

# Collector

Вольнов Сергей

# Задача

ДИТ МОСКВЫ

МИКРОЦИРКУЛЯЦИЯ

COLLECTOR



**Разработка алгоритма прогнозирования выполнения задачи**

timemanagement

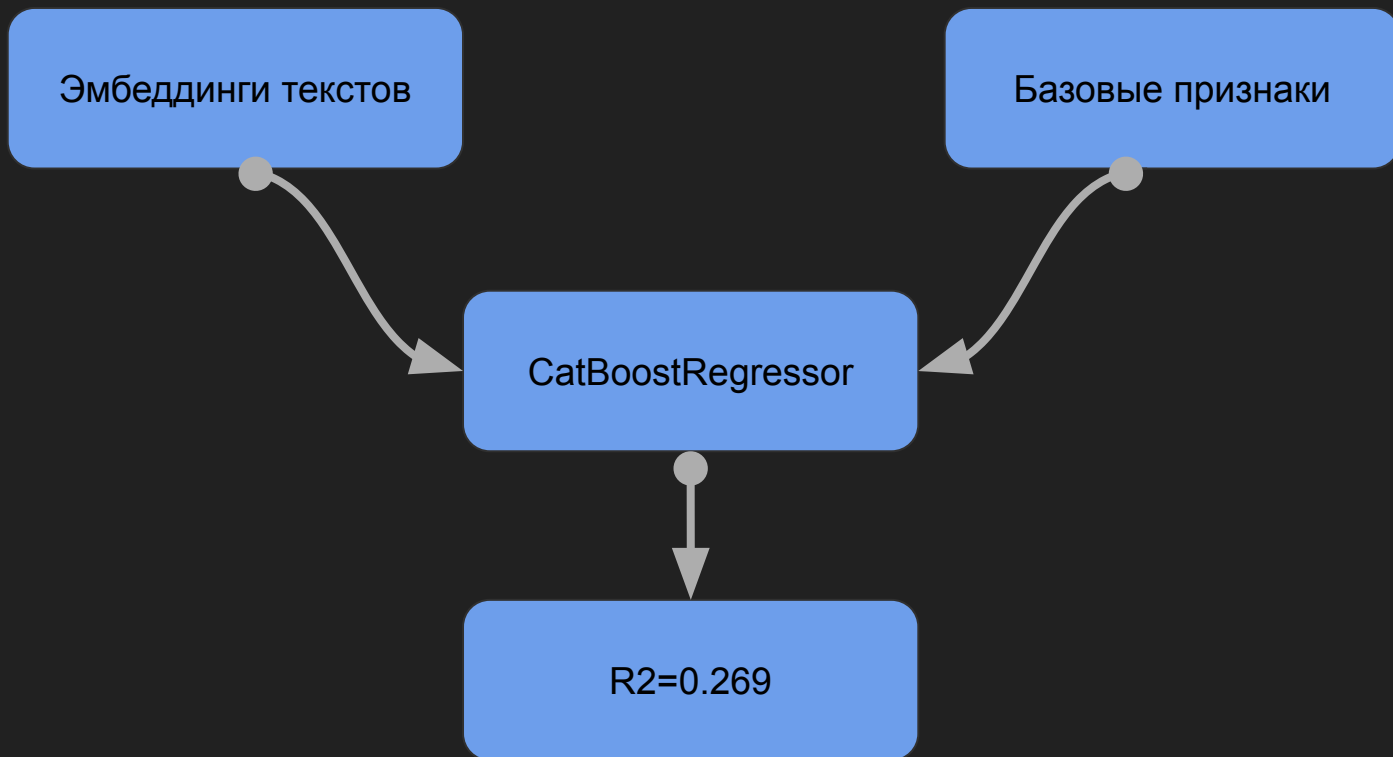
itsoft

# Данные

index	id		created	key	summary	project_id	assignee_id	creator_id	overall_worklogs	
0	7250	700722	2017-10-05 10:01:14	UN-1	Изучить API для приложения Элкарт	15	22	37	14400	
1	7251	700721	2017-10-05 10:08:53	UN-2	Изучить дизайн приложения по скриншотам в прил...	15	1	37	25200	
2	7252	700719	2017-10-06 07:45:43	UN-4	Дизайн экрана регистрации	15	1	37	10800	
3	7253	700718	2017-10-06 08:02:02	UN-5	Верстка формы регистрации нового пользователя	15	13	37	32400	
4	7254	700716	2017-10-06 08:37:59	UN-7	Ввод одноразового кода	15	13	37	10800	

	comment_id	text	issue_id	author_id
0	11779	[https://www.youtube.com/watch?v=tuhOdtsoNY h...	669666	1
1	10601	OK [~accountid:557058:3f7ab89a-8969-4547-90df-...	669670	1
2	76101	I encountered a problem with access to `/users...	670930	2
3	76102	I have learned that `users/:id/emails` endpoin...	670930	2
4	76213	We have decided with Andrew to set member's em...	670930	2

# Общий пайплайн



# Базовые признаки

- Кол-во комментариев
- Категориальные: assignee\_id, creator\_id, project\_id
- Агрегации по категориальным признакам
- Временные агрегации
- Парные агрегации



# Работа с текстом

- Summary
- Bert-base-multilingual-cased
- Finetune



# Модель

```
[42] params = {'n_estimators' : 3000,  
              'learning_rate': .03,  
              'max_depth' : 6,  
              'cat_features' : cat_cols,  
              # 'border_count' : 64,  
              'l2_leaf_reg' : 2,  
              # 'bagging_temperature' : 2,  
              # 'rsm' : .1,  
              'use_best_model': True,  
              'task_type': 'GPU',  
              'random_state': 42,  
              }
```

```
[31] model = CatBoostRegressor(**params)
```

# Выжимаем максимум

- Усредняем значения модели по разным random\_seed
- Feature selection
- Удаляем из обучающей выборки выбросы
- Обучаем на всем датасете
- Final r2: **0.269**





# Использованные инструменты



Не использовано импортное ПО

Спасибо за внимание