

ОБРАБОТКА ЕСТЕСТВЕННОГО ЯЗЫКА. АНАЛИЗ ДАННЫХ

Оглавление

ОБЩИЕ ТРЕБОВАНИЯ	2
Пояснения к выполнению задания	2
Требования к отчету по проекту	2
СТРУКТУРА ОТЧЕТА ПО ПРОЕКТУ	3
ЗАДАНИЕ	3
Часть 1. Обработка естественного языка	3
Часть 2. Анализ академических данных	5
Часть 3. Визуализация результатов анализа академических данных	6

ОБЩИЕ ТРЕБОВАНИЯ

Пояснения к выполнению задания

Задание является частью проекта. Тема проекта и состав группы соавторов проекта должны быть внесены в Яндекс-форму в элементе курса *Темы проектов, авторские группы. Запись*. Требования к теме и структуре проекта приведены в элементе курса *Проект. Требования*.

В процессе выполнения задания оформляется отчет по проекту (продолжается оформление отчета по проекту, в следующих заданиях продолжается оформление отчета – файл дополняется). Структура отчета по проекту (полный вариант) приведена в элементе курса *Проект. Требования*.

По итогам выполнения работы оформляется:

- 1) отчет по проекту в Word, название файла должно содержать фамилию студента (фамилии студентов), далее текст «_Проект»;
- 2) файлы с работами для оценки и анализа (не менее 15), упакованные в архив, название файла должно содержать фамилии студентов и текст [_РаботыДляОценки];
- 3) файл Excel с результатами оценки; название файла должно содержать фамилии студентов и текст [_ОценкаРабот];
- 4) файл Excel с результатами анализа данных; название файла должно содержать фамилии студентов и текст [_АнализДанных].

На платформе <https://prepod.nspu.ru> прикрепляются: 1) файл-отчет по проекту, 2) архив с работами для оценки и анализа, 3) файл Excel с результатами оценки, 4) файл Excel с результатами анализа данных.

Требования к отчету по проекту

Стиль изложения текста отчета по проекту – научный, текст связанный, однородный, должна прослеживаться логическая линия последовательного выполнения работ. Текст дополняется скриншотами, демонстрирующими ход работ (ссылками на рисунки) и приложениями (ссылками на приложения).

Оформление отчета по проекту должно соответствовать требованиям, указанным в стандарте по оформлению выпускной квалификационной работы, который размещен на сайте НГПУ в разделе Студентам / Система менеджмента качества, доступен после авторизации в сервисе НГПУ.Паспорт.

В стандарте Выпускная квалификационная работа. Требования, нужно изучить раздел 8 (п. 8.1 с общими требованиями к шрифтам, отступам), приложение А (п. 3 Нумерация страниц, п. 8 Таблицы).

Пояснения к требованиям по оформлению учебной документации приведены в элементе курса ***Пояснения к требованиям по оформлению отчета по проекту и другой учебной документации в MS Word.***

Критерии оценки отчета по проекту: соответствие требованиям по оформлению учебной документации (указанным в стандарте по Выпускная квалификационная работа. Требования); соответствие структуры отчета требованиям к структуре (указанным в данном задании и в элементе курса *Проект. Требования*; полнота и корректность наполнения пунктов отчета по проекту (выполнения пунктов задания); соответствие темы проекта, указанной на титульном листе требованиям задания на проект (*Проект. Требования*), профессиональному стандарту и теме, внесенной в регистрационную форму (*Темы проектов, авторские группы. Запись*); соответствие фамилий студентов, указанных на титульном листе отчета по проекту, фамилиям, внесенным в регистрационную форму (*Темы проектов, авторские группы. Запись*).

СТРУКТУРА ОТЧЕТА ПО ПРОЕКТУ

(часть, оформляемая в данном задании):

2.6 Обработка (текста) естественного языка

2.7 Анализ академических данных

2.8 Визуализация результатов анализа академических данных

Библиографический список [формируется при первом создании отчета по проекту (в первом задании), в процессе выполнения проекта дополняется и перемещается в заключительную часть отчета по проекту]

Приложения

ЗАДАНИЕ

Часть 1. Обработка естественного языка

Используя нейросети разработать программный код, позволяющий проверять и оценивать работы учеников, представленные в отдельных файлах (формат файлов: docx, txt, pdf, наименование файла – фамилия ученика или код). Оценку работ осуществлять по заданным критериям: 1 критерий – соответствие структуры (в задаче описать требуемую структуру, например: фамилия и имя ученика; формулировка задания; ответ на задание и др.), другие критерии – не менее 4. Результаты оценки выгрузить в файл в формате CSV, который можно импортировать в Excel-файл с двумя листами: 1 лист с оценкой структуры – наличие каждого элемента структуры (раздела) оценивается в 1 балл для всех учеников; 2 лист с итогами оценки по ученикам – общий балл и балл по каждому критерию оценки для каждого ученика.

Для выполнения задания необходимо:

1.1) придумать задание по исследуемой теме и предмету (например, проверочная работа на знание основных определений по изучаемой теме, заданы 5 терминов, нужно привести их определения; структура отчета: фамилия и имя ученика; наименование темы, по

которой проводится проверочная работа; термины и определения; источник, которым эти термины размещены (учебник, дополнительная литература – указать наименование источника, цифровая образовательная платформа – указать наименование и пр.). Рекомендуется выделить 5 требуемых структурных элементов отчета, чтобы максимальная оценка работы по 1 критерию была 5 (аналогично стандартной 5-балльной шкале);

1.2) сформулировать критерии оценки выполнения задания по правилу, указанному выше (например, критерий 2 – верность (правильность) определений; критерий 3 – все заданные термины рассмотрены; 4 – соответствие изложения определения учебнику; 5 – наличие авторских уточнений и пояснений);

1.3) подготовить файлы с работами учеников (не менее 15); файлы с работами, упакованные в архив, являются отчётными материалами по проекту и прикрепляются на сайте к заданию; название файла должно содержать фамилии студентов и текст [_РаботыДляОценки];

1.4) сформулировать запрос для нейросети, запустить диалог, получить нужный ответ – файл с программным кодом и инструкцией по его реализации;

1.5) запустить выполнение программного кода, указать путь к папке с исходными файлами, к папке с результатами;

1.6) импортировать данные из файла CSV в файл Excel; файл Excel с результатами оценки является отчётным материалом по проекту и прикрепляется на сайте к заданию; название файла должно содержать фамилии студентов и текст [_ОценкаРабот].

В отчете по проекту п. 2.6 «Обработка (текста) естественного языка» разместить подробное описание выполненных работ:

1) формулировку задания по исследуемой теме и предмету – задание, которое выполняли ученики;

2) критерии оценки выполнения задания (по правилу, указанному выше);

3) описание содержания подготовленных файлов с работами учеников, их количество, процедур их создания;

4) формулировка запроса для нейросети;

5) наименование нейросети, которая использовалась для выполнения задания, ссылка;

6) скриншот запроса, остальные скриншоты диалога разместить в приложении к отчету;

5) описание выполнения программного кода, скриншоты;

6) описание процесса импортирования данных из файла CSV в файл Excel;

7) таблицы с полученными результатами;

8) вывод о полученных результатах, корректности и успешности примененных алгоритмов, о дальнейших перспективах применения данного алгоритма, направлениях его улучшения.

Часть 2. Анализ академических данных

Используя нейросети провести анализ ответов учеников с применением алгоритмов классификации, кластеризации, поиск аномалий (выбросов). Задание выполнять на основе работ учеников и результатов оценки предыдущей части задания (продолжаем диалог с нейросетью). Классифицировать работы учеников по признакам: 1) успешность: «успешная», «требующая доработки» или «неудовлетворительная»; 2) самостоятельность: «выполнена самостоятельно», «имеет признаки заимствования». При кластеризации использовать 5 признаков, из которых 3 самостоятельно выделяет нейросеть, 2 признака заданы: 1) выделить элементы работы (определения – для рассмотренного выше примера), которые вызвали затруднения у большинства учащихся; 2) выделить элементы работы (определения – для рассмотренного выше примера), которые все учащиеся выполнили верно. Осуществить поиск аномалий (выбросов). Результаты оценки выгрузить в файл в формате CSV, который можно импортировать в Excel-файл с тремя листами (1 лист – классификация; 2 лист – кластеризация, 3 лист – поиск аномалий).

Для выполнения задания необходимо:

2.1) сформулировать запрос для нейросети, запустить диалог, получить нужный ответ – файл с программным кодом и инструкцией по его реализации;

2.2) запустить выполнение программного кода;

2.3) импортировать данные из файла CSV в файл Excel; файл Excel с результатами анализа данных относится к отчётным материалом по проекту и прикрепляется на сайте к заданию; название файла должно содержать фамилии студентов и текст [_АнализДанных].

В отчете по проекту п. 2.7 «Анализ академических данных» разместить подробное описание выполненных работ:

- 1) формулировка запроса для нейросети;
- 2) наименование нейросети, которая использовалась для выполнения задания, ссылка;
- 3) скриншот запроса, остальные скриншоты диалога разместить в приложении к отчету;
- 4) описание выполнения программного кода, скриншоты;
- 5) описание процесса импортирования данных из файла CSV в файл Excel;
- 6) таблицы с полученными результатами;
- 7) вывод о полученных результатах, корректности и успешности примененных алгоритмов, о дальнейших перспективах применения данного алгоритма, направлениях его улучшения.

Часть 3. Визуализация результатов анализа академических данных

Используя нейросети создать визуализации для результатов оценки и анализа данных, полученных при выполнении предыдущих частей задания. Создать по 5 вариантов визуализации данных для: 1) результатов оценки, 2) для результатов анализа данных. Выбрать по 1 наиболее удачному варианту, которые невозможно было создать без применения нейросетей или их создание вызвало бы ряд сложностей. Обосновать выбор.

В отчете по проекту п. 2.8 «Визуализация данных» разместить подробное описание выполненных работ:

- 1) формулировка запроса для нейросети;
- 2) наименование нейросети, которая использовалась для выполнения задания, ссылка;
- 3) скриншот запроса, остальные скриншоты диалога разместить в приложении к отчету;
- 4) 5 вариантов визуализации результатов оценки – скриншоты, пояснения;
- 5) наиболее удачный вариант визуализации результатов оценки, обоснование выбора;
- 6) 5 вариантов визуализации результатов анализа данных – скриншоты, пояснения;
- 7) наиболее удачный вариант визуализации результатов анализа данных, обоснование выбора;
- 8) вывод о полученных результатах, корректности и успешности применения нейросетей для задач данного вида; о дальнейших перспективах применения нейросетей для задач данного вида.