

Аннотация дисциплины Б.1.2.3 Дисциплина. Анализ данных

Дисциплина "Анализ данных" изучается обучающимися по основной профессиональной образовательной программе "Искусственный интеллект в бизнес-аналитике" направления подготовки "09.03.03 Прикладная информатика".

Дисциплина изучается в 3 семестре. Общая трудоемкость дисциплины составляет 108/3 часов/з.ед. Самостоятельная работа заключается в выполнении работ, указанных в разделе 4.

В ходе изучения дисциплины осуществляется текущий контроль в форме технологии рейтингового контроля в соответствии с технологической карты дисциплины, размещенной на электронном курсе, а также промежуточный контроль в форме балльно-рейтинговый контроль.

Целью изучения дисциплины является формирование следующих компетенций:

1. BD-1 Способен осуществлять поиск, сбор, очистку и предварительный анализ данных
2. ML-2 Способен применять фундаментальные принципы и методы машинного обучения включая подготовку данных оценку качества моделей и работу с признаками
3. ПК-4 Способность моделировать прикладные (бизнес) процессы и предметную область

В ходе изучения дисциплины последовательно рассматриваются темы:

1. Лекция №1. Введение в анализ данных.
 - 1.Статистические данные и их роль в ИИ.
 - 2.Классификация и структура наборов данных для ИИ-моделей.
 - 3.Типы данных и переменных в контексте машинного обучения.
 - 4.Требования к наборам данных для обучения ИИ.
2. Лекция № 2. Разведочный анализ данных.
 - 1.Анализ структуры и характеристики данных.
 - 2.Выявление аномалий и выбросов в данных.
 - 3.Идентификация связи и корреляции между переменными.
 - 4.Подготовка данных к обучению моделей машинного обучения.
3. Лекция № 3. Статистические характеристики рядов распределения с ИИ-анализом.
 - 1.Расчет и визуализация статистик средствами ИИ.
 - 2.Использование AI-алгоритмов для создания гистограмм, кумулят и полигонов.
4. Лекция № 4. Анализ статистических характеристик рядов распределения с ИИ-алгоритмами. Типические значения.
 - 1.Структурные средние.
 - 2.Расчет моды и медианы для дискретного ряда.
 - 3.Расчет моды и медианы для интервального ряда.
 - 4.Среднеквадратические отклонения.
 - 5.Дисперсия.
 - 6.Коэффициент вариации.
5. Лекция № 5. Анализ эмпирических и теоретических моментов для повышения точности, обработки выбросов и нормализации данных.
 - 1.Понятие моментов.
 - 2.Условные эмпирические моменты.
 - 3.Эмпирические и теоретические частоты.
6. Лекция № 6. Методы дисперсионного анализа с поддержкой ИИ.
 - 1.Сущность дисперсионного анализа.
 - 2.Математическая постановка.
 - 3.Проверка гипотезы о нормальном распределении.
7. Лекция № 7. Анализ статистических рядов распределения и их видов на основе ИИ-

алгоритмов.

1. Оценка и моделирование атрибутивных и вариационных рядов распределения.
2. Оценка и моделирование дискретных и интервальных рядов распределения.
3. Построение интервального вариационного ряда.
8. Лекция № 8. Анализ временных данных и динамических рядов с применением методов ИИ.
 1. Основные элементы временного ряда.
 2. Основные аналитические характеристики динамических рядов: Абсолютный прирост, темп роста, темп прироста.
 3. Использование ИИ для анализа трендов и сезонных эффектов.
9. Лекция № 9. Определение и оценка зависимости признаков с использованием корреляционного и регрессионного анализа в ИИ.
 1. Степень линейной зависимости количественных признаков-корреляционный анализ.
 2. Виды коэффициентов корреляции. Оценка на значимость.
10. Лекция № 10. Непараметрические методы в анализе данных на основе AI-алгоритмов.
 1. Ранговый коэффициент Спирмена.
 2. Ранговый коэффициент Кендалла.
 3. Коэффициент конкордации Кендалла.
11. Лекция № 11. Индексный метод анализа данных с ИИ-алгоритмами.
 1. Количественные индексы.
 2. Качественные индексы.
 3. Агрегатные индексы средних величин.
12. Лекция № 12. Типы задач машинного обучения.
 1. Классификация объектов исследования.
 2. Регрессионный анализ.
 3. Кластеризация многомерных данных.
 4. Обработка естественного языка.
13. Лекция № 13. Кластерный анализ и алгоритмы метод k-средних с оптимизацией и автоматизацией на базе ИИ.
 1. Иерархические агломеративные методы.
 2. Меры сходства, расстояния (или меры близости между объектами).
 3. Расстояние между группами объектов и меры близости двух групп объектов.
14. Лекция № 14. Кластерный анализ. Метод k-средних. Функционалы качества разбиения.
 1. Алгоритм метода k-средних.
 2. Анализ качества кластеризации- функционалы.
15. Лекция № 15. Снижение размерности анализируемого признакового пространства.
 1. Факторный анализ.
 2. Метод главных компонент.
16. Лекция № 12. Анализ степени зависимости многомерных данных.
 1. Корреляционный анализ многомерных данных.
 2. Математическая модель множественного корреляционного анализа.
 3. Разбор примера.

Основными стратегическими образовательными технологиями являются: лекционные занятия, практические и лабораторные занятия, дискуссионные, процедуры самообучения, обеспечено разнообразие форматов и методов обучения (лекции, семинары с ит-практиками, конкурсы, хакатоны, кейс-чемпионаты, мастер-классы, командные проектные задания и др.) в каждом семестре, доступ к учебным материалам, методическим указаниям, контрольно-измерительным тестам обеспечивается по сети интернет с помощью электронного курса и сервисов образовательного портала пгту, интегрированного с федеральной платформой гис ссос, активно планируется использовать ресурсы открытых образовательных платформ..

В рамках указанных технологий применяются тактические образовательные технологии: задания, классическая лекция, лекция с элементами мозгового штурма, лекция-провокация, проблемная лекция, обеспечено разнообразие форматов и методов обучения (лекции, семинары с ит-практиками, конкурсы, хакатоны, кейс-чемпионаты, мастер-классы, командные проектные задания и др.) в каждом семестре, доступ к учебным материалам, методическим указаниям, контрольно-измерительным тестам обеспечивается по сети интернет с помощью электронного курса и сервисов образовательного портала пгту, интегрированного с федеральной платформой гис ссос, активно планируется использовать ресурсы открытых образовательных платформ..