*MSC: 68Т09* Original article

**Educational data mining for predicting the academic**

**performance of university students**

**N.A. Popova1, E.S. Egorova2**

1 Penza State University

440026, Russia, Penza, 40 Krasnaya street

2 Penza State Technological University

440039, Russia, Penza, 1a/11 Baidukova passage/Gagarina street

***Abstract.***Progress in the field of data mining makes it possible to use educational data to improve the quality of educational processes. This article examines various methods of analyzing student achievement data. The focus is on two aspects: first, predicting students' academic achievements at the end of a four-year undergraduate curriculum; second, examining typical student progressions and combining them with the prediction results. Approximately 10 classification algorithms were used in the prediction process. An approach to improving the performance of classification methods is proposed where classifier attributes are selected during their training. Two important groups of students were identified: low-achieving and high-achieving students. The results show that by focusing on a small number of courses that are indicators of particularly good or poor performance, it is possible to prevent and support low-achieving students in a timely manner, and to provide advice and opportunities to high-achieving students.

***Keywords:*** analysis of educational data, decision tree, clustering, forecasting, academic performance, dissociation

**REFERENCES**

1. Belonozhko P.P., Karpenko A.P., Khramov D.A. Analysis of educational data: directions and prospects for application. *Internet-zhurnal “Naukovedeniye”* [SCIENCE online journal]. Vol. 9. No. 4 (2017). URL: http://naukovedenie.ru/PDF/15TVN417.pdf (In Russian)

2. Mamontova M.Yu. Quality of educational achievements: assessment and forecast based on the results of criteria-oriented testing. *Education and science. News of the Ural Branch of the Russian Academy of Education*. 2009. No. 3(60). Pp. 18–26. (In Russian)

3. Rusakov S.V., Nakaryakova N.N. Forecasting the progress of first-year students using a decision tree based on their results of passing the exam. *Nauka. Informatizatsiya. Tekhnologii. Obrazovaniye* [Science. Informatization. Technologies. Education]: *Materialy XI mezhdunarodnoy nauchno-prakticheskoy konferentsii.* Yekaterinburg: Rossiyskiy gosudarstvennyy professional'no-pedagogicheskiy universitet, 2018. Pp. 589–594. (In Russian)

4. Firstov V. E. Sociometric and informational aspects of clustering the student contingent in the organization and optimization of group cooperation in the educational process at school and university. *Izvestiya of Saratov University. New series. Series: Philosophy. Psychology. Pedagogy*. 2014. Vol. 14. No. 1. Pp. 110–118. (In Russian)

5. Medvedev D., D’yakonov A. New Properties of the Data Distillation Method When Working with Tabular Data. Conference proceedings “Analysis of Images, Social Networks and Texts”. *Lecture Notes in Computer Science*. Vol. 12602. Springer, Cham. 2021. https://doi.org/ 10.1007/978-3-030-72610-2\_29

6. Sucholutsky I., Schonlau M. Soft-Label Dataset Distillation and Text Dataset Distillation. *International Joint Conference on Neural Networks*, Shenzhen, China, 2021. Pp. 1–8. DOI: 10.1109/IJCNN52387.2021.9533769.

7. Nikonorova M.L. Computer model for solving classification problems in the Rapid Miner software environment. *Medical education and professional development*. 2017. No. 2–3(28–29). Pp. 24–33. (In Russian)

8. Filyak P.Yu., Vinogradov M.A. Application of Rapid miner and open environments as data mining tools for security. *Informatsiya i bezopasnost'* [Information and security]. 2017. Vol. 20. No. 4. Pp. 552–555. (In Russian)

9. Maimon O., Rokach L. Data Mining and Knowledge Discovery Handbook. Springer Science, Business Media, 2010. 1285 p. ISBN: 978-0-387-09822-7.

**Information about the authors**

**Popova Nataliya Aleksandrovna,** Candidate of Technical Sciences, Associate Professor of the Department of Mathematical Support and Computer Use, Penza State University;

440026, Russia, Penza, 40 Krasnaya street;

popov.tasha@yandex.ru, ORCID: https://orcid.org/0000-0001-9713-4897

**Egorova Ekaterina Sergeevna,** Candidate of Economic Sciences, Associate Professor of the Department of Applied Informatics, Penza State Technological University;

440039, Russia, Penza, 1a/11 Baidukova passage/Gagarina street;

katepost@yandex.ru, ORCID: https://orcid.org/0000-0002-0816-0944