

PCS5H002 DATA MINING AND DATA WAREHOUSING (4-0-0)

Module 1

The Compelling Need for data warehousing: Escalating Need for strategic information, failures of Past decision-support systems, operational versus decision-support systems, data warehousing – the only viable solution, data warehouse defined

Data warehouse – The building Blocks: Defining Features, data warehouses and data marts, overview of the components, metadata in the data warehouse

Defining the business requirements: Dimensional analysis, information packages – a new concept, requirements gathering methods, requirements definition: scope and content.

Module 2

OLAP in the Data Warehouse:

Demand for Online analytical processing, need for multidimensional analysis, fast access and powerful calculations, limitations of other analysis methods, OLAP is the answer, OLAP definitions and rules, OLAP characteristics, major features and functions, general features, dimensional analysis, what are hypercube? Drill-down and roll-up, slice-and-dice or rotation, OLAP models, overview of variations, the MOLAP model, the ROLAP model, ROLAP versus MOLAP, OLAP implementation considerations

Module 3

Data Mining Basics: What is Data Mining, Data Mining Defined, The knowledge discovery process, OLAP versus data mining, data mining and the data warehouse, Major Data Mining Techniques, Cluster detection, decision trees, memory-based reasoning, link analysis, neural networks, genetic algorithms, moving into data mining, Data Mining Applications, Benefits of data mining, applications in retail industry, applications in telecommunications industry, applications in banking and finance

Module 4

Applications of Data mining-Social Impacts of Data mining-Tools- Mining the World Wide Web- Spatial Data Mining – Multimedia Data Mining – Text Mining. Programming assignments are mandatory.

Text Book:

1. Jiawei Han, Micheline Kamber, and Jian Pei, "Data Mining Concepts and Techniques", Third Edition, Elsevier.
2. Vikram Pudi & P. Radha Krishna, Data Mining, Oxford University Press.
3. Reema Thareja, Data Warehousing, Oxford University Press.

Reference Books:

1. Alex Berson and Stephen J. Smith "Data Warehousing, Data Mining & OLAP", Tata McGraw – Hill Edition, Tenth Reprint 2007
2. K.P. Soman, ShyamDiwakar and V. Ajay "Insight into Data mining Theory and Practice", Easter Economy Edition, Prentice Hall of India, 2006
3. G. K. Gupta "Introduction to Data Mining with Case Studies", Easter Economy Edition, Prentice Hall of India, 2006
4. Pang-Ning Tan, Michael Steinbach and Vipin Kumar "Introduction to Data Mining", Pearson Education, 2007.