



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

УТВЕРЖДАЮ
И.о. директора института

Фонд оценочных средств
(приложение к рабочей программе модуля)
«АНАЛИЗ ДАННЫХ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ»

основной профессиональной образовательной программы бакалавриата
по направлению подготовки
09.03.03 ПРИКЛАДНАЯ ИНФОРМАТИКА

ИНСТИТУТ

цифровых технологий

РАЗРАБОТЧИК

кафедра прикладной математики и информационных
технологий

1 РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ, ПЕРЕЧЕНЬ ОЦЕНОЧНЫХ СРЕДСТВ И КРИТЕРИИ ОЦЕНИВАНИЯ

1.1 Результаты освоения дисциплины

Таблица 1 – Планируемые результаты обучения по дисциплине, соотнесенные с установленными компетенциями

Код и наименование компетенции	Дисциплина	Результаты обучения (владения, умения и знания), соотнесенные с компетенциями
<p>УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач</p>	<p>Анализ данных и искусственный интеллект</p>	<p><u>Знать:</u></p> <ul style="list-style-type: none"> – основные методы анализа и оценки информации, полученной в том числе с помощью цифровых средств; – классификацию задач систем искусственного интеллекта; – методы и средства искусственного интеллекта для решения профессиональных задач. <p><u>Уметь:</u></p> <ul style="list-style-type: none"> – критически оценивать надежность источников информации в условиях неопределенности и избытка/недостатка информации для решения поставленных задач, в том числе в цифровой среде; – осуществлять поиск данных в открытых источниках и специализированных библиотеках; – выявлять и анализировать проблемную ситуацию, выделяя ее структурные составляющие и связи между ними; – применять методы анализа данных при решении профессиональных задач; – выбирать современные технологии искусственного интеллекта для решения профессиональных задач. <p><u>Владеть</u></p> <ul style="list-style-type: none"> – аналитическими и системными навыками, способностью к поиску информации; – навыками подготовки и очистки данных, предназначенных для обработки системами искусственного интеллекта; – современными программными средствами для решения задач анализа данных в профессиональной деятельности.

1.2 К оценочным средствам текущего контроля успеваемости относятся:

- тестовые задания открытого и закрытого типов;
- контрольная работа (заочной формы обучения).

Промежуточная аттестация в форме дифференцированного зачета (зачета с оценкой) в третьем семестре проходит по результатам прохождения всех видов текущего контроля успеваемости. В отдельных случаях (при не прохождении всех видов текущего контроля) зачет может быть проведен в виде тестирования.

1.3 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	2	3	4	5
	0-40%	41-60%	61-80 %	81-100 %
Критерий	«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	В состоянии осуществлять научно корректный анализ	В состоянии осуществлять систематический и научно корректный анализ	В состоянии осуществлять систематический и научно-корректный анализ

Система оценок	2	3	4	5
	0-40%	41-60%	61-80 %	81-100 %
Критерий	«неудовлетворительно»	«удовлетворительно»	«хорошо»	«отлично»
	«не зачтено»	«зачтено»		
	проанализировать только некоторые из имеющихся у него сведений	предоставленной информации	предоставленной информации, вовлекает в исследование новые релевантные задаче данные	предоставленной информации, вовлекает в исследование новые релевантные поставленной задаче данные, предлагает новые ракурсы 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

1.4 Оценивание тестовых заданий закрытого типа осуществляется по системе зачтено/ не зачтено («зачтено» – 41-100% правильных ответов; «не зачтено» – менее 40 % правильных ответов) или пятибалльной системе (оценка «неудовлетворительно» - менее 40 % правильных ответов; оценка «удовлетворительно» - от 41 до 60 % правильных ответов; оценка «хорошо» - от 61 до 80% правильных ответов; оценка «отлично» - от 81 до 100 % правильных ответов).

Тестовые задания открытого типа оцениваются по системе «зачтено/ не зачтено». Оценивается верность ответа по существу вопроса, при этом не учитывается порядок слов в словосочетании, верность окончаний, падежи.

2 ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач.

Тестовые задания открытого типа:

1. Автором термина «искусственный интеллект» является _____

Ответ: Маккарти (Дж. Маккарти) (Джон Маккарти)

2. Метод для определения способности машины (компьютера) проявлять интеллектуальное поведение, неотличимое от человеческого в 1950 году предложил _____

Ответ: Тьюринг (А. Тьюринг) (Алан Тьюринг)

3. _____ (англ.) - технология обнаружения в «сырых» данных ранее неизвестных, нетривиальных, практически полезных и доступных для интерпретации знаний, необходимых для принятия решения в различных сферах человеческой деятельности.

Ответ: Data Mining (интеллектуальный анализ данных)

4. основополагающие работы в области искусственного интеллекта – кибернетике – создал _____

Ответ: Н. Винер (Норберт Винер)

5. В основе имитационного подхода построения систем ИИ лежит модель: _____

Ответ: черный ящик

6. Основой логического подхода построения систем ИИ является: _____

Ответ: булева алгебра (алгебра логики)

7. В контексте анализа данных оптимизация представлений данных и их форматов с точки зрения решаемых задач и целей анализа – это _____ данных.

Ответ: трансформация (преобразование)

8. Метод предобработки числовых признаков в обучающих наборах данных с целью приведения их к некоторой общей шкале измерения без потери информации о различии диапазонов – это _____ данных.

Ответ: нормализация

9. Процесс обработки данных, который преобразует непрерывные данные в дискретные путем замены значений диапазонами, может использоваться для сокращения размерности данных, а именно, для уменьшения числа разнообразных значений признака – это _____

Ответ: квантование

10. _____ — это способ представления переменных (признаков, атрибутов) и их группировки в различные категории, определяющий характер значений, присвоенных переменным в наборе данных.

Ответ: шкала измерения

11. В машинном обучении процесс прогнозирования категориальной метки на основе входных данных – это _____

Ответ: классификация

12. Автоматическое разбиение элементов некоторого множества на группы по принципу схожести – это: _____

Ответ: кластеризация

13. При оценке ассоциативных правил важны три метрики: _____, _____, _____.

Ответ: поддержка, достоверность, лифт (support, confidence, lift)

14. Метод в статистике и машинном обучении, используемый для моделирования и анализа отношений между зависимой и одной или несколькими независимыми переменными – это _____ (вид) анализ

Ответ: регрессионный

15. Тесноту связи и направление линейной зависимости между двумя переменными определяет коэффициент корреляции _____

Ответ: Пирсона (Пирсон)

16. Одним из первых попробовал смоделировать структуру человеческого мозга: _____ (фамилия ученого)

Ответ: Розенблатт (Ф. Розенблатт) (Фрэнк Розенблатт) (Розенблат)

17. В контексте нейронных сетей обучение модели на основе пары входных данных и соответствующих выходных данных – это обучение _____ (тип)

Ответ: с учителем

18. В терминах НС простейший вид искусственного нейрона с пороговой функцией активации – это _____

Ответ: перцептрон (персептрон)

19. Текущее состояние нейрона определяется как взвешенная сумма его _____

Ответ: входов

20. В нейронной сети коэффициент, определяющий важность входных данных для выхода нейрона – это _____

Ответ: вес

21. Тип нейронной сети, чаще всего используемый для анализа визуальных данных, например, для распознавания изображений: _____

Ответ: сверточная (CNN) (СНС)

22. _____ - процесс разделения текста на составляющие (слова, предложения, группы слов), является важным обязательным процессом при применении алгоритмов машинного обучения и нейронных сетей для анализа текста.

Ответ: токенизация

23. В АП Loginom средство визуализации многомерных данных и получения необходимых форм отчетов – это _____

Ответ: куб (OLAP- куб) (гиперкуб)

24. 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 АП Loginom в задаче кластеризации алгоритмом k-means для поля «sepal_length» (длина лепестка) наиболее подходящим будет _____ тип данных.

Введите название типа

Ответ: вещественный

Тестовые задания закрытого типа:

25. Характерная особенность систем искусственного интеллекта:

1. обработка данных в символьной форме
2. обработка данных в числовой форме
3. наличие четкого алгоритма
- 4. необходимость выбора из многих вариантов**

26. Методы искусственного интеллекта применяются для:

1. синтаксического анализа текста
- 2. восприятия и распознавания образов**
- 3. понимания естественного языка**
4. построения базы данных

27. Установление соответствия:

Машинное обучение		Действия	
1	с учителем	а	Действовать с целью добиться максимального результата. Создавать автономных агентов, которые умеют учиться.
2	без учителя	б	Обучить необходимому поведению через размеченные данные. Разбираться с новыми данными на основе предыдущих данных
3	с подкреплением	в	Делать выводы без размеченных данных. Открывать неизвестные или скрытые закономерности.

Ответ: 1б, 2в, 3а

28. Для бинарной задачи обычно используется результат классификации:

1. числовой
- 2. категорийный**
3. временной ряд
4. мультиклассовый

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

1. классификация

2. кластеризация

3. регрессия

4. поиск ассоциативных правил

30. Установление соответствия:

Модель интеллектуального анализа данных		Бизнес-сценарий	
1	Прогнозирование	а	выбор наиболее подходящих заказчиков для целевой рассылки
2	Риск и вероятность	б	определение продуктов, которые с высокой долей вероятности могут быть проданы вместе
3	Рекомендации	в	разделение заказчиков или событий на кластеры связанных элементов
4	Поиск последовательностей	г	оценка продаж
5	Группировка	д	анализ выбора заказчиков во время совершения покупок

Ответ: 1г, 2а, 3б, 4д, 5в






31. Порядок шагов при создании модели регрессионного анализа:

Шаг	
1	определение модели
2	расчет коэффициентов
3	оценка допущений модели

4	интерпретация результатов
---	---------------------------

Ответ: 1, 3, 2, 4

32. В АП Loginom установление соответствие обозначений типов данных:

Знак (АП Loginom)		Тип данных (АП Loginom)	
1		а	Логический
2		б	Дата/Время
3		в	Вещественный
4		г	Целый
5		д	Строковый
6		е	Переменный

Ответ: 1г, 2а, 3б, 4в, 5е, 6д

3 ТИПОВЫЕ ЗАДАНИЯ НА КОНТРОЛЬНУЮ РАБОТУ, КУРСОВУЮ РАБОТУ/ КУРСОВОЙ ПРОЕКТ, РАСЧЕТНО-ГРАФИЧЕСКУЮ РАБОТУ

Учебным планом предусмотрено выполнение одной контрольной работы (для заочной формы обучения).

КОНТРОЛЬНАЯ РАБОТА

Задание 1. Подготовить реферат по заданной теме (по вариантам):

- Понятие анализа данных. Задачи анализа данных. Методы анализа данных. Принципы анализа данных.
- Инструменты анализа данных. Области применения анализа данных.
- Очистка данных. Преобразование данных. Интеграция данных. Разделение данных. Определение типов данных.

- Обработка отсутствующих значений. Обработка выбросов. Масштабирование данных. Нормализация данных.
- Кодирование категориальных признаков. Слияние данных. Удаление дубликатов. Индексирование данных.
- Визуализация данных. Оценка качества данных. Сохранение предобработанных данных.
- Линейная регрессия. Модель линейной регрессии. Коэффициенты линейной регрессии.
- Метод наименьших квадратов.
- Регрессионный анализ. Критерий оптимальности. Предсказание.
- Множественная линейная регрессия.
- Регуляризация.
- Лассо-регрессия.
- Ридж-регрессия.
- Кросс-валидация. Статистическая значимость. Коэффициент детерминации. Остатки. Анализ выбросов.
- Интерпретация результатов. Применение линейной регрессии в реальных задачах.
- Классификация. Класс. Объект. Признак. Алгоритм.
- Кластеризация. Разбиение. Дерево решений.
- Нейронные сети.
- Метод опорных векторов. Метрики качества. Переобучение. Регуляризация.
- Иерархическая кластеризация.
- Алгоритм k-средних.
- Алгоритм DBSCAN.
- Агломеративная кластеризация.
- Спектральная кластеризация.
- Метрики качества кластеризации. Определение оптимального количества кластеров.
- Проблемы кластеризации и их решения.
- Применение кластеризации в машинном обучении.
- Применение кластеризации на практике в различных областях.
- Ассоциативные правила.
- Правила ассоциации первого порядка. Правила ассоциации высоких порядков. Методы извлечения правил ассоциации.

- Алгоритм Apriori. Алгоритм FP-growth. Параллельные алгоритмы извлечения правил ассоциации.
- Построение дерева принятия решений на основе правил ассоциации.
- Анализ последовательностей событий на основе правил ассоциации.
- Оценка качества правил ассоциации.
- Искусственные нейронные сети. Математические модели.
- Распознавание образов. Классификация данных.
- Многослойный перцептрон.
- Сверточные нейронные сети.
- Компьютерное зрение.
- Обработка естественного языка.
- Архитектура нейронных сетей.
- Распознавание речи.
- Обработка изображений.

Задание 2. Использование возможностей платформы Loginom Community и MS Excel для обработки данных.

Создать сценарий Loginom, включающий в себя:

- импорт данных (формат *.csv, *.xlsx) на платформу Loginom Community;
- обработку данных в соответствии с заданием (данные о продажах товаров по группам и категориям):
 - а) исключить из набора записи с отсутствующими данными;
 - б) определить количество продаж товаров по категориям, сумму выручки в каждой категории за определенный период (по вариантам);
 - в) рассчитать прибыль (N% - по вариантам) от продаж определенной (по вариантам) группы товаров за определенный период (по вариантам)
- визуализацию результатов обработки соответствующих узлов сценария;
- экспорт результатов обработки в файл MS Excel.

Задание 3. Использование возможностей платформы Loginom Community и MS Excel для проведения регрессионного анализа исследовательских данных, планирования и обработки результатов факторного эксперимента.

Обработать и оценить результаты исследований по изучению влияния различных факторов на стоимость жилья.

Создать сценарий Loginom, включающий в себя:

- импорт данных (формат *.xlsx) на платформу *Loginom Community*;
- обработку данных в соответствии с заданием (исходный набор данных - по вариантам), использовав модель линейной регрессии может (компонент *Loginom* Линейная регрессия (парная)).

Условия отбора данных:

1. отсутствие записей с пустыми полями;
 2. год постройки дома не ранее 1940 г. и не позже 2018 г;
 3. средний класс жилья (от 2 до 7);
 4. независимые переменные: Площадь и Класс, зависимая переменная Цена, \$.
- визуализацию результатов обработки соответствующих узлов сценария;
 - экспорт результатов обработки в файл MS Excel;
 - сделать вывод.

4 СВЕДЕНИЯ О ФОНДЕ ОЦЕНОЧНЫХ СРЕДСТВ И ЕГО СОГЛАСОВАНИИ

Фонд оценочных средств для аттестации по дисциплине «Анализ данных и искусственный интеллект» представляет собой компонент основной профессиональной образовательной программы бакалавриата по направлению подготовки 09.03.03 Прикладная информатика.

Преподаватель-разработчик – Руденко А.И., к.ф.-м.н., Романов М.А.

Фонд оценочных средств рассмотрен и одобрен и.о. заведующего кафедрой прикладной математики и информационных технологий.

И.о. заведующего кафедрой



А.И. Руденко

Фонд оценочных средств рассмотрен и одобрен заведующим кафедрой прикладной информатики.

Заведующий кафедрой



М.В. Соловей

Фонд оценочных средств рассмотрен и одобрен методической комиссией института цифровых технологий (протокол №5 от 29.08.2024 г).

Председатель методической комиссии



О.С. Витренко