

Кредитный риск

ПРОЕКТ ПО РИСК-МЕНЕДЖМЕНТУ

Александров Никита

Бажмин Сергей

Гармаева Дари-Ханда

Дранга Даниил

Данные

КРАТКИЙ ОБЗОР ДАННЫХ:

Предоставлены данные по выданным кредитам на различные цели со сроком 36 и 60 месяцев:

- обучающая выборка (1M)
- валидационная выборка (0.25M)



Данные

ОБРАБОТКА ДАННЫХ:

- Убрали признаки с долей NaNов > 20%
- Убрали шумные признаки ['id', 'policy_code', 'zip_code']
- Заполнили NaNы:
 - скользящим средним,
 - самым популярным значением

```
le[b]()))var c=function(b){this.element=a(b)};...  
opdown-menu");d=b.data("target");if(d||(d=b.attr("href"),d=d&&d.replace(/.*(?=#[^\s"]+)/g,"#")){  
st a"),f=a.Event("hide.bs.tab",{relatedTarget:b[0]}),g=a.Event("show.bs.tab",{relatedTarget:e[0]}),  
FaultPrevented()){var h=a(d);this.activate(b.closest("li"),c),this.activate(h,h.parent(),function()  
rigger({type:"shown.bs.tab",relatedTarget:e[0]}))}}},c.prototype.activate=function(b,d,e){func  
u > .active").removeClass("active").end().find("[data-toggle="tab"]').attr("aria-expanded",!1),  
ia-expanded",!0),h?(b[0].offsetWidth,b.addClass("in")):b.removeClass("fade"),b.parent(".dropdown  
().find("[data-toggle="tab"]').attr("aria-expanded",!0),e&&e())var g=d.find("> .active"),h=e&&  
le")||!d.find("> .fade").length;g.length&&h?g.one("bsTransitionEnd",f).emulateTransitionEnd  
;var d=a.fn.tab;a.fn.tab=b,a.fn.tab.Constructor=c,a.fn.tab.noConflict=function(){return a.fn.t  
"show");a(document).on("click.bs.tab.data-api",[data-toggle="tab"],e).on("click.bs.tab.data  
se strict";function b(b){return this.each(function(){var d=a(this),e=d.data("bs.affix"),f="ob  
=typeof b&&b[0]()))var c=function(b,d){this.options=a.extend({},c.DEFAULTS,d),this.$target=a  
!,a.proxy(this.checkPosition,this)).on("click.bs.affix.data-api",a.proxy(this.checkPositionWi  
null,this.pinnedOffset=null,this.checkPosition());c.VERSION="3.3.7",c.RESET="affix affix-top  
State=function(a,b,c,d){var e=this.$target.scrollTop(),f=this.$element.offset(),g=this.$targ  
"bottom"==this.affixed)return null!=c?!(e+this.unpin<=f.top)&&"bottom":!(e+g<=a-d)&&"bottom"  
!=c&&e<="top":null!=d&&i+j>=a-d&&"bottom"},c.prototype.getPinnedOffset=function(){if  
:RESET).addClass("affix");var a=this.$target.scrollTop(),b=this  
WithEventLoop=function(){setTimeout(a,0)};a.height().d=this
```

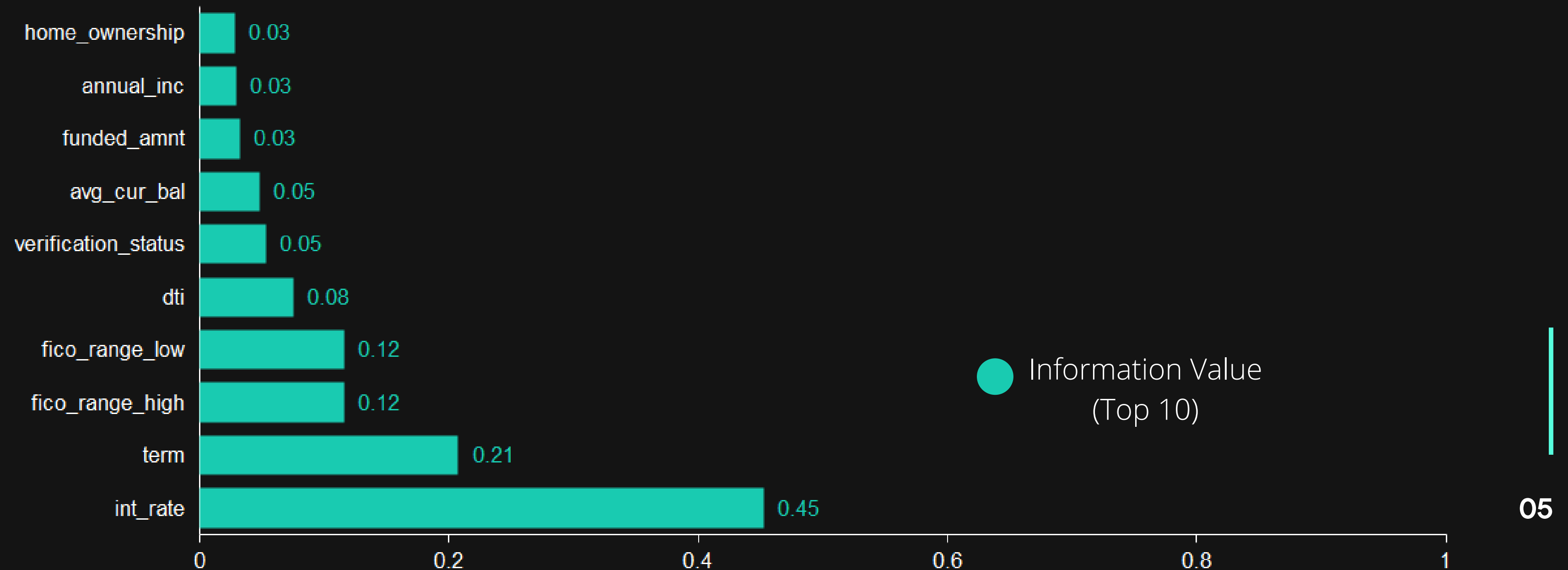

Интерпретируемый скоринг

ОДНОФАКТОРНЫЙ АНАЛИЗ ПЕРЕМЕННЫХ: INFORMATION VALUE

- 1) Непрерывные данные разбили на интервалы и превратили их в ординальные с помощью функции `MonotonicBinning` из библиотеки `xverse`, а категориальные с помощью `category_encoders.woe`
- 2) Для каждой категории вычислили `Weight of Evidence` и, соответственно, посчитали `Information Value` по всем признакам

Интерпретируемый скоринг

ОДНОФАКТОРНЫЙ АНАЛИЗ ПЕРЕМЕННЫХ: INFORMATION VALUE

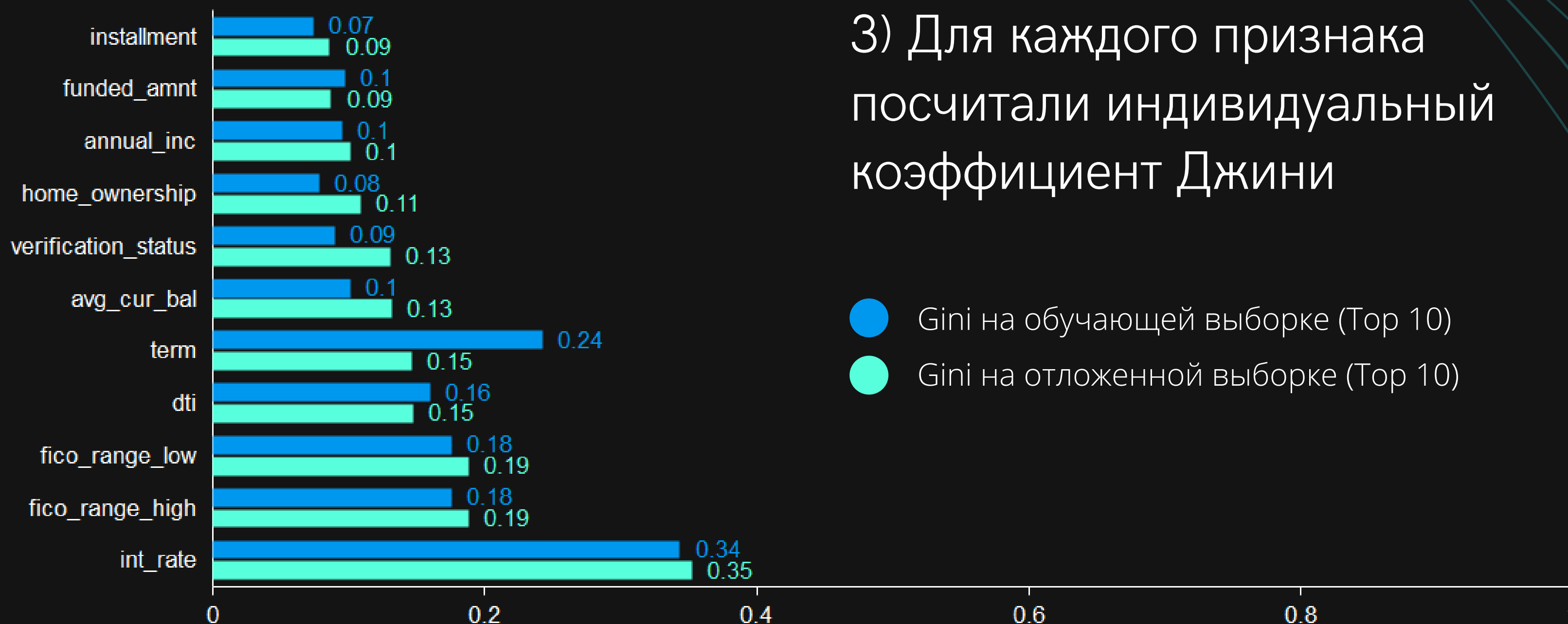


Интерпретируемый скоринг

ОДНОФАКТОРНЫЙ АНАЛИЗ ПЕРЕМЕННЫХ: КОЭФФИЦИЕНТ ДЖИНИ

3) Для каждого признака посчитали индивидуальный коэффициент Джини

- Gini на обучающей выборке (Top 10)
- Gini на отложенной выборке (Top 10)



Интерпретируемый скоринг

Логистическая регрессия
с последовательным
включением и
исключением
отобранных переменных

В итоге в конечную выборку не
вошли следующие признаки:

['application_type', 'tax_liens',
'pub_rec',
'disbursement_method']

backward:

0.39981

0.40019

0.40179

0.40207

forward:

0.35179

0.38618

0.39275

0.40207

Интерпретируемый скоринг

ОТБОР ПЕРЕМЕННЫХ

На основе коэффициент Джини, Information Value и здравого смысла были отобраны следующие 11 признаков для построения модели скоринга:

```
LR_FEATURES =  
['int_rate', 'fico_range_high', 'fico_range_low',  
'dti', 'term', 'avg_cur_bal', 'verification_status',  
'home_ownership', 'annual_inc', 'funded_amnt',  
'installment']
```

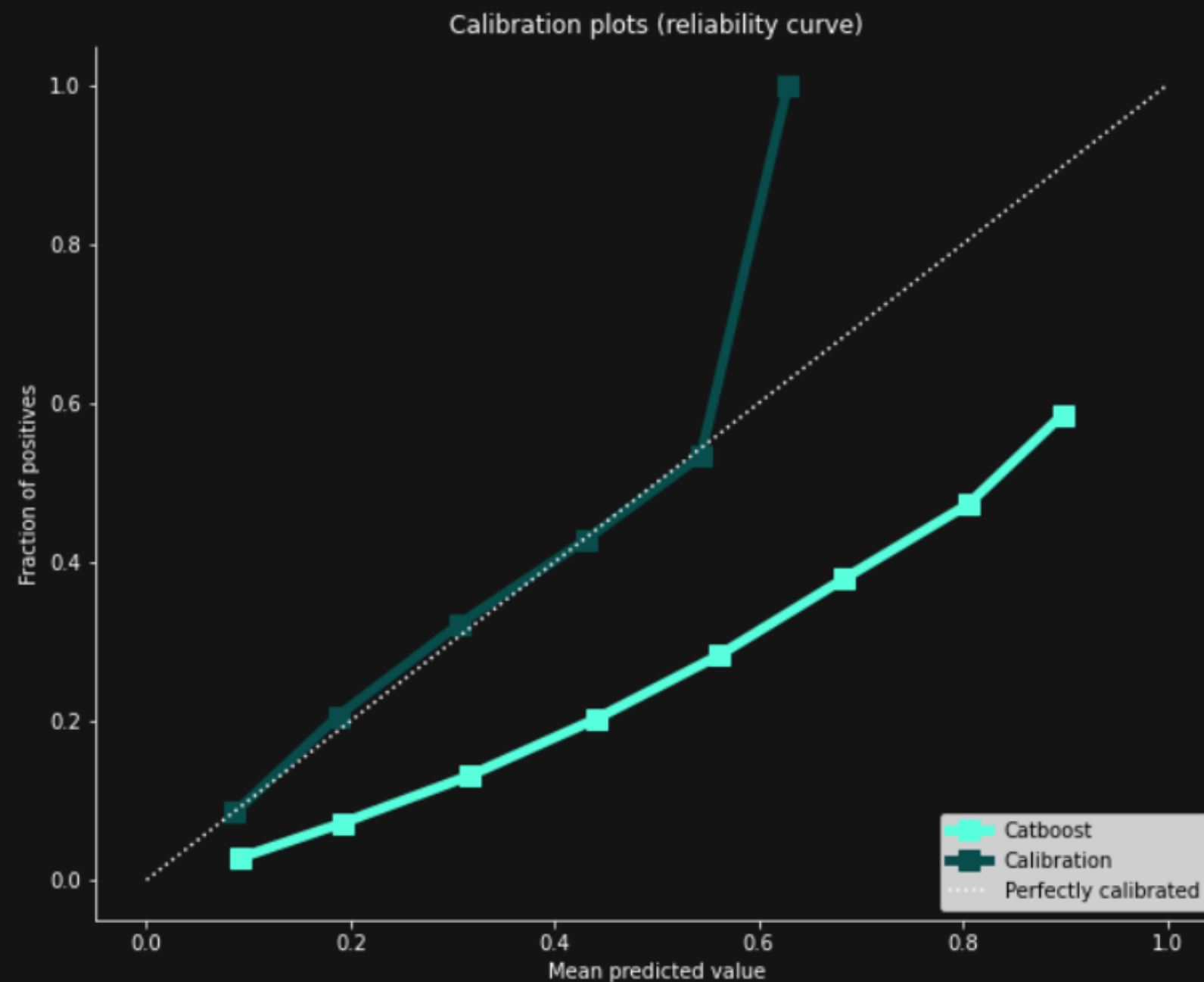
Обучающая выборка: 0.41

Отложенная выборка: 0.39

Валидационная выборка: 0.37

Альтернативная модель

МОДЕЛЬ МАШИННОГО ОБУЧЕНИЯ - CATBOOST CLASSIFIER



- Catboost+Калибровка
- Коэффициент Джини на валидационной выборке 0.3947

Оценка внутренних рейтингов

Высший рейтинг

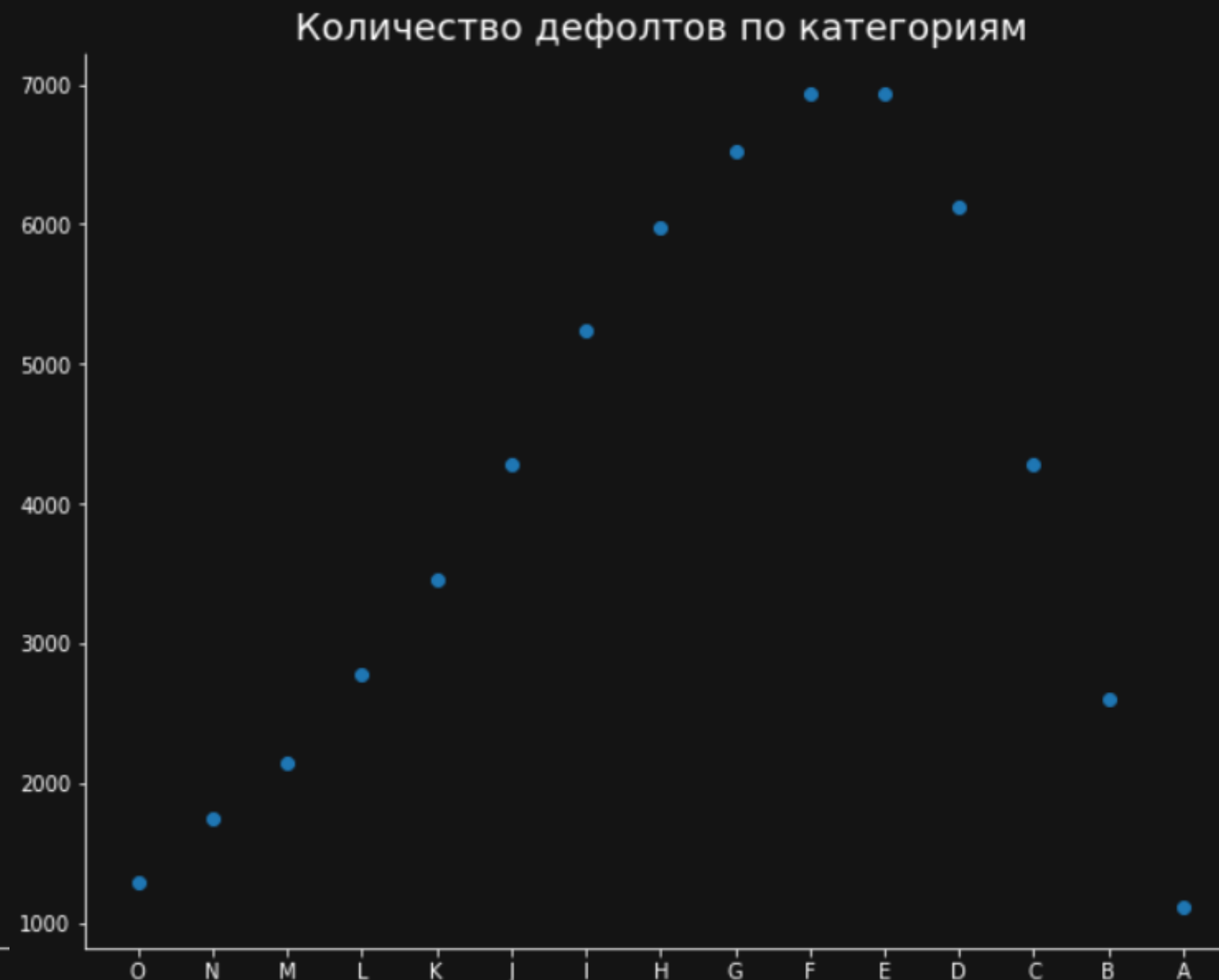
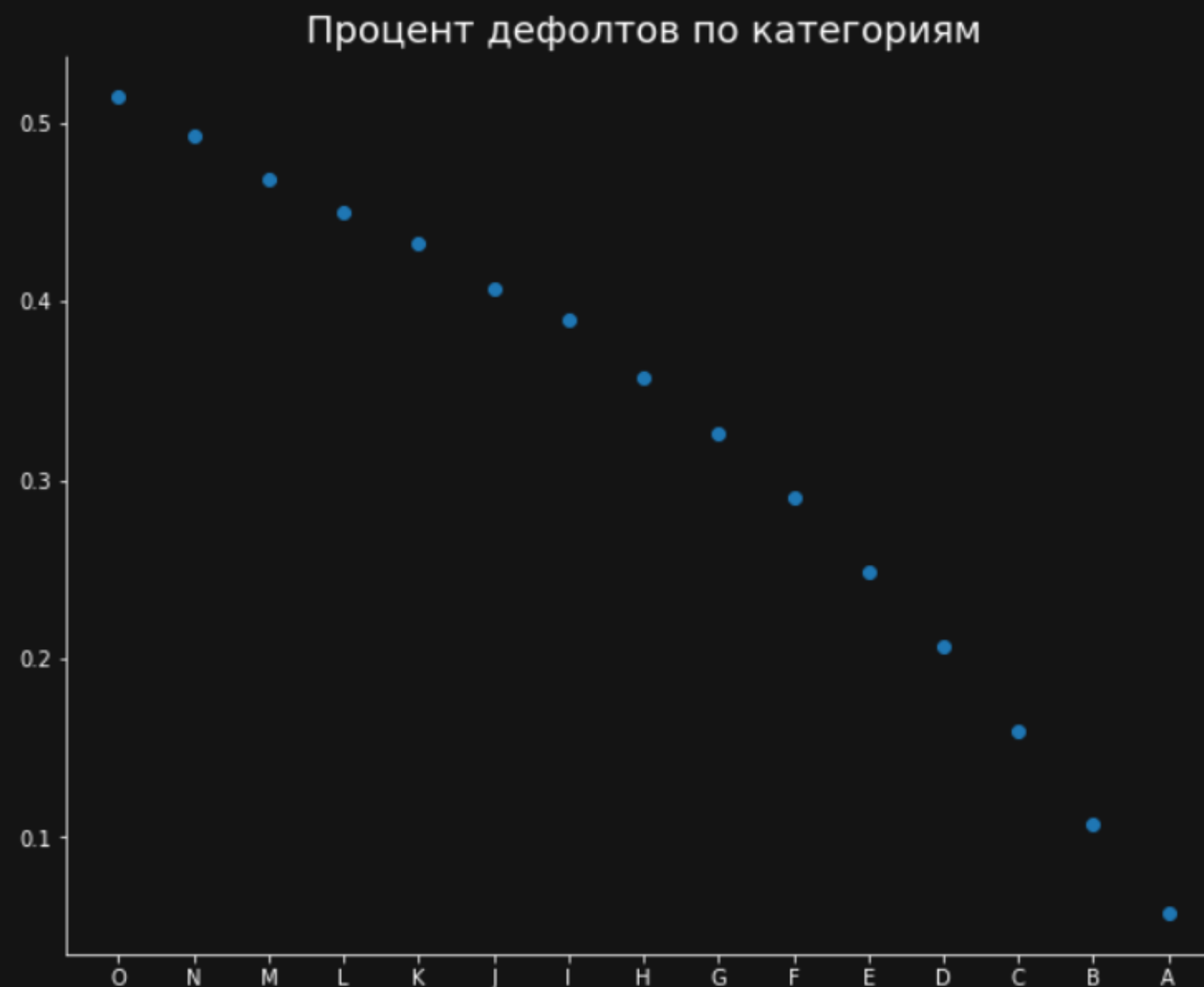
●	●	●
A	B	C
●	●	●
D	E	F
●	●	●
G	H	I
●	●	●
J	K	L
●	●	●
M	N	O

Низший рейтинг

Для обеих моделей - LogReg и Catboost был построен внутренний рейтинг, для этого взяли не-дефолтные категории с вероятностью дефолта меньше 50% с начальным количеством бинов равным 15

Оценка внутренних рейтингов

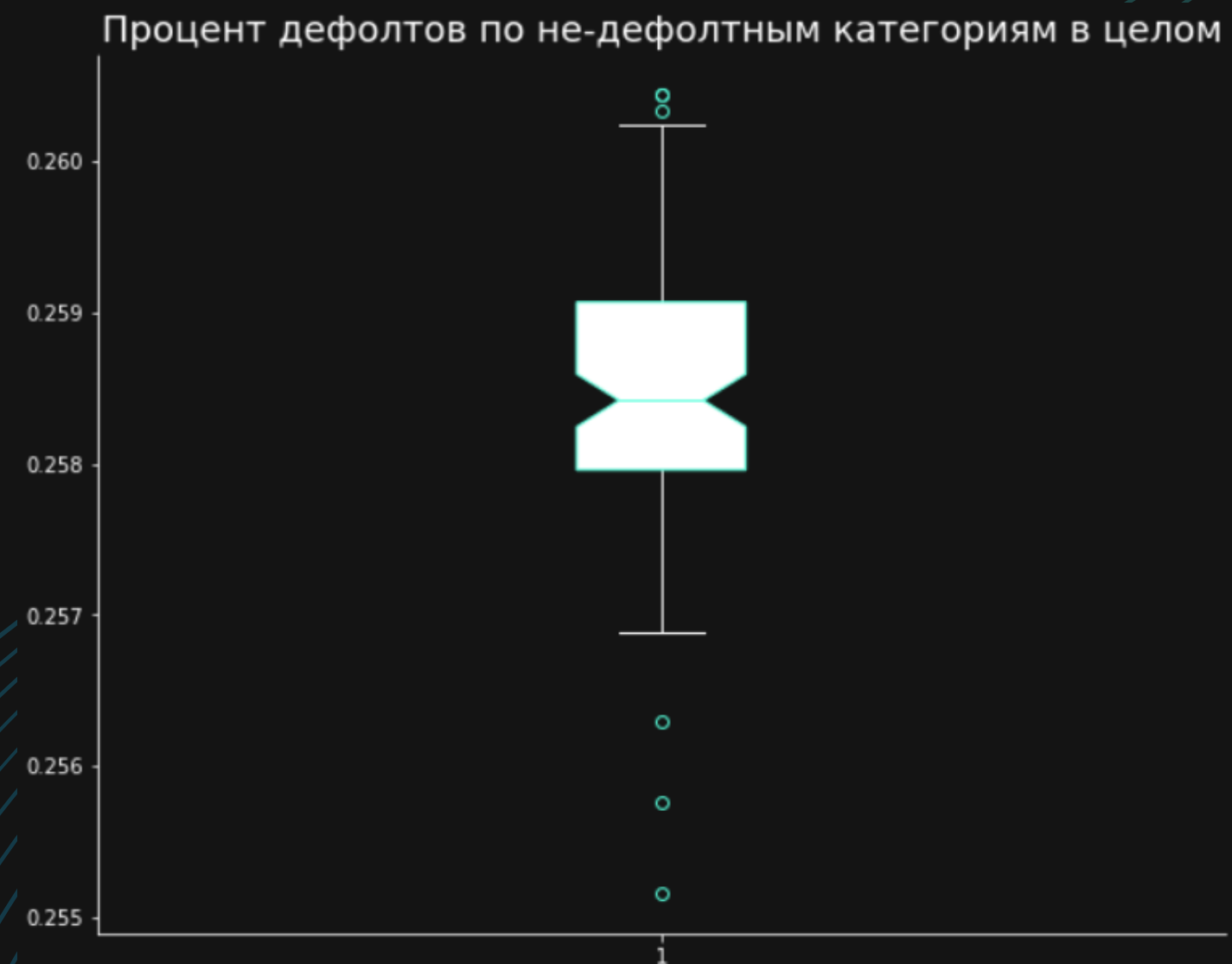
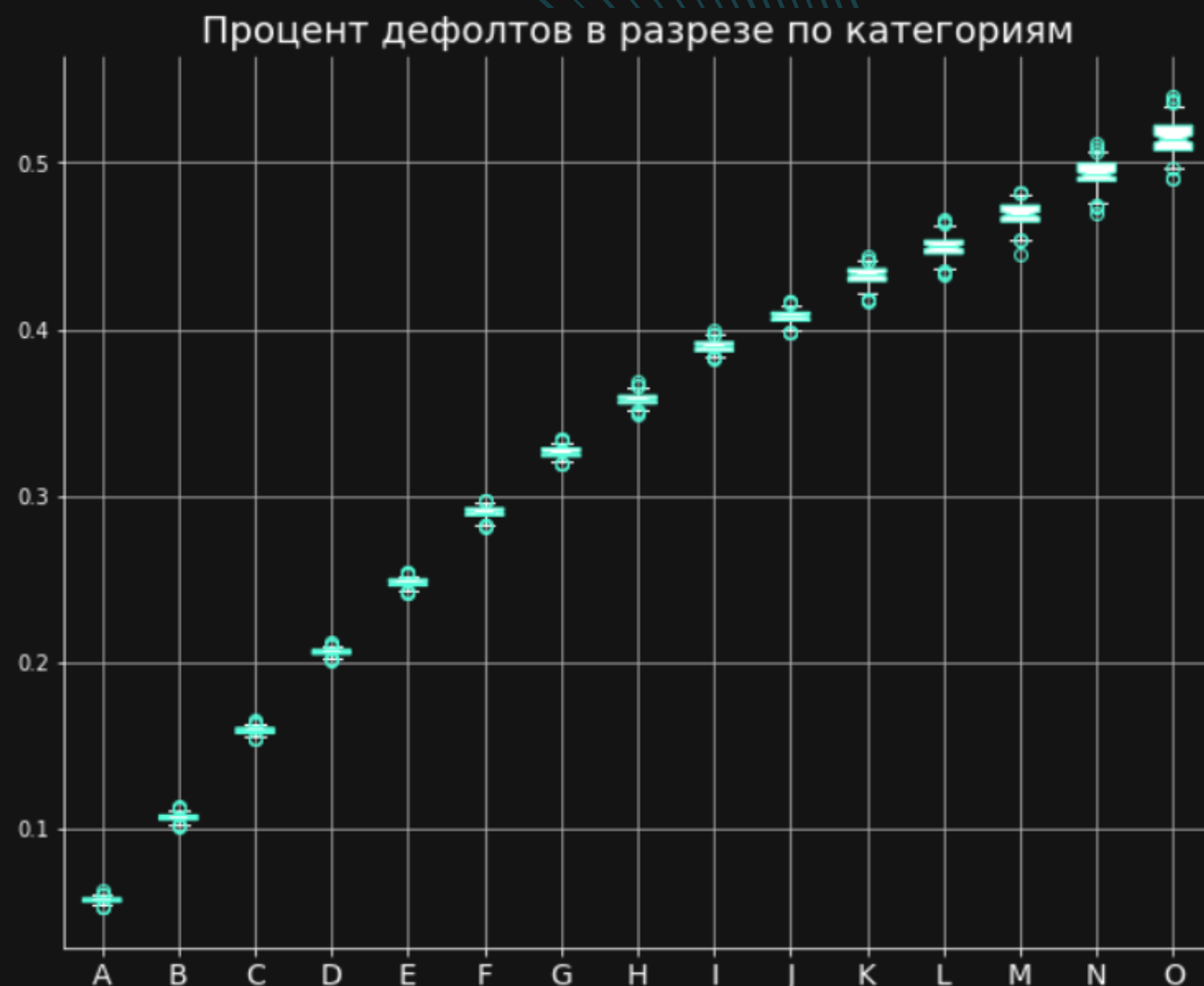
МОДЕЛЬ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ



Оценка внутренних рейтингов

МОДЕЛЬ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ

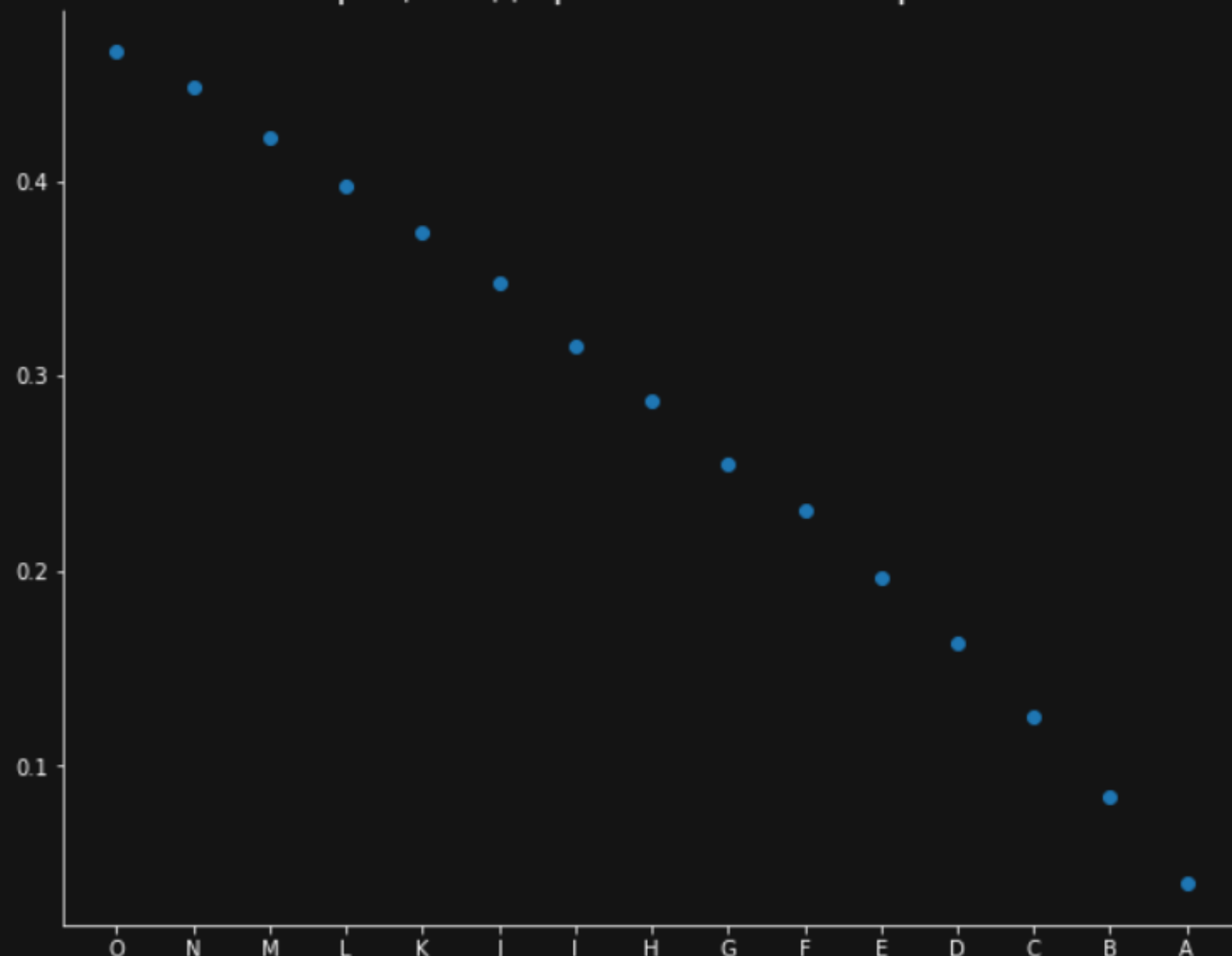
Доля дефолтов во всех не-дефолтных категориях на валидационной выборке:
0.2585



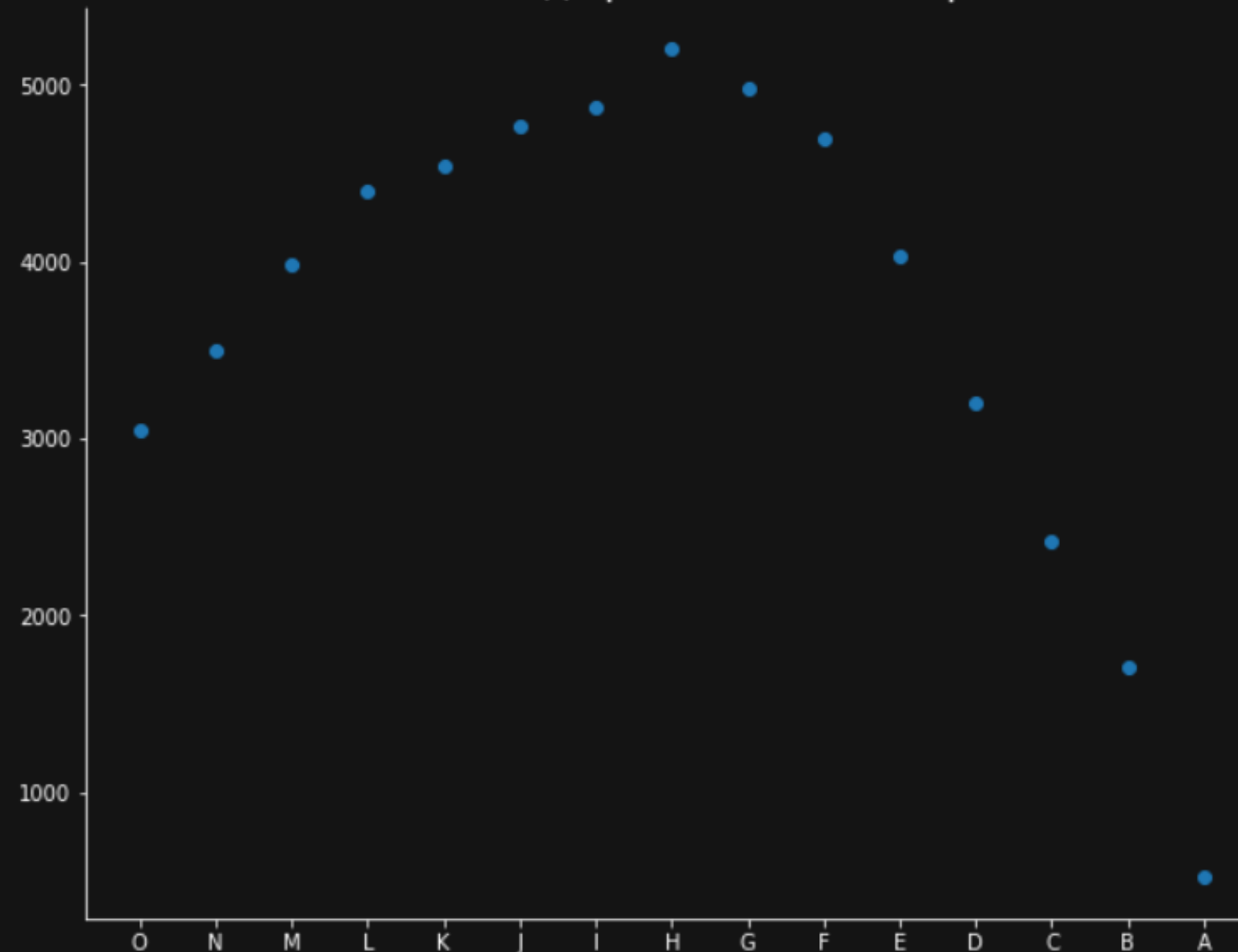
Оценка внутренних рейтингов

МОДЕЛЬ МАШИННОГО ОБУЧЕНИЯ CATBOOST

Процент дефолтов по категориям



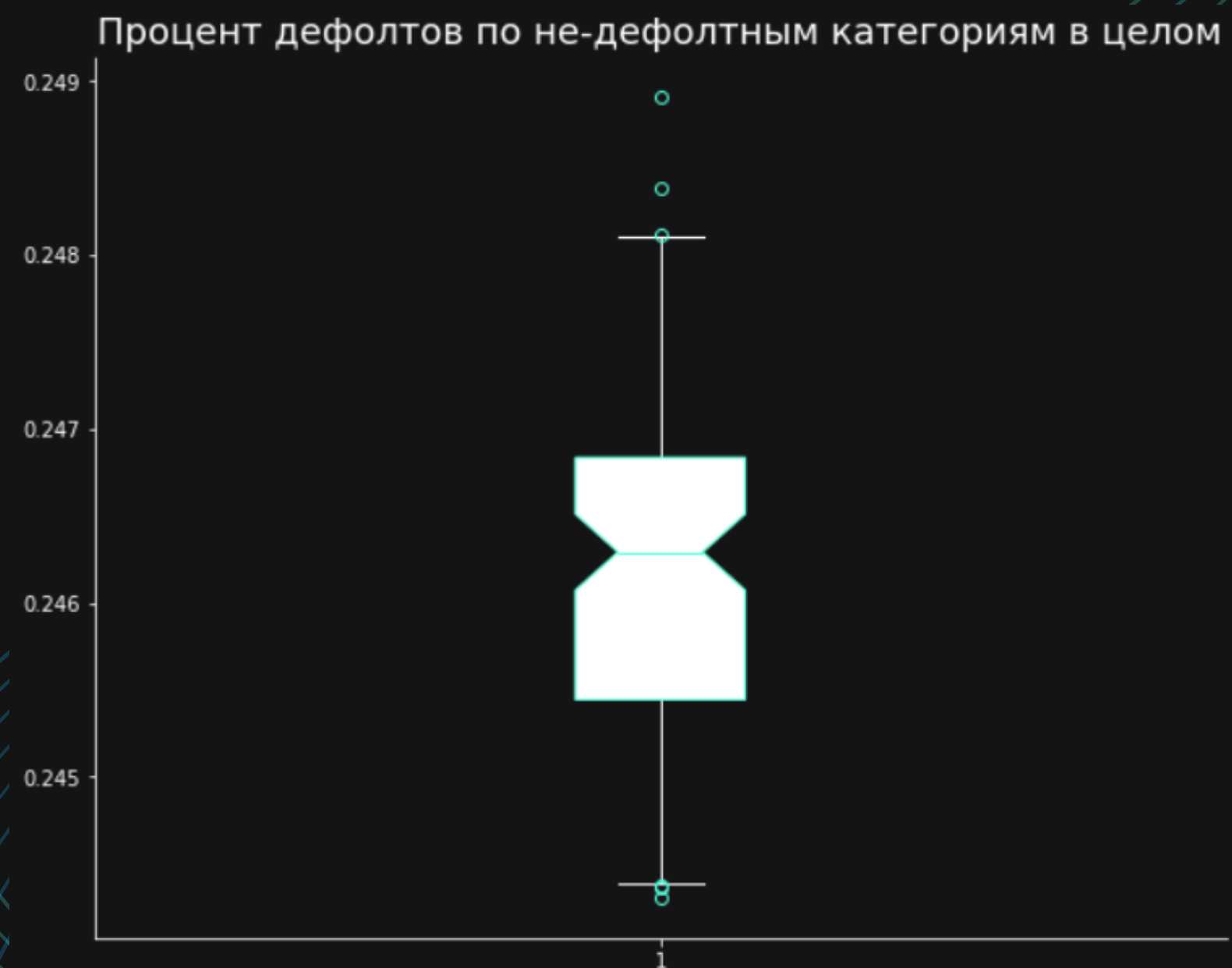
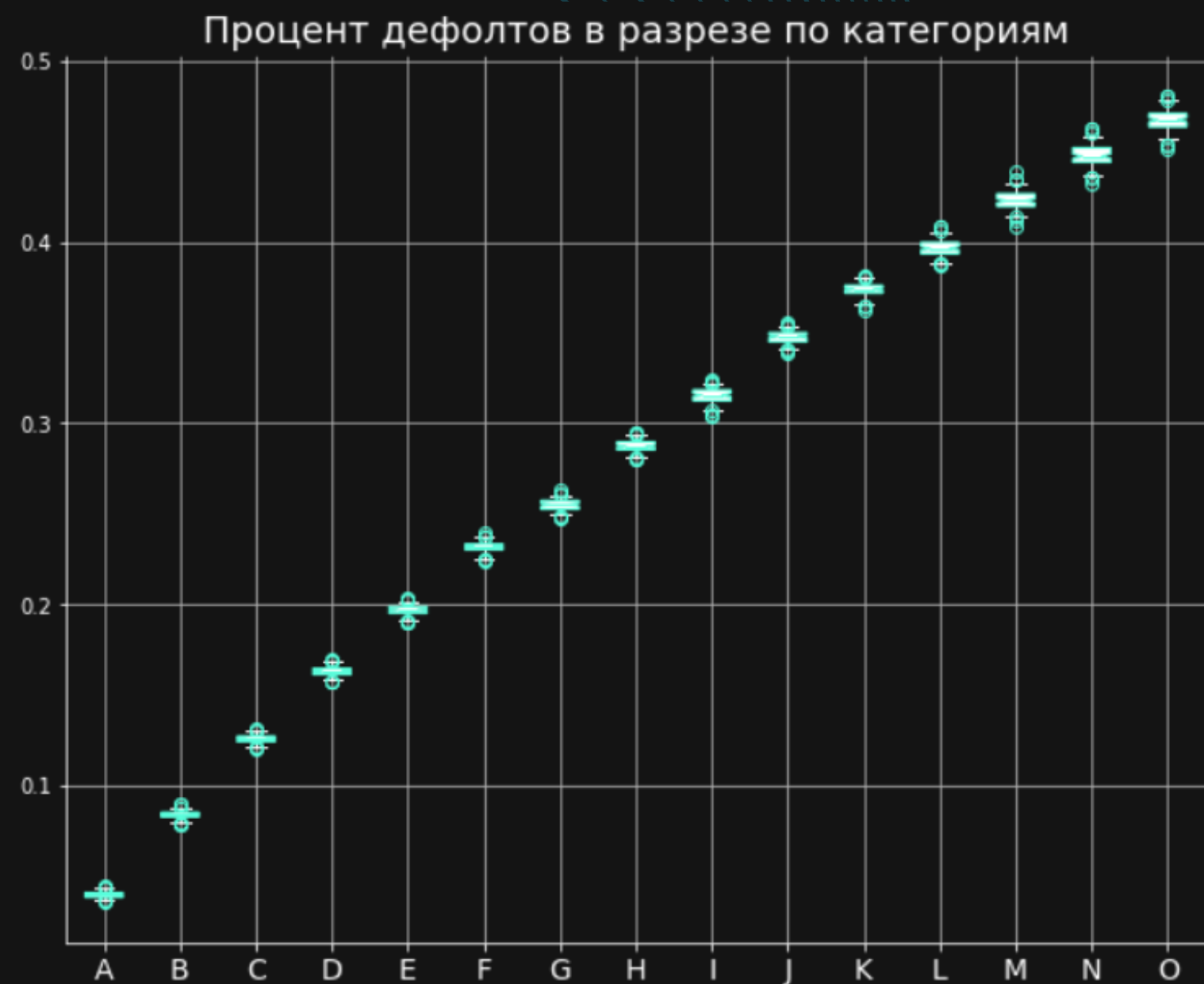
Количество дефолтов по категориям



Оценка внутренних рейтингов

МОДЕЛЬ МАШИННОГО ОБУЧЕНИЯ CATBOOST

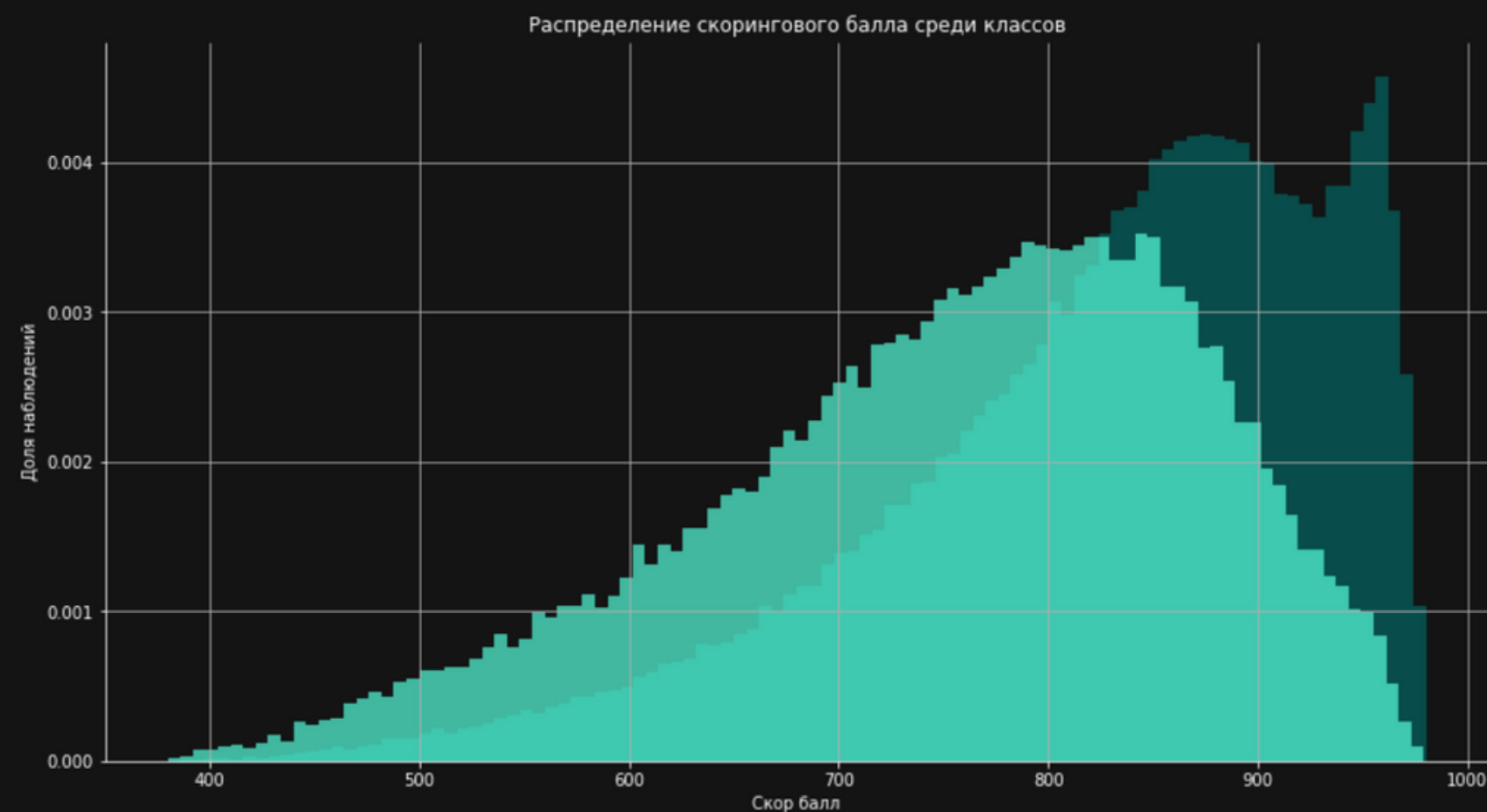
Доля дефолтов во всех не-дефолтных категориях на валидационной выборке:
0.2462



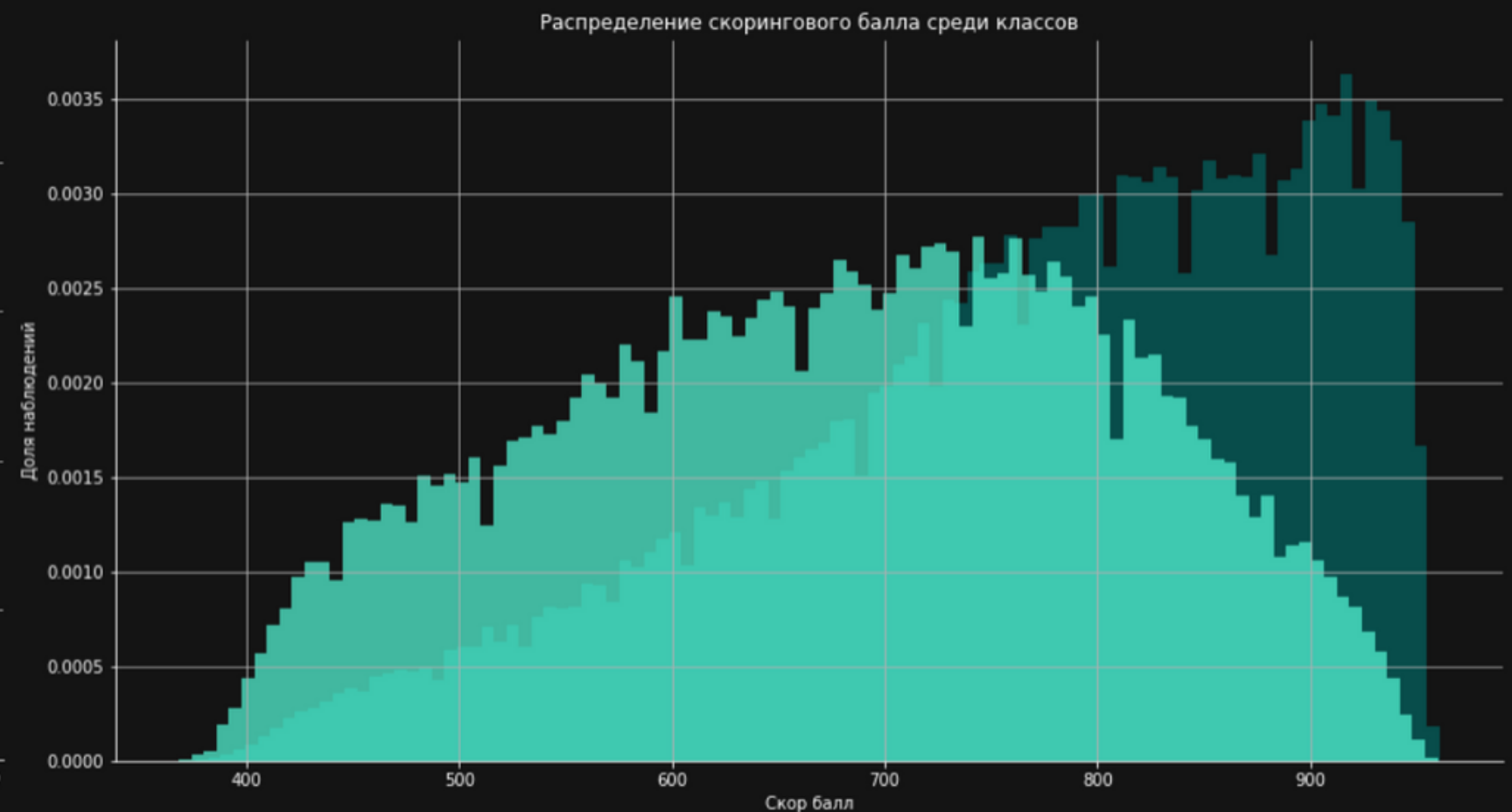
Валидация моделей

ГИСТОГРАММЫ РАСПРЕДЕЛЕНИЯ СКОРИНГОВЫХ БАЛЛОВ ДЛЯ ДЕФОЛТЕРОВ И НЕ-ДЕФОЛТЕРОВ: МОДЕЛЬ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ И CATBOOST

● Bads ● Goods

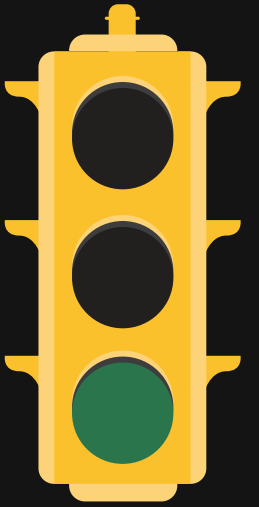


LogReg

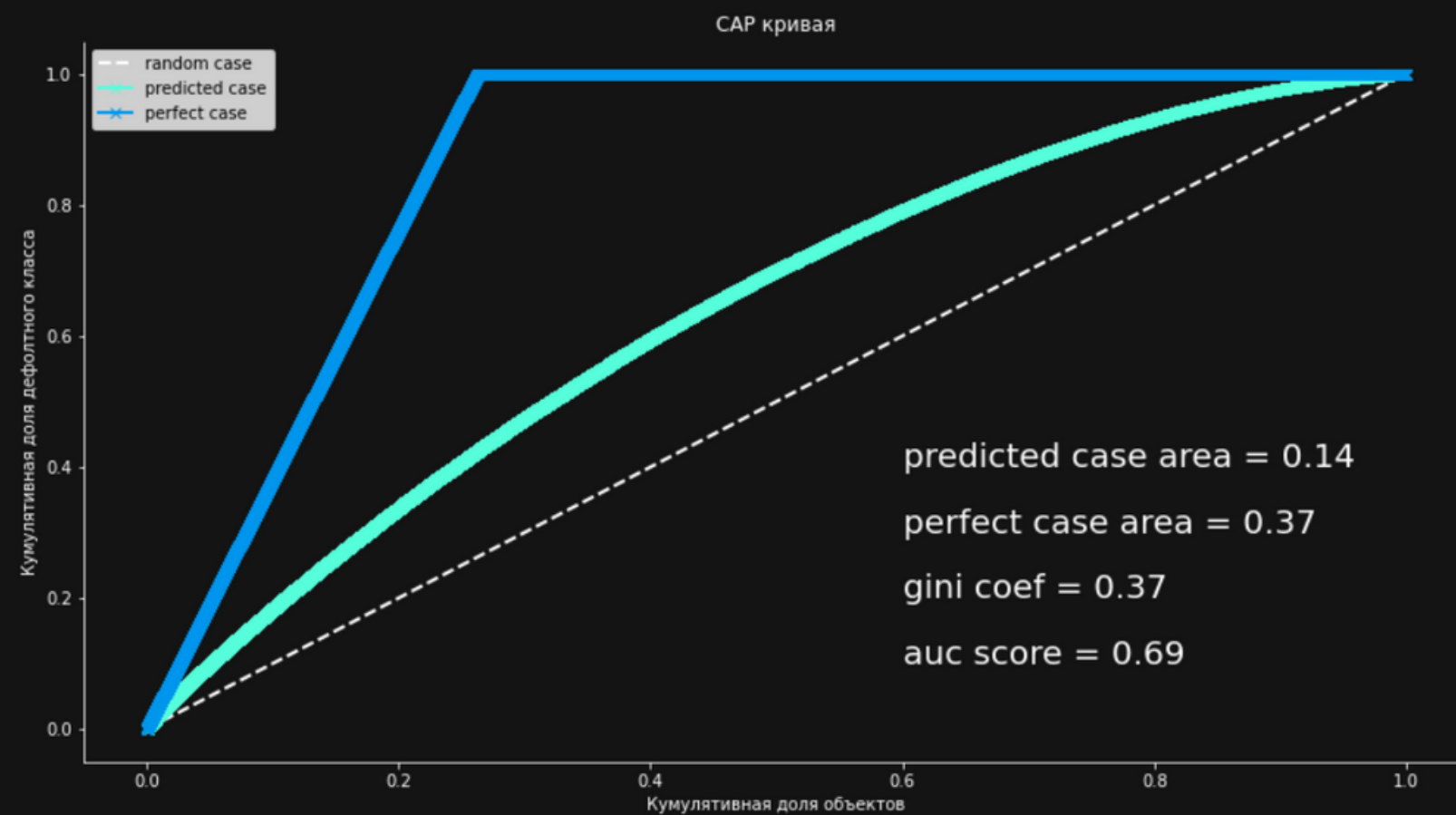


Catboost

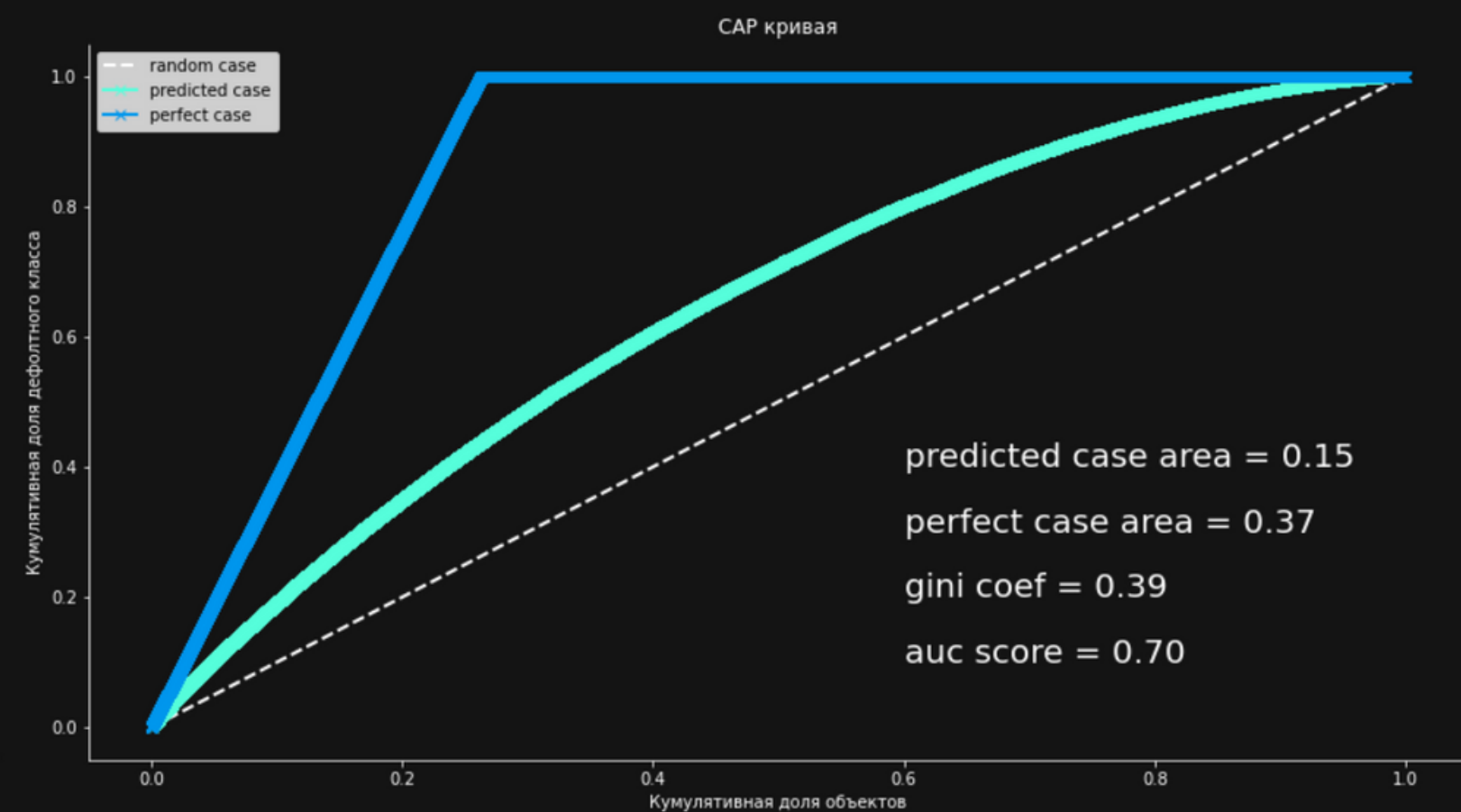
Валидация моделей



САР-КРИВАЯ



LogReg

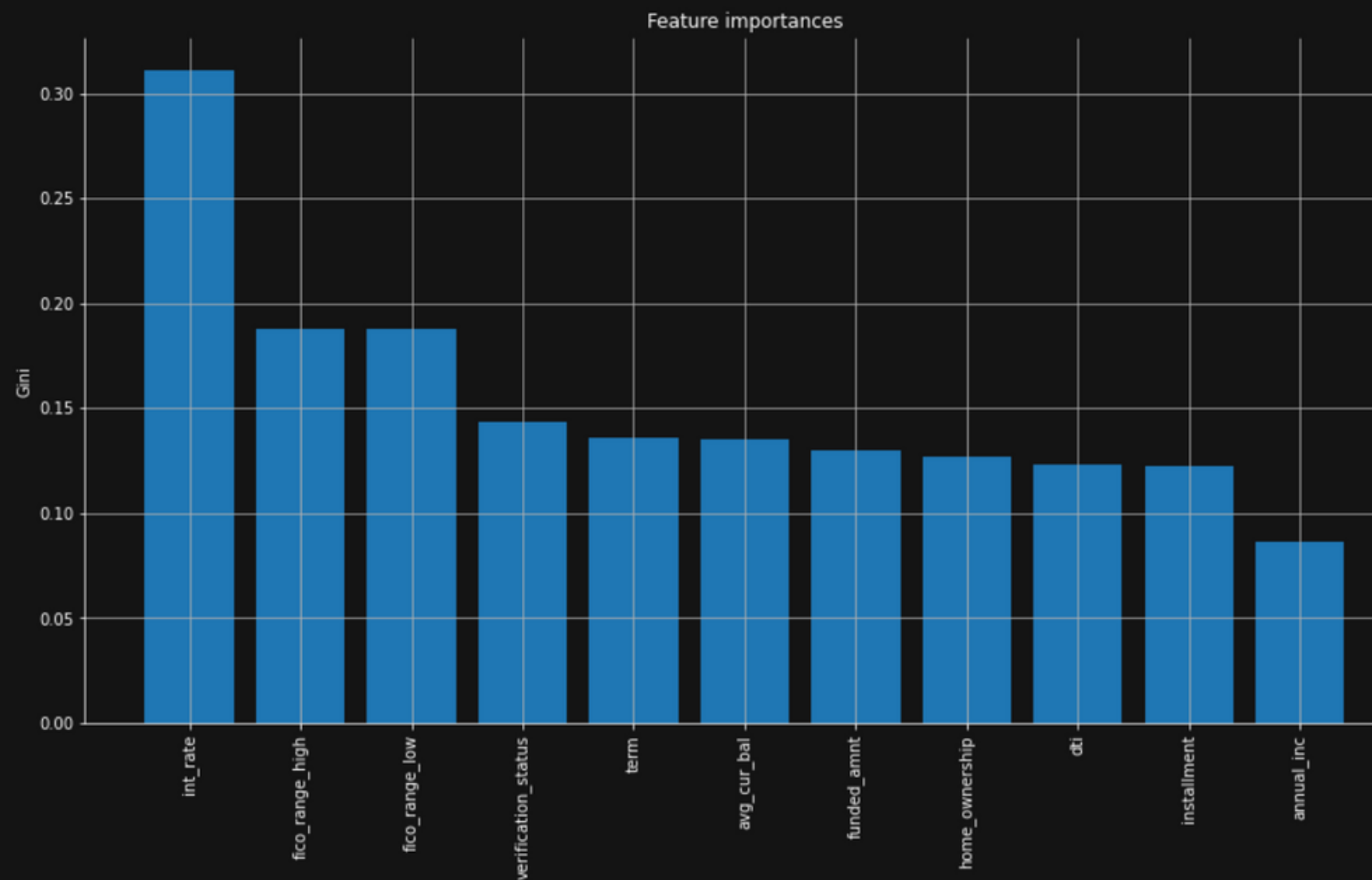


Catboost

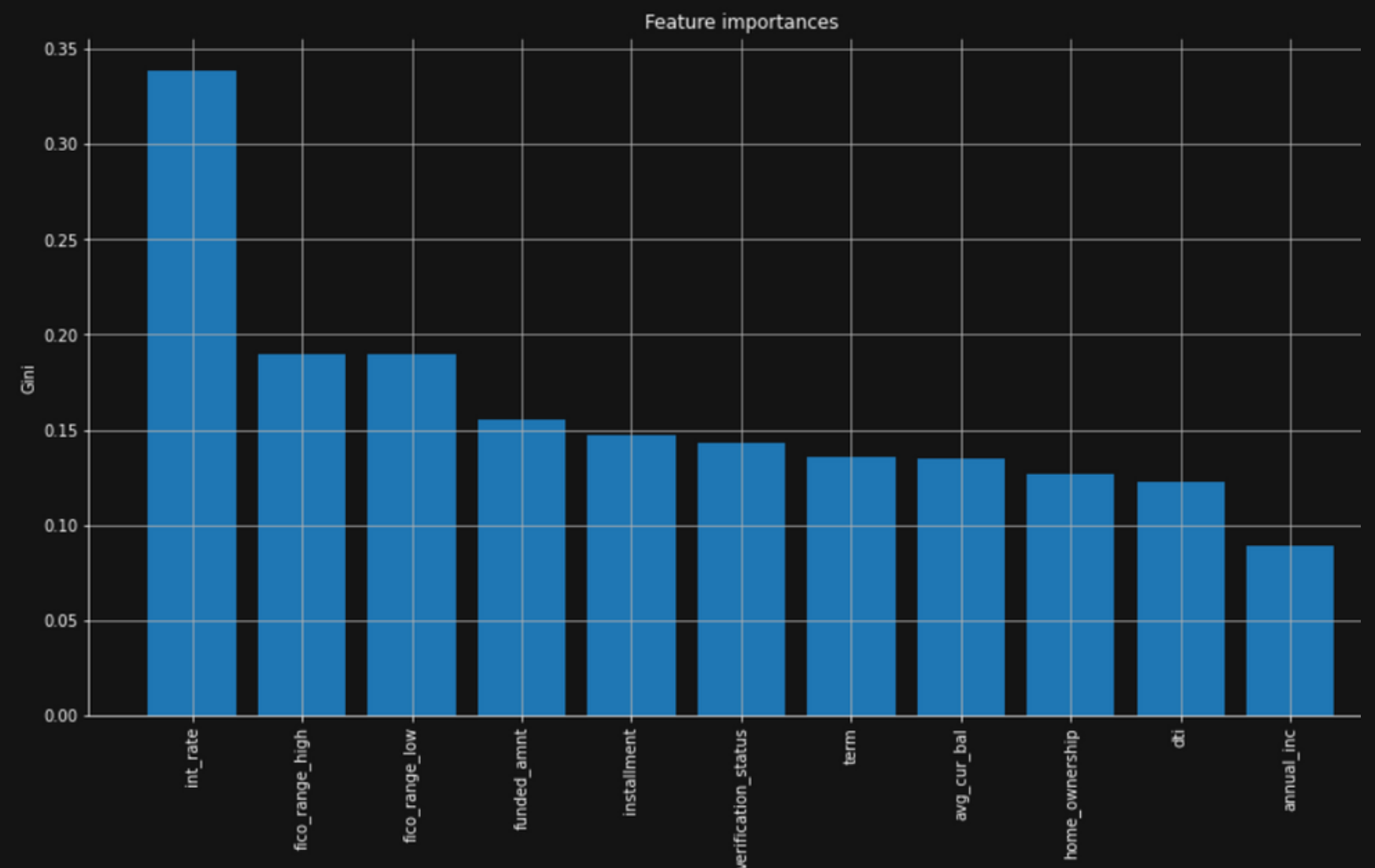
Валидация моделей



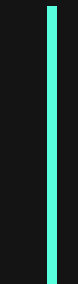
ИНДИВИДУАЛЬНАЯ ЗНАЧИМОСТЬ ПО КОЭФФИЦИЕНТУ ДЖИНИ



LogReg

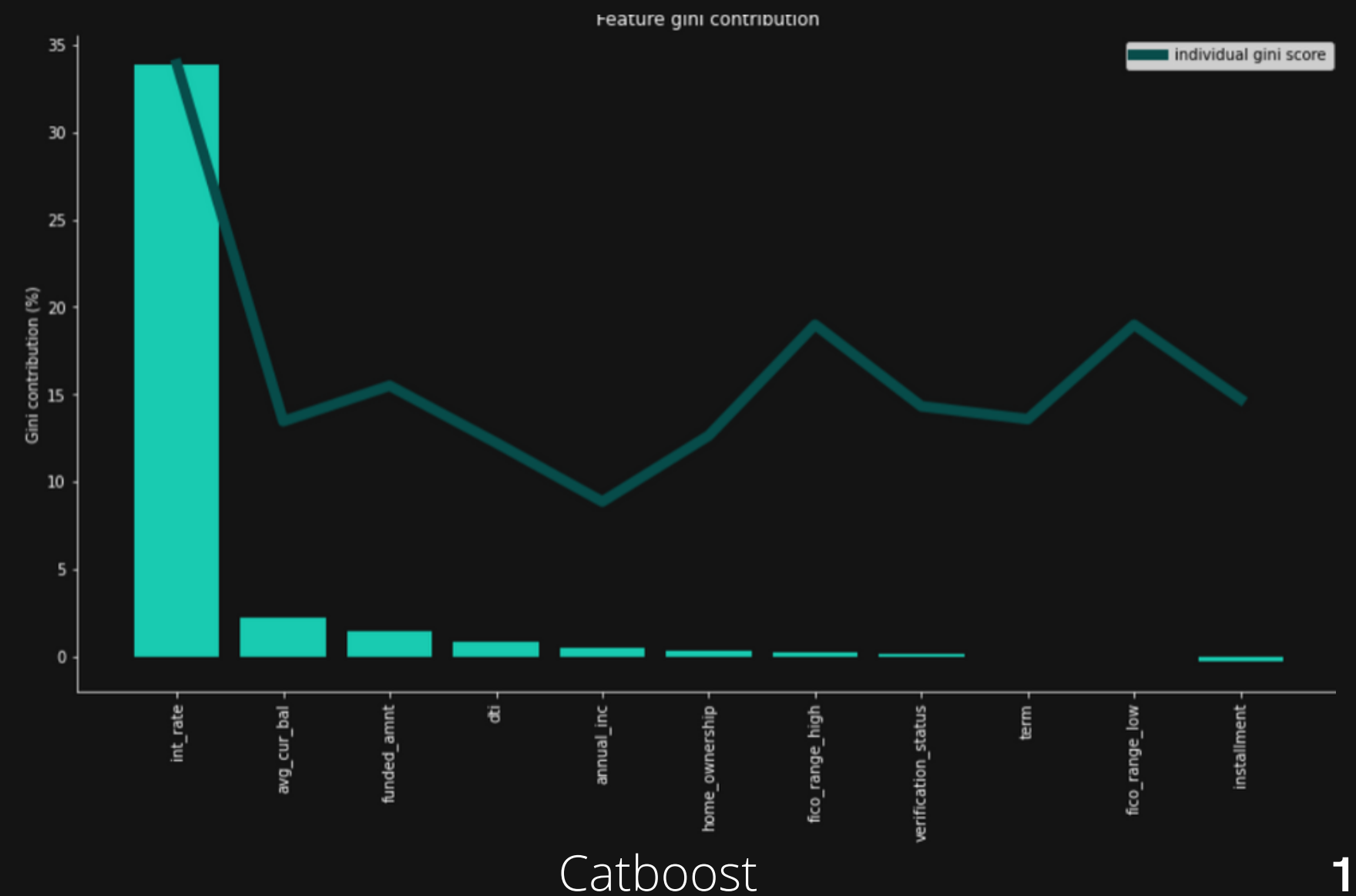
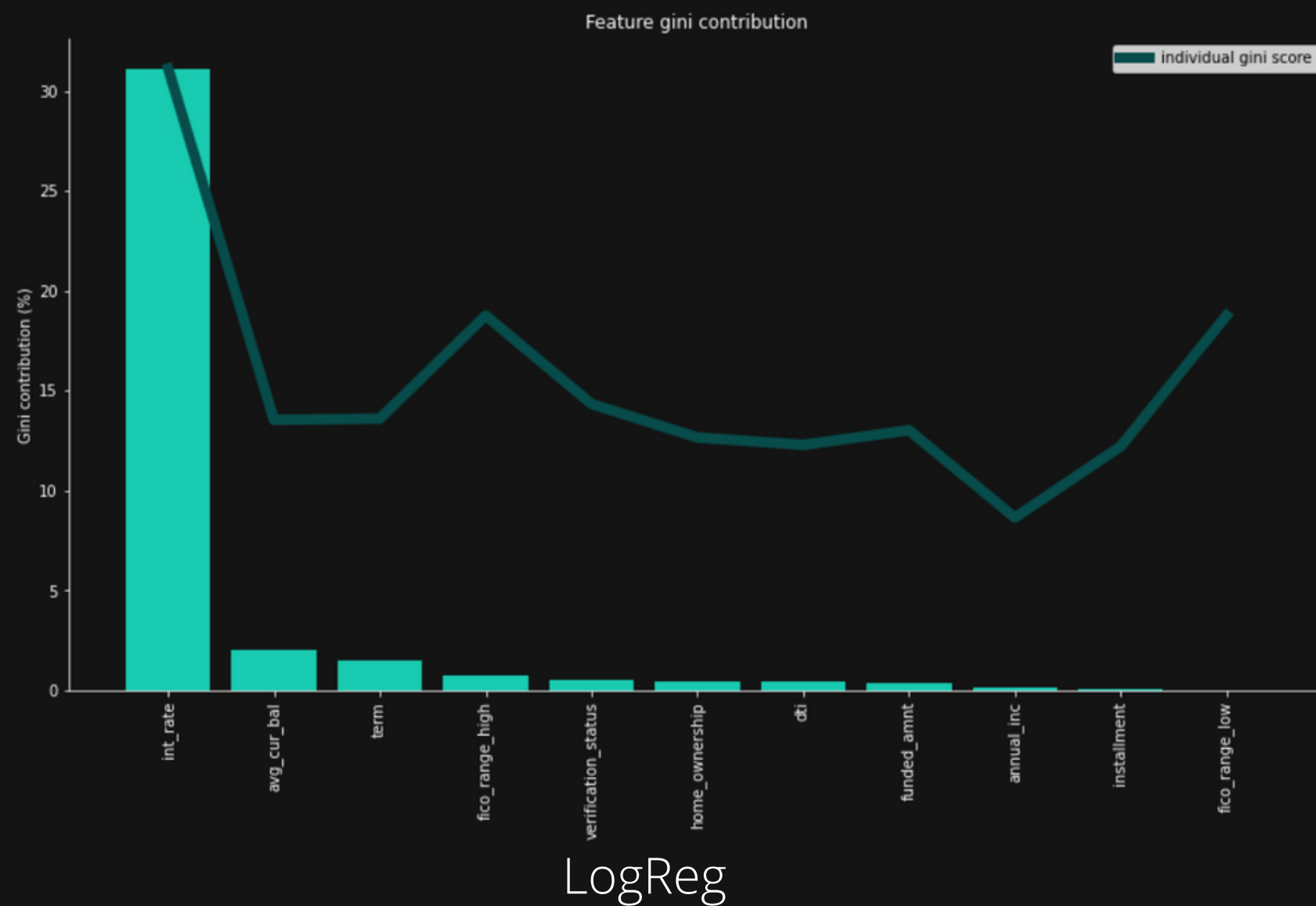


Catboost



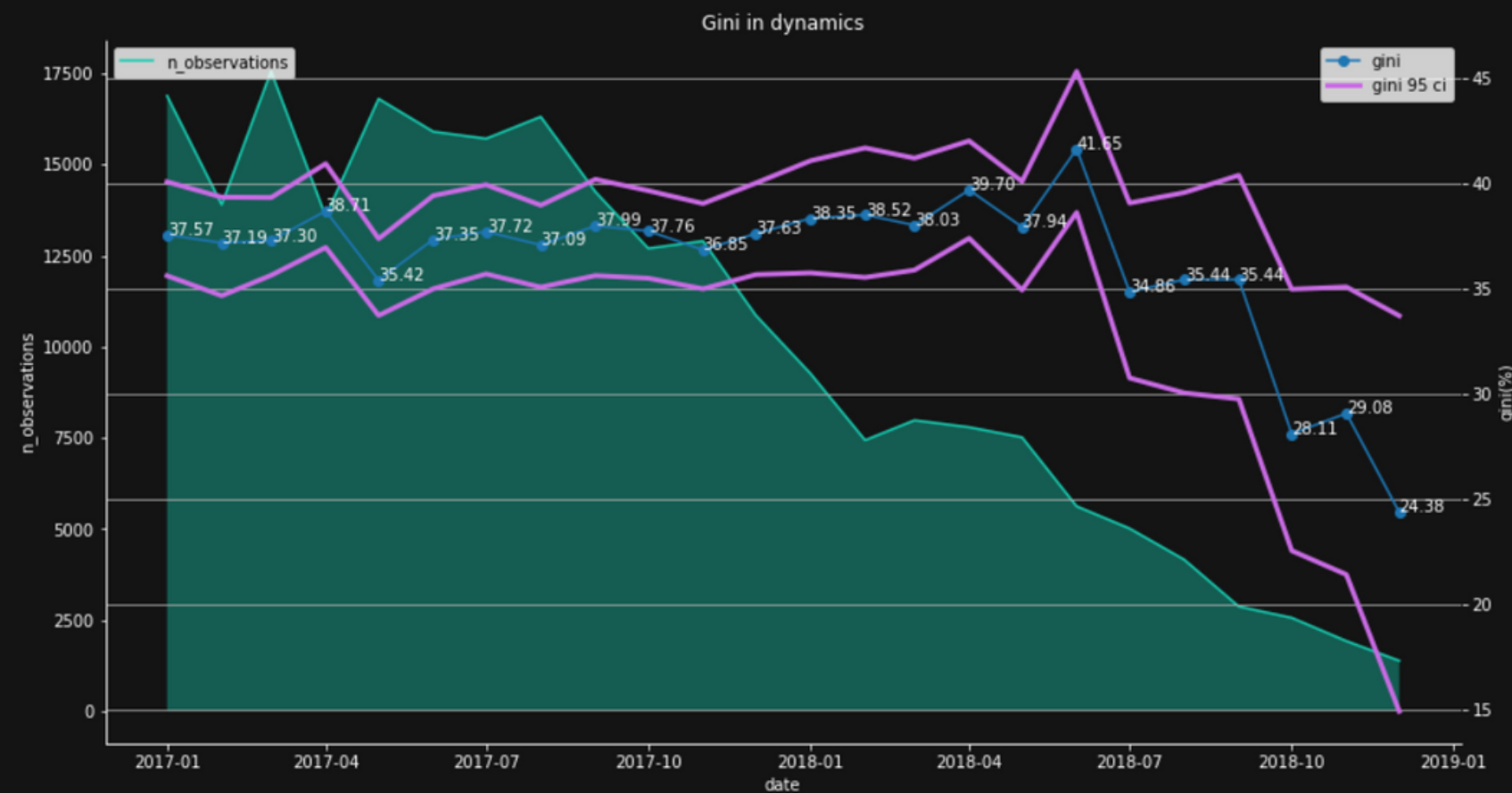
Валидация моделей

ИНКРЕМЕНТАЛЬНЫЙ ВКЛАД ФАКТОРОВ ПО КОЭФФИЦИЕНТУ ДЖИНИ

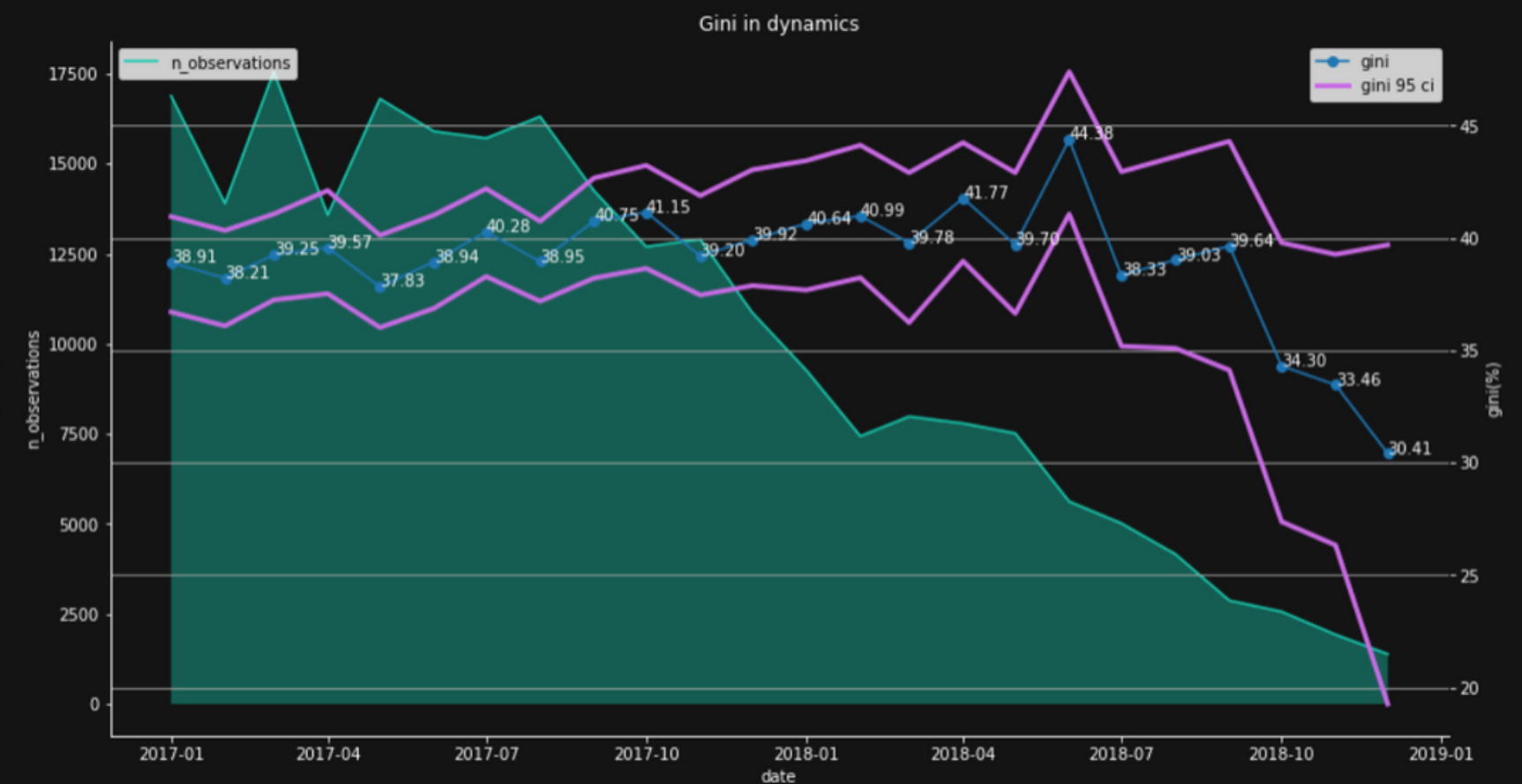


Валидация моделей

ДИНАМИКА КОЭФФИЦИЕНТА ДЖИНИ ВО ВРЕМЕНИ



LogReg

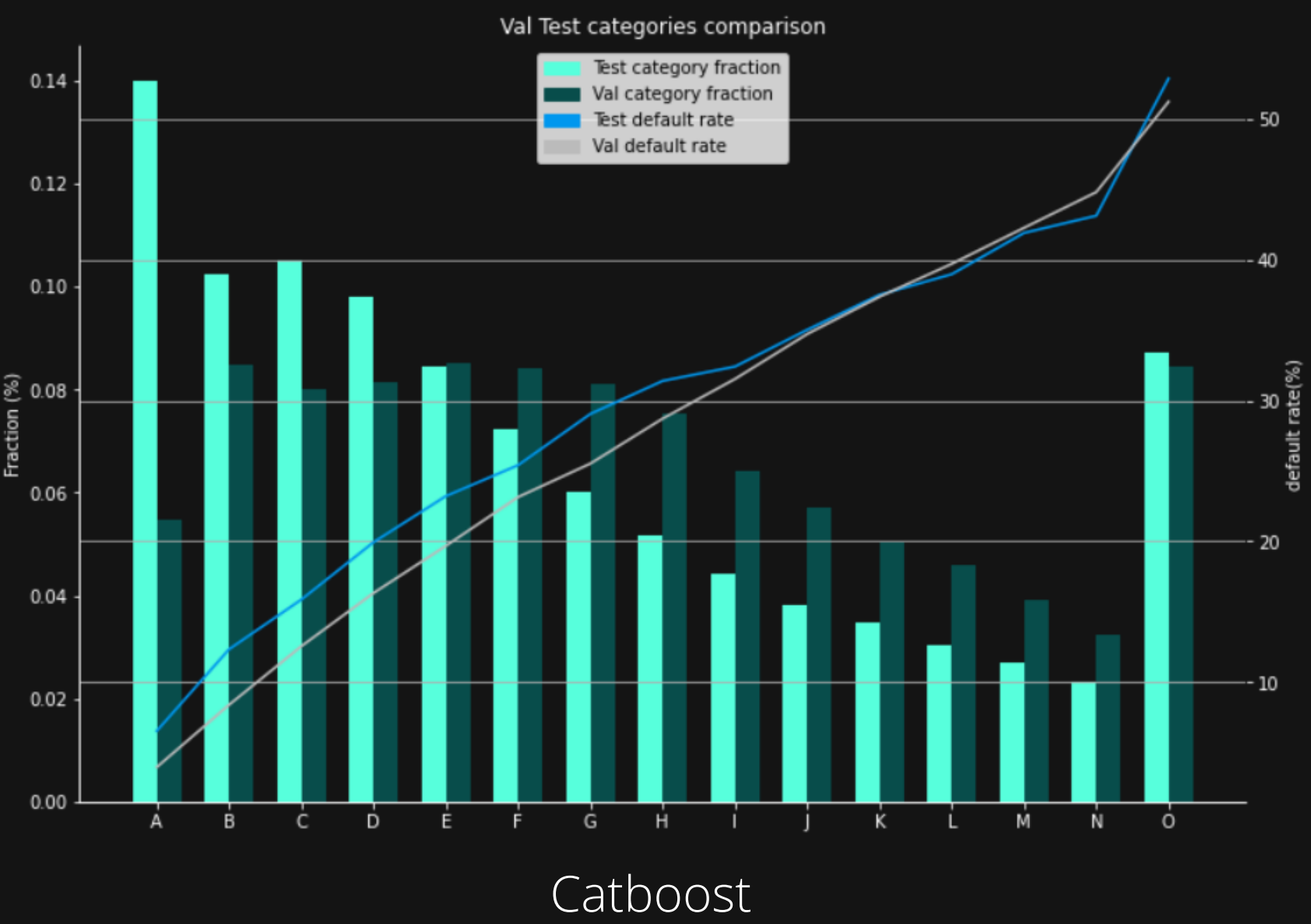
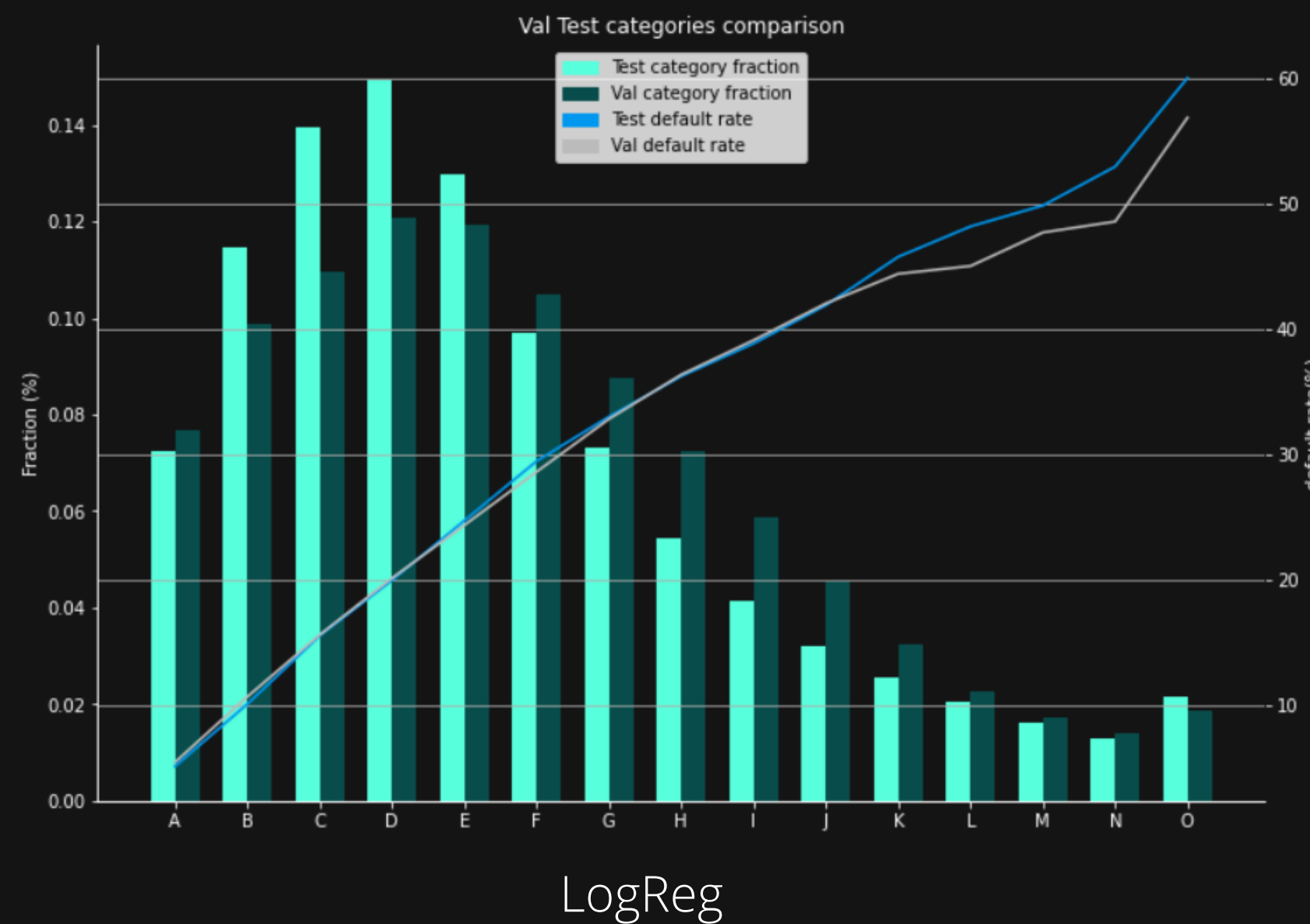


Catboost

Валидация моделей



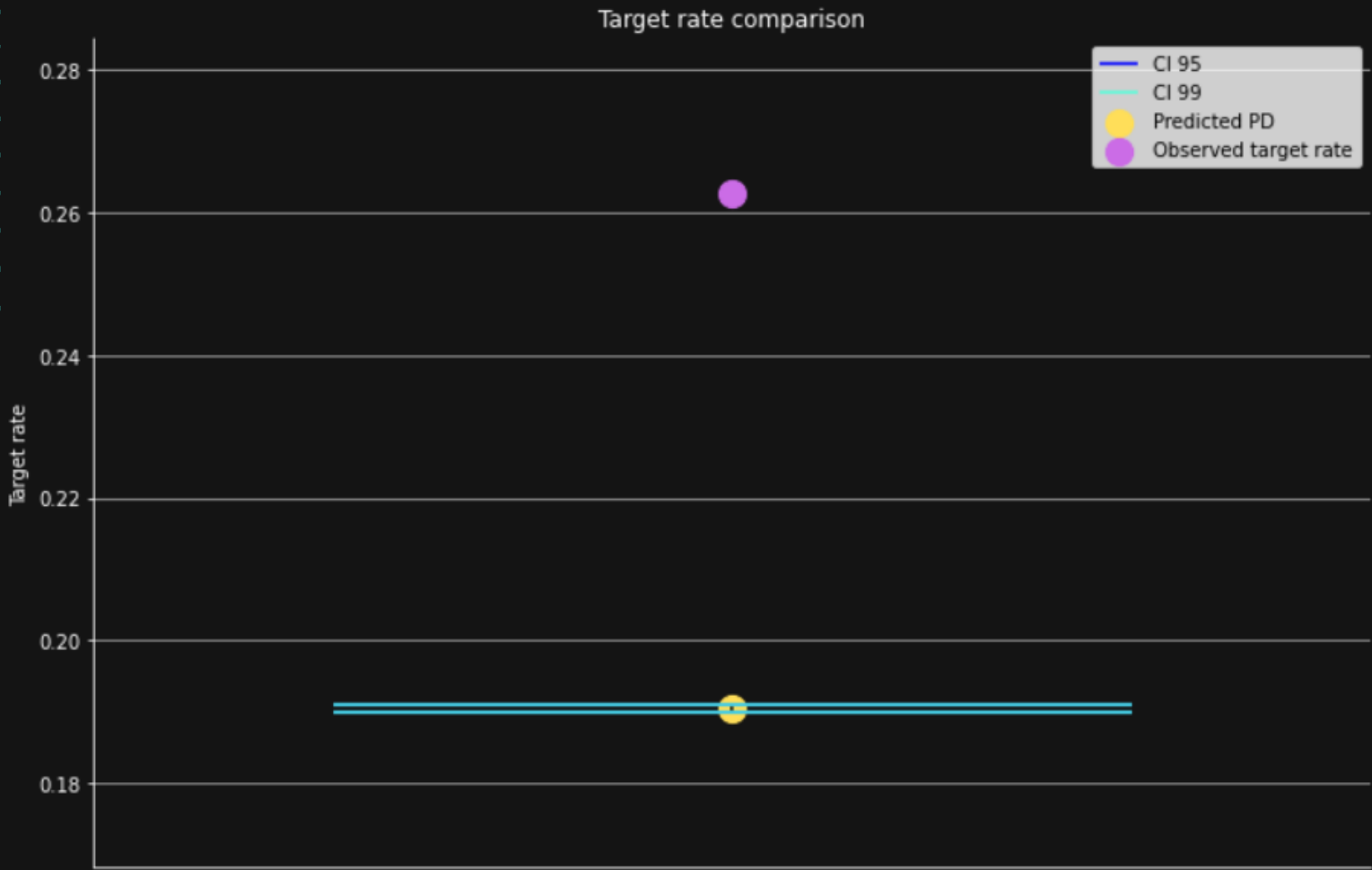
ЭКОНОМИЧЕСКИЙ СМЫСЛ РАЗБИЕНИЯ ПО КАТЕГОРИЯМ



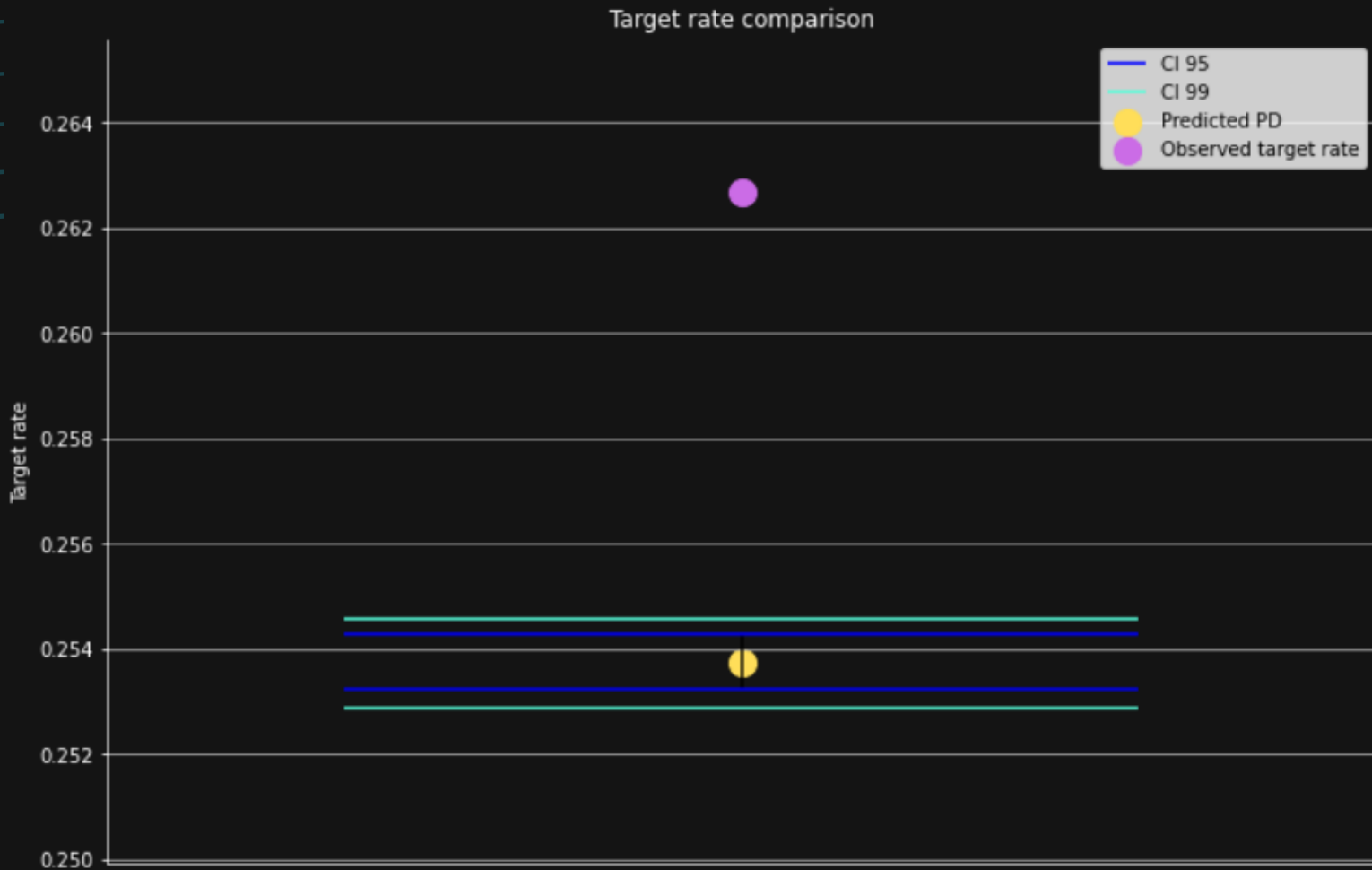
Валидация моделей



РЕАЛИЗОВАННАЯ ДЕФОЛТНОСТЬ (В %)



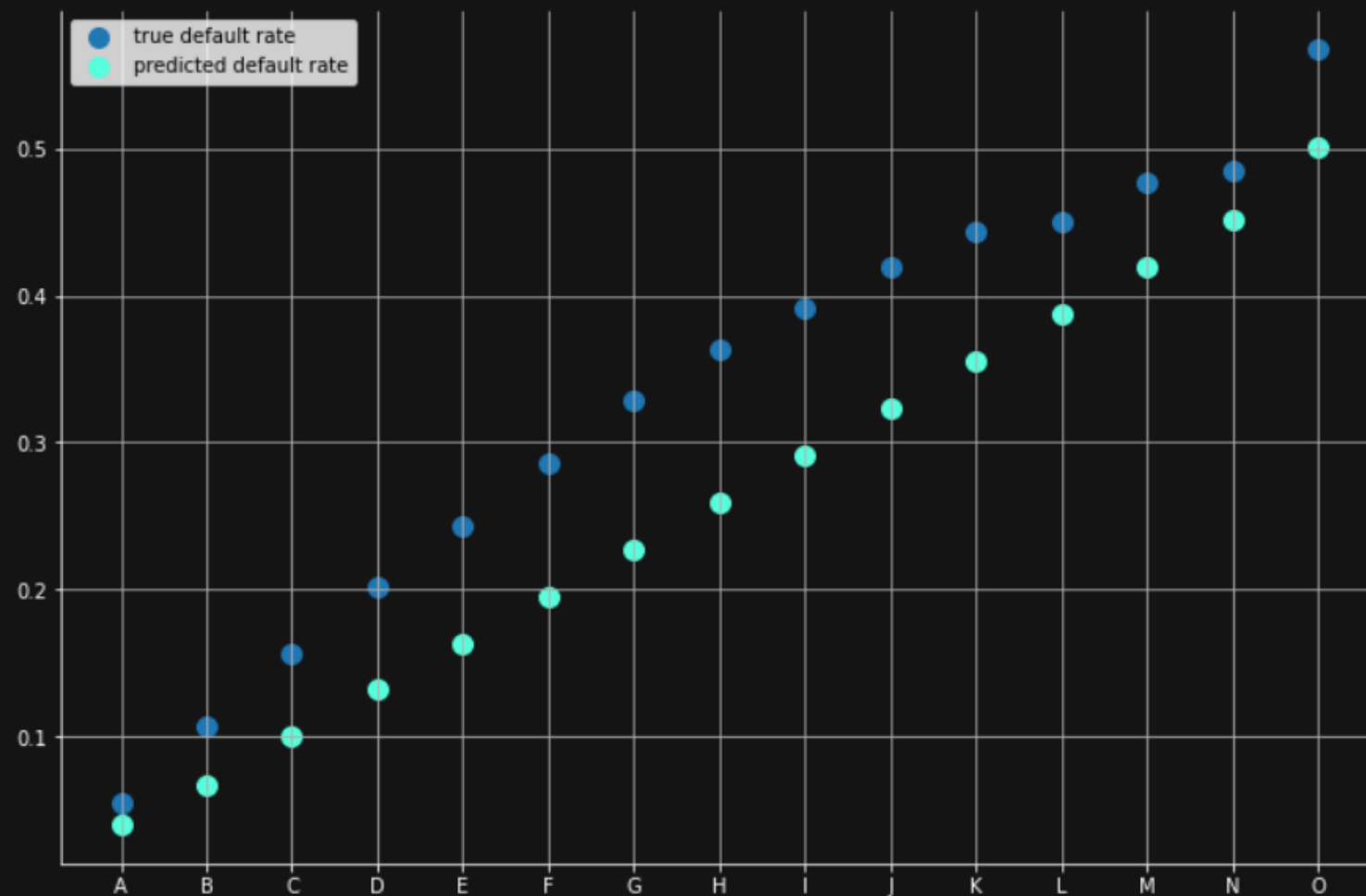
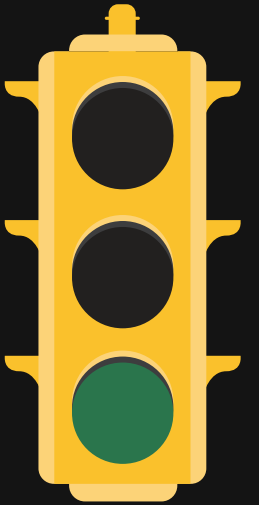
LogReg



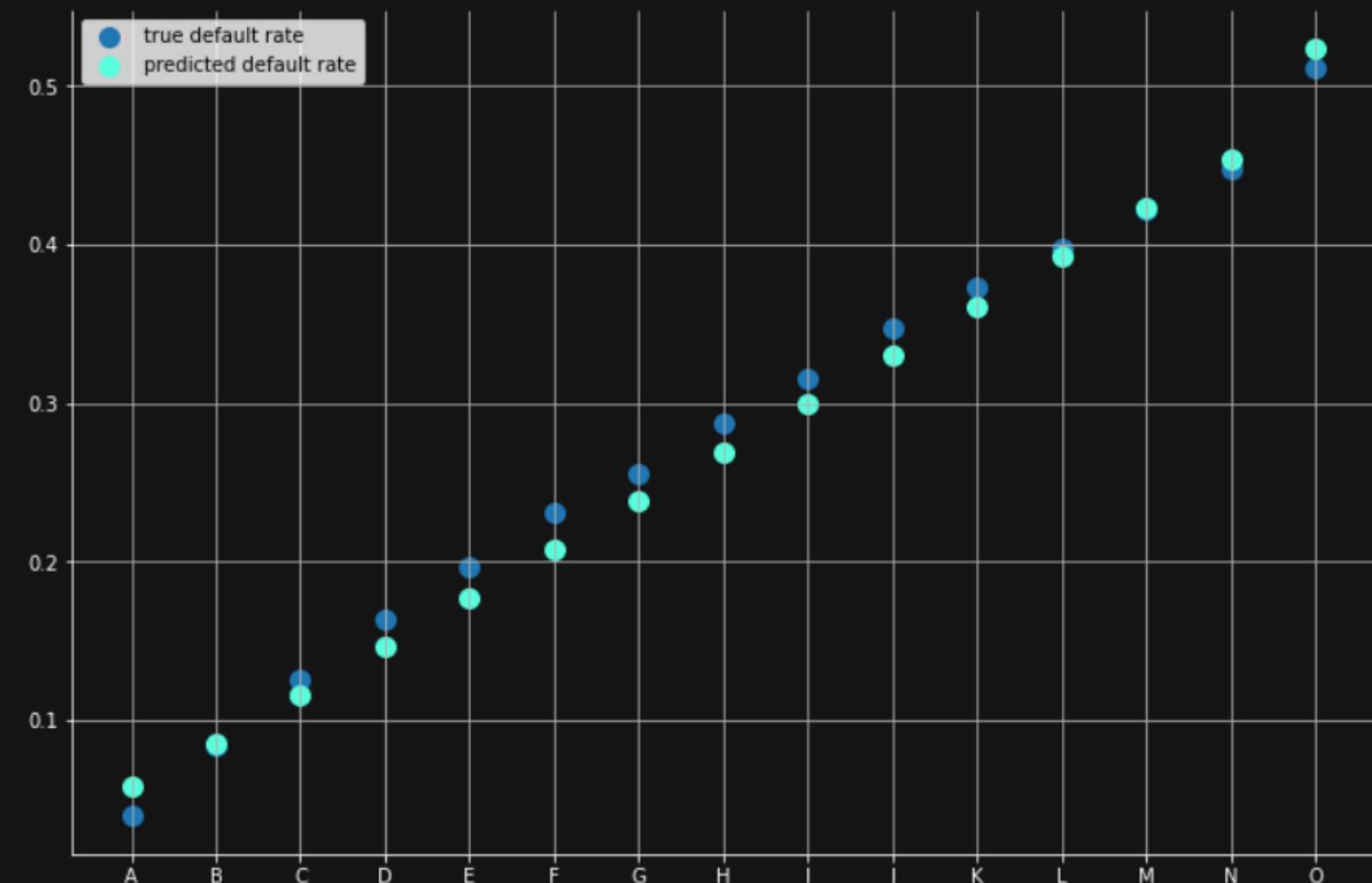
Catboost

Валидация моделей

РЕАЛИЗОВАННАЯ ДЕФОЛТНОСТЬ ПО КАТЕГОРИЯМ



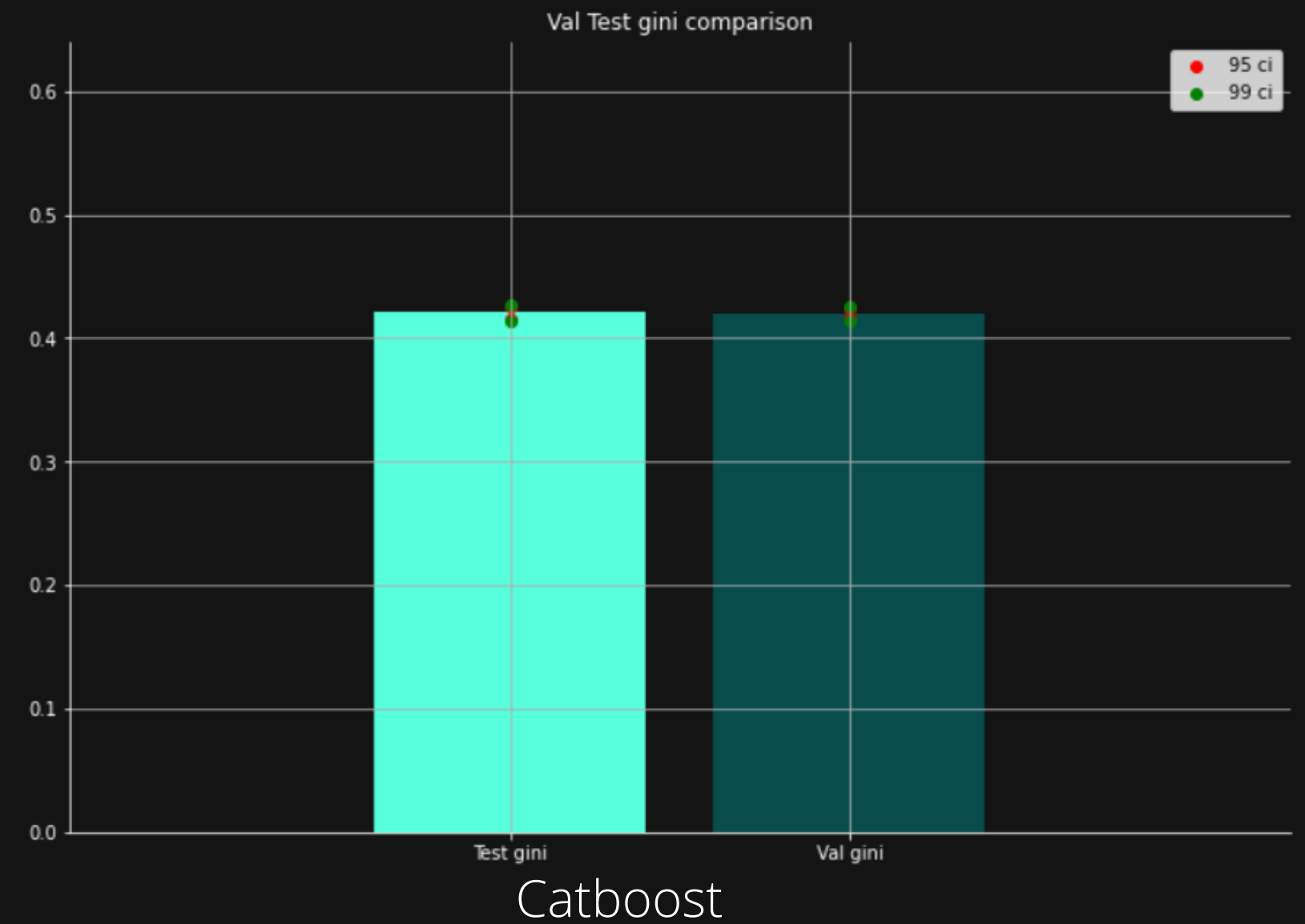
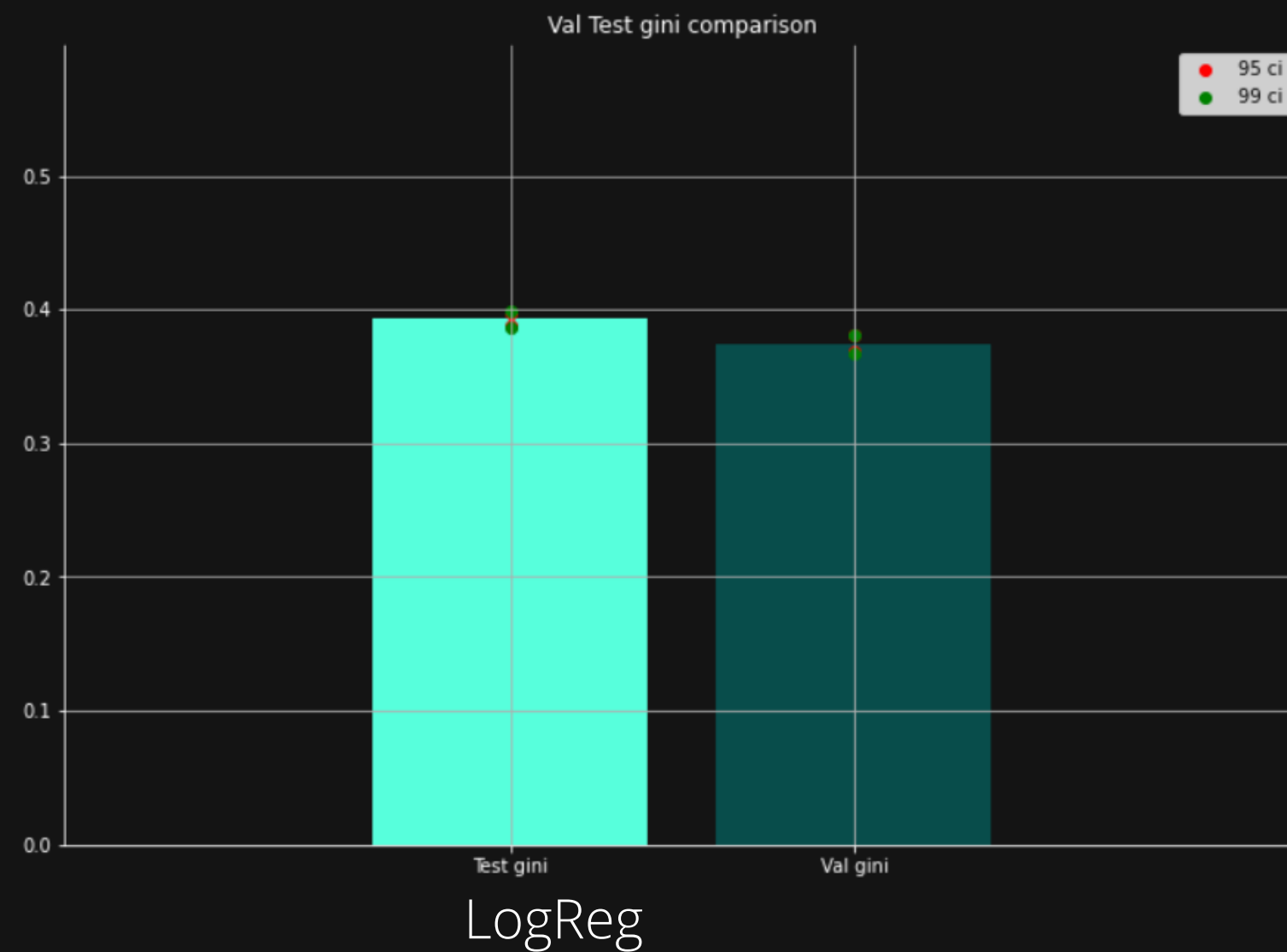
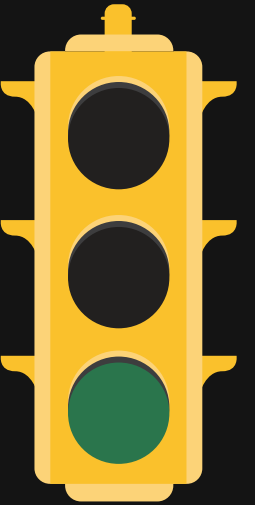
LogReg



Catboost

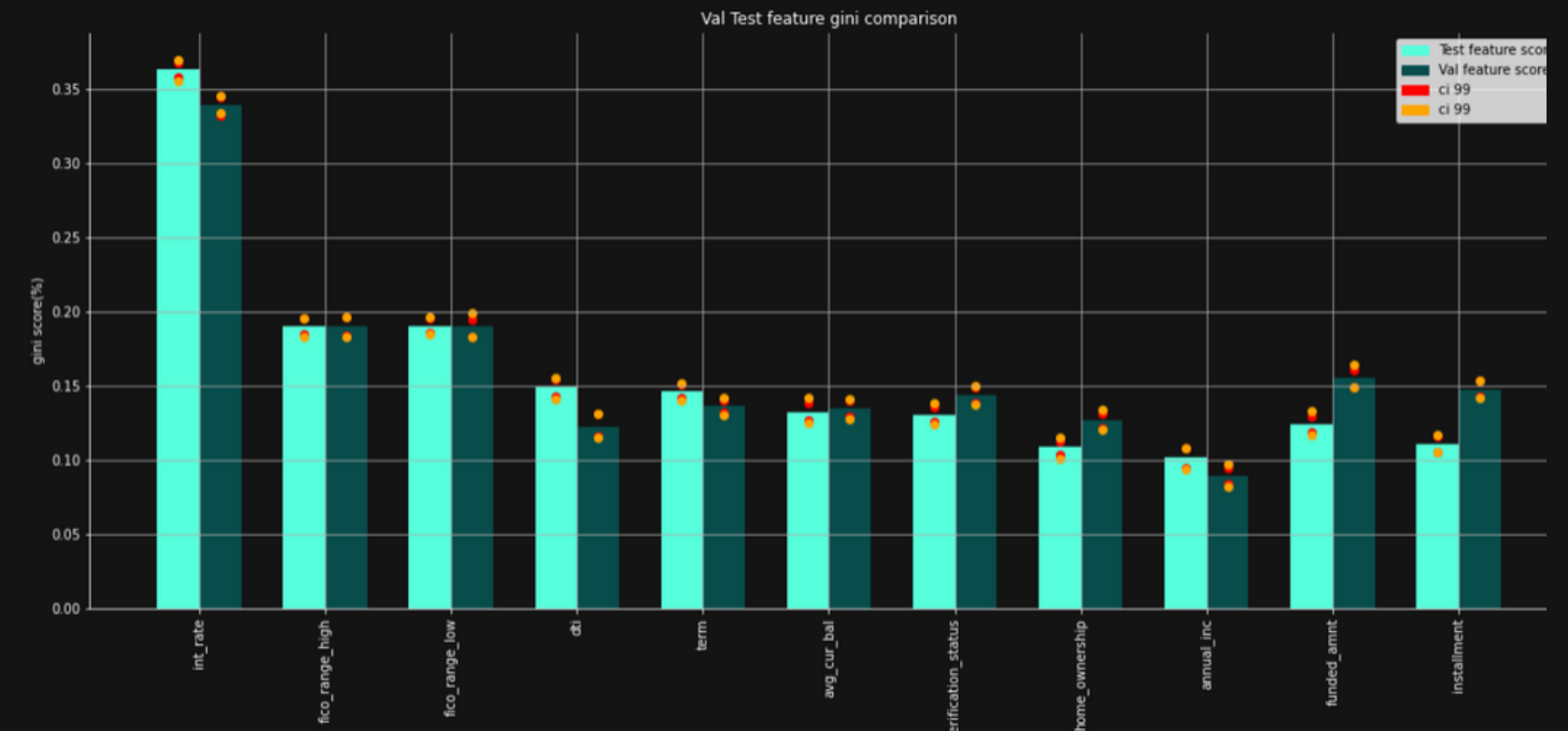
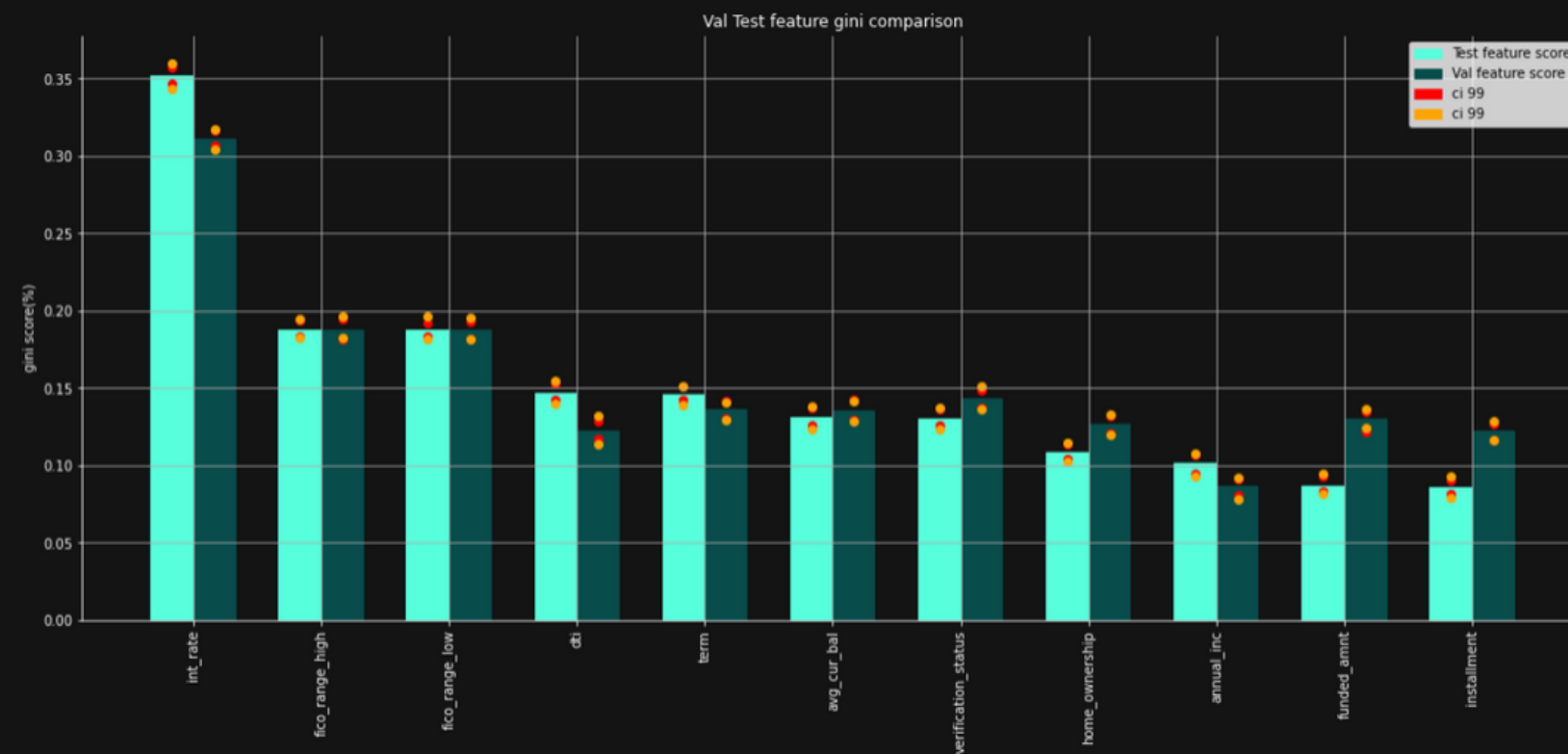
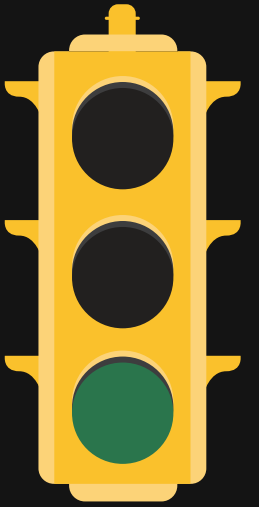
Валидация моделей

РАЗДЕЛЯЮЩАЯ СПОСОБНОСТЬ МОДЕЛЕЙ



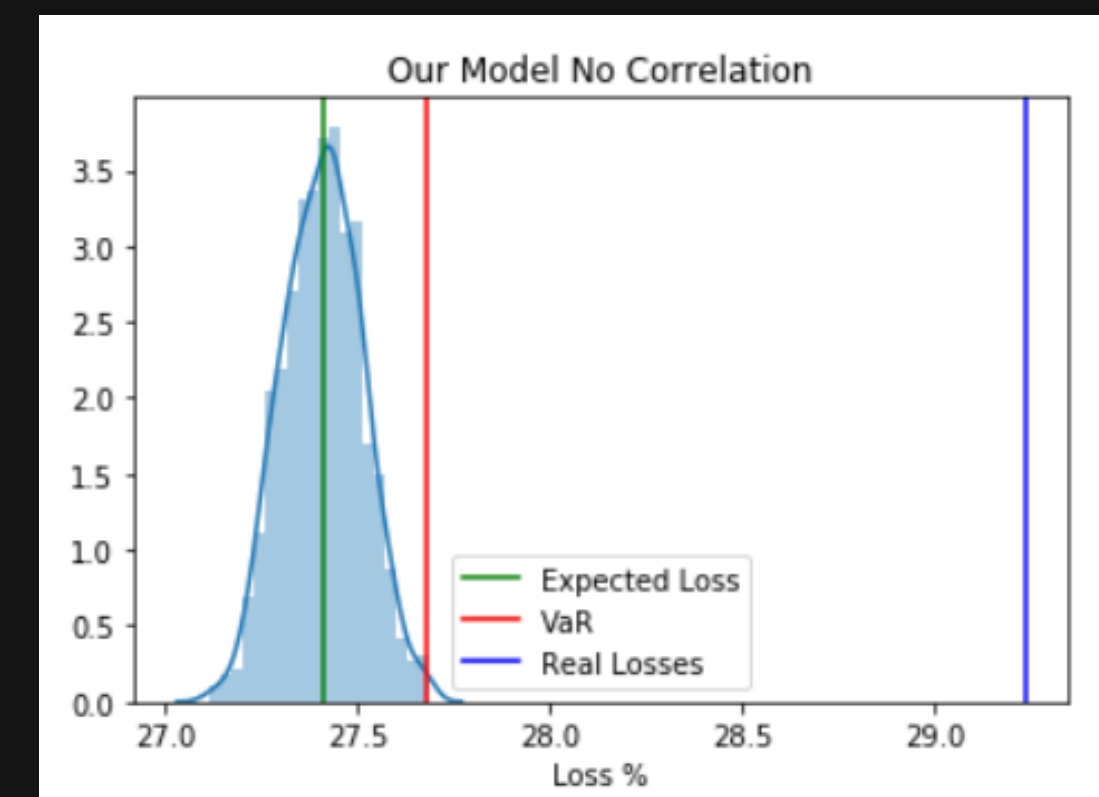
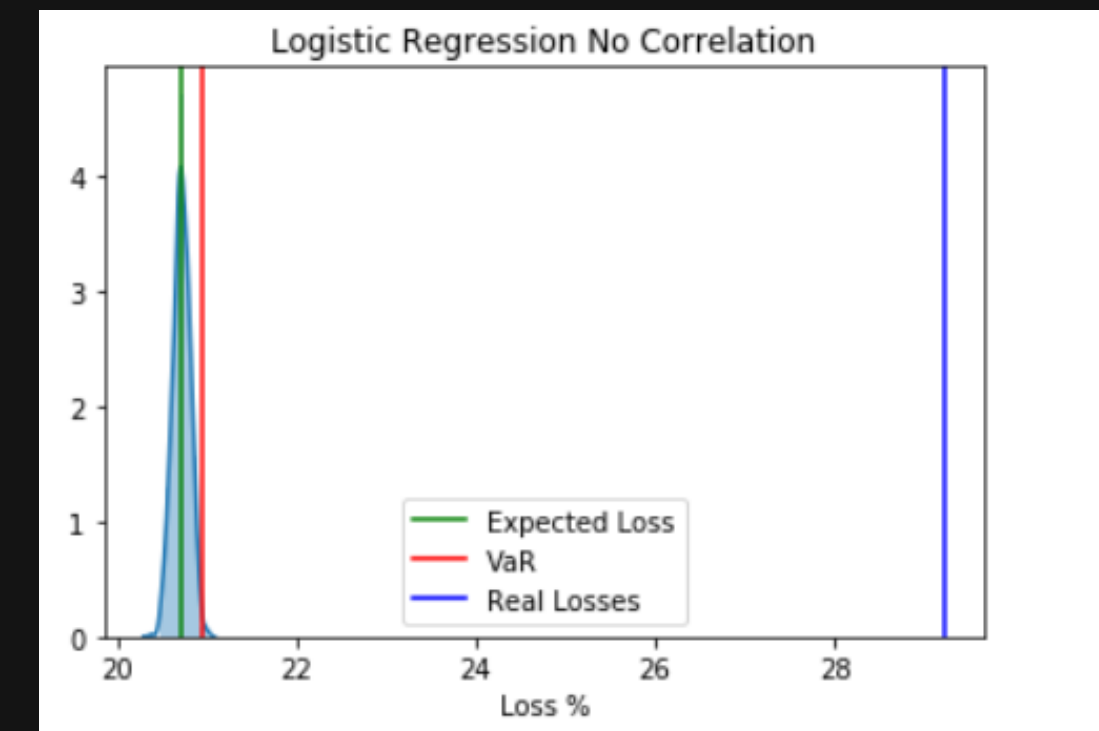
Валидация моделей

ИНДИВИДУАЛЬНАЯ РАЗДЕЛЯЮЩАЯ СПОСОБНОСТЬ МОДЕЛЕЙ



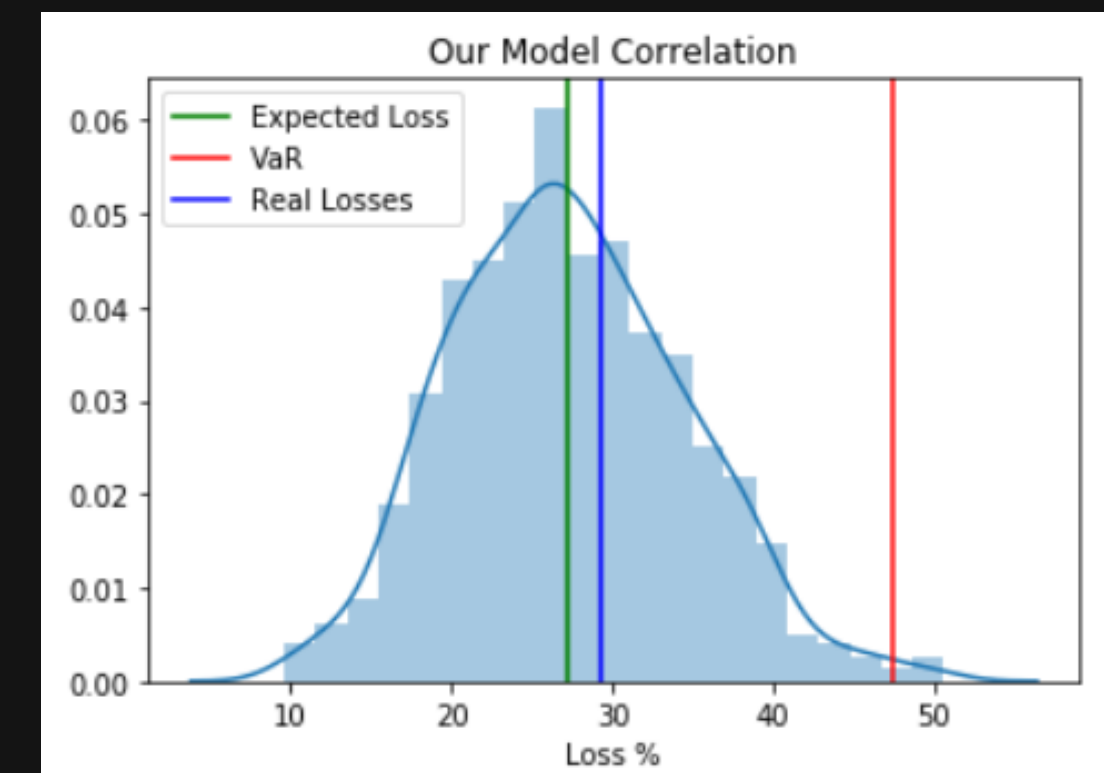
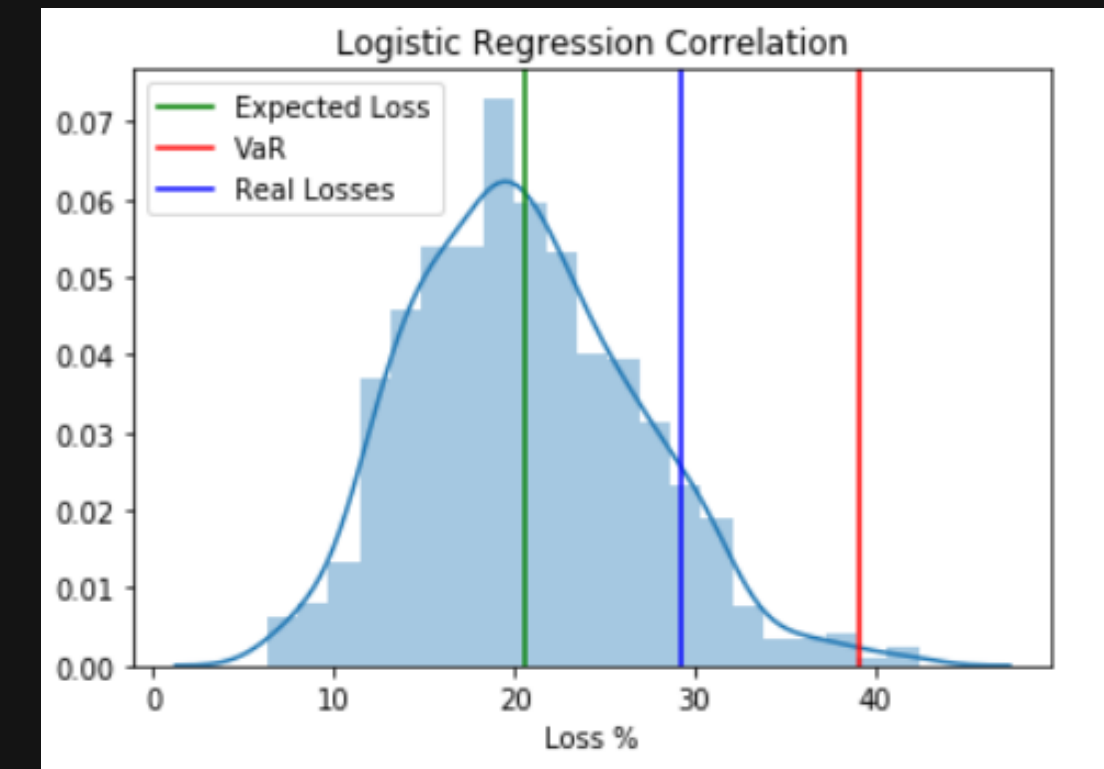
Ожидаемые и неожиданные потери

	Logistic Regression	Our Model
Expected Capital (%)	20.71	27.74
VaR (%)	20.71	27.74
Required Capital (%)	0	0



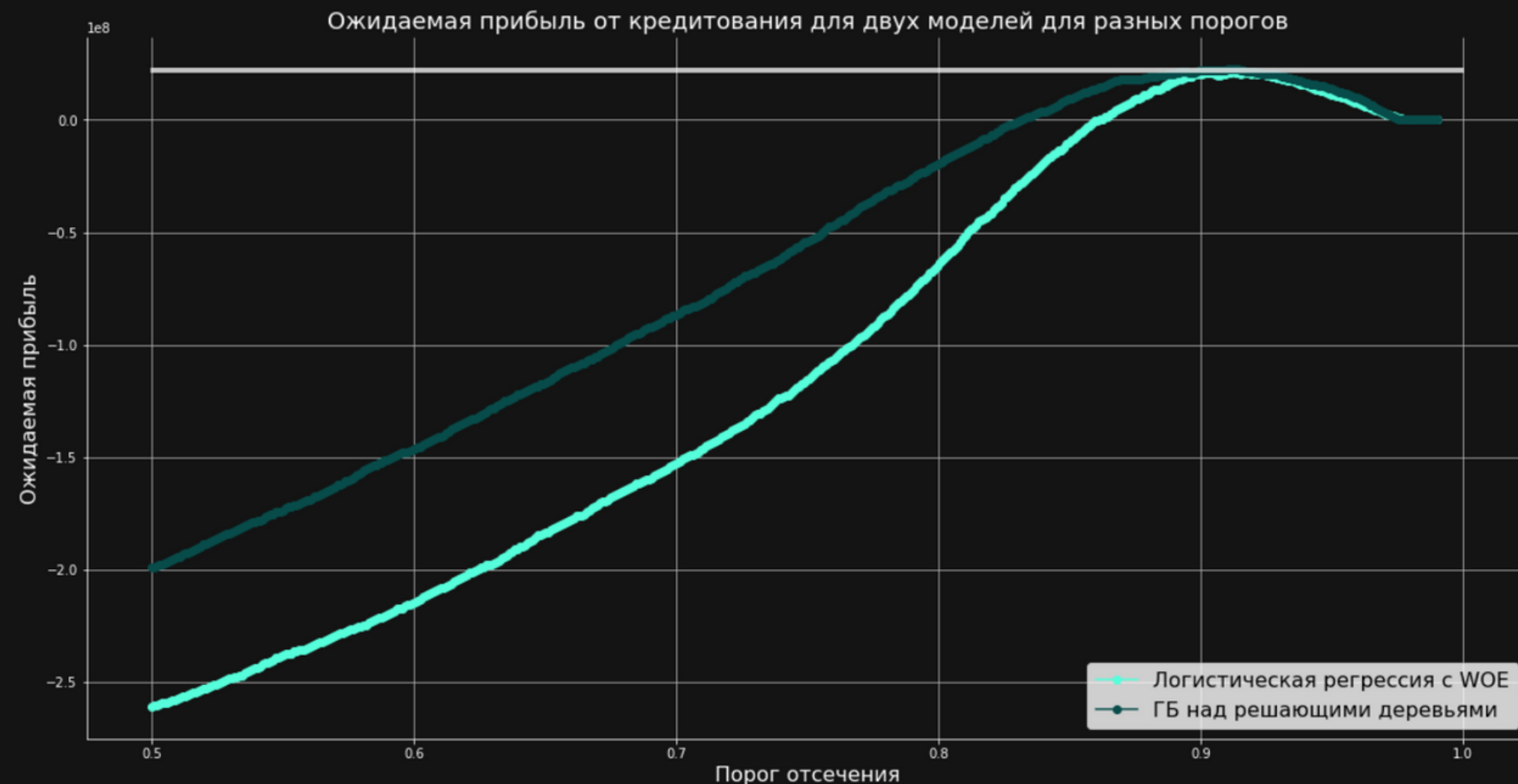
Ожидаемые и неожиданные потери

	Logistic Regression	Our Model
Expected Capital (%)	20.71	27.74
VaR (%)	40.17	49.22
Required Capital (%)	12.43	21.48



Ожидаемая прибыль от кредитования

- Оптимальный порог отсечения для логистической регрессии: 0.9124. Ожидаемая прибыль при этом пороге: 21.2346 миллионов у.е.
- Оптимальный порог отсечения для гб над решающими деревьями: 0.9105. Ожидаемая прибыль при этом пороге: 22.7081 миллионов у.е.



**Спасибо за
внимание**

