

# Кредитный скоринг физических лиц

Подчерцев Алексей  
Самоделкина Мария  
Солодянкин Андрей

---

# СОДЕРЖАНИЕ

Данные

Интерпретируемый скоринг

Бустинг

Валидация

Необходимый капитал и прибыль



# Предобработка данных

Поставляется метка дефолта

Дефолт: Charged Off, Default, Late (90-120 days)

Не дефолт: Fully Paid

Текущие кредиты: Current, In Grace Period, Late (< 90 days)

Удаляются признаки, не известные на момент подачи заявки. А также признаки, связанные с процентной ставкой и оценкой заемщика

Генерируются признаки: debt load, log annual income, fractional annual income

Предобработаны признаки: term, emp\_length, desc, emp\_title

Проводится разбиение на Train, Test, Out-of-time

# Данные

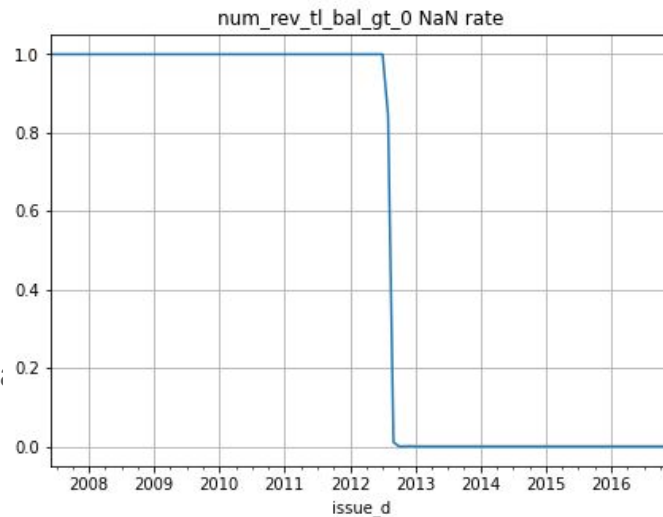
Выборка	Размер	Доля дефолтов	Минимальная дата	Максимальная дата
train	737965	0.201404	2012-09-01	2016-12-01
oos	315873	0.202464	2012-09-01	2016-12-01
oot	230990	0.236499	2017-01-01	2018-11-01



# Качество данных

Некоторые факторы не заполнялись до 2013 или 2016 годов

- Сократили обучающую выборку до 2013 года  
(mort\_acc, num\_rev\_tl\_bal\_gt\_0, total\_bc\_limit)
- Убрали из анализа факторы, не заполненные до 2016 год;  
(inq\_fi, total\_cu\_tl, all\_util)





# Интерпретируемая модель - AutoWoE

Обучение модели логистической регрессии на WoE-биннинге:

- Монотонный биннинг
- Максимальное кол-во бинов - 4
- Минимальный размер бина - 5%

Отбор факторов:

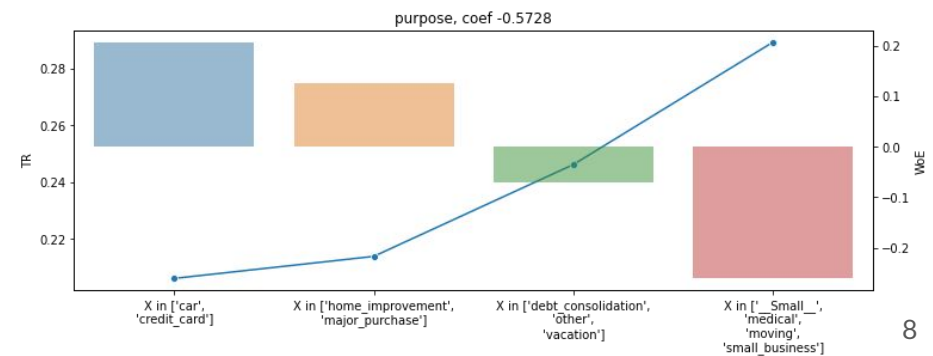
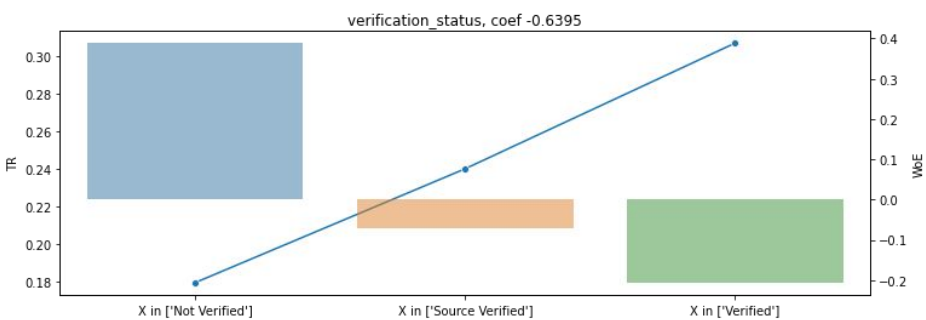
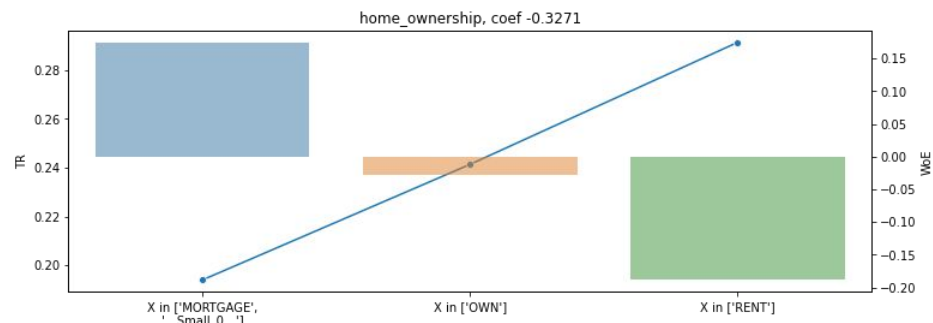
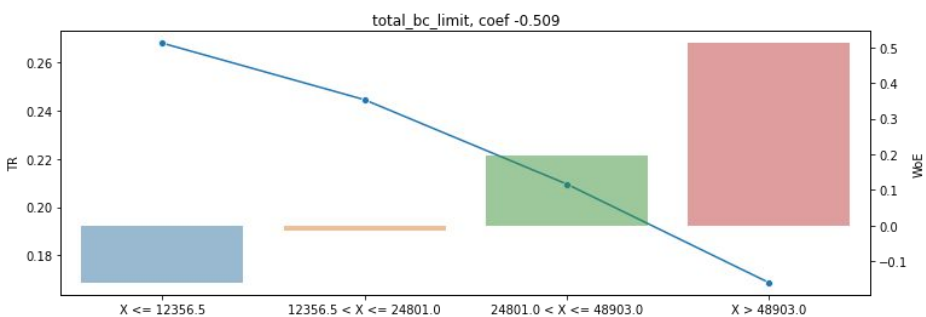
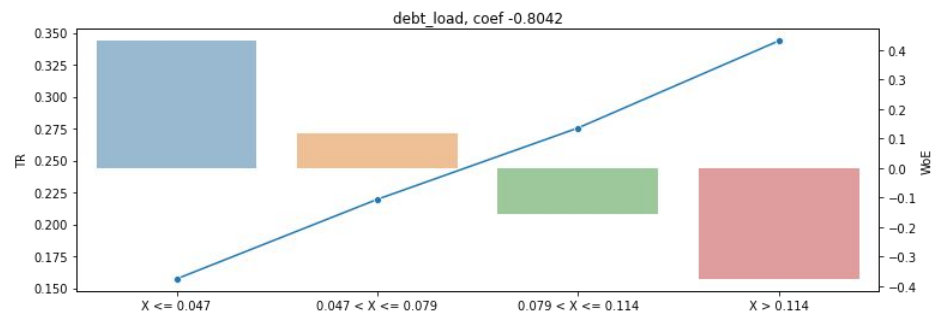
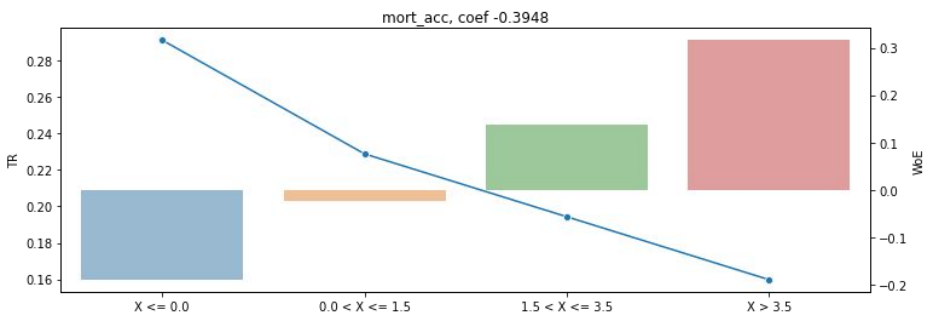
- Порог по корреляции Пирсона - 0.9
- Порог по VIF - 10
- Порог по индивидуальной ранжирующей силе фактора - 6 Джини
- Отбор по важности фактора

**Итоговое количество факторов: 10.** fico\_range\_high, debt\_load, verification\_status, bc\_open\_to\_buy, tot\_hi\_cred\_lim, mort\_acc, total\_bc\_limit, home\_ownership, num\_rev\_tl\_bal\_gt\_0, purpose

# Скоркарта

feature	weight	condition	WoE
fico_range_high	-0.671762	X <= 686.5	-0.266948
		686.5 < X <= 706.5	-0.007008
		706.5 < X <= 741.5	0.339203
		X > 741.5	0.902799
debt_load	-0.804218	X <= 0.047	0.431023
		0.047 < X <= 0.079	0.119727
		0.079 < X <= 0.114	-0.156906
		X > 0.114	-0.375850
verification_status	-0.639472	X in ['Not Verified']	0.388572
		X in ['Source Verified']	-0.070383
		X in ['Verified']	-0.206533
		X <= 2736.5	-0.186837
bc_open_to_buy	-0.084891	2736.5 < X <= 9419.5	-0.033002
		9419.5 < X <= 28307.5	0.226299
		X > 28307.5	0.636757
		NaN	-0.033002

tot_hi_cred_lim	-0.039512	X <= 115526.5	-0.162590
		115526.5 < X <= 175756.5	-0.020699
		175756.5 < X <= 385213.0	0.192889
		X > 385213.0	0.459973
mort_acc	-0.394780	X <= 0.0	-0.189377
		0.0 < X <= 1.5	-0.022894
		1.5 < X <= 3.5	0.138003
		X > 3.5	0.316089
total_bc_limit	-0.508963	X <= 12356.5	-0.159603
		12356.5 < X <= 24801.0	-0.012889
		24801.0 < X <= 48903.0	0.196881
		X > 48903.0	0.511830
home_ownership	-0.327084	X in ['MORTGAGE', '__Small_0__']	0.174495
		X in ['OWN']	-0.028567
		X in ['RENT']	-0.188653
		X <= 3.5	0.208814
num_rev_tl_bal_gt_0	-0.809744	3.5 < X <= 7.5	0.016485
		7.5 < X <= 11.5	-0.190475
		X > 11.5	-0.402753
		X in ['car', 'credit_card']	0.207370
purpose	-0.572758	X in ['home_improvement', 'major_purchase']	0.126524
		X in ['debt_consolidation', 'other', 'vacation']	-0.071530
		X in ['__Small__', 'medical', 'moving', 'small_business']	-0.260544
intercept	-1.373419		nan







# Модель бустинга - CatBoost

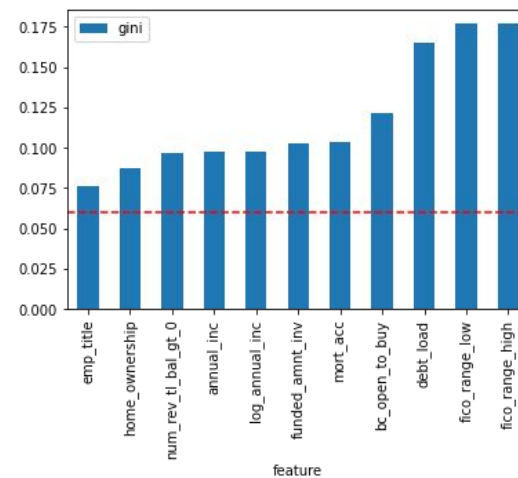
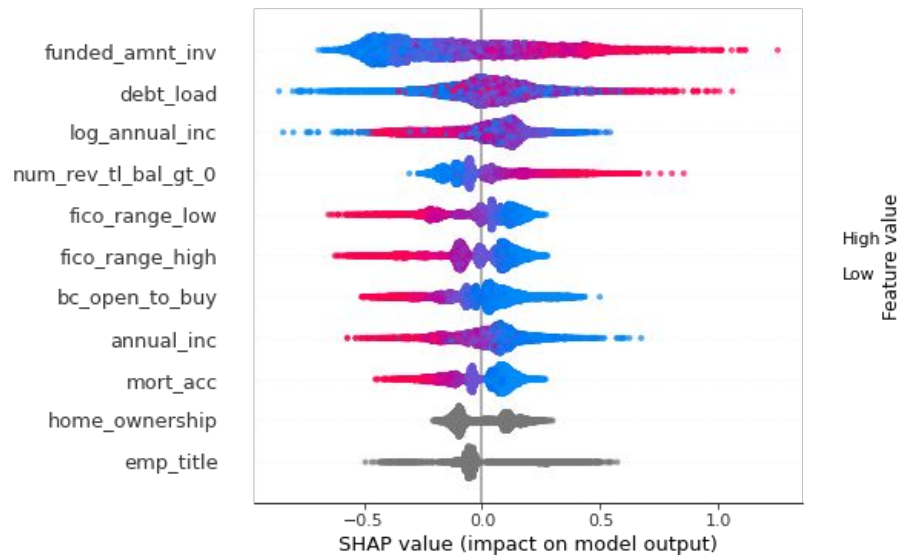
Отбор факторов:

- Порог по индивидуальной ранжирующей силе фактора - 6 Джини
- Отбор алгоритмом recursive feature elimination

Подбор гиперпараметров с помощью библиотеки optuna

**Итоговое количество факторов: 11.** home\_ownership, log\_annual\_inc, annual\_inc, num\_rev\_tl\_bal\_gt\_0, mort\_acc, funded\_amnt\_inv, emp\_title, bc\_open\_to\_buy, debt\_load, fico\_range\_low, fico\_range\_high

# Модель бустинга - CatBoost

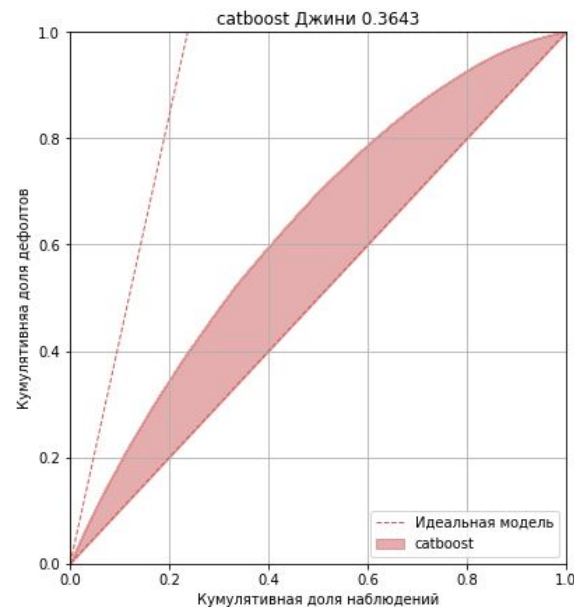
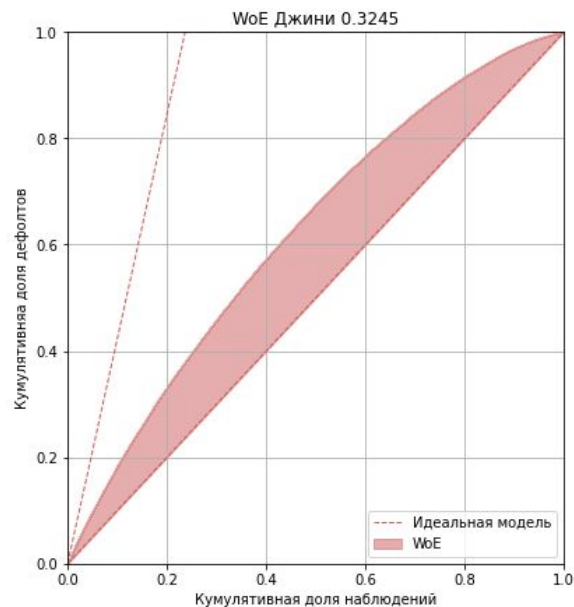


# Валидация

---

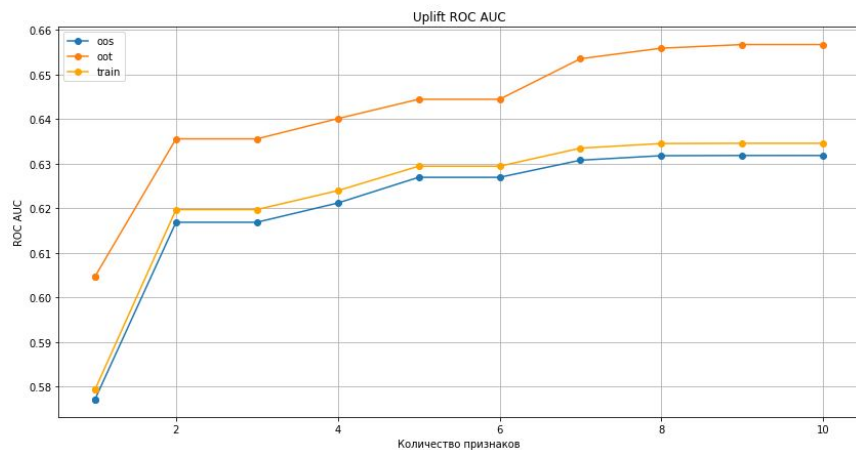
# Ранжирующая способность

Кривая Джини

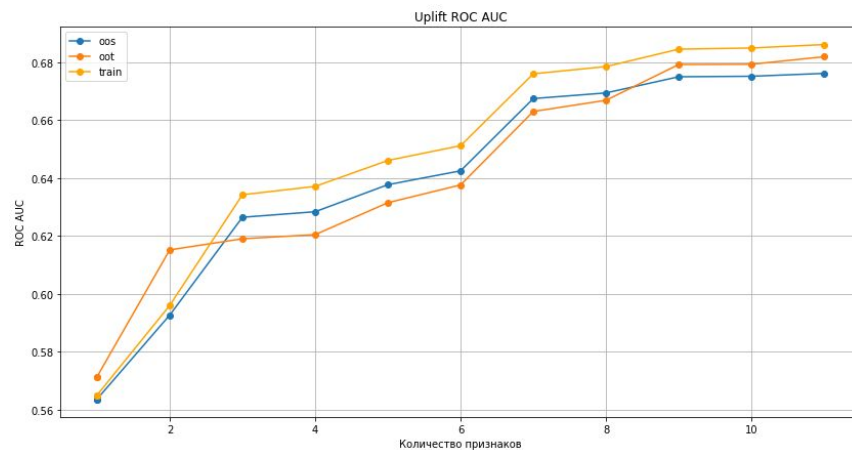


# Проверка на избыточность факторов

## Интерпретируемая модель

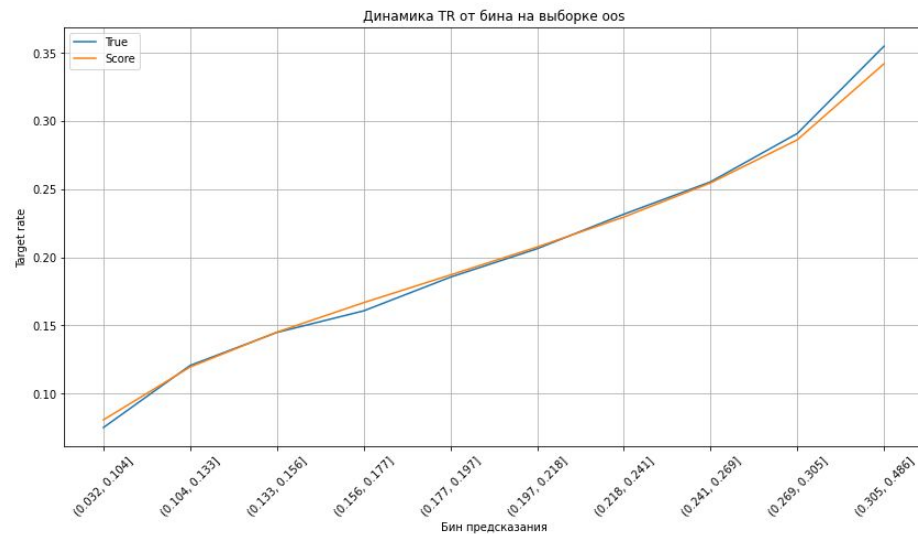


## Бустинг



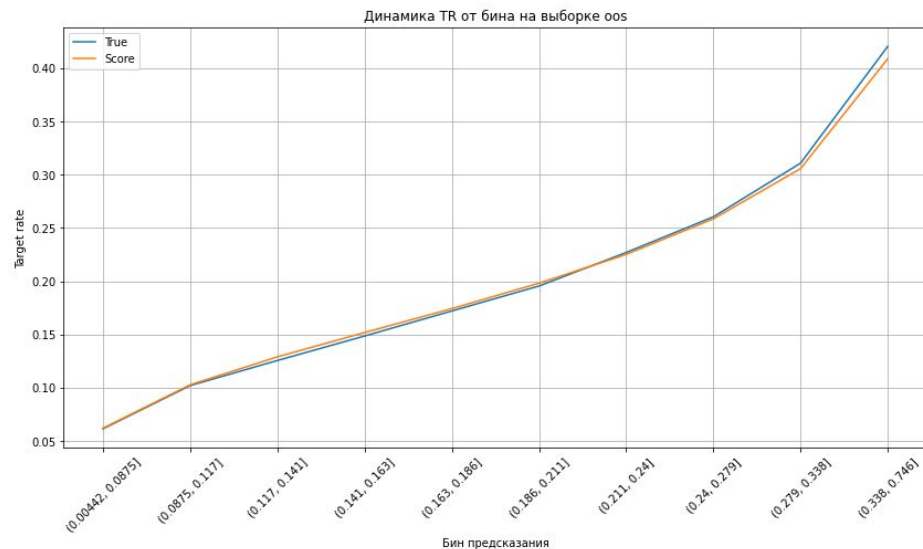
# Калибровка - WoE

sample	TR	SR	LB	UB
train <span style="color: green;">●</span>	0.2014	0.2014	0.2005	0.2023
oos <span style="color: green;">●</span>	0.2025	0.2018	0.2004	0.2032
oot <span style="color: red;">●</span>	0.2365	0.1827	0.1811	0.1843

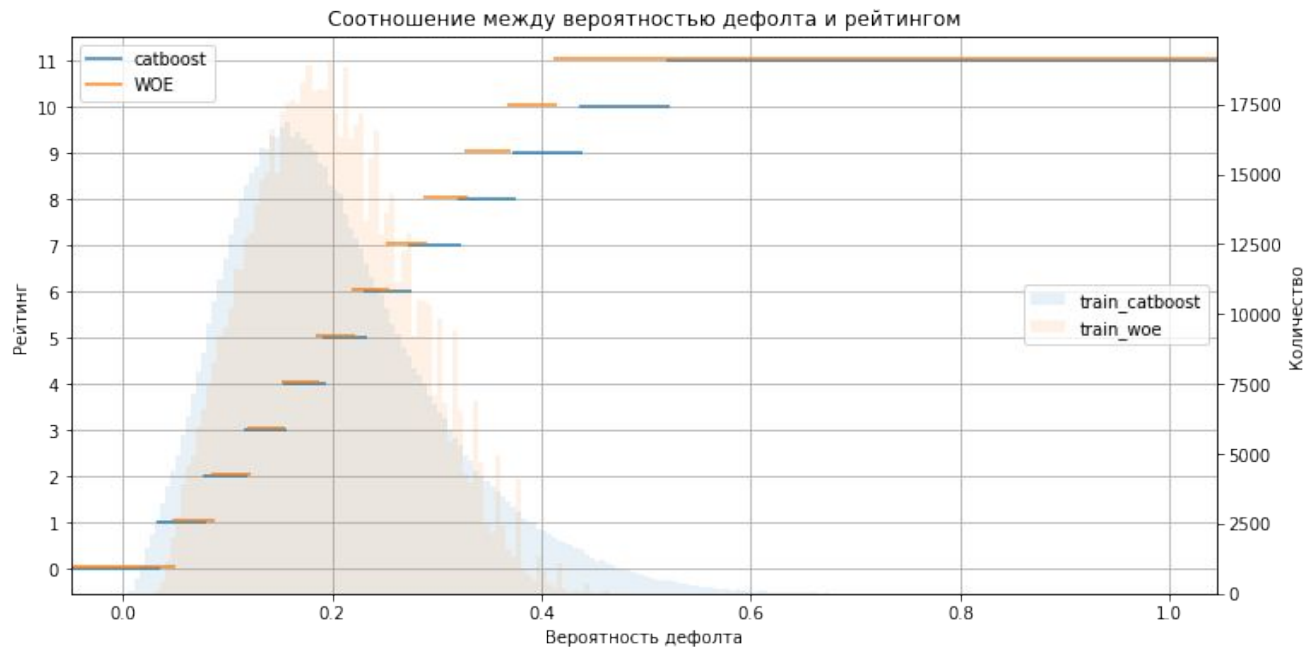


# Калибровка - CatBoost

sample	TR	SR	LB	UB
train <span style="color: green;">●</span>	0.2014	0.2013	0.2004	0.2022
oos <span style="color: green;">●</span>	0.2025	0.2017	0.2003	0.2031
oot <span style="color: red;">●</span>	0.2365	0.1850	0.1834	0.1866



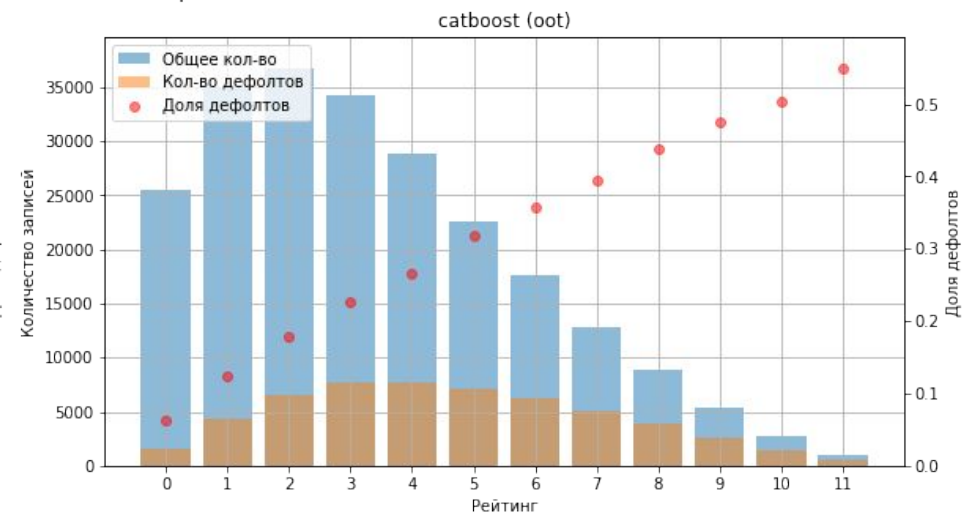
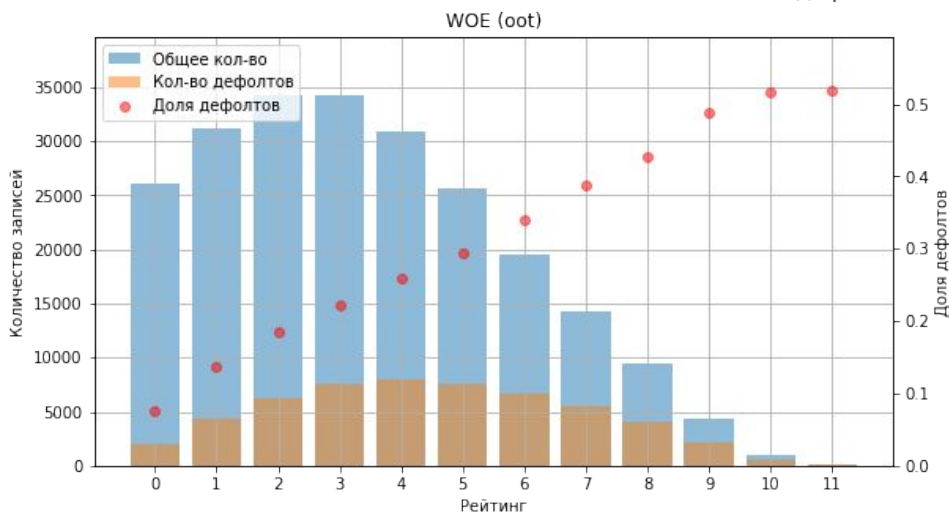
## Присвоение рейтингов (KMeans)





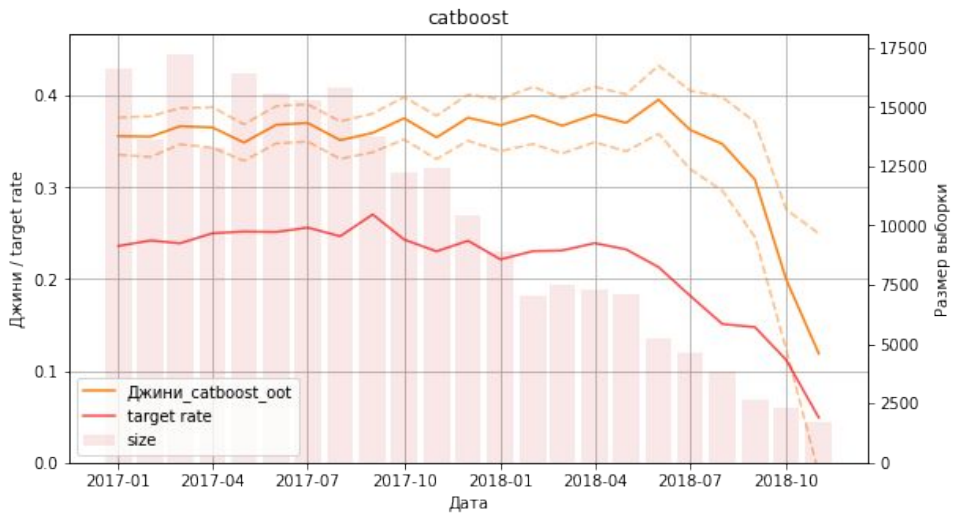
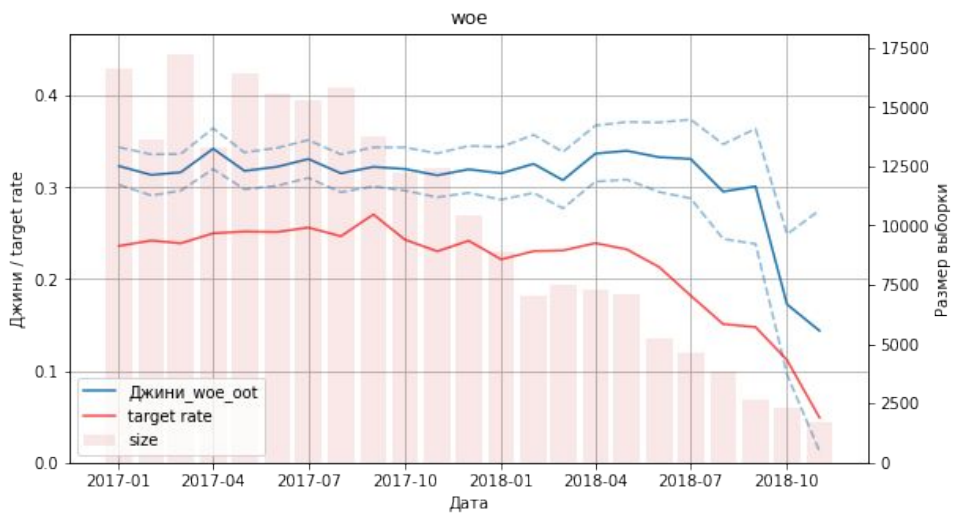
# Исследование рейтингов

Количество дефолтов в зависимости от рейтинга



# Устойчивость модели во времени

График динамики коэффициента Джини (oot)





## Исследование модели на переобучение

WOE

	Метрики Джини	Абс. откл.	Отн. откл. (%)
<b>train/oos</b>	0.2883/0.2823	0.0060	2.06
<b>train/oot</b>	0.2883/0.3245	-0.0362	-12.55
<b>oos/oot</b>	0.2823/0.3245	-0.0421	-14.92

catboost

	Метрики Джини	Абс. откл.	Отн. откл. (%)
<b>train/oos</b>	0.3731/0.3533	0.0198	5.30
<b>train/oot</b>	0.3731/0.3643	0.0088	2.36
<b>oos/oot</b>	0.3533/0.3643	-0.0110	-3.10



# PSI

	train/oos	train/oot	oos/oot
<b>woe_score</b>	0.0015	0.0707	0.0711
<b>catboost_score</b>	0.0019	0.0366	0.0371

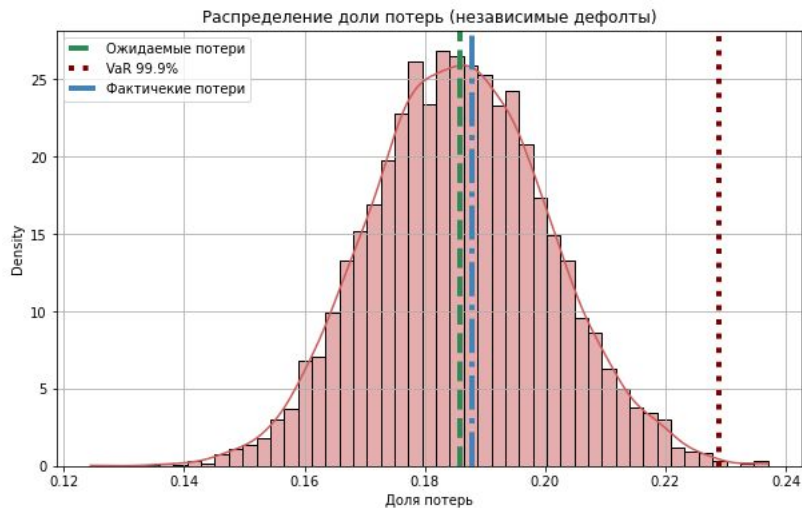
	train/oos	train/oot	oos/oot
<b>funded_amnt_inv</b>	0.0019	0.0856	0.0840
<b>debt_load</b>	0.0012	0.0778	0.0789
<b>fico_range_low</b>	0.0005	0.0541	0.0565
<b>fico_range_high</b>	0.0010	0.0518	0.0548
<b>purpose</b>	0.0001	0.0383	0.0367
<b>total_bc_limit</b>	0.0018	0.0265	0.0257
<b>emp_title</b>	0.0012	0.0152	0.0172
<b>mort_acc</b>	0.0005	0.0172	0.0165
<b>verification_status</b>	0.0000	0.0155	0.0159
<b>log_annual_inc</b>	0.0018	0.0105	0.0105
<b>annual_inc</b>	0.0025	0.0101	0.0096
<b>home_ownership</b>	0.0001	0.0067	0.0054
<b>bc_open_to_buy</b>	0.0000	0.0016	0.0014
<b>num_rev_tl_bal_gt_0</b>	0.0001	0.0000	0.0001

# Проведение вычислений

- Проведение симуляции по батчам заемщиков
- Учитываем заемщиков с текущим кредитом
- Симуляция скоррелированных дефолтов с помощью разложения Холецкого матрицы корреляций
- Объединение симуляций для получения итоговых результатов

# Расчет капитала (независимые дефолты)

## Интерпретируемая модель

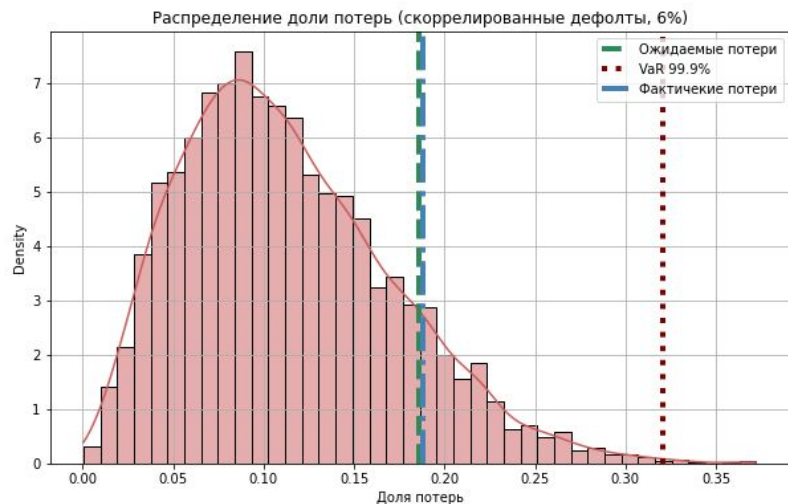


## Бустинг

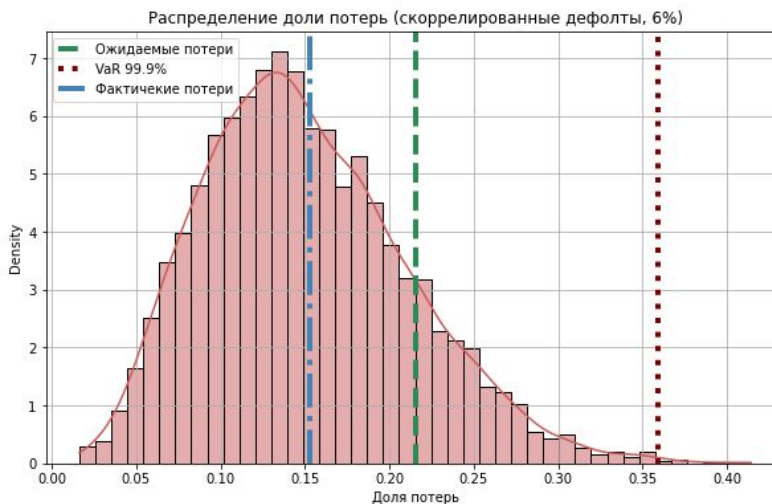


# Расчет капитала (скоррелированные дефолты)

## Интерпретируемая модель



## Бустинг

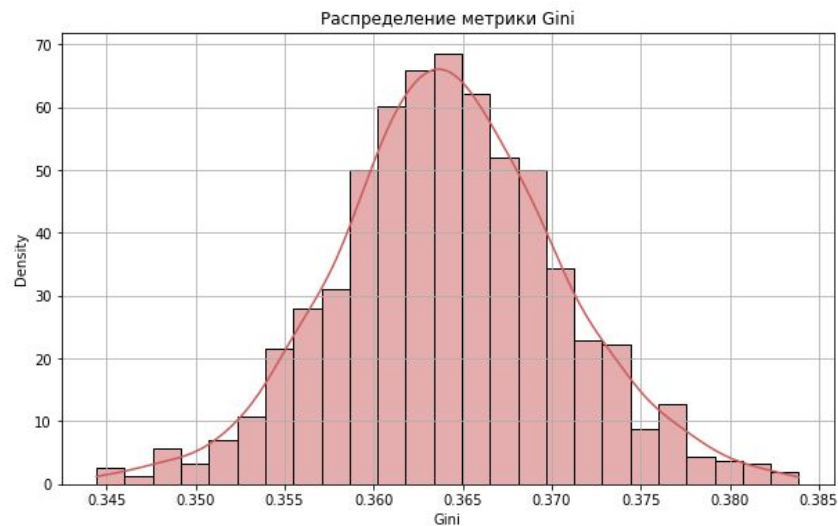


# Качество ранжирования

## Интерпретируемая модель



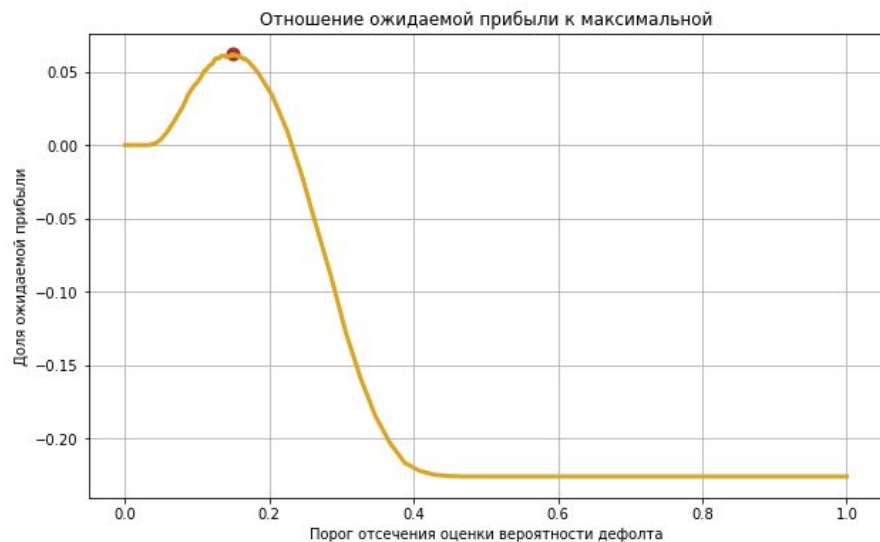
## Бустинг



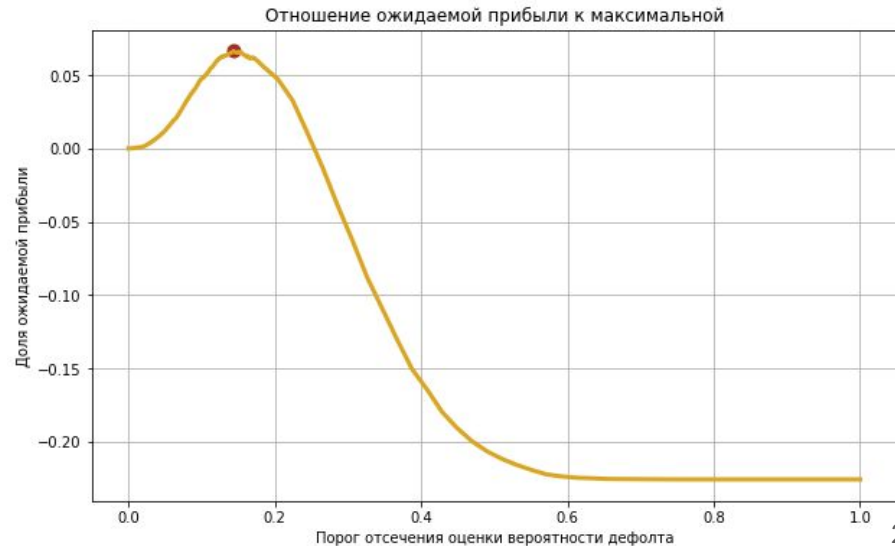


# Максимизация ожидаемой прибыли

## Интерпретируемая модель

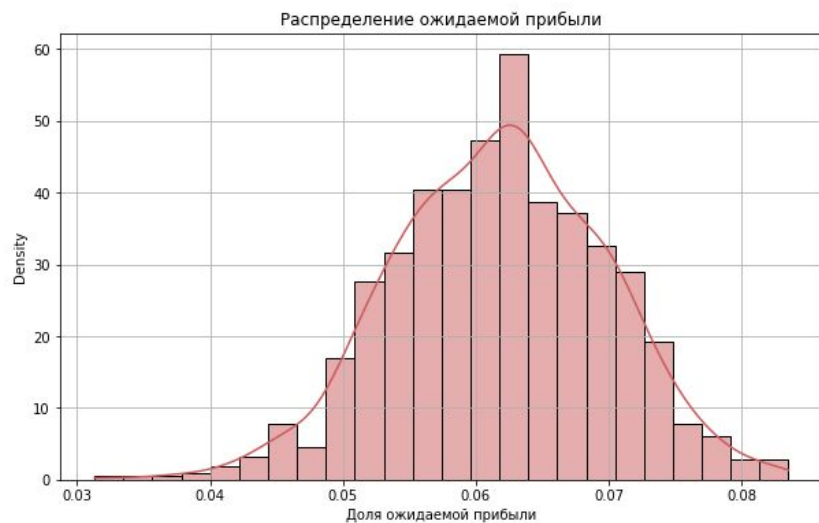


## Бустинг

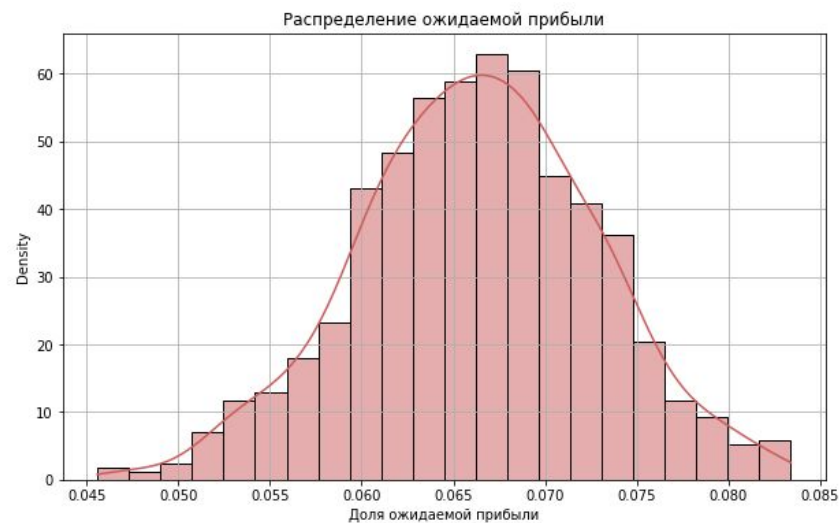


# Ожидаемая прибыль

## Интерпретируемая модель



## Бустинг





# Итоговый результат

## Интерпретируемая модель

	mean	LB 99	UB 99
Капитал (независимые дефолты)	0.0432	0.0363	0.0521
Капитал (коррелированные дефолты)	0.1349	0.1060	0.1869
Коэффициент Джини	0.3246	0.3078	0.3413
Ожидаемая прибыль (доля)	0.0618	0.0409	0.0818

## Бустинг

	mean	LB 99	UB 99
Капитал (независимые дефолты)	0.0456	0.0386	0.0541
Капитал (коррелированные дефолты)	0.1436	0.1092	0.1967
Коэффициент Джини	0.3642	0.3476	0.3815
Ожидаемая прибыль (доля)	0.0663	0.0498	0.0821

---

**Спасибо за внимание!**