

# Технологии разработки ПО / Software Engineering

## Weekly report Recommendation Applicant Service

19.04.2021

A decorative graphic at the bottom of the slide. It features a light blue area on the right, a black horizontal band in the middle, and a large area on the left filled with diagonal grey lines. The boundaries between these areas are wavy and organic.

# Основные Достижения Отчетного Периода / Key Accomplishments For Reporting Period

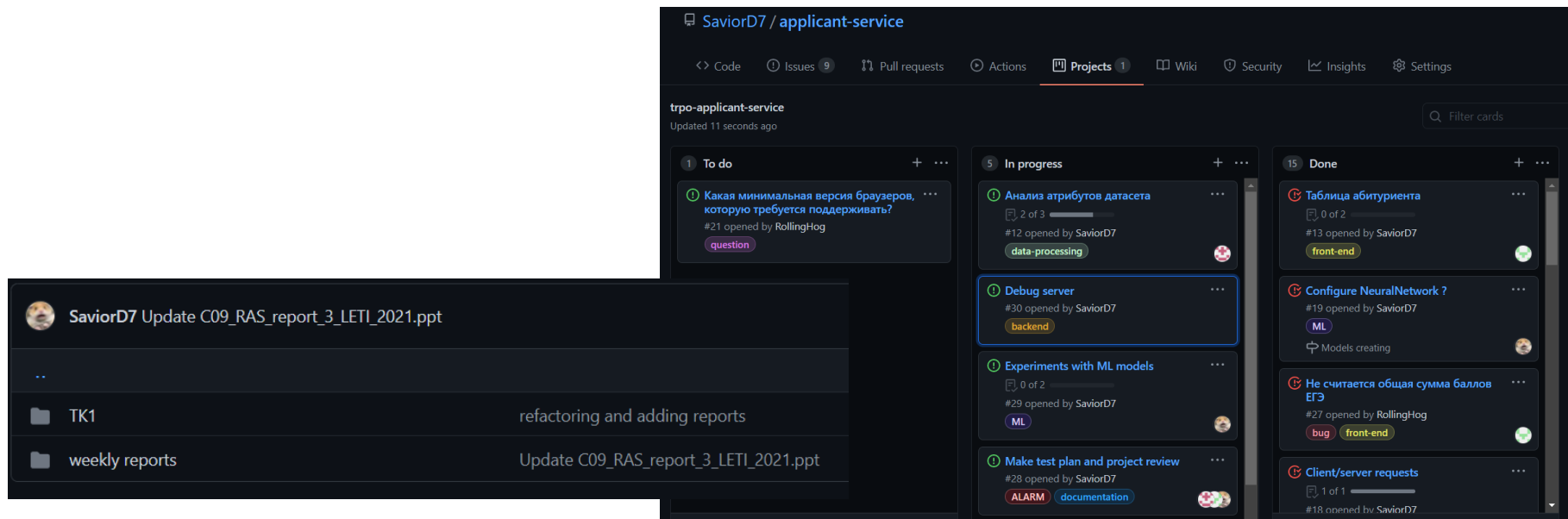
- ▶ Обновлена обработка данных на сервере
- ▶ Добавлены модели нейронной сети для предсказания среднего балла

# Текущие задачи / Current Tasks

- ▶ Провести эксперименты с параметрами для достижения наилучших результатов
- ▶ Отладка работы сервера
- ▶ Планирование и реализация тестирования

# Репозиторий/ Repository

- ▶ Для распределения и отслеживания задач используется доска Kanban
- ▶ Документация отчетных периодов и контрольных точек хранится в репозитории



# Репозиторий/ Repository

Используется две ветки



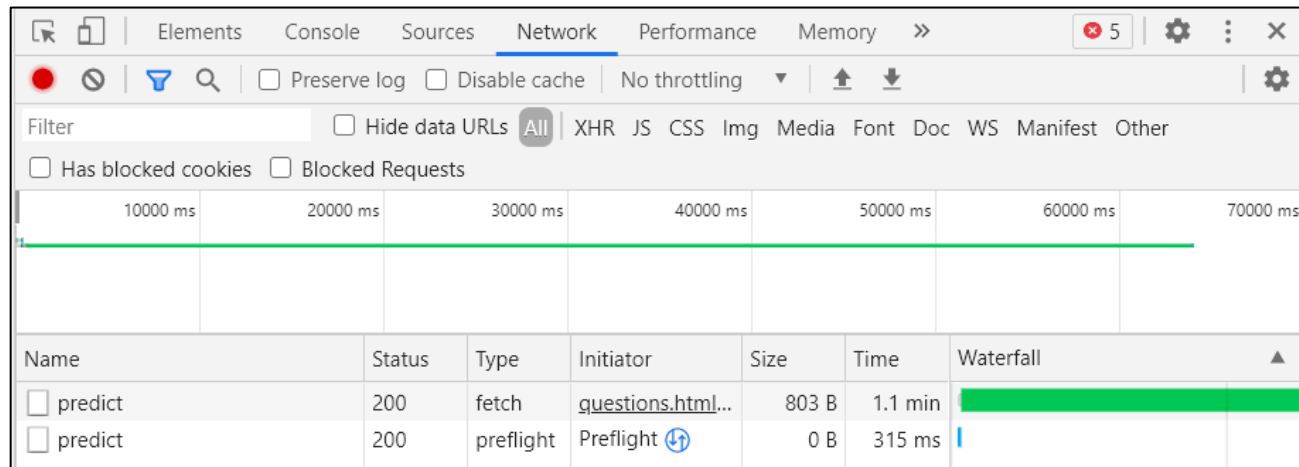
# Текущие проблемы / Current problems

- ▶ Из-за небольшого объема датасета некоторые модели линейной регрессии не справляются с задачей прогнозирования
- ▶ Это может быть также связано с тем, что набор атрибутов для определения одной модели отличается от набора атрибутов другой модели, в связи с чем некоторые из них попросту никогда «не видели» каких-то параметров.
- ▶ Нормализация помогает добиться чуть лучшего результата

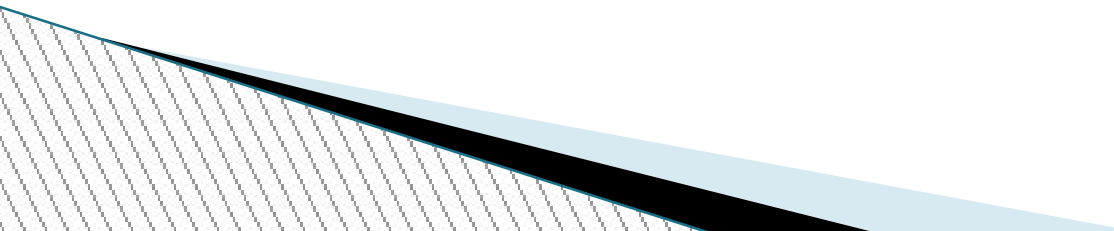
Код	Специальность	Предп. средняя оценка
45.03.02	Лингвистика	874830238298.01
38.03.01	Экономика	5.99
10.05.01	Компьютерная безопасность	4.98
27.03.02	Управление качеством	4.96
12.03.01	Приборостроение	4.90
27.03.05	Инноватика	4.81
09.03.01	Информатика и вычислительная техника. Искусственный интеллект	4.76
27.03.03	Системный анализ и управление	4.71
20.03.01	Техносферная безопасность	4.62
11.03.01	Радиотехника. Интеллектуальные радиотехнические системы	4.48
42.03.01	Реклама и связи с общественностью	4.46
11.03.03	Конструирование и технология электронных средств	4.44
13.03.02	Электроэнергетика и электротехника	4.40
11.03.04	Электроника и нанoeлектроника. Опoeлектроника и фотоника	4.36
27.03.04	Управление в технических системах	4.30
12.03.04	Биотехнические системы и технологии	4.27
38.03.02	Менеджмент	3.77
09.03.02	Информационные системы и технологии	3.30
28.03.01	Нанотехнологии и микросистемная техника	-9291982121013.17
11.03.02	Инфокоммуникационные технологии и системы связи	-66049564885.79
01.03.02	Прикладная математика и информатика	-590785423916.65
09.03.04	Программная инженерия	-4884165200608.46

# Текущие проблемы / Current problems

- ▶ Прогнозирование среднего балла через нейронные сети занимает большое количество времени, так как требуется время на загрузку каждой модели.



# Текущие проблемы / Current problems

- ▶ Нет идей для реализации других рекомендательных кейсов
  - ▶ С точки зрения рекомендательного сервиса самый оптимальный кейс – это предсказание среднего балла всех направлений, его реализация – это проектирование нескольких моделей, отвечающих за свое направление, что отрицательно сказывается на работе моделей, так как объем данных (и так не велик) значительно уменьшается.
  - ▶ Датасет также содержит параметр «Отчислен\ Не отчислен», но кейс, который будет предсказывать такую информацию выглядит некомпетентно. Тем более обстоятельства бывают разные и студент мог учиться хорошо, поэтому реализованный уже кейс опять же выигрывает.
  - ▶ Так как данные студентов являются реальными данными, собранными за несколько лет, то проводить аугментацию не является целесообразным. Такие данные могут не отражать реальные случаи.
  - ▶ На данный момент обдумываются еще возможные идеи
- 



# Датасет / Dataset

- ▶ В обеих моделях используются следующие атрибуты

```
arr_field_names = [  
    'БАЛЛ_ЗА_ДОСТИЖЕНИЯ',  
    'БАЛЛ_ОЛИМПИАДЫ_И_КОНКУРСЫ',  
    'ЕСТЬ_АТТЕСТАТ_С_ОТЛИЧИЕМ',  
    'ЕСТЬ_ДИПЛОМ_С_ОТЛИЧИЕМ',  
    'ОЦЕНКА_ЗА_СОЧИНЕНИЕ',  
    'УРОВЕНЬ_ДИПЛОМА',  
    'ОБЩЕЖИТИЕ_ТИП_ЗАСЕЛЕНИЯ',  
    'НАПРАВЛЕНИЕ_В_ПРИКАЗЕ_БАЛЛ',  
    'НАПРАВЛЕНИЕ_В_ПРИКАЗЕ_ПРЕДМЕТ_1_ПРИОРИТЕТ',  
    'НАПРАВЛЕНИЕ_В_ПРИКАЗЕ_ПРЕДМЕТ_2_ПРИОРИТЕТ',  
    'НАПРАВЛЕНИЕ_В_ПРИКАЗЕ_ПРЕДМЕТ_3_ПРИОРИТЕТ',  
    'НАПРАВЛЕНИЕ_В_ПРИКАЗЕ_ОЦЕНКА_1',  
    'НАПРАВЛЕНИЕ_В_ПРИКАЗЕ_ОЦЕНКА_2',  
    'НАПРАВЛЕНИЕ_В_ПРИКАЗЕ_ОЦЕНКА_3'
```

1

# Датасет / Dataset

- ▶ Текстовые признаки кодируются следующим образом

```
subject = {'Математика': 0, 'Английский язык': 1, 'Информатика': 2, 'Физика': 3, 'Обществознание': 4,  
           'Русский язык': 5, 'Химия': 6}  
education = {'Среднее': 0, 'Начальное профессиональное': 1, 'Среднее профессиональное': 2,  
             'Высшее профессиональное': 3}  
loc_student = {'Не требуется': 0, 'Общежитие': 1, 'Только регистрация': 2}
```

- ▶ Реализованная нейронная сеть является простой нейронной сетью прямого распространения (FeedForward). Параметры сети еще корректируются.

**Спасибо!**