M.Sc. (Computer Science)
Semester-3
MSCS-3-02T: Data Mining and Visualization

Total Marks: 100
External Marks: 70
Internal Marks: 30
Credits: 4
Pass Percentage: 40%

INSTRUCTIONS FOR THE PAPER SETTER/EXAMINER
1. The syllabus prescribed should be strictly adhered to.
2. The question paper will consist of three sections: A, B, and C. Sections A and B will have four questions from the respective sections of the syllabus and will carry 10 marks each. The candidates will attempt two questions from each section.
3. Section C will have fifteen short answer questions covering the entire syllabus. Each question will carry 3 marks. Candidates will attempt any ten questions from this section.
4. The examiner shall give a clear instruction to the candidates to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.
5. The duration of each paper will be three hours.

INSTRUCTIONS FOR THE CANDIDATES
Candidates are required to attempt any two questions each from the sections A and B of the question paper and any ten short q questions from Section C. They have to attempt questions only at one place and only once. Second or subsequent attempts, unless the earlier ones have been crossed out, shall not be evaluated.

SECTION-A

Unit 1: Data Mining: Introduction, Scope of Data Mining; How does Data Mining Works, Predictive Modeling: Data Mining and Data Warehousing: Architecture for Data Mining: Profitable Applications: Data Mining Tools.

Unit II: Data Pre-processing: Overview, Data Cleaning, Data Integration and Transformation, Data Reduction, Discretization and Concept Hierarchy Generation.

Unit III: Data Mining Techniques: An Overview, Data Mining Versus Database Management System, Data Mining Techniques- Association rules, Classification, Regression, Clustering, Neural networks.

Unit IV: Clustering: Introduction, Cluster Analysis, Clustering Methods- K means, Hierarchical clustering, Agglomerative clustering, Divisive clustering, evaluating clusters.

SECTION-B

Unit V: Applications of Data Mining: Introduction, Business Applications Using Data Mining-Risk management and targeted marketing, Customer profiles and feature construction, Medical applications (diabetic screening), Scientific Applications using Data Mining, Other Applications.

Unit VI: Data Visualization: Introduction, Acquiring and Visualizing Data, Simultaneous acquisition and visualization, Applications of Data Visualization, Keys factors of Data Visualization (Control of Presentation, Faster and Better Java Script processing, Rise of HTML 5, Lowering the implementation Bar).

Unit VII: Exploring the Visual Data Spectrum: Charting Primitives (Data Points, Line Charts, Bar Charts, Pie Charts, Area Charts), Exploring advanced Visualizations (Candlestick Charts, Bubble Charts, Surface Charts, Map Charts, Infographics).

Unit VIII: Visualizing data Programmatically: starting with Google charts (Google Charts API Basics, A Basic bar chart, A basic Pie chart, Working with Chart Animations)

Reference Books:
- Jiawei Han, Micheline Kamber and Jian Pei, "Data Mining Concepts and Techniques", 3rd Edition, 2000.

- Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, "Introduction to Data Mining", Pearson 2005.

- M. Kantardzic, "Data Mining: Concepts, Models, Methods, and Algorithms", 2nd edition, Wiley-IEEE Press, 2011.

- Jon Raasch, Graham Murray, Vadim Ogievetsky, Joseph Lowery, "Java Script and j Query for Data Analysis and Visualization", 2014.

- Ben Fry, "Visualizing data: Exploring and explaining data with the processing environment", O'Reilly, 2007.