Prediction of Students' Educational Status Using CART Algorithm, Neural Network, and Increase in Prediction Precision Using Combinational Model

Mohammad Tari, Tehran Payame Noor University

Behrouz Minaei , Iran University of Science and Technology

Ahmad Farahi, Payame Noor University

Mohammad Niknam Pirzadeh Tehran Payame Noor University

Summary
In this paper, using CART and Neural Network data mining algorithms we have dealt with the study of effective factors in the rate of becoming unqualified among students of Payame Noor University of Qom Province. After creating models using algorithms separately, we have combined models in Clementine 12.0 software through Ensemble Node and have created the combinational model and finally with Analysis Node we have evaluated models and compared results. In the created model using neural network algorithm, in input layer 44 neurons, in hidden layer 3 neurons, and in output layer 1 neuron are created, also in CART model a decision tree with depth of 5 is created. Considering that the number of fields in data bank is high, using Feature Selection Node and selecting the target, we have deleted those fields that had less influence on the target. This matter has decreased the model complexity to some extent.
Key words:
mining, neural network, CART, Payame Noor University of Qom Province, Ensemble

1. Introduction

that education always encounters too much data and information about universities, students, faculty members, personnel, pecuniary resources, etc and most of time, this data can contain valuable information and patterns, so it seems that one of the most important applications of data mining is in higher education. Nowadays, there are wide data banks of students' characteristics that include family, educational, and … characteristics. Finding patterns and knowledge concealed in this information can considerably help decision-makers of higher education arena.
Finding concealed patterns and knowledge from educational systems' data can considerably help decision-makers of higher education arena in advancement and improvement of educational processes like planning, enrollment, evaluation, and consultation.

Future prediction in various fields has ever been interesting and attractive for human being. Surely, it can be said that future prediction and transformations trend in all areas is one of the basic and constant distresses of high and intermediate-level managers. However, there have always been numerous difficulties against it that have made precise and reliable predictions almost impossible. Existence of too many and in many cases hidden parameters have changed such instances into highly complicated problems that mathematical gigantic algorithms have become incapable of presenting an appropriate technique to build an efficient prediction model.
In two recent decades, simultaneous with emergence of artificial intelligence and its combination with well-established science of statistics along with advanced and innovative algorithms like genetic algorithm, metaheuristic methods, artificial neural networks, and … a wide revolution has been made in this arena.
Data mining as a new and flourishing area in the field of presenting predictor models have applied and combined different kinds of statistical techniques, artificial intelligent and innovative algorithms so that it can discover reliable algorithms and models from among of gigantic data storages to predict intended parameters.

2. CRISP Methodology

Data mining1 techniques are among scientific modern techniques that are used in description, explanation, prediction, and control of phenomena. These techniques measure, explain, and predict the correlation degree among variables. Data mining methods influence not only on analytic aspects of studies but also on designing and tools of data collection for decision-making and solving problems. The most successful data mining projects are implemented in the framework of a standard process, which is presented by a work team in SPPS Company in the form of projects named CRISP-DM. According to CRISP-DM, a certain data-mining project has a six-stage life cycle that shows succession of stages. Each stage of

Manuscript received June 5, 2011
Manuscript revised June 20, 2011

stages' succession often includes the result of correlation of previous stages. The most important correlation among stages is showing arrows. Recurrent feature of CRISP indicates external cycle that often leads to a solution for research or business problem with additional remarkable questions. In figure 1, you can observe stages of this methodology [1].

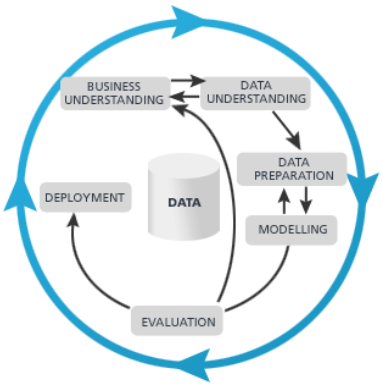


Figure (1): Stages of CRISP-DM Methodology

3. Table Applied

Table (1): Basic information

Student Number	First name and surname	Kind of admission in university	Admission Term	Birth Date	native

Some unnecessary fields have been deleted.

Table (2): Lessons

Lesson's code	Lesson's name	Term	Score	Theoretical credit	Practical credit	Explanations

4. Data preparation

Pre processing2: The importance of data preparation is because of the reality that "lack of data of quality is equal to lack of quality in results of mining" and improper input leads to improper output. In table 2, the importance of data preparation is compared to the importance of other steps of knowledge discovery using data mining. [2]

Data mining step	Percent of time spent from total work	Percent of importance in work's final success
Data preparation	75	75
Data investigation	20	15
Data modeling	5	10

Table (2): comparison of data preparation importance

4.1. Tables Combination

In this stage that is a part of data preparation, we combine tables and the result is a new table with the specification of table 3 that is used as data mine for algorithms.
Through nodes that are data preparation special, we implement some changes on this data mine.

Table (3): List of created combinational table fields

Field name	Field type	Field name	Field type
Gender	Male/female	unqualified	numerical
Native	Yes/no	successively unqualified	numerical
Field of study	Fields of study list	Passed credits	numerical
Course	Formal/comprehensive	Kind of admission	ration
Total average	numerical	age	numerical

4.2. Data anomaly[3] recognition

One of the methods of preparing data is to find anomalous data from the tables and delete them in order to gain results that are more precise through models. To do this, we attach data mine to Anomaly node and next, we add the created model to the project, and then by adjustment of model, anomalous data in models creation are deleted. In the figure, you observe the way of model creation.

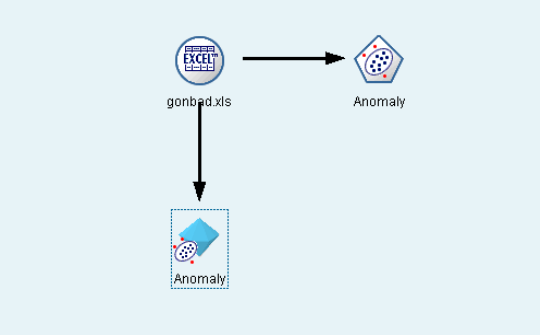


Figure (2): Data Anomaly Model

As you can see in figure number 2, first, we attached data mine to Anomaly algorithm and after creating the model, we added it to the project. Indeed, the technique of anomalous data recognition is use of clustering methods.

4.3. Feature Selection Technique

A technique is used to decrease the number of variables before implementation of data mining algorithm. Irrelevant

variables can have negative effect on prediction duty, or can complicate the calculations.
In order to implement algorithms on data banks, we decreased the number of required fields through this technique.
To do this work, first, we create a mine node (figure 3) and then link data mine, which Anomaly technique has been implemented on it to that feature selection.

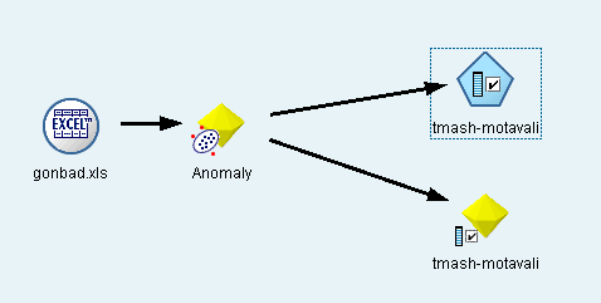


Figure (3): implementation of Feature Selection model

For this model, successively unqualified field was selected as target field and after algorithm implementation that 28 fields have been selected as inputs, 3 fields were deleted from list of fields and 25 fields were selected as inputs of neural network and CART algorithms.
This method is one of the supervisory learning techniques. When outcomes are two classes, linear regression is not so efficient; in this state, applying this technique is more appropriate. The other point is that this method is a nonlinear regression technique and it is not necessary for data to have linear state. If we want to say the reason for use of logistic regression, we should argue that in linear regression not only outcomes should be in numerical forms, but also variables should be in numerical forms as well. Hence, the states that are in sort form should change to numerical forms.

4.4. Filtration and deleting wide data

To make a better model, in addition to fields deleted by Feature Selection algorithm, using filter node we delete those fields that do not influence on models' output.
Therefore, from 28 total fields, 8 fields were deleted, and finally 15 fields were transferred to algorithms for model implementation.
In addition, using Select node, we deleted from the table those data, which their values were null.

5. Neural Network[4]

Neural Network or ANN is an instance of processing system, which is inspired from biological neural networks like brain. Key organ of this new structure is data processor system that a large number of them in the form of a complex work together like brain hormones in order to solve special problems like pattern detection or data classification through learning process. [3]

6. Clementine software

First, Ulshan, Friedman, Barmyman, and Stone designed CART algorithm for trees of regression and classification in 1984. The operation method of this algorithm is named Surrogate Spiliting. This algorithm includes a recurrent method. In every stage, CART algorithm divides instructional records into two subsets so that the records of each subset are more homogeneous than previous subsets. These divisions continue until conditions of stop are established. In CART, the best breaking point or assignment of the value of impurity parameter is determined. If the best break for a branch makes impurity less than the defined extent, that split will not be made. Here, the concept of impurity is assigned to the degree of similarity between target field value and records arrived to a node. If 100 percent of the samples in a node are placed in a specific category of the target field, that node is named pure. It is remarkable that in CART algorithm, a foreteller field may be often used in different levels of decision tree. In addition, this algorithm supports categorical and continues types of foreteller and target fields. [3]

7. Model Creation

Because of lengthiness of data preparation part, we have given up to cite its stages in this section.
To create the model, first we add a data mine node to the project and then add a partition node to the project in order to divide data to two parts of test and learning that 30 percent of data are test data and its 70 percent are instructional data.
Next stage after data conversion is data type determination using type node; after this stage, feature selection capability is used for determination of fields' importance in prediction. This leads to the selection of those fields that have more importance in prediction.
In subsequent stage, we add models, which are CART and Neural Net to the project.
When models are attached, we implement them and then link models' outputs to each other; and after linking models, using Ensemble node we combine them with each other. In figure 4, you observe the created model.

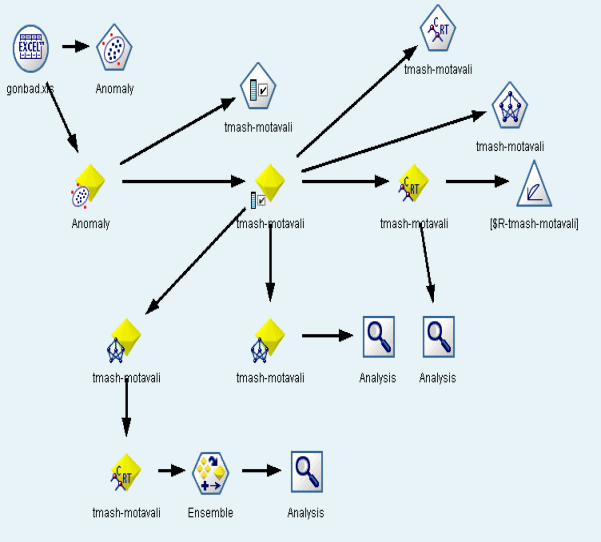


Figure (4): the created combinational model

7.1. Evaluation of Neural Network Model

To evaluate neural network model, first we add an analysis node to the project and then attach it to the created model and implement it.

Minimum Error	-1.68
Maximum Error	2.722
Mean Error	-0.045
Mean Absolute Error	0.28
Standard Deviation	0.35
Linear correlation	0.987
Occurrences	1,678

Minimum Error: the least difference between predicted data and observed data
Maximum Error: the most difference between predicted data and observed data
Mean Error: the observance of errors' average among the total of records
Mean Absolute Error: the observance of the average of errors' absolute value among the total of records
Standard Deviation: observance of Standard deviation of the total of errors
Linear Correlation: linear Correlation between observed data and real data
Occurrences: the number of records used in prediction

In figure 6, the degree of correlation or the importance of fields in target field has been shown that fields of successively unqualified and number of unqualified are the most important and field of term type is the least important one in target field prediction.

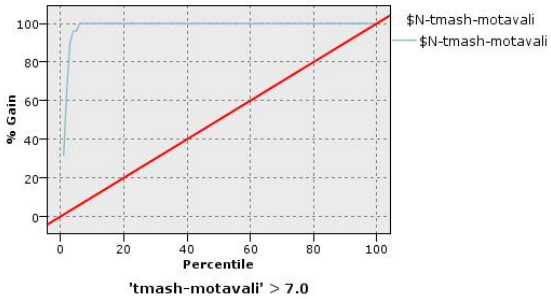


Figure (5): Neural Network's Gain Graph

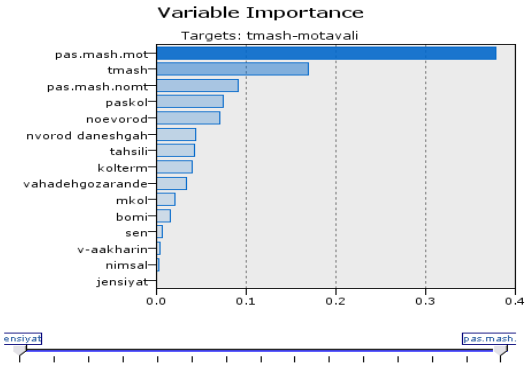


Figure (6): the importance of variables in neural network model

7.2. Evaluation of CART Model

In order to evaluate CART model, first we add an analysis node to the created model and implement it.

Minimum Error	-3.531
Maximum Error	4.469
Mean Error	-0.0
Mean Absolute Error	0.276
Standard Deviation	0.516
Linear correlation	0.971
Occurrences	1,678

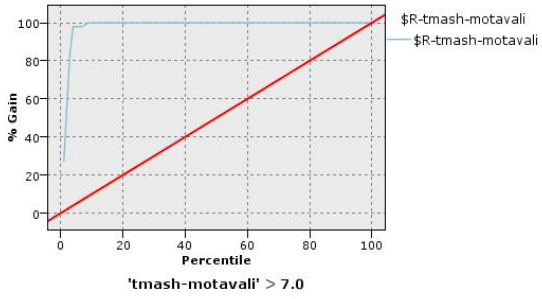


Figure (7): CART Model's Gain Graph

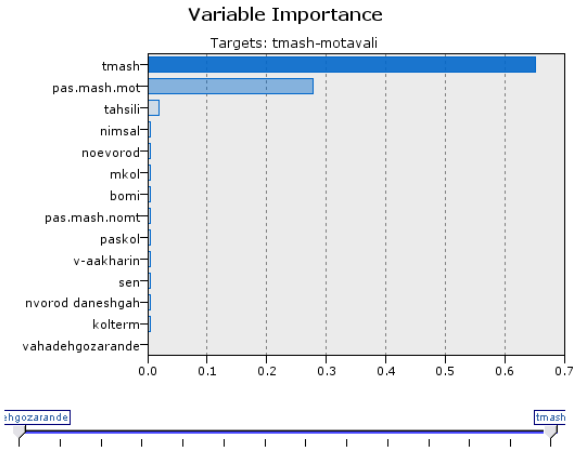
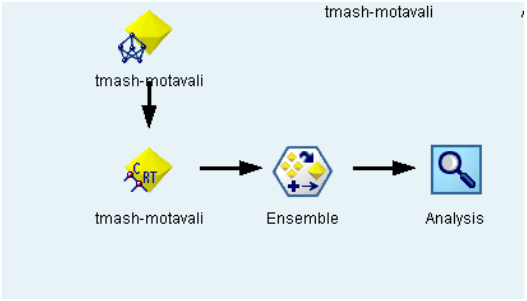


Figure (8): the importance of variables in CART Model

In this stage, using table No.1 that is a table with almost one million records for 5 successive years log of students' use of internet, we can find maximum, minimum, and average values in terms of kilobytes for students' monthly use.

7.3. Evaluation of Combinational Model

Ensemble node combines 2 or more models so that a more precision is gained compared to models that are implemented separately; through combining predictions from prediction models, limitations that are created in single models are deleted and a more precision can be achieved. Most of time combinational models have a better result than separate models. After using Ensemble node, we can use Analysis node to compare the results of combinational model to separate model.



Figure(9):Models' combination through ensemble node

As you see in figure number 9, after creation of 2 models we link them to each other and then combine them through Ensemble node. After implementation of combinational model, in order to evaluate model, we attach it to Analysis mode.

Minimum Error	-1.818
Maximum Error	3.596
Mean Error	-0.022
Mean Absolute Error	0.24
Standard Deviation	0.348
Linear correlation	0.992
Occurrences	1,678

As we see, the degree of linear correlation has increased to 0.992; that the precision of model has increased eminently and model's errors have decreased.

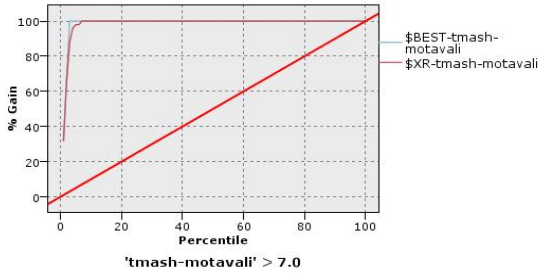


Figure (10): Combinational model's Gain Graph

8. Conclusion

Contemporary universities and higher education institutions are drowned in a mass of data and information that in most cases, their use is limited to current affairs and data are not yet used in strategic decision-making. Data mining which its use is developing day-to-day can lead to the use of existing information in higher education institutions and centers in strategic decision-making domains. In addition, created models from instructional data, using data mining algorithms can be utilized as a decision support system in educational systems and play an important role in advancement of universities' scientific level.

REFERENCES

[1] http://www.crisp-dm.org
[2] Andrew W. Moore. "Regression and Classification with Neural Networks". School of Computer Science Carnegie Mellon University. 2001.
[3] Mosavi, M.R., "A Comparative Study between Performance of Recurrent Neural Network and Kalman Filter for DGPS Corrections Prediction", IEEE Conference on Signal Processing (ICSP 2004), China, Vol.1, pp.356-359, August 31-4, September, 2004.
[4] Ramos, V. and Abraham, A., "Evolving a Stigmergic Self-Organized Data-Mining", Conf. on Intelligent Systems, Design and Applications (ISDA-04), 2004, pp.725-730.