

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Узунов Федор Владимирович

Должность: Ректор

Дата подписания: 24.09.2026 12:52:38

Уникальный программный ключ:

fd935d10451b860e912264c0378b448457bfdb603f94388008e39877e6bcb6f

**АВТОНОМНАЯ НЕКОММЕРЧЕСКАЯ ОРГАНИЗАЦИЯ
«ОБРАЗОВАТЕЛЬНАЯ ОРГАНИЗАЦИЯ ВЫСШЕГО ОБРАЗОВАНИЯ»
«УНИВЕРСИТЕТ ЭКОНОМИКИ И УПРАВЛЕНИЯ»
Факультет экономики, управления и юриспруденции
Кафедра «Управление и бизнес-информатика»**

УТВЕРЖДАЮ

Проректор по учебно-методической работе

Г.П. Узунова / Г.П. Узунова

«02» февраля 2026 г.



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Анализ больших данных

Направление подготовки

09.03.02 «Информационные системы и технологии»

Профиль

«Специалист информационных систем и технологий»

Квалификация выпускника

Бакалавр

Для всех

форм обучения

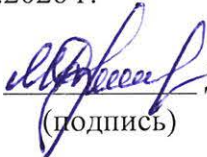
Симферополь, 2026

Рабочая программа составлена в соответствии с ФГОС ВО – бакалавриат по направлению подготовки 09.03.02 «Информационные системы и технологии», утвержденного приказом Министерства науки и высшего образования Российской Федерации от 19.09.2017 №926 (зарегистрировано в Министерстве юстиции РФ 12.10.2017 №48535) с изменениями и дополнениями.

Программу составил Фурин А.Д., преподаватель

Рабочая программа дисциплины «Анализ больших данных» утверждена на заседании кафедры «Управление и бизнес-информатика».

Протокол № 6 от 29.01.2026 г.

Заведующий кафедрой  Д.В. Моторина
(подпись)

АННОТАЦИЯ	
Индекс дисциплины по учебному плану	Наименование дисциплины
Б1.О.30	Анализ больших данных
Цель изучения дисциплины	Цель дисциплины «Анализ больших данных» заключается в формировании у студентов системных представлений о технологиях обработки информации огромного размера (Big Data) и приобретении практических навыков извлечения полезных знаний из массивов, которые невозможно обработать традиционными инструментами
Место дисциплины в структуре ОПОП	Дисциплина Анализ больших данных относится к базовой части ОПОП и является обязательной для освоения.
Компетенции, формируемые в результате освоения дисциплины	ОПК-7
Содержание дисциплины	Тема 1. Введение в Big Data: характеристики, источники и архитектура. Тема 2. Технологии хранения и пакетной обработки данных. Тема 3. Анализ данных средствами Python: библиотеки Pandas и NumPy. Тема 4. Визуализация и исследовательский анализ данных (EDA). Тема 5. Методы машинного обучения на больших данных. Тема 6. Поточковая обработка данных и NoSQL-базы.
Общая трудоемкость дисциплины	Общая трудоемкость дисциплины составляет 6 зачетных единиц (144 часов)
Форма промежуточной аттестации	Зачет с оценкой

Содержание

1. Цель и перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения программы бакалавриата	5
2. Место дисциплины в структуре ОПОП бакалавриата	5
3. Объем дисциплины в зачетных единицах с указанием количества академических часов, выделенных на контактную работу обучающихся с преподавателем (по видам занятий) и на самостоятельную работу обучающихся	5
4. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий	6
5. Контроль качества освоения дисциплины	11
6. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	11
7. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины	12
8. Методические указания для обучающихся по освоению дисциплины	13
9. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем (при необходимости)	13
10. Описание материально-технического обеспечения, необходимого для осуществления образовательного процесса по дисциплине	14
Приложение к РПД	15

1. Цель и перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения программы бакалавриата

Цель дисциплины «Анализ больших данных» заключается в формировании у студентов системных представлений о технологиях обработки информации огромного размера (Big Data) и приобретении практических навыков извлечения полезных знаний из массивов, которые невозможно обработать традиционными инструментами

В результате освоения ОПОП бакалавриата обучающийся должен овладеть следующими результатами обучения по дисциплине:

Коды компетенции	Результаты освоения ОПОП	Перечень планируемых результатов обучения по дисциплине
ОПК-7	Способен осуществлять выбор платформ и инструментальных программно-аппаратных средств для реализации информационных систем	ОПК-7.1. Знать: методы настройки, наладки программно-аппаратных комплексов. ОПК-7.2. Уметь: анализировать техническую документацию, производить настройку, наладку и тестирование программно-аппаратных комплексов. ОПК-7.3. Владеть: навыками проверки работоспособности программно-аппаратных комплексов

2. Место дисциплины в структуре ОПОП бакалавриата

Дисциплина Анализ больших данных относится к базовой части ОПОП и является обязательной для освоения, изучается обучающимися очной формы обучения в 6 семестре, очно-заочной формы обучения – в 7 семестре.

Требования к «входным» знаниям, умениям и навыкам обучающегося, необходимым для освоения дисциплины: структуры данных, алгоритмы обработки информации, знание основ структурного и процедурного программирования, умение работать с логическими операторами, умение работать с циклами, массивами, строками, файлами.

3. Объем дисциплины в зачетных единицах с указанием количества академических часов, выделенных на контактную работу обучающихся с преподавателем (по видам занятий) и на самостоятельную работу обучающихся

Общая трудоемкость (объем) дисциплины составляет 4 зачетных единицы (з.е.), 144 академических часа.

3.1. Объём дисциплины по видам учебных занятий (в часах)

Для очной формы обучения

Общая трудоёмкость дисциплины составляет 4 зачётных единицы 144 часа

Объём дисциплины	Всего часов
Общая трудоемкость дисциплины	144

Контактная работа	44
Аудиторная работа (всего):	44
Лекции	22
Семинары, практические занятия	22
Самостоятельная работа обучающихся (всего)	100
Зачет с оценкой	+

Для очно-заочной формы обучения

Общая трудоёмкость дисциплины составляет 6 зачётных единицы 216 часа

Объём дисциплины	Всего часов
Общая трудоёмкость дисциплины	144
Контактная работа	28
Аудиторная работа (всего):	28
Лекции	10
Семинары, практические занятия	18
Самостоятельная работа обучающихся (всего)	116
Зачет с оценкой	+

4. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1. Разделы дисциплины и трудоёмкость по видам учебных занятий (в академических часах)

№ темы	Наименование темы	Всего		Количество часов					
		ОФО	ОЗФО	Контактная работа				Внеаудит. работа	
				Лекции		Практические		Самост. работа	
				ОФО	ОЗФО	ОФО	ОЗФО	ОФО	ОЗФО
1.	Тема 1. Введение в Big Data: характеристики, источники и архитектура.	34	32	6	2	4	4	24	26
2.	Тема 2. Технологии хранения и пакетной обработки данных.	34	32	6	2	4	4	24	26
3.	Тема 3. Анализ данных средствами Python: библиотеки Pandas и NumPy.	30	32	6	2	2	4	22	26
4.	Тема 4. Визуализация и исследовательский анализ данных (EDA).	30	30	6	2	2	2	22	26
5.	Тема 5. Методы машинного обучения на больших данных.	26	27	2	1	2	2	22	24
6.	Тема 6. Поточковая обработка	26	27	2	1	2	2	22	24

	данных и NoSQL-базы.								
	Всего по дисциплине	144	144	22	10	22	18	100	116
	Контроль	-	-						
	Итого	144	144						

4.2. Содержание дисциплины, структурированное по темам (разделам)

Разделы, темы, дидактические единицы
Тема 1. Введение в Big Data: характеристики, источники и архитектура. Определение понятия «большие данные». Разбор характеристик 5V (Volume — объем, Velocity — скорость, Variety — многообразие, Veracity — достоверность, Value — ценность). Классификация источников данных: внутренние транзакционные системы, логи серверов, потоковые данные датчиков (IoT), социальные медиа. Обзор экосистемы Hadoop (<i>HDFS</i> , <i>YARN</i>) и архитектуры распределенных вычислений.
Тема 2. Технологии хранения и пакетной обработки данных. Принципы работы файловой системы <i>HDFS</i> . Модель программирования <i>MapReduce</i> : логика разделения задач на этапы отображения (<i>map</i>) и свертки (<i>reduce</i>). Особенности написания эффективных <i>MapReduce</i> -задач. Распределенная база данных <i>Apache Hive</i> для выполнения SQL-подобных запросов к данным в <i>HDFS</i> (концепция <i>Data Warehouse</i> поверх кластера).
Тема 3. Анализ данных средствами Python: библиотеки Pandas и NumPy. Эффективная работа со структурами данных в библиотеке <i>NumPy</i> (многомерные массивы, векторизация операций). Использование <i>Pandas</i> для манипуляций с табличными данными (<i>Series</i> и <i>DataFrame</i>): чтение из различных форматов (<i>CSV</i> , <i>JSON</i> , <i>SQL</i>), индексация, группировка (<i>groupby</i>), агрегация и объединение таблиц (<i>merge</i> , <i>join</i>). Очистка данных от пропусков и выбросов.
Тема 4. Визуализация и исследовательский анализ данных (EDA). Построение графиков для первичного анализа распределения признаков с помощью библиотек <i>Matplotlib</i> и <i>Seaborn</i> . Создание гистограмм, диаграмм рассеяния, тепловых карт корреляций и сводных статистик. Выявление аномалий и скрытых зависимостей в данных перед этапом машинного обучения. Принципы построения информативных дашбордов.
Тема 5. Методы машинного обучения на больших данных. Применение алгоритмов классификации и регрессии (линейная/логистическая регрессия, деревья решений) с использованием <i>Scikit-learn</i> . Подготовка признаков (<i>Feature Engineering</i>): масштабирование, кодирование категориальных переменных. Оценка качества моделей (метрики <i>Accuracy</i> , <i>Precision</i> , <i>Recall</i> , <i>F1-score</i> , <i>ROC-AUC</i>). Введение в технологии распределенного машинного обучения (<i>Spark MLlib</i>).
Тема 6. Потоковая обработка данных и NoSQL-базы. Основы работы с нереляционными базами данных. Сравнение типов <i>NoSQL</i> : ключ-значение (<i>Redis</i>), документы (<i>MongoDB</i>), колонки (<i>Cassandra</i>), графы (<i>Neo4j</i>). Технологии потоковой обработки событий в реальном времени. Конвейер данных на базе <i>Apache Kafka</i> (брокеры сообщений) и <i>Apache Spark Streaming</i> для онлайн-аналитики входящих потоков информации.

4.3. Содержание практических занятий (очная форма обучения)

Разделы, темы, дидактические единицы

Тема 1. Введение в Big Data: характеристики, источники и архитектура. Определение понятия «большие данные». Разбор характеристик 5V (Volume — объем, Velocity — скорость, Variety — многообразие, Veracity — достоверность, Value — ценность).
Тема 2. Технологии хранения и пакетной обработки данных. Принципы работы файловой системы <i>HDFS</i> . Модель программирования <i>MapReduce</i> : логика разделения задач на этапы отображения (map) и свертки (reduce). Особенности написания эффективных MapReduce-задач.
Тема 3. Анализ данных средствами Python: библиотеки Pandas и NumPy. Эффективная работа со структурами данных в библиотеке <i>NumPy</i> (многомерные массивы, векторизация операций).
Тема 4. Визуализация и исследовательский анализ данных (EDA). Построение графиков для первичного анализа распределения признаков с помощью библиотек <i>Matplotlib</i> и <i>Seaborn</i> . Создание гистограмм, диаграмм рассеяния, тепловых карт корреляций и сводных статистик.
Тема 5. Методы машинного обучения на больших данных. Применение алгоритмов классификации и регрессии (линейная/логистическая регрессия, деревья решений) с использованием <i>Scikit-learn</i> . Подготовка признаков (<i>Feature Engineering</i>): масштабирование, кодирование категориальных переменных.
Тема 6. Поточковая обработка данных и NoSQL-базы. Основы работы с нереляционными базами данных. Сравнение типов NoSQL: ключ-значение (<i>Redis</i>), документы (<i>MongoDB</i>), колонки (<i>Cassandra</i>), графы (<i>Neo4j</i>). Технологии потоковой обработки событий в реальном времени.

4.4. Содержание самостоятельной работы

Разделы, темы, дидактические единицы
Тема 1. Введение в Big Data: характеристики, источники и архитектура. Классификация источников данных: внутренние транзакционные системы, логи серверов, потоковые данные датчиков (IoT), социальные медиа. Обзор экосистемы Hadoop (<i>HDFS</i> , <i>YARN</i>) и архитектуры распределенных вычислений.
Тема 2. Технологии хранения и пакетной обработки данных. Распределенная база данных <i>Apache Hive</i> для выполнения SQL-подобных запросов к данным в <i>HDFS</i> (концепция <i>Data Warehouse</i> поверх кластера).
Тема 3. Анализ данных средствами Python: библиотеки Pandas и NumPy. Использование <i>Pandas</i> для манипуляций с табличными данными (<i>Series</i> и <i>DataFrame</i>): чтение из различных форматов (<i>CSV</i> , <i>JSON</i> , <i>SQL</i>), индексация, группировка (<i>groupby</i>), агрегация и объединение таблиц (<i>merge</i> , <i>join</i>). Очистка данных от пропусков и выбросов.
Тема 4. Визуализация и исследовательский анализ данных (EDA). Выявление аномалий и скрытых зависимостей в данных перед этапом машинного обучения. Принципы построения информативных дашбордов.
Тема 5. Методы машинного обучения на больших данных. Оценка качества моделей (метрики Accuracy, Precision, Recall, F1-score, ROC-AWS). Введение в технологии распределенного машинного обучения (<i>Spark MLlib</i>).
Тема 6. Поточковая обработка данных и NoSQL-базы. Основы работы с нереляционными базами данных. Конвейер данных на базе <i>Apache Kafka</i> (брокеры сообщений) и <i>Apache Spark Streaming</i> для онлайн-аналитики входящих потоков информации.

5. Контроль качества освоения дисциплины

Текущий контроль и промежуточная аттестация осуществляются в соответствии с «Положением о текущем контроле успеваемости и промежуточной аттестации обучающихся в Автономной некоммерческой организации «Образовательная организация высшего образования» «Университет экономики и управления».

Вид промежуточной аттестации – зачет с оценкой. Форма проведения промежуточной аттестации – письменный зачет.

Фонд оценочных средств по дисциплине приведен в приложении к РПД.

6. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

а) основная литература:

1. Волк, В. К. Базы данных : учебник / В. К. Волк, В. Ю. Осеев, О. С. Черепанов. — Москва, Вологда : Инфра-Инженерия, 2025. — 544 с. — ISBN 978-5-9729-2594-0. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/154413.html> (дата обращения: 17.09.2025). — Режим доступа: для авторизир. пользователей

2. Градусов А.Б Базы данных. Проектирование баз данных : учебно-практическое пособие / Градусов А.Б, Шутов А.В.. — Владимир : Издательство Владимирского государственного университета, 2024. — 259 с. — ISBN 978-5-9984-1888-4. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/152317.html> (дата обращения: 26.04.2025). — Режим доступа: для авторизир. пользователей

3. Наместников, А. М. Базы данных : учебное пособие / А. М. Наместников. — Санкт-Петербург : Санкт-Петербургский государственный архитектурно-строительный университет, ЭБС АСВ, 2024. — 100 с. — ISBN 978-5-9227-1392-4. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/148944.html> (дата обращения: 16.03.2025). — Режим доступа: для авторизир. пользователей

б) дополнительная литература:

4. Бужинская, Н. В. Базы данных : практикум / Н. В. Бужинская, Д. М. Гребнева, Е. А. Кокшарова. — Москва : Ай Пи Ар Медиа, 2025. — 177 с. — ISBN 978-5-4497-4405-0. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/151053.html> (дата обращения: 10.06.2025). — Режим доступа: для авторизир. пользователей

5. Прокушев, Я. Е. Базы данных : учебник с практикумом / Я. Е. Прокушев. — 2-е изд. — Санкт-Петербург : Интермедия, 2022. — 264 с. — ISBN 978-5-4383-0250-6. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/120171.html> (дата обращения: 14.06.2025). — Режим доступа: для авторизир. пользователей

7. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. Информационно-правовой портал «Гарант»: официальный сайт. — URL: <http://www.garant.ru> — Текст: электронный.

2. Цифровой образовательный ресурс «IPRsmart»: официальный сайт. — URL: <http://www.iprbookshop.ru>

www.iprbookshop.ru/ – Текст: электронный.

3. Научная электронная библиотека «КиберЛенинка»: официальный сайт. – URL: <https://cyberleninka.ru/> – Текст: электронный.

4. Российский интернет-портал и аналитическое агентство TAdviser: официальный сайт. – URL: <https://www.tadviser.ru/> – Текст: электронный.

8. Методические указания для обучающихся по освоению дисциплины

При проведении лекций, семинарских (практических) занятий, самостоятельной работе обучающихся применяются интерактивные формы проведения занятий с целью погружения обучающихся в реальную атмосферу профессионального сотрудничества по разрешению проблем, оптимальной выработки навыков и качеств будущего специалиста. Интерактивные формы проведения занятий предполагают обучение в сотрудничестве. Все участники образовательного процесса (преподаватель и обучающиеся) взаимодействуют друг с другом, обмениваются информацией, совместно решают проблемы, моделируют ситуацию.

В учебном процессе используются интерактивные формы занятий:

- творческое задание. Выполнение творческих заданий требует от обучающегося воспроизведение полученной ранее информации в форме, определяемой преподавателем, и требующей творческого подхода;

- групповое обсуждение. Групповое обсуждение кого-либо вопроса направлено на достижение лучшего взаимопонимания и способствует лучшему усвоению изучаемого материала.

В ходе освоения дисциплины при проведении контактных занятий используются следующие формы обучения, способствующие формированию компетенций: лекции-дискуссии; кейс-метод; решение задач; ситуационный анализ; обсуждение рефератов и докладов; разработка групповых проектов; встречи с представителями государственных и общественных организаций.

9. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем (при необходимости)

В процессе лекционных и практических занятий используется следующее программное обеспечение:

- *программы, обеспечивающие доступ в сеть «Интернет» (например, «Microsoft Edge», «Google Chrome»);

- *программы, демонстрации видео материалов (например, проигрыватель «Windows Media Player»);

- *текстовые редакторы и процессоры (например, «Microsoft Office Word»);

- *табличные процессоры (например, «Microsoft Office Excel»);

- *системы управления базами данных (например, «Microsoft Office Access»);

- *программы для демонстрации и создания презентаций (например, «Microsoft PowerPoint»);

- *проблемно-ориентированные пакеты прикладных программ по отраслям и сферам деятельности (например, «1С: Управление нашей фирмой», «Loginom Community Edition»).

10. Описание материально-технического обеспечения, необходимого для осуществления образовательного процесса по дисциплине

Учебная аудитория

Оборудование учебной аудитории:

рабочее место преподавателя ; посадочные места по количеству обучающихся ;

доска классная ;

стенды информационные ;

Учебно-наглядные пособия: компьютеры с лицензионным программным обеспечением и возможностью подключения к информационно-телекоммуникационной сети Интернет ; мультимедийная установка .

Помещение для самостоятельной работы обучающихся

оснащено компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду Университета.