Здравствуйте.

По-прежнему остался вопрос с предыдущего ДЗ по теме: **Оценка точности модели, переобучение, регуляризация.**

Для логической регрессионной модели решил оставить признак – **год рождения спортсмена**. Для этого предварительно выполнил следующий код:

df['dob\_year'] = df['dob'].apply(lambda x: pd.to\_datetime(x,format='%m/%d/%y').year)

У 65 спортсменов из 11 538 год определялся как 2068, 2067, 2056... Такая особенность отмечается у всех спортсменов год рождения которых старше 1969.

Для исправления данной ситуации преподавателем было рекомендовано использовать параметр yearfirst=True:

«Когда вы используете %y, pandas.to\_datetime() переводит двухзначные годы следующим образом: числа от 00 до 68 интерпретируются как 2000-2068, числа от 69 до 99 интерпретируются как 1969-1999. Чтобы гарантировать, что двухзначные годы относятся к XX веку (1900-е), можно использовать параметр yearfirst=True»

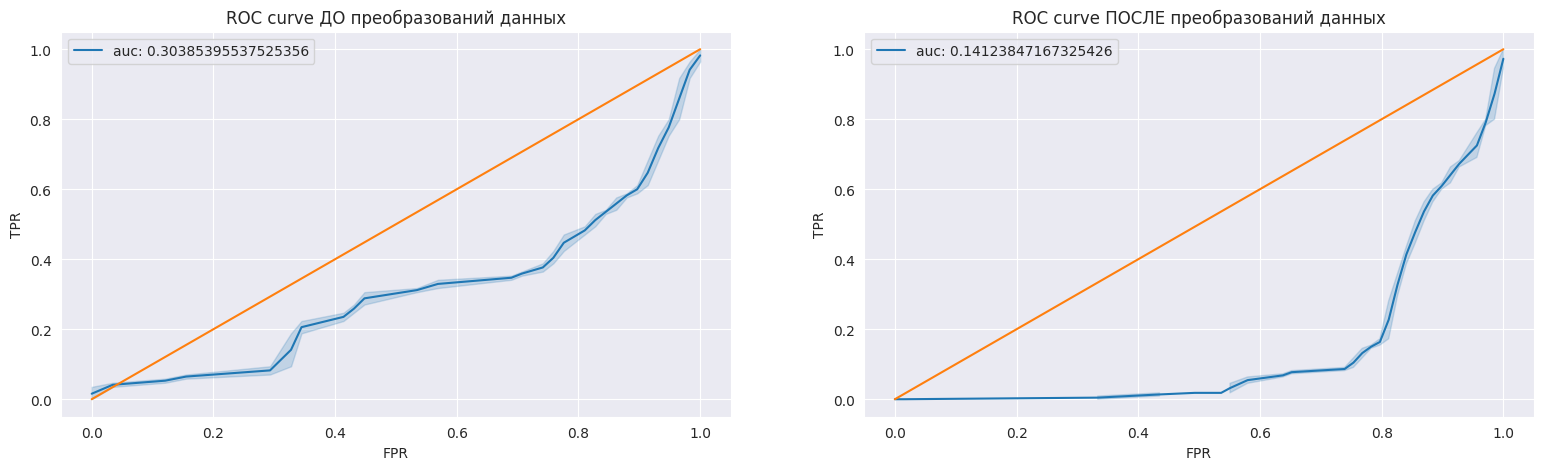
Я использовал данный параметр, код:

df['dob\_year'] = df['dob'].apply(lambda x: pd.to\_datetime(x, format='%m/%d/%y', yearfirst=True).year)

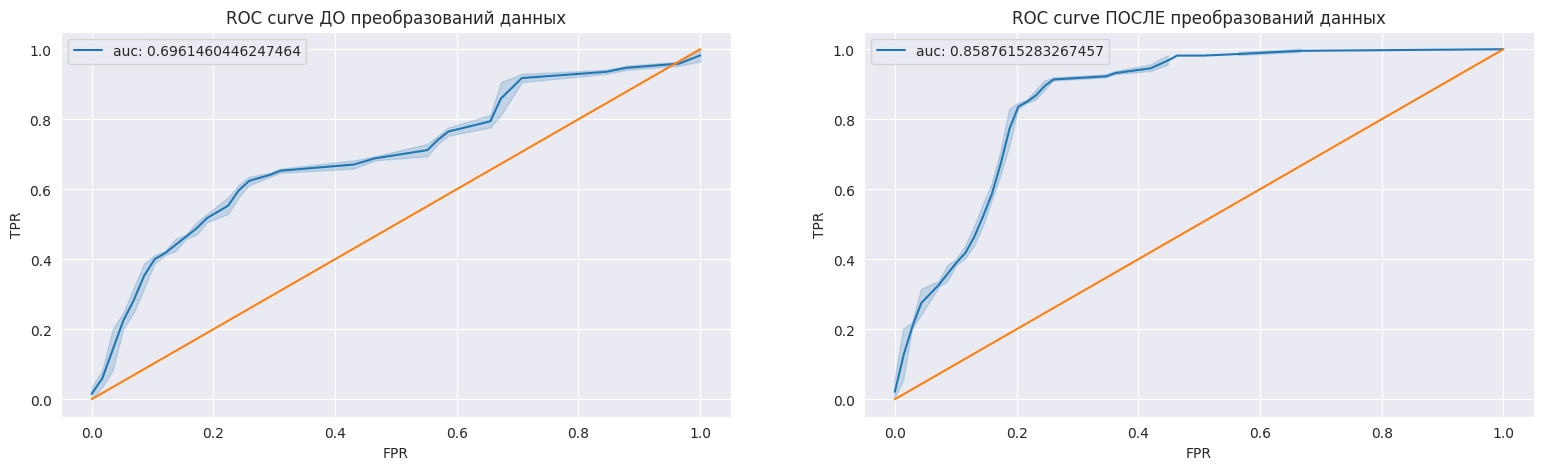
К сожалению, по никоторым спортсменам год конвертировался так же некорректно. У 65 спортсменов из 11 538 год определялся как 2068, 2067, 2056...и такая особенность отмечается у всех спортсменов год рождения которых старше 1969 г. Таким образом результат тот же что и без использования параметра yearfirst=True.

При выполнении домашнего задания на тему **Проблема качества данных** возник сложный для меня вопрос :

Если использовать вероятности **1 класса (выжившие**) : predict[:1], predict\_2[:,1] соответственно, то у обоих моделей кривая выстраивается с противоположной стороны – ниже оранжевой линии, площадь так считается некорректно график:



Если для построения использовать вероятности 0 класса: predict[:0], predict\_2[:,0], то графики выстраиваются так как нужно.



Возникают сомнения. Означает ли это то, что модели путают классы: 0 класс предсказывается как 1, а 1 как 0. Если это так, то как исправить данную ситуацию?