

# Система типизации ОТЗЫВОВ КЛИЕНТОВ



# Система типизации отзывов клиентов

В современном бизнесе обратная связь от клиентов играет ключевую роль в улучшении качества продуктов и услуг. Система типизации отзывов позволяет структурировать и анализировать большие объемы данных, получаемых из различных источников, таких как записи звонков, формы на сайте и Excel-таблицы. Это необходимо для выявления ключевых проблемных зон компании и повышения удовлетворенности клиентов.

## Типизация отзывов производится по следующим классам

**NPS (Net Promoter Score)** — готовность клиентов рекомендовать компанию

**CES (Customer Effort Score)** — удобство использования сайта: (1) проблемы и (2) что можно улучшить

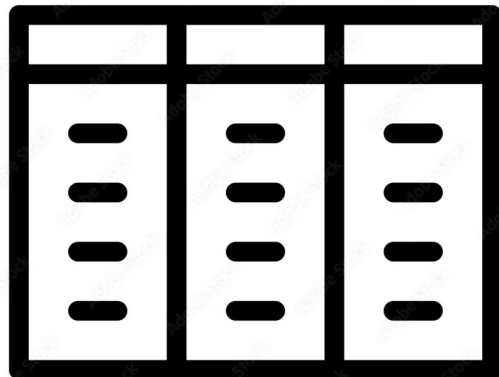
**CSI (Customer Satisfaction Index)** — удовлетворенность клиентов по этапам урегулирования убытков

Для каждой метрики существуют принятый в компании стандартизированный перечень типизаций.

# Разметка данных для обучения моделей

## Ручная разметка

Ручная разметка отзывов, записанных операторами в таблицах, является основой для обучения моделей. Этот процесс обеспечивает высокую точность и позволяет учитывать нюансы, которые могут быть упущены при автоматической обработке.



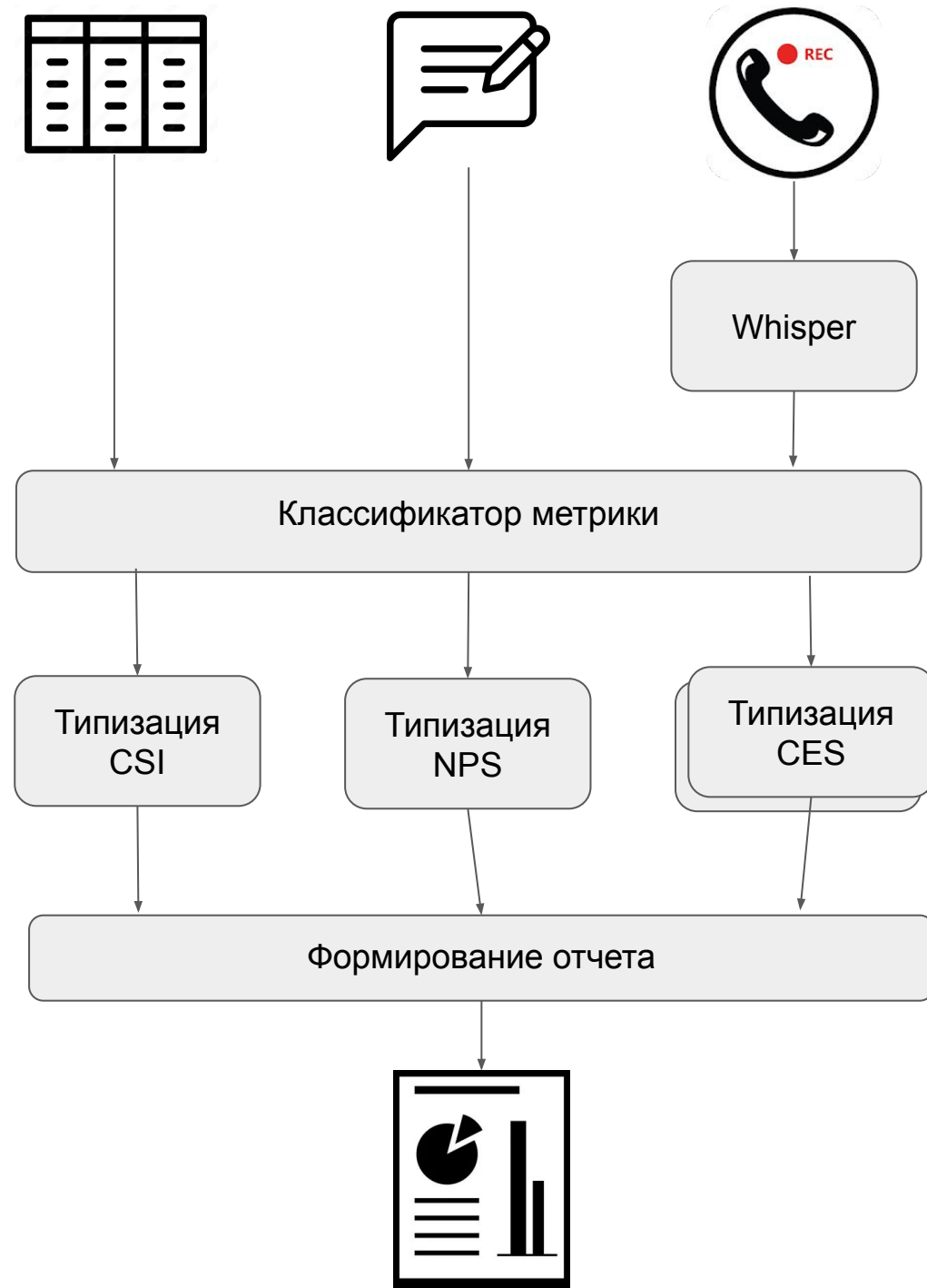
## Разметка текстовых отзывов и записей звонков

Текстовые отзывы из других источников и записи звонков требуют дополнительной разметки. Особое внимание уделяется выделению элементов диалога, не относящихся к непосредственной оценке метрики, чтобы избежать искажений при анализе.

Кроме того, качество записей звонков не всегда высокое, что осложняет транскрибацию.



# Общий пайплайн системы



## 1 Предобработка данных

Загрузка текстов и таблиц, транскрибация записей звонков (например, с использованием Whisper). Этот этап включает очистку данных и приведение их к формату, удобному для дальнейшей обработки.

## 2 Классификация метрики (пока не реализовано)

Классификация отзыва или комментария по типу метрики (NPS, CES, CSI) и фильтрация неинформативных отзывов. Этот шаг позволяет направлять отзывы в соответствующие модели типизации.

## 3 Типизация отзыва

Передача отзыва на модель типизации, соответствующую данной метрике. Модель анализирует отзыв и определяет его тип (например, проблема с доставкой, неудобный интерфейс).

## 4 Формирование результата

Система формирует итоговый отчет с типизированными отзывами. Данные представляются в удобном для анализа формате, что позволяет быстро выявлять ключевые тенденции и проблемные области.

# Предобработка данных: транскрибация с использованием Whisper

## Формат диалога

На выходе Whisper предоставляет диалог в формате:

operator: ...

client: ...

operator: ...

client: ...

## Пример диалога

operator: Здравствуйте, чем могу помочь?

client: У меня проблема с доставкой, заказ задерживается на три дня.

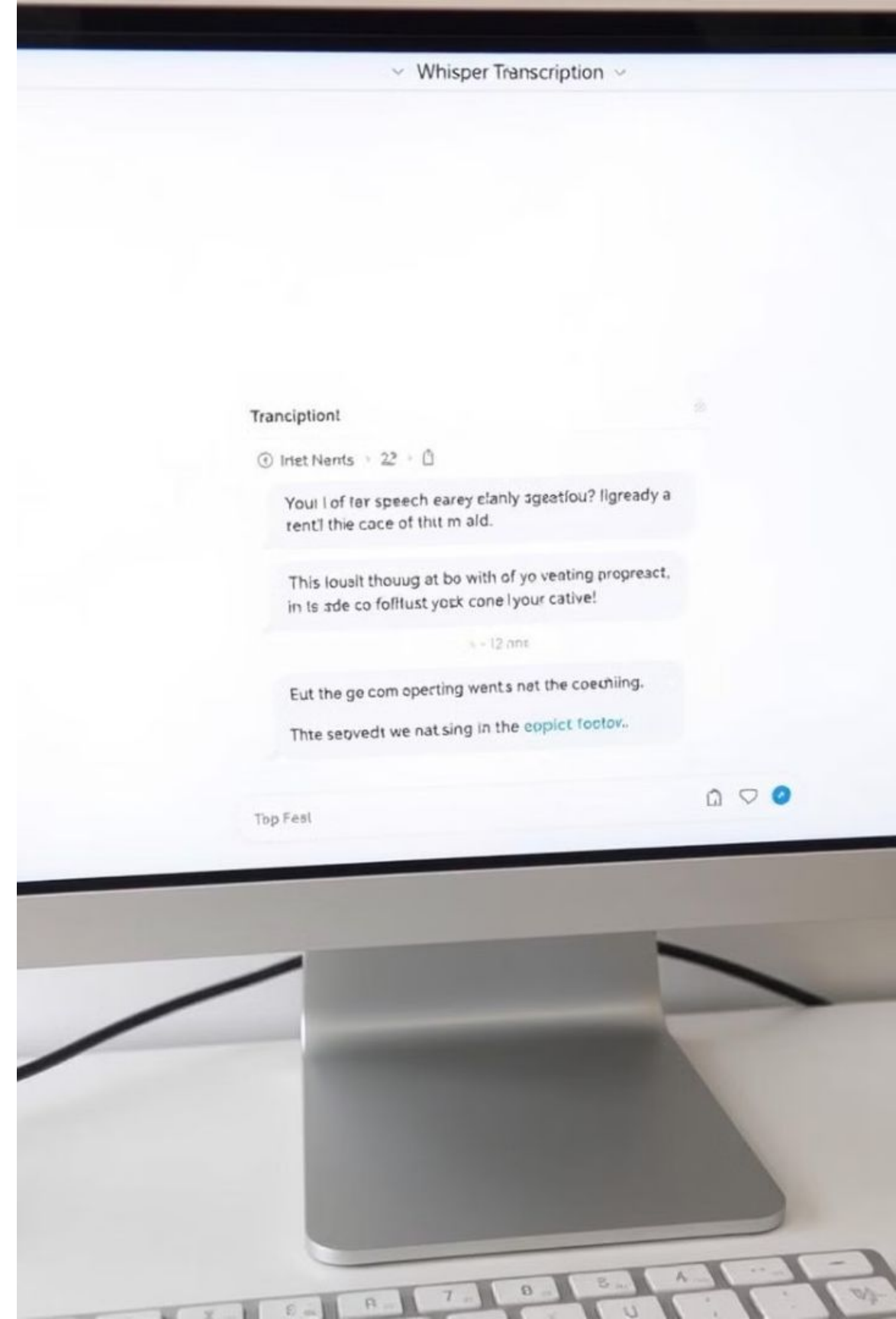
operator: Приношу свои извинения, сейчас уточню информацию.

client: Спасибо, жду.

## Обработка диалога

Необходимо отделить реплики, относящиеся к делу, от приветствий и прощаний.

Важно выделить суть проблемы, указанной клиентом, для дальнейшей типизации.





# Модель типизации отзывов

- 1

Дообучение ruBert

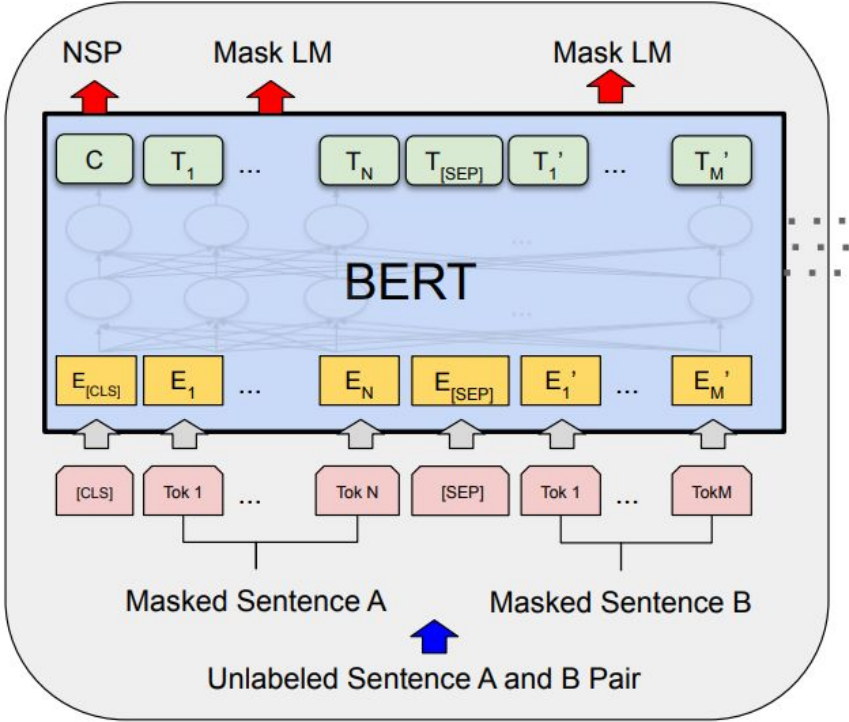
Дообучение модели ruBert для получения семантических эмбеддингов, чтобы отзывы одного типа были близки друг к другу в векторном пространстве.
- 2

Снижение размерности и метод ближайших соседей

Подбор метода снижения размерности (PCA) и метода ближайших соседей на валидационных данных для получения итогового предсказания типизации отзыва. Возможна выдача нескольких наиболее вероятных типизаций.
- 3

Метрики

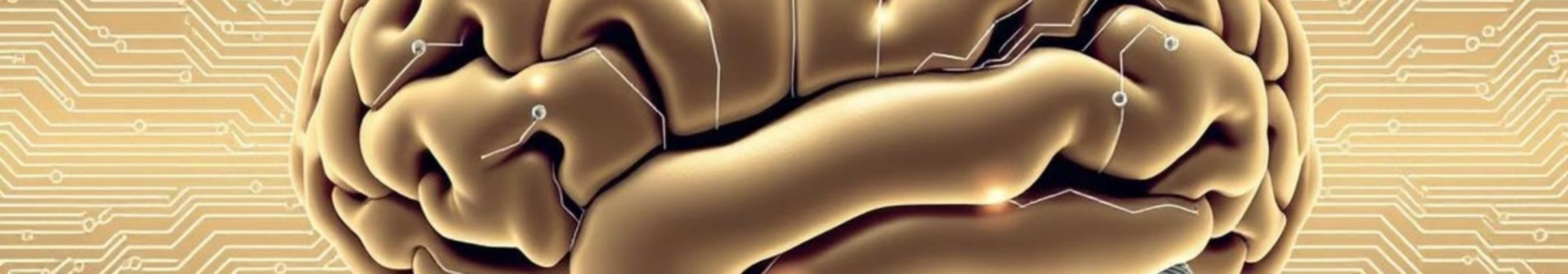
С учетом количества возможных типизаций (до 250 видов) на нынешнем этапе этот подход не может давать высокие метрики (SOTA-модели для русского языка на geo-reviews-dataset-2023 с 272 классами дают максимум ассигасу 59%). Проблемой являются также редко встречающиеся типизации, для которых недостаточно примеров в обучающей выборке. Так, из 245 типизаций по CSI только 176 репрезентативно представлены в выборке.



Ассигасу по метрике	Тестовый датасет	Продовые данные
CSI	0.67	0.38-0.46
NPS	0.73	-
CES 1	0.64	-
CES 2	0.54	-

# Возможные подходы для улучшения качества предсказаний

1. Дополнение обучающей выборки для достижения достаточного количества (не менее 50-100) примеров каждого класса типизации.
2. Использование большой языковой модели для типизации ОТЗЫВОВ.
3. Пересмотр стандартизированной типизации на основе семантической кластеризации ОТЗЫВОВ.



# Перспективная модель типизации с использованием большой языковой модели Phi-3/3.5



## Дообучение языковой модели

Дообучение языковой модели (Phi-3/3.5) для получения краткого описания проблематики, указанной в отзыве клиента.



## Суммаризация

Получение краткого описания проблемы в свободной форме (суммаризация) для более точного понимания контекста отзыва.



## Определение ближайшей типизации

Определение наиболее близкой к этому краткому описанию стандартизированной типизации для автоматической категоризации отзыва.





# Бизнес-эффекты от внедрения системы

1

Улучшение качества продукции и услуг

Систематический анализ

отзывов клиентов позволяет выявлять проблемные зоны и вносить оперативные улучшения, повышая общую удовлетворенность клиентов.

2

Оптимизация бизнес-процессов

Типизация отзывов позволяет оптимизировать процессы, такие как доставка, обслуживание клиентов и взаимодействие с сайтом, сокращая издержки и повышая эффективность.

3

Повышение лояльности клиентов

Оперативное реагирование на проблемы, выявленные в отзывах, способствует повышению лояльности клиентов и укреплению позитивного имиджа компании.