

УТВЕРЖДАЮ
Первый проректор, проректор по
УР

_____ А.Е. Рудин

«30» 06 2020 года

Рабочая программа дисциплины

Б1.В.ДВ.03.01 Технологии анализа больших данных

Учебный план: ФГОС 3++2020-2021_09.03.03_ИИТА_ОО_ИТ-решения.plx

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

Направление подготовки:
(специальность) 09.03.03 Прикладная информатика

Профиль подготовки: ИТ-решения и обеспечение аналитических бизнес-процессов
(специализация)

Уровень образования: бакалавриат

Форма обучения: очная

План учебного процесса

Семестр (курс для ЗАО)		Контактная работа обучающихся		Сам. работа	Контроль, час.	Трудоёмкость, 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:

- Освоить методологии обработки больших данных;
- Изучить методами и технологиями работы с данными и знаниями;
- Сформировать представление о структуризации данных и технологиях Knowledge Discovery in Databases и Data Mining;
- Освоить основы 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
- Облачные и виртуальные технологии
- Прикладная статистика
- Аналитика бизнес-процессов

2 КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ

ПКп-1: Способен проводить обследование организаций, выявлять информационные потребности пользователей
Знать: 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.
Уметь: Разрабатывать программный код для эффективной обработки распределенных данных большого объема с использованием языков программирования; Использовать программные средства визуализации и интерактивного исследования больших данных.
Владеть: Программными средствами автоматической загрузки и обработки больших данных; Навыками работы с системой интерактивного исследования и визуализации больших данных.
ПКп-8: Способен организовывать работы по созданию и редактированию контента
Знать: Основные законы естественнонаучных дисциплин при тестировании программного обеспечения с учетом обработки больших объемов данных.
Уметь: Строить управляющий граф программы для тестирования, на основании использования основных законов естественнонаучных дисциплин, и современные информационно - коммуникационные технологии в профессиональной деятельности.
Владеть: Навыками разработки эффективных наборов тестов для простых и крупных информационных систем.

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					O
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	O

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	
Всего контактная работа и СР по дисциплине	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
ПКп-1	Формулирует основные понятия, терминологию и особенности языков программирования в области технологий обработки и анализа больших данных.	Вопросы устного собеседования
	Строит алгоритм обработки больших данных на основе языков программирования.	Практико-ориентированные задания
	Демонстрирует примеры обработки данных с помощью программных средств.	Практико-ориентированные задания
ПКп-8	Раскрывает сущности законов естественнонаучных дисциплин для обработки больших объемов данных с использованием программных средств.	Вопросы устного собеседования
	Моделирует управляющий граф для тестирования программы по обработке больших объемов данных.	Вопросы устного собеседования
	Формулирует основные принципы разработки эффективных наборов тестов для простых и крупных информационных систем.	Вопросы устного собеседования

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. Учитываются баллы, накопленные в течение семестра.	Качество исполнения всех элементов практико-ориентированного задания полностью соответствует всем требованиям. Учитываются баллы, накопленные в течение семестра.
4 (хорошо)	Ответ полный, основанный на проработке всех обязательных источников информации. Подход к материалу ответственный, но стандартный. Учитываются баллы, накопленные в течение семестра.	Практико-ориентированное задание выполнено в достаточном объеме, но ограничивается только основными подходами. Учитываются баллы, накопленные в течение семестра.
3 (удовлетворительно)	Ответ стандартный, в целом качественный, основан на всех обязательных источниках информации. Присутствуют небольшие пробелы в знаниях или несущественные ошибки. Учитываются баллы, накопленные в течение семестра	Практико-ориентированное задание выполнено в соответствии с заданием. Имеются отдельные несущественные ошибки или отступления от правил оформления работы. Учитываются баллы, накопленные в течение семестра.
2 (неудовлетворительно)	Неспособность ответить на вопрос без помощи экзаменатора. Незнание значительной части принципиально важных элементов дисциплины. Многочисленные грубые ошибки. Не учитываются баллы, накопленные в течение семестра.	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
6.1.1 Основная учебная литература				
Мельниченко А. С.	Математическая статистика и анализ данных	Москва: Издательский Дом МИСиС	2018	http://www.iprbookshop.ru/78563.html
Адлер Ю. П., Черных Е. А.	Статистическое управление процессами. «Большие данные»	Москва: Издательский Дом МИСиС	2016	http://www.iprbookshop.ru/64199.html
Пальмов С. В.	Интеллектуальный анализ данных	Самара: Поволжский государственный университет телекоммуникаций и информатики	2017	http://www.iprbookshop.ru/75376.html
6.1.2 Дополнительная учебная литература				
Федин Ф. О., Федин Ф. Ф.	Анализ данных. Часть 1. Подготовка данных к анализу	Москва: Московский городской педагогический университет	2012	http://www.iprbookshop.ru/26444.html
Федин Ф. О., Федин Ф. Ф.	Анализ данных. Часть 2. Инструменты Data Mining	Москва: Московский городской педагогический университет	2012	http://www.iprbookshop.ru/26445.html
Спицкий С. В.	Эффективная аудиторная и самостоятельная работа обучающихся	СПб.: СПбГУПТД	2015	http://publish.sutd.ru/tp_ext_inf_publish.php?id=2015811
Караулова И. Б., Мелешкова Г. И., Новоселов Г. А.	Организация самостоятельной работы обучающихся	СПб.: СПбГУПТД	2014	http://publish.sutd.ru/tp_ext_inf_publish.php?id=2014550
Жуковский О. И.	Информационные технологии и анализ данных	Томск: Томский государственный университет систем управления и радиоэлектроники, Эль Контент	2014	http://www.iprbookshop.ru/72106.html
Шифф В. К.	Корреляционный, регрессионный и дисперсионный анализ	СПб.: СПбГУПТД	2019	http://publish.sutd.ru/tp_ext_inf_publish.php?id=201950

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
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): ИТ-решения и обеспечение аналитических бизнес-процессов

5.2.3 Типовые практико-ориентированные задания (задачи, кейсы)

№ п/п	Условия типовых практико-ориентированных заданий (задач, кейсов)																																																	
Семестр 7																																																		
1	<p>Провести регрессионный анализ зависимости выручки от числа торговых точек.</p> <p style="text-align: right;">Таблица 1</p> <p style="text-align: center;">Исходные данные</p> <table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse;"> <thead> <tr> <th style="width: 10%;">№</th> <th style="width: 40%;">Выручка, тыс. руб.</th> <th style="width: 50%;">Число торговых точек</th> </tr> </thead> <tbody> <tr><td style="text-align: center;">1</td><td style="text-align: center;">1598</td><td style="text-align: center;">2</td></tr> <tr><td style="text-align: center;">2</td><td style="text-align: center;">2644</td><td style="text-align: center;">5</td></tr> <tr><td style="text-align: center;">3</td><td style="text-align: center;">2197</td><td style="text-align: center;">4</td></tr> <tr><td style="text-align: center;">4</td><td style="text-align: center;">1959</td><td style="text-align: center;">5</td></tr> <tr><td style="text-align: center;">5</td><td style="text-align: center;">1052</td><td style="text-align: center;">3</td></tr> <tr><td style="text-align: center;">6</td><td style="text-align: center;">1922</td><td style="text-align: center;">3</td></tr> <tr><td style="text-align: center;">7</td><td style="text-align: center;">2385</td><td style="text-align: center;">5</td></tr> <tr><td style="text-align: center;">8</td><td style="text-align: center;">2581</td><td style="text-align: center;">5</td></tr> <tr><td style="text-align: center;">9</td><td style="text-align: center;">3105</td><td style="text-align: center;">5</td></tr> <tr><td style="text-align: center;">10</td><td style="text-align: center;">3896</td><td style="text-align: center;">4</td></tr> <tr><td style="text-align: center;">11</td><td style="text-align: center;">1510</td><td style="text-align: center;">4</td></tr> <tr><td style="text-align: center;">12</td><td style="text-align: center;">1880</td><td style="text-align: center;">2</td></tr> <tr><td style="text-align: center;">13</td><td style="text-align: center;">3620</td><td style="text-align: center;">5</td></tr> <tr><td style="text-align: center;">14</td><td style="text-align: center;">5002</td><td style="text-align: center;">6</td></tr> <tr><td style="text-align: center;">15</td><td style="text-align: center;">2819</td><td style="text-align: center;">5</td></tr> </tbody> </table>	№	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	
№	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 style="text-align: right;">Таблица 1</p> <p style="text-align: center;">Исходные данные</p> <table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse;"> <thead> <tr> <th style="width: 10%;">№</th> <th style="width: 40%;">Выручка, тыс. руб.</th> <th style="width: 50%;">Число торговых точек</th> </tr> </thead> <tbody> <tr><td style="text-align: center;">1</td><td style="text-align: center;">1598</td><td style="text-align: center;">2</td></tr> <tr><td style="text-align: center;">2</td><td style="text-align: center;">2644</td><td style="text-align: center;">5</td></tr> <tr><td style="text-align: center;">3</td><td style="text-align: center;">2197</td><td style="text-align: center;">4</td></tr> <tr><td style="text-align: center;">4</td><td style="text-align: center;">1959</td><td style="text-align: center;">5</td></tr> <tr><td style="text-align: center;">5</td><td style="text-align: center;">1052</td><td style="text-align: center;">3</td></tr> <tr><td style="text-align: center;">6</td><td style="text-align: center;">1922</td><td style="text-align: center;">3</td></tr> <tr><td style="text-align: center;">7</td><td style="text-align: center;">2385</td><td style="text-align: center;">5</td></tr> <tr><td style="text-align: center;">8</td><td style="text-align: center;">2581</td><td style="text-align: center;">5</td></tr> <tr><td style="text-align: center;">9</td><td style="text-align: center;">3105</td><td style="text-align: center;">5</td></tr> <tr><td style="text-align: center;">10</td><td style="text-align: center;">3896</td><td style="text-align: center;">4</td></tr> <tr><td style="text-align: center;">11</td><td style="text-align: center;">1510</td><td style="text-align: center;">4</td></tr> <tr><td style="text-align: center;">12</td><td style="text-align: center;">1880</td><td style="text-align: center;">2</td></tr> <tr><td style="text-align: center;">13</td><td style="text-align: center;">3620</td><td style="text-align: center;">5</td></tr> <tr><td style="text-align: center;">14</td><td style="text-align: center;">5002</td><td style="text-align: center;">6</td></tr> <tr><td style="text-align: center;">15</td><td style="text-align: center;">2819</td><td style="text-align: center;">5</td></tr> </tbody> </table>	№	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	
№	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																																																