

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Узунов Федор Владимирович
Должность: Ректор
Дата подписания: 13.09.2023 19:01:17
Уникальный программный ключ:
fd935d10451b860e912264c0378f8448451ba6f438608e117a6b5c5

**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

Е.В. Бебешко

«31» марта 2023 г.

Рабочая программа дисциплины

Анализ больших данных

Направление подготовки

38.03.05 Бизнес-информатика

Профиль

Специалист по информационным системам и технологиям

Квалификация выпускника

Бакалавр

Для всех

форм обучения

Симферополь 2023

АННОТАЦИЯ	
Индекс дисциплины по учебному плану	Наименование дисциплины
Б1.О.30	Анализ больших данных
Цель изучения дисциплины	Формирование у обучающихся представления об основных методах и программных средствах обработки и анализа информации для информационно-аналитической поддержки принятия управленческих решений, навыков работы с большими объемами информации.
Место дисциплины в структуре ОПОП	Дисциплина относится к обязательной части блока 1. «Дисциплины (модули)» программы бакалавриата
Компетенции, формируемые в результате освоения дисциплины	ОПК-4
Содержание дисциплины	Тема 1. Интеллектуальный анализ данных Тема 2. Сбор, хранение, обработка, извлечение и анализ больших потоков данных Тема 3. Статистическое мышление и статистический анализ Тема 4. Методы классификации и прогнозирования Тема 5. Введение в машинное обучение и разработку данных. Тема 6. Применение технологий больших данных для бизнес-анализа Тема 7. Способы визуализации, интерпретации, проектная организация и бизнес рекомендации
Общая трудоемкость дисциплины	Общая трудоемкость дисциплины составляет 4 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	10
6. Учебно-методическое обеспечение дисциплины	10
7. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	10
8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины.....	11
9. Методические указания для обучающихся по освоению дисциплины	12
10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем (при необходимости)	12
11. Описание материально-технического обеспечения, необходимого для осуществления образовательного процесса по дисциплине	11

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

Целью изучения дисциплины «Анализ больших данных» является формирование у обучающихся представления об основных методах и программных средствах обработки и анализа информации для информационно-аналитической поддержки принятия управленческих решений, навыков работы с большими объемами информации.

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-4	Способен понимать принципы работы информационных технологий; использовать информацию, методы и программные средства ее сбора, обработки и анализа для информационно-аналитической поддержки принятия управленческих решений;	ОПК-4.1 Знает методы и программные средства сбора, обработки и анализа информации; ОПК-4.2 Умеет работать с большими объемами информации, использовать методы и программные средства обработки и анализа информации для информационно-аналитической поддержки принятия управленческих решений; ОПК-4.3 Владеет программными средствами сбора, обработки и анализа информации для информационно-аналитической поддержки принятия управленческих решений

2. Место дисциплины в структуре ОПОП бакалавриата

Дисциплина Б1.О.30 «Анализ больших данных» относится к обязательной части блока 1. «Дисциплины (модули)» учебного плана ОПОП бакалавриата по направлению подготовки 38.03.05 Бизнес-информатика. Дисциплина «Анализ больших данных» изучается обучающимися очной формы обучения в 6 семестре, очно-заочной формы обучения – в 7 семестре.

При изучении данной дисциплины обучающийся использует знания, умения и навыки, которые сформированы в процессе изучения предшествующих дисциплин: «Эконометрика», «Специальные разделы математики», «Экономико-математическое моделирование», «Информационные системы и технологии в экономике и управлении», «Моделирование бизнес-процессов», «Нечеткая логика и нейронные сети».

Знания, умения и навыки, полученные при изучении дисциплины «Анализ больших данных», будут необходимы для углубленного и осмысленного восприятия дисциплин: «Информационная безопасность и защита информации», «Системы поддержки и методы принятия решений», «Основы цифровой экономики».

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 часов.

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

Для очной формы обучения

Общая трудоёмкость дисциплины составляет 4 зачётных единицы 144 часов

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
Курсовая работа	-
Зачет с оценкой	+
Экзамен	-

Для очно-заочной формы обучения

Общая трудоёмкость дисциплины составляет 4 зачётных единицы 144 часов

Объём дисциплины	Всего часов
Общая трудоемкость дисциплины	144
Контактная работа	34
Аудиторная работа (всего):	34
Лекции	16
Семинары, практические занятия	18
Самостоятельная работа обучающихся (всего)	110
Курсовая работа	-
Зачет с оценкой	+
Экзамен	-

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

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

те м	Наименование темы	Всего	Количество часов
------	-------------------	-------	------------------

		ОФО	ОЗФО	Контактная работа				Внеаудит. работа	
				Лекции		Практические		Самост. работа	
				ОФО	ОЗФО	ОФО	ОЗФО	ОФО	ОЗФО
1.	Интеллектуальный анализ данных	24	24	4	4	4	4	16	16
2.	Сбор, хранение, обработка, извлечение и анализ больших потоков данных	24	24	4	2	4	2	16	20
3.	Статистическое мышление и статистический анализ	20	20	2	2	2	2	16	16
4.	Методы классификации и прогнозирования	24	24	4	2	4	2	16	20
5.	Введение в машинное обучение и разработку данных	24	24	4	2	4	4	16	18
6.	Применение технологий больших данных для бизнес-анализа	14	14	2	2	2	2	10	10
7.	Способы визуализации, интерпретации, проектная организация и бизнес рекомендации	14	14	2	2	2	2	10	10
	Всего по дисциплине	144	144	22	16	22	18	100	110
	Контроль								
	Итого	144	144	22	16	22	18	100	110

4.2. Содержание дисциплины, структурированное по темам (разделам)

Тема 1. Интеллектуальный анализ данных

Цели и задачи дисциплины. Введение в DataMining. Определение больших данных. Методы и стадии DataMining. Задачи DataMining. Извлечение, данных, информации и знаний. Типы задач, решаемые методами DataMining: классификация, кластеризация, регрессия, ассоциация, поиск последовательных шаблонов. Алгоритмы, получившие наибольшее распространение для каждого типа задач: самоорганизующиеся карты, деревья решений, линейная регрессия, нейронные сети, ассоциативные правила.

Сферы применения DataMining. Научные проблемы в области больших данных. Современные программные средства анализа больших данных.

Тема 2. Сбор, хранение, обработка, извлечение и анализ больших потоков данных

Основные характеристики и типы больших данных. Технологии хранения больших данных. Обзор технологий хранения больших данных. Базы данных. Системы управления базами данных. Модели данных. Подготовка исходных данных для анализа: первичная обработка и визуализация имеющихся данных.

Технологии обработки больших данных. Технологии анализа больших данных.

Технологии Business Intelligence и реляционные системы управления базами данных.

Тема 3. Статистическое мышление и статистический анализ

Статистическое мышление. Обработка данных для контроля качества процессов и технологии sixsigma. Статистический подход к анализу данных. Точечное и интервальное оценивание. Методы статистического анализа данных: дескриптивная статистика, параметрические, непараметрические, номинальные методы (корреляционный, регрессионный, дисперсионный анализы, кластерный, дискриминантный, факторный анализы).

Специализированные статистические пакеты.

Тема 4. Методы классификации и прогнозирования

Анализ последовательностей. Отбор признаков и снижение размерности. Работа с выбросами и пропущенными значениями. Классификация и кластеризация. Поиск признаковых зависимостей и частых множеств. Мультимодальная кластеризация. Предиктивные алгоритмы.

Прогнозирование и визуализация данных.

Тема 5. Введение в машинное обучение и разработку данных.

Основные аналитические методы обработки данных. Машинное обучение и майнинг больших данных (BigData).

Нейронные сети как реализация алгоритмов машинного обучения. Основные понятия теории нейронных сетей. Основные парадигмы нейронных сетей. Многослойный перцептрон: класс решаемых задач, архитектура.

Тема 6. Применение технологий больших данных для бизнес-анализа

Применение технологий больших данных для задач управления в банковской, страховой, финансовой индустриях, в масс медиа, FinTech как область применения бизнес анализа.

Задача сегментации клиентов фирмы. Оценка кредитоспособности физических лиц. Задача прогнозирования потребности в продукции. Задачи прогнозирования продаж, поступления финансовых средств и др. Примеры комбинации методов DataMining.

Тема 7. Способы визуализации, интерпретации, проектная организация и бизнес рекомендации

Способы визуального представления данных. Методы визуализации. Базовые принципы визуализации. Инструменты визуализации данных.

Визуальное управление проектами.

Основные тенденции в области визуализации

Когнитивная информатика, экономика знаний, требования к специалистам в области интеллектуальной обработки данных для бизнеса.

4.3.Содержание практических занятий (очная форма обучения)

Тема 1. Интеллектуальный анализ данных
Практическое занятие1-2. Интеллектуальный анализ данных(4 часа)
<ol style="list-style-type: none"> 1. Основные понятия и определения DataMining. 2. Методы и стадии DataMining. Задачи DataMining. Типы задач, решаемые методами DataMining 3. Сферы применения DataMining. Современные программные средства анализа больших данных.
Тема 2. Сбор, хранение, обработка, извлечение и анализ больших потоков данных
Практическое занятие 3-4. Сбор, хранение, обработка, извлечение и анализ больших потоков данных(4 часа)
<ol style="list-style-type: none"> 1. Обзор технологий хранения больших данных. 2. Модели данных.

3. Технологии обработки больших данных. Технологии анализа больших данных. Технологии Business Intelligence и реляционные системы управления базами данных.
Тема 3. Статистическое мышление и статистический анализ
Практическое занятие 5. Статистический анализ(2 часа)
1. Методы статистического анализа данных 2. Специализированные статистические пакеты
Тема 4. Методы классификации и прогнозирования
Практическое занятие 6-7. Методы классификации и прогнозирования (4 часа)
1. Анализ последовательностей. 2. Классификация и кластеризация. Поиск признаковых зависимостей и частых множеств. Мультимодальная кластеризация. Предиктивные алгоритмы. 3. Прогнозирование и визуализация данных.
Тема 5. Введение в машинное обучение и разработку данных
Практическое занятие 8-9. Введение в машинное обучение и разработку данных (4 часа)
1. Машинное обучение и майнинг больших данных (BigData). 2. Нейронные сети как реализация алгоритмов машинного обучения. 3. Многослойный перцептрон: класс решаемых задач, архитектура. 4. Построение нейросетевой прогнозной модели спроса.
Тема 6. Применение технологий больших данных для бизнес-анализа
Практическое занятие 10. Применение технологий больших данных для бизнес-анализа(2 часа)
1. Применение технологий больших данных для задач управления в банковской, страховой, финансовой индустриях, в масс медиа, FinTech как область применения бизнес анализа. 2. Построение скоринговой модели кредитования (дерева решений);
Тема 7. Способы визуализации, интерпретации, проектная организация и бизнес рекомендации
Практическое занятие 11. Способы визуализации, интерпретации, проектная организация и бизнес рекомендации. (2 часа)
1. Способы визуального представления данных. Базовые принципы визуализации. Инструменты визуализации данных. Методы визуализации. 2. Визуальное управление проектами. 3. Когнитивная информатика, экономика знаний, требования к специалистам в области интеллектуальной обработки данных для бизнеса.

4.4. Содержание самостоятельной работы

Тема 1. Интеллектуальный анализ данных
1. Алгоритмы, получившие наибольшее распространение для каждого типа задач: самоорганизующиеся карты, деревья решений, линейная регрессия, нейронные сети, ассоциативные правила. 2. Сферы применения DataMining. Научные проблемы в области больших данных.
Тема 2. Сбор, хранение, обработка, извлечение и анализ больших потоков данных
1. Технологии обработки больших данных. Технологии анализа больших данных. 2. Технологии Business Intelligence и реляционные системы управления базами данных.
Тема 3. Статистическое мышление и статистический анализ
1. Методы статистического анализа данных: дескриптивная статистика,

параметрические, непараметрические, номинальные методы (корреляционный, регрессионный, дисперсионный анализы, кластерный, дискриминантный, факторный анализы).
Тема 4. Методы классификации и прогнозирования 1. Мультимодальная кластеризация. Предиктивные алгоритмы. 2. Прогнозирование и визуализация данных.
Тема 5. Введение в машинное обучение и разработку данных 1. Нейронные сети как реализация алгоритмов машинного обучения. Основные понятия теории нейронных сетей. Основные парадигмы нейронных сетей. Многослойный персептрон: класс решаемых задач, архитектура.
Тема 6. Применение технологий больших данных для бизнес-анализа 1. Задача сегментации клиентов фирмы. Оценка кредитоспособности физических лиц. Задача прогнозирования потребности в продукции. Задачи прогнозирования продаж, поступления финансовых средств и др. Примеры комбинации методов DataMining.
Тема 7. Способы визуализации, интерпретации, проектная организация и бизнес рекомендации 1. Основные тенденции в области визуализации 2. Когнитивная информатика, экономика знаний, требования к специалистам в области интеллектуальной обработки данных для бизнеса.

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.

Фонд оценочных средств по дисциплине приведен в приложении к РПД.

6. Учебно-методическое обеспечение дисциплины

1. Шнарева, Г. В. Анализ данных : учебно-методическое пособие / Г. В. Шнарева, Ж. Г. Пономарева. — Симферополь : Университет экономики и управления, 2019. — 129 с. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/89482.html>. — Режим доступа: для авторизир. пользователей

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

а) основная литература:

1. Чубукова, И. А. Data Mining : учебное пособие / И. А. Чубукова. — 3-е изд. — Москва, Саратов : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. — 469 с. — ISBN 978-5-4497-0289-0. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/89404.html>. — Режим доступа: для авторизир. пользователей
2. Орлов, А. И. Прикладной статистический анализ : учебник / А. И. Орлов. — Москва : Ай Пи Ар Медиа, 2022. — 812 с. — ISBN 978-5-4497-1480-0. — Текст : электронный //

- Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/117038.html>. — Режим доступа: для авторизир. пользователей. - DOI: <https://doi.org/10.23682/117038>
3. Замятин, А. В. Интеллектуальный анализ данных : учебное пособие / А. В. Замятин. — Томск : Издательский Дом Томского государственного университета, 2020. — 194 с. — ISBN 978-5-94621-898-6. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/116889.html>. — Режим доступа: для авторизир. пользователей

б) дополнительная литература:

4. Маккинли, Уэс Python и анализ данных / Уэс Маккинли ; перевод А. Слинкина. — 2-е изд. — Саратов : Профобразование, 2019. — 482 с. — ISBN 978-5-4488-0046-7. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/88752.html>. — Режим доступа: для авторизир. пользователей
5. Интеллектуальный предиктивный мультимодальный анализ слабоструктурированных больших данных / Н. Г. Ярушкина, И. А. Андреев, Г. Ю. Гуськов [и др.]. — Ульяновск : Ульяновский государственный технический университет, 2020. — 221 с. — ISBN 978-5-9795-2088-9. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/106136.html>. — Режим доступа: для авторизир. пользователей
6. Шахова, О. А. Статистическая обработка результатов исследований : учебное пособие / О. А. Шахова. — Тюмень : Издательство «Титул», 2022. — 103 с. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/119099.html>. — Режим доступа: для авторизир. пользователей

8. 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/> — Текст: электронный.
3. Центр Инженерных Технологий и Моделирования «Экспонента»: официальный сайт. — URL: <https://exponenta.ru> — Текст: электронный.
4. Бесплатная электронная библиотека онлайн «Единое окно доступа к образовательным ресурсам»: официальный сайт. — URL: <http://www.window.edu.ru> — Текст: электронный.

9. 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.

10. 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, «Googlechrome»);

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

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

- *табличный процессор (например, «Microsoft Excel»);

- *системы для анализа данных (например, «Loginom»).

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

Для преподавания дисциплины требуются специальные материально-технические средства (компьютерные классы и т.п.). Однако во время лекционных занятий, которые проводятся в большой аудитории, использовать проектор для демонстрации слайдов, схем, таблиц и прочего материала, мультимедийные проекторы Epson, BenqViewSonic; экраны для проекторов; ноутбуки Asus, Lenovo, микрофоны.