|  |  |  |
| --- | --- | --- |
| РУТ(МИИТ)  Академия «Высшая инженерная школа»  2023/2024 учебный год | **ЭКЗАМЕНАЦИОННЫЙ**  **БИЛЕТ №1**  по дисциплине  «Анализ больших данных и текстовый поиск» для студентов образовательной программы «IT-сервисы и технологии обработки данных на транспорте» | УТВЕРЖДАЮ  Руководитель образовательной программы  \_\_\_\_\_\_\_\_\_\_\_\_  к.т.н., Проневич О.Б. |
| 1. *Основные этапы предварительной обработки текста. Удаление нерелевантных символов, числовая кодировка текстовой информации, снижение уникальности, токенизация.* 2. *LSTM-сети. Задачи. Слой сохранения. Вектор контекста, параметр С.* 3. *Слои нейронной сети. Conv1D, SpatialDropout1D. Гиперпараметры слоев, inputs, outouts. Результаты преобразования.*   *Задача: Классификация “Spam-сообщения”.* | | |

|  |  |  |
| --- | --- | --- |
| РУТ(МИИТ)  Академия «Высшая инженерная школа»  2023/2024 учебный год | **ЭКЗАМЕНАЦИОННЫЙ**  **БИЛЕТ №2**  по дисциплине  «Анализ больших данных и текстовый поиск» для студентов образовательной программы «IT-сервисы и технологии обработки данных на транспорте» | УТВЕРЖДАЮ  Руководитель образовательной программы  \_\_\_\_\_\_\_\_\_\_\_\_  к.т.н., Проневич О.Б. |
| 1. *Токенизация текста. 6 основных характеристик Tokenizer в tensorflow.* 2. *Модели Seq2Seq. Задачи. Общая структура. Encoder и Decoder. Вектор контекста.* 3. *Суммаризация текстов. Виды. TextRank. Метод Луна. LSA.*   *Задача: Классификация “Spam-сообщения”.* | | |

|  |  |  |
| --- | --- | --- |
| РУТ(МИИТ)  Академия «Высшая инженерная школа»  2023/2024 учебный год | **ЭКЗАМЕНАЦИОННЫЙ**  **БИЛЕТ №3**  по дисциплине  «Анализ больших данных и текстовый поиск» для студентов образовательной программы «IT-сервисы и технологии обработки данных на транспорте» | УТВЕРЖДАЮ  Руководитель образовательной программы  \_\_\_\_\_\_\_\_\_\_\_\_  к.т.н., Проневич О.Б. |
| 1. *Токенизация теста. Преобразование текста в вектор. text\_to\_matrix(), texts\_to\_sequences() отличия, условия применения. Примеры преобразования текста. Переход от векторной формы к словам.* 2. *Работа с многомерными данными. Процесс перехода от текста к многомерным данным. Свертка. Ядро свертки. Max Pooling, Avg Pooling. Фильтры при свертке n мерных данных.* 3. *Трансформеры при анализе тестовых данных. Вычисление внутреннего внимания. Три основных параметра.*   *Задача: Классификация “Spam-сообщения”.* | | |

|  |  |  |
| --- | --- | --- |
| РУТ(МИИТ)  Академия «Высшая инженерная школа»  2023/2024 учебный год | **ЭКЗАМЕНАЦИОННЫЙ**  **БИЛЕТ №4**  по дисциплине  «Анализ больших данных и текстовый поиск» для студентов образовательной программы «IT-сервисы и технологии обработки данных на транспорте» | УТВЕРЖДАЮ  Руководитель образовательной программы  \_\_\_\_\_\_\_\_\_\_\_\_  к.т.н., Проневич О.Б. |
| 1. *Сеть прямого распространения. Принцип работы, основные понятия. Архитектура сети прямого распространения. Dense слой.* 2. *Двунаправленные рекуррентные сети. Bidirectional.* 3. *Слои нейронной сети. Conv1D, SpatialDropout1D. Гиперпараметры слоев, inputs, outouts. Результаты преобразования.*   *Задача: Классификация “Spam-сообщения”.* | | |

|  |  |  |
| --- | --- | --- |
| РУТ(МИИТ)  Академия «Высшая инженерная школа»  2023/2024 учебный год | **ЭКЗАМЕНАЦИОННЫЙ**  **БИЛЕТ №5**  по дисциплине  «Анализ больших данных и текстовый поиск» для студентов образовательной программы «IT-сервисы и технологии обработки данных на транспорте» | УТВЕРЖДАЮ  Руководитель образовательной программы  \_\_\_\_\_\_\_\_\_\_\_\_  к.т.н., Проневич О.Б. |
| 1. *Полносвязный слой нейронной сети. Слой Dropout. Функция активации.* 2. *Слои нейронной сети. Conv1D. Окно свертки, ядро свертки.* 3. *Слои нейронной сети. Conv1D, SpatialDropout1D. Гиперпараметры слоев, inputs, outouts. Результаты преобразования.*   *Задача: Классификация “Spam-сообщения”.* | | |

|  |  |  |
| --- | --- | --- |
| РУТ(МИИТ)  Академия «Высшая инженерная школа»  2023/2024 учебный год | **ЭКЗАМЕНАЦИОННЫЙ**  **БИЛЕТ №6**  по дисциплине  «Анализ больших данных и текстовый поиск» для студентов образовательной программы «IT-сервисы и технологии обработки данных на транспорте» | УТВЕРЖДАЮ  Руководитель образовательной программы  \_\_\_\_\_\_\_\_\_\_\_\_  к.т.н., Проневич О.Б. |
| 1. *4 основных оптимизатора keras. Принципы работы. Этап обучение нейронной сети на котором осуществляется выбор оптимизатора.* 2. *Задача NER. Типовые сущности. Метки начала и окончания сущностей. Метод перехода к векторной форме при решении задачи NER при токенизации. Варьируемые параметры при обучении нейронной сети для решения задачи NER.* 3. *Модели Seq2Seq. Общие проблемы классической Seq2Seq. Seq2Seq, общий принцип работы. Отличие от простой Seq2Seq. Вектор внимания, вектор контекста. Размерность вектора контекста.*   *Задача: Классификация “Spam-сообщения”.* | | |