

УДК 004.891.3

ББК 74.58

DOI 10.51955/2312-1327_2023_3_158

ПРИМЕНЕНИЕ ИЕРАРХИЧЕСКОГО КЛАСТЕРНОГО АНАЛИЗА ПРИ ПОДГОТОВКЕ АВИАЦИОННЫХ СПЕЦИАЛИСТОВ

*Денис Александрович Евсевичев,
orcid.org/0000-0003-2234-427X,
кандидат технических наук, доцент
Ульяновский институт гражданской авиации
имени Главного маршала авиации Б. П. Бугаева,
ул. Можайского, 8/8
Ульяновск, 432071, Россия
denistk_87@mail.ru*

*Евгений Александрович Костиков,
orcid.org/0000-0001-8119-4676,
аспирант
Ульяновский институт гражданской авиации
имени Главного маршала авиации Б. П. Бугаева,
ул. Можайского, 8/8
Ульяновск, 432071, Россия
kostikov_128@mail.ru*

*Юрий Владимирович Штырлов,
orcid.org/0000-0002-6010-4720,
аспирант
Ульяновский институт гражданской авиации
имени Главного маршала авиации Б. П. Бугаева,
ул. Можайского, 8/8
Ульяновск, 432071, Россия
yura.shtyrlov@mail.ru*

*Екатерина Сергеевна Алякина,
orcid.org/0000-0003-1791-4362,
аспирант
Ульяновский институт гражданской авиации
имени Главного маршала авиации Б. П. Бугаева,
ул. Можайского, 8/8
Ульяновск, 432071, Россия
ka1105@yandex.ru*

Аннотация. В статье рассматривается подход к оценке качества сформированности компетенций по осваиваемым студентом или курсантом учебного заведения гражданской авиации соответствующим дисциплинам. Оценка формируется за счет применения одного из методов машинного обучения без учителя – иерархического кластерного анализа. Сбор данных для оценки успеваемости курсантов высшего учебного заведения гражданской авиации осуществлялся при освоении дисциплины «Радиотехническое оборудование аэродромов». Экспериментальное исследование проходило в течение семестра. Тесты по дисциплине были сформированы в специальных google формах, что позволило упростить процесс сбора данных и обеспечить удобство контроля за выполнением тестов. Обработка данных и дальнейшая работа с ними осуществлялась в среде разработки Jupyter Notebook с

использованием высокоуровневого языка объектно-ориентированного программирования Python. В программе для осуществления кластеризации был использован метод `cluster.hierarchy.linkage` из библиотеки SciPy. Для графического определения оптимального количества кластеров, на которые следует делить выборку, был использован критерий каменистой осыпи Кеттела. Описанный подход позволяет выделять в отдельные группы (кластеры) обучающихся с целью автоматизации проверки освоения компетенций.

Ключевые слова: образование, компетенции, оценка, статистика, машинное обучение, кластеризация, кластерный анализ, дендрограмма, автоматизация, программа.

APPLICATION OF HIERARCHICAL CLUSTER ANALYSIS IN THE TRAINING OF AVIATION SPECIALISTS

*Denis A. Evsevichev,
orcid.org/0000-0003-2234-427X,
Candidate of Technical Sciences, Associate Professor
Ulyanovsk Civil Aviation Institute,
8/8, Mozhayskiy street
Ulyanovsk, 432071, Russia
denistk_87@mail.ru*

*Evgeniy A. Kostikov,
orcid.org/0000-0001-8119-4676,
Postgraduate Student
Ulyanovsk Civil Aviation Institute,
8/8, Mozhayskiy street
Ulyanovsk, 432071, Russia
kostikov_128@mail.ru*

*Yury V. Shtyrlov,
orcid.org/0000-0002-6010-4720,
Postgraduate Student
Ulyanovsk Civil Aviation Institute,
8/8, Mozhayskiy street
Ulyanovsk, 432071, Russia
yura.shtyrlov@mail.ru*

*Ekaterina S. Alyakina,
orcid.org/0000-0003-1791-4362,
Postgraduate Student
Ulyanovsk Civil Aviation Institute,
8/8, Mozhayskiy street
Ulyanovsk, 432071, Russia
ka1105@yandex.ru*

Abstract. The article discusses an approach to assessing the quality of the formation of competencies in the relevant disciplines mastered by a student or cadet of an educational institution of civil aviation. The assessment is formed by applying one of the unsupervised machine learning methods - hierarchical cluster analysis. Data collection to assess the cadets' performance of a higher educational institution of civil aviation was carried out during the development of the discipline «Radio equipment of airfields». The experimental study took place during the semester. Tests for the discipline were formed in special Google forms, which made it possible to simplify the process of data collection and provide convenient control over the execution of tests. Data processing and further work with them was carried out in the Jupyter Notebook development environment using the high-level object-oriented programming language Python. In the program for the implementation of clustering, the `cluster.hierarchy.linkage` method from the SciPy library was used. For a graphical

determination of the optimal number of clusters into which the sample should be divided, the Cattell's scree criterion was used. The described approach makes it possible to single out students into separate groups (clusters) in order to automate the verification of the development of competencies.

Keywords: education, competencies, grade, statistics, machine learning, clustering, cluster analysis, dendrogram, automation, program.

Введение

Проверка сформированности компетенций является важным этапом проверки освоения основной профессиональной образовательной программы. Данное утверждение является справедливым также для программ подготовки авиационных специалистов, что было отмечено, в том числе, в документах Международной организации гражданской авиации ИКАО [Doc 10056, 2017; Doc 10057, 2017; Doc 9868, 2016].

Для качественной оценки сформированности компетенций в результате освоенной студентом или курсантом учебной дисциплины формируется фонд оценочных средств, в который могут входить контрольные работы, рефераты, тесты и др. Из перечисленных разновидностей средств именно последнее – тестирование – обладает наибольшей объективностью оценивания, однако здесь есть свой недостаток: полученные результаты не дифференцируются в зависимости от сложности изучаемой темы, а также направления подготовки авиационного специалиста. Существует несколько подходов к разделению наборов данных, каковыми и являются результаты тестирования. Широкое применение во многих сферах жизни находит подход с использованием метода машинного обучения без учителя – кластеризация. Однако в системе образования чаще применяются неавтоматизированные средства оценивания, основанные прежде всего на мнении преподавателя. Предлагаемое в статье решение позволяет использовать кластеризацию для автоматизации оценки сформированности компетенций обучающимися учебных заведений гражданской авиации.

На основе вышеизложенного была сформулирована цель предлагаемой работы – автоматизация оценивания ответов обучающихся с учетом объективности компетентностного подхода. Для реализации поставленной цели следует решить задачи сбора и анализа данных, выбора и реализации метода кластерного анализа в программном продукте. В результате проведенного исследования был осуществлен сбор данных тестирования среди курсантов Ульяновского института гражданской авиации имени Главного маршала авиации Б. П. Бугаева при освоении дисциплины «Радиотехническое оборудование аэродромов». Обработка и анализ дифференцированности результатов выполнялся в отдельном программном продукте.

Материалы и методы

Материалом исследования послужили результаты прохождения курсантами Ульяновского института гражданской авиации имени Главного маршала авиации Б. П. Бугаева тестов после каждого лекционного занятия по дисциплине «Радиотехническое оборудование аэродромов». Выполнялись

тесты по пройденной теме и состояли из пяти вопросов. Экспериментальное исследование проходило в течение семестра. Список тем, рассматриваемых в тестах, включал в себя следующие:

- Тема 1. Физические основы радиолокации.
- Тема 2. Свойства радиоволн.
- Тема 3. Дальность действия радиосистем.
- Тема 4. Радиотехнические методы измерения координат.
- Тема 5. Обзорные радиолокационные станции.
- Тема 6. Специальные радиолокационные станции.
- Тема 7. Вторичные радиолокационные станции.
- Тема 8. Радиотехнические системы ближней навигации.
- Тема 9. Радиотехнические системы дальней навигации.
- Тема 10. Системы посадки.

Тесты были сформированы в специальных google формах, что позволило упростить процесс сбора данных и обеспечить удобство контроля за выполнением тестов. Вопросы в тестах включали как общие физические особенности функционирования средств радиотехнического обеспечения полетов и авиационной электросвязи, так и последовательности действий со стороны пользователя (пилота или диспетчера) за конкретной установкой.

Методы исследования: методы математического анализа, машинного обучения без учителя, математического программирования, теории оптимизации, объектно-ориентированного подхода к разработке программных средств.

Дискуссия

Поставленная в работе цель автоматизации оценивания ответов обучающихся может быть решена за счет использования методов машинного обучения и выявления кластеров, наиболее отвечающих обобщенной структуре ответов [Big Data clustering..., 2019; Kumar et al., 2016; Lenard et al., 2000]. Такая кластеризация нашла применение в таких сферах деятельности, как коммерция, маркетинг, социальные сети, рекомендательные системы фильмов, службы знакомств [Silven et al., 2003; Tsurko et al., 2022]. Другими словами, кластеризация позволяет группировать людей на основании их персональных особенностей. Аналогично такой подход позволит объективно дифференцировать результаты тестов обучающихся с учетом формируемых компетенций.

Распределение полученных результатов прохождения описанных выше тестов представлено на рисунке 1. Как можно увидеть, распределения ответов курсантов на различных тестированиях весьма неоднородны. Хотя общая картина распределения по всем тестам (рисунок 2) показывает, что ответы курсантов преобладают в зоне положительных оценок.

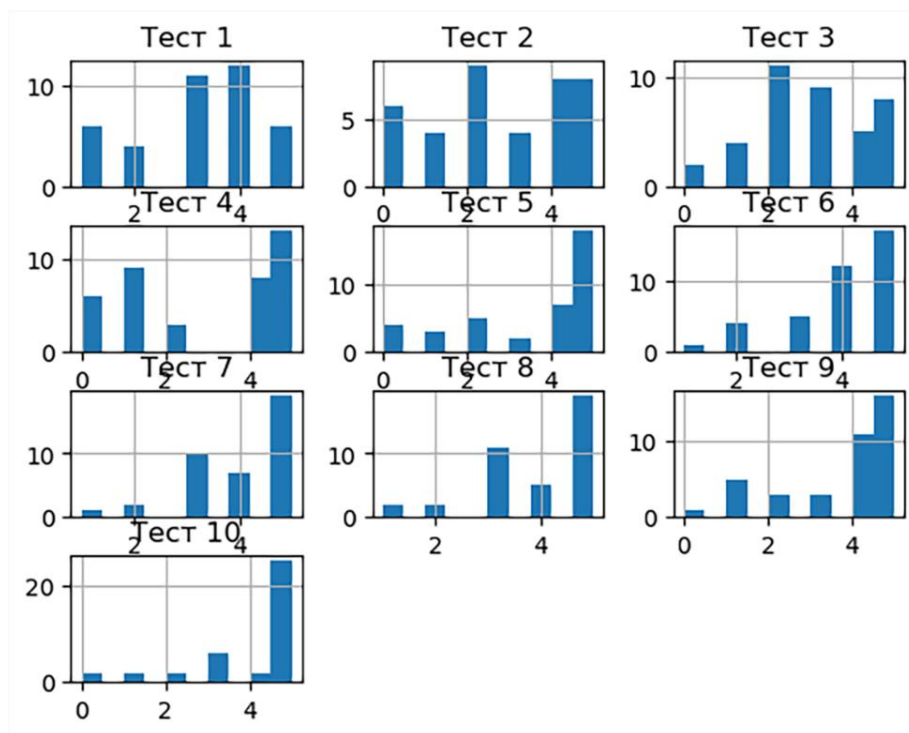


Рисунок 1 – Гистограммы распределения ответов учащихся по тестированиям

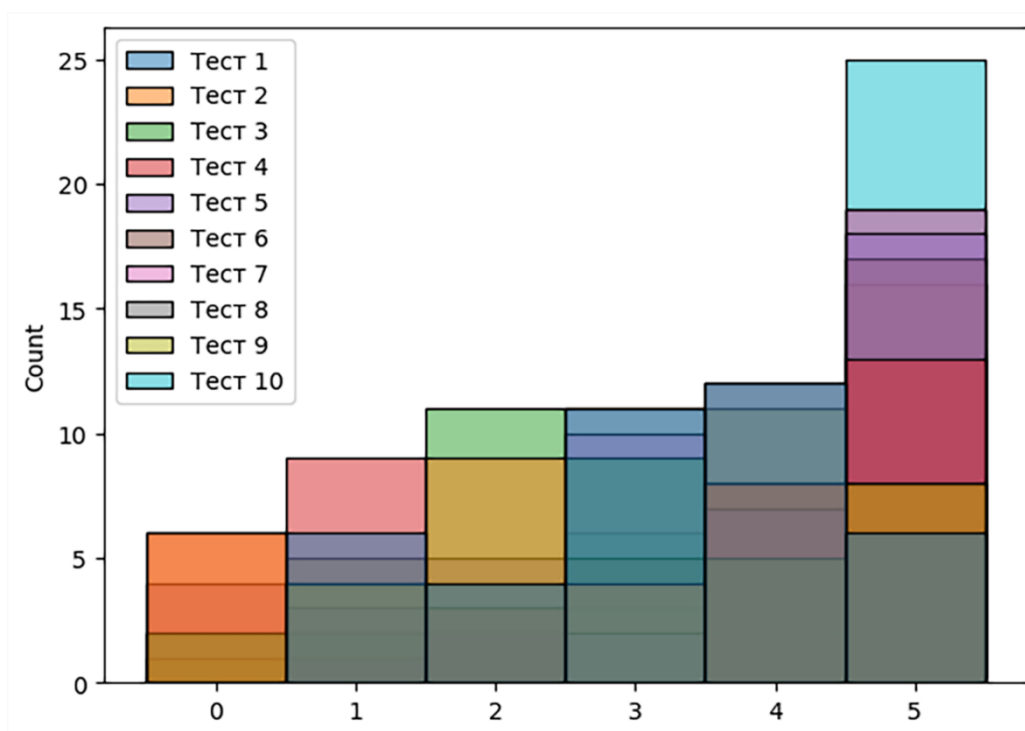


Рисунок 2 – Гистограмма распределения всех ответов учащихся по всем тестированиям

Для анализа полученных результатов был составлен датафрейм, у которого были выделены статистические характеристики, представленные в таблице 1. Обработка данных и дальнейшая работа с ними осуществлялась в среде разработки Jupyter Notebook с использованием высокоуровневого языка

объектно-ориентированного программирования Python, отличительной особенностью которого является динамическая строгая типизация и автоматическое управление памятью. Язык Python ориентирован на повышение производительности разработчика, читаемости кода и его качества, а также на обеспечение переносимости написанных на нём программ [Маккини, 2020, с. 23].

Оценка информации о полученном наборе данных велась с помощью инструментов программной библиотеки pandas, являющейся надстройкой над другой известной библиотекой NumPy, написанной на языке Python и служащей для обработки и анализа данных. Библиотека pandas позволяет использовать специальные структуры данных и операции для манипулирования числовыми таблицами и временными рядами [Маккини, 2020, с. 143].

Таблица 1 – Математическая статистика набора данных по результатам тестирования по дисциплине «Радиотехническое оборудование аэродромов»

Тест	Среднее значение	Среднеквадратическое отклонение	Медиана	Мода
Тест 1	3,21	1,28	3	4
Тест 2	2,72	1,75	3	2
Тест 3	2,89	1,47	3	2
Тест 4	2,87	2,00	4	5
Тест 5	3,51	1,79	4	5
Тест 6	4,03	1,11	4	5
Тест 7	4,05	1,10	4	5
Тест 8	3,95	1,21	4	5
Тест 9	3,69	1,52	4	5
Тест 10	4,03	1,53	5	5

При рассмотрении таблицы 1 можно отметить, что разница в средних значениях, медианах и модах указывает на то, что использование одинаковых подходов к оцениванию на разные тесты не может быть адекватным и объективным. Основным способом повышения точности оценивания курсанта является ручная корректировка получаемых за тест баллов со стороны преподавателя. Однако такое решение подразумевает большую трудоемкость процесса, субъективность из-за эмпиричности выставления весов заданий, отсутствие унифицированности в алгоритме выставления весов, а как следствие отсутствие возможности автоматизации. Кроме того, это никак не решит проблему индивидуального подхода к оцениванию курсанта и выявлению его слабых мест (недостаточности знаний, умений, навыков) в определенных областях. Это, в свою очередь, становится особенно актуально при выполнении тренажерной подготовки, в направлении которой видится дальнейшее развитие представленной работы.

Результаты

К машинному обучению относятся методы и алгоритмы искусственного интеллекта, позволяющие не явно решать поставленные задачи. В зависимости от решаемых задач по поиску выходной информации можно определить следующие парадигмы машинного обучения: обучение с учителем (supervised learning), обучение без учителя (unsupervised learning) и обучение с подкреплением (reinforcement learning). Исходя из стоящей задачи обобщения результатов тестирования, которую можно обозначить как кластеризация, следует воспользоваться одним из алгоритмов машинного обучения без учителя.

Концепция кластеризации предполагает группирование объектов на основе их взаимного сходства. В этом случае при сравнении данных одного наблюдения с данными других наблюдений можно определить степень их сходства без использования меток [Чио и др., 2020, с. 83]. Такой подход может быть применен при оценивании результатов тестов и группирования курсантов по выделенным заданиям.

В эксперименте при выполнении заданий тестов приняли участие 38 испытуемых (обучающихся). В результате был сформирован обучающий набор данных $X \in [0, 5]$, включающий 38 наблюдений, каждое из которых описывается 10 признаками (по количеству тестов). Для визуализации такого набора данных можно перевести значения из десятимерного пространства в двухмерное с помощью метода нелинейного снижения размерности и визуализации многомерных переменных t-SNE (t-distributed stochastic neighbor embedding). Методика моделирует каждое наблюдение высокой размерности так, что похожие наблюдения будут близко располагаться в двухмерном пространстве [t-SNE ..., s. a.]. Распределение точек выборки X представлено на рисунке 3.

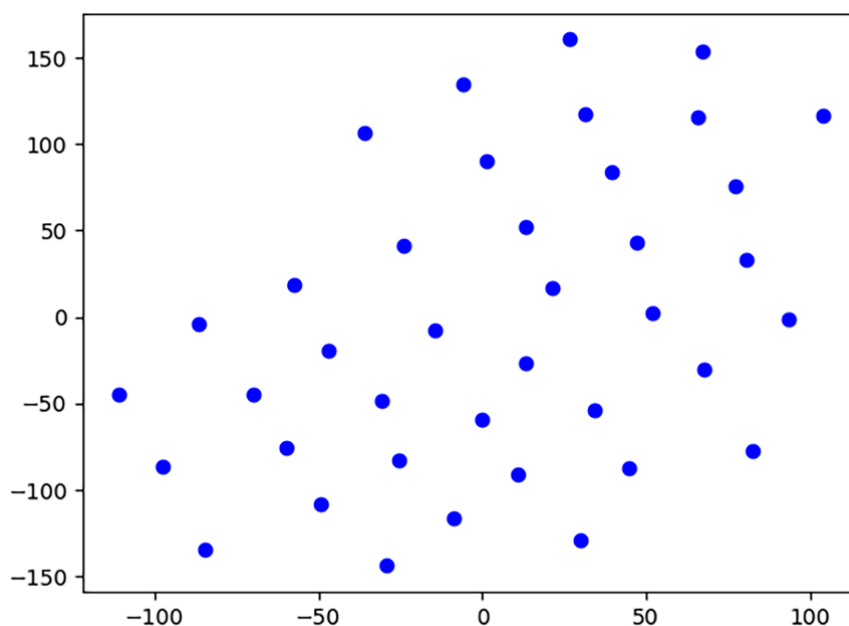


Рисунок 3 – Распределение заданных точек набора данных наблюдения с уменьшенной размерностью

Иерархическая кластеризация представляет собой совокупность алгоритмов упорядочивания данных, обеспечивающих формирование графа из вложенных кластеров. К алгоритмам, обеспечивающим такое упорядочивание, относятся алгоритмы вычисления расстояний между кластерами и алгоритмы вычисления расстояний между объектами.

Для определения различия между кластерами необходимо задать меру расстояния между двумя кластерами. Методами, позволяющими это сделать, являются:

- метод одиночной связи (метод ближайшего соседа), в котором расстояние между двумя кластерами высчитывается по минимальному расстоянию между двумя объектами из этих кластеров;
- метод полной связи (метод дальнего соседа), в котором расстояние между двумя кластерами высчитывается по максимальному расстоянию между двумя объектами из этих кластеров;
- метод средней связи, в котором расстояние между двумя кластерами высчитывается по усредненному значению расстояния между объектами из этих кластеров;
- центроидный метод, в котором расстояние между двумя кластерами высчитывается по расстоянию между их центрами массы (центроидами);
- метод Уорда, в котором расстояние между двумя кластерами высчитывается по приросту суммы квадратов расстояний объектов до центра кластера, получаемого в результате их объединения.

Использование дисперсионного анализа в методе Уорда выгодно выделяет его на фоне других подходов, так как это приводит к тому, что на каждом шаге алгоритма объединяются такие два кластера, которые приводят к минимальному увеличению дисперсии [Пател, 2020, с. 196]. Это означает, что метод является более эффективным при решении задач с близко расположенными кластерами. Метод может быть описан формулой [scipy.cluster.hierarchy.linkage..., s. a.]:

$$d(u, v) = \sqrt{\frac{|v|+|s|}{T} d(v, s)^2 + \frac{|v|+|t|}{T} d(v, t)^2 - \frac{|v|}{T} d(s, t)^2}, \quad (1)$$

где d – расстояние между кластерами;

u – новый кластер, состоящий из кластеров s и t ;

T – сумма мощностей кластеров v, s, t ($T = |v| + |s| + |t|$).

В качестве алгоритма вычисления расстояния между объектами выбрана евклидова метрика:

$$\sqrt{\sum_{i=1}^n (p_i - q_i)^2}, \quad (2)$$

где p_i, q_i – две точки пространства, расстояние между которыми ищется.

Для вычисления расстояния между соседними кластерами в зависимости от числа кластеров по алгоритму иерархического кластерного анализа была

написана программа на языке Python, в которой был использован метод `cluster.hierarchy.linkage` из библиотеки SciPy. Описанное соотношение представлено на рисунке 4.

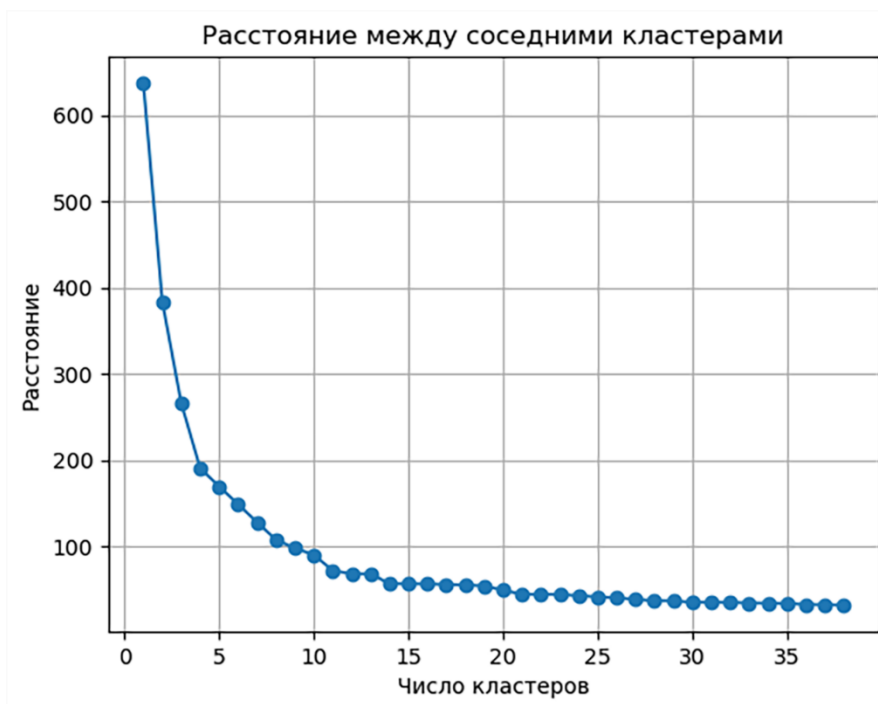


Рисунок 4 – Зависимость расстояния между соседними кластерами в зависимости от числа кластеров, вычисленная по алгоритму иерархического кластерного анализа

Для графического определения оптимального количества кластеров, на которые следует делить выборку, можно использовать так называемый критерий «каменистой осыпи», который состоит в поиске точки, где убывание собственных значений замедляется наиболее сильно [Метод каменистой осыпи..., 2017]. Как можно увидеть из графика (рисунок 4), наилучшим можно считать деление на 4 кластера.

Графически такое деление может быть представлено с помощью дендрограммы (рисунок 5), показывающей степень близости отдельных объектов и кластеров [Plot ..., s. a.].

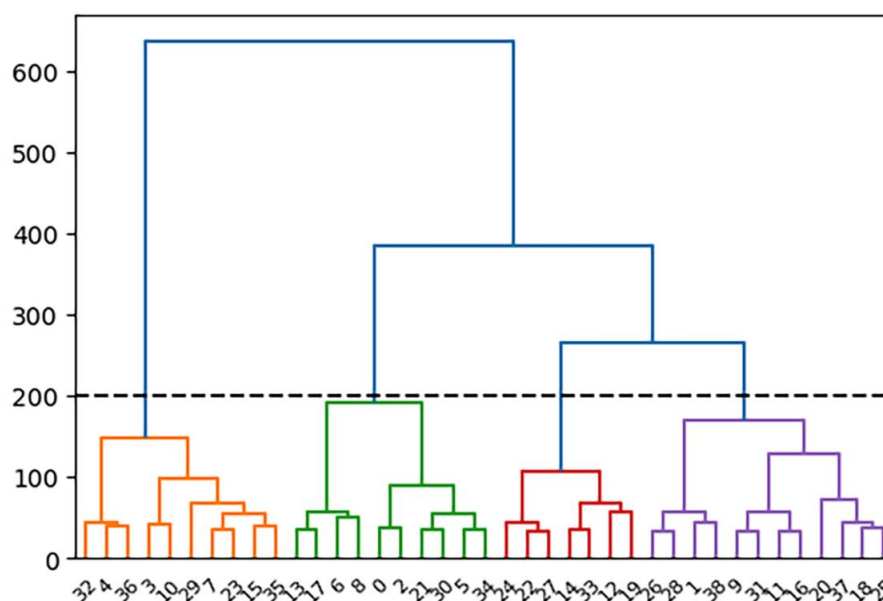


Рисунок 5 – Дендрограмма иерархической кластеризации

Усредненные значения величин баллов в каждом из 4 кластеров представлены в таблице 2.

Как можно увидеть из таблицы 2, к первому кластеру относятся обучающиеся, имеющие низкие оценки по всем темам, а следовательно, можно утверждать, что лекционный курс ими не был усвоен, из-за чего понадобилось проводить для них дополнительные занятия. Обучающихся из второго кластера можно охарактеризовать как выполнивших минимальные требования по освоению компетенций, способных работать с радиотехническим оборудованием, но имеющих пробелы по общефизическим темам, как в первом кластере. Обучающихся из третьего кластера можно охарактеризовать как отличников, а обучающихся из четвертого кластера как хорошистов. Такое распределение говорит прежде всего о готовности обучающихся, освоивших компетенции, решать практические задачи на рабочем месте пилота или диспетчера.

Таблица 2 – Распределение значений в кластерах

Номер кластера	Номер теста										Число объектов в кластере
	1	2	3	4	5	6	7	8	9	10	
1	2,3	1,1	1,4	0,9	1,2	2,9	3,2	2,5	1,6	2,2	10
2	3,6	2,6	2,8	1,7	3,5	4,1	3,8	4,3	4,3	4,7	10
3	4,3	4,3	4,1	4,9	4,9	4,9	4,6	4,3	4,4	4,3	7
4	3	3,3	3,5	4,3	4,6	4,4	4,7	4,7	4,5	4,8	12

Распределение точек выборки X с учетом проведенного иерархического кластерного анализа представлено на рисунке 6.

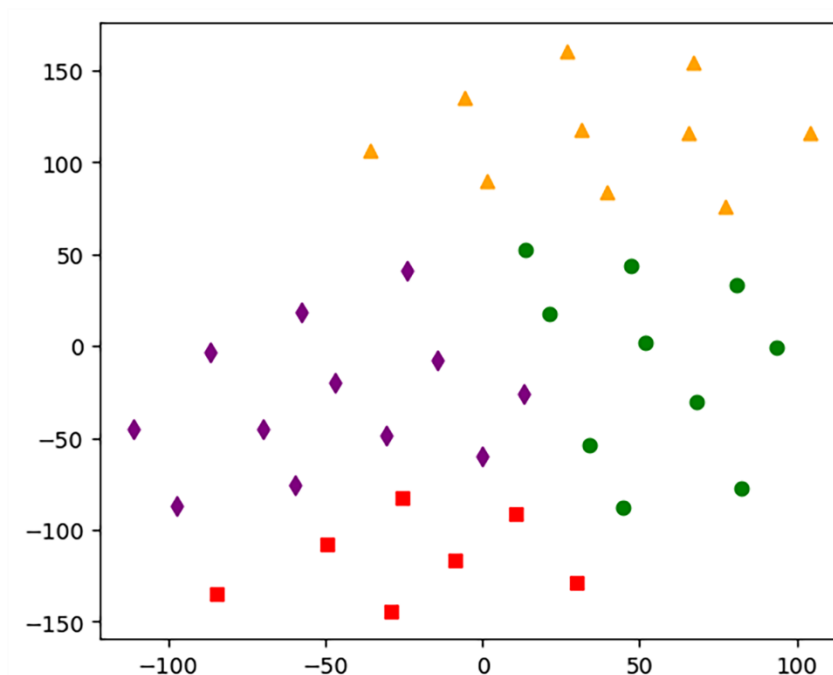


Рисунок 6 – Распределение заданных точек набора данных наблюдения с уменьшенной размерностью с учетом проведенного иерархического кластерного анализа. Треугольникам соответствует кластер 1, кругам – кластер 2, квадратам – кластер 3, ромбам – кластер 4

Таким образом, можно утверждать, что заявленная цель исследования по обеспечению объективного автоматизированного подхода к оцениванию результатов тестов обучающихся была достигнута. Для этого был применен алгоритм иерархического кластерного анализа, использующий метод Уорда для нахождения расстояний между кластерами. Описанное решение было представлено в виде программного продукта, написанного на языке Python.

Заключение

Приведенная в работе методика оценки освоения компетенций обучающимися в учебном заведении гражданской авиации позволяет автоматизировать процесс выставления оценок и найти те вопросы, в которых обучающийся имеет пробелы. Такой подход может найти применение не только при освоении дисциплины «Радиотехническое оборудование аэродромов», но и при освоении прочих дисциплин, а также при проведении тренажерной подготовки персонала на основании анализа фактических данных.

Сформированный на основе описанных методов и алгоритма программный продукт реализует экспертную систему оценивания результатов тестирования. Однако недостатком такой системы видится необходимость дополнительного обучения модели при изменении условий тестирования. Дальнейшие разработки авторского коллектива будут направлены на исключение данного недостатка.

Библиографический список

- Маккини У. Python и анализ данных / перевод с английского А. А. Слинкина. М.: ДМК Пресс, 2020. 540 с.
- Метод каменистой осыпи как основа решения метрологических задач в социально-гуманитарных областях знания (на примере задач экономики, педагогики и социологии) / В. И. Лойко, Д. А. Романов, В. Л. Шапошников [и др.] // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2017. № 129. С. 1382-1406. DOI 10.21515/1990-4665-129-099. EDN WFPZXC.
- Пател А. Прикладное машинное обучение без учителя с использованием Python / перевод с английского А. Г. Гузикович. Санкт-Петербург: ООО «Диалектика», 2020. 432 с.
- Чио К. Машинное обучение и безопасность / К. Чио, Д. Фримэн // перевод с английского А. В. Снастина. М.: ДМК Пресс, 2020. 388 с.
- Big Data clustering with varied density based on mapreduce / S. Heidari, M. Alborzi, M. A. Afsharkazemi [et al.] // Journal of Big Data. 2019. № 1. С. 6-22.
- Doc 10056. Manual on Air Traffic Controller Competency-based Training and Assessment. Монреаль, Канада: ICAO, 2017. 392 с.
- Doc 10057. Manual on Air Traffic Safety Electronics Personnel Competency-based Training and Assessment. Монреаль, Канада: ICAO, 2017. 184 с.
- Doc 9868. Правила аэронавигационного обслуживания. Подготовка персонала. Монреаль, Канада: ICAO, 2016. 254 с.
- Kumar S. Analysis of hourly road accident counts using hierarchical clustering and cophenetic correlation coefficient (CPCC) / S. Kumar, D. Toshniwal // Journal of Big Data. 2016. № 1. С. 3-13.
- Lenard M. J. An analysis of fuzzy clustering and a hybrid model for the auditor's going concern assessment / M. J. Lenard, P. Alam, D. Booth // Decision sciences. 2000. № 4. С. 861-884.
- Plot Hierarchical Clustering Dendrogram // [Электронный ресурс]. – URL: https://scikit-learn.org/stable/auto_examples/cluster/plot_agglomerative_dendrogram.html (дата обращения: 25.01.2023).
- scipy.cluster.hierarchy.linkage – SciPy v1.10.0 Manual // [Электронный ресурс]. – URL: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.cluster.hierarchy.linkage.html> (дата обращения: 19.01.2023).
- Silven O. Wood inspection with non-supervised clustering / O. Silven, N. Niskanen, H. Kauppinen // Machine vision and applications. 2003. № 5. С. 275-285.
- t-SNE в машинном обучении // [Электронный ресурс]. – URL: <https://biconsult.ru/products/t-sne-v-mashinnom-obuchenii> (дата обращения: 19.01.2023).
- Tsurko V. V. Evaluation of statistical relationship of random variables via mutual information / V. V. Tsurko, A. I. Mikhalskii // Automation and remote control. 2022. Vol. 83, No. 5. P. 734-742. DOI 10.1134/S000511792205006X. EDN PUCQOX.

References

- Chio K., Freeman D. (2020). Machine learning and security. Moscow: DMK Press, 2020. 388 p. (in Russian)
- Doc 10056. Manual on Air Traffic Controller Competency-based Training and Assessment. Montreal: ICAO, 2017. 392 p.
- Doc 10057. Manual on Air Traffic Safety Electronics Personnel Competency-based Training and Assessment. Montreal: ICAO, 2017. 184 p.
- Doc 9868. Procedures for Air Navigation Services. Staff training. Montreal: ICAO, 2016. 254 p. (in Russian)
- Heidari S., Alborzi M., Afsharkazemi M. A., Ghatari A. R. (2019). Big Data clustering with varied density based on mapreduce. *Journal of Big Data*. 1: 6-22.
- Kumar S., Toshniwal D. (2016). Analysis of hourly road accident counts using hierarchical clustering and cophenetic correlation coefficient (CPCC). *Journal of Big Data*: 1: 3-13.

- Lenard M. J., Alam P., Booth D. (2000). An analysis of fuzzy clustering and a hybrid model for the auditor's going concern assessment. *Decision sciences*. 4: 861-884.
- Loiko V. I., Romanov D. A., Shaposhnikov V. L., Kushnir N. V., Kushnir A. V. (2017). The scree method as a basis for solving metrological problems in social and humanitarian fields of knowledge (on the example of problems of economics, pedagogy and sociology). *Scientific journal of KubSAU*. 129: 1382-1406. (in Russian)
- Mackini W. (2020). Python and data analysis. Moscow: DMK Press, 2020. 540 p. (in Russian)
- Patel A. (2020). Unsupervised Applied Machine Learning with Python. Saint Petersburg: OOO Dialectika, 2020. 432 p. (in Russian)
- Plot Hierarchical Clustering Dendrogram. Available at: https://scikit-learn.org/stable/auto_examples/cluster/plot_agglomerative_dendrogram.html (accessed 25 January 2023).
- scipy.cluster.hierarchy.linkage — SciPy v1.10.0 Manual. Available at: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.cluster.hierarchy.linkage.html> (accessed 19 January 2023).
- Silven O., Niskanen N., Kauppinen H. (2003). Wood inspection with non-supervised clustering. *Machine vision and applications*. 5: 275-285.
- t-SNE in machine learning. Available at: <https://biconsult.ru/products/t-sne-v-mashinnom-obuchenii> (accessed 19 January 2023). (in Russian)
- Tsurko V. V., Mikhalskii A. I. (2022). Evaluation of statistical relationship of random variables via mutual information. *Automation and remote control*. 83(5): 734-742.