

Б1.В.ДВ.02.01 ПРИКЛАДНОЙ АНАЛИЗ ТЕКСТОВЫХ ДАННЫХ НА PYTHON

Объем трудоемкости: 3 зачетные единицы

Цель дисциплины - развитие навыков программирования на языке Python и обучение методам прикладного анализа текстовых данных на Python.

Задачи дисциплины.

- 1) Развитие навыков программирования на языке Python;
- 2) Развитие способности решать поставленные задачи по импорту, анализу и визуализации данных в Python;
- 3) Развитие способности представлять результаты в интерактивном формате Jupyter Notebook;
- 4) Формирование навыков применения прикладного анализа текстовых данных на Python в профессиональной деятельности лингвиста.

Место дисциплины (модуля) в структуре образовательной программы.

Дисциплина Б1.В.ДВ.02.01 ПРИКЛАДНОЙ АНАЛИЗ ТЕКСТОВЫХ ДАННЫХ НА PYTHON относится к части, формируемой участниками образовательных отношений, Блока 1 «Дисциплины (модули)» учебного плана.

В соответствии с рабочим учебным планом дисциплина изучается на 4 курсе на очной форме обучения в 7 и 8 семестрах. Вид промежуточной аттестации: зачет (7 семестр), экзамен (8 семестр).

Для изучения дисциплины студент должен обладать знаниями по дисциплинам: «Теория вероятности и математическая статистика», «Веб-ресурсы как практика коммуникации», «Основы программирования», «Методы лингвистического анализа». «Введение в анализ лингвистических данных с помощью Python и методов машинного обучения».

Полученные в процессе обучения по данной дисциплине знания могут быть использованы при изучении таких дисциплин как «Интеллектуальный анализ больших текстовых данных» и при проведении эмпирического исследования в ходе написания ВКР.

Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы.

Изучение данной учебной дисциплины направлено на формирование у обучающихся следующих компетенций: ПК-1

Код и наименование индикатора*	Результаты обучения по дисциплине (знает, умеет, владеет (навыки и/или опыт деятельности))
ПК-1 Способен разрабатывать и реализовывать проекты в области автоматизации исследований по теоретической и прикладной лингвистике	
ИПК-1.1. Демонстрирует способность к разработке и реализации проектов в области автоматизации исследований по теоретической и прикладной	Знает как разработать и реализовать проекты в области автоматизации исследований по теоретической и прикладной лингвистике
	Умеет разрабатывать и реализовывать проекты

Код и наименование индикатора*	Результаты обучения по дисциплине (<i>знает, умеет, владеет (навыки и/или опыт деятельности)</i>)
лингвистике	в области автоматизации исследований по теоретической и прикладной лингвистике
	Владеет методами автоматизации исследований по теоретической и прикладной лингвистике

Содержание дисциплины.

Распределение видов учебной работы и их трудоемкости по разделам дисциплины.
Разделы (темы) дисциплины, изучаемые в 7 семестре (*очная форма*)

№	Наименование разделов (тем)	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа
			Л	ПЗ	ЛР	
1	2	3	4	5	6	7
1.	Основы программирования на языке Python (refresher).	7			5	2
2.	Математический аппарат (refresher). Введение в модуль NumPy. Основы работы с Pandas. Разведывательный анализ данных.	7			5	2
3.	Визуализация данных. Представление результатов исследования.	7			5	2
4.	Работа с текстовыми данными. Сбор данных из открытых источников.	7			5	2
5.	Предварительная обработка текстовых данных. Текстовый анализ.	7,8			6	1,8
	<i>Итого по дисциплине:</i>	<i>35,8</i>			<i>26</i>	<i>9,8</i>
	Контроль самостоятельной работы (КСР)					
	Промежуточная аттестация (ИКР)	0,2				
	Подготовка к текущему контролю					
	Общая трудоемкость по дисциплине	36				

Примечание: Л – лекции, ПЗ – практические занятия / семинары, ЛР – лабораторные занятия, СРС – самостоятельная работа студента

Разделы (темы) дисциплины, изучаемые в 8 семестре (*очная форма*)

№	Наименование разделов (тем)	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа
			Л	ПЗ	ЛР	
1	2	3	4	5	6	7
1.	Анализ сетей	11			4	7
2.	Введение в машинное обучение. Модуль sklearn. Задачи классификации и линейные модели.	11			4	7

3.	Деревья решений. Случайный лес. Ансамбли моделей.	11			4	7
4.	Программирование на Python в профессии лингвиста	12			4	8
<i>Итого по дисциплине:</i>		45			16	29
	Контроль самостоятельной работы (КСР)					
	Промежуточная аттестация (ИКР)	0,3				
	Подготовка к текущему контролю	26,7				
	Общая трудоемкость по дисциплине	72				

Примечание: Л – лекции, ПЗ – практические занятия / семинары, ЛР – лабораторные занятия, СРС – самостоятельная работа студента

Курсовые работы: *не предусмотрены*

Форма проведения аттестации по дисциплине: *зачет, экзамен*

Автор: Рябченко Н.А., канд. полит. наук., доц.