

Санкт–Петербургский государственный университет

АНТОНОВ Никита Сергеевич

Выпускная квалификационная работа

Построение рациона питания с учетом индивидуальных физиологических особенностей на основе методов машинного обучения

Уровень образования: бакалавриат

Направление 02.03.02 «Фундаментальная информатика и информационные технологии»

Основная образовательная программа СВ.5003.2016 «Программирование и информационные технологии»

Профиль «Автоматизация научных исследований»

Научный руководитель:

Крылатов Александр Юрьевич

Рецензент:

Севрюков Сергей Юрьевич

Санкт-Петербург

2020 г.

Содержание

Введение	3
Постановка задачи	4
Обзор литературы	5
Глава 1. Теоретические основы разработки плана питания	6
1.1. Показатель основного обмена	6
1.2. Уровень физической активности и энергетические затраты	7
1.3. Тестовые данные	8
1.4. База продуктов	8
Глава 2. Обзор существующих методов оценки ПОО	10
Глава 3. Создание нейронной сети	13
3.1. Структура нейронной сети	13
3.2. Сравнительный анализ методов оценки ПОО	14
Глава 4. Поиск оптимального плана питания	17
Глава 5. Разработка веб-приложения	20
5.1. Используемые технологии	20
5.2. Архитектура приложения	21
5.3. Структура базы данных	22
5.4. Представления приложения	24
Заключение	27
Список литературы	28
Приложение	30

Введение

Одна из главных задач, стоящих перед современной наукой, — поиск способов и средств для увеличения продолжительности жизни человека, в особенности — продолжительности здоровой жизни (healthy life expectancy). По всему миру врачи различных специализаций: от генетиков до иммунологов, работают над тем, чтобы выявить механизмы возникновения смертельно опасных заболеваний с целью их профилактики и лечения. Хотя их заслуги неизмеримы, и в будущем многие из текущих проблем найдут свои решения, уже сейчас мы можем многое сделать, чтобы улучшить качество своей жизни.

Ни для кого не секрет, что соблюдение правил здорового питания способствует профилактике заболеваний и укреплению общего состояния здоровья. Более того, здоровая диета в сочетании с регулярными физическими нагрузками сокращает риск таких хронических заболеваний и расстройств, как ожирение, диабет, остеопороз, а также сердечно-сосудистых заболеваний. Хотя врачи регулярно проводят исследования, направленные на оптимизацию рациона питания человека, одна из неизменных рекомендаций — соблюдение энергетического баланса.

Данная работа нацелена на создание приложения, рассчитывающего энергетические потребности человека по его физиологическим характеристикам, а затем генерирующего индивидуализированный план питания. В работе были исследованы существующие методы оценки энергетических затрат человека, а также проанализирована точность модели предсказания на основе нейронной сети.

Постановка задачи

В данной выпускной квалификационной работе были выделены несколько основных задач:

1. Рассмотреть существующие методы оценки показателя основного обмена человека (ПОО).
2. Создать модель предсказания ПОО на основе многослойной нейронной сети.
3. Провести сравнительный анализ точности результатов предсказания.
4. Сформулировать задачу составления оптимального плана питания по заданным показателям пищевой ценности как задачу линейного программирования.
5. Разработать веб-приложение, генерирующее рекомендуемый рацион питания на основе физиологических параметров человека.

Обзор литературы

При написании данной работы были использованы учебно-методическая, научная и справочная литература, публикации в различных научных изданиях и электронные ресурсы.

Наиболее часто применяемые методы оценки показателя основного обмена описаны в работах [1][2][3][4]. Уравнения, разработанные в результате этих исследований, были использованы в проведении сравнительного анализа в главе 3.

Отчет [5], опубликованный ВОЗ и ФАО ООН в 2004 году, содержит обзор показателей, влияющих на энергетические потребности человека, и текущие рекомендации потребления энергии. В публикации [6] представлены рекомендуемые ВОЗ суточные нормы потребления основных витаминов и минералов.

В качестве источника данных о пищевой ценности продуктов был использован справочник [7].

В [8] описаны общая структура и принцип работы нейронных сетей. [9] рассматривает задачи математического программирования, в частности задачу смешанного целочисленного линейного программирования с логическими ограничениями, описанную в главе 4.

В процессе разработки веб-приложения была использована официальная документация фреймворка Django [10], библиотек Keras [11], PuLP [12] и React [13].

Глава 1. Теоретические основы разработки плана питания

1.1 Показатель основного обмена

Наиболее важным фактором, определяющим потребности человека в пище, являются его энергетические затраты. Общие энергетические затраты, в свою очередь, зависят от основного обмена человека и уровня его физической активности.

Для дальнейшего исследования необходимо дать некоторые определения. Показатель основного обмена человека (ПОО, basal metabolic rate, BMR) — это минимальное количество энергии, необходимое для обеспечения нормальной жизнедеятельности организма в стандартных условиях. Под стандартными условиями понимают: состояние бодрствования, натощак, в лежащем положении, в условиях эмоционального покоя, при температуре воздуха 18–20 °С.

Измерения основного обмена человека, как правило, производятся с помощью прямой или непрямой калориметрии. Прямая калориметрия подразумевает помещение человека в калориметр — специальную камеру, которая улавливает тепло, отдаваемое телом. Непрямая калориметрия основана на зависимости между выделенным человеком теплом за определенный промежуток времени и объемами поглощенного кислорода и выделенного за этот промежуток углекислого газа. Хотя непрямая калориметрия дешевле прямой, оба метода объединяет необходимость в специализированном оборудовании (часто дорогостоящем) и квалифицированном медицинском персонале.

Однако важность определения показателя основного обмена человека как для клинической медицины, так и для диетологии, способствовала проведению большого числа исследований, целью которых было определение характеристик, способных с достаточной точностью предсказывать ПОО. Некоторые из наиболее значимых исследований рассмотрены в следующей главе.

Не вдаваясь в подробности, стоит отметить, что физиологические характеристики, которые наиболее сильно коррелируют со значением ПОО, это: масса, процент массы жировой ткани, рост, пол и возраст. Сложность изме-

рения процента жировой массы человека стала причиной, по которой большинство разработанных уравнений для предсказания показателя основного обмена не используют данный параметр.

1.2 Уровень физической активности и энергетические затраты

Как было отмечено ранее, общие энергетические затраты (total energy expenditure, TEE) человека складываются из основного обмена человека и его уровня физической активности.

Уровень физической активности (physical activity level, PAL) традиционно выражается коэффициентом показателя основного обмена, т.е. справедлива формула:

$$TEE = BMR \times PAL$$

Хотя значения уровня физической активности могут варьироваться в пределах от 1,4 до 2,4, руководствуясь отчетом ВОЗ/ФАО/УООН [5], их можно классифицировать следующим образом:

Таблица 1: Уровни физической активности

Характеристика образа жизни	Значение PAL	Описание	Примеры профессий
Сидячий	1,55	Люди профессий, не требующих значительных физических усилий, преодоления больших расстояний пешком. Люди, как правило, использующие личный автотранспорт для передвижения, не занимающиеся спортом регулярно, предпочитающие свободное время проводить сидя или лежа (чтение, просмотр ТВ).	Офисные работники
Средней активности	1,75	Люди профессий, не предполагающих изматывающий физический труд, но требующих больших энергетических затрат, чем в описании сидячего образа жизни. Люди, как правило, пользующиеся общественным транспортом, проводящие много времени на ногах.	Продавцы розничной торговли, кладовщики

Активный	2	Люди профессий, требующих регулярных физических нагрузок. Люди, в течение часа в день занимающиеся изматывающими физическими упражнениями.	Строители
Крайне высокой активности	2,25	Люди, занимающиеся изматывающим физическим трудом или физическими упражнениями в течение нескольких часов в день.	Спортсмены в период интенсивных тренировок, сельскохозяйственные работники

1.3 Тестовые данные

Набор данных, используемый в работе для оценки точности уравнений и обучения нейронной сети, был собран доктором С. J. К. Henry в процессе написания статьи [1], и любезно предоставлен для использования в данной работе доктором Т. J. Cole.

Данные состоят из 13910 измерений роста, массы тела, возраста, пола и ПОО мужчин и женщин в возрасте от 0 до 106 лет. Набор данных был агрегирован из 174 научных публикаций, изданных в период с 1914 по 2001 год.

1.4 База продуктов

Для составления рациона питания необходима база продуктов. Источником данных о продуктах и их пищевой ценности послужил справочник [7]. Отобранные данные были вручную размечены на принадлежность к некоторым категориям: «веганский продукт», «вегетарианский продукт», «халяльный продукт», «низколактозный продукт», «не содержит глютена». Такая классификация позволит генерировать план питания только из продуктов, соответствующих пользовательским предпочтениям.

Каждый продукт характеризуется следующими атрибутами: **name** (название продукта), **categories** (категории, к которым относится продукт), **water**

(содержание воды, г), **energy** (энергетическая ценность, ккал), **protein** (содержание белка, г), **fat** (содержание жира, г), **carbohydrate** (содержание углеводов, г), **sfa** (насыщенные жирные кислоты, г), **pufa** (полиненасыщенные жирные кислоты, г), **cholesterol** (холестерин, мг), **mds** (моно- и дисахариды, г), **starch** (крахмал, г), **fiber** (пищевые волокна, г), **sodium** (натрий, мг), **potassium** (калий, мг), **calcium** (кальций, мг), **magnesium** (магний, мг), **phosphorus** (фосфор, мг), **iron** (железо, мг), **retinol** (витамин А, мкг), **beta_carotene** (β -каротин, мкг), **tocopherol_eq** (токоферолэквивалент, мг), **thiamine** (витамин В₁, мг), **riboflavin** (витамин В₂, мг), **niacin** (витамин РР, мг), **niacin_eq** (ниациновый эквивалент, мг), **vitamin_c** (витамин С, мг).

Глава 2. Обзор существующих методов оценки ПОО

В данной главе будут рассмотрены наиболее значимые исследования в предсказании показателя основного обмена человека. Результаты сравнительного анализа точности рассмотренных уравнений описаны в следующей главе.

Уравнения Харриса-Бенедикта (1918)

Первые значимые исследования в области предсказания ПОО были проведены J. A. Harris и F. G. Benedict в начале 20 века. Результаты исследований и разработанные уравнения были опубликованы в 1918-1919 годах [3]. Уравнения Харриса-Бенедикта на протяжении десятилетий оставались основой диетологии и, несмотря на целый ряд проведенных впоследствии исследований, до сих пор крайне часто применяется диетологами по всему миру.

$$BMR = 66,4730 + 13,7516w + 500,3300h - 6,7550a \text{ (для мужчин),}$$

$$BMR = 655,0955 + 9,5634w + 184,9600h - 4,6756a \text{ (для женщин),}$$

где BMR – показатель основного обмена (ккал), w – масса тела (кг), h – рост (м), a – возраст.

Уравнения ВОЗ/ФАО/УООН (1981)

Целью Объединенного консультативного совещания ФАО, ВОЗ и УООН, проходившего в 1981 году, было определение потребностей человека в энергии и белке. В подготовленном докладе, опубликованном в 1985 г. [2], были освещены основные факторы, влияющие на энергетические потребности человека, а также представлены рекомендуемые к использованию уравнения для расчета ПОО.

Данные уравнения были разработаны W.N. Schofield в 1980 году и незначительно скорректированы, используя набор данных большего объема. Уравнения ВОЗ/ФАО/УООН представлены в таблице 2.

Таблица 2: Уравнения ВОЗ/ФАО/УООН (1981)

Возраст	Показатель основного обмена (ккал/день)	
	мужчины	женщины
0–3	$60,9w - 54$	$61,0w - 51$
3–10	$22,7w + 495$	$22,5w + 499$
10–18	$16,6w + 77h + 572$	$7,4w + 482h + 217$
18–30	$15,4w - 27h + 717$	$13,3w + 334h + 35$
30–60	$11,3w + 16h + 901$	$8,7w - 25h + 865$
> 60	$8,8w + 1128h - 1071$	$9,2w + 637h - 302$

w – масса (кг), h – рост (м)

Уравнение Миффлина-Сан Жеор (1990)

Уравнение Миффлина-Сан Жеор – еще одна широко применяемая формула расчета показателя основного обмена.

$$BMR = 9,99w + 625h - 4,92a + 166s - 161,$$

где w – масса тела (кг), h – рост (м), a – возраст, s – пол (= 1 для мужчин, = 0 для женщин) [4].

«Оксфордские» уравнения (2001)

На 2001 год было запланировано очередное заседание консультативно-го совещания ФАО/ВОЗ/УООН по энергетическим потребностям человека. В рамках подготовки к совещанию, ФАО поручила доктору С. J. К. Генри подготовить работу [1], суммирующую результаты исследований, направленных на оценку основного обмена. В ходе работы им был собран набор данных, включающий 10552 измерения из 166 научных публикаций, на основе которого впоследствии были разработаны так называемые «оксфордские» уравнения (таблица 3).

Консенсусом заседания стало «отсутствие необходимости» перехода от установленных в 1981 рекомендуемых уравнений к уравнениям, выверенным Генри.

Таблица 3: «Оксфордские» уравнения

Возраст	Показатель основного обмена (ккал/день)	
	мужчины	женщины
0–3	$28,2w + 859h - 371$	$30,4w + 703h - 287$
3–10	$15,1w + 74,2h + 306$	$15,9w + 210h + 349$
10–18	$15,6w + 266h + 299$	$9,4w + 249h + 462$
18–30	$14,4w + 313h + 113$	$10,4w + 615h - 282$
30–60	$11,4w + 541h - 137$	$8,18w + 502h - 11,6$
> 60	$11,4w + 541h - 256$	$8,52w + 421h + 10,7$

w – масса (кг), h – рост (м)

Глава 3. Создание нейронной сети

Искусственные нейронные сети – математические модели, основанные на принципах работы человеческого мозга. Они состоят из нейронов – вычислительных единиц, взаимодействующих друг с другом.

3.1 Структура нейронной сети

Многослойная нейронная сеть состоит из слоя входных нейронов, скрытых слоев и выходного нейрона. Входные нейроны – физиологические параметры, выходной нейрон – результат предсказания ПОО, а нейроны скрытых слоев вычисляют промежуточные результаты и передают их на нейроны следующего слоя. Структура многослойной нейронной сети представлена на рисунке 1.

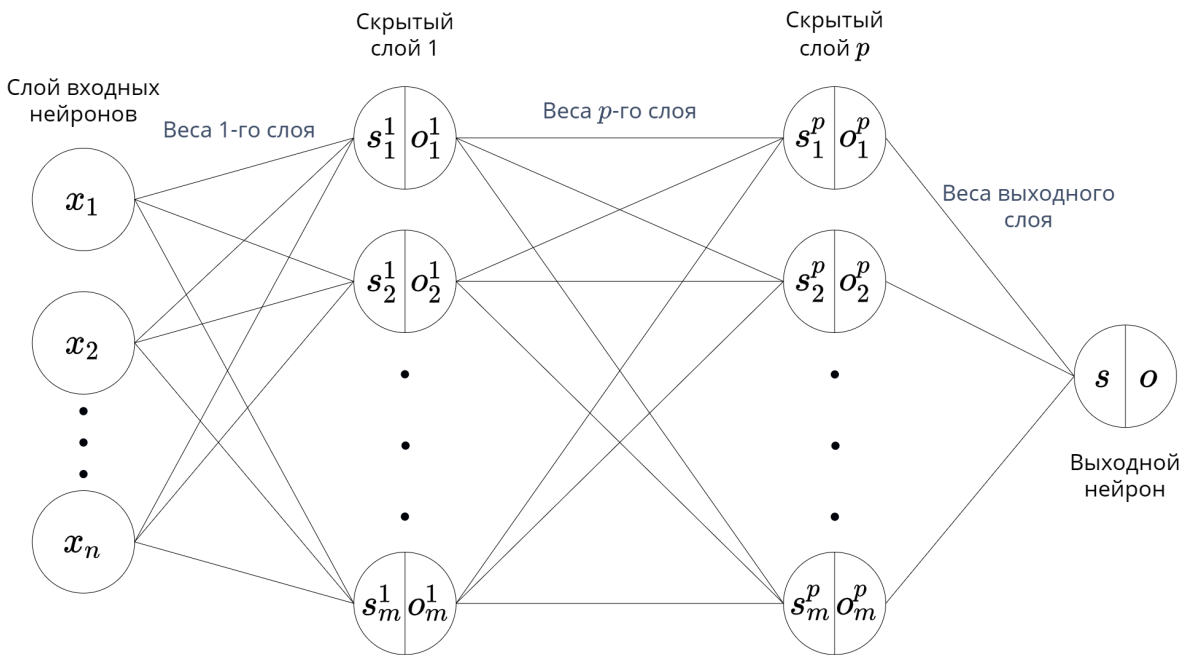


Рис. 1: Структура многослойной нейронной сети.

На вход каждому нейрону k -го слоя подаются выходные значения нейронов $k - 1$ слоя, из которых формируется взвешенная сумма (формула 1).

$$s_j^k = \sum_{i=1}^m o_i^{k-1} w_{i,j}^k, \tag{1}$$

где s_j^k – значение взвешенной суммы для j -го нейрона k -го слоя, o_i^{k-1} – выходное значение i -го нейрона $(k - 1)$ -го слоя, $w_{i,j}^k$ – вес, соответствующий связи между i -м нейроном $(k - 1)$ -го слоя и j -м нейроном k -го слоя.

Выходное значение нейрона рассчитывается как значение функции активации от взвешенной суммы. Выбор функции активации зависит от решаемой задачи, в данной сети была использована сигмоида:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Таким образом,

$$o_j^k = \sigma(s_j^k)$$

Цель нейронной сети в процессе обучения – минимизировать функцию ошибки, в нашей задаче имеющую вид:

$$L = (o - y)^2,$$

где o – значение выходного нейрона, y – фактическое значение ПОО.

3.2 Сравнительный анализ методов оценки ПОО

В качестве метрик для оценки точности предсказания с помощью различных методов были выбраны: среднеквадратичная ошибка, средняя абсолютная ошибка и квадрат коэффициента корреляции.

Среднеквадратичная ошибка (mean squared error, MSE) – усредненный квадрат разности реальных и предсказанных значений:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_{pred}^i - y_{real}^i)^2$$

Средняя абсолютная ошибка (mean absolute error, MAE) – усредненный модуль разности реальных и предсказанных значений:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_{pred}^i - y_{real}^i|$$

Коэффициент корреляции Пирсона (Pearson correlation coefficient) – статистика, показывающая степень линейной связи между двумя величинами (формула 2). Квадрат коэффициента корреляции r^2 – метрика, используемая в рассмотренных работах для оценки точности разработанных уравнений.

$$r = \frac{\sum_{i=1}^n (y_{pred}^i - \bar{y}_{pred})(y_{real}^i - \bar{y}_{real})}{\sqrt{\sum_{i=1}^n (y_{pred}^i - \bar{y}_{pred})^2 \sum_{i=1}^n (y_{real}^i - \bar{y}_{real})^2}} \quad (2)$$

$$\bar{y}_{pred} = \frac{1}{n} \sum_{i=1}^n y_{pred}^i, \bar{y}_{real} = \frac{1}{n} \sum_{i=1}^n y_{real}^i$$

Для оценки точности предсказания нейронной сети был использован метод скользящего контроля (cross-validation). Скользящий контроль заключается в разделении данных на k непересекающихся равных частей случайным образом, $k - 1$ из которых служат в качестве тренировочной выборки, а оставшаяся – в качестве тестовой. Процедура производится k раз, в результате чего каждая из частей используется для тестирования. В данной работе k положено равным 5. Метрики вычисляются как среднее арифметическое значений метрик на каждом этапе.

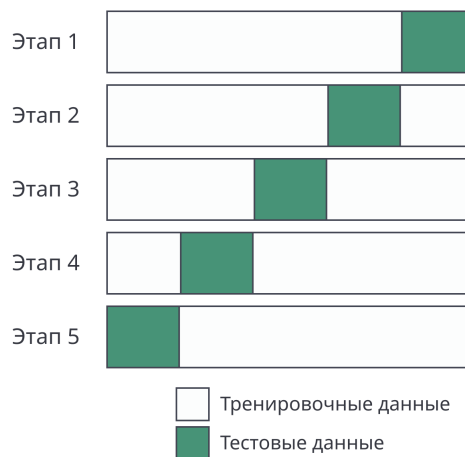


Рис. 2: Скользящий контроль

Результаты оценки точности предсказания показателя основного обмена с помощью уравнений, применяемых в диетологии, и многослойной нейронной сети представлены в таблице 4.

Таблица 4: Метрики методов предсказания ПОО

Метрики Методы	MSE	MAE	r^2
Уравнения Харриса-Бенедикта	0,5087	0,5540	0,7595
Уравнения ВОЗ/ФАО/ЦООН	0,4460	0,5028	0,7951
Уравнения Миффлина-Сан Жеор	0,4949	0,5502	0,7790
«Оксфордские» уравнения	0,4162	0,4805	0,8050
Нейронная сеть	0,3651	0,4533	0,8251

Как можно видеть, нейронная сеть показала результат, превосходящий все уравнения из рассмотренных. Значения метрик для нейронной сети: MSE – 0,3651, MAE – 0,4533, r^2 – 0,8251. Графики реальных и предсказанных значений представлены в приложении.

Глава 4. Поиск оптимального плана питания

Задачу нахождения оптимального плана питания целесообразно представить в виде задачи линейного программирования.

Линейное программирование (ЛП, линейная оптимизация) – частный случай более широкой дисциплины – математического программирования. Задача линейного программирования – нахождение минимума (максимума) линейной целевой функции при действии линейных ограничений в виде равенств (неравенств). Общая задача линейной минимизации может быть записана в следующем виде:

$$\min z = \min \sum_{i=1}^n c_i x_i$$
$$\begin{cases} \sum_{i=1}^n a_{ij} x_i \geq b_j, & j = \overline{1, m}, \\ x_i \geq 0, & i = \overline{1, n}, \end{cases}$$

где

$x_i, i = \overline{1, n}$ – переменные задачи,

$z = \sum_{i=1}^n c_i x_i$ – целевая функция,

$\sum_{i=1}^n a_{ij} x_i \geq b_j$ – ограничения задачи ($j = \overline{1, m}$).

Вектор $X^* = (x_1^*, \dots, x_n^*)$, при котором достигается минимум (максимум) целевой функции и соблюдаются ограничения задачи, называется оптимальным решением задачи линейного программирования.

Так как для создания плана питания необходимо учитывать количество приемов пищи и количество продуктов на каждый прием пищи, целесообразно ввести логические переменные (logical variables, binary variables), принимающие значения 0 и 1, и логические ограничения (logical constraints, indicator constraints). Таким образом, задачу следует сформулировать как задачу смешанного целочисленного программирования (mixed-integer programming).

Задача ЛП преобразуется в задачу смешанного целочисленного программирования, если на некоторые из переменных задачи накладывается

условие целочисленности. Таким образом исходная задача принимает следующий вид:

$$\min z = \min \sum_{i=1}^n c_i x_i$$

$$\begin{cases} \sum_{i=1}^n a_{ij} x_i \geq b_j, & j = \overline{1, m}, \\ x_i \geq 0, & i = \overline{1, n}, \\ x_j \in \mathbb{Z}, & j = \overline{1, p}, p < n \end{cases}$$

Определим целевую функцию, переменные и ограничения задачи поиска оптимального плана питания.

Введем постоянные задачи. Пусть E – энергетические потребности, \underline{F} , \overline{F} – нижняя и верхняя границы нормы потребления жира, \underline{C} , \overline{C} – границы нормы потребления углеводов, \underline{P} , \overline{P} – границы нормы потребления белка. Количество порций положим равным Q , минимальный и максимальный размеры порции – \underline{S} и \overline{S} .

Пусть для пользователя, согласно его предпочтениям, выбрано n допустимых продуктов. Тогда пусть x_i ($i = \overline{1, n}$) – переменные задачи (масса i -го продукта в плане питания), а e_i , f_i , c_i , p_i – энергетическая ценность, содержание жиров, углеводов и белков в продукте соответственно.

Задачу представим как задачу минимизации, где в качестве целевой функции выступает разность энергетических потребностей и суммарной энергетической ценности продуктов.

$$\min z = \min \left(\sum_{i=1}^n e_i x_i - E \right)$$

$$\sum_{i=1}^n e_i x_i \geq E$$

Ограничения потребления питательных веществ приобретают следующий вид:

$$\begin{aligned} \sum_{i=1}^n f_i x_i &\geq \underline{F}, & \sum_{i=1}^n f_i x_i &\leq \overline{F} \\ \sum_{i=1}^n c_i x_i &\geq \underline{C}, & \sum_{i=1}^n c_i x_i &\leq \overline{C} \\ \sum_{i=1}^n p_i x_i &\geq \underline{P}, & \sum_{i=1}^n p_i x_i &\leq \overline{P} \end{aligned}$$

Введем логические переменные для ограничения количества отбираемых продуктов. Пусть каждой переменной x_i соответствует b_i – целочисленная переменная, принимающая значения 0 и 1. Логические ограничения позволяют придавать вес только тем переменным x_i , для которых b_i принимает значение 1. Для нашей задачи логические ограничения будут выглядеть следующим образом:

$$\begin{aligned} x_i - b_i \underline{S} &\geq 0, & i &= \overline{1, n} \\ x_i - b_i \overline{S} &\leq 0, & i &= \overline{1, n} \\ b_i &\in \{0, 1\} \end{aligned}$$

Из уравнения 0 следует, что при $b_i = 0$ переменная x_i принимает значение 0, а при $b_i = 1$ ограничивается снизу и сверху значениями минимального и максимального размера порции (\underline{S} и \overline{S}). Для установки количества отбираемых продуктов достаточно ввести дополнительное ограничение:

$$\sum_{i=1}^n b_i = Q$$

Глава 5. Разработка веб-приложения

5.1 Используемые технологии

Python + Django

Выбор Python в качестве языка программирования обусловлен большим количеством инструментов для решения задач многих областей, в частности анализа данных и машинного обучения. В данной работе для моделирования, подбора параметров и обучения нейронной сети были использованы библиотеки NumPy, Pandas и Keras, а для решения задачи линейного программирования – библиотека PuLP.

Для более простого и эффективного процесса разработки часто прибегают к использованию фреймворков. Два самых популярных веб-фреймворка на языке Python: Flask и Django. В ходе разработки был использован Django в силу ряда преимуществ:

- Подробная и обширная документация
- Встроенная ORM (object-relational mapping) – технология, связующая реляционные базы данных с объектами языка программирования
- Встроенная система аутентификации
- Автоматическое создание каркаса приложения, способствующее более простому написанию структурированного кода.

PostgreSQL

В качестве системы управления базами данных была выбрана PostgreSQL – бесплатная объектно-реляционная СУБД с открытым исходным кодом. PostgreSQL поддерживает практически все возможности стандарта SQL:2016, крайне надежна и легко расширяема. Благодаря большому сообществу, она имеет подробную документацию и регулярно обновляется.

React

Для удобства разработки фронтэнд части приложения была использована JavaScript библиотека React. React предоставляет инструменты для удобного разбиения кода на компоненты – инкапсулированные элементы, которые легко повторно использовать, передавая разные параметры. Также она контролирует рендеринг страницы, обновляя только те компоненты, состояние которых поменялось.

5.2 Архитектура приложения

После определения требуемого функционала приложения необходимо приступить к проектированию его архитектуры – важнейшему этапу разработки. Структуру пользовательских приложений удобно представлять в виде схемы «Модель-Представление-Контроллер» (Model-View-Controller, MVC). Согласно данному принципу, приложение можно разделить на три связанные составляющие:

- **«Модель»** – данные приложения
- **«Представление»** – компонент, отвечающий за визуализацию данных модели пользователю
- **«Контроллер»** – звено, обрабатывающее запросы пользователя и вносящее изменения в «Модель».

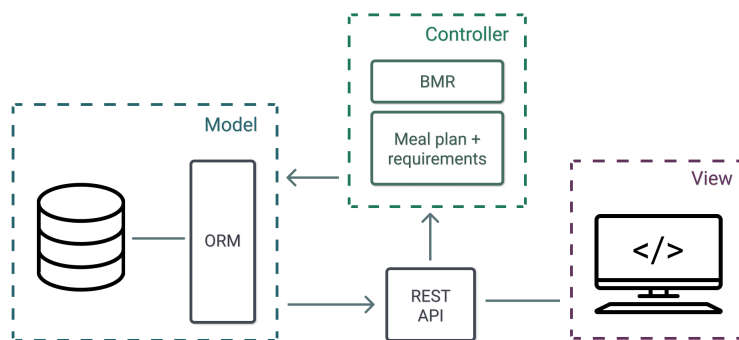


Рис. 3: Архитектура приложения.

Архитектура приложения изображена на рисунке 3. «Контроллер» разделен на два основных модуля:

- **BMR**, по запросу пользователя предсказывающий показатель основного обмена с помощью нейронной сети (или одного из уравнений)
- **Meal plan + requirements**, производящий остальные вычисления, в том числе расчет расхода энергии и рекомендуемых норм потребления питательных веществ, поиск оптимального плана питания.

5.3 Структура базы данных

База данных разработанного приложения состоит из 9 отношений, которые можно разделить на 3 категории: «пользователи», «продукты» и «планы питания». Схема базы данных изображена на рисунке 4.

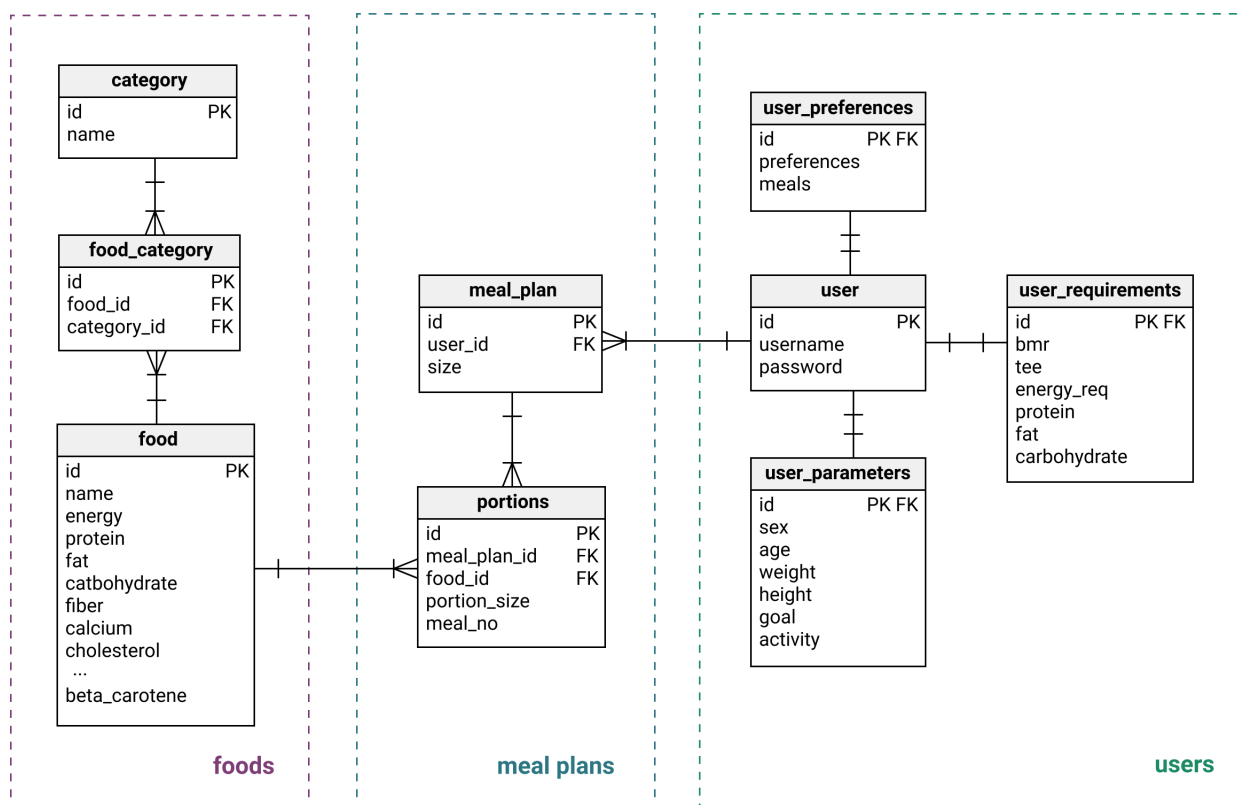


Рис. 4: Структура базы данных.

К «пользовательским» относятся 4 отношения: **user**, **user_parameters**, **user_preferences**, **user_requirements**. Отношение **user** связано с остальными

связью 1:1 с целью разделения атрибутов, относящихся к пользователям, на таблицы в соответствии с выполняемой функцией, что обусловлено особенностями Django ORM. Опишем «пользовательские» отношения и их атрибуты:

- **user** – таблица пользователей, основана на модели пользователя Django
 - username – имя пользователя
 - password – хешированный пароль пользователя
- **user_parameters** – параметры, необходимые для расчета энергетических потребностей
 - sex – пол
 - age – возраст
 - weight – масса тела (кг)
 - height – рост (м)
 - goal – цель. Может принимать 3 значения: «maintain» – поддержание текущей массы тела, «lose» – снижение массы, «gain» – набор массы
 - activity – уровень физической активности (PAL)
- **user_preferences** – предпочтения пользователя (необходимы для составления плана питания)
 - preferences – предпочтения в питании (вегетарианство, низкокалорийная диета, халяльная пища и т.д.)
 - meals – количество приемов пищи в день (допустимые значения: от 3 до 7)
- **user_requirements** – энергетические затраты и рекомендации потребления питательных веществ
 - bmr – показатель основного обмена пользователя (МДж/день)
 - tee – общие энергетические затраты пользователя (МДж/день)
 - energy_req – энергетические потребности (МДж/день)
 - protein – потребность в белке (г/день)
 - fat – потребность в жирах (г/день)

- carbohydrate – потребность в углеводах (г/день)

Отношения категории «продукты» хранят информацию о продуктах, их пищевой ценности и категориях. Атрибуты таблицы **food** дублируют структуру базы продуктов, описанную в главе 1, за исключением категорий продуктов. Список категорий продуктов содержится в таблице **category**, имеющей связь «многие-ко-многим» (many-to-many) с **food**, где в качестве связующей таблицы выступает **food_category**.

Информация о сгенерированных планах питания представлена в таблицах **meal_plan** и **portions**. Рассмотрим атрибуты этих отношений:

- **meal_plan** – отношение, содержащее планы питания
 - user_id – внешний ключ на таблицу user (пользователь, для которого создан план)
 - size – число порций в плане питания
- **portions** – отношение, содержащее отдельные порции
 - meal_plan_id – внешний ключ на таблицу meal_plan (план питания, к которому принадлежит порция)
 - food_id – внешний ключ на таблицу food (продукт, из которого состоит порция)
 - portion_size – размер порции
 - meal_no – порядковый номер приема пищи, к которому относится порция

5.4 Представления приложения

Приложение с точки зрения пользовательского интерфейса можно разделить на 3 представления (views): «Модели», «Параметры», «План питания», каждое из которых выполняет определенную задачу.

Страница «**Модели**» – вспомогательное представление. Оно дает возможность пользователю сравнить рассмотренные в данной работе методы предсказания показателя основного обмена. Вкладка каждой модели содержит:

- Название модели и ее краткое описание
- График, отображающий фактические и предсказанные значения модели на тестовой выборке
- Статистика: объем тестовой выборки, квадрат коэффициента корреляции между фактическими и предсказанными значениями, средняя абсолютная ошибка, среднеквадратичная ошибка
- Форма ввода параметров (рост, масса, пол, возраст)
- Окно результатов предсказания (доступно после отправки формы)

После введения физиологических параметров и отправки формы на сервер, пользователь получает результат предсказания ПОО, вычисленный с помощью выбранного метода.

Nutrition helper | Модели | Параметры | План питания | test1

Физиологические параметры

возраст: 35 | масса (кг): 71 | рост (м): 1.72 | пол: М Ж

Физическая активность & цель

Физическая активность: Средняя активность
Люди профессий, не предполагающих изматывающий физический труд, но требующих большого энергетического затрат.

Цель: Поддержание веса
Сохранить текущую массу тела

Предпочтения

Предпочтения в питании: веганство вегетарианство халяль без глютена низколактозная диета

Количество приемов пищи: 4 приема пищи
План: 4 приема пищи в день

Потребности и рекомендации

ПОО: 7.028 МДж, 1679 ккал

энергетические затраты: 12.299 МДж, 2938 ккал

энергетические потребности: 12.299 МДж, 2938 ккал

Белки	145g	Витамин А	1 мг
Жиры	83g	Витамин В1	2 мг
Углеводы	398g	Витамин В2	2 мг
Питательные вещества		Витамин В3	20 мг
Пищ. волокна	30g	Витамин В6	1 мг
НЖК	< 25g	Витамин В9	400 µg
ПНЖК	> 11g	Витамин В12	2 мг
Холестерин	< 300mg	Витамин С	70 мг
Натрий	< 2400mg	Витамин D	15 µg
Калий	3500mg	Витамин Е	10 мг
Кальций	1000mg	Витамин К	120 µg
Фосфор	1000mg		
Йод	150 µg		
Магний	400mg		
Железо	14mg		
Цинк	10mg		

Рис. 5: Страница «Параметры».

Страница «Параметры» – представление приложения, отвечающее за сохранение пользовательских параметров и предпочтений. Представление состоит из 3 секций: форма физиологических параметров, форма предпочтений и окно с информацией об энергетических потребностях и рекомендуемых нормах потребления питательных веществ. Для составления плана питания пользователю необходимо ввести следующие данные:

- Физиологические параметры (рост, масса тела, пол, возраст)
- Уровень физической активности
- Цель (поддержание текущей массы тела, снижение массы, набор массы)

Также пользователь имеет возможность выбрать различные предпочтения в питании и указать количество приемов пищи в день.

При сохранении параметров пользователем приложение вычисляет показатель основного обмена, используя нейронную сеть, рассчитывает энергетические затраты и потребности согласно указанному уровню физической активности и цели, отображает вычисленные показатели наряду с рекомендуемыми нормами потребления основных питательных веществ.

Назначение страницы «План питания» заключается в отображении подобранного в соответствии с рассчитанными потребностями пользователя рациона питания. Страница содержит информацию о суммарной пищевой ценности рациона и таблицу продуктов, разделенную на приемы пищи, со следующими данными: размер порции, содержание белков, жиров, углеводов, энергетическая ценность.

Nutrition helper | Модели | Параметры | План питания | test1

Пищевая ценность

Масса 1857г
 Белки 131г
 Жиры 87г
 Углеводы 320г
 Энергетическая ценность . . . 2650ккал
 Вода 1261мл
 Насыщенные жирные кислоты . . . 34г
 Mono- и дисахариды 100г
 Крахмал 219г
 Пищевые волокна 23г
 Натрий 2190мг
 Калий 2237мг
 Кальций 1709мг
 Магний 367мг
 Фосфор 1938мг
 Железо 13мг
 Витамин В1 1мг
 Витамин В2 2мг
 Витамин РР 24мг
 Бета-каротин 1279µг
 Витамин С 79мг
 Холестерин 339мг
 Витамин А 341µг

План питания

Нажмите сюда, чтобы скачать план питания

Завтрак

Продукт	Порция	Б	Ж	У	Энергия
Брусника	327.9	2.3	1.6	26.9	150.8
Варенье из клубники	50.0	0.1	0.1	37.0	142.5
Кефир 3,2 % жирности	200.0	5.8	6.4	8.0	118.0
Клюква	50.0	0.3	0.1	1.9	14.0
Итого	627.9	8.5	8.2	73.7	425.3

Обед

Крупа рисовая	298.9	20.9	3.0	221.2	995.5
Ряженка 1,0 % жирности	200.0	6.0	2.0	8.4	80.0
Смородина красная	50.0	0.3	0.1	3.9	21.5
Сыр Голландский (брусокый)	50.0	13.2	13.3	0.0	175.0
Итого	598.9	40.4	18.4	233.5	1272.0

Ужин

Сыр Российский	50.0	11.6	14.8	0.0	182.0
----------------	------	------	------	-----	-------

Рис. 6: Страница «План питания».

Заключение

В результате проделанной работы была разработана нейронная сеть, показавшая большую точность предсказания показателя основного обмена, чем наиболее часто применяемые в диетологии уравнения. С учетом рекомендаций ВОЗ и других компетентных медицинских организаций, на основе нейронной сети было разработано веб-приложение, генерирующие план питания на день в соответствии с индивидуальными потребностями и предпочтениями человека. Исходный код приложения доступен по ссылке: <https://github.com/zisest/bachelor-paper-code/>. С демо-версией можно ознакомиться на <https://nutrition.zisest.ru/>. В дальнейшей работе предполагается совершенствование разработанной системы, которое возможно разбить на следующие этапы:

1. Расширение базы продуктов. Объем использованной в работе базы данных ограничивает разнообразие генерируемых планов питания, поэтому поиск базы большего размера – приоритетная задача.
2. Создание подсистемы в веб-приложении для расчета уровня физической активности, что позволит более точно оценивать энергетические затраты человека.
3. Расширение функционала веб-приложения, например: добавление возможности сохранения нескольких профилей одним пользователям (для врачей, у которых несколько пациентов).

Список литературы

- [1] Henry C.J.K. Basal metabolic rate studies in humans: Measurement and development of new equations // *Public Health Nutrition*. 2005. Vol. 8(7a). P. 1133–1152
- [2] Joint FAO/WHO/UNU Expert Consultation on Energy and Protein Requirements (Rome, 5–17 Oct. 1981): report / World Health Organization. 1985. 206 p.
- [3] Harris J.A., Benedict F.G. A Biometric Study of Human Basal Metabolism // *Proceedings of the National Academy of Sciences*. 1918. Vol. 4, N. 12. P. 370–373.
- [4] Mifflin M.D., St Jeor S.T. et al. A new predictive equation for resting energy expenditure in healthy individuals // *The American Journal of Clinical Nutrition*. 1990. Vol. 51, N. 2. P. 241–247.
- [5] Joint FAO/WHO/UNU Expert Consultation on Human Energy Requirements (Rome, 17–24 Oct. 2001): report / Food and Agriculture Organization of the United Nations. 2004. 107 p.
- [6] Vitamin and mineral requirements. 2nd ed. / World Health Organization, Food and Agriculture Organization of the United Nations. 2004. 341 p.
- [7] Химический состав российских пищевых продуктов: Справочник / под ред. И.М. Скурихина, В.А. Тутельяна. М.: ДеЛи принт, 2002. 236 с.
- [8] Bishop C.M. *Pattern Recognition and Machine Learning*. New York: Springer-Verlag, 2006. 738 p.
- [9] Bradley S.P., Нax A.C., Magnati T.L. *Applied Mathematical Programming*. Reading: Addison-Wesley, 1977. 716 p.
- [10] Official Django documentation [Электронный ресурс]. URL: <https://docs.djangoproject.com/> (дата обращения: 10.04.2020).

- [11] Official Keras documentation [Электронный ресурс]. URL: <https://keras.io/> (дата обращения: 15.04.2020).
- [12] Official PuLP documentation [Электронный ресурс]. URL: <https://coin-or.github.io/pulp/> (дата обращения: 15.05.2020).
- [13] Official React documentation [Электронный ресурс]. URL: <https://reactjs.org/docs/> (дата обращения: 10.04.2020).

Приложение

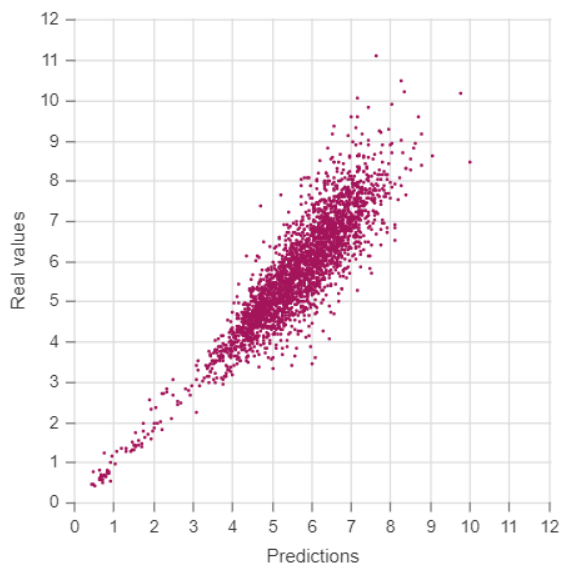


Рис. 7: «Оксфордские» уравнения.

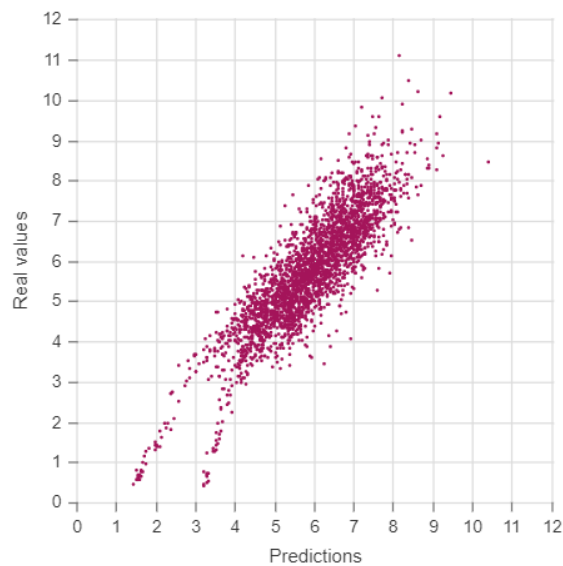


Рис. 8: Уравнения Харриса-Бенедикта.

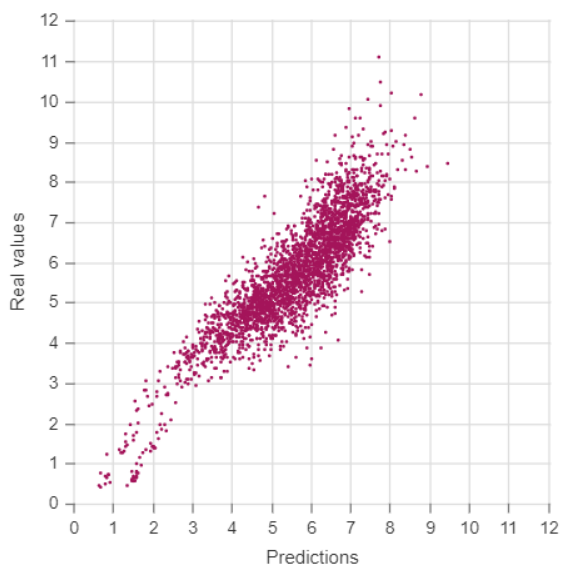


Рис. 9: Уравнение Миффлина-Сан Жеор.

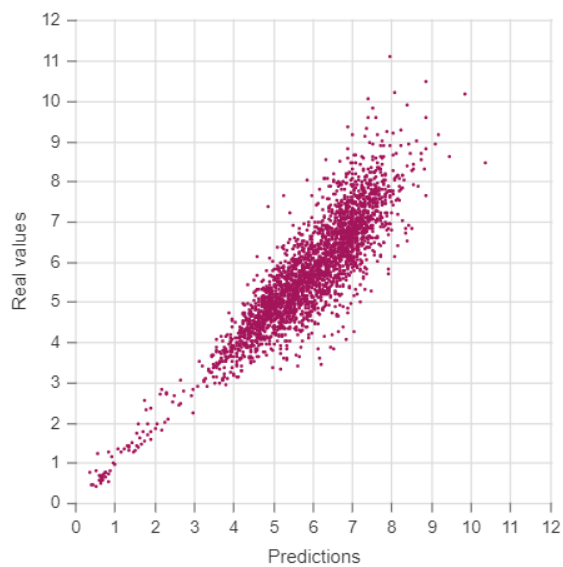


Рис. 10: Уравнения ВОЗ/ФАО/УООН.

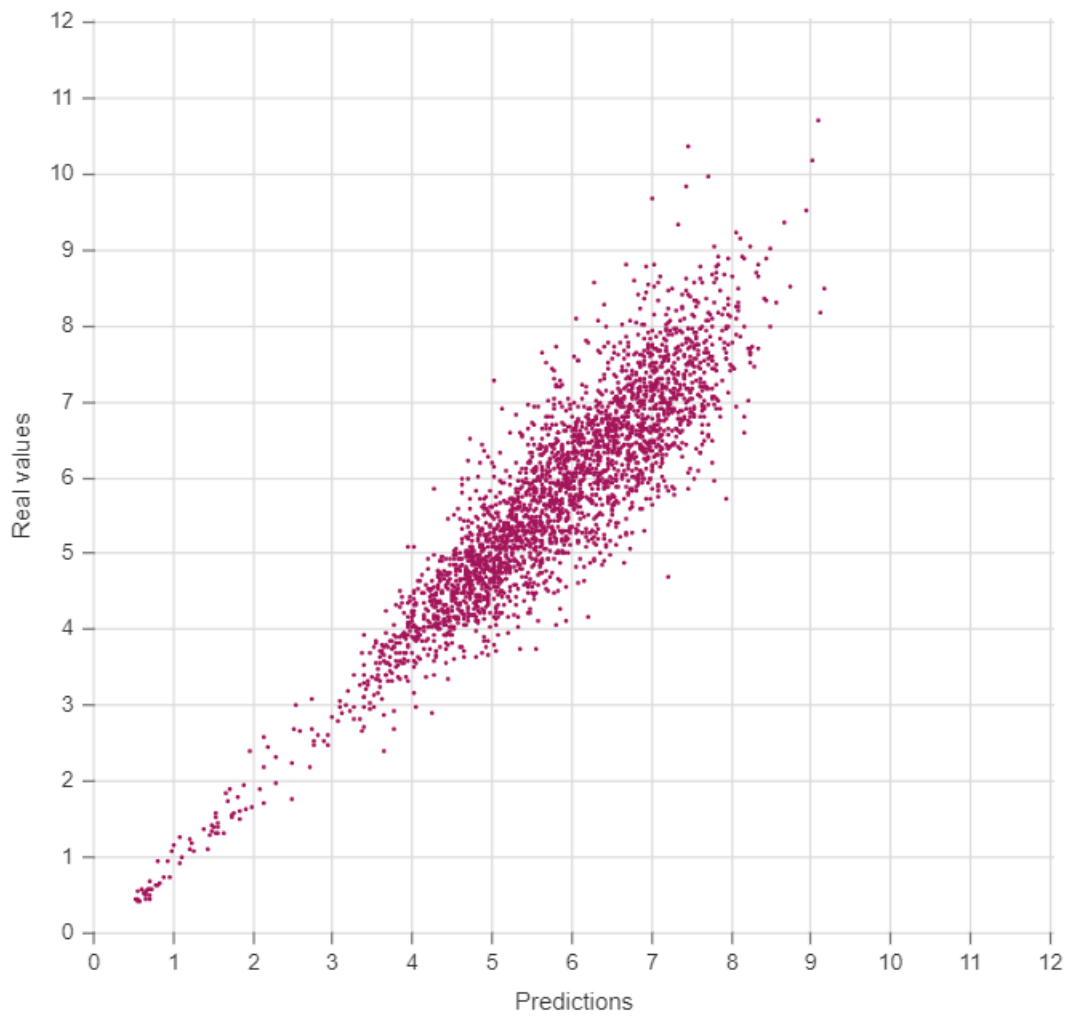


Рис. 11: Многослойная нейронная сеть.