Министерство науки и высшего образования
Российской Федерации

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Донецкий государственный университет»

Факультет математики и информационных технологий
Кафедра информационных систем управления

УТВЕРЖДАЮ
проректор
_____ П.А. Машаров
«29» марта 2024 г.
МП

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

АНАЛИЗ БОЛЬШИХ ДАННЫХ

Укрупненная группа направлений подготовки	46.00.00 История и археология
Программа высшего образования	Программа магистратуры
Направление подготовки	46.04.02 Документоведение и архивоведение
Магистерская программа	Информационно-документационное обеспечение управления
Квалификация	Магистр
Форма обучения	Очная, заочная

Рабочая программа адаптирована для лиц
с ограниченными возможностями здоровья и инвалидов

Донецк 2024

Рабочая программа дисциплины **«Анализ больших данных»** для обучающихся по направлению подготовки 46.04.02 Документоведение и архивоведение (Магистерская программа: Информационно-документационное обеспечение управления), составлена на основании Федерального государственного образовательного стандарта высшего образования – магистратура по направлению подготовки 46.04.02 Документоведение и архивоведение, утвержденного приказом Министерства науки и высшего образования Российской Федерации от 29 октября 2020 г. № 1345, Порядка организации и осуществления образовательной деятельности по образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры, утвержденного приказом Министерства науки и высшего образования Российской Федерации от 06 апреля 2021 г. № 245 (с изм. и доп.), в соответствии с учебным планом, утвержденным Ученым советом ФГБОУ ВО «ДонГУ» для набора 2024 года.

Разработчик:
доцент кафедры информационных систем управления, канд. экон. наук, доцент — А. М. Гизатулин

Рабочая программа одобрена на заседании кафедры информационных систем управления.
Протокол от 22.03.2024 г. № 6а

Заведующий кафедрой — Н. Ш. Пономаренко

СОГЛАСОВАНО:

Декан факультета математики и информационных технологий
28.03.2024 г. — И. А. Моисеенко

Учебно-методическая комиссия факультета математики и информационных технологий.
Протокол от 27.03.2024 г. № 7.
Председатель — Л. И. Селякова

Руководитель основной профессиональной образовательной программы,
д-р экон. наук, доцент
26.03.2024 г. — Н. Ш. Пономаренко

1. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

1.1. Требования к предварительной подготовке обучающихся, предшествующие и сопутствующие дисциплины, на которых основывается изучение данной.

Изучение данной дисциплины основывается на базе программы бакалавриата: Информационные технологии, Информационные технологии, Введение в профессиональную деятельность.

1.2. Дисциплины, курсовые работы и практики, для которых освоение данной дисциплины необходимо как предшествующее:

Подготовка к процедуре защиты и защита выпускной квалификационной работы.

2. ОПИСАНИЕ ДИСЦИПЛИНЫ

2.1. Общая характеристика

Наименование показателя	Значение показателя
Название образовательной программы	46.04.02 Документоведение и архивоведение (Магистерская программа: Информационно-документационное обеспечение управления)
Шифр и название в соответствии с учебным планом	Б1.В.ОД.5 Анализ больших данных
Часть образовательной программы	Вариативная часть: выбор вуза
Количество зачетных единиц / всего часов	2 / 72

2.2. Распределение часов по формам и периодам обучения

Форма обучения	курс	семестр	Общее количество часов					
			лекционных	лабораторных	практических	самостоятельной работы + контроль	всего	Форма контроля
Очная	1	1	17	–	17	38	72	зачет
Заочная	1	1	2	–	4	66	72	зачет

3. ЦЕЛИ ДИСЦИПЛИНЫ

Цель дисциплины «Анализ больших данных» – формирование теоретических знаний и практических навыков, позволяющих использовать современные математико-статистические методы и программные средства для обработки и анализа данных.

4. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ КОМПОНЕНТА ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ, ИХ ИНДИКАТОРЫ И ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Компетенции	Индикаторы	Результаты обучения
ПК-4. Способен осуществлять управление цифровой трансформацией	ПК-4.2. Осуществляет информационно-аналитическую поддержку анализа больших данных и	ПК-6.2.1 Знает основы теории анализа больших данных. ПК-6.2.2 Умеет решать задачи профессиональной деятельности на основе анализа данных.

документированных сфер деятельности организации	интерпретирует его результаты	ПК-6.2.3 Владеет практическими методами анализа больших данных.

5. ПРОГРАММА ДИСЦИПЛИНЫ

Название темы	Краткое содержание темы (вопросы темы)
1. Методы классификации данных.	Постановка задачи классификации. Представление результатов классификации. Методы построения правил классификации. Деревья решений.
2. Кластерный анализ.	Постановка задачи. Меры близости, основанные на расстояниях. Иерархические методы. Итерационные методы.
3. Ассоциативные правила.	Постановка задачи. Представление результатов. Алгоритмы поиска ассоциативных правил.
4. Визуальный анализ данных (Visual Mining).	Процесс визуализации данных. Анализ средств визуализации. Методы визуализации.
5. Анализ текстовой информации (Text Mining).	Задача анализа текстов. Извлечение ключевых понятий из текста. Классификация текстовых документов. Средства анализа текстовой информации.
6. Стандарты Data Mining.	Стандарт CWM (Common Warehouse Metamodel). Стандарт CRISP-DM (CRoss-Industry Standard Process for Data Mining). Стандарт PMML (Predicted Model Markup Language). Стандарт DMX (Microsoft Data Mining eXtensions). Стандарт JDM (Java Data Mining).
7. Извлечение знаний в реальном времени.	Концепция извлечения знаний в реальном времени. Рекомендательные машины. Инструменты извлечения знаний в реальном времени.
8. Извлечение знаний из сети Интернет (Web Mining).	Проблемы анализа информации из сети Интернет. Этапы извлечения знаний из сети Интернет. Классификация задач извлечения знаний из сети Интернет. Классификация методов извлечения знаний из сети Интернет.

6. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

6.1. Форма обучения – очная, курс – 1, семестр – 1

Наименования разделов и тем	Количество часов				
	Лекц.	Лабор.	Практ.	СРС+К	Всего
1. Методы классификации данных.	2	–	2	5	9
2. Кластерный анализ.	2	–	2	5	9
3. Ассоциативные правила.	2	–	2	5	9
4. Визуальный анализ данных (Visual Mining).	2	–	2	5	9
5. Анализ текстовой информации (Text Mining).	2	–	2	5	9
6. Стандарты Data Mining.	2	–	2	5	9
7. Извлечение знаний в реальном времени.	2	–	2	5	9
8. Извлечение знаний из сети Интернет (Web Mining).	3	–	3	3	9
ИТОГО ЗА СЕМЕСТР	17	–	17	38	72

6.2. Форма обучения – заочная, курс – 1, семестр – 1

Наименования разделов и тем	Количество часов				
	Лекц.	Лабор.	Практ.	СРС+К	Всего
1. Методы классификации данных.	0,25	–	0,5	8,25	9
2. Кластерный анализ.	0,25	–	0,5	8,25	9
3. Ассоциативные правила.	0,25	–	0,5	8,25	9
4. Визуальный анализ данных (Visual Mining).	0,25	–	0,5	8,25	9
5. Анализ текстовой информации (Text Mining).	0,25	–	0,5	8,25	9
6. Стандарты Data Mining.	0,25	–	0,5	8,25	9
7. Извлечение знаний в реальном времени.	0,25	–	0,5	8,25	9
8. Извлечение знаний из сети Интернет (Web Mining).	0,25	–	0,5	8,25	9
ИТОГО ЗА СЕМЕСТР	2	–	4	66	72

7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ (СРЕДСТВА) ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ, ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

7.1. Контрольные вопросы

1. Постановка задачи классификации.
2. Представление результатов классификации.
3. Методы построения правил классификации.
4. Деревья решений.
5. Постановка задачи кластеризации.
6. Меры близости, основанные на расстояниях.
7. Иерархические методы.
8. Итерационные методы.
9. Постановка задачи ассоциации.
10. Представление результатов.
11. Алгоритмы поиска ассоциативных правил.
12. Процесс визуализации данных.
13. Анализ средств визуализации.
14. Методы визуализации.
15. Задача анализа текстов.
16. Извлечение ключевых понятий из текста.
17. Классификация текстовых документов.
18. Средства анализа текстовой информации.
19. Стандарт CWM (Common Warehouse Metamodel).
20. Стандарт CRISP-DM (CRoss-Industry Standard Process for Data Mining).
21. Стандарт PMML (Predicted Model Markup Language).
22. Стандарт DMX (Microsoft Data Mining eXtensions).
23. Стандарт JDM (Java Data Mining).
24. Концепция извлечения знаний в реальном времени.
25. Рекомендательные машины.
26. Инструменты извлечения знаний в реальном времени.
27. Проблемы анализа информации из сети Интернет.
28. Этапы извлечения знаний из сети Интернет.
29. Классификация задач извлечения знаний из сети Интернет.
30. Классификация методов извлечения знаний из сети Интернет.

7.2. Темы письменных работ (типы задач)

Практические работы по темам:

1. Методы классификации данных.
2. Кластерный анализ.
3. Ассоциативные правила.
4. Визуальный анализ данных (Visual Mining).
5. Анализ текстовой информации (Text Mining).
6. Стандарты Data Mining.
7. Извлечение знаний в реальном времени.
8. Извлечение знаний из сети Интернет (Web Mining).

8. РАСПРЕДЕЛЕНИЕ БАЛЛОВ, КОТОРЫЕ ПОЛУЧАЮТ ОБУЧАЮЩИЕСЯ

Общая оценка знаний обучающихся по дисциплине проводится по 100-балльной шкале исходя из максимума, приведенного в таблице ниже. Организационно-учебная работа в аудитории оценивается на основе таких критериев как посещаемость занятий, своевременное и качественное выполнение домашних заданий, активность во время проведения лекционных и практических занятий (участие в обсуждении текущего и пройденного материала, решение задач и т.п.).

8.1. Семестр 1

Номера разделов	Виды работ	Максимальное количество баллов
1	Организационно-учебная работа в аудитории	15
	Самостоятельная работа (выполнение практических работ по варианту)	80
	Контрольная работа по теоретическому материалу	5
Общий итог за семестр		100

Соответствие баллов оценке

Количество баллов из 100	ECTS	Оценка по пятибалльной шкале	
		Экзамен, дифференцированный зачет	Зачет
90-100	A	отлично	зачтено
80-89	B	хорошо	зачтено
75-79	C		зачтено
70-74	D	удовлетворительно	зачтено
60-69	E		зачтено
35-59	FX	неудовлетворительно	не зачтено
0-34	F		не зачтено

9. ОБЕСПЕЧЕНИЕ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ДЛЯ ЛИЦ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ И ИНВАЛИДОВ

В ходе реализации дисциплины используются следующие дополнительные методы обучения, текущего контроля успеваемости и промежуточной аттестации обучающихся в зависимости от их индивидуальных особенностей:

1) для слепых и слабовидящих:

– лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;

– для выполнения задания при необходимости предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств;

– письменные задания оформляются увеличенным шрифтом.

2) для глухих и слабослышащих:

– лекции оформляются в виде электронного документа;

– письменные задания выполняются на компьютере в письменной форме;

– экзамен проводится в письменной форме на компьютере; возможно проведение в форме тестирования.

3) для лиц с нарушениями опорно-двигательного аппарата:

– лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;

– письменные задания выполняются на компьютере;

– экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

При необходимости предусматривается увеличение времени для подготовки ответа.

Процедура проведения промежуточной аттестации для обучающихся устанавливается с учётом их индивидуальных психофизических особенностей. Промежуточная аттестация может проводиться в несколько этапов.

Проведение процедуры оценивания результатов обучения допускается с использованием дистанционных образовательных технологий.

Обеспечивается доступ к информационным и библиографическим ресурсам в сети Интернет для каждого обучающегося в формах, адаптированных к ограничениям их здоровья и восприятия информации:

1) для слепых и слабовидящих:

– в печатной форме увеличенным шрифтом;

– в форме электронного документа;

2) для глухих и слабослышащих:

– в печатной форме;

– в форме электронного документа.

3) для обучающихся с нарушениями опорно-двигательного аппарата:

– в печатной форме;

– в форме электронного документа.

10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОГО ПРОЦЕССА

Учебные занятия проводятся в в 8-м учебном корпусе (г. Донецк, ул. Челюскинцев, д. 198 а) университета. Для проведения лабораторных занятий требуется аудитория, оборудованная меловой или маркерной доской, мультимедийный проектор и экран, ноутбук, комплект учебной мебели для студентов, рабочее место преподавателя, выход в Интернет – проводной или с использованием Wi-Fi.

Для самостоятельной работы используются текстовые и электронные ресурсы Научной библиотеки университета и других электронных библиотечных баз данных, учебно-методическое обеспечение, представленное в учебно-методическом кабинете Главного корпуса (ауд.405).

Обучающиеся имеют возможность использовать учебные материалы по дисциплине, размещенные на платформе Moodle Центра дистанционного образования ФГБОУ ВО «ДонГУ». При изучении дисциплины применяются электронное обучение и дистанционные образовательные технологии.

С использованием ресурсов платформы дистанционного образования осуществляется текущий контроль знаний обучающихся на основе тестирования и проверки результатов самостоятельной работы.

11. РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

11.1. Основная литература

1. Алексеев, Д. С. Технологии интеллектуального анализа данных / Д. С. Алексеев, О. В. Щекочихин. — 2-е изд., стер. — Санкт-Петербург : Лань, 2024. — 176 с. — ISBN 978-5-507-48763-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/362915 (дата обращения: 01.09.2023).

2. Алетдинова, А. А. Интеллектуальный анализ больших данных : учебное пособие / А. А. Алетдинова, М. Ш. Муртазина. — Новосибирск : НГТУ, 2023. — 66 с. — ISBN 978-5-7782-4899-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/404567 (дата обращения: 01.09.2023).

11.2. Дополнительная литература

1. Кобзаренко, Д. Н. Учебное пособие дисциплины «Анализ больших данных» для направления подготовки 38.03.05 «Бизнес-информатика», профиль «Электронный бизнес» : учебное пособие / Д. Н. Кобзаренко, А. Г. Мустафаев ; составитель Д. Н. Кобзаренко. — Махачкала : ДГУНХ, 2019. — 107 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/246542 (дата обращения: 01.09.2023).

2. Ланских, Ю. В. Интеллектуальный анализ данных : учебное пособие / Ю. В. Ланских, В. Г. Ланских. — Киров : ВятГУ, 2023. — 240 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/408569 (дата обращения: 01.09.2023).

3. Маккинни, У. Python и анализ данных / У. Маккинни ; перевод с английского А. А. Слинкина. — 2-ое изд., испр. и доп. — Москва : ДМК Пресс, 2020. — 540 с. — ISBN 978-5-97060-590-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/131721 (дата обращения: 01.09.2023).

4. Макшанов, А. В. Технологии интеллектуального анализа данных : учебное пособие / А. В. Макшанов, А. Е. Журавлев. — 2-е изд., стер. — Санкт-Петербург : Лань, 2022. — 212 с. — ISBN 978-5-8114-4493-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/206711 (дата обращения: 01.09.2023).

12. ИНФОРМАЦИОННЫЕ РЕСУРСЫ

1. **Национальная электронная библиотека** (НЭБ): федеральная государственная информационная система / Министерство Культуры РФ; Российская государственная библиотека. – Москва, 2019- . – URL: https://rusneb.ru/ (дата обращения: 01.09.2023). – Режим доступа: свободный, подписка. Необходима установка программного обеспечения. – Текст: электронный.

2. **eLIBRARY.RU**: научная электронная библиотека: сайт. – Москва, 2000- . – URL: https://elibrary.ru (дата обращения: 01.09.2023). – Режим доступа: для авторизов. пользователей. –Текст: электронный.

3. Научная электронная библиотека **«КиберЛенинка»**: сайт / Ассоциация «Открытая наука». – Москва, 2014- . – URL: https://cyberleninka.ru/. – Режим доступа: свободный. – Текст: электронный.

4. Электронно-библиотечная система **«Лань»**: [сайт]. – URL: https://e.lanbook.com (дата обращения: 01.09.2023). – Режим доступа: для авторизов. пользователей. – Текст: электронный.

5. **ЭБС Юрайт**: электронная библиотечная система: сайт. – Москва, 2013. – URL: https://biblio-online.ru (дата обращения: 01.09.2023). – Режим доступа: для авторизов. пользователей. – Текст: электронный.

6. **Электронно-библиотечная система ДонГУ**: сайт / ФГБОУ ВО «ДонГУ». – Донецк, 2016- . – URL: http://library.donnu.ru/ (дата обращения: 01.09.2023). – Режим доступа: свободный. – Текст: электронный.

7. **Электронный каталог** Научной библиотеки ДонГУ: раздел сайта / НБ ДонГУ. – Текст: электронный // ЭБС ДонГУ: сайт. – URL: http://library.donnu.ru/catalog/ (дата обращения: 01.09.2023). – Режим доступа: поиск свободный, электронные документы – для пользователей ДонГУ.

8. **Электронный архив ДонГУ**: раздел сайта / НБ ДонГУ. – Текст: электронный // ЭБС ДонГУ: сайт. – URL: http://repo.donnu.ru/ (дата обращения: 01.09.2023). – Режим доступа: свободный.

13. ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

1. Windows 7 PRO (корпоративная лицензия ДонГУ № 46484614)
2. Microsoft Office (корпоративная лицензия ДонГУ № 46472919)
3. Microsoft Visual Studio (лицензия программы Dream Spark для высших учебных заведений)
4. Антивирус Касперского, Adobe Acrobat Reader, xPDF (лицензии GPL, Apache, BSD для свободного программного обеспечения).