



Задача «Разработка модели прогнозирования дорожно-транспортных происшествий»

Введение

Ежегодно в мире от дорожно-транспортных происшествий (ДТП) погибает более миллиона человек. В ряде стран смертность от автомобильных аварий является одной из главных причин уменьшения населения, особенно детей и молодежи в возрасте от 5 до 29 лет. По данным Госавтоинспекции в России в 2021 году произошло 133 тысячи автомобильных аварий, в которых пострадали более 170 тысяч человек.

Вероятность дорожно-транспортных происшествий динамична и зависит от большого количества параметров. Наряду с несоблюдением правил дорожного движения к причинам травматизма на дорогах относятся: некачественная разметка дорожного полотна, ненадлежащее состояние дорожных знаков, нехватка барьерных ограждений для разделения встречных полос, отсутствие освещения, недостаточное количество регулируемых пешеходных переходов. Существующая на сегодняшний день система по анализу и мониторингу ДТП в РФ позволяет просматривать статистику по происшествиям на дорогах, однако она недостаточно удобна для принятия управленческих решений по улучшению дорожно-транспортной инфраструктуры.

Для повышения безопасности на дорогах участникам чемпионата предстоит создать модель прогнозирования локализации дорожно-транспортных происшествий на основе накопленных многолетних данных. Разрабатываемая модель позволит реализовывать более эффективные мероприятия по снижению факторов, влияющих на возникновение дорожно-транспортных происшествий, тем самым снижая уровень аварий и травматизма.

Условие задачи

Необходимо разработать модель машинного обучения, принимающую на вход место, категорию аварии и предсказывающую количество ранений/летальных исходов у взрослых и подростков.

Описание входных значений

train.csv — содержит в себе ~3600 зарегистрированных случаев ДТП в Нижегородской области;

test.csv — тестовая выборка для предсказания;

sample_submission.csv — образец файла для отправки.

Метрика

В качестве метрики выступает Recall.

$$Recall = \frac{TP}{TP + FN}$$

Так как конечная задача предполагает мультиклассовое предсказание, то для каждой категории будет рассчитано свое значение Recall , которое будет усреднено. Полученный таким образом Recall и будет являться оценкой точности участника.