

DEPARTMENT: CSE

SEMESTER: 7

SUBJECT NAME: DATA MINING AND

BUSINESS INTELLIGENCE



SUBJECT CODE: 2170715

FACULTY NAME: PROF. NENSI

KANSAGARA

IMPORTANT QUESTIONS

1. Define Following Terms:

- Data mart
- Enterprise Warehouse
- Virtual Warehouse

2. Differentiate between Operational Database System and Data Warehouse.

3. What is data mining? List challenges to data mining regarding data mining methodology and user interaction issues.

4. What is Cuboid? Explain Various OLAP Operations on Data Cube with Example.

5. Define Data Mining? List and Describe the Major Issues in Data Mining.

6. Explain different types of data on which mining can be performed.

7. What is Market Basket Analysis? Explain Association Rules with Confidence & Support.

8. Explain Mean, Median, Mode, Variance, Deviation, Standard Deviation in brief and solve all above all for below dataset.

13,15,16,16,19,23,29,35,41,44,53,62,69,72

9. List and Describe the methods for handling the missing values in data cleaning.

10. Write and explain algorithm for mining frequent item sets without candidate item set generation with suitable example.

11. Explain Baye's Theorem and Naive Bayesian Classification.

Using Naive Bayesian classification method, predict class label of X = (age = youth, income = medium, student = yes, credit_rating = fair) using following training dataset.

age	income	Student	credit_rating	Class: buys_computer
youth	high	no	Fair	no
youth	high	no	excellent	no
middle_aged	high	no	fair	yes
senior	medium	no	fair	yes
senior	low	yes	fair	yes
senior	low	yes	excellent	no
middle_aged	low	yes	excellent	Yes
youth	medium	no	fair	no
youth	low	yes	fair	yes
senior	medium	yes	fair	yes
youth	medium	yes	excellent	yes
middle_aged	medium	no	excellent	yes
middle_aged	high	yes	fair	yes
senior	medium	no	excellent	no



12. What is decision tree induction? Write Basic algorithm for inducing a decision tree from training tuples.

13. What are neural networks? Describe the various factors for which make them useful for classification and prediction in data mining. Explain how topology of neural network is designed.

14. Explain data marts. What are the different types of data marts? Give advantages of data marts over data warehouses.

15. Explain metadata repository.

16. What is OLAP? Give difference between OLAP and OLTP.

17. Explain Star, Snowflake, and Fact Constellation Schema for Multidimensional Database.

18. Define KDD. Explain KDD process in detail.

19. Explain three-tier data Warehouse architecture

20. Draw and Explain the data mining architecture.

21. Consider following database of ten transactions. Let $\text{min_sup} = 30\%$ and $\text{min_confidence} = 60\%$. A) Find all frequent itemsets using Apriori algorithm. B) Generate strong association rules.

TID	items bought
T1	pen, pencil
T2	book, eraser, pencil
T3	book, chalk, eraser, pen
T4	chalk, eraser, pen
T5	book, pen, pencil
T6	book, eraser, pen, pencil
T7	ink, pen
T8	book, pen, pencil
T9	eraser, pen, pencil
T10	book, chalk, pencil

22. Explain Following terms with Proper Diagram

- Concept Hierarchy
- Histogram
- Data Transformation
- Data Discretization
- Clustering
- Aggregation
- Forecasting.

23. Explain gain, Gain Ratio, Gini Index.

24. What is noise? List the data smoothing techniques. Explain data smoothing methods as noise removal technique to divide given data into bins of size 3 by bin partition (equal frequency), by bin means, by bin medians and by bin boundaries. Consider the data: 10, 2, 19, 18, 20, 18, 25, 28, 22.

25. Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order):

13, 15, 16, 16, 19, 20, 23, 29, 35, 41, 44, 53, 62, 69, 72

Use min-max normalization to transform the value 45

for age onto the range [0:0, 1:0] Use z-score

normalization to transform the value 45 for age, where the standard deviation of age is 20.64 years.

26. Transaction item list is given below. Draw FP tree.

TID	LIST OF ITEM_Ids
T100	I1,I2,I5
T200	I2,I4
T300	I2,I3
T400	I1,I2,I4
T500	I1,I3
T600	I2,I3
T700	I1,I3
T800	I1,I2,I3,I5
T900	I1,I2,I3

