

Security
Band

Changing the world for the better

Kvirin КРДА

АВТОМАТИЗИРОВАННЫЙ АЛГОРИТМ
ОБЕЗЛИЧИВАНИЯ ДАННЫХ



Проблема

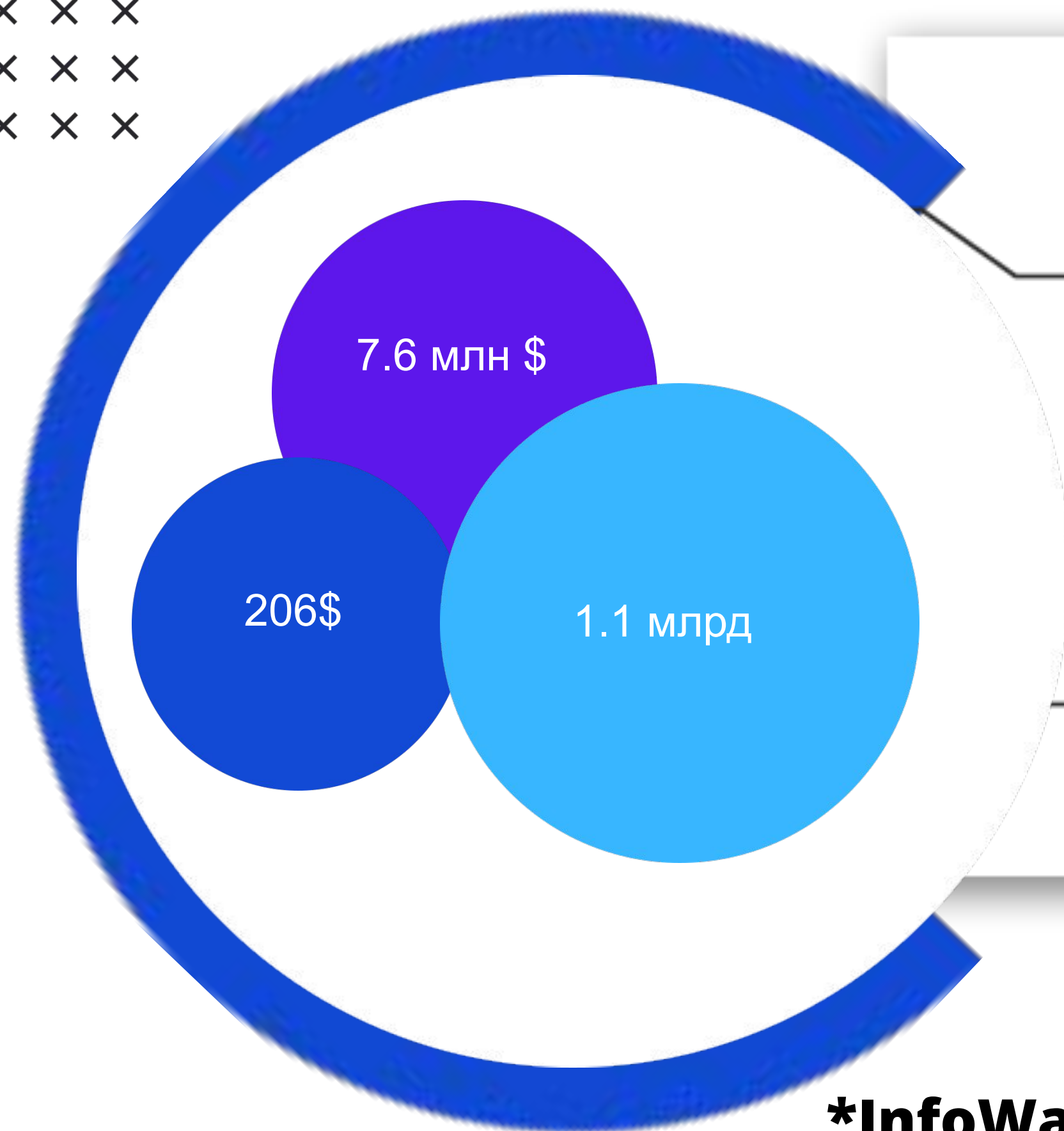
- Утечки ПДн;
- Исполнение 152 ФЗ;
- Исполнение приказа № 996 РКН;
- Введение уголовной ответственности;
- Введение оборотных штрафов.



× × ×
× × ×
× × × ×
× × × ×



Аналитика



В 2020 году было зафиксировано более 1,1 млрд нарушений, что в 2,5 раза больше, чем в 2019 году. *

Средний размер ущерба от нарушений данных в России составлял 7,6 млн. долларов за одно нарушение**;

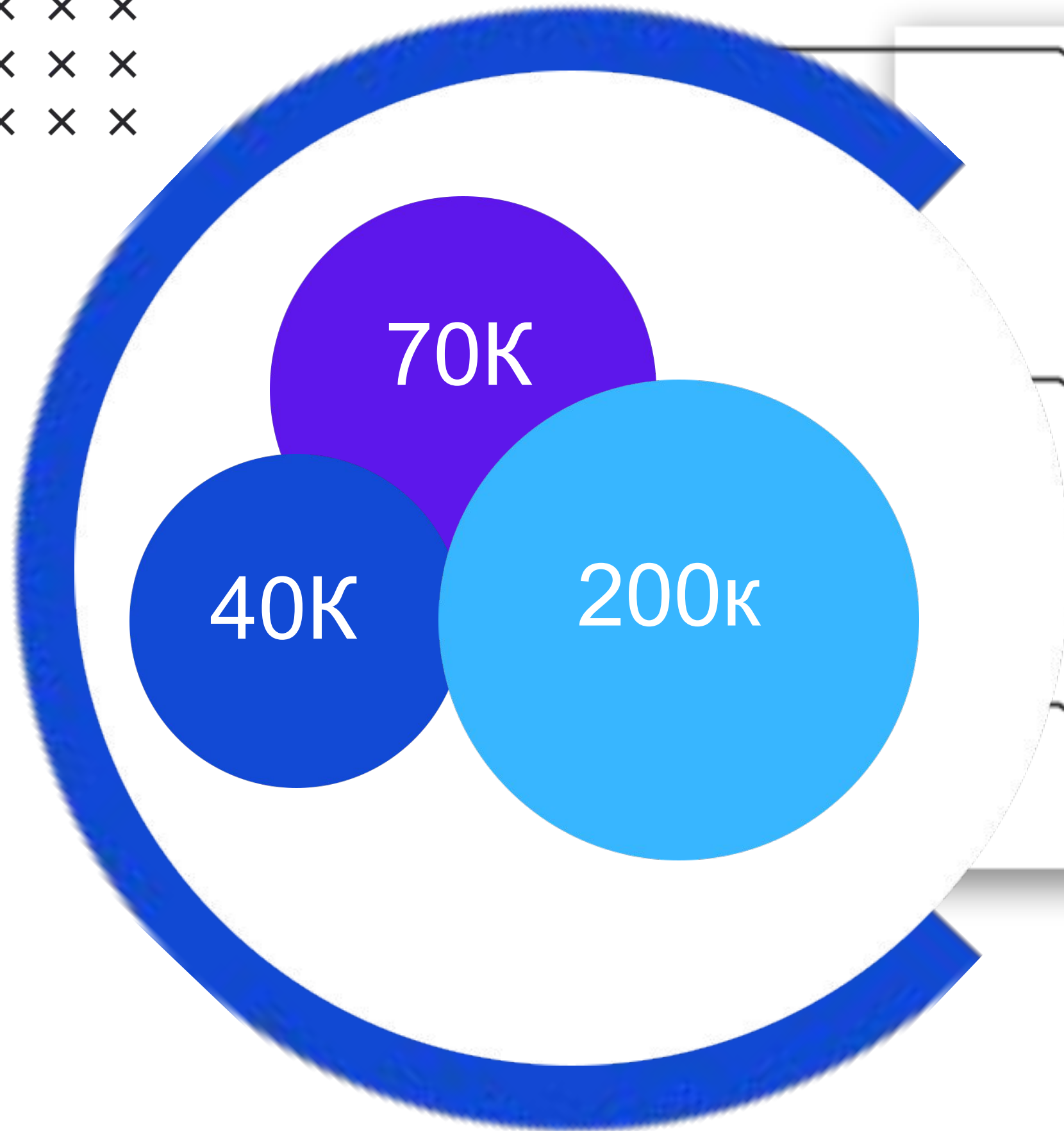
Стоимость потери одной записи с персональными данными - 206 долларов

***InfoWatch**

****Ponemon Institute**



× × ×
× × ×
× × ×
× × ×



• 40 тыс имен

• 70 тыс отчеств

• 200 тыс фамилий

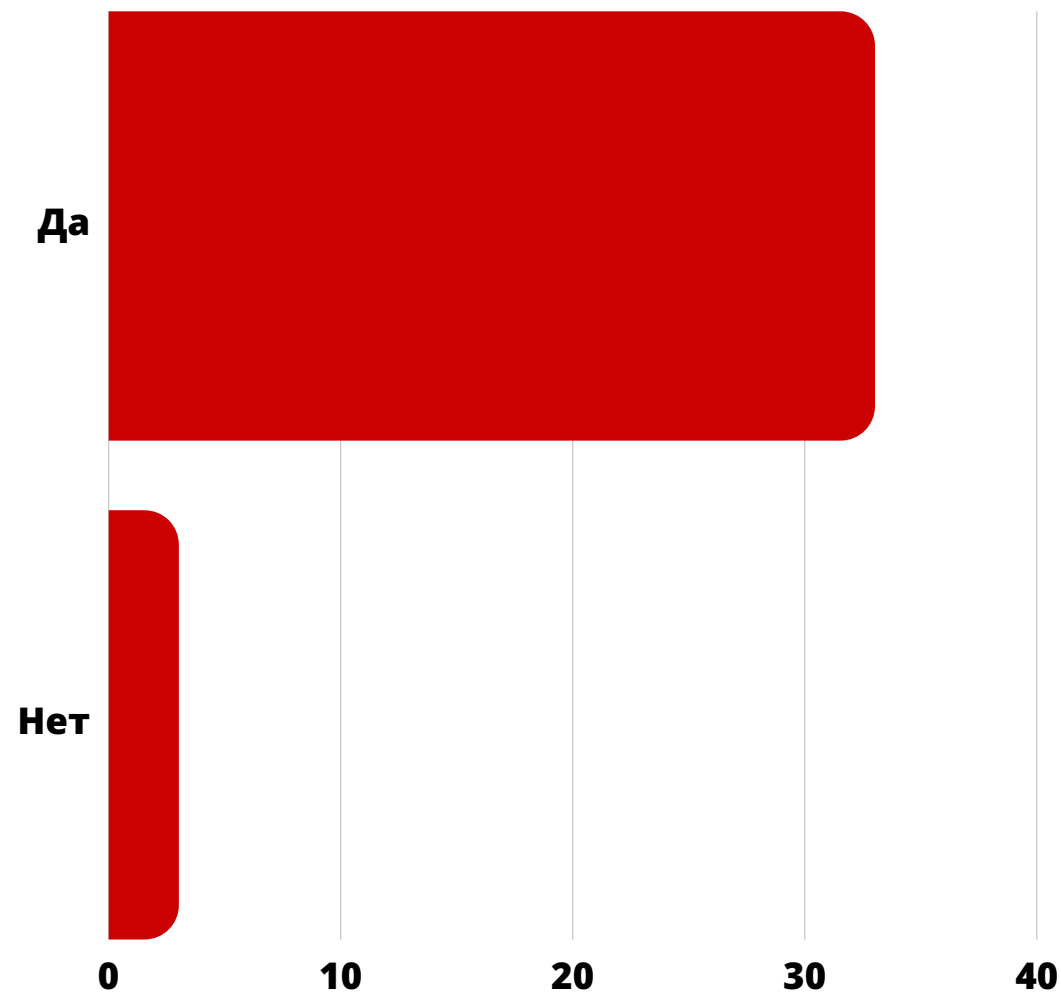
С учетом разных склонений, мужского, женского рода



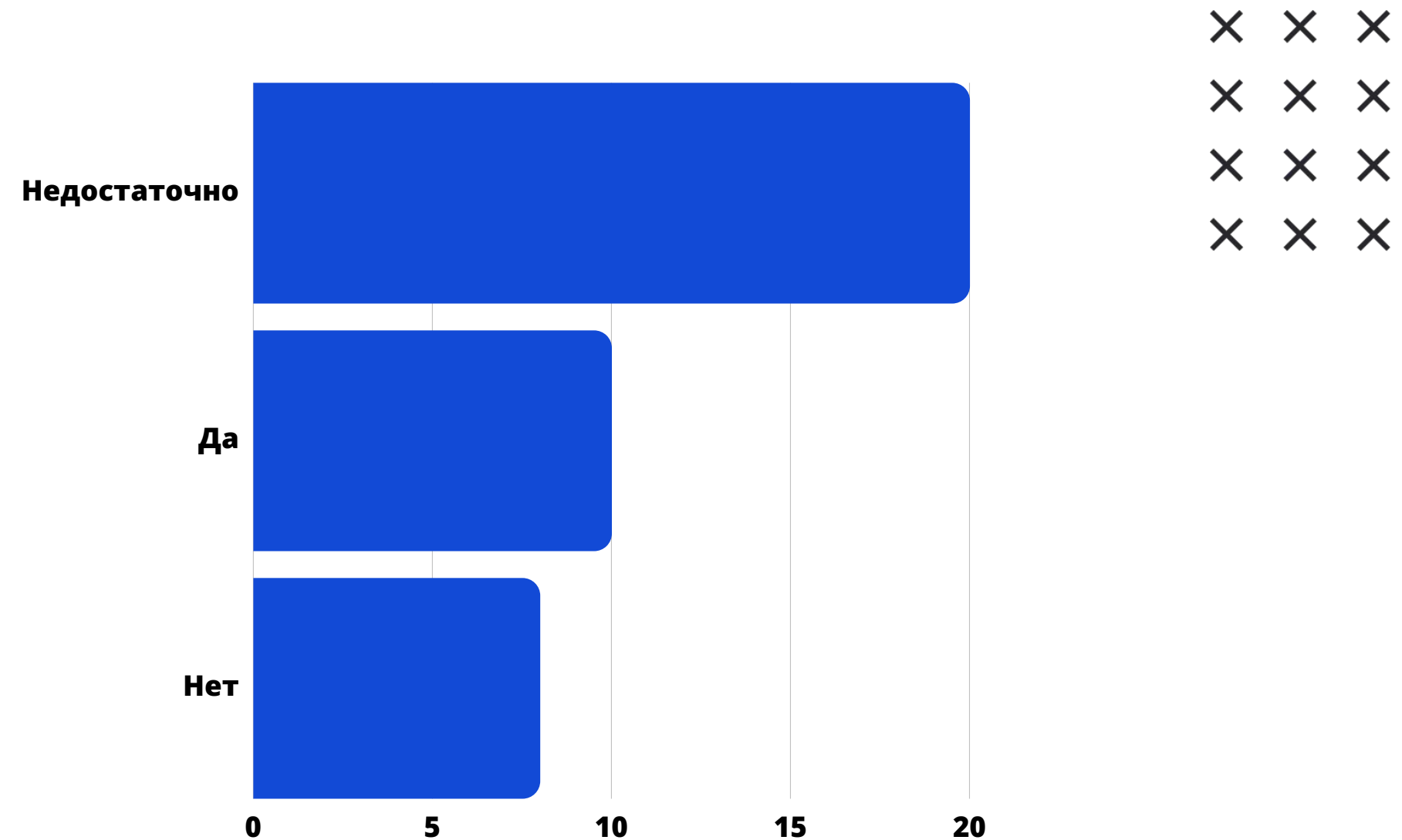
Результаты CustDev

05

Есть ли внутри компании (в которой вы работаете) документы, содержащие персональные данные, которые нужно защищать для выполнения требований регуляторов?

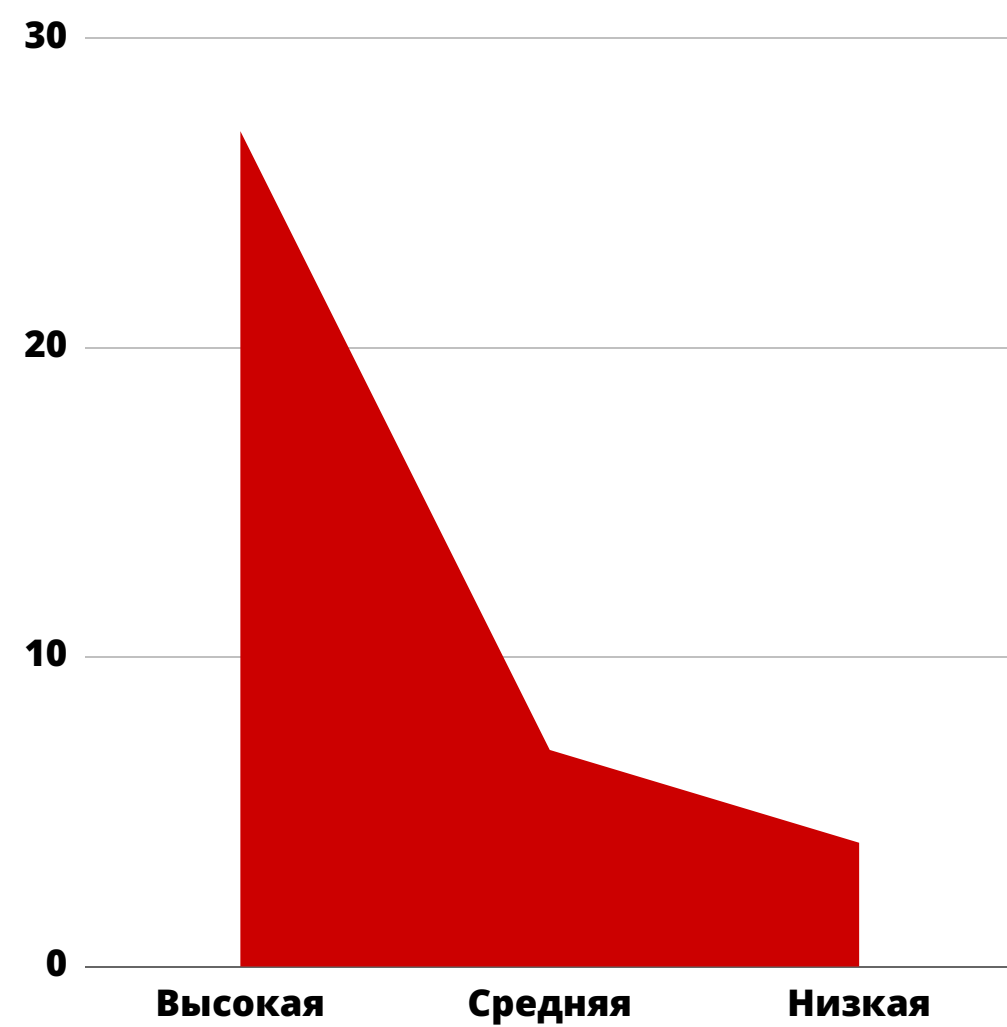


Полностью ли компания (в которой вы работаете) реализует требования по защите персональных данных регуляторов?

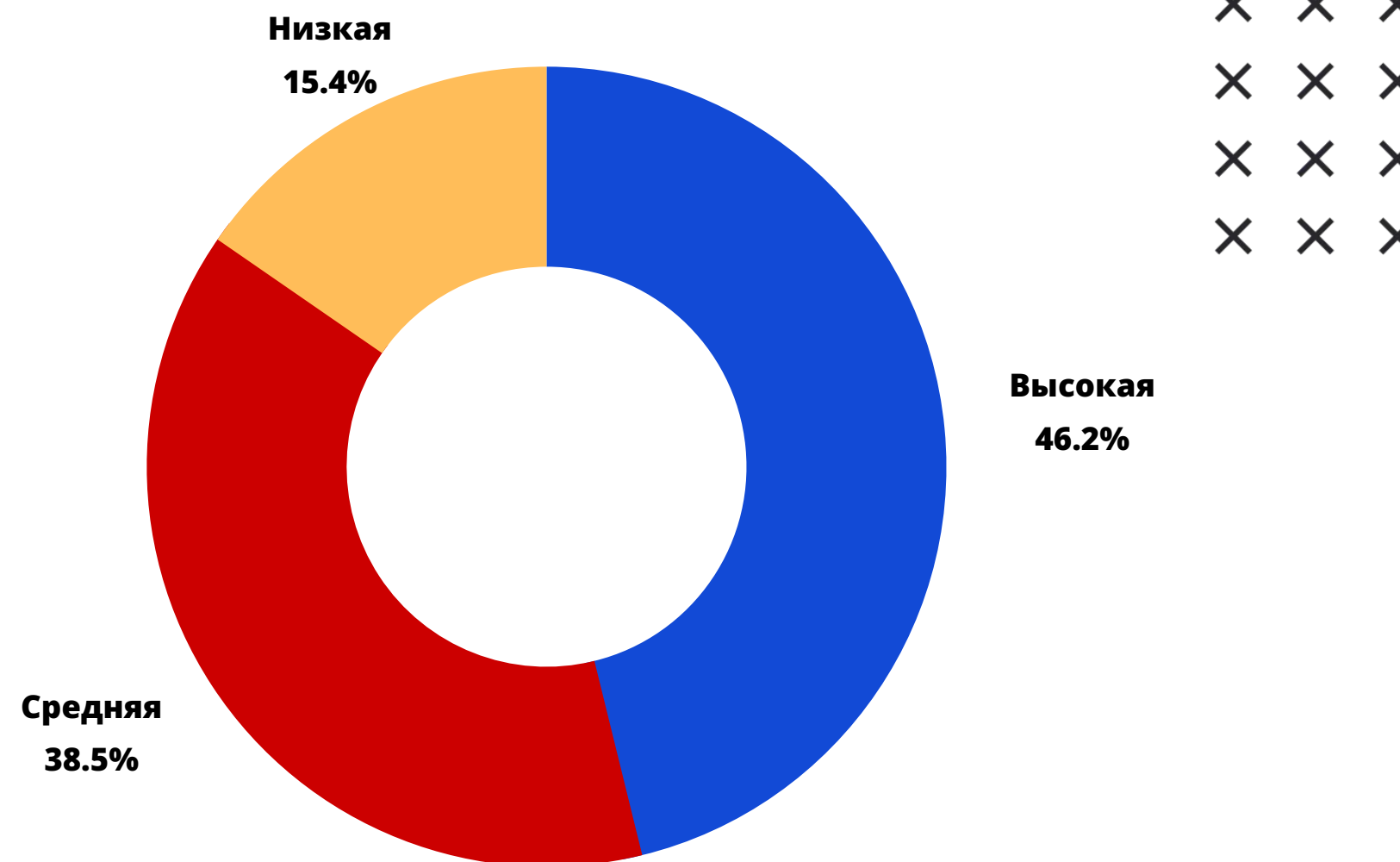


× × ×
× × ×
× × ×
× × ×

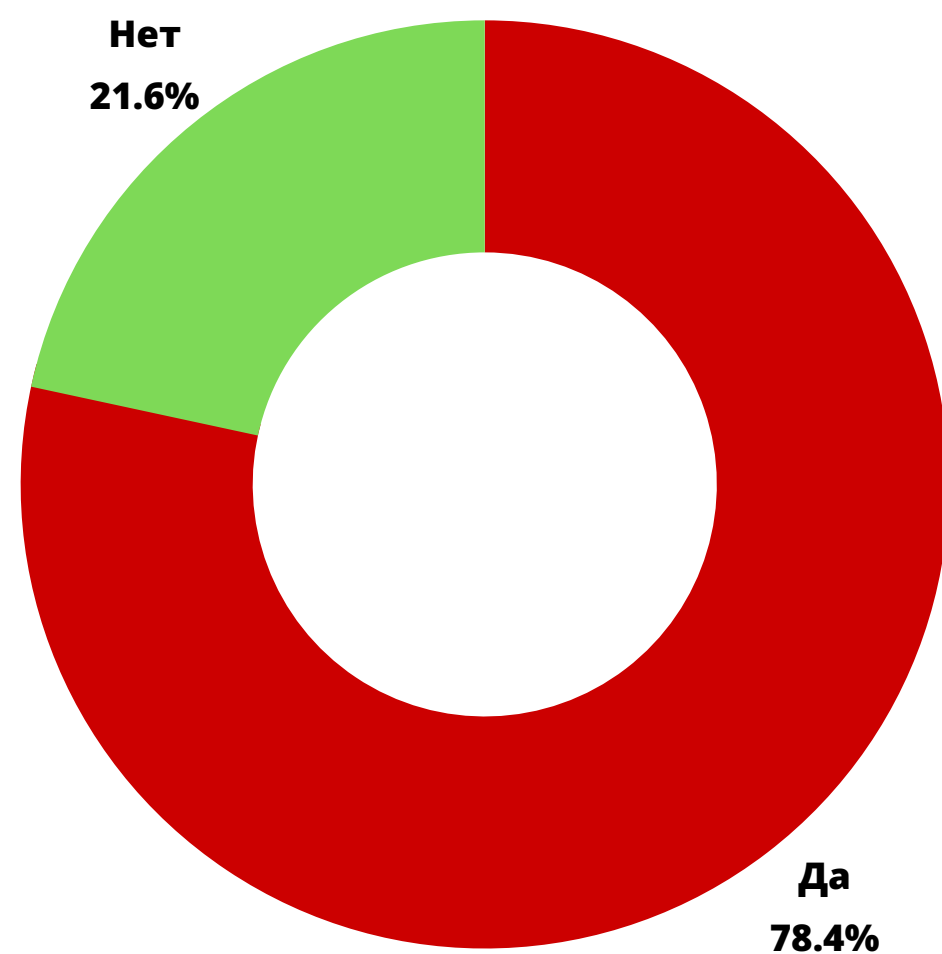
На сколько вы считаете актуальной проблему
защиты персональных данных в России?



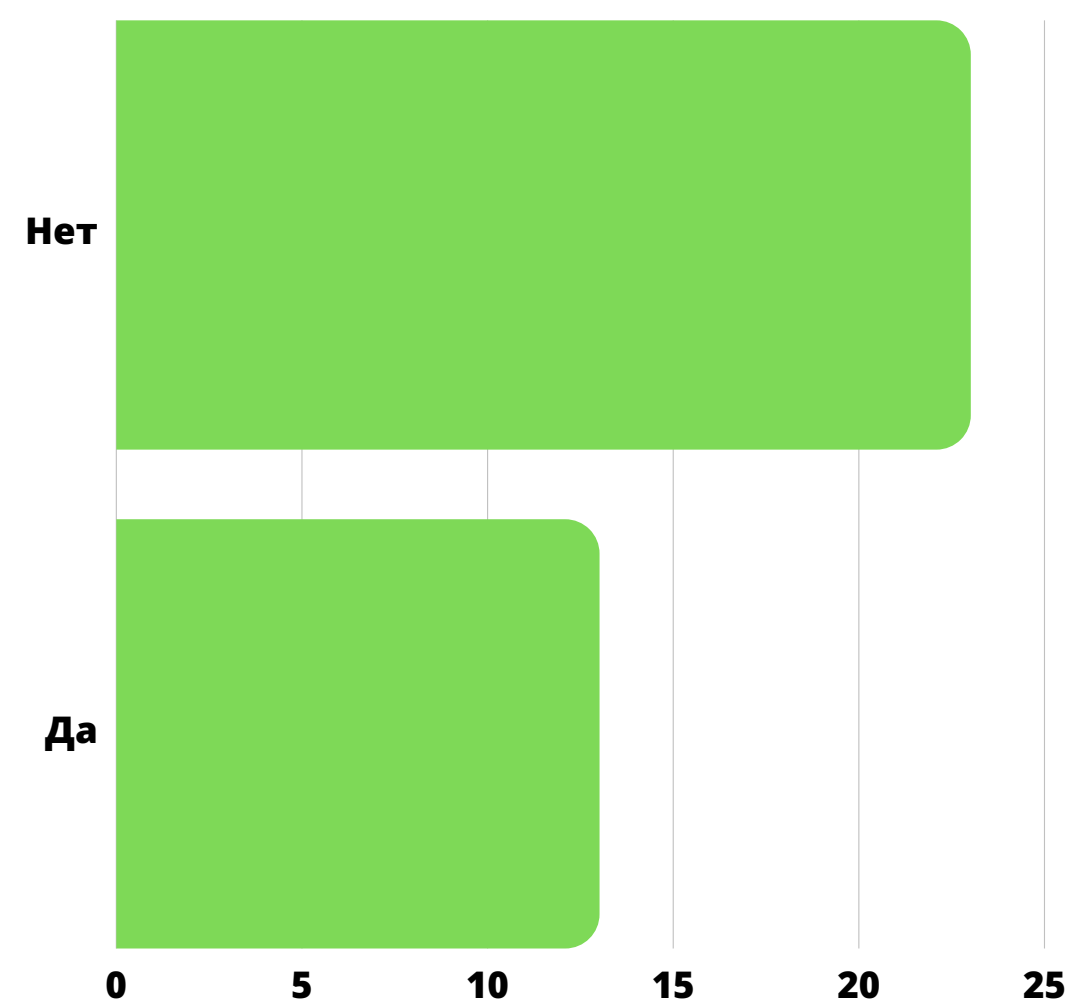
На сколько вы считаете актуальной проблему
защиты персональных данных в компании (в
которой вы работаете)?



Пользуетесь ли Вы средствами защиты персональных данных в компании?

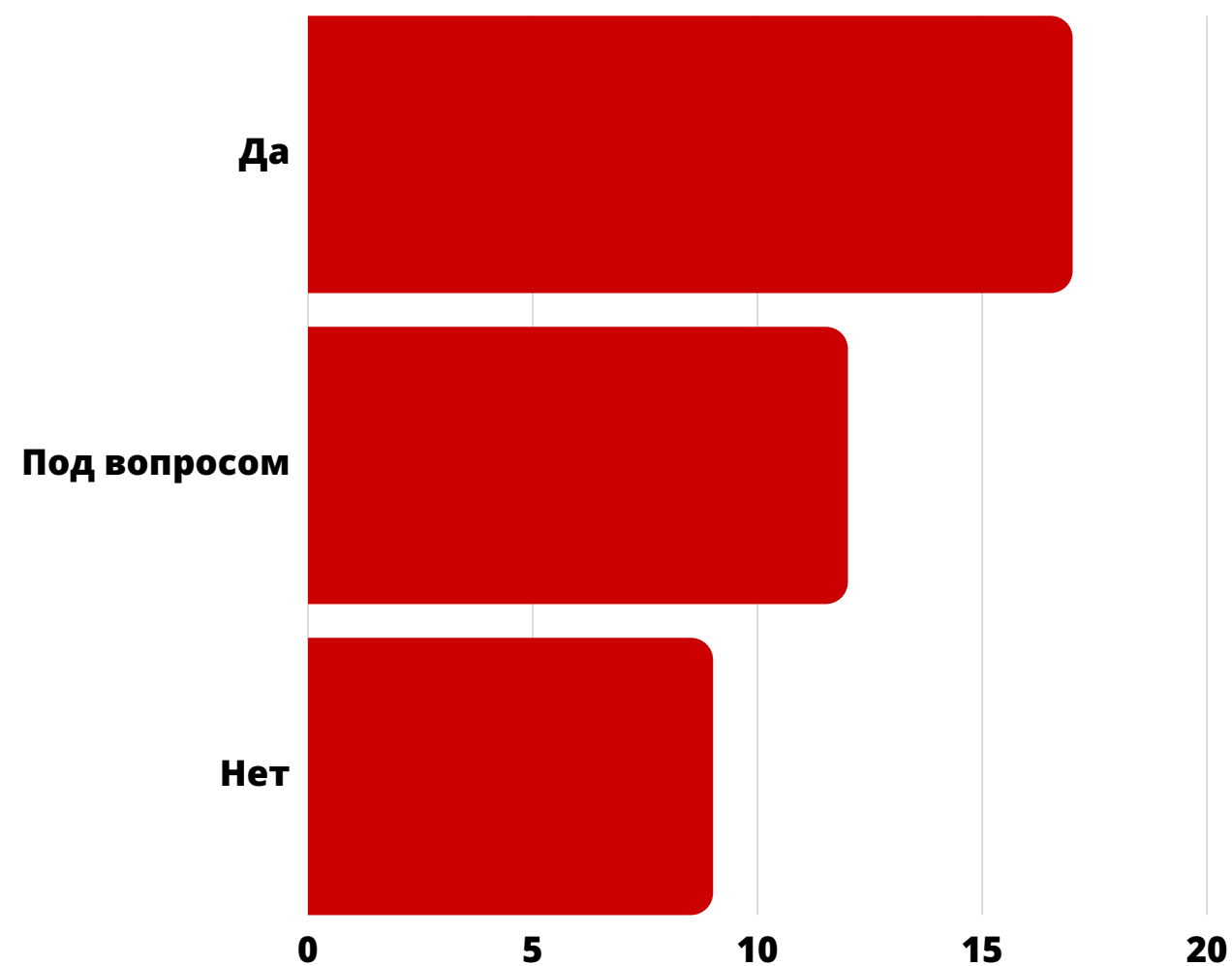


Имеются ли в бюджете компании (в которой Вы работаете) не освоенные деньги на защиту персональных данных на 2023 год?

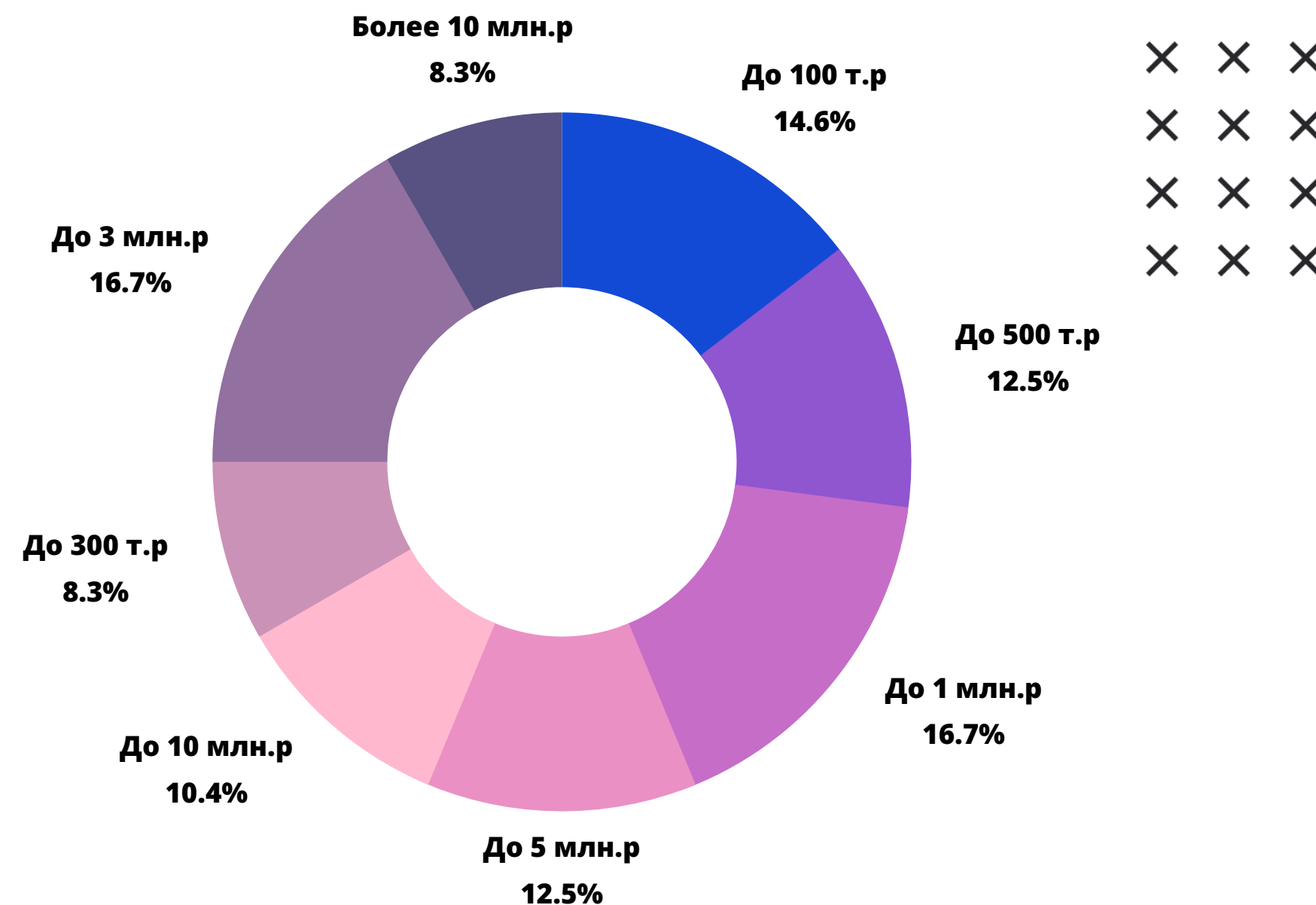


× × ×
× × ×
× × ×
× × ×

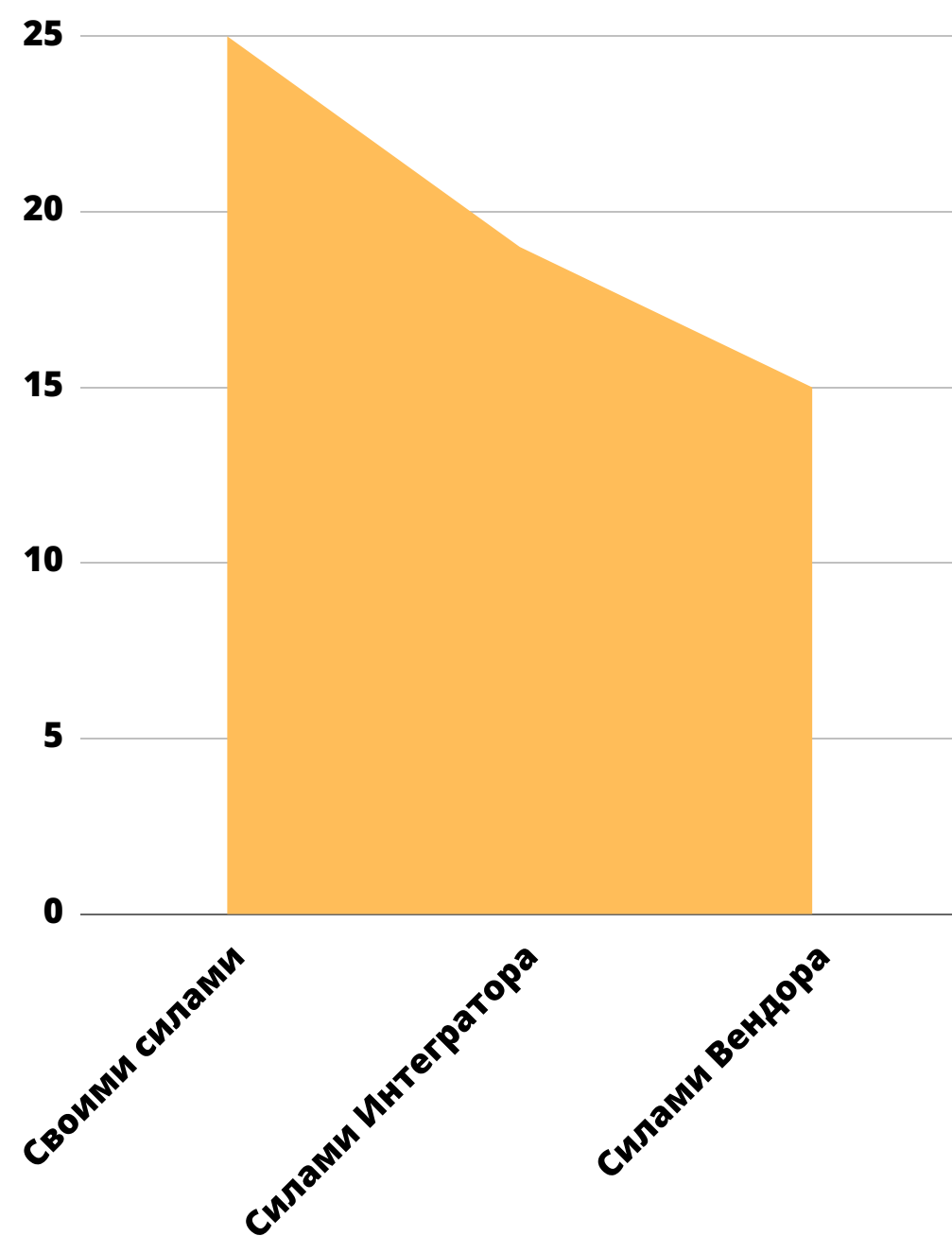
Планируете ли Вы закладывать в бюджет на 2024 год деньги на покупку средств защиты персональных данных?



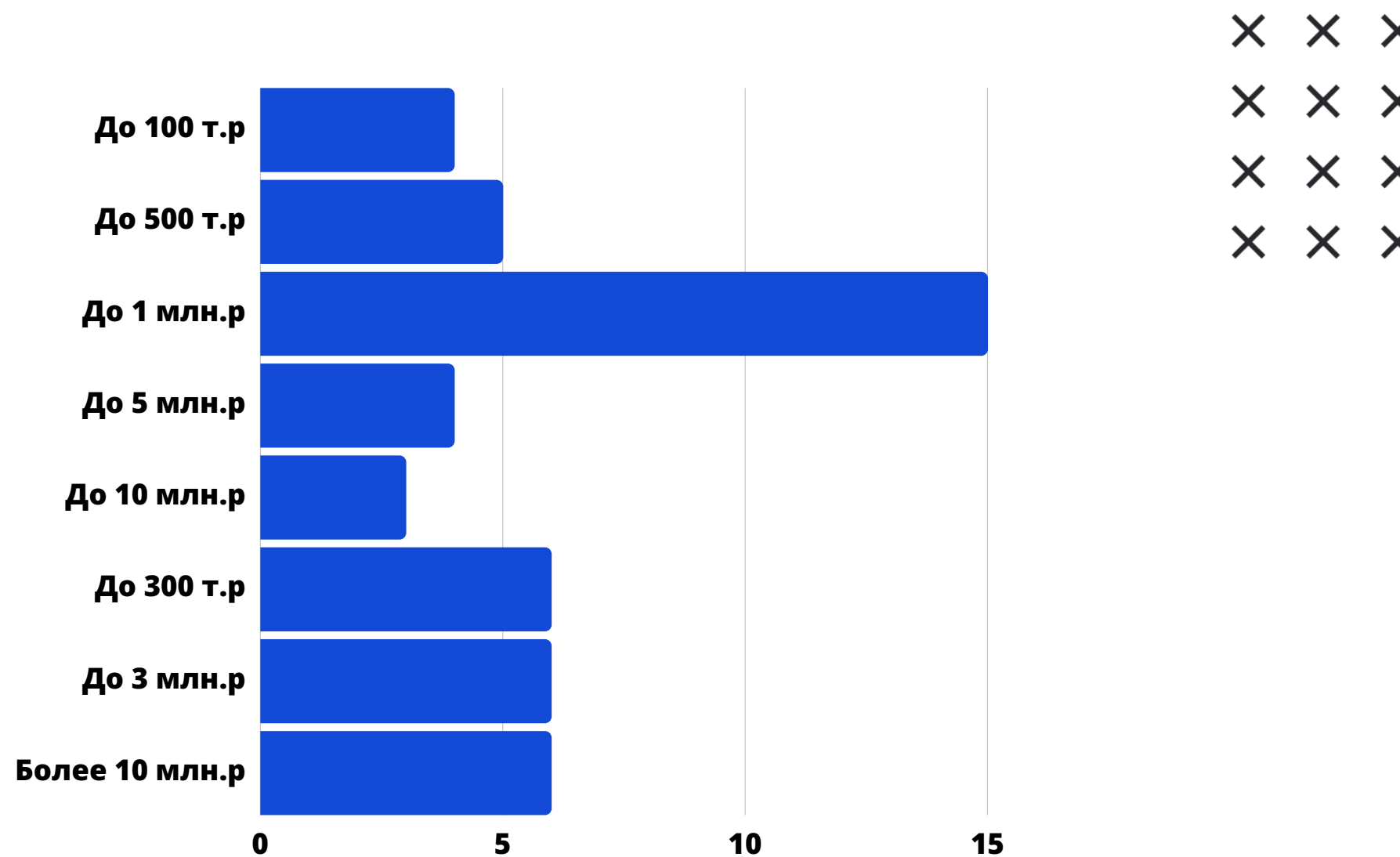
В какую сумму компании (в которой вы работаете) обошлась покупка средств защиты персональных данных?



Как Вы готовы внедрять средства защиты персональных данных в компании (в которой Вы работаете)?

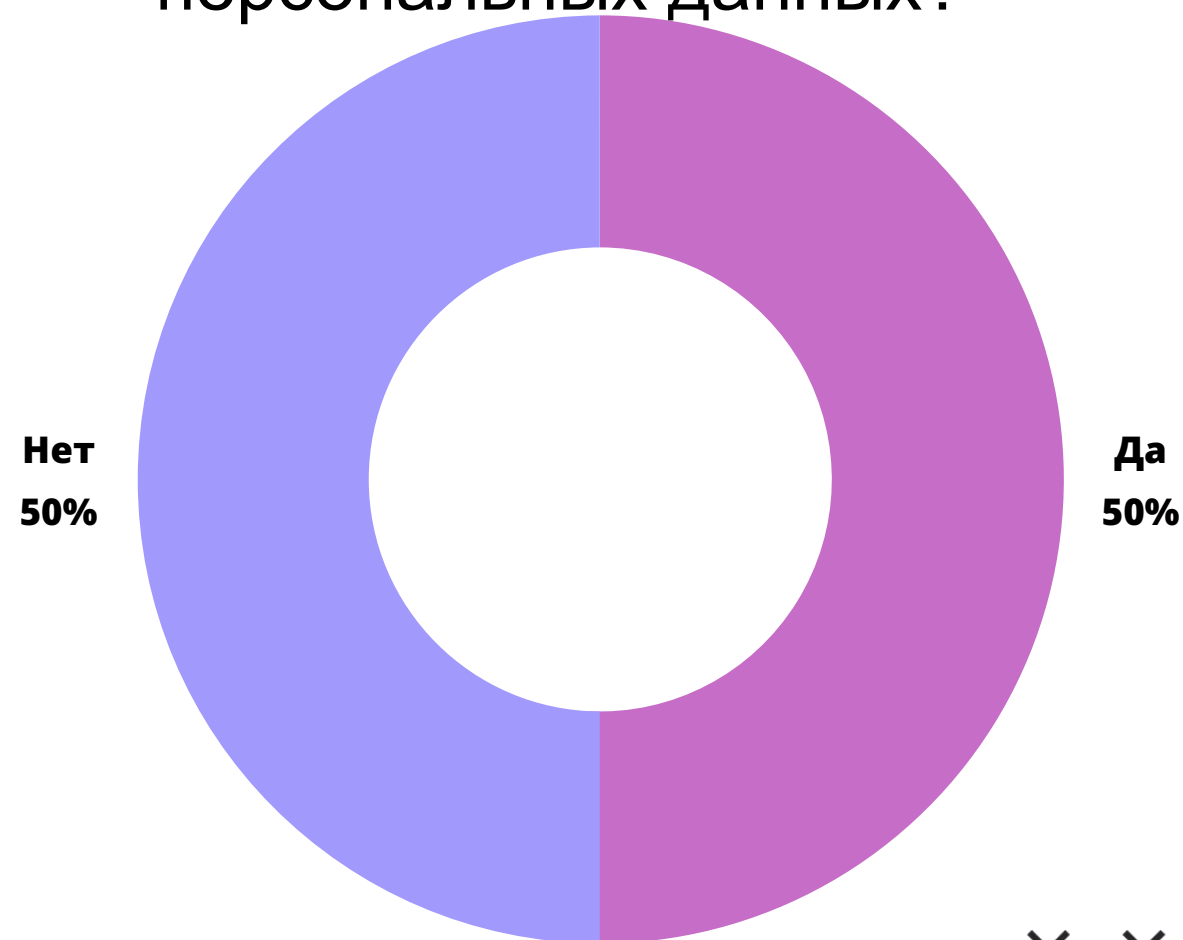


Как Вы считаете, сколько должно стоить решение для обеспечения защиты персональных данных в компании (в которой вы работаете)?



× × ×
× × ×
× × ×
× × ×

Готова ли Ваша компания платить
за оперативную техническую
поддержку купленного Вами
решения для защиты
персональных данных?

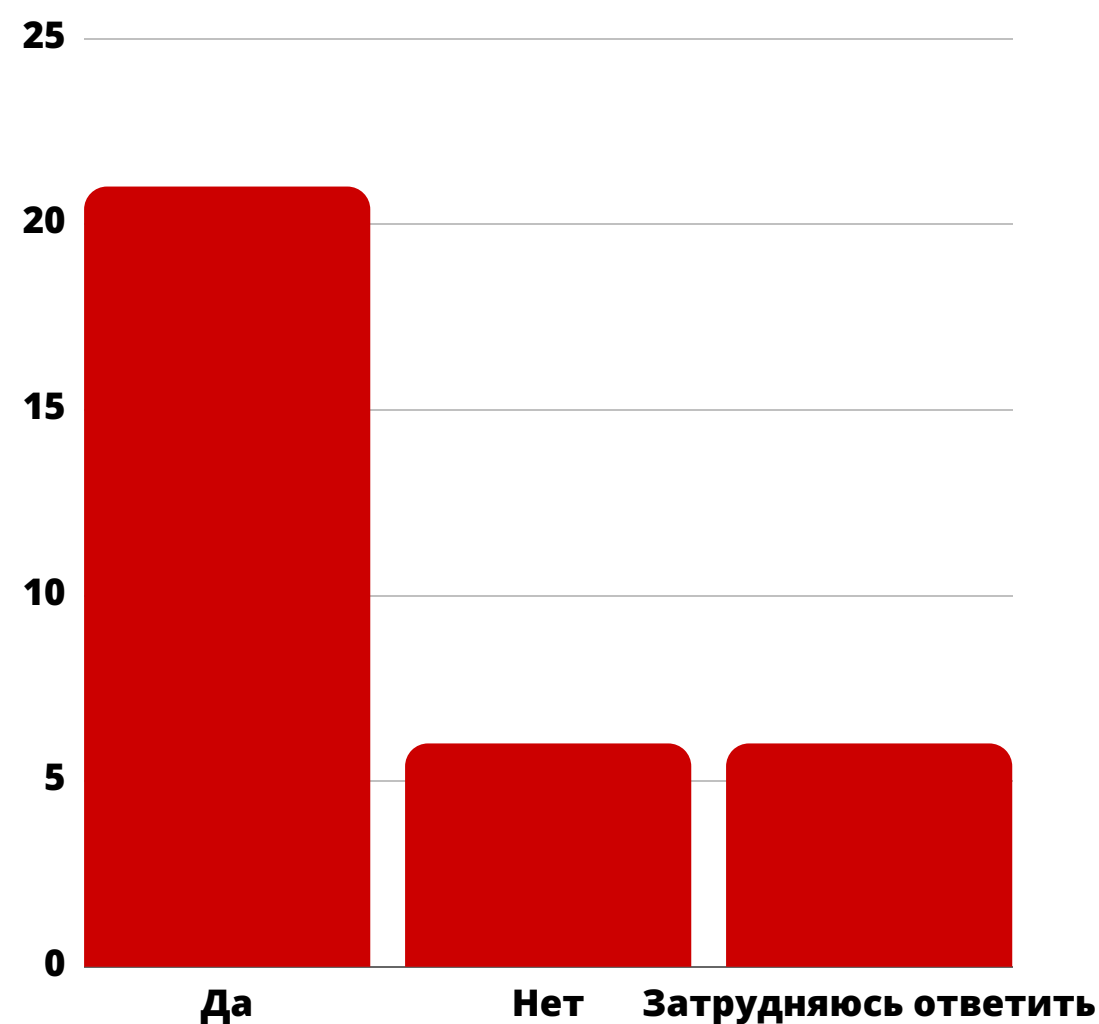


× × ×
× × ×
× × ×
× × ×

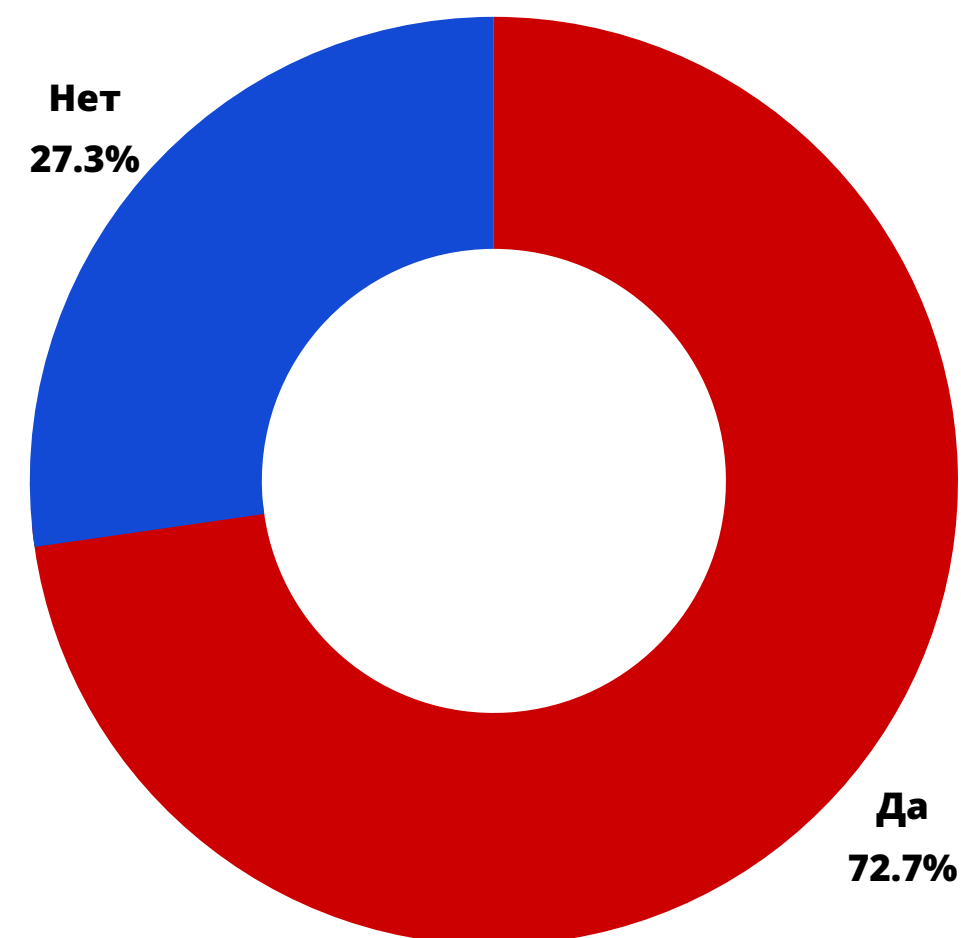
Какие функции средств защиты
персональных данных для Вас важны
в первую очередь?

- | | |
|----------------------------|-----------------------------|
| • Надёжность | • Защита от утечки |
| • Гарантия | • Конфиденциальность данных |
| • Цена | • Простота интеграции |
| • Конфиденциальность | • Отслеживание |
| • Юзабилити | целенаправленных действий с |
| • МСЭ НСД | целью получения ПД |
| • Соответствие требованиям | • Контроль целостности |
| регуляторов | • Контроль доступа |
| • Контроль доступа | • Разграничение доступа |
| • Маскирование | • Защита от утечек |
| • Защита от НСД | • Эффективность защиты |
| • Бесплатность | • Контроль и управление |
| | доступом |
| | • Удобство для проведения |
| | аудита |
| | • Надежность |
| | • Скорость и простота |

Как вы считаете полезен ли функционал по поиску персональных данных в документе и их «заблюриванию» с выдачей исходного документа без персональных данных в нем?



Нужно ли Вашим сотрудникам обучение для работы с системой защиты персональных данных?



Индивидуальное общение
аналитиков с ЛПР из группы:
"Результативный CISO"



Проведение опроса в
специализированном канале,
посвященном защите ПДн:
"Листок бюрократической
защиты информации"



Проведение опроса в
специализированном канале,
посвященном защите ПДн:
"Privacy Advocates"



Kvirin KPDA АВТОМАТИЗИРОВАННЫЙ АЛГОРИТМ ОБЕЗЛИЧИВАНИЯ ДАННЫХ

Протестировать MVP:



x x x
x x x
x x x
x x x

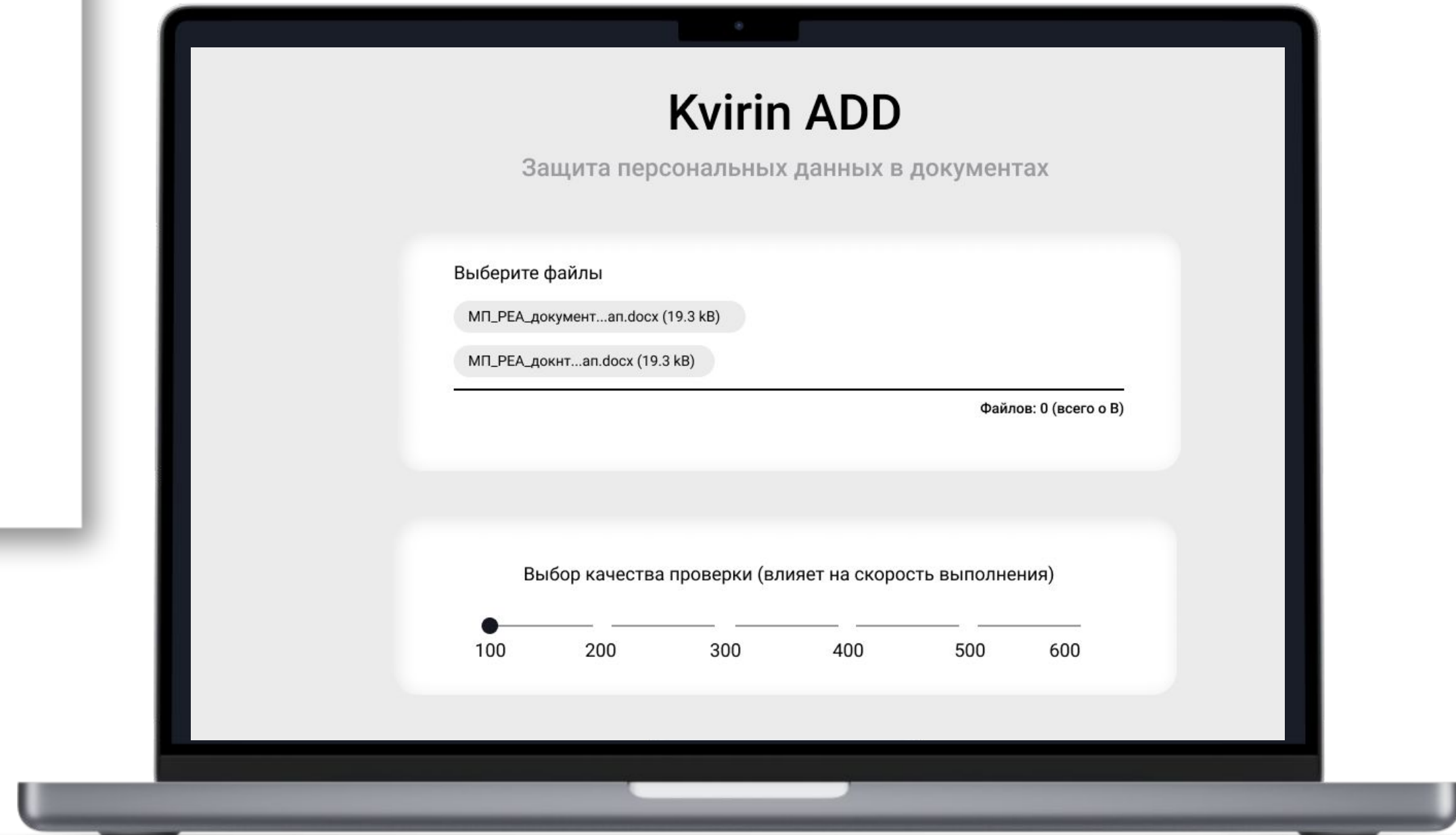
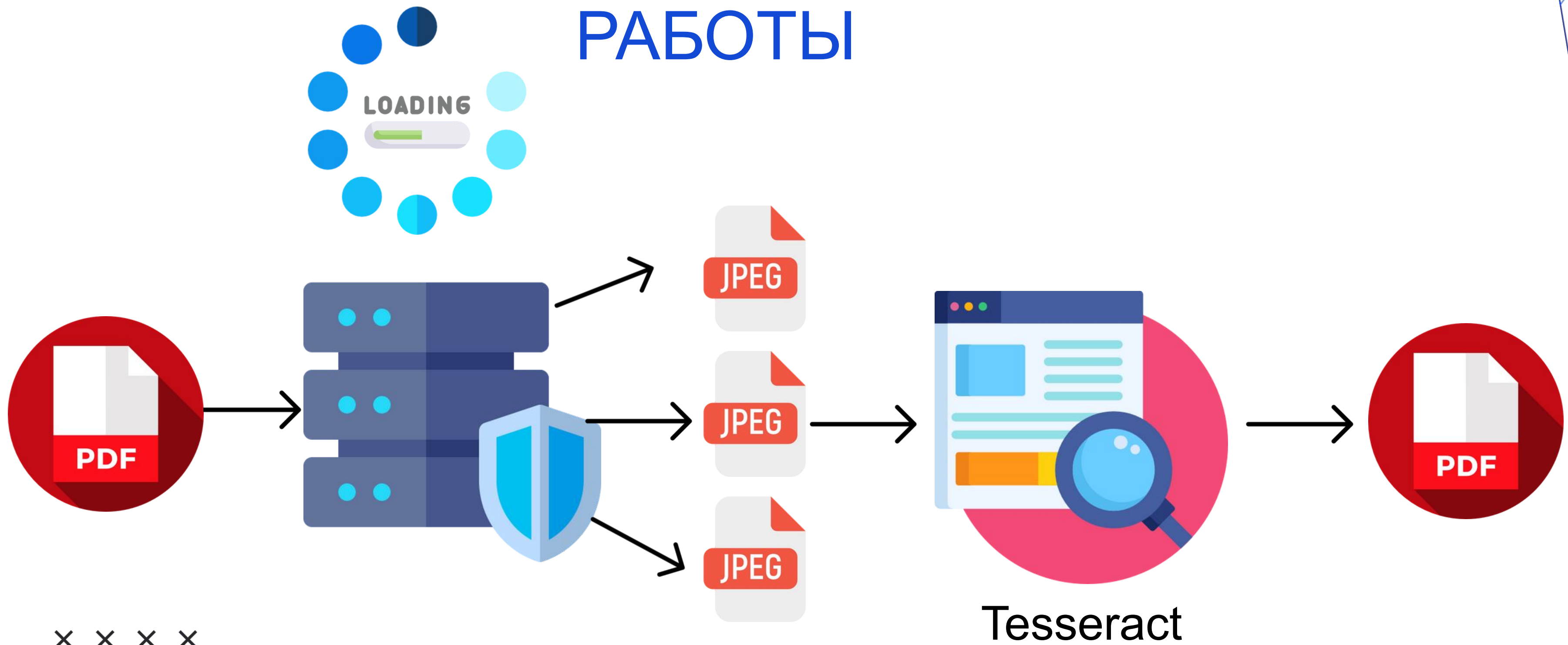


СХЕМА РАБОТЫ



Kvirin

Personal Data Analiser

Достоинства

- Распознавания с помощью Tesseract.
- Для классификатора использовался алгоритм градиентного бустинга xgboost.
- В качестве датасета использовался набор с ФИО в качестве одного класса и орфографический словарь в качестве другого класса.
- Датасет с ФИО был расширен путем добавления некоторых склонений.
- Датасет был приведен в нижний регистр и токенизирован на уровне символов токенизатором из библиотеки Tensorflow.
- Загруженный документ временно хранится только в оперативной памяти. Физически документ не хранится в файловом хранилище.

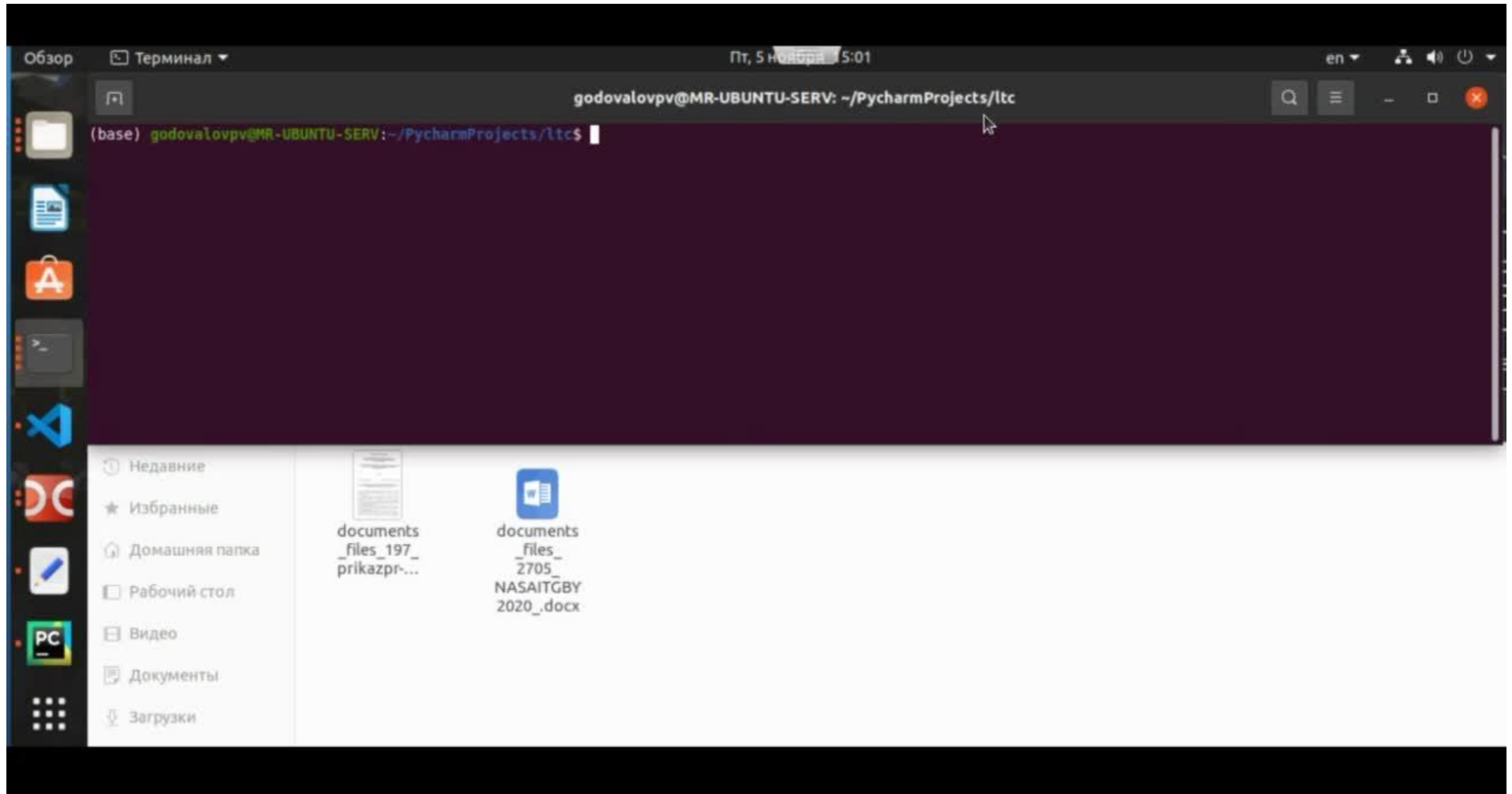
x x x x
x x x x

15



ДЕМОНСТРАЦИЯ РАБОТЫ НА

16



ПРАВИТЕЛЬСТВО МОСКВЫ

ГОСУДАРСТВЕННАЯ ИНСПЕКЦИЯ ПО [REDACTED] ЗА
ИСПОЛЬЗОВАНИЕМ [REDACTED] НЕДВИЖИМОСТИ
ГОРОДА МОСКВЫ

ГОСУДАРСТВЕННОЕ [REDACTED] УЧРЕЖДЕНИЕ ГОРОДА МОСКВЫ
[REDACTED] КОНТРОЛЬНО-МОНИТОРИНГОВЫЙ ЦЕНТР
НЕДВИЖИМОСТИ»

ПРИКАЗ

от 24 [REDACTED] 2020г.

№ [REDACTED] 113/201 ПР- [REDACTED]

«О назначении [REDACTED] уполномоченных на проведение мероприятий по
определению вида фактического использования [REDACTED] (строений, сооружений)
и [REDACTED] помещений [REDACTED] налогообложения»

В соответствии с постановлением [REDACTED]ительства Москвы от 14 [REDACTED] 2014 г.
№ 257-ПП «О [REDACTED] дел [REDACTED] фактического использования [REDACTED]
(строений, сооружений) и [REDACTED] помещений [REDACTED] целей налогообложения»
[REDACTED] в а е м:

1. Утвердить список государственных гражданских служащих Госинспекции
недвижимости ([REDACTED] – государственные служащие) и [REDACTED]
Государственного бюджетного учреждения города Москвы [REDACTED]
контрольно-мониторинговый центр недвижимости» [REDACTED] – [REDACTED]
уполномоченных на проведение мероприятий [REDACTED] определению вида фактического
использования зданий (строений, сооружений) и [REDACTED] [REDACTED] [REDACTED]
налогообложения (приложение).

2. [REDACTED] структурных подразделений довести настоящий приказ
[REDACTED] сведения государственных служащих Госинспекции по недвижимости и
[REDACTED] [REDACTED] [REDACTED] в приложении к [REDACTED] под роспись и
обеспечить его выполнение.

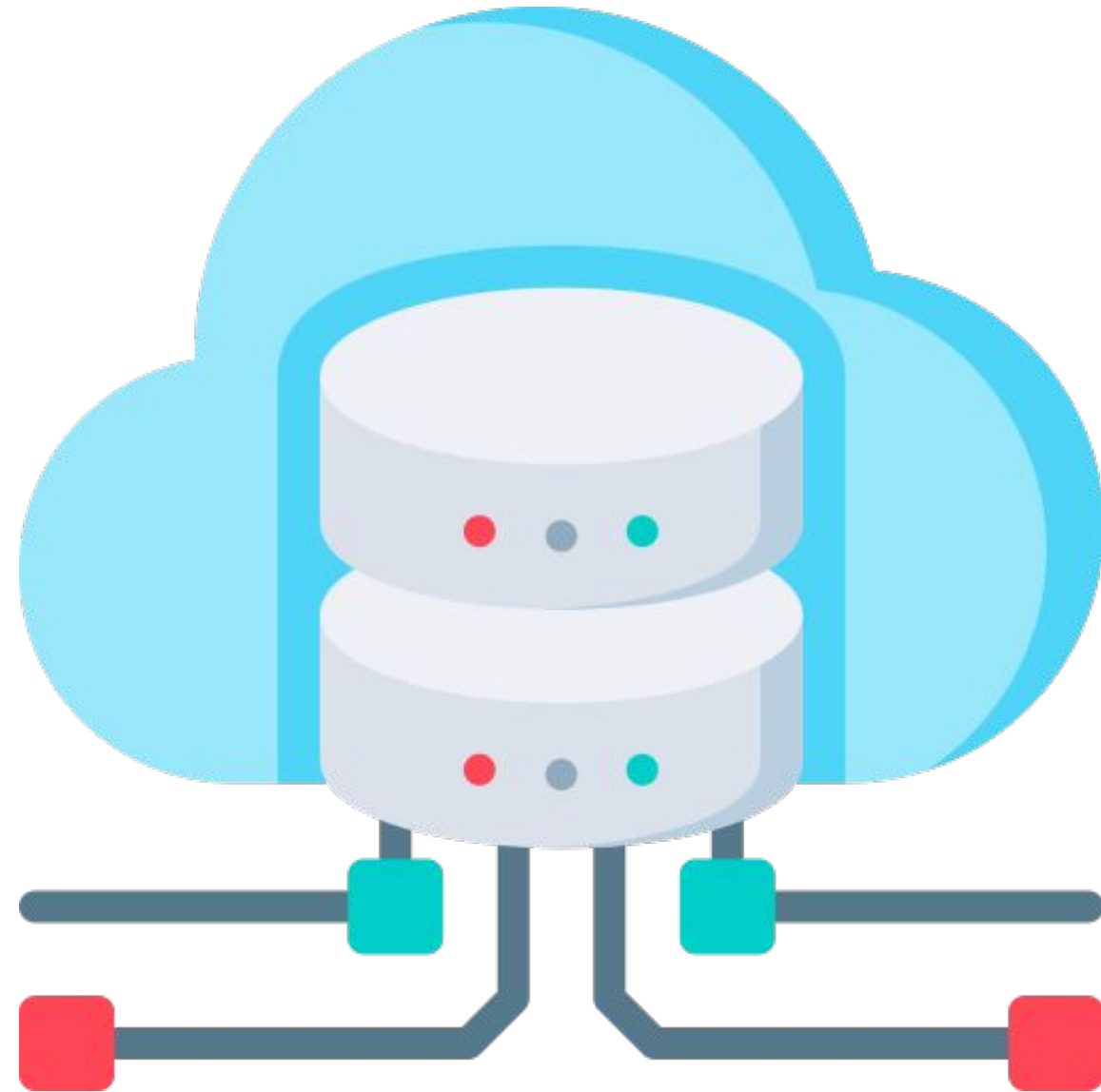
3. Управлению организационного обеспечения Госинспекции по
недвижимости осуществить размещение [REDACTED] настоящего приказа с приложением
в [REDACTED] «Результаты инспекционных мероприятий по [REDACTED] делению
фактического использования» [REDACTED] а «Документы» [REDACTED] интернет-странице

```
[127] print(classification_report(Y_test, predict))
```

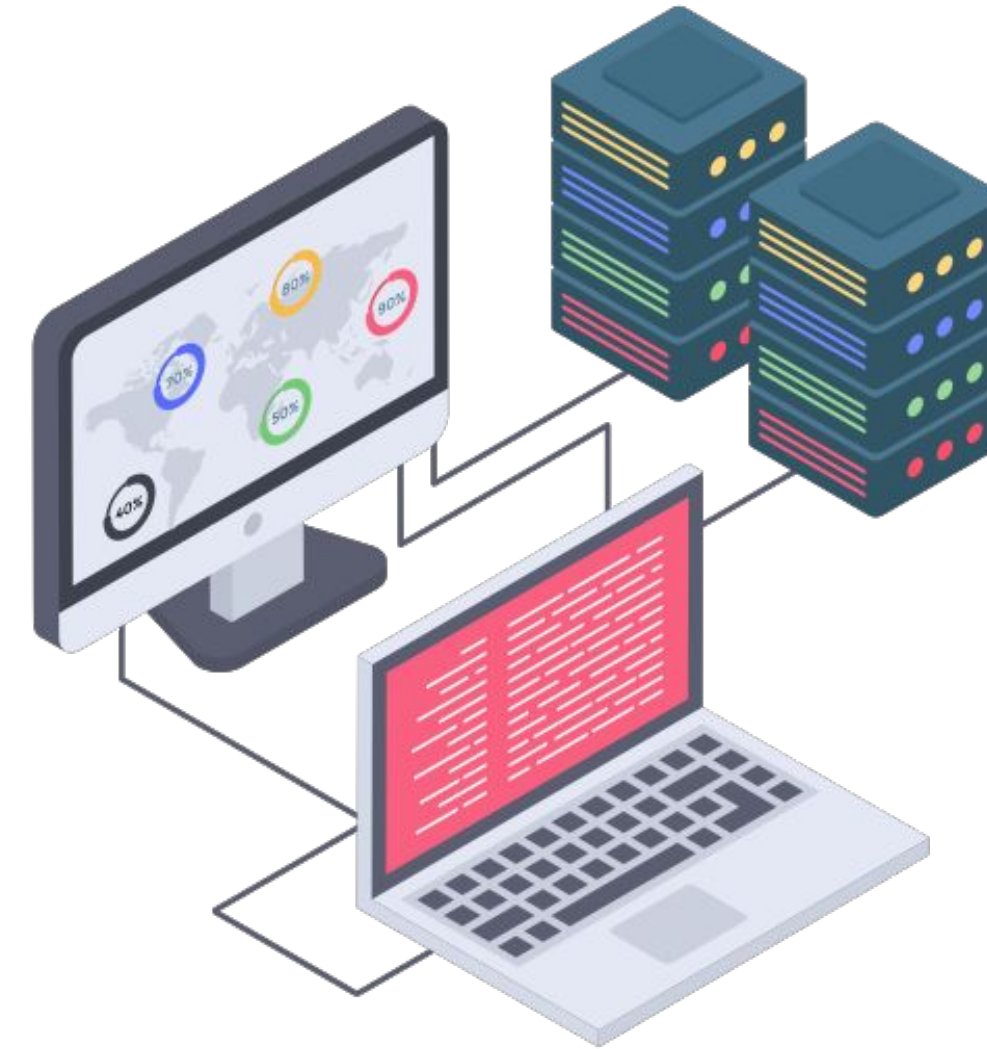
	precision	recall	f1-score	support
0	0.94	0.90	0.92	4562
1	0.98	0.99	0.98	19358
accuracy			0.97	23920
macro avg	0.96	0.94	0.95	23920
weighted avg	0.97	0.97	0.97	23920

1-й класс с ФИО, с точностью
классификации в 98%, полнотой 99% и
средней гармонической score f1 98%,
что является достаточно высоким
показателем

Схема внедрения



Облако



Локальный
Сервер

× × ×
× × ×
× × × ×
× × × ×



Требования к функционированию

Рекомендуемые системные требования:

Сервер
обработки
16 Core
RAM: 16 GB
HDD: 150 GB



Сервер
обработки
16 Core
RAM: 16 GB
HDD: 150 GB

API-Server
4 Core
RAM: 8 GB
HDD: 50

Минимальные системные требования:

All-in-one
8 Core
RAM: 8 GB
HDD: 150
GB



Ubuntu 22.0

docker - 20.10.17, build 100c701

docker-compose version - 1.29.2 и выше

× × × ×
× × × ×
× × × ×
× × × ×



ЮРИДИЧЕСКОЕ ОБОСНОВАНИЕ

- 152 ФЗ. Статья 19. Меры по обеспечению безопасности персональных данных при их обработке;
- Приказ РКН № 996;
- 21 Приказ ФСТЭК.



ДОРОЖНАЯ КАРТА



Интегрировать детектирование
других категорий ПДн (30.07.2023)

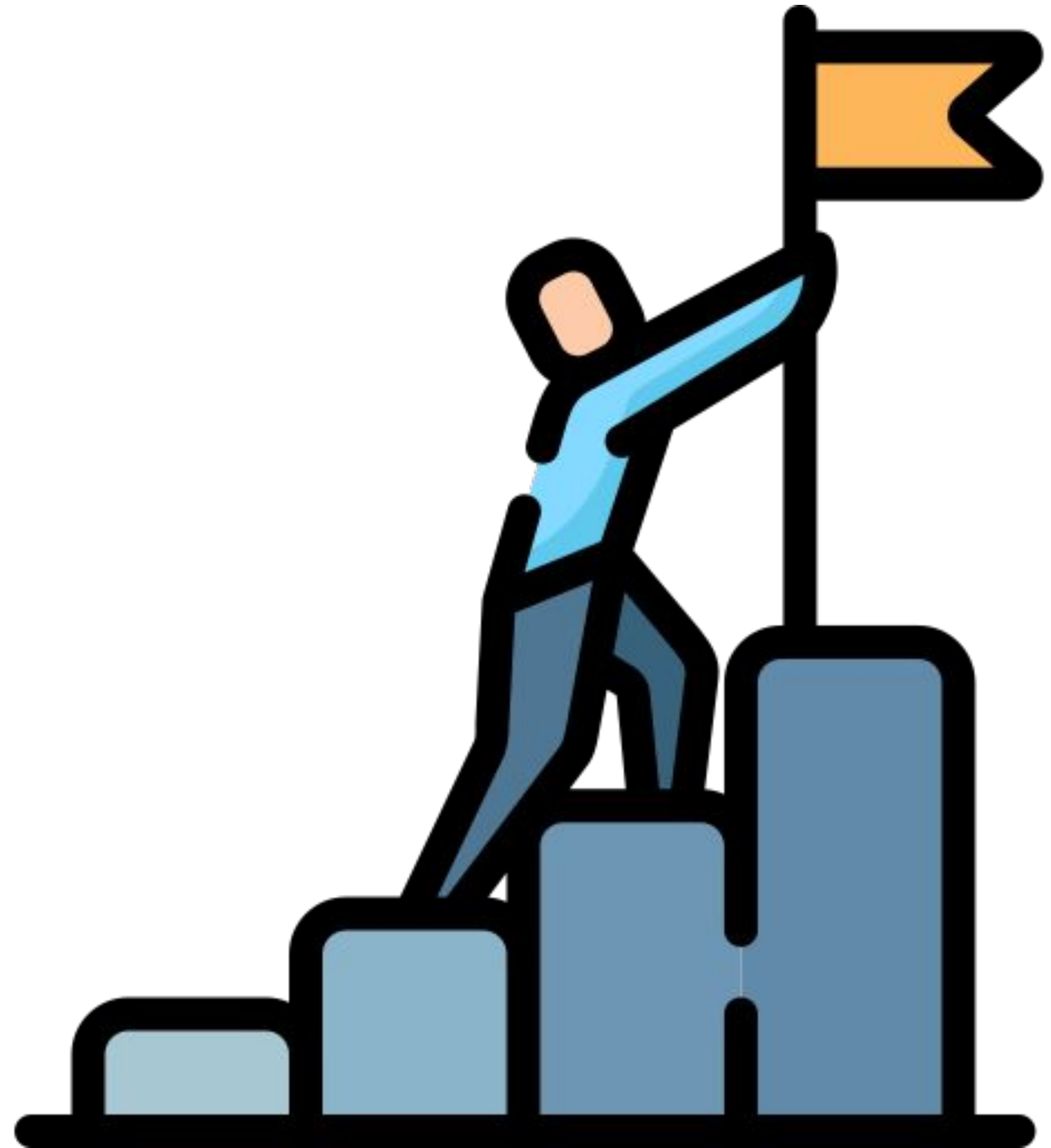
Сделать мультizaгрузку
файлов для анализа (30.08.2023)

Тесты и поиск багов (30.09.2023)

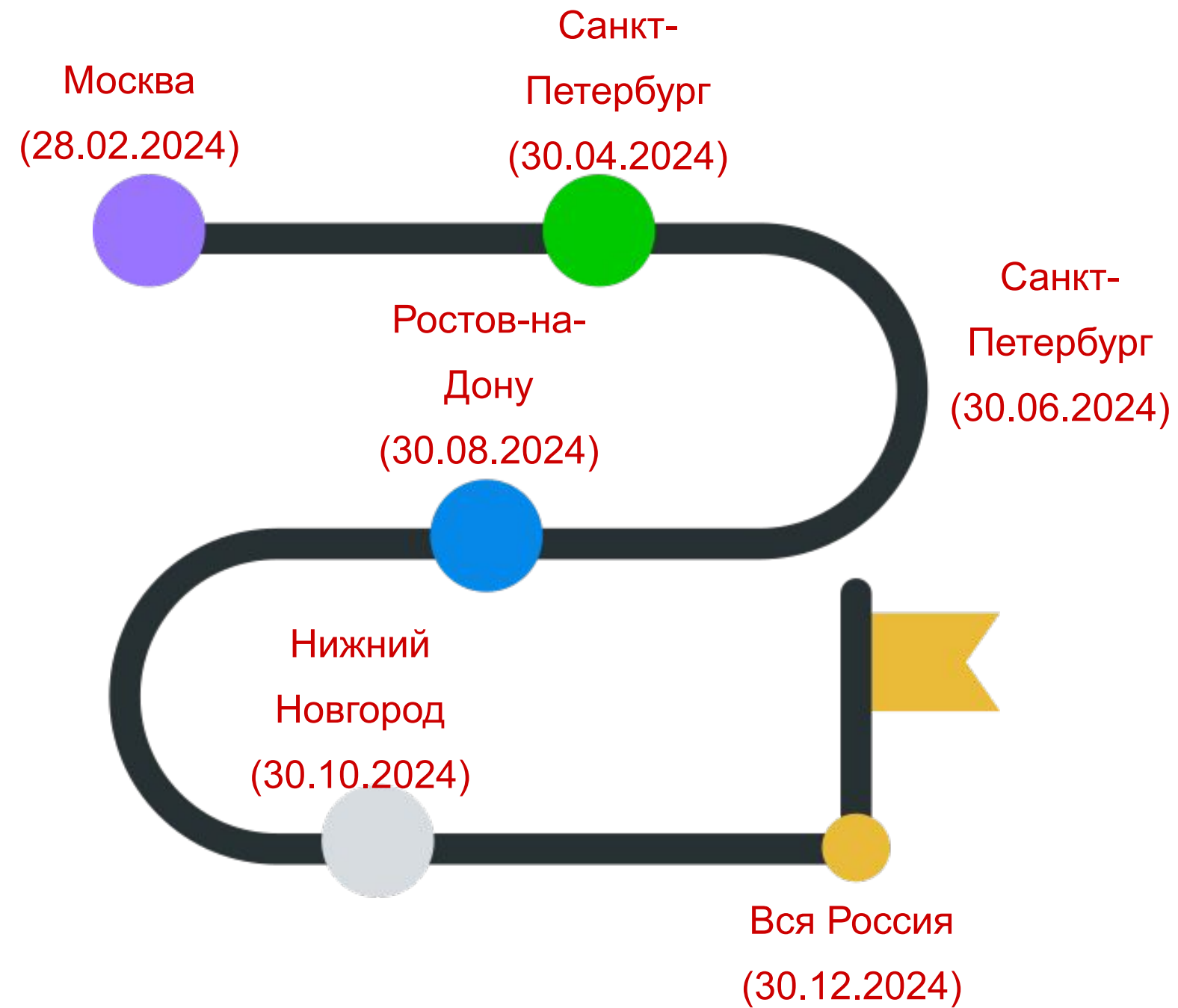
Доработка продукта (30.10.2023)

Пилот у Заказчика (30.11.2023)

Первая продажа (28.02.2024)



ПЛАН ВЫХОДА НА РЫНОК



БЮДЖЕТ ДОРОЖНОЙ

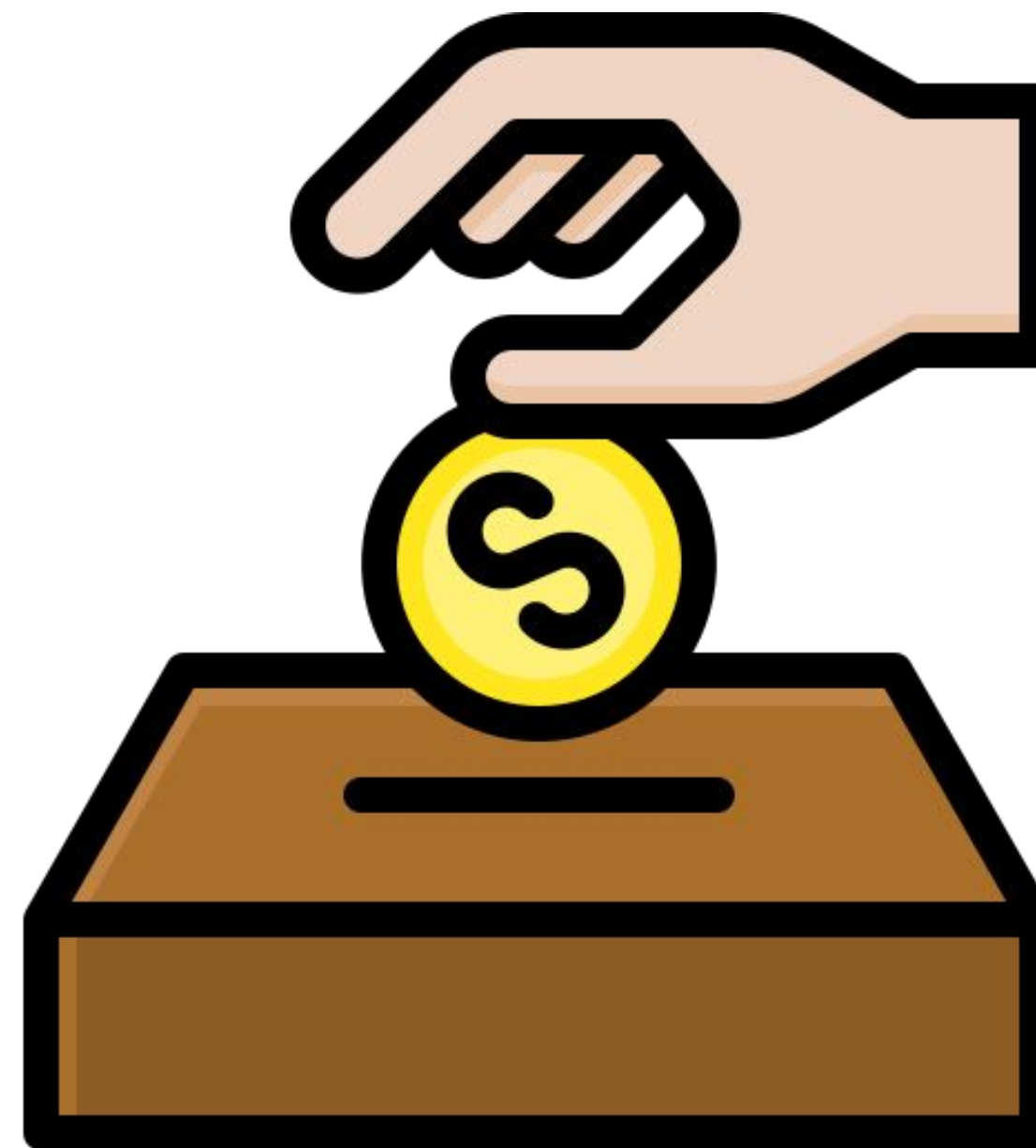
КАРТЫ

- ФОТ -676 т.р/м
- Затраты на серверные ресурсы
-150 т.р/г
- Затраты на защиту сервера - 50

т.р/г

x x x
x x x
x x x
x x x

Итого: 5 608 000 рублей



МОНЕТИЗАЦИЯ

ТАРИФ - ЛАЙТ

До 50 загрузок в год

Цена: 100.000 р/в год



ТАРИФ -

До 500 загрузок в год

Цена: 1.000.000 р/в год



ТАРИФ -

До 1.000 загрузок в год

Цена: 3.000.000 р/в год



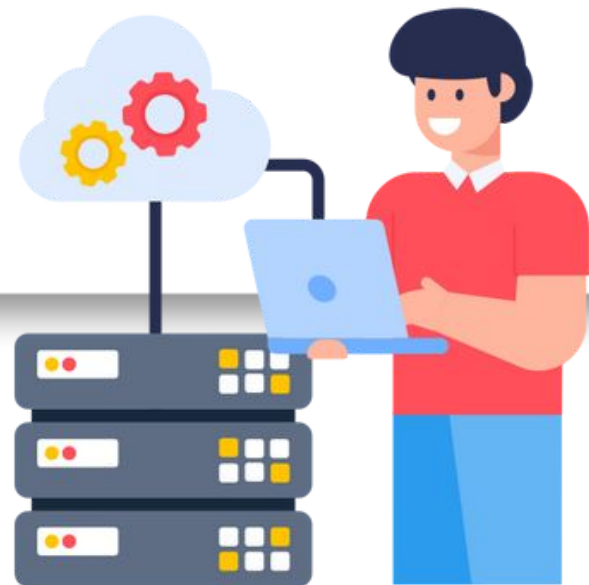
МОНЕТИЗАЦИ Я

ВНЕДРЕНИ

Тариф - Лайт: 100.000 т.р

Тариф - ПРО: 300.000 т.р

Тариф - Премиум: 500.000 т.р



SUPPOR

Тариф - Лайт: 100.000 т.р

(SLA=3 дня)/20 о.г

Тариф - ПРО: 300.000 т.р

(SLA=8 часов)/50 о.г

Тариф - Премиум: 500.000 т.р

(SLA=3 часа)/100 о.г



SUPPORT

ОБУЧЕНИ

50.000 т.р - 1 человек





Стэк технологий

- Python;
- Vue;
- JavaScript;
- Ubuntu 22.04;
- Docker.



НАША КОМАНДА

Беляев Дмитрий

Роль
Product Owner
Кибербезопасность
Вид сотрудничества:
Подряд/ГПХ
Выходные и Праздники



Сафин Илья

Роль
Project Manager
Frontend
Вид сотрудничества:
Подряд/ГПХ
Выходные и Праздники



Годовалов Павел

Роль
Data Science
Вид сотрудничества:
Подряд/ГПХ
Выходные и Праздники



Аббакумов Валерий

Роль
Fullstack
DevOps
Вид сотрудничества:
Подряд/ГПХ
Выходные и Праздники



Работько Анастасия

Роль
UX/UI Дизайнер
Вид сотрудничества:
Подряд/ГПХ
Выходные и Праздники



Шкурат Кристина

Роль
Аналитик
Вид сотрудничества:
Подряд/ГПХ
Выходные и Праздники



Белугина Юлия

Роль
Юрист
Вид сотрудничества:
Подряд/ГПХ
Выходные и Праздники



Стабуль Тимофей

Роль
Аналитик
Вид сотрудничества:
Подряд/ГПХ
Выходные и Праздники



ГОТОВА К