



SNS COLLEGE OF TECHNOLOGY

(An Autonomous Institution)

COIMBATORE-35



Department of Information Technology

Unit – 1

Data mining and warehousing

DATA MINING APPLICATIONS

- Business transactions
 - Eg: Customer segmentat
- Web search engines
 - Eg: SEO analysis
- Financial data analysis
 - Eg: Stock market prediction
- Telecommunication industry
 - Eg: Predicting whether customers will buy additional products like cellular services, call waiting or basic services
- Retail industry
 - Eg: Ads to reach certain segment of customers
- Healthcare and biomedical research
 - Disease prediction based on symptoms
- Science and engineering
 - Eg: Cosmology, molecular genetics, protein sequencing and macro – molecular structure



Department of Information Technology

Unit – 1

Data mining and warehousing

DATA MINING AS KNOWLEDGE DISCOVERY FROM DATA (KDD)

The steps involved are

1. Data cleaning – to remove noise and inconsistent data
2. Data integration – where multiple data sources may be combined
3. Data selection – retrieving relevant data
4. Data transformation – transforming data into a form appropriate for mining
5. Data mining – applying intelligent methods to extract data pattern
6. Pattern evaluation – to identify truly interesting pattern
7. Knowledge presentation – using visualization and knowledge representation techniques to present mined knowledge to users

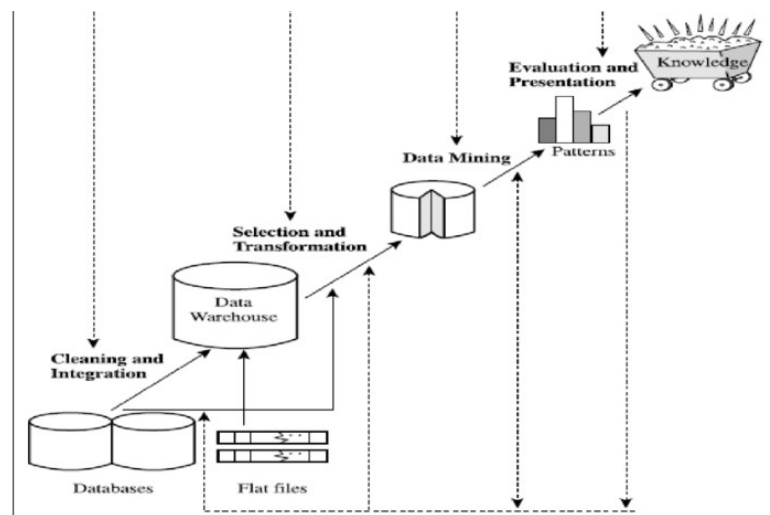


Figure: Data mining as a step in knowledge discovery of data

Real world examples of data mining

1. Retail sector
 - a. Helps to know choice of customers by looking at purchase history – buying preferences can be predicted
 - b. Data mining can be used for product recommendations

Types of data that can be mined

- Relational database
- Data warehouse
- Transactional database



SNS COLLEGE OF TECHNOLOGY

(An Autonomous Institution)

COIMBATORE-35



Department of Information Technology

Unit – 1

Data mining and warehousing

DATA MINING ISSUES

- Human interaction
- Over fitting of data
- Outlier
- Large datasets
- Missing / irrelevant data
- Mining methodology
- Diversified data types