

THE INTERNATIONAL JOURNAL OF SCIENCE & TECHNOLEDGE

A Study on Security & Privacy in “Mining Market Basket Database by a Service Provider”

Uma Rani

Assistant Professor, School of IT, JNTU, Hyderabad, Telangana, India

Dr. M. Srinivasa Rao

Professor, School of IT, JNTU, Hyderabad, Telangana, India

Prakash Soma

4th Sem, M.Tech, Computer Science, School of IT, JNTU, Hyderabad, Telangana, India

Abstract:

Data mining provides useful information for decision making in marketing development strategies. Now a days, many software companies are offering data mining services to various customers. At the same time most of the customers like small and medium size store's owners are not able to perform their data mining tasks on their own. These customers are willing to get data mining services from service providers. Sharing sales information with service providers involves losing privacy of the customers' personal information and valuable business details. In this paper we discuss possible security and privacy threats and some possible solutions to them.

Keywords: Data mining, Decision making, service providers, security and privacy threats.

1. Introduction

Data mining allows us to find useful information from a large database. In retail industry, the database consist data on sales, customer purchasing history, goods transportation, consumption and services. Mining information in market basket database provides the customer buying patterns and trends which can be used to improve customer service quality, good customer retention, store alignment strategy. Data mining resulting information also useful in financial data analysis, retail industry, telecommunication industry, biological data analysis, intrusion detection and so on. So most of the small and medium size retail industries are willing to mine their store's databases. As they are not capable of having technology and man power, they want to get data mining service from service providers. Many software companies are providing data mining services to their customers.

2. Security Issues

When a retail shop owner wants data mining service from a particular service provider, located geographically at remote location, it involves various security and privacy issues.

2.1. Confidentiality and Integrity

As the service provider located at different location, the customer must send his input database files through network. At the same time, the service provider must send his mining result through network only. There might be a number of security attacks on database files and mining result in network. Intruders may disclose the data base file's information and mining result or modify the input data and mining result. The confidentiality and integrity of the files that are going to be sent over the network will be in doubt.

3. Privacy Issues

All service providers are may not be trust worthy. Some service providers may keep input data provided by the retail shop owner and result of the mining result with them and may be used by them in future for their business purpose. They may also sell our valuable mining results to out market competitors. So, The privacy of input data and mining results in dilemma. This is a serious drawback in getting mining result from service providers.

4. Possible Solutions

4.1. Confidentiality

Confidentiality of input data and mining result can be achieved by using the following techniques.

4.1.1. Encryption

To provide absolute confidentiality, data encryption technique is used. The encryption process take input data (plaintext) and produce new cipher text data from which the original data cannot be recovered without the use of some particular information i.e. key. There are two types of encryption algorithms namely symmetric and asymmetric algorithms. The input data that is going to be sent from customer is to be encrypted and that is to be decrypted by service provider. The mining result is also to be sent over network only after encryption by service provider and that will be decrypted by the customer. Password based encryption is more suitable as data is transferred between two parties. Examples for symmetric encryption algorithms are DES,3DES,AES.Examples for asymmetric encryption algorithms are RSA, Diffie-Hellman.

4.1.2. Data Transformation

Data Transformation techniques provides a statistical guarantee of data confidentiality. The main goal of transformation is to make irreversible data modification and destroy actual values and correlations among them. The main idea in data transformation is to preserve aggregate trends in original data while changing the original data. For example, data may be interchanged between different rows to hide exact mapping between fields of a given record, noise data may be added to the data up to some limit.

4.2. Integrity

To provide integrity to the input data and mining results, message diagnostic algorithms like MD5 and SHA-1 are used.

4.3. Privacy

If store owner sends original sales database to service providers, there may be a chance of misusing the customers’ personal details and sales information of the store. Sales information may be shared with business competitors. To preserve privacy in the database, we can use various mapping and replacement techniques. Some of them are mentioned below.

The below table consist 5 transactions each having different items.

Transaction	Items
T1	Bread, Milk, Sugar, Rice
T2	Milk, Sugar, Rice, Salt
T3	Bread, Rice, Salt
T4	Bread, Milk, Sugar
T5	Milk, Rice

Table 1: General representation of the sales database

In the above general retail database the items in transactions are represented with original items’ names but we can use generic symbols like True/False,0/1 or Yes/No instead of using exact names of items.

	Bread	Milk	Sugar	Rice	Salt
T1	1	1	1	1	0
T2	0	1	1	1	1
T3	1	0	0	1	1
T4	1	1	1	0	0
T5	0	1	0	1	0

Table 2: Alternative representation of sales database

Items’ names can be mapped with some generic variables so that service provider cannot understand the association among the items.

Items	Symbols
Bread	A
Milk	B
Sugar	C
Rice	D
Salt	E

Table 3: Mapping between items and symbols

After performing mapping and replacement activities the sales database that is to be sent to the service providers is to be as below.

	A	B	C	D	E
T1	1	1	1	1	0
T2	0	1	1	1	1
T3	1	0	0	1	1
T4	1	1	1	0	0
T5	0	1	0	1	0

Table 4: Abstractive representation sales database

By sending only generalized sales database to service providers we can preserve privacy of customer's personal information and secrecy of valuable business transaction database.

5. Conclusion

In this paper, we discussed various security problems and privacy issues in getting mining service from service providers and also proposed some suitable solutions to them.

6. References

1. Jiawei Han & Micheline Kamber & Jian Pei (2011). Data Mining Concepts and Techniques
2. William Stallings. Network security essentials applications and standards (4th edition)
3. S.Kasturi & T.Meyyappan (2013) Detection of sensitive items in market basket database using association rule mining for privacy preserving, Proceedings of the 2013 international conference in Feb 21-22