o Data Mining (knowledge discovery from data) – Extraction of interesting (non–trivial, implicit, previously unknown and potentially useful) patterns or knowledge from huge amount of data.Advanced algorithms, multiprocessor computers, massive databases Pilot, Lockheed, IBM, SGI, numerous startups (nascent industry) Prospective, proactive information delivery DATA MINING: THEORY AND PRACTICE 11 Potential Applications o Data analysis and decision support – Market analysis and management o Target marketing, customer relationship management (CRM), market basket analysis, cross selling, market segmentation – Risk analysis and management o Forecasting, customer retention, improved underwriting, quality control, competitive analysis – Fraud detection and detection money laundering.On–line analytic processing (OLAP), multidimensional databases, data warehouses Pilot, Comshare, Arbor, Cognos, Microstrategy Retrospective, dynamic data delivery at multiple levels Data Mining (Emerging Today) "What's likely to happen to Boston unit sales next month? Why?"Relational databases (RDBMS), Structured Query Language (SQL), ODBC Oracle, Sybase, Informix, IBM, Microsoft Retrospective, dynamic data delivery at record level Data Warehousing & Decision Support (1990s) "What were unit sales in New England last March? Drill down to Boston."– Model selection criteria 9 Sanad AL–Maskari–Data Mining Database Vs Data Mining o Query – Well defined – SQL o Data – Operational data o Output – Precise – Subset of database ?o Origins of Data Mining – Statistics: more theory–based, focused on testing hypotheses – Database system: information retrieval, deductive query answering – Data Mining: focuses on the algorithms to extract patterns from data 3 Why Data Mining?8 8 Sanad AL–Maskari–Data Mining Data Mining Algorithm o Purpose is to fit data to a model – Descriptive (Clustering, association, summarization, etc ) – Predictive (Classification, regression , etc.) o Model selection – Which model is the best ?– Knowledge discovery through data mining (rules, regularities, patterns, constraints) 7 7 Sanad AL–Maskari–Data Mining Potential Applications o Data analysis and decision support – Market analysis and management o Target marketing, customer relationship management (CRM), market basket analysis, cross selling, market segmentation.o Other Applications – Text mining (news group, email, documents) and Web mining – Medical applications (Cancer detection) – DNA and bio–data analysis 12 Sanad AL–Maskari–Data Mining Market Basket Analysis Example Bread, Coke, Milk Juice, Coke, Diapers, Milk Juice, Bread, Milk Juice, Bread, Diapers, Milk Anything interesting?Diapers, Milk Bread Milk (100%) Diapers Juice(66%) Diapers Milk (100%) Customers who buy diapers also tend to buy Juice Indentify potential cross–selling opportunities among related items 13 Sanad AL–Maskari–Data Mining Bank Loan Risk Example salary loan k IF salary

Summarized by © lakhasly.com