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 ФГАОУ ВО
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 ФГАОУ ВО РУТ (МИИТ)
ID подписи: 1343395
Подписал: И.о. заведующего кафедрой Тищенко Сергей Александрович
Дата: 01.09.2026

1. Общие сведения о дисциплине (модуле).

Целями освоения дисциплины (модуля) является:

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

- приобретение навыков построения прогностических моделей (линейные модели, деревья решений, ансамблевые методы, Uplift-моделирование, временные ряды);

- обеспечение студентов прочными знаниями в области анализа данных, формирование основ математической подготовки, необходимых для профессиональной деятельности в области Data Science, бизнес-аналитики и принятия решений на основе данных.

Задачами освоения дисциплины (модуля) являются:

- формирование личности студента, развитие его интеллекта и умения логически и алгоритмически мыслить применительно к задачам анализа данных и машинного обучения;

- формирование умений и навыков, необходимых для практического применения методов статистического вывода (доверительные интервалы, проверка гипотез, корреляционный анализ);

формирование умений и навыков, необходимых для использования обучающимися в профессиональной деятельности современных алгоритмов регрессии, классификации, ансамблевых методов, Uplift-моделирования и анализа временных рядов.

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:

- методы построения доверительных интервалов, основы проверки статистических гипотез, виды корреляционной связи, методы её измерения и принципы работы линейных моделей, деревьев решений и ансамблевых методов;

- основные понятия анализа временных рядов, типы рядов, методы прогнозирования.

Уметь:

- строить и интерпретировать доверительные интервалы, формулировать нулевые и альтернативные гипотезы, выбирать критерии и рассчитывать уровень значимости;
- обучать линейные модели, деревья решений и ансамблевые методы, строить uplift-модели и проводить анализ временных рядов.

Владеть:

- проверкой параметрических и непараметрических гипотез, применения поправок на множественное тестирование;
- построением и валидацией линейных моделей, деревьев решений и ансамблевых алгоритмов и методами анализа, визуализации и прогнозирования временных рядов.

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; - непараметрические критерии.
3	Множественная проверка гипотез Рассматриваемые вопросы: - постановка задачи множественной проверки гипотез; - методы контроля ошибок первого рода FWER; - методы контроля ошибок первого рода FDR.
4	Корреляционный анализ Рассматриваемые вопросы: - виды связей и типы переменных: - основные виды корреляций: - корреляция Пирсона; - корреляция Спирмена; - корреляция Кенделла; - корреляция Мэтьюса; - корреляция Крамера; - бисериальный коэффициент корреляции.
5	AB тестирование (часть 1) Рассматриваемые вопросы: - постановка задачи AB тестирования; - минимаоьный объем выборки; - стратифицированное семплирование.

№ п/п	Тематика лекционных занятий / краткое содержание
6	АВ тестирование (часть 2) Рассматриваемые вопросы: - применение бутстрапирования в АВ тестировании; - построение синтетических АА и АВ тестов; - проверка корректности статистического критерия.
7	АВ тестирование (часть 3) Рассматриваемые вопросы: - метрики отношения; - дельта-метод для оценки значимости изменения метрик отношения; - методы понижения дисперсии данных.
8	Линейная регрессия Рассматриваемые вопросы: - постановка задачи; - функционалы качества; - обработка выбросов и пропусков в данных; - регуляризация; - практическое применение.
9	Линейная классификация Рассматриваемые вопросы: - постановка задачи; - функционалы качества; - работа с несбалансированными классами; - практическое применение.
10	Деревья решений Рассматриваемые вопросы: - основные принципы работы деревьев решений; - критерии разбиения: энтропийный критерий и критерий Джини; - методы поиска гиперпараметров; - практическое применение.
11	Ансамблиевые методы (часть 1) Рассматриваемые вопросы: - постановка задачи; - алгоритм бэггинга; - методы поиска гиперпараметров; - сравнение с деревом решений; - практическое применение.
12	Ансамблиевые методы (часть 2) Рассматриваемые вопросы: - постановка задачи; - алгоритм бустинга; - сравнение со случайным лесом; - практическое применение.
13	Uplift моделирование (часть 1) Рассматриваемые вопросы: - отличие uplift моделирования от традиционных методов; - понятие uplift эффекта и его интерпретация; - критерии оценки качества; - визуализация uplift кривой.
14	Uplift моделирование (часть 2) Рассматриваемые вопросы:

№ п/п	Тематика лекционных занятий / краткое содержание
	- основные алгоритмы прогнозирования; - практическое применение.
15	Временные ряды (часть 1) Рассматриваемые вопросы: - общие сведения о временных рядах; - основные задачи анализа временных рядов; - коэффициенты автокорреляции и частичной автокорреляции; - свойство стационарности.
16	Временные ряды (часть 2) Рассматриваемые вопросы: - дифференцирование временных рядов; - основные алгоритмы прогнозирования 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	Множественная проверка гипотез В результате выполнения практического задания студент получает навыки множественной проверки статистических гипотез.
4	Корреляционный анализ В результате выполнения практического задания студент получает навыки оценивания степени связи между переменными и применимости корреляционного анализа в различных ситуациях.
5	AB тестирование (часть 1) В результате освоения темы студент овладевает навыками постановки задачи A/B тестирования, расчёта минимально необходимого объёма выборки и применения стратифицированного семплирования для повышения точности эксперимента.
6	AB тестирование (часть 2) В результате освоения темы студент приобретает навыки применения бутстрапирования для оценки неопределённости, построения синтетических AA и AB тестов и проверки корректности используемого статистического критерия.
7	AB тестирование (часть 3) В результате освоения темы студент осваивает методы анализа метрик отношения, применение дельта-метода для оценки значимости их изменений и техники понижения дисперсии данных (например, CUPED).
8	Линейная регрессия В результате выполнения практического задания студент получает навыки понимания основных принципов модели линейной регрессии и ее применимости к различным типам данных.

№ п/п	Тематика практических занятий/краткое содержание
9	Линейная классификация В результате выполнения практического задания студент получает навыки понимания основных принципов модели линейной классификации и ее применимости к различным типам данных.
10	Деревья решений В результате выполнения практического задания студент получает навыки построения и обучения деревьев решений для задач классификации и регресии.
11	Ансамблиевые методы – бэггинг В результате выполнения практического задания студент получает навыки построения модели случайного леса для задач классификации и регрессии.
12	Ансамблиевые методы – бустинг В результате выполнения практического задания студент получает навыки построения модели градиентного бустинга для задач классификации и регрессии.
13	Uplift моделирование – построение модели В результате выполнения практического задания студент получает навыки построения uplift моделей и их применимости в различных областях.
14	Uplift моделирование – оценка качества В результате выполнения практического задания студент получает навыки расчета и интерпретации ключевых метрик качества uplift моделей для оценки эффективности елевого воздействия.
15	Временные ряды – построение модели В результате выполнения практического задания студент получает навыки построения временных рядов и их применимости в различных областях.
16	Временные ряды – оценка качества В результате выполнения практического задания студент приобретает навыки оценки точности прогнозов временных рядов, а также построения доверительных интервалов и валидации моделей на отложенных периодах.

4.3. Самостоятельная работа обучающихся.

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

4.4. Примерный перечень тем курсовых проектов

1. Сравнительный анализ алгоритмов заполнения пропусков в многомерных временных рядах.

2. Применение методов снижения размерности для визуализации и кластеризации высокоразмерных данных.

3. Разработка и исследование ансамблевых методов классификации на основе градиентного бустинга.

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

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

6. Сравнительный анализ методов кластеризации на данных со сложной геометрической структурой.

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

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

9. Разработка рекомендательной системы на основе коллаборативной фильтрации и матричных разложений.

10. Анализ тональности отзывов пользователей с помощью рекуррентных нейронных сетей.

11. Применение методов отбора признаков для повышения интерпретируемости моделей машинного обучения.

12. Прогнозирование оттока клиентов телекоммуникационной компании методами классификации.

13. Распознавание рукописных символов с использованием сверточных нейронных сетей.

14. Корреляционный и регрессионный анализ социально-экономических показателей региона.

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

16. Восстановление плотности распределения случайных величин с помощью ядерных оценок.

17. Применение моделей выживаемости для прогнозирования надежности технических систем.

18. Тестирование статистических гипотез методами повторной выборки на реальных данных.

19. Анализ временных рядов энергопотребления предприятия для краткосрочного прогнозирования нагрузок.

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

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

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

№ п/п	Библиографическое описание	Место доступа
1	Casella G., Berger R.L. Statistical Inference. – Cengage Learning, 2008 – 700 с.	https://faculty.ksu.edu.sa/sites/default/files/casella_berger_statistical_inference1.pdf (дата обращения: 24.11.2026)
2	Wasserman L. All of Statistics: A Concise Course in Statistical Inference. – New York: Springer, 2004 – 446 с.	https://egrcc.github.io/docs/math/all-of-statistics.pdf (дата обращения: 24.11.26)
3	Кельберт М.Я., Сухов Ю.М. Основные понятия теории вероятностей и математической статистики – Москва: МЦНМО, 2007 – 456 с.	https://djvu.online/file/MxVgEXrKO8jR1 (дата обращения: 24.11.26)
4	Андерсон Т. Статистический анализ временных рядов – Москва: Мир, 2012 - 751 с. ISBN 978-5-458-26140-1	https://libcats.org/book/719600 (дата обращения: 24.11.26)

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).

Для проведения занятий лекционного типа требуются аудитории, оснащенные компьютерной техникой и наборами демонстрационного оборудования. Для практических занятий – наличие персональных компьютеров. Для лабораторных занятий – наличие персональных компьютеров.

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:

старший преподаватель, к.н.
кафедры «Математическое
моделирование сложных систем»
Института железнодорожного
транспорта С.В. Дрожжин

Согласовано:

и.о. заведующего кафедрой ПМ С.А. Тищенко

и.о. заведующего кафедрой ММСС С.А. Тищенко

Председатель учебно-методической
комиссии Н.А. Андриянова