A Study of Application of Data Mining and Analytics in Education Domain

Sahil P. Karkhanis
B.Tech in Computer Science and Engineering student
Vellore Institute of Technology

Shweta S. Dumbre, PhD
Project Manager
Larsen and Toubro InfoTech
Ltd., Mumbai

ABSTRACT
Data Mining techniques and algorithms have been used on a large scale in almost all the sectors which range from computer science, manufacturing industry, and healthcare industry. A recently introduced concept of academic analytics uses the data mining algorithms on the educational data of students and gives certain insights about the expected performances of the students, expected retention rate of students and percentage of resources properly utilized. These results also help administrators in decision making and answering certain questions like whether the faculty v/s student's ratio is giving satisfactory results or there is a change needed in the teaching methodology. Educational Data Mining is also another upcoming field and is an allied field of Academic analytics, but it focuses on the data mining algorithm outputs being given back to the faculties in order to properly assess the student's performance. Educational data mining basically helps the tutors modify their teaching strategies if the results with the current teaching model are not satisfactory. This paper basically is a study of certain research experiments which aim to apply data mining algorithms to educational data and contribute to the field of Academic analytics and Educational Data Mining.

General Terms
Machine Learning, Data Mining.

Keywords
Academic Analytics, Educational Data Mining, Learning Analytics, Classification.

1. INTRODUCTION
Business analytics (BA) includes the skills, technologies, applications and practices used for continuous iterative exploration and investigation of past business performance to gain insight and drive business planning. Different techniques used for BA include OLAP, data mining techniques, web analytics and statistical analytics. BA is being used in the banking, manufacturing, healthcare, telecommunication industries in order to extract knowledge from the enormous amount of data being recorded every day. BA is used to cater to customer requirements and provide optimal services to enhance profits. The different types of BA include descriptive analytics, predictive analytics, prescriptive analytics, diagnostic analytics and decisive analytics.

Social networking pioneer Facebook uses the pages liked by every user to suggest people he might know or be interested in. Google uses web history and recent searches to recommend related web pages. BA technology leaders include SAP, Oracle, IBM, SAS Institute and Microsoft. The concept of Video analytics is also being used to constantly keep a track of the surrounding events and then use this data to make the enhance decision making. A new upcoming trend in the field of BA is of crowdsourcing which includes putting information related to business tasks up on the web. Interested skilled people attempt to find the answers to the problem and then revert back to the company through the crowdsourcing websites. The company can check all the answers submitted by the different participants and then go ahead with the most satisfactory answer.

Recently, data mining techniques are being used on educational data and this gives rise to the domains of academic analytics and learning analytics. The main difference between the domains of academic analytics and learning analytics is the stakeholders which benefit from these procedures. Academic analytics aims at extracting knowledge from the educational data which relates to the operations of the administrators and high level authorities of the educational institution. Answers to queries like what is the retention rate of the institution, factors which determine the retention rate, availability of the resources and their usage statistics are the parameters of interest to the above mentioned stake-holders which can be catered to by academic analytics. Learning analytics on the other hand is a field which aims to benefit the students as well as the tutors. It deals with analyzing the academic performance of students in order to find their weaknesses and the methodology of teaching which appeals to them the most. This analysis can then be used to improve the teaching methods and in turn help students grasp content in a better way.

The research paper aims to study the different applications of data mining techniques on the educational da-ta. The research paper is further arranged as follows. The second section elaborates upon the state of the art techniques in the field of educational analytics. The third section briefly describes the various research works published in the field of educational analytics. The fourth section is the summary of the various studies which have been carried out on educational data till date. The fifth section is the conclusion which sheds light on the various future enhancements in which data mining techniques can be applied on educational data to get proper insights and aid in decision making.

2. CURRENT STATE OF ART
A large number of experiments and models have been developed and tested on educational data in order to extract knowledge about the quality of learning and related parameters. The models are being further re-searched upon to increase the success rate of prediction of the target parameters in turn helping the students in acquiring knowledge in an easier way, to aid faculty in understanding the progress of

their students. The bene-fits of educational analytics include aiding in predicting the success rate of students, factors which results in students opting for a transfer to other universities, generating hierarchical list of timetables with an increasing order of difficulty based on the prior knowledge of the student. The educational data needs proper preprocessed before being fed as input to the model. This is because the model is trained first in order to be put to use. Usually the data is divided into two sets, training set and test set. The training set is used to train the model and the test data is then used to check the accuracy of the resultant model.

Various data mining approaches which include classification algorithms, regression algorithms, and association rule mining algorithms, clustering algorithms are being used to extract knowledge from the educational data. Many data analytics tools like SPSS Clementine 12.0, Rapid Miner, Weka data mining tool, R data mining software tool etc. are widely being employed for mining educational data.

3. STUDY OF ANALYTICS IN THE EDUCATIONAL DOMAIN

The different data mining techniques used for mining educational data are explained in the next section. In the next section the research papers have been grouped according to the usage of these techniques for the purposes of curriculum and timetable scheduling and predicting student's performance. Then the next section describes the software tools used for performing data mining on educational data.

3.1 Data Mining Techniques

This research paper [1] elaborately explained the various techniques of data mining which were used on educational data. The paper simplified three techniques used in data mining which were prediction, clustering and relationship mining. Prediction involved predicting certain target parameters based on related information which influenced the target parameters. Classification, Regression and Density Estimation were the three sub techniques under the prediction approach. Classification classified a parameter into a known set of labels. Regression techniques were used on continuous variables of which the classification labels were not known. In density estimation, the predicted variable was a probability density function. The clustering approach basically involved grouping similar data based on a certain criteria. Clustering was used when the most common categories within the data set were not known in advance. The third approach, relationship mining included four sub types which were association rule mining, correlation mining, and sequential pattern mining and casual data mining. Association rule mining involved formation of certain rules based on the variables in the input data in the form of X=>Y. Correlation mining involved finding the correlation between the variables and hence was used to find the most strongly connected attributes in a dataset. Sequential pattern mining was used to find patterns in the sequence of occurrences of events. Casual data mining aimed to find out whether an event Y was triggered by the occurrence of the event X. Outlier analysis, text mining and social network analysis were other methods which were talked about in the paper to extract meaningful data from the educational data sets.

The various research papers published [2] using different mediums of delivering knowledge were analyzed. Certain objectives of educational data mining in which data mining techniques could be employed were listed. Analysis and visualization of data, providing feedback to faculties, student

modelling were the discussed. The various data mining techniques which were employed for these objectives were listed.

The particle swarm optimization algorithm [3] was used to cluster students based on certain parameters. The main drawback of the clustering algorithms like k-means, centroid clustering was the random selection of initial centroids and their inability to deal with continuous arrival of data. Hence an unsupervised clustering algorithm - Particle Swarm Optimization (PSO) was employed for clustering students based on continuously arriving data. The experiment considered a set of M students who were clustered in a set of P clusters. Each cluster had a centroid which had certain values of parameters associated with that cluster. Each student also had certain parameters which included student ID, accuracy, efficiency and quality. This algorithm ceased when all the students were successfully a part of one of the P clusters. The data was of 500 students over a total of 60 problems. Then k-means algorithms and PSO were run using this data and the results were compared based on the criteria - intra cluster distance, inter cluster distance and accuracy. It was found out that the PSO algorithm outweighs the k-means algorithm in all the categories mentioned above and was hence considered as superior to K-means algorithm.

The paper [4] worked on finding whether a difference exists between the sequences of computer interactions that Field Dependent (FD) and Field Independent (FI) learners employed during complex problem solving as well as in terms of their problem solving performance. A total of 101 participants whose average age was 18.56 years participated in the experiment. All the students were first subjected to a test and then categorized into FD, Field Moderate (FM) or FI category. Then these students were given a complex problem to solve, using the Model-It software and then their actions were recorded. The data from only 60 students was taken for analysis as the research focus was on FD and FI students. K-means algorithm and association rule mining techniques were used on the data. The results showed that the performance of the FI and FM learners was better than FD learners.

3.2 Usage of Data Mining Techniques

3.2.1 Predicting student's academic performance

An experiment [5] was conducted to predict the performance of middle school students and dropout rate in Zacatecas Mexico. The input data was extracted from three sources, a survey which was conducted for all the first year students, CENEVAL an agency also gathered certain information about the students and the department of school services provided certain additional parameters. The parameters which were gathered included number of students in a group, attendance during morning and evening sessions, number of hours spent studying daily. The selection algorithms available in the software tool WEKA were used to find out the 15 most influential attributes amongst the 80 whose data was available. Then 10 classification algorithms were applied on these 15 attributes to find out the output in the form of association rules. Also SMOTE algorithm was used to balance the data in order to get more accurate results. These 10 classification algorithms were also applied on all the 80 attributes and the results were analyzed. The first experiment gave a more accurate prediction of the progress and the results of the student due to more emphasis on the most influential attributes and the SMOTE approach to balance the data.

The results of a set of 60 students in their 5th semester were predicted based on their performance in the previous semesters [6] in this experiment. The student's marks i.e. the total marks of the 4 semesters which were out of 1030 were recorded. The decision tree algorithm was used on the data. Its results incorrectly predicted 5 students as pass and 1 student as fail. Then the one R algorithm was applied on the same data. The results were the same as that of the decision tree algorithm. In depth analysis was conducted which involved re-searching about certain parameters like time taken to evaluate the model and true and false positive values obtained. It was found that in both the categories the one R algorithm performed better than the decision tree algorithm and hence was concluded as superior among both the algorithms.

The Open Academic Analytics Initiative (OAAI) [7] [8] [9], supported by a grant from EDUCAUSE's next generation aimed at answering three research questions: What are potential challenges, solutions and benefits associated with developing a completely open-source early alert solution for higher education? To what degree can predictive models be imported from the academic context (e.g., a four-year private liberal arts college) in which they were developed to new and potentially very different academic contexts (e.g., two-year community colleges)? What intervention strategies are most effective in helping academically at-risk students succeed?

The student event data generated from the Sakai collaboration and learning environment was analyzed using the Pentaho business intelligence suite. Initial set of predictive models were developed and then executed on the open source WEKA and IBM's SPSS modeler. To explore the effectiveness of the models pilot runs were executed on three colleges, college of redwoods, Cerritos College and Savannah State University. Similar techniques were used to record data and then analyze it, which were used by John Campbell at Purdue University in a similar approach. The data recorded from Purdue University as well as Marist University was compared and their correlation and influence to the final target parameters was found out. The attributes which were considered by Marist University also showed a certain influence in the model followed at Purdue and hence it was concluded that these models could be portable to a certain extent. Student demographic and academic data was collected from the SAKAI platform. Academic risk was set as the target attribute. This data was partitioned into training data and test data in the ratio of 70% to 30%. Then the data was balanced to give enough focus on the students weak in academics.

Three classifiers for comparison purposes were selected which were logistic regression, support vector machines and C4.5 decision trees. These trained models were used to evaluate the results of these three algorithms by taking into consideration the sensitivity and specificity parameters. The experiment included the data being randomly distributed in 5 sets by varying the random seed. After balancing the data these 5 sets were run on each of the three algorithms giving out results for the 15 runs. Logistic regression and support vector machines algorithm gave satisfactory results with an accuracy rate of around 90%.

After the phase one results were out the models were further enhanced by adding additional data through gradebook data from CMS. Also the models were run thrice every semester by taking into account the level of intervention in the academic preparation of students after completion of 25%,

50% and 75% of the semester. As the model was run these three times the accuracy of the results became more accurate.

Two approaches were decided to deal with at-risk students, which were awareness messaging approach and the next one was online academic support environment (OASE) approach. The awareness messaging approach included alerting the student and suggesting him certain ways to improve which included meeting the tutor, attending a study group session. The OASE approach followed the awareness messaging approach but it also included an invitation to the Sakai based online support system which gave students access to open educational resources. The results showed that there was a significant improvement in the treatment group as compared to the control group. Another analysis showed that the at-risk students who had received at least one intervention showed significant improvement as compared to the control group students. Chi square analysis was used to test the students in terms of content mastery and retention rate. Results showed that in terms of content mastery treatment students achieved better learning outcomes than the control group and students in the treatment group had lower retention rates as compared to control group.

The research objective of the experiment [10] was to get a grip on the retention rate of students from the college of Mumbai 2007 to 2010. The data mining techniques used for the data analysis were the naïve Bayesian algorithm for classification, the k-means clustering for clustering purposes and multiple linear regression technique for prediction purposes. Two set of data were studied, one the student data from 2007-2009 and the student data from 2010-2012. The analysis results after running the algorithms on XL miner software showed the retention rate around 93.0% percent for the first data set and 95.5% for the second data set. The software used for the study was a demo version and could accommodate only 600 records, this was one of the major draw-backs of the study. Based on this data as input the accuracy of the results to the actual retention was around 75-80%.

3.2.2 Curriculum and Timetable Scheduling
The Apriori algorithm [11] was used to find the correlations between certain subjects like Mathematics, Programming, Database, Networks, English and Project. The data of the students who failed to perform well in these subjects was collected. Then using the apriori algorithm, association rules were formed. Apriori algorithm was used to find out relations between two or more entities in a particular data environment. Using the support and confidence parameters of these rules, strong association rules were recognized. These rules showed that if a particular student failed to perform well in a certain subject X, then he was bound to fail in the subject Y. This was because the subject Y required certain pre-requisites which were taught in subject X.

This study [12] basically involved taking variables of first year students at the Iran Institute of Science and Technology which included unfinished units, semester education status ID, failed units, passed units and effective units. Then the techniques of regression analysis and the C5.0 a type of decision tree algorithm were used to find association rules. The SPSS Clementine 12.0 software was used to conduct the experiment. The result was a set of analytical graphs which showed the relationship between the set of variables entered as input. This relationship between variables helped in the

formation of certain association rules in the form of Antecedent => Consequent.

The concepts of correlation [13] and decision trees [14] were employed to compare subjects provided at a particular institute. The correlation results gave an idea about the related subjects in a curriculum of a particular program. Further this helped in formation of level wise format of offering subjects for a particular program. This concept of correlation also gave information about which subjects are negatively correlated and this parameter could also be helpful in the arrangement of subjects for a particular semester for a particular program. The strengths of a particular student were also taken into consideration while the analysis of subjects was done and a list of subjects was recommended to the student. Further focus was upon preparing the timetables for students based on their strengths and weak-nesses by comparing their performances in the basic courses. This helped in balancing the load of academics on the students for a particular semester and through-out the tenure of the program.

The second experiment which was conducted kept a record of the subjects chosen by all the students in every semester of their program. The experiment started with the process of data preprocessing which included taking all the data input. Then for all these different paths the decision trees were formed. These decision trees stored all the different subjects in the order which they had been cleared, the number of students who had followed the path, the overall performance of the student who took that route was recorded. This analysis was then used to guide the students currently pursuing the program to guarantee better results. This experiment helped students to form a proper map to complete the list of courses in a proper order. It also helped students who had chosen a relatively new track to complete the program get back on track.

The 'reinforcement learning algorithm' [15] based time-table scheduling model was demonstrated in this pa-per. The different constraints which are encountered when a timetable needs to be scheduled were listed. These complications included number of hours required per week, whether the lectures need to continuous compulsorily or separate, type of classroom required for teaching and other academic related constraints. Odd and even week timetables are used by certain institutions. Some institutions also employed different number of hours per week for the same course depending upon the program's requirement. The reinforcement learning algorithm took in all the different constraints for each and every course which had to be scheduled. In this study, the principle of five teaching days a week and eight teaching periods a day was used. The Eigen vector concept was used for a particular subject and it consisted of four components - teaching period arranged, weeks required to teach the course, type of classroom and ratio of seats to students.

The first step in the procedure of the model was to sort out the timetable historic data for a certain period. Then those timetables were analyzed and the model was trained according to the rules used in the earlier timetables, formed according to the expert's rules. The next step was to calculate the eigenvector and the timetable scheduling data in a series. Next a timetable was extracted from the calculated eigenvectors. From the time-table mode all the different parameter related information such as the grades, course, teachers, majors, and classes was then extracted. The Naïve Bayesian algorithm was then used to calculate the order and schedule the timetable depending on the input eigenvectors of the vectors. Based on

this the courses were sorted in one of the categories timetable for routine courses, timetable for art courses, timetable for courses with special requirements, timetable for missed lessons, timetable for others , timetable for teachers with special requirements

3.3 Software tools for learning analytics

In this research paper [16] a new 'eLearning' software was proposed which aimed at grouping the data from the course management system employed by an educational institute. The data from various sources that includes the student who opted for the classroom teaching method as well as those learning online was combined, transformed and stored. The resultant data set was called 'elearndataset.data'. The data preprocessing and transformation was done by using the rapid miner software tool. Then k-means clustering algorithm was employed to cluster data into groups which included common words like mining, post-doctoral, introduction etc. Then these clustered groups of data were classified into web based, e-class room based, network based, and virtual class room. These classified sets were assigned a key index value for high speed data retrieval. The aim of the experiment was to classify the data in a way such that the users could find data and classify it with optimum accuracy value. The accuracy rate in the experiment was found out to be 89.988%.

The importance of data mining techniques in the field of education in order to enhance student understanding was highlighted in this research paper [17]. A list of analytical software tools used by universities was pro-vided. Purdue University used the signals course management system with features like signal feedback, self-notification, and early feedback in order to keep a check on the progress of its students. The Sakai institute employed the learning management system which bran-dished features like predicting course outcomes, predicting standard achievement etc. Albany Technical College used the STARS, student racking and retention services which contains features like identifying at risk students and personally connecting them with a counsellor by taking students results, demographics and social data as input. The requirements of the model included, that the system should automatically give feed-back after the system runs the algorithms on the relevant data. Student's reaction plan should also be strictly monitored by the system after found at risk. The data input should contain multiple parameters which could be demographics, academic ability, academic performance, academic history, financial support etc. Warning and alert system should be self-triggered so as to warn the students and alert the faculties.

Then a model was proposed which employed techniques of neural networks and decision trees. The first feed forward network passed the input information from the input layer to the hidden layer. The second network contained a bi-directional flow of data with feedback. This layer basically used the IF-THEN rules to sort the student's data in the form of a decision tree. The third layer was where the warning and alert system were triggered. The rules of the decision tree used in the model were in the form of 'current Student ABC got low marks in first quiz and assignment and he is weak in prerequisite subject, Math, English' message will be sent : Your expected grade is D, please review prerequisite subjects and see the tutor to get more help.'

The LCMS [18] blackboard technological platform was discussed in this paper. It supported the teaching and learning activities of students making it important to understand the

integration of the teaching and learning process. The issues which came up within the organizational, educational and technological dimension were listed. The proposed system was supposed to restrict data, i.e. a faculty should have access to only his students and not the others. The faculty should be given authority to change the course content of the subjects only which were taught by him. Educational issues included integrating the data from various sources of the LMS which included tests and questionnaires evaluation and access material, participation in online forums and blogs, online course content and multimedia con-tent. Technological constraints included ways to trans-form all the data and perform detailed analysis on it and then present the results in a proper format.

The periodical reports which gave blackboard provides were elaborately explained. The importance of the LCMS in the five stages which included entry, adoption, adaptation, immersion and transformation was also explained. The academic analytics domain could ex-tract the data from the various sources which included administrative sources, educational sources, financial sources, research sources, alumni sources and employ-ability sources.

Intelligent tutoring systems (ITS) and adaptive hyper-media systems [19] are adaptive software's from which data is extracted for data mining purposes. A new up-coming concept is personal learning environment (PLE) which allows the user to build customized learning environments, self-adapted to their individual needs. Further the paper elaborates on the different steps in learning analytics which are data collection and pre-processing, analytics and action and post-processing. Of these the post-processing step is of prime importance as it involves compiling new data, determining new attributes for the new iteration or choosing a new analytics method for additional insight in a particular direction.

The first criteria for analysis was what kind of LA sys-tem has been researched on. The different types of systems included open data set, PLE, LMS (learning management system), adaptive systems/ ITS, web based courses, social media and others. Study results showed that most of the research carried out was on Adaptive systems and intelligent tutoring systems. The next part of the study was to find out which techniques were frequently used in the research experiments out of statistics, visualizations, classifications and predictions, clustering, association rule mining, social network analysis and others. Classification and prediction was found to be the most extensively used technique in the field of learning analytics. The next part of the study focused upon what was the objective of the study, monitoring and analysis, prediction and intervention, assessment and feedback, intelligent tutoring and adaptation, personalization and recommendation and others. It was found that most of the research experiments focused on achieving the objective of intelligent tutoring and adaptation with monitoring and analysis finishing a close second.

An experiment [20] was conducted to find out which were the course tracking variables which correlated significantly with academic performance. Majority of Tai-wan universities have adopted learning management systems (LMS) to support student leaning for years. The questions answered were: Which course tracking variables of LMS correlate to academic performance in blended online courses? How accurately can measures of significant course tracking variables of LMS data predict academic courses in blended

courses? The LMS captured and stored large amounts of data from user activities and interaction data. Number and duration of online sessions, messages read or posted and content pages visited were some of the course tracking variables provided by the LMS. The input data was LMS tracking data of one semester of fifteen blended courses offered by one national university in Taiwan from Feb 2012 to July 2012. The five parameters taken into consideration were total number of online sessions, total number of original posts created, total number of follow-up posts created, total number of posts read and total number of content pages viewed. The correlation coefficient was calculated between these five and the academic score of the student by multiple regression analysis using SPSS software tool. It was found out that the number of online sessions variable explained 15% of the variation in the final grade of the student. The other four parameters also influenced the student's grade and explained the variation to about 2% to 8%.

The advantages of e-learning systems [21] in the educational environment which include provision of learning material with no geographical and time constraints, ability to track and monitor the progress are highlighted in this research paper. Social network analysis technique of data mining was used in this experiment. The network graphs which consists of nodes which represent students and edges which represent connections between them were the basis of the study. A personal learning environment (PLE) HOUSELEARN was introduced using the traditional LMS environment tools. This tool was built on the elgg framework which consisted of tools like blogs, social bookmarks, file sharing, polls and status updates. The PLE allowed the formation of groups with private or public access based on common interests. It also allowed users to keep a track of all the activities of another user by 'following' him. Next the students taking the course "Special issues on software engineering" used the HOUSELEARN PLE. A group PLH42 was made specifically for this course and to put up contents of this course on the PLE. All the users in the group could follow all the other group members, start discussions on topics related to the course, could share blog posts and other articles. HOUSELEARN used H2L Database using Navicat Premium and Toad from MySQL 6.0.1. Tools like Gephi and NodeXL were used for social network analysis. Initially in September 2011, only 18 members out of 80 had more than one connection, and the rest were just single nodes in the network graph. After some assignments were given to all the students in December 2011, this was done in order to increase the participation in the PLE group. This increased the number of connections to 73, a 232% increase. In May 2012, the connected nodes were 54 and the total number of connection was 404. The paper also mentioned a list of metrics which were used to analyze the data from the PLE.

4. SUMMARY
The field of Educational analytics and Academic analytics has reached its adolescence and is no longer in its nascent stages. The search for development of more ac-curate models to predict the student's performance, provide relevant and accurate feedback to the faculty, to understand the requirements of the student and then recommend programs available in the institute, to properly arrange and schedule timetable for institutions taking all the constraints into consideration, to develop LMS and e-learning courses to focus on improving the student's learning experience are all research areas which constitute the domain of academic analytics.

A variety of tools some of which are open source are available and widely used to perform analysis of the educational data. These tools include SPSS clementine 12.0, WEKA data mining tool, Rapid Miner, Pentaho Business Intelligence Suite, XL miner, Borland C++ builder6 tool, Hadoop file distribution system (HDFS), Model-It software tool, Moodle serve version

5. CONCLUSION

The educational institutions need to develop models to give in depth information to the students and the faculty related to performance, attendance, program requirements and course related information. After segregating the students in need of attention, treatment procedures designed after carefully assessing the requirements of the students and identifying the weakness of the students should be implemented. The performance after implementing these procedures should also be strictly monitored and changes should be made if performance is not up to the mark.

Data mining techniques like outlier analysis, naïve Bayesian algorithm should be used to research into are-as like finding students with similar interests in order to pursue research projects in that sector and to identify groups of students in need of extra guidance in particular areas to schedule extra coaching sessions. Analysis of students should be carried out on a weekly basis by continuously assessing them through assignments and research projects.

Academic analytics is another field with vast potential and demand in today's world. Administrative staff are always inquisitive to know about whether their investment and resources are being utilized optimally. Faculty qualifications and research portfolios, discipline-wise performance in terms of academics, research and extra-curricular achievements are parameters which can be analyzed. These can be presented in the form of graphs, charts and other explanatory figures to the administrative staff and keep them well informed about the progress of the student and the faculties of the institute.

6. REFERENCES

[1] Barahate Sachin R. and Shelake Vijay M., A Survey and Future Vision of Data mining in Educational Field in Proceedings of the IEEE Symp. Second International Conference on Advanced Computing & Communication Technologies.

[2] Crist´obal Romero and Sebasti´an Ventura, Educational Data Mining: A Review of the State of the Art in Proceedings of IEEE transacations on systems, man and cybernetics.

[3] Kannan Govindarajan, Thamarai Selvi Somasundaram, Vivekanandan S Kumar, Kinshuk, Continuous Clustering in Big Data Learning Analytics in Proceedings IEEE Fifth International Conference on Technology for Education.

[4] Charoula Angeli , Nicos Valanides Using educational data mining methods to assess field-dependent and field-independent learners' complex problem solving in Springer Association for Educational Communications and Technology

[5] Carlos Márquez-Vera, Cristóbal Romero Morales, and Sebastián Ventura Soto, Predicting School Failure and Dropout by Using Data Mining Techniques in

[6] S.Anupama Kumar and Vijayalakshmi M.N, Mining of Student Academic Evaluation Records in Higher Education in Proceedings of International conference on recent advances in computing and Software Systems

[7] Eitel J.M. Lauría, Joshua D. Baron, Mallika Devireddy, Venniraiselvi Sundararaju and Sandeep M. Jayaprakash, Mining academic data to improve college student retention: An open source perspective in Proceedings of ACM Second International Conference on Learning Analytics and Knowledge.

[8] Eitel J.M. Lauría, Joshua D. Baron, Mallika Devireddy, Venniraiselvi Sundararaju and Sandeep M. Jayaprakash, Open Academic Analytics Initiative: Initial Research Findings in Proceedings of ACM Third International Conference on Learning Analytics and Knowledge.

[9] Eitel J.M. Lauría, Joshua D. Baron, Mallika Devireddy, Ven-niraiselvi Sundararaju and Sandeep M. Jayaprakash, Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative in International Journal of Learning Analytics

[10] Reshma Desai, Academic Analytics in Customer Relationship Management Perspective using Data Mining in Proceedings of International Conference in Recent Trends in Information Technology and Computer Science

[11] Mohiuddin Ali Khan, Wajeb Gharibi, Sateesh Kumar Pradhan, Data Mining Techniques for Business Intelligence in Educational System: A Case Mining in Proceedings of IEEE Computer Applications and Information Systems

[12] Mahdi Nasiri, Fereydoon Vafaei, Behrouz Minaei, Predicting GPA and Academic Dismissal in LMS Using Educational Data Mining: A Case Mining in Proceedings of IEEE 6th National and 3rd International conference of e-Learning and e-Teaching

[13] Rainer Knauf, Kinshuk, Kouhei Takada, and Yoshitaka Sakurai, Takashi Kawabe, Setsuo Tsuruta, Personalized and Adaptive Curriculum Optimization Based on a Performance Correlation Analysis in Proceedings of IEEE 2012 Eighth International Conference on Signal Image Technology and Internet Based Systems

[14] Rainer Knauf, Yoshitaka Sakurai, Kouhei Takada, and Setsuo Tsuruta, A Case Study on Using Personalized Data Mining for University Curricula in Proceedings of 2012 IEEE International Conference on Systems, Man, and Cybernetics

[15] Fangming Guo and hua Song, Research and Application of Data-Mining Technique in Timetable Scheduling in Proceedings of Computer Engineering and Technology (ICCET),2nd International Conference

[16] Dr. B. L. Shivakumar and Mr. V. Murugananthan, A Novel Application Framework for Educational Data Mining towards Automated Learning System in

Proceedings of IEEE 2014 International Conference on Intelligent Computing Applications.

[17] Bin Mat, U. Shah Alam, Buniyamin, N. Arsad, P.M., Kassim R, An overview of using academic analytics to predict and improve students' achievement: A proposed proactive intelligent intervention in Proceedings of Engineering Education (ICEED), 2013 IEEE 5th Conference.

[18] Sérgio André Ferreira and António Andrade, Academic Analytics: Mapping the Genome of the University in Proceedings of IEEE Revista Iberoamericana de tecnologias Del aprendizaje.

[19] M.A. Chatti, A.L. Dyckhoff, U. Schroeder, and H. Thüs, A Reference Model for Learning Analytics in International Journal of Technology Enhanced Learning (IJTEL) – Special Issue on State-of-the-Art in TEL.

[20] Che-Cheng Lin and Chiung-Hui Chiu, Correlation between Course Tracking Variables and Academic Performance in Blended Online Courses in Proceedings of 2013 IEEE 13th International Conference on Advanced Learning Technologies.

[21] Eleni Koulocheri Alexandros Soumplis Michalis Xenos, Applying Learning Analytics in an Open Personal Learning Environment in Proceedings of IEEE 2012 16th Panhellenic Conference on Informatics.