

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

УТВЕРЖДАЮ

Директор ИЭС

Н.З. Солодилова

(подпись)

29.06.2023

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

(50579) Информационные технологии (продвинутый уровень) "Инструментарий математического моделирования и анализа данных"

Направление подготовки (специальность): **41.03.01 Зарубежное регионоведение**

Направленность: **профиль «Международные отношения»**

Уровень высшего образования: **бакалавриат**

Форма обучения: **очная;**

Кафедра, обеспечивающая преподавание дисциплины: **Цифровые технологии и моделирование (ЦТиМ);**

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

Уфа 2023

Рабочую программу дисциплины разработал(и):

_____ к.ф.-м.н., доцент Сафин Э.М.

Рецензент

_____ к.ф.-м.н., доцент Захаров А.В.

Рабочая программа дисциплины рассмотрена и одобрена на заседании кафедры Цифровые технологии и моделирование (ЦТиМ);, обеспечивающей преподавание дисциплины 07.04.2023, протокол №7.

Заведующий кафедрой

Цифровые технологии и моделирование (ЦТиМ)_____Д.В. Попов

СОГЛАСОВАНО

Заведующий кафедрой МОИВ_____С.Д. Галиуллина

Год приема 2023 г.

Рабочая программа зарегистрирована 08.06.2023 № 6 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): Концепции современного естествознания

Дисциплины, для которых освоение данной дисциплины необходимо как предшествующее (исходя из формирования этапов по компетенциям): Ознакомительная практика; Профессиональная практика; Управление данными

Блок: Блок 1. Дисциплины (модули);

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

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	
2	3	96	42	54	экзамен;
ИТОГО:	3	96	42	54	

2. Компетенции обучающегося, формируемые в результате освоения дисциплины

№ пп.	Формируемые компетенции	Шифр/ индекс компетен- ции
1	Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности	ОПК-2-22г.- 1
2	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1-2

В результате освоения дисциплины обучающийся должен:

Шифр компетенции	Индикаторы достижения компетенций	Шифр результата обучения	Результат обучения
УК-1	УК-1.1 Выполняет сбор и обра-	З(УК-1)	Знать:

Шифр компетенции	Индикаторы достижения компетенций	Шифр результата обучения	Результат обучения
	ботку информации по изучаемой теме УК-1.2 Производит анализ и синтез информации, полученной из различных источников УК-1.3 Применяет методики системного анализа при решении поставленных задач		принципы сбора и обработки информации по изучаемой теме; методики анализа и синтеза информации из различных источников; принципы системного анализа и их применение при решении задач.
		У(УК-1)	Уметь: выполнять сбор и обработку информации по изучаемой теме с использованием современных информационных технологий; производить анализ и синтез информации, полученной из различных источников, с применением методик системного анализа.
		В(УК-1)	Владеть: навыками системного подхода к решению задач, включая умение оценивать и анализировать сложные ситуации с использованием информационных технологий.
ОПК-2-22г.	ОПК-2.1. Навыки применения современных информационных технологий и программных средств, при решении задач профессиональной деятельности	З(ОПК-2-22г.)	Знать: современные информационные технологии и программные продукты, используемые в области математического моделирования и анализа данных.
		У(ОПК-2-22г.)	Уметь: применять современные информационные технологии и программные продукты при решении профессиональных задач в области математиче-

Шифр компетенции	Индикаторы достижения компетенций	Шифр результата обучения	Результат обучения
			ского моделирования и анализа данных.
		В(ОПК-2-22г.)	Владеть: навыками применения современных информационных технологий и программных средств, включая технологии искусственного интеллекта, при решении задач профессиональной деятельности в контексте математического моделирования и анализа данных.

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:	46		46										
лекции (всего)	14		14										
-в т.ч. лекции on-line курс	0												
практические занятия (ПЗ)	0												
-в т.ч. практические занятия on-line курс	0												
лабораторные работы (ЛР)	22		22										
контролируемая самостоятельная работа (защита курсового проекта, курсовой работы и др. работ (при наличии))	0												
-в т.ч. лабораторные работы on-line курс	0												
иная контактная работа (сдача зачета, экзамена, консультации)	6		6										
проектная деятельность (ПД)	4		4										
Самостоятельная работа обучающихся (СРО), всего в том числе: (указать конкретный вид СРО)	62		62										
выполнение и подготовка к защите курсового проек-	0												

та или курсовой работы													
выполнение и подготовка к защите РГР работы, реферата, патентных исследований, аналитических исследований и т.п	0												
изучение учебного материала, вынесенного на самостоятельную проработку	11		11										
подготовка к лабораторным и/или практическим занятиям	20		20										
подготовка к сдаче зачета, экзамена	23		23										
иные виды работ обучающегося (при наличии)	0												
освоение on-line курса	0												
самостоятельная проектная деятельность (СПД)	8		8										
ИТОГО ПО ДИСЦИПЛИНЕ	108		108										

4. Содержание дисциплины

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

Форма обучения: очная

(раздела)Номер темы	Название темы (раздела)	Семестр	Трудоемкость, часы					Шифр результата обучения
			Л	ПЗ	ЛР	СРО	Всего	
1	Статистический анализ.	2	6		6	26	38	З(УК-1) У(УК-1) В(УК-1)
2	Методы машинного обучения.	2	8	4	16	36	64	З(ОПК-2-22г.) У(ОПК-2-22г.) В(ОПК-2-22г.)
ИТОГО:			14	4	22	62	102	

4.2. Содержание лекционного курса

№ пп.	Номер раздела	Название темы	Трудоемкость, часы
-------	---------------	---------------	--------------------

			очная	очно- заочная	заочная
1	1-Статистический анализ.	Данные в экономике, их визуализация и предварительная обработка. Данные в экономике: объекты, признаки и таблицы, гistogramмы и условное форматирование. Визуализация данных в Microsoft Excel: графики и диаграммы рассеяния, инструменты описательной статистики, диаграммы размаха и сводные таблицы, предварительная обработка данных, выбросы, пропущенные значения.	2		
2	1-Статистический анализ.	Основы математической статистики. Дескриптивная (описательная) статистика: эмпирические распределения и их графические представления; статистическое оценивание параметров; статистическая проверка гипотез. Однофакторный дисперсионный анализ.	2		
3	1-Статистический анализ.	Основы эконометрики. Двумерная регрессионная модель: спецификация парной линейной регрессии; оценивание параметров методом наименьших квадратов; интерпретация и верификация модели; прогнозирование	2		

		на основе линейной модели; нелинейная регрессия. Многомерная регрессионная модель: спецификация модели; линейная модель множественной регрессии; доверительные интервалы и проверка статистических гипотез; верификация модели и интерпретация коэффициентов; прогнозирование в многомерной линейной регрессии.			
4	2-Методы машинного обучения.	Введение в науку о данных (Data Science). Понимание проекта и данных (Project & Data Understanding). Методы визуализации данных в Knime. Определение Data Science. Методологии работы с данными: SEMMA (Sample, Explore, Modify, Model, Assess); CRISP-DM (Cross Industry Standard Process for Data Mining); KDD (Knowledge Discovery in Databases). Пример проекта в науке о данных. Категории задач: классификация; регрессия; кластеризация, сегментация; ассоциативный анализ; анализ отклонений. Понимание проекта и данных (Project & Data Understanding). Примеры применения науки о	2		

		данных. Снижение размерности (Dimensionality Reduction). Методы снижения размерности: анализ главных компонент (PCA); линейный дискриминантный анализ (LDA); и др.			
5	2-Методы машинного обучения.	Основные принципы машинного обучения. Подготовка данных (Data Preparation). Рекомендательные системы. Кластеризация. Основные стратегии для алгоритмов машинного обучения: закрытые формулы; градиентный спуск; стратегии поиска. Измерения ошибок для валидации моделей: матрица ошибок (Confusion matrix); статистические метрики классов; метрики точности; ROC-кривые; коэффициент Коэна; числовые метрики ошибок. Валидация модели: типы ошибок модели; стратегии для более надежной оценки модели; кросс-валидация; работа с несбалансированными наборами данных; меры сложности модели. Подготовка данных: выбор признаков; снижение размерности; выбор и отбор записей; обработка пропущенных значений; выявление	2		

		выбросов; очистка данных и стандартизация; нормализация; инженерия данных; интеграция данных. Ассоциативные правила: майнинг наборов элементов; генерация ассоциативных правил. Коллаборативная фильтрация: кластеризация; меры сходства; иерархическая кластеризация; методы DIANE и AGNES; прототипы и модельные методы кластеризации; K-Means; меры качества; GMM (Gaussian Mixture Model); кластеризация на основе плотности; DBScan.			
6	2-Методы машинного обучения.	Байесовские классификаторы. Деревья принятия решений и регрессии. Байесовские классификаторы: различные виды байесовских классификаторов, включая наивные и полные; сравнение эффективности наивных и полных байесовских классификаторов. Деревья принятия решений: процесс построения дерева принятия решений. возможные меры оценки эффективности деревьев принятия решений; обработка числовых атрибутов и пропущенных значений в деревьях принятия решений; техники об-	2		

		резки деревьев принятия решений, такие как обрезка по снижению ошибки, оптимистичная обрезка, обрезка на уровне уверенности и минимальная длина описания. Деревья регрессии: применение деревьев принятия решений к задачам регрессии; обзор задачи регрессии; линейная регрессия в контексте деревьев принятия решений; другие методы регрессии, такие как логистическая регрессия, робастная регрессия и регрессия для классификации; практический пример, иллюстрирующий использование деревьев регрессии.			
7	2-Методы машинного обучения.	Нейронные сети. Глубокое обучение. Ансамблевое обучение. Метод k ближайших соседей. Метод опорных векторов. Многослойные перцептроны: однослойный перцептрон; многослойные перцептроны; алгоритм обратного распространения ошибки; рекурсивность для обучения скрытых нейронов; черные ящики в машинном обучении; проблема переобучения и методы её предотвращения; специальные ар-	2		

		хитектуры. Глубокое обучение: обзор глубокого обучения и его разновидностей; рекуррентные нейронные сети и их применение; сети с долгосрочной и краткосрочной памятью; сверточные нейронные сети; генеративные состязательные сети (GANs). Методы ансамблевого обучения: Bagging, Boosting. Стекинг и каскадная генерализация. Ансамбли деревьев и случайные леса. AdaBoost и градиентный бустинг. Практические примеры использования. Метод k ближайших соседей в прогнозировании, выбор параметра k. Метод опорных векторов.			
	-	ИТОГО:	14		

4.3. Перечень лабораторных работ

Номер раздела	№ ЛР	Название лабораторной работы	Трудоемкость, часы		
			очная	очно-заочная	заочная
1-Статистический анализ.	1	Описательная статистика в MS Excel. Изучение методов описательной статистики в Microsoft Excel с использованием рядов и графиков для анализа экономических данных. Выработка навыков	2		

		визуализации, интерпретации и числового описания выборочных характеристик.			
1-Статистический анализ.	2	Анализ регрессии и корреляции в Excel. Изучение методов анализа регрессии и корреляции в Microsoft Excel, приобретение навыков построения прогнозов и определения взаимосвязей между переменными на основе реальных данных.	2		
1-Статистический анализ.	3	Введение в аналитическую платформу Loginom: основы и практическое применение. Ознакомление с базовыми принципами и функциональностью аналитической платформы Loginom. Приобретение практических навыков работы с данными, реализации расчетов и загрузки данных из различных источников с использованием Loginom.	2		
2-Методы машинного обучения.	4	Введение в KNIME Analytics Platform. Ознакомление с KNIME Analytics Platform, включая установку программы, изучение ее интерфейса и	2		

		основных функций. Знакомство с процедурой создания рабочих процессов (workflow) используя узлы (node) на примере визуального анализа данных продаж.			
2-Методы машинного обучения.	5	Оценка точности моделей классификации и регрессии. Исследование и практическое применение алгоритма k-Means для кластеризации данных с использованием KNIME	2		
2-Методы машинного обучения.	6	Кластеризация. Исследование и практическое применение алгоритма k-Means для кластеризации данных с использованием KNIME.	2		
2-Методы машинного обучения.	7	Деревья классификации и регрессии. Изучение и применение линейной и логистической регрессии для решения задач предсказания цен и классификации по типу в KNIME Analytics Platform.	2		
2-Методы машинного обучения.	8	Линейная и логистическая регрессии. Овладеть навыками в обучении и применении модели дерева решений для задачи бинарной класси-	2		

		фикации и модели регрессионного дерева для предсказания цены на основе числовых признаков с последующей оценкой их производительности в программе Knime.			
2-Методы машинного обучения.	9	Случайный лес, градиентный бустинг деревьев. Овладеть навыками применения случайного леса и градиентного бустинга деревьев для задачи классификации, а также анализа важности признаков и оптимизации параметров моделей в среде KNIME.	2		
2-Методы машинного обучения.	10	Метод опорных векторов SVM. В программе KNIME – разработать и оценить модель машинного обучения (SVM) для классификации типов транспортных средств на основе числовых атрибутов, оптимизируя параметры для достижения максимальной точности классификации.	2		
2-Методы машинного обучения.	11	Нейронные сети в задачах классификации и предсказания. Нейронные сети в задачах классификации и предсказания.	2		

-		ИТОГО:	22		
---	--	---------------	----	--	--

4.4. Перечень практических занятий

Номер раздела	№ ПЗ	Тема практического занятия	Трудоемкость, часы		
			очная	очно-заочная	заочная
2-Методы машинного обучения.	1	Расширенный анализ данных: применение знаний в реальных проектах. Применение знаний и навыков, полученных в ходе обучения по дисциплине "Информационные технологии (продвинутый уровень): Инструментарий математического моделирования и анализа данных" в реальных практических задачах.	4		
-		ИТОГО:	4		

4.5. Виды СРО

Номер раздела	Вид СРО	Трудоемкость, часы		
		очная	очно-заочная	заочная
1-Статистический анализ.	подготовка к сдаче зачета, экзамена	10		
1-Статистический анализ.	подготовка к лабораторным и/или практическим занятиям	5		
1-Статистический анализ.	изучение учебного материала, вынесенного на самостоятельную	11		

	проработку			
2-Методы машинного обучения.	подготовка к сдаче зачета, экзамена	13		
2-Методы машинного обучения.	самостоятельная проектная деятельность (СПД)	8		
2-Методы машинного обучения.	подготовка к лабораторным и/или практическим занятиям	15		
-	ИТОГО:	62		

Темы для самостоятельной работы обучающихся

Раздел 1. Статистический анализ.

- Проверка гипотез и сопутствующие риски.
- Проверка гипотез с помощью доверительных интервалов.
- Наблюдаемый уровень значимости (р-значение).
- Проверка гипотезы о числовом значении математического ожидания нормального закона распределения.
- Проверка гипотезы о числовом значении вероятности.
- Проверка гипотезы о числовом значении дисперсии нормального закона распределения.
- Проверка гипотезы о равенстве дисперсий двух генеральных совокупностей.
- Проверка гипотезы о равенстве математических ожиданий двух генеральных совокупностей.
- Проверка гипотезы о равенстве генеральных долей.
- Критерии согласия.
- Критерии независимости качественных признаков.
- Критерии сдвига.
- Однофакторный дисперсионный анализ.
- Двухфакторный дисперсионный анализ.

Раздел 2. Методы машинного обучения.

- Подробное изучение SEMMA, CRISP-DM и KDD.
- Изучение различных категорий задач в науке о данных и их применение.
- Анализ классических примеров применения науки о данных.
- Понимание процесса ETL и его роль в науке о данных.
- Глубокое изучение методов визуализации данных.
- Понимание статистических дескрипторов и их роли в анализе данных.
- Применение методов снижения размерности в реальных проектах науки о данных.
- Глубокое изучение стратегий машинного обучения.
- Работа с несбалансированными данными и меры сложности модели.
- Основные принципы и методы подготовки данных.
- Техники обработки пропущенных значений и выявления выбросов.
- Изучение ассоциативных правил и их применение.
- Применение кластеризации в рекомендательных системах.
- Более подробное изучение наивных и полных байесовских классификаторов.
- Практическое применение деревьев принятия решений с использованием различных методов оценки эффективности.

- Изучение различных методов регрессии и их применение в контексте машинного обучения.
- Подробное изучение различных типов нейронных сетей и их применение.
- Глубокое понимание концепции ансамблевого обучения и его методов.
- Исследование влияния параметра k в методе ближайших соседей (kNN).
- Исследование различных ядерных методов.

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
ЭБС Znanium.com	http://znanium.com/
ЭБС Лань	https://e.lanbook.com/
Электронная библиотека УГНТУ	http://www.bibl.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	12-312	Коммутатор(1);Компьютер i3-2120(2);Компьютер i3-3240 21.5" Acer(1);Монитор BENQ E 2210 HDA(3);Моноблок(1);Пр-р HPLaserJet(1);Системный блок Intel Core 2(3);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.
2	12-403	Доска магнит-марке(1);Интерактив. доска(1);Компьютер в комп. (17);Проектор(1);Системный блок ПК тип2 ПЭВМ Кламас G5400(1);Учебно-наглядные пособия по дисциплине,набор демонстрационного оборудования; Столы, стулья;	Учебная аудитория для проведения занятий лекционного типа – укомплектована специализированной (учебной) мебелью, набором демонстрационного оборудования и учебно-наглядными пособиями, обеспечивающими тематические иллюстрации, соответствующие рабочим учебным программам дисциплин (модулей).
3	12-404	Компьютер в комп.(3);Компьютер-моноблок(1);МФУ HP LJ Pro M(1);МФУ Kyocera(1);МФУ лазерное HP LaserJet Pro MFP M134a RU(G3Q66A) (1);Моноблок(1);Принтер HP LJ 1100(1);ПринтерSamsung SCX(1);Системный блок 1 ПЭВМ Кламас Navan IS001-BK 450W(1);Системный блок ПК(1);Шкаф(ы) для хранения	Помещения для хранения и профилактического обслуживания учебного оборудования
4	12-406	Доска магнит-марке(1);Компьютер в комп.(10);Системный блок ПК тип2 ПЭВМ Кламас G5400(1);Столы, стулья	Лаборатория – оснащенная лабораторным оборудованием, в зависимости от степени сложности.
5	12-406	Доска магнит-марке(1);Компьютер в комп.(10);Системный блок ПК тип2 ПЭВМ Кламас G5400(1);Столы, стулья	Учебная аудитория для проведения групповых и индивидуальных консультаций
6	12-406	Доска магнит-марке(1);Компьютер в комп.(10);Системный блок ПК тип2 ПЭВМ Кламас G5400(1);Столы, стулья	Учебная аудитория для текущего контроля и промежуточной аттестации – укомплектована специализированной (учебной) мебелью, техническими средствами обучения.

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	Anaconda	Дата выдачи лицензии 01.01.2006, Поставщик: Свободное программное обеспечение
2	LibreOffice	Дата выдачи лицензии 01.01.2006, Поставщик: Свободное программное обеспечение
3	Office Professional Plus	Дата выдачи лицензии 29.09.2019, Поставщик: АО «СофтЛайн Трейд»
4	PyCharm	Дата выдачи лицензии 01.01.2006, Поставщик: Свободное программное обеспечение
5	Python	Дата выдачи лицензии 01.01.1991, Поставщик: Свободное программное обеспечение
6	Python v 3.7.1	Дата выдачи лицензии 01.01.2000, 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: (50579)(50579)Информационные технологии (продвинутый уровень) "Инструментарий математического моделирования и анализа данных"

Направление подготовки (специальность): 41.03.01 Зарубежное регионоведение

Направленность: профиль«Международные отношения»

Форма обучения: очная;

Кафедра, обеспечивающая преподавание дисциплины: Цифровые технологии и моделирование (ЦТиМ);

Тип	Назначение учебных изданий	Семестр			Библиографическое описание	Кол-во экз.	Адрес нахождения электронного	Коэффициент
		очная	очно-заочная	заочная				
1	2	3	4	5	6	7	8	9
Основная литература	Для изучения теории;	2			Жуковский, О. И. Информационные технологии и анализ данных: учебное пособие / О. И. Жуковский. — Москва: ТУСУР, 2014. — 130 с. — https://e.lanbook.com/book/110351	1	http://www.e.lanbook.com	1.00

Основная литература	Для выполнения лабораторных работ; Для изучения теории;	2			Кузьмин, В. И. Методы анализа данных: учебное пособие / В. И. Кузьмин, А. Ф. Гадзаов. — 2-е изд., перераб. и доп. — Москва: РТУ МИРЭА, 2020. — 155 с. — https://e.lanbook.com/book/171433	1	http://www.e.lanbook.com	1.00
Основная литература	Для выполнения СРО; Для изучения теории;	2			Низаметдинов, Ш. У. Анализ данных: учебное пособие / Ш. У. Низаметдинов, В. П. Румянцев. — Москва: НИЯУ МИФИ, 2012. — 288 с. — https://e.lanbook.com/book/75847	1	http://www.e.lanbook.com	1.00
Основная литература	Для выполнения лабораторных работ; Для изучения теории;	2			Борисова, Л. Р. Математика и анализ данных с поддержкой MS Excel и языка R : учебное пособие / Л. Р. Борисова, Н. А. Светлова, И. Ю. Седых ; под. ред. И. Ю. Седых. - Москва: Прометей, 2023. - 728 с. https://znanium.com/catalog/product/2124863	1	http://www.znanium.com	1.00
Дополнительная литература	Для выполнения СРО; Для выполнения лабораторных работ;	2			Крюкова, А. А. Работа с аналитической платформой Loginom: методические указания / А. А. Крюкова. — Самара: ПГУТИ, 2021. — 82 с. — https://e.lanbook.com/book/301124	1	http://www.e.lanbook.com	1.00

Дополнительная литература	Для выполнения СРО; Для изучения теории; Для проектной деятельности;	2			Юре, Л. Анализ больших наборов данных / Л. Юре, Р. Ананд, Д. У. Джеффри; перевод с английского А. А. Слинкин. — Москва: ДМК Пресс, 2016. — 498 с. — https://e.lanbook.com/book/93571	1	http://www.e.lanbook.com	1.00
Примечание – Графы 1-5,8 заполняются кафедрой, графы 7 и 9 - библиотекой								

Составил:

_____ к.ф.-м.н., доцент Сафин Э.М.

Год приема 2023 г.

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

Наименование дисциплины: (50579)(50579)Информационные технологии (продвинутый уровень) "Инструментарий математического моделирования и анализа данных"

Направление подготовки (специальность): 41.03.01 Зарубежное регионоведение

Направленность профиль«Международные отношения»

Форма обучения очная;

Кафедра, обеспечивающая преподавание дисциплины: Цифровые технологии и моделирование (ЦТиМ);

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;Для проектной деятельности;	2			Информационные технологии (продвинутый уровень): инструментальный математического моделирования и анализа данных : учебно-методическое пособие к выполнению лабораторных и самостоятельных работ, проектной деятельности / УГНТУ, каф. ЦТиМ ; сост. Э. М. Сафин. - Уфа : УГНТУ, 2023. - 1805 Кб. https://bibl.rusoil.net/base_docs/UGNTU/IES/Safin17644.pdf	1	0	http://bibl.rusoil.net	1.00
Примечание – Графы 1-5,8 заполняются кафедрой, графы 6,7 и 9 - библиотекой								

Составил:

_____ к.ф.-м.н., доцент Сафин Э.М.

—
Год приема 2023 г.

Приложение Б

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

УТВЕРЖДАЮ

Директор ИЭС

_____ Н.З. Солодилова
29.06.2023

**Фонд оценочных средств по текущей успеваемости и промежуточной аттестации по дисциплине
(50579) Информационные технологии (продвинутый уровень) "Инструментарий математического моделирования и анализа данных"**

Направление подготовки (специальность): 41.03.01 Зарубежное регионоведение

Направленность: профиль «Международные отношения»

Уровень высшего образования: бакалавриат

Форма обучения: очная;

Кафедра, обеспечивающая преподавание дисциплины: Цифровые технологии и моделирование (ЦТиМ);

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

Уфа 2023

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

_____ к.ф.-м.н., доцент Сафин Э.М.

—
Рецензент

_____ к.ф.-м.н., доцент Захаров А.В.

—

ФОС по текущей успеваемости и промежуточной аттестации по дисциплине рассмотрен и одобрен на заседании кафедры Цифровые технологии и моделирование (ЦТиМ);, обеспечивающей преподавание дисциплины 07.04.2023, протокол №7.

Заведующий кафедрой Цифровые технологии и моделирование (ЦТиМ); _____ Д.В. Попов

СОГЛАСОВАНО

Заведующий кафедрой МОИВ _____ С.Д. Галиуллина

Год приема 2023 г.

ФОС по текущей успеваемости и промежуточной аттестации по дисциплине
зарегистрирован 08.06.2023 № 6 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	Статистический анализ.	В(УК-1)	принципы сбора и обработки информации по изучаемой теме; методики анализа и синтеза информации из различных источников; принципы системного анализа и их применение при решении задач.	УК-1.1 Выполняет сбор и обработку информации по изучаемой теме	Владеет: инструментами Microsoft Excel для визуализации данных и проведения анализа.	Лабораторная работа Письменный и устный опрос
				УК-1.2 Производит анализ и синтез информации, полученной из различных источников	Владеет: навыками интерпретации и верификации моделей, а также проведения прогнозирования на их основе.	Лабораторная работа Письменный и устный опрос
				УК-1.3 Применяет методики системного анализа при решении поставленных задач	Владеет: навыками работы с многомерной линейной регрессией, включая создание модели, проверку статистических гипотез и интерпретацию коэффициентов.	Лабораторная работа Письменный и устный опрос
		З(УК-1)		УК-1.1 Выполняет сбор и обработку информации по изучаемой теме	Знает: процесс предварительной обработки данных, включая обработку выбросов, пропущенных значений, а	Лабораторная работа Письменный и

					также создание синтетических признаков и иерархии признаков.	устный опрос
			УК-1.2 Производит анализ и синтез информации, полученной из различных источников		Знает: основы дескриптивной статистики, включая эмпирические распределения, методы статистического оценивания параметров и проверки гипотез.	Лабораторная работа Письменный и устный опрос
			УК-1.3 Применяет методики системного анализа при решении поставленных задач		Знает: методы визуализации данных в Microsoft Excel, в том числе создание графиков и диаграмм рассеяния, использование инструментов описательной статистики, а также создание диаграмм размаха и сводных таблиц.	Лабораторная работа Письменный и устный опрос
		У(УК-1)	УК-1.1 Выполняет сбор и обработку информации по изучаемой теме		Умеет: проводить сбор и обработку информации по изучаемой теме.	Лабораторная работа Письменный и устный опрос
			УК-1.2 Производит анализ и синтез информации, полученной из различных источников		Умеет: визуализировать данные в Microsoft Excel, включая создание графиков и диаграмм рассеяния.	Лабораторная работа Письменный и

						устный опрос
				УК-1.3 Применяет методики системного анализа при решении поставленных задач	Умеет: выполнять однофакторный дисперсионный анализ с использованием MS Excel.	Лабораторная работа Письменный и устный опрос
10	Методы машинного обучения.	В(ОПК-2-22г.)	современные информационные технологии и программные продукты, используемые в области математического моделирования и анализа данных.	ОПК-2.1. Навыки применения современных информационных технологий и программных средств, при решении задач профессиональной деятельности	Владеет: инструментами Kpime для визуализации данных и анализа; теоретическими основами методов машинного обучения, включая понимание процесса Data Science, этапы оценки ситуации и целей анализа.	Лабораторная работа Письменный и устный опрос
		З(ОПК-2-22г.)		ОПК-2.1. Навыки применения современных информационных технологий и программных средств, при решении задач профессиональной деятельности	Знает: интеллектуальные основы Data Science и их применение в профессиональной деятельности; решающие деревья, байесовские классификаторы, регрессия, обучение правилам, прогнозирование методом ближайших соседей, искусственные нейронные сети, глубокое обучение, метод опор-	Лабораторная работа Письменный и устный опрос

					ных векторов, ансамблевые методы.	
		У(ОПК-2-22г.)		ОПК-2.1. Навыки применения современных информационных технологий и программных средств, при решении задач профессиональной деятельности	Умеет: реализовывать основные этапы процесса Data Science, включая понимание проекта, понимание данных, принципы моделирования, подготовку данных, поиск закономерностей и обеспечение предиктивности; применять методы решающих деревьев, байесовских классификаторов, регрессии, обучение правилам, прогнозирование методом ближайших соседей, использование искусственных нейронных сетей, глубокого обучения, метода опорных векторов и ансамблевых методов.	Лабораторная работа Письменный и устный опрос

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	Темы, задания для выполнения лабораторных работ; вопросы и требования к их защите	оценка «отлично» выставляется обучающемуся, если обучающийся выполняет задания лабораторной работы, оформляет отчет в соответствии с требованиями по выполнению работы, отвечает на контрольные вопросы. оценка «хорошо» выставляется обучающемуся, если обучаю-

		по лабораторным исследованиям		щийся выполняет задания лабораторной работы, оформляет отчет в соответствии с требованиями по выполнению работы, с ошибками отвечает на контрольные вопросы. оценка «удовлетворительно» выставляется обучающемуся, если обучающийся выполняет задания лабораторной работы, оформляет отчет в соответствии с требованиями по выполнению работы, не отвечает на контрольные вопросы. оценка «неудовлетворительно» выставляется обучающемуся, если обучающийся не выполняет задания лабораторной работы, не оформляет отчет в соответствии с требованиями по выполнению работы, не отвечает на контрольные вопросы.
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 «отлично» выставляется обучающемуся, если обучающийся проявил всестороннее, систематическое и глубокое знание учебного материала; демонстрирует понимание взаимосвязей между разделами дисциплины и со смежными дисциплинами; умеет успешно выполнять предусмотренные программой задания; в работе нет существенных ошибок. оценка «хорошо» выставляется обучающемуся, если обучающийся обладает знанием основного учебного материала; умеет успешно выполнять предусмотренные программой задания; возможны небольшие, несущественные ошибки в ответах или при выполнении заданий, которые самостоятельно исправлены; общий уровень знаний и навыков выше среднего. оценка «удовлетворительно» выставляется обучающемуся, если обучающийся знает основной учебный материал и выполняет задания на уровне, удовлетворяющем минимальным требованиям; допускает некоторые ошибки, которые, однако, не существенны и не влияют на общее понимание материала; требуется дополнительное внимание к некоторым аспектам учебной программы. оценка «неудовлетворительно» выставляется обучающемуся, если обучающийся демонстрирует недостаточно полный объем знаний в рамках образовательного стандарта; знание части основной литературы, рекомендованной учебной программой; использование научной терминологии, изложение ответа на вопросы с существенными лингвистическими и логическими ошибками; низкий уровень, культуры исполнения заданий.

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):

Список вопросов к экзамену

1. Данные. Представление данных.
2. Инструменты описательной статистики в Microsoft Excel.
3. Визуализация качественных признаков в Microsoft Excel.
4. Предварительная обработка данных.
5. Дескриптивная статистика.
6. Однофакторный дисперсионный анализ.
7. Двумерная регрессионная модель.
8. Многомерная регрессионная модель.
9. Прогнозирование и дополнительные аспекты.
10. Data Science. Инструменты реализации.
11. Методология работы с данными: SEMMA.
12. Методология работы с данными: CRISP-DM.
13. Методология работы с данными: KDD.
14. Машинное обучение: суть, принципы, задачи.
15. Метрики в машинном обучении, ROC-кривые.
16. Модели классификации с обучением.
17. Модели кластерного анализа.
18. Решающие деревья.
19. Байесовские классификаторы.
20. Обучение правилам: предикативные и индуктивные логические программы.
21. Прогнозирование методом ближайших соседей.
22. Метод опорных векторов.
23. Ансамблевые методы.
24. Искусственные нейронные сети: основные понятия, модель персептрона.
25. Глубокое обучение: рекуррентные нейронные сети, сверточные нейронные сети, генеративно-состязательные сети.

Пример экзаменационного билета:

Уфимский государственный нефтяной технический университет
Кафедра «Цифровые технологии и моделирование»

Экзаменационный билет № 1

по дисциплине «Информационные технологии (продвинутый уровень): инструментарий математического моделирования и анализа данных»

по направлению подготовки 41.03.01 Зарубежное регионоведение
профиль «Международные отношения»

1. Методология работы с данными: SEMMA.
2. Метод опорных векторов.

Заведующий кафедрой ЦТиМ

_____ Д.В. Попов

07.04.2023

Лектор

Э.М. Сафин

07.04.2023

Ответ на вопросы Билета №1

1. Методология работы с данными: SEMMA:

SEMMA представляет собой методологию, используемую в области данных и аналитики, и состоит из следующих этапов:

1. Sample (Выборка):

Выборка данных – первый этап методологии SEMMA. На этом этапе аналитики выбирают представительное подмножество данных для дальнейшего анализа. Пример. Предположим, у нас есть база данных клиентов банка. Мы можем взять случайную выборку из этой базы, чтобы провести более детальный анализ характеристик клиентов.

2. Explore (Исследование):

На этом этапе происходит исследование данных с использованием визуализации и описательной статистики. Пример. Проведем исследование распределения возраста клиентов и их доходов с помощью гистограмм и диаграмм рассеяния, чтобы выявить возможные закономерности или аномалии.

3. Modify (Предварительная обработка):

Этап предварительной обработки данных включает в себя устранение выбросов, заполнение пропущенных значений и создание новых признаков. Пример. Если у нас есть пропущенные значения в столбце "Доход", мы можем заполнить их средним значением дохода по всей выборке.

4. Model (Моделирование):

На этапе моделирования строятся математические модели для анализа данных. Пример. Мы можем применить линейную регрессию для прогнозирования вероятности клиента вернуть кредит, используя его возраст, доход и другие характеристики.

5. Assess (Оценка):

Оценка модели и результатов анализа данных. Пример. Используя метрики, такие как точность предсказаний или коэффициент детерминации, мы оцениваем эффективность модели и ее способность объяснить вариации в данных.

2. Метод опорных векторов:

Метод опорных векторов (Support Vector Machines – SVM) – это алгоритм машинного обучения, используемый для задач классификации и регрессии.

Основная идея SVM заключается в поиске оптимальной гиперплоскости, которая максимально разделяет точки данных разных классов. В случае классификации, эта гиперплоскость служит разделяющей границей, а опорные векторы – точки данных, ближайшие к этой границе.

Основные концепции SVM.

Максимизация зазора. Основная цель SVM – найти гиперплоскость с максимальным зазором между точками разных классов. Зазор – это расстояние от гиперплоскости до ближайших точек каждого класса.

Ядерные функции. SVM может использовать ядерные функции для преобразования данных в пространство более высокой размерности, где линейная гиперплоскость может эффективно разделять классы. Распространенные ядра включают полиномиальные, радиальные базисные функции (RBF) и сигмоидальные.

Метод опорных векторов. Опорные векторы – это точки данных, которые находятся ближе всего к гиперплоскости. Их позиция определяет распределение зазора. Опорные векторы играют ключевую роль в построении модели.

C-параметр. Параметр C в SVM управляет балансом между максимизацией зазора и минимизацией ошибок классификации. Большее значение C приводит к более сложной модели, которая может

адаптироваться к шуму в данных, но может привести к переобучению.

Лабораторная работа.

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

Информационные технологии (продвинутый уровень): инструментарий математического моделирования и анализа данных : учебно-методическое пособие к выполнению лабораторных и самостоятельных работ, проектной деятельности / УГНТУ, каф. ЦТиМ ; сост. Э. М. Сафин. - Уфа : УГНТУ, 2023. - 1805 Кб. https://bibl.rusoil.net/base_docs/UGNTU/IES/Safin17644.pdf

Список лабораторных работ

Лабораторная работа № 1. Описательная статистика в MS Excel

Лабораторная работа № 2. Анализ регрессии и корреляции в Excel

Лабораторная работа № 3. Введение в аналитическую платформу Logiplot: основы и практическое применение

Лабораторная работа № 4. Введение в KNIME Analytics Platform

Лабораторная работа № 5. Оценка точности моделей классификации и регрессии

Лабораторная работа № 6. Кластеризация

Лабораторная работа № 7. Деревья классификации и регрессии

Лабораторная работа № 8. Линейная и логистическая регрессии

Лабораторная работа № 9. Случайный лес, градиентный бустинг деревьев

Лабораторная работа № 10. Метод опорных векторов SVM

Лабораторная работа № 11. Нейронные сети в задачах классификации и предсказания

Аннотация к рабочей программе дисциплины (50579) Информационные технологии (продвинутый уровень) "Инструментарий математического моделирования и анализа данных"

Направление подготовки (специальность): 41.03.01 Зарубежное регионоведение

Направленность: профиль «Международные отношения»

Уровень высшего образования: бакалавриат

Форма обучения: очная

Кафедра, обеспечивающая преподавание дисциплины: Цифровые технологии и моделирование (ЦТиМ)

Компетенции, формируемые в результате освоения дисциплины

УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач:

- УК-1.1 Выполняет сбор и обработку информации по изучаемой теме
- УК-1.2 Производит анализ и синтез информации, полученной из различных источников
- УК-1.3 Применяет методики системного анализа при решении поставленных задач

ОПК-2-22г. Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности:

- ОПК-2.1. Навыки применения современных информационных технологий и программных средств, при решении задач профессиональной деятельности

Результат обучения

Знать:

УК-1-2 принципы сбора и обработки информации по изучаемой теме; методики анализа и синтеза информации из различных источников; принципы системного анализа и их применение при решении задач.
ОПК-2-22г.-1 современные информационные технологии и программные продукты, используемые в области математического моделирования и анализа данных.

Уметь:

УК-1-2 выполнять сбор и обработку информации по изучаемой теме с использованием современных информационных технологий; производить анализ и синтез информации, полученной из различных источников, с применением методик системного анализа.
ОПК-2-22г.-1 применять современные информационные технологии и программные продукты при решении профессиональных задач в области математического моделирования и анализа данных.

Владеть:

УК-1-2 навыками системного подхода к решению задач, включая умение оценивать и анализировать сложные ситуации с использованием информационных технологий.
ОПК-2-22г.-1 навыками применения современных информационных технологий и программных средств, включая технологии искусственного интеллекта, при решении задач профессиональной деятельности в контексте математического моделирования и анализа данных.

Краткая характеристика дисциплины

Статистический анализ.; Методы машинного обучения.;

Трудоёмкость (з.е. / часы)

3 з.е. (108час)

экзамен;

Вид промежуточной аттестации

Разработчик(и):

к.ф.-м.н., доцент Сафин Э.М.

СОГЛАСОВАНО

Заведующий кафедрой МОИВ _____ С.Д. Галиуллина