

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Ястребов Олег Александрович
Должность: Ректор
Дата подписания: 02.06.2025 12:22:08
Уникальный программный ключ:
sa953a01204891083f939673078ef1a989dae18a

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

СТАТИСТИЧЕСКИЕ МЕТОДЫ И ПЕРВИЧНЫЙ АНАЛИЗ ДАННЫХ

(наименование дисциплины/модуля)

Рекомендована МССН для направления подготовки/специальности:

**02.03.02 ФУНДАМЕНТАЛЬНАЯ ИНФОРМАТИКА И ИНФОРМАЦИОННЫЕ
ТЕХНОЛОГИИ,
09.03.03 ПРИКЛАДНАЯ ИНФОРМАТИКА**

(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)

2025 г.

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Статистические методы и первичный анализ данных» 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 Кафедра прикладного искусственного интеллекта. Дисциплина состоит из 3 разделов и 45 тем и направлена на изучение формирования у студентов базовых теоретических и практических компетенций в области описательной и вводной статистики, а также анализа и визуализации данных на ранних этапах работы с ними. Курс закладывает фундамент для последующего освоения продвинутых методов машинного обучения, связанных дисциплин аналитики и искусственного интеллекта.

Целью освоения дисциплины является научить студентов основам статистического анализа данных — от понимания структуры и особенностей исходных массивов, через корректный расчет статистик, до их интерпретации и первичной визуализации. Овладение этими инструментами решает задачи первичной разведки данных, выявления отклонений, закономерностей и подготовки данных для дальнейшего глубокого анализа или обучения моделей.

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Статистические методы и первичный анализ 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)
УК-1	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1 Знает принципы сбора, отбора и обобщения информации, методики системного подхода для решения профессиональных задач; УК-1.2 Умеет анализировать и систематизировать разнородные данные, оценивать эффективность процедур анализа проблем и принятия решений в профессиональной деятельности; УК-1.3 Владеет навыками научного поиска и практической работы с информационными источниками; методами принятия решений;
УК-12	Способен: искать нужные источники информации и данные, воспринимать, анализировать, запоминать и передавать информацию с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования полученной информации для решения задач; проводить оценку информации, ее достоверность, строить логические умозаключения на основании поступающих информации и данных	УК-12.1 Способен: искать нужные источники информации и данные, воспринимать, анализировать, запоминать и передавать информацию с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования полученной информации для решения задач; проводить оценку информации, ее достоверность, строить логические умозаключения на основании поступающих информации и данных;

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
ОПК-1	Способен применять в профессиональной деятельности общеинженерные и фундаментальные знания, полученные в области математических и (или) естественных наук, в том числе методы математического анализа и моделирования, теоретического и экспериментального исследования	ОПК-1.1 Знает основные положения и концепции в области математических и естественных наук; знает основную терминологию; ОПК-1.2 Умеет осуществлять первичный сбор и анализ материала, интерпретировать различные математические объекты; ОПК-1.3 Имеет практический опыт работы с решением стандартных математических задач и применяет его в профессиональной деятельности;
ОПК-7	Способен анализировать и разрабатывать организационно-технические и экономические процессы с применением методов системного анализа и математического моделирования	ОПК-7.1 Знает методы системного анализа и математического моделирования;
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 Умеет применять широкий спектр методов предварительной обработки данных, включая сложные методы работы с различными типами данных;

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Статистические методы и первичный анализ 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 «Статистические методы и первичный анализ 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*
УК-12	Способен: искать нужные источники информации и данные, воспринимать, анализировать, запоминать и передавать информацию с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного	Технологическая (проектно-технологическая) практика (учебная); Введение в искусственный интеллект;	Преддипломная практика; Эксплуатационная практика (учебная); Рекомендательные системы**; Информационный поиск**; Программирование на языке NodeJS**; Программирование на языке Go**;

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
	использования полученной информации для решения задач; проводить оценку информации, ее достоверность, строить логические умозаключения на основании поступающих информации и данных		
УК-1	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	Введение в искусственный интеллект; Технологическая (проектно-технологическая) практика (учебная);	Преддипломная практика; <i>Информационный поиск**</i> ;
ОПК-1	Способен применять в профессиональной деятельности общинженерные и фундаментальные знания, полученные в области математических и (или) естественных наук, в том числе методы математического анализа и моделирования, теоретического и экспериментального исследования	Технологическая (проектно-технологическая) практика (учебная); Дискретная математика; Линейная алгебра; Математический анализ;	Эксплуатационная практика (учебная); Технологическая (проектно-технологическая) практика (производственная); Эксплуатационная практика (производственная); Прикладные задачи машинного обучения; Дифференциальные уравнения;
ОПК-7	Способен анализировать и разрабатывать организационно-технические и экономические процессы с применением методов системного анализа и математического моделирования	Программирование на языке Python;	Эксплуатационная практика (учебная); Эксплуатационная практика (производственная); Программирование на языке C++;
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	Технологическая (проектно-технологическая) практика (учебная); Программирование на языке Python;	Эксплуатационная практика (учебная); Преддипломная практика; Технологическая (проектно-технологическая) практика (производственная); Эксплуатационная практика (производственная); <i>Цифровые двойники**</i> ; <i>Информационный поиск**</i> ; <i>Анализ временных рядов**</i> ; <i>Основы больших языковых моделей**</i> ; Введение в базы данных; Нейронные сети; Hadoop, SPARK; Анализ естественного языка с помощью методов искусственного интеллекта; Оптимизация моделей машинного обучения; Лингвистические основы

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
			анализа естественного языка; Практикум по обработке естественного языка (NLP); Основы глубокого обучения; Введение в компьютерное зрение; Обработка и анализ изображений и видео с помощью методов искусственного интеллекта; Проектирование и разработка систем компьютерного зрения; Программирование на языке C++; <i>Программирование на языке NodeJS**;</i> <i>Программирование на языке Go**;</i> <i>Основы робототехники**;</i>

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

** - элективные дисциплины /практики

4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Статистические методы и первичный анализ 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
Контактная работа, ак.ч.	90		90
Лекции (ЛК)	36		36
Лабораторные работы (ЛР)	18		18
Практические/семинарские занятия (СЗ)	36		36
Самостоятельная работа обучающихся, ак.ч.	36		36
Контроль (экзамен/зачет с оценкой), ак.ч.	18		18
Общая трудоемкость дисциплины	ак.ч.	144	144
	зач.ед.	4	4

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	Вводный анализ данных и описательная статистика	1.1	Введение в анализ данных, источники и форматы данных	ЛК
		1.2	Основные шаги разведочного анализа данных (EDA)	ЛК
		1.3	Типы данных: количественные и категориальные	ЛК
		1.4	Меры центральной тенденции (среднее, медиана, мода)	ЛК
		1.5	Меры разброса (дисперсия, стандартное отклонение, диапазон, квартиль)	ЛК
		1.6	Графическое представление одномерных данных: гистограмма, boxplot	ЛК
		1.7	Импорт и очистка различных форматов данных (csv, xlsx)	ЛР
		1.8	Расчёт базовых статистик (среднее, медиана, мода, минимум, максимум)	ЛР
		1.9	Формирование и визуализация гистограмм, boxplot для реальных данных	ЛР
		1.10	Выбор методов для анализа разных типов данных	СЗ
		1.11	Цикл EDA: формулировка вопросов и первоначальный разбор данных	СЗ
		1.12	Практика группировки и категоризации	СЗ
		1.13	Решение задач на вычисление и интерпретацию описательных статистик	СЗ
		1.14	Разбор примеров ошибок анализа из-за некорректного расчёта базовых статистик	СЗ
		1.15	Обсуждение роли визуализации данных на этапе разведки	СЗ
Раздел 2	Одномерные и двумерные статистические методы и их применение	2.1	Корреляция: понятие, коэффициенты, ограничения	ЛК
		2.2	Основы простейшей линейной регрессии	ЛК
		2.3	Проверка гипотез о средних значениях (t-тест)	ЛК
		2.4	Введение в доверительные интервалы	ЛК
		2.5	Проблема выбросов и способы работы с ними	ЛК
		2.6	Критерии независимости (хи-квадрат, Фишера)	ЛК
		2.7	Анализ корреляции между двумя числовыми переменными	ЛР
		2.8	Проведение t-теста на реальных выборках	ЛР
		2.9	Поиск и визуализация выбросов в данных	ЛР
		2.10	Практика постановки и проверки гипотез	СЗ
		2.11	Построение и интерпретация доверительных интервалов	
		2.12	Разбор типичных примеров корреляции и причинно-следственной связи	СЗ
		2.13	Практика работы с выбросами по реальным данным	СЗ
		2.14	Сравнительный анализ результатов t-теста и других проверок	СЗ
		2.15	Кейсы применения критериев независимости в различных областях	СЗ
Раздел 3	Введение в многомерный анализ, визуализация и статистика в прикладных задачах	3.1	Первичный анализ многомерных данных: парные диаграммы, матрицы корреляций	ЛК
		3.2	Описательные методы выделения групп (кластеризация, без детализации алгоритмов)	ЛК

Номер раздела	Наименование раздела дисциплины	Содержание раздела (темы)		Вид учебной работы*
		3.3	Проблема мультиколлинеарности, прямая/обратная связь признаков	ЛК
		3.4	Основы визуализации: pairplot, тепловая карта, scatter matrix	ЛК
		3.5	Особенности первичной подготовки данных: нормализация, стандартизация	ЛК
		3.6	Интерпретация статистических результатов в реальных задачах (бизнес, наука, технологии)	ЛК
		3.7	Визуализация многомерных данных (pairplot, тепловые карты)	ЛР
		3.8	Первичный кластерный анализ с помощью простых инструментов	ЛР
		3.9	Практика нормализации и стандартизации данных	ЛР
		3.10	Построение комплексной EDA-структуры для заданного проекта	СЗ
		3.11	Кейсы некорректной интерпретации статистики в реальных проектах	СЗ
		3.12	Работа с мультиколлинеарностью и проблемами многомерности	СЗ
		3.13	Групповая разработка схемы подготовки данных для ML	СЗ
		3.14	Дискуссия: грамотная визуализация как часть отчётности	СЗ
		3.15	Обсуждение типичных ошибок в анализе данных и подведение итогов курса	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)
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 25 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. Кулаичев, А. П. Методы и средства комплексного статистического анализа данных: учебное пособие / А.П. Кулаичев. — 5-е изд., перераб. и доп. — Москва: ИНФРА-М, 2025. — 484 с. — (Высшее образование). — DOI 10.12737/25093. - ISBN 978-5-16-020053-8. - Текст: электронный. - URL: <https://znanium.ru/catalog/product/2155997>

2. Григорьев, А. А. Методы и алгоритмы обработки данных: учебное пособие / А.А. Григорьев, Е.А. Исаев. — 2-е изд., перераб. и доп. — Москва: ИНФРА-М, 2024. — 383 с. + Доп. материалы [Электронный ресурс]. — (Высшее образование: Бакалавриат). — DOI 10.12737/1032305. - ISBN 978-5-16-015581-4. - Текст: электронный. - URL: <https://znanium.ru/catalog/product/2084190>

Дополнительная литература:

1. Мыльников Л. А. Статистические методы интеллектуального анализа данных. — СПб.: БХВ-Петербург, 2021. — 240 с.: ил. ISBN 978-5-9775-6733-6

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

- ЭБС «Знаниум» <https://znanium.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 «Статистические методы и первичный анализ данных».

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

РАЗРАБОТЧИК:

Заведующий кафедрой
прикладного искусственного
интеллекта

Должность, БУП

Подпись

Подолько Павел
Михайлович

Фамилия И.О.

РУКОВОДИТЕЛЬ БУП:

Заведующий кафедрой
прикладного искусственного
интеллекта

Должность, БУП

Подпись

Подолько Павел
Михайлович

Фамилия И.О.

РУКОВОДИТЕЛЬ ОП ВО:

Заведующий кафедрой
прикладного искусственного
интеллекта

Должность, БУП

Подпись

Подолько Павел
Михайлович

Фамилия И.О.