

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЯДЕРНЫЙ УНИВЕРСИТЕТ «МИФИ»
(НИЯУ МИФИ)



«УТВЕРЖДАЮ»

Первый проректор НИЯУ МИФИ

Нагорнов О.В.

15 мая 2022 г.

**ДОПОЛНИТЕЛЬНАЯ ПРОФЕССИОНАЛЬНАЯ ПРОГРАММА
ПРОФЕССИОНАЛЬНОЙ ПЕРЕПОДГОТОВКИ
Основы анализа данных и машинное обучение**

(наименование программы)

Москва, 2022

1. ОБЩАЯ ХАРАКТЕРИСТИКА ПРОГРАММЫ

Дополнительная профессиональная программа (программа профессиональной переподготовки) ИТ-профиля «Основы анализа данных и машинное обучение» (далее – Программа) разработана в соответствии с нормами:

- Федерального закона РФ от 29 декабря 2012 года № 273-ФЗ «Об образовании в Российской Федерации», с учетом требований приказа 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»;
- приказа Министерства образования и науки РФ от 23 августа 2017 г. N 816 «Об утверждении Порядка применения организациями, осуществляющими образовательную деятельность, электронного обучения, дистанционных образовательных технологий при реализации образовательных программ»;
- паспорта федерального проекта «Развитие кадрового потенциала ИТ-отрасли» национальной программы «Цифровая экономика Российской Федерации»;
- постановления Правительства Российской Федерации от 13 мая 2021 г. № 729 «О мерах по реализации программы стратегического лидерства «Приоритет-2030» (в редакции постановления Правительства Российской Федерации от 14 марта 2022 г. № 357 «О внесении изменений в постановление Правительства Российской Федерации от 13 мая 2021 г. № 729»);
- приказа Министерства цифрового развития, связи и массовых коммуникаций Российской Федерации от 28 февраля 2022 г. № 143 «Об утверждении методик расчета показателей федеральных проектов национальной программы «Цифровая экономика Российской Федерации» и признании утратившими силу некоторых приказов Министерства цифрового развития, связи и массовых коммуникаций Российской Федерации об утверждении методик расчета показателей федеральных проектов национальной программы «Цифровая экономика Российской Федерации» (далее – приказ Министерства цифрового развития, связи и массовых коммуникаций Российской Федерации № 143);

- устава Национального исследовательского ядерного университета «МИФИ» от 28.12.2018 (с учетом редакций от 26.02.2020 и 20.12.2021);
- положения о разработке и реализации программ дополнительного профессионального образования НИЯУ МИФИ СМК-ДП-7.5-04 от 17.03.2017;
- положения о применении электронного обучения и дистанционных образовательных технологий в НИЯУ МИФИ СМК-ДП-7.5-14 от 01.09.2017;
- положения о практической подготовке обучающихся НИЯУ МИФИ СМК-ДП-7.5-02 от 21.04.2021;
- стратегии инновационного развития и технологической модернизации Госкорпорации «Росатом» на период до 2030 года;
- Единой цифровой стратегии Госкорпорации «Росатом» 4.0;
- федерального государственного образовательного стандарта высшего образования по направлению подготовки 09.03.04 «Программная инженерия» (уровень бакалавриата), утвержденного приказом Минобрнауки России от 12 марта 2015 г. № 229, (далее вместе – ФГОС ВО));
- профессионального стандарта 06.001 «Программист», утвержденного приказом Министерства труда и социальной защиты РФ от 18 ноября 2013 г. № 679н;
- профессионального стандарта 06.019 «Технический писатель», утвержденного приказом Министерства труда и социальной защиты РФ от 8 сентября 2014 года N 612н;
- профессионального стандарта 06.022 «Системный аналитик», утвержденного приказом Министерства труда и социальной защиты РФ от от 28 октября 2014 года N 809н.

1.1. Цель реализации программы

Целью программы является ознакомление будущих специалистов с современными математическими и алгоритмическими подходами к построению систем, обучающихся или самообучающихся по располагаемой выборке данных, а также освоение студентами практических приемов машинного обучения, выбора обучающихся моделей и алгоритмов обучения и оценки качества функционирования, а также приобретение квалификации «Системный аналитик».

Программа разработана для обучающихся по специальностям и направлениям подготовки, не отнесенным к ИТ-сфере.

Полученные в ходе программы компетенции и навыки могут быть применены при использовании российского сертифицированного программного обеспечения, в том числе

сервиса для распознавания речи Yandex SpeechKit Vox, сервиса DataSphere Yandex Cloud для запуска моделей машинного обучения, редактора документов “Мой офис”. Указанное ПО, а также ряд других программных решений можно использовать на операционной системе Astra Linux, что позволяет говорить о технологической независимости будущих разработок специалистов, прошедших обучение по программам Цифровой кафедры НИЯУ МИФИ.

1.2. Характеристика нового вида профессиональной деятельности, новой квалификации

Область профессиональной деятельности слушателя, прошедшего обучение по программе профессиональной переподготовки для выполнения нового вида профессиональной деятельности «Анализ данных и машинное обучение», включает

- разработка, использование и исследование методов статистического анализа данных и машинного обучения;
- разработка систем машинного обучения.

Объектами профессиональной деятельности являются:

- модели и методы машинного обучения;
- программные системы, использующие модели машинного обучения;
- среды программной реализации моделей машинного обучения.

Программа профессиональной переподготовки ориентирована на подготовку слушателя к следующим видам профессиональной деятельности:

- производственно-технологическая.

Слушатель, успешно завершивший обучение по данной программе, должен решать следующие профессиональные задачи в соответствии с видами профессиональной деятельности:

производственно-технологическая деятельность:

- **Задача 1:** разрабатывать программные системы, использующие модели и методы машинного обучения;

- **Задача 2:** формулировать практические задачи в терминах машинного обучения, оценивать целесообразность применения машинного обучения для решения практических задач и выбирать подходы к их решению;
- **Задача 3:** проектировать и обучать модели машинного обучения для решения практических задач, оценивать качество обученных моделей;
- **Задача 4:** проводить экспериментальные исследования используемых моделей и методов анализа данных и машинного обучения.

1.3. Требования к результатам освоения программы

1.3.1 Перечень компетенций слушателя программы

Слушатель в результате освоения программы должен обладать следующими профессиональными компетенциями:

Код компетенции	Вид профессиональной деятельности	Формулировка компетенции
ПК-1	Производственно-технологическая	Способен структурировать общую схему решения прикладной задачи в соответствии с ее спецификой, а также определить совокупность и особенности применения математических методов для каждого из этапов полученной схемы. Способен определить совокупность математических методов для отдельного этапа решения прикладной задачи в рамках заданной схемы.
ПК-2		Способен модифицировать и применять актуальные алгоритмы компьютерной математики, а также реализовывать их в современных программных комплексах. Способен применять актуальные алгоритмы компьютерной математики и реализовывать их в современных программных комплексах.
ПК-3		Способен разрабатывать системное и

		прикладное программное обеспечение.
ПК-4		Применяет языки программирования для решения профессиональных задач.
ПК-5		Разрабатывает программное обеспечение.
ПК-6		Применяет принципы и основы алгоритмизации.
ПК-7		Применяет интегрированные среды разработки (IDE).
ПК-8		Применяет искусственный интеллект и машинное обучение.
ПК-9		Применяет математический аппарат для решения задач по оценке и разработки моделей.
ПК-10		Разрабатывает и реализует архитектуру ансамбля моделей.
ПК-11		Использует программные и технические средства для визуализации больших данных.
ПК-12		Решает задачи искусственного интеллекта (ИИ).
ПК-13		Осуществляет сбор и подготовку данных для обучения моделей искусственного интеллекта.
ПК-14		Разрабатывает и применяет методы машинного обучения (МО) для решения задач.
ПК-15		Применяет принципы защиты информации.
ПК-16		Применяет программное обеспечение для защиты информации.

1.3.2 Характеристика новой квалификации, связанной с видом профессиональной деятельности и трудовыми функциями в соответствии с профессиональным стандартом

Область профессиональной деятельности	Задача профессиональной деятельности	Код и наименование профессиональной компетенции	Профессиональный стандарт	Код и наименование ОТФ (ТФ)
<p><i>разработка, использование и исследование методов статистического анализа данных и машинного обучения</i> <i>разработка систем машинного обучения</i> <i>разработка, использование и исследование методов статистического анализа данных и машинного обучения</i></p>	<p>Задача 2: формулировать практические задачи в терминах машинного обучения, оценивать целесообразность применения машинного обучения для решения практических задач и выбирать подходы к их решению</p>	ПК-1	<p>06.022 «Системный аналитик»</p>	<p>Планирование разработки или восстановления требований к системе С/01.6 Анализ проблемной ситуации заинтересованных лиц С/02.6 Постановка целей создания системы С/04.6 Разработка концепции системы С/05.6 Разработка технического задания на систему С/06.6</p>
			<p>06.019 «Технический писатель»</p>	<p>Разработка технических документов, адресованных специалисту по информационным технологиям D/01.6 Описание технических решений с точки зрения специалиста по информационным технологиям D/02.6</p>

	<p>Задача 3: проектировать и обучать модели машинного обучения для решения практических задач, оценивать качество обученных моделей</p>	ПК-2, ПК-3	<p>06.022 «Системный аналитик»</p>	<p>Постановка целей создания системы С/04.6 Разработка концепции системы С/05.6</p>
	<p>Задача 4: проводить экспериментальные исследования используемых моделей и методов анализа данных и машинного обучения</p>	ПК-2, ПК-3	<p>06.022 «Системный аналитик»</p>	<p>Анализ проблемной ситуации заинтересованных лиц С/02.6</p>
	<p>Задача 1: разрабатывать программные системы, использующие модели и методы машинного обучения</p>	ПК-2, ПК-3	<p>06.001 «Программист»</p>	<p>Анализ требований к программному обеспечению D/01.6 Проектирование программного обеспечения D/03.6</p>

1.3.3 Характеристика новой и развиваемой цифровой компетенции в ИТ-сфере, связанной с уровнем формирования и развития в результате освоения Программы (на основе Модели цифровых компетенций, разрабатываемых университетом ИННОПОЛИС)

Наименование сферы	Код и наименование профессиональной компетенции	Пример инструментов	0 — способность не проявляется/ проявляется в степени, недостаточной для отнесения к 1 уровню сформированности компетенции	1 — способность проявляется под внешним контролем / при внешней постановке задачи/ обучающийся пользуется готовыми, рекомендованным и продуктами	2 — способность проявляется, но обучающийся эпизодически прибегает к экспертной консультации/ самостоятельно подбирает и пользуется готовыми продуктами	3 — способность проявляется системно / обучающийся модифицирует способность под определенные задачи / создает новый продукт, обучает других
Средства программной разработки	ПК-4 Применяет языки программирования для решения профессиональных задач	Python		+		
	ПК-5 Разрабатывает программное обеспечение	Python		+		
	ПК-6 Применяет принципы и основы алгоритмизации	Вычислительные алгоритмы		+		
	ПК-7 Применяет интегрированные среды разработки (IDE)	PyCharm,		+		

Искусственный интеллект и машинное обучение	ПК-8 Применяет искусственный интеллект и машинное обучение	Python		+		
	ПК-9 Применяет математический аппарат для решения задач по оценке и разработки моделей			+		
	ПК-10 Разрабатывает и реализует архитектуру ансамбля моделей	Python		+		
	ПК-11 Использует программные и технические средства для визуализации больших данных	Matplotlib		+		
	ПК-12 Решает задачи искусственного интеллекта (ИИ)	Python		+		
	ПК-13 Осуществляет сбор и подготовку данных для обучения моделей искусственного интеллекта	Python		+		
	ПК-14 Разрабатывает и применяет методы машинного обучения (МО) для решения задач	Python		+		
Защита информации	ПК-15 Применяет принципы защиты информации	Средства анализа защищенности, противодействия утечке информации, межсетевого		+		

		экранирования, шифрования и туннелирования, криптографической защиты информации, защиты web-приложений и др.				
	ПК-16 Применяет программное обеспечение для защиты информации	Антивирусы, firewall, Dr.Web, Kaspersky и т.д.		+		

1.3.4. Знания, умения и навыки, приобретаемые слушателем в рамках освоения материала Программы

В результате освоения программы слушатель должен:

Знать:

- основы языка программирования Python;
- научные библиотеки языка Python и их возможности для проведения научных вычислений, обработки и визуализации данных;
- основные парадигмы, понятия и принципы машинного обучения, а также типы задач, решаемых с помощью машинного обучения;
- математические основы регрессионного анализа данных и методов обучения и диагностики регрессионных моделей;
- математические основы классификации данных;
- непараметрические методы машинного обучения и особенности их использования для решения задач регрессии и классификации;
- основные подходы к обучению ансамблей и повышению качества моделей машинного обучения;
- основы кластерного анализа данных.

Уметь:

- применять язык программирования Python для решения вычислительных задач;
- применять научные библиотеки языка Python для обработки и визуализации данных, создания моделей машинного обучения и исследования их качества;
- формулировать практические задачи в терминах машинного обучения, оценивать целесообразность применения машинного обучения для решения практических задач и выбирать подходы к их решению;
- проектировать и обучать модели машинного обучения, оценивать качество обученных моделей и их обобщающие способности;
- использовать на практике методы обучения и диагностики регрессионных моделей;
- использовать методы классификации данных для решения практических задач;
- использовать непараметрические методы машинного обучения для решения задач обучения с учителем;

- применять методы обучения ансамблей для повышения качества моделей машинного обучения;
- проводить кластерный анализ данных и применять методы кластеризации для решения практических задач.

Владеть:

- языком программирования Python;
- научными библиотеками языка Python (numpy, scipy, scikit-learn и др.);
- средствами разработки на языке Python (PyCharm, Jupyter Notebook, Google Colab).

1.4. Требования к уровню подготовки поступающего на обучение, необходимому для освоения программы

К освоению Программы в рамках проекта допускаются лица получающие высшее образование по очной (очно-заочной) форме, лица, освоившие основную профессиональную образовательную программу (далее – ОПОП ВО) бакалавриата – в объеме не менее первого курса (бакалавры 2-го курса), ОПОП ВО специалитета – не менее первого и второго курсов (специалисты 3-го курса), а также магистратуры, обучающиеся по ОПОП ВО, не отнесенным к ИТ-сфере.

1.5. Трудоемкость обучения

Нормативная трудоемкость обучения по данной программе - 324 часа, включая все виды аудиторной и внеаудиторной (самостоятельной) учебной работы слушателя.

1.6. Форма обучения

Форма обучения – очная с использованием дистанционных образовательных технологий.

При реализации настоящей программы используются следующие образовательные технологии:

- **Информационно-коммуникативные технологии (ИКТ)**

Программа реализуется дистанционно при помощи LMS, различных электронных ресурсов и интерактивных средств.

- **Модульная технология**

Обучение на программе предполагает модульное освоение дисциплин.

○ **Кейс-технология**

Кейсы и практические задачи, рассматриваемые в ходе обучения, обеспечиваются организацией, представители которой участвовали в разработке программы.

○ **Технология уровневой дифференциации**

Настоящая программа рассчитана на обучение слушателей разных уровней подготовки и предполагает возможность выполнения заданий разного уровня сложности.

○ **Смешанное обучение**

В ходе обучения слушателям предлагается посещать не только лекционные и семинарские занятия, но и изучать самостоятельно материалы, представленные в формате онлайн-курсов, вспомогательных видеозаписей и электронных пособий.

○ **Онлайн курсы**

Также для студентов, в процессе обучения доступны онлайн курсы НИЯУ МИФИ, размещенные на платформе «Открытое образование». Данные курсы могут быть использованы как дополнительный методический материал по различным темам в области информационных технологий. Помимо этого, данные курсы могут быть крайне полезны студентам, отнесенными к не ИТ направлениям. Рекомендуемый перечень онлайн курсов приведен ниже:

№	КУРС	АВТОРЫ	З.Е.	ССЫЛКА
1	Управление разработкой корпоративных информационных систем	Зыков С.В.	2	https://openedu.ru/course/mephi/mephi_007_urkis/?session=spring_2022
2	Применение механизмов операционных систем в разработке программного обеспечения		2	https://openedu.ru/course/mephi/mephi_imosrpo/?session=spring_2022
3	Основы интеллектуального анализа данных и машинного обучения	Запечников С.В.	2	https://openedu.ru/course/mephi/mephi_oiadmo/
4	Основы антикризисной разработки корпоративных информационных систем	Зыков С.В.	2	https://openedu.ru/course/mephi/mephi_arkis/?session=spring_2022

5	Обработка данных на языке Python	Киреев В.С.	3	https://openedu.ru/course/mephi/mephi_python/
6	Нейронные сети	Трофимов А.Г.	2	https://openedu.ru/course/mephi/mephi_ns/?session=spring_2022
7	Машинное обучение	Трофимов А.Г.	2	https://openedu.ru/course/mephi/mephi_mo/?session=spring_2022
8	Математические и инструментальные методы машинного обучения	Киреев В.С.	4	https://openedu.ru/course/mephi/mephi_011_machinelearning/?session=spring_2022
9	Биологически мотивированные когнитивные архитектуры (БИКА)	Самсонович А.В.	3	https://openedu.ru/course/mephi/mephi_bica/?session=spring_2022

2. СОДЕРЖАНИЕ ПРОГРАММЫ

2.1. Учебный план

Наименование дисциплин	Общая трудоемкость, час.	Всего, ауд. час.	Аудиторные занятия, час.			СРС, час.	Текущий контроль* (шт.)			Промежуточная аттестация	
			Лек.	Лаб.	Прак.		ЛР	Тест	ДЗ	Зачет	Экзамен
1	2	3	4	5	6	7	8	9	10	11	12
Модуль 1. Научное программирование в Python	46	32	16	16	0	14	2	0	0	+	
Модуль 2. Введение в машинное обучение	62	48	32	0	16	14	0	0	1	+	
Модуль 3. Машинное обучение. Часть 1	62	48	32	0	16	14	0	0	1	+	
Модуль 4. Машинное обучение. Часть 2	62	48	16	16	16	14	2	0	0	+	

Модуль 5. Обеспечение информационной безопасности	20	9	9	0	0	11	0	1	0	+	
Практика, подготовка и защита выпускной квалификационн ой работы	72	0	0	0	0	72	0	0	0		
Итого	324	185	105	32	48	139	4	1	2		
Итоговая аттестация	Выпускная аттестационная работа										
<i>* КП - курсовой проект, КР - курсовая работа, Тест – тестирование, ЛР – лабораторные работы</i>											

2.2. Дисциплинарное содержание программы

МОДУЛЬ 1. НАУЧНОЕ ПРОГРАММИРОВАНИЕ В PYTHON.

1.1. Основы языка программирования Python

Особенности языка Python, области применения Python, философия Python ("The Zen of Python"), среда исполнения Python, начало работы с Python, физические строки, логические строки и блоки кода Python, управление потоком команд.

Типизация в Python, основные типы данных: NoneType, логический, числовой, строковый типы, коллекции, инициализация переменных, изменяемые и неизменяемые типы, преобразование типов, строки и байтовые последовательности, форматирование строк, срезы (slices), интернирование строк.

Функции в Python, анонимные функции и лямбда-выражения, глобальные, нелокальные и локальные переменные, переопределение глобальных переменных, модули, подключение модулей, пакеты, создание пакетов, импорт функций и модулей из пакетов. IDE для научных вычислений на примере Jupyter Notebook. Система контроля версий git.

1.2. Коллекции, функции и объектно-ориентированное программирование в Python.

Списки (lists), создание списков и доступ к элементам списка, срезы (slices), функции и методы списков, псевдонимы (aliases) в Python, поверхностное и глубокое копирование, списки и строки, диапазоны (ranges), итерируемые классы и классы-итераторы.

Декораторы функций, передача параметров в функцию, call by sharing, функции с переменным числом аргументов, упаковка и распаковка последовательностей, функция zip, оператор *, оператор **, функция map.

Ввод данных из стандартного потока ввода, работа с файлами, итерирование строк файла, обработка исключений, конструкция try-except-else-finally, создание собственных исключений, менеджер контекста, конструкция with-as.

Классы и объекты, конструктор и деструктор класса, значения по умолчанию, переменные класса и переменные объекта, методы класса и методы объекта, статические методы, геттеры и сеттеры свойств, использование декораторов свойств, наследование классов, переопределение (overriding) и перегрузка (overloading) методов, перегрузка операторов, абстрактные классы.

1.3. Научные вычисления в Python.

Библиотека NumPy, массив NumPy (ndarray), представления массивов, создание массивов и представлений, копирование массивов, структурированные массивы, типы элементов массива, форма массива, стратегии хранения массивов в памяти, итерирование массивов, страйды массива, изменение формы и страйдов, индексирование и срезы (slices) массива, fancy indexing, функция where.

Операции над массивами, конкатенация, разбиение и дублирование массивов, арифметические операции с массивами, broadcasting, универсальные функции (ufunc), векторно-матричные операции, математические и статистические методы массивов, методы линейной алгебры, модуль linalg, матрицы NumPy, классы matrix и ndarray, генерация случайных массивов, модуль random, загрузка и сохранение массивов в файл.

Возможности библиотеки pandas работы с данными. Возможности библиотек scipy, scikit-learn для проведения научных вычислений. Понятие датафрейма (DataFrame) и серии (Series). Базовые операции со структурами данных pandas. Возможности библиотек scipy, scikit-learn для проведения научных вычислений.

1.4. Визуализация данных в Python.

Этапы визуализации данных. Разведочный анализ данных. Первичная обработка данных. Важность визуализации в анализе данных. Инструменты визуализации данных. Библиотеки языка Python для визуализации. Библиотека matplotlib. Архитектура matplotlib. Бэкенд-слой. Класс Artist. Использование объектов класса Artist для визуализации. Фигуры, субфигуры и оси. Контейнеры графических объектов. Жизненный цикл фигуры. Основные типы графиков в matplotlib. Визуализация на скриптовом уровне. Библиотека seaborn.

МОДУЛЬ 2. ВВЕДЕНИЕ В МАШИННОЕ ОБУЧЕНИЕ.

2.1. Основные парадигмы машинного обучения.

Машинное обучение и науки о данных. Data-driven и model-based подходы. История машинного обучения. Общая схема решения задач машинного обучения. Стандарт CRISP-DM. Типы задач машинного обучения. Примеры прикладных задач. Основные понятия и определения. Обзор основных парадигм машинного обучения.

2.2. Обучение с учителем: основные принципы.

Методы обучения с учителем. Обучающийся алгоритм. Индуктивный порог. Проблема переобучения и понятие обобщающей способности. Функция потерь. Теоретический и эмпирический риск. Принцип минимизации эмпирического риска. Смещение и дисперсия ошибки обучаемой модели. Разложение «bias-variance». Связь сложности модели, индуктивного порога и способности к обобщению. Оценивание точности обученных моделей. Методы кросс-валидации: hold-out, k-fold, LOOCV. Стратификация выборки при кросс-валидации, внутренняя кросс-валидация.

2.3. Обучение с учителем: регрессия.

Понятие статистической модели. Теоретические и статистические модели. Примеры. Регрессионные модели. Постановка задачи обучения регрессионной модели. Риск модели с квадратичной функцией потерь. Функция регрессии. Оптимальность регрессионных моделей. Задачи, виды и этапы регрессионного анализа. Метод наименьших квадратов. Матрица плана и система нормальных уравнений. МНК-оценки параметров простейшей линейной регрессии. Ошибки и остатки модели. Предположения регрессионного анализа. Свойства МНК-оценок: линейность, состоятельность, несмещенность, эффективность, нормальность. Теорема Гаусса-Маркова. Метод наименьших квадратов. Проекционная матрица. МНК-оценки параметров множественной линейной регрессии. Статистические характеристики МНК-оценок. Проблема гетероскедастичности. Статистические тесты на гетероскедастичность. Визуальный анализ гетероскедастичности остатков регрессионной модели. Подходы к устранению гетероскедастичности.

2.4. Диагностика регрессионных моделей.

Цели и виды диагностики регрессионных моделей. Проверка линейности. Графики остатков и частичной регрессии. Смещение оценок вследствие неучтенных переменных (omitted variable bias). Проверка гетероскедастичности, независимости и нормальности регрессионных остатков. Визуальная и количественная диагностика. Квантильные диаграммы. Понятие регрессионного выброса. Показатели аномальности наблюдения. Диаграмма box-and-whisker. Оценка влияния наблюдений на регрессионную модель. Связь

аномальности с леввериджами и стандартизованными регрессионными остатками. Расстояние Кука.

2.5. Бинарная классификация.

Постановка задачи классификации. Бинарная классификация. Функция потерь и эмпирический риск классификатора. Виды функций потерь. Статистический взгляд на задачу обучения. Принцип минимизации эмпирического риска и метод максимального правдоподобия. Регуляризованный эмпирический риск. Оценка точности бинарного классификатора. Матрица ошибок. Специфичность и чувствительность, точность и полнота. ROC-анализ и PR-анализ. Показатели ROC AUC и PR AUC.

2.6. Многоклассовая классификация.

Постановка задачи многоклассовой классификации. Подходы к многоклассовой классификации. Метод ЕСОС. Схемы кодирования. Методы декодирования. Оценка точности многоклассового классификатора. Матрица ошибок. Микро- и макро- показатели точности. Специфичность и чувствительность, точность и полнота. Микро- и макро- ROC-анализы и PR-анализы.

МОДУЛЬ 3. МАШИННОЕ ОБУЧЕНИЕ. ЧАСТЬ 1.

3.1. Байесовская классификация.

Порождающие и разделяющие модели классификации. Бинарная функция потерь. Байесовское решающее правило. Дискриминантные функции. Нормальный байесовский классификатор. Расстояние Махаланобиса. Наивный байесовский классификатор. Регуляризованный байесовский классификатор. Байесовская классификация бинарных и мультиномиальных признаков. Пример применения к задаче категоризации документов. Непараметрический байесовский классификатор. Наивная оценка плотности распределения. Ядерная оценка плотности распределения. Виды ядерных функций. Парзеновское окно. Показатель точности оценки плотности. Разложение «bias-variance» оценки. Подходы к выбору ширины парзеновского окна. Метод Сильвермана. Оценивание многомерных плотностей распределения. Многомерные ядерные функции.

3.2. Непараметрическое оценивание плотностей.

Особенности непараметрических методов машинного обучения. Оценивание плотности с помощью гистограммы частот. Наивное оценивание плотности. Понятие ядерной функции (парзеновского окна). Виды ядер. Разложение bias-variance в ядерном оценивании. Влияние ширины парзеновского окна на свойства восстановленной плотности. Подходы к

выбору ширины окна. Правило Сильвермана. Ядерное оценивание многомерных плотностей.

3.3. Логистическая регрессия.

Понятие функции связи (link function). Виды функций связи, их свойства. Логит и пробит-функции связи. Логистическая функция. Логистическая модель апостериорных вероятностей классов. Отношение вероятностей (odds ratio). Отношение вероятностей для нормального байесовского классификатора. Связь нормального байесовского классификатора и логистической регрессии. Постановка задачи обучения бинарной логистической регрессии. Критерий обучения и метод обучения. Логистическая функция потерь. Регуляризованная логистическая регрессия.

3.4. Непараметрическая регрессия.

Виды регрессионного анализа. Особенности непараметрического регрессионного анализа. Регрессограмма. Наивное оценивание функции регрессии. Ядерная регрессия Надарая-Уотсона. Виды ядер. KNN-регрессия. Линейное сглаживание. Связь матрицы линейного сглаживания и проекционной матрицы линейной регрессионной модели. Локальная линейная регрессия. Локальный метод наименьших квадратов. Использование взвешенного метода наименьших квадратов для локального линейного оценивания.

3.5. Деревья решений.

Понятие дерева решений. Регрессионные и классификационные деревья. Типы вершин. Обучение деревьев. Метрики неомогенности вершин. Метрика Джини. Энтропия распределения данных в вершинах. Критерий ветвления. Выбор решающего правила в узлах. Использование решающих деревьев для отбора признаков. Оценивание важности признаков. Переобучение деревьев решений. Критерий останковки обучения.

3.6. Обучение ансамблей.

Обучение ансамблей. Понятие слабой модели. Подходы к обучению ансамблей: бэггинг, бустинг, стеккинг. Бутстреппинг выборки. Бэггинг деревьев решений. Out-of-bag (OOB) ошибка. Использование OOB-ошибки для кросс-валидации деревьев. Случайный лес. Преимущества и недостатки бэггинга. Техники бустинга моделей машинного обучения. Принципы адаптивного бустинга (AdaBoost). Адаптивный бустинг моделей регрессии и классификации. Алгоритмы AdaBoost.R2 и AdaBoost.M1. Адаптивный бустинг деревьев решений. Пошаговое аддитивное обучение моделей машинного обучения. Постановка задачи обучения слабой модели. Виды функций потерь в аддитивном обучении.

Оптимальная слабая модель. Алгоритм градиентного бустинга обучения ансамблей. Инициализация алгоритма. Градиентный бустинг деревьев решений. Регуляризованный градиентный бустинг. Градиентный бустинг с шринкажем. Стохастический градиентный бустинг. Основные алгоритмы градиентного бустинга: XGBoost, LightGBM, CatBoost.

МОДУЛЬ 4. МАШИННОЕ ОБУЧЕНИЕ. ЧАСТЬ 2.

4.1. Сокращение размерности данных и метод главных компонентов.

Понятие машинного обучения без учителя. Виды и задачи обучения без учителя. Примеры практических задач. Постановка задачи сокращения размерности данных. Подходы к сокращению размерности. Метод главных компонентов. Факторные нагрузки и факторные очки. Выбор числа главных компонентов. Scree-диаграмма и метод локтя. Визуализация факторных нагрузок. Biplot. Анализ ошибок реконструкции. Статистика Хоттелинга. Метод главных компонентов и факторный анализ данных. Сингулярное разложение матрицы. Использование сингулярного разложения для эффективного выделения главных компонентов.

4.2. Кластерный анализ данных.

Кластерный анализ данных: цели, история развития, постановка задачи. Практические применения кластерного анализа. Подходы к кластеризации данных. Виды кластерного анализа. Общая схема решения задачи кластеризации. Метод К-средних. Инициализация метода. Алгоритм K-means++. Сходимость метода К-средних. Модификации метода. Недостатки метода. Агломеративные и дивизивные методы кластеризации. Агломеративная иерархическая кластеризация. Визуализация результатов кластеризации. Дендрограмма. Методы расчета расстояния между кластерами. Статистическое расстояние. Метод Уорда. Формула Ланса-Вильямса. Плотностные алгоритмы кластеризации на примере DBSCAN. Метод распространения близости (Affinity Propagation).

МОДУЛЬ 5. ОБЕСПЕЧЕНИЕ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ.

5.1 Введение в информационную безопасность.

Цели и задачи информационной безопасности. Подходы к обеспечению информационной безопасности.

5.2 Безопасность компьютерных сетей.

Симметричные и асимметричные шифры. Схемы электронной подписи. Инфраструктура открытых ключей. Современные средства защиты информации. Сетевые атаки.

5.3 Безопасность приложений.

Безопасность-веб-приложений. Атаки на приложения. Анализ защищенности и тестирование на проникновение.

5.4 Безопасность интеллектуальных систем.

Конфиденциальное машинное обучение. Доверенный искусственный интеллект.

3. УСЛОВИЯ РЕАЛИЗАЦИИ ПРОГРАММЫ

3.1. Материально-технические условия реализации программы

Наименование специализированных аудиторий, кабинетов, лабораторий	Вид занятий	Наименование оборудования, программного обеспечения
1	2	3
<i>Аудитория</i>	<i>Лекции, семинары</i>	<i>компьютер, Zoom, веб-браузер с доступом в Интернет</i>
<i>Лаборатория</i>	<i>Семинары, лабораторные работы</i>	<i>Компьютер, веб-браузер с доступом в Интернет</i>

3.2. Учебно-методическое обеспечение программы

I. Научное программирование в Python

1. Маккини Уэс «Python и анализ данных». – М.: ДМК Пресс, 2020.
2. Лутц М. Изучаем Python, том 1. – СПб: ООО "Диалектика". 2019.
3. Хеллман Д. Стандартная библиотека python 3. Справочник с примерами. – СПб: ООО "Диалектика". 2019.
4. G. van Rossum. Python Tutorial. Python Software Foundation. 2012.

II. Введение в машинное обучение

1. Бринк Х., Ричардс Д., Феверолф М.: Машинное обучение. – СПб: Издательский дом "Питер", 2017.
2. Плас Дж. Вандер. Python для сложных задач: наука о данных и машинное обучение. – СПб: Издательский дом "Питер". 2018.
3. Alpaydin, E. (2014). Introduction to machine learning. MIT press.
4. Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding machine learning: From theory to algorithms. Cambridge university press.

III. Машинное обучение. Часть 1

а) ОСНОВНАЯ ЛИТЕРАТУРА:

1. Бринк Х., Ричардс Д., Феверолф М.: Машинное обучение. – СПб: Издательский дом "Питер", 2017.

2. Плас Дж. Вандер. Python для сложных задач: наука о данных и машинное обучение. – СПб: Издательский дом "Питер". 2018.

3. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / пер. с англ. А. А. Слинкина. – М.: ДМК Пресс, 2015.

б) ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА:

1. Жерон О. Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем. Пер. с англ. – СПб.: ООО «Альфа-книга», 2018.

2. Мюллер А., Гвидо С. Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными. – М.: Вильямс, 2017.

3. Alpaydin, E. (2014). Introduction to machine learning. MIT press.

4. Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding machine learning: From theory to algorithms. Cambridge university press.

5. Raschka, S. (2015). Python machine learning. Packt Publishing Ltd.

IV. Машинное обучение. Часть 2

1. L. Rokach. Ensemble Learning: Pattern Classification Using Ensemble Methods. Wors Scientific Publishing, 2019.

2. A. Kumar, M. Jain. Ensemble learning for AI developers. Apress, 2020.

3. J. Brownlee. Ensemble learning algorithms with Python. Machine Learning Mastery, 2021.

4. T. Jo. Machine learning foundation. Springer nature, 2021.

V. Обеспечение информационной безопасности

1. Запечников С.В., Криптографические методы защиты информации. – М.: Юрайт, 2022.

3.3. Организация практической подготовки обучающихся

НИЯУ МИФИ сотрудничает с рядом IT-компаний в части проведения практической подготовки студентов на предприятии. Слушатели программ Цифровой кафедры в ходе обучения смогут пройти практику в одной из организаций-партнеров. Среди прочих НИЯУ МИФИ имеет соглашения о практической подготовке со следующими организациями:

1. ООО "Ростелеком информационные технологии"
2. АО "Атомстройэкспорт"
3. АО "Гринатом"
4. АО "Лаборатория Касперского"
5. АО "Русатом Автоматизированные системы управления"
6. АО "ТВЭЛ"
7. АО "Федеральный центр науки и высоких технологий "Специальное научно-производственное объединение "Элерон"
8. ООО "ВР Концепт"
9. ООО "Крипто-Про"
10. ООО "БПЦ Девелопмент"
11. ООО "Безопасная информационная зона"
12. ООО "Код Безопасности"
13. ООО "ЦБИ "Маском"
14. ПАО "Московская биржа"
15. . ФГУП "Всероссийский научно-исследовательский институт автоматики им. Н.Л. Духова"

4. ОЦЕНКА КАЧЕСТВА ОСВОЕНИЯ ПРОГРАММЫ

В рамках каждого учебного модуля программы в качестве оценочной шкалы используется 100 бальная система. Промежуточный контроль для каждого учебного модуля программы выставляется на основе суммарной оценки, полученной слушателями за выполненные лабораторные/домашние работы и тестирование (текущий контроль).

В ходе изучения модулей 2 и 3 слушателям выдается домашняя работа, результаты выполнения которой оформляются в виде презентации.

Максимальный балл за лабораторные/домашние работы по каждому модулю равен 50 баллам. Максимальный балл за тестовое задание в рамках каждого модуля программы равен 50 баллам. Промежуточный контроль в рамках каждого модуля считается пройденным, если слушатель набрал суммарно за все виды текущего контроля не менее 60 баллов. Таким образом, зачет в рамках каждого модуля программы выставляется по результатам прохождения промежуточного контроля.

Слушатели, успешно сдавшие промежуточный контроль по всем модулям программы, допускаются к итоговой аттестации. Итоговая аттестация по Программе проводится в форме защиты выпускной аттестационной работы.

Лицам, успешно освоившим Программу и прошедшим итоговую аттестацию в рамках

проекта «Цифровые кафедры», выдается документ о квалификации: диплом о профессиональной переподготовке.

При освоении ДПП ПП 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 Университета, выдается справка об обучении или о периоде обучения по образцу, самостоятельно устанавливаемому Университетом.

5. СОСТАВИТЕЛИ И ПРЕПОДАВАТЕЛИ ПРОГРАММЫ

ФИО	Образование	Должность, место работы	Опыт работы
<p>Трофимов Александр Геннадьевич</p> <p>(руководитель программы)</p>	<p>НИЯУ МИФИ, «Прикладная математика и информатика», 2005</p> <p>Кандидат технических наук</p>	<p>Доцент Института интеллектуальных кибернетических систем НИЯУ МИФИ</p>	<p>Стаж работы в IT-сфере или в отрасли цифровой экономики 14 лет</p>
			<p>Педагогический стаж 14 лет</p>
			<p>Перечень основных публикаций</p> <ol style="list-style-type: none"> 1. Cross-Modal Transfer Learning for Image and Sound // Studies in Computational Intelligence, 2022 Vol. 1008 SCI, Q4 pp. 238-245 2. A method of choosing a pre-trained convolutional neural network for transfer learning in image classification problems // Studies in Computational Intelligence, 2020 Vol. 856, Q4 pp. 263-270 3. Predictive Model for Calculating Abnormal Functioning Power Equipment // Studies in Systems, Decision and Control, 2020 Vol. 260, Q2 pp. 249-259 4. Модель раннего обнаружения аварийных ситуаций на оборудовании электростанций на основе методов машинного обучения // Теплоэнергетика, 2019г. Вып. 3 Стр. 49-56 5. Прогнозирование аварии на энергетическом оборудовании // XXVIII Международная научно-техническая конференция Современные технологии в задачах управления, автоматики и обработки информации, 2019г. Стр. 190 6. Программно-аналитический модуль предиктивной аналитики // Математические методы в технике и технологиях - ММТТ, 2019г. Т. 10 Стр. 122-126 7. Количество цитирований Scopus: 8 Model for Early Detection of Emergency Conditions in Power Plant Equipment Based on Machine Learning Methods // Thermal Engineering, 2019 Vol. 66, No. 3, Q3 pp. 189-195 8. Abnormal operation detection in heat power plant using ensemble of binary classifiers // Studies in Computational Intelligence, 2019 Vol. 799, Q3 pp. 227-233

			<p>9. An Approach to Use Convolutional Neural Network Features in Eye-Brain-Computer-Interface // Studies in Computational Intelligence, 2018 Vol. 736, Q3 pp. 132-137</p> <p>10. A greedy feature selection algorithm for brain-computer interface classification committees // Procedia Computer Science, 2018 Vol. 123, Q2 pp. 488-493</p>
Киреев Василий Сергеевич	<p>НИЯУ МИФИ, Факультет кибернетики, квалификация «математик, системный программист» по специальности «Прикладная математика», 2005</p> <p>Кандидат технических наук</p>	<p>Доцент Института интеллектуальных кибернетических систем НИЯУ МИФИ</p>	<p>Стаж работы в IT-сфере или в отрасли цифровой экономики 14 лет</p> <p>1. Разработка экономико-математических моделей, методов и инструментария для создания и реализации программ лояльности российских международных мегапроектов; 2. Разработка технологической подосновы развития инновационной социальной прежде всего жилищной инфраструктуры на базе смарт-автоматов и принципах Интернета вещей; 3. Принципы создания алгоритмического обеспечения для многомерной классификации на примере анализа научных направлений; 4. Разработка метода и программно-технических решений повышения пертинентности информации в научных и аналитических рекомендательных системах.</p> <p>Педагогический стаж 18 лет</p>
Комаров Тимофей Ильич	<p>НИЯУ МИФИ, Специальность «Вычислительные машины, комплексы, системы и сети», 2014</p>	<p>Ведущий специалист, АО "Рамэк-ВС"</p>	<p>Стаж работы в IT-сфере или в отрасли цифровой экономики Комаров Т.И. является специалистом компании-системного интегратора полного цикла, которая предлагает своим заказчикам весь спектр услуг от консалтинга и проектирования информационных систем до реализации проектов, поставки техники и сервисного сопровождения. Компания является разработчиком уникальных вычислительных машин "РАМЭК"</p>

Коллектив разработчиков настоящей программы имеет опыт участия в реализации следующих проектов IT-профиля:

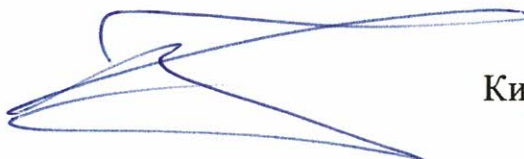
1. Создание и развитие Международного научного-методического центра НИЯУ МИФИ в рамках реализации федерального проекта «Кадры для цифровой экономики» национальной программы «Цифровая экономика в Российской Федерации»;
2. Создание Диджитал-Центра НИЯУ МИФИ в рамках программы развития НИЯУ МИФИ при поддержке ГК «Росатом»;
3. Цифровая трансформация НИЯУ МИФИ.

И.о. руководителя МНМЦ



Когос К.Г.

Декан ФПКПК



Киреев С.В.