DATA MINING WITH NEURAL NETWORKS TO PREDICT STUDENTS ACADEMIC ACHIEVEMENTS

[1]Archana.P, [2]Anbarasan.R,

[1]PG Student, Department of Computer Science Engineering, Bharathidasan Engineering College, Vellore,

[2]Assistant Professor, Department of Computer Science Engineeringng, Bharathidasan Engineering College, Vellore.

Abstract:

In this paper we proposes to apply data mining techniques to How to Improve Listening skill. We use real data on 100 students from Tamilnadu college, such as induction rules and decision trees. In experiments attempt to improve their accuracy for predicting which students Personality Test using all the available attributes. Next selecting the best attributes; and finally, rebalancing data and using cost sensitive classification. The outcomes have been compared and the models with the best results are shown. We focus on the educational data mining and classification techniques. In this study we analyze attributes for the prediction of student's behaviour and academic performance by using WEKA open source data mining tool and various classification methods like decision trees, C4.5 algorithm, ID3 algorithm etc. The classification algorithms naïve bayes,J48, Random forest,AD3, Random tree are grouped with analysis was experimented using Weka tool. The resultant analysis are compared learning behavior of the student data based on the dataset PTQ and Classification methods.

1. INTRODUCTION

Recent years have shown a growing interest and concern in many countries about problem of college students How to class listen or not , and the determination of its main contributing factors . The great deal of research has been done on identifying the factors that affect the low performance of students (Listening Skill) at different educational levels (Anna Univ, deemed univ and Govenment) using the large amount of information that current computers can store in databases. All these data are a "gold mine" of valuable information about students. Identify and find useful information hidden in large databases is a difficult task. A very promising , solution to achieve this goal is the use of knowledge discovery in databases techniques or data mining in education, called educational data mining, EDM. This new area of research focuses on the development of methods to better understand students and the settings in which they learn. In fact, there are good examples of how to apply EDM techniques to create models that Listening skill and Performance test specifically. These works have shown promising results with respect to those sociological, economic, or educational characteristics that may be more relevant in the prediction of low academic performance. It is also important to notice that most of the research on the application of EDM to resolve the problems. more specifically to online or distance education. However, very little information about specific research on part time and full time education has been found, and what has been found uses only statistical methods, not DM techniques .

There are several important differences and/or advantages between applying data mining with respect to only using statistical models : Data mining is a broad process that consists of several stages and includes many techniques, among them the statistics. This knowledge discovery process

comprises the steps of pre-processing, the application of DM techniques and the evaluation and interpretation of the results.

[1] Statistical techniques (data analysis) are often used as a quality criterion of the verisimilitude of the data given the model. DM uses a more direct approach, such asto use the percentage of well classified data.

[2] In statistics, the search is usually done by modeling based on a hill climbing algorithm in combination with a verisimilitude ratio test-based hypothesis. DM is often used a meta-heuristics search.

[3] DM is aimed at working with very large amounts of data (millions and billions). The statistics does not usually work well in large databases with high dimensionality.

2.PROPOSED SYSTEM:

We propose that once students were found at risk, they would be assigned to a tutor in order to provide them with both academic support and guidance for motivating and trying to prevent student failure.We have shown that classification algorithms cab be used successfully in order to predict a student's academic performance and, in particular, to model the difference between Fail and Pass students.

ADVANTAGES:

➢ Data mining is a broad process that consists of several stages and includes many techniques, among them the information.
➢ This knowledge discovery process comprises the steps of pre-processing, the application of DM techniques and the evaluation and reading of the results.
➢ DM is aimed at working with very large amounts of data (millions and billions).
➢ The statistics does not usually work well in large databases with high dimensionality.

3. SYSTEM DESIGN

❖ The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of the input data to the system, various processing carried out on these data, and the output data is generated by the system
❖ The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.
❖ DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.
❖ DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.

4. RESULT ANALYSIS

ANN"s are non-linear predictive models that learn through training and resemble biological neural networks in structure. Neural network is a set of connected input/output units and each connection has a weight present with it. During the learning phase, network learns by adjusting weights so as it will be able to predict the correct class labels of the input data. Neural net works have the remarkable ability to derive meaning from complicated or imprecise data and can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques.. Neural networks are best at identifying patterns or trends in data and well suited for prediction or forecasting needs. Some of the algorithms that are used in the above techniques are listed below:
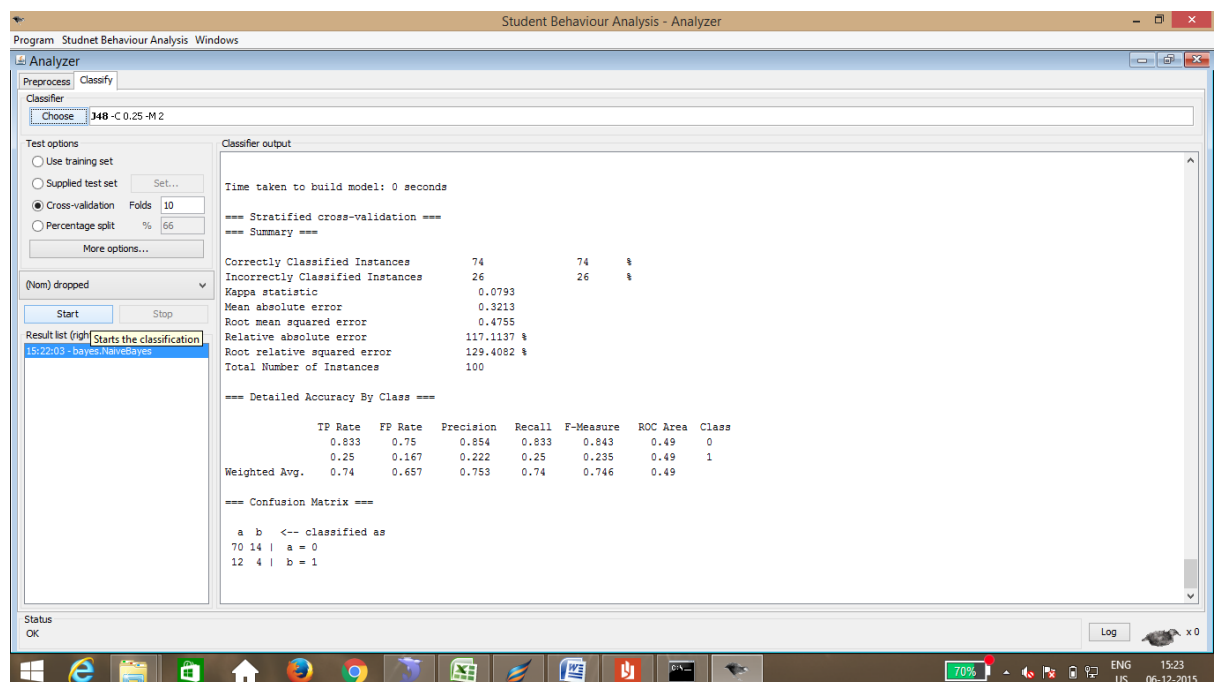

Fig.1.Result analysis

Genetic algorithms: Optimization techniques that use process such as genetic combination, mutation, and natural selection in a design based on the concepts of natural evolution.

Decision trees: Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. The main purpose of the decision tree is to expose the structural information contained in the data. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID). CART and CHAID are decision tree techniques used for classification of a dataset. They provide a set of rules that can be applied to a new (unclassified) dataset to predict which records will have a given outcome. CART segments a dataset by creating 2-way splits while CHAID segments using chi square tests to create multi-way splits. CART typically requires less data preparation than CHAID. ID3 (Induction Decision Tree), a recursive procedure that is used to construct a decision tree from data .A series of

improvements to ID3 culminated to the decision tree approach called C4.5 which deals with numeric attributes, missing values and noisy data.

CONCLUSION

As the growth of education is going beyond the expectations, it is must to enrich the career of the students by providing them a valuable education which would best meet their educational and career motto. This paper hopes to provide a survey of the techniques available in Data mining, how the data are mined, how it can be used in the higher education, identifying the best pattern for the defined problem and what type of methodology should be used to resolve it.An alternative means by which the academic performance of undergraduate students can be predicted; and comparison of this study with the use of machine learning technique based on several metrics shows that, the proposed model generalizes well and performs much better. We hope to extend this study to encompass the integration of mechanisms for the construction of a more robust prediction model that can interact well with the students' portal for real-time exploration in order to strengthen the institution's information system.

REFERENCES

[1] R. S. Naik B, "Using Neural Network to Predict MBA Student Success" College Student Journal, 2004.

[2] O. M. N. Alfan E, "Undergraduate Students' Performance: a case of University Malaya," Emerald Quality Assurance in Education vol 13 no 4, vol. 13 2005.

[3] R. D. Ibrahim Z, "Predicting Students' Academic Performance: Comparing Artificial Neural Network, Decision Tree and Linear Regression, 2007," in 21st Annual SAS Malaysia Forum, 5th September 2007, Shangri-La Hotel, Kuala Lumpur, 2007.

[4] I. Z. Rusli N M, Janor R M "Predicting Students' Academic Achievement: Comparison between Logistic Regression, Artificial Neural Network, and Neuro-Fuzzy," in IEEE, 2008.

[5] N. M. M. Kadirgama K, Sani M S M, M.M.Rahman, M M, Rejab M R M, M.Y.Taib M Y, Sulaiman A, Ibrahim A , Bakar R A "Development of Genetic Algorithms and (MPNN) model to Study The Student

[6] Alan Cheah Kah Hoe, Mohd Sharifuddin Ahmad, Tan Chin Hooi, Mohana Shanmugam, Saraswathy Shamini Gunasekaran, Zaihisma Che Cob, Ammuthavali Ramasamy College of Information and Technology, Universiti Tenaga Nasional, Jalan IKRAM-UNITEN 43009 Kajang, Selangor, Malaysia. Grant Scheme no.: J510050264. Analyzing Students records to Identify Patterns of Student's performance.