

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Ястребов Олег Александрович
Должность: Ректор
Дата подписания: 22.05.2026 10:38:35
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

ПРИКЛАДНАЯ СТАТИСТИКА И АНАЛИЗ 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 «Прикладная статистика и анализ данных» входит в программу магистратуры «Управление данными и искусственный интеллект» по направлению 02.04.02 «Фундаментальная информатика и информационные технологии» и изучается в 1 семестре 1 курса. Дисциплину реализует Кафедра прикладного искусственного интеллекта. Дисциплина состоит из 3 разделов и 18 тем и направлена на изучение формирования у студентов компетенций по глубокому анализу и интерпретации сложных данных, освоение методов работы с многомерными и высокоразмерными выборками, а также на развитие навыков построения статистических моделей для отраслевых, исследовательских и инженерных задач в искусственном интеллекте. Дисциплина фокусируется на современных, практико-ориентированных аспектах статистики, которые выходят за рамки классических основ бакалавриата и требуют владения сложными структурами данных и подходами к анализу.

Целью освоения дисциплины является научить студентов выбирать и применять современные методы прикладной статистики и анализа данных для решения исследовательских и практических задач в ИИ, интерпретировать результаты в условиях неопределённости, корректно использовать выводы для принятия решений и построения моделей, грамотно критиковать и защищать полученные выводы.

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Прикладная статистика и 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 проблемных ситуаций на основе системного подхода, выработать стратегию действий	УК-1.1 Знает принципы сбора, отбора и обобщения информации; УК-1.2 Умеет соотносить разнородные явления и систематизировать их в рамках избранных видов профессиональной деятельности; УК-1.3 Имеет практический опыт работы с информационными источниками, опыт научного поиска, создания научных текстов.;
УК-2	Способен управлять проектом на всех этапах его жизненного цикла	УК-2.2 Умеет определять круг задач в рамках избранных видов профессиональной деятельности, планировать собственную деятельность исходя из имеющихся ресурсов; соотносить главное и второстепенное, решать поставленные задачи в рамках избранных видов профессиональной деятельности; УК-2.3 Имеет практический опыт применения нормативной базы и решения задач в области избранных видов профессиональной деятельности;
ОПК-1	Способен находить, формулировать и решать актуальные проблемы прикладной математики, фундаментальной информатики и информационных технологий	ОПК-1.1 Обладает фундаментальными знаниями в области математических и естественных наук, теории коммуникаций; ОПК-1.2 Умеет осуществлять первичный сбор и анализ материала, интерпретировать различные математические объекты; ОПК-1.3 Имеет практический опыт работы с решением математических задач и применяет его в

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
		профессиональной деятельности;
ОПК-3	Способен проводить анализ математических моделей, создавать инновационные методы решения прикладных задач профессиональной деятельности в области информатики и математического моделирования	ОПК-3.1 Знает методы теории алгоритмов, методы системного и прикладного программирования, основные положения и концепции в области математических, информационных и имитационных моделей;

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Прикладная статистика и 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 «Прикладная статистика и 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-2	Способен управлять проектом на всех этапах его жизненного цикла		Технологическая (проектно-технологическая) практика (учебная); Технологическая (проектно-технологическая) практика (производственная); Обучение с подкреплением; Управление проектами в сфере искусственного интеллекта; Методы оптимизации;
УК-1	Способен осуществлять поиск, критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий		Преддипломная практика; Научно-исследовательская работа (учебная); Научно-исследовательская работа (производственная); Глубокое обучение для NLP; Искусственный интеллект по отраслям**; Вайб-кодинг**;
ОПК-1	Способен находить, формулировать и решать актуальные проблемы прикладной математики, фундаментальной информатики и информационных технологий		Обучение с подкреплением; Глубокое обучение для NLP; Машинное обучение на больших данных;
ОПК-3	Способен проводить анализ математических моделей, создавать инновационные методы решения прикладных задач профессиональной		Глубокое обучение для NLP; Методы оптимизации; Глубокое обучение в компьютерном зрении;

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
	деятельности в области информатики и математического моделирования		

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

** - элективные дисциплины /практики

4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Прикладная статистика и 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)
			1
Контактная работа, ак.ч	51		51
Лекции (ЛК)	17		17
Лабораторные работы (ЛР)	0		0
Практические/семинарские занятия (СЗ)	34		34
Самостоятельная работа обучающихся, ак.ч.	66		66
Контроль (экзамен/зачет с оценкой), ак.ч.	27		27
Общая трудоемкость дисциплины ак.ч.	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	Наименование темы		Содержание темы	Вид учебной работы*
Раздел 1	Современные методы описательной статистики и разведочного анализа сложных данных	1.1	Описательные и визуальные методы анализа многомерных данных	Изучение базовых статистических показателей (среднее, медиана, мода, дисперсия, стандартное отклонение) для данных с множеством переменных. Построение графиков, диаграмм рассеяния, тепловых карт корреляций, box-plot диаграмм для визуализации взаимосвязей между переменными.	ЛК
		1.2	Исследование распределений и аномалий в больших и несимметричных выборках	Анализ форм распределений данных (нормальное, экспоненциальное, биномиальное), работа с асимметрией и эксцессом. Методы выявления выбросов и аномальных значений в больших объёмах данных с использованием статистических критериев.	ЛК
		1.3	Методы очистки и препроцессинга данных в прикладных проектах	Техники обработки пропущенных значений (удаление, импутация), нормализация и стандартизация данных, кодирование категориальных переменных, работа с дубликатами. Практические проекты по подготовке реальных датасетов.	ЛК
		1.4	Визуализация сложных и мультииндустриальных данных (heatmaps, pairplot, PCA-проекция)	Продвинутое методы визуализации для данных из различных отраслей (финансы, медицина, маркетинг). Построение интерактивных дашбордов, многомерных визуализаций, временных рядов и геопространственных данных.	СЗ
		1.5	Групповой анализ стратегий поиска и классификации аномалий (outliers, missing data)	Коллективная работа над выявлением паттернов аномального поведения в данных. Сравнение различных подходов: статистические методы, методы машинного обучения (Isolation Forest, LOF), их эффективность в разных сценариях.	СЗ
		1.6	Разбор реальных кейсов очистки и трансформации больших датасетов	Практический анализ конкретных примеров работы с крупными массивами данных: от сырых данных до готовых к анализу структур. Обсуждение проблем и решений, встречающихся в промышленных проектах.	СЗ
Раздел 2	Статистическое моделирование, гипотезы и интерпретация результатов	2.1	Многомерное статистическое моделирование: регрессия, ANOVA, современные критерии	Построение множественных регрессионных моделей для предсказания количественных переменных на основе нескольких предикторов. Дисперсионный анализ (ANOVA) для сравнения средних значений между несколькими группами. Современные критерии оценки качества моделей: AIC, BIC для выбора оптимальной сложности модели. Проверка предпосылок регрессионного анализа: линейность, гомоскедастичность, независимость остатков.	ЛК
		2.2	Проверка сложных гипотез на больших данных, коррекция ошибок, мультиколлинеарность	Множественное тестирование гипотез и проблема накопления ошибок первого рода. Методы коррекции: поправка Бонферрони, метод Холма, FDR (False Discovery Rate). Выявление мультиколлинеарности между предикторами с помощью VIF (Variance Inflation Factor), матриц корреляций. Влияние мультиколлинеарности на стабильность оценок коэффициентов и способы её устранения.	ЛК

		2.3	Байесовские подходы и вероятностные выводы в бизнесе и индустрии ИИ	Применение байесовского подхода для обновления вероятностей с получением новых данных. Априорные распределения на основе экспертных знаний и предыдущего опыта, апостериорные выводы. Байесовская оптимизация гиперпараметров моделей машинного обучения. Практическое применение: A/B тестирование с байесовским подходом, оценка рисков в финансах, персонализация рекомендаций.	ЛК
		2.4	Решение задач по	Практикум по созданию регрессионных моделей на реальных данных. Интерпретация	СЗ

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
			построению и интерпретации моделей множественной регрессии, выявление влияющих факторов	коэффициентов регрессии: понимание величины и направления влияния каждого фактора. Стандартизованные коэффициенты для сравнения важности предикторов в разных единицах измерения. Выявление наиболее значимых факторов, влияющих на целевую переменную. Построение доверительных интервалов для коэффициентов.	
		2.5	Обсуждение типовых ошибок мультиколлинеарности и их устранения (VIF, регуляризация)	Детальный разбор проблем, возникающих при наличии сильно коррелированных предикторов: неустойчивость коэффициентов, завышенные стандартные ошибки, неверные выводы о значимости. Диагностика с помощью VIF, анализ собственных значений матрицы. Методы решения: удаление коррелирующих переменных, объединение в композитные индексы, методы регуляризации (Ridge, Lasso, Elastic Net), которые штрафуют большие коэффициенты.	СЗ
		2.6	Байесовские выводы на реальных данных (бизнес/медицина/индустрия)	Практические кейсы применения байесовского подхода в разных областях. Бизнес: оценка эффективности маркетинговых кампаний с учетом неопределенности, прогнозирование спроса. Медицина: оценка эффективности лечения, диагностические тесты с учетом априорных вероятностей заболеваний. Промышленность: предиктивное обслуживание оборудования, контроль качества продукции. Построение байесовских моделей и интерпретация результатов для принятия решений.	СЗ
Раздел 3	Авангардные подходы прикладной статистики в анализе данных для ИИ	3.1	Методы статистического вывода в задачах с высокоразмерными и неструктурированными данными	Работа с данными, где количество признаков сопоставимо или превышает количество наблюдений (проблема "большой размерности"). Особенности статистического вывода в таких условиях: нестабильность оценок, переобучение. Методы снижения размерности (PCA, t-SNE, UMAP) перед применением статистических тестов. Работа с неструктурированными данными: тексты, изображения, временные ряды — извлечение статистических признаков для дальнейшего анализа.	ЛК
		3.2	Современные подходы к оценке доверия в моделях (bootstrap, permutation, cross-validation)	Bootstrap-методы для оценки распределения статистик и построения доверительных интервалов без предположений о нормальности данных. Permutation tests для проверки гипотез путем случайной перестановки меток. Cross-validation (кросс-валидация) для оценки обобщающей способности моделей: k-fold, stratified, leave-one-out. Сравнение моделей с использованием статистических тестов на основе результатов кросс-валидации.	ЛК
		3.3	Статистические проблемы интерпретации выводов в моделях ИИ (feature importance, SHAP, causal inference)	Проблема "черного ящика" в сложных моделях машинного обучения. Методы оценки важности признаков: permutation importance, gain/split importance в деревьях. SHAP (SHapley Additive exPlanations) для объяснения предсказаний отдельных наблюдений и глобальной важности признаков. Различие между корреляцией и причинно-следственной связью. Введение в каузальный вывод (causal inference): выявление истинных причинных связей, а не просто статистических ассоциаций.	ЛК

		3.4	<p>Применение bootstrapping и permutation tests к задачам построения доверительных интервалов и диагностики моделей</p>	<p>Практическое применение bootstrap для оценки неопределенности в метриках качества моделей (accuracy, RMSE и др.), построения доверительных интервалов для коэффициентов регрессии в нестандартных ситуациях. Permutation tests для проверки значимости улучшения одной модели над другой, для тестирования значимости признаков. Диагностика моделей: оценка стабильности предсказаний, выявление влиятельных наблюдений через bootstrap.</p>	С3
		3.5	<p>Анализ методов оценки важности признаков и</p>	<p>Глубокое изучение различных подходов к интерпретируемости моделей. Feature importance в различных алгоритмах (случайный лес, gradient boosting). LIME (Local</p>	С3

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
			причинно-следственных связей (interpretability, causal impact)	Interpretable Model-agnostic Explanations) для локальной интерпретации. Переход от интерпретируемости к каузальности: методы оценки причинного влияния (Propensity Score Matching, Difference-in-Differences, Instrumental Variables). Causal Impact анализ для оценки эффекта вмешательств на временных рядах.	
		3.6	Кейсы критического разбора публикаций: статистические ловушки, неверные интерпретации, reproducibility crisis	Анализ реальных научных публикаций и бизнес-отчетов с точки зрения корректности статистических выводов. Типичные ошибки: p-hacking (подгонка результатов под значимость), HARKing (выдвижение гипотез после анализа данных), игнорирование множественного тестирования, переоценка силы эффекта. Кризис воспроизводимости в науке: причины и пути решения. Практика критического чтения статистических отчетов, выявление манипуляций с данными и некорректных выводов. Принципы открытой науки и воспроизводимых исследований.	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.	
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. Анализ данных: учебник для вузов / под редакцией В. С. Мхитаряна. — Москва: Издательство Юрайт, 2024. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст: электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/536007>

2. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва: Издательство Юрайт, 2024. — 448 с. — (Высшее образование). ISBN 978-5-534-19964-2

Дополнительная литература:

1. Криволапов, С. Я. Анализ данных. Методы теории вероятностей и математической статистики на языке Python: учебное пособие / С.Я. Криволапов. — Москва: ИНФРА-М, 2025. — 678 с. — (Высшее образование). — DOI 10.12737/2034420. - ISBN 978-5-16-018616-0. - Текст: электронный. - URL: <https://znanium.ru/catalog/product/2034420>

2. Солодушкин, С. И. Прикладная статистика: учебное пособие / С. И. Солодушкин, И. Ф. Юманова. — 2-е изд. — Москва: ФЛИНТА, 2024. — 100 с. — ISBN 978-5-9765-5410-8. — Текст: электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/370718>

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 «Прикладная статистика и анализ 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