

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



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
ПРИЕМ 2019 г.
ФОРМА ОБУЧЕНИЯ очная

Специальные разделы больших данных

Направление подготовки/ специальность	09.04.04 Программная инженерия		
Образовательная программа (направленность (профиль))	Big Data Solutions / Технологии больших данных (Язык обучения: Английский)		
Специализация	Big Data Solutions / Технологии больших данных		
Уровень образования	высшее образование - магистратура		
Курс	2	семестр	3
Трудоемкость в кредитах (зачетных единицах)	6		
Виды учебной деятельности	Временной ресурс		
Контактная (аудиторная) работа, ч	Лекции	24	
	Практические занятия	0	
	Лабораторные занятия	24	
	ВСЕГО	48	
Самостоятельная работа, ч		168	
ИТОГО, ч		216	

Вид промежуточной аттестации	Экзамен	Обеспечивающее подразделение	ОИТ ИШИТР
Заведующий кафедрой- руководитель ОИТ на правах кафедры			Шерстнев В.С.
Руководитель ООП			Савельев А.О..
Преподаватель			Губин Е.И..

2020г.

1. Цели освоения дисциплины

Целями освоения дисциплины является формирование у обучающихся определенного ООП (п. 5.4 Общей характеристики ООП) состава компетенций для подготовки к профессиональной деятельности.

Код компетенции	Наименование компетенции	Индикаторы достижения компетенций		Составляющие результатов освоения (дескрипторы компетенций)	
		Код индикатора	Наименование индикатора достижения	Код	Наименование
ПК(У)-2	Способен разрабатывать системы управления базами данных	И.ПК (У)-2.1	Разрабатывает компоненты системы управления базами данных	ПК(У)-2.1В1	Владеет навыками разработки структуры управления базами данных в целом и ее отдельных компонентов
				ПК(У)-2.1У1	Умеет идентифицировать класс разрабатываемой СУБД в зависимости от выполняемых ею задач и аппаратных средств, определенных в техническом задании на разработку систему СУБД
				ПК(У)-2.131	Знает теорию БД
				ПК(У)-2.132	Знает основные структуры и модели данных
				ПК(У)-2.133	Знает методы обработки данных
ОПК(У)-1	Способен самостоятельно приобретать, развивать и применять математические, естественно-научные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте	И.ОПК (У)-1.1	Применяет математические, естественнонаучные и социально-экономические методы в профессиональной деятельности	ОПК(У)-1.1В1	Владеет опытом применения математических методов в своей профессиональной сфере
				ОПК(У)-1.1У1	Умеет использовать математические методы и алгоритмы для решения прикладных задач в различных областях практических приложений
				ОПК(У)-1.131	Имеет математические, естественно-научные, социально-экономические и профессиональные знания
ОПК(У)- 4	Способен применять на практике новые научные принципы и методы исследований	И.ОПК (У)-4.2	Осуществляет методологическое обоснование научного исследования; планирует и проводит научные исследования	ОПК(У)-4.2В1	Владеет методами проведения исследований для решения практических задач профессиональной деятельности
				ОПК(У)-4.2У1	Умеет планировать и контролировать научно-исследовательские и проектные работы, проводимые в период производственной практики
				ОПК(У)-4.231	Знает методы анализа и оценки уровня новых принципов и методы исследований при решении исследовательских и проектных задач

2. Место дисциплины (модуля) в структуре ООП

Дисциплина относится к Вариативной части Междисциплинарного профессионального модуля учебного плана образовательной программы.

3. Планируемые результаты обучения по дисциплине

После успешного освоения дисциплины будут сформированы результаты обучения:

Планируемые результаты обучения по дисциплине		Индикатор достижения компетенции
Код	Наименование	
РД1	Осуществляет поиск, отбор и систематизацию информации	И.ПК (У)-2.1
РД2	Определения альтернативные варианты решений в проблемной ситуации	И.ОПК (У)-1.1
РД 3	Осуществлять выбор методов анализа данных в зависимости от природы данных, объема и качества	И.ОПК (У)-4.2
РД 4	Выполнять обработку и анализ данных, полученных при теоретических и экспериментальных исследованиях	И.ОПК (У)-4.2

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. Постановка бизнес-задачи. Целевая функция. Зависимые переменные и их тип.	РД1	Лекции	6
		Практические занятия	0
		Лабораторные занятия	6
		Самостоятельная работа	42
Раздел 2. SAS Base –как базовый инструмент подготовки исходных данных. SAS Data Mining	РД2	Лекции	6
		Практические занятия	0
		Лабораторные занятия	6
		Самостоятельная работа	42
Раздел 3. Методология подготовка исходных данных в SAS. Тренировочная, валидационная и тестовая выборки	РД3	Лекции	6
		Практические занятия	0
		Лабораторные занятия	6
		Самостоятельная работа	42
Раздел 4. Предиктивные модели в SAS. Лог. регрессии, деревья решений, нейронные сети. Решение практических задач	РД4	Лекции	6
		Практические занятия	0
		Лабораторные занятия	6
		Самостоятельная работа	42

Содержание разделов дисциплины:

Раздел 1. Постановка бизнес-задачи. Целевая функция. Зависимые переменные.

Введение, основные понятия бизнес-целей. Формулировка целевой функции и зависимых переменных.

Темы лекций:

- 1. Введение, основные понятия и цели бизнес проекта. (2 ч.)
- 2. Формулировка целевой функции. (2ч.)
- 3. Выбор и предварительный анализ входных переменных (2 ч.)

Темы лабораторных занятий:

1. Первичный анализ исходных данных (4 ч.)
2. Форматы целевой функции и зависимых переменных (2 ч.)

Раздел 2. SAS Base –как базовый инструмент подготовки исходных данных. SAS Data Mining

Предварительная обработка исходных данных. Методы описательной статистики. Анализ исходных данных на основе использования методов визуализации. Использование методологии SAS Data Mining и SEMMA при работе с данными. Форматы данных.

Темы лекций:

1. Основные методы описательной статистики в SAS. (2 ч.)
2. Форматы исходных данных в SAS: (2 ч.)
3. Определение целевой функции и объясняющих переменных (2ч.)

Темы лабораторных занятий:

1. Описательная статистика входных переменных (N, Min, Max, Div, Mean, Freq) (2 ч.)
2. Форматы входных переменных, включая числовые и текстовые (2 ч.)
3. Использование SAS Data Mining для подготовки исходных данных (2 ч.)

Раздел 3. Методология подготовка исходных данных в SAS. Тренировочная валидационная и тестовые выборки

Проверка входных данных на ошибки (описки), пропущенные значения, выбросы, дублирующие строки и мультиколлинеарность. Замена (удаление) не корректных данных в соответствие с принятой моделью.

Понятие тренировочной, тестовой и валидационной выборок. Даются основные принципы формирования тренировочной, тестовой и валидационной выборок.

Темы лекций:

1. Выявление ошибочных и пропущенных данных в SAS. (2 ч.)
2. Выявление мультиколлинеарности, дублирующих строк и выбросов в SAS (2 ч.)
3. Разбиение исходных данных на тренировочную, валидационную и тестовую выборки в SAS (2 ч.)

Темы лабораторных занятий:

1. Определение ошибочных и пропущенных данных (2 ч.)
2. Выявление мультиколлинеарности, дублирующих строк и выбросов (2 ч.)
3. Разбиение исходных данных на тренировочную, валидационную и тестовую выборки (2 ч.)

Раздел 4. Предиктивные модели в SAS. Лог. регрессии, деревья решений, нейронные сети. Решение практических задач.

Обзор и решение предиктивных задач анализа данных из различных прикладных областей, включая финансы и маркетинг.

Темы лекций:

1. Концепция предиктивных задач (2 ч.).
2. Логистические модели и их сильные и слабые стороны. SAS коды для практических задач (2 ч.)
3. Обзор решения задач анализа данных (примеры решений реальных задач в SAS) (2 ч.)

Темы лабораторных занятий:

1. Подготовка данных к анализу, выбор методов анализа (6 ч.).

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;
- Работа в электронном курсе (изучение теоретического материала, выполнение индивидуальных заданий и контролирующих мероприятий и др.);
- Изучение тем, вынесенных на самостоятельную проработку;
- Поиск, анализ, структурирование и презентация информации;
- Подготовка к лабораторным работам, к практическим и семинарским занятиям;
- Работа над междисциплинарным проектом;
- Исследовательская работа и участие в научных студенческих конференциях, семинарах и олимпиадах;
- Анализ научных публикаций по заранее определенной преподавателем теме;
- Подготовка к оценивающим мероприятиям.

6. Учебно-методическое и информационное обеспечение дисциплины

6.1. Учебно-методическое обеспечение

Основная литература:

1. Пантелеев, Е. Р. Методы научных исследований в программной инженерии: учебное пособие для вузов / Е. Р. Пантелеев. — 2-е изд., стер. — Санкт-Петербург: Лань, 2021. — 136 с. — ISBN 978-5-8114-6781-5. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/152439> (дата обращения: 18.11.2020).
2. Пантелеев, Е. Р. Методы научных исследований в программной инженерии: учебное пособие / Е. Р. Пантелеев. — Санкт-Петербург: Лань, 2018. — 136 с. — ISBN 978-5-8114-3220-2. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/110936> (дата обращения: 18.11.2020).

Дополнительная литература:

1. Методы и модели исследования сложных систем и обработки больших данных: монография / И. Ю. Парамонов, В. А. Смагин, Н. Е. Косых, А. Д. Хомоненко ; под редакцией В. А. Смагина и А. Д. Хомоненко. — Санкт-Петербург: Лань, 2020. — 236 с. — ISBN 978-5-8114-4006-1. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/126938> (дата обращения: 19.09.2020).
2. Форман, Д. Много цифр: Анализ больших данных при помощи Excel / Д. Форман ; перевод А. Соколовой. — Москва: Альпина Паблишер, 2016. — 461 с. — ISBN 978-5-9614-5032-3. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/87871> (дата обращения: 19.09.2020).

6.2. Информационное и программное обеспечение

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

1. Персональный сайт Губина Е.И.
<https://portal.tpu.ru/SHARED/g/GUBINE/academics/Tal>

2. SAS Academic Programs

https://www.sas.com/ru_ru/learn/academic-programs.html

3. Продукты, технологии и решения SAS

https://www.sas.com/ru_ru/software/all-products.html/

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

1. SAS лицензионное программное обеспечение для анализа данных
2. СУБД Microsoft SQL Server 2017 Enterprise
3. Microsoft SQL Server Management Studio 17.9.1 Freeware

7. Особые требования к материально-техническому обеспечению дисциплины

В учебном процессе используется следующее:

№	Наименование специальных помещений	Наименование оборудования
1.	Аудитория для проведения учебных занятий всех типов, курсового проектирования, консультаций, текущего контроля и промежуточной аттестации (компьютерный класс) Компьютер - 15 шт. 634034, Томская область, г. Томск, Советская улица, 84/3 212	Доска аудиторная настенная - 1 шт.; Шкаф для документов - 1 шт.; Комплект учебной мебели на 10 посадочных мест; Комплект (стол, кресло) - 4 шт.; Компьютер - 15 шт.
	Аудитория для проведения учебных занятий всех типов, курсового проектирования, консультаций, текущего контроля и промежуточной аттестации (лекционная аудитория) 634034 Томская область, Томск, Советская улица, д. 84/3, учебный корпус «Кибернетический центр», аудитория 313.	Комплект оборудования для проведения лекционных занятий: ПК, 1 шт., Проектор, 1 шт.

Рабочая программа составлена на основе Общей характеристики образовательной программы по направлению 09.04.04 Программная инженерия/ профиль Big Data Solutions / Технологии больших данных /специализации Big Data Solutions/Технологии больших данных (приема 2019 г., очная форма обучения).

Разработчик(и):

Должность	Подпись	ФИО
Доцент ОИТ ИШИТР		Губин Е.И.

Программа одобрена на заседании Отделения информационных технологий ИШИТР (протокол №13 от 19.06.2019 г.).

Заведующий кафедрой –
руководитель ОИТ на правах кафедры


_____ /Шерстнев В.С.
подпись

Лист изменений рабочей программы дисциплины:

Учебный год	Содержание /изменение	Обсуждено на заседании ИШИТР (протокол)
2020/2021	1. Обновлено программное обеспечение 2 Обновлен список литературы, в том числе ссылок ЭБС	от 01.09.2020г. № 19