МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА»
(РУТ (МИИТ)

Рабочая программа дисциплины (модуля),
как компонент образовательной программы
высшего образования - программа бакалавриата
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: 30.04.2025

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-6 - Способен обосновывать принятие технического решения при разработке инновационного проекта, выбирать технические средства и технологии, в том числе с учетом экологических последствий их применения;

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

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

Уметь:

- способен применять научные законы;
- интерпретировать полученные математические результаты на практике;
- способен проводить анализ исходных данных для поектирования подсистем и средств обеспечения информационной безопасности.

Знать:

- математические методы и методики расчетов
- международные и отечественные стандарты, лучшие практики и фреймворки по управлению информационной безопастностью
- теоретические основы методов анализа больших данных и машинного обучения

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

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 - 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 В результате работы на практическом занятии обучающиеся ознакомятся с примерами реализации алгоритмов на MapReduce (включая умножение матриц, операции реляционной алгебры) и изучат колоночные форматы хранения (на примере ORC).

№ п/п	Тематика практических занятий/краткое содержание
2	Интерактивное исследование данных с помощью SAS VA/VS В результате работы на практическом занятии обучающиеся ознакомятся с примерами визуализаций базового статистического анализа, доступных в SAS VA/VS.
3	Обработка текстов в SAS Text Miner В результате работы на практическом занятии обучающиеся ознакомятся с применением технологий текстовой аналитики: обогащение информации по клиентам компаний, мониторинг потоков сообщений, изучат принципы статистического анализа текста (Text Mining): лингвистическая предобработка текста, статистическая фильтрация текста, автоматическое выявление тематик, кластеризация текстов.
4	Анализ временных рядов (SAS Forecast Studio). В результате работы на практическом занятии обучающиеся ознакомятся с использованием интерактивного интерфейса SAS Forecast Studio для создания и использования прогнозных моделей для временных рядов, в том числе для автоматического создания и построения 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.

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:

заведующий кафедрой, профессор,
д.н. кафедры «Управление
инновациями на транспорте» В.Н. Тарасова

Согласовано:

Заведующий кафедрой УИТ В.Н. Тарасова

Заведующий кафедрой ТМСУИ М.Ю. Куликов

Председатель учебно-методической
комиссии Н.А. Андриянова