

Test 1

Question 1: Choose one answer

Что является основной целью Stakeholder'a ML team?

- ☐ Максимизировать прибыль
- ☐ Продать как можно больше
- ☒ Высокая точность модели
- ☐ Вывести продукт как можно быстрее

Question 2: Choose one answer

Почему важна скорость работы ML-моделей в индустрии?

- ☒ Потому что лишняя доля секунды может привести к отказам от покупки
- ☐ Потому что это влияет на точность модели
- ☐ Потому что это влияет на стоимость прогнозирования
- ☐ Потому что это влияет на интерпретируемость модели

Question 3: Choose one answer

В чем отличие данных в исследованиях от данных в индустрии?



В исследованиях данные чистые и статичные, а в индустрии - грязные, постоянно меняющиеся, данные



В исследованиях данные грязные и постоянно меняющиеся, а в индустрии - чистые и статичные



В исследованиях данные исторические, а в индустрии - потоковые



В исследованиях и в индустрии данные одинаковые

Question 4: Choose one answer

Какое значение имеет интерпретируемость в исследованиях и индустрии?



В исследованиях интерпретируемость не всегда важна, а в индустрии - важна



В исследованиях интерпретируемость важна, а в индустрии - не важна



В исследованиях и в индустрии интерпретируемость не важна



В исследованиях и в индустрии интерпретируемость важна

Test 2

Question 1: Choose one answer

Что является узким местом при проектировании системы машинного обучения?

☐ Разработка модели

☐ Вывод модели

☒ Обучение модели

☐ Запуск модели

Question 2: Choose one answer

Какое из следующих утверждений верно относительно пропускной способности?

☐ Она относится к времени, которое требуется от получения запроса до возвращения результата.

☒ Она относится к количеству запросов, обрабатываемых за определенный период времени.

☐ Она относится к прогнозированию оценки модели.

☐ Она относится к времени обработки запросов в сети.

Question 3: Choose one answer

Какое из утверждений верно относительно задержки по времени?

- ☒ Она относится к времени, которое требуется от получения запроса до возвращения результата.
- ☐ Она относится к количеству запросов, обрабатываемых за определенный период времени.
- ☐ Она относится к прогнозированию оценки модели.
- ☐ Она относится к времени обработки запросов в сети.

Question 4: Choose one answer

Какой процентиля используется для определения медианы?

- ☐ p10
- ☒ p50
- ☐ p90
- ☐ d) p99

Question 5: Choose one answer

Какие проценти обычно рассматриваются для обнаружения выбросов?

☐ p10 и p90

☐ p50 и p99

☒ p90 и p95

☐ p99 и p99,9

Test 3

Question 1: Choose one answer

Что представляет собой принцип "Separation of Concerns" в традиционной раз

☒ Разделение программы на отдельные модули, каждый из которых занимается своей задачей

☐ Интеграция кода и данных в один модуль

☐ Выделение кода, приоритетного для тестирования

☐ Разработка программы в команде из специалистов разных профессий

Question 2: Choose one answer

Какое из утверждений больше всего соответствует принципам разработки машинного обучения?

- ☐ Код и данные жестко разделены
- ☐ Важно только версионирование и тестирование кода.
- ☒ ML модель представляет собой интеграцию кода и данных
- ☐ Все точки данных в модели МО равно важны

Question 3: Choose one answer

Какую проблему при работе с данными в контексте машинного обучения иллюстрирует утверждение "git diff обычно не работает"?

- ☒ Объем данных обычно больше, чем кода
- ☐ Данные меняются слишком быстро.
- ☐ Код больше, чем данные
- ☐ Данные всегда рандомны

Question 4: Choose one answer

Какую основную проблему иллюстрирует утверждение "Невозможно наивно копировать большие наборы данных" при версионировании в машинном обучении?

- ☐ Проблема объема хранения данных
- ☐ Проблема внесения изменений в данные
- ☒ Проблема в отслеживании изменений в наборах данных
- ☐ Проблема в отсутствии инструментов для работы с данными

Question 5: Choose one answer

Какую проблему иллюстрирует утверждение "Если на юнит/CI-тесты уходят часы, то разработка застынет" в контексте работы с большими ML-моделями?

- ☐ Недостаточное количество GPU
- ☐ Перегрузка сети
- ☒ Проблему с эффективностью рабочего процесса
- ☐ Проблемы с энергопотреблением

Test 4

Question 1: Choose one answer

Какими двумя основными типами задач являются задачи машинного обучения?

- ☐ Кластеризация и регрессия
- ☒ Классификация и регрессия
- ☐ Дерево решений и нейронные сети
- ☐ Логистическая регрессия и валидация

Question 2: Choose one answer

Что определяет модель классификации?

- ☒ Категории, в которые должны быть классифицированы входные данные
- ☐ Количество входных и выходных данных
- ☐ Скорость обучения
- ☐ Технологию, используемую для обучения модели

Question 3: Choose one answer

Что такое задачи бинарной классификации?

- ☐ Задачи с большим количеством классов
- ☒ Задачи, где есть только два возможных класса
- ☐ Задачи, где важна скорость классификации.
- ☐ Задачи, в которых необходимо использовать нейронные сети.

Question 4: Choose one answer

Что такое "регрессия" в контексте машинного обучения?

- ☒ Регрессия - это модель, которая выводит непрерывное значение
- ☐ Регрессия - это модель, которая классифицирует входные данные в разные категории.
- ☐ Регрессия - это мера эффективности обучения модели
- ☐ Регрессия - это метод устранения ошибок в данных.

Question 5: Choose one answer

Что такое многоклассовая классификация?

- ☒ Задача, где входные данные могут быть классифицированы в более двух классов
- ☐ Задача, где прогностический интерес управляется более чем двумя различными переменными.
- ☐ Задача, где структура данных представляет собой более двух уровней.
- ☐ Задача, где исследуется влияние более двух факторов на прогнозируемый исход.

Question 6: Choose one answer

**Какой подход позволяет сочетать результаты двух моделей и ранжировать со
совокупным оценкам без необходимости переобучения моделей?**

- ☒ Использование двух разных моделей, каждая из которых оптимизирует одну из потерь
- ☐ Использование одной модели с адаптивной функцией потерь
- ☐ Использование комплексной модели с двумя выходами.
- ☐ Использование двух моделей с общей функцией потерь.

Test 5

Question 1: Choose one answer

Что определяют модели данных?

- ☐ Способы хранения данных в базе данных.
- ☐ Способы обработки данных
- ☒ Структуру данных, хранящихся в определенном формате данных.
- ☐ Способы подключения к устройствам для чтения данных

Question 2: Choose one answer

Что такое обучение с подкреплением в машинном обучении?



Это задача понимания естественного языка



Это метод обучения без учителя.



Это метод искусственного интеллекта (AI), где агент изучает, как действовать в среде, выполняя задачи, изучая результаты.



Это метод классификации текстов

Question 3: Choose one answer

Что такое "большие данные"?



Это большие объемы хранимых данных.



Это случайно сгенерированные данные.



Это данные, которые при обработке создают большие объемы выходных данных.



Это очень большие и сложные наборы данных, которые традиционные методы обработки данных не могут обрабатывать.

Question 4: Choose one answer

Что такое данные первого лица (первой стороны)?



Это данные, которые ваша компания уже собирает о своих пользователях или клиентах.



Это данные, собранные другой компанией о своих собственных клиентах, которые они предоставят за них заплатить



Это компании, собирающие данные о тех, кто не является их клиентами.



Это данные, предоставленные государственными органами

Question 5: Choose one answer

Что такое данные второго лица (второй стороны)?



Это данные, которые ваша компания уже собирает о своих пользователях или клиентах.



Это данные, собранные другой компанией о своих собственных клиентах, которые они предоставят за них заплатить



Это компании, собирающие данные о тех, кто не является их клиентами



Это данные, предоставленные государственными органами

Question 6: Choose one answer

Что такое данные третьего лица (третьей стороны)?

- ☐ Это данные, которые ваша компания уже собирает о своих пользователях или клиентах
- ☐ Это данные, собранные другой компанией о своих собственных клиентах, которые они предоставят за них заплатить
- ☒ Это компании, собирающие данные о тех, кто не является их клиентами
- ☐ Это данные, предоставленные государственными органами.

Question 7: Choose one answer

Что такое данные, генерируемые системой?

- ☐ Это данные, получаемые от пользователей.
- ☒ Данные, создаваемые компонентами системы, включая логи и прогнозы моделей.
- ☐ Это данные, полученные из внешних источников.
- ☐ Это данные, хранимые в облаке

Question 8: Choose one answer

Зачем необходимы логи системы?



Логи системы показывают, как работает система, и помогают устранить ошибки и улучшить при



Логи системы необходимы для управления доступом



Логи системы предназначены для долгосрочного хранения данных.



Логи системы используются для резервного копирования данных.

Question 9: Choose one answer

Какие типы данных используются системами для записи поведения пользоват



Ввод данных пользователем, клики, выбор предложения, прокрутка, масштабирование, игнори
окна, долгое время на страницах.



Использование ресурсов системы.



Логи ошибок



Размеры загруженного контента

Test 6

Question 1: Choose one answer

Какие два основных типа выборки существуют?

- ☐ Обучающая и тестовая выборка
- ☐ Обучающая и валидационная выборка
- ☐ Непрерывная и дискретная выборка
- ☒ Вероятностная и невероятностная выборка

Question 2: Choose one answer

Что такое стратифицированная выборка?

- ☐ Это выборка, в которой каждому образцу присваивается вес, определяющий вероятность его в
- ☒ Это выборка, в которой вы сначала делите свою популяцию на группы, которые вас интересуют, а затем выбираете образцы из каждой группы отдельно
- ☐ Это выборка, в которой вы выбираете все образцы в популяции с равными вероятностями их в
- ☐ Это выборка, в которой выбор данных не основан на каких-либо вероятностных критериях

Question 3: Choose one answer

Что такое взвешенная выборка?



Это выборка, в которой выбор данных не основан на каких-либо вероятностных критериях.



Это выборка, в которой вы сначала делите свою популяцию на группы, которые вас интересуют, а затем выбираете образцы из каждой группы отдельно



Это выборка, в которой каждому образцу присваивается вес, определяющий вероятность его выбора



Это выборка, в которой вы выбираете все образцы в популяции с равными вероятностями их выбора

Question 4: Choose one answer

Что такое экспертная выборка?



Это выборка, основанная на доступности данных.



Это выборка, основанная на квотах для определенных сегментов данных.



Это выборка, в которой будущие образцы выбираются на основе существующих образцов



Это выборка, основанная на экспертном мнении.

Question 5: Choose one answer

Что такое простая случайная выборка?

- ☐ Это выборка, в которой каждому образцу присваивается вес, определяющий вероятность его выбора
- ☐ Это выборка, в которой вы сначала делите свою популяцию на группы, которые вас интересуют, а затем выбираете образцы из каждой группы отдельно
- ☒ Это выборка, в которой вы выбираете все образцы в популяции с равными вероятностями их выбора
- ☐ Это выборка, в которой выбор данных не основан на каких-либо вероятностных критериях.

Question 6: Choose one answer

Что такое выборка резервуара?

- ☒ Это метод выборки, который используется при обработке непрерывно поступающих данных.
- ☐ Это метод выборки, который используется при обработке больших объемов данных
- ☐ Это метод выборки, который используется при обработке малых объемов данных
- ☐ Это метод выборки, который используется при обработке статических данных

Question 7: Choose one answer

Что такое резервуар в контексте выборки резервуара?

- ☒ Это массив, в который помещаются выбранные элементы
- ☐ Это функция, которую мы хотим оптимизировать.
- ☐ Это распределение, из которого мы выбираем
- ☐ Это вес, который мы присваиваем выборке

Question 8: Choose one answer

Что такое $Q(x)$ в контексте важности выборки?

- ☒ Это распределение, из которого мы выбираем.
- ☐ Это распределение, которое мы хотим выбрать.
- ☐ Это вес, который мы присваиваем выборке
- ☐ Это функция, которую мы хотим оптимизировать

Test 7 (broken)

Question 1: Choose one answer

Какой пример разбирается в главе "Ручная разметка"?

- ☒ Создание модели ML для распознавания персонажей мультфильма 'Симпсоны'
- ☐ Разработка игры на основе алгоритмов машинного обучения
- ☐ Создание мобильного приложения с использованием машинного обучения
- ☐ Разработка системы безопасности на основе машинного обучения

Question 2: Choose one answer

Какие сервисы в СНГ могут помочь в разметке данных?

- ☐ Google, Yahoo, Bing.
- ☒ ABC Elemetrary и Толока.
- ☐ Facebook, Twitter, Instagram
- ☐ Amazon, eBay, Alibaba

Question 3: Choose one answer

Какое препятствие может возникнуть при ручной разметке данных?

- ☐ Недостаток интернета
- ☒ Проблемы с конфиденциальностью или персональными данными
- ☐ Недостаток программного обеспечения
- ☐ Проблемы с электричеством

Question 4: Choose one answer

Что такое **Weak supervision** в контексте машинного обучения?



Это метод обучения, который использует менее точные, или 'слабые', методы для разметки данных.



Это метод обучения, основанный на использовании большого количества точно размеченных данных.



Это метод обучения, основанный на использовании алгоритмов глубокого обучения.



Это метод обучения, основанный на использовании нейронных сетей.

Question 5: Choose one answer

Какой инструмент с открытым исходным кодом является популярным для **Weak supervision**?



TensorFlow



Keras



Snorkel



PyTorch

Question 6: Choose one answer

Что такое функция разметки (LF) в контексте Weak supervision?

- ☒ Функция, которая кодирует эвристики.
- ☐ Функция, которая кодирует данные для обучения модели.
- ☐ Функция, которая кодирует результаты обучения модели.
- ☐ Функция, которая кодирует ошибки модели.

Question 7: Choose one answer

Какой недостаток может иметь Weak supervision?

- ☒ Метки, полученные при слабом наблюдении, могут быть слишком шумными.
- ☐ Метки, полученные при слабом наблюдении, могут быть слишком точными.
- ☐ Метки, полученные при слабом наблюдении, могут быть слишком простыми.
- ☐ Метки, полученные при слабом наблюдении, могут быть слишком сложными.

Question 8: Choose one answer

Что такое Semi-supervision в контексте машинного обучения?

- ☒ Это метод обучения, который использует структурные предположения для генерации новых меток на основе набора начальных меток.
- ☐ Это метод обучения, который использует эвристики для получения шумных меток.
- ☐ Это метод обучения, основанный на использовании большого количества точно размеченных данных.
- ☐ Это метод обучения, основанный на использовании алгоритмов глубокого обучения.

Question 9: Choose one answer

Когда Semi-supervision наиболее полезен?

- ☒ Когда количество обучающих меток ограничено.
- ☐ Когда есть большое количество точно размеченных данных.
- ☐ Когда используются алгоритмы глубокого обучения.
- ☐ Когда используются нейронные сети.

Question 10: Choose one answer

Что такое transfer learning?

- ☐ Метод обучения с нуля
- ☒ Метод, где модель, разработанная для одной задачи, используется в качестве отправной точки
- ☐ Метод, где модель обучается на одних и тех же данных для разных задач
- ☐ Метод, где модель обучается на разных данных для одной и той же задачи

Question 11: Choose one answer

Что такое моделирование языка?

- ☒ Задача предсказания следующего токена на основе данной последовательности токенов
- ☐ Задача классификации текста на основе его содержания
- ☐ Задача генерации текста на основе заданной темы
- ☐ Задача определения тональности текста

Question 12: Choose one answer

Что такое HuggingFace?

- ☐ Алгоритм машинного обучения
- ☐ База данных для обучения моделей
- ☒ Библиотека, содержащая обширный набор моделей, предобученных для выполнения разнообразных задач
- ☐ Инструмент для визуализации данных

Question 13: Choose all answers that apply

Что означает использование предобученных моделей для бизнеса?

- ☐ Увеличение затрат на обучение моделей
- ☒ Снижение затрат на обучение моделей
- ☐ Увеличение времени на обучение моделей
- ☐ Снижение времени на обучение моделей

Question 14: Choose one answer

Что такое активное обучение?

- ☐ Метод обучения на большом количестве обучающих меток
- ☐ Метод обучения на малом количестве обучающих меток
- ☐ Метод, улучшающий точность ML моделей за счет выборочного использования обучающих данных
- ☒ Метод, где модели машинного обучения могут достигнуть большей точности с большим количеством данных, выбирая из каких образцов данных учиться

Question 15: Choose one answer

Что такое реальное распределение в контексте активного обучения?

- ☒ Это распределение, где у вас есть поток данных, поступающих, как в проде
- ☐ Это распределение, где все данные уже размечены
- ☐ Это распределение, где вы еще не собрали никаких данных
- ☐ Это распределение, где данные поступают постоянно и в большом количестве

Question 16: Choose one answer

Что такое синтетические данные?

- ☐ Данные, полученные путем прямого измерения
- ☒ Данные, которые были искусственно сгенерированы
- ☐ Данные, полученные из реальных источников
- ☐ Данные, которые были получены путем копирования из других источников

Question 17: Choose one answer

Для чего часто используются синтетические данные в машинном обучении?

- ☒ Для увеличения объема обучающих данных или для создания обучающих данных, когда реальных данных недостаточно
- ☐ Для замены реальных данных
- ☐ Для проверки модели на переобучение
- ☐ Для снижения точности модели

Test 8

Question 1: Choose one answer

Почему важна масштабируемость в машинном обучении?

- ☐ Для снижения стоимости
- ☒ Для обработки больших данных
- ☐ Для упрощения моделей
- ☐ Для ускорения обучения

Question 2: Choose one answer

Как решается проблема больших данных в машинном обучении?

- ☐ Используя более мощные компьютеры
- ☐ Уменьшая размер данных
- ☒ Разбивая данные на батчи для параллельной обработки
- ☐ Переносом данных в облако

Question 3: Choose one answer

Что делает техника контрольных точек градиента?

- ☐ Увеличивает скорость обучения
- ☐ Уменьшает объем данных
- ☒ Балансирует использование памяти и вычислений
- ☐ Повышает точность модели

Question 4: Choose one answer

Что такое параллелизм данных в машинном обучении?

- ☐ Использование разных данных для разных моделей
- ☒ Разделение данных между машинами для параллельной обработки
- ☐ Анализ одних и тех же данных на разных машинах
- ☐ Сбор данных с разных источников

Question 5: Choose one answer

В чем разница между ASGD и SSGD?

- ☐ ASGD быстрее
- ☒ ASGD не требует градиентов от всех машин одновременно
- ☐ ASGD использует меньше данных
- ☐ ASGD не поддерживает параллелизм

Question 6: Choose one answer

Какая проблема возникает с ASGD?

- ☐ Требуется больше памяти
- ☒ Застаревание градиента
- ☐ Низкая точность модели
- ☐ Медленное обучение

Question 7: Choose one answer

Почему увеличение размера батча дает убывающую отдачу?

- ☐ Из-за уменьшения точности модели
- ☐ Из-за увеличения времени обучения на шаг
- ☐ Из-за переобучения модели
- ☒ Из-за ограничений вычислительных ресурсов

Question 8: Choose one answer

В чем разница между параллелизмом моделей и данных?

- ☐ Меньше ресурсов для моделей
- ☒ Разные части модели на разных машинах
- ☐ Одна машина для всех вычислений
- ☐ Нет поддержки распределенного обучения

Question 9: Choose one answer

Что такое параллелизм конвейера?

- ☐ Использование разных алгоритмов параллельно
- ☒ Разделение данных на микро-батчи для последовательной обработки
- ☐ Одна модель на всех этапах обработки
- ☐ Обучение нескольких моделей одновременно

Test 9

Question 1: Choose one answer

Что важно знать при оценке ML модели?

- ☐ Её скорость
- ☒ Базовое значение оценки
- ☐ Только её точность
- ☐ Сложность алгоритма

Question 2: Choose one answer

Что такое FID в оценке моделей?

- ☐ Точность классификации
- ☐ Скорость обработки
- ☒ Метрика качества изображений
- ☐ Сложность модели

Question 3: Choose one answer

Зачем использовать базовое значение случайного выбора при оценке модели?

- ☐ Для повышения точности
- ☐ Для проверки предвзятости
- ☒ Сравнить с произвольным угадыванием
- ☐ Для ускорения обучения

Question 4: Choose one answer

Какова цель простого эвристического метода в оценке модели?

- ☐ Упростить модель
- ☒ Проверить базовую эффективность
- ☐ Уменьшить ошибки
- ☐ Ускорить обучение

Question 5: Choose one answer

Что показывает базовое значение нулевого правила?

- ☐ Сложность модели
- ☐ Эффективность наихудшего случая
- ☒ Производительность при предсказании самого распространённого класса
- ☐ Скорость обработки

Question 6: Choose one answer

Почему важна оценка на основе срезов данных?

- ☐ Для уменьшения времени обучения
- ☐ Для повышения общей точности
- ☒ Для выявления предвзятости и недостатков
- ☐ Для уменьшения сложности модели

Question 7: Choose one answer

Что показывает тест возмущения?

- ☐ Быстродействие модели
- ☒ Как модель справляется с шумными данными
- ☐ Общую точность
- ☐ Сложность модели

Question 8: Choose one answer

Зачем нужны тесты инвариантности?

- ☐ Для проверки скорости обработки
- ☒ Для выявления предвзятости в модели
- ☐ Для улучшения точности
- ☐ Для уменьшения сложности

Question 9: Choose one answer

Что такое калибровка модели в ML?

- ☐ Упрощение модели
- ☐ Ускорение обучения
- ☒ Соответствие предсказаний вероятности реальности
- ☐ Увеличение точности

Question 10: Choose one answer

Почему важны тесты направленных ожиданий?

- ☐ Для повышения скорости модели
- ☒ Для проверки логической последовательности предсказаний
- ☐ Для уменьшения ошибок
- ☐ Для упрощения модели