



## 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
<b>Тема 1.</b> Большие данные и методы машинного обучения в естественнонаучных, медико-биологических и общественно-гуманитарных науках	13	2	2	9
<b>Тема 2.</b> Введение в язык программирования python.	13	2	2	9
<b>Тема 3.</b> Автоматический сбор данных из интернета	13	2	2	9
<b>Тема 4.</b> Использование методов машинного обучения для предсказания характеристик пользователей на основании их цифровых следов	13	2	2	9
<b>Тема 5.</b> Интеллектуальный	14	2	2	10

анализ текстов				
<b>Тема 6.</b> Анализ геопространственных данных	14	2	2	10
<b>Тема 7.</b> Этика использования больших данных	14	2	2	10
<b>Тема 8.</b> Презентация индивидуального исследовательского проекта	14	2	2	10
<b>Итого</b>	<b>108</b>	<b>16</b>	<b>16</b>	<b>76</b>

## 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. ЛАБОРАТОРНЫЕ РАБОТЫ (ЛАБОРАТОРНЫЙ ПРАКТИКУМ)**

Данный вид работы не предусмотрен УП

### **8. ПРИМЕРНАЯ ТЕМАТИКА КУРСОВЫХ, КОНТРОЛЬНЫХ РАБОТ, РЕФЕРАТОВ**

Данный вид работы не предусмотрен УП

### **9. 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 и недостатки.

### 10. САМОСТОЯТЕЛЬНАЯ РАБОТА АСПИРАНТОВ

Название разделов и тем	Вид самостоятельной работы	Объем в часах	Форма контроля
<b>Тема 1.</b> Большие данные и методы машинного обучения в социальных науках и исследованиях образования.	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
<b>Тема 2.</b> 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
<b>Тема 3.</b> 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
<b>Тема 4.</b> 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
<b>Тема 5.</b> Интеллектуальный	Проработка учебного материала с использованием ресурсов учебно-	10	Опрос, проверка

анализ текстов.	методического и информационного обеспечения дисциплины. Подготовка к устному опросу. Подготовка к сдаче зачета.		конспектов, вопросы на зачёте
<b>Тема 6.</b> 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
<b>Тема 7.</b> 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
<b>Тема 8.</b> 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

## 11. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

## а) Список рекомендуемой литературы

### Основная:

Федоров, Д. Ю. Программирование на языке высокого уровня Python : учебное пособие для вузов / Д. Ю. Федоров. — 4-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2022. — 214 с. — (Высшее образование). — ISBN 978-5-534-15733-8. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/509562>

Платонов, А. В. Машинное обучение : учебное пособие для вузов / А. В. Платонов. — Москва : Издательство Юрайт, 2023. — 85 с. — (Высшее образование). — ISBN 978-5-534-15561-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/520544>

Черткова, Е. А. Статистика. Автоматизация обработки информации : учебное пособие для вузов / Е. А. Черткова. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2023. — 195 с. — (Высшее образование). — ISBN 978-5-534-01429-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/513393>

### Дополнительная:

Маккинли У., Python и анализ данных [Электронный ресурс] / Уэс Маккинли - М. : ДМК Пресс, 2015. - 482 с. - ISBN 978-5-97060-315-4 - Режим доступа: <http://www.studentlibrary.ru/book/ISBN9785970603154.html>

Рощин, С. М. Современные интернет-технологии. Семь главных трендов : научно-популярное издание / С. М. Рощин. - 2-е изд. - Москва : Дашков и К, 2022. - 124 с. - ISBN 978-5-394-04846-3. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1927306>

Цифровой бизнес : учебник / под науч. ред. О.В. Китовой. — Москва : ИНФРА-М, 2023. — 418 с. — (Высшее образование). — DOI 10.12737/textbook\_5a0a8c777462e8.90172645. - ISBN 978-5-16-013017-0. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1917620>

Статистическая обработка экспериментальных данных. Регрессионный анализ в языке R : учебное пособие / В. Ю. Потапова, А. С. Тарасов, Е. С. Геращенко, М. Б. Никифоров. — Рязань : РГРТУ, 2018. — 52 с. — ISBN 978-5-6041320-7-4. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/168238>

Согласовано:

**ДИРЕКТОР НБ**

Должность сотрудника НБ

**/ БУРХАНОВА М.М. /**

ФИО



подпись

**/ 15.04.2022**

дата

## б) Профессиональные базы данных, информационно-справочные системы



## **1. Электронно-библиотечные системы:**

1.1. Цифровой образовательный ресурс IPRsmart : электронно-библиотечная система : сайт / ООО Компания «Ай Пи Ар Медиа». - Саратов, [2022]. - URL: <http://www.iprbookshop.ru>. - Режим доступа: для зарегистрир. пользователей. - Текст : электронный.

1.2. Образовательная платформа ЮРАЙТ : образовательный ресурс, электронная библиотека : сайт / ООО Электронное издательство ЮРАЙТ. - Москва, [2022]. - URL: <https://urait.ru>. - Режим доступа: для зарегистрир. пользователей. - Текст : электронный.

1.3. Консультант врача. Электронная медицинская библиотека : база данных : сайт / ООО Высшая школа организации и управления здравоохранением-Комплексный медицинский консалтинг. - Москва, [2022]. - URL: <https://www.rosmedlib.ru>. - Режим доступа: для зарегистрир. пользователей. - Текст : электронный.

1.4. Большая медицинская библиотека : электронно-библиотечная система : сайт / ООО Букап. - Томск, [2022]. - URL: <https://www.books-up.ru/ru/library/>. - Режим доступа: для зарегистрир. пользователей. - Текст : электронный.

1.5. ЭБС Лань : электронно-библиотечная система : сайт / ООО ЭБС Лань. - Санкт-Петербург, [2022]. - URL: <https://e.lanbook.com>. - Режим доступа: для зарегистрир. пользователей. - Текст : электронный.

1.6. ЭБС Znanium.com : электронно-библиотечная система : сайт / ООО Знаниум. - Москва, [2022]. - URL: <http://znanium.com>. - Режим доступа : для зарегистрир. пользователей. - Текст : электронный.

1.7. Clinical Collection : научно-информационная база данных EBSCO // EBSCOhost : [портал]. - URL: <http://web.b.ebscohost.com/ehost/search/advanced?vid=1&sid=9f57a3e1-1191-414b-8763-e97828f9f7e1%40sessionmgr102>. - Режим доступа : для авториз. пользователей. - Текст : электронный.

1.8. База данных «Русский как иностранный» : электронно-образовательный ресурс для иностранных студентов : сайт / ООО Компания «Ай Пи Ар Медиа». - Саратов, [2022]. - URL: <https://ros-edu.ru>. - Режим доступа: для зарегистрир. пользователей. - Текст : электронный.

**2. КонсультантПлюс** [Электронный ресурс]: справочная правовая система. /ООО «Консультант Плюс» - Электрон. дан. - Москва : КонсультантПлюс, [2022].

## **3. Базы данных периодических изданий:**

3.1. База данных периодических изданий EastView : электронные журналы / ООО ИВИС. - Москва, [2022]. - URL: <https://dlib.eastview.com/browse/udb/12>. - Режим доступа : для авториз. пользователей. - Текст : электронный.

3.2. eLIBRARY.RU: научная электронная библиотека : сайт / ООО Научная Электронная Библиотека. - Москва, [2022]. - URL: <http://elibrary.ru>. - Режим доступа : для авториз. пользователей. - Текст : электронный

3.3. Электронная библиотека «Издательского дома «Гребенников» (Grebinnikon) : электронная библиотека / ООО ИД Гребенников. - Москва, [2022]. - URL: <https://id2.action-media.ru/Personal/Products>. - Режим доступа : для авториз. пользователей. - Текст : электронный.

**4. Федеральная государственная информационная система «Национальная электронная библиотека»** : электронная библиотека : сайт / ФГБУ РГБ. - Москва, [2022]. - URL: <https://нэб.рф>. - Режим доступа : для пользователей научной библиотеки. - Текст : электронный.

**5. SMART Imagebase** : научно-информационная база данных EBSCO // EBSCOhost : [портал]. - URL: <https://ebsco.smartimagebase.com/?TOKEN=EBSCO->

[1a2ff8c55aa76d8229047223a7d6dc9c&custid=s6895741](http://1a2ff8c55aa76d8229047223a7d6dc9c&custid=s6895741). – Режим доступа : для авториз. пользователей. – Изображение : электронные.

#### **6. Федеральные информационно-образовательные порталы:**

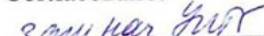
6.1. Единое окно доступа к образовательным ресурсам : федеральный портал . – URL: <http://window.edu.ru/> . – Текст : электронный.

6.2. Российское образование : федеральный портал / учредитель ФГАУ «ФИЦТО». – URL: <http://www.edu.ru>. – Текст : электронный.

#### **7. Образовательные ресурсы УлГУ:**

7.1. Электронная библиотечная система УлГУ : модуль «Электронная библиотека» АБИС Мега-ПРО / ООО «Дата Экспресс». – URL: <http://lib.ulsu.ru/MegaPro/Web>. – Режим доступа : для пользователей научной библиотеки. – Текст : электронный.

Согласовано:

  
Должность сотрудника УИТИТ

  
ФИО

  
подпись

19.04.22  
дата

## **12. 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.

## **13. 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



доцент ММТС Павлов П.Ю.

ПОДПИСЬ

ДОЛЖНОСТЬ

ФИО