

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Волков В.В.

Должность: Ректор

Дата подписания: 14.08.2025 17:03:06

Уникальный программный ключ:

ed68fd4b85b778e0f0b1bfea5dbc56cf4148f1225917e739a70e315177f88591

**Автономная некоммерческая образовательная организация высшего образования
«Европейский университет в Санкт-Петербурге»**

Школа вычислительных социальных наук

УТВЕРЖДАЮ:

Ректор

В.В. Волков

« 26 » *августа* 2025 г.

Протокол УС № 2 от 26.02 2025 г.



**Рабочая программа дисциплины
Введение в машинное обучение**

**образовательная программа
направление подготовки
39.04.01 Социология**

**направленность (профиль) программы
«Вычислительная социология»
уровень высшего образования – магистратура**

Программа двух квалификаций:

- «магистр» по направлению подготовки 39.04.01 Социология;
- дополнительная квалификация – «магистр» по направлению подготовки 09.04.03 Прикладная информатика

язык обучения – русский
форма обучения - очная

Санкт-Петербург

Автор:

Котельников Е.В., доктор технических наук, профессор, Школа вычислительных социальных наук, АНООВО «ЕУСПб»

Рецензент:

Соколов М.В., кандидат экономических наук, старший научный сотрудник Школы вычислительных социальных наук АНООВО «ЕУСПб»

Рабочая программа дисциплины **«Введение в машинное обучение»**, входящей в образовательную программу уровня магистратуры «Вычислительная социология», утверждена на заседании Совета Школы вычислительных социальных наук

Протокол заседания № 4 от 25.02.2025 года.

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

Дисциплина **«Введение в машинное обучение»** является дисциплиной базовой части Блока 1 «Дисциплины (модули)» образовательной программы «Вычислительная социология».

Содержание дисциплины **«Введение в машинное обучение»** охватывает круг вопросов связанных с применением современных алгоритмов машинного обучения к различным экономическим наборам данных (в т.ч., «большим данным») для обеспечения поддержки принимаемых решений. В рамках курса рассматриваются техники предобработки данных, а также современные классификационные алгоритмы. Обсуждаются проблемы и особенности, возникающие при работе с данными, а также подходы к их решению. Преподавание дисциплины предусматривает следующие формы организации учебного процесса: лекции, практические занятия, самостоятельная работа магистранта.

Общая трудоемкость освоения дисциплины составляет 3 зачетных единицы, 108 часов.

Содержание

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ	5
2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ	5
3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ.....	6
4. ОБЪЕМ ДИСЦИПЛИНЫ.....	6
5. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ	6
5.1 Содержание дисциплины	7
5.2 Структура дисциплины.....	8
6. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ.....	9
6.1 Общие положения	9
6.2 Рекомендации по распределению учебного времени по видам самостоятельной работы и разделам дисциплины	9
6.3 Перечень основных вопросов по изучаемым темам для самостоятельной работы обучающихся по дисциплине	9
6.4 Перечень литературы для самостоятельной работы обучающегося:.....	11
6.5 Перечень учебно-методического обеспечения для самостоятельной работы.....	11
7. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ.....	11
7.1 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации.....	11
7.2 Контрольные задания для текущей аттестации.....	13
7.3 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации	13
7.4 Типовые задания к промежуточной аттестации.....	16
7.5 Средства оценки индикаторов достижения компетенций.....	21
8. ОСНОВНАЯ И ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА.....	22
9. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, ИСПОЛЬЗУЕМЫЕ ПРИ ОСУЩЕСТВЛЕНИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА.....	22
9.1 Программное обеспечение	22
9.2 Перечень информационно-справочных систем и профессиональных баз данных информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины:	22
9.3 Лицензионные электронные ресурсы библиотеки Университета	23
9.4 Электронная информационно-образовательная среда Университета.....	23
10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА.....	24
ПРИЛОЖЕНИЕ 1	25

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель дисциплины «Введение в машинное обучение» состоит в ознакомлении студентов с современными методами машинного обучения; развитии навыков работы с данными и извлечения из них новой информации; рассмотрении различных экономических и управленческих ситуаций, в которых машинное обучение может быть применено для обеспечения поддержки принятия решений; знакомстве студентов с современным языком статистического программирования.

Задачи:

1. Сформировать представление о современных алгоритмах машинного обучения.
2. Обсудить особенности использования подходов машинного обучения в различных экономических и управленческих ситуациях для обеспечения поддержки принятия решения.
3. Продемонстрировать применение изучаемого материала на практических примерах с использованием современного языка статистического программирования.

2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ

В результате изучения учебной дисциплины обучающийся должен овладеть следующими компетенциями: общепрофессиональными (ОПК). Планируемые результаты формирования компетенций и индикаторы их достижения в результате освоения дисциплины представлены в Таблице 1.

Таблица 1

Планируемые результаты освоения дисциплины, соотнесенные с индикаторами достижения компетенций обучающихся

Код и наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (знать, уметь, владеть)
ОПК-3 (ПИ) Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров обоснованными выводами и рекомендациями	ИД.ОПК-3.1. Анализирует и структурирует профессиональные данные с использованием современных методов прикладного анализа данных	Знать: принципы, методы и средства анализа и структурирования профессиональной информации З (ОПК-3)
	ИД.ОПК-3.2. Содержательно интерпретирует данные и формулирует выводы и теоретические подходы для решения профессиональных задач	Уметь: интерпретировать данные и формулировать выводы и теоретические подходы для решения профессиональных задач, представляя результаты работы в виде аналитических обзоров У (ОПК-3)
	ИД.ОПК-3.3. Выявляет значимые проблемы и разрабатывает рекомендации по их решению ИД.ОПК-3.4. Оформляет и представляет результаты анализа в виде аналитических обзоров	Владеть: навыками разработки рекомендаций по результатам анализа профессиональной информации В (ОПК-3)
ОПК-4 (ПИ) Способен применять на практике новые научные принципы и методы исследований	ИД.ОПК-4.1. На основе современных теорий и концепций обосновывает актуальность постановки целей и задач научных исследований в профессиональной области знаний	Знать: актуальные направления применения новых научных принципов и методов исследований в профессиональной деятельности З (ОПК-4)
	ИД.ОПК-4.2. Анализирует новые научные принципы и методы исследований в профессиональной области знаний	Уметь: самостоятельно формировать планы и программы научных исследований с применением новых принципов и методов, характерных для выбранной отрасли науки У (ОПК-4)
	ИД.ОПК-4.3. Применяет новые научные принципы и методы исследований в профессиональной области знаний	Владеть: навыками системного использования различных новых научных принципов и

Код и наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (знать, уметь, владеть)
	ИД.ОПК-4.4. Разрабатывает предложения и рекомендации по использованию новых научных принципов и методов исследований в профессиональной области знаний	методов исследований для различных направлений науки В (ОПК-4)

В результате освоения дисциплины магистрант должен:

- **знать:** встроенные типы данных; основные элементы структурного программирования; основы объектно-ориентированного и функционального программирования; стандартные и внешние библиотеки для работы с данными;
- **уметь:** писать несложные вычислительные программы, реализовывать алгоритмы на базовых типах данных; писать процессы, взаимодействующие с внешними API; использовать стандартные и внешние библиотеки;
- **владеть:** навыками работы со стандартной библиотекой языка и внешними модулями, формирования своего инструментария для задач автоматизации; навыками практического применения основных пакетов обработки данных для реализации собственных исследовательских задач.

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина «Введение в машинное обучение» является дисциплиной обязательной части Блока 1 «Дисциплины (модули)» образовательной программы «Вычислительная социология». Курс читается в третьем модуле, форма промежуточной аттестации – зачет с оценкой.

Знания, умения и навыки, полученные при освоении данной дисциплины, применяются магистрантами в процессе прохождения учебной и производственной практики, выполнения выпускной квалификационной работы.

4. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость освоения дисциплины составляет 3 зачетных единицы, 108 часов.

Таблица 2

Объем дисциплины												
Типы учебных занятий и самостоятельная работа		Объем дисциплины										
		Всего	Модуль									
			1	2	3	4	5	6	7	8	9	10
Контактная работа обучающихся с преподавателем в соответствии с УП:		28	-	-	28	-	-	-	-	-	-	-
Лекции (Л)		14	-	-	14	-	-	-	-	-	-	-
Практические занятия (ПЗ)		14	-	-	14	-	-	-	-	-	-	-
Самостоятельная работа (СР)		80	-	-	80	-	-	-	-	-	-	-
Промежуточная аттестация	форма	Зачет с оценкой	-	-	Зачет с оценкой	-	-	-	-	-	-	-
	час.	-	-	-	-	-	-	-	-	-	-	-
Общая трудоемкость дисциплины (час./з.е.)		108/3	-	-	108/3	-	-	-	-	-	-	-

5. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Содержание дисциплины соотносится с планируемыми результатами обучения по дисциплине: через задачи, формируемые компетенции и их компоненты (знания, умения, навыки – далее ЗУВ) по средствам индикаторов достижения компетенций в соответствии с Таблицей 3.

5.1 Содержание дисциплины

Таблица 3

Содержание дисциплины

№ п/п	Наименование тем (разделов)	Содержание тем (разделов)	Коды компетенций	Индикаторы компетенций (в соот.с Таблицей 1)	Коды ЗУВ (в соот.с Таблицей 1)
1	Метод главных компонент	Основные идеи снижения размерности Принципы работы МГК Критерии отбора главных компонент	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)
2	Иерархический кластерный анализ (ИКА)	Виды кластерного анализа Принципы построения кластеров по ИКА. Дендрограмма Интерпретация результатов кластеризации	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)
3	Пропуски и атипичные наблюдения в данных	Причины пропусков в данных Причины возникновения выбросов Основные подходы к решению проблем пропущенных данных и атипичных наблюдений	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)
4	Логистическая регрессия и проблема двухклассовой классификации	Двухклассовая классификация. Матрица неточностей. Несбалансированность классов в выборке. Особенности генерации переменных (скошенность, мультиколлинеарность) Метрики качества классификации (Sens, Spec, F1, ROC AUC) Логит-регрессия как классификатор Многоклассовая логит-регрессия: замечание	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)
5	Многоклассовая классификация. Подходы на основе деревьев решений	Метрики многоклассовой классификации Проблема переобучения «Слабые» классификаторы Классические деревья решений. Параметр сложности Ансамблевые методы. Случайный лес, XGBoost	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)

№ п/п	Наименование тем (разделов)	Содержание тем (разделов)	Коды компетенций	Индикаторы компетенций (в соот.с Таблицей 1)	Коды ЗУВ (в соот.с Таблицей 1)
6	Современные алгоритмы нейронных сетей	Принципы и базовые архитектуры нейронных сетей Примеры решения задач нейронными сетями Продвинутое архитектуры нейронных сетей. Сверточные сети Основы «компьютерного зрения»	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)

5.2 Структура дисциплины

Таблица 4

Структура дисциплины

№ п/п	Наименование тем (разделов)	Объем дисциплины, час.				Форма текущего контроля успеваемости *, промежуточной аттестации	
		Всего	Контактная работа обучающихся с преподавателем по типам учебных занятий в соответствии с УП		СР		
			Л	Пр.З			
Очная форма обучения							
Тема 1	Метод главных компонент	17	2	2	13	Т	ДЗ1
Тема 2	Иерархический кластерный анализ (ИКА)	17	2	2	13	Т	
Тема 3	Пропуски и атипичные наблюдения в данных	17	2	2	13	Т	ДЗ2
Тема 4	Логистическая регрессия и проблема двухклассовой классификации	17	2	2	13	Т	
Тема 5	Многоклассовая классификация. Подходы на основе деревьев решений	17	2	2	13	Т	
Тема 6	Современные алгоритмы нейронных сетей	23	4	4	15	Т	
Промежуточная аттестация						Зачет с оценкой	
Всего:		108	14	14	80		

*Примечание: формы текущего контроля успеваемости: домашнее задание (ДЗ), тест (Т)..

6. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

6.1 Общие положения

Знания и навыки, полученные в результате лекций и семинарских занятий, закрепляются и развиваются в результате повторения материала, усвоенного в аудитории, путем чтения текстов и исследовательской литературы (из списков основной и дополнительной литературы) и их анализа.

Самостоятельная работа является важнейшей частью процесса высшего образования. Ее следует осознанно организовать, выделив для этого необходимое время и соответственным образом организовав рабочее пространство. Важнейшим элементом самостоятельной работы является проработка материалов прошедших занятий (анализ конспектов, чтение рекомендованной литературы) и подготовка к следующим лекциям/семинарским занятиям. Литературу, рекомендованную в программе курса, следует, по возможности, читать в течение всего семестра, концентрируясь на обусловленных программой курса темах.

Существенную часть самостоятельной работы магистранта представляет самостоятельное изучение вспомогательных учебно-методических изданий, лекционных конспектов, интернет-ресурсов и пр. Подготовка к семинарским занятиям является важной формой работы магистранта. Самостоятельная работа может вестись как индивидуально, так и при содействии преподавателя.

6.2 Рекомендации по распределению учебного времени по видам самостоятельной работы и разделам дисциплины

Тема 1. Метод главных компонент:

1. Повторение пройденного на лекциях и практических занятиях материала – 1 час.
 2. Самостоятельная работа с рекомендованной литературой, поиск ответов на возникшие в ходе подготовки вопросы, подготовка к тесту – 8 часов.
 3. Выполнение домашнего задания – 4 часа.
- Итого: 13 часов.

Тема 2. Иерархический кластерный анализ (ИКА):

1. Повторение пройденного на лекциях и практических занятиях материала – 1 час.
 2. Самостоятельная работа с рекомендованной литературой, поиск ответов на возникшие в ходе подготовки вопросы, подготовка к тесту – 8 часов.
 3. Выполнение домашнего задания – 4 часа.
- Итого: 13 часов.

Тема 3. Пропуски и атипичные наблюдения в данных:

1. Повторение пройденного на лекциях и практических занятиях материала – 3 часа.
 2. Самостоятельная работа с рекомендованной литературой, поиск ответов на возникшие в ходе подготовки вопросы, подготовка к тесту – 6 часов.
 3. Выполнение домашнего задания – 4 часа.
- Итого: 13 часов.

Тема 4. Логистическая регрессия и проблема двухклассовой классификации:

1. Повторение пройденного на лекциях и практических занятиях материала – 3 часа.

2. Самостоятельная работа с рекомендованной литературой, поиск ответов на возникшие в ходе подготовки вопросы, подготовка к тесту – 6 часов.
 3. Выполнение домашнего задания – 4 часа.
- Итого: 13 часов.

Тема 5. Многоклассовая классификация. Подходы на основе деревьев решений:

1. Повторение пройденного на лекциях и практических занятиях материала – 3 часа.
 2. Самостоятельная работа с рекомендованной литературой, поиск ответов на возникшие в ходе подготовки вопросы, подготовка к тесту – 6 часов.
 3. Выполнение домашнего задания – 4 часа.
- Итого: 13 часов.

Тема 6. Современные алгоритмы нейронных сетей:

1. Повторение пройденного на лекциях и практических занятиях материала – 7 часов.
 2. Самостоятельная работа с рекомендованной литературой, поиск ответов на возникшие в ходе подготовки вопросы, подготовка к тесту – 8 часов.
- Итого: 15 часов.

6.3 Перечень основных вопросов по изучаемым темам для самостоятельной работы обучающихся по дисциплине

Тема 1. Метод главных компонент

1. Что такое «снижение размерности» данных?
2. Приведите примеры методов/способов снижения размерности.
3. Поясните принципы работы метода главных компонент.
4. Что отражает график «каменистая осыпь»?

Тема 2. Иерархический кластерный анализ (ИКА)

1. Каковы основные виды кластерного анализа?
2. Перечислите основные шаги алгоритма ИКА.
3. Укажите критерии формирования кластеров в рамках ИКА.
4. Каковы основные подходы к интерпретации кластеров?

Тема 3. Пропуски и атипичные наблюдения в данных

1. Дайте классификацию основных причин возникновения пропусков в данных.
2. Какие способы заполнения пропусков в данных Вы знаете?
3. Назовите любые два способа выявления атипичных наблюдений в данных.
4. Перечислите подходы к решению проблемы атипичных наблюдений в данных.

Тема 4. Двухклассовая классификация. Логит-регрессия

1. Сформулируйте две роли логит-регрессии.
2. Что означает «несбалансированность классов» в выборке?
3. Какие метрики применяются для оценивания качества классификации?
4. Поясните, как формируется матрица неточностей.

Тема 5. Многоклассовая классификация. Подходы на основе деревьев решений

1. В каких случаях возникает проблема переобучения?
2. В каких случаях может использоваться «слабый» классификатор?
3. Сформулируйте принципы построения и применения дерева решений.
4. В чем сильные и слабые стороны алгоритма «случайного леса»?

Тема 6. Современные алгоритмы нейронных сетей

1. Перечислите основные компоненты нейронной сети.
2. Для каких решения каких задач может применяться нейронная сеть?
3. Приведите несколько примеров продвинутых архитектур нейронных сетей.
4. Поясните, что понимается под термином «компьютерное зрение»?.

6.4 Перечень литературы для самостоятельной работы обучающегося:

1. Богданов, Е. П. Интеллектуальный анализ данных: практикум для магистрантов направления 09.04.03 «Прикладная информатика» профиль подготовки «Информационные системы и технологии корпоративного управления» / Е. П. Богданов. - Волгоград : ФГБОУ ВО Волгоградский ГАУ, 2019. - 112 с. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1087885> .

2. Горелов, В.И. Анализ статистических данных: практикум: / В.И. Горелов, Т.Н. Ледашева; Российская международная академия туризма. – Москва: Университетская книга, 2015. – 120 с.: ил. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book_red&id=574944.

6.5 Перечень учебно-методического обеспечения для самостоятельной работы

Для обеспечения самостоятельной работы магистрантов по дисциплине «Введение в машинное обучение» разработано учебно-методическое обеспечение в составе:

1. Контрольные задания для подготовки к процедурам текущего контроля (п. 7.2 Рабочей программы).
2. Типовые задания для подготовки к промежуточной аттестации (п. 7.4 Рабочей программы).
3. Рекомендуемые основная, дополнительная литература, Интернет-ресурсы и справочные системы (п. 8, 9 Рабочей программы).
4. Рабочая программа дисциплины размещена в электронной информационно-образовательной среде Университета на электронном учебно-методическом ресурсе АНООВО «ЕУСПб» — образовательном портале LMS Sakai — Sakai@EU.

7. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

7.1 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации

Информация о содержании и процедуре текущего контроля успеваемости, методике оценивания знаний, умений и навыков обучающегося в ходе текущего контроля доводятся научно-педагогическими работниками Университета до сведения обучающегося на первом занятии по данной дисциплине.

Текущий контроль предусматривает подготовку магистрантов к каждому практическому занятию, подготовку домашних заданий, прохождение тестов, активное слушание на лекциях. Магистрант должен присутствовать на практических занятиях, отвечать на поставленные вопросы, показывая, что прочитал разбираемую литературу, представлять содержательные реплики по обсуждаемым вопросам.

Текущий контроль проводится в форме оценивания выполненных домашних заданий и тестов, демонстрирующих степень знакомства с дополнительной литературой.

Таблица 5

**Показатели, критерии и оценивание компетенций и индикаторов их
достижения в процессе текущей аттестации**

Наименование тем (разделов)	Коды компетенций	Индикаторы компетенций	Коды ЗУВ (в соот. с Таблицей 1)	Формы текущего контроля успеваемости		Результаты текущего контроля
Метод главных компонент	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Тестирование 1	Домашнее задание 1	зачтено/ не зачтено
Иерархический кластерный анализ (ИКА)	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Тестирование 2		зачтено/ не зачтено
Пропуски и атипичные наблюдения в данных	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Тестирование 3	Домашнее задание 2	зачтено/ не зачтено
Логистическая регрессия и проблема двухклассовой классификации	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Тестирование 4		зачтено/ не зачтено
Многоклассовая классификация. Подходы на основе деревьев решений	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Тестирование 5		зачтено/ не зачтено
Современные алгоритмы нейронных сетей	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Тестирование 6		зачтено/ не зачтено

При освоении дисциплины каждая из форм текущего контроля (тест, домашнее задание) оценивается с использованием 100-балльной шкалы с последующим переводом в бинарную систему для получения результатов текущего контроля, фиксирующих ход образовательного процесса, согласно Таблице 6

Таблица 6

Критерии оценивания

Формы текущего контроля успеваемости	Описание	Показатели	Количество баллов по 100-балльной шкале	Результаты текущего контроля
Тест	Тесты на занятиях состоят из нескольких вопросов открытого типа, задающихся обучающимся с целью выявить степень понимания ими основных понятий дисциплины и материалов занятий.	Обучающийся ответил на все заданные ему вопросы правильно или в целом правильно, но возможно неполно, или с малозначительными погрешностями.	41–100	зачтено
		Обучающийся ответил на заданные вопросы с грубыми ошибками и/или существенными упущениями, либо вообще не смог ответить на поставленные вопросы.	0–40	не зачтено
Домашнее задание	Домашнее задание подразумевает письменное изложение тезисов по предложенным преподавателем темам либо решение выданных преподавателем задач.	Обучающийся уверенно владеет материалом, грамотно выделяет основные факты, увязывает их с изучаемыми теориями	81–100	зачтено
		Обучающийся демонстрирует хорошее владение материалом, в целом умеет выделить основные факты, но не всегда увязывает их с изучаемыми теориями	61–80	
		Обучающийся демонстрирует недостаточное владение материалом, представляет данные неупорядоченно или несвязно, не увязывает представляемый материал с теорией.	41–60	
		Обучающийся не в состоянии связно изложить материал, выделить суть, не понимает, как увязать материал с теорией, либо допускает грубые ошибки в такой увязке.	0–40	не зачтено

7.2 Контрольные задания для текущей аттестации**7.2.1. Примеры тестов**Тема 1. Метод главных компонент

1. Каковы основные задачи снижения размерности?
2. Каков принцип работы метода главных компонент?
3. По какому критерию отбираются главные компоненты для последующего анализа?

Тема 2. Иерархический кластерный анализ (ИКА)

1. Перечислите основные виды кластерного анализа.
2. Каков принцип построения дендрограммы в ИКА?
3. По какому критерию формируются кластеры в рамках ИКА?

Тема 3. Пропуски и атипичные наблюдения в данных

1. Сформулируйте принципиальные способы заполнения пропусков в данных.
2. Назовите любые два способа выявления атипичных наблюдений в данных.
3. Поясните особенности применения этих способов.

Тема 4. Двухклассовая классификация. Логит-регрессия

1. Поясните, в чем особенность роли логит-регрессии, как классификатора.
2. Каковы последствия несбалансированности классов в выборке?
3. Назовите базовые метрики качества классификации.

Тема 5. Многоклассовая классификация. Подходы на основе деревьев решений

1. Раскройте суть проблемы переобучения.
2. Что такое «слабый» классификатор?
3. Каковы принципы построения и применения ансамблевых методов? Приведите пример.

Тема 6. Современные алгоритмы нейронных сетей

1. Каковы основные элементы нейронной сети?
2. Какие принципы работы реализованы в свёрточных нейронных сетях?
3. Что понимается под термином «компьютерное зрение»?

7.2.2. Примеры домашних заданий

ДЗ №1 по Темам 1 и 2

Вам дан некоторый набор реальных данных, содержащий результаты опроса населения. На этом наборе данных проведите анализ по методу главных компонент. Проинтерпретируйте полученные результаты с помощью biplot. Сколько главных компонент Вы бы отобрали для последующего анализа? Какие названия Вы бы дали отобраным компонентам?

На этом же наборе данных проведите иерархический кластерный анализ. Какое количество кластеров Вы бы сформировали? Почему? Кроме этого, проинтерпретируйте полученные кластеры.

ДЗ №2 по темам 3, 4, 5

Вам дан некоторый набор данных, содержащий (не сбалансированную по классам) выборку наблюдений по клиентам некоторой компании. Для этого набора данных сделайте следующее.

1. Несколькими известными Вам методами идентифицируйте все множественные выбросы в данных. Что это за наблюдения? Проанализируйте их. Также, примите решение о необходимости их исключения из набора данных.

2. На (модифицированном) наборе данных постройте дерево решений с учетом несбалансированности классов. Контролируйте размер дерева с помощью параметра сложности. Какое значение параметра сложности Вы выбрали? Какие факторы и в каком

порядке выстроило полученное дерево решений? Проинтерпретируйте полученный результат с точки зрения экономического смысла.

3. На (модифицированном) наборе данных постройте логит-регрессию с учетом несбалансированности классов. Какие факторы в данной регрессии оказались статистически значимыми? Проинтерпретируйте полученный результат с точки зрения экономического смысла. Какова предсказательная сила построенной модели? Как можно улучшить полученный результат?

7.3 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

Форма промежуточной аттестации – зачет с оценкой в форме тестирования.

Перед зачетом с оценкой проводится консультация, на которой преподаватель отвечает на вопросы магистрантов.

В результате промежуточного контроля знаний студенты получают оценку по дисциплине.

Таблица 7

Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

Вид промежуточной аттестации	Показатели	Количество баллов
Тест	100-81% правильных ответов	81–100
	80-61% правильных ответов	61–80
	60-41% правильных ответов	41–60
	40-0% правильных ответов	0–40

Максимальная оценка **по итогам освоения дисциплины** составляет 100 баллов (30% оценки составляют результаты тестов на занятиях, 20% — результаты домашних заданий, 50% – результаты итогового тестирования). При вычислении оценки обучающегося по итогам освоения дисциплины используется процедура округления до целого.

В результате промежуточного контроля знаний студенты получают аттестацию по дисциплине. На основании оценки обучающегося по итогам освоения дисциплины, выраженной в 100-балльной шкале, выставляется **зачет с оценкой** в соответствии с Таблицей 8.

Таблица 8

Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

Форма промежуточной аттестации/вид промежуточной аттестации	Коды компетенций	Индикаторы компетенций (в соот. с Таблицей 1)	Коды ЗУВ (в соот. с Таблицей 1)	Оценка по итогам освоения дисциплины (в 100-балльной шкале)	Результаты текущего контроля
зачет с оценкой / тест	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1.	З (ОПК-3)	81–100	Зачтено, отлично
		ИД.ОПК-3.2.	У (ОПК-3)	61–80	Зачтено, хорошо
		ИД.ОПК-3.3.	В (ОПК-3)	41–60	Зачтено, удовлетворительно
		ИД.ОПК-3.4.	З (ОПК-4)	0–40	Не зачтено, неудовлетворительно
		ИД.ОПК-4.1.	У (ОПК-4)		
		ИД.ОПК-4.2.	В (ОПК-4)		
		ИД.ОПК-4.3.			
		ИД.ОПК-4.4.			

Результаты промежуточного контроля по дисциплине, выраженные в оценках «зачтено, удовлетворительно», «зачтено, хорошо», «зачтено, отлично» показывают уровень сформированности у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Вычислительная социология» по направлению подготовки 39.04.01 Социология (уровень магистратуры).

Результаты промежуточного контроля по дисциплине, выраженные в оценке «не зачтено, неудовлетворительно», показывают не сформированность у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Вычислительная социология» по направлению подготовки 39.04.01 Социология (уровень магистратуры).

7.4 Типовые задания к промежуточной аттестации

Требования к тестам

Тест включает 25 вопросов по всем компетенциям дисциплины, 10 из них вопросы закрытого типа, 5 – комбинированного типа, 10 – открытого типа, все вопросы разного уровня сложности.

Тест оценивается в баллах в соответствии со следующими критериями:

Задания закрытого типа

Базовый уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте - 1 балл; ответ отличен от эталонного - 0 баллов.

Повышенный уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, лишние символы в ответе отсутствуют - 2 балла; если на любой одной позиции ответа записан не тот символ, который представлен в эталоне ответа - 1 балл; во всех других случаях выставляется 0 баллов

Комбинированные задания

Базовый уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, обоснование по смыслу соответствует эталонному (допускаются различные формулировки ответа, не искажающие его смысла) - 1 балл; ответ отличен от эталонного - 0 баллов.

Повышенный уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, обоснование по смыслу соответствует эталонному (допускаются различные формулировки ответа, не искажающие его смысла) - 2 балла; дан верный ответ, обоснование отсутствует или приведено неверно – 1 балл; во всех остальных случаях - 0 баллов.

Задания открытого типа

Повышенный уровень сложности: ответ соответствует эталонному (допускаются различные формулировки ответа, не искажающие его смысла); правильно названы все запрашиваемые составляющие вопроса, даны верные обоснования - 2 балла; ответ имеет незначительные отклонения от эталонного, правильно названы на все запрашиваемые составляющие вопроса, но для названных даны верные обоснования - 1 балл; ответ значительно отличается от эталонного, имеются фактические ошибки, искажающие его смысл или ответ сформулирован неверно или не сформулирован - 0 баллов.

Высокий уровень сложности: магистрант демонстрирует умение применять знания в нестандартной ситуации, решать нетиповые задачи, приводит корректные обоснования и доказательства, ответ полный, в ответе отсутствуют фактические ошибки, изложение связное, структура прозрачная, логика изложения прослеживается - 3 балла; ответ значительно отличается от эталонного, имеются фактические ошибки, искажающие его смысл или ответ сформулирован неверно или не сформулирован - 0 баллов.

Итоговый балл за тест рассчитывается по формуле:

$$F = \frac{100}{K} * \left(\frac{x_1}{k_1} + \frac{x_2}{k_2} + \dots + \frac{x_n}{k_n} \right),$$

где F – итоговое количество баллов за тест,
 K – количество осваиваемых в рамках дисциплины компетенций,
 k_n – максимально возможное количество баллов за вопросы по компетенции,
 x_n – количество баллов, набранное магистрантом, за правильные ответы на вопросы по соответствующей компетенции.

Примеры тестовых заданий для промежуточной аттестации

ОПК-3 (ПИ) Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями

Задание 1

Уровень сложности: базовый

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Что такое атипичное наблюдение (выброс)?

- А) Наблюдение с пропущенными данными
- В) Наблюдение, сильно отличающееся от остальных
- С) Наблюдение с высокой корреляцией
- D) Наблюдение с нулевой дисперсией

Поле для ответа:

--	--	--

Задание 2

Уровень сложности: базовый

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: В задаче классификации модель дала Accuracy=0.85 для датасета из 100 наблюдений. Сколько наблюдений было предсказано неверно?

- А) 10
- В) 15
- С) 20
- D) 25

Поле для ответа:

--	--	--

Задание 3

Уровень сложности: базовый

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Что такое conjoint-анализ и для чего он нужен?

Поле для ответа _____.

Задание 4

Уровень сложности: базовый

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как метод главных компонент (РСА) помогает сократить размерность данных?

Поле для ответа _____.

Задание 5

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Какой метод использует ансамбли классификаторов?

- А) Линейная регрессия
- В) Бустинг
- С) Сингулярное разложение
- D) Метод k-средних

Поле для ответа:

--	--	--

Задание 6

Инструкция: прочитайте задание и выберите правильный вариант ответа

Уровень сложности: высокий

Вопрос: В случайном лесе с $p=16$, число признаков для разбиения:

- А) 2
- В) 4
- С) 8
- D) 16

Поле для ответа:

--	--	--

Задание 7

Уровень сложности: базовый

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как линейная регрессия может быть использована для импутации пропусков?

Поле для ответа _____.

Задание 8

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Что характеризует главные компоненты?

- А) Высокая корреляция между собой
- В) Ортогональность и убывание дисперсии
- С) Максимальная размерность данных
- D) Случайный порядок дисперсий

Поле для ответа:

--	--	--

Задание 9

Уровень сложности: базовый

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Что такое линейный дискриминантный анализ (LDA)?

- А) Метод снижения размерности

- В) Метод классификации
- С) Использует информацию о классах для проекции данных
- D) Работает только с двумя классами

Поле для ответа:

--	--	--

Задание 10

Уровень сложности: повышенный

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Что такое ансамблевые методы в машинном обучении?

Поле для ответа _____.

ОПК-4 (ПИ) Способен применять на практике новые научные принципы и методы исследований

Задание 1

Уровень сложности: повышенный

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как использовать инструментальные переменные в регрессии?

Поле для ответа _____.

Задание 2

Уровень сложности: повышенный

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как работает алгоритм Random Forest?

Поле для ответа _____.

Задание 3

Уровень сложности: высокий

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Что такое иерархическая кластеризация?

Поле для ответа _____.

Задание 4

Уровень сложности: повышенный

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как работает алгоритм k-средних в кластеризации?

Поле для ответа _____.

Задание 5

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: Какой метрикой измеряется расстояние между объектами в ИКА?

- А) Евклидово расстояние
- В) F1-мера
- С) Энтропия
- D) Коэффициент корреляции

Поле для ответа:

--	--	--

Задание 6

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: Диапазон коэффициента силуэта:

- A) $(-\infty, \infty)$
- B) $[0,1]$
- C) $[-1,1]$
- D) $[-\infty,1]$

Поле для ответа:

--	--	--

Задание 7

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: В бэггинге с 10 моделями, сколько участвует в предсказании?

- A) 1
- B) 5
- C) 8
- D) Все

Поле для ответа:

--	--	--

Задание 8

Уровень сложности: базовый

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: Что такое переобучение в контексте деревьев решений?

- A) Недостаточная глубина дерева
- B) Чрезмерная подгонка под обучающие данные
- C) Игнорирование пропусков в данных
- D) Использование слабых классификаторов

Поле для ответа:

--	--	--

Задание 9

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: В задаче кластеризации вы применили метод агломеративной кластеризации к 5 точкам и остановились на 2 кластерах. Сколько слияний произошло?

- A) 2
- B) 3
- C) 4
- D) 5

Поле для ответа:

--	--	--

Задание 10

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильные ответы

Вопрос: Какие типы ансамблевых методов существуют?

- А) Бэггинг
- В) Бустинг
- С) Стекинг
- D) kNN

Поле для ответа:

--	--	--

7.5 Средства оценки индикаторов достижения компетенций

Таблица 8

Средства оценки индикаторов достижения компетенций

Коды компетенций	Индикаторы компетенций (в соот.с Таблицей 1)	Средства оценки (в соот. с Таблицами 5, 7)
ОПК-3 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4.	тест (Т), домашнее задание (ДЗ), письменная зачетная работа (ПЗР)
ОПК-4 (ПИ)	ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	тест (Т), домашнее задание (ДЗ), письменная зачетная работа (ПЗР)

Таблица 9

Описание средств оценки индикаторов достижения компетенций

Средства оценки (в соот. с Таблицами 5, 7)	Рекомендованный план выполнения работы
Тест	Магистрант в ходе подготовки и выполнения теста по темам дисциплины, показывает способность совершать следующий набор профессиональных действий, получивший развитие в рамках данной дисциплины: 1. Анализировать, структурировать, интерпретировать профессиональные данные с использованием современных методов прикладного анализа данных, формулировать выводы и теоретические подходы для решения профессиональных задач, разрабатывать рекомендации по решению выявленных значимых проблем, представляя результаты анализа в виде аналитических обзоров 2. На основе современных теорий и концепций обосновывать актуальность, анализировать, применять и разрабатывать предложения и рекомендации по использованию новых научных принципов и методов исследований в профессиональной области знаний
Домашнее задание	Магистрант в ходе выполнения домашнего задания по темам дисциплины, показывает способность совершать следующий набор профессиональных действий, получивший развитие в рамках данной дисциплины: 1. Анализировать, структурировать, интерпретировать профессиональные данные с использованием современных методов прикладного анализа данных, формулировать выводы и теоретические подходы для решения профессиональных задач,

Средства оценки (в соот. с Таблицами 5, 7)	Рекомендованный план выполнения работы
	разрабатывать рекомендации по решению выявленных значимых проблем, представляя результаты анализа в виде аналитических обзоров 2. На основе современных теорий и концепций обосновывать актуальность, анализировать, применять и разрабатывать предложения и рекомендации по использованию новых научных принципов и методов исследований в профессиональной области знаний
Письменная зачетная работа	Магистрант в ходе подготовки и выполнения письменной зачетной работы, показывает способность совершать следующий набор профессиональных действий, получивший развитие в рамках данной дисциплины: 1. Анализировать, структурировать, интерпретировать профессиональные данные с использованием современных методов прикладного анализа данных, формулировать выводы и теоретические подходы для решения профессиональных задач, разрабатывать рекомендации по решению выявленных значимых проблем, представляя результаты анализа в виде аналитических обзоров 2. На основе современных теорий и концепций обосновывать актуальность, анализировать, применять и разрабатывать предложения и рекомендации по использованию новых научных принципов и методов исследований в профессиональной области знаний

8. ОСНОВНАЯ И ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

8.1. Основная литература

1. Андрейчиков А. В. Интеллектуальные информационные системы и методы искусственного интеллекта: учебник / А.В. Андрейчиков, О.Н. Андрейчикова. Москва: ИНФРА-М, 2021. 530 с. + Доп. материалы [Электронный ресурс]. (Высшее образование: Магистратура). DOI 10.12737/1009595. ISBN 978-5-16-014883-0. Текст: электронный. URL: <https://znanium.com/catalog/product/1009595> . Режим доступа: по подписке.

8.2 Дополнительная литература

1. Болотова Ю. А. Методы и алгоритмы интеллектуальной обработки цифровых изображений: учеб. пособие / Ю.А. Болотова, А.А. Друки, В.Г. Спицын ; Томский политехнический университет. Томск: Изд-во Томского политехнического университета, 2016. 208 с. ISBN 978-5-4387-0710-3. Текст: электронный. URL: <https://znanium.com/catalog/product/1043928> . Режим доступа: по подписке.

2. Трегуб И. В. Имитационные модели принятия решений: учебное пособие / И. В. Трегуб, Т. А. Горошникова. Москва: ИНФРА-М, 2022. 193 с. + Доп. материалы [Электронный ресурс]. (Высшее образование: Магистратура). ISBN 978-5-16-015393-3. Текст: электронный. URL: <https://znanium.com/catalog/product/1864087> . Режим доступа: по подписке.

9. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, ИСПОЛЬЗУЕМЫЕ ПРИ ОСУЩЕСТВЛЕНИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

9.1 Программное обеспечение

1. ABBYY FineReader 11 Corporate Edition
2. ABBYY Lingvo x5
3. Adobe Acrobat Professional 11.0 MLP AOO License RU
4. Adobe CS5.5 Design Standart Win IE EDU CLP
5. Adobe Acrobat Reader – бесплатно
6. Git (версия 2.40 и выше)
7. Google Chrome
8. Mozilla – бесплатно
9. MS Office (OVS Office Platform)
10. Opera – бесплатно
11. OS Microsoft Windows (OVS OS Platform)

12. VLC – бесплатно
13. Яндекс.Браузер (Yandex Browser) – бесплатно
14. Python – бесплатно

9.2 Перечень информационно-справочных систем и профессиональных баз данных информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины:

Информационно-справочные системы

1. Гарант.Ру. Информационно-правовой портал: <http://www.garant.ru>
2. Информационная система «Единое окно доступа к образовательным ресурсам»: <http://window.edu.ru/>
3. Открытое образование. Ассоциация «Национальная платформа открытого образования»: <http://npoed.ru>
4. Официальная Россия. Сервер органов государственной власти Российской Федерации: <http://www.gov.ru>
5. Официальный интернет-портал правовой информации. Государственная система правовой информации: <http://pravo.gov.ru>
6. Правовой сайт КонсультантПлюс: <http://www.consultant.ru/sys>
7. Российское образование. Федеральный портал: <http://www.edu.ru>

Профессиональные базы данных информационно-телекоммуникационной сети «Интернет»:

1. Национальная электронная библиотека НЭБ: <http://www.rusneb.ru>
2. Президентская библиотека: <http://www.prlib.ru>
3. Российская государственная библиотека: <http://www.rsl.ru/>
4. Российская национальная библиотека: <http://www.nlr.ru/poisk/>

9.3 Лицензионные электронные ресурсы библиотеки Университета

Профессиональные базы данных:

Полный перечень доступных обучающимся профессиональных баз данных представлен на официальном сайте Университета <https://eusp.org/library/electronic-resources>, включая следующие базы данных:

1. **eLIBRARY.RU** — Российский информационно-аналитический портал в области науки, технологии, медицины и образования, содержащий рефераты и полные тексты научных статей и публикаций, наукометрическая база данных: <http://elibrary.ru>;
2. **Университетская информационная система РОССИЯ** — база электронных ресурсов для учебных программ и исследовательских проектов в области социально-гуманитарных наук: <http://www.uisrussia.msu.ru/>;
3. Электронные журналы по подписке (текущие номера научных зарубежных журналов).

Электронные библиотечные системы:

1. **Znaniyum.com** — Электронная библиотечная система (ЭБС) — <http://znaniyum.com/>;
2. Университетская библиотека онлайн — Электронная библиотечная система (ЭБС) — <http://biblioclub.ru/>

9.4 Электронная информационно-образовательная среда Университета

Образовательный процесс по дисциплине поддерживается средствами электронной информационно-образовательной среды Университета, которая включает в себя электронный учебно-методический ресурс АНООВО «ЕУСПб» — образовательный портал LMS Sakai — Sakai@EU, лицензионные электронные ресурсы библиотеки Университета, официальный сайт Университета (Европейский университет в Санкт-Петербурге

[<https://euspr.org/>]), локальную сеть и корпоративную электронную почту Университета, и обеспечивает:

- доступ к учебным планам, рабочим программам дисциплин (модулей), практик и к изданиям электронных библиотечных систем и электронным образовательным ресурсам, указанным в рабочих программах;

- фиксацию хода образовательного процесса, результатов промежуточной аттестации и результатов освоения основной образовательной программы;

- формирование электронного портфолио обучающегося, в том числе сохранение работ обучающегося, рецензий и оценок за эти работы со стороны любых участников образовательного процесса;

- взаимодействие между участниками образовательного процесса, в том числе синхронное и (или) асинхронное взаимодействие посредством сети «Интернет» (электронной почты и т.д.).

Каждый обучающийся в течение всего периода обучения обеспечен индивидуальным неограниченным доступом к электронным ресурсам библиотеки Университета, содержащей издания учебной, учебно-методической и иной литературы по изучаемой дисциплине.

10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

В ходе реализации образовательного процесса используются специализированные многофункциональные аудитории для проведения занятий лекционного типа, занятий семинарского типа (практических занятий, лабораторных работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, укомплектованные специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Проведение занятий лекционного типа обеспечивается демонстрационным оборудованием.

Помещения для самостоятельной работы оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду организации.

Для лиц с ограниченными возможностями здоровья и инвалидов предоставляется возможность присутствия в аудитории вместе с ними ассистента (помощника). Для слабовидящих предоставляется возможность увеличения текста на экране ПК. Для самостоятельной работы лиц с ограниченными возможностями здоровья в помещении для самостоятельной работы организовано одно место (ПК) с возможностями бесконтактного ввода информации и управления компьютером (специализированное лицензионное программное обеспечение – Camera Mouse, веб камера). Библиотека университета предоставляет удаленный доступ к электронным ресурсам библиотеки Университета с возможностями для слабовидящих увеличения текста на экране ПК. Лица с ограниченными возможностями здоровья могут при необходимости воспользоваться имеющимся в университете креслом-коляской. В учебном корпусе имеется адаптированный лифт. На первом этаже оборудован специализированный туалет. У входа в здание университета для инвалидов оборудована специальная кнопка, входная среда обеспечена информационной доской о режиме работы университета, выполненной рельефно-точечным тактильным шрифтом (азбука Брайля).

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ПО ДИСЦИПЛИНЕ
«Введение в машинное обучение»

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

1 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации

Информация о содержании и процедуре текущего контроля успеваемости, методике оценивания знаний, умений и навыков обучающегося в ходе текущего контроля доводятся научно-педагогическими работниками Университета до сведения обучающегося на первом занятии по данной дисциплине.

Текущий контроль предусматривает подготовку магистрантов к каждому практическому занятию, подготовку домашних заданий, прохождение тестов, активное слушание на лекциях. Магистрант должен присутствовать на практических занятиях, отвечать на поставленные вопросы, показывая, что прочитал разбираемую литературу, представлять содержательные реплики по обсуждаемым вопросам.

Текущий контроль проводится в форме оценивания выполненных домашних заданий и тестов, демонстрирующих степень знакомства с дополнительной литературой.

Таблица 1

Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации

Наименование тем (разделов)	Коды компетенций	Индикаторы компетенций	Коды ЗУВ (в соот. с Таблицей 1)	Формы текущего контроля успеваемости		Результаты текущего контроля
Метод главных компонент	ОПК-3-П ОПК-4-П	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Тестирование 1	Домашнее задание 1	зачтено/ не зачтено
Иерархический кластерный анализ (ИКА)	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Тестирование 2		зачтено/ не зачтено
Пропуски и атипичные наблюдения в данных	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Тестирование 3	Домашнее задание 2	зачтено/ не зачтено
Логистическая регрессия и проблема двухклассовой классификации	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Тестирование 4		зачтено/ не зачтено
Многоклассовая классификация. Подходы на основе деревьев решений	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4)	Тестирование 5		зачтено/ не зачтено

Наименование тем (разделов)	Коды компетенций	Индикаторы компетенций	Коды ЗУВ (в соот. с Таблицей 1)	Формы текущего контроля успеваемости	Результаты текущего контроля
		ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	В (ОПК-4)		
Современные алгоритмы нейронных сетей	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	Тестирование 6	зачтено/ не зачтено

При освоении дисциплины каждая из форм текущего контроля (тест, домашнее задание) оценивается с использованием 100-балльной шкалы с последующим переводом в бинарную систему для получения результатов текущего контроля, фиксирующих ход образовательного процесса, согласно Таблице 2.

Таблица 2

Критерии оценивания

Формы текущего контроля успеваемости	Описание	Показатели	Количество баллов по 100-балльной шкале	Результаты текущего контроля
Тест	Тесты на занятиях состоят из нескольких вопросов открытого типа, задающихся обучающимся с целью выявить степень понимания ими основных понятий дисциплины и материалов занятий.	Обучающийся ответил на все заданные ему вопросы правильно или в целом правильно, но возможно неполно, или с малосущественными погрешностями.	41–100	зачтено
		Обучающийся ответил на заданные вопросы с грубыми ошибками и/или существенными упущениями, либо вообще не смог ответить на поставленные вопросы.	0–40	не зачтено
Домашнее задание	Домашнее задание подразумевает письменное изложение тезисов по предложенным преподавателем темам либо решение выданных преподавателем задач.	Обучающийся уверенно владеет материалом, грамотно выделяет основные факты, увязывает их с изучаемыми теориями	81–100	зачтено
		Обучающийся демонстрирует хорошее владение материалом, в целом умеет выделить основные факты, но не всегда увязывает их с изучаемыми теориями	61–80	
		Обучающийся демонстрирует недостаточное владение материалом, представляет данные неупорядоченно или несвязно, не увязывает	41–60	

Формы текущего контроля успеваемости	Описание	Показатели	Количество баллов по 100-балльной шкале	Результаты текущего контроля
		представляемый материал с теорией.		
		Обучающийся не в состоянии связно изложить материал, выделить суть, не понимает, как увязать материал с теорией, либо допускает грубые ошибки в такой увязке.	0–40	не зачтено

2 Контрольные задания для текущей аттестации

2.1. Перечень тестов

Тема 1. Метод главных компонент

1. Каковы основные задачи снижения размерности?
2. Каков принцип работы метода главных компонент?
3. По какому критерию отбираются главные компоненты для последующего анализа?

Тема 2. Иерархический кластерный анализ (ИКА)

1. Перечислите основные виды кластерного анализа.
2. Каков принцип построения дендрограммы в ИКА?
3. По какому критерию формируются кластеры в рамках ИКА?

Тема 3. Пропуски и атипичные наблюдения в данных

1. Сформулируйте принципиальные способы заполнения пропусков в данных.
2. Назовите любые два способа выявления атипичных наблюдений в данных.
3. Поясните особенности применения этих способов.

Тема 4. Двухклассовая классификация. Логит-регрессия

1. Поясните, в чем особенность роли логит-регрессии, как классификатора.
2. Каковы последствия несбалансированности классов в выборке?
3. Назовите базовые метрики качества классификации.

Тема 5. Многоклассовая классификация. Подходы на основе деревьев решений

1. Раскройте суть проблемы переобучения.
2. Что такое «слабый» классификатор?
3. Каковы принципы построения и применения ансамблевых методов? Приведите пример.

Тема 6. Современные алгоритмы нейронных сетей

1. Каковы основные элементы нейронной сети?
2. Какие принципы работы реализованы в свёрточных нейронных сетях?
3. Что понимается под термином «компьютерное зрение»?

2.2. Перечень домашних заданий

ДЗ №1 по Темам 1 и 2

Вам дан некоторый набор реальных данных, содержащий результаты опроса населения. На этом наборе данных проведите анализ по методу главных компонент. Проинтерпретируйте полученные результаты с помощью biplot. Сколько главных компонент Вы бы отобрали для последующего анализа? Какие названия Вы бы дали отобранному компонентам?

На этом же наборе данных проведите иерархический кластерный анализ. Какое количество кластеров Вы бы сформировали? Почему? Кроме этого, проинтерпретируйте полученные кластеры.

ДЗ №2 по темам 3, 4, 5

Вам дан некоторый набор данных, содержащий (не сбалансированную по классам) выборку наблюдений по клиентам некоторой компании. Для этого набора данных сделайте следующее.

1. Несколькими известными Вам методами идентифицируйте все множественные выбросы в данных. Что это за наблюдения? Проанализируйте их. Также, примите решение о необходимости их исключения из набора данных.

2. На (модифицированном) наборе данных постройте дерево решений с учетом несбалансированности классов. Контролируйте размер дерева с помощью параметра сложности. Какое значение параметра сложности Вы выбрали? Какие факторы и в каком порядке выстроило полученное дерево решений? Проинтерпретируйте полученный результат с точки зрения экономического смысла.

3. На (модифицированном) наборе данных постройте логит-регрессию с учетом несбалансированности классов. Какие факторы в данной регрессии оказались статистически значимыми? Проинтерпретируйте полученный результат с точки зрения экономического смысла. Какова предсказательная сила построенной модели? Как можно улучшить полученный результат?

3 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

Форма промежуточной аттестации – зачет с оценкой в форме тестирования.

Перед зачетом с оценкой проводится консультация, на которой преподаватель отвечает на вопросы магистрантов.

В результате промежуточного контроля знаний студенты получают оценку по дисциплине.

Таблица 3

Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

Вид промежуточной аттестации	Показатели	Количество баллов
Тест	100-81% правильных ответов	81–100
	80-61% правильных ответов	61–80
	60-41% правильных ответов	41–60
	40-0% правильных ответов	0–40

Максимальная оценка **по итогам освоения дисциплины** составляет 100 баллов (30% оценки составляют результаты тестов на занятиях, 20% — результаты домашних заданий, 50% – результаты итогового тестирования). При вычислении оценки обучающегося по итогам освоения дисциплины используется процедура округления до целого.

В результате промежуточного контроля знаний студенты получают аттестацию по дисциплине. На основании оценки обучающегося по итогам освоения дисциплины, выраженной в 100-балльной шкале, выставляется **зачет с оценкой** в соответствии с Таблицей 4.

Таблица 4

Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

Форма промежуточной аттестации/вид промежуточной аттестации	Коды компетенций	Индикаторы компетенций (в соот. с Таблицей 1)	Коды ЗУВ (в соот. с Таблицей 1)	Оценка по итогам освоения дисциплины (в 100-балльной шкале)	Результаты текущего контроля
зачет с оценкой / тест	ОПК-3 (ПИ) ОПК-4 (ПИ)	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4. ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	З (ОПК-3) У (ОПК-3) В (ОПК-3) З (ОПК-4) У (ОПК-4) В (ОПК-4)	81–100	Зачтено, отлично
				61–80	Зачтено, хорошо
				41–60	Зачтено, удовлетворительно
				0–40	Не зачтено, неудовлетворительно

Результаты промежуточного контроля по дисциплине, выраженные в оценках «зачтено, удовлетворительно», «зачтено, хорошо», «зачтено, отлично» показывают уровень сформированности у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Вычислительная социология» по направлению подготовки 39.04.01 Социология (уровень магистратуры).

Результаты промежуточного контроля по дисциплине, выраженные в оценке «не зачтено, неудовлетворительно», показывают не сформированность у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Вычислительная социология» по направлению подготовки 39.04.01 Социология (уровень магистратуры).

4 Задания к промежуточной аттестации

Требования к тестам

Тест включает 25 вопросов по всем компетенциям дисциплины, 10 из них вопросы закрытого типа, 5 – комбинированного типа, 10 – открытого типа, все вопросы разного уровня сложности.

Тест оценивается в баллах в соответствии со следующими критериями:

Задания закрытого типа

Базовый уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте - 1 балл; ответ отличен от эталонного - 0 баллов.

Повышенный уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, лишние символы в ответе отсутствуют - 2 балл; если на любой одной позиции ответа записан не тот символ, который представлен в эталоне ответа - 1 балл; во всех других случаях выставляется 0 баллов

Комбинированные задания

Базовый уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, обоснование по смыслу соответствует эталонному (допускаются различные формулировки ответа, не искажающие его смысла) - 1 балл; ответ отличен от эталонного - 0 баллов.

Повышенный уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, обоснование по смыслу соответствует эталонному (допускаются различные формулировки ответа, не искажающие его смысла) - 2 балла; дан верный ответ, обоснование отсутствует или приведено неверно – 1 балл; во всех остальных случаях - 0 баллов.

Задания открытого типа

Повышенный уровень сложности: ответ соответствует эталонному (допускаются различные формулировки ответа, не искажающие его смысла); правильно названы все запрашиваемые составляющие вопроса, даны верные обоснования - 2 балла; ответ имеет незначительные отклонения от эталонного, правильно названы на все запрашиваемые составляющие вопроса, но для названных даны верные обоснования - 1 балл; ответ значительно отличается от эталонного, имеются фактические ошибки, искажающие его смысл или ответ сформулирован неверно или не сформулирован - 0 баллов.

Высокий уровень сложности: магистрант демонстрирует умение применять знания в нестандартной ситуации, решать нетиповые задачи, приводит корректные обоснования и доказательства, ответ полный, в ответе отсутствуют фактические ошибки, изложение связное, структура прозрачная, логика изложения прослеживается - 3 балла; ответ значительно отличается от эталонного, имеются фактические ошибки, искажающие его смысл или ответ сформулирован неверно или не сформулирован - 0 баллов.

Итоговый балл за тест рассчитывается по формуле:

$$F = \frac{100}{K} * \left(\frac{x_1}{k_1} + \frac{x_2}{k_2} + \dots + \frac{x_n}{k_n} \right),$$

где F – итоговое количество баллов за тест,

K – количество осваиваемых в рамках дисциплины компетенций,

k_n – максимально возможное количество баллов за вопросы по компетенции,

x_n – количество баллов, набранное магистрантом, за правильные ответы на вопросы по соответствующей компетенции.

Перечень тестовых заданий для промежуточной аттестации

ОПК-3 (ПИ) Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями

Задание 1

Уровень сложности: базовый

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Что такое атипичное наблюдение (выброс)?

- А) Наблюдение с пропущенными данными
- В) Наблюдение, сильно отличающееся от остальных
- С) Наблюдение с высокой корреляцией
- D) Наблюдение с нулевой дисперсией

Поле для ответа:

--	--	--

Задание 2

Уровень сложности: базовый

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: В задаче классификации модель дала $\text{Assigasy}=0.85$ для датасета из 100 наблюдений. Сколько наблюдений было предсказано неверно?

- A) 10
- B) 15
- C) 20
- D) 25

Поле для ответа:

--	--	--

Задание 3

Уровень сложности: базовый

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Что такое conjoint -анализ и для чего он нужен?

Поле для ответа _____.

Задание 4

Уровень сложности: базовый

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как метод главных компонент (PCA) помогает сократить размерность данных?

Поле для ответа _____.

Задание 5

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Какой метод использует ансамбли классификаторов?

- A) Линейная регрессия
- B) Бустинг
- C) Сингулярное разложение
- D) Метод k-средних

Поле для ответа _____.

Задание 6

Инструкция: прочитайте задание и выберите правильный вариант ответа

Уровень сложности: высокий

Вопрос: В случайном лесе с $p=16$, число признаков для разбиения:

- A) 2
- B) 4
- C) 8
- D) 16

Поле для ответа:

--	--	--

Задание 7

Уровень сложности: базовый

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как линейная регрессия может быть использована для импутации пропусков?

Поле для ответа _____.

Задание 8

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Что характеризует главные компоненты?

- A) Высокая корреляция между собой
- B) Ортогональность и убывание дисперсии
- C) Максимальная размерность данных
- D) Случайный порядок дисперсий

Поле для ответа:

--	--	--

Задание 9

Уровень сложности: базовый

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Что такое линейный дискриминантный анализ (LDA)?

- A) Метод снижения размерности
- B) Метод классификации
- C) Использует информацию о классах для проекции данных
- D) Работает только с двумя классами

Поле для ответа:

--	--	--

Задание 10

Уровень сложности: повышенный

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Что такое ансамблевые методы в машинном обучении?

Поле для ответа _____.

Задание 11

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Какое минимальное количество итераций требуется алгоритму k-means для сходимости, если центроиды не изменяются после первой итерации?

- A) 1
- B) 2
- C) 3
- D) 4
- E) 5

Поле для ответа:

--	--	--

Задание 12

Уровень сложности: базовый

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Что происходит при несбалансированности классов?

- А) Модель переобучается
- В) Ухудшается качество классификации
- С) Увеличивается дисперсия данных
- D) Уменьшается число признаков

Поле для ответа:

--	--	--

Задание 13

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Вы применили one-hot encoding к категориальной переменной с 4 уникальными значениями. Сколько новых столбцов появилось?

- А) 1
- В) 3
- С) 4
- D) 5

Поле для ответа:

--	--	--

Задание 14

Уровень сложности: базовый

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как работает градиентный спуск?

Поле для ответа _____.

Задание 15

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Что может быть следствием игнорирования выбросов?

- А) Смещение результатов классификации
- В) Увеличение числа кластеров
- С) Снижение размерности
- D) Ускорение алгоритма

Поле для ответа:

--	--	--

Задание 16

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Какой индекс оценивает разделимость кластеров?

- А) Силуэтный коэффициент

- В) F1-мера
- С) Коэффициент корреляции
- D) Среднеквадратичная ошибка

Поле для ответа:

--	--	--

Задание 17

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: В дереве решений вы ограничили максимальную глубину до 3. Сколько максимум листьев может быть?

- A) 4
- B) 8
- C) 12
- D) 16

Поле для ответа:

--	--	--

Задание 18

Уровень сложности: высокий

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Объясните, как работает алгоритм AdaBoost для классификации.

Поле для ответа _____.

Задание 19

Уровень сложности: повышенный

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Объясните, как использовать метод Монте-Карло для оценки неопределенности в предсказаниях регрессионной модели.

Поле для ответа _____.

Задание 20

Уровень сложности: базовый

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как регуляризация помогает в машинном обучении?

Поле для ответа _____.

Задание 21

Уровень сложности: базовый

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Какие типы регуляризации используются в бустинге?

- A) L1 (Lasso)
- B) L2 (Ridge)
- C) Elastic Net
- D) K-means

Поле для ответа:

--	--	--

Задание 22

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Какие методы импутации подходят для временных рядов с пропусками?

- А) Линейная интерполяция
- В) Замена предыдущим значением (forward fill)
- С) Использование скользящего среднего
- D) Замена случайными значениями

Поле для ответа:

--	--	--

Задание 23

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Какой метод используется для извлечения главных компонент?

- А) Линейная регрессия
- В) Сингулярное разложение (SVD)
- С) Кластеризация k-средних
- D) Градиентный спуск

Поле для ответа:

--	--	--

Задание 24

Уровень сложности: повышенный

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как бэггинг улучшает качество классификации по сравнению с одним деревом решений?

Поле для ответа _____.

Задание 25

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Что верно о логистической регрессии?

- А) Используется для бинарной классификации
- В) Выводит вероятность принадлежности к классу
- С) Может быть использована для мультиклассовой классификации
- D) Требуется нормальное распределение признаков

Поле для ответа:

--	--	--

Задание 26

Уровень сложности: базовый

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как реализовать бэггинг для улучшения модели?

Поле для ответа _____.

Задание 27

Уровень сложности: базовый

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Какие преимущества имеет LightGBM?

- А) Быстрая скорость обучения
- В) Низкое использование памяти
- С) Высокая точность
- D) Игнорирование пропусков

Поле для ответа:

--	--	--

Задание 28

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Зачем масштабировать данные перед кластеризацией?

- А) Увеличить число кластеров
- В) Устранить влияние разных масштабов признаков
- С) Ускорить вычисления
- D) Снизить размерность

Поле для ответа:

--	--	--

Задание 29

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Какой тест проверяет пригодность данных для анализа выбросов?

- А) Тест Шапиро-Уилка
- В) Тест Бартлетта
- С) КМО-тест
- D) Z-тест

Поле для ответа:

--	--	--

Задание 30

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Какой метод коррекции выбросов применяется в временных рядах?

- А) Удаление наблюдений
- В) Сглаживание данных
- С) Логарифмирование признаков
- D) Кластеризация

Поле для ответа:

--	--	--

Задание 31

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильные варианты ответа

Вопрос: Какие метрики расстояния используются в кластеризации?

- А) Евклидово
- В) Манхэттенское
- С) Косинусное
- D) Корреляция Пирсона

Поле для ответа:

--	--	--

Задание 32

Уровень сложности: базовый

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Какое значение AUC указывает на идеальную классификацию?

- А) 0.0
- В) 0.25
- С) 0.5
- D) 0.75
- E) 1.0

Поле для ответа:

--	--	--

Задание 33

Уровень сложности: базовый

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Что отображает дендрограмма в ИКА?

- А) Распределение данных
- В) Иерархию объединения кластеров
- С) Линейные зависимости
- D) Дисперсию признаков

Поле для ответа:

--	--	--

Задание 34

Уровень сложности: повышенный

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Что такое переобучение модели и как его предотвратить?

Поле для ответа: _____

Задание 35

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильные варианты ответа

Вопрос: Какие метрики используются в ROC-кривой?

- A) True Positive Rate
- B) False Positive Rate
- C) Precision
- D) Area Under Curve (AUC)

Поле для ответа:

--	--	--

Задание 36

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильные варианты ответа

Вопрос: Что такое бэггинг?

- A) Обучение нескольких моделей на разных подвыборках
- B) Объединение предсказаний голосованием
- C) Последовательное обучение моделей на ошибках предыдущих
- D) Использование разных алгоритмов для предсказания

Поле для ответа:

--	--	--

Задание 37

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: В K-means с $K=3$ на датасете из 90 точек, сколько кластеров будет сформировано?

- A) 1
- B) 2
- C) 3
- D) 90

Поле для ответа:

--	--	--

Задание 38

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Что показывает матрица неточностей?

- A) Дисперсию данных
- B) Распределение ошибок классификации
- C) Корреляцию признаков
- D) Число пропусков

Поле для ответа:

--	--	--

Задание 39

Уровень сложности: базовый

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Если первые две главные компоненты объясняют 60% и 30%, общий процент:

- A) 30%
- B) 60%
- C) 90%
- D) 100%

Поле для ответа:

--	--	--

Задание 40

Уровень сложности: базовый

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Что такое несбалансированность классов и как с ней справиться?

Поле для ответа _____.

Задание 41

Уровень сложности: повышенный

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как параметр `max_depth` влияет на деревья в бустинге?

Поле для ответа _____.

Правильный ответ: `Max_depth` ограничивает глубину деревьев, контролируя их сложность.

Задание 42

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильные варианты ответа

Вопрос: Какие методы связывания используются в иерархической кластеризации?

- A) Одиночное связывание
- B) Полное связывание
- C) Среднее связывание
- D) Метод Уорда

Поле для ответа:

--	--	--

Задание 43

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Сколько классов предполагает бинарная логистическая регрессия?

- A) 1
- B) 2
- C) 3
- D) 4
- E) 5

Поле для ответа:

--	--	--

Задание 44

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильные варианты ответа

Вопрос: Какие метрики оценивают качество классификации?

- A) Precision
- B) Recall
- C) F1-score
- D) Среднеквадратичная ошибка

Поле для ответа:

--	--	--

Задание 45

Уровень сложности: высокий

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как метод главных компонент помогает устранить проблему мультиколлинеарности при анализе данных?

Поле для ответа _____.

Задание 46

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: В нейронной сети с 4 входами и 3 нейронами в скрытом слое, сколько весов (без смещений)?

- A) 4
- B) 7
- C) 12
- D) 15

Поле для ответа:

--	--	--

Задание 47

Компетенция: ОПК-3-П

Уровень сложности: базовый

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как early stopping улучшает обучение бустинговой модели?

Поле для ответа _____.

Задание 48

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Какой метод предотвращает переобучение в нейронных сетях?

- A) Dropout
- B) Увеличение числа слоёв
- C) Уменьшение размера выборки
- D) Игнорирование пропусков

Поле для ответа:

--	--	--

Задание 49

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Если PCA объясняет 85% дисперсии, не объясненная часть:

- A) 5%
- B) 15%
- C) 85%
- D) 100%

Поле для ответа:

--	--	--

Задание 50

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Если TP = 15, FP = 5, TN = 70, FN = 10, точность (ассигасу)?

- A) 0.85
- B) 0.75
- C) 0.65
- D) 0.95

Поле для ответа:

--	--	--

Задание 51

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Вы применили Min-Max нормализацию к признаку с диапазоном [5, 25]. Какое значение примет 15 после нормализации?

- A) 0.2
- B) 0.4
- C) 0.5
- D) 0.6

Поле для ответа:

--	--	--

Задание 52

Уровень сложности: базовый

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Почему XGBoost считается более эффективным, чем классический GBM?

Поле для ответа: _____

Задание 53

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: В регрессии с L2-регуляризацией штрафной параметр $\lambda=0.5$, а сумма квадратов коэффициентов равна 8. Чему равен штраф?

- A) 2
- B) 4
- C) 6
- D) 8

Поле для ответа:

--	--	--

Задание 54

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Что является основным элементом нейронной сети?

- A) Нейрон
- B) Кластер
- C) Регрессор
- D) Компонента

Поле для ответа:

--	--	--

Задание 55

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Какая метрика оценивает качество классификации?

- A) Среднеквадратичная ошибка
- B) F1-мера
- C) Коэффициент корреляции
- D) Энтропия

Поле для ответа:

--	--	--

Задание 56

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильные варианты ответа

Вопрос: Что является целью кластеризации?

- A) Группировка похожих объектов
- B) Прогнозирование значений
- C) Выявление закономерностей
- D) Снижение размерности

Поле для ответа:

--	--	--

Задание 57

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Какое количество случайных эффектов оценивается в модели с 6 объектами?

- A) 1
- B) 2
- C) 3
- D) 4
- E) 6

Поле для ответа:

--	--	--

Задание 58

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Какова роль логит-регрессии в классификации?

- A) Прогнозирование непрерывных величин
- B) Оценка вероятности принадлежности к классу
- C) Снижение размерности данных
- D) Кластеризация данных

Поле для ответа:

--	--	--

Задание 59

Уровень сложности: базовый

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: Что показывает дендрограмма?

- A) Распределение данных
- B) Иерархию объединения кластеров
- C) Дисперсию признаков
- D) Качество классификации

Поле для ответа:

--	--	--

Задание 60

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный вариант ответа

Вопрос: В наборе данных [2, 3, 4, 5, 12], какое значение является выбросом по IQR?

- A) 2
- B) 5
- C) 12
- D) Нет выбросов

Поле для ответа:

--	--	--

ОПК-4 (ПИ) Способен применять на практике новые научные принципы и методы исследований

Задание 1

Уровень сложности: повышенный

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как использовать инструментальные переменные в регрессии?

Поле для ответа _____.

Задание 2

Уровень сложности: повышенный

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как работает алгоритм Random Forest?

Поле для ответа _____.

Задание 3

Уровень сложности: высокий

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Что такое иерархическая кластеризация?

Поле для ответа _____.

Задание 4

Уровень сложности: повышенный

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как работает алгоритм k-средних в кластеризации?

Поле для ответа _____.

Задание 5

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: Какой метрикой измеряется расстояние между объектами в ИКА?

- А) Евклидово расстояние
- В) F1-мера
- С) Энтропия
- D) Коэффициент корреляции

Поле для ответа:

--	--	--

Задание 6

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: Диапазон коэффициента силуэта:

- А) $(-\infty, \infty)$
- В) $[0,1]$
- С) $[-1,1]$
- D) $[-\infty,1]$

Поле для ответа:

--	--	--

Задание 7

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: В бэггинге с 10 моделями, сколько участвует в предсказании?

- A) 1
- B) 5
- C) 8
- D) Все

Поле для ответа:

--	--	--

Задание 8

Уровень сложности: базовый

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: Что такое переобучение в контексте деревьев решений?

- A) Недостаточная глубина дерева
- B) Чрезмерная подгонка под обучающие данные
- C) Игнорирование пропусков в данных
- D) Использование слабых классификаторов

Поле для ответа:

--	--	--

Задание 9

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: В задаче кластеризации вы применили метод агломеративной кластеризации к 5 точкам и остановились на 2 кластерах. Сколько слияний произошло?

- A) 2
- B) 3
- C) 4
- D) 5

Поле для ответа:

--	--	--

Задание 10

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильные ответы

Вопрос: Какие типы ансамблевых методов существуют?

- A) Бэггинг
- B) Бустинг
- C) Стекинг
- D) kNN

Поле для ответа:

--	--	--

Задание 11

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: Какое минимальное количество периодов времени требуется для панельных данных?

- A) 1
- B) 2
- C) 3
- D) 4
- E) 5

Поле для ответа:

--	--	--

Задание 12

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: В задаче с несбалансированными классами (90 объектов класса 0, 10 класса 1) модель предсказала все как класс 0. Чему равна точность (ассигасу)?

- A) 0.1
- B) 0.5
- C) 0.9
- D) 1.0

Поле для ответа:

--	--	--

Задание 13

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильные ответы

Вопрос: Какие методы регуляризации применяются в XGBoost?

- A) L1 (Lasso)
- B) L2 (Ridge)
- C) Elastic Net
- D) Dropout

Поле для ответа:

--	--	--

Задание 14

Уровень сложности: высокий

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как интерпретировать собственные значения в PCA?

Поле для ответа _____.

Задание 15

Уровень сложности: высокий

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как градиентный спуск используется в Gradient Boosting Machine (GBM)?

Поле для ответа _____.

Задание 16

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильные ответы

Вопрос: Какие алгоритмы являются примерами бустинга?

- A) AdaBoost
- B) Gradient Boosting Machine
- C) Random Forest
- D) XGBoost

Поле для ответа:

--	--	--

Задание 17

Уровень сложности: высокий

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как собственные значения связаны с главными компонентами, и как их использовать?

Поле для ответа _____.

Задание 18

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: В K-means с $K=5$, минимальное число точек в кластере?

- A) 0
- B) 1
- C) 5
- D) 10

Поле для ответа:

--	--	--

Задание 19

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильные ответы

Вопрос: Какие методы используются для выявления выбросов?

- A) Boxplot
- B) Z-Score
- C) Изоляционный лес (Isolation Forest)
- D) Линейная регрессия

Поле для ответа:

--	--	--

Задание 20

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: Какой метод помогает выбрать число кластеров?

- А) Критерий Кайзера
- В) Метод локтя
- С) F-тест
- D) Тест Шапиро-Уилка

Поле для ответа:

--	--	--

Задание 21

Уровень сложности: повышенный

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как регуляризация применяется в бустинге для предотвращения переобучения?

Поле для ответа _____.

Задание 22

Уровень сложности: повышенный

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как заполнить пропущенные значения в наборе данных?

Поле для ответа _____.

Задание 23

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: В PCA с собственными значениями 5.0, 2.0, 1.0, 0.5, доля дисперсии первой компоненты?

- А) 0.5
- В) 0.625
- С) 0.833
- D) 1.0

Поле для ответа:

--	--	--

Задание 24

Уровень сложности: базовый

Инструкция: прочитайте задание и выберите правильные ответы

Вопрос: Что такое градиентный бустинг?

- А) Последовательное добавление моделей для коррекции ошибок
- В) Использование градиентного спуска для оптимизации
- С) Обучение на случайных подмножествах признаков
- D) Использование бэггинга для снижения variance

Поле для ответа:

--	--	--

Задание 25

Уровень сложности: повышенный

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Почему необходимо масштабировать данные перед применением РСА, и что может произойти, если этого не сделать?

Поле для ответа _____.

Задание 26

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильные ответы

Вопрос: Какие методы относятся к ансамблевым классификаторам?

- А) Бэггинг
- В) Бустинг
- С) Линейная регрессия
- D) К-средних

Поле для ответа:

--	--	--

Задание 27

Уровень сложности: повышенный

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как РСА можно использовать для визуализации данных с высокой размерностью?

Поле для ответа _____.

Задание 28

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильные ответы

Вопрос: Какие особенности характерны для случайного леса как ансамблевого метода?

- А) Использует случайный выбор признаков для каждого дерева
- В) Может быть применён для регрессии и классификации
- С) Требуется тщательной настройки гиперпараметров
- D) Устойчив к несбалансированным данным

Поле для ответа:

--	--	--

Обоснование _____

Задание 29

Уровень сложности: базовый

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: В задаче классификации с двумя классами (50 объектов каждого) модель дала 40 ТР и 10 FN для класса 1. Чему равна полнота (recall) для класса 1?

- А) 0.6
- В) 0.8
- С) 0.9
- D) 1.0

Поле для ответа:

--	--	--

Обоснование _____

Задание 30

Уровень сложности: высокий

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: В чем основная цель кластерного анализа?

Поле для ответа _____

Правильный ответ: Цель кластерного анализа — разделить объекты на группы (кластеры) так, чтобы объекты внутри одного кластера были похожи друг на друга больше, чем на объекты из других кластеров.

Задание 31

Уровень сложности: базовый

Инструкция: прочитайте задание и выберите правильные ответы

Вопрос: Что такое стекинг?

- А) Объединение предсказаний нескольких моделей
- В) Обучение мета-модели на предсказаниях базовых моделей
- С) Использование бэггинга для снижения variance
- D) Последовательное обучение моделей на ошибках

Поле для ответа:

--	--	--

Задание 32

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: В задаче классификации модель дала precision=0.8 и recall=0.6 для класса 1. Чему равно значение F1-меры?

- А) 0.48
- В) 0.68
- С) 0.8
- D) 1.0

Поле для ответа:

--	--	--

Задание 33

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: Какой тип переменных требует использования манхэттенского расстояния?

- А) Непрерывные
- В) Дискретные
- С) Номинальные
- D) Бинарные

Поле для ответа:

--	--	--

Задание 34

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: В задаче регрессии МАН составило 4 для 20 наблюдений. Чему равно общее абсолютное отклонение?

- A) 20
- B) 40
- C) 80
- D) 100

Поле для ответа:

--	--	--

Задание 35

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: Как выбираются начальные центроиды в k-средних?

- A) Всегда фиксированные точки
- B) Случайный выбор объектов
- C) Минимизация дисперсии
- D) Максимальная корреляция

Поле для ответа:

--	--	--

Задание 36

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: Если $\text{ассигасу} = 0.9$, а данные сбалансированы, доля правильных предсказаний?

- A) 0.1
- B) 0.5
- C) 0.9
- D) 1.0

Поле для ответа:

--	--	--

Задание 37

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: Какой метод выявляет атипичные наблюдения?

- A) Boxplot
- B) Линейная регрессия
- C) Метод k-средних
- D) Тест Шапиро-Уилка

Поле для ответа:

--	--	--

Задание 38

Уровень сложности: базовый

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: В градиентном бустинге с 3 деревьями и скоростью обучения 0.2 каждое дерево предсказывает 0.5. Каково итоговое предсказание?

- A) 0.3
- B) 0.5
- C) 0.7
- D) 1.0

Поле для ответа:

--	--	--

Задание 39

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: В задаче регрессии MSE составило 50 для 100 наблюдений. Чему примерно равно RMSE?

- A) 5
- B) 7
- C) 10
- D) 50

Поле для ответа:

--	--	--

Задание 40

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильные ответы

Вопрос: Какие особенности отличают кластеризацию DBSCAN от k-средних?

- A) Не требует задания числа кластеров
- B) Может выделять выбросы как отдельную категорию
- C) Основан на плотности данных
- D) Всегда даёт сферические кластеры

Поле для ответа:

--	--	--

Задание 41

Уровень сложности: высокий

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как провести анализ главных компонент (PCA) для набора данных?

Поле для ответа: _____

Задание 42

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: Какой критерий используется для объединения кластеров в ИКА?

- A) Минимизация внутрикластерной дисперсии
- B) Максимизация корреляции
- C) Минимизация числа кластеров
- D) Максимизация числа признаков

Поле для ответа:

--	--	--

Задание 43

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: Сколько главных компонент можно извлечь из датасета с 6 признаками?

- A) 1
- B) 3
- C) 6
- D) 9

Поле для ответа:

--	--	--

Задание 44

Уровень сложности: повышенный

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как Random Forest может быть использован для оценки важности признаков?

Поле для ответа: _____.

Задание 45

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: Вы применили L1-регуляризацию к модели с 5 признаками, и 2 коэффициента стали равны 0. Сколько признаков осталось в модели?

- A) 2
- B) 3
- C) 4
- D) 5

Поле для ответа:

--	--	--

Задание 46

Уровень сложности: базовый

Инструкция: прочитайте задание и выберите правильные ответы

Вопрос: Какие методы используются для снижения размерности?

- A) PCA
- B) t-SNE
- C) K-means
- D) LDA

Правильный ответ: A, B, D.

Поле для ответа:

--	--	--

Задание 47

Уровень сложности: повышенный

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Что такое ROC-кривая и как ее интерпретировать?

Поле для ответа _____.

Задание 48

Уровень сложности: повышенный

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как ROC-кривая помогает оценить качество классификации в ансамблевых методах?

Поле для ответа _____.

Задание 49

Уровень сложности: базовый

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Что такое выбросы в данных и как их можно обнаружить?

Поле для ответа _____.

Задание 50

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: Вы применили дропаут с вероятностью 0.3 к слою с 10 нейронами. Сколько нейронов в среднем активно?

- A) 3
- B) 5
- C) 7
- D) 10

Поле для ответа:

--	--	--

Задание 51

Уровень сложности: базовый

Инструкция: прочитайте задание и выберите правильные ответы

Вопрос: Какие особенности алгоритма k-means?

- A) Требуется задания числа кластеров заранее
- B) Чувствителен к начальной инициализации центроидов
- C) Работает только с евклидовым расстоянием
- D) Предполагает сферическую форму кластеров

Поле для ответа:

--	--	--

Задание 52

Уровень сложности: базовый

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: Какое количество узлов в дереве решений с глубиной 2, если каждый узел делится на 2 дочерних?

- A) 3
- B) 4
- C) 5
- D) 6
- E) 7

Поле для ответа:

--	--	--

Задание 53

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильные ответы

Вопрос: Какие ограничения есть у PCA?

- A) Неспособность учитывать нелинейные зависимости
- B) Требование к центрированию данных
- C) Чувствительность к масштабу признаков
- D) Невозможность обработки категориальных данных напрямую

Поле для ответа:

--	--	--

Задание 54

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильные ответы

Вопрос: Какие варианты градиентного спуска существуют?

- A) Batch Gradient Descent
- B) Stochastic Gradient Descent
- C) Mini-batch Gradient Descent
- D) K-means Descent

Поле для ответа:

--	--	--

Задание 55

Уровень сложности: базовый

Инструкция: прочитайте задание и выберите правильные ответы

Вопрос: Какие методы кластеризации существуют?

- A) Иерархическая кластеризация

- B) K-means
- C) DBSCAN
- D) Логистическая регрессия

Поле для ответа:

--	--	--

Задание 56

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: В датасете 50 наблюдений и 3 признака. Вы удалили 1 признак из-за мультиколлинеарности, а затем применили PCA. Сколько главных компонент можно теперь извлечь?

- A) 1
- B) 2
- C) 3
- D) 50

Поле для ответа:

--	--	--

Задание 57

Уровень сложности: высокий

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: Какие виды кластерного анализа существуют?

- A) Иерархический и k-средних
- B) Линейный и нелинейный
- C) Одноклассовый и многоклассовый
- D) Регрессионный и факторный

Поле для ответа:

--	--	--

Задание 58

Уровень сложности: повышенный

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Что такое бинарная логистическая регрессия?

Поле для ответа _____

Задание 59

Уровень сложности: базовый

Инструкция: прочитайте задание и дайте развернутый ответ

Вопрос: Как AdaBoost адаптируется для задач классификации?

Поле для ответа _____

Задание 60

Уровень сложности: повышенный

Инструкция: прочитайте задание и выберите правильный ответ

Вопрос: В дереве решений максимальная глубина задана как 3, а число признаков — 2. Какое максимальное число листьев возможно?

- A) 4
- B) 6
- C) 8
- D) 10

Поле для ответа:

--	--	--

5 Средства оценки индикаторов достижения компетенций

Таблица 5

Средства оценки индикаторов достижения компетенций

Коды компетенций	Индикаторы компетенций (в соот.с Таблицей 1)	Средства оценки (в соот. с Таблицами 5, 7)
ОПК-3-П	ИД.ОПК-3.1. ИД.ОПК-3.2. ИД.ОПК-3.3. ИД.ОПК-3.4.	тест (Т), домашнее задание (ДЗ), письменная зачетная работа (ПЗР)
ОПК-4-П	ИД.ОПК-4.1. ИД.ОПК-4.2. ИД.ОПК-4.3. ИД.ОПК-4.4.	тест (Т), домашнее задание (ДЗ), письменная зачетная работа (ПЗР)

Таблица 6

Описание средств оценки индикаторов достижения компетенций

Средства оценки (в соот. с Таблицами 5, 7)	Рекомендованный план выполнения работы
Тест	Магистрант в ходе подготовки и выполнения теста по темам дисциплины, показывает способность совершать следующий набор профессиональных действий, получивший развитие в рамках данной дисциплины: 1. Анализировать, структурировать, интерпретировать профессиональные данные с использованием современных методов прикладного анализа данных, формулировать выводы и теоретические подходы для решения профессиональных задач, разрабатывать рекомендации по решению выявленных значимых проблем, представляя результаты анализа в виде аналитических обзоров 2. На основе современных теорий и концепций обосновывать актуальность, анализировать, применять и разрабатывать предложения и рекомендации по использованию новых научных принципов и методов исследований в профессиональной области знаний
Домашнее задание	Магистрант в ходе выполнения домашнего задания по темам дисциплины, показывает способность совершать следующий набор профессиональных действий, получивший развитие в рамках данной дисциплины: 1. Анализировать, структурировать, интерпретировать профессиональные данные с использованием современных методов прикладного анализа данных, формулировать выводы и теоретические подходы для решения профессиональных задач, разрабатывать рекомендации по решению выявленных значимых проблем, представляя результаты анализа в виде аналитических обзоров 2. На основе современных теорий и концепций обосновывать актуальность, анализировать, применять и разрабатывать предложения и рекомендации по использованию новых научных принципов и методов исследований в профессиональной области знаний

Средства оценки (в соот. с Таблицами 5, 7)	Рекомендованный план выполнения работы
Письменная зачетная работа	Магистрант в ходе подготовки и выполнения письменной зачетной работы, показывает способность совершать следующий набор профессиональных действий, получивший развитие в рамках данной дисциплины: 1. Анализировать, структурировать, интерпретировать профессиональные данные с использованием современных методов прикладного анализа данных, формулировать выводы и теоретические подходы для решения профессиональных задач, разрабатывать рекомендации по решению выявленных значимых проблем, представляя результаты анализа в виде аналитических обзоров 2. На основе современных теорий и концепций обосновывать актуальность, анализировать, применять и разрабатывать предложения и рекомендации по использованию новых научных принципов и методов исследований в профессиональной области знаний