

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**
федеральное государственное автономное образовательное учреждение
высшего образования «Балтийский федеральный университет имени Иммануила
Канта»
Высшая школа компьютерных наук и прикладной математики

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Прикладной статистический анализ данных»

Шифр: 01.03.02

Направление подготовки: «Прикладная математика и информатика»

Профиль: «Искусственный интеллект и анализ данных»

Квалификация (степень) выпускника: бакалавр

Калининград
2023

Лист согласования

Составители:

1. Верещагин Сергей Дмитриевич, к. ф.-м.н., доцент
2. Верещагин Михаил Дмитриевич, к. ф.-м.н., доцент
3. Мищук Богдан Ростиславович, к. ф.-м.н., доцент

Рабочая программа утверждена на заседании
Ученого совета ОНК «Институт высоких технологий»

Протокол № 4 от «24» января 2023 г.

Председатель Ученого совета ОНК
«Институт высоких технологий»

Профессор, д.ф.-м.н.

А.В. Юров

Руководитель ОПОП ВО

Е.П. Ставицкая

Содержание

1. Наименование дисциплины «Прикладной статистический анализ данных».
2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы.
3. Место дисциплины в структуре образовательной программы.
4. Виды учебной работы по дисциплине.
5. Содержание дисциплины, в том числе практической подготовки в рамках дисциплины, структурированное по темам.
6. Рекомендуемая тематика учебных занятий в форме контактной работы.
7. Методические рекомендации по видам занятий
8. Фонд оценочных средств
 - 8.1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы в рамках учебной дисциплины
 - 8.2. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений и навыков и (или) опыта деятельности в процессе текущего контроля
 - 8.3. Перечень вопросов и заданий для промежуточной аттестации по дисциплине
 - 8.4. Планируемые уровни сформированности компетенций обучающихся и критерии оценивания
9. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины
11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине.
12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

1. Наименование дисциплины: «Прикладной статистический анализ данных».

Целью курса «Прикладной статистический анализ данных» - изучение основных задач статистического анализа. А именно, будет дано описание математических моделей и методов таких разделов математической статистики как корреляционный анализ, регрессионный анализ, дисперсионный анализ, дискриминантный анализ, кластерный анализ. Предложенные методы и алгоритмы иллюстрируются с помощью более-менее реальных примеров

2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

Код компетенции	Результаты освоения образовательной программы (ИДК)	Результаты обучения по дисциплине
ПК-1. Способен классифицировать и идентифицировать задачи искусственного интеллекта, выбирать адекватные методы и инструментальные средства решения задач искусственного интеллекта	ПК-1.1. Классифицирует и идентифицирует задачи систем искусственного интеллекта в зависимости от особенностей проблемной и предметной областей ПК-1.2. Выбирает методы и инструментальные средства искусственного интеллекта для решения задач в зависимости от особенностей проблемной и предметной областей	Знать: основные методы построения математических моделей, потенциальные возможности и особенности современных статистических процедур анализа и обработки данных Уметь решать базовые задачи анализа и оценивания многомерных стохастических систем и разрабатывать практические алгоритмы для решения этих задач Владеть навыками оптимального выбора методов и параметров при решении задач анализа и оценивания многомерных стохастических систем

3. Место дисциплины в структуре образовательной программы

«Прикладной статистический анализ данных» представляет собой дисциплину Части, формируемой участниками образовательных отношений (Б1.В.08.02) направления подготовки бакалавриата 01.03.02 «Прикладная математика и информатика», профиль «Искусственный интеллект и анализ данных».

4. Виды учебной работы по дисциплине.

Виды учебной работы по дисциплине зафиксированы учебным планом основной профессиональной образовательной программы по указанному направлению и профилю, выражаются в академических часах. Часы контактной работы и самостоятельной работы обучающегося и часы, отводимые на процедуры контроля, могут различаться в учебных планах ОПОП по формам обучения. Объем контактной работы включает часы контактной

аудиторной работы (лекции/практические занятия/ лабораторные работы), контактной внеаудиторной работы (контроль самостоятельной работы), часы контактной работы в период аттестации. Контактная работа, в том числе может проводиться посредством электронной информационно-образовательной среды университета с использованием ресурсов сети Интернет и дистанционных технологий

5. Содержание дисциплины, структурированное по темам (разделам)

Исходя из рамок, установленных учебным планом по трудоемкости и видам учебной работы по дисциплине, преподаватель самостоятельно выбирает тематику занятий по формам и количеству часов проведения контактной работы: лекции и иные учебные занятия, предусматривающие преимущественную передачу учебной информации преподавателем и (или) занятия семинарского типа (семинары, практические занятия, практикумы, лабораторные работы, коллоквиумы и иные аналогичные занятия), и (или) групповые консультации, и (или) индивидуальную работу обучающихся с преподавателем, в том числе индивидуальные консультации (по курсовым работам/проектам – при наличии курсовой работы/проекта по данной дисциплине в учебном плане). Рекомендуемая тематика занятий максимально полно реализуется в контактной работе с обучающимися очной формы обучения. В случае реализации образовательной программы в заочной / очно-заочной форме трудоемкость дисциплины сохраняется, однако объем учебного материала в значительной части осваивается обучающимися в форме самостоятельной работы. При этом требования к ожидаемым образовательным результатам обучающихся по данной дисциплине не зависят от формы реализации образовательной программы.

№ п/п	Наименование разделов (тем) дисциплины	Содержание разделов (тем) дисциплин
1.	Основные задачи и методы прикладного одномерного и многомерного статистического анализа	Обзор основных задач статистического анализа. Задача о наилучшей линейной оценке. Коэффициенты корреляции и их свойства. Типы данных. Одномерное и многомерное нормальное распределение и его свойства.
2.	Регрессионный анализ	Классическая модель регрессии. Метод наименьших квадратов (МНК). Свойства оценок параметров. Проверка линейных гипотез. Коэффициент детерминации и его свойства. Выбор оптимальной модели. Модели с гетероскедастичностью и автокорреляцией ошибок и их оценивание.
3.	Дисперсионный анализ.	Однофакторный дисперсионный анализ. Двухфакторный дисперсионный анализ.
4.	Дискриминантный анализ.	Постановка задачи, описание общего метода решения задачи дискриминации. Решение задачи дискриминации для случая нормального распределения с известными параметрами. Решение задачи дискриминации для случая нормального распределения с неизвестными параметрами.

5.	Кластерный анализ.	Постановка задачи, описание общего метода решения задачи построения кластеров. Описание методов близости и различия кластеров. Описание алгоритма построения кластеров.
----	--------------------	---

6. Рекомендуемая тематика учебных занятий в форме контактной работы

Рекомендуемая тематика учебных занятий лекционного типа (предусматривающих преимущественную передачу учебной информации преподавателями):

№ п/п	Наименование разделов (тем) дисциплины	Содержание разделов (тем) дисциплин
1	Основные задачи и методы прикладного одномерного и многомерного статистического анализа	Лекция 1. Обзор основных задач статистического анализа. Лекция 2. Типы данных. Одномерное и многомерное нормальное распределение и его свойства.
2	Регрессионный анализ	Лекция 3. Классическая модель регрессии. Лекция 4. Метод наименьших квадратов (МНК). Лекция 5. Свойства оценок параметров. Проверка линейных гипотез. Лекция 6 Коэффициент детерминации и его свойства. Лекция 7. Выбор оптимальной модели. Лекция 8. Модели с гетероскедастичностью и автокорреляцией ошибок и их оценивание.
3	Дисперсионный анализ.	Лекция 9. Однофакторный дисперсионный анализ. Лекция 10-11. Двухфакторный дисперсионный анализ.
4	Дискриминантный анализ.	Лекция 12. Постановка задачи, описание общего метода решения задачи дискриминации. Лекция 13. Решение задачи дискриминации для случая нормального распределения с известными параметрами. Лекция 14. Решение задачи дискриминации для случая нормального распределения с неизвестными параметрами.
5	Кластерный анализ.	Лекция 15-16. Постановка задачи, описание общего метода решения задачи построения кластеров. Лекция 17. Описание методов близости и различия кластеров. Лекция 18. Описание алгоритма построения кластеров.

Рекомендуемая тематика *практических* занятий:

№ п/п	Наименование разделов (тем) дисциплины	Содержание разделов (тем) дисциплин
1	Основные задачи и методы прикладного одномерного и многомерного статистического анализа	Занятия 1-2. Типы данных. Одномерное и многомерное нормальное распределение и его свойства.
2	Регрессионный анализ	Занятие 3. Классическая модель регрессии. Занятие 4. Метод наименьших квадратов (МНК). Занятие 5. Проверка линейных гипотез. Занятие 6 Коэффициент детерминации и его свойства. Занятие 7. Выбор оптимальной модели. Занятие 8. Модели с гетероскедастичностью и автокорреляцией ошибок и их оценивание.
3	Дисперсионный анализ.	Занятие 9. Однофакторный дисперсионный анализ. Занятие 10-11. Двухфакторный дисперсионный анализ.
4	Дискриминантный анализ.	Занятие 12. Постановка задачи, описание общего метода решения задачи дискриминации. Занятие 13. Решение задачи дискриминации для случая нормального распределения с известными параметрами. Занятие 14. Решение задачи дискриминации для случая нормального распределения с неизвестными параметрами.
5	Кластерный анализ.	Занятие я 15-16. Постановка задачи, описание общего метода решения задачи построения кластеров. Занятие 17. Описание методов близости и различия кластеров. Занятие 18. Описание алгоритма построения кластеров.

Требования к самостоятельной работе обучающихся

1. Работа с лекционным материалом, предусматривающая проработку конспекта лекций и учебной литературы, по всем темам из п. 6 настоящей рабочей программы.
2. Выполнение домашнего задания, предусматривающего решение задач, выполнение упражнений, выдаваемых на практических занятиях, по всем темам из п. 6 настоящей рабочей программы.

Руководствуясь положениями статьи 47 и статьи 48 Федерального закона от 29 декабря 2012 г. N 273-ФЗ «Об образовании в Российской Федерации» научно-

педагогические работники и иные лица, привлекаемые университетом к реализации данной образовательной программы, пользуются предоставленными академическими правами и свободами в части свободы преподавания, свободы от вмешательства в профессиональную деятельность; свободы выбора и использования педагогически обоснованных форм, средств, методов обучения и воспитания; права на творческую инициативу, разработку и применение авторских программ и методов обучения и воспитания в пределах реализуемой образовательной программы и отдельной дисциплины.

Исходя из рамок, установленных учебным планом по трудоемкости и видам учебной работы по дисциплине, преподаватель самостоятельно выбирает тематику занятий по формам и количеству часов проведения контактной работы: лекции и иные учебные занятия, предусматривающие преимущественную передачу учебной информации преподавателем и (или) занятия семинарского типа (семинары, практические занятия, практикумы, лабораторные работы, коллоквиумы и иные аналогичные занятия), и (или) групповые консультации, и (или) индивидуальную работу обучающихся с преподавателем, в том числе индивидуальные консультации (по курсовым работам/проектам – при наличии курсовой работы/проекта по данной дисциплине в учебном плане).

Рекомендуемая тематика занятий максимально полно реализуется в контактной работе с обучающимися очной формы обучения. В случае реализации образовательной программы в заочной / очно-заочной форме трудоемкость дисциплины сохраняется, однако объем учебного материала в значительной части осваивается обучающимися в форме самостоятельной работы. При этом требования к ожидаемым образовательным результатам обучающихся по данной дисциплине не зависят от формы реализации образовательной программы.

7. Методические рекомендации по видам занятий

Лекционные занятия.

В ходе лекционных занятий обучающимся рекомендуется выполнять следующие действия. Вести конспектирование учебного материала. Обращать внимание на категории, формулировки, раскрывающие содержание тех или иных явлений и процессов, научные выводы и практические рекомендации по их применению. Задавать преподавателю уточняющие вопросы с целью уяснения теоретических положений, разрешения спорных ситуаций.

Желательно оставить в рабочих конспектах поля, на которых во внеаудиторное время можно сделать пометки из рекомендованной литературы, дополняющие материал прослушанной лекции, а также подчеркивающие особую важность тех или иных теоретических положений.

Практические и семинарские занятия.

На практических и семинарских занятиях в зависимости от темы занятия выполняется поиск информации по решению проблем, практические упражнения, контрольные работы, выработка индивидуальных или групповых решений, итоговое обсуждение с обменом знаниями, участие в круглых столах, разбор конкретных ситуаций, командная работа, представление портфолио и т.п.

Самостоятельная работа.

Самостоятельная работа осуществляется в виде изучения литературы, эмпирических данных по публикациям и конкретным ситуациям из практики, подготовке индивидуальных работ, работа с лекционным материалом, самостоятельное изучение отдельных тем дисциплины; поиск и обзор литературы и электронных источников; чтение и изучение учебника и учебных пособий.

8. Фонд оценочных средств

8.1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы в рамках учебной дисциплины

Основными этапами формирования указанных компетенций при изучении обучающимися дисциплины являются последовательное изучение содержательно связанных между собой тем учебных занятий. Изучение каждой темы предполагает овладение обучающимися необходимыми компетенциями. Результат аттестации обучающихся на различных этапах формирования компетенций показывает уровень освоения компетенций.

Контролируемые разделы (темы) дисциплины	Индекс контролируемой компетенции (или её части)	Оценочные средства по этапам формирования компетенций
		текущий контроль по дисциплине
Основные задачи и методы прикладного одномерного и многомерного статистического анализа	ПК-1	Решение задач
Регрессионный анализ	ПК-1	Решение задач
Дисперсионный анализ.	ПК-1	Решение задач
Дискриминантный анализ.	ПК-1	Решение задач
Кластерный анализ.	ПК-1	Решение задач

8.2. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений и навыков и (или) опыта деятельности процессе текущего контроля

Текущий контроль успеваемости осуществляется путем оценки результатов выполнения заданий практических (семинарских) занятий, самостоятельной работы, предусмотренных учебным планом и посещения занятий/активность на занятиях.

В качестве оценочных средств текущего контроля успеваемости предусмотрены:

выполнение заданий на практических (семинарски) занятиях

Примеры заданий:

1. Промоделировать $N=300$ дискретных случайных величин, принимающих значения 0, 1, 2 с вероятностями $(1/3, 1/3, 1/3)$. Проверить гипотезу согласия полученных модельных данных с этим законом распределения: 1) по критерию отношения правдоподобия, 2) по критерию согласия хи-квадрат.

2. Построить дерево кластеризации по набору категориальных переменных,

1) используя индекс Джини, 2) информационный выигрыш от объединения групп.

8.3. Перечень вопросов и заданий для промежуточной аттестации по дисциплине

1. Основные задачи многомерного статистического анализа.
2. Гильбертово пространство случайных величин. Задача о наилучшей линейной оценке.
3. Корреляционный и регрессионный анализ.
4. Коэффициенты корреляции.
5. Простая линейная регрессия. Метод наименьших квадратов. Свойства оценок.
6. Множественная линейная регрессия. МНК. Свойства оценок.
7. Т-критерий значимости влияния фактора.
8. Проверка линейных гипотез. F-критерий.
9. Проверка адекватности модели. Коэффициент детерминации.
10. Равенство уравнений регрессии. Тест Чоу.
11. Фиктивные переменные.
12. Модель линейной регрессии с гетероскедастичностью.
13. Модель линейной регрессии с автокорреляцией в ошибках. Критерий Дарбина-Уотсона.
14. Однофакторный дисперсионный анализ.
15. Двухфакторный дисперсионный анализ.
16. Дискриминантный анализ: постановка задачи и ее решение в случае известных параметров.
17. Решение задачи дискриминантного анализа в случае неизвестных параметров.
18. Кластерный анализ: постановка задачи и основные понятия.
19. Кластерный анализ: схема последовательного построения факторов.

8.4. Планируемые уровни сформированности компетенций обучающихся и критерии оценивания

Уровни	Содержательное описание уровня	Основные признаки выделения уровня (этапы формирования компетенции, критерии оценки сформированности)	Пятибалльная шкала (академическая) оценка	Двухбалльная шкала, зачет	БРС, % освоения (рейтинговая оценка)
Повышенный	Творческая деятельность	<i>Включает нижестоящий уровень.</i> Умение самостоятельно принимать решение, решать проблему/задачу теоретического и прикладного характера на основе изученных методов, приемов, технологий	отлично	зачтено	86-100
Базовый	Применение знаний и	<i>Включает нижестоящий уровень.</i>	хорошо		71-85

	умений в более широких контекстах учебной и профессиональной деятельности, нежели по образцу с большей степени самостоятельности и инициативы	Способность собирать, систематизировать, анализировать и грамотно использовать информацию из самостоятельно найденных теоретических источников и иллюстрировать ими теоретические положения или обосновывать практику применения			
Удовлетворительный (достаточный)	Репродуктивная деятельность	Изложение в пределах задач курса теоретически и практически контролируемого материала	удовлетворительно		55-70
Недостаточный	Отсутствие признаков удовлетворительного уровня		неудовлетворительно	не зачтено	Менее 55

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

Основная литература

1. Многомерный статистический анализ в экономических задачах: компьютерное моделирование в SPSS : учебное пособие / под ред. И.В. Орловой. — Москва : Вузовский учебник : ИНФРА-М, 2022. — 310 с. + Доп. материалы [Электронный ресурс]. — (Высшее образование: Бакалавриат). - ISBN 978-5-9558-0108-7. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1850713> (дата обращения: 04.04.2023). – Режим доступа: по подписке.

Дополнительная литература

1. Зарова, Е. В. Прикладной многомерный статистический анализ: Презентации для лекций и примеры решений с использованием пакета R: Учебное пособие на английском языке / Зарова Е.В. - Москва : НИЦ ИНФРА-М, 2016. - 370 с. ISBN 978-5-16-012133-8. - Текст : электронный. - URL: <https://znanium.com/catalog/product/557578> (дата обращения: 04.04.2023). – Режим доступа: по подписке.
2. Козлов, А. Ю. Статистический анализ данных в MS Excel : учебное пособие / А.Ю. Козлов, В.С. Мхитарян, В.Ф. Шишов. — Москва : ИНФРА-М, 2023. — 320 с. — (Высшее образование: Бакалавриат). — DOI 10.12737/2842. - ISBN 978-5-16-004579-5. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1907518> (дата обращения: 04.04.2023). – Режим доступа: по подписке.
3. Ниворожкина, Л. И. Многомерные статистические методы в экономике : учебник / Л.И. Ниворожкина, С.В. Арженковский. — Москва : РИОР : ИНФРА-М, 2023. — 203 с. — (Высшее образование). — DOI: <https://doi.org/10.12737/21773>. - ISBN 978-5-369-01621-3. - Текст : электронный. - URL:

<https://znanium.com/catalog/product/1048326> (дата обращения: 04.04.2023). – Режим доступа: по подписке.

10. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля).

- НЭБ Национальная электронная библиотека, диссертации и прочие издания
- ЭБС Консультант студента
- ПРОСПЕКТ ЭБС
- ЭБС ZNANIUM.COM
- ЭБС IBOOKS.RU
- Электронно-библиотечная система (ЭБС) Кантитана (<https://elib.kantiana.ru/>)

11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине.

Программное обеспечение обучения включает в себя:

- система электронного образовательного контента БФУ им. И. Канта – www.lms3.kantiana.ru, обеспечивающую разработку и комплексное использование электронных образовательных ресурсов;
- серверное программное обеспечение, необходимое для функционирования сервера и связи с системой электронного обучения через Интернет;
- корпоративная платформа webinar.ru;
- установленное на рабочих местах обучающихся ПО: Microsoft Windows 7, Microsoft Office Standart 2010, антивирусное программное обеспечение Kaspersky Endpoint Security.
- СУБД PostgreSQL (Свободное ПО, лицензия - Freeware).
- MongoDB (Свободное ПО, лицензия - Freeware).
- Python 2.7.15 (Anaconda2 5.2.0 64-bit)
- Python 3.6.5 (Anaconda3 5.2.0 64-bit)

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

Для проведения занятий лекционного типа, практических и семинарских занятий используются специальные помещения (учебные аудитории), оборудованные техническими средствами обучения – мультимедийной проекционной техникой. Для проведения занятий лекционного типа используются наборы демонстрационного оборудования.

Для проведения лабораторных работ, (практических занятий – при необходимости) используются специальные помещения (учебные аудитории), оснащенные специализированным лабораторным оборудованием: персональными компьютерами с возможностью выхода в интернет и с установленным программным обеспечением, заявленным в п.11.

Для проведения групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации используются специальные помещения (учебные аудитории), оборудованные специализированной мебелью (для обучающихся), меловой / маркерной доской.

Для организации самостоятельной работы обучающимся предоставляются помещения, оснащенные компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду университета.

Для обучения инвалидов и лиц с ограниченными возможностями здоровья университетом могут быть представлены специализированные средства обучения, в том числе технические средства коллективного и индивидуального пользования.