

МИНОБРНАУКИ РОССИИ  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
**«Тверской государственный технический университет»**  
(ТвГТУ)

УТВЕРЖДАЮ  
Проректор по учебной работе  
\_\_\_\_\_ Э.Ю. Майкова  
« \_\_\_\_ » \_\_\_\_\_ 20 г.

**РАБОЧАЯ ПРОГРАММА**  
дисциплины обязательной части Блока 1 «Дисциплины (модули)»  
**«Анализ больших данных»**

Направление подготовки бакалавров – 09.03.02. Информационные системы и технологии

Направленность (профиль) - Разработка, внедрение и сопровождение информационных систем

Типы задач профессиональной деятельности: организационно-управленческий; проектный

Форма обучения – очная, заочная

Факультет информационных технологий  
Кафедра «Информационных систем»

Тверь 20\_\_

Рабочая программа дисциплины соответствует ОХОП подготовки бакалавров в части требований к результатам обучения по дисциплине и учебному плану.

Разработчик программы: к.т.н., доцент

Е.Е. Фомина

Программа рассмотрена и одобрена на заседании кафедры ИС  
«\_\_\_\_\_» \_\_\_\_\_ 20\_\_\_\_ г., протокол № \_\_\_\_\_.

Заведующий кафедрой

Б.В. Палюх

Согласовано  
Начальник учебно-методического  
отдела УМУ

Д.А. Барчуков

Начальник отдела  
комплектования  
зональной научной библиотеки

О.Ф. Жмыхова

## 1. Цель и задачи дисциплины

**Целью** изучения дисциплины «Анализ больших данных» является формирование компетенций в области обработки, анализа и использования информации для информационно-аналитической поддержки принятия управленческих решений. Знания, умения и навыки полученные в результате освоения дисциплины, помогут при сборе и анализе больших объемов структурированной и неструктурированной информации, при разработке моделей данных, и получении новых знаний.

**Задачами дисциплины** являются:

приобретение студентами знаний о технологиях подготовки, хранения, обработки и анализа больших данных;

приобретение практических навыков работы большими данными.

## 2. Место дисциплины в структуре ООП

Дисциплина относится к обязательной части Блока 1 ОП ВО. Для изучения курса требуются знания, умения и навыки, полученные в процессе изучения дисциплин «Информатика», «Информационные технологии», «Технологии программирования» и «Математика».

Приобретенные знания в рамках данной дисциплины помимо их самостоятельного значения являются основой для изучения курсов «Инструментальные средства информационных систем», «Управление ИТ-проектами», «Геоинформационные системы и технологии» и других дисциплин, профессиональная подготовка по которым предполагает использование технологий анализа больших данных при решении задач, а также при выполнении выпускной квалификационной работы.

## 3. Планируемые результаты обучения по дисциплине

### 3.1. Планируемые результаты обучения по дисциплине

**Компетенция, закрепленная за дисциплиной в ОХОП:**

**ОПК-1.** *Способен применять естественнонаучные и общинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности.*

**Индикаторы компетенции, закреплённых за дисциплиной в ОХОП:**

**ИОПК-1.1.** *Использует знания основ математики, физики, вычислительной техники, программирования и методов математического анализа и моделирования при решении стандартных профессиональных задач.*

**Показатели оценивания индикаторов достижения компетенций**

**Знать:**

31.1. Теоретические основы методов анализа больших данных и машинного обучения.

31.2. Типы задач, решаемые с использованием методов анализа больших данных и машинного обучения.

31.3. Современные технологии анализа больших данных и машинного обучения.

**Уметь:**

У1.1. Ставить задачи и разрабатывать программу исследования с использованием анализа больших данных и машинного обучения.

У1.2. Выбирать соответствующие методы решения экспериментальных и теоретических задач.

У1.3. Применять методы и подходы анализа больших данных и машинного обучения для решения экспериментальных и теоретических задач.

У1.4. Интерпретировать, представлять и применять результаты, полученные с использованием анализа больших данных и машинного обучения.

**Компетенция, закреплённая за дисциплиной в ОХОП:**

**ОПК-8.** *Способен применять математические модели, методы и средства проектирования информационных и автоматизированных систем.*

**Индикаторы компетенции, закреплённых за дисциплиной в ОХОП:**

**ИОПК-8.1.** *Использует полученные знания математики, методологии и основных методов математического моделирования, классификации и условий применения моделей, методов и средств проектирования информационных и автоматизированных систем, инструментальных средств моделирования и проектирования в профессиональной деятельности.*

**Показатели оценивания индикаторов достижения компетенций**

**Знать:**

32.1. Основные понятия термина Большие Данные.

32.2. Основные понятия Data Mining.

32.3. Основные способы хранения и анализа данных.

32.4. Методы извлечения знаний из данных (кластеризация, классификация, прогнозирование, регрессия, нейронные сети).

32.5. Научные методы обработки и визуализации данных.

**Уметь:**

У2.1. Работать с программными средствами для хранения и анализа данных;

У2.2. Разрабатывать и адаптировать программные компоненты работы с данными для нужд предприятия.

У2.3. Реализовывать приложения для бизнес аналитики больших данных

У2.4. Обрабатывать и анализировать большие объёмы данных с помощью современных метрик

У2.5. Визуализировать, интерпретировать и давать рекомендации на основании результатов интеллектуального анализа больших данных

**Индикаторы компетенции, закреплённых за дисциплиной в ОХОП:**

**ИОПК-8.2.** *Моделирует процессы и систем, проектирует информационные и автоматизированные системы с применением современных инструментальных средств.*

**Показатели оценивания индикаторов достижения компетенций**

**Знать:**

ЗЗ.1. Основные технологии, применяемые для хранения, извлечения, поиска и анализа больших данных.

ЗЗ.2. Современные инструментальные средства.

**Уметь:**

УЗ.1. Анализировать современные потоки данных;

УЗ.2. Находить, извлекать и структурировать данные;

УЗ.3. Моделировать процессы и системы, проектировать информационные и автоматизированные системы, используя возможности Больших данных.

**3.2. Технологии, обеспечивающие формирование компетенций**

Проведение лекционных и лабораторных занятий.

**4. Трудоемкость дисциплины и виды учебной работы****ОЧНАЯ ФОРМА ОБУЧЕНИЯ**

*Таблица 1а.* Распределение трудоемкости дисциплины по видам учебной работы

Вид учебной работы	Зачетные единицы	Академические часы
<b>Общая трудоемкость дисциплины</b>	4	144
<b>Аудиторные занятия (всего)</b>		75
В том числе:		
Лекции		30
Практические занятия (ПЗ)		не предусмотрены
Лабораторные работы (ЛР)		45
<b>Самостоятельная работа обучающихся (всего)</b>		69
В том числе:		
Курсовая работа		не предусмотрена
Курсовой проект		не предусмотрен
Расчетно-графические работы		не предусмотрены
Другие виды самостоятельной работы: - подготовка к лабораторным работам		50
Текущий контроль успеваемости и промежуточная аттестация (зачет)		19
<b>Практическая подготовка при реализации дисциплины (всего)</b>		0

**ЗАОЧНАЯ ФОРМА ОБУЧЕНИЯ**

*Таблица 1б.* Распределение трудоемкости дисциплины по видам учебной работы

Вид учебной работы	Зачетные единицы	Академические часы
<b>Общая трудоемкость дисциплины</b>	4	144
<b>Аудиторные занятия (всего)</b>		10
В том числе:		
Лекции		4
Практические занятия (ПЗ)		не предусмотрены

Вид учебной работы	Зачетные единицы	Академические часы
Лабораторные работы (ЛР)		6
<b>Самостоятельная работа обучающихся (всего)</b>		130+4 (контроль)
В том числе:		
Курсовая работа		не предусмотрена
Курсовой проект		не предусмотрен
Расчетно-графические работы		не предусмотрены
Другие виды самостоятельной работы: - подготовка к лабораторным работам		80
Текущий контроль успеваемости и промежуточная аттестация (зачет)		50+4 (контроль)
<b>Практическая подготовка при реализации дисциплины (всего)</b>		0

## 5. Структура и содержание дисциплины.

### 5.1. Структура дисциплины ОЧНАЯ ФОРМА ОБУЧЕНИЯ

Таблица 2а. Модули дисциплины, трудоемкость в часах и виды учебной работы

№	Наименование модуля	Труд-ть часы	Лекции	Практич. занятия	Лаб. работы	Сам. работа
1	Введение в Большие данные	6	2	-	-	4
2	Библиотеки работы с датафремами в python	18	6	-	8	8
3	Статистический анализ	29	6	-	10	17
4	Визуализация данных и результатов анализа	31	6	-	10	20
5	Интеллектуальный анализ данных	38	10	-	17	20
<i>Всего на дисциплину</i>		<b>144</b>	<b>30</b>	-	<b>45</b>	<b>69</b>

### ЗАОЧНАЯ ФОРМА ОБУЧЕНИЯ

Таблица 2б. Модули дисциплины, трудоемкость в часах и виды учебной работы

№	Наименование модуля	Труд-ть часы	Лекции	Практич. занятия	Лаб. работы	Сам. работа
1	Введение в Большие данные	21	0,5	-	-	20+0,5 (контроль)
2	Библиотеки работы с датафремами в python	22	0,5	-	1	20+0,5 (контроль)
3	Статистический анализ	23	1	-	1	20+1 (контроль)
4	Визуализация данных и результатов анализа	33	1	-	1	30+1 (контроль)
5	Интеллектуальный анализ данных	45	1	-	3	40+1 (контроль)
<i>Всего на дисциплину</i>		<b>144</b>	<b>4</b>	-	<b>6</b>	<b>130+4 (контроль)</b>

## 5.2. Содержание дисциплины

### МОДУЛЬ 1 «Введение в Большие данные»

Определение понятия Большие данные (Big Data). Источники Big Data. Характеристики Больших данных. Примеры использования Больших данных для решения практических задач. Виды Больших данных. Методы обработки больших данных. Инструменты обработки больших данных.

### МОДУЛЬ 2 «Библиотеки работы с датафреймами в python»

Синтаксис языка Python для основных алгоритмических конструкций, литералов, выражений. Основные структуры данных python. Описание встроенных типов данных, особенности общепринятого в Python стиля программирования. Разнообразие типов-коллекций в языке Python. Знание основных технологий, применяемых для хранения, извлечения, поиска и анализа больших данных. Методы сбора Больших данных. Преобразование данных. Библиотека Pandas для работы с датафреймами. Выгрузка данных из csv, xls файлов.

### МОДУЛЬ 4 «Статистический анализ»

Показатели описательной статистики. Библиотеки статистического анализа данных в python. Применение статистического анализа для решения прикладных задач аналитики.

### МОДУЛЬ 5 «Визуализация данных и результатов анализа»

Библиотека matplotlib в python. Методы визуализации данных в python: графики гистограммы, круговые диаграммы. Визуализация при решении прикладных задач аналитики.

### МОДУЛЬ 6 «Интеллектуальный анализ данных»

Библиотеки pandas, sklearn, scipy. Многомерный корреляционно-регрессионный анализ. Классификация регрессионных зависимостей. Основные этапы и задачи регрессионного анализа. Классическая модель множественной линейной регрессии (МЛР). Оценивание параметров модели МЛР. Оценка адекватности модели. Частный коэффициент корреляции и его использование в МЛР. Причины мультиколлинеарности и методы её устранения. Обобщенная линейная модель множественной регрессии и оценивание её параметров. Взвешенный метод наименьших квадратов. Методы построения оптимальных регрессионных моделей. Достоинства и недостатки пошаговых алгоритмов. Понятие о гребневой регрессии. Нелинейный регрессионный анализ. Методы анализа остатков регрессионных моделей. Реализация в python. Решение задач классификации. Деревья классификации. Основные понятия. Алгоритмы построения деревьев классификации: CART, CHAID. Оценка качества построенной модели. Реализация в python. Кластерный анализ. Сущность кластерного анализа и его геометрическая интерпретация. Основные этапы

кластерного анализа: выбор системы признаков, снижение размерности признакового пространства, выбор меры близости, выбор алгоритма классификации и оценка её качества. Классификация методов кластерного анализа. Эвристические алгоритмы классификации. Иерархические алгоритмы классификации. Метод k-средних. Критерии качества классификации. Реализация в python. Нейронные сети. История развития и области применимости НС. Основные компоненты НС. Правило распространения сигналов в сети. Правило вычисления сигнала активности. Простейшее правило обучения, корректирующее связи. Функция выбора решения. Корректировка весов. Минимизация квадрата ошибки. Линейные и нелинейные задачи. Динамические нейронные сети. Однослойный и многослойный персептрон. Сети Кохонена и Хопфилда. Основные идеи. Самоорганизующаяся карта признаков. Алгоритм работы СКП. Обучение сети СКП. Примеры использования СКП.

### 5.3. Лабораторные работы ОЧНАЯ ФОРМА ОБУЧЕНИЯ

Таблица 3а. Лабораторные работы и их трудоемкость

Порядковый номер модуля. Цели лабораторных работ	Наименование лабораторных работ	Трудоемкость в часах
<b>Модуль 2</b> <b>Цель:</b> освоить работу с основными операторами и структурами данных в python и библиотекой pandas	Линейные процессы, ветвление, циклы Списки, словари, кортежи, множества Обработка исключений Работа с библиотекой pandas: считывание данных различных форматов Методы работы с датафреймами	<b>8</b>
<b>Модуль 3</b> <b>Цель:</b> научиться проводить статистический анализ в python	Расчет показателей описательной статистики Применение статистического анализа для решения прикладных задач аналитики	<b>10</b>
<b>Модуль 4</b> <b>Цель:</b> освоить методы визуализации данных в python	Визуализация данных: гистограммы, круговые диаграммы, графики Визуализация при решении прикладных задач аналитики	<b>10</b>
<b>Модуль 5</b> <b>Цель:</b> освоит методы интеллектуального анализа данных в python	Кластерный анализ Корреляционно-регрессионный анализ Деревья классификации Нейронные сети	<b>17</b>

## ЗАОЧНАЯ ФОРМА ОБУЧЕНИЯ

Таблица 3б. Лабораторные работы и их трудоемкость

Порядковый номер модуля. Цели лабораторных работ	Наименование лабораторных работ	Трудоемкость в часах
<b>Модуль 2</b> <b>Цель:</b> освоить работу с основными операторами и структурами данных в python и библиотекой pandas	Линейные процессы, ветвление, циклы Списки, словари, кортежи, множества Обработка исключений Работа с библиотекой pandas: считывание данных различных форматов Методы работы с датафреймами	<b>1</b>
<b>Модуль 3</b> <b>Цель:</b> научиться проводить статистический анализ в python	Расчет показателей описательной статистики Применение статистического анализа для решения прикладных задач аналитики	<b>1</b>
<b>Модуль 4</b> <b>Цель:</b> освоить методы визуализации данных в python	Визуализация данных: гистограммы, круговые диаграммы, графики Визуализация при решении прикладных задач аналитики	<b>1</b>
<b>Модуль 5</b> <b>Цель:</b> освоит методы интеллектуального анализа данных в python	Кластерный анализ Корреляционно-регрессионный анализ Деревья классификации Нейронные сети	<b>3</b>

### 5.4. Практические и (или) семинарские занятия

Учебным планом практические занятия не предусмотрены.

## 6. Самостоятельная работа обучающихся и текущий контроль успеваемости

### 6.1. Цели самостоятельной работы

Формирование способностей к самостоятельному познанию и обучению, поиску литературы, обобщению, оформлению и представлению полученных результатов, их критическому анализу, поиску новых и неординарных решений, аргументированному отстаиванию своих предложений, умений подготовки выступлений и ведения дискуссий.

### 6.2. Организация самостоятельной работы

Самостоятельная работа заключается в изучении отдельных тем курса по заданию преподавателя по рекомендуемой им учебной литературе, в подготовке к лабораторным работам, к текущему контролю успеваемости и подготовке к зачету.

После вводных лекций, в которых обозначается содержание дисциплины,

ее проблематика и практическая значимость, студентам выдаются задания на лабораторные работы. Лабораторные работы охватывают модули 2-6.

В рамках дисциплины выполняется 20 лабораторных работ, которые защищаются устным опросом.

Выполнение всех лабораторных работ обязательно.

В случае невыполнения лабораторной работы по уважительной причине студент должен выполнить пропущенные лабораторные занятия в часы, отведенные на консультирование с преподавателем.

## **7. Учебно-методическое и информационное обеспечение дисциплины**

### **7.1. Основная литература по дисциплине**

1. Воронов, В.И. Data Mining - технологии обработки больших данных : учебное пособие / В.И. Воронов, Л.И. Воронова, В.А. Усачев. - Москва : Московский технический университет связи и информатики, 2018. - ЦОР IPR SMART. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - URL: <https://www.iprbookshop.ru/81324.html> . - (ID=145723-0)

2. Интеллектуальный предиктивный мультимодальный анализ слабоструктурированных больших данных : учебное пособие / Н.Г. Ярушкина [и др.]. - Ульяновск : Ульяновский государственный технический университет, 2020. - ЦОР IPR SMART. - Текст : непосредственный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - ISBN 978-5-9795-2088-9. - URL: <https://www.iprbookshop.ru/106136.html> . - (ID=145145-0)

3. Макшанов, А.В. Большие данные. Big Data : учебник / А.В. Макшанов, А.Е. Журавлев, Л.Н. Тындыкарь. - 2-е изд. - Санкт-Петербург [и др.] : Лань, 2022. - ЭБС Лань. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - ISBN 978-5-8114-9690-7. - URL: <https://e.lanbook.com/book/198599> . - (ID=145911-0)

### **7.2. Дополнительная литература по дисциплине**

1. Анализ данных : учебник для академического бакалавриата / В.С. Мхитарян [и др.]; под ред. В.С. Мхитаряна ; Национальный исследовательский ун-т - Высшая Школа Экономики. - Москва : Юрайт, 2022. - (Бакалавр. Академический курс). - Образовательная платформа Юрайт. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - ISBN 978-5-534-00616-2. - URL: <https://urait.ru/bcode/489100> . - (ID=100456-0)

2. Андрианова, Е.Е. Управление данными. Интеллектуальный анализ данных : учебное пособие / Е.Е. Андрианова, И.А. Липанова, О.Ю. Сабинин; Санкт-Петербургский государственный университет телекоммуникаций имени М.А. Бонч-Бруевича. - Санкт-Петербург : Санкт-Петербургский государственный университет телекоммуникаций имени М.А. Бонч-Бруевича, 2016. - ЭБС Лань. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - URL: <https://e.lanbook.com/book/180046> . - (ID=145914-0)

3. Вольфсон, М.Б. Анализ данных : учебное пособие / М.Б. Вольфсон; Санкт-Петербургский государственный университет телекоммуникаций имени М.А. Бонч-Бруевича. - Санкт-Петербург : Санкт-Петербургский государственный университет телекоммуникаций имени М.А. Бонч-Бруевича, 2015. - ЭБС Лань. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - URL: <https://e.lanbook.com/book/180254> . - (ID=145915-0)
4. Головенчик, Г.Г. Цифровая экономика : учебно-методический комплекс для студентов, обучающихся по специальности «Мировая экономика» / Г.Г. Головенчик; Белорусский государственный университет. - Минск : Белорусский государственный университет, 2020. - ЭБС Лань. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - ISBN 978-985-566-847-4. - URL: <https://e.lanbook.com/book/180524> . - (ID=145905-0)
5. Железнов, М.М. Методы и технологии обработки больших данных : учебно-методическое пособие / М.М. Железнов. - Москва : МИСИ-МГСУ : ЭБС АСВ, 2020. - ЦОР IPR SMART. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - ISBN 978-5-7264-2193-3. - URL: <https://www.iprbookshop.ru/101802.html> . - (ID=145721-0)
6. Каленик, А.И. Использование новых возможностей Microsoft SQL Server 2005 / А.И. Каленик. - СПб. [и др.] ; М. : Питер : Русская редакция, 2006. - 321 с. : ил. - Текст : непосредственный. - ISBN 5-469-01404-5 (Питер) : 180 p. - (ID=66318-2)
7. Кобзаренко, Д. Н. Учебное пособие дисциплины «Анализ больших данных» для направления подготовки 38.03.05 «Бизнес-информатика», профиль «Электронный бизнес» : учебное пособие / Д. Н. Кобзаренко, А. Г. Мустафаев ; составитель Д. Н. Кобзаренко. — Махачкала : ДГУНХ, 2019. — 107 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/246542> (дата обращения: 04.10.2022). — Режим доступа: для авториз. пользователей. - (ID=150643-0)
8. Методы и модели исследования сложных систем и обработки больших данных : монография / И.Ю. Парамонов [и др.]; Парамонов И.Ю., Смагин В.А., Косых Н.Е., Хомоненко А.Д. ; под редакцией В.А. Смагина и А.Д. Хомоненко. - Санкт-Петербург [и др.] : Лань, 2020. - ЭБС Лань. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - ISBN 978-5-8114-4006-1. - URL: <https://e.lanbook.com/book/126938> . - (ID=136026-0)
9. Методы и модели исследования сложных систем и обработки больших данных : монография / И.Ю. Парамонов [и др.]; Парамонов И.Ю., Смагин В.А., Косых Н.Е., Хомоненко А.Д. ; под редакцией В.А. Смагина и А.Д. Хомоненко. - Санкт-Петербург [и др.] : Лань, 2020. - ЭБС Лань. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - ISBN 978-5-8114-4006-1. - URL: <https://e.lanbook.com/book/126938> . - (ID=136026-0)
10. Миркин, Б.Г. Введение в анализ данных : учебник и практикум / Б.Г. Миркин. - Москва : Юрайт, 2022. - (Высшее образование). - Образовательная платформа Юрайт. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.09.2022. - ISBN 978-5-9916-5009-0. - URL: <https://urait.ru/book/vvedenie-v-analiz-dannyh-469306> . - (ID=134346-0)

11. Новикова, О.А. Анализ данных : учебное пособие. Часть 1 / О.А. Новикова, Е.Г. Андрианова; МИРЭА - Российский технологический университет. - Москва : МИРЭА - Российский технологический университет, 2020. - ЭБС Лань. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - URL: <https://e.lanbook.com/book/167597> . - (ID=145912-0)
12. Просто о больших данных / Д. Гурвиц [и др.]. - Москва : Эксмо, 2015. - 394 с. - (Библиотека Сбербанка. Т.58). - Текст : непосредственный. - ISBN 978-5-699-85806-4 : 414 p. - (ID=114809-5)
13. Рындина, С.В. Цифровая трансформация бизнеса: использование аналитики на основе больших данных : учебное пособие / С.В. Рындина; Пензенский государственный университет. - Пенза : Пензенский государственный университет, 2019. - ЭБС Лань. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - ISBN 978-5-907262-04-1. - URL: <https://e.lanbook.com/book/162301> . - (ID=145921-0)
14. Синева, И.С. Анализ данных в среде R : учебное пособие. Ч. 1 / И.С. Синева. - Москва : Московский технический университет связи и информатики, 2018. - ЦОР IPR SMART. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - Текст : электронный. - URL: <https://www.iprbookshop.ru/92422.html> . - (ID=146948-0)
15. Федин, Ф.О. Анализ данных : учебное пособие. Ч. 2 : Инструменты Data Mining / Ф.О. Федин, Ф.Ф. Федин. - Москва : Московский городской педагогический университет, 2012. - ЦОР IPR SMART. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - URL: <https://www.iprbookshop.ru/26445.html> . - (ID=146080-0)
16. Целых, А.Н. Современные методы прикладной информатики в задачах анализа данных : учебное пособие по курсу «Методы интеллектуального анализа данных» / А.Н. Целых, А.А. Целых, Э.М. Котов. - Ростов-на-Дону ; Таганрог : Издательство Южного федерального университета, 2021. - ЦОР IPR SMART. - Текст : электронный. - Режим доступа: по подписке. - Дата обращения: 07.07.2022. - ISBN 978-5-9275-3783-9. - URL: <https://www.iprbookshop.ru/117165.htm> . - (ID=145142-0)

### 7.3. Методические материалы

1. Вопросы по дисциплине "Управление и анализ данных", направление подготовки 09.04.01 Информатика и вычислительная техника, профиль - Информационное и программное обеспечение автоматизированных систем : в составе учебно-методического комплекса / Тверской гос. техн. ун-т, Каф. ЭВМ ; разработ. А.В. Попов. - Тверь : ТвГТУ, 2016. - (УМК-В). - Сервер. - Текст : электронный. - URL: <https://elib.tstu.tver.ru/MegaPro/GetDoc/Megapro/124790> . - (ID=124790-0)
2. Верпатова, О.Ю. Анализ данных в SPSS : учеб. пособие : в составе учебно-методического комплекса / О.Ю. Верпатова; Тверской гос. техн. ун-т. - Тверь : ТвГТУ, 2016. - 80 с. : ил. - (УМК-У). - Текст : непосредственный. - Текст : непосредственный. - ISBN 978-5-7995-0824-1 : [б. ц.]. - (ID=112484-75)

3. Методические указания по курсу "Компьютерный анализ данных в SPSS" для студентов направления подготовки бакалавров 040100 Социология и 080400 Управление персоналом дневной, заочной форм обучения и института заочного и дополнительного профессионального образования ТвГТУ / Тверской гос. техн. ун-т, Каф. СиСТ ; сост. А.В. Вайсбург. - Тверь : ТвГТУ, 2014. - 31 с. - Текст : непосредственный. - 33 р. 20 к. - (ID=102493-95)

4. Методические указания по курсу "Компьютерный анализ данных в SPSS" для студентов направления подготовки бакалавров 040100 Социология и 080400 Управление персоналом дневной, заочной форм обучения и института заочного и дополнительного профессионального образования ТвГТУ / Тверской гос. техн. ун-т, Каф. СиСТ ; сост. А.В. Вайсбург. - Тверь : ТвГТУ, 2014. - Сервер. - Текст : электронный. - 0-00. - URL: <https://elib.tstu.tver.ru/MegaPro/GetDoc/Megapro/102275> . - (ID=102275-1)

5. Комиссарчик, В.Ф. Анализ данных и планирование эксперимента : учеб. пособие : в составе учебно-методического комплекса / В.Ф. Комиссарчик; Тверской гос. техн. ун-т. - Тверь : ТвГТУ, 2000. - 135 с. - (УМК-У). - Текст : непосредственный. - 30 р. - (ID=5522-5)

#### **7.4. Программное обеспечение по дисциплине**

Операционная система Microsoft Windows: лицензии № ICM-176609 и № ICM-176613 (Azure Dev Tools for Teaching).

Microsoft Office 2007 Russian Academic: OPEN No Level: лицензия № 41902814.

WPS Office: MPL 1.1/GPL 2.0/LGPL 2.1.

Libre Office: MPL 2.0.

LMS Moodle: GPL 3.0.

Python 3.8.10

#### **7.5. Специализированные базы данных, справочные системы, электронно-библиотечные системы, профессиональные порталы в Интернет**

ЭБС и лицензионные ресурсы ТвГТУ размещены:

1. Ресурсы: <https://lib.tstu.tver.ru/header/obr-res>
2. ЭКТвГТУ: <https://elib.tstu.tver.ru/MegaPro/Web>
3. ЭБС "Лань": <https://e.lanbook.com/>
4. ЭБС "Университетская библиотека онлайн": <https://www.biblioclub.ru/>
5. ЭБС «IPRBooks»: <https://www.iprbookshop.ru/>
6. Электронная образовательная платформа "Юрайт" (ЭБС «Юрайт»): <https://urait.ru/>
7. Научная электронная библиотека eLIBRARY: <https://elibrary.ru/>
8. Информационная система "ТЕХНОРМАТИВ". Конфигурация "МАКСИМУМ" : сетевая версия (годовое обновление): [нормативно-технические, нормативно-правовые и руководящие документы (ГОСТы, РД,

СНиПы и др.]. Диск 1,2,3,4. - М. :Технорматив, 2014. - (Документация для профессионалов). - CD. - Текст : электронный. - 119600 р. – (105501-1)

9. База данных учебно-методических комплексов:<https://lib.tstu.tver.ru/header/umk.html>

УМК размещен: <https://elib.tstu.tver.ru/MegaPro/GetDoc/Megapro/150644>

## **8. Материально-техническое обеспечение**

Кафедра «Информационных систем» имеет аудитории для проведения лекций и лабораторных занятий по дисциплине; специализированные учебные классы, оснащенные современной компьютерной техникой, необходимым программным обеспечением, электронными учебными пособиями для проведения лабораторных работ и самостоятельной работы.

В наличии имеются презентационные мультимедийные лекционные курсы, разработанные преподавателями кафедры ИС и внешними разработчиками.

Для проведения лабораторных работ имеются лаборатории с персональными компьютерами (наличие локальной вычислительной сети необязательно).

## **9. Оценочные средства для проведения промежуточной аттестации**

### **9.1. Оценочные средства для проведения промежуточной аттестации в форме экзамена**

Учебным планом не предусмотрен.

### **9.2. Оценочные средства для проведения промежуточной аттестации в форме зачета**

1. Шкала оценивания промежуточной аттестации – «зачтено», «не зачтено».

2. Вид промежуточной аттестации в форме зачета.

Промежуточная аттестация в форме зачета устанавливается преподавателем по результатам текущего контроля знаний и умений, обучающегося без дополнительных контрольных испытаний.

При промежуточной аттестации без выполнения дополнительного итогового контрольного испытания студенту в обязательном порядке описываются критерии проставления зачёта:

«зачтено» - выставляется обучающемуся при условии выполнения им всех контрольных мероприятий, посещения лекций и лабораторных занятий в объеме, соответствующем не менее чем 80% от количества часов, отведенного на контактную работу с преподавателем.

3. Для дополнительного итогового контрольного испытания студенту в обязательном порядке предоставляется:

база заданий, предназначенных для предъявления обучающемуся на дополнительном итоговом контрольном испытании (типовой образец задания приведен в Приложении);

методические материалы, определяющие процедуру проведения дополнительного итогового испытания и проставления зачёта.

Задание выполняется письменно и с использованием ЭВМ.

***Перечень вопросов дополнительного итогового контрольного испытания:***

1. Определение понятия Большие данные (Big Data).
2. Источники Big Data.
3. Характеристики Больших данных.
4. Примеры использования Больших данных для решения практических задач.
5. Виды Больших данных.
6. Методы обработки больших данных. Инструменты обработки больших данных.
7. Синтаксис языка Python для основных алгоритмических конструкций, литералов, выражений.
8. Основные структуры данных в python: списки, кортежи, словари, множества.
9. Библиотека Pandas для работы с датафреймами.
10. Выгрузка данных из csv, xls файлов.
11. Показатели описательной статистики.
12. Библиотеки статистического анализа данных в python.
13. Применение статистического анализа для решения прикладных задач аналитики.
14. Библиотека matplotlib в python.
15. Методы визуализации данных в python: графики гистограммы, круговые диаграммы.
16. Визуализация при решении прикладных задач аналитики.
17. Многомерный корреляционно-регрессионный анализ: основные определения.
18. Классификация регрессионных зависимостей.
19. Основные этапы и задачи регрессионного анализа. Классическая модель множественной линейной регрессии (МЛР).
20. Оценивание параметров модели МЛР.
21. Оценка адекватности модели.
22. Частный коэффициент корреляции и его использование в МЛР.
23. Причины мультиколлинеарности и методы её устранения.
24. Обобщенная линейная модель множественной регрессии и оценивание её параметров.

25. Взвешенный метод наименьших квадратов.
26. Методы построения оптимальных регрессионных моделей.
27. Достоинства и недостатки пошаговых алгоритмов.
28. Нелинейный регрессионный анализ.
29. Методы анализа остатков регрессионных моделей.
30. Деревья классификации: основные понятия.
31. Алгоритмы построения деревьев классификации: CART, CHAID.
32. Оценка качества построенной модели.
33. Кластерный анализ.
34. Сущность кластерного анализа и его геометрическая интерпретация.
35. Основные этапы кластерного анализа: выбор системы признаков.
36. Основные этапы кластерного анализа: снижение размерности признакового пространства.
37. Основные этапы кластерного анализа: выбор меры близости.
38. Выбор алгоритма классификации и оценка её качества.
39. Классификация методов кластерного анализа.
40. Иерархические алгоритмы кластеризации.
41. Метод k-средних.
42. Критерии качества кластеризации.
43. История развития и области применимости НС.
44. Основные компоненты НС.
45. Правило распространения сигналов в сети.
46. Правило вычисления сигнала активности.
47. Простейшее правило обучения, корректирующее связи.
48. Функция выбора решения.
49. Корректировка весов.
50. Минимизация квадрата ошибки.
51. Линейные и нелинейные задачи.
52. Динамические нейронные сети.
53. Однослойный и многослойный персептрон.
54. Сети Кохонена и Хопфилда. Основные идеи. Самоорганизующаяся карта признаков.
55. Алгоритм работы СКП.
56. Обучение сети СКП.

Пользование различными техническими устройствами, кроме ЭВМ компьютерного класса и программным обеспечением, необходимым для решения поставленных задач, не допускается.

Преподаватель имеет право после проверки письменных ответов на экзаменационные вопросы задавать студенту в устной форме уточняющие вопросы в рамках содержания экзаменационного билета, выданного студенту.

Иные нормы, регламентирующие процедуру проведения экзамена, представлены в Положении о текущем контроле успеваемости и промежуточной аттестации студентов.

Число заданий для дополнительного итогового контрольного испытания - 20.

Число вопросов – 3 (2 вопроса для категории «знать» и 1 вопрос для категории «уметь»).

Продолжительность – 60 минут.

4. Критерии выполнения контрольного испытания и условия проставления зачёта:

для категории «знать» (бинарный критерий):

ниже базового - 0 балл;

базовый уровень – 1 балла;

критерии оценки и ее значение для категории «уметь» (бинарный критерий):

отсутствие умения – 0 балл;

наличие умения – 1 балла.

Критерии итоговой оценки за зачет:

«зачтено» - при сумме баллов 2 или 3;

«не зачтено» - при сумме баллов 0 или 1.

### **9.3. Оценочные средства для проведения промежуточной аттестации в форме курсовой работы или курсового проекта**

Учебным планом не предусмотрена.

## **10. Методические рекомендации по организации изучения дисциплины**

Студенты перед началом изучения дисциплины ознакомлены с системами кредитных единиц и балльно-рейтинговой оценки.

Студенты, изучающие дисциплину, обеспечиваются электронными изданиями или доступом к ним, учебно-методическим комплексом по дисциплине, включая методические указания к выполнению лабораторных, и всех видов самостоятельной работы.

В учебный процесс рекомендуется внедрение субъект-субъектной педагогической технологии, при которой в расписании каждого преподавателя определяется время консультаций студентов по закрепленному за ним модулю дисциплины.

## **11. Внесение изменений и дополнений в рабочую программу дисциплины**

Протоколами заседаний кафедры ежегодно обновляется содержание рабочих программ дисциплин, по утвержденной «Положением о рабочих программах дисциплин» форме.

Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Тверской государственный технический университет»

Направление подготовки бакалавров – 09.03.02. Информационные системы и технологии

Направленность (профиль) – Разработка, внедрение и сопровождение информационных систем

Кафедра «Информационных систем»

Дисциплина «Анализ больших данных»

**ЗАДАНИЕ ДЛЯ ДОПОЛНИТЕЛЬНОГО ИТОГОВОГО  
КОНТРОЛЬНОГО ИСПЫТАНИЯ № 1**

*1. Вопрос для проверки уровня «ЗНАТЬ» – 0 или 1 балл:*

Дайте определение понятию Большие данные

*2. Вопрос для проверки уровня «ЗНАТЬ» – 0 или 1 балл:*

Алгоритм иерархического кластерного анализа.

*3. Задание для проверки уровня «УМЕТЬ» – 0 или 1 балл:*

Написать программу для считывания данных из csv файла и расчета показателей описательной статистики предложенной величины.

**Критерии итоговой оценки за зачет:**

«зачтено» - при сумме баллов 2 или 3;

«не зачтено» - при сумме баллов 0, или 1.

Составитель: к.т.н, доцент \_\_\_\_\_ Е.Е. Фомина

Заведующий кафедрой: д.т.н., профессор \_\_\_\_\_ Б.В. Палюх