Минобрнауки России

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Оренбургский государственный университет имени В.А. Бондаренко»

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 2026

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

Кафедра математических методов и моделей в экономике
наименование кафедры

протокол № 11 от "03" марта 2026 г.

Заведующий кафедрой
Кафедра математических методов и моделей в экономике _____ Н.П. Фот
наименование кафедры *подпись* *расшифровка подписи*

Исполнители:
Доцент кафедры ММиМЭ _____ Л.М. Туктамышева
должность *подпись* *расшифровка подписи*

СОГЛАСОВАНО:
Председатель методической комиссии по направлению подготовки
01.03.04 Прикладная математика _____ Н.П. Фот
код наименование *личная подпись* *расшифровка подписи*

Заведующий отделом формирования фонда и научной обработки документов
_____ С.А. Бихтимирова
личная подпись *расшифровка подписи*

Уполномоченный по качеству института
_____ С.Н. Морозова
личная подпись *расшифровка подписи*

№ регистрации 190801

© Туктамышева Л.М., 2026
© ОГУ, 2026

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 навыки использования языка программирования Python для обработки больших данных;

– сформировать навыки работы со специальными подходами и инструментами, позволяющими хранить и обрабатывать огромные объёмы данных.

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.24 Математическая статистика и анализ данных, Б1.Д.Б.27 Программирование, Б1.Д.Б.28 Объектно-ориентированный анализ и программирование, Б1.Д.Б.29 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, откладку, 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, откладки, тестирования, документирования наукоемкого программного обеспечения ПК*-3-В-2 Использует современные информационные технологии для сопровождения этапов жизненного цикла информационных систем и программных комплексов	**Знать:** базовые средства разработки, отладки, тестирования программ и ИС для хранения, обработки и анализа больших данных **Уметь:** применять программные средства и ИКТ для сопровождения этапов жизненного цикла ПО (разработки, отладки, тестирования, документирования ПО), используемого для хранения, обработки и анализа больших данных **Владеть:** навыками применения программных средств и ИКТ для сопровождения этапов жизненного цикла ПО (разработки, отладки, тестирования, документирования ПО) для хранения, обработки и анализа больших данных

Код и наименование формируемых компетенций	Код и наименование индикатора достижения компетенции	Планируемые результаты обучения по дисциплине, характеризующие этапы формирования компетенций
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 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 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:** источники информации и способы поиска информации, обладающего свойствами больших данных для решения практических задач анализа данных, математического и компьютерного моделирования процессов и систем; реализацию методов хранения и обработки больших данных в специализированных программных комплексах, фреймворках, а также языка программирования Python; структуру отчета, доклада, презентации результатов исследования и стандарты их оформления; **Уметь:** осуществлять поиск, анализ и структурирование информации представленного большими данными для исследования, направленного на решение практических задач анализа данных, математического и компьютерного моделирования процессов и систем; пользоваться открытыми источниками данных для формирования информационной базы исследования в форме больших данных; проводить сравнительный анализ источников информации по указанной тематике и осуществлять выбор наиболее релевантных источников информации; применять программные комплексы, специализированные фреймворки, а также языка программирования Python для решения задач в сфере моделирования и прогнозирования на основе больших данных; оформлять и представлять результаты исследования в виде отчета, доклада, презентации **Владеть:** навыками поиска, анализа и структурирования информации представленного большими данными для исследования, направленного на решение практических задач анализа данных, математического и компьютерного моделирования процессов и систем; навыками использования открытых источников данных для формирования информационной базы исследования; навыками сравнительного анализа источников информации по указанной тематике и выбора наиболее релевантных источников информации; навыками применения программных комплексов, специализированных фреймворков, языка программирования Python для решения задач в сфере моделирования и прогнозирования на основе больших данных, навыками оформления и представления результатов исследования в виде отчета, доклада, презентации.

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:	**34,25**	**13,25**	**47,5**
Лекции (Л)	18	6	24
Лабораторные работы (ЛР)	16	6	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; - изучение разделов курса в системе электронного обучения; - подготовка к лабораторным занятиям; - подготовка к коллоквиумам; - подготовка к рубежному контролю)	**73,75**	**58,75**	**132,5**
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	Введение в большие данные	14	2		-	12
2	Способы хранения больших данных	26	4		2	20
3	Технологии обработки больших данных: пакетная обработка	34	6		6	22
4	Технологии обработки больших данных: потоковая обработка	34	6		8	20
	Итого:	108	18		16	74

Разделы дисциплины, изучаемые в 8 семестре

№ раздела	Наименование разделов	Количество часов				
		всего	аудиторная работа			внеауд. работа
			Л	ПЗ	ЛР	
5	Язык Python для обработки больших данных	32	2		2	28
6	Машинное обучение в обработке больших данных	40	4		4	32
	Итого:	72	6		6	60
	Всего:	180	24		22	134

4.2 Содержание разделов дисциплины

1 Введение в большие данные. *Основы систем больших данных: Понятие Больших данных. Характеристики больших данных: скорость, ценность, объем, ценность, достоверность и др. Источники больших данных. Использование больших данных в науке, бизнесе, государственном управлении. Жизненный цикл(ы) Больших данных. Основные этапы управления Большими данными.*

2 Способы хранения больших данных. *Типы хранилищ больших данных. Data Lake (Озеро данных). Data Warehouse (Хранилище данных). Файловое, объектное, блочное хранилище. Распределенные системы хранения. HDFS. Методы хранения: локальное, облачное хранение, гибридные решения. OLAP-кубы.*

3 Технологии обработки больших данных: пакетная обработка *Модель программирования MapReduce. Hadoop и его экосистема. Другие технологии и системы для пакетной обработки больших данных (Apache Hive, Apache Pig и др.)*

4 Технологии обработки больших данных: потоковая обработка *Экосистема Apache Kafka. Платформа контейнеризации Docker. Другие фреймворки для потоковой обработки данных. Гибридные варианты, поддерживающие пакетную и потоковую обработку данных: Apache Spark, Apache Flink, Apache Beam и др.*

5 Язык Python для обработки больших данных. *Обработка данных с использованием библиотеки NumPy. Обработка данных с использованием библиотеки SciPy. PySpark, Vaex , Modin, Dask. Стандартные функции для решения задач анализа данных в Python. Обзор дистрибутивов и оболочек для работы с Python: Anaconda, Jupyter Notebooks, Spyder, PyScripter и др.*

6 Машинное обучение в обработке 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.*

4.3 Лабораторные работы

№ ЛР	№ раздела	Наименование лабораторных работ	Кол-во часов
1	2	Поиск и формирование набора данных	2
2	3	Использование Hadoop для аналитики больших данных	6
3	4	Работа с контейнерами Docker	2
4	4	Отправка сообщений в Apache Kafka	2
5	4	Работа с Apache Spark	4
6	5	PySpark для обработки больших данных	2
7	6	Решение задачи кластеризации и классификации по большим данным	4
		Итого:	22

5 Учебно-методическое обеспечение дисциплины

5.1 Основная литература

1 Методы интеллектуального анализа данных. Задача классификации [Электронный ресурс] : учебное пособие для обучающихся по образовательной программе высшего образования по направлению подготовки 01.03.04 Прикладная математика / О. С. Чудинова [и др.]; М-во науки и высш. образования Рос. Федерации, Федер. гос. бюджет. образоват. учреждение высш. образования "Оренбург. гос. ун-т". - Электрон. дан. - Оренбург : ОГУ, 2024. - 1 электрон. опт. диск (CD-R). - Загл. с этикетки диска. - Систем. требования: Intel Core или аналогичн.; Microsoft Windows 7, 8, 10 ; 512 Мб ; монитор, поддерживающий режим 1024x768 ; мышь или аналогич. устройство. - ISBN 978-5-7410-3284-8. - гос. регистрации 0322402552.

2 Ланских, Ю. В. Введение в большие данные : учебное пособие / Ю. В. Ланских, В. Г. Ланских, К. В. Родионов. — Киров : ВятГУ, 2023. — 172 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/408566 (дата обращения: 29.04.2026).

5.2 Дополнительная литература

Лебедев, А. С. Методы Big Data : учебно-методическое пособие / А. С. Лебедев, Ш. Г. Магомедов. — Москва : РТУ МИРЭА, 2021. — 91 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/182452»

5.3 Периодические издания

Информационные технологии в проектировании и производстве: журнал. - М. : ФГУП «НТЦ оборонного комплекса «Компас». 2019-2026.

5.4 Интернет-ресурсы

1. https://intuit.ru/studies/courses/12385/1181/info - Национальный Открытый Университет «ИНТУИТ», онлайн-курс «Введение в аналитику больших массивов данных»
2. https://openedu.ru/course/hse/big_data_processing/ - онлайн-курс «Обработка больших данных»
3. https://stepik.org/course/150/info - - онлайн-курс «Hadoop. Система для обработки больших объемов данных
4. https://openedu.ru/course/spbstu/BIGDATA/?session=spring_2026#about «Открытое образование», Каталог курсов, МООК: «Наука о данных и аналитика больших объемов данных»

5.5 Программное обеспечение, профессиональные базы данных и информационные справочные системы

1. Операционная система РЕД ОС.
2. Пакет офисных приложений LibreOffice
3. Для работы с ресурсами Интернет - веб-браузер Яндекс https://yandex.ru/
4. ГАРАНТ Платформа F1 [Электронный ресурс]: справочно-правовая система. / Разработчик ООО НПП «ГАРАНТ-Сервис», 119992, Москва, Воробьевы горы, МГУ, [1990–2026]. – Режим доступа в сети ОГУ http://garant.net.osu.ru
5. Автоматизированная интерактивная система сетевого тестирования - АИССТ (зарегистрирована в РОСПАТЕНТ, Свидетельство о государственной регистрации программы для ЭВМ №2011610456, правообладатель – Оренбургский государственный университет), режим доступа - http://aist.osu.ru.

6. Университетская платформа электронного обучения «Электронные курсы ОГУ в системе обучения Moodle» (http://moodle.osu.ru)

7. Python - объектно-ориентированный язык программирования. Бесплатное свободно распространяемое программное обеспечение Режим доступа: https://www.python.org

8. Apache Hadoop - свободно распространяемый набор утилит, библиотек и фреймворк для разработки и выполнения распределённых программ, работающих на кластерах из сотен и тысяч узлов. Режим доступа: https://hadoop.apache.org

6 Материально-техническое обеспечение дисциплины

Учебные аудитории для проведения занятий лекционного типа, лабораторных работ, 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 компьютерный класс, оснащенный компьютерной техникой, подключенной к сети "Интернет", и обеспечением доступа в электронную информационно-образовательную среду ОГУ.

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