**Student’s Name: ISHAAN GUPTA**

**Roll Number: B20292**

**Mobile No: 9179242114**

**Branch:** **MECHANICAL ENGINEERING**

# 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 | **83.928** |
| 3 | **88.988** |
| 5 | **89.583** |

# 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 as if there will be more nearest neighbors then there will be more accuracy.
4. As the classification accuracy increases with the increase in value of K, the number of diagonal elements increase.
5. Increase in diagonal elements with increase in k is because more accuracy leads to more number of true values.
6. As the classification accuracy increases with the increase in value of K, the number of off-diagonal elements decrease.
7. Decrease in off-diagonal elements with increase in k is because more accuracy leads to less number of false values.

# a.

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 100 | 8 |
| 8 | 220 |

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

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 100 | 8 |
| 7 | 221 |

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

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 101 | 7 |
| 4 | 224 |

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 | **95.238** |
| 3 | **95.535** |
| 5 | **96.726** |

# Inferences:

1. Data normalization increases classification accuracy.
2. Data normalization increases classification accuracy because bias is present in data which gets removed by normalization and range of different features or attributes become even.
3. The highest classification accuracy is obtained with K=5.
4. Increasing the value of K increases the prediction accuracy.
5. Increasing the value of K increases the prediction accuracy as if there will be more nearest neighbors then there will be more accuracy.
6. As the classification accuracy increases with the increase in value of K, the number of diagonal elements increase.
7. Increase in diagonal elements with increase in k is because more accuracy leads to more number of true values.
8. As the classification accuracy increases with the increase in value of K, the number of off-diagonal elements decrease.
9. Decrease in off-diagonal elements with increase in k is because more accuracy leads to less number of false values.

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 96 | 12 |
| 2 | 226 |

Figure 7 Confusion Matrix obtained from Bayes Classifier

The classification accuracy obtained from Bayes Classifier is 95.833 %.

Table 3 Mean for class 0 and class 1

|  |  |  |  |
| --- | --- | --- | --- |
| **S. No.** | **Attribute Name** | **Mean** | |
| **Class 0** | **Class 1** |
|  | X\_Maximum | 286.3322 | 746.584 |
|  | Y\_Maximum | 1711478 | 1445964 |
|  | Pixels\_Areas | 7268.032 | 583.512 |
|  | X\_Perimeter | 355.6148 | 52.184 |
|  | Y\_Perimeter | 207.1555 | 43.112 |
|  | Sum\_of\_Luminosity | 808615.7 | 61552.41 |
|  | Minimum\_of\_Luminosity | 53.40283 | 94.804 |
|  | Maximum\_of\_Luminosity | 135.8587 | 130.184 |
|  | Length\_of\_Conveyer | 1382.516 | 1486.63 |
|  | Steel\_Plate\_Thickness | 40.24735 | 100.434 |
|  | Edges\_Index | 0.126447 | 0.388864 |
|  | Empty\_Index | 0.449608 | 0.418643 |
|  | Square\_Index | 0.593253 | 0.510322 |
|  | Outside\_X\_Index | 0.108173 | 0.019854 |
|  | Edges\_X\_Index | 0.565851 | 0.625601 |
|  | Edges\_Y\_Index | 0.524692 | 0.837443 |
|  | Outside\_Global\_Index | 0.268551 | 0.611 |
|  | LogOfAreas | 3.599567 | 2.264311 |
|  | Log\_X\_Index | 2.048011 | 1.214075 |
|  | Log\_Y\_Index | 1.825003 | 1.299494 |
|  | Orientation\_Index | -0.32807 | 0.131946 |
|  | Luminosity\_Index | -0.10907 | -0.12263 |
|  | SigmoidOfAreas | 0.91587 | 0.527024 |

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.

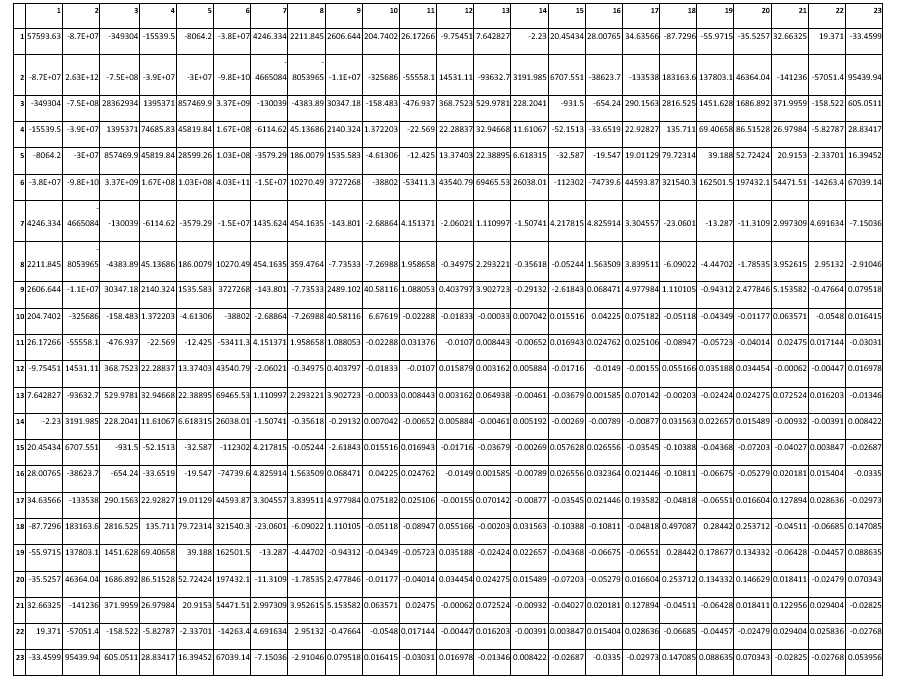


Figure 8: Covariance matrix for class 0

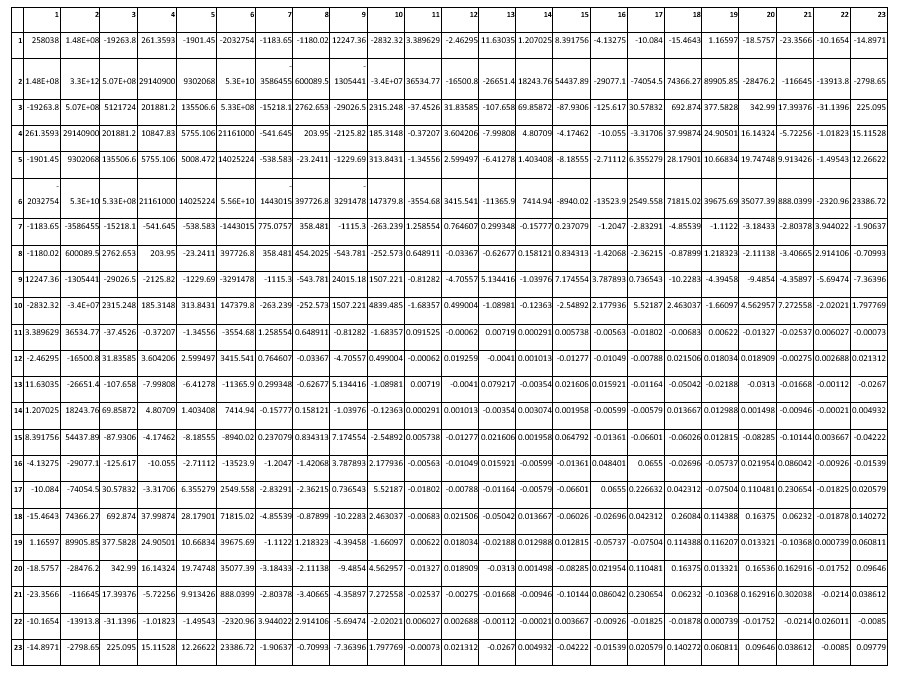


Figure 9: Covariance matrix for class 1

# Inferences:

1. The accuracy of Bayes Classifier is 95.833 %. and state reason why it is lesser / greater than previous classification approaches.
2. Infer from covariance matrix the nature of values along the diagonal. State the reason.
3. Infer from off-diagonal elements the covariance between attributes. Write 2 pair of attributes with maximum and 2 pair of attributes with minimum covariance.

Table 4 Comparison between classifiers based upon classification accuracy

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Classifier** | **Accuracy (in %)** |
|  | KNN | 89.583 |
|  | KNN on normalized data | 96.726 |
|  | Bayes | 95.833 |

# Inferences:

1. The classifier with highest accuracy is KNN on normalized data and lowest accuracy is KNN.
2. The classifiers in ascending order of classification accuracy= KNN < Bayes < KNN on normalized data .