

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ПОВОЛЖСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ



УТВЕРЖДАЮ
Декан ФГО

УТВЕРЖДАЮ /А.В. Артамонова/
(Ф.И.О. декана (директора института))

14.02.2025 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

М.1.2.2 Методы искусственного интеллекта в компьютерной лингвистике

(код и наименование дисциплины по учебному плану)

Направление подготовки
(специальность)

45.04.02 Лингвистика

Квалификация выпускника

Магистр

(бакалавр/магистр/специалист)

Программа магистратуры

Цифровая лингвистика

Курс 1, 2

Семестр 1, 2, 3, 4

Распределение учебного времени

Трудоемкость по учебному плану	288 / 8	часов/зачетных единиц
Лекции	-	часов
Лабораторные работы	-	часов
Практические занятия	58	часов
Иная контактная работа	-	часов
Всего контактной работы (без учета экз.)	58	часов
Контактная работа по экзамену	6	часов
Курсовой проект (работа)	-	семестр
Самостоятельная работа обучающихся (без учета экз.)	194	часов
Самостоятельная работа по подготовке к экзамену	30	часов
Экзамен	4	семестр
Зачет	1, 2, 3	семестр
БРК, ДЗ	-	семестр

(год)

Программа составлена в соответствии с требованиями ФГОС ВО направления подготовки (специальности) 45.04.02 Лингвистика

Программу составили:

доцент	ИЯиЛ	СОГЛАСОВАНО	К.О. Иванов
(должность)	(кафедра)		(И.О. Фамилия)

РАССМОТРЕНА и ОДОБРЕНА на заседании кафедры, за которой закреплена дисциплина
Кафедра иностранных языков и лингвистики

		(наименование кафедры)	
29.01.2025	протокол №	4	
(дата)			
Заведующий кафедрой	СОГЛАСОВАНО	О.В. Филипчук	
		(И.О. Фамилия)	

Рабочая программа СОГЛАСОВАНА с факультетом (институтом), выпускающей(ими) кафедрой(ами).

СООТВЕТСТВУЕТ действующей ОП.

Заведующий кафедрой	СОГЛАСОВАНО	О.В. Филипчук
		(И.О. Фамилия)

Председатель методической комиссии факультета (института), в который входит выпускающая кафедра

СОГЛАСОВАНО	А.В. Артамонова
	(И.О. Фамилия)

Эксперт(ы): Лазарева Е.А. , директор Школы изучения иностранных языков «Лингва»

Рабочая программа проверена и зарегистрирована в УМЦ 28.02.2025 г.

Специалист учебно-методического центра СОГЛАСОВАНО /Т.А. Смирнова/

Раздел 1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целью освоения дисциплины является достижение планируемых результатов обучения, соответствующих установленным в ОПОП индикаторам достижения компетенций:

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Результаты обучения
1. ПК-2 Организационно-педагогическое обеспечение программ дополнительного образования	ИД-1 ПК-2 разрабатывает образовательные программы дополнительного лингвистического образования с учетом современных подходов и технологий лингводидактики, педагогического дизайна, дистанционного и смешанного обучения	знания: Знание принципов образования и психологии обучения взрослых и детей. умения: Умение планировать и оценивать образовательные процессы. навыки: Навыки разработки учебных программ, учета потребностей и уровня подготовки учащихся.
2. ПК-4 Способен осуществлять ведение процедуры медиации и разрабатывать методику выполнения аналитических работ	ИД-1 ПК-4 использует корпуса языков и инструменты их обработки, системы машинного перевода, технологию искусственного интеллекта для анализа текстовой информации	знания: Знание основных языковых явлений и правил, в том числе морфологии и синтаксиса. умения: Умение использовать инструменты машинного обучения для обработки и анализа текстов. навыки: Навыки программирования для работы с инструментами обработки текста и машинного обучения.
	ИД-2 ПК-4 демонстрирует владение технологиями устного и письменного перевода с использованием систем машинного перевода и обработки естественных языков	знания: Знание тематических областей, таких как медицина, юриспруденция, техника и т.д., в которых человек работает. умения: Умение работать с платформами и инструментами для обработки естественных языков (Natural Language Processing, NLP). навыки: Навыки преодоления ограничений систем машинного перевода и улучшения качества переводов.
	ИД-3 ПК-4 использует инструменты ИТ, лингвистические корпуса и базы данных для выявления, анализа и описания мировых практик и проведения аналитических работ	знания: Знание принципов создания, структуры и использования лингвистических корпусов. умения: Умение собирать и синтезировать информацию из различных источников. навыки: Навыки использования инструментов NLP для извлечения информации и выделения ключевых элементов из текстов.

Раздел 2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина относится к части, формируемой участниками образовательных отношений ОПОП.

Дисциплина является обязательной

Изучаемая дисциплина является основой для продолжения формирования указанных компетенций в следующих государственной итоговой аттестации в форме: Подготовка к процедуре защиты и защита выпускной квалификационной работы (ПК-2), Подготовка к процедуре защиты и защита выпускной квалификационной работы (ПК-4)

Раздел 3. ОПИСАНИЕ ОБРАЗОВАТЕЛЬНЫХ ТЕХНОЛОГИЙ

Для формирования заявленных компетенций используются методологические технологии, реализующие деятельностный, личностно-ориентированный, практико-ориентированный подходы.

Основными стратегическими технологиями являются: исследовательские, практические занятия, процедуры самообучения

На достижение конкретных целей обучения направлены применяемые тактические технологии: задания, мини-проекты

Раздел 4. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

1 семестр

Виды и темы занятий	Количество часов	Формируемые компетенции
Работа с модулями Python для автоматической генерации документов	72	ПК-2, ПК-4
Практическое занятие. Основы работы с текстом в Python: Ввод/вывод текстовых данных. Работа с строками и списками. Основы форматирования текста.	2	
Практическое занятие. Работа с библиотеками для генерации документов: Введение в библиотеки для работы с документами (например, docx, reportlab, fpdf). Создание простых документов.	2	
Практическое занятие. Шаблонизация документов: Использование библиотек для создания и использования шаблонов (например, Jinja2). Автоматическое заполнение шаблонов данными.	2	
Практическое занятие. Генерация PDF-документов: Использование библиотек для создания PDF (например, reportlab, fpdf). Вставка изображений и других элементов в PDF.	2	
Практическое занятие. Работа с документами Microsoft Word: Использование библиотеки python-docx для создания и редактирования документов Word. Вставка таблиц, изображений, автоматическая нумерация и маркировка.	2	
Практическое занятие. Генерация отчетов и документации: Создание скрипта для автоматической генерации отчетов по данным.	2	

Использование стилей и разметки для улучшения читаемости.		
Практическое занятие. Интерактивные документы с использованием Jupyter Notebook: Создание интерактивных документов с помощью Jupyter Notebook. Интеграция кода Python для автоматической генерации результатов и графиков.	2	
Практическое занятие. Генерация документов из баз данных: Использование SQL для извлечения данных из базы данных. Автоматическая генерация отчетов на основе данных из базы.	2	
Практическое занятие. Автоматизация процесса подписи документов: Создание скрипта для автоматической вставки электронной подписи в документы. Обеспечение безопасности при работе с подписями.	2	
Задания для самостоятельной работы, в том числе выполнение Подготовка к практическим занятиям	54	
Иная контактная работа:	0	

2 семестр

Виды и темы занятий	Количество часов	Формируемые компетенции
Морфологический анализ текстов на языке Python	72	ПК-2, ПК-4
Практическое занятие. Основы морфологического анализа текста: Использование библиотеки NLTK или SpaCy для выполнения базового морфологического анализа текста. Разбор текста на токены и определение их морфологических характеристик.	2	
Практическое занятие. Лемматизация и стемминг: Сравнение методов лемматизации и стемминга. Прим	2	
Практическое занятие. Частеречная разметка: Использование библиотек для определения частей речи слов в тексте. Анализ текстов на основе частеречной разметки.	2	
Практическое занятие. Извлечение именованных сущностей: Использование инструментов для распознавания именованных сущностей в тексте. Разработка скрипта для извлечения конкретных типов именованных сущностей.	2	
Практическое занятие. Анализ морфологических особенностей разных языков: Исследование и сравнение результатов морфологического анализа для текстов на разных языках. Рассмотрение особенностей морфологии различных языков.	2	
Практическое занятие. Обработка текстов научных статей: Создание программы для морфологического анализа научных текстов. Выделение ключевых слов и терминов с использованием морфологического анализа.	2	

Практическое занятие. Морфологический анализ в области обработки естественного языка (NLP): Интеграция морфологического анализа в NLP-приложения. Применение анализа к задачам машинного обучения, таким как классификация текстов.	2
Практическое занятие. Анализ эмоциональной окраски: Использование морфологического анализа для определения эмоциональной окраски текста.	2
Задания для самостоятельной работы, в том числе выполнение Подготовка к практическим занятиям	56
Иная контактная работа:	0

3 семестр

Виды и темы занятий	Количество часов	Формируемые компетенции
Методы машинного обучения для анализа текстов	72	ПК-2, ПК-4
Практическое занятие. Классификация текстов: Разработка модели машинного обучения для классификации текстовых документов. Применение методов, таких как наивный Байес, метод опорных векторов (SVM) или глубокое обучение для решения задачи.	2	
Практическое занятие. Анализ сентимента: Создание модели для определения тональности текста (положительной, негативной или нейтральной). Использование библиотек, таких как NLTK или TextBlob.	2	
Практическое занятие. Тематическое моделирование: Исследование методов тематического моделирования, таких как Latent Dirichlet Allocation (LDA). Применение тематического моделирования для выявления ключевых тем в коллекции текстов.	2	
Практическое занятие. Извлечение ключевых фраз и слов: Разработка алгоритма для извлечения ключевых фраз или слов из текста. Использование TF-IDF или методов глубокого обучения для этой задачи.	2	
Практическое занятие. Автоматическая обработка естественного языка (NLP): Изучение библиотек NLP, таких как SpaCy или NLTK. Применение методов NLP для предобработки текстов перед обучением модели.	2	
Практическое занятие. Сравнение различных моделей машинного обучения: Сравнение производительности различных алгоритмов машинного обучения для анализа текстов. Использование метрик, таких как точность, полнота и F1-мера.	2	
Практическое занятие. Использование предварительно обученных моделей: Интеграция предварительно обученных моделей, таких как Word2Vec или BERT. Оценка улучшения производительности за счет использования предварительно обученных весов.	2	

Задания для самостоятельной работы, в том числе выполнение Подготовка к практическим занятиям	58	
Иная контактная работа:	0	

4 семестр

Виды и темы занятий	Количество часов	Формируемые компетенции
Глубокое обучение при обработке естественного языка	36	ПК-2, ПК-4
Практическое занятие. Обучение нейронной сети для классификации текстов: Разработка модели глубокого обучения для классификации текстов по заданным категориям. Использование рекуррентных нейронных сетей (RNN) или сверточных нейронных сетей (CNN).	2	
Практическое занятие. Сентимент-анализ с использованием рекуррентных нейронных сетей: Создание модели для определения тональности текста (положительной, негативной или нейтральной) с использованием RNN. Применение долгой краткосрочной памяти (LSTM) или gated recurrent unit (GRU) для учета контекста.	2	
Практическое занятие. Использование предварительно обученных эмбедингов для обработки текста: Интеграция предварительно обученных эмбедингов, таких как Word2Vec или GloVe, в модель глубокого обучения для обработки текста.	2	
Практическое занятие. Многозадачное обучение: Разработка модели для выполнения нескольких задач NLP одновременно (например, классификация именованных сущностей и определение тональности).	2	
Практическое занятие. Генерация текстов с помощью рекуррентных нейронных сетей	2	
Задания для самостоятельной работы, в том числе выполнение Подготовка к практическим занятиям	26	
Иная контактная работа:	0	
Подготовка к экзамену	30	
Проведение экзамена	6	

Раздел 5. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Изучение дисциплины рекомендуется начать с ознакомления с рабочей программой, ее структурой и содержанием разделов. Учебный материал структурирован, изучение дисциплины осуществляется в тематической последовательности. Подготовка к занятиям **семинарского типа** включает ознакомление с планом практического занятия; выполнение домашнего задания, работу с учебной и учебно-методической литературой, научными изданиями и электронными образовательными ресурсами, рекомендованными рабочей программой дисциплины. Содержание **самостоятельной работы** определяется рабочей программой дисциплины, оценочными и методическими материалами, заданиями и указаниями преподавателя. Самостоятельная работа может осуществляться в аудиторной и внеаудиторной формах.

Эффективным средством осуществления самостоятельной работы является электронная информационно-образовательная среда университета, которая обеспечивает доступ к образовательной программе, рабочей программе дисциплины, к электронным библиотечным системам, профессиональным базам данных и информационным справочным системам.

Периодичность проведения, формы текущего контроля успеваемости, система оценивания хода освоения дисциплин представлены в рабочей программе.

Раздел 6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ И УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Учебно-методическое обеспечение

№№ п/п	Список используемой литературы	Количество экземпляров печатных изданий, имеющихся в библиотеке, или электронный адрес издания (ресурса) в сети Интернет
УЧЕБНЫЕ, УЧЕБНО-МЕТОДИЧЕСКИЕ И НАУЧНЫЕ ИЗДАНИЯ		
1.	Барский, А. Б. Логические нейронные сети [Электронный ресурс] / Барский А. Б. 2-е изд. Москва: ИНТУИТ, 2016. - 492 с. ISBN 978-5-94774-646-4.	https://e.lanbook.com/book/100630
2.	Волосова, А. В. Технологии искусственного интеллекта в ULS-системах [Электронный ресурс] / Волосова А. В. Санкт-Петербург: Лань, 2022. - 308 с. ISBN 978-5-8114-8839-1.	https://e.lanbook.com/book/208568
3.	Советов, Б. Я. Информационные технологии: теоретические основы [Электронный ресурс] / Советов Б. Я., Цехановский В. В. 2-е изд., стер. Санкт-Петербург: Лань, 2022. - 444 с. ISBN 978-5-8114-1912-8.	https://e.lanbook.com/book/209876
4.	Романов, П. С. Системы искусственного интеллекта. Моделирование нейронных сетей в системе MATLAB. Лабораторный практикум [Электронный ресурс] : учебное пособие для вузов / Романов П. С., Романова И. П.; Романова И. П. 4-е изд., стер. Санкт-Петербург: Лань, 2024. - 140 с. ISBN 978-5-507-47377-9.	https://e.lanbook.com/book/364964
5.	Кревецкий, Александр Владимирович. Основы технологий искусственного интеллекта [Электронный ресурс] : учебное пособие / А. В. Кревецкий, Ю. А. Ипатов, Н. И. Роженцова ; под общей редакцией А. В. Кревецкого; Министерство науки и высшего образования Российской Федерации, ФГБОУ ВО "Поволжский государственный технологический университет". Йошкар-Ола: ПГТУ, 2023. - 272 с. ISBN 978-5-8158-2358-	https://portal.volgatech.net/books/Krevetskiy_Osnovy_tekhnologii_iskusstvennogo_intellekta_2023.pdf
6.	Сидоркина, Ирина Геннадьевна. Системы искусственного интеллекта [Текст] : [учеб. пособие для вузов по направлению 230100 "Информатика и вычисл. техника"] / И. Г. Сидоркина. М.: Кнорус, 2011. - 245 с. ISBN 978-5-406-00449-4. Экземпляры: всего 88.	88
7.	Остроух, А. В. Интеллектуальные информационные системы и технологии [Электронный ресурс] / Остроух А. В., Николаев А. Б. 3-е изд., стер. Санкт-Петербург: Лань, 2023. - 308 с. ISBN 978-5-507-48511-6.	https://e.lanbook.com/book/354536

8.	Чубукова, И. А. Data Mining [Электронный ресурс] / Чубукова И. А. 2-е изд. Москва: ИНТУИТ, 2016. - 470 с. ISBN 978-5-94774-819-2.	https://e.lanbook.com/book/100582
ЭЛЕКТРОННЫЕ ОБРАЗОВАТЕЛЬНЫЕ РЕСУРСЫ		
1.	Научная электронная библиотека eLIBRARY.RU	http://elibrary.ru
2.	Научная электронная библиотека «Киберленинка»	http://cyberleninka.ru
ПРОФЕССИОНАЛЬНЫЕ БАЗЫ ДАННЫХ И ИНФОРМАЦИОННЫЕ СПРАВОЧНЫЕ СИСТЕМЫ		
1.	Справочно-правовая система Консультант+	http://www.consultant.ru
2.	Информационно-правовой портал Гарант	http://www.garant.ru
3.	Профессиональные справочные системы Техэксперт	http://www.cntd.ru

6.2. Материально-техническая база и программное обеспечение

№№ п/п	Аудитории для проведения учебных занятий, самостоятельной работы и проведения государственной итоговой аттестации	Перечень основного оборудования	Программное обеспечение
1.	531 (III)	ПК B112,2 420W/Intel Celeron Dual-Core E3300/кл,мышь,фильт,мон. VA1931 (5)	Microsoft Windows Enterprise, Справочная правовая система "Консультант Плюс", Microsoft Office Standard, Агент Dr.Web, Комплект ГАРАНТ-Мастер, Microsoft Access, Microsoft Visio Professional, Microsoft Project Professional, Microsoft Visual Studio Enterprise, Комплект ПО для решения основных пользовательских задач

Раздел 7. ФОРМЫ КОНТРОЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ/ ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Критерии оценивания индикаторов достижения компетенций направлены на:

- усвоение теоретического материала (объем знаний, глубина усвоения), предусмотренного рабочей программой;
- умение излагать материал (четкость, грамотность изложения материала, точность и полнота воспроизведения учебного материала);
- умение применять теоретические знания при решении практических заданий.

Шкала оценивания представлена ниже.

Уровень сформированности элементов компетенции	Критерии оценивания	Шкала оценивания
Пороговый уровень	Обучающийся имеет знания основного материала, проявляет умение логично его излагать, но может	удовлетворительно

	допускать неточности в изложении материала, недостаточно правильные формулировки, испытывает затруднения в выполнении практических заданий.	
Продвинутый уровень	Обучающийся твердо знает программный материал, излагает его грамотно и по существу, не допускает существенных неточностей в ответе на вопрос, правильно применяет теоретические положения при решении практических вопросов и задач, владеет необходимыми навыками и приемами их выполнения	хорошо
Высокий уровень	Обучающийся глубоко и прочно усвоил программный материал, грамотно и логически стройно его излагает, дает исчерпывающие ответы на поставленные вопросы. В ответе тесно увязывается теория с практикой, при этом обучающийся не затрудняется с ответом при видоизменении задания, свободно справляется с задачами, вопросами и другими видами применения знаний, показывает знакомство с монографической литературой, периодическими изданиями, правильно обосновывает принятые решения, свободно владеет разносторонними навыками, приемами выполнения практических работ	отлично

7.1. Текущий контроль успеваемости

Текущий контроль успеваемости обеспечивает оценивание хода освоения дисциплины (модуля) и производится с применением технологии рейтингового контроля в соответствии с технологической картой дисциплины. Порядок составления технологической карты и алгоритм проведения процедуры оценивания видов деятельности обучающихся, направленных на освоение знаний, умений, навыков и/или опыта деятельности, по накопительной системе в баллах устанавливается положением о системе РИТМ в ФГБОУ ВО «ПГТУ»

7.2. Промежуточная аттестация обучающихся

Промежуточная аттестация обучающихся направлена на оценивание результатов обучения по дисциплине (модулю) и проводится с использованием фондов оценочных средств.

Примеры типовых контрольных заданий из базы фонда оценочных средств по образовательной программе.

1 семестр:

1. Какая библиотека Python часто используется для генерации документов в формате PDF?

- A. NLTK
- B. Docx
- C. ReportLab
- D. PyPDF2

2. Какой метод предназначен для создания нового документа в библиотеке python-docx?

- A. create_document()
- B. new_document()
- C. Document()
- D. make_document()

3. Какой модуль в Python позволяет работать с шаблонами для генерации документов?

- A. pdfkit
- B. jinja2
- C. docxtemplater
- D. reportlab

4. Как добавить новый абзац в документ с использованием python-docx?

- A. document.new_paragraph()
- B. document.create_paragraph()
- C. document.add_paragraph()
- D. document.insert_paragraph()

5. Как вставить изображение в документ с использованием python-docx?

- A. document.insert_image()
- B. document.add_picture()
- C. document.attach_image()
- D. document.include_image()

6. Какой метод библиотеки ReportLab предназначен для добавления текста в PDF-документ?

- A. draw_text()
- B. add_text()
- C. drawString()
- D. insert_text()

7. Как можно сохранить сгенерированный документ в файл с использованием библиотеки ReportLab?

- A. document.save('output.pdf')
- B. document.write('output.pdf')
- C. save_as_pdf(document, 'output.pdf')
- D. write_pdf('output.pdf', document)

8. Какая команда служит для установки библиотеки PythonDOCX?

- A. apt install pythondocx
- B. sudo install pythondocx
- C. pip install pythondocx
- D. apt-get pythondocx

Правильные ответы:

- 1. C. ReportLab
- 2. C. Document()
- 3. B. jinja2
- 4. C. document.add_paragraph()

- 5. B. document.add_picture()
- 6. C. drawString()
- 7. B. document.write('output.pdf')
- 8. C. pip install pythondocx

2 семестр:

1. Какая библиотека Python часто используется для выполнения морфологического анализа текста?

- A. NLTK
- B. Matplotlib
- C. Django
- D. NumPy

2. Какой метод NLTK предназначен для токенизации текста на слова?

- A. split_words()
- B. word_tokenize()
- C. tokenize_words()
- D. text_to_tokens()

3. Каким образом можно выполнить лемматизацию текста с использованием библиотеки SpaCy?

- A. text.lemmatize()
- B. spacy.lemmatize_text()
- C. nlp(text).lemmatize()
- D. doc.lemma_

4. Какая из следующих библиотек предназначена для определения частей речи в тексте?

- A. NumPy
- B. Django
- C. SpaCy
- D. Requests

5. Как можно определить начальную форму (лемму) слова с использованием библиотеки NLTK?

- A. word.lemma()
- B. lemmatizer.find_root(word)
- C. lemmatizer.lemmatize(word)
- D. word.root()

6. Какая функция NLTK позволяет определить частоту употребления слов в тексте?

- A. word_frequency()
- B. frequency_analysis()
- C. FreqDist()
- D. text_word_count()

7. Каким методом SpaCy можно извлечь именованные сущности из текста?

- A. `extract_entities()`
- B. `get_named_entities()`
- C. `entities.extract()`
- D. `doc.ents`

8. Как провести частеречную разметку текста с использованием библиотеки NLTK?

- A. `pos_tagging()`
- B. `tag_parts_of_speech()`
- C. `nltk.pos_tag()`
- D. `analyze_pos()`

Правильные ответы:

- 1. A. NLTK
- 2. B. `word_tokenize()`
- 3. D. `doc.lemma_`
- 4. C. SpaCy
- 5. C. `lemmatizer.lemmatize(word)`
- 6. C. `FreqDist()`
- 7. D. `doc.ents`
- 8. C. `nltk.pos_tag()`

3 семестр:

1. Какие задачи можно решать с использованием машинного обучения в анализе текстов?

- A. Только классификация
- B. Кластеризация и извлечение информации
- C. Предсказание временных рядов
- D. Только сегментация текста

2. Какие типы признаков могут использоваться при обучении моделей машинного обучения для анализа текстов?

- A. Только числовые
- B. Только категориальные
- C. Текстовые, числовые, категориальные
- D. Только временные

3. Какую задачу решает метод классификации текстов?

- A. Прогнозирование числовых значений
- B. Группировка текстов по схожести
- C. Определение категории текста
- D. Извлечение информации из текста

4. Как провести предобработку текстовых данных перед обучением модели машинного обучения?

- A. Только лемматизация
- B. Только токенизация

C. Лемматизация, удаление стоп-слов, токенизация и векторизация

D. Только удаление стоп-слов

5. Какая библиотека Python часто используется для создания моделей машинного обучения в области анализа текстов?

A. TensorFlow

B. Django

C. Matplotlib

D. Flask

6. Какие методы используются для анализа тональности текста с использованием машинного обучения?

A. Кластеризация

B. Регрессия

C. Классификация

D. Перцептроны

7. Какой метод машинного обучения часто применяется для анализа текстов в нейросетевых моделях?

A. Метод опорных векторов (SVM)

B. Решающие деревья

C. Рекуррентные нейронные сети (RNN)

D. K-средних

8. Каким образом можно оценить производительность модели машинного обучения для анализа текстов?

A. Только accuracy

B. Precision, recall и F1-мера

C. R-squared

D. Только AUC-ROC

Правильные ответы:

1. B. Кластеризация и извлечение информации

2. C. Текстовые, числовые, категориальные

3. C. Определение категории текста

4. C. Лемматизация, удаление стоп-слов, токенизация и векторизация

5. A. TensorFlow

6. C. Классификация

7. C. Рекуррентные нейронные сети (RNN)

8. B. Precision, recall и F1-мера

4 семестр:

1. Какие методы обработки текстов могут применяться с использованием глубокого обучения для определения синонимов, антонимов или ассоциаций между словами?

2. Что такое векторные представления слов (Word Embeddings), и как они используются для

обучения моделей глубокого обучения в NLP?

Перечень вопросов для проведения промежуточной аттестации