МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

им. Н.Э. Баумана

Кафедра «Систем обработки информации и управления»

ОТЧЕТ

**Лабораторная работа №6**

по курсу«Проектирование интеллектуальных систем»

ИСПОЛНИТЕЛЬ:         Арбузов А.П.

группа ИУ5-24М

        \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

подпись

"\_\_" \_\_\_\_\_\_\_\_ 2021 г.

ПРЕПОДАВАТЕЛЬ:       Терехов В.И.

 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

подпись

"\_\_"\_\_\_\_\_\_\_\_\_202\_ г.

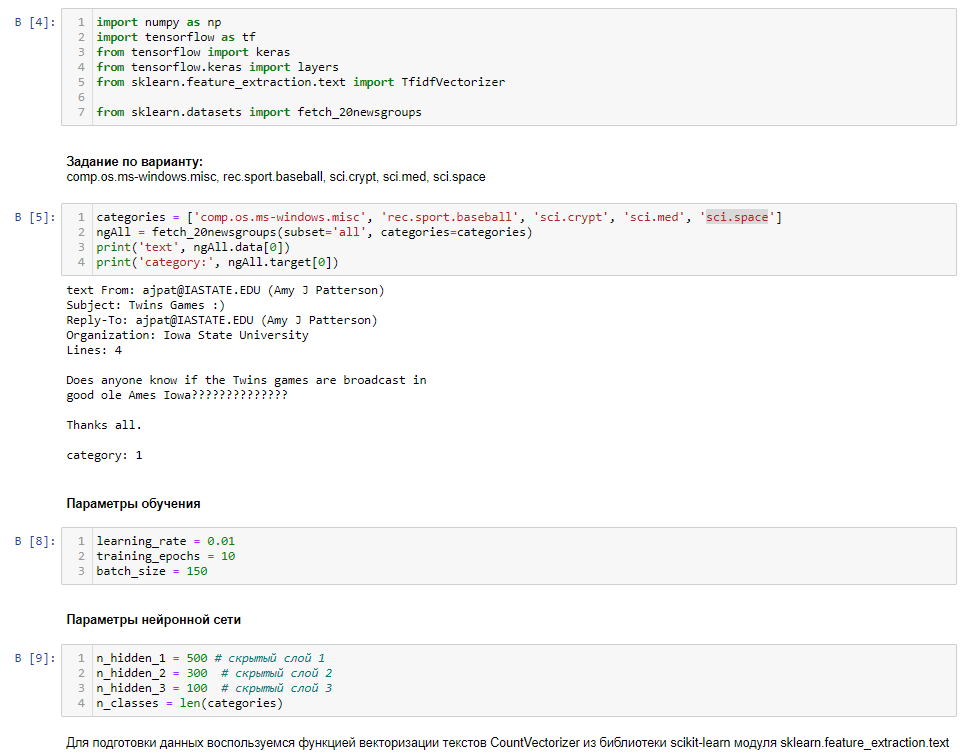
Москва  - 2021

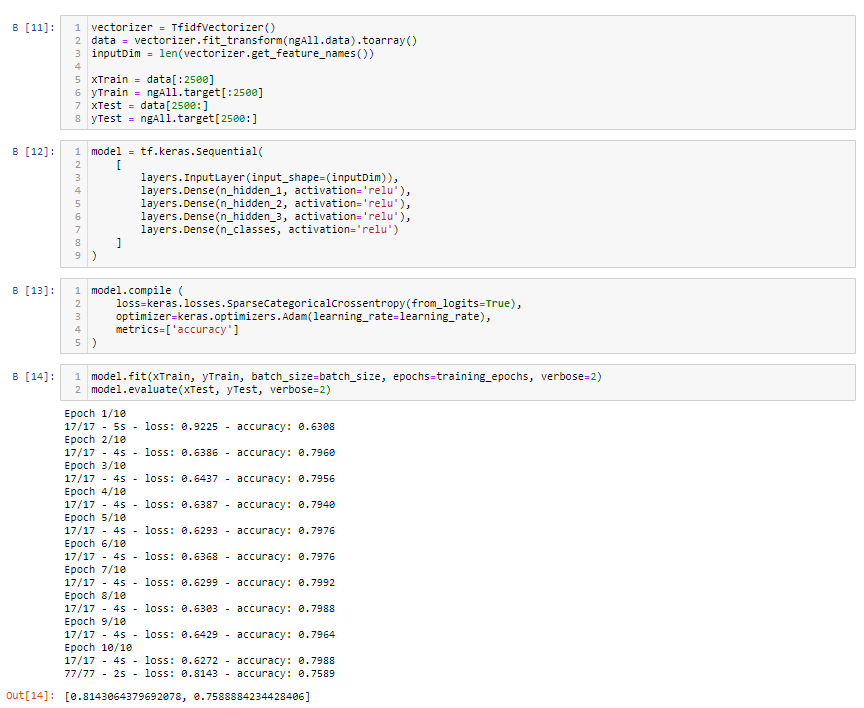
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. **Цель работы**

Необходимо увеличить количество скрытых слоев до 3-ех, а количество нейронов в этих слоях так, чтобы обеспечить точность работы нейросети не менее 75%. Темы текстов необходимо изменить в соответствии с вариантом

**2. Код программы**





**3. Контрольные вопросы**

1. **Какие вы знаете задачи обработки текстов, в чем они заключаются?**
   * Классификация текстов и предложений
   * Машинный перевод
   * Интеллектуальная обработка текстов
   * Генерация описаний по входному объекту
   * Автоматическая обработка текстов
2. **Зачем нужна предобработка текста для машинного обучения?**

Это нужно для использования текстовых данных в нейронных сетях их нужно представить в виде матрицы, каждый элемент будет соответствовать словам.

1. **Какие виды предобработки текста вы знаете?**

Все тексты на естественном языке имеют большое количество слов, которые не несут информации о данном тексте. К примеру, в английском языке такими словами являются артикли, в русском к ним можно отнести предлоги, союзы, частицы. Данные слова называют шумовыми или стоп-словами. Для достижения лучшего качества классификации на первом этапе предобработки текстов обычно необходимо удалять такие слова. Второй этап предобработки текстов — приведение каждого слова к основе, одинаковой для всех его грамматических форм. Это необходимо, так как слова несущие один и тот же смысл могут быть записаны в разной форме. Например, одно и то же слово может встретиться в разных склонениях, иметь различные приставки и окончания.

1. **Что такое стемминг?**

Стемминг — это процесс нахождения основы слова для заданного исходного слова. Основа слова не обязательно совпадает с морфологическим корнем слова. Задача нахождения основы слова представляет собой давнюю проблему в области компьютерных наук. Первая публикация по данному вопросу датируется 1968 годом. Стемминг применяется в поисковых системах для расширения поискового запроса пользователя, является частью процесса нормализации текста.

1. **Что такое 20 Newsgroups?**

20 Newsgroups – это набор, состоящий из примерно 20 тысяч постов.

1. **Чему должно равняться число входных и выходных нейронов в задаче классификации текстов?**

Размер входного слоя должен быть равен количеству уникальных слов в текстах, а размер выходного слоя — количеству классов.

**4. Вывод**

Научился обрабатывать тексты с помощью векторизаторов, научился классифицировать тексты с помощью нейросетей

**5. Список литературы**

1. Документация по tensorflow. https://www.tensorflow.org/
2. J. Brownlee. Encoder-decoder recurrent neural network models for neural machine translation. https://machinelearningmastery.com/ encoder-decoder- recurrentneural-network-models-neural-machine-translation/.
3. Ian Goodfellow, Yoshua Bengio, and Aaron Courville. D), является eep Learning. MIT Press, 2016
4. Архангельская Е. Николенко С., Кадурин А. Глубокое обучение. Погружение в мир нейронных сетей. СПб.: Питер, 2018