

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Институт прикладной математики и компьютерных наук

УТВЕРЖДАЮ:

Директор



А. В. Замятин

« 16 » мая 20 22 г.

Рабочая программа дисциплины

Визуализация данных

по направлению подготовки

01.04.02 Прикладная математика и информатика

Направленность (профиль) подготовки :

Big Data and Data Science

Форма обучения

Очная

Квалификация

Магистр

Год приема

2022

Код дисциплины в учебном плане: Б1.О.09

СОГЛАСОВАНО:

Руководитель ОП

А.В. Замятин

Председатель УМК

С.П. Сущенко

Томск – 2022

1. Цель и планируемые результаты освоения дисциплины

Целью освоения дисциплины является формирование следующих компетенций:

– ПК-1 – способность разрабатывать и применять математические методы, алгоритмы, программное обеспечение для решения задач научно-исследовательской и проектной деятельности.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИПК-1.3 Разрабатывает новые методы, модели, алгоритмы и программное обеспечение для решения задач в области профессиональной деятельности.

2. Задачи освоения дисциплины

– сформировать навыки решения прикладных профессиональных задач с использованием методов и средств визуализации данных и знаний;

– научить анализировать задачи профессиональной деятельности средствами информационных технологий;

– научить интерпретировать результаты визуализации и грамотно их описывать;

– сформировать навыки использования библиотек современных скриптовых языков (Python, R) для визуализации данных и знаний.

3. Место дисциплины в структуре образовательной программы

Дисциплина относится к обязательной части образовательной программы.

4. Семестр(ы) освоения и форма(ы) промежуточной аттестации по дисциплине

Второй семестр, зачет.

5. Входные требования для освоения дисциплины

Для успешного освоения дисциплины требуются результаты обучения по следующим дисциплинам: «Введение в интеллектуальный анализ данных».

6. Язык реализации

Английский.

7. Объем дисциплины

Общая трудоемкость дисциплины составляет 3 з.е., 108 часов, из которых:

-лекции: 10 ч.

-лабораторные: 20 ч.

в том числе практическая подготовка: 0 ч.

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины, структурированное по темам

Тема 1. Визуализация многомерных данных. Задачи визуализации. Способы визуализации

Цель визуализации, задачи визуализации многомерных данных, классификация по визуализируемым объектам. Выбор правильного типа визуализации. Первичный анализ данных с использованием методов визуализации. Библиотеки Python (или R) для решений задач визуализации. Сравнение полученных визуальных образов.

Тема 2. Python-библиотеки для визуализации данных в Data Science

Возможности библиотек Matplotlib, Seaborn, Missingno, Altair, Plotly, Bokeh, Pygal, Networkx. Примеры использования перечисленных библиотек. Визуализация многомерных данных с использованием диаграмм Эндрюса.

Тема 3. Визуализация данных средствами дашбордов.

Что из себя представляет дашборд, его свойства, отличие от отчета. Модули дашборда. Виды дашбордов. Основные инструменты: Google Sheets, Яндекс.Метрика, Google Analytics, Google Data Studio, Qlik, Power Bi, Owox Bi.

Тема 4. Методы визуализации для решения прикладных задач.

Методы визуализации для задач классификации, кластеризации. Визуализация решения задачи временных рядов, демонстрация примеров.

9. Текущий контроль по дисциплине

Текущий контроль по дисциплине проводится путем контроля посещаемости, выполнения лабораторных работ, выполнения индивидуального проекта, и фиксируется в форме контрольной точки не менее одного раза в семестр.

10. Порядок проведения и критерии оценивания промежуточной аттестации

Освоение дисциплины проходит в классической реализации по схеме «Лекция – лабораторные занятия». На лекции, кроме изложения теоретического материала, преподаватель использует активные формы обучения – взаимодействие с аудиторией: вопрос-ответ, учебная дискуссия, работа в группах, проблемное обучение.

Промежуточная аттестация реализуется в форме зачета. Оценка за зачет выставляется на основе представления и защиты индивидуального проекта. Студент выполняет презентацию, а также демонстрирует программный код. Вопросы по результатам могут задавать все студенты группы, не только преподаватель.

11. Учебно-методическое обеспечение

а) Электронный учебный курс по дисциплине в электронном университете «Moodle» - <https://moodle.tsu.ru/course/view.php?id=5464>

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине (Приложение 1).

в) План семинарских / практических занятий по дисциплине.

Не предусмотрены

г) Методические указания по проведению лабораторных работ.

Выполнение лабораторной работы предусмотрено по каждой изучаемой теме дисциплины. Методические пособия размещены в локальной сети ИПМКН в каталоге X:\Workspace\Магистратура\ Представление знаний и визуализация данных.

д) Методические указания по организации самостоятельной работы студентов.

Для укрепления изученного материала предусмотрено выполнение индивидуального проекта в рамках часов самостоятельной работы. Проект может быть выполнен как индивидуально, так и в мини-группе (2-3 чел.), при условии, что объем работы также будет увеличен. В конце семестра по каждому проекту представляется мини-презентация о результатах работы.

12. Перечень учебной литературы и ресурсов сети Интернет

Учебная и учебно-методическая литература, учебно-методические и другие материалы, необходимые для изучения дисциплины:

– Эрик Мэттиз. Изучаем Python: программирование игр, визуализация данных, веб-приложения. – СПб.: Питер, 2021. – 587 с.

- Маккинни, У. Python и анализ данных. – М.: ДМК Пресс, 2020. – 540 с.
- Мастицкий, С.Э., Шитиков В.К. Статистический анализ и визуализация данных с помощью R. – ДМК Пресс, 2015. – 496 с.
- Мастицкий, С.Э. Визуализация данных с помощью ggplot2. – М.: ДМК Пресс, 2017. – 222 с.

Ресурсы сети Интернет, необходимые для изучения дисциплины:

Наименование	Ссылка на ресурс	Доступность (свободный доступ/ ограниченный доступ)
<i>1</i>	<i>2</i>	<i>3</i>
Информационно-справочные системы		
Представление знаний и визуализация данных	https://moodle.ido.tsu.ru/course/view.php?id=1403	Свободный доступ
7 самых популярных Python-библиотек для визуализации данных в Data Science	https://medium.com/@bigdataschool/7-самых-популярных-python-библиотек-для-визуализации-данных-в-data-science-c0a4875df93b	Свободный доступ
Обзор пакетов для визуализации данных на Python	https://medium.com/@alibekashirali/обзор-пакетов-для-визуализации-данных-на-python-a1525b2a5cc7	Свободный доступ
Каталог Визуализации данных	https://datavizcatalogue.com/	Свободный доступ
Зачем и как использовать визуализацию данных?	https://infogra.ru/infographics/zachem-i-kak-ispolzovat-vizualizatsiyu-dannyh	Свободный доступ
Открытый курс машинного обучения	https://habr.com/ru/company/ods/blog/323210/	Свободный доступ
Обзор методов визуализации многомерных данных	https://cyberleninka.ru/article/n/obzor-metodov-vizualizatsii-mnogomernyh-dannyh	Свободный доступ
Как создать дашборд, используя Google	https://blog.sheetgo.com/spreadsheets-tips/how-to-create-a-dashboard-in-google-sheets/	Свободный доступ

Sheets		
Электронно-библиотечные системы		
Научная библиотека ТГУ	https://www.lib.tsu.ru/	Свободный доступ
Электронно-библиотечная система «Лань»	https://e.lanbook.com/	Для авторизованных пользователей
КиберЛенинка	https://cyberleninka.ru/	Свободный доступ
Профессиональные базы данных		
Искусственный интеллект и сферы его применения. Новости разработки квантовых компьютеров. Исследования искусственных нейронных сетей.	https://ai-news.ru	Свободный доступ
Портал мировых соревнований команд по DM	https://www.kaggle.com/	Свободный доступ
Репозиторий по машинному обучению	https://archive.ics.uci.edu/ml/index.php	Свободный доступ

13. Перечень информационных технологий

а) лицензионное и свободно распространяемое программное обеспечение:

– Microsoft Office Standart 2013 Russian: пакет программ. Включает приложения: MS Office Word, MS Office Excel, MS Office PowerPoint, MS Office On-eNote, MS Office Publisher, MS Outlook, MS Office Web Apps (Word Excel MS PowerPoint Outlook);

– публично доступные облачные технологии (Google Docs, Яндекс диск и т.п.).

– средства программирования на Python3 и R:

- Python3
- RStudio.

б) информационные справочные системы:

– Электронный каталог Научной библиотеки ТГУ – <http://chamo.lib.tsu.ru/search/query?locale=ru&theme=system>

– Электронная библиотека (репозиторий) ТГУ – <http://vital.lib.tsu.ru/vital/access/manager/Index>

– ЭБС Лань – <http://e.lanbook.com/>

– ЭБС Консультант студента – <http://www.studentlibrary.ru/>

– Образовательная платформа Юрайт – <https://urait.ru/>

– ЭБС ZNANIUM.com – <https://znanium.com/>

– ЭБС IPRbooks – <http://www.iprbookshop.ru/>

14. Материально-техническое обеспечение

Аудитории для проведения занятий лекционного типа.

Аудитории для проведения лабораторных занятий, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации.

Помещения для самостоятельной работы, оснащенные компьютерной техникой и доступом к сети Интернет, в электронную информационно-образовательную среду и к информационным справочным системам.

Для материально-технического обеспечения дисциплины требуется наличие компьютерной техники с установленным соответствующим программным обеспечением и другого оборудования, поддерживающего проведение презентаций и выходом в сеть Интернет.

15. Информация о разработчиках

Марухина Ольга Владимировна, канд. техн. наук, доцент, доцент кафедры теоретических основ информатики ТГУ.