**SWARNANDHRA COLLEGE OF ENGINEERING & TECHNOLOGY**

DEPARTMENT OF MASTER OF COMPUTR APPLICATIONS

**IV SEMESTER**

SUBJECT: Data Warehousing and Data Mining Subject Code: 16MC4T01

Regulation: R16

**UNIT-I**

1. 1)Discuss knowledge discovery process in data mining
2. 2)List out major issues in data mining
3. 3)Discuss about Data mining functionalities.
4. 4)Write about classification of data mining systems
5. 5)List out the five primitives for specifying a data mining task
6. 6) Illustrate the architecture of a typical data mining system
7. 7) Are all patterns interesting? Analyze

**UNIT-II**

1. Explain Various data cleaning techniques in pre-processing
2. Discuss Data Reduction Techniques in detail
3. Explain Concept hierarchy generation for numerical data
4. Calculate Chi-Square value from the following data samples

|  |  |  |
| --- | --- | --- |
|  | male | female |
| Fiction | 250 | 200 |
| Non fiction | 50 | 1000 |

1. Predict the salary of 10,15 years of experience from the following data set using Linear regression the method of least squares

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X(years Experience) | 3 | 8 | 9 | 13 | 3 | 6 | 11 | 21 |
| Y(salary in 1000’s) | 30 | 57 | 64 | 72 | 36 | 43 | 59 | 90 |

1. Discuss issues to be considered during data integration

**UNIT-III**

1. Define Data warehouse. Differentiate between OLTP and OLAP
2. Explain the Three Tier Data warehouse Architecture with neat diagram.
3. Discuss different OLAP operations with an example
4. Describe about materialization Methods in data warehouse implementation
5. Explain Multidimensional Data model with Example
6. Compare the following
7. i)star schema
8. ii) snow-flake
9. iii)fact-constellation

**UNIT-IV**

1. Explain how to build a decision tree
2. Write about general approach to solving a classification problem
3. Describe the different measure for selecting the best split in decision tree induction
4. Define Hunts algorithm. How is it helpful to construct decision tree.
5. Discuss the Characteristics of Decision Tree Classifiers
6. Discuss the Reasons for Model Over fitting
7. Discuss the methods for Evaluating the Performance of a classifier

**UNIT-V**

1. Explain FP-growth algorithm for the generation of frequent itemsets
2. Write the Apriori Algorithm for Rule Generation
3. CompareApriori and FP-Growth Algorithm
4. Write the steps to construct an FP-Tree with an Example
5. Find The frequent Item sets and Association rules from the following Transactional data set

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| TID | List of Items | TID | Items | TID | List of Items |
| T1 | I1,I2,I5 | T4 | I1,I2,I4 | T7 | I1,I3 |
| T2 | I2,I4 | T5 | I1,I3 | T8 | I1,I2,I3,I5 |
| T3 | I2,I3 | T6 | I2,I3 | T9 | I1,I2,I3 |