

# Educational Data Mining for Prediction of Academically Risky Students Depending on their Temperament

Korenkova M.M.<sup>1,2</sup>, Shadrina E.V.<sup>3</sup>, Oshmarina O.E.<sup>4</sup>

<sup>1</sup> National Research University Higher School of Economics, the Department of Literature and Cross-Cultural Communication, Nizhny Novgorod, Russia.

<sup>2</sup> Center of Applied Psychological and Pedagogical Studies, Moscow State University of Psychology & Education, Moscow, Russia.

<sup>3</sup> National Research University Higher School of Economics Nizhny Novgorod, Department of Applied Mathematics and Informatics, Nizhny Novgorod, Russia.

<sup>4</sup> National Research University Higher School of Economics Nizhny Novgorod, Department of Information Systems and Technologies, Nizhny Novgorod, Russia.

eshadrina@hse.ru

**Abstract.** The article discusses the influence of temperament on the academic performance of the first-year students at HSE-Nizhny Novgorod on the example of the Faculty of Informatics, Mathematics and Computer Science (IM&CS). The analyses were done with the help of statistics and educational data mining. The baseline data for the study is information about students, obtained by a survey: the information about temperament, degree of extraversion, stability, and other personality traits of students. The study involved students of the first and second years of the faculty of the IM&CS 2017-2018 academic year. Further, psychological factors affecting the average score and the probability of re-training for students with different temperaments were identified. A certain connection between temperament and academic success, which makes possible the prediction of "risky" students, was found. Various machine learning methods are used: the kNN-method and decision trees. The best results were shown by decision trees. As a result, first-year students are classified into three groups (Good, Medium, Bad) according to the degree of risk of getting academic debt. The practical result of the research was the recommendations to the educational office of the Faculty of IM&CS to pay attention to risky students and assist them in the educational process. After the end of the summer session, the classification results were checked. The article also presents an algorithm for finding risky students, taking temperament into account.

**Keywords:** psychology of higher education, academic achievements, human temperament, students, data mining, decision tree, on-line questionnaire.

## 1. Introduction

Students' academic performance, successful or unsuccessful leading sometimes to expel from the university is a serious issue that should be carefully studied [1, 3, 12, 14, 24]. Students' performance is the main parameter, on the basis of which it is possible to evaluate how well knowledge has been transferred and understood. If there

is an opportunity for early prediction of examination results, preventive measures can be taken: electives and additional course consultations may be arranged in order to lower the number of students with unsatisfactory results or academic failures and dropouts. In small groups inclinations of students can be determined by a teacher during classes due to personal contact. But if the number of students rises, it becomes more difficult to monitor those who will be likely to fail the examination.

The Faculty of Informatics, Mathematics, and Computer Science (HSE Nizhny Novgorod) grows and accepts more and more first-year students with different motivation level, natural aptitude, perseverance, and other personal features. These personal qualities and attitudes can give them an advantage or create obstacles on the way to the academic success [17].

Despite the fact that students who come enter the faculty have a high score on the Unified State Exam (school leaving tests), the percentage of those who are expelled after the first year is very high - about 30%. It is especially peculiar to the highly selective universities, where the percentage of dropouts is higher than in non-selective universities [18].

Since there is high competition among universities for students, higher education institutions have to take into account the needs of the students more than ever nowadays and do their best to meet these needs. IT faculty, for example, is often chosen by young people who tend to be introverts and are hardly outgoing or communicative (нужна ссылка). This fact requires from the teaching staff a special approach to such type of the audience. Universities that take into account the complex of students' needs (everyday, personal, academic) to a greater extent will have an undoubted competitive advantage. It is of equal importance to be able to understand who the successful students are, what they are interested in, and what their intentions and ambitions are. At The Faculty of Informatics, Mathematics, and Computer Science (HSE Nizhny Novgorod) students with a strong background in computer science have the opportunity to engage in sports (Olympiad) programming and successfully perform at international Olympiads. Consequently, some major goals for a university today are: to find efficient ways of attracting students; to hold their interest; to strike a balance between academic requirements and catering for students' needs and interests; how to retain students without reducing the quality of their studies.

Therefore, we are sure that the problem of predicting students' performance on the basis of their personal features becomes increasingly burning in the field of education.

During the last 20 years different authors conducted studies detecting factors that to a great extent affect the academic performance and students' dropouts [2, 9, 23,15]. Sociological theories take into account the student's university and family environment, mechanisms of socialization, the influence of key people [13].

A large-scale study was conducted by Superby with co-authors at a Belgian university detecting the most influential factors [25]. A newer study was conducted by Lust and co-authors at Belgian and French universities with the help of advanced methods of data mining and machine learning [20]. Finally, a similar study was conducted at the Higher School of Economics in 2014 [7]. All these studies show that factors connected to academic behaviour such as attendance, keeping notes, confidence in selection of a university and profession, doing homework, attending electives, as

well as personal history factors including parents' education, presence of both parents in the family, school grades are of the most importance. They also demonstrate that character features are also important, but not so significantly as factors specified. In his turn, Poldin with co-authors in his article explains that sociable students by means of communication develop personally, and therefore become more successful [22]. The work of Valeeva detects a relationship proving that in the course of time social isolation of students with academic failures takes place, creating additional risks of being expelled from higher education institutions for them [10].

In the works mentioned above the effect of friendship ties on student performance is noted, and the authors of the current study noticed that those who can make friends are also capable of academic work. On the other hand, the authors noticed that students who choose the IT field are more prone to isolation and it is difficult for them to open up in a new team, both with classmates and with teachers. Thus, based on personal experience and the experience of colleagues, the idea arose to study more closely the psychological aspects of the personality of students.

Methods used by the researchers are varied from Statistics to Complex Data and Process Mining. Last years the research focus shifts from static data classification and clustering [2, 11] to the control-flow discovery [5, 6]. Process discovery [4, 26], conformance checking [8], social network mining became more and more popular.

In our research we think not only about classification of students, but also about the educational process in which they are involved. We consider the dropout problem as a relevant outcome of the educational process (and use students score data for three modules) and personal characteristics that were obtained from the survey. So the data and process mining techniques are working together for our purposes.

In this work we will base on the typology of four temperaments singled out by Jung [16]:

- The choleric temperament is characterized with intensity and power of emotional processes. Choleric people are quick-tempered, passionate and energetic;
- A sanguine individual is distinguished by comparatively weak intensity of psyche processes with a quick change of certain processes with other. Sanguine people are cheerful, hard-working, they easily cope with various tasks;
- A phlegmatic person is distinguished by slowness, sluggish movements, lack of energy. Feelings of a phlegmatic person are even and quiet. Phlegmatic people are devoted persons and it is difficult for them to switch into new activity types;
- The melancholic temperament is characterized with depth of emotional expressions, but slow flow of psyche processes. Feelings and emotions of a melancholic person are usually uniform, such people are sensitive to external circumstances and often prove themselves to be passive and sluggish.

This work is aimed at studying the influence of personal features (i.e. temperament as a personality basis) on academic performance and detection of students who will highly likely fail their examination.

Our main hypothesis is that there is a relationship between temperament and student's academic performance. The information about personal features of students received by means of questioning via Google online form was used as a material for the study. First- and second-year students of the Faculty of Informatics, Mathematics and Computer Science of academic year 2017-2018 were the study subject.

## 2. Obtaining and processing data

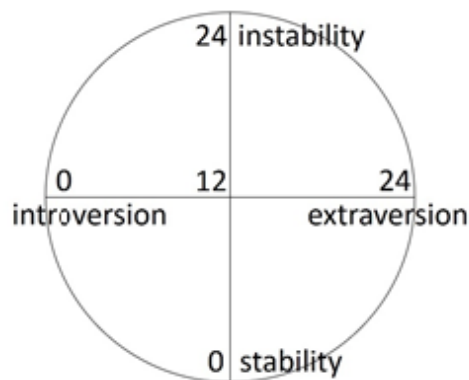
The famous scientist Boris Mirkin draws attention to the importance of classification in data analysis problems. Classification is the construction of a classification that structures the set of phenomena under consideration into a set of separate classes reflecting the important properties of these phenomena. Currently, this term is also applied to the problems of assigning individual objects to predefined classes. [21].

The main task of our research is to classify students into three groups:

- Good students are “not risky” students, they will not have academic debts;
- Medium students are “with a medium likelihood of failing the exam”, they can successfully pass the session, thanks to the support measures adopted at the university;
- Bad is “high risk students” fail the exam, drop out.

Thus, we need to create a dataset in which each student is described according to a set of personality characteristics.

To obtain data on the temperament of students and a number of other character traits, a survey was conducted in the internal system of the HSE-Nizhny Novgorod LMS (Learning Management System). 90 first-year and 50 second-year students of the Faculty of Informatics, Mathematics and Computer Science, NRU HSE-Nizhny Novgorod took part in it.

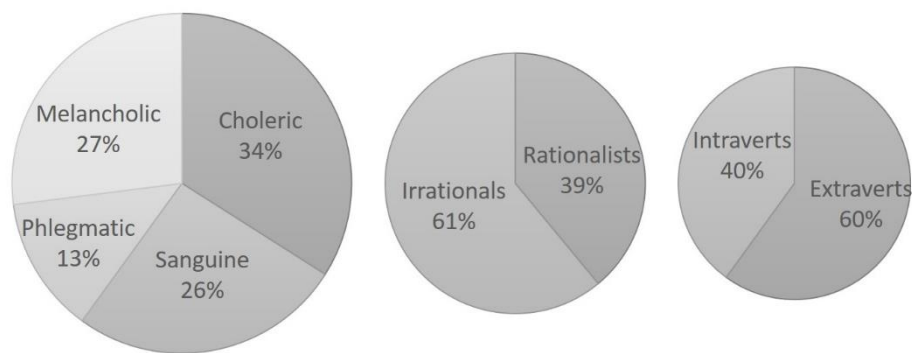


**Figure 1.** Classification of temperaments by extraversion and rationality

The survey was compiled on the basis of a questionnaire test by G. Aysenck [16] to determine the types of temperament of each student in terms of extraversion-introversion and instability-stability. The numbers in Figure 1 indicate the degree of introversion and extraversion from moderate 0-7 to significant 19-24. Questions about school activity, perseverance, the ability to prioritize that are all personal characteristics of the student were also included in the questionnaire. The block for identifying temperament consisted of 12 questions, each of which described one of the temperaments: choleric, sanguine, phlegmatic or melancholic. Using the students' answers, we calculated such parameters as extraversion and rationality (Fig. 1), then

the temperament of each student was determined. In total, students had to answer 17 personal questions in the questionnaire. We took the data on midterm and final control (marks for examinations) after the first half of the academic year from the internal database of the NRU HSE student management system. Next, after combining all the information, we extracted 13 binary variables for each student (variables of the form 0/1). The main variable for making a decision used to validate our model is the average score across all disciplines (GPA) and pass/fail information from the overall student ranking. Based on this variable, the classification into the groups Good, Medium, Bad is made.

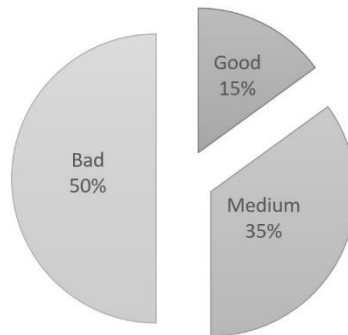
The distribution of temperaments in the sample is shown in Fig. 2. Melancholic (27%) and sanguine (26%) make up about one fourth of the respondents. The number of choleric people (34%) is three times higher than the number of phlegmatic people (13%). It turned out that there were more extroverts than introverts at the Faculty of Informatics, Mathematics, and Computer Science.



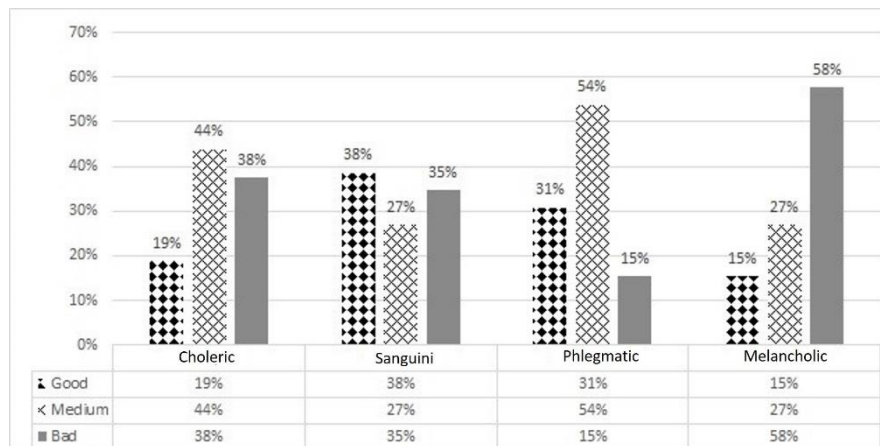
**Figure 2.** Distribution of temperaments in the sample of students

As the number of surveyed first-year students was two times more than the number of second-year students, we decided to create a training sample of the total number of second-year students, and half of the first year, 100 people in total. The sample on which the model will be tested consists of 40 first-year students.

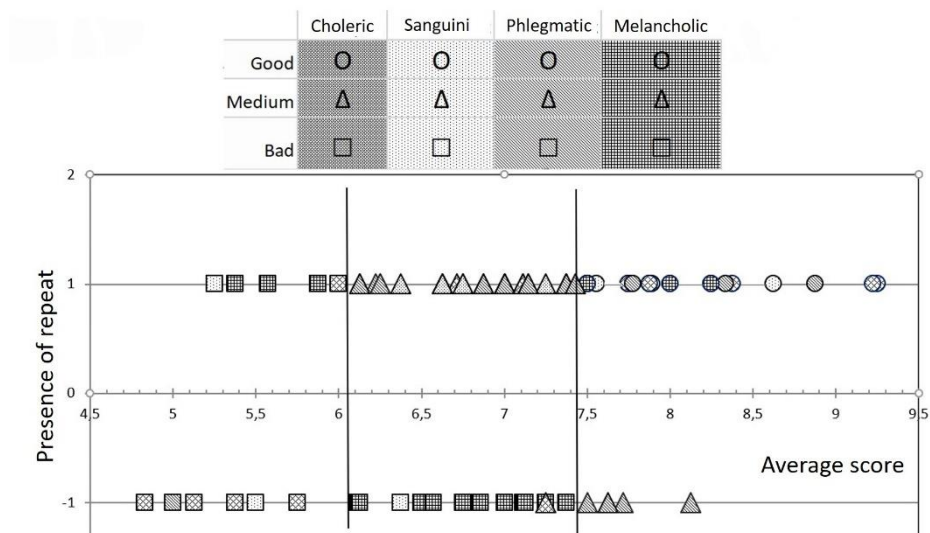
Students from the training sample, depending on the average score and the occurrence of retakes, were divided into categories Good, Medium and Bad, respectively, with a low, medium and high probability of not passing any of the exams (Fig. 3). The category of Good includes students of the first third of rating without retaken exams, the Medium category includes the second third of rating without retaken exams, and the Bad category includes all the rest students (see Figure 3a, Figure 3b).



**Fig 3.** Distribution of students by category



**Fig 3a.** Distribution of temperaments by category (in%)



**Fig 3b.** Distribution of temperaments by categories (Good / Medium / Bad) depending on the average score and the presence / absence of repeats

In Figure 3 triangles represent students from the category Medium, squares represent students from the category Bad, both students marked by triangles and students marked by squares have to retake exams - that is why they are marked at point -1. Triangles have average grade more than 6. Circles are students from category Good -they do not have academic debts. All the circles are at point +1. Figure 3 shows that half of the students fell into the “Bad” category. Such breakdown of the data is explained by the fact that almost each second student of the interviewed had a retaken examination in a term. This is typical for the Faculty of Informatics, Mathematics and Computer Science since a number of disciplines studied at the beginning of the first and second years of education that are rather difficult. It is worth noting that the allocation to category Bad does not mean that the student will not pass three or more exams and is on the verge of expulsion. Many of the students assigned thereto have a high grade average, but an academic debt in one of the disciplines. Since the aim of this study is to define all the students who are expected to have failures, therefore, even students with good progress and some academic failure may be categorized as Bad. From Figure 3a it can be seen that both general tendencies and differences among temperaments can be traced. The pattern of categorization in choleric and phlegmatic students is similar: the most number of students got into Medium category (44% of choleric, 54% of phlegmatic students). Among sanguine people, 38% are students from the Good category. For melancholic people, the largest number of students falls into category Bad.

Let us consider the factor of presence/absence of retake in more detail. Fig. 3b shows the distribution of temperaments by category (Good, Medium, Bad) depending on the average score and the presence/absence of retake. In fig. 3b the following is clearly visible:

- straight A students (Good, marked with a circle figure) have a high average score and no retakes;
- Students category Medium (marked with a triangle figure) have a grade point average of 6.3 to 7.4 and have no retakes;
- students from category Bad (marked with a square figure) have a low average score (below 6) and/or retakes; most Bad students have retakes.

### 3. Methods

The task of dividing students into the categories of high, medium and low probability of having academic failures is a task of classification based on supervised learning [19].

At the first stage of work, we have to select these characteristics that significantly influence the average grade and probable retakes of examinations.

At the second stage of work, we were looking for the algorithm of the optimal classification of students from the training sample (with the greatest number of guessed Bad category students). It was important for our work that the algorithm can assign each student to the necessary category (Bad/Medium/Good) based on each student's formalized characteristics. Machine learning methods were used for this such as the kNN algorithm and the decision tree [11, 25]. As we did not know beforehand which method would give the most exact result, we tested both of them and found the parameters for the maximum precision of guessing unsuccessful students.

At the third stage of the work, the best model was determined and used for predictions for the tested group of first-year students.

#### 3.1. Finding meaningful parameters

The correlation factor helps to understand the degree of dependence between two or among more parameters and is quite successfully used in Sociology of Education and Data Mining [16]. For subsequent analysis and interpretation of the data we represented the data obtained in a convenient form: students' answers were translated into boolean variables: 1 was used if a student agreed with a statement and 0 if otherwise. Using the correlation factor we selected from all the characteristics those that produce the most impact on the grade average and retaken examinations. The calculation results are shown in Table 1. For the purpose of our investigation the value of the correlation coefficient is significant if its absolute value is more than 0.2 (in the table in *italics*) and insignificant, if it is between 0 and 0.2.

**Table 1.** Dependence between academic performance and temperament

Characteristic	Grade average correlation	Retaken examination correlation
Activeness at school	0.06	0.09
Perseverance	0.19	-0.24
Setting of priorities	0.29	-0.18
Living in a dormitory	0.06	-0.05



Living with mates	-0.01	0.01
Living with parents	-0.01	0.02
Living alone	-0.04	0
Extroversion	0.07	-0.18
Rationality	-0.004	-0.08
<i>Choleric</i>	-0.1	0.044
<i>Sanguine</i>	0.05	-0.16
<i>Phlegmatic</i>	0.13	0.017
<i>Melancholic</i>	-0.04	0.2

As it is possible to see from the table 1, the characteristics of “perseverance” and “setting priorities” depend significantly on academic performance, which seems obvious and explained, while temperament and the resulting psychological characteristics are less correlated with the studying success. But since we study the influence of temperaments (Choleric / Sanguine / Phlegmatic / Melancholic) on academic success, they were highlighted in italics, because their absolute value of correlation with at least one of the signs is greater than 0.1.

So, the dependence between academic performance and temperament does exist, and it is rather significant in some cases. For example, choleric have a lower grade average (cor = -0.1) while perseverance of phlegmatics provides for higher marks (cor = 0.13). Sanguine re-take examinations rarely (cor = -0.16), and for melancholics the probability of retaking exams is considerable (cor = 0.2).

### 3.2. Comparison of classification using the kNN method and the decision tree

The first method to test was kNN-method. Three series of calculations were carried out:

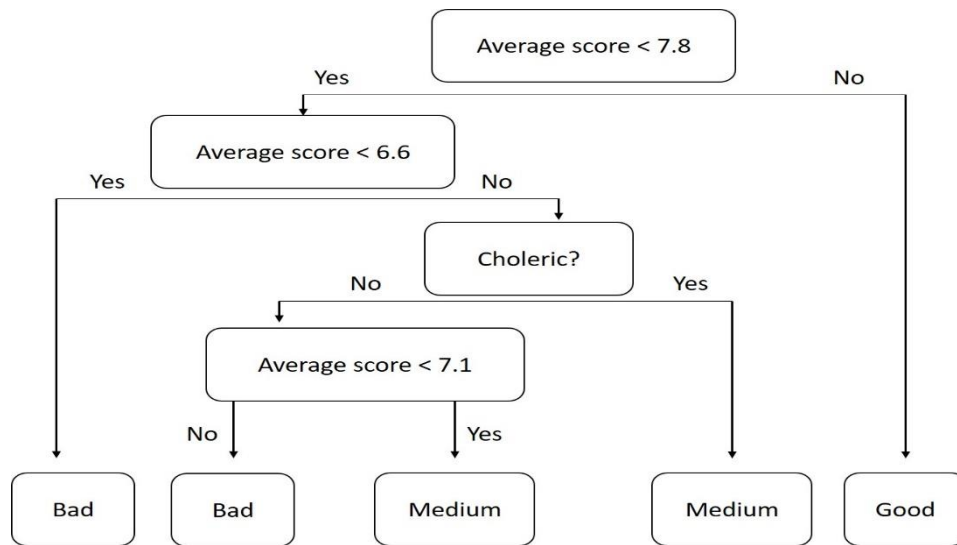
- using all characteristics with high correlation and a grade average. The best result was achieved with 8 nearest neighbors (73%);
- all characteristics without grade average. The best result was shown with 5 neighbors (60%). This is a good indication as probability of allocation to a required category is twice as higher as random guessing of category (which is 33%);
- using only grade average. The best result was obtained with 4 neighbors and was 77% of guessing. This is well-reasoned because a student with a low grade average is more likely to be assigned to the bad students' category and, accordingly, more neighbors of such student with a low grade average will be assigned to Bad.

It is important to mention that when applying kNN method psychological characteristics decrease the accuracy of calculations, therefore, the resulting classification is not suitable for our purposes.

The second method to test was the decision tree method:

- when all the characteristics were used for construction, the accuracy reached 74%;
- without taking into account the average score, the tree was built on the characteristics "Extroversion" and "Rationality" and the accuracy reached 62%;
- based on the average score only, the tree gave an accuracy of 76%;

- the best result was obtained when building a tree on the traits "Perseverance", "Ability to prioritize", "Extraversion", "Rationality" and type of temperament, and the accuracy reached 84%.



**Figure 4.** Decision tree with classification parameters

As it is possible to see, a different combination of parameters can give an increase in guessing accuracy compared to using all features.

The accuracy of answers using both methods fluctuated around 75%, but the decision tree showed the best result. Therefore, this method was chosen for the analysis. The final decision tree with the parameters is shown in Fig. 4. The final distribution included only choleric people, probably because their psychological state can be unstable, while other the temperaments do not significantly affect the results of studies. On the other hand, we conducted a study on two streams of students, the next stage is to analyze more data in order to confirm or deny the influence of temperament on student dropout.

## 4. Research Results

With the help of the decision tree generated at the previous step the students of the first-year were divided into three groups: Bad, Medium and Good. In order not to reveal personal information, the last name of each student was replaced with the symbol Student 1, Student 2, ..., Student N. We found that 22 students were in the Bad category (the category of risky students), in the Medium category - 12 students, in the Good category - 6 students. After the final session in module 4 and all the retakes in the fall, it became possible to check the research results.

Table 2 presents data on the real rating and the occurrence of student retakes at the end of the first year of study.

**Table 2.** Data summary

Predicted category	Total students	Students with retakes		Dropouts or Individual plans		A high score (strictly over 7,5)	
		num	%	num	%	num	%
Bad	22	14	64%	6	27%	0	0%
Medium	12	3	25%	1	8%	4	33%
Good	6	0	0%	0	0%	6	100%

Twenty two students fell into the Bad category, 14 students (64%) had retakes, according to the results of all autumn retakes 6 students (27%) remained with academic debts or were expelled. It should be noted that 4 out of 6 students still had an academic debt in one discipline either Mathematical Analysis or Linear Algebra. In such a case they were offered an Individual Learning Plan (ILP), according to which that study these disciplines again and continue to study on a commercial basis.

Twelve students fell into the Medium category, 3 students (25%) had retakes, but successfully retook the exams and were transferred to the 2nd year without academic debts. One student from the Medium category was expelled of her own free will, the reason was not any academic debts, but she decided to radically change her field of studies and career.

Among the students who fell into the Good category (6 people) there were no retakes and none of the students in this category were expelled.

Thus, we can conclude that the result of the distribution of students by categories is in good agreement with the real situation. It is also important to pay attention to the average score of students: students with a high GPA (strictly over 7.5) are more in the Good category, fewer in the Medium category and none in the Bad category.

- As a result of the study, the following recommendations were given to the educational office of the Faculty of Informatics, Mathematics and Computer Science HSE - Nizhny Novgorod:
- pay close attention to the students from category Bad (e.g. talk with students or their legal representatives);
- provide availability of additional classes for the students of categories Bad and Medium;
- engage training assistants into helping students with understanding and fulfilling home assignments;
- pay additional attention to students from the category Medium with a low average score after the first half of the academic year.

In conclusion, we will describe the algorithm for finding risky students based on temperament:

1. Conducting a survey of 1st and 2nd year students (September-October of the new academic year) in order to obtain data on psychological characteristics.
2. Converting survey results to boolean variables (in accordance with the parameters in table 1).
3. Formation of a training sample from 2nd year students.
4. Obtaining information about the rating and the occurrence of retakes (for students from the training sample).

5. Calculation of the correlation of psychological characteristics with the average score and the occurrence of retakes.
6. The choice of parameters that are significant for our study.
7. Classification of students from the training sample into categories Good, Medium, Bad, depending on the rating and the occurrence of retakes.
8. Building a decision tree for significant parameters in accordance with the resulting classification.
9. Formation of a test sample of the 1st year students.
10. Obtaining information about the rating and the occurrence of retakes after 1 module (November of the current academic year).
11. Classification of the 1st year students using the constructed decision tree.
12. Recommendations for the study office to pay attention to the performance of students who fall into the Bad category, as well as in the Medium category with a low GPA (no later than the beginning of December of the current academic year).

## 5. Conclusion

Studying the influence of temperament on student performance at the NRU HSE-Nizhny Novgorod at the Faculty of the IT Sciences we identified the most important psychological factors. They turned out to be "Perseverance" and "Ability to prioritize" - the presence of these factors sharply increased the average score and reduced the likelihood of retakes. It has been observed that hot-tempered choleric people have a lower GPA, while the calmness and measuredness of phlegmatic people help them study better. Sanguine people are less likely to have retakes, while the chance of a melancholic to retake exams is much higher. The more extraversion a student demonstrates, the higher his GPA.

We believe that our research might be useful to other universities for:

1. identifying academically unsuccessful students and focusing on "risky" students. One of the most important goals of the Faculty of Informatics, Mathematics, and Computer Science (HSE Nizhny Novgorod) is to train and transfer without academic debts to the 2nd as many first-year students as were accepted for the program. Otherwise, the resources of the state (in the case of government-subsidized education) or a student personal funds (in the case of paid education) spent on education will not be used rationally.
2. forming an individual educational trajectory. At HSE-Nizhny Novgorod today, there are flexible opportunities for switching from one educational program to another using Individual Learning Plan (ILP). We assume that forming studying groups, taking into account students personal characteristics' will increase the performance of each student.

We understand that there are certain limitations to our study. In this paper, a small pool of baseline data is presented (140 student responses). And the Faculty of Informatics, Mathematics, and Computer Science (HSE Nizhny Novgorod) is not large enough to talk about "real" Big Data and use all the opportunities of machine learning methods. In the future, we plan to use the longitude data for results verification over

several years. It would also be interesting to try other data mining methods to predict academic failures and dropout.

## 6. Acknowledgements

We would like to thank the education office specialists of HSE University Nizhny Novgorod for providing data for conducting the research.

## References

1. Anuwatvisit S, Tungkasthan A, Premchaiswadi W. Bottleneck mining and Petri net simulation in education situations. In: ICT and Knowledge Engineering (ICT & Knowledge Engineering), 10th International Conference. IEEE; Bangkok, Thailand; 2012, 244–251.
2. Arulselvan A., Boginski, V., Mendoza, P., & Pardalos, P. Predicting the nexus between post-secondary education affordability and student success: An application of network-based approaches. ASONAM'09. International Conference on Advances in Social Networks Analysis and Mining. 2009, pp. 149–154.
3. Bannert M, Reimann P, Sonnenberg C. Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacogn Learn* 2014, 9:161–185.
4. Barreiros BV, Lama M, Mucientes M, Vidal JC. Softlearn: a process mining platform for the discovery of learning paths. In: 14th International Conference on Advanced Learning Technologies. IEEE, Athens, Greece; 2014, 373–375.
5. Bogarin A, Cereso R, Romero C. A survey on educational process mining. *WIREs Data Mining and Knowledge Discovery*. V.9, Is.1, doi:10.1002/widm.1230; 2018
6. Bogarin A, Romero C, Cerezo R, Sánchez-Santillán M. Clustering for improving educational process mining. In: Proceedings of the Fourth International Conference on Learning Analytics And Knowledge. ACM, Indianapolis, USA; 2014, 11–15
7. Bulycheva P., Oshmarina, O., Shadrina, E.: Identifying academically “unsuccessful” first-year students: a case study of Higher School of Economics – Nizhny Novgorod. In: *Vestnik of Lobachevsky State University of Nizhny Novgorod. Series: Social Sciences*, 2016, no. 2(42), pp. 136-143.
8. Cairns AH, Gueni B, Assu J, Joubert C, Khelifa N. Analyzing and improving educational process models using process mining techniques. In: *The Fifth International Conference on Advances in Information Mining and Management*. Brussels, Belgium; 2015, 17–22.
9. Campbell, J.P. Utilizing student data within the course management system to determine undergraduate student academic success: An exploratory study. Purdue University, 2007.
10. Dokuka S.V., Valeeva, D.R., Yudkevich, M.M. How Academic Failures Break Up Friendship Ties: Social Networks and Retakes//*Educational Studies Moscow*, 2017, no. 1 pp.8 -21.
11. Friedl M. A., & Brodley, C. E. Decision tree classification of land cover from remotely sensed data//*Remote sensing of environment*, 1997, no. 61(3), pp. 399–409.
12. Gorbunova E. Vybytiya studentov iz vuzov: issledovaniya v Rossii i SShA [Elaboration of Research on Student Withdrawal from Universities in Russia and the United States]. *Voprosy obrazovaniya / Educational Studies Moscow*, 2018, no. 1, pp. 110-131.

13. Gorbunova E. V., Gruzdev I. A., Frumin I. D. Studencheskii otsev v rossiiskikh vuzakh: k postanovke problemy [Student Dropout in Russian Higher Education Institutions: the Problem Statement]. *Voprosy obrazovaniya* [Educational Studies], 2013, no. 2, pp. 67–81.
14. Howard L, Johnson J, Neitzel C. Examining learner control in a structured inquiry cycle using process mining. In: *Proceedings of the 3th International Conference on Educational Data Mining*, Pittsburgh, USA, 2010:71–80
15. Ignatov D., Mamedova S., Romashkin N., Shamshurin I. What can closed sets of students and their marks say?// In *proceedings of 4th International Conference on Educational Data Mining*, Mykola Pechenizkiy et al. (Eds.), EDM-2011, TU/e Eindhoven, 2011, 223-228.
16. Izenk G. & Vilson G. Kak izmerit lichnost [How to measure a personality]. M.: Kogito tsentr, 2000, pp.156–159.
17. Keyek-Franssen D. Praktiki uspešnosti studentov: ot ochnogo obucheniya k masshtabnomu i obratno [Practices for Student Success: From Face-to-Face to At-Scale and Back]. *Voprosy obrazovaniya / Educational Studies* Moscow, 2018, no. 4, pp. 116-138.
18. Kochergina E., Prakhov I. Vzaimosvyaz' mezhdu otnosheniyem k risku, uspevaemost'yu studentov i veroyatnost'yu otchisleniya iz vuza [Relationships between Risk Attitude, Academic Performance, and the Likelihood of Drop-outs]. *Voprosy obrazovaniya* [Educational Studies] Moscow, 2016, no 4, pp. 206-228.
19. Luan, J. Data mining and its applications in higher education// *New directions for institutional research*, 2002, no. 113, pp. 17–36.
20. Lust T., Meskens, N., & Ahues, M. Predicting academic success in Belgium and France Comparison and integration of variables related to student behavior. *arXiv preprint arXiv:1408.4955*. 2014.
21. Mirkin B. G. Vvedeniye v analiz dannyh: uchebnik I praktikum dlya bakalavriata I magistratury [Introduction to data analyses: theory and practice for bachelor and master courses], M.: Izdatelstvo Urite, 2014, 174 p.
22. Poldin O., Valeeva, D., & Yudkevich, M. How social ties affect peer group effects: Case of university students//*SSRN Electronic Journal*, January 2013. DOI: 10.2139/ssrn.2207666
23. Pal S. Mining educational data using classification to decrease dropout rate of students. *arXiv preprint arXiv:1206.3078*. 2012.
24. Southavilay V, Yacef K, Callvo RA. Process mining to support students' collaborative writing. In: *Proceedings of the 3th International Conference on Educational Data Mining*, Pittsburgh, USA, 2010:257–266.
25. Superby J.-F., Vandamme, J., & Meskens, N. Determination of factors influencing the achievement of the first-year university students using data mining methods. *Workshop on Educational Data Mining*. 2006, no. 32, pp. 234.
26. Vidal JC, Lama M, Bugarín A. Petri net-based engine for adaptive learning. *Expert Syst Appl* 2012, 39:12799–12813