Решение задачи "Ascott group"

Амир Мирас

МГУ имени М.В. Ломоносова

9 января 2018 г.

CatBoost vs lightgbm (до изменения PL)

```
train = pd.read_csv("files-ascott_group/train_set_weeks.csv")
test = pd.read_csv("files-ascott_group/test_set_weeks.csv")

cat = CatBoostRegressor(thread_count=8)
lgb = LGBMRegressor(n_estimators=100, n_jobs=-1)

lgb.fit(train[['idFilial', 'KanalDB', 'idSubGrp', 'N wk']],
        train['value'])
cat.fit(train[['idFilial', 'KanalDB', 'idSubGrp', 'N wk']],
        train['value'], cat_features=[0, 1, 2])
```

Алгоритм	RMSE на PL
lightgbm	58648
CatBoost	94048

Парные признаки (до изменения PL)

```
train['fk'] = train['idFilial'].astype(str) + "_" + train['KanalDB'].astype(str)
train['fs'] = train['idFilial'].astype(str) + "_" + train['idSubGrp'].astype(str)
train['ks'] = train['KanalDB'].astype(str) + "_" + train['idSubGrp'].astype(str)

columns = ['idFilial', 'KanalDB', 'idSubGrp', 'fk', 'fs', 'ks', 'N wk']
cat = CatBoostRegressor(thread_count=8)
cat.fit(train[columns], train['value'], cat_features=[0, 1, 2, 3, 4, 5])
```

Алгоритм	RMSE на PL
seed=0	59358
seed=None	58079
seed=???	57073

Сглаженные средние (до изменения PL)

1. Для категориальных признаков (включая парные) считаем следующие счетчики куммулятивным проходом по данным:

$$SmoothMeanValue = \frac{N \cdot MeanValue + \alpha \cdot GlobalMeanValue}{N + \alpha}$$

2. Обучаем $\mathrm{CatBoost}$

#	Участник	Счет
1	amirassov_MMP_MSU	55309.43174
2	pashakovalenko_MMP_MSU	56294.46284
3	datamove	56356.96667

Сглаженные средние (до изменения PL)

Признаки	**RMSE на CV**
Все	70k
Все – сглаженные средние	72k
Все – N wk	71k
Все – сглаженные средние – N wk	113k

Вывод: сглаженные средние по времени $\neq$ кодирование алгоритма $\mathrm{CatBoost}$

Стекинг (до изменения PL)

Аналогично сглаженным средним проводим стекинг со следующими алгоритмами на первом уровне:

1. *KNN*
2. *XGBoost*
3. *LightGBM*
4. *RandomForest*

Полученные признаки добавляем в исходное признаковое пространство и обучаем *CatBoost*.

7	che	41570.00496	2
8	iggisv9t	41581.57187	116
9	amirassov_MMP_MSU	41905.81069	28