**МОЯ ЛУЧШАЯ МОДЕЛЬ**

**каггл =====> 0.76055**

**При**

**Для линейной регрессии r2 TEST = -2.2352 TRAIN= 0.6524**

**Для случайного леса r2 TEST = 0.7475 TRAIN= 0.9629**

**Для градиентного бустинга r2 TEST = 0.7773 TRAIN= 0.864**

**Для XGB r2 TEST = 0.779 TRAIN= 0.9561**

**Для LGB r2 TEST = 0.7512 TRAIN= 0.7933**

**Для stack r2 TEST = 0.7823 TRAIN= 0.9106**

**Параметры модели:**

**stack = StackingRegressor([('rf', RF\_model), ('gb', GB\_model), ('xgb', xgb\_model), ('lgb', lgb\_model)], final\_estimator=lr)**

**stack.fit(X\_train, y\_train)**

**GB\_model = GradientBoostingRegressor(criterion='mse',**

**max\_depth=6,**

**min\_samples\_leaf=50,**

**random\_state=42,**

**n\_estimators=2250,**

**max\_features='sqrt',**

**loss='huber',**

**learning\_rate=0.025)**

**xgb\_model = xgb.XGBRegressor(colsample\_bytree=0.29,**

**learning\_rate=0.015, max\_depth=8,**

**min\_child\_weight=0.1,**

**n\_estimators=1000,**

**reg\_lambda=0.7,**

**random\_state =7)**

**lgb\_model = lgb.LGBMRegressor(objective='regression',num\_leaves=5,**

**learning\_rate=0.05, n\_estimators=720,**

**max\_bin = 55, bagging\_fraction = 0.8,**

**bagging\_freq = 5, feature\_fraction = 0.2319,**

**feature\_fraction\_seed=9, bagging\_seed=9,**

**min\_data\_in\_leaf =6, min\_sum\_hessian\_in\_leaf = 11)**

**-------------------------------------- фичи ---------------------------------------------------**

**feature\_names = ['Rooms',**

**'Square', 'LifeSquare', 'KitchenSquare',**

**'Floor',**

**'HouseFloor',**

**'HouseYear',**

**'Ecology\_1',**

**#'Ecology\_2',**

**#'Ecology\_3',**

**'Social\_1',**

**'Social\_2',**

**'Social\_3',**

**'Healthcare\_1',**

**# 'Helthcare\_2',**

**'Shops\_1' ,**

**#'Shops\_2'**

**]**

**new\_feature\_names = ['Rooms\_change',**

**#'HouseFloor\_change', ####**

**#'HouseFloor\_0' ###**

**#'HouseYear\_outlier',**

**# 'Square\_min',**

**'LifeSquare\_nan',**

**'DistrictSize',**

**# 'new\_district',**

**#'IsDistrictLarge',**

**#'MedPriceByDistrict',**

**#'MedPriceByFloorYear',**

**'MedPriceStalinka',**

**'MedPrice\_District\_Health',**

**#'MedPrice\_District\_socia',**

**#'HouseFloor\_veryhight',**

**# 'MedPrice\_District\_Shop\_1'#,**

**#'Social'**

**# 'Shops'**

**]**

**Идеи по предобработке:**  
Общие идеи: убрать выбросы, заменить пропущенные значения в признаках на новые значения. На какие значения менять - попробуем проэкспериментировать.

1. Комнаты (Rooms)
   * в квартирах, где комнат больше 6, заменить это значение на среднее
   * добавить новый признак (Rooms\_change), отражающий эти изменения в признаке Rooms
2. Площади (KitchenSquare)
   * в объектах, где площать кухни KitchenSquare пропущена или больше, чем у 99,5 % других объектов - заменить на медиану по признаку (трейновской выборки)
   * в объектах, где площать кухни KitchenSquare меньше, чем 4м2 - заменить на 4м2
3. Этажность (HouseFloor)
   * нулевые значения HouseFloor заменить на медиану.
   * добавить новый признак (HouseFloor\_0), отражающий эти замены в признаке HouseFloor
   * если этажей больше чем 50, то заменить на медиану.
   * добавить новый признак (HouseFloor\_veryhight), отражающий эти замены в признаке HouseFloor
   * если Floor < HouseFloor - поменяем Floor и HouseFloor местами
   * в квартирах, где "Floor" меньше "HouseFloor" поменять местами эти значения
   * HouseFloor не может быть выше 75 этажей (все что выше - меняем на медиану)
4. Год постройки дома (HouseYear)
   * если год постройки дома больше текущей даты, то заменить на текущий год.
   * добавить новый признак (HouseYear\_outlier), отражающий эти замены в признаке HouseYear
5. Здоровье\_медицина (Healthcare\_1)
   * пропуски в признаке Healthcare\_1 заменим медианой, определенной внутри объектов этого же района
6. Этажность (Square и LifeSquare)
   * значения Square меньше 20м заменить на медиану.
   * добавим новый признак (Square\_min), отражающий эти замены в признаке Square
   * Поменяем значения Square и LifeSquare, если одно больше другого
   * пропуски в LifeSquare заменим на медиану
   * добавим новый признак (LifeSquare\_nan), отражающий эти замены в признаке LifeSquare

**Создадим новые признаки:**

1. Признак района (DistrictId)
   * номер DistrictId свяжем с количеством объектов в этом районе
2. Новый признак (MedPriceByDistrict)
   * Построим зависимость стоимости квартиры от количества комнат и от района, в котором она находится.
3. Новый признак (MedPriceByFloorYear)
   * Построим зависимость стоимости квартиры от этажа и года постройки дома.
4. Новый признак (MedPriceStalinka)
   * Построим зависимость стоимости квартиры от типа дома (этажность и год постройки: "сталинки", "хрущевки" и т.д.).
5. Новый признак (MedPrice\_District\_Health)
   * Построим зависимость стоимости квартиры от обеспеченности района медучреждениями (признаки Healthcare).
6. Новый признак (MedPrice\_District\_socia)
   * Построим зависимость стоимости квартиры от обеспеченности района объектами соцкультбыта (признаки Social).
7. Новый признак (MedPrice\_District\_Shop\_1)
   * Построим зависимость стоимости квартиры от обеспеченности района магазинами (признаки Shop)