Digital Assignment - 3

Data mining Research Challenges

☆) Data mining tools handles with large amount of data where two scalability issues are needed.

 a) Most data mining algorithms cannot process the amount of data gathered at websites in reasonable time because they scale nonlinearly.

 b) And generated models are so much complicated for humans to understand.

Major Research Challenges

a) Design classifiers to handle ultra-high -dimensional classification problem.

b) Mining data streams in extremely large data base.

c) Mining complex knowledge from complex data.

d) Mining across multiple heterogeneous data Source; Multi database & multi relational mining:

E) Mining Non-Relational data.

F) Automate Data Cleaning

G) Privacy Preseving data mining.

Illustrate one data mining research issue that may have a strong impact on Market & on society.

d) Mining across multiple heterogeneous data sources: multi database & multi relational mining:

A) The Problem of distributed data mining is very important in network Problems.

A) In distributed environment (such as a sensor or IP network), One has distributed probes placed at strategic locations within the network.

*) The Problem here is to be able to correlate the data seen at the various probes, & discover pattern in the global data seen at all the different probes.

*) There could be different model of distributed data mining here, but One could involve a NOC that collects data from the distributed sites & another in which all sites are treated equally.

*) The goal here obviously would be to minimize the amount of data Shipped between the various sites essentially, to reduce to communication over head.

*) In the distributed mining, one problem is how to mine across multiple heterogeneous data source: multi-database & multi-relational mining[5].