

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Сыров Игорь Анатольевич
Должность: Директор
Дата подписания: 30.10.2023 11:00:29
Уникальный программный ключ:
b683afe664d7e9f64175886cf9626a196149ad36

СТЕРЛИТАМАКСКИЙ ФИЛИАЛ
ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО БЮДЖЕТНОГО ОБРАЗОВАТЕЛЬНОГО
УЧРЕЖДЕНИЯ ВЫСШЕГО ОБРАЗОВАНИЯ
«УФИМСКИЙ УНИВЕРСИТЕТ НАУКИ И ТЕХНОЛОГИЙ»

Факультет Математики и информационных технологий
Кафедра Математического моделирования

Оценочные материалы по дисциплине (модулю)

дисциплина Методы интеллектуального анализа данных

Блок Б1, часть, формируемая участниками образовательных отношений, Б1.В.07
цикл дисциплины и его часть (обязательная часть или часть, формируемая участниками образовательных отношений)

Направление

02.03.03 Математическое обеспечение и администрирование информационных систем

код наименование направления

Программа

Сетевое программирование и администрирование информационных систем

Форма обучения

Очная

Для поступивших на обучение в
2023 г.

Разработчик (составитель)
кандидат химических наук, доцент кафедры математического моделирования
Иремадзе Э. О.

ученая степень, должность, ФИО

1. Перечень компетенций, индикаторов достижения компетенций и описание показателей и критериев оценивания результатов обучения по дисциплине (модулю)	3
2. Оценочные средства, необходимые для оценки результатов обучения по дисциплине (модулю)	6
3. Методические материалы, определяющие процедуры оценивания результатов обучения по дисциплине (модулю), описание шкал оценивания	15

1. Перечень компетенций, индикаторов достижения компетенций и описание показателей и критериев оценивания результатов обучения по дисциплине (модулю)

Формируемая компетенция (с указанием кода)	Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине (модулю)	Показатели и критерии оценивания результатов обучения по дисциплине (модулю)				Вид оценочного средства	
			4					5
			1	2	3	4		
			неуд.	удовл.	хорошо	отлично		
ПК-1. Способен использовать основные методы и средства автоматизации, связанные с разработкой, сопровождением и администрированием программных продуктов и информационных систем	ПК-1.1. Знание	Обучающийся должен знать: основные понятия, методы и технологии в области автоматизированной обработки, визуализации и хранения больших объемов данных	Не обладает знаниями основных понятий, методов и технологий в области автоматизированной обработки, визуализации и хранения больших объемов данных	Фрагментарное знание основных понятий, методов и технологий в области автоматизированной обработки, визуализации и хранения больших объемов данных	В целом успешное, но с пробелами в знании основных понятий, методов и технологий в области автоматизированной обработки, визуализации и хранения больших объемов данных	Успешное знание основных понятий, методов и технологий в области автоматизированной обработки, визуализации и хранения больших объемов данных		
	ПК-1.2. Умение	Обучающийся должен уметь: применять методы обработки больших объемов данных	Не умение применять методы обработки больших объемов данных	Фрагментарное применение умений методов обработки больших объемов данных	В целом успешное, но содержащее отдельные пробелы умение применять	Успешное и систематическое умение применять методы обработки		

		объемов данных и использовать программные средства с технологией Big Data при решении практических задач;	и использовать программные средства с технологией Big Data при решении практических задач.	и использовать программные средства с технологией Big Data при решении практических задач.	методы обработки больших объемов данных и использовать программные средства с технологией Big Data при решении практических задач.	больших объемов данных и использовать программные средства с технологией Big Data при решении практических задач.	
	ПК-1.3. Владение	Обучающийся должен владеть: понятиями и методами обработки больших объемов данных и машинного обучения с применением современных программных продуктов; навыками применения современных методов разработки технического, информационног	Не владение понятиями и методами обработки больших объемов данных и машинного обучения с применением современных программных продуктов; навыками применения современных методов разработки технического, информационног о и	Фрагментарное владение понятиями и методами обработки больших объемов данных и машинного обучения с применением современных программных продуктов; навыками применения современных методов разработки технического, информационног	В целом успешное, но с пробелами владение понятиями и методами обработки больших объемов данных и машинного обучения с применением современных программных продуктов; навыками применения современных методов разработки	Успешное владение понятиями и методами обработки больших объемов данных и машинного обучения с применением современных программных продуктов; навыками применения современных методов разработки технического, информационног	

		о и алгоритмическог о обеспечения Big Data.	алгоритмическог о обеспечения Big Data.	о и алгоритмическог о обеспечения Big Data.	технического, информационног о и алгоритмическог о обеспечения Big Data.	о и алгоритмическог о обеспечения Big Data.	
--	--	---	---	---	--	---	--

2. Оценочные средства, необходимые для оценки результатов обучения по дисциплине (модулю)

Перечень вопросов для оценки уровня сформированности компетенции ПК-1 на этапе «Знания»:

1. Эволюция методов обработки и анализа данных. Переход от статистической обработки к методам OLAP и Data Mining. Принципы, лежащие в основе методов OLAP и Data Mining.

2. Общая характеристика методов интеллектуального анализа данных. Обзор основных задач, решаемых методами Data Mining.

3. Алгоритмы предварительной обработки данных. Очистка данных: факторный анализ, корреляционный анализ, фильтрация.

4. Трансформация данных: метод скользящего окна, квантование, сортировка и группировка, слияние данных.

5. Понятие и основные принципы «извлечения знаний». Основные задачи, решаемые методами Data Mining.

6. Автокорреляция. Линейная и логистическая регрессия.

7. Дерево решений: принципы построения, использование, интерпретация результатов.

8. Метод поиска ассоциативных правил, области его применения.

9. Нейросетевые методы анализа данных. Принципы функционирования искусственных нейронных сетей. Основные области их применения.

10. Многослойные нейросети. Аппроксимация функциональных зависимостей. Прогнозирование временных рядов.

Перечень вопросов для оценки уровня сформированности компетенции ПК-1 на этапе «Умения»:

11. Карты Кохонена. Задачи кластеризации и визуализации больших объемов данных.

Интерпретация результатов.

12. Инструменты многомерного статистического анализа. Статистический анализ данных.

Основные возможности, спектр решаемых задач.

13. Табличные процессоры как инструмент анализа. Эволюция инструментов: от Excel до SPSS и Statistica.

14. Этапы выполнения многомерного статистического анализа и интерпретация результатов

15. Платформа Deductor Studio: основные возможности, сфера применения.

16. Этапы решения задач с помощью Deductor Studio. Импорт/экспорт данных. Визуализация результатов. Возможности автоматизации отдельных этапов анализа данных.

17. Экспертные системы. Составные части экспертной системы. Приобретение знаний.

Извлечение знаний из данных. Участники процесса проектирования: эксперты, инженеры по знаниям, конечные пользователи.

18. Специфика анализа данных в документоведении. Отбор данных для анализа. Предварительная подготовка данных. Понижение размерности, сглаживание аномалий, фильтрация,

группировка. Этапы проведения интеллектуального анализа. Интерпретация результатов анализа

19. Принципы выбора методов и инструментов интеллектуального анализа данных

20. Представление результатов анализа, визуализация, экспорт данных.

Пример практической заданий для оценки уровня сформированности компетенции ПК-1 на этапе «Знания»:

Задача сегментации клиентов фирмы. Оценка кредитоспособности физических лиц. Задача прогнозирования потребности в продукции. Задачи прогнозирования продаж, поступления финансовых средств и др. Примеры комбинации методов Data Mining. Практикум: – построение скоринговой модели кредитования (деревья решений); – построение нейросетевой прогнозной модели спроса.

Пример Вопросов для дискуссии для оценки уровня сформированности компетенции ПК-1 на этапе «Умения»:

Методология анализа сложных систем.

2. Основные концепции построения хранилищ данных.

3. Построение автоматизированных систем предобработки данных.

4. Алгоритмы и технологии DataMining.

5. Построение корпоративных автоматизированных аналитических систем на основе

методология KDD.

6. Обзор алгоритмов построения деревьев решений.

7. Математические основы нейросетевых технологий.

8. Методология построения регрессионных моделей.

9. Применение технологии деревьев решений для оценки кредитоспособности

физических лиц.

10. Применение нейросетевых технологий для оценки кредитоспособности физических

лиц.

11. Ассоциативные правила, как инструмент повышения прибыльности в розничной и

оптовой торговле.

12. Задачи классификации, как инструмент повышения эффективности бизнеса.

13. Методология построения и верификации прогнозных моделей.

14. Основные методы прогнозирования.

15. Методы оценки качества прогнозных моделей.

Пример тестов для оценки уровня сформированности компетенции **ПК-3** на этапе «Владения»:

1. Какие из нижеперечисленных признаков являются количественными:

- а) плотность населения
- б) уровень занятости населения
- в) среднедушевой доход
- г) пол человека
- д) возраст
- е) уровень образования (начальное, среднее, высшее)

2. Какие из нижеперечисленных признаков являются дискретными:

- а) денежные доходы населения
- б) число детей в семье
- в) прибыль предприятия
- г) пол человека
- д) тарифный разряд

3. Типологические группировки применяются для:

- а) характеристики структурных сдвигов
- б) характеристики взаимосвязей между отдельными признаками
- в) разделения совокупности на качественно однородные типы
- г) характеристики структуры совокупности

4. Структурные группировки применяются для:

- а) разделения совокупности на качественно однородные типы
- б) характеристики взаимосвязей между отдельными признаками

в) характеристики структуры совокупности

5. Часть зависимой переменной в регрессионной модели, которая может быть

объяснена значением регрессора:

а) случайное возмущение;

б) отклик;

в) уравнение регрессии;

г) остаток.

6. Гипотеза является сложной, если:

а) она состоит из конечного числа простых гипотез;

б) она состоит из бесконечного числа простых гипотез;

в) Содержит только одно предположение.

7. Коррелированность возмущений с различными номерами называется:

а) гомоскедастичностью;

б) гетероскедастичностью;

в) автокорреляцией.

8. Критической областью называют:

а) совокупность значений критерия, при которых нулевую гипотезу принимают;

б) совокупность значений критерия, при которых нулевую гипотезу отвергают.

9. Причины гетероскедастичности (множественный выбор):

а) исследование неоднородных объектов;

б) характер наблюдений;

в) ошибки спецификации;

г) ошибки измерений.

10. Под мультиколлинеарностью понимается линейная зависимость (единичный выбор):

- а) зависимой переменной с одним или несколькими регрессорами;
- б) двух, или нескольких регрессоров;
- в) зависимой переменной с возмущением;
- г) регрессоров с возмущением.

11. С увеличением объема выборки длина доверительного интервала прогнозируемого

значения зависимой переменной (единичный выбор):

- а) увеличивается;
- б) уменьшается;
- в) не меняется.

12. Как переводится DATA MINING?

- а) “добыча” или “раскопка данных”.
- б) “значение”.
- с) “хранение”.
- д) “перечисление данных”.

13. Какое требование к переработке информации не верно?

- а) Данные имеют неограниченный объем.
- б) Данные являются разнородными.
- с) Результаты должны быть конкретны и понятны.
- д) Инструменты для обработки сырых данных должны быть сложны в

использовании.

14. Какая концепция положена в основу современной технологии Data Mining?

- а) Концепция естествознания.
- б) Концепция управления.
- с) Концепция шаблонов (паттернов).
- д) Концепция становления.

15. Сколько выделяют стандартных типов закономерностей?

- а. 4.
- б. 5.

с. 6.

16. Если несколько событий связаны друг с другом, то это...

- a) Ассоциация.
- b) Последовательность.
- c) Классификация.
- d) Кластеризация.

17. Основой для каких систем служит историческая информация, хранящаяся в БД в

виде временных рядов?

- a) Классификации.
- b) Последовательности.
- c) Прогнозирования.
- d) Ассоциации.

18. Какую иерархическую структуру создают деревья решений?

- a) "ЕСЛИ... ТО...".
- b) "НИ... НИ...".
- c) "КОГДА... ТО...".
- d) "... НИКОГДА...".

19. С чем связано направление эволюционного программирования?

- a) Постановка вопроса вида "значение параметра A больше x ".
- b) Использование метода "ближайшего соседа".
- c) подача значений входных параметров, на основе которых нужно принимать какие-то решения, прогнозировать развитие ситуации.
- d) Поиск зависимости целевых переменных от остальных в форме функций какого-то определенного вида.

20. Что называют хромосомами в генетических алгоритмах?

- a) Кодировку исходных логических закономерностей в базе данных.
- b) Направление эволюционного программирования.
- c) Большой класс систем.
- d) Набор закономерностей

Пример контрольных задач для оценки уровня сформированности компетенции ПК-1 на этапе «Владения»:

1) В файле P1_1.XLS содержатся данные финансовых расчетов с потребителями компании за последние 4 месяца. Каждая строка в приведенной базе данных содержит информацию об одной операции отгрузки товара, а именно, имя потребителя, месяц, категорию отгрузки, сумма отгрузки, сумма поступившей оплаты.

А) создайте сводную таблицу для вычисления количества операций отгрузки по каждому потребителю и по каждой категории за все 4 месяца.

Б) создайте сводную таблицу для вычисления общих сумм поставок по каждому потребителю за каждый месяц. Используя полученные данные, постройте соответствующие временные ряды для каждого потребителя.

В) постройте гистограмму (одну) для поступивших оплат только для двух категорий отгрузки «Оборудование» и «Материалы».

2) Основываясь на данных о продажах из файла «Продажи» и других сопутствующих справочниках сформировать сценарии ежедневных отчетов по долевого объему продаж в

каждом из отделов по группам товаров за последние 10 дней. В каждом из отчетов долже

Примеры контрольной работы

В 1

Необходимо построить сценарий в аналитической платформе Deductor, который ежедневно формирует отчет, показывающий 10 лидеров продаж по сумме продаж по итогам последних 10 дней. Результат необходимо визуализировать в виде упорядоченной по убыванию столбчатой диаграммы. Необходимо произвести визуализацию в двух видах и сформировать соответствующие отчеты, а именно, с отображением сумм продаж, с

отображением долей продаж в общей сумме продаж этих товаров. Исходные данные находятся в файле «Продажи», характеристики товаров в файле «Товары».

В 2

В файле представлены некоторые исходные данные, а именно столбцы «Вход1», «Вход2», «Выход». В данных присутствуют дубликаты и противоречия. Необходимо построить сценария в аналитической платформе Deductor, который исключает дубликаты и противоречия.

Данные представлены в файле «Дубликаты и противоречия».

В3

Коммерческий директор хочет иметь информацию о последних тенденциях в изменении суммовых объемов продаж по товарным группам. Для этого предлагается вычислить относительное изменение объемов продаж за последние 10 дней по отношению к объемам продаж за предыдущие 10 дней. Если это изменение менее $-0,3$, то товарной группе присваиваем категорию «Провал», от $-0,3$ до $-0,1$ – «Падение», от $-0,1$ до 0 – «Уменьшение», от 0 до $0,1$ – «Увеличение», от $0,1$ до $0,3$ – «Подъем», свыше $0,3$ – «Взлет». Необходимо построить соответствующий сценарий в аналитической платформе Deductor. Необходимые

данные находятся в файлах «Продажи», «Товары», «Товарные группы».

В чем состоит цель поиска ассоциативных правил. Дайте определения категориям «транзакция», «поддержка набора», «поддержка правила», «достоверность правила».

Алгоритм поиска ассоциативных правил Apriori: этапы работы. В чем состоит отличие алгоритма Apriori от алгоритмов AIS и SETM.

Организационные факторы при внедрении Data Mining в деятельность компании.

Человеческие факторы при внедрении Data Mining в деятельность компании. Основные роли специалистов в процессах Data Mining.

Перечислите стандарты методологии Data Mining. В чем состоят их особенности.

Вы работаете в небольшой туристической фирме и планируете массовую рассылку рекламного буклета. Ваши средства ограничены, поэтому вы хотите послать ее тем, кто готов тратить на путешествия и отдых в большей степени. В файле P5_1.XLS содержатся данные о случайной выборке клиентов размером 925 (пол, возраст, суммы, затраченные на путешествия и отдых в предыдущем году). Используйте данные, чтобы понять, насколько пол и возраст влияют на объем затрат. Сформулируйте обоснованные рекомендации относительно контингента для рассылки рекламной брошюры.

Основываясь на данных о продажах из файла «Продажи» и других сопутствующих справочниках сформировать сценарий ежедневного отчета по 5 лидерам товарных групп по суммам продаж за последние 15 дней. Отчет визуализировать с помощью столбчатых диаграмм с информацией о сумме продаж и названии товарных групп.

3. Методические материалы, определяющие процедуры оценивания результатов обучения по дисциплине (модулю), описание шкал оценивания

РЕЙТИНГ-ПЛАН

Виды учебной деятельности студентов	Балл за конкретное задание	Число заданий	Баллы	
			Минимальный	Максимальный
Модуль 1				
Текущий контроль			0	20
1) Аудиторная работа	1	2	0	2
2) Выполнение домашнего задания	2	3	0	6
3) Лабораторные работы	4	3	0	12
Рубежный контроль			0	15

Коллоквиум	5	1	0	5
Тестовые задания	5	2	0	10
Модуль 2				
Текущий контроль			0	20
1) Аудиторная работа	1	4	0	4
2) Выполнение домашнего задания	2	2	0	4
3) Лабораторные работы	6	2	0	12
Рубежный контроль			0	15
Коллоквиум	5	1	0	5
Тестовые задания	5	2	0	10
		Итого:	0	70
Поощрительные баллы			0	10
Участие в олимпиаде			0	6
Задания повышенной сложности	2	2	0	4
Посещаемость (баллы вычитаются из общей суммы набранных баллов)				
1. Посещение лекционных занятий			0	-6

2. Посещение практических и лабораторных занятий			0	-10
Итоговый контроль				
Экзамен			0	30
ИТОГО:			0	110

Результаты обучения по дисциплине (модулю) у обучающихся оцениваются по итогам текущего контроля количественной оценкой, выраженной в рейтинговых баллах. Оценке подлежит каждое контрольное мероприятие.

При оценивании сформированности компетенций применяется четырехуровневая шкала «неудовлетворительно», «удовлетворительно», «хорошо», «отлично».

Максимальный балл по каждому виду оценочного средства определяется в рейтинг-плане и выражает полное (100%) освоение компетенции.

Уровень сформированности компетенции «хорошо» устанавливается в случае, когда объем выполненных заданий соответствующего оценочного средства составляет 80-100%; «удовлетворительно» – выполнено 40-80%; «неудовлетворительно» – выполнено 0-40%

Рейтинговый балл за выполнение части или полного объема заданий соответствующего оценочного средства выставляется по формуле:

$$\text{Рейтинговый балл} = k \times \text{Максимальный балл},$$

где $k = 0,2$ при уровне освоения «неудовлетворительно», $k = 0,4$ при уровне освоения «удовлетворительно», $k = 0,8$ при уровне освоения «хорошо» и $k = 1$ при уровне освоения «отлично».

Оценка на этапе промежуточной аттестации выставляется согласно Положению о модульно-рейтинговой системе обучения и оценки успеваемости студентов УУНиТ:

На экзамене выставляется оценка:

- отлично - при накоплении от 80 до 110 рейтинговых баллов (включая 10 поощрительных баллов),
- хорошо - при накоплении от 60 до 79 рейтинговых баллов,
- удовлетворительно - при накоплении от 45 до 59 рейтинговых баллов,
- неудовлетворительно - при накоплении менее 45 рейтинговых баллов.

При получении на экзамене оценок «отлично», «хорошо», «удовлетворительно», на зачёте оценки «зачтено» считается, что результаты обучения по дисциплине (модулю) достигнуты и компетенции на этапе изучения дисциплины (модуля) сформированы.