 | 2.287 |
|  | Log\_X\_Index | 2.057 | 1.227 |
|  | Log\_Y\_Index | 1.848 | 1.318 |
|  | Orientation\_Index | -0.314 | 0.136 |
|  | Luminosity\_Index | -0.115 | -.116 |
|  | SigmoidOfAreas | 0.925 | 0.543 |

In Fig. 8 and 9 representing covariance matrices for class 0 and class 1 respectively the column numbers and row numbers correspond to attribute with serial number as in Table 3.

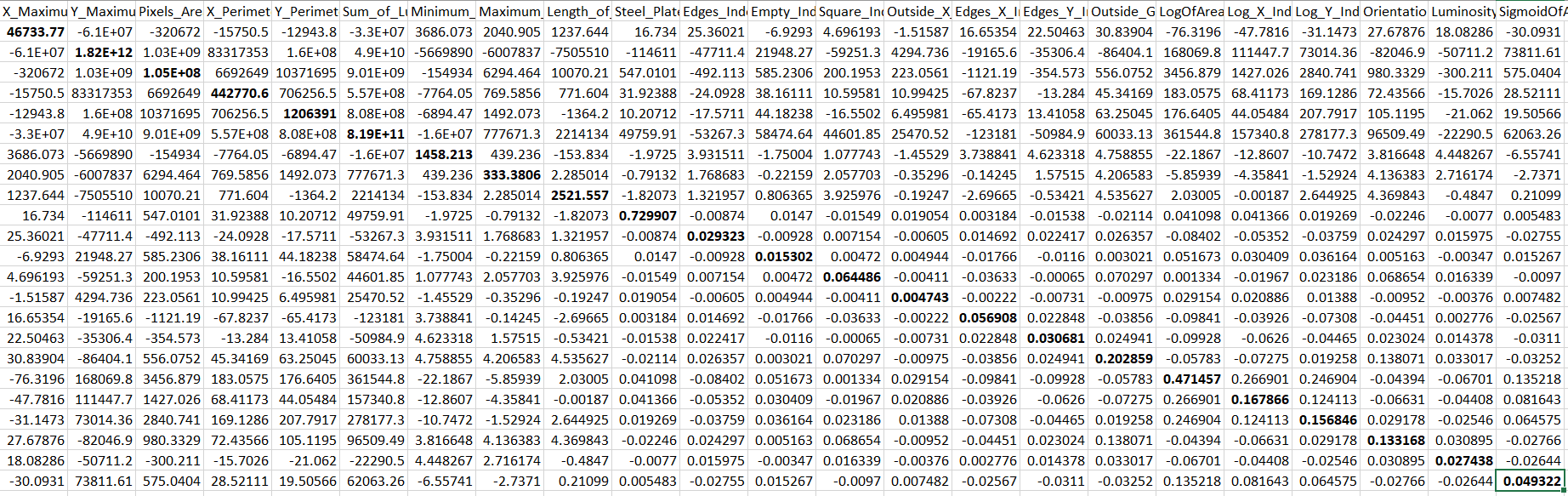


Figure 8: Covariance matrix for class 0

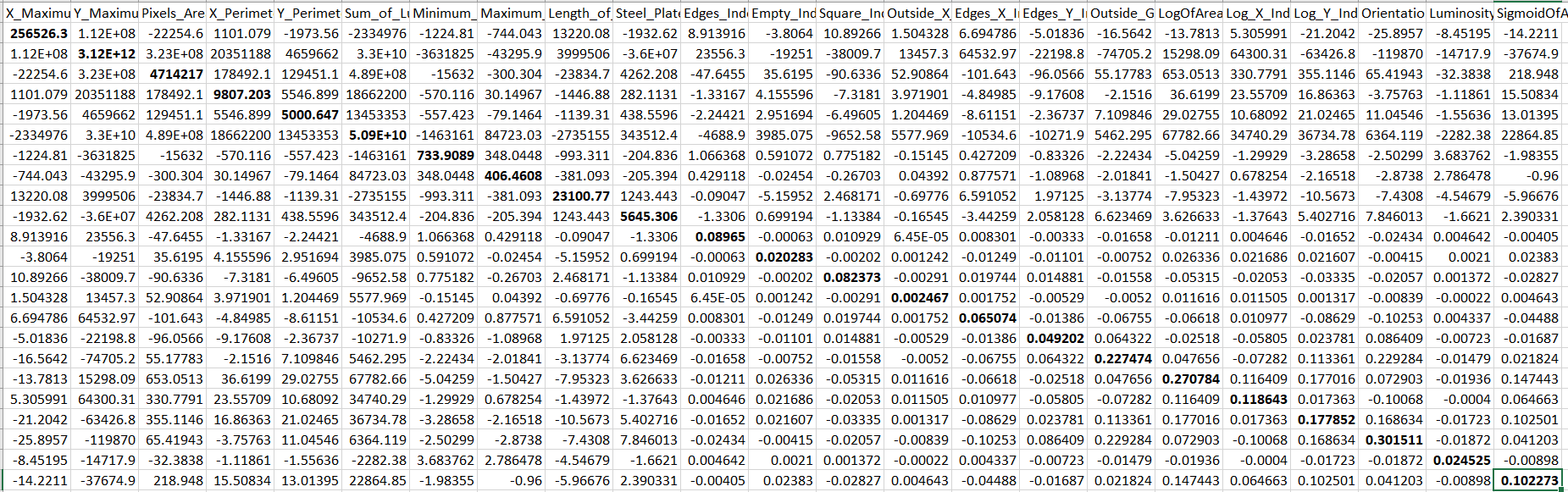


Figure 9: Covariance matrix for class 1

# Inferences:

1. The classification accuracy obtained from Bayes Classifier is 94.6 %.
2. The accuracy of Bayes classifier is 94.6 which is better than the KNN-classifier since Bayes classifier is not affected by outliers and the scale difference between features of dataset. However, it’s slightly less accurate than KNN-classifier on normalized data as Bayes classifier assumes the features to be conditionally independent but they aren’t totally in actuality.
3. From covariance matrix the nature of values along the diagonal is high which tells us that the variance of the attribute is high.
4. Some of the off-diagonal elements show very high covariance which could lead to wrong classification while using Bayes classifier which requires features to be near independent. However, we have dropped the features with max covariance.
5. Max covariance is between (Y\_Maximum and Pixel\_area) and (X\_Maximum & Y\_Maximum)while Min covariance is between (SigmoidOfAreas & Luminosity\_Index) and (Luminosity\_Index & Orientaion).

Table 4 Comparison between classifiers based upon classification accuracy

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Classifier** | **Accuracy (in %)** |
|  | KNN | 0.896 |
|  | KNN on normalized data | 0.970 |
|  | Bayes | 0.946 |

# Inferences:

1. KNN on normalized data has highest accuracy and simple KNN classifier has lowest accuracy.
2. Simple KNN classifier < Bayes Classifier < KNN on normalized data
3. State the reasons behind Inference 1 and 2.