

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

Философский факультет

УТВЕРЖДАЮ:

Декан Философского факультета



Е.В. Сухушина

«04» июля 2023 г.

Рабочая программа дисциплины

Интеллектуальный анализ больших данных

по направлению подготовки

39.04.01 Социология

Направленность (профиль) подготовки:

«Социология управления»

Форма обучения

Очная

Квалификация

Магистр

Год приема

2023

Код дисциплины в учебном плане: Б1.О.03

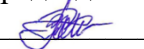
СОГЛАСОВАНО:

Руководитель ОП



А.Ю. Рыкун

Председатель УМК



Т.В. Фаненштиль

Томск – 2023

1. Цель и планируемые результаты освоения дисциплины (модуля)

Целью освоения дисциплины является формирование следующих компетенций:

УК-1 – Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий

ОПК-1 – Способен обоснованно отбирать и использовать современные информационно-коммуникационные технологии для решения профессиональных задач

ОПК-2 – Способен проводить фундаментальные и прикладные социологические исследования и представлять их результаты

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИУК-1.1. Выявляет проблемную ситуацию, на основе системного подхода осуществляет её многофакторный анализ и диагностику

ИОПК-1.1. Определяет и обосновывает релевантные для решения поставленной задачи источники информации, включая национальные и международные базы данных, электронные библиотечные системы, специализированные пакеты прикладных программ

ИОПК-1.2. Проводит поиск социологической информации, необходимой для решения поставленной задачи, получает на ее основе социологические данные и обосновывает их релевантность

ИОПК-1.3. Использует современные сетевые технологии и пакеты прикладных программ для организации и обеспечения профессиональной коммуникации

ИОПК-2.3. Анализирует количественные и качественные данные фундаментальных и прикладных социологических исследований и интерпретирует полученные результаты адекватно поставленным исследовательским задачам

ИОПК-2.4. Составляет аналитические отчеты, презентации, аналитические записки по итогам социологического исследования и представляет полученные результаты разным целевым аудиториям

2. Задачи освоения дисциплины

– Научиться выявлять задачи интеллектуального анализа в социологических данных и применять различные виды интеллектуального анализа данных для исследования практических проблем социологии;

– Освоить концептуальный и понятийный аппарат интеллектуального анализа данных для решения практических задач профессиональной деятельности.

– Понимать возможности и ограничения существующих методов интеллектуального анализа данных;

– Научиться использовать современные инструменты интеллектуального анализа данных для решения практических задач в профессиональной деятельности.

3. Место дисциплины (модуля) в структуре образовательной программы

Дисциплина относится к блоку Б.1.О, обязательная часть образовательной программы

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

Семестр 1, экзамен.

5. Входные требования для освоения дисциплины

Для успешного освоения дисциплины требуются компетенции, сформированные в ходе освоения образовательных программ предшествующего уровня образования.

6. Язык реализации

Русский

7. Объем дисциплины (модуля)

Общая трудоемкость дисциплины составляет 6 з.е., 216 часов, из которых:

– лекции: 20 ч.;

– практические занятия: 32 ч.;

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины (модуля), структурированное по темам

Тема 1. Введение в интеллектуальный анализ данных

Data driven подход. Сферы применения интеллектуального анализа данных (прикладные задачи). Терминология. Методологии DataMining. Общий план анализа данных. Обзор инструментов интеллектуального анализа данных.

Тема 2. Данные: типы, источники, требования и разметка данных

Данные и информация. Типы данных. Типология данных с точки зрения их получения. Способы получения данных. Источники данных. Открытые данные. Методология выбора данных согласно цели исследования. Требования к данным. Валидация данных. Способы формирования маркерных слов. Этапы разметки данных. Типичные задачи разметки данных. Методологии разметки. Практическое применение разметки данных. Инструменты разметки данных. Валидация разметки. Коэффициента межаннотарского согласия.

Тема 3. Разведочный анализ данных

Типы шкал. Этапы разведочного анализа данных. Оценки центрального положения. Оценки вариабельности. Распределение данных. Корреляционный анализ. Визуализация.

Тема 4. Предварительная обработка данных

Этапы предварительной обработки данных. Очистка данных. Восстановление пропущенных значений. Нормализация данных. Кодировка категориальных признаков. Отбор информативных признаков. Уменьшение размерности. Обнаружение выбросов. Обработка дубликатов.

Тема 5. Машинное обучение: введение

Введение в область машинного обучения. Искусственный интеллект: машинное обучение, нейронные сети, глубокое обучение. Задачи машинного обучения (виды). Опасности искусственного интеллекта. Требования для машинного обучения: данные, признаки, алгоритмы.

Тема 6. Машинное обучение с учителем: классификация

Машинное обучение с учителем. Задача классификации. Виды выборок для обучения. Кросс-валидация. Метрики оценки. Алгоритмы классификации: дерево решений, случайный лес, метод опорных векторов, ансамблевые методы.

Тема 7. Машинное обучение с учителем: регрессия

Задачи регрессии. Метрики оценки. Линерализация. Алгоритмы регрессии: линейная регрессия, полиномиальная регрессия, случайный лес для задач регрессии.

Тема 8. Машинное обучение без учителя. Кластеризация

Кластеризация. Задачи кластеризации. Этапы кластерного анализа. Меры расстояний. Методы кластеризации: иерархические, итерационные, плотностные методы.

Тема 9. Обработка естественного языка (NLP)

Обработка естественного языка. Задачи обработки естественного языка. Сложности анализа. Терминология. Этапы обработки естественного языка: предварительная обработка, векторизация, методы, оценки.

9. Текущий контроль по дисциплине

Текущий контроль по дисциплине проводится путем контроля посещаемости и выполнения домашних заданий и фиксируется в форме контрольной точки не менее одного раза в семестр.

Критерии оценивания для текущей аттестации и типовые задания представлены в фонде оценочных материалов. В полном объеме фонд оценочных материалов хранится на кафедре социологии.

10. Порядок проведения промежуточной аттестации

Экзамен в первом семестре предполагает учет следующих форм оценивания:

1. Выполнение самостоятельных практических работ
2. Устный ответ на 2 вопроса экзаменационного билета.

Критерии оценивания для промежуточной аттестации, а также типовые задания представлены в Фонде оценочных материалов. В полном объеме фонд оценочных материалов хранится на кафедре социологии.

11. Учебно-методическое обеспечение

а) Электронный учебный курс по дисциплине в электронном университете «Moodle» - <https://moodle.tsu.ru/course/view.php?id=20721>

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине представлены в Фонде оценочных материалов.

в) План семинарских / практических занятий по дисциплине представлен в Фонде оценочных материалов.

г) Методические указания по организации самостоятельной работы студентов размещены в ЭОИС НИ ТГУ.

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

1. – Маккинни, У. Python и анализ данных / У. Маккинни ; перевод с английского А. А. Слинкина. — 2-ое изд., испр. и доп. — Москва : ДМК Пресс, 2020. — 540 с. — ISBN 978-5-97060-590-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/131721>

2. – Дятлов, А.В. Анализ данных в социологии : учебник / А.В.Дятлов, Д.А.Гугуева ; Южный федеральный университет. - Ростов-на-Дону ; Таганрог : Издательство Южного федерального университета, 2018. - 226 с. - ISBN 978-5-9275-2690-1. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1039664>

3. – Юре, Л. Анализ больших наборов данных / Л. Юре, Р. Ананд, Д. У. Джеффри ; перевод с английского А. А. Слинкин. — Москва : ДМК Пресс, 2016. — 498 с. — ISBN 978-5-97060-190-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/93571>

4. – Федорова О. П. Анализ и визуализация данных : учебно-методический комплекс / О. П. Федорова ; Том. гос. ун-т, [Ин-т дистанционного образования]. - Томск : [ИДО ТГУ], 2010. - . URL: <http://vital.lib.tsu.ru/vital/access/manager/Repository/vtls:000416178>

5. – Миркин Б. Г. Введение в анализ данных : учебник и практикум для бакалавриата и магистратуры : [для студентов вузов, обучающихся по инженерно-техническим, естественно-научным и экономическим направлениям и специальностям] / Б. Г. Миркин ; "Высшая школа экономики" Национальный исследовательский университет. - Москва : Юрайт, 2015. - 173, [1] с.: ил. - (Авторский учебник)

6. – Замятин А. В. Интеллектуальный анализ данных : учебное пособие : [для студентов университетов и вузов] / А. В. Замятин ; Нац. исслед. Том. гос. ун-т. - Томск : Изда-

тельский Дом Томского государственного университета, 2020. - 193 с.: ил., табл.. URL: <http://vital.lib.tsu.ru/vital/access/manager/Repository/vtls:000722107>

б) дополнительная литература:

1. – Аналитическая культура. От сбора данных до бизнес-результатов / Карл Андерсон; пер. с англ. Юлии Константиновой; [науч. ред. Руслан Салахийев]. — М.: Манн, Иванов и Фербер, 2017.
2. – Нолис Жаклин, Робинсон Эмили Data Science для карьериста. — СПб.: Питер, 2021. — 368 с.: ил. — (Серия «Библиотека программиста»).
3. – Брюс, П. Практическая статистика для специалистов Data Science: Пер. с англ. / П. Брюс, Э. Брюс. — СПб.: БХВ-Петербург, 2018. — 304 с.: ил.
4. – Еременко К. Работа с данными в любой сфере: Как выйти на новый уровень, используя аналитику / Кирилл Еременко ; Пер. с англ. — М. : Альпина Паблицер, 2019. — 303 с.
5. – Юре Лесковец, Ананд Раджараман, Джеффри Д. Ульман Анализ больших наборов данных. / Пер. с англ. Слинкин А. А. – М.: ДМК Пресс, 2016. – 498 с.: ил
6. – Монарх (Манро) Р. Машинное обучение с участием человека / пер. с англ. В. И. Бахура. – М.: ДМК Пресс, 2022. – 498 с.: ил.
7. – Ын Анналин, Су Кеннет Теоретический минимум по Big Data. Всё, что нужно знать о больших данных. — СПб.: Питер, 2019. — 208 с.: ил. — (Серия «Библиотека программиста»).
8. – Базалева, Ольга Игоревна. Мастерство визуализации данных.- СПб.: 000 "Диалектика", 2018.-192 с.: ил.
9. – Хобсон Лейн, Ханнес Хапке, Коул Ховард Обработка естественного языка в действии. — СПб.: Питер, 2020. — 576 с.: ил. — (Серия «Для профессионалов»)
10. – Автоматическая обработка текстов на естественном языке и анализ данных : учеб. пособие / Большакова Е.И., Воронцов К.В., Ефремова Н.Э., Клышинский Э.С., Лукашевич Н.В., Сапин А.С. — М.: Изд-во НИУ ВШЭ, 2017. — 269 с.

в) ресурсы сети Интернет:

- Orange Data Mining <https://orangedatamining.com/>
- Хабр <https://habr.com/ru/all/>

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.).

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 занятий семинарского типа, 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.

Аудитории для проведения занятий лекционного и семинарского типа индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации в смешанном формате («Актру»).

15. Информация о разработчиках

Басина Полина Александровна, Центр прикладного анализа больших данных, аналитик