

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель освоения дисциплины «Большие данные и методы машинного обучения в исследованиях» состоит в формировании у аспирантов:

- знаний наиболее актуальных работ в области применения новых типов данных в разных направлениях науки;
- навыков по сбору данных из социальных медиа и других цифровых следов с использованием языка программирования Python;
- навыков обработки и анализа различных типов данных (сетевые, текстовые и геоданные) с использованием языка программирования Python.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ПРОГРАММЫ АСПИРАНТУРЫ

Дисциплина 2.1.3.2. «Большие данные и методы машинного обучения в исследованиях» входит в Блок 2. «Образовательный компонент» и является одной из элективных дисциплин блока Дисциплины (модули) по выбору 3 (ДВ.3). Преподаётся на 2 курсе, во 4 семестре носит комплексный характер. Знания, полученные аспирантами в результате освоения дисциплины «Большие данные и методы машинного обучения в исследованиях», связаны с такими дисциплинами как «Методологии научного исследования», что позволяет аспирантам приобрести компетенции грамотно использовать алгоритмы машинного обучения и статистического анализа для изучения больших данных, развивать творческие способности в научно-познавательной деятельности.

3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

В результате изучения дисциплины аспирант должен:

Знать:

- основные теоретические, методологические и практические подходы к анализу больших данных и новых типов данных;
- источники новых типов данных;
- ключевые исследовательские работы и направления в области применения больших данных и методов машинного обучения в естественнонаучных, медико-биологических и общественно-гуманитарных науках;
- основные этические принципы работы с данными и этические проблемы, связанные с использованием больших данных;

Уметь:

- ставить исследовательские вопросы и формулировать гипотезы, протестировать которые можно с использованием больших данных;
- грамотно использовать алгоритмы машинного обучения и статистического анализа для изучения больших данных;
- проводить исследование полного цикла с использованием новых типов данных;
- интерпретировать и оформлять полученные результаты.

Иметь навыки:

- Постановки исследовательского вопроса в области применения больших данных, данных нового типа и методов машинного обучения в исследованиях образования;
- планирование исследования полного цикла;
- презентации и защиты индивидуального исследовательского проекта.

4. ОБЪЕМ ДИСЦИПЛИНЫ

4.1. Общая трудоемкость дисциплины составляет 3 зачетных единиц (108 часов)

4.2. По видам учебной работы (в часах)

Вид учебной работы	Количество часов (форма обучения - очная)	
	Всего по плану	В т.ч. по семестрам
		2
1	2	3
Лекции	16	16
Практические и семинарские занятия	16	16
Самостоятельная работа	76	76
Текущий контроль (количество и вид: контрольная работа, коллоквиум, реферат)	Опрос, домашнее задание	Опрос, домашнее задание
Виды промежуточной аттестации (экзамен, зачет)	зачет	зачет
Всего часов по дисциплине	108	108

4.3. Содержание дисциплины (модуля.) Распределение часов по темам и видам учебной работы:

Название и разделов, и тем	Всего	Виды учебных занятий		
		Аудиторные занятия		Самостоятельная работа
		лекции	практические занятия, семинар	
1	2	3	4	5
Тема 1. Большие данные и методы машинного обучения в естественнонаучных, медико-биологических и общественно-гуманитарных науках	13	2	2	9
Тема 2. Введение в язык программирования python.	13	2	2	9
Тема 3. Автоматический сбор данных из интернета	13	2	2	9
Тема 4. Использование методов машинного обучения для предсказания характеристик пользователей на основании их цифровых следов	13	2	2	9
Тема 5. Интеллектуальный анализ текстов	14	2	2	10

Тема 6. Анализ геопространственных данных	14	2	2	10
Тема 7. Этика использования больших данных	14	2	2	10
Тема 8. Презентация индивидуального исследовательского проекта	14	2	2	10
Итого	108	16	16	76

5. СОДЕРЖАНИЕ КУРСА

Тема 1. Большие данные и методы машинного обучения в естественнонаучных, медико-биологических и общественно-гуманитарных науках.

Новые типы данных: интернет-данные, другие цифровые следы и возможности их применения. Обсуждение идей индивидуальных исследовательских проектов.

Тема 2. Введение в язык программирования python.

Базовые типы данных. Переменные. Операторы. Условия, циклы и функции. Ошибки и предупреждения.

Тема 3. Автоматический сбор данных из интернета.

Форматы данных. HTML и JSON. Использование API интернет-сервисов на примере социальной сети ВКонтакте. Анализ социальных сетей: основные теоретические понятия и приложения. Изучение сетей дружбы на примере данных «ВКонтакте».

Тема 4. Использование методов машинного обучения для предсказания характеристик пользователей на основании их цифровых следов.

Анализ последовательностей. Прогнозирование и визуализация данных

Тема 5. Интеллектуальный анализ текстов.

Основные теоретические понятия и приложения. Тематическое моделирование. Анализ текстов из социальных сетей.

Тема 6. Анализ геопространственных данных.

Основные теоретические понятия и приложения. Методы сбора, практическое использование и интерпретация результатов.

Тема 7. Этика использования больших данных.

Алгоритмы и дискриминация. Применение технологий больших данных для задач управления в банковской, страховой, финансовой индустриях.

Тема 8. Презентация индивидуального исследовательского проекта.

6. ТЕМЫ ПРАКТИЧЕСКИХ И СЕМИНАРСКИХ ЗАНЯТИЙ

Тема 1. Большие данные и методы машинного обучения в естественнонаучных, медико-биологических и общественно-гуманитарных науках.

Вопросы для дискуссии:

1. Новые типы данных: интернет-данные, другие цифровые следы и возможности их применения.
2. Обсуждение идей индивидуальных исследовательских проектов.

Тема 2. Введение в язык программирования python.

Вопросы для дискуссии:

1. Базовые типы данных. Переменные. Операторы.
2. Условия, циклы и функции. Ошибки и предупреждения.

Тема 3. Автоматический сбор данных из интернета.

Вопросы для дискуссии:

1. Форматы данных. HTML и JSON. Использование API интернет-сервисов на примере социальной сети ВКонтакте.
2. Анализ социальных сетей: основные теоретические понятия и приложения. Изучение сетей дружбы на примере данных «ВКонтакте».

Тема 4. Использование методов машинного обучения для предсказания характеристик пользователей на основании их цифровых следов.

Вопросы для дискуссии:

1. Анализ последовательностей.
2. Прогнозирование и визуализация данных.

Тема 5. Интеллектуальный анализ текстов.

Вопросы для дискуссии:

1. Основные теоретические понятия и приложения.
2. Тематическое моделирование. Анализ текстов из социальных сетей.

Тема 6. Анализ геопространственных данных.

Вопросы для дискуссии:

1. Основные теоретические понятия и приложения.
2. Методы сбора, практическое использование и интерпретация результатов.

Тема 7. Этика использования больших данных.

Вопросы для дискуссии:

1. Алгоритмы и дискриминация.
2. Применение технологий больших данных для задач управления в банковской, страховой, финансовой индустриях.

Тема 8. Презентация индивидуального исследовательского проекта.

7. ТЕМАТИКА КОНТРОЛЬНЫХ РАБОТ, РЕФЕРАТОВ

Данный вид работы не предусмотрен УП

8. САМОСТОЯТЕЛЬНАЯ РАБОТА АСПИРАНТОВ

Название разделов и тем	Вид самостоятельной работы	Объем в часах	Форма контроля
Тема 1. Большие данные и методы машинного обучения в	Проработка учебного материала с использованием ресурсов учебно-методического и информационного обеспечения дисциплины.	9	Опрос, проверка конспектов,

социальных науках и исследованиях образования.	Подготовка к устному опросу. Подготовка к сдаче зачета.		вопросы на зачёте
Тема 2. Введение в язык программирования python.	Проработка учебного материала с использованием ресурсов учебно-методического и информационного обеспечения дисциплины. Подготовка к устному опросу. Подготовка к сдаче зачета.	9	Опрос, проверка конспектов, вопросы на зачёте
Тема 3. Автоматический сбор данных из интернета.	Проработка учебного материала с использованием ресурсов учебно-методического и информационного обеспечения дисциплины. Подготовка к устному опросу. Подготовка к сдаче зачета.	9	Опрос, проверка конспектов, вопросы на зачёте
Тема 4. Использование методов машинного обучения для предсказания характеристик пользователей на основании их цифровых следов.	Проработка учебного материала с использованием ресурсов учебно-методического и информационного обеспечения дисциплины. Подготовка к устному опросу. Подготовка к сдаче зачета.	9	Опрос, проверка конспектов, вопросы на зачёте
Тема 5. Интеллектуальный анализ текстов.	Проработка учебного материала с использованием ресурсов учебно-методического и информационного обеспечения дисциплины. Подготовка к устному опросу. Подготовка к сдаче зачета.	10	Опрос, проверка конспектов, вопросы на зачёте
Тема 6. Анализ геопространственных данных.	Проработка учебного материала с использованием ресурсов учебно-методического и информационного обеспечения дисциплины. Подготовка к устному опросу. Подготовка к сдаче зачета.	10	Опрос, проверка конспектов, вопросы на зачёте
Тема 7. Этика использования больших данных.	Проработка учебного материала с использованием ресурсов учебно-методического и информационного обеспечения дисциплины. Подготовка к устному опросу. Подготовка к сдаче зачета.	10	Опрос, проверка конспектов, вопросы на зачёте
Тема 8. Презентация индивидуального исследовательского проекта.	Проработка учебного материала с использованием ресурсов учебно-методического и информационного обеспечения дисциплины. Подготовка к устному опросу. Подготовка к сдаче зачета.	10	Опрос, проверка конспектов, вопросы на зачёте

ПЕРЕЧЕНЬ ВОПРОСОВ К ЗАЧЕТУ

1. Понятие Большие данные. Роль цифровой информации в 21 веке.
2. Виды массивов данных.
3. Базовые принципы обработки больших данных.
4. Технологии обработки больших данных: NoSQL, MapReduce, Hadoop, R.
5. Технологии Business Intelligence и реляционные системы управления базами данных.
6. Прогнозирование и предвидение: общее и особенное.
7. Виды прогнозов.
8. Вопросы безопасности больших данных.
9. Основные описательные статистики.
10. Регрессионный анализ.
11. Основная идея дисперсионного анализа.
12. Сущность кластерного анализа.
13. Дискриминантный анализ: модель и общая процедура выполнения.
14. Цели факторного анализа.
15. Программные средства анализа данных: Statistica, SPSS, Excel; их преимущества.

9. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

а) Список рекомендуемой литературы

Основная:

1. Неделько, В. М. Основы статистических методов машинного обучения : учебное пособие / В. М. Неделько. — Новосибирск : Новосибирский государственный технический университет, 2010. — 72 с. — ISBN 978-5-7782-1385-2. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/45418.html>

2. Нурматова, Е. В. Управление большими базами данных и высоконагруженными системами : учебное пособие / Е. В. Нурматова, Р. Ф. Халабия, Л. В. Бунина. — Москва : РТУ МИРЭА, 2019. — 120 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/171496>

Дополнительная:

1. Железнов, М. М. Методы и технологии обработки больших данных : учебно-методическое пособие / М. М. Железнов. — Москва : МИСИ – МГСУ, 2020. — 46 с. — ISBN 978-5-7264-2193-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/145102>

2. Сопов, Е. А. Многокритериальные нейроэволюционные системы в задачах машинного обучения и человеко-машинного взаимодействия : монография / Е. А. Сопов, И. А. Иванов. - Красноярск : Сиб. федер. ун-т, 2019. - 160 с. - ISBN 978-5-7638-3969-2. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1818898>

3. Ракитский, А. А. Методы машинного обучения : учебно-методическое пособие / А. А. Ракитский. — Новосибирск : Сибирский государственный университет телекоммуникаций и информатики, 2018. — 32 с. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/90591.html>

4. Шарден, Б. Крупномасштабное машинное обучение вместе с Python / Шарден Б. , Массарон Л. , Боскетти А. , пер. с англ. А. В. Логунова. - Москва : ДМК Пресс, 2018. - 358 с. - ISBN 978-5-97060-506-6. - Текст : электронный // ЭБС "Консультант студента" : [сайт]. - URL : <https://www.studentlibrary.ru/book/ISBN9785970605066.html>

Согласовано:

Специалист Вера Викторовна Стефановичева 01 мая 2023

Должность сотрудника НБ

ФИО

Подпись

Дата

б) Профессиональные базы данных, информационно-справочные системы

10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Учебные аудитории для проведения учебных занятий, предусмотренных программой бакалавриата, оснащенные оборудованием и техническими средствами обучения. Аудитории укомплектованы специализированной мебелью и техническими средствами обучения.

Помещения для самостоятельной работы обучающихся должны быть оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа к электронной информационно-образовательной среде Организации.

11. СПЕЦИАЛЬНЫЕ УСЛОВИЯ ДЛЯ ОБУЧАЮЩИХСЯ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ

В случае необходимости, обучающимся из числа лиц с ограниченными возможностями здоровья (по заявлению обучающегося) могут предлагаться одни из следующих вариантов восприятия информации с учетом их индивидуальных психофизических особенностей:

– для лиц с нарушениями зрения: в печатной форме увеличенным шрифтом; в форме электронного документа; в форме аудиофайла (перевод учебных материалов в аудиоформат); в печатной форме на языке Брайля; индивидуальные консультации с привлечением тифлосурдопереводчика; индивидуальные задания и консультации;

– для лиц с нарушениями слуха: в печатной форме; в форме электронного документа; видеоматериалы с субтитрами; индивидуальные консультации с привлечением сурдопереводчика; индивидуальные задания и консультации;

– для лиц с нарушениями опорно-двигательного аппарата: в печатной форме; в форме электронного документа; в форме аудиофайла; индивидуальные задания и консультации.

В случае необходимости использования в учебном процессе частично/исключительно дистанционных образовательных технологий, организация работа ППС с обучающимися с ОВЗ и инвалидами предусматривается в электронной информационно-образовательной среде с учетом их индивидуальных психофизических особенностей.

Разработчик



подпись

доцент ММТС Павлов П.Ю.

должность

ФИО