

АННОТАЦИЯ рабочей программы дисциплины
К.М.02.02 «Data Science: основы обработки и анализа данных»

Направление подготовки 02.03.03 Математическое обеспечение и администрирование информационных систем

Объем трудоемкости: 4 з.е.

Цели изучения дисциплины определены государственным образовательным стандартом высшего образования и, соотнесены с общими целями ООП ВО по направлению подготовки «Математическое обеспечение и администрирование информационных систем» в рамках которой преподается дисциплина.

Целью освоения учебной дисциплины «Data Science: основы обработки и анализа данных» является изучение основных методов многомерного статистического анализа данных с точки зрения их практического применения; привить навыки работы с соответствующими разделами ППП STATISTICA.

Задачи дисциплины в соответствии с поставленной целью состоят в следующем:

- помочь студентам понять и, освоить методологию многомерного статистического анализа данных;
- привить теоретические и практические знания в области прикладного многомерного анализа данных;
- познакомить студентов и обучить максимально широкому инструментарию многомерного анализа данных в среде ППП STATISTICA;
- выработать в процессе обучения у студентов навыки грамотного использования аппарата вероятностно-статистического моделирования посредством применения передовых информационных технологий.

Место дисциплины в структуре ООП ВО

Дисциплина «Data Science: основы обработки и анализа данных» относится к вариативной части Блока 1 «Дисциплины (модули)» учебного плана. Для успешного освоения данной дисциплины необходимы знания следующих дисциплин: Б1.О.19 Теория вероятностей и математическая статистика. Кроме того, изучение данной дисциплины необходимо для последующего изучения дисциплин: Б1.В.07 Технологии и инструментарий анализа больших данных и Б1.В.06 Машинное обучение.

Требования к уровню освоения дисциплины

Изучение данной учебной дисциплины направлено на формирование у обучающихся следующих компетенций: УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач, УК-2 Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений, ПК-1 Способен демонстрировать базовые знания математических и(или) естественных наук, программирования и информационных технологий, ПК-2 Способен определять структуру программного обеспечения, методы и средства его проектирования на основе требований с учетом существующих ограничений, ПК-5 Способен использовать основные концептуальные положения функционального, логического, объектно-ориентированного и визуального направлений программирования, методы, способы и средства разработки программ в рамках этих направлений.

Основные разделы дисциплины:

№	Наименование разделов (тем)	Количество часов			
		Всего	Аудиторная работа		Внеаудиторная работа
			Л	ЛР	
1	2	3	4	5	6
1.	Работа с данными. Основные операции над случаями и переменными. Основные статистики, корреляционный анализ	4	2	2	
2.	Сравнение средних величин методами параметрической и непараметрической	6	2	2	2

№	Наименование разделов (тем)	Количество часов			
		Всего	Аудиторная работа		Внеаудиторная работа
			Л	ЛР	
1	2	3	4	5	7
	статистики				
3.	Таблицы частот, сопряженности, флагов и заголовков	6	2	2	2
4.	Линейное и нелинейное моделирование взаимосвязей	8	2	2	4
5.	Дискриминантный анализ	6	2	2	2
6.	Кластерный анализ	8	2	2	4
7.	Факторный анализ	6	2	2	2
8.	Деревья классификации и регрессии	6	2	2	2
9.	CHAID модели	8	2	2	4
10.	Интерактивные деревья	8	2	2	4
11.	Стохастический градиентный бустинг	6	2	2	2
12.	Случайные леса регрессии и классификации	6	2	2	2
13.	Опорные вектора	6	2	2	2
14.	k-ближайших соседей	8	2	2	4
15.	Автоматизированные нейронные сети	8	2	2	4
16.	Автоматизированные нейронные сети, кластеризация	4	2	2	
ИТОГО по разделам дисциплины		104	32	32	40
Контроль самостоятельной работы (КСР)		4			
Промежуточная аттестация (ИКР)		0,5			
Подготовка к текущему контролю		35,5			
Общая трудоемкость по дисциплине		144			

Форма проведения аттестации по дисциплине: зачет+экзамен.

Основная литература:

1. Боровиков, В. П. Популярное введение в современный анализ данных в системе STATISTICA. Методология и технология современного анализа данных : учебное пособие / В. П. Боровиков. — Москва : Горячая линия-Телеком, 2018. — 288 с. — ISBN 978-5-9912-0326-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/111023> (дата обращения: 13.06.2023). — Режим доступа: для авториз. пользователей.

2. Чураков, Е. П. Введение в многомерные статистические методы / Е. П. Чураков. — 3-е изд., стер. — Санкт-Петербург : Лань, 2023. — 148 с. — ISBN 978-5-507-47141-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/330530> (дата обращения: 13.06.2023). — Режим доступа: для авториз. пользователей.

3. Ганичева, А. В. Прикладная статистика / А. В. Ганичева. — 4-е изд., стер. — Санкт-Петербург : Лань, 2023. — 172 с. — ISBN 978-5-507-47980-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/336800> (дата обращения: 13.06.2023). — Режим доступа: для авториз. пользователей.

Авторы: Халафян А.А. – д.т.н., профессор КАДИИ; Казаковцева Е.В. – ст.преподаватель КАДИИ