Chapter 4
Secure Data Analysis in Clusters (Iris Database)

Raghvendra Kumar
LNCT College, India

Prasant Kumar Pattnaik
KIIT University, India

Priyanka Pandey
LNCT College, India

ABSTRACT

This chapter used privacy preservation techniques (Data Modification) to ensure Privacy. Privacy preservation is another important issue. A picture, where number of clients owning their clustered databases (Iris Database) wish to run a data mining algorithm on the union of their databases, without revealing any unnecessary information and requires the privacy of the privileged information. There are numbers of efficient protocols are required for privacy preserving in data mining. This chapter presented various privacy preserving protocols that are used for security in clustered databases. The Xln(X) protocol and the secure sum protocol are used in mutual computing, which can defend privacy efficiently. Its focuses on the data modification techniques, where it has been modified our distributed database and after that sanded that modified data set to the client admin for secure data communication with zero percentage of data leakage and also reduce the communication and computation complexity.

INTRODUCTION

In recent years, Agarwal et al. (1993) Agarwal, Imielinski, and Swamy (1993) and Srikant and Agarwal (1994) suggested data mining became a very interesting topic for the researcher due to its vast use in modern technology of computer science but due to its vast use it faces some serious challenges regarding data privacy and data privacy became an interesting topic. Many methods techniques and algorithms are already defined and presented for privacy preserving data mining. These privacy preserving techniques can be classified mainly in two approaches; the authors Agrawal and Srikant (2000) and Lindell and

DOI: 10.4018/978-1-5225-2031-3.ch004

Pinkas (2000) suggested Data modification and Secure Multi-party Computation approach. Data Mining suggested by Kantarcioglu and Clifto (2004) in last few decades has become very useful as the database are increasing day by day many people now connected with the computers by Han and Kamber (2006), so it becomes necessary for computer researchers to make the data so fast to access, also need to find right data. The term Data Mining emphasize on the fact of extracting the knowledge from large amount of data, so data mining is the process through which, we collect knowledgeable data from very large data suggested by Sheikh, Kumar, and Mishra (2010).

Now, the database is very large which consists so much information but what we want to find is the relevant data from large database or want to find some patterns which becomes very difficult with normal DBMS but with the use of data mining techniques we can find the hidden patterns and information from large database system. So, we can also term data mining as the knowledge mining, pattern extraction etc. But before applying data mining techniques we need to apply some processes which we known as pre-processing of data. Although data mining is one of the step involved in process of knowledge discovery, still it becomes more popular by name then that (Jangde, Chandel, & Mishra, 2011).

The data mining technique authored by Sugumar, Jayakumar, and Rengarajan (2012) can be used on Bio-Database (Iris database) for analyzing and acquiring different relations in the food condition of market or environmental conditions authored by Lakshmi and Rani (2012) and many more to find the relations which can tell the cause of any disease at very early stage so that proper precautions can be taken. Bio-Database is the collection of information of medical science by Muthulakshmi and Rani (2012), which contains information about patients, diseases and cause of diseases and many more things related to medical science but this database of Bio-Database contains very huge amount of data or the information which is not easy to analyze and also finding out some useful information from that is also very difficult. We use data mining techniques in order to get some useful information from this huge database. Medical science and market analysis is a field where large amount of data is gathered and collected from many sources now the challenge is to find the appropriate information and pattern from that data so that it can be used for further research to find some valuable results for the patients and customers but security is the major issue we should be very careful while sending data from one place to other otherwise it may create some harmful effects. In this chapter, work is mainly to provide privacy by Vaidya, Clifton, Kantarcioglu, and Patterson (2008) to such type of data so that the information remains safe while transferring data from one place to other. In this chapter, we are going to concentrate on finding the valuable information or patterns or relations between many things from large dataset which can be of any field and then security will be our major concern while transferring data from one environment to other environment for which we will use data modification techniques by Vaidya (2004), which will provide security to database and ensures secure transformation of valuable data.

SECURE ANALYSIS OF IRIS DATABASE USING THE WEKA TOOL

Data mining is the technique which is to be applied to large database to find useful patterns and information. In this chapter, we are going to take an iris database that is iris database which we treat as a centralized database. This centralized database is then divided into distributed databases for this we are going to use K means clustering techniques, in which one centralized data is divided different clusters and each cluster is distributed into different parts. After converting centralized database into distributed database, we now emphasize on security issues as while communicating data from one place to other we