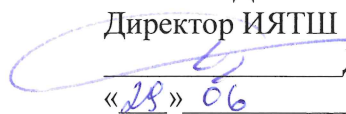


МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

УТВЕРЖДАЮ

Директор ИЯТШ

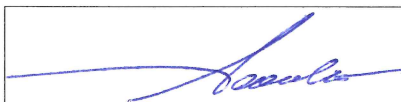
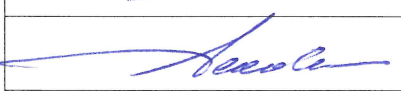
 Долматов О. Ю.
«28» 06 2020 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

ПРИЕМ 2020 г.

ФОРМА ОБУЧЕНИЯ очная

Обработка больших объемов данных			
Направление подготовки/специальность	03.04.02 Физика		
Образовательная программа (направленность (профиль))	Физика конденсированного состояния		
Специализация			
Уровень образования	высшее образование - магистратура		
Курс	2	семестр	3
Трудоемкость в кредитах (зачетных единицах)	3		
Виды учебной деятельности	Временной ресурс		
Контактная (аудиторная) работа, ч	Лекции	8	
	Практические занятия	16	
	ВСЕГО	24	
Самостоятельная работа, ч		84	
ИТОГО, ч		108	

Вид промежуточной аттестации	Диф. зачет	Обеспечивающее подразделение	ОЭФ
Заведующий кафедрой – руководитель отделения на правах кафедры			Лидер А. М.
Руководитель ООП			Лидер А. М.
Преподаватели			Семенов М. Е.

2020 г.

1. Цели освоения дисциплины

Целями освоения дисциплины является формирование у обучающихся определенного ООП (п. 5. Общей характеристики ООП) состава компетенций для подготовки к профессиональной деятельности.

Код компетенции	Наименование компетенции	Составляющие результатов освоения (дескрипторы компетенций)	
		Код	Наименование
ПК(У)-1	Способность самостоятельно ставить конкретные задачи научных исследований в области физики и решать их с помощью современной аппаратуры и информационных технологий с использованием новейшего отечественного и зарубежного опыта	ПК(У)-1.B1	Владеет опытом использования современных компьютерных сетей, баз данных, программных продуктов и ресурсов Internet для решения конкретных задач научных исследований в области физики
ДПК(У)-1	Способность планировать и проводить фундаментальные исследования в проектах в области ядерно-физических исследований, взаимодействия излучения с веществом, модернизации современных и создания методов изучения механических, электрических, магнитных, тепловых свойств твердых тел	ДПК(У)-1.B4	Владеет опытом работы с результатами испытаний и обработки экспериментальных результатов методами статистической и математической обработки

2. Место дисциплины (модуля) в структуре ООП

Дисциплина относится к вариативной части Блока 1 учебного плана образовательной программы.

3. Планируемые результаты обучения по дисциплине

После успешного освоения дисциплины будут сформированы результаты обучения:

Планируемые результаты обучения по дисциплине		Компетенция
Код	Наименование	
РД-1	Выполнять исследования процессов создания, накопления и обработки информации, включая анализ и создание моделей данных и знаний, языков их описания и манипулирования.	ПК(У)-1
РД-2	Владеть методами исследования и обработки данных и их применению в самостоятельной научно-исследовательской деятельности в области профессиональной деятельности.	ДПК(У)-1

Оценочные мероприятия текущего контроля и промежуточной аттестации представлены в календарном рейтинг-плане дисциплины.

4. Структура и содержание дисциплины

Основные виды учебной деятельности

Разделы дисциплины	Формируемый результат обучения по дисциплине	Виды учебной деятельности	Объем времени, ч.
Раздел 1. Стратегии работы с большими объемами данных	РД-1	Лекции	4
		Практические занятия	4
		Самостоятельная работа	20
Раздел 2. Обработка больших	РД-2	Лекции	4

объемов данных		Практические занятия	12
		Самостоятельная работа	64

Содержание разделов дисциплины:

Раздел 1. Стратегии работы с большими объемами данных

Основные типы данных. Основные структуры данных. Структурированный язык запросов. Широкий и длинный формат данных. Предпроцессинг. Манипуляция. Преобразование форматов. Чтение/запись, кэширование. Функции чтения/записи данных из/в текстовый файл. Справочная система R/Python. Стратегии работы с большими массивами данных. Кэширование вычислений.

Темы лекций:

1. Стратегии работы с большими массивами данных
2. Подготовка исходных данных к обработке. Типы данных. Генерация данных

Темы практических занятий:

1. Языки и программные среды для обработки данных.
2. Структурированный язык запросов.

Раздел 2. Обработка данных

Числовые и текстовые данные. Отсутствующие данные. Ошибки в данных. Выбросы в данных. Дублирующие наблюдения (строки). Мультиколлинеарность. Цифровизация данных. Методы фильтрации, обертки, смешанные методы. Предсказание значений. График boxplot. Предиктивные модели. Дискретная, непрерывная, категориальная переменная. Шкалы. Машинное обучение с учителем, без учителя. Переобучение. Метод опорных векторов. Метод k-ближайших соседей. Снижение размерности. Метод главных компонент. Выборочные характеристики положения и их использование для заполнения пропусков. Графический способ определения выбросов. Принципы разделения выборки на обучающую и тестовую. Методы преобразования непрерывных данных в категориальные. Классификация и регрессия. Линейные, нелинейные модели. Кластерный анализ. Дендограммы. Метод главных компонент. Ординационные диаграммы.

Темы лекций:

3. Выбор признаков (Feature Selection). Выбор экземпляров (Instance Selection). Дискретизация для классификации (Discretization).
4. Базовые алгоритмы интеллектуального анализа данных (Data mining).

Темы практических занятий:

3. Статистические методы заполнения пропусков и определения выбросов.
4. Графический разведочный анализ данных.
5. Определение важности признаков.
6. Классификация данных. Обучающая и тестовая выборки.
7. Преобразование и дискретизация данных.
8. Методы ординации многомерных объектов.

5. Организация самостоятельной работы студентов

Самостоятельная работа студентов при изучении дисциплины (модуля) предусмотрена в следующих видах и формах:

- Работа с лекционным материалом, поиск и обзор литературы и электронных источников информации по индивидуально заданной проблеме курса;
- Изучение тем, вынесенных на самостоятельную проработку;

- Поиск, анализ, структурирование и презентация информации;
- Перевод текстов с иностранных языков;
- Выполнение домашних заданий;
- Программные расчеты;
- Подготовка к практическим занятиям;
- Подготовка к оценивающим мероприятиям.

6. Учебно-методическое и информационное обеспечение дисциплины

6.1. Учебно-методическое обеспечение

1. Кабаков Р.И. R в действии. Анализ и визуализация данных в программе R. Москва: ДМК Пресс, 2014. — 588 с. // Лань : электронно-библиотечная система. — Схема доступа: <https://e.lanbook.com/book/58703>
2. Буховец, А. Г. Алгоритмы вычислительной статистики в системе R : учебное пособие / А. Г. Буховец, П. В. Москалев. — 2-е изд., перераб. и доп. — Санкт-Петербург : Лань, 2015. — 160 с. // Лань : электронно-библиотечная система. — Схема доступа: <https://e.lanbook.com/book/68459>
3. Введение в статистическое обучение с примерами на языке R / Г. Джеймс, Д. Уиттон, Т. Хасты, Р. Тибширани ; перевод с английского С. Э. Мاستицкого. — Москва : ДМК Пресс, 2017. — 456 с. // Лань : электронно-библиотечная система. — Схема доступа: <https://e.lanbook.com/book/93580>
4. Москвитин, А. А.. Данные, информация, знания: методология, теория, технологии: монография [Электронный ресурс] / Москвитин А. А.. — Санкт-Петербург: Лань, 2019. — 236 с. Схема доступа: <https://e.lanbook.com/book/113937>

Дополнительная литература

1. X. Wu et al. (2008) Top 10 algorithms in data mining. Knowl. Inf. Syst. 14. 1–37. Схема доступа: <http://www.cs.umd.edu/~samir/498/10Algorithms-08.pdf>
2. Лесковец, Юре. Анализ больших наборов данных : пер. с англ. / Ю. Лесковец, А. Раджараман, Дж. Ульман. — Москва: ДМК Пресс, 2016. — 498 с. Схема доступа: <https://e.lanbook.com/book/93571>

6.2. Информационное и программное обеспечение

Internet-ресурсы (в т.ч. в среде LMS MOODLE и др. образовательные и библиотечные ресурсы):

Персональный сайт преподавателя М.Е. Семенова - <http://portal.tpu.ru/SHARED/s/SME/work>
 Электронно-библиотечная система «Лань» - <https://e.lanbook.com/>
 Сообщество специалистов в области Data Science - <http://ods.ai>
 Центр Инженерных Технологий и Моделирования - <http://www.exponenta.ru>

Лицензионное программное обеспечение (в соответствии с **Перечнем лицензионного программного обеспечения ТПУ**):

1. Acrobat Reader DC ;
2. PTC Mathcad Prime 6 Academic Floating;
3. Mozilla Firefox ESR;
4. WinDjView;
5. Adobe Flash Player;

6. Chrome;
7. Microsoft Office 2007 Standard Russian Academic

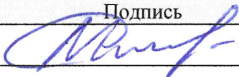
7. Особые требования к материально-техническому обеспечению дисциплины

В учебном процессе используется следующее лабораторное оборудование для практических и лабораторных занятий:

№	Наименование специальных помещений	Наименование оборудования
1.	Аудитория для проведения учебных занятий всех типов, курсового проектирования, консультаций, текущего контроля и промежуточной аттестации (компьютерный класс) 634028, Томская область, г. Томск, Ленина проспект, д. 2 427А	Доска аудиторная настенная - 1 шт.; Шкаф для одежды - 1 шт.; Шкаф для документов - 1 шт.; Комплект учебной мебели на 10 посадочных мест; Компьютер - 11 шт.
2.	Аудитория для проведения учебных занятий всех типов, курсового проектирования, консультаций, текущего контроля и промежуточной аттестации 634028, Томская область, г. Томск, Ленина проспект, д. 2 422	Доска аудиторная настенная - 1 шт.; Комплект учебной мебели на 72 посадочных мест; Компьютер - 1 шт.; Проектор - 1 шт.

Рабочая программа составлена на основе Общей характеристики образовательной программы по направлению 03.04.02 Физика / Физика конденсированного состояния (приема 2020 г., очная форма обучения).

Разработчик(и):

Должность	Подпись	ФИО
доцент		Семенов М.Е.

Программа одобрена на заседании выпускающего отделения экспериментальной физики ИЯТШ

(протокол № 2 от « 04 » июня 2020 г.).

Заведующий кафедрой – руководитель отделения
(на правах кафедры):

д. т. н.

 /Лидер А. М./
подпись