

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
09.03.01 Информатика и вычислительная техника,
утвержденной первым проректором РУТ (МИИТ)
Тимониным В.С.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Основы анализа данных

Направление подготовки: 09.03.01 Информатика и вычислительная техника

Направленность (профиль): IT-сервисы и технологии обработки данных на транспорте

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: 170737
Подписал: заместитель директора академии Паринов Денис Владимирович
Дата: 19.06.2024

1. Общие сведения о дисциплине (модуле).

Курс от простого к сложному погружает слушателей в теорию и практические аспекты работы с большими массивами информации. В рамках курса на простом и понятном языке будет рассказано как в современных условиях производится обобщение статистических данных и поиск скрытых закономерностей, а также как эти закономерности используются при принятии управленческих решений и повышении эффективности работы предприятия. Основная цель данной дисциплины – формирование у студентов основных теоретических знаний и практических навыков в области анализа данных, применении наиболее популярных методов статистического анализа и прогностической аналитики.

2. Планируемые результаты обучения по дисциплине (модулю).

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

ОПК-1 - Способен применять естественнонаучные и инженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности;

ОПК-3 - Способен решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности;

ОПК-6 - Способен разрабатывать бизнес-планы и технические задания на оснащение отделов, лабораторий, офисов компьютерным и сетевым оборудованием;

ОПК-8 - Способен разрабатывать алгоритмы и программы, пригодные для практического применения;

УК-1 - Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач.

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

Знать:

- стратегии управления данными;
- программные обеспечения, применяемые для анализа данных;
- языки программирования, специализированные на анализе данных
- ландшафт альтернативных решений и особенностей их применения;

- терминологию в области анализа данных: программное обеспечение в области анализа данных, методы анализа.

Уметь:

- Выполнять профилирование данных и правильно определить какие методы анализа должны быть использованы в конкретной ситуации;

- выбирать инструменты анализа данных в зависимости от их типов, объем;

- собирать данные из открытых источников.

Владеть:

- языками программирования, предназначенными для анализа и визуализации данных;

- инструментами инженерии данных;

- статистическими инструментами проверки гипотез о данных.

3. Объем дисциплины (модуля).

3.1. Общая трудоемкость дисциплины (модуля).

Общая трудоемкость дисциплины (модуля) составляет 4 з.е. (144 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 №1
Контактная работа при проведении учебных занятий (всего):	32	32
В том числе:		
Занятия семинарского типа	32	32

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 112 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.

Не предусмотрено учебным планом

4.2. Занятия семинарского типа.

Лабораторные работы

№ п/п	Наименование лабораторных работ / краткое содержание
1	<p>Тема 1. Введение в предметную область аналитики больших данных: хайп или нераскрытый потенциал</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - Данные, информация, знания – в чем отличия. - Что есть большие данные. Что есть аналитика. - Что за зверь такой: data-driven организация? - Ожидания рынка vs результаты. - Несколько основных ловушек и извлеченных из них опыт. - Пересмотр приоритетов за последние 10 лет. - Новые игроки на рынке и чем они характеризуются.
2	<p>Стратегия управления данными: искусство видеть лес за деревьями</p> <p>Стратегия управления данными, как составная часть культуры современной организации. Почему это важный вопрос? Что есть стратегия управления данными, кто за неё отвечает в организации. Кто такой владелец данных, его сфера ответственности и полномочия. Модель данных, что это такое и зачем она нужна. Основные заинтересованные стороны. Концепция Self-Service BI. Техническая инфраструктура. Революция open-source и доступность технологий. Как измерить эффективность стратегии управления данными.</p>
3	<p>Хранилища данных: первый шаг к аналитике и зачем все так усложнять</p> <p>Разберемся в предмете: что есть хранилище данных, зачем оно нужно. Что за зверь такой – ETL. Определение источников и загрузка данных. Какие виды преобразования и объединения данных существуют, в чём их принципиальные отличия и на что следует обращать внимание. Что такое витрины данных.</p>
4	<p>Постановка задачи: как не ошибиться с выбором цели и инструментов</p> <p>С чего начать, с какой стороны подойти к задаче анализа данных. Основные термины и понятия. Как правильно ставить задачу и формулировать цель анализа и почему это так важно. Как оценить достаточно ли вам данных для анализа, где и как данные следует добывать, на что обращать внимание. Понятие качества данных. Методики и инструменты обеспечения и контроля качества данных. Выработка гипотез и выбор методов.</p>
5	<p>Подготовка данных: искусство есть слона по частям</p> <p>Какие данные бывают, какие у них свойства, почему важно разобраться в их природе (рассмотрим на примере ошибок, допускаемых аналитиками при работе вслепую). Что такое метаданные или бизнес-гlossарий и почему обязательно сопровождать аналитическую задачу созданием соответствующего репозитория метаданных. Какие есть способы подготовки данных, на что следует обращать внимание, как не ошибиться.</p>

№ п/п	Наименование лабораторных работ / краткое содержание
6	<p>Введение в статистический анализ: разбираемся в основах Основные понятия: генеральная совокупность и выборка. Какие основные статистические показатели используются в аналитике и почему (рассмотрим на известных исторических примерах). Что такое сводная таблица и как она используется в работе (рассмотрим на примере задачи финансового планирования).</p>
7	<p>Обзор описательных статистик: как заставить данные говорить (часть 1) Какие существуют описательные статистики и для чего используются. Почему ими надо аккуратно пользоваться. Почему иногда полезно выполнять нормализацию данных. Какие инструменты нам в этом помогут, как правильно представить результатная статистического анализа.</p>
8	<p>Основы визуализации и story telling: как заставить данные говорить (часть 2) Декодирование результатов аналитической работы. Почему это важно (аналитика не игрушка, она должна быть подчинена реальным бизнес-целям и показателям). Как показывать положительные и отрицательные значения, как выводить данные по категориальным аналитикам, как показывать динамику и прогноз, как отображать выбросы и пропуски, как расставить акценты и выделить самое важное (рассмотрим на реальных кейсах нескольких компаний).</p>
9	<p>Аналитика: сравнение групп данных Для чего и где применяется. Понятие значимости различий при сравнении групп (выборок) данных. Зависимые и независимые выборки. Таблицы перекрестной классификации. Сравнение независимых выборок. Сравнение зависимых выборок. Анализ последовательностей.</p>
10	<p>Аналитика: анализ взаимосвязей, поиск скрытого от человеческого глаза Зачем нужно анализировать связи в данных (примеры задач). Зависимые и независимые переменные. Статистические взаимосвязи и их значимость для анализа. Корреляции. Регрессия и её виды. Факторный анализ. Анализ надежности согласованности. Понятие проклятия размерности.</p>
11	<p>Аналитика: классификация и что это даёт Зачем нужна классификация (примеры задач). Виды классификации. Обзор используемых методов: деревья решений, случайный лес, наивный байесовский классификатор, линейная и логистическая регрессия. Кластеризация (и чем принципиально отличается от классификации). О чем важно помнить при выполнении кластеризации.</p>
12	<p>Аналитика: прогнозирование, хотим управлять будущим Что есть прогнозирование, где и как используется. Временные ряды. Что такое сезонность и как её найти в данных. Методы прогнозирования и их сравнение: линейная регрессия, экспоненциальное сглаживание, фильтр Кальмана, модель сквозящего среднего, авторегрессионная модель, ARIMA.</p>
13	<p>Машинное обучение: где и как применяется Зачем нужно машинное обучение, что скрывается под этим названием. Методы машинного обучения для изучения и подготовки данных. Примеры задач и методов их решения. Обучение с подкреплением, верхняя доверительная граница, A/B тестирование, выборка Томпсона.</p>
14	<p>Машинное обучение и Искусственный интеллект: а есть ли разница Почему современные компании постоянно используют эти термины как будто они значат одно и то же? Есть ли разница на самом деле? Задачи применения искусственного интеллекта, как более широкий класс задач. Понятие нейронных сетей. Проблема интерпретируемости. Примеры применения решений, основанных на искусственном интеллекте, которые окружают нас уже сегодня. Будущее ИИ и ML.</p>
15	<p>Регулирование данных: в любой сфере должны быть свои законы Актуальные вопросы регулирования данных, стоящие на повестке дня в мире и России, в частности. Какие существуют нормативы на сегодня, как контролируется факт их выполнения. Этические проблемы работы с данными и применения машинного обучения / Искусственного интеллекта (рассмотрим на примере нескольких реальных кейсов – как используется и с какими последствиями приходится иметь дело). Формирование этической культуры работы с данными в организации, на что обратить внимание.</p>

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Работа с интернет-источниками и литературой
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 интеллектуального анализа данных. Лабораторный практикум: учебное пособие Нестеров С.А. Издательство "Лань" , 2020	https://reader.lanbook.com/book/130181#1
1	Прогнозное моделирование в IBM SPSS Statistics, R и Python: метод деревьев решений и случайный лес Груздев А.В. Издательство "ДМК Пресс" , 2018	https://reader.lanbook.com/book/123700#484

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

Научно-техническая библиотека Российского университета транспорт:
<http://library.miit.ru/>

Научная электронная библиотека eLibrary.ru. <http://elibrary.ru>

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

e-mail преподавателей для оперативной связи:
s.kulikovskiy@polymatica.com

презентации в среде PowerPoint по теме дисциплины
список сайтов в среде Интернет для поиска научно-технической информации по разделам дисциплины
офисный пакет приложений – Microsoft Office
ПО для анализа данных Polymatica

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

Компьютер преподавателя

Intel Core i7-9700 / Asus PRIME H310M-R R2.0 / 2x8GB / SSD 250Gb / DVDRW

Компьютеры студентов

Intel Core i9-9900 / B365M Pro4 / 2x16GB / SSD 512Gb /

экран для проектора, маркерная доска,

Проектор Optoma W340UST,

9. Форма промежуточной аттестации:

Зачет в 1 семестре.

10. Оценочные материалы.

Оценочные материалы, применяемые при проведении промежуточной аттестации, разрабатываются в соответствии с локальным нормативным актом РУТ (МИИТ).

Авторы:

доцент, к.н. Академии "Высшая
инженерная школа"

Н.В. Атамасов

Согласовано:

Заместитель директора академии

Д.В. Паринов

Председатель учебно-методической
комиссии

Д.В. Паринов