A Comparative Analysis of Techniques for Predicting Student Performance

Hana Bydžovská
CSU and KD Lab Faculty of Informatics
Masaryk University, Brno
bydzovska@fi.muni.cz

ABSTRACT

The problem of student final grade prediction in a particular course has recently been addressed using data mining techniques. In this paper, we present two different approaches solving this task. Both approaches are validated on 138 courses which were offered to students of the Faculty of Informatics of Masaryk University between the years of 2010 and 2013. The first approach is based on classification and regression algorithms that search for patterns in study-related data and also data about students' social behavior. We prove that students' social behavior characteristics improve prediction for a quarter of courses. The second approach is based on collaborative filtering techniques. We predict the final grades based on previous achievements of similar students. The results show that both approaches reached similar average results and can be beneficially utilized for student final grade prediction. The first approach reaches significantly better results for courses with a small number of students. In contrary, the second approach achieves significantly better results for mathematical courses. We also identified groups of courses for which we are not able to predict the grades reliably. Finally, we are able to correctly identify half of all failures (that constitute less than a quarter of all grades) and predict the final grades only with the error of one degree in the grade scale.

Keywords

Student performance prediction, student similarity, classification, regression, collaborative filtering.

1. INTRODUCTION

One of the key problems of educational data mining is to design student models that would predict the student performance. Once we have a reliable performance prediction, it can be used in many contexts: for identifying weak students [14], for guiding the adaptive behavior in intelligent tutoring systems [10], or for providing a feedback to students.

Our specific problem is the following: we have access to data about students, their study achievements and their behavior characteristics stored in the university information system and we want to predict students' final grades. The predictions are useful at the beginning of each semester to help students with planning their workload in the whole semester. We also beneficially use this information to design a course enrollment recommender system. The early grade prediction is more difficult since we have no a priori information about students' knowledge, skills or enthusiasm for particular courses. It has been proven [4] that the data about the activity of students during the semester improves the prediction.

The problem of the student grade prediction in a particular course has recently been addressed using data mining techniques. Researchers usually examine study-related records, e.g. the age, the gender, and the field of study [9] because of their easy availability in university information systems. Moreover, they attempt to identify additional characteristics that can lead to better understanding of students' behavior, e.g. their habits [6] or parents' education [13]. The most typical way how to obtain such data is to conduct questionnaires. Masaryk University has more than 40,000 active students and we try to predict the grades as accurately as possible for all of them. We cannot rely on data obtained by questionnaires since they tend to have a lower response rate. Therefore, only the data originated from the Information System of Masaryk University (IS MU) are employed for our experiments.

The goal of this research is to predict students' grades with the major emphasis on the detection of students who can fail to meet the course requirements. Therefore, we are dealing with the following two main tasks:

- prediction of students' success or failure,
- prediction of the students' final grades.

In this paper, we present two different approaches moving towards our objectives. The first approach is based on the state of the art educational data mining techniques: classification and regression analysis [12]. We created an ensemble learner to utilize the strength of the both techniques. We also present a new type of data about students' social behavior originated from IS MU that can improve the predictions. The second approach is based on collaborative filtering techniques [5] applied to the educational context. We mapped the users-item-rating problem to the student-course-grade problem and predict the final grades based on previous achievements of similar students. This paper describes both approaches in detail, compares them and reports their advantages and disadvantages.

2. DESIGNED METHODS EVALUATION

Historical data were employed for experiments allowing us to evaluate both designed approaches. We processed data about 138 courses which were offered to the students at the Faculty of Informatics. We used only data stored in IS MU in the time of students' enrollments. We omitted freshmen students because we had no data about them in the system. The data comprised of 3,584 students. The two independent data sets were used. The training set consisted of the data collected between the years of 2010 and 2012 (37,005 instances) and was used for the identification of the most suitable methods with their settings. The test set consisted of the data from the year 2013 (11,026 instances) and was used for the validation of the methods on different data.

The following grade scale was used: 1 (excellent), 1.5 (very good), 2 (good), 2.5 (satisfactory), 3 (sufficient), 4 (failed or waived). The value 4 represents student's failure; the others represent a full completion. We evaluated approaches using the *mean absolute error* (MAE). The technique measures how close predictions are to the real outcomes. Lower values represent better results. The measure is commonly used for grade prediction evaluation. In the educational environment, one of the most important issues is to reveal weak students. Therefore, we also computed the *sensitivity* (also called recall). Categorizing students only as successful or unsuccessful, the sensitivity measures the proportion of unsuccessful students who are correctly classified as unsuccessful. For students' success or failure prediction we also utilized *F1 score* that conveys the balance between the precision and the recall.

3. STUDENTS' CHARACTERISTICS

3.1 Study-related Data

Classification and regression are the most often used techniques for student performance prediction [12]. Researchers usually examined study-related (SR) data. Our study-related data contained common attributes such as the gender, the year of birth, the year of admission, the number of credits gained from passed courses, or the average grades. We built a classifier for each investigated course based on the training set and evaluated the results using the 10-fold cross validation. The method that achieved best results was subsequently validated on the test set.

3.1.1 *Student success/failure prediction*

The first task was to reveal unsuccessful students. Two prediction classes were considered: students' success (def. 1: grades 1−3) and failure (def. 2: grade 4). Widely utilized classification algorithms were employed: Support Vector Machines (SVM), Random Forests, Rule-based classifier (OneR), Trees (J48), Part, IB1, and Naive Bayes (NB). As the baseline we defined a model which always predicts failure. Table 1 confirms that SVM achieved the best performance.

Table 1. Classification algorithms results

Rank	Method	F1	MAE	Sensitivity
1	SVM	0.559	0.161	0.444
2	NB	0.554	0.251	0.467
3	J48	0.552	0.182	0.397
4	Random Forests	0.550	0.173	0.362
5	Part	0.543	0.202	0.417
6	IB1	0.536	0.216	0.436
7	OneR	0.508	0.183	0.321
8	Baseline	0.326	0.822	1

3.1.2 *Grade prediction*

The regression is a commonly used technique for student grade prediction. Widely utilized regression algorithms were selected: SVM Reg., Random Forest, IBk, RepTree, Linear Regression, and Additive Regression. The baseline model predicts the average grade of the training set of a given course. The best results (see Table 2) were achieved by support vector machine (SVM Reg.).

3.1.3 *Conclusion*

For each task, the best method was selected and an ensemble learner was built. If the classifiers (SVM or SVM Reg.) predicted the failure or the grade 4, then the ensemble learner also predicted the failure. Otherwise, it resulted in the value of the grade predicted by the SVM Reg. classifier. Finally, the overall performance of this approach could be seen in Table 3. We also evaluated the classifiers on the test set. The results indicated that we were able to reveal almost half of the unsuccessful students even if the task was difficult due to the fact that all unsuccessful students constitute less than a quarter of all students. The prediction error was about 0.75 on average which was almost 1.5 degree in the grade scale.

Table 2. Regression algorithms results

Rank	Method	MAE	Sensitivity
1	SVM Reg.	0.605	0.196
2	Linear Reg.	0.615	0.152
3	Additive Reg.	0.634	0.165
4	RepTree	0.643	0.184
5	Random Forests	0.668	0.216
6	IBk	0.767	0.294
7	Baseline	0.806	0

Table 3. Global SVM results

Data Set	MAE	Sensitivity
Training Set	0.701	0.524
Test Set	0.744	0.414

3.2 Social Behavior Data

Recent researches are often based on finding additional data that can improve the prediction accuracy. Our improvements have been achieved through adding social behavior (SB) data to the original data set [1]. This specific type of data originating from IS MU described the students' behavior characteristics and their mutual cooperation. We focused on statistical data that represented an interaction among students: posts and comments in discussion forums, e-mails statistics, publication co-authoring, or files sharing. This information served as the basis for computing social ties among students and building a sociogram. From this sociogram, new features like weighted average grades of friends can be easily derived. Using Pajek [11], we also computed additional standard graph features [3] like degree (the number of the friends), weighted degree (degree weighted by the strength of ties), centrality or betweenness (the importance measure for each student in the network). Moreover, we collected data about students' disclosure from different system sections. By default, IS MU does not provide a complete list of classmates due to the students' privacy. Students have to actively disclose themselves to become visible for their classmates. We can also calculate how many times students attended courses of a certain teacher. Among others, students can also mark offered courses as favorite.

H1: Hypothesis supposes that students' social ties correlated with the students performance.

Other ensemble learners trained on data sets containing social attributes were built. The other settings were maintained. The comparison of the results can be seen in Table 4. The MAE score was slightly lower on average. However, for 32 courses in the test set, the difference in MAE was significantly better using social behavior data (min: 0.1; average: 0.178; max: 0.734). Only 5 courses achieved worse results (min. 0.1; average: 0.12; max: 0.21). For the rest courses, the difference was negligible.

Table 4. Adding social behavior attributes to the data set

Data Set	Attributes	MAE	Sensitivity
Training Set	SR	0.701	0.524
	SR + SB	0.629	0.528
Test Set	SR	0.744	0.414
	SR + SB	0.688	0.427

The sorted list of selected attributes was constructed. In Table 5, we present the top five social behavior attributes that significantly affected the results.

Table 5. The most interesting social behavior attributes

Rank	Avg. Ord.	Attribute
1	13.328	the betweenness
2	16.252	the information if the course was marked as favorite
3	18.694	the centrality
4	22.464	the weighted degree
5	29.807	the number of times when a student attended any course with the same teacher

H1 was confirmed. Data about students' behavior improved the predictions. Based on the most significant attributes, we assumed that the assistance of students' friends had increased the probability to pass the courses.

4. STUDENTS' GRADES

We also focused on methods utilized in recommender systems [5]. The data about user-item-rating triples were replaced by student-course-grade triples and we focused on the similarities among students' grades.

H2: Our hypothesis supposed that students' knowledge can be characterized by the grades of courses that students enrolled during their studies. Based on this information we could select students with similar interests and knowledge and subsequently predict whether a particular student has sufficient skills needed for a particular course.

4.1 Grade Prediction

Our preliminary work can be found in [2]. However, the approach suffered from several limitations that we overcome in this paper.

The first step was to build a similarity matrix *G* where rows represented students and columns represented courses. Although we predicted grades for 138 courses, the matrix *G* has 499 columns since we analyzed all students' grades (e.g. courses from the other faculties, courses not offered now). Grades obtained by all students from the training set formed the matrix. If a student did not attend a particular course, the corresponding cell remained empty. The aim was to complete cells defining students' grades from the investigated courses enrolled by students in 2012 (marked by symbol *?*).

Using the vectors of grades from the matrix *G*, we computed the similarity between all students enrolled in a course *c* in 2012 and all students previously also enrolled in *c* in 2010 or 2011.

Example of Matrix G

Students / Courses	c_1	c_2	c_3	c_4
s_1	2	?		?
s_2	?	2.5	3	?
s_3	1		2.5	3
s_4		2		1.5

Widely utilized similarity metrics were used for the calculation of the students' similarity: Mean absolute difference (MAD), Root mean squared difference (RMSD), Cosine similarity (COS), and Pearson's correlation coefficient (PC). All metrics compare grades of students' shared courses. The average number of courses shared by students was 10.

Subsequently, the appropriate neighborhood of the most similar students to the examined student could be selected to influence the predicted final grade. We utilize the idea of a baseline user [7]. We selected such students to the neighborhood who were more similar to the investigated student than the investigated student was to the baseline student. We decided to calculate two types of baseline students: an average student (the average grade for each course) and a uniform student (the average grade through all courses: 2.5). The neighborhood of the top 25 students showed reasonable results. However, for smaller courses, 25 students could be all students enrolled in the course in one year. Therefore, we have decided to define three categories of courses with respect to the course occupancy: small (≤30 students), medium (30−70 students), and large (≥70 students). Therefore, we analyzed the suitable size of the neighborhood for courses with the different occupancy. Figure 1 shows the relationship between MAE and the cardinality of *N*. We selected the size of neighborhood as follows: 10 for small courses, 15 for medium courses, and 30 for large courses. In the figure, we can also see that the prediction for smaller courses was the most challenging.

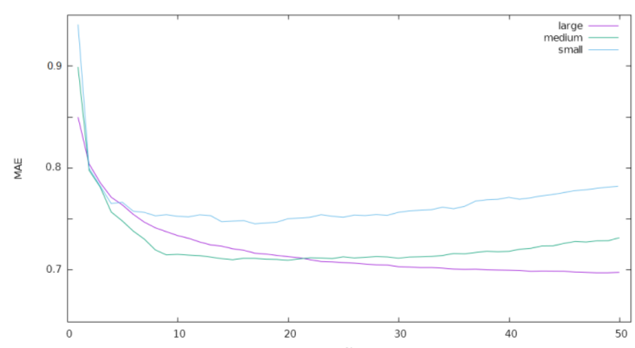

Figure 1. Relationship between MAE and the size of neighborhood with respect to the course occupancy

The final grades were estimated from the grades of similar students belonging to the computed neighborhood. Simple methods as mean, max, median as well as advanced methods utilizing significance weighting were utilized.

Table 6 introduces the top five combinations of the similarity methods, methods for the neighborhood selection and the grade estimation functions. The method utilizing a baseline user needed a large neighborhood for each student ($|N|$ = 376 on average). In the production system, it was very important to lower the ties among students due to the recalculation of all similarities in the system during the course enrollment process to be up to date for students. Therefore, different neighborhood was selected even if the MAE score could be slightly higher. For efficiency reasons, we selected the third one for the implementation in the system.

Table 6. Similarity methods comparison

Rank	Method	|N|	MAE	Sensitivity
1	PC + average student + sig. weighting	376	0.648	0.248
2	PC + uniform student + sig. weighting	378	0.648	0.248
3	PC + Top |N| + sig. weighting	10/15/30	0.650	0.267
4	RMSD + Top |N| + median	10/15/30	0.651	0.211
5	PC + Top 25 + Pred	25	0.657	0.274

4.2 Student Success/Failure Prediction

The majority of students passed examined courses. Therefore, we searched for a smaller neighborhood in order to reveal more unsuccessful students. As you can see in Figure 2, the highest F1 was reached when we included only the most similar student. However, the method suffered by a low precision. Therefore, we predicted failure even if the method for grade prediction (3rd row Table 6) predicted grade worse than 2.4 (average grade). The precision was improved and still we found the sufficient number of unsuccessful students. The final results of methods were: MAE = 0.174, sensitivity = 0.413.

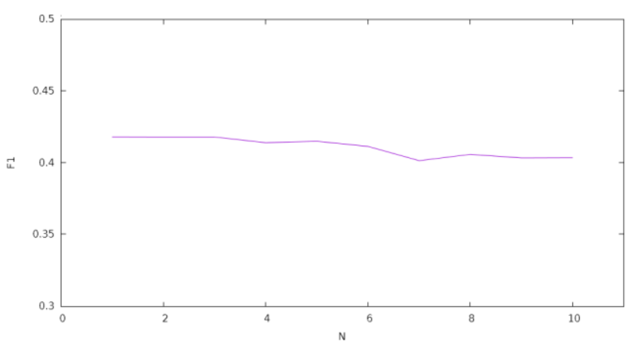

Figure 2. Relationship between F1 and the size of the neighborhood

4.3 Course similarity

Any change in the similarity matrix *G* could lead to the recalculation since the similarity of students was calculated from all students' grades.

H3: Our third hypothesis supposed that similar courses required similar skills of students to pass. It should decrease the computational cost and do not significantly lower the prediction accuracy when we use only grades of similar courses for predictions instead of all attended courses.

4.3.1 Students' grades

The collaborative filtering approach based on similarity of item to item was utilized and the *adjusted cosine similarity* was computed from the previously defined similarity matrix *G* for each pair of courses. Subsequently, we utilized the average link clustering [8] to group the investigated courses based on this similarity measure. The resulted clusters defined the groups of similar courses.

Finally, when we predicted the students' grades of a certain course, we reduced the computations to the grades obtained from courses belonging to the same cluster as the investigated course. 110 of all investigated courses belonged to one of the 37 clusters. The number of courses in one cluster ranged from 2 to 15. The average number of courses in one cluster was 3. The average number of students' shared courses was also 3.

4.3.2 Course Characteristics

Students search for useful information about courses in the Course Catalog that help them to decide whether or not they should enroll the course. We selected different course characteristics and attempted to identify dependencies among courses. Similarity of courses *a* and *b* was defined by the weighted sum of the similarities of the selected course characteristics $t \in T$:

$$sim(a, b) = \sum_{t \in T} w_t \ \mathrm{dist}(a_t, b_t)$$

where *w* defined the weight of the examined characteristic. The weights of the characteristics were set with respect to maximize the grade prediction accuracy. The similarity for each pair of courses was calculated. The selected characteristics and distance metrics *dist* were the following:

Prerequisites define a set of courses that had to be passed before students could enroll a certain course. The similarity was set to the value of 1 if the compared course belonged to the prerequisites; 0 otherwise. The weight of this characteristic was set to 1 because the prerequisites denoted a significant dependence.

Literature contains the recommended literature for particular courses that can be characterized by the set of assigned authors. The similarity of the set of authors *A* and the set of authors *B* is given by Jaccard's coefficient. The characteristics weight was set to the value of 0.9 due to the hypothesis that authors do not frequently publish in different fields. Therefore, the literature could constitute strong ties among courses.

The *course content* was represented by the text about the study subject and outline what students should learn in the course. We cut the STOP words from the text and utilized stemming to get the roots of the words. TF-IDF was utilized for defining the importance of each word in the texts. Subsequently, the Cosine similarity measure was used for the processing of the final vector representation of the words' importance. The characteristics weight was set to the value of 0.7.

Teachers of a course could be divided into two groups: lecturers and tutors. Weighted Jaccard's coefficient was used for comparing the teachers of the two courses. The weight of the lecturers was set to the value of 1 and 0.5 for seminar tutors. The weight of characteristic was set to the value of 0.6.

Course supervisor patronize the courses. The similarity was set to the value of 1 if the compared courses had the same supervisor; 0 otherwise. The characteristics weight was set to the value of 0.4.

When we calculated the similarity of courses by the aforementioned procedure, we could also utilize average link clustering [8]. 340 from all courses (499) belong to one of the 105 created clusters. 93 investigated courses were presented in one of the clusters. The number of courses in one cluster ranged from 2 to 22. The average number of courses in a cluster was 3. The average number of shared courses taken by students was 2.

4.3.3 Comparison of approaches

In comparison with the method using *all grades*, both approaches had positive effects on the number of calculations. 123 courses (from all 138) belonged to some of the created clusters and the final grades could be predicted based on the grades of only 3 other courses on average. 70 of our investigated courses belonged to different clusters using SC_1 and SC_2. A slightly better MAE was obtained by the method utilizing the course characteristics for these courses (see Table 7). Therefore, when a grade is predicted, the corresponding course is searched in SC_2, then SC_1.

Table 7. Comparison of SC_1 and SC_2

Method	MAE	Sensitivity	Average cluster size	Shared Courses
All grades	0.687	0.402	499	10
SC_1	0.681	0.390	3	3
SC_2	0.640	0.386	3	2

4.4 Conclusion

H2 and H3 were confirmed. We described the novel approach for predicting the students performance (see Table 8) using only students' grades and course characteristics. It proved to be as successful as the first described approach (see Table 9). The most important contribution of this approach was that each university information system stores the data about students' grades which were needed for the prediction unlike the data about students' social behavior. We also identified course dependencies that lowered the calculation cost. Moreover, we were able to predict the final grade considering grades from only 3 other courses for the most of the investigated courses.

Table 8. Global results of the approach

Data Set	MAE	Sensitivity
Training Set	0.661	0.470
Test Set	0.685	0.418

5. USAGE OF THE APPROACHES

Both approaches defined in Section 3 (based on students' attributes (SBA)) and Section 4 (based on students' grades (SBG)) reached similar average results (see Table 9). However, they can differ in specific situations. Our goal was to identify course groups for which we could get trustworthy predictions and also to detect when one approach outperforms the other.

Table 9. Comparison of the both approaches

Data Set	Approach	MAE	Sensitivity
Training Set	SBA	0.629	0.528
	SBG	0.661	0.470
Test Set	SBA	0.688	0.427
	SBG	0.685	0.418

H4. Each approach is more suitable for different course groups.

We selected the following categories based on the basic course characteristics:

- difficulty – the average grade of all students' grades is 2.4. Therefore, we divided courses into two categories: easy (≤2.4), and difficult (>2.4),
- occupancy rate – as defined in Section 4.1: small (≤30), medium (30 − 70), and large (≥70),
- specialization – courses divided into four groups: mathematics (M), theoretic informatics (I), applied informatics (P), and others (O).

Each investigated course belonged to one of the groups for each of the defined categories. With respect to the three aforementioned categories, we could define six (3!) tree structures which differ in the splitting order of the categories. We examined each permutation of the categories. We built full trees where courses from the training set were split subsequently by all categories. Each node stored the information about courses that belonged to it with respect to the split. Harmonic mean (HM) was calculated for each node and both approaches in order to get a suitable relationship between the sensitivity and the MAE score.

Subsequently, we examined the trees and merged branches which were not interesting in order to detect significant phenomena. Interesting branches contained at least one of the following situations:

- Difference > 0.1 in HM of SBA and SBG in the node (The rule detected a significant difference in the prediction accuracy of the both approaches for the examined groups of courses.).
- Difference > 0.1 in HM of the sibling nodes (The rule detected course groups that were significantly easily or with difficulties predicted than other courses from this split.).
- Difference > 0.1 in HM of parent and child nodes (The rule detected the course groups that should be separated due to the significant difference in the prediction in comparison with the rest courses from the parent node.).

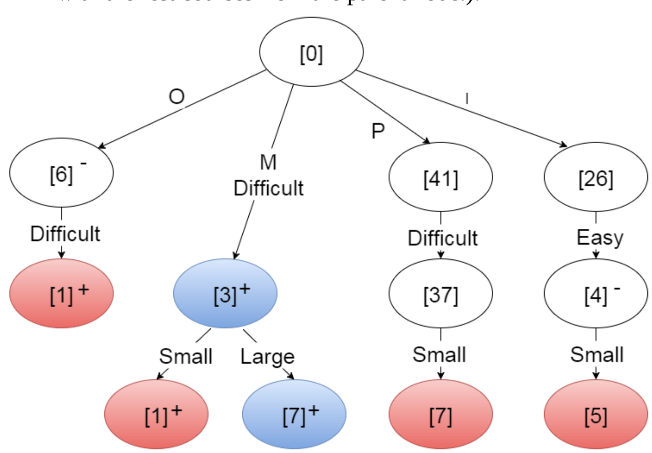

Figure 3. Resulted Tree

One of the resulted trees can be seen in Figure 3. As the figure shows, this approach had several benefits:

- Course groups that were predicted significantly better than average were identified (marked by +). It contains all mathematical courses (the main skill at the faculty of informatics can be easily predicted) and the English course.
- Course groups that were predicted significantly worse than average were identified (marked by −). It contained almost all courses belonged to the category *others* (we do not know students' general knowledge) and medium or large easy theoretic informatics courses (the grade maybe depended on the amount of the effort which could differ for each course and cannot be predicted).

- H4 was confirmed. Course groups that were predicted significantly better by the SBG approach are represented by the blue color. It covered almost all mathematics courses (except one small course). Otherwise, red nodes present better results obtained by the SGA approach. It contained the most of small courses. For the white nodes, the difference in prediction accuracy was negligible.
- Outliers were also identified. One course of the group showed different behavior than others: the course of English (path: O-difficult) was easily predictable in comparison with all courses belonged to the category *others*; one small mathematical course (M-difficult-small) differed in the approach that achieved better results in comparison with all other mathematical courses.

We applied this knowledge for prediction of the students' performance when the test set was utilized. We can easily locate any particular course in the tree and used the suitable approach that led to the better results. We also gave no predictions for courses that we were not able to predict reliably. As the results in Table 10 show, MAE was significantly improved in comparison with the state of the art method utilizing only SVM. Finally, we were able to predict the final grades with an error of one degree in the grade scale. We were also able to reveal almost a half of the unsuccessful students.

Table 10. Final results validated on the Test set

Approach	MAE	Sensitivity	Omitted Courses
Novel	0.609	0.436	10
SVM	0.744	0.414	0

6. CONCLUSION

In this paper, we focused on the problem of predicting final grades of students at the beginning of the semester with the emphasis on identifying unsuccessful students. Two different approaches were presented. Firstly, we utilized widely used classification and regression algorithms. SVM reached the best results. We also proved that data about social behavior of students improve the predictions for a quarter of courses. This approach can be beneficially utilized for the grade prediction of courses with a small number of students.

The second novel approach utilized collaborative filtering techniques and predicted grades based on the similarity of students' achievements. The advantage of this approach was that each university information system stores the data about students' grades which were needed for the prediction unlike the data about students' social behavior. We also succeeded in identifying course dependencies. Finally, we were able to predict the final grades of the investigated course by examining grades of only 3 other courses. The approach can be beneficially used for the grade prediction of mathematical courses.

We also identified groups of courses that are hardly predictable: courses with a different specialization than usual at the Faculty of Informatics, and also large informatics courses which are easy to pass. Finally, we were able to predict the final grade with the error of only one degree in the grade scale for the rest of courses. Half of students' failures were also correctly identified even if the task was difficult due to the fact that all unsuccessful grades constitute less than a quarter of all grades.

7. REFERENCES

[1] Bydžovská, H. and Popelínský, L. 2014. The Influence of Social Data on Student Success Prediction. *Proceedings of the 18th International Database Engineering & Applications Symposium*, pp.374-375.

[2] Bydžovská, H. 2015. Are Collaborative Filtering Methods Suitable for Student Performance Prediction? *Progress in Artificial Intelligence - 17th Portuguese Conference on Artificial Intelligence*, pp. 425-430.

[3] Carrington, P., Scott, J. and Wasserman, S. 2005. Models and methods in social network analysis. Structural analysis in the social sciences. Cambridge University Press.

[4] Koprinska, I., Stretton, J., and Yacef, K. 2015. Students at Risk: Detection and Remediation. *The 8th International Conference on Educational Data Mining (EDM 2015)*, pp. 512 – 515.

[5] Manouselis, N. and Drachsler, H. and Vuorikari, R. and Hummel, H. and Koper, R. 2011. Recommender Systems in Technology Enhanced Learning, Recommender systems Handbook Spriger Verlag 2011, pp 387-415.

[6] Marquez-Vera, C. Romero, C. and S. Ventura. 2011. Predicting school failure using data mining. *In Proceedings of the 4th International Conference on Educational Data Mining (EDM'11)*, pp. 271-276.

[7] Matuszyk, P., and Spiliopoulou, M. 2014. Hoeffding-CF: Neighbourhood-Based Recommendations on Reliably Similar Users. *In User Modeling, Adaptation, and Personalization*, volume 8538 of Lecture Notes in Computer Science, Springer International Publishing.

[8] Murtagh, F. and Contreras, P. 2012. Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(1):86–97.

[9] Nghe, T. N., Janecek, P. and Haddawy, P. 2007. A comparative analysis of techniques for predicting academic performance. *37th ASEE/IEEE Frontiers in Education Conference*, Milwaukee, WI 2007.

[10] Nižnan, J., Pelánek, R., and Řihák, J. 2015. Student Models for Prior Knowledge Estimation. *The 8th International Conference on Educational Data Mining (EDM 2015)*, pp. 109-115.

[11] Nooy, W., Mrvar, A. and Batagelj V. 2011. Exploratory Social Network Analysis with Pajek. Structural Analysis in the Social Sciences. Cambridge University Press.

[12] Strecht, P., Cruz, L., Soares, C., Mendes-Moreira, J., and Abreu, R. 2015. A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance. *The 8th International Conference on Educational Data Mining (EDM 2015)*, pp. 392-395.

[13] Vandamme, J.P., N. Meskens and J.F. Superby, 2007. Predicting academic performance by data mining methods. *Educ. Econ.*, 15, 405-419.

[14] Ventura, S., Romero, C., López, M.-I., and Luna, J.-M. 2013. Predicting students' final performance from participation in on-line discussion forums. *Computers & Education* 68, 458-472.