

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Уфимский государственный нефтяной технический университет»

УТВЕРЖДАЮ

Директор IT-институт

З.Х. Павлова

(подпись)

29.06.2023

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

(626)Статистический анализ данных

Направление подготовки (специальность): **09.03.01 Информатика и вычислительная техника**

Направленность: **профиль«Технологии искусственного интеллекта в нефтегазовой отрасли»**

Уровень высшего образования: **бакалавриат**

Форма обучения: **очная;**

Кафедра, обеспечивающая преподавание дисциплины: **Вычислительная техника и инженерная кибернетика (ВТИК);**

Трудоемкость дисциплины: **3 з.е. (108час)**

Уфа 2023

Рабочую программу дисциплины разработал(и):

_____ старший преподаватель Шварева Е.Н.

Рецензент

_____ доцент, кандидат ф.-м.н. Зарипов Д. М.

Рабочая программа дисциплины рассмотрена и одобрена на заседании кафедры Вычислительная техника и инженерная кибернетика (ВТИК);, обеспечивающей преподавание дисциплины 31.08.2022, протокол №1.

Заведующий кафедрой

Вычислительная техника и инженерная кибернетика (ВТИК)_____..

СОГЛАСОВАНО

Заведующий кафедрой ВТИК_____..

Год приема 2023 г.

Рабочая программа зарегистрирована 19.09.2022 № 1 в ОПЛА и внесена в электронную базу данных

1. Место дисциплины в структуре ОПОП ВО

Дисциплины, предшествующие изучению данной дисциплины (исходя из формирования этапов по компетенциям):

Дисциплины, для которых освоение данной дисциплины необходимо как предшествующее (исходя из формирования этапов по компетенциям): Базы данных; Геоинформационные системы; Интеллектуальные системы управления и автоматизации; Компьютерная графика; Объектно-ориентированное программирование; Операционные системы; Основы теории нейросетевого моделирования; Основы технологии блокчейн; Преддипломная практика; Технологическая (проектно-технологическая) практика

Блок: Блок 1. Дисциплины (модули);

Обязательная или часть, формируемая участниками образовательных отношений (в том числе элективные дисциплины): Часть, формируемая участниками образовательных отношений;

Форма обучения: очная

Семестр, в котором преподается дисциплина	Трудоемкость дисциплины				Вид промежу- точной аттестации
	Зачет- ные едини- цы	Часы			
		Общая	В том числе		
			контакт- ная	СРО	
4	3	108	50	58	экзамен;
ИТОГО:	3	108	50	58	

2. Компетенции обучающегося, формируемые в результате освоения дисциплины

№ пп.	Формируемые компетенции	Шифр/ индекс компетен- ции
1	Способен использовать инструментальные средства для решения задач машинного обучения	ПК-5и-22Г-1
2	Способен разрабатывать системы анализа больших данных	ПК-8и-22Г.- 1

В результате освоения дисциплины обучающийся должен:

Шифр компетенции	Индикаторы достижения компетенций	Шифр результата обучения	Результат обучения
ПК-5и-22Г	ПК-5.1. Осуществляет оценку и выбор инструментальных средств для решения поставленной задачи	З(ПК-5и-22Г)	Знать: знает модели машинного обучения

Шифр компетенции	Индикаторы достижения компетенций	Шифр результата обучения	Результат обучения
	ПК-5.2. Разрабатывает модели машинного обучения для решения задач ПК-5.3. Создает, поддерживает и использует системы искусственного интеллекта, включающие разработанные модели и методы, с применением выбранных инструментов машинного обучения	У(ПК-5и-22Г)	Уметь: умеет оценивать и выбирать средства для решения
		В(ПК-5и-22Г)	Владеть: владеет системами искусственного интеллекта
ПК-8и-22Г.	ПК-8.1. Разрабатывает программные компоненты извлечения, хранения, подготовки больших данных с учётом вариантов использования больших данных, определений, словарей и эталонной архитектуры больших данных ПК-8.2. Разрабатывает программные компоненты обработки удалённой, распределенной и объединённой аналитики, использования результатов анализа, описания и управления качеством и достоверностью больших данных	З(ПК-8и-22Г.)	Знать: знает программные компоненты работы с большими данными
		У(ПК-8и-22Г.)	Уметь: умеет разрабатывать программные компоненты работы с данными
		В(ПК-8и-22Г.)	Владеть: владеет разработкой программных компонентов работы с данными

3. Структура дисциплины

3.1. Виды учебной работы и трудоемкость (всего и по семестрам, в часах)

Форма обучения: очная

Вид учебной работы	Всего и по семестрам, часы	1	2	3	4	5	6	7	8	9	10	11	12
Контактная работа, всего в том числе:	50				50								
лекции (всего)	16				16								
-в т.ч. лекции on-line курс	0												
практические занятия (ПЗ)	4				4								
-в т.ч. практические занятия on-line курс	0												
лабораторные работы (ЛР)	24				24								
контролируемая самостоятельная работа (защита курсового проекта, курсовой работы и др. работ (при наличии))	0												
-в т.ч. лабораторные работы	0												

ты on-line курс													
иная контактная работа (сдача зачета, экзамена, консультации)	6				6								
проектная деятельность (ПД)	0												
Самостоятельная работа обучающихся (СРО), всего в том числе: (указать кон- кретный вид СРО)	58				58								
выполнение и подготовка к защите курсового проек- та или курсовой работы	0												
выполнение и подготовка к защите РГР работы, ре- ферата, патентных иссле- дований, аналитических исследований и т.п	0												
изучение учебного мате- риала, вынесенного на самостоятельную прора- ботку	7				7								
подготовка к лаборатор- ным и/или практическим занятиям	28				28								
подготовка к сдаче зачета, экзамена	23				23								
иные виды работ обучаю- щегося (при наличии)	0												
освоение on-line курса	0												
самостоятельная проект- ная деятельность (СПД)	0												
ИТОГО ПО ДИСЦИПЛИ- НЕ	108				108								

4. Содержание дисциплины

4.1. Темы (разделы) дисциплины и виды занятий (в часах)

Форма обучения: очная

(раздела)Номер темы	Назва- ние темы (разде- ла)	Семестр	Трудоемкость, часы					Шифр ре- зульта- та обу- чения
			Л	ПЗ	ЛР	СРО	Всего	
1	Инстру- менты для ана- лиза данных	4	6	2	10	22	40	З(ПК- 8и-22Г.) У(ПК- 8и-22Г.) В(ПК- 8и-22Г.)
2	Задачи предска- заний	4	10	2	14	36	62	З(ПК- 5и-22Г) У(ПК-

Номер темы	Назва- ние	Семес- тр	Трудоемкость, часы					Шифр ре-
			Л	ПЗ	ЛР	СРО	Всего	
	машин- ного обуче- ния							5и-22Г) В(ПК- 5и-22Г)
	ИТОГО:		16	4	24	58	102	

4.2. Содержание лекционного курса

№ пп.	Номер раздела	Название темы	Трудоемкость, часы		
			очная	очно- заочная	заочная
1	1-Инструменты для анализа данных	Предварительная обработка данных Стандартизация, кодирование категориальных признаков, Дискретизация, Преобразование входных данных для применения алгоритмов машинного обучения	2		
2	1-Инструменты для анализа данных	Понижение размерности Понижение числа случайных переменных для анализа. Визуализация, Увеличение эффективности . Анализ главных компонентов (РСА) , выбор признаков, Факторный анализ, Анализ независимых компонентов (ICA),	2		
3	1-Инструменты для анализа данных	Выбор и оценка модели Сравнение, проверка и выбор параметров и моделей.Повышение точности за счет настройки параметров Перекрестная проверка. Количественная оценка	2		

		качества прогно- зов.			
4	2-Задачи предска- ний машинного обу- чения	Регрессия Прогнозирование атрибута с непре- рывным значени- ем, связанного с объектом.. Ли- нейные модели Метод наимень- ших квадратов. Регрессия и клас- сификация. Лассо. Байе- совская регрес- сия. Логистиче- ская регрессия Обобщенная ли- нейная регрессия	4		
5	2-Задачи предска- ний машинного обу- чения	Классификация Кластеризация Классификация Определение, к какой категории относится объект. Мульти- классовая клас- сификация. SVC, Ближайшие сосе- ди, случайный лес Кластериза- ция Автоматиче- ская группировка похожих объек- тов в наборы. k- Means , спек- тральная класте- ризация , сред- ний сдвиг	4		
6	2-Задачи предска- ний машинного обу- чения	Ансамбли, ги- бридизация Ансамбли алго- ритмов, эвристи- ческие алгорит- мы, гибридиза- ция алгоритмов и метаоптимизация Леса рандомизи- рованных дере- вьев AdaBoost	2		
	-	ИТОГО:	16		

4.3. Перечень лабораторных работ

Номер раздела	№ ЛР	Трудоемкость, часы
---------------	------	-----------------------

			очная	очно- заочная	заочная
1-Инструменты для анализа данных	1	Предварительная обработка данных Стандартизация, кодирование категориальных признаков, Дискретизация, Преобразование входных данных для применения алгоритмов машинного обучения Датасеты с UCI Machine Learning Repository https://archive.ics.uci.edu/ml/index.php Работа с библиотеками Pandas, Scikit-Learn, TensorFlow в облачном сервисе Google Colab	2		
1-Инструменты для анализа данных	2	Понижение размерности Понижение числа случайных переменных для анализа. Визуализация, Увеличение эффективности . Анализ главных компонент (PCA) , выбор признаков, Факторный анализ, Анализ независимых компонент (ICA), Датасеты с UCI Machine Learning Repository https://archive.ics.uci.edu/ml/index.php Работа с библиотеками	4		

		Pandas, Scikit-Learn, TensorFlow в облачном сервисе Google Colab			
1-Инструменты для анализа данных	3	Выбор и оценка модели Сравнение, проверка и выбор параметров и моделей. Повышение точности за счет настройки параметров Перекрестная проверка. Количественная оценка качества прогнозов. Датасеты с UCI Machine Learning Repository https://archive.ics.uci.edu/ml/index.php Работа с библиотеками Pandas, Scikit-Learn, TensorFlow в облачном сервисе Google Colab	4		
2-Задачи предсказаний машинного обучения	4	Регрессия Прогнозирование атрибута с непрерывным значением, связанного с объектом.. Линейные модели Метод наименьших квадратов. Регрессия и классификация. Лассо. Байесовская регрессия. Логистическая регрессия Обобщенная линейная регрессия Датасеты с	4		

		UCI Machine Learning Repository https://archive.ics.uci.edu/ml/index.php Работа с библиотеками Pandas, Scikit-Learn, TensorFlow в облачном сервисе Google Colab			
2-Задачи предсказаний машинного обучения	5	Анализ номинальных данных, классификация Анализ номинальных данных, классификация, логистическая регрессия, кластерный анализ, метод главных компонент, метод ближайших соседей Датасеты с UCI Machine Learning Repository https://archive.ics.uci.edu/ml/index.php Работа с библиотеками Pandas, Scikit-Learn, TensorFlow в облачном сервисе Google Colab	4		
2-Задачи предсказаний машинного обучения	6	Классификация Кластеризация Классификация Определение, к какой категории относится объект. Мультиклассовая классификация. SVC, Ближайшие соседи, случайный лес Кла-	2		

		стеризация Автоматическая группировка похожих объектов в наборы. k-Means , спектральная кластеризация , средний сдвиг Датасеты с UCI Machine Learning Repository https://archive.ics.uci.edu/ml/index.php Работа с библиотеками Pandas, Scikit-Learn, TensorFlow в облачном сервисе Google Colab			
2-Задачи предсказаний машинного обучения	7	Ансамбли алгоритмов, эвристические алгоритмы, гибридизация алгоритмов и метаоптимизация Ансамбли алгоритмов, эвристические алгоритмы, гибридизация алгоритмов и метаоптимизация Леса рандомизированных деревьев AdaBoost Датасеты с UCI Machine Learning Repository https://archive.ics.uci.edu/ml/index.php Работа с библиотеками Pandas, Scikit-Learn, TensorFlow в облачном сервисе Google Colab	4		

-		ИТОГО:	24		
---	--	---------------	----	--	--

4.4. Перечень практических занятий

Номер раздела	№ ПЗ	Тема практического занятия	Трудоемкость, часы		
			очная	очно-заочная	заочная
1-Инструменты для анализа данных	1	Предварительная обработка Предварительная обработка Извлечение признаков и нормализация. Снижение размерности Уменьшение количества случайных переменных для рассмотрения. Выбор модели Сравнение, проверка и выбор параметров и моделей.	2		
2-Задачи предсказаний машинного обучения	2	Корреляционно-регрессионный анализ Корреляционно-регрессионный анализ Прогнозирование атрибута с непрерывным значением, связанного с объектом.	2		
-		ИТОГО:	4		

4.5. Виды СРО

Номер раздела	Вид СРО	Трудоемкость, часы		
		очная	очно-заочная	заочная
1-Инструменты для анализа данных	подготовка к сдаче зачета, экзамена	11		

1-Инструменты для анализа данных	подготовка к лабораторным и/или практическим занятиям	8		
1-Инструменты для анализа данных	изучение учебного материала, вынесенного на самостоятельную проработку	3		
2-Задачи предсказаний машинного обучения	подготовка к сдаче зачета, экзамена	12		
2-Задачи предсказаний машинного обучения	подготовка к лабораторным и/или практическим занятиям	20		
2-Задачи предсказаний машинного обучения	изучение учебного материала, вынесенного на самостоятельную проработку	4		
-	ИТОГО:	58		

Темы для самостоятельной работы обучающихся

Раздел 1. Инструменты для анализа данных

библиотеки для статистического анализа в Python и в языке R

Раздел 2. Задачи предсказаний машинного обучения

Ансамбли алгоритмов машинного обучения и эвристические алгоритмы

5. Формы текущего контроля успеваемости и проведения промежуточной аттестации

Перечень оценочных средств текущего контроля и промежуточной аттестации по дисциплине приведен Фонде оценочных средств (приложение Б).

6. Учебно-методическое и информационное обеспечение дисциплины

6.1. Учебно-методическое обеспечение

Сведения об обеспеченности дисциплины основной, дополнительной и учебно-методической литературой приведены в формах № 1-УЛ и № 2-УЛ (приложение А).

6.2. Перечень современных профессиональных баз данных и информационных справочных систем, рекомендуемых для освоения дисциплины

Названия современных профессиональных баз данных и информационных справочных систем, рекомендуемых для освоения дисциплины	Ссылки на официальные сайты
http://biblioclub.ru/	Университетская библиотека

	онлайн
https://e.lanbook.com/book/	Электронно-библиотечная система Лань
https://www.python.org	Python. The official Python web site. В свободном доступе.
https://znanium.com/	Электронно-библиотечная система
http://znanium.com/	ЭБС Znanium.com
Система дистанционного обучения УГНТУ, учебный курс преподавателя	http://do.rusoil.net

7. Материально-техническое обеспечение дисциплины

7.1. Перечень специальных аудиторий, кабинетов, лабораторий и пр., используемых при реализации дисциплины с перечнем основного оборудования

№ пп.	Номер помещения	Оснащенность помещения (перечень основного оборудования)	Наименование помещения
1	1-431	Винчестер Toshiba Canvio Basics <HDTB305EK3AA> USB 3.0 2.5" 500 Gb(1);Компьютер Nettop Pegatron Walle L6 PV D-SUB(2);Компьютер Pegatron Nettop MiniPC Wall-e L6(1);Компьютер Pegatron Nettop MiniPC Wall-e L6 Pinetrail Atom D510(2);Монитор Philips 27" 273V5LHAB\00(3);Монитор Philips тип5(2);Настенный экран Master Picture 244x244 MW(1);Проектор Acer ProjectorP1203(1);Проектор Epson EB-X49(2);Проектор мультимедийный Sony VPL-CX150(1);Рециркулятор бактерицидный ИБР01(1);Системный блок Powerman 450W\GA-F2A68HM-D3H(1);Экран на треноге Lumien Eco View<LEV-100102> 180*180 см(1);Шкаф(ы) для хранения	Помещения для хранения и профилактического обслуживания учебного оборудования
2	1-432	Компьютер Nettop Pegatron Walle L6 PV D-SUB(15);Компьютер тип K2 i3-3220/21,5" LG 22EA63T-P(1);Проектор мультимедийный Sony VPL-CX150(1);Системный блок UNIVERSAL D1(18);Стол, стулья	Лаборатория – оснащенная лабораторным оборудованием, в зависимости от степени сложности.
3	1-435	Компьютер Pegatron Nettop MiniPC Wall-e L6(12);Монитор Samsung S-LC24F390FHIXCI(9);Монитор Samsung S-LC24FG73FQIXCI(5);Проектор Optoma EH334(1);Системный блок UNIVERSAL D1(14);Стол, стулья	Лаборатория – оснащенная лабораторным оборудованием, в зависимости от степени сложности.
4	1-435	Компьютер Pegatron Nettop MiniPC Wall-e L6(12);Монитор Samsung S-LC24F390FHIXCI(9);Монитор Samsung S-LC24FG73FQIXCI(5);Проектор Optoma EH334(1);Системный блок UNIVERSAL D1(14);Стол, стулья	Учебная аудитория для проведения групповых и индивидуальных консультаций
5	1-441	Компьютер Pegatron Nettop MiniPC Wall-e L6 Pinetrail Atom D510(13);Доступ к электронной информационно-образовательной среде (Корпоративная информационная система УГНТУ); Доступ в интернет;	Помещение для самостоятельной работы – укомплектовано специализированной (учебной) мебелью, оснащено компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечено доступом в электронную информационно-образовательную среду организации.
6	1-444	Компьютер Nettop Pegatron Walle L6 PV D-SUB(1);Настенный экран Master Picture 244x244 MW(1);Проектор Acer ProjectorP1203(1);мультимедиа-проектор;Доступ к электронной ин-	Помещение для самостоятельной работы – укомплектовано специализированной (учебной) мебелью, оснащено компьютерной техникой с возмож-

7.2. Перечень лицензионного и свободно распространяемого программного обеспечения, используемых в учебном процессе при освоении дисциплины

№ пп.	Наименование ПО	Лицензионная чистота (реквизиты лицензии, свидетельства о гос. регистрации и т.п., срок действия)
1	Python	Дата выдачи лицензии 01.01.2006, Поставщик: Свободное программное обеспечение
2	Python	Дата выдачи лицензии 01.01.1991, Поставщик: Свободное программное обеспечение
3	Python v 3.7.1	Дата выдачи лицензии 01.01.2000, Поставщик: Свободное программное обеспечение
4	Python 3	Дата выдачи лицензии 01.01.2006, Поставщик: Свободное программное обеспечение
5	Pyton OnLine	Дата выдачи лицензии 01.01.2006, Поставщик: Свободное программное обеспечение
6	R открытая программная среда для статистических вычислений и графики	Дата выдачи лицензии 01.01.2006, Поставщик: Свободное программное обеспечение

8. Организация обучения лиц с ограниченными возможностями здоровья

Для лиц с ограниченными возможностями здоровья, обучающихся по данной образовательной программе, разрабатывается индивидуальная программа освоения дисциплины с учетом особенностей их психофизического развития, их индивидуальных возможностей и состояния здоровья.

Приложение А

Форма № УЛ-1

СВЕДЕНИЯ

об обеспеченности дисциплины основной и дополнительной учебной литературой

Наименование дисциплины: (626)(626)Статистический анализ данныхНаправление подготовки (специальность): 09.03.01 Информатика и вычислительная техникаНаправленность: профиль«Технологии искусственного интеллекта в нефтегазовой отрасли»Форма обучения: очная;Кафедра, обеспечивающая преподавание дисциплины: Вычислительная техника и инженерная кибернетика (ВТИК);

Тип	Назначение учебных изданий	Семестр			Библиографическое описание	Кол-во экз.	Адрес ания нахождения электронного	остиКоэффициент
		очная	очно-заочная	заочная				
1	2	3	4	5	6	7	8	9
Основная литература	Для выполнения СРО;Для выполнения практических занятий;Для изучения теории;	4			Гмурман, В. Е. Теория вероятностей и математическая статистика : учебное пособие / В. Е. Гмурман. - 12-е изд. - М. : Юрайт, 2013. - 479 с. on-line. - URL: http://bibl.rusoil.net/base_docs/UGNTU/CD/Gmurman1.pdf . - Текст : электронный.	1	http://bibl.rusoil.net	1.00

Основная литература	Для выполнения СРО; Для выполнения лабораторных работ; Для выполнения практических занятий;	4			Гмурман, В. Е. Руководство к решению задач по теории вероятностей и математической статистике : учебное пособие / В. Е. Гмурман. - 11-е изд., перераб. и доп. - М. : Юрайт, 2013. - 404 с. on-line. - URL: http://bibl.rusoil.net/base_docs/UGNTU/CD/Gmurman.pdf . - Текст : электронный.	1	http://bibl.rusoil.net	1.00
Дополнительная литература	Для выполнения СРО; Для выполнения лабораторных работ; Для выполнения практических занятий; Для изучения теории;	4			Гусаров, В. М. Статистика : учебное пособие : [16+] / В. М. Гусаров, Е. И. Кузнецова. – 2-е изд., перераб. и доп. – Москва : Юнити, 2012. – 480 с. –URL: https://biblioclub.ru/index.php?page=book&id=117190 (дата обращения: 22.04.2022). –Текст : электронный.	1	https://biblioclub.ru	1.00
Примечание – Графы 1-5,8 заполняются кафедрой, графы 7 и 9 - библиотекой								

Составил:

_____ старший преподаватель Шварева Е.Н.

Год приема 2023 г.

СВЕДЕНИЯ**об обеспеченности дисциплины учебно-методическими изданиями**

Наименование дисциплины: (626)(626)Статистический анализ данных

Направление подготовки (специальность): 09.03.01 Информатика и вычислительная техника

Направленность профиль«Технологии искусственного интеллекта в нефтегазовой отрасли»

Форма обучения очная;

Кафедра, обеспечивающая преподавание дисциплины: Вычислительная техника и инженерная кибернетика (ВТИК);

Назначение учебных изданий	Семестр			Библиографическое описание	Кол-во экз.		Адрес нахождения электронного	Коэффициент
	очная	очно-заочная	заочная		Всего	в том числе		
1	2	3	4	5	6	7	8	9

Для выполнения СРО;Для выполнения лабораторных работ;	4			Учебно-методический комплекс дисциплины "Математика" : учебно-методический комплекс / Р. Н. Бахтин [и др.] ; УГНТУ, ИАУ, каф. Математики. - Уфа : Изд-во УГНТУ. - URL: http://bibl.rusoil.net/base_docs/UGNTU\Matematika/14UMK2010.PDF . - Текст : электронный. Раздел 14 : Математическая статистика : теоретические основы; метод. указания для студентов; материалы для самостоятельной работы студентов. - 2010.	1	0	http://bibl.rusoil.net	1.00
Для выполнения СРО;Для выполнения лабораторных работ;Для выполнения практических занятий;	4			Статистика на языке программирования R : учебно-методическое пособие для лабораторных, практических и самостоятельных работ / УГНТУ, каф. Математики ; сост.: С. Н. Коледин [и др.]. - Уфа : УГНТУ, 2018. - 1,24 Мб. - URL: http://bibl.rusoil.net/base_docs/UGNTU/Matematika/Koledin5.pdf . - Текст : электронный.	1	0	http://bibl.rusoil.net	1.00
Примечание – Графы 1-5,8 заполняются кафедрой, графы 6,7 и 9 - библиотекой								

Составил:

старший преподаватель Шварева Е.Н.

Год приема 2023 г.

Приложение Б

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Уфимский государственный нефтяной технический университет»

УТВЕРЖДАЮ

Директор IT-институт

_____ 3.Х. Павлова
29.06.2023

**Фонд оценочных средств по текущей успеваемости и промежуточной аттестации по дисциплине
(626)Статистический анализ данных**

Направление подготовки (специальность): 09.03.01 Информатика и вычислительная техника

Направленность: профиль «Технологии искусственного интеллекта в нефтегазовой отрасли»

Уровень высшего образования: бакалавриат

Форма обучения: очная;

Кафедра, обеспечивающая преподавание дисциплины: Вычислительная техника и инженерная кибернетика (ВТИК);

Трудоемкость дисциплины: 3 з.е. (108час)

Уфа 2023

ФОС по текущей успеваемости и промежуточной аттестации по дисциплине разработал (и):

_____ старший преподаватель Шварева Е.Н.

—
Рецензент

_____ доцент, кандидат ф.-м.н. Зарипов Д. М.

—

ФОС по текущей успеваемости и промежуточной аттестации по дисциплине рассмотрен и одобрен на заседании кафедры Вычислительная техника и инженерная кибернетика (ВТИК);, обеспечивающей преподавание дисциплины 31.08.2022, протокол №1.

Заведующий кафедрой Вычислительная техника и инженерная кибернетика (ВТИК);_____..

СОГЛАСОВАНО

Заведующий кафедрой ВТИК_____..

Год приема 2023 г.

ФОС по текущей успеваемости и промежуточной аттестации по дисциплине
зарегистрирован 19.09.2022 № 1 в отделе МСОП и внесен в электронную базу данных

1. Паспорт фонда оценочных средств

№ п/п	Контролируемые разделы (темы) дисциплины	Шифр результата обучения	Результат обучения	Индикатор достижения компетенций	Показатели достижения результатов освоения компетенций	Вид оценочного средства
1	Инструменты для анализа данных	В(ПК-8и-22Г.)	знает программные компоненты работы с большими данными	ПК-8.1. Разрабатывает программные компоненты извлечения, хранения, подготовки больших данных с учётом вариантов использования больших данных, определений, словарей и эталонной архитектуры больших данных	применяет алгоритмы разработки программных компонент работы с данными	Письменный и устный опрос
				ПК-8.2. Разрабатывает программные компоненты обработки удалённой, распределённой и объединённой аналитики, использования результатов анализа, описания и управления качеством и достоверностью больших данных	применяет методы машинного обучения	Письменный и устный опрос
		З(ПК-8и-22Г.)		ПК-8.1. Разрабатывает программные компоненты извлечения, хранения, подготовки больших данных с учётом вариантов использования больших данных, определений, словарей и эталонной архитектуры	описывает программные компоненты работы с большими данными	Тестирование

				больших данных		
				ПК-8.2. Разрабатывает программные компоненты обработки удалённой, распределенной и объединённой аналитики, использования результатов анализа, описания и управления качеством и достоверностью больших данных	описывает методы машинного обучения	Тестирование
				ПК-8.1. Разрабатывает программные компоненты извлечения, хранения, подготовки больших данных с учётом вариантов использования больших данных, определений, словарей и эталонной архитектуры больших данных	разрабатывает программные компоненты работы с данными	Лабораторная работа
				ПК-8.2. Разрабатывает программные компоненты обработки удалённой, распределенной и объединённой аналитики, использования результатов анализа, описания и управления качеством и достоверностью больших данных	демонстрирует методы машинного обучения	Лабораторная работа
7	Задачи предсказаний машинного обучения	В(ПК-5и-22Г)	знает модели машинного обучения	ПК-5.1. Осуществляет оценку и выбор инструментальных средств для решения поставленной	применяет алгоритмы машинного обучения	Письменный и устный опрос

			задачи		
			ПК-5.2. Разрабатывает модели машинного обучения для решения задач	применяет модели предсказания	Письменный и устный опрос
			ПК-5.3. Создает, поддерживает и использует системы искусственного интеллекта, включающие разработанные модели и методы, с применением выбранных инструментов машинного обучения	применяет статистические методы	Письменный и устный опрос
		З(ПК-5и-22Г)	ПК-5.1. Осуществляет оценку и выбор инструментальных средств для решения поставленной задачи	описывает модели машинного обучения	Тестирование
			ПК-5.2. Разрабатывает модели машинного обучения для решения задач	описывает модели предсказания	Тестирование
			ПК-5.3. Создает, поддерживает и использует системы искусственного интеллекта, включающие разработанные модели и методы, с применением выбранных инструментов машинного обучения	описывает статистические методы	Тестирование
		У(ПК-5и-22Г)	ПК-5.1. Осуществляет оценку и выбор инструментальных средств для	разрабатывает модели работы с данными	Лабораторная работа

				решения поставленной задачи		
				ПК-5.2. Разрабатывает модели машинного обучения для решения задач	разрабатывает модели предсказания	Лабораторная работа
				ПК-5.3. Создает, поддерживает и использует системы искусственного интеллекта, включающие разработанные модели и методы, с применением выбранных инструментов машинного обучения	демонстрирует статистические методы	Лабораторная работа

2. Перечень оценочных средств для текущего контроля и промежуточной аттестации по дисциплине

п/п	Вид оценочного средства	Краткая характеристика оценочного средства	Представление оценочного средства в фонде	Шкала оценки
1	2	3	4	5
1	Лабораторная работа	Средство проверки умений применять полученные знания по заранее определенной методике для решения задач или заданий по лабораторным исследованиям	Темы, задания для выполнения лабораторных работ; вопросы и требования к их защите	оценка «отлично» выставляется обучающемуся, если обучающийся правильно выполнил лабораторную работу; при защите ответил на все дополнительные вопросы; оценка «хорошо» выставляется обучающемуся, если обучающийся выполнил лабораторную работу с небольшими неточностями; при защите ответил на большинство дополнительных вопросов; оценка «удовлетворительно» выставляется обучающемуся, если обучающийся выполнил лабораторную работу с небольшими неточностями; при защите ответил на большинство дополнительных вопросов; оценка «хорошо» выставляется обучающемуся, если обучающийся ответил на теоретические вопросы с небольшими неточностями; показал хорошие знания в рамках усвоенного учебного материала; ответил на большинство дополнительных вопросов; оценка «удовлетворительно» выставляется обучающемуся, если обучающийся выполнил лабораторную работу с существенными неточностями; при защите на дополнительные

				ные вопросы было допущено несколько неправильных ответов; оценка « <i>неудовлетворительно</i> » выставляется обучающемуся, если при выполнении лабораторной работы обучающийся продемонстрировал недостаточный уровень владения умениями и навыками при решении профессиональных задач в рамках усвоенного учебного материала; при защите на дополнительные вопросы были даны неправильные ответы;
2	Письменный и устный опрос	Оценочное средство для текущего контроля успеваемости и промежуточной аттестации. Позволяет выявить и восполнить пробелы в знаниях; повторить, закрепить, систематизировать материал; оценить знания, умения, теоретические и практические навыки; определить уровень сформированных у студентов компетенций по дисциплине (модулю)	Совокупность вопросов, заданий, упражнений, тестов для выполнения контрольных работ, домашних заданий, РГР и иных учебных работ. Комплект билетов для текущей и промежуточной аттестации	оценка « <i>отлично</i> » выставляется обучающемуся, если обучающийся правильно ответил на теоретические вопросы; показал отличные знания в рамках усвоенного учебного материала; ответил на все дополнительные вопросы; решены все задачи; оценка « <i>хорошо</i> » выставляется обучающемуся, если обучающийся ответил на теоретические вопросы с небольшими неточностями; показал хорошие знания в рамках усвоенного учебного материала; ответил на большинство дополнительных вопросов; оценка « <i>удовлетворительно</i> » выставляется обучающемуся, если обучающийся ответил на теоретические вопросы с существенными неточностями; показал удовлетворительные знания в рамках усвоенного учебного материала; при ответах на дополнительные вопросы было допущено несколько неправильных ответов; оценка « <i>неудовлетворительно</i> » выставляется обучающемуся, если при ответе на теоретические вопросы обучающийся продемонстрировал недостаточный уровень знаний; на дополнительные вопросы отвечает неправильно;
3	Тестирование	Система стандартизированных простых и комплексных заданий, позволяющая автоматизировать процедуру измерения уровня знаний, умений и владений обучающегося.	Фонд тестовых заданий.	оценка « <i>отлично</i> » выставляется обучающемуся, если обучающийся дал правильные ответы на 90...100 % вопросов; оценка « <i>хорошо</i> » выставляется обучающемуся, если обучающийся дал правильные ответы на 75...89 % вопросов; оценка « <i>удовлетворительно</i> » выставляется обучающемуся, если обучающийся дал правильные ответы на 60...74 % вопросов; оценка « <i>неудовлетворительно</i> » выставляется обучающемуся, если обучающийся дал правильные ответы на 0...59 % вопросов;

Приложение В

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Уфимский государственный нефтяной технический университет»

Письменный и устный опрос.

Перечень вопросов (задач, заданий, тем, комплекта тестовых заданий):

Вопросы:

Стандартизация (Z-преобразование)

Нормализация данных

критерий Манна-Уитни.

тест Шапиро-Уилк

параметры однофакторного дисперсионный анализ данных

параметры двухфакторного дисперсионный анализ

параметры коэффициент корреляция

коэффициенты линии регрессии

параметры анализа номинативных данных

параметры логистической регрессии

Однофакторный дисперсионный анализ

Двухфакторный дисперсионный анализ

Корреляция и регрессия

Одномерный регрессионный анализ

Многофакторный регрессионный анализ

Анализ номинативных данных

Логистическая регрессия

Модель с одним номинативным предиктором

Модель с двумя номинативными предикторами

Смешанная регрессионная модель

Метаоптимизация

Лабораторная работа.

Перечень вопросов (задач, заданий, тем, комплекта тестовых заданий):

Датасеты с UCI Machine Learning Repository <https://archive.ics.uci.edu/ml/index.php>

Работа с библиотеками Pandas, Scikit-Learn, TensorFlow в облачном сервисе Google Colab

ЛР

Предварительнj обработать данные

Извлечь признаки

Нормализовать

Преобразование входные данные для использования с алгоритмами машинного обучения.

ЛР

Снизить размерность

Уменьшить количество случайных переменных

Выбрать признаки

Применить анализ главных компонент (PCA)

Визуализация

ЛР

Выбрать модель

Сравнить, проверить и выбрать параметры и модели.

ЛР

Нормализовать данные

Вычислить критерий Манна-Уитни.

Вычислить тест Шапиро-Уилк

Вычислить параметры однофакторного дисперсионный анализ данных

Вычислить параметры двухфакторного дисперсионный анализ

Вычислить параметры коэффициент корреляция

Выполнить регрессионный анализ

Вычислить коэффициенты линии регрессии

Спрогнозировать атрибуты с непрерывным значением,

Реализовать SVR

Ближайшие соседи

ЛР

Выполнить классификацию

Определить, к какой категории относится объект.

Вычислить параметры анализа номинативных данных

Вычислить параметры логистической регрессии

SVM

Случайный лес

ЛР

Кластеризация

Автоматическая группировка похожих объектов в наборы.

Сгруппировать результаты экспериментов

реализовать k-Means

Статистика на языке программирования R : учебно-методическое пособие для лабораторных, практических и самостоятельных работ / УГНТУ, каф. Математики ; сост.: С. Н. Коледин [и др.]. - Уфа : УГНТУ, 2018. - 1,24 Мб. - URL:

http://bibl.rusoil.net/base_docs/UGNTU/Matematika/Koledin5.pdf. - Текст : электронный

Тестирование.

Перечень вопросов (задач, заданий, тем, комплекта тестовых заданий):

1. Если при исследовании взаимосвязи двух переменных коэффициент детерминации оказался равен 0,25 то чему равен коэффициент корреляции
 - 0,5*
 - -0,5*
2. Если коэффициент корреляции равен 0,7, то такая взаимосвязь всегда будет статистически достоверна (p - уровень значимости обязательно будет меньше 0,05)
 - Да, вероятность получить такую или еще более выраженную взаимосвязь случайно практически равна нулю.
 - Нет, высокая корреляция не обязательно означает статистически значимую взаимосвязь.*
3. Чему равен коэффициент корреляции в данной выборке

X	4	5	2	3	1
Y	2	1	4	3	5

 -1*
4. Укажите основные требования к данным при использовании коэффициента корреляции Пирсона:
 - Линейная взаимосвязь двух переменных*
 - Обе переменные обязательно должны быть представлены в ранговой шкале
 - Значения переменных не должны быть отрицательными
5. Если по результатам исследования мы обнаружили значимую положительную корреляцию уровня величины 1 и уровня величины 2, то какой вывод мы можем сделать о причинно-следственной связи?
 - Величина 1- причина Величины 2.
 - Величина 2- причина Величина 1.
 - Результаты только корреляционного исследования не позволяют нам делать вывод о причинно - следственной связи.*
6. В исследовании приняло участие 11 классов (с 1 по 11 класс, по 30 человек в каждом классе) некоторой школы. Оказалось, что существует значимая положительная корреляция этих двух показателей ($r = 0,7$, $p < 0,05$). Какой вывод он может сделать, основываясь на этих данных?
 - Доказано влияние словарного запаса на рост.
 - Доказано влияние роста на словарный запас.
 - Причиной данных результатов, вероятнее всего, является влияние третьей переменной (в каком классе учится школьник), если учесть этот показатель, то значимая взаимосвязь между ростом и словарным запасом не будет обнаружена.*
7. Проверая нулевую гипотезу об отсутствии взаимосвязи двух наших переменных, мы применяем t - критерий для каких целей?
 - для проверки гипотезы о том, что в генеральной совокупности $b_1 = 0$ *
 - для проверки гипотезы о том, что в генеральной совокупности $b_1 < 0$
 - для проверки гипотезы о том, что в генеральной совокупности $b_0 > 0$
 - для проверки гипотезы о том, что в генеральной совокупности $b_0 = 0$
8. Отклонив нулевую гипотезу, о том что $\beta_1=0$, какой вывод можем сделать:
 - Коэффициент корреляции между двумя переменными обязательно больше нуля
 - Регрессионная прямая пересекает ось Y выше нуля
 - Обнаружена причинно-следственная связь между зависимой и независимой переменной

- Коэффициент корреляции между двумя переменными обязательно меньше нуля
 - Обнаружена значимая взаимосвязь двух переменных*
9. В случае линейной взаимосвязи двух переменных распределение остатков какой вывод можем сделать
- будет иметь несколько мод
 - будет обладать явно выраженной асимметрией
 - будет стремиться к нормальному со средним значением, приблизительно равным нулю*
10. Что такое Гомоскедастичность
- одинаковая изменчивость зависимой переменной на всех уровнях независимой переменной (постоянная изменчивость остатков)*
 - очень сильная корреляция зависимой и независимой переменной
 - нормальное распределение остатков
11. Если в нашей модели коэффициент b_1 оказался не равен нулю, означает ли это, что обнаружена статистически значимая взаимосвязь ($p < 0,05$) между исследуемыми переменными?
- Нет, коэффициент b_1 не отвечает за угол наклона регрессионной прямой, его значение никак не влияет на направление взаимосвязи
 - Нет, коэффициент b_1 , не равный нулю, не всегда означает, что мы отклоняем нулевую гипотезу*
 - Да, коэффициент b_1 , не равный нулю, всегда означает, что мы отклоняем нулевую гипотезу
12. Если исследователь хочет предсказать значение уровня бедности, основываясь на значении уровня образованности, то в каком случае он может больше доверять предсказаниями регрессионной модели
- Взаимосвязь переменных линейная и монотонная, распределение остатков стремится к нормальному*
 - Ни характер связи, ни распределение остатков не может сказаться на точности предсказания.
 - Взаимосвязь переменных нелинейная и немонотонная, распределение остатков значимо отличается от нормального
13. Компания уже потратила 10 миллионов на производство фильма (cost), а также было продано 8 миллионов книг (books). Продюсер интересуется, сколько вы посоветуете ему потратить миллионов на рекламу фильма (promotion), если он хочет, чтобы сборы от фильма составили 150 миллионов? Рассчитайте необходимое количество миллионов на рекламу, основываясь на регрессионном уравнении из прошлого задания. В уравнении регрессии используйте только само число миллионов в качестве коэффициентов: например, 10 миллионов мы запишем в уравнении просто как 10, а не 10 000 000.
- 13*
14. Укажит обязательные условия для применения множественного регрессионного анализа
- Распределение каждой переменной должно значимо отличаться от нормального*
 - Независимые переменные должны быть линейным образом взаимосвязаны с зависимой переменной
 - Остатки должны быть приблизительно нормальным образом распределены вокруг нуля
 - Независимые переменные должны сильно коррелировать между собой*
15. Укажите модификации модели линейной регрессии:
- кластеризация

- лассо*
 - релинеаризация
 - регуляризация*
16. Укажите величины, отвечающие за качество алгоритмов классификации:
- точность*
 - F-значение*
 - MAE
 - полнота*
 - MAPE
17. Чем меньше значения TP и TN в матрице ошибок, тем хуже или лучше
- хуже*
 - лучше
18. Чем больше значения FP и FN в матрице ошибок, тем хуже или лучше
- хуже*
 - лучше
19. По тренировочной выборке из 70 элементов был построен некоторый классификатор. Мы взяли и проверили его качество на тестовой выборке, состоящей из 30 элементов. Сумма чисел TP+FP+FN+TN из матрицы ошибок чему равна?
30*
20. Представим, что было построено дерево решений, и в некоторый лист попало 5 объектов тренировочной выборки. Причем 2 объекта из них принадлежат классу 1. Какую метку дерево будет предсказывать для каждого тестируемого объекта, попавшего в этот лист?
0*
21. Выбор оптимального признака для дерева можно осуществить с помощью каких понятий
- Неопределенность Джини*
 - Энтропия*
 - Статистическая информативность*
22. Модель «изолирующий лес» предназначена для:
- предсказания значения числового признака
 - поиска выбросов*
 - кластеризации
 - предсказания метки класса
23. Искусственный нейрон имеет 2 входа с весами 1, 2 соответственно. Ко входному сигналу прибавляется число 1 и применяется сигмоидная функция. Чему будет равно значение на выходе из нейрона, если на первый вход было подано число 1, а на второй вход было подано число -1?
0,5
24. Выборка состоит из 100 объектов, 70 из них принадлежат классу 1, остальные принадлежат классу 0. Известно, что нецелевой бинарный признак P равен 1 для 10% объектов класса 1 и для

50% объектов из класса 0. Рассмотрим новый объект, у которого признак P равен 1. Чему равна вероятность, что этот объект принадлежит классу (ответ округлить до двух знаков после запятой).
0,32

25. формула Байеса

- $\Pr(B|A) = \Pr(A|B) * \Pr(A) / \Pr(B)$
- $\Pr(B|A) = \Pr(A|B) * \Pr(B) / \Pr(A) *$
- $\Pr(B|A) = \Pr(A|B * B/A)$
- $\Pr(B|A) = \Pr(A|B) / \Pr(B) * \Pr(A)$

26. События A , B называются независимыми, если:

- $\Pr(A) * \Pr(B) = \Pr(AB) *$
- $\Pr(AB) = \Pr(A) + \Pr(B)$
- $\Pr(A|B) = \Pr(B|A)$
- $\Pr(A|B) = \Pr(A) * \Pr(B)$

27. Комитет состоит из трех классификаторов. Окончательное решение принимается по большинству голосов. Классификаторы имеют точность 0,8 и 0,9 и 1 соответственно. Чему равна точность ансамбля алгоритмов? (ответ округлить до двух знаков после запятой).
0,98

28. Для классификации используется схема взвешенного голосования. Допустим, что объект A был определен первым и третьим классификаторами как объект из класса 1, а второй классификатор определил A как объект класса -1. Результат взвешенного голосования, если веса при классификаторах равны 2, 4, 3 соответственно, к какому классу отнесет объект A
1

29. К какому типу ансамбля он относится алгоритма Random Forest (случайный лес)

- голосование по большинству (комитет)*
- градиентный бустинг
- взвешенное голосование

30. Если к модели линейной регрессии применить процедуру бустинга, причем на втором этапе будет также строится линейная модель, то что произойдет с предсказанной точностью

- уменьшится
- возрастет
- останется прежней*

31. Комитет состоит из трех классификаторов. Окончательное решение принимается по большинству голосов. Классификаторы имеют точность 0,8 0,9 0,9 соответственно. Кроме того, второй и третий классификаторы полностью идентичны (то есть их ответы для любого классифицируемого объекта совпадают).

32. Чему равна точность ансамбля алгоритмов
0,9

33. Укажите необходимые условия возникновения проклятия размерности:

- используются данные без нормировки
- признаков очень много*
- используются метрические алгоритмы*
- признаки малоинформативны

- существует зависимость между признаками
- всё вышеперечисленное

34. Дан вектор значений признака $P=(1,0,5,2,2)$. Нормализуем этот вектор по формуле, использующей минимальное и максимальное значение признака P . Значение первой координаты нормализованного вектора будет равно

0,2

35. Дан вектор значений признака $P=(1,0,5,2,2)$. Нормализуем этот вектор по формуле, использующей среднее значение и отклонение признака P . Значение последней координаты нормализованного вектора будет равно

0

36. Первая и третья квартиль значений признака P равны 2, 4 соответственно. Какие из следующих значений будут считаться выбросами?

- 6,5
- 3
- 0
- 0,5
- -1,5*
- 8*
- 7,5*

Аннотация к рабочей программе дисциплины (626)Статистический анализ данных

Направление подготовки (специальность): 09.03.01 Информатика и вычислительная техника

Направленность: профиль«Технологии искусственного интеллекта в нефтегазовой отрасли»

Уровень высшего образования: бакалавриат

Форма обучения: очная;

Кафедра, обеспечивающая преподавание дисциплины: Вычислительная техника и инженерная кибернетика (ВТИК);

Компетенции, формируемые в результате освоения дисциплины

ПК-5и-22Г Способен использовать инструментальные средства для решения задач машинного обучения:

- ПК-5.1. Осуществляет оценку и выбор инструментальных средств для решения поставленной задачи
- ПК-5.2. Разрабатывает модели машинного обучения для решения задач
- ПК-5.3. Создает, поддерживает и использует системы искусственного интеллекта, включающие разработанные модели и методы, с применением выбранных инструментов машинного обучения

ПК-8и-22Г. Способен разрабатывать системы анализа больших данных:

- ПК-8.1. Разрабатывает программные компоненты извлечения, хранения, подготовки больших данных с учётом вариантов использования больших данных, определений, словарей и эталонной архитектуры больших данных
- ПК-8.2. Разрабатывает программные компоненты обработки удалённой, распределенной и объединённой аналитики, использования результатов анализа, описания и управления качеством и достоверностью больших данных

Результат обучения

Знать:

ПК-5и-22Г-1 знает модели машинного обучения

ПК-8и-22Г.-1 знает программные компоненты работы с большими данными

Уметь:

ПК-5и-22Г-1 умеет оценивать и вбирать средства для решения

ПК-8и-22Г.-1 умеет разрабатывать програмные компоннты работы с данными

Владеть:

ПК-5и-22Г-1 владеет системами искусственного интеллекта

ПК-8и-22Г.-1 владеет разработкой програмных компонент работы с данными

Краткая характеристика дисциплины

Инструменты для анализа данных; Задачи предсказаний машинного обучения;

Трудоёмкость (з.е. / часы)

3 з.е. (108час)

Вид промежуточной аттестации

экзамен;

Разработчик(и):

_____ старший преподаватель Шварева Е.Н.

СОГЛАСОВАНО

Заведующий кафедрой ВТИК _____..