

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Ястребов Олег Александрович
Должность: Ректор
Дата подписания: 19.05.2026 18:13:28
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

АНАЛИЗ ДАННЫХ И МАШИННОЕ ОБУЧЕНИЕ

(наименование дисциплины/модуля)

Рекомендована МССН для направления подготовки/специальности:

01.03.02 ПРИКЛАДНАЯ МАТЕМАТИКА И ИНФОРМАТИКА

(код и наименование направления подготовки/специальности)

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

МАТЕМАТИЧЕСКИЕ МЕТОДЫ МЕХАНИКИ КОСМИЧЕСКОГО ПОЛЕТА И АНАЛИЗА ГЕОИНФОРМАЦИОННЫХ ДАННЫХ

(наименование (профиль/специализация) ОП ВО)

2026 г.

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Анализ данных и машинное обучение» входит в программу бакалавриата «Математические методы механики космического полета и анализа геоинформационных данных» по направлению 01.03.02 «Прикладная математика и информатика» и изучается в 5, 6 семестрах 3 курса. Дисциплину реализует Кафедра механики и процессов управления. Дисциплина состоит из 5 разделов и 22 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 «Анализ данных и машинное обучение» 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-4	Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности	ОПК-4.1 Знает основные требования информационной безопасности, существующие информационно коммуникационные технологии; ОПК-4.2 Умеет решать задачи профессиональной деятельности с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности; ОПК-4.3 Владеет навыками использования существующих информационных технологий для решения задач профессиональной деятельности;

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Анализ данных и машинное обучение» 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 «Анализ данных и машинное обучение».

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-4	Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности	Информатика и программирование;	Преддипломная практика; Технологическая практика; Научно-исследовательская работа;

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

** - элективные дисциплины /практики

4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Анализ данных и машинное обучение» составляет «6» 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)	
			5	6
<i>Контактная работа, ак.ч.</i>	108		36	72
Лекции (ЛК)	54		18	36
Лабораторные работы (ЛР)	54		18	36
Практические/семинарские занятия (СЗ)	0		0	0
<i>Самостоятельная работа обучающихся, ак.ч.</i>	81		72	9
<i>Контроль (экзамен/зачет с оценкой), ак.ч.</i>	27		0	27
Общая трудоемкость дисциплины	ак.ч.	216	108	108
	зач.ед.	6	3	3

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.	Определение машинного обучения как научного направления, изучающего методы построения алгоритмов, способных обучаться на основе эмпирических данных без строгого следования заранее заданным правилам. Характеристика основных компонентов: признаки (независимые переменные), целевая переменная (зависимая переменная), обучающая выборка. Классификация задач машинного обучения: обучение с учителем (наличие размеченных данных), обучение без учителя (отсутствие целевой переменной), обучение с подкреплением (формирование поведения на основе взаимодействия со средой).	ЛК
		1.2	Регрессия и классификация; кластеризация, снижение размерности	Определение задачи регрессии как предсказания непрерывной числовой величины. Определение задачи классификации как отнесения объекта к одной из дискретных категорий. Определение задачи кластеризации как группировки объектов на основе их сходства в отсутствие меток классов. Определение задачи снижения размерности как преобразования пространства признаков с целью сохранения существенной информации при уменьшении числа переменных.	ЛК, ЛР
		1.3	Обработка текстов; обработка изображений	Характеристика методов обработки текстовых данных: представление текстовой информации в числовом виде (модель «мешка слов», метод TF-IDF). Основные задачи анализа текстов: классификация документов, тональный анализ, извлечение сущностей. Характеристика методов обработки изображений: представление визуальной информации в виде матриц пикселей. Основные задачи компьютерного зрения: распознавание объектов, сегментация изображений, детектирование границ.	ЛК
Раздел 2	Основной аппарат комбинаторики и мат. статистики. Регрессионный анализ и сжатие данных.	2.1	Основной аппарат комбинаторики и мат. статистики. Основные понятия математической статистики	Изложение элементов комбинаторики: понятия перестановок, сочетаний, размещений. Определение основных категорий математической статистики: генеральная совокупность, выборка, случайная величина, функция распределения, плотность распределения. Характеристика числовых мер	ЛК

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
				случайных величин: математическое ожидание, дисперсия, среднеквадратическое отклонение.	
		2.2	Статистические оценки, их свойства, проверка гипотез. Регрессионный анализ и сжатие данных.	Определение статистической оценки параметров распределения. Характеристика свойств оценок: несмещённость, состоятельность, эффективность. Изложение процедуры проверки статистических гипотез: нулевая и альтернативная гипотезы, уровень значимости, р-значение, доверительный интервал. Определение регрессионного анализа как метода исследования зависимостей между переменными. Понятие сжатия данных как процесса уменьшения объёма данных при сохранении информационной ценности.	ЛК, ЛР
		2.3	Задача регрессии. Минимизация квадрата отклонения. Регрессионная функция: условное мат.ожидание	Формальная постановка задачи регрессии. Обоснование минимизации среднего квадратичного отклонения в качестве стандартного критерия качества. Определение регрессионной функции как условного математического ожидания выходной переменной при заданных значениях входных признаков, представляющего теоретически наилучшее предсказание.	ЛК, ЛР
		2.4	Линейная регрессия и метод k ближайших соседей. Переобучение и недообучение	Характеристика линейной регрессии как модели, предполагающей линейную зависимость предсказываемой величины от признаков. Изложение метода наименьших квадратов для оценки коэффициентов. Описание метода k ближайших соседей, основанного на анализе ближайших объектов обучающей выборки. Определение явлений переобучения (избыточная сложность модели, ведущая к запоминанию шумов) и недообучения (недостаточная сложность модели, не позволяющая выявить закономерности).	ЛК
		2.5	Разложение ошибки на шум, смещение и разброс	Анализ структуры общей ошибки предсказания. Определение шума как неустранимой ошибки, обусловленной случайным характером данных. Определение смещения как ошибки, вызванной упрощающими предположениями модели (приводит к недообучению). Определение разброса как ошибки, связанной с высокой чувствительностью модели к вариациям в обучающих данных (приводит к переобучению). Концепция компромисса между смещением и разбросом.	ЛК, ЛР
Раздел 3	Детектирование выбросов и аномалий. Очистка данных	3.1	Детектирование выбросов и аномалий. Что такое выбросы, типы выбросов	Определение выброса (аномалии) как наблюдения, существенно отклоняющегося от основной совокупности	ЛК, ЛР

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
	и технологии регуляризации.			данных. Классификация типов выбросов: точечные (единичные аномальные значения), контекстуальные (аномалии с учётом условий), коллективные (группы аномальных наблюдений). Анализ причин возникновения выбросов: ошибки измерений, ошибки ввода данных, регистрация редких событий.	
		3.2	Методы обнаружения выбросов. Поиск аномалий	Характеристика статистических методов обнаружения выбросов: правило «трёх сигм», межквартильный размах («ящик с усами»). Описание методов, основанных на расстоянии (расстояние Махаланобиса). Характеристика методов, основанных на плотности (локальный фактор выброса). Описание методов машинного обучения для поиска аномалий: одно-классовая классификация, изолирующий лес.	ЛК
		3.3	Цензурирование выборки. Отсев объектов-выбросов, удаление выбросов	Определение цензурирования выборки как процедуры ограничения крайних значений. Изложение процедуры отсева (удаления) аномальных наблюдений. Обоснование критериев для удаления выбросов: наличие ошибки в данных или редкость события. Анализ рисков, связанных с удалением выбросов (потеря значимой информации).	ЛК, ЛР
		3.4	Очистка данных и технологии регуляризации. Основные виды регуляризации	Характеристика процедур очистки данных: заполнение пропусков, удаление дубликатов, исправление несоответствий. Определение регуляризации как метода предотвращения переобучения посредством добавления штрафного слагаемого за сложность модели. Характеристика основных видов регуляризации: L1-регуляризация (лассо), L2-регуляризация (ридж), комбинированный метод Elastic Net.	ЛК, ЛР
		3.5	Метод редукции размерности. Методы отбора признаков	Определение целей снижения размерности: сокращение вычислительных затрат, снижение риска переобучения, визуализация многомерных данных. Характеристика метода главных компонент как линейного преобразования, проектирующего данные на направления максимальной дисперсии. Описание методов отбора признаков: фильтрационные, встроенные, обёрточные.	ЛК
Раздел 4	Технологии кластеризации и классификации. Нейронные сети. Генетические алгоритмы.	4.1	Технологии кластеризации и классификации. K-means. EM-алгоритм	Определение задачи кластеризации. Характеристика метода K-means: итеративное уточнение центров кластеров и распределение объектов. Описание EM-алгоритма (максимизации ожидания) как вероятностного подхода к	ЛК

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
				кластеризации, предполагающего возможность принадлежности объекта нескольким кластерам с различными вероятностями (смесь гауссовых распределений).	
		4.2	Другие методы кластеризации. Задачи классификации. Байесовский классификатор	Характеристика агломеративной (иерархической) кластеризации. Описание метода DBSCAN, основанного на анализе плотности распределения точек. Определение задач классификации. Характеристика байесовского классификатора, основанного на теореме Байеса и предположении о независимости признаков («наивный» байесовский классификатор).	ЛК, ЛР
		4.3	Линейные методы для классификации. Логистическая регрессия, максимизация правдоподобия	Определение линейных методов классификации. Характеристика логистической регрессии как модели, предсказывающей вероятность принадлежности объекта к классу с использованием сигмоидальной функции преобразования. Изложение метода максимизации правдоподобия для обучения логистической регрессии.	ЛК
		4.4	Нейронные сети: общая архитектура. Многослойные сети. Обратное распространение ошибки	Определение искусственного нейрона как вычислительной единицы. Описание архитектуры нейронной сети: входной слой, скрытые слои, выходной слой. Характеристика многослойного перцептрона. Описание алгоритма обратного распространения ошибки как основного метода обучения многослойных нейронных сетей, основанного на вычислении градиента функции ошибки.	ЛК
		4.5	Стохастический градиентный спуск. Генетические алгоритмы	Характеристика градиентного спуска как итерационного метода оптимизации. Описание модификаций градиентного спуска, использующих усреднение градиентов или моментум. Определение генетических алгоритмов как эвристических методов оптимизации, имитирующих механизмы биологической эволюции. Характеристика основных операторов: отбора, скрещивания, мутации.	ЛК, ЛР
Раздел 5	Выделение особенностей (Feature detection); нормализация данных. Нечеткие множества. Байесовы сети.	5.1	Извлечение признаков / выделение особенностей (Feature detection)	Определение извлечения признаков как процесса преобразования исходных данных в набор информативных характеристик. Отличие извлечения признаков от отбора признаков. Примеры методов извлечения признаков: детекторы границ на изображениях, выделение ключевых точек (SIFT, SURF), извлечение n-грамм из текста, создание	ЛК, ЛР

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
				полиномиальных признаков.	
		5.2	Преобразования признаков. Нормализация данных. Методы нормализации данных	Характеристика преобразований признаков: масштабирование, центрирование, логарифмирование. Определение нормализации данных как процедуры приведения различных признаков к сопоставимому масштабу с целью устранения доминирования признаков с большими числовыми значениями.	ЛК
		5.3	Нормализация по методу минимакса. Нормализация по Z-показателю. Десятичное масштабирование	Характеристика нормализации по методу минимакса (линейное преобразование в заданный диапазон, как правило, от 0 до 1 или от -1 до 1). Характеристика нормализации по Z-показателю (стандартизация): вычитание среднего и деление на стандартное отклонение. Характеристика десятичного масштабирования как деления на степень десяти.	ЛК
		5.4	Нечеткие множества. Байесовы сети. Задачи байесовского вывода. Методика построения нечеткой байесовой сети	Определение нечетких множеств как обобщения классической теории множеств, допускающего частичную принадлежность элементов. Характеристика байесовых сетей как вероятностных графических моделей, представляющих условные зависимости между переменными в виде ориентированного ациклического графа. Определение задач байесовского вывода. Описание методики построения нечеткой байесовой сети как комбинации аппарата нечеткой логики и байесовских сетей.	ЛК

* - 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 20 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 «Анализ данных и машинное обучение».

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

РАЗРАБОТЧИК:

Доцент

Должность, БУП

Подпись

Салтыкова Ольга

Александровна

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.