IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-7, July 2018 E-ISSN: 2347-2693

Predicting Student Performance to Improve their Employability by Applying Data Mining and Machine Learning Techniques

C. Jayasree, K. K. Baseer*

[1] IT, Sree Vidyanikethan Engineering College, JNTUA, Tirupati, India
[2] IT, Sree Vidyanikethan Engineering College, JNTUA, Tirupati, India

Available online at: www.ijcseonline.org

Accepted: 21/Jul/2018, Published: 31/Jul/2018

Abstract- Student performance, dropouts has become an interesting topics in engineering education. To improve student Academic performance, employability and reducing dropouts are the most vital issues in the research. As it has been observed in the literature survey, there are many existing techniques in the Data mining to predict student's GPAs, grades, dropouts, and desertion. If the student desertion issue is underestimated then one cannot cope the student prediction optimally, which can causes significantly high error. This problem can be minimized with appropriate preventive strategies of Data mining techniques like Matrix Factorization, Rador, and Part in advance. However, the results obtained are still erroneous and to overcome this risk of failure some Machine Learning Approaches like Regression, classification and clustering methods are applied along with DMT which are highly effective. To predict the performance of the students accurately, here we considered various datasets like previous grades, study time, parent's status, GPA, school support, higher education, internet usage, travel time etc.,. Which crucially carry out the effective performance, grades for the next term. This can help us for the satisfactory graduation and completion of education on time. The comparative study is done on different algorithms such as linear regression, K-means clustering and neural networks using Weka and Azure tools. This can yield us a better student prediction along with preventive strategies for significantly low error. Further, we can extend our study with few more datasets and it might be possible to find a particular student who can perform effectively up to the mark without any failure. This will help us to reduce the drop outs, failure percentage and increases the confidence levels in the students so that, the progression of student performance can be monitor semester by semester.

Index Terms: Dropouts, Academic performance, Employability, Machine learning and Data mining techniques

I. INTRODUCTION

The retention of students in university as well as the successful completion of their studies is of interest not only for higher education institutions (HEI), but also for society as a whole, because of its many implications. Also making higher education affordable has a significant impact on government and it is the central focus of it, while making Educational policies. In many major countries as the cost of the college education had been increasing constantly, loan debts of the students also increases as they fail to complete their graduation on time. To counteract this student desertion problem, the Education Ministry in major countries had designed policies for its diagnosis, monitoring and prevention. This strategy has opened the way for subsequent researches on the academic, next term grade prediction of the students by collecting data.

Universities accumulate big volumes of data which can be used to obtain information on the development and performance of academic activities. The information that can be obtained from academic databases will serve to look for answers to such questions as: What identifies students with good and poor academic achievement? Which are the causes of the students' retention in the university? Why do students desert? Automatic data mining techniques can be applied to solve these questions and to facilitate the development of strategies for improving the academic processes and the educational programs. One of the main recognized factors for the student desertion is academic risk. This leads to the necessary Universities to design a mechanism which can be able to identify the causes for academic risk, and act upon these causes before the risk has occurred. This would allow the institution to take timely decisions and design better strategies to prevent academic risk as a phenomenon, and as a consequence, diminish the students' probability of desertion.

To overcome them some of data mining methods and techniques are applied. This process of study is known as Educational Data Mining (EDM) [5]. EDM has proven being a useful tool for making predictions in several scenarios and thus will be used to generate a predictive model based on students' academic.

By using advanced Machine Learning Approaches like Linear Regression-Neural network and k-

means clustering and deep SVM methods which are highly effective and can be able to detect the causes for the academic risk and can be able to reduce the student desertion and helps us to take the necessary preventive measures to ensure the completion of the student graduation in time. Hence here we provide useful insights about the causes for academic risk, in order to empower the Institution in the taking of better decisions and the implementation of stronger strategies regarding student desertion.

We also present detailed Algorithms to illustrate how the different algorithms can be applied to determine the best results applied on the datasets that are shown in the next sections of the paper. The rest of this paper is organized as follows: Section II provides a summary of the paper on Literature Survey. Sections III and IV discuss the overall description about Weka and Ms Azure tools, working procedure, algorithm types and their comparisons. Section V shows the results of respective applied algorithm.

II. SYSTEMATIC LITERATURE REVIEW

After going through many articles, papers of scientific and standard publications from the year 2000 to 2017 and from them more than 500 distinct papers are congregated and filtered to the count of Journals (91), Conferences (60), Thesis and others (55).Among them we considered a few approaches which were optimal and precise for the 'prediction of the academic performance of a student'. The key factors which we considered from the articles for the academic prediction were Employability, Academic Performance, Machine Learning Techniques, Drop-outs and Placements.

The issue of student dropout, detection will be reduced by applying the above mentioned algorithms and approaches. Till now the previous study of research is done using Weka tool. As the data will be pre-saved in Weka it won't give that much of worthy information. Further, we can extend our study with possibility to find a particular course in which a student is interested, can perform effectively up to the mark without any failure and to find the risk factors using methods and techniques. MS Azure tool is used for implementation of student dropout and detection. MS Azure is said to be best tool of Microsoft for predicting the real world problems, there are still many open source software's i.e. sales force, oracle OWB, Microsoft SSIS etc. To predict the results accurately we reduce the unnecessary and noisy data attributes. So that the data will be less, time period for the execution will be low and the output will be effective, early warnings which improve positivity in students. For further detailed information [136].

A. Discussion on SLR

Table 1 describes the six aspects of each attribute: Objectives, approaches and algorithms, datasets, implementation, future work/results and Title of the paper. In this paper, mainly focused on Desertion, Dropouts and Poor Academic performance
of a student.

1. Issues on academic performance:

This mainly speaks on which the most of research has been done. Here we find the accuracy and speed on clustering and classification algorithms in the field of prediction. It's also been viewed to atomize the systems to gather more datasets and more problems.

2. Issues on dropouts:

The research done on dropouts is to develop educational stability, deeper mining in studies motivation and to develop long term goals.

3. *Issues on Employability:*

The overall research on EMP says to follow the recommendation models and better tools.

4. Issues on DMT:

After applying DMT, for better results we need proper decision on projects should be taken, and student interested fields to be identified.

*5. Issues on MLT***:**

This survey says to make use of better models to improve performance, accuracy, promotions, increment and carrier advancement should be considered for better prediction in future.

2.2. Critical study on Academic performance, Employability, Machine Learning, Data Mining Techniques, Drop-outs Academic performance

S.No	Issue/Objectives	Solution/ Algorithm/ Techniques/ Approaches	Application/ Datasets/ Real-time attributes	Implementations	Results/ Future work	References
1.	The variable impacts on individual learning processes and of those factors and their different AP (GAP) into: Low, Middle and High, grade-point-average (GPA).	**Techniques:** **A**NN approaches Along with neural networks	**Factors:** g-factor and intelligence.	**Software used**: SSPP v.19. .	**Results:** The identification of the specific influence of each pattern of variables on different level of academic-performance.	**1**
2.	Comparing recommender systems with traditional techniques such as linear regression or logistic regression. From the experimental perspective.	**Algorithms:** Recommender systems, matrix factorization methods, neural networks, SVM	**Datasets:** Problem hierarchy, problem name, step name, and problem view.	**Implementation:** Applying recommender system techniques such as matrix factorization in the educational context, especially for predicting.	**Future work:** In future work, Over student/problem/problem view or, more general student/ problem /time could give better.	11
3.	Clearly, academic performance is a critical factor take into account that, frequently, underachievement is associated with a high dropout rate.	**Models:** DMT approaches.	**Attributes**: Mother Emp, Father Emp, Student Emp, High school	**Tools used:** IBM Data Warehouse Edition (DWE) V.9.5	**Future Work:** In the future, we define specific actions reverse poor academic performance	35
4.	To identify the difference between high learners and slow learner students it is important for student performance to develop predictive data mining model.	**Algorithm:** K-means clustering, regression, NN, artificial intelligence and genetic algorithm, decision tree	**Attributes:** Student family bg, student details, subject marks, sem wise % is recorded.	**Implementation:** This study shows student performance and easy to identify those student which having less mark and poor performance	**Results:** To improve the performance, identify the student who needs special attention, reduce fail ration, take action right time.	40
5.	To find nearest possible a cluster a similar group the turning point India is the performance in higher education for all students	**Approaches:** Naive Bayesian, K-Means Clustering, Clustering,	**Attributes:** Student family background, student details, subjects marks, semester wise %.	**Implementation:** Comparison between algorithms.	**Results:** This study shows student performance and to identify those student who has less mark and poor performance	41
6.	The main aim of the study is to create a model that classifies the instances correctly to predict the performance of students using PAFT methodology	**Algorithm:** Novel genetic NAND PAFT model. Flowcharts.	**Factors:** Various factors are considered.	**Implementation:** Net beans are used for coding the GS-NAND algorithm and this file are converted to a Weka classifier format (.arff). to .csv and given to Weka, it reduces unsupervised filter	**Results:** Weka tool, it is found that the proposed GS-NAND PAFT model results were high and accurate than previous comparisons	57
7.	To overcome the Students who scored lower marks and to improve their performance	**Methods:** Association Rule Mining DT, Bayesian Neural Networks, ensemble Techniques are some of supervised methods.	**Datasets:** Attendance, Seminar, Assignment marks, Student performance, prediction, analysis, early alert, evaluation using DMT.	SVM is considered to be best accurate.	**Results and future work:** Reasons for poor in academics and drop-outs are social media. future work, to do research on classification, clustering to enhance the prediction speed and accuracy	58
8.	To focus on more improvement and accuracy of student performance. The objective of EDM is to develop new methods to explore educational data to determining the usefulness of learning systems	**Algorithms:** Naïve Bayes, Neural Network, and Decision Tree	**Attributes:** Age, section, number of students, program, hours studying, home tuition, sub interest transport, attendance.	**Implementation:** psychological profile, previous schooling, prior academic performance, and student interaction with their classmates and teachers	**Future work:** Research to carry out more with bigger dataset including different courses and levels of educational and also automate system to analyze factors.	62

9.	To overcome the Inconsistencies in determining which students‟ attributes contributes to academic performances	**Algorithm:** ID3, Simple CART, J48, NB Tree, MLP, Bayes net	**Datasets:** socio economic, non-academic and academic factors	Academic performance, Dropouts, course preference, Subject specialization etc.	**Results and Future work:** A new hybrid algorithm used for better classification and prediction using the high influence attributes would be the future work.	66
10.	social, economic and Psychological factors a student faces during their adolescence.	**Algorithms:** Support vector machine (SVM), K-nearest neighbor (K-NN), ANN	**Factors:** Social, economic and psychological factors a student faces during their adolescence.	**Implementation:** WEKA; dimensionality reduction algorithms, namely, PCA, SOM and GHA, were all implemented in Rapid Miner.	**Results:** The dimensionality reduction, discretization and normalization combinations derives the best model to predict the AP.	82
11.	One of the biggest challenges is to improve the quality of the educational processes so as to enhance student's performance	**Algorithm:** J48, DT	**Datasets:** 1. Login credential 2. Student detail record 3. Student outcome analysis	**Implementation:** J48 is a tree based learning approach, based on iterative dichtomiser (ID3) algorithm. It uses divide-and-conquer algorithm to split a root node into a subset of two partitions till leaf node (target node) occur in the tree.	**Results:** Improvement in J48 Is it uses WEKA as API in java code implemented in Net Beans. We used the classifiers (DT) and (RF) to predict student grade to extract rules to predict results	83
12.	To enhance the level of education in the society	**Models:** Multiple regression models are used.	**Datasets:** performance, score or marks, knowledge, guidance. Focused on students 'performance.	**Tools:** (SPSS) analysis tool. MLR to aid Statistical Package for Social-Sciences, Hypothesis testing was used to validate model.	**Results:** An intervention programs that bring student and educators in close mediation. A face to face discussion spark up the students‟ desires to do more.	88
13.	To overcome the lack of quality education	**Algorithms**: Association rule mining, Apriori. used MLT are Naïve Bayes and NN.	**Datasets:** surveys, interviews, focus groups, classroom activities	**Tools:** KNIME Analytics platform tool was used to perform the required work on the data provided	**Future work:** To find alcohol consumption students.	90
14.	To provide insights to our institute regarding various factors which led to the poor performance of the students?	**Algorithms:** classification rules using ID3 (Iterative Dichotomize 3) algorithm	**Attributes:** Poor, Average, Good and Excellent	**Future:** The co-curricular and the extra-curricular activities of students help to excel in academics , masters of all	**Results**: 1. Pre-processed data analysis, 2.Information gain 3.ID3classificatioresult	92
15.	to identify those students who need special attention to reduce failure ration and taking appropriate action for the next academic examination	**Algorithms:** Support Vector Machine (SVM) classifications and kernel k-map	**Attributes:** Class test, Attendance Seminar, innovative activity, Assignment	**Implementation:** The recommendations for additional activities, teaching material and task that would improve his/her learning.	**Results:** K-map is to analyze the relation -ships among student's success, behavior to develop model	98
16.	They should predict which student is at risk of failure and overcome that	**Algorithms:** The Comparison is done on Weka for training and a polynomial kernel function.	**Datasets:** achievement, assign -ments, class presence Personality ,aptitude	**Tools:** Ibk in WEKA Table.1-5 shows the comparisons of all d algorithms.	**Results:** Results are shown in graphical way.	99
17.	To identify the risk and take the special classes for the weakest people. The addiction of student towards alcohol in early life.	**Algorithms:** Classification algorithms Naïve Bayes, Decision Tree, K-NN and NN	**Attributes:** social, gender and study time attributes	**Tools:** The results are done using orange tool	**Results:** Results prove that Naive Bayes technique outperformed other used techniques.	100
18.	To identify the successful and unsuccessful students based on ethnicity, course programme and course block	**Models:** Classification tree models and logistic regression.	**Datasets:** age, edu, gender, work status ethnicity, disability study env- course block program	**Tools:** SPSS 17 and Statistica 8. Pass and Fail are considered.	**Results:** (CART) was the most successful in growing the tree with an overall percentage of correct classification of 60.5%.	102

19.	Conventional techniques may not give accurate solution so use soft techniques to overcome.	**Algorithms:** Fuzzy ARTMAP neural network, and SVM	**Factors:** Course, curriculum, learning, hard work, learning mistakes, notes reference books, revision, ex practicing.	**Implementation:** Fuzzy ARTMAP neural network is a supervised clustering algorithm	**Future work:** Future work in the same area many include exploring additional determinants which affect AP of a student in an institute.	103

Drop-outs

S.No	Issue/Objectives	Solution/ Algorithm/ Techniques/ Approaches	Application/ Datasets/ Real-time attributes	Implementations	Results/ Future work	References
1.	the work of data mining in predicting the drop out feature of students	**Approaches:** DTT,J4.8 implements a later and slightly improved version C4.5	**Datasets:** Identify risk factors that predict results	**Implementation:**	**Results:** New strategies can develop and implement enabling edu stability, system predictability and profits.	9
2.	analyzing educational data to develop models for improving learning experiences and improving institutional effectiveness	**Approaches:** k-NN, NB, NN, ID3, C4.5, CART, ADT.	**Datasets:** Students at risk? Chances of placement? Likely to drop? Quality of parti-cipation? Courses the institute offers to attract.	**Implementation:** The purpose of optimal managerial decision taking, in preventing students' exams failure, improving learning abilities and Scholastic results.	**Results:** Our studies continue with deeper mining of academic Failure, to detect with pragmatic exactness.	25
3.	higher risk of drop out (1) Dropping out before attending first class and (2) dropping out before completing the training.	**Approaches:** AdaBoost algorithm with ROCAUC	**Attributes** : Provider type, student language, language, age, age, class language, total classes residence, income, absences, withdraw count.	**Implementations** : Early Warning Indicator (EWI) systems are rapidly being built and deployed using machine learning algorithms.	**Future work** Future work: more student data like life circumstances to determine, motivations, propensity for self-efficacy.	107
4.	The most vulnerable students are the freshmen, who are at the highest risk of dropping out at the beginning of their study.	**Approaches:** Cox proportional hazards model (Cox). Time-dependent(TD-Cox),	**Datasets:** Family background, financial, high school info, college enrollment and sem-wise credits.	**Implementation:** Implementation is based on Comparison between two models "Predicting dropout student" and "estimating semester of dropout".	increasing student Retention is a long term goal of any academic institution.	77

Employability (J)

S.No	Issue/Objectives	Solution/ Algorithm/ Techniques/ Approaches	Application/ Datasets/ Real-time attributes	Implementations/ Tools	Results/ Future work	References
1.	A good project during the course, aids in the placement of the students as it sharpens the practical skills of the students.	**Approaches:** Bayesian, Multilayer Perceptron Sequential Minimal Optimization (SMO), Ensemble Methods and DT.	**Datasets:** socioeconomic conditions, academic performance, Emotional Skill	**Implementation:** 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.	**Future work** Further employability has been defined as the students' ability to get employment during on Campus drives conducted.	27

2.	To develop models for improving learning experiences and institutional effectiveness.	**Approaches:** statistical and data miming approach, X-Means and Support vector Clustering and Naïve Bayes	**Datasets:** Rank, Gender, Category and Sector and the model will give answer in terms poor Excellent, Good, Avg.	**Implementation:** The algorithms used were implemented and the front end of the tool were developed using PHP and MYSQL as a database.	**Results** X-Means proved to be the best Predicting algorithm representing cluster model, for solving placement chance prediction	81
3.	Employability is a set of achievements, understandings and personal attributes that make individuals more likely to gain and successful in chosen	**Approaches:** K-Nearest Neighbors, Random Forest, DT, Rank estimation using Resumes.	**Datasets:** PrevCo, exp, GPA, Degree/Foreign Degree projects, Achievements, Publications,	**Implementation:** The confusion matrix gives the better idea for the model we are developing it gives the values such as (True Positive), FP, FN, and TN (True Negative).	**Results** It is recommended to collect more proper data of several companies. DB of current employees and past.	85
4.	The survey first explain show higher education has become an exciting field of research and why the prediction of academic performance and employability is beneficial for the institutions.	**Approaches:** Classification, Trees and Regression, CART, ANN, Chi-Squared Automatic Interaction Detection (CHAID)	**Datasets:** Campus placements, ability to job, skills, job within a specified time, period of graduating, willingness to extend graduate at work.	**Implementation:** Implementation is done using the traditional education.	**Future work** The future work includes survey of tools, available for prediction of AP and employability.	94

Data Mining Techniques (J)

S.No	Issue/Objectives	Solution/ Algorithm/ Techniques/ Approaches	Application/ Datasets/ Real-time attributes	Implementations/Tools	Results/ Future work	References
1.	To develop a model which can derive the conclusion on Students' academic success?	**Approaches:** C4.5 , J48, Naïve Bayes, Multilayer perceptron algorithm	**Datasets:** Gender, Distance, GPA, Scholarships, Materials, Grade Importance	**Implementation:** Applied on traditional class room.	**Results/ Future work** The Naïve Bayes, DT, NN methods has been indicated a good model have both accurate, comprehensible.	3
2.	To provide quality education and analyze the performance of students and help them to improve.	**Approaches:** Cluster analysis, Naive Bayes classifier,	**Datasets:** GPA, Workshops, sem marks Conferences, lab Entrance scores, college exams, Attendance.	**Implementation:** RStudio: R (Revolution) for statistical computing and graphics. It provides a variety of statistical and Graphical techniques.	**Future work** K-means helps the students who pursue higher studies in future. Naive Bayes, the various fields a student interested in was identified.	38
3.	Focuses on consolidate the different types of clustering algorithms as applied in Educational Data Mining context.	**Approaches:** Association rules, clustering, sequential pattern, classification, analysis etc.	**Datasets:** final marks, online students marks etc.	**Implementation:** Table.1 shows the comparisons between many types of algorithms.	**Results/ Future work** Various disparate entities that are widely spread across in the educational foray.	41
4.	To pre-recognize student's academic failure.	**Approaches:** (NNge), NB, OneR classification algorithm	**Datasets:** skill-sets of students, short-list the students, arrange exam schedule for students	**Implementation:** The results are done using Weka tool.	**Results/ Future work** real-time project can be used in any edu org for pre-recognizing the failure of students	43
5.	To focuses on designing various	**Approaches:**	**Datasets:**	**Implementation:**	**Results/ Future work**	44

	methods that will help the teachers and the principal (Administrator) of the school to figure out the weak students and improve their educational standards and environment.	NNge, OneR	family, occupation of parents, living with parents, location, illness,	The results are done using Weka tool.	Implemented two rules of induction, DT and NB. Which decreased the complexity and overhead of the system?	
6.	Using c4.5 algorithm, student future exam performance is to be predict based on pass/fail ratio and their rank.	**Approaches:** C4.5, NB, confusion matrix of NB Propagation, NN, Radial Base Supervised Learning.	**Datasets:** assignment, university marks etc. and personal in-formation attributes are like hostler, marital status	**Implementation:** Weka is used for implementing c4.5, naive Bayesian and neural network.	**Results/** Future **work** To help students in the decision making process through our project, and proved neural network is better than naive Bayesian.	61
7.	To improve the quality of the educational processes and to enhance the student's performance.	**Approaches:** Naïve bayes, Multilayer Perception, SMO, J48, REP Tree.	**Datasets:** CGPA, Arrears, Attendance, 12marks, Engineering Cut-off, Medium of Education	**Implementation:** The data collected from Feedback forms and database are entered in excel sheets and converted to ARFF format for further processing in WEKA tool.	**Results/ Future work** We can know the academic status of the students in advance and concentrate to improve their academic results and placements.	68
8.	To identify the techniques to bring co-relation between the student academics and faculty responsibilities.	**Approaches:** Classification, SVM, DT, C4.5, Bayesian, RIPPER.	**Datasets:** ICT, MOODELS, MOOCS	**Implementation:** To find out fitness and accuracy of the algorithm and analyze correct predictive and not correctly predictive rules.	**Results/ Future work** Results are based on how to analyze and predict rules from Educational DB.	79
9.	To measure the performance of students, assessment of students and study students' behavior	**Approaches:** ID3, C4.5, and the Naïve Bayes	**Datasets:** pass/fail	**Implementation:** The results are done using Weka	**Results/ Future work** To identify extraordinary attention to perform well in their discipline	
10.	To develop an application for predicting the student's performance by applying Data Mining Techniques	**Approaches:** Support Vector Machine	**Datasets:** attendance, SSC,HSC or Diploma marks, online test marks	**Implementation:** classification is done by finding the hyper-plane	**Results/ Future work** SVM crucial for cases where very high predictive power is required	87
11.	C4.5improved by the use of L'Hospital Rule, which simplifies the calculation process and improves the efficiency of decision making algorithm.	**Approaches:** C4.5 algorithm	**Datasets:** Information is related to C4.5 algorithm.	**Implementation:** The results are done using tool Weka	**Results/ Future work** C4.5improvesperformance in terms of time saving increased efficiency of lot.	

Machine Learning (J)

S.No	Issue/Objectives	Solution/ Algorithm/ Techniques/ Approaches	Application/ Datasets/ Real-time attributes	Implementations	Results/ Future work	References
1.	To overcome the limitations of the existing techniques; improve reliability and efficiency of instructors' performance evaluation	**Approaches:** Neural networks and decision tree algorithms.	**Datasets:** Teaching, assessment, strengths, weaknesses improvement, class obs,	**Implementation:** 6 criteria's are explained in detail.	**Results/ Future work** Neural networks and decision tree algorithms will be used in search for the best model with high	19

	system,		evaluation report		predictive accuracy	
2.	The main objective is to provide the performance appraisal report of an employee using Decision Tree algorithm	**Approaches:** DT, rule mining, C4.5 clustering**,** J48, ASSISTANT.	**Datasets:** Theory and lab pass %, paper presented, conference participated	**Implementation:** The results are done using Tool Weka.	**Results/ Future work** This study helps to predict whether to give yearly increment, promotion and career advancement for an emp.	30
3.	The main focus is determine a machine learning model for mapping graduates' skills to industry roles using skills profile of emp graduates	**Approaches:** naive Bayes and SVM, K-Nearest Neighbor.	**Datasets:** Related to industrial.	**Implementation:** Using MLT improves skills to industry roles in predicting accuracy for employability, productivity.	**Results/ Future work** Helps in select a suitable machine learning technique that improves performance accuracy.	56
4.	The objective of the educational institution is producing good results in academic exams can be achieved by using the data mining techniques	**Approaches:** Decision Trees, Random Forest, NN and SVM	**Datasets:** Internal Assessment Test(IAT), Assignment, Personal Counseling.	**Implementation:** Apriori is found to be effective in predicting the student under three categories: good, average and poor	**Results/ Future work** The extracted rules help to predict the performance and identify the avg, below avg and good students.	136
5.	To describe the six machine learning algorithms for predicting programming success, using the pre-determined factors.	**Approaches:** Logistic regression, k-nearest neighbor, back propagation, C4.5, naïve Bayes and SVM	**Datasets:** Prev academic experience, computer, game playing, internet, hours studying, job	**Implementation:** The results are done using Tool Weka	**Results/ Future work** Based on accuracy, the successful Algorithms in descending order are NB, SMO, logistic regression, back propagation, C4.5 and 3-NN.	137
6.	To predict the performance of students in an academic organization	**Approaches:** DT(C4.5), Naive Bayes, 3NN RIPPER, WINNOW- NN	**Datasets:** poor, average, good, excellent	**Implementation:** Machine learning has come far from its nascent stages, and can prove to be a powerful tool in academia	**Results/ Future work** The performance of neural networks increases with increase in dataset size	138

III. WEKA TOOL

Weka is Data mining software in java, its open source software issued under the GNU General Public License. Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes. Weka supports several standard data mining tasks, more specifically, data preprocessing, clustering, classification, regression, visualization, and feature selection.

Programming languages:

R, Java, Python, Mat lab, SAS and Rapid Miner.

3.1. Data mining algorithms in Weka

- Classifiers (covers the supervised classification and regression).
- Clusterers (unsupervised learning)
- Associations
- Attribute Selection (evaluators and search methods)
- Preprocessing Filters (supervised and unsupervised data preprocessing).

Categories of Weka algorithms:

1. **Bayes**: Algorithms that use Bayes Theorem, like Naive Bayes.
2. **Function**: Algorithms that estimate a function, like Linear Regression.
3. **Lazy**: Algorithms that use lazy learning, like k-Nearest Neighbors.
4. **Meta**: Algorithms that use or combine multiple algorithms, like Ensembles.
5. **Misc**: Implementations don't neatly fit other, like running a saved model.
6. **Rules**: Algorithms that use rules, like One Rule.
7. **Trees**: Algorithms that use decision trees, like Random Forest.

Weka selects the last attribute in your dataset. If the attribute is nominal, then Weka assumes you are working on a classification problem. If the attribute is numeric, Weka assumes you are working on a regression problem.

This is important, because the type of problem that you are working on determines what algorithms that you can work with. For example, if you are working on a classification problem, you cannot use regression algorithms like Linear Regression. On the other hand, if you are working on a regression problem, you cannot use classification algorithms like Logistic Regression. Fig.1 shows the algorithms which are applicable under Weka tool. Among them our next sections use the j48, neural network, bayes.net, random forest algorithms for comparison.

3.1.2. Weka modules

A collection of machine learning algorithms for solving real-world data mining problems and it also provides data mining functions (regression, association rules and clustering algorithms).

3.1.3. Weka is done using four Modules

1. Explorer: Data processing.
2. Knowledge flow: predictive performances of learning.
3. Command Line Interface: to access Weka functionalities, increments processing of data.
4. Experimenter: scale comparison for predictive performance of learning algorithms.

3.2. Steps to process the WEKA explore

*Step.1***:** Preprocessing Data

Step.2: Uploading a File

Step.3: Once the data is loaded, attributes are shown in the 'Attribute' window.

Step.4: You can visualize the attributes based on selected class by clicking on 'Visualize All' button.

Step.5:Classifiers/Clustering/Associations/Attribute Selection:

If once the data is uploaded click on Classifiers/ Clustering/ Associations/ Attribute Selection then it starts applying the provided algorithms.

Classifiers:

These models are for predicting nominal or numeric quantities.

- Learning classifiers are decision trees and lists, instance-based classifiers, support vector machines, multi-layer perceptions, logistic regression, and Bayes' nets.
- Meta-Classifiers include bagging, boosting, stacking, error-correcting output codes, and locally weighted learning.

Clustering:

Clustering is used to find similar methods, Weka consists these classifiers k-Means, EM, Cobweb, X-means, Farthest First. Same steps are followed to upload data as in classifier.

Once the algorithm is chosen, right click on algorithm, a new dialog box gets opened. Set the value in "numCluster" number it according to the selected attributes and mark to restore the cluster visualization. Then in test options choose classes to cluster evaluation (nom). Now start to run the algorithm.

Associations:

Weka supports Apriori algorithm here, it works only with discrete data and will identify statistical dependencies between groups of attributes. Associate doesn't support the numerical values.

Attribute selection:

It contains two parts: search methods such as best-first, forward selection, random, exhaustive, genetic algorithm, ranking, and an evaluation method such as correlation-based, wrapper, information gain, chi-squared. Test options: Use full training set, cross-validation.

Step.6: Data Visualization
Weka can visualize single attributes (1-d) and pairs of attributes (2-d) rotate 3-d visualizations (Xgobi-style). Weka has "Jitter" option to deal with nominal attributes and to detect "hidden "data points. Here you can select the graphs types (like: rectangle, polygon etc.).

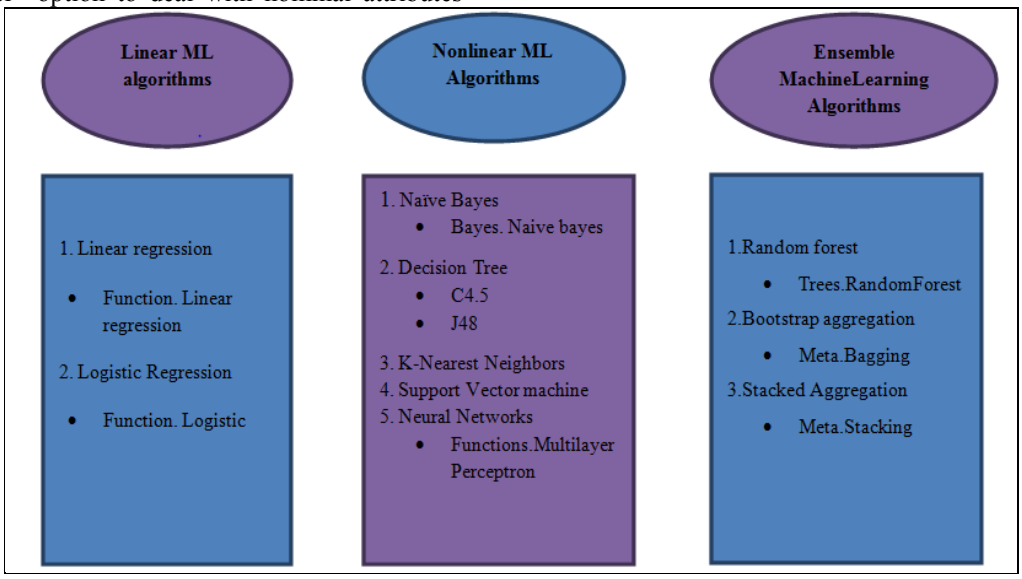

Fig.1: Machine Learning Algorithms under WEKA.

IV. Azure Machine Learning

Azure Machine Learning is a cloud-based predictive analytics service. Azure provides Platform as a Service (PaaS) and Infrastructure as a Service (IaaS), this solution was created by Microsoft in the year 2010. The Cloud Academy has some excellent courses introducing you to the platform. This cloud solves the problem of big data processing.

Azure Machine Learning offers a fairly independent environment to work on. To be precise, in ML Workspace where all ML-related objects live, although we will be able to monitor our ML web services directly from the general Azure dashboard. You can always access your Azure Machine Learning Studio at "https://studio.azureml.net."
The components we have to work on are as follows:

- *Datasets:* These are data containers of our own. We can create a new file from a local file.
- *Experiments:* An experiment is a set of connected components used to create, train, score, and test our model.
- *Trained Models:* Once you the model on which you've experimented, you can save it and can be reuse it in future experiments.
- *Web Services:* In order to obtain new predictions (either online or batch predictions) we'll need to create and query a Web Service.

Supported data sets files are: CSV, TSV, Plain Text (.txt), ZIP (.zip), SvmLight, ARFF, and RData.

4.1. How to create an Azure Machine Learning model:

Step.1: Choose the Data set name and replace the name at top of building block.
Step.2: Data sets> search> select column in dataset (launch the column).
Step.3: Split data (split rows, select the % to split).
Step.4: Machine Learning >Train model (select the column).
Step.5: Machine Learning >Initialize model (select the algorithm to be applied).
Step.6: Score model.
Step.7: Evaluate model.
When we run the experiment we can find the following results that are displayed, here we find model accuracy, precision and recall. More you can graphically visualize how the 6 possible classes have been classified in the testing set by looking at the Confusion Matrix. Fig.2 shows the machine learning approaches in Azure Tool.

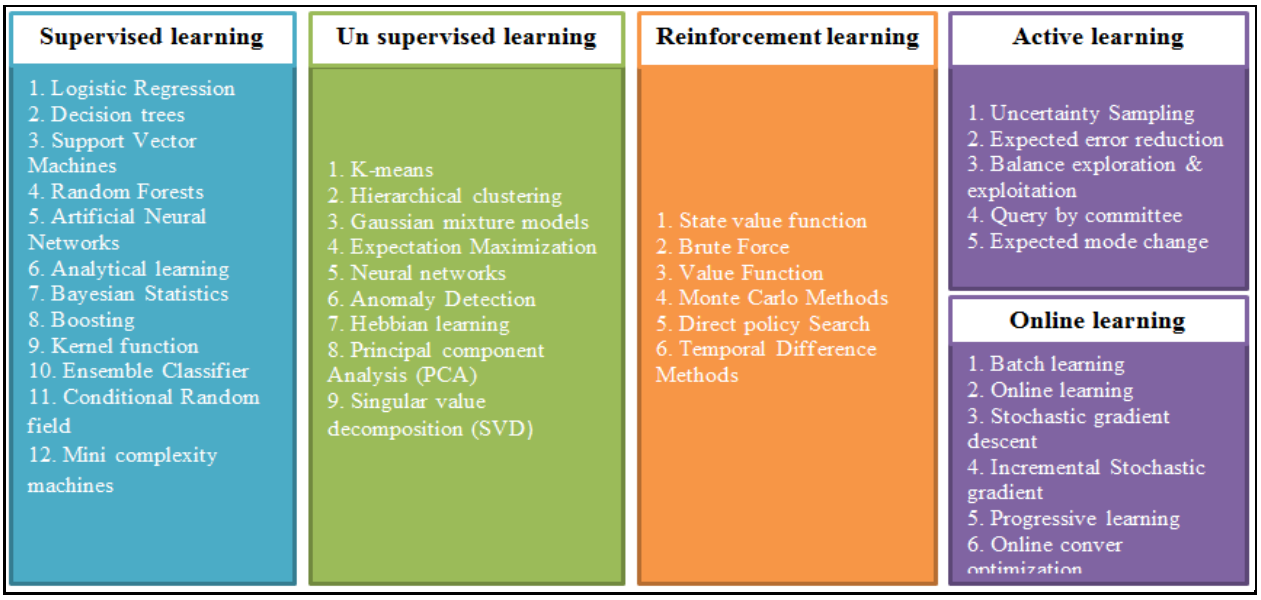

Fig.2: Azure Machine Learning Approaches.

Table.2 shows all the machine learning algorithms in the Azure tool; among these our study goes on three different methods one in Classification (i.e., Two–class neural network), Regression (i.e., Linear Regression) and Clustering (i.e., K-Means Clustering).

Table.2: Azure Machine Learning Algorithms.

Classification	Regression	Clustering
Multi Class Forest	Poisson Regression	K-Means Clustering
Multi Class Decision Jungle	Ordinal Regression	
Multi Class Logistic Regression	Fast Forest Quintile Regression	
Multi Class Neural Networks	Linear Regression	
One-Vs-All Multi Class	Neural Network Regression	
Two-Class Average Percept	Decision Forest Regression	
Two-Class Bayes Point Machine	Bayesian Linear Regression	
Two-Class Boosted Decision	Boosted Decision Tree Algorithm	
Two-Class Decision Forest		
Two-Class Decision Jungle		
Two-Class Locally-Deep SVM		
Two-Class Logistic Regression		
Two-Class Neural Networks		
Two-Class SVM		

Data sets

The following data sets are considered for the algorithm comparisons using Weka and MS Azure tools. The attributes are: School**,** sex, age, address, famsize, Pstatus, Medu, Fedu, Mjob, Fjob, reason, guardian, travel time, studytime, failures, schoolsup, famsup, paid, activities, nursery, higher, internet, romantic, famrel, free time, goout, Dalc, Walc, health, absences, G1, G2,G3,Total, Results.

V. Algorithm Comparisons

5.1. Weka Algorithm results

We have done some experiments using Weka tool in order to evaluate the performance of different algorithms for predicting students' success. The results of the experiments are summarized in the Table.2. The table consists of Time taken (TT), correctly classify instance (CCI), in correctly classify instance (ICCI), kappa static (KS), mean absolute error (MAE), Relative absolute error (RAE), Root mean squared error (RMSE), Total, Test model. It represents the comparison of performance criteria values of the neural networks, J48, Random forest, Naïve bayes, and Bayes.net.

Table.2: Weka algorithms comparisons

Algorithms	TT	CCI	ICCI	KS	MAE	RAE	RMSE	Total	Test model
J48	0.17 sec	394 - 99.7468%	1 - 0.2832	0.9948	0.005	10.19%	0.5	395	NA
Random forest	0.93 sec	134 100%	0 0%	1	0.0502	10.35%	0.0861	134	66.00%
Naïve bayes	0.02 sec	372 94.1772%	23 5.8228	0.8804	0.00649	13.3956	0.1958	395	NA
Neural networks	9.19 sec	NA	NA	NA	NA	NA	NA	NA	NA
Bayes.net	0.02 sec	394 99.746 8%	1 0.2832%	0.9948	0.0028	0.58%	10.22659	395	NA

Visualization:

The results of neural network in graphical way are shown in fig.3, as we increase datasets the size of this gets increased. Fig.4 is the result of j48 it displays pass percentage at one side of tree and failure at other side.

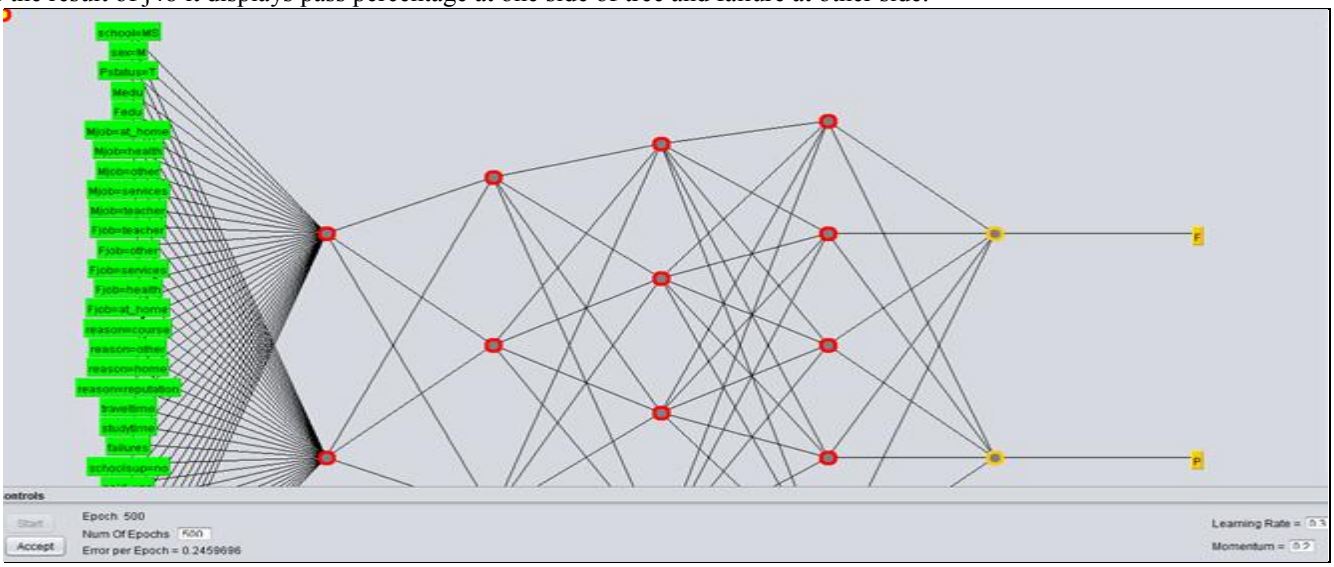

Fig.3: Neural network with Multilayer Perceptron

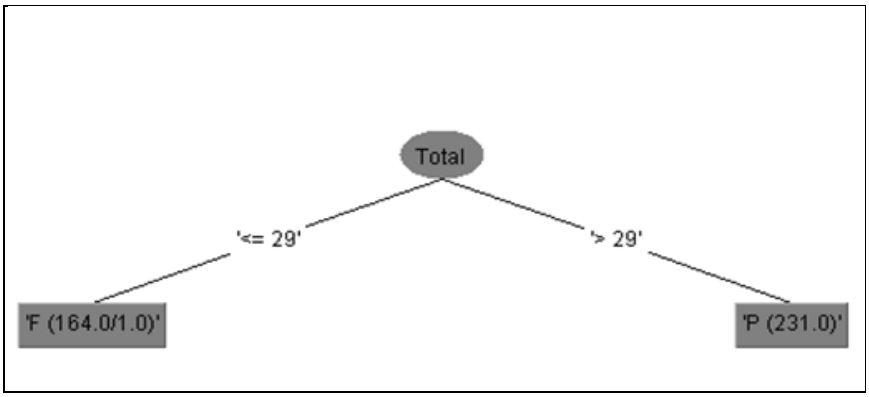

Fig.4: Tree j48 visualization

5.2. Azure Algorithm results

We have done some experiments using Azure tool as well to evaluate the performance of different algorithms for predicting students' success. The results of the experiments are summarized in the Table.3. The table consists of Mean, median, max, min, standard deviation, unique value, missing value, feature type. It represents the comparison of performance criteria values of the two locally neural networks, K- means clustering, linear regression.

Table.3: Azure algorithms comparisons

	Azure Machine Learning Algorithms							
Algorithms	Mean	Median	Min	Max	Standard Dev	Unique Val	Missing Val	Feature Type
Linear Regression	31.6356	32.5	6	54	11.5156	45	0	Numeric Feature
K-means clustering	0	-0.0034	-2.531	2.3439	1.0013	54	0	Numeric Feature
Two locally Neural networks	-0.0195	-0.0034	-2.4411	2.1634	1.0454	48	0	Numeric Label

Visualization of K-means clustering:

The graphical results of K-means are shown in the fig.5 In order to evaluate the performance of the proposed approach, the K-means clustering is one of the methods applied to the studied problem. Coefficients derived in this method are presented.

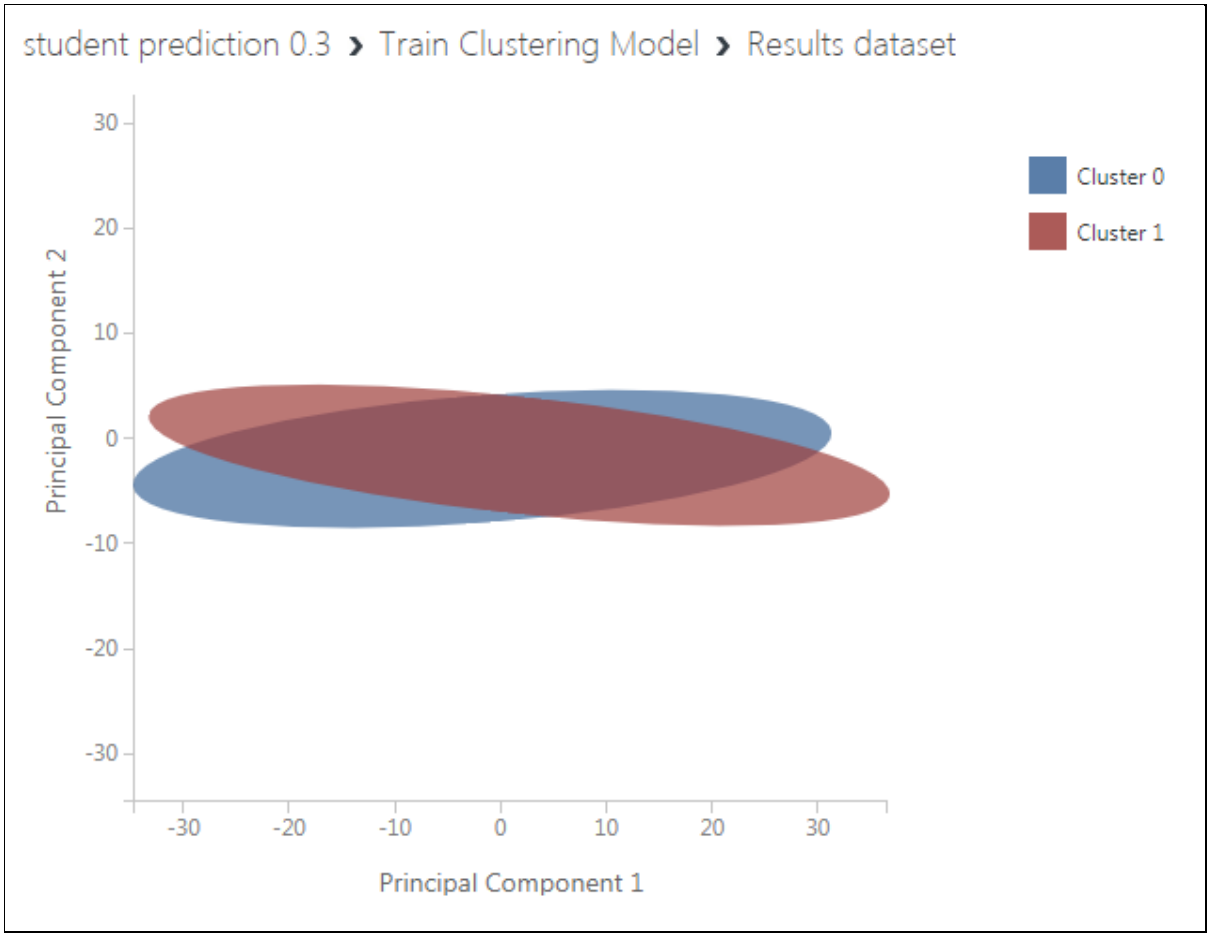

Fig.5: results of train model

VI. Conclusion

The accurate prediction is important as educational services are playing vital role, Present studies shows that academic performances of the students are primarily dependent on their past performances. We have chosen some student past actions for prediction and applied few algorithms to those datasets and the results have been shown. We can say that the performance of neural networks increases with increase in dataset size. Machine learning studio using azure has been best tool for real time applications, and can prove to be a powerful tool in academia. In future, more datasets are taken and prediction is done using k-means and neural networks.

REFERENCES

[1] Mariel F. Musso, Eva Kyndt, Eduardo C. Cascallar, Filip Dochy, "Predicting general academic performance and identifying the differential contribution of participating variables using artificial neural networks," Frontline Learning Research 1,pp.42 – 71, 2013.

[2] A. Dinesh Kumar, Dr.V.Radhika, "A Survey on Predicting Student Performance," International Journal of Computer Science and Information Technologies (IJCSIT), pp. 6147-6149, 2014.

[3] Edin Osmanbegović, Mirza Suljić,"data mining approach for predicting student performance," Economic Review – Journal of Economics and Business, Vol. X, Issue 1, May 2012.

[4] Ahmet TEKIN, "Early Prediction of Students' Grade Point Averages at Graduation: A Data Mining Approach," Eurasian Journal of Educational Research, Issue 54, pp.207-226, 2014.

[5] Mashael A. Al-Barrak and Muna Al-Razgan, "Predicting Students Final GPA Using Decision Trees: A Case Study," International Journal of Information and Education Technology, Vol. 6, No. 7, July 2016.

[6] Bonnie J. Dorr · Craig S. Greenberg · Peter Fontana · Mark Przybocki ·Marion Le Bras · Cathryn Ploehn · Oleg Aulov · Martial Michel ·E. Jim Golden · WO Chang., "A new data science research program: evaluation, metrology, standards, and community outreach," Springer International Publishing Switzerland (outside the USA), 2016.

[7] Young-Jin Lee, "Predicting Students' Problem Solving Performance using Support Vector Machine," Journal of Data Science, pp. 231-244, Issue 14, 2016.

[8] Long Bing Cao, "Data science and analytics: a new era," Int J Data Sci Anal, Springer, Issue 1, pp.1–2, 2106.

[9] Mr. M. N. Quadri, Dr. N.V. Kalyankar, "Drop out Feature of Student Data for Academic Performance Using Decision Tree Techniques," Global Journal of Computer Science and Technology, pp.2, Vol. 10, Issue 2 (Ver 1.0), April 2010.

[10] Ikmal Hisyam Mohamad Paris, Lilly Suriani Affendey, and Norwati Mustapha, "Improving Academic Performance Prediction using Voting Technique in Data Mining," International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol.4, Issue 2, 2010.

[11] Nguyen Thai-Nghe, Lucas Drumond, Artus Krohn-Grimberghe, Lars Schmidt-Thieme, "Recommender System for Predicting Student Performance," Elsevier, pp. 1-9, 2010.

[12] Brijesh Kumar Bhardwaj, Saurabh Pal, "Data Mining: A prediction for performance improvement using classification," International Journal of Computer Science and Information Security (IJCSIS), Vol. 9, Issue. 4, April 2011.

[13] V.Ramesh, P.Parkavi, P.Yasodha, "Performance Analysis of Data Mining Techniques for Placement Chance Prediction," International Journal of Scientific & Engineering Research, Vol. 2, Issue 8, August-2011.

[14] P. Usha, "predicting student performance using Genetic and Svm classifier," international journal of computer engineering (IJCE), Vol. 3, Issue 2, pp. 97–102, July-December 2011.

[15] Diego Garc_a-Saiz, Marta Zorrilla," Comparing classi_cation methods for predicting distance

[16] Students' performance,"JMLR: Workshop and Conference Proceedings 17, 2nd Workshop on Applications of Pattern Analysis, pp. 26-32, 2011.

[17] Sajadin Sembiring, M. Zarlis, Dedy Hartama, Ramliana S, Elvi Wani, "prediction of student academic performance By an application of data mining techniques," International Conference on Management and Artificial Intelligence IPEDR vol.6, 2011.

[18] Erkan Er, "Identifying At-Risk Students Using Machine Learning Techniques: A Case Study with IS 100," International Journal of Machine Learning and Computing, Vol. 2, Issue 4, August 2012.

[19] Bangsuk Jantawan, Cheng-Fa Tsai, "The Application of Data Mining to Build Classification Model for Predicting Graduate Employment, "International Journal of Computer Science and Information Security (IJCSIS), Vol. 11, Issues. 10, October 2013.

[20] Aranuwa Felix Ola, Prof. Sellapan Pallaniappan, "A data mining model for evaluation of instructors' performance in higher institutions of learning using machine learning algorithms," International Journal of Conceptions on Computing and Information Technology, Vol. 1, Issue 2, Dec' 2013.

[21] D. Magdalene Delighta Angeline, " Association Rule Generation for Student Performance Analysis using Apriori Algorithm," The SIJ Transactions on Computer Science Engineering & its Applications (CSEA), Vol. 1, Issue 1, March-April 2013.

[22] Ayinde A.Q, Odeniyi O.A and Sarumi O.A, "Mining Parent Socio-Economic Factors to Predict Students' Academic Performance in Osun State College of Technology, Esa Oke," International Journal of Engineering Research & Technology (IJERT), Vol. 2 Issue 12, December – 2013.

[23] Ayinde A.Q, Dr Adetunji A.B, Bello M and Odeniyi O.A, "Performance Evaluation of Naive Bayes and Decision Stump Algorithms in Mining Students' Educational Data," International Journal of Computer Science Issues (IJCSI), Vol. 10, Issue 4, pp. 1, July 2013.

[24] A.Dinesh Kumar, Dr.V.Radhika, "A Survey on Predicting Student Performance," International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 5 (5), pp. 6147-6149, 2014.

[25] David L. la Red Martinez, Carlos E. Podestá Gomez, "Contributions from Data Mining to Study Academic Performance of Students of a Tertiary Institute, "American Journal of Educational Research, Vol. 2, No. 9, pp.713-726, 2014.

[26] Devikala.D M.phil and Kamalraj.N MCA, M.phil "Data Mining Approaches on Detection of Students' Academic Failure and Dropout: A Brief Survey," International Journal of Computer Trends and Technology (IJCTT) – volume 14 Issue 3 – Aug 2014.

[27] N.Magesh M.E., DR.P.Thangaraj Ph.D., S.Sivagobika, S.Praba, R.Mohana Priya, "Employee performance evaluation using machine learning algorithm," International Journal of Computer Communications and Networks (IJCCN), Vol.4, No.2, April 2014.

[28] Tripti Mishra, Dharminder Kumar, Sangeeta Gupta, "Students' Employability Prediction Model through Data Mining," International Journal of Applied Engineering Research, Vol. 11, Issue 4, pp. 2275-2282, 2016.

[29] Jeng-Fung Chen, Ho-Nien Hsieh and Quang Hung Do, "Predicting Student Academic Performance: A Comparison of Two Meta-Heuristic Algorithms Inspired by Cuckoo Birds for Training Neural Networks," Issue 7, pp. 538-553, 2014.

[30] Mosima Anna Masethe, Hlaudi Daniel Masethe, "Prediction of Work Integrated Learning Placement Using Data Mining Algorithms," Proceedings of the World Congress on Engineering and Computer Science, Vol. 1, pp. 22-24, October 2014.

[31] Camilo E. Lopez G.,Elizabeth León Guzman, Fabio A. Gonzalez, "Data Mining Model to Predict Academic Performance at the Universidad National de Colombia," Twelfth LACCEI Latin American and Caribbean Conference for Engineering and Technology (LACCEI'2014) Excellence in Engineering to Enhance a Country's Productivity pp. 22 - 24, 2014.

[32] S.Rukkumani, Mr.G.Suresh, "Student performance on academic with relative survey of classification and regression algorithms," International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE) Vol. 17, Issue 1,AUGUST 2015.

[33] Pooja Thakar, Anil Mehta, Ph.D., Manisha, Ph.D., "Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue," International Journal of Computer Applications, Vol. 110, Issue 15, pp. 0975 – 8887, January 2015.

[34] Dr. B. Uma Devi, R. Dhanalakshmi, "A Comprehensive Survey of Students Performance Using Various Data Mining Techniques," International Journal of Science and Research (IJSR),pp. 78-96, 2015.

[35] C.R. Durga Devi, "A Survey on forecasting students' performance using EDM," International Journal of Science Technology & Engineering (IJSTE), Vol. 2, Issue 01, 2015.

[36] David L.la Red Martinez, Marcelo Karanik, Mirtha Giovannini, Noelia Pinto," academic performance profiles: A descriptive model based on data mining, "European Scientific Journal, vol.11, Issue.9, March 2015.

[37] Jai Ruby, Dr. K. David, "Analysis of Influencing Factors in Predicting Students Performance Using MLP – A Comparative Study," International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 2, February 2015.

[38] Ashish Dutt, Saeed Aghabozrgi, Maizatul Akmal Binti Ismail, and Hamidreza Mahroeian, "Clustering Algorithms Applied in Educational Data Mining," International Journal of Information and Electronics Engineering, Vol. 5, Issue. 2, March 2015.

[39] Revathy P., Kalaiarasi P., Kavitha J., Madhumita D. A., "Data Mining Approach for Suggesting Higher Education Courses Based on Student's Performance," International Journal of Science & Techno ledge, Vol. 3, Issue 3, 2015.

[40] Ms. Ashna Sethi, Mr. Charanjit Singh, "Data Mining for Prediction and Classification of Engineering Students achievements using Improved Naïve Bayes," International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 6, Issue 7, July 2017.

[41] G. Jayanthi and Dr.V.Ramesh, "Design of Academic Performance Prediction System Using Multi-Layer Perceptron," International Journal of Computer Science and Software Engineering Volume 1, Issue 1, pp. 9-15, 2015.

[42] Solankar Punam Anil, Jagatap Trupti Baban, Rupnawar Sachin Hanumant, Shitole Vibhavari Jayvant, Prof. Kumbhar S. L., "Student Performance Evaluation in Education Sector using Prediction and Clustering Algorithms," IJSRD - International Journal for Scientific Research & Development| Vol. 3, Issue 10, 2015.

[43] Pooja Thakar, Research Scholar, Prof. Dr. Anil Mehta, Dr. Manisha," Role of Secondary Attributes to Boost the Prediction Accuracy of Students' Employability Via Data Mining," International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 6, No. 11, 2015.

[44] Sadaf Fatima Salim Attar, Prof. Y.C Kulkarni, "Precognition of Students Academic Failure Using Data Mining Techniques," International Journal of Engineering Research and General Science Vol. 3, Issue 3, May-June 2015.

[45] Lumbini p. Khobragade, Prof. Pravin Mahadik, "Predicting Students' Academic Failure Using Data Mining Techniques," International Journal of Advance Research in Computer Science and Management Studies, Vol.3, Issue. 5, May 2015.

[46] Mingjie Tan, Peiji Shao, "Prediction of Student Dropout in E-Learning Program Through the Use of Machine Learning Method," IJET, Vol.10, Issue 1, 2015.

[47] Humera Shaziya, Raniah Zaheer, G.Kavitha, "Prediction of Students Performance in Semester Exams using a Naïve bayes Classifier," International Journal of Innovative Research in Science, Engineering and Technology, Vol. 4, Issue 10, October 2015.

[48] Mohammadreza Zahedifard, Iman Attarzadeh, Hadi Pazhokhzadeh, Javad Malekzadeh, "Prediction of students' performance in high school by data mining classification techniques," International Academic Journal of Science and Engineering, Vol. 2, Issue. 7, pp. 25-33, 2015.

[49] Havan Agrawal, Harshil Mavani, "Student Performance Prediction using Machine Learning," International Journal of Engineering Research & Technology (IJERT), Vol. 4, Issue 03, March-2015.

[50] P V V Satya Eswara Rao, S K Sankar, "Survey on Educational Data Mining Techniques," International Journal of Engineering and Computer Science, Vol. 6, Issue 4, pp. 21034-21041, April 2017.

[51] Glyn Hughes and Chelsea Dobbins, "The utilization of data analysis techniques in predicting student performance in massive open online courses (MOOCs)," Hughes and Dobbins Research and Practice in Technology Enhanced Learning, 2015.

[52] Susan Bergin, Aidan Mooney, John Ghent and Keith Quille, "Using Machine Learning Techniques to Predict Introductory Programming Performance," International Journal of Computer Science and Software Engineering (IJCSSE), Volume 4, Issue 12, Page: 323-328, December 2015.

[53] Ms. Priti S. Patel1, Dr. S.G.Desai, "Various Data Mining Techniques used to Study Student's Academic Performance," International Journal of Computer Science and Mobile Applications, Vol.3 Issue. 6, pg. 55-58, June- 2015.

[54] Pimpa Cheewaprakobkit, "Predicting Student Academic Achievement by Using the Decision Tree and Neural Network Techniques," Volume 12, Issue. 2, 2015.

[55] Amirah Mohamed Shahiri, Wahidah Husain, Nur'aini Abdul Rashid, "A Review on Predicting Student's Performance using Data Mining Techniques," The Third Information Systems International Conference,pp.414 – 422, 2015.

[56] Fullgence Mwachoo Mwakondo, Lawrence Muchemi, Elijah Isanda Omwenga, "Automatic Mapping of Graduates' Skills to Industry Roles

[57] Using Machine Learning Techniques: A Case Study of Software Engineering," IJCST Vol. 7, Issue 4, Oct - Dec 2016.

[58] Mojisola G. Asogbon, Oluwarotimi W. Samuel, Mumini O. Omisore, and Bolanle A. Ojokoh, "A Multi-class Support Vector Machine Approach for Students Academic Performance Prediction," International Journal of Multidisciplinary and Current Research, Vol.4, 2016.

[59] Ramanathan l., Angelina geetha, Khalid m., swarnalatha p., "A novel genetic nand paft model for enhancing the student grade performance system in higher educational institutions," IIOAE journal, vol. 7, issue. 5, pp. 1-11, 2016.

[60] P. Kavipriya, "A Review on Predicting Students' Academic Performance Earlier, Using Data Mining Techniques," International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 6, Issue 12, December 2016.

[61] Richa Shambhulal Agrawal, Mitula H. Pandya, "Data Mining With Neural Networks to Predict Students Academic Achievements," IJCST Vol. 7, Issue 2, April - June 2016.
[62] I.A Ganiyu, "Data Mining: A Prediction for Academic Performance Improvement of Science Students using Classification," International Journal of Information and Communication Technology Research, Vol. 6, Issue. 4, April 2016.
[63] Lavannya Varghese, Ms. Christina Joseph, Dr. Vince Paul, "Recommendation System Using Machine Learning and Data Mining Techniques," Vol. 6, Issue. 6, 2016.
[64] Ahmed Mueen, Bassam Zafar, Umar Manzoor, "Modeling and Predicting Students' Academic Performance Using Data Mining Techniques,"I.J. Modern Education and Computer Science, Issue 11, pp. 36-42, 2016.
[65] Neelam peters, Aakanksha s. Choubey, "a survey on data classification and machine learning for forecasting of student performance," international journal of engineering sciences & research technology, december-2016.
[66] Mustafa Agaoglu, "Predicting Instructor Performance Using Data Mining Techniques in Higher Education," IEEE, 2016.
[67] Mashael A. Al-Barrak and Muna Al-Razgan, "Predicting Students Final GPA Using Decision Trees: A Case Study," International Journal of Information and Education Technology, Vol. 6, Issue. 7, July 2016.
[68] Jai Ruby, Dr. K. David, "Prediction Accuracy of Academic Performance of Students using Different Datasets with High Influencing Factors," International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, Issue 2, February 2016.
[69] T.Archana and Usha Devi Gandhi, "prediction of student performance in educational Data mining - a survey," International Journal of Pharmacy & Technology (IJPT), Vol. 8, Issue No.3, Sep-2016.
[70] R. Sumitha, E.S. Vinoth Kumar, "Prediction of Students Outcome Using Data Mining Techniques," International Journal of Scientific Engineering and Applied Science (IJSEAS), Vol. 2, Issue-6, June 2016.
[71] David Otoo-Arthur, Abdulai Jamal-Deen, Ferdinand Apietu Katsriku, "Predictive Modeling and Analysis of Student Academic Performance Using One-Vs-All Logistic Regression Approach," International Journal of Research in Engineering and Applied Sciences (IJREAS),Vol. 6 Issue 12, pp. 81-92, December – 2016.
[72] Akshay Deshpande, Prashant Pimpare, Shashank Bhujbal, Abhishek Kommwar & Prof.Jagruti Wagh," Student Performance Analysis, Visualization and Prediction Using Data Mining Techniques," Imperial Journal of Interdisciplinary Research (IJIR) Vol-2, Issue-5, 2016.
[73] Ramanathan L, Angelina Geetha, Khalid M, Swarnalatha P, "Student Performance Prediction Model Based on Lion-Wolf Neural Network," International Journal of Intelligent Engineering and Systems, Vol.10, Issue.1, 2017.
[74] Manish Kumar," Superiority of Rotation Forest Machine Learning Algorithm in Prediction of Students' Performance," International journal of computer Applications Foundation of Computer Science (FCS), Vol. 137, Issue 7, 2016.
[75] Sheila A. Abaya, Danzel Anerfee D. Orig, Richard S. Montalbo, "using regression analysis in identifying the Performance of students in the board examination," The Online Journal of New Horizons in Education, Vol. 6, Issue 4, October 2016.
[76] Ashwin satyanarayana, Gayathri Ravichandran, "Mining Student data by Ensemble Classification and Clustering for Profiling and Prediction of Student Academic Performance," ASEE Mid-Atlantic Section Conference, 2016.
[77] Carlos villagrá-arnedo, Francisco j. Gallego-durán, Patricia companỹ-rosique, faraón llorens-largo & Rafael molina-carmona, "predicting academic performance from behavioral and learning data," wit conference, vol. 11 issue. 3, 2016.
[78] Binglan Han, Michael J. Watts, "Predicting the Academic Performance of International Students on an Ongoing Basis,"7th annual Conference of Computing and Information Technology Research and Education New Zealand (CITRENZ2016) and the 29th Annual Conference of the National Advisory Committee on Computing Qualifications, July 11-13, 2016.
[79] Sattar Ameri, Mahtab J. Fard, Ratna B. Chinnam, Chandan K. Reddy, "Survival Analysis based Framework for Early Prediction of Student Dropouts," October 2016.
[80] Sourabh Sahu, Prof. Mayank Bhatt, "big data classification of student result prediction," International Journal of Research in Science & Engineering, Vol. 3 Issue: 2 March-April 2017.
[81] P V V Satya Eswara Rao, S K Sankar, "Survey on Educational Data Mining Techniques," International Journal of Engineering and Computer Science, Vol. 6, Issue 4, pp. 21034-21041, April 2017.
[82] Ankita A Nichat, Dr.Anjali B Raut, "Predicting and Analysis of Student Performance Using Decision Tree Technique," International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 4, April 2017.
[83] Dr. Suganthi, G. and Mr. Ashok, M.V., "predicting employability of students using data mining approach," International Journal of Information Research and Review, Vol. 04, Issue, 02, pp.3798-3801, February, 2017.
[84] A. T. M. Shakil Ahamed, Navid Tanzeem Mahmood & Rashedur M Rahman, "An intelligent system to predict academic performance based on different factors during adolescence," journal of information and telecommunication, Vol. 1,Issue 2, pp. 155–175, 2017.
[85] Sana Akhai, Ruchi Karia, Aniket Mahadik, Akshat Shah, Manya Gidwani, "Automated Performance Evaluation System," International Journal of Advance Research, Ideas and Innovations in Technology, Vol. 3, Issue2, pp. 326-329, 2017.
[86] Pooja Thakar, Prof. Dr. Anil Mehta, Dr. Manisha, "A Unified Model of Clustering and Classification to Improve Students' Employability Prediction," I.J. Intelligent Systems and Applications, Vol. 9, pp. 10-18, 2017.
[87] Neeraj Khadilkar, Deepali Joshi, "Predictive Model on Employability of Applicants and Job Hopping using Machine Learning," International Journal of Computer Applications, Vol. 171 Issue. 1, August 2017.
[88] Mudasir Ashraf, Dr. Majid Zaman, Dr. Muheet Ahmed, S. Jahangeer Sidiq, "Knowledge Discovery in Academia: A Survey on Related Literature," International Journal of Advanced Research in Computer Science, Vol. 8, No. 1, Jan-Feb 2017.
[89] Ankita Kadambande, Snehal Thakur, Akshata Mohol, Prof A.M.Ingole, "Predicting Student's Performance System," International Research Journal of Engineering and Technology (IRJET) Vol. 04 Issue: 05, May -2017.
[90] Oyerinde O. D, Chia P. A., "Predicting Students' Academic Performances – A Learning Analytics Approach using Multiple Linear Regression," International Journal of Computer Applications Vol. 157 – No 4, January 2017.
[91] Altyeb Altaher and Omar BaRukab, "Prediction of Student's Academic Performance Based on Adaptive Neuro-Fuzzy Inference," IJCSNS International Journal of Computer Science and Network Security, Vol.17, Issue.1, January 2017.
[92] Prateek Sakaray, Snehal Kankariya, Chandini Lulla, Yash Agarwal, Pankaja Alappanavar, "Review on Student Academic Performance Prediction using Data Mining Techniques," International Journal of Advanced Research in Computer and

Communication Engineering (IJARCC),Vol. 6, Issue 2, February 2017.
[93] Karan Manchandia, Shweta Kondla, Vasudev Lambhate, "review paper on student performance evaluation through Supervised learning using neural network," international journal of engineering sciences & research Technology (IJESRT), March-2017.
[94] Chandini Lulla, Yash Agarwal, Snehal Kankariya, Prateek Sakaray, Pankaja Alappanavar, "Student Academic Performance Prediction using Machine Learning and Data Mining Techniques," International Journal of Computer Science and Mobile Computing(IJCSMC), Vol. 6, Issue. 5, pp.301 – 307, May 2017.
[95] Vinaya Patil, Shiwani Suryawanshi, Mayur Saner and Viplav Patil, "Student Performance Prediction Using Classification Data Mining Techniques," international journal for research in emerging science and technology, Vol -4, Issue-4, APR-2017.
[96] Tripti Mishra, Dharminder Kumar and Sangeeta Gupta, "Students' Performance and Employability Prediction through Data Mining: A Survey," Indian Journal of Science and Technology, Vol 10, June 2017.
[97] Ali Daud, Naif radi aljohani, Rabeeh Ayaz Abbasi, Miltiadis D Lytras, Farhat Abbas, Jalal s.Alowibdi, "Predicting Student Performance using Advanced Learning Analytics," International World Wide Web Conference Committee (IW3C2), April 3-7, 2017.
[98] Barnabas Ndlovu Gatsheni, Olga Ngala Katambwa, "The design of predictive model for the academic performance of students at University based on machine learning," Int'l Conf. Artificial Intelligence,2017.
[99] Nguyen Thai-Nghe, Zeno Gantner, and Lars Schmidt-Thieme, "A New Evaluation Measure for Learning from Imbalanced Data," IEEE.
[100] Mr. Bhushan S. Olokar, Prof. Ms. V.M.Deshmukh, "Application of Data Mining Technique for Prediction of Academic Performance of Student a Literature survey," International Journal on Recent and Innovation Trends in Computing and Communication, Vol.2 Issue. 12.
[101] Thendral Puyalnithi, Madhu Viswanatham V and Mithilesh Kumar Singh," Prediction of Students' Academic Performance based on their lifestyle through Machine Learning Methods," international journal for research in emerging science and technology, Vol-3, Issue-5, MAY-2016.
[102] S.K. Althaf Hussain Basha, Y.R. Ramesh Kumar, A. Govardhan and Mohd. Zaheer Ahmed, "Predicting Student Academic Performance Using Temporal Association Mining," International Journal of Information Science and Education, Vol. 2, Issue 1 pp. 21-41, 2012.
[103] Zlatko J. Kovacic, "Predicting student success by mining enrolment data," Research in Higher Education Journal, pp. 1-20.
[104] Nidhi Arora 1, Jatinder Kumar R. Saini, "Predicting Student Academic Performance using Fuzzy ARTMAP Network," International Journal of Advances in Engineering Science and Technology, Vol. 3, Issue 3, pp. 187-192.
[105] L.S. Affendey, I.H.M. Paris, N. Mustapha, Md. Nasir Sulaiman and Z. Muda, " Ranking of influencing Factors in Predicting Student's Performance," Informational Technology Journal 9, pp. 832-837, 2010.
[106] Jay Bainbridge, James Melitski, Anne Zahradnik, Eitel J. M. Lauría, Sandeep Jaya prakash, and Josh Baron, "Using Learning Analytics to Predict At-Risk Students in Online Graduate Public Affairs and Administration Education, "Journal of Public Affairs Education, pp. 247-262.
[107] Jaya Srivastava, Dr Abhay Kumar Srivastava, "Data Mining in Education Sector: A Review," National Conference on Cloud Computing & Big Data, pp. 184-190.
[108] Wenjun Zeng, Si-Chi Chin, Brenda Zeimet, Rui Kuang, Chih-Lin Chi, "Dropout Prediction in Home Care Training," 10th International Conference on Educational Data Mining, pp. 442-447.
[109] Norman Poh, Ian Smythe, "To What Extend Can We Predict Students' Performance? A Case Study in Colleges in South Africa".
[110] Dorina Kabakchieva, "Predicting Student Performance by Using Data Mining Methods for Classification," cybernetics and information technologies, Vol. 13, Issue 1, 2013.
[111] Cristobel Romero, Sebastian Ventura, Pedro G. Espejo and César Hervás, "Data Mining Algorithms to Classify Students".
[112] R Rathipriya1, Dr. T. Abdul Razak, "predicting the model for academic performance using classification techniques," International Journal of Computer Engineering and Applications, Vol. VIII, Issue I, pp. 46-54, October 14.
[113] Emmanuel N. Ogor, "Student Academic Performance Monitoring and Evaluation Using Data Mining Techniques," IEEE, pp. 354-359.
[114] Tiffany Y. Tang and Gordon McCall, "Student Modeling for a Web-based Learning Environment: a Data Mining Approach," pp. 967-968.
[115] N. Venkatesan, N. Chandru, "Student's Performance Measuring using Assistant Algorithm," International Journal of Soft Computing and Engineering (IJSCE), Vol. 3, Issue-5, November 2013.
[116] Emily H. Thomas, and Nora Gal ambos, "WHAT SATISFIES STUDENTS? Mining Student-Opinion Data with Regression and Decision Tree Analysis," Research in Higher Education, Vol. 45, Issue. 3, May 2004.
[117] Shradha Shet1, Gayathri, "Approach for Predicting Student Performance Using Ensemble Model Method," International Journal of Innovative Research in Computer and Communication Engineering, Vol.2, Issue 5, pp. 161-169, October 2014.
[118] Yi-Chun Chang a,*, Wen-Yan Kao a, Chih-Ping Chu a, Chiung-Hui Chiu, "A learning style classification mechanism for e-learning," Elsevier, pp. 273-285, 2009.
[119] Driyani Rajeshinigo 1, J. Patricia Annie Jebamalar, "Educational Mining: A Comparative Study of Classification Algorithms Using WEKA," International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 3, pp. 5583-5589,March 2017.
[120] Brijesh Kumar Bhardwaj, Saurabh Pal, "Data Mining: A prediction for performance improvement using classification," (IJCSIS) International Journal of Computer Science and Information Security, Vol. 9, No. 4, April 2011.
[121] Fadhilah Ahmad, Nur Hafieza Ismail and Azwa Abdul Aziz, "The Prediction of Students' Academic Performance Using Classification Data Mining Techniques," Applied Mathematical Sciences, Vol. 9, Issue. 129, pp. 6415 – 6426, 2015.
[122] Elizabeth Ayers1, Rebecca Nugent1, and Nema Dean, "A Comparison of Student Skill Knowledge Estimates," Educational Data Mining, 2009.
[123] Suchita Borkar1, K. Rajeswari, "Predicting Students Academic Performance Using Education Data Mining," International Journal of Computer Science and Mobile Computing (IJCSMC), Vol. 2, Issue. 7, pg.273 – 279, July 2013.
[124] Sharon Hardof-Jaffe, Arnon Hershkovitz, Hama Abu-Kishk, Ofer Bergman3, Rafi Nachmias, "How do Students Organize Personal Information Spaces?," Educational Data Mining, pp. 250-258, 2009.

[125] "A Model for Predicting Students' Academic Performance using a Hybrid of K-means and Decision tree Algorithms," International Journal of Computer Applications Technology and Research, Vol. 4, Issue 9, PP. 693 - 697, 2015.
[126] Hamza O. Salami, Esther Y. Mamman, "A Genetic Algorithm for Allocating Project Supervisors to Students," I.J. Intelligent Systems and Applications, Issue 10, pp.51-59, 2016.
[127] Chien-Sing Lee, "Diagnostic, predictive and compositional modeling with data mining in integrated learning environments," Elsevier, pp. 562-580, 2007.
[128] Ioanna Lykourentzou, Ioannis Giannakos, Vassilis Nikolopoulos, George Mpardis, Vassili Loumos, "Dropout prediction in e-learning courses through the combination of machine learning techniques," Elsevier, pp. 950-965, 2009.
[129] Noboru Matsuda, William W. Cohen, Jonathan Sewall, Gustavo Lacerda, and Kenneth R. Koedinger, "Predicting Students' Performance with Sim Student: Learning Cognitive Skills from Observation," pp. 467-476, 2007.
[130] Oyerinde O. D., Chia P. A., "Predicting Students' Academic Performances – A Learning Analytics Approach using Multiple Linear Regression," International Journal of Computer Applications, Vol. 157 – No 4, pp. 37-44, January 2017.
[131] Amjad Abu Saa, "Educational Data Mining & Students' Performance Prediction," (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 5, 2016.
[132] Khairul A. Rasmani · Qiang Shen, "Data-driven fuzzy rule generation and its application for student academic performance evaluation," Springer, pp. 305-319, 2006.
[133] Cristobel Romero, Sebasti´an 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. 40, Issue. 6, NOVEMBER 2010.
[134] Shaleena K.P, Shaiju Paul,"Data Mining Techniques for Predicting Student Performance," IEEE International Conference on Engineering and Technology (ICETECH), 20th March 2015.
[135] Dorina Kabakchieva, "Student Performance Prediction by Using Data Mining Classification Algorithms," International Journal of Computer Science and Management Research, Vol. 1, Issue 4, November 2012.
[136] D. Magdalene Delighta Angeline "Association Rule Generation for Student Performance Analysis using Apriori Algorithm," The SIJ Transactions on Computer Science Engineering & its Applications (CSEA), Vol. 1, No. 1, March-April 2013, pp. 12-16.
[137] Susan Bergin & Ronan Reilly "Predicting introductory programming performance: A multi-institutional multivariate Study," Computer Science Education, Vol. 16, No. 4, December 2006, pp. 303 – 323.
[138] Havan Agrawal, Harshil Mavani, "Student Performance Prediction using Machine Learning," International Journal of Engineering Research & Technology (IJERT) Vol. 4 Issue 03, March-2015, pp. 111-113.

Authors Profile

Dr. K. K. Baseer obtained his Bachelor of Technology and Master of Technology degrees in Computer Science and Engineering from JNTUH, Hyderabad and Ph.D. degree from JNTUA University, Ananthapuram, India. At present working as an Associate Professor in department of Information Technology, Sree Vidyanikethan Engineering College, Tirupati, A.P., INDIA. His areas of interest include Data Science, Software Engineering, Software Architecture, Service Oriented Architecture, Internet of Plants (IoP) and other latest trends in technology. He has more than 10 years of experience in both teaching and industry in the area of Computer Science and Engineering.

C. Jayasree obtained her Masters of Technology degree in Information Technology from Sree Vidyanikethan Engineering College, Tirupati, A.P., India. Her areas of interest Data mining, Machine learning and software engineering.

