

МИНИСТЕРСТВО ПРОСВЕЩЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение высшего
образования
«КРАСНОЯРСКИЙ ГОСУДАРСТВЕННЫЙ ПЕДАГОГИЧЕСКИЙ
УНИВЕРСИТЕТ им. В.П. Астафьева» (КГПУ им. В.П. Астафьева)

Институт математики, физики и информатики

Кафедра Информатики и информационных технологий в образовании

Лебедева Виктория Олеговна

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

**ОРГАНИЗАЦИОННО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ
КУРСА ПО ВЫБОРУ
«ВВЕДЕНИЕ В КЛАССИЧЕСКОЕ МАШИННОЕ ОБУЧЕНИЕ»**

Направление подготовки: 44.03.05 Педагогическое образование

Направленность (профиль) образовательной программы:
Математика и информатика

ДОПУСКАЮ К ЗАЩИТЕ

Зав.кафедрой:
д-р пед. наук, профессор
Пак Н.И.

(дата, подпись)

Руководитель :
канд. пед. наук, доцент
Дорошенко Е.Г.

(дата, подпись)

Дата защиты _____
Обучающийся:
Лебедева В.О.

(дата, подпись)

Красноярск, 2021

Содержание

Введение.....	3
Глава I. Теоретические основы изучения курса по выбору «Введение в классическое машинное обучение» в основной школе.....	7
1.1. Искусственный интеллект и машинное обучение как сквозная технология цифровой экономики	7
1.2. Классическое машинное обучение: данные, параметры, методы	11
1.3. Анализ существующих средств обучения школьников по теме «Введение в классическое машинное обучение»	21
Выводы по 1 главе	27
Глава II. Разработка курса по выбору «Введение в классическое машинное обучение» для обучающихся основной школы.....	28
2.1. Программа курса по выбору «Введение в классическое машинное обучение»	28
2.2. Методические рекомендации по изучению курса по выбору «Введение в классическое машинное обучение»	34
2.3. Апробация курса по выбору и анализ ее результатов.....	74
Выводы по 2 главе	77
Заключение	78
Список использованных источников	80
Приложения	84

Введение

Мы живем в эру стремительно развивающихся информационных технологий. И немалая заслуга в этом одного из самых перспективных направлений развития ИТ-индустрии — искусственного интеллекта.

Искусственный интеллект сегодня плотно вошел в жизнь каждого жителя развитых стран, и необходимость понимания, что это такое и для чего применяется, стала одной из особенностей современного общества. Эта научная сфера и область разработки высокоинтеллектуальных технологий определена в качестве приоритетного направления развития Российской Федерации [1].

Можно сказать, что будущий потенциал любого государства будет определяться тем профессиональным выбором, который сделают сегодняшние школьники. Обучающиеся должны быть мотивированы к выбору ИТ-профессий, понимать их востребованность, осознавать стратегическую важность для личностного самоопределения, чтобы быть конкурентоспособными на рынке труда будущего.

И конечно, школа не должна оставаться в стороне от решения проблем, актуальных для нашей страны, общества в целом и будущего выпускников.

В школьных учебниках по информатике из федерального перечня учебников для основного общего образования: Босова Л.Л., Босова А.Ю., Поляков К.Ю., Еремин Е.А., Семакин И.Г., Залогова Л.А., Русаков С.В., Шестакова Л.В., Гейн А.Г., Юнерман Н.А., теме «Введение в искусственный интеллект и машинное обучение» уделяется недостаточно внимания, ни в теоретическом плане, ни в практическом [27].

В сети Интернет существуют ресурсы для школьников и учителей, предназначенные для начального ознакомления с технологиями искусственного интеллекта. Например, всероссийский образовательный проект в сфере информационных технологий «Урок цифры» [1]. Проект позволяет школьникам, не выходя из дома, знакомиться с основами цифровой экономики, цифровых технологий и программирования. Для

формирования уроков, доступных на сайте проекта, используются образовательные программы в области цифровых технологий от таких компаний, как «Яндекс», Mail.ru, «Лаборатория Касперского», «Сбербанк», «1С». Занятия на тематических тренажёрах проекта «Урок цифры» реализованы в виде увлекательных онлайн-игр и адаптированы для трёх возрастных групп — учащихся младшей, средней и старшей школы. Вместе с «Уроком цифры» школьники могут узнать о принципах искусственного интеллекта и машинном обучении, больших данных, правилах безопасного поведения в интернете и др [8].

Издательство «Просвещение» и Фонд Сбербанка «Вклад в будущее» приглашают учителей математики и информатики, а также обучающихся присоединиться к проекту «Академия искусственного интеллекта» [8]. Это просветительский проект Сбербанка, направленный на формирование интереса учащихся 7–11-х классов к технологиям искусственного интеллекта, машинного обучения и анализа данных, а также развитие их компетенций в программировании и математике. На образовательной платформе AI-Academy собраны уроки и курсы, разработанные экспертами Академии и полностью адаптированные для самостоятельного прохождения обучающимися, а также задачи всех прошедших соревнований для прокачки полученных знаний [8].

Однако эти просветительские образовательные проекты представляют из себя разработки отдельных разрозненных занятий.

Таким образом, анализ методической литературы и информационных ресурсов для школьников по тематике, связанной с изучением технологий искусственного интеллекта, позволил выделить противоречие между важностью формирования представления у обучающихся о технологиях искусственного интеллекта и сферах его применения на этапе обучения в основной школе и недостаточной разработанностью необходимого для этого организационно-методического обеспечения.

Указанное противоречие определило **проблему исследования**: какими должны быть структура, содержание и методическое обеспечение курса по выбору для начального ознакомления обучающихся основной школы с основными методами машинного обучения, лежащими в основе технологий искусственного интеллекта?

Объект исследования: процесс обучения школьников основам технологий искусственного интеллекта.

Предмет исследования: организационно-методическое обеспечение курса по выбору «Введение в классическое машинное обучение» для обучающихся основной школы.

Цель: разработка организационно-методического обеспечения курса по выбору «Введение в классическое машинное обучение» для обучающихся основной школы и методических рекомендаций по его изучению.

Задачи исследования:

1. На основе анализа литературы раскрыть понятие искусственного интеллекта и машинного обучения.
2. Рассмотреть теоретические основы классического машинного обучения.
3. Проанализировать имеющиеся в свободном доступе средства обучения для обучения школьников основам классического машинного обучения.
4. Разработать программу курса по выбору «Введение в классическое машинное обучение», основываясь на выявленном ресурсном обеспечении.
5. Разработать методические рекомендации по изучению курса по выбору «Введение в классическое машинное обучение»
6. Частично апробировать курс по выбору и проанализировать результаты апробации.

Практическая значимость исследования: разработки занятий и методические рекомендации по их проведению могут быть использованы на

уроках информатики, на занятиях курсов по выбору, связанных с изучением технологий искусственного интеллекта.

Глава I. Теоретические основы изучения курса по выбору «Введение в классическое машинное обучение» в основной школе

1.1. Искусственный интеллект и машинное обучение как сквозная технология цифровой экономики

Начало XXI века ознаменовалось прорывным развитием цифровых технологий. Они изменили и облегчили различные сферы, открыли новые рыночные возможности. Современный мир уже невозможно представить себе без информационных технологий. ИТ-индустрия существенно преобразует условия труда и жизни человека. И немалая заслуга в этом одного из самых перспективных направлений развития ИТ-индустрии — искусственного интеллекта.

В настоящее время понятие «искусственный интеллект» представляет собой более емкое понятие, вбирающее в себя и современные технологии разработки интеллектуальных систем. Сегодня эта научная сфера и область разработки высокоинтеллектуальных технологий определена в качестве приоритетного направления развития Российской Федерации.

Указом Президента РФ от 10 октября 2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации» утверждена Национальная стратегия развития искусственного интеллекта на период до 2030 года [26].

Несмотря на все попытки дать точное определение понятию «искусственный интеллект», строгого определения не существует. В Национальной стратегии развития искусственного интеллекта на период до 2030 года используются следующие понятие: «искусственный интеллект — комплекс технологических решений, позволяющий имитировать когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека. Комплекс технологических решений включает в себя информационно-коммуникационную инфраструктуру, программное

обеспечение (в том числе, в котором используются методы машинного обучения), процессы и сервисы по обработке данных и поиску решений» [26]. Мы определим искусственный интеллект (ИИ) как научную область, занимающуюся созданием программ и устройств, имитирующих интеллектуальные функции человека [8].

Искусственный интеллект принято делить на две большие группы: сильный ИИ и слабый ИИ. Под слабым ИИ — понимается алгоритм, способный решать узкоспециализированные задачи [12]. Сильный ИИ — схож с человеческим интеллектом, иными словами, он может успешно выполнять любые умственные задачи, которые под силу людям [23].

В настоящее время все существующие формы искусственного интеллекта ограниченные, то есть слабые. Они способны решать исключительно конкретные прикладные задачи и не могут сравниться с человеческим разумом универсальности в отличие от сильного искусственного интеллекта.

Слабый искусственный интеллект мы часто называем «машинным обучением». Машинное обучение (Machine Learning или МО) — обширный подраздел искусственного интеллекта, изучающий методы построения алгоритмов, способных обучаться [20].

Машинное обучение является одним из направлений искусственного интеллекта. Основной принцип заключается в том, что машины получают данные и «обучаются» на них. В настоящее время это наиболее перспективный инструмент для общества, основанный на искусственном интеллекте. Системы машинного обучения позволяют быстро применять знания, полученные при обучении на больших наборах данных, что позволяет им преуспевать в различных задачах. В отличие от программ с закодированными вручную инструкциями для выполнения конкретных задач, машинное обучение позволяет системе научиться самостоятельно распознавать шаблоны и делать прогнозы.

Технологии искусственного интеллекта уже применяются в различных сферах жизни человека. Например, *в сельском хозяйстве* многие компании разрабатывают и тестируют беспилотные тракторы. Беспилотные транспортные средства помогут повысить эффективность работ в секторе сельского хозяйства. Уже используются устройства LettuceBot, разработанные компанией Blue River Technology, которые, используя технологию машинного обучения, распознают среди здоровых растений сорняки и опрыскивают их гербицидами [14].

Появилось множество приборов для дома, начиная от умных лампочек и заканчивая умными пылесосами, холодильниками и т.д. На сегодняшний день существуют разнообразные устройства, успешно использующие искусственный интеллект *для умного дома*: домашние роботы, ассистенты и камеры безопасности. Если вы используете смартфон, вы взаимодействуете с ИИ, знаете ли вы это или нет. От очевидных функций ИИ, таких как встроенные умные помощники, до не столь очевидных, таких как портретный режим в камере [2].

В образовании появились интеллектуальные обучающие системы, благодаря внедрению искусственного интеллекта. Эти программы симулируют поведение учителя. Они могут проверять уровень знаний обучающихся, анализируя их ответы, давать отзывы и составлять персонализированные планы обучения. На некоторых онлайн-платформах искусственный интеллект оценивает тесты и эссе.

В банках используют технологию машинного обучения для того, чтобы обнаружить подозрительную активность и предотвратить мошеннические транзакции. Также вводятся ассистенты, которые используют искусственный интеллект для того, чтобы помогать сотрудникам банка отвечать на вопросы клиентов, давать персонализированные советы. Создаются сервисы, на основе данных, окружающих пользователя, способные продемонстрировать, каким будет его финансовое положение в будущем и как лучше управлять своими финансами.

В сфере транспорта все слышали о беспилотных автомобилях. В октябре 2015 года с новым обновлением ПО компания Tesla активировала в своих машинах режим автопилота, позволяющий им самостоятельно выполнять ряд действий, но в критических ситуациях передавать управление водителю. В транспортной системе искусственный интеллект составляет расписание движения метро и автобусов, а также определяет справедливую оплату за движение на автомагистралях и мостах. В некоторых городах запустили «умные» светофоры. Эти светофоры анализируют ситуацию на дороге и автоматически переключаются на зеленый свет, если перед ними накопилось много машин.

В области здравоохранения тоже занимаются разработкой и внедрением технологий искусственного интеллекта. Например, технология Watson for Oncology обрабатывает большой объем медицинских данных, в том числе изображения, на основе которых может точно диагностировать рак. Эту технологию используют в ряде стран. Создаются приложения, благодаря технологии машинного обучения, выявляющие потенциальные проблемы со здоровьем человека [14].

Одним из крупнейших пользователей искусственного интеллекта является индустрия *онлайн-рекламы*, которая использует ИИ для того, чтобы не только отслеживать статистику пользователей, но и показывать нам рекламу на основе этой статистики.

Если вы используете социальные сети, на большинство ваших решений влияет искусственный интеллект. От каналов, которые вы видите на своей временной шкале, до уведомлений, которые вы получаете от этих приложений, все курируется искусственным интеллектом. ИИ берет на себя все ваше прошлое поведение, веб-поиски, взаимодействия и все остальное, что вы делаете, когда находитесь на этих веб-сайтах, и адаптирует опыт именно для вас. Так же работают сервисы потоковой передачи музыки и медиа, которые мы используем ежедневно, такие как Spotify, Netflix или YouTube [2].

Сегодня направление искусственного интеллекта превращается в одну из самых передовых отраслей инновационной экономики. Она полна интересными и трудноразрешимыми проблемами, решение которых приведет к очередному взрывному росту технологий, открытию новых рынков, росту благосостояния общества, но камнем преткновения является кадровый вопрос: в мире не существует достаточного количества квалифицированных специалистов для их решения. Очевидно, что в сложившихся условиях развития ИТ-индустрии будущий потенциал любого государства будет определяться тем профессиональным выбором, который сделают сегодняшние школьники.

В Национальной стратегии развития искусственного интеллекта на период до 2030 года отмечается, что «отсутствие понимания того, как искусственный интеллект достигает результатов, является одной из причин низкого уровня доверия к современным технологиям искусственного интеллекта и может стать препятствием для их развития» [26].

Особая роль в просветительской плане принадлежит общеобразовательным учреждениям. Школа, являясь общественным социальным институтом, должна обслуживать социальный заказ общества, государства, поэтому необходима целенаправленная просветительская и профориентированная работа с обучающимися в области наукоемких технологий искусственного интеллекта.

1.2. Классическое машинное обучение: данные, параметры, методы

В 1959 г. специалист по вычислительной технике из компании IBM Артур Самуэль написал компьютерную программу для игры в шашки. Каждому положению на доске присваивался некий вес, базирующийся на вероятности выигрыша. Изначально вероятность определялась по формуле, в которой учитывались такие факторы, как количество шашек на каждой стороне и количество дамек. Подход работал, но Самуэль придумал, каким

образом можно повысить его эффективность. Сыграв с программой тысячу партий, он использовал их результаты для уточнения позиционных весов, К середине 1970-х г. программа достигла уровня хорошо подготовленного непрофессионального игрока.

Самуэль написал компьютерную программу, которая могла по мере накопления опыта улучшать собственные результаты. Программа училась — так зародилось машинное обучение (ML — machine learning или МО) [10].

У людей мы различаем механическое заучивание и интеллектуальное осмысление. Зазубривание телефонных номеров или инструкций, без сомнения, тоже относится к процессу обучения. Но, как правило, под этим понятием мы подразумеваем кое-что другое.

Ребенок, играющий с друзьями, наблюдает реакцию других членов группы на свои действия. Этот опыт влияет на его будущее поведение в социуме. Но он не вспоминает и не проигрывает заново свое прошлое, а опирается на определенные, легко опознаваемые характеристики прошлых взаимодействий: детская площадка, класс, мама, папа, сестры и братья, друзья, незнакомцы, взрослые, дети, в помещении или на улице. Оценка новой ситуации базируется на признаках, с которыми ему доводилось сталкиваться раньше. Обучение при этом является не просто сбором информации. Формируется то, что можно назвать аналитической оценкой.

Представьте, как вы по картинкам учите ребенка отличать собаку от кошки. Показанная картинка кладется в одну из двух стопок, в зависимости от правильности полученного ответа. Чем дольше продолжается процесс, тем выше эффективность распознавания. Что интересно, нет необходимости специально учить ребенка отличать собаку от кошки. Человеческое сознание обладает встроенными механизмами классификации. Ему требуются только образцы. Научившись работать с картинками, ребенок сможет опознать практически любое изображение кошки или собаки, не говоря уже о реальных животных. Эта способность обобщать, применяя полученные в процессе тренировок знания к новым, ранее не встречавшимся образцам,

является ключевой характеристикой как человеческого, так и машинного обучения [10].

Разумеется, процесс получения знаний человеком превосходит своей сложностью самые совершенные алгоритмы машинного обучения, но у компьютера есть преимущество в виде большей емкости для запоминания, извлечения и обработки данных. Накапливаемый им опыт представлен в форме данных за длительный период времени, обработанных с помощью различных техник, причем это представление позволяет получать и оптимизировать алгоритмы, реализующие если не аналитическую оценку, то хотя бы способность к обобщениям.

Аналогия между человеческим и машинным обучением закономерно заставляет вспомнить такое явление, как искусственный интеллект (AI — artificial intelligence или ИИ). При этом естественным образом возникает вопрос: «Чем искусственный интеллект отличается от машинного обучения?». По этому вопросу нет единого мнения, но большинство соглашается с тем, что МО — это одна из форм ИИ, так как ИИ представляет собой куда более обширную область, включающую в числе прочего робототехнику, обработку лингвистической информации и системы машинного зрения. Неоднозначность терминологии усиливает тот факт, что машинное обучение все чаще применяется во многих сопутствующих областях ИИ. Можно сказать, что такая дисциплина, как машинное обучение, относится к специализированной совокупности знаний и связанным с ней техникам. Легко определить, что относится, а что не относится к машинному обучению, в то время как в случае искусственного интеллекта далеко не всегда можно провести такую же четкую границу [10].



Рисунок 1 — Структура искусственного интеллекта

Машинное обучение применяется к широкому кругу экономических задач — от обнаружения мошенничества до выбора целевой аудитории и рекомендаций товара, наблюдения за производством в реальном времени, анализа тональности текстов и медицинской диагностики. Оно может взять на себя задачи, которые невозможно выполнить вручную из-за огромного количества подлежащих обработке данных. В случае больших наборов данных машинное обучение иногда обнаруживает неочевидные зависимости, которые невозможно распознать при сколь угодно скрупулезном ручном рассмотрении. При этом комбинация множества таких «слабых» соотношений дает прекрасно работающие механизмы прогнозирования.

Процесс обучения на основе данных и последующего применения полученных знаний для обоснования будущих решений — чрезвычайно мощный инструмент. Машинное обучение быстро превращается в двигатель современной экономики, управляемой данными [10].

Цель машинного обучения — предсказать результат по входным данным. Чем разнообразнее входные данные, тем проще машине найти закономерности и тем точнее результат. Итак, если мы хотим обучить машину, нам нужны три вещи: данные, параметры (признаки), методы (алгоритмы) [21].

Данные — отдельные факты, характеризующие объекты (а именно предметы, процессы и явления) окружающего мира, в том числе их свойства

[8]. В процессе обучения машина должна найти как можно большее число закономерностей. Для этого она обрабатывает большое количество данных. Другими словами машин обучается на данных, которые ей предоставляются. Приведем примеры. Хотим определять спам — нужны примеры спам-писем, предсказывать курс акций — нужна история цен, узнать интересы пользователя — нужны его лайки или посты. Данных нужно как можно больше. Чем больше входных данных представляется машине и чем они разнообразнее, тем быстрее машина найдет закономерности и тем точнее будут результаты ее поиска. При этом, иногда полученный результат может быть даже таким, о котором люди раньше и не догадывались.

Данные собираются и хранятся в виде баз данных, который называют датасетами. Такая специальная база данных имеет определенный формат и определенную разметку, чтобы хранить соответствующие данные. Под датасетом понимают специальную базу данных, предназначенную для машинного обучения и имеющую определенный формат, определенную разметку [8].

Датасеты могут собираться тремя способами. Во-первых, вручную, то есть человеком. Во-вторых, автоматизировано, то есть с участием человека. В-третьих, автоматически, то есть без участия человека. Датасеты могут быть объектом личной или частной собственности, а могут быть расположены и в открытом доступе. Некоторые компании используют своих же пользователей для бесплатного пополнения датасетов. Например, ReCaptcha, которая иногда требует «найти на фотографии все дорожные знаки» — это оно и есть [21].

В процессе обучения машина должна знать, какие параметры, а именно признаки, величины ей требуется отслеживать. Поэтому чаще всего в названии полей, то есть в столбцах датасетов, указывают те параметры в соответствии с которыми собираются данные. Примеры параметров — это могут быть графические примитивы, счетчики появления слов в тексте, цены

валют, пол пользователя и многое другое. Но что чем больше параметров имеет датасет, тем медленнее будет обучаться машина.

Отличительной чертой машинного обучения является не поиск решения задачи с использованием машины, а постепенное обучение машины в процессе решения множества однотипных задач. Исходя из этого, сделаем общий вывод, что машинное обучение имеет свои методы решения задач. Методы машинного обучения — это совокупность задач, направленных на проверку гипотез, поиск оптимальных решений с помощью искусственного интеллекта [30]. Необходимо понимать какие методы подходят под конкретные задачи. От выбора метода зависит точность, скорость работы и размер используемых ресурсов. Выбор метода зависит от наличия качества сложности отбора данных и параметров. В то же время, если данных мало или они некачественные, то даже с лучшим методом результат получится недостаточно точным [8].

Существуют различные классификации методов машинного обучения. Можно использовать представленную схему.



Рисунок 2 — Направления машинного обучения

Здесь выделены основные подходы к машинному обучению. В зависимости от характеристики данных выделяют четыре основных подхода: классическое обучение, обучение с подкреплением, ансамблевое обучение, обучение нейросетей. В рамках каждого направления используются свои методы для решения различных интеллектуальных задач.

Первые алгоритмы пришли к нам из чистой статистики еще в 1950-х. Они решали формальные задачи — искали закономерности в цифрах, оценивали близость точек в пространстве и вычисляли направления [21]. В направление классического обучения входят достаточно простые методы обучения, другими словами алгоритм обучения. Это направление фактически самое популярное. Так с помощью классических методов выполняется сортировка поисковой выдаче на веб-сайте с учетом предпочтений пользователя, блокировка денег на банковской карточке при определенных действиях с ней в сети и так далее. Классическое обучение используется в том случае, если параметры определить достаточно просто и сами данные являются простыми. В классическом обучении выделяем два вида: обучение с учителем и обучение без учителя.

В рамках *обучения с учителем* машине предлагается правильная выборка, в которой все данные разделены. Это может быть в тех случаях, когда данные численные или заранее категоризированы. Так, если машину обучают отличать на фотографии кошек от собак, то эти фотографии заранее разделены (размечены) на две категории, поэтому машина обучается на конкретных примерах [18].

Обучение без учителя производится на неразмеченных данных. В этом случае, машина в начале выполняет анализ данных, затем объединяет их по выделенным параметрам и лишь потом использует их для обучения. Так, в нашем примере машине предоставляют сразу все фотографии кошек и собак. Она должна самостоятельно найти закономерности и обучиться различать на фотографиях кошек и собак [8].

Обучение с подкреплением используется, если данные отсутствуют и имеется среда для взаимодействия. Данный подход используется тогда, когда необходимо анализировать не данные, а реальную среду, взаимодействуя с которой необходимо минимизировать ошибки [18].

Ансамблевое обучение используется в тех случаях, когда нужен точный результат. Этот подход улучшает результаты, полученные при обучении с использованием других методов, например классического обучения [8].

Обучение нейросетей используется в тех случаях, когда данные являются сложными и у них невозможно определить параметры. Данный метод занимает особое место сегодня. Нейросети способны решить любую задачу, которая обычно решается с использованием других подходов. Нейросеть, основываясь на принципах работы мозга человека, позволяют получить искомый результат, даже если параметры и данные сложно определяются. Несмотря на высокое качество решений и эффективность нейросетей, этот подход требует больших ресурсов и временных затрат. Поэтому если задача решается, например, в рамках классического подхода, то вначале задействовать классические методы обучения и лишь затем рассматривать вопрос об улучшении полученного результата с помощью обучения нейросетей [8].

В нашей работе мы будем рассматривать характеристики классического обучения.

Способ	Обучение с учителем		Обучение без учителя		
Данные и параметры	Параметры определены Данные размечены		Параметры определены Данные не размечены		
Задача	Классификация	Регрессия	Кластеризация	Обобщение	Ассоциация
Цель	Определить класс объекта	Прогнозировать значение	Разделить по схожести	Найти зависимости	Найти правило
Рисунок	 Classification	 Regression	 Clustering	 Dimension Reduction	 Association Rule Learning
Пояснение к рисунку	Классификация объектов по цвету	Распределение объектов по размеру	Группировка похожих объектов	Выборка из объектов	Выявить последовательность
Примеры методов	Наивный Байес, деревья решений, логистическая регрессия, K-ближайших соседей, машины опорных векторов	Линейная регрессия, полиномиальная регрессия	Метод K-средних, Mean-Shift, DBSCAN	Метод главных компонент, сингулярное разложение, латентное размещение, латентно-семантический анализ, t-SNE	Euclat, Apriori, FP-growth

Рисунок 3 — Характеристики классического обучения

Уже было сказано, что в классическом обучении выделяют обучение с учителем и обучение без учителя. Для *обучения с учителем* характерно решение таких задач, как классификация и регрессия. В таком обучении параметры определены, а данные размечены.

Рассмотрим тип *задач классификации*. В этой задаче необходимо определить класс объектов и разделить объекты на основе заранее известных параметров. Данный тип задач лежит в основе решения следующих проблем: разработки спам-фильтров, определение языка, поиска похожих документов, анализа тональности, распознавания рукописных букв и цифр, определения подозрительных транзакций.

Следующей рассмотрим *задачу регрессии*. Эта задача связана с прогнозированием значения непрерывных параметров какого-либо объекта. По своей сути, это задача на классификацию, но в которой вместо категории объектов, другими словами класс объекта, прогнозируется число. Данный тип задач лежит в основе решения следующих проблем: прогнозирование стоимости ценных бумаг, анализа спроса или объема продаж, установление медицинских диагнозов, выявление любых зависимостей числа от времени, определение стоимости автомобиля по его пробегу, прогнозирование

количества пробок на дороге в зависимости от времени суток, выявления объема спроса на товар от роста компании [8].

В классическом *обучении без учителя* чаще встречаются задачи кластеризации, обобщения и ассоциации. В таком обучении параметры определены, но данные еще не размечены.

Задача кластеризации связана с группировкой схожих объектов на основе их параметров. Это похоже на задачу классификации, но в которой заранее неизвестны классы. Машина сама ищет похожие объекты и объединяет их в кластеры [15]. Их количество может быть определено машиной или задано заранее. Данный тип задач лежит в основе решения следующих проблем: сегментация рынка (типов покупателей, лояльности), объединению близких точек на карте, сжатие изображений, анализа и разметки новых данных, разработки детекторов аномального поведения, нахождение лиц людей на фотографиях и группировке их в альбом.

Рассмотрим *задачу обобщения*, связанную с нахождением зависимости между объектами и объединением их параметров в абстракцию, по-другому в параметр более высокого уровня без потери связей между ними. Машина сжимает данные большой размерности в меньшую, это необходимо для визуализации данных или для их использования в других методах машинного обучения. Решая задачи обобщения, мы получаем систему рекомендаций в областях кино, музыки и других. В соответствии с заданными критериями можно осуществить отбор ограниченного множества объектов с большего множества. Например, это представление персонализированного выбора товаров или услуг пользователей магазинов для повышения лояльности. Данный тип задач лежит в основе решения следующих проблем: создание рекомендательных систем, красивых визуализаций, определение тематики и поиска похожих документов, выявление фейковых изображений, анализа риск-менеджмента.

Следующей рассмотрим *задачу ассоциации*. В ней мы находим закономерности в потоке последовательностей, поэтому ее еще называют

поиском правил. Данный тип задач лежит в основе решения следующих проблем: прогноз акций и распродаж, анализа товаров, покупаемых вместе, расстановка товаров на полках [8].

1.3. Анализ существующих средств обучения школьников по теме «Введение в классическое машинное обучение»

В сети Интернет можно найти различные ресурсы, которые помогут учителям и обучающимся в первоначальном ознакомлении с темой классического машинного обучения.

Всероссийский образовательный проект «Урок цифры» позволяет учащимся получить знания от ведущих технологических компаний, развить навыки и компетенции в цифровой экономике, сориентироваться в мире профессий, связанных с компьютерными технологиями и программированием [1].

На сайте проекта доступен урок 2020 г. «Искусственный интеллект и машинное обучение». На уроке ученики пройдут все этапы создания алгоритмов искусственного интеллекта на примере беспилотного автомобиля и больше узнают про профессию Data Scientist.

Сперва предлагается просмотр видеолекции, подготовленной специалистами компании - организатора урока. В содержании ролика можно выделить несколько тематических блоков:

– Введение понятия «искусственный интеллект»: дается определение понятия «искусственный интеллект», показывается принципиальное различие между «сильным» и «слабым» искусственным интеллектом.

– Машинное обучение (профориентационный): анализируется сфера применения систем слабого искусственного интеллекта и подчеркивается, что это стало благодаря возможностям машинного

обучения; акцент делается на значении данных для машинного обучения и специальностях, связанных с обучением машин.

– Мотивационный: акцентируется внимание на базовых предметах и метапредметных результатах (Soft Skills), которые необходимы будут школьникам для получения престижных профессий, связанных с машинным обучением, роли самообразования для профессионального самоопределения.

Затем идет этап получения и отработки практических навыков в процессе работы с тренажером или решения отдельных задач, ориентированных на погружение и знакомство с профессиями в области работы и исследования данных, связанных непосредственно с машинным обучением и искусственным интеллектом.

В практической работе с тренажером по сюжету игры школьники вместе с главными героями Урока цифры (братья Слэши и Скобец) попадают в Институт инновационных разработок, где их подруга Запятыня проходит стажировку на проекте по разработке беспилотного автомобиля. В ходе выполнения заданий обучающимся предстоит познакомиться с устройством беспилотника, попробовать себя в роли специалистов по работе с данными и машинному обучению, чтобы научить автомобиль распознавать различные классы объектов на дороге (дорожные знаки, пешеходов, транспортные средства и элементы разметки) и самостоятельно передвигаться по городу также безопасно, как если бы им управлял опытный водитель.

В альтернативном сценарии практической части занятия «Мы — исследователи: анализ данных в MS EXCEL» показываются на практике некоторые аспекты работы специалиста в области исследования данных (Data Science) (аналитика данных), связанные с обработкой и извлечением смысла из массива данных: их структурированием, нахождением закономерностей, визуализации и построением выводов для принятия решений. Предлагаемая для выполнения лабораторная работа включает две части: анализ данных с помощью сводных таблиц и визуализация и анализ данных с помощью функции условного форматирования [1].

Урок, разработанный в рамках акции «Урок цифры» в 2019 году, познакомит учеников со сферами применения искусственного интеллекта, расширит их представление о технологии машинного обучения и перспективах развития этого направления в IT-индустрии.

На первом этапе урока понятие ИИ рассматривается через призму личного опыта школьников, анализируя который обучающиеся погружаются в предметное поле, связанное с развитием технологий искусственного интеллекта. Также предлагается посмотреть видеоролик.

На втором этапе урока в тренажёре для 1-8 классов перед школьниками стоит цель наладить работу зоопарка будущего, где уход за животными осуществляют роботы-смотрители. По сюжету клетки с животными поменяли местами, появились новые звери, и роботы стали допускать ошибки: например, собакам достается корм для кошек, а игрушки для кошек попадают в секцию для волков. Это происходит из-за того, что роботы не понимают кто где находится, а просто работают по стандартной компьютерной программе. Перемены в зоопарке происходят часто, а постоянно переписывать и отлаживать новые программы для роботов очень долго и дорого.

Перед школьниками, выполняющими роль специалистов по искусственному интеллекту, руководством зоопарка ставится задача: обучить роботов распознавать животных, чтобы машины самостоятельно могли принимать правильные решения и не допускать ошибок при уходе за волками, собаками, кошками и тиграми.

В ходе работы с тренажёром обучающимся предстоит самостоятельно пройти 4 уровня, чтобы создать собственную модель обучения робота, выбрать алгоритм принятия решений и протестировать их работу в виртуальном зоопарке [1].

Еще один ресурс — *академия искусственного интеллекта для школьников*. Это всероссийский образовательный проект, реализуемый Благотворительным фондом Сбербанка «Вклад в будущее» при поддержке

Сбербанка и направленный на просвещение, обучение и работу с талантами в сфере искусственного интеллекта (ИИ) среди школьников [1].

Для обучающихся на сайте академии размещены 3 вводных урока, которые расскажут о роли искусственного интеллекта в современном мире, познакомят с машинным обучением и помогут начать самому решать задачи анализа данных.

На сайте можно найти материалы серии уроков «Знакомство с искусственным интеллектом». Цель уроков — познакомить учащихся с технологиями искусственного интеллекта и их использованием в современном мире. Серия состоит из 7 уроков, которые могут проводиться как учителями информатики, так и преподавателями других дисциплин. Например, урок «Машинное обучение в искусстве» может пройти в рамках МХК, ИЗО или музыки, а «Машинное обучение в спорте» проведен в качестве альтернативного урока по физкультуре.

В ходе урока «Введение в машинное обучение» ученики познакомятся с различными аспектами использования технологий машинного обучения в разных сферах деятельности человека. Активное погружение в мир новых технологических решений, связанных с робототехникой, беспилотными автомобилями и поездами, интеллектуальными играми, голосовыми помощниками и, наконец, произведениями искусства должно помочь создать атмосферу сопричастности учащихся к той удивительной эпохе, в которой мы живем, тем открытиям, которые делаются нашими современниками. Но, что еще важнее, весь «калейдоскоп» открытий, с которыми знакомятся и которые обсуждают школьники, должен помочь осознать им уровень требований к специалистам будущего в разных профессиональных сферах. Это серьезная профориентационная идея урока, задача на перспективу.

Также для обучающихся есть курс по машинному обучению. Он создан, чтобы обучающийся смог получить базовые знания о языке Python, узнать об объектно-ориентированном программировании и научиться работать с моделями машинного обучения. Представлены разные формы

подачи материала для обучения: видео и скринкасты, подсказки, конспекты [8].

Курс состоит из разделов:

- Вводный урок. В нем размещена информация о курсе.
- Основы python. Данный раздел позволит с нуля освоить инструменты и навыки, необходимые для успешного решения будущих задач по анализу данных и, конечно, машинному обучению.
 - Анализ данных. В этом уроке знакомятся с инструментами, которые помогут обрабатывать, анализировать и визуализировать данные, учат работать с ними, делать выводы. Рассматриваются основные понятия из линейной алгебры и математической статистики, которые понадобятся в работе.
 - Машинное обучение. В этом разделе можно научиться строить свои ML-модели.
 - Финальный проект.
 - Итоговый урок.

Академия ИИ предлагает и курсы для учителей. Всего представлены два курса. Курс «Основы Python» познакомит с возможностями языка и позволит с нуля освоить необходимые для начала работы навыки. Данный курс является частью курса по машинному обучению для обучающихся. Курс «Искусственный интеллект: основы теории и методика обучения в основной школе» для повышения квалификации для педагогов — совместная разработка специалистов МГПУ и экспертов Академии искусственного интеллекта для школьников. Он построен таким образом, чтобы педагоги разных специальностей могли познакомиться со сферой искусственного интеллекта и подготовиться к проведению занятий с учениками. После прохождения обучения педагоги смогут не только организовать качественное обучение с использованием передовых методик преподавания данной дисциплины в школе, но и сделать процесс интересным для учеников.

При подготовке к урокам можно использовать книгу *Джона Формана «Много цифр: Анализ больших данных при помощи Excel»* [29]. Автор этой книги доказывает, что анализ данных можно организовать и в простом, понятном, очень эффективном и знакомом многим Excel. Причем не важно, насколько велик ваш массив данных. В книге говорится как научиться видеть в данных нужную информацию и анализировать ее [28].

Полезными главами будут:

- Глава 2. Кластерный анализ с использованием метода k-средних
- Глава 6. Регрессия, как инструмент контролируемого искусственного интеллекта

В них рассказывается о техниках, которые применяются в решении задач классического машинного обучения. Можно понять и увидеть принцип их работы на примерах в привычном MS Excel.

Компания Google запустила *веб-эксперимент Teachable Machine*, наглядно показывающий, как работает машинное обучение [7]. На сайте Teachable Machine можно продемонстрировать какой-либо объект системе при помощи веб-камеры компьютера. После этого с помощью нажатия трёх разноцветных кнопок проводится процесс обучения искусственного интеллекта — система распознаёт объекты. При этом для каждой кнопки (каждого класса обучения) нужно использовать различные объекты. После обучения сервис связывает распознанные объекты с одним из доступных вариантов представления. Соответствие может быть представлено в виде GIF-анимации, звуковых эффектов или голосового сообщения. В дальнейшем пользователь может менять объекты перед камерой устройства, а система будет определять эти изменения и выводить связанное с конкретным объектом соответствие. Несмотря на кажущуюся простоту сервиса Teachable Machine, он демонстрирует некоторые фундаментальные аспекты машинного обучения [6].

Выводы по 1 главе

Современные инструменты искусственного интеллекта (ИИ) и машинного обучения (МО) приносят в нашу жизнь новые возможности, новый контент и новые знания, но при этом порождая соответствующий поток задач. В России применение МО отстает от остального мира. Сложилась парадоксальная ситуация: в ТОП 100 лучших Data Scientists планеты входит достаточно много наших соотечественников, при этом компании, которые могут похвастаться реальными проектами, приносящими пользу в этой сфере, в России можно пересчитать по пальцам одной руки. Это говорит о том, что на рынке существует колоссальная нехватка специалистов, причем даже не гениальных ученых, а простых квалифицированных аналитиков, способных грамотно поставить задачу для машинного обучения, подготовить данные, осознанно выбрать подходящую модель и проверить ее [22]. Поэтому необходимо начинать знакомить обучающихся с технологиями ИИ и машинного обучения в школе. Так обучающиеся будут мотивированы к выбору ИТ-профессий, осознавать востребованность специалистов в этой сфере и понимать важность для личностного самоопределения, чтобы быть всегда необходимым на рынке труда.

Ресурсы для ознакомления с машинным обучением есть, но они не дают полного погружения в эту тему, а лишь частичное. Таким образом, появляется необходимость разработать курс по выбору для обучающихся основной школы «Введение в классическое машинное обучение». Его задачами будет сформировать представление о машинном обучении, актуальных профессиях в области ИИ, основных подходах к машинному обучению и т.д. Программу такого курса по выбору, а также методические рекомендации по его изучению, рассмотрим во второй главе.

Глава II. Разработка курса по выбору «Введение в классическое машинное обучение» для обучающихся основной школы

2.1. Программа курса по выбору «Введение в классическое машинное обучение»

Ниже предлагается разработанная программа курса по выбору «Введение в классическое машинное обучение». Она содержит пояснительную записку, цели изучения курса, планируемые результаты освоения курса, тематическое планирование, содержание курса, требования к уровню подготовки обучающихся и учебно-методический комплект.

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Место курса в образовательном процессе: курс «Введение в классическое машинное обучение» — курс по выбору для обучающихся 9 классов. Курс рассчитан на 9 часов.

В настоящее время процесс информатизации проявляется во всех сферах человеческой деятельности. Использование современных информационных технологий является необходимым условием успешного развития как отдельных отраслей, так государства в целом.

Машинное обучение (МО) все глубже проникает в нашу жизнь посредством пользовательских продуктов, созданных с помощью методов искусственного интеллекта (ИИ). Очевидно, что данные технологии будут развиваться и дальше, постепенно становясь частью повседневной рутины во многих областях человеческой профессиональной деятельности.

Обучение по данной программе создаёт благоприятные условия для развития интеллектуальной деятельности, познавательной активности и творческой самореализации обучающихся.

Отличительная особенность настоящей программы состоит в том, что обучающиеся 9 классов получают представление о классическом машинном обучении, что поможет им сформировать индивидуальный образовательный запрос на развитие в одной из самых перспективных сфер деятельности.

ЦЕЛИ ИЗУЧЕНИЯ КУРСА

Основной *целью* курса по выбору для обучающихся основной школы «Введение в классическое машинное обучение» является формирование представления о классических подходах к машинному обучению.

Задачи курса:

- расширить представление обучающихся о машинном обучении;
- познакомить с профессиями в области искусственного интеллекта и машинного обучения;
- сформировать представление о основных подходах к машинному обучению;
- рассмотреть классические подходы к машинному обучению: обучение с учителем и обучение без учителя;
- сформировать представление о принципах работы методов классического машинного обучения.

ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ КУРСА

Предметные результаты обучения:

- иметь представление о возможностях и перспективах машинного обучения;
- знать основные подходы машинного обучения;
- приводить примеры типовых задач классического машинного обучения;
- знать принципы работы алгоритмов классического машинного обучения.

Метапредметные результаты обучения:

- формировать универсальные учебные действия (познавательные, регулятивные, коммуникативные), обобщенные способы информационной деятельности при взаимодействии с интеллектуальными системами;

– развить познавательные интересы, интеллектуальные и творческие способности при проведении компьютерных экспериментов по машинному обучению;

– приобрести опыт машинного обучения интеллектуальных систем в индивидуальной, групповой и коллективной учебно-познавательной деятельности.

Личностные результаты обучения:

– личностное и предпрофессиональное самоопределение через познавательную мотивацию к получению профессий, связанных с искусственным интеллектом и через познавательный интерес к достижениям в области машинного обучения;

– построение дальнейшей индивидуальной образовательной траектории через получение представления о перспективных направлениях машинного обучения;

– осознание стратегической важности для государства, общества и для своего будущего развития машинного обучения.

ТЕМАТИЧЕСКОЕ ПЛАНИРОВАНИЕ КУРСА

№	Название темы	Кол. час.
1.	Вводные уроки по машинному обучению	4
1.1	Введение в машинное обучение	1
1.2	Большие данные	1
1.3	Профессии в сфере ИИ	1
1.4	Основные подходы к машинному обучению	1
2.	Машинное обучение: классическое обучение	5
2.1	Задачи и методы классического машинного обучения: Обучение с учителем	1
2.2	Обучение с учителем: Линейная регрессия	1

2.3	Обучение с учителем: Дерево решений	1
2.4	Обучение без учителя: Кластеризация	1
2.5	Заключительный урок	1
ВСЕГО:		9

СОДЕРЖАНИЕ КУРСА

Урок 1. Введение в машинное обучение

Теория: Коротко об искусственном интеллекте. Машинное обучение. Отличие от обычного программирования. Понятия «данные» и «знания», «базы данных» и «базы знаний». Работа с базами данных и базами знаний. Сущность процессов работы со знаниями. Важность данных для машинного обучения.

Практика: работа с текстом с теорией, составление схем.

Урок 2. Большие данные

Теория: Понятие больших данных. Характеристика больших данных. Источники больших данных. Процесс обработки данных.

Практика: Работа в группах. Мини-доклады о примерах использования больших данных в разных сферах человеческой деятельности.

Урок 3. Профессии в сфере ИИ

Теория: Как меняется мир профессий с развитием ИТ-технологий. Рассмотрение атласа новых профессии. Профессии в сфере ИИ, МО, БД. Какие знания и навыки нужны для работы. Карьера.

Практика: Лабораторная работа. «Мы — исследователи: анализ данных в MS Excel».

Урок 4. Основные подходы к машинному обучению

Теория: Особенность машинного обучения. Три составляющие машинного обучения. Датасет. Параметры. Методы. Подходы к машинному обучению. Особенности подходов к машинному обучению. Классическое

обучение, обучение с подкреплением, ансамблевое обучение, обучение нейросетей.

Практика: Компьютерный эксперимент по машинному обучению с использованием тренажера от образовательного проекта «Урок цифры».

Урок 5. Задачи и методы классического машинного обучения: Обучение с учителем

Теория: Особенность обучения с учителем. Задачи, решаемые методами обучения с учителем.

Практика: Компьютерный эксперимент по машинному обучению с учителем с использованием тренажеров Teachable Machine.

Урок 6. Обучение с учителем: Линейная регрессия

Теория: Задача регрессии. Линейная регрессия. Математическое основание линейной регрессии.

Практика: Разбор примера с предсказанием оценки для игры Fortnite от Академии ИИ. Построение линейной регрессионной модели по данным в MS Excel.

Урок 7. Обучение с учителем: Дерево решений

Теория: Дерево решений. Принцип построения дерева решений. Понятия «узел», «листья», «глубина дерева».

Практика: Построение дерева решений для конкретной задачи.

Урок 8. Обучение без учителя: Кластеризация

Теория: Особенность обучения без учителя. Задачи, решаемые методами обучения без учителя. Кластеризация. Кластерный анализ. Методы кластерного анализа. Кластеризация методом k-средних.

Практика: Работа в MS Excel. Кластеризация методом k-средних.

Урок 9. Заключительный урок

Практика: Работа в группах. Каждая группа составляет список вопросов по всему курсу, после чего отвечают на вопросы других групп, затем проверяют их ответы. Прохождение теста «Какое направление в IT выбрать?» от образовательной платформы «Нетология».

ТРЕБОВАНИЯ К УРОВНЮ ПОДГОТОВКИ ОБУЧАЮЩИХСЯ

В результате изучения данного курса по выбору обучающиеся должны:

Знать:

- возможности актуальных алгоритмов машинного обучения, которые широко используются на практике, основные сферы их применения;
- классы задач, решаемых с помощью классических алгоритмов машинного обучения;
- принцип работы классических алгоритмов машинного обучения.

Уметь:

- обосновать применение того или иного алгоритма машинного обучения для решения конкретной задачи;
- представить принцип работы классических алгоритмов машинного обучения на примерах.

УЧЕБНО-МЕТОДИЧЕСКИЙ КОМПЛЕКТ

Информационные ресурсы:

1. Академия искусственного интеллекта для школьников [Электронный ресурс]. Режим доступа: <http://www.ai-academy.ru/> (дата обращения: 11.05.2021).
2. Всероссийский образовательный проект в сфере цифровой экономики «Урок цифры» [Электронный ресурс]. Режим доступа: <https://урокцифры.рф/> (дата обращения: 11.05.2021)
3. Джон Форман. Много цифр: Анализ больших данных при помощи Excel. М.: Альпина Паблицер, 2016. 464 с.
4. Teachable Machine [Электронный ресурс]. URL: <https://teachablemachine.withgoogle.com/> (дата обращения: 11.05.2021).
5. Атлас новых профессий [Электронный ресурс]. URL: <https://new.atlas100.ru/> (дата обращения: 11.05.2021).

6. Тест: Какое направление в IT выбрать? [Электронный ресурс].
URL: <https://netology-code.github.io/who2be/> (дата обращения: 11.05.2021).

Программно-аппаратные ресурсы:

- проектор;
- интерактивная доска;
- доступ к Интернет;
- оборудованный компьютерный кабинет для работы с группой/классом.

2.2. Методические рекомендации по изучению курса по выбору «Введение в классическое машинное обучение»

В данном параграфе рассмотрим в какой последовательности и каким образом достигаются запланированные результаты курса, а также приведем предлагаемые планы занятий и некоторые методические рекомендации.

Урок 1. Введение в машинное обучение

Цель урока:

Расширить и развить понятия «данные» и «знания», «базы данных» и «базы знаний».

Задачи урока:

- Ввести понятия «данные» и «знания», «базы данных» и «базы знаний»;
- Обобщить представление о работе с ними;
- Раскрыть сущность процессов работы со знаниями.

Предлагаемый план занятия и методические рекомендации

Предлагается выстроить беседу, в ходе которой обсуждаются вопросы:

Что такое искусственный интеллект?

Определим искусственный интеллект (ИИ) как научную область, занимающуюся созданием программ и устройств, имитирующих интеллектуальные функции человека [8].

Какой состав области искусственного интеллекта?



Рисунок 4 — Структура искусственного интеллекта

Что понимается под машинным обучением?

Машинное обучение (МО) — обширный подраздел искусственного интеллекта, изучающий методы построения алгоритмов, способных обучаться [20].

Что понимается под данными?

Данные — отдельные факты, характеризующие объекты (а именно предметы, процессы и явления) окружающего мира, в том числе их свойства.

Какие существуют этапы последовательной работы с данными?

- данные существуют как результат измерений и наблюдений
- данные содержащиеся на материальных носителях (например, в протоколах, справочниках)
- данные структурно представлены (например, в виде таблиц, списков, диаграмм, графиков, функций)
- данные используются в компьютере и закодированы на языках описания данных
- данные хранятся в базах данных

Что понимается под знаниями?

Знания - совокупность утверждений о мире и свойствах объектов, о различных закономерностях процессов и явлений, правила логического вывода одних утверждений из других и использования результатов вывода для принятия решений [11].

В чем разница между декларативными и процедурными знаниями?

Знания бывают двух типов: декларативные и процедурные. Декларативные знания, то есть описательные, это утверждения, которые отражают законы природы, общества и закономерности в конкретных областях. С развитием систем ИИ часть знаний, являющихся декларативными, сосредотачивается в отдельных структурах. Например, мы знаем, что жесткий диск это внешняя память компьютера. Процедурные знания, то есть алгоритмические, это знания о том, как надо выполнять действия. Эти знания растворены в самом тексте программы, и корректировка таких знаний требует изменения самого текста программы. Алгоритм записи информации на жесткий диск будем считать процедурным знанием.

Какие существуют этапы последовательной работы со знаниями?

- знания существуют в памяти человека как результат обучения и воспитания
- знания содержатся на материальных носителях (например, в учебниках, инструкциях, методических пособиях, книгах)
- знания структурно представлены (например, в виде продукционной модели, фреймовой модели, семантической модели)
- знания функционируют в компьютере и закодированы на языках описания знаний
- знания хранятся в компьютере в базах знаний

Чем отличается база данных от базы знаний?

В отличие от баз данных, которые характеризуются большим объемом и относительно небольшой стоимостью информации, базы знаний имеют сравнительно небольшой объем и высокую стоимость информации.

Формирование знаний человеком или системой искусственного интеллекта включает в себя три последовательных процесса: извлечения (приобретения) знаний, представление знаний и обработку.

Какова сущность процесса извлечения (приобретения) знаний?

Процесс извлечения знаний предполагает взаимодействие с их источниками. Источником знаний могут быть люди, которые хранят знания у себя в голове, информационные ресурсы, которые тоже хранят знания, то есть учебники, энциклопедии, документы и т.д. Есть также формализуемые, плохо формализуемые и неформализуемые знания. Формализуемые - это те, которые хранятся на информационных ресурсах в виде текстов, формул, моделей, алгоритмов. Плохо формализуемыми знаниями являются эмпирические и эвристические правила и приемы, которые формируются в результате наблюдений, опыта работы. Неформализуемые знания являются те, которые не могут быть выражены на языке представления (например, рисунки, фотографии) или на языке описания (например, словесная математическая запись). В результате приобретения знаний, становится ясно, нужны ли эти знания для дальнейшей обработки. Для приобретения знаний из источников существуют различные методы. Системы искусственного интеллекта могут самостоятельно приобретать, извлекать знания, используя методы интеллектуального анализа данных, статистического анализа данных и другие методы [14].

Какова сущность процесса представления знаний?

Он предполагает перевод приобретенных или извлеченных знаний в форму, которая может быть автоматически обработана системой искусственного интеллекта. Форма представления знаний связана с определенной структурой (другими словами, совокупностью связей между элементами) и позволяет структурно представить знания об определенной

предметной области. Знания представляют различными способами в зависимости от моделируемой предметной области, от решаемых интеллектуальных задач.

Какова сущность процесса обработки знаний?

Процесс обработки знаний содержит следующие последовательные процедуры: логический вывод, принятие решений и объяснение принятого решения. Логический вывод заключается в получении новых знаний из имеющихся фактов. При выполнении этой процедуры, имеющаяся ситуация сравнивается с теми, что уже записаны в базе данных. Если ситуация соответствует шаблону из базы знаний, то эти шаблоны применяются, а как результат мы получаем логический вывод. Для различных способов представления знаний процедура логического вывода будет различной. Принятием решений является выбор действий из множества вариантов, которые может выбрать интеллектуальная система. После процедуры логического вывода может получиться несколько разных вариантов. Среди этих вариантов необходимо выбрать один, который применяется. Обычно это происходит в условиях неопределенности и неполноты информации. Чтобы объяснить принятое решение есть процедура, когда рассматриваются гибридные схемы построения системы искусственного интеллекта. Такие системы объясняют свои решения и обладают мощностью нейронных сетей. Сегодня это одно из наиболее динамично развивающихся направлений искусственного интеллекта.

Важность данных для машинного обучения.

Цель машинного обучения — предсказать результат по входным данным. Чем разнообразнее входные данные, тем проще машине найти закономерности и тем точнее результат [21].

В начале урока предложите обучающимся заполнять листы-конспекты с теорией. В них необходимо будет записывать изучаемую информацию.

Частичный пример такого листа:

Тема урока: _____

1. Изобразите, что входит в область искусственного интеллекта:

2. Этапы последовательной работы с данными:

-
-
-
-
-

3. Назовите два типа знаний:

4.

После беседы зафиксируйте результат урока. Задайте ученикам вопросы:

- Каковы связь и отличие между данными и знаниями? Приведите примеры данных и знаний из различных предметных областей.
- Каковы особенности знаний формализуемых, плохо формализуемых и неформализуемых?
- Почему существуют различные способы (методы) извлечения и представления знаний?

Урок 2. Большие данные

Цель урока:

Сформировать у учеников представление о понятии «большие данные» и разобрать основные сферы деятельности, связанные с их применением.

Задачи урока:

- Ввести понятие «большие данные»;
- Обсудить источники больших данных;
- Рассмотреть процесс обработки данных;

- Разобрать примеры применения концепции больших данных в современном мире;
- Обсудить полученный опыт, сформулировать выводы.

Предлагаемый план занятия и методические рекомендации

Некоторые методические рекомендации были взяты с сайта всероссийского образовательного проекта в сфере цифровой экономики «Урок цифры».

1. Анонс занятия

Подведите обучающихся к теме урока «Большие данные».

Мы живем в цифровом мире и каждый день сталкиваемся с большими данными, даже если не подозреваем об этом. Почему какие-то данные называют большими? Например, пользуясь смартфоном, человек приобщается к теме больших данных. Задайте вопрос: каким образом это происходит?

Сформулируйте цель на урок: познакомиться с понятием большие данные, их источниками, сферами использования.

В ходе урока обучающиеся познакомятся с понятием большие данные, узнают, откуда они возникают и как их можно использовать.

2. Обсуждение с обучающимися терминологии: «данные», «большие данные», характеристики больших данных, процесс работы с данными.

Обсудите с обучающимися понятие «данные». Задайте ученикам вопрос, как они могут объяснить, что такое «данные», помнят ли они определение с предыдущего занятия. Выслушайте варианты, предложенные учениками.

Данные — отдельные факты, характеризующие объекты (а именно предметы, процессы и явления) окружающего мира, в том числе их свойства [8].

Спросите обучающихся, в чем измеряются данные (байты, мегабайты, гигабайты и др.). Затем предложите ученикам примерно оценить, сколько сейчас существует данных. После нескольких ответов раскройте реальную оценку — зеттабайты данных.

Сейчас в мире примерно 40 зеттабайтов информации. Спросите, могут ли они представить, что такое «зеттабайт»? Для сравнения расскажите, что в Облаке на 100 Гб может поместиться около 100 фильмов в HD или 10000 фотографий, снятых на смартфон. Зеттабайт это 10 миллиардов по 100 фильмов. Зеттабайт — единица измерения количества информации, равная 10^{21} в степени (секстиллион) байт [1].

Покажите ученикам график. Отметьте неуклонный рост количества информации.

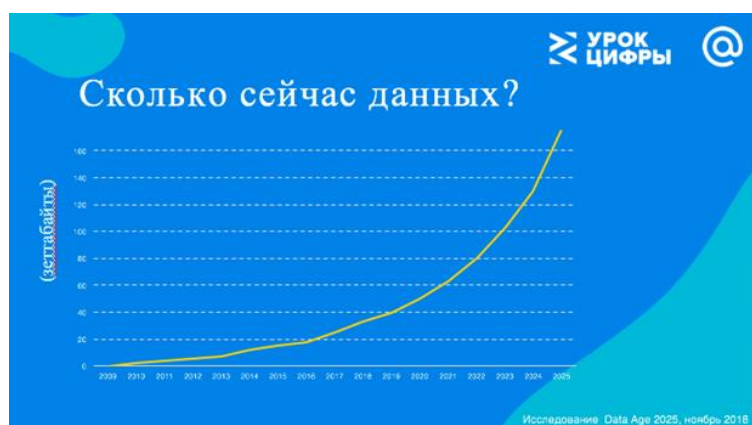


Рисунок 5 — График роста количества информации

Спросите: «Какие данные можно считать большими? Какие данные маленькие и что им нужно сделать, чтобы стать большими?» Обсудите, что четкого разделения между «малыми» и «большими» данными нет. Термину «большие данные» всего 15 лет, и окончательного определения еще не сложилось. Большими данными называют:

- различные инструменты, подходы и методы обработки как структурированных, так и неструктурированных данных для того, чтобы использовать их для конкретных задач и целей.
- данные, которые невозможно обработать на одном компьютере.

Расскажите про характеристики больших данных. Выделяют три основные характеристики больших данных («три V»):

1) объём (volume) — величина физического объёма данных. Большие данные измеряются в десятках терабайт;

2) скорость (velocity) — скорость постоянного прироста данных, а также необходимость высокоскоростной обработки и получения результатов на их основе;

3) многообразие (variety) — возможность одновременной обработки различных типов структурированных и полуструктурированных данных.

К первым трем добавляют:

4) достоверность (veracity) — данные должны быть репрезентативны и непротиворечивы;

5) ценность (value) — данные должны обладать полезностью или потенциальной ценностью (многие данные, собираемые сегодня, не всегда подлежат обработке, большинство просто хранится до нужного времени) [19].

Объясните процесс работы с данными.



Рисунок 6 — Схема процесса работы с данными

Для простоты вся работа с данными объединена в три этапа — генерацию, предобработку и анализ.

Спросите у обучающихся как данные генерируются? Большие данные берутся из самых разных источников. К источникам больших данных относят:

- Интернет — соцсети, блоги, СМИ, форумы, сайты, интернет вещей (IoT).
- Корпоративные данные — транзакционная деловая информация, архивы, базы данных.
- Показания устройств — датчиков, приборов, а также метеорологические данные, данные сотовой связи и т.д. [5].

Можно привести примеры. В момент генерации новых данных используются различные устройства, которые фиксируют происходящее вокруг, например, в этом классе — это датчик пожарной безопасности. В телефоне — это GPS который считывает вашу геолокацию вплоть до полуметра, гироскоп и акселерометр определяют положение вашего телефона в пространстве и так далее. Фиксируется все, что вы делаете в приложениях, соцсетях (мы говорили уже про сообщения, лайки и т. п.), играх и веб-сайтах — это может быть собрано и в будущем проанализировано.

Данные могут быть получены сразу большим объемом от других организаций. Так, нередко многие магазины делятся своими данными между собой.

Все эти данные сохраняются в больших хранилищах — ЦОДах — Центрах обработки данных. И на сегодняшний день ЦОДы — это одни из самых защищенных мест в любых ИТ-компаниях и государствах.

Далее происходит первичная обработка данных, так как их надо структурировать, очистить от ненужных или являющихся выбросами, которые могут смазать картинку. Данные могут между собой соединяться, обогащаться за счет друг друга. Конечно же, для этого необходима специальная инфраструктура (базы данных для умеренно большого объема данных и MapReduce для настоящей Big Data), которая требует поддержки со стороны дата-инженеров.

После этого начинается процесс анализа данных. Дата-аналитики и исследователи строят прогнозные модели, проверяют гипотезы на данных и, если все работает, то запускают их в жизнь. Для этого они используют методы визуализации данных, статистической проверки гипотез и технологии машинного обучения, в частности нейросети [1].

Сегодня не всегда понятно, как обрабатывать многие данные, еще не придуманы все алгоритмы на свете, потому многое пока просто хранится. Это можно сравнить с айсбергом — на поверхности мы видим только 1/9 часть применения данных, а через годы станет ясно, как использовать и оставшуюся часть, что может значительно изменить нашу жизнь. 15 лет назад еще не было специалистов по машинному обучению, как и специалистов по мобильной разработке, а сегодня и тех, и других требуются десятки и сотни тысяч.

3. Работа за компьютером и представление докладов.

Задайте обучающимся вопросы: «Где применяется анализ больших данных? В каких сферах деятельности?». Когда ученики будут выдвигать свои предположения, фиксируйте их на доске. В результате могут быть написаны: погода и климат, сфера развлечений, торговля и реклама, медицина, виртуальные помощники, транспорт, безопасность, образование. Разделите обучающихся на 5 групп. Предложите каждой группе выбрать одну сферу применения. Дайте задание: с помощью сети Интернет найти примеры использования анализа больших данных в заданной сфере человеческой деятельности. По возможности привести примеры популярных сервисов. Подготовить небольшой доклад для класса и выступить с ним.

4. Рефлексия

После всех выступлений зафиксируйте результат урока. Задайте ученикам вопросы:

- Какая информация была для вас новой?

- С какими источниками больших данных мы сталкиваемся в повседневной жизни?
- Каким образом сами участвуем в генерации больших данных?
- Для чего нужен анализ больших данных?

Урок 3. Профессии в сфере ИИ

Цель урока:

Сформировать у учеников представление о профессиях в сфере искусственного интеллекта и показать на практике некоторые аспекты работы специалиста в области исследования данных (Data Science).

Задачи урока:

- Обсудить изменение в мире профессий с развитием ИТ-технологий;
- Разобрать сферы применения технологий ИИ и МО;
- Рассмотреть атлас новых профессий;
- Выявить необходимые умения и навыки для работы в сфере ИИ;
- Сформировать навыки использования базовых инструментальных средств табличного процессора MS Excel для анализа данных;
- Получить опыт использования сводных таблиц и функций условного форматирования для аргументации принятия решений и выдвижения гипотез.

Подготовка и планирование урока

Традиционным и наиболее распространённым вариантом построения структуры занятия является урок, включающий три основных этапа:

Этап 1. На этом этапе обучающиеся должны понять, как меняется мир профессий с развитием ИТ-технологий. Это является важным для личностного и профессионального самоопределения школьников.

Этап 2. Получение и отработка практических навыков в процессе решения отдельных задач, ориентированных на погружение и знакомство с

профессиями в области работы и исследования данных, связанных непосредственно с машинным обучением и искусственным интеллектом.

Этап 3. Подведение итогов и рефлексия.

Ниже приводится специфика организации этапов урока.

Рекомендации по проведению вводной части урока

Ни для кого не секрет, что в ближайшем будущем рынок труда ждут крутые перемены. Уйдут в прошлое многие востребованные ныне профессии, и на их место придут новые. Задайте вопросы: почему это происходит? С чем это связано? Выслушайте ответы обучающихся. Подведите обучающихся к теме урока «Профессии в сфере ИИ».

Цифровая трансформация добралась и до рабочих мест. С помощью технологий искусственного интеллекта, машинного и глубокого обучения, роботизации процессов удастся облегчить работу человека, а иногда — полностью его заменить. Уже сегодня роботам поручают не только скучные, рутинные задачи, отнимающие много времени и сил. Умные механизмы способны избавить людей от необходимости рисковать здоровьем, а порой и жизнью, при выполнении работ и научных исследований, связанных с риском [13]. Обсудите вопросы: как изменится мир, когда мы сможем переложить на плечи роботов всю рутину и обезопасить человека-оператора? Какие новые профессии появятся, а от каких получится полностью отказаться? Помогите сформулировать обучающимся цель на урок.

Обсудите, в каких сферах и каким образом используются технологии искусственного интеллекта и машинного обучения в жизни человека.

Далее обучающимся предлагается рассмотреть Атлас новых профессий [9]. Это альманах профессий будущего, актуальных для российской экономики. Он основан на данных форсайт-сессий, в которых принимали участие ключевые игроки рынка. Атлас поможет понять, какие отрасли будут активно развиваться, какие в них будут рождаться новые технологии, продукты, практики управления и какие новые специалисты потребуются работодателям. Самая последняя версия, Атлас 3.0, написана специально для

подростков, простым и увлекательным языком и полна интересных историй о специалистах будущего [9].

Затем можно разделить обучающихся на группы. Необходимо рассмотреть профессии в ИТ-секторе. Для начала прочесть представленный в Атласе образ будущего. Затем выбрать наиболее понравившуюся профессию, рассмотреть, какие надпрофессиональные навыки и умения необходимы для нее. С помощью сети Интернет узнать о развитии карьеры в выбранной профессии. После этого, спросить у каждой группы о результатах работы с Атласом.

После вводной части учителю следует обратить внимание на то, что в области Data Science (науки о данных) работают люди разных профессий: исследователь данных (Data Scientist, архитекторы данных, эксперты по машинному обучению, администраторы баз данных, специалисты по визуализации данных и др.

Возглавляют этот список, безусловно, исследователи данных, которые, работая с массивами информации, формируют прогнозы и гипотезы о развитии явлений. Такие люди помогают компаниям лучше понимать собственные процессы и ближе узнавать клиентов, принимать управленческие решения. И, что немаловажно, они умеют представлять свои открытия в наглядном и понятном виде.

Далее обучающиеся попробуют себя в роли исследователей с помощью табличного процессора MS Excel.

Рекомендации по проведению практической части урока

Цель практической части урока: показать на практике некоторые аспекты работы специалиста в области исследования данных (Data Science) (аналитика данных), связанные с обработкой и извлечением смысла из массива данных: их структурированием, нахождением закономерностей, визуализации и построением выводов для принятия решений.

Лабораторная работа размещена на сайте всероссийского образовательного проекта в сфере цифровой экономики «Урок цифры». В

материалах для учителя к уроку «Искусственный интеллект и машинное обучение» [1].

Предлагаемая для выполнения лабораторная работа включает две части:

- Часть I. Анализ данных с помощью сводных таблиц.
- Часть II. Визуализация и анализ данных с помощью функции условного форматирования.

Лабораторная работа включает алгоритмы и пошаговые инструкции по выполнению заданий с анализом реакции системы (скриншотами). Кроме того, учитель имеет доступ к файлам с исходными данными, используемыми в задачах, что позволит значительно оптимизировать процесс, не тратя времени на набор данных (ввод информации).

Подробное описание действий в инструкциях (выделите, откройте и т.п.) позволит включить в работу даже школьников, не имеющих опыта работы с табличным процессором.

Примечание: следует обратить внимание школьников, что сводные таблицы — это эффективный инструмент для вычисления, сведения и анализа данных, который упрощает поиск сравнений, закономерностей и тенденций.

Используемый во второй части лабораторной работы пример с используемыми командами условного форматирования ориентирован на визуализацию данных и использование цвета в качестве дополнительного источника информации.

Рекомендации по проведению рефлексии

В конце урока имеет смысл проведение ретроспективной рефлексии, которая служит для анализа уже выполненной в течение всего урока деятельности.

В данном уроке содержательно вписывается групповая рефлексия в формате «4-3-2-1».

Школьникам предлагается создать заполнить следующий бланк, который представлен в таблице 1.

Таблица 1 - Бланк для рефлексии в формате «4-3-2-1»

Предмет рефлексии	Ответ обучающихся
4 новых <i>профессии</i> , о которых я узнал	
3 новых <i>понятия</i> , которые я узнал	
2 <i>области искусственного интеллекта</i> , которые считаются наиболее перспективными	
1 <i>проблема</i> , связанная с развитием искусственного интеллекта	

Урок 4. Основные подходы к машинному обучению

Цель урока:

Сформировать у обучающихся представление о основных подходах к машинному обучению и получить представление о специфике работы специалистов в области исследования данных и машинного обучения.

Задачи урока:

- Обсудить особенность машинного обучения;
- Рассмотреть три составляющие машинного обучения;
- Разобрать подходы к машинному обучению;
- Расширить представление школьников о задачах и основных этапах работы специалиста в области машинного обучения;
- Решить практическую задачу машинного обучения (на примере распознавания изображений и принятия решений беспилотным автомобилем на дороге), создав собственную модель (обучения) и применив её в игровых ситуациях, используя предлагаемые наборы данных.

Подготовка и планирование урока

Традиционным и наиболее распространённым вариантом построения структуры занятия является урок, включающий три основных этапа:

Этап 1. Беседа, которая обеспечивают непосредственную включённость школьников в обсуждаемую тему, осмысление её ключевых элементов, анализ примеров.

Этап 2. Получение и отработка практических навыков в процессе в процессе работы с тренажером, ориентированном на погружение и знакомство с профессиями в области работы и исследования данных, связанных непосредственно с машинным обучением и искусственным интеллектом.

Этап 3. Подведение итогов и рефлексия.

Ниже приводится специфика организации этапов урока.

Рекомендации по проведению вводной части урока

Предлагается выстроить беседу, в ходе которой обсуждаются вопросы:

- В чем особенность машинного обучения?
- Каковы три составляющие машинного обучения?
- Что такое датасет?
- Почему в машинном обучении необходимы параметры, где они содержатся?
- Почему в машинном обучении используются различные методы?
- Какие существуют подходы к машинному обучению?
- В чем особенность классического обучения?
- Чем способ обучения с учителем отличается от способа обучения без учителя в классическом подходе к машинному обучению?
- В чем особенности обучения с подкреплением?
- В чем особенности ансамблевого обучения?
- В чем особенности обучения нейросетей?

В ответах на эти вопросы можно использовать информацию, представленную в параграфе 1.2.

Рекомендации по проведению практической части урока

Основная идея и цель практической работы с тренажером — профориентация школьников в сфере искусственного интеллекта: получение первичных представлений об основных понятиях, задачах и специфике работы специалистов в области исследования данных и машинного обучения (в ходе игры по разработке интеллектуальной системы для беспилотного автомобиля); формирование мотивации и интереса школьников к дальнейшему погружению в данную профессиональную область.

Тренажер размещен на сайте всероссийского образовательного проекта в сфере цифровой экономики «Урок цифры» к уроку «Искусственный интеллект и машинное обучение» [1].

По сюжету игры школьники вместе с главными героями Урока цифры (братья Слэши и Скобец) попадают в Институт инновационных разработок, где их подруга Запятакья проходит стажировку на проекте по разработке беспилотного автомобиля. В ходе выполнения заданий обучающимся предстоит познакомиться с устройством беспилотника, попробовать себя в роли специалистов по работе с данными и машинному обучению, чтобы научить автомобиль распознавать различные классы объектов на дороге (дорожные знаки, пешеходов, транспортные средства и элементы разметки) и самостоятельно передвигаться по городу также безопасно, как если бы им управлял опытный водитель.

Перед началом этапа работы с тренажером учителю рекомендуется погрузить школьников в предметное поле, актуализировать предстоящую работу, а также разобрать с ними новые понятия, которые пригодятся им при работе с тренажером. Для этого учитель может воспользоваться следующим алгоритмом работы:

1. Сообщите обучающимся о том, что одной из самых интересных задач, которые сегодня решают специалисты по искусственному интеллекту и машинному обучению, является компьютерное зрение.

2. Сообщите школьникам о том, что на этом уроке им предстоит попробовать себя в роли специалистов по машинному обучению, создав собственную модель (обучения беспилотного автомобиля) в игровом тренажере Урока цифры и применив её в игровых ситуациях. При этом важно обратить внимание школьников на тот факт, что каждый игровой уровень приближен к реальным этапам работы специалистов по машинному обучению [1].

3. Переходите к самостоятельной работе обучающихся с тренажёром. Важно отметить, что для мотивации обучающихся и установки на достижение наилучших результатов учитель может опираться на три игровых элемента мотивации обучающихся, заложенных в тренажере:

а. Так, в процессе выполнения заданий школьники проходят этапы работы и карьеры специалиста в области машинного обучения, получая карточки-достижения, соответствующие выполненной задаче: «Сборщик данных», «Аналитик данных», «Дата-инженер», «Ведущий инженер».

б. Результат выполнения каждого задания оценивается по шкале из трех звезд, где 1 звезда — самый низкий результат, а 3 звезды — наилучший. Таким образом, они смогут получать от платформы обратную связь (насколько точно они справились с заданием), анализировать и корректировать свои действия, соревноваться с собой или одноклассниками, стремиться улучшать свои результаты. По результатам прохождения тренажера выводится финальная оценка, которая рассчитывается автоматически по среднему арифметическому точности выполнения задания на каждом этапе.

с. По окончании урока или прохождения тренажера школьники могут сохранить и распечатать именной электронный сертификат Урока цифры [1].

4. После завершения работы с тренажёром проведите итоговую рефлексию.

Рекомендации по проведению рефлексии

В конце урока имеет смысл проведение ретроспективной рефлексии, которая служит для анализа уже выполненной в течение всего урока деятельности.

В данном уроке содержательно вписывается таблица ЗИУ. Этот метод является одним из самых простых и действенных приёмов для актуализации знаний в начале урока и рефлексии в конце урока, для повышения мотивации учеников, формирования умений оценивать собственные образовательные достижения. В начале урока обучающиеся заполняют первый столбец таблицы «Знаю». Второй столбец таблицы «Интересуюсь» («Хочу узнать») это определение того, что школьники хотят узнать, пробуждение интереса к новой информации. Последняя часть таблицы «Узнал» заполняется по завершению урока [16].

Урок 5. Задачи и методы классического машинного обучения: Обучение с учителем

Цель урока:

Сформировать представление о задачах и методах машинного обучения с учителем.

Задачи урока:

- Рассмотреть машинное обучение с учителем;
- Раскрыть особенность обучения с учителем;
- Рассмотреть типовые задачи, решаемые методами обучения с учителем;
- Решить практическую задачу машинного обучения (с использованием тренажеров Teachable Machine), создав собственную модель (обучения) и убедиться в правильности ее работы.

Рекомендации по проведению вводной части урока

Расскажите обучающимся, что первые алгоритмы пришли к нам из чистой статистики еще в 1950-х. Они решали формальные задачи — искали

закономерности в цифрах, оценивали близость точек в пространстве и вычисляли направления [21].

В направление классического обучения входят достаточно простые методы обучения, другими словами алгоритм обучения. Это направление фактически самое популярное. Так с помощью классических методов выполняется сортировка поисковой выдаче на веб-сайте с учетом предпочтений пользователя, блокировка денег на банковской карточке при определенных действиях с ней в сети и так далее. Классическое обучение используется в том случае, если параметры определить достаточно просто и сами данные являются простыми.

В классическом обучении выделяем два вида: обучение с учителем и обучение без учителя.

В рамках обучения с учителем машине предлагается правильная выборка, в которой все данные разделены. Это может быть в тех случаях, когда данные численные или заранее категоризированы. Так, если машину обучают отличать на фотографии кошек от собак, то эти фотографии заранее разделены (размечены) на две категории, поэтому машина обучается на конкретных примерах.

Для обучения с учителем характерно решение таких задач, как классификация и регрессия. В таком обучении параметры определены, а данные размечены.

Скажите про тип задач классификации. В этой задачи необходимо определить класс объектов и разделить объекты на основе заранее известных параметров. Данный тип задач лежит в основе решения следующих проблем: разработки спам-фильтров, определение языка, поиска похожих документов, анализа тональности, распознавания рукописных букв и цифр, определения подозрительных транзакций.

Следующей рассмотрим задачу регрессии. Эта задача связана с прогнозированием значения непрерывных параметров какого-либо объекта. По своей сути, это задача на классификацию, но в которой вместо категории

объектов, другими словами класс объекта, прогнозируется число. Данный тип задач лежит в основе решения следующих проблем: прогнозирование стоимости ценных бумаг, анализа спроса или объема продаж, установление медицинских диагнозов, выявление любых зависимостей числа от времени, определение стоимости автомобиля по его пробегу, прогнозирование количества пробок на дороге в зависимости от времени суток, выявления объема спроса на товар от роста компании.

Укажите, что более подробно с этими задачами обучающиеся познакомятся на следующих уроках.

Рекомендации по проведению практической части урока

Предложите компьютерный эксперимент по машинному обучению с учителем с использованием тренажеров Teachable Machine [7].

Пользоваться тренажером очень просто. В первую очередь нужно собрать данные. Вы увидите слева то, что называется классами. Каждый из них представляет категорию, узнавать которую нужно научить компьютер. Необходимо дать название каждому классу и добавить данные. Можно загружать данные со своего компьютера или использовать веб-камеру. Также есть функция добавления данных в форме спектрограммы, которая является другим способом визуализации звука. Можно сделать распознаватель, который различает что-то короткое, например, аплодисменты или другой жест. Чтобы создать позу для модели идет отслеживание ключевых точек на теле.

Остается только обучить модель, нажав соответствующую кнопку. После обучения появится окно предпросмотра справа. С помощью него вы можете убедиться, что модель классифицирует новые примеры правильно.

Рекомендации по проведению рефлексии

После практической части зафиксируйте результат урока. Задайте ученикам вопросы:

- Какая информация была для вас новой?

- В чем особенность обучения с учителем?
- Какие задачи решаются методами обучения с учителем?
- Почему тренажер Teachable Machine относится к обучению с учителем?

Урок 6. Обучение с учителем: Линейная регрессия

Цель урока:

Сформировать у учеников представление о линейной регрессии, принципе ее работы.

Задачи урока:

- Обсудить задачу линейной регрессии;
- Ввести понятие «линейная регрессия»;
- Разобрать принцип работы линейной регрессии;
- Решить практическую задачу машинного обучения (на примере предсказания оценки для игры и по данным в Excel).

Рекомендации по проведению вводной части урока

Из прошлого урока обучающиеся знают, что задача регрессии относится к обучению с учителем. С помощью этой модели устанавливается влияние независимых переменных-признаков на зависимую целевую функцию.

Задача линейной регрессии состоит в прогнозировании (предсказывании) конкретного числа, опираясь на входные данные.

В данных каждый объект описывается парой x, y :

- x — данные (многомерный вектор);
- y — целевое значение (может быть любым действительным числом).

Надо найти функцию $f(x)=y$. Эту функция называется целевой.

Регрессия называется линейной, так как в её основе лежит линейная функция, значит и связь между переменными линейная. График этой

функции — прямая. Точки будут располагаться примерно на одной линии, как показано ниже. Этот график-линия показывает ожидаемое значение для каждого объекта по какой-то отдельной характеристике.

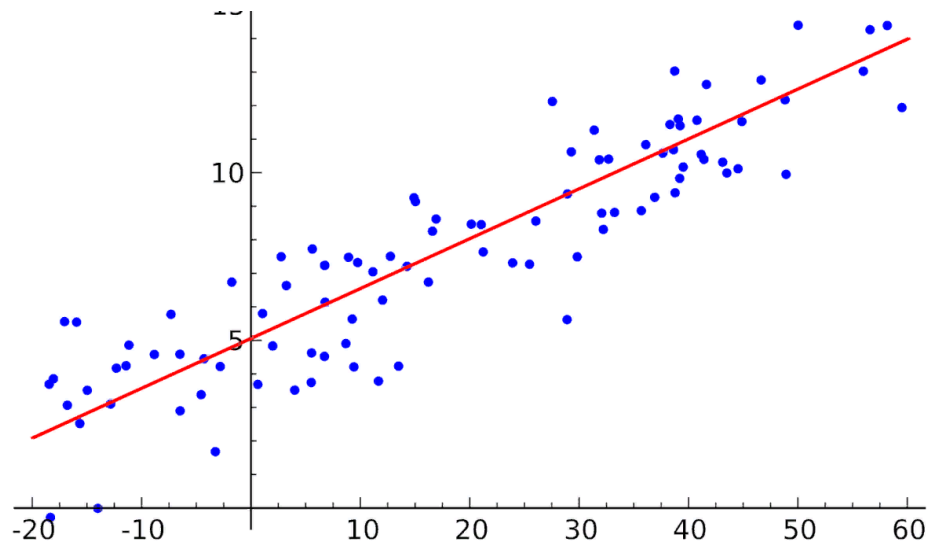


Рисунок 7 — Пример графика линейной регрессии

Линейная регрессия — модель, которая строит прямую линию, опираясь на входные параметры. Формула совпадает с уравнением прямой, которую проходят в школе: $y = kx + b$. Единственное отличие в том, что линейную регрессию часто строят на многомерных данных (то есть количество признаков больше одного, но все они остаются линейными) [8].

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p$$

Константа → b_0 Признаки → x_1, x_2, \dots, x_p
Веса → b_1, b_2, \dots, b_p

Рисунок 8 — Формула линейной регрессии

Опишем обозначения:

Зависимая переменная (\hat{y}) — значение, предсказанное моделью (y с крышечкой). Это переменная, описывающая процесс, который пытаются предсказать или понять.

Независимые переменные (x_i) — i -й признак (например, x_2 — 2-й признак). Это переменные, которые используются для прогнозирования целевого значения, то есть зависимой переменной. В уравнении регрессии они располагаются справа от знака равенства.

Коэффициенты регрессии (b_i) — это значения, которые получают в ходе анализа. Для каждой независимой переменной получают отдельную величину. Эти коэффициенты показывают силу и тип взаимосвязи независимой переменной по отношению к зависимой. Сложность модели будет зависеть именно от числа коэффициентов.

b_i — коэффициент i -го признака (например, b_2 — коэффициент для второго признака).

b_0 — свободный коэффициент.

Все, что нужно сделать, — это умножить каждый признак на коэффициент и сложить их. Коэффициент — это вес, который показывает важность того или иного признака. Чем больше коэффициент у конкретного признака, тем сильнее он влияет на прогнозируемое значение.

Задача машинного обучения — определить коэффициенты перед признаками или, как их чаще называют, веса.

Линейную регрессию нельзя использовать на любых данных. Чтобы построить модель, нужно учесть определенные требования:

1. Набор независимых переменных X должен линейно влиять на целевую переменную. То есть при увеличении признака X линейно увеличивается и целевая переменная.

Например, в магазине сумма покупок линейно и напрямую зависит от того сколько и какие товары кладутся в корзину. Цены на товары — это независимые переменные, а их веса — количество единиц товара. Если взять зависимость роста от количества прожитых лет, то это уже нелинейная

зависимость. В начале жизни люди очень быстро растут, но чем старше становятся, тем медленнее увеличивается рост, а к старости — даже начинает уменьшаться.

2. Разброс данных должен быть постоянным на всем протяжении прямой графика. Это требование к данным называют гомоскедастичность. Если в начале разброс данных очень мал, а в конце, наоборот, велик, то такие данные невозможно описать линейной регрессией.

На рисунке слева — данные гомоскедастичны и подходят для линейной регрессии, а на рисунке справа данные не подойдут для линейной регрессии

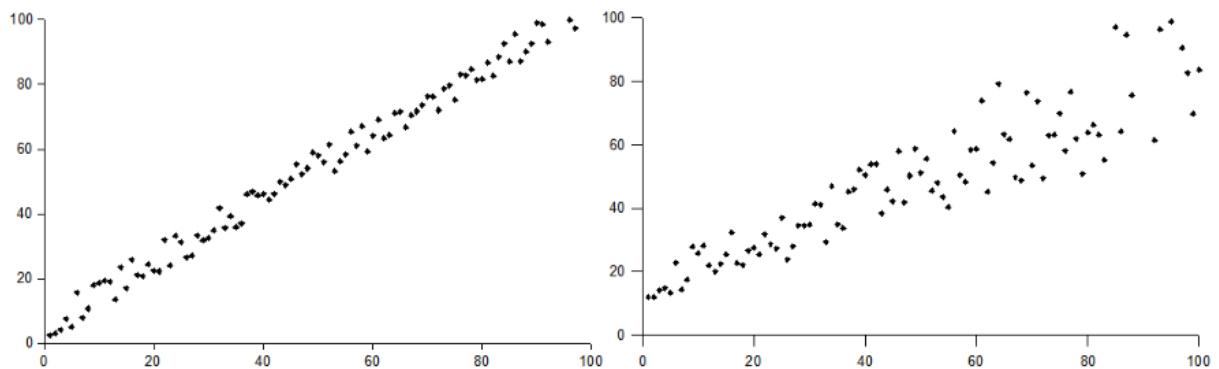


Рисунок 9 — Графики данных

Когда убедились, что данные подходят для линейной регрессии, можно составлять уравнение. Задача — составить такое уравнение, которое вычислит результаты, близкие к правильным, используя входные данные.

Ответы модели не сразу будет идеальными, поэтому для её создания берут веса в уравнении случайным образом. Этот способ позволяет быстро создать рабочую модель, особенно если много признаков [8].

Рекомендации по проведению практической части урока

Разбор примера с предсказанием оценки для игры Fortnite.

Данные будут использоваться из таблицы. Оценка игры от пользователей уже заранее известна.

Игра	Средняя оценка пользователей	Количество часов в игре	Оценка
Minecraft	8.4	72	8
Dota 2	9	118	9.8
Fortnite	9.3	68	8.7

Рисунок 10 — Данные для предсказания оценки для игр

Всего 2 параметра — средняя оценка пользователей и количество часов, проведенных пользователем в игре. Составляется уравнение:

$$\hat{y} = b_1x_1 + b_2x_2, \text{ где:}$$

\hat{y} — это оценка пользователя за игру;

x_1 — средняя оценка других пользователей;

x_2 — количество часов, проведенных в игре;

b_1 и b_2 — веса перед параметрами, которые мы и будем подбирать.

В первом варианте уравнения берем случайные веса. Например, $b_1 = 0.4$, а $b_2 = 0.5$. Составленное уравнение со случайными весами и будет первой моделью: $\hat{y} = 0,4x_1 + 0,5x_2$

Далее совершенствуем модель: изменяем веса для более точного ответа. Чтобы понять, как изменить веса, вычисляется результат и оцениваться. Оценки для первых двух игр, используя веса 0.4 и 0.5:

$$\hat{y}_1 = 0.4 * 8.4 + 0.5 * 72 = 3.36 + 36 = 39.36$$

$$\hat{y}_2 = 0.4 * 9 + 0.5 * 118 = 3.6 + 59 = 62.6$$

Составили первое уравнение и создали модель, теперь можем начать её обучать. Рассмотрим, что именно компьютер скрывает и выполняет без вмешательства человека [8].

На каждом шаге корректировки весов модель оценивает полученные результаты с целевыми при помощи специальной функции. Эта функция показывает, куда двигаться, чтобы прийти к правильному ответу. Оценки пока далеки от реальных. На результат сильно влияет второй параметр, поэтому изменяем его коэффициент в 10 раз и видим результат:

$$\hat{y}_1 = 0.4 * 8.4 + 0.05 * 72 = 3.36 + 3.6 = 6.96$$

$$\hat{y}_2 = 0.4 * 9 + 0.05 * 118 = 3.6 + 5.9 = 9.5$$

В этот раз результат намного лучше. С каждым шагом в обучении все ближе к цели. Как сильно можно приблизиться к идеальному результату, зависит от того, какая ошибка в прогнозе будет устраивать [8].

Далее необходимо изменять коэффициенты, пока результат не станет точнее и протестировать модель на третьей игре. Пусть обучающиеся вычислят наилучшие коэффициенты.

$$\hat{y}_1 = 0.7 * 8.4 + 0.03 * 72 = 5.88 + 2.16 = 8.04$$

$$\hat{y}_2 = 0.7 * 9 + 0.03 * 118 = 6.3 + 3.54 = 9.84$$

$$\hat{y}_3 = 0.7 * 9.3 + 0.03 * 68 = 6.51 + 2.04 = 8.54$$

Теперь можно построить линейную регрессионную модель по данным в Excel.

Предложите следующую задачу. Есть выборка данных за несколько лет о числе ЧП на территории предприятия и количестве трудоустроенных работников. Необходимо выявить взаимосвязь между этими двумя переменными. Есть объясняющая переменная X — это число рабочих и объясняемая переменная Y — это число чрезвычайных происшествий. Исходные данные распределены в двух столбцах.

	А	В
1	X (число работников)	Y (число чрезвычайных ситуаций)
2	10	3
3	11	3
4	12	4
5	14	5
6	10	6
7	11	5
8	16	8
9	18	7
10	20	10
11	21	12
12		

Рисунок 11 — Данные о предприятии

Постройте точечную диаграмму по данным. Затем нужно добавить линейную линию тренда. В дополнительных параметрах поставить галочку

на показ уравнения на диаграмме. Таким образом, функция будет иметь вид:
 $Y = 0,64 * X - 2,84$.

Что это даёт? Это даёт возможность составить прогноз. Допустим, на предприятие нанимается 25 работников и нужно примерно представить, каким при этом будет количество чрезвычайных происшествий. Подставляем в функцию данное значение и получаем результат $Y = 0,64 * 25 - 2,84 = 13,16$. Примерно 13 ЧП будет происходить.

Рекомендации по проведению рефлексии

В заключении урока можно использовать «Метод пяти пальцев». Обучающиеся должны ответить на определенные вопросы.

М (мизинец) — мыслительный процесс. Какие знания, опыт я сегодня получил?

Б (безымянный) — близость цели. Что я сегодня делал и чего достиг?

С (средний) — состояние духа. Каким было сегодня моё преобладающее настроение?

У (указательный) — услуга, помощь. Чем я сегодня помог, чем порадовал или чему поспособствовал?

Б (большой) — бодрость, физическая форма. Каким было моё физическое состояние сегодня? [16]

Урок 7. Обучение с учителем: Дерево решений

Цель урока:

Сформировать общие представления о модели «дерево решений» и принципах его построения.

Задачи урока:

- Рассмотреть дерево решений как модель машинного обучения;
- Ввести понятия «узел», «листья», «глубина дерева»;
- Разобрать принцип построения дерева решений;
- Решить практическую задачу построения дерева решений.

Предлагаемый план занятия и методические рекомендации

План урока построен по модульному принципу.

Модуль 1. Введение понятия «дерево решений».

Скажем, про модели машинного обучения. Модель это математическая формула, которая переводит значения X в значения Y . Моделей очень много, но среди всех моделей можно выделить 3 наиболее популярных:

- линейные модели;
- деревья решений;
- нейронные сети.

С линейной моделью обучающиеся ознакомились на прошлом уроке. Скажем, что деревья — это всего лишь удобная структура для записи правил:

- в узлах вычисляются условия;
- в листьях (в самом низу дерева) — то, что предсказываем.

Дерево решений полностью эквивалентно набору правил, записанных в более компактной форме. Несмотря на простоту, деревья решений — это мощная модель машинного обучения.[1]

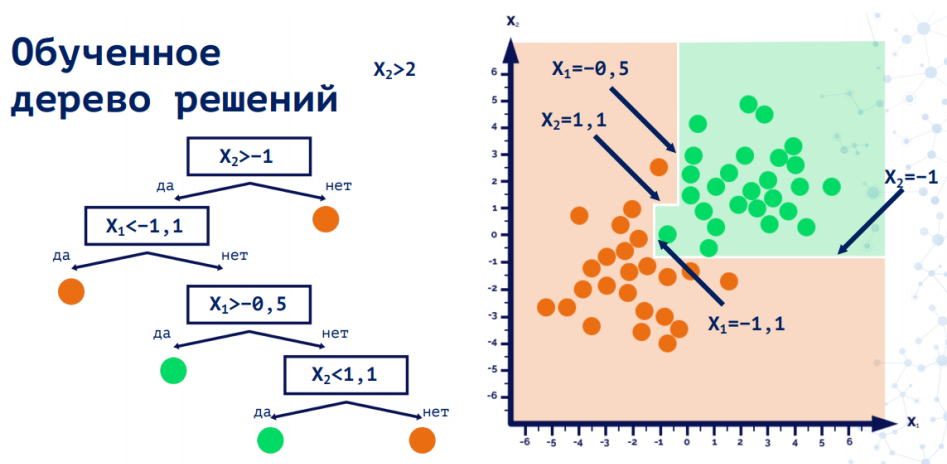


Рисунок 12 — Пример обученного дерева решений

Модуль 2. Дерево решений как модель машинного обучения. На этом этапе обучающиеся на конкретных примерах закрепляют понятия: узел, листья, глубина дерева; на конкретных моделях получают первый опыт построения деревьев решений.

Деревья решений используются и в повседневной жизни в самых разных областях человеческой деятельности, например, для решения задач классификации — предсказания категории объекта.

Зачастую дерево решений служит аналогом обобщения опыта экспертов, средством передачи знаний будущим сотрудникам или моделью бизнес-процесса компании. Например, до внедрения масштабируемых алгоритмов машинного обучения в банковской сфере задача кредитного скоринга (кому выдать кредит, а кому нет) решалась экспертами. Решение о выдаче кредита заемщику принималось на основе некоторых интуитивно (или по опыту) выведенных правил, которые можно представить в виде дерева решений.

Дерево решений эквивалентно набору правил «если то», только записанное в более компактной форме.

При обучении машина сама автоматически вычисляет и задает правила в узлах: разделяет все данные по вопросам, ответы на которые «да» или «нет». Так получается дерево вопросов. Чем выше уровень, тем более общий вопрос. Деревья нашли свою нишу в областях с высокой ответственностью: диагностике, медицине, финансах [1].

Можно попробовать построить деревья решений. Задача машины подобрать такие значения X_1 и X_2 , чтобы выбрать объекты только одного класса по заданному условию.

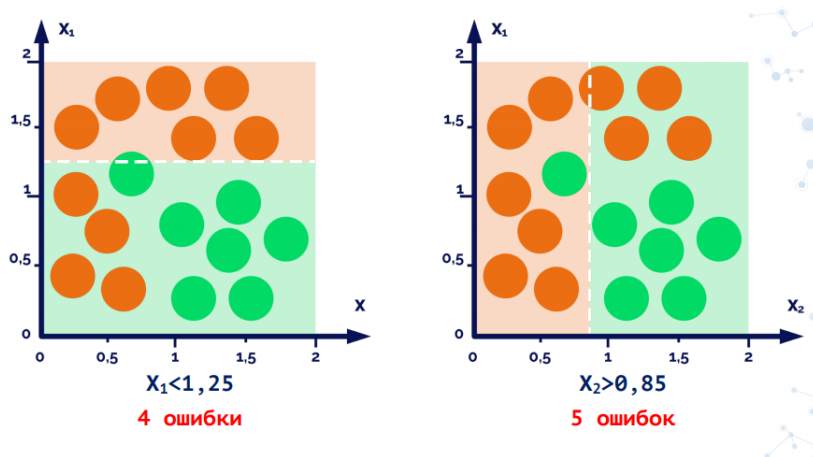


Рисунок 13 — Пример разделения классов

Например, в множестве отелей, выделяем те, в которых хотели бы проживать (это первый класс объектов - зеленые шары), и те, которые не подходят (это второй класс объектов - оранжевые шары). При этом X_1 — это стоимость отеля, X_2 - его уровень.

Машина, задавая диапазон этих числовых параметров, определяет группу объектов, в которой могут оказаться объекты и того, и другого класса.

Задача обучения состоит в том, чтобы подбирая значения X_1 и X_2 выделить отели только одного класса. При работе с деревьями решений человек практически всегда помогает машине, корректируя параметры.

Отделяем прямыми линиями области с зелеными и оранжевыми шарами. Смотри на предыдущий опыт, вопросы на первом и втором уровне очевидны. Затем отделим еще один зеленый шар, определим значения X_1 и X_2 и поставим соответствующие условия. Построенное дерево идеально делит наши шары на 2 класса [1].

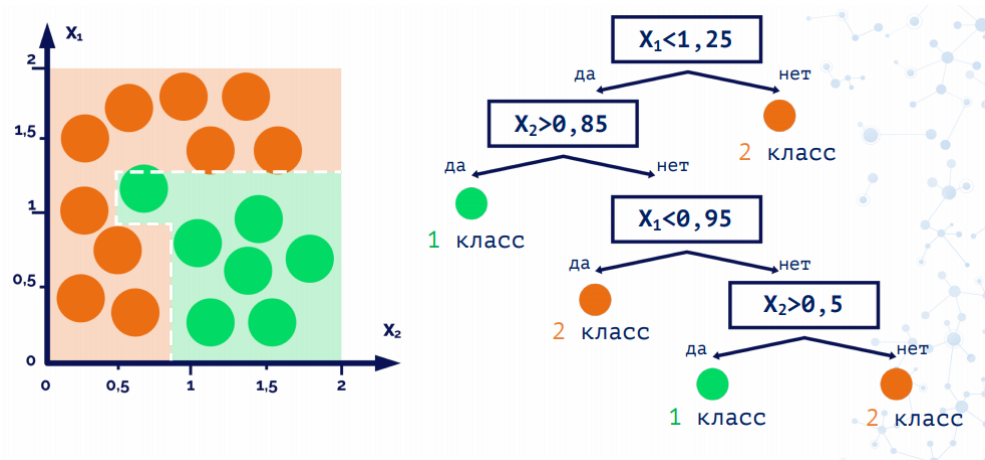


Рисунок 14 — Пример построенного дерева решений

Лист — это узел, не имеющий дочерних элементов.

Узел — любой узел дерева, имеющий потомков, и таким образом, не являющийся листовым узлом.

Глубина дерева — это количество уровней, на которых располагаются его узлы.

Модуль 3. Построение и испытание модели машинного обучения в форме дерева решений.

Обучающимся предлагается конкретное задание — построить модель в виде дерева решений, которая будет предсказывать пойдет сегодня дождь или нет, анализируя только два типа данных: атмосферное давление и относительную влажность воздуха. Им предлагаются реальные данные об атмосферном давлении и относительной влажности за два месяца, а также о том, шел в тот день дождь или нет в одном из городов, представленные на графической модели. Задача школьников построить деревья решений с глубиной 1, 2 и 4.

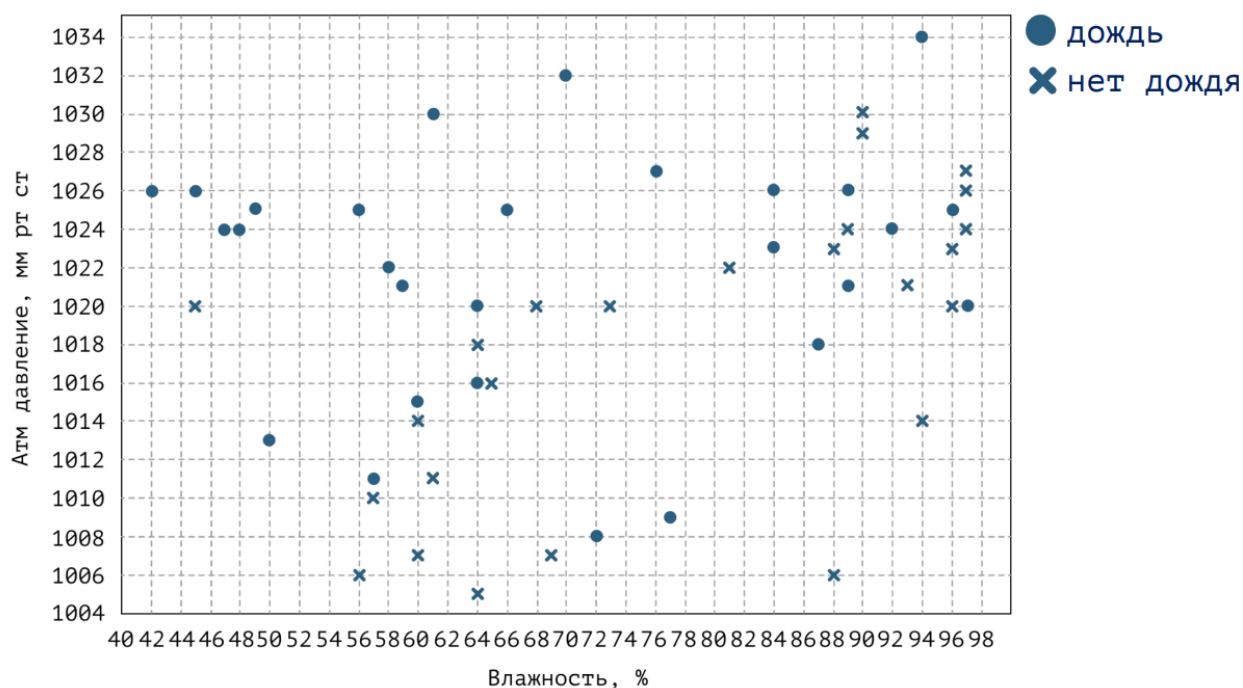


Рисунок 15 — Данные об атмосферном давлении и влажности

На заключительном этапе они анализируют точность всех построенных моделей на реальных полученных данных и делают важный вывод, что самая сложная модель не обязательно показала нам самый лучший результат. На практике всегда приходится искать какой-то компромисс между сложностью и точностью.

В заключении следует еще раз акцентировать внимание на том факте, что деревья нашли свою нишу в областях с высокой ответственностью: диагностике, медицине, финансах и т.п. [1]

Модуль 4. Рефлексия.

«Плюс — минус — интересно». Для письменного выполнения обучающимся предлагается заполнить таблицу из трех граф.

В графу «П» — «плюс» записывается все, что понравилось на уроке, информация и формы работы, которые вызвали положительные эмоции, либо, по мнению ученика, могут быть ему полезны для достижения каких-то целей.

В графу «М» — «минус» записывается все, что не понравилось на уроке, показалось скучным, осталось непонятым.

В графу «И» — «интересно» обучающиеся вписывают все любопытные факты, о которых узнали на уроке, и что бы еще хотелось узнать по данной проблеме, вопросы к учителю [24].

Урок 8. Обучение без учителя: Кластеризация.

Цель урока:

Сформировать у учеников представление о задаче кластеризации, принципе ее решения методом k-средних.

Задачи урока:

- Рассмотреть машинное обучение без учителя;
- Раскрыть особенность обучения без учителя;
- Рассмотреть типовые задачи, решаемые методами обучения без учителя;
- Обсудить задачу кластеризации;
- Разобрать принцип работы метода k-средних;
- Решить практическую задачу машинного обучения (на примере электронного маркетинга по данным в Excel).

Рекомендации по проведению вводной части урока

Обучении без учителя производится на неразмеченных данных. В этом случае, машина в начале выполняет анализ данных, затем объединяет их по выделенным параметрам и лишь потом использует их для обучения. Так,

если машину обучают отличать на фотографии кошек от собак, ей предоставляют сразу все фотографии кошек и собак. Она должна самостоятельно найти закономерности и обучиться различать на фотографиях кошек и собак.

Скажите обучающимся, что в классическом обучении без учителя чаще встречаются задачи кластеризации, обобщения и ассоциации. В таком обучении параметры определены, но данные еще не размечены.

Задача кластеризации связана с группировкой схожих объектов на основе их параметров. Это похоже на задачу классификации, но в которой заранее неизвестны классы. Машина сама ищет похожие объекты и объединяет их в кластеры. Их количество может быть определено машиной или задано заранее. Данный тип задач лежит в основе решения следующих проблем: сегментация рынка (типов покупателей, лояльности), объединению близких точек на карте, сжатие изображений, анализа и разметки новых данных, разработки детекторов аномального поведения, нахождение лиц людей на фотографиях и группировке их в альбом.

Рассмотрите задачу обобщения, связанную с нахождением зависимости между объектами и объединением их параметров в абстракцию, по-другому в параметр более высокого уровня без потери связей между ними. Машина сжимает данные большой размерности в меньшую, это необходимо для визуализации данных или для их использования в других методах машинного обучения. Решая задачи обобщения, мы получаем систему рекомендаций в областях кино, музыки и других. В соответствии с заданными критериями можно осуществить отбор ограниченного множества объектов с большего множества. Например, это представление персонализированного выбора товаров или услуг пользователей магазинов для повышения лояльности. Данный тип задач лежит в основе решения следующих проблем: создание рекомендательных систем, красивых визуализаций, определение тематики и поиска похожих документов, выявление фейковых изображений, анализа риск-менеджмента [8].

Следующей рассмотрим задачу ассоциации. В ней мы находим закономерности в потоке последовательностей, поэтому ее еще называют поиском правил. Данный тип задач лежит в основе решения следующих проблем: прогноз акций и распродаж, анализа товаров, покупаемых вместе, расстановка товаров на полках.

На этом уроке подробно рассматривается задача кластеризации. Кластеризация (или кластерный анализ) — это задача разбиения множества объектов на группы, называемые кластерами. Внутри каждой группы должны оказаться «похожие» объекты, а объекты разных группы должны быть как можно более отличны.



Рисунок 16 — Пример кластеризации

В данном уроке следует рассмотреть самый простой способ кластеризации под названием метод k -средних, который ведет свою историю из 50-х годов и с тех пор стал дежурным в открытии знаний из баз данных (ОЗБД) во всех отраслях и правительственных структурах.

Метод k -средних — не самый математически точный из всех методов. Он создан, в первую очередь, из соображений практичности и здравого смысла. Его несомненным преимуществом является интуитивная простота. Рассмотрим, как работает этот метод, на простом примере [29].

Например, ученики средней школы рассеяны по актовому залу, как показано на рисунке.

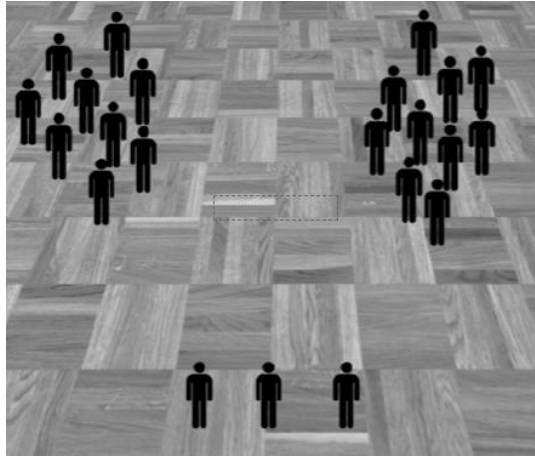


Рисунок 17 — Ученики средней школы рассеяны по актовому залу

Кластеризация по k -средним зависит от количества кластеров, на которое вы желаете поделить присутствующих. Остановимся на трех кластерах. Алгоритм размещает три флажка на полу актового зала некоторым допустимым образом, где вы видите 3 начальных флажка, распределенных по полу и отмеченных черными кружками.

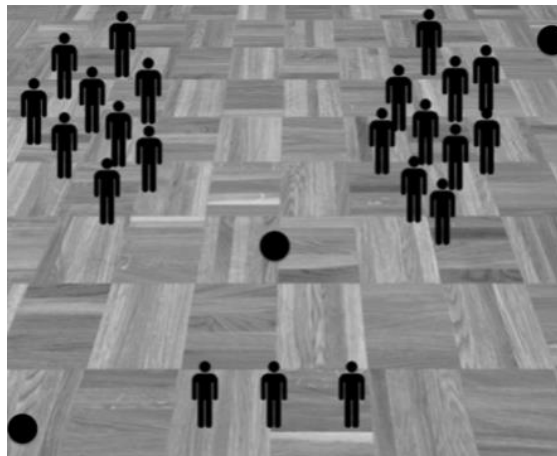


Рисунок 18 — Размещение начальных центров кластеров

В кластеризации методом k -средних танцоры привязаны к ближайшему для них кластерному центру, так что между двумя любыми центрами на полу можно нарисовать демаркационную линию. Таким образом, если танцор находится на одной стороне линии, он принадлежит к одной группе, если по другую сторону — то уже к другой [29].

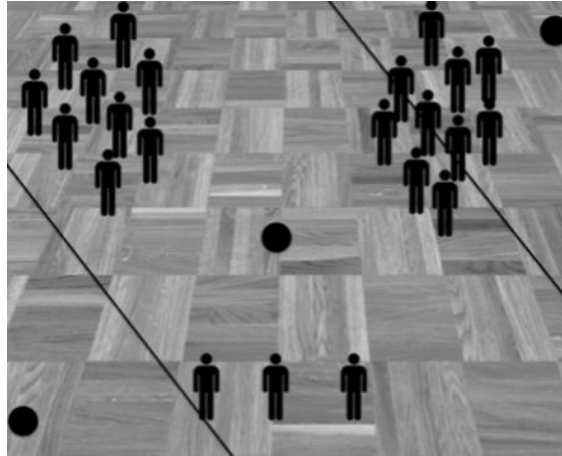


Рисунок 19 — Линии отмечают границы кластеров

Пространство разделено довольно странным образом: нижняя левая группа осталась пустой, а на границе верхней правой группы, напротив, много людей.

Алгоритм кластеризации методом k -средних перемещает кластерные центры по полу, пока не достигнет наилучшего результата. Как определить «наилучший результат»? Каждый присутствующий отстоит на сколько-то от своего кластерного центра. Чем меньше среднее расстояние от участников до центра их группы, тем лучше результат. Этим занимается машина.

Рано или поздно машина вычислит, что центры должны быть размещены в середине каждой группы танцоров.

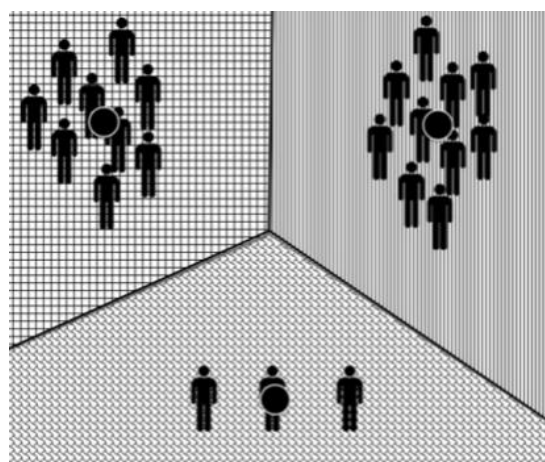


Рисунок 20 — Оптимальная кластеризация

Вот так выглядит идеальная кластеризация. Кластерные центры находятся в центре каждой группы танцоров, минимизируя среднее расстояние между танцором и ближайшим центром. Далее можно переходить

к попытке понять, что же эти кластеры означают, что можно опустить. Таким образом, метод k-средних позволяет разделить множество посетителей танцев на группы и скоррелировать характеристики каждого посетителя с принадлежностью к определенному кластеру, чтобы понять причину разделения [29].

Рекомендации по проведению практической части урока

Чтобы увидеть, как можно применять кластеризацию методом k-средних в реальной жизни, используйте пример «Оптовая Винная Империя Джоуи Бэг О'Донатса» из книги Д. Форман «Много цифр. Анализ больших данных при помощи Excel» [29].

Это импортно-экспортный бизнес, целью которого является доставка огромного количества вина из-за границы и продажи его определенным винным магазинам по всей стране. Этот бизнес работает таким образом, что Джоуи путешествует по всему миру в поисках невероятных сделок с большим количеством вина. Он отправляет его к себе в Джерси, а пристроить присланное в магазины и получить прибыль — будет забота обучающихся.

Покупателей находят разными способами: страница на Facebook, аккаунт в Twitter, порой даже прямая рассылка — ведь электронные письма «раскручивают» большинство видов бизнеса. В прошлом году отправляли одно письмо в месяц. Обычно в каждом письме описываются две или три сделки, скажем, одна с шампанским, а другая с мальбеком.

В итоге заключили около 32 сделок за год и все они прошли более-менее гладко. Но то, что дела идут просто хорошо, не значит, что они не могут идти лучше. Было бы нелишне чуть глубже понять мотивы своих покупателей. Было бы неплохо иметь возможность разбить список клиентов на группы по интересам. Тогда можно было бы отредактировать письма к каждой группе отдельно и, возможно, раскрутить бизнес еще больше. Но как разделить список рассылки? С чего начать? Есть возможность дать компьютеру сделать работу. Используя кластеризацию методом k-средних,

можно найти наилучший вариант разбиения на группы, а затем попытаться понять, почему же он лучший [29].

Документ Excel, который будет использоваться, находится на сайте книги — www.wiley.com/go/datasmart [4]. В нем содержатся все исходные данные. Есть возможность отредактировать данные, чтобы обучающиеся сами выполняли этапы работы над примером.

Рекомендации по проведению рефлексии

После практической части зафиксируйте результат урока. Задайте ученикам вопросы:

- Какая информация была для вас новой?
- В чем особенность обучения без учителя?
- Какие задачи решаются методами обучения без учителя?
- В чем заключается задача кластеризации?
- Какова суть метода k-средних?

Урок 9. Заключительный урок.

Цель урока:

Закрепление знаний, полученных на уроках курса «Введение в классическое обучение»

Предлагаемый план занятия и методические рекомендации

На этом уроке предлагается разделить обучающихся на группы. Работу можно организовать в онлайн-доске Conceptboard [3]. Каждая группа в определенное время составляет на доске список вопросов для других групп, основываясь на изученной информации в течении всего курса. Затем, с помощью комментариев, необходимо ответить на вопросы, которые составляли другие группы. Группы также должны проверять ответы на свои вопросы и отмечать правильные. Засчитывается первый правильный ответ. Чем больше группе засчитали ответов, тем выше шансы занять первое место.

По завершению работы обучающимися, подведите итоги. Далее предложите пройти тест «Какое направление в IT выбрать?» от образовательной платформы «Нетология» [25]. Тест поможет сориентироваться в IT-профессиях и покажет те, которые больше всего подходят обучающимся.

Рекомендации по проведению рефлексии

В конце урока проведите рефлекссию с помощью метода «Плюс-минус-интересно». Обучающимся предлагается заполнить таблицу из трех граф. В графу «П» — «плюс» — записывается все, что понравилось в течении курса, информация и формы работы, которые вызвали положительные эмоции, либо, по мнению обучающегося, могут быть ему полезны для достижения каких-то целей. В графу «М» — «минус» — записывается все, что не понравилось в течении курса, показалось скучным, вызвало неприязнь, осталось непонятным, или информация, которая, по мнению обучающегося, оказалась для него не нужной, бесполезной. В графу «И» — «интересно» — вписывают все любопытные факты, о которых узнали на курсе, что бы еще хотелось узнать в области машинного обучения, вопросы к учителю.

2.3. Апробация курса по выбору и анализ ее результатов

Частичная апробация курса по выбору «Введение в классическое машинное обучение» проводилась на занятиях по дисциплине «Социальная информатика». В апробации принимали участие 31 студент 2 курса Института математики, физики и информатики, обучающиеся по направлению «Педагогическое образование» профиль «Информатика».

В курсе «Социальная информатика» есть тема «Технологии четвертой промышленной революции и их влияние на общество». Основной термин технологической революции — цифровая трансформация. Интернет вещей, искусственный интеллект, машинное зрение, аналитика на основе больших данных, дополненная реальность изменяют абсолютно все сферы жизни


человека: труд, взаимоотношения с государством, учебу, науку, медицину, общение, личные отношения.


В этой теме студенты в том числе знакомятся с основными понятиями и примерами применения технологий искусственного интеллекта. В разделе «Классические методы машинного обучения» курса «Социальная информатика», были использованы разработки занятий курса по выбору по темам «Обучение с учителем» и «Обучение без учителя» [17].


Классические методы машинного обучения

Обучение с учителем


Обучение с учителем: основные понятия


 Обучение с учителем

 Обучение с учителем - тренажер на сайте Teachable Machine


 Прикрепить ответ на задание по использованию тренажера Teachable Machine


Обучение с учителем: линейная регрессия


 Линейная регрессия - описание метода

 Прикрепить ответ на задание по использованию метода линейной регрессии

Обучение с учителем: дерево решений

 Машинное обучение: строим дерево решений

 Прикрепить ответ на задание по построению дерева решений

 Давление-влажность

Обучение без учителя

Обучение без учителя: основные понятия

 Обучение без учителя

Обучение без учителя: Кластеризация методом k-средних


 Кластеризация методом k-средних - описание метода

Рисунок 21 — Фрагмент курса «Социальная информатика» с учебными материалами разработанного курса по выбору

На электронном курсе <https://e.kspu.ru/course/view.php?id=4221> нами были размещены следующие учебные материалы:

1. Презентация «Обучение с учителем». В ней рассмотрены: направления классического машинного обучения, более подробно о обучении с учителем, об особенностях этого обучения, информация о задачах

классификации и регрессии, описано построение модели машинного обучения с использованием тренажера Teachable Machine.

2. Ссылка на сайт Teachable Machine. Это легкий способ создать модель машинного обучения для своего сайта, приложения или другого ресурса. С инструментом справится даже новичок без навыков программирования.

3. Форма для прикрепления ответа на задание по использованию тренажера Teachable Machine. Студентам было необходимо создать свою модель машинного обучения с помощью сайта Teachable Machine.

4. Презентация «Линейная регрессия — описание метода». В ней рассмотрено более подробно задачу линейной регрессии, на примере показано построение линейной модели.

5. Форма для прикрепления ответа на задание по использованию метода линейной регрессии. Необходимо было с помощью линейной линии тренда получить функцию, с помощью которой составить прогноз.

6. Презентация «Машинное обучение: строим дерево решений». В ней рассматривается дерево решений, как модель машинного обучения. На несложном примере, показывается построение дерева.

7. Форма для прикрепления ответа на задание по построению дерева решений. По данной графической модели, необходимо было построить деревья решений с глубиной 1, 2 и 4.

8. Презентация «Обучение без учителя». В ней рассмотрены: направления классического машинного обучения, более подробно о обучении без учителя, об особенностях этого обучения, информация о задачах кластеризации, обобщения и ассоциации.

9. Презентация «Кластеризация методом k-средних — описание метода». В ней, на простом примере, рассмотрен принцип решения задачи кластеризации методом k-средних.

Студенты успешно отвечали на вопросы по информации, представленной в презентациях. С практическими заданиями большинство справилось на отлично.

Нами был проведен опрос студентов, принявших участие в апробации (Приложение А). Студенты отметили простоту и наглядность подачи сложной теоретической информации о машинном обучении, а точнее о классических подходах к машинному обучению и принципах работы алгоритмов классического машинного обучения, интересные практические задания. Студенты сделали вывод о том, что уровень сложности заданий позволяет реализовывать этот курс по выбору в основной школе.

Выводы по 2 главе

В этой главе мы разработали программу курса по выбору «Введение в классическое машинное обучение». Она содержит пояснительную записку, цели изучения курса, планируемые результаты освоения курса, тематическое планирование, содержание курса, требования к уровню подготовки обучающихся и учебно-методический комплект.

Привели планы занятий и некоторые методические рекомендации к ним, чтобы увидеть в какой последовательности и каким образом достигаются запланированные результаты курса.

Некоторые занятия курса по выбору частично апробировали на занятиях по дисциплине «Социальная информатика» на студентах 2 курса Института математики, физики и информатики, обучающихся по направлению «Педагогическое образование» профиль «Информатика».

Заключение

Цель данной работы заключалась в разработке организационно-методического обеспечения курса по выбору «Введение в классическое машинное обучение» для обучающихся основной школы и методических рекомендаций по его изучению.

В ходе выполнения работа были решены следующие задачи:

1. На основе анализа литературы раскрыть понятие искусственного интеллекта и машинного обучения. Искусственный интеллект — научная область, занимающаяся созданием программ и устройств, имитирующих интеллектуальные функции человека. Машинное обучение — обширный подраздел искусственного интеллекта, изучающий методы построения алгоритмов, способных обучаться.

2. Рассмотрены теоретические основы классического машинного обучения. В направление классического обучения входят достаточно простые методы обучения, другими словами алгоритм обучения. Это направление фактически самое популярное. Классическое обучение используется в том случае, если параметры определить достаточно просто и сами данные являются простыми. В классическом обучении выделяем два вида: обучение с учителем и обучение без учителя.

3. Проанализированы имеющиеся в свободном доступе средства обучения для обучения школьников основам классического машинного обучения. Ресурсы для ознакомления с машинным обучением есть, но они не дают полного погружения в эту тему, а лишь частичное. Таким образом появляется необходимость разработать курс по выбору для обучающихся основной школы «Введение в классическое машинное обучение».

4. Разработана программа курса по выбору «Введение в классическое машинное обучение», на основе выявленного ресурсного обеспечения. Программа содержит пояснительную записку, цели изучения курса, планируемые результаты освоения курса, тематическое планирование,

содержание курса, требования к уровню подготовки обучающихся и учебно-методический комплект.

5. Разработаны методические рекомендации по изучению курса по выбору «Введение в классическое машинное обучение». Приведены планы занятий и некоторые методические рекомендации к ним, чтобы увидеть в какой последовательности и каким образом достигаются запланированные результаты курса.

6. Некоторые занятия курса по выбору частично апробированы на занятиях по дисциплине «Социальная информатика». Студенты отметили простоту и наглядность подачи сложной теоретической информации о машинном обучении, а точнее о классических подходах к машинному обучению и принципах работы алгоритмов классического машинного обучения, интересные практические задания. Студенты сделали вывод о том, что уровень сложности заданий позволяет реализовывать этот курс по выбору в основной школе.

Таким образом, задачи работы выполнены, цель достигнута.

Список использованных источников

1. «Урок цифры» — всероссийский образовательный проект в сфере Информационных технологий [Электронный ресурс]. URL: <https://урокцифры.рф/> (дата обращения: 11.05.2021).
2. 10 примеров искусственного интеллекта, который вы используете в повседневной жизни [Электронный ресурс]. URL: <https://ru.gadget-info.com/98593-10-examples-of-artificial-intelligence-youre-using-in-daily-life> (дата обращения: 11.05.2021).
3. Conceptboard | Collaborative Online Whiteboard | Visual Collaboration [Электронный ресурс]. URL: <https://conceptboard.com/> (дата обращения: 11.05.2021).
4. Data Smart: Using Data Science to Transform Information into Insight [Электронный ресурс]. URL: <http://www.wiley.com/go/datasmart> (дата обращения: 11.05.2021).
5. Data-Driven Design: теория. Источники данных и метрики. Толкование данных и принятие решения [Электронный ресурс]. URL: <https://zen.yandex.ru/media/smsnoadress/datadriven-design-teoriiia-istochniki-dannyh-i-metriki-tolkovanie-dannyh-i-priniatie-resheniia-600792f135bbe858550ae799> (дата обращения: 11.05.2021).
6. Teachable Machine - демонстрация машинного обучения [Электронный ресурс]. URL: https://pikabu.ru/story/teachable_machine_demonstratsiya_mashinnogo_obucheniya_5402089 (дата обращения: 11.05.2021).
7. Teachable Machine [Электронный ресурс]. URL: <https://teachablemachine.withgoogle.com/> (дата обращения: 11.05.2021).
8. Академия искусственного интеллекта для школьников [Электронный ресурс]. URL: <http://www.ai-academy.ru/> (дата обращения: 11.05.2021).
9. Атлас новых профессий [Электронный ресурс]. URL: <https://new.atlas100.ru/> (дата обращения: 11.05.2021).

10. Бринк Х., Феверолф М., Ричардс Д. Машинное обучение. СПб.: Питер, 2017. 336 с.
11. Знание [Электронный ресурс]. URL: <https://rusinfo.info/cto-takoe-znanie> (дата обращения: 11.05.2021).
12. Искусственный интеллект: время слабых [Электронный ресурс]. URL: <https://russiancouncil.ru/analytics-and-comments/analytics/iskusstvennyy-intellekt-vremya-slabykh/> (дата обращения: 11.05.2021).
13. Как технологии меняют мир. ИТ-хирург, киберспортсмен и другие профессии будущего [Электронный ресурс]. URL: <https://lenovo.ua/ru/blog/yak-tehnologiyi-zminyuyut-svit-it-hirurg-kibersportsmen-ta-inshi-profesiyi-majbutnogo> (дата обращения: 11.05.2021).
14. Карта применения технологий искусственного интеллекта [Электронный ресурс]. URL: <http://www.iotexpert.ru/node/19282> (дата обращения: 11.05.2021).
15. Классическое машинное обучение: задачи классификации, обобщения, кластеризации данных [Электронный ресурс]. URL: <https://evergreens.com.ua/ru/articles/classical-machine-learning.html> (дата обращения: 10.06.2021).
16. Копилка приёмов рефлексии [Электронный ресурс]. URL: <https://infourok.ru/kopilka-priyomov-refleksii-bolee-priemov-3463151.html> (дата обращения: 11.05.2021).
17. Курс «Социальная информатика» [Электронный ресурс]. URL: <https://e.kspu.ru/course/view.php?id=4221> (дата обращения: 10.06.2021).
18. Левченко, И. В. Модуль «Машинное обучение систем искусственного интеллекта» в общеобразовательном курсе информатики / И. В. Левченко, Д. Б. Абушкин, Л. И. Карташова // Вестник Московского городского педагогического университета. Серия: Информатика и информатизация образования. 2020. № 4(54). С. 27-38.

19. Математическая статистика и большие данные Big Data [Электронный ресурс]. URL: <https://infourok.ru/matematicheskaya-statistika-i-bolshie-dannye-big-data-4322218.html> (дата обращения: 10.06.2021).

20. Машинное обучение [Электронный ресурс]. URL: [www.machinelearning.ru/wiki/index.php?title=Машинное обучение](http://www.machinelearning.ru/wiki/index.php?title=Машинное_обучение) (дата обращения: 11.05.2021).

21. Машинное обучение для людей: Разбираемся простыми словами [Электронный ресурс]. URL: https://vas3k.ru/blog/machine_learning/ (дата обращения: 11.05.2021).

22. Машинное обучение: от теории к практике [Электронный ресурс]. URL: https://www.jetinfo.ru/machine_learning/ (дата обращения: 11.05.2021).

23. Полякова А. Чем различаются слабый, сильный и супер-ИИ [Электронный ресурс]. URL: <https://rb.ru/story/narrow-general-super-ai/> (дата обращения: 11.05.2021).

24. Рефлексия содержания учебного материала. Приёмы [Электронный ресурс]. URL: <https://nsportal.ru/shkola/materialy-metodicheskikh-obedinenii/library/2017/03/06/refleksiya-soderzhaniya-uchebnogo> (дата обращения: 11.05.2021).

25. Тест: Какое направление в IT выбрать? [Электронный ресурс]. URL: <https://netology-code.github.io/who2be/> (дата обращения: 11.05.2021).

26. Указ Президента РФ от 10 октября 2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации» [Электронный ресурс]. - URL: <http://publication.pravo.gov.ru/Document/View/0001201910110003> (дата обращения: 11.05.2021).

27. Федеральный перечень учебников [Электронный ресурс]. URL: <https://fpu.edu.ru/> (дата обращения: 11.05.2021).

28. Форман Дж. Много цифр: Анализ больших данных при помощи Excel [Электронный ресурс]. URL: <https://baguzin.ru/wp/dzhon-forman-mnogo-tsifr-analiz-bolshih/> (дата обращения: 11.05.2021).

29. Форман Дж. Много цифр: Анализ больших данных при помощи Excel. М.: Альпина Пабlishер, 2016. 464 с.

30. Что такое машинное обучение, для чего оно нужно и как работает [Электронный ресурс]. URL: <https://www.calltouch.ru/glossary/mashinnoe-obuchenie/#:~:text=Методы%20машинного%20обучения%20–%20это,решений%20с%20помощью%20искусственного%20интеллекта> (дата обращения: 11.05.2021).

Приложения

Приложение А

Опрос студентов по результатам апробации фрагмента курса по выбору



Оценка раздела

Оставьте отзыв о пройденном разделе «Классические методы машинного обучения». Оцените его структуру и содержание.

* Обязательно

Узнали ли Вы что-то новое для себя? *

- Да
- Нет
- Затрудняюсь ответить

Раздел был хорошо организован и спланирован? *

- Да
- Нет
- Затрудняюсь ответить

Получили ли вы представление о классических подходах к машинному обучению? *

- Да
- Нет
- Затрудняюсь ответить

Понятны ли были Вам принципы работы алгоритмов машинного обучения? *

- Понятны
- Непонятны
- Затрудняюсь ответить

Позволяет ли уровень сложности заданий реализовывать этот курс по выбору в основной школе? *

- Позволяет
- Не позволяет
- Затрудняюсь ответить

Была ли сложной для восприятия теоретическая информация о машинном обучении? *

- Была сложной
- Была несложной
- Затрудняюсь ответить

Насколько сложными показались практические задания? *

- | | | | | | | |
|----------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|--------|
| | 1 | 2 | 3 | 4 | 5 | |
| Несложно | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | Сложно |