

IT WORKSHOP

ON

DATAMINING

Data Warehousing & Data Mining

Collections of databases that work together are called data warehouses. This makes it possible to integrate data from multiple databases. Data mining is used to help individuals and organizations make better decisions.

Data Warehousing: In a typical setting, the database files reside on the server, but they can be accessed from many different computers in the organization. As the number and complexity of databases grows, we start referring to them together as a data warehouse. A data warehouse is a collection of databases that work together. A data warehouse makes it possible to integrate data from multiple databases, which can give new insights into the data.

Data Mining: Once all the data is stored and organized in databases, what's next? Many day-to-day operations are supported by databases. Queries based on SQL, a database programming language, are used to answer basic questions about data. But, as the collection of data grows in a database, the amount of data can easily become overwhelming. How does an Organization get the most out of its data in the details? That's where Data Mining comes in. Data Mining is a Process of analyzing data and summarizing it to produce useful information.

Topics to be covered in Workshop

Introduction to Data Mining

- What is data mining?
- Related technologies - Machine Learning, DBMS, OLAP, Statistics
- Data Mining Goals
- Stages of the Data Mining Process
- Data Mining Techniques
- Knowledge Representation Methods
- Applications
- Example: weather data

Data Warehouse and OLAP

- Data Warehouse and DBMS
- Multidimensional data model
- OLAP operations
- Example: loan data set

Data preprocessing

- Data cleaning
- Data transformation
- Data reduction
- Discretization and generating concept hierarchies
- Installing Weka 3 Data Mining System
- Experiments with Weka - filters, discretization

Data mining knowledge representation

- Task relevant data
- Background knowledge
- Interestingness measures
- Representing input data and output knowledge
- Visualization techniques
- Experiments with Weka - visualization

Attribute-oriented analysis

- Attribute generalization
- Attribute relevance
- Class comparison
- Statistical measures
- Experiments with Weka - using filters and statistics

Data mining algorithms: Association rules

- Motivation and terminology
- Example: mining weather data
- Basic idea: item sets
- Generating item sets and rules efficiently
- Correlation analysis
- Experiments with Weka - mining association rules

Data mining algorithms: Classification

- Basic learning/mining tasks
- Inferring rudimentary rules: 1R algorithm
- Decision trees
- Covering rules
- Experiments with Weka - decision trees, rules

Data mining algorithms: Prediction

- The prediction tasks
- Statistical (Bayesian) classification
- Bayesian networks
- Instance-based methods (nearest neighbor)
- Linear models
- Experiments with Weka - Prediction

Evaluating what's been learned

- Basic issues
- Training and testing
- Estimating classifier accuracy (holdout, cross-validation, leave-one-out)
- Combining multiple models (bagging, boosting, stacking)
- Minimum Description Length Principle (MLD)
- Experiments with Weka - training and testing

Advanced techniques, Data Mining software and applications

- Text mining: extracting attributes (keywords), structural approaches (parsing, soft parsing).
- Bayesian approach to classifying text

- Web mining: classifying web pages, extracting knowledge from the web
- Data Mining software and applications

Duration: The duration of this workshop will be two consecutive days, with 6-7-hour session each day in a total of 12-14 hours.

Certification Policy:

- Certificate of Merit for all the workshop participants from Manac Infotech Pvt Ltd.
- At the end of this workshop, a small competition will be organized among the participating students and winners will be awarded with a '**Certificate of Excellence**'.
- Certificate of Coordination for the coordinators of the campus workshops.

Eligibility: It's a basic level workshop so there are no prerequisites. Anyone interested, can join this workshop.