

**Министерство науки и высшего образования РФ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Национальный исследовательский университет «МЭИ»**

Направление подготовки/специальность: 02.03.02 Фундаментальная информатика и информационные технологии

Наименование образовательной программы: Информационные технологии и системы искусственного интеллекта

Уровень образования: высшее образование - бакалавриат

Форма обучения: Очная

**Оценочные материалы
по дисциплине
Методы и средства анализа данных**

**Москва
2024**

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ РАЗРАБОТАЛ:

Разработчик

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Алексеев Н.П.
	Идентификатор	Rd05f8b92-AlexeevNP-53805acf

Н.П.
Алексеев

СОГЛАСОВАНО:

Руководитель
образовательной
программы

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Чернецов А.М.
	Идентификатор	Re594826f-ChernetsovAM-0080e09

А.М. Чернецов

Заведующий
выпускающей
кафедрой

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Варшавский П.Р.
	Идентификатор	R9a563c96-VarshavskyPR-efb4bbd

П.Р.
Варшавский

ОБЩАЯ ЧАСТЬ

Оценочные материалы по дисциплине предназначены для оценки достижения обучающимися запланированных результатов обучения по дисциплине, этапа формирования запланированных компетенций и уровня освоения дисциплины.

Оценочные материалы по дисциплине включают оценочные средства для проведения мероприятий текущего контроля успеваемости и промежуточной аттестации.

Формируемые у обучающегося компетенции:

1. ОПК-2 Способен применять компьютерные/суперкомпьютерные методы, современное программное обеспечение (в том числе отечественного производства) для решения задач профессиональной деятельности

ИД-3 Анализирует результаты численного решения задач и оценивает необходимые для выполнения работы ресурсы

2. ОПК-5 Способен устанавливать и сопровождать программное обеспечение информационных систем и баз данных, в том числе отечественного происхождения, с учетом информационной безопасности

ИД-1 Использует техническое сопровождение информационных систем и баз данных

и включает:

для текущего контроля успеваемости:

Форма реализации: Компьютерное задание

1. Классификация текстов. Байесовский классификатор. (Лабораторная работа)
2. Машина опорных векторов (SVM) (Лабораторная работа)
3. Реализация алгоритм CART для построения дерева решений (Лабораторная работа)
4. Реализация алгоритма DBSCAN для решения задачи кластеризации (Лабораторная работа)
5. Реализация алгоритма K-means для решения задачи кластеризации (Лабораторная работа)

БРС дисциплины

8 семестр

Перечень контрольных мероприятий текущего контроля успеваемости по дисциплине:

- КМ-1 Реализация алгоритма K-means для решения задачи кластеризации (Лабораторная работа)
- КМ-2 Реализация алгоритма DBSCAN для решения задачи кластеризации (Лабораторная работа)
- КМ-3 Реализация алгоритм CART для построения дерева решений (Лабораторная работа)
- КМ-4 Классификация текстов. Байесовский классификатор. (Лабораторная работа)
- КМ-5 Машина опорных векторов (SVM) (Лабораторная работа)

Вид промежуточной аттестации – Экзамен.

Раздел дисциплины	Веса контрольных мероприятий, %				
	Индекс КМ:	КМ-1	КМ-2	КМ-3	КМ-4

	Срок КМ:	3	6	9	12	16
Интеллектуальный анализ данных (Data Mining). Этапы исследования данных с помощью методов Data Mining. Типы закономерностей. Методы Data Mining.						
Интеллектуальный анализ данных (Data Mining). Этапы исследования данных с помощью методов Data Mining. Типы закономерностей. Методы Data Mining.	+	+		+	+	
Задачи Data Mining. Задача кластеризации. Смесь нормальных распределений. EM-алгоритм. K-means и его модификации.						
Задачи Data Mining. Задача кластеризации. Смесь нормальных распределений. EM-алгоритм. K-means и его модификации.	+					
Байесовская классификация. Иерархическая классификация. Расстояние между кластерами. Алгоритм DBSCAN.						
Байесовская классификация. Иерархическая классификация. Расстояние между кластерами. Алгоритм DBSCAN.		+				
Задачи классификации и регрессии. Деревья решений. Алгоритм CART, другие алгоритмы.						
Задачи классификации и регрессии. Деревья решений. Алгоритм CART, другие алгоритмы.			+			
Классификация текстов.						
Классификация текстов.				+		
Системы аналитической обработки данных. Линейные модели для классификации и регрессии. Машина опорных векторов.						
Системы аналитической обработки данных. Линейные модели для классификации и регрессии. Машина опорных векторов.						+
Вес КМ:	20	20	20	20	20	20

СОДЕРЖАНИЕ ОЦЕНОЧНЫХ СРЕДСТВ ТЕКУЩЕГО КОНТРОЛЯ

I. Оценочные средства для оценки запланированных результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

Индекс компетенции	Индикатор	Запланированные результаты обучения по дисциплине	Контрольная точка
ОПК-2	ИД-3 _{ОПК-2} Анализирует результаты численного решения задач и оценивает необходимые для выполнения работы ресурсы	Знать: методы системного анализа алгоритмы анализа больших объёмов информации основные подходы к реализации методов искусственного интеллекта для анализа данных Уметь: реализовать методы системного анализа реализовать алгоритмы анализа больших объёмов данных современные платформы и языки программирования высокого уровня	КМ-1 Реализация алгоритма K-means для решения задачи кластеризации (Лабораторная работа) КМ-2 Реализация алгоритма DBSCAN для решения задачи кластеризации (Лабораторная работа) КМ-3 Реализация алгоритм CART для построения дерева решений (Лабораторная работа) КМ-4 Классификация текстов. Байесовский классификатор. (Лабораторная работа) КМ-5 Машина опорных векторов (SVM) (Лабораторная работа)
ОПК-5	ИД-1 _{ОПК-5} Использует техническое сопровождение информационных систем и баз данных	Знать: этапы жизненного цикла информационных систем подходы к тестированию информационных систем Уметь:	КМ-2 Реализация алгоритма DBSCAN для решения задачи кластеризации (Лабораторная работа) КМ-4 Классификация текстов. Байесовский классификатор. (Лабораторная работа) КМ-5 Машина опорных векторов (SVM) (Лабораторная работа)

		осуществлять поддержку разработки информационных систем выполнять доработку модулей информационных систем	
--	--	--	--

II. Содержание оценочных средств. Шкала и критерии оценивания

КМ-1. Реализация алгоритма K-means для решения задачи кластеризации

Формы реализации: Компьютерное задание

Тип контрольного мероприятия: Лабораторная работа

Вес контрольного мероприятия в БРС: 20

Процедура проведения контрольного мероприятия: Лабораторная работа на компьютере.

Краткое содержание задания:

Разработать и реализовать алгоритм K-means для решения задачи классификации.
Выполнить тестирование разработанного приложения и написать вывод о работе алгоритма.

Контрольные вопросы/задания:

Запланированные результаты обучения по дисциплине	Вопросы/задания для проверки
Знать: алгоритмы анализа больших объёмов информации	1.Описание метода K-means
Уметь: современные платформы и языки программирования высокого уровня	1.Реализовать метод K-means 2.Обосновать выбор количества кластеров

Описание шкалы оценивания:

Оценка: «зачтено»

Описание характеристики выполнения знания: Студент выполнил лабораторную работу

Оценка: «не зачтено»

Описание характеристики выполнения знания: Студент не выполнил лабораторную работу

КМ-2. Реализация алгоритма DBSCAN для решения задачи кластеризации

Формы реализации: Компьютерное задание

Тип контрольного мероприятия: Лабораторная работа

Вес контрольного мероприятия в БРС: 20

Процедура проведения контрольного мероприятия: Лабораторная работа на компьютере.

Краткое содержание задания:

Разработать и реализовать алгоритм K-means для решения задачи классификации.
Выполнить тестирование разработанного приложения и написать вывод о работе алгоритма.

Контрольные вопросы/задания:

Запланированные результаты обучения по дисциплине	Вопросы/задания для проверки
Знать: основные подходы к реализации методов искусственного интеллекта для анализа данных	1.Описание алгоритма DBSCAN
Знать: этапы жизненного цикла информационных систем	1.Перечислить основные достоинства и недостатки алгоритма DBSCAN
Уметь: выполнять доработку модулей	1.Реализация алгоритма DBSCAN

Запланированные результаты обучения по дисциплине	Вопросы/задания для проверки
информационных систем	

Описание шкалы оценивания:

Оценка: «зачтено»

Описание характеристики выполнения знания: Студент выполнил лабораторную работу

Оценка: «не зачтено»

Описание характеристики выполнения знания: Студент не выполнил лабораторную работу

КМ-3. Реализация алгоритм CART для построения дерева решений

Формы реализации: Компьютерное задание

Тип контрольного мероприятия: Лабораторная работа

Вес контрольного мероприятия в БРС: 20

Процедура проведения контрольного мероприятия: Лабораторная работа на компьютере.

Краткое содержание задания:

Разработать и реализовать алгоритм CART для решения задачи классификации.

Выполнить тестирование разработанного приложения и написать вывод о работе алгоритма.

Контрольные вопросы/задания:

Запланированные результаты обучения по дисциплине	Вопросы/задания для проверки
Уметь: реализовать методы системного анализа	1.Реализация алгоритма CART 2.Привести пример изменения построенного дерева решений при удалении записей из обучающего множества

Описание шкалы оценивания:

Оценка: «зачтено»

Описание характеристики выполнения знания: Студент выполнил лабораторную работу

Оценка: «не зачтено»

Описание характеристики выполнения знания: Студент не выполнил лабораторную работу

КМ-4. Классификация текстов. Байесовский классификатор.

Формы реализации: Компьютерное задание

Тип контрольного мероприятия: Лабораторная работа

Вес контрольного мероприятия в БРС: 20

Процедура проведения контрольного мероприятия: Лабораторная работа на компьютере.

Краткое содержание задания:

Реализовать байесовский наивный классификатор. Выполнить тестирование разработанного приложения и написать вывод о работе алгоритма.

Контрольные вопросы/задания:

Запланированные результаты обучения по дисциплине	Вопросы/задания для проверки
Знать: подходы к тестированию	1.Описание байесовского наивного классификатора

Запланированные результаты обучения по дисциплине информационных систем	Вопросы/задания для проверки
Уметь: реализовать алгоритмы анализа больших объёмов данных	1.Реализация байесовского наивного классификатора 2.Привести пример вычисления условной вероятности объекта согласно алгоритму наивного байесовского классификатора

Описание шкалы оценивания:

Оценка: «зачтено»

Описание характеристики выполнения знания: Студент выполнил лабораторную работу

Оценка: «не зачтено»

Описание характеристики выполнения знания: Студент не выполнил лабораторную работу

КМ-5. Машина опорных векторов (SVM)

Формы реализации: Компьютерное задание

Тип контрольного мероприятия: Лабораторная работа

Вес контрольного мероприятия в БРС: 20

Процедура проведения контрольного мероприятия: Лабораторная работа на компьютере.

Краткое содержание задания:

Реализовать алгоритм SVM. Выполнить тестирование разработанного приложения и написать вывод о работе алгоритма.

Контрольные вопросы/задания:

Запланированные результаты обучения по дисциплине	Вопросы/задания для проверки
Знать: методы системного анализа	1.Описание алгоритма SVM
Знать: основные подходы к реализации методов искусственного интеллекта для анализа данных	1.Перечислить основные достоинства и недостатки алгоритма SVM
Уметь: осуществлять поддержку разработки информационных систем	1.Реализация алгоритма SVM

Описание шкалы оценивания:

Оценка: «зачтено»

Описание характеристики выполнения знания: Студент выполнил лабораторную работу

Оценка: «не зачтено»

Описание характеристики выполнения знания: Студент не выполнил лабораторную работу

СОДЕРЖАНИЕ ОЦЕНОЧНЫХ СРЕДСТВ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

8 семестр

Форма промежуточной аттестации: Экзамен

Пример билета

1. Основы регрессионного анализа. Понятие «регрессия». Уравнение регрессии. Интерпретация параметров регрессии. Проверка модели регрессии.
2. Основные алгоритмы Data-mining. Метод k-средних.

Процедура проведения

Студенту даётся 45 минут на подготовку ответа на билет. Затем он отвечает преподавателю. Преподаватель задаёт дополнительные вопросы. На основе оценки качества и подробности изложения ответа студента выставляется оценка за экзамен.

1. Перечень компетенций/индикаторов и контрольных вопросов проверки результатов освоения дисциплины

1. Компетенция/Индикатор: ИД-Зопк-2 Анализирует результаты численного решения задач и оценивает необходимые для выполнения работы ресурсы

Вопросы, задания

- 1.1. Введение в «анализ данных». Работа с данными. Этапы решения задачи анализа данных и их взаимосвязи
- 2.3. Задачи Data Mining. Задача кластеризации. Смесь нормальных распределений. EM-алгоритм. K-means и его модификации.
- 3.4. Байесовская классификация. Maximum a posteriori. Иерархическая классификация (agglomerative и divisive подходы). Расстояние между кластерами. DBSCAN. Выбор количества кластеров.
- 4.5. Основы регрессионного анализа. Понятие регрессия. Простая линейная взаимосвязь. Уравнение регрессии. Подгонка линии регрессии. Понятие корреляции и ковариации
- 5.6. Задачи классификации и регрессии. Подходы к моделированию. Виды моделей: генеративные модели, дискриминативные модели, функции решения. Функции потерь. Переобучение. «Проклятие размерности». Теория принятия решений. Оценка результатов классификации: разделение выборки, скользящий контроль, bootstrap. Метрики качества. ROC. Деревья решений. Алгоритм CART, другие алгоритмы.
- 6.8. Линейные модели для классификации и регрессии. Линейная регрессия. ML-функция правдоподобия. Регуляризация. Логистическая регрессия. Обобщённые линейные модели. Градиентный спуск и метод Ньютона.
- 7.9. Машина опорных векторов (SVM). Функции ядра.

Материалы для проверки остаточных знаний

- 1.1. Задачи и процесс Data Mining.

Ответы:

Изложение указанных алгоритмов и подходов МиСАД

Верный ответ: Корректное изложение указанных алгоритмов и подходов МиСАД

- 2.4. Задача кластеризации и классификации

Ответы:

Изложение указанных алгоритмов и подходов МиСАД

Верный ответ: Корректное изложение указанных алгоритмов и подходов МиСАД

3.5.1. Методы Data Mining. Методы построения правил классификации

Ответы:

Изложение указанных алгоритмов и подходов МиСАД

Верный ответ: Корректное изложение указанных алгоритмов и подходов МиСАД

4.5.2. Методы построения деревьев решений. Алгоритм CART.

Ответы:

Изложение указанных алгоритмов и подходов МиСАД

Верный ответ: Корректное изложение указанных алгоритмов и подходов МиСАД

5.5.4. Линейные модели классификации и регрессии

Ответы:

Изложение указанных алгоритмов и подходов МиСАД

Верный ответ: Корректное изложение указанных алгоритмов и подходов МиСАД

6.5.4. Метод опорных векторов

Ответы:

Изложение указанных алгоритмов и подходов МиСАД

Верный ответ: Корректное изложение указанных алгоритмов и подходов МиСАД

7.5.5. Алгоритм k-ближайших соседей

Ответы:

Изложение указанных алгоритмов и подходов МиСАД

Верный ответ: Корректное изложение указанных алгоритмов и подходов МиСАД

8.5.6. Алгоритм Apriori

Ответы:

Изложение указанных алгоритмов и подходов МиСАД

Верный ответ: Корректное изложение указанных алгоритмов и подходов МиСАД

9.5.7. EM-алгоритм

Ответы:

Изложение указанных алгоритмов и подходов МиСАД

Верный ответ: Корректное изложение указанных алгоритмов и подходов МиСАД

10.5.9. Алгоритм усиления классификаторов AdaBoost

Ответы:

Изложение указанных алгоритмов и подходов МиСАД

Верный ответ: Корректное изложение указанных алгоритмов и подходов МиСАД

2. Компетенция/Индикатор: ИД-1_{ОПК-5} Использует техническое сопровождение информационных систем и баз данных

Вопросы, задания

1.2. Интеллектуальный анализ данных (Data Mining). Этапы исследования данных с помощью методов Data Mining. Типы закономерностей. Методы Data Mining.

2.7. Классификация текстов. Обработка текстов и Naïve Bayes. Задачи Text Mining. Этапы обработки текстов. Закон Ципфа. Представление документов. Байесовский классификатор.

3.10. Системы аналитической обработки данных. CRM – технология. ERP – системы. OLAP – технология

Материалы для проверки остаточных знаний

1.2. Этапы исследования данных с помощью методов Data Mining

Ответы:

Изложение указанных алгоритмов и подходов МиСАД

Верный ответ: Корректное изложение указанных алгоритмов и подходов МиСАД

2.3. Типы закономерностей в данных

Ответы:

Изложение указанных алгоритмов и подходов МиСАД

Верный ответ: Корректное изложение указанных алгоритмов и подходов МиСАД
3.5.3. Наивный байесовский классификатор

Ответы:

Изложение указанных алгоритмов и подходов МиСАД

Верный ответ: Корректное изложение указанных алгоритмов и подходов МиСАД

4.5.8. Алгоритм ссылочного ранжирования PageRank

Ответы:

Изложение указанных алгоритмов и подходов МиСАД

Верный ответ: Корректное изложение указанных алгоритмов и подходов МиСАД

5.6. Системы аналитической обработки данных. CRM – технология. ERP – системы.
OLAP – технология.

Ответы:

Изложение указанных алгоритмов и подходов МиСАД

Верный ответ: Корректное изложение указанных алгоритмов и подходов МиСАД

II. Описание шкалы оценивания

Оценка: 5 («отлично»)

Нижний порог выполнения задания в процентах: 70

Описание характеристики выполнения знания: Работа выполнена в рамках "продвинутого" уровня. Ответы даны верно, четко сформулированные особенности практических решений

Оценка: 4 («хорошо»)

Нижний порог выполнения задания в процентах: 60

Описание характеристики выполнения знания: Работа выполнена в рамках "базового" уровня. Большинство ответов даны верно. В части материала есть незначительные недостатки

Оценка: 3 («удовлетворительно»)

Нижний порог выполнения задания в процентах: 50

Описание характеристики выполнения знания: Работа выполнена в рамках "порогового" уровня. Основная часть задания выполнена верно. на вопросы углубленного уровня

Оценка: 2 («неудовлетворительно»)

Описание характеристики выполнения знания: Работа не выполнена или выполнена преимущественно неправильно

III. Правила выставления итоговой оценки по курсу

Для получения итоговой оценки баллы промежуточной аттестации суммируются с балами текущей аттестации.