

Программа курса «Middle data scientist: продвинутый специалист по анализу больших данных»

Номер	Название темы	Количество часов	Описание темы
1	Рекуррентные нейронные сети. LSTM слои	3	Рекуррентные нейронные сети, LSTM слои. Особенности строения. Решение задачи с использованием LSTM слоёв. Сравнение LSTM и других RNN слоёв. Практические занятия: Собрать простую LSTM нейронную сеть, решить задачу классификации текстовых данных. Выполнить прогноз на проверочных данных. Снять метрики и ошибки модели.
2	Tensorflow, построение нейронных сетей на уровне графов	6,5	Библиотека tensorflow, понятия графов и построение нейронных сетей с применением графов. Практические занятия: Загрузить датасет (кейс – классификация одежды), подготовить датасет, выполнить исследование и отбор признаков. Собрать и обучить простую нейронную сеть с применением библиотеки tensorflow.
3	Построение архитектуры нейронной сети для задач object detection	6,5	Библиотека pytorch, для создания нейронных сетей. Особенности построения нейронных сетей, как отдельных классов. Практические занятия: Загрузить датасет (кейс – классификация одежды), подготовить датасет, выполнить исследование и отбор признаков. Собрать и обучить простую нейронную сеть с применением библиотеки pytorch.
4	Построение архитектуры нейронной сети для задач object detection	7	Основные подходы к созданию архитектур нейронных сетей, для распознаваний образов, классификации изображений. Подход по детекции объекта в кадре, с применением bounding box. Практические занятия: Загрузить датасет (кейс – обнаружение самолётов), подготовить датасет, выполнить исследование и отбор признаков. Собрать и обучить нейронную сеть для object detection.
5	Построение архитектуры нейронной сети для задач segmentation	4,5	Подход к созданию архитектур нейронных сетей, для решения задач segmentation, оптимальные метрики и принцип оптимизации сетей. Практические занятия: Загрузить датасет, подготовить датасет (датасет содержит размещённые изображения для обучения автопилота в автомобилях), выполнить исследование и отбор признаков. Собрать и обучить нейронную сеть для segmentation автопилота.
6	Задачи NLP. Препроцессинг текстовых неструктурированных данных	4,5	Ключевые задачи при работе с текстами, препроцессинг текстовых данных, преобразование слов в индексы и в вектора, для решения задач NLP с применением нейронных сетей. Области решаемых задач. Практические занятия: Загрузить датасет (датасет содержит реферативные данные научных публикаций) подготовить датасет, выполнить препроцессинга текстовых данных.
7	Построение архитектуры нейронной сети для классификации текстов	4	Основные архитектуры нейронных сетей, для задач классификации текстовых данных, оптимальные метрики и активационные функции. Данные, которые подаются на вход нейронной сети и ошибки, которые минимизируются. Практические занятия: Загрузить датасет (датасет содержит отзывы о автомобилях и отелях. Выбрать любую отрасль и выполнить классификацию по тональности), подготовить датасет, выполнить препроцессинга текстовых данных. Собрать и обучить нейронную сеть для классификации текстовых данных, сделать визуализацию процесса обучения, написать выводы, предложить методы оптимизации и улучшения нейронной сети.
8	Чат-боты и генерация текста. Особенности построения	6	Основные этапы создания чат-бота, генерирующего естественную речь из скрытого многомерного пространства, оптимальные архитектуры, сложности в создании человекоподобного чат-бота. Практические занятия: Загрузить датасет (датасет содержит диалоги в формате вопрос-ответ, для обучения чат-бота), подготовить датасет, выполнить препроцессинга текстовых данных, использовать

	архитектуры нейронной сети		архитектуру нейронной сети, подать на вход разное количество обучающей выборки в формате вопрос - ответ, сделать визуализацию обучения нейронной сети, написать выводы, предложить методы оптимизации нейронной сети.
9	Сегментация текстовых данных	4	Подходы к разметке и подготовке обучающей выборки, при создании архитектур нейронных сетей, для задач сегментации текстовых данных. Практическая работа: Загрузить датасет (датасет содержит 1000 размеченных договоров), подготовить датасет, выполнить препроцессинга текстовых данных, использовать архитектуру нейронной сети, подать на вход разное количество сегментируемых классов, сделать визуализацию обучения нейронной сети, написать выводы, предложить методы оптимизации нейронной сети.
10	Контроль версионности моделей с tensorflow serving	5,5	Вопросы контроля версионности моделей, с применением docker образа tensorflow-serving, и вызов необходимых версий по средствам API. Практические занятия: Загрузить датасет, обучить модель, сохранить в формате save_model. Обучить вторую модель, также сохранить её. Написать программу, которая будет проверять на входе цветное изображение или нет, и отправлять на распознавание в соответствующую версию.
11	Развёртывание облачной инфраструктуры. Обзор облачных платформ Google Cloud Platform, AWS, Sbercloud	5,5	Обзор основных поставщиков облачных решений, как иностранные, так и отечественные. Особенности тарификации и развёртывания облачной инфраструктуры. Практические занятия: Развернуть простое облако на платформе GCP с одной виртуальной машиной и одним хранилищем SSD. Сделать скриншоты рабочего instance.
12	Flask приложение. Выведение моделей в production	7	Описание pipeline, выведения модели машинного обучения в продуктив. Создание сервисов на базе фреймворка flask. Будут разобраны основные этапы, etl процессы и взаимодействие. Практические занятия: Выведение моделей в продуктив.
13	Итоговая аттестация	3	Тестирование

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

УТВЕРЖДАЮ
Первый проректор –
проректор по учебной работе
МГТУ им. Н.Э. Баумана
Б.В. Падалкин
_____ 2021 г.



Дополнительное профессиональное образование

ДОПОЛНИТЕЛЬНАЯ ПРОФЕССИОНАЛЬНАЯ ПРОГРАММА
ПОВЫШЕНИЯ КВАЛИФИКАЦИИ
«Middle data scientist: продвинутый специалист по анализу больших
данных»

Регистрац. № 05.22.21.12.4

Москва, 2021

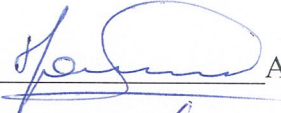
АВТОР ПРОГРАММЫ:

Заместитель директора
Центра НТИ «Цифровое материаловедение:
новые материалы и вещества»



А.Ю. Чесалов

СОГЛАСОВАНО:

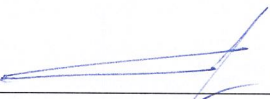
Руководитель стратегического проекта «Bauman U2U»


А.С. Комшин

Начальник УСП
МГТУ им. Н.Э. Баумана


Т.А. Гузева

Директор ИСОТ
МГТУ им. Н.Э. Баумана


В.Г. Брекалов

Директор
Центра НТИ «Цифровое материаловедение:
новые материалы и вещества»


В.А. Нелюб

Оглавление

1. Общая характеристика дополнительной профессиональной программы (ДПП)	4
1.1..Цель ДПП	4
1.2. Планируемые результаты обучения	4
1.3. Дополнительные характеристики ДПП.....	4
1.4. Перечень профессиональных компетенций в рамках имеющейся квалификации, качественное изменение которых осуществляется в результате обучения	5
1.5. Соответствие видов деятельности профессиональным компетенциям и их составляющих	5
2. Учебный план ДПП.....	6
2.1. Категория слушателей ДПП.....	6
2.2. Общая трудоёмкость программы, аудиторная и самостоятельная работа	6
2.3. Форма обучения.....	6
2.4. Учебный план	6
3. Календарный учебный график.....	8
4. Рабочая программа ДПП	10
5. Условия реализации ДПП	26
5.1. Организационные условия реализации ДПП.....	26
5.2. Педагогические условия реализации ДПП	26
5.3. Учебно-методическое обеспечение ДПП.....	27
5.4. Методические рекомендации	27
6. Формы итоговой аттестации ДПП.....	28
7. Оценочные материалы итоговой аттестации.....	28
7.1. Паспорт комплекта оценочных средств	28
7.2. Комплект оценочных средств	28

1. Общая характеристика дополнительной профессиональной программы (ДПП)

Программа подготовлена на основе:

- Федерального закона от 29 декабря 2012 года № 273-ФЗ «Об образовании в Российской Федерации»;
- приказа Минобрнауки России от 23 августа 2017 года № 816 «Об утверждении Порядка применения организациями, осуществляющими образовательную деятельность, электронного обучения, дистанционных образовательных технологий при реализации образовательных программ»;
- требований Приказа Минобрнауки России от 01.07.2013 года № 499 «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам»;
- методических рекомендаций-разъяснений Минобрнауки России по разработке дополнительных профессиональных программ на основе профессиональных стандартов от 22 апреля 2015 года № ВК-1030/06.

Реализация программы дополнительного профессионального образования направлена на получение новой(-ых) компетенции(-ий), необходимой(-ых) для профессиональной деятельности, в рамках реализации программы стратегического академического лидерства «Приоритет-2030».

1.1. Цель ДПП

Сформировать у обучающихся компетенции в области создания информационных технологий нового поколения, обеспечивающих экономически эффективное извлечение полезной информации из больших объемов разнообразных данных путем высокой скорости их сбора, обработки и анализа, и применение этих технологий в информационно-аналитической деятельности, в системах управления и принятия решений, а также для разработки на их основе новых продуктов и услуг.

1.2. Планируемые результаты обучения

Планируемые результаты обучения по ДПП:

- освоение профессиональных компетенций в процессе изучения перечисленных тем в учебном плане;
- успешное освоение программы повышения квалификации;
- успешное прохождение итоговой аттестации (зачет).

Обучающимся, успешно прошедшим обучение, выполнившим текущие контрольные задания и выдержавшим предусмотренное учебным планом итоговую аттестацию, выдается удостоверение о повышении квалификации по ДПП «Middle data scientist: продвинутый специалист по анализу больших данных».

1.3. Дополнительные характеристики ДПП

Перечень профессиональных компетенций в рамках имеющейся квалификации, качественное изменение которых осуществляется в результате обучения, определены в Приказом Министерства труда и социальной защиты Российской Федерации от 06.07.2020 №405н об утверждении профессионального стандарта «Специалист по большим данным».

Вид профессиональной деятельности:

- Создание и применение технологий больших данных (Код 06.042).

Обобщенные трудовые функции:

- Анализ больших данных с использованием существующей в организации методологической и технологической инфраструктуры (ОТФ 06.042_А).

Трудовые функции:

- Проведение аналитического исследования с применением технологий больших данных в соответствии с требованиями заказчика (ТФ 06.042_А/04.6).

1.4. Перечень профессиональных компетенций в рамках имеющейся квалификации, качественное изменение которых осуществляется в результате обучения

Данная программа направлена на совершенствование и (или) получение новой(-ых) компетенции, необходимой для профессиональной деятельности, и (или) повышение профессионального уровня в рамках имеющейся квалификации.

При определении профессиональных компетенций на основе профессиональных стандартов, МГТУ им. Н.Э.Баумана осуществляет выбор профессиональных стандартов, соответствующих профессиональной деятельности выпускников, из числа указанных в приложении к ФГОС ВО и (или) иных профессиональных стандартов, соответствующих профессиональной деятельности выпускников.

Реализуемые компетенции базируются на основании Приказа Министерства образования и науки РФ от 19 сентября 2017 г. N 929 "Об утверждении федерального государственного образовательного стандарта высшего образования - бакалавриат по направлению подготовки 09.03.01 Информатика и вычислительная техника" (с изменениями и дополнениями) Редакция с изменениями N 1456 от 26.11.2020.

Перечень компетенций:

ПК-1. Способен участвовать в исследовательских и опытно-конструкторских разработках в области создания и совершенствования ИТ-систем.

ОПК-3. Способен решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности.

1.5. Соответствие видов деятельности профессиональным компетенциям и их составляющих

Профессиональные компетенции	Практический опыт	Умения	Знания
- Проведение аналитического исследования с применением технологий больших данных в соответствии с требованиями заказчика (А/04.6).			
ПК-1. Способен участвовать в исследовательских и опытно-конструкторских разработках в области создания и совершенствования ИТ-систем	Выбор методов и инструментальных средств анализа больших данных для проведения аналитических работ	Проводить аналитические работы с использованием технологий больших данных, как индивидуально, так и, осуществляя руководство малыми аналитическими группами	Технологии анализа данных: статистический анализ, семантический анализ, анализ изображений, машинное обучение, методы сравнения средних, частотный анализ, анализ соответствий, кластерный анализ, дискриминантный анализ, факторный анализ, деревья классификации, многомерное шкалирование, моделирование

			структурными уравнениями, методы анализа выживаемости, временные ряды, планирование экспериментов, карты контроля качества
ОПК-3. Способен решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности			

2. Учебный план ДПП

2.1. Категория слушателей ДПП

Имеющаяся квалификация (требования к обучающимся) – к освоению ДПП допускаются лица, соответствующего профессиональному стандарту уровню образования или получающие соответствующий уровень образования (бакалавриат).

2.2. Общая трудоёмкость программы, аудиторная и самостоятельная работа

Общая трудоёмкость программы 67 академических часов, из них 26 часов аудиторной работы, 38 часов самостоятельной работы и 3 часа итоговой аттестации.

2.3. Форма обучения

Форма обучения по ДПП – очная, с применением электронного обучения и дистанционных образовательных технологий.

2.4. Учебный план

ДПП «Middle data scientist: продвинутый специалист по анализу больших данных» реализуется одним модулем.

№ п/п	Наименование темы, модуля	Форма контроля	Всего, акад. час*	В том числе		
				Лекции	Практические занятия	Самостоятельная работа
1.	Рекуррентные нейронные сети. LSTM слои	тест	3,5	1	0,5	2
2.	Tensorflow, построение нейронных сетей на уровне графов	тест	6,5	0,5	2	4
3.	Обзор библиотеки PyTorch. Особенности построения нейронных сетей	тест	6,5	0,5	2	4
4.	Построение архитектуры нейронной сети для задач object	тест	7	1	2	4

	detection					
5.	Построение архитектуры нейронной сети для задач segmentation	тест	4,5	0,5	2	2
6.	Задачи NLP. Препроцессинга текстовых неструктурированных данных	тест	4,5	1	1	2,5
7.	Построение архитектуры нейронной сети для классификации текстов	тест	4	0,5	1	2,5
8.	Чат-боты и генерация текста. Особенности построения архитектуры нейронной сети	тест	5,5	0,5	1	4
9.	Сегментация текстовых данных	тест	4	0,5	1,5	2
10.	Контроль версионности моделей с tensorflow serving	тест	5,5	0,5	2	3
11.	Развёртывание облачной инфраструктуры. Обзор облачных платформ Google Cloud Platform, AWS, Sbercloud	устный опрос	5,5	0,5	1	4
12.	Flask приложение. Выведение моделей в production	домашнее задание	7	1	2	4
13.	Итоговая аттестация	зачет	3	-	-	-
	ИТОГО	-	67	8	18	38

*академический час составляет 45 минут

3. Календарный учебный график

№ п/п	Наименование темы, модуля	1 день	2 день	3 день	4 день	5 день	6 день	7 день
1.	Рекуррентные нейронные сети. LSTM слой							
2.	Tensorflow, построение нейронных сетей на уровне графов							
3.	Обзор библиотеки PyTorch. Особенности построения нейронных сетей							
4.	Построение архитектуры нейронной сети для задач object detection							
5.	Построение архитектуры нейронной сети для задач segmentation							
6.	Задачи NLP. Препроцессинг текстовых неструктурированных данных							
7.	Построение архитектуры нейронной сети для классификации текстов							
8.	Чат-боты и генерация текста. Особенности построения архитектуры нейронной сети							
9.	Сегментация текстовых данных							
10.	Контроль версионности моделей с tensorflow serving							
11.	Развёртывание облачной инфраструктуры. Обзор облачных платформ Google Cloud Platform, AWS, Sbercloud							
12.	Flask приложение. Выведение моделей в production							
13.	Итоговая аттестация							

№ п/п	Наименование темы, модуля	7 день	8 день	9 день	10 день	11 день	12 день	13 день	14 день
1.	Рекуррентные нейронные сети. LSTM слои								
2.	Tensorflow, построение нейронных сетей на уровне графов								
3.	Обзор библиотеки PyTorch. Особенности построения нейронных сетей								
4.	Построение архитектуры нейронной сети для задач object detection								
5.	Построение архитектуры нейронной сети для задач segmentation								
6.	Задачи NLP. Препроцессинг текстовых неструктурированных данных								
7.	Построение архитектуры нейронной сети для классификации текстов								
8.	Чат-боты и генерация текста. Особенности построения архитектуры нейронной сети								
9.	Сегментация текстовых данных								
10.	Контроль версионности моделей с tensorflow serving								
11.	Развёртывание облачной инфраструктуры. Обзор облачных платформ Google Cloud Platform, AWS, Sbercloud								
12.	Flask приложение. Выведение моделей в production								

13.	Итоговая аттестация								Зачет
-----	---------------------	--	--	--	--	--	--	--	-------

Рекомендованный срок освоения ДПП — 14 дней.

Фактическое расписание занятий утверждается при заключении договора с обучающимися или при формировании группы.

4. Рабочая программа ДПП

4.1. Рабочая программа модуля

4.1.1. Цель изучения модуля: сформировать у обучающихся компетенции в области создания информационных технологий нового поколения, обеспечивающих экономически эффективное извлечение полезной информации из больших объемов разнообразных данных путем высокой скорости их сбора, обработки и анализа, и применение этих технологий в информационно-аналитической деятельности, в системах управления и принятия решений, а также для разработки на их основе новых продуктов и услуг.

4.1.2. Задача изучения модуля: изучить средства анализа больших данных.

4.1.3. Планируемые результаты обучения

Процесс изучения раздела направлен на формирование следующих компетенций

Код/ наименование компетенции	Перечень планируемых результатов обучения по модулю	Методы и формы обучения, способствующие формированию и развитию компетенции
ПК-1	<p>Знать: Методы поиска научно-технической информации по теме исследования; методы проведения экспериментов и наблюдений, обобщения и обработки их результатов.</p> <p>Уметь: Выполнять поиск и анализ научно-технической информации по теме исследования, проводить эксперименты, обобщать и обрабатывать полученные результаты.</p> <p>Владеть: Практическими навыками поиска научно-технической информации по теме исследования.</p>	<p>Методы обучения: Активные, пассивные, интерактивные. Формы обучения: лекция; практическое занятие; самостоятельная работа.</p>
ОПК-3	<p>Знать: Технологии анализа данных: статистический анализ, семантический анализ, анализ изображений, машинное обучение, методы сравнения средних, частотный анализ, анализ соответствий, кластерный анализ, дискриминантный анализ, факторный анализ, деревья классификации, многомерное шкалирование, моделирование структурными уравнениями, методы анализа выживаемости, временные ряды, планирование</p>	<p>Методы обучения: Активные, пассивные, интерактивные. Формы обучения: лекция; практическое занятие; самостоятельная работа.</p>

	<p>экспериментов, карты контроля качества.</p> <p>Уметь: Проводить аналитические работы с использованием технологий больших данных, как индивидуально, так и, осуществляя руководство малыми аналитическими группами.</p> <p>Владеть: Выбор методов и инструментальных средств анализа больших данных для проведения аналитических работ.</p>	
--	---	--

4.1.4 Содержание курса

Тема 1. Рекуррентные нейронные сети. LSTM слои (3 часа)

Лекции (1 час). В рамках данной темы будут рассмотрены Рекуррентные нейронные сети с памятью. LSTM слои. Особенности строения. Решение задачи с использованием LSTM слоёв. Сравнение LSTM и других RNN слоёв.

Практические занятия (0,5 часа). Собрать простую LSTM нейронную сеть, решить задачу классификации текстовых данных. Выполнить прогноз на проверочных данных. Снять метрики и ошибки модели.

Самостоятельная работа (2 часа).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Рекуррентные нейронные сети. LSTM слои	Рекуррентные нейронные сети	Проработка дополнительных источников информации	Хливненко Л.В., Пятакович Ф.А. Практика нейросетевого моделирования: учебное пособие / Хливненко Л.В., Пятакович Ф.А. – 2-е изд., стер. – СПб.: Лань, 2021. – 196 с. (https://library.bmstu.ru/Catalog/Details/556123)	тест

Тема 2. Tensorflow, построение нейронных сетей на уровне графов (6,5 часа)

Лекции (0,5 часа). В рамках данной темы будет рассмотрена библиотека tensorflow, понятия графов и построение нейронных сетей с применением графов.

Практические занятия (2 часа). Загрузить датасет (кейс – классификация одежды), подготовить датасет, выполнить исследование и отбор признаков. Собрать и обучить простую нейронную сеть с применением библиотеки tensorflow.

Самостоятельная работа (4 часа).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Tensorflow, построение нейронных сетей на уровне графов	Уровень графов	Проработка дополнительных источников информации	Галушкин А.И. Нейронные сети: основы теории / Галушкин А.И. – М.: Горячая линия - Телеком, 2010. – 496 с. (https://library.bmstu.ru/Catalog/Details/219193)	тест

Тема 3. Обзор библиотеки PyTorch. Особенности построения нейронных сетей (6,5 часа)

Лекции (0,5 часа). В рамках данной темы будет рассмотрена библиотека pytorch, для создания нейронных сетей. Разобраны особенности построение нейронных сетей, как отдельных классов.

Практические занятия (2 часа). Загрузить датасет (кейс – классификация одежды), подготовить датасет, выполнить исследование и отбор признаков. Собрать и обучить простую нейронную сеть с применением библиотеки pytorch.

Самостоятельная работа (4 часа).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Обзор библиотеки PyTorch. Особенности построения нейронных сетей	Построение нейронных сетей	Проработка дополнительных источников информации	Галушкин А.И. Нейронные сети: основы теории / Галушкин А.И. – М.: Горячая линия - Телеком, 2010. – 496 с. (https://library.bmstu.ru/Catalog/Details/219193)	тест

Тема 4. Построение архитектуры нейронной сети для задач object detection (7 часов)

Лекции (1 час). В рамках данной темы будут рассмотрены основные подходы к созданию архитектур нейронных сетей, для распознаваний образов, классификации изображений. Будет рассмотрен подход по детекции объекта в кадре, с применением bounding box.

Практические занятия (2 часа). Загрузить датасет (кейс – обнаружение самолётов), подготовить датасет, выполнить исследование и отбор признаков. Собрать и обучить нейронную сеть для object detection.

Самостоятельная работа (4 часа).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Построение архитектуры нейронной сети для задач object detection	Архитектура, object detection	Проработка дополнительных источников информации	Ручкин В.Н., Костров Б.В., Свирина А.Г. Системы искусственного интеллекта. Нейросети и нейрокомпьютеры: учебник для вузов / Ручкин В.Н., Костров Б.В., Свирина А.Г. – М.: КУРС, 2018. – 282 с. (https://library.bmstu.ru/Catalog/Details/509245)	тест

Тема 5. Построение архитектуры нейронной сети для задач segmentation (4,5 часа)

Лекции (0,5 часа). В рамках данной темы будет разобран подход к созданию архитектур нейронных сетей, для решения задач segmentation, оптимальные метрики и принцип оптимизации сетей.

Практические занятия (2 часа). Загрузить датасет, подготовить датасет (датасет содержит размещенные изображения для обучения автопилота в автомобилях), выполнить исследование и отбор признаков. Собрать и обучить нейронную сеть для segmentation автопилота.

Самостоятельная работа (2 часа).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Построение архитектуры нейронной сети для задач segmentation	Segmentation	Проработка дополнительных источников информации	Зозуля Ю.И. Интеллектуальные нейросистемы / Зозуля Ю.И. – М.: Радиотехника, 2003. – 143 с. (https://library.bmstu.ru/Catalog/Details/92733)	тест

Тема 6. Задачи NLP. Препроцессинга текстовых неструктурированных данных (4,5 часа)

Лекции (1 час). В рамках данной темы будут рассмотрены ключевые задачи при работе с текстами, препроцессинг текстовых данных, преобразование слов в индексы и в вектора, для решения задач NLP с применением нейронных сетей. Области решаемых задач.

Практические занятия (1 час). Загрузить датасет (датасет содержит реферативные данные научных публикаций) подготовить датасет, выполнить препроцессинга текстовых данных.

Самостоятельная работа (2,5 часа).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Задачи NLP. Препроцессинга текстовых неструктурированных данных	Препроцессинг, текстовые неструктурированные данные	Проработка дополнительных источников информации	Злобин В.К., Ручкин В.Н. Нейросети и нейрокомпьютеры: учеб. пособие для вузов / Злобин В.К., Ручкин В.Н. – СПб.: БХВ-Петербург, 2011. – 252 с. https://library.bmstu.ru/Catalog/Details/246505	тест

Тема 7. Построение архитектуры нейронной сети для классификации текстов (4 часа)

Практическая работа (0,5 часа). В рамках данной темы будут рассмотрены основные архитектуры нейронных сетей, для задач классификации текстовых данных, оптимальные метрики и активационные функции. Данные, которые подаются на вход нейронной сети и ошибки, которые минимизируются.

Практические занятия (1 час). Загрузить датасет (датасет содержит отзывы о автомобилях и отелях. Выбрать любую отрасль и выполнить классификацию по тональности), подготовить датасет, выполнить препроцессинга текстовых данных. Собрать и обучить нейронную сеть для классификации текстовых данных, сделать визуализацию процесса обучения, написать выводы, предложить методы оптимизации и улучшения нейронной сети.

Самостоятельная работа (0,5 часа).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Построение архитектуры нейронной сети для классификации текстов	Классификация текстов	Проработка дополнительных источников информации	Зозуля Ю.И. Интеллектуальные нейросистемы / Зозуля Ю.И. – М.: Радиотехника, 2003. – 143 с. (https://library.bmstu.ru/Catalog/Details/92733)	тест

Тема 8. Чат-боты и генерация текста. Особенности построения архитектуры нейронной сети (6 часов)

Практическая работа (0,5 часов). В рамках данной темы будут рассмотрены основные этапы создания чат-бота, генерирующего естественную речь из скрытого

многомерного пространства, оптимальные архитектуры, сложности в создании человекоподобного чат-бота.

Практические занятия (1 час). Загрузить датасет (датасет содержит диалоги в формате вопрос-ответ, для обучения чат-бота), подготовить датасет, выполнить препроцессинг текстовых данных, использовать архитектуру нейронной сети, подать на вход разное количество обучающей выборки в формате вопрос - ответ, сделать визуализацию обучения нейронной сети, написать выводы, предложить методы оптимизации нейронной сети.

Самостоятельная работа (4 часа).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Чат-боты и генерация текста. Особенности построения архитектуры нейронной сети	Архитектура нейронной сети	Проработка дополнительных источников информации	Хливненко Л.В., Пятакович Ф.А. Практика нейросетевого моделирования: учебное пособие / Хливненко Л.В., Пятакович Ф.А. – 2-е изд., стер. – СПб.: Лань, 2021. – 196 с. (https://library.bmstu.ru/Catalog/Details/556123)	тест

Тема 9. Сегментация текстовых данных (4 часа)

Лекции (0,5 часа). В рамках данной темы будут рассмотрены основные подходы к разметке и подготовке обучающей выборки, при создании архитектур нейронных сетей, для задач сегментации текстовых данных.

Практическая работа (1,5 часа). Загрузить датасет (датасет содержит 1000 размеченных договоров), подготовить датасет, выполнить препроцессинг текстовых данных, использовать архитектуру нейронной сети, подать на вход разное количество сегментируемых классов, сделать визуализацию обучения нейронной сети, написать выводы, предложить методы оптимизации нейронной сети.

Самостоятельная работа (2 часа).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Сегментация текстовых данных	Сегментация данных	Проработка дополнительных источников информации	Бериков В.Б. Методы кластерного анализа данных и сегментации изображений: учеб. пособие / Бериков В.Б.; Новосибирский гос. ун-т. – Новосибирск: Изд-во РИЦ НГУ,	тест

			2015. – 97 с. (https://library.bmstu.ru/Catalog/Details/475417)	
--	--	--	--	--

Тема 10. Контроль версионности моделей с tensorflow serving (5,5 часов)

Лекции (0,5 часа). В рамках данной темы будут рассмотрены вопросы контроля версионности моделей, с применением docker образа tensorflow-serving, и вызов необходимых версий по средствам API.

Практические занятия (2 часа). Загрузить датасет, обучить модель, сохранить в формате save_model. Обучить вторую модель, также сохранить её. Написать программу, которая будет проверять на входе цветное изображение или нет, и отправлять на распознавание в соответствующую версию.

Самостоятельная работа (3 часа).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Контроль версионности моделей с tensorflow serving	Версионность моделей	Проработка дополнительных источников информации	Злобин В.К., Ручкин В.Н. Нейросети и нейрокомпьютеры: учеб. пособие для вузов / Злобин В.К., Ручкин В.Н. – СПб.: БХВ-Петербург, 2011. – 252 с. (https://library.bmstu.ru/Catalog/Details/246505)	тест

Тема 11. Развёртывание облачной инфраструктуры. Обзор облачных платформ Google Cloud Platform, AWS, Sbercloud (5,5 часов)

Лекции (0,5 часа). В рамках данной темы, будет дан обзор основных поставщиков облачных решений, как иностранные, так и отечественные. Особенности тарификации и развёртывания облачной инфраструктуры.

Практические занятия (1 час). Развернуть простое облако на платформе GCP с одной виртуальной машиной и одним хранилищем SSD. Сделать скриншоты рабочего instance.

Самостоятельная работа (4 часа).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Развёртывание облачной инфраструктуры. Обзор облачных платформ Google Cloud Platform,	Облачные хранилища	Проработка дополнительных источников информации	Докучаев В.А., Кальфа А.А., Маклачкова В.В. Архитектура центров обработки данных /	устный опрос

AWS, Sbercloud			Докучаев В.А., Кальфа А.А., Маклачкова В.В.; ред. Докучаев В.А. – М.: Горячая линия - Телеком, 2021. – 240 с. (https://library.bmstu.ru/Catalog/Details/555242)	
----------------	--	--	--	--

Тема 12. Flask приложение. Выведение моделей в production (7 часов)

Лекции (1 час). В рамках данной темы будет описан pipeline выведения модели машинного обучения в продуктив. Создание сервисов на базе фреймворка flask. Будут разобраны основные этапы, etl процессы и взаимодействие.

Практические занятия (2 часа). Выведение моделей в продуктив.

Самостоятельная работа (4 часа).

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Flask приложение. Выведение моделей в production	Flask-приложения	Проработка дополнительных источников информации	Хливненко Л.В., Пятакович Ф.А. Практика нейросетевого моделирования: учебное пособие / Хливненко Л.В., Пятакович Ф.А. – 2-е изд., стер. – СПб.: Лань, 2021. – 196 с. (https://library.bmstu.ru/Catalog/Details/556123)	домашнее задание

4.1.5. Оценочное средство для текущего контроля (темы для подготовки к тесту и устному опросу, формулировка условий практических заданий):

Тема 1. Тест: рекуррентные нейронные сети. LSTM слои (тест считается сданным при верном ответе на 4 из 5 вопросов).

1. Что является отличительной особенностью рекуррентных нейронных сетей?

- A. наличие перекрестных связей между нейронами одного слоя
- B. это сети только прямого распространения
- C. наличие связей «каждый с каждым»
- D. наличие обратных связей в архитектуре сети

2. Для решения каких задач из перечисленных наиболее подходят рекуррентные нейронные сети (RNN)?

- A. анализ текстов
- B. анализ речи
- C. анализ изображений
- D. компьютерное зрение
- E. машинный перевод

3. Чем различаются рекуррентные нейронные сети Джордана и Элмана?

- A. количеством слоев,
- B. в сети Элмана отсутствуют обратные связи,
- C. местом отведения обратных связей (слоем),
- D. наличием в сети Джордана не только прямых и обратных связей, но еще и перекрестных.

4. К какому классу обучения относятся РНС Хопфилда и Коско?

- A. с учителем,
- B. самоорганизующиеся,
- C. сеть Коско самоорганизующаяся, а сеть Хопфилда обучается с учителем,
- D. сеть Хопфилда самоорганизующаяся, а сеть Коско обучается с учителем

5. Что такое «вентиль забывания» и в каком типе рекуррентных нейросетей он встречается?

- A. вентиль забывания контролирует меру сохранения значений весовых коэффициентов РНС Хеминга,
- B. вентиль забывания нужен для перенастройки весов выходного слоя в РНС Джордана и Элмана,
- C. вентиль забывания предназначен для «затирания» из автоассоциативной памяти образа, который нам больше не нужен, т.е. используется в РНС Хопфилда,
- D. вентиль забывания контролирует меру сохранения значения в памяти РНС LSTM.

Тема 2. Тест: Tensorflow, построение нейронных сетей на уровне графов (тест считается сданным при верном ответе на 4 из 5 вопросов).

1. Что такое Tensorflow?

- A. Среда, созданная на основе python для проектирования, создания и изучения моделей глубокого обучения
- B. Это программная библиотека для численных расчетов с использованием графиков потоков данных
- C. Направление для проектирования, создания и изучения моделей глубокого обучения
- D. Язык для проектирования, создания и изучения моделей глубокого обучения

2. Что такое тензор?

- A. Представление математических операций в Tensorflow
- B. Это узлы, соединяющие ребра
- C. Это центральная единица данных в Tensorflow
- D. Это только одномерные массивы данных

3. **Что в графическом отображении представляют математические операции?**
- A. Узлы
 - B. Связки
 - C. Ребра
 - D. Хрящи
4. **Что такое графы?**
- A. Основа Tensorflow на которой находятся все вычисления и переменные
 - B. Основа Tensorflow на которой находятся только все вычисления
 - C. Основа Tensorflow на которой находятся только все переменные
 - D. Основа Tensorflow на которой находятся только все результаты вычислений
5. **Что делает объект Session?**
- A. Объект Session - это ни на что не влияющий объект, создающийся при каждом запуске сессии
 - B. Объект Session - это буквально то рабочее пространство в среде, в котором вы сейчас работаете с любыми другими объектами
 - C. Объект Session инкапсулирует среду, в которой выполняются объекты Operation, и оцениваются объекты Tensor

Тема 3. Тест: обзор библиотеки PyTorch. Особенности построения нейронных сетей (тест считается сданным при верном ответе на 4 из 5 вопросов).

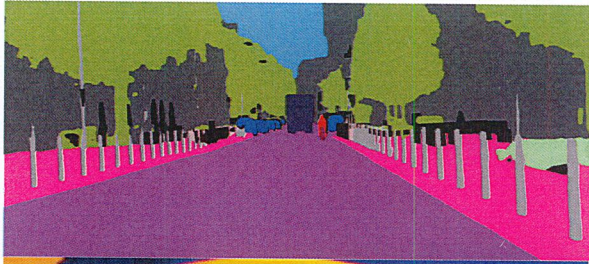
1. **Что такое PyTorch?**
- A. Современная библиотека глубокого обучения от Facebook
 - B. Современная библиотека глубокого обучения от Google
 - C. Библиотека поддержки работы Tensorflow разработанная Google
 - D. Библиотека, созданная на основе Tensorflow для глубокого обучения
2. **Выберете основные характеристики PyTorch (Несколько вариантов ответа)**
- A. Простой интерфейс
 - B. Использование Python
 - C. Вычислительные графы
 - D. Сложный интерфейс
 - E. Отсутствие графов
 - F. Использование быстросействующих функций вычисления
3. **Что из списка является преимуществами PyTorch? (Несколько вариантов ответа)**
- A. Код легко отлаживать и понимать
 - B. Большой функционал настроек

- C. Много метрик
 - D. Интеграция с numpy
 - E. Сложные структуры сетей
4. **Какой фреймворк является старше?**
- A. PyTorch является более старшим, чем Tensorflow
 - B. Tensorflow является более старшим, чем PyTorch
 - C. PyTorch и Tensorflow являются ровесниками
5. **Как определяется граф вычислений в PyTorch?**
- A. Определение графа перед началом обучения
 - B. Определение графа происходит во время обучения
 - C. Определение графа происходит после обучения

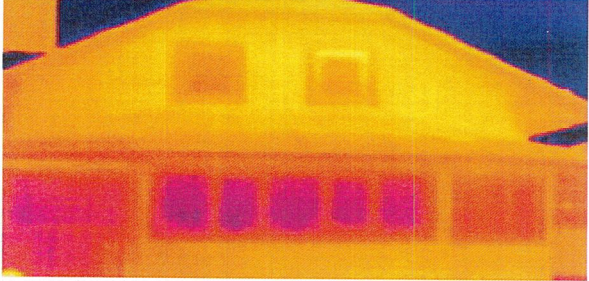
Тема 4. Тест: архитектуры нейронной сети для задач object detection (тест считается сданным при верном ответе на 4 из 5 вопросов).

1. **Какая задача решается в направлении object detection?**
- A. Поиск на видео или на изображении групп пикселей, характеризующих объект.
 - B. Создание архитектур, помогающих в фото и видео редактировании.
 - C. Генерация новых объектов для дальнейшего их внедрения на фото или видео.
 - D. Обнаружение координат объектов на изображении или видео и их классификация.
2. **Какие слои имеются в большем количестве в нейронной сети с задачей object detection? (Несколько вариантов ответа)**
- A. Слои нормализации
 - B. Enbending - Слои
 - C. Свёрточные слои
 - D. Основные слои (Dense, Dropout, Reshape и т.д.)
 - E. Pooling слои
3. **Какое изображение является результатом работы нейронной сети с задачей object detection?**





B.



C.

D. Не одно из показанных вариантов

4. Какие алгоритмы object detection из данного списка являются устаревшими?

- A. RCNN
- B. YOLO
- C. RetinaNet
- D. SPP
- E. FastER-RCNN

5. Что такое Anchors?

- A. Алгоритм, пытающийся найти объекты на изображении для создания ранних гипотез
- B. Алгоритм свёртки, создающий карту признаков для создания ранних гипотез.
- C. Определённые места на изображении или видео, вокруг которых создаются случайные Bounding Box для создания ранних гипотез.
- D. Алгоритм свёртки, создающий карту признаков на основе ранних гипотез.

Тема 5. Тест: архитектуры нейронной сети для задач segmentation (тест считается сданным при верном ответе на 4 из 5 вопросов).

1. Какую задачу выполняют модели в направлении segmentation?

- A. Восстановление повреждённого или обновления старого фото или видео.
- B. Обнаружение координат объектов на изображении или видео, их классификация.
- C. Генерация новых изображений и видео.
- D. Поиск на видео или изображении групп пикселей, характеризующих объект.

2. В чем главные минусы модели, решающей задачу segmentation? (Несколько вариантов ответа)

- A. Низкая практическая ценность итогов работы.
- B. Большие требования к вычислительным мощностям на этапе обучения.

- C. Большие требования к вычислительным мощностям на этапе работы в продакшене.
- D. Необходимость писать мощные классификаторы для модели.
- E. Отсутствие возможности находить большое число классов на изображении или видео.

3. Как должна работать самописная метрика для задачи segmentation?

- A. Считаются площади зон разных пикселей у результата работы и сравниваются с значениями площадей у таргета
- B. Считается количество пикселей разного цвета у результата работы и сравнивается с количеством пикселей этих цветов у таргета
- C. Подсчитываются совпавшие по координатам и цветам пиксели у результата работы и таргета и выводится процент совпавших пикселей в метрике

4. В чем главное преимущество segmentation по сравнению с object detection?

- A. Возможность анализировать видео в большем количестве кадров.
- B. Возможность смотреть, какие объекты стоят на переднем плане, а какие находятся за этими объектами.
- C. Возможность классификации большего числа объектов.
- D. Возможность анализировать несколько изображений за один раз.

5. Какая архитектура модели является верной для задачи segmentation?

- A. Получение обучающей выборки => проход через блок свертки с уменьшением размерности => проход через блок свертки с увеличением размерности и получение всех результатов прошлого блока свёртки => восстановление изображения в сегментированном виде
- B. Получение обучающей выборки => проход через блок свёртки с уменьшением размерности => проход через блок свертки для формирования тензоров, обозначающих объекты => восстановление изображения в сегментированном виде из полученных тензоров
- C. Получение обучающей выборки => проход через блок свёртки с уменьшением размерности => проход через блок классификации => восстановление изображения в сегментированном виде
- D. Получение обучающей выборки => проход через блок свёртки с уменьшением размерности => проход через блок свертки для восстановления изображения в сегментированном виде

Тема 6. Устный опрос: задачи NLP. Препроцессинга текстовых неструктурированных данных. ✓

Тема 7. Тест: архитектуры нейронной сети для классификации текстов (тест считается сданным при верном ответе на 4 из 5 вопросов).

1. Использование функции softmax на последнем слое окажет эффект...

- A. Повышения качества работы обученной нейронной сети
- B. Не окажет никакого эффекта
- C. Понизит качество работы нейронной сети

- 2. У нас есть база данных с 10 писателями. Как можно построить архитектуру модели, чтобы использовать только бинарную классификацию?**
- A. Нужно создать свой слой включающий в себя функционал нескольких классификаторов
 - B. Нужно использовать специальный слой многократной бинарной классификации
 - C. Нужно сделать разветвление на 10 бинарных классификаторов для каждого писателя
 - D. Нужно сделать несколько моделей и специальными методами превратить в одну
- 3. Выберите, почему модели обучают на частях текстов, а не сразу на всём тексте. (Несколько вариантов ответа)**
- A. Если анализировать весь текст, то будет слишком большая нагрузка для вычислительных мощностей
 - B. Если подавать весь текст, модель будет хуже выхватывать признаки
 - C. При работе в продакшн, нейронная сеть не будет работать, ведь нужно будет искать текст для анализа размером поданного во время обучения текста
 - D. Обучение только на частях текста проходит гораздо быстрее, чем на всём тексте сразу
- 4. Возможно ли добиться хорошего результата классификации, используя только полносвязные слои?**
- A. Нет, модель не сможет выхватить нужное количество признаков для классификации
 - B. Да, это можно сделать на любых текстах
 - C. Да, но только не на сложном тексте и при малом количестве классов
- 5. Можно ли использовать автокодировщик для классификации текста?**
- A. Да, отличий в работе от обычных подходов нет
 - B. Да, но текст нужно подавать в виде 2D массива
 - C. Нет, это связано с природой текстовых данных

Тема 8. Тест: чат-боты и генерация текста. Особенности построения архитектуры нейронной сети (тест считается сданным при верном ответе на 4 из 5 вопросов).

- 1. Какая основная задача в генерации текста?**
- A. Создать новый текст любой формы даже неосмысленной
 - B. Давать подсказки при создании текста человеком
 - C. Выявить закономерности в имеющемся тексте и создать новый «осмысленный» текст
 - D. Выявление закономерности в имеющемся тексте для дальнейшего его использования в подходящих ситуациях
- 2. Какие есть основные задачи обработки естественного языка? (Несколько вариантов ответа)**
- A. Расщепление речи
 - B. Распознавание речи

- C. Анализ текста
- D. Синтез речевых правил
- E. Синтез речи
- F. Генерация разделённой речи
- G. Генерация текста

3. Что из списка входит в работу с речью у умных помощников (Алиса, сири и т.д.) (Несколько вариантов ответа)

- A. Распознавание речи
- B. Расщепление речи
- C. Синтез речи
- D. Генерация текста
- E. Генерация разделённой речи
- F. Распознавание намерений

4. Можно ли с помощью генерации текста создать литературное искусство?

- A. Да, такое возможно
- B. Нет, машина не может воспроизвести творческую работу

5. На выходе из модели, генерирующей текст мы получаем?

- A. Необработанные данные понятны только машине, их можно преобразовать в текст
- B. Готовые текстовые данные, которые понятны человеку
- C. Необработанные текстовые данные, которые нужно переформатировать, чтобы они стали понятными
- D. Нет верного варианта ответа

Тема 9. Тест: сегментация текстовых данных (тест считается сданным при верном ответе на 4 из 5 вопросов).

1. Что такое процесс сегментации текста?

- A. Создание нового текста на основе шума
- B. Разделение текстовых данных по значимым частям (слова, предложения, темы)
- C. Разделение слов или предложений к разным темам (одно и то же слово или предложение может относиться к нескольким темам сразу)
- D. Разделение текста, преобразованного в изображение, на важные признаки

2. Как нужно подготовить выборку обучения перед препроцессингом?

- A. Все области текстовых данных нужно разметить в соответствии с их темами с помощью тегов (<t1><t2>Программирование круто</t1>, это правда</t2>)
- B. Все слова нужно разделить по темам с помощью имеющихся в тесте разделителей (: ; . , /)
- C. Нужно разделить текст на отдельные переменные для обучения([[Отдыхать было хорошо][Река была такой холодной]])
- D. Текст нужно подавать в неизменённом виде

3. Важно ли поддерживать равное количество текста для каждого класса в выборке для обучения?

- A. Да, иначе модель не обучится хорошо определять каждый класс в равной степени
 - B. Нет, так как на практике редко встречаются равномерно распределенные данные. Значит модель нужно обучать на похожих по распределению данных
 - C. Не имеет значения, разница при обучении слишком незначительна
- 4. На этапе создания словаря с помощью токеназера, что мы можем сделать для удобства дальнейшего формирования таргета (yTraine)?**
- A. Сохранить словарь на случай, если его будет нужно вернуть к форме без изменений
 - B. Создать переменную с размером нашего словаря
 - C. Создать еще один словарь и поменять индекс со словами местами (dict1 = {'1' : 'привет', '2' : 'Отлично'} => dict2 = {'привет' : '1', 'Отлично' : '2'})
 - D. Создать свой слой для модели, в котором в векторном представлении будут храниться все параметры нашего словаря
- 5. Как формирует таргет (yTrain) для задачи сегментации текста?**
- A. Создаётся список тем, к которым присоединены соответствующие предложения
 - B. Создаётся список индексов слов всего текстового документа, каждому индексу списка присваивается вектор, описывающий принадлежавшую ему тему
 - C. Создаётся список векторов, описывающих тему на месте каждого слова документа
 - D. Создаётся список индексов слов всего текстового документа, который разделен внутри тегами для опознания темы

Тема 10. Тест: контроль версионности моделей с tensorflow serving (тест считается сданным при верном ответе на 4 из 5 вопросов).

1. Какая цель у TensorFlow Serving?

- A. Довести модели машинного обучения до продакшена
- B. Довести модели машинного обучения до лучшей версии обучения
- C. Довести модели машинного обучения до самой оптимизированной версии
- D. Довести модели машинного обучения до стартового рабочего состояния

2. Выберите верный вариант изменения версионности.

- A. Версии библиотек, используемых в программе, подкручиваются до актуальных вручную
- B. Версии библиотек, используемых в программе, подкручиваются до актуальных автоматически
- C. У работника, занесённого в БД может изменяться должность, у должности в свою очередь может меняться оклад

3. Какие области применения верны для TensorFlow Serving? (Несколько вариантов ответа)

- A. Контроль версионности
- B. Репозиторий моделей
- C. Обработка запросов и распараллеливание операций
- D. Маршрутизация запросов на predict

4. Можно ли реализовать TensorFlow Serving с Docker?

- A. Да, такая реализация является одной из основных
- B. Да, но только в очень редких случаях
- C. Нет, так сделать нельзя

5. Для чего используется формат SavedModel?

- A. Для сохранения архитектуры модели без значений весов
- B. Для сохранения промежуточных результатов работы модели, только для просмотра
- C. Для сохранения модели в состоянии, которое уже нельзя будет никак изменить
- D. Для сохранения и загрузки вашей модели - переменных, графа и метаданных графа

Тема 11. Устный опрос: развертывание облачной инфраструктуры. Обзор облачных платформ Google Cloud Platform, AWS, Sbercloud. ✓

Тема 12. Домашнее задание: выведение моделей в продуктив. ✓

5. Условия реализации ДПП

5.1. Организационные условия реализации ДПП

Наименование аудитории	Вид занятия	Наименование оборудования, программного обеспечения
Компьютерный класс	Лекции	Материальное обеспечение: компьютер, мультимедийный проектор, экран, доска, пишущий инструмент, Программное обеспечение: Anaconda
Компьютерный класс	Практические занятия	Материальное обеспечение: компьютер, мультимедийный проектор, экран, доска, пишущий инструмент, Программное обеспечение: Anaconda
Компьютерный класс	Самостоятельная работа	Материальное обеспечение: компьютер, мультимедийный проектор, экран, доска, пишущий инструмент, Программное обеспечение: Anaconda
Компьютерный класс	Итоговая аттестация	Материальное обеспечение: компьютер, мультимедийный проектор, экран, доска, пишущий инструмент, Программное обеспечение: Anaconda

5.2. Педагогические условия реализации ДПП

Реализация программы обеспечивается преподавательским составом, удовлетворяющим следующим условиям:

- наличие высшего образования, соответствующее профилю программы, из числа штатных преподавателей, или привлеченных на условиях почасовой оплаты труда;

- опыт практической деятельности в соответствующей сфере из числа штатных преподавателей или привлеченных на условиях почасовой оплаты труда.

5.3. Учебно-методическое обеспечение ДПП

1. Хливненко Л.В., Пятакович Ф.А. Практика нейросетевого моделирования: учебное пособие / Хливненко Л.В., Пятакович Ф.А. – 2-е изд., стер. – СПб.: Лань, 2021. – 196 с. (<https://library.bmstu.ru/Catalog/Details/556123>)
2. Галушкин А.И. Нейронные сети: основы теории / Галушкин А.И. – М.: Горячая линия - Телеком, 2010. – 496 с. (<https://library.bmstu.ru/Catalog/Details/219193>)
3. Ручкин В.Н., Костров Б.В., Свирина А.Г. Системы искусственного интеллекта. Нейросети и нейрокомпьютеры: учебник для вузов / Ручкин В.Н., Костров Б.В., Свирина А.Г. – М.: КУРС, 2018. – 282 с. (<https://library.bmstu.ru/Catalog/Details/509245>)
4. Зозуля Ю.И. Интеллектуальные нейросистемы / Зозуля Ю.И. – М.: Радиотехника, 2003. – 143 с. (<https://library.bmstu.ru/Catalog/Details/92733>)
5. Злобин В.К., Ручкин В.Н. Нейросети и нейрокомпьютеры: учеб. пособие для вузов / Злобин В.К., Ручкин В.Н. – СПб.: БХВ-Петербург, 2011. – 252 с. (<https://library.bmstu.ru/Catalog/Details/246505>)
6. Бериков В.Б. Методы кластерного анализа данных и сегментации изображений: учеб. пособие / Бериков В.Б.; Новосибирский гос. ун-т. – Новосибирск: Изд-во РИЦ НГУ, 2015. – 97 с. <https://library.bmstu.ru/Catalog/Details/475417>
7. Докучаев В.А., Кальфа А.А., Маклачкова В.В. Архитектура центров обработки данных / Докучаев В.А., Кальфа А.А., Маклачкова В.В.; ред. Докучаев В.А. – М.: Горячая линия - Телеком, 2021. – 240 с. (<https://library.bmstu.ru/Catalog/Details/555242>)

5.4. Методические рекомендации

ДПП построена по тематическому принципу, каждый раздел представляет собой логически завершённый материал.

Преподавание программы основано на личностно-ориентированной технологии образования, сочетающей два равноправных аспекта этого процесса: обучение и учение. Личностно-ориентированный подход развивается при участии слушателей в активной работе на практических занятиях. Личностно-ориентированный подход направлен, в первую очередь, на развитие индивидуальных способностей обучающихся, создание условий для развития творческой активности слушателя и разработке инновационных идей, а также на развитие самостоятельности мышления при решении учебных задач разными способами, нахождение рационального варианта решения, сравнения и оценки нескольких вариантов их решения и т.п. Это способствует формированию приемов умственной деятельности по восприятию новой информации, ее запоминанию и осознанию, созданию образов для сложных понятий и процессов, приобретению навыков поиска решений в условиях неопределенности.

Лекции проводятся для приобретения навыков реализации знаний в предметной области, с использованием активных методов обучения.

Практические занятия проводятся для приобретения навыков решения практических задач в предметной области. Задания, выполняемые на практических занятиях, выполняются с использованием активных и интерактивных методов обучения.

Самостоятельная работа слушателей предназначена для проработки дополнительных источников информации. При изучении ДПП предусмотрены следующие методы организации и осуществления учебно-познавательной деятельности:

- объяснительно-иллюстративный метод;
- репродуктивный метод;
- частично-поисковый метод.

- При изучении ДПП предусмотрены активные формы проведения занятий:
- управляемая дискуссия;
 - разбор конкретных ситуаций.

6. Формы итоговой аттестации ДПП

Итоговая аттестация проводится в форме зачета для проверки сформированности компетенций, полученных в рамках ДПП.

Зачет проводится в формате тестирования. Результатом зачета служат правильные ответы на вопросы билета.

По результатам итоговой аттестации обучающемуся выставляется оценка «ЗАЧТЕНО/НЕ ЗАЧТЕНО»:

Оценка «ЗАЧТЕНО» выставляется обучающемуся, который:

- ответил на 8 из 12 вопросов теста;
- продемонстрировал необходимые систематизированные знания и достаточную степень владения принципами предметной области программы, понимание их особенностей и взаимосвязь между ними в течение всего срока обучения по ДПП.

Оценка «НЕ ЗАЧТЕНО» ставится обучающемуся, который:

- ответил менее, чем на 8 из 12 вопросов теста;
- имеет крайне слабые теоретические и практические знания, обнаруживает неспособность к построению самостоятельных заключений.

7. Оценочные материалы итоговой аттестации

7.1. Паспорт комплекта оценочных средств

Предметы оценивания	Объекты оценивания	Показатели оценки
ПК-1. Способен участвовать в исследовательских и опытно-конструкторских разработках в области создания и совершенствования ИТ-систем	Ответы на вопросы	Количество правильных ответов
ОПК-3. Способен решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности	Ответы на вопросы	Количество правильных ответов

7.2. Комплект оценочных средств

7.2.1. Темы для подготовки к зачету:

1. Рекуррентные нейронные сети.
2. Построение нейронных сетей на уровне графов.

3. Особенности построения нейронных сетей.
4. Препроцессинг текстовых неструктурированных данных.
5. Чат-боты и генерация текста.
6. Сегментация текстовых данных.
7. Контроль версионности моделей.
8. Развёртывание облачной инфраструктуры.
9. Flask приложения.
10. Выведение моделей в production.

7.2.2. Пример билета:

7.2.3.

1 – Тип нейронных сетей, которые специализируются на обработке последовательностей это ...

- A – Полносвязные нейронные сети
- B – Рекуррентные нейронные сети
- C – Сверточные нейронные сети
- D – Последовательные нейронные сети

2 – В какой библиотеке не используются статические расчетные графы – определяемые заранее, сразу и окончательно?

- A – TensorFlow
- B – Caffe2
- C – MXNet
- D – PyTorch

3 – Какая из библиотек лучше справится с сколь угодно масштабной задачей машинного обучения?

- A – PyTorch
- B – TensorFlow
- C – PyTorch и Tensorflow одинаково хорошо справятся

4 – Какой слой более важен для задачи генерации где нужно учитывать контекст?

- A – Conv2D
- B – BatchNormalization
- C – LSTM
- D – MaxPooling

5 – Что из себя представляет задача сегментации изображения?

- A – Это процесс разделения цифрового изображения на несколько сегментов. Цель сегментации заключается в упрощении и/или изменении представления изображения, чтобы его было проще и легче анализировать
- B – Разделение объектов на изображения по группам или фактическое разделение на объекты с последующим сохранением объектов в отдельных изображениях
- C – Это процесс создания нового изображения без использования каких-либо данных. Проще говоря это генерация из ничего

6 – Для чего важен контроль версионности моделей?

- A – Это важно, так как с его помощью можно вносить все изменения по нужному графику
- B – С его помощью всегда можно вернуться к старым версиям и отслеживать все внесённые изменения. При некоторых вариантах можно даже отслеживать кто их вносил

С – Он помогает автоматически вводить новые изменения в продукт, который уже выведен в продакшэн

7 – Что обязательно нужно сделать с текстом перед тем как подавать его в модель для обучения?

A – Его нужно перевести на английский язык

B – Его нужно привести к виду, в котором не будет орфографических ошибок

C – Его нужно разделить на смысловые части

D – Его нужно оцифровать

8 – Что из вариантов ниже является облачной платформой? (Несколько вариантов ответа)

A – Google Cloud Platform

B – AWS

C – Sbercloud

D – Flask

E – NLP

9 – Процесс разбиения текста на части и последующее присвоение слову, букве или предложению своего цифрового кода называется

A – Кодинг

B – Классификация

C – Токенизация

D – Препроцессинг

10 – Определение поэты по тексту — это задача...

A – Классификации текста

B – Генерации текста

C – Сегментация текста

D – Глубокое текстовое погружение

11 – Процесс разделения письменного текста на осмысленные единицы, такие как слова, предложения или темы это...

A – Классификации текста

B – Генерации текста

C – Сегментация текста

D – Глубокое текстовое погружение

12 – Задача, в которой необходимо выделить несколько объектов на изображении посредством нахождения координат, их ограничивающих рамок и классификации этих ограничивающих рамок из множества заранее известных классов.

A – Detection picture

B – Class Detection

C – RGB detection

D – Object detection