Озеров П.В.

ИУ5-65

РК-1

Какие способы обработки пропусков в данных для категориальных и количественных признаков Вы использовали?

В числовых данных был применен способ "Внедрение значений" – импьютация. Для заполнения пропущенных значений использовался столбец “YEAR”. Таким образом, в качестве стратегии была использована медиана, так как она более устойчива к выбросам в данных.

Для категориальных данных был выбран аналогичный способ, но была использована стратегия “most_frequent”-наиболее часто встречающееся значение.

Какие признаки Вы будете использовать для дальнейшего построения моделей машинного обучения и почему?

Для исходного набора данных я бы убрал следующие признаки:

Колонку GSM, так как фактически является не заполненным столбцом (количество пустых значений 99.07%).

Колонка EYE, так как количество пустых значений превышает половину (52.61%)

Остальные признаки, как числовые, так и категориальные, стоит оставить.