УТВУТВЕРЖДАЮ

Зав. кафедрой СП ЮУрГУ

\_\_\_\_\_\_\_\_\_Л.Б. Соколинский

«\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_2022 г.

Фонд оценочных средств ООП «Инженерия информационных и интеллектуальных систем» по направлению 09.03.04 – Программная инженерия

Дисциплина «Подготовка данных для машинного обучения»

| **№ КМ** | **Вид КМ** | **Наименование КМ** | **Оценочные средства** |
| --- | --- | --- | --- |
|  | Текущий контроль | Тест 1 | Вопросы для подготовки к тесту:   1. Мат. ожидание 2. Мода 3. Медиана 4. Среднее (арифметическое, геометрическое) 5. Требования, предъявляемые к данным 6. Международные стандарты 7. Проблемы в данных 8. Способы получения новых данных |
|  | Текущий контроль | Тест 2 | Вопросы для подготовки к тесту:   1. Представление изображений в компьютере 2. Трехмерные массивы и тензоры 3. Способы аугментации изображений 4. Разметка изображений 5. Способы хранения наборов данных с изображениями 6. Способы получения новых данных |
|  | Текущий контроль | Статистические характеристики наборов данных. Работа с табличными данными, текстами и изображениями. | Вопросы для подготовки к устному опросу:   1. Статистические особенности данных 2. Мода 3. Среднее 4. Числовые характеристики данных 5. Выбросы 6. Нормализация 7. Уменьшение размерности 8. Противоречивость данных |
|  | Текущий контроль | Представление изображений в компьютере. Рассмотрение современных библиотек работы с изображениями Pillow, OpenCV, Albumentation. Аугментация. | Вопросы для подготовки к устному опросу:   1. Возможности библиотеки OpenCV 2. Возможности библиотеки Albumentation 3. Возможности библиотеки Pillow 4. Предварительная обработка изображений на этапе создания обучающего набора 5. Методы работы с изображениями в tensorflow |
|  | Текущий контроль | Представление видео данных. Кадрирование и аугментация данных. | Вопросы для подготовки к устному опросу:   1. Способы кадрирования видео 2. Методология применения систем искусственного интеллекта к видео данным 3. Сверточные слои 4. Рекуррентные слои 5. Ядро свертки 6. Память рекуррентных слоев 7. Методология одновременной обработки видео и аудио данных 8. Многоголовые сети |
|  | Текущий контроль | Представление текстов в компьютере. Мешок слов, one-hot-encoding, векторное представление, embedings. Методы работы с текстами sklearn и tensorflow. | Вопросы для подготовки к устному опросу:   1. Представление текстовых данных в компьютере 2. Мешок слов 3. One-hot-encoding 4. Векторное представление 5. Embeddings 6. Способы аугментации текстов 7. Возможности библиотеки sklearn 8. Предварительная обработка тектосвых данных на этапе создания обучающего набора 9. Методы работы с текстами в tensorflow 10. Разметка текстов 11. Способы хранения наборов текстовых данных 12. Способы получения новых текстовых данных |
|  | Текущий контроль | Парсинг текстов. Подготовка текстовых данных для решения различных задач: классификации и распознавания именованных сущностей. | Вопросы для подготовки к устному опросу:   1. Парсинг веб сайтов 2. Ручная разметка 3. Перевод pdf в текст. 4. Хранение размеченных текстовых данных 5. Подготовка наборов данных для машинного обучения 6. Способы аугментации текстов 7. Предварительная обработка тектосвых данных на этапе создания обучающего набора 8. Методы работы с текстами в tensorflow 9. Разметка текстов 10. Способы получения новых текстовых данных |
|  | Текущий контроль | Представление аудио в компьютере. Рассмотрение современной библиотеки работы с аудио Librosa. Подготовка аудио данных с помощью методов tensorflow. Аугментация. | Вопросы для подготовки к устному опросу:   1. Представление аудио в компьютере 2. Представление изображений в компьютере 3. Представление текстов в компьютере 4. Тензоры 5. Преобразование данных в тензоры и массивы 6. Основные подходы к кодированию данных 7. One-hot-encoding 8. Мешок слов 9. Векторные представления 10. Аудио в изображения 11. Мелспектрограммы |
|  | Промежуточный контроль | Итоговый тест | Вопросы для подготовки к экзамену по дисциплине:   1. Мат. ожидание, Мода, Медиана 2. Среднее (арифметическое, геометрическое) 3. Требования, предъявляемые к данным 4. Международные стандарты 5. Проблемы в данных 6. Способы получения новых данных 7. Представление изображений в компьютере 8. Трехмерные массивы и тензоры 9. Способы аугментации изображений 10. Предварительная обработка изображений на этапе создания обучающего набора 11. Разметка изображений 12. Способы хранения наборов данных с изображениями 13. Способы получения новых данных 14. Представление текстовых данных в компьютере 15. Мешок слов, One-hot-encoding, Векторное представление, Embeddings 16. Способы аугментации текстов 17. Предварительная обработка тектосвых данных на этапе создания обучающего набора 18. Разметка текстов 19. Способы хранения наборов текстовых данных 20. Способы получения новых текстовых данных 21. Парсинг веб сайтов 22. Ручная разметка 23. Перевод pdf в текст. 24. Предварительная обработка текстовых данных на этапе создания обучающего набора 25. Представление аудио в компьютере 26. Дискретизация аудио сигнала 27. Хранение размеченных аудио данных 28. Подготовка наборов данных для машинного обучения 29. Способы аугментации аудио 30. Предварительная обработка аудио данных на этапе создания обучающего набора 31. Аудио в изображения 32. Мелспектрограммы |

Паспорт фонда оценочных средств приведен в п. 6.3 РПД.

Разработчик С.А. Иванов

ФГАОУ ВО «Южно-Уральский государственный университет   
(национальный исследовательский университет)»

Кафедра системного программирования

Дисциплина «Подготовка данных для машинного обучения»

ИТОГОВЫЙ ТЕСТ

| № | Вопрос | Варианты ответа |
| --- | --- | --- |
| 1. | Аугментация это | 1. Методика создания дополнительных данных из имеющихся данных  2. Методика нормализации данных  3. Способы представления текстовой информации  4. методы векторизации данных |
| 2 | Совокупность всех объектов\n(единиц), относительно которых предполагается делать выводы при изучении\nконкретной задачи это | 1. генеральная совокупность  2. выборочная совокупность  3. репрезентативная выборка  4. кластерная выборка |
| 3 | Сумма\nвсех значений измеренного признака, деленная на количество измеренных значений это | 1. Среднее значение  2. Дисперсия  3. Мода  4. Гистограмма |
| 4 | Разность\nмаксимального и минимального значений это | 1. размах  2. мода  3. математическое ожидание  4. нормализация |
| 5 | Корпус в задачах обработки естественного языка это | 1. <p>подобранная\nи обработанная по определённым правилам совокупность текстов, используемых в\nкачестве базы для исследования языка.<br><p>  2. это совокупность слов и предложений на которые разбиты исходные тексты.  3. <p>это наборы токенов полученных из исходных текстовых данных.<br><p>  4. это векторное представления токенов, полученных из исходных данных. |
| 6 | Укажите все базовые критерии\nкачества данных по стандарту ISO 9000 | 1. полнота  2. достоверность  3. точность  4. согласованность  5. доступность  6. своевременность  7. долговечность |
| 7 | Какой подход к нормализации данных преобразует исходный набор данных в новый набор со средним значением равным 0 и стандартным отклонением равным 1 | 1. Стандартизация  2. Нормализация  3. Минимакс  4. Среднее квадратичное отклонение |
| 8 | Метод, с помощью которого мы кодируем несколько последовательных слов как один токен, называется | 1. n- grams  2. One hot encoding  3. Lemmatisation  4. TF-IDF |
| 9 | Каким способом из слова "был" можно извлечь правильную базовую форму глагола "быть"? | 1.Стемминг  2. Лемматизация  3. TF-IDF  4. One hot encoding |
| 10 | Какая формула описывает, как получается выходной вектор после применения SimpleRNN? | 1.y \= activation(dot(state\_t, U) + dot(input\_t, W) + b)  2. y \= activation(dot(state\_t, U) + dot(input\_t, W) + X)  3. y \= activation(dot(state\_t, U) + dot(input\_t, W)) + b  4. y \= activation(dot(W, U) + dot(input\_t, state\_t) + b) |
| 11 | Что за метод всегда обрезает слово до основания, так что "was" становится "wa"? | 1. Stemming  2. Lemmatization  3. TF-IDF  4. One hot encoding |
| 12 | Операция свертки извлекает паттерны из входной карты признаков и применяет одинаковые преобразования ко всем паттернам, создавая выходную карту признаков. Каналы в этой карте соответствуют ... | 1.фильтрам  2. цветовым каналам изображения  3. значениям цветов пикселей  4. лицу на изображении |
| 13 | Слой Embedding имеет выход | 1. векторные представления слов  2. категориальные признаки  3. n-граммы |
| 14 | Так ли работает сверточная нейронная сеть? Первый сверточный слой будет изучать небольшие локальные паттерны, такие как края, второй будет изучать более крупные паттерны, состоящие из особенностей, возвращенных первым слоем, и так далее. | 1. да  2. нет |
| 15 | Сверточная нейронная сеть - это класс глубоких нейронных сетей, наиболее часто применяемых для анализа | 1. визуальных образов  2. спортивной области  3. генерации текста |