

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение высшего образования
«Тольяттинский государственный университет»

Кафедра _____ «Прикладная математика и информатика» _____
(наименование)

01.03.02 Прикладная математика и информатика
(код и наименование направления подготовки / специальности)

Компьютерные технологии и математическое моделирование
(направленность (профиль) / специализация)

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА (БАКАЛАВРСКАЯ РАБОТА)

на тему «Исследование и реализация алгоритма формирования рейтинга студентов на
основе их цифрового следа» _____

Обучающийся _____ В.А. Чураков _____
(Инициалы Фамилия) (личная подпись)

Руководитель _____ к.ф.-м.н. О.В. Лелонд _____
(ученая степень (при наличии), ученое звание (при наличии), Инициалы Фамилия)

Консультант _____ к.п.н., доцент А.В. Егорова _____
(ученая степень (при наличии), ученое звание (при наличии), Инициалы Фамилия)

Тольятти 2024

Аннотация

Тема выпускной квалификационной работы – «Исследование и реализация алгоритма формирования рейтинга студентов основе их цифрового следа».

Необходимым условием эффективного процесса обучения является наличие адаптивного и гибкого механизма формирования рейтинга студентов.

Объектом исследования бакалаврской работы является рейтинг студентов.

Предметом исследования бакалаврской работы является алгоритм формирования рейтинга студентов.

Цель бакалаврской работы – исследование и реализация алгоритма формирования рейтинга студентов на основе их цифрового следа.

Методы исследования – методы и алгоритмы интеллектуального анализа данных, технологии реализации алгоритмов на языках высокого уровня.

Практическая значимость бакалаврской работы заключается в разработке и тестировании программы, реализующей эффективный алгоритм формирования рейтинга студентов на основе их цифрового следа.

Результаты бакалаврской работы представляют научно-практический интерес и могут быть рекомендованы для анализа и программной реализации методов и алгоритмов формирования рейтинга студентов основе их цифрового следа.

Бакалаврская работа состоит из 40 страницы текста, 9 рисунков, 3 таблиц и 33 источников.

Abstract

The title of the graduation work is: «Research and implementation of an algorithm for forming a student rating based on their digital footprint».

The study and features of the practical application of algorithms for forming a student rating is relevant and of scientific and practical interest.

The object of study of the bachelor's work is the student rating.

The subject of research of the bachelor's work is the algorithm for forming a student rating.

The purpose of the bachelor's work is the study and implementation of algorithm for forming a student rating based on their digital footprint.

The research methods include methods and algorithms of student rating, technologies for implementing algorithms in a high-level language.

The practical significance of the bachelor's work lies in the development and testing of a program that implements effective algorithm for forming a student rating.

The results of the bachelor's work are of scientific and practical interest and can be recommended for analysis and software implementation of methods and algorithms for forming a student rating.

The bachelor's thesis consists of 40 pages of text, 9 figures, 3 tables and 33 literature sources.

Оглавление

Введение.....	5
Глава 1 Постановка задачи исследования и анализ методов формирования рейтинга студентов	7
1.1 Постановка задачи исследования.....	7
1.2 Методы формирования рейтинга студентов.....	9
1.2.1 Методика расчета индивидуальных рейтингов студентов на основе учебного показателя	9
1.2.2 Метод формирования интегральной рейтинговой оценки деятельности студента	10
1.2.3 Метод формирования итогового рейтингового показателя студента	12
1.2.4 Метод формирования рейтинга студентов по среднему баллу .	13
Глава 2 Обзор и анализ методов ранжирования	16
2.1 Методы детерминированного ранжирования.....	16
2.2 Методы вероятностного ранжирования	17
2.3 Методы ранжирования на основе машинного обучения.....	18
2.4 Анализ методов ранжирования на основе машинного обучения	20
Глава 3 Разработка и тестирование программы ранжирования студентов.....	25
Заключение	36
Список используемой литературы и используемых источников.....	38

Введение

Как показывает практика, необходимым условием эффективного и увлекательного процесса обучения является наличие адаптивного и гибкого механизма оценки образовательной деятельности студента, которая полезна как для студентов, так и для преподавателей.

Таким механизмом является механизм формирования рейтинга студентов, для реализации которого используются специальные алгоритмы.

В настоящее время в процесс обучения вузов активно внедряются технологии дистанционного обучения и LMS (Learning management system) – современные системы управления обучением.

Интеллектуальный анализ образовательных данных показал, что большой объем данных о студентах, собираемых системами электронного обучения, можно использовать для формирования рейтингов студентов для их всестороннего исследования.

Основным источником таких данных является цифровой след студента.

Таким образом, исследование и особенности программной реализации алгоритма формирования рейтинга студентов на основе их цифрового следа являются актуальными и представляют научно-практический интерес.

Объектом исследования бакалаврской работы является рейтинг студентов.

Предметом исследования бакалаврской работы является алгоритм формирования рейтинга студентов.

Цель бакалаврской работы – исследование и реализация алгоритма формирования рейтинга студентов на основе их цифрового следа.

Для достижения данной цели необходимо выполнить следующие задачи:

- выполнить постановку задачи исследования и проанализировать методы формирования рейтинга студентов;
- проанализировать методы ранжирования студентов по успеваемости;

– разработать и протестировать программу, реализующую алгоритм формирования рейтинга студентов на основе их цифрового следа.

Методы исследования – методы формирования рейтинга студентов, методы и алгоритмы интеллектуального анализа данных, технологии реализации алгоритмов на языках высокого уровня.

Практическая значимость бакалаврской работы заключается в разработке и тестировании программы, реализующей эффективный алгоритм формирования рейтинга студентов на основе их цифрового следа.

Данная работа состоит из введения, трех глав, заключения и списка используемой литературы.

Первая глава работы посвящена постановке задачи исследования и анализу методов формирования рейтинга студентов.

Вторая глава работы посвящена обзору и анализу алгоритмов формирования рейтинга студентов.

В третьей главе рассматривается программная разработка и тестирование алгоритма формирования рейтинга студентов на основе их цифрового следа.

В заключении описываются результаты выполнения выпускной квалификационной работы.

Бакалаврская работа состоит из 40 страницы текста, 9 рисунков, 3 таблиц и 23 источников.

Глава 1 Постановка задачи исследования и анализ методов формирования рейтинга студентов

1.1 Постановка задачи исследования

Оценка качества высшего образования состоит из внешней и внутренней оценки качества образовательных услуг.

Внешняя оценка качества образовательных услуг осуществляется всеми экономическими субъектами, которые являются потребителями. Внутренняя оценка качества образовательных услуг должна проводиться самим университетом. Для этих целей университет разрабатывает систему управления качеством и организует отдел, ответственный за организацию и мониторинг качества образовательных услуг.

Внутренняя оценка качества образовательных услуг включает такие параметры, как качество подготовки студентов и качество преподавания в университете. Для решения данной задачи в вузах разрабатывается рейтинговая оценка знаний студентов [16].

Рассмотрим известные определения рейтинга студентов.

Рейтинг студента – это числовой показатель оценивания успешности освоения дисциплины (учебного курса), отражающий место студента в групповом (поточковом, курсовом) рейтинге в соответствии с полученным итоговым рейтинговым баллом [9].

Рейтинг – отсортированный список студентов, демонстрирующий успехи в освоении дисциплин из учебного плана, за определенный период [10].

Алгоритм формирования (составления) рейтинга студентов определяет критерии и порядок оценивания достижений студента в соответствии с выбранным методом формирования рейтинга студентов [2].

Источником данных для формирования рейтинга студентов являются большие образовательные данные (Educational Big Data, EBD) [23].

Как показывают результаты интеллектуального анализа

образовательных данных (Educational Data Mining, EDM), основным источником данных для формирования рейтинга студента является его цифровой след [11].

«Цифровой след n – это запись онлайн-активности пользователя Интернет.

Цифровой след показывает, где был пользователь, данные, которыми он поделился, и информацию, которые он передал или оставил» [22].

«В рассматриваемом случае цифровой след студента – это запись его онлайн-активности в системах управления обучением.

Существует две основные категории цифровых следов: активные следы и пассивные следы. Какой из них применяется, зависит от того, как информация получена.

Активные цифровые следы состоят из данных, которые остаются, когда пользователи Интернета делают преднамеренный выбор. Например, посты, сделанные в социальных сетях, являются типом активного следа.

Примеры активного цифрового следа студента: работа в LMS и др.

Пассивные цифровые следы – это те, которые остаются без намерения пользователя Интернета, а иногда и без ведома об этом. Пассивные цифровые следы – это скрытый процесс, о котором пользователи Интернета могут вообще не знать.

Примером пассивного цифрового следа студента является образовательный портал ТГУ. Студенты оставляют цифровые следы в любое время, когда они находятся в сети, независимо от того, находятся ли они в вузе или за его пределами» [18].

Задача настоящего исследования заключается в математическом описании методов, анализе и реализации алгоритмов формирования рейтинга студентов на основе их цифрового следа.

1.2 Методы формирования рейтинга студентов

«Следует отметить, что в настоящее время отсутствует единая методика формирования рейтинга студентов. Каждый вуз формирует собственный порядок расчета с учетом тех аргументов и коэффициентов, которые считает актуальными для себя» [9].

Рассмотрим основные методы формирования рейтинга студентов.

1.2.1 Методика расчета индивидуальных рейтингов студентов на основе учебного показателя

Данная методика используется в Северо-Западном институте управления РАНХиГС [1].

Учебный показатель рассчитывается по следующей формуле (1):

$$R = \frac{\sum_i x_i K_i}{100 \sum_i K_i}, \quad (1)$$

где x_i – количество баллов по i -й дисциплине;

K_i – весовой коэффициент по i -й дисциплине, равный объему i -й дисциплины в ЗЕТ по рабочему учебному плану.

Индивидуальный рейтинг студента (ИРС) рассчитывается по формуле (2):

$$\text{ИРС} = \text{УП} + \text{НП} + \text{ОП} + \text{КТП} + \text{СП} + \text{БР} - \text{ШБ}, \quad (2)$$

где УП – учебный показатель;

НП – научный показатель;

ОП – общественный показатель;

КТП – культурно-творческий показатель;

СП – спортивный показатель;

БР – бонусы института;

ШБ – штрафные баллы.

По итогам подсчета рейтинга определяется общий рейтинг группы (ОРГ), который складывается из суммы показателей ИРС и делится на количество студентов группы.

Все показатели, кроме УП, представляются в табличной форме.

Утвержденные рейтинги размещаются на сайте Института.

1.2.2 Метод формирования интегральной рейтинговой оценки деятельности студента

«Интегральный рейтинг студента строится на основе следующей структуры (3):

$$R_{\text{инг}} = w_{\text{ак}} R_{\text{ак}} + w_{\text{в}} \quad (3)$$

где $R_{\text{ак}}$, $R_{\text{в}}$ – значения академического и внеучебного рейтингов соответственно;

$w_{\text{ак}}$, $w_{\text{в}}$ – весовые коэффициенты частных рейтингов студентов, причем $w_{\text{ак}} + w_{\text{в}} = 1$.

Условие, ограничивающее размер весов $w_{\text{ак}}$ и $w_{\text{в}}$, обеспечивает единообразие шкал частных и интегрального рейтинга.

Очевидно, что область определения $R_{\text{ак}}$, $R_{\text{в}}$, $R_{\text{инг}}$ – положительные числа от 0 до некоторого Z . Обычно $Z = 100$ » [4].

«В большинстве случаев академический рейтинг определяется по формуле (4):

$$R_{\text{ак}} = \sum_{p=1}^P b_p R_p, \quad (4)$$

где p – номер дисциплины (или другого вида работ, например, практики, курсовой работы);

P – количество дисциплин в отчетном периоде;

b_p – вес дисциплины;

R_p – рейтинг студента по p -й дисциплине.

Формула (4) позволяет рассчитать как полный академический рейтинг за все годы обучения, так и текущий накопительный рейтинг, при этом множитель b_p для еще не изученных дисциплин (невыполненных видов работ) полагается равным 0.

Внеучебный рейтинг определяется по следующей формуле (5):

$$R_B = \sum_{k=1}^K R_k, \quad (5)$$

где k – номер мероприятия;

K – количество мероприятий в отчетном периоде (например, учебном году или семестре);

R_k – рейтинг студента в k -м мероприятии.

В свою очередь, рейтинг студента в мероприятии R_k можно определить по формуле (6).

$$R_k = \alpha_k \sum_{m=1}^M A_k^m, \quad (6)$$

где α_k – весовой коэффициент мероприятия;

A_k^m – вес статуса студента в k -м мероприятии;

M – количество статусов в одном мероприятии;

m – номер статуса студента в одном мероприятии.

Предлагаемый метод расчета интегрального рейтинга студента представляет собой обоснованную оценку деятельности студента, которая может служить инструментом управления образовательным процессом.

Его применение позволяет унифицировать процесс и автоматизировать работу по формированию различных рейтингов студента.

Метод был апробирован в СГПУ» [3].

1.2.3 Метод формирования итогового рейтингового показателя студента

Введем следующие обозначения [9]:

S_i – долей выполнения учебных заданий или доля успешности прохождения студентом контрольных точек;

$(S_{cr})_i$ – пороговая доля прохождения студентом контрольных точек.

«Итоговый рейтинговый показатель (D), характеризующий успешность освоения дисциплины в целом, строится по аддитивной (кумулятивной) схеме.

Он представляет собой линейную комбинацию составляющих, которые включают доли выполнения всех m отдельных видов деятельности ($S_1, \dots S_m$) при том, что выполняются условия $S_i \geq (S_{cr})_i$ (7):

$$D = \sum_{i=1}^m \alpha_i S_i, \quad (7)$$

где α_i – коэффициенты весовой значимости оценки вида i -й деятельности.

При этом, очевидно, что (8):

$$\sum_{i=1}^m \alpha_i = 1 \quad (8)$$

Изначально рейтинговый показатель D нормирован на 1 и обозначает долю выполнения студентом всей учебной работы по дисциплине.

При этом, если какая-либо из составляющих $S_i = 0$, то и $D = 0$.

Это условие соответствует требованию обязательного достижения установленного минимума (S_{cr})_i во всех видах деятельности и невозможности компенсировать один вид за счет другого.

От D по установленной оценочной шкале можно легко перейти к балльной оценке или оценке по ECST grading scale – Европейской системе перевода и накопления баллов» [12].

1.2.4 Метод формирования рейтинга студентов по среднему баллу

Средний балл студента (Grade Point Average, GPA) GPA или средний балл – это число, которое показывает, насколько высоко студенты в среднем набрали баллы на своих курсах.

Используя шкалу от 1,0 до 4,0, можно с помощью среднего балла отслеживать успеваемость студента. Это число используется для оценки того, соответствует ли студент стандартам и ожиданиям, установленным программой получения квалификации [19].

Формула расчета среднего балла студента имеет вид (9):

$$GPA = SO / KD, \quad (9)$$

где SO – сумма оценок студента;

KD – количество зачетных дисциплин.

После этого с помощью выбранного алгоритма производится ранжирование всех студентов одной образовательной программы по GPA.

Ранжирование – это отношение между набором элементов, такое, что для любых двух элементов первый либо «ранжируется выше», либо «ранжируется ниже», либо «ранжируется равно» второму. В математике это известно, как слабый порядок или полный предварительный порядок объектов [30].

Уровень ранжирования используется для определения различных рейтингов, используемых институтом, на основе процента баллов по каждому предмету и на основе количества предметов. На основе определенных уровней ранжирования могут быть сгенерированы отчеты учащегося. Для серий с

системой CWA / Normal следует выставлять оценки в %, а для серий с системой GPA - оценку GPA [20], [29], [33].

Для сравнения методов формирования рейтинга студентов используем таблицу 1 [5].

Таблица 1 – Сравнительный анализ методов формирования рейтинга студентов

Метод	Преимущества	Недостатки
Методика расчета индивидуальных рейтингов студентов на основе учебного показателя	«Повышение объективности оценки качества усвоения знаний по дисциплине.	«Система призвана измерять успеваемость студентов на основании их регулярной работы в течение семестра. Полученные данные не будут представлять действительное соотношение сил, если значительная часть студентов не будет работать регулярно. Сложность ранжирования Сложность реализации» [5]
Метод формирования интегральной рейтинговой оценки деятельности студента	Предсказуемость итоговой оценки. Рейтинг может быть получен как для отдельной дисциплины, так и для группы дисциплин, и может учитывать достижения студента за любой период обучения» [5]	
Метод формирования итогового рейтингового показателя студента		
Метод формирования рейтинга студентов среднему баллу	Широкая область применения. Простота ранжирования студентов. Простота реализации	Пониженная мотивация студентов. Невысокие возможности для анализа успеваемости

Таким образом, недостатки методов расчета индивидуальных рейтингов студентов на основе учебного показателя, формирования интегральной рейтинговой оценки деятельности студента и рейтингового показателя студента обусловлены особенностями балльно-рейтинговой системы (БРС) [8].

Поэтому более предпочтительным представляется метод формирования рейтинга студентов по среднему баллу.

Выводы по главе 1

Результаты проделанной в первой главе работы позволили сделать следующие выводы:

- алгоритм формирования рейтинга студентов определяет критерии и порядок оценивания достижений студента в соответствии с выбранным методом формирования рейтинга студентов;
- источником данных для формирования рейтинга студентов является его цифровой след;
- уровень ранжирования используется для определения различных рейтингов, используемых институтом, на основе процента баллов по каждому предмету и на основе количества предметов;
- недостатки методов расчета индивидуальных рейтингов студентов на основе учебного показателя, формирования интегральной рейтинговой оценки деятельности студента и рейтингового показателя студента обусловлены особенностями БРС.

Поэтому более предпочтительным представляется метод формирования рейтинга студентов по среднему баллу.

Глава 2 Обзор и анализ методов ранжирования

Методы ранжирования используются для ранжирования элементов в наборе данных в соответствии с некоторым критерием. Методы ранжирования можно разделить на две категории: детерминированные и вероятностные.

Методы ранжирования используются в поисковых системах для ранжирования веб-страниц в соответствии с их релевантностью поисковому запросу пользователя, для ранжирования студентов по успеваемости и др. [14].

Рассмотрим свойства известных методов ранжирования [31].

2.1 Методы детерминированного ранжирования

«Метод детерминированного ранжирования – это метод, в котором порядок элементов в ранжированном списке фиксирован и не меняется независимо от входных данных.

Примером детерминированного алгоритма ранжирования является алгоритм ранжирования по признакам.

В этом алгоритме каждому элементу присваивается ранг на основе значения его признака.

Шаг 1. Элементу с наименьшим значением признака присваивается ранг 1, а элементу с наибольшим значением признака присваивается ранг N, где N – количество элементов в наборе данных.

Шаг 2. Если несколько значений равны, им начисляется ранг, представляющий собой среднее значение из тех рангов, которые они получили бы, если бы не были равны.

Шаг 3. Общая сумма рангов должна совпадать с расчетной, которая определяется по формуле (10)» [14]:

$$\sum_{i=1}^n (R_i) = \frac{N(N+1)}{2}, \quad (10)$$

где N – общее количество ранжируемых наблюдений (значений).

«Несовпадение реальной и расчетной сумм рангов будет свидетельствовать об ошибке, допущенной при начислении рангов или их суммировании.

Одним из реальных применений детерминированного алгоритма ранжирования является упорядочение товаров в продуктовом магазине. Товары в продуктовом магазине обычно упорядочены по отделам, таким как продукты, мясо, молочные продукты и т.д.

В каждом отделе элементы обычно организованы в алфавитном порядке. Этот тип организации является примером детерминированного алгоритма ранжирования.

Алгоритмы сортировки используются в алгоритмах детерминированного ранжирования для упорядочивания элементов в ранжированном списке. Существует множество различных типов алгоритмов сортировки, каждый со своим набором преимуществ и недостатков. Одними из наиболее распространенных алгоритмов сортировки являются сортировка вставкой, сортировка слиянием и быстрая сортировка» [9].

2.2 Методы вероятностного ранжирования

В методе вероятностного ранжирования порядок элементов в ранжированном списке может меняться в зависимости от входных данных.

Примером вероятностного алгоритма ранжирования является алгоритм ранжирования по достоверности [24], [26], [27].

В этом алгоритме каждому элементу присваивается ранг на основе его значения достоверности. Элементу с наибольшим значением достоверности присваивается ранг 1, а элементу с наименьшим значением достоверности присваивается ранг N , где N – количество элементов в наборе данных.

Другим примером вероятностного алгоритма ранжирования является

байесовский фильтр нежелательной почты. В этом алгоритме каждому электронному письму присваивается вероятность того, что оно окажется спамом. Электронные письма с наибольшей вероятностью ранжируются первыми, а электронные письма с наименьшей вероятностью – последними.

«Вероятностные алгоритмы ранжирования могут использоваться в поисковых системах для ранжирования веб-страниц в соответствии с их релевантностью поисковому запросу пользователя. Алгоритм ранжирования использует входные данные, такие как количество ссылок на веб-страницу с других веб-сайтов и количество раз, когда ключевое слово встречается на странице, для вычисления показателя релевантности страницы. Чем выше показатель релевантности, тем выше рейтинг страницы в результатах поиска.

Вероятностные алгоритмы ранжирования также могут использоваться в алгоритмах машинного обучения для ранжирования элементов в наборе данных в соответствии с их вероятностью того, что они являются положительным примером» [9].

Алгоритм ранжирования использует входные данные, такие как количество признаков, общих как для положительных, так и для отрицательных примеров, для вычисления показателя релевантности элемента.

Чем выше оценка релевантности, тем больше вероятность того, что элемент является положительным примером. Существует много различных типов вероятностных алгоритмов ранжирования, каждый со своим набором преимуществ и недостатков. Например, наивный байесовский алгоритм.

2.3 Методы ранжирования на основе машинного обучения

Обучение ранжированию (Learning to rank) – «класс задач машинного обучения, суть которых состоит в автоматизированном построении ограничений для ранжирующей модели по обучающей выборке, для их последующего применения к неизвестным объектам со сходной структурой»

[6].

Рассмотрим постановку задачи.

Пусть:

X – множество объектов;

$X^l = \{x_1, \dots, x_l\}$ – обучающая выборка;

$i < j$ – правильный порядок на парах $(i, j) \in \{1, \dots, l\}^2$

Задача:

Построить ранжирующую функцию $a: X \rightarrow R$ такую, что (11):

$$i < j \Rightarrow a(x_i) < a(x_j) \quad (11)$$

Линейная модель ранжирования (12):

$$a(x; w) = \langle x, w \rangle, \quad (12)$$

где $x \mapsto (f(x), \dots, f_n(x)) \in R^n$ – вектор признаков объекта x .

Схема процесса ранжирования на основе машинного обучения показан на рисунке 1.

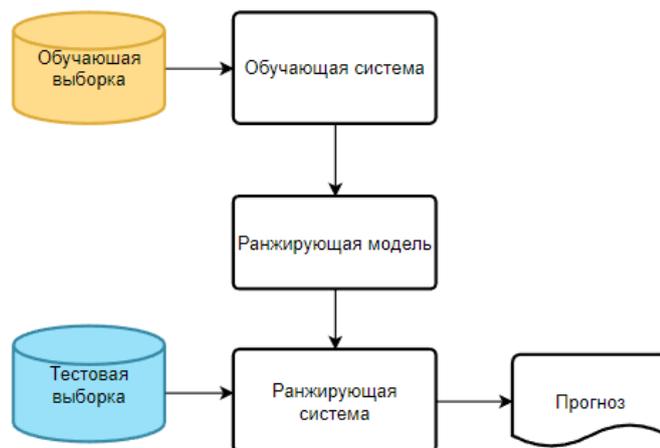


Рисунок 1 – Схема процесса ранжирования на основе машинного обучения

Для сравнения алгоритмов ранжирования построена таблица 2.

Таблица 2 – Сравнение методов ранжирования

Метод	Преимущества	Недостатки
Детерминированное ранжирование	Высокое быстродействие, относительная простота реализации	Сильно деградируют по скорости при неудачных выборах опорных элементов, что может случиться при неудачных входных данных
Вероятностное ранжирование	Высокая точность, эффективны, когда связь между атрибутами набора данных относительно независима	Независимость атрибутов наборов данных во многих случаях трудно обеспечить, поскольку атрибуты наборов данных часто коррелируют друг с другом, эффект классификации будет значительно снижен
Ранжирование на основе машинного обучения	Применение в инновационных приложениях, которые могут улучшить способы выполнения процессов и задач.	Относительная сложность реализации, временные обучения в организации, ограничения прогнозов

Принимая во внимание инновационные возможности методов ранжирования на основе машинного обучения, выбираем указанные методы для ранжирования студентов по среднему баллу.

2.4 Анализ методов ранжирования на основе машинного обучения

Для решения задачи ранжирования на основе машинного обучения используются интеллектуальный анализ данных и следующие методы ранжирования:

- разбиение по Парето;

- правило Борда;
- аналитический иерархический процесс (Analytic Hierarchy Process, АНР) с кластеризацией данных.

Рассмотрим особенности каждого метода.

«Разбиение по Парето: рассмотрим векторное отношение R «больше» такое, что для двух вариантов $x, y \in X$ имеет место xRy если и только если $x_j \geq y_j$ для всех j , причем хотя бы одно неравенство строгое.

На каждом шаге находим множество недоминируемых объектов, которые объединяются в класс C_l . Полученный класс исключается из рассмотрения, и процедура повторяется для оставшихся объектов $X \setminus C_l$, пока множество X не пусто.

Правило Борда: для каждого объекта x_i вычисляются оценки по всем критериям (13)» [7]:

$$r_i(x_j) = |\{x_p \in X: x_{ij} > p_j, p \neq i\}|. \quad (13)$$

Затем итоговый ранг объекта x_i подсчитывается как сумма оценок по всем критериям (14):

$$r(x_j) = \sum_{j=1}^M r_j(x_i). \quad (14)$$

«АНР – это метод организации и анализа сложных решений с использованием математики и психологии. Он был разработан Томасом Л. Саати в 1970-х годах и с тех пор совершенствовался.

Он состоит из трех частей: конечной цели или проблемы, которую необходимо решить, всех возможных решений, называемых альтернативами, и критериев, по которым вы будете оценивать альтернативы.

АНР обеспечивает рациональную основу для необходимого решения, определяя его критерии и альтернативные варианты, а также связывая эти

элементы с общей целью.

Заинтересованные стороны сравнивают важность критериев, по два за раз, путем попарного сравнения. АНР преобразует эти оценки в числа, которые можно сравнить со всеми возможными критериями. Эта возможность количественного определения отличает АНР от других методов принятия решений.

На заключительном этапе процесса рассчитываются численные приоритеты для каждого из альтернативных вариантов. Эти числа представляют наиболее желаемые решения, основанные на значениях всех пользователей» [21].

На рисунке 2 представлена блок-схема процесса ранжирования по методу АНР с применением метода кластеризации [17].



Рисунок 2 – Блок-схема процесса ранжирования по методу АНР с применением метода кластеризации

Для сравнения методов ранжирования на основе машинного обучения построена таблица 3.

Таблица 3 – Характеристики методов ранжирования на основе машинного обучения

Метод	Преимущества	Недостатки
Разбиение по Парето	Возможность четкой структуризации или классификации задач на более или менее значимые	Не универсален и работает не во всех случаях. Сложность реализации с помощью алгоритмов машинного обучения
Правило Борда	Имеется возможность детального выражения целого ряда предпочтений	Необходим какой-то уровень математических знаний, что может препятствовать полному пониманию метода. Сложность реализации с помощью алгоритмов машинного обучения
АНР с кластеризацией данных	Простота реализации с помощью алгоритмов машинного обучения	Формирование структуры модели принятия решения в методе анализа иерархий достаточно трудоемкий процесс

Принимая во внимание простоту реализации с помощью алгоритмов машинного обучения метода АНР, выбираем данный метод для ранжирования студентов по среднему баллу на основе машинного обучения.

Выводы по главе 2

Результаты проделанной работы позволили сделать следующие выводы:

- в методе детерминированного ранжирования порядок элементов в ранжированном списке фиксирован и не меняется независимо от входных данных;
- в методе вероятностного ранжирования порядок элементов в ранжированном списке может меняться в зависимости от входных данных;

- «обучение ранжированию – класс задач машинного обучения, суть которых состоит в автоматизированном построении ограничений для ранжирующей модели по обучающей выборке, для их последующего применения к неизвестным объектам со сходной структурой» [17];
- инновационные возможности методов ранжирования на основе машинного обучения позволяет использовать указанные методы для ранжирования студентов по среднему баллу;
- для решения задачи ранжирования на основе машинного обучения используются интеллектуальный анализ данных и следующие методы ранжирования: разбиение по Парет, правило Борда, аналитический иерархический процесс (Analytic Hierarchy Process, АНР) с кластеризацией данных и др.;

Простота реализации метода АНР с помощью алгоритмов машинного обучения позволяет использовать указанный метод для ранжирования студентов по среднему баллу.

Глава 3 Разработка и тестирование программы ранжирования студентов

На рисунке 3 показан алгоритм ранжирования студентов методом кластеризации.

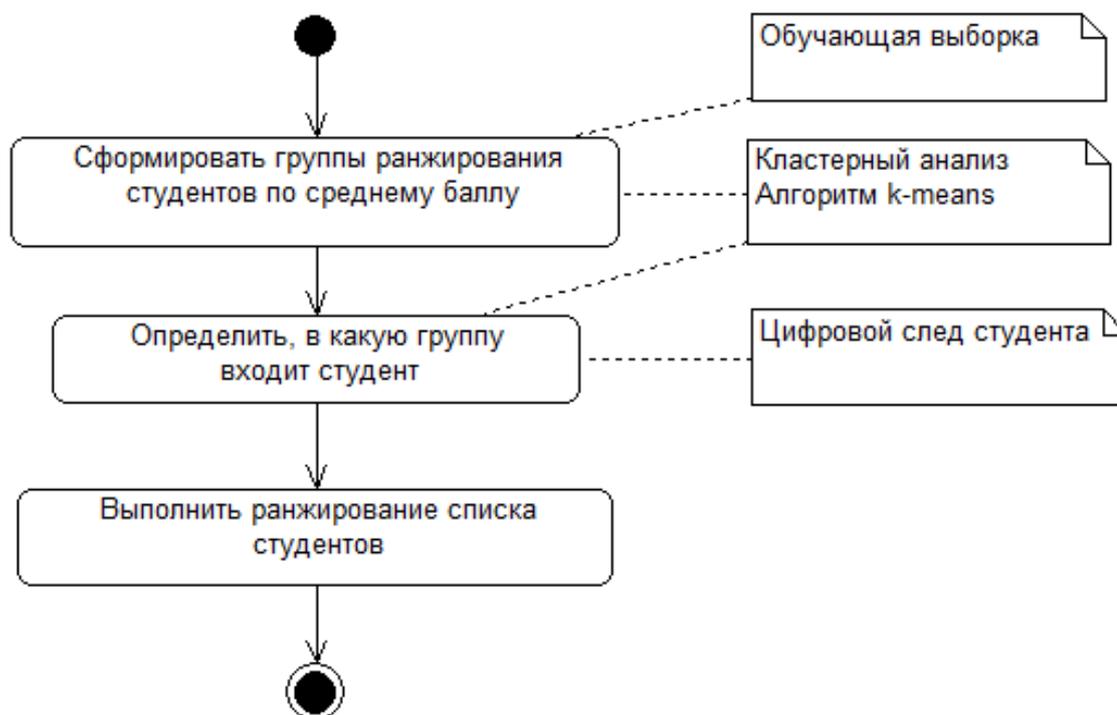


Рисунок 3 – Алгоритм ранжирование студентов по среднему баллу

Для формирования групп ранжирования используем метод кластеризации с помощью алгоритма k-means.

«Алгоритм k-means – это итеративный алгоритм, который пытается разделить набор данных на заранее определенные k -отдельных неперекрывающихся групп-кластеров, где каждая точка данных принадлежит только одной группе.

Блок-схема алгоритма k-means показана на рисунке 4.

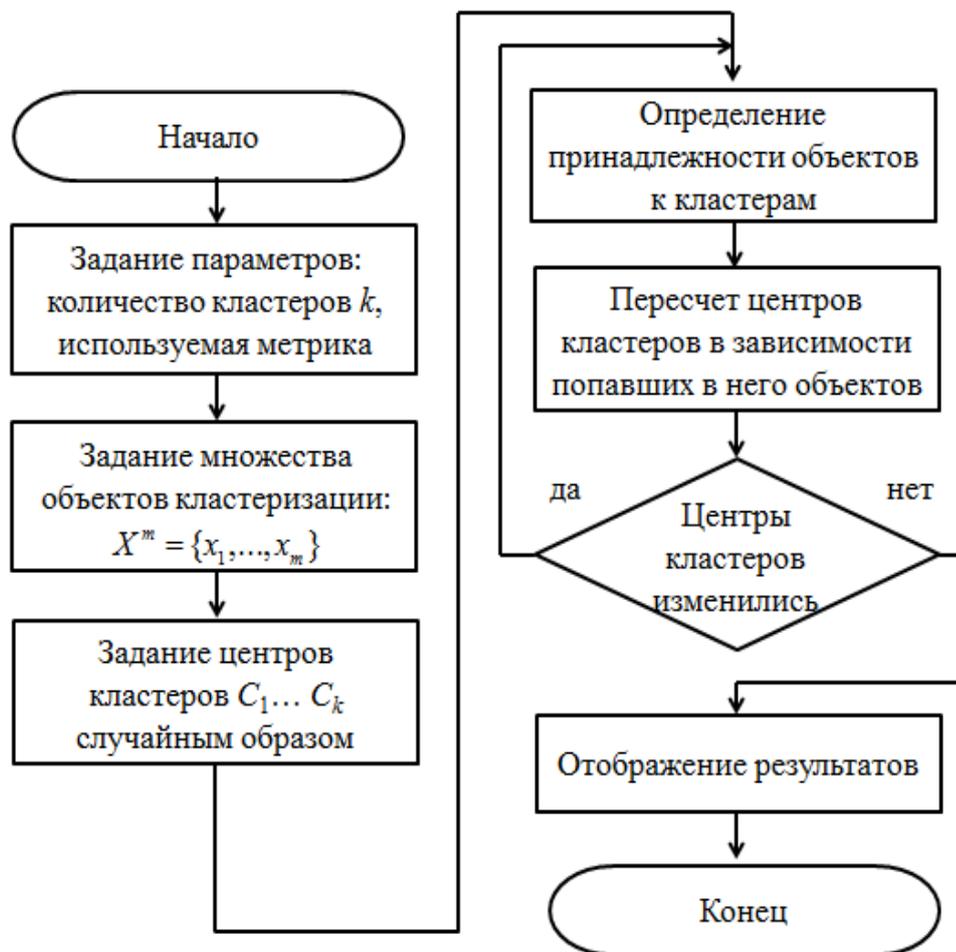


Рисунок 4 – Блок-схема алгоритма k-means

Алгоритм k-means пытается сделать точки данных внутри кластера как можно более похожими, но при этом сохраняя кластеры как можно более разными.

Он назначает точки данных кластеру таким образом, чтобы сумма квадратов расстояния между точками данных и центроидом кластера (среднем арифметическим всех точек данных, принадлежащих этому кластеру) была минимальной.

Чем меньше вариаций внутри кластеров, тем более однородные (похожие) точки данных находятся в одном кластере.

Подход, который использует k-means для решения проблемы, называется EM (Expectation-Maximization)-алгоритмом.

EM-алгоритм состоит из итерационного повторения двух шагов. На E-шаге вычисляется ожидаемое значение (expectation) вектора скрытых переменных G по текущему приближению вектора параметров Θ . На M-шаге решается задача максимизации правдоподобия (maximization) и находится следующее приближение вектора Θ по текущим значениям векторов G и Θ » [4].

«Алгоритм k-means состоит из следующих шагов:

Шаг 1. Случайным образом выбирается k объектов обучающей выборки, которые будут служить начальными центрами кластеров.

Шаг 2. Для каждого объектов обучающей выборки определяется ближайший к ней центр кластера. Для этого вычисляется расстояние между объектами и центрами кластеров. Считается, что объект принадлежит тому кластеру, к которому он ближе. В качестве формулы для оценки близости объектов в многомерном пространстве признаков используется одна из известных метрик.

Шаг 3. Как только состав кластеров на данной итерации известен, производится расчёт новых центров кластеров. Это делается путем расчета средних значений для каждого числового признака по всем объектам рассматриваемого кластера. Например, в двухмерном пространстве координаты центр кластера на основе вошедших в него t объектов рассчитывается следующим образом (15)» [13]:

$$(P_{2ц}, P_{1ц}) = \left(\frac{\sum_1^t P_1(t)}{t}, \frac{\sum_1^t P_2(t)}{t} \right). \quad (15)$$

«Шаг 4. Шаги 2 и 3 повторяются до тех пор, пока не выполнятся один из двух критериев остановки:

- границы кластеров и расположения центров кластеров не перестанет изменяться от итерации к итерации, т.е. на каждой итерации в каждом кластере будет оставаться один и тот же набор записей. На

практике алгоритм k-means обычно находит набор стабильных кластеров за несколько десятков итераций;

- достигнут критерий сходимости. Чаще всего используется критерий суммы квадратов ошибок между центром кластера и всеми вошедшими в него объектами (16):

$$E = \sum_{i=1}^k \sum_{p \in C_i} (p - m_i)^2, \quad (16)$$

где $p \in C_i$ – произвольная точка данных, принадлежащая кластеру C_i , m_i – центр данного кластера.

Иными словами, алгоритм остановится тогда, когда ошибка E достигнет достаточно малого значения.

Преимуществом алгоритма k-means является высокая производительность, поскольку все, что мы на самом деле делаем, это вычисляем расстояния между точками и центроидами групп.

Алгоритм имеет линейную сложность $O(n)$.

С другой стороны, у k-means есть недостатки:

- необходимо выбрать количество групп/классов. Это не всегда тривиально, и в идеале с алгоритмом кластеризации мы хотели бы, чтобы он выяснил это за нас, потому что его цель - получить некоторое представление о данных;
- алгоритм начинается со случайного выбора центров кластеров и, следовательно, может давать разные результаты кластеризации при разных прогонах алгоритма.

Таким образом, результаты могут быть неповторимыми и противоречивыми.

Поэтому большая область исследований в области кластеризации была сосредоточена на улучшении процесса кластеризации в том числе с помощью адаптивных алгоритмов k-means» [13].

Разработана диаграмма вариантов использования программы

формирования рейтинга студентов, показанная на рисунке 5.

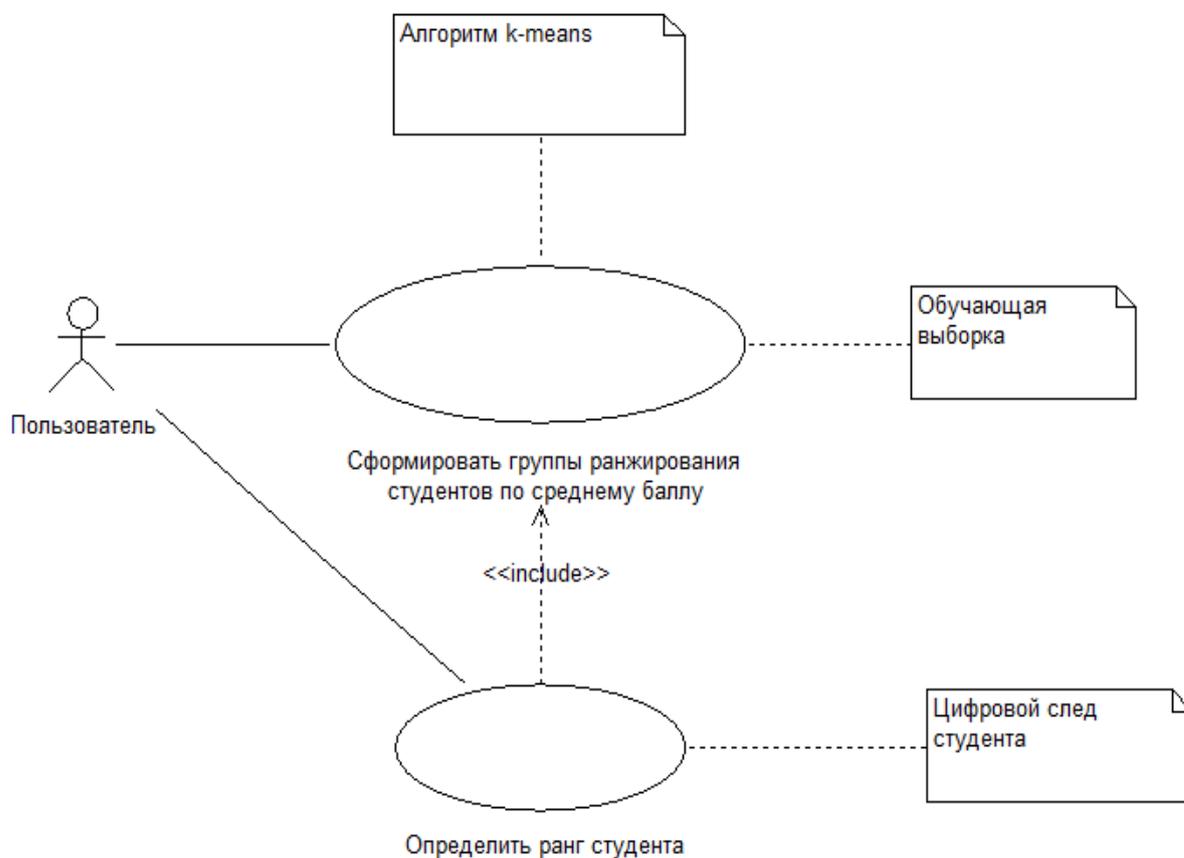


Рисунок 5 – Диаграмма вариантов использования программы формирования рейтинга студентов

Диаграмма вариантов использования отражает функционал разработанной программы.

Также разработана диаграмма классов программы, показанная на рисунке 6.

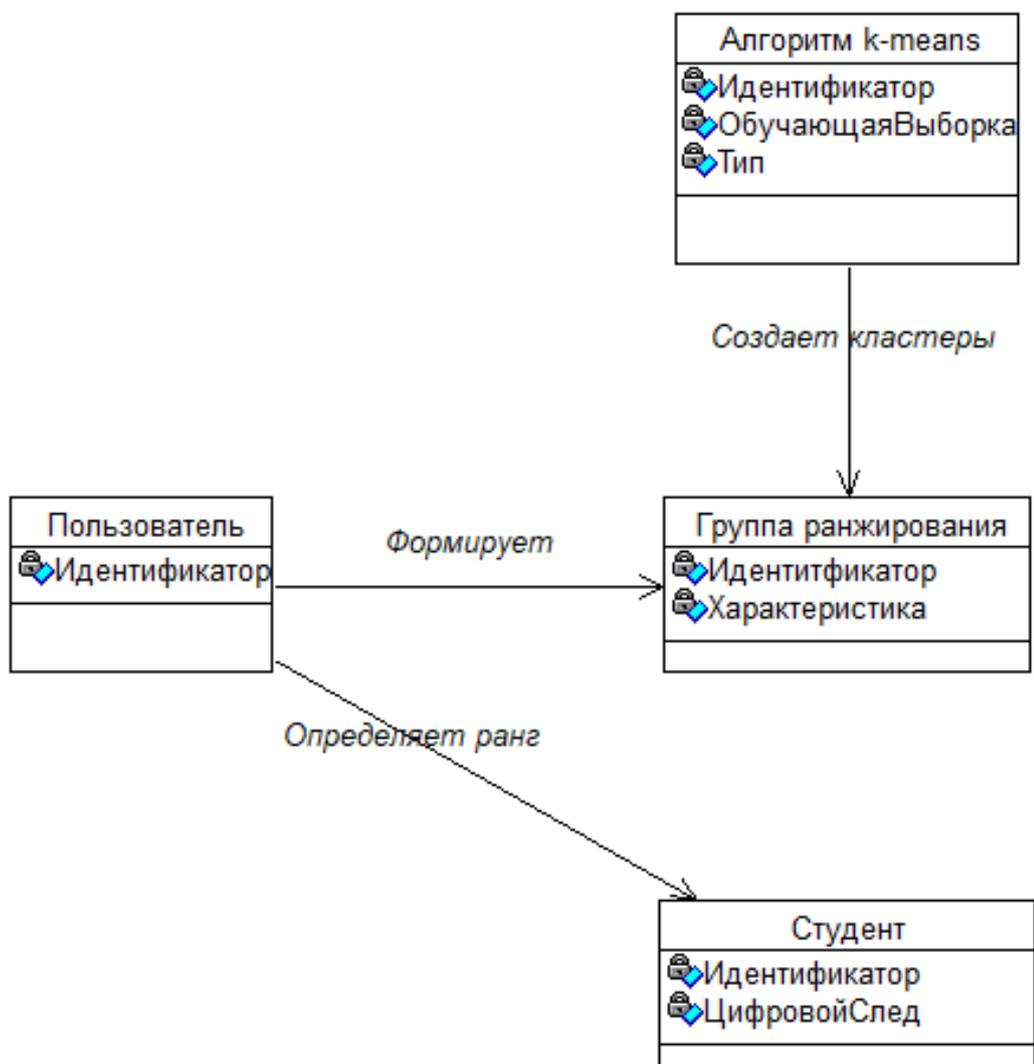


Рисунок 6 – Диаграмма классов программы формирования рейтинга студентов

Диаграмма классов изображает структуру программы.

Для реализации алгоритма машинного обучения использованы язык Python и среда Jupyter Notebook (anaconda 3) с библиотекой sklearn [15], [25].

В качестве обучающей выборки использован текстовый файл CSV, фрагмент которого показан на рисунке 7.

1	x,y	
2	3.0,3	
3	3.1,3	
4	3.2,3	
5	3.3,3	
6	3.4,3	
7	3.5,3	
8	3.6,4	
9	3.7,4	
10	3.8,4	
11	3.9,4	
12	4.0,4	
13	4.1,4	
14	4.2,4	
15	4.3,4	
16	4.4,4	
17	4.5,4	
18	4.6,5	
19	4.7,5	
20	4.8,5	
21	4.9,5	
22	5.0,5	
23	2.0,2	
24	2.1,2	
25	2.2,2	
26	2.3,2	
27	2.4,2	
28	2.5,2	

Рисунок 7 – Фрагмент обучающей выборки

Обучающая выборка создана на основе рабочей книги Excel для ранжирования студентов по следующим группам:

- отличники (средний балл 4.6-5.0);
- хорошисты (средний балл 3.6-4.0);
- троечники (средний балл 3.0-3.6);
- двоечники (средний балл 2.0-2.8).

Код алгоритма кластеризации с помощью алгоритма k-means представлен в листинге 1.

Листинг 1 – Код алгоритма кластеризации с помощью алгоритма k-means

```
from sklearn.cluster import KMeans
import pandas as pd
import numpy as np
import pickle
import pathlib

p = pathlib.Path('E:\Rank_GAP.csv')
from sklearn.cluster import KMeans

indata = pd.read_csv(p)
kmeans = KMeans(4)
kmeans.fit(indata.values)
print(kmeans.labels_)
print(kmeans.cluster_centers_)
#predicted_class = kmeans.predict([[3.0,3]])
#print(predicted_class)

import matplotlib.pyplot as plt
from matplotlib.colors import ListedColormap

%matplotlib inline

customcmap = ListedColormap(["crimson", "mediumblue", "darkmagenta"])
fig, ax = plt.subplots(figsize=(10, 8))
plt.scatter(x=indata['x'], y=indata['y'], s=200,
            c=kmeans.labels_,
            cmap=customcmap)
ax.set_xlabel(r'x', fontsize=14)
ax.set_ylabel(r'y', fontsize=14)
plt.xticks(fontsize=12)
plt.yticks(fontsize=12)
plt.show()
```

Результат кластеризации показан на рисунке 8.

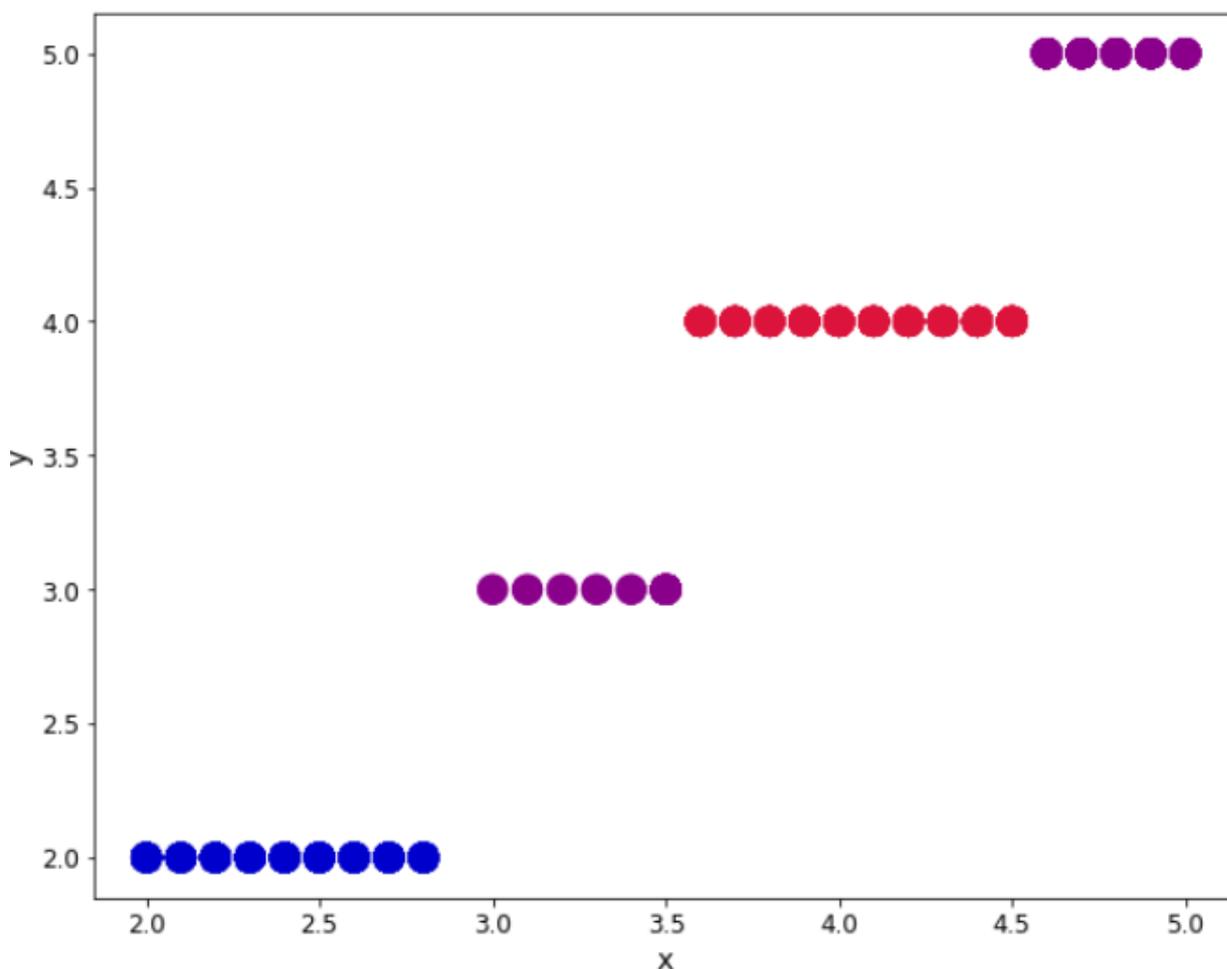


Рисунок 8 – График результата кластеризации

Далее определяем ранг студента по среднему баллу.

Определение ранга заключается в прогнозировании вхождения студента в конкретную группу ранжирования по его цифровому следу.

Для определения ранга студента разработан код, представленный в листинге 2.

Листинг 2 – Код определения ранга студента

```

from sklearn.cluster import KMeans
import pandas as pd
import numpy as np
import pickle

```

```

import pathlib
p = pathlib.Path('E:\Rank_GAP1.csv')
from sklearn.cluster import KMeans
indata = pd.read_csv(p)
kmeans = KMeans(4)
kmeans.fit(indata.values)
print(kmeans.labels_)
print(kmeans.cluster_centers_)
gpa=input()
predicted_class = kmeans.predict([[gpa]])
print(predicted_class)

```

Пример результата выполнения кода по листингу 2 показан на рисунке 9.

```

[2 2 2 2 2 2 2 2 0 0 0 0 0 0 3 3 3 3 3 3 1 1 1 1 1 1 1 1 1]
[[4.1 ]
 [2.4 ]
 [3.35]
 [4.75]]

```

```

4.2
[0]

```

Рисунок 9 – Результат определения ранга студента

Таким образом, используя цифровой след студента, можно определить его ранг и включить его в рейтинг студентов по среднему баллу успеваемости [28], [32].

Выводы по главе 3

Результаты проделанной работы позволили сделать следующие выводы:

- для формирования групп ранжирования выбран метод кластеризации с помощью алгоритма k-means, как наиболее простого в понимании реализации;
- для реализации алгоритма машинного обучения использованы язык Python и среда Jupyter Notebook (anaconda 3) с библиотекой sklearn;
- обучающая выборка создана на основе рабочей книги Excel для ранжирования студентов по группам успеваемости (отличники, хорошисты, троечники и двоечники);
- реализована и протестирована программа ранжирования студентов по среднему баллу;
- определение ранга заключается в прогнозировании вхождения студента в конкретную группу ранжирования по его цифровому следу.

Таким образом, используя цифровой след студента, можно определить его ранг и включить его в рейтинг студентов по среднему баллу успеваемости.

Заключение

Выпускная квалификационная работа посвящена актуальной теме исследования и реализации алгоритма формирования рейтинга студентов основе их цифрового следа.

Необходимым условием эффективного процесса обучения является наличие адаптивного и гибкого механизма формирования рейтинга студентов.

Таким механизмом является механизм формирования рейтинга студентов, для реализации которого используются специальные алгоритмы.

Для достижения поставленной в работе цели решены следующие задачи:

- выполнена постановка задачи исследования и проанализированы методы формирования рейтинга студентов. Как показал анализ, алгоритм формирования рейтинга студентов определяет критерии и порядок оценивания достижений студента в соответствии с выбранным методом формирования рейтинга студентов. Источником данных для формирования рейтинга студентов является его цифровой след. Уровень ранжирования используется для определения различных рейтингов, используемых институтом, на основе процента баллов по каждому предмету и на основе количества предметов. Недостатки методов расчета индивидуальных рейтингов студентов на основе учебного показателя, формирования интегральной рейтинговой оценки деятельности студента и рейтингового показателя студента обусловлены особенностями БРС. Поэтому более предпочтительным представляется метод формирования рейтинга студентов по среднему баллу;
- проанализированы методы ранжирования студентов по успеваемости. Как показал анализ, методы ранжирования используются для ранжирования элементов в наборе данных в соответствии с некоторым критерием. Методы ранжирования можно разделить на две категории: детерминированные и вероятностные. В методе

детерминированного ранжирования порядок элементов в ранжированном списке фиксирован и не меняется независимо от входных данных. В методе вероятностного ранжирования порядок элементов в ранжированном списке может меняться в зависимости от входных данных. Обучение ранжированию – класс задач машинного обучения, суть которых состоит в автоматизированном построении ограничений для ранжирующей модели по обучающей выборке, для их последующего применения к неизвестным объектам со сходной структурой. Инновационные возможности методов ранжирования на основе машинного обучения позволяет использовать указанные методы для ранжирования студентов по среднему баллу. Простота реализации метода АНР с помощью алгоритмов машинного обучения позволяет использовать указанный метод для ранжирования студентов по среднему баллу;

- реализована и протестирована программа ранжирования студентов по среднему баллу. Для формирования групп ранжирования выбран метод кластеризации с помощью алгоритма k-means, как наиболее простого в понимании реализации. Обучающая выборка создана на основе рабочей книги Excel для ранжирования студентов по группам успеваемости. Определение ранга заключается в прогнозировании вхождения студента в конкретную группу ранжирования по его цифровому следу.

Результаты бакалаврской работы представляют научно-практический интерес и могут быть рекомендованы для анализа и программной реализации методов и алгоритмов формирования рейтинга студентов основе их цифрового следа.

Список используемой литературы и используемых источников

1. Актуальная методика расчета ИПС [Электронный ресурс]. URL: https://spb.ranepa.ru/wp-content/uploads/2021/03/metodika-rascheta-irs-2021_compressed.pdf (дата обращения: 10.04.2023).
2. Алгоритм составления рейтинга студентов для назначения государственных академических стипендий в повышенном размере за учебную, научно-исследовательскую, общественную, культурно-творческую и спортивную деятельность [Электронный ресурс]. URL: https://news.sfu-kras.ru/files/algorithm_s_ispravleniyami.pdf (дата обращения: 10.04.2023).
3. Большие данные в образовании [Электронный ресурс]. URL: <https://www.unkniga.ru/vishee/9614-bolshie-dannye-v-obrazovanii.html>
<https://www.hse.ru/> (дата обращения: 10.04.2023).
4. Жданов Д.Н. Анализ успеваемости студентов для оценки деятельности куратора // Гарантии качества профессионального образования. 2010. С. 317-320.
5. Интеллектуальный анализ данных [Электронный ресурс]. URL: <https://wiki.loginom.ru/articles/data-analysis.html> (дата обращения: 10.04.2023).
6. Климович Л. А., Митющенко Е. В. Формирование интегральной рейтинговой оценки деятельности студента образовательного учреждения [Электронный ресурс]. URL: https://elar.urfu.ru/bitstream/10995/52821/1/UM_2011_6_007.pdf (дата обращения: 10.04.2023).
7. Код адаптивного алгоритма k-means на языке Python [Электронный ресурс]. URL: <https://gist.github.com/jianchao-li/7065b510305d4df48063a801e6433213#file-adaptive-kmeans-ipynb> (дата обращения: 15.03.2023).
8. Котов К., Красильников Н. Кластеризация данных [Электронный ресурс]. URL: <https://logic.pdmi.ras.ru/~yura/internet/02ia-seminar-note.pdf> (дата обращения: 17.04.2023).

9. Кулагина Е. Модель компетенций: как разработать для разных должностей [Электронный ресурс]. URL: <https://www.hr-director.ru/article/66647-model-kompetentsiy-19-m4> (дата обращения: 15.03.2023).

10. Леоненков А. В. Объектно-ориентированный анализ и проектирование с использованием UML и IBM Rational Rose : учебное пособие. М. : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. 317 с. [Электронный ресурс]. URL: <https://www.iprbookshop.ru/97554.html> (дата обращения: 24.03.2023).

11. Мазур М. Преимущества и недостатки использования балльно-рейтинговой системы в оценке учебных достижений студентов [Электронный ресурс]. URL: <https://pandia.ru/text/78/079/77773.php> (дата обращения: 10.04.2023).

12. Машинное обучение: Ранжирование [Электронный ресурс]. URL: https://edu.mmcs.sfedu.ru/pluginfile.php/18142/mod_resource/content/4/10%20%D0%A0%D0%B0%D0%BD%D0%B6%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%B8%D0%B5.pdf (дата обращения: 10.04.2023).

13. Миркин Б. Г., Орлов М. А. Методы многокритериальной стратификации и их экспериментальное сравнение. М. : Изд. дом Высшей школы экономики, 2013. 32 с.

14. Порядок организации балльно-рейтинговой системы оценки успеваемости студентов ТГУ [Электронный ресурс]. URL: https://www.tltsu.ru/sveden/document/Formi_sroki_kontrolya_26.10.17.pdf (дата обращения: 10.04.2023).

15. Стариченко Б. Е. Балльно-рейтинговая система оценивания учебной деятельности студентов: вопросы моделирования // Педагогическое образование в России. 2017. № 6. С. 205-215.

16. Степаненко А.А., Фещенко А.В. «Цифровой след студента»: поиск, анализ, интерпретация // Открытое и дистанционное образование. 2017. № 4(68). С. 58-62.

17. Студенческий рейтинг [Электронный ресурс]. URL: <https://www.hse.ru/> (дата обращения: 10.04.2023).

18. Узел Байесовская сеть [Электронный ресурс]. URL: <https://www.ibm.com/docs/ru/spss-modeler/SaaS?topic=models-bayesian-network-node> (дата обращения: 15.03.2023).

19. Узел Байесовская сеть [Электронный ресурс]. URL: <https://www.ibm.com/docs/ru/spss-modeler/SaaS?topic=models-bayesian-network-node> (дата обращения: 15.03.2023).

20. Экзамен GRE: структура и специфика [Электронный ресурс]. URL: <https://www.iqconsultancy.ru/articles/ekzamen-gre-struktura-i-spetsifika/> (дата обращения: 15.03.2023).

21. Data Mining [Электронный ресурс]. URL: <https://educationaldatamining.org/> (дата обращения: 10.04.2022).

22. ESTS Grading System [Электронный ресурс]. URL: <https://www.ensta-bretagne.fr/en/ects-grading-system> (дата обращения: 10.04.2023).

23. K-means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks [Электронный ресурс]. URL: <https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a> (дата обращения: 17.04.2023).

24. Kumar A. Ranking Algorithms and Types: Concepts and Examples [Электронный ресурс]. URL: <https://vitalflux.com/ranking-algorithms-types-concepts-examples/> (дата обращения: 10.04.2023).

25. Project Jupyter [Электронный ресурс]. URL: <https://jupyter.org/> (дата обращения: 17.04.2023).

26. Rating Assessment of Students' Knowledge vis-a-vis Rating Evaluation of University Teachers [Электронный ресурс]. URL: <https://www.igi-global.com/chapter/rating-assessment-of-students-knowledge-vis-a-vis-rating-evaluation-of-university-teachers/236035> (дата обращения: 10.04.2023).

27. Suhailan Safei et al. Targeted Ranking-Based Clustering Using AHP K-Means // Int. J. Advance Soft Compu. Appl, 2015. Vol. 7, No. 3. P. 100-113.

28. Understanding the Digital Footprint in the Classroom and Beyond [Электронный ресурс]. URL: <https://www.goguardian.com/blog/understanding-the-digital-footprint> (дата обращения: 10.04.2023).

29. What Is a GPA and Why Is It So Important? [Электронный ресурс]. URL: <https://www.mastersportal.com/articles/2126/what-is-a-gpa-and-why-is-it-so-important.html> (дата обращения: 10.04.2023).

30. What is Ranking Level? [Электронный ресурс]. URL: <https://support.fedena.com/support/solutions/articles/53900-what-is-ranking-level> (дата обращения: 10.04.2023).

31. What Is the Analytic Hierarchy Process [Электронный ресурс]. URL: <https://www.passagetechology.com/what-is-the-analytic-hierarchy-process> (дата обращения: 10.04.2023).

32. What's a digital footprint and why does it matter? [Электронный ресурс]. URL: <https://www.avg.com/en/signal/what-is-a-digital-footprint> (дата обращения: 10.04.2022).

33. Xiaomei Bai, Fuli Zhang, Jinzhou Li, Teng Guo, Abdul Aziz, Aijing Jin, Feng Xia "Educational Big Data: Predictions, Applications and Challenges", Big Data Research, Vol. 26, 2021, 100270.