Predication of Students Performance Using Data Mining

A. V. Kadu[1], Dr. G. R. Bamnote[2]

[1]M.E Scholar, Department of Computer Sc. & Engineering, PRMIT&R, Badnera

[2]Professior, Department of Computer Sc. & Engineering, PRMIT&R, Badnera

Abstract: *The achievement of the good context of different degrees and education which is in the Undergraduate degrees the important in the context of Higher Education, and both for students and for the institutions that host them. To investigate attempt to use data mining techniques to predict students' outcomes based on early module performance and other student characteristics. In this work on whether the data mining can be used to highlight performance problems early in that and propose remedial actions. Many more are the some of the methods may also form the basis for recommender systems in that may guide students towards their module choices to increase the good outcomes their chances is necessary for good outcome. Here use data collected through the admission process and through the students' degrees. In this predict good context of different degrees and education outcomes based on data at admission also on the first year results of modules. To validate the proposed results modules, we evaluate the system whether the data relating to students with different characteristics from the different schools and the collages. Many universities have specific targets for students achieving good honor degrees.*

Keywords: Classification; Data Mining; Performance Prediction; Recommender System

1. Introduction

The universities have specific targets for students achieving good honor degrees. Achievement in terms of good honours is often reported in league tables. For example, the University has complete guide reports good honours as "The graduate's student percentage of are achieving a first or second class honours degree". For example, the Complete University guide et.al [7] reports good honours as "the percentage of graduates achieving a first or upper second class honours degree". The achievement of good honours in Undergraduate degrees is important in the context of Higher Education (HE), both for students and for the institutions that host them It is also important for students to achieve a good degree as this can impact on their employment prospects. The databases in the most organization hold too much data and information that it become complicated and difficult to analyze those data manually.

It is therefore in the interest of both students and the universities to identify students at risk of not obtaining a good honours degree so that early intervention may improve their outcome. The universities have large numbers of module choices, and it is challenging for students to familiarize themselves with the entire make appropriate choices and possibilities. Appropriate choices may lead to higher student satisfaction and/or better performance. The universities are working in a very dynamic and powerfully viable environment today. They gather large amount of volumes of data with reference to their students in electronic ways.

However, they are data rich but information poor which results in unreliable decision making. The biggest challenge is the effective transformation of large volumes of data into knowledge to improve the quality of managerial decisions. It can be the basis for a recommender system to student's aid in module selection. Recommender systems are currently considered as an advisable automated solution for assisting students with their choices.

Nowadays, the databases in the most organization hold so much data and information that it become complicated and difficult to analyze those data manually. To overcome the human is a limitation on handling data in the manual way, Data Mining (DM) is the suitable techniques to be used for conducting the data analysis process. Its combine machine learning, statistical and visualization techniques to discover and extract knowledge in such a way that humans can easily interpret. DM can be used in many areas such as medical, economic, fraud the main objective of Educational Data Mining (EDM) is to discover the new knowledge and the hidden pattern exists in the students' data.

There is variety of DM methods that available for data experiment such as classification, clustering Baradwaj et.al [5]. This research used the classification techniques to develop a model for predicting the Student Academic Performance (SAP).

As there is no industry definition about the components of LMS but broadly defined as follows (See Fig 1):
- A system that deals with Courses, Examinations, Assessments, and Collaboration features.
- Each presenting specific functionalities via specific tools.
- Stake holder of LMS includes developers, managers, administrator, instructor and learner that are responsible for course development, management, training and teaching learning process etc.

Figure 1: Component of learning management system(LMS)

Figure 1 Components of LMS According to the, following functions are provided by an LMS:

- On demand Content delivery
- Data assessment and future action
- Provide interaction among learners, instructors and course administrator etc.
- Progress tracking and reporting includes courses for usage, relationship in mandated training and extraction of tracked information etc.
- Centralization of learning functions and management system
- Demographic and learning record keeping
- Matching of learner and administrator preferences
- Data exchange with external systems (e.g.: content management systems)
- Reuse of data and
- Ability to find and prefer course curricula by learner and respective supervisors including instructor led training classes. E-LEARNING DEVELOPMENT PHASES.

2. Related Work

There are various previous studies conducted to predict the student good and poor honours performance using Data Mining techniques. The implementation of Data Mining techniques in educational field is giving additional insights in making better decisions and solutions for every issue arise.

Azwa Abdul Aziz.et.al [1] presented the Waikato Environment for Knowledge Analysis. The Waikato Environment for Knowledge Analysis. The open source tool is used for classification model development. Classification is the one of the most commonly applied technique in that predicts group exists in the data set. Classification technique is used by in the educational field to better understand students. The educational field that involves Data Mining techniques is rapidly increasing. Applying Data Mining techniques in an educational environment are known as Educational Data Mining that aims to discover hidden knowledge and patterns about students. The amount of data stored in an educational database at is increasing rapidly by the times.

A number of studies have addressed similar issues within the educational data mining framework. For example, early warning systems were investigated by Arnold et.al [2]. The author suggests warning weak students that they are at risk of failing to achieve satisfactory results; this is reported as a good source of motivation for them. Students are accordingly evaluated through the three commonly known traffic signal lights (green, amber and red). The signal's color indicates the risk level for the student relative to their counterparts. The

algorithm used analyses both the student's module performance and their help seeking behavior.

A number of models have also been used in this context. For example, A. C. K. Hoe et.al [3] employed a CHAID algorithm to identify the important variables that influence the performance of undergraduate students. The study examined the patterns obtained using the data of students demographics and past performances.

Hence, Kondaveeti et.al [4] presented used the GIS techniques with data mining techniques (spatial and association data mining techniques) to model the crime patterns and trends. Reference proposed the use of naive bayes algorithm with the concept of named entity recognition (NER), also known as entity or element extraction, to classify the news articles into the crime type and to create a crime model. In addition to that, A priori algorithm is used to find and create frequent patterns in crime by training crime data from the different web sites. For prediction in crime, they used the decision tree concept. As a tested result, their system can classify and predict the crimes more than 90% accuracy. For a crime predicting model implemented in collaboration with the police department of a United States city in the North east crime, the hotspots are the best method for crime forecasting proposed. To improve the accuracy of clustering technique, the segmented multiple metric similarity measure (SMMSM) is proposed that used to find the crime suspects.

Nowadays, the databases in the most organization hold so much data and information that it become complicated and difficult to analyze those data manually. To overcome the human is a limitation on handling data in the manual way, Data Mining (DM) is the suitable techniques to be used for conducting the data analysis process. Its combine machine learning, statistical and visualization techniques to discover and extract knowledge in such a way that humans can easily interpret. DM can be used in many areas such as medical, economic, fraud the main objective of Educational Data Mining (EDM) is to discover the new knowledge and the hidden pattern exists in the students' data.

3. Proposed Work

The student management system does not provide data in the format ready for an easy and direct statistical analysis and modeling. The same problem was reported for the university. Therefore a data preparation and cleaning as well as creation of variables for analysis were undertaken to prepare database for modeling. Data Preparation and Cleaning the Information Systems course was offered as a distance and online course with a separate course codes. From this course is offered exclusively as an online course. Therefore for the pre-data, the course ID was changed and both distance and online occurrences of the course were coded as one course. From the initial dataset all studens granted cross-credit or credit were excluded because they didn't actually study this course.

The aim to investigate the available features that may be used for prediction, as well as the type of classifiers that may produce the best results. Fig 2 shows the Proposed

Framework of Students Performance Prediction.to the analysis could also be used to influence admission policies given the characteristics of predicted poor performers. The hope to uncover early indicators of poor performance that may be used to target remedial action for the concerned students. That's why develop a predication system by which Students Per-romance Using Data Mining.
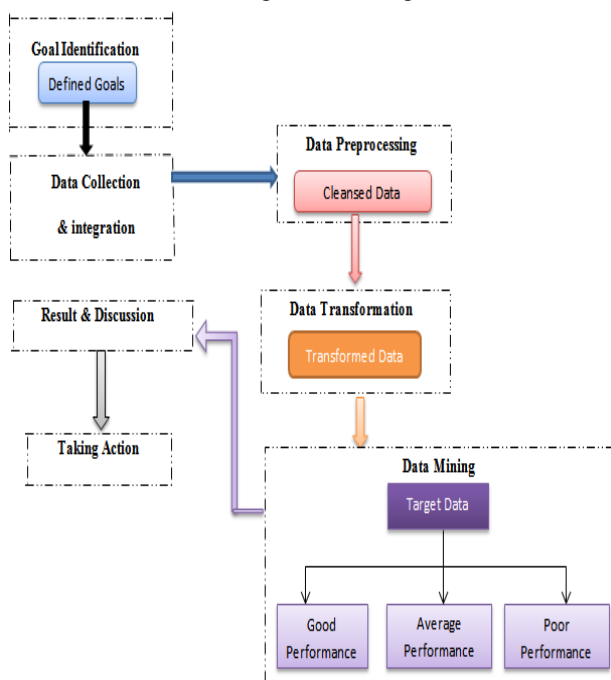

Figure 2: Proposed Framework of Students Performance Prediction

Goal Identification
The goal is to develop the Students Performance prediction model using selected classification techniques and Data mining techniques which are K-mean Clustering algorithms. The accuracy values of the three models are compared and the model with the highest accuracy is selected as a Students Performance predictive.

Data collection and integration
In our system using student performance prediction. The data contain the information about students. All the students' information is combined in a single file using the students' matrix number as a primary key during the integration process. The six parameters were selected In this report; gender, race, hometown location, family income, university entry mode. The data were collected from the Academic Department of the database that is stored in Informix Database Management System.

In present day's educational system, a student's" performance is determined by the internal assessment and end semester examination. The internal assessment is carried out by the teacher based upon students" performance in educational activities such as class test, seminar, assignments, general proficiency, attendance and lab work. The end semester examination is one that is scored by the student in semester examination. Each student has to get minimum marks to pass a semester in internal as well as end semester examination.

These types of data can be collected from the learning management systems or other content management systems and they can be mined using data mining techniques not only to enhance the learning experience but also to facilitate accurate decision making. In order to apply educational data mining to the e-learning courses, the following four steps need to be followed.

They are:
1. Data Collection: Learning Management systems are accessed by hundreds or thousands of employees every day to access various e-learning courses in order to fulfill their learning requirement. These usages and learner data will be stored in a database from where it needs to be extracted.
Students-. xlsx –Table 1 shows the parameters are selected from the database

Sr No		Student Attributes
1.		Students academy
	a.	Subjects and subjects Code
	b.	Subjects Marks
	c.	Students Attendances
	d.	Students Grades
2.		Extra-curriculum Activities
	a.	Paper presentation
	b.	Blood Donation
3.		Intra-curriculum Activities
	a.	Singing Competition
	b.	Dancing Competition
	c.	Drama Competition
	d.	Debate Competition
4.		Other Activities
	a.	Sports

Result-.xlsx –Table 2 shows this excel file contains the students

Sr. No	Results contains
1	Subject and Subject Code
3	Max Unit test Marks
4	Max Marks College Assessment
5	Min Passing Marks

2. Data Preprocessing: The data is cleansed and transformed into an appropriate format to which some suitable data mining tool can be applied. The data set used in this study was obtained from on the sampling method of computer Applications department of course from session. Initially size of the data. In this step data stored in different tables was joined in a single table after joining process errors were removed.

3. Data selection and transformation: - In this step only those fields were selected which were required for data mining. A few derived variables were selected. While some of the information for the variables was extracted from the database.

4. Data Mining: Data mining techniques are applied to the preprocessed data in order to build and execute the model that is required to extract and summarize the data which is of interest for the course ware developer or the learning and development team. In order to do so, here can use a general

or a specific data mining tool which can be chosen by or for the organization under consideration.

$$j = \sum_{j=1}^{k} \sum_{i=1}^{n} \|xi(j) - cj\|_2^2$$

Where,

$\|(xi(j), \mu i)\|_2^2$ Is a chosen distance measure between a

data point xi(j) and the cluster Centre cj, is anindicator of the distance of the n data points from their respective cluster centers.

The algorithm is composed of the following steps:
Step1. Place K points into the space represented by the objects that are beingclustered. These points represent initial group centroids.

Step2. Assign each object to the group that has the closest centroid.

Step3. When all objects have been assigned, recalculate the positions of the K-centroids.

Step4. Repeat Steps 2 and 3 until the centroids no longer move. This produces aseparation of the objects into groups from which the metric to be minimized can be calculated.

5. Evaluation and Interpretation of results : In this step, the results are interpreted and used in order to understand the useful patterns of learning. The SCORM concepts and the steps described till now are applicable to the majority of the leading learning management systems in the market like Saba, Moodle etc. Now here have covered the gist of e-learning development and the ways in which educational data mining can be used in eLearning system.

Table 1: Student Respective Year
- Implementation details

| Value	Description
First Class (1st)	Student Percentage > 75%
Second Class (2nd)	Student Percentage >=60% to 75%
Third Class (3rd)	Student Percentage >=35% to 60%
Fail	Below 35%

- Result and Discussion

Value	Description
Fair	student mark > (module' s average mark + 5%)
Average	(student mark >= module's average mark - 5%) and (student mark <= module's average mark + 5%)
Poor	student mark < (module's average mark - 5%)

4. Result

As the result show the front page of project. What is the Plot chart of student performance and show performance in the bar chart.

4.1.1 Front Page
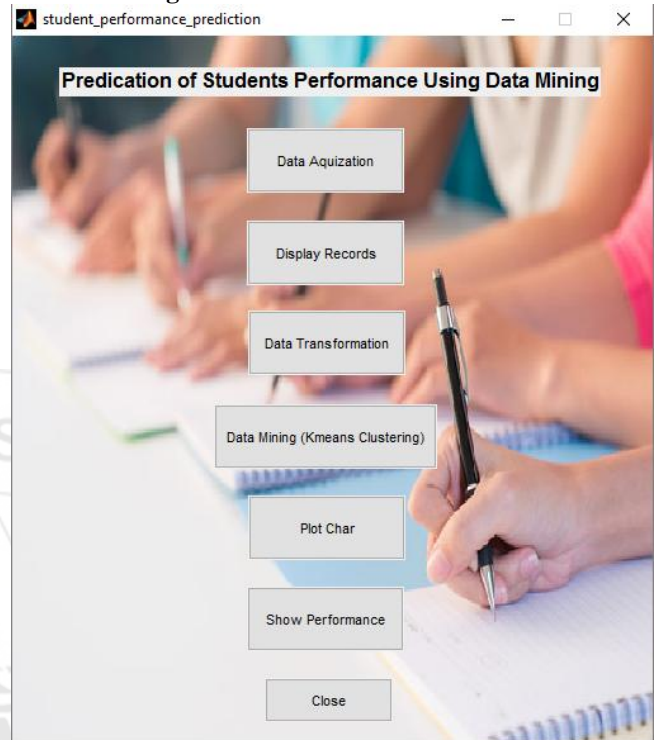
Screen Shots 1:- Front page of predication of student performance

In this page here data have given the 4 year student record in excel. Choose a year as you want for clustering the data. In that if click on final year, third year, second year, first year or all academic years than message will be shown as above **"All Student data is being is clustered".**
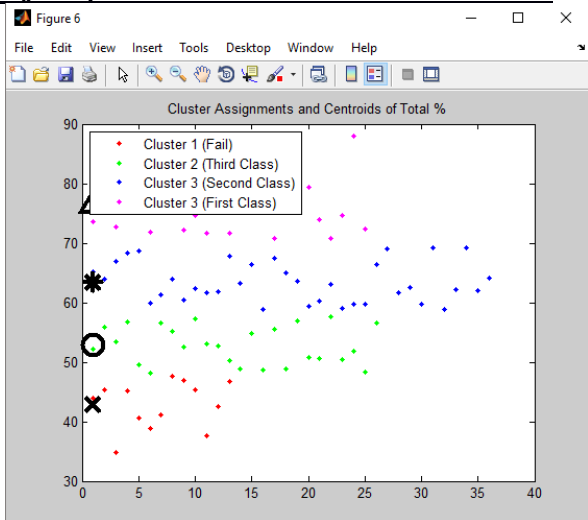
4.2 System Execution details
4.2.1 Plot Chart:-
Student Respective Year of Final Year Student 2016
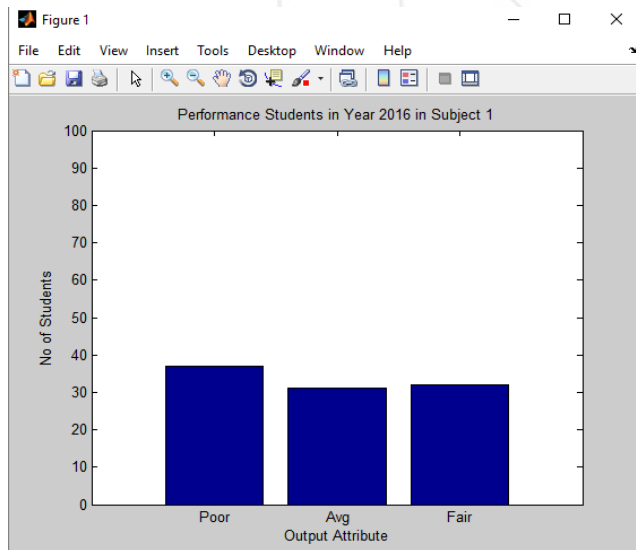
Paper ID: ART20173509
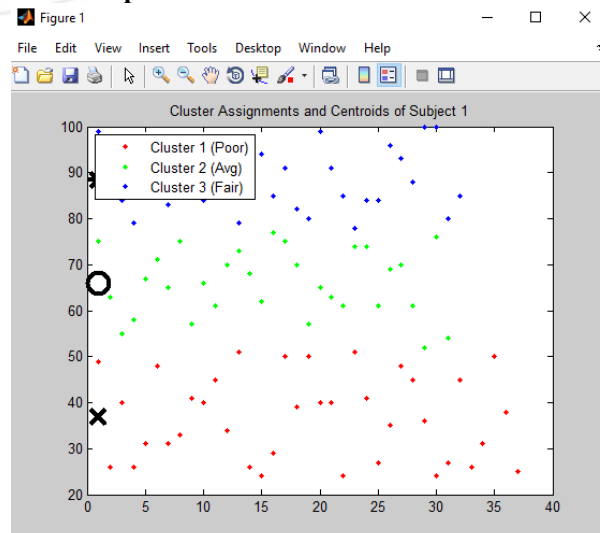
In this plot chart student respective year of final year for 2016.same condition is apply on each and every year i.e 2013 to 2016. Fair is denoted in * symbol,Average is dentoed in 0 symbol and poor is denoted in X symbol.
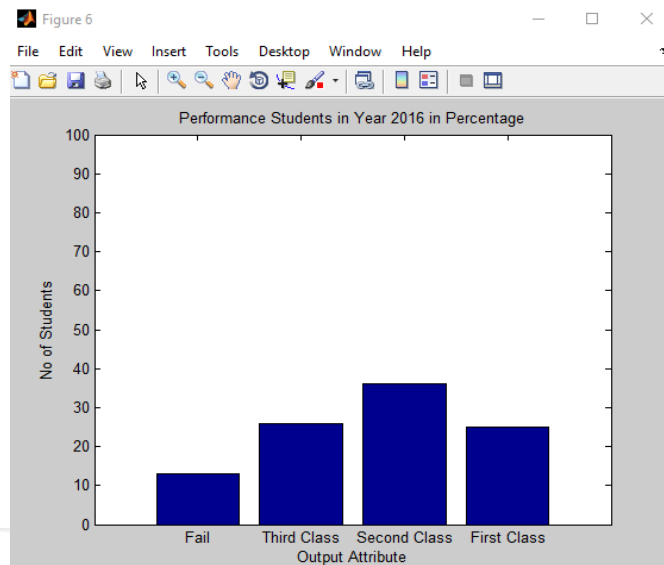
Subject Specified Year of Final Year Student 2016

In this plot chart subject respective year of final year for 2016.same condition is apply on each and every year i.e 2013 to 2016. Fair is denoted in X symbol,third class is denoted in 0 symbol,Second class is dentoed in * symbol and Fair is denoted in symbol.

4.3.1 Show Performance:-
Student Respective Year 2016:-

In this show performance student respective year of final year for 2016.same condition is apply on each and every year i.e 2013 to 2016. Average is denoted in blue bar and writen in below "Avg",Fair is denoted in blue bar and writen in below "Fair".and poor is denoted in blue bar and writen in below "poor".

Subject Specified Year 2016:-

In this show performance student respective year of final year for 2016.same condition is apply on each and every year i.e 2013 to 2016. First class is denoted in blue bar and writen in below "First class", Second class is denoted in blue bar and writen in below "Second class",Fair is denoted in blue bar and writen in below "Second class".and Third class is denoted in blue bar and writen in below "Third class".

Student performance over all four academic year

Year	Poor	Average	Fair
2013-2014	38	73	87
2014-2015	35	74	86
2015-2016	38	75	89

Subject performance over all four academic year

Year	Fail	3rd class	2nd class	1st class
2013-2014	40	50	60	80
2014-2015	39	53	61	87
2015-2016	42	19	70	90

5. Conclusion & Future Scope

The primary goal of this work project is predict students i.e., that are at a high risk of not achieving a good honours degree, but more importantly, to identify this as early as possible in year so that interventions can be proposed. The next stage will be to recommend strategies based on this and measure performance improvements. The stage to recommend strategies based on this and measure performance improvements. It may also be possible to incorporate data on engagement (e.g. attendance monitoring, library loans) which i becoming available in the data warehouse to see its impact on prediction accuracy.

References

[1] Azwa Abdul Aziz, Nor Hafieza Ismail and Fadhilah Ahmad "First Semester Computer Science Students' Academic Performances Analysis by Using Data

Mining Classification Algorithms" *Higher Education Journal,2015-2016.*

[2] A. Pe˜na-Ayala, "Educational data mining: A survey and a data miningbased analysis of recent works," in *Expert systems with applications*, vol. 41(4), no. 4. Elsevier, 2014, pp. 1432–1462.

[3] A. C. K. Hoe, M. S. Ahmad, T. C. Hooi, M. Shanmugam, S. S.Gunasekaran, Z. C. Cob, and A. Ramasamy, "Analyzing students records to identify patterns of students performance," in *Researchand Innovation in Information Systems (ICRIIS), 2013 International Conference* on. IEEE, 2013, pp. 544–547.

[4] A. Kondaveeti, G. Runger, H. Liu, and J. Rowe, "Extracting geographic knowledge from sensor intervention data using spatial association rules," in *Spatial Data Mining and Geographical Knowledge Services (ICSD10), 2011 IEEE International Conference on*, June 2011, pp. 127–130.

[5] Baradwaj, B. K. (2011). *Mining Educational Data to Analyze Students' Performance. (IJACSA)* International Journal of Advanced Computer Science and Applications, 2(6), 63–69.

[6] B. K. Baradwaj and S. Pal, "Mining educational data to analyzestudents' performance,"*arXiv preprint arXiv:*1201.3417, 2012.

[7] "Complete university league tables," February 2015. [Online].Available:http://www.thecompleteuniversityg uide.co. uk/leaguetables/ rankings.

[8] C. Vialardi, J. Chue, J. P. Peche, G. Alvarado, B. Vinatea, J. Estrella,and A´ . Ortigosa, "A data mining approach to guide students through theenrollment process based on academic performance," in *User modelingand user-adapted interaction*, vol. 21(1-2), no. 1-2. Springer, 2011,pp. 217–248.

[9] C. Romero, S. Ventura, E. Vasilyeva, and M. Pechenizkiy, "Classassociation rules mining from students' test data." in *EDM*. ERIC,2010, pp. 317–318.

[10] Dutta Borah, M., Jindal, R., Gupta,D.,Deka,G.C (2011), "*Application of knowledge based decision technique to Predict student's enrolment decision*. in Proc. Int. Conf. on Recent Trends inInformationSystems,India:IEEE,pp.180184,DOI :10.1109/ReTIS.2011.6146864.

[11] Dutta Borah, M., Jindal, R., Gupta,D.,Deka,G.C (2011), "*Application of knowledge based decision technique to Predict student's enrolment decision*. in Proc. Int. Conf. on Recent Trends inInformationSystems,India:IEEE,pp.180184,DOI :10.1109/ReTIS.2011.6146864.

[12] D. Kabakchieva, Predicting Student Performance by Using Data Mining Methods for Classification, *Cybernetics d Information Technologies,* vol. 13, No. 1, pp. 61-72, 2013.

[13] D. Garcıa-Saiz and M. Zorrilla, "Towards the development of a classification service for predicting students' performance." EDM, 2013, p.212.

[14] E. Aguiar, N.V. Chawla, J. Brockman,G.A. Ambrose, V.Goodrich, ―Engagement vs Performance: Using Electronic Portfolios to predict first semester engineering student retention,‖ International Conference on Learning Analytics and Knowledge, ACM, 2014.

[15] F.J. Hair Jr, C.W. Black, J.B. Babin, and E.R. Anderson, *Multivariate data analysis* 7th Ed., Upper Saddle River, NJ: Prentice Hall Inc., 2010, pp. 335-438.

[16] Galit, "*Examining online learning processes based on log files analysis: a case study*". Research, Reflection and Innovations in Integrating ICT in Education 2007.