

0.1. Минуллин Д.А. Сравнительный анализ методов машинного обучения в образовательной аналитике

Данная работа посвящена применению методов анализа больших данных для анализа информации об образовательном процессе в школе. Основу исследования составили данные, собранные через государственную информационную систему «Электронное образование в Республике Татарстан». Система включает в себя базы данных образовательной информации по всем учащимся и всем педагогам общеобразовательных организаций РТ. База данных содержит более двух миллиардов информационных единиц, включая информацию об успеваемости более миллиона учащихся и профессиональной деятельности более 120 000 преподавателей.

Образование является сферой, в которой производится и накапливается большое количество данных. Правильный анализ такой информации может помочь составить более полную картину процесса обучения, выявить полезные и, возможно, неочевидные связи. Методы машинного обучения могут позволить предсказать исход какой-либо ситуации, основываясь на исторических данных. В отличие от традиционных мер измерения результатов учащихся, которые помогают измерять только конечный результат, применение методов машинного обучения может помочь получить ценную информацию о том, как улучшить и персонализировать обучение, составлять прогнозы и рекомендации, проводить изменения в режиме реального времени [1].

Задачей данного исследования являлся анализ процесса перехода учащегося из 9-го класса в 10-й. Для первичной обработки данных с целью группировки учащихся и расчёта средних оценок за четверть был разработан программный комплекс на языке программирования Python с использованием библиотеки для параллельных и распределённых вычислений Dask [2]. Для прогнозирования перехода учащегося использовались методы машинного обучения (Таблица 1) [3], которым на вход подавались оценки ученика и на выход параметр отвечающий за переход ученика в следующий класс.

В результате проведённого анализа видно, что про-

цент точности прогнозирования невысок, все методы показали приблизительно одинаковый результат (в районе 70%), но даже основываясь на таком показателе удалось выделить некоторые взаимосвязи. Из полученных результатов видно, что есть возможность прогнозирования, но одних только оценок для точного ответа недостаточно. В дальнейшем данное исследование будет продолжаться с расширением количества параметров, характеризующих образовательный процесс.

Работа выполнена при финансовой поддержке РФФИ (грант № 19-29-14082).

Научный руководитель — канд. физ.-мат. наук, доцент Гафаров Ф. М.

Список литературы

- [1] ROMERO C., VENTURA S. Educational data mining: A review of the state of the art // IEEE Transactions on Systems Man and Cybernetics. Part C (Applications and Reviews). 2010. Vol. 40. N. 6. P. 601–618.
- [2] ROCKLIN M. Dask: Parallel computation with blocked algorithms and task scheduling // Proc. 14th Python In Science Conf. (SCIPY 2015). P. 126–132.
- [3] SARKER IQBAL H. Machine learning: Algorithms, real-world applications and research directions // SN Computer Science. 2021. Vol. 2. N. 3. P. 160.

Метод	Точность, %
LogisticRegression	70.20
LinearDiscriminantAnalysis	70.45
KNeighborsClassifier	71.12
GaussianNB	70.47
DecisionTreeClassifier	70.12
RandomForestClassifier	70.79
SupportVector(linear)	70.12
SupportVector(rbf)	71.08
Simple neural network	70.14

Таблица 1: Точность прогнозирования.