Συνέδρια της Ελληνικής Επιστημονικής Ένωσης Τεχνολογιών Πληροφορίας & Επικοινωνιών στην Εκπαίδευση

Τόμ. 1 (2004)

4ο Συνέδριο ΕΤΠΕ «Οι ΤΠΕ στην Εκπαίδευση»

Comparing Regression Algorithms for Predicting Students' Marks in Hellenic Open University

S.B. Kotsiantis, P. E. Pintelas

Comparing Regression Algorithms for Predicting Students' Marks in Hellenic Open University

S.B. Kotsiantis, P. E. Pintelas
Department of Mathematics
University of Patras
Educational Software Development Laboratory
Hellas
{sotos, pintelas}@math.upatras.gr

ABSTRACT

The ability to provide assistance for a student at the appropriate level is invaluable in the learning process. Not only does it aid the student's learning process but also prevents problems, such as student frustration and floundering. Students' key demographic characteristics and their marks in a small number of written assignments can constitute the training set for a regression method in order to predict the student's performance. The scope of this work compares some of the state of the art regression algorithms in the application domain of predicting students' marks. A number of experiments have been conducted with six algorithms, which were trained using datasets provided by the Hellenic Open University. Finally, a prototype version of software support tool for tutors has been constructed implementing the M5rules algorithm, which proved to be the most appropriate among the tested regression algorithms.

KEYWORDS: *student performance, machine learning, data mining.*

INTRODUCTION

The application of Machine Learning Techniques in predicting students' performance proved to be helpful for identifying poor performers and it can enable tutors to take remedial measures at an earlier stage, even from the very beginning of an academic year using only students' demographic data, in order to provide additional help to the groups at risk (Kotsiantis et al., 2004). The diagnosis of students' performance is increased as new curriculum data is entered during the academic year, offering the tutors more effective results. Kotsiantis et al. (2004) showed that the most accurate machine learning algorithm for identifying predicted poor performers is the Naïve Bayes Classifier. However, that work could only predict if a student passes a course module or not.

This paper uses existing regression techniques in order to predict the students' marks in a distance learning system. It compares some of the state of the art regression algorithms to find out which algorithm is more appropriate not only to predict student's performance accurately but also to be used as an educational supporting tool for tutors. For the purpose of our study the 'informatics' course of the Hellenic Open University (HOU) provided the data set.

Generally, the usage of regression analysis to classify data can be an extremely useful tool for researchers and Open University administrators. A plethora of data can be utilized simultaneously to classify cases and the resultant model can be evaluated for usefulness relatively easily. The ability to develop a predictive model based on the model produced through the regression analysis procedure increases its usefulness substantially. Open Universities can utilize this dynamic and powerful procedure to target services and interventions to students who need it most, thereby utilizing their resources more effectively.

The following section describes in brief the Hellenic Open University (HOU) distance learning methodology and the data of our study. Some very basic definitions about regression techniques are given in section 3. Section 4 presents the experiment results for all the tested algorithms and at the same time compares these results. Section 5 presents the produced educational decision support tool. Finally, section 6 discusses the conclusions and some future research directions.

HELLENIC OPEN UNIVERSITY AND DATA DESCRIPTION

The mission of the Hellenic Open University (HOU) is to offer university level education using the distance learning methodology. The basic educational unit of the HOU is the course module (referred simply as module from now on) that covers a specific subject in graduate and postgraduate level. A module is equivalent to three semester academic lessons of Hellenic Universities while a student may register with up to three modules per year. The 'informatics' course of HOU is composed of 12 modules and leads to a Bachelor Degree. For the purpose of our study the 'informatics' course provided the training set. A total of 354 instances (student's records) have been collected from the module 'Introduction to Informatics' (INF10) (Xenos et al., 2002).

Regarding the INF10 module of HOU during an academic year students have to hand in 4 written assignments, optional participate in 4 face to face meetings with their tutor and sit for final examinations after an 11-month-period. A student with a mark >=5 'passes' a lesson or a module while a student with a mark <5 'fails' to complete a lesson or a module.

Generally, a student must submit at least three assignments (out of 4). Subsequently, the tutors evaluate these assignments and a mark greater or equal to 20 should be obtained in total in order that each student successfully completes the INF10 module. Students who meet the above criteria may sit the final examination test.

The attributes (features) of our dataset are presented in Table 1 along with the values of every attribute. The set of the attributes was divided in 3 groups. The 'Registry Class', the 'Tutor Class' and the 'Classroom Class'. The 'Registry Class' represents attributes which were collected from the Student's Registry of the HOU concerning students' sex, age, marital status, number of children and occupation. In addition to the above attributes, the previous –post high school– education in the field of informatics and the association between students' jobs and computer knowledge were also taken into account. If a student has attended at least a seminar (of 100 hours or more) on Informatics after high school then he/she would qualify as 'yes' in computer literacy. Moreover, students who use software packages (such as word processor) at their job without having any deep knowledge in informatics were considered as 'junior-users', while students who work as programmers or in data processing departments were considered a 'senior users'. The remaining students' jobs were listed as 'no' concerning association with computers.

'Tutor Class' represents attributes, which were collected from tutors' records concerning students' marks on the written assignments and their presence or absence in face-to-face meetings. Finally, the 'class attribute' represents the result on the final examination test.

According to the data collected in the framework of this research, the students' age follows a normal distribution with an average value 31.1 years (±5.1). It must be noted that no students under the age of 24 years can be accepted according to the regulation of HOU, since it is considered that such students could easily attend conventional Hellenic Universities.

The analysis of the demographic attributes showed that the ratio of men who passed the exams vs. men who failed is 48–52%, while for women this ratio drops to 39–61%. Moreover, it should be noted that the percentage of students below 32 years old that pass the exams is measured 46%, when the corresponding number for older students is 44%. Another interesting fact is related to student performance and their marital status. It is just as possible for a married student to pass the exams (51%) while a single student has only 41% probability to pass the module. A similar situation holds with the existence of children, a student with children has 52% probability to pass the module while a student without children has only 43%. This is probably due to the fact that the family obligations is known and has been taken under consideration prior to the commencement of the studies. It must be also mentioned that the workload separates the probabilities just in the middle.

Student's Registry (demographic) attributes	Sex	male, female
	Age	24-46
	Marital status	single, married, divorced, widowed
	Number of children	none, one, two or more
	Occupation	no, part-time, fulltime
	Computer literacy	no, yes
	Job associated with computers	no, junior-user, senior-user
Attributes from tutors' records	1st face to face meeting	Absent, present
	1st written assignment	no, 0-10
	2nd face to face meeting	absent, present
	2nd written assignment	no, 0-10
	3rd face to face meeting	absent, present
	3rd written assignment	no, 0-10
	4th face to face meeting	absent, present
	4th written assignment	no, 0-10
Class	Final examination test	0-10

Table 1. The attributes used and their values

On the contrary, as far as the demographic attributes are concerned, stronger correlation exists between student performance and the existence of previous education in the field of Informatics. The ratio of students who have previous education in the field of Informatics and pass the exams vs. them who fail is 51–49%, while for the remaining students this ratio drops to 28–72%. A similar correlation exists between the involvements in professional activities demanding the use of computer. The students who use the computer in their job have 52% probability to pass the module while the remaining students have only 32%.

Until now, we have described how each demographic attribute influences the prediction based on our dataset. In order to show in which direction (pass or fail) each of the remaining attributes' values push the induction in Table 2 some practical probabilities are estimated. The interpretation of Table 2 is easy enough and it shows, for example, that a student with a mark more than 6 in WRI-4, has about 4 times more probabilities to pass than fail (0.65/0.17).

Subsequently, in an attempt to show how much each attribute influences the induction, we rank the influence of each one according to a statistical measure – RRELIEF (Sikonja and Kononenko, 1997). The key idea of the RRELIEF algorithm is to estimate the quality of attributes according to how well their values distinguish between the instances that are near to each other. In regression problems the predicted value (class) is continuous, therefore the (nearest) hits and misses cannot be used. Instead of requiring the exact knowledge of whether two instances belong to the same class or not, we can introduce a kind of probability that the predicted values of two instances are different. This probability can be modeled with the relative distance between the predicted (class) values of the two instances.

Attribute	Value	Pass	Fail
WRI-4	Mark<3	0.04	0.68
	3=<Mark=<6	0.31	0.15
	Mark>6	0.65	0.17
WRI-3	Mark<3	0.03	0.61
	3=<Mark=<6	0.21	0.2
	Mark>6	0.66	0.19
WRI-2	Mark<3	0.08	0.52
	3=<Mark=<6	0.15	0.26
	Mark>6	0.77	0.22
FTOF-4	Absent	0.23	0.76
	Present	0.77	0.24
FTOF-3	Absent	0.2	0.65
	Present	0.8	0.35
WRI-1	Mark<3	0.02	0.19
	3=<Mark=<6	0.14	0.35
	Mark>6	0.84	0.46
FTOF-2	Absent	0.22	0.54
	Present	0.78	0.46

Table 2.

The average RRELIEF score of each attribute according to our dataset are presented in Table 4. The larger the value of the RRELIEF scores is, the more influence of the attribute in the induction.

Attribute	RRELIEF
W_ASS-4	0.11799
W_ASS-3	0.09263
W_ASS-2	0.03932
sex	0.01266
F_MEET1	0.0104
F_MEET4	0.00989
children	0.00307
Job associated with computers	0.00102
domestic	-0.00563
F_MEET3	-0.00601
occupation	-0.00903
F_MEET2	-0.00931
Computer Knowledge	-0.01091
age	-0.01416
W_ASS-1	-0.03098

Table 3. The average RRELIEF score of each attribute

Thus, the demographic attributes that mostly influence the induction are the ‘sex’ and the ‘children’. In addition, it was found that 1st written assignment has not a large value of influence. The reason is that almost all students try harder with the first written assignment thus making the offered information of this attribute minimal and maybe confusing.

REGRESSION ISSUES

The problem of regression consists in obtaining a functional model that relates the value of a target continuous variable *y* with the values of variables $x_1, x_2, ..., x_n$ (the predictors). This model is obtained using samples of the unknown regression function. These samples describe different mappings between the predictor and the target variables.

For the propose of our comparison the six most common regression techniques namely Model Trees (Wang & Witten, 1997), Neural Networks (Mitchell, 1997), Linear regression (Fox, 1997), Locally weighted linear regression (Atkeson et al., 1997) and Support Vector Machines (Shevade et al., 2000) are used. In the following we will briefly describe these regression techniques.

Linear regression is the simplest statistical technique used to find the best-fitting linear relationship between the class and its predictors (other features).

$$y = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}$$

Find values of beta that minimize *Q:*

$$Q = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik}))^2$$

Note that nominal features with *n* values are converted into *n*-1 binary features and a Wald test is used to test the statistical significance of each coefficient (β_ι) in the model (Fox, 1997). A standard linear regression method may employ an attribute deletion strategy, which simplifies the prediction task.

Model trees are the counterpart of decision trees for regression tasks. Model trees are trees that classify instances by sorting them based on attribute values. Instances are classified starting at the root node and sorting them based on their attribute values. The most well known model tree inducer is the M5΄ (Wang & Witten, 1997). A model tree is generated in two stages. The first builds an ordinary decision tree, using as splitting criterion the maximization of the intra-subset variation of the target value (Witten & Frank, 2000). The second prunes this tree back by replacing subtrees with linear regression functions wherever this seems appropriate. If this step is omitted and the target is taken to be the average target value of training examples that reach this leaf, then the tree is called a "regression tree" instead. Although the models trees are smaller and more accurate than the regression trees, the regression trees are more comprehensible (Witten & Frank, 2000).

M5rules algorithm produces propositional regression rules in IF-THEN rule format using routines for generating a decision list from M5΄Model trees (Witten & Frank, 2000). The algorithm is able to deal with both continuous and nominal variables, and obtains a piecewise linear model of the data.

Artificial Neural Networks (ANNs) are another method of inductive learning based on computational models of biological neurons and networks of neurons as found in the central nervous system of humans (Mitchell, 1997). A multi layer neural network consists of large number of units (neurons) joined together in a pattern of connections. Units in a net are usually segregated into three classes: input units, which receive information to be processed, output units where the results of the processing are found, and units in between called hidden units. Regression with a neural network takes place in two distinct phases. First, the network is trained on a set of paired data to determine the input-output mapping. The weights of the connections between neurons are then fixed and the network is used to predict the numerical class values of a new set of data.

Locally weighted linear regression (LWR) is a combination of instance-based learning and linear regression (Atkeson et al., 1997). Instead of performing a linear regression on the full, unweighted dataset, it performs a weighted linear regression, weighting the training instances according to their distance to the test instance at hand. This means that a linear regression has to be done for each new test instance, which makes the method computationally quite expensive. However, it also makes it highly flexible, and enables it to approximate non-linear target functions.

The sequential minimal optimization algorithm (SMO) has been shown to be an effective method for training support vector machines (SVMs) on classification tasks defined on sparse data sets (Platt, 1999). SMO differs from most SVM algorithms in that it does not require a quadratic programming solver. Shevade et al. (2000) generalize SMO so that it can handle regression problems. This implementation globally replaces all missing values and transforms nominal attributes into binary ones.

For the regression methods, there isn't only one regressor's criterion. Table 4 represents the most well known. Fortunately, it turns out for in most practical situations the best regression method is still the best no matter which error measure is used.

Mean absolute error	$\frac{\lvert p_1 - a_1 \rvert + \ldots + \lvert p_n - a_n \rvert}{n}$
Root mean squared error	$\sqrt{\frac{(p_1 - a_1)^2 + \ldots + (p_1 - a_1)^2}{n}}$
Relative absolute error	$\frac{\lvert p_1 - a_1 \rvert + \ldots + \lvert p_n - a_n \rvert}{\lvert a_1 - \bar{a} \rvert + \ldots + \lvert a_n - \bar{a} \rvert}$
Root relative squared error	$\sqrt{\frac{(p_1 - a_1)^2 + \ldots + (p_1 - a_1)^2}{(a_1 - \bar{a})^2 + \ldots + (a_1 - \bar{a})^2}}$

Table 4. Regressor criteria (p : predicted values, a : actual values, $\bar{a} = \frac{1}{n}\sum_i a_i$ *)*

EXPERIMENTS RESULTS

The learning algorithms are useful as a tool for identifying predicted poor performers (Kotsiantis et al., 2003). With the help of machine learning the tutors will be in position to know from the beginning of the module, based only on curriculum-based data of the students whose of them will complete the module with enough accurate precision, which reaches 64% in the initial forecasts and exceeds 80% before the middle of the period (Kotsiantis et al., 2004). After the middle of the period, we can use existing regression techniques in order to predict the students' marks.

The experiments took place in two distinct phases. During the first phase (training phase) the algorithms were trained using the data collected from the academic year 2000-1. The training phase was divided in 5 consecutive steps. The 1st step included the demographic data, the two first face-to-face meetings and written assignments as well as the resulting class (final mark). The 2nd step additionally included the third face-to-face meeting. The 3rd step additionally included the third written assignment. The 4th step additionally included the fourth face-to-face meeting and finally the 5th step that included all attributes described in Table 1.

Subsequently, ten groups of data for the new academic year (2001-2) were collected from 10 tutors and the corresponding data from the HOU registry. Each one of these 10 groups was used to measure the accuracy within these groups (testing phase). The testing phase also took place in 5 steps. During the 1st step, the demographic data as well as the two first face-to-face meetings and written assignments of the new academic year were used to predict the class (final student mark) of each student. This step was repeated 10 times (for every tutor's data). During the 2nd step these demographic data along with the data from the third face-to-face meeting were used in order to predict the class of each student. This step was also repeated 10 times. During the 3rd step the data of the 2nd step along with the data from the third written assignment were used in order to predict the student class. The remaining steps use data of the new academic year in the same way as described above. These steps are also repeated 10 times.

It must be mentioned that we used the free available source code by (Witten and Frank, 2000) for our experiments. We have tried to minimize the effect of any expert bias by not attempting to tune any of the algorithms to the specific data set. Wherever possible, default values of learning parameters were used. This naïve approach results in lower estimates of the true mean absolute error, but it is a bias that affects all the learning algorithms equally.

In Table 5, the most easily understandable measure - mean absolute error - of each algorithm for all the testing steps of the experiment is presented.

	M5΄	*BP*	*LR*	*LWR*	*SMOreg*	*M5rules*
WRI-2	1.83	2.15	1.89	1.84	1.84	1.83
FTOF-3	1.74	2.08	1.83	1.79	1.78	1.74
WRI-3	1.55	1.79	1.6	1.53	1.56	1.55
FTOF-4	1.54	1.8	1.56	1.5	1.55	1.54
WRI-4	1.23	1.65	1.5	1.4	1.44	1.21

Table 5. Mean absolute error

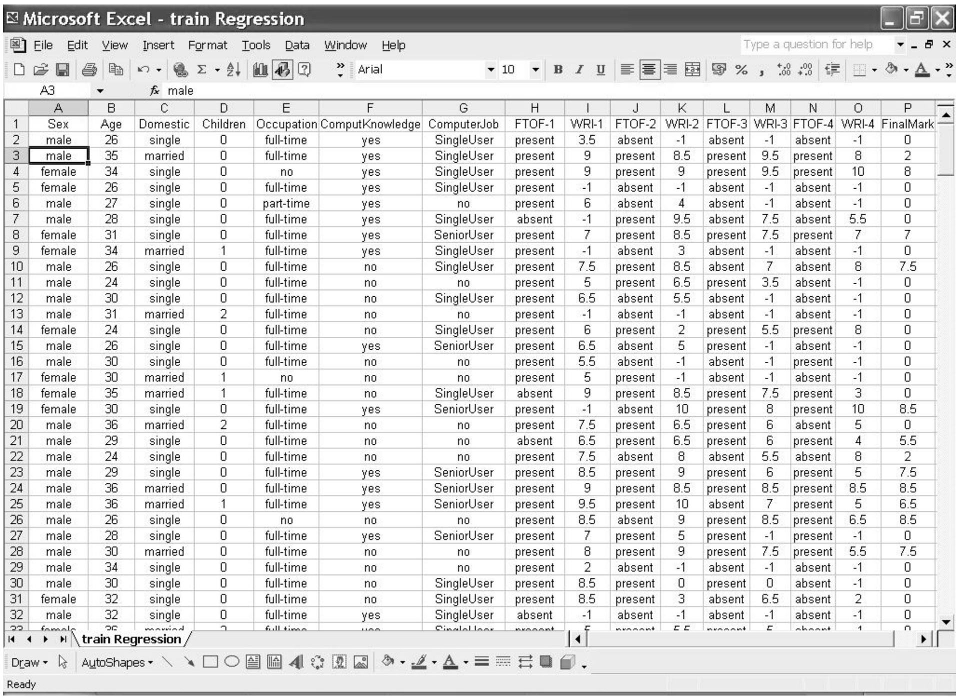
According to the results, the M5rules is the most accurate regression algorithm to be used for the construction of a software support tool. An advantage of M5rules except for its better performance is its comprehensibility.

SOFTWARE SUPPORT TOOL

A prototype version of the software support tool has already been constructed and is in use by the tutors. The tool expects the training set as a spreadsheet in CSV (Comma-Separated Value) file format (Figure 1). The tool assumes that the first row of the CSV file is used for the names of the attributes. There is not any restriction in attributes' order. However, the class attribute must be in the last column. It must be mentioned that the used attributes are not a conclusive list. An extension can introduce new attributes that were not in the current database, but are collectable by tutors and may potentially contribute to the prediction of academic achievement. For example, measures of different intellectual abilities, interests, motivation, and personality traits of students.

Once the database is in a single relation, each attribute is automatically examined to determine its data type (for example, whether it contains numeric or symbolic information). A feature must have the value ? to indicate that no measurement was recorded. After opening the data set that characterizes the problem for which the user wants to take the prediction, the tool automatically uses the corresponding attributes for training.

After the training of the model, the user is able to see the produced regressor. The tool (Figure 2) can also predict the output of either a single instance or an entire set of instances (batch of instances). It must be mentioned that for batch of instances the user must import an Excel cvs file with all the instances he/she wants to have predictions.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Sex	Age	Domestic	Children	Occupation	ComputKnowledge	ComputerJob	FTOF-1	WRI-1	FTOF-2	WRI-2	FTOF-3	WRI-3	FTOF-4	WRI-4	FinalMark
2	male	26	single	0	full-time	yes	SingleUser	present	3.5	absent	-1	absent	-1	absent	-1	0
3	male	35	married	0	full-time	yes	SingleUser	present	9	present	8.5	present	9.5	present	8	2
4	female	34	single	0	no	yes	SingleUser	present	9	present	9	present	9.5	present	10	8
5	female	26	single	0	full-time	yes	SingleUser	present	-1	absent	-1	absent	-1	absent	-1	0
6	male	27	single	0	part-time	yes	no	present	6	absent	4	absent	-1	absent	-1	0
7	male	28	single	0	full-time	yes	SingleUser	absent	-1	present	9.5	absent	7.5	absent	5.5	0
8	female	31	single	0	full-time	yes	SeniorUser	present	7	present	8.5	present	7.5	present	7	7
9	female	34	married	1	full-time	yes	SingleUser	present	-1	absent	3	absent	-1	absent	-1	0
10	male	26	single	0	full-time	no	SingleUser	present	7.5	present	8.5	absent	7	absent	8	7.5
11	male	24	single	0	full-time	no	no	present	5	present	6.5	present	3.5	absent	-1	0
12	male	30	single	0	full-time	no	SingleUser	present	6.5	absent	5.5	absent	-1	absent	-1	0
13	male	31	married	2	full-time	no	no	present	-1	absent	-1	absent	-1	absent	-1	0
14	female	24	single	0	full-time	no	SingleUser	present	6	present	2	present	5.5	present	8	0
15	male	26	single	0	full-time	yes	SeniorUser	present	6.5	absent	5	present	-1	absent	-1	0
16	male	30	single	0	full-time	no	no	present	5.5	absent	-1	absent	-1	present	-1	0
17	female	30	married	1	no	no	no	present	5	present	-1	absent	-1	absent	-1	0
18	female	35	married	1	full-time	no	SingleUser	absent	9	present	8.5	present	7.5	present	3	0
19	female	30	single	0	full-time	yes	SeniorUser	present	-1	absent	10	present	8	present	10	8.5
20	male	36	married	2	full-time	no	no	present	7.5	present	6.5	present	6	absent	5	0
21	male	29	single	0	full-time	no	no	absent	6.5	present	6.5	present	6	present	4	5.5
22	male	24	single	0	full-time	no	no	present	7.5	absent	8	absent	5.5	absent	8	2
23	male	29	single	0	full-time	yes	SeniorUser	present	8.5	present	9	present	6	present	5	7.5
24	male	36	married	0	full-time	yes	SeniorUser	present	9	present	8.5	present	8.5	present	8.5	8.5
25	male	36	married	1	full-time	yes	SeniorUser	present	9.5	present	10	absent	7	present	5	6.5
26	male	26	single	0	no	no	no	present	8.5	absent	9	present	8.5	present	6.5	8.5
27	male	28	single	0	full-time	yes	SeniorUser	present	7	present	5	present	-1	present	-1	0
28	male	30	married	0	full-time	no	no	present	8	present	9	present	7.5	present	5.5	7.5
29	male	34	single	0	full-time	no	no	present	2	absent	-1	absent	-1	absent	-1	0
30	male	30	single	0	full-time	no	SingleUser	present	8.5	present	0	present	0	absent	-1	0
31	female	32	single	0	full-time	no	SingleUser	present	8.5	present	3	absent	6.5	absent	2	0
32	male	32	single	0	full-time	yes	SingleUser	absent	-1	absent	-1	absent	-1	absent	-1	0

Figure 1. The CSV file of the use case

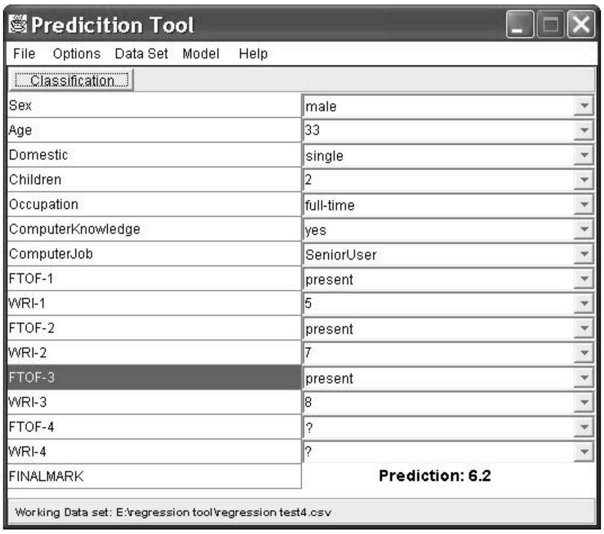


Figure 2. The prototype tool

The ranking of the attributes' influence brought considerable benefits; by helping the tutors to better understand the characteristics of the population that mostly affect academic achievement. For example, the prototype tool for the used dataset shows that the attributes that mostly influence the induction are the 'WRI-4' and the 'WRI-3' (Figure 3).

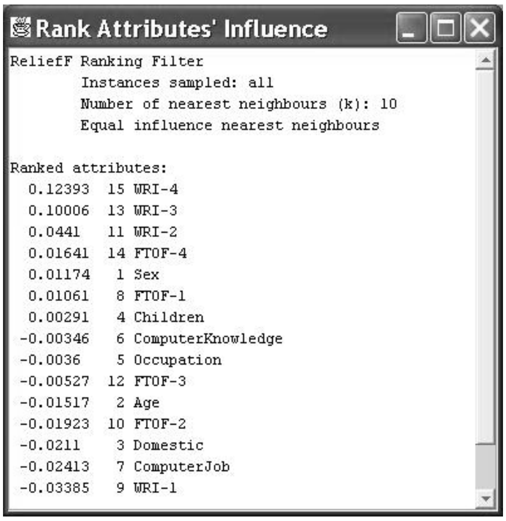


Figure 3. Ranking the attributes' influence to the final prediction in our use case

What is more, the implemented tool can present useful information about the imported data set such as the presence or not of missing attribute values, the frequency of each attribute value etc. Finally, the tool provides on-line help for novice users.

CONCLUSION

This paper aims to fill the gap between empirical prediction of student performance and the existing regression techniques. Our data set is from the module INFO but most of the conclusions are wide-ranging and present interest for the majority of programs of study of Hellenic Open University and more generally for all the distance education programs. It would be interesting to compare our results with those from other open and distance learning programs offered by other open Universities. So far, however, we have not been able to find such results.

Generally, the education domain offers many interesting and challenging applications for data mining. Firstly, an educational institution often has many diverse and varied sources of information. There are the traditional databases (e.g. students' information, teachers' information, class and schedule information, alumni information), online information (online web pages and course content pages) and more recently, multimedia databases. Secondly, there are many diverse interest groups in the educational domain that give rise to many interesting mining requirements. For example, the administrators may wish to find out information such as admission requirements and to predict the class enrollment size for timetabling. The students may wish to know how best to select courses based on prediction of how well they will perform in the courses selected. With so much information and so many diverse needs, it is foreseeable that an integrated data mining system that is able to cater for the special needs of an education institution will be in great demand particularly in the 21st century.

In a next study we intend to apply data mining methods with the goals of answering the following two research questions:
1) Do there exist groups of students who use online resources in a similar way? If so, can we identify the class an individual student belongs to? Based on the usage of the resource by other students in the group, can we help a new student use the resources better?
2) Can we classify the learning difficulties of the students? If so, can we show how different types of problems impact students' achievement? Can we help instructors to develop the homework more effectively and efficiently?

APPENDIX

The tool is available in the web page: http://www.math.upatras.gr/~esdlab/Regression-tool/
The Java Virtual Machine (JVM) 1.2 or newer is needed for the execution of the program.

REFERENCES

Atkeson, C. G., Moore, A.W., & Schaal, S. (1997). Locally weighted learning. *Artificial Intelligence Review*, 11, 11–73.

Fox, J. (1997), *Applied Regression Analysis, Linear Models, and Related Methods*, ISBN: 080394540X, Sage Pubns.

Kotsiantis, S., Pierrakeas, C., Pintelas, P.(2003), Preventing student dropout in distance learning systems using machine learning techniques, Proceedings of *Seventh International Conference on Knowledge-Based Intelligent Information & Engineering Systems*, 3-5 September 2003, Lecture notes in AI, Springer-Verlag Vol 2774, pp 267-274..

Kotsiantis S., Pierrakeas C., Pintelas P. (2004), Predicting Students' Performance in Distance Learning Using Machine Learning Techniques, Accepted for publication in *Applied Artificial Intelligence* (AAI).

Mitchell, T. (1997), *Machine Learning*. McGraw Hill.

Platt, J. (1999). Using sparseness and analytic QP to speed training of support vector machines. In: Kearns, M. S., Solla, S. A. & Cohn D. A. (Eds.), *Advances in neural information processing systems* 11. MA: MIT Press.

Shevade, S., Keerthi, S., Bhattacharyya C., and Murthy, K. (2000). Improvements to the SMO algorithm for SVM regression. *IEEE Transaction on Neural Networks*, 11(5):1188-1183.

Sikonja M. and Kononenko I. (1997), An adaptation of Relief for attribute estimation in regression, *Proceedings of the Fourteenth International Conference (ICML'97)*, ed., Dough Fisher, pp. 296-304. Morgan Kaufmann Publishers.

Wang, Y. & Witten, I. H. (1997). Induction of model trees for predicting continuous classes, In *Proc. of the Poster Papers of the European Conference on ML*, Prague (pp. 128–137). Prague: University of Economics, Faculty of Informatics and Statistics.

Witten, I.H., Frank, E. (2000), *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann, San Mateo, CA, 2000.

Xenos, M., Pierrakeas C. and Pintelas P. (2002). A survey on student dropout rates and dropout causes concerning the students in the course of informatics of the Hellenic Open University, *Computers & Education* (39): 361–377.