

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: 937226
Подписал: руководитель образовательной программы
Проневич Ольга Борисовна
Дата: 10.10.2024

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

ОПК-9 - Способен осваивать методики использования программных средств для решения практических задач;

ПК-1 - Способен анализировать большие данные с использованием существующей в организации методологической и технологической инфраструктуры.

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

Владеть:

языками программирования, предназначенными для анализа и визуализации данных,

инструментами инженерии данных,

статистическими инструментами проверки гипотез о данных.

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.

Знать:

Стратегии управления данными,
программное обеспечения, применяемые для анализа данных,
языки программирования, специализированные на анализе данных
Знание ландшафта альтернативных решений и особенностей их
применения.

терминология в области анализа данных: программное обеспечение в
области анализа данных, методы анализа.

3. Объем дисциплины (модуля).

3.1. Общая трудоемкость дисциплины (модуля).

Общая трудоемкость дисциплины (модуля) составляет 3 з.е. (108 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 76 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.

Практические занятия

№ п/п	Тематика практических занятий/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>Тема 2. Стратегия управления данными: искусство видеть лес за деревьями</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none">- 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.
3	<p>Тема 3. Хранилища данных: первый шаг к аналитике и зачем все так усложнять</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none">- 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.
4	<p>Тема 4. Постановка задачи: как не ошибиться с выбором цели и инструментов</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none">- С чего начать, с какой стороны подойти к задаче анализа данных.- Основные термины и понятия.- Как правильно ставить задачу и сформулировать 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.
5	<p>Тема 5. Подготовка данных: искусство есть слона по частям</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none">- Какие данные бывают, какие у них свойства, почему важно разобраться в их природе (рассмотрим на примере ошибок, допускаемых аналитиками при работе вслепую).

№ п/п	Тематика практических занятий/краткое содержание
	<ul style="list-style-type: none"> - 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.
6	<p>Тема 6. Введение в статистический анализ: разбираемся в основах</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - 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).
7	<p>Тема 7. Обзор описательных статистик: как заставить данные говорить (часть 1)</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - 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 результаты статистического анализа.
8	<p>Тема 8. Основы визуализации и story telling: как заставить данные говорить (часть 2)</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - 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).
9	<p>Тема 9. Аналитика: сравнение групп данных</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - 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.
10	<p>Тема 10. Аналитика: анализ взаимосвязей, поиск скрытого от человеческого глаза</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - 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.
11	<p>Тема 11. Аналитика: классификация и что это даёт</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - Зачем нужна классификация (примеры задач). - Виды классификации.

№ п/п	Тематика практических занятий/краткое содержание
	<ul style="list-style-type: none"> - 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.
12	<p>Тема 12. Аналитика: прогнозирование, хотим управлять будущим</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - 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 скользящего среднего, авторегрессионная модель, ARIMA.
13	<p>Тема 13. Машинное обучение: где и как применяется</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - 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. - A/B тестирование.
14	<p>Тема 14. Машинное обучение и Искусственный интеллект: а есть ли разница</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - 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.
15	<p>Тема 15. Регулирование данных: в любой сфере должны быть свои законы</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - 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.
16	<p>Тема 16. Яндекс.Даталенс</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - Анализ открытых данных ДТП на дорогах России - Построение интерактивного дашборда - Оценка полученного результата
17	<p>Тема 17. Построение визуализации и подготовка данных</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - Решение практических задач по подготовке данных - Анализ качества - Расчетные показатели - Визуализация данных
18	<p>Тема 18. Работа с открытыми данными</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - Загрузка данных

№ п/п	Тематика практических занятий/краткое содержание
	- Анализ данных - Представление результата

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Работа с интернет-источниками и литературой
2	Настройка программного обеспечения
3	Изучение технической документации на программное обеспечение в области анализа данных и отслеживание обновления библиотек
4	Подготовка к промежуточной аттестации.
5	Подготовка к текущему контролю.

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

№ п/п	Библиографическое описание	Место доступа
1	Груздев, А. В. Прогнозное моделирование в IBM SPSS Statistics, R и Python: метод деревьев решений и случайный лес : руководство / А. В. Груздев. — Москва : ДМК Пресс, 2018. — 642 с. — ISBN 978-5-97060-539-4	https://e.lanbook.com/book/123700
2	Нестеров, С. А. Основы интеллектуального анализа данных. Лабораторный практикум : учебное пособие / С. А. Нестеров. — Санкт-Петербург : Лань, 2020. — 40 с. — ISBN 978-5-8114-4509-7	https://e.lanbook.com/book/130181

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

<https://habr.com/ru> - база знаний в виде статей, обзоров

<https://journal.tinkoff.ru/short/ai-for-all/> - база данных нейронных сетей

<https://vc.ru/services/916617-luchshie-neyroseti-bolshaya-podborka-iz-top-200-ii-generatorov-po-kategoriyam> - база данных нейронных сетей

<https://github.com/abalmumcu/bert-rest-api> - профессиональная платформа для командой работы над проектов (нейронная сеть bert)

<http://library.miit.ru/> - электронно-библиотечная система Научно-технической библиотеки МИИТ

<https://proglib.io/p/raspoznavanie-obektov-s-pomoshchyu-yolo-v3-na-tensorflow-2-0-2020-11-08> - профессиональная библиотека программистов

https://yandex.cloud/ru/blog/posts/2022/12/andrey-berger-and-yandex-cloud?utm_referrer=https%3A%2F%2Fyandex.ru%2F – библиотека профессиональных статей разработчиков Яндекс

<https://yandex.cloud/ru/blog> - библиотека профессиональных статей разработчиков Яндекс

<https://tproger.ru/translations/opencv-python-guide> - библиотека основных команд OpenCV

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

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
Компьютеры студентов
экран для проектора, маркерная доска,
Проектор

9. Форма промежуточной аттестации:

Экзамен в 1 семестре.

10. Оценочные материалы.

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

Авторы:

руководитель образовательной
программы

Согласовано:

Директор

Руководитель образовательной
программы

Председатель учебно-методической
комиссии

Н.В. Атамасов

О.Б. Проневич

Б.В. Игольников

О.Б. Проневич

Д.В. Паринов