МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное учреждение высшего образования«Поволжский государственный университет сервиса» (ФГБОУ ВО «ПВГУС»)

|  |  |  |  |
| --- | --- | --- | --- |
| УТВЕРЖДЕНО |  | | |
| на заседании Высшей школы интеллектуальных систем и кибертехнологий | | | |
|  | | | |
| Протокол от | 15.12.2023г. | № | 4 |

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ

(фонд оценочных средств)

для проведения промежуточной аттестации по дисциплине

|  |
| --- |
| **«Интеллектуальный анализ данных»** |

наименование дисциплины

|  |  |
| --- | --- |
| по образовательной программе высшего образования – программе | **магистратуры** |
|  | бакалавриата, специалитета, магистратуры |

**«Информационная безопасность интеллектуальных и информационно-аналитических систем»**

|  |
| --- |
|  |

наименование образовательной программы

|  |
| --- |
| **10.04.01 «Информационная безопасность»** |
| шифр, наименование направления подготовки / специальности |

|  |  |
| --- | --- |
| Составитель | Сыротюк С.Д., доцент, Высшая школа интеллектуальных систем и кибертехнологий, к.п.н., доцент |
|  | ФИО, должность, структурное подразделение,  ученая степень, ученое звание |

Тольятти

2023

**1. Паспорт фонда оценочных средств**

* 1. **Планируемые результаты обучения по дисциплине**

| Код и наименование компетенции | Код и наименование индикатора  достижения компетенции | **Основание (ПС**) \*для профессиональных компетенций |
| --- | --- | --- |
| ОПК-4. Способен осуществлять сбор, обработку и анализ научно-технической информации по теме исследования, разрабатывать планы и программы проведения научных исследований и технических разработок | ИОПК-4.1. Ведет подготовку научных и технических материалов в рамках проведения исследований в сфере информационной безопасности  ИОПК-4.2. Разрабатывает планы и программы проведения научных исследований в сфере информационной безопасности |  |
| ПК-1. Способен провести обследование и анализ деятельности подразделений предприятия, и на их основе выбрать технологии и основные компоненты создаваемых интеллектуальных и информационно-аналитических систем | ИПК-1.1. Проводит предпроектное обследование и анализ деятельности подразделений предприятия и выявляет их потребности, в том числе с применением интеллектуального анализа данных | 06.031 Специалист по автоматизации информационно-аналитической деятельности |

В результате изучения дисциплины обучающийся должен:

владеть: навыками проектирования технических разработок защищенных информационных, информационно-аналитических и экспертных систем,: навыками извлечения и применения знаний проектирования интеллектуальных и информационно-аналитических систем

уметь: разрабатывать планы и программы проведения научных исследований в сфере информационной безопасности, обосновывать выбор технологий и компонент создаваемых интеллектуальных и информационно-аналитических систем и формировать базу знаний и базу правил в области выбора конфигураций и состава информационно-аналитических и экспертных систем

знать: требования к оформлению научных публикаций по теме исследования в сфере информационной безопасности, принципы проведения предпроектного исследования, принципы и методы анализа предметной области исследования; принципы функционирования информационно-аналитических и экспертных систем; преимущества интеллектуальных технологий и систем

**1.2. Содержание дисциплины**

|  |  |  |  |
| --- | --- | --- | --- |
| № | Тема (раздел дисциплины)  (в соответствии с РПД) | Код компетенцияи | Индикатор достижения компетенции |
| 1 | Тема 1. Технологии и средства интеллектуального анализа данных | ПК-1 | ИПК-1.1 |
| 2 | Тема 2. Интеллектуальный анализ деятельности предприятия | ОПК-4 | ИОПК-4.1,  ИОПК-4.2; |
| 3 | Тема 3. Применение систем извлечения знаний методами Data Mining и интеллектуального анализа данных для комплексной оценки безопасности автоматизированных систем управления и разработки систем проактивной защиты информации на основе анализа событий в информационной системе | ПК-1 | ИПК-1.1 |

**1.3. Система оценивания по дисциплине**

Дисциплина изучается в течение одного семестра.

Форма промежуточной аттестации по дисциплине– экзамен

|  |  |  |  |
| --- | --- | --- | --- |
| Шкала оценки уровня сформированности компетенций | | Шкала оценки уровня освоения дисциплины | |
| Уровневая шкала оценки компетенций | 100 балльная шкала, % | 100 балльная шкала, % | 5-балльная шкала,  дифференцированная оценка/балл |
| Допороговый | Ниже 61 | Ниже 61 | «Неудовлетворительно» / 2 |
| Пороговый | 61-85,9 | 61-69,9 | «Удовлетворительно» / 3 |
| 70-85,9 | «Хорошо» / 4 |
| Повышенный | 86-100 | 86-100 | «Отлично» / 5 |

**2. . Перечень оценочных материалов для проведения промежуточной аттестации**

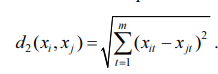
Контроль и оценка результатов освоения дисциплины осуществляется преподавателем в ходе текущего контроля успеваемости (в процессе проведения практических занятий, тестирования, опросов).

В ходе проведения промежуточной аттестации осуществляется контроль и оценка результатов освоения компетенций.

**Перечень вопросов для подготовки к экзамену**

**ОПК-4: ИОПК-4.1, ИОПК-4.2.** **Способен осуществлять сбор, обработку и анализ научно-технической информации по теме исследования, разрабатывать планы и программы проведения научных исследований и технических разработок**

1. Перечислите задачи интеллектуального анализа данных
2. Что понимается под кластером объектов данных, в чем его преимущества перед классификацией
3. Перечислите требования к кластеризации в интеллектуальном анализе данных
4. Назовите категории, на которые можно разделить методы кластеризации
5. Приведите примеры методов кластеризации
6. Перечислите характеристики кластера и дайте их краткое описание:
7. Перечислите основные недостатки методов кластеризации
8. Назовите основные проблемы кластерного анализа
9. Название расстояния, используемое в алгоритмах кластеризации, вычисляемое по данной формуле



1. Название расстояния, используемое в алгоритмах кластеризации, вычисляемое по данной формуле



1. Название расстояния, используемое в алгоритмах кластеризации, вычисляемое по данной формуле



1. Название расстояния, используемое в алгоритмах кластеризации, вычисляемое по данной формуле



1. Раскройте определение Дерево решений (DecisionTree) —
2. По какому алгоритму осуществляется построение дерева
3. Классификация как метода ИАД
4. Дайте характеристику простой классификации
5. Основными этапами процесса классификации являются:
6. Укажите цель процесса классификации
7. Назовите достоинства нейронной сети Кохонена
8. Назовите недостатки нейронной сети Кохонена

**ПК-1:ИПК-1.1. Способен провести обследование и анализ деятельности подразделений предприятия, и на их основе выбрать технологии и основные компоненты создаваемых интеллектуальных и информационно-аналитических систем**

1. Что понимается под Data Mining? В каком году и кем было дано данное определение?
2. Какова классификация задач Data Mining?
3. Дайте краткое описание задаче анализа данных - классификация
4. Дайте краткое описание задаче анализа данных - регрессия
5. Дайте краткое описание задаче анализа данных - кластеризация
6. Дайте краткое описание задаче анализа данных - поиск ассоциативных правил
7. В чем заключается цель технологии Data Mining. Какие существуют виды моделей Data Mining?
8. Какие подходы и методы относят к базовым методам Data Mining? Укажите их недостатки
9. Дайте краткую характеристику генетическим алгоритмам
10. Дайте краткую характеристику нейронным сетям
11. Перечислите этапы интеллектуального анализа данных
12. Каково назначение стандарта CWM (Common WarehouseMetamodel) Data Mining?
13. Перечислите основные стандарты в области DataMining.
14. На какие концептуальные области разделен Пакет Data Mining стандарта CWM?
15. Какова структура стандарта CWM-DM?
16. Каково назначение стандарта Data Mining – CRISP?
17. На какие этапы стандарт CRISP-DM делит жизненный цикл проекта Data Minin ?
18. Каково назначение стандарта Data Mining – PMML?
19. Какие типы моделей Data Mining поддерживает вторая версия PMML?
20. Каково назначение стандарта Data Mining – SQL/MM?
21. Каково назначение стандарта Data Mining – JDMAPI?

**Вопросы (задания) для подготовки к экзамену**

**с «ключами» правильных ответов**

| № | Содержание вопроса | Правильный ответ | |
| --- | --- | --- | --- |
| **ОПК-4. Способен осуществлять сбор, обработку и анализ научно-технической информации по теме исследования, разрабатывать планы и программы проведения научных исследований и технических разработок** | | | |
|  | Перечислите задачи интеллектуального анализа данных | К задачам интеллектуального анализа данных относят:   1. Классификация 2. Кластеризация 3. Выявление ассоциаций 4. Выявление 5. последовательностей 6. Регрессия 7. Прогнозирование | |
|  | Что понимается под кластером объектов данных, в чем его преимущества перед классификацией | Кластер объектов данных можно рассматривать как одну группу.  При выполнении кластерного анализа мы сначала разделяем набор данных на группы на основе сходства данных, а затем присваиваем группам метки.  Основное преимущество кластеризации перед классификацией заключается в том, что она адаптируется к изменениям и помогает выделять полезные функции, которые отличают разные группы | |
|  | Перечислите требования к кластеризации в интеллектуальном анализе данных | Требования, предъявляемые к кластеризации в интеллектуальном анализе данных:  1. Масштабируемость 2. Способность работать с различными видами атрибутов 3. Обнаружение кластеров с атрибутивной формой 4. Высокая размерность 5. Способность работать с зашумленными данными 6. Интерпретируемость | |
|  | Назовите категории, на которые можно разделить методы кластеризации | Методы кластеризации можно разделить на следующие категории   1. Метод секционирования 2. Иерархический метод 3. Метод, основанный на плотности 4. Грид-метод 5. Метод, основанный на моделях 6. Метод, основанный на ограничениях | |
|  | Приведите примеры методов кластеризации | Примеры методов кластеризации   1. CURE (Clustering Using Representatives) 2. CHAMELEON 3. BIRCH (Balanced Iterative Reducing and 4. Clustering using Hierarchies) 5. k-средних 6. LargeItem 7. Самоорганизующиеся карты Кохонена 8. Fuzzy C-means и др. | |
|  | Перечислите характеристики кластера и дайте их краткое описание: | Центр кластера – это среднее геометрическое место точек в пространстве характеристик образующих кластер.  Радиус кластера – это максимальное расстояние до объекта, входящего в кластер, от центра кластера.  Размер кластера – определяется либо по радиусу кластера, либо по среднеквадратичному отклонению объектов для этого кластера. | |
|  | Перечислите основные недостатки методов кластеризации | К основным недостаткам методов кластеризации относят:  1) необходимость задания пороговых значений;  2) необходимость задания количества кластеров;  3) работа только с данными одного типа (числовые или текстовые);  4) выделение кластеров определенной формы;  5) чувствительность к аномалиям в наборе данных;  6) существующие методы являются контекстно-зависимыми;  7) возможно возникновение неопределенностей;  8) медленная работа на больших объемах данных;  9) отсутствие гарантии в нахождении оптимального  решения;  10)нелинейность времени работы алгоритма в зависимости  от объема входных данных;  11)вычислительная сложность | |
|  | Назовите основные проблемы кластерного анализа | Основные проблемы кластерного анализа  1 Выбор метода исследования  2 Оценка качества полученного разбиения  3 Выбор значения параметра «Количество кластеров»  4 Постоянно растущие объемы данных | |
|  | Название расстояния, используемое в алгоритмах кластеризации, вычисляемое по данной формуле | Таким образом вычисляется Евклидово расстояние. Иногда может возникнуть желание возвести в квадрат стандартное евклидово расстояние, чтобы придать большие веса более отдаленным друг от друга объектам. | |
|  | Название расстояния, используемое в алгоритмах кластеризации, вычисляемое по данной формуле | Таким образом вычисляется расстояние по Хеммингу. Это расстояние является средним разностей по координатам. В большинстве случаев данная мера расстояния приводит к таким же результатам, как и для обычного расстояния Евклида, однако для нее влияние отдельных больших разностей (выбросов) уменьшается (т. к. они не возводятся в квадрат). | |
|  | Название расстояния, используемое в алгоритмах кластеризации, вычисляемое по данной формуле | Таким образом вычисляется расстояние Чебышева. Это расстояние может оказаться полезным, когда желают определить два объекта как "различные", если они различаются по какой-либо одной координате (каким-либо одним измерением). | |
|  | Название расстояния, используемое в алгоритмах кластеризации, вычисляемое по данной формуле | Таким образом вычисляется расстояние Махаланобиса, данная мера расстояния плохо работает, если ковариационная матрица высчитывается на всем множестве входных данных. В то же время, будучи сосредоточенной на конкретном классе (группе данных), данная мера расстояния показывает хорошие результаты | |
|  | Раскройте определение Дерево решений (DecisionTree) — | это модель машинного обучения, которая представляет собой дерево с узлами и листьями.  Узлы дерева представляют собой решения, которые необходимо принимать, а листья — конечные результаты. В задачах классификации каждый лист дерева соответствует определённому к классу, а в задачах. | |
|  | По какому алгоритму осуществляется построение дерева | Построение дерева происходит по следующему алгоритму  Выбирается признак, по которому происходит разделение выборки.  Выбирается пороговое значение для этого признака.  Разделяется выборка на две части: объекты, у которых значение признака меньше или равно порогу, и объекты, у которых значение признака больше порога.  Для каждой из двух полученных подвыборок рекурсивно повторяются шаги 1-3, пока не будет достигнут критерий останова (например, глубина дерева, количество объектов в листе и т.д.).  В листе дерева выводится ответ — класс или значение целевой переменной | |
|  | Классификация как метод интеллектуального анализа данных | Классификация - системное распределение изучаемых предметов, явлений, процессов по родам, видам, типам, по каким-либо существенным признакам для удобства их исследования; группировка исходных понятий и расположение их в определенном порядке, отражающем степень этого сходства.  Классификация относится к стратегии обучения с учителем, которое также именуют контролируемым или управляемым обучением. | |
|  | Дайте характеристику простой классификации | Простая классификация-деление родового понятия только по признаку и только один раз до раскрытия всех видов (дихотомия, при которой членами деления бывают только два понятия, каждое из которых является противоречащим другому). | |
|  | Основными этапами процесса классификации являются: | Процесс классификации состоит из двух этапов : конструирования модели и ее использования | |
|  | Укажите цель процесса классификации | Цель процесса классификации состоит в том, чтобы построить модель, которая использует прогнозирующие атрибуты в качестве входных параметров и получает значение зависимого атрибута | |
|  | Назовите достоинства нейронной сети Кохонена | Преимущества нейронных сетей Кохонена:   1. Простота реализации. Не требуют сложных математических выкладок и могут быть реализованы с помощью небольшого количества кода. 2. Быстрая обучаемость. Обучаются быстро, особенно на небольших наборах данных. Это позволяет использовать их для решения задач в реальном времени. 3. Способность кластеризации. Хорошо справляются с задачами кластеризации, позволяя группировать данные по схожим характеристикам. 4. Устойчивость к шуму. Могут обрабатывать данные с некоторым уровнем ошибок и все равно выдавать достаточно точные результаты. | |
|  | Назовите недостатки нейронной сети Кохонена | Недостатки:   1. Ограниченность в сложности задач. Лучше подходят для простых задач кластеризации. 2. Неопределенность в выборе параметров. Необходимо выбрать количество нейронов и скорость обучения. Оптимальный выбор может быть неочевидным и требовать экспериментов. 3. Чувствительность к начальным условиям. Это может привести к различным результатам при каждом запуске. 4. Необходимость предварительной обработки данных. Может потребоваться предварительная обработка данных, такая как нормализация или удаление выбросов. | |
| **ПК-1. Способен провести обследование и анализ деятельности подразделений предприятия, и на их основе выбрать технологии и основные компоненты создаваемых интеллектуальных и информационно-аналитических систем** | | | |
|  | Что понимается под Data Mining? В каком году и кем было дано данное определение? | | Data Mining — исследование и обнаружение "машиной" (алгоритмами, средствами искусственного интеллекта) в сырых данных скрытых знаний, которые ранее не были известны, нетривиальны, практически полезны, доступны для интерпретации человеком.  Классическое определение этого термина дал в 1996 г. один из основателей этого направления Пятецкий-Шапир |
|  | Какова классификация задач Data Mining? | | Методы Data Mining помогают решить многие задачи, с которыми сталкивается аналитик. Из них основными являются: классификация, регрессия, поиск ассоциативных правил и кластеризация. |
|  | Дайте краткое описание задаче анализа данных - классификация | | Задача классификации сводится к определению класса объекта по его характеристикам. Необходимо заметить, что в этой задаче множество классов, к которым может быть отнесен объект, заранее известно. |
|  | Дайте краткое описание задаче анализа данных - регрессия | | Задача регрессии, подобно задаче классификации, позволяет определить по известным характеристикам объекта значение некоторого его параметра. В отличие от задачи классификации значением параметра является не конечное множество классов, а множество действительных чисел. |
|  | Дайте краткое описание задаче анализа данных - кластеризация | | Задача кластеризации заключается в поиске независимых групп (кластеров) и их характеристик во всем множестве анализируемых данных. Решение этой задачи помогает лучше понять данные. Кроме того, группировка однородных объектов позволяет сократить их число, а следовательно, и облегчить анализ. |
|  | Дайте краткое описание задаче анализа данных - поиск ассоциативных правил | | При поиске ассоциативных правил целью является нахождение частых зависимостей (или ассоциаций) между объектами или событиями. Найденные зависимости представляются в виде правил и могут быть использованы как для лучшего понимания природы анализируемых данных, так и для предсказания появления событий. |
|  | В чем заключается цель технологии Data Mining. Какие существуют виды моделей Data Mining? | | Цель технологии Data Mining — нахождение в данных таких моделей, которые не могут быть найдены обычными методами.  Существуют два вида моделей: предсказательные и описательные. |
|  | Какие подходы и методы относят к базовым методам Data Mining? Укажите их недостатки | | К базовым методам Data Mining можно отнести также и подходы, использующие элементы теории статистики. В связи с тем, что Data Mining является развитием статистики, таких методов достаточно много. Основная их идея сводится к корреляционному, регрессионному и другим видам статистического анализа. Основным недостатком является усреднение значений, что приводит к потере информативности данных. Это в свою очередь приводит к уменьшению количества добываемых знаний |
|  | Дайте краткую характеристику генетическим алгоритмам | | Генетические алгоритмы (ГА) относятся к числу универсальных методов оптимизации, позволяющих решать задачи различных типов (комбинаторные, общие задачи с ограничениями и без ограничений) и различной степени сложности. При этом ГА характеризуются возможностью как однокритериального, так и многокритериального поиска в большом пространстве, ландшафт которого является негладким |
|  | Дайте краткую характеристику нейронным сетям | | Нейронные сети — это класс моделей, основанных на биологической аналогии с мозгом человека и предназначенных после прохождения этапа так называемого обучения на имеющихся данных для решения разнообразных задач анализа данных. |
|  | Перечислите этапы интеллектуального анализа данных | | Этапы интеллектуального анализа данных:   1. понимание и формулировка задачи анализа; 2. подготовка данных для автоматизированного анализа (препроцессинг); 3. применение методов Data Mining и построение моделей; 4. проверка построенных моделей; интерпретация моделей человеком. |
|  | Каково назначение стандарта CWM (Common WarehouseMetamodel) Data Mining? | | Стандарт CWM (Common Warehouse Metamodel) — это стандарт, разработанный консорциумом OMG для обмена метаданными между различными программными продуктами и репозиториями, участвующими в создании корпоративных СППР. Он основан на открытых объектно-ориентированных технологиях и стандартах использует UML в качестве языка моделирования, XML и для обмена метаданными и язык программирования JAVA для реализации моделей и спецификаций. |
|  | Перечислите основные стандарты в области DataMining. | | Основными стандартами в области Data Mining являются: CWM Data Mining от OMG, CRISP, PMML, OLE DB for Data Mining корпорации Microsoft, SQL/MM, OLE DB for Data Mining |
|  | На какие концептуальные области разделен Пакет Data Mining стандарта CWM? | | Пакет Data Mining стандарта CWM разделен на три концептуальные области:  Model — описание метаданных моделей, получаемых в результате работы методов Data Mining;  Settings — описание метаданных настроек процесса построения моделей;  Attributes — описание метаданных для атрибутов данных |
|  | Какова структура стандарта CWM-DM? | | CWM имеет модульную структуру, разбитую на четыре основных уровня:  Foundation поддерживает спецификацию базовых структурных элементов;  Resource описывает информационные источники;  Analysis описывает средства анализа, включая многомерный анализ и Data Mining;  Management описывает особенности функционирования ХД. |
|  | Каково назначение стандарта Data Mining – CRISP? | | CRISP-DM — непатентованная, документированная и свободно доступная модель, описывающая основные фазы, выполнение которых позволяет организациям получать максимальную выгоду от использования методов Data Mining |
|  | На какие этапы стандарт CRISP-DM делит жизненный цикл проекта Data Minin ? | | CRISP-DM делит жизненный цикл проекта Data Mining на следующие шесть фаз: понимание бизнес-процессов, понимание данных, подготовка данных, моделирование, оценка, размещение |
|  | Каково назначение стандарта Data Mining – PMML? | | Стандарт PMML (Predicted Model Markup Language) предназначен для обмена построенными mining-моделями между системами Data Mining. Данный стандарт описывает форму представления моделей в виде XML-документа.  PMML — достаточно молодой стандарт и нуждается в дальнейшем совершенствовании. |
|  | Какие типы моделей Data Mining поддерживает вторая версия PMML? | | Вторая версия PMML поддерживает следующие типы моделей: ассоциативные правила, деревья решений, кластеры, регрессия, нейронные сети, результат метода Naive Bayes, последовательность |
|  | Каково назначение стандарта Data Mining – SQL/MM? | | SQL/MM Data Mining обеспечивает стандартный интерфейс к алгоритмам Data Mining. Он может представлять собой как верхний уровень любой объектно-реляционной системы базы данных, так и промежуточный уровень. Ñòàíäàðòû |
|  | Каково назначение стандарта Data Mining – JDMAPI? | | Стандарт JDMAPI будет представлять собой спецификацию API-функций для построения моделей Data Mining, извлечения знаний, их использование, а также создание, хранение, доступ и сохранение данных и метаданных, поддерживающих результаты Data Mining и выбор трансформации данных |