

Development of Students' Academic and Employability Model through Data Mining

**Dr. Tripti Mishra*



The Abstract is of the Thesis "Development of Students' Academic and Employability Model through Data Mining" submitted by Dr. Tripti Mishra for the award of Ph.D. degree from Mewar University, Chittorgarh. The Supervisors were: Dr. C.D. Kumawat, Professor, Department of Computer Science, Mewar University, Chittorgarh and Dr. Sangeeta Gupta, Professor, Management Education and Research Institute, New Delhi. The University awarded the Doctorate to the research scholar in 2019.

**The Research Scholar Dr. Mishra is presently Professor, Delhi Institute of Advanced Studies, Delhi.*

INTRODUCTION

Background

With more than one billion population, in which half of the population is under the age of 25 years, India is all set to become the youngest nation by 2020. This pool of young people can be converted to highly productive human capital, by providing them higher education. Today higher education (HE) is an industry in itself which provides a strong work force to other industries.

Limited opportunities in government-run organizations for higher education has forced the Indian government to encourage private players to set up new institutions and deemed universities. Further, this unprecedented growth of Institutions has created a highly competitive education sector. The major challenges and issues of this sector are

- (i) Increasing students' dropout rate;
- (ii) Migration of students from one Institution to other Institutions;
- (iii) Finding the learning ability of the student by the institutions;
- (iv) Recommending a suitable course to the students in which s/he will excel;
- (v) Developing a curriculum based on students need and capability;
- (vi) Continuous improvement of quality of education;
- (vii) Increasing economic viability of the course by increasing the employability of the students;
- (viii) Curriculum update as per industry requirement

The sudden surge of private Institutions has increased the accessibility of HE to the masses. However, the quality of education being provided has always been questioned, as the unemployment of students having higher education degree, has been rising.

Keeping in view the problems faced by the Institutions and the above challenges, the Institutions are adopting technologies that can help them take timely and informed decision. Today every educational Institution generates a huge amount of data which can be a storehouse of information. Bearing in mind this ever growing data and the prominence of students' assessment, the solution of the problem will turn out to be data mining (DM). Concerning education, a novel application of DM technique known as educational data mining (EDM) can be used for knowledge discovery in educational databases for learning, cognition, and assessment. The research work presented in this thesis makes use of EDM concepts to develop a predictive model for predicting student's performance and their employability.



STATEMENT OF THE PROBLEM

The Institutions providing higher education are thriving in a highly competitive world. Students seek admission in an Institution based on the academic performance of the

students graduating from that Institution in a particular course and the employability of these students. According to (Susila M. et.al, 2013) career prospects and reputation of the Institution plays an important role in selection of Institution by students. A student dissatisfied on these parameters may migrate to another institution resulting in loss of reputation and revenue of the Institution. Thus the primary goal of any higher educational institution is to improve the academic decisions, endow with quality education, enhanced student's performance and elevated employability of students. With the increasing complexity of educational system, the decision making process is becoming all the more difficult. One of the significant facts in higher learning institution is the explosive growth of educational data. The technology like data mining can help in setting new strategies and better the existing process to improve quality of education. Data mining techniques can be used to build models that can help predict the academic performance of student as well as its employability. The performance prediction models can be used by Institute authorities to take timely action and give special attention, training and counseling to students at high risk of poor academic performance and at the risk of unemployment so as to bring much needed improvement in results. Additionally, the prediction can provide excellent guidelines for making better strategy in future.



OBJECTIVES OF THE STUDY

In context to the statement of the problem the objectives of the study are:

1. To identify factors that affect students' academic performance;
2. To design a predictive model for the performance of students using factors identified;
3. To suggest guidelines for timely intervention to improve performance thus increasing success rate;
4. To identify factors that affect students employability;
5. To design a predictive model for employability using factors identified;
6. To suggest a guideline for enhancing the employability and construct a decision support system for prediction of academic performance and employability.



THEORETICAL BASIS: EDUCATIONAL DATA MINING

International Educational Data Mining Society defines Educational Data Mining as "an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in."

EDM is poised to leverage a massive amount of research from the data mining community. It is an interesting research area which extracts useful, previously unknown patterns from educational databases for better understanding, improved

educational performance and assessment of student learning process and provide a system that can help different stakeholders of education system. EDM has been applied in different educational setups like traditional education, E-learning and student centric Intelligent Tutoring System. The stake holders of EDM are students, educators, course developers, and administration. Figure 1.1 depicts a typical data mining set up in educational systems.

However, effects of emotional skill have not been considered by researchers in EDM. In the present work we have considered, emotional skills, academic parameters and socio economic parameters to develop performance and employability prediction model.

Next section discusses the contribution of the thesis.

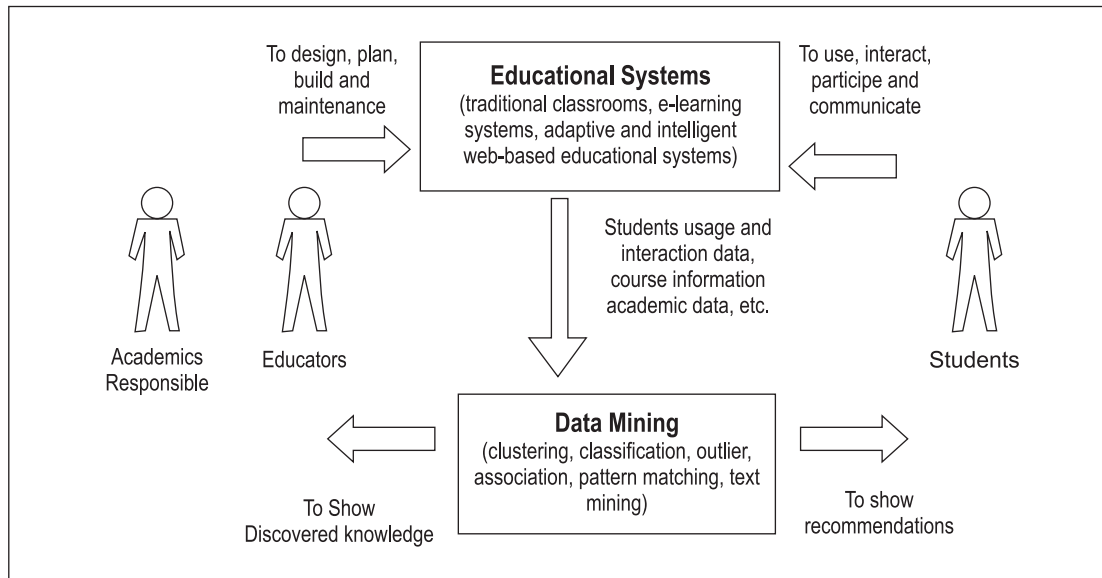


Figure 1: Data Mining in Educational Setup [Source: Romero and Ventura, 2007]



APPLICATIONS OF EDM

Different groups of users or stakeholders view the system provided by EDM from different perspectives, each group having its own goal, vision and objectives. Based on the perspective of various stakeholders, Educational data mining has forayed into various research areas such as

- Students' Modeling
- Recommender Systems
- Studying Pedagogical Support
- Course Management Systems
- Student Retention and Attrition
- Students' Performance Prediction
- Students' Employability Prediction

With the above methods, it has been observed that there is a need of further research that can help improve the quality of education and improve its economic viability.

Academic performance and employability of students build the reputation of an Institute. Performance prediction is a mature field. Researchers have used academic and socioeconomic parameters to predict the performance of students. Emotional skills are very important factors that have been proved to affect the performance and employability by (Potgieter , 2013),(Marc A. Brackett,et.al, 2012) in psychology.



CONTRIBUTION OF THE THESIS

The present research work is an attempt to combine the issues of quality education.

- (i) Students' academic performance prediction has been one of the favorite topics of the research but, not much work is done in the area of employability prediction. Predictive model for both academic performance and employability is developed in this study.
- (ii) The existing research has considered mostly academic and socioeconomic parameters for prediction purposes. The present study uses a standard tool (Emotional skill Assessment Process) to calculate emotional skills like leadership quality, self-esteem, drive strength, empathy, time management, stress management, decision making and assertion and uses them as predictive parameters apart from considering academic and socioeconomic parameters.
- (iii) A comprehensive Decision Support System is developed to predict the academic performance and employability of students to help the management (teachers/ trainers) of the institute.



LITERATURE REVIEW

The review of the past research efforts related to theoretical framework in Educational Data Mining method is organized chronologically

and categorically offer an insight on how past research efforts laid the groundwork for subsequent studies, including the present research efforts. The detailed review has been carried out, so that the present research can be tailored to add the present body of literature, as well as the scope and direction of the present research effort



EDUCATIONAL DATA MINING

Educational data mining develops methods and applies techniques from statistics, machine learning and data mining to analyze data collected from educational settings. From these it should be possible to test learning theories, develop algorithms and models to help education sector.

According to the literature study, it has been observed that EDM research pertains to mainly three heads, according to the way data is collected-

Traditional face to face or the offline education system based on data generated in the classroom. E-learning in which the learning is provided through online content based on online activity logs. Intelligent tutoring system (ITS) and Adaptive Educational Hypermedia System (AEHS) involve online teaching based on students need, his or her progress rather than providing same structured lesson to all the students. For the analysis purpose the data picked up by these systems are log files, user models, etc.

As the present work relates to traditional educational environment a comprehensive view of traditional educational environment is mentioned here.



TRADITIONAL EDUCATIONAL SETUP

Traditional educational set up refers to class room teaching which is still the most popular form of teaching by the Institutions across the globe and needs to be explored. Next two subsections discuss the researcher's inclination towards prediction of both students' performance and their employability. The important factors for performance and employability prediction are the attributes considered and method adopted for prediction. Both academic performance and employability take into consideration students past record, as well as performance in current learning environment and other attributes that can help or affect negatively student's learning.



PREDICTION OF STUDENTS' PERFORMANCE

Students' academic performance is a mature field now with many researchers contributing to it. Moving in chronological increasing order from 2007 onwards the researchers have considered various parameters for prediction and using different algorithm tried to predict the result of specific course of a particular university.

Nghe, Janecek and Haddawy (2007) have considered two diverse population, international, large Can Thao University of Vietnam and Asian Institute of Technology a small international postgraduate institute and achieved similar

levels of accuracy of prediction performance. In this analysis, the decision tree was consistently 3-12% more accurate than the Bayesian network.

Ogor (2007) has developed a simple student performance assessment and monitoring system based on various data mining techniques. The predictor attributes included students demographic details, course average score in 1st to 5th semester overall gain performance, etc. For classification of students in three categories Excellent, Good, Satisfactory. C5 showed highest accuracy (97.3%) followed by CART (92.5%), ANN (91.42%) CHAID (57.77%). Ogor further used clustering to understand groups, non-performance attributes that latently affect performance.

Obdokumet. Al (2008) used records of 112 students, each recording 10 attributes like University matriculation exam, GCE (General Certificate of Education) Score, Senior Secondary Certified Examination score SSCE), grades in O level subject, location of University from, gender and age, Cumulative Grade point was to be classified as Good, Average and Poor. Oladokum has used only Neural networks with 56% data for training 30% for Testing and 14% for cross validation. ANN was able to give 74% accuracy of prediction.

Instead of Higher education, Cortez and Silva (2008) have focused on prediction of school students performances taken from two secondary school students. A total of 33 parameters including socio -demographic details like (parental marital status, father's job, mother's job, quality of family relationship, attitude towards study (No. of hour, past failure) Internet facility, family support, free time after school, health , alcohol consumption etc. were considered and four Data mining algorithms Decision tree, Neural network, Support Vector machine Random forest were used to predict the academic performance. Attendance, parents job, previous year performance affect the current performance.

Feature Selection is the process of choosing a subset of input variables by removing the feature irrelevant to prediction. Ramaswami and Bhaskaran (2009) experimented with 1969 higher secondary student from different schools of Tamilnadu taking 32 variables initially for prediction of secondary exam (pass/fail). Student's gender, Eyesight, Community, Physical handicap, food habit, family details, mode of transport , medium of instruction, sports activity etc. were considered. Using WEKA a set of methods were applied and four reduced subsets obtained with maximum 17 and minimum 12 attributes. It was also observed that classification methods like Naïve Bayes, one R voted perception performed much better with feature selected subset than where all variables were considered.

Paris , Affendy and Musthpha (2010) have compared C4.5, NB Tree, Bayes Net, Hidden Naïve Bayes , and voting techniques of classification based on three weak classifiers (Naïve Bayes, One R and Decision Stump) for improving the accuracy of performance prediction. The combination of HNB method and one weak classifier Decision Stump is used for voting technique. This particular combination has been taken, as HNB works well with most of the classes except for a high distribution class where Decision Stump gives good result.

The students dropping out of an open polytechnic of New Zealand due to failure has been explored by Kovaic. Z (2010). Enrollment data consisting of socio-demographic variables (age, gender, ethnicity, education, work status, and disability) and study environment (course programme and course block) .CART algorithm was found to be best suited with highest accuracy.

Expanding their earlier work, Kovacic and Green(2010) applied classification on data of students enrolled in Open Polytechnic programme. Ethnicity, Course level, Secondary school, age, Course block , Course offer type and Work status have been found most important attributes after feature selection. The logistic regression and discriminant analysis models were found to be a superior prediction model with higher accuracy than the classification tree models (between 1% to 4%) however at the cost of using more variables. The authors therefore recommend the use of the CART classification tree model in the early identification of at risk students.

Ramaswami and Bhaskaran (2010) used Chi-squared Automatic Interaction Detector (CHAID) and HSC result prediction model was derived based on 34 independent variables. Apart from demographic details student health, tuition, care of study at home etc. have been studied. Prediction Accuracy obtained was 44.69, and Potential influences variables were found to be Xth grade marks, location of school and Mother's private tuition .

Guruler,Istanbulla, Karahasan (2010) used SQL server and Analysis services to construct Mugla University student knowledge discovery unit program. The study aims to discover individual characteristics that decide their success using Microsoft Decision Tree. The type of registration and family income were found to be the greatest factors affecting the target.

Affendey, Paris and Mustafa (2010) emphasized attribute importance by ranking them , using correlation based feature subset selection and consistency subset selection (COE) and using them further find accuracy of various classifiers. Unlike other researchers, authors do not consider demographical details of the students, but concentrate only on grades of various courses as attributes to predict categories of students as first class, the other consisting second class upper lower and third class. Using a 10 fold cross validation, it is observed that Naïve Bayes, AODE and RBF Network performed best on the date set (2427 records) followed by RBF Network SMO and Logistic and decision tree.

Oyelade et al (2010) applied cluster analysis on the data set of students of one semester of one of the Nigerian University. The performance index based on GPA was decided into Excellent (>70), Very good (60-69), Good (50-59), Very Fair (45-49), Fair (40-45) and Poor (below 45). By considering k=3,4,5 different cluster sets of students are obtained which help monitor the progression students' performance in higher education.

Semibiring et al (2011) applied smooth support vector machine (SSVM) classification and Kernal K means clustering techniques to develop a model of student academic predictors

by employing psychometric factors such as Interest, Study behavior Engage Time, and Family Support. Successful partitioning was obtained with 5 clusters J48 decision tree was used to generate predictive rules which was implemented into SSVM algorithm to predict the student final grade.

Bharadwaj and Pal (2011) base their experiment only on Previous Semester marks, class test grade, seminar performance, Assignment, attendance , Lab work to predict end semester marks.

Huang S (2011) uniquely considers 6 combinations of predictor variables, three pre requisite courses, scores of three dynamics midterm, to predict academic performance in Engineering Drawing course. The analysis reveals that type of MLR, multiplayer perception MLP network, radial basis function RBF network and support vector machine) has only slight effect on average prediction accuracy or percentage of predictions. Huang recommends use of SVM when individual academic performance of the student is to be predicted while multiple regression technique is best suited when average performance of whole class is to be predicted.

Cheewaprabkhit (2012) applied decision tree and neural network to predict academic performance. Decision tree proves to be a better classifier than the neural network with 1.31% more accuracy. Number of hours worked per semester, additional English course ,no of credits enrolled per semester and marital status of the students are major factors affecting the performance.

Sen, Uçar, and Delen (2012) have ranked importance of 24 predictor variables including demography, scores in Maths, Turkish, religion and ethics, science and technology and level determination exams etc for predicting Turkish secondary education placement result. Application of Artificial Neural Network, Support Vector Machine, Multiple Regression and Decision Tree were considered and decision Tree C5 gives the best result.

Osmanbegovic and Suljic (2012) collected 12 attributes of 1st year students of Economics a course consisting of demographic variable, score of high school entrance exam and attributes related to their attitude towards studying ,Importance of input variables on prediction was obtained using Chi – square test, One-R test, Information Gain test and Gain – ratio test. It is that GPA, score of entrance exam, study material and average weekly hours devoted to studying are having maximum impact while number of household members distance of residence and gender have least impact.

Shah N. S. (2012) applied various algorithms decisions tree (C45 Random Forest, BF Tree, Rep Tree) Functions (logistic RBF Network) Rule (3 Rip) and Bayes Net, Naive Bayes to categorize (predict) students in 5 categories (Very good, Good, Satisfactory, Below Satisfactory and Fail) Random Forest proven to most accurate classifier .

Kabakchieva (2013) used 14 attributes including personal profile, secondary educational score, entrance exam score, admission year and used l the classifier J48, Bayesian, K-nearest neighbor one R and J Rip. Kabakchieva concludes that J48 performs best with highest overall accuracy, followed by

rule based (J Rip) and the K-NN classifier with Naïve Bayes being least accurate classifier.



PREDICTION OF STUDENTS' EMPLOYABILITY

Today, the reputation of an Institution is judged by its academic success, its ability to retain students and to provide employment for its' students. The term "Employability" still has no precise definition. Lee Harvey (2001) has made an attempt to describe it in many ways like the ability to secure a job, getting a job within a specified time period after graduating, the ability to skill map oneself according to the job need, or the willingness of the student to extend the graduate learning at work. According to V.K. Gokuladas (2010) "employability is determined according to students' success in campus recruitment drives by information technology companies, that is, whether they received an offer of employment". The same definition has been followed in the present study.

Research in employability prediction is in nascent stage. It mostly involves identification of skills or attributes required from the perspective of employers and is obtained from employer through questionnaire and interviews. Mostly statistical methods have been applied and research is more of descriptive than predictive.

Rees (2006) in his report by Higher Education Academy with the Council for Industry and Higher Education (CIHE) in United Kingdom concluded that cognitive, personal, technical, practical, generic abilities along with organizational awareness are most important competencies employers look for. Kayha (2007) has considered the effect of working environment on performance of the employee.

Researchers Chein and Chen (2008) have taken attributes from the curriculum vitae, application and interview of the candidate and applied data mining techniques to predict the performance of a new applicant. The model helps the management in deciding the hiring of the employee.

Mukhtar et al (2009) have concluded that as the paradigm is shifting from product based to service based industry specially in Information Technology, hence the curriculum and method of delivering lecture must evolve. The new challenges are faced by considering a combination of approaches.

General studies for identifying factors that affect the job prospects of a student, have been conducted by Shafie and Nayan (2010). A descriptive study in the research indicates that employers look forward to employees with Personal attributes that include loyalty, commitment, honesty, integrity, enthusiasm, reliability, personal presentation, common sense, positive self-esteem, a sense of humor, motivation, adaptability, a balanced attitude to work and home life and ability to deal with pressure.

Gokuladas (2010) has used correlation and ordinal regression to conclude that a students' non-technical education consisting of reasoning, logical ability and soft skills were stronger predictor of their employability than their technical education consisting of their academic performance. Othman et.al.(2010) identified the reasons for the low employability

rate amongst university graduates . Poor English language competency is found to be a major reason for the low employability. This has been supported by Gokuladas in 2011, that knowledge of GPA and English Language competency are required for the students in software industry to continue with their employment and also the female candidates were better performers than male candidates in campus placement drives.

The quest for finding the skills that make a student employable has been explored in the field of psychology. An attempt to relate the participant's personality to employability has been made by authors, Potgieter & Coetzee, in SA Journal of Industrial Psychology, 2013 using methods of psychology. Career competence, Career resilience, Self efficacy, Proactivity etc. were few personality traits considered.

Yusoff et al. (2012) have used statistical methods to calculate the performance score of entry level engineers based on normalized skill weight. Employers' perspective regarding communication skill, team work, lifelong learning, problem solving skill, competency of employees were found to be important parameters .It has been confirmed that soft skills are significant as compared to technical skills. It also provides an equation based on which an engineer can be selected at entry level.

This work was extended by BangsukJantawan, (2013) who has used real data of graduate students of Maejo University in Thailand for three academic years. The parameters considered are socio economic, academics, and job satisfaction, reason for not working, etc were considered for prediction. Algorithms of Bayesian Network, and Decision Tree, both with five variations, have been used to build the classification model for graduate employability.

Pool, L, Dacre, Qualter, P. (2013) found that emotional self efficacy strongly decides the employability of a student. Further, it also affects the person satisfaction from his/her job.

One of the researchers recently (2014) stated that emotional intelligence, self management and life experiences are important factors for Employability development profile. Another researcher describes employability in strong correlation with competences and depositions.

The attributes identified are categorized under three heads. Some of the factors are pertaining to social integration, some others are based on academic integration and rest are on emotional skills that have been obtained in this study using Emotional skill Assessment Process (ESAP) .These attributes can be used for different purposes with respect to a candidate like predicting employability alone, performance alone or combining both of them.

In traditional education, performance prediction is in matured state with contribution from many researchers. However, there is paucity of research in the field of employability prediction. As both performance and employability of students graduating from an institution decide the market value of the institution, research is required to develop comprehensive models for performance and employability tool and develop a system that will be able to predict both performance and employability.

As both performance and employability are associated with each other all the factors that are used for performance prediction are used for employability. From the literature review, it is clear that most commonly used predictors are socio economic /demographic profile and past academic record of the students. Apart from this, number of hours dedicated to studies, distance of the institution from home, loan, internet facility etc has been considered by the individual researchers in their studies. Thus in general researchers in the field of EDM have focused on academic and social integration of students for performance and employability prediction.

Researchers in psychology have shown that the emotional skills of students are also important factors for performance prediction. However, not much work has been done in the field of EDM to validate the existing knowledge or construct new knowledge about emotional skill being predictor of performance. One reason for not considering emotional skill is lack of authentic data. The present work is an attempt to bridge this gap. The present study constructs authentic primary data that has factors of academic integration and social integration and emotional skills.



ETHODOLOGY

Research Design

The research design adopted is exploratory and experimental due to the nature of the study. An exploratory design was needed to conduct literature survey and identify factors used by researchers in past. Further, in order to develop predictive model, the data set has

to go through an experimental set up under which different algorithms are applied on the data set (sample). Then the algorithm that gives the best result is selected for model building. Thus, experimental design follows the exploratory design in this study.

Sampling Technique and Data Collection

A structured questionnaire is constructed using Google Doc which has been administered to the MCA students from all the institutions affiliated to Guru Gobind Singh University over Internet. The sampling technique used fall under the category of convenience sampling. A data of 1545 students was collected.

Performance prediction includes attributes pertaining to academic integration (AI) (Table 1), social integration (SI) (Table 2), and emotional skills (ES) (Table 3). Whereas, employability prediction includes, all the attributes used for performance prediction along with project work, on-campus placement. The attributes regarding academic integration and social integration were obtained by direct responses to questions through check boxes or radio buttons. Based on the exploratory study done in , factors have been identified categorized under three heads. Some of the factors are pertaining to social integration (SI), some others are based on academic integration (AI) and rest are on emotional skills(ES). These attributes can be used for different purposes with respect to a candidate like predicting employability alone, performance alone or combining both. The emotional skills attributes were calculated based on the responses to a set of questions. For this standard tool of Emotional skill assessment Process (ESAP) has been used .

Table 1 Attributes Pertaining to Social Integration

Attribute	Values
Attribute	Values
Gender	Male, Female
Father’s Education (FE)	Secondary, Senior secondary, Grad, Postgrad
Mother’s Education (ME)	Secondary, Senior secondary, Grad, Postgrad
Father’s Occupation (FO)	Government job, Private job, Business, Others
Mother’s Occupation (MO)	Government job, Private job, Business, Housewife
Family Income (FI)	Low income Group (LIG)(<2 lakh per annum)
Loan	Middle income group (MIG) (2 to 4 lakh per annum)Other
	Yes, No
Early Life (Where a student has spent first 15 years of his life	Metro, City, Village

Table 2 Attributes Pertaining to Academic Integration

Attribute	Values
Medium of Instruction at school level	English, others
Percentage of Marks in Secondary, Senior secondary, Graduation, Firstsem, SecondSem, Third Sem, Fourthsem	Below average BLAVG (<60) Average AVG (60 to less than 70) Above Average ABVG (70 to less than 80) Excellent EXCL(>=80)

Type of Graduation degree (GRADDEGTYPE)	Regular, Distance
Graduation Stream (GRADDEGSTREAM)	Computer Science(CS) Non Computer Science(NCS)
Gap year after Graduation (GAPYEAR)	Yes, No
Hours spent on academic activities ACADEMICHRS)	Insufficient(UNSUF) <2Hrs Sufficient(SUF) 2-4Hrs Optimal > 4 Hrs.
On campus placement (ONCAMPUSPLACE)	Yes, No
Relevant work experience (RELWORKEXP)	Yes. No
Project	Yes, No

Emotional Skill attributes are assessed through Emotional Skill Assessment Process (ESAP) tool developed by (Nelson and Low, 2003), consisting of psychometric questions to judge various parameters.

Following attributes were considered

Next objective was to construct a prediction model for academic performance and employability individually and provide guidelines to management for improving academic performance and employability. Further, the research aimed at developing a decision support system tool for the prediction

Table 3 Emotional Skill Attributes

Attribute	Values
1. Assertion (Ability to communicate effectively, honestly, clearly)	D: Needs to develop the skill(Absent)
2. Empathy (Ability to care for others)	S: Skill is present but need to strengthen(Moderate)
3. Decision making (Ability to take informed decisions)	E: Skill is present and Enhanced(Enhanced)
4. Leadership (Ability to influence others)	
5. Drive Strength (Ability to set a goal and strive for it)	
6. Time Management (Ability to manage time for best productive use)	
7. Self Esteem (Ability to regard himself or herself)	
8. Stress management(Ability to work under stress)	

After the collection of data, it needs to be cleaned and formatted so as to make it suitable for modeling.



RESULT AND ANALYSIS

Prediction of academic performance and employability helps the faculty and administrator take proactive measures like special instructional strategies and specialized training for the students who are at risk of poor performance and unemployment.

Prediction of a variable requires exploring the predicting variables or parameters that might play a role in influencing the prediction of target variable. Thus, the foremost objectives were to explore the factors that affect the academic performance and employability of students individually. The factors were identified based on the exploratory study of recent works done. Emotional skill parameters that were not emphasized in previous studies were also included in our study.

of both the academic performance and the employability. We explain all result and analysis pertaining to academic performance and that for employability prediction objectives.



ACADEMIC PERFORMANCE PREDICTION

It takes almost a year for a student to get integrated in the environment of the institution hence for academic performance prediction, MCA third semester results are considered for prediction in category below average (BAVG), average (AVG), above average (ABVG) and excellent(EXCL). Prediction of categorical variable is a classification problem and there is no specific algorithm that will prove to be best in any given situation. Thus it was required to apply various algorithms on the research work at hand and select the best suited after comparison of results obtained. Multilayer Perceptron (MLP), Sequential Minimal Optimization (SMO), Naïve Bayes, Random forest, Random Tree and J48 were selected for the experimentation based on the exploratory study. Results obtained were compared on accuracy, F-measure, Area under ROC curve and time taken by classifier to develop the model.



ANALYSIS OF ACADEMIC RESULTS

This section presents the glimpse of the comparative results obtained in the above section. The classifiers' performances for prediction of academic results are compared using parameters F Measure, ROC area, Accuracy and Time taken to build the model. The comparative results are presented in Table 4 Figure 1 and Figure 2 shows the graphical view of the obtained results.

Analyzing the classifiers, we find that Naive Bayes and Random tree, inspite of their instant model building cannot be selected due to their low performance on other parameters. Naive Bayes has high AUC but lower accuracy . Similarly, Multilayer Perceptron, though having higher accuracy than Naive Bayes and Random tree has less accuracy than J48 and Random Forest. Further model classification time is very high. Accuracy and F measure of Random Forest, SMO, and J48 are comparable but Random forest, and SMO take substantial time to build the model.

Table 4 Comparative Analysis of Classifiers for Academic performance prediction

Algorithm	F Measure	Accuracy (%)	ROC Area	Model Building time(Seconds)
SMO	0.944	94.41	0.973	1.2
Naïve Bayes	0.909	90.75	98.2	0.02
Multilayer Perceptron	0.955	95.47	97.6	50.7
Random Forest	0.946	94.90	0.97	0.12
Random Tree	0.926	92.30	0.962	0.01
J 48	0.945	94.48	0.973	0.01

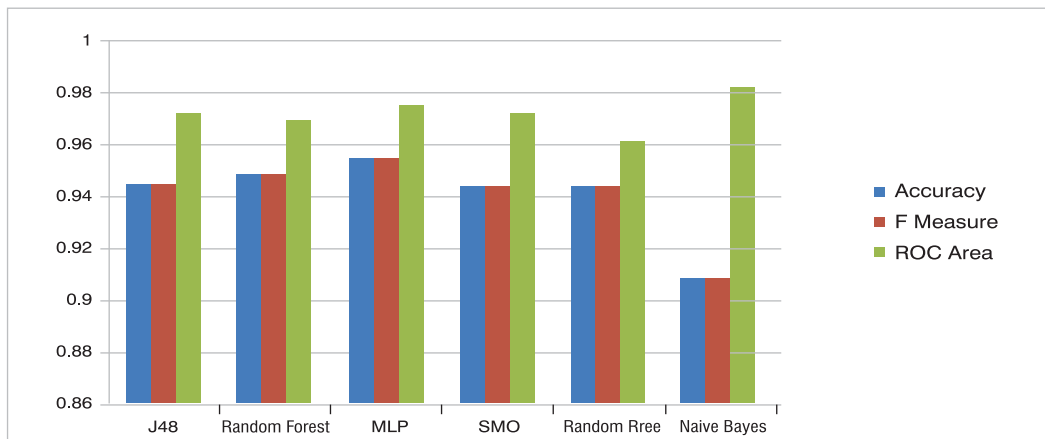


Figure 1: Classifiers comparison for academic performance prediction

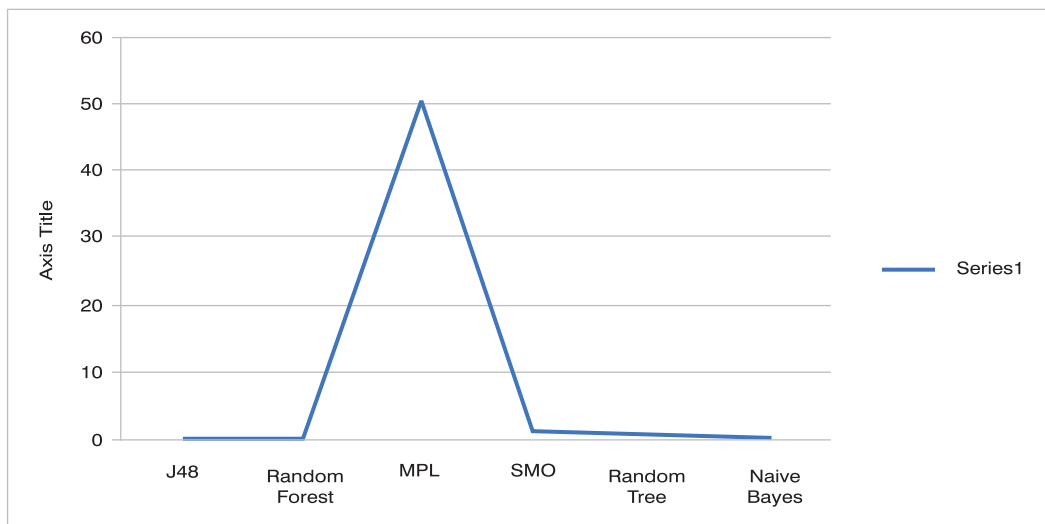


Figure 2: Comparison of time taken for model building for academic performance

Apart from this, the tree generated by J48 can be converted to rules which are easily interpret able. Thus J48 is selected for building the predictive model for academic performance due to its high accuracy, F-Measure, ROC area, low model building time, and easy interpretation.



CADEMIC PERFORMANCE PREDICTION MODEL

The analysis performed in the above section motivated us to use J48 algorithm to obtain the further analysis. Thus, we have used this algorithm to generate the tree, as shown in Figure 3.

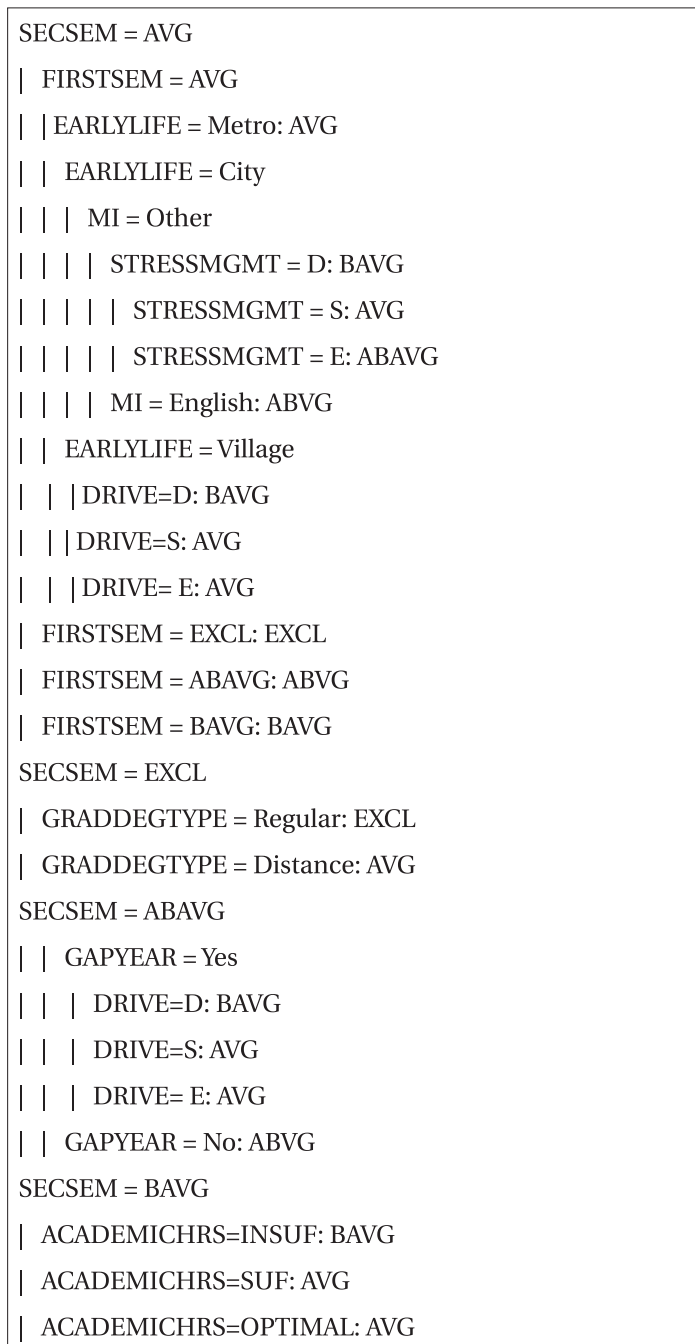


Figure 3: Tree generated by J48 for academic prediction

These set of rules then works as model for predicting the Academic performance. Figure 4 presented the model obtained, which consist of rules to classify the data and predict the students' performance.

1. IF (SECSEM = AVG) AND (FIRSTSEM = AVG) AND (EARLYLIFE = Metro) :AVG
2. IF (SECSEM = AVG) AND (FIRSTSEM = AVG) AND (EARLYLIFE = City) AND (MI = Other) AND (STRESSMGMT = D): BAVG
3. IF (SECSEM = AVG) AND (FIRSTSEM = AVG) AND (EARLYLIFE = City) AND (MI = Other) AND (STRESSMGMT = S): AVG
4. IF (SECSEM = AVG) AND (FIRSTSEM = AVG) AND (EARLYLIFE = City) AND (MI = Other) AND (STRESSMGMT = E): ABAVG
5. IF (SECSEM = AVG) AND (FIRSTSEM = AVG) AND (EARLYLIFE = City) AND (MI = English): ABVG
6. IF (SECSEM = AVG) AND (FIRSTSEM = AVG) AND (EARLYLIFE = Village) AND (DRIVE=D): BAVG
7. IF (SECSEM = AVG) AND (FIRSTSEM = AVG) AND (EARLYLIFE = village) AND (DRIVE=S): AVG
8. IF (SECSEM = AVG) AND (FIRSTSEM = AVG) AND (EARLYLIFE = Village) AND (DRIVE= E): AVG
9. IF (SECSEM=AVG) AND (FIRSTSEM= EXCL): EXCL
10. IF (SECSEM=AVG) AND (FIRSTSEM= ABVG): ABVG
11. IF (SECSEM=AVG) AND (FIRSTSEM= BAVG): BAVG
12. IF (SECSEM = EXCL) AND (GRADDEGTYPE = Regular): EXCL
13. IF (SECSEM = EXCL) AND (GRADDEGTYPE = Distance):AVG
14. IF (SECSEM = ABVG) AND (GAPYEAR =No):ABVG
15. IF (SECSEM = ABAVG) AND (GAPYEAR = Yes) AND (DRIVE=D): BAVG
16. IF (SECSEM = ABAVG) AND (GAPYEAR = Yes) AND (DRIVE=S): AND (DRIVE= E) : AVG
17. IF (SECSEM=BAVG) AND (ACADEMICHRS=INSUF):BAVG
18. IF (SECSEM=BAVG) AND ((ACADEMICHRS=SUF) AND (ACADEMICHRS= OPTIMAL)) :AVG

Figure 4: Model developed for academic prediction



DISCUSSION OF RESULTS FOR ACADEMIC PERFORMANCE

For easy interpretation the tree was converted into a set of rules as shown in Figure 4. The research aimed at developing a predictive model for the academic performance of MCA students after they have spent one year in college (i.e. prediction of Third Semester). The rule set obtained from the tree clearly illustrates the attributes which are actually important for deciding the academic performance of the students. They are,

(i) Second Semester Result (ii) First Semester result (iii) Early life (iv) Medium of Instruction (v) Emotional skills stress management and drive (vi) Graduate degree type (vii) Gap year (viii) Academic hours

The following conclusions can be derived from above rule set

- A student with average performance in first and second semester gives average performance in third semester in spite of his initial upbringing of metro. [Rule 1].
- Stress management skill is imperative for all the students, however, it becomes pertinent especially for a student with upbringing of a city, with vernacular medium of instruction in school and average performance in first and second semester. It is observed that academic performance of third semester is above average for a student with enhanced stress management skill (E), average if Stress management skill is strengthened (S) and below average if it he has poor stress management skill (D). Also under the similar circumstances an English medium student performance will be above average [From Rule 2,3,4,5].
- A student with average performance in first and second semester with the early upbringing of a village and poor drive (D) results in below average academic performance in third semester where as if Drive is strengthened (S) or is enhanced (E) the result can improve to average [Rules 6,7,8].
- It is observed that graduation degree from a regular course is important. A student who has completed his graduation in a regular course and gives excellent performance in second semester also gives excellent performance in third semester but above average performance if the graduate degree is from distance learning [Rule 12,13]. A regular education gives a better edge to the students in comparison to the one from distance learning course
- It is observed that if a student's first semester performance is other than average then with second semester performance as average, the third semester performance depends on first semester result. The third semester result is above average, below average and excellent if first semester performance is above average, below average and excellent respectively. [Rule 9,10,11].
- A student with above average performance in second semester with no gap year also gives above average performance [Rule 14].
- A student with above average performance in second semester having gap year after graduation degree is predicted to perform below average if his drive is poor or absent and average performance if his drive is present but needs to be strengthened (S) or enhanced (E) [Rule

15,16,17].

- A student with below average performance in second semester can improve the result to average by putting in sufficient or optimal number of hours in studies otherwise for insufficient academic hours yield below average result [Rules 18,19,20].



GUIDELINES TO THE MANAGEMENT FOR IMPROVING ACADEMIC PERFORMANCE

From the academic prediction model it was observed that first and second semester results, early life, medium of instruction, academic hours, gap year and graduate degree type and emotional skills like stress management and drive are attributes that affect the third semester result. From the above observations it is suggested that the management of any Institution offering MCA needs to comprehend the student's Socio-economic condition, past academic records as well as emotional skill to predict academic result.

First Semester result plays an important role in students' performance after one year hence care should be taken in building the foundation and basic concepts in the first year itself. Further, below average performer in second semester must be given tutorial classes so as to increase their academic hours and improve the results. Students from non-English medium background should be given extra support for enhancing their proficiency thus enabling them to cope up with the stress arising due to education in English, where they face difficulty in expressing themselves either in written exam or viva. Students from rural background need to be counseled to improve the drive for the higher education and competing with others. Similarly, students with distance learning course in graduation or those who have a gap year after graduation need special attention in the form of extra classes, extra assignment and motivation to improve their performance. Apart from faculty who help the students with academics appointing a counselor can bring improvement to students' emotional skill and thus help give a thrust to academic performance.



EMPLOYABILITY PREDICTION

Further, next objective is to design the predictive model for the student's employability and thereby suggest the guidelines for enhancing the employability.

The employability is defined as the capability to get on-campus placement in terms of YES and NO.

Analysis of Employability Results

The classifiers are compared using parameters F-Measure, ROC area, Accuracy and Time taken to build the model. The results obtained are presented in Table 5 for comparison. The graphical views of obtained results are shown in Figure 5 and Figure 6.

Table 5: Comparative Analysis of Classifiers for Employability prediction

Algorithm	F Measure	Accuracy (%)	ROC Area	Model Building time(Seconds)
SMO	0.839	83.9	0.831	1.88
Naïve Bayes	0.74	73.85	0.818	0.02
Multilayer Perceptron	0.932	92.08	0.971	47.59
Random Tree	0.876	87.63	0.887	0.0
Random Forest	0.886	88.6	0.887	0.17
J 48	0.883	88.3	0.884	0.01

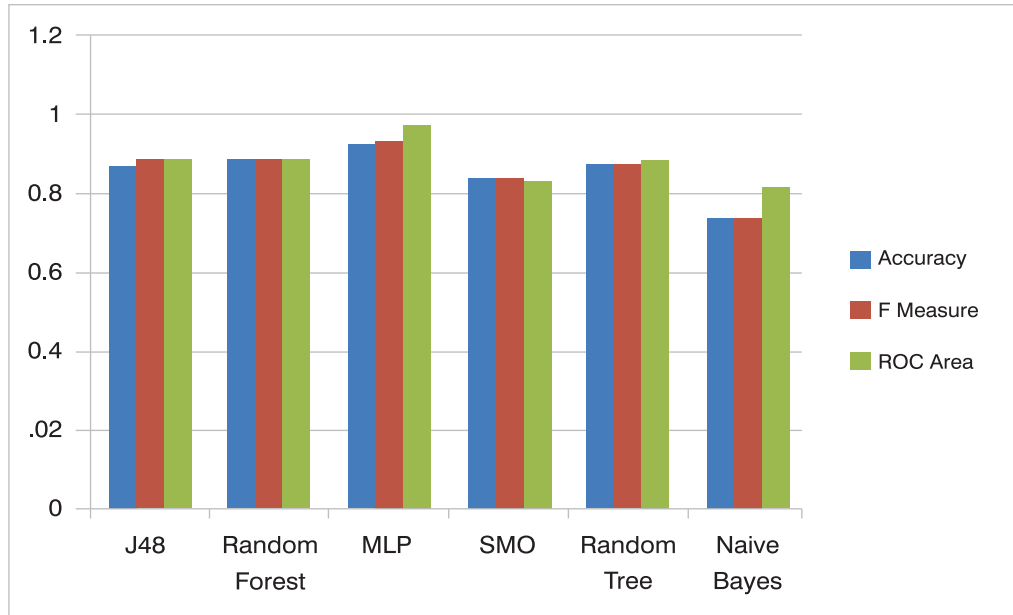


Figure 5: Classifiers comparison Employability prediction

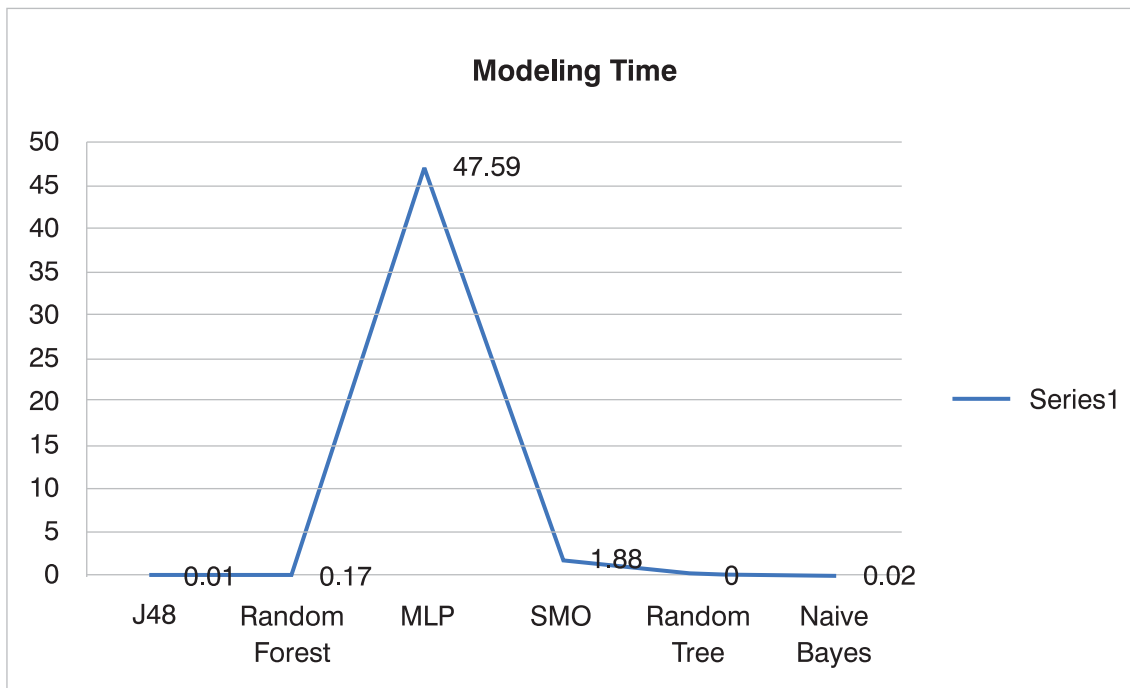


Figure 6: Classifiers comparison of model building time for Employability prediction

Analyzing the classifiers, and by the similar arguments as in academic performance we find that J48 is most suited for the problem of employability prediction.



EMPLOYABILITY PREDICTION MODEL

This section discusses the tree obtained by applying J48 algorithm for predicting employability of the students as shown in Figure 7.

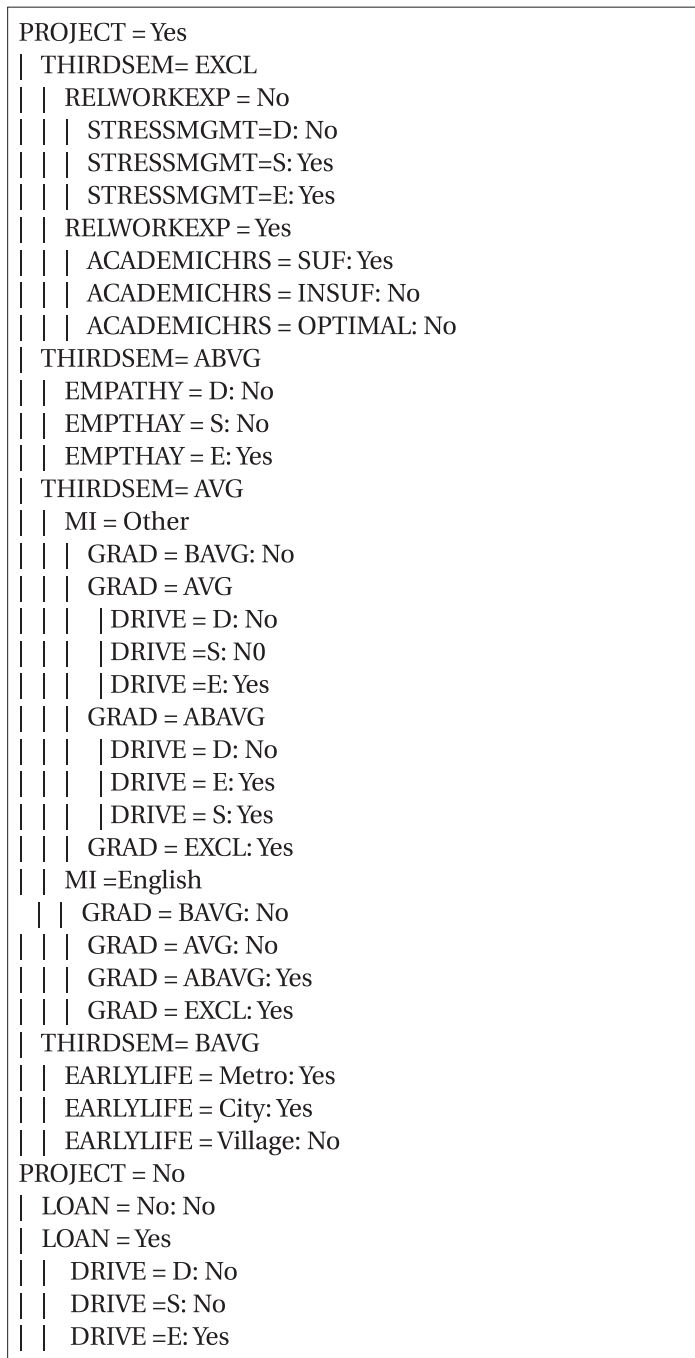


Figure 7: J48 Tree for Employability Prediction

Rules derived from the above tree are used as model to predict the employability of the students are presented in Figure 4.8.

- 1 IF (PROJECT = YES) AND (THIRDSEM = EXCL) AND (RELWORKEXP = NO) AND (STRESSMGMT=D): NO
- 2 IF (PROJECT = YES) AND (THIRDSEM = EXCL) AND (RELWORKEXP = NO) AND (STRESSMGMT=E or STRESSMGMT=S): YES
- 3 IF (PROJECT = YES) AND (THIRDSEM = EXCL) AND (RELWORKEXP = YES) AND (ACADEMICHRS=SUFF): YES
- 4 IF (PROJECT = YES) AND (THIRDSEM = EXCL) AND (RELWORKEXP = YES) AND (ACADEMICHRS=INSUFF): NO
- 5 IF (PROJECT = YES) AND (THIRDSEM = EXCL) AND (RELWORKEXP = YES) AND (ACADEMICHRS=OPTIMAL): NO
- 6 IF (PROJECT = YES) AND (THIRDSEM =ABVG) AND (EMPATHY = D) OR (EMPATHY = S): NO
- 7 IF (PROJECT = YES) AND (THIRDSEM =ABVG) AND (EMPATHY = E): YES
8. IF (PROJECT = YES) AND (THIRDSEM= AVG) AND (MI = OTHER) AND (GRAD = BAVG): NO
9. IF (PROJECT = YES) AND (THIRDSEM= AVG) AND (MI =OTHER) AND (GRAD = AVG) AND (DRIVE = D) OR(DRIVE=S): NO
10. IF (PROJECT = YES) AND (THIRDSEM= AVG) AND (MI = OTHER) AND (GRAD = AVG) AND (DRIVE = E): YES
11. IF (PROJECT = YES) AND (THIRDSEM= AVG) AND (MI = OTHER) AND (GARD =ABVG) AND (DRIVE = D): NO
12. IF (PROJECT = YES) AND (THIRDSEM= AVG) AND (MI =OTHER) AND (GARD =ABVG) AND (DRIVE = E) OR (DRIVE = S): YES
13. IF (PROJECT = YES) AND (THIRDSEM= AVG) AND (MI = OTHER) AND (GRAD =EXCL): YES
14. IF (PROJECT = YES) AND (THIRDSEM= AVG) AND (MI = ENGLISH) AND (GRAD = BAVG) OR (GRAD = AVG): NO
15. IF (PROJECT = YES) AND (THIRDSEM= AVG) AND (MI = ENGLISH) AND (GRAD =ABVG) OR (GRAD = EXCL): YES
16. IF (PROJECT = YES) AND (THIRDSEM =BAVG) AND (EARLYLIFE = METRO OR EARLYLIFE = CITY): YES
17. IF (PROJECT = YES) AND (THIRDSEM =BAVG) AND (EARLYLIFE = VILLAGE): NO
18. IF (PROJECT = NO) AND (LOAN = NO):NO
19. IF (PROJECT = NO) AND (LOAN = YES) AND (DRIVE = S OR DRIVE = D): NO
20. IF (PROJECT = NO) AND (LOAN = YES) AND (DRIVE = E): YES

Figure 8: Model developed for Employability Prediction



DISCUSSION OF RESULTS FOR EMPLOYABILITY

The rule set obtained from the tree it is obvious that out of all the attributes considered, following attributes are important for deciding the employability of the student semester(i) Project (ii) Third semester result (iii) Graduation result (iv) Relevant work experience (v) Emotional skills like stress management.

Following conclusions can be drawn

- A student who has undertaken project ,with excellent third semester result ,with no relevant work experience is not employable if he lacks in stress management skill i.e. Stress management skill is D. However, under the same circumstances, if he possesses strengthened (S) or enhanced stress management skill (E), his employability is assured [Rule 1,2]
- A student with excellent third semester result and project having relevant work experience is employable if he puts in sufficient academic hours in studies whereas he is not employable if his academic hours are insufficient or optimal. [Rule 3,4,5]
- A student who has undertaken project and above average performance in third semester result is employable only if he posses empathy that is enhanced (E) otherwise not. [Rule 6,7]
- For a student who has undertaken project , has average result in thrid semester, with vernacular medium of instruction Graduate result and drive becomes important. A below average performance in graduation makes him unemployable. Further, with an average performance in graduation he is employable only if his/her drive is fully enhanced (E) otherwise not. Under the similar circumstances but having above average performance the students is employable if Drive is enhanced (E) or strengthen (S) otherwise not. [Rule 8,9,10,11,12,13]
- A student who has undertaken project, with average performance in third semester, is unemployable in spite of having medium of instruction as English if his performance in graduation degree is below average or average. However, an above average or excellent result in graduation under these circumstances makes him employable. [Rule 14,15]
- A student who has undertaken project and has below average third semester performance can get employment if he/ she has upbringing of a city or metro but an initial upbringing of village makes him unemployable. [Rule 16,17]
- If the student who has not undertaken a project has and has not taken educational loan is not employable. A student who has taken loan and has enhanced drive(E) is employable otherwise not. [Rule 18,19,20]



GUIDELINES TO THE MANAGEMENT FOR IMPROVING EMPLOYABILITY

Investigation of employability models divulges that project, third semester result, performance in graduate degree, loan taken by the students along with a new emotional skill empathy, stress management and drive plays a pivotal role in employability.

In order to improve the employability the faculty members must ensure that students do live projects and help them learn the subject through practical implementation. Students with average and below average performance in graduation needs to be monitored and trained to improve their academic performance as well as drive to excel in this competitive world. Personality development and Communication classes for the students of non-English medium back ground students or those students with upbringing of a village must be held regularly to groom them and enhance their confidence. Students should be encouraged to form groups and solve each other problems thus enhancing the empathy in them and making them team player. A counselor can be appointed to develop and make student work upon their emotional skills like stress management, drive and empathy.

A decision support system EMPPT (Employability and Performance Prediction Tool), based on the model discussed above, is developed and validated. The description of this system is described in the next section.



MODEL VALIDATION THROUGH DECISION SUPPORT SYSTEM EMPPT

As stated in Chapter 1, after model building for academic performance and employability prediction, the next objective is to validate them through Decision support system (DSS). DSS will help the senior level management in taking well informed decision. It is a computerized system that aids in decision making by highlighting certain facts which may not be visible otherwise.

In this section we present a decision support system prototype Employability and Performance Prediction Tool (EMPTT). EMPTT uses model derived from classification algorithm J48 for prediction of Academic performance and employability performance.



ARCHITECTURE OF EMPPT

Figure 9 shows the architecture of EMPTT, and its interaction with the model building process.

The software captures students attribute (I) after the declaration of second semester result to predict third semester result (ii) in fifth semester to predict the employability of the students

The DSS consists of following components:

- Login Component
- Add student Component

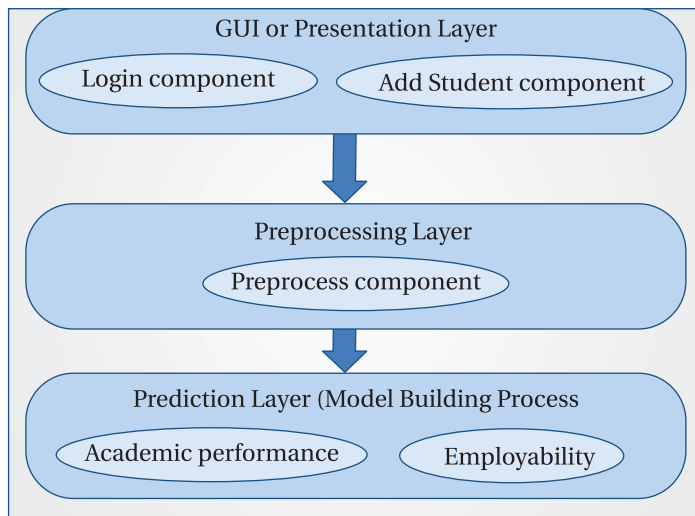


Figure 9: Architecture of EMPPT

- Preprocess Component
- Start Prediction Component
- Academic Performance Prediction
- Employability Prediction

Login Component

This component gives login facility to two users (I) Student (ii) Administrator. Student can only fill in his/her data while admin will be responsible for loading the complete class data, do the preprocessing and then predict the academic performance and employability of whole class. The login component provides facility to register first and then login.

Add Student Component

This component allows addition of records of students. It adds data through a form and hence data entered is in raw form.

File Upload

Once the data is entered in excel file it is uploaded and preprocessed. During preprocessing all numeric ranges are converted to categorical value and emotional traits are calculated based on the questionnaire filled by the students. Once the file is processed the message alerts that file has been processed and we one can proceed for prediction.

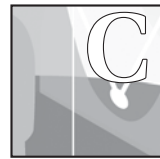
Academic Prediction

This component predicts the third semester result based on the parameters uploaded.

Once academic prediction is selected and Start prediction button is pressed the software gives the predicted value of third semester result in terms of average, above average, good, and Excellent. The results are verified on a new set of data where the third semester values results are known to verify whether the software is working properly or not The result in figure shows a comparison of actual value and predicted value.

Employability Prediction

This component predicts that whether students will get on campus placement or not which starts in Vth semester. For verification a new data set (different from which the model was trained and tested) was applied to the software and result of actual value and predicted value can be seen to be same.



CONCLUSION

The key research objective was to first identify the factors that affect the academic performance and employability. In present work, MCA (Masters in Computer Applications) which is a professional post graduate degree course has been chosen. Exhaustive literature survey conducted revealed that many prediction models with various approaches have been developed that predict the academic performance of the students based on primary attributes related to their academic qualification, social background etc. Not much emphasis has been given to attributes pertaining to emotional skills which could also be factors concerning the performance and employability of a student majorly due to lack of authentic data. Thus, as first step the sample set was collected. A sample of 1504 students of MCA students of various colleges affiliated to Guru Gobind Singh Indra Prastha University was collected through a structured questionnaire, which included attributes that have been categorized into academic integration(all declared past academic results, medium of instruction, no. of hours devoted to study, project work done ,type of graduation degree, graduation stream, gap year, relevant work experience), social integration (parent's education, parent's occupation, family income , loan taken by the student, environment where student has spent his early life) and emotional skills (assertion, empathy, decision making, leadership, drive strength, time management, self esteem, and stress management each measured as D i.e need to develop, S i.e needs to strength and E i.e enhanced). Data is collected thorough google forms. Emotional Skill parameters are assessed through Emotional Skill Assessment Process (ESAP) tool consisting of psychometric questions to judge various parameters. The collected dataset was preprocessed to make it appropriate for the data mining tasks.

The next objectives were to develop prediction model individually for academic performance and employability. Third semester result was selected for prediction based on the logic that a student will take approximately a year(first two semester) to adjust to institution environment. The prediction model was created by applying various classification algorithms available in WEKA data mining tool. The algorithms were compared on various parameters of accuracy, model building time, interpretability, F-measure and ROC Area. It was found that J48 algorithm which is implementation of pruned C4.5 Decision Tree algorithm of WEKA is most suitable for the both academic performance and employability prediction. From the predictions though decision tree, rules were derived that best predicts the academic performance and the employability. The findings were validated though a Decision Support System (DSS) tool called Employability and Performance Prediction Tool (EMPPT) which has been developed based on the predicted model in the last step.

Out of all past academic results first and second semester is key influencer of third semester result. It is expected also, as the programming subjects of second semester form the foundation of programming subjects of third semester. Thus consistently good academic performance is clearly a good indication of good performance in third semester too. Stress management becomes important for an average performer with upbringing of city but vernacular medium of instruction at school, where as a student with early life spent at village and average performance in first and second semester of MCA his Drive strength plays an important role in deciding his third semester result. A student with graduation from regular degree course has a better performance than distance learning graduation degree. Gap year after graduation results in poor performance if the Drive is absent. Sufficient and optimal no. of hours put in study can improve the result.

Some interesting findings related to employability are observed. Third semester of MCA is highly computer programming oriented semester and after that students are supposed to apply it in their projects. The finding reveals that student with an excellent third semester result, relevant work experience and project is more employable if he devotes sufficient number of hours to studies (4 to 8 hrs a day) than the student who devotes optimal number of hours (more than 8 hrs) .This indicates that long hours of studies could bring in saturation adversely affect the performance.

Empathy seems to play a substantial role in employability of the students as student with project and above average third semester result are found successful if they have enhanced empathy. The reason could be that such students do take care of other class mates by solving their difficulty, explaining them concepts that improve their own communication skill, expression power and ability to work in a team etc. much sought after qualities by the recruiters.

Further graduation result, medium of instruction in school (that may decide English language fluency) and drive strength of the student play a major role in on campus placement. A below average graduation degree along with vernacular medium schooling puts a student at risk of poor performance in on campus placement even if s/he is having average third semester result and project however , if his graduation result is above average and drive is strengthened or enhanced the students will be employed. If the student has taken loan and has enhanced drive then in spite of not having a project the student has chances of getting selected in on campus placement.

In case of employability stress management becomes very important. A student with project but below average third Semester result is more employable if his early life is spent in

City or Metro than in village as students from Metro and city have better opportunities of grooming their personality.

This research has identified factors affecting academic performance and employability and the effect of some of the emotional skill parameters on academic performance and employability has been established where others have not shown much effect as per expectation.

The Decision support system developed is capable of predicting academic performance and employability with acceptable accuracy.

The general guidelines for the institutions to enhance performance and employability are to make the foundation very strong in first and second semester and give it practical orientation so that students are capable of doing projects. Monitoring of students with vernacular medium and village back ground is essential. They need more grooming than students brought up in city or metro and those who have studied in English medium schools. Thus special communication classes can be arranged separately for those students. A counselor can be appointed for developing emotional quotient and helping student learn stress management, develop empathy and use their drive strength in positive direction.



UTUREWORK

Future work will include the students of B. Sc. and B. E as well and will try to find out if there is a preference being given in the companies to B.Sc. / B. E. over MCA. Further employability has been defined as the students' ability to get employment during on Campus drives conducted in Vth Semester and does not take into account the pay package offered and the rating of the company in which he or she is placed. These aspects will be considered in our future work. Although academic performance is one of the criteria of student's consistent efforts and perseverance, most of the companies are now concentrating upon employee's logical reasoning, quantitative aptitude, communication skill, etc.. Average students are also capable of getting good employment based on other skills and virtues. These parameters will be included in further research so as to infer these academically.

The research reveals great potential of data mining in education sector. More work is needed to establish it as a customized guiding tool for students. In conclusion, this study can motivate and help institutions to perform data mining tasks for students' data regularly to find out interesting results and patterns which can help both the university as well as the students in many ways.

REFERENCES

- i. Affendey, L. S., Paris, I. H. M., Mustapha, N., Sulaiman, M. N., & Muda, Z. (2010). Ranking of influencing factors in predicting students' academic performance. *Information Technology Journal*, 9(4), 832-837.
- ii. Bangsuk, J., Cheng-Fa, T. (2013), The Application of Data Mining to Build classification Model for Predicting Graduate Employment, (IJCSIS) International Journal of Computer Science and Information Security, Vol. 11, No. 10,
- iii. Bharadwaj, B.K. & Pal, S. (2011). Mining Educational Data to Analyse Students' Performance, International Journal of Advanced Computer Science and Applications, 2(6):63-69
- iv. Cheewaprabkhit, P. (2013). Study of Factors Analysis Affecting Academic Achievement of Undergraduate Students in International Program. In *Proceedings of the International Multi-Conference of Engineers and Computer Scientists* (Vol. 1, pp. 13-15).

- v. Chien, C. F., & Chen, L. F. (2008). Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry. *Expert Systems with applications*, 34(1), 280-290.
- vi. Cortez, P. and Silva, P. (2008) Using data mining to predict secondary school student performance, Proc. 5th Annual Future Business Technology Conference, Porto, pp. 5-12.
- vii. Gokuladas, V. K. (2010). Technical and non technical education and the employability of engineering graduates: an Indian case study. *International Journal of Training and Development*, 14(2), 130-143.
- viii. Gokuladas, V. K. (2011). Predictors of employability of engineering graduates in campus recruitment drives of Indian software services companies. *International Journal of Selection and Assessment*, 19(3), 313-319.
- ix. Guruler, H., Istanbulu, A., & Karahasan, M. (2010). A new student performance analysing system using knowledge discovery in higher educational databases. *Computers & Education*, 55(1), 247-254.
- x. Han, J. (2012). Micheline KAMBER a Jian PEI. *Data mining: concepts and techniques*.
- xi. Harvey, L. (2001). Defining and measuring employability. *Quality in higher education*, 7(2), 97-109.
- xii. Huang, S. (2011) Predictive Modeling and Analysis of Student Academic Performance in an Engineering Dynamics Course.
- xiii. Jantawan, B., & Tsai, C. F. (2013). The application of data mining to build classification model for predicting graduate employment. *arXiv preprint arXiv:1312.7123*.
- xiv. Kabakchieva, D. (2013). Predicting student performance by using data mining methods for classification. *Cybernetics and Information Technologies*, 13(1), 61-72.
- xv. Kahya, E. (2007). The effects of job characteristics and working conditions on job performance. *International journal of industrial ergonomics*, 37(6), 515-523.
- xvi. Kardan, S., & Conati, C. (2010, June). A framework for capturing distinguishing user interaction behaviors in novel interfaces. In *Educational Data Mining 2011*.
- xvii. Kovačević, Z. J., & Green, J. S. (2010). Predictive working tool for early identification of 'at risk' students.
- xviii. Kovacic, Z. (2012) Early prediction of student success: Mining students' enrollment data, Proc. Informing Science & IT Education Conference (InSITE), pp. 647-665.
- xix. Nelson, D. B., LOW, G. R., & Santos, M. D. D. (2004). Professional Manual.
- xx. Nghe, N. T., Janecek, P., & Haddawy, P. (2007, October). A comparative analysis of techniques for predicting academic performance. In *2007 37th Annual Frontiers In Education Conference-Global Engineering: Knowledge Without Borders, Opportunities Without Passports* (pp. T2G-7). IEEE.
- xxi. Ogor, E. N. (2007, September). Student academic performance monitoring and evaluation using data mining techniques. In *Electronics, Robotics and Automotive Mechanics Conference (CERMA 2007)* (pp. 354-359). IEEE.
- xxii. Oladokun, V. O., Adebajo, A. T., & Charles-Owaba, O. E. (2008). Predicting students' academic performance using artificial neural network: A case study of an engineering course. *The Pacific Journal of Science and Technology*, 9(1), 72-79.
- xxiii. Osmanbegović, E., & Suljić, M. (2012). Data mining approach for predicting student performance. *Economic Review*, 10(1).
- xxiv. Othman, Z., Musa, F., Mokhtar, N. H., Ya'acob, A., Latiff, R. A., & Hussin, H. (2010). Investigating University Graduates' English Language Competency towards Employability: A Proposed Research Method. *International Journal of Learning*, 17(7).
- xxv. Paris, I. H. M., Affendey, L. S., & Mustapha, N. (2010). Improving academic performance prediction using voting technique in data mining. *World Academy of Science, Engineering and Technology*, 62, 820-823.
- xxvi. Pool, L., Dacre, Qualter, P. (2013), Emotional Self Efficacy, graduate Employability and Career Satisfaction: Testing the Associations
- xxvii. Potgieter, I., & Coetzee, M. (2013). Employability attributes and personality preferences of postgraduate business management students. *SA Journal of Industrial Psychology*, 39(1), 01-10.
- xxviii. Ramaswami, M., & Bhaskaran, R. (2009). A study on feature selection techniques in educational data mining. *arXiv preprint arXiv:0912.3924*.
- xxix. Ramaswami, M. and Bhaskaran R., (2010), A CHAID based performance prediction model in educational data mining, " International Journal of Computer Science Issues, 7(1) 10-18
- xxx. Rees, C., Forbes, P., & Kubler, B. (2006). Student employability profiles. *The Higher Education Academy, York*.
- xxxi. Romero, C. and S. Ventura (2007). Educational data mining: A survey from 1995 to 2005. *Expert systems with applications* 33(1): 135-146.
- xxxii. Sembiring, Sajadin & Zarlis, M. & Hartama, Dedy & Ramliana, S. & Wani, E.. (2011). Prediction of student academic performance by an application of data mining techniques. 6. 110-114.
- xxxiii. Sen, Baha & Uçar, Emine & Delen, Dursun. (2012). Predicting and analyzing secondary education placement-test scores: A data mining approach. *Expert Systems with Applications: An International Journal*. 39. 9468-9476.
- xxxiv. Sapaat, M. A., Mustapha, A., Ahmad, J., Chamili, K., & Muhamad, R. (2011). A data mining approach to construct graduates employability model in Malaysia. *International Journal of New Computer Architectures and their Applications (IJNCAA)*, 1(4), 1086-1098.
- xxxv. Shafie, L. A., & Nayan, S. (2010). Employability awareness among Malaysian undergraduates. *International Journal of Business and Management*, 5(8), 119-123.
- xxxvi. Shah, N. S. (2012). Predicting factors that affect students' academic performance by using data mining techniques. *Pakistan business review*. 13(4) 631-668
- xxxvii. Susila M., Jaafar M, Nagaraj S (2013), . Does Reputation Matter? Case Study of Undergraduate Choice at a Premier University. *Asia-Pacific Edu Res* 23, 451-462 (2014)
- xxxviii. Yusoff, Y. M., Omar, M. Z., Zaharim, A., Mohamed, A., & Muhamad, N. (2012). Employability skills performance score for fresh engineering graduates in Malaysian industry. *Asian Social Science*, 8(16), 140.