

Аннотация к рабочей программы дисциплины
Б1.О.35 Обработка данных на Python

Объем трудоемкости: 2 зачетные единицы

Цель дисциплины: сформировать у студентов базовые компетенции в области обработки и анализа данных с использованием языка программирования Python. В ходе освоения дисциплины студенты изучат фундаментальные принципы работы с различными типами данных и форматами, освоят основные библиотеки для их загрузки, очистки, преобразования и первичного визуального анализа. Это позволит создать прочную основу для дальнейшего изучения более сложных дисциплин, таких как машинное обучение и глубокое обучение, и сформировать навыки, необходимые для решения практических задач анализа данных.

Задачи дисциплины

- Изучение основ языка Python и его экосистемы библиотек для анализа данных (NumPy, Pandas, Matplotlib, Seaborn).
- Освоение методов загрузки данных из различных источников и форматов (CSV, JSON, таблицы).
- Приобретение практических навыков проверки данных на корректность, выявления и обработки пропусков, дубликатов и аномалий.
- Формирование умения выполнять первичный анализ и визуализацию данных для выявления закономерностей и тенденций.
- Изучение базовых методов предварительной обработки данных, включая преобразование и кодирование признаков.
- Освоение принципов работы с текстовыми данными: их очистка, токенизация и векторизация.
- Формирование понимания принципов формирования и подготовки качественных наборов данных (datasets).

Место дисциплины (модуля) в структуре образовательной программы

Дисциплина «Обработка данных на Python» относится к «Обязательная часть» Блока 1 «Дисциплины (модули)» учебного плана.

Код и наименование индикатора	Результаты обучения по дисциплине
ОПК-3 Способен понимать и применять современные информационные технологии, в том числе отечественные, при создании программных продуктов и программных комплексов различного назначения	
ОПК-3.1 Аргументировано применяет современные информационные технологии, в том числе отечественные, при создании программных продуктов и программных комплексов различного назначения	Знать современные библиотеки и фреймворки Python для обработки данных; основные принципы работы с открытыми и отечественными ИТ-решениями; критерии выбора технологий для решения задач обработки данных. Уметь сравнивать и выбирать подходящие инструменты обработки данных; применять современные библиотеки Python для решения практических задач; обосновывать выбор технологического стека. Владеть навыками работы с основными библиотеками обработки данных

	Python; методами интеграции различных технологий в единое решение.
BD-1 Способен осуществлять поиск, сбор, очистку и предварительный анализ данных	
BD-1.4 Отбирает признаки данных, значимые для исследования	Знать основные типы данных, критерии оценки информативности признаков, методы и алгоритмы отбора признаков. Уметь анализировать распределения признаков, применять методы feature selection, интерпретировать результаты. Владеть навыками использования специализированного программного обеспечения и библиотек для анализа и отбора признаков.
BD-2 Способен определять требования к наборам данных для решения задач машинного обучения, проводить разметку и анализ наборов данных, оценивать качество данных, обеспечивать непрерывную интеграцию данных	
BD-2.2 Работает с данными, в том числе собирает данные из разрозненных источников, проверяет данные на корректность	Знать форматы хранения данных, принципы работы с базами данных и сетевыми API, типовые проблемы качества данных. Уметь загружать данные из различных источников, проводить первичный осмотр выявлять аномалии, осуществлять базовую очистку данных. Владеть технологиями извлечения, преобразования и загрузки данных на базовом уровне.
ML-2 Способен применять фундаментальные принципы и методы машинного обучения, включая подготовку данных, оценку качества моделей и работу с признаками	
ML-2.2 Применяет методы предварительной обработки данных и работы с признаками	Знать методы обработки пропусков, кодирования категориальных признаков, масштабирования числовых признаков, концепцию конвейеров обработки данных. Уметь выбирать и применять методы предобработки в зависимости от типа данных, строить конвейеры обработки Владеть навыками применения библиотек для предварительной обработки данных и создания пайплайнов.
PL-1 Способен применять язык программирования Python для решения задач в области ИИ	
PL-1.4 Проектирует системы распределённых вычислений на Python для эффективной обработки большого количества задач	Знать принципы модульности и компонентного подхода при проектировании программных систем основы оптимизации кода на Python библиотеки для работы с большими объемами данных и их внутренние механизмы Уметь проектировать и структурировать код для обработки данных в виде логических модулей

	<p>выявлять "узкие места" в процессе обработки данных</p> <p>эффективно использовать встроенные методы библиотек для ускорения вычислений</p> <p>Владеть навыками создания оптимизированных конвейеров обработки данных</p> <p>методами профилирования кода для оценки производительности</p> <p>техниками работы с большими datasets с</p>
LLM-2	Дообучение и адаптация генеративных моделей
LLM-2.2 Создаёт обучающие наборы данных	<p>Знать принципы разметки данных, методы аугментации, требования к качеству datasets.</p> <p>Уметь формировать репрезентативные выборки, применять аугментацию, организовывать процесс разметки.</p> <p>Владеть навыками использования инструментов для аугментации, методами проверки согласованности разметки, технологиями версионирования данных.</p>

Содержание дисциплины:

Распределение видов учебной работы и их трудоемкости по разделам дисциплины.
Разделы (темы) дисциплины, изучаемые в 3 семестре

№	Наименование разделов (тем)	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа
			Л	ПЗ	ЛР	
1	2	3	4	5	6	7
1.	Основы работы с данными в python	14	4		6	4
2.	Базовые операции и анализ данных	16	4		8	4
3.	Визуализация и исследовательский анализ	12	2		6	4
4.	Преобразование и обогащение данных	14	2		8	4
5.	Интеграция и автоматизация обработки данных	13,8	4		6	3,8
ИТОГО по разделам дисциплины		69,8	16		34	19,8
Контроль самостоятельной работы (КСР)		2				
Промежуточная аттестация (ИКР)		0,2				
Подготовка к текущему контролю						
Общая трудоемкость по дисциплине		72				

Курсовые работы: не предусмотрена

Форма проведения аттестации по дисциплине: зачет

Автор: Харченко А.В. – к. п. н., доцент кафедры информационных технологий