CHAPTER 5
RESULTS AND DISCUSSIONS

This chapter presents the results and discussions of the implemented methodology of the automated data mining system with the help of software agents in the Object Oriented databases. The student data set of the Karunya University, Campus Management System is used in greater details. The results of the student data set are analyzed and partitioned, ranked, and using the clustering algorithms of the automated data mining system the knowledge is mined. Results are presented for the partitioning Object Oriented Database and clustering. Finally, this chapter will conclude with a summary of the findings.

5.1 Working setup

The ultimate goal of the research work is to develop an automated data mining system for an Object Oriented data with the help of software agents. Any non-expert users will be able to achieve reasonable results with minimum effort. The developed system provides an interface to the non-experts and also hides the data mining concepts away from the users thus helping to bridge the conceptual gap usually associated with data mining.

The student data set of Karunya University Campus Management System is considered for the research work. Any non-expert user can place his high level goal or objectives or user's hint as a query. First, the user interface agent present in the automated data mining system analyzes the Object Oriented data set. It classifies them based on the following classification. It identifies the
relationships between the objects, classes, and attributes. It identifies the importance of each of the attributes. The user interface agent also identifies the type of the attributes. After the user interface agent analyzes and classify the object oriented database based on the user's high level goal or objectives or user's hint the action takes place. The user interface agent identifies the attribute for the action. The analyzed and classified information is communicated to the partitioning agent.

Partitioning agent gets the analyzed information from user interface agent. The Object Oriented data is partitioned vertically with the help of the partitioning processor based on the information given by the user interface agent. The partitioning agent selects the partitioning method that is suitable for the attribute. In addition, the partitioning agent will decide the number of partitions based on the relationship between the attributes, classes, objects, and methods. Once the object oriented data is partitioned the ranking agent is called by the partitioning agent.

Ranking is carried out based on the Query weight and scoring function. Ranking agent ranks all the attributes that is partitioned by the partitioning agent and prompts the highest ranked attribute based on the user hints or query to the data mining agent. The highest ranked attributes are taken for clustering.

Finally, the Data Mining agent lists the possible clustering algorithms for the ranked attribute. The Data Mining agent contains specific clustering algorithms implemented in house, that may have been installed within the
environment. Data Mining agent decides which clustering algorithms are appropriate for the ranked data and it selects the appropriate clustering algorithm. Thus, the Data Mining agent is responsible for performing the actual data mining activity and generating the results. Thus necessary knowledge is mined using the clustering algorithm with less user intervention. The results for partitioning, ranking and clustering are presented in the following section.

5.2 Results and Discussion

The following are the results concerning the effects of vertical partitioning of object oriented databases, ranking of attributes and clustering of object oriented database using software agents.

![Automated Data Mining System](image)

**Figure 5a: User Input Querying Screen**

The above screen is used for getting the query to be solved by the user in terms of sentence.
Figure 5b: User Interface Agent

Figure 5b shows a snapshot of the Automated Data Mining System for an User Interface agent. Typically, the left half of the agent interface panel contains ‘tabs’ that provide specific information related to that agent, while the right half shows the agent communication. The top box on the right hand side, labeled “Incoming messages”, displays the list of messages the agent has received so far, while the lower box labeled “message content” shows the content of a particular message.
Figure 5c: Data Mining Agent

Figure 5c shows a snapshot of the Automated Data Mining System for an User Interface agent. In this screen the partitioning agent messages and the ranking agents messages are seen as incoming messages. The data mining agent chooses the appropriate clustering algorithm suitable for the object oriented data.
Data Mining Agent Method Selection Panel

The parameters are ranked in the previous panel. In this panel please select the appropriate clustering method. If you want to do the mining process automatically please click the "Automatic" button. To use the "User" button. To end the session click on the "End" button. To get help related to a certain option, please click on the "Help" button.

<table>
<thead>
<tr>
<th>Knowledge Mining Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ Automatic</td>
</tr>
<tr>
<td>☐ User</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Clustering Algorithms For Numerical Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐ K-Means Algorithm</td>
</tr>
<tr>
<td>☐ K-Medoids Algorithm (PAM: Partitioning Around Medoids)</td>
</tr>
<tr>
<td>☐ Clustering Large Applications (CLARA)</td>
</tr>
</tbody>
</table>

Figure 5d: Data Mining Agent Method selection panel for numerical attribute

Figure 5d shows a snapshot of the Automated Data Mining System of a data mining method selection panel. In this, the agents will automatically choose the appropriate clustering algorithm based on the user input. Here the automated data mining system has chosen k-means algorithm for clustering numerical attributes. In case, the user want to use other algorithms it can be performed with minimum effort.
Figure 5e: Data Mining Agent Method selection panel for Object Oriented data

Figure 5e shows a snapshot of the Automated Data Mining System of a data mining method selection panel for object oriented data. In this, the agents will automatically choose the appropriate clustering algorithm based on the user preferences.
put. Here the automated data mining system has chosen k-means algorithm for clustering Object Oriented data. In case, the user wants to use other algorithm it can be performed by selection with minimum effort.

![Graph showing ranking values of various attributes](image)

**Figure 5f: Ranking of mixed attributes based on the score function and query weight**

The figure 5f shows ranking of various attributes in the student data set. Ranking factors used in Ranking of attributes. A Scoring function is used to score the attributes. Query weight is also calculated, based on number of times a particular attribute is accessed by the user. The importance and relationship between the attributes are considered while clustering. In the figure 5a the attribute pincode is of numeric data type. Its ranked high among the other attributes. Numeric attributes have high scoring function. In the student data set based on the query given by the user the user interface agent identifies the closely related attributes with the query. In this case the pincode and district are
the closely related attributes with the address of the student. So they are partitioned and the ranking agent ranks the attributes. Among the two attributes the pincode is of numeric type which has the high score and query weight. Also pincode is an important attribute for determining the city and its relationship is close to address so the ranking agent gives a high score to that attribute. From the above graph the numeric attributes will get a high score based on the user query.

**Figure: 5g: Time taken to cluster the attributes**

Figure 5g shows the time taken to cluster the number of attributes used in this research work. As the number of attributes increases in the data set, the time taken to cluster the attributes will also increase. If, the user hint or query has less attributes clustering takes less time. As the number of attributes
Increases, more time is taken for clustering for the given real time data. From this graph, the automated data mining system takes more time to cluster the attributes as the number of attributes increases.

![Response Time Function with No. of Objects for Cactis]

**Figure 5h: Number of Attributes VS Response Time using Cactis Algorithm**

Figure 5h shows the response time with number of attributes using Cactis Clustering Algorithm. The number of attributes in x-axis and response time in milliseconds in y-axis. In the automated data mining system while clustering using the Cactis algorithm as the number of attributes increases the response time also increases. Using Cactis algorithm the time taken to cluster is less when we cluster the objects. Cactis algorithm choose the "hotter" (i.e., more frequently accessed) most referenced object in the data.
Figure 5i: Response Time VS Number of Attributes using ORION Algorithm

Figure 5i shows the response time with a number of attributes using ORION Clustering Algorithm. The number of attributes in x-axis and response time in milliseconds in y-axis as shown in the above Figure 5d. The time taken to cluster the objects using ORION is increased. ORION uses the composite hierarchy method to cluster the objects.
Figure 5j: Clustering time taken with different attributes

Figure 5j shows the time taken to cluster the number of attributes. As shown in the graph, the clustering time is small for clustering a few attributes. And the time taken will be more to cluster a large number of attributes.
Figure 5k: Clustering with numerical attribute using k-means algorithm

Figure 5k shows the clustering of numerical attribute using k-means algorithm. Here students of different classes are taken in the x-axis and the control number are taken in the y-axis and using the k-means algorithm as n=4 the data are clustered.
Figure 51: Cluster representation for attendance from the student data set.

From the figure 51 shows the cluster representation for attendance from the student data set. It shows the students who are regular and irregular for the classes. Cluster 0 has more dropout students because they are irregular and who are having low attendance than any other clusters. Investigation of the data reveals that the parents of the students in this cluster are not educated and belong to low income category.
Figure 5m: Cluster representation for parent income from the student data set.

From the figure 5m shows the cluster representation for parent income from the student data set. It shows the students who belong to medium income of the parents are more and then students who belong to low income parents are also high. Cluster 0 seems to be an interesting cluster from a teacher’s perspective and it requires immediate action. The parents of these students are low income and because of these reasons there are more dropouts and needs attention of teachers.
Figure 5n: Cluster representation for parent education from the student data set.

Investigation of the above student data cluster 0 reveals that the parents of the students in this cluster are not educated and belong to the low-income category (refer figure 5m). Investigation of the above figure shows that cluster 2 shows that parents are educated though they belong to low-income category.
Figure 50: Clustering based on grade values

Figure 50 shows the students clustered based on their grade values. The following analysis is made from this figure. Grading is taken for a set of class strength 120. From the above figure getting the lowest grade and highest grade is less. And the average number of students with average grade is high. Getting the is not depending upon the teachers. Extra coaching should be given to poor students.
Figure 5p: Clustering based on the grade values between different departments

Figure 5p shows the students are clustered based on the grade values. The following analysis is made from this figure. Total number of student is 120. Getting the lowest grade and highest grade is less. Average number of student is high. Getting the grade is not depending upon the teachers. Extra coaching should be given to poor students. EEE, ECE, IT and CST students are getting high marks than other department students.
Figure 5q: Clustering with different papers

Figure 5q shows the clustering of different subjects in a department. From the figure there is a good number of students who has scored high percentage in language paper. And poor percentage in Theory paper because of vast syllabus.