

BEST HACK

Data Science

Вам предлагается 2 задачи. Для успешного прохождения полуфинального этапа вам необходимо решить только ОДНУ задачу на ваш выбор.

Требования к решению:

Перед отправкой решения убедитесь, что оно соответствует следующим критериям:

1. Присутствует презентация вашего решения в формате .pptx или .pdf, в которой присутствует следующая информация:
 - a. Название команды;
 - b. Имя капитана команды;
 - c. ФИО членов команды;
 - d. Подробное текстовое описание решения с обоснованием выполненных действий;
 - e. Визуализация данных и работы модели;
 - f. Результаты, полученные при вычислении указанной метрики качества.
2. Присутствует файл (ноутбук) с программой в формате .ipynb (Jupyter Notebook, Google Collab, Kaggle Notebook и т.п.), где подробно описано решение. Обратите внимание, что требуется именно файл, а не ссылка на него.

Решение отправляйте на почту besthackathon2022@gmail.com с темой **[ПФ][Data Science] <Название команды>**. В теле письма обязательно должны быть указаны ФИО капитана команды. Все решения принимаются до 15 апреля, 21:00.

Задача №1.

Один из самых важных органов человека – это сердце. В древние времена считалось, что именно в сердце заключена душа человека, его сущность, а анализ крови способен рассказать о характере человека. Современная кардиология не полностью отказалась от таких идей, и по сей день большинство медицинских компаний и учреждений используют электрокардиограмму (ЭКГ) для получения сведений о состоянии здоровья пациента. По ЭКГ возможно определить запас сил человека, его усталость, вегетативный баланс и многое другое, однако при установке соответствующего оборудования для снятия сердечного ритма довольно часто необходимо дополнительно настраивать алгоритмические системы для отображения и обработки данных.

В последние годы количество исследований на тему того, что биение сердца у каждого человека индивидуально, многократно возросло.

Компания ООО “Квазар” занимается разработкой программного обеспечения и медицинских датчиков в сфере телемедицины. Наша компания разработала мобильное устройство для снятия ЭКГ в режиме реального времени и хочет помочь организациям здравоохранения ускорить процедуру обработки и персонализации ЭКГ, для этого вам предстоит решить задачу по идентификации человека по его электрокардиограмме.

Задание:

Учитывая все сложности необходимо разработать модель машинного обучения, способную идентифицировать человека из заданной выборки по переданной в нее электрокардиограмме.

Сложности, с которыми придется столкнуться:

1. Сигнал имеет много помех, его необходимо сгладить
2. Визуально кардиограмма может быть у двух человек похожа
3. ЭКГ может содержать различного рода заболевания (аритмии, фибрилляции), что усложняет распознавание ритма
4. Частота Сердечных Сокращений (ЧСС) может увеличиваться и уменьшаться на записи, поэтому необходимо четко выделять удары сердца
5. Возможна дыхательная аритмия (т.е. задержка ритма, связанная с изменением дыхания)

Данные к задаче:

Датасет к задаче и дополнительную информацию вы можете получить по ссылке: https://drive.google.com/drive/folders/14EdcxhmMyVN_aijJ5sC3cT6Q5CL9f2OL?usp=sharing

Критерии оценки:

Качество модели будет оцениваться по метрике F1-взвешенное

Дополнительные баллы:

Дополнительные баллы можно получить, если система будет способна отличить здорового человека от больного (больным считается тот, у кого имеются отклонения в сердечном ритме) или способна по ЭКГ приблизительно дать информацию о человеке (пол, возраст)

Задача №2.

Область цифровизации повседневных вещей в 21 веке находится в тенденции роста популярности. Именно благодаря этой тенденции мы сейчас можем переписываться в мессенджерах, записываться к врачу через ЕМИАС, заказывать такси через мобильное приложение и так далее.

ООО "Квазар" занимается разработкой программного обеспечения в сфере Искусственного Интеллекта, автоматизации бизнес-процессов и биомедицинских технологий. Компания успешно разработала и внедрила «Цифровой завод» на предприятия Московской, Калужской области и других.

Сейчас в нашей Компании развивается проект по цифровизации отделения реанимации в больницах. Мы планируем помочь врачам максимизировать внимание, направленное на лечение конкретного пациента. Для этого используются технологии машинного обучения для предсказания групп здоровья, предиктивной аналитики по диагнозам и другой функционал.

Задание:

В рамках хакатона Вам необходимо решить одну из задач «Цифровой реанимации». В представленном датасете имеются медицинские параметры пациентов, больных раком. Необходимо классифицировать на основе этих данных – пациент жив или погиб.

Для этого задания заранее подготовлен baseline с точностью 50%. Файл можете найти по ссылке ниже.

Сложности, с которыми придется столкнуться:

1. Медицинские данные полны терминологии, которую необходимо изучить для проработки вопроса;
2. Низкая корреляция данных с таргетом, возможно нужно создавать новые фичи;
3. Признаков (фич) много, необходимо выбрать те, которые способны дать максимальный результат.

Данные к задаче:

Датасет к задаче и дополнительную информацию вы можете получить по ссылке: <https://drive.google.com/drive/folders/12fNV9N0ZBqtHVMaKi1gz0KEhtOliZzRC?usp=sharing>

Критерии оценки:

Качество модели будет оцениваться по метрике accuracy.

Дополнительные баллы:

Предсказание других признаков и других таргетов.