



LEADERSTAT

ВСЕГДА НА ШАГ ВПЕРЕДИ

O DATASET

Общий объём данных в CSV

800 Mb

Полезный объём после очистки

800 Кб

Количество с летальным исходом

.078

Количество реабилитированных

78 000



НАША ЦЕЛЬ

ДАТЬ МЕДИКАМ
РЕАНИМАТОЛОГАМ ПОЛЕЗНЫЙ
ИНСТРУМЕНТ КОТОРЫЙ
ПОЗВОЛИТ:



Исключить утомительную ручную
обработку поступивших данных.

Опробовать систему определения
степени достоверности введённых
данных (косвенно).

Снижение числа ошибок по вине
“человеческого фактора”.

Выявление и профилактика рисков
заболевания уже на ранней стадии
(превентивная медицина).

Визуальный вывод обрабатываемой
информации в виде графиков для
удобства восприятия врачом, с
фильтрацией по (место проживания, пол,
стадии заболевания или типу
заболевания).



И ЧТОБЫ ЭТО БЫЛО УДОБНО
ДЛЯ ВРАЧА И ПАЦИЕНТА

Методика

1. Исследование эффективности и применимости алгоритмов ML в медицинской практике.
2. Обязанность врача исполнять рекомендации МЗ РФ, Они постоянно обновляются.
3. Опыт внедрения систем на базе ИИ - Webiomed, Zopinion, Celsius, BotkinAI.
4. Существующие на рынке инструменты разбора большого массива структурированных данных предъявляют высокие требования к мощностям вычислительных систем, квалификации DataScientist.
5. Нейросетевое обучение занимает длительный промежуток времени (от 10ч. до неск. дней). Необходимость постоянной коррекции при внесении изменения в процедуру и регламент собираемых, обрабатываемых данных, введения новых типов нозологий и анамнеза.
6. Разметка данных сейчас несовершенна что влияет на качество предсказания обучаемых нейронных сетей. Проблема разметки Datasets данных решается сейчас неизвестными лицами за солидное вознаграждение.
7. Проблема с безопасностью данных при обработке и хранения персонифицированной информации является одной из важнейших.
8. Большинство специалистов отпугивает юридическая ответственность за использование Dataset содержащий обезличенные данные пациентов без разрешения правообладателя этого набора данных.

ПАМЯТКА

ИМТ от 18,5 кг/м² до 24,9 кг/м²

Нормальная масса тела. Продолжайте придерживаться вашего рациона питания и режима физической активности.

ИМТ ≤ 18,5 кг/м²

Дефицит массы тела. Недостаточная масса тела может плохо отразиться как на здоровье в целом, так и на состоянии опорно-двигательного аппарата.

ИМТ от 25 кг/м² до 29,9 кг/м²

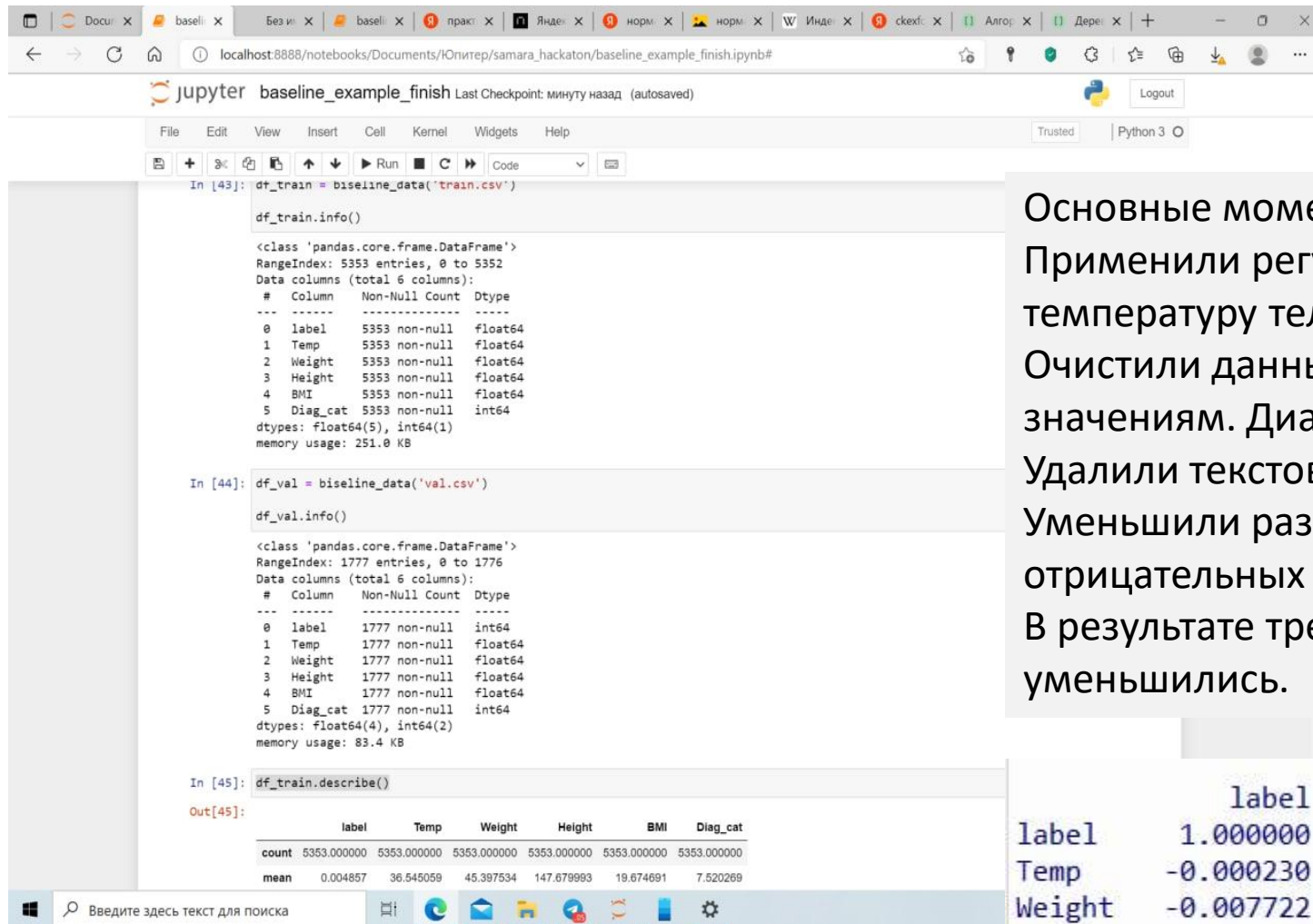
Избыток массы тела. Необходимо снизить вес. Больше двигайтесь и уменьшите калорийность питания.

ИМТ от 30 кг/м² и более

Ожирение. Необходимо как можно скорее нормализовать массу тела. Рекомендуется консультация специалиста.

рост, м масса, кг	1,45	1,50	1,55	1,60	1,65	1,70	1,75	1,80	1,85	1,90	1,95	2,00
40	19,0	17,8	16,6	15,6	14,7	13,8	13,1	12,3	11,7	11,1	10,5	10,0
45	21,4	20,0	18,7	17,6	16,5	15,6	14,7	13,9	13,1	12,5	11,8	11,3
50	23,8	22,2	20,8	19,5	18,4	17,3	16,3	15,4	14,6	13,9	13,1	12,5
55	26,2	24,4	22,9	21,5	20,2	19,0	18,0	17,0	16,1	15,2	14,5	13,8
60	28,5	26,7	25,0	23,4	22,0	20,8	19,6	18,5	17,5	16,6	15,8	15,0
65	30,9	28,9	27,1	25,4	23,9	22,5	21,2	20,1	19,0	18,0	17,1	16,3
70	33,3	31,1	29,1	27,3	25,7	24,2	22,9	21,6	20,5	19,4	18,4	17,5
75	35,7	33,3	31,2	29,3	27,5	26,0	24,5	23,1	21,9	20,8	19,7	18,8
80	38,0	35,6	33,3	31,3	29,4	27,7	26,1	24,7	23,4	22,2	21,0	20,0
85	40,4	37,8	35,4	33,2	31,2	29,4	27,8	26,2	24,8	23,5	22,4	21,3
90	42,8	40,0	37,5	35,2	33,1	31,1	29,4	27,8	26,3	24,9	23,7	22,5
95	45,2	42,2	39,5	37,1	34,9	32,9	31,0	29,3	27,8	26,3	25,0	23,8
100	47,6	44,4	41,6	39,1	36,7	34,6	32,7	30,9	29,2	27,7	26,3	25,0
105	49,9	46,7	43,7	41,0	38,6	36,3	34,3	32,4	30,7	29,1	27,6	26,3
110	52,3	48,9	45,8	43,0	40,4	38,1	35,9	34,0	32,1	30,5	28,9	27,5
115	54,7	51,1	47,9	44,9	42,2	39,8	37,6	35,5	33,6	31,9	30,2	28,8
120	57,1	53,3	49,9	46,9	44,1	41,5	39,2	37,0	35,1	33,2	31,6	30,0
125	59,5	55,6	52,0	48,8	45,9	43,3	40,8	38,6	36,5	34,6	32,9	31,3

ПРОЦЕСС ПОДГОТОВКИ ОБУЧЕНИЯ



```
In [43]: df_train = baseline_data('train.csv')
df_train.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5353 entries, 0 to 5352
Data columns (total 6 columns):
#   Column      Non-Null Count  Dtype  
---  -
0    label      5353 non-null   float64
1    Temp       5353 non-null   float64
2    Weight     5353 non-null   float64
3    Height     5353 non-null   float64
4    BMI        5353 non-null   float64
5    Diag_cat   5353 non-null   int64   
dtypes: float64(5), int64(1)
memory usage: 251.0 KB

In [44]: df_val = baseline_data('val.csv')
df_val.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1777 entries, 0 to 1776
Data columns (total 6 columns):
#   Column      Non-Null Count  Dtype  
---  -
0    label      1777 non-null   int64   
1    Temp       1777 non-null   float64
2    Weight     1777 non-null   float64
3    Height     1777 non-null   float64
4    BMI        1777 non-null   float64
5    Diag_cat   1777 non-null   int64   
dtypes: float64(4), int64(2)
memory usage: 83.4 KB

In [45]: df_train.describe()

Out[45]:
```

	label	Temp	Weight	Height	BMI	Diag_cat
count	5353.000000	5353.000000	5353.000000	5353.000000	5353.000000	5353.000000
mean	0.004857	36.545059	45.397534	147.679993	19.674691	7.520269

Основные моменты:

Применили регулярные запросы и получили: массу, рост, температуру тела, коды диагноза.

Очистили данные: привели показатели к референсным значениям. Диагноз по МКБ10 первое значение.

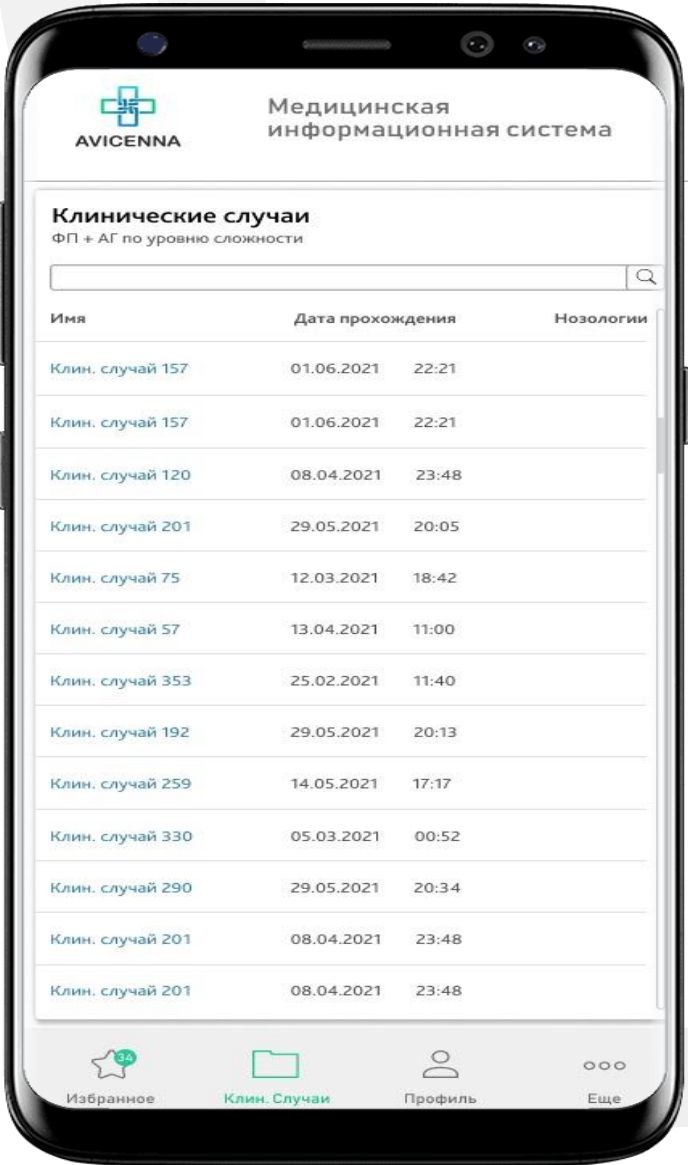
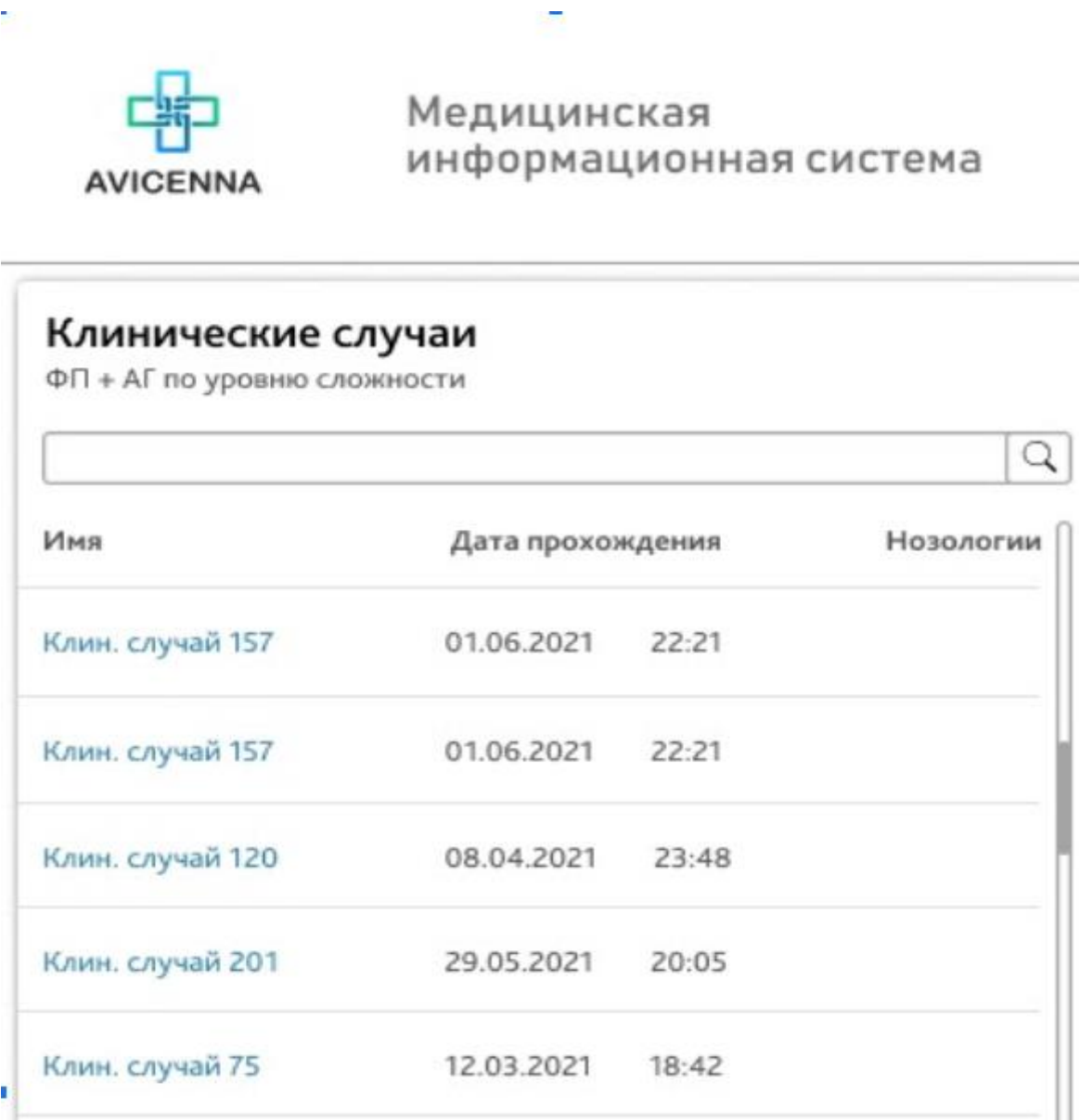
Удалили текстовые данные и id.

Уменьшили размерность выборки 3 раза за счет отрицательных значений.

В результате тренировочные и валидационные датасеты уменьшились.

	label	Temp	Weight	Height	Diag_cat
label	1.000000	-0.000230	-0.007722	-0.000885	-0.027693
Temp	-0.000230	1.000000	0.034326	0.032889	-0.031909
Weight	-0.007722	0.034326	1.000000	0.576994	-0.120055
Height	-0.000885	0.032889	0.576994	1.000000	-0.139607
Diag_cat	-0.027693	-0.031909	-0.120055	-0.139607	1.000000

СИСТЕМА ПОДДЕРЖКИ ПРИНЯТИЯ ВРАЧЕБНЫХ РЕШЕНИЙ. ПРИЛОЖЕНИЕ



КОМАНДА



ГРИБАНОВ ДМИТРИЙ

Технологический лидер ГК РТ

Капитан



СОФЬЯ МАЙТОМАЛОВА

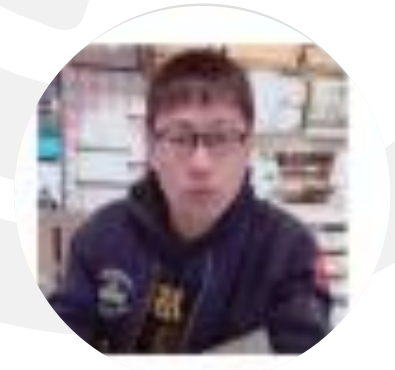
Дизайн. UX. Figma



КУЗНЕЦОВ ЮРИЙ

СТО

Начальник ИТ отдела



КАСЬКОВ АЛЕКСАНДР

Врач анестезиолог-
реаниматолог

ВЫВОДЫ

Методики сбора наборов данных для ML требуют серьёзной доработки

Нужна система контроля референсных значений с подтверждением

если выходит за границу значений более чем 1,5 раза

Классификаторы из Клинических рекомендаций.

Автоматическое предложение кода диагноза. Объективные данные из

текста должны переноситься в базу данных.

Лист назначений должен идти отдельно.