International Conference on Emerging Trends in Engineering, Technology, Science and Management
IIMT College of Engineering, Greater Noida, India (ICETETSM-17)
12th April 2017, www.conferenceworld.in ISBN: 978-93-86171-38-2

ROLE OF PSYCHOLOGICAL FACTORS IN EDUCATIONAL DATA MINING

Iti Burman[1], Subhranil Som[2], Syed Akhter Hossain[3]

[1,2]Amity Institute of Information Technology, Amity University Uttar Pradesh, Noida, UP, India

[3]Daffodil International University, Dhaka, Bangladesh

ABSTRACT

This paper aims to provide a comprehensive insight to Educational Data Mining and its utilization towards understanding an individuals' behaviour. It provides the basic introduction to various psychological factors that help to predict the future prospect choices of an individual. We studied a number of papers, supporting the role of non-intellective constructs in growth of a person. It constitutes the data of the survey carried over 150+ adults to support the study. Further, the paper identifies (by the means of opinion polling) the requirement of a recommender system in this area. Finally, we set our thoughts to summarize and conclude the results revealed by the data.

I. INTRODUCTION

Educational Data Mining (popularly known as EDM) is a growing research area which is interdisciplinary in nature; majorly focusing on generating methods for analysing data coming from educational environment and its impact on students learning capability in order to devise the useful changes in learning settings [1]. EDM , at large, is concerned with studying, researching ,applying and developing automated methods to identify the patterns in enormous volume of data available from educational field which is otherwise, almost, unfeasible to extract and detect due to voluminous nature of the data [2]. Educational data mining (EDM) is a field that works on combination of statistical techniques along with machine-learning concepts and data-mining (DM) algorithms over the varied educational data. The main purpose is to study and analyze data to resolve research concerns related to education [3]. Mining of Educational data is a latest emerging trend which can lay a road to various new horizons in field of education. Therefore, as a result, recent inclination towards mining of student related database to extract valuable information has emerged. This implementation of warehousing and mining techniques on such data can help to reveal hidden valuable information and contribute to improvement in education quality [4]. This technique helps to reveal important and fruitful results about the educational preferences and future career prospects of an individual.

It is evident, from study revealed in [5] every individual's learning behaviour depends not only on the academic growth but also the psychology and various other non-intellective correlates. The proper identification of traits such as intelligence, attitude, personality and beliefs can be a great help to assist a person for more favourable career paths comparatively earlier than their own self-exploration. Psychometric tests comprise personality profiles, motivation questionnaires, reasoning tests and ability assessments. These tests moreover reveal objective data for otherwise subjective measurements. Factors apart from intelligence may be decisive to precise

ISBN: 978-93-86171-38-2

prediction of performance as provided in [6] the relation between intelligence, personality, and interests; also in [7] it is demonstrated that academic performance is allied with personality traits.

Thus results of EDM over Learning Analytics combined with psychometric data study of individual can prove a boon to education system. It may not only help to improve the learning process but also can lead to a better fully exploited education technique which can fetch an overall boost to learner and teaching environment. Also it will cater the major need of understanding learning pattern and career assistance at early stages of life.

The main objective of this paper is to provide with the insights of educational data mining and non-intellective correlates. It also identifies the relevance of various psychological factors that can affect in guiding most favourable options for future aspects of individual. The analysis of non-intellective data along with study of grade and performance pattern can reveal undiscovered horizon of individual learning and interest. At the end we have summarized the results for, the survey carried out to explore the need of a recommender system which can analyse psychological correlation and predict favourable future prospects further.

II. EDUCATIONAL DATA MINING

Data mining, also Knowledge Discovery in Databases (KDD), is the field of discovering original and potentially helpful information from huge data set [8]. The foremost task of data mining is applying various techniques and algorithms to discover and extort patterns of stored data. These may help in predictive analysis as well as decision making .Data mining techniques have gain a hand to new areas including neural networks, patterns recognition, spatial data analysis, image databases and other application fields such as business, economics, and bioinformatics leaving education as no exception.

The educational data mining community defines, “Educational Data Mining (EDM) is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the setting which they learn in”[9]. EDM has gained quite an attention of researchers; The application of data mining in educational system is an interactive cycle of hypothesis formation, testing, and refinement [10].

2.1 Methods of EDM

The following are the methods of EDM. These methods are listed as web mining methods, and are quite prominent in mining web data and in mining other forms of educational data. These categories of educational data mining methods are largely acknowledged to be universal across types of data mining [11].

1. Prediction
2. Clustering
3. Relationship mining

2.1.1 Prediction

Prediction includes developing a model that can deduce a single aspect of the data known as predicted variable; from some grouping of other aspects of data called predictor variables. It has two key uses within educational data mining. Firstly, prediction methods can be used to identify important features of a model for prediction, giving relevant information about core construct. This technique is basically used to analyze the students' performance [12]. In second type of usage, the prediction methods are utilized to foresee what the output value would be in the contexts when it is not feasible to directly attain a label for that construct [13]. Further, prediction can be classified into these categories: Classification, Regression and Density estimation.

A classifier is mapping between space X (which can be discrete or continue) and discrete set of labels Y [13]. Classification analysis helps to predict class labels which describes future situation. It is a supervised technique used for labelling newly encountered, but unlabeled patterns, from a collection of pre-classified already labelled patterns. Some popular classification methods include logistic regression, support vector machines and decision trees. Regression technique can also be considered for prediction. In this, the predicted variable is continuous. Few popular methods of regression within EDM include neural networks, linear regression and support vector machine regression. Density estimation takes the predicted variable in the form of probability density function. This estimation can be based on range of kernel functions, together with Gaussian function.

For each of the mentioned prediction methods, the input may vary to be continuous or categorical. The effectiveness of a particular method depends upon the type of input variable used. Most of the real-world EDM problems are never simple prediction. Hence, more complex techniques may need to be applied to estimate future values using amalgamation of various techniques viz. logistic regression, neural networks or decision trees.

2.1.2 Clustering

Clustering is process of grouping or combining objects into classes of similar objects [14]. This is an unsupervised categorization or partitioning of patterns, like observations, data items or future vectors, into classes, clusters or subset based on their vicinity and connectivity inside N-dimensional space. Clustering is predominantly constructive in the cases where the common categories in the data set haven't been discovered in advance. In educational data mining, clustering technique has been utilized to group students as per their behaviour [15]. The educational institute could also be clustered to find similarities & differences between their environment and teaching patterns. Some of the well known methods of clustering are k-mean and expectation maximization algorithm (EM-clustering) [16].

2.1.3 Relationship Mining

Relationship mining has the goal to determine relationships (most strong association amongst with a single variable of particular interest) variables, in data sets with a huge number of variables. Broadly relationship mining is classified as: association rule, correlation, sequential pattern, and casual data mining.

Association rule mining discovers associations among attributes in data set, generating if-then constructs regarding attribute-values [17]. It is an important technique that focuses on extractions, correlations of interest, repeating patterns, relation or casual structures identified in set of items of the transaction databases. This technique has been applied to EDM for: discovering students' mistakes which often occur while solving exercises [18]; finding out interactions in learners' behavioural patterns [19]; diagnosing learning problem of student and offer rectifying advice [20] etc. Correlation mining deals with the goal to find linear correlations between variables which may positive or negative as well. It is used to discover the most strongly correlated attributes. Sequential pattern mining yet another technique which is a more restrictive type of association rule mining. In this accessed item's order is also taken in consideration. Sequential pattern mining tries to find inter-session patterns like the presence of a particular set of items with other set items in a time-ordered set of episodes [21]. Sequential pattern can disclose which content has stimulated the access to other content, or how tools and contents are entangled in the learning process. In [22], Zaine and Luo propose the discovery of useful patterns based on restrictions, to help educators evaluates students' activities in web courses. In [23] Paul and Donnelan discovered and compared with expected behavioural patterns specified by the mentor that portray an ideal learning path. Casual data mining technique, attempts to identify whether observed construct was the reason of invocation of another event, either by covariance analysis of two events or from the information of how one of the events was actually triggered. A casual association can be anecdotal, if a pedagogical incident is arbitrarily chosen using automated testing, and frequently leads to a optimistic learning result.

2.2 Psychometric factors involved in studying the learning behaviour of individual

Intelligence tests[24][25] taken by age-old methods depict only cognitive capacities of learner, which includes the ability of individual to represent and manipulate abstract relations[26]. These measures assess individual capacity. Other psychological correlates can be taken into consideration to clarify how individual subjects are likely to apply their intellectual capacities[27][28]. Identifying such non-intellective antecedents or correlates of academic performance has been increased manifold over the past some years [29].Studies have assessed and identified the role of personality in academic and overall performance [30]. Dispositional traits of personality(like intelligence) are considered to exert a continuing influence on performance across challenges. Such traits are genetically mediated and tend to remain relatively stable over time [31]. Research in field of learning behaviour identification has also highlighted the relevance of domain-specific, motivational contributions to ones' academic performance [32].It demonstrates that performance-centric beliefs, values, and goals are "dynamic and contextually bound and that learning strategies can be learned and brought under the control of the student" [33].Consequently, models predicting academic performance may have to include expectancies, motivation, goals, as well as use of self-regulatory learning strategies [29][34]. Unlike generally considered intelligence and personality, these indicators are more malleable and context sensitive [35][36].

We have carried an online survey to identify the importance of such intellective and non-intellective factors. The survey was performed to analyse the need of recommender system for students to help them take better decisions regarding studies and career based on their psychological test. They were asked to rate the various factors on a scale of 1-5; based on its contribution to their learning behaviour and future prospects (includes studies and career options). The survey was carried over a group of more than 150 individuals and the results are shown in Table 1.

Factors are as follows:

2.2.1 Background Knowledge (include knowledge gained at home, neighbourhood, and school environment).

In this parameter, we are getting information regarding his knowledge attain from his/ her school, friends, and school environment.

2.2.2 Intelligence (includes your IQ level, things learned from books etc)

In this parameter, we are getting information his/ her intelligence level i.e. how quickly he/ she learn the things come to his/ her contact.

2.2.3 Personality Traits (like Openness, social mixing, etc)

In this parameter, we are getting information regarding his/ her presentation of issues/ problems/ views before friends/ relatives and family persons .

2.2.4 Motivation Factors (include focus to achieve goal, inspiration from others)

In this parameter, we are getting information regarding his/ her targets, routes of targets, inspirational personalities.

2.2.5 Self-Regulatory Learning Strategies (critical thinking, peer learning etc)

In this parameter, we are getting his/ her views regarding the burning topics of societies/ politics/ education and other problems and how he/ she defends his/ her views also.

2.2.6 Psycho-social Contextual Influences (stress, depression etc)

In this parameter, we are getting information regarding how he/ she face stress situation, stress timing and depression situation and timing also.

Factors	Scale-1	Scale-2	Scale-3	Scale-4	Scale-5	Total Response
Background	1	3	21	52	81	158
Intelligence	2	5	31	58	62	158
Personality Traits	0	4	29	57	68	158
Motivation Factors	1	5	22	55	75	158
Self-Regulatory Learning Strategies	3	7	30	57	61	158
Psycho-social Contextual Influences	15	34	41	47	21	158

Table 1: Rating of Various Factors on Scale 1-5

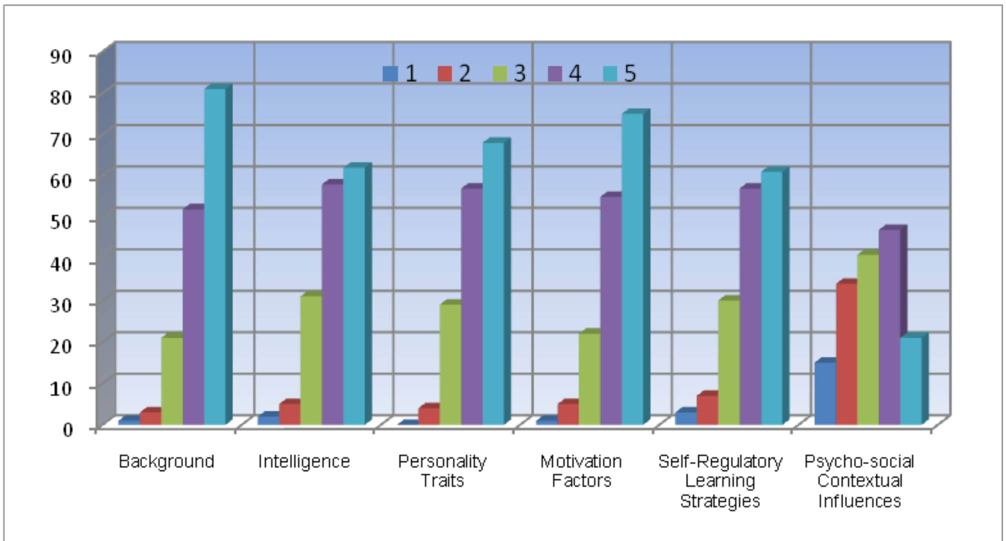
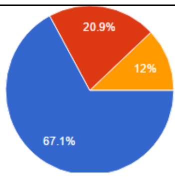
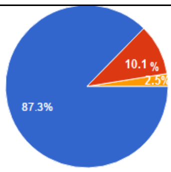
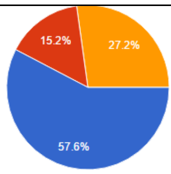


Fig 1: Graph comparing various factors based on Table 1

It can be seen in the results that apart from intelligence other non-intellective factors play equal importance to the individual's growth with highest ranking to background and motivational factors. Personality traits stood next in the line while intelligence and self-regulatory learning strategies played almost equal contribution. This survey depicts that future prospects of an individual depends on the various factors except the intelligence. Hence, if other subjective measures can be depicted objectively in some measurable form; they may prove to be basis for development of a recommender system which can feed all these intellective and non-intellective data and provide results in the form of future prospects assistance.

III. NEED FOR RECOMMENDER SYSTEM

Since in the above sections, we found the contribution of various factors in Educational Data Mining. We have conducted survey to analyse need of a recommender system for favourable career assistance based on psychological study and learning behaviour of individual. This survey depicts the requirement of a recommender system that can help individual in decision making regarding their choices for future. The result from analysis of Table 2 shows that if such a recommender system is developed it will prove to be a great popularity and help to young generation of society who are otherwise confused over multiple choices available to pursue their career. Hence, EDM combined with psychological correlates and academic indicators will help the subject to have a better understanding and guidance for the choice to be made at an earlier stage of life. Such assistance at least can assure that individual has understood all the options based on a measured set of values rather than just interpreting their own behaviour, which may be affected by various emotional factors than actually justified factors. Since, human emotions are vulnerable to many influences; this system may help them to evaluate choices logically with the help of well deduced parameters. The questions and responses are recorded in the form of graph shown in Table 2.

Question	Response
Do you think the learning success depends more on the psychological behaviour than intelligence?	
Do you think the career of an individual is more affected by its attitude towards things than grades in exams?	
Do you think a recommender system based on psychological behaviour of an individual will be beneficial to depict the learning behaviour?	

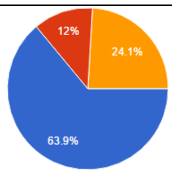
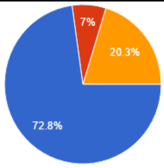
Do you think the recommender system which can analyse your learning behaviour will be helpful to take better decision regarding your future study prospects?	
Do you think the recommender system which can provide you the most favourable career options for you can be helpful?	
YES NO Don't Know	

Table 2: Survey to analyse need of Recommender System

IV. CONCLUSION AND FUTURE WORK

The literature review of the previous research studies and results of survey conducted helped us to understand the importance of non-intellective factors along with intellective constructs towards the growth and enhancement of the individual. We found that EDM in combination psychological factors as input can reveal useful assistance information about an individual. At the end, it is revealed by the survey(in Table 2) that such a recommender system can be developed and prove to be a great help to the young generation in better decision making about the options available to them for future prospects.

REFERENCES

[1] BakerRSJd, Yacef K. The state of educational data mining in 2009: A review and future visions. J EduData Min 2009.

[2] Romero C, Ventura S, Pechenizky M, Baker R. Handbook of Educational Data Mining. Data Mining and Knowledge Discovery Series. Boca Raton,FL: Chapman and Hall/CRCPress; 2010.

[3] T. Barnes, M. Desmarais, C. Romero, and S. Ventura, presented at the 2nd Int. Conf. Educ. Data Mining, Cordoba, Spain, 2009.

[4] Göker, H. , Bülbül, H. I., & Irmak, E. (2013, December). The Estimation of Students' Academic Success by Data Mining Methods. In Machine Learning and Applications (ICMLA), 2013 12th International Conference on (Vol. 2, pp. 535-539). IEEE.

[5] Richardson, Michelle, Charles Abraham, and Rod Bond. "Psychological correlates of university students' academic performance: a systematic review and meta-analysis." Psychological bulletin 138.2 (2012): 353.

[6] Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. Psychological Bulletin, 121, 219–245. doi:10.1037/0033-2909.121.2.219.

[7] Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. psychological Bulletin, 135, 322–338. doi:10.1037/a0014996.

[8] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth, “From Data Mining to Knowledge Discovery in Databases,” American Association for Artificial Intelligence, 1997, (pp. 3754).

[9] www.educationaldatamining.org

[10] Romero, C., & Ventura, S., “Educational data mining: A survey from 1995 to 2005,” Expert Systems with Applications, 33(1), 2007, (pp. 135–146).

[11] Romero, C., Ventura, S., Espejo, P.G. and Hervas, C., “Data Mining Algorithms to Classify Students,” In Proceedings of the 1st International Conference on Educational Data Mining, 2008, (pp. 8-17).

[12] Brijesh Kumar Baradwaj, Saurabh Pal, “Mining Educational Data to Analyze Students’ Performance,” In International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011, (pp. 63-69).

[13] Duda, R. O., Hart, P. E., & Stork, D. G., “Pattern classification,” Wiley Interscience. 2000.

[14] Jain, A. K., Murty, M. N., & Flynn, P. J., “Data clustering: A review,” ACM Computing Surveys, 31(3), 1999, (pp. 264–323).

[15] Amershi, S., Conati, C., “ Automatic Recognition of Learner Groups in Exploratory Learning Environments,” Proceedings of ITS 2006, 8th International Conference on Intelligent Tutoring Systems, 2006.

[16] Dempster A., Larid N., Rubin D., “Maximum Likelihood estimation from incomplete data via EM Algorithm,” Journal of Royal Statistics Society, 39(1), 1997, (pp. 1-38).

[17] Agarwal, R., Imielinski, T., & Swami, A., “ Mining association rules between sets of items in large databases,” In Proceedings of the ACM SIGMOD international conference on management of data,Washington DC, USA, 1993, (pp. 1–22).

[18] Merceron, A., & Yacef, K., “Mining student data captured from a web-based tutoring tool: Initial exploration and results,” Journal of Interactive Learning Research, 15(4), 2004, (pp. 319–346).

[19] Yu, P., Own, C., & Lin, L, “On learning behavior analysis of web based interactive environment,” In Proceedings of the implementing curricular change in engineering education, Oslo, Norway, 2001, (pp. 1–10).

[20] Hwang, G. J., Hsiao, C. L., & Tseng, C. R., “A computer-assisted approach to diagnosing student learning problems in science courses,” Journal of Information Science and Engineering, 19,2003, (pp. 229–248).

[21] Hwang, G. J., Hsiao, C. L., & Tseng, C. R., “A computer-assisted approach to diagnosing student learning problems in science courses,” Journal of Information Science and Engineering, 19,2003, (pp. 229–248).

[22] Zai¨ane, O., & Luo, J., “Web usage mining for a better web-based learning environment,” In Proceedings of conference on advanced technology for education, Banff, Alberta, 2001, (pp. 60–64).

[23] Pahl, C., & Donnellan, C., “Data mining technology for the evaluation of web-based teaching and learning systems,” In Proceedings of the Congress E-learning. Montreal, Canada, 2003, (pp. 1–7).

[24] Harris, D. (1940). Factors affecting college grades: A review of the literature, 1930 –1937. Psychological Bulletin, 37, 125–166. doi: 10.1037/h0055365

[25] Neisser, U., Boodo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J.Urbina, S. (1996). Intelligence: Knowns and unknowns. American Psychologist, 51, 77–101. doi:10.1037/0003-066X.51.2.77

[26] Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. Psychological Review, 97, 404–431. doi:10.1037/ 0033-295X.97.3.404

[27] Barrick, M. R., Mount, M. K., & Strauss, J. P. (1993). Conscientiousness and performance of sales representatives: Test of the mediating effects of goal setting. Journal of Applied Psychology, 78, 715–722. doi: 10.1037/0021-9010.78.5.715

[28] Busato, V. V., Prins, F. J., Elshout, J. J., & Hamaker, C. (1998). The relation between learning styles, the Big Five personality traits and achievement motivation in higher education. Personality and Individual Differences, 26, 129–140. doi:10.1016/S0191-8869(98)00112-3

[29] Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. Annual Review of Psychology, 53, 109 –132. doi:10.1146/annurev.psych.53.100901.135153

[30] Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. Psychological Bulletin, 135, 322–338. doi:10.1037/a0014996

[31] Murphy, P. K., & Alexander, P. (2000). A motivated exploration of motivation terminology. Contemporary Educational Psychology, 25,3–53. doi:10.1006/ceps.1999.1019

[32] Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. Educational Psychology Review, 16, 385–407. doi:10.1007/s10648-004-0006-x

[33] Duncan, T. G., & McKeachie, W. J. (2005). The making of the Motivated Strategies for Learning Questionnaire. Educational Psychologist, 40, 117–128. doi:10.1207/s15326985ep4002_6

[34] Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. Psychological Bulletin, 130, 261–288. doi: 10.1037/0033-2909.130.2.26

[35] Carver, C. S., & Scheier, M. F. (1981). A control-systems approach to behavioral self-regulation. In L. Wheeler (Ed.), Review of personality and social psychology (Vol. 2, pp. 107–140). Beverly Hills, CA: Sage.

[36] Wolters, C., Pintrich, P., & Karabenick, S. (March, 2003). Assessing academic self-regulated learning. Paper presented at Indicators of Positive Development Conference sponsored by Child Trends, Washington, DC.