

Документ подписан простой электронной подписью  
Информация о владельце:  
ФИО: Ястребов Олег Александрович  
Должность: Ректор  
Дата подписания: 22.05.2026 11:38:35  
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**

### **МАШИННОЕ ОБУЧЕНИЕ НА БОЛЬШИХ ДАННЫХ**

(наименование дисциплины/модуля)

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

### **02.04.02 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):**

### **УПРАВЛЕНИЕ ДАННЫМИ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ**

(наименование (профиль/специализация) ОП ВО)

**2026 г.**

## 1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Машинное обучение на больших данных» входит в программу магистратуры «Управление данными и искусственный интеллект» по направлению 02.04.02 «Фундаментальная информатика и информационные технологии» и изучается в 3 семестре 2 курса. Дисциплину реализует Кафедра прикладного искусственного интеллекта. Дисциплина состоит из 2 разделов и 18 тем и направлена на изучение формирования у студентов компетенций по применению и адаптации алгоритмов машинного обучения к распределённым, потоковым, высокообъёмным и разнообразным данным с использованием специализированных платформ анализа (Spark MLlib, H2O, TensorFlow for Big Data и др.). Курс фокусируется на эффективной обработке и анализе данных, недоступных для традиционных ML-инструментов из-за масштаба, распределённости, частой смены структуры и требований к высокой производительности. В магистратуре дисциплина строится на уникальных инженерных, вычислительных и аналитических аспектах, интеграции с инфраструктурой Big Data и не повторяет программирование алгоритмов «вручную», разобранные в бакалавриате. Целью освоения дисциплины является научить студентов разрабатывать, выбирать и внедрять pipeline машинного обучения для анализа больших данных, использовать современные распределённые ML-фреймворки, грамотно обрабатывать real-time и batch данные, анализировать и преодолевать инженерные, вычислительные и методологические вызовы при работе с Big Data, обеспечивать надёжность, масштабируемость, интерпретируемость и бизнес-ценность моделей.

## 2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Машинное обучение на больших данных» 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)
ОПК-1	Способен находить, формулировать и решать актуальные проблемы прикладной математики, фундаментальной информатики и информационных технологий	ОПК-1.3 Имеет практический опыт работы с решением математических задач и применяет его в профессиональной деятельности;
ОПК-5	Способен устанавливать и сопровождать программное обеспечение информационных систем, осуществлять эффективное управление разработкой программных средств и проектов	ОПК-5.1 осуществляет эффективное управление проектом разработки, внедрения и сопровождения программных средств и проектов на всех стадиях; ОПК-5.2 разрабатывает, внедряет и сопровождает программное обеспечение информационных систем;
ПК-2	Способен проектировать, разрабатывать и поддерживать интегрированное программное обеспечение с использованием нейросетевых моделей и сквозных технологий искусственного интеллекта	ПК-2.1 Знает методы математического моделирования и машинного обучения, используемые при разработке требований и проектировании программного обеспечения систем и моделей искусственного интеллекта; ПК-2.2 Выбирает и моделирует архитектурные решения для реализации интегрированного программного обеспечения с использованием нейросетевых моделей и сквозных технологий искусственного интеллекта; ПК-2.3 Имеет навыки применения методов математического моделирования объектов и процессов, машинного обучения при разработке требований и проектировании программного

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
		обеспечения систем и моделей искусственного интеллекта;

### 3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Машинное обучение на больших данных» 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 «Машинное обучение на больших данных».

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

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
ОПК-1	Способен находить, формулировать и решать актуальные проблемы прикладной математики, фундаментальной информатики и информационных технологий	Прикладная статистика и анализ данных; Глубокое обучение для NLP; Компьютерное зрение; SQL и NoSQL базы данных;	
ОПК-5	Способен устанавливать и сопровождать программное обеспечение информационных систем, осуществлять эффективное управление разработкой программных средств и проектов	Методы машинного обучения (продвинутый курс);	
ПК-2	Способен проектировать, разрабатывать и поддерживать интегрированное программное обеспечение с использованием нейросетевых моделей и сквозных технологий искусственного интеллекта	Методы машинного обучения (продвинутый курс); Компьютерное зрение; Глубокое обучение в компьютерном зрении; Программирование на языке C++ (продвинутый курс); Искусственный интеллект по отраслям**; Вайб-кодинг**;	Преддипломная практика; Генеративный искусственный интеллект; Искусственный интеллект в компьютерных играх**;

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

\*\* - элективные дисциплины /практики

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

Общая трудоемкость дисциплины «Машинное обучение на больших данных» 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>	36		36
Лекции (ЛК)	12		12
Лабораторные работы (ЛР)	0		0
Практические/семинарские занятия (СЗ)	24		24
<i>Самостоятельная работа обучающихся, ак.ч.</i>	81		81
<i>Контроль (экзамен/зачет с оценкой), ак.ч.</i>	27		27
<b>Общая трудоемкость дисциплины ак.ч.</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	Наименование темы		Содержание темы	Вид учебной работы*
Раздел 1	Архитектура и инфраструктура ML на больших данных	1.1	Архитектурные принципы: особенности, проблемы, задачи ML на Big Data	Рассмотрение фундаментальных принципов построения ML-систем для больших данных. Анализ основных проблем масштабирования: вычислительная сложность, распределенная обработка, управление ресурсами. Изучение специфики задач машинного обучения в условиях больших объемов данных, требования к архитектуре систем.	ЛК
		1.2	Обзор индустриальных фреймворков: Spark MLlib, H2O, TensorFlowOnSpark	Детальное изучение популярных промышленных инструментов для распределенного машинного обучения. Сравнительный анализ возможностей, производительности и областей применения каждого фреймворка. Практические примеры использования и интеграции в существующую инфраструктуру.	ЛК
		1.3	Построение пайплайнов данных и интеграция потоковых и статических задач ML	Проектирование и реализация конвейеров обработки данных для ML-задач. Методы объединения batch и streaming обработки. Организация ETL-процессов, управление зависимостями между этапами обработки данных.	ЛК
		1.4	Развёртывание кластера для ML-задач (Spark/H2O): настройка рабочей среды	Практическое руководство по установке и конфигурации распределенных систем. Оптимизация параметров кластера под конкретные задачи ML. Настройка сетевого взаимодействия, управление памятью и вычислительными ресурсами.	
		1.5	Импорт и подготовка крупномасштабного датасета в распределённом окружении	Техники эффективной загрузки больших объемов данных из различных источников. Методы партиционирования и распределения данных по узлам кластера. Оптимизация операций чтения и записи для повышения производительности.	
		1.6	Построение пайплайна обработки и загрузки потоковых данных	Создание систем для обработки данных в реальном времени. Интеграция с источниками потоковых данных (Kafka, Kinesis). Реализация windowing операций, обработка событий и управление состоянием.	
		1.7	Кейсы проектирования инфраструктуры ML-проектов для разных бизнес-сценариев	Анализ реальных примеров архитектурных решений для различных индустрий. Выбор оптимальной инфраструктуры в зависимости от требований бизнеса. Учет факторов стоимости, масштабируемости и надежности.	СЗ
		1.8	Обсуждение архитектурных ошибок, bottleneck и диагностик производительности	Типичные антипаттерны в проектировании ML-систем. Методы выявления узких мест в производительности. Инструменты мониторинга и профилирования распределенных приложений. Стратегии оптимизации.	СЗ
		1.9	Разбор схем потоковой и оффлайн обработки ML, их плюсы/минусы	Сравнительный анализ batch и streaming подходов к обработке данных. Гибридные архитектуры (Lambda, Kappa). Критерии выбора подходящей схемы обработки для конкретных задач.	СЗ

Раздел 2	Методы и алгоритмы ML в среде больших данных	2.1	Адаптация и оптимизация классических алгоритмов ML для распределённых и потоковых данных	Модификация традиционных алгоритмов для работы в распределенной среде. Параллелизация вычислений, методы разделения данных и моделей. Техники инкрементального обучения для потоковых данных.	ЛК
		2.2	Online learning: обучение на	Алгоритмы онлайн-обучения и их особенности. Работа с концепт-дрифтом и	ЛК

Номер раздела	Наименование раздела дисциплины	Наименование темы	Содержание темы	Вид учебной работы*
		нестационарных и потоковых данных. Batch vs streaming	изменяющимися распределениями. Сравнение эффективности batch и streaming подходов, выбор оптимальной стратегии.	
		2.3 Вопросы качества, мониторинга и устойчивости ML-моделей в Big Data-проектах	Метрики качества моделей в производственной среде. Системы мониторинга деградации моделей. Обеспечение воспроизводимости результатов и версионирование моделей.	ЛК
		2.4 Обучение и сравнение моделей (регрессия, классификация, кластеризация) в Spark MLlib или H2O	Практическая работа с распределенными реализациями популярных алгоритмов. Настройка гиперпараметров, кросс-валидация на больших данных. Сравнительный анализ производительности различных моделей.	
		2.5 Реализация онлайн-алгоритмов (streaming ML) и их оценка по потоковым метрикам	Разработка моделей для обучения в реальном времени. Методы оценки качества на потоковых данных. Адаптивные алгоритмы и механизмы обновления моделей.	
		2.6 Автоматизация мониторинга качества и drift-детекции в Big Data ML pipeline	Построение систем автоматического обнаружения деградации моделей. Детектирование изменений в распределении данных. Триггеры для переобучения и алерты при падении качества.	
		2.7 Разбор рисков и особенностей качества моделей при работе с Big Data (precision, latency, reproducibility)	Компромиссы между точностью и скоростью работы моделей. Влияние распределенных вычислений на воспроизводимость результатов. Управление latency в реал-тайм системах.	СЗ
		2.8 Семинар-обсуждение стратегий поддержки и реагирования на деградацию моделей	Практические подходы к поддержке моделей в production. Стратегии A/B тестирования и постепенного внедрения обновлений. Процессы реагирования на инциденты и восстановления работоспособности.	СЗ
		2.9 Работа с метриками и интерпретацией моделей на сложных потоках данных	Методы интерпретации моделей машинного обучения в распределенной среде. Feature importance на больших данных. Визуализация и объяснение предсказаний для бизнес-стейкхолдеров.	СЗ

\* - заполняется только по ОЧНОЙ форме обучения: ЛК – лекции; ЛР – лабораторные работы; СЗ – практически/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)
Лекционная	Аудитория для проведения занятий лекционного типа, оснащенная комплектом специализированной мебели; доской (экраном) и техническими средствами мультимедиа презентаций.	
Компьютерный класс	Компьютерный класс для проведения занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная персональными компьютерами (в количестве [Параметр] шт.), доской (экраном) и техническими средствами мультимедиа презентаций.	
Семинарская	Аудитория для проведения занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная комплектом специализированной мебели и техническими средствами мультимедиа презентаций.	
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.	

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

## 7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

### Основная литература:

1. Основы машинного обучения и нейронные сети. учебное пособие [Электронный ресурс] / Бобрикова Екатерина Васильевна [и др.]. - М.: РУДН, 2024. 124 с. ISBN 978-5-209-12322-4 URL: [https://mega.rudn.ru/MegaPro/UserEntry?Action=Link\\_FindDoc&id=517203&idb=0](https://mega.rudn.ru/MegaPro/UserEntry?Action=Link_FindDoc&id=517203&idb=0)

2. Аханова, Марина Анатольевна. Большие данные и машинное обучение [Текст]: учебник / М. А. Аханова, С. В. Овчинникова, О. М. Барбаков. - Тюмень: ТИУ, 2022. - 201 с.: ил.; ISBN 978-5-9961-2960-7

### Дополнительная литература:

1. Параскевов, А. В. Большие данные: учебное пособие / А. В. Параскевов, А. Э. Сергеев. - Москва; Вологда: Инфра-Инженерия, 2024. - 148 с. - ISBN 978-5-9729-2120-1. - Текст: электронный. - URL: <https://znanium.ru/catalog/product/2169699>

2. Machine Learning and Big Data: Concepts, Algorithms, Tools and Applications Share Icon. Uma N. Dulhare (Editor), Khaleel Ahmad (Editor), Khairol Amali Bin Ahmad (Editor). ISBN: 978-1-119-65474-2. 2020.

### 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 <https://mega.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)



## 2. Базы данных и поисковые системы

- Sage <https://journals.sagepub.com/>
- Springer Nature Link <https://link.springer.com/>
- Wiley Journal Database <https://onlinelibrary.wiley.com/>
- Наукометрическая база данных Lens.org <https://www.lens.org>

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

1. Курс лекций по дисциплине «Машинное обучение на больших данных».

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

## РАЗРАБОТЧИКИ

---

Должность

### **РУКОВОДИТЕЛЬ БУП**

Заведующий кафедрой

---

Должность

### **РУКОВОДИТЕЛЬ ОП ВО**

Заведующий кафедрой прикладного  
искусственного интеллекта

---

Должность

---

Подолько П.М.

Фамилия И.О

---

Подолько П.М.

Фамилия И.О

---

Подолько П.М.

Фамилия И.О