

Профориентационное занятие «Пробую профессию в области цифровых технологий»

Введение

Подготовка к уроку Темы 10

Уважаемые педагоги!

Перед проведением профориентационного занятия «Пробую профессию в области цифровых технологий» ознакомьтесь с памяткой во вложении. Профпроба в компьютерном классе предполагает самостоятельное выполнение заданий обучающимися на ПК (индивидуально или в малых группах), в обычном классе — демонстрация заданий педагогом на экране.

Справка для педагога: Data Scientist — это специалист по работе с данными для решения задач бизнеса. Он работает на стыке программирования, машинного обучения и математики. В основные обязанности дата-сайентиста входит сбор и анализ данных, построение моделей, их обучение и тестирование.

Приветствие

Слово педагога: Добрый день, дорогие ребята! Сегодня у нас с вами будет практическое занятие. Именно практика помогает вам примерить на себя профессию и лучше понять, подходит ли она вам. Первый шаг — знакомство с отраслями и профессиями, второй шаг — представить себя на месте конкретного специалиста. Вы уже хорошо знакомы с форматом профессиональных проб. С его помощью вы выполняли рабочие задачи учителя и материаловеда. Вам понравилось выполнять такие задания? Вам это помогло лучше узнать об этих специальностях?

Возможные ответы учеников: «да», «да, стали лучше понимать профессию».

Слово педагога: Отлично! На всякий случай я напомним вам, что все профпробы подготовлены под чутким руководством реальных представителей профессий, конкретно эту профпробу нам помогали делать профессионалы из VK, всем вам известной социальной сети ВКонтакте. Поэтому вы можете быть уверенными, что все задания, которые вам предстоит выполнить, имеют реальное отношение к работе специалиста. Сегодня мы познакомимся со

специалистом, имеющим отношение к цифровым технологиям, посмотрим интересный видеоролик об этой профессии и попробуем выполнить ряд его рабочих задач. Кстати, в этой профессии особенно важно закреплять теорию практикой. Есть идеи, о ком мы сегодня поговорим?

Возможные ответы школьников: «программист», «разработчик мобильных приложений», «робототехник».

Слово педагога: Интересные догадки! Вижу, что вы хорошо запомнили профессии из прошлого урока! Правильный ответ я скажу чуть позже. Сегодняшнюю профпробу мы с вами разберём на уроке все вместе. Также у вас будет возможность пройти эту и другие профпробы самостоятельно в свободное время. В конце урока я расскажу вам, где вы сможете их найти. Вижу, что вам уже не терпится приступить к занятию. Мне тоже! Давайте начнём.

Игра-разминка

Разминка на основе материалов предыдущего занятия (Тема 9. Профориентационное занятие «Россия цифровая: узнаю достижения страны в области цифровых технологий»).

Для проведения игры вы можете использовать Презентацию №1.

Слово педагога: Ребята, о какой профессиональной области мы говорили с вами в прошлый раз?

Вероятный ответ школьников: «область цифровых технологий»

Слово педагога: Верно! А какие профессии вам больше всего запомнились из этой отрасли? Почему?

Возможные ответы школьников: «программист», «робототехник», «ИТ-медик».

Слово педагога: Мне тоже запомнились эти профессии. Спасибо за ответы! Сейчас предлагаю вам провести небольшую игру по мотивам прошлой темы. Не переживайте, это не проверочные задания. При помощи короткой игры «Верю — не верю» мы с вами освежим в памяти факты об области цифровых технологий. На экране появятся утверждения, а вам нужно будет ответить, являются ли они верными или нет. Если да, то поднимайте большие пальцы вверх. Если нет, тогда опускайте большие пальцы вниз. Приступаем!

1) Программистом можно назвать только представителя одной конкретной специальности

НЕВЕРНО. За словом «программист» скрывается множество разнообразных специальностей в области цифровых технологий.

2) Цифровые технологии — это отрасль, которая никак не пересекается с другими отраслями экономики в стране

НЕВЕРНО. Цифровые технологии сегодня применяются практически в любой отрасли.

3) Системы, созданные на основе искусственного интеллекта, могут предотвращать пожары, петь и следить за соблюдением правил безопасности на предприятиях

ВЕРНО. Конечно, всё это — разные системы, но у всех в основе лежит искусственный интеллект.

4) Гуманитариям нет места в мире цифровых технологий

НЕВЕРНО. Наоборот, сегодня всё больше профессий объединяют в себе гуманитарные и ИТ-направления. Например, цифровой лингвист.

5) Система компьютерного зрения позволяет устройствам распознавать объекты и выполнять нужные задачи

ВЕРНО. Это одна из важнейших цифровых технологий, которая необходима, например, робототехникам.

6) Робофутбол создан для развлечения аудитории

НЕВЕРНО. Обучая роботов игре в футбол, робототехники разрабатывают сложнейшие алгоритмы, которые позволяют роботам ориентироваться в пространстве не хуже человека и решать самые разные задачи.

7) В 38 регионах России уже используется беспилотный транспорт в экспериментальном режиме

ВЕРНО. Это позволит приобрести опыт эксплуатации беспилотных автомобилей и подготовить основу для дальнейшего развития этого направления в России.

8) Беспилотную технику можно встретить на российских сельскохозяйственных полях

ВЕРНО. В России в сельском хозяйстве используют беспилотные комбайны и тракторы. Это помогает увеличивать урожай.

9) Крайне редко цифровые технологии используются в медицине

НЕВЕРНО. Медики, наоборот, одними из первых используют прогрессивные технологии, например, для выявления опасных заболеваний у пациентов.

10) Цифровые технологии не могут распознавать эмоции людей

НЕВЕРНО. Специальные программы способны считывать эмоции, и их используют в российских банках, чтобы улучшать клиентский сервис.

Слово педагога: Отлично провели нашу разминку и вспомнили предыдущую тему! Очень радостно, что вы верно ответили на вопросы. Это значит, что вам многое запомнилось из предыдущего урока, и вы заинтересовались сферой цифровых технологий. Давайте разбираться в ней дальше!

Портрет специалиста

Слово педагога: Сегодня мы познакомимся с профессией, которая называется «Специалист по Data Science» (дата саенс). Как вы думаете, кто это?

Ответы обучающихся.

Слово педагога: Не пугайтесь такого сложного названия профессии, на деле всё довольно просто. И сейчас я вам это докажу. «Data Science» (дата саенс), или науку о данных, можно определить как работу с большими данными, или большим объёмом информации. Вот представьте, сколько данных о клиентах хранятся, например, у операторов сотовой связи, если у нас в стране 146,5 миллионов человек? А у некоторых ещё и несколько сим-карт. Человеку не хватит целой жизни, чтобы все эти данные обработать и структурировать. Вот такие данные и называются большими. Их способны анализировать только компьютеры. А теперь усложним задачу. Представьте, что из всего объёма информации о клиентах операторов сотовой связи нужно вычислить только юношей и девушек до 20 лет, определить сколько гигабайт интернета они потребляют в месяц, чтобы предложить им новый тариф? Вот тут в дело и вступают специалисты по Data Science. Они знают как сделать так, чтобы из всего огромного объёма информации компьютер нашёл и проанализировал именно то, что нужно в данный момент. И чтобы компьютер самостоятельно решил, как именно ему это сделать быстрее и качественнее. Это и называется машинное обучение. Когда человек не пишет новую программу всякий раз, когда возникает новая задача, а учит компьютер делать это самостоятельно. Машина сама принимает решения, как именно ей обрабатывать нужные данные, чтобы выдать человеку тот итог, который ему нужен.

И, конечно, специалист по Data Science применяет методы машинного обучения, по сути, цифровые технологии, чтобы работать с большими данными из разных сфер. Сотовая связь — один из многочисленных примеров. Это может быть медицинская статистика или статистика покупок за определённый период, базы данных генетических компаний, метеоданные и многое другое!

Всё равно звучит довольно запутанно, правда? Лучше я передам слово профессионалу в этой области. Внимание на экран!

Видеоролик «Специалист по Data Science»

В видеоролике специалист в области цифровых технологий рассказывает, что же кроется под словосочетанием «искусственный интеллект». А также в нём перечисляются разные профессии, которые связаны с машинным обучением.

Обсуждение ролика

Слово педагога: Какой интересный видеоролик, не правда ли? Ребята, вам понравилось? Вас заинтересовала эта сфера? Что вас больше всего впечатлило? Хотелось бы попробовать себя в этой сфере?

Ответы обучающихся.

Профпроба: «Специалист по Data science»

Формат: Профпроба в обычном классе

Рекомендация

Распределите класс на 3-4 мини-группы, для каждой группы должен быть подготовлен раздаточный материал (см. Приложение к уроку: Раздаточный материал). В сценарии дополнительно будут отмечены задания в раздаточном материале.

Педагог выступает модератором занятия (необходим ПК с доступом в Интернет или заранее установленная программа с пробой на ПК, см. файл в приложении к уроку или в следующем блоке).

Ученики выполняют задания в мини-группах и участвуют в обсуждении заданий.

С целью дополнительного погружения обучающихся в тему вы можете заранее выбрать из класса (или из старших классов, профильных информационно-технологических классов) 1-2 обучающихся, которые смогут выступить в роли модераторов пробы и провести занятие. В этом случае рекомендуется подготовить ребят заранее к данному профориентационному занятию (пройти совместно пробу и проиграть сценарный план).

Доступ к профпробе

Ссылка на профпробу: <https://bvb-kb.ru/ds>.

Введите эту ссылку в браузер компьютера, задействованного для прохождения профпробы или заранее скачайте профпробу в формате .exe. Рекомендуется заранее включить и проверить пробу на ПК.

Важно! Не забудьте ознакомиться с памяткой, размещённой в начале занятия, и организовать рабочее пространство, а также подготовить дополнительные материалы в соответствии с рекомендациями.

Стартовая страница

Слово педагога: Начинаем проходить профпробу. Перед вами стартовая страница новой профессиональной пробы. Прочитаем, что здесь написано и познакомимся с профессией.

Педагог, модератор-ученик или любой желающий зачитывает текст на слайде.

Если вы используете файл в формате .exe, заранее скачайте слайды стартовой страницы, описания задания / что такое искусственный интеллект и заключительного экрана. Файл .exe содержит в себе только этап практики.

Раздаточный материал, страница 1.

Слово педагога: Отлично! Вы наверняка заметили, что в ваших распечатанных материалах остались свободные ячейки, заполните их.

Вопросы для обсуждения:

Почему важны эти специалисты?

Каким компаниям могут помочь специалисты по Data science?

Какие качества нужны специалисту, чтобы хорошо выполнять свою работу?

Задание

Педагог, модератор-ученик или любой желающий зачитывает текст.

Раздаточный материал, страница 2.

Слово педагога: Знаю, что вам хочется поскорее перейти к заданию. Прежде чем мы приступим, хочу обратить ваше внимание на несколько вещей. Среди раздаточных материалов вы найдёте справочник, который станет вашим помощником. С его помощью вы получите ответы на вопросы по теме и найдёте там дополнительную информацию. Вы можете советоваться друг с другом, чтобы находить ответы. Если затрудняетесь дать ответ, вы всегда можете воспользоваться справочником. Помните, что в нашей профпробе не нужно бояться дать неверных ответов. Я не буду ставить вам оценку после профпробы, поэтому не стоит об этом переживать. Всё, что вам нужно сделать, — постараться погрузиться в мир этой профессии и представить, что вы проходите небольшую стажировку. А теперь начинаем!

Что такое искусственный интеллект?

Слово педагога: В сегодняшней профпробе вы потренируетесь выполнять задачи специалиста по Data science. Они будут интересными и важными. Перед тем, как приступить к ним, давайте прочитаем информацию об искусственном интеллекте.

Педагог, модератор-ученик или любой желающий зачитывает текст.

Обратите внимание на «Справочник» — он поможет в прохождении профпробы.

Часть 1: Определение задач

Слово педагога: Приступаем к прохождению профпробы «Специалист по Data science». Ваше первое задание — определить задачи машинного обучения. Перед вами 6 разных задач, внимательно прочитайте их и определите те, которые, как вам кажется, относятся к задачам машинного обучения, и те, которые относятся к продуктовым задачам. Обратите внимание на подсказку, она поможет вам верно распределить задачи. Те варианты, которые наиболее подходят к задачам машинного обучения, отметьте галочкой, а оставшиеся варианты — крестиком.

Раздаточный материал, страница 3.

Ребята отмечают в раздаточном материале задачи машинного обучения галочкой, продуктовые задачи крестиком, потом каждая группа объясняет свой выбор.

Подсказка для педагога:

Задачи машинного обучения:

распознавать данные в аудиодорожке и переводить её в текст

предсказывать наличие знаков препинания в расшифрованных аудиосообщениях

Продуктовые задачи:

увеличивать ежедневное количество пользователей чата на 15%

создать технологию, чтобы люди в любой ситуации могли воспринимать видео- и голосовые сообщения

поднять оценку показателя «Удобный чат» на 10%

дать возможность слабовидящим пользователям проходить «капчу» в виде сгенерированной аудиодорожки

Слово педагога: Ребята, вы успешно справились с первым заданием и теперь вы можете ознакомиться с видами задач машинного обучения.

Педагог, модератор-ученик или любой желающий зачитывает текст.

Часть 2: Сбор и подготовка данных

Слово педагога: Ребята, переходим к следующей странице раздаточных материалов и ко второму заданию. Нам с вами нужно собрать и подготовить данные, которые помогут разработать алгоритм для перевода аудиосообщений в текст. Настоящая мечта многих людей, которые не любят получать аудиосообщения!

Педагог, модератор-ученик или любой желающий зачитывает текст.

Раздаточный материал, страница 4.

Слово педагога: Итак, перед вами шесть аудиозаписей. Ваша задача — понять, какие варианты следует исключить из работы. То есть они не подойдут для машинного обучения. Проще говоря, алгоритм не научится на этих данных выполнять необходимую задачу. Поставьте галочку в ячейки рядом с записями, которые нужно исключить. При необходимости вы можете подглядывать в справочник.

Ученики выполняют задание, потом педагог может попросить учеников из каждой группы объяснить свой выбор. Педагог может исключать названные ребятами ответы в профпробе. Если ответ будет неверный, система это отметит и предложит ответить повторно.

Подсказка для педагога: записи, которые нужно исключить – 2,5,6.

Слово педагога: Отлично, ребята! Вы верно исключили неподходящие аудиодорожки, в которых было много пустых мест. Они были малоинформативными для машинного обучения. Итак, данные отобраны и из них будут сформированы датасеты. Это своего рода учебники для алгоритма, с их помощью он научится выполнять необходимое задание.

Часть 3: Препроцессинг

Осциллограмма и спектрограмма

Слово педагога: Третий этап работы называется препроцессинг. Наши с вами аудиосообщения относятся к звуковой информации, а звук — это волна, колебания которой записываются в виде осциллограммы (*текст со слайда*).

Педагог может зачитать информацию на слайде сам или попросить это сделать ученика.

Слово педагога: Итак, ребята, перед вами настоящая осциллограмма. Ваша задача — найти все пиковые значения на представленной записи и выделить их.

Раздаточный материал, страница 5.

Слово педагога: Посчитайте, сколько пиковых значений тут можно найти, обведите их ручкой и затем назовите цифру.

Ответы школьников.

Слово педагога: Верный ответ — 15 пиковых значений. Отлично!

Отлично справились, ребята! При помощи специального программного обучения вы перевели звуковую дорожку в другой вид отображения. Теперь компьютер может работать с этими данными. А нам с вами нужно проверить полученные данные. Это одна из важных задач специалиста по Data science. Посмотрите на спектрограмму в своих раздаточных материалах и скажите, не упущены ли там пиковые значения? Отметьте или обведите пропущенные пиковые значения. Сколько пропущенных значений нашли?

Ответы школьников. После этого педагог может выделить их на экране.

Подсказка для педагога: 4 пиковых значения были пропущены.

Слово педагога: Вы подготовили и проверили спектрограмму! Именно её специалисты загружают в алгоритм машинного обучения и обучают с её помощью нашего будущего помощника.

Педагог может зачитать текст на слайде или попросить об этом ученика.

Подсказка для педагога: *speech-to-Text* произносится как «спич-ту-текст».

Акустическая модель

Слово педагога: Теперь мы с вами будем работать с акустической моделью. Давайте прочитаем текст на слайде.

Педагог, модератор-ученик или любой желающий зачитывает текст.

Раздаточный материал, страница 6.

Слово педагога: Ваша задача — определить, какие звуки распознал алгоритм. Не верю, что я сейчас это произнесу, но на время выполнения задания — и только на это время — забудьте о правописании. Сейчас вам нужно постараться угадать только звуки, какие мог определить алгоритм. Вам нужно посмотреть на варианты ответов рядом с каждым словом и обвести в круг верную фонему.

Учащиеся вписывают ответы в пропуски в раздаточном материале. После педагог может просить называть ответы каждую группу и выбирать их на экране профпробы.

Подсказка для педагога: верные фонемы — а, а, о, а, ы, э, о, ы, с.

Слово педагога: Вновь вспоминанием правописание и больше никогда не забываем! Тем более, что теперь нам нужно исправить ошибки.

Лингвистическая модель

Слово педагога: После акустической модели нужно поработать с лингвистической моделью.

Педагог, модератор-ученик или любой желающий зачитывает текст.

Слово педагога: Ваше новое задание похоже на работу учителя русского языка, только вам нужно проверить домашнюю работу машинного обучения. Вам нужно исправить ошибки и определить верное правописание для каждого транскрибированного слова. Первое задание на странице 7 раздаточного материала.

Раздаточный материал, страница 7.

Ученики исправляют ошибки и вписывают правильно фразу в раздаточном материале. После педагог может попросить учеников каждой команды выбирать правильные ответы и отмечать их в профпробе.

Подсказка для педагога:

Верные варианты в заданной последовательности:

пойдём

вечером

оно написано верно

в

оно написано верно

оно написано верно

вышел

оно написано верно

класный

оно написано верно

Слово педагога: Не подвели, правописание вспомнили! А как насчёт знаков препинания? Вот вам и доказательство, как важно представителям всех профессий знать русский язык. Если специалист по Data science плохо знает правила грамматики и пунктуации, то он плохо обучит алгоритм. Поэтому хороший специалист развит всесторонне!

Пунктуационная модель

Слово педагога: Вы практически научили свой алгоритм переводить аудиофайлы в текст! Осталось рассказать ему о пунктуации. Если алгоритм будет с ней незнаком, то, увы, многое будет непонятно или вызывать смех в расшифрованных записях. На той же странице раздаточного материала у вас ещё одно задание. Давайте прочитаем текст на карточке (на странице раздаточного материала).

Педагог, модератор-ученик или любой желающий зачитывает текст.

Раздаточный материал, страница 7.

Слово педагога: Перед вами четыре примера расшифрованных аудиосообщений. Ваша задача — поставить галочки там, где должен быть какой-то знак препинания, включая точку.

Комментарии для педагога: Если ученики ещё не знакомы с правилами, которые помогут выполнить эти задания — это не страшно. Они могут выполнять это задание, глядя на экран, или предположить расстановку знаков, а потом сверить с правильным вариантом.

Важно показать сам алгоритм.

Подсказка для педагога:

Правильные варианты ответов:

Прошло жаркое лето, и наступила осень.

Здравствуй, Иван Федорович.

Летом ребята приносили разные цветы: лютики, ромашки, васильки, колокольчики.

Фу, какая мерзость.

Подсказка для педагога:

Акустическая модель:

Гле п А даж к на и на а будет в встре е

Лингвистическая и пунктуационная модели:

Гле Б, зах О ди к П нам на П Д ва па П будет р Д в встре че.

5) Он х Ор шо знает ру С кий Язык П и П это И мущество П П П

Подсказка для педагога:

Акустическая модель:

Он П Ор шо П знает С кий Язык П и П это И мущество П П П

Лингвистическая и пунктуационная модели:

Он х Ор О шо знает ру С С кий Я язык, и это его пр Е имущество.

Слово педагога: Переходим к следующему заданию, страница 10 раздаточного материала, перед вами аудиодорожка. Мы уже выяснили, что «спокойные» участки в аудиодорожке обозначают тишину, а «активные» — звуки. Сейчас вам нужно обвести или заштриховать паузы в аудиодорожке.

Педагог может дать немного времени, чтобы ученики в мини-группах выделили паузу.

Раздаточный материал, страница 10.

Слово педагога: Нашли паузы? Тогда проверяйте за мной!

Педагог выделяет паузы на аудиодорожке в пробе. Всего их 2 штуки.

Слово педагога: Ребята, давайте прочитаем, что тут зашифровано? «Классно быть дата саентистом. Что думаешь?». Ну, как думаете, классно?

Ответы учеников.

Слово педагога: Отлично, ребята! Ваша фантазия безгранична! Вы вполне могли бы создать интересную модель машинного обучения. Это ещё раз показывает, что работа специалиста по Data science тоже является творческой, без этого будет сложно анализировать большие объёмы данных и находить оптимальные решения.

Часть 4: Оценка результата

Слово педагога: Вы долго обучали алгоритм и проделали большую работу: за очень короткий срок рассказали машинному обучению о фонетике, грамматике и пунктуации. Именно по этому принципу работает специалист по Data science, чтобы получить эффективный программный алгоритм. Настало время оценить результат!

Педагог, модератор-ученик или любой желающий зачитывает текст.

Раздаточный материал, страница 11.

Слово педагога: Ваша новая задача — разделить фразу на минимальные цепочки, чтобы алгоритму было проще расшифровать их. Обведите или заштрихуйте необходимые участки.
Педагог может дать немного времени, чтобы ученики в мини-группах выделили участки фразы.

Слово педагога: Готовы? Тогда вновь проверяйте за мной!
Педагог выделяет на аудиодорожке паузы. Всего их – 5 штук.

Заключительный экран

Слово педагога: Вот это да! Почувствовали себя настоящим специалистом по Data science? Давайте мы сейчас подведём итоги нашей интересной профпробы.
Педагог, модератор-ученик или любой желающий зачитывает текст финального слайда.
Раздаточный материал, страница 12.

Заключение

Карточка профессии «Специалист по Data Science»

Дополнительные материалы приложены к сценарию занятия. Заранее распечатать карточку профессии (можно одну на класс, на команду или отдельно для каждого ученика).

Слово педагога: Время урока заканчивается, но мы ещё успеваем выполнить последнее задание на сегодня. Вы могли заметить, что на финальной страничке профпробы был элемент под названием «Артефакт». У нас уже сложилась добрая традиция, когда вы получаете «бонусы» за прохождение профпроб.

При наличии технической возможности, открыть «Артефакт» в классе.

Слово педагога: Вы получили такие «бонусы» в виде карточек после профпробы «Учитель» и «Материаловед». И сегодняшнее занятие не исключение! Вы получаете новую карточку по профессии «Специалист по Data Science». И карточку, на которой вы можете найти новый набор букв. Помните, что эти буквы — фрагменты уже нового слова-пароля. Пройдя все пробы по теме «Цифровые технологии», вы сможете узнать все буквы, составить из них зашифрованный пароль и получить новый свиток знаний. Самостоятельно в свободное время пройти эту и другие профпробы вы можете, переходя по ссылкам, указанным в артефакте или в виртуальном городе профессий Профиград. Советую попробовать себя в новых профессиях. Чем больше знаний о профессиях у вас будет, тем легче вам будет выбрать свой профессиональный путь.

Большая карта отраслей

1) Рекомендуем создать профориентационный уголок (лист / ватман / гирлянда/ коробка — формат на ваше усмотрение), где вы и ребята сможете размещать интересные артефакты по профориентационным занятиям, собирать отзывы ребят и обращать их внимание на приобретённый опыт.

2) Артефакты профпроб. На финальной странице профпроб, справа размещён «Артефакт». Нажав на него, откроется карточка пройденной профессии и набор из букв. Эти буквы — фрагменты слова-пароля. Пройдя все пробы по теме «Цифровые технологии»: «Специалист по Data Science», «Специалист по мехатронике и мобильной робототехнике», «Программист», «Оператор беспилотных летательных аппаратов», «Инженер компьютерного зрения», ребята могут собрать 9 букв, из которых можно составить слово-пароль «ПРОЦЕССОР» (не сообщайте это слово детям, они должны составить его сами). Это слово можно ввести в специальную форму в Профиграде. Она откроет ребятам новый «Свиток знаний» с дополнительными материалами уже об этой отрасли.

3) «Большая карта отраслей». В конце каждого занятия-профпробы ребята будут получать фрагмент карты отрасли. В ней они смогут записывать свои впечатления от занятий, новые знания об отрасли и профессиях, а также спрятанное в профпробах слово-пароль. Соединяя фрагменты друг с другом с помощью скотча, клея или канцелярского степлера, можно постепенно собрать «Большую карту отраслей». Вы можете собирать её в профориентационном уголке или хранить в сложенном виде.

Слово педагога: Ребята, я очень рад(-а), что сегодня мы с вами вместе попробовали себя в роли специалиста по Data Science. Мы все отлично справились, и я думаю, что кто-то из вас в будущем обязательно выберет для себя этот интересный и увлекательный профессиональный путь. В завершении нашего занятия, я хочу, чтобы вы смогли записать то, что вам больше всего запомнилось на этом уроке. Для этого я раздам вам новый фрагмент «Большой карты отраслей». Начиная с первой команды, заполняйте любые пункты и передавайте следующей команде. Слово-пароль вы сможете вписать тогда, когда пройдёте остальные профпробы. Этот листок нам нужно обязательно сохранить, так как на каждом последующем занятии мы будем присоединять к нему новую часть. В конце нашего курса у нас получится большая карта, на которой мы увидим весь путь, который мы прошли. Всем спасибо за занятие!

Задание

Часто для решения сложных задач или даже создания новых IT-продуктов используют искусственный интеллект. Но он появляется не сам по себе: обучением машин занимаются специалисты по Data Science. Именно такую роль вы сегодня и примерите на себя.

Ваша основная задача: разработать модель машинного обучения для перевода аудиоконтента в текст.

Вы разберётесь, откуда вообще появляются алгоритмы и нейросети, на конкретной задаче обучите модель распознавания слов и проверите, насколько эффективно она работает.

А ещё, несмотря на техническую сложность, вам может пригодиться знание русского языка — ведь надо будет указать алгоритму на его грамматические ошибки.

Вперёд, будущие айтишники!

Начинаем!

Этапы

- 1 Определить цель машинного обучения
- 2 Собрать и подготовить данные
- 3 Обучить модель машинного обучения в несколько этапов
- 4 Проверить работу алгоритма с помощью метрик



Я помогу!

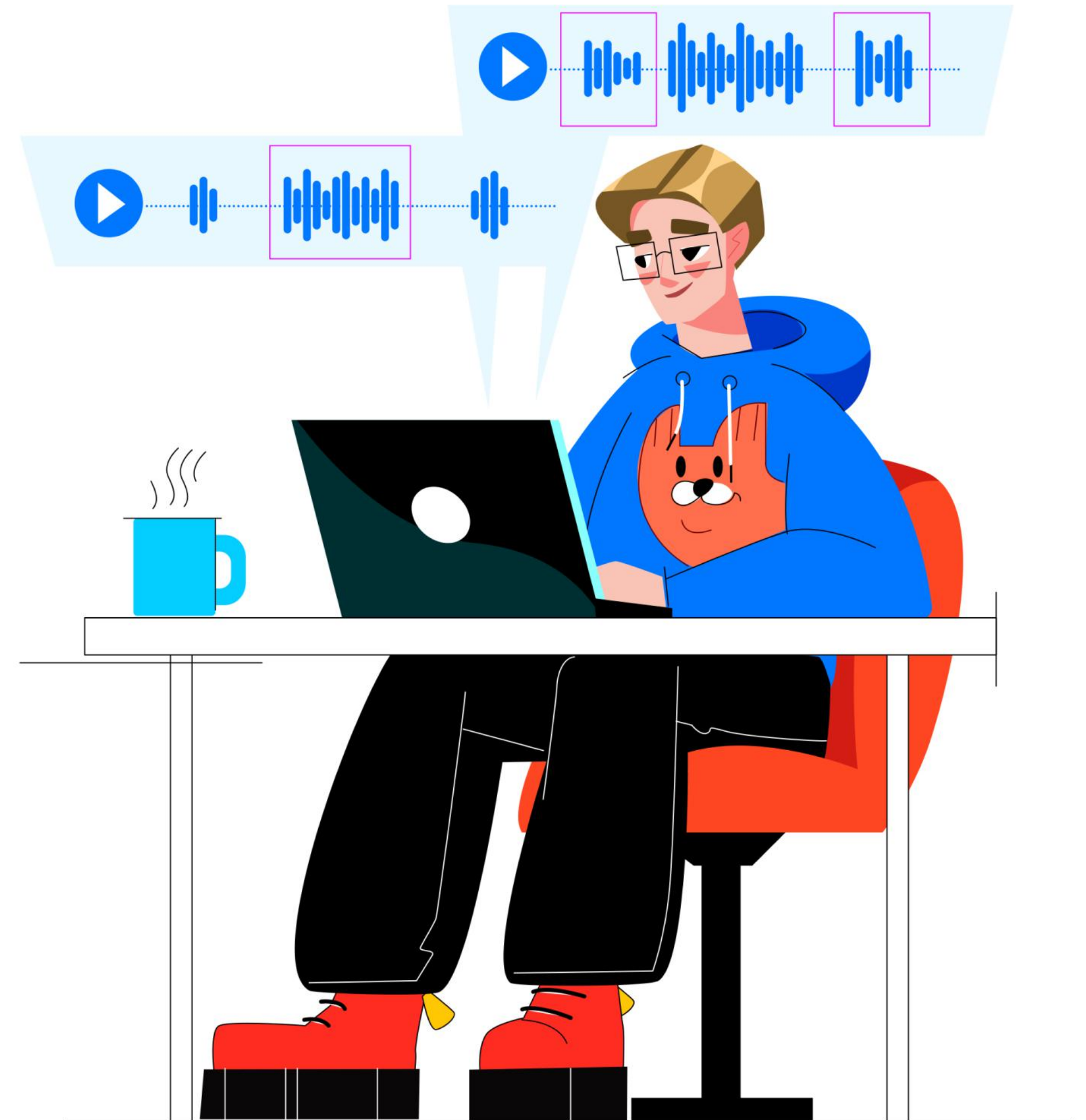
Заглядывайте в справочник, если возникнут трудности или вы захотите узнать больше интересной информации

Отлично, задание выполнено!

Поздравляем! Теперь любое голосовое сообщение или видео можно превратить в разборчивый текст.

Сегодня вы:

- ✓ Познакомились с одной из задач специалиста по Data Science
- ✓ Отличили частоты на осциллограмме от частот на спектрограмме
- ✓ Поработали с несколькими моделями машинного обучения
- ✓ Обучили алгоритм и убедились, что он работает без ошибок



Мне понравилось! Что дальше?

Советы от эксперта



Иван Самсонов
Data Science Lead, VK

1 Участвуйте в хакатонах и чемпионатах по программированию! Это даст колоссальный опыт и знакомства.

2 Развивайте насмотренность. То есть наблюдайте, как другие решают какую-то задачу, какую технологию и где применяют. Это очень важный навык.

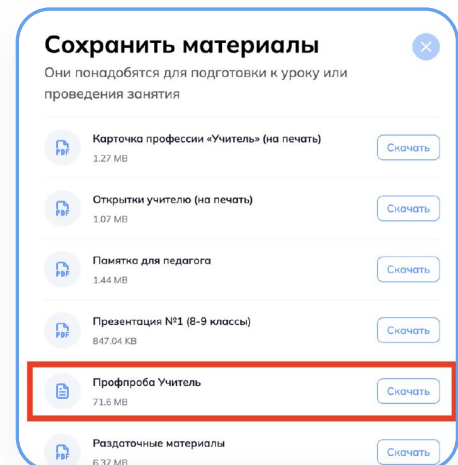
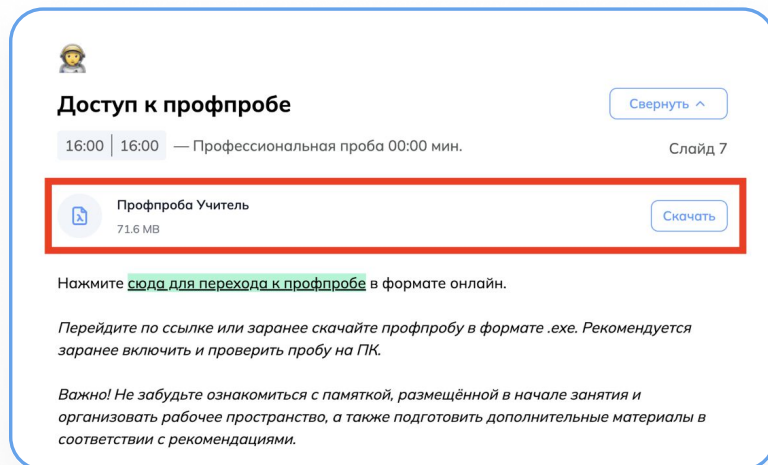
3 Если вы чувствуете, что хотите и ИТ, но пока не определились, начните делать свой первый проект. Прощупайте разные роли: дизайнера, разработчика, менеджера и пр.

4 В интернете много лекций и инструкций, много уже написанных программ. Так вы разберётесь во многих нюансах!

Уважаемые коллеги!

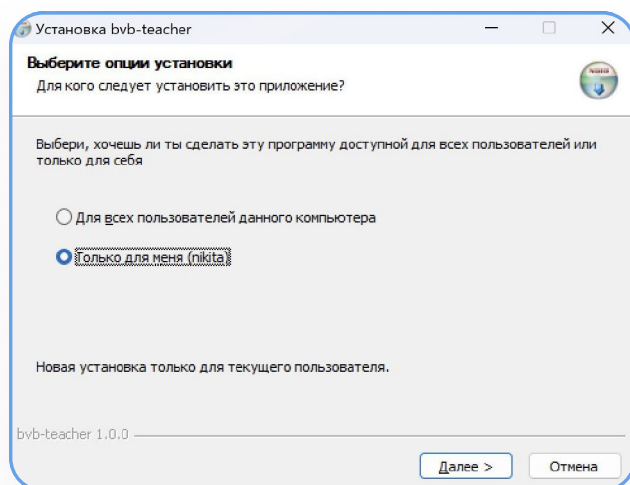


Если в классе, где будет проходить практическое профориентационное занятие, нет доступа в сеть интернет, то **рекомендуется заранее скачать файл в формате .exe**, который воспроизводит виртуальную профпробу без доступа в интернет. Файл Вы можете найти в сценарии занятия в разделе «Доступ к профпробе», или же в общем списке материалов к занятию.

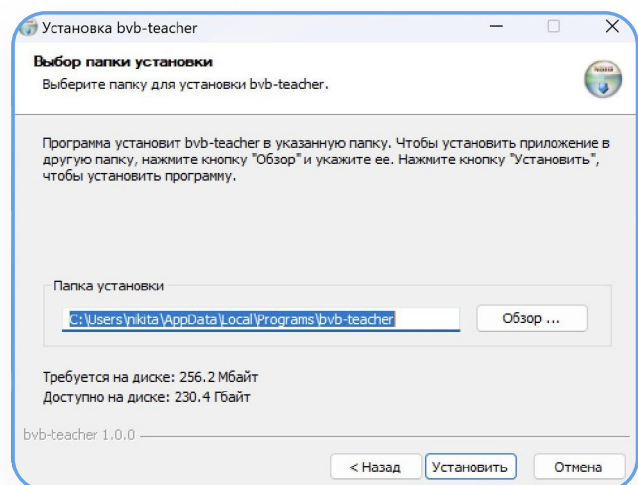


Для того, чтобы открыть приложение с профпробой, необходимо сделать несколько шагов:

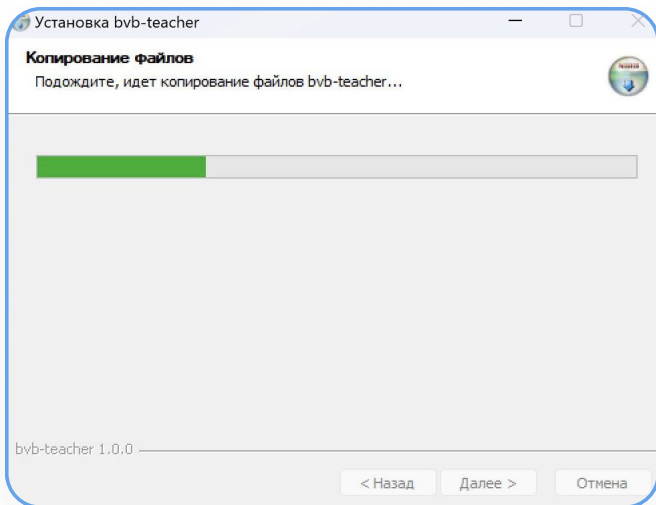
1. Скачайте файл на внешний носитель (флешку) на любом компьютере с доступом в интернет.
2. Запустите файл на компьютере в классе, в котором Вы будете проводить занятие.



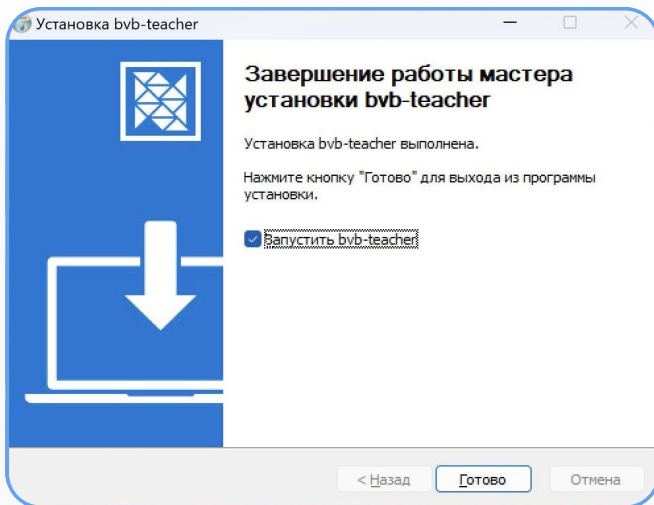
3. Выберите, кто будет иметь доступ к приложению – только Вы (если вы заходите в компьютер по отдельной учетной записи) или любой пользователь этого компьютера.



4. Выберите папку, куда сохранить приложение.

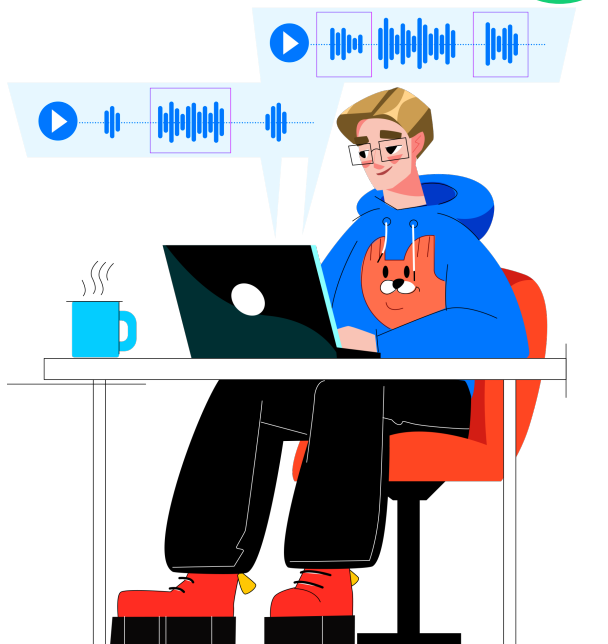


5. Подождите, пока произойдет копирование файлов.



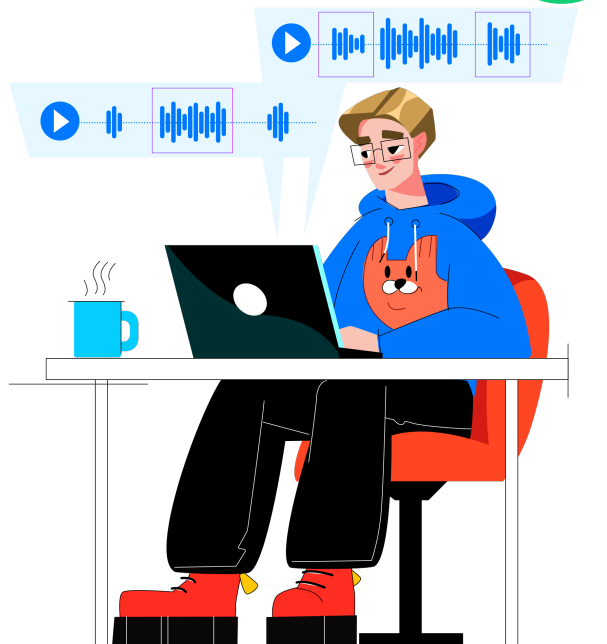
6. Приложение готово к работе.

Обратите внимание, что приложение содержит в себе только практическую часть виртуальной профпробы. Стартовую страницу с общей информацией о профессии, страницу с заданием профпробы, а также финальную страницу Вы можете скачать отдельно в материалах к занятию. Если при работе с профпробой необходимо воспользоваться Справочником – его Вы так же можете найти отдельно.



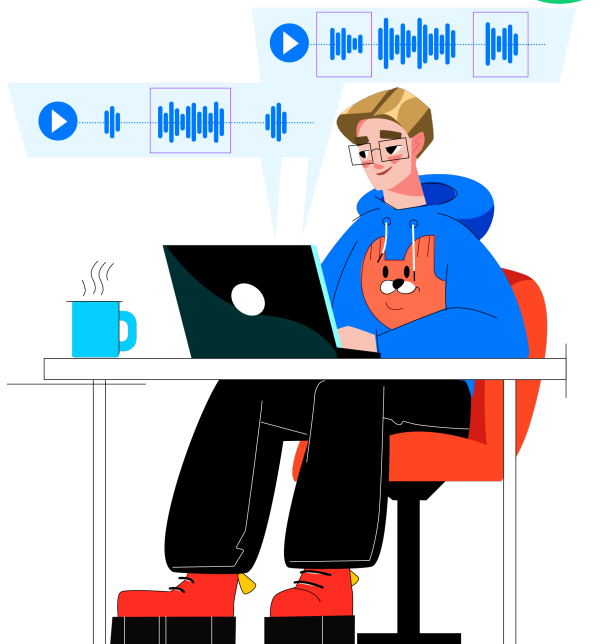
Data Science

Виртуальная профпроба



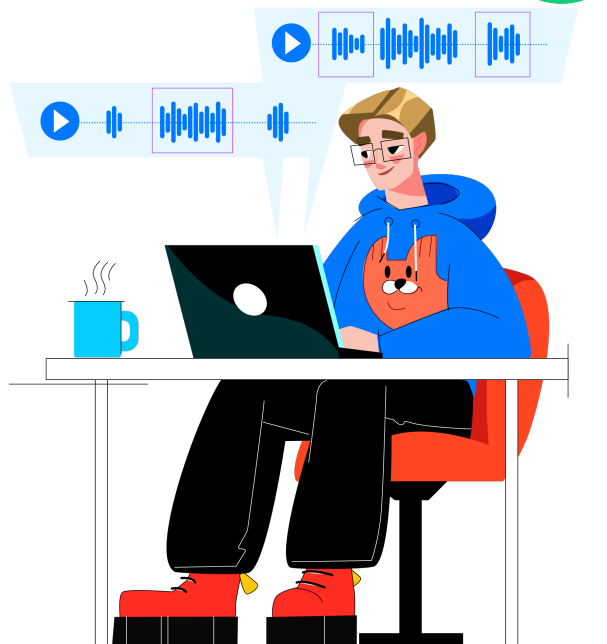
Data Science

Виртуальная профпроба



Data Science

Виртуальная профпроба



Data Science

Виртуальная профпроба

Памятка для педагога

Часть 1

- 1** Моделирующая профпроба на платформе (далее — виртуальная проба, проба) представляет собой моделирование профессиональной деятельности разных специалистов с помощью цифровых интерактивных технологий (приложение-симулятор), что позволяет сформировать представления о компетенциях и особенностях профессий, необходимых навыках для осуществления конкретной профессиональной деятельности. Виртуальные пробы доступны на платформе «Билет в будущее».
- 2** Виртуальные пробы на платформе проводятся в рамках курса внеурочной деятельности отдельным профориентационным занятием. Для данных занятий педагогу будут заранее заданы и рекомендованы к проведению пробы по определённой профессии. После того как вы проведёте эту пробу в школе, ребята смогут самостоятельно её пройти в Профигrade на платформе «Билет в будущее» (<https://profigrad.bvbinfo.ru/>), а также им будут доступны дополнительные пробы из данной тематической среды.
- 3** Все материалы для проведения данного занятия — сценарный план, видеоматериалы, раздаточные материалы, ссылка на пробу будут доступны вам в готовом виде в цифровом инструменте проекта — «Конструкторе будущего».

Мы рекомендуем вам готовиться к занятиям заранее, это позволит в комфортном для себя графике изучить материалы, подготовить раздаточные, презентационные и видеоматериалы и успешно провести занятие.

Дополнительно для подготовки к профориентационным занятиям педагогам доступны образовательные материалы от экспертов проекта «Билет в будущее» (для педагогических работников, ответственных за реализацию Профминимума, материалы доступны в «Конструкторе будущего» в разделе «Инструктаж»). Для педагогов-навигаторов проекта «Билет в будущее» материалы доступны в личном кабинете в разделе «Обучение»).

Памятка для педагога

Часть 2

4

В «Конструкторе будущего» вам будет предложено выбрать один из двух форматов проведения виртуальной пробы:

- Виртуальную пробу рекомендуется проводить в компьютерном классе с доступом всех компьютеров в Интернет. Если учеников больше, чем компьютеров, ребят можно объединять по 2-3 человека (максимум — 5). Пробу обучающиеся решают на компьютере, при этом рекомендуется, чтобы у ребят были бумага и ручка, если возникнет необходимость сделать заметки или расчёты.
- Если нет технической возможности провести данное занятие в классе с доступом ребят к компьютерам, можно провести пробу в формате демонстрации на экране/проекторе или интерактивной доске, где педагог выступает в роли модератора данного процесса согласно рекомендациям сценарного плана.

Для этого вам понадобится:

- экран, проектор, интерактивная доска или телевизор.
- компьютер или ноутбук с выходом в Интернет
- раздаточные и дополнительные материалы из сценарного плана.

Если в вашем кабинете нет технической возможности для выхода в Интернет, то для вас в рамках подготовки к занятию будут даны дополнительные инструкции по технической подготовке к проведению занятия.

Важно! Виртуальные пробы с телефонов пройти нельзя, это нужно учитывать при планировании занятия.

5

Сценарный план занятия рассчитан на 45 мин, который включает в себя: вводную часть — интерактивные форматы взаимодействия в классе — игры, обсуждения, просмотр видеоролика — всё, что направлено на развитие групповой динамики и формирование познавательного интереса у ребят. Далее идёт формат проведения виртуальной пробы. Если у вас есть технические возможности для прохождения ребятами проб самостоятельно или в мини-группах, то на этот этап рекомендуется отвести 15 минут. Если такой возможности нет, то сценарный план занятия будет построен немного иначе. Вы в роли педагога-модератора будете совместно с ребятами проходить пробу на вашем экране, параллельно давать им задания, упражнения для выполнения в мини-группах, обсуждать и дискутировать о ходе выполнения пробы. Конечно, в конце каждого занятия мы рекомендуем провести рефлексию и подведение итогов.

Памятка для педагога

Дополнительные рекомендации

- В качестве дополнительной рекомендации вы можете заранее предложить ребятам выбрать модератора — одного или двух, которые будут помогать вам проводить занятия. Для этого вам заранее надо будет пройти с ребятами пробу, разобрать её и подготовиться к занятию.
- После занятия в школе ребята могут пройти виртуальную пробу в свободное время в Профиграде (<https://profigrad.bvbinfo.ru/>), что формирует у ребёнка дополнительный интерес к профессиям и повышает его осознанность. Профиград находится в свободном доступе для всех желающих. Чтобы мотивировать ребят на самостоятельную работу с пробами, на занятии обучающиеся получают артефакт — это карточка по пройденной профессии и пароль для получения «бонуса» в виртуальном городе профессий Профиграде.



МИНИ-ИГРА

1/10

МОЯ РОССИЯ —
МОИ ГОРИЗОНТЫ

Программистом можно назвать только представителя одной конкретной специальности.

ВЕРНО



НЕВЕРНО





НЕВЕРНО 👎

За словом «программист» скрывается множество разнообразных специальностей в области цифровых технологий.



Цифровые технологии — это отрасль,
которая никак не пересекается
с другими отраслями экономики
в стране.

ВЕРНО



НЕВЕРНО





МИНИ-ИГРА

2/10

МОЯ РОССИЯ —
МОИ ГОРИЗОНТЫ

НЕВЕРНО 👎

Цифровые технологии сегодня
применяются практически в любой
отрасли.



Системы, созданные на основе искусственного интеллекта, могут предотвращать пожары, петь и следить за соблюдением правил безопасности на предприятиях.

ВЕРНО



НЕВЕРНО





МИНИ-ИГРА

3/10

МОЯ РОССИЯ —
МОИ ГОРИЗОНТЫ



ВЕРНО



Конечно, всё это — разные системы,
но у всех в основе лежит **искусственный
интеллект.**



МИНИ-ИГРА

4/10

МОЯ РОССИЯ —
МОИ ГОРИЗОНТЫ

Гуманитариям нет места в мире
цифровых технологий.

ВЕРНО



НЕВЕРНО





НЕВЕРНО 👎

Наоборот, сегодня всё больше профессий объединяют в себе гуманитарные и ИТ-направления. Например, цифровой лингвист.



МИНИ-ИГРА

5/10

МОЯ РОССИЯ —
МОИ ГОРИЗОНТЫ

Система компьютерного зрения
позволяет устройствам распознавать
объекты и выполнять нужные задачи.

ВЕРНО



НЕВЕРНО





МИНИ-ИГРА

5/10

МОЯ РОССИЯ —
МОИ ГОРИЗОНТЫ



ВЕРНО



Это одна из важнейших цифровых технологий, которая необходима, например, робототехникам.



МИНИ-ИГРА

6/10

МОЯ РОССИЯ —
МОИ ГОРИЗОНТЫ

Робофутбол создан для развлечения аудитории.

ВЕРНО



НЕВЕРНО





НЕВЕРНО 👎

Обучая роботов игре в футбол, робототехники разрабатывают сложнейшие алгоритмы, которые позволят роботам ориентироваться в пространстве не хуже человека и решать самые разные задачи.



МИНИ-ИГРА

7/10

МОЯ РОССИЯ —
МОИ ГОРИЗОНТЫ

В 38 регионах России уже используется беспилотный транспорт в экспериментальном режиме.

ВЕРНО



НЕВЕРНО





МИНИ-ИГРА

7/10

МОЯ РОССИЯ —
МОИ ГОРИЗОНТЫ



ВЕРНО



Это позволит приобрести опыт эксплуатации беспилотных автомобилей и подготовить основу для дальнейшего развития этого направления в России.



МИНИ-ИГРА

8/10

МОЯ РОССИЯ —
МОИ ГОРИЗОНТЫ

Беспилотную технику можно встретить на российских сельскохозяйственных полях.

ВЕРНО



НЕВЕРНО





МИНИ-ИГРА

8/10

МОЯ РОССИЯ —
МОИ ГОРИЗОНТЫ



ВЕРНО



В России в сельском хозяйстве используют беспилотные комбайны и тракторы. Это помогает увеличивать урожай.



МИНИ-ИГРА

9/10

МОЯ РОССИЯ —
МОИ ГОРИЗОНТЫ

Крайне редко цифровые технологии
используются в медицине.

ВЕРНО



НЕВЕРНО





НЕВЕРНО 👎

Медики, наоборот, одними из первых используют прогрессивные технологии, например, для выявления опасных заболеваний у пациентов.



МИНИ-ИГРА

10/10

МОЯ РОССИЯ —
МОИ ГОРИЗОНТЫ

Цифровые технологии не могут
распознавать эмоции людей.

ВЕРНО



НЕВЕРНО





МИНИ-ИГРА

10/10

МОЯ РОССИЯ —
МОИ ГОРИЗОНТЫ

НЕВЕРНО 👎

Специальные программы способны считывать эмоции, и их используют в российских банках, чтобы улучшить клиентский сервис.

Data Science

Специалист по Data Science — это человек, который анализирует большие объёмы данных, совершенствует нейросети, создаёт модели машинного обучения и умные алгоритмы. Благодаря этому компании улучшают качество своих продуктов и услуг, оптимизируют расходы и создают прорывные технологии.

Что же такое эти большие данные? Это могут быть тысячи чеков из огромного супермаркета, по которым можно изучить потребительский спрос, сотни рентгеновских снимков, на основе которых компьютер обнаруживает болезни, метеоданные для прогноза погоды, а ещё клиентские базы, видео, аудиофайлы, веб-аналитика и многое другое.

Задача дата-саентиста — так поработать с данными, чтобы на их основании сделать прогнозы или решить конкретную задачу, которая принесёт пользу людям.

Задачи

- 1** Сбор, подготовка и разметка подходящих для задачи данных
- 2** Сбор, подготовка и разметка подходящих для задачи данных
- 3** Анализ данных для выявления закономерностей
- 4** Создание алгоритмов, способных анализировать огромные объёмы информации
- 5** Создание моделей, делающих статистические прогнозы

Вам будет интересна эта профессия, если вы...



Давно освоили компьютер и даже пробуете себя в простых алгоритмах



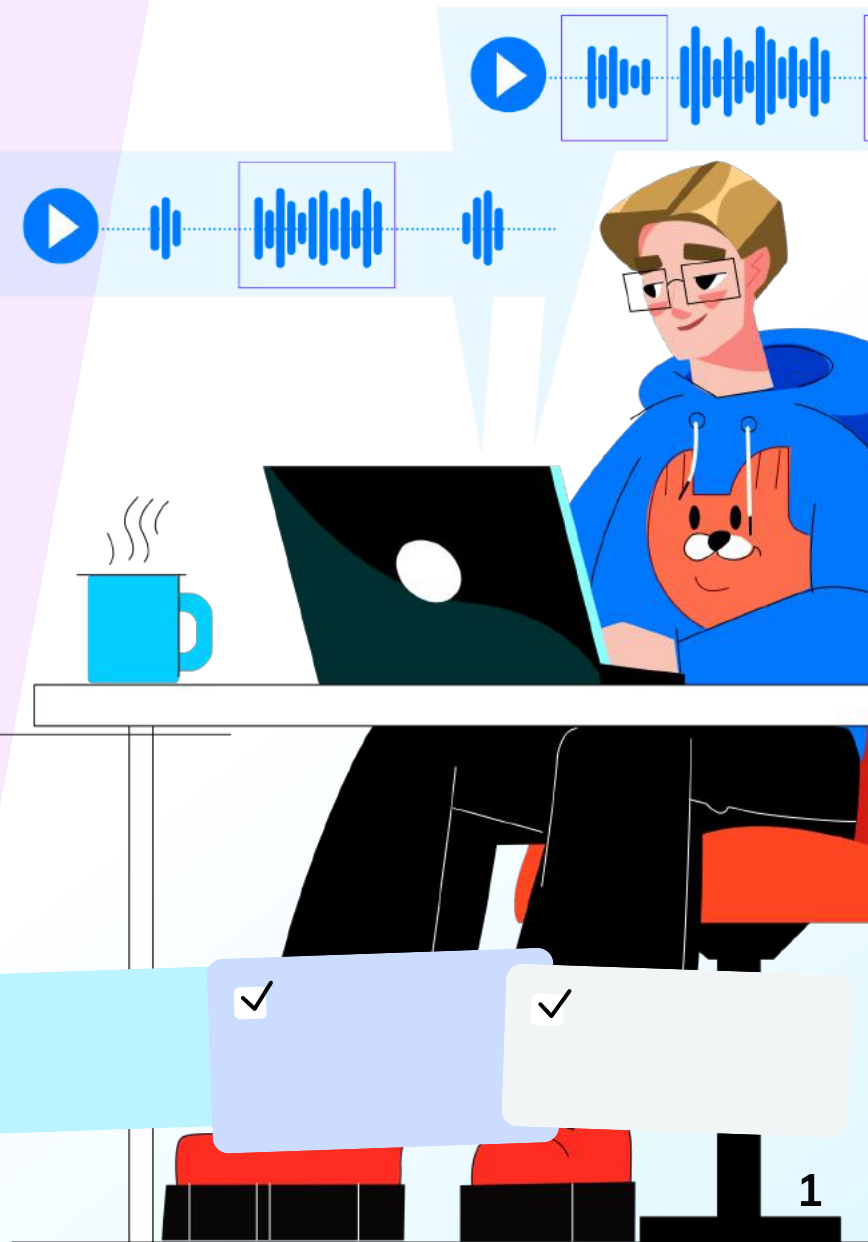
Быстро разбираетесь в математических формулах



Хотите получить современную профессию с достойной зарплатой



Усидчивы, аккуратны и внимательны



Задание

Часто для решения сложных задач или даже создания новых IT-продуктов используют искусственный интеллект. Но он появляется не сам по себе: обучением машин занимаются специалисты по Data Science. Именно такую роль вы сегодня и примерите на себя.

Ваша основная задача: разработать модель машинного обучения для перевода аудиоконтента в текст.

Вы разберётесь, откуда вообще появляются алгоритмы и нейросети, на конкретной задаче обучите модель распознавания слов и проверите, насколько эффективно она работает.

А ещё, несмотря на техническую сложность, вам может пригодиться знание русского языка — ведь надо будет указать алгоритму на его грамматические ошибки.

Вперёд, будущие айтишники!

Этапы

- 1 Определить цель машинного обучения
- 2 Собрать и подготовить данные
- 3 Обучить модель машинного обучения в несколько этапов
- 4 Проверить работу алгоритма с помощью метрик

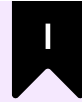
Определение задач

Специалистам по Data Science важно сразу чётко определить, какую технологическую задачу им предстоит решать.

Попробуйте определить, что из предложенного будет задачей машинного обучения, а что — нет?

Подсказка: задача машинного обучения — это тип прогноза или вывода о какой-то проблеме на основе данных.

Задачи бизнеса обычно направлены на оптимизацию процессов в компании, увеличение выручки и количества пользователей, создание новых продуктов и функциональных возможностей.



Задачи машинного обучения



Бизнес-задачи

1. Распознавать данные в аудиодорожке и переводить её в текст.
2. Создать технологию, чтобы люди в любой ситуации могли воспринимать видео- и голосовые сообщения.
3. Увеличивать ежедневное количество пользователей чата на 15%.
4. Дать возможность слабовидящим пользователям проходить «капчу» в виде сгенерированной аудиодорожки.
5. Предсказывать наличие знаков препинания в расшифрованных аудиосообщениях.
6. Поднять оценку показателя «Удобный чат» на 10%.

Часть 1

Определение задач

Специалистам по Data Science важно сразу чётко определить, какую технологическую задачу им предстоит решать.

Попробуйте определить, что из предложенного будет задачей машинного обучения, а что — нет?

Подсказка: задача машинного обучения — это тип прогноза или вывода о какой-то проблеме на основе данных.

Задачи бизнеса обычно направлены на оптимизацию процессов в компании, увеличение выручки и количества пользователей, создание новых продуктов и функциональных возможностей.



Задачи машинного обучения

- ✓ Распознавать данные в аудиодорожке и переводить её в текст
- ✓ Предсказывать наличие знаков препинания в расшифрованных аудиосообщениях



Продуктовые задачи

- ✓ Создать технологию, чтобы люди в любой ситуации могли воспринимать видео- и голосовые сообщения
- ✓ Увеличить ежедневное количество пользователей чата на 15%
- ✓ Дать возможность слабовидящим пользователям проходить «капчу» в виде сгенерированной аудиодорожки
- ✓ Поднять оценку показателя «Удобный чат» на 10%

Сбор и подготовка данных

Цель поставлена: необходимо разработать алгоритм для перевода аудиосообщений в текст. Но как его обучить?

Самое важное для машинного обучения — данные. Данных нужно много, они должны быть разнообразны и качественно отобраны.

В среднем человек за одну секунду говорит два слова. Если сообщение из пяти слов длится 20 секунд, будем считать такую дорожку некачественной, не подходящей для обучения модели.

И если на аудиодорожке много пустых мест без «всплесков», она также не подходит для обучения.

Исключите 3 записи, которые не подойдут для машинного обучения.

Запись 1



Пойдем вечером в кино. Там вышел новый классный фильм. 9 слов

Запись 2



Всем привет. Очень нужны ваши сердечки. Кхе. 6 слов

Запись 3



Спасибо большое. Очень рад, что вам понравилась моя статья. 9 слов

Запись 4



А кто в кино собрался? Я не понял. 8 слов

Запись 5



Ого-о-о! 1 слово

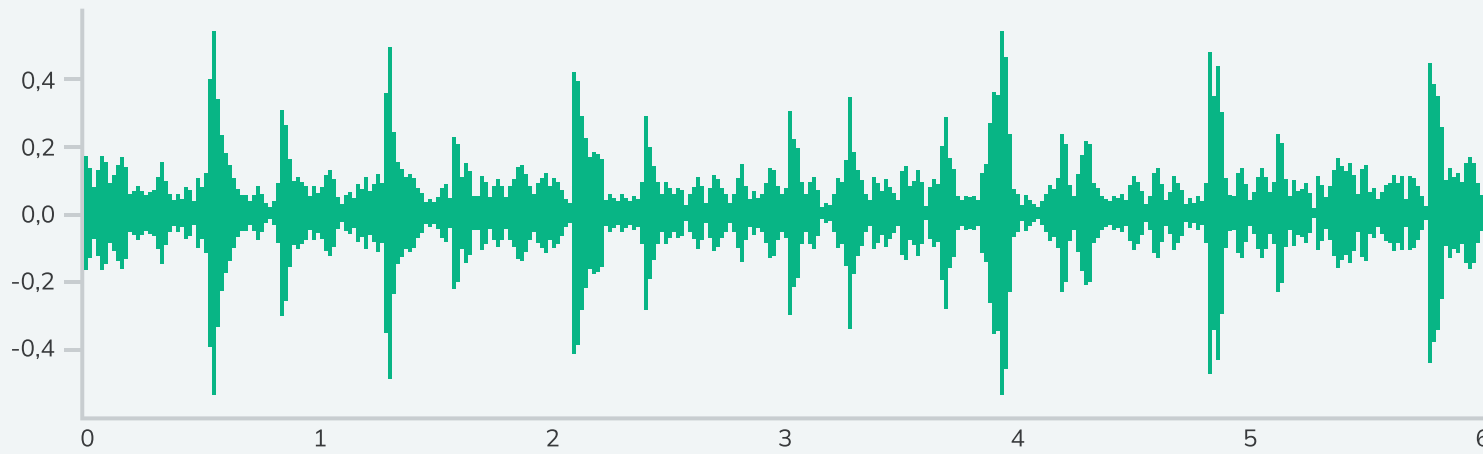
Запись 6



Да... Нет... Угу. 3 слова

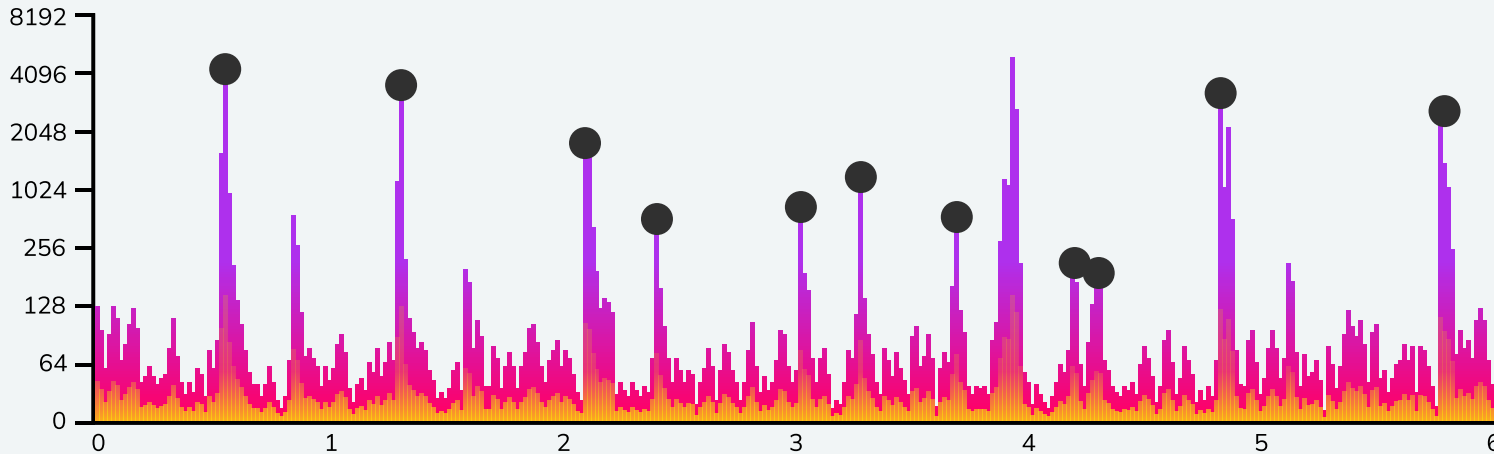
Найдите все пиковые значения на представленной записи и выделите их.

Осциллограмма



Все ли отметки пиковых частот есть на спектрограмме?
Проставьте недостающие, чтобы не потерять никаких данных.

Спектрограмма



Акустическая модель

Основная часть автоматического распознавания речи — работа с акустической моделью. Данные проходят через неё и преобразуются в символы.

Главная задача модели на этом этапе: определить наиболее вероятный звук в отдельный момент времени.

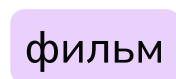
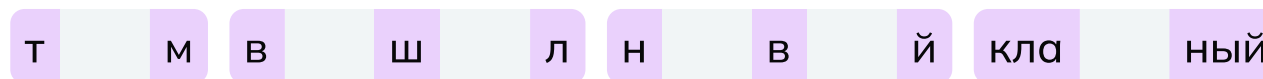
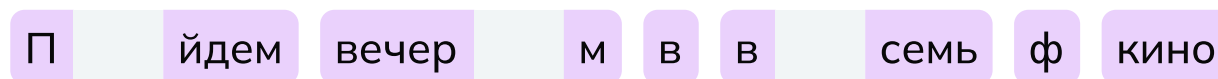
Установите, какие звуки распознал алгоритм.

Будьте внимательны: нужно отметить именно фонему, то есть «как слышится», а не «как пишется».

Расшифровка звуков:



Определите пропущенные звуки по звуковой дорожке:



Часть 3

Акустическая модель

Основная часть автоматического распознавания речи — работа с акустической моделью. Данные проходят через неё и преобразуются в символы.

Главная задача модели на этом этапе: определить наиболее вероятный звук в отдельный момент времени.

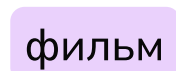
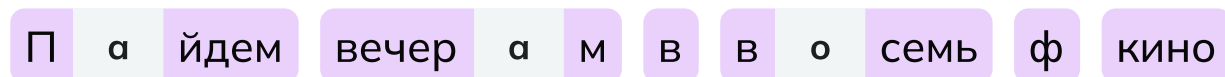
Установите, какие звуки распознал алгоритм.

Будьте внимательны: нужно отметить именно фонему, то есть «как слышится», а не «как пишется».

Расшифровка звуков:



Определите пропущенные звуки по звуковой дорожке:



Лингвистическая модель

Найдите и исправьте все ошибки в написании. Ниже напишите всю фразу правильно.

Пайдем

вечерам

в

восемь

ф

кино

там

вышэл

новый

класный

фильм

Пунктуационная модель

Поставьте галочку там, где должен быть какой-то знак препинания, включая точку.

Пример 1

Прошло жаркое лето и наступила

осень

Пример 2

Здравствуйте Иван Федорович

Пример 3

Летом ребята приносили разные цветы

лютики ромашки васильки колокольчики

Пример 4

Фу какая мерзость

Общее количество знаков 11

1. Распознать звуки в словах и написать их в пропусках для акустической модели.
2. Прописать уже верные буквы по правилам правописания для лингвистической модели.
3. Расставить знаки препинания в те ячейки, где они необходимы.

Пример 1

Акустическая модель

Привет п идем на пл ща ку п играем в футбол

Лингвистическая и пунктуационная модели:

Привет п идем на пл ща ку п играем в футбол

Пример 2

Акустическая модель

Ф нетика это раздел науки о зыке к торый изучает
 звуки речи

Лингвистическая и пунктуационная модели:

Ф нетика это раздел науки о зыке к торый изучает
 звуки речи

Пунктуационная модель

Обведите или заштрихуйте паузы в аудиодорожке.

Дорожка



Классно быть дата саентистом! Что думаешь?

Оценка результата

Разделите фразу на минимальные цепочки, чтобы алгоритму было проще расшифровать их. Обведите или заштрихуйте необходимые участки.



0:05

Прости

что

долго

не

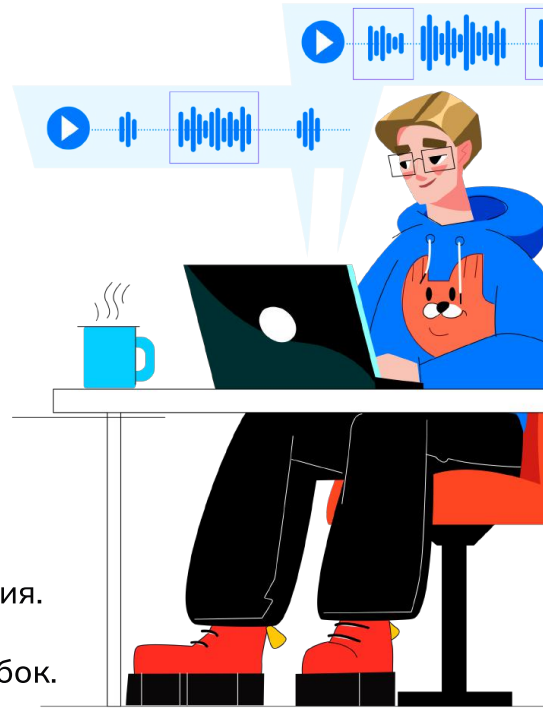
отвечала

Отлично, задание выполнено!

Поздравляем! Теперь любое голосовое сообщение или видео можно превратить в разборчивый текст.

Сегодня вы:

- ✓ Познакомились с одной из задач специалиста Data Science.
- ✓ Отличили частоты на осциллограмме от частот на спектрограмме.
- ✓ Поработали с несколькими моделями машинного обучения.
- ✓ Обучили алгоритм и убедились, что он работает без ошибок.



Мне понравилось! Что дальше? Советы от эксперта

1

Участвуйте в хакатонах и чемпионатах по программированию! Это даст колоссальный опыт и знакомства.

2

Развивайте насмотренность. То есть наблюдайте, как другие решают какую-то задачу, какую технологию и где применяют. Это очень важный навык.

3

Если вы чувствуете, что хотите в ИТ, но пока не определились, начните делать свой проект. Пробуйте разные роли: дизайнера, разработчика, менеджера и пр.

4

В интернете много лекций и инструкций, много уже написанных программ! Так вы разберётесь во многих нюансах!



Иван Самсонов

Data Science Lead, VK

Data Science

Специалист по Data Science — это человек, который анализирует большие объёмы данных, совершенствует нейросети, создаёт модели машинного обучения и умные алгоритмы. Благодаря этому компании улучшают качество своих продуктов и услуг, оптимизируют расходы и создают прорывные технологии.

Что же такое эти большие данные? Это могут быть тысячи чеков из огромного супермаркета, по которым можно изучить потребительский спрос, сотни рентгеновских снимков, на основе которых компьютер обнаруживает болезни, метеоданные для прогноза погоды, а ещё клиентские базы, видео, аудиофайлы, веб-аналитика и многое другое.

Задача дата-саентиста — так поработать с данными, чтобы на их основании сделать прогнозы или решить конкретную задачу, которая принесёт пользу людям.

Задачи

- 1** Сбор, подготовка и разметка подходящих для задачи данных
- 2** Сбор, подготовка и разметка подходящих для задачи данных
- 3** Анализ данных для выявления закономерностей
- 4** Создание алгоритмов, способных анализировать огромные объёмы информации
- 5** Создание моделей, делающих статистические прогнозы

Вам будет интересна эта профессия, если вы...



Давно освоили компьютер и даже пробуете себя в простых алгоритмах



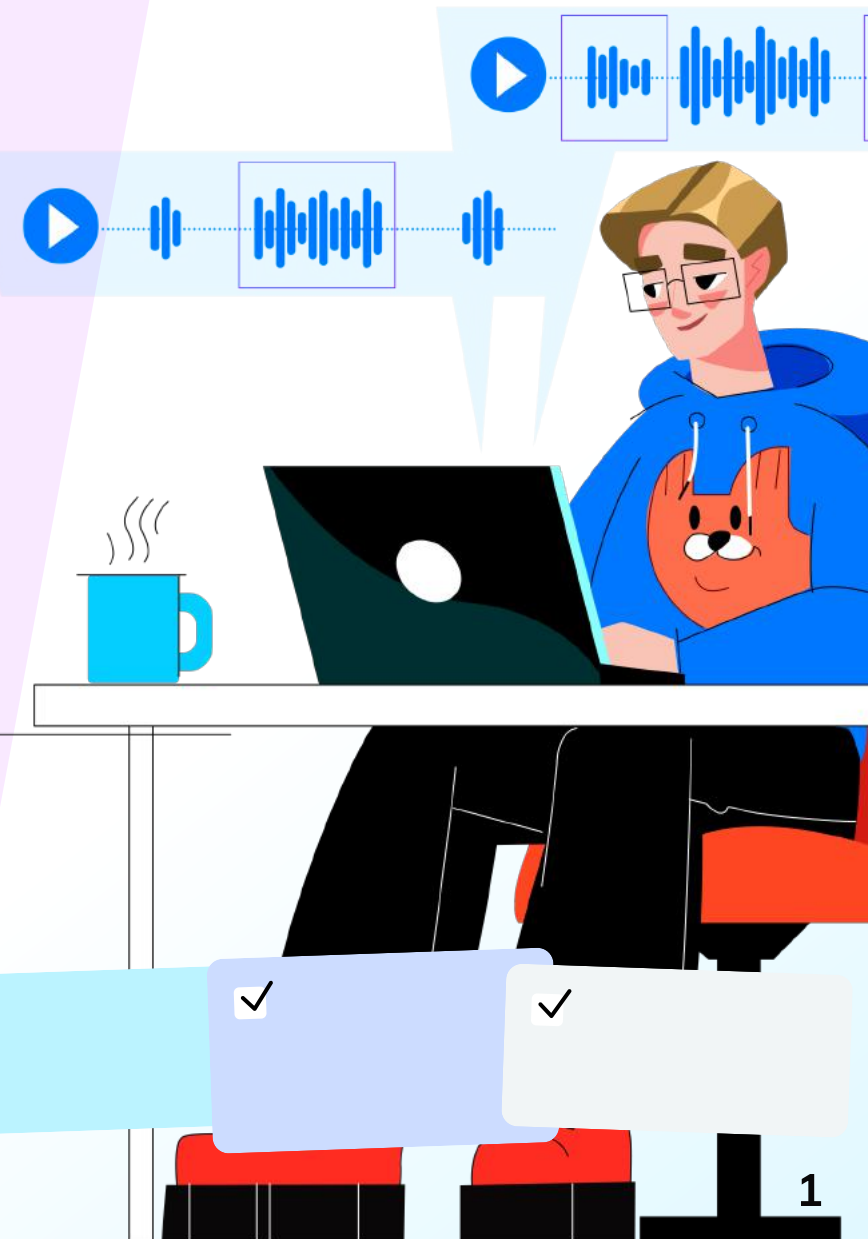
Быстро разбираетесь в математических формулах



Хотите получить современную профессию с достойной зарплатой



Усидчивы, аккуратны и внимательны



Задание

Часто для решения сложных задач или даже создания новых IT-продуктов используют искусственный интеллект. Но он появляется не сам по себе: обучением машин занимаются специалисты по Data Science. Именно такую роль вы сегодня и примерите на себя.

Ваша основная задача: разработать модель машинного обучения для перевода аудиоконтента в текст.

Вы разберётесь, откуда вообще появляются алгоритмы и нейросети, на конкретной задаче обучите модель распознавания слов и проверите, насколько эффективно она работает.

А ещё, несмотря на техническую сложность, вам может пригодиться знание русского языка — ведь надо будет указать алгоритму на его грамматические ошибки.

Вперёд, будущие айтишники!

Этапы

- 1 Определить цель машинного обучения
- 2 Собрать и подготовить данные
- 3 Обучить модель машинного обучения в несколько этапов
- 4 Проверить работу алгоритма с помощью метрик

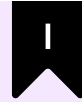
Определение задач

Специалистам по Data Science важно сразу чётко определить, какую технологическую задачу им предстоит решать.

Попробуйте определить, что из предложенного будет задачей машинного обучения, а что — нет?

Подсказка: задача машинного обучения — это тип прогноза или вывода о какой-то проблеме на основе данных.

Задачи бизнеса обычно направлены на оптимизацию процессов в компании, увеличение выручки и количества пользователей, создание новых продуктов и функциональных возможностей.



Задачи машинного обучения



Бизнес-задачи

1. Распознавать данные в аудиодорожке и переводить её в текст.
2. Создать технологию, чтобы люди в любой ситуации могли воспринимать видео- и голосовые сообщения.
3. Увеличивать ежедневное количество пользователей чата на 15%.
4. Дать возможность слабовидящим пользователям проходить «капчу» в виде сгенерированной аудиодорожки.
5. Предсказывать наличие знаков препинания в расшифрованных аудиосообщениях.
6. Поднять оценку показателя «Удобный чат» на 10%.

Часть 1

Определение задач

Специалистам по Data Science важно сразу чётко определить, какую технологическую задачу им предстоит решать.

Попробуйте определить, что из предложенного будет задачей машинного обучения, а что — нет?

Подсказка: задача машинного обучения — это тип прогноза или вывода о какой-то проблеме на основе данных.

Задачи бизнеса обычно направлены на оптимизацию процессов в компании, увеличение выручки и количества пользователей, создание новых продуктов и функциональных возможностей.



Задачи машинного обучения

- ✓ Распознавать данные в аудиодорожке и переводить её в текст
- ✓ Предсказывать наличие знаков препинания в расшифрованных аудиосообщениях



Продуктовые задачи

- ✓ Создать технологию, чтобы люди в любой ситуации могли воспринимать видео- и голосовые сообщения
- ✓ Увеличить ежедневное количество пользователей чата на 15%
- ✓ Дать возможность слабовидящим пользователям проходить «капчу» в виде сгенерированной аудиодорожки
- ✓ Поднять оценку показателя «Удобный чат» на 10%

Сбор и подготовка данных

Цель поставлена: необходимо разработать алгоритм для перевода аудиосообщений в текст. Но как его обучить?

Самое важное для машинного обучения — данные. Данных нужно много, они должны быть разнообразны и качественно отобраны.

В среднем человек за одну секунду говорит два слова. Если сообщение из пяти слов длится 20 секунд, будем считать такую дорожку некачественной, не подходящей для обучения модели.

И если на аудиодорожке много пустых мест без «всплесков», она также не подходит для обучения.

Исключите 3 записи, которые не подойдут для машинного обучения.

Запись 1



Пойдем вечером в кино. Там вышел новый классный фильм. 9 слов

Запись 2



Всем привет. Очень нужны ваши сердечки. Кхе. 6 слов

Запись 3



Спасибо большое. Очень рад, что вам понравилась моя статья. 9 слов

Запись 4



А кто в кино собрался? Я не понял. 8 слов

Запись 5



Ого-о-о! 1 слово

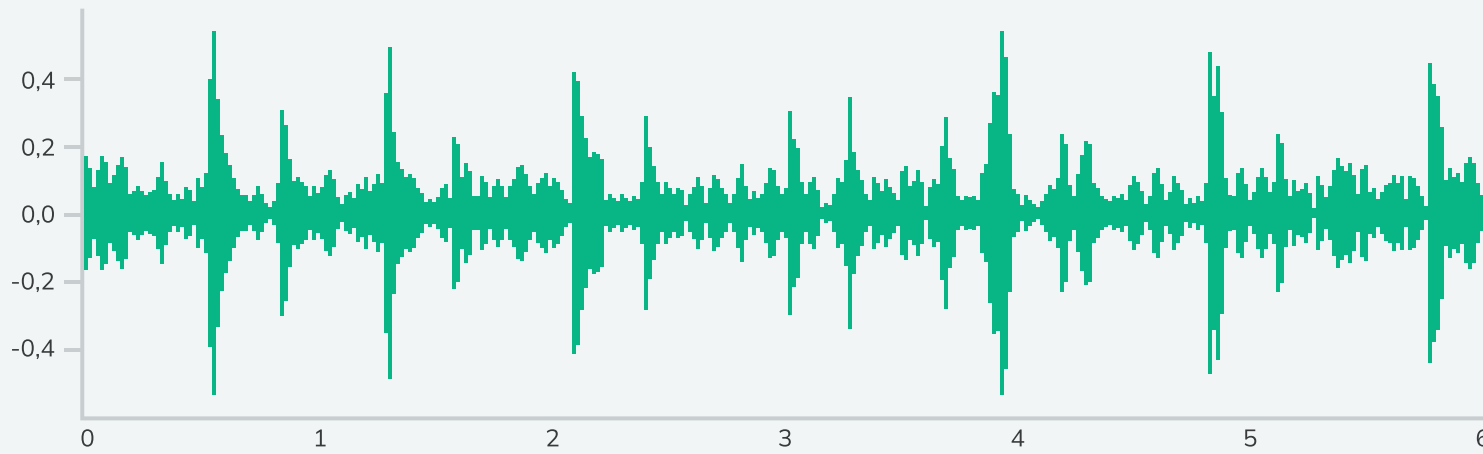
Запись 6



Да... Нет... Угу. 3 слова

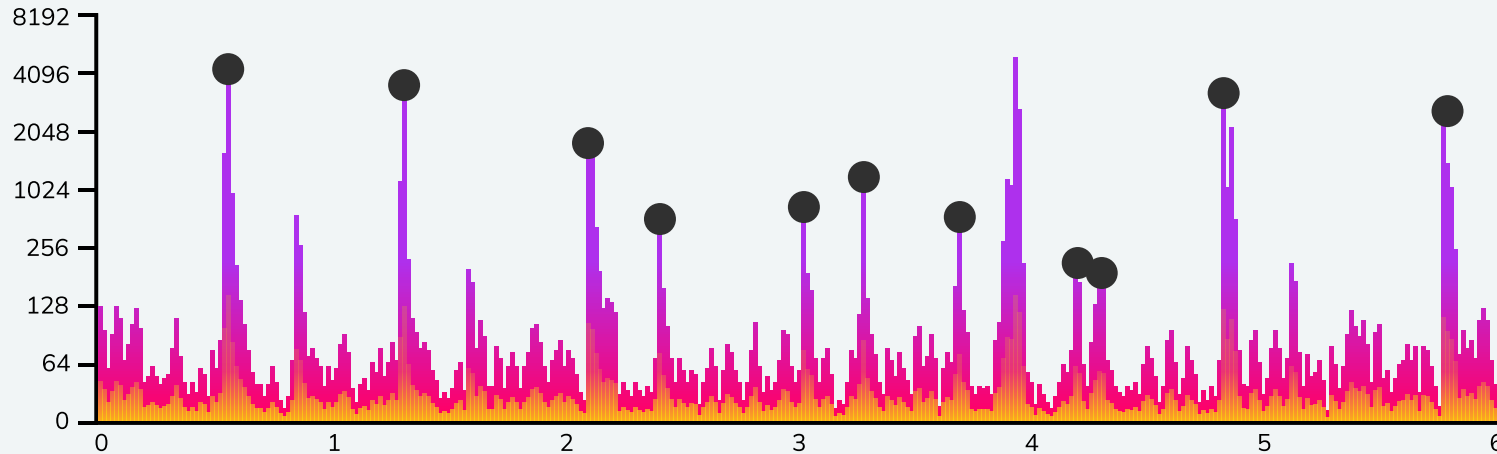
Найдите все пиковые значения на представленной записи и выделите их.

Осциллограмма



Все ли отметки пиковых частот есть на спектрограмме?
Проставьте недостающие, чтобы не потерять никаких данных.

Спектрограмма



Акустическая модель

Основная часть автоматического распознавания речи — работа с акустической моделью. Данные проходят через неё и преобразуются в символы.

Главная задача модели на этом этапе: определить наиболее вероятный звук в отдельный момент времени.

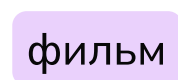
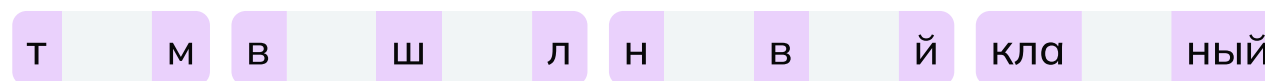
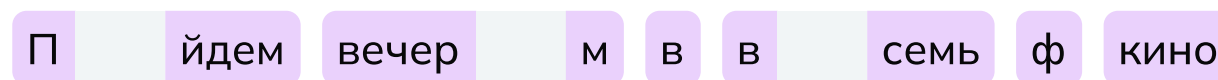
Установите, какие звуки распознал алгоритм.

Будьте внимательны: нужно отметить именно фонему, то есть «как слышится», а не «как пишется».

Расшифровка звуков:



Определите пропущенные звуки по звуковой дорожке:



Часть 3

Акустическая модель

Основная часть автоматического распознавания речи — работа с акустической моделью. Данные проходят через неё и преобразуются в символы.

Главная задача модели на этом этапе: определить наиболее вероятный звук в отдельный момент времени.

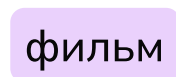
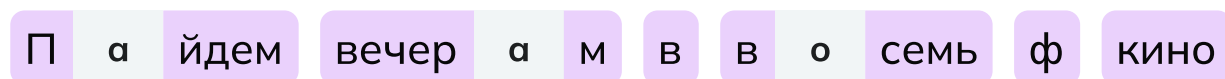
Установите, какие звуки распознал алгоритм.

Будьте внимательны: нужно отметить именно фонему, то есть «как слышится», а не «как пишется».

Расшифровка звуков:



Определите пропущенные звуки по звуковой дорожке:



Лингвистическая модель

Найдите и исправьте все ошибки в написании. Ниже напишите всю фразу правильно.

Пайдем

вечерам

в

восемь

ф

кино

там

вышэл

новый

класный

фильм

Пунктуационная модель

Поставьте галочку там, где должен быть какой-то знак препинания, включая точку.

Пример 1

Прошло жаркое лето и наступила

осень

Пример 2

Здравствуйте Иван Федорович

Пример 3

Летом ребята приносили разные цветы

лютики ромашки васильки колокольчики

Пример 4

Фу какая мерзость

Общее количество знаков 11

1. Распознать звуки в словах и написать их в пропусках для акустической модели.
2. Прописать уже верные буквы по правилам правописания для лингвистической модели.
3. Расставить знаки препинания в те ячейки, где они необходимы.

Пример 1Привет п_йдём на пл_ща_ку п_играем в футбол **Акустическая модель**Привет п а йдем на пл а ща т ку п а играем в футбол **Лингвистическая и пунктуационная модели:**Привет , п о йдем на пл о ща д ку п о играем в футбол .**Пример 2**Ф_нетика это раздел науки о _зыке к_торый изучает звуки речи **Акустическая модель**Ф а нетика это раздел науки о и зыке к а торый изучает
 звуки речи **Лингвистическая и пунктуационная модели:**Ф о нетика — это раздел науки о я зыке , к о торый изучает
 звуки речи .

Пример 3

Как _сегда мой с_се_ пр_спал и _п_здал е_о не пустили в кла_

Акустическая модель

Как ф сегда мой с а се т пр а спал и а п а здал
е в о не пустили в кла с

Лингвистическая и пунктуационная модели:

Как в сегда , мой с о се д пр о спал и о п о здал ,
е г о не пустили в кла сс .

Пример 4

Где_ зах_ди к нам на обе_ папа будет ра_ встрече

Акустическая модель

Где п зах а ди к нам на обе т папа будет
ра т встрече

Лингвистическая и пунктуационная модели:

Где б , зах о ди к нам на обе д , папа будет
ра д встрече .

Пример 5

Он х_р_шо знает ру_кий _зык и это его пр_имущество

Акустическая модель

Он х а р а шо знает ру с кий и зык и это его пр и имущество

Лингвистическая и пунктуационная модели:

Он х о р о шо знает ру сс кий я зык , и это его пр е имущество .

Пунктуационная модель

Обведите или заштрихуйте паузы в аудиодорожке.

Дорожка



Классно быть дата саентистом! Что думаешь?

Оценка результата

Разделите фразу на минимальные цепочки, чтобы алгоритму было проще расшифровать их. Обведите или заштрихуйте необходимые участки.



0:05

Прости

что

долго

не

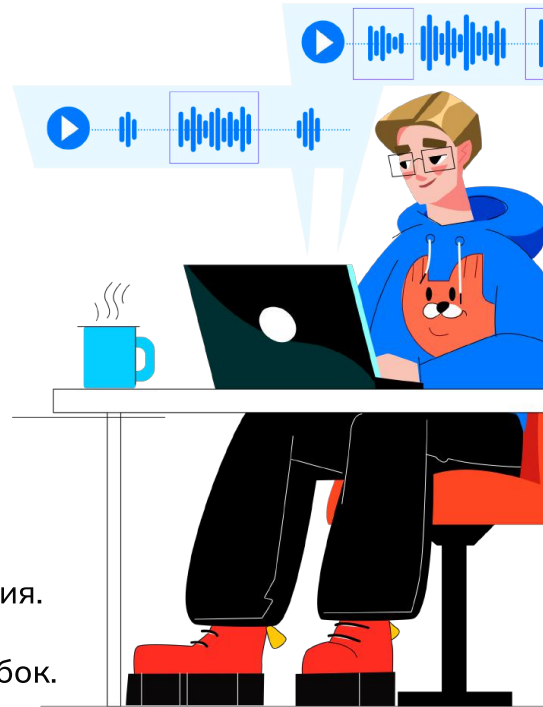
отвечала

Отлично, задание выполнено!

Поздравляем! Теперь любое голосовое сообщение или видео можно превратить в разборчивый текст.

Сегодня вы:

- ✓ Познакомились с одной из задач специалиста Data Science.
- ✓ Отличили частоты на осциллограмме от частот на спектрограмме.
- ✓ Поработали с несколькими моделями машинного обучения.
- ✓ Обучили алгоритм и убедились, что он работает без ошибок.



Мне понравилось! Что дальше? Советы от эксперта

1

Участвуйте в хакатонах и чемпионатах по программированию! Это даст колоссальный опыт и знакомства.

2

Развивайте насмотренность. То есть наблюдайте, как другие решают какую-то задачу, какую технологию и где применяют. Это очень важный навык.

3

Если вы чувствуете, что хотите в ИТ, но пока не определились, начните делать свой проект. Пробуйте разные роли: дизайнера, разработчика, менеджера и пр.

4

В интернете много лекций и инструкций, много уже написанных программ! Так вы разберётесь во многих нюансах!



Иван Самсонов

Data Science Lead, VK

Часто для решения сложных задач или даже создания новых IT-продуктов используют искусственный интеллект. Но он появляется не сам по себе: обучением машин занимаются специалисты по Data Science. Именно такую роль вы сегодня и примерите на себя. Ваша основная задача: разработать модель машинного обучения для перевода аудиоконтента в текст.

Вы разберётесь, откуда вообще появляются алгоритмы и нейросети, на конкретной задаче обучите модель распознавания слов и проверите, насколько эффективно она работает. А ещё, несмотря на техническую сложность, вам может пригодиться знание русского языка — ведь надо будет указать алгоритму на его грамматические ошибки.

Часть 1. Определение задач

Специалистам по Data Science важно сразу чётко определить, какую технологическую задачу им предстоит решать.

Это задание весьма непростое, хоть и не кажется таким на первый взгляд. Всё дело в том, что вам нужно **отличить задачи машинного обучения**, то есть буквально то, над чем будет работать алгоритм, **от бизнес-задач**, то есть задач всей компании, её отдельных департаментов или продуктов.

Обратите внимание: продуктовые задачи, или задачи бизнеса, обычно направлены на оптимизацию процессов в компании, увеличение выручки и количества пользователей, создание новых продуктов и функциональных возможностей. Зачастую они даже имеют какой-то численный показатель, при котором задача считается успешно выполненной.

А вот задача машинного обучения всегда связана с **обработкой данных**, в результате которой получается прогноз или вывод о какой-то проблеме. Например, если вы генерируете изображение по текстовому запросу, то в результате получаете картинку, которая, по мнению алгоритма, лучше всего его отражает. Вопрос лишь в том, насколько точным оказалось «попадание» в этот запрос.

Как пройти это задание?

Поднять оценку «Удобный чат» от пользователей на 10% — это точно не задача машинного обучения, это так называемая продуктовая задача, которую можно решить разными способами. Новый функционал на основе машинного обучения может стать одним из инструментов для её достижения.

Зато чтобы пользователям было удобнее общаться, можно внедрить такой алгоритм, который люди бы оценили по достоинству. Задача такого алгоритма может звучать примерно так: распознавать данные в аудиодорожке и переводить её в текст. Именно такую задачу нужно решить специалисту по Data Science. И когда этот функционал позволит расшифровывать голосовые сообщения, чат точно станет удобнее!

Попробуйте распределить варианты по своим группам самостоятельно, но если вдруг возникают сложности, можете подсматривать в ответы ниже.

Задачи машинного обучения:

1. Распознавать данные в аудиодорожке и переводить её в текст
2. Предсказывать наличие знаков препинания в расшифрованных аудиосообщениях

Продуктовые задачи:

1. Увеличить ежедневное количество пользователей чата на 15%
2. Чтобы люди в любой ситуации могли воспринимать видео- и голосовые сообщения
3. Поднять оценку показателя «Удобный чат» на 10%
4. Добавить возможность расшифровки «капчи*» для слабовидящих

*Капча, или captcha — это специальный защитный код, по которому программа определяет, реальный ли вы человек или бот. Вы наверняка видели такую при входе в некий сервис: нужно ввести цифры на экране, определить объекты на картинке или ответить на вопрос. **1**

Часть 2. Сбор и подготовка данных

Самое важное для машинного обучения — данные. Данных нужно много, они должны быть разнообразны и качественно отобраны.

Если мы работаем над алгоритмом, который будет распознавать текст в аудио, то и данные для обучения нужны соответствующие. Чем больше у вас будет дорожек с записанной речью, тем лучше.

Как пройти это задание?

Посмотрите на предложенные варианты. Обратите внимание: они не воспроизводятся по нажатию на кнопку Play, поэтому под каждой звуковой дорожкой есть текст с расшифровкой.

Визуально оцените плотность данных на каждой дорожке, сравните её длительность и количество слов, а затем сделайте вывод: какие из аудиофайлов НЕ подходят для обучения модели.

Сравнили? Оценили? Нашли некачественные записи? Молодцы!

А если нет, то нажмите «Исключить» напротив записей №2, №5 и №6.

Часть 3. Препроцессинг

Сам по себе звук представляется в памяти компьютера как массив значений, показывающих колебания амплитуды по времени. Этот параметр обычно исчисляется в десятках тысяч точек в секунду (или кГц). Звучит сложно, но суть в том, что такая дорожка получается очень длинной и неудобной для работы.

Поэтому перед подачей в алгоритм звук предобрабатывают: в частности, переводят в спектрограмму — зависимость интенсивности колебания звука на конкретной частоте по времени. Ух, звучит ещё сложнее, да? Если коротко: компьютеру намного проще работать со значениями частот.

Не переживайте, специалистам по Data Science в этой задаче помогает специальное программное обеспечение. А чтобы вам стало немного понятнее, мы предлагаем пройти интерактивное задание.

Как его пройти?

Вам нужно перевести звук из вида осциллограммы в спектрограмму. Нажимайте на все пиковые значения на «зелёном» графике, чтобы они отразились на спектрограмме.

После этого надо убедиться, что точно все значения сохранились, и вновь проставить недостающие точки.

Часть 3. Акустическая модель

После того как сырая дорожка переведена в удобное для нейросети представление, мы готовы распознавать речь: получать из звука распределение вероятностей букв по времени.

Главная задача модели на этом этапе: определить наиболее вероятный звук в отдельный момент времени. То есть модель «смотрит» на спектрограмму, видит там набор звуков и пытается спрогнозировать, где какая фонема расположена.

Как пройти это задание?

Чаще всего такие модели работают по принципу «что слышится, то и пишется». Поэтому вам нужно подобрать в каждом слове не букву, соответствующую словарному написанию, а звук — как если бы вы делали транскрипцию.

Вы можете посмотреть на визуальное обозначение фонемы и подобрать ей соответствующую пару из предложенных сверху вариантов, а затем выбрать его в выпадающем списке. **2**

Часть 3. Лингвистическая модель

Ура! Вы составили фонетические транскрипции, а теперь с помощью отдельной языковой модели «причешите» результат — пора исправить грамматические и орфографические ошибки в расшифровках.

Так как модель предсказывает, какой набор символов вероятнее всего встретить в языке, её обучают на классических текстах, словарях и статьях, сленге, разговорных сообщениях и неологизмах.

Как его пройти?

Сейчас вы на месте алгоритма: нужно исправить все ошибки в словах! Выбирайте подходящие варианты и двигайтесь дальше.

Это, конечно же, не экзамен по русскому языку, поэтому вы можете ошибаться, и ничего вам за это не будет.

Варианты ответов не очень запутанные, но если вдруг не удаётся распознать ошибку, то вот правильный набор слов:

Пойдем вечером в восемь в кино там вышел новый классный фильм

Часть 3. Пунктуационная модель

После того, как мы получили читаемую последовательность слов, можем расставить знаки препинания. Это особенно полезно, когда текст длинный: с точками и запятыми читать его точно будет проще.

Подход в генерации данных для обучения тут похожий: мы берём тексты с расставленными знаками препинания и искусственно «портим» эти данные, убирая пунктуацию. А затем учимся восстанавливать.

Как пройти это задание?

Попробуйте самостоятельно отметить места, в которых должны стоять знаки препинания. Не волнуйтесь, здесь тоже можно ошибаться. Профпроба — это не контрольная работа.

Чтобы не было совсем просто, даём такую подсказку:

- В примере №1 пропущены два знака.
- В примере №2 — два знака.
- В примере №3 — пять знаков.
- В примере №4 — два знака.

Отлично! Теперь нужно отметить на дорожке места, в которых вероятнее всего должны стоять знаки препинания. Ориентируйтесь на паузы между словами и нажимайте на кликабельные зоны звуковой дорожки.

Часть 4. Оценка результата с помощью метрик

Метрика — это какой-то рассчитываемый показатель, который используется для оценки эффективности. Метрики помогают понять, насколько хорошо работает модель, и сравнить её с другими моделями. В нашем случае используется метрика word error rate — чем меньше ошибок делает алгоритм, тем выше её значение.

Представьте аналогию: ваша успеваемость тоже может быть представлена в виде метрики. Допустим, у вас 10 предметов, и за четверть вы получили все пятерки. Получается, что ваш средний балл — 5,0.

А что если сюда затесались тройка по физкультуре и четвёрка по литературе? Тогда средний балл будет уже 4,7.

Примерно так работают метрики — они позволяют оценить самые разные показатели.

Как пройти это задание?

Сейчас на аудиодорожке можно заметить пять отдельных кусочков с какими-то словами. Нажмите поочерёдно на каждый из них, чтобы расшифровать.

Спойлер! Ваша модель работает так круто, что метрика WER равна 100%. В жизни, конечно, так почти не бывает — даже самый крутой алгоритм иногда ошибается. А уж на стадии обучения ошибки — обычное дело, и специалист по Data Science ищет способ их решения и улучшения модели.

Специалист по Data Science

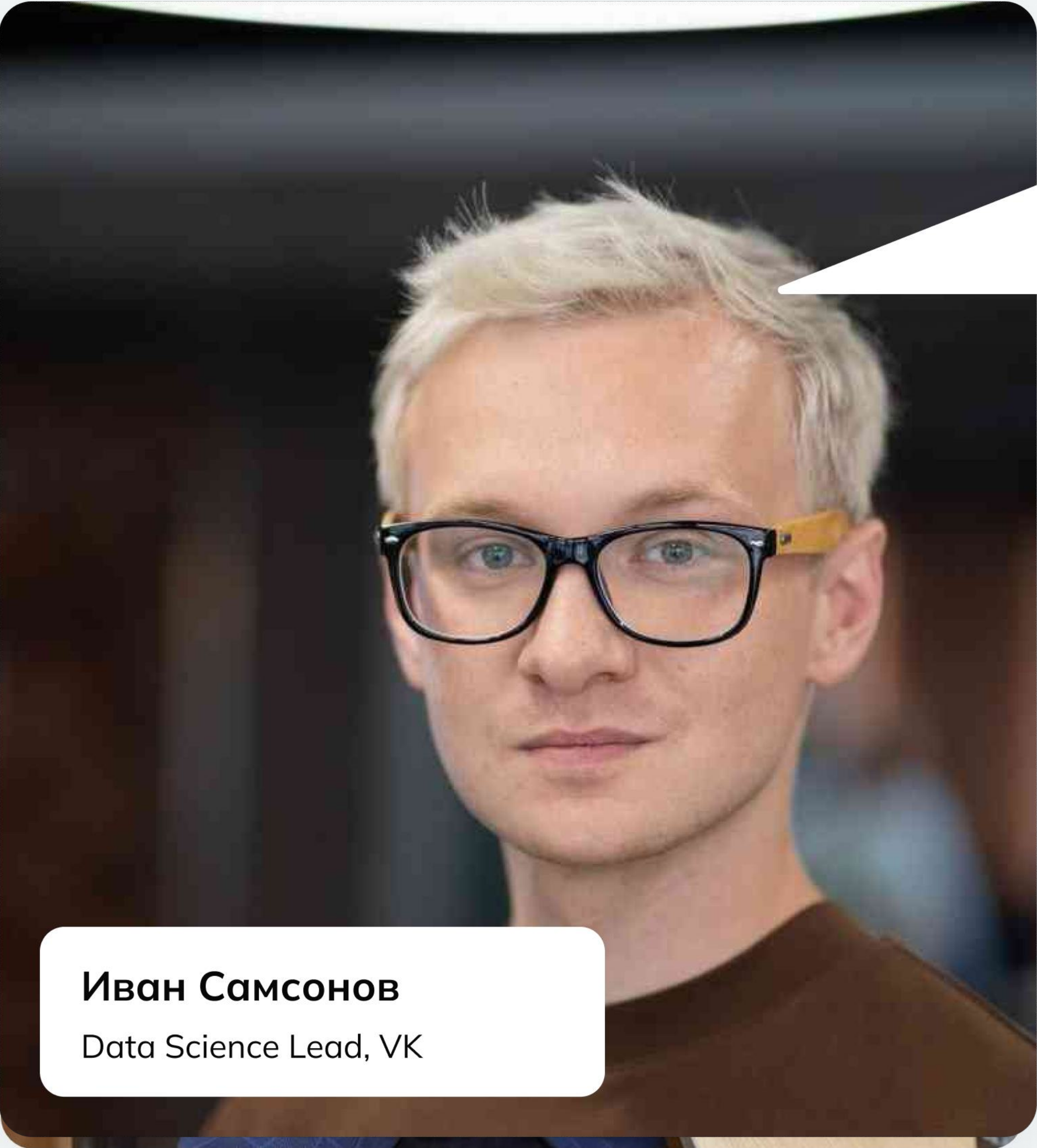
Специалист по Data Science — человек, который анализирует большие объёмы данных, обучает модели и создаёт умные алгоритмы. Благодаря этому компании улучшают качество своих продуктов и услуг, оптимизируют расходы и создают прорывные технологии.

Что же такое эти большие данные? Например, это могут быть сотни тысяч чеков из огромного супермаркета, клиентские базы, видео и аудиофайлы или веб-аналитика, которая подробно рассказывает, откуда на страничку заглянул тот или иной посетитель.

Задача дата-саентиста — так поработать с большими данными, чтобы на их основании сделать прогнозы или решить конкретную задачу.

[К заданию →](#)





Иван Самсонов

Data Science Lead, VK

Напутствие от эксперта

Мы часто слышим в интернете и по телевизору, что к нам вот-вот придёт пугающий искусственный интеллект. Но на самом деле его... не существует!

Зато существует машинное обучение. Это такое сложное преобразование информации на основе статистики и разных данных. И за ним стоят обычные ребята вроде меня и вроде вас. Не думайте, что это какие-то волшебники!

Если вам интересна эта область, обязательно пробуйте свои силы и идите к цели. Не бойтесь ошибаться. В конце концов, вы всегда сможете сменить вектор и стать разработчиком, аналитиком, продакт-менеджером.

Конечно, кроме знаний в инженерной области ещё важны «софт-скиллы» и умение работать в команде. Ведь именно команды создают классные продукты и двигают мир вперёд. Так что желаю вам успехов и найти единомышленников!

Задачи

1

Сбор, подготовка и разметка подходящих для задачи данных

2

Анализ данных для выявления закономерностей

3

Создание алгоритмов, способных анализировать огромные объёмы информации

4

Создание моделей, делающих статистические прогнозы

Вам будет интересна эта профессия, если вы...



Давно освоили компьютер и даже пробуете себя в простых алгоритмах



Быстро разбираетесь в математических формулах



Хотите современную профессию с достойной зарплатой



Усидчивы, аккуратны и внимательны



Увлекаетесь искусственным интеллектом и хотите разобраться, как он работает

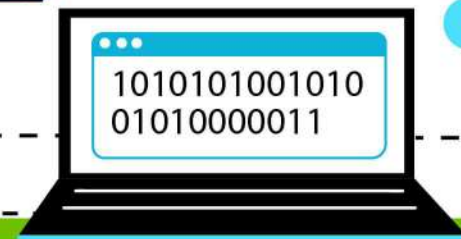
Вперёд к мечте!

Попробуйте себя в роли настоящих специалистов

[К заданию →](#)

Чтобы тут работать, нужно...

Цифровые технологии это...



РОССИЯ ЦИФРОВАЯ

Здесь я могу стать...

ВПИШИТЕ КОДОВОЕ СЛОВО

