

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

Институт прикладной математики и компьютерных наук

УТВЕРЖДЕНО:

Директор

А. В. Замятин

Рабочая программа дисциплины

Анализ данных и социальные медиа

по направлению подготовки

01.04.02 Прикладная математика и информатика

Направленность (профиль) подготовки:

Интеллектуальный анализ больших данных

Форма обучения

Очная

Квалификация

Магистр

Год приема

2024

СОГЛАСОВАНО:

Руководитель ОП

А.В. Замятин

Председатель УМК

С.П. Сущенко

1. Цель и планируемые результаты освоения дисциплины

Целью освоения дисциплины является формирование следующих компетенций:
ПК-5 – способность определять проблемную ситуацию, ставить задачи анализа данных в общественных науках, подбирать математический и аппаратный инструментарий для их решения.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИПК-5.3 Анализирует полученную информацию и находит скрытые закономерности.

ИПК-5.2 Собирает информацию из сети Интернет, в т.ч. из социальных сетей, и обрабатывает ее.

ИПК-5.1 Дает оценку поведения общества или его отдельных групп на основе анализа данных.

2. Задачи освоения дисциплины

- Познакомиться со статистическими моделями и методами, применяемыми при анализе данных и социальных медиа.
- Научиться выявлять и строить математические модели процессов, оценивать их качество и интерпретировать результаты их анализа.
- Научиться использовать программные средства, используемые при анализе данных.

3. Место дисциплины в структуре образовательной программы

Дисциплина относится к Блоку 1 «Дисциплины (модули)».

Дисциплина относится к обязательной части образовательной программы. Дисциплина входит в модуль «Специализация».

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

Второй семестр, экзамен.

5. Входные требования для освоения дисциплины

Для освоения данной дисциплины необходимы знания математического анализа, линейной алгебры, методов оптимизаций, теории вероятностей и математической статистики, основ теории графов и основ программирования. Также требуются результаты обучения по следующим дисциплинам: «Введение в интеллектуальный анализ данных».

6. Язык реализации

Русский

7. Объем дисциплины

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

-лекции: 16 ч.

-практические занятия: 32 ч.

в том числе практическая подготовка: 0 ч.

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины, структурированное по темам Раздел 1. Введение

1.1. Задачи и этапы статистического анализа

1 Раздел 2. Анализ данных

2.1. Типы и структуры данных

2.2. Статистические методы анализа данных

2.3. Методы и модели машинного обучения. Обучение с учителем: регрессия, классификация. Обучение без учителя: кластерный анализ, снижение размерности.

2 Раздел 3. Социальные медиа

3.1. Введение в Social Network Analysis (SNA)

3.2. Элементы теории графов

3.3. Основные термины и понятия SNA

3.4. Случайные сети

3.5. Распределения и метрики

9. Текущий контроль по дисциплине

Текущий контроль по дисциплине осуществляется на основании проверки лабораторных работ, выполняемых студентами на компьютерах в течение семестра.

Студенты получают у преподавателя или собирают самостоятельно данные для дальнейшего анализа и построения математических моделей. Проводят предварительную обработку данных, выбирают адекватный метод анализа в зависимости от целей исследования и типов данных, реализуют выбранные методы на компьютере, делают выводы и интерпретацию полученных результатов.

10. Порядок проведения и критерии оценивания промежуточной аттестации

К зачету студенты должны выполнить все лабораторные работы и выступить на семинаре на любую, выбранную ими тему по анализу социальных медиа из разделов:

1. Введение в Social Network Analysis (SNA);
2. Элементы теории графов;
3. Основные термины и понятия SNA;
4. Случайные сети;
5. Распределения и метрики.

11. Учебно-методическое обеспечение

а) Электронный учебный курс по дисциплине в электронном университете «Moodle»

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

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

- Джеймс Г., Уиттон Д., Хасты Е., Тибширани Р. Введение в статистическое обучение с примерами на языке R. – М.: ДМК Пресс, 2016. – 450 с.
- Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение. – М.: ДМК Пресс, 2018. – 652 с.

б) дополнительная литература:

- Шолле Франсуа. Глубокое обучение на Python. – СПб.: Питер, 2018. – 400 с.
- Шолле Франсуа. Глубокое обучение на R. – СПб.: Питер, 2018. – 400 с.

в) ресурсы сети Интернет:

- Alberto-Laszlo Barabasi. Network Science. URL: <http://networksciencebook.com/>

13. Перечень информационных технологий

а) лицензионное и свободно распространяемое программное обеспечение: – MS Windows, MS Office, R, R Studio, Python

б) информационные справочные системы:

- <http://statsoft.ru/#tab-STATISTICA-link>
- <https://www.r-project.org/>
- [h t t p s : / / r s t u d i o - p u b s - static.s3.amazonaws.com/33285_9ffa5341065d4bff9607fd6ccba75d98.html#-](https://r-studio-pubs-static.s3.amazonaws.com/33285_9ffa5341065d4bff9607fd6ccba75d98.html#-)
- <https://pythonworld.ru/samouchitel-python>
- <https://archive.ics.uci.edu/ml/index.php>
- <https://www.kaggle.com/>

14. Материально-техническое обеспечение

Для реализации дисциплины необходимы лекционные аудитории и аудитории для проведения практических занятий. Специальные технические средства (проектор, компьютер и т.д.) требуются для демонстрации материала в рамках изучаемых разделов, проведения защиты проектов в конце семестра.

15. Информация о разработчиках

Кабанова Татьяна Валерьевна, канд. физ.-мат. наук, доцент кафедры теории вероятностей и математической статистики ИПМКН ТГУ.