

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«ДАГЕСТАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Факультет управления

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
НАУКА О ДАННЫХ ДЛЯ БИЗНЕСА

Кафедра «Бизнес-информатика и высшая математика»
факультета управления

Образовательная программа
38.04.05 Бизнес-информатика

Профиль подготовки
Моделирование и оптимизация бизнес-процессов

Уровень высшего образования
магистратура

Форма обучения
Очно-заочная

Статус дисциплины: входит в обязательную часть ОПОП

Махачкала, 2021

Рабочая программа дисциплины «Наука о данных для бизнеса» составлена в 2021 году в соответствии с требованиями ФГОС ВО по направлению подготовки 38.04.05 Бизнес-информатика (уровень магистратуры), утвержденного приказом Минобрнауки РФ от «12» августа 2020г. №990

Разработчик: кафедра бизнес-информатики и высшей математики, Дадаева Б.Ш. к.э.н., доцент.

Рабочая программа дисциплины одобрена:
на заседании кафедры бизнес-информатики и высшей математики
от «2» 06 2021 г., протокол № 10

Зав. кафедрой  Омарова Н.О.

на заседании Методической комиссии факультета управления от
«1» 07 2021 г., протокол № 10

Председатель  Гашимова Л.Г.

Рабочая программа дисциплины согласована с учебно-методическим
управлением

«08» 07 2021 г.



Гасангаджиева А. Г.

Содержание

Аннотация рабочей программы дисциплины.....	4
1. Цели освоения дисциплины.....	5
2. Место дисциплины в структуре ОПОП магистратуры.....	5
3. Компетенции обучающегося, формируемые в результате освоения дисциплины (перечень планируемых результатов обучения).....	5
4. Объем, структура и содержание дисциплины.....	6
4.1. Объем дисциплины.....	6
4.2. Структура дисциплины.....	6
4.3. Содержание дисциплины, структурированное по темам (разделам).....	7
5. Образовательные технологии.....	13
6. Учебно-методическое обеспечение самостоятельной работы студентов.....	14
7. Фонд оценочных средств для проведения текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины.....	15
7.1. Типовые контрольные задания.....	15
7.2. Методические материалы, определяющие процедуру оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.....	18
8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины.....	19
9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины.....	20
10. Методические указания для обучающихся по освоению дисциплины.....	21
11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем.....	22
12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.....	22

Дисциплина «Наука о данных для бизнеса» входит в обязательную часть ОПОП магистратуры по направлению подготовки 38.04.05 Бизнес-информатика, магистерская программа Моделирование и оптимизация бизнес-процессов.

Дисциплина реализуется на факультете управления кафедрой бизнес-информатики и высшей математики.

Содержание дисциплины охватывает круг вопросов, связанных с интеллектуальным анализом данных, а именно, его применением к большим массивам данных. Курс дает практические рекомендации по разработке и использованию систем обработки и анализа больших массивов данных. Студенты получают знания и навыки, необходимые для постановки задачи анализа данных; предварительной обработки данных; разработки, реализации и применения методов интеллектуального анализа данных к анализу и обработке больших массивов данных для создания соответствующего инструмента и представления результатов работы. Курс адресован магистрам, обучающимся по направлению подготовки «Бизнес-информатика».

Дисциплина нацелена на формирование следующих компетенций выпускника: универсальных – УК-1; УК-2; общепрофессиональных – ОПК-1; ОПК-3; ОПК-5.

Преподавание дисциплины предусматривает проведение следующих видов учебных занятий: лекции, практические занятия и самостоятельная работа.

Рабочая программа дисциплины предусматривает проведение следующих видов контроля успеваемости: текущего контроля в форме контрольной работы, тестов, устного опроса и промежуточного контроля в форме экзамена.

Объем дисциплины 3 зачетные единицы, в том числе в академических часах 108 ч. по видам учебных занятий.

Семестр	Учебные занятия								Форма промежуточной аттестации (зачет, дифференцированный зачет, экзамен)
	Всего	в том числе							
		Контактная работа обучающихся с преподавателем						СРС, в том числе экзамен	
		Всего	из них						
			Лекции	Лабораторные занятия	Практические занятия	КСР	консультации		
1	108	12	6	-	6	-	-	96	Экзамен

1. Цели освоения дисциплины

Цель дисциплины – теоретическая и практическая подготовка студентов к работе с большими данными. Знания, полученные в результате освоения дисциплины, помогут при сборе и анализе огромных объемов структурированной или неструктурированной информации, при разработке моделей данных и получении новых знаний.

Задачи освоения дисциплины:

- приобретение студентами знаний о технологиях подготовки, хранения, обработки и анализа больших данных;
- применение статистических и математических методов для анализа больших объемов информации;
- разработка математических моделей защищаемых процессов и средств защиты информации и систем, обеспечивающих информационную безопасность объектов.

2. Место дисциплины в структуре ОПОП магистратуры

Дисциплина «Наука о данных для бизнеса» входит в обязательную часть ОПОП магистратуры по направлению 38.04.05 Бизнес-информатика, профиль подготовки: «Моделирование и оптимизация бизнес-процессов».

Дисциплина «Наука о данных для бизнеса» базируется на знаниях основ таких дисциплин, как: Интеллектуальные предприятия; Анализ данных.

Изучение данной дисциплины должно предшествовать изучению дисциплин: Системный анализ; Управление жизненным циклом ИС (продвинутый уровень); Система сбалансированных показателей в оптимизации бизнес-процессов.

3. Компетенции обучающегося, формируемые в результате освоения дисциплины (перечень планируемых результатов обучения)

Код компетенции из ФГОС ВО	Формулировка компетенции из ФГОС ВО	Планируемые результаты обучения (показатели достижения заданного уровня освоения компетенций)	Процедура освоения

<p>УК-1 Системное и критическое мышление</p>	<p>УК-1.И-1. Анализирует проблемную ситуацию как целостную систему, выявляя ее составляющие и связи между ними. УК-1.И-2. Разрабатывает варианты решения проблемной ситуации на основе критического анализа доступных источников информации</p>	<p>Знает виды, методы и концепции критического анализа. Умеет применять виды, методы и концепции критического анализа при выработке плана действий в проблемных ситуациях. Владеет основными принципами, определяющими цель и стратегию решения сложных ситуаций. .</p>	<p>Устный опрос, письменный опрос; тестирование, реферат</p>
<p>УК-2. Способен управлять проектом на всех этапах его жизненного цикла</p>	<p>УК-2.И-1. Понимает принципы проектного подхода к управлению. УК-2.И-2. Демонстрирует способность управления проектами.</p>	<p>Знает: организационные и технологические методы, принципы и инструменты, используемые в проектной работе, методы, критерии и параметры представления, описания и оценки результатов/продуктов проектной деятельности. Умеет: разрабатывать техническое задание проекта, его план-график, составляет, проверяет и анализирует проектную документацию, составляет и представляет результаты проекта в виде отчетов, статей, выступлений на конференциях, организывает и координирует работу участников проекта. Владеет: навыками эффективной организации и координации этапов реализуемого проекта с целью достижения наилучшего результата при балансировании между объемом работ и ресурсами</p>	<p>Устный опрос, письменный опрос; тестирование, реферат</p>

<p>ОПК-1. Способен разрабатывать стратегию развития информационных технологий и инфраструктуры предприятия и управлять ее реализацией.</p>	<p>ОПК-1.И-1. Обладает навыками разработки ИТ стратегии предприятия. ОПК-1.И-2. Способен определить портфель проектов, реализующих ИТ-стратегию предприятия. ОПК-1.И-3. Обеспечивает согласование ИТ-стратегии с бизнес-стратегией.</p>	<p>Знает: подходы к построению ИТ стратегии компании. Умеет: разрабатывать основные элементы ИТ стратегии компании Владеет: навыками: аналитическими, управленческими, понимает специфики построения ИТ инфраструктуры компании. Осуществляет разработку стратегии в области информационных технологий за счет использования современных информационных систем и ресурсов, позволяющих решать весь комплекс задач на всех уровнях управления, учитывающих особенности системы управления, цели и перспективы развития предприятия.</p>	<p>Устный опрос, письменный опрос; тестирование, реферат</p>
<p>ОПК-3. Способен принимать решения, осуществлять стратегическое планирование и прогнозирование в профессиональной деятельности с использованием современных методов и программного инструментария сбора, обработки и анализа данных, интеллектуального оборудования и систем искусственного интеллекта.</p>	<p>ОПК-3.И-1. Владеет основными техниками бизнес-анализа ОПК-3.И-2. Проектирует альтернативные решения. ОПК-3.И-3. Выявляет возможности, создаваемые информационными и цифровыми технологиями</p>	<p>Знает: методы сбора и анализа информации для решения экономических задач с применением ИКТ Умеет: осуществлять сбор и анализ данных, необходимых для решения экономических задач с использованием современных ИКТ, в тч. ИИ Владеет: навыками анализа результатов реализации проектов использованием ИКТ, в тч ИИ Выбирает оптимальные пути достижения цели и решения поставленной задачи в соответствии с конкретной экономической проблемой с использованием современных методов и</p>	<p>Устный опрос, письменный опрос; тестирование, реферат</p>

		программного инструментария сбора, обработки и анализа данных. Методами анализа соответствия бизнес-процессов и ИТ-инфраструктуры стратегиям и целям предприятия;	
ОПК-5 Способен проводить исследования, организовывать самостоятельную и коллективную научно-исследовательскую, проектную и учебно-профессиональную деятельность для поиска, выработки и применения новых решений в области информационно-коммуникационных технологий.	ОПК-5.И-1. Формулирует исследовательскую задачу и обеспечивает ее последующее решение. ОПК-5.И-2. Критически оценивает результаты научных исследований, проводит анализ, систематизирует и оценивает результаты научных исследований. ОПК-5.И-3. Систематизирует и обобщает результаты отечественных и зарубежных исследований актуальных проблем бизнес-информатики и смежных наук.	Знает: и применяет на практике теоретические и эмпирические научные методы для выявления потребностей бизнеса и общества в совершенствовании и разработке новых ИКТ, продуктов и услуг на их основе; Умеет: выбирать методы исследования и обработки результатов. осуществлять самостоятельный сбор данных, изучать, анализировать и обобщать научно-техническую информацию по тематике исследования Владеет: умением и навыками самостоятельного использования современных информационно-коммуникационных технологий, глобальных информационных ресурсов в научно-исследовательской и расчетно-аналитической деятельности в области ИКТ.	Устный опрос, письменный опрос; тестирование; реферат

4. Структура и содержание дисциплины

4.1. Объем дисциплины составляет 3 зачетных единиц, 108 академических часов.

4.2. Структура дисциплины.

Форма обучения – очно-заочная
1 семестр

№ п/п	Разделы и темы дисциплины	Семестр	Неделя семестра	Виды учебной работы, включая самостоятельную работу студентов и трудоемкость (в часах)			Самостоятельная работа	Формы текущего контроля успеваемости (<i>по неделям семестра</i>) Форма промежуточной аттестации (<i>по семестрам</i>)
				Лекции	Практическ ие/лаборатор ные занятия	Контроль		
	Модуль 1. Основы больших данных							
1	Введение в большие данные.	1	1-3	1	1		10	Опрос, участие в дискуссиях, представление докладов,
2	Жизненный цикл аналитики данных	1	4-7	1	1		10	Опрос, участие в дискуссиях, представление докладов
3.	Роль корреляции и регрессии в аналитике больших данных	1	8- 10	2	2		8	Опрос, участие в дискуссиях, представление докладов
	<i>Итого по модулю 1:36</i>	1	1-9	4	4	-	28	Контрольная работа
	Модуль 2. Статистические методы анализа данных							
4	Визуализация данных и результатов анализа	1	10- 12	1	1		16	Опрос, участие в дискуссиях, представление докладов,
5	Статистические методы анализа данных	1	13- 17	1	1		16	Опрос, участие в дискуссиях, представление докладов,
	<i>Итого по модулю 2:36</i>	1	10-17	2	2	-	32	Контрольная работа
	Промежуточный контроль	1	-	-	-	-		Экзамен
	ИТОГО: 108	1	1-17	6	6	-	60	36

4.3. Содержание дисциплины, структурированное по темам (разделам).

4.3.1. Содержание лекционных занятий по дисциплине

*Модуль 1. Основы больших данных***Тема 1. Введение в большие данные**

Большие данные: определение, Big Data терминология История появления термина Big Data, Причины появления Big Data. Характеристики Big Data, Рост объемов информации. Источники Big Data, Четыре основных типа данных, Аналитика данных. Задачи, решаемые Big Data, Статистика объема данных. Big data для операторов связи, Big data в банках, Европейские программы, связанные с Big Data.

Тема 2. Жизненный цикл аналитики данных

Введение, Понятие жизненного цикла аналитики данных. Business Intelligence , BI-платформа, ETL – процесс, Средства BI. OLAP, OLAP - Многомерный куб, Пример многомерного куба Продвинутой визуализация, Предиктивное моделирование и DataMining. Инструменты анализа. Power Query, MS Power BI. Pyramid Analytics, Компоненты аналитики MS SQL server (MDS, SSIS, SSAS). Business Intelligence vs. Data Science.

Тема 3. Роль корреляции и регрессии в аналитике больших данных

Корреляция и регрессионный анализ. Коэффициент корреляции. Графическое представление. Постановка задачи регрессионного анализа. Линейная регрессия. Метод наименьших квадратов и их роль в аналитике больших данных

Модуль 2. Статистические методы анализа данных

Тема 4. Визуализация данных и результатов анализа

Типы визуализации. Задачи визуализации. Требования к визуализации, Традиционные виды визуализации. Графики и диаграммы, Инфографика. Презентация и анализ данных, Интерактивный сторителлинг. Дашборды и бизнес аналитика. Визуализация в медицине и науке. Карты и картограммы, Облако тегов, кластерграмма. Исторический поток. Пространственный поток Язык R, Отличия языка R, Основные возможности R. Типы данных, Amazon S3, Достоинства Amazon S3. Многокомпонентная загрузка, Особенности хранения в S3. Дедупликация данных

Тема 5. Статистические методы анализа данных

Статистические гипотезы, Статистические критерии. Машинное обучение. Метрический классификатор. Линейный классификатор, ROC – кривая. Кластерный анализ. Алгоритм K-means, Алгоритм C-means. Поиск ассоциативных правил, Свойство анти- монотонности

Темы семинарских и практических занятий

Модуль 1. Основы больших данных

Тема 1. Введение в большие данные

Цель семинарского (практического) занятия – изучить теоретические и методологические основы больших данных и их информационное обслуживание, понять и усвоить основные принципы появления больших данных, чтобы использовать их при обработке и анализе больших объемов информации.

Вопросы:

Большие данные: определение, Big Data терминология

1. История появления термина Big Data, Причины появления Big Data
2. Задачи, решаемые Big Data, Статистика объёма данных
3. Big data для операторов связи, Big data в банках, Европейские программы, связанные с Big Data.

Литература:

1. Статистическое управление процессами. «Большие данные»: учебное пособие / Ю. П. Адлер, Е. А. Черных. — М. : Издательский Дом МИСиС, 2016 — 52 с. — ISBN 978-5-87623-969-3. — Текст : электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. — URL: <http://www.iprbookshop.ru/64199.html>
2. Data Mining – технологии обработки больших данных : учебное пособие / В. И. Воронов, Л. И. Воронова, В. А. Усачев. — М. : 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>
3. Big Data. Методы и средства анализа : учебное пособие / Л. И. Воронова, В. И. Воронов. — М. : Московский технический университет связи и информатики, 2016 — 33 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/61463.html>
4. Анализ данных: учебник для академического бакалавриата / В. С. Мхитарян [и др.]; под ред. В. С. Мхитаряна. М.: Юрайт, 2016. — 490 с. — ISBN 978-5-9916-5591-0.
5. Барсегян, А.А. Анализ данных и процессов: учеб. пособие / А.А. Барсегян, М.С. . - Купприанов, И.И. Холод, М.Д. Тесс, С.И. Елизаров. - 3-е изд., перераб. и доп. СПб.: БХВПетербург, 2009. – 512 с. - Режим доступа: https://books.google.ru/books?id=tHEhjlgo5gAC&printsec=frontcover&hl=ru&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

Тема 2. Жизненный цикл аналитики данных

Цель семинарского (практического) занятия – ознакомить студентов с теоретическими и методологическими основами жизненного цикла данных, рассмотреть концепцию жизненного цикла и принципы, на которых она базируется, чтобы использовать их при изучении вопросов жизненного цикла аналитики данных.

Вопросы:

1. Введение, Понятие жизненного цикла аналитики данных
2. Business Intelligence , BI-платформа, ETL – процесс, Средства BI
3. Продвинутая визуализация, Предиктивное моделирование и DataMining
4. Инструменты анализа

Литература:

1. Статистическое управление процессами. «Большие данные»: учебное пособие / Ю. П. Адлер, Е. А. Черных. — М. : Издательский Дом МИСиС, 2016 — 52 с. — ISBN 978-5-87623-969-3. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/64199.html>
2. Data Mining – технологии обработки больших данных : учебное пособие / В. И. Воронов, Л. И. Воронова, В. А. Усачев. — М. : 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>
3. Big Data. Методы и средства анализа : учебное пособие / Л. И. Воронова, В. И. Воронов. — М. : Московский технический университет связи и информатики, 2016 — 33 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/61463.html>
5. Майер-Шенбергер, В. Большие данные. Революция которая изменит, то как мы живем, работаем и мыслим / Виктор Майер-Шенбергер, Кеннет Кукьер : пер. с англ. Инны Гайдюк. - М.: Манн, Иванов и Фербер, 2014. – 240 с. – Режим доступа: https://books.google.ru/books?id=n5RTAgAAQBAJ&printsec=frontcover&hl=ru&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false
6. Сэнди Риза, Ури Лезерсон, Шон Оуэн, Джош Уиллс. Spark для профессионалов. Современные паттерны обработки больших данных. – СПб.: Питер, 2017. – 272 с. – Режим доступа:

https://books.google.ru/books?id=xRSLDQAAQBAJ&printsec=frontcover&hl=ru&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

Тема 3. Роль корреляции и регрессии в аналитике больших данных

Цель семинарского (практического) занятия – ознакомить студентов с теоретическими и методологическими основами объективно существующих связей между явлениями, а также получение моделей зависимостей для практического использования, понять и усвоить основные принципы метода наименьших квадратов и их роль в аналитике больших данных. чтобы использовать их при изучении вопросов корреляции и регрессии.

Вопросы:

1. Коэффициент корреляции.
2. Линейная регрессия.
3. Метод наименьших квадратов и их роль в аналитике больших данных

Литература:

1. Статистическое управление процессами. «Большие данные»: учебное пособие / Ю. П. Адлер, Е. А. Черных. — М. : Издательский Дом МИСиС, 2016 — 52 с. — ISBN 978-5-87623-969-3. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/64199.html>
2. Data Mining – технологии обработки больших данных : учебное пособие / В. И. Воронов, Л. И. Воронова, В. А. Усачев. — М. : 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>
3. Big Data. Методы и средства анализа : учебное пособие / Л. И. Воронова, В. И. Воронов. — М. : Московский технический университет связи и информатики, 2016 — 33 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/61463.html>
4. Вуколов Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL: учебное пособие для студентов вузов. – М.: ФОРУМ: ИНФРА-М, 2013 – 464 с.
5. Гмурман В.Е. Теория вероятностей и математическая статистика: учебное пособие для бакалавров. - М.: Юрайт: ИД Юрайт, 2012 – 479 с.

Модуль 2. Статистические методы анализа данных

Тема 4. Визуализация данных и результатов анализа

Цель семинарского (практического) занятия – изучить теоретические и методологические основы визуализации данных, понять и усвоить основные принципы построения графиков и появления инфографики и их роль в аналитике больших данных, чтобы использовать их при изучении визуализации данных и результатов анализа.

Вопросы:

1. Типы задачи визуализации.
2. Графики и диаграммы, инфографика.
3. Визуализация в медицине и науке.
4. Карты и картограммы, Облако тегов, кластерграмма.
5. Исторический поток и пространственный поток

Литература:

1. Статистическое управление процессами. «Большие данные»: учебное пособие / Ю. П. Адлер, Е. А. Черных. — М. : Издательский Дом МИСиС, 2016 — 52 с. — ISBN 978-5-87623-969-3. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/64199.html>
2. Data Mining – технологии обработки больших данных : учебное пособие / В. И. Воронов, Л. И. Воронова, В. А. Усачев. — М. : 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>
3. Big Data. Методы и средства анализа : учебное пособие / Л. И. Воронова, В. И. Воронов. — М. : Московский технический университет связи и информатики, 2016 — 33 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/61463.html>
4. Анализ данных: учебник для академического бакалавриата / В. С. Мхитарян [и др.]; под ред. В. С. Мхитаряна. М.: Юрайт, 2016. — 490 с. — ISBN 978-5-9916-5591-0.
5. Барсегян, А.А. Анализ данных и процессов: учеб. пособие / А.А. Барсегян, М.С. . - Куприянов, И.И. Холод, М.Д. Тесс, С.И. Елизаров. - 3-е изд., перераб. и доп. СПб.: БХВПетербург, 2009. – 512 с. - Режим доступа: https://books.google.ru/books?id=tHEhjlgo5gAC&printsec=frontcover&hl=ru&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false
6. Вуколов Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL: учебное пособие для студентов вузов. – М.: ФОРУМ: ИНФРА-М, 2013 – 464 с.
7. Гмурман В.Е. Теория вероятностей и математическая статистика: учебное пособие для бакалавров. - М.: Юрайт: ИД Юрайт, 2012 – 479 с.

Тема 5. Статистические методы анализа данных

Цель семинарского (практического) занятия – ознакомить студентов с теоретическими и методологическими основами статистических методов сбора, обработки и анализа массивов данных, выработать навыки использования современных информационных технологий и программных средств для интерпретации полученных результатов и их роль в аналитике больших данных.

Вопросы:

1. Статистические гипотезы
2. Статистические критерии. Машинное обучение.
3. Метрический классификатор. Линейный классификатор

Литература:

1. Статистическое управление процессами. «Большие данные»: учебное пособие / Ю. П. Адлер, Е. А. Черных. — М. : Издательский Дом МИСиС, 2016 — 52 с. — ISBN 978-5-87623-969-3. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/64199.html>
2. Data Mining – технологии обработки больших данных : учебное пособие / В. И. Воронов, Л. И. Воронова, В. А. Усачев. — М. : 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>
3. Big Data. Методы и средства анализа : учебное пособие / Л. И. Воронова, В. И. Воронов. — М. : Московский технический университет связи и информатики, 2016 — 33 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/61463.html>
4. Теория вероятностей и математическая статистика: учебное пособие для студентов вузов / под ред. В.И. Ермакова. – М.: ИНФРА-М, 2012 – 287 с.
11. Шумак О.А., Гераськин А.В. Статистика: учебное пособие для студентов вузов. – М.: РИОР: ИНФРА-М, 2012 – 311 с.
5. Фрэнкс, Билл. Укрощение больших данных: как извлекать знания из массивов информации . с помощью глубокой аналитики / Билл Фрэнкс ; пер. с англ. Андрея Баранова. - М. : Манн, Иванов и Фербер, 2014. – 352 с. – Режим доступа: <https://books.google.ru/books?id=Quo2BAAAQBAJ&printsec=frontcover&hl=ru#v=onepage&q&f=false>

5. Образовательные технологии

С целью формирования и развития профессиональных навыков обучающихся в соответствии с требованиями ФГОС ВО по направлению

подготовки предусматривается широкое использование в учебном процессе активных и интерактивных форм проведения занятий.

Лекции проводятся с использованием средств визуализации лекционного материала (мультимедийных презентаций) и применением таких методов и технологий, как дискуссия, проблемная лекция и т.п.

При ведении практических занятий по данной дисциплине используются такие стандартные методы обучения, как тестирование, фронтальный опрос, индивидуальный опрос, выполнение кейс-заданий, метод малых групп и т.п.

При проведении практических занятий в интерактивной форме используются следующие методы: дебаты, круглый стол, тематическая групповая дискуссия, блиц-опрос, научный кружок.

Вузовская лекция должна выполнять не только информационную функцию, но также и мотивационную, воспитательную и обучающую.

Информационная функция лекции предполагает передачу необходимой информации по теме, которая должна стать основой для дальнейшей самостоятельной работы студента.

Мотивационная функция должна заключаться в стимулировании интереса студентов к науке. На лекции необходимо заинтересовывать, озадачить студентов с целью выработки у них желания дальнейшего изучения той или иной экономической проблемы.

Воспитательная функция ориентирована на формирование у молодого поколения чувства ответственности, закладку нравственных, этических норм поведения в обществе и коллективе, формирование патриотических взглядов, мотивов социального поведения и действий, финансово-экономического мировоззрения.

Обучающая функция реализуется посредством формирования у студентов навыков работы с первоисточниками и научной и учебной литературой.

6. Учебно-методическое обеспечение самостоятельной работы студентов.

Возрастает значимость самостоятельной работы студентов в межсессионный период. Поэтому изучение курса «Наука данных для бизнеса» предусматривает работу с основной специальной литературой, дополнительной обзорного характера, а также выполнение домашних заданий.

Самостоятельная работа студентов должна способствовать более глубокому усвоению изучаемого курса, формировать навыки исследовательской работы и ориентировать студентов на умение применять теоретические знания на практике.

Задания для самостоятельной работы, их содержание и форма контроля приведены в форме таблицы.

Наименование тем	Содержание самостоятельной работы	Форма контроля
Тема 1. Введение в большие данные	Изучение разделов дисциплины по учебной литературе, в том числе вопросов, не освещаемых на лекциях; подготовка реферата; подготовка к контролю текущих знаний по дисциплине	Опрос, защита реферата,
Тема 2. Жизненный цикл аналитики данных	Изучение разделов дисциплины по учебной литературе, в том числе вопросов, не освещаемых на лекциях; подготовка реферата; подготовка к контролю текущих знаний по дисциплине	Опрос, защита реферата, письменная контрольная
Тема 3. Роль корреляции и регрессии в аналитике больших данных	Изучение разделов дисциплины по учебной литературе, в том числе вопросов, не освещаемых на лекциях; решение задач и тестов; подготовка к контролю текущих знаний по дисциплине	Опрос, проверка заданий, проверка заданий
Тема 4. Визуализация данных и результатов анализа	Изучение разделов дисциплины по учебной литературе, в том числе вопросов, не освещаемых на лекциях; подготовка к контролю текущих знаний по дисциплине	Опрос, защита реферата,
Тема 5. Статистические методы анализа данных	Изучение разделов дисциплины по учебной литературе, в том числе вопросов, не освещаемых на лекциях; подготовка к контролю текущих знаний по дисциплине	Опрос, проверка заданий

7. Фонд оценочных средств для проведения текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины

7.1. Типовые контрольные задания

Текущий контроль успеваемости в форме опросов, рефератов, дискуссий, тестов, решения задач и промежуточный контроль в форме экзамена.

Примерные задания для проведения текущего контроля

1. Метаданные

1.1. Понятие метаданных

- 1.2. Жизненный цикл метаданных
- 1.3. Оценка требований и анализ контента
- 1.4. Спецификация системных требований
- 1.5. Система метаданных
- 1.6. Сервис и оценка

2. Большие данные. Системы управления Большими данными
- 2.1. Распределенные файловые системы
- 2.2. Распределенные фреймворки
- 2.3. Бенчмаркинг
- 2.4. Серверное программирование
- 2.5. Планирование
- 2.6. Системы развертывания
- 2.7. Интеграция данных
- 2.8. Информационная безопасность
- 2.9. Машинное обучение
- 2.10. Базы данных NoSQL и новые SQL базы данных

Примерная тематика рефератов

1. Информационное общество и большие данные.
2. М. МакЛюэн и роль медиа в больших данных.
3. Современные способы хранения и обработки больших данных.
4. Современные способы анализа массивов данных.
5. Современные концепции больших данных.
6. Блокчейн системы как приложение системы Больших данных.
7. Структурированные и неструктурированные данные.
8. Набор признаков Больших данных (VVV-подход).
9. Информационно-технологическая структура Больших данных
10. Прогнозирование качества процесса переработки нефти при помощи машинного обучения
11. Как выбрать подход к управлению ML проектом
12. Гиперавтоматизация глазами программного робота
13. Воспитание моделей
14. Нейронные Сети в Анализе и Генерации Звука
15. Масштабируем Data Science с помощью Modin
16. Качество Или Количество Данных – Что Важнее?
17. Вероятностная Калибровка
18. Оптимизация Нейросетевой Архитектуры Агента, Выполняющего Обучение с Подкреплением
19. NLP или Как Подключить Тысячи Пользователей к Данным
20. Дайте характеристику Big Data на мировом рынке.

21. Охарактеризуйте Big Data в России.
22. Определите понятие Data Mining.
23. Вопросы безопасности больших данных.

Контрольные вопросы к экзамену

для промежуточного контроля

1. Понятие «Большие данные». Роль цифровой информации в 21 веке.
2. Виды массивов данных.
3. Базовые принципы обработки больших данных.
4. Технологии обработки больших данных: NoSQL, MapReduce, Hadoop, R.
5. Технологии Business Intelligence и реляционные системы управления базами данных.
6. Прогнозирование и предвидение: общее и особенное.
7. Виды прогнозов.
8. Общие методы анализа социально-политических и медиа процессов.
9. Специальные методы анализа социально-политических и медиапроцессов.
10. Предварительный анализ данных.
11. Проверка гипотез о законе распределения случайной величины.
12. Статистические оценки параметров. Доверительные области.
13. Теория моментов.
14. Корреляционный анализ.
15. Использование модели множественной линейной регрессии для прогнозирования экономических показателей.
16. Доверительные интервалы для зависимой переменной.
17. Сглаживание временных рядов.
18. Динамические модели с распределенными лагами.
19. Стационарные временные ряды.
20. Тестирование стационарности.
21. Коинтеграция.
22. Анализ временных рядов.
23. Адаптивные и мультипликативные методы прогнозирования.
24. Экспоненциальное сглаживание.
25. Авторегрессионные модели.
26. Модели скользящего среднего.
27. Интегрированные процессы.
28. Идентификация авторегрессионной модели скользящего среднего.
29. Прогнозирование с моделями временных рядов.
30. Доверительные интервалы прогноза.
31. Предсказание и прогнозирование социально-экономических прогнозов.

32. Дисперсионный анализ влияния качественных факторов.
33. Ранговые методы.
34. Факторный анализ. Метод главных факторов.
35. Многомерное шкалирование.
36. Классическая модель многомерного шкалирования.
37. Неметрические методы.
38. Кластерный анализ.
39. Дискриминантный анализ.
40. Многомерный статистический анализ.
41. Определение больших данных, ключевые характеристики.

Основные виды данных.

42. Роль аналитика по данным (Data Scientist). Ключевые компетенции аналитика.
43. Отличия BI от Data Science.
44. Корреляция и регрессионный анализ. Коэффициент корреляции.
45. Графическое представление. Постановка задачи регрессионного анализа.
46. Линейная регрессия. Метод наименьших квадратов. Привести примеры использования регрессионного анализа.
47. Классификация. Признаковое описание объекта и таблица объект-свойства.
48. Визуализация. Дать определение визуализации. Показать важность визуализации в аналитике больших данных. Привести примеры использования визуализации.
49. «Жизненный цикл» проекта по аналитике больших данных.
50. Типовая архитектура проекта в области больших данных. Перечислить используемые технологии, указать степень вовлеченности каждой из технологий на каждом этапе работы над проектом. Перечислить основные роли исполнителей проекта.

7.2. Методические материалы, определяющие процедуру оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.

Общий результат выводится как интегральная оценка, складывающаяся из текущего контроля - 50% и промежуточного контроля - 50%.

Текущий контроль по дисциплине включает:

- посещение занятий – 10 баллов (на каждом занятии),
- участие на семинарских (практических) занятиях, - 60 баллов (на каждом занятии),
- выполнение домашних (аудиторных) контрольных работ – 15 баллов,
- выполнение рефератов – 15 баллов.

Промежуточный контроль проводится в устной форме

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

а) основная литература:

1. Статистическое управление процессами. «Большие данные»: учебное пособие / Ю. П. Адлер, Е. А. Черных. — М. : Издательский Дом МИСиС, 2016 — 52 с. — ISBN 978-5-87623-969-3. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/64199.html>
2. Data Mining – технологии обработки больших данных : учебное пособие / В. И. Воронов, Л. И. Воронова, В. А. Усачев. — М. : 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>
3. Big Data. Методы и средства анализа : учебное пособие / Л. И. Воронова, В. И. Воронов. — М. : Московский технический университет связи и информатики, 2016 — 33 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/61463.html>

б) дополнительная литература

4. Анализ данных: учебник для академического бакалавриата / В. С. Мхитарян [и др.]; под ред. В. С. Мхитаряна. М.: Юрайт, 2016. — 490 с. — ISBN 978-5-9916-5591-0.
5. Барсегян, А.А. Анализ данных и процессов: учеб. пособие / А.А. Барсегян, М.С. . - Куприянов, И.И. Холод, М.Д. Тесс, С.И. Елизаров. - 3-е изд., перераб. и доп. СПб.: БХВПетербург, 2009. – 512 с. - Режим доступа: https://books.google.ru/books?id=tHEhJlgo5gAC&printsec=frontcover&hl=ru&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false
6. Вуколов Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL: учебное пособие для студентов вузов. – М.: ФОРУМ: ИНФРА-М, 2013 – 464 с.
7. Гмурман В.Е. Теория вероятностей и математическая статистика: учебное пособие для бакалавров. - М.: Юрайт: ИД Юрайт, 2012 – 479 с.
8. Майер-Шенбергер, В. Большие данные. Революция которая изменит, то как мы живем, работаем и мыслим / Виктор Майер-Шенбергер, Кеннет Кукьер : пер. с англ. Инны Гайдюк. - М.: Манн, Иванов и Фербер, 2014. – 240 с. – Режим доступа: https://books.google.ru/books?id=n5RTAgAAQBAJ&printsec=frontcover&hl=ru&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

9. Сэнди Риза, Ури Лезерсон, Шон Оуэн, Джош Уиллс. Spark для профессионалов. Современные паттерны обработки больших данных. – СПб.: Питер, 2017. – 272 с. – Режим доступа:
https://books.google.ru/books?id=xRSLDQAAQBAJ&printsec=frontcover&hl=ru&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false
10. Теория вероятностей и математическая статистика: учебное пособие для студентов вузов / под ред. В.И. Ермакова. – М.: ИНФРА-М, 2012 – 287 с.
11. Шумак О.А., Гераськин А.В. Статистика: учебное пособие для студентов вузов. – М.: РИОР: ИНФРА-М, 2012 – 311 с.
12. Фрэнкс, Билл. Укрощение больших данных: как извлекать знания из массивов информации . с помощью глубокой аналитики / Билл Фрэнкс ; пер. с англ. Андрея Баранова. - М. : Манн, Иванов и Фербер, 2014. – 352 с. – Режим доступа:
<https://books.google.ru/books?id=Quo2BAAAQBAJ&printsec=frontcover&hl=ru#v=onepage&q&f=false>

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

1. Для расчетов статистических показателей возможно использование программы «Excel».
2. Интернет-ресурсы:
 Поисковая система. – URL: <http://www.rambler.ru>.
 Поисковая система. – URL: <http://www.yandex.ru> .
 Поисковая система. – URL: <http://www.google.ru> .
<http://www.gks.ru> – ФСГС РФ
<http://www.un.org/popin> - сайт ООН
<http://www.cbr.ru> – Центральный банк России
<http://www.imf.org> – сайт Международного валютного фонда
<http://www.wto.org> – Всемирная торговая организации
<https://rosmintrud.ru> – Министерство труда и социальной защиты РФ
<http://www.cisstat.com> – Статистический комитет СНГ
<http://www.undp.ru> – Программа развития ООН.
3. Справочно-правовая система «КонсультантПлюс» [Электронный ресурс] – URL: <http://www.consultant.ru> (дата обращения 08.06.2018).
4. Информационно-правовой портал «Гарант.ру» [Электронный ресурс] – URL: <http://www.garant.ru> (дата обращения 05.06.2018).
5. Электронный каталог НБ ДГУ [Электронный ресурс]: база данных содержит сведения о всех видах литературы, поступающих в фонд НБ ДГУ/Дагестанский гос. ун-т. – Махачкала, 2018. – URL: <http://elib.dgu.ru> (дата обращения 21.05.2021).
6. eLIBRARY.RU[Электронный ресурс]: электронная библиотека / Науч. электрон. б-ка. — Москва. – URL: <http://elibrary.ru/defaultx.asp> (дата обращения 05.02.2018).

7. Moodle [Электронный ресурс]: система виртуального обучения: [база данных] / Даг. гос. ун-т. – г. Махачкала. – Доступ из сети ДГУ или, после регистрации из сети ун-та, из любой точки, имеющей доступ в интернет. – URL: <http://moodle.dgu.ru/> (дата обращения 21.03.2021).
8. Электронно-библиотечная система «Университетская библиотека онлайн» (<http://biblioclub.ru/>).

10. Методические указания для обучающихся по освоению дисциплины

Для изучения теоретического курса студентам необходимо использовать лекционный материал, учебники и учебные пособия из списка основной и дополнительной литературы, статьи периодических изданий.

Для развития самостоятельности, умения решать творческие задачи крайне важно научить студентов работать с литературой, находить нужные источники, анализировать прочитанное и делать выводы.

Для теоретического и практического усвоения дисциплины большое значение имеет самостоятельная работа студентов, которая может осуществляться студентами индивидуально и под руководством преподавателя.

Самостоятельная работа студентов, предусмотренная учебным планом в объеме 60 часов, направлена на более глубокое усвоение изучаемого курса, формирование навыков исследовательской работы и ориентирование студентов на умение применять теоретические знания на практике.

Основные направления самостоятельной работы: самостоятельное изучение отдельных тем, доказательства статистических формул, работа со справочниками, участие в научно-исследовательской работе и в научно-студенческих конференциях

Организация самостоятельной работы требует контроля не столько регистрирующего, сколько текущего, поскольку только текущий контроль определяет "болевые точки" обучения, корректируя учебный процесс в нужном направлении. В своей работе мы практикуем такие методы и способы контроля как: устный опрос, короткая письменная работа, контрольная работа, аттестация.

Задания для самостоятельной работы составляются по разделам и темам, по которым не предусмотрены аудиторские занятия, либо требуют дополнительной проработки и анализа материала в объеме запланированных часов.

В зависимости от места и времени проведения, характера руководства со стороны преподавателя и формы контроля, СРС подразделяется на следующие виды:

- самостоятельную работу во время основных аудиторных занятий (лекций, семинаров, лабораторных работ);
- самостоятельную работу под контролем преподавателя в форме плановых консультаций, творческих контактов, зачетов и экзаменов;

– внеаудиторную самостоятельную работу при выполнении студентом домашних заданий учебного и творческого характера (индивидуальных работ по вариантам, рефератов, аналитических записок, сообщений и т.п.)

Самостоятельная работа студентов может осуществляться в виде:

- конспектирования учебной, научной и периодической литературы;
- проработки учебного материала (по конспектам лекций учебной и научной литературы);
- подготовки докладов к семинарам и практическим занятиям, участию в тематических дискуссиях, работе научного кружка и конференциях;
- работы с нормативными документами и законодательной базой, с первичными документами;
- поиска и обзора научных публикаций и электронных источников информации, подготовки заключения по обзору информации;
- выполнения контрольных работ, творческих (проектных) заданий;
- решения практических и ситуационных задач;
- составления аналитических таблиц, графического оформления материала;
- написания рефератов, тезисов докладов;
- работы с тестами и контрольными вопросами для самопроверки;
- обработки и анализа статистической информации;
- моделирования и анализа конкретных проблемных ситуаций;
- написания выводов и предложений на основе проведенного анализа.

Самостоятельная работа должна носить систематический характер, быть интересной и привлекательной для студентов.

Результаты самостоятельной работы контролируются преподавателем и учитываются при модульной и промежуточной аттестации обучающегося (экзамен). При этом проводятся: тестирование, экспресс-опрос на семинарских и практических занятиях, заслушивание докладов, проверка письменных работ и т.д. Оценка самостоятельной работы проводится по каждому дисциплинарному модулю дисциплины в рамках общей системы ранжирования оценки знаний по курсу.

11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем.

При подготовке к практическим занятиям, а также при написании рефератов могут использоваться поисковые сайты сети «Интернет», информационно-справочная система «Консультант+», а также Интернет-ресурсы, перечисленные в разделе 9 данной программы. Кроме того, может использоваться учебный курс, размещенный на платформе Moodle ДГУ, <http://moodle.dgu.ru/> (автор-разработчик Дадаева Б.Ш.) и другие учебные курсы, размещенные на указанной платформе, а также учебные материалы, размещенные на образовательном блоге Дадаевой Б.Ш. <http://barstatistics.blogspot.com/>. Для проведения индивидуальных консультаций может использоваться электронная почта.

12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.

На факультете управления Дагестанского государственного университета имеются аудитории (405, 408, 409, 421, 434), оборудованные интерактивными, мультимедийными досками, проекторами, что позволяет читать лекции в формате презентаций, разработанных с помощью пакета прикладных программ MS Power Point, использовать наглядные, иллюстрированные материалы, обширную информацию в табличной и графической формах, пакет прикладных обучающих программ, а также электронные ресурсы сети Интернет.