(C4.5) algorithms could yield high predictive accuracy. Also, the feature evaluation methods CHI and ING were significantly dominating other feature evaluation methods. The results of the predictive accuracy of the machine learning algorithms further justifies using misclassification cost measure, which confirmed that, both Neural Net and Decision-Tree algorithms were best suited for student performance prediction for higher secondary students.

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

The academic achievement of higher secondary school education in India is a turning point in the life of any student, as it serves as a very important link between the higher and higher secondary education of students. But, there are determinants like demographic, academic and socio-economic factors of students that restrict the students' performance. This necessitates the need for some forecasting systems to predict the academic performance of students at plus two examinations. This is an attempt made first time in this aspect, which is mainly devoted to design and develop a prediction model by taking into account variables pertaining to the Indian society, for Indian educational system. In this connection, (i) Identification and collection of the variables that are related to the academic performance of higher secondary students in Indian context, (ii) Investigation on the highly correlated variables that are influencing student academic performance predominately by means of different
feature selection of data mining and (iii) Proposition of suitable performance prediction model with reasonable predictive accuracy have been framed as the main objectives for the present investigation. As the present investigation pertains to the model prediction of the performance of students in higher secondary education, it will be useful to parents, educators, academics, and policy makers for taking appropriate decisions for the development of student community in the highly competitive world to-day. It has high social relevance due to its linkage with academic achievement of students and hence the development of the society.

Wide literature review on academic performance of students and its prediction by using performance models was carried out. But, it was noticed that limited research investigations have been executed not only on the factors that are influencing the academic performance of the students at high school/ higher secondary level but also on the prediction of the academic performance of the students using different classification algorithm in data mining. So, the specific reviews of literature pertaining to the present investigation were utilized with the categorizations on reviews related to the factors that influence the academic performance and reviews with reference to performance models. From the review of literature, it was found that a number of studies identified different factors that could influence the academic achievement of students. It was also found that various prediction models were proposed in different contexts. It is worth mentioning here that the foundation of the present investigation was strongly based on the information from the review of literature.

An overview about Feature Selection and the main approaches used in the literature to carry out feature evaluations was carried out subsequently. The advantage of feature selection was found to be that it could reduce the dimensionality of the student data. This led to allowing learning algorithms to operate faster and more effectively to produce better classification results with high predictive accuracy. The effect of feature selection was elaborately studied interms of Filter, Wrapper and Hybrid methods on the student data set by varying class values (two-class, three-class, five-class and seven-class) and the results were perceived. It was observed that there were variations in cardinality of the feature subset with variations in class values of the class variable for the same set of independent variables. The efficacy of the feature subsets generated by Filter, Wrapper and Hybrid based methods through series of machine learning algorithms was studied subsequently.

An extensive evaluation of five classifiers with different configurations settings was carried out and it was observed that the predictive accuracy of the classifiers ranged from 75% to 98% for two-class class variable, 55% to 88% for three-class class variable, 40% to 92% for
five-class class variable and 31% to 84% for seven-class variable. In addition, it was observed that the Decision Tree and Neural network models showed better performance based on predictive accuracy as well as misclassification cost measure. In examining the problem of prediction of performance, we observed that it was possible to automatically predict students’ performance. Moreover, by using extensible classification formalism such as Neural Network, Decision-Tree, with Rough Set Theory it may be possible to easily and uniformly integrate such knowledge into the learning task.

From this investigation, we observed that, practical machine learning algorithms often make assumptions or apply heuristics that trade some accuracy of the resulting model for speedy execution, and comprehensibility of the result. While these assumptions and heuristics are reasonable and often yield good results, the presence of irrelevant and redundant information can often mislead them, resulting in reduced accuracy and less understandable results. Feature subset selection can help focus the learning algorithm on the important features for a particular problem. It can also reduce the dimensionality of the data, allowing learning algorithms to operate faster and more effectively.

Based on the outcomes of the present investigation, the future scope of research has been framed. They have been listed below:

As the present study covers only three districts, the study shall be extended all over the state so as to aid the development of the students

As the present study limits the number of features, more number of features shall be incorporated so as to enhance the efficiency of the models

As there is a need for comparison of the experimental estimation and theoretical prediction, students shall be given specific tests and they shall be evaluated so as to compare their performance against the prediction of our system.

As there is a need of several methods to predict performance, it is essential to explore more learning algorithms. In this regard, we feel that techniques based on rough set theory, Neuro-Fuzzy and other hybridization machine learning algorithms may throw light on the subject.

As there is a need of using clusters to predict performance, a systematic study is to be carried out for further beneficial recommendations. While the existing clustering algorithms are based on similarity checking, it shall be essential to explore on difference checking so that automatic composition of groups with heterogeneous nature becomes possible.
It is also worthwhile to include performance prediction into the student model component of an intelligent tutoring system.