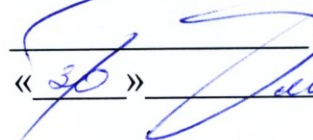




МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«КАЗАНСКИЙ ГОСУДАРСТВЕННЫЙ ЭНЕРГЕТИЧЕСКИЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «КГУ»)

УТВЕРЖДАЮ

Директор Института цифровых
технологий и экономики

 Э.И. Беляев
« 30 » Июль 2023 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1.В.02 Анализ больших данных



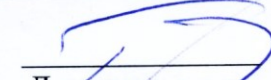
Направление подготовки 01.03.03 Прикладная информатика

Квалификация Бакалавр

г. Казань, 2023

Программу разработал(и):

Наименование кафедры	Должность, уч. степень, уч. звание	ФИО разработчика
Информационные технологии и интеллектуальные системы	доцент, к.т.н., доцент	Гадельшина Г.А.

Согласование	Наименование подразделения	Дата	№ протокола	Подпись
Одобрена	ИТИС	27.04.2023	Протокол №3	 Зав.каф., д.п.н., доц. Торкунова Ю.В.
Согласована	Учебно- методический совет ИЦТЭ	30.05.2023	Протокол №7	 Директор, к.т.н., доц. Беляев Э.И.
Одобрена	Ученый совет ИЦТЭ	30.05.2023	Протокол №9	 Директор, к.т.н., доц. Беляев Э.И.

Рецензия на рабочую программу и оценочные материалы по дисциплине «Анализ больших данных»

Содержание РПД и ОМ соответствует требованиям федерального государственного стандарта высшего образования по направлению подготовки 09.03.03 «Прикладная информатика» и учебному плану.

РПД и ОМ соответствуют требованиям, предъявляемым к структуре, содержанию РПД и ОМ по дисциплине, а именно:

1. Перечень формируемых компетенций, которыми должен овладеть обучающийся в результате освоения дисциплины, соответствует ФГОС ВО и профстандарту, будущей профессиональной деятельности выпускника.

2. Структура и содержание дисциплины соответствует учебному плану.

3. РПД содержит информацию об учебно-методическом, информационном и материально-техническом обеспечении дисциплины; об особенностях организации образовательной деятельности для лиц с ограниченными возможностями здоровья и инвалидов и методические рекомендации для преподавателей по организации воспитательной работы с обучающимися.

4. Показатели и критерии оценивания компетенций в ОМ, а также шкалы оценивания обеспечивают возможность проведения всесторонней оценки результатов обучения, уровней сформированности компетенций.

5. Контрольные задания и иные материалы оценки результатов освоения разработаны на основе принципов оценивания: валидности, определённости, однозначности, надёжности, а также соответствуют требованиям к составу и взаимосвязи оценочных средств, полноте по количественному составу оценочных средств и позволяют объективно оценить результаты обучения, уровни сформированности компетенций.

6. Направленность РПД и ОМ по дисциплине соответствует целям ОП по направлению 09.03.03 «Прикладная информатика», профстандартам.

Заключение. На основании проведенной экспертизы можно сделать заключение, что РПД и ОМ по дисциплине соответствует требованиям ФГОС ВО, профессионального стандарта, современным требованиям рынка труда и рекомендуются для использования в учебном процессе.

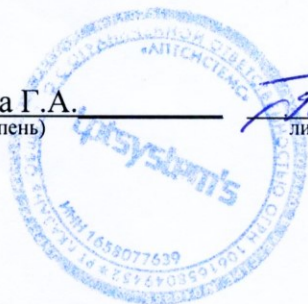
Рецензент

Генеральный директор «ЛПТСИСТЕМС» Фатыхова Г.А.

(Фамилия И.О., место работы, должность, ученая степень)

личная подпись

Дата 23.05.2023



1. Цель, задачи и планируемые результаты обучения по дисциплине

(Цель и задачи освоения дисциплины, соответствующие цели ОП)

Целью освоения дисциплины «Анализ больших данных» является формирование профессиональных компетенций, освоение теоретических знаний основных научных принципов и методов исследования с применением технологий больших данных, методов обработки и анализа данных; получение практических навыков и умений решения задач с помощью машинного обучения и продвинутой аналитики.

Задачами дисциплины являются:

познакомить обучающихся с основными научными принципами и методами исследования, изучить математические методы и алгоритмы обработки и анализа данных;

изучить современные информационные технологии и программное обеспечение для анализа данных;

сформировать практические навыки и умения использования современных информационных технологий, необходимых для анализа данных в различных прикладных областях;

обучить обучающихся применять прикладное программное обеспечение, в которых реализованы математические методы и алгоритмы, осуществляющие решение задач анализа данных, необходимых для профессиональной деятельности.

Компетенции и индикаторы, формируемые у обучающихся:

Код и наименование компетенции	Код и наименование индикатора
ПК-2 Способен осуществлять проектно-аналитические работы с использованием технологий больших данных и искусственного интеллекта	ПК-2.1 Способен планировать и организовывать аналитические работы с использованием технологий больших данных
	ПК-2.2 Выполняет полный цикл решения задач с помощью машинного обучения и продвинутой аналитики

2. Место дисциплины в структуре ОП

Предшествующие дисциплины: «Теория вероятностей и математическая статистика», «Математические модели и методы», «Алгоритмизация и программирование», «Пакеты прикладных программ и анализа данных».

Последующие дисциплины: «Проектный практикум по разработке информационных систем управления и анализу данных», «Производственная практика (преддипломная)».

3. Структура и содержание дисциплины

3.1. Структура дисциплины

Для очной формы обучения

Вид учебной работы	Всего ЗЕ	Всего часов	Семестр(ы)
			7
ОБЩАЯ ТРУДОЕМКОСТЬ ДИСЦИПЛИНЫ	6	216	216
КОНТАКТНАЯ РАБОТА*	2,9	103	103
АУДИТОРНАЯ РАБОТА	1,9	68	68
Лекции	0,95	34	34
Практические (семинарские) занятия		-	-
Лабораторные работы	0,95	34	34
САМОСТОЯТЕЛЬНАЯ РАБОТА ОБУЧАЮЩЕГОСЯ	3,1	112	112
Проработка учебного материала	2,1	76	76
Курсовой проект		0	0
Курсовая работа	1	36	36
Подготовка к промежуточной аттестации	1	36	36
Промежуточная аттестация:			Э
			КР

Для заочной формы обучения

Вид учебной работы	Всего ЗЕ	Всего часов	Семестр(ы)
			5
ОБЩАЯ ТРУДОЕМКОСТЬ ДИСЦИПЛИНЫ	6	216	216
КОНТАКТНАЯ РАБОТА*	1,9	71	70
АУДИТОРНАЯ РАБОТА	0,8	30	30
Лекции	0,5	18	18
Практические (семинарские) занятия		-	-
Лабораторные работы	0,3	12	12
САМОСТОЯТЕЛЬНАЯ РАБОТА ОБУЧАЮЩЕГОСЯ	4,9	177	177
Проработка учебного материала	3,7	141	132
Курсовой проект		-	-
Курсовая работа	1	36	36
Подготовка к промежуточной аттестации	0,3	9	9
Промежуточная аттестация:			Э
			КР

3.2. Содержание дисциплины, структурированное по разделам и видам занятий

Для очной формы обучения

Разделы дисциплины	Всего часов	Распределение трудоемкости по видам учебной работы				Формы и вид контроля	Индексы индикаторов формируемых компетенций
		лекции	лаб. раб.	пр. зан.	сам. раб.		
Раздел 1. Инструменты математической статистики анализа данных.	18	4	4		10	ТК1	ПК-2.1, ПК-2.2
Раздел 2. Визуализация данных	8	2	2		4	ТК1	ПК-2.1, ПК-2.2
Раздел 3. Подготовка данных	8	2	2		4	ТК1	ПК-2.1, ПК-2.2
Раздел 4. Задачи регрессии	48	12	12		24	ТК2	ПК-2.1, ПК-2.2
Раздел 5. Моделирование временных рядов	18	4	4		10	ТК2	ПК-2.1, ПК-2.2
Раздел 6. Снижение размерности	8	2	2		4	ТК2	ПК-2.1, ПК-2.2
Раздел 7. Задачи классификации	18	4	4		10	ТК3	ПК-2.1, ПК-2.2
Раздел 8. Нейронные сети	18	4	4		10	ТК3	ПК-2.1, ПК-2.2
Курсовая работа	36				36	ОМкр	ПК-2.1, ПК-2.2
Экзамен	36				36	ОМ 3	ПК-2.1, ПК-2.2
ИТОГО	216	34	34		112		

Для заочной формы обучения

Разделы дисциплины	Всего часов	Распределение трудоемкости по видам учебной работы				Формы и вид контроля	Индексы индикаторов формируемых компетенций
		лекции	лаб. раб.	пр. зан.	сам. раб.		
Раздел 1. Инструменты математической статистики анализа данных.	14	2	1		11	ТК1	ПК-2.1, ПК-2.2
Раздел 2. Визуализация данных	14	2	1		11	ТК1	ПК-2.1, ПК-2.2

Раздел 3. Подготовка данных	20	2	2		16	TK1	ПК-2.1, ПК-2.2
Раздел 4. Задачи регрессии	53	4	2		47	TK2	ПК-2.1, ПК-2.2
Раздел 5. Моделирование временных рядов	20	2	2		16	TK2	ПК-2.1, ПК-2.2
Раздел 6. Снижение размерности	20	2	2		16	TK2	ПК-2.1, ПК-2.2
Раздел 7. Задачи классификации	20	2	2		16	TK3	ПК-2.1, ПК-2.2
Раздел 8. Нейронные сети	10	2			8	TK3	ПК-2.1, ПК-2.2
Курсовая работа	36				36	ОМкр	ПК-2.1, ПК-2.2
Экзамен	9				9	ОМ 3	ПК-2.1, ПК-2.2
ИТОГО	216	18	12		177		

3.3. Содержание дисциплины

Раздел 1. Инструменты математической статистики анализа данных

Тема 1.1. Точечное и интервальное оценивание характеристик случайных величин

Предмет, цели и задачи дисциплины. Основные научные принципы и методы обработки и анализа данных. Области применения. Основные источники информации, необходимые для профессиональной деятельности. Математические методы обработки и анализа цифровых данных.

Статистическая оценка. Точечное и интервальное оценивание. Свойства оценок. Точность. Надежность. Доверительный интервал.

Тема 1.2. Проверка статистических гипотез

Понятие статистической гипотезы. Критическая область. Критические точки. Ошибки первого и второго рода. Уровень значимости. Мощность критерия. Сравнение дисперсий двух генеральных совокупностей.

Раздел 2. Визуализация данных

Тема 2. Визуализация данных

Графическая визуализация данных с использованием специализированных графических пакетов и Python Notebook для профессиональной визуализации и анализа данных.

Раздел 3. Подготовка данных

Тема 3. Подготовка данных

Использование различных источников данных. Технология получения данных из баз данных с помощью SQL запросов. Предобработка данных. Очистка данных.

Раздел 4. Задачи регрессии

Тема 4.1. Линейная регрессия

Спецификация модели. Построение линейной регрессии методом наименьших квадратов. Оценка качества модели. Свойства, экономическая интерпретация и оценка параметров линейного уравнения регрессии.

Проверка гипотез о значимости регрессионной модели и проверка значимости ее параметров. Критерии Стьюдента и Фишера.

Интервалы прогноза по линейному уравнению регрессии. Построение доверительных интервалов для прогнозируемых значений. Стандартные ошибки коэффициентов регрессии. Средняя ошибка аппроксимации.

Метод градиентного спуска.

Тема 4.2. Нелинейная регрессия. Логистическая регрессия.

Нелинейная регрессия. Основные модели. Выбор зависимости. Линеаризация нелинейных зависимостей. Схема применения метода наименьших квадратов в нелинейных моделях.

Корреляция для нелинейной регрессии. Средняя ошибка аппроксимации.

Раздел 5. Моделирование временных рядов

Тема 5. Моделирование временных рядов

Понятие временного ряда. Основные элементы временного ряда. Автокорреляция уровней временного ряда и выявление его структуры. Коррелограмма.

Выбор вида зависимости. Анализ качества модели.

Моделирование сезонных колебаний. Элиминирование сезонной компоненты. Аддитивная модель. Моделирование сезонных колебаний с помощью мультипликативной модели. Применение фиктивных переменных для моделирования сезонных колебаний

Раздел 6. Снижение размерности

Тема 6. Методы снижения размерности

Задачи снижения размерности. Компонентный анализ. Метод главных компонент. Критерии применимости метода: критерий Кайзера–Мейера–Олкина, критерий Бартлетта. Выбор числа главных компонент: кумулятивная дисперсия, собственные значения главных компонент, график «каменистой осыпи». Метод ортогонального вращения варимакс. Матрица факторных нагрузок. Интерпретация главных компонент.

Раздел 7. Задачи классификации

Тема 7. Задачи классификации

Классификация методами бинарной и множественной регрессии.

Кластерный анализ. Задачи кластерного анализа. Меры расстояний. Алгоритмы кластеризации. Формы кластеров. Иерархическая кластеризация. Кластеризация k-средним.

Оценка качества кластеризации.

Раздел 8. Нейронные сети

Тема 8. Нейронные сети

Основные понятия нейронных сетей и типы сетей. Классы задач, решаемых с помощью нейронных сетей. Построение нейронной сети и обучение нейрона. Пример применения нейронные сети для решения задачи регрессии.

3.4. Тематический план практических занятий

Данный вид работы не предусмотрен учебным планом

3.5. Тематический план лабораторных работ

Тема 1.1. Точечное и интервальное оценивание характеристик случайных величин

Тема 1.2. Проверка статистических гипотез

Тема 2. Визуализация данных

Тема 3. Подготовка данных

Тема 4.1. Линейная регрессия. Отбор факторов, оценка значимости

Тема 4.2. Нелинейная регрессия. Модели бинарного выбора. Модель множественного выбора

Тема 5. Автокорреляционная функция и структура временного ряда

Тема 6. Метод главных компонент

Тема 7. Кластерный анализ

Тема 8. Нейронные сети

3.6. Курсовая работа

Примерный перечень тем для выполнения курсовой работы

1. Исследование зависимости доходов от инвестиций, полученных по внешнеэкономической деятельности от прямых инвестиций за границу

2. Исследование зависимости импорта товаров от экспорта товаров

3. Исследование зависимости импорта товаров и услуг от экспорта товаров и услуг

4. Исследование зависимости средней цены 1 кв. м общей площади квартир на рынке жилья от среднемесячной номинальной начисленной заработной платы

5. Исследование зависимости стоимости фиксированного набора потребительских товаров и услуг от средних потребительских цен на товары, реализуемые на розничных рынках

6. Исследование зависимости средних потребительских цен (тарифов) на товары и услуги от стоимости фиксированного набора потребительских товаров и услуг

7. Исследование зависимости квартальных индексов потребительских цен и ВВП в части оплата труда

8. Изучение зависимости числа преступлений от среднедушевых доходов населения

9. Исследование зависимости между валовой добавленной стоимостью в основных ценах и чистыми налогами на производство и импорт

10. Исследование зависимости задолженности по кредитам индивидуальных предпринимателей от объема выданных кредитов

11. Исследование зависимости между расходами на конечное потребление и средними потребительскими ценами

12. Исследование зависимости между изменением запасов материальных оборотных средств и финансовой деятельностью в структуре ВВП

13. Исследование взаимосвязи между объемом предоставленных кредитов и просроченной задолженностью

14. Исследование динамики рождаемости в зависимости от доходов населения

15. Анализ зависимости оборота розничной торговли всеми предприятиями и организациями от денежных доходов населения

Также студент может предложить собственную тему исследования при условии согласования с руководителем.

4. Оценивание результатов обучения

Оценивание результатов обучения по дисциплине осуществляется в рамках текущего контроля и промежуточной аттестации, проводимых по балльно-рейтинговой системе (БРС).

Шкала оценки результатов обучения по дисциплине:

Код компетенции	Код индикатора компетенции	Запланированные результаты обучения по дисциплине	Уровень сформированности компетенции(индикатора достижения компетенции)			
			Высокий	Средний	Ниже среднего	Низкий
			от 85 до 100	от 70 до 84	от 55 до 69	от 0 до 54
			Шкала оценивания			
			отлично	хорошо	удовлетворительно	неудовлетворительно
			зачтено		не зачтено	
ПК-2 Способен осуществлять проектно-аналитические работы с использованием технологий больших данных и искусственного	ПК-2.1 Способен планировать и организовывать аналитические работы с использованием технологий больших данных	знать:				
		методы организации аналитических расчетов при работе с большими данными	Отличный уровень знаний в объеме, соответствующем программе подготовки, без ошибок	Хороший уровень знаний в объеме, соответствующем программе подготовки, несколько несущественных негрубых ошибок	Удовлетворительный уровень знаний в объеме, соответствующем программе подготовки, много несущественных негрубых ошибок	Уровень знаний ниже предъявляемых минимальных требований, имеют место целый ряд грубых ошибок
		уметь:				
		планировать и организовывать аналитические расчеты при работе с большими данными	Продемонстрированы все основные умения, решены все основные задачи с отдельным	Продемонстрированы все основные умения, решены все основные задачи с негрубыми ошибками,	Продемонстрированы основные умения, решены типовые задачи с негрубыми ошибками, выполнены	При решении стандартных задач не продемонстрированы основные умения, имеют место грубые ошибки

Код компетенции	Код индикатора компетенции	Запланированные результаты обучения по дисциплине	Уровень сформированности компетенции(индикатора достижения компетенции)			
			Высокий	Средний	Ниже среднего	Низкий
			от 85 до 100	от 70 до 84	от 55 до 69	от 0 до 54
			Шкала оценивания			
			отлично	хорошо	удовлетворительно	неудовлетворительно
			зачтено			не зачтено
		и несущественными недочетами и, выполнены все задания в полном объеме	выполнены все задания в полном объеме, но некоторые с недочетами	все задания, но не в полном объеме		
владеть:						
		навыками применения инструментальных средств для осуществления анализа данных при работе с большими данными	Продемонстрированы навыки при решении нестандартных задач без ошибок и недочетов	Продемонстрированы базовые навыки при решении стандартных задач с некоторыми недочетами	Имеется минимальный набор навыков для решения стандартных задач с некоторыми недочетами	При решении стандартных задач не продемонстрированы базовые навыки, имеют место грубые ошибки
ПК-2.2 Выполняет полный цикл решения задач с помощью машинного обучения и продвинутой аналитики	знать:					
			- возможности алгоритмов машинного обучения и продвинутой аналитики; - классы задач, решаемых с помощью моделей машинного обучения и продвинутой аналитики;	Отличный уровень знаний в объеме, соответствующем программе подготовки и, без ошибок	Хороший уровень знаний в объеме, соответствующем программе подготовки, несколько несущественных негрубых ошибок	Удовлетворительный уровень знаний в объеме, соответствующем программе подготовки, много несущественных негрубых ошибок

Код компетенции	Код индикатора компетенции	Запланированные результаты обучения по дисциплине	Уровень сформированности компетенции(индикатора достижения компетенции)			
			Высокий	Средний	Ниже среднего	Низкий
			от 85 до 100	от 70 до 84	от 55 до 69	от 0 до 54
			Шкала оценивания			
			отлично	хорошо	удовлетворительно	неудовлетворительно
			зачтено			не зачтено
уметь:						
- применять на практике алгоритмы машинного обучения и продвинутой аналитики; - обосновывать применение выбранных алгоритмов для решения конкретных задач		Продемонстрированы все основные умения, решены все основные задачи с отдельным и несущественными недочетами, выполнены все задания в полном объеме	Продемонстрированы все основные умения, решены все основные задачи с негрубыми ошибками, выполнены все задания в полном объеме, но некоторые с недочетами	Продемонстрированы основные умения, решены типовые задачи с негрубыми ошибками, выполнены все задания, но не в полном объеме	При решении стандартных задач не продемонстрированы основные умения, имеют место грубые ошибки	
владеть:						
- базовыми технологиями анализа данных		Продемонстрированы навыки при решении нестандартных задач без ошибок и недочетов	Продемонстрированы базовые навыки при решении стандартных задач с некоторыми недочетами	Имеется минимальный набор навыков для решения стандартных задач с некоторыми недочетами	При решении стандартных задач не продемонстрированы базовые навыки, имеют место грубые ошибки	

Оценочные материалы для проведения текущего контроля и промежуточной аттестации приведены в Приложении к рабочей программе дисциплины.

Полный комплект заданий и материалов, необходимых для оценивания результатов обучения по дисциплине, хранится на кафедре разработчика.

5. Учебно-методическое и информационное обеспечение дисциплины

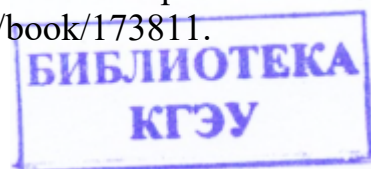
5.1. Учебно-методическое обеспечение

5.1.1. Основная литература

1. Плотников, А. Н. Элементарная теория анализа и статистическое моделирование временных рядов : учебное пособие для вузов / А. Н. Плотников. — 2-е изд., испр. и доп. — Санкт-Петербург : Лань, 2021. — 212 с. — ISBN 978-5-8114-7748-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/179030>.
2. Чураков, Е. П. Введение в многомерные статистические методы / Е. П. Чураков. — 3-е изд., стер. — Санкт-Петербург : Лань, 2023. — 148 с. — ISBN 978-5-507-47141-6. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/330530>.
3. Костромин, А. В., Эконометрика : учебное пособие / А. В. Костромин, Р. М. Кундакян. — Москва : КноРус, 2021. — 228 с. — ISBN 978-5-406-08124-2. — URL: <https://book.ru/book/939179>. — Текст : электронный.

5.1.2.Дополнительная литература

1. Как устроен Python. Гид для разработчиков, программистов и интересующихся : справочное издание / М. Харрисон. - СПб. : Питер, 2019. - 272 с. - URL: <https://ibooks.ru/bookshelf/359217>. - ISBN 978-5-4461-0906-7. - Текст : электронный.
2. Нейронные сети, генетические алгоритмы и нечеткие системы : научное издание / Д. Рутковская, М. Пилиньский, Л. Рутковский ; пер. с пол. И. Д. Рудинского. - 2-е изд. - М. : Горячая линия - Телеком, 2013. - 384 с. - URL: <https://ibooks.ru/bookshelf/334029>. - ISBN 978-5-9912-0320-3. - Текст : электронный.
3. Хливненко, Л. В. Практика нейросетевого моделирования : учебное пособие для вузов / Л. В. Хливненко, Ф. А. Пятакович. — 2-е изд., стер. — Санкт-Петербург : Лань, 2021. — 200 с. — ISBN 978-5-8114-8264-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/173811>.



5.2. Информационное обеспечение

5.2.1. Электронные и интернет-ресурсы

№ п/п	Наименование электронных и интернет-ресурсов	Ссылка
1	Электронно-библиотечная система «Лань»	https://e.lanbook.com/
2	Электронно-библиотечная система «ibooks.ru»	https://ibooks.ru/
3	Электронно-библиотечная система «book.ru»	https://www.book.ru/
4	Портал «Открытое образование»	https://npod.ru
5	Российская национальная библиотека	https://nlr.ru/
6	КиберЛенинка	https://cyberleninka.ru
7	Техническая библиотека	https://techlibrary.ru
8	Национальная электронная библиотека (НЭБ)	https://rusneb.ru/
9	Онлайн-курс “Прикладное программирование на языке Python”.	https://openedu.ru/course/urfu/PYAP
10	Real Python Tutorials	https://realpython.com
11	Jake VanderPlas. Python Data Science Handbook	https://jakevdp.github.io/PythonDataScienceHandbook
12	Jeroen Janssens. Data Science at the Command Line	https://www.datascienceatthecommandline.com/
13	Andrew Ng. MLOps: From Model-centric to Data-centric AI	https://www.deeplearning.ai/wp-content/uploads/2021/06/MLOps-From-Model-centric-to-Data-centric-AI.pdf
14	Spark SQL, DataFrames and Datasets Guide	https://spark.apache.org/docs/latest/sql-programming-guide.html
15	Гугл Академия	https://scholar.google.ru/

5.2.2. Профессиональные базы данных / Информационно-справочные системы

№ п/п	Наименование профессиональных баз данных	Адрес	Режим доступа
1	Официальный интернет-портал правовой информации	http://pravo.gov.ru	http://pravo.gov.ru
2	Справочная правовая система «Консультант Плюс»	http://consultant.ru	http://consultant.ru
3	Справочно-правовая система по законодательству РФ	http://garant.ru	http://garant.ru
4	Applied Science & Technology Source. EBSCO publishing	http://search.ebscohost.com	http://search.ebscohost.com
5	Wiley Online Library	http://onlinelibrary.wiley.com/	http://onlinelibrary.wiley.com/
6	Научная электронная библиотека	http://elibrary.ru	http://elibrary.ru

№ п/п	Наименование профессиональных баз данных	Адрес	Режим доступа
7	Российская государственная библиотека	http://www.rsl.ru	http://www.rsl.ru
8	Международная реферативная база данных научных изданий zbMATH	http://www.zbmath.org	http://www.zbmath.org
9	Международная реферативная база данных научных изданий Springerlink	http:// link.springer.com	http:// link.springer.com
10	Образовательный портал	http://www.ucheba.com	http://www.ucheba.com

5.2.3. Лицензионное и свободно распространяемое программное обеспечение дисциплины

№ п/п	Наименование программного обеспечения	Способ распространения (лицензионное/свободно)	Реквизиты подтверждающих документов
1	Операционная система Microsoft Windows 10	Пользовательская операционная система	Договор №133/2021 от 12.10.2021, лицензиар - ЗАО «Софт Лайн Трейд», тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно
3	Microsoft Office 2019	Пакет офисных приложений	Договор №133/2021 от 12.10.2021, лицензиар - ЗАО «Софт Лайн Трейд», тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно
4	LMS Moodle	ПО для эффективного онлайн-взаимодействия преподавателя и студента	Свободная лицензия, тип (вид) лицензии - неискл. право, срок действия лицензии - бессрочно.
5	Visual Studio Express	Инструмент создания приложений	https://visualstudio.microsoft.com/ru/vs/express/

6. Материально-техническое обеспечение дисциплины

Наименование вида учебной работы	Наименование учебной аудитории, специализированной лаборатории	Перечень необходимого оборудования и технических средств обучения
Лекции	Учебная аудитория для проведения занятий лекционного типа	Специализированная учебная мебель, технические средства обучения, служащие для представления учебной информации большой аудитории (мультимедийный

Наименование вида учебной работы	Наименование учебной аудитории, специализированной лаборатории	Перечень необходимого оборудования и технических средств обучения
		проектор, компьютер (ноутбук), экран), демонстрационное оборудование, учебно-наглядные пособия
Лабораторные работы	Учебная лаборатория программной инженерии, ауд. В-608	Специализированное лабораторное оборудование по профилю лаборатории программной инженерии, специализированная учебная мебель на 50 посадочных мест, 24 компьютера с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, мультимедийная доска, моноблок), необходимое лицензионное программное обеспечение
	Компьютерный класс, ауд. В-610	Специализированная учебная мебель на 42 посадочных места, 17 компьютеров с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, экран для проектора, моноблок), необходимое лицензионное программное обеспечение
	Учебная лаборатория информационной безопасности, ауд. В-615	Специализированное лабораторное оборудование по профилю лаборатории информационной безопасности, специализированная учебная мебель на 35 посадочных мест, 15 компьютеров с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, мультимедийная доска, моноблок), необходимое лицензионное программное обеспечение
	Компьютерный класс, ауд. В-617	Специализированная учебная мебель на 24 посадочных места, 21 компьютер с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, необходимое лицензионное программное обеспечение
	Компьютерный класс, ауд. В-619	Специализированная учебная мебель на 26 посадочных мест, 21 компьютер с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, необходимое лицензионное программное обеспечение

Наименование вида учебной работы	Наименование учебной аудитории, специализированной лаборатории	Перечень необходимого оборудования и технических средств обучения
	Компьютерный класс, ауд. В-621	Специализированная учебная мебель на 35 посадочных мест, 13 компьютеров с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, экран для проектора, моноблок), необходимое лицензионное программное обеспечение
	Учебная лаборатория реинжиниринга и управления бизнес-процессами, ауд. В-623	Специализированное лабораторное оборудование по профилю лаборатории реинжиниринга и управления бизнес-процессами, специализированная учебная мебель на 34 посадочных места, 13 компьютеров с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, мультимедийная доска, моноблок), необходимое лицензионное программное обеспечение
	Компьютерный класс, В-600	Специализированная учебная мебель на 30 посадочных мест, 30 компьютеров, компьютеров с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, технические средства обучения (мультимедийный проектор, ноутбук, экран), видеокамеры, необходимое лицензионное программное обеспечение
Самостоятельная работа	Компьютерный класс с возможностью выхода в Интернет и обеспечением доступа в ЭИОС В-600	Специализированная учебная мебель на 30 посадочных мест, 30 компьютеров, технические средства обучения (мультимедийный проектор, компьютер (ноутбук), экран), видеокамеры, программное обеспечение
	Читальный зал библиотеки	Специализированная мебель, компьютерная техника с возможностью выхода в Интернет и обеспечением доступа в ЭИОС, экран, мультимедийный проектор, программное обеспечение

7. Особенности организации образовательной деятельности для лиц с ограниченными возможностями здоровья и инвалидов

Лица с ограниченными возможностями здоровья (ОВЗ) и инвалиды имеют возможность беспрепятственно перемещаться из одного учебно-лабораторного корпуса в другой, подняться на все этажи учебно-лабораторных корпусов, заниматься в учебных и иных помещениях с учетом особенностей

психофизического развития и состояния здоровья.

Для обучения лиц с ОВЗ и инвалидов, имеющих нарушения опорно-двигательного аппарата, обеспечены условия беспрепятственного доступа во все учебные помещения. Информация о специальных условиях, созданных для обучающихся с ОВЗ и инвалидов, размещена на сайте университета www/kgeu.ru. Имеется возможность оказания технической помощи ассистентом, а также услуг сурдопереводчиков и тифлосурдопереводчиков.

Для адаптации к восприятию лицами с ОВЗ и инвалидами с нарушенным слухом справочного, учебного материала по дисциплине обеспечиваются следующие условия:

- для лучшей ориентации в аудитории, применяются сигналы оповещения о начале и конце занятия (слово «звонок» пишется на доске);
- внимание слабослышащего обучающегося привлекается педагогом жестом (на плечо кладется рука, осуществляется нерезкое похлопывание);
- разговаривая с обучающимся, педагогический работник смотрит на него, говорит ясно, короткими предложениями, обеспечивая возможность чтения по губам.

Компенсация затруднений речевого и интеллектуального развития слабослышащих обучающихся проводится путем:

- использования схем, диаграмм, рисунков, компьютерных презентаций с гиперссылками, комментирующими отдельные компоненты изображения;
- регулярного применения упражнений на графическое выделение существенных признаков предметов и явлений;
- обеспечения возможности для обучающегося получить адресную консультацию по электронной почте по мере необходимости.

Для адаптации к восприятию лицами с ОВЗ и инвалидами с нарушениями зрения справочного, учебного, просветительского материала, предусмотренного образовательной программой по выбранному направлению подготовки, обеспечиваются следующие условия:

- ведется адаптация официального сайта в сети Интернет с учетом особых потребностей инвалидов по зрению, обеспечивается наличие крупношрифтовой справочной информации о расписании учебных занятий;
- педагогический работник, его собеседник (при необходимости), присутствующие на занятии, представляются обучающимся, при этом каждый раз называется тот, к кому педагогический работник обращается;
- действия, жесты, перемещения педагогического работника коротко и ясно комментируются;
- печатная информация предоставляется крупным шрифтом (от 18 пунктов), тотально озвучивается;
- обеспечивается необходимый уровень освещенности помещений;
- предоставляется возможность использовать компьютеры во время занятий и право записи объяснений на диктофон (по желанию обучающихся).

Форма проведения текущей и промежуточной аттестации для обучающихся с ОВЗ и инвалидов определяется педагогическим работником в соответствии с учебным планом. При необходимости обучающемуся с ОВЗ,

инвалиду с учетом их индивидуальных психофизических особенностей дается возможность пройти промежуточную аттестацию устно, письменно на бумаге, письменно на компьютере, в форме тестирования и т.п., либо предоставляется дополнительное время для подготовки ответа.

8. Методические рекомендации для преподавателей по организации воспитательной работы с обучающимися.

Методическое обеспечение процесса воспитания обучающихся выступает одним из определяющих факторов высокого качества образования. Преподаватель вуза, демонстрируя высокий профессионализм, эрудицию, четкую гражданскую позицию, самодисциплину, творческий подход в решении профессиональных задач, в ходе образовательного процесса способствует формированию гармоничной личности.

При реализации дисциплины преподаватель может использовать следующие методы воспитательной работы:

- методы формирования сознания личности (беседа, диспут, внушение, инструктаж, контроль, объяснение, пример, самоконтроль, рассказ, совет, убеждение и др.);

- методы организации деятельности и формирования опыта поведения (задание, общественное мнение, педагогическое требование, поручение, приучение, создание воспитывающих ситуаций, тренинг, упражнение, и др.);

- методы мотивации деятельности и поведения (одобрение, поощрение социальной активности, порицание, создание ситуаций успеха, создание ситуаций для эмоционально-нравственных переживаний, соревнование и др.)

При реализации дисциплины преподаватель должен учитывать следующие направления воспитательной деятельности:

Гражданское и патриотическое воспитание:

- формирование у обучающихся целостного мировоззрения, российской идентичности, уважения к своей семье, обществу, государству, принятым в семье и обществе духовно-нравственным и социокультурным ценностям, к национальному, культурному и историческому наследию, формирование стремления к его сохранению и развитию;

- формирование у обучающихся активной гражданской позиции, основанной на традиционных культурных, духовных и нравственных ценностях российского общества, для повышения способности ответственно реализовывать свои конституционные права и обязанности;

- развитие правовой и политической культуры обучающихся, расширение конструктивного участия в принятии решений, затрагивающих их права и интересы, в том числе в различных формах самоорганизации, самоуправления, общественно-значимой деятельности;

- формирование мотивов, нравственных и смысловых установок личности, позволяющих противостоять экстремизму, ксенофобии, дискриминации по социальным, религиозным, расовым, национальным признакам, межэтнической и межконфессиональной нетерпимости, другим негативным социальным явлениям.

Духовно-нравственное воспитание:

- воспитание чувства достоинства, чести и честности, совестливости, уважения к родителям, учителям, людям старшего поколения;
- формирование принципов коллективизма и солидарности, духа милосердия и сострадания, привычки заботиться о людях, находящихся в трудной жизненной ситуации;
- формирование солидарности и чувства социальной ответственности по отношению к людям с ограниченными возможностями здоровья, преодоление психологических барьеров по отношению к людям с ограниченными возможностями;
- формирование эмоционально насыщенного и духовно возвышенного отношения к миру, способности и умения передавать другим свой эстетический опыт.

Культурно-просветительское воспитание:

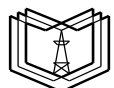
- формирование эстетической картины мира;
- формирование уважения к культурным ценностям родного города, края, страны;
- повышение познавательной активности обучающихся.

Научно-образовательное воспитание:

- формирование у обучающихся научного мировоззрения;
- формирование умения получать знания;
- формирование навыков анализа и синтеза информации, в том числе в профессиональной области.

Вносимые изменения и утверждения на новый учебный год

№ п/п	№ раздела внесения изменений	Дата внесения изменений	Содержание изменений	«Согласовано» Зав. каф. реализующей дисциплину	«Согласовано» председатель УМК института (факультета), в состав которого входит выпускающая кафедра)
1	2	3	4	5	6
1					
2					
3					



КГУ

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«КАЗАНСКИЙ ГОСУДАРСТВЕННЫЙ ЭНЕРГЕТИЧЕСКИЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «КГУ»)

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ
по дисциплине

Анализ больших данных

(Наименование дисциплины в соответствии с РУП)

Направление подготовки

09.03.03 Прикладная информатика

Квалификация

бакалавр

г. Казань, 2023

Оценочные материалы по дисциплине *Анализ больших данных* предназначены для оценивания результатов обучения на соответствие индикаторам достижения компетенций.

Оценивание результатов обучения по дисциплине осуществляется в рамках текущего контроля (ТК) и промежуточной аттестации, проводимых по балльно-рейтинговой системе (БРС).

1.Технологическая карта

Семестр 7

Наименование раздела	Формы и вид контроля	Рейтинговые показатели							
		I текущий контроль	Дополнительные баллы к ТК1	II текущий контроль	Дополнительные баллы к ТК2	III текущий контроль	Дополнительные баллы к ТК3	Итого	Промежуточная аттестация
Раздел 1. «Инструменты математической статистики анализа данных» Раздел 2. «Визуализация данных» Раздел 3. «Подготовка данных»	ТК1	15	0-15					15-30	15-30
Тест		7							
Защита лабораторной работы		4							
Отчет по лабораторной работе		4							
Раздел 4. «Задачи регрессии» Раздел 5. «Моделирование временных рядов» Раздел 6. «Снижение размерности»	ТК2			15	0-15			15-30	15-30
Тест				7					
Защита лабораторной работы				4					
Отчет по лабораторной работе				4					
Раздел 7. «Задачи классификации» Раздел 8. «Нейронные сети»	ТК3					25	0-15	25-40	25-40
Тест						7			
Защита лабораторной работы						4			
Отчет по лабораторной работе						14			
Промежуточная аттестация (зачет, экзамен, КП, КР)	ОМ								0-45

Задание промежуточной аттестации									0-15
В письменной форме по билетам									0-30

2. Оценочные материалы текущего контроля и промежуточной аттестации

Шкала оценки результатов обучения по дисциплине:

Код компетенции	Код индикатора компетенции	Запланированные результаты обучения по дисциплине	Уровень сформированности компетенции(индикатора достижения компетенции)			
			Высокий	Средний	Ниже среднего	Низкий
			от 85 до 100	от 70 до 84	от 55 до 69	от 0 до 54
			Шкала оценивания			
			отлично	хорошо	удовлетворительно	неудовлетворительно
			зачтено		не зачтено	
ПК-2 Способен осуществлять проектно-аналитические работы с использованием технологий больших данных и искусственного интеллекта	ПК-2.1 Способен планировать и организовывать аналитические работы с использованием технологий больших данных	знать:				
		методы организации аналитических расчетов при работе с большими данными	Отличный уровень знаний в объеме, соответствующем программе подготовки, без ошибок	Хороший уровень знаний в объеме, соответствующем программе подготовки, несколько несущественных негрубых ошибок	Удовлетворительный уровень знаний в объеме, соответствующем программе подготовки, много несущественных негрубых ошибок	Уровень знаний ниже предъявляемых минимальных требований, имеют место целый ряд грубых ошибок
		уметь:				
		планировать и организовывать аналитические расчеты при работе с большими данными	Продемонстрированы все основные умения, решены все основные задачи с отдельным и несущественными недочетами.	Продемонстрированы все основные умения, решены все основные задачи с негрубыми ошибками, выполнены все задания в полном объеме, но	Продемонстрированы основные умения, решены типовые задачи с негрубыми ошибками, выполнены все задания, но не в полном объеме	При решении стандартных задач не продемонстрированы основные умения, имеют место грубые ошибки

Код компетенции	Код индикатора компетенции	Запланированные результаты обучения по дисциплине	Уровень сформированности компетенции(индикатора достижения компетенции)			
			Высокий	Средний	Ниже среднего	Низкий
			от 85 до 100	от 70 до 84	от 55 до 69	от 0 до 54
			Шкала оценивания			
			отлично	хорошо	удовлетворительно	неудовлетворительно
		зачтено			не зачтено	
		выполнены все задания в полном объеме	некоторые с недочетами			
	владеть:					
	навыками применения инструментальных средств для осуществления анализа данных при работе с большими данными	Продemonстрированы навыки при решении нестандартных задач без ошибок и недочетов	Продemonстрированы базовые навыки при решении стандартных задач с некоторыми недочетами	Имеется минимальный набор навыков для решения стандартных задач с некоторыми недочетами	При решении стандартных задач не продемонстрированы базовые навыки, имеют место грубые ошибки	
ПК-2.2 Выполняет полный цикл решения задач с помощью машинного обучения и продвинутой аналитики	знать:					
	- возможности алгоритмов машинного обучения и продвинутой аналитики; - классы задач, решаемых с помощью моделей машинного обучения и продвинутой аналитики;	Отличный уровень знаний в объеме, соответствующем программе подготовки, без ошибок	Хороший уровень знаний в объеме, соответствующем программе подготовки, несколько несущественных негрубых ошибок	Удовлетворительный уровень знаний в объеме, соответствующем программе подготовки, много несущественных негрубых ошибок	Уровень знаний ниже предъявляемых минимальных требований, имеют место целый ряд грубых ошибок	
	уметь:					

Код компетенции	Код индикатора компетенции	Запланированные результаты обучения по дисциплине	Уровень сформированности компетенции(индикатора достижения компетенции)			
			Высокий	Средний	Ниже среднего	Низкий
			от 85 до 100	от 70 до 84	от 55 до 69	от 0 до 54
			Шкала оценивания			
			отлично	хорошо	удовлетворительно	неудовлетворительно
		зачтено			не зачтено	
		- применять на практике алгоритмы машинного обучения и продвинутой аналитики; - обосновывать применение выбранных алгоритмов для решения конкретных задач	Продемонстрированы все основные умения, решены все основные задачи с отдельным и несущественными недочетами, выполнены все задания в полном объеме	Продемонстрированы все основные умения, решены все основные задачи с негрубыми ошибками, выполнены все задания в полном объеме, но некоторые с недочетами	Продемонстрированы основные умения, решены типовые задачи с негрубыми ошибками, выполнены все задания, но не в полном объеме	При решении стандартных задач не продемонстрированы основные умения, имеют место грубые ошибки
		владеть:				
		- базовыми технологиями анализа данных	Продемонстрированы навыки при решении нестандартных задач без ошибок и недочетов	Продемонстрированы базовые навыки при решении стандартных задач с некоторыми недочетами	Имеется минимальный набор навыков для решения стандартных задач с некоторыми недочетами	При решении стандартных задач не продемонстрированы базовые навыки, имеют место грубые ошибки

Оценка «**отлично**» выставляется за выполнение *расчетных работ в семестре; тестовых заданий; глубокое понимание методов расчета, полные и содержательные ответы на вопросы билета.*

Оценка **«хорошо»** выставляется за выполнение *расчетных работ в семестре; тестовых заданий; понимание методов расчета, ответы на вопросы билета.*

Оценка **«удовлетворительно»** выставляется за выполнение *расчетных работ в семестре и тестовых заданий; понимание базовых методов расчета, ответы на вопросы билета с частичными ошибками.*

Оценка **«неудовлетворительно»** выставляется за *частичное выполнение расчетных работ в семестре и тестовых заданий; ответы на вопросы билета с ошибками.*

3. Перечень оценочных средств

Краткая характеристика оценочных средств, используемых при текущем контроле успеваемости и промежуточной аттестации обучающегося по дисциплине:

Наименование оценочного средства	Краткая характеристика оценочного средства	Описание оценочного средства
Контрольная работа (КнТР)	Средство проверки умений применять полученные знания для решения задач определенного типа по теме или разделу	Комплект контрольных заданий по вариантам
Курсовая работа (КР)	Конечный продукт, получаемый в результате планирования и выполнения комплекса учебных и исследовательских заданий. Позволяет оценить умения обучающихся самостоятельно конструировать свои знания в процессе решения практических задач и проблем, ориентироваться в информационном пространстве и уровень сформированности аналитических, исследовательских навыков, навыков практического и творческого мышления. Выполняется в индивидуальном порядке	Темы работ
Отчет по лабораторной работе (ОЛР)	Выполнение лабораторной работы, обработка результатов испытаний, измерений, эксперимента. Оформление отчета, защита результатов лабораторной работы по отчету	Перечень заданий и вопросов для защиты лабораторной работы, перечень требований к отчету
Тест (Тест)	Система стандартизированных заданий, позволяющая автоматизировать процедуру измерения уровня знаний и умений обучающегося	Комплект тестовых заданий

4. Перечень контрольных заданий или иные материалы, необходимые для оценки знаний, умений и навыков, характеризующих этапы формирования компетенций в процессе освоения дисциплины

Для текущего контроля ТК1:

Проверяемая компетенция: ПК2.1, ПК2.1

Комплект лабораторных работ

Для оценки сформированности компетенций используются следующие задания для выполнения лабораторных работ. Данные необходимо взять из файлов, размещенных на сетевом ресурсе.

Работа 1. Точечное и интервальное оценивание характеристик случайных величин

Цель работы: освоить методы расчета основных статистических характеристик.

Задание. Дана месячная заработная плата рабочих (тыс.руб.)

Данные взять из файла *rab1.xlsx* в соответствии с вариантом

MS Excel

1.1. По исходным данным:

Вычислить следующие характеристики:

- среднюю заработную плату;
- оценку дисперсии;
- оценку среднеквадратического отклонения.

Проверить. Для проверки использовать функции Excel (СРЗНАЧ, ДИСП.В, СТАНДОТКЛОН.В)

Сделать вывод о средней заработной плате рабочих предприятия.

1.2. На новом листе из исходных данных получить сгруппированные данные. По сгруппированным данным найти Среднюю, Дисперсию и Среднеквадратическое отклонение. Сравнить с результатами предыдущих расчетов.

1.3. Получить данные с интервальной группировкой. Для этого:

- Найти интервал разбиения (интервал группировки) $h = \frac{x_{\max} - x_{\min}}{n^*}$, где n^* определяется по формуле Стерджесса $n^* = [1 + 3.322 \lg n]$, n – объем выборки, а квадратные скобки обозначают округление до целого;
- определить границы интервалов $x_i' = x_{i-1}' + h$
- найти для каждого интервала сумму частот, попавшие в i –ый интервал.

По данным с интервальной группировкой найти Среднюю, Дисперсию и Среднеквадратическое отклонение. Сравнить с результатами предыдущих расчетов.

1.4. Определить значение t_n для нормального распределения с заданной надежностью, пользуясь функцией НОРМ.СТ.ОБР (Мастер функций, категория Статистические).

Рассчитать точность $\delta = t \cdot \frac{s}{\sqrt{n}}$.

1.5. Проверить найденное значение точности с помощью функции ДОВЕРИТ.НОРМ, которая также использует для определения t нормальное распределение. (Мастер функций, категория Статистические).

1.6. Найти границы доверительного интервала $\bar{x} - t \cdot \frac{s}{\sqrt{n}} < a < \bar{x} + t \cdot \frac{s}{\sqrt{n}}$.

Сделать выводы.

1.7. Определить значение t_s для распределения Стьюдента, пользуясь функцией СТЬЮДЕНТ.ОБР.2Х (Мастер функций, категория Статистические). Сравнить с величиной t_n . Повторить пункты 4-6. В п.5 для проверки использовать функцию ДОВЕРИТ.СТЬЮДЕНТ. Сделать выводы.

ППП Statistica

2.1. Запустить Statistica. Создать новую таблицу. Задать имя (X) и свойство поля для данных о зарплате.

2.2. Скопировать данные из файла Excel в таблицу Statistica.

2.3. Вывести данные описательной статистики и сравнить с п.1.1.

2.4. Получить сгруппированные данные. Сравнить с п.1.2.

2.5. Получить данные с интервальной группировкой. Сравнить с п.1.3.

Python

3.1. Запустить Python. Подключить библиотеку Pandas.

3.2. Получить данные из файла Excel с листа со своим вариантом с присвоением имени столбцу с данными

(справка: https://pandas.pydata.org/docs/reference/api/pandas.read_excel.html#)

3.3. Используя функции, рассчитать среднее, дисперсию и стандартное отклонение и сравнить с п.1.1.

(справка:

<https://pandas.pydata.org/docs/reference/api/pandas.Series.html?highlight=series#pandas.Series>)

3.4. Используя групповые операции, получить сгруппированные данные. Сравнить с п.1.2.

(справка:

<https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.groupby.html?highlight=groupby#pandas.DataFrame.groupby>)

3.5 По исходным данным получить интервальную группировку (параметр *bins* метода *value_counts*).

3.6 Рассчитать доверительный интервал

Работа 2. Проверка статистических гипотез

Цель работы: освоить методы проверки статистических гипотез.
Для каждой из задач, представленных ниже:

- сформулировать гипотеза H_0 и H_1
- записать критерий
- обосновать тип критической области
- найти расчетное и критическое значение критерия
- сформулировать вывод

Решение получить в Excel, ППП Statistica и Python

Задача 1.

Двумя методами проведены измерения температуры жидкости. Получены следующие данные:

x	9,6	10	9,8	10,2	10,6
y	10,4	9,7	10	10,3	

Определить, дают ли оба этих метода одинаковую точность (сравнить дисперсии). Уровень значимости α принять равным 0,1. В качестве альтернативной гипотезы принять $D(X) \neq D(Y)$

Задача 2. В результат длительного хронометража времени сборки узла различными сборщиками установлено, что дисперсия этого времени $\sigma_0^2 = 2$ мин². Результаты работы новичка таковы:

x_i	56	58	60	62	64
n_i	1	4	10	3	2

Можно ли при уровне значимости 0,05 считать, что новичок работает ритмично (разброс его времени существенно не отличается от остальных сборщиков)?

Работа 3. Визуализация данных

Цель работы: освоить методы построения графиков.

Работу делать на основе Работы 1!!!

Задание. Дана месячная заработная плата рабочих (тыс.руб.) (открыть файл предыдущего задания)

MS Excel

1.1. По сгруппированным данным (x_i, n_i) построить график зависимости частот от значения зарплаты рабочих x_i .

1.2. По данным с интервальной группировкой построить гистограмму зависимости частот от x_i'

Построить *гистограмму*, используя Сервис, *Анализ данных*, *Гистограмма*. В рабочем окне процедуры Гистограмма задать: Входной интервал – это исходный ряд данных (не сгруппированных!), Выходной интервал - любую свободную ячейку, не имеющую непустых ячеек справа и

снизу; поставить метку в окошке Вывод графика. Сравнить полученную гистограмму с п.2.

Построить еще одну гистограмму. В рабочем окне процедуры Гистограмма задать: Входной интервал (см. п. 3) и Интервал карманов — это первый столбец (левая граница интервалов) из п.1.3 работы 1. Сравнить полученную гистограмму с п.3.

1.3. Построить эмпирическую функцию распределения и ее график (в качестве Типа диаграммы использовать *точечная*).

ППП Statistica

2.1. По сгруппированным данным (x_i, n_i) построить точечный график (Scatterplot) зависимости частот от значения зарплаты рабочих x_i . (Подсказка: в качестве исходных данных взять частотное распределение Statistics→Basic Statistics→FrequencTable...)

2.2. По исходным данным построить гистограмму зависимости частот от x_i . Число интервалов взять такое же, как было по формуле Стерджиса.

2.3. Построить гистограмму эмпирической функции распределения. (Подсказка: в качестве исходных данных взять частотное распределение, столбец Cumulative Percent)

Python

3.1. По сгруппированным данным (x_i, n_i) построить график зависимости частот от значения зарплаты рабочих x_i .

3.2. По исходным данным построить гистограмму зависимости частот от x_i с интервальной группировкой. Число интервалов указать в соответствии с формулой Стерджиса.

3.3. Построить гистограмму эмпирической функции распределения.

Работа 4. Подготовка данных

Цель работы: освоить методы подготовки данных для дальнейшего анализа данных.

Расчеты провести в Excel и Python

1. Открыть файл log.csv. Ознакомиться с данными.
2. Посмотрите все уникальные значения в sample.csv
3. Посмотрите, сколько непустых значений в колонке City
4. Файл log.csv, в колонке user_id есть записи, которые содержат технические ошибки. Укажите, что записано в поле user_id в строках с ошибкой
5. Создайте новый датафрейм sample2, в который будут входить только записи о людях в возрасте меньше 30 лет
6. Создайте новый датафрейм log_win, в который будут входить только записи, где пользователь выиграл. Посчитайте, сколько таких записей, и сохраните в переменной win_count

7. Создайте новый датафрейм `sample2`, в который будут входить только записи о рабочих младше 30 лет
8. Найдите записи, где в городах нет буквы "о", и сохраните в переменную `sample4`. Не забудьте про NaN и параметр `na`
9. Сохраните в переменную `new_log` датафрейм, из которого удалены записи с ошибкой в поле `user_id`
10. С помощью `apply` и лямбда-функции увеличьте возраст во всех записях на 1 год и сохраните в `sample2`. В переменной `sample2` должен содержаться весь датафрейм `sample`
11. С помощью `apply` и `lambda`-функции замените все буквы в поле `City` на маленькие и сохраните в `sample2`. Вам может понадобиться функция `s.lower()`
12. Напишите функцию `profession_code`, которая на вход получает строку, а на выход возвращает:
 - 0 — если на вход поступила строка "Рабочий"
 - 1 — если на вход поступила строка "Менеджер"
 - 2 — в любом другом случае
13. Напишите функцию `age_category`, которая на вход получает число, а на выход отдаёт:
 - "молодой" — если возраст меньше 23
 - "средний" — если возраст от 23 до 35
 - "зрелый" — если возраст больше 35

Работа 5. Линейная регрессия

Цель работы: освоить методы расчета коэффициентов линейной регрессии .

MS Excel

1. Рассчитать параметры уравнения регрессии и записать полученное уравнение.
2. Рассчитать по построенной модели y_p .
3. Построить график зависимости y и y_p (должна получиться прямая!!).
4. Запустить Регрессионный анализ и сравнить полученные значения.
5. Оценить значимость уравнения, сделать вывод.
6. Оценить значимость параметров уравнения, сделать вывод.
7. Оценить качество уравнения по R^2 . Сделать вывод.
8. Рассчитать доверительные интервалы с надежностью 90% и 95% для параметров уравнения регрессии. Попадает ли в интервалы ноль? Сделать выводы.
9. Рассчитать доверительные интервалы для \hat{y} с надежностью 95%. Добавить на график п.3 границы доверительных интервалов. Убедиться, что не менее 95% значений y_i находятся внутри оцененных границ.

Statistica

Получиться все оценки через ППП Statistica

Python

Повторить все пункты, используя библиотеку `scikit-learn`

В отчет

1. Записать уравнения регрессии.
2. Записать, что показывает коэффициент b_1 .
3. Перерисовать (распечатать) график.
4. Оценить значимость уравнения (2 способа!), записать вывод.
5. Оценить значимость параметров уравнения (2 способа!), записать вывод.
6. Оценить качество уравнения по R^2 . Записать вывод.
7. Записать доверительные интервалы для надежности 90% и 95% для параметров уравнения регрессии. Записать выводы.
8. Зарисовать график с доверительными интервалами. Записать сколько процентов значений y_i оказалось внутри оцененных границ.

Контрольные вопросы

1. Парная регрессия
2. Множественная регрессия
3. Парная линейная регрессия
4. МНК
5. Формулы и смысл $SS_{\text{общ}}$, $SS_{\text{ост}}$, $SS_{\text{факт}}$, $S^2_{\text{общ}}$, $S^2_{\text{ост}}$, $S^2_{\text{факт}}$
6. Оценка значимости уравнения (включая формулировку гипотез)
7. Формулы m_{b0} , m_{b1} , m_y , t_b
8. Оценка значимости параметров уравнения (включая формулировку гипотез)
9. Коэффициент детерминации.
10. Доверительные интервалы для b_j и \hat{y} .

Работа 6. Отбор факторов, оценка значимости

Цель работы: освоить методы отбора факторов при построении модели линейной регрессии.

Задание 1.

1. По исходным данным из файла текущей работы сформировать матрицу X и вектор Y .
2. Найти оценку вектора B коэффициентов уравнения множественной линейной регрессии: $B = (X'X)^{-1}X'Y$.
3. Найти вектор E и $SS_{\text{ост}}$: $E = Y - XB$, $SS_{\text{ост}} = E^T \cdot E$
4. Найти стандартные ошибки коэффициентов регрессии: $m_b = \sqrt{S_b^2}$, где $S_b^2 = S_{\text{ост}}^2 (X'X)^{-1}$.
5. Рассчитать t-статистики для коэффициентов регрессии. Рассчитать p-значение для t-статистик.
6. Выполнить Анализ данных – Регрессия, сравнить результаты.

В отчет

9. Записать уравнение регрессии и t-статистики.

Задание 2.

1. Перейти к стандартизованным переменным.
2. Найти оценку вектора β коэффициентов уравнения множественной линейной регрессии в стандартизованных переменных, используя матричный метод.
3. Используя полученные β коэффициенты, найти коэффициенты b_j исходного уравнения. Сравнить с результатами из Задания 1.

В отчет

1. Записать уравнение регрессии в стандартизованных переменных и t-статистики.

Задание 3.

1. Построить частные уравнения регрессии
2. Рассчитать частные коэффициенты эластичности и их средние. Сделать выводы.
3. Рассчитать частные коэффициенты корреляции. Сделать выводы.
4. Рассчитать частные F-критерии.
5. Рассчитать t-критерии, используя частные F-критерии. Сравнить с результатами из Задания 1.

В отчет

1. Записать частные уравнение регрессии.
2. Записать средние частные коэффициенты эластичности и выводы.
3. Записать частные коэффициенты корреляции и выводы.
4. Записать частные F-критерии.
5. Записать t-критерии.

Задание 4.

Провести корреляционный анализ. Сделать выводы.

В отчет

Записать таблицу корреляционного анализа. записать выводы.

Задание 5.

Метод пошагового исключения

1. Определить параметры уравнения множественной линейной регрессии с включением всех факторов.
2. Оценить значимость параметров уравнения регрессии
3. Исключить наиболее незначимый фактор
4. Перестроить уравнение регрессии по оставшимся факторам.
5. Повторять пп 2-6 до тех пор, пока в уравнении не останутся только значимые факторы. Построить уравнение регрессии по значимым факторам (с учетом результатов корреляционного анализа).
6. Провести оценку значимости уравнения и параметров уравнения регрессии.

Метод пошагового включения

7. Построить уравнения линейной регрессии с одним фактором.
8. Оценить значимость уравнения регрессии

9. Добавить следующий фактор. Перестроить регрессию с дополнительным фактором.

10. Провести оценку значимости. Если включенные факторы значимы, то добавить следующий фактор. Если есть незначимый фактор, то его следует исключить. Повторять пп. 14-15 пока в уравнении не останутся только значимые факторы.

11. Построить уравнение регрессии по значимым факторам (с учетом результатов корреляционного анализа).

12. Провести оценку значимости уравнения и параметров уравнения регрессии.

Регрессия в стандартизованных переменных $t_{xy} = \frac{(y - \bar{y})}{\sigma_y} \cdot \frac{(x - \bar{x}_j)}{\sigma_{x_j}}$

13. Перейти к стандартизованным переменным.

14. Построить уравнения линейной регрессии в стандартизованных переменных.

15. Оценить значимость уравнения регрессии и параметров

16. Сделать выводы.

17. Получить расчетные значения $y_{\text{расч}}$.

18. Рассчитать $SS_{\text{общ}}$, $SS_{\text{ост}}$, $SS_{\text{факт}}$, $S^2_{\text{общ}}$, $S^2_{\text{ост}}$, $S^2_{\text{факт}}$

19. Рассчитать коэффициент множественной детерминации, коэффициент множественной корреляции, нормированный коэффициент множественной детерминации, F и $F_{\text{кр}}$. Сделать выводы.

20. Построить графики y_i (точками) и $y_{\text{расч}}$ (линией).

В отчет

1. Записать полученную последовательность шагов по методу пошагового исключения. На каждом шаге указать R^2 , нормированный R^2 , t критерий и его значимость для каждого коэффициента регрессии.

2. Записать полученную последовательность шагов по методу пошагового включения. На каждом шаге указать R^2 , нормированный R^2 , t критерий и его значимость для каждого коэффициента регрессии.

3. Записать итоговое уравнение регрессии, t критерии для каждого коэффициента, P значение, выводы о значимости.

4. Записать уравнение регрессии в стандартизованных переменных и сделать выводы.

5. Записать полученные $SS_{\text{общ}}$, $SS_{\text{ост}}$, $SS_{\text{факт}}$, $S^2_{\text{общ}}$, $S^2_{\text{ост}}$, $S^2_{\text{факт}}$.

6. Записать полученные: коэффициент множественной детерминации, коэффициент множественной корреляции, нормированный коэффициент множественной детерминации, F и $F_{\text{кр}}$. Сделать выводы.

7. Зарисовать график y_i (точками) и $y_{\text{расч}}$ (линией)..

1. По результатам текущей работы рассчитать ошибки в MS Excel

2. По результатам текущей работы рассчитать ошибки в Python

3. Сравнить

В отчет

8. Записать полученные значения ошибок.

Контрольные вопросы

1. Абсолютные и относительные ошибки.
2. Формулы расчета:
 - Средняя абсолютная ошибка (Mean Absolute Error)
 - Средняя квадратичная ошибка (Mean Squared Error)
 - Среднеквадратическая ошибка (Root Mean Squared Error)
 - Средняя абсолютная процентная ошибка, средняя ошибка аппроксимации (Mean Absolute Percentage Error)
 - Коэффициент детерминации

Контрольные вопросы

3. Множественная линейная регрессия.
4. Отбор факторов.
5. Корреляционный анализ
6. Коллинеарность и мультиколлинеарность факторов
7. Оценка значимости в множественной линейной регрессии.
8. Метод исключения
9. Метод включения
10. Коэффициент множественной детерминации, коэффициент множественной корреляции, нормированный коэффициент множественной детерминации.
11. Уравнение регрессии в стандартизованных переменных.
12. Формулы расчета вектора B , $SS_{\text{ост}}$, стандартных ошибок, t -статистик.
13. Формулы стандартизованных переменных, их связь с b коэффициентами.
14. Формулы частных коэффициентов корреляции и частных F -критериев.

Работа 7. Нелинейная регрессия

Цель работы: освоить методы линеаризации и оценки коэффициентов нелинейной регрессии .

Задание

10. Построить график (точечный) по исходным данным.
 11. Выбрать подходящие для моделирования зависимости.
- Для каждой из выбранных зависимостей*
12. Сделать линеаризацию, предварительно ввести новые переменные.
 13. По новым переменным найти параметры преобразованной модели.
 14. Найти параметры исходной нелинейной зависимости.
 15. Рассчитать $Y_{\text{расч}}$ (по нелинейной! модели).
 16. Найти индекс детерминации, индекс корреляции, среднюю ошибку аппроксимации.
 17. Построить графики исходных и расчетных данных (исходные – точками, расчетные – линия без маркеров).

18.Сделать вывод.

Метод средних точек

19.Для каждой из стандартных зависимостей рассчитать координаты средних точек.

20.Добавить средние точки на график п.1. Сделать вывод, какая из функций лучше будет описывать данные. Совпадает ли этот вывод с выводом п.9?

В отчет

10.Зарисовать графики исходных и расчетных данных.

11.Рядом с каждым графиком записать:

- уравнения регрессии,
- процесс линеаризации (сведения нелинейной зависимости к линейной),
- индекс детерминации,
- индекс корреляции,
- среднюю ошибку аппроксимации.

12.Зарисовать график со средними точками. Сделать вывод.

13.Записать общий вывод по работе с учетом п.2 и п.3.

Контрольные вопросы

15.Нелинейная регрессия

16.Классификация моделей нелинейной регрессии

17.В чем заключается линеаризация?

18.Индекс детерминации.

19.Индекс корреляции.

20.Средняя ошибка аппроксимации.

21.Метод средних точек.

Работа 8. Метод градиентного спуска

Цель работы: освоить метод градиентного спуска .

Задание 1. Разбиение выборки

1. Открыть файл с датасетом Iris:

```
from sklearn.datasets import load_iris
```

```
np=load_iris()
```

```
X=pd.DataFrame(np.data, columns=['sepal length (cm)',  
                                'sepal width (cm)',  
                                'petal length (cm)',  
                                'petal width (cm)'])
```

```
y=pd.DataFrame(np.target,columns=['sort'])
```

где sort – это

0 - 'setosa', 1 - 'versicolor', 2 - 'virginica'

2. Разбить выборку с помощью функции train_test_split на обучающую и тестовую без параметров. Вывести y_train.

3. Разбить выборку с помощью функции `train_test_split` на обучающую и тестовую с долей `train 0,6`. Вывести `y_train`.
4. Разбить выборку с помощью функции `train_test_split` на обучающую и тестовую с числом `test 20`. Вывести `y_test`.
5. Разбить выборку с помощью функции `train_test_split` на обучающую и тестовую с долей `train 0,1`. Обеспечить равное вхождение всех сортов ирисов. Вывести `y_train`.

Задание 2. Построение модели

1. Открыть файл с датасетом `boston`: `from sklearn.datasets import load_boston`
 2. Изучить признаки: какие имеет смысл включать в модель?
 3. Провести корреляционный анализ. Выявить коллинеарные факторы.
 4. Используя библиотеку `statsmodels` и метод пошагового исключения, отобрать факторы в модель (не забудьте про свободный член!!). Учесть результаты корреляционного анализа.
 5. Разделить выборку на тренировочную и тестовую случайным(!) образом.
 6. Обучить модель на тренировочной выборке.
 7. Рассчитать значение ошибок `MSE` и `RMSE` и R^2 для тренировочных данных:
 8. Рассчитать значение ошибок `MSE` и `RMSE` и R^2 для тестовых данных: Повторить шаги 5-7 несколько раз (отдельный код писать не надо – просто несколько раз запустить). Что можно сказать по поводу построенной модели?
 9. Какие факторы вошли в модель (словами, а не именами переменных)?
 10. Записать модель
 11. Дать экономическую интерпретацию коэффициентам регрессии.
- пп. 9-11 записать словами как комментарии (Markdown) в файл блокнота.

Задание 3. Градиентный спуск

12. Продолжить работу с датасетом `boston`
13. Подключить библиотеку `SGDRegressor`
14. Построить модель по исходным данным (без нормализации). Вывести коэффициенты. Сравнить результаты с Заданием 2.
15. Нормализовать факторы `X`.
16. Построить модель по нормализованным переменным.
17. Вывести коэффициенты. Вывести R^2 . Сравнить результаты с Заданием 2.
18. Рассчитать время обучения моделей `SGDRegressor`, `LinearRegression` и `statsmodels`. Какая в данном случае работает быстрее?

Работа 9. Модели бинарного выбора

Цель работы: освоить методику расчета с использованием модели бинарного выбора .

1. Загрузить исходные данные в специализированный ППП.

2. Получить оценки параметров модели бинарного выбора с помощью логит-модели.
3. Получить оценки параметров модели бинарного выбора с помощью пробит-модели.
4. По полученным результатам выбрать подходящую модель.
5. В MS Excel получить расчетные значения. Оценить точность модели.
6. По построенной модели получить прогнозные значения для заданных условий.
7. Построить модель в Python и сравнить

В отчет

1. Записать полученную логит-модель.
2. Записать полученную пробит-модель.
3. Сделать вывод.
4. Записать результаты прогнозирования.

Контрольные вопросы

1. Модели бинарного выбора.
2. Свойства функции моделирования вероятности.
3. Логит-модель.
4. Пробит-модель.
5. Оценка качества модели.
6. Оценка значимости коэффициентов модели.

Работа 10. Модель множественного выбора

Цель работы: освоить методику расчета с использованием модели бинарного выбора .

1. Загрузить исходные данные в специализированный ППП.
2. Получить оценки параметров модели множественного выбора с помощью логит-модели.
3. В MS Excel получить расчетные значения. Оценить точность модели.
4. По построенной модели получить прогнозные значения для заданных условий.

В отчет

1. Записать полученную логит-модель.
2. Оценить качество и значимость параметров модели. Сделать вывод.
3. Записать результаты прогнозирования.

Контрольные вопросы

1. Модели множественного выбора.
2. Логит-модель множественного выбора.
3. Пробит-модель множественного выбора.
4. Оценка качества модели.
5. Оценка значимости коэффициентов модели.
6. Предельные эффекты.

Работа 11. Автокорреляционная функция и структура временного ряда

Цель работы: освоить методику расчета с использованием модели бинарного выбора .

Исходные данные представлены на трех рабочих листах Excel. Для каждого набора данных выполнить следующие задания.

1. Считать данные из файла.
2. Построить график. Сделать предположения о структуре временного ряда.
3. Рассчитать значения автокорреляционной функции.
4. Построить коррелограмму. На основании графика и коррелограммы сделать вывод о структуре временного ряда.

В отчет

5. Зарисовать полученный график временного ряда.
6. Зарисовать коррелограмму.
7. Записать вывод.

Контрольные вопросы

8. Структура временного ряда.
9. Понятия лага и порядка коэффициента автокорреляции.
10. Понятие автокорреляционной функции и коррелограммы.
11. До какого порядка следует рассчитывать автокорреляционную функцию.
12. Связь коррелограммы и структуры временного ряда.

Работа 12. Модели ARIMA

1. Считать данные из файла.
2. Построить график.
3. При необходимости убрать зависимость от времени
4. Найти мат. ожидание и дисперсию процесса.
5. Найти частные коэффициенты автокорреляции r_k^{part} .
6. Найти коэффициенты автокорреляции r_k .
7. Построить автокорреляционную и частную автокорреляционную функции.
8. Сделать предположения о порядке процесса p и q .
9. Рассчитать параметры модели
10. Провести оценку модели.
11. Подобрать оптимальную модель, изменяя p и q .
12. Записать уравнение. Построить график.

В отчет (для каждого из процессов)

1. Записать уравнения процесса для различных сочетаний p и q .
2. Для каждой модели сравнить критерии Шварца и Акайке. Сделать выводы.

Контрольные вопросы

1. Идентификация стационарной модели ARMA.
2. Q-статистика Лjung – Бокса.

3. Критерий Шварца.
4. Критерий Акайке.
5. Оценивание коэффициентов модели.
6. Диагностика оцененной модели.

Работа 13. Методы снижения размерности

Присоединитесь к соревнованию на платформе Kaggle.com, ознакомьтесь с описанием данных.

1. Проведите первичный анализ данных соревнования
2. Выполните предварительную обработку данных
3. Используя метод главных компонент, убедитесь в целесообразности снижения размерности, выберите оптимальное количество главных компонент
4. Рассчитать главные компоненты и дать им характеристику.
5. Используя полученные главные компоненты, провести иерархическую кластеризацию и построить дендрограмму. Определить оптимальное количество кластеров.
6. Используя имеющиеся в Scikit-Learn методы кластеризации, классифицировать объекты различными методами.
7. Проанализировать результаты

Работа 14. Задачи классификации

Присоединитесь к соревнованию на платформе Kaggle.com, ознакомьтесь с описанием данных.

1. Проведите первичный анализ данных соревнования
2. Выполните предварительную обработку данных
3. Построить модель логистической регрессии. Учтите, что в модели должны остаться только значимые факторы.
4. Какие факторы вошли в модель (словами, а не именами переменных)?
5. Записать модель
6. Получить расчетные значения
7. Проанализировать метрики качества модели
8. Дать экономическую интерпретацию коэффициентам регрессии.
9. Проанализировать результаты

Работа 15. Построение нейронной сети и обучение нейрона

1. Дан вектор входных данных и вектор весов нейрона. Применить сумматорную функцию и ввести верный ответ.
 $X = [15, 1, 25, -18, -12, -12, -9, 4, -22, 2]$
 $w = [0.1, -0.4, -0.9, -1.0, -0.7, -1.6, -0.8, -1.2, -1.2, -0.5]$
2. На вход функции активации Хэвисайда подано число 5. Известно, что ее смещение (bias) равно 10. Функция активации вернет...
3. На вход сигмоидальной функции активации пришло число -0,89. Функция вернет...

4. На вход функции активации гиперболический тангенс пришло число -0,89. Функция вернет...
5. На вход функции активации ReLU пришло число -0,89. Функция вернет...

Работа 16. Нейронные сети для решения задач регрессии

Присоединитесь к соревнованию на платформе Kaggle.com, ознакомьтесь с описанием данных.

Скопировав базовое решение и запустив его у себя вы сможете достигнуть значения метрики качества $MAE=12465$. Ваша задача изменяя любые параметры архитектуры нейронной сети или параметры обучения нейронной сети достигнуть наибольшего качества модели. Результат своих экспериментов вы сможете наблюдать на вкладке «Leaderboard» сразу после загрузки ответа во вкладке «My Submissions». Ваша цель достигнуть качества модели на тестовых данных $MAE<37000$. Тогда вы проходите задание. Когда используя LeaderBoard убедитесь, что результат вас устраивает, скачайте файл своего решения, который вы можете увидеть открыв свой рабочий блокнот в режиме просмотра (но не редактирования) и перейдя справа во вкладку «Output». Скачанный файл в формате «.csv» загрузите сюда, на платформу курса.

Работа 17. Нейронные сети для решения задач классификации

Присоединитесь к соревнованию на платформе Kaggle.com, ознакомьтесь с описанием данных.

Скопировав базовое решение и запустив его у себя вы сможете достигнуть значения метрики качества $accuracy=0.85$. Ваша задача изменяя любые параметры архитектуры нейронной сети или параметры обучения нейронной сети достигнуть наибольшего качества модели. Результат своих экспериментов вы сможете наблюдать на вкладке «Leaderboard» сразу после загрузки ответа во вкладке «My Submissions». Ваша цель достигнуть качества модели на тестовых данных $MAE>0.9$. Тогда вы проходите задание. Когда используя LeaderBoard убедитесь, что результат вас устраивает, скачайте файл своего решения, который вы можете увидеть открыв свой рабочий блокнот в режиме просмотра (но не редактирования) и перейдя справа во вкладку «Output». Скачанный файл в формате «.csv» загрузите сюда, на платформу курса.

Работа 18. Нейронные сети для решения задач обработки изображений

Присоединитесь к соревнованию на платформе Kaggle.com, ознакомьтесь с описанием данных.

Применить к размещенному там набору данных архитектуры нейронных сетей, которые указаны ниже. Выбрать архитектуру, которая дает лучшее качество решения задачи. Вписать точность, которую дала каждая архитектура (с точностью до 3 знаков после точки). Разделителем целой части числа от дробной является точка.

VGG16=...
VGG19=...
ResNet50=...
InceptionV3=...
DenseNet201=...
Xception=...

При оценке отчетов по лабораторным работам учитываются следующие критерии:

1. Знание теоретического материала
2. Выполнение самостоятельных заданий
3. Ответы на вопросы
4. Отчет о выполненной работе

Критериями оценки выполнения задания, согласно достигнутого уровня, являются: Высокий уровень. Знания теоретического материала, правильно выполнены все задания в соответствии с требованиями, полные ответы на вопросы, правильно выполнены домашние задания, своевременно предоставлен отчет о выполнении работы - 4 балла. Средний уровень. Теоретический материал знает, правильно выполнены все задания, ответы на вопросы не полные, домашние задания выполнены не в полном объеме, предоставлен отчет о выполнении работы, либо в случае несвоевременного предоставления отчета или с наличием несущественных ошибок в выполнении лабораторных заданиях - 3 балла. Ниже среднего уровень. Выполнено не все, но более 50% заданий лабораторной работы, домашнее задание не выполнены, несвоеременно предоставлен отчет о выполнении работы - 2 балла. Низкий уровень. Выполнено менее 50% лабораторной работы, не выполнено домашнее задание, отчет о выполнении работы не предоставлен – 1 балл. Количество баллов за Отчёт по лабораторной работе: минимум – 1 б. Количество баллов за Отчёт по лабораторной работе (с учетом коэффициента сложности): максимум – 4 б.

Для текущего контроля ТК2:

Проверяемая компетенция: ПК2.1, ПК2.1

Задания для контрольной работы

Задание

Исходные данные взять из файла в соответствии с вариантом. Задание сделать в Notebook Python. Формулы / модель написать на бумаге. Выводы писать в комментариях в файле.

Типовой вариант

1. Построить модель множественной линейной регрессии с включением всех факторов. Провести оценку параметров и оценку значимости. Сделать выводы.

2. Построить уравнение регрессии в стандартизованных переменных. Сделать выводы.

3. Провести корреляционный анализ. Выявить пары коллинеарных факторов. Проверить наличие мультиколлинеарности. Сделать выводы.

4. Выбрать подходящий метод последовательного включения или исключения факторов. Построить модель множественной линейной регрессии с учетом выводов, сделанных по предыдущим пунктам.

Критерии оценки.

Максимальная оценка составляет 15 баллов. Каждое задание оценивается максимум в 4 баллов. Минимальное количество баллов за работу – 8.

Для текущего контроля ТКЗ:

Проверяемая компетенция: ПК2.1, ПК2.1

Тест

Примерный вариант

Линейный коэффициент корреляции предназначен для измерения...

Выберите один ответ:

- ☐ a. гомоскедастичности
- ☐ b. значимости уравнения
- ☒ c. тесноты связи между факторами ✓
- ☐ d. значимости факторов
- ☐ e. гетероскедастичности

Дан вариационный ряд, где x_i – варианты, n_i – частоты, w_i – относительные частоты, n – объем совокупности. \bar{x} – среднее

Отметить формулу (формулы) для расчета дисперсии

Выберите один или несколько ответов:

- ☐ a. $\frac{\sum x_i \cdot n_i}{n}$
- ☐ b. $\frac{\sum (x_i - \bar{x})^2 \cdot w_i}{n}$
- ☒ c. $\frac{\sum (x_i - \bar{x})^2}{n}$ ✓
- ☐ d. $\sum (x_i - \bar{x})^2 \cdot w_i$
- ☐ e. $\sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$

В таблице представлена выборка. Отметить вариант с исходными данными

x_i	1	2	3
n_i	2	3	5

Выберите один ответ:

- ☒ a.
3, 1, 3, 1, 2, 3, 2, 2, 3, 1 ✕
- ☐ b.
3, 3, 2, 3, 3, 2, 1, 1, 2, 3
- ☐ c.
2, 3, 2, 2, 1, 2, 2, 3, 1, 1
- ☐ d.
2, 1, 2, 2, 1, 1, 3, 3, 1, 2

Коэффициент эластичности - это...

Выберите один или несколько ответов:

- ☐ a. $\frac{1}{n} \cdot \sum \left| \frac{y - y_p}{y} \right| \cdot 100$
- ☐ b. $f'(x) \frac{x}{y}$
- ☒ c. $\sqrt{1 - \frac{SS_{ост}}{SS_{общ}}} \times$
- ☐ d. $\frac{100}{\bar{y}} \cdot \sqrt{\frac{\sum (y - y_p)^2}{n}}$
- ☐ e. $\frac{R^2}{1 - R^2} \cdot \frac{n - m}{m - 1}$
- ☐ f. $\frac{100}{\bar{y}} \cdot \frac{\sum (y - y_p)^2}{n}$

Общая дисперсия рассчитывается по формуле

n - объем выборки

m - число параметров уравнения регрессии

Выберите один ответ:

- ☐ a. $\sum (y_i - \hat{y}_i)^2$
- ☐ b. $\sum (y_i - \bar{y})^2$
- ☐ c. $\frac{\sum (y_i - \hat{y}_i)^2}{n - m}$
- ☐ d. $\sum (\hat{y}_i - \bar{y})^2$
- ☒ e. $\frac{\sum (y_i - \bar{y})^2}{n - 1} \checkmark$
- ☐ f. $\frac{\sum (\hat{y}_i - \bar{y})^2}{m - 1}$

По результатам регрессионного анализа отметить верные утверждения

Регрессионная статистика					
Множественный R	0,940				
R-квадрат	0,884				
Нормированный R-квадрат	0,826				
Стандартная ошибка	43,227				
Наблюдения	16				
Дисперсионный анализ					
	df	SS	MS	F	Значимость F
Регрессия	5	142693	28539	15,27315	0,000210
Остаток	10	18586	1859		
Итого	15	161379			
	Коэффициенты	Станд. ошибка	t-статистика	P-Значение	
Y-пересечение	-3008,975	1136,175	-2,648	0,024	
x1	-6,674	5,383	-1,240	0,243	
x2	6,678	3,153	2,118	0,0001	
x3	30,664	11,975	2,561	0,028	
x4	13,949	15,157	0,920	0,379	
x5	-10,216	17,045	-0,599	0,562	

Выберите один или несколько ответов:

- ☐ a. факторная дисперсия 142693
- ☒ b. объем выборки 14 ✕
- ☒ c. факторная дисперсия 15,27 ✕
- ☐ d. факторная дисперсия 28539
- ☐ e. фактор x3 значим
- ☒ f. фактор x4 значим ✕

Для чего используется метод наименьших квадратов?

Выберите один ответ:

- ☒ a. для оценки значимости уравнения регрессии ✕
- ☐ b. для оценки качества модели
- ☐ c. для нахождения y расчетного
- ☐ d. для оценки значимости параметров уравнения регрессии
- ☐ e. для нахождения числа параметров уравнения регрессии
- ☐ f. для нахождения параметров уравнения регрессии

Каков порядок частного коэффициента корреляции?

$$r_{yx1,x2,x3,x6}$$

Ответ ввести цифрой.

Свойства стандартизованных переменных:

Выберите один или несколько ответов:

- ☒ a. $\sigma[t] = 1$ ✓
- ☐ b. $M[t] = 1$
- ☐ c. $D[t] = 0$
- ☒ d. $\sigma[t_1] = \sigma[t_{30}] + \sum \sigma[t_{3i}]$ ✕
- ☐ e. $M[t] = 0$

Получено уравнение регрессии:

$$y = 60 - 3 \cdot x$$

где x - число автобусов на маршруте

y - время ожидания автобуса, мин

Сколько автобусов можно снять с маршрута, чтобы время ожидания увеличилось не более, чем на 3 мин?

Ответ ввести числом.

Отметить подходящее линеаризующее преобразование

$$y = b_0 \cdot e^{b_1 x}$$

Выберите один ответ:

- ☒ a. $\ln y = b_0 + b_1 x$ ✕
- ☐ b. $\ln y = \ln b_0 + b_1 \ln x$
- ☐ c. $y = b_0 + b_1 x_1$
- ☐ d. $\ln y = \ln b_0 + b_1 x$
- ☐ e. $y = b_0 + b_1 \frac{1}{x}$
- ☐ f. $y = b_0 + b_1 x_1 + b_2 x_2$

Ранговый коэффициент корреляции Спирмена может изменяться в интервале

Выберите один ответ:

- ☐ a. должен быть больше t критического
- ☐ b. должен быть меньше t критического
- ☐ c. от 0 до 1
- ☐ d. от 0 до бесконечности
- ☒ e. от -1 до 1 ✓

В множественной линейной регрессии каким должно быть соотношение между объемом выборки n и числом факторов m ?

Выберите один ответ:

- ☐ a. $n > m$
- ☒ b. $n > m + 1$ ✕
- ☐ c. $n < m$
- ☐ d. $n > 2 \cdot m$
- ☐ e. $n > 5 \cdot m$

В модели бинарного выбора можно оценить...

Выберите один ответ:

- ☐ a. математическое ожидание $M(Y)$ где $Y=\{-1;0;+1\}$
- ☒ b. вероятность $P(Y=1)$ где $Y=\{-1;+1\}$ ✓
- ☐ c. вероятность $P(Y=0)$ где $Y=\{-1;0;+1\}$
- ☐ d. математическое ожидание $M(Y)$ где $Y=\{-1;+1\}$
- ☐ e. вероятность $P(Y=1)$ где $Y=\{-1;0;+1\}$

В результате оценки параметров множественной **logit-модели** в ППП Statistica была получена следующая таблица

y - Parameter estimates (множеств.sta)									
Distribution : MULTINOMIAL									
Link function: LOGIT									
Effect	Level of Effect	Level of Response	Column	Estimate	Standard Error	Wald Stat.	Lower CL 95, %	Upper CL 95, %	p
Intercept 1		0	1	5,35856	3,473264	2,380242	-1,44891	12,16604	0,122878
x2		0	2	-0,08802	0,942748	0,008717	-1,93577	1,75973	0,925612
x3		0	3	-2,37098	1,095301	4,685864	-4,51773	-0,22423	0,030412
Intercept 2		1	4	5,84634	3,381311	2,989492	-0,78091	12,47358	0,083806
x2		1	5	-0,63935	0,903627	0,500603	-2,41042	1,13173	0,479235
x3		1	6	-1,93432	1,016099	3,623960	-3,92583	0,05720	0,056953
Scale				1,00000	0,000000		1,00000	1,00000	

Рассчитать значение линейной формы $u(y_i=0)$ для $x_2=0$ и $x_3=0$.

Ответ округлить до сотых, ввести через точку.

Пример: 0.24

Пример: 0.40

Значение линейного коэффициента корреляции $r(x,y)=0,095$

Выберите один ответ:

- ☐ a. связь логарифмическая
- ☐ b. связь обратная
- ☐ c. существует линейная зависимость от времени
- ☐ d. связь статистически значима
- ☐ e. связь статистически не значима
- ☒ f. связь полиномиальная ✗
- ☐ g. возможна нелинейная зависимость от времени

Дана выборка. Рассчитать значение линейного коэффициента корреляции. Среднеквадратические отклонения принять

$S_x=1,6$

$S_y=2,6$

Ответ округлить до двух знаков после запятой.

x	0	-2	2	0
y	-2	0	2	4

В результате построения модели множественного выбора была получена следующая матрица ошибок

Observed	Predicted 0	Predicted 1	Predicted 2
0	6	1	1
1	3	4	1
2	1	1	3

Чему равняется точность в категории 2?

Ответ округлить до сотых, ввести через точку, ответ вводить НЕ в процентах!

Пример: 0.24

Пример: 0.40

Пусть выборка задана рядом распределения относительных частот признака X.

Вычислите среднее арифметическое выборки

x_i	1	2	4	7
w_i	0,4	0,4	0,1	0,1

К методам снижения размерности относятся...

Выберите один или несколько ответов:

- ☒ a. метод главных компонент ✓
- ☐ b. регрессионный анализ
- ☐ c. многофакторный дисперсионный анализ
- ☐ d. кластерный анализ
- ☒ e. дискриминантный анализ ✓

Имеется следующий ряд данных.

0
1
0
-1
-1
0
2
2

Определить ранг для числа "0"

Для контроля ОМкр:

Проверяемая компетенция: ПК2.1, ПК2.1

Требования к выполнению курсовой работы

Для оценки сформированности компетенций используется курсовая работа.

Целью курсовой работы является систематизация теоретических знаний и практических навыков в области компьютерного анализа данных, а также оценка уровня полученных знаний и степени подготовленности студента к использованию полученных знаний в будущей профессиональной деятельности.

Темы курсовых работ связаны с основными разделами изучаемой дисциплины и по характеру выполнения являются индивидуальными. Выбранную тему из предлагаемого и утвержденного руководителем списка курсовых работ студент должен зарегистрировать с указанием даты получения темы. С этого момента студент несет ответственность за качество и своевременность выполнения курсовой работы. Контроль за ходом выполнения работы, а также необходимые консультации по ходу выполнения осуществляет руководитель.

Курсовая работа обязательно включает в себя следующие структурные элементы:

- титульный лист
- содержание
- введение
- разделы (главы)
- заключение
- список использованной литературы
- приложения (при необходимости)

Содержание должно включать в себя перечень разделов и подразделов, приложений с указанием номеров страниц, на которых они помещены.

Во введении указывают область знаний, которой посвящена работа, перечень исследуемых проблем или задач и их актуальность. Обязательным во введении является формулирование цели и задач курсовой работы и логики изложения материала.

Основная часть работы должна состоять из двух глав:

в первой главе следует дать логическое и обзорное изложение исследуемой проблемы с рассмотрением различных теоретических концепций и подходов к исследованиям по заданной тематике, существующих методов решения задач в рассматриваемой области;

во второй главе следует выполнить математическую формализацию и самостоятельный расчет по поставленной задаче с последующим анализом полученного решения. Расчеты должны быть выполнены с применением одного из Notebook Python.

В заключении курсовой работы должны быть сделаны необходимые обобщения и выводы по проведенному исследованию.

Список используемой литературы должен содержать не менее 15 наименований, оформленных по соответствующим ГОСТам.

Примерный перечень тем для выполнения курсовой работы

1. Использование модели линейной регрессии для выявления влияния факторов в *исследуемой предметной области*
2. Построение нелинейной регрессии для выявления влияния факторов в *исследуемой предметной области*
3. Сравнение модификаций метода градиентного спуска

4. Использование модели бинарного выбора для классификации объектов в *исследуемой предметной области*
5. Использование модели множественного выбора для классификации объектов в *исследуемой предметной области*
6. Моделирование динамики *ряда* с помощью модели ARIMA
7. Методы снижения размерности
8. Задачи кластеризации
9. Использование нейронной сети в *исследуемой предметной области*

Также студент может предложить собственную тему исследования при условии согласования с руководителем.

Шкала оценивания курсовой работы

Оценка	Характеристика результатов
Отлично	<p>В курсовой работе дано развернутое обоснование актуальности, целей и задач тематики исследования – 9-10 баллов</p> <p>Содержание работы четко структурировано, материал изложен последовательно и логично – 9-10 баллов</p> <p>Работа имеет высокую степень проработанности теоретического материала (полный и актуальный библиографический список, актуальные ссылки на используемый материал в ходе изложения теоретической части) – 9-10 баллов</p> <p>Работа имеет глубокое теоретическое обоснование и обзор соответствующего терминологического аппарата по выбранной тематике. Теоретическая глава структурирована по подразделам. Приведен обзор накопленного практического опыта решения проблем по выбранной тематике. – 21-25 баллов</p> <p>Практическая часть работы выполнена полностью: приведено полное структурирование изучаемой проблемной ситуации, дано расширенное обоснование математической модели и методики решения, проведен глубокий и критический анализ результатов моделирования – 21-25 баллов</p> <p>В процессе защиты курсовой работы обучаемый показал всестороннее, систематическое и глубокое знание учебно-программного материала, усвоил основную и дополнительную литературу, рекомендованную программой. Дал полные развернутые ответы на поставленные вопросы. При ответе могут быть допущены незначительные недочеты в определении понятий, оговорки, поправленные студентом самостоятельно в процессе ответа – 19-20 баллов</p>
Хорошо	<p>В курсовой работе приведено обоснование актуальности, целей и задач тематики исследования – 7-9 баллов</p>

Оценка	Характеристика результатов
	<p>Содержание работы структурировано, материал изложен в основном последовательно и логично – 7-9 баллов</p> <p>Работа имеет достаточно полную степень проработанности теоретического материала (полный библиографический список, ссылки на используемый материал в ходе изложения теоретической части) – 7-9 баллов</p> <p>Работа имеет достаточно полное теоретическое обоснование и обзор соответствующего терминологического аппарата по выбранной тематике. Теоретическая глава структурирована по подразделам. Приведен обзор накопленного практического опыта решения проблем по выбранной тематике. – 18-21 баллов</p> <p>Практическая часть работы выполнена полностью: приведено структурирование изучаемой проблемной ситуации, дано обоснование математической модели и методики решения, проведен типовой анализ результатов моделирования – 18-21 баллов</p> <p>В процессе защиты курсовой работы обучаемый показал достаточно уверенное знание учебно-программного материала, усвоил основную и дополнительную литературу, рекомендованную программой. Дал полные ответы на поставленные вопросы. При ответе могут быть допущены незначительные недочеты в определении понятий, оговорки, поправленные студентом самостоятельно в процессе ответа – 17-18 баллов</p>
Удовлетворительно	<p>В курсовой работе приведено обоснование актуальности, цели и задачи тематики исследования четко не сформулированы – 5-7 баллов</p> <p>Содержание работы структурировано частично, материал изложен без соблюдения логики исследования – 5-7 баллов</p> <p>Работа имеет удовлетворительную степень проработанности теоретического материала (присутствует минимально необходимый в соответствии с требованиями библиографический список, ссылки на используемый материал в ходе изложения теоретической части встречаются частично) – 5-7 баллов</p> <p>Работа имеет неполное теоретическое обоснование и краткий обзор соответствующего терминологического аппарата по выбранной тематике. Обзор накопленного практического опыта решения проблем по выбранной тематике отсутствует. – 16-18 баллов</p> <p>Практическая часть работы выполнена в основном: приведено структурирование изучаемой проблемной ситуации, приведена математическая модель и методика</p>

Оценка	Характеристика результатов
	<p>решения, отсутствует типовой анализ результатов моделирования – 16-18 баллов</p> <p>В процессе защиты курсовой работы обучаемый показал низкий уровень знаний учебно-программного материала, плохо ориентируется в знании основной и дополнительной литературы, рекомендованной программой. Дал частичные ответы на поставленные вопросы. При ответе были допущены недочеты в определении понятий, студент самостоятельно не может обосновать выбор модели и методики решения в процессе ответа –13-16 баллов</p>
Неудовлетворительно	<p>В курсовой работе не приведено обоснование актуальности, цели и задачи тематики исследования не выделены или не соответствуют выбранной тематике– менее 5 баллов</p> <p>Содержание работы структурировано частично, материал изложен без соответствующей стилистики и логики – менее 5 баллов</p> <p>Работа не имеет теоретической проработки (присутствует формальный, неактуальный и/или несоответствующий тематике библиографический список, ссылки на используемый материал в ходе изложения теоретической части отсутствуют) – менее 5 баллов</p> <p>Работа не имеет соответствующего тематике теоретического обоснования и обзора или теоретическая часть представлена слабо и содержит частичные, разрозненные сведения о терминологии по тематике исследования – менее 16 баллов</p> <p>В практической части работы отсутствует структурирование изучаемой проблемной ситуации, формализация математической модели приведена с ошибками, методика решения отсутствует, отсутствует анализ результатов моделирования – менее 16 баллов</p> <p>В процессе защиты курсовой работы обучаемый показал низкий уровень знаний учебно-программного материала, плохо ориентируется в знании основной и дополнительной литературы, рекомендованной программой. Не дал ответы на два и более поставленных вопроса. При ответе были допущены ошибки в определении понятий, студент самостоятельно не может обосновать выбор модели и методики решения в процессе ответа –менее 13 баллов</p>

Для промежуточной аттестации ОМ1:

Варианты экзаменационных билетов

МИНОБРНАУКИ РФ

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Казанский государственный энергетический университет»

Факультет цифровых технологий и экономики
Кафедра ИТИС
Направление 09.03.03

УТВЕРЖДАЮ
Зав. кафедрой _____ Торкунова Ю.В.
«__» _____ 202_/2_ г.

Экзаменационный билет № 1 по дисциплине «Компьютерный анализ данных»

Построить уравнение линейной регрессии.
Включить только значимые некоррелированные факторы.
Получить метрику MAPE.
Сделать вывод

Данные взять из файла

МИНОБРНАУКИ РФ

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Казанский государственный энергетический университет»

Факультет цифровых технологий и экономики
Кафедра ИТИС
Направление 09.03.03

УТВЕРЖДАЮ
Зав. кафедрой _____ Торкунова Ю.В.
«__» _____ 202_/2_ г.

Экзаменационный билет № 2 по дисциплине «Компьютерный анализ данных»

Построить уравнение регрессии **второго порядка** для следующих исходных данных.
Оба фактора должны войти в уравнение.
Включить только значимые факторы.
Получить метрику MAE.

Данные взять из файла

МИНОБРНАУКИ РФ
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Казанский государственный энергетический университет»

Факультет цифровых технологий и экономики
Кафедра ИТИС
Направление 09.03.03

УТВЕРЖДАЮ
Зав. кафедрой _____ Торкунова Ю.В.
«__» _____ 202_/2_ г.

Экзаменационный билет № 4 по дисциплине «Компьютерный анализ данных»

Построить логит-модель бинарного выбора, предсказывающую появление просроченной задолженности по кредиту.
Включить только значимые факторы
Рассчитать метрику точности.

Данные взять из файла

МИНОБРНАУКИ РФ
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Казанский государственный энергетический университет»

Факультет цифровых технологий и экономики
Кафедра ИТИС
Направление 09.03.03

УТВЕРЖДАЮ
Зав. кафедрой _____ Торкунова Ю.В.
«__» _____ 202_/2_ г.

Экзаменационный билет № 5 по дисциплине «Компьютерный анализ данных»

Выбрать подходящий вид зависимости для нелинейной регрессии
Найти коэффициенты
Получить расчетные значения
Получить метрику R^2 .

Данные взять из файла

МИНОБРНАУКИ РФ
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Казанский государственный энергетический университет»

Факультет цифровых технологий и экономики
Кафедра ИТИС
Направление 09.03.03

УТВЕРЖДАЮ
Зав. кафедрой _____ Торкунова Ю.В.
«__» _____ 202_/2_ г.

Экзаменационный билет № 6 по дисциплине «Компьютерный анализ данных»

Выбрать подходящий вид зависимости для нелинейной регрессии
Найти коэффициенты
Получить расчетные значения
Получить метрику R^2 .

Данные взять из файла

МИНОБРНАУКИ РФ
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Казанский государственный энергетический университет»

Факультет цифровых технологий и экономики
Кафедра ИТИС
Направление 09.03.03

УТВЕРЖДАЮ
Зав. кафедрой _____ Торкунова Ю.В.
«__» _____ 202_/_2_ г.

Экзаменационный билет № 7 по дисциплине «Компьютерный анализ данных»

Построить пробит-модель бинарного выбора, предсказывающую появление просроченной задолженности по кредиту.
Включить только значимые факторы
Рассчитать меру F1.

Данные взять из файла

МИНОБРНАУКИ РФ
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Казанский государственный энергетический университет»

Факультет цифровых технологий и экономики
Кафедра ИТИС
Направление 09.03.03

УТВЕРЖДАЮ
Зав. кафедрой _____ Торкунова Ю.В.
«__» _____ 202_/_2_ г.

Экзаменационный билет № 8 по дисциплине «Компьютерный анализ данных»

Построить уравнение линейной регрессии.
Включить только значимые некоррелированные факторы.
Получить метрику MAPE.
Сделать вывод

Данные взять из файла

МИНОБРНАУКИ РФ
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Казанский государственный энергетический университет»

Факультет цифровых технологий и экономики
Кафедра ИТИС
Направление 09.03.03

УТВЕРЖДАЮ
Зав. кафедрой _____ Торкунова Ю.В.
«__» _____ 202_/_2_ г.

Экзаменационный билет № 9 по дисциплине «Компьютерный анализ данных»

Построить уравнение линейной регрессии по стандартизованным переменным.
Включить только значимые некоррелированные факторы.
Получить метрику MSE.
Сделать вывод

Данные взять из файла

МИНОБРНАУКИ РФ
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Казанский государственный энергетический университет»

Факультет цифровых технологий и экономики
Кафедра ИТИС
Направление 09.03.03

УТВЕРЖДАЮ
Зав. кафедрой _____ Торкунова Ю.В.
«__» _____ 202_/2_ г.

Экзаменационный билет № 10 по дисциплине «Компьютерный анализ данных»

Вычислить частные коэффициенты корреляции
Сделать вывод

Данные взять из файла

МИНОБРНАУКИ РФ
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Казанский государственный энергетический университет»

Факультет цифровых технологий и экономики
Кафедра ИТИС
Направление 09.03.03

УТВЕРЖДАЮ
Зав. кафедрой _____ Торкунова Ю.В.
«__» _____ 202_/2_ г.

Экзаменационный билет № 11 по дисциплине «Компьютерный анализ данных»

По данным обследования ВШЭ 2015 года дан фрагмент выборки, по которой надо определить, завит ли степень удовлетворенностью жизнью от группы дохода индивидуума. Для этого вычислить коэффициент корреляции рангов Кендела, оценить его значимость и сделать выводы.

Данные взять из файла