

УДК 004.94

## ПОСТРОЕНИЕ ПОРТРЕТОВ УСПЕВАЕМОСТИ ОБУЧАЮЩИХСЯ НА ОСНОВЕ МЕТОДОВ АНАЛИЗА ОБРАЗОВАТЕЛЬНЫХ ДАННЫХ

Тирский Б.Д. (Университет ИТМО), Рунушкина О.Р. (Санкт-Петербургский политехнический университет Петра Великого)

Научный руководитель – к. п. н., доцент Авксентьева Е. Ю.  
(Университет ИТМО)

Работа выполнена в рамках темы НИР №620164 «Методы искусственного интеллекта для киберфизических систем».

Данная научно-исследовательская работа посвящена разработке алгоритма кластеризации и созданию интерактивной визуализации данных. Были проведены исследования современных методов и инструментов анализа образовательных данных. Выбран оптимальный инструментарий анализа данных и построения интерактивных визуализаций.

**Ключевые слова:** Интеллектуальный анализ данных, аналитика образовательных данных, кластеризация.

С увеличением количества учащихся возникла потребность в реализации алгоритма, который бы увеличил эффективность обучения, дав необходимую информацию для персонализации учебного плана для каждого студента. Также представление всех данных в легко-интерпретируемой для последующего анализа платформе повышает практическую значимость данной работы.

Процесс сбора данных выглядит следующим образом.



Рис. 1. Процесс сбора данных

Таким образом была сформирована выборка, которую можно использовать для тестирования различных методов анализа образовательных данных.

Из всех методов обработки данных был выбран наиболее подходящий, так как кластеризация работает лучше остальных с большими массивами данных и с небольшим количеством кластеров.

Кластеризация — это разделение множества входных объектов на группы (кластеры) по степени «схожести» друг на друга.

Один из популярных алгоритмов кластеризации – алгоритм k-средних. Он разбивает множество элементов векторного пространства на заранее известное число кластеров k. Действие алгоритма таково, что он стремится минимизировать среднеквадратичное

отклонение на точках каждого кластера. Основная идея заключается в том, что на каждой итерации пересчитывается центр масс для каждого кластера, полученного на предыдущем шаге, затем векторы разбиваются на кластеры вновь в соответствии с тем, какой из новых центров оказался ближе по выбранной метрике. Алгоритм завершается, когда на какой-то итерации не происходит изменения кластеров [1].

В качестве объекта исследования были получены данные, в которых содержалась информация об успеваемости студентов университета по различным предметам и направлениям. После очистки получился некоторый набор параметров за последние три года [2].

Далее был реализован алгоритм. Сделали кодировку категориальных переменных, чтобы в дальнейшем была возможность анализировать данные с помощью кластеризации. С целью улучшения эффективности был использован метод главных компонент для уменьшения размерности данных [3]. После этого определяем количество кластеров – используется «Elbow method», или метод локтя, при котором определяется оптимальное разбиение данных.

Полученные кластеры выглядят следующим образом.

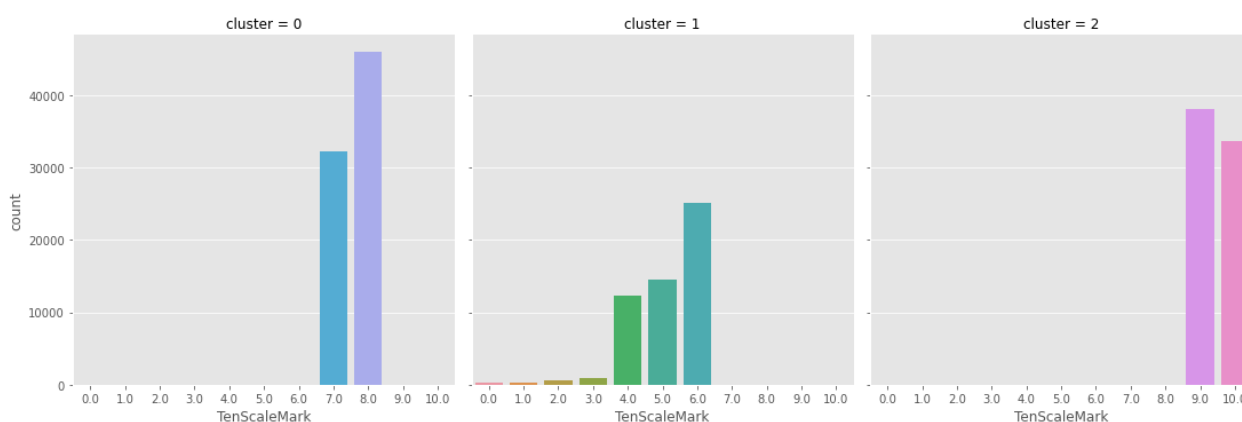
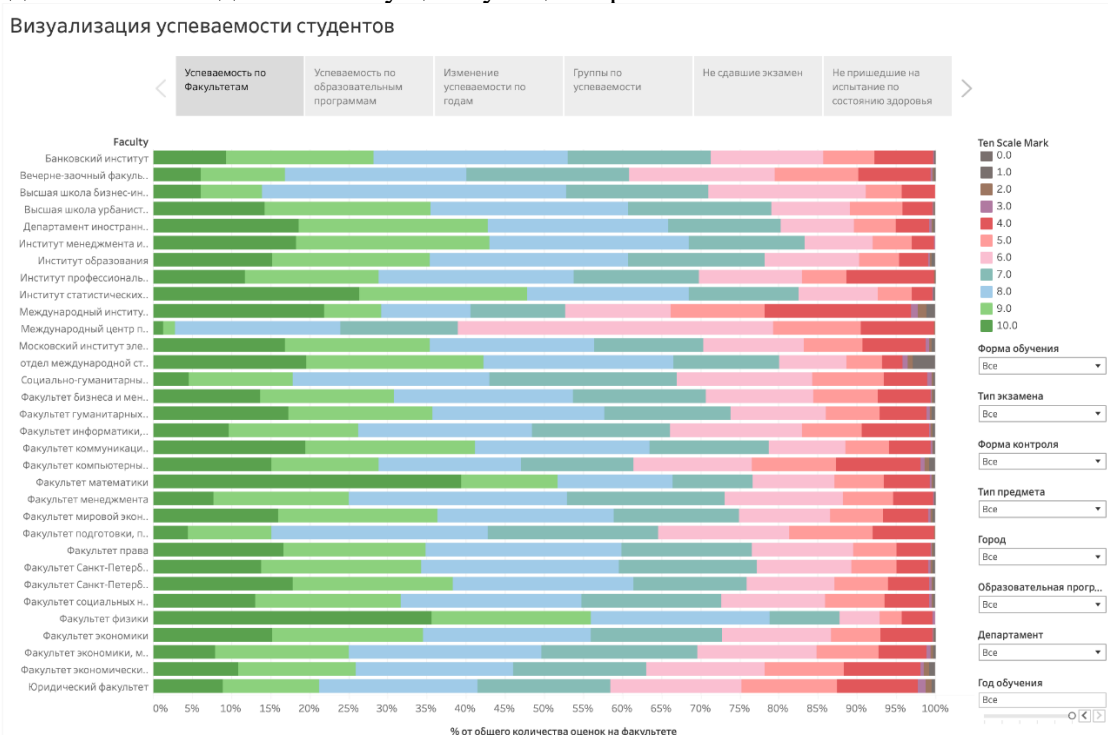


Рис. 2. Разбиение по кластерам

Заключительным этапом исследования была интерактивная визуализация. С помощью Tableau реализована возможность получать срезы интересующей информации, визуально наблюдать отличия в данных по существующим признакам.



### Рис. 3. Интерактивная визуализация

В ходе работы был реализован алгоритм для задачи исследования образовательных данных. Результаты приведены и выгружены в интерактивную среду аналитики и визуализации, то есть создан инструмент для последующего изучения полученных результатов.

#### Литература

1. CLUSTERING ALGORITHMS IN EDUCATIONAL DATA MINING: A REVIEW [Электронный ресурс]. Режим доступа: <https://www.ijcns.com/pdf/ijpcsvol7no12015-9.pdf> (дата обращения: 20.01.2022).
2. Data mining techniques for data cleaning [Электронный ресурс]. Режим доступа: [https://link.springer.com/chapter/10.1007/978-0-85729-320-6\\_91](https://link.springer.com/chapter/10.1007/978-0-85729-320-6_91) (дата обращения: 18.01.2022).
3. Study of Data Mining in Higher Education-A Review [Электронный ресурс]. Режим доступа: <http://ijcsit.com/docs/Volume%206/vol6issue01/ijcsit20150601101.pdf> (дата обращения: 23.01.2022).