Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Институт прикладной математики и компьютерных наук

УТВЕРЖДЕНО:
Директор
А. В. Замятин

Оценочные материалы по дисциплине

Представление знаний и визуализация данных

по направлению подготовки

01.04.02 Прикладная математика и информатика

Направленность (профиль) подготовки:
Интеллектуальный анализ больших данных

Форма обучения
Очная

Квалификация
Магистр

Год приема
2024

СОГЛАСОВАНО:
Руководитель ОП
А.В. Замятин

Председатель УМК
С.П. Сущенко

Томск – 2024

1. Компетенции и индикаторы их достижения, проверяемые данными оценочными материалами

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-1 Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте;.

ПК-4 Способен управлять получением, хранением, передачей, обработкой больших данных.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-1.1 Владеет фундаментальными математическими, естественнонаучными, социально-экономическими и профессиональными понятиями в контексте решения задач в области информационных технологий

ИПК-4.1 Осуществляет мониторинг и оценку производительности обработки больших данных

ИПК-4.2 Использует методы и инструменты получения, хранения, передачи, обработки больших данных.

2. Оценочные материалы текущего контроля и критерии оценивания

Элементы текущего контроля:

- тесты;
- защита лабораторных работ.

2.1.Примеры тестов

Пример теста (ИОПК-2.2)

1. Квартет Энскомба был составлен в 1973 году английским математиком Ф. Дж. Энскомбом и представляет собой …
    a. четыре набора данных, которые почти идентичны по описательным характеристикам, но имеют разное распределение и при графическом представлении дают совершенно одинаковую картину
    b. четыре набора данных, которые почти идентичны по описательным характеристикам, и имеют одинаковое распределение, но при графическом представлении дают совершенно разную картину
    c. четыре набора данных, которые не идентичны по описательным характеристикам, и имеют разное распределение и при графическом представлении дают совершенно разную картину
    d. четыре набора данных, которые почти идентичны по описательным характеристикам, но имеют разное распределение и при графическом представлении дают совершенно разную картину
2. Квартет Энскомба демонстрирует важность
    a. графического представления при изучении данных
    b. вычисление описательные статистики перед началом анализа данных
    c. очистки данных и удаления ошибок перед построением «читаемых» графиков
    d. визуализации для определения точности модели анализа данных
3. Выбор типа визуализации, прежде всего, определяется:
    a. используемыми данными

b. желанием исследователя
c. тематикой исследования
d. обычно задается заранее, вместе с описанием задачи

4. Кто призывал рассматривать анализ данных как самостоятельное направление в статистике, отдельное от математической статистики?
    a. Михаэль Флоран ван Лангрен
    b. Уильям Плейфэр
    c. Шарль Жозеф Минар
    d. Джон Тьюки
5. Кто придумал диаграмму «петушиный гребень»?
    a. Михаэль Флоран ван Лангрен
    b. Уильям Плейфэр
    c. Флоренс Найтингейл
    d. Генри Бек
6. Для чего используются графики?
    a. Для отображения зависимости одной величины от другой
    b. Для отображения величин частей целого
    c. Для сравнения нескольких величин, изображённых в виде вертикальных или горизонтальных столбцов
7. Какой вид визуализации лучше всего использовать для отображения величин частей целого?
    a. График
    b. Гистограмма
    c. Круговая диаграмма
8. Для каких целей используется гистограмма?
    a. Для отображения величин частей целого
    b. Для сравнения нескольких величин, изображённых в виде вертикальных или горизонтальных столбцов
    c. Для отображения зависимости одной величины от другой
9. Что из перечисленного позволяют отслеживать динамику изменения данных?
    a. Гистограммы
    b. Круговые диаграммы
    c. Графики
10. Традиционные методы визуализации могут применяться в решении следующих задач
    a. представление информации в наглядном виде
    b. выявление и оценка закономерностей, присущих исходному набору данных
    c. снижение размерности или сжатие информации
    d. упрощение расчетов для построения модели
    e. поиск пробелов и ошибок в наборе данных

Ключи: 1 d), 2 a), 3 a), 4d), 5 c), 6 a), 7 c), 8 b), 9 c), 10 a,b,c,e)

Пример теста (ИПК-8.1)

1. Использование способов визуального представления поможет …

a. иллюстрировать результаты работы модели
b. интерпретировать полученный результат
c. настроить точные параметры модели машинного обучения
d. быть средством оценки качества построенной модели

2. Нахождение шумов и выбросов в данных …
   a. возможно при помощи средств визуализации
   b. невозможно при помощи средств визуализации
   c. не является функцией визуализации
3. Для определения корреляции чаще всего используют:
   a. тепловые карты
   b. гистограммы
   c. облако слов
   d. диаграммы рассеяния
4. Метод визуализации, который демонстрирует частотность появления слов в определенном тексте, определяя размер каждого слова пропорционально его частотности – это…:
   a. тепловая карта
   b. радиальная диаграмма
   c. облако слов
   d. гистограмма, определяющая частоту появления слов
5. Интерактивная визуализация – это …
   a. синоним понятия «анимация»
   b. способ, позволяющий представить большой объем информации в структурированном и систематизированном виде, выявить ключевые слова темы
   c. способ графического представления информации, позволяющий пользователю взаимодействовать с системой отображения информации и наблюдать ответную реакцию системы
   d. удобный инструмент для отображения процесса мышления и структурирования информации в визуальной форме
6. Визуализация в виде параллельных координат используете для визуализации
   a. многомерных данных, но размерность не должна превышать 3
   b. многомерных данных, но размерность не должна превышать 10
   c. многомерных данных, размерность формально не ограничена
7. Визуализация в виде "лиц Чернова" является представлением информации в …
   a. двухмерном измерении
   b. трехмерном измерении
   c. более, чем в трехмерном измерении
8. Строка многомерной таблицы данных обычно интерпретируется как:
   a. запись
   b. атрибут
   c. объект
   d. переменная
   e. признак
9. Продолжите фразу: "Кластеризация и классификация относятся к...
   a. стратегии обучения с учителем"

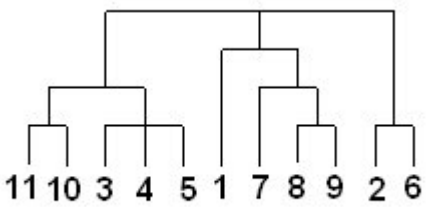
b. стратегии обучения без учителя"
c. к двум разным стратегиям: обучения без учителя и обучения с учителем"

10. Метод деревьев решений (и соответствующая ему визуализация) применяется для решения задач …
    a. классификации
    b. кластеризации
    c. классификации и кластеризации

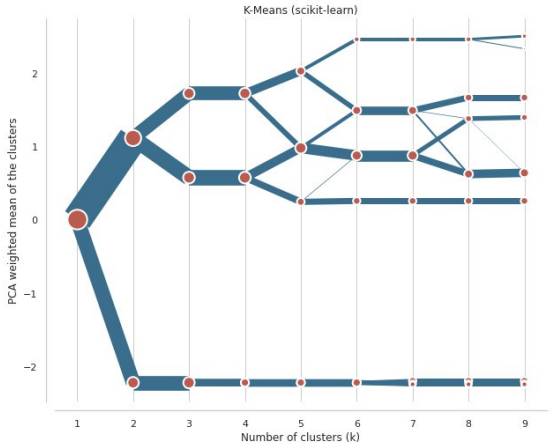
Ключи: 1 a,b,d), 2 a), 3 a,d), 4 c), 5 c), 6 c), 7 c), 8 a,c), 9 c), 10 a)

Пример теста (ИУК-12.1)

1. Возможности визуализации включают:
    a. поддержку интерактивного и согласованного исследования
    b. помощь в представлении результатов
    c. формализацию задач машинного обучения
    d. повышение точности моделей машинного обучения
2. Визуализация результатов регрессионного анализа помогает в получении сведений о том:
    a. какую форму имеет зависимость между исследуемыми переменными
    b. какой характер имеет зависимость между исследуемыми переменными
    c. какую количественную взаимосвязь имеют два набора данных
3. Что помогает сделать визуализация в задачах кластеризации:
    a. Отыскать "скрытую структуру" данных.
    b. Разделить набор данных на предопределенные классы.
    c. Определить центр набора данных
    d. Упорядочить объекты в статистически однородные группы
4. Как называется данный граф?



    a. Граф состояний
    b. Дендрограмма
    c. Гистограмма
    d. Древо кластеров
5. Как называется данный граф?



a. Граф состояний
b. Дендрограмма
c. Гистограмма
d. Кластерограмма

6. Такие данные как температура воздуха относятся к …
   a. непрерывным данным
   b. дискретным данным
   c. целочисленным данным
   d. бинарным данным
7. Какая из следующих техник визуализации используется для отображения многомерных данных в виде плавных кривых на плоскости?
   a. Лица Чернова
   b. Гистограммы
   c. Кривые Эндрюса
   d. Древовидные диаграммы
8. Что представляет собой "лицо Чернова" в контексте визуализации данных?
   a. Набор параллельных осей для отображения многомерных данных
   b. Способ отображения многомерных данных через лица, где каждая характеристика изменяет определённую часть лица
   c. Математическая модель для анализа временных рядов
   d. Техника кластеризации многомерных данных
9. Какие данные особенно удобно визуализировать с помощью кривых Эндрюса?
   a. Одномерные временные ряды
   b. Двумерные выборки с разными распределениями
   c. Многомерные данные с большим числом переменных
   d. Наборы данных с категориальными переменными
10. Что представляет собой каждое пересечение линий в графике параллельных координат?
    a. Пересечение двух переменных с одинаковыми значениями
    b. Проекция одномерных данных на двухмерное пространство
    c. Значение каждой переменной для одной наблюдаемой точки
    d. Описание зависимости между двумя случайными величинами

Ключи: 1 a,b), 2 a,b), 3 a), 4 b), 5 d), 6 a), 7 c), 8 b), 9 c), 10 c)
Критерии оценивания: тест считается пройденным, если обучающий ответил правильно как минимум на половину вопросов.

2.2. Примеры заданий для лабораторных работ (ИОПК-2.2, ИПК-8.1, ИУК-12.1)

Лабораторная работа № 1

1. Постройте график функции y(x).

Вариант	y(x)
1	$y(x) = e^{-2x} \cdot \cos(20x)$, $x \in [0,1]$, $\Delta x = 0{,}01$
2	$y(x) = e^{-x} \cdot sin^2(10x)$, $x \in [0,1]$, $\Delta x = 0{,}01$
3	$y(x) = e^{-x} \cdot sin^3(20x)$, $x \in [0,1]$, $\Delta x = 0{,}01$
4	$y(x) = e^{x} \cdot \cos(20x)$, $x \in [0,2]$, $\Delta x = 0{,}01$
5	$y(x) = e^{x} \cdot sin^2(20x)$, $x \in [0,2]$, $\Delta x = 0{,}01$

2. Постройте в одном окне графики двух функции y1(x) и y2(x).

Вариант	y1(x)	y2(x)	Интервал, шаг
1	$y1(x) = 10e^{-2x}$	$y2(x) = \cos(20x)$	$x \in [0,1]$, $\Delta x = 0{,}01$
2	$y1(x) = 10e^{-x}$	$y2(x) = sin^2(10x)$	$x \in [0,1]$, $\Delta x = 0{,}01$
3	$y1(x) = 10e^{-x}$	$y2(x) = sin^3(20x)$	$x \in [0,1]$, $\Delta x = 0{,}01$
4	$y1(x) = 10e^{x}$	$y2(x) = \cos(20x)$	$x \in [0,1]$, $\Delta x = 0{,}01$
5	$y1(x) = 10e^{x}$	$y2(x) = sin^2(20x)$	$x \in [0,1]$, $\Delta x = 0{,}01$

3. Добавьте к указанным двум графикам третий – по своему усмотрению. И постройте эти три графика в разделенных окнах:
- Используя функцию subplot();
- Используя функцию subplots();
- Свободная компоновка с GridSpec.

Лабораторная работа № 2

1. Выберите функцию по номеру варианта.
2. Постройте график функции. График должен содержать заголовок и подписи осей.
3. Скопируйте код для построения графика и измените внешний вид заголовка и подписей осей, используя необходимые параметры.
4. Используя аннотацию – выделите: локальный минимум (или максимум) функции, точку разрыва, точку перегиба. Используйте разные типы стрелок.
5. Добавьте на график надпись (текст придумайте самостоятельно).
6. Добавьте еще один график (любой), оформите легенду.

Варианты функций для построения графиков

1	$y = \frac{x+6}{x^2+13}$	3	$y = cos\,x + \frac{\sqrt{3}}{2}x - \frac{1}{2}$
2	$y = \frac{x}{2} + cos\,x$	4	$y = sin\,2x - x$

Лабораторная работа № 3

1. Постройте в одном окне графики двух функции y1(x) и y2(x), $x \in [0,2\pi]$, $\Delta x = 0{,}2$. При построении графиков вид функций и параметры выбрать согласно варианту. Толщину линии, размер и цвет маркера, а также контура маркера подобрать самостоятельно.

Вариант	Функции	Параметры (цвет линии, тип маркера, тип линии)

1	$y1(x) = e^{-0,5x}$	Желтый, точка, сплошная
	$y2(x) = \cos(3x)$	Розовый, кружок, пунктирная
2	$y1(x) = e^{-x}$	Голубой, крестик, штрих-пунктирная
	$y2(x) = sin^2(x)$	Красный, знак плюс, штриховая
3	$y1(x) = e^{-x}$	Зеленый, треугольник вершиной влево, сплошная
	$y2(x) = sin^3(x)$	Синий, треугольник вершиной вправо, пунктирная
4	$y1(x) = 0,05e^{x}$	Черный, пятиконечная звезда, штрих-пунктирная
	$y2(x) = 60\cos(2x)$	Желтый, знак плюс, штриховая
5	$y1(x) = 0,1e^{-x}$	Желтый, квадрат, штрих-пунтктирная
	$y2(x) = 60sin^2(x)$	Розовый, ромб, штриховая

2. Постройте один из графиков функций для вашего варианта (y1 или y2), используйте для построения не менее 500 точек.
3. Подберите наилучший вариант маркировки (на основе визуализации и сравнения нескольких вариантов).
4. Продемонстрируйте навыки построения графиков с заливкой (2-3 варианта).
5. Продемонстрируйте навыки обрезки графика (ограничения для обрезки придумайте самостоятельно).

Лабораторная работа № 4

1. Постройте stairs-график (ступенчатый) для функции y(x). При построении графиков вид функций и параметры выбрать согласно варианту. Толщину линии, размер и цвет маркера, а также контура маркера подобрать самостоятельно.

Вариант	y(x)	Интервал, шаг	Параметры (цвет линии, тип маркера, тип линии)
1	$y(x) = 10e^{-2x}$	$x \in [0,5]$, $\Delta x = 0,2$	Черный, треугольник вершиной вверх, сплошная
2	$y(x) = e^{x}$	$x \in [0,4]$, $\Delta x = 0,2$	Черный, шестиконечная звезда, пунктирная
3	$y(x) = \sqrt{x}$	$x \in [0,100]$, $\Delta x = 5$	Красный, ромб, штриховая

2. Постройте stem-график для функции y(x), $x \in [0,2\pi]$, $\Delta x = 0,2$. При построении графиков вид функций и параметры выбрать согласно варианту. Толщину линии, размер и цвет маркера, а также контура маркера подобрать самостоятельно.

Вариант	Функции	Параметры (цвет линии, тип маркера, тип линии)
1	$y(x) = \cos(x)$	Розовый, кружок, пунктирная
2	$y(x) = sin^2(x)$	Красный, знак плюс, штриховая
3	$y(x) = sin^3(x)$	Синий, треугольник вершиной вправо, пунктирная

3. Вам необходимо установить зависимость между двумя переменными (согласно варианту), а также исследовать некоторую динамику по годам.
   - Постройте диаграмму рассеяния и сделайте по результатам построения вывод о наличии и характере связи.
   - Постройте столбчатую диаграмму, согласно варианту.
   - Постройте круговую диаграмму согласно варианту.
   - Все параметры диаграмм подобрать самостоятельно.

Варианты 1,2,3. В таблице 1 приведены данные по одной очень активной на рынке компании. Исследуйте зависимость ***Y*** от (***X_1* – вариант 1, *X_2* – вариант 2, *X_3* – вариант 3**) за 20 лет– постройте **диаграмму рассеяния** и сделайте вывод о наличии/отсутствии зависимости и ее характере.

Постройте и заполните таблицу:

	Пятилетка, гг.	**Средний *Y***	**Средний *X_1* – вариант 1, *X_2* – вариант 2, *X_3* – вариант 3**
1	2004-2008		
2	2009-2013		
3	2014-2018		
4	2019-2023		
	Итого в среднем:		

Постройте **столбчатые диаграммы**: 1) для **среднего *Y*** по пятилеткам; 2) для **среднего (*X_1* – вариант 1, *X_2* – вариант 2, *X_3* – вариант 3**) по пятилеткам. Расположите две диаграммы рядом – по горизонтали.

Постройте **круговую диаграмму,** отражающую процентное соотношение ***Y*** по пятилеткам.

Сделайте общий вывод.

Таблица 1.

	Год	**Чистая прибыль ($, млрд)**	**Оборачиваемость капитала ОК ($, млрд)**	**Вложенный капитал K ($, млрд)**	**Количество работающих L (тыс. чел.).**
		Y	***X_1***	***X_2***	***X_3***
1	2004	6,6	6,9	83,6	222,0
2	2005	3,0	18	6,5	32,0
3	2006	6,5	107,9	50,4	82,0
4	2007	3,3	16,7	15,4	45,2
5	2008	0,1	79,6	29,6	299,3
6	2009	3,6	16,2	13,3	41,6
7	2010	1,5	5,9	5,9	17,8
8	2011	5,5	53,1	27,1	151,0
9	2012	2,4	18,8	11,2	82,3
10	2013	3,0	35,3	16,4	103,0
11	2014	4,2	71,9	32,5	225,4
12	2015	2,7	93,6	25,4	675,0
13	2016	1,6	10,0	6,4	43,8
14	2017	2,4	31,5	12,5	102,3
15	2018	3,3	36,7	14,3	105,0
16	2019	1,8	13,8	6,5	49,1
17	2020	2,4	64,8	22,7	50,4
18	2021	1,6	30,4	15,8	480,0
19	2022	1,4	12,1	9,3	71,0
20	2023	0,9	31,3	18,9	43,0

Лабораторная работа № 5

1. Выберите датасет из встроенного в seaborn для дальнейшего анализа (опционально – можно взять любые открытые данные, используя любой открытый репозиторий).
2. Найдите описание датасета (о чем данные, где взят и т.п.) и данных (описание – что характеризует, тип) и добавьте в отчет. Изучение предметной области – немаловажный шаг для аналитика данных.
3. Проверьте – есть ли отсутствующие данные, если есть – удалите.

4. Используя различного рода визуализации проведите исследование данных: закономерности, зависимости, распределения. Снабдите каждый график необходимыми подписями (заголовок, оси, легенда и т.п.), выберите подходящую цветовую гамму. В качестве образца можно взять данный Урок.
5. По желанию можно добавить графики и диаграммы, не описанные в Уроке.
6. По каждому графику должен быть вывод – что исследовано и что получено.
7. Сделайте общий вывод по работе.

Лабораторная работа № 6

1. Постройте визуализацию ваших данных (ЛР 5), используя «Лица Чернова».
2. Интерпретируйте полученный результат.
3. Сгруппируйте Ваши данные по какому-либо признаку, найдите центральные значения о остальным признакам для каждой полученной группы. Постройте Лица Чернова для полученных групп и вновь интерпретируйте результат.
4. В каких случаях Вы рекомендуете использовать Лица Чернова?
5. Постройте график параллельных координат для для несгруппированных и сгруппированных данных.
6. Постойте график RadViz для для несгруппированных и сгруппированных данных.
7. Постойте диаграммы «Кривые Эндрюса» для несгруппированных и сгруппированных данных. Какие можно сделать выводы?
8. Какая, на Ваш взгляд, визуализация «нагляднее» всего демонстрирует различие в группах?
9. Сделайте выводы в целом по ЛР 6.

Лабораторная работа № 7

ЧАСТЬ 1. Используйте набор данных (например, представленный в курсе **Fish Market** – набор данных включает виды рыб, вес, длину, высоту и ширину.) для следующих упражнений.

Внимание! Если студенты имеют уровень «начинающий», допускается выполнение только пп. 1-8 этого задания. Данное решение согласуется с преподавателем.

№	Задание	Ожидаемый результат
1. ПЕВИЧНЫЙ АНАЛИЗ ДАННЫХ. ВИЗУАЛИЗАЦИЯ		
1	Загрузите датасет fish.csv (в папке курса). Укажите тип переменных датасета	Программный код с описанием, вывод структуры датасета с описанием. Вывод нескольких записей датасета.
2	Выведите описательную статистику для переменных датасета	Вывод
3.	Используя визуализацию, покажите, сколько различных видов рыб включает ваш набор данных. Используйте интерактивную визуализацию.	2-3 диаграммы разного вида, используются, в том числе, интерактивные диаграммы
4.	Исследуйте, есть ли какие-либо зависимые переменные.	Визуализация корреляционных зависимостей, вывод о наличии или отсутствии корреляции, описание вида зависимости.

№	Задание	Ожидаемый результат
5.	Определите тип распределения количественных переменных	Визуализация для определения типа распределения, выводы
6.	Существуют ли в этом наборе данных аномальные наблюдения?	Визуализация наличия выбросов, выводы
7.	Проверьте данные на наличие пропущенных значений.	Визуализация пропусков в данных. Выводы.
8.	Используя знания, полученные в ходе выполнения лабораторной работы № 2, реализуйте визуализацию многомерных данных.	Можно построить кривые Эндрюса, графики в параллельных координатах, лица Чернова, графики RadViz и т.п.
2. **РЕГРЕССИОННЫЙ АНАЛИЗ**		
9.	Пусть переменная *Weight* (*y*) – зависимая (целевая) переменная. А *одна из других* – независимая (*x*). Постройте модель линейной регрессии. Дайте общую оценку созданной модели.	Визуализация результатов построения линейной регрессии. Выводы.
10.	Пусть переменная *Weight* (*y*) – зависимая (целевая) переменная. А *все другие* – независимые (x_i). Постройте модель линейной регрессии. Постройте график рассеяния (scatterplot) остатков (residuals), проверьте распределение остатков на нормальность. Дайте общую оценку созданной модели.	Визуализация результатов. Требуемые графики, выводы. Какая модель, на ваш взгляд, точнее (однофакторная (п.8) или многофакторная (п.9))?
11.	Найдите выбросы, наблюдения с высокой степенью напряженности, влиятельные наблюдения для модели, построенной в п.9. Визуализируйте результаты.	Требуемые графики, выводы.
12.	Удалите все, найденные в п.10 «плохие» точки из датасета. Постройте модель на обновленных данных. Сравните результат с п. 9	Графики, выводы.
3. ДИСПЕРСИННЫЙ АНАЛИЗ		
13.	Есть ли статистическая значимость различия веса рыб по видам? Выяснить на уровне значимости 5%.	Графики, выводы.

ЧАСТЬ 2. Выберите набор данных, согласно вашего варианта, по ссылкам, приведенным ниже.

Студент может использовать СВОЙ вариант набора данных – в случае, например, если данные связаны с его научной работой или другими исследованиями.

Необходимо провести исследования, близкие к тем, что проводились в Части 1. Возможно проведение других видов анализа, вместо/вместе с регрессионным: кластеризация, классификация, факторный анализ и т.п. с обязательным представлением визуальных элементов.

3. Оценочные материалы итогового контроля (промежуточной аттестации) и критерии оценивания

Схема определения итоговой балльно-рейтинговой оценки

	Виды учебной деятельности	**Виды оценивания (балльные оценки)**

1	Изучение дисциплины (форма контроля - **зачет, диф. зачет**)	Мероприятия текущего контроля в семестре Максимум 100 баллов			=	Итоговая рейтинговая оценка
2	Изучение дисциплин (форма контроля - **экзамен**)	Мероприятия текущего контроля в семестре Максимум 80 баллов	+	Мероприятия промежуточной аттестации (экзамен) Максимум 20 баллов	=	Итоговая рейтинговая оценка

Тематический план курса

№	Модуль, тема	Вид занятия (ЛК/ЛБ)	Максимальное количество баллов	
			Зачет, дифзачет	Экзамен
1	**Тема 1. Концептуальные основы визуализации данных.**			
	Тема 1.1. Концептуальные основы визуализации. Тема 1.2. Методы визуализации многомерных данных. История визуализации данных. Как визуализация данных помогает в принятии решений?	ЛК		
	Тема 1.2. Способы визуализации. Библиотеки Python для решений задач визуализации	ЛК		
	Лабораторная работа № 1.	ЛБ	10	10
2	**Тема 2. Способы и инструменты визуализации данных**			
	Тема 2.1. Библиотеки Python для решений задач визуализации. Сравнение полученных визуальных образов.	ЛК		
	Тема 2.2. Визуализация в задачах машинного обучения	ЛК		
	Тема 2.3. Инфографика и пиктография			
	Лабораторная работа № 2.	ЛБ	10	10
	Лабораторная работа № 3.	ЛБ	15	10
3	**Тема 3. Управление компоновкой диаграмм. Интерактивная визуализация**			
	Тема 3.1. Интерактивная визуализация, библиотеки Python для интерактивной визуализации.	ЛК		
	Тема 3.2. Примеры задач, разбор примеров. Научная визаулизация	ЛК		
	Лабораторная работа № 4.	ЛБ	15	10
	Лабораторная работа № 5.	ЛБ	15	10
4	**Тема 4. Способы визуализации многомерных данных**			
	Тема 4.1. Визуализация многомерных данных: график параллельных координат, лица Чернова, радиальные диаграммы.	ЛК		

№	Модуль, тема	Вид занятия (ЛК/ЛБ)	Максимальное количество баллов	
			Зачет, дифзачет	Экзамен
	Лабораторная работа № 6.	ЛБ	15	15
	Тема 5. Научная визуализация. Визуализация в научных исследованиях			
	Лабораторная работа № 7.	ЛБ	20	15
	ЭКЗАМЕН (ФИНАЛЬНЫЙ ПРОЕКТ)			20
	ИТОГО		**100**	**100**

Критерии оценки отчетов по лабораторным работам и финальному проекту (экзамену)

№	Критерий	Максимальный балл
		Экзамен
1.	На проверку преподавателю представлено:	
	1) файл-отчет с исходным кодом (например, *.ipynb)	0,5
	2) отчет в формате html	0,5
	3) все массивы данных, на основе которых была проведена работа (можно в архиве).	0,5
2.	Отчет содержит подробные комментариями к коду и результатам визуализации.	2
3.	Комментарии содержат не только факт выполнения – построена такая-то диаграмма, но также описание результата: что показывает диаграмма или график.	3
4.	Отчет содержит общий вывод по работе.	1
5.	Общий вывод по работе отражает не только факты (выполнено, построено и т.п.), но и аргументированные выводы.	3
6	Графики, диаграммы, а также другие объекты визуализации сопровождаются всеми необходимыми подписями (оси, название, легенда (если необходимо)), понятны для интерпретации.	2
7.	Все выводы построены грамотно, аргументированно. Если требуется в работе – представлены гиперссылки (например, на дашборды).	3
8.	В программном коде в отчете нет ошибок.	0,5
Выбор: устная защита или оценка по отчету		
9.1	При защите студент ответил на все дополнительные вопросы	2
9.2	В целом, отчет выполнен аккуратно, порядок этапов выполнения работы логичен.	2
	ВСЕГО:	**20**

4. Оценочные материалы для проверки остаточных знаний (сформированности компетенций)

4.1. Тест

1. Использование способов визуального представления поможет …
    a. иллюстрировать результаты работы модели
    b. интерпретировать полученный результат
    c. настроить точные параметры модели машинного обучения
    d. быть средством оценки качества построенной модели
2. Нахождение шумов и выбросов в данных …
    a. возможно при помощи средств визуализации
    b. невозможно при помощи средств визуализации
    c. не является функцией визуализации
3. Для определения корреляции чаще всего используют:
    a. тепловые карты
    b. гистограммы
    c. облако слов
    d. диаграммы рассеяния
4. Метод визуализации, который демонстрирует частотность появления слов в определенном тексте, определяя размер каждого слова пропорционально его частотности – это…:
    a. тепловая карта
    b. радиальная диаграмма
    c. облако слов
    d. гистограмма, определяющая частоту появления слов
5. Интерактивная визуализация – это …
    a. синоним понятия «анимация»
    b. способ, позволяющий представить большой объем информации в структурированном и систематизированном виде, выявить ключевые слова темы
    c. способ графического представления информации, позволяющий пользователю взаимодействовать с системой отображения информации и наблюдать ответную реакцию системы
    d. удобный инструмент для отображения процесса мышления и структурирования информации в визуальной форме
6. Визуализация в виде параллельных координат используете для визуализации
    a. многомерных данных, но размерность не должна превышать 3
    b. многомерных данных, но размерность не должна превышать 10
    c. многомерных данных, размерность формально не ограничена
7. Визуализация в виде "лиц Чернова" является представлением информации в …
    a. двухмерном измерении
    b. трехмерном измерении
    c. более, чем в трехмерном измерении
8. Строка многомерной таблицы данных обычно интерпретируется как:
    a. запись
    b. атрибут
    c. объект
    d. переменная
    e. признак
9. Продолжите фразу: "Кластеризация и классификация относятся к...

a. стратегии обучения с учителем"
b. стратегии обучения без учителя"
c. к двум разным стратегиям: обучения без учителя и обучения с учителем"

10. Метод деревьев решений (и соответствующая ему визуализация) применяется для решения задач …
    a. классификации
    b. кластеризации
    c. классификации и кластеризации

Ключи: 1 a,b,d), 2 a), 3 a,d), 4 c), 5 c), 6 c), 7 c), 8 a,c), 9 c), 10 a)

Информация о разработчиках

Марухина Ольга Владимировна, канд. техн. наук, доцент, кафедра теоретических основ информатики ИПМКН ТГУ, доцент