МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение высшего образования

«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ

ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

УТВЕРЖДАЮ

Директор ИШИТР \_\_\_\_\_\_\_\_\_\_\_ / Демин А.Ю. «\_26\_\_»\_ \_\_\_\_06\_\_\_\_\_\_2021 г.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

**ПРИЕМ 2023 г.**

**ФОРМА ОБУЧЕНИЯ очная**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Введение в большие данные** | | | | | | |
|  |  | | | | | |
| 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 | **1** | семестр | | **1** | | | |
| Трудоемкость в кредитах (зачетных единицах) | **3** | | | | | | |
| Виды учебной деятельности | Временной ресурс | | | | | | |
| Контактная (аудиторная) работа, ч | Лекции | | | | **24** | |
| Практические занятия | | | | **24** | |
| Лабораторные занятия | | | | **0** | |
| ВСЕГО | | | | **48** | |
| Самостоятельная работа, ч | | | | | **60** | |
| ИТОГО, ч | | | | | **108** | |
|  |  | | | | | |
|  |  | | | | | |
| Вид промежуточной аттестации | **Зачет** | | Обеспечивающее подразделение | | | **ОИТ ИШИТР** |
|  |  | | | | | |
| Заведующий кафедрой- руководитель ОИТ на правах кафедры |  | | | | | Шерстнев В.С. |
| Руководитель ООП |  | | | | | Губин Е.И. |
| Преподаватель |  | | | | | Губин Е.И.. |

2023г.

# 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 компетенции)** | |
| --- | --- | --- | --- | --- | --- |
| **Код индикатора** | **Наименование**  **индикатора достижения** | **Код** | **Наименование** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ПК(ОУ)-5 | Способен разрабатывать и модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем | И.ОПК (У)-5.1 | Применяет знания современного программного обеспечения информационных и автоматизированных систем для решения профессиональных задач | ОПК(У)-5.1В1 | Владеет средствами проектирования баз данных |
| ОПК(У)-5.1У1 | Умеет проектировать и разрабатывать бизнес-логику БД с использованием современных инструментальных сред |
| ОПК(У)-5.1З1 | Знает алгоритмы работы бизнес-логики баз данных и систем управления базами данных |
| И.ОПК (У)-5.2 | Осуществляет разработку и модернизацию программного обеспечения информационных и автоматизированных систем | ОПК(У)- 5.2В1 | Владеет опытом разработки и тестирования программного обеспечения |
| ОПК(У)- 5.2У1 | Умеет применить методы и способы эффективного управления разработкой программных средств и проектов |
| ОПК(У)- 5.2З1 | Знает методы и способы эффективного управления разработкой программных средств и проектов, алгоритмы оптимизации/профилирования запросов |
| ОПК(У)-6 | Способен использовать методы и средства системной инженерии в области получения, передачи, хранения, переработки и предоставления информации посредством информационных технологий | И.ОПК (У)-6.2 | Анализирует техническое задание, разрабатывает и оптимизирует программный код для решения задач обработки информации и автоматизированного проектирования | ОПК(У)- 6.2В1 | Владеет опытом создания стратегии проектирования и критериев эффективности новых методов проектирования и разработки программных систем |
| ОПК(У)- 6.2У1 | Умеет организовывать взаимодействие коллективов разработчика и заказчика при внедрении и сопровождении (модернизации и интеграции) программных систем |
| ОПК(У)- 6.2З1 | Знает способы проектирования компонентов информационных систем |

# 2. Место дисциплины (модуля) в структуре ООП

Дисциплина относится к **Базовой части Общепрофессионального 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.ОПК (У)-5.1 |
| РД2 | Определения альтернативные варианты решений в проблемной ситуации | И.ОПК (У)-5.2 |
| РД 3 | Осуществлять выбор методов анализа данных в зависимости от природы данных, объема и качества | И.ОПК (У)-6.2 |

# 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. Понятие Big Data. Основные источники Big Data и их приложения** | РД1 | Лекции | **6** |
| Практические занятия | **6** |
| Лабораторные занятия | **0** |
| Самостоятельная работа | **15** |
| **Раздел 2. Описательная статистика исходных данных. Форматы. Целевая функция и объясняющие переменные** | РД2 | Лекции | **6** |
| Практические занятия | **6** |
| Лабораторные занятия | **0** |
| Самостоятельная работа | **15** |
| **Раздел 3. Подготовка исходных данных для прогнозного анализа. Тренировочная, валидационная и тестовая выборки** | РД2 | Лекции | **6** |
| Практические занятия | **6** |
| Лабораторные занятия | **0** |
| Самостоятельная работа | **15** |
| **Раздел 4. Предиктивные модели. Решение практических задач** | РД3 | Лекции | **6** |
| Практические занятия | **6** |
| Лабораторные занятия | **0** |
| Самостоятельная работа | **15** |

Содержание разделов дисциплины:

**Раздел 1. Понятие Big Data. Основные источники Big Data и приложения**

Введение, основные понятия больших данных. Кто работает с большими данными. Источники больших данных. Использование методологии Data Mining и SEMMA при работе с большими данными. Форматы данных.

**Темы лекций:**

1. Введение, основные понятия больших данных. (2 ч.)
2. Кто использует большие данные. (2ч.)
3. Использование методологии Data Mining и SEMMA (2 ч.)

**Темы практических занятий:**

1. Первичный анализ больших данных (4 ч.)
2. Форматы исходных данных (2 ч.)

**Раздел 2. Описательная статистика исходных данных.**

Типы данных. Предварительная обработка данных. Методы описательной статистики. Анализ исходных данных на основе использования методов визуализации.

Определение бизнес целей. Объясняющие переменные и целевая функция. Основные методы статистической обработки данных: Проверка характера распределения данных. Проверка распределения на «нормальность».

**Темы лекций:**

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

**Темы практических 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. Связь целевой функции и объясняющих переменных (2 ч.)

**Раздел 3. Подготовка исходных данных для прогнозного анализа**

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

Разбиение исходных данных на тренировочную, валидационную и тестовую выборки.

**Темы лекций:**

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

**Темы практических занятий:**

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

**Раздел 4. *Предиктивные модели***

Обзор и решение предиктивных задач анализа данных из различных прикладных областей.

**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) (2 ч.)

**Темы практических 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. Юре, Л. Анализ больших наборов данных / Л. Юре, Р. Ананд, Д. У. Джеффри; перевод с английского А. А. Слинкин. — Москва: ДМК Пресс, 2016. — 498 с. — ISBN 978-5-97060-190-7. — Текст: электронный // Лань: электронно-библиотечная система. — URL: https://e.lanbook.com/book/93571 (дата обращения: 19.09.2020).
2. MySQL 8 для больших данных / Ш. Чаллавала, Д. Лакхатария, Ч. Мехта, К. Патель; перевод с английского А. В. Логунова. — Москва: ДМК Пресс, 2018. — 226 с. — ISBN 978-5-97060-653-7. — Текст: электронный // Лань: электронно-библиотечная система. — URL: https://e.lanbook.com/book/131684 (дата обращения: 19.09.2020).

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. Смагинаи 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. М. Флёнов. Transact-SQL В подлиннике.

<http://www.flenov.info/books/read/transact-sql>

1. Справочник по Transact-SQL

<https://docs.microsoft.com/ru-ru/sql/t-sql/language-reference?view=sql-server-2017>

1. Brent Ozar Unlimited. T-SQL Level Up.

<https://www.brentozar.com/training/t-sql-level/>

1. <https://portal.tpu.ru/SHARED/g/GUBINE/academics>

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 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) 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. |
| 2 | 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/профильBig Data Solutions **/** Технологии больших данных /специализации Big Data Solutions**/**Технологии больших данных (приема 2020 г., очная форма обучения).

Разработчик(и):

|  |  |  |
| --- | --- | --- |
| Должность | Подпись | ФИО |
| Доцент ОИТ ИШИТР |  | Губин Е.И. |

Программа одобрена на заседании Отделения информационных технологий ИШИТР (протокол №18 от 09.06.2020 г.)

Заведующий кафедрой –

руководитель ОИТ на правах кафедры \_\_\_\_\_\_\_\_\_\_\_\_\_/ В.С. Шерстнев

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 |
|  |  |  |