

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«СЕВЕРО-КАВКАЗСКАЯ ГОСУДАРСТВЕННАЯ АКАДЕМИЯ»

«УТВЕРЖДАЮ»

Проректор по учебной работе

« 30 » 03

2022 г.

Г.Ю. Нагорная



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Анализ данных и машинное обучение

Уровень образовательной программы _____ бакалавриат _____

Направление подготовки _____ 09.03.04 Программная инженерия _____

Направленность (профиль) _____ Программная инженерия _____

Форма обучения _____ очная _____

Срок освоения ОП _____ 4 года _____

Институт _____ Прикладной математики и информационных технологий _____

Кафедра разработчик РПД _____ Прикладная информатика _____

Выпускающая кафедра _____ Прикладная информатика _____

Начальник
учебно-методического управления _____ Семенова Л.У.

Директор института ПМ и ИТ _____ Тебурев Д.Б.

Заведующий выпускающей кафедрой _____ Хапаева Л.Х.

г. Черкесск, 2022 г.

СОДЕРЖАНИЕ

1. Цели освоения дисциплины	
2. Место дисциплины в структуре образовательной программы	
3. Планируемые результаты обучения по дисциплине	
4. Структура и содержание дисциплины	
4.1. Объем дисциплины и виды учебной работы.....	
4.2. Содержание дисциплины	
4.2.1. Разделы (темы) дисциплины, виды учебной деятельности и формы контроля.....	
4.2.2. Лекционный курс	
4.2.3. Лабораторный практикум	
4.2.4. Практические занятия.....	
4.3. Самостоятельная работа обучающегося.....	
5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине	
6. Образовательные технологии	
7. Учебно-методическое и информационное обеспечение дисциплины	
7.1. Перечень основной и дополнительной учебной литературы.....	
7.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет».....	
7.3. Информационные технологии, лицензионное программное обеспечение	
8. Материально-техническое обеспечение дисциплины	
8.1. Требования к аудиториям (помещениям, местам) для проведения занятий	
8.2. Требования к оборудованию рабочих мест преподавателя и обучающихся	
8.3. Требования к специализированному оборудованию.....	
9. Особенности реализации дисциплины для инвалидов и лиц с ограниченными возможностями здоровья	
Приложение 1. Фонд оценочных средств	

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целями освоения дисциплины «Анализ данных и машинное обучение» являются: формирование у обучающихся понимания современных методов и инструментальных средств применения машинного обучения для анализа данных.

При этом задачами дисциплины являются:

- знать типы задач, модели и алгоритмы машинного обучения, особенности нейронных сетей;
- выбирать современные инструментальные средства, устанавливать необходимые библиотеки и пакеты для анализа данных;
- применять методы и алгоритмы, используемые в сфере машинного обучения для разработки приложений на языке Python.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

2.1. Дисциплина «Анализ данных и машинное обучение» относится к части, формируемой участниками образовательных отношений Блока 1. Дисциплины (модули), имеет тесную связь с другими дисциплинами.

2.2. В таблице приведены предшествующие и последующие дисциплины, направленные на формирование компетенций дисциплины в соответствии с матрицей компетенций ОП.

Предшествующие дисциплины, направленные на формирование компетенций

№ п/п	Предшествующие дисциплины	Последующие дисциплины
1	Алгоритмы и структуры данных	Информационная безопасность и защита информации

3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Планируемые результаты освоения образовательной программы (ОП) – компетенции обучающихся определяются требованиями стандарта по направлению подготовки 09.03.04 Программная инженерия и формируются в соответствии с матрицей компетенций ОП

№ п/п	Номер/ индекс компетенции	Наименование компетенции (или ее части)	В результате изучения дисциплины обучающиеся должны:
1	2	3	4
1.	ПК-1	Способен использовать методы и инструментальные средства исследования объектов профессиональной деятельности	ПК-1.1. Выбирает современные инструментальные средства для проектирования и конструирования программного обеспечения ПК-1.3. Осуществляет поиск, хранение обработку и анализ информации из различных источников, представляет в требуемом формате с использованием информационных технологий ПК-1.4. Применяет прикладные аспекты и инструментальные средства и методы в современных программных комплексах
	ПК-2	Способен обосновать принимаемые проектные решения, осуществлять постановку и выполнение экспериментов по проверке их корректности и эффективности	ПК-2.1. Разрабатывает и управляет проектами, и реализует их с использованием современного программного обеспечения. ПК-2.4. Оценивает временную и емкостную сложность программного обеспечения ПК-2.6. Работает с современными системами программирования и проектирования

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

4.1. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Вид учебной работы	Всего часов	Семестр	Семестр	
		№ 7	№ 8	
		Часов	Часов	
1	2	3		
Аудиторная контактная работа (всего)	102	42	60	
В том числе:				
Лекции (Л)	34	14	20	
Практические занятия (ПЗ), Семинары (С)	-	-	-	
Лабораторные работы (ЛР)	68	28	40	
Контактная внеаудиторная работа, в том числе:	3,7	1,7	2	
Групповые и индивидуальные консультации	3,7	1,7	2	
Самостоятельная работа обучающегося (СРО) (всего)	119	64	55	
Работа с книжными источниками	22	12	10	
Подготовка к лабораторным занятиям	23	12	11	
Выполнение индивидуальных заданий	26	14	12	
Работа с электронными источниками	26	14	12	
Подготовка к тестированию	22	12	10	
Промежуточная аттестация	Зачет	3	3	
	экзамен (Э) в том числе:	Э (27)	Э (27)	
	Прием экз., час.	0,8	0,3	
	Консультация, час.	2	2	
	СРО, час.	24,5	24,5	
ИТОГО:	Часов	252	108	144
Общая трудоемкость	зачетных единиц	7	3	4

СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

4.2.1. Разделы (темы) дисциплины, виды учебной деятельности и формы контроля

№ п/п	Наименование раздела дисциплины	Виды учебной деятельности, включая самостоятельную работу обучающихся (в часах)					Формы текущей и промежуточной аттестации)
		Л	ЛР	ПЗ	СРО	Всего	
1	2	3	4	5	6	7	8
Семестр 7							
1.	Раздел 1. Среда параллельного программирования Python	8	8	-	20	36	лабораторные работы индивидуальные задания
2.	Раздел 2. Встроенные и дополнительные функции и методы для математических вычислений в Python	4	10	-	22	36	Лабораторные работы, индивидуальные задания
3	Раздел 3. Параллельное программирование в среде OpenMP	2	10	-	22	34	тестирование, индивидуальные задания
4.	Контактная внеаудиторная работа					1,7	Индивидуальные и групповые консультации
5.	Промежуточная аттестация					0,3	зачет
Итого часов в 7 семестре:		14	28	-	64	108	

№ п/п	Наименование раздела дисциплины	Виды учебной деятельности, включая самостоятельную работу обучающихся (в часах)					Формы текущей и промежуточн ой аттестации)
		Л	ЛР	ПЗ	СРО	Всего	
1	2	3	4	5	6	7	8
Семестр 8							
1.	Раздел 1. Введение. О задачах машинного обучения.	8	-	-	16	34	Лабораторные работы, индивидуальные задания
2.	Раздел 2. Модели и алгоритмы машинного обучения.	4	-	-	17	35	Лабораторные работы, индивидуальные задания
3	Раздел 3. Разработка приложений в сфере машинного обучения.	8	40	-	22	46	тестирование, индивидуальные задания
4.	Контактная внеаудиторная работа					2	Индивидуальные и групповые консультации
5.	Промежуточная аттестация					27	экзамен
Итого часов в 8 семестре:		20	40	-	55	144	
Всего:		34	68		119	252	

4.2.2. Лекционный курс

№ п/п	Наименование раздела дисциплины	Наименование темы лекции	Содержание лекции	Всего часов
Семестр 7				
1	2	3	4	5
1.	Раздел 1. Среда параллельного программирования Python	Тема 1.1 История языка программирования Python. Типы данных, определение переменной, логические выражения.	Python – интерпретируемый язык программирования. Синтаксис. Интерактивный режим, интерпретатор. Создание скриптов. Типы данных: целые числа, числа с плавающей точкой, строки. Операции присвоения. Логический тип данных. Логические операторы. Сложные логические выражения.	2
		Тема 1.2 Условный оператор, инструкция if. Множественное ветвление. Цикл While.	Конструкции if — заголовок, в котором проверяется условие выполнения строк кода после двоеточия. Сложная форма ветвления if–else. Инструкция if-elif-else. Циклы и счетчики.	2
		Тема 1.3 Ввод данных с клавиатуры. Строки как последовательность и символов. Списки. Словари.	Функция print. Функция input. Аргумент- приглашение строкового типа. Функция len, позволяющая измерить длину строки. Операции конкатенации (+) и дублирования (*). Оператор индексирования. Словарь - изменяемый неупорядоченный набор пар "ключ:значение". Функция del.	2
		Тема 1.4 Цикл for языке программирования Python. Функции. Параметры и аргументы функций. Локальные и глобальные переменные.	Цикл for - цикл обхода, заданного множества элементов и выполнения в своем теле различных операций над ними. Инструкция def - команда языка программирования Python, позволяющая создавать функцию.	2

2	Раздел 2. Встроенные и дополнительные функции и методы для математических вычислений в Python	Тема 2.1 Линейные программы. Разветвляющиеся вычислительные процессы. Организация циклов.	Модуль math. Условные операторы ветвления. Графики функций. Базовые конструкции структурного программирования - операторы цикла.	2
		Тема 2.2 Одномерные массивы. Двумерные массивы и функции. Файлы.	Обработка одномерных массивов: поиск максимумов и минимумов, сортировка. Обработка двумерных массивов с использованием функций. Операции с текстовыми файлами - чтение, запись.	2
3.	Раздел 3. Параллельное программирование в среде OpenMP	Тема 3.1 Модель программирования в общей памяти Модель "пульсирующего" параллелизма FORK-JOIN. Параллельные алгоритмы	Нити разделяющие общую память. Обмены между нитями посредством чтения/записи данных в общей памяти. Выполнение нити на различных ядрах одного процессора. Алгоритм parallel_for Алгоритм parallel_for_each Алгоритм parallel_invoke Алгоритмы parallel_transform и parallel_reduce Алгоритм parallel_transform Алгоритм parallel_reduce	2
Итого часов в 7 семестре:				14

№ п/п	Наименование раздела дисциплины	Наименование темы лекции	Содержание лекции	Всего часов
Семестр 8				
1	2	3	4	5
1.	Раздел 1. Введение. О задачах машинного обучения.	Тема 1.1 Алгоритм машинного обучения	Пространство признаков. Формальное определение понятия «обучение». Общий алгоритм машинного обучения.	2
		Тема 1.2 Задачи машинного обучения	Типы задач машинного обучения. Способы обучения и оценки его качества.	2

		Тема 1.2 Типовые задачи при подготовке данных и обучении моделей. Переобучение	Учет пропусков. Кодирование нечисловых признаков. Приведение данных к единому масштабу и стандартизация. Разметка данных.	4
2	Раздел 2. Модели и алгоритмы машинного обучения.	Модели и алгоритмы машинного обучения.	Методы теории вероятностей. Деревья решений. Статистические модели и методы. Модели и методы нечеткой логики	4
3	Раздел 3. Разработка приложений в сфере машинного обучения.	Разработка приложений в сфере машинного обучения.	Основы работы с Python. Элементарные операции с данными. Работа с DataFrame. Предобработка данных. Работа с деревьями решений. Сохранение и загрузка обученной модели. Решение задачи ранжирования признаков. Работа с простейшими моделями нейронных сетей.	8
Итого часов в 8 семестре:				20
Всего часов				34

4.2.3. Лабораторный практикум

№ п/п	Наименование раздела дисциплины	Наименование лабораторного занятия	Содержание лабораторного занятия	Всего часов
Семестр 7				
1	2	3	4	5
1.	Раздел 1. Среда параллельного программирования Python	Интерактивный режим для написания программ	Присвоив значения 3 переменным, написать программу используя операции конкатенации (+) и повторения строки (*). Создайте скрипт (файл data.py), который бы запрашивал у пользователя - его имя.	8
2.	Раздел 2. Встроенные и Дополнительные функции и методы для Математических вычислений в Python	Структурное программирование	Линейные программы. Разветвляющиеся вычислительные процессы. Организация циклов Одномерные массивы. Двумерные массивы и подпрограммы.	10
3.	Раздел 3. Параллельное программирование в среде OpenMP	Разработка параллельных программ с применением технологии OpenMP	Модель параллельной программы OpenMP. Директивы и функции OpenMP. Компиляция и выполнение OpenMP программы. Параллельные секции, переменные среды замер времени.	4
		Варианты распределения работы между нитями параллельной программы	Низкоуровневое программирование. Распараллеливание оператора цикла.	6
Итого часов в 7 семестре:				28

№ п/п	Наименование раздела дисциплины	Наименование лабораторного занятия	Содержание лабораторного занятия	Всего часов
Семестр 8				
1	2	3	4	5
1.	Раздел 3 Разработка приложений в сфере машинного обучения.	Выбор аппаратной и программной составляющей	Установка пакетов Python: AnacondaEnvironment, PyCharm, CommunityEdition, TDM-GCC	4
2.		Элементарные операции с данными	Изучение способов получения данных и возможностей библиотеки Numpy по операциям с матрицами	2
3.		Работа с DataFrame	Изучение способов работы с таблицами DataFrame	4
		Предобработка данных. Стандартизация и нормализация	Изучение процедур стандартизации и нормализации с помощью библиотеки scikit-learn(класса StandardScaler)	4
		Работа с деревьями решений	Изучение особенностей решающих деревьев с помощью библиотеки scikit-learn	4
		Работа с линейной регрессией	Решение задачи прогнозирования уровня зарплаты по данным вакансий.	2
		Работа с логистической регрессией	Решение задачи классификации, знакомство с моделью логистической регрессии	4
		Сохранение и загрузка обученной модели	Изучение практики сохранения и загрузки ее параметров в\из файл(а).	2
		Решение задачи ранжирования признаков	Применение регрессионных моделей для определения важности признаков	4
		Работа с полиномиальной регрессией	Знакомство с переобучением модели	2
		Работа с простейшими и моделями нейронных сетей	Изучение простейших моделей нейронной сети на примере библиотеки Sklearn	4
		Реализация алгоритма обучения нейронной сети	Изучение алгоритма обратного распространения ошибки	2
		Регуляция и сеть	Изучение общих рекомендаций	2

		прямого распространения	по выполнению экспериментов	
Итого часов в 8 семестре:				40
Всего часов:				68

4.2.4. Практические занятия не предполагаются

4.3. САМОСТОЯТЕЛЬНАЯ РАБОТА ОБУЧАЮЩЕГОСЯ

№ п/п	Наименование раздела (темы) дисциплины	№ п/п	Виды СРО	Всего часов
1	2	3	4	5
Семестр 7				
1.	Раздел 1. Среда параллельного программирования Python	1.1.	Работа с лекциями, книжными источниками, подготовка к тестированию	8
		1.2.	Выполнение домашнего задания по темам лабораторных занятий	8
		1.3.	Подготовка к лабораторным занятиям	4
2.	Раздел 2. Встроенные и дополнительные функции и методы для математических вычислений в Python	2.1.	Самоподготовка: внеаудиторное чтение, работа с электронными источниками	6
		2.2.	Выполнение индивидуальных заданий	8
		2.3.	Контактная внеаудиторная работа	8
3.	Раздел 3. Параллельное программирование в среде OpenMP	3.1.	Работа с лекциями, книжными источниками, подготовка к текущему контролю	6
		3.2.	Выполнение домашнего задания по темам лабораторных занятий	8
		3.3.	Подготовка к лабораторным занятиям	8
Итого часов в 7 семестре:				64

№ п/п	Наименование раздела (темы) дисциплины	№ п/п	Виды СРО	Всего часов
1	2	3	4	5
Семестр 8				
1.	Раздел 1. Введение. О задачах машинного обучения.	1.1.	Работа с лекциями, книжными источниками, подготовка к текущему контролю	6
		1.2.	Выполнение домашнего задания по темам лабораторных занятий	6
		1.3.	Подготовка к лабораторным занятиям	4
2.	Раздел 2. Модели и алгоритмы машинного обучения.	2.1.	Самоподготовка: внеаудиторное чтение, работа с электронными источниками	6
		2.2.	Выполнение индивидуальных заданий Тестирование	6

		2.3	Контактная внеаудиторная работа	5
3.	Раздел 3. Разработка приложений в сфере машинного обучения.	3.1.	Работа с лекциями, книжными источниками, подготовка к текущему тестовому контролю	8
		3.2.	Выполнение домашнего задания по темам лабораторных занятий	8
		3.3.	Подготовка к лабораторным занятиям	6
Итого часов в 8 семестре:				55
Всего часов:				119

5. ПЕРЕЧЕНЬ УЧЕБНО-МЕТОДИЧЕСКОГО ОБЕСПЕЧЕНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

5.1. Методические указания для подготовки обучающихся к лекционным занятиям

Лекция является основной формой обучения в высшем учебном заведении. Записи лекций в конспектах должны быть избирательными, полностью следует записывать только определения. В конспекте рекомендуется применять сокращение слов, что ускоряет запись. Вопросы, возникающие в ходе лекции, рекомендуется записывать на полях и после окончания лекции обратиться за разъяснением к преподавателю.

Работа над конспектом лекции осуществляется по этапам:

- повторить изученный материал по конспекту;
- непонятные положения отметить на полях и уточнить;
- неоконченные фразы, пропущенные слова и другие недочеты в записях устранить, пользуясь материалами из учебника и других источников;
- завершить техническое оформление конспекта (подчеркивания, выделение главного, выделение разделов, подразделов и т.п.).

Самостоятельную работу следует начинать с доработки конспекта, желательно в тот же день, пока время не стерло содержание лекции из памяти. Работа над конспектом не должна заканчиваться с прослушивания лекции. После лекции, в процессе самостоятельной работы, перед тем, как открыть тетрадь с конспектом, полезно мысленно восстановить в памяти содержание лекции, вспомнив ее структуру, основные положения и выводы.

С целью доработки необходимо прочитать записи, восстановить текст в памяти, а также исправить опiski, расшифровать не принятые ранее сокращения, заполнить пропущенные места, понять текст, вникнуть в его смысл. Далее прочитать материал по рекомендуемой литературе, разрешая в ходе чтения, возникшие ранее затруднения, вопросы, а также дополнения и исправляя свои записи. Записи должны быть наглядными, для чего следует применять различные способы выделений. В ходе доработки конспекта углубляются, расширяются и закрепляются знания, а также дополняется, исправляется и совершенствуется конспект. Еще лучше, если вы переработаете конспект, дадите его в новой систематизации записей. Это, несомненно, займет некоторое время, но материал вами будет хорошо проработан, а конспективная запись его приведена в удобный для

запоминания вид. Введение заголовков, скобок, обобщающих знаков может значительно повысить качество записи. Этому может служить также подчеркивание отдельных мест конспекта красным карандашом, приведение на полях или на обратной стороне листа краткой схемы конспекта и др.

Подготовленный конспект и рекомендуемая литература используется при подготовке к практическому занятию. Подготовка сводится к внимательному прочтению учебного материала, к выводу с карандашом в руках всех утверждений и формул, к решению примеров, задач, к ответам на вопросы, предложенные в конце лекции преподавателем или помещенные в рекомендуемой литературе. Примеры, задачи, вопросы по теме являются средством самоконтроля.

Непременным условием глубокого усвоения учебного материала является знание основ, на которых строится изложение материала. Обычно преподаватель напоминает, какой ранее изученный материал и в какой степени требуется подготовить к очередному занятию. Эта рекомендация, как и требование систематической и серьезной работы над всем лекционным курсом, подлежит безусловному выполнению. Потери логической связи как внутри темы, так и между ними приводит к негативным последствиям: материал учебной дисциплины перестает основательно восприниматься, а творческий труд подменяется утомленным переписыванием. Обращение к ранее изученному материалу не только помогает восстановить в памяти известные положения, выводы, но и приводит разрозненные знания в систему, углубляет и расширяет их. Каждый возврат к старому материалу позволяет найти в нем что-то новое, переосмыслить его с иных позиций, определить для него наиболее подходящее место в уже имеющейся системе знаний. Неоднократное обращение к пройденному материалу является наиболее рациональной формой приобретения и закрепления знаний. Очень полезным, но, к сожалению, еще мало используемым в практике самостоятельной работы, является предварительное ознакомление с учебным материалом. Даже краткое, беглое знакомство с материалом очередной лекции дает многое. Обучающиеся получают общее представление о ее содержании и структуре, о главных и второстепенных вопросах, о терминах и определениях. Все это облегчает работу на лекции и делает ее целеустремленной.

5.2. Методические указания для подготовки обучающихся к лабораторным занятиям

Ведущей дидактической целью лабораторных занятий является систематизация и обобщение знаний по изучаемой теме, приобретение практических навыков по тому или другому разделу курса, закрепление полученных теоретических знаний. Лабораторные работы сопровождают и поддерживают лекционный курс. Подготовка к лабораторным занятиям и практикумам носит различный характер, как по содержанию, так и по сложности исполнения.

Многие лабораторные занятия требуют большой исследовательской работы, изучения дополнительной научной литературы. Прежде чем приступить к выполнению такой работы, обучающемуся необходимо ознакомиться обстоятельно с содержанием задания, уяснить его, оценить с точки зрения восприятия и запоминания все составляющие его компоненты. Это очень важно, так как при проработке соответствующего материала по конспекту лекции или по рекомендованной литературе могут встретиться определения, факты, пояснения, которые не относятся непосредственно к заданию. Обучающийся должен хорошо знать и понимать содержание задания, чтобы быстро оценить и отобрать нужное из читаемого. Далее, в соответствии со списком рекомендованной литературы, необходимо отыскать материал к данному заданию по всем пособиям.

Весь подобранный материал нужно хотя бы один раз прочитать или внимательно просмотреть полностью. По ходу чтения помечаются те места, в которых содержится ответ на вопрос, сформулированный в задании. Читая литературу по теме, обучающийся должен мысленно спрашивать себя, на какой вопрос задания отвечает тот или иной абзац прорабатываемого пособия. После того, как материал для ответов подобран, желательно хотя бы мысленно, а лучше всего устно или же письменно, ответить на все вопросы. В случае если обнаружится пробел в знаниях, необходимо вновь обратиться к литературным источникам и проработать соответствующий раздел. Только после того, как преподаватель убедится, что обучающийся хорошо знает необходимый теоретический материал, что его ответы достаточно аргументированы и доказательны, можно считать обучающегося подготовленным к выполнению лабораторных работ.

5.3. Методические указания для подготовки обучающихся к практическим занятиям

Учебным планом не предусмотрено

5.4. Методические указания по самостоятельной работе обучающихся

Работа с литературными источниками и интернет ресурсами

В процессе подготовки к лабораторным занятиям, обучающимся необходимо обратить особое внимание на самостоятельное изучение рекомендованной учебно-методической (а также научной и популярной) литературы.

Самостоятельная работа с учебниками, учебными пособиями, научной, справочной и популярной литературой, материалами периодических изданий и Интернета, статистическими данными является наиболее эффективным методом получения знаний, позволяет значительно активизировать процесс овладения информацией, способствует более глубокому усвоению изучаемого материала, формирует у обучающихся свое отношение к конкретной проблеме.

Более глубокому раскрытию вопросов способствует знакомство с дополнительной литературой, рекомендованной преподавателем по каждой теме семинарского или практического занятия, что позволяет студентам проявить свою индивидуальность в рамках выступления на данных занятиях, выявить широкий спектр мнений по изучаемой проблеме.

Тестирование

Тестирование представляет собой средство контроля усвоения учебного материала темы или раздела дисциплины. При самостоятельной подготовке к тестированию обучающемуся необходимо:

а) проработать информационный материал по дисциплине, проконсультироваться с преподавателем по вопросу выбора учебной литературы;

б) выяснить все условия тестирования заранее, узнать, сколько тестов будет предложено, сколько времени отводится на тестирование, какова система оценки результатов и т.д.

в) приступая к работе с тестами, внимательно и до конца прочитать вопрос и предлагаемые варианты ответов, выбрать правильные (их может быть несколько).

г) в процессе решения желательно применять несколько подходов в решении задания, это позволяет максимально гибко оперировать методами решения, находя каждый раз оптимальный вариант.

д) если встретился чрезвычайно трудный вопрос, не тратить много времени на него, перейти к другим тестам, вернуться к трудному вопросу в конце.

е) обязательно оставить время для проверки ответов, чтобы избежать механических ошибок.

Методические указания по выполнению индивидуальных заданий

Индивидуальные задания оформляются в распечатанном или рукописном варианте. Номер варианта выбирается по порядковому номеру списка обучающихся. Задания с другим номером

варианта не зачитываются. Работа выполняется аккуратно, в случае рукописного оформления чтение ее не должно вызывать затруднений.

Задания должны состоять из титульного листа и основной части. Допускается включение в работу приложений, содержащих таблицы, рисунки, полученные на компьютере. На титульном листе обязательно указывается наименование дисциплины, ФИО обучающегося, группа, вариант задания, ФИО преподавателя. Выполненная и оформленная работа должна быть представлена преподавателю не позднее, чем за 10 дней до начала сессии.

В основной части заданий до решения каждой задачи должны быть представлены собственные данные: вариант задания, формулировка задания, численные значения, соответствующие своему варианту. Далее должно быть представлено решение с расшифровкой формул и последовательности действий.

Промежуточная аттестация

По итогам 7 семестра проводится зачет, 8 семестра проводится экзамен. При подготовке к сдаче зачета, экзамена рекомендуется пользоваться материалами практических занятий и материалами, изученными в ходе текущей самостоятельной работы.

Зачет, экзамен проводится в устной форме, включает подготовку и ответы обучающегося на теоретические вопросы. По итогам экзамена выставляется оценка.

По итогам обучения проводится экзамен, к которому допускаются студенты, имеющие положительные результаты по защите лабораторных работ.

6. Образовательные технологии

№ п/п	Виды учебной работы	Образовательные технологии	Всего часов
1	2	3	4
Семестр 7			
1	Лекция. История языка программирования Python. Типы данных, определение переменной, логические выражения.	Лекция –презентация	2
2	Лекция. Условный оператор, инструкция if. Множественное ветвление. Цикл While.	Лекция –презентация	2
3	Лекция Ввод данных с клавиатуры. Строки как последовательности символов. Списки. Словари.	Диалоговые технологии	2
4	Лекция Цикл for в языке программирования Python. Функции. Параметры и аргументы функций. Локальные и глобальные переменные.	Диалоговые технологии	4
5	Лабораторная работа Интерактивный режим для написания программ	Технология развития критического мышления	2
Итого часов в 7 семестре:			12
Семестр 8			
1	Лекция Типовые задачи при подготовке данных и обучении моделей. Переобучение.	Лекция –презентация	4
2	Лекция Модели и алгоритмы машинного обучения.	Лекция –презентация	2
3	Лекция Разработка приложений в сфере машинного обучения.	Диалоговые технологии	2
5	Лабораторная работа Работа с простейшими моделями нейронных сетей	Технология развития критического мышления	4
Итого часов в 8 семестре:			12
Всего часов:			24

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

7.1. Перечень основной и дополнительной учебной литературы

Основная литература

1. Теория и практика машинного обучения : учебное пособие / В. В. Воронина, А. В. Михеев, Н. Г. Ярушкина, К. В. Святков. — Ульяновск : Ульяновский государственный технический университет, 2017. — 291 с. — ISBN 978-5-9795-1712-4. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/106120.html>.
2. Воронова, Л. И. Machine Learning: регрессионные методы интеллектуального анализа данных : учебное пособие / Л. И. Воронова, В. И. Воронов. — Москва : Московский технический университет связи и информатики, 2018. — 82 с. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/81325.html>.
3. Замятин, А. В. Интеллектуальный анализ данных : учебное пособие / А. В. Замятин. — Замятин, А. В. Введение в интеллектуальный анализ данных : учебное пособие / А. В. Замятин. — Томск : Издательский Дом Томского государственного университета, 2016. — 119 с. — ISBN 978-5-94621-531-2. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/109021.html>.
4. Жуковский, О. И. Информационные технологии и анализ данных : учебное пособие / О. И. Жуковский. — Томск : Томский государственный университет систем управления и радиоэлектроники, Эль Контент, 2014. — 130 с. — ISBN 978-5-4332-0158-3. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/72106.html>. — Режим доступа: для авторизир. пользователей
5. Томск : Издательский Дом Томского государственного университета, 2020. — 194 с. — ISBN 978-5-94621-898-6. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/116889.html>.
6. Целых, А. Н. Современные методы прикладной информатики в задачах анализа данных : учебное пособие по курсу «Методы интеллектуального анализа данных» / А. Н. Целых, А. А. Целых, Э. М. Котов. — Ростов-на-Дону, Таганрог : Издательство Южного федерального университета, 2021. — 130 с. — ISBN 978-5-9275-3783-9. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/117165.html>.
7. Целых А.Н. Извлечение знаний методами машинного обучения [Электронный ресурс]: учебное пособие по курсам «Модели и методы инженерии знаний», «Методы машинного обучения»/ Целых А.Н., Котов Э.М.— Электрон. текстовые данные.— Ростов-на-Дону, Таганрог: Издательство Южного федерального университета, 2022.— 105 с.— Режим доступа: <https://ipr-smart.ru/131448> .— IPR SMART, по паролю
8. Целых, А. Н. Принятие решений на основе методов машинного обучения : учебное пособие по курсам «Модели и методы инженерии знаний», «Методы анализа больших данных» / А. Н. Целых, Н. В. Драгныш, Э. М. Котов. — Ростов-на-Дону, Таганрог : Издательство Южного федерального университета, 2022. — 113 с. — ISBN 978-5-9275-4246-8. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/131458.html>.

Дополнительная литература

1. Запечников, С. В. Основы интеллектуального анализа данных и машинного обучения: конспект лекций : учебное пособие / С. В. Запечников. — Москва : Национальный исследовательский ядерный университет «МИФИ», 2022. — 136 с. — ISBN 978-5-7262-2856-3. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/132687.html>.
2. Обработка данных магниторазведки: обратные задачи и машинное обучение : учебное пособие / Д. С. Киселев, А. М. Гриф, М. Г. Персова, Ю. Г. Соловейчик. — Новосибирск

- : Новосибирский государственный технический университет, 2021. — 76 с. — ISBN 978-5-7782-4556-3. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/126576..html>.
3. Шнарева, Г. В. Анализ данных : учебно-методическое пособие / Г. В. Шнарева, Ж. Г. Пономарева. — Симферополь : Университет экономики и управления, 2019. — 129 с. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/89482.html>.
 4. Сараев, П. В. Методы машинного обучения : методические указания и задания к лабораторным работам по курсу / П. В. Сараев. — Липецк : Липецкий государственный технический университет, ЭБС АСВ, 2017. — 48 с. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/83183.html>.

Методическая литература

нет

7.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»
<https://intuit.ru/> Сайт национального открытого университета ИНТУИТ

7.3. Информационные технологии, лицензионное программное обеспечение.

Лицензионное программное обеспечение:

Лицензионное программное обеспечение	Реквизиты лицензий/ договоров
Microsoft Azure Dev Tools for Teaching 1. Windows 7, 8, 8.1, 10 2. Visual Studio 2008, 2010, 2013, 2019 5. Visio 2007, 2010, 2013 6. Project 2008, 2010, 2013 7. Access 2007, 2010, 2013 и т. д.	Идентификатор подписчика: 1203743421 Срок действия: 30.06.2022 (продление подписки)
MS Office 2003, 2007, 2010, 2013	Сведения об Open Office: 63143487, 63321452, 64026734, 6416302, 64344172, 64394739, 64468661, 64489816, 64537893, 64563149, 64990070, 65615073 Лицензия бессрочная
Антивирус Dr.Web Desktop Security Suite	Лицензионный сертификат Серийный № 8DVG-V96F-H8S7-NRBC Срок действия: с 20.10.2022 до 22.10.2023
Цифровой образовательный ресурс IPRsmart	Лицензионный договор № 10423/23П от 30.06.2023 г. Срок действия: с 01.07.2022 г. до 01.07.2023г.

Свободное ПО: Python

8. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

8.1. Требования к аудиториям (помещениям, местам) для проведения занятий

1. Учебная аудитория для проведения занятий лекционного типа

Специализированная мебель:

Стол преподавательский - 1 шт., компьютерные столы - 10 шт., парты - 7 шт., стулья - 24 шт., доска меловая - 1 шт.

Лабораторное оборудование, технические средства обучения, служащие для предоставления учебной информации большой аудитории:

Персональный компьютер – 10 шт.

Экран настенный рулонный – 1 шт.

2. Лаборатория компьютерной графики

Специализированная мебель:

Стол преподавательский - 1 шт., компьютерные столы - 10 шт., парты - 7 шт., стулья - 24 шт., доска меловая - 1 шт.

Лабораторное оборудование, технические средства обучения, служащие для предоставления учебной информации большой аудитории:

Персональный компьютер – 10 шт.

Экран настенный рулонный – 1 шт.

3. Учебная аудитория для проведения занятий семинарского типа, курсового проектирования (выполнение курсовых работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации

Специализированная мебель:

Доска меловая - 1 шт., парты - 35 шт., стулья - 66 шт., кафедра настольная - 1 шт.

Набор демонстрационного оборудования и учебно-наглядных пособий, обеспечивающих тематические иллюстрации:

Настенный экран - 1 шт.

Проектор - 1 шт.

Ноутбук – 1 шт.

Специализированная мебель:

Стол преподавательский - 1 шт., компьютерные столы - 13 шт., стул мягкий - 1 шт., парты - 7 шт., стулья - 21 шт., доска меловая - 1 шт.

Технические средства обучения, служащие для предоставления учебной информации большой аудитории:

Сист. блок iRu Ergo-Corp 1294 E8400(3000)1024 80 DVD-RW FDD K – 1 шт.;

Сист. Блок Intel P Dual E5700/HDD 500Gb/DDR3 2Gb – 9 шт.;

Монитор LCD e`machines LCD 18.5' – 8 шт.;

Монитор LOC 18.5' – 2 шт.;

Экран настенный рулонный 200x200 – 1 шт.

4. Помещение для самостоятельной работы. Библиотечно-издательский центр.

Отдел обслуживания печатными изданиями

Специализированная мебель:

Рабочие столы на 1 место – 21 шт.

Стулья – 55 шт.

Набор демонстрационного оборудования и учебно-наглядных пособий, обеспечивающих тематические иллюстрации:

Экран настенный – 1 шт.

Проектор – 1 шт.

Ноутбук – 1 шт.

Информационно-библиографический отдел.

Специализированная мебель:

Рабочие столы на 1 место - 6 шт.

Стулья - 6 шт.

Компьютерная техника с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду ФГБОУ ВО «СевКавГА»:

Персональный компьютер – 1 шт.

Сканер – 1 шт.

МФУ – 1 шт.

Отдел обслуживания электронными изданиями

Специализированная мебель:

Рабочие столы на 1 место – 24 шт.

Стулья – 24 шт.

Набор демонстрационного оборудования и учебно-наглядных пособий, обеспечивающих тематические иллюстрации:

Интерактивная система - 1 шт.

Монитор – 21 шт.

Сетевой терминал -18 шт.

Персональный компьютер -3 шт.

МФУ – 2 шт.

Принтер –1шт.

8.2. Требования к оборудованию рабочих мест преподавателя и обучающихся

1. Рабочее место преподавателя, оснащенное компьютером с доступом в Интернет.

2. Рабочие места обучающихся, оснащенные компьютерами с доступом в Интернет, предназначенные для работы в электронной образовательной среде.

8.3. Требования к специализированному оборудованию

- нет

9. ОСОБЕННОСТИ РЕАЛИЗАЦИИ ДИСЦИПЛИНЫ ДЛЯ ИНВАЛИДОВ И ЛИЦ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ

Для обеспечения образования инвалидов и обучающихся с ограниченными возможностями здоровья разрабатывается (в случае необходимости) адаптированная образовательная программа, индивидуальный учебный план с учетом особенностей их психофизического развития и состояния здоровья, в частности применяется индивидуальный подход к освоению дисциплины, индивидуальные задания: рефераты, письменные работы и, наоборот, только устные ответы и диалоги, индивидуальные консультации, использование диктофона и других записывающих средств для воспроизведения лекционного и семинарского материала.

В целях обеспечения обучающихся инвалидов и лиц с ограниченными возможностями здоровья комплектуется фонд основной учебной литературой, адаптированной к ограничению электронных образовательных ресурсов, доступ к которым организован в БИЦ Академии. В библиотеке проводятся индивидуальные консультации для данной категории пользователей, оказывается помощь в регистрации и использовании сетевых и локальных электронных образовательных ресурсов, предоставляются места в читальном зале.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

ПО ДИСЦИПЛИНЕ «Анализ данных и машинное обучение»

1. ПАСПОРТ ФОНДА ОЦЕНОЧНЫХ СРЕДСТВ ПО ДИСЦИПЛИНЕ

«Анализ данных и машинное обучение»

Компетенции, формируемые в процессе изучения дисциплины

Индекс	Формулировка компетенции
ПК-1	Способен использовать методы и инструментальные средства исследования объектов профессиональной деятельности
ПК-2	Способен обосновать принимаемые проектные решения, осуществлять постановку и выполнение экспериментов по проверке их корректности и эффективности

2. Этапы формирования компетенции в процессе освоения дисциплины

Основными этапами формирования указанных компетенций при изучении обучающимися дисциплины являются последовательное изучение содержательно связанных между собой разделов (тем) учебных занятий. Изучение каждого раздела (темы) предполагает овладение обучающимися необходимыми компетенциями. Результат аттестации обучающихся на различных этапах формирования компетенций показывает уровень освоения компетенций студентами.

Этапность формирования компетенций прямо связана с местом дисциплины в образовательной программе.

Разделы (темы) дисциплины	Формируемые компетенции (коды)	
	ПК-1	ПК-2
1	2	3
Раздел 1. Введение. О задачах машинного обучения.	+	+
Раздел 1. Среда параллельного программирования Python	+	+
Раздел 2. Модели и алгоритмы машинного обучения.	+	+
Раздел 2. Встроенные и дополнительные функции и методы для математических вычислений в Python	+	+
Раздел 3. Разработка приложений в сфере машинного обучения.	+	+
Раздел 3. Параллельное программирование в среде OpenMP	+	+

3.

Показатели, критерии и средства оценивания компетенций, формируемых в процессе изучения дисциплины

ПК-1 Способен использовать методы и инструментальные средства исследования объектов профессиональной деятельности

Индикаторы достижения компетенции	Критерии оценивания результатов обучения				Средства оценивания результатов обучения	
	неудовлетв	удовлетв	хорошо	отлично	Текущий контроль	Промежуточная аттестация
1	2	3	4	5	6	7
ПК-1.1. Выбирает современные инструментальные средства для проектирования и конструирования программного обеспечения	Не знает основные принципы применения современных инструментальных средств для проектирования и конструирования программного обеспечения	Неполные представления о принципах применения современных инструментальных средств для проектирования и конструирования программного обеспечения	Сформированные, но содержащие отдельные пробелы представления о принципах применения современных инструментальных средств для проектирования и конструирования программного обеспечения	Сформированы систематические представления о принципах применения современных инструментальных средств для проектирования и конструирования программного обеспечения	выполнение индивидуальных заданий, лабораторные работы, тестирование	7 семестр – зачет 8 семестр-экзамен
ПК-1.3. Осуществляет поиск, хранение обработку и анализ информации из различных источников, представляет в требуемом формате с использованием информационных технологий	Не умеет и не готов осуществлять поиск, хранение обработку и анализ информации из различных источников, представляет в требуемом формате с использованием информационных технологий	В целом успешное, но с систематическими ошибками осуществляет поиск, хранение обработку и анализ информации из различных источников, представляет в требуемом формате с использованием информационных технологий	В целом успешное, но содержащее отдельные пробелы умение осуществлять поиск, хранение обработку и анализ информации из различных источников, представляет в требуемом формате с использованием информационных технологий	Сформированное умение осуществлять поиск, хранение обработку и анализ информации из различных источников, представляет в требуемом формате с использованием информационных технологий		7 семестр– зачет 8 семестр -экзамен
ПК-1.4. Применяет прикладные аспекты и инструментальные средства и методы в современных программных комплексах	Не умеет применять прикладные аспекты и инструментальные средства и методы в современных программных комплексах	Не достаточно хорошо применяет прикладные аспекты и инструментальные средства и методы в современных программных комплексах	В целом успешно, но с отдельными пробелами применяет прикладные аспекты и инструментальные средства и методы в современных программных комплексах	Достаточно успешно применяет прикладные аспекты и инструментальные средства и методы в современных программных комплексах		7 семестр– зачет 8 семестр -экзамен

ПК-2 Способен обосновать принимаемые проектные решения, осуществлять постановку и выполнение экспериментов по проверке их корректности и эффективности

Индикаторы достижения компетенции	Критерии оценивания результатов обучения				Средства оценивания результатов обучения	
	неудовлетв	удовлетв	хорошо	отлично	Текущий контроль	Промежуточная аттестация
1	2	3	4	5	6	7
ПК-2.1. Разрабатывает и управляет проектами, и реализует их с использованием современного программного обеспечения	Не знает основ управления проектами, и не реализует их с использованием современного программного обеспечения	Не разрабатывает и управляет проектами, и реализует их с использованием современного программного обеспечения	Сформировано, но с отдельными пробелами разрабатывает и управляет проектами, и реализует их с использованием современного программного обеспечения	Успешно разрабатывает и управляет проектами, и реализует их с использованием современного программного обеспечения	выполнение индивидуальных заданий, лабораторные работы, тестирование	7 семестр – зачет 8 семестр – экзамен
ПК-2.4. Оценивает временную и емкостную сложность программного обеспечения	Не умеет оценивать временную и емкостную сложность программного обеспечения	В целом успешно, но с систематическими ошибками оценивает временную и емкостную сложность программного обеспечения	В целом успешно, но содержащее отдельные пробелы умение оценивать временную и емкостную сложность программного обеспечения	В достаточной форме сформированное умение оценивать временную и емкостную сложность программного обеспечения		7 семестр – зачет 8 семестр – экзамен
ПК-2.6. Работает с современными системами программирования и проектирования	Не умеет работать с современными системами программирования и проектирования	Не достаточно хорошо работает с современными системами программирования и проектирования	В целом успешно, но с отдельными пробелами работает с современными системами программирования и проектирования	Достаточно успешно работает с современными системами программирования и проектирования		7 семестр – зачет 8 семестр – экзамен

4. Комплект контрольно-оценочных средств по дисциплине «Анализ данных и машинное обучение»

Вопросы к зачету по дисциплине Анализ данных и машинное обучение

1. История языка программирования Python.
2. Типы данных, определение переменной, логические выражения.
3. Python –интерпретируемый язык программирования.
4. Синтаксис. Интерактивный режим, интерпретатор.
5. Создание скриптов.
6. Типы данных: целые числа, числа с плавающей точкой, строки.
7. Операции присвоения.
8. Логический тип данных. Логические операторы. Сложные логические выражения.
9. Условный оператор, инструкция if. Множественное ветвление. Цикл While.
10. Конструкции if — заголовок, в котором проверяется условие выполнения строк кода после двоеточия. Сложная форма ветвления if–else. Инструкция if-elif-else. Циклы и счетчики.
11. Ввод данных с клавиатуры. Строки как последовательности символов. Списки.
12. Словари. Функция print. Функции input. Аргумент-приглашение строкового типа. Функция len, позволяющая измерить длину строки.
13. Операции конкатенации (+) и дублирования (*). Оператор индексирования. Словарь – изменяемый неупорядоченный набор пар "ключ:значение". Функция del.
14. Цикл for в языке программирования Python. Функции. Параметры и аргументы функций. Локальные и глобальные переменные.
15. Цикл for - цикл обхода, заданного множества элементов и выполнения в своем теле различных операций над ними. Инструкция def -команда языка программирования Python, позволяющая создавать функцию.
16. Основы работы с Python.
17. Элементарные операции с данными. Работа с DataFrame.
18. Предобработка данных.
19. Работа с деревьями решений.
20. Сохранение и загрузка обученной модели.
21. Решение задачи ранжирования признаков.
22. Работа с простейшими моделями нейронных сетей.

Вопросы к экзамену по дисциплине Анализ данных и машинное обучение

1. История языка программирования Python.
2. Типы данных, определение переменной, логические выражения.
3. Python –интерпретируемый язык программирования.
4. Синтаксис. Интерактивный режим, интерпретатор.
5. Создание скриптов.
6. Типы данных: целые числа, числа с плавающей точкой, строки.
7. Операции присвоения.

8. Логический тип данных. Логические операторы. Сложные логические выражения.
9. Условный оператор, инструкция if. Множественное ветвление.
10. Цикл While.
11. Конструкции if — заголовок, в котором проверяется условие выполнения строк кода после двоеточия.
12. Сложная форма ветвления if-else. Инструкция if-elif-else.
13. Циклы и счетчики.
14. Ввод данных с клавиатуры. Строки как последовательности символов. Списки.
15. Словари. Функция print.
16. Функции input. Аргумент-приглашение строкового типа. Функция len, позволяющая измерить длину строки.
17. Операции конкатенации (+) и дублирования (*). Оператор индексирования. Словарь — изменяемый неупорядоченный набор пар "ключ:значение". Функция del.
18. Цикл for в языке программирования Python.
19. Функции. Параметры и аргументы функций. Локальные и глобальные переменные.
20. Цикл for - цикл обхода, заданного множества элементов и выполнения в своем теле различных операций над ними.
21. Инструкция def - команда языка программирования Python, позволяющая создавать функцию.
22. Пространство признаков.
23. Формальное определение понятия «обучение». Общий алгоритм машинного обучения.
24. Типы задач машинного обучения. Способы обучения и оценки его качества.
25. Модуль math. Условные операторы ветвления.
26. Графики функций. Базовые конструкции структурного программирования - операторы цикла.
27. Обработка одномерных массивов: поиск максимумов и минимумов, сортировка.
Обработка двумерных массивов с использованием функций. Операции с текстовыми файлами - чтение, запись.
28. Учет пропусков. Кодирование нечисловых признаков. Приведение данных к единому масштабу и стандартизация.
29. Разметка данных.
30. Модель программирования в общей памяти.
31. Модель "пульсирующего" параллелизма FORK-JOIN
32. Нити разделяющие общую память. Обмены между нитями посредством чтения/записи данных в общей памяти.
33. Параллельные алгоритмы.
34. Основы работы с Python.
35. Элементарные операции с данными. Работа с DataFrame.
36. Предобработка данных.
37. Работа с деревьями решений.
38. Сохранение и загрузка обученной модели.
39. Решение задачи ранжирования признаков.
40. Работа с простейшими моделями нейронных сетей.

Задачи к экзамену
по дисциплине Анализ данных и машинное обучение

1. Дан двумерный массив X , состоящий из $n \times m$ элементов. Найти максимальный элемент двумерного массива .
2. Транспонированная матрица— матрица A^T , полученная из исходной матрицы A заменой строк на столбцы. Формально, транспонированная матрица для матрицы A размеров $m \times n$ — матрица A^T размеров $n \times m$, определённая как $A^T_{ij} = A_{ji}$. То есть для получения транспонированной матрицы из исходной нужно каждую строчку исходной матрицы записать в виде столбца в том же порядке.
3. Результатом перемножения матриц A и B является матрица C , каждый элемент которой есть скалярное произведение соответствующих строк матрицы A и столбцов матрицы B . Задача состоит в получении результата перемножения двух случайно заданных квадратных матриц. Для решения используются последовательный и параллельный алгоритмы, результаты которых нужно сравнить. Параллельный алгоритм разрабатывается на основе последовательного, с целью ускорения выполнения процесса умножения матриц больших размеров.
4. Написать параллельную реализацию предложенной последовательной программы. В случае, когда распараллеливание возможно по нескольким индексам, выбрать наиболее эффективный вариант. Параллельная и последовательная реализации должны генерировать одинаковый выходной файл.

СЕВЕРО-КАВКАЗСКАЯ ГОСУДАРСТВЕННАЯ АКАДЕМИЯ

Кафедра Прикладной информатики

20 - 20 учебный год

Экзаменационный билет № 1

по дисциплине Анализ данных и машинное обучение
для обучающихся направления подготовки 09.03.04 Программная инженерия

1. Условный оператор, инструкция if. Множественное ветвление.
2. Типы задач машинного обучения. Способы обучения и оценки его качества.
3. Дан двумерный массив X , состоящий из $n \times m$ элементов. Найти максимальный элемент двумерного массива .

Зав. кафедрой

Хапаева Л.Х.

Задания на лабораторные работы
По дисциплине «Анализ данных и машинное обучение»

Вывести на экран в графическом режиме графики двух функций на интервале от $x_{нач}$ до $x_{кон}$ с шагом dx . Первая функция задана с помощью ряда Тейлора, ее вычисление должно выполняться с точностью ϵ . Значение параметра b для второй функции вводится с клавиатуры. Графики должны быть плавными и различаться цветами.

Вариант 1
$$= 2 \cdot \left(1 + \frac{1}{x} + \frac{1}{3 \cdot x^3} + \frac{1}{5 \cdot x^5} + \dots \right),$$

$$y(x) = 2 \cdot \sum_{n=0}^{\infty} \frac{1}{(2 \cdot n + 1) \cdot x^{2 \cdot n + 1}} \quad \left(\frac{1}{x} + \frac{1}{3 \cdot x^3} + \frac{1}{5 \cdot x^5} + \dots \right) \quad \left. \begin{array}{l} x > 1; \\ || \\ || \end{array} \right\}$$

$$z(x) = \ln \frac{x+1}{x-1} + b.$$

Вариант 2

$$y(x) = \sum_{n=0}^{\infty} \frac{(-1)^n \cdot x^n}{n!} = \left(1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \frac{x^4}{4!} - \dots \right), \quad |x| < \infty;$$

$$z(x) = e^{-x} + b;$$

Вариант 3

$$y(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} = \left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots \right), \quad |x| < \infty$$

$$z(x) = e^x + b.$$

Вариант 4

$$y(x) = \frac{1}{2} + \sum_{n=0}^{\infty} \frac{(-1)^{n+1}}{(2 \cdot n + 1) \cdot x^{2 \cdot n + 1}} = \frac{1}{2} - \frac{1}{x} + \frac{1}{3 \cdot x^3} - \frac{1}{5 \cdot x^5} + \dots, \quad x > 1;$$

$$z(x) = \arctg x + b.$$

Вариант 5

$$y(x) = \sum_{n=0}^{\infty} \frac{(-1)^n \cdot x^{2 \cdot n + 1}}{(2 \cdot n + 1)} = \frac{x^3}{3} - \frac{x^5}{5} + \frac{x^7}{7} - \dots, \quad |x| \leq 1$$

$$z(x) = \arctg x + b.$$

Вариант 6
$$= 2 \cdot \left(1 + \frac{1}{x} + \frac{1}{3 \cdot x^3} + \frac{1}{5 \cdot x^5} + \dots \right)$$

$$y(x) = 2 \cdot \sum_{n=0}^{\infty} \frac{1}{(2 \cdot n + 1) \cdot x^{2 \cdot n + 1}} \quad \left(\frac{1}{x} + \frac{1}{3 \cdot x^3} + \frac{1}{5 \cdot x^5} + \dots \right) \quad \left. \begin{array}{l} x > 1; \\ || \\ || \end{array} \right\}$$

$$z(x) = \operatorname{Arth} x + b.$$

Вариант 7

$$y(x) = -\frac{1}{2} + \sum_{n=0}^{\infty} \frac{(-1)^{n+1}}{(2 \cdot n + 1) \cdot x^{2 \cdot n + 1}} = -\frac{1}{2} - \frac{1}{x} + \frac{1}{3 \cdot x^3} - \frac{1}{5 \cdot x^5} + \dots, \quad x < -1$$

$$z(x) = \arctg x + b.$$

Вариант 8

$$y(x) = \sum_{n=0}^{\infty} \frac{(-1)^n \cdot x^{2n}}{n!} = 1 - x^2 + \frac{x^4}{2!} - \frac{x^6}{3!} + \frac{x^8}{4!} - \dots, \quad |x| < \infty$$

$$z(x) = e^{-x^2} + b.$$

Вариант 9

Написать программу, которая выводит на экран секторную диаграмму. Диаграмму снабдить заголовком и наименованием для каждого сектора. Исходные данные сформировать в текстовом файле. Количество секторов задавать в программе в виде именованной константы.

Построение секторной диаграммы оформить в виде процедуры. Параметры процедуры: координаты центра диаграммы; радиус; количество секторов; массив процентов; массив наименований. Пример исходных данных см. Таблица 1.

Вариант 10

Написать программу, которая выводит на экран две секторные диаграммы, расположив их рядом. Диаграмму снабдить заголовком и наименованием для каждого сектора. Исходные данные сформировать в текстовом файле. Количество секторов задавать в программе в виде именованной константы.

Построение секторной диаграммы оформить в виде процедуры. Параметры процедуры: координаты центра диаграммы; радиус; количество секторов; массив процентов; массив наименований. Пример исходных данных см. Таблица 1.

Вариант 11

Написать программу, которая выводит на экран две столбиковые диаграммы. На экране диаграммы расположить рядом, каждую в своих координатных осях. Каждую диаграмму снабдить заголовком и наименованием единиц измерений по осям X и Y. Исходные данные сформировать в текстовом файле. Количество столбцов задавать в программе в виде именованной константы. Построение диаграммы оформить в виде процедуры. Пример исходных данных см. Таблица 1.

Вариант 12

Написать программу, которая выводит на экран две столбиковые диаграммы водной координатной плоскости. Диаграмму снабдить градацией осей и заголовком. Исходные данные сформировать в текстовом файле. Количество столбцов задавать в программе в виде именованной константы. Построение диаграммы оформить в виде процедуры. Пример исходных данных см. Таблица 1.

Вариант 13

Написать программу, которая выводит на экран трехмерную столбиковую диаграмму. Диаграмму снабдить градацией осей и заголовком. Исходные данные сформировать в текстовом файле. Количество столбцов задавать в программе в виде именованной константы. Построение диаграммы оформить в виде процедуры.

Вариант 14

Написать программу, которая выводит на экран столбиковую диаграмму, представляющую оптовые и розничные цены на различные наименования кофе. Исходные данные сформировать в текстовом файле.

Построение диаграммы оформить в виде процедуры. Параметры процедуры: количество наименований; массив значений оптовых цен; массив значений розничных цен; массив наименований. Наименования товаров разместить вертикально под осью абсцисс.

Вариант 15

Написать программу, которая выводит на экран столбиковую диаграмму, представляющую максимальную и среднюю норму прибыли при реализации различных сортов шоколада. Исходные данные сформировать в текстовом файле самостоятельно.

Построение диаграммы оформить в виде процедуры. Параметры процедуры: количество наименований; массив значений оптовых цен; массив значений розничных цен; массив наименований. Наименования товаров разместить вертикально под осью абсцисс.

Вариант 16

Написать программу, которая выводит на экран графики динамики изменения максимального, минимального и среднего курса доллара за заданное количество дней. Исходные данные сформировать в текстовом файле самостоятельно.

Построение графика оформить в виде процедуры. Параметры процедуры: массив дат; количество дней; массивы максимальных, минимальных и средних значений.

Вариант 17

Написать программу, которая выводит на экран трехмерную столбиковую диаграмму курса немецкой марки по отношению к рублю за заданное количество дней. Исходные данные сформировать в текстовом файле самостоятельно.

Построение диаграммы оформить в виде процедуры. Параметры процедуры: массив дат; количество дней; массив значений по оси Y; код заполнителя.

Индивидуальные задания

Работа с файлом данных «Титаника»

Приведенные ниже задания основаны на данных 'titanic.csv', где содержатся сведения о пассажирах «Титаника».

Задачи:

1. Какое количество мужчин и женщин плыло на корабле?
2. Какой части пассажиров удалось выжить? Посчитайте долю выживших пассажиров. Ответ приведите в процентах (число в интервале от 0 до 100, знак процента не нужен), округлив до двух знаков.
3. Какую долю пассажиры первого класса составляли среди всех пассажиров? Ответ приведите в процентах (число в интервале от 0 до 100, знак процента не нужен), округлив до двух знаков.
4. Какого возраста были пассажиры? Посчитайте среднее и медиану возраста пассажиров. В качестве ответа приведите два числа через пробел.
5. Коррелируют ли число братьев/сестер/супругов с числом родителей/детей? Посчитайте корреляцию Пирсона между признаками SibSp и Parch.

Тестовые вопросы
по дисциплине: «Анализ данных и машинное обучение»

1. Укажите правильное определение понятия «Пространство признаков» (ПК-1)
 - N-мерное пространство, где N – число измеряемых характеристик объектов, выделенное для конкретной задачи;
 - добавление некоторой дополнительной информации к условию минимизации ошибки
 - свойство натренированного алгоритма на объектах тренировочной выборки давать существенно меньшую вероятность ошибки, чем на объектах тестовой
 - совокупность объектов, принадлежащих ему с определенной степенью.

2. Впишите правильный ответ. Файлы на языке Python в среде Windows имеют расширение _____(ПК-2)

3. Впишите правильный ответ(два слова). « Отличительной особенностью методов _____ является то, что они способны решать широкий спектр задач без явного кодирования алгоритма решения».(ПК-1)

4. Что отражает данное изображение, показанное на рисунке 1 (ПК-2)

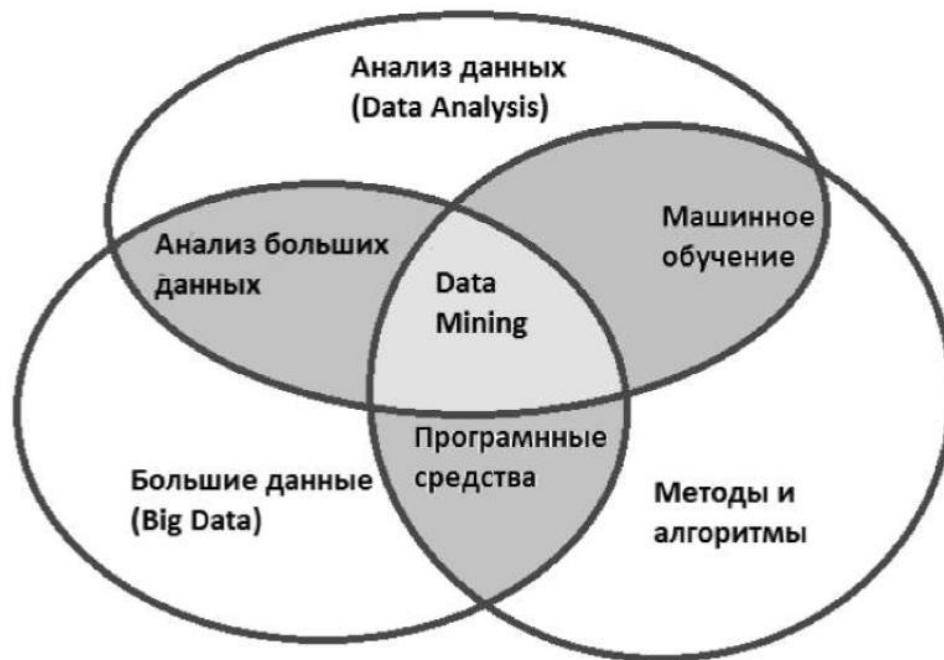


Рисунок 1.

- Области науки о данных
 - Этапы машинного обучения
 - Этапы анализа данных
 - Области обработки больших данных.
5. Укажите особенности асинхронной модели параллельных вычислений (ПК-2)
 - все процессы выполняют одни и те же действия с собственными данными
 - различные процессы решают разные задачи
 - все процессы используют общую память
 - все процессы выполняются в своих критических секциях

6. Укажите особенности синхронной модели параллельных вычислений (ПК-2)

- все процессы выполняют одни и те же действия с собственными данными
- различные процессы решают разные задачи
- все процессы используют общую память
- все процессы выполняются в своих критических секциях

7. Выберите операции, которые могут выполняться параллельно (ПК-2)

- независимые
- зависимые
- элементарные
- неделимые

8. Впишите правильный ответ. _____ имеет собственную память для данных (ПК-1)

9. Что из перечисленного характеризует язык программирования PYTHON (ПК-2)

- машинно - ориентированный язык (низкого уровня)
- язык высокого уровня
- объектно - ориентированный язык

10. Укажите типы переменных, которые используются в языке программирования PYTHON: (ПК-2)

- float
- list
- num
- int
- bool
- integer

11. Из перечисленного ниже выберите состояние, которым характеризуется состояние параллельной программы (ПК-1)

- адресами выполняемых команд
- последовательностью состояний $s_0 \rightarrow s_1 \rightarrow \dots \rightarrow s_n$.
- значениями переменных в некоторый момент времени
- объемом занимаемой оперативной памяти

12. Выберите вариант ответа, который характеризует MPI – системы (ПК-1)

- модуль параллельной обработки в системе
- специальная ОС для параллельного программирования
- интерфейс, содержащий набор функций, типов и констант для параллельного программирования
- организация, координирующая разработку параллельных интерфейсов

13. Укажите области параллельного программирования, для которых наиболее удобна технология MPI: (ПК-2)

- программирование для систем с общей памятью
- программирование для кластеров
- научные вычисления
- программирование клиент-серверных приложений

14. Впишите правильный ответ. Операция _____ является наиболее популярной моделью параллельных вычислений (ПК-1)

15. Укажите операции в модели параллельных вычислений, которые могут выполняться параллельно (ПК-2)

- смежные в графе «операции-операнды»
- связанные путем в графе «операции-операнды»
- не связанные путем в графе «операции-операнды»
- не смежные в графе «операции-операнды»

16. Выберите способы распределения данных, которые используются при разработке матричных параллельных алгоритмов (выберите несколько вариантов ответов) (ПК-1)

- ленточное разбиение
- блочное разбиение
- диагональное разбиение
- циклическое разбиение

17. Впишите правильный ответ. Оператор _____ определяет функцию в языке программирования (ПК-2)

18. Выберите вариант ответа характеризующий многопроцессность (concurrency) в контексте параллельных вычислений (выберите несколько вариантов ответов) (ПК-2)

- обеспечение минимального времени выполнения одной программы
- первичность пропускной способности
- не требуется обеспечение максимальной изоляции процессов друг от друга
- обеспечение как можно более равномерного распределения ресурсов между процессами

19. Впишите правильный ответ. База данных _____ работает на основе стека Hadoop. (ПК-1)

20. Впишите правильный ответ. Директива _____ маркирует начало параллельной секции (OpenMP, язык C) (ПК-1)

21. Впишите правильный ответ. Функция _____ возвращает количество процессов в группе в MPI (язык C) (ПК-1)

22. Укажите верное утверждение характеризующее СУБД (ПК-1)

- реляционные СУБД изначально ориентированы на работу с большими объемами данных в распределенном режиме
- в реляционных СУБД для обеспечения отказоустойчивости применяется репликация данных
- по мере усложнения структуры реляционных СУБД производительность сложных запросов возрастает нелинейно
- в реляционных СУБД для обеспечения отказоустойчивости применяется статическое разбиение данных

23. Впишите правильный ответ. Стандарт _____ параллельного программирования предназначен для использования в системах с распределенной памятью. (ПК-1)

24. Выберите правильный результат выполнения параллельной секции следующей параллельной программы (язык C, OpenMP, больше 1 параллельного потока): (ПК-2)

<pre>#include <omp.h> #include <stdio.h> int main(int argc, char **argv) { #pragma omp parallel printf("Hi!\n"); printf("Bye!\n"); return 0; }</pre>	<ul style="list-style-type: none"> - Hi! ... Hi! Bye! ... Bye! - Hi! ... Hi! Bye! - Hi! Bye! Hi! ... Bye! - Hi! Hi! Bye! Bye! Hi! ... Bye!
--	--

25. Впишите правильное определение _____ характеризуется однородным доступом к памяти всех процессорных устройств. (ПК-1)

26. Выберите верную характеристику ускорения параллельного алгоритма (ПК-1)

- минимальное время выполнения последовательного алгоритма
- отношение количества процессоров к количеству потоков исполнения
- минимальное время выполнения параллельного алгоритма
- размер входных данных

27. Укажите способы использования распределенных вычислительных систем (ПК-1):

- могут быть использованы для параллельных вычислений только для программ с низкой интенсивностью потоков межпроцессорных передач данных
- не могут быть использованы для организации параллельных вычислений
- ориентированы на проведение параллельных вычислений

28. Выберите вариант ответа, характеризующий производительность многопроцессорной вычислительной системы (ПК-2)

- количеством операций, производимых за единицу времени
- количеством байт информации, переданных в единицу времени
- числом импульсов, генерируемых в единицу времени
- объемом располагаемой для вычислений памяти

29. Укажите к какому классу относятся многоядерные системы (ПК-2)

- матричные процессоры
- распределенные системы
- кластеры
- системы с общей памятью

30. Укажите к какому классу относятся кластерные системы (ПК-2)

- матричные процессоры
- распределенные системы
- симметричные мультипроцессоры
- системы с общей памятью

5. Методические материалы, определяющие процедуры оценивания компетенции

5.1 Критерии оценивания качества выполнения лабораторных работ

Оценка **«зачтено»** выставляется обучающемуся, если лабораторная работа выполнена правильно и студент ответил на все вопросы, поставленные преподавателем на защите.

Оценка **«не зачтено»** выставляется обучающемуся, если лабораторная работа выполнена не правильно или студент не проявил глубоких теоретических знаний при защите работы.

5.2 Критерии оценивания тестирования

При тестировании все верные ответы берутся за 100%.

90%-100% отлично

75%-90% хорошо

60%-75% удовлетворительно

менее 60% неудовлетворительно

5.3 Критерии оценивания качества индивидуального задания

Оценка **«отлично»** выставляется за глубокое знание предусмотренного программой материала, за умение четко, лаконично и логически последовательно отвечать на поставленные вопросы.

Оценка **«хорошо»** – за твердое знание основного (программного) материала, за грамотные, без существенных неточностей ответы на поставленные вопросы.

Оценка **«удовлетворительно»** – за общее знание только основного материала, за ответы, содержащие неточности или слабо аргументированные, с нарушением последовательности изложения материала.

Оценка **«неудовлетворительно»** – за незнание значительной части программного материала, за существенные ошибки в ответах на вопросы, за неумение ориентироваться в материале, за незнание основных понятий дисциплины.

5.4 Критерии оценивания ответа на зачете

Оценка **«зачтено»** выставляется обучающемуся за общее знание основного материала, включая расчеты (при необходимости), за грамотные, без существенных неточностей ответы на поставленные вопросы, за умение применять теоретические положения для решения практических задач.

Оценка **«не зачтено»** выставляется обучающемуся за незнание значительной части программного материала, за существенные ошибки в ответах на вопросы, за неумение ориентироваться в расчетах, за незнание основных понятий дисциплины.

5.5 Критерии оценивания результатов освоения дисциплины на экзамен

Оценка **«отлично»** выставляется за глубокое знание предусмотренного программой материала, содержащегося в основных и дополнительных рекомендованных литературных источниках, за умение четко, лаконично и логически последовательно отвечать на поставленные вопросы, за умение анализировать изучаемые явления в их взаимосвязи и диалектическом развитии, применять теоретические положения при решении практических задач.

Оценка **«хорошо»** – за твердое знание основного (программного) материала, включая расчеты (при необходимости), за грамотные, без существенных неточностей ответы на поставленные вопросы, за умение применять теоретические положения для решения практических задач.

Оценка **«удовлетворительно»** – за общее знание только основного материала, за ответы, содержащие неточности или слабо аргументированные, с нарушением последовательности изложения материала, за слабое применение теоретических положений при решении практических задач.

Оценка **«неудовлетворительно»** – за незнание значительной части программного материала, за существенные ошибки в ответах на вопросы, за неумение ориентироваться в расчетах, за незнание основных понятий дисциплины.