ABSTRACT

Distributed Data Mining (DDM) is a branch of the field of data mining that offers a framework to mine distributed data paying careful attention to the distributed data and computing resources. This process is carried out with parallel data mining algorithms. Association rules are used in this research for the mining process. The distributed workstations are identified and the CPU utilization is retrieved from the identified workstations. The CPU idle time retrieval is carried out and the job is allocated to distributed workstations. The job allocation is performed based on the CPU idle time and a task allocation algorithm the simulated annealing. After the task allocation, the mining process is carried out with parallel mining algorithms. The system also integrates the mined results obtained from different geographical locations. The proposed research work carried out the above task by using the intelligent agents to monitor the performance during the mining process and also for integrating the knowledge. Dynamic remote memory utilization is used in this research so that the memory is utilized effectively without any overload.

This research work deals with developing a parallel and distributed data mining system which will manage the memory efficiently using dynamic remote memory utilization. The new system proposed gives an improved response time.
The major strengths of this research work are listed as follows:


b. Dynamic remote memory utilization for utilizing the memory effectively.

c. Knowledge Integration Process with relevant models.

d. Optimized Response Time.

e. XML dataset for performing mining.

f. Intelligent agents algorithms are developed to monitor the mining process.

The architecture used for this research work is the client server model along with the agents. This model combines the best aspects of the agent model and the client-server approach by building an agent framework with a data mining server. It brings with it the advantage of combining the concept of dedicated data mining resources. Hence distributed data mining in parallel environment provides a framework for scalability, which allows the splitting up of larger datasets with high dimensionality into smaller subsets that require less computational resources individually.