

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 - программы 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: 5665
Подписал: заведующий кафедрой Нугович Вероника
Евгеньевна
Дата: 25.10.2023

1. Общие сведения о дисциплине (модуле).

Целью дисциплины "Анализ больших данных" является изучить методы и подходы анализа больших данных. Задачей дисциплины является научить обучающихся критически анализировать современные проблемы инноватики, ставить задачи и разрабатывать программу исследования, выбирать соответствующие методы решения экспериментальных и теоретических задач, интерпретировать, представлять и применять полученные результаты.

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-11 - Способность проводить исследования инновационной инфраструктуры.

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

Уметь:

ставить задачи и разрабатывать программу исследования с использованием анализа больших данных;

выбирать соответствующие методы решения экспериментальных и теоретических задач;

применять методы и подходы анализа больших данных для решения экспериментальных и теоретических задач;

интерпретировать, представлять и применять результаты, полученные с использованием анализа больших данных.

Знать:

теоретические основы методов анализа больших данных;

типы задач, решаемые с использованием методов анализа больших данных;

современные технологии анализа больших данных.

Владеть:

основными методами, подходами и инструментарием анализа больших данных;

навыком критически анализировать современные проблемы инноватики

с использованием анализа больших данных.

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	Сем. №8
Контактная работа при проведении учебных занятий (всего):	48	48
В том числе:		
Занятия лекционного типа	24	24
Занятия семинарского типа	24	24

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 24 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 интеллектуального анализа больших данных Рассматриваемые вопросы: - аналитические методы, применимые к большим объемам данных - многомерные связи, ассоциации, корреляции - непрерывность поступающих обновлений как характерная черта анализа больших данных. - примеры обработки неструктурированных данных - понятия неоднозначности и недостоверности 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 объема, содержащих временные ряды.

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Самостоятельное изучение разделов (тем) дисциплины (модуля).
2	Работа с литературой. Подготовка к текущим занятиям.
3	Подготовка к текущему контролю.
4	Подготовка к промежуточной аттестации.
5	Подготовка к промежуточной аттестации.
6	Подготовка к текущему контролю.

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

№ п/п	Библиографическое описание	Место доступа
1	Глубокое обучение Гудфеллоу Я., Бенджио И., Курвилль А.е / пер. с англ. Слинкина А.А. – 2-е изд., испр. М.: ДМК Пресс, 2018	https://klex.ru/157x . Текст: электронный.
2	Основы эконометрики	http://elib.osu.ru/bitstream/123456789/11339/1/3794_20130919.pdf?ysclid=l . Текст: электронный.

	Реннер А.Г., Стебунова О.И., Туктамышева Л.М. Учебное пособие Оренбург: ГОУ ОГУ , 2009	
3	Диффузионные модели в финансовом анализе Медведев Г. А. Книга Минск : Белорусский государственный университет , 2010	https://e-catalog.nlb.by/Record/BY-NLB-br0000503786?ysclid=lbbw03ybze520230613 . Текст: электронный.
4	Методы распознавания Горелик А.Л., Скрипкин В.А. - 4- е изд., испр. Учебное пособие Москва : Высшая школа , 2004	https://search.rsl.ru/ru/record/01002451622?ysclid=lbbw4fdrrmm460013909 . Текст: электронный.
5	Нейронные сети на персональном компьютере Горбань А.Н., Россиев Д.А.; Отв. ред. Быков В.И. Книга Новосибирск : Наука : Сиб. изд. фирма , 1996	https://search.rsl.ru/ru/record/01001770680?ysclid=lbbw6vpzqc390131576 . Текст: электронный.
6	Анализ данных и процессов Барсегян А.А., Куприянов М.С., Холод И.И., Тесс М.Д., Елизаров С.И. - 3-е изд., перераб. и доп. Учебное пособие СПб.: БХВ- Петербург , 2009	НТБ РУТ (МИИТ)

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>);
- Интернет-университет информационных технологий <http://www.intuit.ru/>;
- Научно-техническая библиотека РУТ (МИИТ) (<http://library.miit.ru>);
- SAS Programming I: Essentials (<https://support.sas.com/edu/schedules.html?ctry=us&id=277>).

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;
- ChartJS, библиотека для создания интерактивных графиков и диаграмм в веб-приложениях, и она распространяется под лицензией MIT, которая является лицензией с открытым исходным кодом. Это позволяет бесплатно использовать, изменять и распространять Chart.js в своих образовательных проектах;
- Яндекс.Облако. Облачные сервисы Yandex Cloud, технологии для решения актуальных задач образования, которые отвечают современным запросам студентов и преподавателей. Современные инфраструктурные решения помогут построить цифровую образовательную среду и реализовать концепцию цифрового университета. Цена договорная по академической лицензии)
- R7 – отечественный аналог Microsoft Office (90% скидка образовательным организациям).

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

Лекционная аудитория должна быть оборудована персональным компьютером и мультимедийным проектором для демонстрации презентационных материалов, лазерной указкой. Учебные видеофильмы и прочие видеоматериалы. Информационные слайды, презентации. В случае проведения занятий с применением электронного обучения и дистанционных образовательных технологий необходимо наличие компьютерной техники, для организации коллективных и индивидуальных форм общения педагогических работников со студентами.

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:

старший преподаватель кафедры
«Цифровые технологии управления
транспортными процессами»

И.С. Разживайкин

Согласовано:

Заведующий кафедрой УИТ

В.Н. Тарасова

Заведующий кафедрой ЦТУТП

В.Е. Нутович

Председатель учебно-методической
комиссии

С.В. Володин