

## Контрольная работа 2. Обработка текстов

### 1. Подготовка базовых моделей (40%)

**Вариант 1.** Используя [шаблон ноутбука](#) [1], напишите четыре нейронные сети (НС) для распознавания писателей. Для этого выполните следующее:

1. Загрузите данные и разделите на слова.
2. Создайте словарь и превратите данные в индексы.
3. Превратите данные в обучающую выборку.
4. Напишите НС.

Четыре НС, которые нужно написать:

1. Embedding + Dense сеть.
2. Embedding + LSTM сеть.
3. Embedding + Conv1D сеть.
4. Embedding + сложная сеть из Dense, LSTM и Conv1D.

1 ссылка на ноутбук шаблона [https://github.com/Irina-64/Deep-Learning-Algorithms/blob/main/%D0%9B%D0%B5%D0%BA%D1%86%D0%B8%D1%8F%20%D0%BA%D1%83%D1%80%D1%81%D0%90%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC%D1%8B%D0%B3%D0%BB%D1%83%D0%B1%D0%BE%D0%BA%D0%BE%D0%B3%D0%BE%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D1%8F\(2022\).ipynb](https://github.com/Irina-64/Deep-Learning-Algorithms/blob/main/%D0%9B%D0%B5%D0%BA%D1%86%D0%B8%D1%8F%20%D0%BA%D1%83%D1%80%D1%81%D0%90%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC%D1%8B%D0%B3%D0%BB%D1%83%D0%B1%D0%BE%D0%BA%D0%BE%D0%B3%D0%BE%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D1%8F(2022).ipynb)

### Вариант 2

Используя [шаблон ноутбука](#) [1], напишите классификацию заболеваний по базе симптомов, используя любые подходы, связанные с обработкой текстов.

1. Embedding + LSTM сеть.
2. Embedding + Conv1D сеть.
3. Embedding + сложная сеть из Dense, LSTM и Conv1D.

Добейтесь максимальной точности распознавания.

Используйте xLen = 50.

### База данных “болезни”

[https://drive.google.com/drive/folders/1HFijLcwKFUOA31\\_HtvTFq87vINqIM4Ce?usp=share\\_link](https://drive.google.com/drive/folders/1HFijLcwKFUOA31_HtvTFq87vINqIM4Ce?usp=share_link)

## Контрольное задание (60%)

Сделайте оценку того, как блоки слов в словаре влияют на точность классификации (на любой базе, например, распознавание писателей). Для этого выполните следующее:

- Сформируйте словарь.
- Используйте первые 50000 слов.
- Разбейте все слова на блоки по 100 слов последовательно.
- Поочередно отключайте блоки по одному, то есть сделайте так, чтобы эти слова не использовались при обучении сети, например, их индексы всегда приравнивались к 0.
- Оцените, как меняется точность классификации на проверочной выборке.
- Сделайте выводы, какие блоки слов влияют в плюс, а какие в минус.
- Соберите итоговую подборку слов, в которую войдут только те блоки по 100 слов, которые улучшили точность (исключение которых давало падение точности).
- Обучите сеть только на этих блоках слов.
- Оцените изменение точности в сравнении с обучением на полном словаре (первые 50000 слов).