# Отчёт

В ходе выполнения работы был сформирован набор данных пациентов реанимации. Изначально, для каждого параметра пациента бралось средние значение и стандартное отклонение, но этого оказалось недостаточно, классификаторы показывали низкую точность. Поэтому я добавил больше признаков: минимум, максимум и квантили 25 и 75. Затем были обработаны пропущенные значения, если пропусков в столбце было более 30 процентов, столбец выкидывался, остальные пропуски заменялись средним. Категориальные признаки были обработаны. Аномальные значения убраны. Забыл сделать матрицу корреляций и выкинуть признаки с высокой корреляцией.

Для всего набора данных была проведена стандартизация. Были проведены эксперименты с методами библиотеки imblearn, т.к. выборка была не сбалансирована.

Затем с помощью GridSearch подбирались лучшие гиперпараметры каждой из 4 моделей. Лучшие результаты были у XGB и RandomForest.

Для случайного леса и логистической регрессии были выведены важности признаков.

Потом для каждой модели была выведена ROC кривая, средняя Roc и её стандартное отклонение. Для случайного леса и логистической регрессии посчитана и визуализирована средняя важность признаков.

Попытался вывести в пространстве 2 наиболее важных признаков регионы решений, но классы линейно не разделимы и удалось вывести только элементы классов в этом пространстве.