

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»
(МФТИ, Физтех)

УТВЕРЖДАЮ

Ректор МФТИ

д-р физ.-мат. наук, профессор

Д.В. Ливанов



« 29 » января 2024 г.

**Дополнительная профессиональная
программа повышения квалификации**

«Инженер данных»

УГСН 09.00.00 Информатика и вычислительная техника
Направление подготовки 09.03.01 Информатика и вычислительная техника
ОКВЭД 63 - Деятельность в области информационных технологий

Москва 2024

СОДЕРЖАНИЕ

РАЗДЕЛ 1. Характеристика программы «Инженер данных»	3
1.1 Цель реализации программы	3
1.2 Характеристика новой квалификации и связанных с ней видов профессиональной деятельности, трудовых функций и/или уровней квалификации	3
1.3 Совершенствуемые и/или приобретаемые компетенции	4
1.4 Планируемые результаты обучения	5
1.5 Требования к уровню подготовки поступающего на обучение, необходимому для освоения программы	7
1.6 Форма, объем и режим обучения	8
РАЗДЕЛ 2. Содержание программы «Инженер данных»	9
2.1 Учебный план программы	9
2.2 Календарный учебный график	10
2.3 Рабочие программы дисциплин (модулей)	11
2.3.1 Рабочая программа Модуля 1	11
2.3.2 Рабочая программа Модуля 2	16
2.3.3 Рабочая программа Модуля 3	21
2.3.4 Рабочая программа Модуля 4	27
2.3.5 Рабочая программа Модуля 5	33
2.3.6 Рабочая программа Модуля 6	38
РАЗДЕЛ 3. Оценка качества освоения программы «Инженер данных»	44
РАЗДЕЛ 4. Организационно-педагогические условия реализации программы «Инженер данных»	48
4.1 Материально-технические условия и организационное обеспечение реализации дополнительной профессиональной программы повышения квалификации	48
4.2 Учебно-методическое обеспечение реализации дополнительной профессиональной программы повышения квалификации	49
4.3 Составители программы	50

РАЗДЕЛ 1. ХАРАКТЕРИСТИКА ПРОГРАММЫ

«Инженер данных»

Общая характеристика программы

Реализация программы повышения квалификации направлена на совершенствование и (или) получение новой компетенции, необходимой для профессиональной деятельности, и (или) повышение профессионального уровня в рамках имеющейся квалификации. Программа разрабатывается на основании установленных квалификационных требований профессионального стандарта 06.042 «Специалист по большим данным», требований федерального государственного образовательного стандарта высшего образования 09.03.01 «Информатика и вычислительная техника» к результатам освоения образовательной программы, требований к компетенциям Универсальной модели компетенций в сфере искусственного интеллекта, разработанная РЭУ им. Плеханова в 2021 году в рамках результата Федерального проекта «Разработаны образовательные программы бакалавриата и магистратуры по профилю «Искусственный интеллект», повышена квалификация преподавателей высшего образования в сфере искусственного интеллекта, обеспечена подготовка специалистов по программам высшего образования по профилю «Искусственный интеллект» по специальности «Инженер данных (Data Engineer)».

Дополнительная профессиональная программа повышения квалификации является модульной. Каждый модуль представляет собой завершённую программу, направленную на формирование/совершенствование у слушателей конкретных компетенций. Объём программы каждого модуля превышает 16 академических часов. Обучение по каждому модулю завершается промежуточной аттестацией для проверки сформированности заявленных компетенций.

Слушателям, успешно прошедшим итоговую аттестацию по дополнительной профессиональной программе повышения квалификации, выдается удостоверение установленного образца.

1.1 Цель реализации программы

Целью реализации дополнительной профессиональной программы повышения квалификации «Инженер данных» является получение и совершенствование компетенций, необходимых для профессиональной деятельности инженеров данных – специалистов, которые готовят качественные данные для обучения моделей искусственного интеллекта.

1.2. Характеристика новой квалификации и связанных с ней видов профессиональной деятельности, трудовых функций и/или уровней квалификации

Настоящая дополнительная профессиональная программа повышения квалификации разработана с учётом соответствующих ей профессиональных стандартов и направлена на совершенствование компетенций, необходимых для выполнения вида профессиональной деятельности «Инженер данных».

Таблица 1

Профессиональный стандарт 06.042 «Специалист по большим данным»	
Обобщенная трудовая функция А. Анализ больших данных с использованием существующей в организации методологической и технологической инфраструктуры	
Трудовая функция	Компетенции
1. Подготовка данных для проведения аналитических работ по исследованию больших данных	А/03.6

Программа повышения квалификации разрабатывается на основании квалификационных требований, установленных Единым квалификационным справочником должностей руководителей, специалистов и служащих (ЕКТС).

Настоящая дополнительная профессиональная программа разработана в соответствии с ФГОС ВО 09.03.01 «Информатика и вычислительная техника».

1.3 Характеристика совершенствуемых и/или получаемых новых компетенций, необходимых для профессиональной деятельности и/или повышения профессионального уровня в рамках имеющейся квалификации

Компетенции, формируемые и совершенствуемые в результате обучения, представлены в таблицах 2 и 3.

Таблица 2

№	Компетенция в соответствии с профессиональным стандартом 06.042 «Специалист по большим данным»	Код компетенции
1	Способен осуществлять подготовку данных для проведения аналитических работ по исследованию больших данных	A/03.6

таблица 3

№	Компетенции в соответствии с направлением подготовки 09.03.01 Информатика и вычислительная техника	Код компетенции
1.	Способен использовать современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности	ОПК-2

таблица 4

№	Компетенции в соответствии с универсальной моделью компетенций в сфере искусственного интеллекта*	Код компетенции
1.	Способен классифицировать и идентифицировать задачи искусственного интеллекта, выбирать адекватные методы и инструментальные средства решения задач искусственного интеллекта	ПК-1.p
2.	Способен разрабатывать и применять методы машинного обучения для решения задач	ПК-4.p
3.	Способен использовать инструментальные средства для решения задач машинного обучения	ПК-5.p
4.	Способен создавать и поддерживать системы искусственного интеллекта на основе нейросетевых моделей и методов	ПК-6.p
5.	Способен разрабатывать системы анализа больших данных	ПК-8.p
6.	Способен создавать и внедрять одну или несколько сквозных цифровых субтехнологий искусственного интеллекта	ПК-9.p

*Универсальная модель компетенций в сфере искусственного интеллекта, разработанная РЭУ им. Плеханова в 2021 году в рамках результата Федерального проекта «Разработаны образовательные программы бакалавриата и магистратуры по профилю «Искусственный интеллект», повышена квалификация преподавателей высшего образования в сфере искусственного интеллекта, обеспечена подготовка специалистов по программам высшего образования по профилю «Искусственный интеллект»

1.4. Планируемые результаты обучения

Планируемые результаты обучения соответствуют результатам освоения соответствующей основной профессиональной образовательной программы.

Планируемые результаты обучения представлены в таблице 5.

таблица 5

	Уметь - знать	09.03.01 «Информатика и вычислительная техника»
		Квалификация: бакалавр
		Код компетенции
1	<p>Знать:</p> <ul style="list-style-type: none"> - современные информационные технологии и программные средства, в том числе отечественного производства при решении задач профессиональной деятельности <p>Уметь:</p> <ul style="list-style-type: none"> - выбирать современные информационные технологии и программные средства, в том числе отечественного производства при решении задач профессиональной деятельности <p>Владеть:</p> <p>Методиками использования программных средств для решения практических задач</p>	ОПК-2
		Профессиональный стандарт 06.042 «Специалист по большим данным»
		Код компетенции
2	<p>Знать:</p> <ul style="list-style-type: none"> - Современные методы и инструментальные средства анализа больших данных <p>Уметь:</p> <ul style="list-style-type: none"> - Осуществлять взаимодействие с внутренними и внешними поставщиками данных из гетерогенных источников <p>Владеть:</p> <p>Инструментами определения источников больших данных для анализа, идентификации внешних и внутренних источников данных для проведения аналитических работ; получения и фильтрации больших объемов данных из гетерогенных источников; извлечения, проверки и очистки больших объемов данных из гетерогенных источников; агрегации и разработки представления больших объемов данных из гетерогенных источников; оценки соответствия набора данных предметной области и задачам аналитических работ.</p>	A/03.6
		Универсальная модель компетенций в сфере искусственного интеллекта
		Код компетенции

3	<p>Знать:</p> <ul style="list-style-type: none"> - Типы задач искусственного интеллекта - Методы и инструменты систем искусственного интеллекта <p>Уметь:</p> <p>ПК-1.1.р. Классифицирует и идентифицирует задачи систем искусственного интеллекта в зависимости от особенностей проблемной и предметной областей</p> <p>ПК-1.2.р. Выбирает методы и инструментальные средства искусственного интеллекта для решения задач в зависимости от особенностей проблемной области</p> <p>Владеть:</p> <p>Python, sklearn, HuggingFace, TensorFlow</p>	ПК-1.р
4	<p>Знать:</p> <ul style="list-style-type: none"> - Классы задач машинного обучения и особенности их применения <p>Уметь:</p> <p>ПК-4.1.р. Проводит анализ требований и определяет необходимые классы задач машинного обучения</p> <p>ПК-4.3.р. Принимает участие в оценке и выборе используемых методов машинного обучения</p> <p>Владеть:</p> <p>Python, sklearn, HuggingFace, TensorFlow</p>	ПК-4.р
5	<p>Знать:</p> <ul style="list-style-type: none"> - Классы задач машинного обучения и особенности их применения - Методы и алгоритмы машинного обучения <p>Уметь:</p> <p>ПК-5.1.р. Осуществляет оценку и выбор инструментальных средств для решения поставленной задачи</p> <p>ПК-5.2.р. Разрабатывает модели машинного обучения для решения задач</p> <p>ПК-5.3.р. Создает, поддерживает и использует системы искусственного интеллекта, включающие разработанные модели и методы, с применением выбранных инструментов машинного обучения</p> <p>Владеть:</p> <p>Python, sklearn, HuggingFace, TensorFlow, SQL</p>	ПК-5.р
6	<p>Знать:</p> <ul style="list-style-type: none"> - Классы задач машинного обучения и особенности их применения - Методы и алгоритмы машинного обучения <p>Уметь:</p> <p>ПК-6.1.р. Осуществляет оценку и выбор моделей искусственных нейронных сетей и инструментальных средств для решения поставленной задачи</p>	ПК-6.р

	<p>ПК-6.2.р. Разрабатывает системы искусственного интеллекта на основе моделей искусственных нейронных сетей и инструментальных средств Владеть: Python, sklearn, HuggingFace, TensorFlow, SQL</p>	
7	<p>Знать: - Типы задач искусственного интеллекта - Методы и инструменты систем искусственного интеллекта - Классы задач машинного обучения и особенности их применения Уметь: ПК-8.1.р. Разрабатывает программные компоненты извлечения, хранения, подготовки больших данных с учетом вариантов использования больших данных, определений, словарей и эталонной архитектуры больших данных ПК-8.2.р. Разрабатывает программные компоненты обработки, удаленной, распределенной и объединенной аналитики, использования результатов анализа, описания и управления качеством и достоверностью больших данных Владеть: Python, Apache Spark, ClickHouse</p>	ПК-8.р
8	<p>Знать: - Методы и инструменты обработки больших данных Уметь: ПК-9.1.р. Участвует в реализации проектов в области сквозной цифровой субтехнологии «Компьютерное зрение» ПК-9.2.р. Участвует в реализации проектов в области сквозной цифровой субтехнологии «Обработка естественного языка» ПК-9.3.р. Участвует в реализации проектов в области сквозной цифровой субтехнологии «Рекомендательные системы и системы поддержки принятия решений» ПК-9.4.р. Участвует в реализации проектов в области сквозной цифровой субтехнологии «Распознавание и синтез речи» ПК-9.5.р. Участвует в реализации проектов в области перспективных сквозных цифровых субтехнологий искусственного интеллекта Владеть: Python, sklearn, HuggingFace, TensorFlow, SQL, dvc, ClearML</p>	ПК-9.р

1.5 Требования к уровню подготовки поступающего на обучение, необходимого для освоения программы

Программа предназначена для специалистов, имеющих высшее или среднее профессиональное образование, а также лиц, получающих высшее или среднее профессиональное образование.

Для успешного освоения программы требуется квалификация «Разработчик начального или среднего уровня».

Для успешного освоения программы требуется опыт профессиональной деятельности в области разработки программного обеспечения на Python (или другом языке программирования) от 1 года.

Для успешного освоения программы необходимы знания начального уровня в следующих областях:

- Язык программирования Python.
- Реляционные базы данных.
- Компьютерные сети, в первую очередь протоколы HTTP и DNS.

Необходимо владеть следующими профессиональными компетенциями:

- Программирование на Python
- Работа с реляционными базами данных
- Работа с компьютерными сетями

Рекомендуется изучение курса по Python для искусственного интеллекта, например: <https://stepik.org/course/110361/promo>, или знания в объеме материалов данного курса

Для поступления на обучение абитуриенты проходят обязательное вступительное испытание, которое представляет собой тест из 20 вопросов.

Оценивается:

- знание языка программирования Python;
- знание реляционных баз данных и основ языка SQL;
- знание компьютерных сетей.

Используется 2-хбалльная система оценивания.

"Зачтено" – поступающий дал правильные ответы не менее, чем по 65% заданий теста.

"Не зачтено" – поступающий дал правильные ответы менее, чем по 65% заданий теста.

1.6 Форма, объем, режим и стоимость обучения

Форма обучения: Очно-заочная с использованием дистанционных образовательных технологий

Объем программы: 256 академических часов

Режим обучения: 4 месяца

Количество обучающихся на потоке: от 30 до 100 человек.

Стоимость обучения: 110 000 рублей.

РАЗДЕЛ 2. СОДЕРЖАНИЕ ПРОГРАММЫ

2.1. Учебный (тематический) план

Учебный (тематический) план программы представлен в таблице 6.

таблица 6

№ п/п	Наименование модулей, разделов и тем	Всего, час.	Лекции	Практич. работа	Аттестация	Самост. работа	Форма контроля
УЧЕБНЫЙ ПЛАН							
	Входное тестирование	1				1	
1.	Жизненный цикл систем искусственного интеллекта	25	4	8	2	11	Зачет
2.	Подготовка данных для систем искусственного интеллекта	32	4	16	2	10	Зачет
3.	Язык SQL и работа с базами данных	32	4	16	2	10	Зачет
4.	Основы машинного обучения и нейронные сети	44	8	18	2	16	Зачет
5.	Инструменты работы с большими данными	36	4	10	2	20	
6.	Основы автоматизация машинного обучения MLOps	36	6	14	2	14	
	Итоговая аттестация	50			4	46	Защита проекта
	ИТОГО	256	30	82	16	128	
УЧЕБНО-ТЕМАТИЧЕСКИЙ ПЛАН							
	Входное тестирование	1				1	
Модуль 1. Жизненный цикл систем искусственного интеллекта							
1.1	Жизненный цикл систем искусственного интеллекта	11	2	4		5	
1.2	Роль данных в жизненном цикле проекта искусственного интеллекта	12	2	4		6	
	Промежуточная аттестация по модулю 1	2			2		Зачет
	Итого по модулю 1	25	4	8	2	11	
Модуль 2. Подготовка данных для систем искусственного интеллекта							
2.1	Работа с данными в Python	14	2	8		4	
2.2	Загрузка данных из интернет и социальных сетей	16	2	8		6	
	Промежуточная аттестация по модулю 2	2			2		Зачет
	Итого по модулю 2	32	4	16	2	10	
Модуль 3. Язык SQL и работа с базами данных							
3.1	SQL и реляционные базы данных	14	2	8		4	
3.2	Работа с базами данных из Python	16	2	8		6	
	Промежуточная аттестация по модулю 3	2			2		Зачет
	Итого по модулю 3	32	4	16	2	10	
Модуль 4. Основы машинного обучения и нейронные сети							
4.1	Машинное обучение	6	2	2		2	
4.2	Основы классических алгоритмов машинного обучения	14	2	6		6	

4.3	Классические алгоритмы машинного обучения	6	2	2		2	
4.4	Основы нейронных сетей	16	2	8		6	
	Промежуточная аттестация по модулю 4	2			2		Зачет
	Итого по модулю 4	44	8	18	2	16	
Модуль 5. Инструменты работы с большими данными							
5.1	Apache Spark	18	2	6		10	
5.2	База данных Clickhouse	16	2	4		10	
	Промежуточная аттестация по модулю 5	2			2		Зачет
	Итого по модулю 5	36	4	10	2	20	
Модуль 6. Основы автоматизация машинного обучения MLOps							
6.1	Автоматизация машинного обучения MLOps	6	2	2		2	
6.2	Управление версиями кода и данных	12	2	4		6	
6.3	MLOps платформы	16	2	8		6	
	Промежуточная аттестация по модулю 6	2			2		Зачет
	Итого по модулю 6	36	6	14	2	14	
	Итоговая аттестация по программе	50			4	46	Защита проекта
	Итого:	256	30	82	16	128	

2.2. КАЛЕНДАРНЫЙ УЧЕБНЫЙ ГРАФИК

Календарный учебный график составляется при сформированной группе с учетом уровня подготовки слушателей.

Календарный учебный график отражает периоды теоретических занятий, процедур промежуточной и итоговой аттестаций и т.д.

КАЛЕНДАРНЫЙ УЧЕБНЫЙ ГРАФИК

Таблица 7

Учебные недели	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Учебные занятия														
Самостоятельная работа														
Промежуточная аттестация														
Итоговая аттестация														

2.3. РАБОЧИЕ ПРОГРАММЫ ДИСЦИПЛИН (МОДУЛЕЙ)

Рабочие программы дисциплин (модулей) отражают дисциплинарное содержание дополнительной профессиональной программы повышения квалификации.

Модуль 1. Жизненный цикл систем искусственного интеллекта

Модуль 2. Подготовка данных для систем искусственного интеллекта

Модуль 3. Язык SQL и работа с базами данных

Модуль 4. Основы машинного обучения и нейронные сети

Модуль 5. Инструменты работы с большими данными

Модуль 6. Основы автоматизация машинного обучения MLOps

2.3.1. РАБОЧАЯ ПРОГРАММА МОДУЛЯ 1

1. Цель и задачи освоения Модуля 1

Целью изучения модуля «Жизненный цикл систем искусственного интеллекта» является знакомство с жизненным циклом систем искусственного интеллекта.

Задачи Модуля 1:

1. Познакомиться с основными этапами жизненного цикла систем искусственного интеллекта.
2. Рассмотреть примеры проектов создания систем искусственного интеллекта и их продвижения по жизненному циклу.
3. Познакомиться с подходом к созданию систем искусственного интеллекта Data-centric AI.

2. Характеристика совершенствуемых и/или получаемых новых компетенций, необходимых для профессиональной деятельности и/или повышения профессионального уровня в рамках имеющейся квалификации Модуль 1 «Жизненный цикл систем искусственного интеллекта»

Таблица 8.1

№	Компетенция в соответствии с профессиональным стандартом 06.042 «Специалист по большим данным»	Код компетенции
1	Способен осуществлять подготовку данных для проведения аналитических работ по исследованию больших данных	A/03.6

Таблица 9.1

№	Компетенции в соответствии с направлением подготовки 09.03.01 Информатика и вычислительная техника	Код компетенции
1.	Способен использовать современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности	ОПК-2

Таблица 10.1

№	Компетенции в соответствии с универсальной моделью компетенций в сфере искусственного интеллекта*	Код компетенции
1.	Способен классифицировать и идентифицировать задачи искусственного интеллекта, выбирать адекватные методы и инструментальные средства решения задач искусственного интеллекта	ПК-1.р.

2.2. Учебная программа

Содержание учебной программы приведено в таблице 9.

таблица 11.1

№ п/п	Наименование модуля, разделов и тем	Содержание обучения, наименование и тематика практических занятий (семинаров), самостоятельной работы	Объем, ак.час.
Модуль 1 Жизненный цикл систем искусственного интеллекта			
1.1	Жизненный цикл систем искусственного интеллекта	Лекция Этапы жизненного цикла систем искусственного интеллекта. Состояние отрасли разработки систем искусственного интеллекта. Роль данных в жизненном цикле проекта искусственного интеллекта. Практическая работа Рассмотрение кейсов жизненного цикла систем искусственного интеллекта из различных предметных областей. Самостоятельная работа Самостоятельный поиск и разбор кейсов жизненного цикла разработки систем искусственного интеллекта.	11
1.2	Роль данных в жизненном цикле проекта искусственного интеллекта	Лекция Влияние объема и качество данных для обучения на реализацию проектов искусственного интеллекта. Подход к реализации проектов искусственного интеллекта Data-centric AI. Практическая работа Рассмотрение кейсов применения подхода Data-centric AI для повышения качества работы систем на основе искусственного интеллекта. Самостоятельная работа Самостоятельный поиск и разбор кейсов использования данных для повышения качества работы систем искусственного интеллекта без усложнения используемых моделей машинного обучения.	12
	Промежуточная аттестация по модулю 1	зачет	2
	Итого:		25

3. Формы аттестации и оценочные материалы

3.1. Формы аттестации

Промежуточная аттестация по модулю является обязательной.

Оценка качества освоения модуля проводится по 2-х-балльной системе по результатам текущего контроля (тестирование, проверочные задания на взаимную оценку), контроля посещаемости практических занятий и результатам промежуточной аттестации.

Форма промежуточной аттестации: зачет.

Слушатель считается аттестованным в случае положительных результатов работы (не менее 65% баллов от итоговой оценки) в процессе обучения и успешной сдачи зачета.

Критерии оценивания:

- выполнение / не выполнение практических заданий по темам лекции (1 балл за каждое задание);
- выполнение / не выполнение итогового задания по модулю (1/0 баллов соответственно).

таблица 12.1/1

Наименование модуля, разделов и тем	Основные показатели оценки	Формы и методы контроля и оценки
Жизненный цикл систем искусственного интеллекта	ПК-1.р, ОПК-2, А/03.6	Зачет

Оценивание

таблица 12.1/2

Критерий	Соответствует	Не соответствует
выполнение / не выполнение практических заданий по теме лекции	1 балл	0 баллов
выполнение / не выполнение итоговой задачи	1 балл	0 баллов

Оценивание проводится преподавателем на основе представленных критериев и шкалы оценки.

Зачет выставляется слушателю, если по результатам оценивания слушатель получает не менее 2 баллов.

3.2. Оценочные материалы

таблица 13.1

Наименование раздела, темы	Основные показатели оценки	Формы и методы контроля и оценки	Вес задания
Жизненный цикл систем искусственного интеллекта	ПК-1.р, ОПК-2, А/03.6	Практические задания по темам лекций	25
Роль данных в жизненном цикле проекта искусственного интеллекта	ПК-1.р, ОПК-2, А/03.6	Практические задания по темам лекций	25
Промежуточная аттестация	ПК-1.р, ОПК-2, А/03.6	Итоговое задание по модулю	50

Практические задания

1. Найдите в интернет примеры успешных кейсов разработки и внедрения системы искусственного интеллекта. Опишите жизненный цикл разработки и внедрения такой системы.

2. Поиск и разбор кейсов использования данных для повышения качества работы систем искусственного интеллекта без усложнения используемых моделей машинного обучения.

Примеры заданий для проведения промежуточной аттестации по Модулю 1

1. Разбор кейса создания системы искусственного интеллекта для выявления «сомнительных» клиентов банка (<https://rb.ru/opinion/somnitelnye-klienty/>). Исследование данных, которые могут быть полезны для решения задачи кейса.

4. Организационно-педагогические условия реализации программы

4.1. Учебно-методическое обеспечение и информационное обеспечение программы

4.1.1. Список литературы:

Основная литература:

1. Daswin De Silva, Daminda Alahakoon. An artificial intelligence life cycle: From conception to production // Patterns, Volume 3, Issue 6, 2022, ISSN 2666-3899, <https://doi.org/10.1016/j.patter.2022.100489>.

2. Andrew Ferlitsch. Making the machine: the machine learning lifecycle. <https://cloud.google.com/blog/products/ai-machine-learning/making-the-machine-the-machine-learning-lifecycle>

3. Data-centric AI Resource Hub. <https://datacentricai.org/>

4.1.2. Электронные информационные ресурсы

1. Андрей Созыкин. Анализ больших данных помогает нам выявлять «сомнительных» клиентов. <https://rb.ru/opinion/somnitelnye-klienty/>

2. Why do 87% of data science projects never make it into production? <https://venturebeat.com/ai/why-do-87-of-data-science-projects-never-make-it-into-production/>

3. Data-centric AI Resource Hub. <https://datacentricai.org/>

4.1.3. Электронные образовательные ресурсы

Управление проектами по анализу данных <https://sberuniversity.ru/learning/courses/digital-skills/upravlenie-proektami-po-analizu-dannykh/>

Дополнительная литература:

1. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных – <http://www.machinelearning.ru/>

4.2. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

таблица 14.1

Наименование специализированных аудиторий, кабинетов, лабораторий	Вид занятий	Наименование оборудования, программного обеспечения
Аудитория	Лекция	Ноутбук / стационарный компьютер, гарнитура Программное обеспечение: VK Звонки / Яндекс телемост, Браузер Chrome или Яндекс.Браузер
Аудитория	Практическая работа	Ноутбук / стационарный компьютер, гарнитура Программное обеспечение: VK Звонки / Яндекс телемост, Браузер Chrome или Яндекс.Браузер

Аудитория	Промежуточная аттестация	Ноутбук / стационарный компьютер, гарнитура Программное обеспечение: VK Звонки / Яндекс телемост, Браузер Chrome или Яндекс.Браузер
-----------	--------------------------	--

4.3. Организация образовательного процесса

Таблица 15.1

Методы обучения	Методы контроля	Формы организации учебных занятий	Формы организации учебной деятельности	Дистанционные образовательные технологии
Лекция Практическая работа под руководством преподавателя Самостоятельная практическая работа Изучение литературы по теме	Выполнение практических занятий по темам лекций Выполнение итогового задания по модулю	Лекция Вебинар с элементами практической работы и разбора теоретического материала	Групповая работа Индивидуальная работа	Использование образовательных интернет-ресурсов Использование ресурсов, созданных преподавателем WEB-консультации и другие

2.3.2. РАБОЧАЯ ПРОГРАММА МОДУЛЯ 2

1. Цель и задачи освоения Модуля 2

Целью изучения модуля «Подготовка данных для систем искусственного интеллекта» является формирование навыков использования Python для подготовки данных для систем искусственного интеллекта.

Задачи Модуля 2:

1. Познакомиться с библиотеками работы с данными в Python.
2. Научиться работать с файлами разного типа в Python.
3. Научиться загружать данные из интернет и социальных сетей.

2. Характеристика совершенствуемых и/или получаемых новых компетенций, необходимых для профессиональной деятельности и/или повышения профессионального уровня в рамках имеющейся квалификации Модуль 2 «Подготовка данных для систем искусственного интеллекта»

Таблица 8.2

№	Компетенция в соответствии с профессиональным стандартом 06.042 «Специалист по большим данным»	Код компетенции
1	Способен осуществлять подготовку данных для проведения аналитических работ по исследованию больших данных	A/03.6

Таблица 9.2

№	Компетенции в соответствии с направлением подготовки 09.03.01 Информатика и вычислительная техника	Код компетенции
1.	Способен использовать современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности	ОПК-2

Таблица 10.2

№	Компетенции в соответствии с универсальной моделью компетенций в сфере искусственного интеллекта*	Код компетенции
1.	Способен классифицировать и идентифицировать задачи искусственного интеллекта, выбирать адекватные методы и инструментальные средства решения задач искусственного интеллекта	ПК-1.р.

2.2. Учебная программа

Содержание учебной программы приведено в таблице 11.

таблица 11.2

№ п/п	Наименование модуля, разделов и тем	Содержание обучения, наименование и тематика практических занятий (семинаров), самостоятельной работы	Объем, ак.час.
Модуль 2 Подготовка данных для систем искусственного интеллекта			
2.1	Работа с данными в Python	Лекция Работа с файлами разного типа в Python. Библиотеки для работы с данными Pandas, numpy.	14

		<p>Практическая работа Работа с файлами и каталогами в Python. Открытие и закрытие файлов. Чтение и запись файлов. Хранение файлов, работа с каталогами, модули pathlib и os.path. Работа с файлами CSV. Работа с файлами JSON. Сохранение и восстановление структур данных, библиотечный модуль pickle. Библиотеки Pandas и numpy для работы с таблицами в Python.</p> <p>Самостоятельная работа Обработка данных из кейса поиска «сомнительных» клиентов банка: единый государственный реестр юридических лиц (ЕГРЮЛ), единый государственный реестр индивидуальных предпринимателей (ЕГРИП), финансовая отчетность от Росстата и ФНС, общероссийский классификатор видов экономической деятельности ОКВЭД 2</p>	
2.2	Загрузка данных из интернет и социальных сетей	<p>Лекция Работа с сетью на Python. Загрузка Web-страниц. Парсинг Web-страниц. Использование REST API. Загрузка данных из социальных сетей с помощью REST API.</p> <p>Практическая работа Web-страницы, HTML. Протокол HTTP. Загрузка Web-страниц из интернет, библиотеки requests и httpx. Распознавание Web-страниц, библиотека BeautifulSoup. Сохранение распознанных данных. Использование REST API в Python. Инструмент для работы с API Postman. Работа по API с популярными сайтами (hh.ru) и социальными сетями (vk.com). Авторизация в REST API.</p> <p>Самостоятельная работа Загрузка данных о вакансиях Инженеров данных с сайт hh.ru через API с помощью Python. Поиск в соцсети vk.com пользователей, интересующихся искусственным интеллектом.</p>	16
	Промежуточная аттестация по модулю 2	зачет	2
	Итого:		32

3. Формы аттестации и оценочные материалы

3.1. Формы аттестации

Промежуточная аттестация по модулю является обязательной.

Оценка качества освоения модуля проводится по 2-х-балльной системе по результатам текущего контроля (тестирование, проверочные задания на взаимную оценку), контроля посещаемости практических занятий и результатам промежуточной аттестации.

Форма промежуточной аттестации: зачет.

Слушатель считается аттестованным в случае положительных результатов работы (не менее 65% баллов от итоговой оценки) в процессе обучения и успешной сдачи зачета.

Критерии оценивания:

- выполнение / не выполнение практических заданий по темам лекции (1 балл за каждое задание);
- выполнение / не выполнение итогового задания по модулю (1/0 баллов соответственно).

таблица 12.2/1

Наименование модуля, разделов и тем	Основные показатели оценки	Формы и методы контроля и оценки
Подготовка данных для систем искусственного интеллекта	ПК-1.р, ОПК-2, А/03.6	Зачет

Оценивание

таблица 12.2/2

Критерий	Соответствует	Не соответствует
выполнение / не выполнение практических заданий по теме лекции	1 балл	0 баллов
выполнение / не выполнение итоговой задачи	1 балл	0 баллов

Оценивание проводится преподавателем на основе представленных критериев и шкалы оценки.

Зачет выставляется слушателю, если по результатам оценивания слушатель получает не менее 2 баллов.

3.2. Оценочные материалы

таблица 13.2

Наименование раздела, темы	Основные показатели оценки	Формы и методы контроля и оценки	Вес задания
Работа с данными в Python	ПК-1.р, ОПК-2, А/03.6	Практические задания по темам лекций	25
Загрузка данных из интернет и социальных сетей	ПК-1.р, ОПК-2, А/03.6	Практические задания по темам лекций	25
Промежуточная аттестация	ПК-1.р, ОПК-2, А/03.6	Итоговое задание по модулю	50

Практические задания

1. Обработка данных из кейса поиска «сомнительных» клиентов банка:
 - Единый государственный реестр юридических лиц (ЕГРЮЛ).
 - Единый государственный реестр индивидуальных - предпринимателей (ЕГРИП).
 - Финансовая отчетность от Росстата и ФНС.
 - Общероссийский классификатор видов экономической деятельности ОКВЭД 2
2. Загрузка данных о вакансиях Инженеров данных с сайт hh.ru через API с помощью Python.
3. Поиск в соцсети vk.com пользователей, интересующихся искусственным интеллектом.

Примеры заданий для проведения промежуточной аттестации по Модулю 2

1. Разработка программы на Python для сбора информации о компании или ИП из открытых источников, включая:

- Регистрационная информация
- Финансовая отчетность
- Социальные сети
- Сайты вакансий
- Популярные сайты по выбору обучающихся

4. Организационно-педагогические условия реализации программы

4.1 Учебно-методическое обеспечение и информационное обеспечение программы

4.1.1. Список литературы:

Основная литература:

1. Daswin De Silva, Daminda Alahakoon. An artificial intelligence life cycle: From conception to production // Patterns, Volume 3, Issue 6, 2022, ISSN 2666-3899, <https://doi.org/10.1016/j.patter.2022.100489>.

2. Andrew Ferlitsch. Making the machine: the machine learning lifecycle. <https://cloud.google.com/blog/products/ai-machine-learning/making-the-machine-the-machine-learning-lifecycle>

3. Data-centric AI Resource Hub. <https://datacentricai.org/>

4.1.2. Электронные информационные ресурсы

1. Открытые данные – <https://ofdata.ru/open-data>
2. Андрей Созыкин. Анализ больших данных помогает нам выявлять «сомнительных» клиентов. <https://rb.ru/opinion/somnitelnye-klienty/>
3. Pandas – <https://pandas.pydata.org/>
4. Numpy - <https://numpy.org/>
5. Requests – <https://requests.readthedocs.io/en/latest/>
6. httpx – <https://www.python-httpx.org/>
7. BeautifulSoup – <https://www.crummy.com/software/BeautifulSoup/>
8. Postman – <https://www.postman.com/>

4.1.3. Электронные образовательные ресурсы

Python для искусственного интеллекта – <https://stepik.org/course/110361/promo>

Дополнительная литература:

1. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных – <http://www.machinelearning.ru/>

4.2. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

таблица 14.2

Наименование специализированных аудиторий, кабинетов, лабораторий	Вид занятий	Наименование оборудования, программного обеспечения
Аудитория	Лекция	Ноутбук / стационарный компьютер, гарнитура

		Программное обеспечение: VK Звонки / Яндекс телемост, Браузер Chrome или Яндекс.Браузер Интерпретатор языка Python Библиотеки Pandas, numpy, requests, httpx, BeautifulSoup, Postman
Аудитория	Практическая работа	Ноутбук / стационарный компьютер, гарнитура Программное обеспечение: VK Звонки / Яндекс телемост, Браузер Chrome или Яндекс.Браузер Интерпретатор языка Python Библиотеки Pandas, numpy, requests, httpx, BeautifulSoup, Postman
Аудитория	Промежуточная аттестация	Ноутбук / стационарный компьютер, гарнитура Программное обеспечение: VK Звонки / Яндекс телемост, Браузер Chrome или Яндекс.Браузер Интерпретатор языка Python Библиотеки Pandas, numpy, requests, httpx, BeautifulSoup, Postman

4.3. Организация образовательного процесса

Таблица 15.2

Методы обучения	Методы контроля	Формы организации учебных занятий	Формы организации учебной деятельности	Дистанционные образовательные технологии
Лекция Практическая работа под руководством преподавателя Самостоятельная практическая работа Изучение литературы по теме	Выполнение практических занятий по темам лекций Выполнение итогового задания по модулю	Лекция Вебинар с элементами практической работы и разбора теоретического материала	Групповая работа Индивидуальная работа	Использование образовательных интернет-ресурсов Использование ресурсов, созданных преподавателем WEB-консультации и другие

2.3.3. РАБОЧАЯ ПРОГРАММА МОДУЛЯ 3

1. Цель и задачи освоения Модуля 3

Целью изучения модуля «Язык SQL и работа с базами данных» является научиться эффективно работать с реляционными базами данных.

Задачи Модуля 3:

1. Научиться выгружать данные из реляционных баз с помощью SQL.
2. Научиться работать с реляционными базами данных из Python.

2. Характеристика совершенствуемых и/или получаемых новых компетенций, необходимых для профессиональной деятельности и/или повышения профессионального уровня в рамках имеющейся квалификации Модуль 3 «Язык SQL и работа с базами данных»

Таблица 8.3

№	Компетенция в соответствии с профессиональным стандартом 06.042 «Специалист по большим данным»	Код компетенции
1	Способен осуществлять подготовку данных для проведения аналитических работ по исследованию больших данных	A/03.6

Таблица 9.3

№	Компетенции в соответствии с направлением подготовки 09.03.01 Информатика и вычислительная техника	Код компетенции
1.	Способен использовать современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности	ОПК-2

Таблица 10.3

№	Компетенции в соответствии с универсальной моделью компетенций в сфере искусственного интеллекта*	Код компетенции
1.	Способен использовать инструментальные средства для решения задач машинного обучения	ПК-5.р.

2.2. Учебная программа

Содержание учебной программы приведено в таблице 11.

таблица 11.3

№ п/п	Наименование модуля, разделов и тем	Содержание обучения, наименование и тематика практических занятий (семинаров), самостоятельной работы	Объем, ак. час.
Модуль 3 Язык SQL и работа с базами данных			
3.1	SQL и реляционные базы данных	Лекция Реляционные базы данных, таблицы. Язык SQL. Оператор SELECT. Фильтрация, группировка и сортировка в SQL. Объединение данных из нескольких таблиц. Индексы. Представления. Практическая работа	14

		<p>Реляционная база данных PostgreSQL. Создание баз данных и таблиц в PostgreSQL. Загрузка данных в PostgreSQL. Клиенты и инструменты разработки SQL: psql, pgAdmin, DataGrip. Выполнение SQL запросов. Фильтрация и сортировка данных в SQL. Группировка данных и агрегатные функции. Фильтрация результатов группировки, HAVING. Объединение данных из нескольких таблиц. Внутреннее и внешнее объединение. Левое, правое и полное внешнее объединение. Объединение данных из трех и более таблиц. Создание и использование представлений. Создание и использование индексов. Самостоятельная работа Загрузка в PostgreSQL открытых данных для кейса поиска «сомнительных» клиентов банка: – Единый государственный реестр юридических лиц (ЕГРЮЛ). – Единый государственный реестр индивидуальных предпринимателей (ЕГРИП). – Финансовая отчетность от Росстата и ФНС. – Общероссийский классификатор видов экономической деятельности ОКВЭД 2 Исследование открытых данных в PostgreSQL с помощью SQL.</p>	
3.2	Работа с базами данных из Python	<p>Лекция Работа с базами данных в Python. Библиотеки для запуска SQL запросов. Работа с базами данных из Pandas. Object-Relational Mapping (ORM) на Python. Практическая работа Python Database-API. Библиотеки для работы с разными базами данных: sqlite3 (SQLite), psycopg2 (PostgreSQL), mysql.connector (MySQL), pyodbc (работа через интерфейс ODBC). Запросы к данным в базе. Запись данных в базу. Изменение данных в базе. Использование курсоров. Работа с базами данных из Pandas. Реляционная и объектная модели данных. Object-Relational Mapping (ORM). Работа с SQLAlchemy в Python. Самостоятельная работа Исследование открытых данных для кейса поиска «сомнительных» клиентов банка в PostgreSQL на Python: – Единый государственный реестр юридических лиц (ЕГРЮЛ). – Единый государственный реестр индивидуальных предпринимателей (ЕГРИП). – Финансовая отчетность от Росстата и ФНС.</p>	16

		Общероссийский классификатор видов экономической деятельности ОКВЭД 2	
	Промежуточная аттестация по модулю 3	зачет	2
	Итого:		32

3. Формы аттестации и оценочные материалы

3.1. Формы аттестации

Промежуточная аттестация по модулю является обязательной.

Оценка качества освоения модуля проводится по 2-х-балльной системе по результатам текущего контроля (тестирование, проверочные задания на взаимную оценку), контроля посещаемости практических занятий и результатам промежуточной аттестации.

Форма промежуточной аттестации: зачет.

Слушатель считается аттестованным в случае положительных результатов работы (не менее 65% баллов от итоговой оценки) в процессе обучения и успешной сдачи зачета.

Критерии оценивания:

- выполнение / не выполнение практических заданий по темам лекции (1 балл за каждое задание);
- выполнение / не выполнение итогового задания по модулю (1/0 баллов соответственно).

таблица 12.3/1

Наименование модуля, разделов и тем	Основные показатели оценки	Формы и методы контроля и оценки
Язык SQL и работа с базами данных	ПК-5.р, ОПК-2, А/03.6	Зачет

Оценивание

таблица 12.3/2

Критерий	Соответствует	Не соответствует
выполнение / не выполнение практических заданий по теме лекции	1 балл	0 баллов
выполнение / не выполнение итоговой задачи	1 балл	0 баллов

Оценивание проводится преподавателем на основе представленных критериев и шкалы оценки.

Зачет выставляется слушателю, если по результатам оценивания слушатель получает не менее 2 баллов.

3.2. Оценочные материалы

таблица 13.3

Наименование раздела, темы	Основные показатели оценки	Формы и методы контроля и оценки	Вес задания
SQL и реляционные базы данных	ПК-5.р, ОПК-2, А/03.6	Практические задания по темам лекций	25
Работа с базами данных из Python	ПК-5.р, ОПК-2, А/03.6	Практические задания по темам лекций	25

Промежуточная аттестация	ПК-5.р, ОПК-2, А/03.6	Итоговое задание по модулю	50
--------------------------	-----------------------	----------------------------	----

Практические задания

1. Загрузка в PostgreSQL открытых данных для кейса поиска «сомнительных» клиентов банка:

- Единый государственный реестр юридических лиц (ЕГРЮЛ).
 - Единый государственный реестр индивидуальных предпринимателей (ЕГРИП).
 - Финансовая отчетность от Росстата и ФНС.
 - Общероссийский классификатор видов экономической деятельности ОКВЭД 2
2. Исследование открытых данных в PostgreSQL с помощью SQL.
3. Исследование открытых данных в PostgreSQL из Python.

Примеры заданий для проведения промежуточной аттестации по Модулю 3

1. Продолжение разработки программы на Python для сбора информации о компании или ИП из открытых источников: программа должна сохранять все загруженные данные в реляционной базе PostgreSQL.

4. Организационно-педагогические условия реализации программы

4.1 Учебно-методическое обеспечение и информационное обеспечение программы

4.1.1. Список литературы:

Основная литература:

1. Моргунов, Е. П. PostgreSQL. Основы языка SQL: учеб. пособие / Е. П. Моргунов; под ред. Е. В. Рогова, П. В. Лузанова. — СПб.: БХВ-Петербург, 2018. — 336 с.: ил. <https://postgrespro.ru/education/books/sqlprimer>
2. Новиков Б. А.. Основы технологий баз данных: учеб. пособие / Б. А. Новиков, Е. А. Горшкова, Н. Г. Графеева; под ред. Е. В. Рогова. — 2-е изд. — М.: ДМК Пресс, 2020. — 582 с. <https://postgrespro.ru/education/books/dbtech>
3. Рогов Е. В. PostgreSQL 15 изнутри. — М.: ДМК Пресс, 2023. — 662 с. <https://postgrespro.ru/education/books/internals>

Дополнительная литература:

1. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных – <http://www.machinelearning.ru/>

4.1.2. Электронные информационные ресурсы

1. Открытые данные – <https://ofdata.ru/open-data>
2. Андрей Созыкин. Анализ больших данных помогает нам выявлять «сомнительных» клиентов. <https://rb.ru/opinion/somnitelnye-klienty/>
3. Pandas – <https://pandas.pydata.org/>
4. PostgreSQL – <https://www.postgresql.org/>
5. pgAdmin – <https://www.pgadmin.org/>
6. DataGrip – <https://www.jetbrains.com/ru-ru/datagrip/>
7. SQLAlchemy – <https://www.sqlalchemy.org/>

4.1.3. Электронные образовательные ресурсы

1. Администрирование PostgreSQL. Базовый курс – <https://postgrespro.ru/education/courses/DBA1>
2. PostgreSQL. Оптимизация запросов – <https://postgrespro.ru/education/courses/QPT>

4.2. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

таблица 14.3

Наименование специализированных аудиторий, кабинетов, лабораторий	Вид занятий	Наименование оборудования, программного обеспечения
Аудитория	Лекция	Ноутбук / стационарный компьютер, гарнитура Программное обеспечение: VK Звонки / Яндекс телемост, Браузер Chrome или Яндекс.Браузер Интерпретатор языка Python Библиотеки Pандас, sqlalchemy.org Реляционная база данных PostgreSQL Клиент SQL: psql, pgAdmin или DataGrip
Аудитория	Практическая работа	Ноутбук / стационарный компьютер, гарнитура Программное обеспечение: VK Звонки / Яндекс телемост, Браузер Chrome или Яндекс.Браузер Интерпретатор языка Python Библиотеки Pандас, sqlalchemy.org Реляционная база данных PostgreSQL Клиент SQL: psql, pgAdmin или DataGrip
Аудитория	Промежуточная аттестация	Ноутбук / стационарный компьютер, гарнитура Программное обеспечение: VK Звонки / Яндекс телемост, Браузер Chrome или Яндекс.Браузер Интерпретатор языка Python Библиотеки Pандас, sqlalchemy.org Реляционная база данных PostgreSQL Клиент SQL: psql, pgAdmin или DataGrip

4.3. Организация образовательного процесса

Таблица 15.3

Методы обучения	Методы контроля	Формы организации учебных занятий	Формы организации учебной деятельности	Дистанционные образовательные технологии
Лекция Практическая работа под руководством преподавателя Самостоятельная практическая работа	Выполнение практических занятий по темам лекций Выполнение итогового задания по модулю	Лекция Вебинар с элементами практической работы и разбора теоретического материала	Групповая работа Индивидуальная работа	Использование образовательных интернет-ресурсов Использование ресурсов, созданных преподавателем

Изучение литературы по теме				WEB-консультации и другие
-----------------------------	--	--	--	---------------------------

2.3.4. РАБОЧАЯ ПРОГРАММА МОДУЛЯ 4

1. Цель и задачи освоения Модуля 4

Целью изучения модуля «Основы машинного обучения и нейронные сети» является знакомство с понятием машинного обучения.

Задачи Модуля 4:

1. Познакомиться с понятием машинного обучения.
2. Изучить основы классических алгоритмов машинного обучения.
3. Изучить основы нейронных сетей.

2. Характеристика совершенствуемых и/или получаемых новых компетенций, необходимых для профессиональной деятельности и/или повышения профессионального уровня в рамках имеющейся квалификации Модуль 4 «Основы машинного обучения и нейронные сети»

Таблица 8.4

№	Компетенция в соответствии с профессиональным стандартом 06.042 «Специалист по большим данным»	Код компетенции
1	Способен осуществлять подготовку данных для проведения аналитических работ по исследованию больших данных	A/03.6

Таблица 9.4

№	Компетенции в соответствии с направлением подготовки 09.03.01 Информатика и вычислительная техника	Код компетенции
1.	Способен использовать современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности	ОПК-2

Таблица 10.4

№	Компетенции в соответствии с универсальной моделью компетенций в сфере искусственного интеллекта*	Код компетенции
1.	Способен разрабатывать и применять методы машинного обучения для решения задач	ПК-4.р.
2.	Способен создавать и поддерживать системы искусственного интеллекта на основе нейросетевых моделей и методов	ПК-6.р.
3.	Способен создавать и внедрять одну или несколько сквозных цифровых субтехнологий искусственного интеллекта	ПК-9.р.

2.2. Учебная программа

Содержание учебной программы приведено в таблице 11.

таблица 11.4

№ п/п	Наименование модуля, разделов и тем	Содержание обучения, наименование и тематика практических занятий (семинаров), самостоятельной работы	Объем, ак. час.
Модуль 4 Основы машинного обучения и нейронные сети			
4.1	Машинное обучение	<i>Лекция</i>	6

		<p>Искусственный интеллект и машинное обучение. Отличие машинного обучения от программирования. Типы задач машинного обучения. Данные для машинного обучения. Популярные модели машинного обучения. Оценка качества обучения моделей, метрики качества. Переобучение. Популярные библиотеки машинного обучения.</p> <p>Практическая работа Установка и настройка библиотек машинного обучения.</p> <p>Самостоятельная работа Исследование возможностей библиотек машинного обучения.</p>	
4.2	Основы классических алгоритмов машинного обучения	<p>Лекция Линейная регрессия. Логистическая регрессия. k-средних. Деревья решений и случайный лес.</p> <p>Практическая работа Библиотека sklearn. Решение задачи регрессии на табличных данных с помощью sklearn. Решение задачи классификации на табличных данных с помощью sklearn. Оценка качества работы моделей машинного обучения</p> <p>Самостоятельная работа Обучение модели классификации на размеченных данных для обнаружения атак на киберфизическую систему.</p>	14
4.3	Классические алгоритмы машинного обучения	<p>Лекция Ансамблевые методы. Бэггинг. Бустинг.</p> <p>Практическая работа Бэггинг и бустинг в библиотеке sklearn. Библиотеки бустинга Catboost и XGBoost.</p> <p>Самостоятельная работа Обучение модели классификации на размеченных данных для обнаружения атак на киберфизическую систему с помощью бустинга.</p>	6
4.4.	Основы нейронных сетей	<p>Лекция Модель искусственного нейрона. Нейронные сети. Популярные архитектуры нейронных сетей: полносвязные, сверточные, рекуррентные, сети с архитектурой трансформер. Библиотеки для работы с нейронными сетями: TensorFlow, HuggingFace. Предварительно обучение нейронные сети.</p> <p>Практическая работа Библиотека HuggingFace. Использование предварительно обученных моделей из библиотеки HuggingFace для решения задач обработки естественного языка и компьютерного зрения. Библиотека TensorFlow. Создание и обучение нейронных сетей в TensorFlow. Архитектура нейронных сетей автокодировщик. Обучение автокодировщика для решения задачи обнаружения выбросов.</p>	16

		Самостоятельная работа Разработка нейронной сети для обнаружения атак на киберфизическую систему на данных без разметки с помощью подхода обнаружения выбросов.	
	Промежуточная аттестация по модулю 4	зачет	2
	Итого:		44

3. Формы аттестации и оценочные материалы

3.1. Формы аттестации

Промежуточная аттестация по модулю является обязательной.

Оценка качества освоения модуля проводится по 2-х-балльной системе по результатам текущего контроля (тестирование, проверочные задания на взаимную оценку), контроля посещаемости практических занятий и результатам промежуточной аттестации.

Форма промежуточной аттестации: зачет.

Слушатель считается аттестованным в случае положительных результатов работы (не менее 65% баллов от итоговой оценки) в процессе обучения и успешной сдачи зачета.

Критерии оценивания:

- выполнение / не выполнение практических заданий по темам лекции (1 балл за каждое задание);
- выполнение / не выполнение итогового задания по модулю (1/0 баллов соответственно).

таблица 12.4/1

Наименование модуля, разделов и тем	Основные показатели оценки	Формы и методы контроля и оценки
Основы машинного обучения и нейронные сети	ПК-4.р, ПК-6.р, ПК-9.р, ОПК-2, А/03.6	Зачет

Оценивание

таблица 12.4/2

Критерий	Соответствует	Не соответствует
выполнение / не выполнение практических заданий по теме лекции	1 балл	0 баллов
выполнение / не выполнение итоговой задачи	1 балл	0 баллов

Оценивание проводится преподавателем на основе представленных критериев и шкалы оценки.

Зачет выставляется слушателю, если по результатам оценивания слушатель получает не менее 2 баллов.

3.2. Оценочные материалы

таблица 13.4

Наименование раздела, темы	Основные показатели оценки	Формы и методы контроля и оценки	Вес задания
Машинное обучение	ПК-4.р, ПК-6.р, ПК-9.р, ОПК-2,	Практические задания по темам лекций	15

	А/03.6		
Основы классических алгоритмов машинного обучения	ПК-4.р, ПК-6.р, ПК-9.р, ОПК-2, А/03.6	Практические задания по темам лекций	15
Классические алгоритмы машинного обучения	ПК-4.р, ПК-6.р, ПК-9.р, ОПК-2, А/03.6	Практические задания по темам лекций	15
Основы нейронных сетей	ПК-4.р, ПК-6.р, ПК-9.р, ОПК-2, А/03.6	Практические задания по темам лекций	15
Промежуточная аттестация	ПК-4.р, ПК-6.р, ПК-9.р, ОПК-2, А/03.6	Итоговое задание по модулю	40

Практические задания

1. Исследование возможностей библиотек машинного обучения.
2. Обучение модели классификации на размеченных данных для обнаружения атак на киберфизическую систему.
3. Разработка нейронной сети для обнаружения атак на киберфизическую систему на данных без разметки с помощью подхода обнаружения выбросов.

Примеры заданий для проведения промежуточной аттестации по Модулю 4

Задание для промежуточного контроля разработал Чернышев Юрий Юрьевич, исследователь в компании Сайберлимфа (<https://cyberlympha.ru/>).

Необходимо научиться обнаруживать атаки на киберфизическую систему: производственное предприятие, включающее как физическое оборудование, так и информационные системы, объединенные компьютерной сетью.

Атаки могут быть двух видов:

- Кибератака с использованием компьютерной сети предприятия и его информационных систем.
- Физическая атака, связанная с манипулированием оборудованием на предприятии.

Набор данных для задачи включает показатели датчиков оборудования в разные моменты времени. Набор данных размеченный: указано, в какие моменты времени происходила атака.

Необходимо подобрать подходящую модель и алгоритм машинного обучения для определения атак на киберфизическую систему. Также необходимо выбрать метрики оценки качества модели машинного обучения.

4. Организационно-педагогические условия реализации программы

4.1 Учебно-методическое обеспечение и информационное обеспечение программы

4.1.1. Список литературы:

Основная литература:

1. Andrew Ferlitsch. Making the machine: the machine learning lifecycle. <https://cloud.google.com/blog/products/ai-machine-learning/making-the-machine-the-machine-learning-lifecycle>
2. Data-centric AI Resource Hub. <https://datacentricai.org/>
3. Шолле Франсуа. Глубокое обучение на Python. Библиотека программиста. 2018. 400 с.
4. Орельен Жерон: Прикладное машинное обучение с помощью Scikit-Learn, Keras и TensorFlow. Концепции, инструменты. Диалектика. 2020. 1040 с.

Дополнительная литература:

1. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных – <http://www.machinelearning.ru/>

4.1.2. Электронные информационные ресурсы

1. Sklearn – <https://scikit-learn.org/>
2. HuggingFace – <https://huggingface.co/>
3. TensorFlow – <https://www.tensorflow.org/>

4.1.3. Электронные образовательные ресурсы

Быстрый старт в искусственный интеллект – <https://stepik.org/course/80782/promo>

4.2. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

таблица 14.4

Наименование специализированных аудиторий, кабинетов, лабораторий	Вид занятий	Наименование оборудования, программного обеспечения
Аудитория	Лекция	Ноутбук / стационарный компьютер, гарнитура Программное обеспечение: VK Звонки / Яндекс телемост, Браузер Chrome или Яндекс.Браузер Интерпретатор языка Python Библиотеки sklearn, HuggingFace, TensorFlow
Аудитория	Практическая работа	Ноутбук / стационарный компьютер, гарнитура Программное обеспечение: VK Звонки / Яндекс телемост, Браузер Chrome или Яндекс.Браузер Интерпретатор языка Python Библиотеки sklearn, HuggingFace, TensorFlow
Аудитория	Промежуточная аттестация	Ноутбук / стационарный компьютер, гарнитура Программное обеспечение: VK Звонки / Яндекс телемост, Браузер Chrome или Яндекс.Браузер Интерпретатор языка Python Библиотеки sklearn, HuggingFace, TensorFlow

4.3. Организация образовательного процесса

Таблица 15.4

Методы обучения	Методы контроля	Формы организации учебных занятий	Формы организации учебной деятельности	Дистанционные образовательные технологии
Лекция Практическая работа под	Выполнение практических занятий по темам лекций	Лекция Вебинар с элементами	Групповая работа	Использование образовательных интернет-ресурсов

руководством преподавателя	Выполнение итогового задания по модулю	практической работы и разбора теоретического материала	Индивидуальная работа	Использование ресурсов, созданных преподавателем
Самостоятельная практическая работа				
Изучение литературы по теме				

2.3.5. РАБОЧАЯ ПРОГРАММА МОДУЛЯ 5

1. Цель и задачи освоения Модуля 5

Целью изучения модуля «Инструменты работы с большими данными» является знакомство с инструментами работы с большими данными.

Задачи Модуля 5:

1. Познакомиться с инструментом распределенной обработки больших данных Apache Spark.
2. Познакомиться с базой данных для обработки больших данных ClickHouse

2. Характеристика совершенствуемых и/или получаемых новых компетенций, необходимых для профессиональной деятельности и/или повышения профессионального уровня в рамках имеющейся квалификации Модуль 5 «Инструменты работы с большими данными»

Таблица 8.5

№	Компетенция в соответствии с профессиональным стандартом 06.042 «Специалист по большим данным»	Код компетенции
1	Способен осуществлять подготовку данных для проведения аналитических работ по исследованию больших данных	A/03.6

Таблица 9.5

№	Компетенции в соответствии с направлением подготовки 09.03.01 Информатика и вычислительная техника	Код компетенции
1.	Способен использовать современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности	ОПК-2

Таблица 10.5

№	Компетенции в соответствии с универсальной моделью компетенций в сфере искусственного интеллекта*	Код компетенции
1.	Способен разрабатывать системы анализа больших данных	ПК-8.р.

2.2. Учебная программа

Содержание учебной программы приведено в таблице 11.

таблица 11.5

№ п/п	Наименование модуля, разделов и тем	Содержание обучения, наименование и тематика практических занятий (семинаров), самостоятельной работы	Объем, ак.час.
Модуль 5 Инструменты работы с большими данными			
5.1	Apache Spark	Лекция Кластеры для распределенной обработки данных. Apache Hadoop и Apache Spark. Распределенная обработка данных в Apache Spark. DataFrame API в Apache Spark. Практическая работа Установка и настройка Apache Spark. Кластер из одного узла. Загрузка данных в Apache Spark.	18

		<p>Обработка данных в Apache Spark с помощью DataFrame API. Использование SQL в Apache Spark.</p> <p>Самостоятельная работа</p> <p>Загрузка в Apache Spark открытых данных для кейса поиска «сомнительных» клиентов банка:</p> <ul style="list-style-type: none"> - Единый государственный реестр юридических лиц (ЕГРЮЛ). - Единый государственный реестр индивидуальных предпринимателей (ЕГРИП). - Финансовая отчетность от Росстата и ФНС. - Общероссийский классификатор видов экономической деятельности ОКВЭД 2 <p>Анализ открытых данных в Apache Spark с помощью Spark DataFrame API и Spark SQL.</p>	
5.2	База данных Clickhouse	<p>Лекция</p> <p>База данных Clickhouse. Особенности хранения данных в Clickhouse. Обработка больших данных в Clickhouse.</p> <p>Практическая работа</p> <p>Установка и настройка Clickhouse. Загрузка данных в Clickhouse. Запуск SQL запросов в Clickhouse. Работа с Clickhouse из Python.</p> <p>Самостоятельная работа</p> <p>Загрузка в Clickhouse открытых данных для кейса поиска «сомнительных» клиентов банка:</p> <ul style="list-style-type: none"> - Единый государственный реестр юридических лиц (ЕГРЮЛ). - Единый государственный реестр индивидуальных предпринимателей (ЕГРИП). - Финансовая отчетность от Росстата и ФНС. - Общероссийский классификатор видов экономической деятельности ОКВЭД 2 <p>Анализ открытых данных Clickhouse с помощью SQL и Python.</p>	16
	Промежуточная аттестация по модулю 5	зачет	2
	Итого:		36

3. Формы аттестации и оценочные материалы

3.1. Формы аттестации

Промежуточная аттестация по модулю является обязательной.

Оценка качества освоения модуля проводится по 2-х-балльной системе по результатам текущего контроля (тестирование, проверочные задания на взаимную оценку), контроля посещаемости практических занятий и результатам промежуточной аттестации.

Форма промежуточной аттестации: зачет.

Слушатель считается аттестованным в случае положительных результатов работы (не менее 65% баллов от итоговой оценки) в процессе обучения и успешной сдачи зачета.

Критерии оценивания:

– выполнение / не выполнение практических заданий по темам лекции (1 балл за каждое задание);

– выполнение / не выполнение итогового задания по модулю (1/0 баллов соответственно).

таблица 12.5/1

Наименование модуля, разделов и тем	Основные показатели оценки	Формы и методы контроля и оценки
Инструменты работы с большими данными	ПК-8.р, ОПК-2, А/03.6	Зачет

Оценивание

таблица 12.5/2

Критерий	Соответствует	Не соответствует
выполнение / не выполнение практических заданий по теме лекции	1 балл	0 баллов
выполнение / не выполнение итоговой задачи	1 балл	0 баллов

Оценивание проводится преподавателем на основе представленных критериев и шкалы оценки.

Зачет выставляется слушателю, если по результатам оценивания слушатель получает не менее 2 баллов.

3.2. Оценочные материалы

таблица 13.5

Наименование раздела, темы	Основные показатели оценки	Формы и методы контроля и оценки	Вес задания
Apache Spark	ПК-8.р, ОПК-2, А/03.6	Практические задания по темам лекций	20
База данных Clickhouse	ПК-8.р, ОПК-2, А/03.6	Практические задания по темам лекций	20
Промежуточная аттестация	ПК-8.р, ОПК-2, А/03.6	Итоговое задание по модулю	60

Практические задания

1. Загрузка в Apache Spark и ClickHouse открытых данных для кейса поиска «сомнительных» клиентов банка:

- Единый государственный реестр юридических лиц (ЕГРЮЛ).
- Единый государственный реестр индивидуальных предпринимателей (ЕГРИП).
- Финансовая отчетность от Росстата и ФНС.
- Общероссийский классификатор видов экономической деятельности ОКВЭД

2. Анализ открытых данных в Apache Spark с помощью Spark DataFrame API и Spark SQL.

3. Анализ открытых данных в Apache Spark и ClickHouse.

Примеры заданий для проведения промежуточной аттестации по Модулю 5

Расширение возможностей программы, загружающей информацию о компании или ИП из открытых источников таким образом, чтобы загруженные данные сохранялись в

Apache Spark и ClickHouse.

4. Организационно-педагогические условия реализации программы

4.1 Учебно-методическое обеспечение и информационное обеспечение программы

4.1.1. Список литературы:

Основная литература:

1. Andrew Ferlitsch. Making the machine: the machine learning lifecycle. <https://cloud.google.com/blog/products/ai-machine-learning/making-the-machine-the-machine-learning-lifecycle>
2. Data-centric AI Resource Hub. <https://datacentricai.org/>
3. Шолле Франсуа. Глубокое обучение на Python. Библиотека программиста. 2018. 400 с.
4. Орельен Жерон: Прикладное машинное обучение с помощью Scikit-Learn, Keras и TensorFlow. Концепции, инструменты. Диалектика. 2020. 1040 с.

Дополнительная литература:

1. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных – <http://www.machinelearning.ru/>

4.1.2. Электронные информационные ресурсы

1. Apache Spark – <https://spark.apache.org/>
2. ClickHouse – <https://clickhouse.com/>

4.1.3. Электронные образовательные ресурсы

- Big Data и Data Science: перейди на новый уровень – <https://stepik.org/course/101689/info>

4.2. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

таблица 14.5

Наименование специализированных аудиторий, кабинетов, лабораторий	Вид занятий	Наименование оборудования, программного обеспечения
Аудитория	Лекция	Ноутбук / стационарный компьютер, гарнитура Программное обеспечение: VK Звонки / Яндекс телемост, Браузер Chrome или Яндекс.Браузер Интерпретатор языка Python Apache Spark База данных ClickHouse
Аудитория	Практическая работа	Ноутбук / стационарный компьютер, гарнитура Программное обеспечение: VK Звонки / Яндекс телемост, Браузер Chrome или Яндекс.Браузер Apache Spark База данных ClickHouse

Аудитория	Промежуточная аттестация	Ноутбук / стационарный компьютер, гарнитура Программное обеспечение: VK Звонки / Яндекс телемост, Браузер Chrome или Яндекс.Браузер Apache Spark База данных ClickHouse
-----------	--------------------------	--

4.3. Организация образовательного процесса

Таблица 15.5

Методы обучения	Методы контроля	Формы организации учебных занятий	Формы организации учебной деятельности	Дистанционные образовательные технологии
Лекция Практическая работа под руководством преподавателя Самостоятельная практическая работа Изучение литературы по теме	Выполнение практических занятий по темам лекций Выполнение итогового задания по модулю	Лекция Вебинар с элементами практической работы и разбора теоретического материала	Групповая работа Индивидуальная работа	Использование образовательных интернет-ресурсов Использование ресурсов, созданных преподавателем WEB-консультации и другие

2.3.6. РАБОЧАЯ ПРОГРАММА МОДУЛЯ 6

1. Цель и задачи освоения Модуля 6

Целью изучения модуля «Основы автоматизация машинного обучения MLOps» является знакомство с основы автоматизация машинного обучения Machine Learning Operations (MLOps).

Задачи Модуля 6:

1. Познакомиться с концепцией MLOps.
2. Научиться использовать dvc для управления версиями наборов данных.
3. Научиться использовать платформу MLOps ClearML.

2. Характеристика совершенствуемых и/или получаемых новых компетенций, необходимых для профессиональной деятельности и/или повышения профессионального уровня в рамках имеющейся квалификации Модуль 6 «Основы автоматизация машинного обучения MLOps»

Таблица 8.6

№	Компетенция в соответствии с профессиональным стандартом 06.042 «Специалист по большим данным»	Код компетенции
1	Способен осуществлять подготовку данных для проведения аналитических работ по исследованию больших данных	A/03.6

Таблица 9.6

№	Компетенции в соответствии с направлением подготовки 09.03.01 Информатика и вычислительная техника	Код компетенции
1.	Способен использовать современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности	ОПК-2

Таблица 10.6

№	Компетенции в соответствии с универсальной моделью компетенций в сфере искусственного интеллекта*	Код компетенции
1.	Способен разрабатывать и применять методы машинного обучения для решения задач	ПК-4.р
2.	Способен создавать и поддерживать системы искусственного интеллекта на основе нейросетевых моделей и методов	ПК-6.р
3.	Способен создавать и внедрять одну или несколько сквозных цифровых субтехнологий искусственного интеллекта	ПК-9.р

2.2. Учебная программа

Содержание учебной программы приведено в таблице 11.

таблица 11.6

№ п/п	Наименование модуля, разделов и тем	Содержание обучения, наименование и тематика практических занятий (семинаров), самостоятельной работы	Объем, ак.час.
Модуль 6 Основы автоматизация машинного обучения MLOps			
6.1	Автоматизация машинного обучения MLOps	Лекция Основы автоматизации машинного обучения MLOps. Поток работ машинного обучения. Версионирование кода и данных для обучения моделей. Отслеживание экспериментов обучения моделей. Практическая работа Ручная организация потоков работ машинного обучения. Самостоятельная работа Ручная организация потока работ машинного обучения для обнаружения атак на киберфизическую систему.	6
6.2	Управление версиями кода и данных	Лекция Необходимость отслеживания версий кода и данных при обучении моделей. Система dvc (data version control). Практическая работа Установка и настройка dvc. Интеграция dvc с git. Настройка различных типов хранилищ данных в dvc. Самостоятельная работа Настройка dvc для хранения версий данных об атаках на киберфизическую систему.	12
6.3	MLOps платформы	Лекция Обзор популярных систем и облачных платформ для реализации MLOps. Платформа ClearML. Управление версиями данных в ClearML. Отслеживание экспериментов в ClearML. Оценка качества обучаемых моделей в ClearML. Практическая работа Работа с облачной платформой ClearML. Отправка данных об обучении моделей на ClearML. Интеграция с ClearML библиотек sklearn и TensorFlow. Управление версиями наборов данных в ClearML. Отслеживание экспериментов по обучению моделей машинного обучения в ClearML. Настройка потоков работ машинного обучения в ClearML. Самостоятельная работа Настройка потока работ в ClearML для обучения модели обнаружения атак на киберфизическую систему.	16

Промежуточная аттестация по модулю 6	зачет	2
Итого:		36

3. Формы аттестации и оценочные материалы

3.1. Формы аттестации

Промежуточная аттестация по модулю является обязательной.

Оценка качества освоения модуля проводится по 2-х-балльной системе по результатам текущего контроля (тестирование, проверочные задания на взаимную оценку), контроля посещаемости практических занятий и результатам промежуточной аттестации.

Форма промежуточной аттестации: зачет.

Слушатель считается аттестованным в случае положительных результатов работы (не менее 65% баллов от итоговой оценки) в процессе обучения и успешной сдачи зачета.

Критерии оценивания:

- выполнение / не выполнение практических заданий по темам лекции (1 балл за каждое задание);
- выполнение / не выполнение итогового задания по модулю (1/0 баллов соответственно).

таблица 12.6/1

Наименование модуля, разделов и тем	Основные показатели оценки	Формы и методы контроля и оценки
Основы автоматизация машинного обучения MLOps	ПК-4.п, ПК-6.п, ПК-9.п, ОПК-2, А/03.6	Зачет

Оценивание

таблица 12.6/2

Критерий	Соответствует	Не соответствует
выполнение / не выполнение практических заданий по теме лекции	1 балл	0 баллов
выполнение / не выполнение итоговой задачи	1 балл	0 баллов

Оценивание проводится преподавателем на основе представленных критериев и шкалы оценки.

Зачет выставляется слушателю, если по результатам оценивания слушатель получает не менее 2 баллов.

3.2. Оценочные материалы

таблица 13.6

Наименование раздела, темы	Основные показатели оценки	Формы и методы контроля и оценки	Вес задания
Автоматизация машинного обучения MLOps	ПК-4.п, ПК-6.п, ПК-9.п, ОПК-2, А/03.6	Практические задания по темам лекций	20
Управление версиями кода и данных	ПК-4.п, ПК-6.п, ПК-9.п, ОПК-2, А/03.6	Практические задания по темам лекций	20

MLOps платформы	ПК-4.р, ПК-6.р, ПК-9.р, ОПК-2, А/03.6	Практические задания по темам лекций	20
Промежуточная аттестация	ПК-4.р, ПК-6.р, ПК-9.р, ОПК-2, А/03.6	Итоговое задание по модулю	40

Практические задания

1. Настройка dvc для хранения версий данных об атаках на киберфизическую систему.
2. Ручная организация потока работ машинного обучения для обнаружения атак на киберфизическую систему.
3. Настройка потока работ в ClearML для обучения модели обнаружения атак на киберфизическую систему.

Примеры заданий для проведения промежуточной аттестации по Модулю 6

Задание для промежуточного контроля разработал Чернышев Юрий Юрьевич, исследователь в компании Сайберлимфа (<https://cyberlympha.ru/>).

Необходимо научиться обнаруживать атаки на киберфизическую систему: производственное предприятие, включающее как физическое оборудование, так и информационные системы, объединенные компьютерной сетью.

Атаки могут быть двух видов:

- Кибератака с использованием компьютерной сети предприятия и его информационных систем.
- Физическая атака, связанная с манипулированием оборудованием на предприятии.

Набор данных для задачи включает показатели датчиков оборудования в разные моменты времени. Набор данных размеченный: указано, в какие моменты времени происходила атака.

Необходимо организовать с помощью ClearML отслеживание экспериментов по подбору подходящую модель и алгоритма машинного обучения для определения атак на киберфизическую систему.

4. Организационно-педагогические условия реализации программы

4.1 Учебно-методическое обеспечение и информационное обеспечение программы

4.1.1. Список литературы:

Основная литература:

1. Andrew Ferlitsch. Making the machine: the machine learning lifecycle. <https://cloud.google.com/blog/products/ai-machine-learning/making-the-machine-the-machine-learning-lifecycle>
2. Data-centric AI Resource Hub. <https://datacentricai.org/>
3. Шолле Франсуа. Глубокое обучение на Python. Библиотека программиста. 2018. 400 с.
4. Орельен Жерон: Прикладное машинное обучение с помощью Scikit-Learn, Keras и TensorFlow. Концепции, инструменты. Диалектика. 2020. 1040 с.

Дополнительная литература:

1. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных – <http://www.machinelearning.ru/>

4.1.2. Электронные информационные ресурсы

1. sklearn – <https://scikit-learn.org/>
2. TensorFlow – <https://www.tensorflow.org/>
3. MLOps платформа ClearML – <https://clear.ml/>

4.1.3. Электронные образовательные ресурсы

MLOps with ClearML – <https://youtu.be/lowmDcR5qL8>

4.2. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

таблица 14.6

Наименование специализированных аудиторий, кабинетов, лабораторий	Вид занятий	Наименование оборудования, программного обеспечения
Аудитория	Лекция	Ноутбук / стационарный компьютер, гарнитура Программное обеспечение: VK Звонки / Яндекс телемост, Браузер Chrome или Яндекс.Браузер Интерпретатор языка Python Библиотеки sklearn, TensorFlow Data version control (dvc) MLOps платформа ClearML.
Аудитория	Практическая работа	Ноутбук / стационарный компьютер, гарнитура Программное обеспечение: VK Звонки / Яндекс телемост, Браузер Chrome или Яндекс.Браузер Интерпретатор языка Python Библиотеки sklearn, TensorFlow Data version control (dvc) MLOps платформа ClearML.
Аудитория	Промежуточная аттестация	Ноутбук / стационарный компьютер, гарнитура Программное обеспечение: VK Звонки / Яндекс телемост, Браузер Chrome или Яндекс.Браузер Интерпретатор языка Python Библиотеки sklearn, TensorFlow Data version control (dvc) MLOps платформа ClearML.

4.3. Организация образовательного процесса

Таблица 15.6

Методы обучения	Методы контроля	Формы организации учебных занятий	Формы организации учебной деятельности	Дистанционные образовательные технологии
-----------------	-----------------	-----------------------------------	--	--

<p>Лекция</p> <p>Практическая работа под руководством преподавателя</p> <p>Самостоятельная практическая работа</p> <p>Изучение литературы по теме</p>	<p>Выполнение практических занятий по темам лекций</p> <p>Выполнение итогового задания по модулю</p>	<p>Лекция</p> <p>Вебинар с элементами практической работы и разбора теоретического материала</p>	<p>Групповая работа</p> <p>Индивидуальная работа</p>	<p>Использование образовательных интернет-ресурсов</p> <p>Использование ресурсов, созданных преподавателем</p> <p>WEB-консультации и другие</p>
---	--	--	--	---

РАЗДЕЛ 3. ОЦЕНКА КАЧЕСТВА ОСВОЕНИЯ ПРОГРАММЫ

«Инженер данных»

Оценка качества освоения программы включает текущую, промежуточную и итоговую аттестацию обучающихся.

Текущая аттестация – служит для оценки объема и уровня усвоения слушателем учебного материала одного модуля программы и применяется в рамках текущего контроля успеваемости слушателя. Учебные достижения обучающихся по всем видам учебных заданий в ходе текущего контроля.

Промежуточная аттестация слушателей включает сдачу зачетов по модулям, предусмотренным учебно-тематическим планом.

Итоговая аттестация выполняется в виде защиты проекта.

Текущий контроль успеваемости слушателей проводится во время учебных занятий и по результатам выполнения заданий для самостоятельной работы.

Промежуточная аттестация проводится по окончании каждого модуля программы.

При проведении итоговой аттестации применяется 2-хбалльная шкала оценивания. Содержание промежуточной аттестации, критерии и процедуры оценивания определяются рабочими программами модулей.

Итоговая аттестация слушателей проводится в сроки, предусмотренные учебным планом и графиком учебного процесса.

Примерные темы проектов

1. Разработка автоматизированного потока работ машинного обучения для обнаружения киберфизических атак.

2. Разработка автоматизированного потока работ машинного обучения для поиска объектов заданного типа в видео.

При проведении итоговой аттестации применяется 2-хбалльная шкала оценивания. Результаты аттестационных испытаний определяются отметками «Зачтено», «Незачтено» и объявляются после оформления в установленном порядке протоколов заседаний итоговой аттестационной комиссии.

При проведении итоговой аттестации предлагается следующая шкала оценок:

оценка «Зачтено» выставляется обучающемуся, если он показал всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений;

оценка «Незачтено» выставляется обучающемуся, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных понятий дисциплины и не умеет использовать полученные знания при решении типовых практических задач.

Оценивается:

- объем и структура данных, подготовленных для обучения.

- качество обученной модели машинного обучения;

- уровень автоматизации потока работ машинного обучения, использование платформ MLOps.

В случае несогласия с результатами итоговой аттестации слушатель имеет право подать апелляцию в аттестационную комиссию на имя председателя.

Апелляция принимается в день защиты проекта.

В ходе рассмотрения апелляции проверяется правильность и аргументированность выставленной оценки. При необходимости ИАК принимает решение о возможности повторной защиты проекта. Итоги рассмотрения апелляции вносятся в соответствующий протокол в раздел «Особое мнение комиссии».

ИАК рассматривает апелляцию и сообщает свое решение не позднее, чем через сутки с момента подачи апелляции. Решение ИАК по апелляции является окончательным.

Самостоятельная работа слушателей подразделяется на аудиторную и внеаудиторную. Аудиторную самостоятельную работу составляют самостоятельно выполняемые задания во время учебных занятий, поиск информации в интернете и др.

Внеаудиторная самостоятельная работа включает формы: самостоятельный поиск информации по заданным темам, разбор кейсов, выполнение домашних заданий, разработка плана реализации проекта, разработка предложений и др.

Основными критериями качества организации самостоятельной работы служит наличие контроля результатов самостоятельной работы и технических условий выполнения заданий.

Основными современными формами организации самостоятельной работы слушателей являются творческие работы и работа с информационными компьютерными технологиями.

Примеры практико-ориентированных кейсов

1. Обнаружение атак на киберфизические системы.

Кейс предоставлен компанией Сайбероимфа (<https://cyberlymph.ru/>).

Вы можете попробовать применить машинное обучение для решения задачи информационной безопасности: обнаружение атак на киберфизические системы. Это системы для решения критически важных задач, которые состоят из физических компонентов и цифровых алгоритмов.

Киберфизическая система – это производственное предприятие, включающее как физическое оборудование, так и информационные системы, объединенные компьютерной сетью.

Атаки могут быть двух видов:

- Кибератака с использованием компьютерной сети предприятия и его информационных систем.

- Физическая атака, связанная с манипулированием оборудованием на предприятии.

Демонстрационные данные для кейса взяты от стенда для безопасной очистки воды (Secure Water Treatment, SWaT), которая является уменьшенным вариантом реального завода. Стенд работал 11 дней, из которых 7 дней прошло в нормальном режиме (без атак и сбоев), а в оставшиеся дни проводились атаки двух типов:

- Кибератаки через компьютерную сеть стенда.
- Физические атаки, связанные с выводом из строя компонентов стенда (например, двигатель или насос).

Набор данных включает как состояние оборудования и характеристики очищаемой воды, так и сетевой трафик. Атаки в обучающем наборе данных размечены вручную. С помощью размеченных данных вам нужно научиться определять атаки для тестового набора данных.

Фото установки, на которой проводились эксперименты:



2. Поиск артефакта в видеоряде

Кейс предоставлен компанией Сайбероимфа (<https://cyberlympha.ru/>).

Важная задача кибербезопасности – обнаружение объектов заданного типа на изображении в видео. Таким объектом может быть человек без каски на промышленном предприятии, логотип компании, который использовали с нарушением авторских прав и т.п.

Задача кейса: научиться находить объекты заданного типа в видео. Усложненный вариант задания – интеграция по API с видеохостингами, RuTube YouTube.

3. Выявление сомнительных клиентов банка

Открытое описание кейса – <https://rb.ru/opinion/somnitelnye-klienty/>

Банки, которые обслуживают юридические лица и индивидуальных предпринимателей, должны выполнять требования регулирования. Одно из важных требований содержится в законе 115-ФЗ «О противодействии легализации (отмыванию) доходов, полученных преступным путем, и финансированию терроризма». Банк должен проверять платежи всех клиентов на соответствие этому закону. Если он делает это плохо, то у него могут отозвать лицензию. Таким образом очень важно выявлять клиентов, которые выполняют сомнительные платежи, и применять к ним меры противодействия легализации доходов, вплоть до закрытия счета.

Как по операциям определить, что клиент «сомнительный»? Центральный Банк выпускает методические рекомендации. Их достаточно много, среди наиболее важных признаков сомнительных клиентов можно указать:

- налоговая нагрузка меньше 0,9% к дебетовому обороту (сумма исходящих платежей);
- не платится зарплата, аренда и другие типичные платежи;
- остатки денег на счете малы по сравнению с оборотом;
- назначения поступлений на счет не совпадают с назначениями затрат;
- деньги приходят с НДС, а уходят без.

Необходимо обучить модель машинного обучения и реализовать приложение на его основе, которое определяет «сомнительных» клиентов банка.

РАЗДЕЛ 4. ОРГАНИЗАЦИОННО-ПЕДАГОГИЧЕСКИЕ УСЛОВИЯ РЕАЛИЗАЦИИ ПРОГРАММЫ «Инженер данных»

4.1. Материально-технические условия и организационное обеспечение реализации дополнительной профессиональной программы повышения квалификации

Для эффективного осуществления образовательного процесса необходимы

таблица 16

Наименование специализированных аудиторий, кабинетов, лабораторий	Вид занятий	Наименование оборудования, программного обеспечения
Система дистанционного обучения провайдера массовых открытых онлайн курсов	Лекции	Слушателю необходимо наличие доступа в сеть интернет, компьютер. Преподавателю курса необходимо наличие доступа администратора курса на LMS-платформе к материалам курса
Информационно-коммуникационная платформа дистанционных семинаров	Практические занятия (дистанционные семинары)	Слушателю необходимо наличие доступа в сеть интернет, компьютер. Преподавателю курса необходимо оборудование для проведения дистанционных семинаров (вебинаров), качественный отказоустойчивый доступ в сеть интернет
Система дистанционного обучения провайдера массовых открытых онлайн курсов	Самостоятельная работа	Наличие компьютера и доступа в сеть интернет
Система дистанционного обучения провайдера массовых открытых онлайн курсов	Текущий контроль Промежуточная аттестация Итоговая аттестация	Наличие компьютера и доступа в сеть интернет

В таблице 17 указаны образовательные технологии.

Таблица 17

№	Вид занятия	Форма проведения занятий	Цель
1	Лекция	Очные лекции, самостоятельный просмотр записей видеолекций	Ознакомление слушателей с базовым материалом по тематике курса
2	Практические занятия	Очное посещение занятий. Выполнение практических заданий, получение обратной связи от преподавателя. Обсуждение вопросов, возникших в результате просмотра видеолекций и изучения литературы.	Практическое освоение теоретических знаний, а также углубление знаний по курсу
3	Самостоятельная работа	Самостоятельное изучение дополнительных материалов и литературы.	Углубление знаний по курсу.

4	Контрольные задания	Выполнение тренировочных заданий и тестов	Углубление знаний и умений по курсу.
5	Выполнение проектного задания	Реализация проекта по выбору на технологиях, которые преподаются в процессе курса.	Практическое освоение теоретических знаний, контроль освоения материалов.
6	Итоговая аттестация	Выполнение итогового теста и подготовка итоговой аттестационной работы.	Практическое освоение теоретических знаний, контроль освоения материалов.

4.2. Учебно-методическое обеспечение реализации дополнительной профессиональной программы повышения квалификации

4.2.1. Список литературы:

Основная литература:

1. Daswin De Silva, Daminda Alahakoon. An artificial intelligence life cycle: From conception to production // Patterns, Volume 3, Issue 6, 2022, ISSN 2666-3899, <https://doi.org/10.1016/j.patter.2022.100489>.

2. Andrew Ferlitsch. Making the machine: the machine learning lifecycle. <https://cloud.google.com/blog/products/ai-machine-learning/making-the-machine-the-machine-learning-lifecycle>

3. Data-centric AI Resource Hub. <https://datacentricai.org/>

4. Моргунов, Е. П. PostgreSQL. Основы языка SQL: учеб. пособие / Е. П. Моргунов; под ред. Е. В. Рогова, П. В. Лузанова. — СПб.: БХВ-Петербург, 2018. — 336 с.: ил. <https://postgrespro.ru/education/books/sqlprimer>

5. Новиков Б. А.. Основы технологий баз данных: учеб. пособие / Б. А. Новиков, Е. А. Горшкова, Н. Г. Графеева; под ред. Е. В. Рогова. — 2-е изд. — М.: ДМК Пресс, 2020. — 582 с. <https://postgrespro.ru/education/books/dbtech>

6. Рогов Е. В. PostgreSQL 15 изнутри. — М.: ДМК Пресс, 2023. — 662 с. <https://postgrespro.ru/education/books/internals>

7. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных – <http://www.machinelearning.ru/>

8. Шолле Франсуа. Глубокое обучение на Python. Библиотека программиста. 2018. 400 с.

9. Орельен Жерон: Прикладное машинное обучение с помощью Scikit-Learn, Keras и TensorFlow. Концепции, инструменты. Диалектика. 2020. 1040 с.

4.2.2. Электронные информационные ресурсы

Андрей Созыкин. Анализ больших данных помогает нам выявлять «сомнительных» клиентов. <https://rb.ru/opinion/somnitelnye-klienty/>

Why do 87% of data science projects never make it into production? <https://venturebeat.com/ai/why-do-87-of-data-science-projects-never-make-it-into-production/>

Data-centric AI Resource Hub. <https://datacentricai.org/>

Открытые данные – <https://ofdata.ru/open-data>

Pandas – <https://pandas.pydata.org/>

Numpy - <https://numpy.org/>

Requests – <https://requests.readthedocs.io/en/latest/>

httplib – <https://www.python-httpx.org/>

BeautifulSoup – <https://www.crummy.com/software/BeautifulSoup/>

Postman – <https://www.postman.com/>

Открытые данные – <https://ofdata.ru/open-data>

PostgreSQL – <https://www.postgresql.org/>
pgAdmin – <https://www.pgadmin.org/>
DataGrip – <https://www.jetbrains.com/ru-ru/datagrip/>
SqlAlchemy – <https://www.sqlalchemy.org/>
Sklearn – <https://scikit-learn.org/>
HuggingFace – <https://huggingface.co/>
TensorFlow – <https://www.tensorflow.org/>
Apache Spark – <https://spark.apache.org/>
ClickHouse – <https://clickhouse.com/>
TensorFlow – <https://www.tensorflow.org/>
MLOps платформа ClearML – <https://clear.ml/>

4.2.3 Электронные образовательные ресурсы

Управление проектами по анализу данных
<https://sberuniversity.ru/learning/courses/digital-skills/upravlenie-proektami-po-analizu-dannykh/>
Python для искусственного интеллекта – <https://stepik.org/course/110361/promo>
Администрирование PostgreSQL. Базовый курс –
<https://postgrespro.ru/education/courses/DBA1>
PostgreSQL. Оптимизация запросов – <https://postgrespro.ru/education/courses/QPT>
Быстрый старт в искусственный интеллект – <https://stepik.org/course/80782/promo>
Big Data и Data Science: перейди на новый уровень –
<https://stepik.org/course/101689/info>
MLOps with ClearML – <https://youtu.be/lowmDcR5qL8>

4.3. Составители программы

Кошелев Антон Александрович, к. физ.-мат. н., руководитель группы математического моделирования, Банк Уралсиб
Чернышев Юрий Юрьевич, к. физ.-мат. н., исследователь, Сайберлимфа
Долганов Антон Юрьевич, к. т. н., доцент, Уральский федеральный университет
Созыкин Андрей Владимирович, к. т. н., заместитель директора центра развития ИТ-образования, Московский физико-технический институт
Борисов Василий Ильич, к. т. н., доцент, Уральский федеральный университет

Согласовано
Эксперт отдела сопровождения
образовательных программ


Ж.И. Зубцова

Согласовано
Директор ЦРИТО


А. В. Малеев

Федеральное государственное автономное образовательное учреждение
высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»
(МФТИ, Физтех)

ВЫПИСКА ИЗ ПРОТОКОЛА № 6

заседания учебно-методического совета от 29 января 2024 года.

ПОВЕСТКА:

Рассмотрение дополнительных общеобразовательных и профессиональных программ.

Проректор по учебной работе А. А. Воронов.

СЛУШАЛИ: заместителя директора (Центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск") А. И. Рыбакову о представлении дополнительных общеобразовательных и профессиональных программ (Центр «Пуск», МФТИ).

ПОСТАНОВИЛИ:

Рекомендовать к утверждению в установленном порядке дополнительную профессиональную программу профессиональной переподготовки «Инженер данных».

Решение принято единогласно.

Форма проведения заседания: заочная.

Председатель УМС МФТИ

Ученый секретарь УМС МФТИ




А.А. Воронов


М.В. Березникова