ФПМИ Альфа Банк

GEEK BATTLE

Решение бизнес-задач, связанных с CLTV

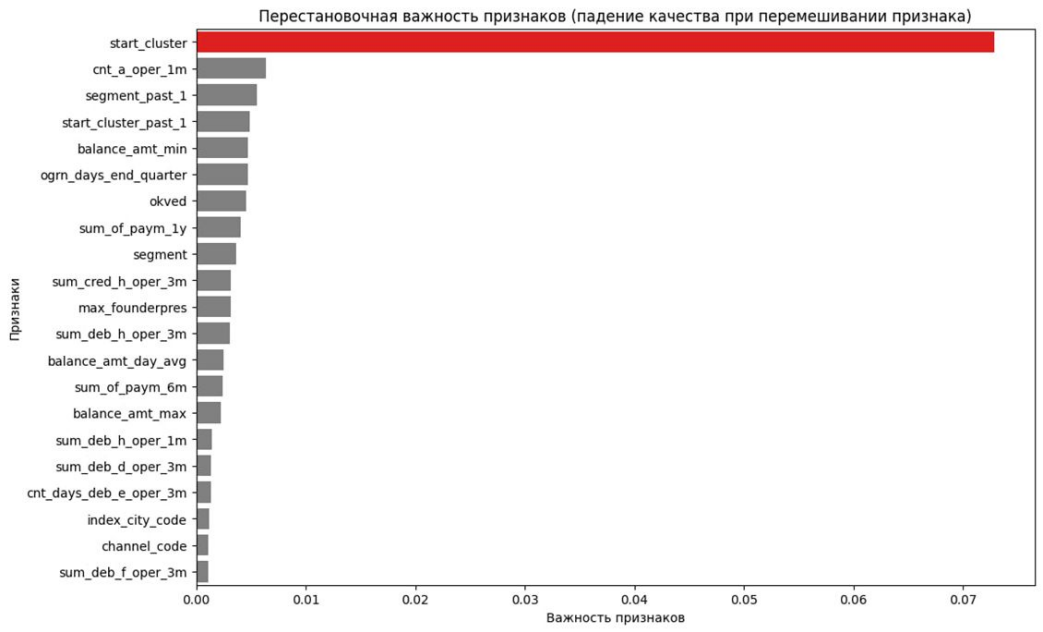
No Loss March

ФПМИ Альфа Банк

GEEK BATTLE

Start cluster – самый важный признак.

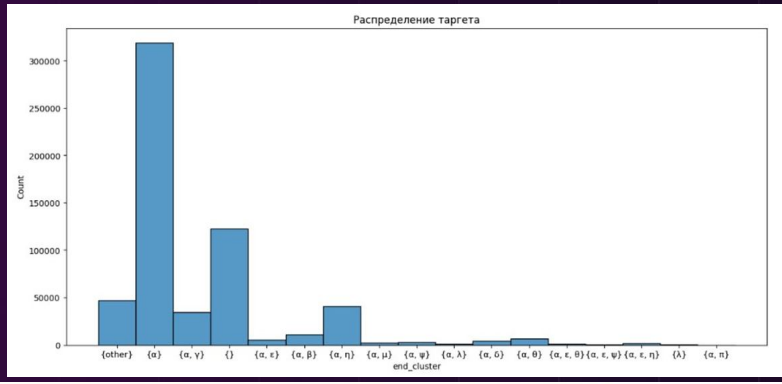
82% - точность константного baseline, основанного на Start_cluster



ФПМИ Альфа Банк

GEEK BATTLE

Частотный анализ Start cluster

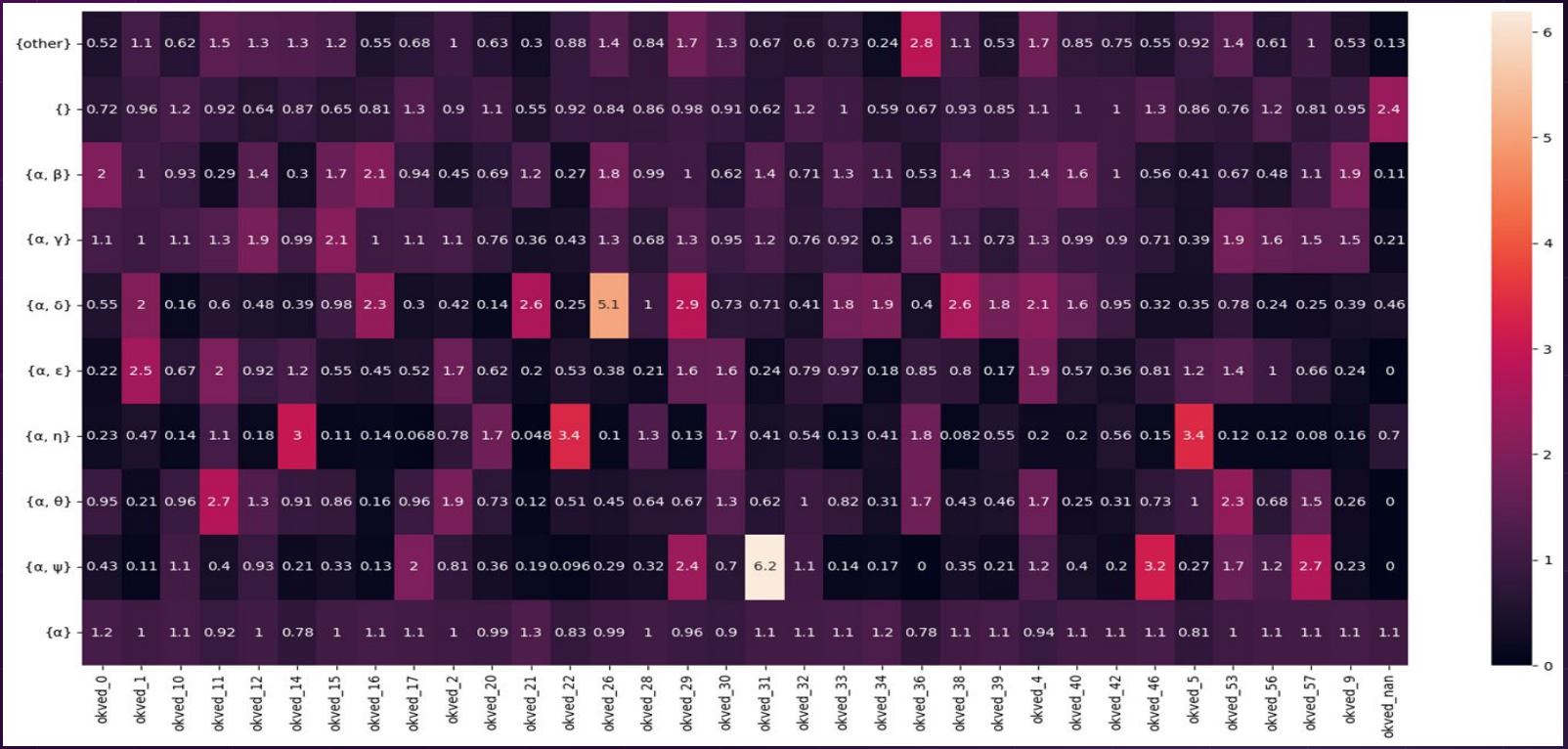


	{λ}	{α}	{α, ψ}	{α, π}	{α, μ}	{α, λ}	{α, θ}	{α, η}	{α, ε}	{α, ε, ψ}	{α, ε, θ}	{α, ε, η}	{α, δ}	{α, γ}	{α, β}	{}	{other}
{λ}	0.71	0	0	0	0	0.21	0	0	0	0	0	0	0	0	0	0	0.083
{α}	0	0.97	0.00016	0	0.00049	0.0003	0.00032	0.0064	0.00039	0	0	1.8e-05	0.0017	0.0046	0.004	0.01	0.0013
{α, ψ}	0	0.014	0.95	0	0	0	0	0	0	0.0023	0	0	0	0	0	0	0.032
{α, π}	0	0.67	0	0.33	0	0	0	0	0	0	0	0	0	0	0	0	0
{α, μ}	0	0.057	0	0	0.89	0	0	0	0	0	0	0	0	0	0	0	0.057
{α, λ}	0.0083	0.083	0	0	0	0.88	0	0	0	0	0	0	0	0	0	0	0.025
{α, θ}	0	0.019	0	0	0	0	0.93	0.0015	0	0	0.0073	0	0	0.0015	0	0	0.037
{α, η}	0	0.015	0	0	0.00015	0	0	0.97	0	0	0	0.00061	0	0	0	0.00031	0.012
{α, ε}	0	0.044	0	0	0	0	0	0	0.91	0	0.0037	0.0085	0	0	0.0012	0	0.028
{α, ε, ψ}	0	0	0.048	0	0	0	0	0	0	0.93	0	0	0	0	0	0	0.024
{α, ε, θ}	0	0	0	0	0	0	0.061	0	0	0	0.9	0	0	0	0	0	0.043
{α, ε, η}	0	0	0	0	0	0	0	0.015	0.015	0	0	0.96	0	0	0	0	0.015
{α, δ}	0	0.065	0	0	0	0	0	0	0	0	0	0	0.91	0	0.00071	0	0.024
{α, γ}	0	0.025	0	0	0	0	0	0.0002	0	0	0	0	0	0.94	0	0.0002	0.031
{α, β}	0	0.09	0	0	0	0	0	0	0	0	0	0	0.00056	0	0.88	0.00056	0.025
{}	0.00028	0.28	0.00019	0	0.004	9.3e-05	0.0014	0.031	0.00037	0	0	0	0.00074	0.0065	0.0021	0.66	0.014
{other}	0	0.004	0.0025	0	0.0015	0.0012	0.0037	0.0081	0.0038	0.0012	0.0012	0.00038	0.0033	0.03	0.004	0.00058	0.93

ФПМИ Альфа Банк

GEEK BATTLE

Предпочтение различных ОКВЭД к кластеру



ФПМИ Альфа Банк GEEK BATTLE

Другие интересные замечания из EDA

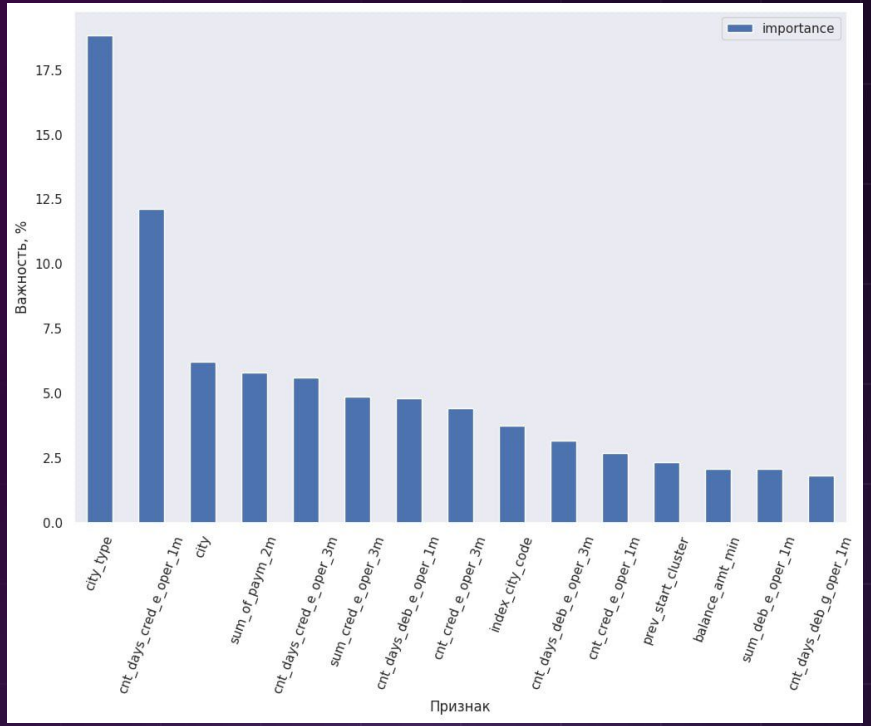
- У **новых** пользователей свой **ограниченный набор кластеров**
- Среднее время до получения ОГРН в трэйн ~0, в тест > 0 => **train идет после test** по времени
- Город, индекс города и тип города, **перекрывают пропуски** в данных друг у друга. Выбрали **оставить** все **несмотря на** кажущуюся **дублируемость** информации.
- В крупных городах России распределение категорий доходности отличается от малых.

ФПМИ Альфа Банк

GEEK BATTLE

Train vs Test

- Сильные различия в признаках указывающих на прошедшее время с момента.
- В тесте есть новые города
- Числовые признаки испытывают небольшой сдвиг по времени(Операция “E” изменяется сильнее всего, возможно в купе с типом города)



ФПМИ Альфа Банк GEEK BATTLE

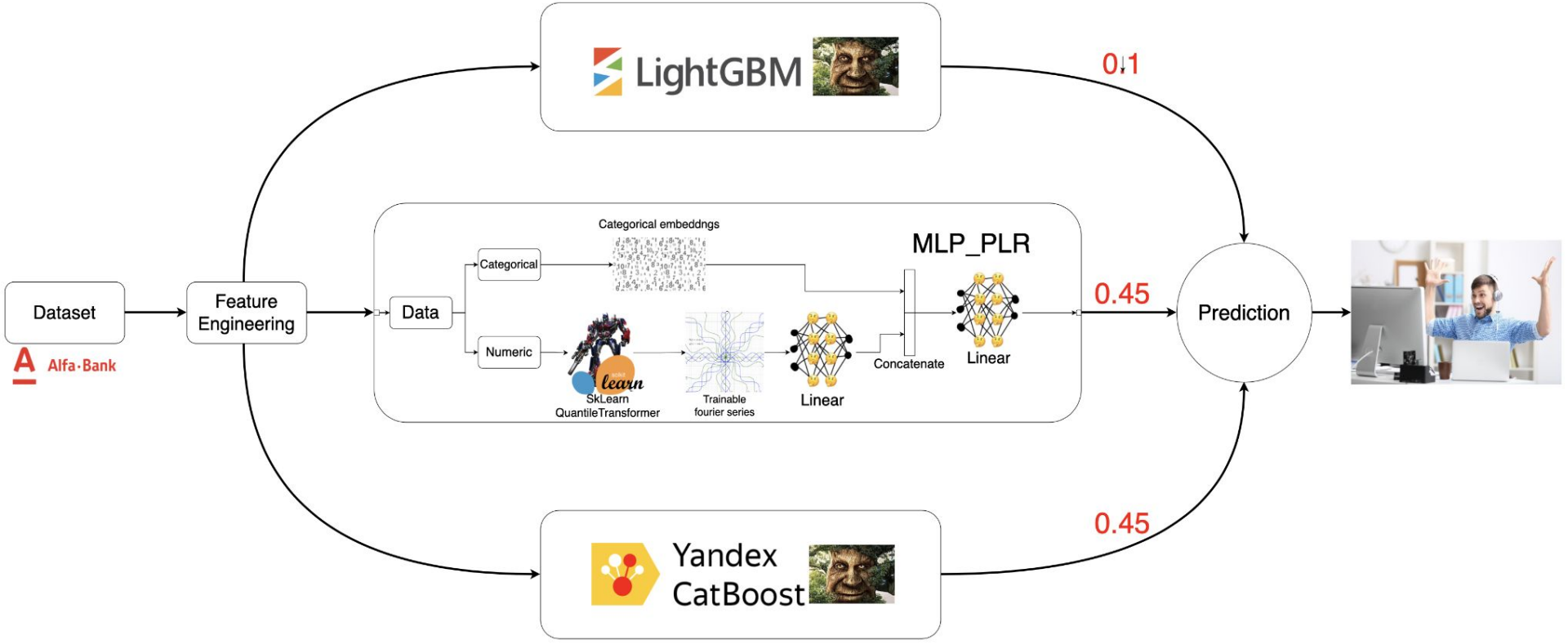
Неудачные эксперименты

- Добавление истории. RNN.
- Tied Embeddings.
- Комбинирование признаков.
- Доменная адаптация(Предобучение)
- Добавление весов примерам

ФПМИ Альфа Банк

GEEK BATTLE

Модель



ФПМИ Альфа Банк

GEEK BATTLE

Конкурентные Преимущества.

Высокая точность

Легко встроить и поддерживать

ФПМИ Альфа Банк

GEEK BATTLE

Конкурентные Преимущества.

Легкая масштабируемость

- Легко применить к большему количеству данных
- Борьба с ковариационным сдвигом
- Извлечение пользы из неразмеченного набора данных

ФПМИ Альфа Банк

GEEK BATTLE

Над проектом работали

Дмитрий Харчев

ML-разработчик

@KharDim08

Тохчуков Данил

ML-разработчик

@makriot

Черемискин Егор

ML-разработчик

@he_is_already_here

Лапиков Владислав

ML-разработчик

@What_is_Love_iss

Кадченко Иван

Аналитик

@KadchenkoIE