

Student's Name: Gajraj Singh Chouhan

Mobile No: +91-9351159849

Roll Number: B19130

Branch: DSE

1 a.

Prediction Outcome			
True Label		Negative Class	Positive Class
	Negative Class	671 (TN)	54 (FP)
	Positive Class	46 (FN)	5 (TP)

Figure 1 KNN Confusion Matrix for K = 1

Prediction Outcome			
True Label		Negative Class	Positive Class
	Negative Class	707	18
	Positive Class	47	4

Figure 2 KNN Confusion Matrix for K = 3

Prediction Outcome			
True Label		Negative Class	Positive Class
	Negative Class	718	7
	Positive Class	46	5

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	87.113
3	91.623
5	93.170

Inferences:

1. The highest classification accuracy is obtained with K=5.
2. Increasing the value of K increases the prediction accuracy.
3. If we increase K, the areas predicting each class will be more smoothened.. Thus there would be an increase in the accuracy.
4. As the classification accuracy increases with the increase in value of K, the number of diagonal elements showing correct positive and negative predictions increases.
5. Since accuracy is equal to percent of correct predictions (of negative or positive), thus the diagonal elements representing these will increase (like they have from 671 to 718 for True Negative).

6. As the classification accuracy increases with the increase in value of K, the number of off-diagonal elements decreases.
7. Since sum of total number of elements = Correct + Incorrect predictions, as correct predictions increase and sum is remaining constant thus incorrect predictions will start to decrease as the True Positive has decreased to just 5 for k=5.
8. We can see False Negatives are not changing much. This could probably be due to the outliers in the train and test.

2 a.

Prediction Outcome			
True Label		Negative Class	Positive Class
	Negative Class	676	47
	Positive Class	42	9

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

Prediction Outcome			
True Label		Negative Class	Positive Class
	Negative Class	703	20
	Positive Class	44	7

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

Prediction Outcome			
True Label		Negative Class	Positive Class
	Negative Class	716	7
	Positive Class	48	3

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

b.

Table 2 KNN Classification Accuracy for K = 1,2,3,4 and 5 post data normalization

K	Classification Accuracy (in %)
1	88.501
3	91.731
5	92.894

Inferences:

1. Data normalization decreases classification accuracy.
2. Sometimes after data normalization, some features may get removed this can lead to decrease in accuracy.
3. The highest classification accuracy is obtained with K=5.
4. Increasing the value of K increases the prediction accuracy.
5. If we increase K, the areas predicting each class will be more smoothened.

6. As the classification accuracy increases with the increase in value of K, the number of diagonal elements increases.
7. Diagonal elements represent no. of correctly predicted values (TP/TN), which is directly proportional to accuracy score
8. As the classification accuracy increases with the increase in value of K, the number of off-diagonal elements decreases.
9. Values of off-diagonal elements represent no. of FP/FN values, which is inversely proportional to accuracy score.

3

Prediction Outcome			
True Label		Negative Class	Positive Class
	Negative Class	669	56
	Positive Class	39	12

Figure 7 Confusion Matrix obtained from Bayes Classifier

The classification accuracy obtained from Bayes Classifier is 87.757 %.

Table 3 Mean for Class 0 and Class 1

S. No.	Attribute Name	Class 0 Mean	Class 1 Mean
1.	seismic	1.335	1.496
2.	seismoacoustic	1.403	1.445
3.	shift	1.389	1.101
4.	genergy	76209.828	198697.395
5.	gpuls	490.057	944.824
6.	gdenergy	12.082	17.202
7.	gdpuls	3.542	10.639
8.	ghazard	1.107	1.076
9.	energy	4941.741	10278.992
10.	maxenergy	4374.600	8246.218

Table 4 Covariance Matrix for Class 0

	seismic	seismoacoustic	shift	genergy	gpuls	gdenenergy	gdpuls	ghazard	energy	maxenergy
seismic	0.223	0.016	-0.058	341.106	53.938	5.440	4.665	0.016	1306.739	1133.043
seismoacoustic	0.016	0.285	-0.018	2326.935	34.331	8.157	7.394	0.091	-34.790	5.745
shift	-0.058	-0.018	0.238	-20720.277	-108.223	-2.791	-2.712	-0.008	-967.727	-765.351
genergy	341.106	2326.935	-20720.277	43147695877.399	76016422.409	808600.411	1021197.415	-3538.720	343322901.721	271746897.295
gpuls	53.938	34.331	-108.223	76016422.409	253960.842	12700.784	13244.251	18.993	2346354.498	2013481.006
gdenenergy	5.440	8.157	-2.791	808600.411	12700.784	6834.718	4165.206	8.992	279011.669	270563.881
gdpuls	4.665	7.394	-2.712	1021197.415	13244.251	4165.206	3928.186	6.550	278212.480	267202.824
ghazard	0.016	0.091	-0.008	-3538.720	18.993	8.992	6.550	0.124	-160.341	-120.558
energy	1306.739	-34.790	-967.727	343322901.721	2346354.498	279011.669	278212.480	-160.341	468144388.200	443099212.485
maxenergy	1133.043	5.745	-765.351	271746897.295	2013481.006	270563.881	267202.824	-120.558	443099212.485	426402725.329

Table 5 Covariance Matrix for Class 1

	seismic	seismoacoustic	shift	genergy	gpuls	gdenenergy	gdpuls	ghazard	energy	maxenergy
seismic	0.252	0.006	-0.033	629.014	88.588	3.281	1.664	0.005	3384.233	2889.603
seismoacoustic	0.006	0.300	-0.011	-1728.237	-8.963	7.342	7.154	0.059	1681.470	1108.902
shift	-0.033	-0.011	0.091	-15394.057	-74.846	-3.444	-0.777	0.001	-539.389	-389.446
genergy	629.014	-1728.237	-15394.057	98499436799.088	180520099.706	-794559.640	69419.220	-8909.632	1436182.097	103759960.405
gpuls	88.588	-8.963	-74.846	180520099.706	615028.282	7514.434	9052.453	3.700	997000.499	1235626.022
gdenenergy	3.281	7.342	-3.444	-794559.640	7514.434	4734.518	3430.124	6.315	-168083.863	-162052.621
gdpuls	1.664	7.154	-0.777	69419.220	9052.453	3430.124	3425.453	6.078	-127216.978	-136438.242
ghazard	0.005	0.059	0.001	-8909.632	3.700	6.315	6.078	0.071	805.840	854.102
energy	3384.233	1681.470	-539.389	1436182.097	997000.499	-168083.863	-127216.978	805.840	409162012.534	341912419.883
maxenergy	2889.603	1108.902	-389.446	103759960.405	1235626.022	-162052.621	-136438.242	854.102	341912419.883	300604879.647

Inferences:

1. Accuracy of Bayes Classifier is 87.757% and state reason why it is lesser / greater than previous classification approaches.
2. Along the diagonal, some of attributes like seismic, seismicacoustic, shift have lowest magnitude of covariance so they have lowest variance whereas others have large variance. Correlation can be used determining the relationship.
3. Genergy and shift has the most number of negative covariance in classes, rest of the attributes mostly have positive covariances.
4. Maximum -> 443099212.485 b.w energy-maxenergy in class 0, 341912419.883 b/w energy-maxenergy in class 1.
5. Minimum -> 0.008 for shift-ghazard in class 0 and 0.001 for shift-ghazard in class 1.

4

Table 6 Comparison between Classifier based upon Classification Accuracy

S. No.	Classifier	Accuracy (in %)
1.	KNN	93.170
2.	KNN on normalized data	92.894
3.	Bayes	87.757

Inferences:

1. KNN has the highest accuracy, and Bayes has the lowest.
2. Bayes < KNN on normalized < KNN.
3. Bayes classifier has good accuracy only for large no. of datasets, whereas normalizing the data sometimes removes important feature differences, hence reducing the accuracy.