

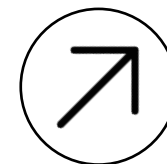
РОСГОССТРАХ

Комплексное развитие ИИ в Росгосстрахе

Еремеев Дмитрий
Шихалиев Фрэнк



О чем поговорим сегодня



Концепции, которых придерживаемся

1

Концепция фиче вектора-эмбединга
вокруг сущности

2

Демократизация
и transfer learning

3

Supervised
и unsupervised learning

4

Экономика и клиенты
в первую очередь

5

Венчурная модель
в экспериментах
по машинному обучению

Команда и подходы к управлению

- OKR
- Agile / SAFe и команда
- Scrum и Канбан
- Подрядчики часть команды



Задачи

Моделирование риска

Моделирование спроса
Динамическое ценообразование

Anti-fraud

- Анализ графов, модели на графах, графовые БД
- Векторные БД, алгоритмы схожести текста

Распознавание документов

- Полисы
- Документы

Автоматизация колл-центра

- Чат-бот
- Анализ отзывов (текст, аудио)

Анализ изображений

- Фото с предосмотра авто
- Детекция повреждений

Гео-аналитика

Эмбеддинги локаций

Моделирование риска (авто)

Каско

- Виновник
- Потерпевший
- Стекло
- ПДТЛ
- Тоталь
- Угон

ОСАГО

Типы урегулирования

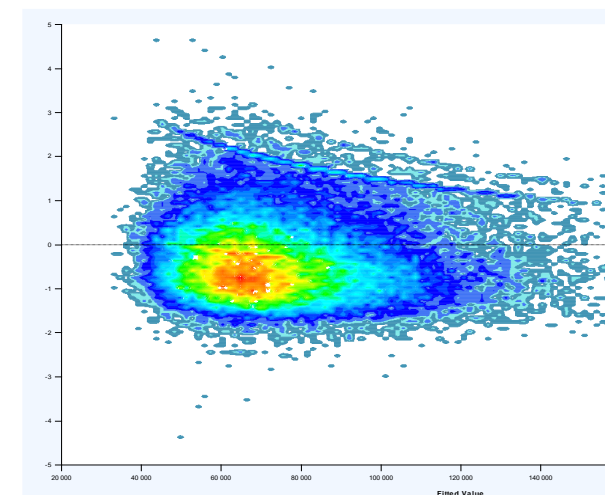
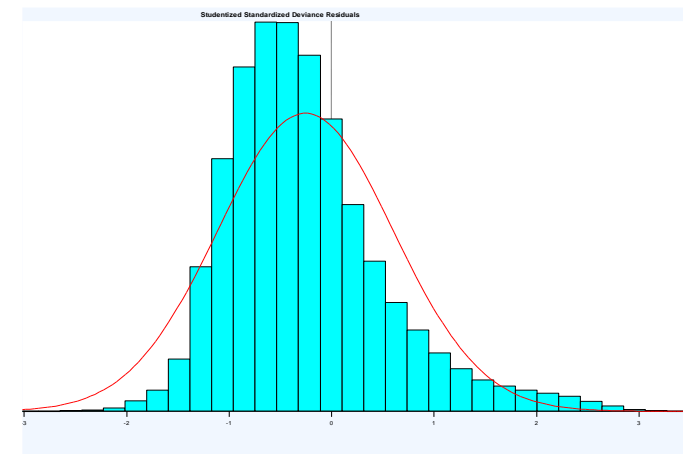
Риск

Частота

(Poisson regression)

Тяжесть

(Gamma regression)

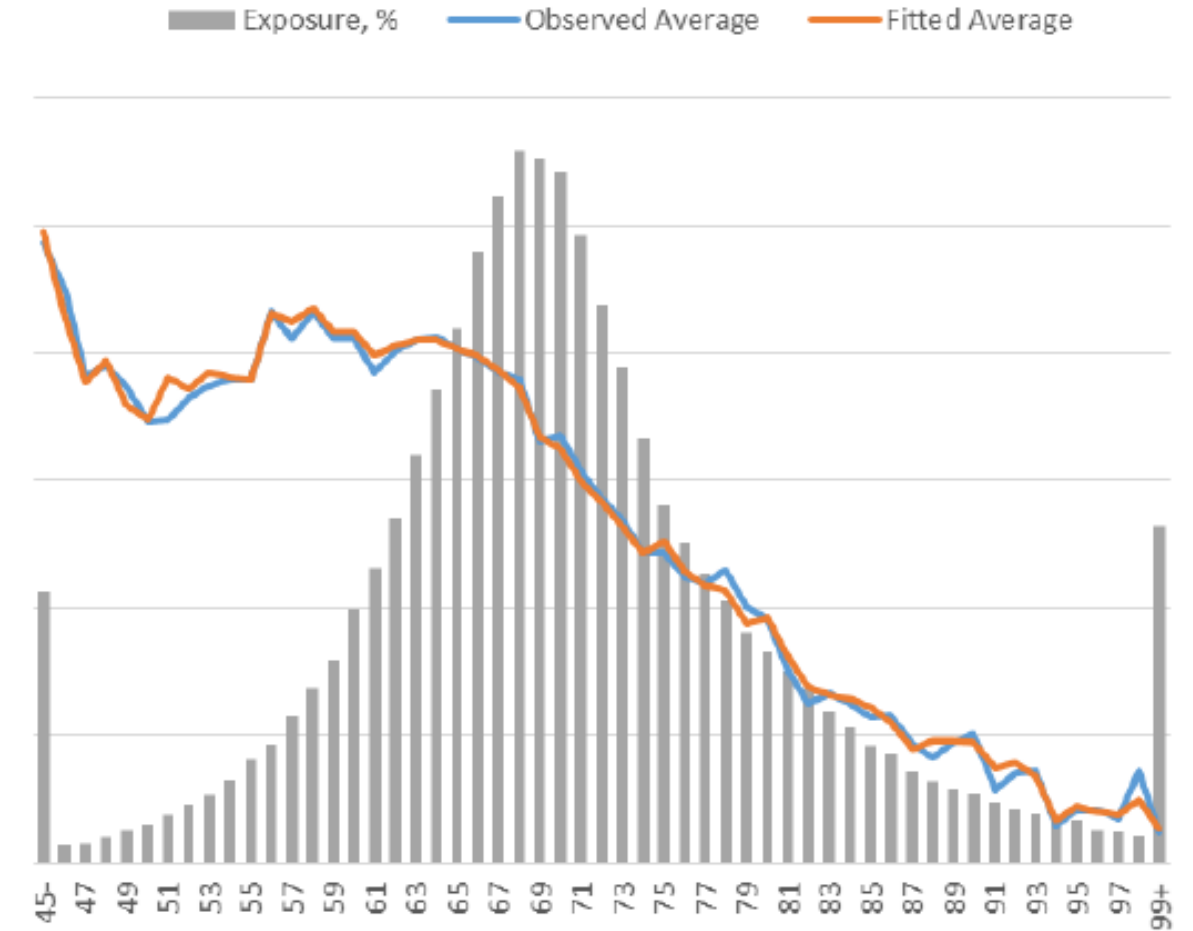


Моделирование спроса

Модели спроса

- Пролонгация (расторжение)
- Новый бизнес (конверсия)
- Досрочное расторжение
- Смена покрытия / продукта
- Прогноз вероятности (logistic regression)

Зависимость конверсии от ценовой привлекательности компании



GLM

Generalized Linear Model

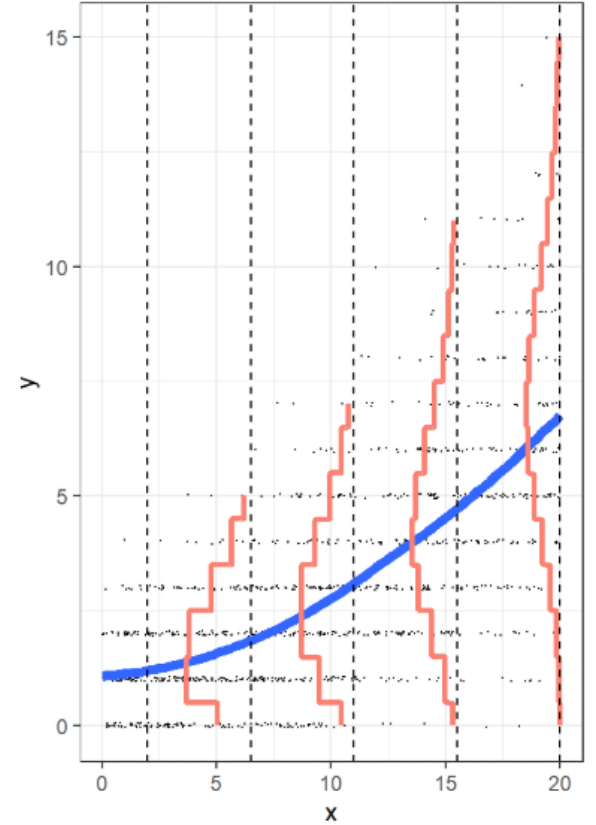
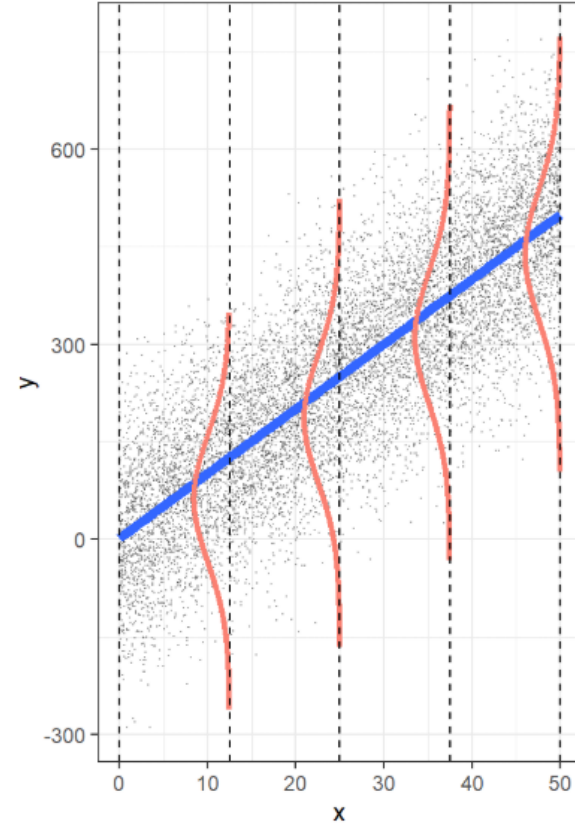
Наблюдения Y_i из экспоненциального распределения:

$$Y_i \sim p(y_i|\theta_i, \varphi) = c(y_i, \varphi) \exp\left(\frac{y_i\theta_i - a(\theta_i)}{\varphi}\right).$$

Мат. ожидание $\mu = \mathbb{E}(Y_i)$:

$$\mu = g^{-1}(\mathbf{w}_1 X_1 + \mathbf{w}_2 X_2 + \dots + \mathbf{w}_n X_n),$$

где g - функция связи.

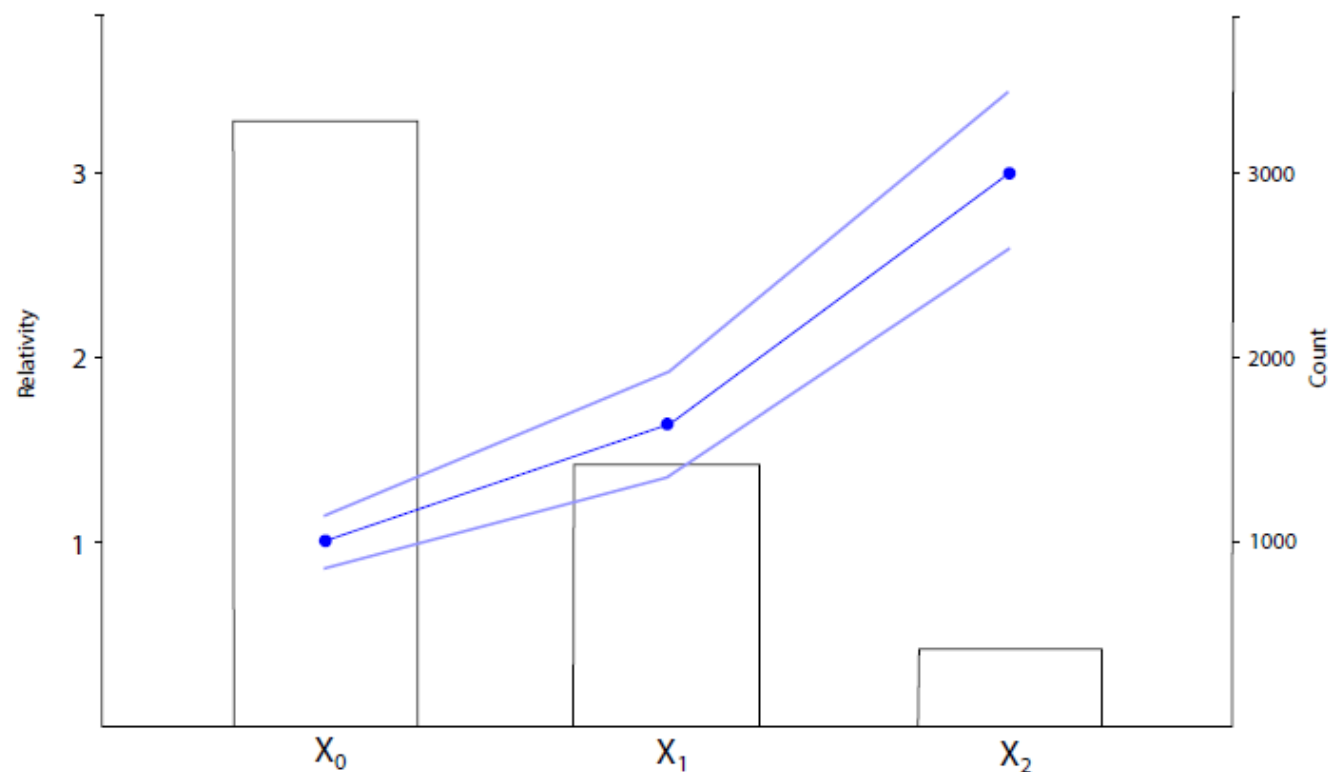


Likelihood

Параметры $\mathbf{w} = \{w_1, w_2, \dots, w_n\}$ определяются максимизацией правдоподобия:

$$l(\mathbf{w}) = \log p(\mathcal{D}|\mathbf{w}) = \sum \log p(y_i|x_i, \theta_i, \varphi),$$

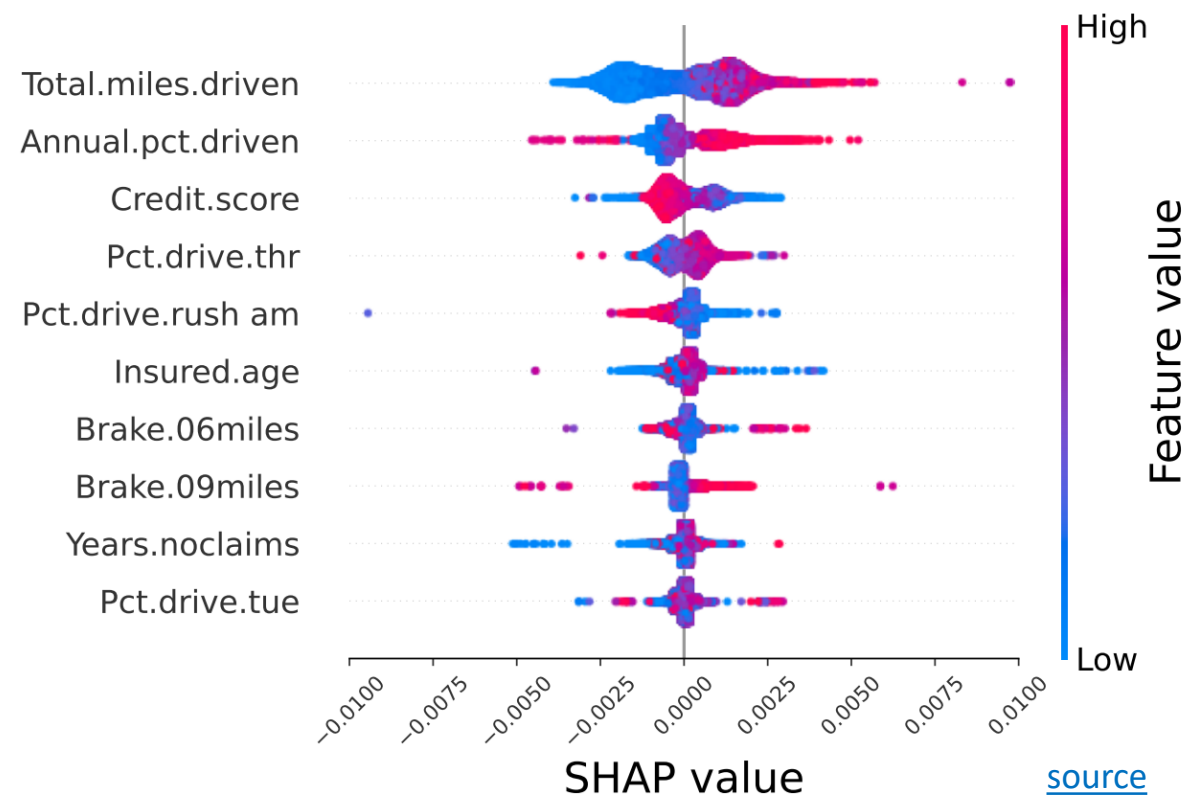
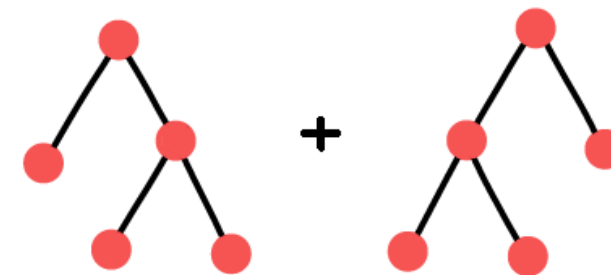
$$\hat{\mathbf{w}} = \operatorname{argmax} l(\mathbf{w}).$$



GLM -> GBM

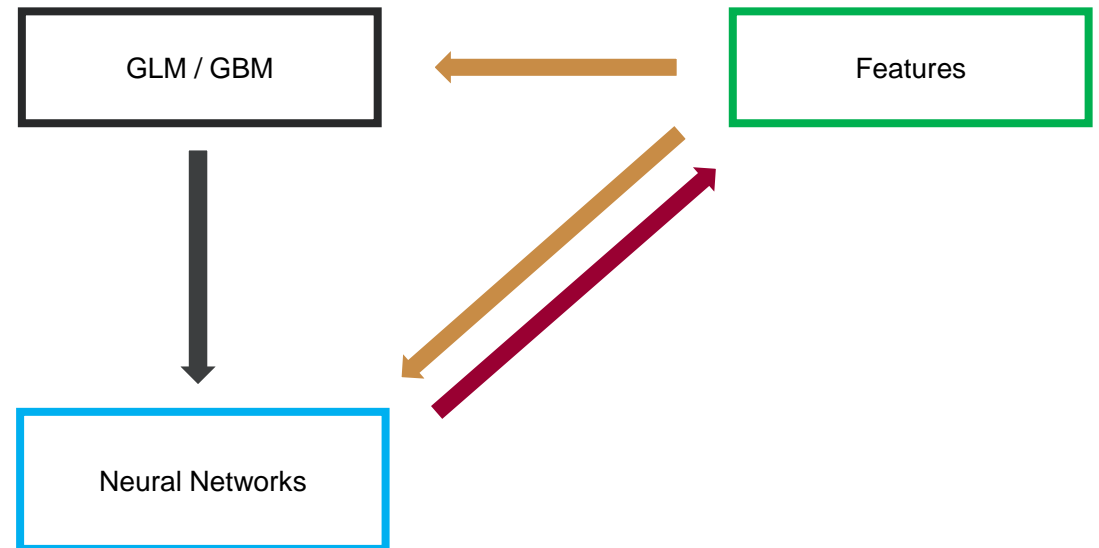
Градиентный Бустинг

- Позволяет лучше учесть нелинейности, что влияет на точность предсказаний.
- Облегчает работу с большим количеством признаков.
- Дает выигрыш в скорости построения моделей.
- Хуже интерпретируемость SHAP, Partial Dependence



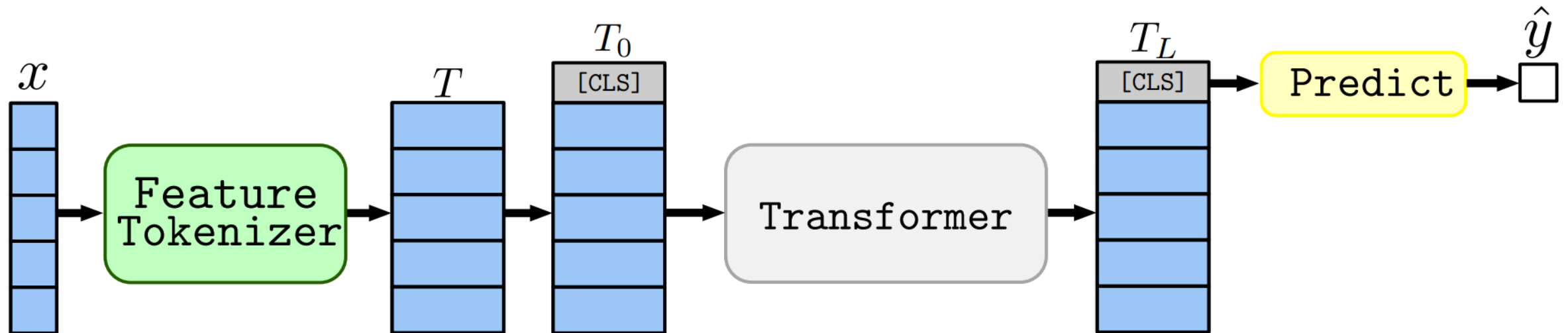
GBM -> Нейронные сети

- Нейронные сети могут как заменять GLM / GBM, так и служить для усиления этих моделей.
- Замена рискованных моделей возможна с помощью «табличных» нейросетей.
- Усиление GLM / GBM возможно с помощью генерации признаков нейросетями.



Табличные данные

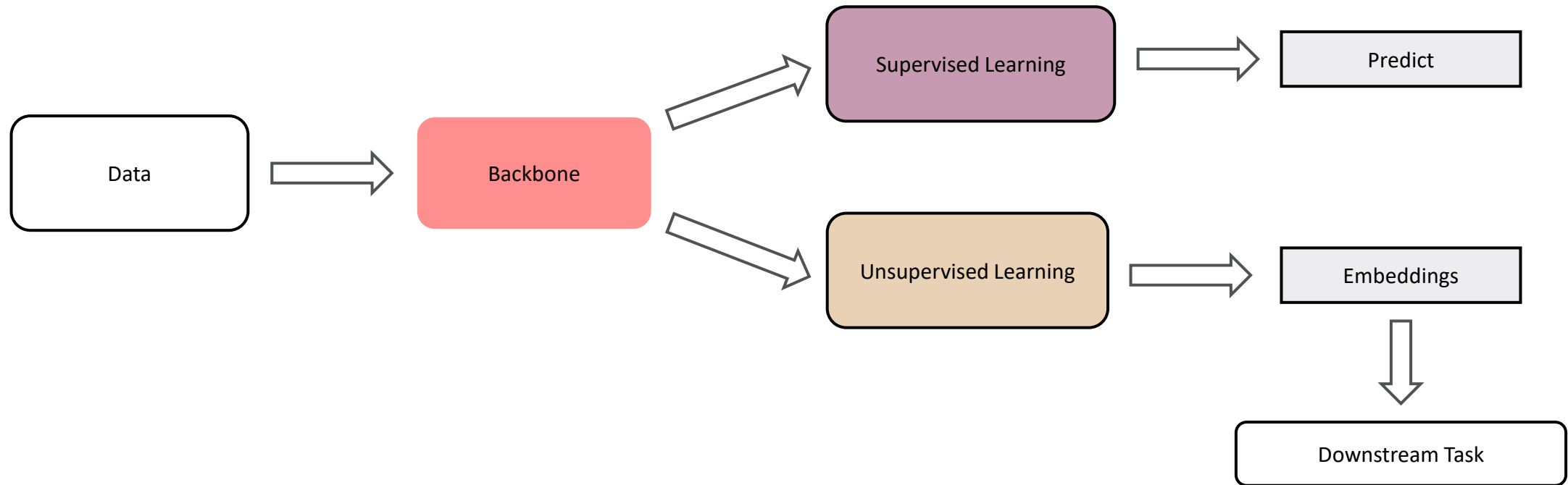
- Модификации MLP
- Табличные трансформеры: TabTransformer, FT-Transformer, ...



[FT-Transformer](#)

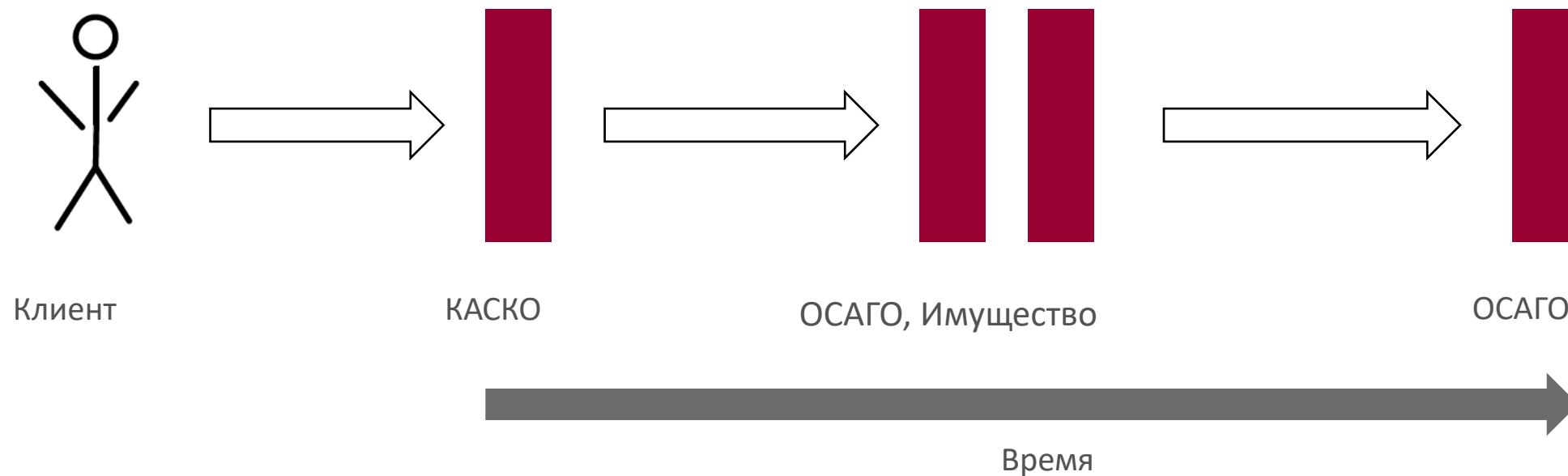
Векторные представления

- Нейросети предоставляют удобные средства работы со структурированными данными.
- Тексты, картинки, аудио, последовательности, графы представляются в виде обучаемых векторных представлений (эмбеддингов).



История клиента

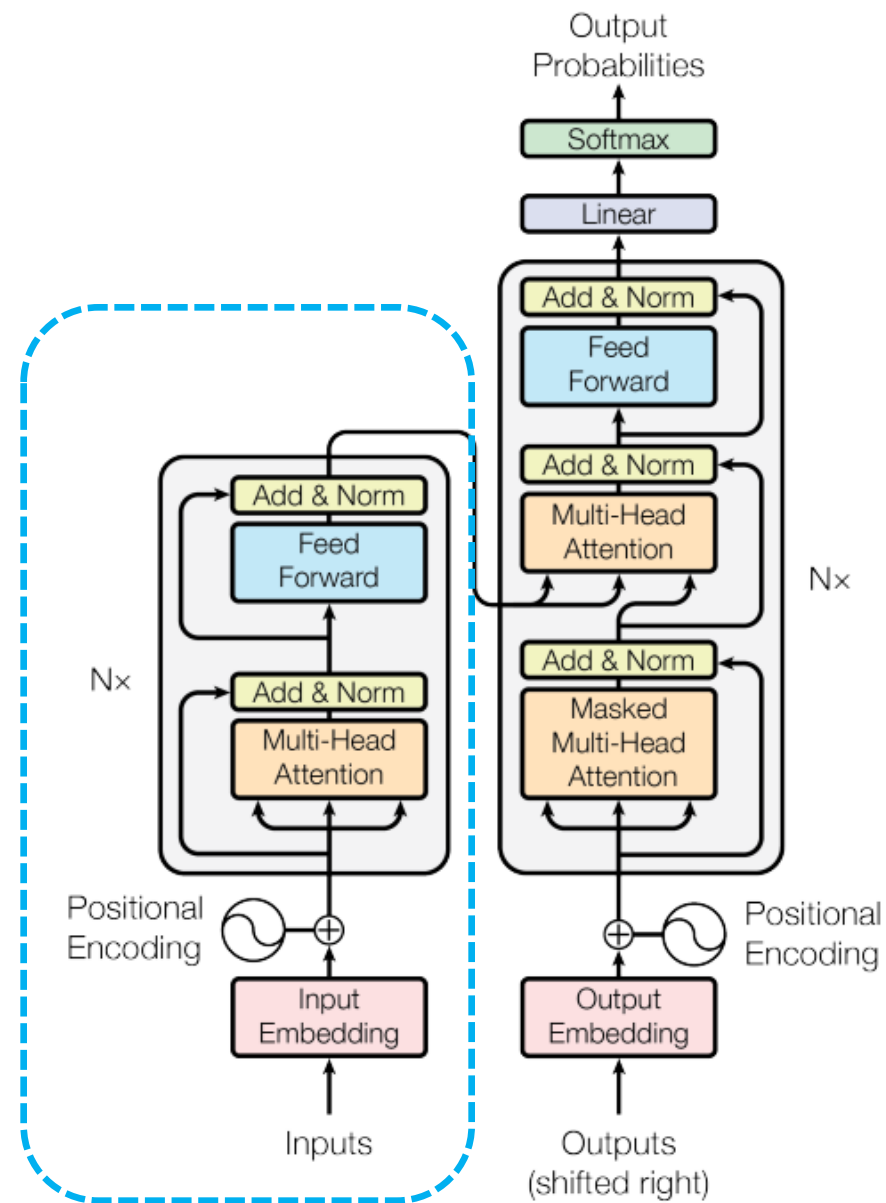
- Учесть влияние страховой истории клиента
- Event-sequence задача
- Каждый полис имеет свои характеристики и историю убытков



История клиента

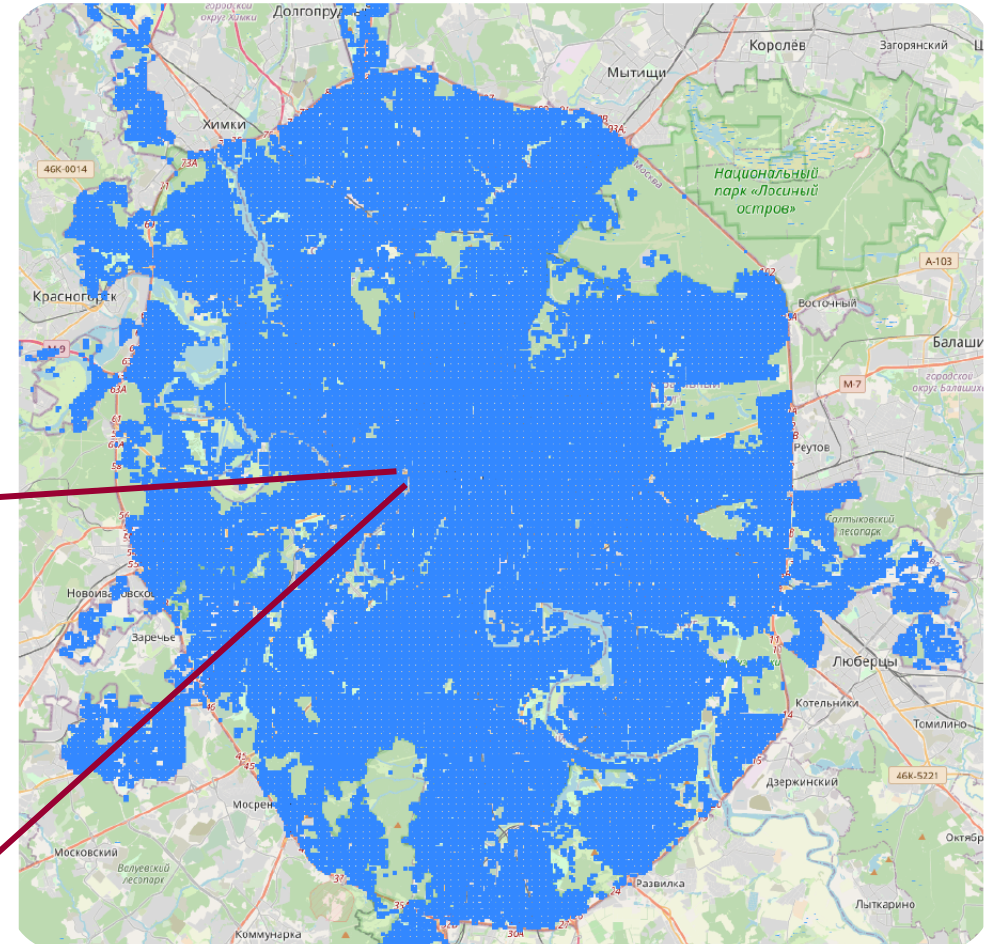
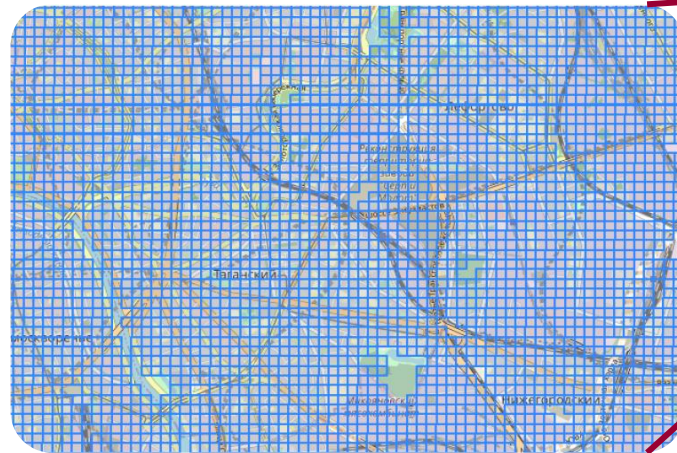
- Для получения векторного представления последовательности событий используем Transformer Encoder.
- Supervised learning:
[CLS] токен + log-likelihood
- Self-supervised learning:
 - CoLES (Contrastive Learning for Event Sequences with Self-Supervision)
 - Основанные на подходах из LLM.

[Transformer](#)



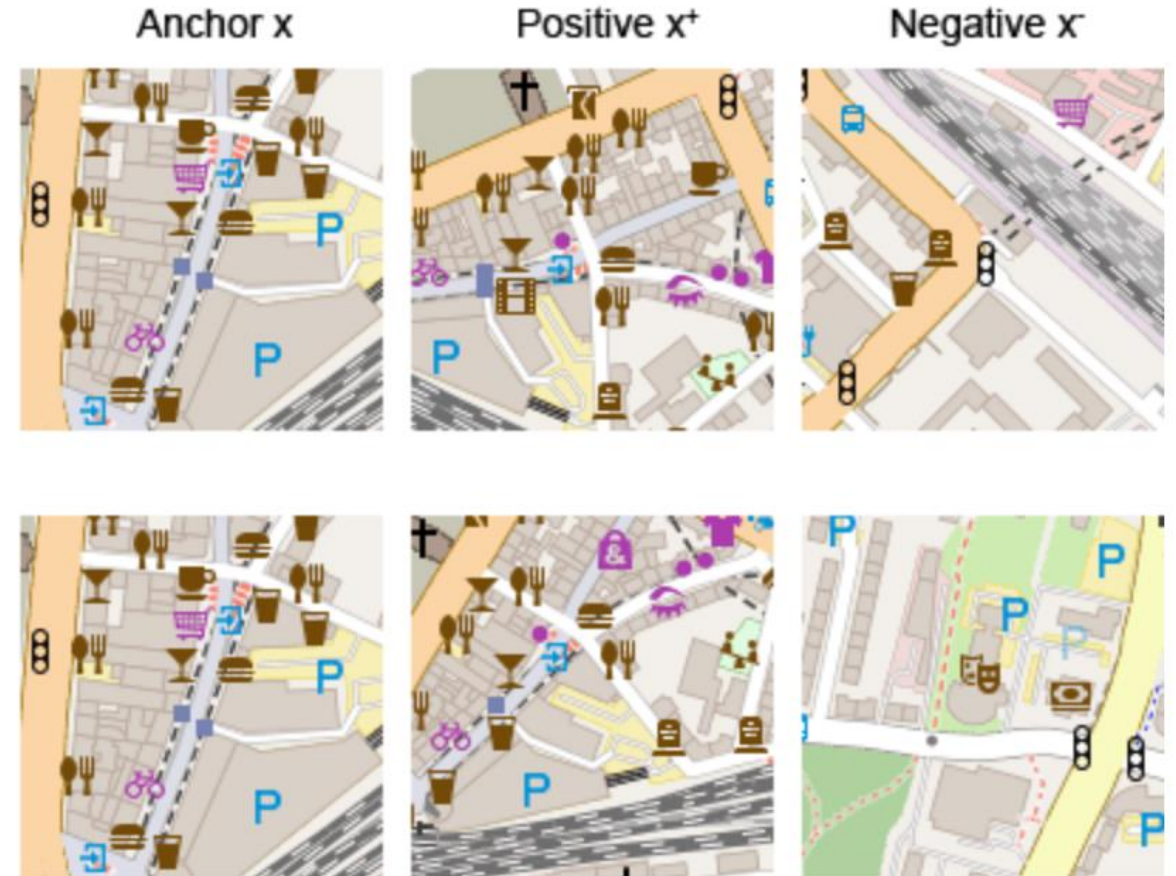
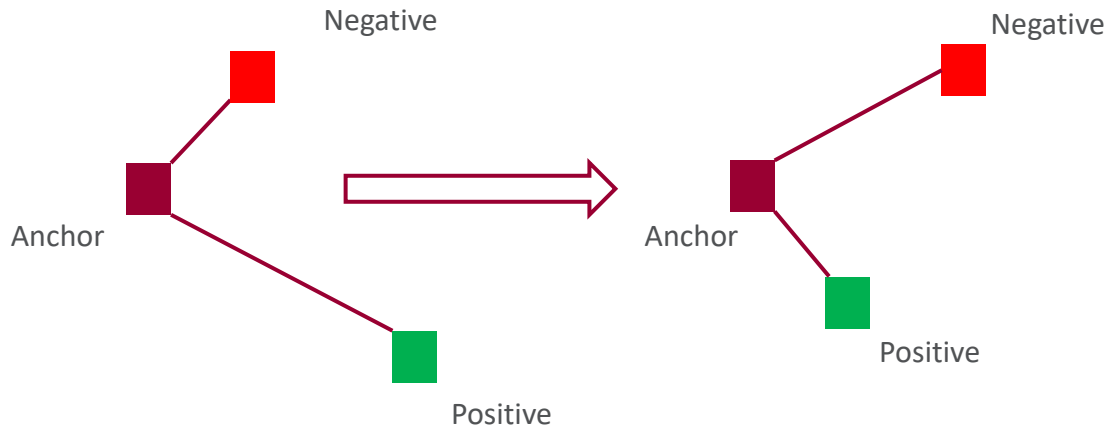
Гео

- Геосетка: полигоны 100 x 100 метров
- Признаки: основаны на POI, местности в окрестности, свойствах городов
- Слои в карты в стандартном стиле
- Эмбединги полигонов
- Геокодирование адресов



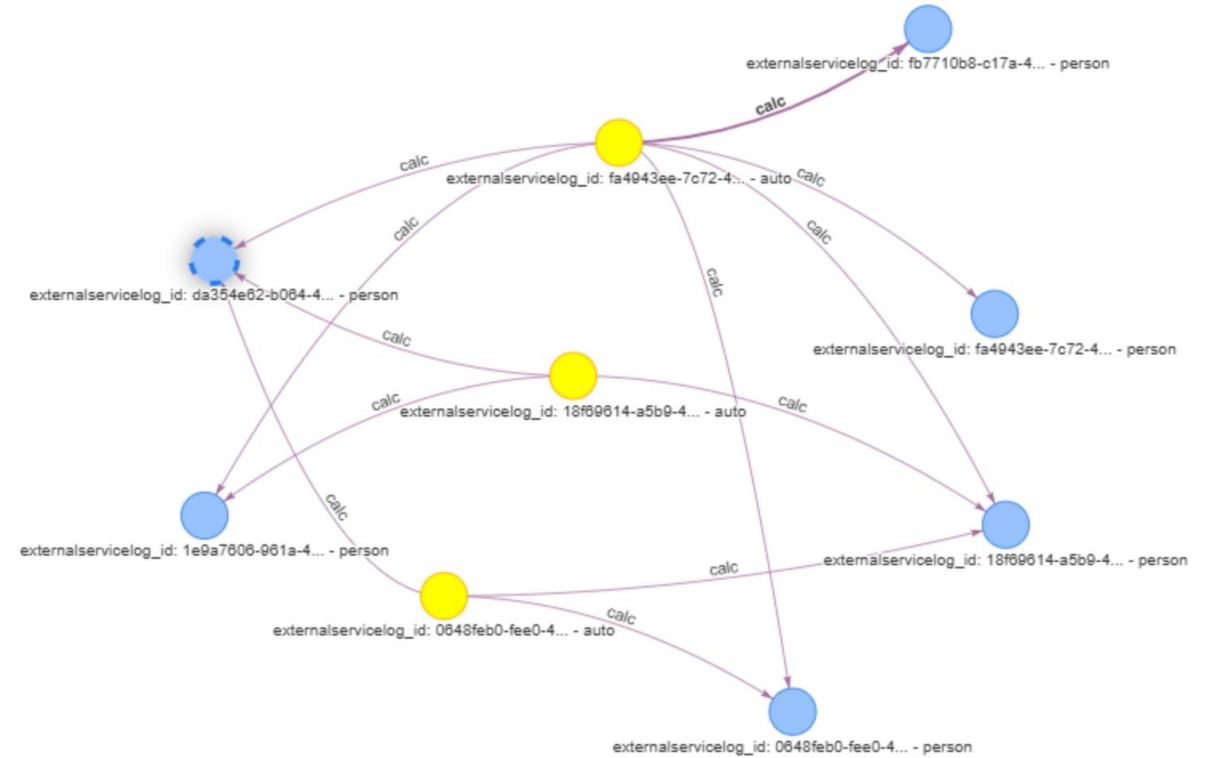
Векторные представления полигонов

- Энкодер: сверточная нейросеть (CNN)
- Каналы изображения: слои карты
- Contrastive learning



Anti-Fraud

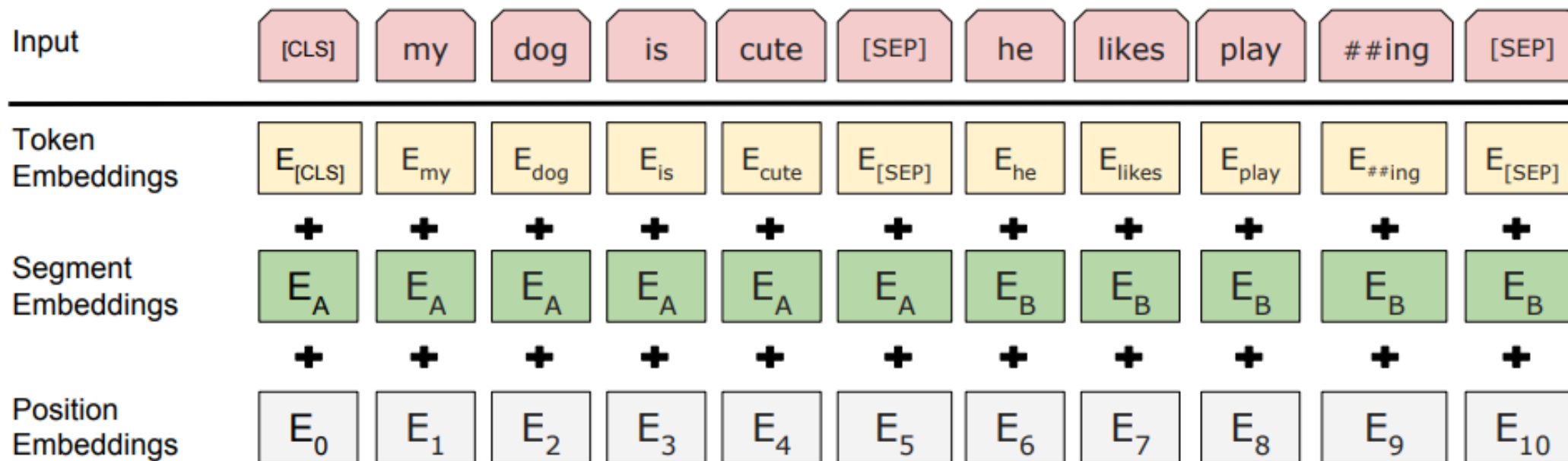
- Графы: убытки + котировки
- Вершины: персоны, авто + признаки
- Ребра: связи соответствующего типа
- Semi-Supervised / Unsupervised модели
- Графовые БД



Обработка отзывов клиентов

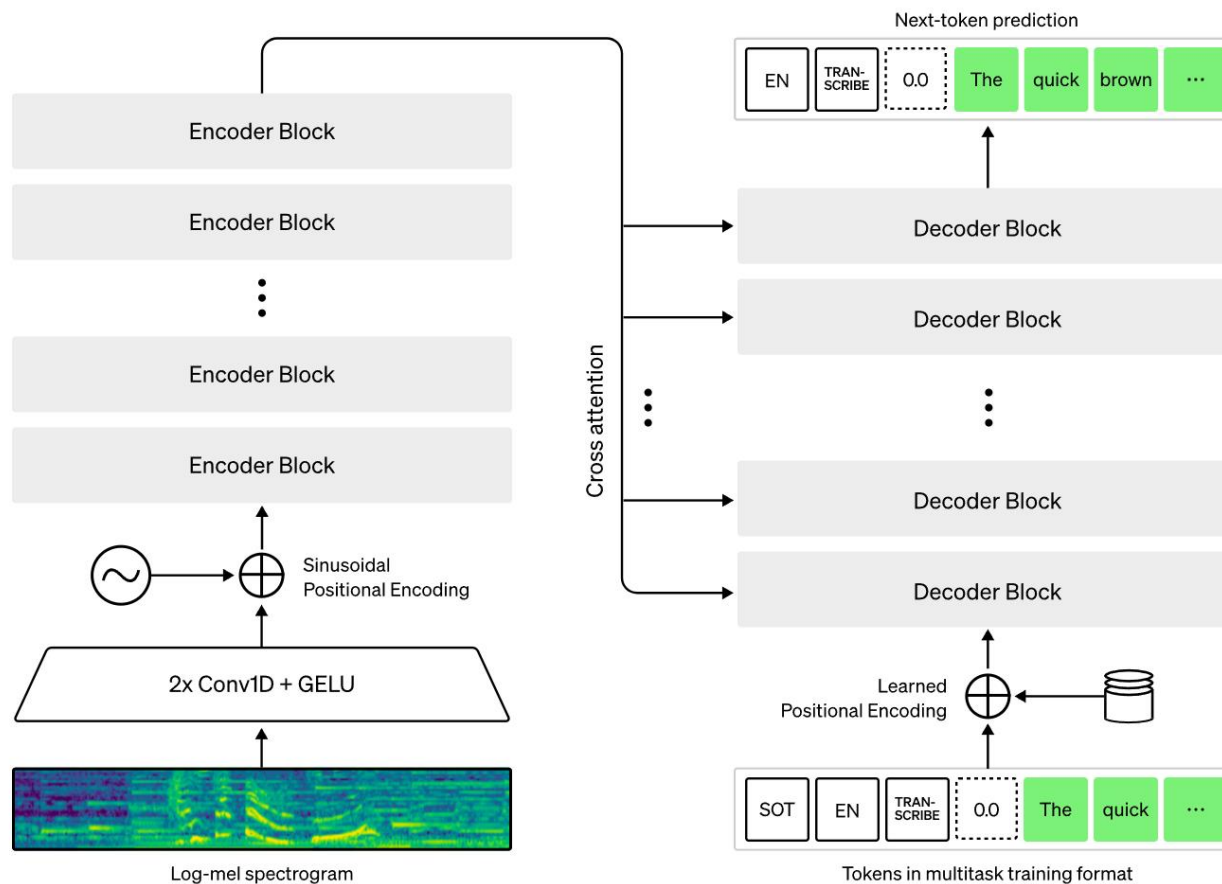
- Классификация / типизация / суммаризация отзывов, обращений.
- Удовлетворенность клиента, готовность рекомендовать компанию, удобство клиента.
- Текстовые отзывы Transformer-based модели

BERT



Обработка звука

- Модели распознавания звука (ASR) позволяют перевести речь в текст
- Видный представитель: Whisper



OCR

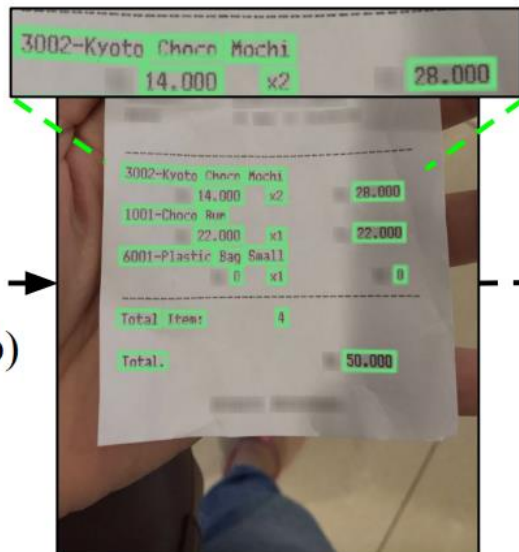
- Распознавание документов
- Паспорта, полисы, ...

Document Image --- Structured Information

(a)



(b)



(c)

```
{ "words": [  
  {  
    "bbox": [[0.11,0.21],..., [0.19,0.22]],  
    "text": "3002-Kyoto"  
  }, {  
    "bbox": [[0.21,0.22],..., [0.45,0.23]],  
    "text": "Choco"  
  }, {  
    "bbox": [[0.46,0.22],..., [0.52,0.23]],  
    "text": "Mochi"  
  }, ..., {  
    "bbox": [[0.66,0.31],..., [0.72,0.32]],  
    "text": "50.000"  
  }  
]
```

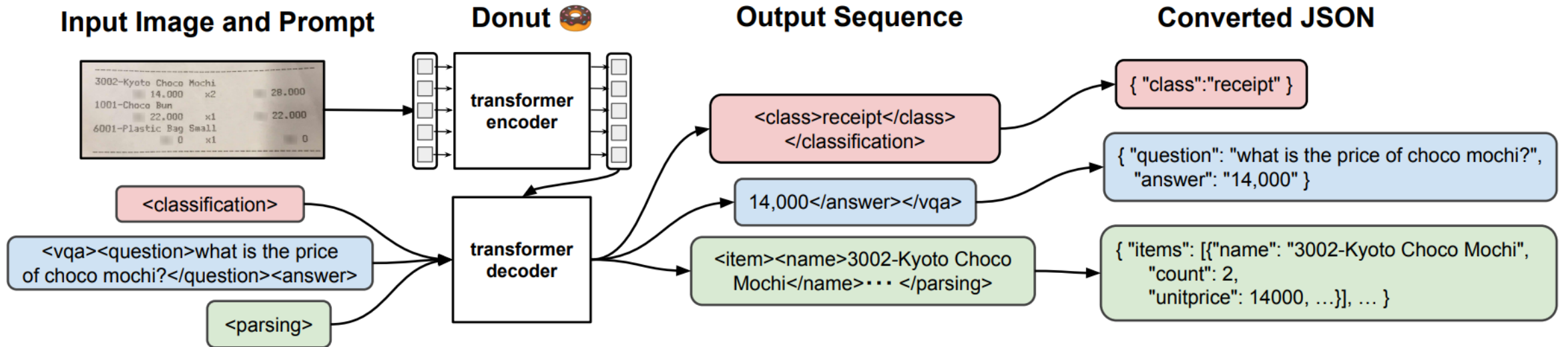
(d)

```
{ "items": [  
  {  
    "name": "3002-Kyoto Choco Mochi",  
    "count": 2,  
    "priceInfo": {  
      "unitPrice": 14000,  
      "price": 28000  
    }  
  }, ...  
],  
"total": [ {  
  "menuqty_cnt": 4,  
  "total_price": 50000  
} ]  
}
```

OCR

- Donut
- Phi-3-Vision

Donut



Распознавание повреждений

- Детекция повреждений по фото.
- Можно использовать как непосредственно детекцию / оценку повреждений, так и оценивать влияние на риск при принятии на страхование.



Метрики

300  млн
прибыли в год

Запросы и будущее

Мультимодальные модели
в периметре компании

Корпоративные платформы чат ботов на
базе LLM и RAG

Система исходящего персонального взаимодействия на
базе истории
с применением LLM

Новые подходы борьбы
с мошенничеством на базе ИИ

Корпоративная система
АБ тестирования

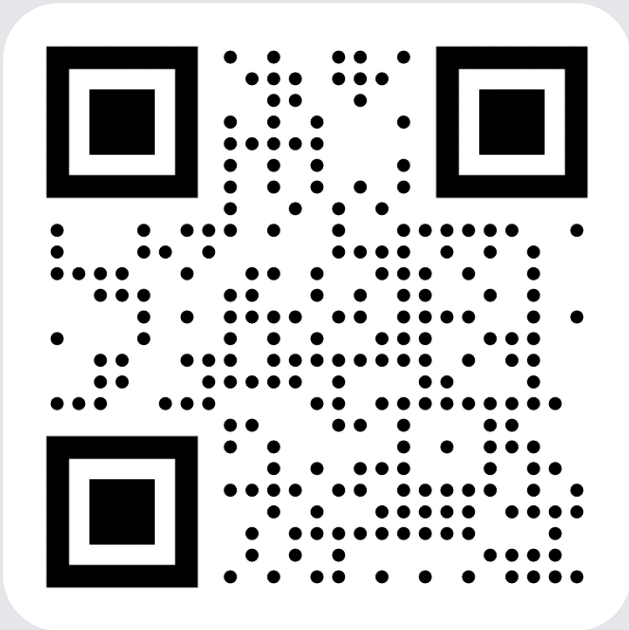
Корпоративные системы работы с
различными внешними моделями

Системы контроля
за моделями разных видов

Для партнеров

Предложения по ИИ
от партнеров собираем централизованно:

ai@rgs.ru



- Аутстаф - достижения в страховании и ИИ, запрос на резюме свободных сотрудников для привлечения с опытом ИИ в страховании
- **Аутсорс** – примеры проектов в страховании в области ИИ, чем подрядчик гордится
- **Интеграторы и стартапы** - продукты по ИИ с подтвержденным эффектом и успешным кейсом внедрения, подходящим под страхование