TRINITY COLLEGE FOR WOMEN NAMAKKAL

Department of Computer Science

DATA MINING

EVEN Semester

Presented by

Mrs.A.KAVITHA

Assistant Professor

Department of Computer Science

http://www.trinitycollegenkl.edu.in/

What is Data Mining

- Data mining is the process of understanding data through cleaning raw data, finding patterns, creating models, and testing those models. It includes statistics, machine learning, and database systems.

Benefits of data mining

- Data mining is most effective when deployed strategically to serve a business goal, answer business or research questions, or be a part of a solution to a problem.

- Data mining assists with making accurate predictions, recognizing patterns and outliers, and often informs forecasting. Further, data mining helps organizations identify gaps and errors in processes, like bottlenecks in supply chains or improper data entry.

How data mining works

- The first step in data mining is almost always data collection. Today's organizations can collect records, logs, website visitors' data, application data, sales data, and more every day.

- Collecting and mapping data is a good first step in understanding the limits of what can be done with and asked of the data in question.

- The Cross-Industry Standard Process for Data Mining (CRISP-DM) is an excellent guideline for starting the data mining process. This standard was created decades ago and is still a popular paradigm for organizations that are just starting.

Types of data mining techniques

- Data mining includes multiple techniques for answering the business question or helping solve a problem.

1. Classification
2. Clustering

Classification

- The most common technique is classification. To do this, identify a target variable and then divide that variable into appropriate level of detail categories. For example, the variable 'occupation level' might be split into 'entry-level', 'associate', and 'senior'.

- With other fields such as age and education level, you can train your data model to predict what occupation level a person is more likely to have.

- Insurance or financial institutions such as PEMCO Insurance used classification to train their Algorithms to flag fraud and to monitor claims.

Clustering

- Clustering is another common technique, grouping records, observations, or cases by similarity. There won't be a target variable like in classification. Instead, clustering just means separating the data set into subgroups.

- This method can include grouping records of users by geographic area or age group. Typically, clustering the data into subgroups is preparation for analysis. The subgroups become inputs for a different technique.

How to avoid data mining mistakes

- Data mining is a powerful and useful process for exploring data to predict patterns or outcomes. Unfortunately, it's easy to do data mining incorrectly. You shouldn't use data mining if your leaders do not have analytical or statistical knowledge to oversee the software.

- Inaccurate mining techniques can create incorrect models, resulting in inaccuracies. Further, if the team is using personally identifiable information in data mining activities, they must ensure they are following compliance regulations and governance standards.

Who does data mining in an organization?

- Data mining specialization is most often a function or capability of data scientist or data analyst roles. Data mining tends to require large projects with far-reaching, cross-functional project management, and it can ladder up to analytics or business analysis teams.

- Some organizations look to data mining specialists to build machine learning or artificial intelligence scripts, so proficiency and knowledge of these is often a core competency.

THANK YOU

http://www.trinitycollegenkl.edu.in/