

Отчет

Игнатович Пётр

23 марта 2025 г.

Содержание

1	Анализ данных	2
1.1	Описание набора данных	2
2	Описание набора данных	2
2.1	Визуализация распределений	3
2.2	Проверка пропусков и выбросов	3
3	Выбор модели	4
3.1	Описание архитектуры нейронной сети	4
4	Описание архитектуры нейронной сети	4
4.1	Обоснование выбора активационных функций, оптимизатора и loss-функции	5
4.2	Гиперпараметры	5
5	Результаты	6
5.1	Таблица с метриками	6
5.2	График обучения	6
5.3	Ссылка на submission в leaderboard	6

1 Анализ данных

1.1 Описание набора данных

2 Описание набора данных

Данный набор данных содержит из соревнования по определению типа ожирения 20 758 записей и 17 столбцов, включающих признаки участников исследования и целевую переменную, определяющую тип веса. Ниже приведено подробное описание каждого столбца:

Gender Пол участника. Возможные значения: 'Male', 'Female'.

Age Возраст участника.

Height Рост участника (например, в метрах).

Weight Вес участника (например, в килограммах).

family_history_with_overweight Семейная история лишнего веса. Возможные значения: 'yes', 'no'.

FAVC Частое потребление калорийной пищи. Возможные значения: 'yes', 'no'.

FCVC Частота потребления овощей. Значения: от 1 до 3 (например, 1 — редко, 3 — часто).

NCP Количество основных приёмов пищи. Значения: от 1 до 4.

CAEC Перекусы между приёмами пищи. Возможные значения: 'no', 'Sometimes', 'Frequently', 'Always'.

SMOKE Курит ли человек. Возможные значения: 'yes', 'no'.

CH2O Потребление воды (приблизительное количество литров в день). Значения: от 1 до 3.

SCC Контроль калорий. Возможные значения: 'yes', 'no'.

FAF Физическая активность. Количество часов физической активности в неделю. Значения: от 0 до 3+.

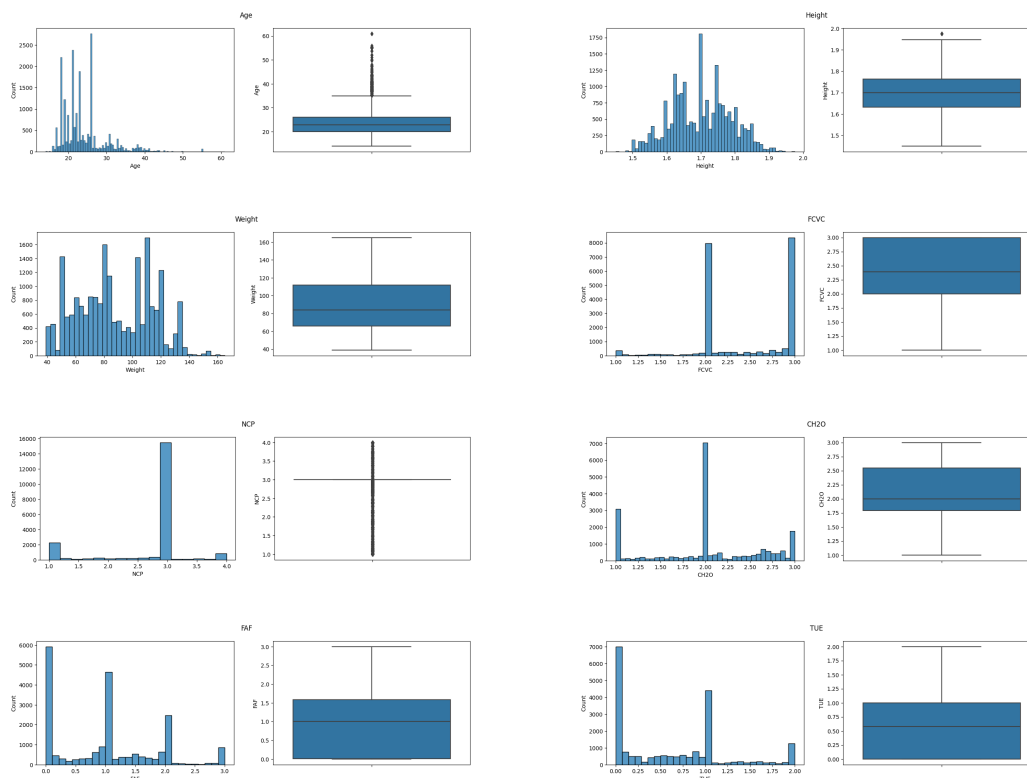
TUE Использование технологий. Количество часов в день, в течение которых участник использует электронные устройства. Значения: от 0 до 3+.

CALC Употребление алкоголя. Возможные значения: 'no', 'Sometimes', 'Frequently'.

MTRANS Транспорт. Основной способ передвижения участника. Возможные значения: 'Walking', 'Bike', 'Motorbike', 'Automobile', 'Public_Transportation'.

NObeyesdad Целевая переменная (тип веса), предсказываемая моделью. Возможные значения: 'Insufficient_Weight', 'Normal_Weight', 'Overweight_Level_I', 'Overweight_Level_II', 'Obesity_Type_I', 'Obesity_Type_II', 'Obesity_Type_III'.

2.1 Визуализация распределений



2.2 Проверка пропусков и выбросов

Пропусков не было изначально, а что касается выбросов, то при сравнении нескольких результатов с разным ограничением на предел того, что считать выбросом стало понятно, что лучше их все сохранять, т.к. иначе модель не сможет нормально обучиться, ведь данных не хватит.

3 Выбор модели

3.1 Описание архитектуры нейронной сети

4 Описание архитектуры нейронной сети

В данной работе используется полносвязная нейронная сеть, состоящая из следующих слоев:

1. **Входной слой:** Принимает вектор из 22 признаков.
2. **Первый скрытый слой (FC1):**
 - Полносвязный слой, преобразующий 22 входных признака в 100 нейронов.
 - **Batch Normalization** для стабилизации и ускорения обучения.
 - Функция активации **ReLU** для введения нелинейности.
 - **Dropout** с вероятностью 0.3 для регуляризации и предотвращения переобучения.
3. **Второй скрытый слой (FC2):**
 - Полносвязный слой, преобразующий 100 нейронов в 200 нейронов.
 - **Batch Normalization.**
 - Функция активации **ReLU**.
 - **Dropout** с вероятностью 0.4.
4. **Третий скрытый слой (FC3):**
 - Полносвязный слой, преобразующий 200 нейронов в 100 нейронов.
 - **Batch Normalization.**
 - Функция активации **ReLU**.
 - **Dropout** с вероятностью 0.4.
5. **Выходной слой (FC4):**
 - Полносвязный слой, отображающий 100 нейронов в 7 выходов, соответствующих 7 классам целевой переменной.

4.1 Обоснование выбора активационных функций, оптимизатора и loss-функции

Функция активации ReLU была выбрана из-за ее надежности и практичности, оптимизатор Adam чтобы сходиться быстрее и не застревать в локальных минимумах и CrossEntropyLoss в качестве функции потерь, т.к. первое что в голову пришло.

4.2 Гиперпараметры

Укажите гиперпараметры, использованные в обучении:

- Число слоев: 3
- Размер батча: 128
- Количество эпох: 25
- Скорость обучения: 0.01

5 Результаты

5.1 Таблица с метриками

Ниже представлена таблица с основными метриками (например, ассигасу, loss, RMSE и др.):

Метрика	Значение
Accuracy	0.8814
Loss	0.3268

5.2 График обучения

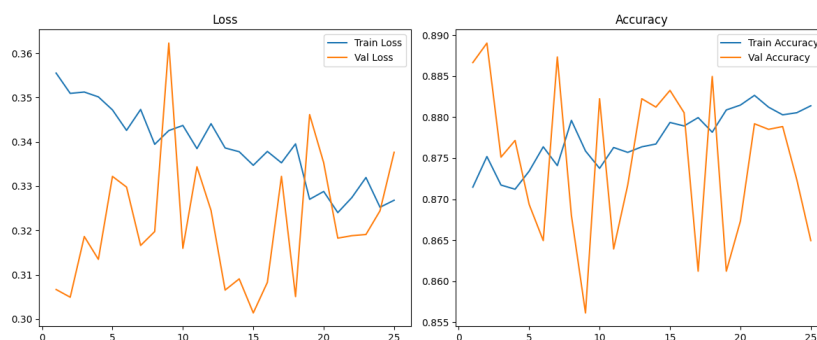


Рис. 1: График обучения: loss и ассигасу на train/val

5.3 Ссылка на submission в leaderboard

Так соревнование же закончилось уже, мы никак в лидерборд не попадем