

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Ястребов Олег Александрович
Должность: Ректор
Дата подписания: 27.05.2026 08:22:31
Уникальный программный ключ:
ca953a0120d891083f939673078ef1a989dae18a

**Федеральное государственное автономное образовательное учреждение высшего образования
«Российский университет дружбы народов имени Патриса Лумумбы»**

Инженерная академия

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

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

МАШИННОЕ ОБУЧЕНИЕ И АНАЛИЗ БОЛЬШИХ ДАННЫХ

(наименование дисциплины/модуля)

Рекомендована МССН для направления подготовки/специальности:

27.04.04 УПРАВЛЕНИЕ В ТЕХНИЧЕСКИХ СИСТЕМАХ

(код и наименование направления подготовки/специальности)

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

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И РОБОТОТЕХНИЧЕСКИЕ СИСТЕМЫ

(наименование (профиль/специализация) ОП ВО)

2026 г.

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Машинное обучение и анализ больших данных» входит в программу магистратуры «Искусственный интеллект и робототехнические системы» по направлению 27.04.04 «Управление в технических системах» и изучается в 1 семестре 1 курса. Дисциплину реализует Кафедра механики и процессов управления. Дисциплина состоит из 5 разделов и 20 тем и направлена на изучение фундаментальных основ аппарата комбинаторики и мат. статистики, регрессионного анализа и сжатия данных, детектированных выбросов и аномалий, очистки данных и технологий регуляризации, технологий кластеризации и классификации, нейронных сетей, генетических алгоритмов, выделения особенностей (feature detection), нормализации данных, нечетких множеств, байесовых сетей; разбор основных методов решения типовых задач и знакомство с областью их применения в профессиональной деятельности.

Целью освоения дисциплины является формирование фундаментальных знаний и навыков применения методов решения задач, необходимых для профессиональной деятельности (например, последующая разработка стратегий продвижения продуктов и услуг в цифровом пространстве), повышение общего уровня цифровой грамотности студентов,

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Машинное обучение и анализ 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)
ПК-1	Способен формулировать цели, задачи научных исследований в области искусственного интеллекта, выбирать методы и средства решения задач	ПК-1.1 Знает методы и средства решения задач научных исследований в области систем искусственного интеллекта и робототехнических систем;; ПК-1.2 Умеет формулировать цель и задачи научных исследований в профессиональной области;; ПК-1.3 Владеет приемами для формулировки цели и задач научных исследований, умеет выбирать методы и средства решения задач профессиональной деятельности.;
ПК-2	Способен применять современные теоретические и экспериментальные методы разработки математических моделей исследуемых объектов и процессов, относящихся к профессиональной деятельности по направлению подготовки	ПК-2.1 Знает современные теоретические и экспериментальные методы, применяемые для разработки математических моделей исследуемых объектов и процессов профессиональной деятельности;; ПК-2.2 Умеет определять эффективность применяемых методов для разработки математических моделей исследуемых объектов и процессов;; ПК-2.3 Владеет современными теоретическими и экспериментальными методами для разработки математических моделей объектов и процессов профессиональной деятельности по направлению подготовки.;

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Машинное обучение и анализ 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 и анализ 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*
ПК-1	Способен формулировать цели, задачи научных исследований в области искусственного интеллекта, выбирать методы и средства решения задач		Технологическая практика; Искусственные нейронные сети (Глубокое обучение); Искусственные нейронные сети (Обучение с подкреплением); Интеллектуальные информационные системы; Geoinformation Systems and Applications; Проектирование робототехнических систем; Преддипломная практика;
ПК-2	Способен применять современные теоретические и экспериментальные методы разработки математических моделей исследуемых объектов и процессов, относящихся к профессиональной деятельности по направлению подготовки		Научно-исследовательская работа; Технологическая практика; Преддипломная практика; Искусственные нейронные сети (Глубокое обучение); Искусственные нейронные сети (Обучение с подкреплением); <i>Dynamics and Control of Space Systems</i> **; <i>Динамика и управление космическими системами</i> **;

* - 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 и анализ больших данных» составляет «5» зачетных единиц.

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)
			1
<i>Контактная работа, ак.ч.</i>	34		34
Лекции (ЛК)	17		17
Лабораторные работы (ЛР)	17		17
Практические/семинарские занятия (СЗ)	0		0
<i>Самостоятельная работа обучающихся, ак.ч.</i>	119		119
<i>Контроль (экзамен/зачет с оценкой), ак.ч.</i>	27		27
Общая трудоемкость дисциплины	ак.ч.	180	180
	зач.ед.	5	5

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	Введение в машинное обучение и обработку данных. Программный инструментарий Data Mining и Machine Learning.	1.1	Введение в машинное обучение и обработку данных. Постановка основных классов задач в машинном обучении.	Машинное обучение как научное направление, изучающее методы построения алгоритмов, способных обучаться на основе эмпирических данных без строгого следования заранее заданным правилам. Основные компоненты: признаки как независимые переменные, целевая переменная как зависимая переменная, обучающая выборка. Классификация задач машинного обучения: обучение с учителем с использованием размеченных данных, обучение без учителя в отсутствие целевой переменной, обучение с подкреплением для формирования поведения на основе взаимодействия со средой.	ЛК, ЛР
		1.2	Регрессия и классификация; кластеризация, снижение размерности	Задача регрессии как предсказание непрерывной числовой величины. Задача классификации как отнесение объекта к одной из дискретных категорий. Задача кластеризации как группировка объектов на основе их сходства при отсутствии меток классов. Задача снижения размерности как преобразование пространства признаков с сохранением существенной информации при уменьшении числа переменных.	ЛК, ЛР
		1.3	Обработка текстов; обработка изображений	Методы обработки текстовых данных: представление текстовой информации в числовом виде через модель мешка слов и метод TF-IDF. Основные задачи анализа текстов: классификация документов, тональный анализ, извлечение сущностей. Методы обработки изображений: представление визуальной информации в виде матриц пикселей. Основные задачи компьютерного зрения: распознавание объектов, сегментация изображений, детектирование границ.	ЛК, ЛР
Раздел 2	Регрессионный анализ и сжатие данных.	2.1	Задача регрессии. Минимизация квадрата отклонения. Регрессионная функция: условное мат.ожидание	Задача регрессии как предсказание числового значения целевой переменной на основе входных признаков. Минимизация квадрата отклонения как стандартный подход к оценке качества регрессии. Регрессионная функция как условное математическое ожидание целевой переменной при заданных значениях признаков. Связь между квадратичной функцией потерь и условным математическим ожиданием.	ЛК, ЛР

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
		2.2	Линейная регрессия и метод k ближайших соседей. Переобучение и недообучение	Линейная регрессия как модель, предполагающая линейную зависимость между признаками и целевой переменной с нахождением коэффициентов через минимизацию среднеквадратичной ошибки. Метод k ближайших соседей как непараметрический метод регрессии и классификации с предсказанием по ближайшим объектам обучающей выборки. Переобучение как чрезмерная подстройка модели под обучающие данные с потерей способности к обобщению. Недообучение как чрезмерное упрощение модели, неспособной уловить закономерности в данных.	ЛК, ЛР
		2.3	Разложение ошибки на шум, смещение и разброс	Разложение общей ошибки предсказания на три составляющие: шум как неустраняемая ошибка из-за случайности данных, смещение как систематическая ошибка из-за неверных предположений модели о данных, разброс как ошибка из-за чувствительности модели к малым колебаниям обучающей выборки. Компромисс смещения и разброса: простые модели имеют высокое смещение и низкий разброс, сложные модели наоборот.	ЛК, ЛР
Раздел 3	Детектирование выбросов и аномалий. Очистка данных и технологии регуляризации.	3.1	Детектирование выбросов и аномалий. Что такое выбросы, типы выбросов	Выбросы как объекты данных, существенно отличающиеся от основной массы наблюдений. Типы выбросов: точечные выбросы с аномальным значением одного признака, контекстуальные выбросы, аномальные в конкретных условиях, коллективные выбросы как группы объектов, отклоняющиеся от общего распределения. Причины возникновения выбросов: ошибки измерений, редкие события, сбой оборудования.	ЛК, ЛР
		3.2	Методы обнаружения выбросов. Поиск аномалий	Статистические методы обнаружения выбросов на основе правила трёх сигма и межквартильного размаха. Методы на основе расстояний: расстояние Махаланобиса, метод ближайших соседей. Методы на основе плотности: локальный выброс по плотности. Машинное обучение для поиска аномалий: одноклассовая классификация, изолирующий лес, автокодировщики.	ЛК, ЛР
		3.3	Цензурирование выборки. Отсев объектов-выбросов, удаление выбросов	Цензурирование выборок как процесс ограничения экстремальных значений для уменьшения влияния выбросов. Отсев объектов-выбросов с полным удалением аномальных записей из набора данных. Критерии удаления выбросов:	ЛК, ЛР

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
				превышение пороговых значений, несоответствие логическим ограничениям, невозможность верификации. Риски при удалении выбросов: потеря полезной информации при ошибочном отсеве редких, но значимых событий.	
		3.4	Очистка данных и технологии регуляризации. Основные виды регуляризации	Очистка данных как комплекс мер по выявлению и исправлению ошибок, пропусков и несоответствий в данных. Регуляризация как технология борьбы с переобучением путём добавления штрафного члена к функции потерь. Основные виды регуляризации: L1-регуляризация лассо, приводящая к разреженным решениям, L2-регуляризация ридж, равномерно уменьшающая веса, Elastic Net как комбинация L1 и L2.	ЛК, ЛР
		3.5	Метод редукции размерности. Методы отбора признаков	Редукция размерности как процесс уменьшения числа признаков при сохранении существенной информации. Методы отбора признаков: фильтрационные методы с оценкой важности каждого признака независимо, обёрточные методы с перебором подмножеств признаков на основе качества модели, встроенные методы с отбором признаков внутри процесса обучения. Метод главных компонент как линейное преобразование исходных признаков в новые некоррелированные компоненты.	ЛК, ЛР
Раздел 4	Технологии кластеризации и классификации. Нейронные сети. Генетические алгоритмы.	4.1	Технологии кластеризации и классификации. K-means. EM-алгоритм	Кластеризация как задача группировки объектов без учителя на основе их сходства. Метод k-средних с итеративным отнесением объектов к ближайшим центрам кластеров и пересчётом центров. Выбор числа кластеров через метод локтя и силуэтный коэффициент. EM-алгоритм для вероятностной кластеризации с предположением о смеси гауссовых распределений. Чередование E-шага с оценкой принадлежности объектов и M-шага с пересчётом параметров распределений.	ЛК, ЛР
		4.2	Другие методы кластеризации. Задачи классификации. Байесовский классификатор	Агломеративная иерархическая кластеризация с последовательным объединением ближайших объектов в кластеры. DBSCAN как метод кластеризации на основе плотности с выделением областей высокой плотности. Классификация как задача отнесения объекта к одному из заранее известных классов. Байесовский классификатор, основанный на теореме Байеса и предположении о независимости признаков наивный байесовский классификатор.	ЛК, ЛР

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
		4.3	Линейные методы для классификации. Логистическая регрессия, максимизация правдоподобия	Линейные методы классификации с разделением классов линейной границей в пространстве признаков. Логистическая регрессия как метод бинарной классификации с предсказанием вероятности принадлежности к классу через сигмоидальную функцию. Максимизация правдоподобия как критерий обучения логистической регрессии вместо минимизации квадрата ошибки. Интерпретация коэффициентов логистической регрессии как влияния признаков на логарифм отношения шансов.	ЛК, ЛР
		4.4	Нейронные сети: общая архитектура. Многослойные сети. Обратное распространение ошибки	Общая архитектура нейронной сети: входной слой признаков, один или несколько скрытых слоёв с нелинейными функциями активации, выходной слой с размерностью, соответствующей задаче. Многослойные сети как универсальные аппроксиматоры любых непрерывных функций. Прямой проход для вычисления выходов сети при заданных входных данных. Обратное распространение ошибки как алгоритм вычисления градиентов функции потерь по весам сети через последовательное применение правила цепочки от выходного слоя к входному.	ЛК, ЛР
		4.5	Стохастический градиентный спуск. Генетические алгоритмы	Стохастический градиентный спуск как оптимизационный метод для обучения нейронных сетей с обновлением весов по случайному мини-батчу примеров, а не по всей выборке. Преимущества стохастического градиентного спуска: более быстрая сходимость, возможность работы с большими наборами данных, выход из локальных минимумов. Генетические алгоритмы как эволюционные методы оптимизации, имитирующие процессы естественного отбора. Основные операции генетического алгоритма: селекция лучших решений, скрещивание для обмена частями хромосом, мутация для случайного изменения.	ЛК, ЛР
Раздел 5	Выделение особенностей (Feature detection); нормализация данных. Нечеткие множества. Байесовы сети.	5.1	Извлечение признаков / выделение особенностей (Feature detection)	Извлечение признаков как процесс создания новых информативных признаков из сырых данных. Отличие извлечения признаков от отбора признаков: извлечение создаёт новые признаки, отбор выбирает подмножество существующих. Методы извлечения признаков для изображений: детекторы углов и границ, гистограммы	ЛК, ЛР

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
				ориентированных градиентов. Методы извлечения признаков для текстов: n-граммы, тематическое моделирование, векторные представления слов.	
		5.2	Преобразования признаков. Нормализация данных. Методы нормализации данных	Преобразования признаков как приведение данных к виду, удобному для обучения моделей. Нормализация данных как масштабирование признаков к определённому диапазону или распределению. Необходимость нормализации для методов, чувствительных к масштабу признаков: градиентный спуск, метод главных компонент, методы на основе расстояний.	ЛК, ЛР
		5.3	Нормализация по методу минимакса. Нормализация по Z-показателю. Десятичное масштабирование	Нормализация по методу минимакс с линейным преобразованием признаков в заданный диапазон, обычно от нуля до единицы. Сохранение формы исходного распределения при минимаксной нормализации. Нормализация по Z-показателю как центрирование и масштабирование к единичной дисперсии: вычитание среднего и деление на стандартное отклонение. Преобразование признаков к стандартному нормальному распределению. Десятичное масштабирование как деление всех значений признака на степень десятки для смещения десятичной точки. Простота и обратимость десятичного масштабирования.	ЛК, ЛР
		5.4	Нечеткие множества. Байесовы сети. Задачи байесовского вывода. Методика построения нечеткой байесовой сети	Нечёткие множества как расширение классической теории множеств с принадлежностью элементов, задаваемой непрерывной функцией от нуля до единицы. Функции принадлежности для описания лингвистических переменных: высокий, средний, низкий. Байесовы сети как вероятностные графовые модели, представляющие зависимости между переменными через направленные ациклические графы. Задачи байесовского вывода: вычисление апостериорных вероятностей переменных при наличии наблюдений. Методика построения нечёткой байесовой сети: объединение нечёткой логики для работы с неточными данными и байесовских сетей для вероятностного вывода. Применение в системах поддержки принятия решений при неполной или нечёткой информации.	ЛК, ЛР

* - 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)
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 ____ шт.), доской (экраном) 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. James, G. et al. An introduction to statistical learning. – Springer, 2013. – 426 pp
2. Trevor Hastie, Robert Tibshirani, et al., The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edition, 2017
3. Вьюгин, В. В. Математические основы машинного обучения и прогнозирования : учебное пособие / В. В. Вьюгин. — Москва : МЦНМО, 2014. — 304 с.

Дополнительная литература:

1. Bruce, P. C., & Bruce, A. (2017). Practical Statistics for Data Scientists : 50 Essential Concepts (Vol. First edition). Sebastopol, CA: O'Reilly Media
2. Molnar, C. (2018). iml: An R package for Interpretable Machine Learning
3. Explainable and interpretable models in computer vision and machine learning. (2018)
4. Комбинаторика и теория вероятностей, учебное пособие, 99 с., Райгородский, А. М., 2013

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
- 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/scopus/>

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

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

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

РАЗРАБОТЧИКИ:

Доцент

Должность, БУП

Подпись

Салтыкова Ольга
Александровна

Фамилия И.О.

Ассистент

Должность, БУП

Подпись

Дам Ван Ньить

Фамилия И.О.

РУКОВОДИТЕЛЬ БУП:

Заведующий кафедрой

Должность, БУП

Подпись

Разумный Юрий
Николаевич

Фамилия И.О.

РУКОВОДИТЕЛЬ ОП ВО:

Профессор

Должность, БУП

Подпись

Разумный Юрий
Николаевич

Фамилия И.О.