Date
Page 8

Assignment - 2

1702801160 18

1) What is data mining? Explain different types of data on which mining can be performed.

- Data mining is a process of discovering pattern in large data sets

i) Flat files: Data files in text form or binary form with a structure

2) Relational database: Collection of data organized in tables with rows and column

3) Data warehouse: Collection of data integrated from multiple sources that will queries & decision making.

2) what kind of pattern can be mined

- Data mining functionality are used to specify the kind of patterns to be found in data mining

- concept discrimination
- classification & regression
- clustering analysis
- outlier analysis

170280116018

3) Explain integration of data mining system to database or data ware house.

- 1) No coupling : DM system will not utilize any function of DB or DW system.

2) loose coupling : DM system will use some functionality of DB or DW system.

3) semi tight coupling : Besides linking a DM system to DB/DW system efficient implimentation of few essential data mining primitives can be provided in the DB/DW system.

4) tight coupling : DM system is smoothly integrated into the DB/DW system.

4) Explain major issues in DM.

- 1) Mining methodology & user interaction.
  - mining different kind of knowledge in databases
  - Interactive mining of knowledge at multiple levels of abstraction.
  - Incorporation of background knowledge.

2) Performance issues:
  - efficiency and scalability of data mining eg algorithm.
  - parallel, distributed & incremental mining algo.

170280116018

5) Explain data mining as step of knowledge discovered process.

- 1) Data cleaning: cleaning in case of missing value.
  - cleaning noisy data, where noise random.

2) Data Integration: Using Data Mining tools.
  - using ETL process.

3) Data selection: where data relevant to the decided and retrival from data collection

4) Data transformation: process of terms forming data into appropriate form required by mining procedure.