

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Ястребов Олег Александрович

Должность: Ректор

Дата подписания: 29.05.2024 10:35:38

Уникальный программный ключ:

ca953a0120d891083f939673078ef1a989dae18a

**Федеральное государственное автономное образовательное учреждение высшего образования**

**«Российский университет дружбы народов имени Патриса Лумумбы»**

**Факультет искусственного интеллекта**

(наименование основного учебного подразделения (ОУП)-разработчика ОП ВО)

## **РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

### **СБОР И РАЗМЕТКА ДАННЫХ ДЛЯ МАШИННОГО ОБУЧЕНИЯ**

(наименование дисциплины/модуля)

**Рекомендована МССН для направления подготовки/специальности:**

### **02.03.02 ФУНДАМЕНТАЛЬНАЯ ИНФОРМАТИКА И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ**

(код и наименование направления подготовки/специальности)

**Освоение дисциплины ведется в рамках реализации основной профессиональной образовательной программы высшего образования (ОП ВО):**

### **ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: РАЗРАБОТКА И ОБУЧЕНИЕ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ**

(наименование (профиль/специализация) ОП ВО)

**2024 г.**

## 1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Сбор и разметка данных для машинного обучения» входит в программу бакалавриата «Искусственный интеллект: разработка и обучение интеллектуальных систем» по направлению 02.03.02 «Фундаментальная информатика и информационные технологии» и изучается в 3 семестре 2 курса. Дисциплину реализует NOT DEFINED. Дисциплина состоит из 8 разделов и 13 тем и направлена на изучение полного цикла работы с данными — от сбора обучающего набора данных до проверки и поддержания качества работы модели на меняющихся данных.

Целью освоения дисциплины является формирование компетенций, позволяющих собирать, обрабатывать и использовать данные для работы с ML-моделями.

## 2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Сбор и разметка данных для машинного обучения» направлено на формирование у обучающихся следующих компетенций (части компетенций):

*Таблица 2.1. Перечень компетенций, формируемых у обучающихся при освоении дисциплины (результаты освоения дисциплины)*

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
ПК-1	Способен создавать и оценивать различные модели машинного обучения и алгоритмы искусственного интеллекта с целью выбора наиболее эффективных решений для конкретных профессиональных задач	ПК-1.1 Умеет определять типы задач (классификация, регрессия, кластеризация и т.д.) и подходящие для них методы машинного обучения.; ПК-1.2 Владеет методами создания и обучения моделей с использованием различных алгоритмов и техник ML и AI; ПК-1.3 Применяет метрики для оценки производительности моделей, таких как точность, полнота, F1-мера и другие; ПК-1.4 Демонстрирует навыки обработки, представления и анализа данных для построения моделей машинного обучения;
ПК-2	Способен эффективно работать с большими объемами данных, включая их предварительную обработку, анализ и визуализацию, с целью извлечения полезной информации для обучения моделей искусственного интеллекта	ПК-2.1 Демонстрирует навыки анализа данных с использованием статистических методов и инструментов, таких как Python, R, SQL, NoSQL; ПК-2.2 Владеет методами работы с различными алгоритмами машинного обучения и глубинного обучения для решения различных задач;

## 3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Сбор и разметка данных для машинного обучения» относится к обязательной части блока 1 «Дисциплины (модули)» образовательной программы высшего образования.

В рамках образовательной программы высшего образования обучающиеся также осваивают другие дисциплины и/или практики, способствующие достижению запланированных результатов освоения дисциплины «Сбор и разметка данных для машинного обучения».

*Таблица 3.1. Перечень компонентов ОП ВО, способствующих достижению запланированных результатов освоения дисциплины*

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
ПК-1	Способен создавать и оценивать различные модели машинного обучения и алгоритмы искусственного интеллекта с целью выбора наиболее эффективных решений для конкретных профессиональных задач		<p>Преддипломная практика;  Технологическая (проектно-технологическая) практика (производственная);  Нейронные сети;  Управление проектами;  Практикум по программированию;  Методы машинного обучения;  Глубинное обучение;  Обработка естественного языка и векторизация текста;  Методы оптимизации в ML;  <i>Типовые задачи и применение ИИ в лингвистике**</i>;  <i>Типовые задачи и применение ИИ в финансах**</i>;  <i>Практикум по программированию в лингвистике**</i>;  <i>Практикум по программированию в финансах**</i>;  <i>Нетиповые задачи ИИ в лингвистике**</i>;  <i>Нетиповые задачи ИИ в финансах**</i>;  <i>Программирование на C++**</i>;  <i>Функциональное и параллельное программирование**</i>;</p>
ПК-2	Способен эффективно работать с большими объемами данных, включая их предварительную обработку, анализ и визуализацию, с целью извлечения полезной информации для обучения моделей искусственного интеллекта	Искусственный интеллект и когнитивная психология; Программирование на языке Python;	<p>NoSQL базы данных;  Нейронные сети;  Онтология и графы знаний;  Hadoop, SPARK;  Обработка сигналов;  Анализ временных рядов;  Распознавание образов и компьютерное зрение;  Обработка естественного языка и векторизация текста;  Методы оптимизации в ML;  <i>Типовые задачи и применение ИИ в лингвистике**</i>;  <i>Типовые задачи и применение ИИ в финансах**</i>;  <i>Практикум по программированию в лингвистике**</i>;  <i>Практикум по программированию в финансах**</i>;  <i>Нетиповые задачи ИИ в</i></p>

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
			<i>лингвистике**;</i> <i>Нетиповые задачи ИИ в финансах**;</i> Глубинное обучение; Преддипломная практика; Технологическая (проектно-технологическая) практика (производственная);

\* - заполняется в соответствии с матрицей компетенций и СУП ОП ВО

\*\* - элективные дисциплины /практики

#### 4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Сбор и разметка данных для машинного обучения» составляет «4» зачетные единицы.

Таблица 4.1. Виды учебной работы по периодам освоения образовательной программы высшего образования для очной формы обучения.

Вид учебной работы	ВСЕГО, ак.ч.		Семестр(-ы)
			3
<i>Контактная работа, ак.ч.</i>	48		48
Лекции (ЛК)	0		0
Лабораторные работы (ЛР)	32		32
Практические/семинарские занятия (СЗ)	16		16
<i>Самостоятельная работа обучающихся, ак.ч.</i>	69		69
<i>Контроль (экзамен/зачет с оценкой), ак.ч.</i>	27		27
<b>Общая трудоемкость дисциплины</b>	<b>ак.ч.</b>	<b>144</b>	<b>144</b>
	<b>зач.ед.</b>	<b>4</b>	<b>4</b>

## 5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Таблица 5.1. Содержание дисциплины (модуля) по видам учебной работы

Номер раздела	Наименование раздела дисциплины	Содержание раздела (темы)		Вид учебной работы*
Раздел 1	Введение.	1.1	Работа с данными в машинном обучении	ЛР, СЗ
Раздел 2	Общая концепция краудсорсинга.	2.1	Декомпозиция задачи	ЛР, СЗ
Раздел 3	Проектирование интерфейсов и инструкций для краудсорсинговых проектов	3.1	Проектирование интерфейсов и инструкций для краудсорсинговых проектов	ЛР, СЗ
Раздел 4	Контроль качества.	4.1	Этапы и инструменты	ЛР, СЗ
Раздел 5	Автоматизация конвейера по сбору данных	5.1	Автоматизация конвейера по сбору данных	ЛР, СЗ
Раздел 6	Проектная работа: сбор данных для обучения модели с целевым качеством	6.1	Проектная работа: сбор данных для обучения модели с целевым качеством	ЛР, СЗ
Раздел 7	Подходы к оптимизации бюджета	7.1	Подходы к оптимизации бюджета	ЛР, СЗ
		7.2	Агрегация полученных результатов: категориальная разметка, попарная разметка, агрегация сложных ответов	ЛР, СЗ
		7.3	Human-in-the-loop и другие подходы к оптимизации процессов разметки	ЛР, СЗ
		7.4	Проектная работа: сбор данных для обучения модели с целевым качеством и ограничением по бюджету	ЛР, СЗ
		7.5	Оценка качества работы модели	ЛР, СЗ
		7.6	Сбор данных для дообучения	ЛР, СЗ
Раздел 8	Проектная работа: сбор данных для обучения и поддержания качества модели в условиях меняющегося мира	8.1	Кейс-стади: работа с данными в Поиске, Алисе, Переводчике, Беспилотных технологиях, тестировании и других продуктах	ЛР, СЗ

\* - заполняется только по **ОЧНОЙ** форме обучения: ЛК – лекции; ЛР – лабораторные работы; СЗ – практические/семинарские занятия.

## 6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Таблица 6.1. Материально-техническое обеспечение дисциплины

Тип аудитории	Оснащение аудитории	Специализированное учебное/лабораторное оборудование, ПО и материалы для освоения дисциплины (при необходимости)
Компьютерный класс	Компьютерный класс для проведения занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная персональными компьютерами (в количестве 11 шт.), доской (экраном) и техническими средствами мультимедиа	Компьютерный класс для проведения занятий лекционного типа, практических занятий, групповых и индивидуальных консультаций, текущего

	презентаций.	<p>контроля и промежуточной аттестации. Комплект специализированной мебели; технические средства: Моноблок MSI (по) - 13  Моноблок Lenovo c560 – 3, Моноблок hp pro one – 1, Интерактивная система Smart Unifi45 – 1, Проектор Notevision – 1, Проекционный экран Cactus – 1, Компьютерная гарнитура -17, Портативная камера для документов – 1. Имеется выход в интернет.  Программное обеспечение: продукты Microsoft (ОС, пакет офисных приложений, в т.ч. MS Office/Office 365, Teams,).</p>
Семинарская	<p>Аудитория для проведения занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная комплектом специализированной мебели и техническими средствами мультимедиа презентаций.</p>	<p>Компьютерный класс для проведения занятий лекционного типа, практических занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации. Комплект специализированной мебели; технические средства: Моноблок MSI (по) - 13  Моноблок Lenovo c560 – 3, Моноблок hp pro one – 1, Интерактивная система Smart Unifi45 – 1, Проектор Notevision – 1, Проекционный экран Cactus – 1, Компьютерная гарнитура -17, Портативная камера для документов – 1. Имеется выход в интернет.  Программное обеспечение: продукты Microsoft (ОС, пакет</p>

		офисных приложений, в т.ч. MS Office/Office 365, Teams,).
Для самостоятельной работы	Аудитория для самостоятельной работы обучающихся (может использоваться для проведения семинарских занятий и консультаций), оснащенная комплектом специализированной мебели и компьютерами с доступом в ЭИОС.	Компьютерный класс для самостоятельной работы обучающихся. Комплект специализированной мебели (11 посадочных мест); технические средства: Моноблок Lenovo Idea Centre, проектор ViewSonic p9d6253, имеется выход в интернет, Проекционный экран Cactus, гарнитуры Plantronics Audio 655 Dsp Программное обеспечение: продукты Microsoft (ОС, пакет офисных приложений, в т.ч. MS Office/Office 365, Teams)

\* - аудитория для самостоятельной работы обучающихся указывается **ОБЯЗАТЕЛЬНО!**

## 7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

### Основная литература:

1. Сузи Р.А. Язык программирования Python [Электронный ресурс]/ Сузи Р.А.— Электрон. текстовые данные.— Москва: Интернет-Университет Информационных Технологий (ИНТУИТ), 2016.— 350 с.

2. Васильев А.Н. Python на примерах [Электронный ресурс]: практический курс по программированию/ Васильев А.Н.— Электрон. текстовые данные.— Санкт-Петербург: Наука и Техника, 2017.

### Дополнительная литература:

1. Горожанина Е.И. Нейронные сети [Электронный ресурс]: учебное пособие/Горожанина Е.И.— Электрон. текстовые данные.— Самара: Поволжский государственный университет телекоммуникаций и информатики, 2017.— 84 с.

2. Павлова А.И. Информационные технологии: основные положения теории искусственных нейронных сетей [Электронный ресурс]: учебное пособие/ Павлова А.И.— Электрон. текстовые данные.— Новосибирск: Новосибирский государственный университет экономики и управления «НИНХ», 2017.— 191 с.

### Ресурсы информационно-телекоммуникационной сети «Интернет»:

1. ЭБС РУДН и сторонние ЭБС, к которым студенты университета имеют доступ на основании заключенных договоров

- Электронно-библиотечная система РУДН – ЭБС РУДН

<http://lib.rudn.ru/MegaPro/Web>

- ЭБС «Университетская библиотека онлайн» <http://www.biblioclub.ru>

- ЭБС Юрайт <http://www.biblio-online.ru>

- ЭБС «Консультант студента» [www.studentlibrary.ru](http://www.studentlibrary.ru)

- ЭБС «Троицкий мост»

2. Базы данных и поисковые системы

- электронный фонд правовой и нормативно-технической документации

<http://docs.cntd.ru/>

- поисковая система Яндекс <https://www.yandex.ru/>
- поисковая система Google <https://www.google.ru/>
- реферативная база данных SCOPUS

<http://www.elsevier.com/locate/elsevier/scopus/>

*Учебно-методические материалы для самостоятельной работы обучающихся при освоении дисциплины/модуля\*:*

1. Курс лекций по дисциплине «Сбор и разметка данных для машинного обучения».

\* - все учебно-методические материалы для самостоятельной работы обучающихся размещаются в соответствии с действующим порядком на странице дисциплины **в ТУИС!**

## **8. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ И БАЛЛЬНО-РЕЙТИНГОВАЯ СИСТЕМА ОЦЕНИВАНИЯ УРОВНЯ СФОРМИРОВАННОСТИ КОМПЕТЕНЦИЙ ПО ДИСЦИПЛИНЕ**

Оценочные материалы и балльно-рейтинговая система\* оценивания уровня сформированности компетенций (части компетенций) по итогам освоения дисциплины «Сбор и разметка данных для машинного обучения» представлены в Приложении к настоящей Рабочей программе дисциплины.

\* - ОМ и БРС формируются на основании требований соответствующего локального нормативного акта РУДН.

**РАЗРАБОТЧИК:**

_____	_____	_____
<i>Должность, БУП</i>	<i>Подпись</i>	Андреева Марина Андреевна <i>Фамилия И.О.</i>

**РУКОВОДИТЕЛЬ БУП:**

_____	_____	_____
Заведующий кафедрой <i>Должность БУП</i>	<i>Подпись</i>	<i>Фамилия И.О.</i>

**РУКОВОДИТЕЛЬ ОП ВО:**

_____	_____	_____
<i>Должность, БУП</i>	<i>Подпись</i>	Воробьева Александра Андреевна <i>Фамилия И.О.</i>