

Determining the Factors of the Students using Factor Analysis in EDM

Ruchi Jain, Neha Nagpal
Assistant Professor, Assistant Professor

Abstract - *In educational environment, a plethora of data is available. The number of factors are generated during the analysis of the educational database. This paper determine the factors which makes a great impact on the academic performance and the precise model must be developed for enhancing the performance of academics. Tinto's model have been discussed to analyze the interaction between institute and student in order to analyse the drop out decisions. The geometric model has been suggested to study the various factors of the student. Factor analysis is introduced for reducing the numerous factors. Factor analysis is incorporated in the process of development of DSS model.*

Keywords - *Factor Analysis, EDM, Students, Academics.*

INTRODUCTION

Educational data mining is the way which has the ability to discover the knowledge from the educational environment. Educational data mining aims to analyze the prediction system of academic performance. In EDM, during the learning process, the data is gathered from the various resources and then further studies can be made using the techniques such as statistics, machine learning and other data mining techniques.

It is still a hot issue to determine the factors which makes a great impact on the academic performance as well as the development of a precise and deterministic model for the enhancing the performance of the academics.

There can be variety of assessments applied to student performance and can be aggregated to generate a single measure to study the performance of student academically. There can be various factors which can be estimated from enrolment to tertiary in order to develop a deterministic model of academic performance. When the value of the factors weakens down the student is unable to continue in the course to which they are enrolled.

BACKGROUND

Different modelling methods have been introduced in educational data to predict the retention of student. Tinto's model proposed the theoretical framework for evaluating factors in student's academic success[2].

In 1987, Vincent Tinto states that the process of student's attrition is a socio-psychological interaction between the characteristics of student enrolling in the university and the experiences in the institute. The various key factors which are responsible for the student attrition were stated as: lack of ability to associate the new knowledge with the previous one, a feeling of isolation, problem of adaptation. As over 35 years, Tinto's model have been undergoing great amendments according to the present scenario. Over the period of time, Tinto's model was upgraded and was named as student's integration model, which states that "Students who are academically and socially associated to campus community results in increase in the commitment to the institution, have the higher probability to be graduate." Tinto and many other researchers have concluded that student expectation and institutes objectives should coincide with each other.[3]

GEOMETRIC MODEL FOR IDENTIFYING FACTORS

Various models have been introduced regarding the selection of the factors affecting the student persistence to the particular course. Among them, one of the model for the said concept was geometric model. It is one of the simple model which precisely express the relationship between college and student. In this model, the student has been considered as the core element and the three factors make an impact on that element such that Cognitive factors, Social factors and institutional factors which plays a substantial role in the development, persistence and growth of the student. The persistence process in this model illustrates to maintain the intricate balance between the student resources and institution resources. The student resources describes the stuff brought to the campus by the student and the institution resources depicts the facilities provided by the institution by the student. The factors discussed in the geometric model can be explained as: cognitive factors forms the academic ability

of the student such as proficiency in reading, writing and comprehension. Social factors describes the ability to efficiently deal with other persons, cultural background and personal attitudes. The institutional factors refers to the

strategies, practices and culture of the college which makes a great impact directly or indirectly on the students' growth, development and persistence. [4]

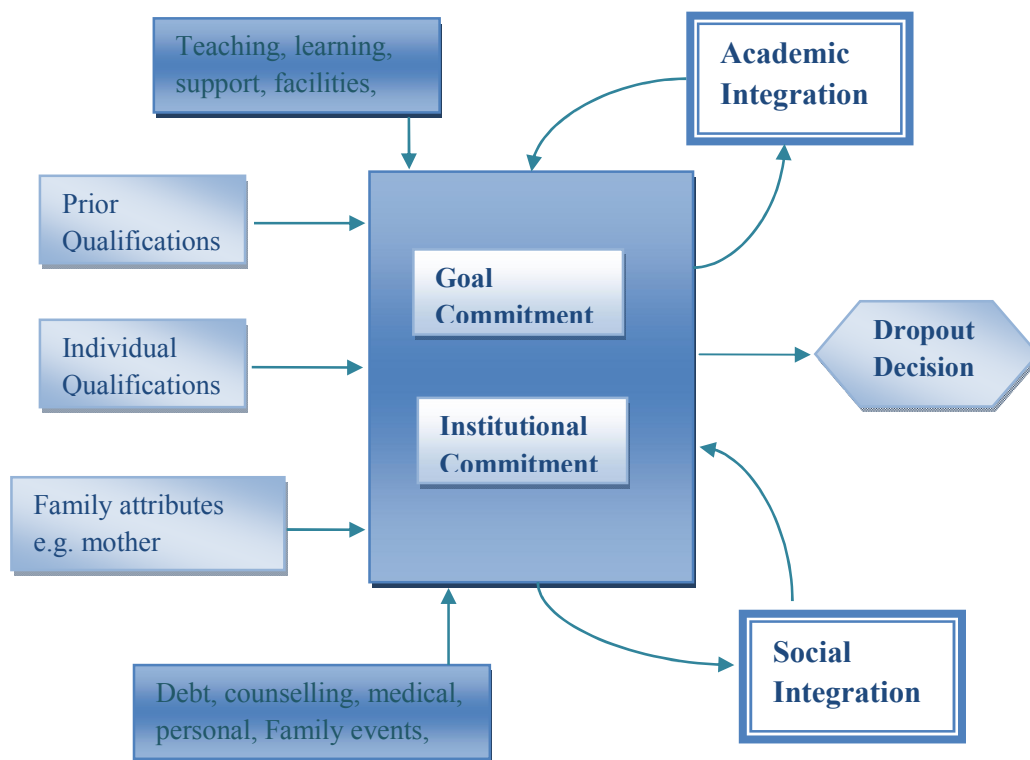


Figure 1 Tinto's Model

FRAMEWORK FOR THE STUDY [7]

Data mining techniques is used in educational field to predict the academic performance of the student. A plethora of data is available and knowledge has to be discovered from the hidden patterns. The data can be academic or personal which is used to enhance the performance in student's behavior and also to improve the learning environment. The effective data mining technique has to be applied for the analysis of factors which makes the way for better academic planning and management. The educational data is refined at each stage and the model has been developed which increases the performance. The following are the four phases of the study:- [1]

METHODOLOGY FOR DATA PRE-PROCESSING

Factor analysis(FA) is the method used to discover the factors of the student attributes as there is excessive data and may have redundant dimensions in the data which can induce the negative impact in discovering the performance of the data mining methods. Factor analysis, generally preferred

technique, to reduce the voluminous data into meaningful and significant data. This method also perceives the correlation structure between the number of considered variables using the definition of a set of common hidden elements termed as factors. The main objectives of FA are: first, to reduce data dimensions into substantial characteristics and second, to categorize variables using the discovered relationship between the variables. Consequently, there are two types of factor analysis: Exploratory FA and Confirmatory FA. In EDM, various factors are discovered which are identified as independent structure factors and then task of determining the legitimate limit for the said factors is to be performed. Thus, the related descriptions for each variable must be deduced. After completion of these steps the *data abstraction* and *data volume reduction* can be applied at initial stage of the factor analysis. Data abstraction implies that the factors which are concealed in the data, then these factors are elucidated so that all the factors are revealed and then reduction of variable can be performed. Data volume reduction can be done by evaluating the scores for hidden factors and then replacing it with the initial variables. [5]

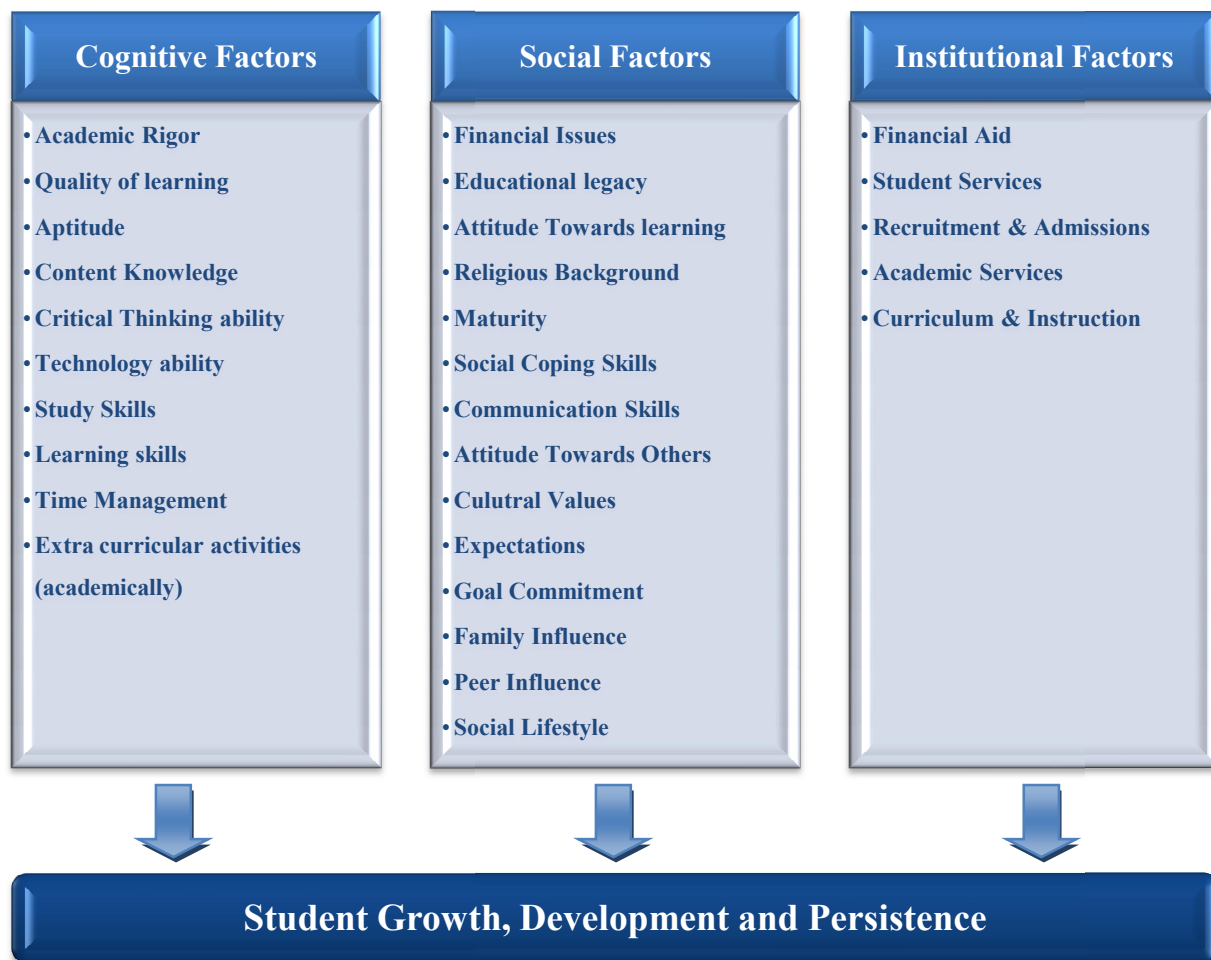


Figure 2The Geometric model of student persistence and achievement



Figure 2 Framework for the study of student development

[6]The initial assumption of FA is to consider the measured variables as a linear combination of the latent factors which is immeasurable explicitly. The purpose of FA is to recognize the immeasurable factors, analogous behavior of the variables and the related description of the variables.

The general model of FA can be expressed mathematically as ,

Let X be column vector containing p measured variables, μ be the mean vector of X , F be the column vector contains q latent factors, L be the $p \times q$ matrix that transforms F to X .

The elements of L represents the weight of each measured variable contributed by each factor. ϵ be the column vector, which contains the p uncorrelated random errors. The condition imposed is $q < p$. So the model of FA can be expressed as by Johnson and Wichern 1998:

$$X - \mu = LF + \epsilon \quad \text{.....eq 5.1}$$

Rummel, 2002, stated that " It untangles the linear relationships into their separate patterns as each pattern will appear as a factor delineating a distinct cluster of interrelated data". The above statement can be interpreted as that FA considers all the interdependent variables and categorize them into a separate group, finally those variables can be reduced to the number

of variables required in the experiment.

Algorithm to analyze the factors of student persistence using Factor Analysis:[5]

Step 1: For a certain group select the primal variables based on the geometric model.

Step 2: The test should be applied for the suitability of data such as Kaiser-meyer-Olkin measure of sampling adequacy or Bartlett's test of Sphericity for FA applications.

Step 3: Determine the number derivation of factors using Eigen value etc. i.e generation of scores.

Step 4: The factor scores are applied to the new limited set of variables by replacing the original values.

CLASSIFICATION ALGORITHM

Classification algorithm basically predicts the membership of the group for the given instance of data. The classification process can be used to the reduced set of independent variables. The algorithm discovers the hidden relationships between the variables, so that the outcome can be predicted. There can be various classification process which can be used such as SVM, genetic algorithm, Bayesian Classification etc.

EVALUATION OF CLASSIFIER:

Accuracy, TP (True Positive) rate, FP (False-Positive) rate, TN (True Negative) rate, FN (False Negative) rate has to be evaluated using different models like 10- fold cross validation and supplied test set.

GENERATED ACADEMIC MODEL:

Classification Algorithm generate the rules which is used to develop the Decision Support System model for student development, growth and persistence.

CONCLUSION:

Tinto's model was found to be very useful to analyze the interaction between the institute and student and can reduce the number of drop out students. Geometric model identified the various factors which makes a great impact on the student development, growth and persistence. Those factors have been categorized as Cognitive, Social and Institutional which plays a vital role for academic growth of the student. Factor analysis has been found to be the most effective way to reduce the number of variables. At the initial stage of the study, the preprocessing of data has to be performed so reduction of the factors falls under this category. Then any one of the classification algorithm can be applied to the pre-processed data. Various classifiers generated can be evaluated which helps in the development of the DSS model for the development of the student.

REFERENCES

- [1] Nitya Upadhyay,Vinodini Katiyar;"A survey on the classification techniques on Educational Data Mining";International Journal of Computer Applications Technology and Research Volume 3- Issue 11, 725 - 728, 2014, ISSN: 2319-8656.
- [2] Sajadin Sembiring, M. Zarlis,Dedy Hartama,Ramlana S,Elvi Wani;"Prediction of Student Academic Performance By An Application of Data Mining Techniques";2011 International Conference on Management and Artificial Intelligence IPEDR vol.6 (2011) © (2011) IACSIT Press, Bali, Indonesia.
- [3] Steve Draper;"Tinto's model of student retention"; <http://www.psy.gla.ac.uk/~steve/loaled/tinto.html>
- [4] Dr. Watson Scott Swail;"The art of student retention";www.educationalpolicy.org

- [5] Ali Asghar Anvary Rostamy, Tahereh Aliheidri Bioki, Farideh Bakhshi Takanlou, Amin Anvary Rostamy; "Utilizing Data Mining and Factor Analysis for Identifying Activity Base Costing Cost Drivers in Iranian Bank"; Universal Journal of Accounting and Finance 1(1): 1-8, 2013 DOI: 10.13189/ujaf.2013.010101

- [6] Zu-Hsu Lee, Richard L. Peterson, Chen-Fu Chien, Ruben Xing;"Factor Analysis in Data Mining"; Copyright © 2006, Idea Group Inc.

- [7] WILAIRAT YATHONGCHAI, CHUSAK YATHONGCHAI, KITTISAK ERDPRASOP, NITTAYA KERDPRASOP;"Factor Analysis with data mining technique in Higher Educational Student Drop out";ISBN: 978-1-61804-093-0