Research Methodology

The primary intention of my research is to design and develop a technique for privacy protection for the utility patterns in knowledge discovery process. The important consideration in privacy preservation is to provide the proper balance between the privacy protection and knowledge discovery. In order to handle these scenarios carefully, the protection can be done through Three ways, such as, sanitization, measure reduction in mining and post-processing. The data sanitization is the process of identifying and reducing the sensitive attributes from database and in the second way, the sensitive attributes will be identified in knowledge discovery process and the measure will be reduced. In the third way, the utility patterns mined from the database will be converted to insensitive utility patterns. These three ways will be effectively handled by the proposed approach. The proposed approach will be designed to do the privacy protection in these three ways. Here, utility pattern mining algorithm will be devised utilizing the tree-based data structure and then, the privacy protection schemes will be applied. The performance of the proposed approach will be evaluated with the help of benchmark databases and three different evaluation metrics such as, hiding failure, miss cost and database difference ratio. The proposed approach will be implemented using JAVA/Matlab..

A preliminary list of evaluation parameters to be used for assessing the quality of privacy preserving data mining algorithms, is given below:

- the *performance* of the proposed algorithm in terms of time requirements, that is the time needed by each algorithm to hide a specified set of sensitive information;
- the *data utility* after the application of the privacy preserving technique, which is equivalent with the minimization of the information loss or else the loss in the functionality of the data;
- the *level of uncertainty* with which the sensitive information that have been hidden can still be predicted;
- the *resistance* accomplished by the privacy algorithms, to different data mining techniques.
**Work Plan:**

First Three Months (1 to 3): Literature Review.

Next Six Month (3 to 9): Designing of approach

Next Three Months (9 to 12): Development of approach

Next Three Months (12 to 15): Implementation, Testing and Analyzing the performance

Next Three Month (15 to 18): Findings, Conclusion and Suggestion.

**Scope and significance**

It is often the case that no privacy preserving algorithm exists that outperforms all the others on all possible criteria. Rather, an algorithm may perform better than another one on specific criteria, such as performance and/or data utility. It is thus important to provide users with a set of metrics which will enable them to select the most appropriate privacy preserving technique for the data at hand, with respect to some specific parameters they are interested in optimizing.