



Федеральное агентство по рыболовству
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Калининградский государственный технический университет»
(ФГБОУ ВО «КГТУ»)

УТВЕРЖДАЮ
Начальник УРОПСИ

Фонд оценочных средств
(приложение к рабочей программе дисциплины)
«ТЕХНОЛОГИИ DATA MINING»

вариативной части профессиональной образовательной программы магистратуры
по направлению подготовки

09.04.01 ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА

ИНСТИТУТ

цифровых технологий

РАЗРАБОТЧИК

кафедра прикладной математики и информационных технологий

1 РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Таблица 1 – Планируемые результаты обучения по дисциплине, соотнесенные с установленными индикаторами достижения компетенций

Код и наименование компетенции	Индикаторы достижения компетенции	Дисциплина	Результаты обучения (владения, умения и знания), соотнесенные с компетенциями/индикаторами достижения компетенции
ПК-1: Разработка и внедрение новых методов и технологий исследования больших данных	ПК-1.2: Проведение испытаний и разработка рекомендаций по внедрению и использованию усовершенствованных или разработанных новых методов, моделей, алгоритмов, технологий и инструментальных средств работы с большими данными	Технологии Data Mining	<p><u>Знать:</u></p> <ul style="list-style-type: none"> - содержание технологии Data Mining; - основные методы Data Mining; <p><u>Уметь:</u></p> <ul style="list-style-type: none"> - понимать основные проблемы, возникающие при анализе больших данных, и пути их решения в рамках технологии Data Mining; - выбирать методы интеллектуального анализа исходя из практической задачи; <p><u>Владеть:</u></p> <ul style="list-style-type: none"> - навыками анализа данных различной природы с использованием современных инструментальных средств.

Обучение проводится с использованием академической программы "Loginom Academic"

2 ПЕРЕЧЕНЬ ОЦЕНОЧНЫХ СРЕДСТВ И КРИТЕРИИ ОЦЕНИВАНИЯ

2.1 К оценочным средствам текущего контроля успеваемости относятся:

- тестовые задания открытого и закрытого типов.

2.2 Промежуточная аттестация по дисциплине проводится в форме зачета, который выставляется по результатам прохождения всех видов текущего контроля успеваемости. При необходимости тестовые задания закрытого и открытого типов могут быть использованы для проведения промежуточной аттестации.

2.3 Критерии оценки результатов освоения дисциплины

Универсальная система оценивания результатов обучения включает в себя системы оценок: 1) «отлично», «хорошо», «удовлетворительно», «неудовлетворительно»; 2) «зачтено», «не зачтено»; 3) 100 – балльную/процентную систему и правило перевода оценок в пятибалльную систему (табл. 2).

Таблица 2 – Система оценок и критерии выставления оценки

Система оценок Критерий	2	3	4	5
	0-40%	41-60%	61-80 %	81-100 %
	«неудовлетворительно»	«удовлетворительно»	«хорошо»	«отлично»
	«не зачтено»	«зачтено»		
1 Системность и полнота знаний в отношении изучаемых объектов	Обладает частичными и разрозненными знаниями, которые не может научно-корректно связывать между собой (только некоторые из которых может связывать между собой)	Обладает минимальным набором знаний, необходимым для системного взгляда на изучаемый объект	Обладает набором знаний, достаточным для системного взгляда на изучаемый объект	Обладает полнотой знаний и системным взглядом на изучаемый объект
2 Работа с информацией	Не в состоянии находить необходимую информацию, либо в состоянии находить отдельные фрагменты информации в рамках поставленной задачи	Может найти необходимую информацию в рамках поставленной задачи	Может найти, интерпретировать и систематизировать необходимую информацию в рамках поставленной задачи	Может найти, систематизировать необходимую информацию, а также выявить новые, дополнительные источники информации в рамках поставленной задачи
3. Научное осмысление изучаемого явления, процесса, объекта	Не может делать научно корректных выводов из имеющихся у него сведений, в состоянии проанализировать только некоторые из имеющихся у него сведений	В состоянии осуществлять научно корректный анализ предоставленной информации	В состоянии осуществлять систематический и научно корректный анализ предоставленной информации, вовлекает в исследование новые релевантные задаче данные	В состоянии осуществлять систематический и научно-корректный анализ предоставленной информации, вовлекает в исследование новые релевантные поставленной задаче данные, предлагает новые ракурсы

Система оценок Критерий	2	3	4	5
	0-40%	41-60%	61-80 %	81-100 %
	«неудовлетворительно»	«удовлетворительно»	«хорошо»	«отлично»
	«не зачтено»	«зачтено»		
				поставленной задачи
4. Освоение стандартных алгоритмов решения профессиональных задач	В состоянии решать только фрагменты поставленной задачи в соответствии с заданным алгоритмом, не освоил предложенный алгоритм, допускает ошибки	В состоянии решать поставленные задачи в соответствии с заданным алгоритмом	В состоянии решать поставленные задачи в соответствии с заданным алгоритмом, понимает основы предложенного алгоритма	Не только владеет алгоритмом и понимает его основы, но и предлагает новые решения в рамках поставленной задачи

3 ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

ПК-1: Разработка и внедрение новых методов и технологий исследования больших данных.

Индикатор ПК-1.2: Проведение испытаний и разработка рекомендаций по внедрению и использованию усовершенствованных или разработанных новых методов, моделей, алгоритмов, технологий и инструментальных средств работы с большими данными.

Тестовые задания закрытого типа:

1. К основным характеристикам Big Data относятся:

- а. Virtualization, Volume, Variability, Vehicle
- б. Variety, Velocity, Volume, Value**
- в. Verification, Volume, Velocity, Visualization
- г. Video, Value, Variety, Volume

2. **НЕВЕРНО** утверждение, что MapReduce:

- а. интерфейс для массово-параллельной обработки данных, где вычисления производятся на узлах, где информация изначально была сохранена
- б. две операции: распределения и сборки данных
- в. придуман разработчиками Hadoop**
- г. анонсирован разработчиками Google

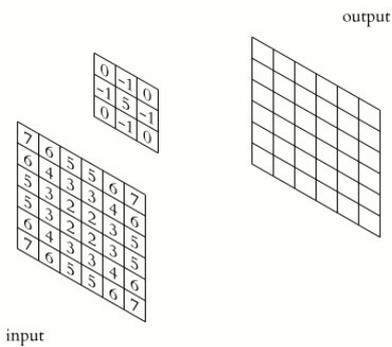
3. Hadoop – это:

- а. набор утилит, и программный каркас для выполнения распределённых программ, работающих на кластерах**
- б. распределённая СУБД, позволяющая обрабатывать большие данные
- в. язык выполнения заданий в парадигме MapReduce
- г. распределённая файловая система, предназначенная для хранения файлов большого объёма

4. Задача поиска групп объектов в не размеченном исходном наборе данных - это:

- а. классификация
- б. кластеризация**
- в. регрессия
- г. ассоциация

5. На рисунке



представлена нейронная сеть:

- а. прямого распространения
- б. сверточная**
- в. рекуррентная
- г. обратного распространения

6. К классу «обучение с учителем» относят задачи интеллектуального анализа данных:

Возможно несколько вариантов ответа

- а. классификация**
- б. кластеризация
- в. регрессия**
- г. поиск ассоциативных правил

7. В процессе управляемого обучения нейронной сети минимизации требует целевая функция:

- а. ошибок**
- б. переобучения
- в. активации
- г. корреляции

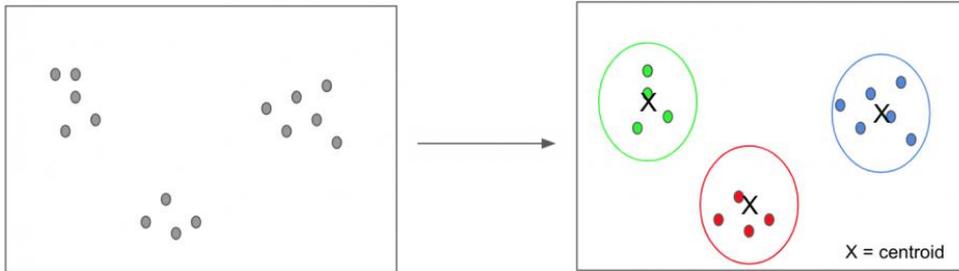
Тестовые задания открытого типа:

8. В основе алгоритма _____ лежит построение (обучение) бинарных деревьев решений.

Введите название заглавными буквами (аббревиатура, англ.) Допускаются иные формулировки ответа, не искажающие его смысла.

Ответ: CART

9. На рисунке



представлена иллюстрация алгоритма _____

Введите общепринятое название алгоритма (англ.) Допускаются иные формулировки ответа, не искажающие его смысла.

Ответ: k-means

10. Нежелательное явление, возникающее при решении задач обучения по прецедентам, когда вероятность ошибки обученного алгоритма на объектах тестовой выборки оказывается существенно выше, чем средняя ошибка на обучающей выборке - это _____.

Введите название явления в именительном падеже, регистр не важен..

Ответ: переобучение

11. Евклидово расстояние является частным случаем метрики Миньковского $D(x, y) = (\sum_{i=1}^n |x_i - y_i|^p)^{1/p}$ при значении p , равном ____.

Введите число.

Ответ: 2

12. Торговая компания хочет на основе имеющихся у нее анкетных данных получить группы клиентов, обладающих общими характеристиками. В технологии Data Mining для решения применяется тип задачи: _____.

Определите название в именительном падеже, регистр не важен..

Ответ: кластеризация

13. Доступны данные каталога землетрясений, содержащие сведения о дате, времени, месте, магнитуде и пр. Необходимо спрогнозировать магнитуду следующего землетрясения. В технологии Data Mining для решения применяется тип задачи: _____.

Определите название в именительном падеже, регистр не важен..

Ответ: регрессия

14. В технологии Data Mining для фильтрации спам-писем применяется тип задачи: _____.

Определите название в именительном падеже, регистр не важен..

Ответ: классификация

15. В теории нейронных сетей функция _____ определяет выходное значение нейрона в зависимости от результата взвешенной суммы входов и порогового значения.

Допускаются иные формулировки ответа, не искажающие его смысла.

Ответ: активации

16. В признаковом пространстве объекты, существенно отличающиеся от большинства остальных - это _____ .

Ответ: выброс

17. Имеется набор данных:

	sepal_length	sepal_width	petal_length	petal_width
0	5.1	3.5	1.4	0.2
1	4.9	3.0	1.4	0.2
2	4.7	3.2	1.3	0.2
3	4.6	3.1	1.5	0.2
4	5.0	3.6	1.4	0.2
5	5.4	3.9	1.7	0.4
6	4.6	3.4	1.4	0.3
7	5.0	3.4	1.5	0.2
8	4.4	2.9	1.4	0.2
9	4.9	3.1	1.5	0.1

При использовании системы LogiNot в задаче кластеризации алгоритмом k-means для поля «sepal_length» (длина лепестка) наиболее подходящим будет _____ тип данных.

Введите название типа. Допускаются иные формулировки ответа, не искажающие его смысла.

Ответ: вещественный

18. В сценарии LogiNot осуществляется фильтрация набора данных, содержащего сведения о рыночной стоимости квартир. Узел «фильтрация» настроен, как указано на рисунке.

Фильтрация данных

Состояние входа: Активировано

2 000 000,00 <= 90 Стоимость (т.руб.) <= 3 000 000,00 × и ab Тип планировки = брежневка ×

и ab Тип планировки = хрущевка × +

Для корректного набора данных условиям фильтра будет удовлетворять количество записей:

_____.

Введите число.

Ответ: 0

19. В системе LogiNot с помощью узла «Слияние» выполнено правое соединения (RIGHT JOIN) Таблицы 1 и Таблицы 2 по полю Артикул с главной таблицей «Таблица 1».

Таблица 1			Таблица 2		
Клиент	Артикул	Кол-во	Артикул	Товар	Цена
Клиент 1	1234	1	1234	Молоко	80
Клиент 1	2475	3	2475	Хлеб	30
Клиент 2	6488	2	1275	Йогурт	35
Клиент 3	1275	1	4168	Кофе	150

Итоговая таблица будет иметь пустые значения в: _____.

Введите имя поля (имена полей через запятую, без пробелов, регистр не важен)

Ответ: Клиент, Кол-во

20. Для решения задач бинарной классификации, в которых выходная переменная может принимать только два значения, используется модель регрессии: _____.

Введите вид модели в именительном падеже, регистр не важен.

Ответ: логистическая

21. На рисунке приведена матрица ошибок бинарной классификации (событие/не событие). Качество алгоритма оценивается по тестовой выборке.

Классифицировано	Фактически		Итого
	Событие	Не-событие	
Обучающее	4	10	
Событие	3	3	6
Не-событие	1	7	8
Тестовое	3	3	
Событие	2	1	3
Не-событие	1	2	3

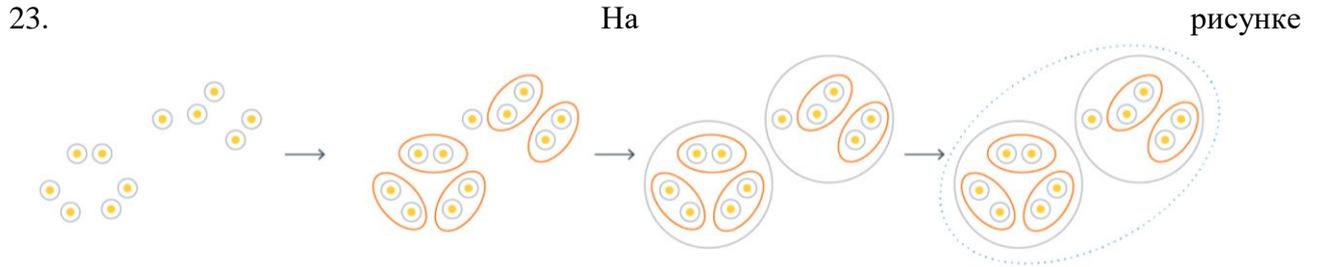
Ошибка первого рода равна _____.

Введите число (результат округлить до двух знаков, разделитель – запятая)

Ответ: 0,17

22. Метод _____ позволяет определить оптимальное число кластеров для алгоритма k-means. Допускаются иные формулировки ответа, не искажающие его смысла.

Ответ: локтя | Elbow



представлена _____ иерархическая кластеризация.

Определите вид, в именительном падеже, регистр не важен.

Ответ: агломеративная

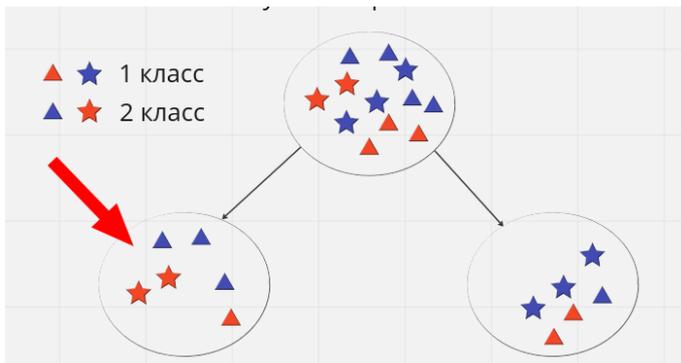
24. Отношение числа истинно-отрицательных классификаций (TN) к общему числу отрицательных наблюдений (N) – это _____.

Допускаются иные формулировки ответа, не искажающие его смысла.

Ответ: специфичность | Specify

25. На рисунке представлено случайно разбиение множества объектов на 2 группы. Индекс Джини отмеченной красной стрелкой группы равен _____.

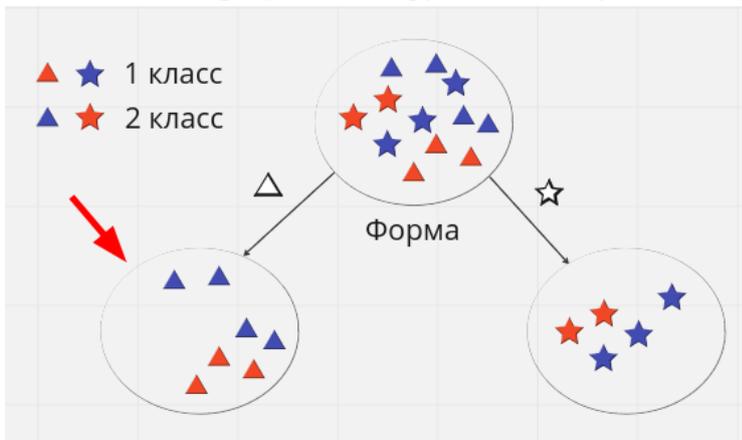
Введите число (результат округлить до двух знаков, разделитель – запятая).



Ответ: 0,72

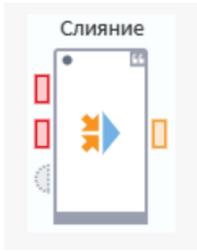
26. На рисунке представлено разбиение множества объектов на 2 группы по признаку «форма». Индекс Джини отмеченной красной стрелкой группы равен _____.

Введите число (результат округлить до двух знаков, разделитель – запятая).



Ответ: 0,51

27. В Loginom количество обязательных входных портов узла



равно ____

Ответ: 2

28. В Logiном при импорте данных (см. рис)

Обновить все		Обновить данные		Кол-во строк для анализа
Поля	7	ab	0/1	25
Имя	COL1	COL2	COL3	
Метка	Дата	Объект	Потребление эл энергия,кВт час	
Тип данных	7 Дата/Время	ab Строковый	0/1 Логический	
Вид данных	0 Непрерывный	0 Дискретный	0 Дискретный	
Использовать	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
1	01.05.2010, 0:00	Объект 2		
2	01.06.2010, 0:00	Объект 2		
3	01.07.2010, 0:00	Объект 2	true	
4	01.08.2010, 0:00	Объект 2		
5	01.09.2010, 0:00	Объект 2	true	
6	01.10.2010, 0:00	Объект 2	true	
7	01.11.2010, 0:00	Объект 2		
8	01.12.2010, 0:00	Объект 2	true	
9	01.01.2011, 0:00	Объект 2	true	
25	01.02.2011, 0:00	Объект 2		

неверно распознанные данные находятся в поле _____

Введите имя (регистр важен)

Ответ: COL3

29. Из упорядоченного множества принимают значения _____ признаки.

Введите вид признаков (мн. число)

Ответ: ординальные

30. Проверить значимость различия двух реализаций случайной переменной (например, оценить влияние изменения на сайте на реакцию пользователя) позволяет метод _____

Допускаются иные формулировки ответа, не искажающие его смысла.

Ответ: АВ-тестирование

4 ТИПОВЫЕ ЗАДАНИЯ НА КОНТРОЛЬНУЮ РАБОТУ, КУРСОВУЮ РАБОТУ/КУРСОВОЙ ПРОЕКТ

Данный вид контроля по дисциплине не предусмотрен учебным планом.

5 СВЕДЕНИЯ О ФОНДЕ ОЦЕНОЧНЫХ СРЕДСТВ И ЕГО СОГЛАСОВАНИИ

Фонд оценочных средств для аттестации по дисциплине «Технологии Data Mining» представляет собой компонент основной профессиональной образовательной программы магистратуры по направлению подготовки 09.04.01 Информатика и вычислительная техника.

Фонд оценочных средств рассмотрен и одобрен на заседании методической комиссии института цифровых технологий (протокол № 2 от 26.04.2022 г.).

Фонд оценочных средств актуализирован. Изменения, дополнения рассмотрены и одобрены на заседании методической комиссии института цифровых технологий (протокол № 3 от 24.03.2023 г.).

Директор института



А.Б. Тристанов