

МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА»
(РУТ (МИИТ))



Рабочая программа дисциплины (модуля),
как компонент образовательной программы
базового высшего образования
по направлению подготовки
01.03.02 Прикладная математика и информатика,
утвержденной первым проректором РУТ (МИИТ)
Тимониным В.С.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Анализ данных

Направление подготовки: 01.03.02 Прикладная математика и информатика

Направленность (профиль): Математическое моделирование и системный анализ

Форма обучения: Очная

Рабочая программа дисциплины (модуля) в виде электронного документа выгружена из единой корпоративной информационной системы управления университетом и соответствует оригиналу

Простая электронная подпись, выданная РУТ (МИИТ)
ID подписи: 1343395
Подписал: И.о. заведующего кафедрой Тищенко Сергей Александрович
Дата: 25.06.2026

1. Общие сведения о дисциплине (модуле).

Целью освоения дисциплины (модуля) является:

- ознакомление студентов с элементами математической теории анализа данных и временных рядов.

Задачами освоения дисциплины (модуля) являются:

- формирование личности студента, развитие его интеллекта и умения логически и алгоритмически мыслить, формирование умений и навыков, необходимых при практическом применении теории комплексного анализа;

- обучение использованию в своей деятельности современных статистико-математические методов и моделей.

2. Планируемые результаты обучения по дисциплине (модулю).

Перечень формируемых результатов освоения образовательной программы (компетенций) в результате обучения по дисциплине (модулю):

ПК-4 - Уметь ставить цели создания системы, разрабатывать концепцию системы и требования к ней, выполнять декомпозицию требований к системе.

Обучение по дисциплине (модулю) предполагает, что по его результатам обучающийся будет:

Знать:

- имеющиеся методы и средства решения прикладных задач;
- понятия и факты из области математических, а также других естественно-научных дисциплин.

Уметь:

- формулировать постановку задачи и излагать ее;
- применять на практике изученные методы и подходы.

Владеть:

- навыками классификации и выделения существенных подзадач при анализе данных;
- навыками сбора и обработки данных.

3. Объем дисциплины (модуля).

3.1. Общая трудоемкость дисциплины (модуля).

Общая трудоемкость дисциплины (модуля) составляет 4 з.е. (144 академических часа(ов)).

3.2. Объем дисциплины (модуля) в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

| Тип учебных занятий | Количество часов | |
|---|------------------|------------|
| | Всего | Семестр №8 |
| Контактная работа при проведении учебных занятий (всего): | 64 | 64 |
| В том числе: | | |
| Занятия лекционного типа | 32 | 32 |
| Занятия семинарского типа | 32 | 32 |

3.3. Объем дисциплины (модуля) в форме самостоятельной работы обучающихся, а также в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении промежуточной аттестации составляет 80 академических часа (ов).

3.4. При обучении по индивидуальному учебному плану, в том числе при ускоренном обучении, объем дисциплины (модуля) может быть реализован полностью в форме самостоятельной работы обучающихся, а также в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении промежуточной аттестации.

4. Содержание дисциплины (модуля).

4.1. Занятия лекционного типа.

| № п/п | Тематика лекционных занятий / краткое содержание |
|-------|--|
| 1 | Доверительные интервалы Рассматриваемые вопросы: - квантили и интерквартильный размах; - предсказательный интервал; - доверительный интервал; - минимально необходимые объемы выборок; - минимальный обнаружимый эффект. |
| 2 | Проверка гипотез Рассматриваемые вопросы: - достигаемый уровень значимости и критическая область; - ошибки I и II рода; - связь с доверительными интервалами; - параметрические критерии; |

| № п/п | Тематика лекционных занятий / краткое содержание |
|----------|---|
| | <ul style="list-style-type: none"> - непараметрические критерии; - множественная проверка гипотез: методы контроля FWER и FDR |
| 3 | <p>Корреляционный анализ Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - виды связей и типы переменных; - основные виды корреляций; - корреляция Пирсона; - корреляция Спирмена; - корреляция Кенделла; - корреляция Мэтьюса; - корреляция Крамера; - бисериальный коэффициент корреляции. |
| 4 | <p>Линейная модели Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - основные подходы линейного анализа: регрессия и классификация; - функционалы качества; - обработка выбросов и пропусков в данных; - регуляризация; - работа с несбалансированными классами; - практическое применение. |
| 5 | <p>Деревья решений Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - основные принципы работы деревьев решений; - критерии разбиения: энтропийный критерий и критерий Джини; - методы поиска гиперпараметров; - практическое применение. |
| 6 | <p>Бэггинг и бустинг Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - принцип работы алгоритмов; - основные алгоритмы: XGBoost, LightGBM, CatBoost, Random Forest; - методы поиска гиперпараметров; - сравнение с деревом решений; - практическое применение. |
| 7 | <p>Uplift моделирование Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - отличие uplift моделирования от традиционных методов; - основные алгоритмы; - понятие uplift эффекта и его интерпретация; - критерии оценки качества; - визуализация uplift кривой; - практическое применение. |
| 8 | <p>Временные ряды Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - общие сведения о временных рядах и задачах анализа; - анализ временных рядов; - прогнозирование временных рядов; - типы временных рядов; - практическое применение. |

4.2. Занятия семинарского типа.

Практические занятия

| № п/п | Тематика практических занятий/краткое содержание |
|-------|--|
| 1 | Доверительные интервалы В результате выполнения практического задания студент получает навыки построения доверительных интервалов и оценки точности и достоверности оценок. |
| 2 | Проверка гипотез В результате выполнения практического задания студент получает навыки формулирования гипотез, интерпретации результатов статистического теста и умение выбрать подходящий критерий для проверки гипотезы. |
| 3 | Корреляционный анализ: В результате выполнения практического задания студент получает навыки оценивания степени связи между переменными и применимости корреляционного анализа в различных ситуациях |
| 4 | Линейные модели В результате выполнения практического задания студент получает навыки понимания основных принципов линейных моделей и их применимости к различным типам данных. |
| 5 | Деревья решений В результате выполнения практического задания студент получает навыки построения и обучения деревьев решений для задач классификации и регрессии. |
| 6 | Бэггинг и бустинг В результате выполнения практического задания студент получает навыки построения и обучения случайного леса и градиентного бустинга для задач классификации и регрессии. |
| 7 | Uplift моделирование В результате выполнения практического задания студент получает навыки построения uplift моделей и их применимости в различных областях. |
| 8 | Временные ряды В результате выполнения практического задания студент получает навыки построения временных рядов и их применимости в различных областях. |

4.3. Самостоятельная работа обучающихся.

| № п/п | Вид самостоятельной работы |
|-------|---|
| 1 | Изучение учебной литературы из приведенных источников |
| 2 | Подготовка к практическим занятиям. |
| 3 | Самостоятельное изучение лекционного материала. |
| 4 | Выполнение курсового проекта. |
| 5 | Подготовка к промежуточной аттестации. |
| 6 | Подготовка к текущему контролю. |

4.4. Примерный перечень тем курсовых проектов

Сравнительный анализ алгоритмов заполнения пропусков в многомерных временных рядах.

Применение методов снижения размерности для визуализации и кластеризации высокоразмерных данных.

Разработка и исследование ансамблевых методов классификации на основе градиентного бустинга.

Прогнозирование финансовых временных рядов с использованием статистических моделей и рекуррентных нейронных сетей.

Обнаружение аномалий в сетевом трафике методами одноклассовой классификации и изолирующего леса.

Сравнительный анализ методов кластеризации на данных со сложной геометрической структурой.

Оценка кредитного скоринга с использованием моделей логистической регрессии и случайного леса.

Классификация текстовых документов методами машинного обучения с использованием векторных представлений слов.

Разработка рекомендательной системы на основе коллаборативной фильтрации и матричных разложений.

Анализ тональности отзывов пользователей с помощью рекуррентных нейронных сетей.

Применение методов отбора признаков для повышения интерпретируемости моделей машинного обучения.

Прогнозирование оттока клиентов телекоммуникационной компании методами классификации. Распознавание рукописных символов с использованием сверточных нейронных сетей.

Корреляционный и регрессионный анализ социально-экономических показателей региона.

Сегментация рынка клиентов интернет-магазина на основе анализа давности, частоты, стоимости покупок и кластеризации.

Восстановление плотности распределения случайных величин с помощью ядерных оценок.

Применение моделей выживаемости для прогнозирования надежности технических систем.

Тестирование статистических гипотез методами повторной выборки на реальных данных.

Анализ временных рядов энергопотребления предприятия для краткосрочного прогнозирования нагрузок.

Определение авторства текста на основе стилометрического анализа и классификаторов машинного обучения.

Распознавание спам-сообщений методами наивного байесовского классификатора и опорных векторов.

Моделирование пространственных данных методами геостатистического анализа.

Применение методов тематического моделирования для анализа коллекций текстовых документов.

Оценка эффективности сплит-тестирования методами математической статистики в условиях высокой дисперсии.

Прогнозирование спроса на товары в ритейле с учетом сезонности и акций.

5. Перечень изданий, которые рекомендуется использовать при освоении дисциплины (модуля).

| № п/п | Библиографическое описание | Место доступа |
|-------|--|---|
| 1 | «Иванов, Б. Н. Теория вероятностей и математическая статистика : учебное пособие для вузов / Б. Н. Иванов. — 3-е изд., стер. — Санкт-Петербург : Лань, 2024. — 224 с. — ISBN 978-5-507-49479-8.» (Иванов, Б. Н. Теория вероятностей и математическая статистика : учебное пособие для вузов / Б. Н. Иванов. — 3-е изд., стер. — Санкт-Петербург : Лань, 2024. — ISBN 978-5-507-49479-8 | https://e.lanbook.com/book/393053 (дата обращения: 25.06.2025) |
| 2 | Катаргин, Н. В. Анализ временных рядов : учебник для вузов / Н. В. Катаргин, Е. А. Качалина. — Санкт-Петербург : Лань, 2024. — 180 с. — ISBN 978-5-507-50162-5. — Текст : электронный // Лань : электронно-библиотечная система | https://e.lanbook.com/book/433307 (дата обращения: 25.06.2025) |

6. Перечень современных профессиональных баз данных и информационных справочных систем, которые могут использоваться при освоении дисциплины (модуля).

- Научно-техническая библиотека РУТ (МИИТ) (<http://library.miit.ru>);
- Информационный портал Научная электронная библиотека eLIBRARY.RU (www.elibrary.ru);
- Образовательная платформа «Юрайт» (<https://urait.ru/>);
- Электронно-библиотечная система издательства «Лань» (<http://e.lanbook.com/>);
- Интернет-университет информационных технологий (<http://www.intuit.ru/>).

7. Перечень лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства, необходимого для освоения дисциплины (модуля).

- Операционная система Windows;
- Microsoft Office;
- MS Teams;
- Поисковые системы.

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

Учебные аудитории для проведения учебных занятий, оснащенные компьютерной техникой и наборами демонстрационного оборудования.

9. Форма промежуточной аттестации:

Зачет в 8 семестре.

Курсовой проект в 8 семестре.

10. Оценочные материалы.

Оценочные материалы, применяемые при проведении промежуточной аттестации, разрабатываются в соответствии с локальным нормативным актом РУТ (МИИТ).

Авторы:

профессор, профессор, д.н. кафедры
«Математическое моделирование
сложных систем» Института
железнодорожного транспорта

А.М. Филимонов

Согласовано:

и.о. заведующего кафедрой ПМ
Председатель учебно-методической
комиссии

С.А. Тищенко

Н.А. Андриянова