Combined Data Mining and Decision Support Approach to the Prediction of Academic Achievement

Silvana Gasar [1], Marko Bohanec [2,3], Vladislav Rajkovič [4,2]

[1]High School Jesenice, Ruparjeva 2, SI-4270 Jesenice, Slovenia
silvana.gasar@telesat.si
[2]Institute Jožef Stefan, Jamova 39, SI-1000 Ljubljana, Slovenia
[3]University of Ljubljana, School of Public Administration Ljubljana, Slovenia
marko.bohanec@ijs.si
[4]University of Maribor, Faculty of Organisational Sciences, Kranj, Slovenia
vladislav.rajkovic@fov.uni-mb.si

Abstract. We present the development of multi-attribute hierarchical models for the prediction of final academic achievement in a particular high-school educational program. The models were developed by a sequential application of data mining (DM) and decision support (DS) techniques. A database of pupils' achievements was first analyzed by DM methods: statistical analysis, clustering, decision trees and hierarchical multi-attribute models. The findings were incorporated into expert-developed DS models. Predictive accuracy of these models is comparable to that of experienced human experts.

1 Introduction

Data Mining (DM) and Decision Support (DS) are complementary modeling disciplines; DS [1] tends to rely on knowledge acquired from experts, while DM [2] attempts to extract it from data. Recently, Bohanec and Zupan [3] proposed an approach that combines DM and DS for the development of qualitative hierarchical multi-attribute models. The approach combines two methods: DEX [4] as a DS method for model development based on expert knowledge, and HINT [5] as a DM method that discovers concepts and models from data. Since both methods share a common model representation, they can be combined in a number of ways, such as supervised, serial, or parallel. These modes of operations were demonstrated on a real-life case of housing loan allocation [3], indicating a considerable improvement of classification accuracy and comprehensibility of models developed in a combined way. However, that study had a weak point: it was based on a case that had been originally approached only by DEX, and for the integration of DEX and HINT, the case was revisited several years later in a somewhat hypothetical setting. Thus, the study explicitly recommended further practical evaluation of the approach.

In this paper, we present such a real-life case in which the combination of DM and DS methods has taken place from the beginning. The case is in the area of education: the aim was to develop a hierarchical multi-attribute decision model for the prediction of final academic achievement in a particular high-school educational program. We used a database of pupils collected in one of Slovenian high schools. In order to discover the patterns and indicators that determine academic success or failure, this database was analyzed by a number of DM methods, including basic statistical analysis, clustering, and machine learning of decision trees and hierarchical multi-attribute models. These findings were then taken into account in developing a predictive multi-attribute model, which was done in a DS way by involving an expert and using DEX. In the final stage, the model was thoroughly evaluated from the viewpoint of its predictive accuracy and suitability for practice.

This paper is organized as follows. Section 2 describes the problem of academic achievement prediction, and formulates research questions, goals and methodology. The results of data mining are presented in section 3. These results were combined with expert knowledge to develop two models, which are presented in section 4. The quality of the models is assessed in section 5. The paper is concluded by a summary and recommendations for further research.

2 Problem

Academic achievement depends on the consistency between individual's features and demands of school. Therefore, the problem of high school failure has its roots mostly in an inappropriate choice of school. The choice of school or profession is a multi-attribute decision-making process in which the choice takes place at both sides. The goals of pupils and schools are often in conflict: pupils wish to choose the most appropriate school for themselves, and schools want to select only the best candidates. Because of uncertainty involved in the process, there is always a risk of inadequate choice of school with negative consequences. Furthermore, when passing from primary to high school, pupils are still immature; they do not know themselves and they are not fully aware of different opportunities and demands of further education. They need professional advice for choosing schools and educational programs [6].

Educational and professional counseling in our country, Slovenia, is provided mostly by schools' counseling services. A study of their work [7] revealed a number of problems. Counselors are too busy for educational counseling of high quality. Their suggestions are usually intuitive and based on very few and often incomplete data. All schools try to reject "bad" pupils. There is little time and there is a lack of alternatives. At the time of high schools enrolment it is already too late for appropriate activities to prevent academic failure. Because of the conflict situation and subjective advice, pupils and their parents rarely consider counselors' warnings.

Therefore, counselors need a tool for the prediction of final academic achievement in an easy understanding way with high accuracy. Such a tool would provide an opportunity to predict school failure and react properly on time. Indirectly, it would also help to decrease some undesirable social phenomena such as unemployment, delinquency, drug addiction, violence, etc. [8]. In general, academic achievement depends on the interaction of physical, physiological, social and psychological factors [9], which provide a basis for multi-attribute modeling.

Our research goal was to develop a multi attribute model for the prediction of final academic achievement on an individual high school educational program. We wanted to validate the model, to show its strengths and weaknesses, and give recommendations for its application. In particular, we were looking for the answers on the following questions:

- Is it possible to discover general patterns and rules of academic achievements from a database of pupils, such as the one commonly used in Slovenian primary and high schools?
- On this ground, is it possible to build a multi-attribute model for the prediction of final academic achievement on an individual high school educational program?
- How accurate is the prediction of such models?
- If and how can models contribute to the quality of achievements' estimations and quality of school choices?
- Could they improve achievements in general?
- How to implement the approach in practice?

The methodology involved is a combination of DM and DS methods, which were used sequentially as presented in the following two sections.

3 Data Mining

The DM stage was carried out using statistical methods, visualization, clustering, and machine learning. Among machine learning methods we used the development of classification decision trees and development of multi-attribute hierarchical models. We used the tools SPSS [10], Weka [11] and Orange [12]. In accordance with a general DM methodology [2], the analysis was preceded by data preparation and pre-processing, and followed by the interpretation and evaluation of results.

3.1 Data Preparation and Pre-Processing

The analysis was based on a pupils' database that was created in one of Slovenian high schools using a computer program Evidenca [13]. The database was exported to Microsoft SQL Server 2000 [14], which was used as a tool for data preparation. After normalization, the database was integrated into a single table, in which each record contained all data available about one pupil. In total, there are 96 attributes. A part, 19 attributes, is known before enrolment in high school, while the remaining 77 attributes represent school marks and other data obtained in successive grades of the high school. The main groups of attributes are:

- Pupil's personal and demographic data: gender, date and town of birth, citizenship, primary school name, etc.
- Data on academic achievements in primary school: individual subject marks and general achievement marks in the last two years of primary education.

- Data on achievements and behavior in the first, second, third and fourth high school grade: individual subject marks, general achievement mark, discipline sanctions, hours of excused and unexcused absence from school, etc.

For each pupil in the database, his or her academic achievement is already known. It is represented by the following five categories

- 5: graduates with general achievement mark 4 or 5 (B or A) after four years;
- 4: graduates with general achievement mark 3 or 2 (C or D) after four years;
- 3: graduates after five or six years (prolonged time of education);
- 2: fails and stops educating after one or two years;
- 1: fails and stops educating after tree or more years.

This database–hereafter referred to as DB1–contained data about $N = 1794$ pupils. All the records contained complete data known before the enrolment and data about final academic achievements. However, a considerable proportion of data for all school years was incomplete. Therefore, we also created a smaller database, DB2, of $N = 889$ pupils for which complete data was available for all school years.

3.2 Basic Statistical Analysis

We started the analysis with establishing general statistics, measures of correlations between variables and visualization in SPSS. Descriptive statistics and frequency distributions of variables were assessed for both databases, and Spearman coefficients of rank correlation were computed between numerical variables and final academic achievement. The relation between nominal variables and final achievement was established using chi-square test and contingency coefficient.

Table 1: Frequency distribution of achievement categories in DB1.

Category	**1**	**2**	**3**	**4**	**5**
Frequency [%]	12.3	11.9	14.6	51.2	9.9

The distribution of achievement categories is shown in Table 1. It turned out that academic failure (categories 1 and 2) was quite common – more than 20% pupils left the school and never graduated. Such level of failure would be considered a disaster in every work organization [8], but surprisingly not in schools. High percentage of failure in our high schools confirms a poor performance of professional counseling.

The majority of pupils graduate in time with general achievement mark 3 or 2 (C or D in American system). In our analysis, they represent the majority category 4 with apriori classification accuracy of 51.2 %. General achievement marks in the first high school grade are on average one or two marks lower than general achievement marks in primary school. Most marks are between 1 and 3 (E and C), and pupils with 4 and 5 (B and A) are very rare.

Absences from school and discipline sanctions are in negative, while different marks in primary and high school are in positive correlation with achievements. Achievements most strongly correlate with general achievement marks at the end of grades. Pupils' age at high school enrolment negatively correlates with achievements. Obviously, predictions based on "later" data are more accurate and valid, but also less useful in practice, as they come too late for a successful corrective action.

Statistical analyses also revealed some interesting and unexpected findings, such that particular subject marks from a particular teacher have no validity at all, what deserves serious consideration and appropriate actions of school's management.

3.3 Machine Learning of Decision Trees

Classification decision trees were built with the program Weka, using the algorithm J4.8 [11], a version of well-known Quinlan's algorithm C4.5 [15]. An initial decision tree was built from all attributes. Different decision trees were then developed on the basis of different selections of attributes. Classification accuracy was estimated by 10-fold cross validation.

When using all the attributes, the classification accuracy turned out as high as 99.39%. This means that at the end of high-school education the prediction of final achievement is almost certain. However, this is not practical as we wish to make predictions several years earlier, possibly at the enrolment in high school or at least after the first high school grade. When we limit the prediction to attributes that are known at that time, the classification accuracy of corresponding decision trees degrades considerably. Figure 1 shows that the classification accuracy on the basis of data known before enrolment (16 attributes) only slightly exceeds 50%. At the end of the first grade (30 attributes), the classification accuracy improves to about 60%, and one year later (46 attributes) to slightly below 70%.

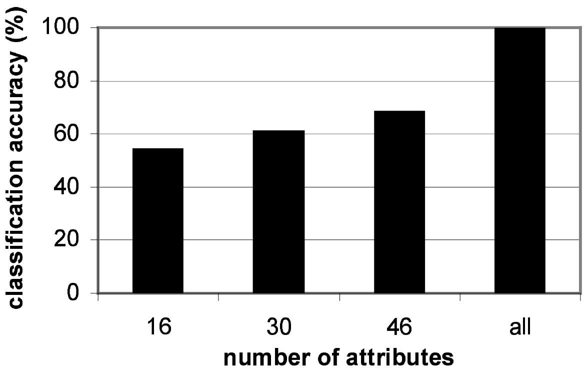

Fig. 1. Classification accuracy of decision trees developed from different attributes

In order to improve classification accuracy, we conducted a number of additional experiments: (1) developing trees on data corresponding to specific educational programs, (2) using different achievement classifications, (3) using cost-sensitive classification, and (4) developing trees on DB2 instead of DB1. The results revealed that developing trees with respect to educational program with attributes known at least at the end of first high school grade seem the most rational. Different achievement classifications and cost-sensitive classification did not improve classification accuracy. Developing trees on DB2 resulted in smaller trees with significantly better classification accuracy – often over 10% better than with DB1.

We concluded that for developing multi-attribute decision models, the best basis provide the trees developed on DB2, corresponding to specific educational program, and using 16 or 30 expert-selected attributes. Despite considerable differences between trees, it seems that the best predictor among the attributes known before enrolment is the final general achievement mark of primary school, and among attributes known at the end of first high school grade the final general achievement mark of the first high school grade.

Other good predictors depend on the educational program. For example, the best tree using 30 attributes for educational program referred to as "L" is shown in Figure 2. Among attributes known at the end of the first grade the most important are the first grade general achievement mark, subject marks in Slovene language, History and Physics in the first high school grade, age at the high school enrolment and unexcused absence in the third semester. These attributes indicate that the program "L" demands general intelligence, verbal abilities, work habits, memory and logical abilities and that the absence diminishes the possibility of success.

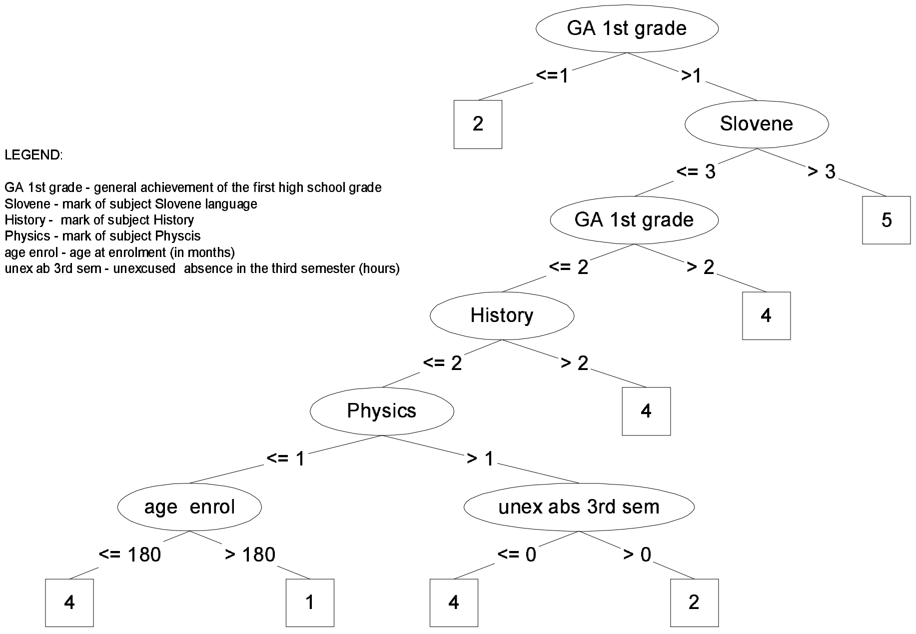

Fig.2. The best tree for educational program "L" (*n* = 468, classification accuracy is 69.7%)

In summary, classification accuracy of trees using attributes known before enrolment is generally low (about 60%) and slightly exceeds the apriori accuracy (51.2%). Somewhat better, about 70%, is the classification accuracy of decision trees developed from attributes known at the end of first high school grade.

3.4 Clustering

In Weka, pupils were clustered into three, four and five groups using the *k*-means algorithm [16]. Two sets of attributes were used: (1) 16 expert-selected attributes, all known before enrolment in high school, and (2) 30 attributes, known at the end of first high school grade.

Clustering reflected different abilities and motivation of pupils. The best results were achieved by clustering into five groups on the basis of attributes known at the end of first high school grade, because some differences manifest only on the higher and more demanding level of education. The five clusters differ mostly in subject grades and general achievement grades. It is interesting that "worst" pupils in primary school often chose a less demanding educational program, indicating they had at least partially considered advice of educational counselors. But pupils from large cities seem to have higher educational aspirations and despite the same achievement tend to choose more demanding programs, resulting in the lowest achievements of the first high school grade. When abilities are low, high aspirations are not enough for success.

3.5 Machine Learning of Multi-Attribute Hierarchical Models with HINT

Finally, we used HINT to construct multi-attribute models from data. We used its implementation within the DM suite Orange [12], where HINT is limited to discrete data without missing values. Thus, continuous attributes were first discretized and missing values were replaced with majority values. Models were built using 16 and 30 expert-selected attributes, using unsupervised minimal-error decomposition with the default bound set size of two. Again, the classification accuracy was estimated by 10-fold cross validation.

The classification accuracy of HINT models was quite low and unsatisfactory – most often it was close to the apriori accuracy. Such results are partly due to a relatively small learning sample ($N = 889$) and complexity of data.

Although the models were not appropriate for direct use, they provided a number of important guidelines for composing attributes and defining decision rules in the forthcoming manually development of DEX models. For example, among the attributes known before the enrolment, the marks of Math and Physics turned out to be very good predictors. The analysis also revealed the so-called "rule of chain": the marks of Math and Physics mostly limit the highest general achievement. Even more obviously than decision trees, HINT models revealed interesting patterns of absence: pupils are not absent coincidentally or because of health reasons, but systematically and related to exams taking place in school.

4 Development of Decision Support Models

The final multi-attribute decision models were developed manually using DEX and following the three typical DS development stages [4]: (1) acquisition of attributes, (2) development of the hierarchy of attributes, and (3) defining decision rules. The expert tried to incorporate her previous knowledge about academic achievement prediction and supplement it by the results of the DM stage.

First, a list of attributes was created on the basis of expert opinion; the best decision trees and best models developed by HINT. Beside demographical variables, which represent extenuating circumstances or difficulties in reaching a high achievement, the list mostly included different school marks, hours of absence and discipline sanctions, which reflect pupil's knowledge, abilities, motivation and other personality traits. Continuous attributes were discretized in the same way as with HINT.

Next, attributes were organized into a hierarchical structure according to their dependence and interaction, introducing new aggregate attributes (internal nodes) whenever necessary. This turned out to be the most difficult task, since each new attribute is a result of complex interaction of multiple basic attributes. Sometimes it is difficult to accurately estimate the contribution of each single factor, because it is changeable and may depend on an individual pupil.

Eventually, two attribute hierarchies were developed in this way (Figure 3). The first one uses only attributes that are known before the enrolment in high school. The second hierarchy contains an extended set of attributes for the prediction at the end of first high school grade. Both structures are meant to be general in the sense that they do not address any particular educational program.

In the final stage, the expert defined decision rules, i.e., rules that determine the aggregation of values from the leaves towards the root of the hierarchy. The expert judged the importance of attributes and determined their weights. The results of data mining were taken into account; for example, due to their important influence to the final achievement, relatively high weights have been given to school-marks of Math and Physics.

Finally, we fine-tuned the models to the requirements of a single educational program "L". Decision rules were appropriately modified; all rules in decision tables were reviewed in detail and changed, if necessary. We considered the accuracy of used predictors and the "rule of chain" by which one single weak link or sub-criteria is enough for failure. The structure of model 2 was simplified by excluding 8 less relevant attributes.

Table 2 shows an example of decision rules for model 1 and program "L". The rules predict pupils' final academic achievement on the basis of their abilities and previous knowledge, motivation, and circumstances. Only rules for the achievement categories 1 and 2 are shown in Table 2. Rule 1, for example, states that if pupils' abilities and previous knowledge are inappropriate and their motivation is lowered, regardless on circumstances, they will achieve the category 1. According to rule 4, the category 2 will be achieved if their abilities are at least appropriate, motivation is lowered and circumstances are negative.

Table 2: An example of decision rules for model 1(program "L").

	Abilities and p. knowledge	**Motivation**	**Circumstances**	**Fin. achievement**
1	Inappropriate	Lowered	*	**1**
2	Inappropriate	*	<= Appropriate	**1**
3	Inappropriate	Appropriate	Positive	**2**
4	>= Appropriate	Lowered	Negative	**2**

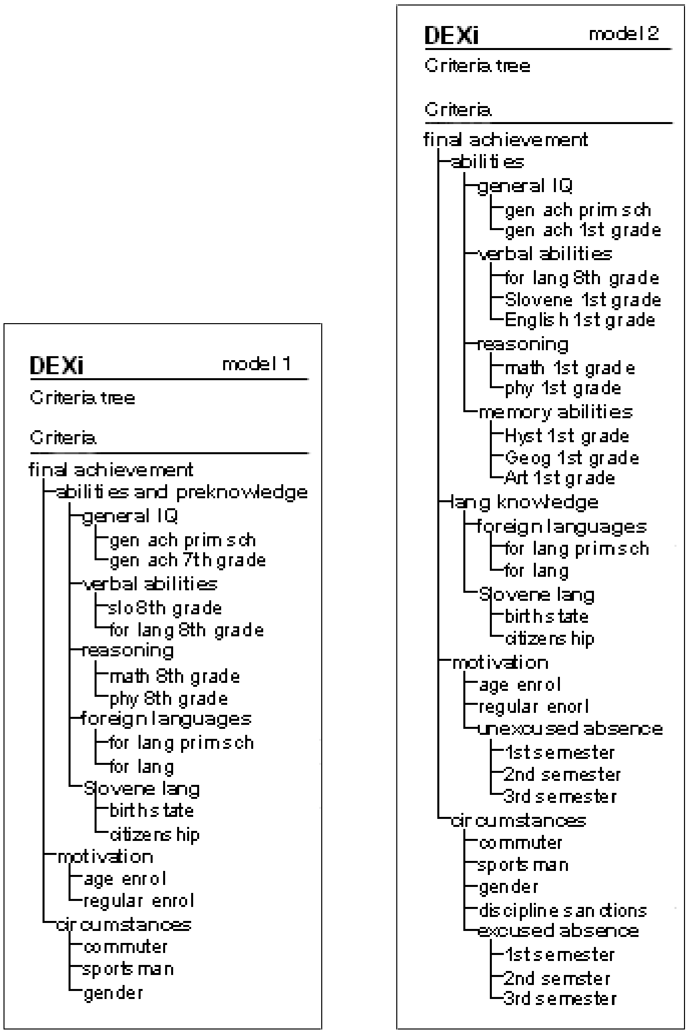

Fig. 3. Structure of the two multi-attribute models

5 Evaluation of Models

A 10% stratified sample (N = 47) was extracted from the database DB1 and used for the evaluation of developed models. We compared the classification accuracy of: (1) both DEX models, (2) decision trees developed in the DM stage, and (3) the classification of the 47 cases, which was provided by an expert counselor.

All the three groups of models achieved almost the same classification accuracy: about 60% using data known at the time of enrolment in high school, and about 70% after one year of study. The models almost never predict failure to really successful pupils, while pupils with bad prediction usually really fail. Occasionally (i.e., in 12.8% and 8.5% of cases, respectively) the models wrongly predict success to unsuccessful pupils. They recognize typically successful and typically unsuccessful pupils relatively well. Wrong predictions, which are more than one category apart from real achievements, are relatively rare, occurring in 12.8% and 6% of cases, respectively.

In particular, the models perform best in recognizing pupils from the achievement category 4, and worst in recognizing pupils from the category 3. They often intermix cases from the categories 1 and 3; this distinction is indeed difficult because both categories correspond to pupils who educate a long time and progress slowly. But almost until the end we cannot say if they will successfully pass the program (category 3) or fail (category 1).

In summary, with regard to a relatively small number and low quality of used attributes, which are not "pure" neither accurate measures of pupil's features, the predictive accuracy of the models is surprisingly high. We must notice that experts in practice also take into consideration other data, obtained in personal meetings with pupils and their parents, what usually increases the accuracy of their predictions. Our expert counselor was very experienced, but his predictive accuracy does not tell much about average prediction accuracy of school counselors, which differ a lot in experience, abilities, knowledge, intuition, and motivation. As highly experienced and capable experts are rare, the use of predictive models for educational counseling seems meaningful. It may also provide a step toward more consistent, objective and systematic estimation, evidence and evaluation of predictions.

For further work, it is interesting to compare DM and DS models developed in this study and notice an important difference in the treatment of preference-ordered attributes: while DS models do take into account the ordering information, DM ones do not. For example, the achievement categories themselves are preference ordered, because 5 (graduates as A or B) is better than 4 (graduates as C or D), which is better than 3, etc. The rule acquisition mechanism of DEX ensures their consistency, so that if object x is better than or equal to y on all considered preferentially-ordered attributes, then x is not assigned a worse class than y. Neither decision trees nor HINT can ensure this kind of consistency. In further integration of DM and DS, it is thus important to include DM methods that can deal with preference-ordered attributes, for example methods based on rough sets [17].

6 Conclusion

Using a combination of data mining and decision support techniques, we developed and evaluated multi-attribute models for the prediction of final academic achievement in high schools. Their predictive accuracy is close–although in practice probably lower than–the accuracy of an experienced human expert, but may be better than the accuracy of an inexperienced expert. Thus, we believe that at this stage the models are appropriate for experimental use by school counselors, but not yet by pupils and their parents. We also believe that their application would increase the accuracy of predictions and quality of educational counseling, helping to prevent inappropriate choices of school and improve academic achievements.

Both DEX models have a number of strong points. They work with data that are practically always available or easy to get. They facilitate a consistent and systematical prediction and evaluation of estimates. Re-estimation at the end of first high school grade increases the reliability of prediction. The use of models would probably improve pupils' and parents' trust in estimates and consequently contribute to more serious selection of schools and educational programs.

On the weak side, the models occasionally wrongly predict good achievement to pupils who will really fail. By that they encourage them to persist in school that is too demanding for them. Also, the models have not been validated on data other than that used in the analysis, so their general prediction accuracy in practice is unknown, but it is probably lower than established. The predictive accuracy of models can easily degrade due to changes of school system and generational changes of pupils. Thus, the models should be adapted and validated perpetually.

Methodologically, the models were developed by a combination of data mining and decision support techniques. From the viewpoint of DS modeling, which is usually carried out without extensive data analysis and relies mainly on expert knowledge, the preceding DM stage brought considerable benefits. Although difficult to quantify, it is clear that the DM stage helped the expert to better understand the characteristics of the population and to discover some important rules and patterns that affect academic achievement. In particular, important contributions of DM to DS were the following:

- Machine learning of decision trees (section 3.3) provided evidence of what was achievable with the available attributes and data, and established the target classification accuracies to be achieved by DS models under various conditions.
- At the intersection of DM and DS, we used HINT to develop parts of DS models from data (section 3.5). Although HINT's models themselves exhibited poor classification accuracy, some developed subtrees turned out extremely useful and revealed important patterns, such as the "rule of chain". Some concepts discovered by HINT were used almost unchanged in the final model. Thus, HINT has been particularly useful as a knowledge exploration and feature discovery tool.
- After the DS models have been developed, the available DM database facilitated a relatively easy evaluation of the quality of DS models (section 5), which is usually very difficult and thus too often omitted from DS modeling.

From the DM viewpoint, the benefits of combining it with DS in this project are somewhat less clear. Using only DM, we would still obtain a number of models (decision trees) of sufficient classification accuracy. Bringing in the expert and developing a DS model clearly did not improve the accuracy in this case. What we did obtain using DS, however, is a general hierarchy of attributes (Figure 3) that can be easily adapted to various educational programs and used for typical DS tasks, such as what-if and sensitivity analysis, and generation of alternatives. Probably its greatest advantage is that it can be relatively easily extended further using DS tools such as DEX. The extension can introduce new attributes that have not been available in the current database, but are achievable by counselors in schools and may potentially contribute to the prediction of academic achievement, for example measures of different intellectual abilities, interests, motivation and personality traits of pupil. Such an extension of our models remains a challenge for the future.

Acknowledgment

The work reported here was in part supported by the Slovenian Ministry of Education, Science and Sport, and by the EU project SolEuNet, IST-11495.

References

1. Mallach, E.G.: Understanding Decision Support Systems and Expert Systems. Irwin, Burr Ridge (1994)
2. Han, J., Kamber, M.: Data Mining: Concepts and Techniques. Morgan Kaufman, San Francisco (2001)
3. Bohanec, M., Zupan, B.: Integrating decision support and data mining by hierarchical multi-attribute decision models. ECML/PKDD-2001 Workshop Integrating Aspects of Data Mining, Decision Support and Meta-Learning (IDDM-2001) (eds. Giraud-Carrier, C., Lavrač, N., Moyle, S., Kavšek, B.), Freiburg, (2001) 25–36
4. Bohanec, M., Rajkovič, V.: DEX: An expert system shell for decision support. Sistemica 1(1) (1990) 145–157
5. Zupan, B., Bohanec, M., Demšar, J., Bratko, I.: Learning by discovering concept hierarchies. Artificial Intelligence 109 (1999) 211–242
6. Hughes, M., Wikeley, F., Nash, T.: Parents and their children's schools. Oxford, Cambridge, Blackwell (1994)
7. Resman, M., Bečaj, J., Bezić, T., Čačinovič-Vogrinčič, G., Musek, J.: Svetovalno delo v vrtcih, osnovnih in srednjih šolah (Educational counseling in nursery, primary and high schools). The National Education Institute of Slovenia, Ljubljana (1999)
8. Dryden, G., Vos, J.: Revolucija učenja (Revolution of the learning). Educy, Ljubljana (2001)
9. Hayes, N., Orell, S.: Psychology: an Introduction. Longman, Harlow (1993)
10. Einstein, G., Abernethy, K.: SPSS Tutorial: Statistical Package for the Social Science: SPSS Version 10.0. Furman University. http://s9000.furman.edu/mellonj/spss1.htm (2000)
11. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann, San Francisco (2000)
12. Zupan, B., Demšar, J.: Orange. http://magix.fri.uni-lj.si/orange/ (2002)
13. Galle, R.: Priročnik za uporabo programa: vodenje evidence učencev (Evidenca 3) (User's manual for the program Evidenca 3). High School of Electrotechnical and Computer Science, Ljubljana (1996)
14. Vieira, R.: Professional SQL Server 2000 Programming. Wrox Press, Birmingham (2000)
15. Quinlan, R.J.: C4.5: Programs for Machine Learning. Morgan Kaufman, San Francisco (1993)
16. Kaufman, L., Rousseeuw, P.J.: Finding groups in data: an introduction to cluster analysis. New York: J. Wiley & Sons (1990).
17. Greco, S., Matarazzo, B., Slowinski, R.: Rough sets methodology for sorting problems in presence of multiple attributes and criteria. European Journal of Operational Research 138(2) (2002) 247–259.