

**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: 2221  
Подписал: заведующий кафедрой Тарасова Валентина  
Николаевна  
Дата: 07.10.2024

## 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;

**ПК-3** - Способность выполнять работы по осуществлению финансово-экономической деятельности структурного подразделения.

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 проектирования подсистем и средств обеспечения информационной безопасности.

### **Знать:**

- математические методы и методики расчетов
- международные и отечественные стандарты, лучшие практики и фреймворки по управлению информационной безопасностью
- 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	<b>Введение в большие данные</b> Рассматриваемые вопросы: - терминология, история появления; - технические сложности работы с большими данными; - распределенные файловые системы;
2	<b>Модель вычислений MapReduce</b> Рассматриваемые вопросы: - модель вычислений MapReduce - реализация MapReduce в Hadoop - планирование вычислений.
3	<b>Основы языка SAS base</b> Рассматриваемые вопросы: - изучение основ программирования на языке SAS - структуры языка, запуска и отладки программы - простейшие аналитические процедуры и приёмы для трансформации данных
4	<b>Макропроцессор SAS. SAS SQL.</b> Рассматриваемые вопросы: - изучение основ программирования на языке SAS Macro - использование макропрограмм и макропеременных для написания программ SAS со сложной структурой и логикой.
5	<b>Обзор SAS Data Loader for Hadoop</b> 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	<b>Обзор задач и методов интеллектуального анализа больших данных</b> 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	<b>SAS Enterprise Miner</b> 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	<b>Реализации алгоритмов на MapReduce</b> 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	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	Реннер, А.Г. Основы эконометрики: учебное пособие/ А.Г. Реннер, О.И. Стебунова, Л.М. Туктамышева. – Оренбург: ГОУ ОГУ, 2009. – 156 с.	URL: <a href="http://elib.osu.ru/bitstream/123456789/11339/1/3794_20130919.pdf?ysclid=1bbvpvh5hz44505028">http://elib.osu.ru/bitstream/123456789/11339/1/3794_20130919.pdf?ysclid=1bbvpvh5hz44505028</a> (дата обращения 06.12.2023г.)
2	Медведев Г. А. Диффузионные модели в финансовом анализе / Г. А.	URL: <a href="https://e-catalog.nlb.by/Record/BY-NLB-br0000503786?ysclid=lbbw03ybze520230613">https://e-catalog.nlb.by/Record/BY-NLB-br0000503786?ysclid=lbbw03ybze520230613</a> (дата обращения 06.12.2023г.)

	Медведев. - Минск : БГУ, 2010. - 159 с. ISBN 978-985- 518-275-8	
3	Горелик, Александр Леопольдович Методы распознавания [Текст] : учебное пособие для студентов вузов, обучающихся по специальности "Автоматизирова нные системы обработки информации и управления" направления подготовки дипломированны х специалистов "Информатика и вычислительная техника" / А. Л. Горелик, В. А. Скрипкин. - 4-е изд., испр. - Москва : Высшая школа, 2004 (ГУП Смол. обл. тип. им. В.И. Смирнова). - 260, [1] с. : ил., табл.; 22 см.; ISBN 5- 06-004396-7	URL: <a href="https://search.rsl.ru/ru/record/01002451622?ysclid=lbbw4fdrmm460013909">https://search.rsl.ru/ru/record/01002451622?ysclid=lbbw4fdrmm460013909</a> (дата обращения 06.12.2023г.)
4	Горбань, Александр Николаевич Нейронные сети на персональном компьютере / А. Н. Горбань, Д. А.	URL: <a href="https://search.rsl.ru/ru/record/01001770680?ysclid=lbbw6vpzqc390131576">https://search.rsl.ru/ru/record/01001770680?ysclid=lbbw6vpzqc390131576</a> (дата обращения 06.12.2023г.)

	<p>Россиев; Отв. ред. В. И. Быков; Рос. акад. наук, Сиб. отд-ние, Вычисл. центр (г. Красноярск). - Новосибирск : Наука : Сиб. изд. фирма, 1996. - 270,[5] с.; 21 см.; ISBN 5-02-031196-0 (В пер.)</p>	
5	<p>Барсегян, А. А. Анализ данных и процессов: учеб. пособие / А. А. Барсегян, М. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров. — 3-е изд., перераб. и доп. — СПб.: БХВ-Петербург, 2009. — 512 с. ISBN 978-5-9775-0368-6</p>	<p>URL: <a href="http://www.kist.ntu.edu.ua/textPhD/AnalizDannyhIProcessov.pdf">http://www.kist.ntu.edu.ua/textPhD/AnalizDannyhIProcessov.pdf</a> (дата обращения 06.12.2023г.)</p>

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](http://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;

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:

доцент, доцент, к.н. кафедры  
«Управление инновациями на  
транспорте»

В.Б. Ручкин

Согласовано:

Заведующий кафедрой УИТ  
Председатель учебно-методической  
комиссии

В.Н. Тарасова

С.В. Володин