Пензенский государственный университет

Факультет вычислительной техники

Кафедра «Вычислительная техника»

**Отчёт**

по лабораторной работе № 4

## на тему «Изучение основных принципов построения обучаемых алгоритмов»

по дисциплине «Программные средства кибертехнических систем»

Вариант № 6

выполнили: ст-ты гр. 22ВВП1

Захаров А. С.

Беляев Д. И.

Демин М. С

Проверили:

Карамышева Н.С.

Зинкин С.А.

2025

**Цель:** изучить основные принципы построения обучаемых алгоритмов

**Задание:**

1.Получить описание заданной предметной области и набор соответствующих ей данных.

2. Провести исследование в соответствии с планом работы в рамках первой фазы исследования.

3. Подготовить отчёт по результатам первой фазы работ.

Отчёт по изучению материала по теме 1.1 должен содержать информацию в соответствии со следующей структурой:

1. Бизнес цель.

2. Ресурсы и риски.

3. Цель анализа данных.

4. План проекта.

Вариант 6

#### Задача определения качества преподавания предмета ассистентом преподавателя (Class attribute: 6-й столбец набора дан ных {1 – Низкое, 2 – Среднее, 3 – Высокое}) по пяти атрибутам:

#### Родной английский язык {1 – Родной, 2 – Не родной}

#### Номер преподавателя {25 категорий}

#### Номер предмета {26 категорий}

#### Тип семестра {1 – Летний, 2 – Обычный}

#### Размер группы {численное значение}

#### Ссылка на репозиторий:

#### <https://archive.ics.uci.edu/ml/datasets/Teaching+Assistant+Evaluation>

**1. Бизнес-цель**

Целью исследования является определение качества преподавания предмета ассистентом преподавателя на основе пяти атрибутов. Данный анализ позволит выявить закономерности и определить, какие факторы оказывают наибольшее влияние на уровень восприятия материала студентами. Это поможет в принятии решений по улучшению образовательного процесса и оптимизации преподавания. Кроме того, результаты анализа могут быть использованы для оценки эффективности различных методик преподавания и разработки рекомендаций по их усовершенствованию.

**2. Ресурсы и риски**

**2.1 Доступные ресурсы**

* Источники данных: информация о преподавателях, предмете, размере группы, типе семестра и родном языке студентов. Эти данные позволят провести комплексный анализ факторов, влияющих на качество преподавания.
* Технические ресурсы: вычислительные мощности, программное обеспечение для анализа данных, такие как Python, R, SQL и специализированные библиотеки машинного обучения (Scikit-learn, TensorFlow и др.).
* Человеческие ресурсы: специалисты по анализу данных, преподаватели, администрация учебного заведения, а также студенты, предоставляющие обратную связь.

**2.2 Возможные риски**

* Неполнота или низкое качество данных: возможны пропуски в записях, ошибки при вводе данных, а также ограниченность выборки.
* Дисбаланс классов в данных: неравномерное распределение меток качества преподавания может привести к смещенным результатам модели.
* Субъективность оценок: качество преподавания может оцениваться студентами субъективно, что приведет к искажениям в данных.
* Ограниченность вычислительных мощностей: сложные модели требуют значительных ресурсов, что может затруднить их обучение.
* Возможное отсутствие значимых закономерностей в данных: если данные не содержат достаточной информации для выявления полезных закономерностей, эффективность моделей может быть низкой.

**3. Цель анализа данных**

* Определить наиболее значимые факторы, влияющие на качество преподавания, и выявить ключевые зависимости между параметрами преподавания и оценками студентов.
* Построить модель, способную предсказывать уровень качества преподавания, что позволит автоматизировать процесс оценки и выявления проблемных зон.
* Выбрать подходящую метрику оценки качества модели (например, точность, F1-мера, ROC-AUC) для объективного сравнения различных методов анализа данных.
* Установить пороговое значение точности, обеспечивающее практическую применимость модели в образовательных процессах и позволяющее принимать обоснованные решения по оптимизации преподавания.

**4. План проекта**

Проект строится в соответствии с методологией CRISP-DM и включает следующие этапы:

4**.1 Изучение данных (Data Understanding)**

1. Сбор данных из доступных источников, включая базы данных университета, опросы студентов и административные записи.
2. Описание данных (число записей, атрибуты, статистики по признакам, распределение значений).
3. Исследование данных: анализ возможных закономерностей, выявление пропусков, оценка корреляций между переменными, построение графиков распределения данных.
4. Документирование процесса работы с данными: фиксирование всех этапов, включая выявленные проблемы, принятые решения и предварительные выводы.

**4.2 Подготовка данных (Data Preparation)**

1. Отбор релевантных признаков, влияющих на целевую переменную.
2. Очистка данных: удаление выбросов, обработка пропущенных значений, устранение ошибок в данных.
3. Преобразование данных: нормализация числовых переменных, кодирование категориальных признаков, создание новых переменных при необходимости.
4. Генерация новых признаков: применение методов feature engineering для улучшения предсказательной способности модели.

**4.3 Моделирование (Modeling)**

1. Выбор алгоритмов: логистическая регрессия, деревья решений, случайный лес, градиентный бустинг, нейронные сети и другие подходы.
2. Разделение данных на обучающую, валидационную и тестовую выборки с целью предотвращения переобучения.
3. Обучение моделей и их оценка по выбранным метрикам.
4. Анализ важности признаков: определение факторов, наиболее влияющих на предсказания модели.

**4.4 Оценка (Evaluation)**

1. Оценка точности моделей с точки зрения бизнеса: насколько предсказания модели соответствуют реальным оценкам качества преподавания.
2. Анализ результатов: выявление сильных и слабых сторон моделей, анализ возможных источников ошибок.
3. Выбор наилучшей модели, обеспечивающей баланс между точностью, интерпретируемостью и вычислительной эффективностью.

**4.5 Внедрение (Deployment)**

1. Разработка прототипа приложения или интеграция модели в существующие системы.
2. Мониторинг качества модели в реальных условиях.
3. Обновление модели в случае изменения стандартов оценки качества преподавания или появления новых данных.

## **Критерии устаревания модели**

1. **Снижение точности предсказаний**: если модель начинает показывать результаты ниже установленного порога качества, ее необходимо пересмотреть.
2. **Изменение характеристик данных**: появление новых факторов, влияющих на качество преподавания, или изменения в образовательном процессе могут сделать модель устаревшей.
3. **Обнаружение новых методик анализа**: развитие машинного обучения и появление более эффективных алгоритмов могут потребовать обновления модели.
4. **Обратная связь от пользователей**: если преподаватели или студенты отмечают снижение точности оценок модели, необходимо провести повторный анализ.
5. **Регулярный мониторинг производительности**: периодическая проверка модели на актуальность и корректность ее работы позволит своевременно выявлять устаревание и принимать меры по её обновлению.

**Вывод:** изучили основные принципы построения обучаемых алгоритмов, установили цели и задачи, разработали план проекта.