



DATA MINING (PSIT101) SYLLABUS

Unit I	Introduction: Basics of data mining, related concepts, Data mining techniques	Data Mining: Introductory and Advanced Topics -- M H Dunham	Chapter 1,2,3 –
	Data: Introduction, Attributes, Data Sets, and Data Storage, Issues Concerning the Amount and Quality of Data, Knowledge Representation: Data Representation and their Categories: General Insights, Categories of Knowledge Representation, Granularity of Data and Knowledge Representation Schemes, Sets and Interval Analysis, Fuzzy Sets as Human-Centric Information Granules, Shadowed Sets, Rough Sets, Characterization of Knowledge Representation Schemes, Levels of Granularity and Perception Perspectives, The Concept of Granularity in Rules.	Data Mining- A Knowledge Discovery Approach - Krzysztof J. Cios	Chapter 3, 5
Unit II	Data Preprocessing: Descriptive Data Summarization, Data Cleaning, Data Integration and Transformation, Data Reduction, Data Discretization and Concept Hierarchy Generation. Mining Frequent Patterns, Associations, and Correlations: Basic Concepts, Efficient and Scalable Frequent Itemset Mining Methods, Mining Various Kinds of Association Rules, From Association Mining to Correlation Analysis, Constraint-Based Association Mining	J. Han and M. Kamber Data Mining: Concepts and Techniques	Chapter 2,5
Unit III	Classification and Prediction: What Is Classification?, What Is Prediction?, Issues Regarding Classification and Prediction, Classification by Decision Tree Induction, Bayesian Classification, Rule-Based Classification, Classification by Back-propagation, Support Vector Machines, Associative Classification: Classification by Association Rule Analysis, Lazy Learners, Other Classification Methods, Prediction, Accuracy and Error Measures, Evaluating the Accuracy of a Classifier or Predictor, Ensemble Methods Increasing the Accuracy, Model Selection.	J. Han and M. Kamber, Data Mining: Concepts and Techniques”	Chapter 6
Unit IV	Cluster Analysis: What Is Cluster Analysis?, Types of Data in Cluster Analysis, A Categorization of Major Clustering Methods, Partitioning Methods, Hierarchical Methods, Density-Based Methods, Grid-Based Methods, Model-Based Clustering Methods, Clustering High-Dimensional Data, Constraint-Based Cluster Analysis, Outlier Analysis	J. Han and M. Kamber, ” Data Mining: Concepts and Techniques”	Chapter 7
Unit V	Graph Mining, Social Network Analysis, and Multirelational Data Mining: Graph Mining, Social Network Analysis, Multirelational Data Mining. Mining Object, Spatial, Multimedia, Text, and Web Data: Multidimensional Analysis and Descriptive Mining of Complex Data Objects, Spatial Data Mining, Multimedia Data Mining, Text Mining, Mining the World Wide Web.	J. Han and M. Kamber, ” Data Mining: Concepts and Techniques”	Chapter 9, 10

DATA MINING PRACTICAL LIST

1. Prepare the Analysis services for Adventure Works Cycles or (any other database) . Build the data mining model structure and built the decision tree with proper decision nodes. And infer atleast five different types of reports.
2. Prepare the Analysis services for Adventure Works Cycles or (any other database) . Build the data mining model structure and Implement Naïve Bayes Algorithm.
3. Prepare the Analysis services for Adventure Works Cycles or (any other database) .Build the data mining model structure. Implement the clustering Algorithm.
4. Prepare the Analysis services for Adventure Works Cycles or (any other database) .Build the basic Time series model structure and create the predictions.
5. Prepare the Analysis services for Adventure Works Cycles or (any other database) . Build the data mining model and implement k-nearest neighbour.
6. Prepare the Analysis services for Adventure Works Cycles or (any other database) . Build the data mining model and implement Apriori algorithm.
7. Prepare the Analysis services for Adventure Works Cycles or (any other database) .Build the basic data mining model and show the implementation of Association algorithm. And also apply the DMX queries.
8. Consider the suitable data for text mining and Implement the Text Mining technique using R-Tool.
9. Design the spatial data model and apply the data mining techniques for extracting the information from spatial data base using R-Tool.
10. Using R-Tool , show the analysis for social networking sites.

Software List:

Sql Server 2008/2012 [www.microsoft.com]

Weka 3.4/ 3.6 [

R-3 [rdataming.com]