Students' Performance and Employability Prediction through Data Mining: A Survey

Article *in* Indian Journal of Science and Technology · June 2017

DOI: 10.17485/ijst/2017/v10i24/110791

Dharminder Kumar
Guru Jambheshwar University of Science & Technology
41 PUBLICATIONS **424** CITATIONS
SEE PROFILE

Sangeeta Gupta
Management Education & Research Institute
12 PUBLICATIONS **173** CITATIONS
SEE PROFILE

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 4 (2016) pp 2275-2282
© Research India Publications. http://www.ripublication.com

Students' Employability Prediction Model through Data Mining

Tripti Mishra
Department of Computer Science, Mewar University, Rajasthan, India.

Dharminder Kumar
Department of Computer Science, G. J. University, Hisar, Haryana, India.

Sangeeta Gupta
Department of Management, Guru Nanak Institute of Management, Delhi, India.

Abstract

The students' employability is a major concern for the institutions offering higher education and a method for early prediction of employability of the students is always desirable to take timely action. The paper uses various classification techniques of data mining, like Bayesian methods, Multilayer Perceptrons and Sequential Minimal Optimization (SMO), Ensemble Methods and Decision Trees, to predict the employability of Master of Computer Applications (MCA) students and find the algorithm which is best suited for this problem. For this purpose, a data set is developed with the traditional parameters like socioeconomic conditions, academic performance and some additional emotional skill parameters. A comparative analysis concludes that J48(a pruned C4. 5 decision tree) is most suitable for employability prediction with 70. 19% accuracy, easy interpretation and model building time(0. 02Sec) less than Random Forest, which has slightly better prediction accuracy (71. 30%), higher building time(0. 11) and difficult interpretation. Further, Empathy, Drive and Stress Management abilities are found to be the major emotional parameters that affect employability.

Keywords: Classification, Data Mining, Employability Prediction, Emotional Skills Modeling.

Introduction

Higher education plays a key role in strengthening a nation's economy as it is an industry in itself and it supports the rest of the industry by providing a trained workforce. Earlier, the major concern for these Institutions were the decrease in the student success rate, decrease in retention of students, increase in students moving to other competitive institution and lack of counseling to students in subject selection. However, with education becoming more and more employment oriented, employment of students, graduating from any Institution has become a major factor in building the reputation of the Institution and hence a major concern.

Educational institutions generate and collect huge amount of data. This may include students' academic records, their personal profile, observations of their behavior, their web log activities and also faculty profile. This large data set is basically a storehouse of information and must be explored to have a strategic edge among the Educational Organizations.

Predicting student employability can help identify the students who are at risk of unemployment and thus management can intervene timely and take essential steps to train the students to improve their performance.

Before making predictions, it is essential to find out an algorithm that is best suited for the problem, which requires comparison of algorithms based on certain metrics. In this paper, we have applied different classification algorithms and compared them based on Area under Receiver Operating Characteristic (ROC) curve, and F Measure, apart from accuracy, TP rate, Precision and recall as considered byresearchers earlier.

In past, researchers have tried to establish links between academic performance, socioeconomic conditions, job skills of the students and employability, however, mostly by deploying statistical method. The unique contribution of this paper is that apart from above stated parameters, it also explores the link between and emotional skills like assertion, empathy, decision making, leadership, drive, stress management to predict employability using data mining techniques. The emotional skills like assertion, leadership, management, empathy, decision making have been included using standard ESAP [9]. ESAP stands for "Emotional Skills Assessment Process" which is a comprehensive technique to evaluate a student's Emotional Quotient (EQ). Further details of this technique are provided under Experimental Setting section.

In this paper MCA stands for "Master of Computer Applications". It is an advanced six semester degree course in applied Computer Science offered by several universities across India.

The 'employability' of a student has been defined as whether the student was capable of getting an on campus placement offered in V semester.

The paper is organized as follows: The Introduction section gives a brief idea of the subject followed by the related work in Literature Review section. Next, Experimental Settings section discusses the data collection, data pre-processing, and data set development. This is followed by the Result and Discussion which compares different classifier's ability in predicting employability. Model developed section discusses prediction model obtained, followed by Conclusions and Future work.

Related Work

Application of data mining techniques to educational data has provided support to the institutions in many activities. Most recent survey paper by Pena-Ayala [26] has observed that most of the researches fall in one of the six categories: students' behavior modeling, students' performance modeling, students' support and feedback, curriculum designing, using different techniques of data mining. A comprehensive view of the work in educational data mining is presented by Romero and Ventura [32] in which they have summarized work in Educational Data Mining field till 2005 and extend it further to a survey paper [33] that discusses the work till 2010.

Researchers have analyzed students' web logs using statistical tools and visualization to help understand students learning capabilities, for example, Hwang et. al [18] used statistical analysis to provide information about the kind of pages visited and the time spent by students on these pages, Juan et. al. [20] traced weekly students' activity information for instructors. However, these analyses were statistical and did not provide any hidden information. Instructors have been helped in decision making based on feedback obtained through techniques of data mining. Psaromiligkos [29] proposed a continuous feedback from the logging activities of the students using data mining techniques, Cackir [5] traced interaction pattern of students in chats. Similarly, using association rule mining and sequential pattern analysis on the web usage of students, researchers have been able to recommend personalized learning scenarios, H. Ba-Omar [2] and personalized activities Wang [40]. V. Rus et. al [34] have compared various Data mining algorithms in order to provide a mental model for students. The majority of work has been done in the area of performance prediction using different attribute set and different mining algorithms.

As the academic performance of the students not only decides the reputation of the institution but, also the dropout rate, the retention rate, as well as the employability of the students we find that a great deal of work in Educational Data Mining is focused on the performance prediction using different attribute set and different mining algorithms.

Kabakchieva [21] used 14 attributes including personal profile, secondary educational score, entrance exam score, admission year, etc. to predict performance using classifier J48, Bayesian, K-nearest neighbor one R and J Rip. Kabakchieva concludes that J48 performs best with highest overall accuracy.

Cheewaprakobkit [6] found that 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 and decision tree was best classifier. Sen et al [37] have ranked importance of 24 predictor variables, including demography, scores in mathematics, Turkish, religion and ethics, science and technology and level determination exams etc. for predicting Turkish secondary education placement result. Osmanbegovic and Suljic [10] collected 12 attributes and applied Chi square test, One-R test, Information Gain test and Gain ratio test and found that GPA, score of entrance exam, study material and average weekly hours devoted to studying are having maximum impact while the number of household members, the distance of residence, and gender have the least impact. N. S. Shah [39] applied various algorithms, decisions to categorize (predict) students in 5 categories (Very good, Good, Satisfactory, Below Satisfactory and Fail). Sembiring et al [35] 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. Bharadwaj and Pal [4] base their experiment only on Previous Semester marks, class test grade, seminar performance, Assignment, attendance, Lab work to predict end semester marks. The paper also calculates Split info, gain ratio of each predictor and products prediction rules. S. Huang [16] uniquely considers 6 combinations of predictor variables, three pre-requisite courses, scores of three dynamics midterm, to predict academic performance in theEngineering Drawing course. The analysis reveals that the 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 the percentage of predictions. Huang recommends the use of SVM when individual academic performance of the student is to be predicted while the multiple regression technique is best suited when average performance of the whole class is to be predicted.

Paris et al. [1] have compared C4. 5, NB Tree, Bayes Net, Hidden Nave Bayes, and voting techniques of classification based on three weak classifiers (Nave Bayes, One R and Decision Stump) for improving the accuracy of performance prediction. The students dropping out of an open polytechnic of New Zealand due to failure has been explored by Z. Kovaic [23] Ramaswami and Bhaskaran [30] used Chi-squared Automatic Interaction Detector (CHAID) to make an 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. Wook et al. [41] considered demographics (age, gender, religion, home town, etc.), Educational Background (Previous Qualifications, results) computer skill, name and number of courses taken, Total credit taken etc.) and personality (motivation of study, reading level, learning environment style and interests etc.) and uses secondary data bout CPA, CGPA grade points for 85 student undergraduate students of computer science department from National Defense University of Malaysia. Cortez and Silva [8] have focused on prediction of school students' performances taken from two secondary schools.

The ultimate aim of the academic excellence of the students is to seek employment and this also decides the reputation of the organization.

Although a lot of research has been done in the academic performance prediction that may lead to employability, the employability prediction is still in a nascent state. Even the term "Employability" still has no precise definition. Lee Harvey [15] 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, it may be the ability to skill map oneself according to the job need, or the willingness of the student to extend the graduate learning at work. In this paper we have taken employability as securing a job while on campus i. e. while students are in the fifth semester and get placement offers from companies.

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 and Coetzee [28] using methods of psychology. Yusoff et al. [42] have used statistical methods to calculate the performance score of entry level engineers based on normalized skill weight. Employers perspective regarding various skills was collected through questionnaires. 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.

Bangsuk Jantawan [19] has used real data of graduate students of Maejo University in Thailand for three academic years. Various algorithms of Bayesian Network and Decision Tree have been used to build the classification model for graduate employability. The study Finch et al [11] was conducted to increase our understanding of factors that influence the employability of university graduates. Using both qualitative and quantitative methods it was concluded that employers find soft skill more significant in employment as compared academic excellence. This exploratory study further explores the attributes like emotional intelligence, life experience work life balance has been linked to employability, however, the findings are limited to theoretical aspects. A recent study by Sapaat et al [36] included almost the same methodology as above in which the data was sourced from the Tracer Study database. Application of data mining algorithms indicates the superiority of the Decision Tree classification model over Bayes Network Classification Models. Lack of communication skill, creative and critical thinking, problem solving were indicated by Noor Aieda et al [3] and highlight why they are important for employment. Researchers Chein and Chen [7] 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. General studies for identifying factors that affect the job prospects of a student, have been conducted by Shafie and Nayan [38], Mukhtar et al [25] and Kayha [22] where communication skill, enthusiasm, attitude, employee position have come out as major factors essential for employment. More recently a new Malaysian Engineering Employability Framework has been proposed which is based on accrediting and professional bodies recommendations and suggests qualifications as well as training guidelines to make students employable in Malaysia. Rees [30] in his report by the Higher Education Academy with the Council for Industry and Higher Education (CIHE) in the United Kingdom concluded that cognitive, personal, technical, practical, generic abilities along with organizational awareness are the most important competencies employers look for. V. K. Gokuladas in the year 2010 [12] and 2011 [13] respectively establishes that apart from academic excellence certain other skills are must in order to secure jobs. He further concludes that GPA and proficiency in the English language are important predictors of employability.

Experimental Setting

The major objective of the proposed methodology is to apply various classification algorithms on the data set and build the prediction model based on the most suitable algorithm that classifies student's employability (i. e. on-campus placement as Yes or No) indicating whether the student was placed during the on-campus placements or not.

Data Collection

A sample of 1400 students of MCA students of various colleges in India was collected through a structured questionnaire, which included following attributes that have been categorized into:

1. Demographic profile/Social Integration consists of GENDER(Male, Female), Fathers Education (FE) and Mothers Education (ME) (SECONDARY, SENIORSECONDARY, GRAD, POSTGRAD), Father's and Mother's Occupation(FO, MO)(Govtjob, Pvtjob, Business, Others), Family Income FI (LIG Low income Group i. e. less than 2 lac per annum) MIG i. e. Middle income group, (2 to less than 4 lacs per annum), HIG High Income Group(4 lacs to 6 lacs per annum), VHIG Very High Income Group(more than 6 lacs per annum), where 1 lac is equivalent to 100000, Loan(Yes, No) depending upon whether the student has taken Educational loan at any level of education, Early Life (Where a student has spent first 15 years of his life (Metro, City, Village)
2. Academic Integration consists of MI(Medium of instruction at school level. (English, Other) Percentage marks in SECONDARY, SENIORSECONDARY, GRADUATION, FIRSTSEM, SECSEM, THIRDSEM, FOURTHSEM are categorized into BLAVG(Less than 60), AVG(60 to less than 70), ABAVG(70 to less than 80), EXCL(more than 80), GRADDEGTYPE(Type of Graduation Degree)(Regular, Distance), GRADDEGSTREAM(Computer Science, Non Computer Science)(CS, NCS), GAPYEAR(Gap year in education)(Yes, No), RELWORKEXP((Relevant work experience in the gap year)(Yes, No), ONCAMPUSPLACE (Yes, No), ACADEMICHRS (Hours spent on academic activities)(INSUF i. e. less than 2 Hrs., SUF i. e. 2-4Hrs, OPTIMAL i. e. more than 4Hrs), PROJECT(whether a project was done by the student or not)(Yes, No)
3. Emotional Skill parameters are assessed through Emotional Skill Assessment Process(ESAP) tool developed by Darwin Nelson and Garry Low, consisting of psychometric questions to judge various parameters [9]. The parameters selected are as follows:
   i. 'ASSERTION' which is the ability of a student to communicate effectively, honestly and clearly.
   ii. 'LEADERSHIP SKILL' which is the ability to lead in any situation
   iii. 'EMPATHY' is the ability to identify oneself with others' situation and pay attention to the needs of others
   iv. 'DECISION MAKING' is the ability to take quick and informed decision
   v. 'DRIVE STRENGTH' evaluates whether or not a student has some well-defined goals and aims in life

vi. ‘TIME MANAGEMENT’ evaluates the skill of a student to view time as a valuable resource and make judicious use of it and put it to the best possible use.

vii. ‘SELF ESTEEM’ is a person’s opinion, regard and respect for one’s own self and shows a student’s faith in one’s own abilities

viii. ‘STRESS MANAGEMENT’ evaluates the ability of a student to cope with the stress of responsibilities and demands of daily life and work.

Data Pre-Processing and Data Development

Cleaning the data involved eliminating data with missing values or inputting it where possible, rectifying inconsistent data, identifying outliers and removing them, as well as removing duplicate data. This process reduced the data set to 1359 MCA students, which we use for further analysis. The Excel data set obtained was converted to CSV(Comma Separated Variables) file as required by WEKA.

The Dataset developed appears as presented in table 1, 2 and 3.

Table 1: Partial Dataset with First 10 Attributes

GENDER	FE	ME	FO	MO	FI	LOAN	EARLYLIFE	MI	SECONDRY
M	GRAD	GRAD	GOVTJOB	HOUSEWIFE	MIG	No	Metro	English	AVG
F	GRAD	SENIORSECONDARY	BUSINESS	HOUSEWIFE	HIG	No	City	English	EXCL
M	GRAD	GRAD	PVTJOB	GOVTJOB	HIG	Yes	Metro	English	ABVG
M	POSTGRAD	GRAD	GOVTJOB	HOUSEWIFE	MIG	Yes	City	English	ABVG
F	SECONDARY	SENIORSECONDARY	BUSINESS	HOUSEWIFE	MIG	Yes	City	English	ABVG
M	GRAD	SENIORSECONDARY	GOVTJOB	HOUSEWIFE	HIG	No	Metro	English	EXCL
M	POSTGRAD	POSTGRAD	PVTJOB	GOVTJOB	HIG	No	Village	English	AVG
M	GRAD	GRAD	GOVTJOB	OTHER	MIG	Yes	Metro	English	AVG
M	GRAD	GRAD	GOVTJOB	HOUSEWIFE	MIG	No	City	English	ABVG
M	SECONDARY	SENIORSECONDARY	OTHER	HOUSEWIFE	LIG	No	Village	Other	AVG

Table 2: Partial Dataset with Second 10 Attributes

SENIOR SECONDRY	GRAD	FIRST SEM	SECOND SEM	THIRD SEM	FOURTH SEM	GRADDEG TYPE	GRADDEG STREAM	GAP YEAR	RELWORK EXP
AVG	BLAVG	AVG	AVG	AVG	AVG	REGULAR	CS	Yes	No
EXCL	AVG	EXCL	EXCL	EXCL	EXCL	REGULAR	CS	No	No
AVG	BLAVG	AVG	AVG	AVG	AVG	REGULAR	CS	Yes	No
ABVG	AVG	AVG	AVG	AVG	AVG	REGULAR	CS	Yes	Yes
AVG	EXCL	AVG	AVG	AVG	AVG	REGULAR	CS	Yes	No
AVG	AVG	AVG	AVG	AVG	AVG	DISTANCE	CS	Yes	No
AVG	AVG	AVG	AVG	AVG	AVG	DISTANCE	CS	No	No
ABVG	AVG	AVG	AVG	AVG	AVG	REGULAR	CS	No	No
AVG	BLAVG	AVG	AVG	AVG	AVG	REGULAR	CS	No	No
ABVG	EXCL	BLAVG	AVG	AVG	AVG	REGULAR	CS	No	No

Table 3: Partial Dataset with Last 11Attributes

ACADEMICHRS	PROJECT	ASSERTION	EMPATHY	DECISION MAKING	LEADER SHIP	DRIVE STRENGTH	TIME MANAGEMENT	STRESS MANAGEMENT	SELF ESTEEM
SUFF	Yes	S	D	S	E	D	E	D	D
SUFF	Yes	E	E	E	E	D	D	E	E
INSUFF	No	S	E	S	E	D	D	D	E
INSUFF	Yes	E	E	S	E	D	S	E	D
SUFF	Yes	D	D	D	D	D	D	E	E
INSUFF	Yes	S	E	E	E	E	D	E	E
SUFF	Yes	S	D	E	E	D	S	D	D
SUFF	Yes	S	D	S	E	D	E	D	D
SUFF	Yes	S	E	E	E	E	D	E	E
SUFF	Yes	S	D	S	D	D	D	D	D

Modeling

The classification models have been built to predict the campus placement as a binary Yes/No output variable, using Waikato Environment for Knowledge Analysis (WEKA), an open-source data mining tool, that provides various learning algorithms that can be applied to the data set.

Assigning the target attribute (in this study on campus placement) a class/group (yes or no) based on other attributes is called classification. A number of classification techniques are available, each having its own advantages and disadvantages.

All the attributes of our data set are nominal, hence we have considered the algorithms which could handle the nominal data like Decision tree, Bayesian classifiers, Support Vector Machine classifiers, Multilayer Perceptrons and ensemble methods like Random Forests and Random Tree. Before proceeding further, a brief description of each algorithm is given.

In a decision tree we start with all instances at the root node. Then the attribute that gives best discrimination is used at the root node, and branches out to inner nodes based on the splitting attribute. The process continues till all the instances at a node belong to the same class or some threshold criteria is met or there are no attributes left.

One of the most useful characteristics of decision trees is their comprehensibility and easy interpretation in the form of rules. The assumption made in the decision trees is that instances belonging to different classes have different values in at least one of their features [24]. We have considered J48 algorithm which is a class for generating a pruned C4. 5 decision tree in WEKA.

Bayesian classifiers are based on Naive Bayes theorem and assume that each attribute is independent of the other attributes of the instance. The conditional probability of a class label is estimated, and the assumption is then made on the model to decompose this probability into a product of conditional probabilities. According to [23] a Bayesian Network (BN) is a graphical model for probability relationships among a set of variable's features. The Bayesian network structure S is a directed acyclic graph (DAG) and the nodes in S are in one-to-one correspondence with the features X. The arcs represent casual influences among the features while the lack of possible arcs in S encodes conditional independence. Bayesian classifiers, Nave Bayes have been considered.

Support vector machines are based on statistical learning and work with high dimension data and represents the decision boundaries using a subset of the training set called support vectors. Sequential Minimal Organization (SMO) algorithm hasbeen used in this paper.

Ensemble techniques are one in which a set of base classifiers is constructed from the training dataset and perform classification taking a vote on the predictions made by each of the classifiers. Random forest and random tree have been used for ensemble method. We have used Cross-validation for testing as it has been proved to be more suitable for limited data set. [13].

Results and Discussion

Decision tree J48, Bayesian classifiers Naïve Bayes, Support Vector Machines algorithm SMO and Ensemble Methods Random Tree and Random Forestand Multilayer Perceptronswere used on the data set using 10-fold cross validation.

Researchers have mostly considered accuracy, TP rate, FP rate, etc. for comparing classifiers but these measurements are not used for unbalanced class. We have used area under ROC curve, Accuracy and F Measure as the evaluation metricsin our experimental setting as these measures are the most comprehensive and evaluates the classifier's performance fairly. The time to build the model also plays an important role in real time application with huge data sets hence it is included the comparative analysis.

Table 4: Performance comparison of classifiers

Algorithm	ROC Area	Accuracy (%)	F Measure	Time to build
J48	0. 661	70. 19	0. 702	. 02
Random forest	0. 745	71. 304	0. 704	. 11
Random Tree	0. 615	63. 35	0. 629	0. 0
SMO	0. 608	63. 7	0. 635	0. 39
Multilayer Perceptron	0. 725	70. 64	0. 706	34. 05
Naive Bayes	0. 725	62. 87	0. 629	0. 0

Naive Bayes algorithm is not suitable due to low accuracy 62. 87%. SMO and Multilayer Perceptron have a comparatively higher accuracy of 63. 7% and 70. 64% respectively but higher model building times too requiring 0. 39 seconds and 34. 04 seconds respectively, and this factor will become all the more significant with large data set.

Ensemble method Random Tree generates a tree that is easily interpret able and takes no time in building this model but has low accuracy of 63. 35% and area under ROC curve is also less (. 615) hence, it is not suitable for the present problem.

The other Ensemble Method Random Forest has highest accuracy of 71. 304%, highest ROC area (0. 745) and highest F measure (0. 704). The drawback of the method is that it is least interpret able and has model building time of 0. 11 seconds. Since the aim is to understand the parameters affecting the employability Random Forest is not considered suitable for this problem.

The algorithm J48 which is implementation of decision tree C4. 5 generates an easily interpretable decision tree from which useful rules can be derived. Further, it's accuracy 70. 19% is comparable with that of Random Forest 71. 3 % as well as the F measure of 0. 702 comparable with 0. 704 of Random Forest and time to build the model 0. 02 seconds as compared to 0. 11 seconds of Random Forest.

We further consider ROC curve generated by J48, Random Forest and Random Tree as presented in Fig. 1, 2 and 3 respectively. All the three curves are confined to top left corner above the diagonal of random performance, indicating that the predictions are better than by chance

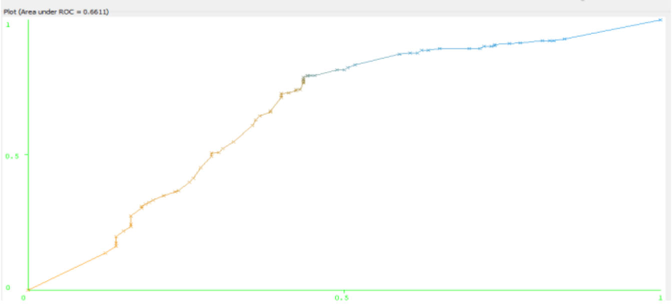

Figure 1: ROC curve for J48

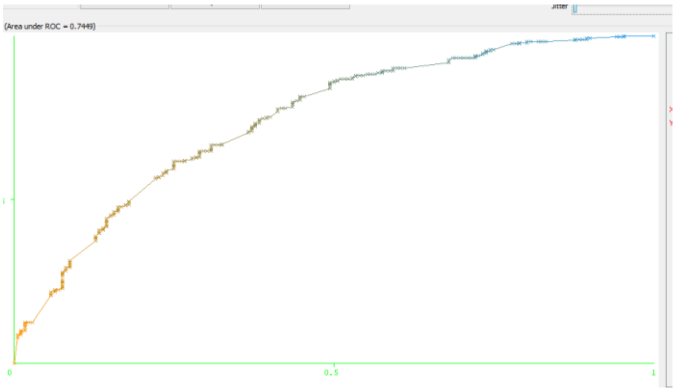

Figure 2: ROC curve for Random Forest

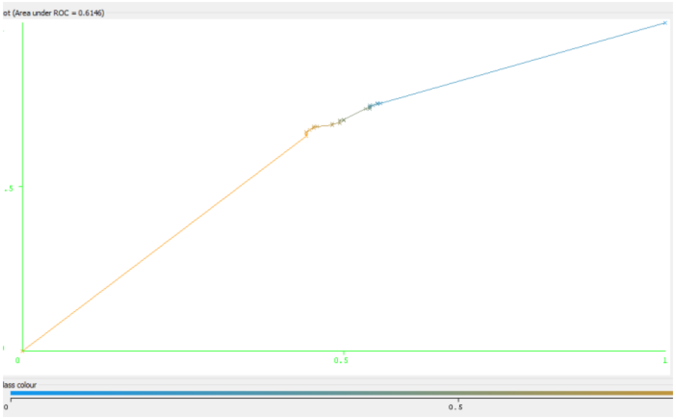

Figure 3: ROC curve for Random Tree

Model Developed

J48suits the problem of identifying students at risk of not getting employment, hence, it is worthwhile to consider the rules generated by J48. These rules give an insight of the attributes that affect the employability of the students.

Rules derived from J48

1. If (PROJECT=Yes) and (THIRDSEM=EXCL) and (RELWORKEX=No) and (STRESSMGMT=D): No
2. If (PROJECT=Yes) and (THIRDSEM=EXCL) and (RELWORKEX=No) and (STRESSMGMT=E or STRESSMGMT=S): Yes
3. If (PROJECT=Yes) and (THIRDSEM=EXCL) and (RELWORKEX=yes) and (ACADEMICHRS=SUFF): Yes
4. If (PROJECT=Yes) and (THIRDSEM=EXCL) and (RELWORKEX=yes) and (ACADEMICHRS=INSUFF): No
5. If (PROJECT=Yes) and (THIRDSEM=ABVG) and (EMPATHY=D or S): Yes
6. If (PROJECT=Yes) and (THIRDSEM=ABVG) and (EMPATHY=E):Yes
7. If (PROJECT=Yes) and (THIRDSEM=BAVG) and (EARLYLIFE=Metro or EARLYLIFE=City): Yes
8. If (PROJECT=Yes) and (THIRDSEM=BAVG) and (EARLYLIFE=Village): No
9. If(PROJECT=No)and (MI=English)and(LOAN=No):No
10. If (PROJECT=No) and (MI=English) and (LOAN=Yes) and (DRIVE=S or DRIVE=D): No
11. If (PROJECT=No) and (MI=English) and (LOAN=Yes) and (DRIVE=E): Yes

Conclusion

The aim of this paper was to apply various classifiers to find the employability of students and develop employability model based on the suitable classifier. It was found that J48 algorithm which is implementation of pruned C4. 5 Decision Tree algorithm of WEKA is most suitable for the employability prediction. Further, from the rules derived from the model emphasize following points

- A good project during the course, aids in the placement of the students as it sharpens the practical skills of the students.
- In the syllabus of MCA, third Semester consist of two major computer Languages JAVA and C Sharp which are very much in demand by recruiters. A student having project, and Strengthened(S) or Enhanced (E) Stress management skill is employable even if he does not have relevant work experience. Under similar situation a student whose stress management skill needs to be developed(D) is not employable.
- Academic Hours put into study are important, as a student with project and excellent third semester result is employable if ACADEMICHRS are sufficient (SUFF i. e. 2to 4 Hrs.) than a student who puts in Insufficient (INSUFF i. e. less than 2 Hrs.) ACADEMICHRS into study.
- A student's empathy towards other students is an unexpected factor which emerges from these rules as we find that a student with project and Above Average Third Semester result is employable when he has enhanced(E) EMPATHY whereas, student who need to develop (D) or strengthen (S) their EMPATHY predicted to remain unplaced through on-campus placement. One possible reason behind this factor can be that a student with high empathy will be helpful to other students by sharing his knowledge and teaching his peers which in turn sharpens his soft skills like expressiveness, convincing ability and communication. Apart from this it makes him/her a team player which is most sought after quality in the industry.
- 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.
- A student who has enhanced(E) DRIVE and has taken loan is predicted to get on campus placement even if he does not have a project but his medium of instruction has been English whereas, under similar conditions A student lacking in DRIVE, that is having DRIVE equal to (D) or (S) is at risk of not getting On campus placement.

This paper identified factors affecting employability and then has applied and compared various classifiers. The effect of some of the emotional skill parameters on placement has been established where others have not shown so much effect as per expectation.

Future Work

This research work has considered only MCA students, whereas Bachelor of Science (B. Sc) and Bachelor of Engineering (B. E) are also popular professional courses among employers. 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 of B. S. / 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 employees logical reasoning, quantitative aptitude, communication skill, etc. These parameters will be included in further research so as to infer if academically average students are also capable of getting good employment based on other skills and virtues.

References

[1] L. Affendey, I. Paris, N. Mustapha, M. Sulaiman, and Z. Muda, "Ranking of influencing factors in predicting students' academic performance, " Information Technology Journal, Vol. 9 No. 4, pp. 832-837, 2010.

[2] H. Ba-Omar, I. Petrounias, and F. Anwar, "A framework for using web usage mining for personalize e-learning, " Proc. IEEE International Conference in Advanced Learning Technologies Seventh, pp. 937-938, 2007.

[3] A. Bakar, A. Noor, A. Mustapha, and K. M. Nasir, "Clustering analysis for empowering skills in graduate employability model, " Australian Journal of Basic and Applied Sciences, Vol. 7, No. 14, pp. 21-24, 2013.

[4] B. Bharadwaj, and S. Pal, "Mining educational data to analyze students' performance, " International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, pp. 63-69, 2011.

[5] M. Cakir, F. Xhafa, N. Zhou, and G. Stahl, "Thread-based analysis of patterns of collaborative interaction in chat, " Proc. Int. Conf. AI Educ., Vol. 125, pp. 120-127, 2007.

[6] P. Cheewaprakobkit, "Study of factor analysis affecting achievements of undergraduate, " International Multi Conference of Engineers and Computer Scientists, 2013.

[7] C. Chien, and L. Chen, "Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry, " Expert Systems with applications, Vol. 34, No. 1, pp. 280-290, 2008.

[8] P. Cortez, and A. Silva, "Using data mining to predict secondary school student performance, " Proc. 5th Annual Future Business Technology Conference, Porto, pp. 5-12, 2008.

[9] N. Darwin and L. Garry, "Emotional skill assessment process. "

[10] E. Osmanbegovic and M. Suljic. "Data mining approach for predicting student performance, " Economic Review-Journal of Economics and Business, Vol. 10, No. 1, pp. 3-12, 2012

[11] D. Finch, L. Hamilton, R. Baldwin, and M. Zehner, "An exploratory study of factors affecting undergraduate employability, " Education Training, Vol. 55, No. 7, pp. 681-704, 2013.

[12] V. K. Gokuladas, "Technical and non-technical education and the employability of engineering graduates: An Indian case study, " International Journal of Training and Development, Vol. 14, No. 2, pp. 130-143, 2010.

[13] V. K. Gokuladas, "Predictors of employability of engineering graduates in campus recruitment drives of indian software services companies, " International Journal of Selection and Assessment, Vol. 19, No. 3, pp. 313-319, 2011.

[14] H. Guruler, A. Istanbullu, and M. Karahasan, "A new student performance analysing system using knowledge discovery in higher educational databases, " Computers and Education, Vol. 55, No. 1, pp. 247-254, 2010.

[15] L. Harvey, "Defining and Measuring Employability, " Quality in Higher Education, Vol. 7, No. 2, pp. 97-109, 2001.

[16] S. Huang, "Predictive Modeling and Analysis of Student Academic Performance in an Engineering Dynamics Course, " 2011.

[17] http://digitalcommons. usu. edu/etd/1086

[18] G. J. Hwang, P. S. Tsai, C. C. Tsai, and J. C. R. Tseng, "A novel approach for assisting teachers in analyzing student web-searching behaviors, " Computer Education, Vol. 51, No. 2, pp. 926-938, 2008.

[19] B. Jantawan, and C. Tsai, "The application of data mining to build classification model for predicting graduate employment, " International Journal of Computer Science and Information Security, 2013.

[20] A. A. Juan, T. Daradoumis, J. Faulin, and F. Xhafa, "SAMOS: a model for monitoring students' and groups' activities in collaborative e-learning, " International Journal of Learning Technology, Vol. 4, No. 1-2, pp. 53-72, 2009.

[21] D. Kabakchieva, "Predicting student performance by using data mining methods for classification, " Cybernatics and Information Technologies, Vol. 13, No. 1, pp. 61-71, 2013.

[22] E. Kahya, "The effects of job characteristics and working conditions on job performance, " International Journal of Industrial Ergonomics, Vol. 37, No. 6, pp. 515-523, 2007.

[23] Z. Kovacic, "Early prediction of student success: Mining students' enrollment data, " Proc. Informing Science & IT Education Conference (InSITE), pp. 647-665, 2010.

[24] Mitchell, T. Machine learning. McGraw Hill (1997).

[25] M. Mukhtar, Y. Yahya, S. Abdullah, A. N. Hamdan, N. Jailani, and Z. Abdullah, "Employability and service science: Facing the challenges via curriculum design and restructuring, " Proc. International Conference on Electrical Engineering and Informatics, Vol. 2, pp. 357-361, 2009.

[26] A. Pena-Ayala, "Educational data mining: A survey

and a data mining-based analysis of recent works, " Expert systems with applications, Vol. 41, No. 4, pp. 1432-1462, 2014.

[27] T. N. Phyu, "Survey of classification techniques in data mining, " Proc. International Multi Conference of Engineers and Computer Scientists, Vol. 1, pp. 18-20, 2009.

[28] I. Potgieter, and M. Coetzee, "Employability attributes and personality preferences of postgraduate business management students, " SA Journal of Industrial Psychology, Vol. 39, No. 1, pp. 1-10, 2013.

[29] Y. Psaromiligkos, M. Orfanidou, C. Kytagias, and E. Zafiri, "Mining log data for the analysis of learners' behaviour in web-based learning management systems, " Operational Research, Vol. 11, No. 2, pp. 187-200, 2009.

[30] M. Ramaswami, and R. Bhaskaran, "A chaid based performance prediction model in educational data mining, " International Journal of Computer Science Issues, Vol. 7, No. 1, pp. 10-18, 2010.

[31] C. Rees, P. Forbes, and B. Kubler, "Student employability profiles, " A Guide for Higher Education Practitioners, 2006.

[32] C. Romero, and S. Ventura, "Educational data mining: A survey from 1995 to 2005, " Expert systems with applications, Vol. 33, No. 1, pp. 135-146, 2007.

[33] C. Romero, and S. Ventura, "Educational data mining: a review of the state of the art, " IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, Vol. 6, pp. 601-618, 2010.

[34] V. Rus, M. Lintean, and R. Azevedo, "Automatic Detection of Student Mental Models During Prior Knowledge Activation in Metatutor", Proc. International Conference on Educational Data Mining, pp. 161-170, 2009.

[35] S. Sembiring, M. Zarlis, D. Hartama, S. Ramliana, andE. Wani, "Prediction of student academic performance by an application of data mining techniques, " Proc. International Conference in Management and Artificial Intelligence, Vol. 6, pp. 110-114, 2011.

[36] M. Sapaat, A. Mustapha, J. Ahmad, K. Chamili, and R. Muhamad, "A data mining approach to construct a graduates employability model in Malaysia, " International Journal of New Computer Architectures and their Applications, Vol. 1, No. 4, pp. 1086-1098, 2011.

[37] B. Sen, E. Ucar, and D. Delen, "Predicting and analyzing secondary education placement-test scores: A data mining approach, " Expert Systems with Applications, Vol. 39, No. 10, pp. 9468-9476, 2012.

[38] L. Shaffie, and S. Nayan, "Employability awareness among Malaysian under-graduates, " International Journal of Business and Management, Vol. 5, No. 8, pp. 119-123, 2010.

[39] N. Shah, "Predicting factors that affect students academic performance by using data mining techniques, "Pakistan Business Review, Vol. 13, No. 4. pp. 631-668, 2012.

[40] F. H. Wang, "Content recommendation based on education-contextualized browsing events for web-based personalized learning, "Journal of Educational Technology and Society, Vol. 11, No. 4, pp. 94-112, 2008.

[41] M. Wook, Y. Yahaya, N. Wahab, M. Isa, N. Awang, and H. Seong, "Predicting NDUM student's academic performance using data mining techniques, " Proc. Second International Conference in Computer and Electrical Engineering, Vol. 2, pp. 357-361, 2009.

[42] Y. Yusoff, M. Omar, A. Zaharim, A. Mohamed, and N. Muhamad, "Employability skills, performance score for fresh engineering graduates in Malaysian industry, " Asian Social Science, Vol. 18, No. 16, pp. 140-145, 2012.