

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное
учреждение высшего образования
«Мурманский арктический государственный университет»
(ФГБОУ ВО «МАГУ»)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Б1.В.ДВ.05.02 R: программирование и анализ данных

(название дисциплины (модуля) в соответствии с учебным планом)

**основной профессиональной образовательной программы
по направлению подготовки**

**01.03.02 Прикладная математика и информатика
направленность Управление данными и машинное обучение**

(код и наименование направления подготовки
с указанием направленности (наименования магистерской программы))

высшее образование – бакалавриат

уровень профессионального образования: высшее образование – бакалавриат / высшее образование –
специалитет, магистратура / высшее образование – подготовка кадров высшей квалификации

бакалавр

квалификация

очная

форма обучения

2021

год набора

Составитель(и):

Лазарева Ирина Михайловна,
доцент, к.ф.-м.н.,
зав. кафедрой МФиИТ

Утверждено на заседании кафедры
математики, физики и информационных
технологий факультета
математических и естественных наук
(протокол № 07 от 12.04.2021)

Переутверждено на заседании кафедры
математики, физики и информационных
технологий факультета
математических и естественных наук
(протокол № 09 от 02.07.2021)

Зав. кафедрой _____ Лазарева И.М.
подпись Ф.И.О.

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ).

Цель – формирование понимания подходов к основам программирования на языке R, и всех этапов анализа данных.

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

В результате освоения дисциплины (модуля) формируются следующие компетенции:

Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с индикаторами достижения компетенций

Компетенция	Индикаторы компетенций	Результаты обучения
УК-2. Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм и имеющихся ресурсов и ограничений ПК-3: Способен к разработке и применению алгоритмических и программных решений в области системного и прикладного программного обеспечения	УК-2.1 Формулирует в рамках поставленной цели проекта совокупность взаимосвязанных задач, обеспечивающих ее достижение. Определяет ожидаемые результаты решения выделенных задач. УК-2.2 Проектирует решение конкретной задачи исходя из правовых и(или) экономических знаний в различных сферах жизнедеятельности. ПК-3.1 Разрабатывает алгоритм решения поставленной задачи выбранным методом ПК-3.2 Выбирает и обосновывает выбор языковой среды ПК-3.3 Использует современную языковую среду для реализации сложных алгоритмов ПК-3.4 Решает задачу тестирования программного продукта	<i>Знать:</i> <ul style="list-style-type: none">– основные способы получения и обработки информации, необходимой для профессиональной деятельности;– основные парадигмы машинного обучения;– язык программирования R;– принципы и методы разработки системного и прикладного программного обеспечения
		<i>Уметь:</i> <ul style="list-style-type: none">– решать задачи производственной и технологической деятельности на профессиональном уровне, включая разработку алгоритмических и программных решений в области системного и прикладного программирования;– применять методы машинного обучения для решения задач анализа данных;– оценивать качество моделей машинного обучения;
		<i>Владеть:</i> <ul style="list-style-type: none">– навыком исследования и моделирования предметной области;– владеть терминологией машинного обучения и искусственных нейронных сетей;– владеть инструментальными средствами для построения моделей машинного обучения с учителем;– приемами целеполагания во временной перспективе, способами планирования, организации, самоконтроля и самооценки деятельности;– навыков самоподготовки и освоения параллельного направления профессиональной деятельности

3. УКАЗАНИЕ МЕСТА ДИСЦИПЛИНЫ (МОДУЛЯ) В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ.

Дисциплина «R: программирование и анализ данных» относится к обязательным дисциплинам части учебного плана, формируемой участниками образовательных отношений

(Б1.В.ДВ.05.02) по направлению подготовки 01.03.02 Прикладная математика и информатика, направленность (профиль) Управление данными и машинное обучение.

4. ОБЪЕМ ДИСЦИПЛИНЫ (МОДУЛЯ) В ЗАЧЕТНЫХ ЕДИНИЦАХ С УКАЗАНИЕМ КОЛИЧЕСТВА АКАДЕМИЧЕСКИХ ИЛИ АСТРОНОМИЧЕСКИХ ЧАСОВ, ВЫДЕЛЕННЫХ НА КОНТАКТНУЮ РАБОТУ ОБУЧАЮЩИХСЯ С ПРЕПОДАВАТЕЛЕМ (ПО ВИДАМ УЧЕБНЫХ ЗАНЯТИЙ) И НА САМОСТОЯТЕЛЬНУЮ РАБОТУ ОБУЧАЮЩИХСЯ.

Общая трудоемкость дисциплины (модуля) составляет 3 зачетные единицы или 108 часа (из расчета 1 ЗЕ = 36 часов).

Курс	Семестр	Трудоемкость в ЗЕ	Общая трудоемкость (час.)	Контактная работа			Всего контактных часов	Из них:		Кол-во часов на СРС		Кол-во часов на контроль	Форма контроля
				ЛК	ПР	ЛБ		В интерактивной форме*	В форме практической подготовки*	Общее количество часов на СРС	из них – на курсовую работу		
4	7	3	108	18	-	36	54	8	18	54	-	-	зачет
Итого:		3	108	18	-	36	54	8	18	54	-	-	зачет

Интерактивная форма реализуется в виде проблемных лекций и проектной деятельности по темам дисциплины.

Практическая подготовка реализуется в виде лабораторных работ.

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ), СТРУКТУРИРОВАННОЕ ПО ТЕМАМ (РАЗДЕЛАМ) С УКАЗАНИЕМ ОТВЕДЕННОГО НА НИХ КОЛИЧЕСТВА АКАДЕМИЧЕСКИХ ИЛИ АСТРОНОМИЧЕСКИХ ЧАСОВ И ВИДОВ УЧЕБНЫХ ЗАНЯТИЙ.

№ п/п	Наименование раздела, темы	Контактная работа			Всего контактных часов	Из них		Кол-во часов на СРС	Кол-во часов на контроль
		ЛК	ПР	ЛБ		В интерактивной форме	В форме практической подготовки		
Раздел 1. Основы программирования в R									
1.	Введение в R: основные элементы, функции, циклы	2	-	4	6	-	2	-	
2.	Продвинутая обработка данных: пакеты tidyr и dplyr Работа со строками: строки в R, регулярные выражения	2	-	4	8	2	2	10	
3.	Визуализация данных: base R vs. ggplot2 Интерактивная визуализация: rmarkdown, plotly, lingtypology	2	-	4	6	-	2	10	

№ п/п		Контактная работа			КОНТ актн лы	Из них		во часо в	часо в	в
Раздел 2. Анализ данных в R										
4.	Корреляция и линейная регрессия	4	-	8	10	2	4	10		
5.	Кластеризация	4	-	8	12	2	4	12		
6.	Логистическая регрессия	4	-	8	12	2	4	12		
	Зачет									-
	ИТОГО:	18	-	36	54	8	18	54		-

Содержание дисциплины (модуля)

Тема 1. Введение в R: основные элементы, функции, циклы.

Введение. Rstudio, арифметика. Базовые объекты в R. Переменные. Фактор. Векторизованные функции, последовательности, распределения, рандомизатор. Матрицы, списки, конверсия. Data.frame. Функции. Пакеты.

Тема 2. Продвинутой обработке данных: пакеты tidy и dplyr. Работа со строками: строки в R, регулярные выражения.

Пакеты tidy и dplyr. Импортирование данных в R. Получение строк в R. Изменение регистра, подсчет длины строки, сортировка. Скорость сортировки. Соединение строк. Поиск по подстроке. Замена подстроки, разделение подстроки.

Тема 3. Визуализация данных: base R vs. ggplot2. Интерактивная визуализация: rmarkdown, plotly, lingtypology.

Квартет Анскомба. Гистограмма и функция плотности. Rmarkdown: markdown, оформление кода, yaml-шапка, презентация, Java-script библиотеки, Shiny.

Тема 4. Корреляция и линейная регрессия.

Данные, типы переменных, описательные статистики. Работа с пропущенными значениями. Встроенные распределения, правило трёх сигм. Z-преобразование. Визуализация z-преобразования. Доверительные интервалы. Тест Стьюдента. Проверка нормальности. Биномиальный тест. Введение в регрессию. Shiny приложение. Предсказания на основании модели. Ограничения на применение линейной регрессии.

Тема 5. Кластеризация.

Двухвыборочные тесты. Парные тесты. Ошибки первого и второго рода. Мощность теста. Тест хи-квадрат. Поправка на множественные сравнения. Дисперсия, ковариация, коэффициент корреляции

Тема 6. Логистическая регрессия.

Введение в логистическую регрессию. Связь логарифма шансов и вероятности. Логистическая модель без предикторов. Логистическая модель с предикторами. Предсказание на основании модели. Логистическая модель с dummy переменной. Сравнение моделей.

6. ПЕРЕЧЕНЬ УЧЕБНО-МЕТОДИЧЕСКОГО ОБЕСПЕЧЕНИЯ, НЕОБХОДИМОГО ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ).

Основная литература:

1. Анализ данных : учебник для академического бакалавриата / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2018. — 490 с. — (Бакалавр. Академический курс). — ISBN 978-5-534-00616-2. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/412967>
2. Миркин, Б. Г. Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — Москва : Издательство Юрайт, 2018. — 174 с. — (Авторский учебник). — ISBN 978-5-9916-5009-0. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/413060>
3. Чубукова, И.А. Data Mining : учебное пособие : [16+] / И.А. Чубукова. — 2-е изд., испр. — Москва : Интернет-Университет Информационных Технологий (ИНТУИТ) : Бином.

Лаборатория знаний, 2008. – 383 с. – (Основы информационных технологий). – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=233055>. – ISBN 978-5-94774-819-2. – Текст : электронный.

Дополнительная литература:

1. Каган, Е.С. Прикладной статистический анализ данных : учебное пособие : [16+] / Е.С. Каган ; Кемеровский государственный университет. – Кемерово : Кемеровский государственный университет, 2018. – 235 с. : ил., табл. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=573550>. – Библиогр.: с. 184-186. – ISBN 978-5-8353-2413-2. – Текст : электронный.
2. Келлехер, Д. Наука о данных: базовый курс : [16+] / Д. Келлехер, Б. Тирни ; науч. ред. З. Мамедьяров ; пер. с англ. М. Белоголовского. – Москва : Альпина Паблишер, 2020. – 224 с. : схем., табл. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=598235>. – ISBN 978-5-9614-3170-4. – Текст : электронный.
3. Парфенов, Ю. П. Постреляционные хранилища данных : учебное пособие для вузов / Ю. П. Парфенов. — М. : Издательство Юрайт, 2018. — 121 с. — (Серия : Университеты России). — ISBN 978-5-534-03408-0. — Режим доступа : www.biblio-online.ru/book/628DAC6C-ECBF-45B3-BD23-F6B57148D18F.

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ).

В образовательном процессе используются:

- учебные аудитории для проведения учебных занятий, оснащенные оборудованием и техническими средствами обучения: учебная мебель, ПК, оборудование для демонстрации презентаций, наглядные пособия;
- помещения для самостоятельной работы, оснащенные компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду МАГУ.

7.1 ПЕРЕЧЕНЬ ЛИЦЕНЗИОННОГО И СВОБОДНО РАСПРОСТРАНЯЕМОГО ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ:

Лицензионное программное обеспечение:

- Операционная система: MS Windows версии 7 и выше;
- Программные средства, входящие в состав офисного пакета: MS Office (Word, Excel, Access, Publisher, PowerPoint);
- Программное обеспечение: MS Office Visio, MS ACCESS, MS SQL SERVER 2008, Visual Studio 2010.

Свободно распространяемое программное обеспечение:

- Программное обеспечение: MongoDB.
- Программы для просмотра документов: Adobe Acrobat Reader, DJVU Reader;
- Среда логического проектирования структуры базы данных Erwin;
- Браузер: Google Chrome;
- Архиватор: 7Zip.

7.2 ЭЛЕКТРОННО-БИБЛИОТЕЧНЫЕ СИСТЕМЫ:

- ЭБС «Издательство Лань» [Электронный ресурс]: электронная библиотечная система / ООО «Издательство Лань». – Режим доступа: <https://e.lanbook.com/>;
- ЭБС «Электронная библиотечная система ЮРАЙТ» [Электронный ресурс]: электронная библиотечная система / ООО «Электронное издательство ЮРАЙТ». – Режим доступа: <https://biblio-online.ru/>;
- ЭБС «Университетская библиотека онлайн» [Электронный ресурс]: электронно-периодическое издание; программный комплекс для организации онлайн-доступа к лицензионным материалам / ООО «НексМедиа». – Режим доступа: <https://biblioclub.ru/>

7.3 СОВРЕМЕННЫЕ ПРОФЕССИОНАЛЬНЫЕ БАЗЫ ДАННЫХ:

- Информационно-аналитическая система SCIENCE INDEX
- Электронная база данных Scopus
- Базы данных компании CLARIVATE ANALYTICS

7.4. ИНФОРМАЦИОННЫЕ СПРАВОЧНЫЕ СИСТЕМЫ:

- Справочно-правовая информационная система Консультант Плюс <http://www.consultant.ru/>
- ООО «Современные медиа технологии в образовании и культуре». <http://www.informio.ru/>

8. ИНЫЕ СВЕДЕНИЯ И МАТЕРИАЛЫ НА УСМОТРЕНИЕ ВЕДУЩЕЙ КАФЕДРЫ.

1. <http://www.machinelearning.ru/> - Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных
2. https://courses.openedu.ru/courses/course-v1:hse+RLING+fall_2020/course/

9. ОБЕСПЕЧЕНИЕ ОБРАЗОВАНИЯ ДЛЯ ЛИЦ С ОВЗ.

Для обеспечения образования инвалидов и лиц с ограниченными возможностями здоровья реализация дисциплины может осуществляться в адаптированном виде, с учетом специфики освоения и дидактических требований, исходя из индивидуальных возможностей и по личному заявлению обучающегося.