

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

Объем трудоемкости: 4 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.

Пререквизитом для освоения дисциплины «Обработка данных на Python» является фундаментальная подготовка по дисциплине «Основы программирования» (1 семестр). Постреквизитами выступают такие дисциплины как «Технологии обработки больших данных» (6 семестр), «Глубокое обучение» (6 семестр) и «Машинное обучение» (5 семестр). Кроме того, данная дисциплина тесно связана со смежными курсами: «Базы данных» (4 семестр) и «Алгоритмы и структуры данных» (2 семестр), формируя единый технологический стек для анализа данных.

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 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	Знать принципы модульности и компонентного подхода при проектировании программных систем

для эффективной обработки большого количества задач	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 выявлять "узкие места" в процессе обработки данных эффективно использовать встроенные методы библиотек для ускорения вычислений Владеть навыками создания оптимизированных конвейеров обработки данных методами профилирования кода для оценки производительности техниками работы с большими datasets с
LLM-2 Дообучение и адаптация генеративных моделей	
LLM-2.1 Понимает принципы дообучения	Знать определение дообучения, отличие от обучения с нуля, понятия обучающей и валидационной выборок, признаки переобучения. Уметь разделять данные на обучающую и проверочную части, запускать дообучение в Colab или на компьютере, выбирать скорость обучения, понимать график ошибок Владеть навыками запуска готового кода дообучения (ноутбук/скрипт), загрузкой модели из библиотеки, сохранением результатов, проверкой ответов модели на тестовых примерах.
LLM-2.2 Создает обучающие наборы данных	Знать принципы разметки данных, методы аугментации, требования к качеству datasets. Уметь формировать репрезентативные выборки, применять аугментацию, организовывать процесс разметки. Владеть навыками использования инструментов для аугментации, методами проверки согласованности разметки, технологиями версионирования данных.

Содержание дисциплины:

Распределение видов учебной работы и их трудоемкости по разделам дисциплины.
Разделы (темы) дисциплины, изучаемые в 2 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	20	4		6	10
2.	Базовые операции и анализ данных	22	4		8	10
3.	Визуализация и исследовательский анализ	20	2		6	12
4.	Преобразование и обогащение данных	22	2		8	12
5.	Интеграция и автоматизация обработки данных	22	4		6	12
ИТОГО по разделам дисциплины		106	16		34	56
Контроль самостоятельной работы (КСР)		2				
Промежуточная аттестация (ИКР)		0,3				
Подготовка к текущему контролю		35,7				
Общая трудоемкость по дисциплине		144				

Курсовые работы: *не предусмотрена*

Форма проведения аттестации по дисциплине: *экзамен*

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