

**Компонент ОПОП 09.03.01 Информатика и вычислительная техника
Направленность (профиль) Технологии разработки веб-приложений**

Б1.В.01.02

РАБОЧАЯ ПРОГРАММА

**Дисциплины
(модуля)**

Первичная обработка и анализ данных

Разработчик:
Лазарева Ирина Михайловна,
доцент кафедры информационных
технологий,
канд. физ.-мат. наук, доцент

Утверждено на заседании кафедры
Информационных технологий
протокол № 6 от 01.02.2024

Заведующий кафедрой ИТ



О.И. Ляш

**Мурманск
2024**

Пояснительная записка

Объем дисциплины 3 з.е.

1. Результаты обучения по дисциплине (модулю), соотнесенные с индикаторами достижения компетенций, установленными образовательной программой

Компетенции	Индикаторы достижения компетенций	Результаты обучения по дисциплине (модулю)
<p>ПК-2 Способен разрабатывать модели компонентов информационных систем, включая модели баз данных и модели пользовательских интерфейсов</p>	<p>ИД-2_{ПК-2} Формулирует задачи в рамках проекта и определяет ожидаемые результаты ИД-3_{ПК-2} Умеет грамотно отбирать значимые данные ИД-4_{ПК-2} Обеспечивает модульность выполнения задачи с учетом имеющихся ресурсов ИД-5_{ПК-2} Обеспечивает пользовательскую привлекательность создаваемого программного продукта ИД-6_{ПК-2} Умеет представлять результаты своей деятельности с учетом уровня аудитории</p>	<p><i>Знать:</i></p> <ul style="list-style-type: none"> – типы шкал, в которых могут быть представлены данные, и ограничения на задачи, которые могут быть решены с учётом этих шкал; – основные способы графического и табличного представления данных; – критерии определения аномальных значений в выборочных данных; – основные понятия корреляционного и регрессионного анализа; – основные правила проверки значимости и интервального оценивания уравнения и коэффициентов регрессии; – понятия и методы кластерного и классификационного анализа; – основные виды расстояний между объектами (метрик пространства) и кластерами, их особенности и возможности применения; – формулировку задачи о построении базиса признакового пространства из главных компонент и задачи о снижении размерности пространства; – особенности методов интеллектуального анализа текстовых данных. <p><i>Уметь:</i></p> <ul style="list-style-type: none"> – конкретных данных с учётом их природы, погрешности, пространственного и временного разрешения, а также задач исследования; – правильно понимать и интерпретировать полученные результаты исследования; – уметь критически оценивать возможности и ограничения используемых методов; – применять методы первичной обработки данных; – осуществлять дискретизацию непрерывных данных с учётом решаемой задачи; – выбирать наиболее подходящий способ табличного или графического представления данных, исходя из целей исследования; – проверять наличие статистически значимой линейной связи между переменными; выполнять их параметрическую идентификацию, оценивать качество аппроксимации реальных данных выбранной моделью; – выбирать метод классификации или кластеризации в зависимости от цели исследования и характера имеющихся данных; проводить сравнительный анализ различных способов классификации и кластеризации множества объектов; – использовать сингулярный анализ для получения главных компонент и оценки объяснённой дисперсии каждой компоненты. <p><i>Владеть:</i></p> <ul style="list-style-type: none"> – технологиями поиска данных и оценки их качества;

		<ul style="list-style-type: none"> – навыками сбора первичной информации, организации и хранения данных для конкретного исследования; – навыками самостоятельного проведения исследований; – основными категориями, понятиями, методами современной описательной статистики, методами корреляционного анализа; – технологией получения уравнения регрессии и проверки его статистической значимости; – технологией работы с программными средствами, позволяющими осуществлять статистический анализ; – основными приёмами и методами классификации, кластерного и дискриминантного анализа в зависимости от характера используемой информации; – технологией работы с программными средствами, позволяющими осуществлять кластерный и классификационный анализ⁷
--	--	--

2. Содержание дисциплины (модуля)

Раздел 1. Основные понятия технологий сбора и систематизации данных.

Данные, информация, знания. Цели и задачи анализа данных. Источники данных, мониторинг как информационная технология. Суть многомерности данных. Линейное пространство признаков. Пространственно-временные данные. Основные понятия Data Science.

Раздел 2. Предобработка данных, визуализация, первичный статистический анализ.

Детерминированные и случайные составляющие в данных. Анализ качества данных. Дискретизация, удаление выбросов, заполнение пропусков, фильтрация, сглаживание данных. Выборка и генеральная совокупность. Расчёт описательных статистик выборки.

Раздел 3. Корреляционный и регрессионный анализы.

Цель корреляционного анализа, выборочный парный коэффициент корреляции Пирсона (статистическая значимость, согласованность для генеральной совокупности, интервальная оценка). Цель регрессионного анализа. Метод наименьших квадратов. Получение уравнения одномерной линейной регрессии. Проверка значимости уравнения линейной регрессии.

Раздел 4. Классификация.

Обучение с учителем. Размеченная выборка. Метод ближайших соседей. Деревья решений. Логистическая регрессия. Наивный байесовский классификатор. Метрики качества классификации.

Раздел 5. Кластерный анализ.

Обучение без учителя. Кластерный анализ многомерной выборки: основные понятия, метрики, функционалы и критерии качества, критерии останова. Метод K-средних. Иерархическая кластеризация.

Раздел 6. Метод главных компонент.

Задача снижения размерности данных. Стандартизация матрицы «объект-признак». Сингулярный анализ. Анализ главных компонент и факторный анализ.

Раздел 7. Аффинитивный анализ.

Задача поиска ассоциативных правил. Метод Apriori. Основные показатели метода: поддержка и достоверность.

3. Перечень учебно-методического обеспечения дисциплины (модуля)

- мультимедийные презентационные материалы по дисциплине (модулю) представлены в электронном курсе в ЭИОС МАУ;
- методические указания к выполнению лабораторных/практических/контрольных работ представлены в электронном курсе в ЭИОС МАУ;
- методические материалы для обучающихся по освоению дисциплины (модуля) представлены на официальном сайте МАУ в разделе «Информация по образовательным программам, в том числе адаптированным».

4. Фонд оценочных средств по дисциплине (модулю)

Является отдельным компонентом образовательной программы, разработан в форме отдельного документа, включает в себя:

- перечень компетенций с указанием этапов их формирования в процессе освоения дисциплины (модуля);
- задания текущего контроля;
- задания промежуточной аттестации;
- задания внутренней оценки качества образования.

5. Перечень основной и дополнительной учебной литературы (печатные издания, электронные учебные издания и (или) ресурсы электронно-библиотечных систем).

Основная литература

1. Анализ данных. Учебник для академического бакалавриата/ В.С. Мхитарян - Отв. ред.-М.:Издательство Юрайт: Национальный исследовательский университет "Высшая школа экономики" (НИУ ВШЭ) (г. Москва), 2016. – 490 с.; То же [Электронный ресурс]. – URL: <http://www.urait.ru/book/AF1D197F-1759-422E-9593-8B43E2D1093B>
2. Большаков, А. А. Методы обработки многомерных данных и временных рядов : Учебн. пособие для вузов / А. А. Большаков, Р. Н. Каримов. – 2-е изд., стереотип.. – М. : Горячая линия – Телеком, 2016. – 522 с.

Дополнительная литература:

1. Мхитарян, В.С. Анализ данных : Учебник для академического бакалавриата/ В.С. Мхитарян - Отв. ред.-М.: Издательство Юрайт: Национальный исследовательский университет "Высшая школа экономики" (НИУ ВШЭ) (г. Москва), 2016. – 490 с.; То же [Электронный ресурс]. – URL: <http://www.urait.ru/book/AF1D197F-1759-422E-9593-8B43E2D1093B>
2. Яцков Н. Н. Интеллектуальный анализ данных : учеб. пособие для студ. вузов, обуч. по спец. "Радиофизика", "Прикл. информатика", "Компьютерная безопасность" / Яцков Н. Н. ; Белорусский гос. ун-т. – Минск : БГУ, 2014. – 151 с.

6. Профессиональные базы данных и информационные справочные системы

- 1) Государственная система правовой информации - официальный интернет-портал правовой информации- URL: <http://pravo.gov.ru>
- 2) Информационная система «Единое окно доступа к образовательным ресурсам» - URL: <http://window.edu.ru>
- 3) Справочно-правовая система. Консультант Плюс - URL: <http://www.consultant.ru/>

7. Перечень лицензионного и свободно распространяемого программного

обеспечения, в том числе отечественного производства

- 1) Офисный пакет Microsoft Office 2007
- 2) Система оптического распознавания текста ABBYY FineReader
- 3) Adobe Reader
- 4) FlashPlayer
- 5) Git
- 6) Google Chrome
- 7) K-Lite_Codec_Pack
- 8) LibreOffice.org
- 9) Mozilla FireFox
- 10) Google Chrome
- 11) NetBeans
- 12) Notepad++
- 13) Paint.NET
- 14) Python 3.6 Anaconda
- 15) StarUML
- 16) The Gimp
- 17) VirtualBox
- 18) VLC

8. Обеспечение освоения дисциплины лиц с инвалидностью и ОВЗ

Обучающиеся из числа инвалидов и лиц с ОВЗ обеспечиваются печатными и (или) электронными образовательными ресурсами в формах, адаптированных к ограничениям их здоровья.

9. Материально-техническое обеспечение дисциплины (модуля) представлено в приложении к ОПОП «Материально-технические условия реализации образовательной программы» и включает:

- учебные аудитории для проведения учебных занятий, предусмотренных программой бакалавриата, оснащенные оборудованием и техническими средствами обучения;

- помещения для самостоятельной работы обучающихся, оснащенные компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде МАУ;

Допускается замена оборудования его виртуальными аналогами.

10. Распределение трудоемкости по видам учебной деятельности

Таблица 1 - Распределение трудоемкости

Вид учебной учебной деятельности	Распределение трудоемкости дисциплины (модуля) по формам обучения									
	Очная			Очно-заочная				Заочная		
	Семестр		Всего часов	Семестр		Всего часов	Семестр/Курс		Всего часов	
	7									
Лекции	18									18
Лабораторные работы	36									36
Самостоятельная работа	54									54

Всего часов по дисциплине	108											108
/ из них в форме практической подготовки	18											18

Формы промежуточной аттестации и текущего контроля

зачет	За											
-------	----	--	--	--	--	--	--	--	--	--	--	--

Перечень лабораторных работ по формам обучения

№ п/п	Темы лабораторных работ
1	2
	Очная форма
1	Предобработка данных, визуализация, первичный статистический анализ
2	Корреляционный и регрессионный анализы
3	Метод главных компонент
4	Кластерный анализ методом К-средних
5	Классификация методом К-ближайших соседей
6	Классификация методом построения дерева принятия решений
7	Поиск ассоциативных правил