



WebioMed

Обзор и план развития ML направления 2022
Материалы к совещанию 20.01.2022

Результаты работы в 2021 году

2

Встроены в платформу:

- **9** моделей:
 - 5 моделей госпитализации в течении 12 месяцев (для пациентов с ССЗ, сахарным диабетом, наркологическими, гинекологическими и легочными заболеваниями)
 - 2 модели по гранту ПетрГУ (атеросклеротические бляшки, преэклампсия)
 - 2 модели по предсказания смертности при диабете 2го типа (в течение 1 года и 5 лет)

Разработаны по другим проектам:

- 4 модели прогнозирования смертности (по проекту ЦНИИОЗ)
- 7 моделей для прогнозирования выживаемости после операций стентирования, шунтирования коронарных артерий (проект Бакулева)

Команда: 4 разработчика



Затраты:

1. Общий бюджет на ML команду: **7 млн. рублей**
2. Затрата на разработку 1 модели: **350 тыс. рублей**
3. Времени на 1 модель в среднем: **54 дня**

Результаты работы в 2021 году

Модели встроенные в платформу

№	Название модели	Набор данных	Метрики
1	Модель прогнозирования вероятности госпитализации пациентов, имеющих в анамнезе сердечно-сосудистые заболевания (ССЗ), в течение ближайших 12 месяцев	4356	ACCURACY: 0.885 ROC AUC: 0.946
2	Модель прогнозирования вероятности госпитализации пациентов, имеющих в анамнезе легочные заболевания, в течение ближайших 12 месяцев	4190	ACCURACY: 0.77 ROC AUC: 0.79
3	Модель прогнозирования вероятности госпитализации пациентов, имеющих в анамнезе сахарный диабет (СД), в течение ближайших 12 месяцев	14409	ACCURACY: 0.78 ROC AUC: 0.68
4	Модель прогнозирования вероятности госпитализации пациентов, имеющих в анамнезе гинекологические заболевания, в течение ближайших 12 месяцев	1141	ACCURACY: 0.81 ROC AUC: 0.87
5	Модель прогнозирования вероятности госпитализации пациентов, имеющих в анамнезе наркологические заболевания, в течение ближайших 12 месяцев	345	ACCURACY: 0.80 ROC-AUC: 0.80
6	Модель прогнозирования вероятности смерти пациентов, имеющих в анамнезе сахарный диабет 2 типа, в течение 1 года	7812	ACCURACY: 0.88 ROC-AUC: 0.85
7	Модель прогнозирования вероятности смерти пациентов, имеющих в анамнезе сахарный диабет 2 типа, в течение 5 лет	20942	ACCURACY: 0.8 ROC-AUC: 0.8
8	Модель прогнозирования вероятности развития преэклампсии	3930	ACCURACY: 0.71 ROC AUC: 0.77
9	Модель оценки вероятности наличия атеросклеротических бляшек брахиоцефальных артерий у пациентов с ожирением	447	ACCURACY: 0.96 ROC AUC: 0.97

Текущая команда и обязанности

4



Разработчики моделей

- Анализ и подготовка данных
- Разработка моделей
- Подготовка отчетности по моделям



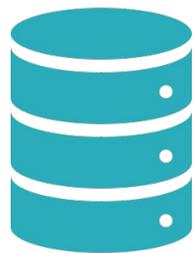
Аналитик

- Работа с документацией (подготовка ТЗ, проверка паспортов, отчетов, ведение статистики по разработанным моделям)
- AD-HOC, POST-HOC анализ данных (подготовка и анализ данных для нужд команды, техническое тестирование разработанных моделей)



Лидер команды

- Контроль выполнения задач
- Разработка моделей
- Выбор технических подходов для реализации алгоритмов (исследование SOTA подходов, их тестирование и внедрение, выбор и внедрение DS инструментов для разработки)
- Развитие разработчиков моделей (код ревью, разбор ошибок)



Доступность и качество данных

Проблемы:

1. Качество данных
2. Структура базы данных

Предложения:

1. Создание нового хранилища на основе Webiomed.Dataset с более продвинутой фильтрацией и удобной структурой хранения данных



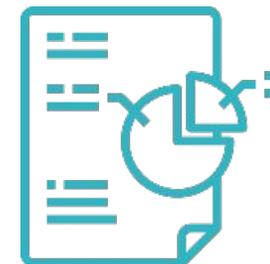
Тестирование моделей

Проблемы:

1. Нет тестирования

Предложения:

1. Выделить время медицинского эксперта на тестирование моделей (2 дня на модель)



Организация процессов

Проблемы:

1. Нет структуры отчетов, презентаций
2. Нет четкого плана и приоритета по разработке моделей
3. “Бюрократия” (долгая выдача доступов к внутренним ресурсам, трудно пробовать что-то новое)

Предложения:

1. Разработка структуры типового отчета, презентации
2. Разработка плана по моделям выбор ключевых направлений
3. Упростить внутреннее взаимодействие

Процесс разработки моделей

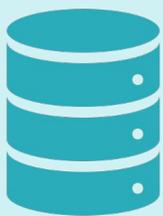
6



Разработка ТЗ

1. Аналитик
2. Врач

2 дня + 1 день
на передачу
разработчику



Сбор данных

1. Разработчик

5 дней



Анализ и подготовка данных

1. Разработчик

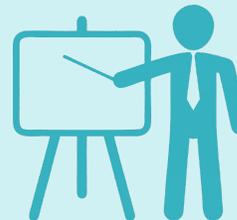
5 дней



Разработка модели

1. Разработчик

5 дней



Мини презентация

1. Разработчик

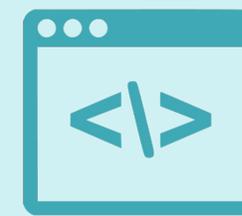
1 день



Тестирование модели

1. Врач

2 дня



Написание кода

1. Разработчик

2 дня



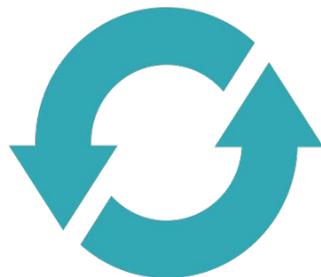
Отчет, паспорт, презентация

1. Разработчик
2. Аналитик

5 дней

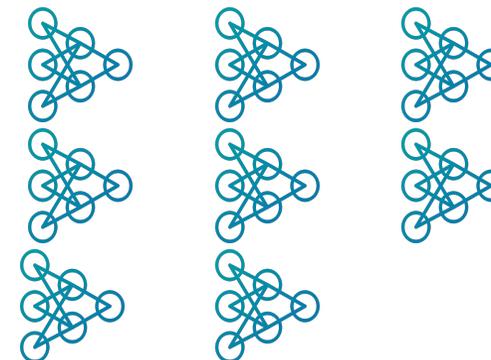
Подсчет времени разработки модели

7



Цикл разработки

1 модель - 28 рабочих дней (1.5 календарных месяца)



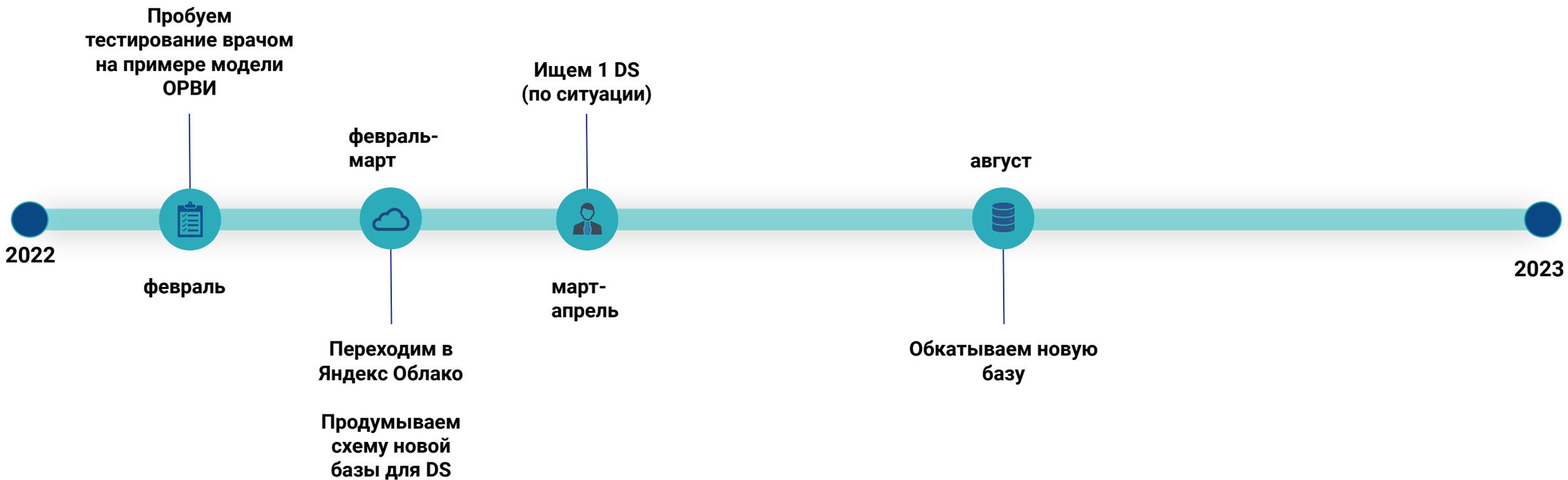
Производительность

В 2022 году 247 рабочих дней или 219 дней, вычитая отпуск

1 разработчик - 8 моделей в год

Roadmap 2022

8



1. Формат команды изменился, нужно время (~1-1.5 месяца) чтобы понять производительность
2. В 2021 многие из моделей не были доведены до платформы, необходимо ввести тестирование на этапе разработки и разработать список разрабатываемых моделей
3. Улучшив базу данных мы получим прирост в качестве моделей и скорости их разработки
4. Наш теоретический максимум - 20 моделей в год
5. Нужно упростить взаимодействие внутри компании (выдача прав и доступа к внутренним ресурсам)
6. Активно пробовать новые инструменты (это не всегда бесплатно, следует об этом помнить при планировании бюджета)