

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Усынин Максим Валерьевич
Должность: Ректор
Дата подписания: 07.07.2023 16:05:54
Уникальный программный ключ:
f498e59e83f65dd7c3ce7bb8a25cbbabb33ebc58

**Частное образовательное учреждение высшего образования
«Международный Институт Дизайна и Сервиса»
(ЧОУВО МИДиС)**

Кафедра математики и информатики

**ФОНД
ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕГО
КОНТРОЛЯ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНОЙ
АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ
АНАЛИЗ И ВИЗУАЛИЗАЦИЯ ДАННЫХ**

Направление подготовки: 09.03.03 Прикладная информатика
Направленность (профиль): Разработка веб и мобильных приложений
Квалификация выпускника: Бакалавр
Год набора: 2023

Автор-составитель: Овсяницкая Л.Ю.

СОДЕРЖАНИЕ

1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы.....	3
2. Показатели и критерии оценивания компетенций на различных этапах их формирования, описание шкал оценивания.....	4
3. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы.....	6
4. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы.....	25

1. ПЕРЕЧЕНЬ КОМПЕТЕНЦИЙ С УКАЗАНИЕМ ЭТАПОВ ИХ ФОРМИРОВАНИЯ В ПРОЦЕССЕ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Процесс изучения дисциплины «Анализ и визуализация данных» направлен на формирование следующих компетенций:

Код и наименование компетенций выпускника	Код и наименование индикатора достижения компетенций
ПК-3 Способен проектировать и разрабатывать информационные системы в соответствии с требованиями заказчика	ПК-3.1. Выполнять действия разработке прототипов информационных систем, мобильных и веб приложений ПК-3.2. Выполнять действия по проектированию, верификации информационных систем, мобильных и веб приложений в соответствии с требованиями заказчика. ПК 3.3. Владеть инструментами и методами разработки и тестирования баз данных информационных систем
ПК-4 Способен проектировать информационные ресурсы (веб, мобильных приложений) составлять формализованные описания решений, поставленных задач, в соответствии с требованиями, принятыми в организации нормативных документов	ПК-4.1. Составлять формализованные описания решений поставленных задач в соответствии с требованиями, принятыми в организации нормативных документов, выполнять действия по проектированию структур баз данных и дизайну программных интерфейсов. ПК-4.2. Использовать существующие типовые решения и шаблоны информационных ресурсов (веб, мобильных приложений), применять методы и средства проектирования и дизайна информационных ресурсов, баз данных и программных интерфейсов. ПК-4.3. Применять типовые решения, библиотеки программных модулей, шаблоны, классы объектов, используемые при разработке информационных ресурсов (веб, мобильных приложений)
ПК-5 Способен разрабатывать мобильные приложения	ПК-5.1. Осуществлять деятельность по разработке и отладке мобильных приложений. ПК-5.2. Работать со стандартными сервисами платформ и со встроенными устройствами для получения данных, использовать технологии для работы с различными протоколами обмена данными. ПК-5.3. Применять программные средства, технологии и платформы для разработки мобильных приложений, знать основы информационной безопасности

№ п/п	Код компетенции	Наименование компетенции	Этапы формирования компетенций
1.	ПК-3	Способен проектировать и разрабатывать информационные системы в соответствии с требованиями заказчика	<i>1 Этап – Знать:</i> ПК-3.1. - способы разработки прототипов информационных систем, мобильных и веб приложений <i>2 Этап – Уметь:</i> ПК-3.2. - выполнять действия по проектированию, верификации информационных систем, мобильных и веб приложений в соответствии с требованиями заказчика <i>3 Этап – Владеть:</i>

			ПК-3.3. - инструментами и методами разработки и тестирования баз данных информационных систем
2.	ПК-4	Способен проектировать информационные ресурсы (веб, мобильных приложений) составлять формализованные описания решений, поставленных задач, в соответствии с требованиями, принятыми в организации нормативных документов	<i>1 Этап – Знать:</i> ПК-4.1. - формализованные способы описания решений поставленных задач в соответствии с требованиями, принятыми в организации нормативных документов; - технологии проектирования структур баз данных и дизайна программных интерфейсов <i>2 Этап – Уметь:</i> ПК-4.2. - использовать существующие типовые решения и шаблоны информационных ресурсов (веб, мобильных приложений); - применять методы и средства проектирования и дизайна информационных ресурсов, баз данных и программных интерфейсов <i>3 Этап – Владеть:</i> ПК-4.3. - способами применения типовых решений, библиотек программных модулей, шаблонов, классов объектов, используемых при разработке информационных ресурсов (веб, мобильных приложений)
3.	ПК-5	Способен разрабатывать мобильные приложения	<i>1 Этап – Знать:</i> ПК-5.1. - технологии разработки и отладки мобильных приложений <i>2 Этап – Уметь:</i> ПК-5.2. - работать со стандартными сервисами платформ и со встроенными устройствами для получения данных, использовать технологии для работы с различными протоколами обмена данными <i>3 Этап – Владеть:</i> ПК-5.3. - программными средствами, технологиями для разработки мобильных приложений; - основами информационной безопасности;

2. ПОКАЗАТЕЛИ И КРИТЕРИИ ОЦЕНИВАНИЯ КОМПЕТЕНЦИЙ НА РАЗЛИЧНЫХ ЭТАПАХ ИХ ФОРМИРОВАНИЯ, ОПИСАНИЕ ШКАЛ ОЦЕНИВАНИЯ

№ п/п	Код компетенции	Наименование компетенции	Критерии оценивания компетенций на различных этапах формирования	Шкала оценивания
1.	ПК-3	Способен проектировать и разрабатывать информационные си-	<i>1 Этап – Знать:</i> ПК-3.1. - способы разработки прототипов информационных систем, мобильных и веб приложений <i>2 Этап – Уметь:</i>	«ЗАЧТЕНО»: 1. Усвоение программного материала. 2. Умение применять основные приемы и

		стемы в соответствии с требованиями заказчика	ПК-3.2. - выполнять действия по проектированию, верификации информационных систем, мобильных и веб приложений в соответствии с требованиями заказчика <i>3 Этап – Владеть:</i> ПК-3.3. - инструментами и методами разработки и тестирования баз данных информационных систем	методы обработки данных. 3. Выполнение практических заданий и самостоятельной работы за семестр. 4. Точность и обоснованность выводов. 5. Точные, полные и логичные ответы на дополнительные вопросы. «НЕ ЗАЧТЕНО»: 1. Незнание значительной части программного материала 2. Невыполнение практических заданий и самостоятельной работы за семестр. 3. Грубые ошибки при выполнении практических заданий и самостоятельной работы. 4. Неумение выделить главное, сделать выводы и обобщения. 5. Неправильные ответы на дополнительные вопросы.
2.	ПК-4	Способен проектировать информационные ресурсы (веб, мобильных приложений) составлять формализованные описания решений, поставленных задач, в соответствии с требованиями, принятыми в организации нормативных документов	<i>1 Этап – Знать:</i> ПК-4.1. - формализованные способы описания решений поставленных задач в соответствии с требованиями, принятыми в организации нормативных документов; - технологии проектирования структур баз данных и дизайна программных интерфейсов <i>2 Этап – Уметь:</i> ПК-4.2. - использовать существующие типовые решения и шаблоны информационных ресурсов (веб, мобильных приложений); - применять методы и средства проектирования и дизайна информационных ресурсов, баз данных и программных интерфейсов <i>3 Этап – Владеть:</i> ПК-4.3. - способами применения типовых решений, библиотек программных модулей, шаблонов, классов объектов, используемых при разработке информационных ресурсов (веб, мобильных приложений)	«НЕ ЗАЧТЕНО»: 1. Незнание значительной части программного материала 2. Невыполнение практических заданий и самостоятельной работы за семестр. 3. Грубые ошибки при выполнении практических заданий и самостоятельной работы. 4. Неумение выделить главное, сделать выводы и обобщения. 5. Неправильные ответы на дополнительные вопросы.
3	ПК-5	Способен разрабатывать мобильные приложения	<i>1 Этап – Знать:</i> ПК-5.1. - технологии разработки и отладки мобильных приложений <i>2 Этап – Уметь:</i> ПК-5.2. - работать со стандартными сервисами платформ и со встроенными устройствами для получения данных, использовать технологии для работы с различ-	

			ными протоколами обмена данными	
			3 Этап – Владеть: ПК-5.3. - программными средствами, технологиями для разработки мобильных приложений; - основами информационной безопасности;	

3. ТИПОВЫЕ КОНТРОЛЬНЫЕ ЗАДАНИЯ ИЛИ ИНЫЕ МАТЕРИАЛЫ, НЕОБХОДИМЫЕ ДЛЯ ОЦЕНКИ ЗНАНИЙ, УМЕНИЙ, НАВЫКОВ И (ИЛИ) ОПЫТА ДЕЯТЕЛЬНОСТИ, ХАРАКТЕРИЗУЮЩИХ ЭТАПЫ ФОРМИРОВАНИЯ КОМПЕТЕНЦИЙ В ПРОЦЕССЕ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

1 ЭТАП – ЗНАТЬ

Вопросы тестирования по теоретическим основам:

1. Шкала, содержащая только категории; данные в ней не могут упорядочиваться, с ними не могут быть произведены никакие арифметические действия, называется:

Номинальная

Порядковая

Интервальная

Относительная

Дихотомическая

2. Шкала, в которой числа присваивают объектам для обозначения относительной позиции объектов, но не величины различий между ними, называется:

Номинальная

Порядковая

Интервальная

Относительная

Дихотомическая

3. Шкала, разности между значениями которой могут быть вычислены, однако их отношения не имеют смысла, называется:

Номинальная

Порядковая

Интервальная

Относительная

Дихотомическая

4. Шкала, в которой есть определенная точка отсчета и возможны отношения между значениями шкалы, называется:

Номинальная

Порядковая

Интервальная

Относительная

Дихотомическая

5. Шкала, содержащая только две категории, называется:

Номинальная

Порядковая

Интервальная

Относительная

Дихотомическая

6. Группа методов визуального анализа данных, к которой относится календарь, называется:

Табличные методы

Графики

Диаграммы

Тепловая карта

7. Какой способ визуализации информации целесообразно использовать для описания ряда объектов, обладающих одинаковыми наборами свойств?

Таблицы

Списки

Диаграммы

Графики

8. Какой вид диаграмм отражает данные, суммарное значение которых образуют 100%:

Линейные диаграммы

Круговые диаграммы

Гистограммы

9. Для чего используются графики?

Для отображения зависимости одной величины от другой

Для отображения величин частей целого

Для сравнения нескольких величин, изображённых в виде горизонтальных столбцов

Для сравнения нескольких величин, изображённых в виде вертикальных столбцов

10. Лучше всего отображает дискретные временные значения тип диаграммы:

Круговая

Столбчатая

Линейчатая

Точечная

Пузырьковая

11. При отображении круговой диаграммы рекомендуется использовать не более N секторов:

2

4

6

8

12. Для отображения доли каждого значения в общем наборе данных следует использовать диаграмму:

Круговую

Столбчатую

Линейчатую

Точечную

Пузырьковую

13. Тип диаграммы, которую в MS Excel можно построить с помощью формул условного форматирования, называется:

Плоское дерево

Коробчатая диаграмма

Тепловая карта

Полярная диаграмма

Диаграмма с параллельными координатами

14. Для визуализации непрерывных изменений во времени следует использовать:

Плоское дерево

Линейную диаграмму

Тепловую карту

Полярную диаграмму

15. Одновременно визуализировать входящие и исходящие потоки данных позволяет диаграмма

Плоское дерево

Линейная диаграмма

Санкей

Полярная диаграмма

16. Легенду на диаграмме нужно отображать:

Всегда

Никогда

По желанию

При использовании различных вариантов кодирования данных

При многомерном анализе данных

17. Основным элементами кодирования данных на диаграмме являются форма, размер и:

Оптимальность

Цвет

Размерность

Непрерывность

18. Метод, позволяющий предсказывать значения той или иной непрерывной числовой величины для входных данных называется:

Кластеризация

Классификация

Регрессия

Метод опорных векторов

19. Какой из видов машинного обучения основывается на взаимодействии обучаемой системы со средой?

С учителем

Без учителя

С подкреплением

Глубинное

20. Метод, позволяющий прогнозировать выходы с двумя возможными значениями, помеченными как «0» или «1», называется:

Логистическая регрессия

Метод опорных векторов

Метод k-ближайших соседей

Нейросеть

21. Ошибку 1-го рода иногда называют:

Точность модели

Ложная тревога

Вероятность отказа

Пропуск цели

22. Ошибку 2-го рода иногда называют:

Пропуск цели

Ложная тревога

Вероятность отказа

Точность модели

23. Доля объектов, названных положительными и являющиеся положительным, отражает метрика:

Recall

Precision

Accuracy

Exactly

24. Доля объектов положительного класса из всех объектов положительного класса определяется метрикой

Ошибка 1-го рода

Ошибка 2-го рода

Ошибка 3-го рода

Ошибка 4-го рода

25. "Пропуск" уходящего абонента и ошибочное принятие нулевой гипотезы называется:

Разделяющие

Опорные

Решающие

Гипервектора

26. Под «соседями» в методе k-NN понимаются:

Параметры модели, лучше всего описывающие объект

Рядом находящиеся объекты

Объекты, близкие к исследуемому в том или ином смысле

Объекты, находящиеся напротив исследуемого объекта

27. Кластеризация относится к методу обучения:

С учителем

Без учителя

С подкреплением

С предсказанием

28. Классификация относится к методу обучения:

С учителем

Без учителя

С подкреплением

С предсказанием

29. В чем заключается цель машинного обучения?

Предсказать входные данные по результату

Предсказать результат по входным данным

Обучить машину ездить без водителя

Проводить арифметические вычисления

30. Обучение, основанное на маркированных обучающих данных, называется:

Обучение с учителем

Обучение без учителя

Обучение с подкреплением

Машинное обучение

31. Библиотека языка Python, добавляющая поддержку больших многомерных массивов и матриц, вместе с большой библиотекой высокоуровневых (и очень быстрых) математических функций для операций с этими массивами, называется:

pandas

numpy

matplotlib

sklearn

32. Библиотека Python, позволяющая строить сводные таблицы, выполнять группировки, предоставляет удобный доступ к табличным данным, называется:

pandas

numpy

matplotlib

sklearn

33. Библиотека Python, которая предоставляет множество возможностей, таких как многоступенчатый анализ, регрессия и алгоритмы кластеризации, называется:

pandas

numpy

matplotlib

sklearn

34. Библиотека Python, предназначенная для визуализации данных, называется

pandas

numpy

matplotlib

sklearn

35. Метод, который задает начальные условия для генератора случайных чисел, называется:

random.seed()

np.median()

np.median()

np.arange()

36. "Надстройка» над Matplotlib, которая предоставляет лучшую графику и большее количество возможностей её настройки, называется

plot

graphic

seaborn

diagrams

37. Открытая программная библиотека для машинного обучения, разработанная компанией Google для решения задач построения и тренировки нейронной сети, называется:

NeiroNet

NeiroLib

TensorFlow

FlowKeras

38. В каком отношении обычно делят выборку на обучающую и тестовую:

20:80

80:20

50:50

90:10

39. Что такое "переобучение" модели?

Модель содержит чрезмерно большое число переменных

Модель слишком часто участвовала в обучении

Модель излишне точно соответствует сети конкретному набору обучающих примеров и теряет способность к обобщению.

40. Среда программирования Python, включающая набор свободных библиотек машинного обучения, называется:

Cobra

Anaconda

MachineLearning

PythonMLLibrary

41. Инструмент для разработки и представления проектов Data Science в интерактивном виде, объединяющий код, текст, математические уравнения и визуализации, называется:

Mars Notebook

Jupyter Notebook

Venera Notebook

Pluton Notebook

42. Датасет - это

Сет данных

Обработанная и структурированная информация в табличном виде

Усредненная выборка данных по строкам

Усредненная выборка данных по столбцам

43. Как называется бесплатная среда Google для создания ноутбуков Jupyter, которая полностью работает в облаке?

Laboratory

Googlaboratory

Colaboratory

Neirabolatory

44. Какой тип ячеек в ноутбуках Jupyter предназначен для ввода текста и изображений?

Code

Markdown

Memo

Image

45. Какой тип ячеек в ноутбуках Jupyter предназначен для ввода программного кода?

Markdown

Memo

Image

Code

2 ЭТАП – УМЕТЬ

Комплект практических работ

Практические работы служат для работы студентов над учебными задачами с целью выработки и закрепления практических навыков.

Тема 1. Основы анализа данных.

Задание

Постройте диаграмму Паретто. Проанализируйте использование основной и вспомогательной оси в MS Excel.

Построение диаграммы Парето

В бизнес-анализе часто используют диаграмму Парето, позволяющую показать, например, какие товары или клиенты приносят наибольшую прибыль.

Диаграмма Парето – это инструмент, позволяющий распределить усилия для решения возникающих проблем и выявить основные причины, с которых нужно начинать действовать. Метод анализа Парето заключается в классификации проблем качества на немногочисленные, но существенно важные и многочисленные, но несущественные. Он позволяет распределить усилия и установить основные факторы, с которых нужно начинать действовать.

Построение диаграммы Парето начинают с классификации возникающих проблем по отдельным факторам (например, проблемы, относящиеся к браку; проблемы, относящиеся к работе оборудования или исполнителей, и т.д.) Затем следуют сбор и анализ статистического материала по каждому фактору, чтобы выяснить, какие из этих факторов являются преобладающими при решении проблем.

Вариант 1. Построение диаграммы Парето по готовым данным

Введите в MS Excel таблицу, показанную ниже. В таблице представлены данные по товарам, приносящим определенную прибыль магазину.

	А	В
1	Товар	Прибыль
2	Абрикос	75855
3	Лук	76734
4	Нектарин	97275
5	Картофель	12797
6	Апельсин	88043
7	Свекла	14997
8	Баклажан	70102
9	Киви	30012
10	Капуста	107820
11	Салат	25487
12	Мандарины	76721
13	Банан	106757
14	Малина	64750
15	Банан	30507
16	Огурец	98480
17	Баклажан	55019

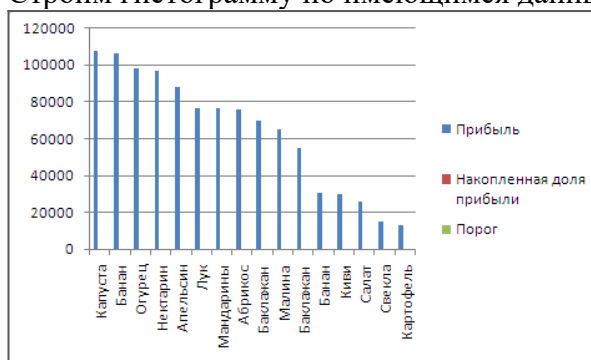
Отсортируйте таблицу по убыванию прибыли (вкладка **Данные - Сортировка**).

Добавьте столбец с формулой для расчета накопленного процента прибыли. Эта формула делит суммарную накопленную прибыль с начала списка до текущего товара на общую прибыль по всей таблице.

	А	В	С
1	Товар	Прибыль	Накопленная доля прибыли
2	Капуста	107820	10,45%
3	Банан	106757	20,81%
4	Огурец	98480	30,35%
5	Нектарин	97275	39,79%
6	Апельсин	88043	48,32%
7	Лук	76734	55,76%
8	Мандарины	76721	63,20%
9	Абрикос	75855	70,56%
10	Баклажан	70102	77,35%
11	Малина	64750	83,63%
12	Баклажан	55019	88,97%
13	Банан	30507	91,92%
14	Киви	30012	94,83%
15	Салат	25487	97,31%
16	Свекла	14997	98,76%
17	Картофель	12797	100,00%

Добавьте столбец с константой 80% для создания в будущей диаграмме горизонтальной пороговой линии.

Строим гистограмму по имеющимся данным.



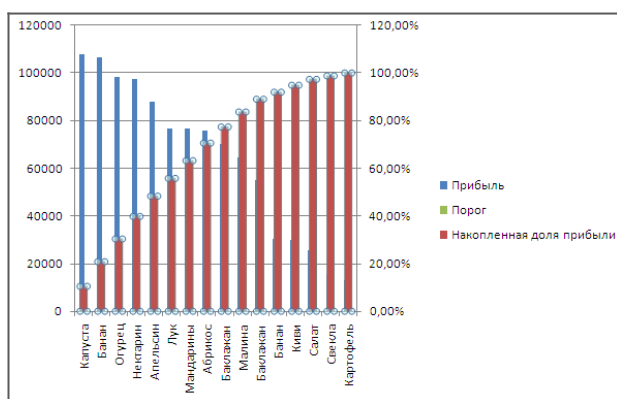
Далее ряды с процентами на полученной диаграмме необходимо разместить на вторичной (правой) оси. Для этого нужно выделить ряды накопленной доли прибыли мышью, но это может быть сложно, поскольку их плохо видно на фоне больших столбцов прибыли.

Воспользуемся для выделения выпадающим списком на вкладке **Работа с диаграммами – Макет** и выбираем ряд **Накопленная доля прибыли**:

Затем щелкните по выделенному ряду правой кнопкой мыши и выберите команду **Формат ряда данных**

В появившемся окне выбрать опцию **По вторичной оси**.

В итоге наша диаграмма начнет выглядеть так:



Аналогично разместите на вторичной оси ряд Порог.

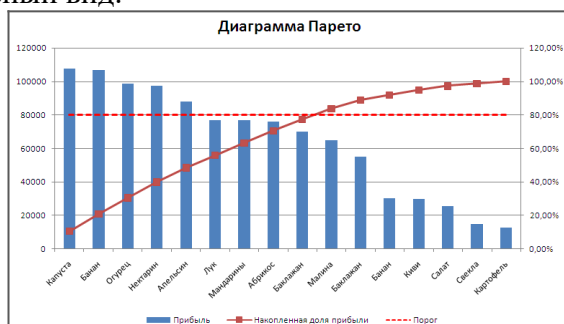
Для рядов Накопленная доля прибыли и Порог необходимо поменять тип диаграммы со столбцов на линию. Для этого щелкните по каждому из этих рядов и выберите команду Изменить тип диаграммы для ряда и выберите тип График.

Аналогично поступите с рядом Порог.

Отформатируем ряд Порог так, чтобы он стал похож на граничную линию, а не на данные (уберем маркеры, сделаем линию красной, пунктирной и т.д.).

Все это можно сделать, щелкнув по ряду правой кнопкой мыши и выбрав команду **Формат ряда**.

Добавьте название, перенесите легенду вниз. Теперь диаграмма примет окончательный вид:



По диаграмме можно сделать вывод, какие товары приносят 80% всей прибыли, а на какие товары приходится только 20% прибыли.

Задание для самостоятельной работы

Исследуется проблема дефектных изделий – печатных плат. В качестве типа данных для анализа выбраны дефекты, которые детализированы по видам. Единица измерений дефектов – процент от общего числа дефектов.

Постройте диаграмму Парето и проанализируйте, на что нужно направить максимальные усилия и ресурсы.

Вид дефектов на печатной плате	Процент от общего числа дефектов
Недостаток припоя	58
Трещины	20
Царапины	10
Пятна припоя	6
Тепловая деформация	3
Расслоение платы	2
Прочее	1
Итого:	100

Тема 2. Визуализация данных

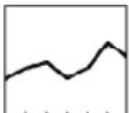
Задание 1

В настоящее время выделяют пять основных типов сравнения данных (метод Джина Желязны):

- покомпонентное – подойдёт в том случае, если нужно сравнить между собой распределение составных частей какого-то целого значения, например, вклад каждой реализуемой товарной группы в общий доход от продаж;
- позиционное – позволяет сравнить значения отдельных позиций друг с другом, например, выявить самый прибыльный филиал за прошлый месяц;
- временное – отражает динамику данных, например, изменение уровня конверсии сайта по дням за текущий месяц;
- частотное – показывает распределение общего массива данных по заданным диапазонам, например, сколько сотрудников компании получает зарплату от 50 до 100 тыс. рублей;
- корреляционное – выявляет наличие зависимости между данными, например, влияет ли количество рекламных роликов на объём продаж.

Для отображения зависимостей можно обойтись пятью базовыми диаграммами:

- линейчатой диаграммой;
- круговой диаграммой;
- линейным графиком;
- столбиковой диаграммой;
- точечной диаграммой (рассеивания).

		ТИПЫ СРАВНЕНИЯ				
		ПОКОМПОНЕНТНОЕ	ПОЗИЦИОННОЕ	ВРЕМЕННОЕ	ЧАСТОТНОЕ	КОРРЕЛЯЦИОННОЕ
ОСНОВНЫЕ ТИПЫ ДИАГРАММ	КРУГОВАЯ					
	ЛИНЕЙЧАТАЯ					
	ГИСТОГРАММА					
	ГРАФИК					
	ТОЧЕЧНАЯ					

Выберите подходящий тип диаграммы в соответствии с матрицей и составьте диаграммы, которые вы бы использовали для иллюстрации соответствующих идей.

1. В течение следующих десяти лет прогнозируется увеличение объема продаж Временное	2. Большинство сотрудников получает от 30 до 35 тыс. долл. Частотное
3. Более высокие цены на отдельные марки бензина не означают более высокое качество Корреляционное	4. В сентябре уровень текучести кадров в шести подразделениях был примерно одинаков Позиционное
5. Менеджер по продажам проводит с клиентами лишь 15 % своего времени Покомпонентное	6. Размер прибавки по результатам работы не зависит от выслуги лет Корреляционное

7. В прошлом году наибольшая текучесть кадров наблюдалась в возрастной группе от 30 до 35 лет Частотное i	8. Центральный регион занимает последнее место по производительности Позиционное
9. Доходность акций нашей компании падает Временное	10. Большая часть всех фондов задействована в производстве Покомпонентное
11. Наблюдается связь между рентабельностью и зарплатами Корреляционное	12. В августе два завода обогнали по производительности шесть других Позиционное

Задание 2

На основе приведенных данных начертите диаграмму, показывающую долю «Кричигрома» в общем объеме продаж воображаемой отрасли игрушечной промышленности в 2001 г. - в сравнении с конкурентами.

Подсказка

Ключевыми являются слова «доля в общем объеме продаж отрасли в 2001 г.». Они указывают на покомпонентное сравнение (размеры каждой доли в процентах от целого), а также говорят о том, что нужно использовать круговую диаграмму, поскольку мы говорим о целом. Компоненты здесь расположены по часовой стрелке - от компании с наибольшей долей продаж до компании с наименьшей долей. «Кричигром» в соответствии с данными занимает второе место. Для того, чтобы выделить долю нашей компании, мы использовали закрашивание соответствующего ей сегмента.

Доля каждой компании в общем объеме продаж отрасли в 2001 г.:

Кричигром 19,3%
Компания А 10,1 %
Компания Б 16,6%
Компания В 12,4 %
Компания Г 31,8%
Компания Д 9,8 %.

Задание 3

Начертите диаграмму, которая показывает, какое место в отрасли в 2001 г. занимал «Кричигром» по показателю доходности активов.

Подсказка

В задании 2 слова «какое место» четко указывают на позиционное сравнение. Здесь мы хотим знать, у какой из компаний наибольший показатель доходности активов, а у какой - наименьший. В данном случае у «Кричигрома» – наименьший. Обратите внимание, насколько эффективно линейчатая диаграмма отражает основную идею: «Кричигром» расположен внизу списка. Эффект усилен при помощи темной заливки.

Коэффициент доходности активов в 2001 г.:

«Кричигром» 8,3 %
Компания А 9,8 %
Компания Б 15,9%
Компания В 22,4 %
Компания Г 14,7%
Компания Д 19,1 %

Задание 4

Начертите диаграмму, демонстрирующую, существует ли взаимосвязь между долей продаж и показателем доходности активов в воображаемой отрасли игрушечной промышленности в 2001 г.

Подсказка

В большинстве случаев мы предполагаем наличие взаимосвязи между долей продаж и показателем доходности активов: чем больше доля, тем больше показатель доходности. В данном же случае подобной взаимозависимости не наблюдается.

Например, доля продаж «Кричигрома» – вторая по величине, однако коэффициент доходности активов у этой компании намного меньше, чем, скажем, у компании В, которая занимает лишь четвертое место по доле продаж. В данном случае можно было бы использовать и точечную диаграмму, но двусторонняя линейчатая диаграмма позволяет подписать данные более наглядно.

Доля продаж каждой компании в 2001 г.

«Кричигром» 19,3%

Компания А 10,1 %
 Компания Б 16,6%
 Компания В 12,4%
 Компания Г 31,8%
 Компания Д 9,8 %
 Коэффициент доходности активов в 2001 г.
 «Кричигром» 8,3 %
 Компания А 9,8 %
 Компания Б 15,9%
 Компания В 22,4 %
 Компания Г 14,7%
 Компания Д 19,1 %

Задание 5

Начертите диаграмму, показывающую тенденции изменения объема продаж и прибыли компании «Кричигром» в 1997-2001 гг. При этом 1997 г. примите за основу, а последующие годы покажите в процентном соотношении с ним.

Подсказка

В задании требуется показать изменения во времени - временное сравнение, - и для этой цели лучше всего подходит график. Мы перевели абсолютные числа в проценты от базовой величины (в данном случае - данные за 1997 г.), чтобы обеспечить общую и более понятную основу для сравнения двух различных величин - 530 и 29 млн. долл. Это отличный пример преимущества диаграмм перед табличными данными. На графике ясно видна нестабильная модель изменения прибыли, которая была бы не столь наглядной в таблице.

Год	Объем продаж	Прибыль
	Млн. %	Млн. %
	долл. 1997 = 100%	долл. 1997 = 100%
1997	387 100	1997 24 100
1998	420 109	1998 39 162
1999	477 123	1999 35 146
2000	513 133	2000 45 188
2001	530 137	2001 29 121

Задание 6

Начертите диаграмму, показывающую, что в 2001 г. основную долю наших продаж, в отличие от компании Х, составили наименее дорогие модели трансформеров.

Подсказка

Задание отсылает нас к частотному сравнению, поскольку мы сравниваем количество единиц товара, проданного в пределах определенных ценовых рамок. В данном случае мы использовали ступенчатую гистограмму, выделив столбцы, относящиеся к "Кричигрому", заливкой. Конечно, можно было бы показать то же самое и на графике с двумя линиями, но с таким небольшим количеством показателей гистограмма предпочтительнее.

Количество проданных единиц товара, тыс.

Цена, долл.	«Кричигром»	Компания Х
Менее 5,00	320	280
5,00-9,99	770	340
10,00-14,99	410	615
15,00-19,99	260	890
20,00 и более	105	550

Задание 7

Начертите диаграмму, показывающую, насколько доля различной продукции в

объеме продаж нашей компании отличались от тех же показателей нашего конкурента в 2001 г.

Процент общих продаж по видам продукции в 2001 г.

Подсказка

В этом задании мы возвращаемся к покомпонентному сравнению, т. е. отображению процентов от общих продаж. Как показывает матрица, мы можем использовать круговую диаграмму. Однако, поскольку мы имеем дело не с одним целым (продажи «Кричигрома» и компании X), лучше воспользоваться гистограммой, нормированной на 100 %. В этом случае нам не понадобятся лишние подписи и легенда, а взаимосвязи между сегментами будут более наглядными.

	Продукция «Кричигром»	Компания Г
Куклы	15.0%	25.3 %
Мячики	8.4 %	21.3%
Покемоны	20.6 %	19.9%
Хаги Ваги	16.2%	18.6 %
Бакуган	39.8 %	14.9%

Тема 3. Машинное обучение и интеллектуальный анализ данных

Написание реферата является обязательным элементом работы студентов в рамках освоения курса «Обработка естественного языка и компьютерное зрение». Реферат (от лат. «refere» - докладываю, сообщаю) - это самостоятельная исследовательская работа, в которой автор раскрывает суть исследуемой проблемы; приводит различные точки зрения, а также собственные взгляды на нее. Содержание реферата должно быть логичным; изложение материала носит проблемно-тематический характер.

Темы для рефератов:

1. Интеллектуальный анализ данных.
2. Искусственные нейронные сети в интеллектуальном анализе данных.
3. Деревья решений как метод интеллектуального анализа данных.
4. Метод «к-ближайших соседей».
5. Метод опорных векторов.
6. Байесовская классификация.
7. Линейная регрессия.
8. Корреляционно-регрессионный анализ.
9. Иерархические методы кластерного анализа.
10. Неиерархические методы кластерного анализа, в том числе алгоритмы k-средних и k-медианы.
11. Поиск ассоциативных правил.
12. Метод ограниченного перебора.
13. Эволюционное программирование и генетические алгоритмы.
14. Основные технологии интеллектуального анализа данных в маркетинге.
15. Основные понятия и определение нечеткой логики.
16. Основные определения и свойства генетических алгоритмов, их отличия от традиционных методов.

Структура реферата:

- 1) Ключевые слова.
- 2) Аннотация содержания (2-3 предложения).
- 3) Введение (не более 2 страниц). Во введении необходимо обосновать актуальность темы, очертить область исследования, объект исследования, основные цели и задачи исследования, сформулировать выдвигаемые гипотезы.
- 4) Основная часть состоит из 2-3 разделов. В них раскрывается суть исследуемой

проблемы, проводится обзор информации по предмету исследования. Изложение материала не должно ограничиваться лишь описательным подходом к раскрытию выбранной темы. Оно также должно содержать собственное видение рассматриваемой проблемы.

5) Заключение (1-2 страницы). В заключении кратко излагаются выводы, а также предполагаемые научные результаты и прогнозы.

6) Библиографический список (от 5 до 10 источников) в алфавитном порядке. В данный список рекомендуется включать работы отечественных и зарубежных авторов. Библиографический список содержит только те произведения, на которые есть сноски в тексте.

7) Приложение (при необходимости).

Создание презентации по заданной теме

Мультимедийные презентации используются для того, чтобы обучающийся смог наглядно продемонстрировать визуальные (аудио, видео, графические) материалы, освоенные в ходе самостоятельной и практической работы по предмету.

Общие требования к презентации:

Презентация не должна быть меньше 10 слайдов.

Первый слайд – титульный лист, на котором обязательно должны быть представлены: тема; фамилия, имя, автора, номер учебной группы;

Второй слайд – содержание, где представлены основные вопросы разобранные в ходе изучения темы. Желательно, чтобы из содержания по гиперссылке можно перейти на необходимую страницу и вернуться вновь на содержание.

В структуре презентации необходимо использовать: графическую и анимационную информацию: видео и аудио фрагменты, таблицы, диаграммы, инфографику и т.д.

Последний слайд демонстрирует список ссылок на, используемые информационные ресурсы.

Лабораторная работа «Кластерный анализ с помощью метода k-means»

Создадим исходные точки.

1. Импортируем библиотеки

```
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
from sklearn.datasets import make_blobs
from sklearn.cluster import KMeans
```

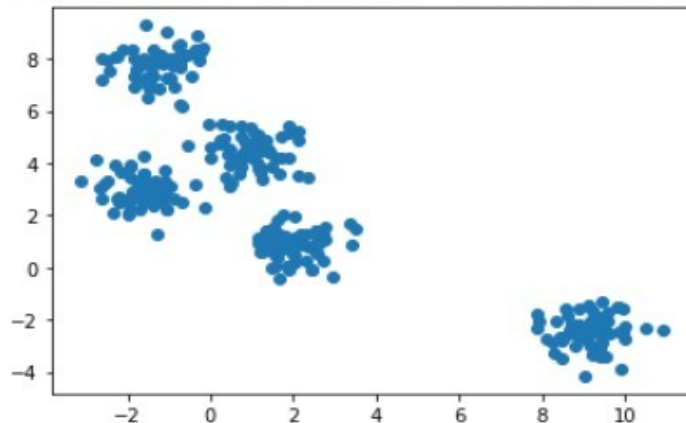
2. Создаем тестовый набор данных для моделирования Будем использовать функцию make_blobs().

Функция make_blobs() генерирует данные из изотропных гауссовых распределений. В качестве аргумента можно указать **количество объектов, количество центров и стандартное отклонение каждого кластера**.

Проанализируйте построение графика.

```
X, labels = make_blobs(n_samples=300, centers=5, cluster_std=0.60, random_state=0)
plt.scatter(X[:,0], X[:,1])
```

<matplotlib.collections.PathCollection at 0x7f3e78990150>



3. Запускаем алгоритм кластеризации – обучения модели методом k-means

```
#Запускаем алгоритм кластеризации – обучения модели методом k-means
kmeans = KMeans(n_clusters=5, init='k-means++', max_iter=300, n_init=10, random_state=0)
pred_y = kmeans.fit_predict(X)
```

Параметры обучения:

n_clusters целое, по умолчанию = 8

Количество кластеров, которые нужно сформировать, а также количество центроидов, которые нужно сгенерировать.

init {'k-means++', 'random'}, вызываемый или подобный массиву формы (n_clusters, n_features), default='k-means++' Способ инициализации:

'k-means++': разумно выбирает начальные центры кластеров для кластеризации k-mean, чтобы ускорить сходимость. См. раздел Примечания в k_init для более подробной информации.

«случайный»: выбор n_clusters наблюдений (строк) случайным образом из данных для начальных центроидов.

max_iter целое, по умолчанию = 300

Максимальное количество итераций алгоритма k-средних за один прогон.

n_init целое, по умолчанию = 10

Количество раз, когда алгоритм k-средних будет запускаться с разными начальными центроидами. Окончательные результаты будут лучшим результатом n_init последовательных запусков с точки зрения инерции.

random_state int, экземпляр RandomState или None, по умолчанию = None

Определяет генерацию случайных чисел для инициализации центроида.

После обучения получим атрибуты:

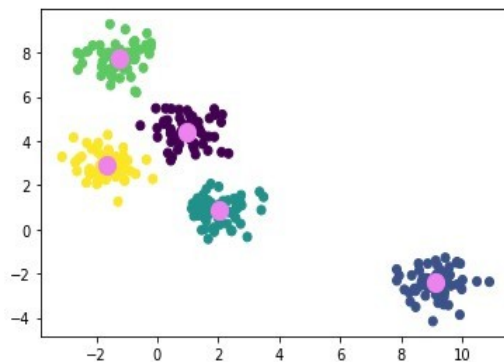
cluster_centers_ ndarray формы (n_clusters, n_features)

Координаты центров кластеров. Если алгоритм останавливается до полной сходимости (см. toli max_iter), они не будут согласовываться с labels_.

Выражение pred_y = kmeans.fit_predict(X) предсказывает ближайший кластер, к которому принадлежит каждый образец в X.

4. Визуализируем результат обучения

```
plt.scatter(X[:,0], X[:,1], c=pred_y)
plt.scatter(kmeans.cluster_centers_[0], kmeans.cluster_centers_[1], s=150, c='violet')
plt.show()
```



Проанализируйте, как мы добились того, что кластеры окрашены в разные цвета.

Для того, чтобы посмотреть, к какому кластеру по прогнозу относится конкретная точка, можно ввести:

```
Labels=kmeans.labels_
```

```
print(Labels[230])
```

```
3
```

5. Для того, чтобы получить массив центров кластеров, необходимо преобразовать его к типу DataFrame

```
list_tuples=list(zip(kmeans.cluster_centers_[0], kmeans.cluster_centers_[1]))
```

```
print(list_tuples)
```

```
[(-0.37461184624978294, 3.673352543706669), (9.122391035184616, -2.4282218470473773),
```

```
dframe=pd.DataFrame(list_tuples, columns=['x', 'y'])
```

```
print(dframe)
```

```

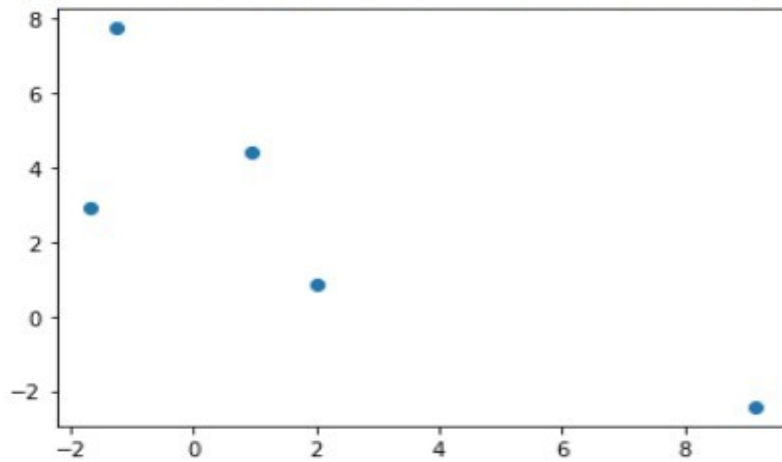
      x      y
0 -0.374612  3.673353
1  9.122391 -2.428222
2  2.009493  0.886137
3 -1.242689  7.753172

```

Визуализируем результат:

```
plt.scatter(dframe['x'], dframe['y'])
```

```
<matplotlib.collections.PathCollection at 0x7fe0d1ff0dd0>
```



Задания для самостоятельной работы

1. Создайте новые кластеры, указав параметр `random_state` равным любому значению (например, 5, 10, 25, 30 и т.д.).
2. Задайте 1 итерацию, 2 итерации и т.д. Посмотрите на положение центров кластеров.
3. Изучите генерацию различных датасетов, меняя количество кластеров и их дисперсию.

```

fig,ax = plt.subplots(nrows=1, ncols=3,figsize=(16,5))
plt_ind_list = np.arange(3)+131

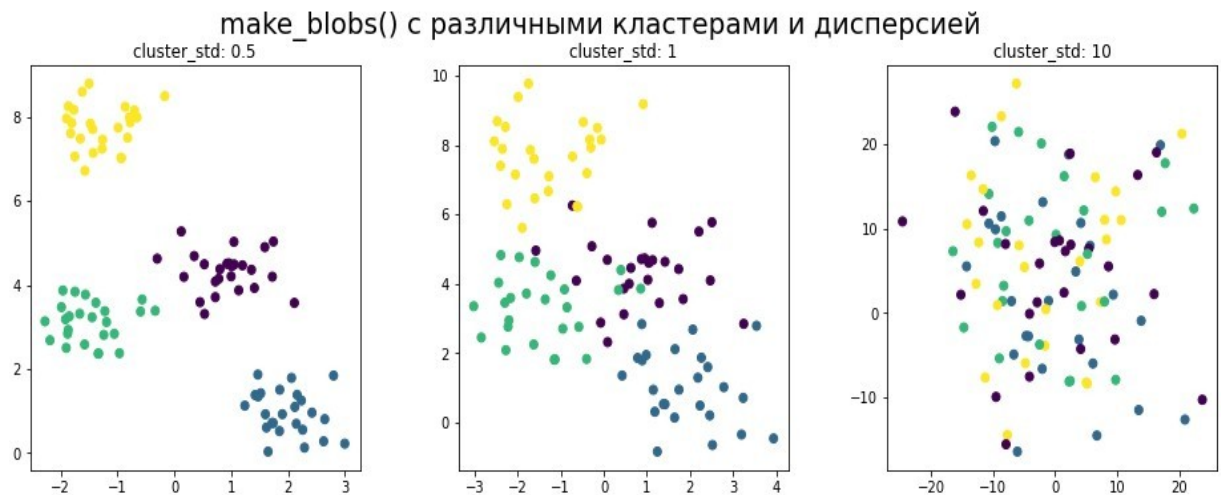
for std,plt_ind in zip([0.5,1,10],plt_ind_list):
    x, label = make_blobs(n_features=2,
                          centers=4,
                          cluster_std=std,
                          random_state=0)

    plt.subplot(plt_ind)
    my_scatter_plot = plt.scatter(x[:,0],
                                  x[:,1],
                                  c=label,
                                  vmin=min(label),
                                  vmax=max(label))

    plt.title('cluster_std: '+str(std))

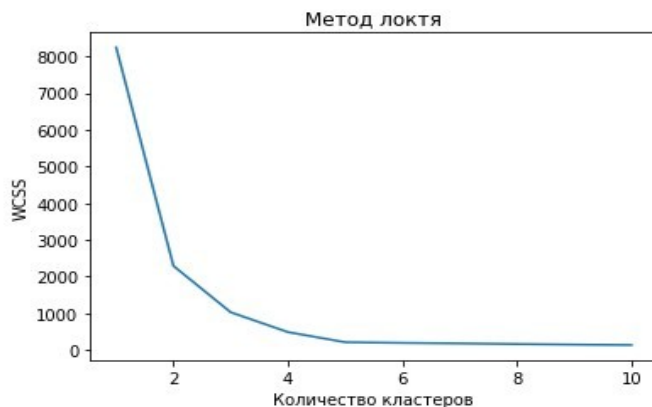
fig.subplots_adjust(hspace=0.3,wspace=.3)
plt.suptitle('make_blobs() с различными кластерами и дисперсией ',fontsize=20)
plt.show()

```



4. Изучите метод локтя для определения оптимального количества кластеров.

```
# методом локтя определим оптимальное количество кластеров
wcss = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters=i, init='k-means++', max_iter=300, n_init=10, random_state=0)
    kmeans.fit(X)
    wcss.append(kmeans.inertia_)
plt.plot(range(1, 11), wcss)
plt.title('Метод локтя')
plt.xlabel('Количество кластеров')
plt.ylabel('WCSS')
plt.show()
```



3 ЭТАП – ВЛАДЕТЬ

Вопросы к зачету

1. Интеллектуальный анализ данных.
2. Искусственные нейронные сети в интеллектуальном анализе данных.
3. Деревья решений как метод интеллектуального анализа данных.
4. Метод «к-ближайших соседей».
5. Метод опорных векторов.
6. Байесовская классификация.
7. Линейная регрессия.
8. Информационно-аналитическая модель (дашборд). Принципы построения.
9. Проектирование дашборда.
10. Пять основных типов сравнения данных (метод Джина Желязны).
11. Облачные сервисы анализа и визуализации данных.
12. Программное обеспечение для анализа и визуализации данных.

Контрольные вопросы

Тема 1. Основы анализа данных

1. Роль анализа данных в современном мире.
2. Понятие «данные» и «анализ данных». Этапы анализа данных.
3. Основы обработки данных. Объекты и признаки. Типы шкал, наборы данных.
4. Линейные и нелинейные модели.
5. Сбор и подготовка данных. Источники данных.
6. Оценка точности полученных результатов.

Тема 2. Визуализация данных

1. Визуализация как инструмент представления аналитической информации.

Основные понятия и определения визуального анализа данных Цели и задачи визуализации данных.

2. Визуализация бизнес-информации.
3. Информационная аналитическая панель (дашборд).
4. Работа с моделью данных.
5. Нестандартная визуализация.

Тема 3. Машинное обучение и интеллектуальный анализ данных

1. Основы искусственного интеллекта. Интеллектуальный анализ данных.
2. Основы программирования для задач анализа данных. Основы машинного обучения.
3. Обучение «с учителем», «без учителя», «с подкреплением». Задачи классификации, кластеризации, регрессии.
4. Анализ качества построенных моделей в Python

4. МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ, ОПРЕДЕЛЯЮЩИЕ ПРОЦЕДУРЫ ОЦЕНИВАНИЯ ЗНАНИЙ, УМЕНИЙ, НАВЫКОВ И (ИЛИ) ОПЫТА ДЕЯТЕЛЬНОСТИ, ХАРАКТЕРИЗУЮЩИХ ЭТАПЫ ФОРМИРОВАНИЯ КОМПЕТЕНЦИЙ В ПРОЦЕССЕ ОСВОЕНИЯ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

1 ЭТАП – ЗНАТЬ

Методические рекомендации по решению тестовых заданий

1. Отвечая на вопрос с одним правильным ответом, щелкните на радиокнопке рядом с правильным ответом.
2. Отвечая на вопрос с несколькими правильными вариантами ответа, щелкните на чекбоксах рядом со всеми правильными ответами.
3. Отвечая на вопросы, требующие определить правильную последовательность, впишите по порядку номера в поле ввода рядом с ответом.
4. Вопросы в тесте – 30.
5. Время на выполнение теста – 40 минут.

Критерии оценивания решения тестовых заданий

Оценка	Правильность (ошибочность) выполнения задания
«отлично»	91-100% правильных ответов
«хорошо»	74-90% правильных ответов
«удовлетворительно»	59-73% правильных ответов
«неудовлетворительно»	58% и менее правильных ответов

2 ЭТАП – УМЕТЬ

Критерии оценивания лабораторных работ

Оценка	Правильность (ошибочность) выполнения задания
«решение зачтено»	Использованы все графические ресурсы. Выполнены все этапы технического задания
«решение не зачтено»	Не выполнены этапы технического задания

Критерии оценки по выполнению практических заданий

Оценка «отлично» выставляется студенту, если он выполнил работу на 90%, при защите работы студент показывает глубокое знание вопросов темы.

Оценка «хорошо» выставляется студенту, если он выполнил работу на 75%, при защите работы студент без затруднений отвечает на вопросы.

Оценка «удовлетворительно» выставляется студенту, если он выполнил работу на 60% с незначительными ошибками; при защите показывает слабое знание вопросов темы,

не всегда дает исчерпывающие и обоснованные вопросы на заданные вопросы, допускает существенные ошибки.

Оценка «неудовлетворительно» – работа не выполнена.

Задание на реферативное изложение материала

Требования к оформлению реферата:

- текст на формате А4, с одной стороны листа;
- шрифт Times New Roman;
- кегль шрифта 14;
- межстрочное расстояние 1,5;
- поля: сверху 2 см, снизу - 2 см, слева - 3 см, справа 1,5 см;
- реферат должен быть представлен в сброшюрованном виде;
- формат абзаца текста должен быть выровнен «по ширине» положения на странице. Абзацный отступ первой строки каждого абзаца должен быть равен 1,25 см;
- номер страницы проставляется арабскими цифрами в центре верхней части листа без точки, начиная с введения (3 страница). На титульном листе и на Содержании страницы не ставится;
- титульный лист оформляется в соответствии с образцом оформления реферата, курсовой работы, выпускной квалификационной работы, принятым в ЧОУВО МИДиС

Критерии оценивания реферата

Оценка	Правильность (ошибочность) выполнения задания
«зачтено»	Содержание реферата соответствует теме; Правильное использования источников литературы; Соответствие оформления реферата стандартом; Знание учащимся изложенного в реферате материала (для ответа на устные вопросы); Умение грамотно и аргументировано изложить суть проблемы в тексте и устно; Тезисы и предположения аргументированы, сделаны выводы и умозаключения; В тексте отражено умение анализировать фактический материал.
«не зачтено»	Тема реферата раскрыта недостаточно полно; Отсутствует библиографический список; Тезисы и предположения не аргументированы, не сделаны выводы и умозаключения; В тексте отсутствует фактологический анализ; Ответы на устные вопросы не отражают умение грамотно и аргументировано изложить суть реферата.

Критерии оценивания заданий для самостоятельной работы

Оценка	Правильность (ошибочность) выполнения задания
«решение зачтено»	Использованы все графические ресурсы. Выполнены все этапы технического задания
«решение не зачтено»	Не выполнены этапы технического задания

3 ЭТАП – ВЛАДЕТЬ

Критерии оценивания знаний на зачете

Оценка «ЗАЧТЕНО»:

1. Усвоение программного материала.
2. Умение применять основные приемы и методы обработки данных.
3. Выполнение практических заданий и самостоятельной работы за семестр.
4. Точность и обоснованность выводов.
5. Точные, полные и логичные ответы на дополнительные вопросы.

Оценка «НЕ ЗАЧТЕНО»:

1. Незнание значительной части программного материала
2. Невыполнение практических заданий и самостоятельной работы за семестр.
3. Грубые ошибки при выполнении практических заданий и самостоятельной работы.
4. Неумение выделить главное, сделать выводы и обобщения.
5. Неправильные ответы на дополнительные вопросы.

Критерии оценки ответов студентов на контрольные вопросы

Оценка	Правильность (ошибочность) выполнения задания
«отлично»	полно и аргументировано отвечает по содержанию задания; понимает материал, обосновывает свои суждения, применяет знания на практике, иллюстрирует ответ примерами не только по предложенной литературе; излагает материал последовательно и правильно.
«хорошо»	полно и аргументировано отвечает по содержанию задания; понимает материал, обосновывает свои суждения, применяет знания на практике, иллюстрирует ответ примерами не только по предложенной литературе; излагает материал последовательно и правильно, но допускает 1-2 ошибки, которые исправляет сам.
«удовлетворительно»	знает и понимает основные положения по содержанию задания; излагает материал неполно, но допускает неточности в определении понятий или формулировке правил; не достаточно глубоко и доказательно обосновать свои суждения иллюстрирует ответ примерами только по предложенной литературе; излагает материал непоследовательно и допускает 3-4 ошибки.
«неудовлетворительно»	допускает существенные ошибки в формулировке определений и правил, искажающие их смысл; излагает материал непоследовательно, неуверенно и не по существу задания; допускает существенные ошибки, не позволяющие раскрыть смысл задания, являющиеся серьезным препятствием к успешному овладению следующим материалом.