

Минобрнауки России

Федеральное государственное бюджетное образовательное учреждение
высшего образования

«Оренбургский государственный университет»

Кафедра математических методов и моделей в экономике

РАБОЧАЯ ПРОГРАММА

ДИСЦИПЛИНЫ

«Б1.Д.В.Э.4.2 Технологии обработки больших данных»

Уровень высшего образования

БАКАЛАВРИАТ

Направление подготовки

01.03.04 Прикладная математика
(код и наименование направления подготовки)

Математическое и компьютерное моделирование
(наименование направленности (профиля) образовательной программы)

Квалификация

Бакалавр

Форма обучения

Очная

Год набора 2023

Рабочая программа дисциплины «Б1.Д.В.Э.4.2 Технологии обработки больших данных» рассмотрена и утверждена на заседании кафедры

Кафедра математических методов и моделей в экономике
наименование кафедры

протокол № 8 от "6" 02 2023 г.

Заведующий кафедрой

Кафедра математических методов и моделей в экономике
наименование кафедры



Н.П. Фот

подпись

расшифровка подписи

Исполнители:

Доцент кафедры ММиМЭ

должность



подпись

А.В. Раменская

расшифровка подписи

должность

подпись

расшифровка подписи

СОГЛАСОВАНО:

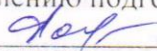
Председатель методической комиссии по направлению подготовки

01.03.04 Прикладная математика

код наименование

личная подпись

расшифровка подписи



Н.П. Фот

Заведующий отделом формирования фонда и научной обработки документов

личная подпись

расшифровка подписи

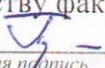


Уполномоченный по качеству факультета

личная подпись

Н.А. Тычинина

расшифровка подписи



№ регистрации _____

© Раменская А.В., 2023

© ОГУ, 2023

1 Цели и задачи освоения дисциплины

Цель (цели) освоения дисциплины:

теоретическая и практическая подготовка обучающихся к работе с большими данными. Знания, полученные в результате освоения дисциплины, помогут при сборе и анализе больших объемов структурированной или неструктурированной информации, при разработке моделей данных и извлечении из них новых знаний в процессе решения задач практической и научно-исследовательской деятельности

Задачи:

- сформировать теоретические знания о технологиях подготовки, хранения, обработки и анализа больших данных;
- сформировать теоретические знания методах обработки БД с использованием Python;
- сформировать теоретические знания о методах визуализации данных, о визуальной аналитике и дата-сторителлингу;
- сформировать навыки работы с программными продуктами, применяемыми для анализа, обработки, визуализации больших данных;
- сформировать навыки создания интерактивных дэшбордов, визуальной аналитике и дата-сторителлингу с использованием современных инструментальных средств.

2 Место дисциплины в структуре образовательной программы

Дисциплина относится к дисциплинам (модулям) по выбору вариативной части блока Д «Дисциплины (модули)»

Пререквизиты дисциплины: *Б1.Д.Б.13 Информационные технологии и программирование, Б1.Д.Б.25 Математическая статистика и анализ данных, Б1.Д.Б.28 Программирование, Б1.Д.Б.29 Объектно-ориентированный анализ и программирование, Б1.Д.Б.30 Базы данных и системы управления базами данных, Б1.Д.В.7 Технологии разработки программного обеспечения*

Постреквизиты дисциплины: *Отсутствуют*

3 Требования к результатам обучения по дисциплине

Процесс изучения дисциплины направлен на формирование следующих результатов обучения

Код и наименование формируемых компетенций	Код и наименование индикатора достижения компетенции	Планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенций
ПК*-3 Способен осуществлять проектирование, разработку, отладку, тестирование, документирование наукоемкого программного обеспечения, принимать участие в управлении проектами создания информационных систем и программных комплексов на различных стадиях жизненного цикла	ПК*-3-В-1 Применяет знания современных технологий проектирования, разработки, отладки, тестирования, документирования наукоемкого программного обеспечения	Знать: современные технологии проектирования, разработки, отладки, тестирования прикладных программ для обработки больших данных Уметь: применять современные технологии разработки программ и ИКТ для обработки больших данных Владеть: навыками применения современных технологий разработки программ, программных средств и ИКТ для обработки больших данных
	ПК*-3-В-2 Использует современные	Знать: особенности технологии хранения и обработки больших данных,

Код и наименование формируемых компетенций	Код и наименование индикатора достижения компетенции	Планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенций
	информационные технологии для сопровождения этапов жизненного цикла информационных систем и программных комплексов	этапы жизненного цикла информационных систем по работе с большими данными Уметь: применять современные технологии разработки программ и ИКТ для сопровождения этапов их жизненного цикла Владеть: навыками применения современных технологий разработки программ, программных средств и ИКТ для обработки больших данных и сопровождения этапов жизненного цикла ПО (проектирования, разработки, отладки, тестирования, документирования)
ПК*-4 Способен использовать знания современных языков программирования, стандартных пакетов прикладных программ, информационно-телекоммуникационной сети "Интернет", инструментальных средств анализа данных при решении практических задач управления информацией	ПК*-4-В-1 Применяет знания современных языков программирования при решении практических задач на ЭВМ	Знать: современные методы сбора информации, инструментарий и технологии обработки БД на языке Python, методы анализа и визуализации информации при решении исследовательских задач моделирования процессов и систем на ЭВМ. Уметь: применять современные методы сбора информации, инструментарий и технологии обработки БД на языке Python, методы анализа и визуализации информации при решении исследовательских задач моделирования процессов и систем на ЭВМ Владеть: навыками применения современных методов сбора информации, инструментария и технологий обработки БД на языке Python, методов анализа и инструментов визуализации информации при решении исследовательских задач моделирования процессов и систем на ЭВМ
	ПК*-4-В-2 Применяет знания стандартных и специализированных ППП, информационно-телекоммуникационных технологий для решения практических задач анализа данных, математического и компьютерного моделирования процессов и систем	Знать: современные методы сбора информации, инструментарий и технологии обработки БД в стандартных средствах майнинга (RapidMiner), Power BI, методы анализа и визуализации информации при решении исследовательских задач моделирования процессов и систем на ЭВМ. Уметь: применять современные методы сбора информации,

Код и наименование формируемых компетенций	Код и наименование индикатора достижения компетенции	Планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенций
		инструментарий и технологии обработки БД в стандартных средствах майнинга (RapidMiner), Power BI, методы анализа и визуализации информации при решении исследовательских задач моделирования процессов и систем на ЭВМ Владеть: навыками применения современных методов сбора информации, инструментария и технологий обработки БД в стандартных средствах майнинга (RapidMiner), Power BI, методов анализа и инструментов визуализации информации при решении исследовательских задач моделирования процессов и систем на ЭВМ
	ПК*-4-В-3 Применяет знания стандартных ППП, информационно-телекоммуникационных технологий, сети "Интернет" для поиска, обработки и систематизации информации, оформления отчетов по результатам выполненных исследований, представления их в виде презентаций, докладов	Знать: современные методы сбора информации, инструментарий и технологии обработки Уметь: оформлять результаты работы в виде презентаций, докладов, отчетов. Владеть: навыками оформления результатов работы в виде презентаций, докладов, отчетов.

4 Структура и содержание дисциплины

4.1 Структура дисциплины

Общая трудоемкость дисциплины составляет 5 зачетных единиц (180 академических часов).

Вид работы	Трудоемкость, академических часов		
	7 семестр	8 семестр	всего
Общая трудоёмкость	108	72	180
Контактная работа:	26,25	21,25	47,5
Лекции (Л)	14	10	24
Лабораторные работы (ЛР)	12	10	22
Консультации		1	1
Промежуточная аттестация (зачет, экзамен)	0,25	0,25	0,5
Самостоятельная работа: - выполнение индивидуального творческого задания (ИТЗ); - самоподготовка (проработка и повторение лекционного материала и материала учебников и учебных пособий;	81,75	50,75	132,5

Вид работы	Трудоемкость, академических часов		
	7 семестр	8 семестр	всего
- подготовка к лабораторным занятиям.)			
Вид итогового контроля (зачет, экзамен, дифференцированный зачет)	диф. зач.	экзамен	

Разделы дисциплины, изучаемые в 7 семестре

№ раздела	Наименование разделов	Количество часов				
		всего	аудиторная работа			внеауд. работа
			Л	ПЗ	ЛР	
1	Основы систем больших данных		2		-	
2	Современные инструменты работы с большими данными		4		4	
3	Базы и банки данных		2		2	
4	Современные языки программирования для обработки больших данных		6		6	
	Итого:	108	14		12	82

Разделы дисциплины, изучаемые в 8 семестре

№ раздела	Наименование разделов	Количество часов				
		всего	аудиторная работа			внеауд. работа
			Л	ПЗ	ЛР	
5	Методы обработки больших данных		4		4	
6	Введение в рекомендательные системы		2		2	
7	Инструменты визуализации больших данных		4		4	
	Итого:	72	10		10	52
	Всего:	180	24		22	134

4.2 Содержание разделов дисциплины

Раздел № 1 Основы систем больших данных: Понятие Больших данных. 3V и 7V больших данных. Стандартизация Больших данных. Особенности сбора, хранения, обработки и анализа больших массивов данных. Источники больших данных. Использование больших данных в науке, бизнесе, государственном управлении. Жизненный цикл(ы) Больших данных. Основные этапы управления Большими данными.

Раздел № 2 Современные инструменты работы с большими данными: Архитектура Больших данных. Руководство Больших данных. Open source инструменты для анализа данных. Программы для визуализации собранных данных. Сентимент-инструменты. Программы для парсинга данных в Интернете.

Раздел № 3 Базы и банки данных: Основные понятия: банк данных и база данных. История развития баз и банков данных. Обзор платформ и языков баз данных (SQL БД, SQL Hadoop, New SQL, NoSQL(графовые, столбцовые, для сложных документов, ключ-значение)).

Раздел № 4 Современные языки программирования для обработки больших данных

Обзор Библиотеки Python, применяемые для обработки и анализа БД. Обработка данных с использованием библиотеки NumPy. Обработка данных с использованием библиотеки SciPy. Стандартные функции для решения задач анализа данных в Python.

Раздел № 5 Методы обработки больших данных Методы машинного обучения Введение в data mining и machine learning. Основные задачи: классификация (Classification), кластеризация

(Clustering), ассоциация (Associations), последовательность (Sequence) или последовательная ассоциация (sequential association), прогнозирование (Forecasting), определение отклонений или выбросов (Deviation Detection), оценивание (Estimation), анализ связей (Link Analysis). Классификация и обзор методов DM и ML. Постановка задачи поиска ассоциативных правил, исходная информация для анализа. Характеристики правил: поддержка, достоверность, улучшение. Алгоритмы поиска ассоциативных правил. Определение последовательных шаблонов, понятие последовательности. Клиенториентированные базы данных.

Раздел № 6 Введение в рекомендательные системы Понятие рекомендательной системы. Подходы к построению рекомендательных систем. Использование в web-аналитики.

Раздел № 7 Визуализация больших данных

Основы визуализации данных, принципы эффективного представления данных с учётом психологии восприятия, методы визуальной аналитики, создания интерактивных дэшбордов, датасторителлинга. Инструментарий для визуализации больших данных: PowerBI, tableau и др. Отечественные системы BI.

4.3 Лабораторные работы

№ ЛР	№ раздела	Наименование лабораторных работ	Кол-во часов
1	1	Современные инструменты работы с большими данными	4
2	2	Работа с базами данных	2
3	2	Применение стандартных библиотек NumPy, SciPy для решения задач анализа данных в Python	6
4	3	Построение ассоциативных правил и последовательных шаблонов с использованием специализированных ППП.	4
5			2
6	4	Построение дашбордов	4
		Итого:	22

5 Учебно-методическое обеспечение дисциплины

5.1 Основная литература

1. Программные системы статистического анализа: обнаружение закономерностей в данных с использованием системы R и языка Python [Электронный ресурс] – Новосибирск : Новосибирский государственный технический университет, 2017. – 74 с. – Режим доступа: <https://biblioclub.ru/index.php?page=book&id=576496>
2. Шелудько, В. М. Язык программирования высокого уровня Python: функции, структуры данных, дополнительные модули [Электронный ресурс] – Ростов-на-Дону ; Таганрог : Южный федеральный университет, 2017. – 108 с. – Режим доступа: <https://biblioclub.ru/index.php?page=book&id=500060>

5.2 Дополнительная литература

1. Марков, А. С. Базы данных : введение в теорию и методологию / А.С. Марков, К.Ю. Лисовский. - М. : Финансы и статистика, 2006. - 512 с.
2. Советов, Б. Я. Базы данных : учебник для прикладного бакалавриата / Б. Я. Советов, В.В. Цехановский, В. Д. Чертовской; С.-Петерб. гос. электротехн. ун-т "ЛЭТИ" им. В. И. Ульянова (Ленина).- 2-е изд. - Москва : Юрайт, 2015. - 463 с.

Методические указания

1. Стебунова, О. И. Среда разработки "RStudio" в эконометрическом моделировании [Электронный ресурс] : методические указания для обучающихся по образовательным программам высшего образования по направлениям подготовки 01.03.04 Прикладная математика, 38.03.05 Бизнес-информатика, 38.04.01 Экономика / О. И. Стебунова, К. В. Пивоварова; М-во образования и науки Рос. Федерации, Федер. гос. бюджет. образоват. учреждение высш. образования "Оренбург. гос. ун-т", Каф. мат. методов и моделей в экономике. - Электрон. текстовые дан. (1 файл: 1.20 Мб). - Оренбург : ОГУ, 2018. - 37 с. – Режим доступа: http://artlib.osu.ru/web/books/metod_all/78666_20180705.pdf

5.3 Периодические издания

1. Информационные технологии : журнал. - М. : Агентство "Роспечать", 2022, 2023.
2. Вестник компьютерных и информационных технологий : журнал. - М. : Агентство "Роспечать", 2022, 2023

5.4 Интернет-ресурсы

<http://www.machinelearning.ru> – профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных;
<https://r-analytics.blogspot.com> – R: Анализ и визуализация данных;
<https://yandexdataschool.ru/edu-process/courses/machine-learning> – видеолекции курса «Машинное обучение»;
<https://www.intuit.ru/studies/courses/6/6/info> Национальный Открытый Университет «ИНТУИТ», онлайн-курс «Data Mining» (Автор: И. Чубукова)
<https://rapidminer.com/> сайт платформы RapidMiner, включающий в учебные материалы Academy RapidMiner
<http://www.basegroup.ru/> сайт разработчика аналитической платформы Deductor
<https://scholarspace.manoa.hawaii.edu/bitstream/10125/41281/1/paper0132.pdf> – Faizullov, I., Yablonsky, S. Modern Advanced Analytics Platforms and Predictive Models for Stock Price Forecasting: IBM Watson Analytics Case // In: The Proceedings of the Hawaii International Conference on System Sciences HICSS-50: January 4-7, 2017.
https://docs.microsoft.com/ru-ru/learn/browse/?products=power-bi&WT.mc_id=sitertzn_learntab_mslearn-banner-powerbi Работа с MS Power BI обучающие ресурсы
https://needfordata.ru/manual_power_bi MS Power BI полезные материалы: учебники, коннекторы, ссылки и другое

5.5 Программное обеспечение, профессиональные базы данных и информационные справочные системы

1. Операционная система РЕД ОС
2. Пакет офисных приложений LibreOffice
3. Программная система для организации видео-конференц-связи Webinar.ru
4. Python 3.* - язык программирования с открытым исходным кодом. Доступен бесплатно. Режим доступа: <https://www.python.org/downloads/>
5. Свободная среда разработки программного обеспечения с открытым исходным кодом для языка программирования Python. PyCharm Community Edition. Доступна бесплатно. Режим доступа: <https://www.jetbrains.com/ru-ru/pycharm/download/#section=windows>
6. Консультант Плюс [Электронный ресурс] : справочно-правовая система / Компания Консультант Плюс. – Электрон. дан. – Москва. – Режим доступа : в локальной сети ОГУ <\\fileserver1\\CONSULT\\cons.exe>

7. Гарант [Электронный ресурс] : справочно-правовая система / НПП Гарант-Сервис. – Электрон. дан. - Москва. – Режим доступа <\\fileserv1\GarantClient\garant.exe> в локальной сети ОГУ.

6 Материально-техническое обеспечение дисциплины

Учебные аудитории для проведения занятий лекционного типа, семинарского типа, для проведения групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации.

Аудитории оснащены комплектами ученической мебели, техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Для проведения лабораторных занятий используется лаборатория компьютерный класс, оснащенный персональными компьютерами.

Помещение для самостоятельной работы обучающихся оснащены компьютерной техникой, подключенной к сети "Интернет", и обеспечением доступа в электронную информационно-образовательную среду ОГУ.