

БАНКИ: ВЧЕРА И СЕГОДНЯ

Развитие DS в банковской сфере
AI-трансформация банкинга

Масютин Алексей

Центр ИИ НИУ ВШЭ
ex-Сбер

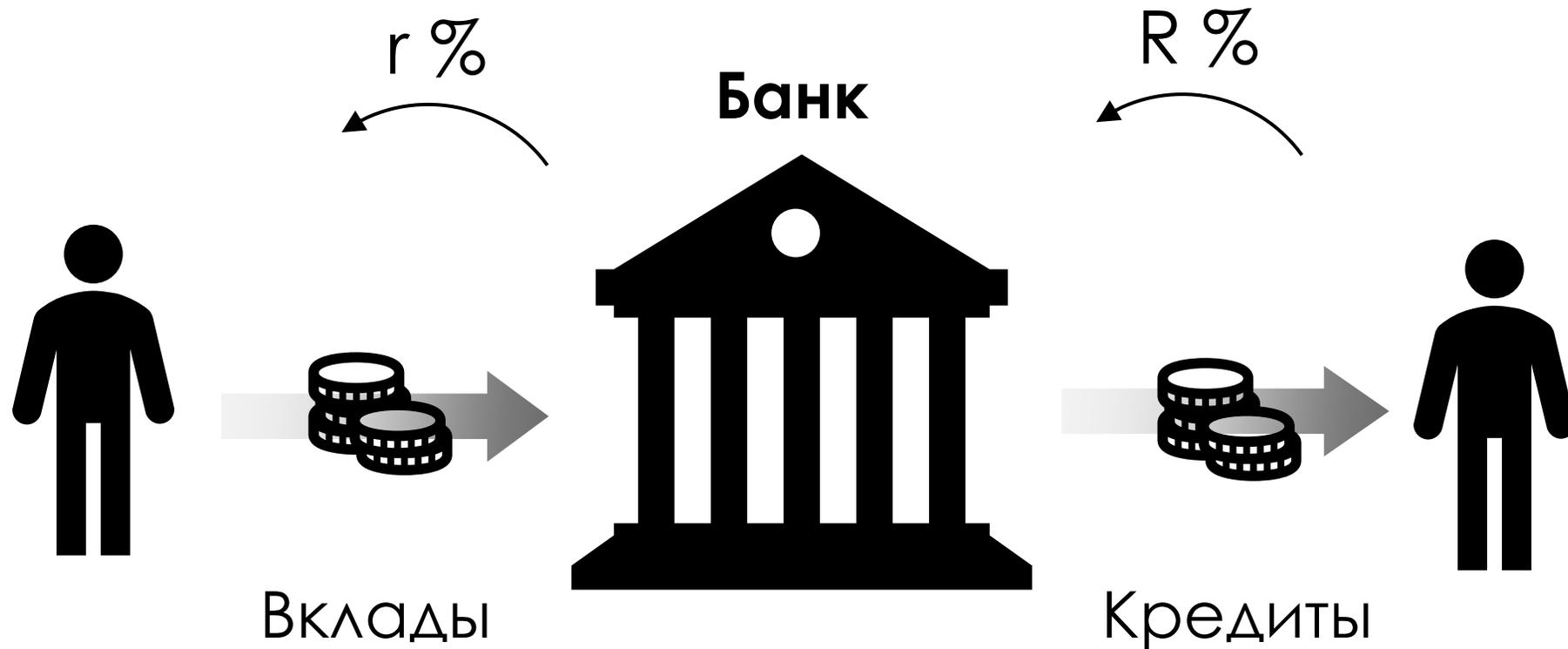


I ВЕК Н.Э. Древний Рим **Аргентáрии**

Изначально «аргентарии» были простыми менялами, занимались проверкой и обменом монет.

По мере развития Римской империи «аргентарии» превратились в настоящих банкиров, принимавших вклады, выдававших кредиты

За 2000 лет принципы банкинга существенно **не изменились**



Прибыль: $(R-r)\%$

✓ Оценка рисков; ✓ Финансирование

ЧТО делают банки осталось **неизменным**.
Изменилось, то **КАК** банки это делают.

Принятие решения о кредитовании

1950-е



Инструмент оценки риска:

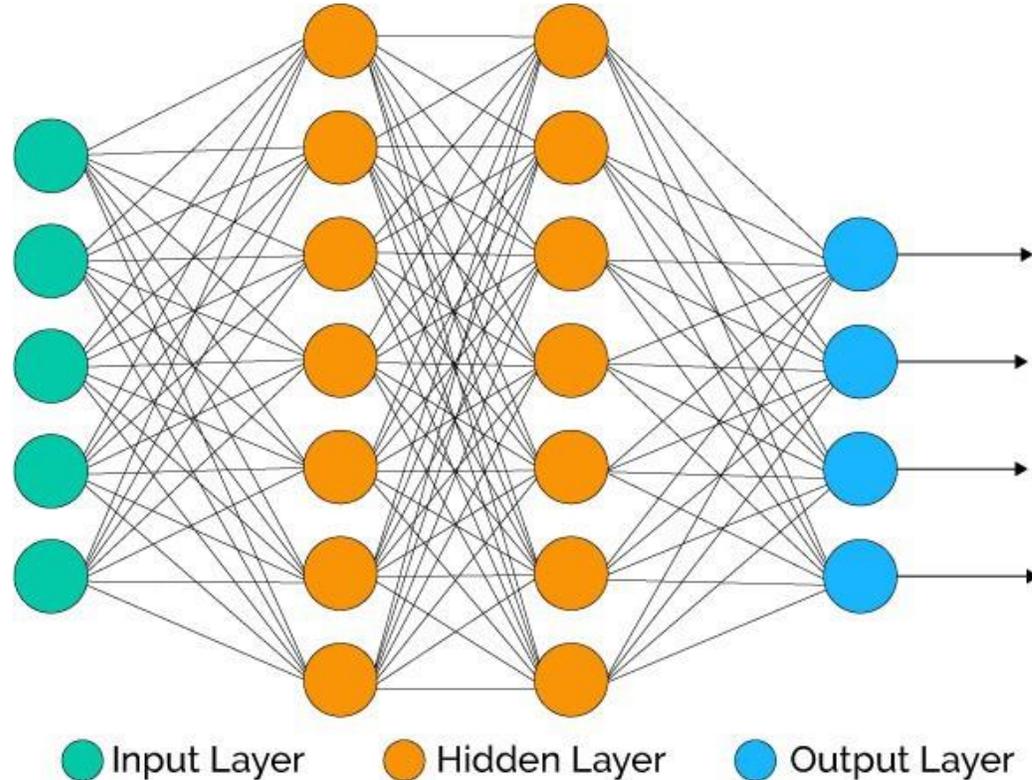
Очное интервью:

- ✓ Внешний вид, поведение
- ✓ Рекомендации от коллег
- ✓ Бумажные справки

Чем плохи эти критерии?

Принятие решения о кредитовании

2019



Инструменты оценки риска:

Статистические модели на основе данных клиента:

- ✓ кредитная история
- ✓ транзакции
- ✓ поведение в соц.сетях

<http://futurebanking.ru/post/3576>

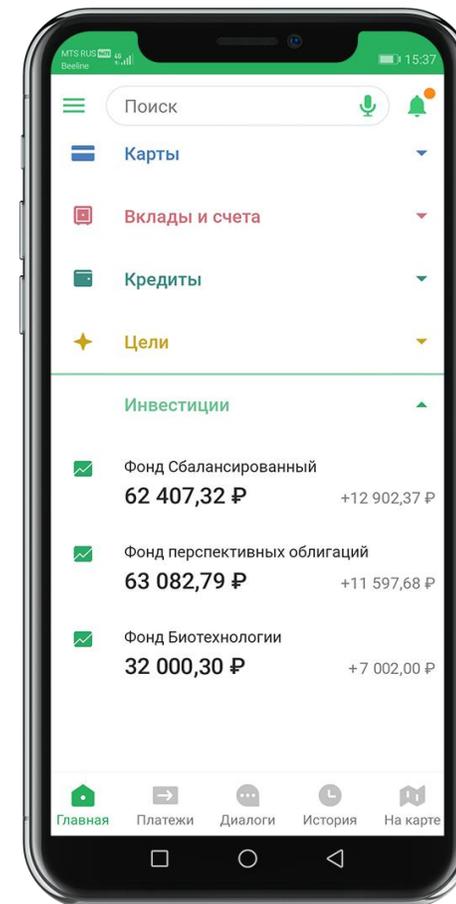
AI-инструмент: нейронные сети

Предложения и каналы предоставления продуктов

1950-е



2010-е



AI-инструмент: рекомендательные системы

Банкоматы и управление наличностью

1960-е



Лондон, 1967 год, очередь у первого банкомата

2010-е



AI-инструмент: прогноз временных рядов, теория очередей

Торговля финансовыми инструментами

1990-е



Инструмент: трейдеры

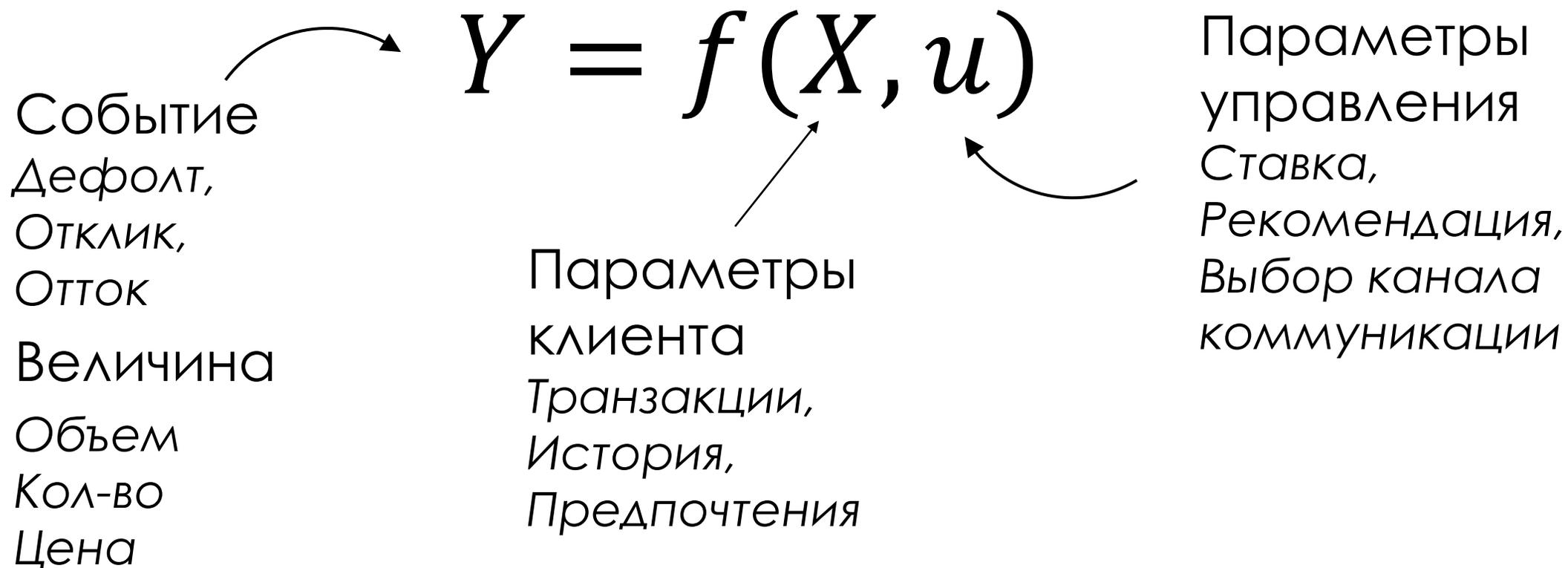
2010-е

```
13 cashBalanceList = []
14
15 timeSteps = len(dateList) # It can be helpful to
16 barIterator = 0
17
18 while barIterator < timeSteps:
19     for symbol in backtestSymbolList:
20         # Historical data input has to be adjusted
21         # Simple moving average cross strategy
22         price = data[symbol]["close"]
23         SMA20 = data[symbol]["SMA20"]
24         SMA50 = data[symbol]["SMA50"]
25
26         if SMA20 > SMA50:
27             openPosition = backtester.retu
```

AI-инструмент: Time-series анализ,
Monte-Carlo

Как работает AI/DS/ML/эконометрика/статистика?

$f = ?$ Какая закономерность?



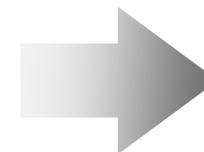
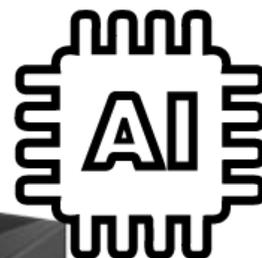
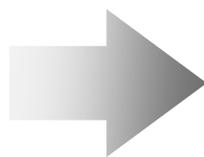
Как работает AI?

Статистика

X *u* *Y*

34	ЗВОНОК	ОТКЛИК
20	SMS	ОТКЛИК
21	SMS	ОТКАЗ
50	ЗВОНОК	ОТКЛИК
41	SMS	ОТКАЗ
25	SMS	ОТКАЗ
19	SMS	ОТКЛИК
82	ЗВОНОК	ОТКЛИК

Гб данных



Оценка закономерности

\hat{f}

AI-алгоритм подбирает параметры закономерности на основе исторических данных



Эволюция скоринговых систем
Алгоритмы

Задача скоринга

Исторические данные

$X_1 \dots X_k \quad Y$

34		выплатил
20		выплатил
21		дефолт
50	...	выплатил
41		выплатил
25		дефолт
19		выплатил
82		выплатил

«Выплатил»: $Y = 0$

«Дефолт»: $Y = 1$

$$\Pr(Y = 1 | X_1, \dots, X_k) = ?$$

Факторы для принятия решений

1. Социально-демографические

- Образование (среднее, высшее, степени)
- Семейное положение
- Возраст
- Стаж
- Категория должности
- Наличие детей
- Оценка дохода

С 1900-х.
1899 Retail Credit Company
будущий Equifax

2. Бюро кредитных историй

- Частота и глубина просрочек в прошлом
- Число активных займов
- Число закрытых займов

С 1950-х.
Fair Isaac Company,
Experian

3. Транзакционная активность

- Динамика операций
- Типы операций

С 2010-х.
Благодаря развитию
LSTM

4. Данные внешних поставщиков:

С 2010-х.

- скорбаллы социальных сетей
- скорбаллы мобильных операторов

5. Неструктурированные данные:

2015-2019 гг.

- Коммуникация с клиентом
- Аудиозаписи,...расшифровка текста

Логистическая регрессия

Модель:

$$\mathbb{P}\{y = 1 \mid x\} = \frac{1}{1 + e^{-z}}$$

$$z = \theta^T x = \theta_0 + \theta_1 x_1 + \dots + \theta_n x_n$$

Подбор параметров:

$$\sum_{i=1}^m \log \mathbb{P}\{y^{(i)} \mid x^{(i)}, \theta\} - \lambda \|\theta\|_1 \rightarrow \max$$

- Простая в имплементации, обучении
- Легко интерпретируемы
- Подвержены выбросам в данных
- Предоставляет сценарный анализ (чувствительность прогноза модели к изменению фактора)

Гиперпараметры:

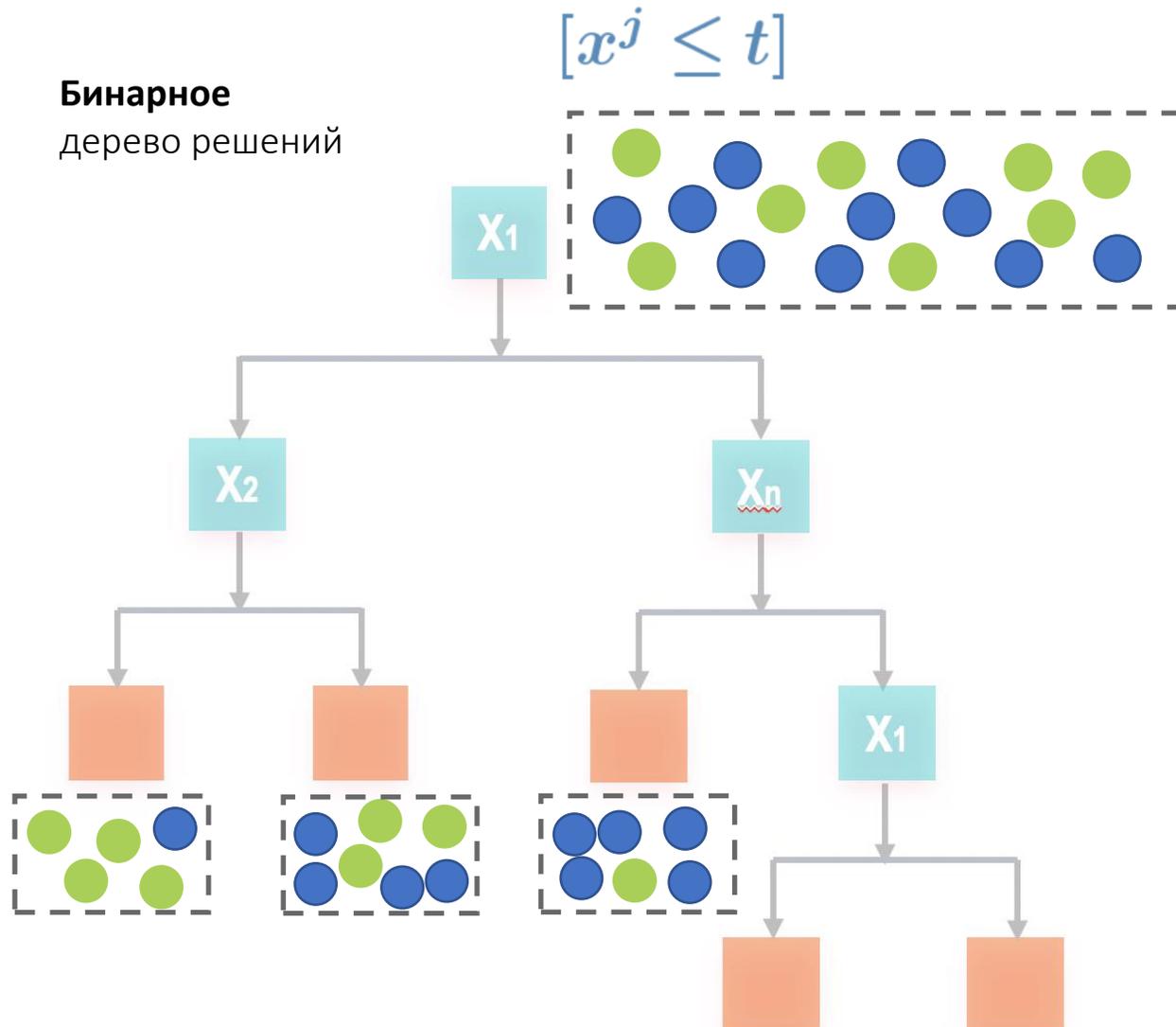
- Регуляризация (λ)

Параметры:

- Коэффициенты факторов (θ)
-

Дерево решений

Бинарное
дерево решений



- Последовательно проверяют простые условия
- Легко интерпретируемы
- Позволяют восстановить нелинейные зависимости
- Легко переобучаются
- Отсутствие локальной чувствительности прогноза модели к изменению фактора

Гиперпараметры (критерии остановки):

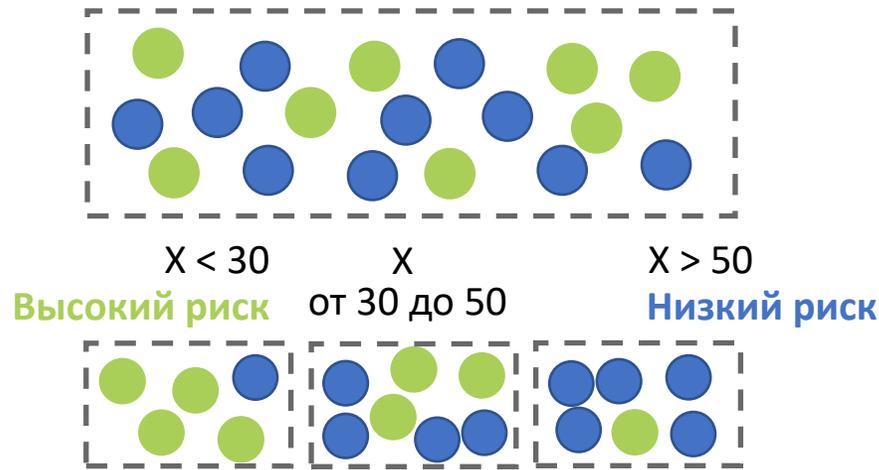
- Глубина дерева
- Минимальное число объектов в листе

Параметры:

- Пороги разбиений

WOE BINNING

WoE (Weight Of Evidence) – метод преобразования исходных факторов модели, позволяющий выделить стабильные и высокопредиктивные диапазоны значений (бины) факторов в модели



$$N_0 = 10$$

$$N_1 = 8$$

Идея метода:

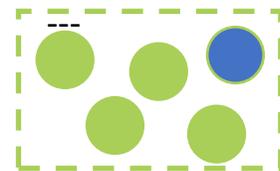
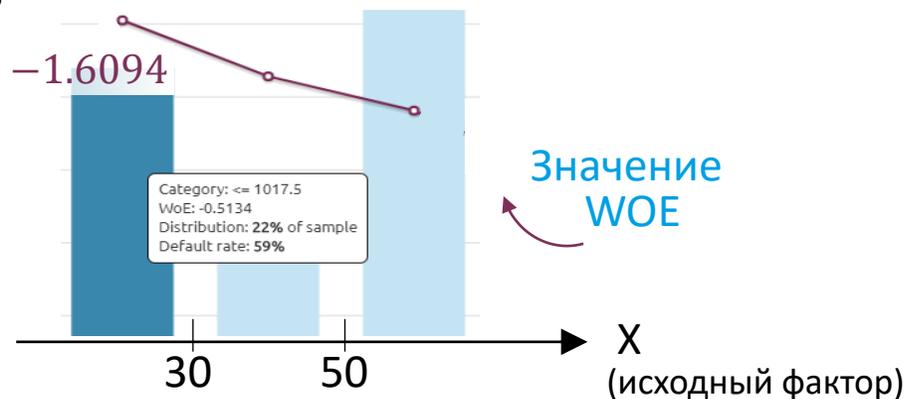
Строятся однофакторные деревья решений, решающие задачу бинарной классификации на уровне каждого фактора

Каждому листу дерева i присваиваются значение трансформированного фактора WOE:

$$WOE(i) = \ln \left(\frac{\frac{N_0(i)}{N_1(i)}}{\frac{N_0}{N_1}} \right)$$

Где $N_0(i)$ и $N_0/N_1(i)$ и N_1 – количество наблюдений, где target=0/target=1 в группе i и по всей выборке, соответственно

$i = 1, 2, 3$



$X < 30$

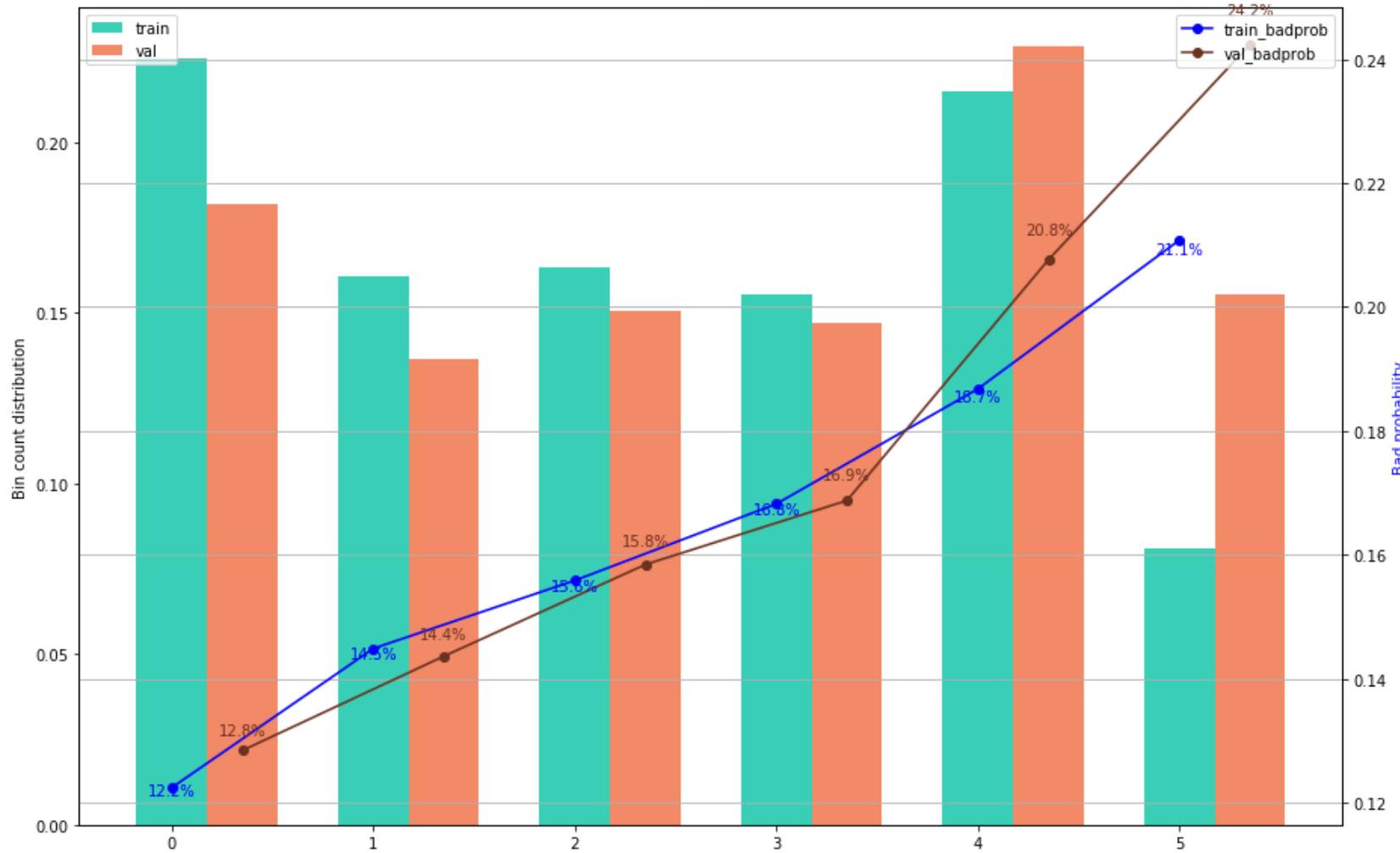
$$N_0(i) = 1 \quad N_0 = 10$$

$$N_1(i) = 4 \quad N_1 = 8$$

$$WOE(i) = \ln \left(\frac{1/4}{10/8} \right) = -1.6094$$

WOE BINNING пример

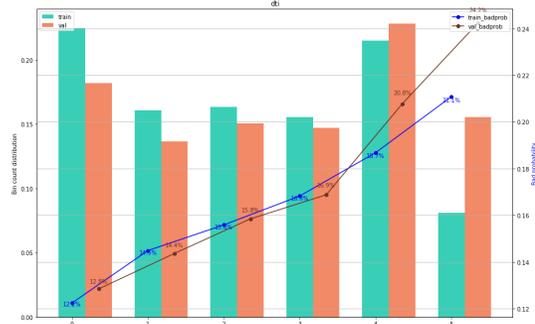
dti



Уровень дефолта

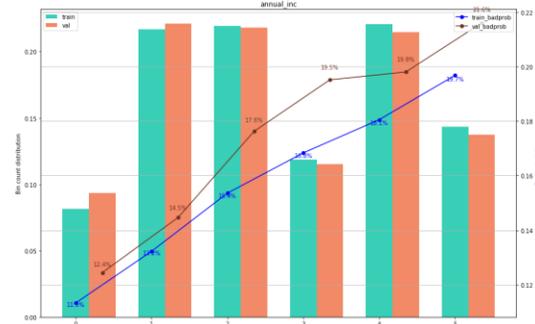
WOE BINNING + Лог.рег = Скоринговая карта

фактор1



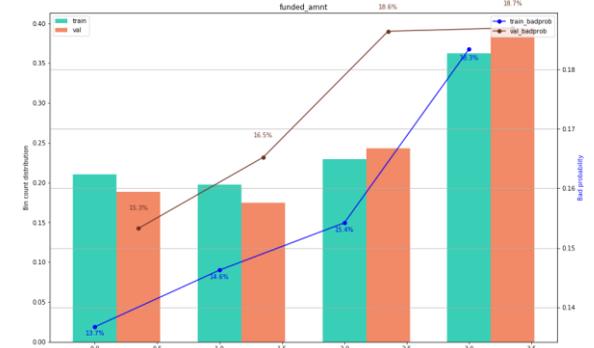
θ_1

фактор2



θ_2

фактор k



θ_k

$$z = \theta^T x = \theta_0 + \theta_1 x_1 + \dots + \theta_n x_n$$



$$\mathbb{P}\{y = 1 \mid x\} = \frac{1}{1 + e^{-z}}$$



ЭВОЛЮЦИЯ ИСТОЧНИКОВ ДАННЫХ



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Факультет компьютерных наук
Финансовые технологии и анализ данных

РАСПОЗНАВАНИЕ И АНАЛИЗ ЭМОЦИЙ В ЧЕЛОВЕЧЕСКОЙ РЕЧИ НА ОСНОВЕ АУДИОЗАПИСЕЙ

Студентка: Ковалева Ю.Д.

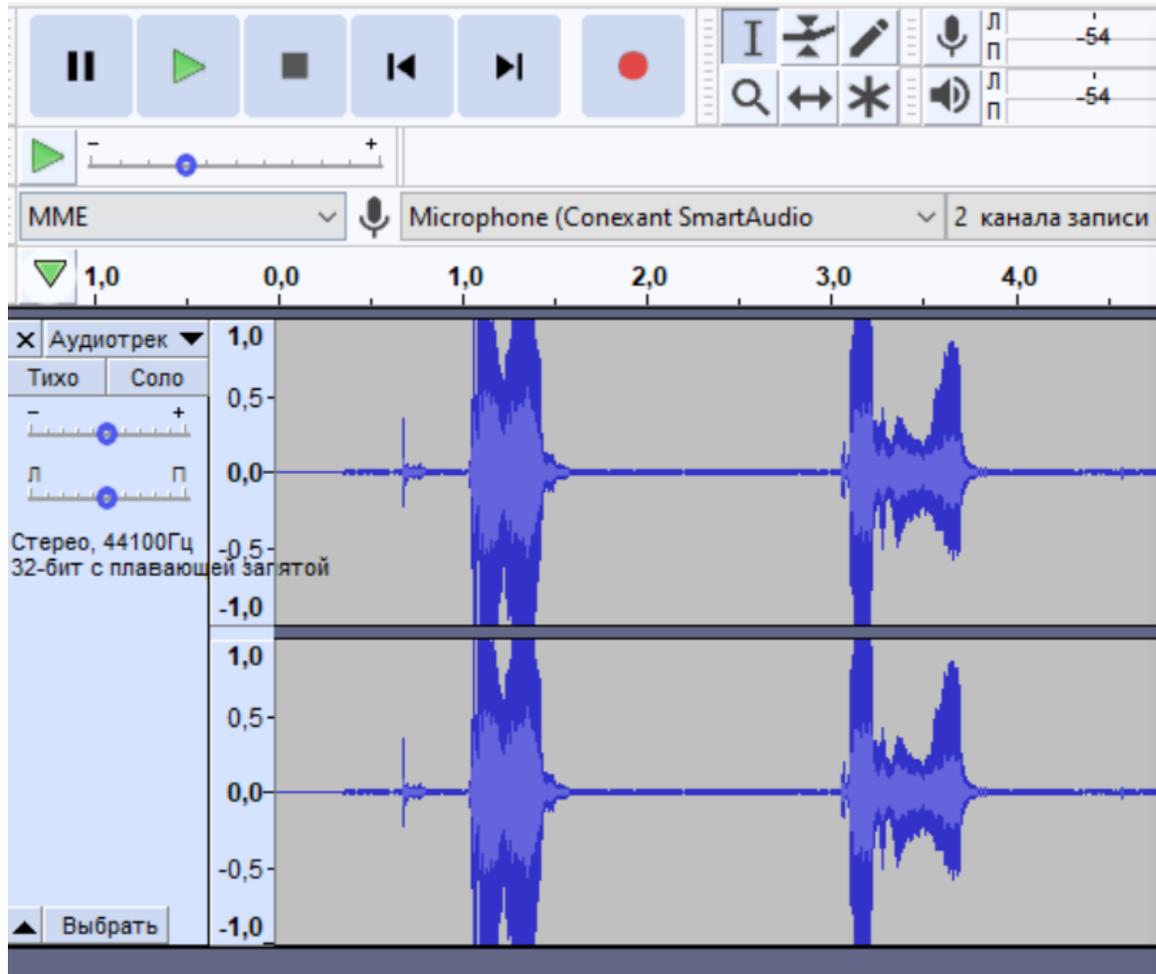
Москва, 2019

ПРИЗНАКОВЫЕ ПРОСТРАНСТВА

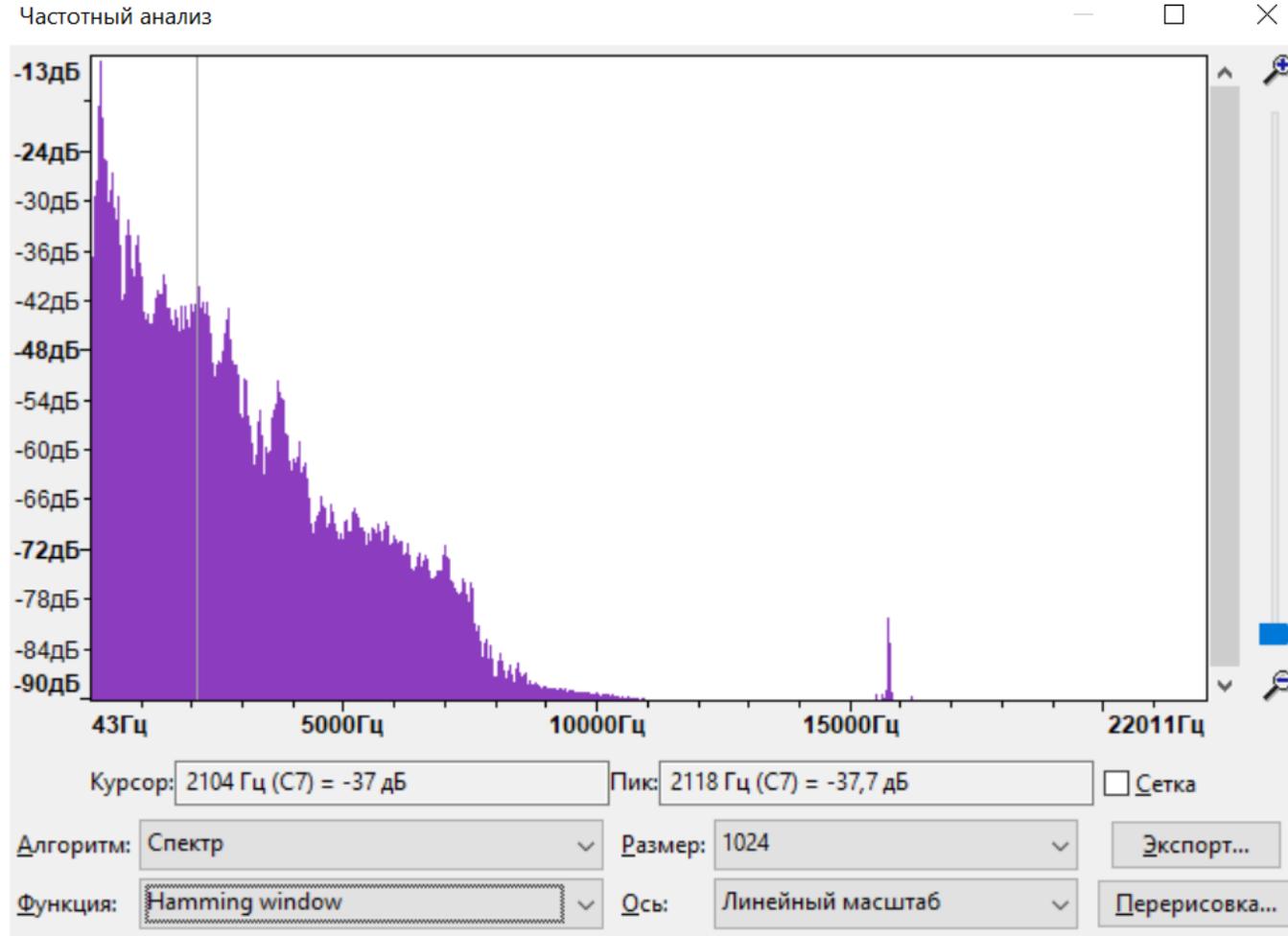
Звуковой волна эмоционально окрашенной речи -> признаки для классификации эмоций

- Среднее и нормализованные средние значения спектра;
- Медианное значение спектра речи и тд;
- Относительная мощность спектра речи;
- Интенсивность, амплитуда;
- Энергия;
- Частота основного тона F_0 и др;
- Кепстральные признаки (MFCC, LPC и тд)
- Дискретное Вейвлет преобразование (DWT)

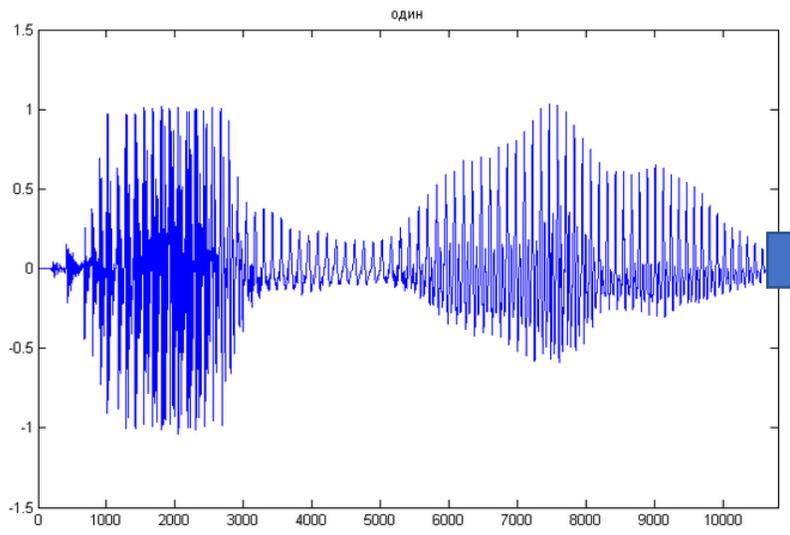
Звуковой сигнал во времени



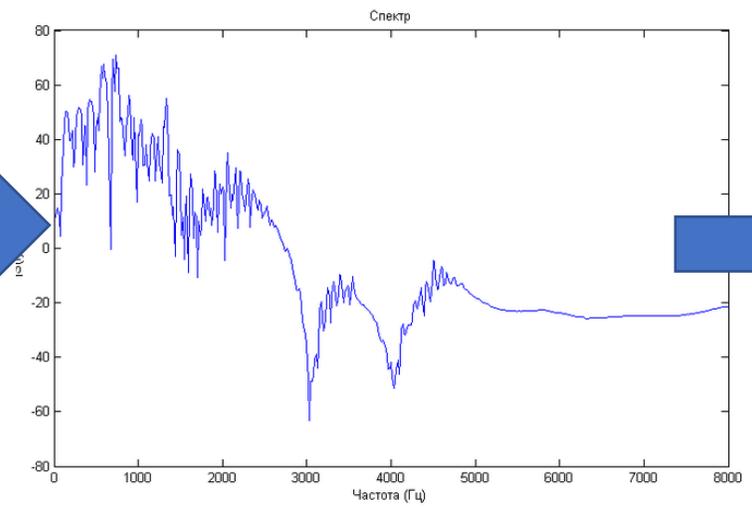
Спектр звукового сигнала: Распределение амплитуды по частоте



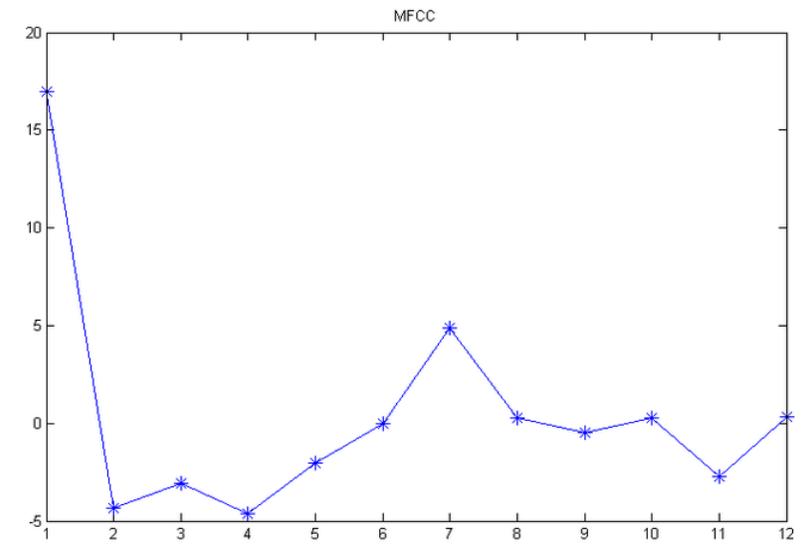
Звуковой сигнал во времени



Выделение спектра



Расчет MFCC



$$M = 1127 * \log(1 + F/700)$$

Сложный неструктурированный аудиоряд
теперь представим 12 числами (MFCC)!



Задача скоринга + обогащение MFCC данными

Исторические данные

$X_1 \dots X_k \quad Y$

34		ВЫПЛАТИЛ
20		ВЫПЛАТИЛ
21		дефолт
50	...	ВЫПЛАТИЛ
41		ВЫПЛАТИЛ
25		дефолт
19		ВЫПЛАТИЛ
82		ВЫПЛАТИЛ

+

$MFCC_1 \dots k$

-0.5	
0.7	
...	

- Признаки, полученные на основе аудиозаписей разговоров клиентов, имеют высокую дискриминирующую способность и могут быть использованы в процессе принятия решения о выдаче кредита;

ИНТЕРПРЕТАЦИЯ

MFCC коэффициенты:

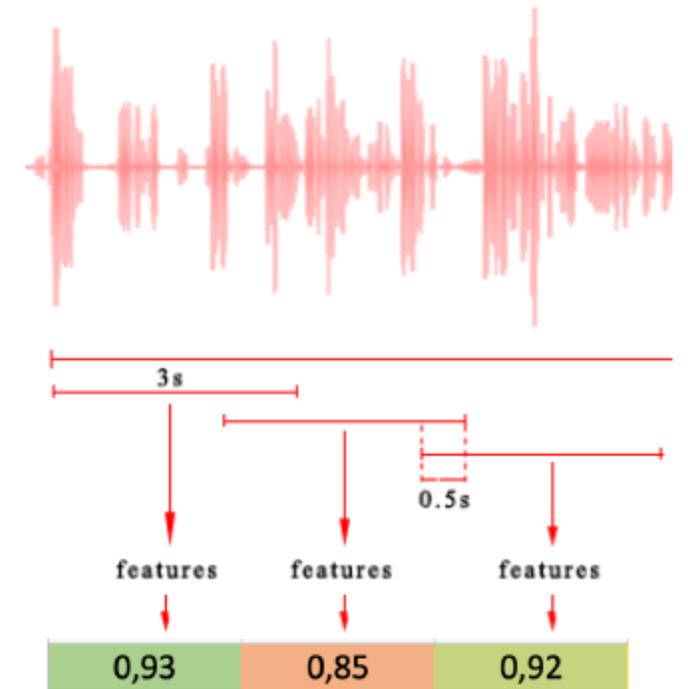
- C0 - мощность всего частотного диапазона
- C1 - баланс между низко- и высокочастотными компонентами во фрейме

Остальные признаки, в большинстве своем, не интерпретируемые

Для выявления факторов, влияющих в лучшую или худшую сторону, были прослушаны фрагменты записи, где происходил большой скачок предсказания модели, что позволило сделать субъективные выводы:

- Факторы, соответствующие **ухудшению** оценки:
 - тянущие звуки ('ну-у-у' 'э-э-э');
 - резкое изменение тона разговора (повышение/понижение);
 - презрение;
 - музыка на заднем плане.

- Факторы, соответствующие **улучшению** оценки:
 - спокойный тон разговора;
 - доброжелательность;



*изменение предсказания модели
в течение разговора*

В 2017 г. открыта
магистратура
ФТиАД

- Фокус на **3** области:
 - машинное обучение,
 - экономика
 - и программирование



11
июл
2017

Состоялось подписание договора о сотрудничестве между НИУ ВШЭ и Сбербанком



Документ подписали президент, председатель правления Сбербанка Герман Греф и

<https://www.hse.ru/ma/fintech/news/207382675.html>

Graduation ceremony



Agile Home Сбербанк, первый выпуск



ВОЗМОЖНОСТИ
ЛЕТНИХ ШКОЛ
ДЛЯ ШКОЛЬНИКОВ

<https://cs.hse.ru/sber/>

Контакты

amasyutin@hse.ru – Алексей Масютин