Course name: Predictive Analytics   Course code: MPBA G513
Time: 3 PM – 6 PM (3 hours)   Total marks: 35
Note: Attempt all the questions

1. ROC curve is plotted between 'Recall' and '1-Specificity'. (True/False - Briefly explain)   **1**

2. Briefly explain 'F1-score' with its relation to 'Precision' and 'Sensitivity'.   **1**

3. For the given confusion matrix calculate the Diagnostic Odds Ratio and Matthews Correlation Coefficient   **4**

		Fecal occult blood screen test outcome	
Total population (pop.) = 2030		Test outcome **positive**	Test outcome **negative**
Patients with bowel cancer (as confirmed on endoscopy)	Actual condition positive	**True positive** (TP) = 20	**False negative** (FN) = 10
	Actual condition negative	**False positive** (FP) = 180	**True negative** (TN) = 1820

4. What is a Voronoi diagram and its relation with the K-Means Clustering algorithm? Explain with a pictorial presentation.   **2**

5. The tree obtained upon hierarchical clustering is known as a _____.   **1**

6. Flipping the sign of PCA loadings vector has no effect or meaning. (True/False - Briefly explain)   **1**

7. Hierarchical clustering using complete linkage measures the minimal intercluster dissimilarity, whereas single linkage measures the maximal intercluster dissimilarity. (True/False - Briefly explain)   **1**

8. The tree obtained after single linkage has more number of splittings as compared to the tree generated through hierarchical clustering using complete linkage. (True/False - Briefly explain)   **1**

9. Maximal Margin Classifier technique is widely used method for classification as it is robust successor of Support Vector Classifier. (True/False - Briefly   **1**

explain)

10. Briefly explain what is predictor space   1

        1

$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} > 0 \text{ if } y_i = 1$$

$$\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} < 0 \text{ if } y_i = -1$$

11. For binary classification (class A and B), A maximal margin classifier has above given properties of its separating hyperplane. The application of the sign (±) of f(x) is if $f(x_i) > 0$ then $y_i = 1$ (class A) and if $f(x_i) < 0$ then $y_i = -1$ (class B). What is the application of the magnitude/value of $f(x_i)$?

12. Maximal Margin Classifier (MMC) uses soft margin hyperplane as compared   2
to Support Vector Classifier (SVC - uses hard margin) as MMC is robust in the sense that adding new data points doesn't affect its prediction. (True/False - Briefly explain)

Briefly explain what is hard/soft margin and write comments on the sensitivity/robustness of both models.

13. Write the tuning hyperparameters of Support Vector Machines and their   1
effect on modeling performance.

14. Briefly explain the relation between LDA (Linear Discriminant Analysis) and   1
Support Vector Classifer (SVC)

15. A non-linear polynomial kernel is what makes a support vector classifier to   1
act like a Support Vector Machine also known as 'Kernal trick'. A polynomial kernal of degree 'd' = 1 is therefore a Support Vector Machine which can be used to separate data points which are not linearly separable. (True/False - Briefly explain)

16. Explain the use of **LabelEncoder()** function and the Python library in which   2
it is included. Also write how this function will encode two strings 'Churners' and 'Non-churners'.

17. If the decision boundary is linear, decision trees are most commonly found   1
to outperform linear regression. (True/False - Briefly explain)

18. Briefly explain what is Bootstrap aggregation and whether it allows data   2
point selection with/without replacement and how it is different from 'Random forests (RF)' with the relation as to how RF avoids creation of highly correlated trees.

19. Association rule mining can be implemented only as unsupervised learning   1
manner. (True/False - Briefly explain)

20. Briefly explain what is WCSS score and its application.   1

21. Briefly explain what is SSE, SSR, MSR and MSE in regression analysis   2

22. Briefly explain steps of CRISP-DM model   1

23. Write the Bayes Theorem and why Naive Bayes Classifier is called Naive with relation to its limitations/assumptions.   2

24. Briefly explain the application of **random_state** argument in following Python code snippet.   1

```
from sklearn.model_selection import train_test_split
X_train, X_test, Y_train, Y_test = train_test_split(X, Y,
    test_size=1, random_state = 31)
```

25. Briefly explain what is Q-Q plot and its graphical interpretation.   1

26. Write how z-scaling is performed in Python and using which Python function , and the module in which it is available.   1