

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ
(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)»**

Журнал практики

Институт № 8 «Компьютерные науки и прикладная математика»

Кафедра 806 Учебная группа М8О-408Б-20

ФИО обучающегося Зубко Дмитрий Валерьевич

Направление подготовки/ 01.03.02 Прикладная математика и информатика
специальность

шифр, наименование направления подготовки/специальности

Вид практики Преддипломная
учебная, производственная, преддипломная или другой вид практики

Оценка за практику Крылов С.С.

Москва

2024

1. Место и сроки проведения практики:

Наименование организации: ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ
БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ «МОСКОВСКИЙ
АВИАЦИОННЫЙ ИНСТИТУТ (НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)»

Сроки проведения практики

дата начала практики: 09.02.2024

дата окончания практики: 09.05.2024

2. Инструктаж по технике безопасности:

_____ /Крылов С.С. / 09.02.2024 г.
подпись проводившего расшифровка подписи дата проведения

3. Индивидуальное задание обучающегося:

Кластеризация образовательных организаций с использованием машинного обучения на основе контекстных данных

4. План выполнения индивидуального задания обучающегося:

№ п/ п	Место проведения	Тема	Период выполнения
1	Кафедра 806	Инструктаж	09.02.2024-09.02.2024
2	Кафедра 806	Изучение предметной области	10.02.2024-12.02.2024
3	Кафедра 806	Изучение задачи кластеризации	13.02.2024-01.03.2024
4	Кафедра 806	Группировка данных учащихся	06.03.2024-13.03.2024
5	Кафедра 806	Применение алгоритма кластеризации	14.03.2024-14.03.2024
7	Кафедра 806	Разделение ОО на кластера	15.03.2024-15.03.2024
8	Кафедра 806	Построение графиков	16.03.2024-19.03.2024
9	Кафедра 806	Построение тепловых карт	20.03.2024-14.04.2024
10	Кафедра 806	Анализ полученных кластеров	15.04.2024-20.04.2024
11	Кафедра 806	Создание названий кластеров	21.04.2024-30.04.2024

12	Кафедра 806	Расчет внутрекластерных корреляций	01.05.2024-05.05.2024
13	Кафедра 806	Анализ полученных результатов	05.05.2024-07.05.2024

Утверждаю

подпись руководителя от МАИ

/ Крылов С.С. /
расшифровка подписи

09.02.2024 г.
дата утверждения*

подпись руководителя от
организации/предприятия

/ Крылов С.С. /
расшифровка подписи

09.02.2024 г.
дата утверждения*

Ознакомлен

подпись обучающегося

/ Зубко Д. В. /
расшифровка подписи

09.02.2024 г.
дата ознакомления*

*Дата утверждения и ознакомления – дата начала практики

5. Отзыв руководителя практики от организации/предприятия:

Студент группы М8О-408Б-20 Зубко Дмитрий Валерьевич выполнил все задачи, поставленные индивидуальным заданием практики по теме: «Кластеризация образовательных организаций с использованием машинного обучения». Практика была пройдена в рамках кафедральной работы на кафедре 806 Федерального Бюджетного Образовательного Учреждения Высшего Образования «Московского Авиационного Института» (Национального Исследовательского Университета). В результате проделанной работы были получены кластера образовательных организаций. Полученные кластеры были подробно изучены, что позволило выявить ключевые особенности и тенденции в данных. При выполнении данной работы студент продемонстрировал все необходимые навыки и умения. Материалы, изложенные в отчёте обучающегося, полностью соответствуют индивидуальному заданию. Рекомендуемая оценка «отлично».

*подпись руководителя от
организации/предприятия*

/ Крылов С.С. /
расшифровка подписи

07.05.2024 г.
дата

6. Отчет обучающегося по практике

Данная работа является второй частью проекта «Кластеризация общеобразовательных организаций с учетом психологического портрета учащихся». Первая часть проекта включает разбор свободных ответов учащихся. В этой работе используются данные, полученные в первой части, такие как тип семьи и индексы, относящиеся к истории, культуре и семейной жизни учащихся.

В России создана Единая система оценки качества образования, известная как ЕСОКО. Эта система помогает отслеживать академические достижения учащихся на различных этапах школьного образования. Она также способствует своевременному обнаружению и решению возникающих проблем в образовательной сфере, учитывая предметные, школьные и региональные аспекты.

Эта система предоставляет полный обзор состояния образования в стране и способствует анализу разнообразных аспектов, оказывающих влияние на результаты деятельности школ. Она дает возможность школьным учреждениям самостоятельно оценивать свою деятельность и определять существующие проблемы. Кроме того, она предоставляет родителям данные о качестве знаний их детей.

Система оценки качества школьного образования в России состоит из: единого государственного экзамена (ЕГЭ), основного государственного экзамена (ОГЭ). Промежуточные срезы знаний проводятся при помощи разных работ, например, НИКО и ВПР.

Центр проведения национальных и международных исследований качества образования, который входит в состав ФИОКО, реализует на территории России НИКО, а также организует проведение всероссийских проверочных работ.

Актуальность данной работы связана с тем, что ФИОКО сможет её использовать для автоматизации анализа НИКО и ВПР.

Цель работы – применение методов машинного обучения для кластеризации образовательных организаций на основе контекстных данных с последующим анализом полученных профилей образовательных организаций. Для достижения поставленной цели в работе были решены следующие задачи:

- преобразование данных об учащихся в образовательных организациях в данные об образовательных организациях;
- дополнение данных об образовательных организациях;
- нормализация и очистка данных;
- применение выбранного метода кластеризации к данным и создание «профилей» (кластеров) образовательных организаций;

- поиск корреляций полученных профилей образовательных организаций с успеваемостью учащихся;
- создание тепловых карт для каждого кластера.

6.1 Профили образовательных организаций

Профили образовательных организаций представляют собой описание образовательных организаций, выделенных в различные кластеры на основе контекстных данных.

Имея профили образовательных организаций, предоставляется множество возможностей для работы и поддержки этих образовательных организаций:

- анализ и улучшение учебных планов: используя данные об успеваемости учащихся образовательных организаций, можно определить, какие предметы или темы требуют дополнительного внимания или изменения в методах преподавания;
- профессиональное развитие учителей: анализируя, какие образовательные организации имеют наилучшие результаты, можно организовывать обмен опытом и наставничество среди педагогического состава;
- прогнозирование успеваемости: используя аналитические инструменты, можно прогнозировать будущую успеваемость учащихся в выбранной образовательной организации, что позволяет заранее предотвращать возможные учебные трудности;
- распределение ресурсов: данные о профилях образовательных организаций могут помочь определить, в каких областях и классах может потребоваться дополнительная поддержка, что позволяет более эффективно распределять ресурсы и помощь;
- выявление резильентных школ: резильентные или устойчивые школы – это образовательные учреждения, которые показывают хорошие учебные результаты несмотря на то, что они расположены в экономически неблагополучных районах или сталкиваются с другими внешними трудностями. Сбор и анализ данных о профилях школ могут помочь выявить такие учреждения и понять, какие методы и подходы к управлению и преподаванию они используют. Это позволит распространить успешные практики среди других школ, а также дать возможность для проведения целенаправленных исследований эффективности этих методов.

6.2 Кластеризация образовательных организаций

Данные изначально распределены по нескольким датасетам, данные в датасетах неперсонифицированы. Нужно совместить эти датасеты в единый датасет, учитывая только те поля, которые необходимы для проведения кластеризации и последующего исследования.

После объединения датасетов с помощью python и pandas проведена очистка данных. Чтобы не проводить дополнительные проверки, удалено немногочисленное количество записей, в которых тип семьи содержит значение «Нет четкого ответа» или нет отметки по математике или русскому языку.

После кластеризации образовательных организаций интересно посмотреть, какие средние зарплаты у учителей образовательных организаций. Для этого были взяты данные об средних зарплатах из открытого источника и добавлены в датасет.

После получения всех необходимых данных учащихся необходимо описать образовательные организации. Перед этим удаляются данные об учащихся, в школах которых меньше 10 учащихся, чтобы полученные значения были более точными.

Для создания датасета, состоящего из образовательных организаций, нужно охарактеризовать школы по разным показателям, а затем объединить эти показатели в один датасет.

Первыми посчитаны средние оценки по математике и русскому языку и первичный балл за диагностическую работу. Для этого данные группируются по логину образовательной организации.

После группировки данных по логину образовательной организации и применения агрегирующей функции mean получился новый датасет.

Вторыми посчитаны средние значения индексов истории, культуры и семейной жизни.

Третьими посчитаны отношения типа семьи. Для этого применены агрегирующие функции во время группировки данных. Необходимо знать количество всех учащихся, учащихся с неполной, полной, многодетной полной семьями.

Для кластеризации образовательных организаций из датасета выделены поля, по которым происходит кластеризация: Average History Knowledge Index, Average Culture Index, Average Family Life Index, Отношение_неполных, Отношение_полных.

Отношение_многодетных_полных. Остальные поля, такие как оценки по математику и русскому языку и первичный балл за диагностическую работу не участвуют в кластеризации.

Далее проведена нормализация данных Average History Knowledge Index, Average Culture Index, Average Family Life Index. Для нормализации этих данных используется StandartSkaler из библиотеки sklearn. Этот скейлер реализует алгоритм z-нормализации.

Нормализация данных перед кластеризацией становится необходимой в случаях, когда признаки имеют разный масштаб или разные единицы измерения. Нормализация помогает уравнивать важность всех признаков, что особенно важно для алгоритмов кластеризации, таких как K-means или иерархической, где расстояния между точками данных используются для определения принадлежности к кластерам.

После нормализации была построена визуализация алгоритма TSNE из библиотеки sklearn. TSNE (t-distributed Stochastic Neighbor Embedding) – это алгоритм машинного обучения для визуализации, разработанный Лауренсом ван дер Маатеном и Джефффри Хинтоном. Он широко используется для уменьшения размерности больших многомерных наборов данных в 2D. Основная задача TSNE – отобразить объекты с высокоразмерного пространства на плоскость так, чтобы похожие объекты были отображены близко друг к другу, а непохожие – далеко друг от друга.

Профили образовательных организаций получены при помощи иерархической кластеризации с методом дальнего соседа.

6. 3 Вывод

Проведена работа, ориентированная на решение актуальной задачи кластеризации образовательных организаций. На языке программирования Python разработаны специализированные модули, что позволило автоматизировать процесс группировки школ.

Для аналитического обзора полученных профилей был применен комплексный подход, включающий в себя исследование основных параметров и особенностей каждого выделенного кластера. Описание составленных профилей демонстрирует разнообразие образовательной среды и способствует повышению понимания структурных особенностей школьного образования в исследуемых регионах.

Одним из наиболее информативных инструментов визуализации данных стали тепловые карты. Они визуально демонстрируют распределение школ по кластерам в пространственном измерении, что не только наглядно представляет исходные данные, но и может служить основой для дальнейших исследований и разработки рекомендаций по оптимизации работы образовательных учреждений и планированию новых.

Важной частью работы является и оценка качества проведенной кластеризации, что позволяет убедиться в эффективности выбранных методов и корректности интерпретации результатов. Возможности по улучшению методологии могут включать в себя глубокий анализ и внедрение различных метрик для оценки качества кластеров, а также использование дополнительных данных и параметров.

Таким образом, работа несет в себе значительный аналитический потенциал и может быть использована как в органах управления образованием, так и при разработке стратегий развития образовательной инфраструктуры на муниципальном или региональном уровне.

Работа имеет потенциал для дальнейшего расширения. Расширение диапазона данных, использованных при кластеризации, может способствовать улучшению детализации и разнообразия получаемых профилей образовательных учреждений. Кроме того, с применением технологий искусственного интеллекта можно не только группировать школы по кластерам, но и осуществлять более глубокое описание и анализ индивидуальных характеристик каждого учебного заведения.

подпись обучающегося

/Зубко Д. В. /
расшифровка подписи

07.05.2024 г.
дата