

УТВЕРЖДАЮ  
Первый проректор, проректор по  
УР

\_\_\_\_\_ А.Е. Рудин

«28» 06 2022 года

## Рабочая программа дисциплины

**Б1.В.ДВ.03.01** Технологии анализа больших данных

Учебный план: 2022-2023 09.03.03 ИИТА ЦТвФ ОО №1-1-144.plx

Кафедра: **33** Информационных систем и компьютерного дизайна

Направление подготовки:  
(специальность) 09.03.03 Прикладная информатика

Профиль подготовки: Цифровые технологии в финансах  
(специализация)

Уровень образования: бакалавриат

Форма обучения: очная

### План учебного процесса

Семестр (курс для ЗАО)		Контактная работа обучающихся		Сам. работа	Контроль, час.	Трудоё мкость, ЗЕТ	Форма промежуточной аттестации
		Лекции	Практ. занятия				
7	УП	17	17	47	27	3	Экзамен
	РПД	17	17	47	27	3	
Итого	УП	17	17	47	27	3	
	РПД	17	17	47	27	3	

Рабочая программа дисциплины составлена в соответствии с федеральным государственным образовательным стандартом высшего образования по направлению подготовки 09.03.03 Прикладная информатика, утверждённым приказом Министерства образования и науки Российской Федерации от 19.09.2017 г. № 922

Составитель (и):

Доцент

\_\_\_\_\_

Лебедева Светлана  
Викторовна

От кафедры составителя:  
Заведующий кафедрой информационных систем и  
компьютерного дизайна

\_\_\_\_\_

Сошников Антон  
Владимирович

От выпускающей кафедры:  
Заведующий кафедрой

\_\_\_\_\_

Сошников Антон  
Владимирович

Методический отдел:

\_\_\_\_\_

## 1 ВВЕДЕНИЕ К РАБОЧЕЙ ПРОГРАММЕ ДИСЦИПЛИНЫ

**1.1 Цель дисциплины:** Сформировать компетенции обучающегося в области теоретических и практических основ технологий анализа больших данных, получение знаний об особенностях анализа данных для решения разнообразных бизнес-аналитических задач, возникающих в профессиональной деятельности.

**1.2 Задачи дисциплины:**

Освоение методологии обработки больших данных;  
 Владение методами и технологиями работы с данными и знаниями;  
 Иметь представление о структуризации данных и технологиях Knowledge Discovery in Databases и Data Mining;

Освоение основ статистического анализа данных.

**1.3 Требования к предварительной подготовке обучающегося:**

Предварительная подготовка предполагает создание основы для формирования компетенций, указанных в п. 2, при изучении дисциплин:

- Операционные системы, сети и телекоммуникации
- Теория систем и системный анализ
- Алгоритмизация и программирование
- Исследование операций и методы оптимизации
- Программная инженерия
- Прикладная статистика
- Облачные и виртуальные технологии
- Аналитика бизнес-процессов

## 2 КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ

<b>ПК-8: Способен организовывать оценку соответствия требований существующих систем и их аналогов.</b>
<b>Знать:</b> Основные законы естественнонаучных дисциплин при тестировании программного обеспечения с учетом обработки больших объемов данных.
<b>Уметь:</b> Строить управляющий граф программы для тестирования, на основании использования основных законов естественнонаучных дисциплин, и современные информационно - коммуникационные технологии в профессиональной деятельности.
<b>Владеть:</b> Навыками разработки эффективных наборов тестов для простых и крупных информационных систем.
<b>ПК-1: Способен разрабатывать процедуры интеграции программных модулей.</b>
<b>Знать:</b> Основные понятия и терминологию в области технологий обработки и анализа больших данных; языки программирования для обработки данных и методы организации взаимодействия аналитических платформ с распределенными хранилищами информации.
<b>Уметь:</b> Разрабатывать программный код для эффективной обработки распределенных данных большого объема с использованием языков программирования; использовать программные средства визуализации и интерактивного исследования больших данных.
<b>Владеть:</b> Программными средствами автоматической загрузки и обработки больших данных; навыками работы с системой интерактивного исследования и визуализации больших данных.

## 3 РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Наименование и содержание разделов, тем и учебных занятий	Семестр (курс для ЗАО)	Контактная работа		СР (часы)	Инновац. формы занятий	Форма текущего контроля
		Лек. (часы)	Пр. (часы)			
Раздел 1. Технологии анализа данных	7					О
Тема 1. Введение в анализ больших данных. Методология анализа данных. Практические занятия: Аналитический и информационный подходы к моделированию.		2	2	4	ИЛ	
Тема 2. Основные принципы анализа данных. Извлечение и визуализация данных. Практические занятия: Этапы моделирования, формы представления данных. Подготовка данных к анализу		2	2	6	ИЛ	

Раздел 2. Технологии анализа больших объемов данных					
Тема 3. Технологии KDD и Data Mining. Практические занятия: Основные подходы к использованию технологий KDD и Data Mining.	2	2	4	ИЛ	0
Тема 4. Корреляционный анализ. Ассоциативные правила. Практические занятия: Значимость и поиск ассоциативных правил.	2	2	6	ИЛ	
Тема 5. Кластеризация. Алгоритм кластеризации k-means. Меры расстояний. Сети Кохонена. Карты Кохонена. Практические занятия: Проблемы алгоритмов кластеризации.	2	2	6	ИЛ	
Тема 6. Классификация данных. Линейная и логистическая регрессия. Практические занятия: Статистические методы.	1	1	4	ИЛ	
Тема 7. Машинное обучение. Введение в деревья решений. Алгоритмы построения. Нейронные сети. Искусственный нейрон. Практические занятия: Принципы построения нейронных сетей. Алгоритмы	2	2	6	ИЛ	
Тема 8. Анализ и прогнозирование временных рядов. Практические занятия: Модели прогнозирования.	2	2	4	ИЛ	
Тема 9. Аналитические платформы. Программное обеспечение в области анализа данных. Язык R. Практические занятия: Аналитика для неструктурированных данных - MapReduce и Hadoop.	2	2	7	ИЛ	
Итого в семестре (на курсе для ЗАО)	17	17	47		
Консультации и промежуточная аттестация (Экзамен)	2,5		24,5		
<b>Всего контактная работа и СР по дисциплине</b>	36,5		71,5		

#### 4 КУРСОВОЕ ПРОЕКТИРОВАНИЕ

Курсовое проектирование учебным планом не предусмотрено

#### 5. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

##### 5.1 Описание показателей, критериев и системы оценивания результатов обучения

##### 5.1.1 Показатели оценивания

Код компетенции	Показатели оценивания результатов обучения	Наименование оценочного средства
ПК-1	Формулирует основные понятия и терминологию в области больших данных.	Вопросы устного собеседования
	Классифицирует языки программирования для обработки больших объемов данных с использованием средств визуализации и интерактивного исследования.	Практико-ориентированные задания
	Демонстрирует программный код для организации автоматической загрузки и обработки больших данных.	Практико-ориентированные задания

ПК-8	Раскрывает сущность основных естественнонаучных законов используемых при тестировании программного обеспечения с учетом обработки больших объемов данных.	Вопросы устного собеседования
	Строит алгоритм программы для тестирования, на основании использования основных законов естественнонаучных дисциплин.	Практико-ориентированные задания
	Демонстрирует тест для тестирования информационной системы.	Практико-ориентированные задания

### 5.1.2 Система и критерии оценивания

Шкала оценивания	Критерии оценивания сформированности компетенций	
	Устное собеседование	Письменная работа
5 (отлично)	Полный, исчерпывающий ответ, явно демонстрирующий глубокое понимание предмета и широкую эрудицию в оцениваемой области. Критический, оригинальный подход к материалу. Качество исполнения всех элементов практико-ориентированного задания полностью соответствует всем требованиям.	
4 (хорошо)	Ответ полный, основанный на проработке всех обязательных источников информации. Подход к материалу ответственный, но стандартный. Практико-ориентированное задание выполнено в достаточном объеме, но ограничивается только основными подходами.	
3 (удовлетворительно)	Ответ стандартный, в целом качественный, основан на всех обязательных источниках информации. Присутствуют небольшие пробелы в знаниях или несущественные ошибки. Практико-ориентированное задание выполнено в соответствии с заданием. Имеются отдельные несущественные ошибки или отступления от правил оформления работы.	
2 (неудовлетворительно)	Неспособность ответить на вопрос без помощи экзаменатора. Незнание значительной части принципиально важных элементов дисциплины. Многочисленные грубые ошибки. Отсутствие одного или нескольких обязательных элементов практико-ориентированного задания, либо многочисленные грубые ошибки в работе, либо грубое нарушение правил оформления или сроков представления работы. Не учитываются баллы, накопленные в течение семестра.	

### 5.2 Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности

#### 5.2.1 Перечень контрольных вопросов

№ п/п	Формулировки вопросов
Семестр 7	
1	Основные принципы анализа данных.
2	Извлечение и визуализация данных.
3	Этапы моделирования и формы представления данных.
4	Этапы подготовки данных к анализу.
5	Технологии KDD.

6	Технологии Data Mining.
7	Корреляционный анализ.
8	Ассоциативные правила. Значимость и поиск ассоциативных правил.
9	Кластеризация. Алгоритм кластеризации k-means.
10	Меры расстояний.
11	Сети Кохонена. Карты Кохонена.
12	Классификация данных.
13	Линейная и логистическая регрессия.
14	Статистические методы обработки данных.
15	Нейронные сети. Искусственный нейрон.
16	Принципы построения нейронных сетей. Алгоритмы обучения нейронных сетей.
17	Анализ и прогнозирование временных рядов.
18	Модели прогнозирования.
19	Аналитические платформы.
20	Программное обеспечение в области анализа данных.
21	Возможности языка R для обработки данных.
22	Аналитика для неструктурированных данных - MapReduce и Hadoop.

### 5.2.2 Типовые тестовые задания

Не предусмотрено.

### 5.2.3 Типовые практико-ориентированные задания (задачи, кейсы)

Типовые тестовые задания находятся в Приложении к данной РПД

### 5.3 Методические материалы, определяющие процедуры оценивания знаний, умений, владений (навыков и (или) практического опыта деятельности)

#### 5.3.1 Условия допуска обучающегося к промежуточной аттестации и порядок ликвидации академической задолженности

Проведение промежуточной аттестации регламентировано локальным нормативным актом СПбГУПТД «Положение о проведении текущего контроля успеваемости и промежуточной аттестации обучающихся»

#### 5.3.2 Форма проведения промежуточной аттестации по дисциплине

Устная  Письменная  Компьютерное тестирование  Иная

#### 5.3.3 Особенности проведения промежуточной аттестации по дисциплине

- время на подготовку к устному собеседованию составляет 30 минут;
- выполнение кейс-задания осуществляется на компьютере за 60 минут.

## 6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

### 6.1 Учебная литература

Автор	Заглавие	Издательство	Год издания	Ссылка
<b>6.1.1 Основная учебная литература</b>				
Адлер, Ю. П., Черных, Е. А.	Статистическое управление процессами. «Большие данные»	Москва: Издательский Дом МИСиС	2016	<a href="http://www.iprbookshop.ru/64199.html">http://www.iprbookshop.ru/64199.html</a>
Пальмов, С. В.	Интеллектуальный анализ данных	Самара: Поволжский государственный университет телекоммуникаций и информатики	2017	<a href="http://www.iprbookshop.ru/75376.html">http://www.iprbookshop.ru/75376.html</a>
Мельниченко, А. С.	Математическая статистика и анализ данных	Москва: Издательский Дом МИСиС	2018	<a href="http://www.iprbookshop.ru/78563.html">http://www.iprbookshop.ru/78563.html</a>
<b>6.1.2 Дополнительная учебная литература</b>				
Спицкий С. В.	Эффективная аудиторная и самостоятельная работа обучающихся	СПб.: СПбГУПТД	2015	<a href="http://publish.sutd.ru/tp_ext_inf_publish.php?id=2015811">http://publish.sutd.ru/tp_ext_inf_publish.php?id=2015811</a>
Шифф В. К.	Корреляционный, регрессионный и дисперсионный анализ	СПб.: СПбГУПТД	2019	<a href="http://publish.sutd.ru/tp_ext_inf_publish.php?id=201950">http://publish.sutd.ru/tp_ext_inf_publish.php?id=201950</a>
Караулова И. Б., Мелешкова Г. И., Новоселов Г. А.	Организация самостоятельной работы обучающихся	СПб.: СПбГУПТД	2014	<a href="http://publish.sutd.ru/tp_ext_inf_publish.php?id=2014550">http://publish.sutd.ru/tp_ext_inf_publish.php?id=2014550</a>

Федин, Ф. О., Федин, Ф. Ф.	Анализ данных. Часть 1. Подготовка данных к анализу	Москва: Московский городской педагогический университет	2012	<a href="http://www.iprbookshop.ru/26444.html">http://www.iprbookshop.ru/26444.html</a>
Федин, Ф. О., Федин, Ф. Ф.	Анализ данных. Часть 2. Инструменты Data Mining	Москва: Московский городской педагогический университет	2012	<a href="http://www.iprbookshop.ru/26445.html">http://www.iprbookshop.ru/26445.html</a>
Жуковский, О. И.	Информационные технологии и анализ данных	Томск: Томский государственный университет систем управления и радиоэлектроники, Эль Контент	2014	<a href="http://www.iprbookshop.ru/72106.html">http://www.iprbookshop.ru/72106.html</a>

### 6.2 Перечень профессиональных баз данных и информационно-справочных систем

Информационная система «Единое окно доступа к образовательным ресурсам. Раздел. Информатика и информационные технологии» [Электронный ресурс]. URL: [http://window.edu.ru/catalog/?p\\_rubr=2.2.75.6](http://window.edu.ru/catalog/?p_rubr=2.2.75.6)

Электронно-библиотечная система IPRbooks [Электронный ресурс]. URL: <http://www.iprbookshop.ru/>

### 6.3 Перечень лицензионного и свободно распространяемого программного обеспечения

MicrosoftOfficeProfessional

Microsoft Windows

R

Deductor Academic

### 6.4 Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Аудитория	Оснащение
Лекционная аудитория	Мультимедийное оборудование, специализированная мебель, доска
Компьютерный класс	Мультимедийное оборудование, компьютерная техника с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду

**Приложение**  
 рабочей программы дисциплины Технологии анализа больших данных  
наименование дисциплины

по направлению подготовки 09.03.03 Прикладная информатика  
 наименование ОП (профиля): Цифровые технологии в финансах

**5.2.3 Типовые практико-ориентированные задания (задачи, кейсы)**

№ п/п	Условия типовых практико-ориентированных заданий (задач, кейсов)																																																			
1	<p>Провести регрессионный анализ зависимости выручки от числа торговых точек.</p> <p align="right">Таблица 1</p> <table border="1" data-bbox="539 398 1206 1205"> <thead> <tr> <th colspan="3" data-bbox="539 398 1206 409">Исходные данные</th> </tr> <tr> <th data-bbox="539 409 683 499">№</th> <th data-bbox="683 409 970 499">Выручка, тыс. руб.</th> <th data-bbox="970 409 1206 499">Число торговых точек</th> </tr> </thead> <tbody> <tr><td>1</td><td>1598</td><td>2</td></tr> <tr><td>2</td><td>2644</td><td>5</td></tr> <tr><td>3</td><td>2197</td><td>4</td></tr> <tr><td>4</td><td>1959</td><td>5</td></tr> <tr><td>5</td><td>1052</td><td>3</td></tr> <tr><td>6</td><td>1922</td><td>3</td></tr> <tr><td>7</td><td>2385</td><td>5</td></tr> <tr><td>8</td><td>2581</td><td>5</td></tr> <tr><td>9</td><td>3105</td><td>5</td></tr> <tr><td>10</td><td>3896</td><td>4</td></tr> <tr><td>11</td><td>1510</td><td>4</td></tr> <tr><td>12</td><td>1880</td><td>2</td></tr> <tr><td>13</td><td>3620</td><td>5</td></tr> <tr><td>14</td><td>5002</td><td>6</td></tr> <tr><td>15</td><td>2819</td><td>5</td></tr> </tbody> </table>	Исходные данные			№	Выручка, тыс. руб.	Число торговых точек	1	1598	2	2	2644	5	3	2197	4	4	1959	5	5	1052	3	6	1922	3	7	2385	5	8	2581	5	9	3105	5	10	3896	4	11	1510	4	12	1880	2	13	3620	5	14	5002	6	15	2819	5
Исходные данные																																																				
№	Выручка, тыс. руб.	Число торговых точек																																																		
1	1598	2																																																		
2	2644	5																																																		
3	2197	4																																																		
4	1959	5																																																		
5	1052	3																																																		
6	1922	3																																																		
7	2385	5																																																		
8	2581	5																																																		
9	3105	5																																																		
10	3896	4																																																		
11	1510	4																																																		
12	1880	2																																																		
13	3620	5																																																		
14	5002	6																																																		
15	2819	5																																																		
2	<p>Провести корреляционный анализ зависимости выручки от числа торговых точек.</p> <p align="right">Таблица 1</p> <table border="1" data-bbox="539 1283 1206 2087"> <thead> <tr> <th colspan="3" data-bbox="539 1283 1206 1294">Исходные данные</th> </tr> <tr> <th data-bbox="539 1294 683 1384">№</th> <th data-bbox="683 1294 970 1384">Выручка, тыс. руб.</th> <th data-bbox="970 1294 1206 1384">Число торговых точек</th> </tr> </thead> <tbody> <tr><td>1</td><td>1598</td><td>2</td></tr> <tr><td>2</td><td>2644</td><td>5</td></tr> <tr><td>3</td><td>2197</td><td>4</td></tr> <tr><td>4</td><td>1959</td><td>5</td></tr> <tr><td>5</td><td>1052</td><td>3</td></tr> <tr><td>6</td><td>1922</td><td>3</td></tr> <tr><td>7</td><td>2385</td><td>5</td></tr> <tr><td>8</td><td>2581</td><td>5</td></tr> <tr><td>9</td><td>3105</td><td>5</td></tr> <tr><td>10</td><td>3896</td><td>4</td></tr> <tr><td>11</td><td>1510</td><td>4</td></tr> <tr><td>12</td><td>1880</td><td>2</td></tr> <tr><td>13</td><td>3620</td><td>5</td></tr> <tr><td>14</td><td>5002</td><td>6</td></tr> <tr><td>15</td><td>2819</td><td>5</td></tr> </tbody> </table>	Исходные данные			№	Выручка, тыс. руб.	Число торговых точек	1	1598	2	2	2644	5	3	2197	4	4	1959	5	5	1052	3	6	1922	3	7	2385	5	8	2581	5	9	3105	5	10	3896	4	11	1510	4	12	1880	2	13	3620	5	14	5002	6	15	2819	5
Исходные данные																																																				
№	Выручка, тыс. руб.	Число торговых точек																																																		
1	1598	2																																																		
2	2644	5																																																		
3	2197	4																																																		
4	1959	5																																																		
5	1052	3																																																		
6	1922	3																																																		
7	2385	5																																																		
8	2581	5																																																		
9	3105	5																																																		
10	3896	4																																																		
11	1510	4																																																		
12	1880	2																																																		
13	3620	5																																																		
14	5002	6																																																		
15	2819	5																																																		