

**Федеральное государственное автономное образовательное
учреждение высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»**

УТВЕРЖДЕНО

**Директор по цифровизации
образования**

Д.И. Гриц

	Рабочая программа дисциплины (модуля)
по дисциплине:	Современные методы анализа данных
по направлению:	Бизнес-информатика
профиль подготовки:	Финансовые технологии и аналитика центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск" центр дополнительного, дополнительного профессионального и онлайн-образования "Пуск"
курс:	1
квалификация:	магистр

Семестры, формы промежуточной аттестации:

- 2 (весенний) - Дифференцированный зачет
- 3 (осенний) - Экзамен

Аудиторных часов: 72 всего, в том числе:

- лекции: 30 час.
- семинары: 42 час.
- лабораторные занятия: 0 час.

Самостоятельная работа: 123 час.

Подготовка к экзамену: 30 час.

Всего часов: 225, всего зач. ед.: 5

Программу составили:

Е.А. Савицкая
О.А. Культепина

Программа обсуждена на заседании центра дополнительного, дополнительного профессионального и онлайн-образования "Пуск" 13.06.2022

Аннотация

В рамках дисциплины «Современные методы анализа данных» обучающиеся познакомятся с современными методами анализа данных, получат представление об автоматизированных моделях анализа данных, научатся применять полученные умения и навыки для решения прикладных задач.

1. Цели и задачи

Цель дисциплины

- овладеть умениями и практическими навыками построения автоматизированных аналитических моделей.

Задачи дисциплины

- изучение базовых понятий и методов анализа данных;
- освоение технологий анализа данных;
- развитие навыков применения полученных знаний и умений для решения практических задач.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
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	ОПК-3.1 Владеет методами стратегического планирования и прогнозирования в профессиональной деятельности
	ОПК-3.2 Самостоятельно выбирает и обосновывает выбор современных методов и программного инструментария сбора, обработки и анализа данных
	ОПК-3.3 Владеет аналитическими и вычислительными методами решения, задач, понимает и учитывает на практике границы применимости получаемых решений
ПК-16 Способен готовить аналитические материалы для оценки мероприятий и выработки стратегических решений в области ИКТ	ПК-16.1 Эффективно применяет в ходе профессиональной деятельности методы и инструментарий анализа данных
	ПК-16.2 Знает методы подготовки аналитических материалов для оценки мероприятий и выработки стратегических решений в области ИКТ

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны

знать:

- основные понятия анализа данных;
- статистические методы анализа данных;
- интеллектуальные методы анализа данных;
- основные этапы внедрения систем анализа данных;
- процесс подготовки аналитических материалов.

уметь:

- формировать аналитические модели;
- анализировать качество полученных моделей;
- выбирать аналитические платформы и пакеты.

владеть:

- навыками внедрения систем анализа данных;
- навыками применения инструментов анализа данных.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1. Разделы дисциплины (модуля) и трудоемкости по видам учебных занятий

№	Тема (раздел) дисциплины	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.			
		Лекции	Семинары	Лаборат. работы	Самост. работа
1	Роль методов анализа данных в научно-исследовательской и практической деятельности	2	2		17
2	Типы данных	2	4		16
3	Основные положения математической статистики	4	2		17
4	Направления и области применения методов анализа данных	4	4		16
5	Одномерный статистический анализ данных	2	6		10
6	Многомерный анализ данных	4	6		9
7	Технологии анализа данных	2	6		9
8	Методы интеллектуального анализа данных	4	4		9
9	Программные средства анализа данных	4	4		10
10	Практические аспекты использования систем анализа данных	2	4		10
Итого часов		30	42		123
Подготовка к экзамену		30 час.			
Общая трудоёмкость		225 час., 5 зач.ед.			

4.2. Содержание дисциплины (модуля), структурированное по темам (разделам)

Семестр: 2 (Весенний)

1. Роль методов анализа данных в научно-исследовательской и практической деятельности

Классические и современные методы анализа данных. Основные особенности методов классической математической статистики. Типы статистических задач.

2. Типы данных

Матрицы данных. Типы объектов или наблюдений. Пространственная выборка и временной ряд. Классификации показателей. Шкалы измерения.

3. Основные положения математической статистики

Понятия детерминированной и случайной величин. Закон распределения случайной величины. Виды представления закона распределения дискретной и непрерывной случайных величин. Задачи статистической оценки параметров и проверки статистических гипотез. Виды статистических оценок. Свойства точечных статистических оценок.

4. Направления и области применения методов анализа данных

Агрегирование и коррелирование. Методы одномерного анализа данных. Методы двумерного анализа данных. Методы многомерного анализа данных. Методы анализа временных рядов.

5. Одномерный статистический анализ данных

Основные типы задач, решаемые в процессе статистического анализа данных. Графический анализ. Интегральные характеристики случайных величин. Начальные и центральные моменты. Интегральные характеристики центра и разброса. Вариационный ряд. Статистические гистограммы.

6. Многомерный анализ данных

Графический анализ двух показателей (точечные диаграммы). Корреляционный анализ. Ковариация и ее свойства. Коэффициент корреляции и его свойства. Матрица попарных коэффициентов корреляции. Регрессионный анализ количественных признаков. Парная регрессия. Метод наименьших квадратов. Анализ таблиц сопряженности (качественных признаков). Множественная регрессия. Дискриминантный анализ. Кластерный анализ. Факторный анализ.

7. Технологии анализа данных

Оперативный анализ данных (OLAP). Технологии анализа данных. Оперативный анализ данных – OLAP. Многомерный оперативный анализ данных (MOLAP). MOLAP (Multidimensional OLAP). Пространственная схема данных. Реляционный оперативный анализ данных. ROLAP. ROLAP (Relational OLAP). Работа с реляционным хранилищем. Дополнительные реляционные таблицы компьютерных сетей. Гибридный оперативный анализ данных (HOLAP). HOLAP (Hybrid OLAP). Многомерные таблицы для агрегатов. WOLAP, Web-based OLAP – OLAP ориентированный на Web.

8. Методы интеллектуального анализа данных

Интеллектуальный анализ данных. Основные понятия, особенности, тенденции, перспективы. Data Mining. Visual Mining. Text Mining.

9. Программные средства анализа данных

Пакеты прикладных программ STATISTICA, BMDP, SPSS, Statgraphics, Systat в решении задач анализа данных. Современные программные средства для оперативного анализа данных. SAP OLAP Server. SAP BW. IBM Cognos PowerPlay. Microstrategy Intelligence Server. Mondrian.

10. Практические аспекты использования систем анализа данных

Критерии выбора аналитических платформ и пакетов. Основные этапы внедрения систем анализа данных. Категории пользователей аналитических систем. Требования, предъявляемые к различным группам пользователей. Способы снижения рисков проектов.

5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Занятия по учебной дисциплине проводятся с использованием дистанционных образовательных технологий. Каждый обучающийся обеспечен доступом к образовательной платформе <https://netology.ru/>

6. Перечень рекомендуемой литературы

Основная литература

1. Python и анализ данных, Электрон. версия печ. публикации / У. Маккини. — Москва, ДМК Пресс, 2020

Дополнительная литература

1. Прикладной многомерный статистический анализ [Текст], сб. работ/Акад. наук СССР, Центр. экономико-матем. ин-т, -М., Наука, 1973

7. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

1. Анализ данных: учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва: Издательство Юрайт, 2022. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст: электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/489100>
2. Клячкин, В. Н. Статистические методы анализа данных: учебное пособие / В. Н. Клячкин, Ю. Е. Кувайскова, В. А. Алексеева. — Москва: Финансы и статистика, 2021. — 242 с. — ISBN 978-5-00184-057-2. — Текст: электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/>
3. Кузьмин, В. И. Методы анализа данных: учебное пособие / В. И. Кузьмин, А. Ф. Гадзаов. — 2-е изд., перераб. и доп. — Москва: РТУ МИРЭА, 2020. — 155 с. — Текст: электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/171433>
4. Брусенцев А.Г. Анализ данных и процессов. Ч.1. Методы статистического анализа данных: учебное пособие / Брусенцев А.Г.. — Белгород: Белгородский государственный технологический университет им. В.Г. Шухова, ЭБС АСВ, 2017. — 63 с. — ISBN 978-5-361-00540-6. — Текст: электронный // IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/92237.html>
5. Кулаичев, А. П. Методы и средства комплексного анализа данных / А. П. Кулаичев. - 4-е изд., перераб. и доп. - Москва: НИЦ ИНФРА-М, 2016. - 511 с. - ISBN 978-5-16-104593-0 - Текст: электронный. - URL: <https://znanium.com/catalog/product/548836>

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)

1. Образовательная платформа <https://netology.ru/>
2. Webinar.ru
3. Zoom
4. Google Drive

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Студент, изучающий дисциплину, должен с одной стороны, овладеть общим понятийным аппаратом, а с другой стороны, должен научиться применять теоретические знания на практике. В результате изучения дисциплины студент должен знать основные определения дисциплины, уметь применять полученные знания для решения различных задач.

Успешное освоение курса требует:

- посещения всех занятий, предусмотренных учебным планом по дисциплине;
- ведения конспекта занятий;
- напряжённой самостоятельной работы студента.

Самостоятельная работа включает в себя:

- чтение рекомендованной литературы;
- проработку учебного материала, подготовку ответов на вопросы, предназначенных для самостоятельного изучения;
- решение задач, предлагаемых студентам на занятиях;
- подготовку к выполнению заданий текущей и промежуточной аттестации.

Показателем владения материалом служит умение без конспекта отвечать на вопросы по темам дисциплины.

Важно добиться понимания изучаемого материала, а не механического его запоминания. При затруднении изучения отдельных тем, вопросов, следует обращаться за консультациями к преподавателю.

Возможен промежуточный контроль знаний студентов в виде решения задач в соответствии с тематикой занятий.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению:	Бизнес-информатика		
профиль подготовки:	Финансовые технологии и аналитика	▲	▲
	онлайн-образования "Пуск"	▲	▲
	онлайн-образования "Пуск"		
курс:	1		
квалификация:	магистр		

Семестры, формы промежуточной аттестации:

- 2 (весенний) - Дифференцированный зачет
- 3 (осенний) - Экзамен

Разработчики:

Е.А. Савицкая
О.А. Культепина

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
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	ОПК-3.1 Владеет методами стратегического планирования и прогнозирования в профессиональной деятельности
	ОПК-3.2 Самостоятельно выбирает и обосновывает выбор современных методов и программного инструментария сбора, обработки и анализа данных
	ОПК-3.3 Владеет аналитическими и вычислительными методами решения, задач, понимает и учитывает на практике границы применимости получаемых решений
ПК-16 Способен готовить аналитические материалы для оценки мероприятий и выработки стратегических решений в области ИКТ	ПК-16.1 Эффективно применяет в ходе профессиональной деятельности методы и инструментари анализа данных
	ПК-16.2 Знает методы подготовки аналитических материалов для оценки мероприятий и выработки стратегических решений в области ИКТ

2. Показатели оценивания компетенций

В результате изучения дисциплины «Современные методы анализа данных» обучающийся должен:

знать:

- основные понятия анализа данных;
- статистические методы анализа данных;
- интеллектуальные методы анализа данных;
- основные этапы внедрения систем анализа данных;
- процесс подготовки аналитических материалов.

уметь:

- формировать аналитические модели;
- анализировать качество полученных моделей;
- выбирать аналитические платформы и пакеты.

владеть:

- навыками внедрения систем анализа данных;
- навыками применения инструментов анализа данных.

3. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю

Во время текущего контроля студент должен уметь ответить на следующие вопросы:

1. Основные типы статистических задач.
2. Матрица данных
3. Пространственная выборка.
4. Временной ряд.
5. Основные шкалы измерения.
6. Что такое генеральная совокупность?
7. Что такое выборочная совокупность?
8. Основные виды статистических оценок.
9. Закон распределения случайной величины.
10. В каком виде представляется закон распределения дискретной случайной величины?
11. В каком виде представляется закон распределения непрерывной случайной величины?
12. Основные свойства точечных оценок.
13. Агрегатирование показателей
14. Задачи коррелирования

15. Методы одномерного анализа данных
16. Методы двумерного анализа данных.
17. Методы многомерного анализа данных.
18. Модели временных рядов.
19. Статистическая оценка параметров.
20. Проверка статистической гипотезы.
21. Интегральные характеристики разброса.
22. Применимость интегральных характеристик для различных шкал измерения.
23. Корреляция и ее свойства.
24. Что такое регрессия?
25. Суть метода наименьших квадратов.
26. Основные характеристики качества регрессионной модели.
27. Основная идея кластерного анализа.
28. Основные технологии анализа данных
29. Что такое оперативный анализ данных?
30. Основные функции OLAP-систем.
31. Какие преимущества дает использование OLAP систем?
32. Что такое Data Mining?
33. Основные задачи Data mining.
34. Определение Visual mining.
35. Что такое Text Mining?
36. Пакеты прикладных программ математической статистики
37. Основные возможности SAP Crystal Server
38. Что такое SAP BW?
39. Преимущества Pentaho-BI Suite.
40. Критерии выбора аналитических платформ.

Во время занятий могут проходить интерактивные обсуждения в чатах курса, что будет являться домашним заданием. Успешное выполнение всех заданий по курсу и выполнение контрольных срезов знаний дает преимущество на экзамене.

4. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся

Пример итогового задания для дифференцированного зачета:

- 1) В исходной таблице содержатся данные финансовых расчетов с потребителями компании за последние 4 месяца. Каждая строка в базе данных содержит информацию об одной операции отгрузки товара, а именно, имя потребителя, месяц, категорию отгрузки, сумму отгрузки, сумму оплаты.
 - Создать сводную таблицу для вычисления количества операций отгрузки по каждому потребителю и по каждой категории за все 4 месяца.
 - Создать сводную таблицу для вычисления общих сумм поставок по каждому потребителю за каждый месяц.
 - Построить соответствующие временные ряды для каждого потребителя.
 - Построить гистограмму для поступивших оплат только для двух категорий отгрузки «Оборудование» и «Материалы».
- 2) Основываясь на данных о продажах сформировать сценарии ежедневных отчетов по долевого объему продаж в каждом из отделов по группам товаров за последние 10 дней. В каждом из отчетов должен присутствовать полный перечень товарных групп, упорядоченных по возрастанию кода.

Пример итогового задания для экзамена:

Собрать недельные данные о ценах закрытия и объемах торгов по пяти акциям, обращающимся на Московской бирже. Для каждой акции:

- рассчитать недельные доходности и натуральные логарифмы объемов торгов;

- построить и описать гистограммы всех показателей;
- проверить гипотезы о нормальном и логарифмически нормальном законе распределения всех показателей;
- построить 95% доверительный интервал для доходности - проверить гипотезу о том, что доходность по нему равна среднерыночной при альтернативной гипотезе о том, что доходность превышает среднерыночную;
- построить модель прогнозирования цен закрытия, дав точечные и интервальные прогнозы.

Критерии оценивания

Оценка отлично (10 баллов) - выставляется обучающемуся, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины, проявляющему интерес к данной предметной области, продемонстрировавшему умение уверенно и творчески применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично (9 баллов) - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка отлично (8 баллов) - выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, правильное обоснование принятых решений, с некоторыми недочетами.

Оценка хорошо (7 баллов) - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но недостаточно грамотно обосновывает полученные результаты.

Оценка хорошо (6 баллов) - выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности.

Оценка хорошо (5 баллов) - выставляется студенту, если он в основном знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач достаточно большое количество неточностей.

Оценка удовлетворительно (4 балла) - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он освоил основные разделы учебной программы, необходимые для дальнейшего обучения, и может применять полученные знания по образцу в стандартной ситуации.

Оценка удовлетворительно (3 балла) - выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, допускающему ошибки в формулировках базовых понятий, нарушения логической последовательности в изложении программного материала, слабо владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и с трудом применяет полученные знания даже в стандартной ситуации.

Оценка неудовлетворительно (2 балла) - выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных принципов и не умеет использовать полученные знания при решении типовых задач.

Оценка неудовлетворительно (1 балл) - выставляется студенту, который не знает основного содержания учебной программы дисциплины, допускает грубейшие ошибки в формулировках базовых понятий дисциплины и вообще не имеет навыков решения типовых практических задач.

5. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Экзамен и дифференцированный зачет по дисциплине проводятся в форме выполнения итогового задания.