

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Минцаев Магомед Шавалович

Должность: Ректор

Дата подписания: 02.03.2020

Уникальный программный ключ:

236bcc35c296f119d6aafdc22836b21db52dbc07971a88865a56231a4504cc

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ГРОЗНИНСКИЙ ГОСУДАРСТВЕННЫЙ НЕФТЯНОЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
имени академика М.Д. Миллионщикова

«УТВЕРЖДАЮ»

Первый проректор

И.Г. Гайрабеков



РАБОЧАЯ ПРОГРАММА

дисциплины

«Анализ больших данных»

Направление подготовки

09.03.02 Информационные системы и технологии

Направленности (профили)

«Информационные системы и технологии»

«Информационные технологии в образовании»

«Информационные технологии в дизайне»

Квалификация

бакалавр

Грозный – 2020

1. Цели и задачи дисциплины

Цель преподавания дисциплины «Анализ больших данных» состоит в изучение теоретических основ анализа больших данных, включая базовые элементы статистического программирования и интеллектуального анализа больших наборов данных.

Задачи дисциплины – научить производить расчеты с применением технологий анализа больших данных и решать широкий спектр прикладных задач обработки больших наборов данных.

2. Место дисциплины в структуре образовательной программы

Учебная дисциплина «Анализ больших данных» относится к формируемой части ФГОС ВО по направлению подготовки 09.03.02 Информационные системы и технологии (квалификация «бакалавр»).

Для освоения дисциплины «Анализ больших данных» студент должен обладать знаниями и умениями, приобретенными в результате освоения предшествующих дисциплин:

- Безопасность информационных технологий и систем.
- Технологии обработки информации.
- Вычислительные машины, сети и телекоммуникации.
- Управление данными.
- Базы данных.
- Представление знаний в информационных системах.

Дисциплина «Анализ больших данных» является предшествующей и необходимой для дисциплины «Проектирование информационных систем» и выполнения и защиты выпускной квалификационной работы.

3. Требования к результатам освоения дисциплины

Процесс изучения дисциплины направлен на формирование следующих компетенций и индикаторов их достижения:

ОПК-1. Способность применять естественнонаучные и общетехнические знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности:

– **ИД-1 ОПК-1-знать:** основы математики, физики, вычислительной техники и программирования.

– **ИД-2 ОПК-1-уметь:** решать стандартные профессиональные задачи с применением естественнонаучных и общетехнических знаний, методов математического анализа и моделирования.

– **ИД-3 ОПК-1-иметь навыки:** теоретического и экспериментального исследования объектов профессиональной деятельности.

ОПК-6. Способность разрабатывать алгоритмы и программы, пригодные для практического применения в области информационных систем и технологий:

– **ИД-1 ОПК-6-знать:** основные языки программирования и работы с базами данных, операционные системы и оболочки, современные программные среды разработки информационных систем и технологий.

– **ИД-2 ОПК-6-уметь:** применять языки программирования и работы с базами данных, современные программные среды разработки информационных систем и

технологий для автоматизации бизнес-процессов, решения прикладных задач различных классов, ведения баз данных и информационных хранилищ.

– **ИД-3 ОПК-6-иметь навыки:** программирования, отладки и тестирования прототипов программно-технических комплексов задач.

ПК-1. Способность разрабатывать программное обеспечение (ПО), включая проектирование, отладку, проверку работоспособности и модификацию ПО:

- **ИД-1 ПК-1-знать:** процедуры интеграции программных модулей;
- **ИД-2 ПК-уметь:** проводить анализ требований к программному обеспечению;
- **ИД-3 ПК-1-иметь навыки:** разработки технических спецификаций на программные компоненты и их взаимодействие.

4. Объем дисциплины и виды учебной работы

Таблица 1

Вид учебной работы	Всего		Семестры	
	часов/ зач. ед.		8	8
	ОФО	ЗФО	ОФО	ЗФО
Контактная работа (всего)	72/2	20/0,4	72/2	20/0,4
В том числе:				
Лекции	24/0,7	10/0,3	24/0,7	10/0,3
Практические занятия	-	-	-	-
Семинары	-	-	-	-
Лабораторные работы	48/1,3	10/0,3	48/1,3	10/0,3
Самостоятельная работа (всего)	108/3	160/4,4	108/3	160/4,4
В том числе:				
Доклады с презентациями	36/1	52/1,4	36/1	52/1,4
<i>И (или) другие виды самостоятельной работы:</i>				
Подготовка к лабораторным работам	36/1	72/2	36/1	72/2
Подготовка к экзамену	36/1	36/1	36/1	36/1
Вид отчетности				
Общая трудоемкость дисциплины	ВСЕГО в часах	180	180	144
	ВСЕГО в зач. единицах	5	5	4

5. Содержание дисциплины

5.1. Разделы дисциплины и виды занятий

Таблица 2

№ п/п	Наименование раздела дисциплины по семестрам	Часы лекционн ых занятий		Часы лаборатор ных занятий		Часы практичес ких (семинарс ких) занятий		Всего часов	
		ОФ О	ЗФ О	ОФ О	ЗФ О	ОФ О	ЗФ О	ОФ О	ЗФ О
2 семестр									
1.	Технологии анализа данных	10	4	16	4	-	-	26	8
2.	Интеллектуальный анализ данных	14	6	32	6	-	-	46	12
Итого:		24	10	48	10	-	-	72	20

5.2. Лекционные занятия

Таблица 3

№ п/п	Наименование раздела дисциплины	Содержание раздела
1.	Технологии анализа данных	<p>Большие данные (Big Data): современные подходы к обработке и хранению.</p> <p>Процесс анализа. Общая схема анализа. Извлечение и визуализация данных. Этапы моделирования. Процесс построения моделей. Формы представления данных, типы и виды данных. Представления наборов данных.</p> <p>Технологии KDD и Data Mining. Подготовка данных к анализу. Методика извлечения знаний. Data Mining. Мультидисциплинарный характер Data Mining. Причины распространения KDD и Data Mining. Актуальность технологий Data Mining как средств обработки больших объемов информации.</p> <p>Программное обеспечение в области анализа данных. Аналитические платформы: классификация и особенности применения. Языки визуального моделирования.</p>

2.	Интеллектуальный анализ данных	<p>Ассоциативные правила. Аффинитивный анализ, предметный набор. Поддержка и достоверность ассоциативного правила. Значимость ассоциативных правил, лифт и левередж. Поиск ассоциативных правил. Частые предметные наборы и их обнаружение. Алгоритм генерации ассоциативных правил. Иерархические ассоциативные правила. Методы поиска иерархических ассоциативных правил.</p> <p>Определение кластеризации. Постановка задачи кластеризации. Цели кластеризации в Data Mining. Примеры кластеризации в различных областях. Виды метрик. Шаги алгоритма. Меры расстояний. Пример работы алгоритма k-means. Проблемы алгоритмов кластеризации.</p> <p>Применение классификации и регрессии. Обзор методов классификации и регрессии. Статистические методы. Методы, основанные на обучении, разнообразие подходов.</p>
----	--------------------------------	--

5.3. Лабораторные занятия

Таблица 4

№ п/п	Наименование раздела дисциплины	Наименование лабораторных работ
1.	Технологии анализа данных	<p>Хранение больших данных. Предварительный анализ больших наборов данных</p>
2.	Интеллектуальный анализ данных	<p>Алгоритмизация больших данных. Метод главных компонент для сжатия больших данных. Иерархические методы кластерного анализа больших данных. Неиерархические методы кластерного анализа больших данных.</p>

5.4. Практические (семинарские) занятия: нет

6. Самостоятельная работа студентов по дисциплине

Способ организации самостоятельной работы: подготовка презентации на 12-15 слайдов с устным докладом по заданной тематике; разработка приложений по заданию преподавателя.

Тематика докладов с презентациями:

1. Ассоциативные правила. Поиск ассоциативных правил.
2. Кластеризация. Алгоритм кластеризации k-means.
3. Прогнозирование с помощью линейной регрессии.
4. Классификация с помощью нейросети.
5. Классификация с помощью деревьев решений.

6. Программы статистической обработки информации.
7. Прогнозирование и предвидение в социально-политических и медиа процессах

Учебно-методическое обеспечение для самостоятельной работы студентов:

1. Виктор Майер-Шенбергер Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим / Виктор Майер-Шенбергер, Кеннет Кукьер. — Москва : Манн, Иванов и Фербер, 2014. — 220 с. — ISBN 978-5-91657-936-9. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/39150.html> (дата обращения: 09.04.2020).
2. Воронов В.И. Data Mining - технологии обработки больших данных : учебное пособие / Воронов В.И., Воронова Л.И., Усачев В.А.. — Москва : Московский технический университет связи и информатики, 2018. — 47 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/81324.html> (дата обращения: 09.04.2020).
3. Железнов М.М. Методы и технологии обработки больших данных : учебно-методическое пособие / Железнов М.М.. — Москва : МИСИ-МГСУ, ЭБС АСВ, 2020. — 46 с. — ISBN 978-5-7264-2193-3. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/101802.html> (дата обращения: 09.04.2020).
4. Яцков, Н. Н. Интеллектуальный анализ данных: пособие / Н. Н. Яцков. — Минск: БГУ, 2014. — 151 с.

7. Оценочные средства

Вопросы к рубежной аттестации

К 1-ой рубежной аттестации:

1. Понятие Большие данные. Роль цифровой информации в 21 веке.
2. Проблемы анализа и обработки большого объема данных.
3. Базовые принципы обработки больших данных.
4. Определение модели. Свойства модели.
5. Аналитический подход к моделированию.
6. Информационный подход к моделированию.
7. Лица, участвующие в информационном моделировании.
8. Общая схема анализа.
9. Определение тиражирования знаний. Процесс построения модели.
10. Технологии обработки больших данных: NoSQL
11. Технологии обработки больших данных: MapReduce
12. Технологии обработки больших данных: Hadoop, R.
13. Методика извлечения знаний Knowledge Discovery in Databases (KDD). Этапы KDD.
14. Data Mining. Постановка основных задач.
15. Машинное обучение.

Ко 2-ой рубежной аттестации:

1. Бизнес-решения с помощью алгоритмов Data Mining.
2. Классификация ПО в области Data Mining и KDD.

3. Типовая схема системы на базе аналитической платформы.
 4. Понятие ассоциативного правила и транзакции.
 5. Определение поддержки и достоверности.
 6. Определение значимости и полезности ассоциативных правил, показатели их характеризующие.
 7. Формальная постановка задачи кластеризации.
 8. Цели кластеризации.
 9. Основные шаги алгоритма k-means. Условие останковки алгоритма k-means.
- Преимущества и недостатки алгоритма k-means.
10. Кластеризация с помощью самоорганизующейся карты Кохонена
 11. Этапы проведения классификации.
 12. Обзор методов классификации и регрессии.
 13. Задачи линейной и логистической регрессии.
 14. Определение дерева решений. Структура дерева решений. Выбор атрибута разбиения в узле.

Вопросы к экзамену

1. Понятие Большие данные. Роль цифровой информации в 21 веке.
 2. Проблемы анализа и обработки большого объема данных.
 3. Базовые принципы обработки больших данных.
 4. Определение модели. Свойства модели.
 5. Аналитический подход к моделированию.
 6. Информационный подход к моделированию.
 7. Лица, участвующие в информационном моделировании.
 8. Общая схема анализа.
 9. Определение тиражирования знаний. Процесс построения модели.
 10. Технологии обработки больших данных: NoSQL,
 11. Технологии обработки больших данных: MapReduce,
 12. Технологии обработки больших данных: Hadoop, R.
 13. Методика извлечения знаний Knowledge Discovery in Databases (KDD). Этапы KDD.
 14. Data Mining. Постановка основных задач.
 15. Машинное обучение.
 15. Бизнес-решения с помощью алгоритмов Data Mining.
 16. Классификация ПО в области Data Mining и KDD.
 17. Типовая схема системы на базе аналитической платформы.
 18. Понятие ассоциативного правила и транзакции.
 19. Определение поддержки и достоверности.
 20. Определение значимости и полезности ассоциативных правил, показатели их характеризующие.
 21. Формальная постановка задачи кластеризации.
 22. Цели кластеризации.
 23. Основные шаги алгоритма k-means. Условие останковки алгоритма k-means.
- Преимущества и недостатки алгоритма k-means.
24. Кластеризация с помощью самоорганизующейся карты Кохонена
 25. Этапы проведения классификации.

26. Обзор методов классификации и регрессии.
27. Задачи линейной и логистической регрессии.
28. Определение дерева решений. Структура дерева решений. Выбор атрибута разбиения в узле.

Помимо проверки знания теоретического материала, на аттестации / экзамене студентам предлагаются практические задания по разделам дисциплины.

Образец билетов рубежной аттестации:

<p style="text-align: center;">Грозненский Государственный Нефтяной Технический Университет им. акад. М.Д. Миллионщикова Кафедра «Информационные технологии» Дисциплина «Анализ больших данных» 1-я рубежная аттестация Группа: _____ Семестр: 2 Билет № _____</p> <ol style="list-style-type: none">1. Базовые принципы обработки больших данных.2. Технологии обработки больших данных: NoSQL, <p>Преподаватель _____</p>
--

<p style="text-align: center;">Грозненский Государственный Нефтяной Технический Университет им. акад. М.Д. Миллионщикова Кафедра «Информационные технологии» Дисциплина «Анализ больших данных» 2-я рубежная аттестация Группа: _____ Семестр: 2 Билет № _____</p> <ol style="list-style-type: none">1. Базовые принципы обработки больших данных2. Классификация ПО в области Data Mining и KDD <p>Преподаватель _____</p>
--

Образец билета к экзамену:

**Грозненский Государственный Нефтяной Технический Университет
им. акад. М.Д. Миллионщикова**

Кафедра «Информационные технологии»

Дисциплина «Анализ больших данных»

Группа:

Семестр: 2

Билет №

1. Определение модели. Свойства модели.
2. Этапы проведения классификации.

Подпись преподавателя _____ **Подпись заведующего**
кафедрой _____

Текущий контроль

Образец типового задания для лабораторных занятий

Лабораторная работа на тему «Неиерархические методы кластерного анализа больших данных»

Используя механизм кластеризации, реализованный на алгоритме k-means, основываясь на данных предложенных преподавателем, решите задачу распределения данных на кластеры и выявления скрытых закономерностей. Проанализируйте получившуюся картину распределения.

8. Учебно-методическое и информационное обеспечение дисциплины

Основная литература:

1. Виктор Майер-Шенбергер Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим / Виктор Майер-Шенбергер, Кеннет Кукьер. — Москва : Манн, Иванов и Фербер, 2014. — 220 с. — ISBN 978-5-91657-936-9. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/39150.html> (дата обращения: 09.04.2020).

2. Воронов В.И. Data Mining - технологии обработки больших данных : учебное пособие / Воронов В.И., Воронова Л.И., Усачев В.А.. — Москва : Московский технический университет связи и информатики, 2018. — 47 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/81324.html> (дата обращения: 09.04.2020).

Дополнительная литература:

3. Железнов М.М. Методы и технологии обработки больших данных : учебно-методическое пособие / Железнов М.М.. — Москва : МИСИ-МГСУ, ЭБС АСВ, 2020. — 46 с. — ISBN 978-5-7264-2193-3. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/101802.html> (дата обращения: 09.04.2020).

4. Яцков, Н. Н. Интеллектуальный анализ данных: пособие / Н. Н. Яцков. — Минск: БГУ, 2014. — 151 с

9. Материально-техническое обеспечение дисциплины

Перечень материально-технических средств учебной аудитории для проведения занятий по дисциплине:

- учебная аудитория, доска;
- стационарные компьютеры;
- мультимедийный проектор;
- настенный экран.

Составитель:

Старший преподаватель кафедры
«Информационные технологии»



/ Шабазов И. М. /

СОГЛАСОВАНО:

Зав. выпускающей кафедрой
«Информационные технологии»



/ Моисеенко Н.А. /

Директор ДУМР



/ Магомаева М.А. /