## Тестовое задание

Тестовое задание состоит из нескольких задач по различным темам. Необходимо решить хотя бы одну задачу до технического собеседования. Решение задач допускается продолжить после собеседования. Все задачи имеют разную сложность и разное количество баллов за решение задачи. Чем больше баллов набрано за решение задач, тем лучше.

Решение задач необходимо предоставить в виде воспроизводимого отчета (например: .*ipynb*), если не указано иное. Все файлы решения необходимо упаковать в архив zip. Приветствуется описание приведенных решений и их предпосылок в блокноте решения.

Внимание! Не прикрепляйте к заданиям всю среду виртуального окружения (папки *.venv* и т.п.). Для описания зависимостей используйте файл *requirement.txt*.

### Задача 1

**Описание задачи со слов бизнеса**: Распознать аудиозаписи разговоров записанные с микрофонов для дальнейшей обработки

**Входные данные**: набор аудиофайлов для задачи 1

**Выходные данные**: текст извлеченный из аудио

**С чем предстоит столкнуться**: реальные аудиофайлы записаны в низком качестве с помехами и искажениями. Ваша задача как можно успешнее осуществить полную транскрипцию содержимого аудиофайла.

Оценка задачи:

|  |  |
| --- | --- |
| 2 балла | Транскрипция файла без применения каких-либо улучшений. Это минимально возможное решение задачи |
| 3 балла | Применение улучшений к аудио и последующее распознавание с доказанным эффектом повышения качества |
| 4 балла | Применение нестандартных подходов и сложных методов для улучшения качества аудио или работы процесса распознавания с доказанным эффектом повышения качества |

### Задача 2

**Описание со слов бизнеса**: Получить цвет авто по его фотографии для отдела покраски как в формате для оборудования (RGB), так и для визуального отображения сотрудникам.

**Входные данные**: изображения вероятно содержащие автомобили

**Выходные данные**: список кортежей каждый из которых содержит значения каналов RGB, а также отчет в котором можно визуально увидеть схожесть определенного цвета и цвета автомобиля на картинке (можно прямо внутри .ipynb файла сделать визуализацию)

**С чем предстоит столкнуться**: реальные изображения могут быть искажены, обьекты на них не всегда имеют нужную перспективу и не всегда на самом изображении может присутствовать искомый обьект

Оценка задачи:

|  |  |
| --- | --- |
| 1 балл | Обработка изображения в результате которой можно получить цвета выделенные из него. Минимально возможное решение, не учитывающее особенности данных |
| 2 балла | Решение учитывает обьекты на изображении и применяет для них корректный метод извлечения цвета |
| 3 балла | Решение учитывает перспективу и геометрию обьектов на изображении и применяет не стандартные/сложные методы обработки. Для извлечения цвета используется не стандартный/сложный алгоритм |

### Задача 3

**Описание со слов бизнеса**: Необходимо автоматизировать оборот документации. Для этого нужно понимать автоматически, что перед нами за документ. Сделать это надо как можно более экономно и только на своих мощностях для сохранения коммерческой тайны.

**Входные данные**: изображения вероятно содержащие документы

**Выходные данные**: метка принадлежности документа

**С чем предстоит столкнуться**: реальные изображения могуть иметь некорректную обрезку, искаженную геометрию или иные дефекты.

Оценка задачи:

|  |  |
| --- | --- |
| 3 балла | Решение позволяет классифицировать изображение. Минимально возможное решение |
| 4 балла | Применяется обработка для устранения искажений с доказанной эффективностью |
| 6 баллов | Применяется обработка для вырезания изображения, нормализация и иные приемы повышения качества изображения с доказанной эффективностью. Применяется не стандартный подход к классификации |
| 7 баллов | Применяется комплекс приемов устраняющих дефекты на изображении. Решение учитывает сдвиги классификатора |

### Задача 4

**Описание со слов бизнеса**: Необходимо снизить нагрузки на колл-центры частично автоматизировав обработку обращений. Клиенты могут звонить по разным вопросам и нужно перенаправлять их звонки к операторам или отвечать заранее заготовленной фразой. Для этого нужно понимать что сказал клиент и куда его направить. Сделать это надо как можно более экономно и только на своих мощностях для сохранения коммерческой тайны.

**Входные данные**: аудиофайлы звонков

**Выходные данные**: метки характеризующие звонок

**С чем предстоит столкнуться**: реальные записи звонком имеют искажения и высокую долю шума, а так же могут содержать нецелевой контент. Кроме того, нужно учитывать, что намерение может быть распознано некорректно и тогда, лучше передать звонок оператору. То же самое касается обратного: если мы уверены, что это нецелевой звонок, то он не должен попасть к оператору

Оценка задачи:

|  |  |
| --- | --- |
| 5 баллов | Решение корректно выставляет метки для звонков с низким уровнем искажений |
| 6 баллов | Применены методы снижения искажений аудиодорожки, нестандартный подход к классификации с доказанной эффективностью |
| 8 баллов | Решение учитывает контекст аудиофайла и его содержимое. Применены методы существенно повышающие качество классификации с доказанной эффективностью |

### Задача 5

**Описание со слов бизнеса**: Необходимо внедрить сервис с использованием машинного обучения в бизнес-процесс. В систему будут попадать записи звонков, которые нужно транскрибировать и текст отправлять дальше на другой сервис. В случае, если что-то пошло не так, необходимо уведомлять внешние системы об ошибке. Внешние апи для распознавания использовать нельзя, из-за коммерческой тайны.

**Входные данные**: аудиофайлы из задания 1

**Выходные данны**е: вызов метода внешней системы в котором будет передан результат работы системы

**С чем предстоит столкнуться**: реальные требования бизнеса не всегда четко указывают на технические ограничения будущей системы. Это задание не сосредоточено на качестве распознавания аудио, а рассматривает то, насколько гибкий и настраиваемый сервис вы можете спроектировать и реализовать, насколько хорошо вы понимаете как система будет работать под реальной нагрузкой и как встроить ее в бизнес-процессы. Поэтому качество транскрипции оцениваться не будет.

Оценка задачи:

|  |  |
| --- | --- |
| 4 балла | Создан сервис с апи содержащим только метод для транскрипции аудиофайла, который принимает путь к файлу. Это минимально возможное решение |
| 5 баллов | Сервис поддерживает многопоточную обработку, имеет низкую степень зависимости между компонентами. |
| 6 баллов | Сервис подготовлен к развертыванию как конечное решение в виде контейнеров. Сервис просто расширяем. |
| 8 баллов | Осуществлена корректная декомпозиция сервера на компоненты, учтена возможность распределенного развертывания как контейнера или пода. Учтены и описаны проблемы, которые могут возникнуть при масштабировании и примерные нагрузки, которые может выдержать сервис. Код покрыт тестами. |

### Задание 6

**Описание со слов бизнеса**: Внешние заказчики хотят увидеть демо проекта. Их не интересуют технические подробности решения, хотят оценить процесс работы

**Суть задачи**: В этом задании необходимо реализовать демо-стенд для одного из предыдущих заданий в виде веб-приложения. Для этого можно использовать библиотеки интерактивного прототипирования: *Streamit, Gradio, Marimo, Solara* и т.п.

Оценка задачи:

|  |  |
| --- | --- |
| 2 балла | Реализован интерактивный дашборд/страница с использование одной из библиотек прототипирования. Это минимально возможное решение для этой задачи. |
| 3 балла | Решение подготовлено к развертыванию в виде контейнера или пода |
| 4 баллов | Решение содержит методы снижения нагрузки на бэкенд (кеширование и т.п.), |

### Задание 7

**Описание со слов бизнеса**: Очень надо понять, когда напомнить клиенту о ТО для автомобиля. Мы выгрузили вам данные из систем, скажите что сможете сделать и в какие сроки?

**Входные данные**: набор данных содержащий данные об автомобилях

**Суть задачи:** В этой задаче необходимо понять степень ее сложности, декомпозировать и оценить сроки ее реализации. Не требуется совершать поиск готового решения, а лишь указать на гипотезы, которые могут помочь, произведя разведывательный анализ данных.

Оценка задачи:

|  |  |
| --- | --- |
| 4 балла | Произведена очистка и агрегация данных. Выдвинута одна гипотеза, предпосылки или прогнозы отсутствуют. Это минимальный вариант решения этой задачи |
| 6 баллов | Произведены трансформации и статистические проверки данных, учтены их особенности. Выдвинуто несколько гипотез с указанием их различий |
| 8 баллов | Учтена доменная область задания. Произведена декомпозиция задачи. Выдвинуто несколько гипотез с разными подходами к решению. Произведена оценка сложности по шкале (легко, нелегко, сложно, невыполнимо) |
| 10 баллов | Осуществлено планирование экспериментов, оценено примерное время выполнения кажого этапа или эксперимента. Выдвинуты оценки перспективности каждой гипотезы на основе анализа данных, предложены метрики качества и приемы их улучшения |