

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ
федеральное государственное автономное образовательное учреждение высшего
образования
"САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
АЭРОКОСМИЧЕСКОГО ПРИБОРОСТРОЕНИЯ"

Кафедра № 63

УТВЕРЖДАЮ
Ответственный за образовательную
программу

к.ф.н., доц. _____
(должность, уч. степень, звание)

М.А. Чиханова _____
(инициалы, фамилия)

_____ (подпись)
«20» февраля 2025 г

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Машинное обучение»
(Наименование дисциплины)

Код направления подготовки/ специальности	45.03.02
Наименование направления подготовки/ специальности	Лингвистика
Наименование направленности	Теоретическая и прикладная лингвистика
Форма обучения	заочная
Год приема	2025

Санкт-Петербург– 2025

Лист согласования рабочей программы дисциплины

Программу составил (а)

зав. каф., доц., к. фил. н.
(должность, уч. степень,
звание)

14.02.2025

(подпись, дата)

М. А. Чиханова

(инициалы, фамилия)

Программа одобрена на заседании кафедры № 63

«14» февраля 2025 г, протокол № 7

Заведующий кафедрой № 63

к.ф.н., доц.
(уч. степень, звание)

14.02.2025

(подпись, дата)

М.А. Чиханова

(инициалы, фамилия)

Заместитель декана факультета №6 по методической работе

проф., д.и.н., доц.
(должность, уч. степень,
звание)

14.02.2025

(подпись, дата)

Л.Ю. Гусман

(инициалы, фамилия)

Аннотация

Дисциплина «Машинное обучение» входит в образовательную программу высшего образования – программу бакалавриата по направлению подготовки/ специальности 45.03.02 «Лингвистика» направленности «Теоретическая и прикладная лингвистика». Дисциплина реализуется кафедрой «№63».

Дисциплина нацелена на формирование у выпускника следующих компетенций:

УК-2 «Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений»

ПК-6 «Готовность овладеть методами формального и когнитивного моделирования естественного языка и методами создания метаязыков»

ПК-7 «Готовность овладеть основными математико-статистическими методами обработки лингвистической информации с учетом элементов программирования и автоматической обработки лингвистических корпусов»

ПК-8 «Способность решать основные задачи в области лингвистического обеспечения информационных систем»

Содержание дисциплины охватывает круг вопросов, связанных с изучением некоторых алгоритмов обучения с использованием соответствующих функций библиотеки компьютерного зрения OpenCV и IPP. В курсе рассматриваются основные задачи обучения по прецедентам: классификация, кластеризация, регрессия, понижение размерности. Изучаются методы их решения, как классические, так и новые, созданные за последние годы. Упор делается на глубокое понимание математических основ, взаимосвязей, достоинств и ограничений рассматриваемых методов.

Преподавание дисциплины предусматривает следующие формы организации учебного процесса: *лекции, практические занятия, самостоятельная работа студента.*

Программой дисциплины предусмотрены следующие виды контроля: текущий контроль успеваемости, промежуточная аттестация в форме зачета.

Общая трудоемкость освоения дисциплины составляет 2 зачетных единицы, 72 часа.

Язык обучения по дисциплине «русский/английский».

1. Перечень планируемых результатов обучения по дисциплине

1.1. Цели преподавания дисциплины

Цели преподавания дисциплины: 1) ознакомить студентов с основными концепциями и алгоритмами машинного обучения; 2) научить студентов применять алгоритмы машинного обучения для решения практических задач; 3) развить навыки программирования и анализа данных с использованием Python и библиотек машинного обучения.

1.2. Дисциплина входит в состав части, формируемой участниками образовательных отношений, образовательной программы высшего образования (далее – ОП ВО).

1.3. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения ОП ВО.

В результате изучения дисциплины обучающийся должен обладать следующими компетенциями или их частями. Компетенции и индикаторы их достижения приведены в таблице 1.

Таблица 1 – Перечень компетенций и индикаторов их достижения

Категория (группа) компетенции	Код и наименование компетенции	Код и наименование индикатора достижения компетенции
Универсальные компетенции	УК-2 Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.В.2 владеть навыками выбора оптимального способа решения задач с учетом имеющихся условий, ресурсов и ограничений УК-2.В.3 владеть навыками использования цифровых средств для решения поставленной задачи
Профессиональные компетенции	ПК-6 Готовность овладеть методами формального и когнитивного моделирования естественного языка и методами создания метаязыков	ПК-6.3.1 знать принципы построения больших языковых моделей, принципы функционирования нейросетей и систем машинного перевода
Профессиональные компетенции	ПК-7 Готовность овладеть основными математико-статистическими методами обработки лингвистической информации с учетом элементов программирования	ПК-7.3.1 знать основные математико-статистические методы обработки лингвистической информации ПК-7.В.1 владеть навыками программирования и навыками автоматической обработки корпусов

	и автоматической обработки лингвистических корпусов	
Профессиональные компетенции	ПК-8 Способность решать основные задачи в области лингвистического обеспечения информационных систем	<p>ПК-8.3.1 знать терминологию, алгоритмы и схемы, приемы и законы создания документации по аппаратным и программным компонентам информационных систем</p> <p>ПК-8.У.1 уметь определять назначение технического описания информационных систем, нормативной документации и методических указаний к выполнению работы</p> <p>ПК-8.В.1 владеть навыками демонстрации результатов анализа и расчётов, выполненных согласно изученному техническому описанию, ПО, нормативной документации; владеть навыками проведения научных исследований, связанных с автоматизированной обработкой текста</p>

2. Место дисциплины в структуре ОП

Дисциплина может базироваться на знаниях, ранее приобретенных обучающимися при изучении следующих дисциплин:

- ~ «Обработка текстовых массивов»
- ~ «Технологии искусственного интеллекта в лингвистических исследованиях»
- ~ «Когнитивные технологии в прикладной лингвистике»

Знания, полученные при изучении материала данной дисциплины, имеют как самостоятельное значение, так и используются при изучении других дисциплин:

- ~ «Информационные языки»
- ~ «Компьютерные программы в лингвистических исследованиях»
- ~ «Производственная преддипломная практика».

3. Объем и трудоемкость дисциплины

Данные об общем объеме дисциплины, трудоемкости отдельных видов учебной работы по дисциплине (и распределение этой трудоемкости по семестрам) представлены в таблице 2.

Таблица 2 – Объем и трудоемкость дисциплины

Вид учебной работы	Всего	Трудоемкость по семестрам
		№10
1	2	3
Общая трудоемкость дисциплины, ЗЕ/ (час)	2/ 72	2/ 72
Из них часов практической подготовки	10	10
Аудиторные занятия, всего час.	24	24
в том числе:		
лекции (Л), (час)	10	10
практические/семинарские занятия (ПЗ),	14	14

(час)		
лабораторные работы (ЛР), (час)		
курсовой проект (работа) (КП, КР), (час)	*	*
экзамен, (час)		
Самостоятельная работа , всего (час)	48	48
Вид промежуточной аттестации: зачет, дифф. зачет, экзамен (Зачет, Дифф. зач, Экз.**)	Зачет	Зачет

Примечание: ** кандидатский экзамен

* - часы , не входящие в аудиторную нагрузку

4. Содержание дисциплины

4.1. Распределение трудоемкости дисциплины по разделам и видам занятий.

Разделы, темы дисциплины и их трудоемкость приведены в таблице 3.

Таблица 3 – Разделы, темы дисциплины, их трудоемкость

Разделы, темы дисциплины	Лекции (час)	ПЗ (СЗ)	ЛР (час)	КП (час)	СРС (час)
Семестр 10					
Раздел 1. Введение в машинное обучение Тема 1.1. Основные понятия и области применения машинного обучения Тема 1.2. Установка и настройка Python, Jupyter Notebook, основные библиотеки (NumPy, Pandas, Matplotlib)	2	2 2			4 4
Раздел 2. Основы Python и библиотек для машинного обучения Тема 2.1 Обзор основных библиотек для анализа данных и машинного обучения Тема 2.2 Методы очистки и подготовки данных, работа с пропущенными значениями Тема 2.3 Линейная регрессия, метод наименьших квадратов. Логистическая регрессия для бинарной классификации Тема 2.4 Принципы работы k-NN, метрические методы. Реализация алгоритма k-NN для задач классификации и регрессии.	4	5			20
Раздел 3. Классификаторы Тема 3.1 Теория наивного байеса, применение в классификации текстов. Реализация наивного байесовского классификатора для задач классификации Тема 3.2 Принципы построения деревьев решений, критерии расщепления. Реализация деревьев решений для задач классификации и регрессии. Тема 3.3 Ансамблевые методы, случайный лес, градиентный бустинг Реализация случайного леса и градиентного бустинга с использованием Scikit-Learn Тема 3.4 Основы кластеризации, алгоритмы k-	4	5			20

means и иерархической кластеризации. Реализация алгоритмов кластеризации и анализ результатов. Тема 3.5 Методы снижения размерности, PCA, t-SNE. Применение PCA и t-SNE для визуализации и анализа данных Тема 3.6 Основы нейронных сетей, перцептрон, многослойные перцептроны (MLP). Реализация простых нейронных сетей с использованием Keras Тема 3.7 Основы обработки естественного языка, мешок слов, TF-IDF. Применение методов NLP для анализа текстовых данных					
Выполнение курсовой работы					
Итого в семестре:	10	14			48
Итого	10	14			48

Практическая подготовка заключается в непосредственном выполнении обучающимися определенных трудовых функций, связанных с будущей профессиональной деятельностью.

4.2. Содержание разделов и тем лекционных занятий.

Содержание разделов и тем лекционных занятий приведено в таблице 4.

Таблица 4 – Содержание разделов и тем лекционного цикла

Номер раздела	Название и содержание разделов и тем лекционных занятий
1	Введение в машинное обучение Основные понятия и области применения машинного обучения Установка и настройка Python, Jupyter Notebook, основные библиотеки (NumPy, Pandas, Matplotlib)
2	Основы Python и библиотек для машинного обучения 1. Обзор основных библиотек для анализа данных и машинного обучения 2. Методы очистки и подготовки данных, работа с пропущенными значениями 3. Линейная регрессия, метод наименьших квадратов. 4. Логистическая регрессия для бинарной классификации 5. Принципы работы k-NN, метрические методы.
3	Классификаторы 1. Теория наивного байеса, применение в классификации текстов. 2. Принципы построения деревьев решений, критерии расщепления. 3. Ансамблевые методы, случайный лес. 4. Основы кластеризации, алгоритмы k-means и иерархической кластеризации. 5. Методы снижения размерности, PCA, t-SNE.

	6. Основы нейронных сетей, перцептрон, многослойные перцептроны (MLP). 7. Основы обработки естественного языка, мешок слов, TF-IDF.
--	--

4.3. Практические (семинарские) занятия

Темы практических занятий и их трудоемкость приведены в таблице 5.

Таблица 5 – Практические занятия и их трудоемкость

№ п/п	Темы практических занятий	Формы практических занятий	Трудоемкость, (час)	Из них практической подготовки, (час)	№ раздела дисциплины
Семестр 10					
1	Основные понятия и области применения машинного обучения	Решение ситуационных задач	1	1	1
2	Установка и настройка Python, Jupyter Notebook, основные библиотеки (NumPy, Pandas, Matplotlib)	Решение ситуационных задач	1	1	1
3	Обзор основных библиотек для анализа данных и машинного обучения	Решение ситуационных задач	1	1	2
4	Методы очистки и подготовки данных, работа с пропущенными значениями	Решение ситуационных задач	1	1	2
5	Линейная регрессия, метод наименьших квадратов. Логистическая регрессия для бинарной классификации	Решение ситуационных задач	1	1	2
6	Принципы работы k-NN, метрические методы. Реализация алгоритма k-NN для задач классификации и регрессии.	Решение ситуационных задач	1	1	2
7	Теория наивного	Решение	1	1	3

	байеса, применение в классификации текстов. Реализация наивного байесовского классификатора для задач классификации	ситуационных задач			
8	Принципы построения деревьев решений, критерии расщепления. Реализация деревьев решений для задач классификации и регрессии.	Решение ситуационных задач	1	1	3
9	Ансамблевые методы, случайный лес, градиентный бустинг. Реализация случайного леса и градиентного бустинга с использованием Scikit-Learn	Решение ситуационных задач	1	1	3
10	Основы кластеризации, алгоритмы k-means и иерархической кластеризации. Реализация алгоритмов кластеризации и анализ результатов	Решение ситуационных задач	1	1	3
11	Методы снижения размерности, PCA, t-SNE. Применение PCA и t-SNE для визуализации и анализа данных	Решение ситуационных задач	1	1	3
12	Основы нейронных сетей, перцептрон, многослойные перцептроны (MLP). Реализация простых нейронных сетей с использованием Keras	Решение ситуационных задач	1	1	3
13	Основы обработки естественного языка,	Решение ситуационных задач	2	2	3

	мешок слов, TF-IDF. Применение методов NLP для анализа текстовых данных				
Всего			14		

4.4. Лабораторные занятия

Темы лабораторных занятий и их трудоемкость приведены в таблице 6.

Таблица 6 – Лабораторные занятия и их трудоемкость

№ п/п	Наименование лабораторных работ	Трудоемкость, (час)	Из них практической подготовки, (час)	№ раздела дисциплины
Учебным планом не предусмотрено				
Всего				

4.5. Курсовое проектирование/ выполнение курсовой работы

Цель курсовой работы:

Часов практической подготовки:

Примерные темы заданий на курсовую работу приведены в разделе 10 РПД.

Обязательно указать темы на курсовую работу и выделить для неё время в СРС

4.6. Самостоятельная работа обучающихся

Виды самостоятельной работы и ее трудоемкость приведены в таблице 7.

Таблица 7 – Виды самостоятельной работы и ее трудоемкость

Вид самостоятельной работы	Всего, час	Семестр 10, час
1	2	3
Изучение теоретического материала дисциплины (ТО)	18	18
Курсовое проектирование (КП, КР)	14	14
Расчетно-графические задания (РГЗ)		
Выполнение реферата (Р)		
Подготовка к текущему контролю успеваемости (ТКУ)		
Домашнее задание (ДЗ)	10	10
Контрольные работы заочников (КРЗ)		
Подготовка к промежуточной аттестации (ПА)	6	6
Всего:	48	48

5. Перечень учебно-методического обеспечения

для самостоятельной работы обучающихся по дисциплине (модулю)

Учебно-методические материалы для самостоятельной работы обучающихся указаны в п.п. 7-11.

6. Перечень печатных и электронных учебных изданий

Перечень печатных и электронных учебных изданий приведен в таблице 8.

Таблица 8– Перечень печатных и электронных учебных изданий

Шифр/ URL адрес	Библиографическая ссылка	Количество экземпляров в библиотеке (кроме электронных экземпляров)
URL: https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf	Christopher M. Bishop «Pattern Recognition and Machine Learning»	
URL: https://www.clc.hcmus.edu.vn/wp-content/uploads/2017/11/Hands_On_Machine_Learning_with_Scikit_Learn_and_TensorFlow.pdf	Aurélien Géron "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow"	
URL: https://znanium.com/catalog/product/1908430	Чио, К. Машинное обучение и безопасность: защита систем с помощью данных и алгоритмов: практическое руководство / К. Чио, Д. Фримэн. – Москва: ДМК Пресс, 2020. - 388 с. - ISBN 978-5-97060-713-8. - Текст: электронный. – Режим доступа: по подписке.	
URL: https://znanium.com/catalog/product/2085038	Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных: учебник / П. Флах. - 2-е изд. - Москва: ДМК Пресс, 2023. - 401 с. - ISBN 978-5-89818-300-4. - Текст: электронный. . – Режим доступа: по подписке.	
	Барсегян, А. А., Куприянов М. С., Степаненко В.В., Холод А.А., Методы и модели анализа данных: OLAP и Data Mining. БХВ-Петербург, 2004.	

	- 336 с.	
URL: https://e.lanbook.com/book/131686	Шалев-Шварц Ш. , Бен-Давид Ш. Идеи машинного обучения, ДМК Пресс, 2019	
	<u>Себастьян</u> Рашка, А. <u>В. Логунов</u> Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения, ДМК Пресс, 2017. https://znanium.com/catalog/document?id=341047	
	<u>Коэльо Луис Педро, Вилли Ричарт, Слинкин А. А.</u> Построение систем машинного обучения на языке Python, ДМК Пресс, 2016. https://znanium.com/catalog/document?id=341080	
	В. В. Вьюгин Математические основы машинного обучения и прогнозирования, МЦНМО, 2014, https://znanium.com/catalog/document?id=328779	
ЭБС "ИНФРА-М" http://www.znanium.com ЭБС «Лань» http://www.elan.com	1. J. Han and M. Kamber, Data Mining: Concepts and Techniques, Second Edition, Morgan Kaufman, 2006. (Глава 1) 2. Tom Mitchell, Machine Learning, McGraw Hill, 1997. (Глава 1) 3. Ian H. Witten and Eibe Frank. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, San Francisco, 2 editions, 2005. (Главы 1, 2, 3) 4. Биркгоф Г. Теория решеток. - М.: Наука, 1984.	

	<p>5. Szathmary, L. Symbolic Data Mining Methods with the Coron Platform. PhD thesis, University Henri Poincaré -- Nancy 1, France, Nov 2006. (глава The Coron User Toolkit)</p> <p>6. Janez Demsar, Blaz Zupan, Gregor Leban, Tomaz Curk: Orange: From Experimental Machine Learning to Interactive Data Mining. J.-F. Boulicaut et al. (Eds.): PKDD 2004, LNAI 3202, pp. 537–539, Springer, 2004</p> <p>7. Николенко, С.И., Тулупьев, А.Л., Самообучающиеся системы. – М.: МЦНМО, 2009. – 288 с. (Глава 6)</p> <p>8. L. Zhukov, Spectral Clustering of Large Advertiser Datasets. Technical report, Overture R&D, 2003.</p> <p>9. Автоматическое порождение гипотез в интеллектуальных системах/ Сост. Е.С. Панкратова, В.К. Финн; Под общ. ред. В.К.Финна. – М.: Книжный дом «ЛИБРОКОМ», 2009. – 528 с. (Главы 1,2)</p>	
Библиотеки	Scikit-Learn, TensorFlow, Keras, PyTorch	

7. Перечень электронных образовательных ресурсов
информационно-телекоммуникационной сети «Интернет»

Перечень электронных образовательных ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины приведен в таблице 9.

Таблица 9 – Перечень электронных образовательных ресурсов информационно-телекоммуникационной сети «Интернет»

URL адрес	Наименование
http://www.machinelearning.ru/wiki/index.php?title=ML	Машинное обучение (курс лекций, К.В.Воронцов)
http://www.machinelearning.ru/wiki/index.php?title=29	Машинное обучение (семинары, ВМК МГУ)
http://www.machinelearning.ru/wiki/index.php?title=29	Машинное обучение (курс лекций, Н.Ю.Золотых)
http://www.machinelearning.ru/wiki/index.php?title=29	Машинное обучение (курс

	лекций, СГАУ, С.Лисицын)
https://www.deeplearning.ai/courses/machine-learning-specialization/	Курс "Машинное обучение" (курс лекций Andrew Ng)

8. Перечень информационных технологий

8.1. Перечень программного обеспечения, используемого при осуществлении образовательного процесса по дисциплине.

Перечень используемого программного обеспечения представлен в таблице 10.

Таблица 10– Перечень программного обеспечения

№ п/п	Наименование
	Не предусмотрено

8.2. Перечень информационно-справочных систем, используемых при осуществлении образовательного процесса по дисциплине

Перечень используемых информационно-справочных систем представлен в таблице 11.

Таблица 11– Перечень информационно-справочных систем

№ п/п	Наименование
1.	Электронно-библиотечная система Лань URL: https://e.lanbook.com/
2.	Электронно-библиотечная система Znanium URL: https://znanium.com/
3.	Информационно-правовой портал ГАРАНТ.РУ URL: http://www.garant.ru/
4.	Правовая поддержка КОНСУЛЬТАНТПЛЮС URL: http://www.consultant.ru

9. Материально-техническая база

Состав материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине, представлен в таблице 12.

Таблица 12 – Состав материально-технической базы

№ п/п	Наименование составной части материально-технической базы	Номер аудитории (при необходимости)
1	Аудитория для практических занятий Компьютерный класс	Аудитория укомплектована специализированной мебелью, техническими средствами обучения, служащими для представления учебной информации большой аудитории Ауд. 34-10
2	Аудитории общего пользования (для групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации)	Аудитория укомплектована специализированной мебелью, техническими средствами обучения, служащими для предоставления учебной информации большой аудитории
3	Библиотека, Интернет-класс ГУАП (для самостоятельной работы)	Помещения укомплектованы специализированной мебелью, оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечено

	доступом в электронную информационно-образовательную среду ГУАП
--	---

10. Оценочные средства для проведения промежуточной аттестации

10.1. Состав оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине приведен в таблице 13.

Таблица 13 – Состав оценочных средств для проведения промежуточной аттестации

Вид промежуточной аттестации	Перечень оценочных средств
Зачет	Список вопросов; Тесты;
Выполнение курсовой работы	Экспертная оценка на основе требований к содержанию курсовой работы по дисциплине.

10.2. В качестве критериев оценки уровня сформированности (освоения) компетенций обучающимися применяется 5-балльная шкала оценки сформированности компетенций, которая приведена в таблице 14. В течение семестра может использоваться 100-балльная шкала модульно-рейтинговой системы Университета, правила использования которой, установлены соответствующим локальным нормативным актом ГУАП.

Таблица 14 – Критерии оценки уровня сформированности компетенций

Оценка компетенции	Характеристика сформированных компетенций
5-балльная шкала	
«отлично» «зачтено»	<ul style="list-style-type: none"> – обучающийся глубоко и всесторонне усвоил программный материал; – уверенно, логично, последовательно и грамотно его излагает; – опираясь на знания основной и дополнительной литературы, тесно привязывает усвоенные научные положения с практической деятельностью направления; – умело обосновывает и аргументирует выдвигаемые им идеи; – делает выводы и обобщения; – свободно владеет системой специализированных понятий.
«хорошо» «зачтено»	<ul style="list-style-type: none"> – обучающийся твердо усвоил программный материал, грамотно и по существу излагает его, опираясь на знания основной литературы; – не допускает существенных неточностей; – увязывает усвоенные знания с практической деятельностью направления; – аргументирует научные положения; – делает выводы и обобщения; – владеет системой специализированных понятий.
«удовлетворительно» «зачтено»	<ul style="list-style-type: none"> – обучающийся усвоил только основной программный материал, по существу излагает его, опираясь на знания только основной литературы; – допускает несущественные ошибки и неточности; – испытывает затруднения в практическом применении знаний направления; – слабо аргументирует научные положения; – затрудняется в формулировании выводов и обобщений; – частично владеет системой специализированных понятий.
«неудовлетворительно» «не зачтено»	<ul style="list-style-type: none"> – обучающийся не усвоил значительной части программного материала; – допускает существенные ошибки и неточности при рассмотрении проблем в конкретном направлении; – испытывает трудности в практическом применении знаний;

Оценка компетенции	Характеристика сформированных компетенций
5-балльная шкала	
	<ul style="list-style-type: none"> – не может аргументировать научные положения; – не формулирует выводов и обобщений.

10.3. Типовые контрольные задания или иные материалы.

Вопросы (задачи) для экзамена представлены в таблице 15.

Таблица 15 – Вопросы (задачи) для экзамена

№ п/п	Перечень вопросов (задач) для экзамена	Код индикатора
	Учебным планом не предусмотрено	

Вопросы (задачи) для зачета / дифф. зачета представлены в таблице 16.

Таблица 16 – Вопросы (задачи) для зачета / дифф. зачета

№ п/п	Перечень вопросов (задач) для зачета / дифф. зачета	Код индикатора
1	Объясните понятие "обучение с подкреплением".	УК-2.В.2
2	Что такое модель, параметр и гиперпараметр в машинном обучении?	УК-2.В.3
3	Объясните принцип работы метода главных компонент (PCA). В чем заключается метод t-SNE и где он применяется? Почему снижение размерности важно для работы с высокоразмерными данными?	ПК-6.3.1
4	Какие методы вы знаете для предобработки текстовых данных? Какие задачи решаются с помощью методов обработки естественного языка?	ПК-7.3.1
5	Какие библиотеки для машинного обучения вы использовали и какие их основные особенности? Объясните, как проводится кросс-валидация и зачем она нужна. Какие методы вы знаете для борьбы с переобучением модели? Как интерпретировать результаты модели машинного обучения?	ПК-7.В.1
6	NN). Что такое наивный байесовский классификатор и какие его основные предположения?	ПК-8.3.1
7	Как работают деревья решений и в чем их преимущества и недостатки? Что такое ансамблевые методы и какие их основные виды?	ПК-8.У.1
8	Что такое перцептрон и как он работает? Объясните архитектуру многослойного перцептрона (MLP). Какие функции активации вы знаете и каковы их особенности? Объясните, что такое обратное распространение ошибки (backpropagation).	ПК-8.В.1

Перечень тем для курсового проектирования/выполнения курсовой работы представлены в таблице 17.

Таблица 17 – Перечень тем для курсового проектирования/выполнения курсовой работы

№ п/п	Примерный перечень тем для курсового проектирования/выполнения курсовой работы
1	Морфологический анализ текстов при помощи машинного обучения (на материале одного или нескольких языков)
2	Система разработки данных и машинного обучения с помощью программных средств Анализа Формальных Понятий
3	Частичные порядки и решетки. Диаграммы порядка. Объектно-признаковые данные и формальные контексты. Операторы Галуа. Решетки формальных понятий. Признаковые импликации
4	Concept Explorer – программная система анализа данных на основе АФП.
5	Согон – система поиска частых множеств признаков и ассоциативных правил.
6	Информационная энтропия и прирост информации. Алгоритмы ID3 и C4.5.
7	Критерии остановки и отсеечения.
8	Меры и методы оценки качества обучения (скользящий контроль).
9	Системы машинного обучения WEKA и Orange и ее средства для работы с деревьями решений. Средства оценки качества классификации в этих системах.
10	Методы кластеризации в системах Weka и Orange.
11	Спектральная кластеризация средствами пакета MatLab и библиотеки NumPy для языка Python.
12	Rough Set Exploration System (RSES) – программная система для исследования объектно-признаковых данных на основе неточных множеств.
13	QUDA – система разработки данных и машинного обучения. ДСМ-метод в системе QUDA
14	Наивная байесовская классификация в системе Orange. Реализация наивной байесовской классификации на языке Python.

Вопросы для проведения промежуточной аттестации в виде тестирования представлены в таблице 18.

Таблица 18 – Примерный перечень вопросов для тестов

№ п/п	Примерный перечень вопросов для тестов	Код индикатора
1	<p>What is the purpose of data normalization?</p> <ul style="list-style-type: none"> A. To remove outliers from the dataset B. To convert data to a standard scale without distorting differences in the ranges of values C. To increase the number of features in the dataset D. To reduce the number of missing values 	<p>УК-2. В. 2 УК-2. В. 3</p>
2	<p>Which technique is used to handle missing data?</p> <ul style="list-style-type: none"> A. One-hot encoding B. Imputation C. Standardization D. Feature scaling 	<p>УК-2. В. 2 УК-2. В. 3</p>
3	<p>What does one-hot encoding do?</p>	<p>УК-2. В. 2 УК-2. В. 3</p>

	<ul style="list-style-type: none"> • A. Normalizes the data • B. Encodes categorical variables as binary vectors • C. Imputes missing values • D. Reduces dimensionality 	
4	<p>Which library in Python is commonly used for data manipulation and analysis?</p> <ul style="list-style-type: none"> • A. NumPy • B. Pandas • C. Matplotlib • D. Scikit-learn 	YK-2. B. 2 YK-2. B. 3
5	<p>What is feature engineering?</p> <ul style="list-style-type: none"> ○ A. The process of creating new features from existing ones to improve model performance ○ B. The process of selecting the best model for a given dataset ○ C. The process of training a model on a dataset ○ D. The process of splitting the data into training and testing sets 	YK-2. B. 2 YK-2. B. 3
6	<p>Which algorithm is used for regression tasks?</p> <ul style="list-style-type: none"> ○ A. k-Nearest Neighbors (k-NN) ○ B. Linear Regression ○ C. Decision Trees ○ D. Naive Bayes 	YK-2. B. 2 YK-2. B. 3
7	<p>Which of the following is a tree-based ensemble method?</p> <ul style="list-style-type: none"> • A. k-Means Clustering • B. Principal Component Analysis (PCA) • C. Random Forest • D. Support Vector Machine (SVM) 	YK-2. B. 2 YK-2. B. 3
8	<p>What is the key idea behind Support Vector Machines (SVM)?</p> <ul style="list-style-type: none"> • A. To minimize the distance between points and their cluster center • B. To maximize the margin between different classes • C. To reduce dimensionality of the data • D. To calculate the probability of class membership 	YK-2. B. 2 YK-2. B. 3
9	<p>Which of the following is true about Naive Bayes classifiers?</p> <ul style="list-style-type: none"> • A. They are based on the k-nearest neighbors algorithm • B. They assume that features are conditionally independent given the class label • C. They are always better than decision trees • D. They cannot handle categorical data 	YK-2. B. 2 YK-2. B. 3
10	<p>Which algorithm is commonly used for clustering tasks?</p> <ul style="list-style-type: none"> • A. k-Means • B. Linear Regression • C. Logistic Regression • D. Random Forest 	YK-2. B. 2 YK-2. B. 3

11	<p>What is cross-validation used for in machine learning?</p> <ul style="list-style-type: none"> • A. To train the model • B. To evaluate the model's performance on unseen data • C. To reduce the number of features • D. To increase the dataset size 	<p>УК-2. Б. 2 УК-2. Б. 3</p>
12	<p>What does a confusion matrix show in classification tasks?</p> <ul style="list-style-type: none"> • A. The distribution of data points • B. The relationship between true and predicted classes • C. The correlation between features • D. The dimensionality of the data 	<p>УК-2.Б.2 УК-2. Б.3</p>
13	<p>What metric would you use to evaluate a regression model?</p> <ul style="list-style-type: none"> • A. Accuracy • B. Precision • C. Recall • D. Mean Squared Error (MSE) 	<p>УК-2.Б.2 УК-2.Б.3</p>
14	<p>What is the purpose of a validation set?</p> <ul style="list-style-type: none"> ○ A. To train the model ○ B. To test the model's performance during training ○ C. To fine-tune hyperparameters ○ D. To store unprocessed data 	<p>УК-2. Б.2 УК-2. Б.3</p>
15	<p>What is the ROC curve used for?</p> <ul style="list-style-type: none"> • A. To visualize the performance of a regression model • B. To visualize the trade-off between true positive rate and false positive rate • C. To reduce the dimensionality of data • D. To increase the number of features in the dataset 	<p>УК-2. Б.2 УК-2. Б.3</p>
16	<p>Как называется процесс преобразования текста на естественном языке в формат, пригодный для обработки алгоритмами?</p> <ul style="list-style-type: none"> ○ A. Компиляция ○ B. Токенизация ○ C. Парсинг ○ D. Нормализация 	<p>ПК-6. 3. 1</p>
16	<p>Какой метод машинного обучения часто используется для анализа тональности текста?</p> <ul style="list-style-type: none"> ○ A. K-means ○ B. Naive Bayes ○ C. PCA ○ D. KNN 	<p>ПК-6. 3. 1</p>
17	<p>Что такое word2vec в контексте обработки естественного языка?</p> <ul style="list-style-type: none"> ○ A. Метод для кластеризации текстов ○ B. Алгоритм для уменьшения размерности текстовых данных ○ C. Метод для представления слов в виде векторных представлений 	<p>ПК-6. 3. 1</p>

	<ul style="list-style-type: none"> ○ D. Техника для предсказания следующего слова в предложении 	
18	<p>Как называется метод, использующийся для выявления синтаксической структуры предложения?</p> <ul style="list-style-type: none"> ○ A. Лемматизация ○ B. Парсинг ○ C. Стемминг ○ D. Токенизация 	ПК-6. 3. 1
19	<p>Что такое нейронные сети в контексте обработки естественного языка?</p> <ul style="list-style-type: none"> ○ A. Методы для уменьшения размерности текстовых данных ○ B. Алгоритмы, которые моделируют работу человеческого мозга для обработки текста ○ C. Программы для перевода текста с одного языка на другой ○ D. Базы данных для хранения текстовой информации 	ПК-6. 3. 1
20	<p>Как называется архитектура нейронных сетей, специально предназначенная для обработки последовательностей данных, таких как текст?</p> <ul style="list-style-type: none"> ○ A. Convolutional Neural Networks (CNN) ○ B. Recurrent Neural Networks (RNN) ○ C. Fully Connected Networks ○ D. Generative Adversarial Networks (GAN) 	ПК-6. 3. 1
21	<p>Какой метод используется для преобразования текста в последовательности чисел для дальнейшей обработки нейронными сетями?</p> <ul style="list-style-type: none"> ○ A. One-Hot Encoding ○ B. Min-Max Scaling ○ C. Standardization ○ D. Data Augmentation 	ПК-6. 3. 1
22	<p>Какой подход используется в системах машинного перевода, основанных на нейронных сетях?</p> <ul style="list-style-type: none"> ○ A. Рулевой перевод (Rule-based translation) ○ B. Статистический машинный перевод (Statistical Machine Translation, SMT) ○ C. Нейронный машинный перевод (Neural Machine Translation, NMT) ○ D. Фразовый перевод (Phrase-based translation) 	ПК-6. 3. 1
23	<p>Что такое трансформер (Transformer) в контексте обработки естественного языка?</p> <ul style="list-style-type: none"> ○ A. Метод для кластеризации текстов ○ B. Архитектура нейронных сетей для обработки последовательностей данных ○ C. Алгоритм для уменьшения размерности 	ПК-6. 3. 1

	<p>текстовых данных</p> <ul style="list-style-type: none"> ○ D. Программа для предсказания следующего слова в предложении 	
24	<p>Какой алгоритм используется для создания больших языковых моделей, таких как GPT (Generative Pre-trained Transformer)?</p> <ul style="list-style-type: none"> ○ A. SVM ○ B. CNN ○ C. RNN ○ D. Transformer 	ПК-6. 3. 1
25	<p>Что такое лемматизация в контексте обработки естественного языка?</p> <ul style="list-style-type: none"> ○ A. Процесс разделения текста на отдельные слова ○ B. Преобразование слов к их базовой или начальной форме ○ C. Удаление стоп-слов из текста ○ D. Определение части речи для каждого слова 	ПК-6. 3. 1
26	<p>Что такое машинный перевод?</p> <ul style="list-style-type: none"> ○ A. Автоматическое создание текстов на естественном языке ○ B. Процесс перевода текста с одного языка на другой с помощью алгоритмов ○ C. Определение части речи для каждого слова в тексте ○ D. Объединение текстов из нескольких источников 	ПК-6. 3. 1
27	<p>Какой подход используется для улучшения качества перевода в нейронных системах машинного перевода?</p> <ul style="list-style-type: none"> ○ A. Использование синтаксических правил ○ B. Обучение на большом объеме параллельных текстов ○ C. Применение токенизации и лемматизации ○ D. Удаление всех стоп-слов 	ПК-6. 3. 1
28	<p>Какой метод часто используется для предсказания следующего слова в предложении?</p> <ul style="list-style-type: none"> ○ A. K-means ○ B. Naive Bayes ○ C. N-grams ○ D. PCA 	ПК-6. 3. 1
29	<p>Опишите основные подходы к когнитивному моделированию естественного языка. Какие когнитивные процессы лежат в основе понимания и производства речи?</p>	ПК-6. 3. 1
30	<p>Назовите формальные методы, которые используются для моделирования естественного языка? Приведите примеры.</p>	ПК-6. 3. 1
31	<p>Опишите принципы построения больших языковых моделей (LLM). Как они обучаются?</p>	ПК-6. 3. 1
32	<p>Приведите примеры и кратко опишите работу системы обработки естественного языка и машинного перевода.</p>	ПК-6. 3. 1
33	<p>Опишите в общих чертах, какие принципы функционирования</p>	ПК-6. 3. 1

	нейросетей важны для понимания систем машинного перевода.	
34	<p>Какой метод используется для удаления стоп-слов в библиотеке NLTK?</p> <ul style="list-style-type: none"> • a) <code>nltk.remove_stopwords()</code> • b) <code>nltk.corpus.stopwords.words('english')</code> • c) <code>nltk.clean_stopwords()</code> • d) <code>nltk.tokenize_stopwords()</code> 	ПК-7. 3. 1
35	<p>Какая библиотека Python используется для обработки естественного языка?</p> <ul style="list-style-type: none"> ○ a) NumPy ○ b) Pandas ○ c) NLTK ○ d) Matplotlib 	ПК-7. 3. 1
36	<p>Какой метод в библиотеке SpaCy используется для токенизации текста?</p> <ul style="list-style-type: none"> ○ a) <code>tokenize()</code> ○ b) <code>spacy.tokenize()</code> ○ c) <code>nlp()</code> ○ d) <code>text.split()</code> 	ПК-7. 3. 1
37	<p>Какой из следующих методов используется для лемматизации в NLTK?</p> <ul style="list-style-type: none"> ○ a) <code>WordNetLemmatizer</code> ○ b) <code>PorterStemmer</code> ○ c) <code>LancasterStemmer</code> ○ d) <code>SnowballStemmer</code> 	ПК-7. 3. 1
38	<p>Какое назначение у TF-IDF (term frequency-inverse document frequency) в анализе текста?</p> <ul style="list-style-type: none"> ○ a) Определение синтаксических структур ○ b) Извлечение ключевых слов ○ c) Перевод текста ○ d) Лемматизация текста 	ПК-7. 3. 1
39	<p>Какая из следующих библиотек предназначена для работы с векторами слов в Python?</p> <ul style="list-style-type: none"> ○ a) Matplotlib ○ b) Gensim ○ c) Pandas ○ d) Scikit-learn 	ПК-7. 3. 1
40	<p>Опишите процесс предобработки текста для анализа, включающий удаление стоп-слов, токенизацию и лемматизацию. Приведите пример кода на Python с использованием библиотек NLTK или SpaCy.</p>	ПК-7. 3. 1
41	<p>Объясните принцип работы TF-IDF и как он используется для выделения ключевых слов в тексте.</p>	ПК-7. 3. 1

42	Приведите пример использования библиотеки Gensim для обучения модели Word2Vec на текстовом корпусе.	ПК-7. В. 1
43	Опишите основные этапы построения синтаксического анализатора текста с использованием библиотеки SpaCy.	ПК-7. В. 1
44	Объясните, как можно использовать машинное обучение для автоматической классификации текстов. Приведите пример алгоритма и кода.	ПК-7. В. 1
45	Проанализируйте результаты работы кода и объясните, что представляет собой каждый из параметров, выводимых для каждого токена (например, <code>token.text</code> , <code>token.lemma_</code> и т.д.).	ПК-7. В. 1
46	Опишите процесс лемматизации и его важность в обработке текстовых данных.	ПК-7. В. 1
47	Какие шаги включены в предобработку текста перед анализом с использованием алгоритмов машинного обучения?	ПК-7. В. 1
48	Объясните различия между TF-IDF и Bag-of-Words моделями.	ПК-7. В. 1
49	Как нейронные сети используются для обработки естественного языка? Приведите примеры применения.	ПК-7. В. 1
50	Опишите процесс создания и использования языковой модели на основе нейросетей для задач машинного перевода.	ПК-7. В. 1
51	Объясните, что представляет собой ER-диаграмма и как она используется при проектировании баз данных.	ПК-8. З. 1
52	Опишите процесс создания и чтения блок-схем, включая основные символы и их значение.	ПК-8. З. 1
53	Проиллюстрируйте, какие основные компоненты включаются в архитектурную документацию информационной системы? Приведите примеры.	ПК-8. З. 1
54	Опишите, как диаграммы последовательности используются для моделирования взаимодействия между объектами в программной системе.	ПК-8. З. 1
55	Объясните, каким образом стандарты, такие как ANSI Y14.5, помогают в создании и чтении чертежей аппаратных компонентов.	ПК-8. З. 1
56	Как называется графическое представление структур данных и алгоритмов, используемое для документирования информационных систем? а) Диаграмма состояния б) Блок-схема в) ER-диаграмма г) Диаграмма классов	ПК-8. З. 1
57	Какой стандарт используется для создания и чтения чертежей электронных схем? а) IEEE 802.11 б) ISO 9001 в) ANSI Y14.5 г) UML	ПК-8. З. 1
58	Какой тип документации описывает взаимодействие между различными компонентами программного обеспечения?	ПК-8. З. 1

	a) Техническое задание b) Пользовательская документация c) Схема последовательности d) Архитектурная документация	
59	<p>Что такое DFD (Data Flow Diagram)?</p> a) Диаграмма потока данных b) Диаграмма взаимодействия c) Диаграмма сущность-связь d) Диаграмма состояний	ПК-8. 3. 1
60	<p>Как называется процесс создания схем и чертежей, описывающих физическую и логическую организацию информационной системы?</p> a) Моделирование b) Программирование c) Кодирование d) Архитектурное проектирование	ПК-8. У. 1
61	<p>Какой тип документации включает описание требований к функциональности и производительности программного обеспечения?</p> <ul style="list-style-type: none"> • a) Пользовательская документация • b) Техническая документация • c) Спецификация требований • d) Дизайн-документация 	ПК-8. У. 1
62	<p>Какой метод используется для анализа и визуализации последовательностей выполнения алгоритмов и программных процессов?</p> <ul style="list-style-type: none"> • a) Блок-схема • b) ER-диаграмма • c) Диаграмма классов • d) Диаграмма активностей 	ПК-8. У. 1
63	<p>Какой тип диаграммы используется для описания структурных элементов и их отношений в программном обеспечении?</p> a) Диаграмма потоков данных b) Диаграмма классов c) Диаграмма активностей d) Диаграмма последовательностей	ПК-8. У. 1
64	<p>Какой термин описывает метод структурного анализа, включающий диаграммы классов и диаграммы последовательности?</p> a) CASE b) PERT c) UML d) SSADM	ПК-8. У. 1
65	<p>Какой метод позволяет улучшить читаемость и точность документации путем стандартизации терминологии и оформления?</p> a) Использование естественного языка b) Применение диаграмм c) Документирование в формате ISO	ПК-8. В. 1

	d) Автоматизация документирования	
66	Как называется процесс, в результате которого чертежи и схемы приводятся в соответствие с требованиями стандарта? a) Сертификация b) Нормирование c) Форматирование d) Кодификация	ПК-8. В. 1
67	Что из перечисленного относится к аппаратным компонентам информационной системы? a) Операционная система b) Сервер c) База данных d) Программное обеспечение	ПК-8. В. 1
68	Как называется схема, которая показывает компоненты системы и связи между ними? a) Диаграмма классов b) Диаграмма компонент c) Диаграмма развертывания d) Диаграмма объектов	ПК-8. В. 1
69	Какой метод проверки программного обеспечения используется для обнаружения и исправления ошибок в коде? a) Компиляция b) Дебаггинг c) Интерпретация d) Верификация	ПК-8. В. 1
70	Что из перечисленного является основным компонентом технической документации по программным системам? a) Рекламный буклет b) Пользовательское соглашение c) Техническое задание (ТЗ) d) Мануал по установке	ПК-8. В. 1
	Диагностическая работа	
1	<p>Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа:</p> <p>Методы формального и когнитивного моделирования естественного языка играют ключевую роль в разработке систем обработки текста и речи. Формальные модели включают использование математических и логических методов для описания грамматических структур и синтаксических правил языка. В то же время когнитивное моделирование учитывает ментальные процессы и структуру человеческого мышления, что позволяет создавать модели, более близкие к естественной обработке языка человеком. Создание метаязыков, таких как регулярные выражения или грамматики контекстно-свободных языков, также является важным инструментом в описании и анализе языковых данных.</p> <p>Какой из следующих методов наиболее подходит для создания формальных моделей синтаксиса языка?</p> <p>a) Использование регулярных выражений</p>	ПК-6

	<p>b) Анализ ментальных процессов, связанных с обработкой языка</p> <p>c) Моделирование семантических полей</p> <p>d) Создание моделей машинного перевода</p>											
2	<p>Прочитайте текст, выберите правильные варианты ответа и запишите аргументы, обосновывающие выбор ответов.</p> <p>В современном лингвистическом анализе широко применяются методы формального и когнитивного моделирования. Формальное моделирование включает в себя использование математических и логических инструментов для описания грамматик и синтаксических структур языка. Такие методы, как конечные автоматы и контекстно-свободные грамматики, позволяют строить модели, которые формально описывают структуру языка и его синтаксические правила.</p> <p>С другой стороны, когнитивное моделирование сосредоточено на понимании того, как человеческий мозг воспринимает и обрабатывает языковую информацию. Здесь часто применяются искусственные нейронные сети, которые имитируют работу человеческого мозга, обучаясь на больших объемах текстов и выявляя скрытые семантические и синтаксические зависимости.</p> <p>Какие методы наиболее соответствуют задачам формального и когнитивного моделирования языка?</p> <p>a) Использование конечных автоматов для описания регулярных языков</p> <p>b) Применение метода наивного байесовского классификатора для анализа текста</p> <p>c) Построение контекстно-свободных грамматик для описания синтаксиса языка</p> <p>d) Применение искусственных нейронных сетей для моделирования процессов восприятия языка</p>	ПК-6										
3	<p>Прочитайте текст и установите соответствие. К каждой позиции, данной в левом столбце, подберите соответствующую позицию в правом столбце.</p> <p>Моделирование естественного языка требует использования различных методов и подходов, которые помогают анализировать и обрабатывать текстовую информацию.</p> <p>Установите соответствие между методами и их описанием:</p> <table><thead><tr><th>Метод</th><th>Описание</th></tr></thead><tbody><tr><td>A. Контекстно-свободные грамматики</td><td>1. Инструмент для поиска и сопоставления шаблонов в текстах, обычно используемый для простых синтаксических операций.</td></tr><tr><td>B. Искусственные нейронные сети</td><td>2. Статистический метод для классификации текста на основе вероятностных предположений.</td></tr><tr><td>C. Регулярные выражения</td><td>3. Метод, использующий математические модели для описания грамматических структур языка.</td></tr><tr><td>D. Наивный байесовский классификатор</td><td>4. Метод, позволяющий моделировать процессы обработки информации, сходные с теми, что происходят в человеческом мозге.</td></tr></tbody></table>	Метод	Описание	A. Контекстно-свободные грамматики	1. Инструмент для поиска и сопоставления шаблонов в текстах, обычно используемый для простых синтаксических операций.	B. Искусственные нейронные сети	2. Статистический метод для классификации текста на основе вероятностных предположений.	C. Регулярные выражения	3. Метод, использующий математические модели для описания грамматических структур языка.	D. Наивный байесовский классификатор	4. Метод, позволяющий моделировать процессы обработки информации, сходные с теми, что происходят в человеческом мозге.	ПК-6
Метод	Описание											
A. Контекстно-свободные грамматики	1. Инструмент для поиска и сопоставления шаблонов в текстах, обычно используемый для простых синтаксических операций.											
B. Искусственные нейронные сети	2. Статистический метод для классификации текста на основе вероятностных предположений.											
C. Регулярные выражения	3. Метод, использующий математические модели для описания грамматических структур языка.											
D. Наивный байесовский классификатор	4. Метод, позволяющий моделировать процессы обработки информации, сходные с теми, что происходят в человеческом мозге.											

4	<p>Прочитайте текст и установите последовательность. Запишите соответствующую последовательность букв слева направо:</p> <p>Моделирование естественного языка с использованием методов машинного обучения требует соблюдения определенной последовательности действий. Этот процесс включает в себя несколько ключевых этапов.</p> <p>Установите правильную последовательность действий при моделировании естественного языка с использованием машинного обучения:</p> <ul style="list-style-type: none"> • А. Тестирование модели • В. Сбор данных • С. Обучение модели • D. Предобработка данных • Е. Оптимизация модели 	ПК-6
5	<p>Прочитайте текст и запишите развернутый обоснованный ответ.</p> <p>Машинное обучение становится все более важным инструментом в лингвистике, особенно в области обработки естественного языка (NLP). Одним из ключевых аспектов является использование алгоритмов глубокого обучения, таких как нейронные сети, которые способны анализировать огромные объемы текстовых данных и выявлять сложные взаимосвязи между словами и фразами. Эти модели могут быть использованы для задач, таких как автоматический перевод, распознавание речи, анализ тональности текста и генерация текста. Однако успешное применение этих технологий требует правильной настройки и оптимизации моделей, включая выбор архитектуры сети, обучение на большом количестве данных, регуляризацию для предотвращения переобучения, и использование современных методов оценки, таких как кросс-валидация.</p> <p>Почему настройка и оптимизация моделей машинного обучения, используемых в обработке естественного языка, играют ключевую роль в успешности их применения?</p>	ПК-6
6	<p>Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа:</p> <p>Вы анализируете лингвистический корпус текстов с целью построения модели автоматического перевода. Корпус включает параллельные тексты на двух языках. Для достижения лучшего результата вы решили использовать статистические методы машинного перевода (SMT), основанные на фразовых соответствиях и их вероятностях. Перед вами стоит задача выбора подходящего метода оценки качества перевода, который будет использоваться для сравнения различных моделей. Существуют такие метрики, как BLEU (Bilingual Evaluation Understudy), METEOR, TER (Translation Error Rate) и ROUGE. Каждая из них имеет свои особенности и применимость в различных ситуациях.</p> <p>Выберите правильную метрику для оценки качества машинного перевода в данном контексте и запишите аргументы,</p>	ПК-7

	<p>обосновывающие ваш выбор:</p> <p>A. BLEU B. METEOR C. TER D. ROUGE</p>											
7	<p>Прочитайте текст, выберите правильные варианты ответа и запишите аргументы, обосновывающие выбор ответов.</p> <p>При разработке системы машинного перевода перед вами встает вопрос выбора подходящих методов и инструментов для предобработки текста и оценки качества перевода. Вы работаете с параллельным корпусом текстов на английском и русском языках. Основная цель — повысить точность перевода, особенно в условиях вариативности лексики и синтаксиса. Вы можете использовать различные методы предобработки данных, такие как лемматизация, стемминг, и оценочные метрики, такие как BLEU, METEOR, и ROUGE. Также необходимо учесть возможность применения метода перекрестной проверки (cross-validation) для повышения надежности модели.</p> <p>Какие методы предобработки текста наиболее целесообразно использовать?</p> <p>A. Лемматизация B. Стемминг C. Токенизация D. Удаление стоп-слов</p>	ПК-7										
8	<p>Прочитайте текст и установите соответствие. К каждой позиции, данной в левом столбце, подберите соответствующую позицию в правом столбце.</p> <p>Вы разрабатываете модель машинного перевода, использующую методы машинного обучения. Основными этапами в процессе разработки являются: сбор и предобработка данных, выбор алгоритма, обучение модели, оценка качества модели и ее оптимизация. На каждом этапе используются специфические методы и инструменты, которые наиболее подходят для выполнения поставленных задач.</p> <p>Установите соответствие между этапами разработки модели и применяемыми на них методами.</p> <table><tr><td>Этапы разработки модели</td><td>Применяемые на них методы</td></tr><tr><td>1. Сбор и предобработка данных</td><td>A. Градиентный бустинг</td></tr><tr><td>2. Выбор алгоритма</td><td>B. Токенизация и лемматизация</td></tr><tr><td>3. Обучение модели</td><td>C. BLEU, METEOR</td></tr><tr><td>4. Оценка качества модели</td><td>D. Рекуррентные нейронные сети (RNN)</td></tr></table>	Этапы разработки модели	Применяемые на них методы	1. Сбор и предобработка данных	A. Градиентный бустинг	2. Выбор алгоритма	B. Токенизация и лемматизация	3. Обучение модели	C. BLEU, METEOR	4. Оценка качества модели	D. Рекуррентные нейронные сети (RNN)	ПК-7
Этапы разработки модели	Применяемые на них методы											
1. Сбор и предобработка данных	A. Градиентный бустинг											
2. Выбор алгоритма	B. Токенизация и лемматизация											
3. Обучение модели	C. BLEU, METEOR											
4. Оценка качества модели	D. Рекуррентные нейронные сети (RNN)											
9	<p>Прочитайте текст и установите последовательность. Запишите соответствующую последовательность букв слева направо:</p> <p>При разработке модели машинного перевода необходимо последовательно выполнить несколько ключевых этапов.</p> <p>Установите правильную последовательность этапов разработки модели машинного перевода:</p>	ПК-7										

	<p>A. Оценка качества модели</p> <p>B. Сбор и подготовка данных</p> <p>C. Выбор алгоритма</p> <p>D. Обучение модели</p> <p>E. Оптимизация модели</p>	
10	<p>Прочитайте текст и запишите развернутый обоснованный ответ на тему: "Роль математико-статистических методов в автоматической обработке лингвистических данных".</p> <p>В последнее время машинное обучение и автоматическая обработка лингвистических данных становятся всё более важными в различных областях, включая анализ текста, обработку естественного языка (NLP) и создание интеллектуальных систем. Методы машинного обучения применяются для различных задач, таких как классификация текстов, извлечение информации, и автоматический перевод. Одним из ключевых аспектов успешной обработки текстов является правильное использование математико-статистических методов, таких как статистическое моделирование, анализ частотности и методы снижения размерности. Эти методы помогают в создании эффективных алгоритмов для обработки и анализа текстовых данных.</p>	ПК-7
11	<p>Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа:</p> <p>В области лингвистического обеспечения информационных систем важной задачей является правильное распознавание и обработка текста. Одной из ключевых задач является извлечение информации, которая позволяет выделить полезные данные из неструктурированного текста. Для этой цели могут использоваться различные методы машинного обучения.</p> <ol style="list-style-type: none"> 1. Модели на основе правил 2. Статистические модели 3. Глубокие нейронные сети 4. Метод опорных векторов <p>Вопрос: Какой метод лучше всего подходит для извлечения информации из неструктурированного текста и почему?</p>	ПК-8
12	<p>Прочитайте текст, выберите правильные варианты ответа и запишите аргументы, обосновывающие выбор ответов.</p> <p>В области обработки естественного языка (NLP) и лингвистического обеспечения информационных систем используются различные методы и алгоритмы для обработки текстовых данных. Основные подходы включают:</p> <ol style="list-style-type: none"> 1. Частотный анализ 2. Модели на основе правил 3. Статистические методы 4. Модели глубокого обучения <p>Вопрос: Какие методы являются наиболее эффективными для обработки больших объемов текста и почему?</p>	ПК-8

13	<p>Прочитайте текст и установите соответствие. К каждой позиции, данной в левом столбце, подберите соответствующую позицию в правом столбце.</p> <p>Вам нужно установить соответствие между методами обработки текста и их описаниями. Соедините каждую позицию из левого столбца с подходящим описанием из правого столбца.</p> <table><tr><th>Методы обработки текста</th><th>Описание</th></tr><tr><td>1. Частотный анализ</td><td>А. Метод, который использует заранее заданные правила и шаблоны для обработки и интерпретации текста. Эффективен в узких областях с четко определенными правилами.</td></tr><tr><td>2. Модели на основе правил</td><td>В. Метод, использующий глубокие нейронные сети, такие как трансформеры, для захвата сложных контекстов и взаимосвязей в текстах. Обеспечивает высокую точность в задачах извлечения информации и генерации текста.</td></tr><tr><td>3. Статистическое моделирование</td><td>С. Метод, основанный на подсчете частоты появления слов и фраз в текстах, часто используется для выделения ключевых терминов.</td></tr><tr><td>4. Глубокие нейронные сети</td><td>Д. Метод, который применяет алгоритмы, такие как скрытые марковские процессы и латентное семантическое анализирование, для выявления скрытых тем и зависимостей в тексте.</td></tr></table>	Методы обработки текста	Описание	1. Частотный анализ	А. Метод, который использует заранее заданные правила и шаблоны для обработки и интерпретации текста. Эффективен в узких областях с четко определенными правилами.	2. Модели на основе правил	В. Метод, использующий глубокие нейронные сети, такие как трансформеры, для захвата сложных контекстов и взаимосвязей в текстах. Обеспечивает высокую точность в задачах извлечения информации и генерации текста.	3. Статистическое моделирование	С. Метод, основанный на подсчете частоты появления слов и фраз в текстах, часто используется для выделения ключевых терминов.	4. Глубокие нейронные сети	Д. Метод, который применяет алгоритмы, такие как скрытые марковские процессы и латентное семантическое анализирование, для выявления скрытых тем и зависимостей в тексте.	ПК-8
Методы обработки текста	Описание											
1. Частотный анализ	А. Метод, который использует заранее заданные правила и шаблоны для обработки и интерпретации текста. Эффективен в узких областях с четко определенными правилами.											
2. Модели на основе правил	В. Метод, использующий глубокие нейронные сети, такие как трансформеры, для захвата сложных контекстов и взаимосвязей в текстах. Обеспечивает высокую точность в задачах извлечения информации и генерации текста.											
3. Статистическое моделирование	С. Метод, основанный на подсчете частоты появления слов и фраз в текстах, часто используется для выделения ключевых терминов.											
4. Глубокие нейронные сети	Д. Метод, который применяет алгоритмы, такие как скрытые марковские процессы и латентное семантическое анализирование, для выявления скрытых тем и зависимостей в тексте.											
14	<p>Прочитайте текст и установите последовательность. Запишите соответствующую последовательность букв слева направо:</p> <p>Рассмотрим процесс обработки текста в системе машинного обучения. Вам нужно установить правильный порядок этапов, чтобы корректно выполнить анализ текста и извлечение информации. Определите последовательность этапов обработки текста, основываясь на их логическом порядке.</p> <ol style="list-style-type: none">1. Обучение модели2. Применение модели3. Извлечение признаков4. Предварительная обработка текста5. Оценка модели	ПК-8										
15	<p>Прочитайте текст и запишите развернутый обоснованный ответ.</p> <p>В последние годы методы обработки естественного языка (NLP) значительно развились, и многие современные системы используют различные подходы для улучшения качества анализа текста. Одним из ключевых аспектов является выбор между</p>	ПК-8										

	статистическими методами и методами глубокого обучения. Какой подход — статистические методы или методы глубокого обучения — более эффективен для обработки текстовых данных и почему?	
--	---	--

Примечание: Система оценивания тестовых заданий:

1 тип) Задание комбинированного типа с выбором одного верного ответа из четырех предложенных и обоснованием выбора считается верным, если правильно указана цифра и приведены конкретные аргументы, используемые при выборе ответа. Смысловое совпадение с верным ответом оценивается 1 баллом, неверный ответ или его отсутствие — 0 баллов. *Полного совпадения быть не может, т. к. ответ должен быть на английском языке.*

2 тип) Задание комбинированного типа с выбором нескольких вариантов ответа из предложенных и развернутым обоснованием выбора считается верным, если правильно указаны цифры и приведены конкретные аргументы, используемые при выборе ответов. Полное совпадение с верным ответом оценивается 1 баллом, если допущены ошибки или ответ отсутствует — 0 баллов.

3 тип) Задание закрытого типа на установление соответствия считается верным, если установлены все соответствия (позиции из одного столбца верно сопоставлены с позициями другого столбца). Смысловое совпадение с верным ответом оценивается 1 баллом, неверный ответ или его отсутствие — 0 баллов. *Полного совпадения быть не может, т. к. ответ должен быть на английском языке.*

4 тип) Задание закрытого типа на установление последовательности считается верным, если правильно указана вся последовательность цифр. Полное совпадение с верным ответом оценивается 1 баллом, если допущены ошибки или ответ отсутствует — 0 баллов.

5 тип) Задание открытого типа с развернутым ответом считается верным, если ответ совпадает по смыслу с эталонным по содержанию и полноте. *Полного совпадения быть не может, т. к. ответ должен быть на английском языке.*

Правильный ответ за задание оценивается в 3 балла, если допущена одна ошибка \ неточность \ ответ правильный, но не полный - 1 балл, если допущено более 1 ошибки \ ответ неправильный \ ответ отсутствует — 0 баллов.

Перечень тем контрольных работ по дисциплине обучающихся заочной формы обучения, представлены в таблице 19.

Таблица 19 – Перечень контрольных работ

№ п/п	Перечень контрольных работ
	Не предусмотрено

10.4. Методические материалы, определяющие процедуры оценивания индикаторов, характеризующих этапы формирования компетенций, содержатся в локальных нормативных актах ГУАП, регламентирующих порядок и процедуру проведения текущего контроля успеваемости и промежуточной аттестации обучающихся ГУАП.

11. Методические указания для обучающихся по освоению дисциплины

11.1. Методические указания для обучающихся по освоению лекционного материала.

Основное назначение лекционного материала – логически стройное, системное, глубокое и ясное изложение учебного материала. Назначение современной лекции в рамках дисциплины не в том, чтобы получить всю информацию по теме, а в освоении фундаментальных проблем дисциплины, методов научного познания, новейших достижений научной мысли. В учебном процессе лекция выполняет методологическую, организационную и информационную функции. Лекция раскрывает понятийный аппарат конкретной области знания, её проблемы, дает цельное представление о дисциплине, показывает взаимосвязь с другими дисциплинами.

Планируемые результаты при освоении обучающимися лекционного материала:

- ~ получение современных, целостных, взаимосвязанных знаний, уровень которых определяется целевой установкой к каждой конкретной теме;
- ~ получение опыта творческой работы совместно с преподавателем;
- ~ развитие профессионально-деловых качеств, любви к предмету и самостоятельного творческого мышления.
- ~ появление необходимого интереса, необходимого для самостоятельной работы;
- ~ получение знаний о современном уровне развития науки и техники и о прогнозе их развития на ближайшие годы;
- ~ научиться методически обрабатывать материал (выделять главные мысли и положения, приходить к конкретным выводам, повторять их в различных формулировках);
- ~ получение точного понимания всех необходимых терминов и понятий.

Лекционный материал может сопровождаться демонстрацией слайдов и использованием раздаточного материала при проведении коротких дискуссий об особенностях применения отдельных тематик по дисциплине.

Структура предоставления лекционного материала:

- ~ Постановка проблемы;
- ~ Обзор предметной области;
- ~ Обзор литературы;
- ~ Разработка проекта решения;
- ~ Реализация и анализ результатов;
- ~ Дальнейшие направления исследований.

11.2. Методические указания для обучающихся по прохождению практических занятий

Практическое занятие является одной из основных форм организации учебного процесса, заключающаяся в выполнении обучающимися под руководством преподавателя комплекса учебных заданий с целью усвоения научно-теоретических основ учебной дисциплины, приобретения умений и навыков, опыта творческой деятельности.

Целью практического занятия для обучающегося является привитие обучающимся умений и навыков практической деятельности по изучаемой дисциплине.

Планируемые результаты при освоении обучающимися практических занятий:

- ~ закрепление, углубление, расширение и детализация знаний при решении конкретных задач;
- ~ развитие познавательных способностей, самостоятельности мышления, творческой активности;
- ~ овладение новыми методами и методиками изучения конкретной учебной дисциплины;
- ~ выработка способности логического осмысления полученных знаний для выполнения заданий;
- ~ обеспечение рационального сочетания коллективной и индивидуальной форм обучения.

Требования к проведению практических занятий

Начинать подготовку к занятию надо с изучения рекомендованной литературы.

Необходимо помнить, что на лекции обычно рассматриваются основные вопросы и наиболее сложные проблемы. Остальная его часть восполняется в процессе самостоятельной работы. В связи с этим работа с рекомендованной литературой обязательна. Особое внимание при этом необходимо обратить на содержание основных положений и выводов, объяснение явлений и фактов, уяснение практического приложения рассматриваемых теоретических вопросов. В процессе этой работы студент должен стремиться понять и запомнить основные положения, поясняющие примеры, разобраться в иллюстративном материале.

Заканчивать подготовку следует составлением плана (конспекта) по изучаемому материалу (вопросу). Это позволяет составить концентрированное, сжатое представление по изучаемым вопросам.

В процессе подготовки к занятиям рекомендуется взаимное обсуждение материала, во время которого закрепляются знания, а также приобретает практика в изложении и разъяснении полученных знаний, развивается речь.

При необходимости следует обращаться за консультацией к преподавателю. Идя на консультацию, необходимо хорошо продумать вопросы, которые требуют разъяснения.

В начале занятия студенты под руководством преподавателя более глубоко осмысливают теоретические положения по теме занятия, раскрывают и объясняют основные положения.

11.3. Методические указания для обучающихся по прохождению курсового проектирования/выполнения курсовой работы

Курсовой проект/ работа проводится с целью формирования у обучающихся опыта комплексного решения конкретных задач профессиональной деятельности.

Курсовой проект/ работа позволяет обучающемуся: систематизировать и закрепить полученные теоретические знания и практические умения в соответствии с требованиями к уровню подготовки, установленными программой учебной дисциплины, программой подготовки специалиста соответствующего уровня, квалификации;

применить полученные знания, умения и практический опыт при решении комплексных задач, в соответствии с основными видами профессиональной деятельности по направлению Лингвистика, теоретическая и прикладная лингвистика;

сформировать умения применять теоретические знания при решении нестандартных задач;

приобрести опыт аналитической, работы и сформировать соответствующие умения;

сформировать умения работы со специальной литературой, справочной, нормативной и правовой документацией и иными информационными источниками;

сформировать умения формулировать логически обоснованные выводы, предложения и рекомендации по результатам выполнения работы;

~ развить профессиональную письменную и устную речь обучающегося;

~ развить системное мышление, творческую инициативу, самостоятельность, организованность и ответственность за принимаемые решения;

~ сформировать навыки планомерной регулярной работы над решением поставленных задач.

Цель выполнения курсовой работы - самостоятельное решение студентом профессиональных

Структура пояснительной записки курсового проекта/ работы

Курсовая работа по дисциплине «Машинное обучение» предполагает следующую структуру:

1. Введение

- Определение темы и актуальность выбранной проблемы или задачи.
- Цели и задачи курсовой работы.
- Обзор структуры работы.

2. Обзор литературы

- Обзор основных концепций и методов машинного обучения, связанных с выбранной темой.
- Рассмотрение существующих исследований и публикаций по теме.
- Анализ применения методов машинного обучения в рассматриваемой области.

3. Методология

- Описание используемых методов и моделей машинного обучения.
- Обоснование выбора конкретных методов для решения задачи.
- Описание источников данных и методов их предобработки.
- План оценки и экспериментального исследования.

4. Эксперименты и результаты

- Описание проведенных экспериментов и использованных данных.
- Анализ полученных результатов с использованием метрик оценки качества моделей.
- Выводы по каждому эксперименту и их интерпретация с точки зрения целей работы.

5. Обсуждение

- Критический анализ результатов и методологии работы.
- Сравнение существующих подходов и результатов других исследований.
- Рассмотрение ограничений и возможных направлений для будущих исследований.

6. Заключение

- Подведение итогов и достижение поставленных целей.
- Основные выводы и выводы на основе полученных результатов.
- Рекомендации для практического применения результатов работы.
-

7. Список литературы: перечисление всех использованных источников, упомянутых в тексте курсовой работы.

8. Приложения (по необходимости)

- Дополнительные материалы, которые могут быть полезны для понимания работы (код, таблицы данных, дополнительные графики и т.д.).

Общие рекомендации:

- **Выбор темы:** выберите тему, которая вас интересует и которую вы готовы глубоко изучить.
- **Использование методов машинного обучения:** примените несколько методов, сравните их результаты и обоснуйте выбор.
- **Эксперименты и оценка:** провести несколько экспериментов с различными конфигурациями моделей, используя подходящие метрики для оценки результатов.
- **Критический подход:** работа должна включать критический анализ результатов и методологии, а также обсуждение ограничений.

Требования к оформлению пояснительной записки курсового проекта/ работы

Пояснительная записка должна быть оформлена в соответствии с требованиями СТО ГУАП.СМК 3.170

11.4. Методические указания для обучающихся по прохождению самостоятельной работы

В ходе выполнения самостоятельной работы, обучающийся выполняет работу по заданию и при методическом руководстве преподавателя, но без его непосредственного участия.

Для обучающихся по заочной форме обучения, самостоятельная работа может включать в себя контрольную работу.

В процессе выполнения самостоятельной работы, у обучающегося формируется целесообразное планирование рабочего времени, которое позволяет им развивать умения и навыки в усвоении и систематизации приобретаемых знаний, обеспечивает высокий уровень успеваемости в период обучения, помогает получить навыки повышения профессионального уровня.

Методическими материалами, направляющими самостоятельную работу обучающихся, являются:

- ~ учебно-методический материал по дисциплине;
- ~ методические указания по выполнению контрольных работ (для обучающихся по заочной форме обучения).

Записи имеют первостепенное значение для самостоятельной работы студентов. Они помогают понять построение изучаемого материала, выделить основные положения, проследить их логику.

Ведение записей способствует превращению чтения в активный процесс, мобилизует, наряду со зрительной, и моторную память. Следует помнить: у студента, систематически ведущего записи, создается свой индивидуальный фонд подсобных материалов для быстрого повторения прочитанного, для мобилизации накопленных знаний. Особенно важны и полезны записи тогда, когда в них находят отражение мысли, возникшие при самостоятельной работе.

11.5. Методические указания для обучающихся по прохождению текущего контроля успеваемости.

Текущий контроль успеваемости предусматривает контроль качества знаний обучающихся, осуществляемого в течение семестра с целью оценивания хода освоения дисциплины.

Главное условие успешности в освоении учебной дисциплины - систематические занятия. Для полного понимания изучаемого материала следует задавать вопросы непосредственно на практических занятиях, чтобы не оставлять пробелов в изучении. За дополнительными разъяснениями и рекомендациями студент может обращаться к преподавателю во время консультаций. Систематическая работа и выполнение практических заданий обеспечит высокую оценку при прохождении промежуточной аттестации.

11.6. Методические указания для обучающихся по прохождению промежуточной аттестации.

Промежуточная аттестация обучающихся предусматривает оценивание промежуточных и окончательных результатов обучения по дисциплине. Она включает в себя:

зачет – это форма оценки знаний, полученных обучающимся в ходе изучения учебной дисциплины в целом или промежуточная (по окончании семестра) оценка знаний

обучающимся по отдельным разделам дисциплины с аттестационной оценкой «зачтено» или «не зачтено».

На зачете в процессе подготовки к ответу прежде, чем приступить к подробному изложению ответа на вопрос, следует составить (письменно или устно) план предстоящего ответа.

Промежуточная аттестация проводится не только в традиционном формате «вопрос-ответ», но и в форме дискуссии, в процессе которой определяется умение студента быстро мыслить, формулировать свой ответ при линейном развитии речи, владение устной и письменной версией официально-деловой нормы современного русского языка.

Лист внесения изменений в рабочую программу дисциплины

Дата внесения изменений и дополнений. Подпись внесшего изменения	Содержание изменений и дополнений	Дата и № протокола заседания кафедры	Подпись зав. кафедрой