

A Survey on Factors Affecting Education Society using Data Mining

Het Trivedi¹

Department of Computer Science,
Hemchandracharya North Gujarat University,
Patan, Gujarat, – India

Dr. Ajay Patel²

Department of Computer & Science,
Ganpat University,
Kherva, Gujarat, – India

Abstract: Forecasting the presentation of a learner is a huge worry to the superior instruction mainframes. The range of this article is to recognize the elements affecting the presentation of schoolchilids in ultimate inspections and discover out a appropriate data mining algorithm to forecast the rank of schoolchilids so as to a provide conveniently and an suitable forewarning to schoolchilids those who are at danger. In the current exploration, a review cum investigational methodology was accepted to produce a record and it was organized from a main and a inferior foundation. The accessed outcome since proposition trying admits that type of school is not influence apprentice concert and parents' profession plays a chief task in forecasting categorys. This effort will facilitate the edifying organizations to recognize the learners who are at danger and to and offer well again extra guidance for the weedy learners.

Key Words: Educational Data Mining, Decision Tree, Multilayer Perception, Student performance.

I. INTRODUCTION

Students' educational appearance elbows on various elements similar to individual, socio-monetry, mental and extra ecological durables[4]. forecast layouts that contain all these valuables are necessitated for the helpful forecast of the presentation of the learners. The forecast of learner presentation with high precision is favorable to recognize the learners with short educational achievements originally[6]. The recognized learners can be independently assisted by the educators so that their presentation is healthier in upcoming[2]. Student's performance such as grades and test score had been identified as one of the measures to determine the effectiveness of programs or activities conduct by educational institutes[10]. The student performance are not the result of simple cause and effect relationship but of interactions among factors such as gender, age, learning ability, learning support, motivation of learning and achievement[11]. Education is a key factor for achieving a long-term economic progress[1]. Data mining methods have been functional in several submission domains such as banking, scam finding, and telecommunications[5]. freshly the data mining methodologies were used to improve and estimate the superior schooling errands[3]. a few researchers have planned some methods and architectures for using data mining for superior schooling[12].

II. REVIEW OF LITERATURE

A figure of reviews pertaining to not just the various factors like private, socio-economic, emotional and extra ecological variables that manipulate the presentation of learners but also the models that have been used for the presentation forecast are accessible in the writing and a little precise studies are scheduled under for orientation.

M.Ramaswami and R.Bhaskaran have used CHAID forecast replica to examine the interrelation connecting variables that are used to forecast the result of the presentation at senior inferior school instruction. The features like standard of lessons, symbols obtained in inferior schooling, mark of educate, existing region and kind of inferior schooling were the strongest indicators for the learner presentation in upper inferior schooling. The CHAID forecast form of learner presentation was constructed with seven division forecaster changeable[8].

Nguyen Thai-Nghe, Andre Busche, and Lars Schmidt-Thieme have functional mechanism scholarship techniques to get better the forecast outcome of educational performances for two the actual case studies. Three methods have been used to covenant with the division inequality trouble and all of them prove acceptable outcome. They primary re unbiased the datasets and then used both price-insensible and receptive culture with SVM for the little datasets and with conclusion Tree for the superior datasets. The models are originally deployed on the restricted net[7].

Arockiam et al. used FP Tree and K-means clustering method for discovery the likeness among town and rustic students encoding skills. FP Tree taking out is functional to colander the patterns from the dataset. K-means clustering is used to decide the training skills of the students. The lessons visibly indicates that the country and the town learner change in their encoding skills. The vast scope of town students are superior in encoding ability compared to country students. It divulges that academicians give additional teaching to town students in the encoding theme[9].

III. METHODOLOGY

During broad hunt of the text and conversation with experts on scholar presentation, a figure of factors that are measured to have authority on the presentation of a learner were documented. These influencing factors were categorized as effort variables. The harvest variables on the further hand characterize some probable grades. The main data were collected from the senior inferior train students and the minor facts were composed from the schools and from internet.

For this study, we have take the sample from the north Gujarat district. A sample of 900 students will be taken from a group of schools. Students were grouped in a classroom where they were briefed clearly about the questionnaire and it took on average half an hour to fill the questionnaire. Selection of students was at random.

The primary data was collected using a questionnaire which includes questions related to several personal, socio- economic, psychological and school related variables that were expected to affect student performance. The questionnaire was reviewed by the professionals and tested on a small set of 45 students in order to get a feedback. The final version contained 50 questions and it was answered by more than 900 students. Latter a sample of 500 were selected from the whole.

The secondary data such as mark details were collected from the schools. All the predictor and response variables which were derived from the questionnaire are given in Table 1 for reference. The domain values for some of the nativity variables were defined for the present investigation as follows:

PS: Parental status. It determines the parental status of the student. The possible values are mother only, father only and both.

MTTS: Mode of transportation to school. It determines student. The possible values are by walk, bicycle, two wheeler, and town bus.

COMM: Community- Even though India has defined itself as a secular state, religion and caste are deeply entrenched in the identity of Indians across ages. These factors play a direct or indirect role in the daily lives including the education of young people. In terms of social status, the population is grouped into five categories: Scheduled Castes (SC), Scheduled Tribes (ST), Most Backward Classes (MBC), Backward Classes (BC) and Others (OC). Possible values are OC, BC, MBC, SC and ST.

PTUI-SEC

Private tuition at secondary level. Most of the parents send their wards for private tutoring after school hours. The number of subjects taught at secondary level is five. Therefore the number private tutoring subjects can vary from zero to five.

X-GRA: Marks obtained at secondary level. Students who are in state board stream appear for five subjects each carry 100 marks, transform the marks in percentage into grades by mapping O – 90% to 100%, A – 80% - 89%, B – 70% - 79%, C – 60% -

69%, D – 50% - 59%, E – 40% - 49%, and F

- < 40% }.

GRP-HSC:

Six types of group of study, based on core subjects, is offered at higher secondary level comprising first group (Maths, Physics, Chemistry, Biology), second group (Maths, Chemistry, Physics, Computer Science) third group (Physics, Chemistry, Botany, Zoology), fourth group (History, Economics, Commerce, Accountancy) fifth group (Computer Science, Economics, Commerce, Accountancy) and sixth group (Commerce, Accountancy, Practical- type writing, Office management).

TOS: Type of school. This determines the type of school that the student studied in higher secondary level. It includes the possible values, co-education, boys, and girls.

PTUI-HSEC:

Private tuition at higher secondary level. The number of subjects taught at higher secondary level is six. Therefore, the number private tutoring subjects can vary from zero to six.

HSCG:

Marks/Grade obtained at higher secondary level and it is declared as response variable. It is also split into seven class values: O – 90% to 100%, A – 80%- 89%, B – 70% - 79%, C – 60% - 69%, D – 50% -59%, E – 40% - 49%, F - < 40%.

TABLE 1 STUDENT RELATED VARIABLES

| VARIABLE NAME | DESCRIPTION | DOMAIN |
|----------------------|----------------------------------|--|
| SEX | SEX | {M, F} |
| CO | COMM | {OC, BC, MBC, SC, ST} |
| PS | PARENTAL STATUS | {BOTH, ONLY MOTHER , ONLY FATHER } |
| FOHT | FOOD HABIT | {VEG , NON-VEG } |
| LA | LIVING AREA | {CORPORATION, MUNICIPAL, RURAL } |
| FA-SIZE | FAMILY SIZE | {SMALL, MEDIUM, LARGE } |
| MOTTS | MODE OF TRANSPORTATION TO SCHOOL | {BY WALK, BICYCLE, TWO WHEELER, TOWN BUS, SCHOOL BUS, AUTO } |
| P-E | HAD PRIMARY EDUCATION | {YES, NO } |
| SA-EL | SCHOOL AREA AT ELEMENTARY LEVEL | {CORPORATION, MUNICIPAL, RURAL } |
| I-E | INSTITUTION AT ELEMENTARY LEVEL | {PRIVATE, GOVERNMENT } |
| SA-S | SCHOOL AREA AT SECONDARY LEVEL | {CORPORATION, MUNICIPAL, RURAL } |
| I-S | INSTITUTION AT ELEMENTARY | {PRIVATE, GOVERNMENT } |

| | | |
|--------------------------------|--|--|
| | LEVEL | |
| S-S | SECONDARY SYLLABUS | { MATRIC, CBSE, STATE BOARD } |
| MI | MEDIUM OF INSTRUCTION AT SECONDARY LEVEL | { ENGLISH, TAMIL } |
| TS | TYPE OF SCHOOL | { CO-ED, BOYS, GIRLS } |
| PT-S | PRIVATE TUITION AT SECONDARY LEVEL | { YES, NO } |
| X-G | GRADE OBTAINED AT SECONDARY LEVEL | { O – 90% - 100%, A – 80% - 89%, B – 70% - 79%, C – 60% - 69%, D – 50% - 59%, E – 40% - 49%, F - < 40% } |
| G-HS | GROUP OF STUDY | { FIRST, SECOND, THIRD, FOURTH, FIFTH, SIXTH } |
| HA | SCHOOL AREA AT HIGHER SECONDARY LEVEL | { CORPORATION, MUNICIPAL, RURAL } |
| MO | STUDENT'S HAVING MOBILE | { YES, NO } |
| IS | INTEREST IN SPORTS | { YES, NO } |
| C-H | COMPUTER AT HOME | { YES,NO } |
| N-AC | INTERNET ACCESS | { YES, NO } |
| CA-H | CARE AT HOME | { MOTHER, FATHER, SISTER, BROTHER, OTHER } |
| P-E | PARENT'S EDUCATION | { BOTH EDUCATED, MOTHER EDUCATED, FATHER EDUCATED, BOTH UNEDUCATED } |
| F-O | FATHER OCCUPATION | { COOLEY, FARMER, WEAVER, PRIVATE, GOVERNMENT, BUSINESS, NOT APPLICABLE } |
| M-O | MOTHER OCCUPATION | { HOUSE WIFE, COOLEY, FARMER, WEAVER, PRIVATE, GOVERNMENT, NOT APPLICABLE } |
| P-S | PARENTS SALARY | { {0 .. 0.9K, 1K .. 2.9K, 3K 4.9K, 5K ..9K, 10K..20K, ABOVE 20K, £NOT-APPLICABLE} } |
| HSCG(RESPONSE VARIABLE) | MARK OBTAINED | { O – 90% - 100%, A – 80% - 89%, B – 70% - 79%, C – 60% - 69%, D – 50% - 59%, E – 40% - 49%, F - < 40% } |

IV. CONCLUSION AND FUTURE WORK

Data mining methodes permit a tall rank withdrawal of information since rare facts, contribution motivating potential for the teaching area. In this learning a replica was urbanized based on a few chosen enter variables composed through Questionnaire technique. After annoying some suggestion, a few of nearly all influencing factors were recognized and in use to predict Data mining methods are applied to forecast the presentation of the students and establish that Multi Layer awareness algorithm is best matched to forecast the grades. After this study we will get mainly exaggerated factors which will influence the educations of the learners.

References

1. S. B. Kotsiantis, C. J. Pierrakeas, and P. E. Pintelas, "Preventing student dropout in distance learning using machine learning techniques", In proceedings of 7th International Conference on Knowledge-Based Intelligent Information and Engineering Systems (KES 2003), pp. 267- 274, 2003. ISBN 3-540-40803-7.
2. B. Minaei-Bidgoli, G. Kortemeyer, and W. F. Punch, "Enhancing Online Learning Performance: An Application of Data Mining Method", In proceedings of The 7th IASTED International Conference on Computers and Advanced Technology in Education (CATE 2004), Kauai, Hawaii, USA, pp. 173-8, August 2004.
3. Kalles D., Pierrakeas C., Analyzing student performance in distance learning with genetic algorithms and decision trees, Hellenic Open University, Patras, Greece,2004.
4. Larose, D. T., "Discovering Knowledge in Data: An Introduction to Data Mining", ISBN 0-471-66657-2, John Wiley & Sons, Inc, 2005.
5. Z. N. Khan, "Scholastic Achievement of Higher Secondary Students in Science Stream", Journal of Social Sciences, Vol. 1, No. 2, 2005, pp. 84-87.
6. P. Cortez, and A. Silva, "Using Data Mining To Predict Secondary School Student Performance", In EUROSIS, A. Brito and J. Teixeira (Eds.), 2008, pp.5-12.
7. Nguyen Thai-Nghe, Andre Busche, and Lars Schmidt- Thieme, "Improving Academic Performance Prediction by Dealing with Class Imbalance", 2009 Ninth International Conference on Intelligent Systems Design and Applications.
8. M.Ramaswami and R.Bhaskaran, "A CHAID Based Performance Prediction Model in Educational Data Mining", International Journal of Computer Science Issues Vol. 7, Issue 1, No. 1, January 2010.
9. L.Arockiam, S.Charles, I.Carol, P.Bastin Thiyagaraj, S. Yosuva, V. Arulkumar, "Deriving Association between Urban and Rural Students Programming Skills", International Journal on Computer Science and Engineering Vol. 02, No. 03, 2010, 687-690
10. Woodman, R. (2014). Investigation of factors that influence student retention and success rate on Open University courses in the East Anglia region. M.Sc. Dissertation, Sheffield Hallam University, UK.
11. Vandamme, J.-P., Meskens, N., & Superby, J.-F. (2015). Predicting academic performance by data mining methods. *Education Economics*, 15(4), 405-419.
12. Chapman, P., Clinton, J., Kerber, R., Khabaza, T.,Reinartz, T., Shearer, C. and Wirth, R.. "CRISP-DM 1 : Step-by-step data mining guide, NCR Systems Engineering Copenhagen (USA and Denmark), DaimlerChrysler AG (Germany), SPSS Inc. (USA) and OHRA Verzekeringen Bank Group B.V (The Netherlands), 2015".