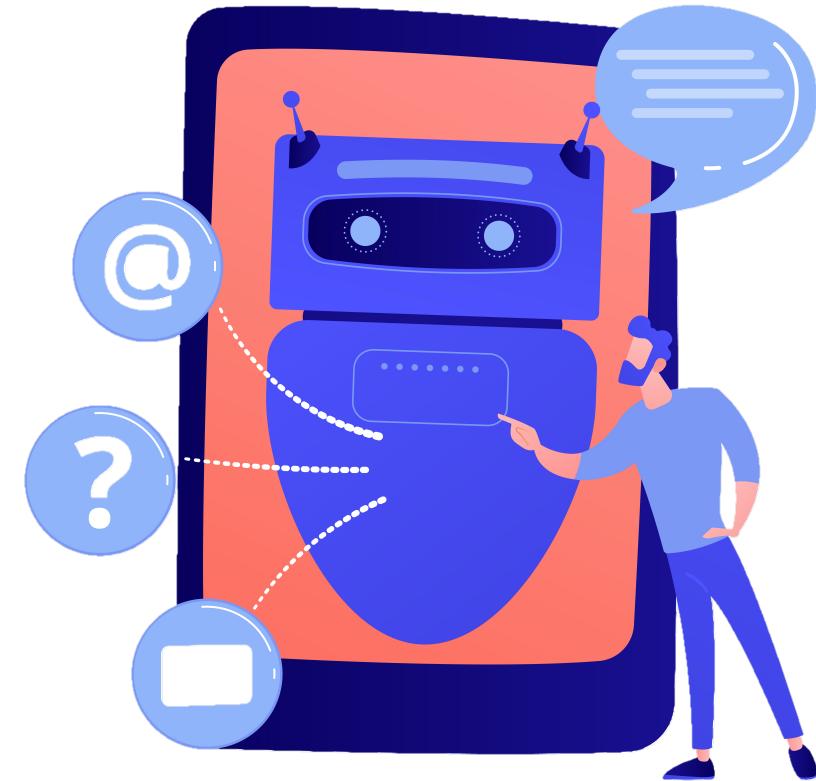


# ОПРЕДЕЛЕНИЕ ЭМОЦИОНАЛЬНОГО ОКРАСА СООБЩЕНИЯ В ЧАТЕ

**FinTech-bomb:**

Данилова Екатерина  
Зорина Кристина  
Лепёхина Анфиса  
Плевако Матвей

Резюме Альтернативы Описание Эффект Внедрение Риски



**ПОЧТА  
БАНК**

**Проблема**

Ежедневно клиенты оставляют большое количество сообщений и отзывов о работе Почта Банка. Однако, на данный момент нет технологий, которые позволяли бы оперативно анализировать потребности клиентов и определять тональность их обращений.

**Цель**

**Автоматизировать** процесс определения эмоционального фона сообщений, классифицировать их в одну из трех категорий: негативные, позитивные, нейтральные. Настроить рекомендации для каждой из трех категорий.

**Решение**

**Разработка** сервиса, который с помощью машинного обучения сможет определять эмоциональный окрас входящих обращений и классифицировать их по категориям.

**Разработка** «Кризисного помощника», который оперативно анализирует обращение клиента и предоставляет резюмированную информацию менеджеру.

**Эффект**

Увеличение доли положительно настроенных клиентов до 71%.

Сокращение времени обработки обращения клиента

Оптимизация работы кризисных менеджеров.

Срок окупаемости проекта 1 год.

# На основании сравнения нескольких альтернатив по основным критериям и их результатам была выбрана логическая регрессия

Критерии	Logistic Regression	SVM	Bayes models	Random Forest	Gradient Boosting	Recurrent Neural Networks
Качество работы с текстом	4	4	4	2	3	5
Устойчивость к выбросам	4	3	4	5	5	4
Интерпретируемость	5	3	5	5	4	3
Переобучение	0 Низкая вероятность	0 Низкая вероятность	0 Низкая вероятность	1 Высокая вероятность	1 Высокая вероятность	1 Высокая вероятность
Скорость работы	5	4	5	2	3	2
Масштабируемость, работа с большим количеством признаков	1	1	1	0	0	1
Работоспособность в условиях ограниченных машинных ресурсах	5	5	5	4	4	2
Качество на обучающей выборке, кросс-валидация	0,79	0,78	0,78	0,75	0,69	0,86
Качество на тестовой выборке	0,81	0,80	0,84	0,71	0,70	0,52
Итого	25,6	21,58	25,6	20,46	21,39	19,38

## Основные этапы создания модели для определения эмоционального окраса сообщений в чате

## Предобработка данных

Для обучения был выбран датасет - **Sentiment140**, который состоит из размеченных твитов с различной тональностью



**Конвертация** данных в формат **.parquet** для более быстрой и простой загрузки данных в **GoogleColab**, где проводилась вся работа и обучались модели.

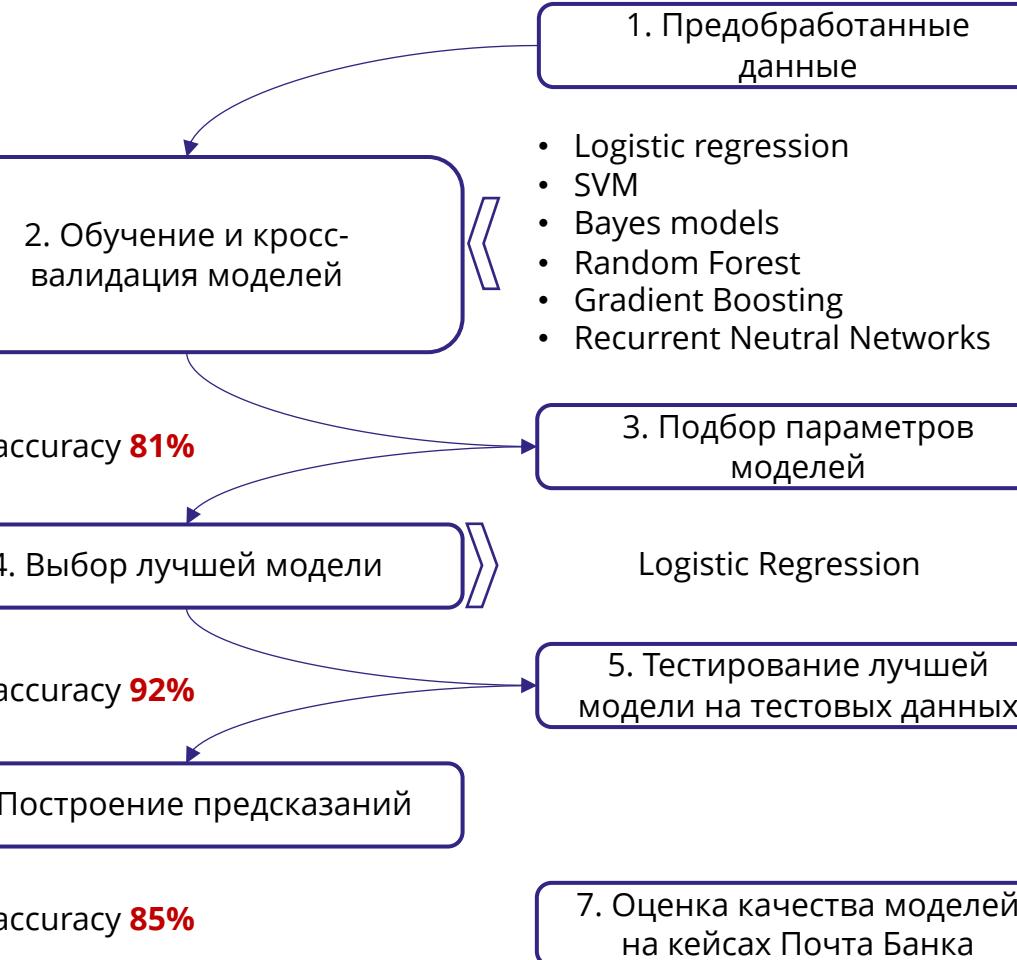


## Применение **CountVectorizer()** для преобразования коллекции текстовых документов в вектор.



Применение **TfidfTransformer()**, для повышения влияния уникальных слов в документе.

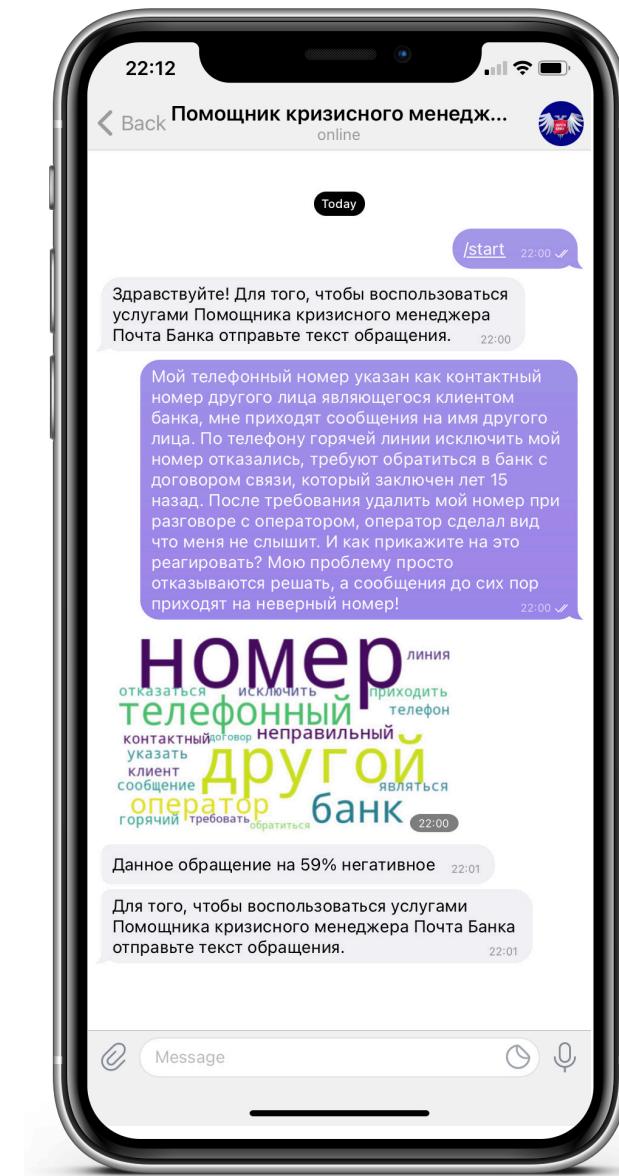
## Обучение модели



## Схема взаимодействия с клиентом, которая позволит увеличить долю положительно настроенных клиентов



- анализируется вся история** обращения и его контекст, а не отдельные сообщения клиента;
- автоматически** определяется **тональность обращения** клиента;
- в случае классификации обращения как «негативное» менеджеру предоставляется **резюмированная информация об обращении** клиента с помощью бота «Помощник кризисного менеджера»;
- система **«Помощник кризисного менеджера»** - предоставление резюмированной информации об обращении менеджеру в виде облака слов, процент негативной настроенности клиента.



## Преимущества внедрения ML сервиса для банка

### Повышение эффективности работы с клиентами

- сегментация клиентов
- сокращение оттока недовольных клиентов
- рекомендация подходящего продукта

### Получение конкурентного преимущества

- преимущества времени и эффективности
- оптимизация человеческих ресурсов, работа только с недовольными клиентами
- прямое взаимодействие с клиентами

### Улучшение бизнес-метрик

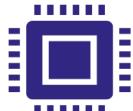
- Повышение рентабельности капитала
- Увеличение доходов бизнеса
- Увеличение числа лояльных клиентов

Прогнозируемые выгоды за два года от внедрения проекта в 8,5 раз превышают первоначальные затраты. Период окупаемости равен одному году при равномерном приросте клиентов на 50% от планируемого.

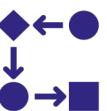
Стоимость разработки и внедрения модели ~ **10 740 000 ₽**



Сбор данных



Данные и аналитическая платформа



Ресурсы проекта



Внедрение проекта и эксплуатационные расходы

Почта Банк имеет собственные каналы сбора данных о клиентах из онлайн-банка и приложения. Значительные затраты на сбор отсутствуют.

Jupiter Notebook - бесплатная среда разработки.



Единовременные расходы на инфраструктуру, интеграцию и настройку системы для работы кризисного менеджера.

~ 2 100 000 ₽

Ежегодные затраты на контроль, дальнейшее обучение и улучшение модели (з/п сотрудников) в течение двух лет.

~ 4 320 000 ₽

### Прогнозируемые доходы

N (выборка)	Predicted: negative	Predicted: positive/neutral	Всего
Actual: negative	True Negative (TN) 32,86%	False Positive (FP) 0,48%	33,33%
Actual: positive/neutral	False Negative (FN) 2,23%	True Positive (TP) 64,44%	66,67%
Всего	35,09%	64,91%	

\*Модель предсказала, что 32,86% клиентов оставят негативный отзыв. Из них предположительно 20% (6,57%) после общения с кредитным менеджером оставят положительную оценку.

Совокупный прирост дохода от клиентов ~ **Σ 91 315 151,72 ₽**

## Этапы разработки и внедрения сервиса для его запуска во втором квартале 2022 года

**Изучение бизнес-контекста** для корректного понимания требований к реализуемому проекту, какие **проблемы пользователей** он должен решать и функции необходимо в нем обеспечить.

**Сентябрь 2021 - Октябрь 2021**

**1**

**Октябрь 2021 - Ноябрь 2021**  
**Сбор и анализ требований к сервису**, которые позволят сделать то, что необходимо компании и отрасли в целом. На этом этапе будут проведены **исследование существующих предложений**, частота обращений в банк с различными целями, потребностей целевой аудитории.

**Дизайн окна** непосредственно с обращениями, также небольшие изменения в дизайн существующего приложения банка для добавления новой иконки мессенджера.

**Декабрь 2021 - Январь 2022**

**2**

**3**

**Тестирование модели**, полученного сервиса, чтобы выяснить, насколько он удобен в использовании для клиентов и его **точность в решении поставленной задачи**.

**Февраль-Март 2022**

**4**

**5**

**Сопровождение сервиса**, чтобы обеспечить возможность своевременного устранения возможных неисправностей

**Апрель 2022 - Декабрь 2022**

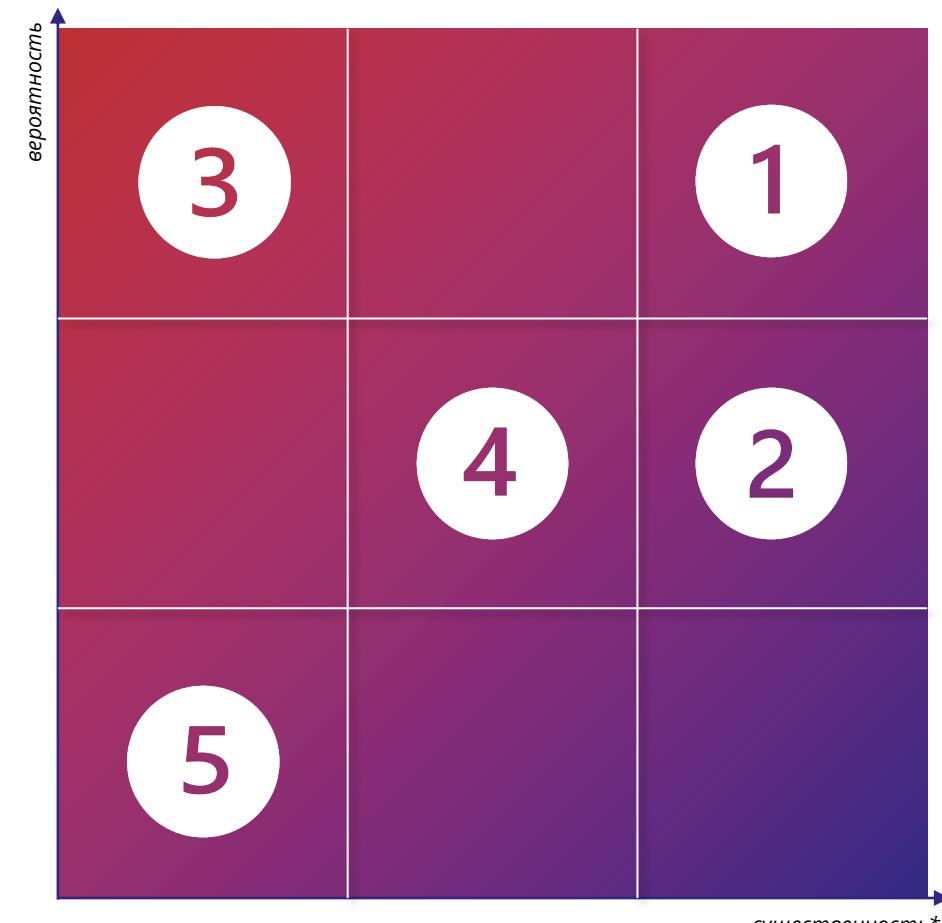
**6**

**7**

**Апрель 2022**

**Внедрение разработанного алгоритма в онлайн-банк и мобильное приложение**, оповещение пользователей о появившейся функции. Еженедельный анализ, насколько корректно сервис работает на практике.

Возможные риски	Меры для снижения влияния рисков
1. Неверное определение смысла и тональности сообщения в случаях употребления пользователем сарказма или иронии.	Дополнительное обучение используемой модели на данных, содержащих сарказм, иронию, разговорные фразы, сленг и нецензурных высказываний, чтобы распознать их в момент обработки обращения клиента.
2. Употребление сокращений, сленгов, разговорных фраз, нецензурных высказываний, вследствие чего смысл обращения может быть определен неверно, проблема клиента не будет решена.	Использование предобученных сетей, которые способны различать сарказм и иронию.
3. Использование клиентом слишком большого количества слов для описания обращения, трудности с выделением сущности проблемы	Дополнительное построение облака слов, что позволит выделить сущность проблемы на основе часто используемых слов
4. Сложность определения тональности обращения, содержащих информацию и факты без эмоциональной оценки.	Ввод дополнительного класса «нейтральный», в который будут относиться обращения без эмоциональной окраски.
5. Использование клиентом фактов, не относящихся к работе банка, жалобы на посторонние вещи.	Использование рекурентных нейронных сетей, при которых обращений анализируется целиком, смысл и эмоциональный окрас распознаются с учетом контекста.



\*Зависит от того, насколько критично то, что может быть пропущено негативное обращение клиента, из-за чего Банк не решит проблему клиента

## Приложение: расчет экономического эффекта

Затраты на внедрения сервиса состоят из двух частей: инвестиционные затраты (являются единовременными, единственная переменная часть з/п программистов, работающих с моделью, может быть спрогнозирована) и затраты на з/п кризисного менеджера. Последнее учитывается в расчете потенциального дохода от клиента, так как возможен вариант резкого роста негативно настроенных клиентов, что повлечет пропорциональный рост данной составляющей затрат.

Объем выборки = 9 681	Predicted: negative	Predicted: neutral	Predicted: positive	Всего
	True Negative (TN)	False Positive (FP)	False Positive (FP)	
Actual: negative	32,86%	0,29%	0,19%	33,3%
Actual: neutral	1,80%	28,86%	2,68%	33,3%
Actual: positive	False Negative (FN)	True Positive (TP)	True Positive (TP)	
	0,43%	1,54%	31,36%	33,3%
Всего	35,1%	30,7%	34,2%	100,0%

\*Для расчета экономических выгод, нейтральное и позитивное сообщение от клиента оценивается одинаково: клиент доволен сервисом и работой банка

## Допущения для расчета потенциальных доходов

Количество клиентов рост за 2 года (2021-2023), чел.	16 500 000 -> 21 000 000
Доходность одного клиента, руб*	389
Зарплата кризисного менеджера, руб	30 000
Доля клиентов, использующих онлайн банк и приложение	30%
Количество клиентов, пользующихся онлайн банком и приложением, чел.	6 300 000
Доля людей, которые после общения с кредитным сотрудником изменили отрицательный комментарий на положительный	15%
Клиенты перешедшие из категории негативные в позитивные	6,57%
Количество клиентов, пользующихся онлайн банком и приложением	5 250 000
Доходность одного клиента	389,82
Доходы на клиентов	134 491 500,69
Расходы на кредитного сотрудника, руб/мин	117,19
Среднее время работы с клиентом, мин	15
Расходы на клиентов	43 176 348,97
Итого доход (с учетом инвестиционных затрат)	91 315 151,72

\*Доходность одного клиента рассчитывается как прибыль банка деленная на общее количество клиентов

## Приложение: команда



**Кристина Зорина**

Проект  
НИУ ВШЭ г. Санкт-Петербург  
[kszorina@bk.ru](mailto:kszorina@bk.ru)



**Анфиса Лепёхина**

Разработчик  
СПбГЭУ  
[anfisa\\_lepekhina5@mail.ru](mailto:anfisa_lepekhina5@mail.ru)



**Матвей Плевако**

Разработчик  
Иннополис  
[matveyplevako@mail.ru](mailto:matveyplevako@mail.ru)



**Екатерина Данилова**

Аналитик  
ФУ при Правительстве РФ  
[keth.danilova@yandex.ru](mailto:keth.danilova@yandex.ru)