CHAPTER-10

CONCLUSION
CHAPTER-10
CONCLUSION

Defines the work privacy as someone’s right to keep their personal matters and relationships secret. Today, every form of commerce leaves and electronic trail, and acts that were once considered private or at least quickly forgotten, are stored for future reference. The world over complain that they are overwhelmed by the amount of data available to them, but that they are unable to make any sense of this data. The changing business environment and the fact that customers are becoming more and more demanding highlight the need for organizations to be able to adapt faster and more effectively to those changes.

Privacy violation may occur legal liability that could result in expensive law suits. Data mining developed as a direct result of the natural evolution of information technology. The increased organizational use of computer based system has resulted in the accumulation of vast amounts of data, and the need for decision makers to have efficient access to knowledge, and not only data, has resulted in more and more organizations adopting the use of data mining. Privacy violations may result in bad press that can do considerable damage to corporate or brand image.

Privacy in data mining is a novel research direction in data mining and statistical databases, where data mining algorithms are analysed for the side-effects they incur in data privacy. The main objective or privacy in data mining is to develop algorithms for modifying the original data in some way, so that the private data and private knowledge remain private even after the mining process. The promise of data mining is to return the focus of large, impersonal organizations to serving their customers.
better and to providing more efficient business processes. Indeed, for some organizations data mining offers the potential for gaining a competitive advantage, but for others it has become a matter of survival.

On thing that information technology experts and business professionals must realize is that following ethical practices and respecting the privacy of individuals makes good business sense. The bad publicity associated with a single incident can taint and organization’s reputation for years, even when the organization has followed the law and done everying that it perceives possible to ensure the privacy of those from whome the data was collected. The literature is filled with examples of the successful application of data mining, not only to specific business functions, but also in specific industries. Undoubtedly, certain industries, such as those dealing with huge amounts of data, and those exposed to many diverse customers, stand to benefit more from data mining than others.

The benefits associated with data mining, for organization, individuals and society as a whole, far exceed its drawbacks, but the biggest issue facing organizations that want to employ data mining, is its cost. The other drawbacks of data mining relate to the threat that it poses to privacy, and any data mining efforts mush not only be done within the framework of the relevant laws, but must also be done in an ethical manner.

Although data mining is probably beyond the financial ability of most organizations, its main principle, the fact that there might be value in organizational data, should not be forgotten. Organizations much Endeavour to treat their data with the same respect it has for all its other corporate assets.
As more and more business organizations adopt Web services, ensuring secure communication between communicating partners is becoming even more important. For Web services environments, security is becoming even more important due to the Web services' unique characteristics. In the dissertation, I had discussed the Web services' characteristics, technologies & standards; however the focus had been on the Web services' security.

In the dissertation, there has been two main parts, technologies study and implementation. Under the technologies study, I had been introduce the main ideas & concepts of the core Web services' technologies, explain why SSL falls short when it comes to Web services and XML-based Web services security schemes. And I had been also discuss the importance of the federated identity. Then I will implement the Web services security under the technologies of Web Sphere and J2EE by deploying a Web service on a case of e-marketplace so to show how these technologies might be used together in future work.

We have presented how web mining (in a broad sense, DM applied to e-commerce) is applicable to improving the services provided by e-commerce based enter-prizes. Specifically, we first discussed some popular tools and techniques used in data mining. Statistics, AI and database methods were surveyed and their relevance to DM in general was discussed. We then presented a host of applications of these tools to DM in e-commerce. Later, we also highlighted architectural and implementation issues.

We now present some ways in which web mining can be extended for further research. With the growing interest in the notion of semantic web, an increasing number of sites use structured semantics and
domain ontologies as part of the site design, creation, and content delivery. The notion of Semantic Web Mining was introduced by Berendt et al (2002). The primary challenge for the next-generation of personalization systems is to effectively integrate semantic knowledge from domain ontology into the various parts of the process, including the data preparation, pattern discovery, and recommendation phases. Such a process must involve some or all of the following tasks and activities.

(1) Ontology learning, extraction, and preprocessing: Given a page in the web site, we must be able extract domain-level structured objects as semantic entities contained within this page.

(2) Semantic data mining: In the pattern discovery phase, data mining algorithms must be able to deal with complex semantic objects.

(3) Domain-level aggregation and representation: Given a set of structured objects representing a discovered pattern, we must then be able to create an aggregated representation as a set of pseudo objects, each characterizing objects of different types occurring commonly across the user sessions.

(4) Ontology-based recommendations: Finally, the recommendation process must also incorporate semantic knowledge from the domain ontologisms.

Some of the challenges in e-commerce DM include the following.

**Crawler/bot/spider/robot identification:** Bots and crawlers can dramatically change click stream patterns at a web site. For example,
some websites like (www.keynote.com) provide site performance measurements. The Keynote bot can generate a request multiple times a minute, 24 hours a day, 7 days a week, skewing the statistics about the number of sessions, page hits, and exit pages (last page at each session). Search engines conduct breadth-first scans of the site, generating many requests in short duration. Tools need to have mechanisms to automatically sieve such noisy data in order for DM algorithms to yield sensible and pragmatic proposals.

**Data transformations:** There are two sets of transformations that need to take place:

(i) Data must be brought in from the operational system to build a data warehouse, and (ii) data may need to undergo transformations to answer a specific business question, a process that involves operations such as defining new columns, binning data, and aggregating it. While the first set of transformations needs to be modified infrequently (only when the site changes), the second set of transformations provides a significant challenge faced by many data mining tools today.

**Scalability of data mining algorithms:** With a large amount of data, two scalability issues arise: (i) most data mining algorithms cannot process the amount of data gathered at web sites in reasonable time, especially because they scale nonlinearly; and (ii) generated models are too complicated for humans to comprehend.

**The above challenges need to be better addressed in real world tools:**

Episode mining involves mining not one-time events, but mining for a historical pattern of events. Episode-mining methods rely on extensions
of rule-mining methods. Alternate approaches could be explored here. Support vector machines have taken the centre stage of late, in learning linear and nonlinear relationships. Their applications in episode mining could be an exciting area for further work.