Patrician College of Arts and Science

Department of Computer Science

DATA MINING

Subject Code SEE6A

Even Semester

Presented By JOHN VASEEKARAN S

https://www.patriciancollege.ac.in/

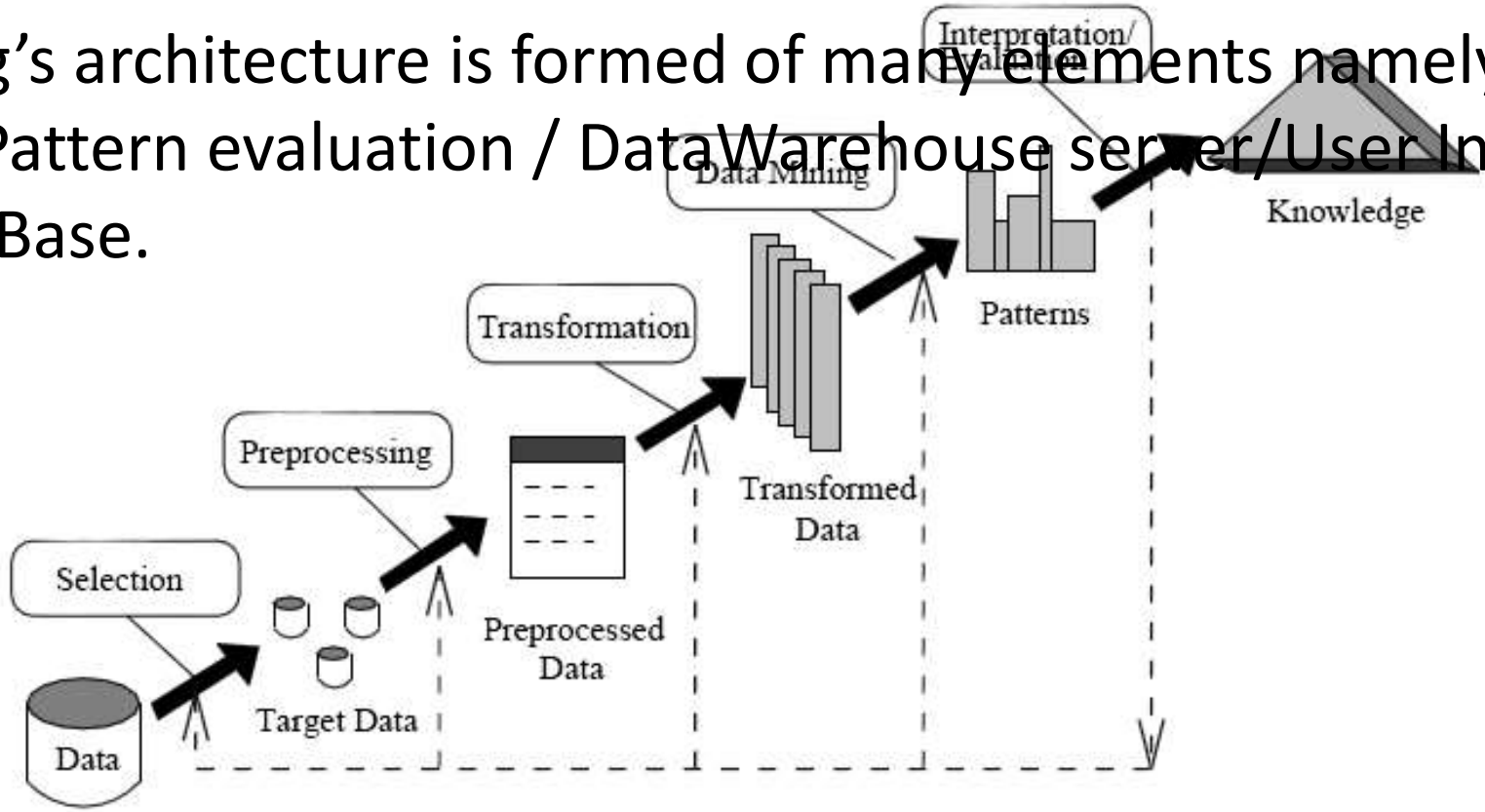
INTRODUCTION TO DATA MINING

Data Mining is the process of locating potentially practical, interesting and previously unknown patterns from abig volume of data. It plays an important role in result orientation.

Data mining can be used in each and every aspect of life. The same is similarly significant in other areas including sales/marketing, revenue services, sports, health care and insurance etc. The said paper implies general idea of data mining system, functionalities and its applications.

ARCHITECTURE OF DATA MINING

Data Mining's architecture is formed of many elements namely Data Mining Engineer / Pattern evaluation / DataWarehouse server/User Interface and Knowledge Base.



PARTS OF THE DATA MINING

Data Mining's architecture is formed of many elements namely,

1. Data Mining Engineer
2. Pattern evaluation
3. Data Warehouse server
4. User Interface and Knowledge Base.

EXPLANATION OF THE PARTS OF DATA MINING

- *Knowledge Base:*

  Centralized storage of Knowledge Base are used to collect the information and to evaluate the pattern.

- *Data Mining Engine:*

  An essential element of data mining system and consists of functional elements that perform various tasks namely clustering, classification, prediction, association and correlation analysis, characterization.

- *Pattern Evaluation Module:*

  The element performs interesting measures and communicates with the data mining engine module to find out interesting pattern.

- *User Interface:*

  User interface module interacts between user and data exploring system. It allows the subscriber to do interaction with the system by explaining his query and simultaneously by identifying information in order to help in search and to carry out exploratory data mining based on the intermediate data mining results.

FUNCTIONALITIES OF DATA MINING

The functionalities are measured to perceive the type of patterns to be found in data mining tasks. Data Mining tasks can be categorized in to two categories namely,

- *Descriptive Task:*

  These tasks present the general properties of data stored in database. The descriptive tasks are used to find out patterns in data i.e. cluster, correlation, trends and anomalies etc.

- *Predictive Tasks:*

  Predictive data mining tasks predict the value of one attribute on the bases of values of other attributes, which is known as target or dependent variable and the attributes used for making the prediction are known as independent variables.

APPLICATIONS OF DATA MINING

- *Data mining applications in sales/ marketing*
- *Data mining applications in banking / finance*
- *Data mining applications in Health Care and Insurance*
- *Data Mining for the Retail Industry*
- *Data mining for the Telecommunications industry*
- *Data Mining Application in Higher Education*
- *Data mining for instruction Detection*

CHALLENGES IN DATA MINING

In current situation of affairs data mining research is “too”ad-hoc” and their are so many challenges to unify different data mining tasks. Some of the challenges in the area are as under,

- *Scalability*
- *High Dimensional Data and High Data Streams*
- *Complex and Heterogeneous Data*
- *Data Ownership, Security and Privacy*
- *Data Distribution*

CHALLENGES IN DATA MINING

- *Scalability:*

  One important challenge is mining data from huge data bases. Computer data network and satellite data can easily be of this scale but to-days technology in data mining are too slow to handle data of this scale.

- *High Dimensional Data and High Data Streams:*

  One challenge is to design classifiers to control ultra high dimensional classification problems for mining vast, enormous and high dimensional data set out-of-memory, parallel and distributed algorithms, algorithm is need to be developed.

CHALLENGES IN DATA MINING

- *Complex and Heterogeneous Data:*

  Another challenge erupted in these years is emergence of more data complex. A good system must scale the complexity from users. Previous analysis data mining method deal with the data set consisting attribute of similar type.

- *Data Ownership, Security and Privacy:*

  It is a big challenge to find out data for an analysis at one location or to be owned by one location or to be owned by one entity. An automatic data mining in distributed environment can develop serious issues in terms of data privacy or its security.

CHALLENGES IN DATA MINING

- *Data Distribution:*

  This challenge in data mining is very important in network problems. This can be addressed by the development of distributed data mining techniques. The key challenges in distributed data mining are:

  a) To minimize the amount of communication needed to perform the distributed computation.

  b) To consolidate the data mining results obtains from multiple sources in a efficient manner.

  c) To tackle data security issues.

Thank you

https://www.patriciancollege.ac.in/