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
27.03.05 Инноватика,
утвержденной первым проректором РУТ (МИИТ)
Тимониным В.С.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Анализ больших данных

Направление подготовки: 27.03.05 Инноватика

Направленность (профиль): Управление цифровыми инновациями

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: 87771
Подписал: заведующий кафедрой Куликов Михаил Юрьевич
Дата: 10.07.2026

1. Общие сведения о дисциплине (модуле).

Основная цель данной дисциплины – формирование у студентов базовых теоретических знаний в области теории вероятностей и математической статистики и способности к применению технологий обработки данных, в том числе big data и машинного обучения к решению прикладных задач, связанных с оказанием услуг для цифровизации и цифровой трансформации.

Задача курса – изучение основных методов хранения и анализа больших данных.

2. Планируемые результаты обучения по дисциплине (модулю).

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

ОПК-3 - Способен применять базовые цифровые и информационные технологии, включая методы искусственного интеллекта и машинного обучения, для сбора, обработки, хранения, передачи и анализа данных, прогнозирования, оптимизации и автоматизации процессов в профессиональной деятельности;

ПК-2 - Способность управлять операционной деятельностью организации в области ИТ;

ПК-3 - Способность разрабатывать план комплексного развития пассажирского транспорта в городских агломерациях.

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

Знать:

- международные и отечественные стандарты, лучшие практики и фреймворки по управлению информационной безопасностью;
- методы, используемые при сборе, систематизации и анализе социологических, экономических данных;
- методы обработки и анализа материалов исследований транспортных объектов;
- средства и методы сбора и обработки данных об объективных условиях участка застройки, включая обмеры, фотофиксацию, вычерчивание генерального плана местности, макетирование, графическую фиксацию;
- теоретические основы методов анализа больших данных и машинного обучения;

- математические методы и методики расчетов;

- методы и модели оценки компетенций ИТ-персонала;

- основы системного анализа.

Уметь:

- интерпретировать полученные математические результаты на практике;

- проводить анализ исходных данных для поектирования подсистем и средств обеспечения информационной безопасности;

- выбирать методы оценки компетенций персонала ИТ-подразделения;

- использовать методы обучения и развития персонала ИТ-подразделения для формирования целевой модели компетенций;

- осуществлять мониторинг и контроль управления ИТ-активами;

- рассчитывать экономическую, экологическую, эксплуатационную эффективность внедрения цифровых решений на пассажирском транспорте в городской агломерации;

- обрабатывать запросы пассажиров и предложения перевозчиков по поводу эффективности и качества функционирования существующих цифровых решений на пассажирском транспорте в городской агломерации.

Владеть:

- навыком участвовать в проведении ТЭО соответствующих проектных решений;

- основными методами, подходами и инструментарием анализа больших данных и машинного обучения;

- навыком критически анализировать современные проблемы инноватики с использованием анализа больших

данных и машинного обучения;

- контролем качества и управлением ИТ-проектами;

- определением состава методов и средств обеспечения безопасности ИТ, соответствующих критериям оценки безопасности ИТ;

- определением затрат на внедрение, эксплуатацию, обслуживание и ремонт цифровых решений на пассажирском транспорте в городской агломерации;

- подготовкой аналитических документов о транспортных системах городских агломераций.

3. Объем дисциплины (модуля).

3.1. Общая трудоемкость дисциплины (модуля).

Общая трудоемкость дисциплины (модуля) составляет 2 з.е. (72 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):	42	42
В том числе:		
Занятия лекционного типа	28	28
Занятия семинарского типа	14	14

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 30 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	Введение в большие данные Рассматриваемые вопросы: - терминология, история появления; - технические сложности работы с большими данными; - распределенные файловые системы.
2	Модель вычислений MapReduce Рассматриваемые вопросы: - модель вычислений MapReduce;

№ п/п	Тематика лекционных занятий / краткое содержание
	- реализация MapReduce в Hadoop; - планирование вычислений.
3	Основы языка SAS base Рассматриваемые вопросы: - изучение основ программирования на языке SAS; - структуры языка, запуска и отладки программы; - простейшие аналитические процедуры и приёмы для трансформации данных.
4	Макропроцессор SAS. SAS SQL. Рассматриваемые вопросы: - изучение основ программирования на языке SAS Macro; - использование макропрограмм и макропеременных для написания программ SAS со сложной структурой и логикой.
5	Обзор SAS Data Loader for Hadoop Рассматриваемые вопросы: - Hadoop как технология хранения и обработки больших данных. Способы загрузки данных в Hadoop; - базовые операции с данными. Профилирование, дедупликация. Выполнение процессов Data Quality внутри Hadoop; - интеграция с SAS In-Memory Analytics Server.
6	Обзор задач и методов иинтеллектуального анализа больших данных Рассматриваемые вопросы: - аналитические методы, применимые к большим объёмам данных; - многомерные связи, ассоциации, корреляции; - непрерывность поступающих обновлений как характерная черта анализа больших данных. - примеры обработки неструктурированных данных; - понятия неоднозначности и недостоверно сти данных.
7	SAS Enterprise Miner Рассматриваемые вопросы: - регрессионные модели и деревья решений; - поиск стандартных путей по истории процессов, частотный анализ и выявление отклонений; - диагностика и оценка качества моделей. Примеры: кредитные модели и оценка рисков.

4.2. Занятия семинарского типа.

Практические занятия

№ п/п	Тематика практических занятий/краткое содержание
1	Реализации алгоритмов на MapReduce Рассматриваемые вопросы: - примеры реализации алгоритмов на MapReduce (включая умножение матриц, операции реляционной алгебры); - колоночные форматы хранения (на примере ORC).
2	Интерактивное исследование данных с помощью SAS VA/VS Рассматриваемые вопросы: - примеры визуализаций базового статистического анализа, доступных в SAS VA/VS.
3	Обработка текстов в SAS Text Miner Рассматриваемые вопросы: - применение технологий текстовой аналитики: обогащение информации по клиентам компаний, мониторинг потоков сообщений;

№ п/п	Тематика практических занятий/краткое содержание
	- принципы статистического анализа текста (Text Mining): лингвистическая предобработка текста, статистическая фильтрация текста, автоматическое выявление тематик, кластеризация текстов.
4	Анализ временных рядов (SAS Forecast Studio) Рассматриваемые вопросы: - использование интерактивного интерфейса SAS Forecast Studio; - создание и использование прогнозных моделей для временных рядов; - автоматическое создание и построение прогнозных моделей для данных большого объёма, содержащих временные ряды.

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Подготовка к практическим занятиям
2	Подготовка к промежуточной аттестации
3	Подготовка к текущему контролю
4	Подготовка к промежуточной аттестации.
5	Подготовка к текущему контролю.

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

№ п/п	Библиографическое описание	Место доступа
1	Основы эконометрики Березинец И. В. Высшая школа менеджмента , 2011	https://znanium.ru/catalog/document?id=184249
2	Модели финансового рынка и прогнозирование в финансовой сфере Новиков А.И. ИНФРА-М , 2022	https://znanium.ru/catalog/document?id=399971
3	Эконометрика: теоретические основы Соколов Г.А. ИНФРА-М , 2022	https://znanium.ru/catalog/document?id=414305
4	Анализ данных: моделирование инвестиционного портфеля Цогоева А.Р., Цогоев А.Ю., Волик М.В. Прометей , 2023	https://znanium.ru/catalog/document?id=443483
5	Интеллектуальный анализ больших данных Алетдинова А.А., Муртазина М.Ш. Новосибирский государственный технический университет , 2023	https://znanium.ru/catalog/document?id=479552

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

- Информационный портал Научная электронная библиотека eLIBRARY.RU (www.elibrary.ru);
- Единая коллекция цифровых образовательных ресурсов (http://window.edu.ru);
- Научно-техническая библиотека РУТ (МИИТ) (http://library.miit.ru).

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

- Операционная система Windows;
- Microsoft Office;
- Microsoft Azura;
- SAS University Edition;
- SAS OnDemand for Academics (Enterprise Miner, Text Miner, Forecast Server, Enterprise
- Guide + SAS OR + SAS ETS);
- SAS Access for Hadoop.

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

- персональный компьютер и мультимедийный проектор;
- лазерная указка;
- учебный видеофильм и прочие видеоматериалы;
- информационные слайды, презентации;
- Microsoft office ОС.

9. Форма промежуточной аттестации:

Зачет 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 «Транспортное машиностроение, сертификация и управление инновациями»	В.Н. Тарасова

Согласовано:

Заведующий кафедрой УИТ	В.Н. Тарасова
Заведующий кафедрой ТМСУИ	М.Ю. Куликов
Председатель учебно-методической комиссии	Н.А. Андриянова