

УДК 004.415.25

UDC 004.415.25

5.2.2. Математические, статистические и инструментальные методы экономики (физико-математические науки, экономические науки)

5.2.2. Mathematical, statistical and instrumental methods of economics (physical and mathematical sciences, economic sciences)

МЕТОДЫ ПРОГНОЗИРОВАНИЯ И ОЦЕНКИ КРИТЕРИЕВ В КОНТЕКСТЕ СОВРЕМЕННЫХ ТЕНДЕНЦИЙ И ИННОВАЦИЙ

FORECASTING AND EVALUATION METHODS FOR CRITERIA IN THE CONTEXT OF MODERN TRENDS AND INNOVATIONS

Параскевов Александр Владимирович
РИНЦ SPIN-код: 2792-3483
paraskevov.a@kubsau.ru
Кубанский государственный аграрный университет имени И.Т. Трубилина, Краснодар, Россия

Paraskevov Alexander Vladimirovich
RSCI SPIN-code: 2792-3483
paraskevov.a@kubsau.ru
Kuban State Agrarian University named after I. T. Trubilin, Krasnodar, Russia

Уварова Алина Геннадьевна
бакалавр 3 курса
факультет прикладной информатики
alinauv02@mail.ru
Кубанский государственный аграрный университет имени И.Т. Трубилина, Краснодар, Россия

Uvarova Alina Gennadievna
bachelor of the 3rd year
Department of Applied Informatics
alinauv02@mail.ru
Kuban State Agrarian University named after I. T. Trubilin, Krasnodar, Russia

Методы, основанные на искусственном интеллекте (ИИ) и машинном обучении (ML) достаточно новы, но при этом очень надежны в прогнозировании. Они особенно полезны с данными, имеющими слабую или вовсе отсутствующую структуру. Популярным алгоритмом являются нейронные сети - математические модели, организованные по принципу нервной сети живого организма. К методам обучения таких сетей относятся: обучение с учителем, то есть контролируемое человеком, с предварительной разметкой данных, разбиением выборки на обучающую и тестовую; обучение без учителя возможность выявить закономерности, которые человек не смог бы предположить; обучение с подкреплением, то есть на решения машины дается обратная связь человека, которая корректирует выводы. Также в категорию методов, основанных на ИИ и ML можно отнести дерево решений - способ представления правил в иерархической последовательной логической структуре, который позволяет соотнести объект или ситуацию на входе с одним или несколькими выходными узлами. Последняя категория основана на применении графических средств анализа данных. Сюда входят такие методы, как контрольный листок, диаграмма Парето, схема Исикавы, гистограмма, диаграмма разброса, расслоение, контрольная карта, график временного ряда и др. Преимуществом таких методов является простота освоения, использования, возможность применения их в комплексе с другими методами. Они достаточно удобны для восприятия даже не

Methods based on artificial intelligence (AI) and machine learning (ML) are quite new, but at the same time they are very reliable in forecasting. They are especially useful with data that has weak or no structure at all. A popular algorithm is neural networks, mathematical models organized according to the principle of the nervous network of a living organism. The methods of training such networks include: learning with a teacher, that is, supervised by a person, with preliminary data markup, splitting the sample into training and test; learning without a teacher the opportunity to identify patterns that a person would not be able to guess. Reinforcement learning, that is, human feedback is given to the decisions of the machine, which corrects the conclusions. Also included in the category of AI and ML-based methods is a decision tree, a way of representing rules in a hierarchical sequential logical structure that allows you to correlate an object or situation at the input with one or more output nodes. The latter category is based on the use of graphical means of analyzing the submitted data. This includes methods such as checklist, Pareto diagram, Ishikawa diagram, histogram, scatter diagram, layering, control map, time series graph, etc. The advantage of such methods is the ease of mastering, use, and the possibility of using them in combination with other methods. They are quite convenient for the perception of even a person who has nothing to do with information technology, so this is the simplest, but reliable enough way to make decisions

имеющего общего с информационными технологиями человека, поэтому это самый простой, но достаточно надежный способ для принятия решений

Ключевые слова: АНАЛИЗ ДАННЫХ, БОЛЬШИЕ ДАННЫЕ, ВЫСШЕЕ ОБРАЗОВАНИЕ, МАШИННОЕ ОБУЧЕНИЕ, ДАТАСЕТ

Keywords: DATA ANALYSIS, BIG DATA, HIGHER EDUCATION, MACHINE LEARNING, DATASET

<http://dx.doi.org/10.21515/1990-4665-198-038>

Введение.

Одной из основных характеристик высших учебных заведений России является рейтинг. Ранжирование формируется на таких критериях, как территориальное расположение, качество преподавательского состава, количество направлений подготовки, их содержание, соотношение бюджетных и коммерческих мест, а также научная, культурная и спортивная активность образовательного обучения, уровень трудоустройства выпускников, работа с иностранными студентами и многие другие показатели. На основании данного рейтинга базируется как финансирование, что очень важно для существования учебного заведения, так и сама известность ВУЗа среди школьников, желающих получить наилучшее образование.

Поскольку рейтинг зависит от деятельности студентов, учебным заведениям очень важно, чтобы принимаемые абитуриенты были потенциально одаренными и активными учащимися, которые будут не только поддерживать, но и улучшать данный показатель.

Получить образ сегодняшнего студента, бывшего абитуриента, предугадать, каким он будет в ближайшем будущем возможно с помощью данных о поступающих прошлых лет, которые хранятся в ВУЗе. А ведь их не обрабатывают. Если корректно собрать, обработать, применить методы и техники анализа Big Data, то можно найти скрытые закономерности, узнать пользу от внутренних процессов, на которые тратятся ресурсы (например, подготовительные занятия перед поступлением), сформировать

портрет абитуриента прошлых и будущих лет, понять как он меняется во времени. Иначе говоря, технологии Big Data позволят принимать решения высшим учебным заведениям, основываясь не на догадках, а на фактах и числах.

Для демонстрации возможностей анализа создадим набор данных абитуриентов учетно-финансового и экономической факультетов ВГБОУ ВО Кубанский ГАУ им. И. Т. Трубилина за период с 2015 по 2023 год со следующими их характеристиками:

- пол;
- возраст;
- район/населенный пункт проживания;
- сданные дисциплины и баллы;
- наличие/отсутствие достижений;
- наличие/отсутствие курсов довузовской подготовки;
- средний балл аттестата;
- аттестат с отличием или нет;
- поступление или нет;
- поступление на бюджет или коммерцию;
- на какие факультеты подавал документы;
- на какой факультет поступил.

Данные для исследования сгенерированы случайным образом с помощью программы на языке Python с использованием библиотеки random. Набор данных не содержит персональные данные и может быть использован для анализа. Точно также, как и набор реальных данных может быть обезличен и законно использован для анализа. [1]

Целью исследования является демонстрация возможности использования высшими учебными заведениями методов и технологий Big Data для проведения анализа и аудита профориентационной работы,

развития направлений подготовки, принятия решений по модернизации образовательных процессов и совершенствованию образовательного продукта.

Основная часть.

Полученный набор данных был сгенерирован при помощи программного кода, предварительно обработан, закодирован и отфильтрован по 2023 году. Он был исследован с помощью:

- визуализации и графиков;
- нейросетевого алгоритма – самоорганизующиеся карты Кохонена;
- корреляционного анализа;
- регрессионного анализа.

Визуализация данных – это процесс применения визуальных элементов для представления данных, которые легче обрабатывать. Главное преимущество визуализации данных – простота использования. На основе сырых данных уже можно извлечь практические выводы при помощи диаграмм, сводных таблиц. Для удобного представления данных и их анализа был подготовлен статический дашборд (dashboard от англ. — это интерактивная аналитическая панель) со статистикой поступавших, поступивших и не поступивших.

Для начала изучим данные поступающих. Как видно из 1 диаграммы, на 1 место в ВУЗе конкурс примерно 2 человека. Это невысокий конкурс, что говорит о невысокой сложности поступления абитуриента. Возможно, стоит пересмотреть количество мест на факультетах, так как существует повышенный спрос. Самыми популярными экзаменами являются обществознание, география, иностранный язык. Все экзамены по выбору популярны для сдачи, так что не требуется их пересмотр для поступления с точки зрения количества набора.

Dashboard

Статистика поступавших

Поступление

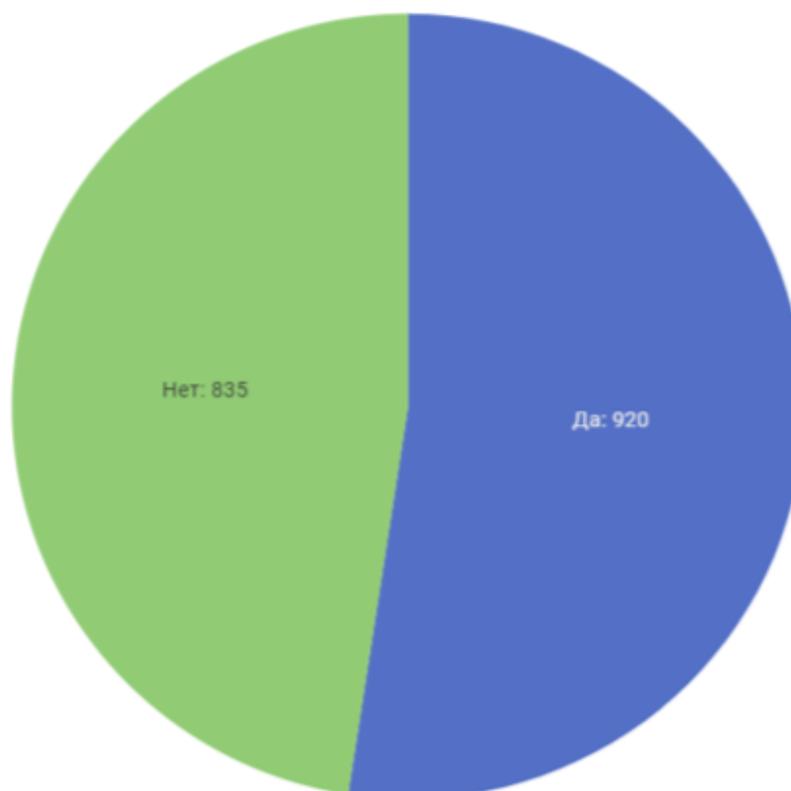


Рисунок 1 – Результаты поступавших.

Далее рассмотрим статистику по поступившим абитуриентам. Большинство из них - представители районов края, при этом больше всего из Динского, Кавказского, Кореновского районов и Усть-Лабинска. Стоит задуматься о сотрудничестве со школами, привлекая потенциально одаренную молодежь в ВУЗ. Самые распространенные экзамены среди них: география, иностранный язык, но отрыв от других небольшой, так что нельзя говорить о простоте сдачи данных испытаний. Либо все они одинаково несложные, что с большой вероятностью противоречит логике.

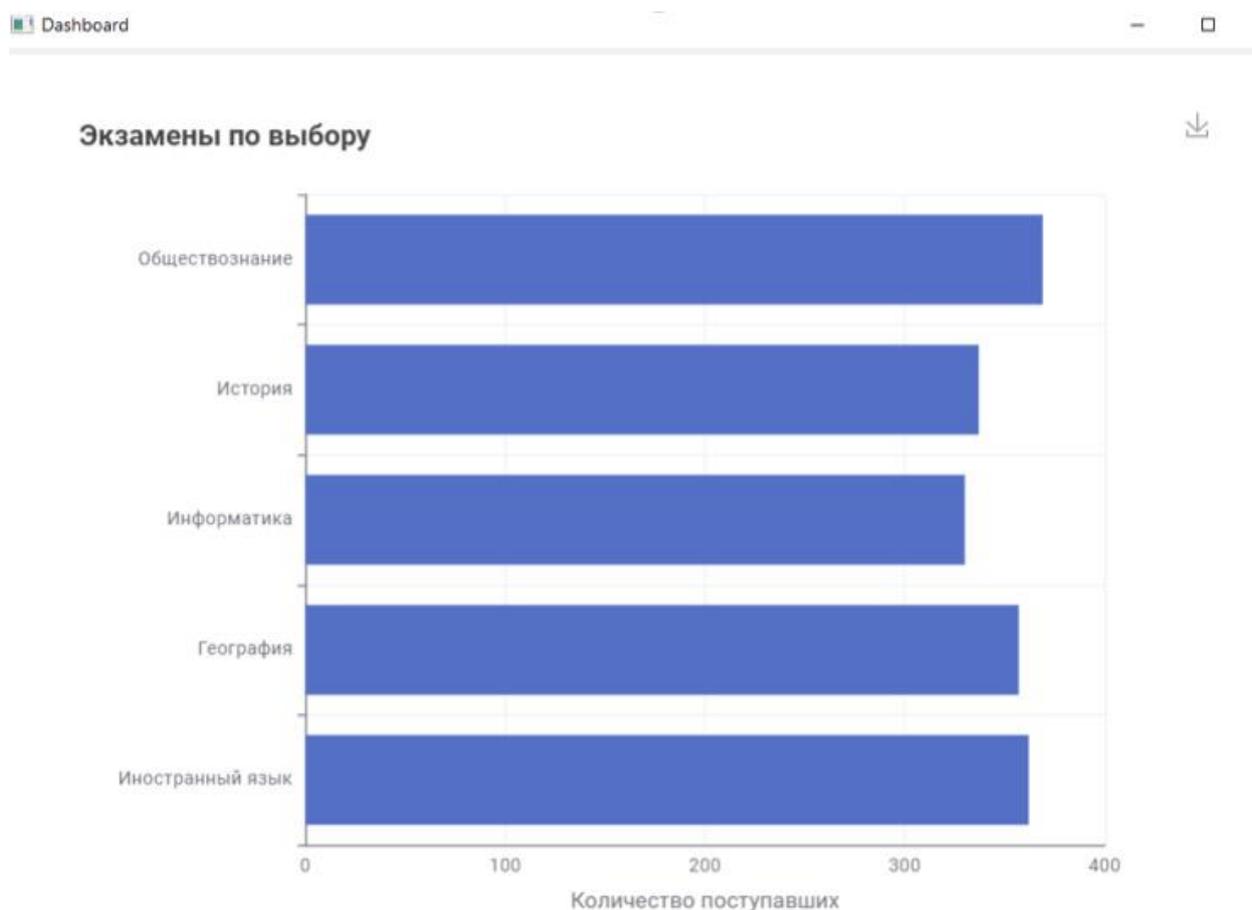


Рисунок 2 – Экзамены по выбору поступающих

Рассматривая статистику не поступивших, приходим к выводу, что также больше всего не поступило из районов. Наибольшее количество провалившихся дали Лабинск, Туапсе, Ленинградский район. Популярный экзамен среди них – обществознание, с большим отрывом, что может говорить о сложности экзамена, следует обратить внимание и пересмотреть испытание, также это говорит о слабой подготовке к экзамену.

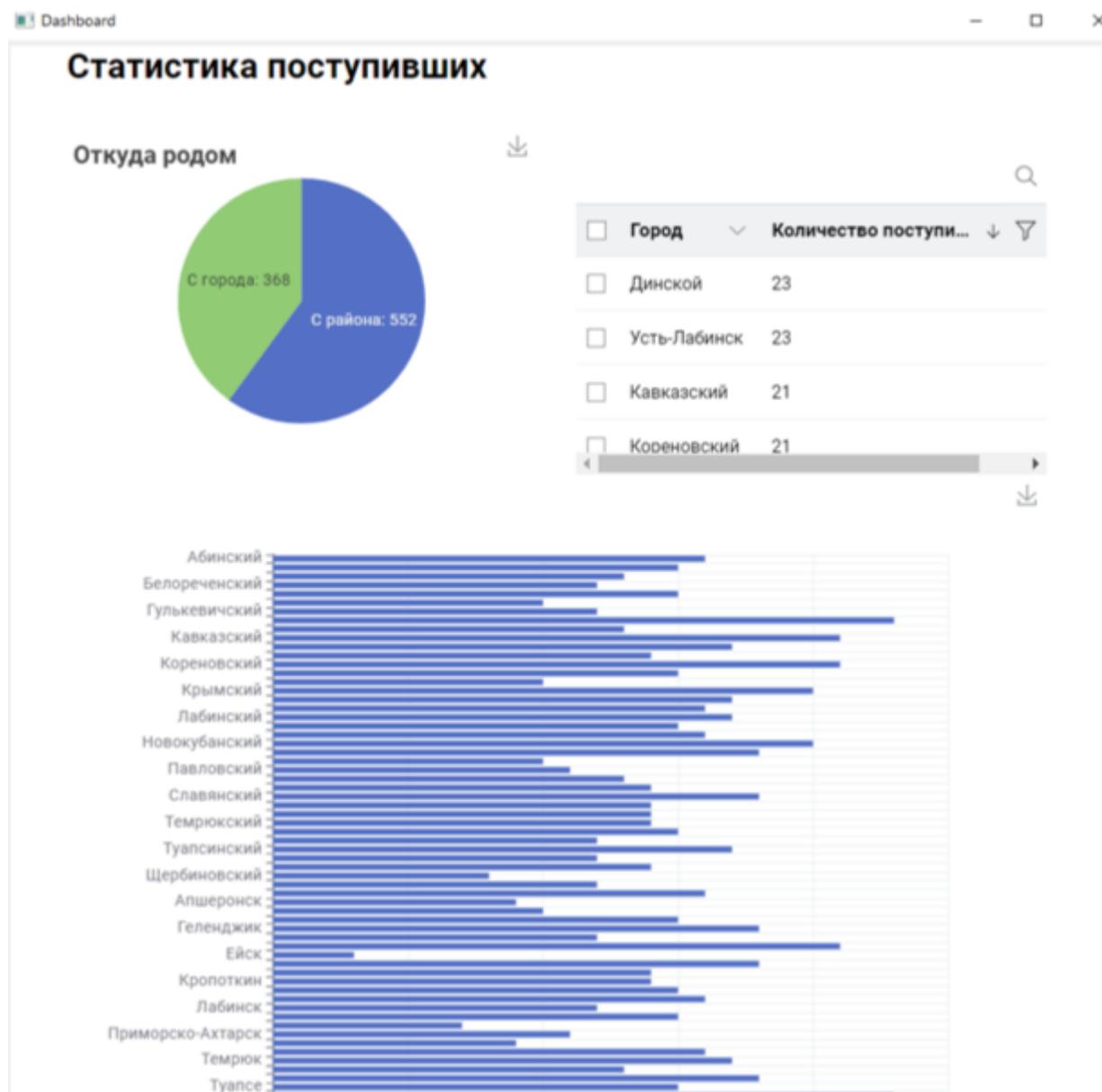


Рисунок 3 – Статистика поступивших

Подытоживая статистику, хочется отметить, что ВУЗ больше популярен у людей с района, чем с города. Это может связано как с рейтингом образовательного учреждения в крае, так и осведомленности абитуриентов о данном заведении.

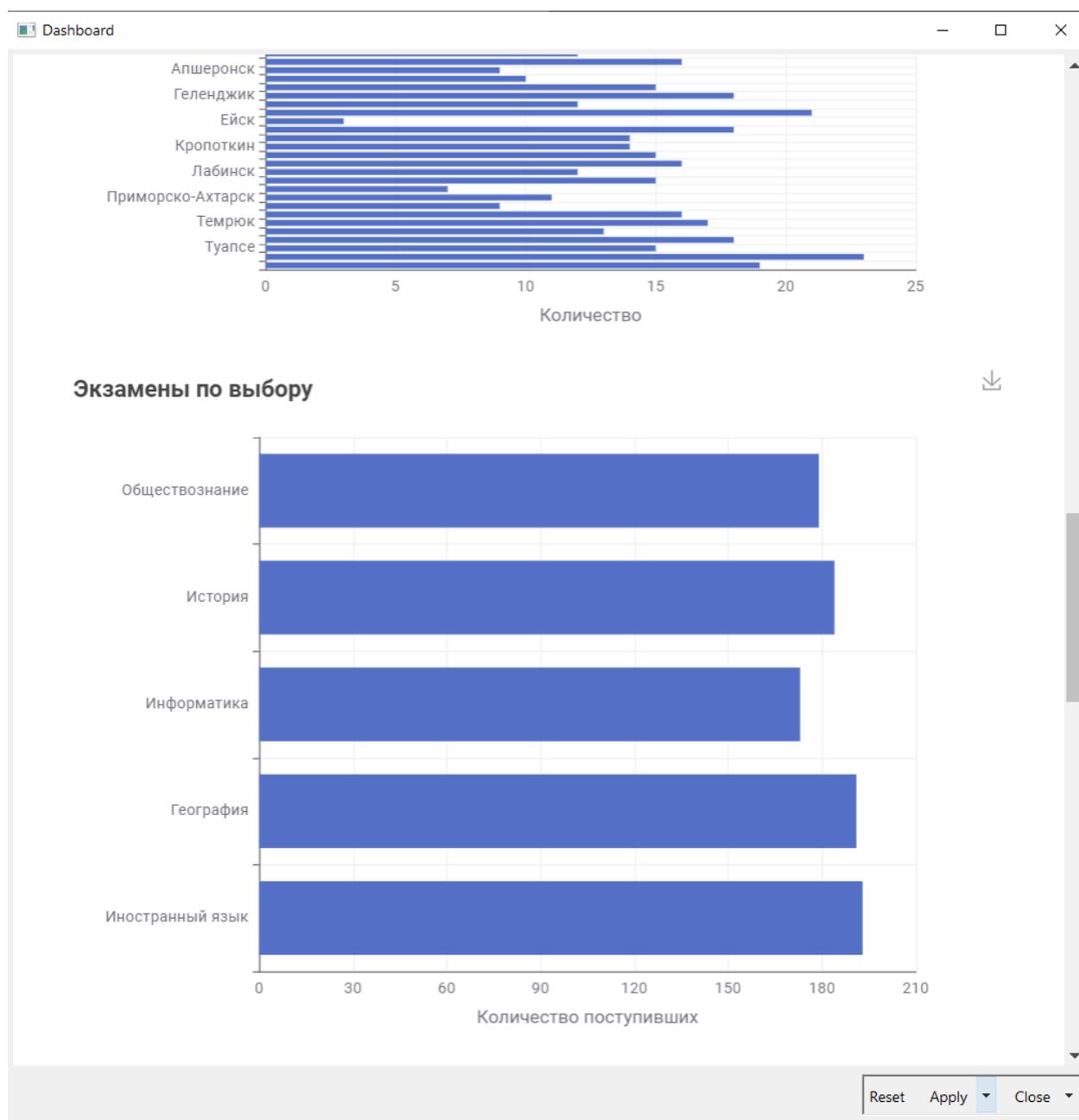


Рисунок 4 – Экзамены по выбору поступивших

В первом случае стоит обратить внимание на внутреннюю политику заведения, стараясь повысить рейтинг, во втором случае на внешнюю, можно сотрудничать со школами, проводить мероприятия для повышения известности ВУЗа.

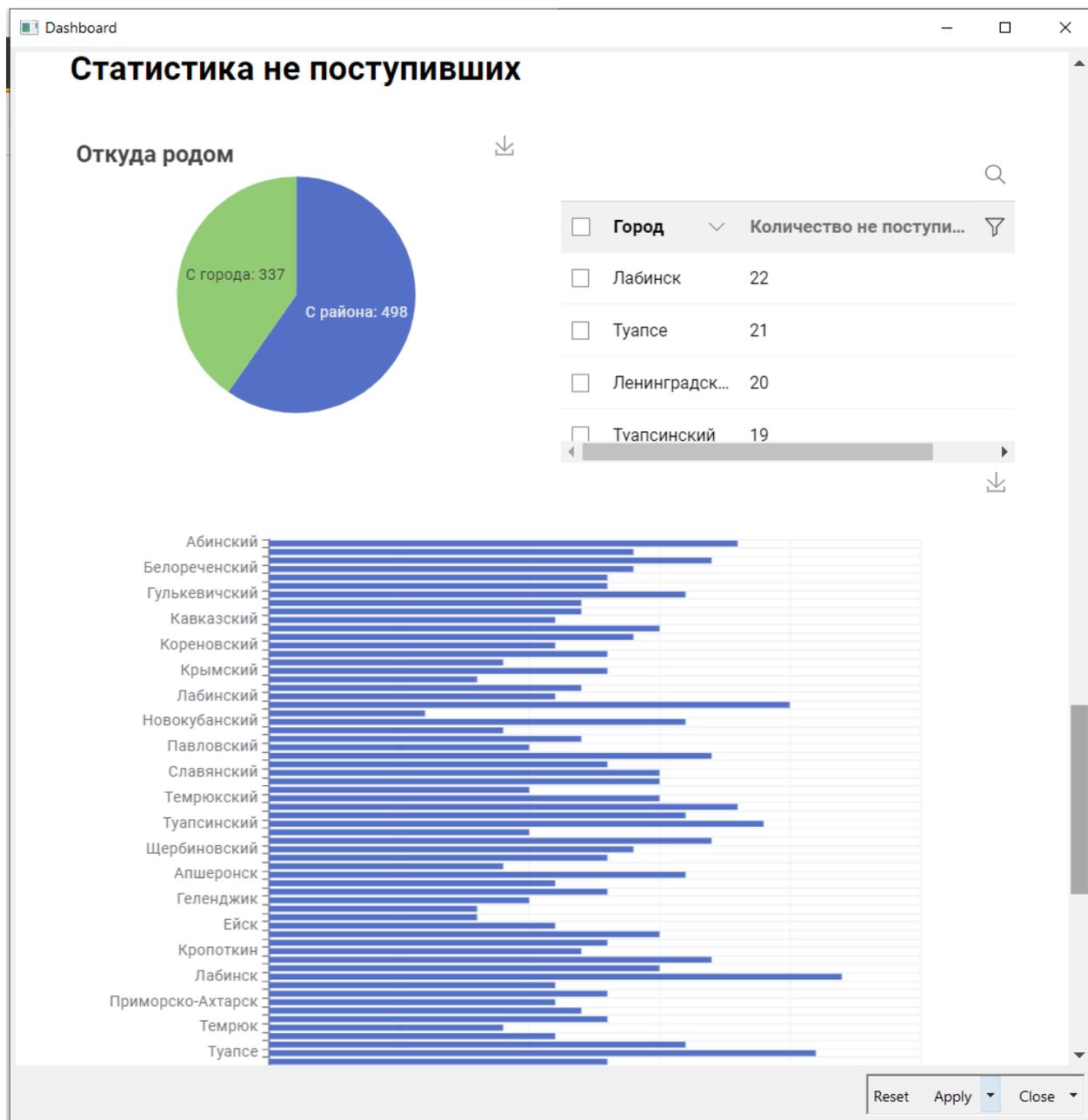


Рисунок 5 – Статистика не поступивших

В программном обеспечении Deductor Studio имеется очень интересный инструмент кластеризации, в результате которого получаются компактные и удобные для интерпретации и анализа карты двумерные карты. Основными преимуществами являются: устойчивость к зашумленным данным, быстрое обучение без учителя (результат зависит

только от структуры входных данных), визуализация многомерных входных данных.

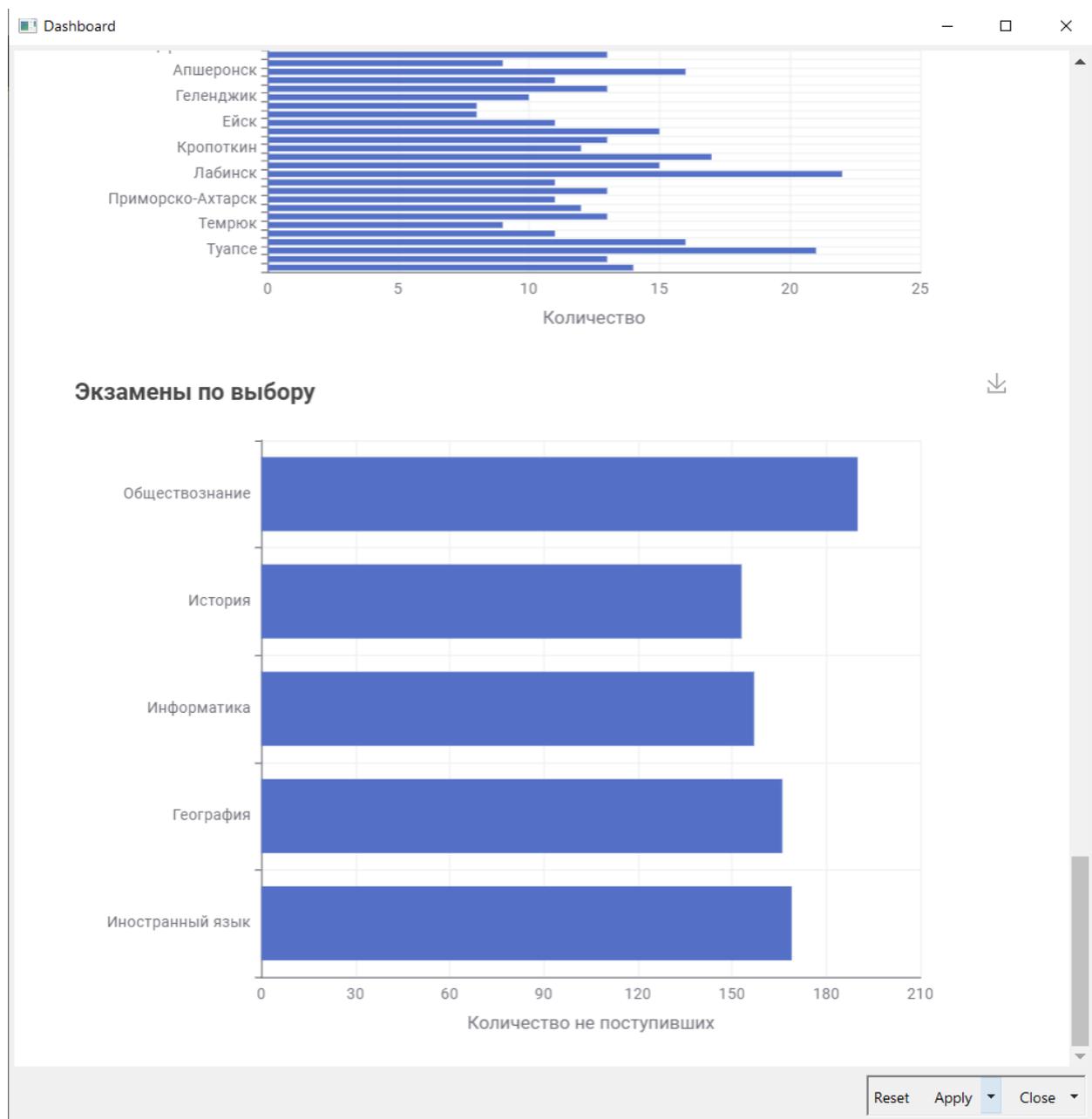


Рисунок 6 – Экзамены по выбору не поступивших

С помощью карт Кохонена можно оценить большой массив данных, улучшить понимание его структуры, найти закономерности. В данном случае, алгоритм делит абитуриентов на группы по схожим признакам.

Кластером будет являться группа векторов, расстояние между которыми внутри группы меньше, чем расстояние до соседних групп.

Каждая шестиугольная ячейка карты несет в себе определенное значение определенного поля. Для наглядности они окрашены, вариацию цвета в зависимости от значения можно увидеть на горизонтальной полосе под картой. Точное значение будет выделено на этой полосе. Выделив определенную ячейку в одной карте, автоматически подсветятся ячейки в других. Их совокупность представляет из себя группу записей. В нашем случае, данные о группе абитуриентов со схожими признаками.

Распределение на обучающее и тестовое множество 75% на 25%.
Размер карт 16x12. Количество кластеров 4.

Первая карта была построена с использованием полей город/район, где 0 – район, 1 – город, экзамены по выбору, достижения, подготовительные курсы, средний балл аттестата, отличие в аттестате. Выходными полями являются поступление, бюджет/коммерция, где 0 – непоступление, 1 – коммерция, 2 – бюджет. Настройки обработчика: распознавание примеров при ошибке меньше 0,4, длительность обучения 600 эпох. Способ инициализации карты: из обучающего множества, скорость обучения в начале 1, в конце 0,5 с радиусом обучения в начале 2, в конце 0,01.

Рассмотрев внимательно карту «Кластеры», можно сказать, что:

- 0 кластер – «городские» поступающие без подготовительных курсов;
- 1 кластер – поступающие из районов края без подготовительных курсов;
- 2 кластер – «городские» поступающие, проходившие подготовительные курсы;

– 3 кластер – поступающие с района, проходившие подготовительные курсы.

На рисунке ниже можно увидеть данные абитуриентов, которые поступили на коммерцию. Они «городские», сдали историю на 77-78 баллов, имеют достижения, средний балл 4,07, проходили обучение на подготовительных курсах. По итогам работы алгоритма он был определен во 2 кластер.

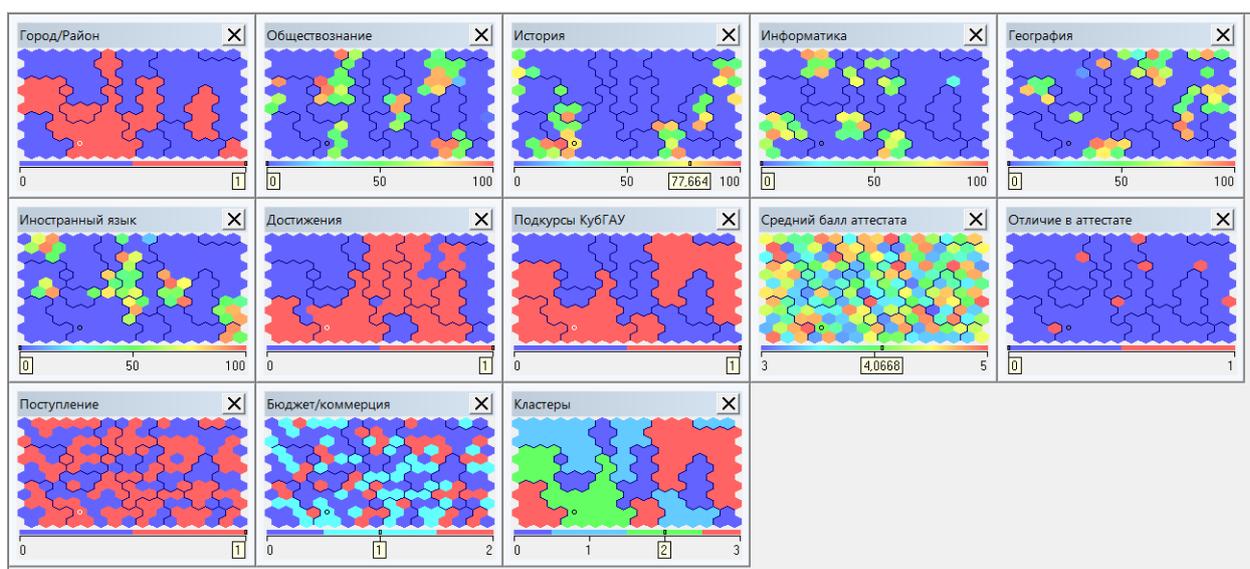


Рисунок 7 – Карты Кохонена, общее (1 группа).

Переместив точку в другое место, можно наблюдать данные по другой группе абитуриентов. Они родом из районов, сдавали информатику на 45-46 баллов, не имеют достижений, но проходили подготовительные курсы. Средний балл аттестата высокий – 4,81. Как итог, они не поступили. Алгоритм отнес их к 3 кластеру.

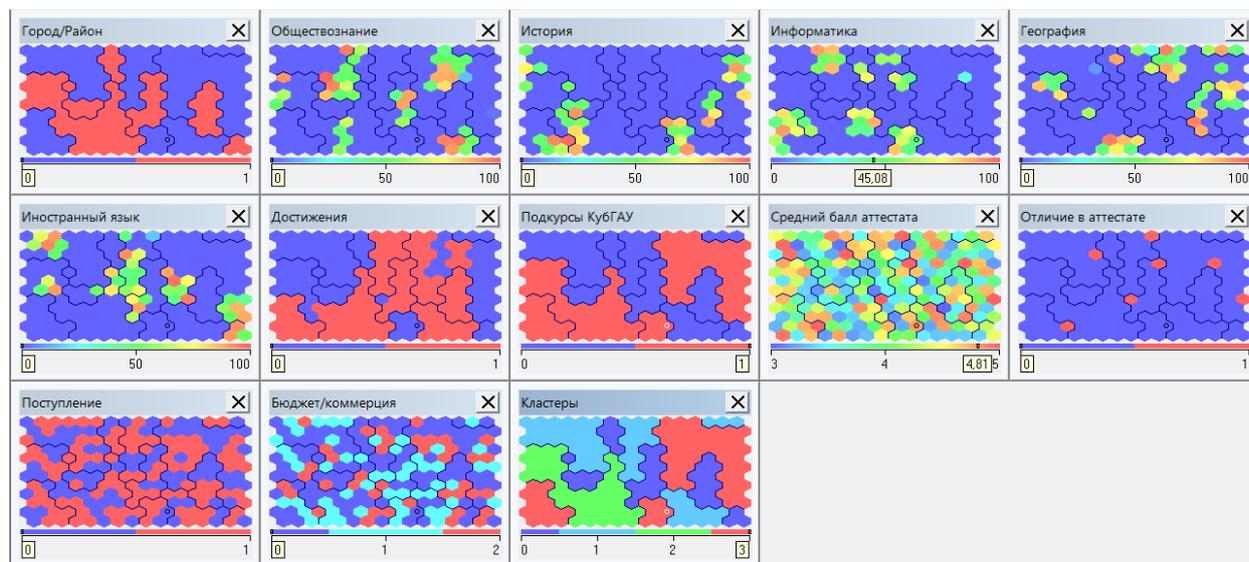


Рисунок 8 – Карты Кохонена, общее (2 группа).

Следующая карта была построена с использованием полей город/район, где 0 – район, 1 – город, экзамены по выбору, сумма баллов, подготовительные курсы. Выходными полями являются поступление, бюджет/коммерция, где 0 – непоступление, 1 – коммерция, 2 – бюджет. Настройки обработчика: распознавание примеров при ошибке меньше 0,4, длительность обучения 600 эпох. Способ инициализации карты: из собственных векторов, скорость обучения в начале 1, в конце 1 с радиусом обучения в начале 4, в конце 0,01.

Здесь тяжелее заметить четкую связь 2 факторов на формировании кластеров, явно задействованы экзамены по выбору в итоговом решении алгоритма кластеризации.

Распределение баллов среди «городских» обучающихся и с районов одинаково, при этом подготовительными курсами были заинтересованы больше первые, чем вторые. Это может быть следствием их очного проведения, можно задуматься о дистанционном формате проведения занятий.

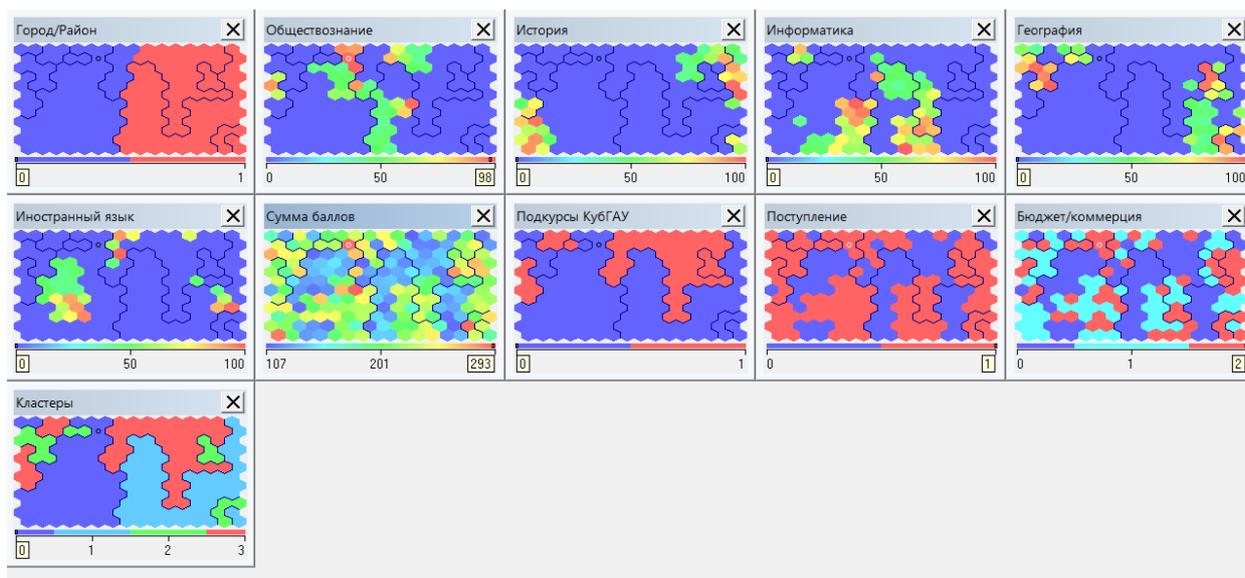


Рисунок 9 – Карты Кохонена: экзамены и подготовительные курсы по району/городу.

Последняя карта была построена с использованием полей город/район, где 0 – район, 1 – город, сумма баллов, достижения, средний балл аттестата, отличие в аттестате. Выходными полями являются поступление, бюджет/коммерция, где 0 – непоступление, 1 – коммерция, 2 – бюджет. Настройки обработчика: распознавание примеров при ошибке меньше 0,3, длительность обучения 600 эпох. Способ инициализации карты: из обучающего множества, скорость обучения в начале 1, в конце 1 с радиусом обучения в начале 4, в конце 0,01.

В данном случае были выделены следующие кластеры:

- 0 – «городские» абитуриенты без достижений;
- 1 – «городские» абитуриенты с достижениями;
- 2 – абитуриенты с района без достижений;
- 3 – абитуриенты с района с достижениями.

Рассматривая карты, можно сказать, что средний балл выше у школьников с города. Достижения чаще встречаются также у них. Поступают на бюджет и коммерцию как «городские», так и с районов

равнозначно. Отличие в аттестате встречается только у 1 и 2 кластера. 0 кластер чаще всего имеет высокий средний балл.

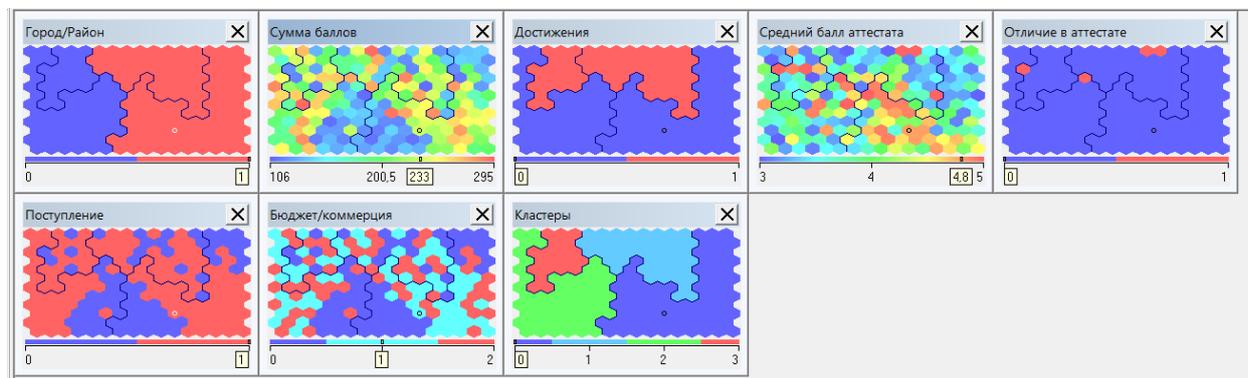


Рисунок 10 – Карты Кохонена: достижения по району/городу.

Корреляционный анализ – статистический метод, помогающий выявить взаимосвязь между двумя и более переменными, оценить силу и направление. Это позволяет определить факторы, влияющие на исследуемые явления и предсказать поведение переменных. Такой анализ основан на понятии корреляции, которая описывает степень зависимости между объектами. Основные методы: расчет коэффициента корреляции и построение корреляционной матрицы.

Коэффициент корреляции позволяет выяснить силу линейной зависимости. Его значения варьируются от -1 (полная отрицательная корреляция) до 1 (полная положительная корреляция). Значение 0 указывает на отсутствие взаимосвязи, положительное отражает прямую зависимость, а отрицательное – обратную. То есть положительное значение говорит о том, что при увеличении значений одной переменной, значения взаимосвязанной также увеличатся, а отрицательное – при увеличении значений одной, значения другой уменьшатся.

Корреляционная матрица представляет собой таблицу, в которой каждая ячейка содержит коэффициент корреляции между

соответствующими переменными. Она позволяет оценить не только связь между парами, но и между несколькими факторами одновременно.

В данном примере изучается влияние таких факторов, как сумма баллов, средний балл аттестата, отличие в аттестате, подготовительных курсов и других достижений на поступление, поступление на бюджет или коммерцию и на желаемый факультет.

Для начала исследуем зависимости между отдельными факторами. Исходя из полученных значений, отраженных в таблице, можно сделать вывод, что существует сильная зависимости между поступлениями и суммой баллов. Интересна отрицательная слабая взаимосвязь между средним баллом аттестата и поступлениями. Это может говорить о том, что в ВУЗ поступают абитуриенты, которые впоследствии могут быть отчислены из-за плохой успеваемости.

Между поступлениями и подготовительными курсами связь ничтожно мала, что говорит о важности пересмотра, изменения подхода к подготовке абитуриентов. У достижений и отличием в аттестате также нет влияния на поступление. Это впоследствии может отразиться на слабой активности учащихся во внеучебное время. Также выводы о взаимосвязи подтверждают корреляционные матрицы, отражающие цветом влияние факторов друг на друга.

Фактор	Сумма баллов	Средний балл аттестата	Отличие в аттестате	Подкурсы КубГАУ	Достижения
String	Number (double)	Number (double)	Number (double)	Number (double)	Number (double)
Поступление	0.933	-0.044	-0.016	-0.004	0.007
Бюджет/коммерция	0.848	-0.033	0.013	0.007	-0.001
Поступление на желаемый факультет	0.529	-0.061	-0.003	0.038	0

Рисунок 11 – Коэффициенты корреляции факторов.

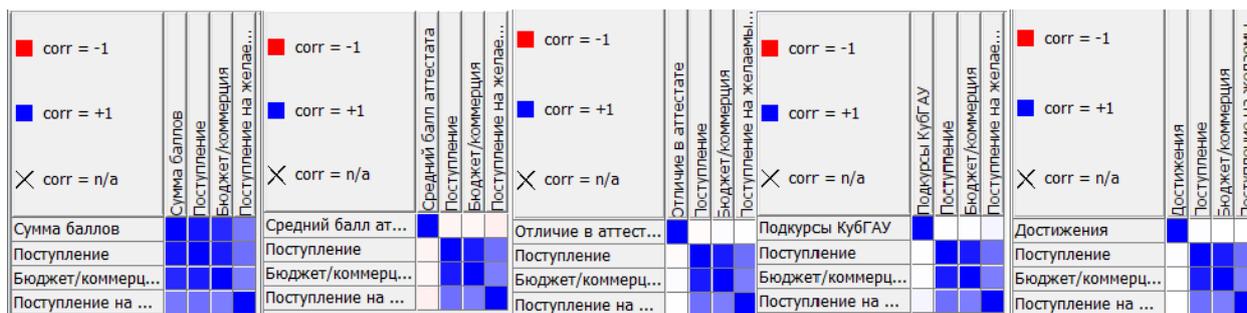


Рисунок 12 – Корреляционные матрицы факторов.

Далее рассмотрим влияние пар факторов на результат. Исходя из таблицы коэффициентов корреляции ниже, можно сказать, что наличие подготовительных курсов в купе с другими либо не оказывает влияния, либо имеет противоположный эффект на желаемый результат, что еще раз доказывает важность пересмотра программы обучения перед поступлением. Наличие отличия в аттестате и других достижений увеличивает шансы поступления на бюджет, однако в ВУЗ поступает меньшее число будущих студентов с предположительно положительной успеваемостью, чем следовало бы. Для наглядности также приведены матрицы корреляций.

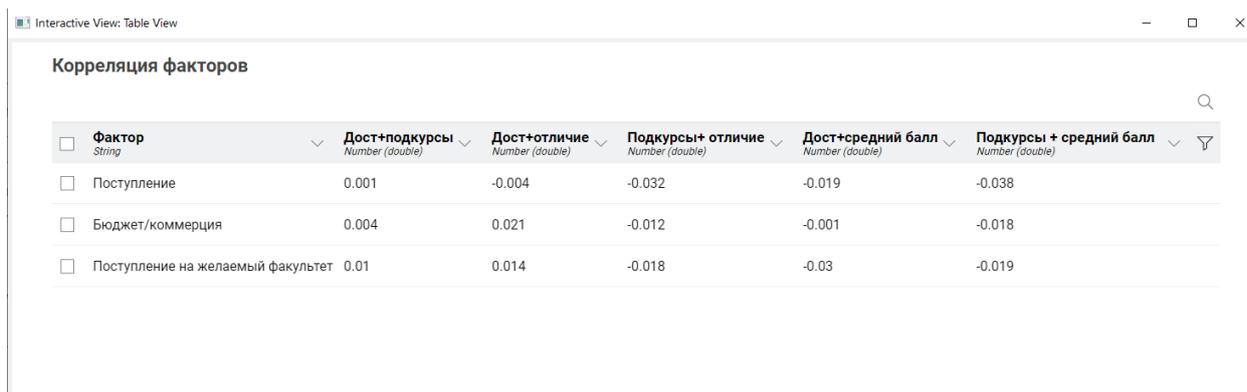


Рисунок 13 – Коэффициенты корреляции пар факторов.

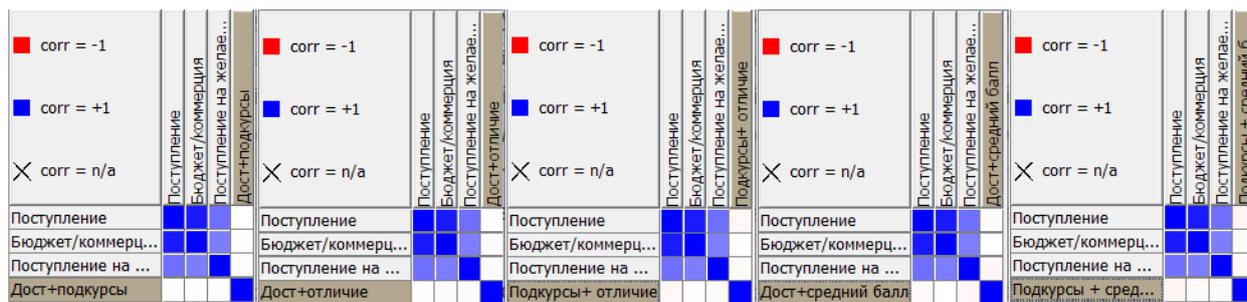


Рисунок 14 – Матрицы корреляций пар факторов.

Также были проанализированы влияние троек факторов на поступление абитуриентов. Данные отражены в таблице ниже, а также приведены корреляционные матрицы. Как можно увидеть из данных, отличники поступают хуже, ВУЗу стоит задуматься о том, как они могут привлечь к себе больше потенциально одаренной молодежи.

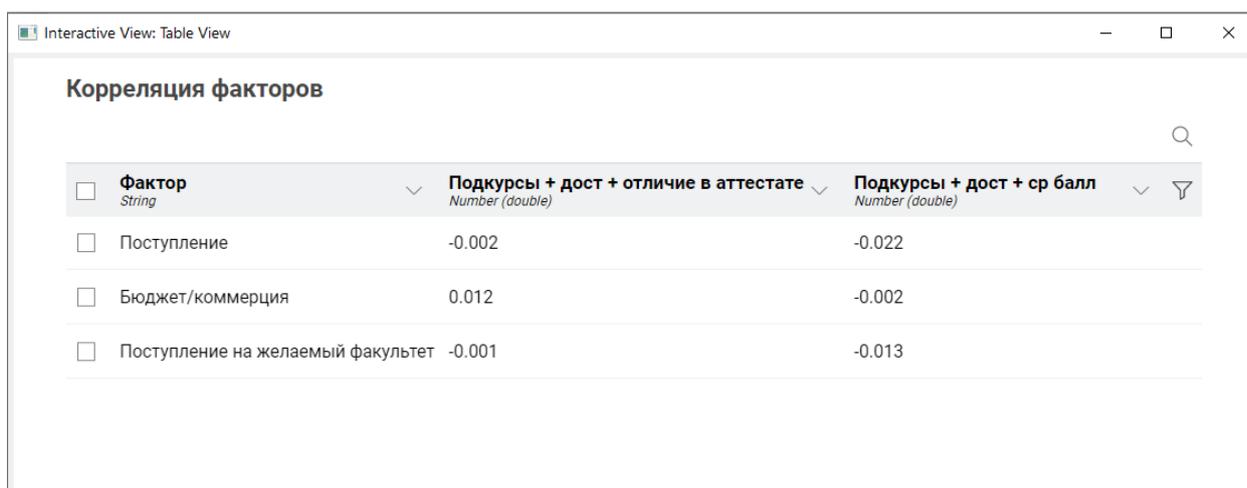


Рисунок 15 – Коэффициенты корреляции троек факторов.

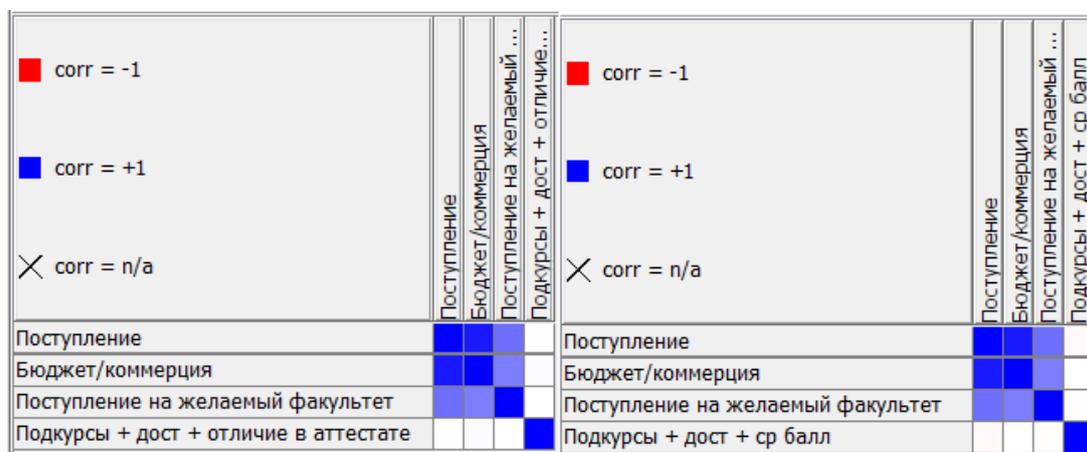


Рисунок 16 – Матрицы корреляций троек факторов

P-значение – это статистическая величина, отражающая вероятность получить результаты, схожие по значимости с теми, которые замечены в исследовании. При небольшом значении данной величины маловероятна случайность полученных результатов, что дает возможность отвергнуть нулевую гипотезу (отсутствие зависимости между переменными). Она будет использоваться в качестве объективной оценки взаимосвязи между факторами. Влияющим, или неслучайным, будет фактор, р-значение взаимосвязи которого меньше 0,07. Полученные результаты р-значений для факторов отражены в таблице ниже.

Исходя из значений таблицы, можно отметить, что реальная зависимость существует между парами: поступление-сумма баллов, бюджет/коммерция-сумма баллов, поступление на желаемый факультет-сумма баллов, средний балл аттестата-поступление, средний балл аттестата-поступление на желаемый факультет.

Таблица 1 – P-значения для взаимосвязи между факторами

	Поступление	Бюджет/коммерция	Поступление на желаемый факультет
Сумма баллов	0	0	0
Средний балл аттестата	0,065	0,163	0,01
Отличие в аттестате	0,515	0,597	0,907
Подкурсы КубГАУ	0,868	0,779	0,112
Достижения	0,766	0,962	0,994
Дост + подкурсы	0,967	0,879	0,667
Дост + отличие	0,866	0,371	0,564
Подкурсы + отличие	0,184	0,619	0,458
Дост + средний балл	0,42	0,968	0,208
Подкурсы + средний балл	0,112	0,446	0,42
Подкурсы + дост + отличие в аттестате	0,923	0,606	0,962
Подкурсы + дост + средний балл	0,365	0,95	0,591

Далее следует изучить, как будет изменяться переменная (поступление) на основании другой (сумма баллов). Для этого проводится регрессионный анализ – статистический метод выявления и количественной оценки зависимости между переменными. Он помогает понять, как значение зависимой переменной изменяется при изменении значения независимой, где зависимая – переменная для анализа или прогноза (результат) и независимая – переменные, которые влияют на зависимую, используемые для объяснения изменений. Такой анализ применяется для предсказания, классификации и прогнозирования. Целью регрессионного анализа является определение наиболее подходящей линии или кривой, которая отражает связь между независимыми и зависимой переменными.

В данном случае используется одиночная линейная регрессия, отражающая связь между парой переменных. Наилучшее соответствие представляется графиком $y = a + bx$. На основании подобранных значений коэффициент по обучающей выборке сформирован до 2023 года. Был сделан прогноз для 2023 года, результаты которого отражены на рисунке 17. Слева находится реальный исход, а справа спрогнозированное значение Поступившим считается человек, у которого значение прогноза выше 0,66.

#	Row...	Поступление Number (integer)	Prediction (Поступление) Number (double)
1	Row0	0	-0.066
2	Row1	1	1.062
3	Row2	0	0.127
4	Row3	1	1.079
5	Row4	0	0.092
6	Row5	0	0.066
7	Row6	0	0.075
8	Row7	0	0.119
9	Row8	0	0.092
10	Row9	0	0.136
11	Row10	1	0.709
12	Row11	1	0.771
13	Row12	1	1.229
14	Row13	1	1.141
15	Row14	0	0.057
16	Row15	0	0.039
17	Row16	0	0.039

Рисунок 17 – Результаты прогноза для записей за 2023

Чтобы применять полученную модель на практике, следует оценить ее качество. Для этого существует оценка качества регрессии, а также оценка качества классификации. В результате получаем таблицу, отраженную на рисунке ниже. Это матрица ошибок, которая показывает, где полученная на основе регрессии модель справилась, а где нет.

Слева находятся реальные значения, а сверху отклик построенной регрессионной модели. Если реальный результат положительный (поступит) и модель предсказала поступление, то это истинно положительные значения, если предсказано непоступление, то – ложно отрицательный. Если реальность – «не поступит» и решение модели «не поступит», то истинно отрицательный, если решение «поступит» – ложно положительный.

Исходя из таблицы, можно сказать, что обученная модель лучше справляется с предсказанием отрицательных исходов, что логично. Все ошибки приходятся на предсказание положительного результата, а общая точность модели составила 94,3%.

RowID	Не поступит <i>Number (integer)</i>	Поступит <i>Number (integer)</i>
Не поступит	835	0
Поступит	100	820

Рисунок 18 – Матрица ошибок

Также ниже приведена таблица показателей для оценки, насколько повышение сложности модели увеличит ее точность. Сюда входят следующие меры.

– Коэффициент детерминации измеряет долю дисперсии, объясненную моделью, в общей дисперсии целевой переменной. Если значение близко к единице, то модель хорошо объясняет данные.

– Средняя абсолютная ошибка отражает допущение моделью больших ошибок прогноза. Данная величина стремится к нулю, чем ее значение меньше, тем модель меньше допускает грубых ошибок.

– Средняя квадратичная ошибка гораздо сильнее штрафует функцию за большие отклонения, чем предыдущий показатель, поэтому более чувствительна к выбросам. Ошибки наиболее заметны за счет квадратичной зависимости.

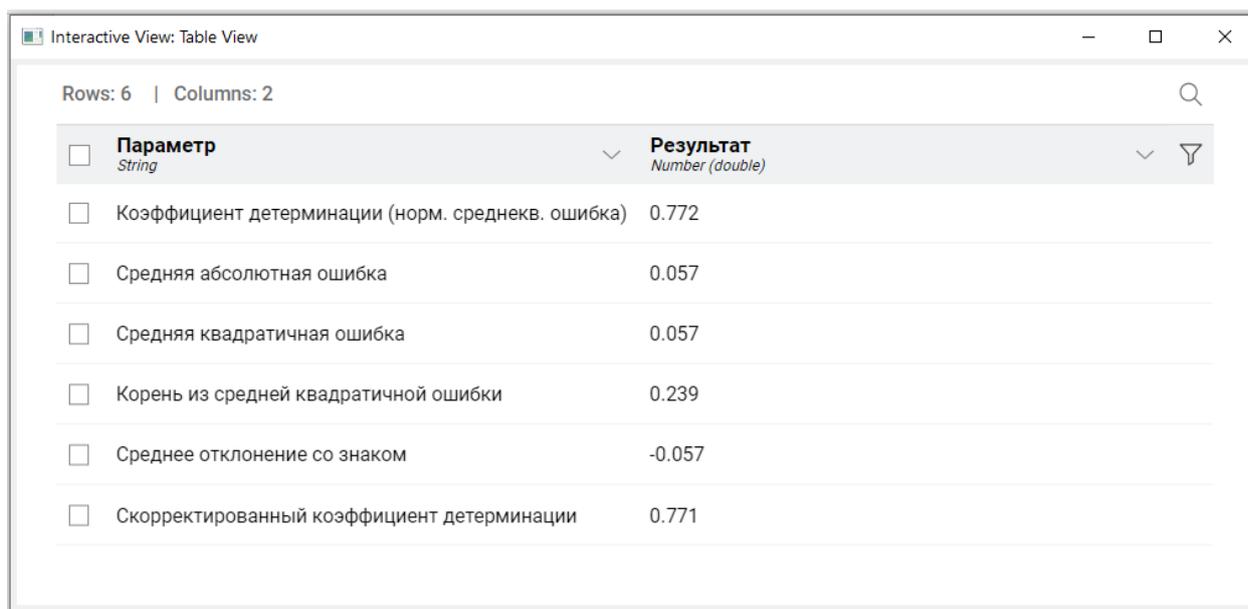
– Корень из средней квадратичной ошибки более чувствителен к аномальным значениям.

– Среднее отклонение со знаком вычисляет среднее различие между фактическим и предсказанным значениями.

– Скорректированный коэффициент детерминации является преодолением проблемы использования коэффициента детерминации, который увеличивается при добавлении в обучающую выборку новых переменных. Он дает возможность сравнивать модели с разным числом переменных так, чтобы они не влияли на коэффициент.

Исходя из полученных значений в таблице, можно сказать, что полученная модель объясняет данные на 77,2%, ошибаясь в своих предсказаниях в меньшую сторону от реального поступления (ложно отрицательные результаты), допуская 5,7% грубых ошибок.

Эти результаты впоследствии могут быть применены для подбора лучшей регрессии, однако стоит опасаться переобучения (ошибка крайне близка или равна 0), когда модель полностью описывает исходные данные, так как это приведет к повышению грубых ошибок в предсказаниях. Таким образом, полученная регрессия является подходящей для наших данных.



Параметр <i>String</i>	Результат <i>Number (double)</i>
Кoeffициент детерминации (норм. среднекв. ошибка)	0.772
Средняя абсолютная ошибка	0.057
Средняя квадратичная ошибка	0.057
Корень из средней квадратичной ошибки	0.239
Среднее отклонение со знаком	-0.057
Скорректированный коэффициент детерминации	0.771

Рисунок 19 – Оценка качества подобранной регрессии.

Вывод.

Рейтинг ВУЗов основывается как на качестве и разнообразии преподавания (преподавательский состав, количество направлений подготовки, их содержание), так и на своих студентах, которые являются лицом образовательного учреждения. Их жизненная позиция, активность напрямую отражает подход университета к развитию не только специалистов, но и личностей. От этого зависят известность, репутация в глазах школьников, будущих студентов, так и финансирование, дающее более качественное образование.

Для того, чтобы улучшить понимание образа бывших и будущих абитуриентов и провести ряд решений по его улучшению, заинтересовывать больше потенциально одаренной молодежи, могут быть применены технологии Big Data.

Таким образом, на сгенерированном наборе данных были продемонстрированы возможности анализа с применением методов Big Data, основанных на визуализации и графиках, математике и статистике, искусственном интеллекте и машинном обучении. Они могут помочь

высшим учебным заведениям России принимать решения, основываясь на данных и фактах.

При использовании реальных данных, применяя методы регрессионного, корреляционного анализов, самоорганизующихся карт Кохонена и визуализации ВУЗ может качественно повлиять не только на свой рейтинг, но и на имидж:

- проанализировать и улучшить работу внутренних процессов;
- построить образ абитуриента, будущего студента;
- принять решения о взаимодействии со школами районов и городов, богатыми потенциально одаренной молодежью, проведение мероприятий по продвижению и популяризации ВУЗа.

Список литературы

1. Пенкина Ю. Н. Адаптивная система поддержки принятия оперативных решений в управлении IT-проектами / Ю. Н. Пенкина, А. В. Параскевов // Научное обеспечение агропромышленного комплекса, сборник статей по материалам 71-й научно-практической конференции студентов по итогам НИР за 2015 год. Краснодар – Кубанский государственный аграрный университет имени И. Т. Трубилина, 2016.

2. Параскевов А. В. Разработка системы принятия решений при работе с динамическими показателями / А. В. Параскевов, А. Ф. Алексеев // Математические методы и информационно-технические средства. Материалы X Всероссийской научно-практической конференции, Краснодар, 2014.

3. Параскевов А. В. Оценка компетенций бакалавра: возможное решение проблем / А. В. Параскевов, Г. О. Монин // статья в открытом архиве www.researchgate.net, 10.13140/RG.2.2.26177.58721, 2020.

4. Параскевов А. В. Необходимость внедрения информационных технологий / А. В. Параскевов, Д. А. Махлушев, А. А. Ахлѣстова // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2022. – №09(183). С. 212 – 222. – IDA [article ID]: 1832209021. – Режим доступа: <http://ej.kubagro.ru/2022/09/pdf/21.pdf>, 0,688 у.п.л.

5. Параскевов А. В. Анализ опыта внедрения информационных систем в образовательную деятельность / А. В. Параскевов // Точки научного роста: на старте десятилетия науки и технологии. Материалы ежегодной научно-практической конференции преподавателей по итогам НИР за 2022г. Краснодар, 2023.

References

1. Penkina Ju. N. Adaptivnaja sistema podderzhki prinjatija operativnyh reshenij v upravlenii IT-proektami / Ju. N. Penkina, A. V. Paraskevov // Nauchnoe obespechenie agropromyshlennogo kompleksa, sbornik statej po materialam 71-j nauchno-prakticheskoj konferencii studentov po itogam NIR za 2015 god. Krasnodar – Kubanskij gosudarstvennyj agrarnyj universitet imeni I. T. Trubilina, 2016.
2. Paraskevov A. V. Razrabotka sistemy prinjatija reshenij pri rabote s dinamicheskimi pokazateljami / A. V. Paraskevov, A. F. Alekseev // Matematicheskie metody i informacionno-tehnicheskie sredstva. Materialy X Vserossijskoj nauchno-prakticheskoj konferencii, Krasnodar, 2014.
3. Paraskevov A. V. Ocenka kompetencij bakalavra: vozmozhnoe reshenie problem / A. V. Paraskevov, G. O. Monin // stat'ja v otkrytom arhive www.researchgate.net, 10.13140/RG.2.2.26177.58721, 2020.
4. Paraskevov A. V. Neobhodimost' vnedrenija informacionnyh tehnologij / A. V. Paraskevov, D. A. Mahlushev, A. A. Ahljostova // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2022. – №09(183). S. 212 – 222. – IDA [article ID]: 1832209021. – Rezhim dostupa: <http://ej.kubagro.ru/2022/09/pdf/21.pdf>, 0,688 u.p.l.
5. Paraskevov A. V. Analiz opyta vnedrenija informacionnyh sistem v obrazovatel'nuju dejatel'nost' / A. V. Paraskevov // Tochki nauchnogo rosta: na starte desjatiletija nauki i tehnologii. Materialy ezhegodnoj nauchno-prakticheskoj konferencii prepodavatelej po itogam NIR za 2022g. Krasnodar, 2023.