Unit 03: Various Data Pre-processing Operations

1. _____ is the process of converting data to either 0 or 1 based on a threshold value.

   `Binarizing data`   Label encoding   One hot encoding   Standardizing data

2. Which of these represents the techniques of data transformation?

   Binarizing data   Label encoding   One hot encoding   `All of the above`

3. The process of _____ is used to convert textual labels into numeric form in order to prepare it to be used in a machine-readable form.

   Binarizing data   `Label encoding`   One hot encoding   Standardizing data

4. Where is the data integration carried out?

   Data warehousing   Data migration   Information integration   `All of the above`

5. The redundant data can be detected using

   `Concept of correlation analysis`   Concept of regression analysis   Concept of classification analysis   Concept of clustering analysis

6. The value +1 of correlation coefficient indicates

   No relationship   `A perfect positive relationship`   A perfect negative relationship   Just a number

7. The value -1 of correlation coefficient indicates

   No relationship   A perfect positive relationship   `A perfect negative relationship`   Just a number

8. Which of these techniques of data integration is also known as Uniform Data Access?

   `Virtual integration`   Physical data integration   Application based integration   Manual integration

9. Which of these are data integration tools?

   Microsoft   Talend   IBM   `All of the above`

10. Which of the following are data pre-processing operations?

    Data reduction   Data cleaning   Data transformation   `All of the above`

11. The data is cleaned by:

| Filling in the missing values | Smoothing noisy data | Detecting and removing outliers | All of the above |

12. The methods for smoothing noisy data are:

| Regression | Binning | Both of the above | None of the above |

13. Which of these can be used for smoothing of noisy data?

| Equal width binning | Equal frequency binning | Both of the above | None of the above |

14. _____ works by performing a greedy search to acquire the best feature subset from a given dataset.

| Recursive feature elimination | Univariate selection | Stepwise forward selection | Stepwise backward elimination |

15. Which process of discretization is also known as splitting?

| Top-down discretization | Bottom-up discretization | Left-right discretization | Right-left discretization |

16. Which of these represents the techniques of data transformation?

| A. Binarizing data | B. Label encoding | C. One hot encoding | D. All of the above |

17. _____ is the process of converting data to either 0 or 1 based on a threshold value.

| A. Binarizing data | B. Label encoding | C. One hot encoding | D. Standardizing data |

18. The process of _____ is used to convert textual labels into numeric form in order to prepare it to be used in a machine-readable form.

| A. Binarizing data | B. Label encoding | C. One hot encoding | D. Standardizing data |

19. Where is the data integration carried out?

| A. Data warehousing | B. Data migration | C. Information integration | D. All of the above |

20. The redundant data can be detected using

| A. Concept of correlation analysis | B. Concept of regression analysis | C. Concept of classification analysis | D. Concept of clustering analysis |

21. The value +1 of correlation coefficient indicates

| A. No relationship | **B. A perfect positive relationship** | C. A perfect negative relationship | D. Just a number |

22. The value -1 of correlation coefficient indicates

| A. No relationship | B. A perfect positive relationship | **C. A perfect negative relationship** | D. Just a number |

23. Which of these techniques of data integration is also known as Uniform Data Access?

| **A. Virtual integration** | B. Physical data integration | C. Application based integration | D. Manual integration |

24. Which of these are data integration tools?

| A. Microsoft | B. Talend | C. IBM | **D. All of the above** |

25. Which of the following are data pre-processing operations?

| A. Data reduction | B. Data cleaning | C. Data transformation | **D. All of the above** |

26. The data is cleaned by:

| A. Filling in the missing values | B. Smoothing noisy data | C. Detecting and removing outliers | **D. All of the above** |

27. The methods for smoothing noisy data are:

| A. Regression | B. Binning | **C. Both of the above** | D. None of the above |

28. Which of these can be used for smoothing of noisy data?

| A. Equal width binning | B. Equal frequency binning | **C. Both of the above** | D. None of the above |

29. _____ works by performing a greedy search to acquire the best feature subset from a given dataset.

| **A. Recursive feature elimination** | B. Univariate selection | C. Stepwise forward selection | D. Stepwise backward elimination |

30. Which process of discretization is also known as splitting?

| **A. Top-down discretization** | B. Bottom-up discretization | C. Left-right discretization | D. Right-left discretization |