CLASSIFICATION TECHNIQUES BASED DATA MINING IN STUDENT PERFORMANCE PREDICTION

S.Bakiyalakshmi[1], T.K.Anusuya[2]

[1]Department of Computer Science, Bon Secours College for Women, Thanjavur.
[2]Head of the Department & Assistant Professor, Department of Computer Science, Bon Secours College for Women, Thanjavur.

ABSTRACT
In displayed models best classification algorithms in instructive data mining towards exhibiting predominant foreseeing models have been given. Dissect and approval done on models demonstrates the got outcomes are exact and dependable. In such manner individual, natural and instructive variables influencing successful and unsuccessful students have been examined and as per them proficient models dependent on choice tree strategies like c4.5 tree calculation, bolster vector machine techniques and calculated relapse have been displayed. The outcomes can help directors of instructive frameworks towards a right instructive arranging an advancing of instructive procedures in secondary schools. Assessment and expectation of students' execution in secondary school help to discover imperative variables influencing students' achievement in training and also they can have a vital job in helping instructive administrators in enhancing the nature of schools. As indicated by this reality that data mining science has dependably been a reasonable procedure to remove information from data, this can be utilized for giving a decent methodology. This article attempts to introduce prevalent models in foreseeing students' execution. The referenced data of this article are taken from 386 students of secondary schools in Bushehr territory.

KEYWORDS instructive data mining,c4.5 tree calculation, bolster vector machine techniques,datamining science

INTRODUCTION

Data mining or the strategsy for learning extraction from data bases prompts find new and valuable examples from substantial sizes of data in data bases Educational data mining is accustomed to tending to various issues identifying with instruction. As indicated by the substantial size of put away natural data identifying with understudies in instruction office and everyday expanding in this data, getting concealed information from them is a need in enhancing instructive quality and in a right arranging. Particularly in instruction field, these techniques are utilized to examine key inquiries and find experimental confirmations for enhancing and creating of instructive hypotheses which thus brought about new area of research called instructive data mining. Consequently, by utilizing data mining procedures and dependent on individual, ecological and instructive variables influencing on understudies' execution we endeavored to make a valuable model for them. a model in which, people pick a right arranging dependent on their circumstances to enhance the nature of their instruction and more than that directors of instructive framework by taking a gander at the understudy's circumstances can have a superior control and present instructive counsel to them and toward the end by a right instructive arranging assist them with improving their instructive quality. In such manner by gathering and investigating the natural and individual highlights of successful and unsuccessful people in second and third year of secondary school in various fields we separate the components and properties influencing understudies' execution which has been ignored by instructive chiefs, in this way we utilized data mining classification systems to display appropriate models in foreseeing successful or unsuccessful execution of secondary schools understudies. as a rule, the reason for this article is to introduce effective models dependent on anticipating strategies and by

accepting parameters like individual and family includes, understudy action, individual highlights and By utilizing data mining classification algorithms.

Since the goal of this paper is to predict students' execution in secondary schools dependent on individual, natural and instructive factors and as indicated by this reality that put away data has a deficiency in tending to individual and ecological elements, we chose to plan a poll with help of social and mental specialists and schools counselors so as to extricate affecting highlights. Predicting models on each arrangement of information are made by the related variables. The gathering of information utilized in this investigation are gotten from 386 successful and unsuccessful students of secondary schools which has been gathered to predict students execution .data about students are gathered independently in 5 field, mathematic and physic, regular science, humanities field, specialized and word related field and Work And Knowledge field, which all of these records or classifications has 69 predicting variables comprised of individual and family includes, exercises ,wellbeing and other individual highlights of understudy.

METHODOLOGY

This strategy has got some selective highlights which one of them is its execution free of some random space. This implies this procedure can be utilized in any industry and application. Crisp methodology and process is more finished than SAMA methodology and as per the measurements and numbers it has better application, and more speed and plausibility in progress for execute in complicated business. This procedure incorporates 6 phases which these stages covers every single required advance from first program to conclusive assessment. So as to apply data mining strategies to foresee understudy execution in secondary schools by best data mining models the crisp calculation was utilized. Institutionalization procedure of data mining in postindustrial use which are extremely intrigued among specialists.

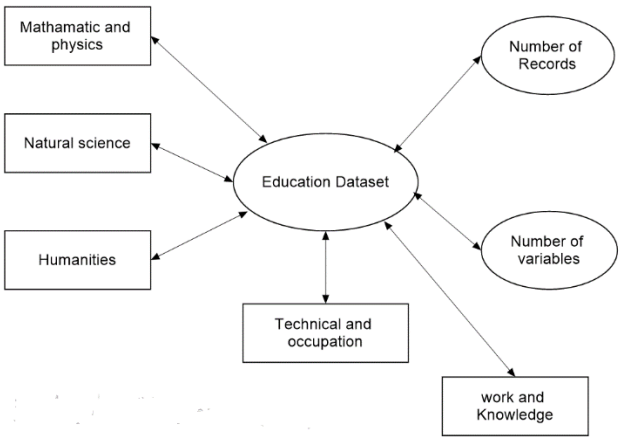


Figure 1. Block Diagram

Data preparation

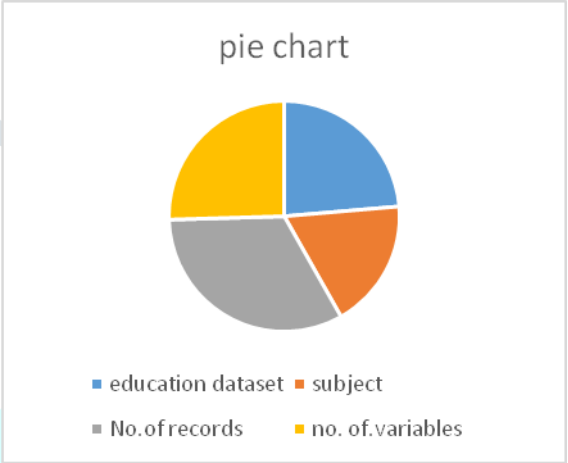
Every gathered variable are discrete or ended up discrete, in order to, applying grouping procedures should be possible in an increasingly exact and proficient. Since the quantity of factors is huge, it was chosen that some element choice strategies would be utilized to lessen factors. Since we had informational index for 5 instructive field, determination of influencing highlights for each set was finished. at this stage 3 gathering of highlight choice has been utilized. First gathering highlight choice techniques are done dependent on weight, these gatherings incorporates GINI index, Information Gain, Chi Squared Statistic, Information gain ratio, Maximum Relevance and One R. Second technique is improved the situation highlight choice among first gathering strategies by casting a ballot and third strategy has been connected by upgrading strategy like hereditary calculation.

At first, objective is to imagine information and make their vital highlights and the relationship of factors clear. Box plot is utilized to demonstrate the connection between two factors, bashfulness as an understudy highlights and GPA. Above we can see that bashful individuals with five rate point (extremely timid) have a lower GPA contrasting with different understudies. Understudies who are generally bashful have most noteworthy GPA. Before bringing in information to the displaying stage, information must be amended. since information gathering was finished by poll, number of lost amounts for factors was low, thus the rectification of information hasn't been done on this stage. in essential

determination arrange, factors which no amount hasn't been entered for them in over 70% records or entered amount was invariant for them in over 70% of cases has been wiped out. The main target variable is understudy's normal incorporates a couple of missing qualities. It's been chosen to pick an esteem that give the most recurrence in the variable, and afterward swap new an incentive for these missing estimations of variable. Understudies' prosperity or disappointment dependent on GPA variable was dictated by supposition of instructive counselors of schools. At that point target variable (GPA) for understudies with GPA higher than 17 was an effective circumstance and for understudies lower than 17 was a disappointment circumstance.

RESULTS AND EVALUATION

In this examination, n is a number between 8 to 40. Subsequent to modeling, we ought to assess the performance of exhibited display in expectation of understudy s achievement or disappointment. The most important criterion in efficiency determination of an algorithm is precision index rating. This index indicates what percent of all informational indexes was evaluated precisely. Distinction in instructive information and decent variety in understudies asks for in each instructive framework make instructive scientists utilize diverse methods for finding shrouded designs in information bases. In this investigation, modeling is finished by using 10-fold cross validation strategy which is one of the most astute methods among order assessment and along by various and best algorithms of arrangement. algorithms utilized at this stage are logistic regression methods, naïve Bayes, Bayes nets, nearest neighborhood, decision tree (CART, C4.5, J48, CHAID and bayes), support vector machines, optimized support vector machines, artificial neural networks, bagging and boosting aggregate methods.in this respect all referenced algorithms was connected to each every last one of 5 informational collection.



At that point, just n variable with most weight has been decided for modeling. According to the table 8 and by assessment and correlation of same models we presume that all models with 90% precision have in excess of 1000 examples. Thus, because of the constrained gathered records for each datum gathering, obtained exactness of exhibited models are wonderful. As a rule, assessment of 5 model of each datum set in this investigation demonstrates that weight-task methods to factors for choosing highlight determination is a superior strategy than highlight choice dependent on optimizing like hereditary algorithm. Additionally among weight-task methods, chi-squared strategy nearly in each datum set can pick best factors. As a general appraisal of predominant grouping models we can say that methods dependent on decision tree like c4.5 tree, support vector machine and logistic regression were best methods in order. Likewise plainly in all informational collection, by choosing no less than 8 to 23 affecting factors among all factors most exact in modeling can be accomplished. An exact estimation of future proceeding of understudies for each instructive framework is a need. Obtained results can enhance nature of training and learning. Counselors

and instructors can take an important commitment in these outcomes and urge understudies to take an interest in various exercises to enhance their performance. What's more, managers additionally can utilize the consequences of this order to organize instructive strategies and enhance asset distribution. on the off chance that we gather more information from schools information base, an increasingly exact outcomes can be accomplished .in light of the fact that a low information volume can influence the came about exactness.

CONCLUSION

The outcomes can help chiefs of instructive frameworks towards a right instructive arranging an upgrading of instructive procedures in secondary schools. In displayed models best order algorithms in instructive information mining towards introducing prevalent anticipating models have been given. Break down and approval done on models demonstrates the acquired outcomes are exact and dependable. In such manner individual, ecological and instructive variables influencing successful and unsuccessful students have been broke down and as indicated by them effective models dependent on choice tree strategies like c4.5 tree algorithm, bolster vector machine techniques and calculated relapse have been exhibited.

REFERENCES

[1] M. H. Dunham, "Data mining introductory and advanced," ISBN 0-13-088892-3, Prentice Hall, 2003.

[2] T. Mitchell, "Machine learning," ISBN 0070428077, McGraw Hill, 1997.

[3] S. Russell and P. Norvig, "Artificial intelligence, a modern approach," Third Edition, Pearson, ISBN-13: 978-0-13-604259-4, ISBN-10: 0-13- 604259-7, 2010.

[4] Y. S. Abu-Mostafa, M. Magdon-Lsmail, and H. Lin, "Learning from data," ISBN 10:1-60049-006-9, AMLbook, 2012.

[5] S. L. Hagedorn, "How to define retention: a new look at an old problem," In Alan Seidman (Ed), College student retention: Formula for student Success, Westport, CT: Praeger Publishers, 2005.

[6] R. Alkhasawneh and R. Hobson, "Modeling student retention in science and engineering disciplines using neural networks," IEEE Global Engineering Education Conference (EDUCON), 660-663, 2011.

[7] D. B. Stone, "African-American males in computer science – examining the pipeline for clogs," The School of Engineering and Applied Science of the George Washington University, Thesis, 2008.

[8] E. Frank, "Pruning Decision Trees and Lists," Department of Computer Science, University of Waikato, Thesis, 2000.

[9] C. H. Yu, S. Digangi, A. Jannasch-pennell, and C. Kaprolet, "A data mining approach for identifying predictors of student retention from sophomore to junior year," Journal of Data Science 8, 307-325, 2010.

[10] S. K. Yadav, B. Bharadwaj, and S. Pal, "Mining educational data to predict student's retention: a comparative study," International Journal of Computer Science and Information Security, 10(2), 113-117, 2012.

[11] A. Nandeshwara, T. Menziesb, and A. Nelson, "Learning patterns of university student retention," Expert Systems with Applications, 38(12), 14984–14996, 2011.

[12] S. A. Kumar and M. V. N., "Implication of classification techniques in predicting student's recital," International Journal of Data Mining & Knowledge Management Process (IJDKP), 1(5), 2011.

[13] D. Kabakchieva, "Student performance prediction by using data mining classification algorithms," International Journal of Computer Science and Management Research, 1(4), 686-690, 2012.

[14] S. Lin, "Data mining for student retention management," The Consortium for Computing Sciences in Colleges, 2012.

[15] S. Singh and V. Kumar, "Classification of student's data using data mining techniques for training & placement department in technical education," International Journal of Computer Science and Network.