
Искусственный интеллект (Передовые технологии). Обучение в записи

Урок 6. Семинар. Создание проекта машинного обучения

Оглавление

1. Подумайте о реальной проблеме, которую, по вашему мнению, можно решить с помощью машинного обучения. Четко определите эту проблему и поставьте SMART-цели для ее решения. Помните, что это должна быть проблема, в которой машинное обучение может помочь сделать прогноз, классифицировать данные или дать рекомендации.....	2
2. Определите потенциальные источники данных, которые могут быть использованы для решения вашей проблемы. Какие данные вам понадобятся? Где вы можете их получить? Напишите краткое резюме вашей стратегии сбора данных.	3
3. Даже если у вас нет реальных данных для анализа, представьте, что они у вас есть. Как бы вы стали исследовать и понимать эти данные? Что бы вы искали? Напишите краткое описание вашей стратегии анализа исследовательских данных.	3
4. Вспомните различные модели машинного обучения, которые мы обсуждали, в частности модель "Дерево решений". Какой тип модели машинного обучения будет наиболее подходящим для вашей проблемы и почему? Объясните свой выбор.	5
5. Напишите краткое описание того, как вы будете обучать и оценивать выбранную вами модель машинного обучения, используя воображаемый набор данных.	5
6. Наконец, подумайте о том, как вы могли бы развернуть свою обученную модель. Где она будет использоваться? Кто будет ее использовать? Какую пользу она принесет им? Напишите краткое резюме вашей стратегии развертывания.	6
Список источников	7

<p><i>Помните, что целью этого задания является не создание реальной модели машинного обучения (пока!), а применение концепций, которые вы изучили на этом семинаре, к реальной проблеме.</i></p>

1. Подумайте о реальной проблеме, которую, по вашему мнению, можно решить с помощью машинного обучения. Четко определите эту проблему и поставьте SMART-цели для ее решения. Помните, что это должна быть проблема, в которой машинное обучение может помочь сделать прогноз, классифицировать данные или дать рекомендации.

Проблема: прогнозирование оценки деятельности образовательных организаций в современных условиях.

За годы работы образовательная организация накапливает данные о своих обучающихся: их пол, возраст, успеваемость по различным предметам, результаты олимпиады и внеучебные достижения, посещаемость, какие типы заданий давались лучше или хуже, какие преподаватели вели курс.

Правильный анализ такой информации может помочь составить более полную картину процесса обучения, выявить полезные и, возможно, неочевидные связи: как уровень первоначальной подготовки влияет на успеваемость по конкретному предмету, зависит ли успех в освоении дисциплины от пола, посещаемости или преподавателя, обучающиеся каких преподавателей показывают лучшие результаты.

Методы машинного обучения могут позволить предсказать исход какой-либо ситуации, основываясь на аналитических данных. В отличие от традиционных мер измерения результатов учащихся применение методов машинного обучения может помочь педагогам и исследователям получить ценную информацию о том, как улучшить и персонализировать обучение, составлять прогнозы и рекомендации, проводить изменения в режиме реального времени, когда это имеет смысл и необходимо.

SMART-цели для решения:

- 1) **Конкретная (Specific):** Разработать модель машинного обучения, способную выявить учащихся из группы риска, прогнозировать успешность в освоении дисциплин каждым обучающимся. Система оповещает учителя о трудностях в понимании материала.
- 2) **Измеримая (Measurable):** Достичь точности прогноза уровня освоения дисциплин обучающимися не менее 90% по сравнению с фактическими данными в тестовом наборе данных. Программа может давать рекомендации по конкретным вмешательствам, которые должен использовать учитель, чтобы адаптировать обучающегося.
- 3) **Достижимая (Achievable):** ИИ помогает автоматизировать часть рутинной работы преподавателя, освобождая ему время для более важных или творческих задач.
- 4) **Релевантная (Relevant):** ИИ может анализировать демографические данные, данные о посещаемости, успеваемости, регистрации, социально-эмоциональном обучении, финансовые обстоятельства, семейная ситуация, мотивация и т. д., а также профессиональные навыки учителей.

5) **Временная (Time-bound):** Завершить разработку и тестирование модели за 2-4 месяца до окончания учебного года.

2. Определите потенциальные источники данных, которые могут быть использованы для решения вашей проблемы. Какие данные вам понадобятся? Где вы можете их получить? Напишите краткое резюме вашей стратегии сбора данных.

Для построения модели необходимо использовать данные об обучающихся, которые хранятся в образовательных учреждениях:

- Фамилия, имя и отчество
- Успеваемость по предметам (для решения поставленной проблемы возможно использовать данные не по всем, а по наиболее сложным предметам)
- Уровень качества знаний по результатам контрольных работ и промежуточной аттестации
- Участие обучающихся в олимпиадах и внеурочной деятельности по предмету
- и т.д.

Для построения модели необходимо использовать данные о преподавателях, которые хранятся в образовательных учреждениях:

- Фамилия, имя и отчество
- Образование
- Возраст, стаж работы по специальности
- Уровень квалификации

3. Даже если у вас нет реальных данных для анализа, представьте, что они у вас есть. Как бы вы стали исследовать и понимать эти данные? Что бы вы искали? Напишите краткое описание вашей стратегии анализа исследовательских данных.

Обработка данных. Здесь решается вопрос, какие методы будут использованы для обработки данных, – нормирование, стандартизация, преобразование, агрегирование, а также создаются новые показатели или выбираются самые значимые показатели для использования в модели машинного обучения.

Подготовка данных – это процесс подготовки «сырых» (необработанных) данных для их дальнейшей обработки и анализа.

Предобработка данных включает следующие процедуры:

- проверка данных;
- очистка данных;

- трансформация данных;
- трансформация данных;
- дополнение;
- оптимизация.

Проверка данных включает выявление:

- дубликатов, противоречий, ошибок;
- аномальных наблюдений;
- пропусков.

Очистка данных содержит:

- устранение дубликатов, противоречий и ошибок;
- обработку аномальных наблюдений;
- обработку пропусков.

При решении различных задач следует обрабатывать пропуски. Их можно удалять, игнорировать или заполнять. Однако нужно осторожно использовать удаление пропусков. Если они составляют большую долю по отношению к объему всей выборки данных, то прибегать к удалению пропусков не стоит.

Заполнять пропуски можно с помощью:

- нулей;
- моды, медианы или среднего значения;
- индикаторных переменных.

Трансформация данных включает:

- переименование признаков;
- сортировка, группировка данных;
- кодирование переменных;
- нормировка данных.

Дополнение данных подразумевает создание новых признаков, агрегирование признаков.

Оптимизация данных включает:

- снижение размерности;
- выявление и исключение незначительных признаков.

Подготовка данных может занимать до 80 % времени, затрачиваемого на проект машинного обучения. Чтобы оптимизировать этот процесс, важно использовать инструменты, специально разработанные для подготовки данных.

4. Вспомните различные модели машинного обучения, которые мы обсуждали, в частности модель "Дерево решений". Какой тип модели машинного обучения будет наиболее подходящим для вашей проблемы и почему? Объясните свой выбор.

На этом этапе создается математическая или компьютерная модель образовательного процесса. В модели нужно определить основные элементы образовательной системы, их взаимодействие и зависимости, а также параметры и переменные, которые будут использоваться для анализа и прогнозирования.

Для обучения модели применен алгоритм множественной линейной регрессии. Для этого мы разделим наш набор данных на две части: обучающую и тестовую выборки. Обучающая выборка будет использоваться для обучения модели, а тестовая выборка – для проверки ее точности.

Алгоритм множественной линейной регрессии – это модель машинного обучения, которая используется для предсказания значения некоторой целевой переменной на основе значений набора независимых переменных (признаков). Цель состоит в том, чтобы найти линейную связь между независимыми переменными и целевой переменной.

5. Напишите краткое описание того, как вы будете обучать и оценивать выбранную вами модель машинного обучения, используя воображаемый набор данных.

1. Настройка модели. На этом этапе производится настройка параметров модели, чтобы она соответствовала реальной образовательной системе. Калибровка может осуществляться путем анализа и сопоставления моделируемых данных с реальными данными.
2. Валидация модели. На этом этапе проверяется точность и достоверность модели. Это может быть достигнуто путем сравнения результатов моделирования с реальными результатами образовательного процесса.
3. Анализ и интерпретация результатов моделирования. На этом этапе производится анализ моделирования и интерпретация полученных результатов. Может проводиться комплексный анализ различных параметров и переменных модели, чтобы выявить влияние различных факторов на процесс обучения.
4. Прогнозирование и оптимизация. На этом этапе используются результаты моделирования для прогнозирования будущих изменений и оптимизации образовательного процесса. На основе моделирования могут быть предложены конкретные меры и стратегии для улучшения образовательной системы.

6. Наконец, подумайте о том, как вы могли бы развернуть свою обученную модель. Где она будет использоваться? Кто будет ее использовать? Какую пользу она принесет им? Напишите краткое резюме вашей стратегии развертывания.

1. Внедрение и управление изменениями. На этом этапе разрабатываются и внедряются изменения в образовательный процесс на основе результатов моделирования. Это может включать изменение программы обучения, методов и форм обучения, использование новых технологий и т.д.
2. Оценка и обновление модели. После внедрения изменений необходимо провести оценку и обновление модели. Это может включать сбор новых данных, анализ изменений в образовательном процессе и обновление модели с учетом полученной информации.

Каждый этап моделирования образовательного процесса является важным и выполняется последовательно для достижения желаемых результатов.

К преимуществам моделирования образовательного процесса можно отнести следующее:

1. Улучшение понимания. Моделирование позволяет участникам образовательного процесса лучше понять и увидеть взаимосвязи между различными аспектами обучения и образования. Это помогает учащимся и преподавателям видеть, как различные факторы влияют на образовательные результаты.
2. Планирование и прогнозирование. Моделирование образовательного процесса позволяет разработать детальные планы занятий и курсов, а также прогнозировать результаты обучения. Это позволяет лучше организовать и структурировать образовательную программу.
3. Оптимизация ресурсов. Моделирование позволяет оптимизировать использование ресурсов, таких как время, деньги и персонал. При помощи моделирования можно определить оптимальные стратегии, например, распределение учебных нагрузок или использование различных образовательных технологий.
4. Прогнозирование результатов. Моделирование помогает прогнозировать результаты обучения, позволяя участникам образовательного процесса более рационально планировать свои образовательные цели и стратегии достижения этих целей.
5. Инновации и улучшение качества образования. Моделирование позволяет проводить эксперименты и тестирования различных образовательных методик и подходов, что может привести к улучшению качества образования и инновациям в образовательной сфере.
6. Анализ данных. Моделирование образовательного процесса позволяет собирать и анализировать данные, что помогает выявить тренды и закономерности в образовательной сфере. Это может способствовать принятию обоснованных решений и улучшению образовательной практики.

В целом, моделирование образовательного процесса предлагает множество преимуществ, которые помогают более эффективно организовать и улучшить образовательное обучение и достичь лучших результатов для учащихся.

Список источников

1. Платформа Experfy.
Статья: Применение машинного обучения в образовании.
Автор: Рич Хьюбнер
URL: <https://resources.experfy.com/ai-ml/applications-of-machine-learning-in-education/#:~:text=Personalized%20learning,is%20known%20as%20adaptive%20learning>.
2. Известия Кабардино-Балкарского научного центра РАН № 3(113) 2023
Научно-исследовательская статья: СИСТЕМНЫЙ АНАЛИЗ, УПРАВЛЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ
URL: <https://cyberleninka.ru/article/n/primenenie-metodov-mashinnogo-obucheniya-pri-otsenke-deyatelnosti-obrazovatelnoy-organizatsii-vysshey-shkoly/viewer>
3. Менеджмент образовательных организаций
Научно статья: ИСПОЛЬЗОВАНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗА УСПЕВАЕМОСТИ УЧАЩИХСЯ ОСНОВНОЙ ШКОЛЫ
URL: <https://cyberleninka.ru/article/n/ispolzovanie-algoritmov-mashinnogo-obucheniya-dlya-prognoza-uspevaemosti-uchaschihsya-osnovnoy-shkoly/viewer>
4. Мир технологий.
Статья: Что такое моделирование образовательного процесса?
URL: https://dzen.ru/a/ZWWmh9e_O0-QMA8H