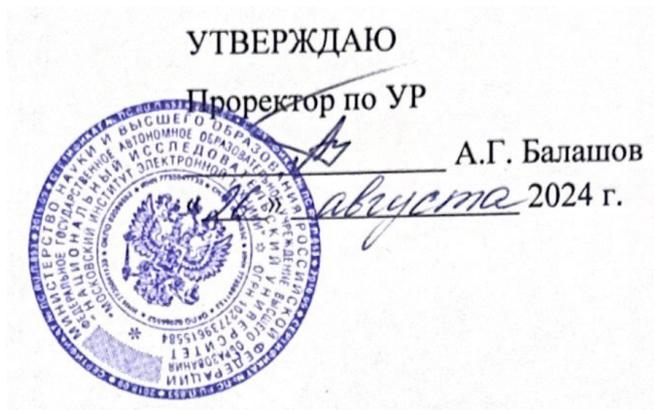


Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение высшего образования
«Национальный исследовательский университет
«Московский институт электронной техники»



**Дополнительная профессиональная программа
(программа профессиональной переподготовки)**

Анализ данных с применением технологий машинного обучения
(наименование программы)

Информационно-коммуникационные технологии

Москва, 2024 г.

Аннотация

Дополнительная профессиональная программа профессиональной переподготовки ИТ-профиля (далее – Программа) предназначена для обучающихся по очной или по очно-заочной форме за счет бюджетных средств или по договорам об оказании платных образовательных услуг, освоившие программы бакалавриата в объеме не менее 1 курса (бакалавры 2 курса), специалитета в объеме не менее 1 курса (специалисты 2 курса) и магистратуры (магистры) по специальностям и направлениям подготовки, отнесенным к ИТ-сфере.

Целью программы профессиональной переподготовки является формирование у слушателей профессиональных компетенций, необходимых для выполнения нового вида профессиональной деятельности – проектирование и разработка компьютерного программного обеспечения для анализа больших наборов данных при помощи технологий машинного обучения; получение актуальной для отрасли информационно-коммуникационных технологий дополнительной ИТ-квалификации «Программист».

Нормативный срок освоения программы 320 часов при очно-заочной форме подготовки в формате электронного обучения с применением дистанционных образовательных технологий.

Авторы и преподаватели:

- Рычагов Михаил Николаевич, профессор Института СПИНТех МИЭТ, д.ф.-м.н., профессор, Senior Manager on SW Development «Align Technology»;
- Гайдук Игорь Олегович, доцент Института СПИНТех МИЭТ, к.т.н., инженер НПК «Технологический центр»;
- Гаращенко Антон Витальевич, доцент Института СПИНТех МИЭТ, к.т.н., руководитель направления департамента перспективных вычислительных технологий ПАО «СБЕРБАНК»;
- Власова Алина Олеговна, старший преподаватель Института СПИНТех МИЭТ;
- Доронина Анна Александровна, старший преподаватель Института СПИНТех МИЭТ.

Содержание

Аннотация.....	2
I. Общие положения	4
1. Нормативная правовая основа Программы	4
2. Термины и определения, используемые в Программе.....	5
3. Требования к поступающим	7
II. Планируемые результаты обучения и структура Программы.....	8
1. Структура образовательных результатов	9
2. Структура Программы	10
III. Учебный план Программы.....	11
IV. Календарный учебный график	12
V. Рабочие программы модулей.....	13
1. Рабочая программа учебной дисциплины «Программный инструментарий высокопроизводительных вычислений».....	13
2. Рабочая программа учебной дисциплины «Методы машинного обучения»	18
3. Рабочая программа учебной дисциплины «Основы нейросетевых технологий»	22
4. Рабочая программа учебной дисциплины «Аналитика больших данных» ..	26
5. Рабочая программа практики.....	30
VI. Итоговая аттестация по Программе.....	33
1. Требования к содержанию и оформлению ИАР	33
2. Критерии начисления баллов за ИАР	35
VII. Завершение обучения по Программе	37

I. Общие положения

1. Нормативная правовая основа Программы

- Федеральный закон от 29 декабря 2012 года №273-ФЗ «Об образовании в Российской Федерации»;
- постановление Правительства Российской Федерации от 13 мая 2021 г. № 729 «О мерах по реализации программы стратегического лидерства «Приоритет-2030»;
- паспорт федерального проекта «Развитие кадрового потенциала ИТ-отрасли» национальной программы «Цифровая экономика Российской Федерации»;
- приказ Минцифры России от 29.12.2023 № 1180 «Об утверждении методик расчета показателей федеральных проектов «Развитие кадрового потенциала ИТ-отрасли» и «Обеспечение доступа в Интернет за счет развития спутниковой связи» национальной программы «Цифровая экономика Российской Федерации», а также внесении изменений в некоторые приказы Министерства цифрового развития, связи и массовых коммуникаций Российской Федерации об утверждении методик расчета показателей федеральных проектов национальной программы «Цифровая экономика Российской Федерации» (далее – приказ Минцифры России № 1180);
- приказ Минобрнауки России от 1 июля 2013 г. № 499 «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам» (с изменениями, внесенными приказом Минобрнауки России от 15 ноября 2013 г. № 1244 «О внесении изменений в Порядок организации и осуществления образовательной деятельности по дополнительным профессиональным программам, утвержденный приказом Министерства образования и науки Российской Федерации от 1 июля 2013 г. № 499»);
- приказ Минтруда России от 12 апреля 2013 г. № 148н «Об утверждении уровней квалификации в целях разработки проектов профессиональных стандартов»;
- методические рекомендации по разработке основных профессиональных образовательных программ и дополнительных профессиональных программ с учетом соответствующих профессиональных стандартов (утв. Минобрнауки России 22 января 2015 г. № ДЛ-1/05вн);
- постановление Правительства Российской Федерации от 11 октября 2023 г. № 1678 «Об утверждении Правил применения организациями, осуществляющими образовательную деятельность, электронного обучения, дистанционных образовательных технологий при реализации образовательных программ»;
- федеральный государственный образовательный стандарт высшего образования по направлению подготовки 09.03.04 Программная инженерия (уровень бакалавриата), утвержденный приказом Минобрнауки России от 19 сентября 2017 г. № 920 (далее вместе – ФГОС ВО);
- профессиональный стандарт «Программист», утвержденный приказом Министерства труда и социальной защиты РФ от 20 июля 2022 г. N 424н (далее – профессиональный стандарт).

2. Термины и определения, используемые в Программе

Дополнительная ИТ-квалификация – квалификация, приобретаемая в ходе освоения Программы обучающимися:

1) специальностей и направлений подготовки, отнесённых к ИТ-сфере, – в части формирования навыков использования и формирования цифровых компетенций, необходимых для выполнения нового вида профессиональной деятельности в соответствии с перечнем областей цифровых компетенций согласно приложению 1 к Методике расчета показателя «Количество обученных, получивших дополнительную ИТ-квалификацию на «цифровых кафедрах», утверждённой приказом Минцифры России № 1180 (далее – Методика расчета Показателя);

2) специальностей и направлений подготовки, не отнесённых к ИТ-сфере, – в области создания алгоритмов и компьютерных программ, пригодных для практического применения.

Специальности и направления подготовки, отнесённые к ИТ-сфере, – специальности и направления подготовки, перечисленные в перечне направлений подготовки (бакалавриат) и специальностей (специалитет) высшего образования в приложении 2 к Методике расчета Показателя.

Специальности и направления подготовки, не отнесённые к ИТ-сфере, – специальности и направления подготовки (бакалавриат, специалитет, магистратура, ординатура), не указанные в перечне направлений подготовки (бакалавриат) и специальностей (специалитет) высшего образования в приложении 2 к Методике расчета Показателя.

Цифровая компетенция (компетенция) – образовательный результат, формируемый при освоении Программы, необходимый для приобретения дополнительной ИТ-квалификации и выражающийся в осуществлении деятельности в области создания алгоритмов и компьютерных программ, пригодных для практического применения, выполнении нового вида профессиональной деятельности.

Целевой уровень сформированности компетенций – установленный Программой уровень сформированности компетенций в соответствии с Матрицей компетенций, актуальных для цифровой экономики, с приоритетом компетенций в ИТ-сфере.

Матрица цифровых компетенций – матрица компетенций, актуальных для цифровой экономики, с приоритетом компетенций в ИТ-сфере, разработанная Университетом Иннополис при участии ИТ-компаний и университетов-участников программы «Приоритет-2030», представляющая собой перечень компетенций, структурированный по сферам применения, типу компетенций, уровням их сформированности и характеристикам.

Знание (З) – информация о свойствах объектов, закономерностях процессов и явлений, правилах использования этой информации для принятия решений, присвоенная обучающимся на одном из уровней, позволяющих выполнять над ней мыслительные операции.

Умение (У) – освоенный субъектом способ выполнения действия, обеспечиваемый совокупностью приобретенных знаний и навыков; операция (действие), выполняемая определенным способом и с определенным качеством.

Опыт практической деятельности (ОПД) – образовательный результат, включающий выполнение обучающимся деятельности, завершающейся получением результата / продукта (элемента продукта), значимого при выполнении трудовой функции, в условиях реального производства или в модельной ситуации.

Дополнительная профессиональная программа профессиональной переподготовки (Программа) – комплекс основных характеристик образования (объем, содержание, планируемые результаты) и организационно-педагогических условий, который представлен в виде учебного плана, календарного учебного графика, рабочих программ учебных курсов, дисциплин (модулей), оценочных и методических материалов, а также программ учебной и производственной практик, стажировок и форм аттестации, иных компонентов и обеспечивает приобретение дополнительной квалификации. Программа может разрабатываться с учетом положений профессиональных стандартов, федеральных государственных образовательных стандартов, требований рынка труда (индустрии).

Рабочая программа – нормативный документ в составе Программы, регламентирующий взаимодействие преподавателя и обучающихся в ходе учебного процесса при реализации структурных элементов Программы (модуль, дисциплина, курс).

Профессиональный модуль (ПМ) – структурный элемент Программы, предназначенный для формирования определенных компетенций.

Учебная дисциплина (УД) – структурный элемент Программы, предназначенный для формирования знаний и умений в соответствующей сфере профессиональной деятельности.

Практика (практическая подготовка) – форма организации образовательной деятельности при освоении образовательной программы в условиях выполнения обучающимися определенных видов работ, связанных с будущей профессиональной деятельностью и направленных на формирование, закрепление, развитие практических навыков и компетенции по профилю соответствующей образовательной программы.

Дистанционные образовательные технологии – это образовательные технологии, реализуемые в основном с применением информационно- телекоммуникационных сетей при опосредованном (на расстоянии) взаимодействии обучающихся и педагогических работников.

Фонды оценочных средств (ФОС) – совокупность оценочных средств, используемых на различных этапах педагогической диагностики.

Оценочные средства (ОС) – дидактические средства для оценки качества подготовленности обучающихся.

Оценка цифровых компетенций (ассесмент) – проводимая на платформе Минцифры России оценка уровня сформированности цифровых компетенций, состоящая из трёх этапов:

1) входная оценка – оценка входного уровня цифровых компетенций обучающихся, которая проводится на этапе зачисления и начала обучения по Программе.

2) промежуточная оценка – это оценка уровня сформированности цифровых компетенций обучающихся, которая проводится в процессе обучения по Программе.

3) итоговая оценка – оценка достижения обучающимися целевого уровня сформированности цифровых компетенций, которая проводится на этапе завершения обучения по Программе.

3. Требования к поступающим

К обучению по Программе допускаются обучающиеся по очной или по очно-заочной форме за счет бюджетных средств или по договорам об оказании платных образовательных услуг, освоившие программы бакалавриата в объеме не менее 1 курса (бакалавры 2 курса), специалитета в объеме не менее 1 курса (специалисты 2 курса) и магистратуры (магистры) по специальностям и направлениям подготовки, отнесённым к ИТ-сфере.

4. Квалификационная характеристика выпускника

Выпускникам Программы присваивается дополнительная ИТ-квалификация в области формирования навыков использования и формирования цифровых компетенций, необходимых для выполнения нового вида профессиональной деятельности.

Выпускник Программы будет готов к выполнению нового вида трудовой деятельности – разработка компьютерного программного обеспечения для анализа данных – в соответствии с областью профессиональной деятельности: 06 Связь, информационные и коммуникационные технологии.

Объекты профессиональной деятельности: компьютерное программное обеспечение для анализа данных.

Задачи профессиональной деятельности:

- проектирование программных интерфейсов компьютерного программного обеспечения для анализа больших наборов данных при помощи технологий машинного обучения;

- проектирование структур данных компьютерного программного обеспечения;

- разработка архитектуры компьютерного программного обеспечения для анализа данных.

Квалификация: Программист.

Вид экономической деятельности: деятельность в области информации и связи.

Компетенции и трудовые функции определены на основании профессионального стандарта 06.001 «Программист».

Обобщенная трудовая функция: Д6 «Разработка требований и проектирование программного обеспечения».

Трудовая функция: Д/03.6 «Проектирование компьютерного программного обеспечения».

Квалификационный уровень по национальной рамке квалификаций: 6.

II. Планируемые результаты обучения и структура Программы

Получение дополнительной ИТ-квалификации «Программист» обеспечивается формированием приведённых в таблице цифровых компетенций:

Наименование сферы	ID и наименование компетенции	Инструменты профессиональной деятельности	Целевой уровень формирования компетенций в Программе			
			Минимальный (исходный)	Базовый	Продвинутый	Экспертный
Средства программной разработки	ПК-1 (ID 29) Разрабатывает программное обеспечение	Python	Решает учебные задачи по программированию, руководствуясь подробной пошаговой задачей под контролем опытных наставников.	-	Участствует в разработке ПО, самостоятельно применяет языки программирования и для решения технически сложных задач, не способен решать бизнес задачи с применением языков программирования.	-
Искусственный интеллект и машинное обучение	ПК-2 (ID 171) Разрабатывает модули машинного обучения (МО) для решения задач	Python, GNU Octave	Не применяет	-	Применяет методы и критерии оценки качества моделей МО, определяет критерии и метрики оценки результатов моделирования. Самостоятельно решает задачи анализа, прогнозирования, планирования, синтеза и принятия решений	-
Искусственный интеллект и машинное обучение	ПК-3 (ID 37) Применяет Искусственный интеллект и машинное обучение	NVIDIA Jetson, DataRobot	Не применяет Искусственный интеллект и машинное обучение	-	Разрабатывает отдельные части проектов по применению искусственного интеллекта и машинного обучения	-

1. Структура образовательных результатов

Формирование цифровых компетенций, необходимых для получения обучающимися дополнительной ИТ-квалификации, обеспечивается последовательным формированием промежуточных образовательных результатов, начиная со знаний.

ID и формулировка целевого уровня формирования компетенций	Промежуточные образовательные результаты		
	Опыт практической деятельности (ОПД)	Умения (У)	Знания (З)
ПК-1 (ID 29) Разрабатывает программное обеспечение	ОПД_1 разработки и отладки параллельных процессов; применения средств виртуализации вычислений.	У_1 работать со средствами запуска параллельных процессов, навыками проведения эксперимента по оценке эффективности вычислений.	З_1 средства виртуализации вычислений; архитектуру высокопроизводительных вычислительных кластеров, программные и аппаратные компоненты кластерных систем.
ПК-2 (ID 171) Разрабатывает модули машинного обучения (МО) для решения задач	ОПД_2 применения методов машинного обучения, построения и оценки качества моделей	У_2 визуализировать результаты работы алгоритмов машинного обучения, выбирать метод машинного обучения, соответствующий исследовательской задаче, интерпретировать полученные результаты.	З_2 ключевые понятия, цели и задачи использования машинного обучения; методологические основы применения алгоритмов машинного обучения.
ПК-3 (ID 37) Применяет Искусственный интеллект и машинное обучение	ОПД_3 использования нейронной сети для построения системы принятия решений.	У_3 использовать современные технологии для построения нейронных сетей, для выбора стратегии обучения и самообучения нейронной сети.	З_3 современных технологий построения и использования нейронных сетей.
	ОПД_4 применения библиотек современных языков программирования для анализа, визуализации данных.	У_4 применять современные информационные системы и технологии для реализации моделей многократного использования при обработке больших данных методами машинного и глубокого обучения.	З_4 методики и алгоритмы анализа и обработки больших данных с использованием современных языков программирования высокого уровня.

2. Структура Программы

Структура Программы регулирует образовательные траектории обучающихся, последовательность освоения структурных элементов (разделов) Программы, соответственно, последовательность формирования всех образовательных результатов.

Структурные элементы (разделы Программы)	Шифры образовательных результатов	Вариатив / инвариант и целевые группы обучающихся
Профессиональный цикл		
Модуль 1. Программный инструментарий высокопроизводительных вычислений	Знания: З_1 Умения: У_1	Инвариант для всех групп обучающихся
Модуль 2. Методы машинного обучения	Знания: З_2 Умения: У_2	Инвариант для всех групп обучающихся
Модуль 3. Основы нейросетевых технологий	Знания: З_3 Умения: У_3	Инвариант для всех групп обучающихся
Модуль 4. Аналитика больших данных	Знания: З_4 Умения: У_4	Инвариант для всех групп обучающихся
Практика	Опыт практической деятельности: ОПД_1, ОПД_2, ОПД_3, ОПД_4	Инвариант для всех групп обучающихся

III. Учебный план Программы

Объем Программы составляет 320 часов.

Учебный план Программы определяет перечень, последовательность, общую трудоемкость разделов и формы контроля знаний.

Структурные элементы (разделы Программы)	Общая трудоемкость, часов	Обязательная аудиторная учебная нагрузка		Самостоятельная работа, часов	Практики, стажировки, часов	Промежуточная аттестация, часов
		всего, часов	в т.ч. практические занятия, часов			
Модуль 1. Программный инструментарий высокопроизводительных вычислений	94	44	24	48		2
Модуль 2. Методы машинного обучения	64	30	16	32		2
Модуль 3. Основы нейросетевых технологий	60	26	16	32		2
Модуль 4. Аналитика больших данных	58	24	16	32		2
Практика	32				32	
Итоговая аттестация в формате защиты итоговой аттестационной работы (включая подготовку к аттестации)	12			12		
Итого:	320	124	72	156	32	8

V. Рабочие программы модулей

1. Рабочая программа учебной дисциплины «Программный инструментарий высокопроизводительных вычислений»

1.1. Область применения рабочей программы

Рабочая программа «Программный инструментарий высокопроизводительных вычислений» (далее – рабочая программа) является частью дополнительной профессиональной программы профессиональной переподготовки ИТ-профиля «Анализ данных с применением технологий машинного обучения» и направлена на *получение студентами знаний и умений в области высокопроизводительных вычислений и опыта применения стека программного обеспечения для параллелизации процессов.*

1.2. Структура и краткое содержание рабочей программы

№ п/п	Наименование тем, виды учебной работы и краткое содержание учебного материала	Объем, часов
1.	<p>Тема 1. Python-библиотеки для высокопроизводительных вычислений</p> <p>Лекция 1: Введение в HPC на Python. Обзор существующих фреймворков (TensorFlow, Numpy, Keras).</p> <p>Лекция 2: NumPy с точки зрения производительности. Применение программирования массивов. Тонкая настройка внутренней архитектуры NumPy.</p> <p>Лекция 3: Высокопроизводительные pandas и Apache Arrow. Оптимизация памяти и времени при загрузке данных. Методики для увеличения скорости анализа данных. Считывание данных в pandas при помощи Arrow. Применение Arrow Interop для делегирования работы более эффективным языкам и системам.</p> <p>Лекция 4: Матричные вычисления с использованием пакета Numba. Декораторы в Numba. PyCUDA. PyOpenCL.</p> <p>Лабораторная работа 1: Выполнение векторно-матричных операций в пакете NumPy.</p> <p>Лабораторная работа 2: Матричные умножения с применением пакета Numba.</p> <p>Самостоятельная работа: Текущая проработка теоретического материала (изучение учебного материала по конспектам лекций, литературным источникам и составление конспекта, развернутого плана). Выполнение практического задания по теме №1 в целях подготовки к выполнению лабораторных работ №1-2.</p>	32
2.	<p>Тема 2. Данные и системы их хранения</p> <p>Лекция 5: Основные сведения о данных. История развития баз данных. Основные компоненты баз данных. Модели данных. Реляционная модель. NoSQL модель. Иерархическая модель. Сетевая модель.</p> <p>Лекция 6: Язык определения данных (DDL). Язык манипулирования данными (DML). Язык запросов (DQL). Язык управления данными (DCL).</p> <p>Лекция 7: Задачи нормализации. Основные нормальные формы. Нормализация выше третьей формы. Денормализация.</p> <p>Лекция 8: Хранимые процедуры. Синтаксис хранимых процедур. Python в хранимых процедурах. Триггеры. NoSQL базы данных. Key-Value базы данных. Wide-column базы данных. Документоориентированные базы данных. Графовые базы данных.</p>	32

№ п/п	Наименование тем, виды учебной работы и краткое содержание учебного материала	Объем, часов
	<p>Лабораторная работа 3: Проектирование и нормализация базы данных.</p> <p>Лабораторная работа 4: SQL и NoSQL базы данных.</p> <p>Самостоятельная работа: Текущая проработка теоретического материала (изучение учебного материала по конспектам лекций, литературным источникам и составление конспекта, развернутого плана). Выполнение практического задания по теме №2 в целях подготовки к выполнению лабораторных работ №3-4.</p>	
3.	<p>Тема 3. Программно-аппаратная акселерация</p> <p>Лекция 9: Архитектура нейросетевых акселераторов: GPU, TPU, GROQ. Систематические массивы. Матричные вычисления. Векторизация.</p> <p>Лекция 10: Модели распределенных вычислений. Специализированные типы данных. Бинарная трансляция.</p> <p>Лабораторная работа 5: Матричные вычисления с использованием пакета Numba.</p> <p>Лабораторная работа 6: Побитовая арифметика специализированных типов данных с плавающей запятой.</p> <p>Самостоятельная работа: Текущая проработка теоретического материала (изучение учебного материала по конспектам лекций, литературным источникам и составление конспекта, развернутого плана). Выполнение практического задания по теме №3 в целях подготовки к выполнению лабораторных работ №5-6.</p>	28
4.	Промежуточная аттестация	2

1.3. Учебно-тематический план рабочей программы

№ п/п	Наименование и краткое содержание структурного элемента (раздела) Программы	Количество часов		
		аудиторных		самостоятельной работы
		лекции, семинары	практические занятия	
1	Тема 1. Python-библиотеки для высокопроизводительных вычислений	8	8	16
2	Тема 2. Данные и системы их хранения	8	8	16
3	Тема 3. Программно-аппаратная акселерация	4	8	16
5	Промежуточная аттестация	2		
	Итого	94		

1.4. Контроль и оценка результатов освоения рабочей программы

Образовательная организация высшего образования, реализующая рабочую программу, обеспечивает организацию и проведение *текущего, промежуточного* контроля.

Текущий контроль проводится преподавателем на основе оценивания результатов лабораторных работ и самостоятельной работы обучающихся. Промежуточный контроль проводится в форме демонстрационного экзамена, состоящего из двух частей: тестирования

и решения практического задания. Формы и методы текущего и промежуточного контроля, критерии оценивания доводятся до сведения обучающихся в начале обучения.

1.4.1. Примеры оценочных средств

Слушатель знает средства виртуализации вычислений; архитектуру высокопроизводительных вычислительных кластеров, программные и аппаратные компоненты кластерных систем.

Примеры тестовых заданий:

Вопрос	Варианты ответов
1. Какой метод используется для объединения массивов по горизонтали?	A) hstack(); B) stack(); C) hor_stack(); D) stackH().
2. Какая функция отвечает за создание массива из заданного количества чисел?	A) space(); B) range(); C) num_range(); D) linspace().
3. Таблицы в базах данных предназначены:	A) для хранения данных базы; B) для отбора и обработки данных базы; C) для ввода данных базы и их просмотра; D) для автоматического выполнения группы команд; E) для выполнения сложных программных действий.
4. В каком диалоговом окне создают связи между полями таблиц базы данных?	A) таблица связей; B) схема связей; C) схема данных; D) таблица данных.
5. Распределенные системы — это системы, которые предназначены для:	A) работы только на одном персональном компьютере или рабочей станции; B) на одном процессоре либо на интегрированной группе процессоров; C) на слабо интегрированной группе параллельно работающих процессоров, связанных через сеть.
6. Возможность распределенного размещения данных, логически представляющих собой единое целое, называется:	A) локальная автономия; B) непрерывные операции; C) прозрачная фрагментация.

Проверка умений слушателей работать со средствами запуска параллельных процессов, проводить эксперименты по оценке эффективности вычислений.

Примеры типовых практических заданий:

Написать программу на языке Python с применением пакета NumPy в соответствии с номером варианта.

1. Создать вектор размерностью 25 элементов, состоящими из чисел от 1 до 25. Преобразовать этот вектор в матрицу 5x5 элементов.

2. Выделить из полученной матрицы элементы главной диагонали и выполнить умножение с сатурацией и вывести результат на экран.

Данное практическое задание предполагает выполнение следующих этапов:

1. Сформировать вектор с числовой последовательностью;
2. Привести вектор к двумерной матрице;

3. Выделить необходимые элементы из матрицы и выполнить операции MAC;
4. Вывести полученный результат на экран;
5. Оформить и защитить отчет по практическому заданию.

1.5. Учебно-методические материалы для обучающихся и преподавателей

Для теоретической подготовки используются учебная литература и комплект теоретических материалов, в виде электронных презентаций к лекциям, содержащие основные определения, математический аппарат и иллюстративный материал.

Для подготовки к лабораторным работам используется лабораторный практикум, содержащий описание методов и подходов к выполнению лабораторных заданий.

Электронные презентации по дисциплине, методические рекомендации, а также лабораторный практикум размещаются в информационной системе iSpring <https://edu.miet.ru/>.

1.6. Материально-техническое обеспечение реализации рабочей программы

Компьютерная техника с возможностью подключения к сети «Интернет» и установленным программным обеспечением: Операционная система Microsoft Windows от 7 версии и выше, Мой Офис, браузер (Яндекс и пр.), среда разработки (Jupiter, Notepad++, VS CODE), Git, Gitlab, Gitverse, виртуальная машина/контейнер (Virtual Box, Docker) с установленной debian-based операционной системой (ОС AstraLinux, ОС Ubuntu), интерпретатор языка Python, системы визуализации данных PlantUML, база данных PostgreSQL.

1.7. Информационное обеспечение реализации рабочей программы

Перечень учебной литературы

1. Тоуманнен, Б. Программирование GPU при помощи Python и CUDA : исследуйте высокопроизводительные параллельные вычисления с помощью CUDA : практическое пособие : [16+] / Б. Тоуманнен ; пер. с англ. А. В. Борескова. – Москва : ДМК Пресс, 2020. – 255 с. : ил. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=602198> (дата обращения: 17.07.2024). – ISBN 978-5-97060-821-0. – Текст : электронный.

2. Мартишин, С. А. Базы данных. Практическое применение СУБД SQL и NoSQL-типа для проектирования информационных систем : учебное пособие / С.А. Мартишин, В.Л. Симонов, М.В. Храпченко. — Москва : ФОРУМ : ИНФРА-М, 2024. — 368 с. — (Высшее образование). - ISBN 978-5-8199-0946-1. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2096940> (дата обращения: 17.07.2024). – Режим доступа: по подписке.

Информационные ресурсы

1. Лань: Электронно-библиотечная система Издательства Лань. - СПб., 2011-. - URL: <https://e.lanbook.com> (дата обращения: 15.07.2024). - Режим доступа: для авторизованных пользователей МИЭТ

2. eLIBRARY.RU: Научная электронная библиотека: сайт. - Москва, 2000 -. -

URL: <https://www.elibrary.ru/defaultx.asp> (дата обращения: 15.07.2024). - Режим доступа: для зарегистрированных пользователей

3. Единое окно доступа к информационным ресурсам: сайт /ФГАУ ГНИИ ИТТ "Информика". - Москва, 2005-2010. - URL: <http://window.edu.ru/catalog/> (дата обращения: 15.07.2024).

2. Рабочая программа учебной дисциплины «Методы машинного обучения»

2.1. Область применения рабочей программы

Рабочая программа «Методы машинного обучения» (далее – рабочая программа) является частью дополнительной профессиональной программы профессиональной переподготовки ИТ-профиля «Анализ данных с применением технологий машинного обучения» и направлена на *получение студентами знаний технологий и умений реализации функционала интеллектуальных систем для решения профессиональных задач.*

2.2. Структура и краткое содержание рабочей программы

№ п/п	Наименование тем, виды учебной работы и краткое содержание учебного материала	Объем, часов
1.	<p>Тема 1. Искусственный интеллект как научное направление</p> <p>Лекция 1: Интеллектуальные системы как научное направление. Термины и определения. История искусственного интеллекта.</p> <p>Лекция 2: Обучение на основе наблюдений. Линейная регрессия как задача контролируемого (индуктивного) обучения.</p> <p>Лекция 3: Предсказание вероятности возникновения события по значениям множества признаков (логистическая регрессия).</p> <p>Лабораторная работа 1: Линейная регрессия как задача контролируемого обучения.</p> <p>Лабораторная работа 2: Предсказание вероятности возникновения события по значениям множества признаков.</p> <p>Самостоятельная работа: Текущая проработка теоретического материала (изучение учебного материала по конспектам лекций, литературным источникам и составление конспекта, развернутого плана). Выполнение практического задания по теме №1 в целях подготовки к выполнению лабораторных работ №1-2.</p>	30
2.	<p>Тема 2. Инструментальные средства проектирования систем искусственного интеллекта</p> <p>Лекция 4: Статистические методы обучения. Метод опорных векторов. «Ядерные» машины.</p> <p>Лекция 5: Интеллектуальные агенты. Агенты и варианты среды. Структура агентов. Обучающиеся агенты.</p> <p>Лекция 6: Уменьшение размерности данных с помощью метода главных компонент.</p> <p>Лекция 7: Решение проблем посредством поиска. Примеры задач. Измерение производительности решения задачи.</p> <p>Лабораторная работа 3: Изучение статистических методов обучения. Метод опорных векторов.</p> <p>Лабораторная работа 4: Изучение кластеризации данных методом k-средних. Уменьшение размерности данных с помощью метода главных компонент.</p> <p>Самостоятельная работа: Текущая проработка теоретического материала (изучение учебного материала по конспектам лекций, литературным источникам и составление конспекта, развернутого плана). Выполнение практического задания по теме №2 в целях подготовки к выполнению лабораторных работ №3-4.</p>	32
3.	Промежуточная аттестация	2

2.3. Учебно-тематический план рабочей программы

№ п/п	Наименование и краткое содержание структурного элемента (раздела) Программы	Количество часов		
		аудиторных		самостоятельной работы
		лекции, семинары	практические занятия	
1	Тема 1. Искусственный интеллект как научное направление	6	8	16
2	Тема 2. Инструментальные средства проектирования систем искусственного интеллекта	8	8	16
3	Промежуточная аттестация	2		
	Итого	64		

2.4. Контроль и оценка результатов освоения рабочей программы

Образовательная организация высшего образования, реализующая рабочую программу, обеспечивает организацию и проведение *текущего, промежуточного* контроля.

Текущий контроль проводится преподавателем на основе оценивания результатов лабораторных работ и самостоятельной работы обучающихся. Промежуточный контроль проводится в форме демонстрационного экзамена, состоящего из двух частей: тестирования и решения практического задания. Формы и методы текущего и промежуточного контроля, критерии оценивания доводятся до сведения обучающихся в начале обучения.

2.4.1. Примеры оценочных средств

Слушатель знает ключевые понятия, цели и задачи использования машинного обучения; методологические основы применения алгоритмов машинного обучения.

Примеры тестовых заданий:

Задание	Варианты ответов
1. Какое из перечисленных понятий НЕ входит в понятие искусственного интеллекта?	А) Машинное обучение В) Глубинное обучение С) Экспертные системы D) Аналитика данных
2. Как определяется индикатор ошибки в задачах обучения по прецедентам?	А) Через целевую функцию $y^*(x)$ В) Через функцию потерь $I(a, x)$ С) Через регрессию при $Y=R$
3. Если алгоритмы a и a' допускают ошибки на одних и тех же объектах, то их называют:	А) Неразличимыми на выборке X^L В) Индикатором ошибок С) Средней частотой ошибок

Проверка умений слушателей визуализировать результаты работы алгоритмов машинного обучения, выбирать метод машинного обучения, соответствующий исследовательской задаче, интерпретировать полученные результаты.

Примеры типовых практических заданий:

Разработать, используя язык Python, программу, выполняющую следующие функции:

1. Определить три образа: круг, треугольник и квадрат;
2. Определить обучающую выборку для каждого из образов, состоящую из 5 отличающихся друг от друга изображений соответствующей геометрической фигуры;
3. Задать распознаваемое изображение;
4. Определить расстояние по Хэммингу от распознаваемого изображения до изображений обучающей выборки;
5. Определить потенциал, создаваемый каждым из образов в распознаваемой точке и выбрать образ с наибольшим потенциалом.

2.5. Учебно-методические материалы для обучающихся и преподавателей

Для теоретической подготовки используются учебная литература и комплект теоретических материалов, в виде электронных презентаций к лекциям, содержащие основные определения, математический аппарат и иллюстративный материал.

Для подготовки к лабораторным работам используется лабораторный практикум, содержащий описание методов и подходов к выполнению лабораторных заданий.

Электронные презентации по дисциплине, методические рекомендации, а также лабораторный практикум размещаются в информационной системе iSpring <https://edu.miet.ru/>.

2.6. Материально-техническое обеспечение реализации рабочей программы

Компьютерная техника с возможностью подключения к сети «Интернет» и установленным программным обеспечением: Операционная система Microsoft Windows от 7 версии и выше, Мой Офис, браузер (Яндекс и пр.), среда разработки (Jupiter, Notepad++, VS CODE), Git, Gitlab, Gitverse, виртуальная машина/контейнер (Virtual Box, Docker) с установленной debian-based операционной системой (ОС AstraLinux, ОС Ubuntu), интерпретатор языка Python, системы визуализации данных PlantUML, база данных PostgreSQL.

2.7. Информационное обеспечение реализации рабочей программы

Перечень учебной литературы

1. Уатт Д., Борхани Р., Катсагелос А. Машинное обучение: основы, алгоритмы и практика применения: Пер. с англ //СПб.: БХВ Петербург. – 2022.
2. Коэльо Л., Ричарт В. Построение систем машинного обучения на языке Python. 2-е издание / пер. с англ. Слинкин А. А. – Москва : ДМК Пресс , 2022. — 302 с. - ISBN 978-5-97060-330-7. - Текст : электронный. – URL: <https://www.litres.ru/book/villi-richart/postroenie-sistem-mashinnogo-obucheniya-na-yazyke-python-22873803> (дата обращения: 17.07.2024). – Режим доступа: по подписке.

Информационные ресурсы

1. Лань: Электронно-библиотечная система Издательства Лань. - СПб., 2011-. - URL: <https://e.lanbook.com> (дата обращения: 15.07.2024). - Режим доступа: для авторизованных пользователей МИЭТ
2. eLIBRARY.RU: Научная электронная библиотека: сайт. - Москва, 2000 -. -

URL: <https://www.elibrary.ru/defaultx.asp> (дата обращения: 15.07.2024). - Режим доступа: для зарегистрированных пользователей

3. Единое окно доступа к информационным ресурсам: сайт /ФГАУ ГНИИ ИТТ "Информика". - Москва, 2005-2010. - URL: <http://window.edu.ru/catalog/> (дата обращения: 15.07.2024).

3. Рабочая программа учебной дисциплины «Основы нейросетевых технологий»

3.1. Область применения рабочей программы

Рабочая программа «**Основы нейросетевых технологий**» (далее – рабочая программа) является частью дополнительной профессиональной программы профессиональной переподготовки ИТ-профиля «**Анализ данных с применением технологий машинного обучения**» и направлена на *получение студентами знаний и умений использования нейронных сетей для решения задач профессиональной деятельности в области разработки компьютерного программного обеспечения.*

3.2. Структура и краткое содержание рабочей программы

№ п/п	Наименование тем, виды учебной работы и краткое содержание учебного материала	Объем, часов
1.	<p>Тема 1. Основы нейронных сетей и методы их обучения</p> <p>Лекция 1: Классификация нейронных сетей. Модель перцептрона Розенблатта. Сверточные нейронные сети. Рекуррентные нейронные сети. Элементарный перцептрон. Составные элементы перцептрона.</p> <p>Лекция 2: Факторизация данных. Факторизация в машинном обучении. Нормализация данных. Точность и полнота. Матрица неточностей. F-мера. ROC-кривая и AUC ROC. Функции активации. Пороговая функция. Линейная функция. Сигмоидная функция. Гиперболический тангенс. Линейный выпрямитель. Функции потерь. Бинарная кросс-энтропия. Категориальная кросс-энтропия. Оптимизаторы. Механизм обратного распространения. Стохастический градиентный спуск. Моментум. Ускоренный градиент Нестерова. Adagrad. RMSProp. Adam.</p> <p>Лабораторная работа 1: Нейронные сети для бинарной классификации. Лабораторная работа 2: Нейронные сети для мультиклассовой классификации.</p> <p>Самостоятельная работа: Текущая проработка теоретического материала (изучение учебного материала по конспектам лекций, литературным источникам и составление конспекта, развернутого плана). Выполнение практического задания по теме №1 в целях подготовки к выполнению лабораторных работ №1-2.</p>	28
2.	<p>Тема 2. Топология нейронных сетей. Обучение с подкреплением</p> <p>Лекция 3: Сверточные нейронные сети. Локализация R-CNN. Семантическая сегментация Mask R-CNN. Основные составляющие. Основные виды свертки. Дополнение краевых значений. Bitmap vs. RGB. Изменение размеров. Pooling-слой. Субдискретизация, подвыборка. Варианты пулинга. Полносвязный слой. Функция активации.</p> <p>Лекция 4: Обработка естественного языка. Инструменты для анализа текстов. n-граммы. Векторизация. Токенизация. Регулярные выражения. Лемматизация. Стеммизация. Нормализация. Рекуррентные нейронные сети. Рекурсивная нейронная сеть. Обучение RNN. Улучшения RNN. LSTM.</p> <p>Лекция 5: Обучение с подкреплением. Жадная стратегия. ε-жадная стратегия. Оптимистические оценки. Стратегия softmax. Марковское свойство. Марковское свойство в обучении с подкреплением. Ценность состояний. Q-learning. Метод Монте-Карло. Метод обучения на каждом временном шаге. Сравнение TD и MC методов.</p> <p>Лабораторная работа 3: Нейронная сеть для предсказания непрерывной величины. Лабораторная работа 4: Проведение кластеризации над множеством.</p> <p>Самостоятельная работа: Текущая проработка теоретического материала (изучение учебного</p>	30

№ п/п	Наименование тем, виды учебной работы и краткое содержание учебного материала	Объем, часов
	материала по конспектам лекций, литературным источникам и составление конспекта, развернутого плана). Выполнение практического задания по теме №2 в целях подготовки к выполнению лабораторных работ №3-4.	
3.	Промежуточная аттестация	2

3.3. Учебно-тематический план рабочей программы

№ п/п	Наименование и краткое содержание структурного элемента (раздела) Программы	Количество часов		
		аудиторных		самостоятельной работы
		лекции, семинары	практические занятия	
1	Тема 1. Основы нейронных сетей и методы их обучения	4	8	16
2	Тема 2. Топология нейронных сетей. Обучение с подкреплением	6	8	16
3	Промежуточная аттестация	2		
	Итого	60		

3.4. Контроль и оценка результатов освоения рабочей программы

Образовательная организация высшего образования, реализующая рабочую программу, обеспечивает организацию и проведение *текущего, промежуточного* контроля.

Текущий контроль проводится преподавателем на основе оценивания результатов лабораторных работ и самостоятельной работы обучающихся. Промежуточный контроль проводится в форме демонстрационного экзамена, состоящего из двух частей: тестирования и решения практического задания. Формы и методы текущего и промежуточного контроля, критерии оценивания доводятся до сведения обучающихся в начале обучения.

3.4.1. Примеры оценочных средств

Слушатель знает современные технологии построения и использования нейронных сетей.

Примеры тестовых заданий:

Задание	Варианты ответов
1. В чем отличие алгоритма кластеризации DBSCAN от K-means при работе с неструктурированными данными?	<p>A) DBSCAN имеет фиксированное количество кластеров, определяемое заранее.</p> <p>B) DBSCAN требует явного задания начальных центроидов для кластеров.</p> <p>C) DBSCAN работает только с числовыми данными, не подходящими для категориальных или текстовых данных.</p> <p>D) DBSCAN способен обнаруживать кластеры любой формы и размера, в отличие от k-means, который предполагает сферические кластеры.</p>
2. Какое утверждение наиболее точно описывает ограничения многослойного перцептрона (по Румельхарту)?	<p>A) Перцептрон по Румельхарту легко справляется с нелинейными задачами классификации благодаря своей гибкой структуре.</p> <p>B) Перцептрон по Румельхарту может автоматически адаптироваться к различным формам данных без дополнительной предварительной обработки.</p> <p>C) Одним из основных ограничений перцептрона Розенблатта является его способность работать только с линейно разделимыми данными.</p> <p>D) Перцептрон по Румельхарту может эффективно обрабатывать многоклассовые задачи классификации без необходимости модификаций.</p>
3. Какой вывод можно сделать о качестве модели машинного обучения, если значение AUC-ROC равно 0.8?	<p>A) Модель обладает хорошей способностью различать между классами, так как значение AUC-ROC близко к 1.</p> <p>B) Модель обладает плохой способностью различать между классами, так как значение AUC-ROC близко к 0.</p> <p>C) Невозможно произвести оценку модели на основе значения AUC-ROC.</p> <p>D) Значение AUC-ROC равно 0.8 указывает на случайный выбор класса, модель неинформативна.</p>
4. Какой вывод можно сделать о качестве модели машинного обучения, если коэффициент детерминации равен 0.8?	<p>A) Модель обладает хорошей предсказательной способностью, так как коэффициент детерминации близок к 1.</p> <p>B) Модель имеет низкую предсказательную способность, так как квадрат коэффициента детерминации близок к 0.5.</p> <p>C) Невозможно произвести оценку модели на основе коэффициента детерминации.</p> <p>D) Модель имеет низкую предсказательную способность, так как коэффициент детерминации меньше 1.</p>
5. Сколько операций свертки происходит на каждом шаге при применении операции Valid свертки с фильтром размером 3x3 к изображению размером 5x5 с шагом (stride) равным 1?	<p>A) 6</p> <p>B) 9</p> <p>C) 25</p> <p>D) 1</p>

Проверка умений слушателей использовать современные технологии для построения нейронных сетей, для выбора стратегии обучения и самообучения нейронной сети.

Примеры типовых практических заданий:

1. Разработать программу, моделирующую поведение искусственного трехходового нейрона. Ниже представлена таблица вариантов заданий.

Номер варианта	Тип нейрона
1	Перцептрон
2	Сигмоидальный нейрон
3	Радиальный нейрон
4	Инстар Гроссберга
5	Нейроны WTA

2. Отладить модель нейрона и процедуру его обучения на произвольных двухмерных

данных. Рекомендуется, в тех ситуациях, когда это возможно, использовать режим обучения "оффлайн".

3.5. Учебно-методические материалы для обучающихся и преподавателей

Для теоретической подготовки используются учебная литература и комплект теоретических материалов, в виде электронных презентаций к лекциям, содержащие основные определения, математический аппарат и иллюстративный материал.

Для подготовки к лабораторным работам используется лабораторный практикум, содержащий описание методов и подходов к выполнению лабораторных заданий.

Электронные презентации по дисциплине, методические рекомендации, а также лабораторный практикум размещаются в информационной системе iSpring <https://edu.miet.ru/>.

3.6. Материально-техническое обеспечение реализации рабочей программы

Компьютерная техника с возможностью подключения к сети «Интернет» и установленным программным обеспечением: Операционная система Microsoft Windows от 7 версии и выше, Мой Офис, браузер (Яндекс и пр.), среда разработки (Jupiter, Notepad++, VS CODE), Git, Gitlab, Gitverse, виртуальная машина/контейнер (Virtual Box, Docker) с установленной debian-based операционной системой (ОС AstraLinux, ОС Ubuntu), интерпретатор языка Python, системы визуализации данных PlantUML, база данных PostgreSQL.

3.7. Информационное обеспечение реализации рабочей программы

Перечень учебной литературы

1. Прикладные задачи свёрточных нейронных сетей : учебное пособие / Л. Г. Гагарина, А. Р. Федоров, П. А. Федоров, А.А. Доронина [и др.]; Министерство образования и науки РФ, Национальный исследовательский университет "МИЭТ". - Москва : МИЭТ, 2020. – 76с. - ISBN 978-5-7256-0950-9.

2. Основы нечеткой логики и нейросетевые алгоритмы: учебно-методическое пособие /А. П. Ширяев, А. Ф. Петрова, Е. Н. Петров, А.А. Доронина [и др.]; Министерство образования и науки РФ, Национальный исследовательский университет "МИЭТ"; Под ред. Л.Г. Гагариной.- М.: МИЭТ, 2020. - 88 с

Информационные ресурсы

1. Лань: Электронно-библиотечная система Издательства Лань. - СПб., 2011-. - URL: <https://e.lanbook.com> (дата обращения: 15.07.2024). - Режим доступа: для авторизованных пользователей МИЭТ

2. eLIBRARY.RU: Научная электронная библиотека: сайт. - Москва, 2000 -. - URL: <https://www.elibrary.ru/defaultx.asp> (дата обращения: 15.07.2024). - Режим доступа: для зарегистрированных пользователей

3. Единое окно доступа к информационным ресурсам: сайт /ФГАУ ГНИИ ИТТ "Информика". - Москва, 2010. -URL: <http://window.edu.ru/catalog/> (дата обращения: 15.07.2024).

4. Рабочая программа учебной дисциплины «Аналитика больших данных»

4.1. Область применения рабочей программы

Рабочая программа «Аналитика больших данных» (далее – рабочая программа) является частью дополнительной профессиональной программы профессиональной переподготовки ИТ-профиля «Анализ данных с применением технологий машинного обучения» и направлена на *получение студентами знаний теоретических и практических аспектов технологий, в основе которых лежат принципы получения, преобразования, обработки, хранения и анализа больших объемов данных.*

4.2. Структура и краткое содержание рабочей программы

№ п/п	Наименование тем, виды учебной работы и краткое содержание учебного материала	Объем, часов
1.	<p>Тема 1. Задачи классификации и регрессии</p> <p>Лекция 1: Основные определения больших данных. Сферы использования. Области внедрения. Технологии для работы с большими данными. Перспективы технологий машинного обучения. Этапы и ошибки работы с большими данными.</p> <p>Лекция 2: Ансамблевые методы. Задачи классификации. Нейронные сети. Решающее дерево. Задачи регрессии. Случайный лес и его параметры. Преимущества и недостатки случайного леса. Бустинг и его виды.</p> <p>Лабораторная работа 1: Построение бинарного классификатора.</p> <p>Лабораторная работа 2: Построение регрессора для предсказания непрерывной величины.</p> <p>Самостоятельная работа: Текущая проработка теоретического материала (изучение учебного материала по конспектам лекций, литературным источникам и составление конспекта, развернутого плана). Выполнение практического задания по теме №1 в целях подготовки к выполнению лабораторных работ №1-2.</p>	28
2.	<p>Тема 2. Нейронные сети как инструмент решения прикладных задач</p> <p>Лекция 3: Типы признаков. Конструирование и отбор признаков. Основные категории алгоритмов отбора признаков (методы фильтрации, обертывания, вложения). Глубокий синтез признаков.</p> <p>Лекция 4: Анализ временных рядов. Составляющие временных рядов. Наивное предсказывание. Экспоненциальное сглаживание. Динамические линейные модели и их оценка.</p> <p>Лабораторная работа 3: Определение типа дорожного знака.</p> <p>Лабораторная работа 4: Предсказание эмоциональной окраски отзыва на основе анализа естественного языка</p> <p>Самостоятельная работа: Текущая проработка теоретического материала (изучение учебного материала по конспектам лекций, литературным источникам и составление конспекта, развернутого плана). Выполнение практического задания по теме №2 в целях подготовки к выполнению лабораторных работ №3-4.</p>	28
3.	Промежуточная аттестация	2

4.3. Учебно-тематический план рабочей программы

№ п/п	Наименование и краткое содержание структурного элемента (раздела) Программы	Количество часов		
		аудиторных		самостоятельной работы
		лекции, семинары	практические занятия	
1	Тема 1. Задачи классификации и регрессии	4	8	16
2	Тема 2. Нейронные сети как инструмент решения прикладных задач	4	8	16
3	Промежуточная аттестация	2		
	Итого	58		

4.4. Контроль и оценка результатов освоения рабочей программы

Образовательная организация высшего образования, реализующая рабочую программу, обеспечивает организацию и проведение *текущего, промежуточного* контроля.

Текущий контроль проводится преподавателем на основе оценивания результатов лабораторных работ и самостоятельной работы обучающихся. Промежуточный контроль проводится в форме демонстрационного экзамена, состоящего из двух частей: тестирования и решения практического задания. Формы и методы текущего и промежуточного контроля, критерии оценивания доводятся до сведения обучающихся в начале обучения.

4.4.1. Примеры оценочных средств

Слушатель знает методики и алгоритмы анализа и обработки больших данных с использованием современных языков программирования высокого уровня.

Примеры тестовых заданий:

Задание	Варианты ответов
1. Как влияет глубина решающего дерева на его способность к обобщению и сложность модели?	<p>А) Увеличение глубины решающего дерева всегда улучшает способность модели к обобщению.</p> <p>В) Увеличение глубины решающего дерева может привести к переобучению модели.</p> <p>С) Глубина решающего дерева не влияет на сложность модели.</p> <p>Д) Глубина решающего дерева имеет линейное влияние на сложность модели.</p>
2. Какой принцип лежит в основе алгоритма адаптивного бустинга (AdaBoost), который делает его особенно эффективным для обучения слабых моделей?	<p>А) Адаптивный бустинг использует градиентный спуск для оптимизации весов моделей.</p> <p>В) Адаптивный бустинг активно изменяет структуру моделей в процессе обучения.</p> <p>С) Адаптивный бустинг фокусируется на ошибках предыдущих моделей, чтобы улучшить обобщающую способность.</p> <p>Д) Адаптивный бустинг добавляет случайные компоненты к предсказаниям, чтобы повысить разнообразие моделей.</p>
3. Каковы основные преимущества многоуровневого стекинга по сравнению с одноуровневым стекингом при построении ансамблей моделей?	<p>А) Многоуровневый стекинг обладает более высокой вычислительной эффективностью.</p> <p>В) Многоуровневый стекинг позволяет более гибко управлять зависимостями между базовыми моделями.</p>

	<p>С) Многоуровневый стекинг имеет более низкую степень сложности и переобучения.</p> <p>Д) Многоуровневый стекинг обеспечивает более высокую точность предсказаний без необходимости дополнительной настройки моделей.</p>
<p>4. Что представляет собой метод кросс-валидации (cross-validation) и какая его роль при оценке и выборе моделей?</p>	<p>А) Метод кросс-валидации используется для разделения данных на обучающую и тестовую выборки для оценки обобщающей способности модели.</p> <p>В) Метод кросс-валидации позволяет оценить качество модели на различных подмножествах данных и уменьшить риск переобучения.</p> <p>С) Метод кросс-валидации применяется для определения оптимальной степени полинома и настройки гиперпараметров модели.</p> <p>Д) Метод кросс-валидации осуществляет проверку модели на новых данных для подтверждения её стабильности и точности предсказаний.</p>
<p>5. Какие основные принципы лежат в основе метода гребневой регрессии (ridge regression), и в чем заключается его отличие от обычной линейной регрессии?</p>	<p>А) Гребневая регрессия минимизирует сумму квадратов остатков модели при условии ограничения на сумму квадратов весов параметров модели.</p> <p>В) Гребневая регрессия рассматривает только линейные зависимости между признаками и целевой переменной, в то время как обычная линейная регрессия может учитывать нелинейные взаимосвязи.</p> <p>С) Основным принципом гребневой регрессии заключается в добавлении регуляризационного члена к функции потерь, чтобы сократить разброс оценок параметров модели.</p> <p>Д) Отличие гребневой регрессии от обычной линейной регрессии заключается в том, что в гребневой регрессии используется метод градиентного спуска для оптимизации весов модели.</p>

Проверка умений слушателей применять современные информационные системы и технологии для реализации моделей многократного использования при обработке больших данных методами машинного и глубокого обучения.

Примеры типовых практических заданий:

Необходимо проанализировать данные датасета.

План этапов:

1. Приведение набора данных к виду, доступному для обработки.
2. Устранение проблем по недостаточному представлению части классов в наборе данных.
3. Построение ансамблевых моделей, сравнение эффективности разных подходов для определенного класса задач.
4. Выявление скрытых зависимостей в данных с помощью изученных методов обработки, построение аналитической модели процесса.

4.5. Учебно-методические материалы для обучающихся и преподавателей

Для теоретической подготовки используются учебная литература и комплект теоретических материалов, в виде электронных презентаций к лекциям, содержащие основные определения, математический аппарат и иллюстративный материал.

Для подготовки к лабораторным работам используется лабораторный практикум, содержащий описание методов и подходов к выполнению лабораторных заданий.

Электронные презентации по дисциплине, методические рекомендации, а также лабораторный практикум размещаются в информационной системе iSpring <https://edu.miet.ru/>.

4.6. Материально-техническое обеспечение реализации рабочей программы

Компьютерная техника с возможностью подключения к сети «Интернет» и установленным программным обеспечением: Операционная система Microsoft Windows от 7 версии и выше, Мой Офис, браузер (Яндекс и пр.), среда разработки (Jupiter, Notepad++, VS CODE), Git, Gitlab, Gitverse, виртуальная машина/контейнер (Virtual Box, Docker) с установленной debian-based операционной системой (ОС AstraLinux, ОС Ubuntu), интерпретатор языка Python, системы визуализации данных PlantUML, база данных PostgreSQL.

4.7. Информационное обеспечение реализации рабочей программы

Перечень учебной литературы

1. Брантон С. Л., Куц Дж. Н. Анализ данных в науке и технике / пер. с англ. А. А. Слинкина. - М.: ДМК Пресс, 2021. - 542 с. ISBN 978-5-97060-910-1
2. Хайндман Р., Атанасопулос Дж. / пер. с англ. Логунов А. - М.: ДМК Пресс, 2023. – 458 с. ISBN: 978-5-93700-151-1

Информационные ресурсы

1. Лань: Электронно-библиотечная система Издательства Лань. - СПб., 2011-. - URL: <https://e.lanbook.com> (дата обращения: 15.07.2024). - Режим доступа: для авторизованных пользователей МИЭТ
2. eLIBRARY.RU: Научная электронная библиотека: сайт. - Москва, 2000 -. - URL: <https://www.elibrary.ru/defaultx.asp> (дата обращения: 15.07.2024). - Режим доступа: для зарегистрированных пользователей
3. Единое окно доступа к информационным ресурсам: сайт /ФГАУ ГНИИ ИТТ "Информика". - Москва, 2005-2010. - URL: <http://window.edu.ru/catalog/> (дата обращения: 15.07.2024).

5. Рабочая программа практики

5.1. Область применения рабочей программы

Рабочая программа практики является частью дополнительной профессиональной программы профессиональной переподготовки ИТ-профиля **«Анализ данных с применением технологий машинного обучения»** и нацелена на *получение умений и практических навыков разработки программного обеспечения для аналитики больших данных с использованием нейросетевых моделей.*

5.2. Объем практики

Объём практики — 32 ак. часа.

Практика организуется с 33 по 36 неделю.

5.3. Содержание практики

Практическая подготовка слушателей проходит в индустриальных партнерах ИТ-сферы. На практике слушателям предлагается индивидуальное задание на разработку программного обеспечения для анализа данных с использованием нейронных сетей по теме, выданной на месте практики и согласованной с руководителем практики. Разработка осуществляется слушателями самостоятельно и курируется представителем с места практики.

При прохождении учебной практики слушатели развивают приобретенные навыки, знакомятся с имеющимся технологическим оборудованием на производстве, изучают инструкции по работе с оборудованием и технике безопасности, нормативную документацию, самостоятельно проводят поиск научно-технической информации по тематике проекта. Принимают участие в выполнении проектов индивидуально или в составе проектной группы, под руководством руководителя. Планируют этапы выполнения проекта, учатся оформлять и утверждать техническое задание.

В ходе прохождения практической подготовки слушателям необходимо разработать программное обеспечение для анализа данных по согласованной теме, содержащие следующие реализации:

1. Разработана концепция проекта.
2. Проведена предварительная оценка рисков проекта.
3. Проведен анализ существующих решений.
4. Спроектирована нейросетевая модель.
5. Разработана или использована/изменена существующая архитектура нейронной сети.

По результатам выполнения этапов индивидуального задания, слушателем формируется итоговая работа, представляемая на итоговую аттестацию по настоящей программе.

5.4. Пример типового задания по практике

Проверка слушателей на опыт деятельности:

- разработки и отладки параллельных процессов; применения средств виртуализации вычислений;
- применения методов машинного обучения, построения и оценки качества моделей;
- использования нейронной сети для построения системы принятия решений;
- применения библиотек современных языков программирования для анализа, визуализации данных.

Пример типового задания на практику:

Требуется разработать приложение для анализа и прогнозирования климатических данных. С интуитивно понятным графическим интерфейсом.

Технические требования:

1. Загрузка и хранение данных:

- возможность загрузки и хранения разнообразных климатических данных, включая температуру, осадки, влажность, скорость ветра и др (максимальный объем данных не должен превышать 10 гб);
- поддержка различных форматов данных для импорта и экспорта (CSV и XLSX);

2. Визуализация данных:

- отображение климатических данных на графиках, диаграммах (с возможностью задать временной отрезок и место);
- возможность выбора временного интервала и места (города или метеостанции) для анализа;

3. Анализ и обработка данных:

- выявление трендов (долгосрочные изменения в погодных условиях), и корреляций между различными погодными параметрами;
- проведение статистического анализа данных для выявления закономерностей (нахождение средних значений, медиан, отклонений);

4. Прогнозирование погоды:

- реализация моделей машинного обучения для прогнозирования погоды на ближайшие дни или недели (точность не оценивается, метод прогнозирования на усмотрение участников);

5. Мониторинг погодных явлений:

- возможность подключения к внешним источникам данных о погоде и климате (рекомендуется использовать следующий API: <https://www.gismeteo.ru/api/>);
- вывод текущей информации о климатической ситуации в отдельной вкладке (окне);

6. Настройки и персонализация:

- возможность настройки параметров анализа и прогноза в соответствии с потребностями пользователя (например: изменение промежутка времени для анализа и прогнозирования);

7. Система уведомлений:

- вывод Push-уведомлений: для предупреждения пользователя на основе

обнаруженных погодных изменений (вывод на экран или в самом приложении уведомлений);

8. Безопасность и доступ:

- авторизация пользователей через БД (хранение логина и пароля в БД).

В процессе выполнения задания должны быть выполнены следующие задачи:

1. Сформулировать цели и задачи проекта.
2. Разработать концепцию будущего проекта.
3. Провести предварительную оценку рисков будущего проекта.
4. Провести анализ существующих решений.
5. Спроектировать нейросетевую модель.
6. Разработать или использовать/изменить существующую архитектуру нейронной сети для выполнения поставленных задач аналитики данных.

5.5. Учебно-методические материалы для обучающихся

Для выполнения индивидуального задания в рамках практики используется учебная литература и примеры выполнения практических заданий, представленные в соответствующих разделах учебных дисциплин, в рамках которых происходит обучение подзадачам разработки программного обеспечения на основе нейросетевых моделей для анализа данных.

5.6. Материально-техническое обеспечение реализации рабочей программы

Компьютерная техника с возможностью подключения к сети «Интернет» и установленным программным обеспечением: Операционная система Microsoft Windows от 7 версии и выше, Мой Офис, браузер (Яндекс и пр.), среда разработки (Jupiter, Notepad++, VS CODE), Git, Gitlab, Gitverse, виртуальная машина/контейнер (Virtual Box, Docker) с установленной debian-based операционной системой (ОС AstraLinux, ОС Ubuntu), интерпретатор языка Python, системы визуализации данных PlantUML, база данных PostgreSQL.

5.7. Информационное обеспечение реализации рабочей программы

Информационные ресурсы

1. Лань: Электронно-библиотечная система Издательства Лань. - СПб., 2011-. - URL: <https://e.lanbook.com> (дата обращения: 15.07.2024). - Режим доступа: для авторизованных пользователей МИЭТ
2. eLIBRARY.RU: Научная электронная библиотека: сайт. - Москва, 2000 -. - URL: <https://www.elibrary.ru/defaultx.asp> (дата обращения: 15.07.2024). - Режим доступа: для зарегистрированных пользователей
3. Единое окно доступа к информационным ресурсам: сайт /ФГАУ ГНИИ ИТТ "Информика". - Москва, 2005-2010. - URL: <http://window.edu.ru/catalog/> (дата обращения: 15.07.2024).

VI. Итоговая аттестация по Программе

После завершения обучения по Программе и прохождения итоговой оценки сформированности цифровых компетенций обучающиеся допускаются к итоговой аттестации.

Итоговая аттестация проводится с участием представителей профильных индустриальных партнёров в форме защиты итоговой аттестационной работы (далее – ИАР). На итоговую аттестацию представляется итоговая аттестационная работа, включающая текстовые и графические результаты, полученные в ходе выполнения индивидуального задания по согласованной теме, описанного в программе практики.

1. Требования к содержанию и оформлению ИАР

ИАР оформляется в соответствии с требованиями ГОСТ 7.32-2001 «Отчет о научно-исследовательской работе. Структура и правила оформления», ГОСТ 7.1-2003 «Библиографическая запись. Библиографическое описание», ГОСТ Р7.05-2008 «Библиографическая ссылка. Общие требования и правила составления» и содержит следующие разделы:

- титульный лист с указанием темы (названия индивидуального задания);
- содержание;
- введение с описанием технического задания;
- основная часть описания алгоритма работы ПО;
- заключение;
- приложение А: листинг кода разработанного программного обеспечения.

Основными требованиями к работе являются:

- четкость и логическая последовательность изложения материала;
- краткость и точность формулировок, исключающая возможность неоднозначного их толкования;
- конкретность изложения полученных результатов, их анализа и теоретических положений;
- обоснованность выводов, рекомендаций и предложений.

Содержание ИАР должно соответствовать названию темы.

К расположению материала на каждой странице текста предъявляются следующие требования:

- текст печатается на одной стороне листа белой бумаги формата А4;
- листы ИАР оформляются без рамок;
- цвет шрифта черный;
- шрифт TimesNewRoman;
- размер шрифта 12;
- абзацный отступ: слева и справа 0 см, первая строка 1,25 см
- абзацный интервал: перед и после 0 см, междустрочный 1,5 см;
- выравнивание текста на странице «по ширине»;

- поля разметки страницы «Обычные»: верхнее поле страницы 2см, нижнее поле страницы 2см, расстояние от левого края страницы до границы текста 3см, расстояние от правой границы текста до края страницы 1,5см;

- все рисунки, таблицы, формулы и программный код (или часть программного кода) в зависимости от их размера располагаются в тексте непосредственно после того абзаца, в котором они были впервые упомянуты, или на следующей странице, а при необходимости – в приложении;

- нумерация страниц сквозная по всему тексту (титульный лист не нумеруется): в нижнем колонтитуле по центру без абзацных отступов, арабские цифры, шрифтом Times New Roman, размером шрифта 10.

Оформление иллюстраций:

- в качестве иллюстраций в работах могут быть представлены чертежи, схемы, диаграммы, рисунки и т.п;

- все иллюстрации обозначают в тексте словом «рисунок»;

- иллюстрации могут быть выполнены как в черно-белом, так и в цветном варианте;

- рисунки в зависимости от их размера располагают в тексте непосредственно после того абзаца, в котором данный рисунок был впервые упомянут, или на следующей странице, а при необходимости – в приложении;

- рисунок должен располагаться в центре;

- все рисунки должны иметь наименование.

Оформленная работа должна быть сброшюрована.

Объем ИАР должен составлять не менее 15 страниц без учета приложений.

К оформлению таблиц предъявляются следующие требования:

- все таблицы должны иметь название;

- название таблицы перед самой таблицей в формате «Таблица 1 – Название», в отличие от основного текста оформлено с выравниванием по правому краю, без абзацного отступа, с междустрочным интервалом «Одинарный»;

- таблица выровнена по ширине окна;

- текст в таблице в отличие от основного текста оформлен междустрочным интервалом «Одинарный» без абзацного отступа;

- один абзацный отступ перед названием и после таблицы.

К оформлению рисунков предъявляются следующие требования:

- все рисунки должны иметь название;

- рисунок выравнивается по центру без абзацного отступа;

- название после рисунка в формате «Рисунок 1 – Название» в отличие от основного текста оформлено с выравниванием по центру, без абзацного отступа, с междустрочным интервалом «Одинарный»;

- один абзацный отступ перед рисунком и после его названия.

К оформлению формул предъявляются следующие требования:

- формула должна быть выполнена в редакторе формул Microsoft Office Word;

- текст формулы в отличие от основного текста оформлен с выравниванием по центру, без абзацного отступа, с междустрочным интервалом «Одинарный»;

- каждая формула, на которую идёт ссылка по тексту, должна быть пронумерована справа от формулы в формате «(1)», «(2)» и т.д.

К оформлению программного кода предъявляются следующие требования:

- текст кода в отличие от основного текста оформлен с выравниванием по левому краю, без абзацного отступа, с междустрочным интервалом «Одинарный», шрифтом Courier New, размером шрифта 10.

2. Критерии начисления баллов за ИАР

Показатель оценки	Критерии оценивания достижения показателя	Условия начисления баллов по критерию	Количество баллов
Разработано компьютерное программное обеспечение для анализа и прогнозирования по согласованной теме	Вид приложения соответствует техническим требованиям по согласованной теме	Все поля имеют нужный тип и параметры, вводимые данные корректно проверяются, выходные данные представлены в требуемом виде	10
		Есть несоответствия требуемым параметрам	5
Разработано компьютерное программное обеспечение для анализа и прогнозирования по согласованной теме	Вид приложения соответствует техническим требованиям по согласованной теме	Большая часть полей имеют не соответствующий тип и параметры, вводимые данные не проверяются, выходные данные не формируются	0
		Типы переменных соответствуют хранимым данным	20
	Структура данных соответствует техническим требованиям, выполняемым по согласованной теме	Типы не всех переменных соответствуют хранимым данным	10
		Типы переменных не соответствуют хранимым данным	0
	Алгоритм разработан корректно с учетом всех требований и ограничений по согласованной теме	Выполнение алгоритма выдает верный результат на всем диапазоне входных данных или функционал пользовательского интерфейса реализован корректно	10
		Выполнение алгоритма выдает неверный результат или функционал пользовательского интерфейса реализован некорректно	0
	Разработаны программные интерфейсы программного обеспечения	С использованием принципов объектно-ориентированного подхода	20
		Без использования принципов объектно-ориентированного подхода	0
	Были разработаны модели машинного обучения для прогнозирования по согласованной теме	Были разработаны модели машинного обучения для прогнозирования по согласованной теме	20

Показатель оценки	Критерии оценивания достижения показателя	Условия начисления баллов по критерию	Количество баллов
		Не были разработаны модели машинного обучения для прогнозирования по согласованной теме	0
Разработано компьютерное программное обеспечение для анализа и прогнозирования по согласованной теме	Была разработана или изменена существующая архитектура программного обеспечения для выполнения поставленных задач анализа и прогнозирования	Разработана или изменена существующая архитектура программного обеспечения	20
		Архитектура программного обеспечения не была разработана или изменена	0
Суммарный балл по показателю:			100

При выставлении оценки по итоговой аттестационной работе членом аттестационной комиссии, используется шкала, приведенная ниже в таблице.

Сумма баллов	Оценка
Менее 50	2 (неудовлетворительно)
50 – 69	3 (удовлетворительно)
70 – 85	4 (хорошо)
86 – 100	5 (отлично)

Оценка по итоговой аттестационной работе формируется как среднее арифметическое оценок членов аттестационной комиссии в соответствии с правилами округления математики.

VII. Завершение обучения по Программе

Лицам, завершившим обучение по Программе и достигших целевого уровня сформированности цифровых компетенций по результатам итоговой оценки и прошедших итоговую аттестацию, присваивается дополнительная ИТ-квалификация, установленная Программой.

При освоении Программы параллельно с получением высшего образования диплом о профессиональной переподготовке выдается не ранее получения соответствующего документа об образовании и о квалификации (за исключением лиц, имеющих среднее профессиональное или высшее образование).

Лицам, не прошедшим итоговую аттестацию или получившим на итоговой аттестации неудовлетворительные результаты, а также лицам, освоившим часть Программы и (или) отчисленным из образовательной организации высшего образования, реализующей Программу, выдается справка об обучении или о периоде обучения по образцу, самостоятельно устанавливаемому образовательной организацией высшего образования.

РАЗРАБОТЧИКИ ПРОГРАММЫ:

Ст. преподаватель Института СПИНТех



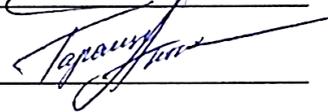
А.А. Доронина

Ст. преподаватель Института СПИНТех



А.О. Власова

Доцент Института СПИНТех, к.т.н.



А.В. Гаращенко

СОГЛАСОВАНО:

Директор ДРОП



Н.Ю. Соколова

Директор Института СПИНТех



Л.Г. Гагарина

Руководитель ЦК



В.В. Кокин