



# Пилотный проект создания оператора биомедицинских данных

---

Ассоциация «Национальная база медицинских знаний» совместно с АНО «Врачебная палата ФМБА России»

# Этапность создания Оператора

Этап №1.  
Пилотный проект

Этап №2.  
Пробная  
эксплуатация

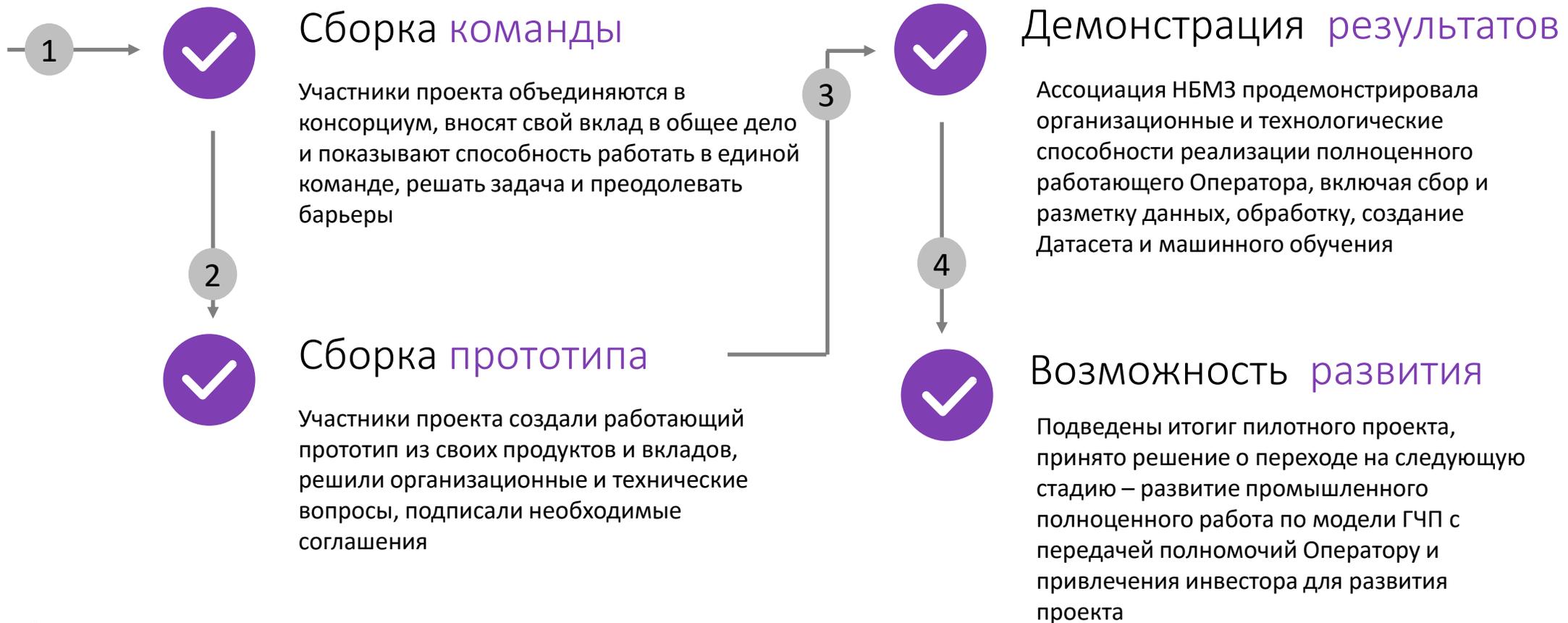
Этап №3.  
Промышленная  
стадия

Создается на основе безвозмездного вклада участников, согласившихся самостоятельно создать прототип Оператора

В случае успешного создания прототипа производится его пилотная эксплуатация и развитие. Для этого привлекаются средства Инвестора

Оператор работает за счет зарабатываемых средств, в том числе от коммерческих заказчиков и государственных контрактов

# Задачи проведения пилотного проекта



# Основные принципы пилотного проекта



Реализация без бюджетного финансирования



Моделирование основных процессов работы



Измеримые результаты и демонстрация ценности для участников рынка



Короткие сроки пилота – до 6 месяцев

# Участники проекта



## Врачебная палата ФМБА

Подключение медицинских организаций, сбор и обработка биомедицинских данных



## SOL STRIX

Предоставление инфраструктуры (ЦОДа) и общесистемного ПО



## Инсофт

Интеграционная шина, сбор данных от МИС МО



## К-МИС

Система Webiomed для сбора ЭМК и производства DataSet



## Robomed

Сервис для учета собранных биомедицинских данных



## MedAboutMe

Медицинский портал о медицине, здоровье и красоте



## CoBrain Analytics

Платформа для сбора и обработки медицинских данных и машинного обучения

# Основные принципы платформы оператора биомедицинских данных



## Интеграция

Инструменты для эффективной интеграции с медицинскими информационными системами



## Деперсонафикация

Удаление персональных данных из состава передаваемой информации



## Готовность к ML

Данные должны передаваться и храниться в структуре и форматах, пригодных для их последующей подготовки и очистки, разметки и производства Датасетов для машинного обучения



## Создание Датасетов

Создание датасетов и доступ к ним должны контролироваться ПО оператора для недопущения выкачивания информации и ее неконтролируемого использования





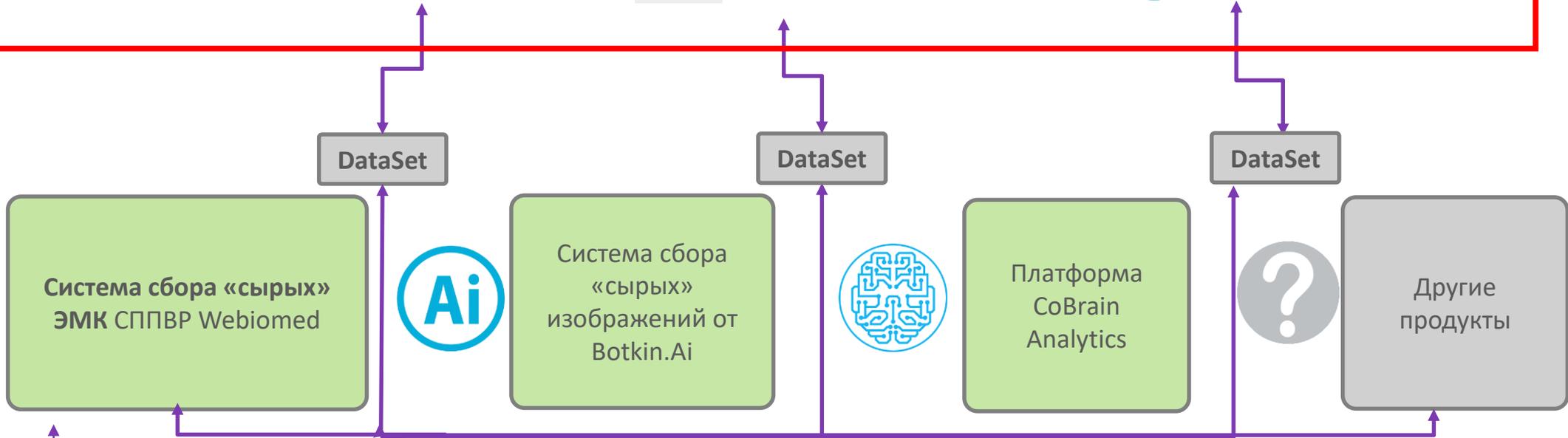
# Архитектура

Обученный ИИ / СППВР

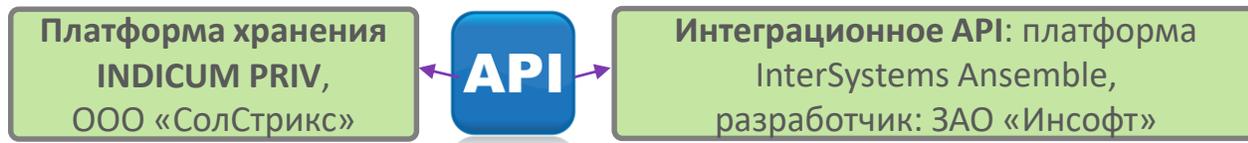
Результаты исследований

Прогнозы / аналитика

Программная платформа Оператора



Система учета биомедицинских данных и их децентрализованного хранения



Источники данных



Медицинские изображения из пилотных МО



«Медеора» МО ФМБА

Медицинские данные из MedAbout медицина обо мне

# Отработан сбор и анализ диагностических изображений

Была проведена интеграция с PACS (архивами изображений) в 4 субъектах РФ

Получено 1279 исследований компьютерной томографии

По результатам анализа изображений с использованием технологий искусственного интеллекта было выявлено 15 человек с подозрением на рак легких

**Результат.** Создана и протестирована технология сбора и обработки диагностических изображений для проекта Оператор биомедицинских данных

oncure.botkin.ai demo2 ^

Список запросов ↻

Общее количество запросов: 17

ID	Карта пациента	Статус патологии	Дата исследования	Модальность	Риск	Статус
3586667	2569	●	22-02-2019	СТ	0.9987	Завершен
3583810	1768	●	22-02-2019	СТ	0.9998	Завершен
3586663	1744	●	22-02-2019	СТ	0.9999	Завершен
3583806	2239	●	22-02-2019	СТ	0.9997	Завершен
3583796	2537	●	22-02-2019	СТ	0.9997	Завершен
3583804	1900	●	22-02-2019	СТ	0.9983	Завершен
3583798	1957	●	22-02-2019	СТ	0.9999	Завершен
3582660	3038	●	22-02-2019	СТ	0.9998	Завершен
3583812	1774_12	●	22-02-2019	СТ	0.9999	Завершен
3583800	1607	●	22-02-2019	СТ	0.9998	Завершен
3583802	2335_95	●	22-02-2019	СТ	0.9958	Завершен
3583808	1369	●	22-02-2019	СТ	0.9998	Завершен
3586665	1494	●	22-02-2019	СТ	0.9999	Завершен
3586657	3002	●	22-02-2019	СТ	0.9999	Завершен
3583814	2547	●	22-02-2019	СТ	0.9998	Завершен
3586659	2795	●	22-02-2019	СТ	0.9999	Завершен
3586661	2835	●	22-02-2019	СТ	0.9999	Завершен

Фильтры

Пациент

Модальность

Начало диапазона

Конец диапазона

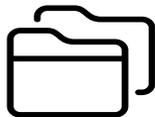
Статус

# Собранны «сырые» данные из ЭМК

## Данные МО ФМБА:



**298** тыс. пациентов



**9** тыс. случаев лечения

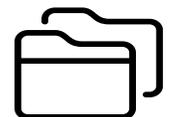


**58** тыс. СЭМДов

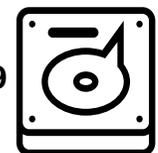
## Данные «Открытой клиники»:



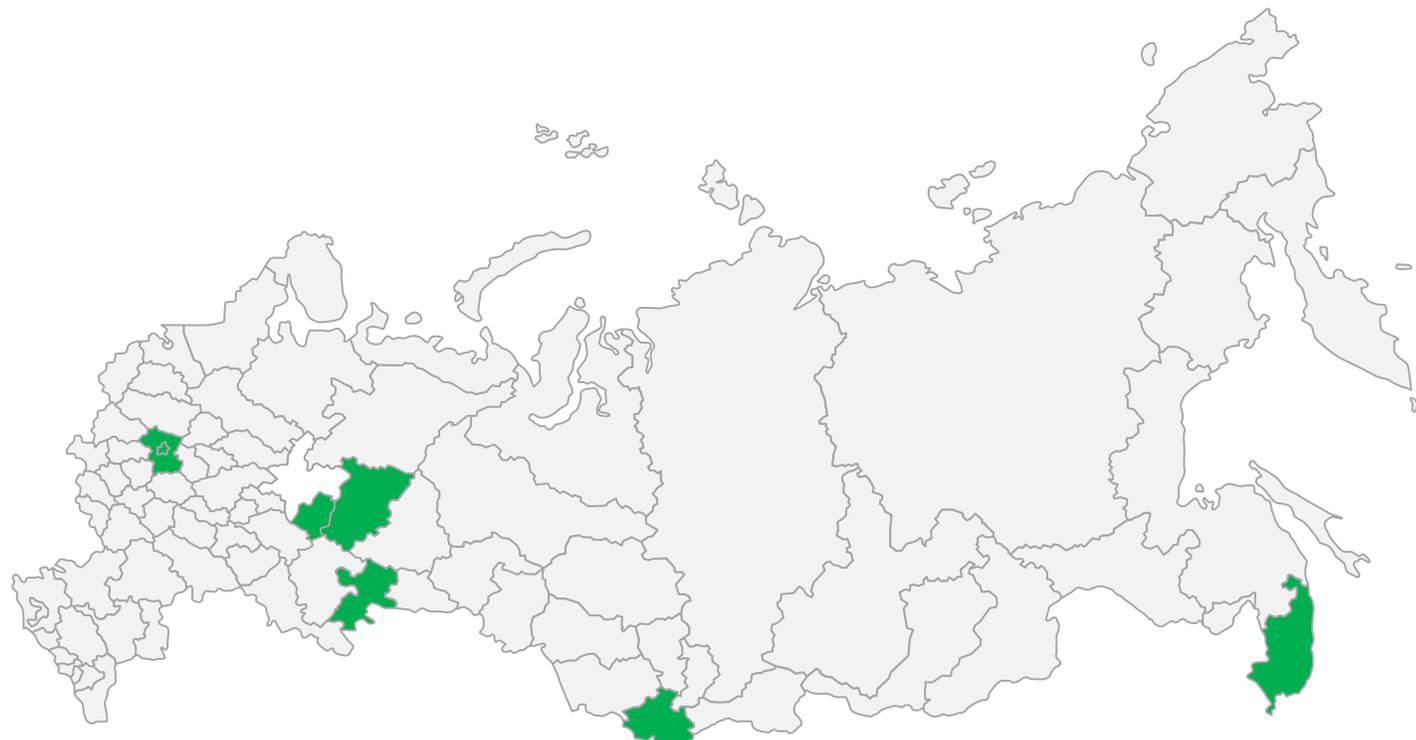
**864** тыс. пациентов



**12,5** млн. протоколов



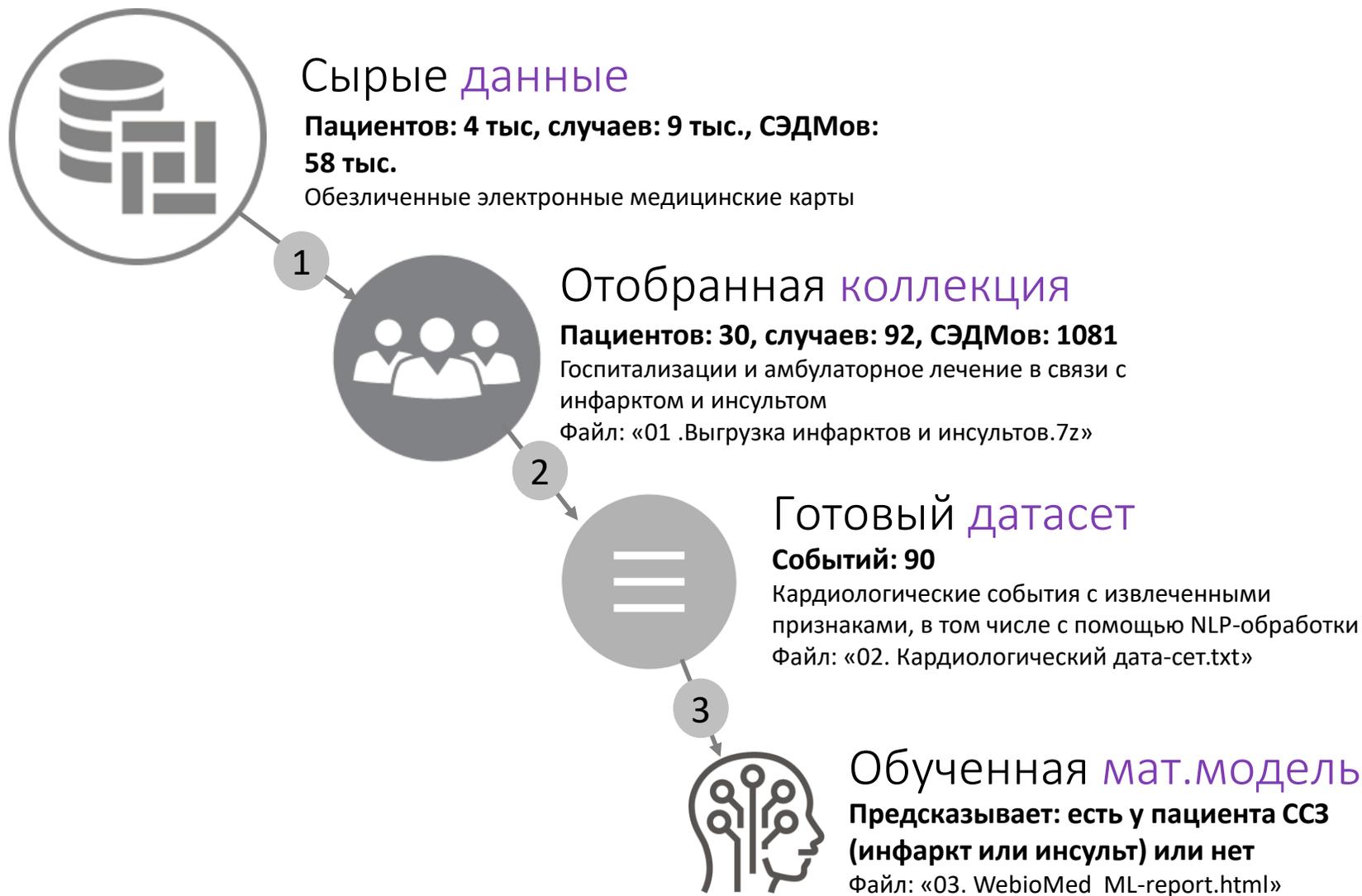
**200** Ть DICOM-изображений



Источники данных

Имя учетной записи	Регион	Название МО	Объем информации		
			Пациентов	Случаев	СЭМДов
msch21	Московская	ФГБУЗ ЦМСЧ № 21 ФМБА России	48441	0	0
msch174	Московская	Медико-санитарная Часть № 174 ФМБА России	13304	0	0
msch164	Московская	Медико-санитарная часть № 164 ФМБА России	15852	0	0
ozersk	Челябинская	ФГБУЗ "Клиническая больница №71" ФМБА России	500	2614	20407
medsanchast100	Приморский	ФГБУЗ "Медико-санитарная часть № 100 ФМБА России	500	817	3043
medsanchast140	Пермский	ФГБУЗ "Медико-санитарная часть № 140 ФМБА России	500	1231	9260
medsanchast41	Удмуртская	ФГБУЗ "Медико-санитарная часть № 41" ФМБА России	500	1124	7531
medsanchast9	Москва	ФБУЗ "Медико-санитарная часть № 9" ФМБА России	502	14	14
medsanchast98	Приморский	ФГБУЗ "Медико-санитарная часть № 98 ФМБА России	478	677	849
medsanchast128	Алтай	ФГБУЗ "Медико-санитарная часть № 128 ФМБА России	501	2324	15987
medsanchast15	Челябинская	ФГБУЗ "Центральная медико-санитарная часть № 15 ФМБА России	502	522	1034
msch92	Челябинская	Медико-санитарная Часть № 92 ФМБА России	216927	0	0

# Отлажен процесс создания датасетов и машинного обучения

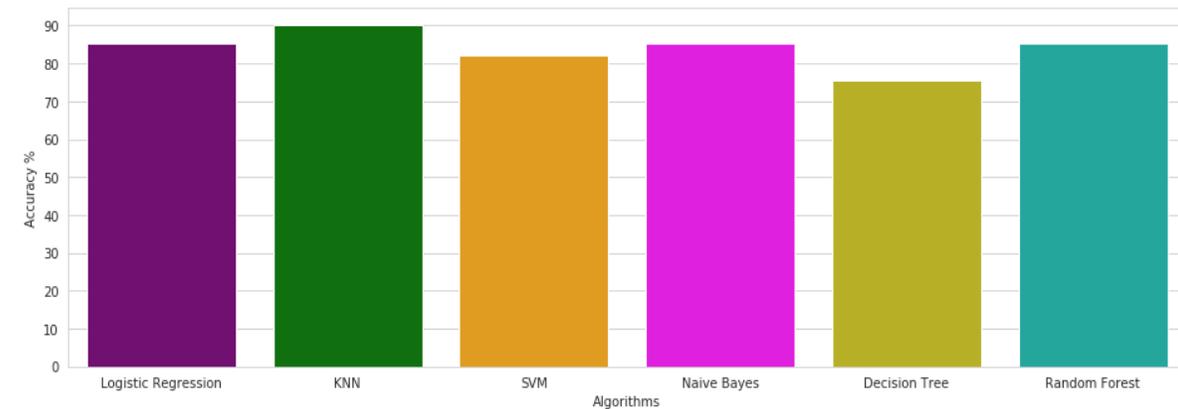


# Результаты машинного обучения

Задача: предсказать на основе ЭМК наличие развивающегося инфаркта или инсульта

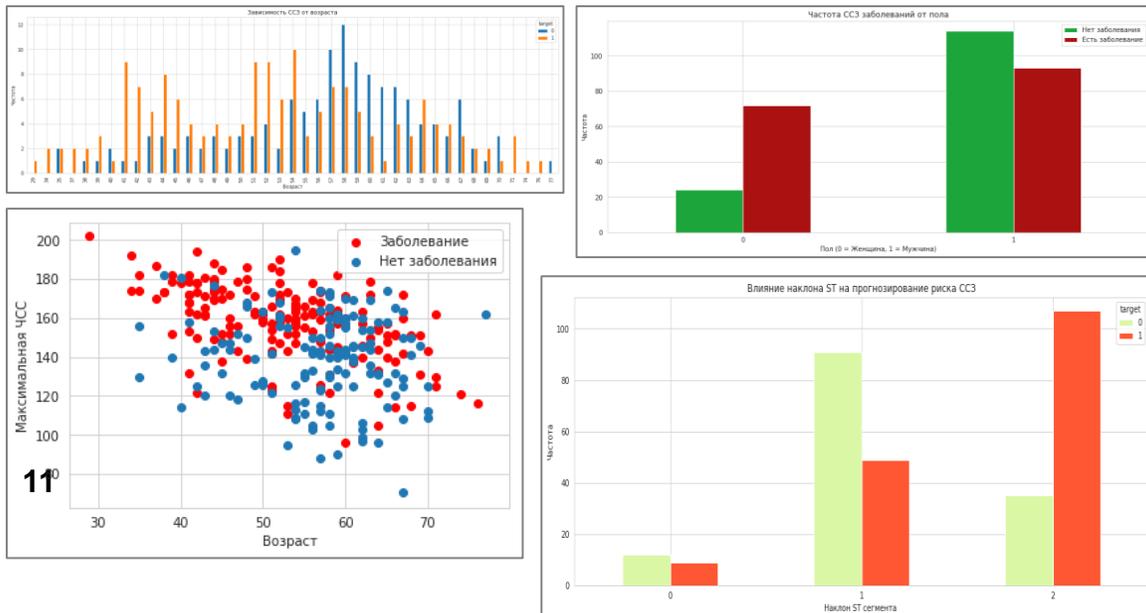
- `age` - возраст
- `sex` - (1 = мужчина; 0 = женщина)
- `cp` - тип грудной боли (1 - стенокардия, 2 - нестабильная стенокардия, 3 - другое, 4 - безсимптомная)
- `trestbps` - систолическое кровяное давление (mm Hg при поступлении)
- `fbs` - (быстрый сахар > 120 mg/dl) (1 = true; 0 = false)
- `restecg` - resting electrocardiographic results 0 = normal; 1 = having ST-T; 2 = hypertrophy
- `thalach` - максимальная частота сердечных сокращений
- `exang` - индуцированная стенокардия (1 = yes; 0 = no)
- .....
- `target` - have disease or not (1=yes, 0=no)

Сравнение Моделей:

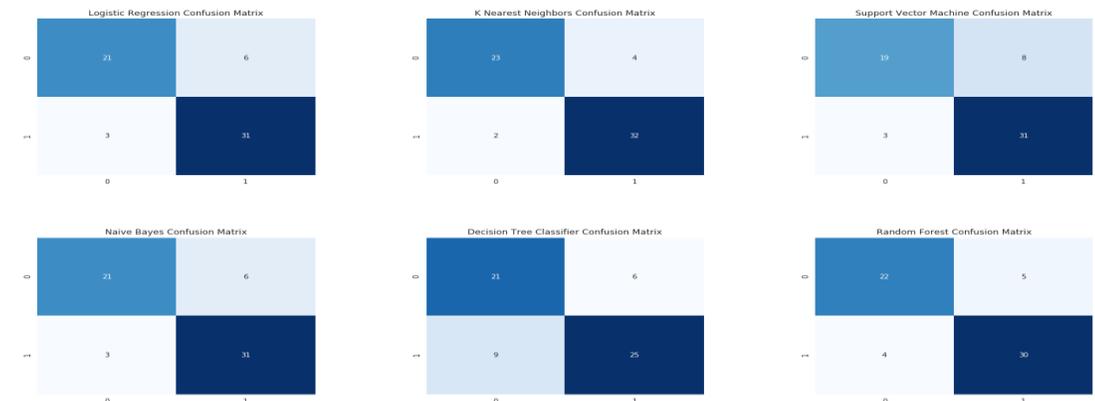


Accuracy (качество обучения): **85,25%**

Confusion Matrix (матрица неточностей):



Confusion Matrixes



Спасибо за  
внимание!