IT-AD05 ADD ON DIPLOMA COURSE IN DATA MINING

Objective: To provide in depth professional knowledge in the area of Data mining by introducing students to different data mining methods. This course is mainly designed as an add-on course for post graduate students

Eligibility: Graduate in Science subjects with 50% marks

Reservation: SC/ST/OBC as per university rules.

Duration: One Year part time, 80 hrs of teaching

Fee: Regular students from Constituent colleges: Rs 2500/- for Students of courses where add-on courses are compulsory, Rs 3000/- for others. Students from other institutions: Rs 3500/-

Seats: Thirty. The course will be offered only against admission of a minimum of 15 candidates

Examination: Examination will be conducted by a board consisting of an internal examiner and an external examiner on the basis of a MCQ on-line /off-line test of 1 hr duration (50 questions, 100 marks) and practical test of 3hrs (100 Marks). Total marks of the examination will be 200. Rs 200/- per candidate will be collected by the computer center towards the MCQ based OMR/Online test in addition to the university examination fee.

In order to eligible for the diploma, candidate is required to score 50% of the total marks. Those who fail to get 50% marks will be required to reappear in the examination as and when conducted by the university. A

Certificate will be awarded instead of Diploma to the candidates who are allowed to take this course independently, not as an Add-on course

SYLLABUS

DATA MINING

Introduction to data mining: basic data mining tasks, Data mining versus knowledge discovery in database, data mining issues and metrics, practical applications of data mining.

Basic concepts: Database/OLTP systems, Fuzzy sets and Fuzzy logic, information retrivel,Decision support systems, Dimensional modeling, Data warehousing, OLAP,Web search engines, Statistics, Machine learning, pattern matching

Data mining techniques: Statistical perspective on data mining, similarity measurements, decision trees, Neural networks, Genetic algorithms

Classification: Issues in classification, Statistical based algorithms, distance based algorithms, decision tree based algorithms, Neural network based algorithms, rule based algorithms, combining techniques

Clustering: Similarity and distance measures, outliers, **hierarchical algorithms**: *Agglomerative and divisive algorithms*, **partitional algorithms**: *Minimum spanning tree, Squared error clustering,K-Means clustering,Nearest neighbour,PAM,Bond energy, clustering with genetic, clustering with neural networks*.
Clustering large databases: *BRCH,DBSCAN,CURE*, clustering with categorical attributes, comparison.

Association Rules:

Large itemsets, basic algorithms : *Apriori algorithms, sampling algorithm,partitioning,* parallel and distributed algorithms*: Data parallelism and Task parallelism*, comparing approaches,incremental rules, Advanced association rule techniques: *Generalised Association rules,Multiple level,Quantitative association rules, Using Multiple minimum supports,Correlation rules,* Measuring the quality of rules.

Web Mining: Web content mining: *Crawlers, Harvest system, Virtual Web view,personalization,*Web structure mining*: PageRank,Clever*, Web usage mining: *Preprocessing, data structures, pattern discovery ,pattern analysis.*

Spatial Mining: Spatial data Overview: *Spatial Queries, Spatial Data Structures, Thematic maps, Image databases.* Spatial data mining primitives, Generalization and Specialization:*Progressive refinement,Generalisation,Nearest Neighbour,STING*

Spatial rules, Spatial classification algorithm:*ID3 extension and Spatial Decision tree.*

Spatial clustering Algorithms:

CLARANS,SD(CLARANS),DBCLASD,BANG,Wave cluster.

Temporal Mining: Modelling Temporal Events, Time series: *Time series analysis,Trend Analysis,Transformation,Similiarity,Prediction*, Pattern Detection, introductory concepts of Sequences and , Temporal association rules.

Introduction to data mining softwares

Reference Book:

Data Mining : Introductory and Advanced Topics, Margaret H Dunham, Pearson Education 2003