Data Mining with Oracle using either Clustering or Classification Algorithms

Abstract

Data mining also known as knowledge discovery, involves finding unexpected but interesting patterns within enormous amounts of data that are normally stored in databases and data warehouses. Data Mining has three major components Clustering or Classification, Association Rules and Sequence Analysis which come in form of many algorithms proposed. Some of the algorithms have had better success than the others.

However, the commercial world is fast reacting to the growth and potential in this area as a wide range of tools are marketed under the label of data mining. The main objective of this project is to investigate two types of algorithms available in Oracle for data mining. Apply the two algorithms to actual data. Then, analyses the results and compare outcome in terms of accuracy, efficiency and effectiveness.

Literature Survey

The resources listed below will basically serve as a starting point for my research project. As the project progresses, more detailed resources will be required. These may help in the understanding of Oracle data mining algorithms.

Michael J.A. Berry and Gordon S. Linoff. *Mastering Data Mining.* The Art and Science of Customer Relationship Management, Wiley Computer Publishing, 2000. This will allow for the understanding and of uses of data mining. It also

discusses different methodologies used in data mining, thus it will help with evaluations of results.

Jiawei Han and Micheline Kamber, *Data mining:* concepts and techniques. San Francisco, California, Morgan Kauffmann, 2001. This book gives a wide range of information on the different algorithms used in data mining. It also discusses different uses for the algorithms. Therefore, it will be useful for analyzing performance of the data mining algorithms.

Timeline for Implementation

This is just an initial timeline that will help me focus on major events in the course of the project. The project will however be researched and developed using an iterative approach, where analysis, design and implementation will continuously be taking places.