**Student’s Name: Gayatri Shridhar Kapse**

**Roll Number: B20199**

**Mobile No: 8623916561**

**Branch:EE**

# a.

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 93 | 25 |
| 19 | 200 |

Figure 1 KNN Confusion Matrix for K = 1

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 92 | 26 |
| 9 | 210 |

Figure 2 KNN Confusion Matrix for K = 3

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 92 | 26 |
| 10 | 209 |

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 | 86.9436 |
| 3 | 89.6142 |
| 5 | 89.3175 |

# Inferences:

1. The highest classification accuracy is obtained with K =.3
2. Increasing the value of K at first it increases and then decreases again.
3. By increasing the value of K we are actually considering the Euclidian distances of K points from the test point that’s why we get high accuracy on increasing the K value.
4. As the classification accuracy increases with the increase in value of K the number of diagonal elements increase.
5. The reason for increase in the number of diagonal element is because as the accuracy increase more number of predicted values are equal to the actual value or we can say our prediction is right.

# a.

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 111 | 7 |
| 6 | 213 |

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

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 113 | 5 |
| 4 | 215 |

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

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 109 | 9 |
| 4 | 215 |

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 | 96.1424 |
| 3 | 97.3293 |
| 5 | 96.1424 |

# Inferences:

1. Normalization increases classification accuracy.
2. There is an increase in classification accuracy after normalization because as we are scaling the values of the attributes so that their will not be any data point that will be dominating the output.
3. The highest classification accuracy is obtained with K =3.
4. The classification accuracy first increases and the decreases with the increase in value of K.
5. By increasing the value of K we are actually considering the Euclidian distances of K points from the test point that’s why we get high accuracy on increasing the K value thereby calculating the eulcidian distance between the test data point and K number of train data points.

|  |  |  |
| --- | --- | --- |
|  | **Prediction Outcome** | |
| **True Label** | 105 | 13 |
| 6 | 213 |

Figure 7 Confusion Matrix obtained from Bayes Classifier

The classification accuracy obtained from Bayes Classifier is 94.4%.

Table 3 Mean for class 0 and class 1

|  |  |  |  |
| --- | --- | --- | --- |
| **S. No.** | **Attribute Name** | **Mean** | |
| **Class 0** | **Class 1** |
|  | X\_Minimum |  |  |
|  | X\_Maximum | 273.418 | 723.656 |
|  | Y\_Minimum |  |  |
|  | Y\_Maximum | 1583169.659 | 1431588.690 |
|  | 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 |
|  | TypeOfSteel\_A300 |  |  |
|  | TypeOfSteel\_A400 |  |  |
|  | 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 | 0.020 |
|  | Edges\_X\_Index | 0.550 | 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 | -0.116 |
|  | 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.

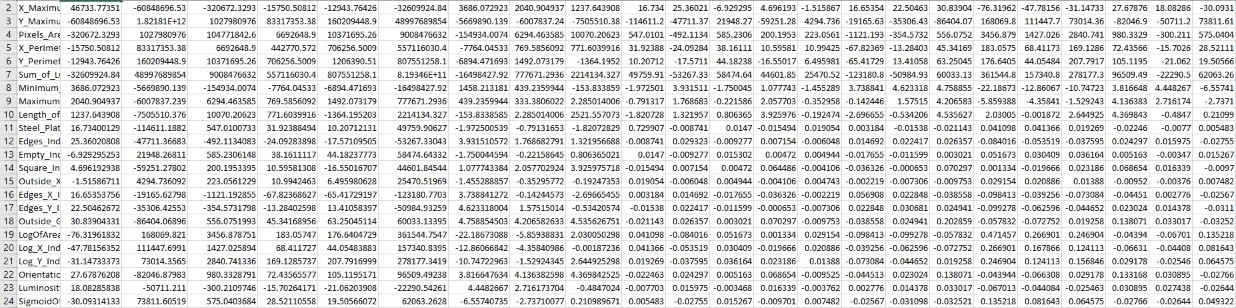
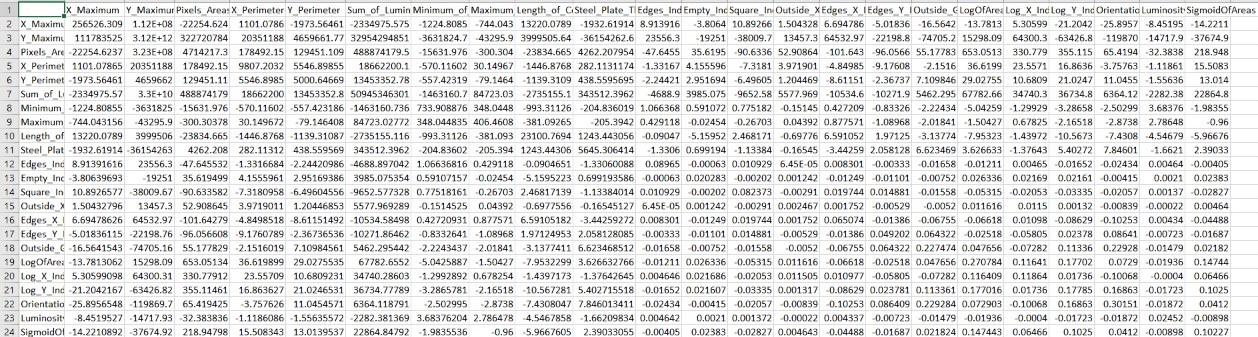


Figure 8: Covariance matrix for class 0

class\_0

Figure 9: Covariance matrix for class 1



Class\_1

# Inferences:

1. The accuracy of bayes classifier is 94.4% and it is lesser than previous classification approach because the previous KNN Classifier used Normal distribution.
2. The nature of values along the diagonal for some attribute is high and for some other its low because some attribute follow standard normal distribution.
3. The off-diagonal values have varied values. The two pair of attributes having maximum covariance is (Y\_Maximum, Sum of Luminosity) and (Y\_Maximum,Pixel\_Area). The two pair of attribute with minimum covariance is (Outside\_X,Edges\_X) and (Outside\_X , Empty\_index) .

Table 4 Comparison between classifiers based upon classification accuracy

|  |  |  |
| --- | --- | --- |
| **S. No.** | **Classifier** | **Accuracy (in %)** |
|  | KNN | 89.6 |
|  | KNN on normalized data | 97.3 |
|  | Bayes | 94.4 |

# Inferences:

1. KNN classifier with normalized data has highest accuracy and KNN classifier has the lowest accuracy.
2. KNN< Bayes< KNN on normalized data.
3. Usually Bayes classifier has higher accuracy but in this case the KNN classifier used is using Normalised data points that is why its accuracy is high.