

МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА»

УТВЕРЖДАЮ:

Директор ИУЦТ



С.П. Вакуленко

06 октября 2020 г.

Кафедра «Цифровые технологии управления транспортными процессами»

Автор Лецкий Эдуард Константинович, д.т.н., профессор

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Анализ данных

Направление подготовки:	09.04.01 – Информатика и вычислительная техника
Магистерская программа:	Технологии разработки информационных систем
Квалификация выпускника:	Магистр
Форма обучения:	очная
Год начала подготовки	2019

Одобрено на заседании Учебно-методической комиссии института Протокол № 3 05 октября 2020 г. Председатель учебно-методической комиссии  Н.А. Клычева	Одобрено на заседании кафедры Протокол № 2 02 октября 2020 г. Заведующий кафедрой  В.Е. Нутович
--	---

Москва 2020 г.

1. ЦЕЛИ ОСВОЕНИЯ УЧЕБНОЙ ДИСЦИПЛИНЫ

Цель изучения дисциплины – освоении методов и приемов анализа данных в информационно-аналитических системах, осуществляемого для поддержки принятия решений в различных областях деятельности (бизнес, транспорт, финансы и пр.).

Задачи дисциплины:

- дать знания в области методов решения широкого круга задач анализа данных, в том числе, «больших данных» (Big Data);
- освоить приемы применения программных платформ (Data Mining, Business Intelligence) при анализе данных.

Дисциплина предназначена для получения знаний для решения следующих профессиональных задач (в соответствии с видами деятельности):

Научно-исследовательская деятельность:

- разработка математических моделей исследуемых процессов и изделий;
- сбор, обработка, анализ и систематизация научно-технической информации по теме исследования, выбор методик и средств решения задачи.

2. МЕСТО УЧЕБНОЙ ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Учебная дисциплина "Анализ данных" относится к блоку 1 "Дисциплины (модули)" и входит в его базовую часть.

2.1. Наименования предшествующих дисциплин

2.2. Наименование последующих дисциплин

3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ), СООТНЕСЕННЫЕ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

В результате освоения дисциплины студент должен:

№ п/п	Код и название компетенции	Ожидаемые результаты
1	ОПК-3 Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями	ОПК-3.1 Знать основные методы поиска, хранения, обработки и анализа информации из различных источников и баз данных, представления её в требуемом формате с использованием информационных, компьютерных и сетевых технологий. ОПК-3.2 Уметь анализировать профессиональную информацию, выделять в ней главное. ОПК-3.3 Владеть навыками анализа и структурирования информации, оформления и составления научно-технических отчётов и докладов; по информационному обслуживанию и обработке данных в области производственной деятельности.
2	ОПК-7 Способен адаптировать зарубежные комплексы обработки информации и автоматизированного проектирования к нуждам отечественных предприятий	ОПК-7.1 Владеет навыками разработки алгоритмов решения типовых профессиональных задач и навыками применения технических и программных средств моделирования и проектирования. ОПК-7.2 Может работать с зарубежными комплексами обработки информации и автоматизированного проектирования, проводить сравнительный анализ их эффективности и надежности. ОПК-7.3 Знает способы адаптации зарубежных комплексов обработки информации и автоматизированного проектирования к нуждам отечественных предприятий.
3	ПКО-6 Владение существующими методами и алгоритмами решения задач распознавания и обработки данных	ПКО-6.1 Знать методы распознавания и обработки данных; основные тенденции развития информационных технологий в области БД. ПКО-6.2 Уметь выявлять проблемы организации, связанные с информационным обеспечением и особенностями распознавания и обработки данных. ПКО-6.3 Владеть навыками сбора и анализа данных.
4	ПКР-2 Применение перспективных методов исследования и решения профессиональных задач на основе знания мировых тенденций развития вычислительной техники и информационных технологий	ПКР-2.1 Знать методы исследования и решения профессиональных задач; мировые тенденции развития вычислительной техники; знать перспективные тенденции развития информационных технологий. ПКР-2.2 Уметь применять перспективные методы исследования для решения. ПКР-2.3 Владеть навыками применения перспективных методов исследования и решения профессиональных задач на основе знания мировых тенденций развития вычислительной техники и информационных технологий.
5	УК-3 Способен организовать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели	УК-3.1 Знать принципы и методы организации и руководства работой команды. УК-3.2 Уметь ставить цели и определять пути их достижения. УК-3.3 Владеть навыками организатора и

№ п/п	Код и название компетенции	Ожидаемые результаты
		руководителя команды.

4. ОБЪЕМ ДИСЦИПЛИНЫ (МОДУЛЯ) В ЗАЧЕТНЫХ ЕДИНИЦАХ И АКАДЕМИЧЕСКИХ ЧАСАХ

4.1. Общая трудоемкость дисциплины составляет:

6 зачетных единиц (216 ак. ч.).

4.2. Распределение объема учебной дисциплины на контактную работу с преподавателем и самостоятельную работу обучающихся

Вид учебной работы	Количество часов	
	Всего по учебному плану	Семестр 3
Контактная работа	44	44,15
Аудиторные занятия (всего):	44	44
В том числе:		
лекции (Л)	18	18
практические (ПЗ) и семинарские (С)	18	18
лабораторные работы (ЛР)(лабораторный практикум) (ЛП)	8	8
Самостоятельная работа (всего)	172	172
ОБЩАЯ трудоемкость дисциплины, часы:	216	216
ОБЩАЯ трудоемкость дисциплины, зач.ед.:	6.0	6.0
Текущий контроль успеваемости (количество и вид текущего контроля)	КР (1), ПК1, ПК2	КР (1), ПК1, ПК2
Виды промежуточной аттестации (экзамен, зачет)	ЗаО	ЗаО

4.3. Содержание дисциплины (модуля), структурированное по темам (разделам)

№ п/п	Семестр	Тема (раздел) учебной дисциплины	Виды учебной деятельности в часах/ в том числе интерактивной форме						Формы текущего контроля успеваемости и промежуточной аттестации
			Л	ЛР	ПЗ/ПП	КСР	СР	Всего	
1	2	3	4	5	6	7	8	9	10
1	3	Раздел 1 Компьютерный анализ данных: цели, возможности, прикладное значение, инструментальные средства	4		2		28	34	ПК1, ПК2, (тестирование)
2	3	Тема 1.1 Определения основных понятий. Задачи анализа. Особенности технологий анализа «больших данных».	2					2	
3	3	Тема 1.2 Основные положения теории статистического оценивания. Свойства оценок. Критерии оценивания. Статистическая проверка гипотез.	2					2	
4	3	Раздел 2 Методы и алгоритмы предварительного анализа данных	3		4		28	35	ПК1, ПК2, (тестирование)
5	3	Тема 2.1 Фильтрация массивов наблюдаемых данных. Выявление грубых ошибок наблюдения.	1					1	
6	3	Тема 2.2 Проверка независимости результатов последовательных наблюдений.	2					2	
7	3	Раздел 3 Анализ случайных последовательностей	3	4	6		44	57	ПК1, ПК2, (тестирование)
8	3	Тема 3.1 Проверка стационарности случайного процесса.	1					1	

№ п/п	Семестр	Тема (раздел) учебной дисциплины	Виды учебной деятельности в часах/ в том числе интерактивной форме						Формы текущего контроля успеваемости и промежуточной аттестации
			Л	ЛР	ПЗ/ТП	КСР	СР	Всего	
1	2	3	4	5	6	7	8	9	10
9	3	Тема 3.2 Выявление скрытых закономерностей (тренды, сезонные колебания и пр.). Обоснование выбора модели потока событий.	2					2	
10	3	Раздел 4 Построение и анализ зависимостей	4	4	6		36	50	ПК1, ПК2, (тестирование)
11	3	Тема 4.1 Методы выявления влияющих факторов.	2					2	
12	3	Тема 4.2 Построение и статистический анализ регрессионных моделей. Задачи классификации.	2					2	
13	3	Раздел 5 Инструментальные средства поддержки анализа данных	4				36	40	ПК1, ПК2, (тестирование)
14	3	Тема 5.1 Особенности анализа «больших данных». Общая характеристика платформ поддержки анализа данных.	2					2	
15	3	Тема 5.2 Рекуррентные алгоритмы вычисления оценок. Технологии анализа данных в информационно-аналитических системах.	2					2	
16	3	Раздел 6 Зачет с оценкой						0	ЗаО, КР
17		Всего:	18	8	18		172	216	

4.4. Лабораторные работы / практические занятия

Практические занятия предусмотрены в объеме 18 ак. ч.

№ п/п	№ семестра	Тема (раздел) учебной дисциплины	Наименование занятий	Всего часов/ из них часов в интерактивной форме
1	2	3	4	5
1	3	РАЗДЕЛ 1 Компьютерный анализ данных: цели, возможности, прикладное значение, инструментальные средства	Статистические методы обработки данных	2
2	3	РАЗДЕЛ 2 Методы и алгоритмы предварительного анализа данных	Методы и алгоритмы предварительного анализа данных	4
3	3	РАЗДЕЛ 3 Анализ случайных последовательностей	Проверка стационарности случайной последовательности	2
4	3	РАЗДЕЛ 3 Анализ случайных последовательностей	Выявление скрытых закономерностей (тренды, сезонные колебания)	4
5	3	РАЗДЕЛ 4 Построение и анализ зависимостей	Построение и анализ зависимостей методом регрессионного анализа	4
6	3	РАЗДЕЛ 4 Построение и анализ зависимостей	Методы выявления влияющих факторов	2
ВСЕГО:				18/0

Лабораторные работы предусмотрены в объеме 8 ак. ч.

№ п/п	№ семестра	Тема (раздел) учебной дисциплины	Наименование занятий	Всего часов/ из них часов в интерактивной форме
1	2	3	4	5
1	3	РАЗДЕЛ 3 Анализ случайных последовательностей	Применение статистических пакетов при предварительном анализе данных	4
2	3	РАЗДЕЛ 4 Построение и анализ зависимостей	Применение статистических пакетов при построении зависимостей	4
ВСЕГО:				8/0

4.5. Примерная тематика курсовых проектов (работ)

Курсовая работа выполняется с целью приобретения практических навыков анализа данных при решении реальных проблем в различных сферах деятельности. Все проекты выполняются с использованием стандартных инструментальных средств анализа данных и предполагают работу с конкретными данными.

Примерные темы:

1. Анализ тенденций изменения цены на нефть.
2. Анализ факторов, влияющих на надежность подвижного состава метрополитена.
3. Оценка степени влияния различных факторов на показатели безопасности процессов перевозок на железнодорожном транспорте.
4. Построение и анализ многофакторных моделей для показателей, характеризующих эффективность функционирования организации (предприятия, отрасли и т.п.)

В качестве курсовой работы может быть выполнен аналитический обзор или реферат по заданной преподавателем теме с докладом на научном семинаре магистрантов.

5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Лекционные занятия должны проходить в аудиториях, оборудованных видеотехникой (экран, проектор, компьютер), в при наличии у студентов опорного конспекта, который лектор размещает на сайте кафедры, а студенты имеют возможность скачать и распечатать.

Методические материалы по курсу (опорный конспект, описания лабораторных работ, рекомендации по выполнению курсового проекта и пр.) размещаются на сайте кафедры. Студентам предоставляется возможность общения с преподавателями в режиме off line через сайт кафедры.

Проведении занятий по дисциплине (модулю) возможно с применением электронного обучения и дистанционных образовательных технологий, реализуемые с применением информационно-телекоммуникационных сетей при опосредованном (на расстоянии) взаимодействии обучающихся и педагогических работников.

В процессе проведения занятий с применением электронного обучения и дистанционных образовательных технологий применяются современные образовательные технологии, такие как (при необходимости):

- использование современных средств коммуникации;
- электронная форма обмена материалами;
- дистанционная форма групповых и индивидуальных консультаций;
- использование компьютерных технологий и программных продуктов, необходимых для сбора и систематизации информации, проведения требуемых программой расчетов и т.д.

6. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

№ п/п	№ семестра	Тема (раздел) учебной дисциплины	Вид самостоятельной работы студента. Перечень учебно-методического обеспечения для самостоятельной работы	Всего часов
1	2	3	4	5
1	3	РАЗДЕЛ 1 Компьютерный анализ данных: цели, возможности, прикладное значение, инструментальные средства	Подготовка к тестированию по теории вероятности и математической статистике	28
2	3	РАЗДЕЛ 2 Методы и алгоритмы предварительного анализа данных	Подготовка к тестированию по теории вероятности и математической статистике	28
3	3	РАЗДЕЛ 3 Анализ случайных последовательностей	Подготовка к защите лабораторной работы №1. Выполнение курсовой работы	44
4	3	РАЗДЕЛ 4 Построение и анализ зависимостей	Подготовка к защите лабораторных работ №2. Выполнение курсовой работы	36
5	3	РАЗДЕЛ 5 Инструментальные средства поддержки анализа данных	Выполнение и защита курсовой работы	36
ВСЕГО:				172

7. ПЕРЕЧЕНЬ ОСНОВНОЙ И ДОПОЛНИТЕЛЬНОЙ ЛИТЕРАТУРЫ, НЕОБХОДИМОЙ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

7.1. Основная литература

№ п/п	Наименование	Автор (ы)	Год и место издания Место доступа	Используется при изучении разделов, номера страниц
1	Теория статистики	Р.А. Шмойлова, В.Г. Минашкин, Н.А. Садовникова, Е.Б. Шувалова ; Под ред. Р.А. Шмойловой.	Финансы и статистика, 2007 НТБ МИИТ: ФБ (ауд. 1230)-3, ЧЗ№2-2, УБ№2 (ауд. 3115)-60	Все разделы
2	Интеллектуальные системы и технологии	Б. Я. Советов, В. В. Цехановский, В. Д. Чертовской	Академия, 2013 НТБ МИИТ: ФБ (ауд. 1230)-2	Все разделы

7.2. Дополнительная литература

№ п/п	Наименование	Автор (ы)	Год и место издания Место доступа	Используется при изучении разделов, номера страниц
3	Анализ результатов наблюдений. Разведочный анализ	Дж. Тьюки; Пер.: А.Ф. Кушнир, А.Л. Петросян, Е.Л. Резников; Под ред. В.Ф. Писаренко	Мир, 1981 НТБ (фб.)	Все разделы

8. ПЕРЕЧЕНЬ РЕСУРСОВ ИНФОРМАЦИОННО-ТЕЛЕКОММУНИКАЦИОННОЙ СЕТИ "ИНТЕРНЕТ", НЕОБХОДИМЫЕ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

- Интернет: <http://www.intuit.ru/goto/cou>
- Материалы по дисциплине на сайте кафедры АСУ: <http://www.miitasu.ru>

9. ПЕРЕЧЕНЬ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ И ИНФОРМАЦИОННЫХ СПРАВОЧНЫХ СИСТЕМ, ИСПОЛЬЗУЕМЫХ ПРИ ОСУЩЕСТВЛЕНИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

Учебный процесс включает освоение программных продуктов компании SAS - SAS OnDemand for Academics. Так же необходимо: Windows 7, Microsoft Office 2013, Microsoft Office 2007, Microsoft Essential Security 2012.

При организации обучения по дисциплине (модулю) с применением электронного обучения и дистанционных образовательных технологий необходим доступ каждого студента к информационным ресурсам – библиотечному фонду Университета, сетевым ресурсам и информационно-телекоммуникационной сети «Интернет».

В случае проведения занятий с применением электронного обучения и дистанционных образовательных технологий может потребоваться наличие следующего программного обеспечения (или их аналогов): ОС Windows, Microsoft Office, Интернет-браузер, Microsoft Teams и т.д.

В образовательном процессе, при проведении занятий с применением электронного обучения и дистанционных образовательных технологий, могут применяться следующие

средства коммуникаций: ЭИОС РУТ(МИИТ), Microsoft Teams, электронная почта, скайп, Zoom, WhatsApp и т.п.

10. ОПИСАНИЕ МАТЕРИАЛЬНО ТЕХНИЧЕСКОЙ БАЗЫ, НЕОБХОДИМОЙ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

Лекционные аудитории оборудуются видеопроекционной аппаратурой, устройствами для затемнения окон.

Лабораторные занятия проводятся в компьютерном классе с использованием пакета SAS OnDemand for Academics.

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, посредством используемых средств коммуникации.

Допускается замена оборудования его виртуальными аналогами.

11. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

Необходимым условием освоения дисциплины является знание теории вероятностей, теории случайных процессов, основ математической статистики. В случае недостаточного владения указанными дисциплинами (это проверяется при тестировании студентов на втором практическом занятии) студенту необходимо самостоятельно изучить материалы по соответствующим курсам, выложенные на сайте кафедры (www.miitasu.ru).

Во время лекций студент обязан иметь распечатанный опорный конспект (скачать с сайта кафедры), в котором, при необходимости, можно делать пояснения, замечания и пр.

Задания на курсовую работу размещаются на сайте кафедры. Срок сдачи пояснительной записки по курсовой работе устанавливает преподаватель. Задержки с представлением пояснительной записки учитываются при оценке курсовой работы.

Вопросы для экзамена помещены на сайт кафедры.