# **Обучение нейросетевых регрессора и классификаторов**

**Цель** работы – получить навыки создания полносвязного бинарного и многоклассового классификаторов, а также нейросетевого регрессора с помощью библиотеки построения нейронных сетей.

**Задачи** – Построить три нейросетевые модели:

a) Бинарный Классификатор

b) Многоклассовый классификатор

c) Регрессор

Варианты (выборки):

Для своего варианта необходимо посмотреть последнюю цифру номера своей зачетной книжки (или студенческого билета):

• если последняя цифра – число четное:

1. Оценка вероятности, того, что клиент откроет банковский депозит в результате маркетинговой акции: <https://archive.ics.uci.edu/ml/datasets/Bank+Marketing> Класс: атрибут 21 - y - has the client subscribed a term deposit? (binary: 'yes', 'no').
2. Оценка здоровья внутриутробного развития плода: <https://www.kaggle.com/datasets/andrewmvd/fetal-health-classification> Метка класса: fetal\_health.
3. Оценка уровня потребления ресурсов домохозяйством. Набор данных DS\_2019\_public.csv. Описание атрибутов – файл recs2009\_public\_codebook.xlsx. Предсказываемый признак – один из атрибутов (использовать только один из): TOTALBTU, TOTALBTUSPH, TOTALBTUCOL, TOTALBTUWTH, TOTALBTURFG, TOTALBTUOTH (Примечание BTU – Британская тепловая единица BTU ≈ 252 калории).

• если последняя цифра – число нечетное:

1. Оценка вероятности диагностики диабета у человека <https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset> Класс: Diabetes\_012. Класс отрицательный – 0 (no diabetes – нет диабета), класс положительный – 1 & 2 (prediabetes – преддиабетическое состояние & diabetes - диабет).
2. Оценка уровня физического развития людей разного возраста: <https://www.kaggle.com/datasets/kukuroo3/body-performance-data> Метка класса: class.
3. Оценка расходов домохозяйства на один из видов потребляемых ресурсов. Набор данных DS\_2019\_public.csv. Описание атрибутов – файл recs2009\_public\_codebook.xlsx. Предсказываемый признак – один из атрибутов (использовать только один из): TOTALDOL, TOTALDOLSPH, TOTALDOLCOL, TOTALDOLWTH, TOTALDOLOTH.

Выполнить загрузку и предварительную обработку данных из наборов.

Разделить каждую выборку на обучающую, тестовую и валидационную. Произвести обучение набора нейросетевых архитектур, отличающихся разным набором параметров: число слоёв, количество нейронов в слоях, функции активации в слоях, процедур оптимизации:

* Подобрать архитектуры нейронных сетей, которые с одной стороны позволяют получить модели с лучшими метриками качества работы, с другой стороны не являются избыточными и не переобученными.
* Вычислить следующие метрики работы:

Для бинарного классификатора: Recall, Precision, Weighted Accuracy, AUC для всех исследованных моделей.

Для многоклассового классификатора: Recall, Precision, Weighted Accuracy, AUC для всех классов всех исследованных моделей. Вывести ROC-кривые для каждого класса в лучшем классификаторе.

Для регрессора: MSE, MAE, R2 для всех полученных моделей.

Сделать выводы по результатам построения моделей.

**Описание отчета.**

1. Описание наборов данных.

2. Признаки, которые были использованы для анализа.

2. Параметры архитектур и обучения нейронных сетей, использованные для обучения.

3. Графики обучения для архитектур нейронных сетей с лучшими характеристиками эффективности

5. ROC-кривые классов для лучших классификаторов.

6. Оценки моделей на тестовых выборках в виде таблиц/ диаграмм, отображающих метрики качества.

7. Программный код.