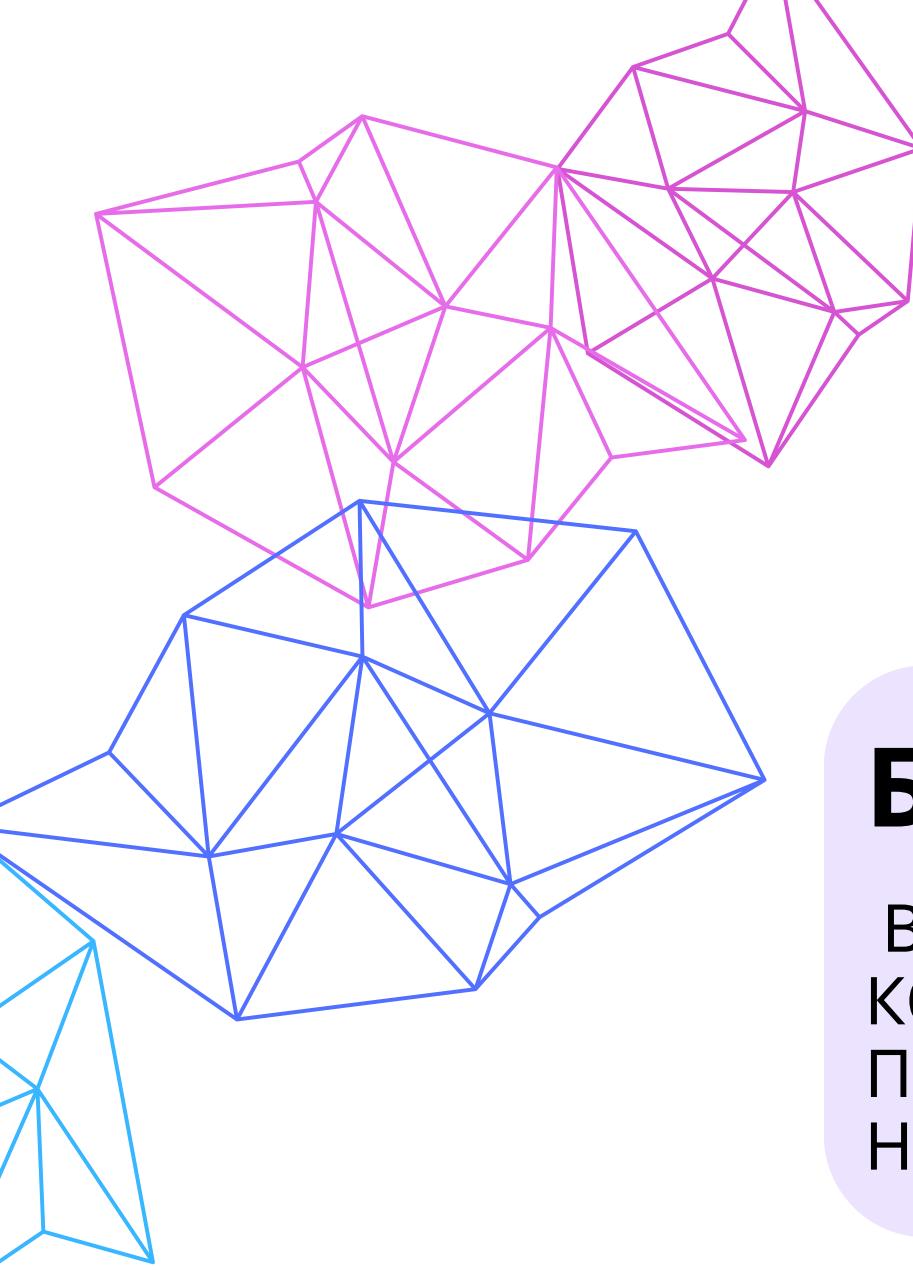


Larana, Inc.



# Models I'd like to Finetune



## НЕУВЕРЕННОСТЬ В ПОДРЯДЧИКАХ

НАЛИЧИЕ НЕНАДЁЖНЫХ ПОДРЯДЧИКОВ ПРИВОДИТ К ФИНАНСОВЫМ ПОТЕРЯМ И ЗАДЕРЖКАМ ПО ПРОЕКТУ

## БОЛЬШИЕ ОБЪЁМЫ ДАННЫХ

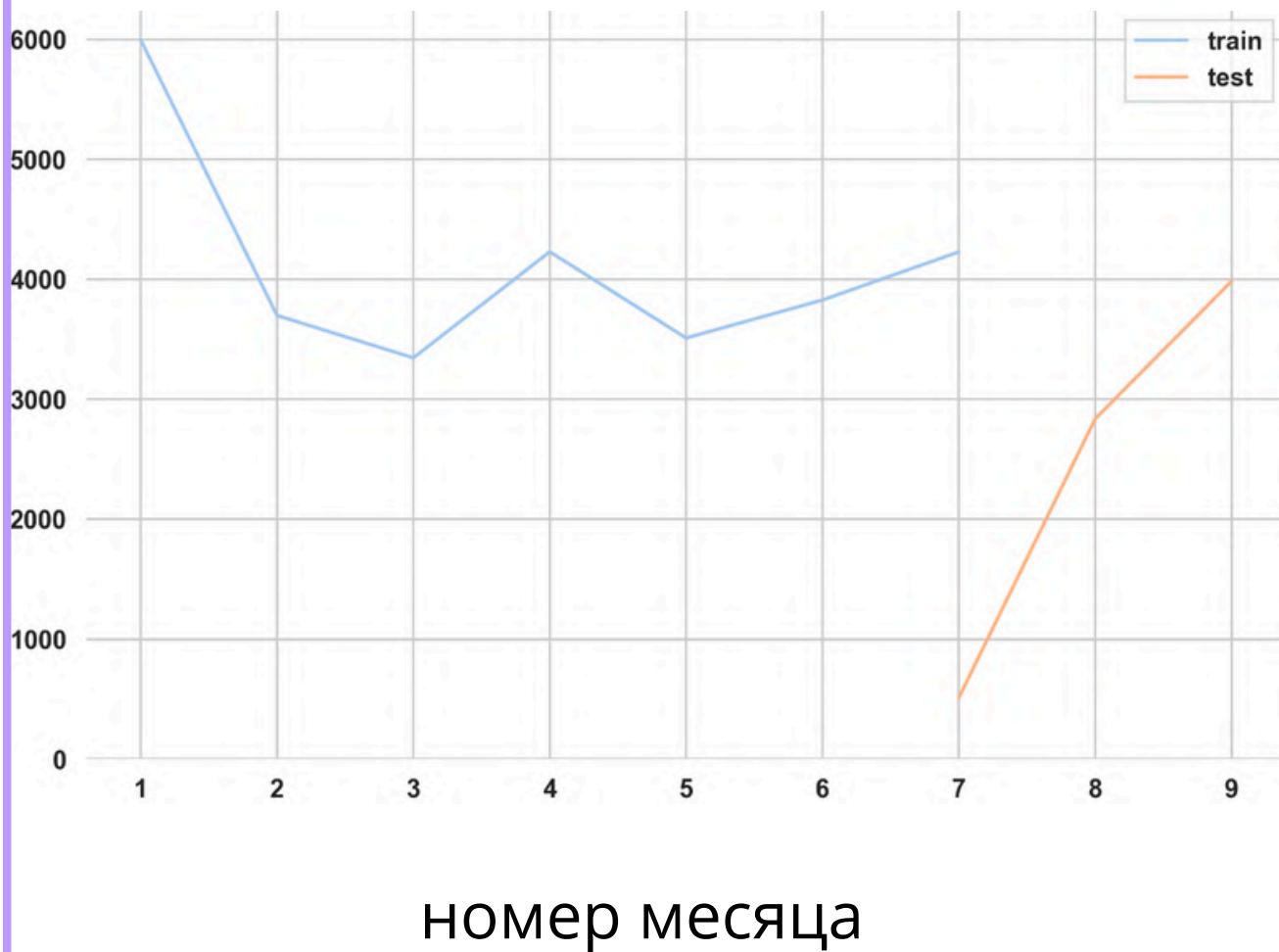
ВНУТРИ КОМПАНИИ ЕСТЬ МНОЖЕСТВО ДАННЫХ ПО ВЕДЕНИЮ КОНТРАКТОВ С РАЗЛИЧНЫМИ ПОДРЯДЧИКАМИ, ГДЕ ЕСТЬ ПОЛЕЗНЫЕ ПРИЗНАКИ, ТАКИЕ КАК НАПРИМЕР КОЛ-ВО РАБОЧИХ ОТ ПОДРЯДЧИКА НА ПЛОЩАДКЕ ИЛИ ИЗМЕНЕНИЕ СУММЫ КОНТРАКТА

## ПОТЕНЦИАЛ ПРИМЕНЕНИЯ МЛ

МОЖНО ПОСТРОИТЬ СКОРИНГ ПОДРЯДЧИКОВ ДЛЯ ВЫЯВЛЕНИЯ ТЕХ, КТО НЕ ВЫПОЛНЯЕТ СВОИ ОБЯЗАННОСТИ, ЧТОБЫ В ДАЛЬНЕЙШЕМ СНИЗИТЬ ВЕРОЯТНОСТЬ ДЕФОЛТА

# ПЕРВИЧНЫЙ АНАЛИЗ ДАННЫХ

РАСПРЕДЕЛЕНИЕ ПО МЕСЯЦАМ  
НАБЛЮДЕНИЙ В ТРЕЙНЕ/ТЕСТЕ



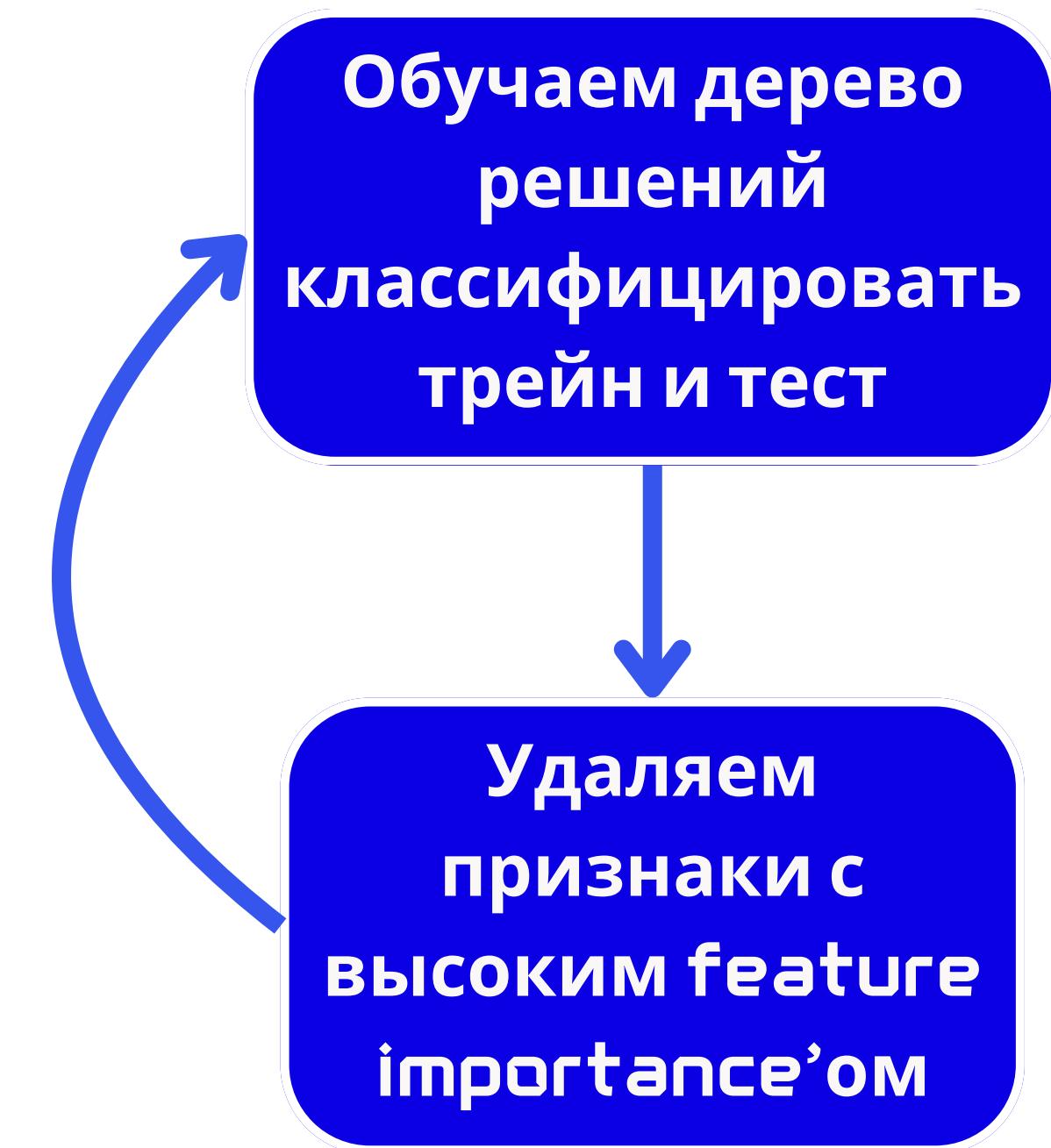
ПОЯВИЛАСЬ ГИПОТЕЗА, ЧТО  
ПРИЗНАКИ, СВЯЗАННЫЕ С  
НОМЕРОМ МЕСЯЦА, МОГУТ БЫТЬ  
СМЕЩЕНЫ ПО ОТНОШЕНИЮ К  
ТЕСТУ

НЕОБХОДИМО РАЗРАБОТАТЬ  
МЕТОДЫ ПОИСКА ПРИЗНАКОВ СО  
СМЕЩЕНИЕМ

# ПОИСК СМЕЩЕНИЯ ПО ТЕСТУ

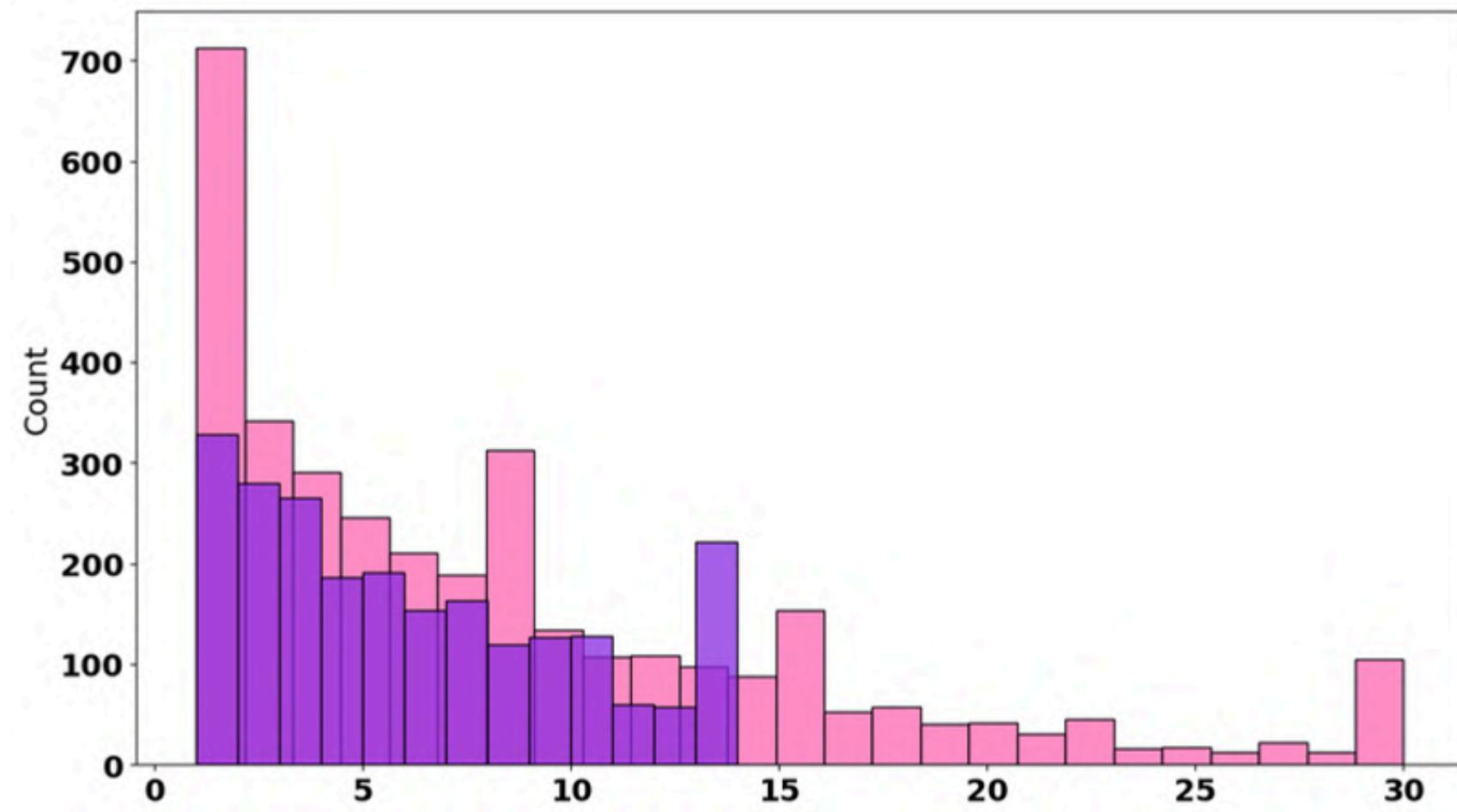
## ИДЕЯ

ОБУЧИМ НЕБОЛЬШУЮ МОДЕЛЬ  
(НАПРИМЕР, МАЛЕНЬКОЙ ДЕРЕВО РЕШЕНИЙ ГЛУБИНЫ),  
КЛАССИФИЦИРОВАТЬ ТРЕЙН И ТЕСТ.  
БУДЕМ ИТЕРАТИВНО УДАЛЯТЬ ПРИЗНАКИ,  
НА КОТОРЫЕ ЧАЩЕ ВСЕГО СМОТРИТ  
МОДЕЛЬ, ПОКА ТОЧНОСТЬ  
КЛАССИФИКАЦИИ НЕ БУДЕТ  
МИНИМАЛЬНОЙ. В ПЕРСПЕКТИВЕ, ЕСЛИ  
ИСПОЛЬЗОВАТЬ БУСТИНГОВЫЕ МОДЕЛИ,  
ТАКОЙ ПОДХОД ПОМОЖЕТ УВЕЛИЧИТЬ  
СТАБИЛЬНОСТЬ КАЖДОГО ДЕРЕВА



# НЕЯВНОЕ СМЕЩЕНИЕ

НА ТРЕЙНЕ ЕСТЬ ЧАСТЬ ГОРАЗДО БОЛЕЕ ДЛИННЫХ КОНТРАКТОВ, ЧЕМ НА ТЕСТЕ. ЭТО ВАЖНО УЧИТАВАТЬ ПРИ ГЕНЕРАЦИИ ПРИЗНАКОВ, А ТАКЖЕ УЧЕСТЬ, ЧТО НЕКОТОРЫЕ ПРИЗНАКИ В ПОСЛЕДНИХ ОТЧЁТАХ У БОЛЕЕ ДЛИННЫХ КОНТРАКТОВ МОГУТ ОТЛИЧАТЬСЯ



РАСПРЕДЕЛЕНИЕ ДЛИН  
КОНТРАКТОВ НА ТЕСТЕ И ТРЕЙНЕ

# АНАЛИЗ ЦЕЛЕВОЙ ПЕРЕМЕННОЙ

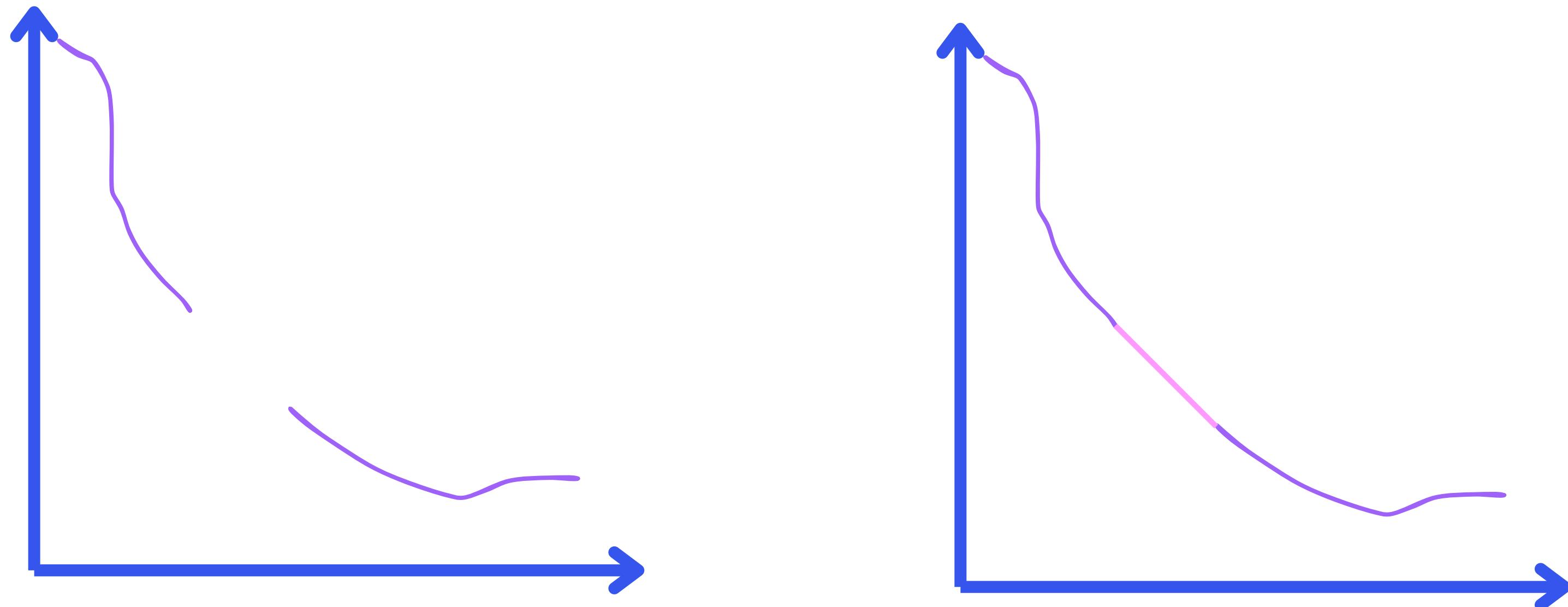
НА ОБУЧАЮЩЕЙ ВЫБОРКЕ ВСЕГО 2% СЛУЧАЕВ, В КОТОРЫХ ЦЕЛЕВАЯ ПЕРЕМЕННАЯ МЕНЯЕТСЯ “ПО ХОДУ” ОТЧЁТОВ В КОНТРАКТЕ. ПОЭТОМУ ИМЕЕТ СмысЛ ПОСТРОИТЬ БИНАРНУЮ МОДЕЛЬ, ОПРЕДЕЛЯЮЩУЮ СОСТОЯНИЕ КОНТРАКТА ЦЕЛИКОМ. Т.К. МЫ НЕ МОЖЕМ ГАРАНТИРОВАТЬ ТАКОЕ-ЖЕ РАСПРЕДЕЛЕНИЕ В ТЕСТЕ, ТПКУЮ МОДЕЛЬ БУДЕТ ЭФФЕКТИВНЕЕ ИСПОЛЬЗОВАТЬ ВМЕСТЕ С ДРУГИМИ, ПРЕДСКАЗЫВАЮЩИМИ ВЕРОЯТНОСТЬ СРЫВА КАЖДОГО ОТЧЁТА В КОНТРАКТЕ.



# КЛАССИФИКАЦИЯ КОНТРАКТА ЦЕЛИКОМ

КЛАССИФИКАЦИЯ ПО ВСЕМУ КОНТРАКТУ ОТКРЫВАЕТ БОЛЬШОЕ ПРОСТРАНСТВО ДЛЯ EDA: МЫ МОЖЕМ ГЕНЕРИРОВАТЬ МНОГО АГГРЕГИРОВАННЫХ ПРИЗНАКОВ ПО ВСЕМУ ОТЧЁТУ. ТАКЖЕ, ПОДОБНАЯ МОДЕЛЬ БУДЕТ НАМНОГО БОЛЕЕ УСТОЙЧИВА К ПЕРЕОБУЧЕНИЮ

# БОРЬБА С ПРОПУСКАМИ



ИСПОЛЬЗУЕМ ЛИНЕЙНУЮ ИНТЕРПОЛЯЦИЮ ДЛЯ ВРЕМЕННЫХ  
ПРИЗНАКОВ, В КОТОРЫХ ИЗВЕСТНА ЧАСТЬ ЗНАЧЕНИЙ. ДЛЯ ОСТАЛЬНЫХ  
ЗАМЕНИМ МЕДИАННОЙ, ЕСЛИ ПРОПУСКОВ ПО ЭТОМУ ПРИЗНАКУ  
МЕНЬШЕ 60%.

# EDA

Бинаризация  
временных  
признаков

Нормализация  
константных  
признаков

Производные  
временных  
признаков

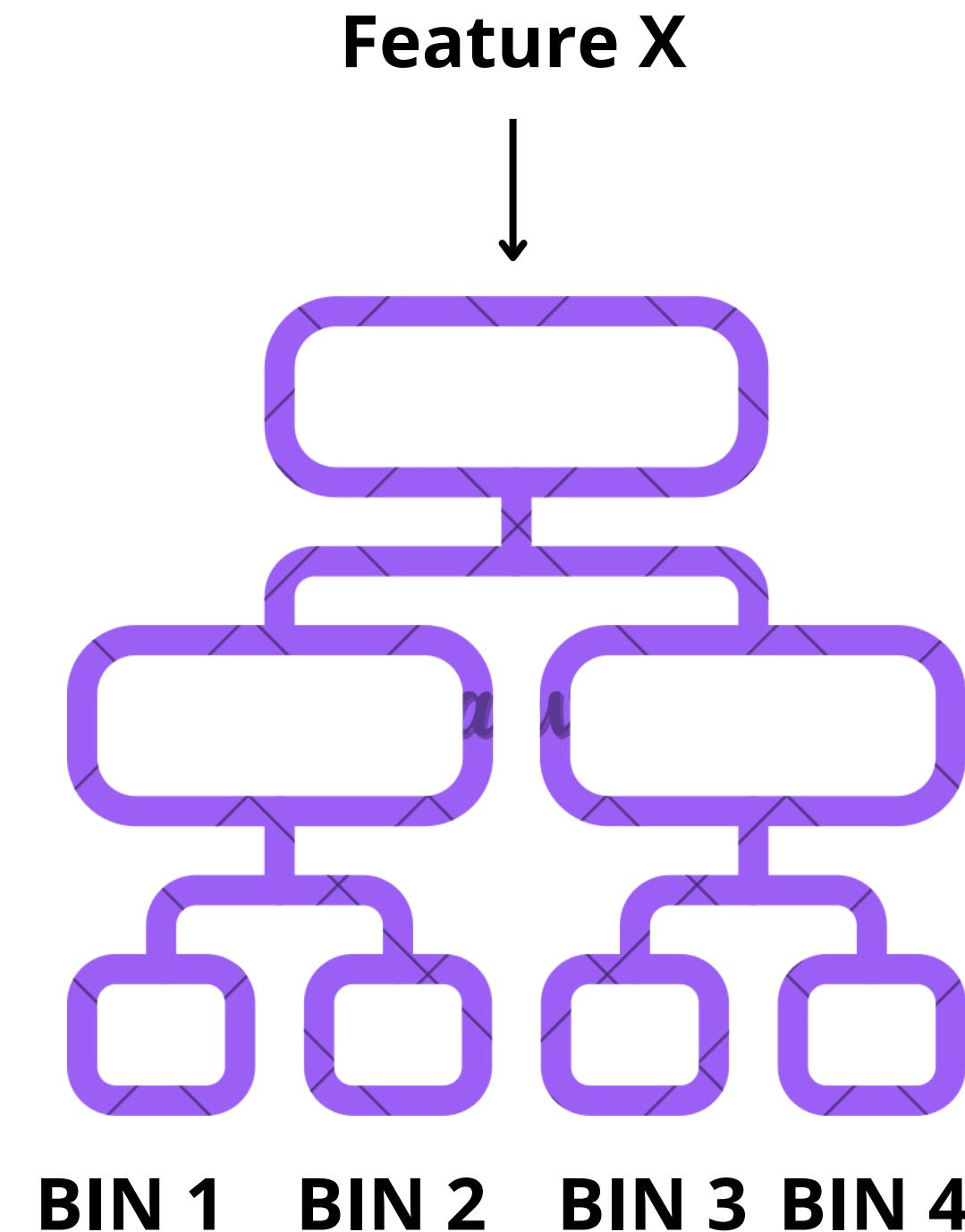
**TSFRESH** для  
временных  
признаков

Графовые признаки  
[о них чуть позже]

# БИНАРИЗАЦИЯ ПЕРЕМЕННЫХ

для каждой числовой фичи мы:

- 1.** Строим дерево решений на предсказания таргета по фиче
- 2.** Далее смотрим для каждого сэмпла в какой он лист попадает
- 3.** Это позволяет нам теоретически разбить на более выгодные бины, относительно задачи минимизации энтропии целевого класса

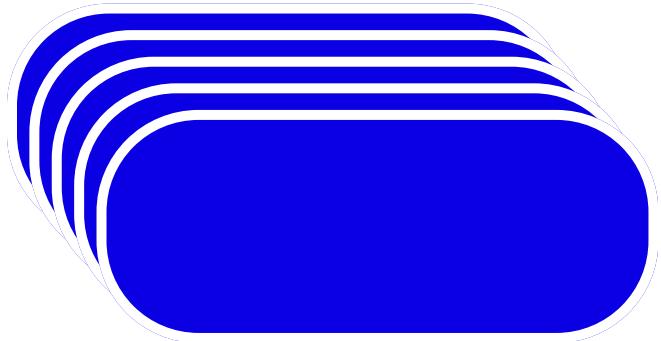


# МОДЕЛЬ

КЛАССИФИКАЦИЯ КОНТРАКТОВ

## TimeseriesStratified5Fold

Предобработанные  
данные отчётом

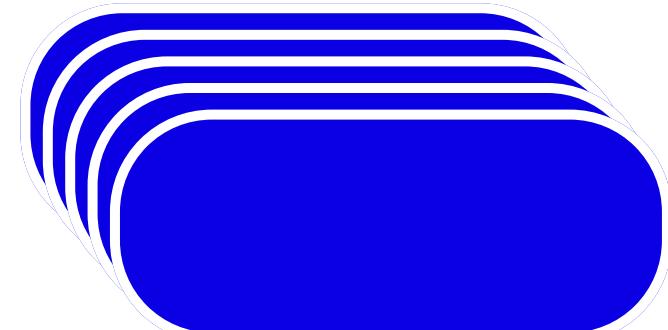


Optuna



модель/модели

## 5 моделей



Взвешенное усреднение по фолдам

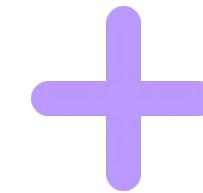
На каждом фолде:

# КЛАССИФИКАЦИЯ ОТЧЁТОВ

ОСНОВНАЯ ПРОБЛЕМА КЛАССИФИКАЦИИ ПО КАЖДОМУ ОТЧЁТУ -  
ПЕРЕОБУЧЕНИЕ, Т.К. ОТЧЁТЫ ВНУТРИ КОНТРАКТА СХОЖИ МЕЖДУ  
СОБОЙ, ЧТО СИЛЬНО СУЖАЕТ ДОМЕН ОБУЧАЮЩЕЙ ВЫБОРКИ

## РЕШЕНИЕ

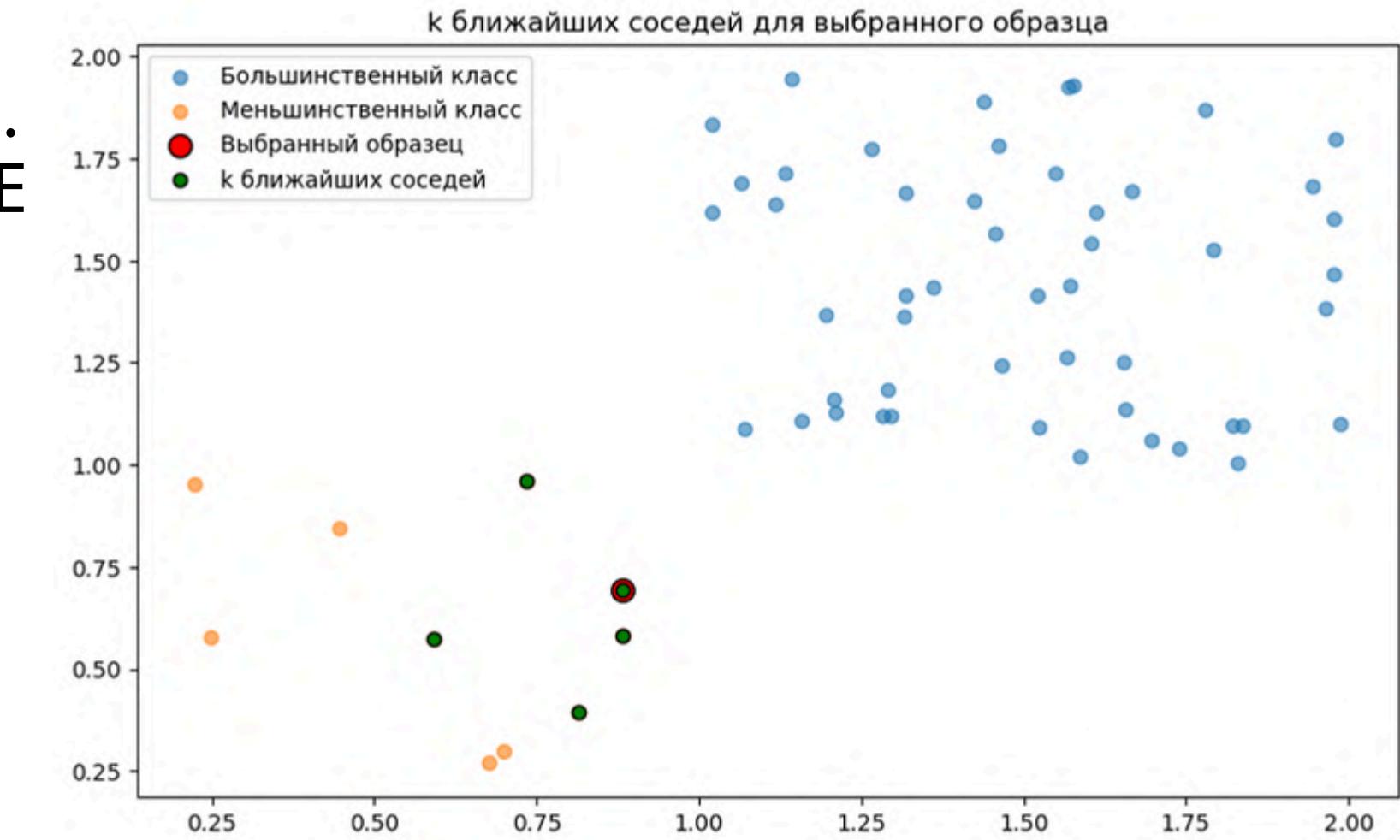
**StratifiedGroup  
Kfold validation**



**продвинутый  
OVERSAMPLING**

# SMOTENC

OVERSAMPLING - ТЕХНОЛОГИЯ РАСШИРЕНИЯ ОБУЧАЮЩЕЙ ВЫБОРКИ.  
SMOTE - МЕТОД, ПРИ КОТОРОЙ НОВЫЕ СЕМПЛЫ В ДАННЫХ ГЕНЕРИРУЮТСЯ С ПОМОЩЬЮ УСРЕДНЕНИЯ К-БЛИЖАЙШИХ СОСЕДЕЙ ОБЪЕКТОВ ТОГО-ЖЕ КЛАССА.  
SMOTENC - ПРОДВИНУТАЯ ВЕРСИЯ SMOTE, УЧИТЫВАЮЩАЯ КАТЕГОРИАЛЬНЫЕ ПРИЗНАКИ, ЧТО АКТУАЛЬНО ДЛЯ НАШЕЙ ЗАДАЧИ.

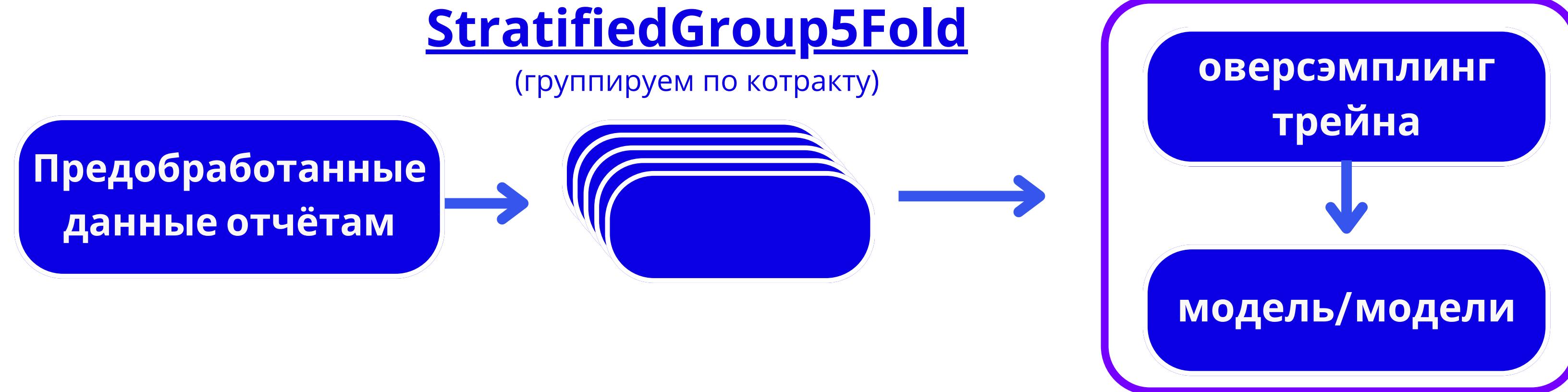


ВО ВРЕМЯ ОБУЧЕНИЯ МЫ ПРИМЕНЯЕМ OVERSAMPLING ВНУТРИ ОБУЧАЮЩЕЙ ВЫБОРКИ КАЖДОГО ФОЛДА, ЧТО ПОЗВОЛЯЕТ ИЗБЕЖАТЬ СИТУАЦИИ, КОГДА НОВЫЕ СЕМПЛЫ ГЕНЕРИРУЮТСЯ НА ОСНОВЕ СЕМПЛОВ ИЗ ВАЛИДАЦИИ

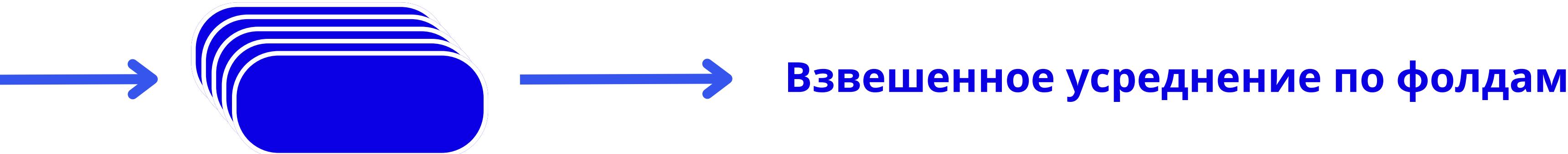
# БАЗОВЫЙ ПАЙПЛАЙН

КЛАССИФИКАЦИЯ ОТЧЁТОВ

На каждом фолде:



5 моделей



# МОДЕЛИ. БУСТИНГИ

CatBoost

LightGBM

XGBoost

+

Optuna

+

Iterative feature selection on CV

&

Shap values

# ГРАФОВЫЕ ПРИЗНАКИ

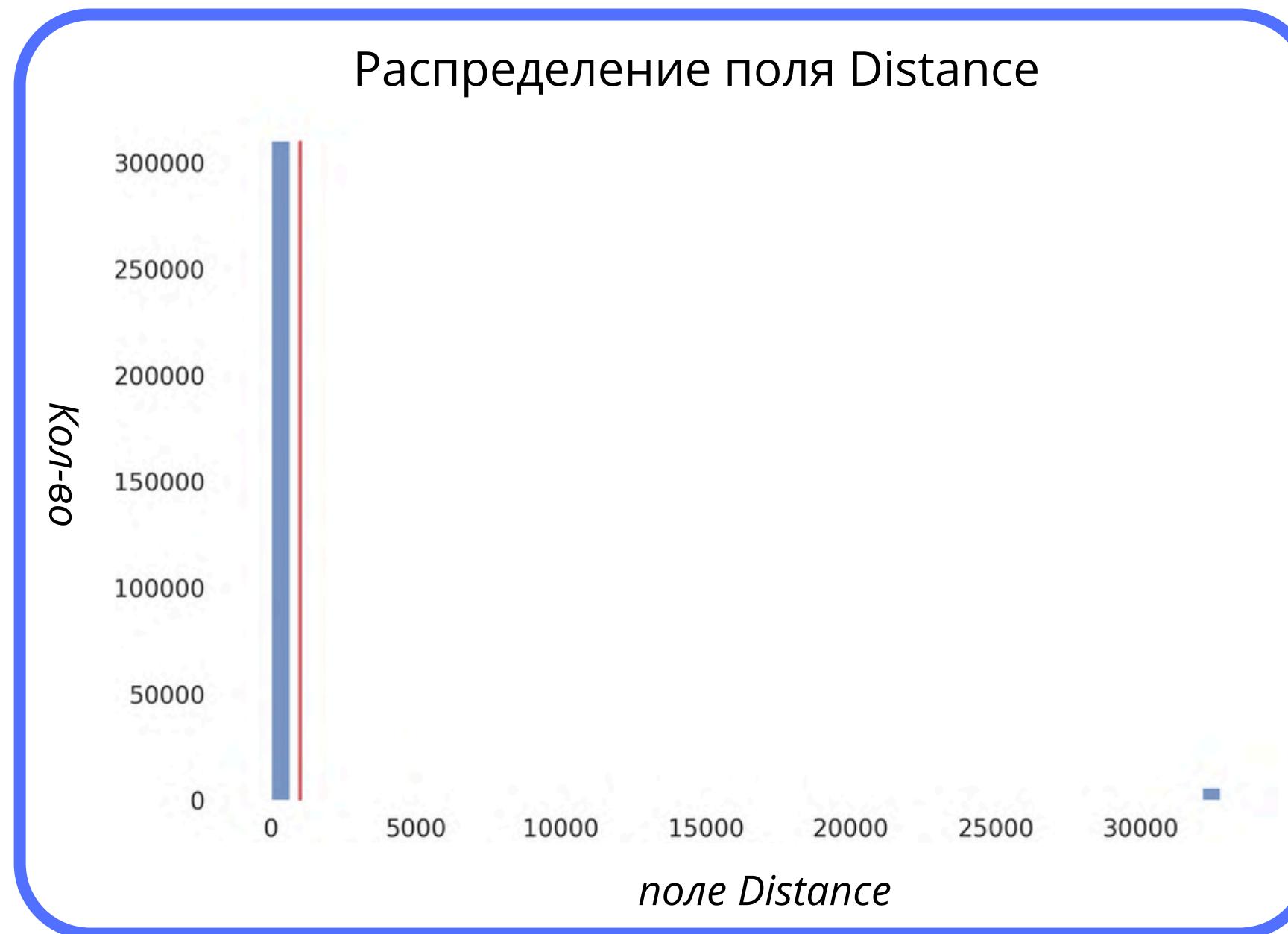
ПОСТАВЩИКИ В ОБУЧАЮЩЕЙ И ТЕСТОВОЙ ВЫБОРКЕ ЧАСТИЧНО  
ПЕРЕСЕКАЮТСЯ, ЭТО МОЖЕТ ГАРАНТИРОВАТЬ ЭФФЕКТИВНОСТЬ  
КАЧЕСТВЕННО СГЕНЕРИРОВАННЫХ ГРАФОВЫХ ПРИЗНАКОВ.

# Граф поставщиков

Кол-во поставщиков в трейне/тесте: **847**

Кол-во поставщиков в графе: **803**

Кол-во поставщиков, которые есть и в трейне/тесте, и в графе: **764**



Удалили выбросы

# Генерация графовых фичей

*Меры центральности вершин*

Betweenness Centrality

Closeness Centrality

Eigenvector Centrality

PageRank

And other...

*Признаки соседей*

Mean neighbors features

Neighbors ids

Mean edge features

Averaging target

And other...

*Другие признаки*

Mean edge centrality

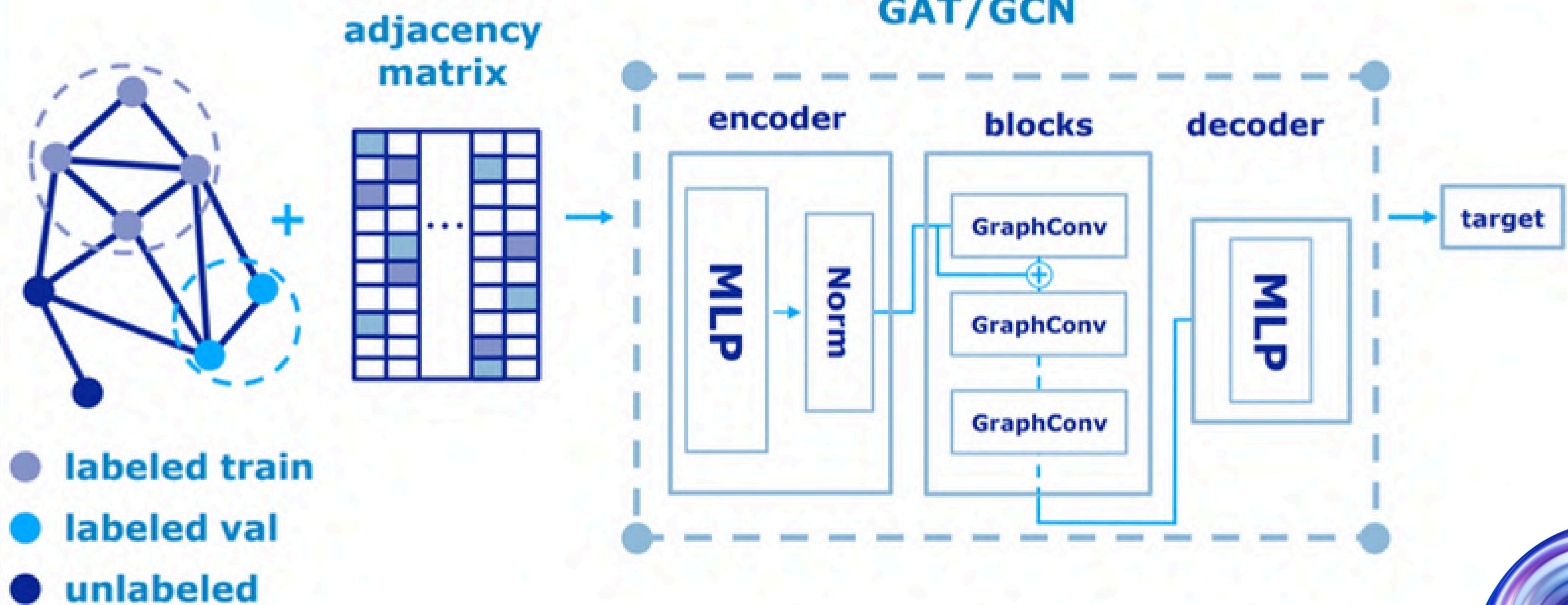
Clustering coef

Triangles in graph

And other...

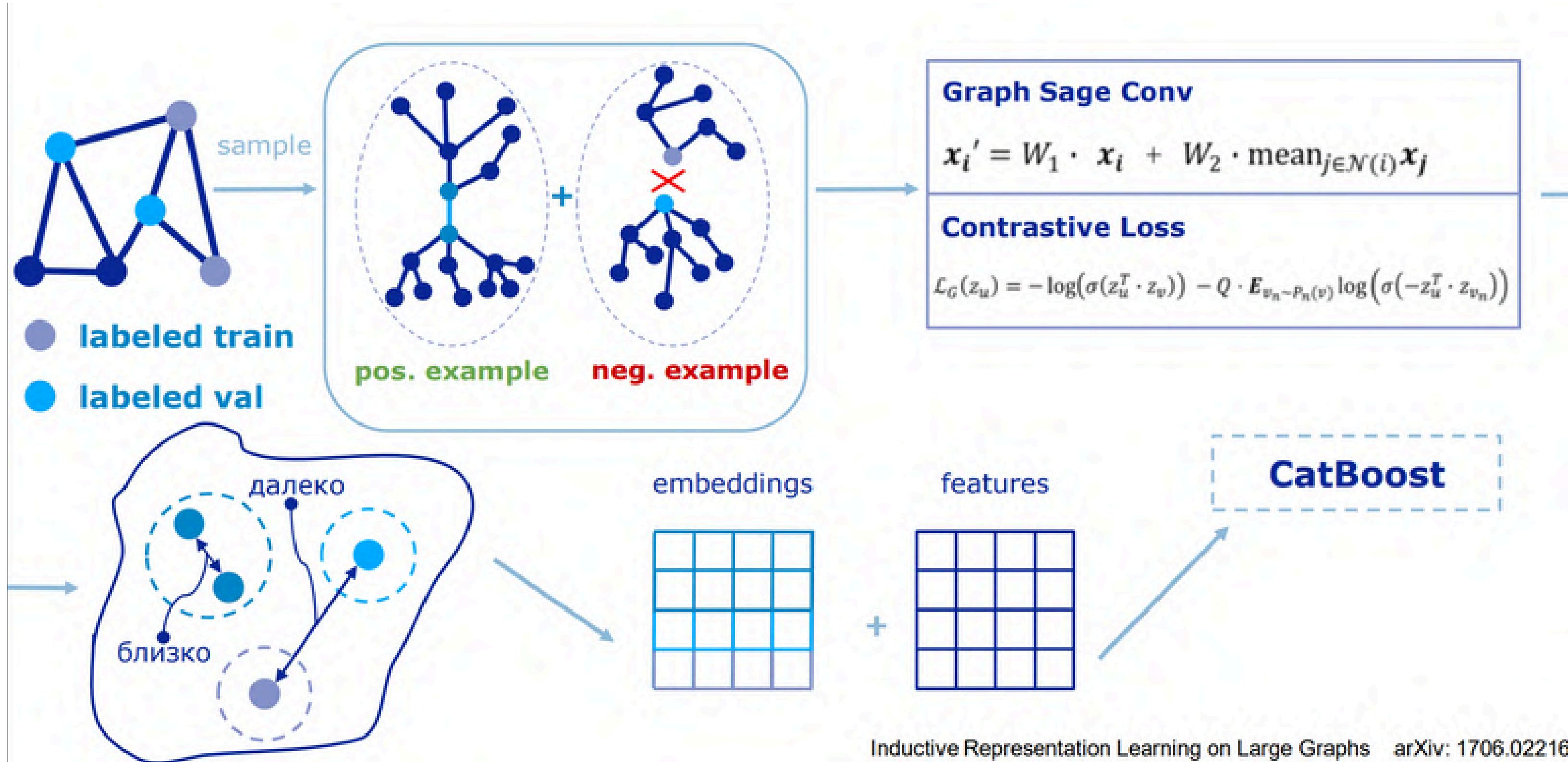
# Продвинутые фичи

генерация эмбеддингов с помощь supervised transductive learning on GAT/GCN



# Продвинутые фичи

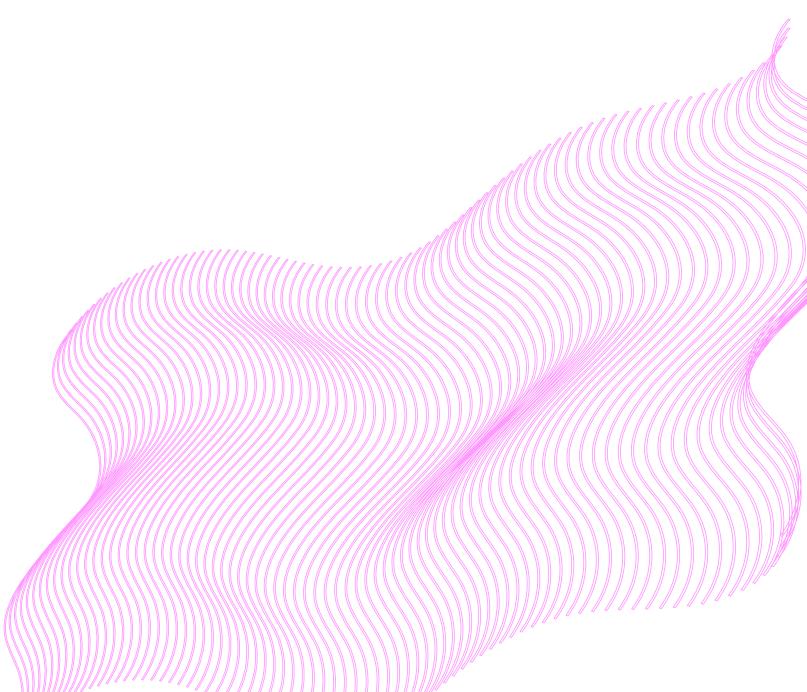
генерация эмбеддингов с помощь unsupervised inductive learning with GraphSage



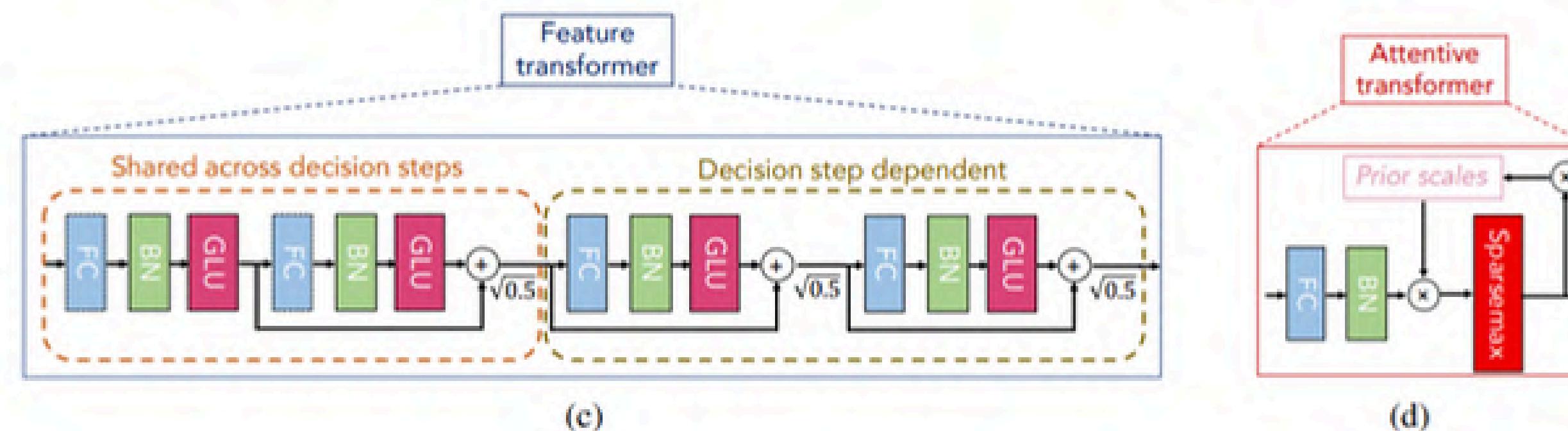
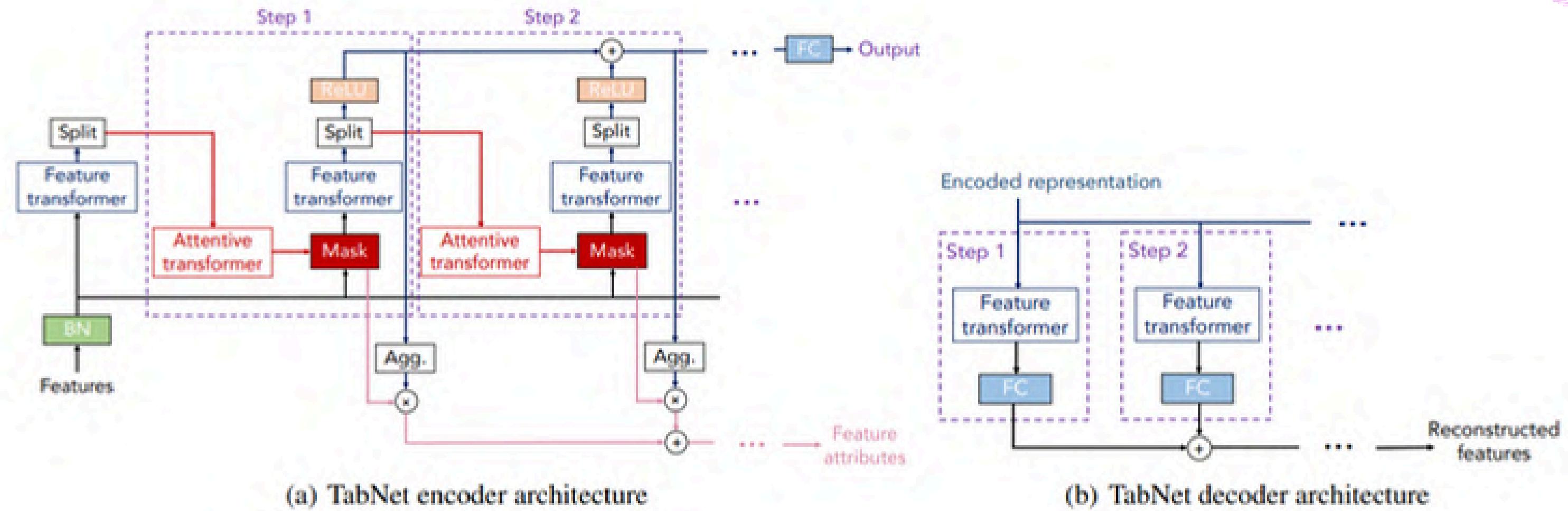


# **КАКИЕ ЕЩЁ ПОДХОДЫ БЫЛИ ОПРОБОВАНЫ**

НО ОНИ ПОКАЗАЛИ БОЛЕЕ НИЗКУЮ ТОЧНОСТЬ



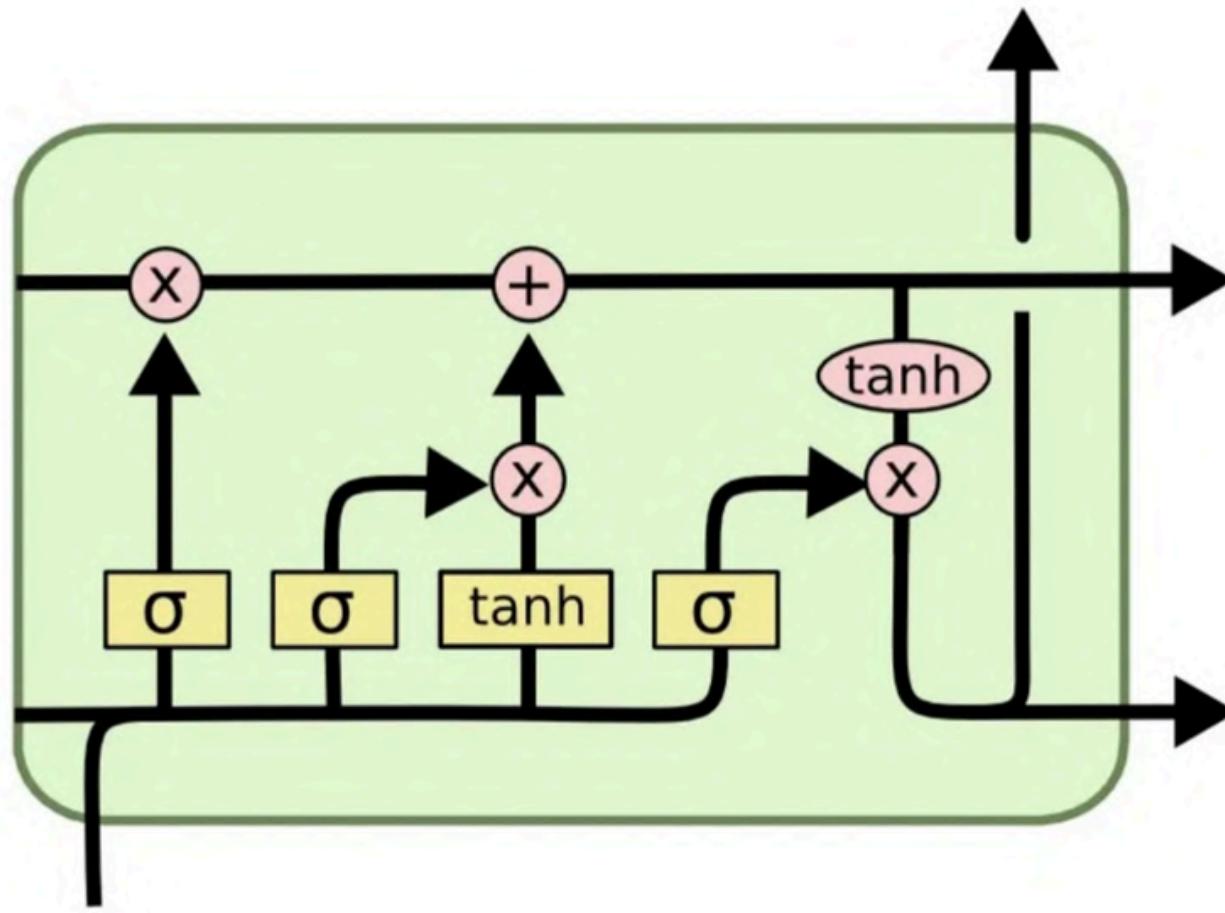
# TABNET



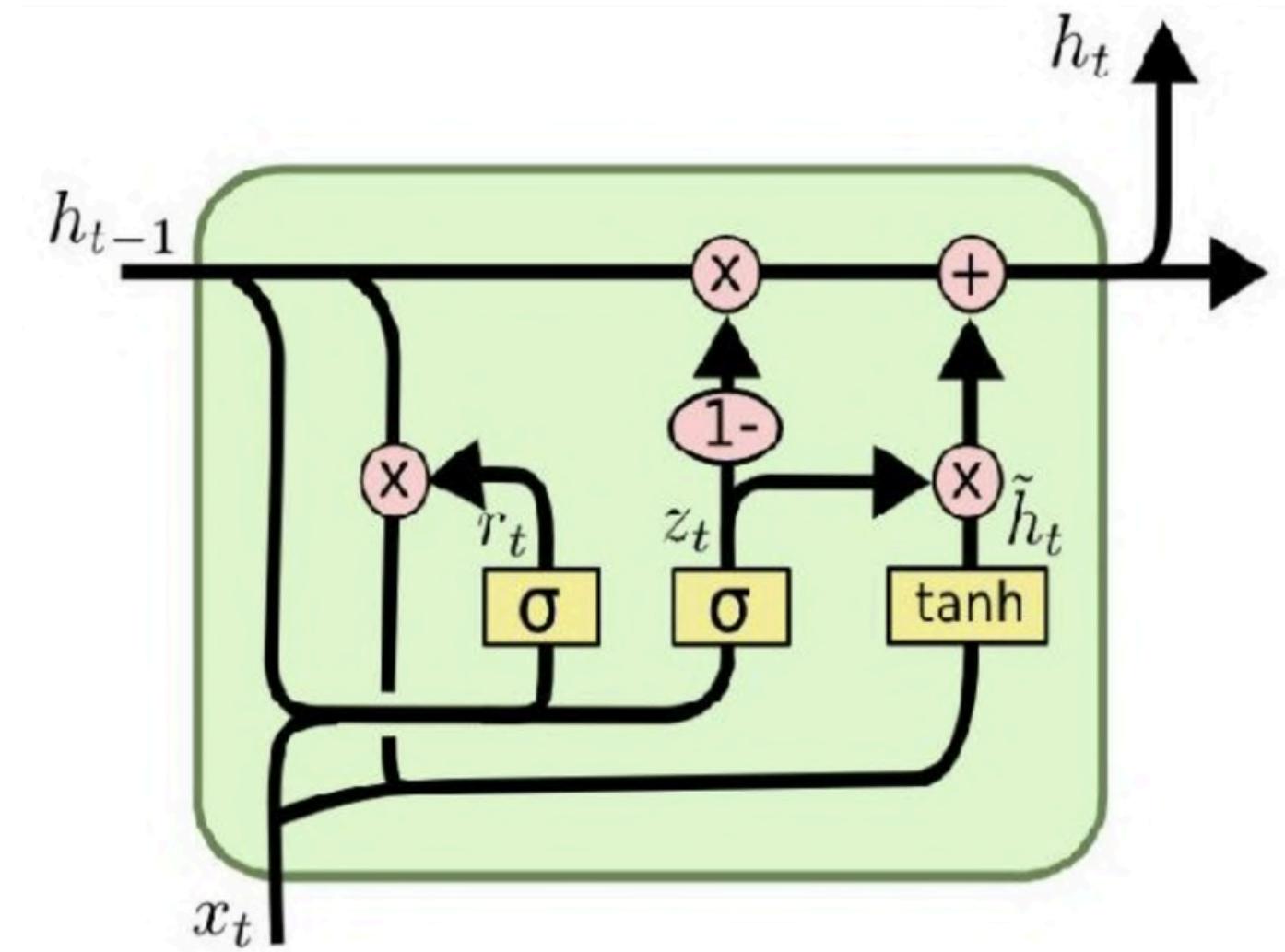
# RNNs

(анализ временных последовательностей с помощью DL)

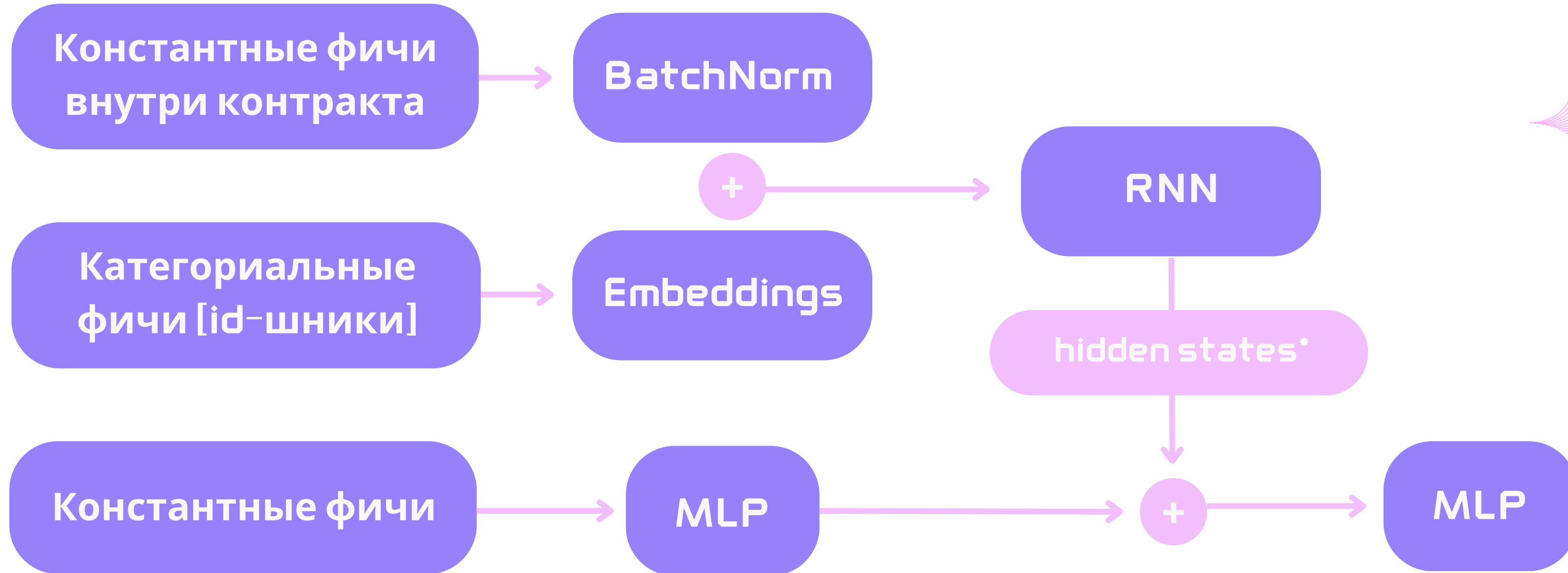
## LSTM



## GRU



# RNNs. Модель классификации отчётов



- \* • В случае классификации отчётов берутся все hidden-state'ы
- В случае классификации контрактов берётся последний hidden-state

# Что ещё попробовали?

## 1. PytorchLifestream(CoLES)

- *unsupervised* эмбеддинги последовательностей с помощью по-сущи *Contrastive Learning*

## 2. MLP + эмбеддинги из неё

## 3. Batch Training для борьбы с переобучением

1 Batch - 1й элемент из каждой последовательности

2 Batch - 2й элемент из каждой последовательности

i Batch - i-ый элемент из каждой последовательности

## 4. Псевдолейблинг



# **СРАВНЕНИЕ МЕТОДИК**

# Результаты. Сравнение моделей

	Классификация отчётов	Классификация контрактов
Бустинги	<b>0.477</b>	<b>0.502</b>
RNNs	<b>0.28</b>	<b>0.33</b>



ВСЕ РЕЗУЛЬТАТЫ - МЕТРИКА НА ЛИДЕРБОРДЕ

# Результаты. Классификация отчётов

Графы	Табличные фичи	Другое	
Эмбеддинги из unsupervised/unsupervised подхода	Лучшие эмпирические фичи	Добавление лучших фичей	Feature Selection
			0.44
			0.502
			0.42
			0.45

\* ВСЕ ЭКСПЕРИМЕНТЫ ПРОВЕДЕНЫ НА БУСТИНГАХ

# Результаты. Классификация контрактов

Графы	Табличные фичи	Другое	
Эмбеддинги из unsupervised/unsupervised подхода	Лучшие эмпирические фичи	Добавление лучших фичей	Feature Selection
			0.44
			0.477
			0.42
			0.45

\* ВСЕ ЭКСПЕРИМЕНТЫ ПРОВЕДЕНЫ НА БУСТИНГАХ

# BLENDING

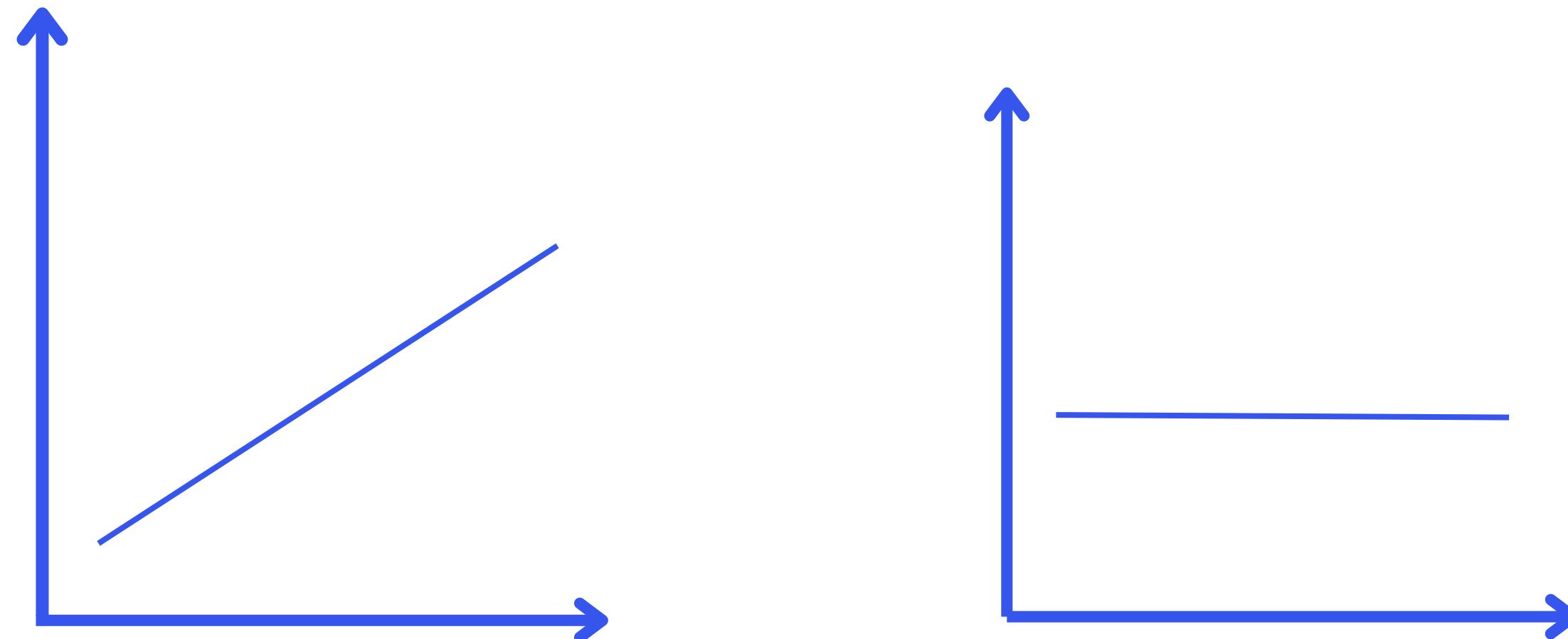
Усреднение по  
фолдам с весами

+

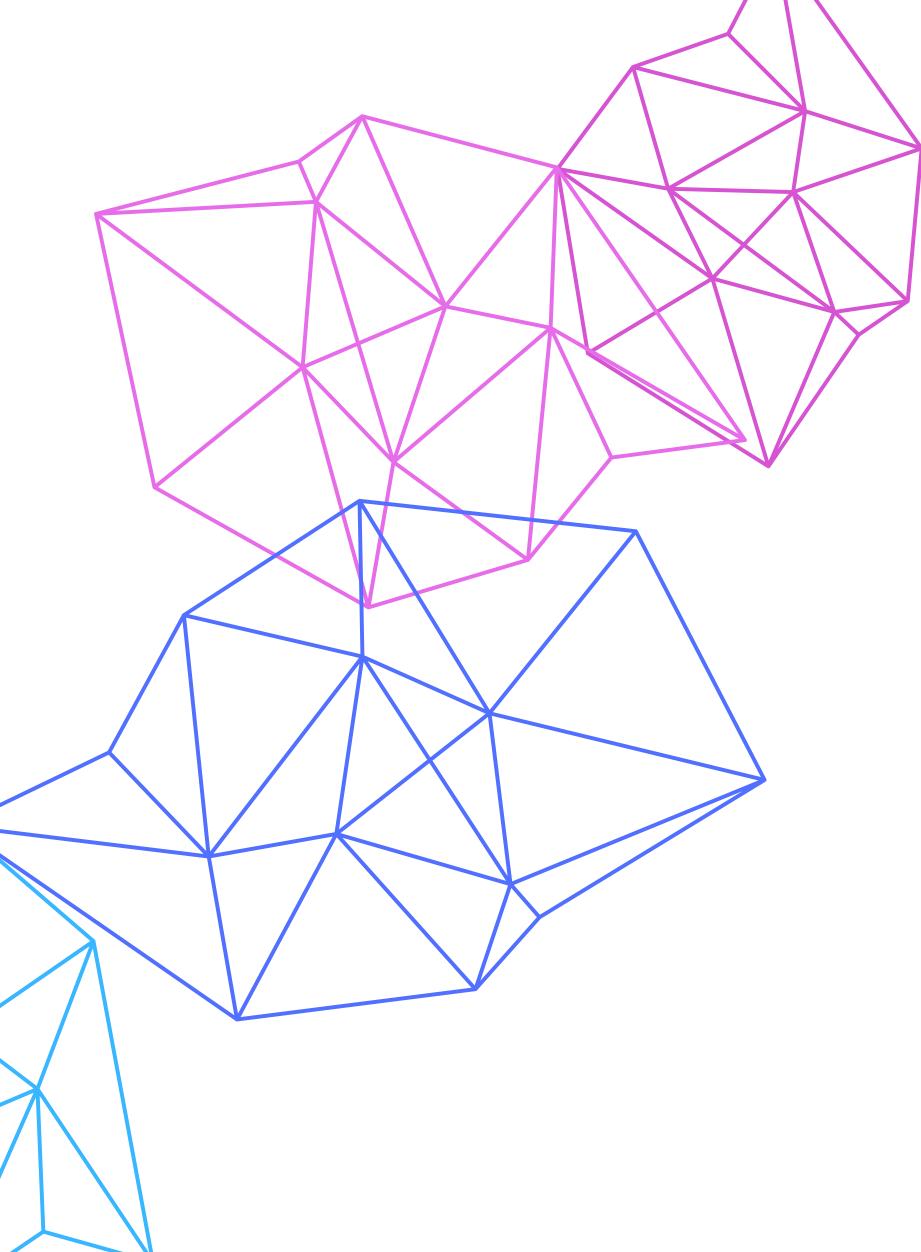
Пайплайн  
классификации  
контрактов

+

Пайплайн  
классификации  
отчётов



Итоговый скор: **0.5208**



# **ИНТЕРПРЕТАЦИЯ**

**= BUSINESS VALUE**



# Идея

## Интерпретация

Интерпретация,  
которая будет  
полезна всегда

Интерпретация на  
основе константных  
признаков

Интерпретация на  
основе изменения  
временных  
признаков

Индекс  
доверенности к  
поставщику

Анализ похожих контрактов по  
признаках, показывающим  
высокий индекс Шепли

Сегментация промежутков  
контракта с наиболее  
характерными для срыва  
контракта участками

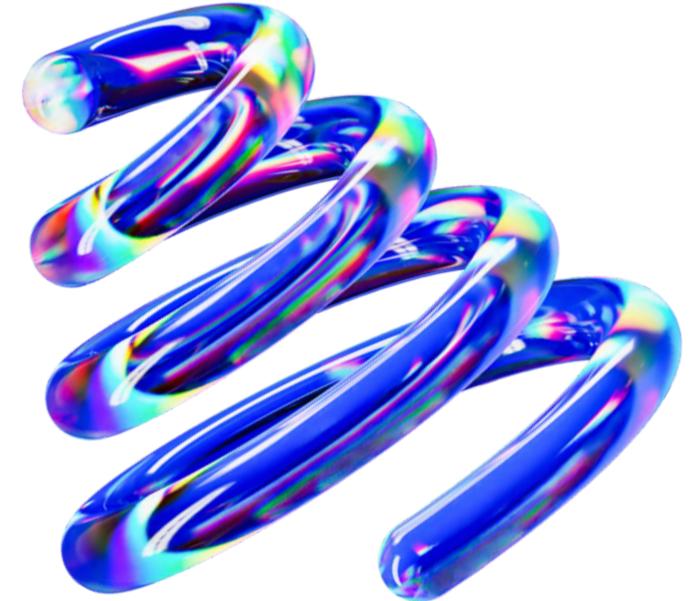
Индекс  
“безопасности”  
специализации

Модель  
предсказания  
срыва по одному  
контракту

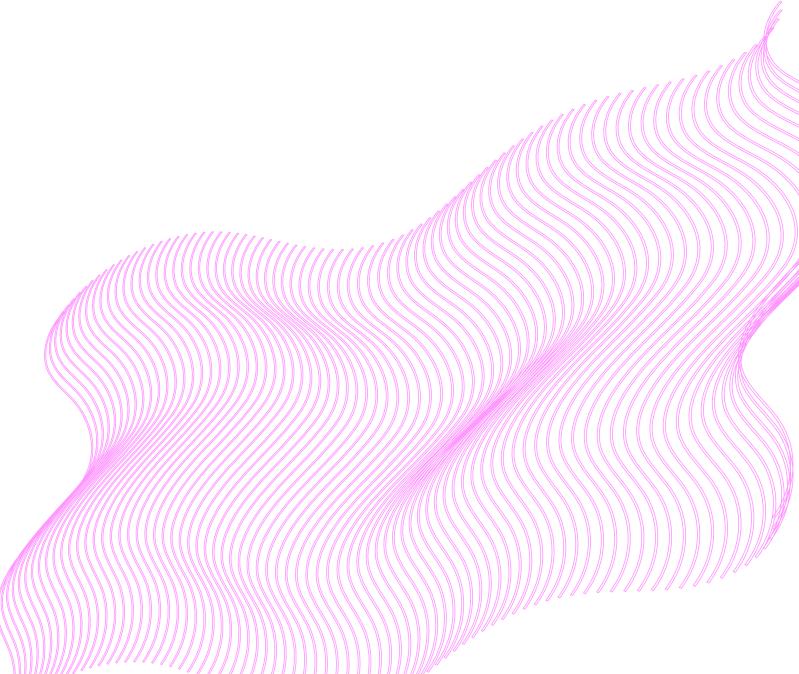
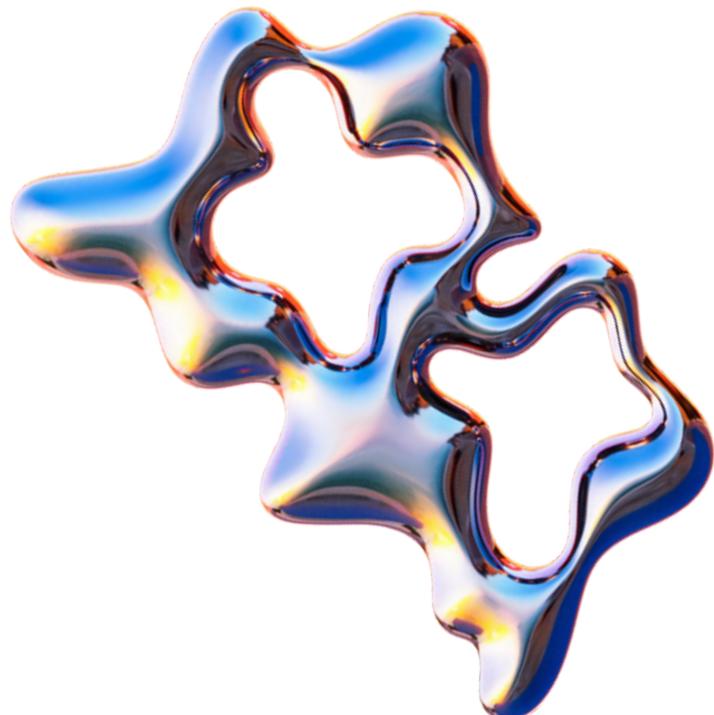
Выделение временных  
признаков, наиболее вляющих  
на повышение вероятности  
срыва

Модель  
предсказания  
срыва по всему  
контракту

Интерпретация  
влияния временных  
изменений  
признаков на  
повышения  
вероятности срыва



**МЫ УЧИТЫВАЕМ НЕ ТОЛЬКО ИНДЕКС ШЕПЛИ  
ПРИЗНАКОВ В МОМЕНТЕ, НО И ИХ ИЗМЕНЕНИЕ И  
ТОПОЛОГИЮ ВО ВРЕМЕНИ**



# Интерпритация

## Индекс доверия поставщику



$$\text{contractor\_index}_u = \text{contractor\_pagerank}_u * \frac{\sum_{v \in N(u)} (D(u, v) * \text{contractor\_success\_rate}_v))}{\sum_{v \in N(u)} D(u, v)} + \text{contractor\_success\_rate}_u$$

- 1.** Учитывает успешность выполнения заказов поставщиком
- 2.** Также учитывает окружение поставщика и его свойства

Позволяет оценить “ответственность” поставщика  
на начале выполнения работ



# Интерпритация

## “Освоенность” специализации

ГИПОТЕЗА В ТОМ, ЧТО РЕДКИЕ СПЕЦИАЛИЗАЦИИ  
СИЛЬНО ОТЛИЧАЮТСЯ ПО ДРУГИМ ПРИЗНАКАМ И МОГУТ УХУДШАТЬ  
СТАБИЛЬНОСТЬ ПРЕДСКАЗАНИЙ, ПОЭТОМУ ИХ НУЖНО ОТСЛЕЖИВАТЬ  
БОЛЕЕ ВНИМАТЕЛЬНО

### 1 РЕДКИЕ СПЕЦИАЛИЗАЦИИ

- СИЛЬНО ОТЛИЧАЮТСЯ ПО ДРУГИМ ПРИЗНАКАМ И МОГУТ УХУДШАТЬ СТАБИЛЬНОСТЬ ПРЕДСКАЗАНИЙ

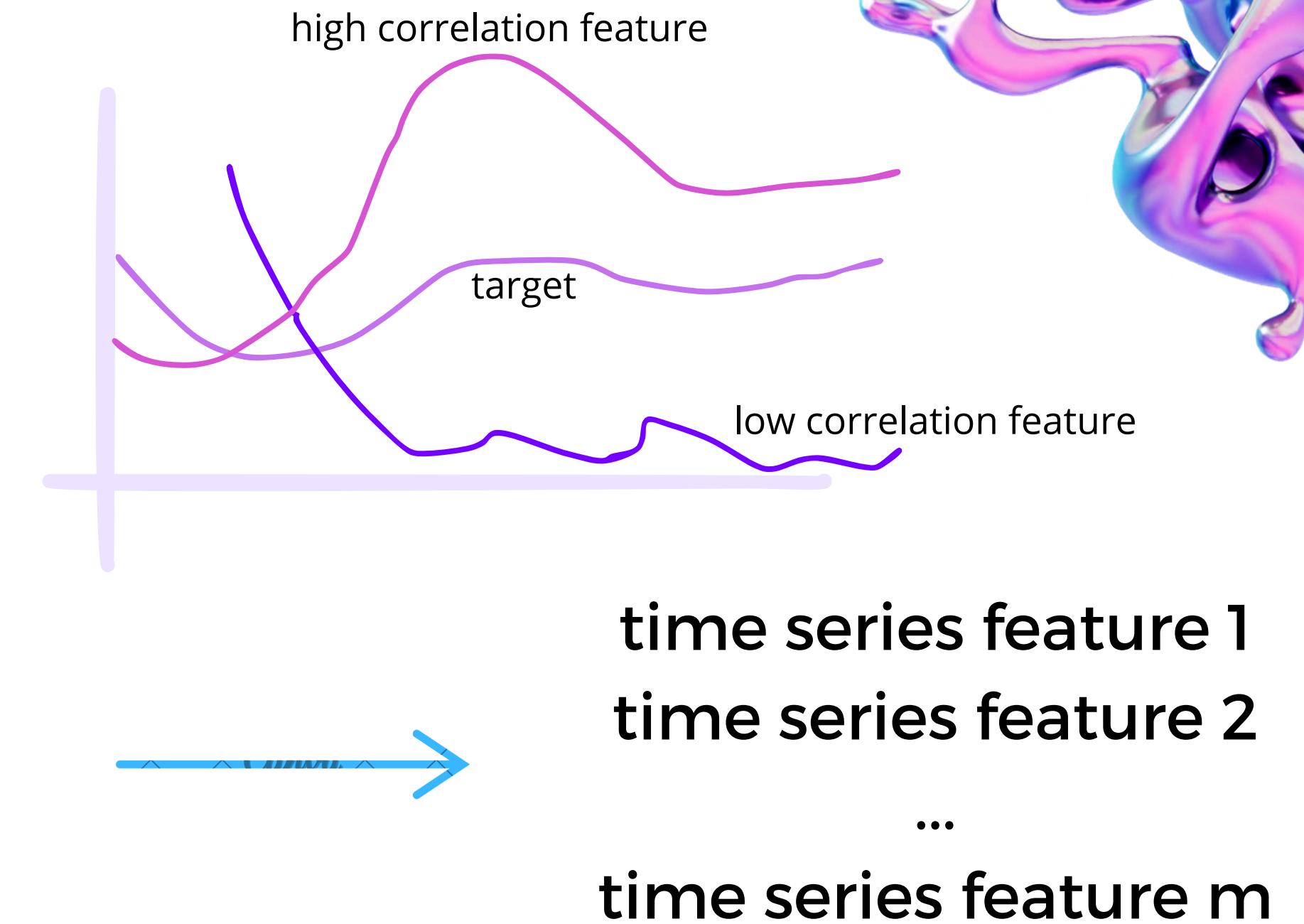
### 2 Исполнители хуже работают с более редкими специализациями, их навыки выполнения таких работ • сложнее оценить

Позволяет увидеть некоторые особенности  
относительно специализаций

# Интерпритация

## Временные признаки

Predict proba	TIME SERIES FEATURE 1	...	TIME SERIES FEATURE N
0.45	shap value2	....	shap value n
...	...	....	
0.89	shap value	....	



отбираем только те признаки, shap values которых коррелируют с вероятностью внутри каждого контракта

# Интерпритация

## Временные признаки

1 sample	2 sample	....	i sample
2 sample	<code>sum([dist[i] - dist[k] for k in range(1, j)])</code>	....	<code>sum([dist[i] - dist[k] for k in range(1, j)])</code>
...	<code>sum([dist[i] - dist[k] for k in range(1, j)])</code>	....	<code>sum([dist[i] - dist[k] for k in range(1, j)])</code>
j sample	<code>sum([dist[i] - dist[k] for k in range(1, j)])</code>	<code>sum([dist[i] - dist[k] for k in range(1, j)])</code>	<code>sum([dist[i] - dist[k] for k in range(1, j)])</code>

ВНУТРИ ОДНОГО КОНТРАКТА  
СЧИТАЕМ МАТРИЦУ С СУММАРНЫМ  
ИЗМЕНЕНИЕМ ВЕРОЯТНОСТИ СРЫВА  
ДЛЯ ВСЕХ ВОЗМОЖНЫХ  
ПРОМЕЖУТКОВ

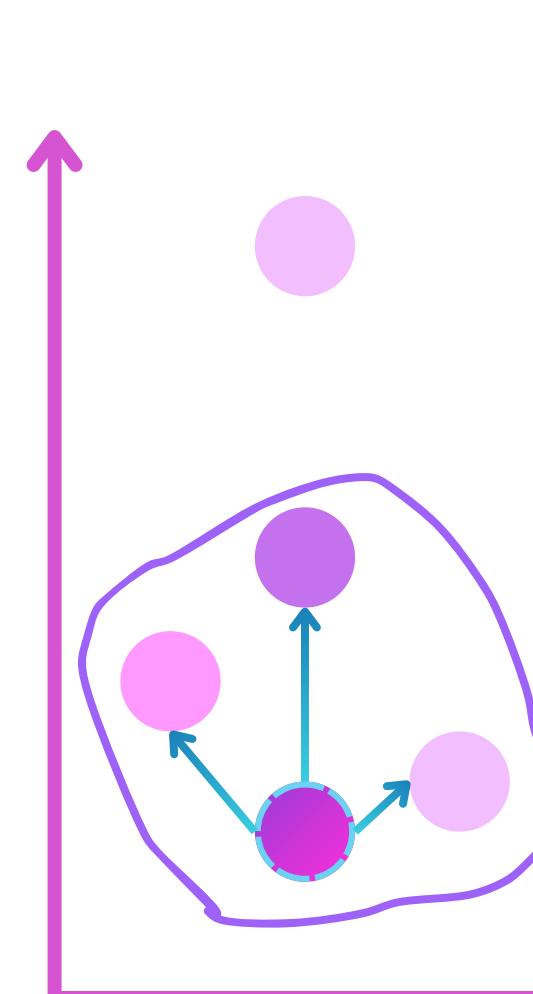
НАХОДИМ ПРОМЕЖУТКИ, В  
КОТОРЫХ СИЛЬНЕЕ ВСЕГО  
МЕНЯЛАСЬ ВЕРОЯТНОСТЬ СРЫВА  
КОНТРАКТА

- ГИПОТЕЗА В ТОМ, ЧТО ДЛЯ ПОИСКА ГЛУБОКИХ ИНСАЙТОВ НАМ ВАЖНЫ НЕ ПРОСТО ПРОМЕЖУТКИ С МАКСИМАЛЬНОЙ ВЕРОЯТНОСТЬЮ СРЫВА КОНТРАКТА, А ПРОМЕЖУТКИ, НА КОТОРЫХ ВЕРОЯТНОСТЬ СИЛЬНЕЕ ВСЕГО ИЗМЕНИЛАСЬ.

# Интерпритация

## Временные признаки

ВЫЧИСЛЯЕМ АГГРЕГИРОВАННЫЕ ПРИЗНАКИ ДЛЯ ВЫБРАННЫХ СЕГМЕНТОВ



НАХОДИМ ПОХОЖИЕ СЕГМЕНТЫ ИЗ ИСТОРИЧЕСКОЙ ВЫБОРКИ, СЧИТАЕМ ИХ ВЕРОЯТНОСТИ СРЫВА

ИНТЕРПРЕТИРУЕМ НА ОСНОВЕ АГГРЕГИРОВАННОГО ПРИЗНАКИ

Модель предсказания срыва по всему контракту

ВЫБИРАЕМ ПРИЗНАК С САМЫМ ВЫСОКИМ ИНДЕКСОМ ШЕПЛИ СРЕДИ СОРВАВШИХСЯ КОНТРАКТОВ

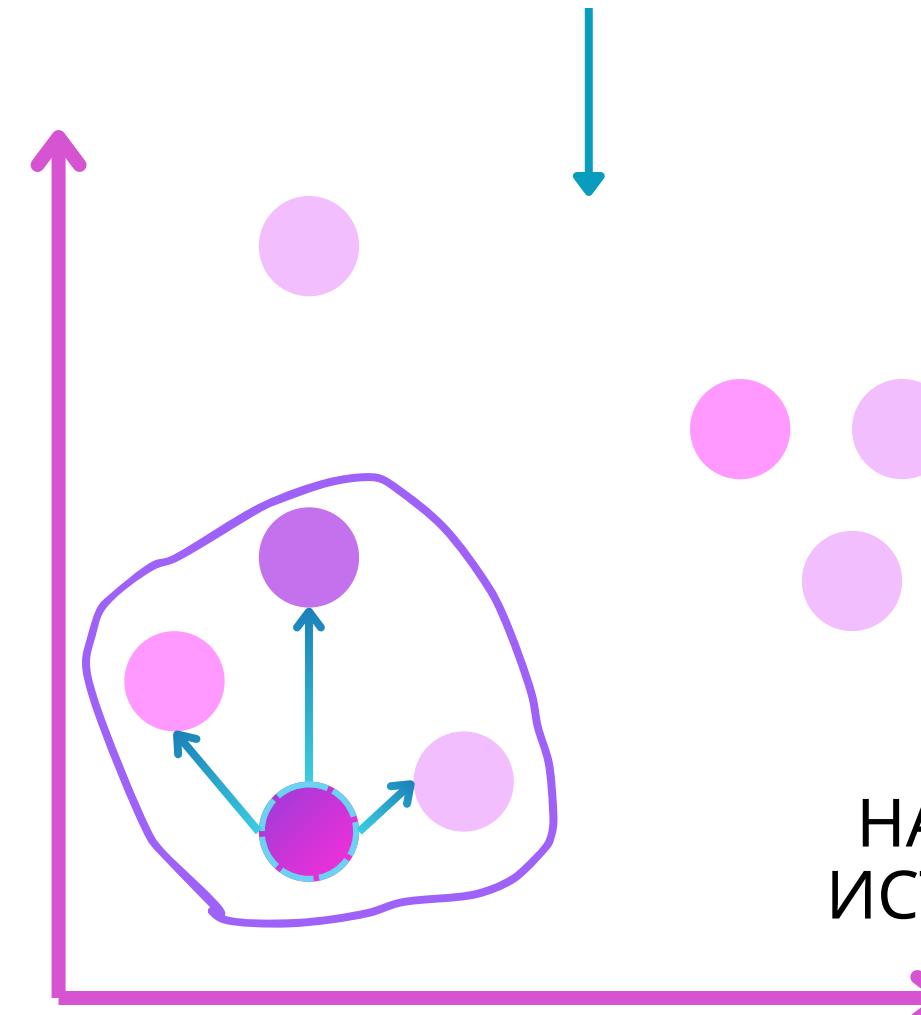
Среди похожих отрезков много сорвавшихся контрактов?

ДА

# Интерпритация

## Константные признаки

НАХОДИМ КОНСТАНТНЫЕ ПРИЗНАКИ С САМЫМ ВЫСОКИМ ИНДЕКСОМ ШЕПЛИ НА ИССЛЕДУЕМЫХ СЕГМЕНТАХ



НАХОДИМ ПОХОЖИЕ СЕГМЕНТЫ ИЗ ИСТОРИЧЕСКОЙ ВЫБОРКИ, СЧИТАЕМ ИХ ВЕРОЯТНОСТИ СРЫВА

ДАННОЕ КОНСТАНТНОГО ПРИЗНАКА ПРИ ТЕКУЩЕЙ КОНФИГУРАЦИИ ДРУГИХ ПАРАМЕТРОВ НЕГАТИВНО ВЛИЯЕТ НА ВЕРОЯТНОСТЬ СРЫВА

Среди похожих отрезков много сорвавшихся контрактов?

ДА

# ПРЕИМУЩЕСТВА И ОСОБЕННОСТИ

- ОРИЕНТИРУЕМСЯ НЕ ТОЛЬКО НА КОНКРЕТНЫЙ СЕМПЛ, НО И НА ДРУГИЕ ПОХОЖИЕ СЕМПЛЫ, ЧТО ДЕЛАЕТ ОЦЕНКУ БОЛЕЕ **ОБЪЕКТИВНОЙ** И УМЕНЬШАЕТ ШУМ;
- В НЕСКОЛЬКО ЭТАПОВ ОТБИРАЕМ ПРИЗНАКИ ДЛЯ АНАЛИЗА, ЧТОБЫ ИЗБЕЖАТЬ **ЛОЖНЫХ КОРРЕЛЯЦИЙ**;
- УЧИТЫВАЕМ **СЛОЖНЫЕ СТРУКТУРНЫЕ ОСОБЕННОСТИ** ДАННЫХ, КОТОРЫЕ НЕ ВСЕГДА МОЖНО УВИДЕТЬ ЧЕРЕЗ ИНДЕКС ШЕПЛИ НАПРЯМУЮ.
- ДЛЯ  $i$ -ОГО ОТЧЁТА МЫ УЧИТЫВАЕМ ТОЛЬКО ИНФОРМАЦИЮ ОТ 0 ДО  $i$ -ОГО ОТЧЁТА, ЧТО ДЕЛАЕТ РЕШЕНИЕ ПРИМЕНИМЫМ В **РЕАЛЬНОЙ АНАЛИТИКЕ**;
- ОТЧЁТЫ ИМЕЮТСЯ ЛЕГКОИНТЕРПРЕТИРУЕМЫЙ ТЕКСТОВЫЙ ВИД.



# примеры работы

обнаружены контракты с похожим значением Кредитном лимите. срываются в 42.0% случаев. Это значение сильно выше нормы, обратите на это внимание.

-Обратите внимание на Среднее кол-во заявок на пропуск за последние 2 недели настройплощадку. За последнее время найдено 152 контрактов с похожей временной динамикой этого параметра, которые не были завершены. Значения этого параметра в диапазоне значений, характерным для часто срывающихся контрактов

-Обратите внимание на Текущую сумму контракта за последние 2 недели. За последнее время найдено 144 контрактов с похожей временной динамикой этого параметра, которые не были завершены.

-Индекс к доверия исполнителю: 1.0. Это в пределах допустимой нормы, а значит исполнитель или другие исполнители, с которыми он связан, имели низкую долю похожих сорвавшихся сделок и высокий кредит доверия.

-Контракты с такой специализацией срываются в среднем в 13.0% случаев. Это допустимое значение, показывающее, что специализация мало влияет на вероятность срыва.



# примеры работы

Обратите внимание на Кол-во заказов материалов по договору за последние 2 надел{т}. За последнее время найдено 101 контрактов с похожей временной динамикой этого параметра, которые не были завершены. Аномальные скачки в изменении этого параметра в этом промежутке встречаются в похожих контрактах, которые сорвались в дальнейшем

-Обратите внимание на Среднее кол-во заявок на пропуск за последние 2 недели на стройплощадку. За последнее время найдено 166 контрактов с похожей временной динамикой этого параметра, которые не были завершены. Значения этого параметра в диапазоне значений, характерным для часто срывающихся контрактов

-Обратите внимание на Текущую сумму контракта за последние 2 недели. За последнее время найдено 211 контрактов с похожей временной динамикой этого параметра, которые не были завершены.

-Индекс к доверия к исполнителю: 1.0. Это в пределах допустимой нормы, а значит исполнитель или другие исполнители, с которыми он связан, имели низкую долю похожих сорвавшихся сделок и высокий кредит доверия.

-Контракты с такой специализацией срываются в среднем в 27.0% случаев. Это допустимое значение, показывающее, что специализация мало влияет на вероятность срыва.



# примеры работы

Контракты с похожими значениями Начальной сумме контракта, срываются в 13.0% случаев. Это допустимое значение, что может говорить о стабильности и надёжности этого параметра.

-Обратите внимание на Среднее кол-во заявок на пропуск за последние 11 недель настройплощадку. За последнее время найдено 168 контрактов с похожей временной динамикой этого параметра, которые не были завершены. Значения этого параметра в диапазоне значений, характерным для часто срывающихся контрактов

-Обратите внимание на Текущую сумму контракта за последние 11 недель. За последнее время найдено 122 контрактов с похожей временной динамикой этого параметра, которые не были завершены.

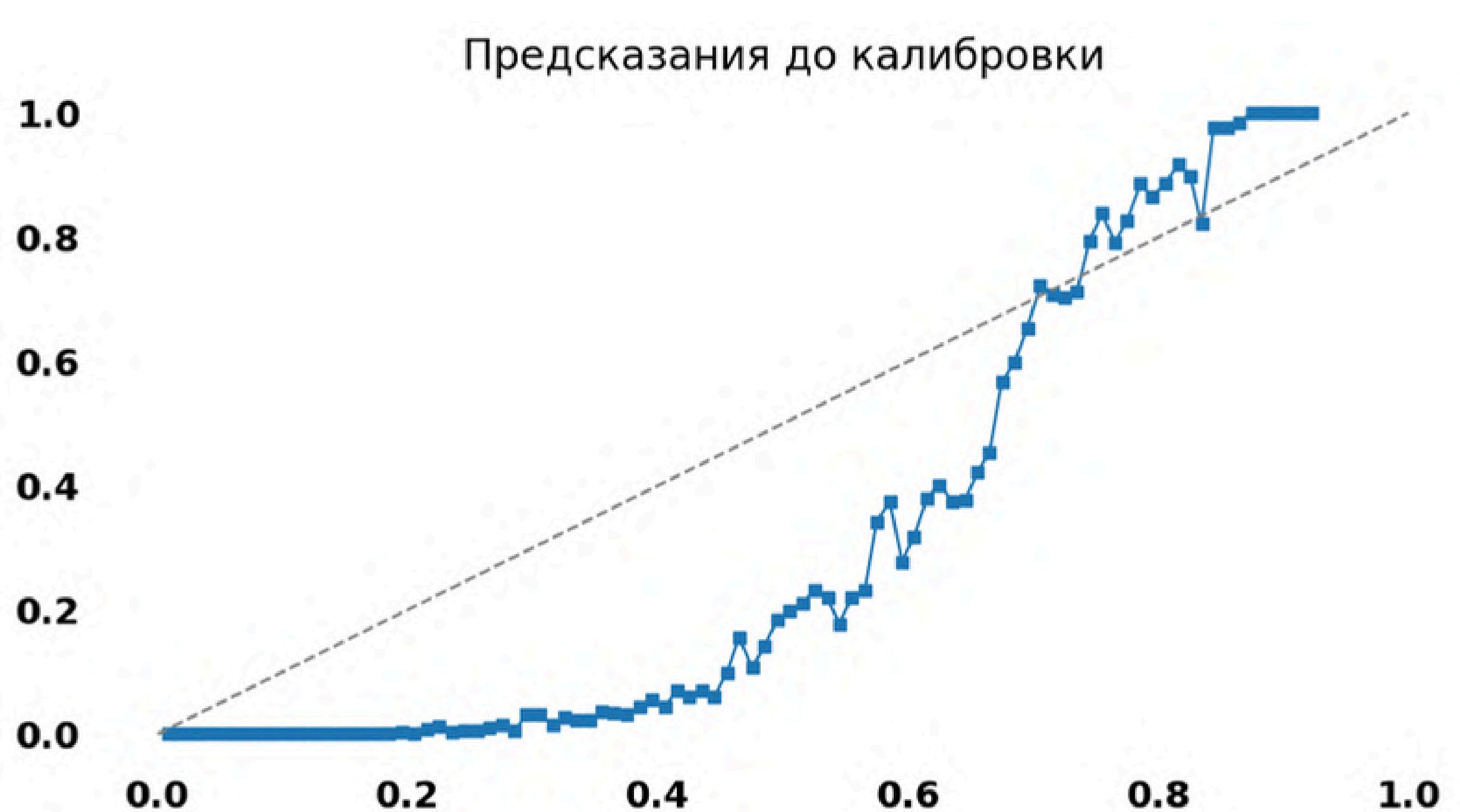
-Индекс к доверия к исполнителю: 0.73. Это в пределах допустимой нормы, а значит исполнитель или другие исполнители, с которыми он связан, имели низкую долю похожих сорвавшихся сделок и высокий кредит доверия.

-Контракты с такой специализацией срываются в среднем в 0.0% случаев. Это допустимое значение, показывающее, что специализация мало влияет на вероятность срыва.



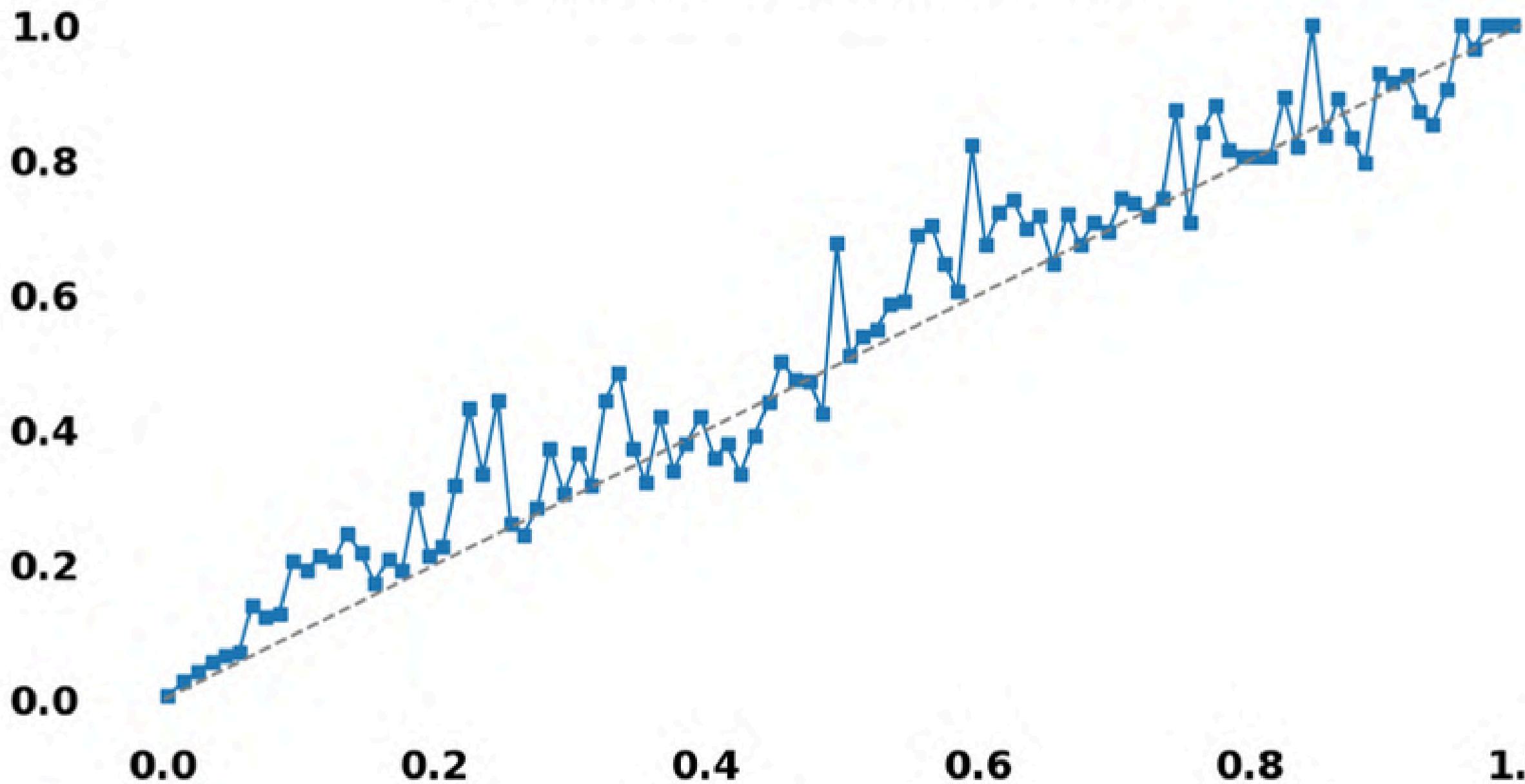
# Калибровка вероятностей

важно знать настоящую вероятность ухода поставщика, не ее подобие



# Калибровка вероятностей

Предсказания после калибровки



# Наш сервис



# Перспективы и масштабирование

- Перевод графа в **гетерогенный**, так как некоторые рёбра повторяются с разными весами
- Попробовать продвинутое использование **AutoML**
- Добавить более продвинутый **стекинг**
- Расширить выборку, для более качественной проверки сезонности

# Команда



Данис  
Динмухаметов



Артём  
Селезнёв



**СПАСИБО ЗА  
ВНИМАНИЕ**