6. CONCLUSION AND FUTURE WORK

Data mining – the non-trivial extraction of novel, potentially useful and ultimately understandable patterns from large data - has recently seen an upsurge in research community. Data mining has been applied to multiple domains to solve hitherto unsolved problems and results have been quite promising.

Software engineering is the application of systematic and quantifiable approach for the development and delivery of quality software. Development of quality software is fraught with a lot of challenges. Researches in the domain of software engineering have attempted to develop sound techniques, principles and tools that can aid in the engineering of quality software.

Given its success in other domains, the natural question that arises is whether data mining can be applied to solve problems that have been confronting software engineers and organizations for a long time. This research work is to investigate how data mining techniques can be applied to solve problems in software engineering.

In this context, the researcher first applied clustering for the discovery of fault-prone modules. Clustering is the process of organizing data into clusters so that objects within a cluster exhibit a high degree of similarity. The proposed research work is to cluster modules so that fault-prone modules form a single cluster and non-fault prone modules form another cluster. The researcher
applied genetic k-means clustering algorithm for the problem and found that genetic k-means performs well compared to the k-means algorithm.

An attempt is also made to mine association rules pertaining to human factors involved in software development. Human factors have a profound impact on the development of quality software. Many researchers have attempted to list such human factors. It is always not clear as to which of the human factors plays a pivotal role in the quality of the developed software. Here genetic algorithm is applied to mine association rules pertaining to human factors and software quality, and it is found that motivation, commitment and domain experience are the most crucial human factors. While 50 teams involved in the development of 50 websites are used here as subjects of analysis and therefore is small-scale, it throws light on the prospects of discovering significant human factors using association rule mining.

Software reliability estimation has been a great challenge to the software engineering community, and various approaches like usage of static code analyzers and testing techniques have been proposed. While each of the approaches has its own advantages, none of these approaches in isolation can fully result in reliable software. Taking cognizance of the need for additional techniques for ensuring software reliability, GA is applied in this study to the problem of software reliability classification, and it is found that GA was successfully able to classify around 90% of modules correctly. Such an
identification of modules likely to be unreliable can greatly aid the manager in allocation of resources for testing of such modules.

Data mining techniques like clustering, association rule mining and classification have been applied for solving some issues confronted in the domain of software engineering. Software engineering has many more issues that need to be resolved, and data mining techniques can greatly aid in discovering solutions to such pressing problems in software engineering. As a part of future work various other data mining techniques like spatial data mining and time series analysis can be applied to solve problems in software engineering. The software engineering domain is yet to benefit fully from the potential of data mining techniques, and this study is a first step in this direction.