

МИНОБРНАУКИ РОССИИ

**Федеральное государственное автономное образовательное
учреждение высшего образования "Пермский
государственный национальный исследовательский
университет"**

Кафедра информационных систем и математических методов в экономике

**Авторы-составители: Назаров Дмитрий Михайлович
Радионова Марина Владимировна
Русакова Ольга Леонидовна**

Рабочая программа дисциплины
ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И АНАЛИЗ ДАННЫХ
Код УМК 98795

Утверждено
Протокол №9
от «24» мая 2023 г.

Пермь, 2023

1. Наименование дисциплины

Искусственный интеллект и анализ данных

2. Место дисциплины в структуре образовательной программы

Дисциплина входит в Блок « ОК.В.00 » образовательной программы по научным специальностям:

Научная специальность: **5.9.8** Теоретическая, прикладная и сравнительно-сопоставительная лингв

3. Планируемые результаты обучения по дисциплине

В результате освоения дисциплины **Искусственный интеллект и анализ данных** у обучающегося должны быть сформированы следующие планируемые результаты обучения:

5.9.8 Теоретическая, прикладная и сравнительно-сопоставительная лингвистика

ИРО.1 Проводит анализ и теоретическое обобщение научных данных в соответствии с задачами исследования

ИРО.2 Проводит анализ результатов экспериментов и наблюдений

4. Объем и содержание дисциплины

Научная специальность	5.9.8 Теоретическая, прикладная и сравнительно-сопоставительная лингвистика
форма обучения	очная
№№ триместров, выделенных для изучения дисциплины	5,6
Объем дисциплины (з.е.)	6
Объем дисциплины (ак.час.)	216
Контактная работа с преподавателем (ак.час.), в том числе:	72
Проведение лабораторных работ, занятий по иностранному языку	72
Самостоятельная работа (ак.час.)	144
Формы промежуточной аттестации	Экзамен (5 триместр) Экзамен (6 триместр)

5. Аннотированное описание содержания разделов и тем дисциплины

Тема 1. Основы анализа данных и проверка статистических гипотез

Содержание, цель и значение дисциплины в подготовке аспирантов, ее связь с другими дисциплинами и подготовкой кандидатской диссертации. Общая классификация решаемых задач.

Тема 2. Задачи классификации и кластеризации

Общая постановка задачи классификации, классификации без обучения, понятия кластерного анализа. Метрики расстояния и близости между объектами, расстояния между кластерами. Основные типы задач и алгоритмов кластерного анализа. Иерархические, параллельные и последовательные процедуры кластерного анализа. Метод k-средних. Визуализация.

Тема 3. Методы корреляционного и регрессионного анализа

Основные показатели исследования взаимосвязей. Коэффициенты корреляции Пирсона и Спирмена. Таблицы сопряженности и их интерпретация. Предпосылки линейной регрессии. Построение уравнения простой и множественной линейной регрессии. Оценка качества построенной модели. Модели бинарного выбора.

Тема 4 Основные технологии искусственного интеллекта для анализа данных: машинное обучение и нейронные сети

Введение в искусственные нейронные сети. Понятие искусственного нейрона. Обучение базового нейрона. Обучение нейронных сетей. Переобучение (overfitting). Использование нейронных сетей в качестве классификатора. Компьютерное зрение. Обработка естественного языка. Проводится серия лабораторных работ, демонстрирующая возможности методов искусственного интеллекта для анализа научных данных, в частности, для извлечения знаний. Приводятся примеры выполнения проектов анализа данных и извлечения знаний, относящиеся к различным предметным областям: промышленности, медицине, социологии, политологии, спорту и др. Основное внимание уделяется проектам, выполненным учеными и студентами ПГНИУ.

Тема 5 Программные инструменты искусственного интеллекта для анализа данных

Проводится серия лабораторных работ, в ходе которых осваиваются и приобретаются навыки работы с программными инструментами искусственного интеллекта. Это библиотеки Python, а также пермский инструмент для генерации, обучения и применения нейросетей «Нейросимулятор-5.0» для анализа данных и извлечения знаний.

Тема 6 Проекты создания и применения интеллектуальных систем для анализа данных по теме диссертаций

Каждый аспирант выбирает тему самостоятельной работы, являющуюся проектом создания системы искусственного интеллекта и ее применения для анализа данных и извлечения знаний в какой-либо предметной области. Проекты, по возможности, должны быть связаны (или даже быть частью) с темой диссертационной работы аспиранта. Основное внимание уделяется интеллектуальным системам, основанным на нейронных сетях, как на наиболее эффективной технологии машинного обучения и способа анализа данных и извлечения из них знаний. Индивидуальные проекты, выполняемые аспирантами, заканчиваются созданием действующей интеллектуальной системы и ее применением для анализа данных и извлечения знаний в выбранной им предметной области: промышленности, экономике, политологии, социологии, лингвистике, медицине, педагогике, криминалистике, спорте и т.д.

6. Методические указания для обучающихся по освоению дисциплины

Освоение дисциплины требует систематического изучения всех тем в той последовательности, в какой они указаны в рабочей программе.

Основными видами учебной работы являются аудиторские занятия. Их цель - расширить базовые знания обучающихся по осваиваемой дисциплине и систему теоретических ориентиров для последующего более глубокого освоения программного материала в ходе самостоятельной работы. Обучающемуся важно помнить, что контактная работа с преподавателем эффективно помогает ему овладеть программным материалом благодаря расстановке необходимых акцентов и удержанию внимания интонационными модуляциями голоса, а также подключением аудио-визуального механизма восприятия информации.

Самостоятельная работа преследует следующие цели:

- закрепление и совершенствование теоретических знаний, полученных на лекционных занятиях;
- формирование навыков подготовки текстовой составляющей информации учебного и научного назначения для размещения в различных информационных системах;
- совершенствование навыков поиска научных публикаций и образовательных ресурсов, размещенных в сети Интернет;
- самоконтроль освоения программного материала.

Обучающемуся необходимо помнить, что результаты самостоятельной работы контролируются преподавателем во время проведения мероприятий текущего контроля и учитываются при промежуточной аттестации.

Обучающимся с ОВЗ и инвалидов предоставляется возможность выбора форм проведения мероприятий текущего контроля, альтернативных формам, предусмотренным рабочей программой дисциплины. Предусматривается возможность увеличения в пределах 1 академического часа времени, отводимого на выполнение контрольных мероприятий.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации.

При проведении текущего контроля применяются оценочные средства, обеспечивающие передачу информации, от обучающегося к преподавателю, с учетом психофизиологических особенностей здоровья обучающихся.

7. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине

При самостоятельной работе обучающимся следует использовать:

- конспекты лекций;
- литературу из перечня основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля);
- текст лекций на электронных носителях;
- ресурсы информационно-телекоммуникационной сети "Интернет", необходимые для освоения дисциплины;
- лицензионное и свободно распространяемое программное обеспечение из перечня информационных технологий, используемых при осуществлении образовательного процесса по дисциплине;
- методические указания для обучающихся по освоению дисциплины.

8. Перечень основной и дополнительной учебной литературы

Основная:

1. Маккинли, Уэс Python и анализ данных / Уэс Маккинли ; перевод А. Слинкина. — 2-е изд. — Саратов : Профобразование, 2019. — 482 с. — ISBN 978-5-4488-0046-7. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. <http://www.iprbookshop.ru/88752.html>

Дополнительная:

1. Маркус, Г. Искусственный интеллект: перезагрузка: как создать машинный разум, которому действительно можно доверять / Г. Маркус, Э. Дэвис ; перевод В. Скворцов ; под редакцией А. Марченковой. — Москва : Альпина ПРО, 2022. — 300 с. — ISBN 978-5-907394-93-3. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. <https://www.iprbookshop.ru/122525>

2. Пальмов, С. В. Интеллектуальный анализ данных : учебное пособие / С. В. Пальмов. — Самара : Поволжский государственный университет телекоммуникаций и информатики, 2017. — 127 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. <http://www.iprbookshop.ru/75376.html>

3. Орешков, В. И. Интеллектуальный анализ данных : учебное пособие / В. И. Орешков. — Рязань : Рязанский государственный радиотехнический университет, 2017. — 161 с. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. <https://www.iprbookshop.ru/121842>

9. Перечень ресурсов сети Интернет, необходимых для освоения дисциплины

<https://intuit.ru/studies/courses/13844/1241/info> Школа анализа данных (Яндекс): Машинное обучение

<https://intuit.ru/studies/courses/3521/763/info> Нейросетевые технологии искусственного интеллекта

<https://intuit.ru/studies/courses/1153/318/info> Статистические методы анализа данных

<https://learn.microsoft.com/ru-ru/training/modules/explore-analyze-data-with-python/> Изучение и анализ данных с помощью Python

<https://learn.microsoft.com/ru-ru/training/modules/introduction-to-ai-technology/> Общие сведения о технологиях искусственного интеллекта

<https://learn.microsoft.com/ru-ru/training/career-paths/ai-engineer> Обучение инженеров ИИ

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине

Образовательный процесс по дисциплине **Искусственный интеллект и анализ данных** предполагает использование следующего программного обеспечения и информационных справочных систем:

- доступ в режиме on-line в Электронную библиотечную систему (ЭБС);
- доступ в электронную информационно-образовательную среду университета.
- IDE для Python версии 3 и выше.

При освоении материала и выполнении заданий по дисциплине рекомендуется использование материалов, размещенных в Личных кабинетах обучающихся ЕТИС ПГНИУ (student.psu.ru).

При организации дистанционной работы и проведении занятий в режиме онлайн могут использоваться:

система видеоконференцсвязи на основе платформы BigBlueButton (<https://bigbluebutton.org/>).

система LMS Moodle (<http://e-learn.psu.ru/>), которая поддерживает возможность использования текстовых материалов и презентаций, аудио- и видеоконтент, а так же тесты, проверяемые задания, задания для совместной работы.

система тестирования Indigo (<https://indigotech.ru/>).

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Для лекционных занятий требуется аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук) с соответствующим программным обеспечением и маркерной доской.

Для проведения лабораторных занятий - компьютерный класс. Состав оборудования определен в Паспорте компьютерного класса (требуется доступ в Интернет и должна быть установлена IDE для Python).

Для групповых (индивидуальных) консультаций - аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук) с соответствующим программным обеспечением и маркерной доской.

Для проведения текущего контроля - компьютерный класс. Состав оборудования определен в Паспорте компьютерного класса (требуется доступ в Интернет и должна быть установлена IDE для Python).

Самостоятельная работа студентов: аудитория, оснащенная компьютерной техникой с возможностью

подключения к сети «Интернет», с обеспеченным доступом в электронную информационно-образовательную среду университета, помещения Научной библиотеки ПГНИУ.

Помещения научной библиотеки ПГНИУ для обеспечения самостоятельной работы обучающихся:

1. Научно-библиографический отдел, корп.1, ауд. 142. Оборудован 3 персональными компьютера с доступом к локальной и глобальной компьютерным сетям.

2. Читальный зал гуманитарной литературы, корп. 2, ауд. 418. Оборудован 7 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

3. Читальный зал естественной литературы, корп.6, ауд. 107а. Оборудован 5 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

4. Отдел иностранной литературы, корп.2 ауд. 207. Оборудован 1 персональным компьютером с доступом к локальной и глобальной компьютерным сетям.

5. Библиотека юридического факультета, корп.9, ауд. 4. Оборудована 11 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

6. Читальный зал географического факультета, корп.8, ауд. 419. Оборудован 6 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

Все компьютеры, установленные в помещениях научной библиотеки, оснащены следующим программным обеспечением:

Операционная система ALT Linux;

Офисный пакет Libreoffice.

Справочно-правовая система «КонсультантПлюс»

**Фонды оценочных средств для аттестации по дисциплине
Искусственный интеллект и анализ данных**

Планируемые результаты обучения по дисциплине и критерии их оценивания

Планируемый результат обучения	Знания, умения и навыки	Критерии оценивания результатов обучения
<p>ИРО.1 Проводит анализ и теоретическое обобщение научных данных в соответствии с задачами исследования</p>	<p>Знает методы интеллектуального анализа данных, теоретического обобщения Проводит анализ и теоретическое обобщение научных данных в соответствии с задачами исследования</p>	<p align="center">Неудовлетворител</p> <p>Не знает понятия и методы интеллектуального анализа данных, не умеет применять эти методы для анализа и обобщения научных данных в соответствии с задачами исследования. Допускает грубые ошибки при анализе данных. Не владеет навыками теоретического обобщения научных данных.</p> <p align="center">Удовлетворительн</p> <p>Удовлетворительно знает понятия и методы интеллектуального анализа данных, умеет на удовлетворительном уровне применять эти методы для анализа и обобщения научных данных в соответствии с задачами исследования. Допускает ошибки при анализе данных. Слабо владеет навыками теоретического обобщения научных данных.</p> <p align="center">Хорошо</p> <p>Знает понятия и методы интеллектуального анализа данных. Умеет применять эти методы для анализа и обобщения научных данных в соответствии с задачами исследования . Допускает незначительные ошибки при анализе данных. Владеет навыками теоретического обобщения научных данных.</p> <p align="center">Отлично</p> <p>Отлично знает понятия и методы интеллектуального анализа данных. Умеет анализировать данные. Показывает хорошее владение навыками теоретического обобщения научных данных.</p>

Планируемый результат обучения	Знания, умения и навыки	Критерии оценивания результатов обучения
<p>ИРО.2 Проводит анализ результатов экспериментов и наблюдений</p>	<p>Владеет навыками проведения анализа результатов экспериментов и наблюдений.</p>	<p>Неудовлетворител Не знает методы анализа результатов экспериментов и наблюдений. Допускает грубые ошибки при анализе результатов. Не владеет навыками анализа результатов и наблюдений.</p> <p>Удовлетворительн Удовлетворительно знает методы анализа результатов экспериментов и наблюдений. Допускает ошибки при анализе результатов. Слабо владеет навыками анализа результатов и наблюдений.</p> <p>Хорошо Знает методы анализа результатов экспериментов и наблюдений. Допускает незначительные ошибки при анализе результатов. Владеет навыками анализа результатов и наблюдений.</p> <p>Отлично Отлично знает методы анализа результатов экспериментов и наблюдений. Умеет выполнять анализ результатов. Демонстрирует хорошее владение навыками анализа результатов и наблюдений</p>

Оценочные средства

Схема доставки : Базовая

Вид мероприятия промежуточной аттестации : Экзамен

Способ проведения мероприятия промежуточной аттестации : Защищаемое контрольное мероприятие

Продолжительность проведения мероприятия промежуточной аттестации : время отводимое на доклад 60

Показатели оценивания

<ul style="list-style-type: none"> - Не демонстрирует знание основного содержания дисциплины; - Не владеет основными понятиями, законами и теорией, необходимыми для объяснения явлений, закономерностей и т.д.; – не умеет выполнять типовые задания и задачи предусмотренные 	<p>Неудовлетворител</p>
---	--------------------------------

программой	Неудовлетворител
<ul style="list-style-type: none"> - Демонстрирует знание основного содержания дисциплины и его элементов в соответствии с прослушанным лекционным курсом; - Владение основными понятиями, законами и теорией, необходимыми для объяснения явлений, закономерностей и т.д.; –показывает умение выполнять типовые задания и задачи предусмотренные программой; – выполняет расчеты с ошибками 	Удовлетворительн
<ul style="list-style-type: none"> - ответ по вопросу или заданию аргументированный, демонстрирующий знание основного содержания дисциплины и его элементов в соответствии с прослушанным лекционным курсом и с учебной литературой; - демонстрирует понимание материала, приводит примеры; - Владение основными понятиями, законами и теорией, необходимыми для объяснения явлений, закономерностей и т.д.; –показывает владение методологией дисциплины, умение выполнять типовые задания и задачи предусмотренные программой; – выполняет расчеты с ошибками 	Хорошо
<ul style="list-style-type: none"> - ответ по вопросу или заданию аргументированный, логически выстроенный, полный, демонстрирующий знание основного содержания дисциплины и его элементов в соответствии с прослушанным лекционным курсом и с учебной литературой; - демонстрирует полное понимание материала, выводы доказательны, приводит примеры; - свободное владение основными понятиями, законами и теорией, необходимыми для объяснения явлений, закономерностей и т.д.; – - показывает владение методологией дисциплины, умение выполнять типовые задания и задачи предусмотренные программой; – выполняет расчеты без ошибок; - демонстрирует способность творчески применять знание теории к решению профессиональных практических задач 	Отлично

Перечень вопросов для проведения промежуточной аттестации

1. Инструменты описательной статистики. Измерение центра распределения. Измерение разброса данных.
2. Визуализация качественных признаков. Сводные таблицы и сводные диаграммы. Таблицы сопряженности и парадокс Симпсона. Иерархия признаков.
3. Предварительная обработка данных. Выбросы и их обработка. Пропущенные значения и их обработка.
4. Проверка статистических гипотез.
5. Общая постановка задачи машинного обучения.
6. Обучение с учителем и без учителя, обучение с подкреплением.
7. Задачи классификации, восстановления регрессии, предсказания.
8. Модели алгоритмов. Признаки. Типы признаков. Понятие функционала качества.
9. Вероятностная постановка задачи. Оценка обобщающей способности.

10. Примеры практических задач машинного обучения
11. Обобщенный метрический классификатор. Виды метрик.
12. Метод ближайшего соседа. Алгоритм k-ближайших соседей.
13. Взвешенная версия алгоритма k-ближайших соседей.
14. Классификация, деревья решений и метод ближайших соседей
15. Постановка задач линейной регрессии и линейной классификации.
16. Метод наименьших квадратов в матричной форме. Аналитическое решение.
17. Мультиколлинеарность. Гребневая регрессия. Метод лассо.
18. Линейные классификаторы.
19. Логистическая регрессия.
20. Метод опорных векторов.
21. Линейные модели классификации и регрессии
22. Логистическая регрессия и случайный лес.
23. Линейная регрессия, Lasso и RF-регрессия. Построение и отбор признаков.
24. Методы кластеризации. Типы кластерных структур.
25. Функционал качества кластеризации. EM-алгоритм.
26. Метод k-средних. Иерархическая кластеризация.
27. Обучение без учителя. Метод главных компонент. Кластеризация
28. Проблема переобучения. Критерии оценки качества работы алгоритмов машинного обучения. ROC-кривые.
29. Основы анализа временных рядов

Вид мероприятия промежуточной аттестации : Экзамен

Способ проведения мероприятия промежуточной аттестации : Защищаемое контрольное мероприятие

Продолжительность проведения мероприятия промежуточной аттестации :
время отводимое на доклад 60

Показатели оценивания

<ul style="list-style-type: none"> - Не демонстрирует знание основного содержания дисциплины; - Не владеет основными понятиями, законами и теорией, необходимыми для объяснения явлений, закономерностей и т.д.; – не умеет выполнять типовые задания и задачи предусмотренные программой 	Неудовлетворител
<ul style="list-style-type: none"> - Демонстрирует знание основного содержания дисциплины и его элементов в соответствии с прослушанным лекционным курсом; - Владение основными понятиями, законами и теорией, необходимыми для объяснения явлений, закономерностей и т.д.; –показывает умение выполнять типовые задания и задачи предусмотренные программой; – выполняет расчеты с ошибками 	Удовлетворительн
<ul style="list-style-type: none"> - ответ по вопросу или заданию аргументированный, демонстрирующий знание основного содержания дисциплины и его элементов в соответствии с прослушанным лекционным курсом и с учебной литературой; - демонстрирует понимание материала, приводит примеры; - Владение основными понятиями, законами и теорией, необходимыми для объяснения явлений, закономерностей и т.д.; 	Хорошо

<p>–показывает владение методологией дисциплины, умение выполнять типовые задания и задачи предусмотренные программой; – выполняет расчеты с ошибками</p>	Хорошо
<p>- ответ по вопросу или заданию аргументированный, логически выстроенный, полный, демонстрирующий знание основного содержания дисциплины и его элементов в соответствии с прослушанным лекционным курсом и с учебной литературой; - демонстрирует полное понимание материала, выводы доказательны, приводит примеры; - свободное владение основными понятиями, законами и теорией, необходимыми для объяснения явлений, закономерностей и т.д.; – - показывает владение методологией дисциплины, умение выполнять типовые задания и задачи предусмотренные программой; – выполняет расчеты без ошибок; - демонстрирует способность творчески применять знание теории к решению профессиональных практических задач</p>	Отлично

Перечень вопросов для проведения промежуточной аттестации

1. Понятие искусственного интеллекта. Проблематика задач искусственного интеллекта (ИИ). Основные направления исследований в области ИИ.
2. Моделирование эвристических методов. Биологическое направление ИИ. Генетические алгоритмы и их назначение. Нейроны и их моделирование.
3. Общая схема генетического алгоритма.
4. Моделирование мутации и кроссовера в генетическом алгоритме.
5. Активные и пассивные методы получения знаний. Метод Делфи изучения предметной области.
6. Система знаний. Модели представления знаний: логические модели. Понятие о нечеткой логике.
7. Система знаний. Модели представления знаний: фреймовая и продукционная.
8. Система знаний. Модели представления знаний: семантические сети. Тезаурус и его использование в ИИ. Машинное представление знаний.
9. Задача распознавания образов в ИИ. Лингвистический и геометрический подход.
10. Задача распознавания образов в ИИ. Методы классификации.
11. Задача распознавания образов в ИИ. Методы кластеризации.
12. Основные модели нейронов – модели персептрона и сигмоидального нейрона.
13. Основные модели нейронов – адалайн, инстар-оутстар, WTA, модель Хебба, стохастическая модель.
14. Понятие нейронной сети. Основные виды нейронных сетей и их использование в системах ИИ. Случайные сети. Байсовские сети и сети Петри.
15. Обучение нейронной сети.
16. Экспертные системы. Общая характеристика, структура и основные элементы экспертных систем.
17. Экспертные системы. Интеллектуальные информационные ЭС. Понятие о ИАД.
18. Экспертные системы. Классификация ЭС по назначению. Основные направления приложения ЭС. Классификация ЭС по методам построения.
19. Инженерия знаний. Метод мозгового штурма.
20. Классификация компьютерных средств разработки систем ИИ. Роль программирования в развитии методов представления знаний. Агентные системы