

# ОБЗОР ЭТАПА ОПРЕДЕЛЕНИЯ ЗНАЧИМЫХ ФАКТОРОВ В ЗАДАЧИ СОЗДАНИЯ СИСТЕМЫ ПРОГНОЗИРОВАНИЯ АКАДЕМИЧЕСКОЙ УСПЕВАЕМОСТИ

*К.Г. Виноградов, аспирант гр.А9-36  
Томский политехнический университет  
E-mail: kgv1@tpu.ru*

## **Введение**

Сохранность контингента студентов в наше время – это фактор конкурентного преимущества образовательных учреждений. Низкие показатели сохранности контингента образовательного учреждения влекут за собой уменьшение бюджетного финансирования и снижение рейтинговых показателей. Один из способов решения проблемы сохранности контингента в образовательном учреждении – это создание системы прогнозирования успеваемости студентов, которая даст возможность предоставить своевременную помощь отстающим студентам. В исследованиях систем прогнозирования успеваемости студентов выделяют несколько важных этапов, одним из которых является этап выделения значимых факторов, влияющих на успеваемость.

## **Обзор этапа определения значимых факторов**

Прогнозирование успеваемости основывается на анализе больших данных, для выполнения которого обычно используют методы машинного обучения. Методы машинного обучения чувствительны по отношению подаваемых на анализ данных. Когда выборка данных, подаваемая на вход методам машинного обучения, содержит шумовой признак (избыточный признак, который не влияет на конечный ответ), то происходит ухудшение качества создаваемой модели [1]. Такая ситуация может возникнуть еще на этапе постановки задачи во время формирования данных, когда еще неизвестно, какие атрибуты бесполезны или дублируют другие данные. Поэтому при создании системы прогнозирования успеваемости одним из важнейшим этапом является отбор значимых факторов, влияющих на успеваемость, которые используются для предварительного ранжирования данных.

Выделить универсальную выборку данных для выявления академической успеваемости невозможно. Это объясняется тем, что в каждом образовательном учреждении учебный процесс построен своим индивидуальным образом. Поэтому каждое учебное учреждение требует выявления своей собственной выборки данных, включающей в себя только те параметры, что влияют на успеваемость студентов именно этого учреждения, учитывая все особенности учебного процесса. Однако, алгоритм для определения переменных и атрибутов, которые оказывают наибольшее влияние на успеваемость студентов, также отсутствует, что является одной из основных проблем в задаче определения академической успеваемости.

В [2] была составлена классификация параметров, используемых при прогнозировании академической успеваемости. В классификации были выделены 9 категорий факторов: электронное обучение студентов; предыдущая успеваемость студентов; студенческая среда; студенческая демография; атрибуты преподавателя; атрибуты курса; социальная информация студентов; информация оценки курсов студентами, собранная путем опроса студентов; опыт студентов.

В [2, 3] были предприняты попытки создания обобщенного набора факторов и атрибутов, влияющих на успеваемость и результаты обучения студентов в сфере высшего образования, путем выявления наиболее распространенных и широко используемых категорий факторов. В результате таких исследований выявлены следующие категории. Самой важной категорией факторов, влияющей на успеваемость студентов, определена категория показателей предыдущей успеваемости учащихся. Следующей по значимости выделяют категорию демографических факторов. Также к важнейшим относят факторы, связанные с социальной активностью студентов. В [2] еще выделяют факторы активности электронного обучения учащихся.

Проанализировав результаты работ [4, 5], где сравниваются эффективность предсказаний успеваемости, построенных на разных комбинациях факторов, можно сделать вывод, что использование комбинации, сочетающей в себе значимые факторы, относящиеся к разным категориям, приводит к более эффективным результатам, нежели использование комбинации, включающей значимые факторы из одной категории. В тех же [4, 5], можно наблюдать и обратную ситуацию: комбинация значимых факторов, содержащая максимально доступное количество факторов, не всегда дает самый эффективный результат. Таким образом, для каждой систем прогнозирования

успеваемости студентов необходимо подбирать свою оптимальную комбинацию факторов, которая в результате предсказания выдаст самый эффективный ответ.

В исследованиях [3, 5, 6] рассматриваются психометрические факторы (интерес студента, учебное поведение, поддержка семьи). Данные факторы относятся к категории социальной информации о студентах и являются очень эффективными при прогнозировании успеваемости. Однако применение психометрических факторов имеет ряд сложностей. В первую очередь, это трудности сбора такого рода данных. Сбор проводится путем опроса студентов, во время которого респонденты могут предоставить недостоверную информацию, искажая или скрывая ее по личным причинам. Также осложнением при сборе информации является отсутствие точной шкалы оценивания психометрических факторов, в связи с чем можно получить различные данные при одном уровне проявления фактора. Например, один человек может оценить свой уровень заинтересованности в учебе как высокий, а другой как средний, а на деле оба заинтересованы в учебе в равной степени, но выдают разные данные из-за отсутствия конкретной шкалы оценивания. Психометрические факторы сложны не только в сборке, но и в обработке. Чаще всего, психометрические факторы представляют собой совокупность качественных данных. Данные в качественном представлении хуже обрабатываются методами анализа данных, нежели количественные.

Существуют исследования, в которых прогностическую модель успеваемости строят на нестандартных данных. В статье [7] рассматривалось влияние факторов, связанных со временем (время изучения теоретического материала, время прохождения тестов, время общего посещения и т.д.) на успеваемость студентов, проходящих электронное обучение. В [8] экспериментальным путем подтверждают зависимость между успеваемостью и поведением студентов в интернете.

## **Заключение**

Был проведен анализ научно-технической литературы для освещения вопросов выявления факторов, влияющих на успеваемость студентов. В результате проведенного анализа был сделан вывод, что для каждого учебного заведения должна подбираться личная выборка факторов, учитывающая особенности образовательного процесса данного заведения. Также была приведена классификация факторов по категориям, были рассмотрены особенности использования психометрических факторов и выделены нестандартные факторы, которые использовались для предсказания успеваемости в некоторых работах.

## **Список использованных источников**

1. Воронцов К.В. Лекции по методам оценивания и выбора моделей // 2010. – С.2-3. – Режим доступа: <http://www.machinelearning.ru/wiki/images/2/2d/Voron-ML-Modeling.pdf>, свободный (дата обращения: 04.11.2020)
2. Saa A.A. Factors Affecting Students' Performance in Higher Education: A Systematic Review of Predictive Data Mining Techniques / A.A. Saa, M. Al-Emran, K. Shaalan // *Technology, Knowledge and Learning*. – 2019. – Vol. 24. – P. 567 598. – doi: 10.1007/s10758-019-09408-7
3. Shahiria A.M. A Review on Predicting Student's Performance using Data Mining Techniques / A.M. Shahiria, W. Husaina, N.A. Rashida // *Procedia Computer Science*. – 2015. – Vol. 72. – P. 414 422. – doi: 10.1016/j.procs.2015.12.157
4. Tomasevic N. An overview and comparison of supervised data mining techniques for student exam performance prediction / N. Tomasevic, N. Gvozdenovic, S. Vranes // *Computers & Education*. – January 2020. – Vol. 143, Article 103676. – doi: 10.1016/j.compedu.2019.103676
5. Francis B.K. Predicting Academic Performance of Students Using a Hybrid Data Mining Approach / B.K. Francis, S.S. Babu // *Journal of Medical Systems*. – 2019. – Vol. 43, Iss. 162. – doi: 10.1007/s10916-019-1295-4
6. Ajibade S.-S.M. A Data Mining Approach to Predict Academic Performance of Students Using Ensemble Techniques / S.-S.M. Ajibade, N.B. Ahmad, S.M. Shamsuddin // *Intelligent Systems Design and Applications. ISDA. – Advances in Intelligent Systems and Computing*. – 2018. – vol. 940. – p. 749 760. – doi: 10.1007/978-3-030-16657-1\_70
7. Hu Y.-H. Developing early warning systems to predict students' online learning performance / Y.-H. Hu, Ch.-L. Lo, Sh.-P. Shih // *Computers in Human Behavior*. – July 2014. – Vol. 36. – P. 469 478. – doi: 10.1016/j.chb.2014.04.002
8. Prediction of academic performance associated with internet usage behaviors using machine learning algorithms / X. Xu, J. Wang, H. Peng, R. Wu // *Computers in Human Behavior*. – 2019. – Vol. 98. – P. 166 173. – doi: 10.1016/j.chb.2019.04.015