



Задача «Создание модели прогнозирования заболеваемости населения»

Введение

Использование технологий искусственного интеллекта — один из главных трендов современного здравоохранения. Технологии ИИ могут повысить точность диагностики, поспособствовать разработке новых лекарственных препаратов, снизить расходы на лечение, повысить качество медицинских услуг, развить механизмы дистанционного мониторинга состояния здоровья и даже прогнозировать возможные вспышки заболеваний.

Известно, что еще в конце 2019 года алгоритм одного из ИИ-стартапов предсказал распространение коронавирусной инфекции на основе публикаций в интернете о новой болезни, появляющихся в интернете, и анализе авиасообщения между странами. Создание и развитие подобных аналитических инструментов в будущем позволит предупредить пандемии и сохранить жизни миллионов людей.

Условие задачи (требования к решению)

В рамках чемпионата участникам предлагается создать математическую модель, которая способна прогнозировать количество пациентов по каждому виду заболевания согласно международной классификации болезней в привязке к населенному пункту, месяцу и половозрастным характеристикам на основе многолетних данных.

Описание входных значений

train.csv – набор данных для обучения

PATIENT_SEX – пол группы пациентов

MKB_CODE – первичный диагноз группы пациентов, код МКБ-10

ADRES – населенный пункт группы пациентов

VISIT_MONTH_YEAR – месяц и год постановки диагноза

AGE_CATEGORY – возрастная категория группы пациентов (Классификация ВОЗ)

PATIENT_ID_COUNT – кол-во пациентов группе, которая характеризуется вышеперечисленными признаками

test.csv – набор данных для тестирования

PATIENT_ID_COUNT – отсутствует, его необходимо определить

sample_solution.csv – формат отправки решения

Метрика

Коэффициент детерминации (R2):

$$R2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$$

Правила чемпионата:

1. С момента открытия датасета до момента завершения приема решений репозиторий участника, в котором он ведет разработку по задаче текущего чемпионата, должен оставаться закрытым.
2. Участник обязан открыть доступ к репозиторию на чтение по ссылке (которая была прикреплена в ЛК в поле «Ссылка на код (гитхаб)») не позднее чем в течение 12 часов с момента окончания дедлайна отправки решений на региональном чемпионате.
3. Согласно п. 5.8 Положения в процессе верификации решений организаторы и технические эксперты, проверяющие решения участников, в праве назначить интервью с участниками чемпионата. Участник получит приглашение и ссылку на интервью не позднее чем за 12 часов до публикации итогового лидерборда. Пропуск интервью участником является поводом для дисквалификации.
4. Организаторы вправе исключить участника из призовых позиций лидерборда за непредоставление одного из артефактов решения задачи: тизера, скринкаста, презентации, ссылки на репозиторий.

5. Организаторы вправе дисквалифицировать участника в случае выявления плагиата кода или несоблюдение Положения конкурса.

6. Участник, получивший 2 дисквалификации за сезон конкурса, попадает в чёрный список с дальнейшим отстранением от участия в чемпионатах до конца сезона.