CHAPTER 2
LITERATURE REVIEW

To present the thesis entitled “Computational model for prediction of learning performance” related available information in general has been reviewed and presented below.

2.1. Literature survey

Research work of Yehuala Muluken Alemu [103] investigated the potential applicability of data mining technology to predict student success and failure cases. Classification and prediction data mining functionalities are used to extract hidden patterns from students’ data. The classification rule generation process is based on the decision tree and Bayes as a classification technique and the generated rules were studied and evaluated. Using WEKA 3.7 application software does analysis. The research results offer a helpful and constructive recommendations to the academic planners in universities of learning to enhance their decision making process. So, on the bases of the research findings the level of student success will increase and it is possible to prevent educational institutions from serious financial strains.

Tiwari Ravi, Awadhesh Kumar Sharma [96] build a prediction model for placement prediction for students. It takes education data i.e. student data as an input and predicts the status of placement for the student. It also suggests the best field according to nature of the student data for particular student, which have a chance to place. So through this educational organization can predict the placement of each student and can work according to it to improve the placement of the organization. By applying the various algorithms on the given dataset they find following results. After applying the various algorithms on the student information data set find Id3 is 71% accurate, BayesNet is 70% accurate, J48 is 71% accurate, RBF network algorithm is
65 % accurate and Random Tree algorithm is most accurate for the prediction process i.e 73%. According to them random tree algorithm is best suited for prediction of placement status for a student.

Pawel Bujnowski and others [77] construct a new classifier called an intuitionistic fuzzy decision tree is presented. Data is used to analyze the performance of the classifier. The results are compared to some other popular classification algorithms. Finally, the classifier behavior is verified while solving a real-world classification problem. They presented and tested an extension of the fuzzy ID3 decision tree algorithm, namely, a new intuitionistic fuzzy decision tree. The new classifier was tested on real-world data examples giving very encouraging results.

Fadhilah Ahmad and others [6] proposes a framework for predicting students’ academic performance of first year bachelor students in computer science course. For work they collect students’ demographics, previous academic records, and family background information. Decision tree, Naïve Bayes, and Rule based classification techniques are applied to the students’ data in order to produce the best students’ academic performance prediction model. The experiment result shows the rule based is a best model among the other techniques by receiving the highest accuracy value of 71.3%. The extracted knowledge from prediction model will be used to identify and profile the student to determine the students’ level of success in the first semester.

R.R. Kabra and R. S. Bichkar [83] discussed use of decision trees in educational data mining. Decision tree algorithms are applied on engineering students’ past performance data to generate the model and this model can be used to predict the students’ performance. It will enable to identify the students in advance who are likely to fail and allow the teacher to provide appropriate inputs. The model developed by them shows 69.94 % accuracy that is out of 346 instances 242 are correctly classified. One of the aims of presenting work is increase accuracy also so that treating is as base paper.
Sonia Joseph and Laya Devadas [58] present a classification model based on decision tree approach to predict student’s academic performance. This method is useful in identifying those students who are probable to fail in the semester examinations and allow the teachers to provide appropriate assistance in timely manner. It also helps the weaker students to improve and bring out betterment in the result. They model a modified ID3 weighted decision tree. According to experiment accuracy comparison is C4.5 45.8333%, ID3 52.0833%, CART 56.25%, MODIFIED ID3 76%.

According to Bansode Jyoti [14] state that using decision tree students' performance can be predicted. The students, whose performance is poor, can be warned. The management can take necessary action to improve their performance by giving more attention, taking extra lectures etc. Due to such measures student performance can be improved. The number of failures can be reduced. Ultimately college results also get improved. Presently only traditional learning is considered.

Abeer Badr El Din Ahmed and others [3] used classification task to predict the final grade of students and as there are many approaches that are used for data classification. They used ID3 algorithm. This study will help the student's to improve the student's performance, to identify those students, which needed special attention to reduce failing ration and taking appropriate action at right time.

Namdeo Jyoti and Others [70] is discusses the basic approach and concepts of the rough set theory in the field of academic domain for the performance prediction of students in course works. They used weka tools and different algorithms. On the basis of previous student result they predict the future student result using J48, NB, MLP and random forest. According to result classification accuracy is not very high. Although NB is giving more correct classification accuracy compared to rest three.

Adhatrao Kalpesh and others [5] developed a system, which can predict the performance of students from their previous performances using concepts of data mining techniques under Classification. They analyzed the data set containing
information about students, such as gender, marks scored in the board examinations of classes X and XII, marks and rank in entrance examinations and results in first year of the previous batch of students. By applying the ID3 (Iterative Dichotomiser 3) and C4.5 classification algorithms data, they predicted the general and individual performance of freshly admitted students in future examinations. In result for a total of 182 students, the average percentage of accuracy they achieved in bulk and singular evaluations is approximately 75.275. This paper act as a base paper for the work.

Yongrong He and others [40] researched comprehensive quality evaluation system of college student, and introduces data mining method to solve mess students’ data. Firstly, they construct students’ data warehouse, then construct comprehensive multi-dimensions OLAP (Online Analysis Process) model, and through classification deterministic method. They get useful message from the system to guide college manage students, and some Employers can find their satisfied employees. The system was divided as three isolated aero: moral, academic, extracurricular practice. The moral data often come from evaluation and peer assessment and evaluated by instructor; The academic data includes core courses, required courses, and elective courses, when computing scores, can give different weight to Corresponding course, and standardized all the score.

V.Ramesh and others [99] identify the factors which are influencing the performance of students in final examinations and find out a suitable data mining algorithm to predict the grade of students so as to a give timely and an appropriate warning to students those who are at risk. In the investigation they do survey and experimental methodology adopted to generate a database and it was constructed from a primary and a secondary source. The obtained results from hypothesis testing reveals that type of school is not influence student performance and parents’ occupation plays a major role in predicting grades. This work will help the educational institutions to identify the students who are at risk and to and provide better additional training for the weak students.
A.F.Elgamal [1] presents an educational data-mining model for predicting student performance in programming courses. Work by ElGamal taken as base paper for the present task. Identifying variables that predict student-programming performance may help educators. These variables are influenced by various factors. The study engages factors like students' mathematical background, programming aptitude, problem solving skills, gender, prior experience, high school mathematics grade, locality, previous computer programming experience, and e learning usage. Curriculum committees can use prediction results to guide changes to the curriculum and evaluation of the effects of those changes. This study investigates the potential of data mining for enhancing the effectiveness of academic planners and level advisers in higher institutions of learning. The machine-learning algorithm extracts only the rules, which have the antecedent(s) satisfy 100 % of the conditional probability within certain class. The extracted rules confirm the important of variables such as High School Mathematics Grade and programming aptitude. Educators should tack these variables into account in the qualifying exam to join computer science departments at universities.

Adeleke Raheem Ajiboye [4] provides a model using fuzzy logic approach to predict the risk status of students based on some predictive factors is proposed. Some basic information that has some correlations with students’ academic achievement and other predictive variables were modeled, the simulated model shows some degree of risk associated with their past academic achievement. The result of study would enable the teacher to pay more attention to student’s weaknesses and could also help school management in decision making, especially for the purpose of giving scholarship to talented students whose risk of failure was found to be very low; while students identified as having high risk of failure, could be counseled and motivated with a view to improving their learning ability. The researchers modeled the transformed input predictive variables using the approach of fuzzy logic. The technique of fuzzy logic applied in research shows its capability of handling uncertainty. The results segmented the students according to their risk status, the model can be applied to predict the academic performance of all applicants seeking admission to Nigerian institutions of higher learning and the technique used can be generalized to make similar prediction in any institution outside Nigeria.
Olufunke O. Oladipupo [73] examines the relationship between students’ pre admission academic profile and academic performance. The pre admission academic profile considered includes ‘O’ level grades, University Matriculation Examination (UME) scores, and Post-UME scores. Fuzzy association rule mining (FARM) was used to identify the hidden relationships that exist between students’ pre-admission profile and academic performance. It determines students’ performance ratings as against their pre-admission academic profile. This can serve as a predictor for admission committee to enhance the quality of the new in-take and guide for the academic advisers. It would intimate the level advisers the basis upon which they could monitor each student academic performance appropriately.

M.Geethalakshmi and others [66] analyses the results of three students using Fuzzy Matrix Solution (FMS) with the help of product of fuzzy matrices. They introduce fuzzy matrix by introducing max-min operation & min-max operation. They conclude the results of three students as pass or fail using fuzzy matrix. Using Visual Basic they develop software to calculate Fuzzy relation. The software calculates from R1 to R4gives the status of results of the students.

Goel Gagan and others [33] have developed a student model that respects the process of knowledge tracing. The model can handle data at a higher level of abstraction and it also has the ability to deal with uncertainty. As the inputs and the rules are particularly comprehensible for humans due to the linguistic levels expressed in natural language, it is quite easy to adapt and refine the model (for instance by experts). In the past, knowledge tracing has been implemented with Hidden-Markov model and logistic regression. Determination of the parameters (the thresholds of each membership function) may be done either by experts or by machine learning algorithms. Results show some over-fitting and a lack of precision in the beginning, so constructing the model with machine learning techniques is also important, this would help us improve the accuracy of the model during the initial stages.

M'hammed and others [2] describes a hybrid approach which uses EDM and regression analysis to analyze live video streaming (LVS) students’ online learning behaviors and their performance in their courses. Students’ participation and login
frequency, as well as the number of chat messages and questions that they submit to their instructors, were analyzed, along with students’ final grades. Results of the study show a considerable variability in students’ questions and chat messages. They provide a strong and coherent analytical framework capable of enabling a deeper and richer understanding of students’ learning behavior and experiences. Their results revealed several students’ learning behaviors, ranging from active participation and interaction with the instructor to a lack of participation or even of attendance.

Osmanbegovic Edin with other [74] different methods and techniques of data mining were compared during the prediction of students' success, applying the data collected from the surveys conducted during the summer semester at the University of Tuzla. The success was evaluated with the passing grade at the exam. Three supervised data mining algorithms were applied on the preoperative assessment data to predict success in a course (either passed or failed) and the performance of the learning methods were evaluated based on their predictive accuracy, ease of learning and user friendly characteristics. The results indicate that the Naïve Bayes classifier outperforms in prediction decision tree and neural network methods. It has also been indicated that a good classifier model has to be both accurate and comprehensible for professors. In future investigations, with identifying and evaluating variables associated with process of studying, and with the sample increase, it would be possible to produce a model which would stand as a foundation for the development of decision support system in higher education.

M Narayana Swamy [94] present the data mining technique to predict the performance of the students based on the enrollment data. It helps the teacher to take remedial measure for slow learners to improve the performance in the university examination. This study shows that student’s enrolment data of MCA course can be used to create the model using decision tree algorithm that can be used for prediction of student’s performance in MCA degree. From the confusion matrix it is clear that out of four actual categories, the accuracy of the model for the FAIL class is 92.5% that means model is successfully identifying the students who are likely to fail. These students can be considered for proper counseling so as to improve their result.
Kalyani M. Moroney and others [64] presents a system that can be used for a performance improvisation of students in their academic studies. The system accumulates a vast amount of information, which is very valuable for analyzing the student’s performance. The outcome of the result indicates that some students are good in subject or topic but some are really poor and need the help to improve in it. Using this system they suggest admin to necessarily concentrate on these students and to appoint experts in these subjects for teaching or preparing better study material than present. In the proposed system, regression is used for classification. Division technique of regression is used to differentiate between poor and good performers. The solution set for every question paper is provided so that students can understand where they are wrong.

Kabakchieva Dorina [24] research is focused on the development of data mining models for predicting student performance, based on their personal, pre-university and university-performance characteristics. The dataset used for the research purposes includes data about students admitted to the university in three consecutive years. Several well-known data mining classification algorithms, including a rule learner, a decision tree classifier, a neural network and a nearest neighbor classifier, are applied on the dataset. The performance of these algorithms is analyzed and compared. The highest accuracy is achieved for the neural network model (73.59%), followed by the Decision Tree model (72.74%) and the k-NN model (70.49%). The neural network model predicts with higher accuracy the “strong” class, while the other three models perform better for the “weak” class. The data attributes related to the students’ university admission score and numbers of failures at the first-year university exams are among the factors influencing most the classification process.

Parashar Hem Jyotsana and others [76] discuss on the shortcoming of ID3’s inclining to choose attributes with many values, and then a new decision tree algorithm that is improved version of ID3. In our proposed algorithm attributes are divided into groups and then they apply the selection measure 5 for these groups. If information gain is not good then again divide attributes values into groups.
These steps are done until get good classification/misclassification ratio. The proposed algorithms classify the data sets more accurately and efficiently.

Mlambo Victor [68] designed to identify and analyze some determinants of academic performance in an introductory biochemistry (AGRI 1013) course plagued by chronic high failure rates. A survey instrument was administered to a random sample of 66 registered students of AGRI 1013 (representing a 40% sampling fraction) to generate data on demographics (gender and age), learning preference, and entry qualifications. The effect of learning preference, age, gender, and entry qualifications on academic performance (measured as the final coursework mark obtained) was determined. Relationships/associations between gender and learning styles, gender and entry qualifications, age and learning preferences, and age and entry qualifications were analyzed using Pearson’s chi-square test. There were significant (P < 0.05) associations between entry qualifications and both gender and age. However, since entry qualifications did not significantly (P > 0.05) affect academic performance, this association should be of limited concern. Author was concluded that more determinants of academic performance need to be investigated and that students who are admitted based on a diploma in agriculture may need a remedial course given that their coursework grades, though statistically insignificant were consistently lower than that of the other students.

Sembiring Sajadin and others [90] was applying the kernel method as data mining techniques to analyze the relationships between students’ behavioral and their success and to develop the model of student performance predictors. Using smooth support vector machine (SSVM) classification does this and kernel k-means clustering techniques. The results reported a model of student academic performance predictors by employing psychometric factors as variables predictors. The result indicates that data mining techniques (DMT) capabilities provided effective improving tools for student performance. It showed how useful data mining could be in higher education in particularly to predict the final performance of student.

They expressed the strong correlation between mental condition of student and their final academic performance.
Bharadwaj Brijesh Kumar And other [16] designed to justify the capabilities of data mining techniques in context of higher education by offering a data mining model for higher education system in the university. In there research, the classification task is used to evaluate student’s performance and as there are many approaches that are used for data classification, the decision tree method is used in the work. They extract knowledge that describes students’ performance in end semester examination. It helps earlier in identifying the dropouts and students who need special attention and allow the teacher to provide appropriate advising/counseling.

Bhardwaj Brijesh Kumar and other [17] present an experimental methodology that was adopted to generate a database. The raw data was preprocessed in terms of filling up missing values, transforming values in one form into another and relevant attribute/ variable selection. As a result, they had 300 student records, which were used for by Bayes classification prediction model construction. In this paper, Bayesian classification method is used on student database to predict the students division on the basis of previous year database. Study shows that academic performances of the students are not always depending on their own effort. Investigation shows that other factors have got significant influence over students’ performance.

Diego García-Saiz and other [23] they compare the performance and interpretation level of the output of the different classification techniques applied on educational datasets and propose a meta-algorithm to preprocess the datasets and improve the accuracy of the model. Experimentation shows that there is not one algorithm that obtains a significantly better classification accuracy. In fact, the accuracy depends on the sample size and the type of attributes. When the sample size is very small (less than 100 instances) and contains numeric attributes, Naive Bayes perform adequately, on the other hand, when the dataset is bigger, BayesNet TAN is a better alternative. J48 is suitable for datasets with more instances and/or with the presence of nominal attributes with missing data, although in this last context Naive Bayes is the best but less interpretable. Due to the special characteristics of the datasets used, the best results are obtained with the meta-algorithm proposed using
both Naive Bayes and J48 to preprocess and to predict, being better if the preprocessed task is carried out according the most significant attribute for the algorithm used to preprocess.

B.Nithyasri and others [11] investigate the accuracy of data mining techniques in a particular environment. The first step of the study is to gather student’s data on technical, analytical, communicational and problem solving abilities. The second step is to clean the data and choose the relevant attributes. Attributes were classified into two groups “Demographic Attributes” and “Performance Attributes”. In the third step, Decision tree and Naive Bayes algorithms were constructed and their performances were evaluated. The study revealed that the Decision tree algorithm is more accurate than the Naïve Bayes algorithm. Their work will help the institute to accurately predict the performance of the students. From the results it is proven that ID3 algorithm is most appropriate for predicting student performance. The error rate is very high for Naïvebayes classifier. ID3 gives 98% prediction for 50 instances which is relatively higher than Naïve Bayes classifier.

Hesham A. Hefny [42] presented a technique to build fuzzy decision Tree by employing the ambiguity of attributes and classify ability of instance. Our technique builds a reduced FDT, which does not need for applying the pruning algorithms to reduce the size. The paper also presents the results of a set of empirical studies conducted on a dataset of UCI repository of machine learning database that evaluate the effectiveness of our technique compared to fuzzy iterative dichotomiser 3 (FID3), ambiguity, and FID3 with classify ability techniques. The studies show the effectiveness of technique in reducing the number of the extracted rules without loosing of the rules accuracy.

Tree induction has become an important technique for machine learning, expert system and prediction analysis and so on. Most existing methods are crisp and fuzzy decision tree induction. When choosing a decision tree induction method to classify unseen instance, mainly consider the generalization capability of tree induction. This paper analyzes and compares the generalization capability of decision tree between fuzzy and crisp tree algorithms.
The initial conclusion is that, for the classification problem of numerical attributes; the fuzzy decision tree has the stronger generalization capability than crisp one.

Sahay Amar and others [89] develop a software system to assist higher education in assessing and predicting key issues related to student success. The software uses several data mining algorithms and quality tool such as, quality function deployment to study and predict issues including but not limited to enrollment management, dropout rate, time to degree, and suggest ways to improve courses and programs. Data mining knowledge discovery and predictive modeling tools along with quality function deployment (QFD) are used to uncover and understand hidden patterns in vast databases to understand student related issues and suggest ways to improve them.

M. Ramaswami and others [65] investigation, a survey cum experimental methodology was adopted to generate a database and it was constructed from a primary and a secondary source. While the primary data was collected from the regular students, the secondary data was gathered from the school and office of the chief educational officer (CEO). As a result, they had 772 student records, which were used for CHAID prediction model construction. A set of prediction rules were extracted from CHIAD prediction model and the efficiency of the generated CHIAD prediction model was found. The accuracy of the present model was compared with other model and it has been found to be satisfactory.

S.K. Althaf Hussain Basha and others [9], proposed a method to predict the intra-year academic performance of the student using the historic data. The idea is to identify existing patterns in the historic data, and maintain a database for it. Comparing the current performance of a student with the existing patterns, they predict the possible performance of the student in the future. In the process, they identify any new patterns that come across. For the purpose of identifying the interestingness of the pattern look into the percentage of increase or decrease make, appropriate methodologies are used. The method has been proved to be effective from the error percentage calculated is less than 7.
M. N. Quadri and other [80] say that academic performance in this study is measured by their cumulative grade point average (CGPA) upon graduating. They presented the work of data mining in predicting the drop out feature of students. This study applies decision tree technique to choose the best prediction and analysis.

The list of students who are predicted as likely to drop out from college by data mining is then turned over to teachers and management for direct or indirect intervention. This study introduced the data mining approach to modeling drop out feature and some implementation of this approach. The key to gaining a competitive advantage in the educational industry is found in recognizing that student databases, if properly managed, analyzed and exploited, are unique, valuable assets.